Chapter 6
Unipotent Dynamics and Ratner's Theorems

In this chapter we discuss unipotent dynamics and prove several special cases of Ratner's theorems. We will not discuss the history in detail, and refer to the survey papers of Kleinbock, Shah and Starkov [88], Ratner [138], Margulis [111], 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.

6.1 Unipotent Invariance

Unipotent dynamics and Ratner's theorem (as discussed in greater detail in the next section) are of great importance for various applications. To see this we explain how unipotent invariance may arise naturally in applications of various types.

6.1.1 Factor Rigidity

For simplicity we let $G = \mathrm{SL}_2(\mathbb{R})$ or $G = \mathrm{SL}_3(\mathbb{R})$. Let $\Gamma_1, \Gamma_2 < G$ be lattices and write $X_j = G/\Gamma_j$ for $j = 1, 2$. Moreover, let $H = \{u_t \mid t \in \mathbb{R}\}$ be a one-parameter subgroup.

Definition 6.1. A *factor map* from the action of H on X_1 to the action of H on X_2 is a measurable almost everywhere defined map $\phi\colon X_1 \to X_2$ such that

$$\phi(u_t \cdot x) = u_t \cdot \phi(x) \tag{6.1}$$

for every $t \in \mathbb{R}$ and almost every $x \in X_1$ and $\phi_* m_{X_1} = m_{X_2}$. Moreover, ϕ is called an *isomorphism* if there exists a measurable set $X_1' \subseteq X_1$ of full measure such that $\phi|_{X_1'}$ is injective.

In general factor maps or isomorphisms are potentially indeed only measurable, of fractal nature, and may not respect (for example) topological dimension. However, if $H = U$ is a unipotent subgroup then factor maps are essentially algebraic, and in particular are smooth maps. That is, the equivariance condition in (6.1) along one direction of G forces good behaviour along all directions of G. This is quite surprising and is an instance of a 'rigidity' phenomenon. The case of $G = \mathrm{SL}_2(\mathbb{R})$ is due to earlier work of Ratner [131] and more general cases have been obtained by Witte Morris [172] using similar arguments.

Theorem 6.2 (Factor rigidity). *Let X_1, X_2 be as above, let U be a one-parameter unipotent subgroup, and let $\phi\colon X_1 \to X_2$ be a factor map for U. Then there exists an automorphism $\varphi\colon G \to G$ whose restriction to U is the identity map and some element $p \in G$ so that*

$$\phi(g\Gamma_1) = \varphi(g)p\Gamma_2$$

for m_{X_1}-almost every $g\Gamma_1$. In particular, the lattices are related by

$$\varphi(\Gamma_1) \subseteq p\Gamma_2 p^{-1}$$

and the factor map ϕ is an isomorphism if and only if Γ_1 and Γ_2 have the same covolume.

The proof of Theorem 6.2 starts by encoding the map ϕ in terms of the probability measure

$$\mu = \mu_\phi = (\mathrm{id}, \phi)_* m_{X_1}$$

on $X_1 \times X_2$. We note that μ is concentrated[†] on $\mathrm{Graph}(\phi)$, meaning that

$$\mu\bigl(\mathrm{Graph}(\phi)\bigr) = 1.$$

Moreover, the equivariance of ϕ in (6.1) shows that μ is invariant under the action of (u_t, u_t) on $X_1 \times X_2$ for $u_t \in U$. Ergodicity of this action follows as it is measurably isomorphic to the action of U on X_1 with respect to m_{X_1}. The projection of μ to X_2 is $\phi_*(m_{X_1}) = m_{X_2}$.

Theorem 6.3 (Ratner's joining classification). *Let μ be a probability measure on $X_1 \times X_2$ that projects to m_{X_1} and m_{X_2} under the coordinate projections and that is invariant and ergodic under the action of (u_t, u_t) for $u_t \in U$. Then μ is the L-invariant probability measure on a closed L-orbit for a closed connected subgroup $L < G \times G$. Moreover, either $L = G \times G$ or L is the graph of an automorphism $\varphi\colon G \to G$.*

We will prove Theorems 6.2 and 6.3 in this chapter and use the latter to further understand all possible 'abstract' factors of U acting on X_1.

Exercise 6.4. Try to prove Theorem 6.3 using Theorem 6.2 as a black box.

[†] We avoid saying 'supported' as we do not know whether $\mathrm{Graph}(\phi)$ is closed.

6.1.2 Oppenheim's Conjecture

We note that for a linear form L in $d \geqslant 2$ variables it is easy to determine whether $L(\mathbb{Z}^d) \subseteq \mathbb{R}$ is dense or not. In fact $L(\mathbb{Z}^d)$ is dense in $\mathbb{R}$ if and only if L is not a multiple of a form with rational coefficients. As a generalization Oppenheim [126] conjectured in 1929 that a non-degenerate indefinite quadratic form Q in $d \geqslant 5$ variables that is not a multiple of a form with integer coefficients has $Q(\mathbb{Z}^d)$ dense in $\mathbb{R}$.

Raghunathan [130] noticed(32) in the mid 1970s the connection to homogeneous dynamics and, motivated by this, formulated far-reaching conjectures concerning orbit closures for subgroups generated by unipotent subgroups. Margulis developed these ideas to prove the Oppenheim conjecture in the following stronger form in 1986 [112].

Theorem 6.5 (Margulis' solution of Oppenheim conjecture). *Let Q be a non-degenerate indefinite quadratic form in $d \geqslant 3$ variables that is not a multiple of a form with integer coefficients. Then $Q(\mathbb{Z}^d_{\mathrm{prim}})$ is dense in $\mathbb{R}$, where*

$$\mathbb{Z}^d_{\mathrm{prim}} = \{v = (v_1, \ldots, v_d)^{\mathrm{t}} \in \mathbb{Z}^d \mid \gcd(v_1, \ldots, v_d) = 1\}.$$

We will prove Oppenheim's conjecture later, but for now let us point out that it also follows quickly from the following dynamical result.

Theorem* 6.6 (Orbit closure by Dani–Margulis). Let

$$Q_0(x_1, x_2, x_3) = x_2^2 - 2x_1x_3$$

and $H = \mathrm{SO}_{Q_0}(\mathbb{R})^o$. For any $x_0 \in \mathsf{X}_3$ either $H\boldsymbol{\cdot}x_0$ is closed or $H\boldsymbol{\cdot}x_0$ is dense in X_3.

We note that an orthogonal group in 3-dimensions is 3-dimensional. Moreover, $\mathrm{SO}_{2,1}(\mathbb{R})$ is locally isomorphic to $\mathrm{SL}_2(\mathbb{R})$. To see this† consider the adjoint representation of $g \in \mathrm{SL}_2(\mathbb{R})$ on the Lie algebra $\mathfrak{sl}_2(\mathbb{R})$ by $v \mapsto gvg^{-1}$ and the indefinite quadratic form $\det v = -a^2 - bc$ for $v = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}$. This shows that $H \cong \mathrm{PSL}_2(\mathbb{R})$ is simple and generated by unipotent one-parameter subgroups.

We have chosen the above quadratic form Q_0 so that its corresponding orthogonal group SO_{Q_0} is easy to work with. For instance, $\mathrm{SO}_{Q_0}(\mathbb{R})$ contains the diagonal matrices of the form $\mathrm{diag}(a, 1, a^{-1})$ for $a \in \mathbb{R}^\times$. Moreover, it contains the unipotent one-parameter subgroup

$$\left\{ \begin{pmatrix} 1 & t & \frac{t^2}{2} \\ & 1 & t \\ & & 1 \end{pmatrix} \,\middle|\, t \in \mathbb{R} \right\} \leqslant \mathrm{SO}_{Q_0}(\mathbb{R}).$$

† This is an instance of a low-dimensional accident: In essence there is only one real non-compact simple Lie algebra of dimension 3 up to isomorphism.

To see this we calculate

$$\begin{aligned}
Q_0\left(\begin{pmatrix}1 & t & \frac{t^2}{2}\\ & 1 & t\\ & & 1\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}\right) &= Q_0\left(\begin{pmatrix}x_1+tx_2+\frac{t^2}{2}x_3\\ x_2+tx_3\\ x_3\end{pmatrix}\right)\\
&= (x_2+tx_3)^2-2\bigl(x_1+tx_2+\tfrac12 t^2x_3\bigr)x_3\\
&= x_2^2+2tx_2x_3+t^2x_3^2-2x_1x_3-2tx_2x_3-t^2x_3^2\\
&= x_2^2-2x_1x_3. \qquad (6.2)
\end{aligned}$$

Similarly, we also have

$$\begin{pmatrix}1 & & \\ t & 1 & \\ \frac{t^2}{2} & t & 1\end{pmatrix}\in \mathrm{SO}_{Q_0}(\mathbb{R})$$

for $t\in\mathbb{R}$. Finally we note that the Lie algebra elements corresponding to the diagonal group and these two one-parameter unipotent subgroups can be chosen to match the $\mathfrak{sl}_2$-triple (2.6) on page 65.

Proof that Theorem 6.6 implies Theorem 6.5 for $d=3$. Suppose first that Q is a non-degenerate indefinite quadratic form in $d=3$ variables and is not a multiple of a form with integer coefficients. By our discussion of signatures of quadratic forms in Theorem 3.5 it follows that there exist $\lambda\in\mathbb{R}^\times$ and $g_0\in\mathrm{SL}_3(\mathbb{R})$ so that $Q=\lambda Q_0\circ g_0$. In particular, $Q(\mathbb{Z}^3_{\mathrm{prim}})=\lambda Q_0(g_0\mathbb{Z}^3_{\mathrm{prim}})$. We define $x_0=g_0\mathbb{Z}^3\in\mathsf{X}_3$, $H=\mathrm{SO}_{Q_0}(\mathbb{R})^o$, and consider the two cases in Theorem 6.6.

Density implies Density. Suppose that the orbit $H\raisebox{0.2ex}{.}x_0$ is dense in X_d. Fix some $a\in\mathbb{R}$ and set $g=\begin{pmatrix}1&0&0\\0&1&0\\-a&0&1\end{pmatrix}$ so that $a\in Q_0(g\mathbb{Z}^3_{\mathrm{prim}})$. If now $h\in H$ is such that $h\raisebox{0.2ex}{.}x_0$ is very close to x then the values in $\lambda Q_0(g_0\mathbb{Z}^3_{\mathrm{prim}})=\lambda Q_0(hg_0\mathbb{Z}^3_{\mathrm{prim}})$ can be used to approximate $\lambda a=\lambda Q_0(ge_1)$ arbitrarily well. In other words $Q(\mathbb{Z}^3_{\mathrm{prim}})$ is dense in $\mathbb{R}$.

Closed implies Rationality. Suppose now that the orbit $H\raisebox{0.2ex}{.}x_0$ is closed. As $\mathrm{SO}_{2,1}(\mathbb{R})^o\cong\mathrm{PSL}_2(\mathbb{R})$ is non-compact and simple, Theorem 4.18 applies and shows that $H\raisebox{0.2ex}{.}x$ actually has finite volume. By Borel density (Theorem 3.50) it follows that $g_0^{-1}Hg_0\cap\mathrm{SL}_d(\mathbb{Z})$ is Zariski dense in $g_0^{-1}\,\mathrm{SO}_{Q_0}\,g_0=\mathrm{SO}_Q$. We now show that this implies that a multiple of Q has integer coefficients, giving a contradiction to our assumptions.

We start by showing that the linear hull $\langle Q_0\rangle$ of $Q_0((x_1,x_2,x_3)^{\mathrm{t}})=x_2^2-2x_1x_3$ is the subspace of all quadratic forms q satisfying $q\circ h=q$ for all $h\in H$. Indeed for $a=\mathrm{diag}(\mathrm{e},1,\mathrm{e}^{-1})\in H$ every quadratic monomial is an eigenvector but only x_1x_3 and x_2^2 have eigenvalue 1. Hence $q\circ a=q$ implies that $q((x_1,x_2,x_3)^{\mathrm{t}})=\alpha x_2^2+\beta x_1x_3$ for two scalars α and β. Going through the calculation in (6.2) that H contains the unipotent elements u_t again, we see that the required cancellation of term in the expression $q \circ u_t$ only happens if $\beta = -2\alpha$.

By conjugation with g_0 we also see that the linear hull $\langle Q \rangle$ of Q is the subspace of all quadratic forms q satisfying $q \circ h = q$ for all $h \in \mathrm{SO}_Q(\mathbb{R})^o$, or equivalently for all $h \in \mathrm{SO}_Q(\mathbb{R})^o \cap \mathrm{SL}_3(\mathbb{Z})$. As the latter are rational equations that have a nontrivial solution, we obtain that a multiple of Q has integer coefficients. $\square$

One may wonder why it might be advantageous to study three-dimensional orbits inside the eight-dimensional space X_3 in Theorem 6.6 to prove a mere density statement in $\mathbb{R}$ as in Theorem 6.5. As a partial answer to this we note that the set $Q(\mathbb{Z}^3) \subseteq \mathbb{R}$ has very little structure and, in particular, has no invariance properties. However, in the eight-dimensional X_3 applying the orthogonal group to a lattice does not change the values of the quadratic form. Moreover, as the quadratic form is indefinite we obtain in this way the powerful tool of unipotent invariance.

Exercise 6.7. Show that Oppenheim's conjecture cannot hold for binary quadratic forms (that is, forms in two dimensions). For this let $\alpha \in \mathbb{R}$ be badly approximable with $\alpha^2 \notin \mathbb{Q}$ and consider the form $x^2 - \alpha^2 y^2 = (x - \alpha y)(x + \alpha y)$.

6.1.3 Distorted Orbits

In Sections 5.4–5.7 we have seen the importance of studying 'distorted orbits' of the form $gH{\boldsymbol\cdot}\Gamma$ while varying $gH \in G/H$. We wish to explain—under suitable assumptions—why weak* limits of the Haar measures on such distorted orbits often have unipotent invariance.

Lemma 6.8 (Unipotent invariance for limits of distorted orbits). *Let G be a closed linear group, $\Gamma < G$ a lattice, and $H < G$ a closed subgroup with Lie algebra $\mathfrak{h} = \mathrm{Lie}\, H$ so that $\Gamma \cap H < H$ is also a lattice. We assume moreover that (a_n) is a sequence in G with*

$$\lim_{n\to\infty} \big\| \mathrm{Ad}_{a_n}|_{\mathfrak{h}} \big\| = \infty. \tag{6.3}$$

Then any weak limit of the Haar measures on $a_n H{\boldsymbol\cdot}\Gamma$ inside $X = G/\Gamma$ is invariant under a one-parameter unipotent subgroup.*

Proof. We assume without loss of generality that the Haar measures $m_{a_n H{\boldsymbol\cdot}\Gamma}$ on $a_n H{\boldsymbol\cdot}\Gamma$ converge in the weak* topology to a measure μ. By the assumption (6.3) there exists a sequence (v_n) in $\mathfrak{h}$ so that $v_n \to 0$ but $\|\mathrm{Ad}_{a_n} v_n\| = 1$ for all $n \geqslant 1$. By choosing a subsequence once more we may assume that $\mathrm{Ad}_{a_n} v_n$ converges to an element w in the Lie algebra $\mathfrak{g}$ of G. Let $f \in C_c(X)$ and $t \in \mathbb{R}$. Then

$$\int f\big(\exp(tw){\boldsymbol\cdot}x\big)\,\mathrm{d}\mu(x) = \lim_{n\to\infty} \int f\big(\exp(tw){\boldsymbol\cdot}x\big)\,\mathrm{d}m_{a_n H{\boldsymbol\cdot}\Gamma}(x)$$

by definition. Moreover, $\mathsf{d}\bigl(\exp(tw), a_n \exp(tv_n) a_n^{-1}\bigr) \to 0$ as $n \to \infty$. Hence uniform continuity of f allows us to replace $\exp(tw)$ by $a_n \exp(tv_n) a_n^{-1}$, giving

$$\int f\bigl(\exp(tw)\boldsymbol{\cdot} x\bigr) \,\mathrm{d}\mu(x) = \lim_{n\to\infty} \int f\bigl(\underbrace{a_n \exp(tv_n) a_n^{-1}}_{\in a_n H a_n^{-1}} \boldsymbol{\cdot} x\bigr) \,\mathrm{d}m_{a_n H\boldsymbol{\cdot}\Gamma}(x)$$
$$= \lim_{n\to\infty} \int f \,\mathrm{d}m_{a_n H\boldsymbol{\cdot}\Gamma} = \int f \,\mathrm{d}\mu,$$

where we have used the fact that $m_{a_n H\boldsymbol{\cdot}\Gamma}$ is invariant under $a_n H a_n^{-1}$. As the function $f \in C_c(X)$ and $t \in \mathbb{R}$ were arbitrary we deduce that μ is invariant under the one-parameter subgroup determined by w. Finally note that the eigenvalues of $w = \lim_{n\to\infty} \mathrm{Ad}_{a_n} v_n$ vanish as $\lim_{n\to\infty} v_n = 0$ and the conjugation Ad_{a_n} does not change eigenvalues. □

We note that Lemma 6.8 and the powerful theorems due to Ratner from the next section can be used for proving equidistribution (and counting) results in situations where the banana mixing argument from Chapter 5 does not apply.

6.1.4 Orbits Arising From Expanding Curves

We wish to explain another way in which unipotent invariance may arise. For this we suppose that $\mathcal{I} \subseteq \mathbb{R}$ is a compact interval and that $\gamma\colon \mathcal{I} \to \mathbb{R}^d$ has continuous second derivative. We also assume that $\gamma'(s) \neq 1$ for $s \in \mathcal{I}$ and that $\mathcal{I} \ni s \mapsto g_s \in \mathrm{GL}_d(\mathbb{R})$ is continuous so that $g_s \gamma'(s) = \mathbf{e}_1$ for all $s \in \mathcal{I}$. We identify $g \in \mathrm{GL}_d(\mathbb{R})$ with

$$\begin{pmatrix} (\det g)^{-1} & \\ & g \end{pmatrix} \in \mathrm{SL}_{d+1}(\mathbb{R})$$

to simplify the notation. For $v \in \mathbb{R}^d$ and $t \in \mathbb{R}$ we also define $u_v = \begin{pmatrix} 1 & \\ v & I_d \end{pmatrix}$ and $a_t = \begin{pmatrix} t^{-d} & \\ & tI_d \end{pmatrix}$.

Lemma 6.9 (Twisting trick, first step). *Let $x_0 \in \mathsf{X}_{d+1}$. Using the above assumptions and notation, any weak* limit of*

$$\frac{1}{|\mathcal{I}|} \int_{\mathcal{I}} \delta_{a_t g_s u_{\gamma(s)} x_0} \,\mathrm{d}s \tag{6.4}$$

for $t \to \infty$ is invariant under the one-parameter subgroup $U = \{u_{r\mathbf{e}_1} \mid r \in \mathbb{R}\}$.

Proof. Let $f \in C_c(\mathsf{X}_{d+1})$, $r \in \mathbb{R}$, and $t \geqslant 0$. We define $\kappa = 1 + \frac{1}{d}$ and note that $a_t u_v a_t^{-1} = u_{\mathrm{e}^{-\kappa t} v}$ for $v \in \mathbb{R}^d$. Then we have

$$\int_{\mathcal{I}} f(u_{r\mathbf{e}_1} a_t g_s u_{\gamma(s)} x_0)\,\mathrm{d}s = \int_{\mathcal{I}} f(a_t u_{\mathrm{e}^{-\kappa t} r\mathbf{e}_1} g_s u_{\gamma(s)} x_0)\,\mathrm{d}s$$
$$= \int_{\mathcal{I}} f(a_t g_s u_{\mathrm{e}^{-\kappa t} r\gamma'(s)+\gamma(s)} x_0)\,\mathrm{d}s$$

by the way a_t and g_s interact with $u_{\mathbf{e}_1}$. Fixing r and thinking of a large $t > 0$, we have

$$\gamma(s) + \mathrm{e}^{-\kappa t} r\gamma'(s) = \gamma\left(s + \mathrm{e}^{-\kappa t} r\right) + \varepsilon(s)$$

for an error term $\varepsilon(s) = \varepsilon_{r,t}(s) = \mathrm{O}\big(\mathrm{e}^{-2\kappa t}\big)$ as γ is assumed to be twice continuously differentiable. Let

$$\mathcal{I}' = \mathcal{I} \cap \left(\mathcal{I} + \mathrm{e}^{-2\kappa t} r\right)$$

which (for fixed $r \in \mathbb{R}$ and large $t > 0$) is basically equal to $\mathcal{I}$. We let

$$s' = s + \mathrm{e}^{-\kappa t} r$$

for $s \in \mathcal{I}$, and obtain

$$\int_{\mathcal{I}} f(u_{r\mathbf{e}_1} a_t g_s u_{\gamma(s)} x_0)\,\mathrm{d}s = \int_{\mathcal{I}'} f(a_t g_s u_{\gamma(s')+\varepsilon(s)} x_0)\,\mathrm{d}s' + \mathrm{O}\big(r\mathrm{e}^{-\kappa t}\big).$$

We now take the error term $\varepsilon(s)$ and again move it across g_s (which will rotate and mildly stretch or contract it) and a_t (which will expand it) to the left. Defining $\varepsilon_{\mathrm{new}}(s) = g_s \varepsilon(s)$, we obtain

$$\int_{\mathcal{I}} f(u_{r\mathbf{e}_1} a_t g_s u_{\gamma(s)} x_0)\,\mathrm{d}s = \int_{\mathcal{I}'} f(u_{\mathrm{e}^{\kappa t}\varepsilon_{\mathrm{new}}(s)} a_t g_s u_{\gamma(s')})\,\mathrm{d}s' + \mathrm{O}\big(r\mathrm{e}^{-\kappa t}\big).$$

As $\varepsilon(s) = \mathrm{O}\big(\mathrm{e}^{-2\kappa t}\big)$, we know that

$$\mathrm{e}^{\kappa t}\varepsilon_{\mathrm{new}}(s) = \mathrm{O}\big(\mathrm{e}^{-\kappa t}\big)$$

is tiny. Using continuity of f, we see that the term $u_{\mathrm{e}^{\kappa t}\varepsilon_{\mathrm{new}}(s)}$ does not change the value of f much. For a weak* limit μ of (6.4) as $t \to \infty$ this shows that

$$\int f(u_{r\mathbf{e}_1} x)\,\mathrm{d}\mu(x) = \int f\,\mathrm{d}\mu.$$

As $f \in C_c(\mathsf{X}_{d+1})$ was arbitrary, it follows that μ is invariant under U, as desired. □

Suppose now we can use unipotent magic (meaning Ratner's theorem and related results) to show that the measures in (6.4) equidistribute. We next explain what this has to do with the expanded curves $a_t u_{\gamma(s)} x_0$ for $s \in \mathcal{I}$.

Lemma 6.10 (Twisting trick, second step). *Suppose now in addition that for any interval $\mathcal{I}$ and any $x_0 \in \mathsf{X}_{d+1}$ we have*

$$\frac{1}{|\mathcal{I}|}\int_{\mathcal{I}}\delta_{a_t g_s u_{\gamma(s)}x_0}\,\mathrm{d}s \longrightarrow m_{\mathsf{X}_{d+1}}.$$

Then for any interval $\mathcal{I}$ *and* $x_0\in\mathsf{X}_{d+1}$ *we also have that*

$$\frac{1}{|\mathcal{I}|}\int_{\mathcal{I}}\delta_{a_t u_{\gamma(s)}x_0}\,\mathrm{d}s \longrightarrow m_{\mathsf{X}_{d+1}}.$$

In other words, the additional matrix g_s that was so helpful above to obtain unipotent invariance can simply be forgotten.

Proof of Lemma 6.10. Let $\varepsilon>0$ and $f\in C_c(\mathsf{X}_{d+1})$. By the assumed continuity of the map $\mathcal{I}\ni s\mapsto g_s$, the function $\mathcal{I}\times\mathsf{X}_{d+1}\ni(s,x)\mapsto f(g_s^{-1}x)$ is uniformly continuous. Hence there exists some $\delta>0$ so that

$$\left|f(g_{s_1}^{-1}x)-f(g_{s_2}^{-1}x)\right|<\varepsilon \tag{6.5}$$

whenever $x\in\mathsf{X}_{d+1}$ and $s_1,s_2\in\mathcal{I}$ satisfy $|s_1-s_2|<\delta$.

We split $\mathcal{I}$ into finitely many sub-intervals $\mathcal{I}_\ell$ for $\ell=1,\ldots,L$ of equal length and length less than δ. Using the assumed equidistribution for each of these intervals, the above continuity property of f, and the fact that g_s commutes with a_t, we can now obtain the desired conclusion up to 2ε. Indeed, fix some $s_\ell\in\mathcal{I}_\ell$ for $\ell=1,\ldots,L$ and apply our assumption to the interval $\mathcal{I}_\ell$ and the function $\mathsf{X}_{d+1}\ni x\mapsto f(g_{s_\ell}^{-1}x)$. As $m_{\mathsf{X}_{d+1}}$ is invariant under g_{s_ℓ}, we therefore have

$$\left|\frac{1}{|\mathcal{I}_\ell|}\int_{\mathcal{I}_\ell} f\left(g_{s_\ell}^{-1}g_s a_t u_{\gamma(s)}\boldsymbol{\cdot}x_0\right)\,\mathrm{d}s-\int_{\mathsf{X}_{d+1}} f\,\mathrm{d}m_{\mathsf{X}_{d+1}}\right|<\varepsilon$$

for $\ell=1,\ldots,L$ and all sufficiently large t. Now we may use the estimate (6.5) for $x=g_s a_t u_{\gamma(x)}\boldsymbol{\cdot}x_0$, $s,s_\ell\in\mathcal{I}_\ell$, and $\ell=1,\ldots,L$ to obtain

$$\begin{aligned}
&\Big|\frac{1}{|\mathcal{I}|}\int_I f(a_t u_{\gamma(s)}x_0)\,\mathrm{d}s-\int_{\mathsf{X}_{d+1}} f\,\mathrm{d}m_{\mathsf{X}_{d+1}}\Big|\\
&\qquad=\left|\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{|\mathcal{I}_\ell|}\int_{\mathcal{I}_\ell} f(g_s^{-1}g_s a_t u_{\gamma(s)}\boldsymbol{\cdot}x_0)\,\mathrm{d}s-\int_{\mathsf{X}_{d+1}} f\,\mathrm{d}m_{\mathsf{X}_{d+1}}\right|\\
&\qquad\leqslant\left|\frac{1}{L}\sum_{\ell=1}^{L}\left(\frac{1}{|\mathcal{I}_\ell|}\int_{\mathcal{I}_\ell} f(g_{s_\ell}^{-1}g_s a_t u_{\gamma(s)}\boldsymbol{\cdot}x_0)\,\mathrm{d}s-\int_{\mathsf{X}_{d+1}} f\,\mathrm{d}m_{\mathsf{X}_{d+1}}\right)\right|+\varepsilon\\
&\qquad\leqslant 2\varepsilon
\end{aligned}$$

for all sufficiently large $t>0$. As $f\in C_c(\mathsf{X}_{d+1})$ and $\varepsilon>0$ were arbitrary, we obtain the lemma. □

6.1.5 Totally Geodesic Submanifolds in Hyperbolic Quotients

(to come)

6.1.6 Integer Points on Spheres

We end this motivational section with another application of Ratner's theorem in the context of integer points on spheres. For this fix $d \geqslant 2$ and let $D \geqslant 1$ be an integer such that

$$\mathbb{S}^{d-1}(D) = \left\{ \frac{1}{\sqrt{D}} v \;\middle|\; v \in \mathbb{Z}^d_{\mathrm{prim}}, \|v\| = \sqrt{D} \right\}$$

is non-empty. Then one can ask about the behaviour of this finite set as $D \to \infty$. For $d = 2$ this question is related to factorization in the Gaussian integers $\mathbb{Z}[\mathrm{i}]$ and, if $D = p$ is a prime congruent to 1 modulo 4, the set $\mathbb{S}^{d-1}(p)$ contains 8 points. It follows that we cannot expect density or equidistribution as $D \to \infty$.

For $d = 3$ this problem has been studied by Linnik using the so-called 'ergodic method' under a congruence condition on D, with many subsequent developments.(33) In a sense this is the most difficult dimension in which to study the problem. The full equidistribution theorem for $\mathbb{S}^2(D)$ is the content of Duke's theorem [39], proved using analytic number theory and a breakthrough result of Iwaniec [77].

For $d \geqslant 4$ the distribution problem of $\mathbb{S}^{d-1}(D)$ is easier to study by various methods. In particular, it can be approached using unipotent dynamics, and this was done by Aka, Einsiedler, and Shapira [2]. We note that for $d = 4$ the set $\mathbb{S}^{d-1}(D)$ is non-empty precisely when D is not divisible by 8, and that for $d \geqslant 5$ it is non-empty for every $D \in \mathbb{N}$. Below we always restrict to the set of $D \in \mathbb{N}$ for which $\mathbb{S}^{d-1}(D)$ is non-empty. For $v \in \mathbb{S}^{d-1}(D)$ the intersection $v^\perp \cap \mathbb{Z}^d$ is a lattice in the $(d-1)$-dimensional subspace $v^\perp$. After rotating this hyperplane to $\mathbb{R}^{d-1} \times \{0\}$ and rescaling the lattice one obtains for each $v \in \mathbb{S}^{d-1}(D)$ an element $[v^\perp \cap \mathbb{Z}^d] \in \mathrm{SO}_{d-1}(\mathbb{R})\backslash \mathsf{X}_d$—the 'shape' of the lattice $v^\perp \cap \mathbb{Z}^d$ in the orthogonal complement of v.

Theorem 6.11 (Equidistribution of integer points and their orthogonal shapes). *For $d \geqslant 4$ the set*

$$\left\{ \left(v, [v^\perp \cap \mathbb{Z}^d]\right) \;\middle|\; v \in \mathbb{S}^{d-1}(D) \right\} \subseteq \mathbb{S}^{d-1} \times \mathrm{SO}_{d-1}(\mathbb{R})\backslash \mathsf{X}_d$$

equidistributes as $D \to \infty$ for $d > 4$ and with $D \in \mathbb{N} \smallsetminus 8\mathbb{N}$ for $d = 4$.

It should be surprising that unipotent dynamics can be useful for this problem. Indeed the homogeneous space $\mathbb{S}^{d-1}$ is associated to its isometry group $\mathrm{SO}_d(\mathbb{R})$, which is compact and contains no unipotent elements. However, it is possible to define an enlarged locally homogeneous space by using $\mathbb{Q}_p$ for the prime $p = 5$ (for example). In this enlarged space one can find one-parameter unipotent subgroups (with the parameter belonging to $\mathbb{Q}_p$) and apply theorems of Ratner [137] or Margulis and Tomanov [113] (see Chapter 10). We also note that in [2] a congruence condition on D was assumed for $d \in \{4, 5\}$, but that this condition is unnecessary due to an observation by Michael Bersudsky.

6.2 The Main Theorems

We let $X = G/\Gamma$, where G is a closed subgroup of $\mathrm{SL}_d(\mathbb{R})$ 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 [132], [133], [134]. As it turned out, this was a powerful starting point for the other results that follow.

The classification results will generally take the form of asserting that an initially unknown measure has 'algebraic structure': A probability measure μ on G/Γ is called *algebraic*, *homogeneous*, or *periodic* 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 on a closed finite volume orbit $L\bullet x_0$ (for any $x_0 \in \operatorname{supp}\mu$).

Theorem* 6.12 (Dani's conjecture and Ratner's measure classification). If $X = G/\Gamma$ with Γ discrete 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 5.18, Theorem 6.12, and the general non-divergence property of unipotent orbits in Chaper 4 suggest other results. Ratner [135] generalized all of these results in the following theorem.

Theorem* 6.13 (Ratner's equidistribution theorem). Let Γ be a lattice, let $X = G/\Gamma$, 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 $U \leqslant L$,

- $L\bullet x_0$ is closed with finite L-invariant volume, and
- $\displaystyle\frac{1}{T}\int_0^T f(u_s\bullet x_0)\,\mathrm{d}s \longrightarrow \frac{1}{\mathrm{vol}(L\bullet x_0)}\int_{L\bullet x_0} f\,\mathrm{d}m_{L\bullet x_0}$ as $T \longrightarrow \infty$ for any function $f \in C_c(X)$.

It is interesting to note that Theorem 6.13 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 $\mathsf{d}(x, u_{t_k} \cdot x) \to 0$ 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 6.13 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[(34)] that motivated all of the theorems above, and was proved by Ratner [134] using the above results as stepping stones. The orbit closure $\overline{H \cdot x_0}$ of a point x_0 in G/Γ under the action of a closed subgroup H is similarly called *algebraic*, *homogeneous*, or *periodic* if there exists some closed connected unimodular subgroup L with $H \leqslant L \leqslant G$ such that $\overline{H \cdot x_0} = L \cdot x_0$, and $L \cdot x_0$ supports a finite L-invariant measure.

Theorem* 6.14 (Raghunathan's conjecture; Ratner's orbit closure theorem). Let Γ be a lattice, let $X = G/\Gamma$, and 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 5.8) 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 (for example, asymptotic counting results). A satisfying answer to this question is given by Mozes and Shah [123].

Theorem* 6.15 (Mozes–Shah equidistribution theorem). †Let Γ be a lattice, let $X = G/\Gamma$, 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 either $\mu = 0$ or μ is an algebraic measure, where in each case more can be said:

(a) If $\mu = 0$ then $\operatorname{supp} \mu_n \to \infty$ 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 $n \geqslant N$.
(b) Otherwise $\mu = m_{L \cdot y}$ is the L-invariant probability measure on a closed finite volume orbit $L \cdot y$ for the closed connected group $L = \operatorname{Stab}_G(\mu)^o \leqslant G$. Moreover, μ is invariant and ergodic for the action of a one-parameter unipotent

† 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.

subgroup. Furthermore, suppose that $x_n = \varepsilon_n \boldsymbol{\cdot} x \in \operatorname{supp}\mu_n$ 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 $\mu_n = m_{L_n \boldsymbol{\cdot} x_n}$ for $n \geqslant 1$. Then $Lx = Ly = \operatorname{supp}\mu$ 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 (a), 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 (b), 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 (b) says that the convergence to the limit measure $m_{L\boldsymbol{\cdot} x}$ is almost from within the orbit $L\boldsymbol{\cdot} x$. In fact, after modifying the measures in the sequence only slightly by the elements ε_n we get

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

for $n \geqslant N$.

We will prove special cases of the theorems above.

6.2.1 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 $x \in X = G/\Gamma$. In this section we explain how this kind of question is intimately related to questions of rationality.

This relationship is elementary in the abelian setting of $G = \mathbb{R}^d$, $\Gamma = \mathbb{Z}^d$, 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

$$U = \left\{ \begin{pmatrix} 1 & s \\ & 1 \end{pmatrix} \;\middle|\; s \in \mathbb{R} \right\} < \mathrm{SL}_2(\mathbb{R})$$

and $\mathsf{X}_2 = \mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$, which we already understand in some detail (see Section 1.2, Chapter 5, and [46, Sec. 11.7]). If $x = g\Gamma$ 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 $\mathrm{SL}_2(\mathbb{Z}) \cdot (z, v)$.

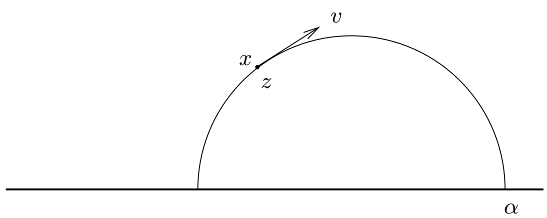

Fig. 6.1: The geodesic determined by x.

In general the answer is given by the following result found by Borel and Prasad [11]. A more general version of this result was obtained more recently by Tomanov [161].

Theorem* 6.16 (Borel–Prasad). Let $\mathsf{X}_d = \mathrm{SL}_d(\mathbb{R})/\mathrm{SL}_d(\mathbb{Z})$, $x = g\Gamma \in X$, 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 6.12 and 6.13 (respectively Theorem 6.14) is the connected component of $g\mathbb{F}(\mathbb{R})g^{-1}$, 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 $g\mathbb{F}(\mathbb{R})g^{-1}$ contains U (respectively H).

Similarly, the group L appearing in Theorem 6.15 is the connected component of $g\mathbb{F}(\mathbb{R})g^{-1}$ where $x = g\Gamma$ and $\mathbb{F}$ is the smallest algebraic group $\mathbb{F} \leqslant \mathrm{SL}_d$ defined over $\mathbb{Q}$ for which $g\mathbb{F}(\mathbb{R})g^{-1}$ contains $\varepsilon_n L_n \varepsilon_n^{-1}$ for $n \geqslant N$, where N is as in Theorem 6.15.

For this result, one needs some understanding of the mechanisms that make orbits $\mathbb{F}(\mathbb{R})\,\mathrm{SL}_d(\mathbb{Z})$ 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 and the required language further in Section 9.7.

6.3 First Ideas in Unipotent Dynamics

The structure of the proof of Ratner's measure classification (Theorem 6.12) involves studying

$$\mathrm{Stab}_G(\mu) = \{g \in G \mid g_*\mu = \mu\}$$

and showing that the measure μ on $X = G/\Gamma$ is supported on a single orbit of this subgroup. This will often be achieved indirectly in the following sense: 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 6.12, once we have shown that μ gives a single orbit of $\mathrm{Stab}_G(\mu)$ positive measure, we actually obtain that μ is supported on a single closed orbit of $\mathrm{Stab}_G(\mu)^o$.

Lemma 6.17. *Let $X = G/\Gamma$ be a quotient of a closed linear 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 positive measure to a single orbit of its stabilizer subgroup* $\mathrm{Stab}_G(\mu)$*, then μ is the Haar measure on a closed orbit of the subgroup* $\mathrm{Stab}_G(\mu)^o$*.*

Proof. If $\mu\big(\mathrm{Stab}_G(\mu)\cdot x_0\big) > 0$ for some $x_0 \in X$, then $\mu\big(\mathrm{Stab}_G(\mu)\cdot x_0\big) = 1$ by ergodicity. As the index of $\mathrm{Stab}_G(\mu)^o$ in $\mathrm{Stab}_G(\mu)$ is at most countable, there exists a point x_1 so that $\mu\big(\mathrm{Stab}_G(\mu)^o\cdot x_1\big) > 0$. This implies once more that $\mu\big(\mathrm{Stab}_G(\mu)^o\cdot x_1\big) = 1$ as H is assumed to be connected. It follows that μ is the Haar measure on $\mathrm{Stab}_G(\mu)^o\cdot x_1$, which is also closed by Corollary 1.36. □

6.3.1 Generic Points with Parallel Orbits

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.(35)

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

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

as $T \to \infty$ for all $f \in C_c(X)$. Using the pointwise ergodic theorem [46, 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 also Lemma 6.23 for a stronger claim).

Lemma 6.18 (Parallel orbits). *If x and $y = h{\bullet}x$ in X are generic for μ and $h \in C_G(U) = \{g \in G \mid gu = ug$ for all $u \in U\}$, then h preserves μ.*

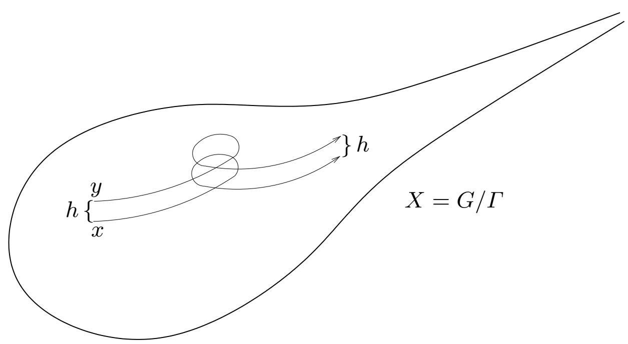

Fig. 6.2: If $y = h{\bullet}x$ with $h \in C_G(U)$, then the two orbits are parallel. If in addition both x and y are generic, then the orbits equidistribute (that is, approximate μ), which gives Lemma 6.18.

Proof of Lemma 6.18. We refer to Figure 6.2 for a depiction of the proof. We know that

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

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

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

where $f^h\colon X \ni z \mapsto f(h\boldsymbol{\cdot} z)$. As this holds for any $f \in C_c(X)$ we see that μ is h-invariant. □

Lemma 6.18 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 Factor Rigidity for Quotients of $\mathbf{SL_3}(\mathbb{R})$

Following Witte Morris [172] we wish to prove the Factor Rigidity Theorem 6.2 for the case $G = \mathrm{SL}_3(\mathbb{R})$. For this we note that $\mathrm{SL}_3(\mathbb{R})$ contains up to conjugation two fundamentally different one-parameter unipotent subgroups: One defined by the Lie algebra element x_2 in (2.5) and one defined by x_3 in (2.6). The result holds for either of the two. In the proof it will be useful to even allow factor maps between these two types of one-parameter unipotent subgroups—even though we will see in the end that no such factor maps exist.

We let $G = \mathrm{SL}_3(\mathbb{R})$, let $\Gamma_1, \Gamma_2 < G$ be lattices, and define $X_j = G/\Gamma_j$ for $j = 1, 2$. Let $U_j = \{u_j(t) \mid t \in \mathbb{R}\} < G$ be a one-parameter unipotent subgroup for $j = 1, 2$. Finally, suppose that $\phi\colon X_1 \to X_2$ is a measurable map defined almost everywhere with respect to the probability measure m_{X_1} satisfying $m_{X_2} = \phi_* m_{X_1}$ and

$$\phi\big(u_1(t)\boldsymbol{\cdot} x\big) = u_2(t)\boldsymbol{\cdot}\phi(x) \tag{6.6}$$

for $t \in \mathbb{R}$ and almost every $x \in X_1$. As in Section 6.1.1 we use ϕ to define the probability measure $\mu = (\mathrm{id}, \phi)_* m_{X_1}$ on $X_1 \times X_2$ concentrated on $\mathrm{Graph}(\phi)$. By (6.6) the measure μ is invariant under $\big(u_1(t), u_2(t)\big)$ for $t \in \mathbb{R}$. Ergodicity of μ for this action is inherited from ergodicity of m_{X_1} under U_1.

Lemma 6.19 (Partial homomorphism). *The closed subgroup*

$$L = \mathrm{Stab}_{G\times G}(\mu)$$

is the graph of a homomorphism $\varphi\colon H \to G$ defined on a subgroup $H < G$. For $h \in H$ acting ergodically with respect to m_{X_2} we have that $\varphi(h)$ also acts ergodically with respect to m_{X_2}.

Proof. Suppose $(h,\ell)\in L$ for some $h,\ell\in G$. Then (h,ℓ) preserves μ and hence sends the set $\mathrm{Graph}(\phi)$ (which is of full measure) to itself up to a null set. Equivalently,

$$\phi(h\cdot x)=\ell\cdot\phi(x) \tag{6.7}$$

for m_{X_1}-almost every $x\in X_1$.

Suppose now in addition that $h=I$. Then

$$\ell\cdot\phi(x)=\phi(I\cdot x)=\phi(x)$$

for m_{x_1}-almost every $x\in X_1$. Using $\phi_*m_{X_1}=m_{X_2}$ this gives $\ell g\Gamma_2=\Gamma_2$ for m_{X_2}-almost every $g\Gamma_2\in X_2$. By continuity it follows that $g^{-1}\ell g\in\Gamma_2$ for all $g\in G$, which in turn implies $\ell\in\Gamma_2\cap C_G$. However, for $G=\mathrm{SL}_3(\mathbb{R})$ the centre C_G is trivial. Summarizing, we have shown that $(I,\ell)\in L$ implies $\ell=I$ also. Setting $H=\pi_1(L)$ (the projection onto the first coordinate) we obtain that $L=\mathrm{Graph}(\varphi)$ for a homomorphism $\varphi\colon H\to G$.

Suppose now that $\ell=\varphi(h)$ does not act ergodically with respect to m_{X_2}. Then there exists a non-trivial ℓ-invariant subset $B\subseteq X_2$. Taking the pre-image under ϕ we obtain together with (6.7) a non-trivial h-invariant subset in X_1. That is, h does not act ergodically with respect to m_{X_1}. □

By our initial setup in (6.6) we know that φ is defined on $U_1\subseteq H$ and $\varphi(u_1(t))=u_2(t)$ for $t\in\mathbb{R}$. Our goal is to show that φ is actually defined on all of G. This is achieved in several steps, with the following lemma being the key.

Lemma 6.20 (Centralizer). *We have $C_G(U_1)^o\subseteq H$ and*

$$\varphi\bigl(C_G(U_1)^o\bigr)\subseteq C_G(U_2)^o.$$

Proof. We are going to use several sets of almost full measure. Let X_{gen} be the set of points $x\in X_1$ such that $\bigl(x,\phi(x)\bigr)\in\mathrm{Graph}(\phi)$ is generic for $\{(u_1(t),u_2(t))\mid t\in\mathbb{R}\}$, and we note that $m_{X_1}(X_{\mathrm{gen}})=1$. By Lusin's theorem there exists a compact set $X_{\mathrm{LT}}\subseteq X_1$ with $m_{X_1}(X_{\mathrm{LT}})>0.99$ on which ϕ is defined and continuous. This implies in particular that $\phi(X_{\mathrm{LT}})$ is compact. The maximal ergodic theorem for U_1 acting on X_1 shows that the set

$$X_{\mathrm{MET}}=\Bigl\{x\in X_1\Bigm| |\{t\in[0,T]\mid u_1(t)\cdot x\in X_{\mathrm{LT}}\}|\geqslant\tfrac{9}{10}T\text{ for all }T>0\Bigr\}$$

has $m_{X_1}(X_{\mathrm{MET}})>0.9$ (we are again using $|A|$ to denote the Lebesgue measure of a set A). As $X_{\mathrm{LT}}\times\phi(X_{\mathrm{LT}})$ is compact there exists a uniform injectivity radius r on it.

Note that

$$p_g(t)^2=\bigl\|u_2(t)gu_2(-t)-I\bigr\|_2^2$$

is, for any fixed $g\in G$, a polynomial in $t\in\mathbb{R}$. Moreover, there is a uniform bound $2D$ on the degree of this polynomial. Applying the Remes inequality in Lemma 4.8 gives some $\varepsilon_0>0$ so that for any $T>0$ and $g\in G$ we have

$$\left|\left\{t\in[0,T]\mid \|p_g(t)\|_2<\varepsilon_0\|p_g\|_{T,\infty}\right\}\right|<\tfrac{1}{10}T,$$

where $\|p_g\|_{T,\infty}=\sup_{t\in[0,T]}\|p_g(t)\|_2$. We convert this into geometric terms as follows. There exists $\varepsilon>0$ such that if $x', y'=g{\boldsymbol\cdot}x'\in X_2$, and $T>0$ is minimal such that $\mathsf{d}\bigl(u_2(T)gu_2^{-1}(-T),I\bigr)=r$, then

$$\left|\left\{t\in[0,T]\;\middle|\;u_2(t)x', u_2(t)y'\in\phi(X_{\mathrm{LT}}), \mathsf{d}(u_2(t)x',u_2(t)y')<\varepsilon\right\}\right|<\tfrac{1}{10}T. \quad (6.8)$$

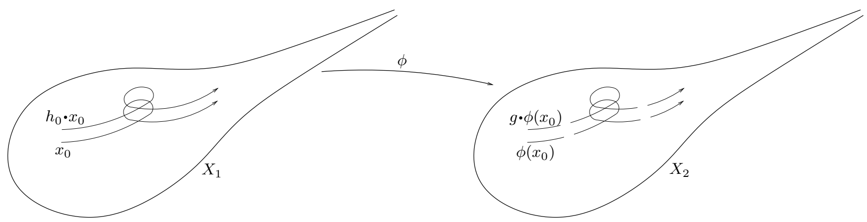

Fig. 6.3: We study two orbits that are completely parallel and close together in the domain X_1. Moreover, they spend together 80% of their time in a Lusin set for ϕ. This gives for the orbits of their image that for 80% of the time they are close together (the gaps in the orbits indicate our incomplete knowledge about other times).

As $\phi|_{X_{\mathrm{LT}}}$ is uniformly continuous, there exists some $\delta\in(0,r)$ so that if

$$x, y=h{\boldsymbol\cdot}x\in X_{\mathrm{LT}}$$

and $\mathsf{d}_{X_1}(x,y)=\mathsf{d}_G(I,h)<\delta$ then $\mathsf{d}_{X_2}(\phi(x),\phi(y))<\varepsilon$.

We now suppose that $h_0\in B_\delta^{C_G(U_1)}$. Note that

$$m_{X_1}\bigl(X_{\mathrm{gen}}\cap X_{\mathrm{LT}}\cap X_{\mathrm{MET}}\bigr)>0.8$$

and hence there exists some

$$x_0\in X_{\mathrm{gen}}\cap X_{\mathrm{LT}}\cap X_{\mathrm{MET}}\cap h_0^{-1}\bigl(X_{\mathrm{gen}}\cap X_{\mathrm{LT}}\cap X_{\mathrm{MET}}\bigr).$$

Using the definition of X_{LT}, $\delta>0$, and h_0 it follows that $\phi(h_0{\boldsymbol\cdot}x_0)=g_0{\boldsymbol\cdot}\phi(x_0)$ for some $g_0\in G$ with $\mathsf{d}_G(g_0,I)<\varepsilon$. We claim that $g_0\in C_G(U_2)$.

Using the claim and the fact that $(x_0,\phi(x_0))$ and

$$(h_0{\boldsymbol\cdot}x_0,\phi(h_0{\boldsymbol\cdot}x_0))=(h_0,g_0){\boldsymbol\cdot}(x_0,\phi(x_0))$$

are generic, Lemma 6.18 implies that $(h_0,g_0)\in L$ preserves μ. As any neighbourhood of $I\in C_G(U_2)$ generates a subgroup containing $C_G(U_2)^o$ the claim implies the lemma.

To prove the claim we suppose that $g_0 \notin C_G(U_2)$. This implies that $\phi(x_0)$ and $g_0 \cdot \phi(x_0)$ have orbits that eventually move away from each other. Therefore we can find a minimal $T > 0$ satisfying (6.8). This leads to a contradiction by the following argument. The bound (6.8) says that only 10% of all $t \in [0, T]$ have the property that the images of $\phi(x_0)$ and of $\phi(h_0 \cdot x_0) = g_0 \phi(x_0)$ under $u_2(t)$ belong to $\phi(X_{\mathrm{LT}})$ and have distance apart less than ε. However, the definitions of X_{LT} and X_{MET} together with the equivariance of ϕ show that this holds for at least 80% of all $t \in [0, T]$. □

Proof of Theorem 6.2 for $G = \mathrm{SL}_3(\mathbb{R})$. Let $\phi\colon X_1 \to X_2$ be a factor map for U as in the statement of the theorem. We define $\mu = (I \times \phi)_* m_{X_1}$ and apply Lemmas 6.19 and 6.20. For the following it is necessary to know the centralizers of the one-parameter unipotent subgroups. For a root unipotent subgroup, say $U_{1,3} = \{u_{1,3}(t) = I + tE_{1,3} \mid t \in \mathbb{R}\}$ the connected component of the centralizer is four-dimensional and is given by

$$H_{1,3} = C_G(U_{1,3})^o = \left\{ \begin{pmatrix} a & * & * \\ & a^{-2} & * \\ & & a \end{pmatrix} \middle| \, a > 0 \right\}.$$

On the other hand, for the 'long' unipotent subgroup $U_{\mathrm{long}} < \mathrm{SO}_{2,1}(\mathbb{R})$ (defined by the Lie algebra element x_3 in (2.6)) the centralizer is two-dimensional and is given by

$$H_{\mathrm{long}} = C_G(U_{\mathrm{long}})^o = \left\{ \begin{pmatrix} 1 & t & * \\ & 1 & t \\ & & 1 \end{pmatrix} \middle| \, t \in \mathbb{R} \right\} = U_{\mathrm{long}} U_{1,3}.$$

We leave these two claims as exercises but note that $U_{1,3}$ is normalized by the full diagonal subgroup, which implies the same for $H_{1,3}$. This helps in the argument needed.

This already implies that a factor map from the $U_{1,3}$-action to the U_{long}-action cannot exist: In this scenario the partial homomorphism φ would be defined on the four-dimensional subgroup $H_{1,3}$ with $\varphi(H_{1,3})$ belonging to the two-dimensional H_{long}. This contradicts the final claim in Lemma 6.19.

Suppose now that ϕ is a factor map for the action of U_{long} on both spaces. Then φ is defined on $U_{1,3}$, which implies that ϕ is also a factor map from the action of $U_{1,3}$ on X_1 to the action of $\varphi(U_{1,3})$ on X_2. By the previous paragraph $\varphi(U_{1,3})$ cannot be conjugated to U_{long}, which implies that it is conjugated to $U_{1,3}$ instead. This allows us to reduce the case of U_{long} to the case of $U_{1,3}$ considered next.

Suppose therefore that ϕ is a factor map for $U_{1,3}$ (acting on both X_1 and X_2). Applying Lemmas 6.19 and 6.20 we obtain that φ is defined on the four-dimensional $H_{1,3}$ satisfying $\varphi(H_{1,3}) \subseteq H_{1,3}$. Now notice that the commutator subgroup of $H_{1,3}$ is given by

$$[H_{1,3}, H_{1,3}] = \left\{ \begin{pmatrix} 1 & * & * \\ & 1 & * \\ & & 1 \end{pmatrix} \right\}.$$

Hence φ sends the upper-triangular subgroup to itself. In particular, φ is defined on the unipotent one-parameter subgroups $U_{1,2}$ and $U_{2,3}$, with $\varphi(U_{1,2})$ and $\varphi(U_{2,3})$ again two one-parameter unipotent subgroups. Applying the centralizer Lemma 6.20 twice more shows that φ is also defined on $U_{3,2}$ (commuting with $U_{1,2}$) and $U_{2,1}$ (commuting with $U_{2,3}$). However, as $U_{1,3}$, $U_{1,2}$, $U_{2,3}$, $U_{3,2}$, and $U_{2,1}$ together generate $\mathrm{SL}_3(\mathbb{R})$ we see that φ is defined on all of $\mathrm{SL}_3(\mathbb{R})$: For every $g \in \mathrm{SL}_3(\mathbb{R})$ we have

$$\phi(g \cdot x) = \varphi(g)\phi(x)$$

for m_{X_1}-almost every $x \in X_1$. Applying Fubini's theorem we find some $x = h_1 \varGamma_1$ and $\phi(x) = h_2 \varGamma_2$ so that

$$\phi(g h_1 \varGamma_1) = \varphi(g) h_2 \varGamma_2$$

for m_G-almost every $g \in \mathrm{SL}_3(\mathbb{R})$. Replacing g by $g h_1^{-1}$ and setting

$$p = \varphi(h_1)^{-1} h_2$$

we can also write this as

$$\phi(g \varGamma_1) = \varphi(g) p \varGamma_2$$

for m_G-almost every $g \in \mathrm{SL}_3(\mathbb{R})$. For g and $g\gamma_1$ with $\gamma_1 \in \varGamma_1$ this also implies

$$\varphi(g) p \varGamma_2 = \varphi(g\gamma_1) p \varGamma_2$$

or, equivalently,

$$\varphi(\gamma_1) \in p \varGamma_2 p^{-1}.$$

□

6.3.3 Polynomial divergence leading to invariance

Let $U = \{u(s) \mid s \in \mathbb{R}\}$ be a unipotent one-parameter subgroup of G, then $\mathrm{Ad}_{u(s)}$ is unipotent for all $s \in \mathbb{R}$ and is a (matrix-valued) polynomial in s. This polynomial (as opposed to exponential) structure of unipotent subgroups has the following consequence. Given a nearby pair of points x and $y = \varepsilon \cdot x$, let $v = \log \varepsilon$ and consider the $\mathfrak{g}$-valued polynomial $\mathrm{Ad}_{u(s)}(v)$. 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 non-constant (equivalently, ε does not lie in $C_G(U)$), 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[†] $\|\operatorname{Ad}_{u(T)}(v)\| = 1$, 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.

In order to state the principle that gives additional invariance, we will need the following refinement of the notion of genericity.

Definition 6.21 (Uniformly generic points). 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 time $T_0 = T_0(f, \varepsilon)$ with

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

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

Proposition 6.22 (Polynomial divergence leads to invariance). *Suppose that $(x_n), (y_n)$ are sequences of uniformly generic points with $y_n = \varepsilon_n \bullet x_n$ 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*

$$p_n(r) = \operatorname{Ad}_{u(T_n r)}(v_n),$$

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 $r \in \mathbb{R}$, where

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

is a polynomial with entries in the Lie algebra $\mathfrak{g}$. Then μ is invariant under $\exp(p(r))$ 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

[†] It does not matter which norm on $\mathfrak{g}$ is used.

[‡] 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.

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 6.18. Part of the argument for Proposition 6.22 is illustrated in Figure 6.4.

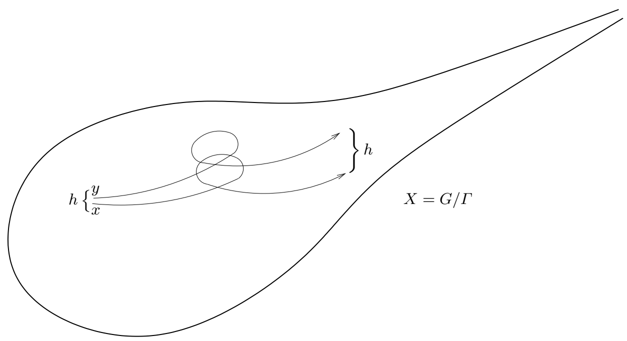

Fig. 6.4: If $y = \varepsilon \cdot x$ with $\varepsilon \notin C_G(U)$ close to the identity, then the orbits of x and y move away from each other at polynomial speed. If x and y are generic then the last 1% of these pieces of orbits are almost parallel and equidistribute.

Proof of Proposition 6.22. 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

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

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[†]

$$\mathsf{d}(\exp p_n(r), \exp p(r_0)) < \delta \tag{6.9}$$

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

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

[†] This is the formal version of the statement in Figure 6.4 that the last 1% are parallel.

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

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

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

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

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

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

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

We now restrict ourself to the range of $s \in \mathbb{R}$ with $\frac{s}{T_n} \in [r_0-\kappa, r_0]$. Together with (6.9), 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 $s \in [(r_0-\kappa)T_n, r_0T_n]$. 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 6.23 (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 [46, Cor. 8.15] for every $f_\ell \in D$ we have

[†] In Figure 6.4 we referred to this as the equidistribution of the last 1% of the orbit.

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

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

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

as $T_0 \to \infty$. Now choose, for every $f_\ell \in D$ and for every $\varepsilon = \frac{1}{n}$, a time $T_{\ell,n}$ so that

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

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 function $f \in D$. 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 6.25); 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 1% argument in Proposition 6.22. This will lead to a strengthening of Dani's theorem (Theorem 5.7), due to Ratner, and is an important step towards the general case.

Exercise 6.24. Show that the limit polynomial in Proposition 6.22 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 6.25. Use the results from Section 6.3.3 to prove the measure classification theorem (Theorem 6.12) 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 $X = G/\Gamma$ is called a *nilmanifold*.

6.4.1 Measure Classification for Nilmanifolds

Theorem 6.26. *Let $\Gamma < G$ be a discrete subgroup of a connected nilpotent Lie group G and let $X = G/\Gamma$. 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 6.22. 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 centre 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 $g \in \operatorname{Stab}_G(\mu)$ by Lemma 6.18 and $y \in \operatorname{Stab}_G(\mu){\cdot}x_0$ also. It follows that $X' \subseteq \operatorname{Stab}_G(\mu){\cdot}x_0$ has full measure, and we deduce that μ must be the Haar measure on $\operatorname{Stab}_G(\mu){\cdot}x_0$ 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 $j \in \{0, \dots, k-1\}$. We may also assume that $U \not\subseteq C_G$. 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 6.23. Choose some

$$x_0 \in K \cap \operatorname{supp}(\mu|_K). \tag{6.10}$$

We distinguish between two possible scenarios.

It could be that there is some $\delta > 0$ such that $y = h{\cdot}x_0 \in K$ with $\mathsf{d}(h, I) < \delta$ implies that $h \in C_G(U)$ and so also $h \in \operatorname{Stab}_G(\mu)$ by Lemma 6.18. In this case (6.10) implies that $\operatorname{Stab}_G(\mu){\cdot}x_0$ has positive measure, and so we may apply Lemma 6.17 to conclude.

In the second case we find a sequence $(y_n = \varepsilon_n{\cdot}x_0)$ in K with $\varepsilon_n \to I$ 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 6.22 converges to a non-constant polynomial $p\colon \mathbb{R} \to \mathfrak{g}$. By Proposition 6.22 we deduce that μ is invariant under $\exp(p(r))$ 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 6.22)

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

for all r, where

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

Since $\varepsilon_n \to I$ 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

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

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

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

We are going to apply the inductive hypothesis to $G' = G$, $\Gamma' = \Gamma$, and U'. However, as μ may not be[†] ergodic with respect to U' we first have to decompose μ into U'-ergodic components. Recall from [46, Th. 6.2, 8.20] that the ergodic decomposition allows us to write

$$\mu = \int_X \mu_x^{\mathcal{E}'} \,\mathrm{d}\mu, \tag{6.11}$$

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

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

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 $L_x{\scriptstyle\bullet} x$. 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

$$(u_t)_* \mu_x^{\mathcal{E}'} = \mu_{u_t{\scriptstyle\bullet} x}^{\mathcal{E}'} \tag{6.12}$$

for every $t \in \mathbb{R}$ and μ-almost every x by [46, Cor. 5.24]. Since $\mu_x^{\mathcal{E}'}$ is L_x-invariant, it follows from (6.12) that $(u_1)_* \mu_x^{\mathcal{E}'}$ is $u_1 L_x u_1^{-1}$-invariant, which implies that

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

and, by a similar argument for the reverse inclusion,

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

Iterating this relationship shows that

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

$$u_1^n L_x u_1^{-n} = L_{u(n)\boldsymbol{\cdot} x} \tag{6.13}$$

for μ-almost every x. Now either L is normalized by u_1, or the sequence of subgroups in (6.13) converges to a subgroup that is normalized by u_1 (to see this, apply the argument from the proof of Lemma 3.51 to any element of $\bigwedge^{\dim L_x}(\mathrm{Lie}(L_x))$. 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 6.27 below. Equation (6.11) 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 nontrivial connected subgroup $L \leqslant \mathrm{Stab}_G(\mu)$ containing U' such that the orbit $L\boldsymbol{\cdot} x$ is for μ-almost every x closed, with finite L-invariant measure and with the property that $U' \leqslant L$ acts ergodically on $L\boldsymbol{\cdot} x$. Since $L \leqslant G$ is nilpotent, simply connected and connected, $M = C_L(L)$ is a nontrivial connected subgroup. We claim that the orbit $M\boldsymbol{\cdot} x$ is compact for μ-almost every x and postpone the proof to Lemma 6.29.

Next we claim that $N_G^1(M)\boldsymbol{\cdot} x$ is a closed orbit for μ-almost every x, see Lemma 6.30. This implies that μ is supported on a single orbit $N_G^1(M)\boldsymbol{\cdot} x_0$ 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

$$\operatorname{supp}\mu = \overline{U\boldsymbol{\cdot} x_0} \subseteq N_G^1(M)\boldsymbol{\cdot} x_0.$$

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 $M\boldsymbol{\cdot}\Gamma$ 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

$$\pi_X\colon G/\Gamma \longrightarrow \pi_M(G)/\pi_M(\Gamma)$$

defined by

$$\pi_X\colon g\bullet\Gamma \longmapsto \pi_M(g)\bullet\pi_M(\Gamma).$$

The fibers of this map are precisely the M-orbits in the sense that

$$\pi_X^{-1}\left(\pi_X(g\bullet\Gamma)\right)=\{h\bullet\Gamma \mid hM\bullet\pi_M(\Gamma)=gM\bullet\pi_M(\Gamma)\}=gM\bullet\Gamma$$

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 $\pi_M(g)H'\bullet\pi_M(\Gamma)$ for some $\pi_M(g)\in G'$. Finally, we claim that μ is the H-invariant probability measure on the closed orbit $gH\bullet\Gamma$ where $H=\pi_M^{-1}(H')$.

Since $\pi_M(\Gamma)\pi_M(g)H'$ is closed we also obtain that

$$\Gamma gH=\pi_X^{-1}\left(H'\pi_M(g)\pi_M(\Gamma)\right)$$

is closed. Now let $f\in C(X)$. Then

$$\int_X f(x)\,\mathrm{d}\mu(x)=\int_X f(m\bullet x)\,\mathrm{d}\mu(x) \tag{6.14}$$

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 (6.14) 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 $H{\scriptstyle\bullet}x$ 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 6.27 (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 6.27. 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

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

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 6.28. *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 $C_G(\eta_1,\dots,\eta_k)\Gamma$ is closed in $X = G/\Gamma$.*

Proof. The proof is similar to the proof of Proposition 3.1 or Proposition 3.11. So suppose that $g_n\boldsymbol{\cdot}\Gamma \to g\boldsymbol{\cdot}\Gamma$ 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 $g_n\gamma_n \to g$ 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 6.29. *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 3.50; also see the argument on p. 143) there exist elements $\lambda_1, \ldots, \lambda_k \in \Lambda$ with

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

Thus Lemma 6.28 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 1.35 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. $\square$

Lemma 6.30. *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*

$$N_G^1(M) = \{g \in G \mid gMg^{-1} = M \text{ and } gm_Mg^{-1} = m_M\}$$

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 6.28, 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 γg lies 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 1.45, equivalently via the argument in the proof of Mahler's compactness criteria in Theorem 1.51) there exist elements

$$\eta_{n,1}, \ldots, \eta_{n,\dim M} \in \Gamma$$

such that

$$(\gamma_n g_n)^{-1}\,\eta_{n,i}\,(\gamma_n g_n) \in M \tag{6.15}$$

is of bounded size (independent of n) and gives a basis of

$$(g_n^{-1}\Gamma g_n) \cap M$$

for $i = 1, \ldots, \dim(M)$. Therefore, we may choose a subsequence such that for every $i = 1, \ldots, \dim(M)$ we have (after renaming the indexing variable in the sequence) that

$$(\gamma_n g_n)^{-1}\,\eta_{n,i}\,(\gamma_n g_n) \longrightarrow m_i \in M \tag{6.16}$$

as $n\to\infty$. Since we also have $\gamma_n g_n \to g$ we may conjugate by $\gamma_n g_n$ in (6.16) 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 (6.15) 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

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

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,

$$\gamma_N^{-1} g\big(\gamma_n^{-1} g\big)^{-1} = \gamma_N^{-1}\gamma_n \in N_G(M)$$

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 1.44). 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 6.26 we can establish the equidistribution theorem (Theorem 6.13) and the orbit closure theorem (Theorem 6.14) 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 6.31. *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 let x_0 lie 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 Semisimple Groups

Using Section 6.3.3 we are also ready to prove the special case of Ratner's measure classification theorem where the acting group is semisimple.† 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 [134], the observation that this particular case has a short and relatively easy proof was made in [41].

Theorem 6.32 (Ratner measure classification; the semisimple case). *Let G be a closed linear group, $\Gamma < G$ a discrete subgroup, and assume that H is a semisimple subgroup of G 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 $\operatorname{Stab}(\mu) = \{g \in G \mid g_*\mu = \mu\}$, the connected component $L = \operatorname{Stab}(\mu)^o$, 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 $H \leqslant L$.

† This case is interesting as the proof is relatively straightforward, even though there may be a large gap between 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 $\overline{H\bullet x} \subseteq X$.

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 nontrivially 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 2.25 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 [49, Th. 4.11], Fulton and Harris [57], or Knapp [91, 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 semisimple.

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

$$p_n(r) = \mathrm{Ad}_{u(T_n r)}(\log g_n), \tag{6.17}$$

assume that these converge as $n \to \infty$, and apply Proposition 6.22. 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 invariant under $\exp p(r)$ 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

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

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

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

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

$$x' = \ell_x\boldsymbol{\cdot} x \in K, \tag{6.18}$$

$$y' = \ell_y\boldsymbol{\cdot} y \in K, \tag{6.19}$$

and

$$y' = \exp(v)\boldsymbol{\cdot} x' \tag{6.20}$$

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 neighbourhood O of the identity in G. Let now $g \in B_\kappa^G(I)$ be chosen so that $y = g\boldsymbol{\cdot} x$. Then we would like to find $\ell_x, \ell_y \in B_\delta^L$ with $g\ell_x^{-1} = \ell_y^{-1}\exp(v)$, which will give (6.20). 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

$$\psi^{-1}(g\ell_x^{-1}) = (\ell_y^{-1}, v). \tag{6.21}$$

However, we still have to worry about the conditions (6.18) and (6.19).

For this, we are going to see that most points $\ell_x \in B_\delta^L$ (and the corresponding ℓ_y) will satisfy this. Indeed, by definition of Y, at least 90% of all $\ell_x \in B_\delta^L$ satisfy $x' = \ell_x\boldsymbol{\cdot} x \in K$, and at least 90% of all $\ell_y \in B_\delta^L$ satisfy $y' = \ell_y\boldsymbol{\cdot} y \in K$. However, we need to do this while ensuring that (6.21) (or equivalently, (6.20)) 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 (6.21). 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

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

Together with

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

we see that there are many points $\ell_x \in B_\delta^L$ with $\ell_x{\boldsymbol\cdot}x \in K$ for which ℓ_y defined by (6.21) also satisfies $\ell_y{\boldsymbol\cdot}y \in K$.

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

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

Then for every $\kappa = \frac{1}{n}$ there exist $x_n = z, y_n = g_n{\boldsymbol\cdot}x_n \in Y$ 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

$$x_n', y_n' = \exp(v_n){\boldsymbol\cdot}x_n' \in K, v_n \in V, v_n \neq 0, v_n \longrightarrow 0$$

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 6.18 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, we have $\exp(tw) \in \mathrm{Stab}(\mu)$ 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 the eigenspace for any $n \geqslant 1$ (by deleting finitely many terms). In this case we may define T_n such that the polynomials in (6.17) 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 6.22 to see that μ is invariant under 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. □

6.5.1 A Case of the Mozes–Shah Theorem

We wish now to explain the following case of the Mozes–Shah theorem (Theorem 6.15).

Theorem 6.33. *Let $G = \mathbb{G}(\mathbb{R})^o$ for a simple algebraic group $\mathbb{G}$, $\Gamma < G$ a uniform lattice, and $X = G/\Gamma$. Let $\mathbb{H} < \mathbb{G}$ be a semisimple subgroup and let H be $\mathbb{H}(\mathbb{R})^o$. Suppose furthermore that H has no compact factors and is maximal among all the connected subgroups of G. Then any sequence $(H \cdot x_n)$ of closed H-orbits in X without repetitions equidistributes to the Haar measure on X.*

The proof involves more knowledge of simple and semisimple Lie groups, for which we refer to Knapp [91, Ch. VI]. It may be helpful to simply assume that $G = \mathrm{SL}_3(\mathbb{R})$ and $H = \mathrm{SO}_{2,1}(\mathbb{R})^o$.

Proof of Theorem 6.33. By Theorem 6.32 any H-invariant and ergodic probability measure μ on X is the Haar measure on a closed orbit $L \cdot x$ for a closed connected subgroup $L < G$ containing H. By assumption we have $L = H$ or $L = G$, where the latter corresponds to the normalized Haar measure $\mu = m_X$. In the former case μ is the normalized Haar measure $\mu = m_{Hg\Gamma}$ of a closed H-orbit.

Countably many orbits. As H is semisimple without compact factors the Borel density theorem (Theorem 3.50) implies that $\Gamma \cap g^{-1}Hg$ is Zariski dense in $g^{-1}\mathbb{H}g$. If $\mathbb{G}$ is defined over $\mathbb{Q}$ and $\Gamma = G \cap \mathrm{SL}_d(\mathbb{Z})$, then $g^{-1}\mathbb{H}g = \mathbb{L}$ is defined over $\mathbb{Q}$, which implies that there are only countably many possibilities for $\mathbb{L}$. In general the lattice $\Gamma \cap g^{-1}Hg$ is finitely generated, which implies once more that for $\Gamma \cap g^{-1}H$ and for $\mathbb{L}$ there are only countably many possibilities. If now $Hg_1\Gamma$ and $Hg_2\Gamma$ are two closed orbits for which $g_1^{-1}\mathbb{H}g_1 = \mathbb{L} = g_2^{-1}\mathbb{H}g_2$ is the same algebraic subgroup, then $g_2 g_1^{-1}$ belongs to the normalizer of H. As H is semisimple, the connected component of its normalizer is given by $H \cdot C_G(H)^o$. As G is assumed to be simple and H is assumed to be a maximal connected subgroup, it follows that $C_G(H)^o = \{I\}$. Therefore H has countable† index in its normalizer. Indeed, $N_{\mathbb{G}}(\mathbb{H}) = N_{\mathbb{G}}(H)$ is an algebraic group defined over $\mathbb{R}$ and the connected component of its $\mathbb{R}$-points has at most countable index. Put together these show that there are at most countably many closed H-orbits in X.

Applying the ergodic decomposition. Unfortunately we do not know at this stage whether the limit measure μ is ergodic with respect to H. Hence we have to apply Theorem 6.32 in conjunction with the ergodic decomposition

† With a little more work it may be shown that H in fact has finite index in its normalizer.

(see [46, Th. 8.20]). As we have already shown that there are only countably many H-invariant and ergodic probability measures we obtain

$$\mu = c_0 m_X + \sum_{n=1}^{\infty} c_n m_{H\bullet z_k}$$

with:

- $c_k \in [0,1]$ for $k \in \mathbb{N}_0$;
- $\sum_{k=0}^{\infty} c_k = 1$; and
- $m_{H\bullet z_k}$ the normalized Haar measure on the closed orbit $H\bullet z_k$ for $k \in \mathbb{N}$.

Our goal is to show that $\mu = m_X$, which is achieved by showing that $c_k = 0$ for all $k \in \mathbb{N}$.

Put differently, our goal is to show that $\mu(H\bullet y_0) = 0$ for a given closed orbit $Y = H\bullet y_0$ (which is compact by our assumption on X). Let $V \subseteq \mathfrak{g} = \operatorname{Lie} G$ be an H-invariant linear complement to $\mathfrak{h} = \operatorname{Lie} H$. By a standard compactness argument there exists $\eta > 0$ so that the map

$$B_\eta^V \times Y \ni (v, y) \longmapsto \exp(v)x \in X$$

is an injective homomorphism. This defines a smooth 'tubular' coordinate system in a neighbourhood of Y.

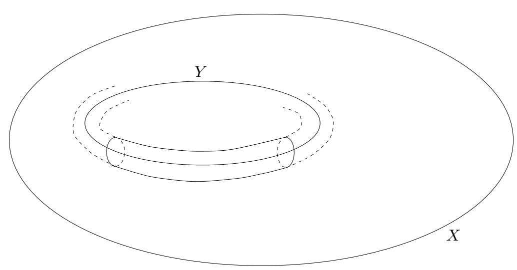

Fig. 6.5: The set $\exp\big(B_\eta^V\big)Y$ is a tubular neighbourhood having the canonical coordinate system $B_\eta^V \times Y$. As we are only concerned with the distance of points to Y we will focus on the V component of this coordinate system.

We note that for $h \in H$, $v \in V$, and $y \in Y$ we have

$$h\bullet(\exp(v)y) = \exp(\operatorname{Ad}_h v)h\bullet y,$$

where $\operatorname{Ad}_h v \in V$ by invariance of V. This allows us to study the behaviour transverse to the H-orbit, where we will be using once more the Remes inequality (Lemma 4.8). For this let $U_1, \ldots, U_\ell < H$ be unipotent one-parameter subgroups that together generate H: As H is assumed to be semisimple and without compact factors, we can find a unipotent subgroup projecting nontrivially to each simple factor of H. Conjugates of this unipotent subgroup will generate a normal subgroup of H and hence all of H.

As the sequence (μ_n) converging to μ is assumed to have no repetitions we may suppose that n is large enough to ensure that $H\centerdot x_n = \operatorname{supp}\mu_n \neq Y$. We may also suppose without loss of generality that the point x_n satisfies the ergodic theorem for $U_1,\dots,U_\ell$ and the characteristic functions of $\exp\bigl(B_\eta^V\bigr)Y$ and $\exp\bigl(B_\varepsilon^V\bigr)Y$ for $\varepsilon \in (0,\eta)\cap\mathbb{Q}$. If $x_n \notin \exp\bigl(B_\eta^V\bigr)Y$ we set $U = U_1$ in the argument below. So suppose that $x_n = \exp(v_1)y_1$ for some $v_1 \in B_\eta^V$ and $y_1 \in Y$. Then $v_1 \neq 0$ as $H\centerdot x_n \neq Y$ and v_1 is not fixed by H as $C_G(H)^o = \{I\}$. Hence there exists a unipotent one-parameter subgroup $U \in \{U_1,\dots,U_\ell\}$ that does not have v_1 as an eigenvector.

Let $T > 0$ and split

$$\Bigl\{t \in [0,T] \Bigm| u_t x_n \in \exp\bigl(B_\varepsilon^V\bigr)Y\Bigr\} = P_1 \sqcup \cdots \sqcup P_J$$

into its connected components (the ‘protecting intervals’). We note that for each $j \in \{1,\dots,J\}$ there is a vector $v_j \in V$ so that $x_n = \exp(v_j)y_j$ for some choice of $y_j \in Y$ and P_j is one of the connected components of

$$\bigl\{t \in [0,T] \bigm| \|\operatorname{Ad}_{u_t} v_j\| < \eta\bigr\}\,.$$

If $0 \in P_1$ our choice of U above ensures, for large enough T, that $T \notin P_1$. Equivalently we have

$$\sup_{t\in P_j} \|\operatorname{Ad}_{u_t} v_j\| = \eta$$

for $j = 1,\dots,J$. This allows us to apply the Remes inequality in each of the intervals P_j. As the intervals are disjoint we obtain as a result

$$\Bigl|\Bigl\{t \in [0,T] \Bigm| u_t x_n \in \exp\bigl(B_\varepsilon^V\bigr)Y\Bigr\}\Bigr| \ll \bigl(\frac{\varepsilon}{\eta}\bigr)^{\frac{1}{D}},$$

where the degree D depends on the action of the unipotent subgroups $U_1,\dots,U_\ell$ on V. We let $T \to \infty$ and it follows that for any $\delta > 0$ we can find a neighbourhood O_δ of Y so that $\mu_n(O_\delta) < \delta$ for all large enough n.

By Urysohn’s lemma and the definition of weak* convergence, this implies $\mu(Y) = 0$ for any (fixed) closed H-orbit Y. As explained earlier, we obtain from this that $\mu = m_X$ as desired. □

6.6 Transverse Divergence for the Horocycle Flow

[†]We will reprove the classification of invariant measures for the horocycle flow. Note that the assumption made in Theorem 6.34 below is weaker than the assumption in Theorem 5.7, 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 6.12 and the survey [136]).

Theorem 6.34 (Invariant measures for the horocycle flow). *Let Γ be a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$, let $X = \mathrm{SL}_2(\mathbb{R})/\Gamma$, and let*

$$U = \left\{ u_s = \begin{pmatrix} 1 & s \\ & 1 \end{pmatrix} \;\middle|\; s \in \mathbb{R} \right\}.$$

Suppose that μ is a U-invariant and ergodic probability measure on X. Then either

- *μ is supported on a single periodic orbit of U; or*
- *$\mu = m_X$ and Γ is a lattice.*

We want to apply an argument similar to the one in Section 6.3.3. 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 6.24). We start by generalizing Section 6.3.1.

Lemma 6.35 (Parallel orbits with a stretch). *Let $X = G/\Gamma$ 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 $x, y = g\centerdot x \in X$ 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

$$g = \begin{pmatrix} \lambda & b \\ & \lambda^{-1} \end{pmatrix}$$

for some $\lambda \neq 0$ and $b \in \mathbb{R}$. Indeed, if

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R})$$

[†] 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 a proof of Oppenheim's conjecture involving unipotent flows on X_3.

normalizes U, then we may calculate

$$\mathrm{Ad}_g \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} -ac & a^2 \\ -c^2 & ac \end{pmatrix}$$

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 6.6 for the geometrical picture of the proof.

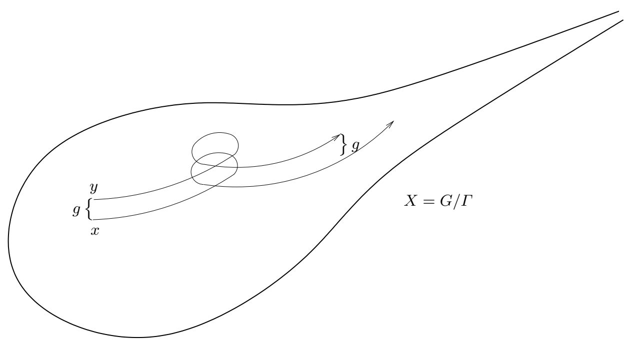

Fig. 6.6: If $y = g{\boldsymbol{\cdot}}x$ with $g \in N_G(U)$, then the two orbits are again parallel as in Figure 6.2, but $u_s{\boldsymbol{\cdot}}x$ may not be close to $u_s{\boldsymbol{\cdot}}y$ but may instead be close to $u_t{\boldsymbol{\cdot}}y$ for some t (depending on g and s).

Proof of Lemma 6.35. 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

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

is mapped under g to

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

As before, the normalized Lebesgue measure on $u_{[-T,T]}{\boldsymbol{\cdot}}x$ and $u_{[-|\lambda|T,|\lambda|T]}{\boldsymbol{\cdot}}y$ 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 6.34 we cannot hope in general for this propitious situation—the requirement that g lie 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

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

close to I with $ad - bc = 1$. However no other constraint will be assumed, and in particular c will be permitted to be non-zero. As we will see in the proof, we will be able to use 'transverse divergence' to produce additional invariance.

Proof of Theorem 6.34. Suppose $K \subseteq X$ is a set of uniformly generic points with measure $\mu(K) > 0.99$. We may assume that the genericity holds not only for forward orbits (as in Definition 6.21) but also for backward orbits. Suppose that $z \in \operatorname{supp} \mu|_K$. We split the proof into three stages.

Periodic orbit measures. If for some $\delta > 0$ we have

$$B_\delta(z) \cap K \subseteq U \centerdot x_0,$$

then $U \centerdot x_0$ has positive measure. This shows that x_0 is therefore a periodic orbit and μ is its normalized periodic orbit measure.

Additional invariance under the diagonal subgroup. Otherwise, it follows that we can choose $x_n \in K$ and $y_n \in K$ with

$$y_n = g_n \centerdot x_n$$

and $g_n \notin U$ for all $n \geqslant 1$ and $g_n \to I$ as $n \to \infty$. If we have

$$g_n = \begin{pmatrix} \lambda_n & b_n \\ 0 & \lambda_n^{-1} \end{pmatrix} \in N_G(U)$$

for infinitely many n (and with $\lambda_n \neq 1$ converging to 1 as $n \to \infty$) then we have parallel stretched orbits. By Lemma 6.35 we know that

$$A = \left\{ \begin{pmatrix} e^t & \\ & e^{-t} \end{pmatrix} \;\middle|\; t \in \mathbb{R} \right\}$$

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 would like to argue along the lines of Proposition 6.22, but we already learned from the proof of Lemma 6.35 that we might have to use different clocks for the parametrization of the orbits of x and y. We have seen before (see (2.8) in the proof of Proposition 2.17 on p. 69) the calculation that lies behind this:

$$\begin{aligned} u_t g_n u_{-s} &= \begin{pmatrix} 1 & t \\ & 1 \end{pmatrix} \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix} \begin{pmatrix} 1 & -s \\ & 1 \end{pmatrix} \\ &= \begin{pmatrix} a_n + c_n t & b_n + d_n t - s(a_n + c_n t) \\ c_n & d_n - c_n s \end{pmatrix}. \end{aligned} \tag{6.22}$$

If we set $s = t$ 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

$$s_n(t) = \frac{d_n t}{a_n + c_n t}.$$

Having made this choice, we obtain the simpler formula

$$u_t g_n u_{-s_n(t)} = \begin{pmatrix} a_n + c_n t & b_n \\ c_n & d_n - c_n s_n(t) \end{pmatrix}.$$

Once again we want to speed up the time parameter $t = T_n r$ by defining T_n to be $\frac{1}{c_n}$ and

$$\phi_n(r) = u_{T_n r} g_n u_{-s_n(T_n r)} = \begin{pmatrix} a_n + r & b_n \\ c_n & * \end{pmatrix}$$

for all $n \geqslant 1$. Here the $(2, 2)$ entry is also determined by the fact that (ϕ_n) is 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 $r = -a_n$, which is approximately -1. It follows that (ϕ_n) converges uniformly on compact subsets of $(-1, 1)$ to the rational function

$$\phi(r) = \begin{pmatrix} 1 + r & 0 \\ 0 & \frac{1}{1+r} \end{pmatrix}.$$

We claim that

$$\mu \text{ is invariant under } \phi(r) \quad \text{for any } r \in (-1, 1). \tag{6.23}$$

Once we know this we are at the same stage as in the previous special case.

To prove the claim in (6.23) we again apply a version of the 'last 1% argument'. Fix some $\varepsilon > 0$ and $f \in C_c(X)$. Then there exists a $\delta > 0$ with

$$\mathsf{d}(h_1, h_2) < \delta \implies |f(h_1 \centerdot x) - f(h_2 \centerdot x)| < \varepsilon \tag{6.24}$$

for all $x \in X$. Fix $r_0 > 0$ and notice that there exists some $\kappa \in (0, r_0)$ such that

$$\mathsf{d}\big(\phi(r), \phi(r_0)\big) < \tfrac{\delta}{2}$$

for all $r \in [r_0 - \kappa, r_0]$, which implies that

$$\mathsf{d}\big(\phi_n(r), \phi(r_0)\big) < \delta \tag{6.25}$$

for all sufficiently large n. By taking a linear combination of two ergodic averages while keeping κ fixed, we can deduce (just as in Section 6.3.3) that

$$\mathsf{A}_n = \frac{1}{\kappa T_n}\int_{(r_0-\kappa)T_n}^{r_0 T_n} f(u_t\bullet y_n)\,\mathrm{d}t \longrightarrow \int_X f\,\mathrm{d}\mu$$

as $n\to\infty$. We note that if $c_n<0$ and hence also $T_n<0$ we have used here the fact that points in K are also generic for the backward orbits. Now

$$f(u_t\bullet y_n) = f(u_t g_n u_{-s_n(t)}u_{s_n(t)}\bullet x_n) = f\bigl(\phi_n(c_n t)u_{s_n(t)}\bullet x_n\bigr)$$

is within ε of

$$f\left(\phi(r_0)u_{s_n(t)}\bullet x_n\right) = f^{\phi(r_0)}(u_{s_n(t)}\bullet x_n)$$

since we assume $c_n t\in[r_0-\kappa,r_0]$ and because of (6.25) and (6.24). Together we deduce that

$$\left|\mathsf{A}_n - \frac{1}{\kappa T_n}\int_{(r_0-\kappa)T_n}^{r_0T_n} f^{\phi(r_0)}\bigl(u_{s_n(t)}\bullet x_n\bigr)\,\mathrm{d}t\right| < \varepsilon.$$

The integral in this estimate is almost of the same form for x_n as the ergodic average A_n for y_n—except that the orbit is run through non-linearly. For that reason we now use the substitution $s_n(t)=\frac{d_n t}{a_n+c_n t}$. Its derivative is given by

$$\frac{\mathrm{d}s_n}{\mathrm{d}t} = \frac{a_n d_n}{(a_n+c_n t)^2},$$

which for large n and sufficiently small κ satisfies

$$\left|\phi'(r_0)^{-1}\frac{\mathrm{d}s_n}{\mathrm{d}t} - 1\right| < \varepsilon. \tag{6.26}$$

This shows that

$$\left|\mathsf{A}_n - \frac{1}{\phi'(r_0)\kappa T_n}\int_{(r_0-\kappa)T_n}^{r_0T_n} f^{\phi(r_0)}\bigl(u_{s_n(t)}\bullet x_n\bigr)\frac{\mathrm{d}s_n}{\mathrm{d}t}\,\mathrm{d}t\right| < \varepsilon\bigl(1+\|f\|_\infty\bigr),$$

and, equivalently,

$$\left|\mathsf{A}_n - \frac{1}{\phi'(r_0)\kappa T_n}\int_{\frac{(r_0-\kappa)d_nT_n}{a_n+r_0-\kappa}}^{\frac{r_0d_nT_n}{a_n+1}} f^{\phi(r_0)}\left(u_s\bullet x_n\right)\,\mathrm{d}s\right| < \varepsilon\bigl(1+\|f\|_\infty\bigr).$$

From (6.26) we also deduce that the length of the interval for the integral is asymptotic to $\phi'(r_0)\kappa T_n$ as $\varepsilon\to 0$. More precisely, integrating (6.26) with respect to $t\in[(r_0-\kappa)T_n, T_n]$ and dividing by κT_n gives

$$\left|\frac{\text{total length}}{\phi'(r_0)\kappa T_n} - 1\right| < \varepsilon.$$

Applying the linear combination argument to the intervals

$$\left[0, \frac{r_0 d_n T_n}{a_n + r_0}\right] \text{ and } \left[0, \frac{(r_0 - \kappa) d_n T_n}{a_n + r_0 - \kappa}\right],$$

the initial point x_n and the function $f^{\phi(r_0)}$ gives

$$\frac{1}{\text{total length}} \int_{\frac{(r_0-\kappa)d_n T_n}{a_n+r_0-\kappa}}^{\frac{r_0 d_n T_n}{a_n+1}} f^{\phi(r_0)}(u_s \cdot x_n)\, \mathrm{d}s \longrightarrow \int_X f^{\phi(r_0)}\, \mathrm{d}\mu$$

as $n \to \infty$. Together we see after taking the limits as $n \to \infty$ that

$$\left| \int_X f^{\phi(r_0)}\, \mathrm{d}\mu - \int_X f\, \mathrm{d}\mu \right| < O(\varepsilon)\big(1 + \|f\|_\infty\big),$$

and this holds for any $\varepsilon > 0$ and $f \in C_c(X)$. Hence we have shown (unless μ is the Lebesgue measure on a periodic U-orbit) that μ is invariant under the diagonal subgroup A.

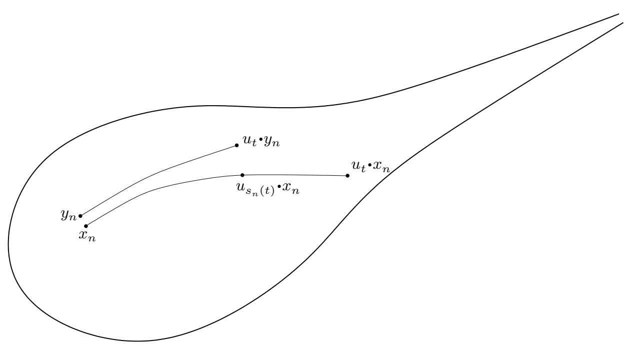

Fig. 6.7: Using the same time for x_n and y_n would make one orbit overtake the other in the orbit direction. By using different times we can ensure that they instead diverge transversally. In the last 1% the orbits are almost parallel and differ in the direction of A.

Conclusion. Using the fact that

$$A = \left\{ a_t = \begin{pmatrix} \mathrm{e}^t & \\ & \mathrm{e}^{-t} \end{pmatrix} \;\middle|\; t \in \mathbb{R} \right\}$$

preserves μ we wish to show that $\mu = m_X$ is the normalized Haar measure on $X = \mathrm{SL}_2(\mathbb{R})/\Gamma$.

Recall that μ was assumed to be ergodic under U. Also note that U belongs to the Auslander ideal of $B = AU$ with respect to any non-trivial element of A.

Hence the Mautner phenomenon in Proposition 2.22 implies that A also acts ergodically. Let

$$X_{\mu,A}^{+} = \left\{ x \in X \;\middle|\; \lim_{N\to\infty} \frac{1}{N}\int_0^N f(a_t \boldsymbol{\cdot} x)\,\mathrm{d}t = \int f\,\mathrm{d}\mu \text{ for all } f \in C_c(X) \right\}$$

be the set of μ-generic points for the forward orbit under A. Note that both $X_{\mu,A}^{+}$ and

$$X_\mu = \left\{ x \in X \;\middle|\; |\{t \in (-1,1) \mid u_t \boldsymbol{\cdot} x \in X_{\mu,A}^{+}\}| = 2 \right\}$$

have full measure by invariance under U and Fubini's theorem. We fix some point $x_\mu \in X_\mu$.

Suppose for a moment that Γ is a lattice. Let $X_{m_X,A}^{+}$ be the set of m_X-generic points for the forward orbit under A. Then $X_{m_X,A}^{+}$ has full measure for m_X and is invariant under the diagonal subgroup as well as the opposite stable horocyclic subgroup. Together with the product structure for $m_{\mathrm{SL}_2(\mathbb{R})}$ in Lemma 5.13 this implies that $u_t \boldsymbol{\cdot} x_\mu \in X_{m_X,a}^{+}$ for almost every $t \in (-1,1)$. This shows that we can find a point that is simultaneously generic for μ and for m_X, which shows that $\mu = m_X$.

Suppose now that Γ is not a lattice. We wish to mimic the argument above using only certain subsequences of ergodic averages under A, and end up with the contradiction that $\mu = 0$ instead. To begin, let $f \in C_c(X)$ and note that $L^2_{m_X}(X)$ does not contain any non-trivial invariant vectors. Hence the Howe–Moore theorem (Theorem 2.42) applies in a stronger form and gives $\langle \pi_{a_t} f, f \rangle \to 0$ as $t \to \infty$. We define the function

$$\mathsf{A}_N^f = \frac{1}{N}\int_0^N f \circ a_t \,\mathrm{d}t$$

and calculate

$$\begin{aligned}
\left\| \mathsf{A}_N^f \right\|_{L^2}^2 &= \left\langle \frac{1}{N}\int_0^{N-1} f\circ a_t\,\mathrm{d}t, \frac{1}{N}\int_0^{N-1} f \circ a_s\,\mathrm{d}s \right\rangle \\
&= \frac{1}{N^2}\iint_{[0,N]^2} \left\langle \pi_{a_{-t}} f, \pi_{a-s} f \right\rangle \mathrm{d}s\,\mathrm{d}t \\
&= \frac{1}{N^2}\iint_{[0,N]^2} \left\langle \pi_{a_{s-t}} f, f \right\rangle \mathrm{d}s\,\mathrm{d}t \longrightarrow 0
\end{aligned}$$

as $N \to \infty$. We choose a subsequence $N_k \nearrow \infty$ for $k \to \infty$ so that

$$\left\| \mathsf{A}_{N_k}^f \right\|_{L^2}^2 = \int_X \left| \mathsf{A}_{N_k}^f \right|^2 \mathrm{d}m_X < \frac{1}{k^4}$$

for all $k \geqslant 1$. This implies that

$$m_X\left(\left\{x \in X \;\middle|\; |\mathsf{A}^f_{N_k}| \geqslant \tfrac{1}{k}\right\}\right) < \frac{1}{k^2}$$

and so allows us to apply Borel–Cantelli. Define

$$X^+_{f,A} = \left\{x \in X \;\middle|\; \lim_{k\to\infty} \mathsf{A}^f_{N_k}(x) = 0\right\}$$

and obtain that $m_X\bigl(X \smallsetminus X^+_{f,A}\bigr) = 0$. Moreover, by definition $X^+_{f,A}$ is invariant under the diagonal subgroup (which would only shift our integrals defining A^f_N by a bounded amount) and the opposite stable horocyclic subgroup in $\mathrm{SL}_2(\mathbb{R})$ (which would create asymptotic orbits). Once more we see that for almost every $t \in (-1,1)$ we have $u_t{\boldsymbol\cdot}x_\mu \in X^+_{\mu,A} \cap X^+_{f,A}$. This implies that the ergodic average under a of $u_t{\boldsymbol\cdot}x_\mu$ converges to $\int f\,\mathrm{d}\mu$ and (along a subsequence) converges to 0. As this argument applies to any $f \in C_c(X)$ we obtain the contradiction $\mu = 0$. □

6.7 A Non-Horospherical Case and 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 the quotient space $X = G/\Gamma$ where $\Gamma < G$ is a lattice. Up to conjugation, G allows three different choices of one-parameter unipotent flows:

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

The first two are actually horospherical subgroups so the discussion in Chapter 5 applies to these cases. Thus we will only consider the third (most difficult) case (which of course is a special case of Ratner's measure classification in Theorem 6.12).

Theorem 6.36 (Diagonally embedded horocycles). *Let*

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

let Γ be a lattice, and define the quotient space $X = G/\Gamma$. 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 note that the word 'algebraic' subsumes here several different possibilities. To see this we suppose for a moment that $\Gamma_1, \Gamma_2 < \mathrm{SL}_2(\mathbb{R})$ are lattices and let $\Gamma = \Gamma_1 \times \Gamma_2$ be their product. Write $X_i = \mathrm{SL}_2(\mathbb{R})/\Gamma_i$ and consider the projections $\pi_1\colon (x,x') \mapsto x$ and $\pi_2\colon (x,x') \mapsto x'$ from

$$X = \mathrm{SL}_2(\mathbb{R})/\Gamma_1 \times \mathrm{SL}_2(\mathbb{R})/\Gamma_2$$

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 5.7 these measures are therefore known to be algebraic (on a periodic U-orbit or Haar on X_i), which leads us to three cases.

(i) The measures 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 the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$.
(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. By 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 Appendix C this implies that

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

almost surely. To summarize we have shown 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.

Hence it remains to study the joinings. As it turns out, in this case the stabilizer of μ is either G or the graph of conjugation by $\begin{pmatrix} 1 & s \\ & 1 \end{pmatrix}$ for some $s \in \mathbb{R}$. To see this we have to study transverse divergence again.

6.7.1 Using Transverse Divergence

In the following we let $X = G/\Gamma$, $U < G$, and let μ be as in Theorem 6.36. Moreover, we always let K be a set of uniformly generic points in X of measure $\mu(K) > 0.99$. As in the last section, we now want to study how the U-orbits of points in K move apart, where we allow different time parameters for the two orbits.

For the argument to follow it will be convenient to write displacements in a particular coordinate system that favours the diagonally embedded copy of $\mathrm{SL}_2(\mathbb{R})$. Hence we define the subgroup

$$\Delta = \{\Delta(g) = (g,g) \mid g \in \mathrm{SL}_2(\mathbb{R})\} < G = \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$$

and note that any element of G can be written in the form

$$(g,h) = \Delta(g)(I, g^{-1}h) \in \Delta V,$$

where

$$V = \left\{ v = (I, g') \bigm| g' \in \mathrm{SL}_2(\mathbb{R}) \right\}$$

is an invariant complement to Δ and in fact a normal subgroup of G. To simplify notation we write

$$v\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \left(I, \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \right) \in V$$

for the elements of the second copy of $\mathrm{SL}_2(\mathbb{R})$. With this any element $g \in G$ can be written in the form

$$g = \Delta\begin{pmatrix} a & b \\ c & d \end{pmatrix} v \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \tag{6.27}$$

with $\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R})$.

Using these notations we now calculate for $x, y \in K$ with $y = g{\scriptstyle\bullet}x$ that

$$u_t{\scriptstyle\bullet}y = u_t g u_{-s}{\scriptstyle\bullet}(u_s{\scriptstyle\bullet}x)$$

and for g as in (6.27) we obtain (by applying the calculation (6.22) twice)

$$\begin{aligned} u_t g u_{-s} &= u_t \Delta \begin{pmatrix} a & b \\ c & d \end{pmatrix} u_{-s} u_s v \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} u_{-s} \\ &= \Delta \begin{pmatrix} a+ct & b+dt-s(a+ct) \\ c & d-cs \end{pmatrix} v \begin{pmatrix} a'+c's & b'+(d'-a')s-c's^2 \\ c' & d'-c's \end{pmatrix}. \end{aligned}$$

We again set $s = \frac{dt}{a+ct}$, so that the above simplifies to

$$\phi(t) = \Delta \begin{pmatrix} a+ct & b \\ c & d-cs \end{pmatrix} v \begin{pmatrix} a'+c's & b'+(d'-a')s-c's^2 \\ c' & d'-c's \end{pmatrix}, \tag{6.28}$$

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

Once again we need to speed up the time parameter t by setting $t = Tr$ 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 $\phi(t)$ 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 we can also do this abstractly.

Lemma 6.37. *Let $G \leqslant \mathrm{SL}_d(\mathbb{R})$ be a linear group, let $U = \{u(s) \mid s \in \mathbb{R}\}$ be a one-parameter unipotent subgroup. Fix some $g \in G$. Suppose also that $s = s(t)$ is defined on an open neighbourhood of $0 \in \mathbb{R}$ (for example, by a rational function) such that $s(0) = 0$ and*

$$\phi(s) = u(t) g u(-s(t))$$

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

Proof. Let $t \in \mathbb{R}$ be close to 0. Then

$$\phi(t) = u(t) g u(-s(t)) = \phi(0) = g$$

is equivalent to

$$g^{-1} u(t) g = u(s(t)) \in U.$$

As this holds for all t near 0 and U is connected, the lemma follows. □

Suppose for a moment that $\phi(t)$ as in (6.28) is indeed constant. Then we have $g \in N_G(U)$, and by Lemma 6.35 we also have $g \in \mathrm{Stab}_G(\mu)$. Suppose that $x_0 \in \operatorname{supp} \mu|_K$, and that we are in this case for all $y = g \cdot x_0 \in K$ sufficiently close to x. Then $\mathrm{Stab}_G(\mu) \cdot x$ has positive measure, and hence has full measure and the theorem follows from Lemma 6.17.

It remains to consider the case where there is a sequence (y_n) with

$$y_n = g_n \cdot x_0 \in K$$

[†] For example, if $c > 0$ is much smaller than c' this will happen.

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 (using g_n) is not constant. This also requires us to define s_n as a rational function of t with denominator $a_n + c_n t$. Moreover, ϕ_n only contains (apart from the constants and t) the terms s_n and s_n^2. Therefore

$$\Phi_n(t) = \Big(a_n + c_n t, (a_n + c_n t)^2 \phi_n(t)\Big)$$

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

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

where we use the maximum norm on all entries of Φ_n. We note that this implies in particular that $|c_n| T_n \leqslant 1$ (by looking at the first component of $\Phi_n(T) - \Phi(0)$).

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,

$$\frac{1}{(1 + \alpha r)^2} \psi_0(r) = \psi(r)$$

is the limit of the sequence of functions $(r \mapsto \phi_n(T_n r))$ uniformly on compact subsets of $(-1, 1)$. Now we can ask for the behaviour of the function

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

for $r \in (-1, 1)$, without calculating it explicitly.

Lemma 6.38. *Let $G \leqslant \mathrm{SL}_d(\mathbb{R})$ be a closed linear group. Let*

$$U = \{u(t) \mid t \in \mathbb{R}\} < G$$

be a one-parameter unipotent subgroup. 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 $\big(s_n \colon \mathbb{R} \to \mathbb{R}\big)$ such that

$$\psi_n(r) = u(T_n r) g_n u(-s_n(T_n r))$$

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

converges uniformly on some open subset $O \subseteq \mathbb{R}$ to some function $\psi\colon O \to G$. Then $\psi(O) \subseteq N_G(U)$.

Proof. Let $t_0 \in \mathbb{R}$ and define $\varepsilon_n = \frac{t_0}{T_n}$ so that $\varepsilon_n \to 0$ as $n \to \infty$. By uniform convergence this implies for $r \in O$ that

$$\begin{aligned}\psi(r) &= \lim_{n\to\infty} \psi_n(r+\varepsilon_n)\\ &= \lim_{n\to\infty} u(T_n r + t_0) g_n u(-s_n(T_n r + t_0))\\ &= \lim_{n\to\infty} \underbrace{u(T_n r) g_n u(-s_n(T_n r))}_{\psi_n(r)} u\big(s_n(T_n r) - s_n(T_n r + t_0)\big)\\ &= u(t_0)\psi(r) u_{r,t_0}\end{aligned}$$

for some $u_{r,t_0} \in U$. In particular we have $\psi(r)^{-1} u(t_0)\psi(r) \in U$. As this holds for all $u \in U$, the lemma follows. □

Using the same last 1% argument as in the proof of Theorem 6.34 it now follows that μ is invariant under all elements $\psi(r)$ for all $r \in (-1,1)$. Let us use (6.28) to be more specific concerning the structure of ψ. By definition of Φ_n we have that $a_n + c_n T_n r$ converges uniformly on compact subsets to $1 + \alpha r$ with $|\alpha| \leqslant 1$ as $n \to \infty$. This already determines the Δ-portion of ψ to be

$$\Delta\left(\begin{pmatrix} 1+\alpha r & \\ & \frac{1}{1+\alpha r}\end{pmatrix}\right).$$

For the v-portion we note that

$$V \cap N_G(U) = \left\{ \pm \begin{pmatrix} 1 & * \\ & 1 \end{pmatrix} \right\}.$$

Moreover the sequence of rational functions in the upper right entry has common denominator $a_n + c_n T_n r \to 1 + \alpha r$ and converges uniformly on $[-\frac{1}{2}, \frac{1}{2}]$ if and only if the coefficients in the numerator converge. To summarize, it follows that

$$\psi(r) = \Delta \begin{pmatrix} 1+\alpha r & \\ & \frac{1}{1+\alpha r}\end{pmatrix} v \begin{pmatrix} 1 & \beta(r) \\ & 1 \end{pmatrix}$$

for the rational function $\beta(r) = \beta_1 \frac{r}{1+\alpha r} + \beta_2 \big(\frac{r}{1+\alpha r}\big)^2$ defined by some $\beta_1, \beta_2 \in \mathbb{R}$. We now consider two separate cases as follows.

Case I: Assume first that $\alpha = 0$ so that

$$\psi(r) = \Delta \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} v \begin{pmatrix} 1 & \beta(r) \\ & 1 \end{pmatrix}$$

for a nonconstant polynomial $\beta(r)$. This shows 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 the arguments in Chapter 5. We note that if along our sequence the displacement always has the form $g = \Delta(I)v(h)$ for some

$$h \neq \begin{pmatrix} 1 & * \\ & 1 \end{pmatrix},$$

then we will be in this Case I.

Case II: Suppose now $\alpha \in [-1, 1]$ is non-zero so that

$$\psi(\tfrac{1}{2}) = \Delta \begin{pmatrix} \lambda & \\ & \lambda^{-1} \end{pmatrix} v \begin{pmatrix} 1 & c \\ & 1 \end{pmatrix}$$

for some positive $\lambda \neq 1$ and $c \in \mathbb{R}$. We claim that we may assume in the following that c is zero. In fact, replacing μ by

$$v \begin{pmatrix} 1 & s \\ & 1 \end{pmatrix}_* \mu$$

gives a new measure that is still invariant under U and is also invariant under

$$v \begin{pmatrix} 1 & s \\ & 1 \end{pmatrix} \left(\begin{pmatrix} \lambda & \\ & \lambda^{-1} \end{pmatrix}, \begin{pmatrix} \lambda & c \\ & \lambda^{-1} \end{pmatrix}\right) v \begin{pmatrix} 1 & -s \\ & 1 \end{pmatrix} = \left(\begin{pmatrix} \lambda & \\ & \lambda^{-1} \end{pmatrix}, \begin{pmatrix} \lambda & c + \lambda^{-1}s' - \lambda s' \\ & \lambda^{-1} \end{pmatrix}\right).$$

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 2.22 and obtain that f is also U-invariant and so constant by our assumption on μ.

6.7.2 A Proof for Product Spaces

We note that in Section 6.7.3 we will give a second independent and general proof of Theorem 6.36, but we first wish to complete our discussions of joinings in (iii) from page 276.

Classifying joinings. We assume that $X = X_1 \times X_2$ and that

$$(\pi_i)_* \mu = m_{X_i}$$

for $i = 1, 2$. By the beginning of the proof in Section 6.7.1, 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

$$g_n = v \begin{pmatrix} a'_n & b'_n \\ c'_n & d'_n \end{pmatrix} \neq I$$

with

$$\begin{pmatrix} a'_n & b'_n \\ c'_n & d'_n \end{pmatrix} \longrightarrow I$$

as $n \to \infty$. We now apply the argument from Section 6.7.1 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 6.18), 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

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

by [46, 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

[†] 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'.

[§] 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.

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

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 $\mu_{(x,\cdot)}^{\mathcal{A}}$ 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

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

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$f_1(x) \in \overline{Q_n}$,

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

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 $f_1, \dots, f_m$. To simplify the notation for these arguments we set

$$u = \begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix} \text{ and } a = \begin{pmatrix} \lambda & \\ & \lambda^{-1} \end{pmatrix}$$

with $\lambda \in (0, 1)$ as before. Since we already know that (u, u) and (a, a) preserve μ (which is determined by m_{X_1} and the functions $f_1, \dots, f_m$), we get the 'equivalence properties' for the functions $f_1, \dots, f_m$. In fact

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

almost surely, and similarly with u replaced by a. Indeed, by [46, 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

$$\{f_1(v_t{\bullet}x), \dots, f_m(v_t{\bullet}x)\} = v_t{\bullet}\{f_1(x), \dots, f_m(x)\}, \tag{6.29}$$

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 [46, 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

$$\begin{aligned} f_1(a^{-n}v_t\bullet x) &= a^{-n}gf_1(x) = a^{-n}ga^n\bullet f_1(a^{-n}\bullet x) \\ &= f_1(v_{\lambda^{2n}t}a^{-n}\bullet x). \end{aligned} \tag{6.30}$$

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 (6.30) 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 $n\to\infty$.

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

$$u^\ell\bullet f_1(v_t\bullet x) = \bigl(u^\ell g u^{-\ell}\bigr)u^\ell\bullet f_1(x)$$

is very close to $u^\ell\bullet f_1(x)$. 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

$$|c| \ll |t|, |d-a| \ll \sqrt{|t|}, |b| \ll 1.$$

† The geodesic flow has many pairs of orbits that are close for a large percentage of time without being close for a good (meaning algebraic) reason.

‡ It might appear disadvantageous to use u instead of a^{-1}, since $a^{-n}v_ta^n$ 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.

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

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

and

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

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

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

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 6.39 (Linearization for the correspondence). *Depending on K there exists some $\delta > 0$ such that for all*

$$y = v_t \boldsymbol{\cdot} x, x \in Y_1 \cap K$$

with $t \in (-\delta, \delta)$ and all i there exists j such that

$$f_i(y) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \boldsymbol{\cdot} f_j(x)$$

for some $a, b, c, d \in \mathbb{R}$ with $|c| \ll |t|$, $|a - 1|, |d - 1| \ll \sqrt{|t|}$ and $|b| \ll 1$.

In the proof we will use the fact that $y = v_t \boldsymbol{\cdot} x$ and x satisfy that $u^\ell \boldsymbol{\cdot} x$ and $u^\ell \boldsymbol{\cdot} y$ are close together as long as $\ell^2 t$ is small. Applying $f_1, \dots, f_m$ we have the weaker property that the image points are close in X_2 for some fixed percentage of this time window (if $m = 1$ this percentage would be 80%). Here we will need the following lemma.

Lemma 6.40 (Linearization for two orbits). *Let $X = \Gamma \backslash \mathrm{SL}_2(\mathbb{R})$ be a quotient by a lattice. For any $p \in (0,1)$ and any compact subset $K \subseteq X$ there exists some $\kappa \in (0,1]$ with the following property. Suppose that $L > 1$, the points $x \in K$ and $y \in X$ satisfy*

$$\frac{1}{L}\left|\left\{\ell \in \{0,\dots,L-1\} \;\middle|\; u^\ell \bullet x \in K \text{ and } \mathsf{d}(u^\ell \bullet x, u^\ell \bullet y) < \kappa\right\}\right| \geqslant p.$$

Then $y = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \bullet x$ 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

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

For $\ell \in S$ we let $g_\ell \in \mathrm{SL}_2(\mathbb{R})$ be the unique matrix satisfying $u^\ell \bullet y = g_\ell u^\ell \bullet x$ and

$$\mathsf{d}\left(g_\ell, I\right) = \mathsf{d}\left(u^\ell \bullet y, u^\ell \bullet x\right) < \rho.$$

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

$$|b + (d-a)\ell - c\ell^2| \ll \rho$$

† 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.

for some absolute implied constant, which depends only on the Riemannian metric. By Lemma† 4.8 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 $|(d - a)L| \ll_p 1$.

To prove that S contains only one equivalence class, we assume the opposite, choose κ sufficiently small and will again use Lemma 4.8 to derive a contradiction. In fact by that lemma we may choose $\kappa < \rho$ so that

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

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

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

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 $\{0, \dots, L\}$ on which $\mathsf{d}(u^{k-\ell} g_\ell u^{-(k-\ell)}, I) \leqslant \rho$ for all $k \in P_{[\ell]}$. By definition $[\ell] \subseteq P_{[\ell]}$. We may also assume that for each equivalence class $[\ell]$ and its interval $P_{[\ell]}$ we have $\mathsf{d}(u^{k-\ell} g_\ell u^{-(k-\ell)}, I) \geqslant \rho$ for k equal to the left end point minus one or

† 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.

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,

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

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

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

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 6.36 and apply the lemma above.

Proof of Lemma 6.39. Since K is compact and the functions $f_1, \ldots, f_m$ 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 6.40 to

$$X = X_2 = \Gamma_2 \backslash \mathrm{SL}_2(\mathbb{R})$$

and the compact set K'. This defines some $\kappa > 0$. Since $f_i(x) \neq f_j(x)$ for $i \neq j$ and all x in the domain of these functions by construction, we may also suppose that

$$\mathsf{d}(f_i(x), f_j(x)) > 2\kappa$$

for $x \in K$ and $i \neq j$. Again since f_i restricted to K is continuous we see that there exists a $\delta > 0$ such that

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

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

Suppose now that $t \in (-\delta, \delta)$ and $x, v_t\cdot x \in K \cap Y_1$. We can now find an interval $I_{x,y}$ of length $\gg_\delta 1/\sqrt{|t|}$ such that for $\ell \in I_{x,y}$ we have

$$\mathsf{d}(u^\ell v_t u^{-\ell}, I) = \mathsf{d}\left(\begin{pmatrix} 1+\ell t & \ell^2 t \\ t & 1-\ell t \end{pmatrix}, I\right) < \delta,$$

and (by definition of Y_1) for $\frac{8}{10}$ of all $\ell \in I_{x,y}$ we have $u^\ell\cdot x, u^\ell\cdot y \in K$ and so

$$\mathsf{d}\big(f_i(u^\ell\cdot y), f_i(u^\ell\cdot x)\big) < \kappa$$

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

$$\mathsf{d}\big(u^\ell\cdot f_i(y), u^\ell\cdot f_j(x)\big) < \kappa. \tag{6.31}$$

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 (6.31) holds (with j independent of s). Applying Lemma 6.40 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 6.36. 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 6.39 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

$$\begin{aligned} |\lambda^{-2n} c_n| &\ll t, \\ |a_n - 1| &\ll \lambda^n \sqrt{|t|}, \\ |d_n - 1| &\ll \lambda^n \sqrt{|t|}, \end{aligned}$$

and

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

Here it is crucial that the entries of the matrix $a^n g_n a^{-n}$ are uniformly bounded. Hence we may choose a subsequence such that $a^n g_n a^{-n}$ converges and $j = j(n)$ is constant along this subsequence. It follows that we have shown that for every i and every pair $y = v_t \cdot x, x \in Y_2$ there exists some $j = j(x,t,i), c = c(x,t,i) \in \mathbb{R}$ 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 every set of sufficiently large measure we find points

$$(x, x') = (x, f_j(x))$$

and

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

with $t \neq c$. Applying the transverse divergence argument from the beginning of the proof on pages 277–281 we end up in Case I and 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 (6.29) holds almost surely. Let us now show that this implies

[†] As was mentioned before, we do not know at this stage any relationship between these displacement g_n for different values of n.

that μ is invariant under (v_t, v_t) for any $t \in \mathbb{R}$. So let $f \in C_c(X_1 \times X_2)$. Then

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

where we used in order the definition of the conditional measures, our description of them, (6.29) for m_{X_1}-almost every x, and the fact that v_t preserves m_{X_1}.

Now note that U as in Theorem 6.36 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 6.32 applies and shows that μ is algebraic.

6.7.3 Proof in the General Case

We will follow the arguments of Ratner [132] in the case $G = \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$. Let $\Gamma < G$ be an arbitrary lattice and suppose that μ is a U-invariant and ergodic probability measure on $X = G/\Gamma$ as in Theorem 6.36. If μ is invariant under the horospherical subgroup

$$\left\{\left(\begin{pmatrix}1 & *\\ & 1\end{pmatrix}, \begin{pmatrix}1 & *\\ & 1\end{pmatrix}\right)\right\}$$

then we can apply our discussions of Chapter 5.

Hence we will assume from now on that μ is *not* invariant under the horospherical subgroup and so is also *not* invariant under

$$\left\{ v\begin{pmatrix}1 & b\\ & 1\end{pmatrix} \;\middle|\; b \in \mathbb{R}\right\}.$$

We apply the discussions of Section 6.7.1. Unless μ is the Lebesgue measure on a periodic U-orbit this implies that μ has additional invariance. As we assume that μ is not invariant under the horosphere we must be in Case II of page 281. As explained there we may replace μ by a push forward of μ under an element of the form

$$v\begin{pmatrix}1 & *\\ & 1\end{pmatrix}$$

and assume that μ is invariant under the connected component $\Delta(A)$ of the diagonally embedded diagonal subgroup.

Analyzing Section 6.7.1 more carefully in combination with our assumption that μ is not invariant under the horosphere we obtain the following 'geometric' statement.

Lemma 6.41 (Ratner's basic lemma). *Let $K \subseteq X$ be a set of uniformly generic points. Suppose that μ is a U-invariant and ergodic probability measure that is invariant under $\Delta(A)$ but not invariant under the horospherical subgroup. Then for any sufficiently small $\varepsilon > 0$ we have that $x, y = g{\scriptstyle\bullet} x \in K$ and*

$$g = \Delta\begin{pmatrix} a & b \\ c & d \end{pmatrix} v \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}$$

with $\max\bigl(|a|, |d|, |b|, |b'|\bigr) \leqslant 2$, $|c| \leqslant \varepsilon$, $\max\bigl(|d'-a'|, |c'|\bigr) \leqslant 10\varepsilon^{\frac{1}{2}}$ *imply that* $|d'-a'| < \frac{1}{10}\varepsilon$ *and* $|c'| < \frac{1}{100}\varepsilon^2$. *Moreover, as $\varepsilon \to 0$ we also have that $b' \to 0$.*

Proof. Suppose first that the final claim fails. Then by compactness there exist sequences (x_n) and (y_n) in K with $x_n \to x \in K$ and $y_n = g_n{\scriptstyle\bullet} x_n \to y \in K$ as $n \to \infty$ and

$$g_n = \Delta\begin{pmatrix} a & b \\ c & d \end{pmatrix} v \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \longrightarrow \Delta\begin{pmatrix} a & b \\ 0 & \frac{1}{a} \end{pmatrix} v \begin{pmatrix} 1 & b' \\ 0 & 1 \end{pmatrix}$$

as $n \to \infty$ with $b' \neq 0$. By Lemma 6.35 this shows that g preserves μ. Recalling that $\Delta(A)$ also preserves μ, this also implies that μ is preserved under the horospherical subgroup, contradicting our assumption.

The proof of the main estimates in the lemma is similar in spirit (but using Section 6.7.1). Indeed, if the lemma fails then we can find sequences $\varepsilon_n \searrow 0$, (x_n) and (y_n) in K with $y_n = g_n{\scriptstyle\bullet} x_n$ satisfying

$$\max\bigl(|a_n|, |d_n|, |b_n|, |b_n'|\bigr) \leqslant 2, \quad |c_n| \leqslant \varepsilon_n \tag{6.32}$$

and

$$\frac{1}{10}\varepsilon_n \leqslant \max\bigl(|d_n' - a_n'|, |c_n'|^{1/2}\bigr) \ll \varepsilon_n^{\frac{1}{4}} \tag{6.33}$$

for all $n \geqslant 1$. We use these sequences to define ϕ_n as in (6.28), $s_n = \frac{d_n t}{a_n + c_n t}$, and the speed-up parameter $T_n \to \infty$ as $n \to \infty$. Writing simply s_n for

$$s_n(T_n r) = \frac{d_n T_n r}{a_n + c_n T_n r}$$

and choosing a subsequence we may assume that

$$\phi_n(T_n r) = \Delta\begin{pmatrix} a_n + c_n T_n r & b_n \\ c_n & d_n - c_n s_n \end{pmatrix} v \begin{pmatrix} a_n' + c_n' s_n & b_n' + (d_n' - a_n') s_n - c_n' s_n^2 \\ c_n' & d_n' - c_n' s_n \end{pmatrix}$$

converges to

$$\varDelta \begin{pmatrix} 1+\alpha r & \\ & \frac{1}{1+\alpha r} \end{pmatrix} v \begin{pmatrix} 1 & \beta_1 \frac{r}{1+\alpha r} + \beta_2 \left(\frac{r}{1+\alpha r}\right)^2 \\ & 1 \end{pmatrix}$$

as $n \to \infty$, where

$$\begin{aligned} \alpha &= \lim_{n\to\infty} c_n T_n, \\ \beta_1 &= \lim_{n\to\infty} (d_n' - a_n') T_n, \text{ and} \\ \beta_2 &= \lim_{n\to\infty} c_n' T_n^2. \end{aligned}$$

By our non-invariance assumption we must have $\beta_1 = \beta_2 = 0$, which implies $|\alpha| = 1$ and hence $T_n > \frac{1}{2|c_n|} \geqslant \frac{1}{2}\varepsilon_n$ for sufficiently large n by (6.32). Comparing this with $\beta_1 = \beta_2 = 0$ and the lower bound in (6.33) gives a contradiction. □

Notes to Chapter 6

(32) (Page 231) Arguably this was a rediscovery of a connection between Diophantine problems of this sort and homogeneous dynamics used earlier by Artin [3] and by Cassels and Swinnerton-Dyer [14]. We refer to a survey by Margulis [112] for a motivated history.

(33) (Page 237) We refer to Linnik's original book [106] and to later work by Einsiedler, Lindenstrauss, Michel, and Venkatesh [44], by Ellenberg, Michel, and Venkatesh [51] and by Wieser [169].

(34) (Page 239) This appeared in print in the work of Dani [20, Conjecture II].

(35) (Page 243) This is an instance of a more general result due to Weyl [168] giving equidistribution modulo one for the values on the natural numbers of any polynomial with an irrational coefficient. Furstenberg [58] showed that this followed from a general result extending unique ergodicity from irrational circle rotations to certain maps on tori. We refer to [46, Sec. 4.4.3] for a detailed discussion.