# Chapter 5
# Horospherical Subgroups and Counting Results

The inheritance property of ergodicity of the Mautner phenomenon in Proposition 2.25 and in the general Theorem 2.55 (also see Exercise 2.24 and 2.45) established in Chapter 2 already gives the equidistribution of many orbits.

Indeed, if a simple Lie group $G$ acts ergodically on $(X, \mu)$ and

$$\{g_t \mid t \in \mathbb{R}\} \subseteq G$$

is an unbounded one-parameter subgroup, then

$$\frac{1}{T}\int_0^T f(g_t \centerdot x)\,\mathrm{d}t \longrightarrow \int_X f\,\mathrm{d}\mu$$

for $\mu$-almost every $x \in X$, for any $f \in C_c(X)$ as $T \to \infty$. A point $x \in X$ with this property is called *generic* for $\mu$ and the one-parameter subgroup $\{g_t \mid t \in \mathbb{R}\}$.

In this chapter we start the discussion of unipotent dynamics by considering the case of horospherical actions. For those actions we will show 'unique ergodicity', and sometimes 'almost unique ergodicity', and we will understand precisely which points are generic for $m_X$. The method of proof also gives other equidistribution results of certain 'distorted orbits', which in turn can be used to prove asymptotic counting results. Hence in the second half of the chapter we will explain the set-up of Duke, Rudnick, and Sarnak and its dynamical interpretation by Eskin and McMullen.

## 5.1 Dynamics on Hyperbolic Surfaces

Let us start by discussing briefly the case of the geodesic flow and the horocycle flow on quotients of $\mathrm{SL}_2(\mathbb{R})$ as introduced in Section 1.2.

### 5.1.1 The Geodesic Flow

We note first that for the geodesic flow defined by the diagonal subgroup it is not possible to make a more general statement about the equidistribution of orbits by relaxing the requirement that the point be $\mu$-typical. Indeed, in this case the flow is partially hyperbolic and as a result $X$ contains many irregular orbits. As this result can be considered of negative type we will not prove it here, but refer to [45, Sec. 9.7.2] for a more detailed discussion of the case of the geodesic flow on the modular surface.

*Example 5.1.* For a compact quotient $X$ of $\mathrm{SL}_2(\mathbb{R})$ by a uniform lattice as in Figure 5.1, the action of the one-parameter subgroup

$$A = \left\{ a_t = \begin{pmatrix} \mathrm{e}^{-t/2} & \\ & \mathrm{e}^{t/2} \end{pmatrix} \,\middle|\, t \in \mathbb{R} \right\} \tag{5.1}$$

has many orbits that, for example, stay on one side of the dotted line.[27]

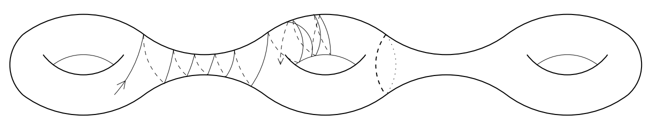

**Fig. 5.1:** There are many orbits under the action of $A$ that stay on one side of the dotted line furthest to the right.

We also refer to Exercises 5.2–5.5 for the behaviour of the geodesic flow and higher dimensional analogues.

**Exercise 5.2 (Anosov shadowing for $\mathrm{SL}_2(\mathbb{R})$).** Let $X$ be the quotient of $\mathrm{SL}_2(\mathbb{R})$ by a discrete subgroup $\Gamma < \mathrm{SL}_2(\mathbb{R})$.
(a) Let $x \in X$, $T > 0$, $\varepsilon > 0$ and $y \in X$ be chosen with $\mathsf{d}(a_T{\bullet}x, y) < \varepsilon$. Then there exists a point $z \in X$ with $\mathsf{d}(x, z) \ll \mathrm{e}^{-T}\varepsilon$ (and so $\mathsf{d}(a_t{\bullet}x, a_t{\bullet}z) \ll \varepsilon$ for $t \in [0, T]$) and $\mathsf{d}(a_t{\bullet}y, a_{T+t}{\bullet}z) \ll \varepsilon$ for all $t \geqslant 0$. Also show that there exists some $\delta$ with $|\delta| \ll \varepsilon$ such that $\mathsf{d}(a_{t+\delta}{\bullet}y, a_{T+t}{\bullet}z) \ll \mathrm{e}^{-t}$ for all $t \geqslant 0$.
(b) Assume now that $X$ is compact (for example, as in Figure 5.1) and use (a) to construct non-periodic orbits as in Example 5.1.

**Exercise 5.3 (Anosov closing for $\mathrm{SL}_2(\mathbb{R})$).** Let $X$ be as in Exercise 5.2. Let $x$ in $X$ and $T \geqslant 1$ be chosen so that $\mathsf{d}(a_T{\bullet}x, x) \leqslant \varepsilon < 1$. Show that there exists a point $z \in X$ which is periodic with period $T_z$ satisfying

$$|T_z - T| \ll \varepsilon$$

and

$$\mathsf{d}(a_t{\bullet}x, a_t{\bullet}z) \ll \varepsilon$$

for all $t \in [0, T]$.

**Exercise 5.4 (Anosov shadowing for $G$).** Let $G$ be a connected Lie group, let $\Gamma < G$ be a discrete subgroup, let $X = G/\Gamma$, and let $a \in G$ be such that $\mathrm{Ad}_a$ is diagonalizable with positive eigenvalues.
(a) Let $x \in X$, $N > 1$, $\varepsilon > 0$ and $y \in X$ be such that $\mathsf{d}(a^N \bullet x, y) < \varepsilon$. Then there exists a point $z \in X$, some $\lambda < 1$ (independent of $x, y$ and $\Gamma$) with

$$\mathsf{d}(a^n \bullet x, a^n \bullet z) \ll \lambda^{N-n} \varepsilon$$

for $n = 0, \dots, N$ and

$$\mathsf{d}(a^{N+n} \bullet z, a^n \bullet y) \ll \varepsilon$$

for all $n \geqslant 0$.
(b) Assume that $X$ has finite volume and $a$ acts mixingly on $X$ with respect to $m_X$. Construct non-periodic irregular orbits by iterating (a).

**Exercise 5.5 (Anosov closing for $X = \mathrm{SL}_d(\mathbb{R})/\Gamma$).** We let $X$ be any quotient of the group $G = \mathrm{SL}_d(\mathbb{R})$ by a discrete subgroup $\Gamma < G$, and let $A$ be the subgroup of $G$ of positive diagonal matrices. Let $a \in A$ be a nontrivial element.
(a) Suppose that $x \in X$ and $N \geqslant 1$ are such that $\mathsf{d}(a^N, I) \geqslant 1$ but $\mathsf{d}(a^N \bullet x, x) \leqslant \varepsilon < 1$. Assume that $\varepsilon$ is sufficiently small and that $N$ is sufficiently large. Show that there exists some $z \in X$ and some $c \in \mathrm{SL}_d(\mathbb{R})$ with $ac = ca$, $\mathsf{d}(a^N, c) \ll \varepsilon$, $c \bullet z = z$ and

$$\mathsf{d}(a^n \bullet x, a^n \bullet z) \ll \varepsilon$$

for $n = 0, \dots, N$.
(b) Suppose that $a$ is regular (that is, no two eigenvalues are the same) and $X$ is compact. Show that $z$ as in (a) is a periodic point for $A$.
(c) Suppose $d = 3$ and $a$ is regular and does not have 1 as an eigenvalue, and

$$X = \mathsf{X}_3 = \mathrm{SL}_3(\mathbb{R}) / \mathrm{SL}_3(\mathbb{Z}).$$

Show again that the point $z$ as in (a) is periodic for $A$.
(d) Repeat (c) for

$$X = \mathsf{X}_d = \mathrm{SL}_d(\mathbb{R}) / \mathrm{SL}_d(\mathbb{Z}),$$

assuming that $a \in A$ has the property that no product over a proper non-empty subset of the eigenvalues of $a$ equals 1.
(e) In the setting of (b), (c), and of (d), show that periodic $A$-orbits are dense in $X$.
(f) Generalize the statement in (b) to any semisimple group.[†]

### 5.1.2 The Horocycle Flow

The discussion above for the geodesic flow is in stark contrast to the behaviour of horocycle orbits defined by the unipotent subgroup

$$U = G_{a_1}^{+} = \left\{ u_s = \begin{pmatrix} 1 & \\ s & 1 \end{pmatrix} \;\middle|\; s \in \mathbb{R} \right\}$$

in the compact quotient $X$: The orbit of every point under this group action visits the right-hand side in Figure 5.1 at some point (indeed much more is true).

[†] In that sense Poincaré recurrence can be used to construct anisotropic tori (see Section 9.3).

In fact Hedlund [68] showed in 1936 that the horocycle flow on any compact quotient of $\mathrm{SL}_2(\mathbb{R})$ is *minimal* (that is, has no nontrivial closed invariant subsets) and that Haar measure is ergodic. This was strengthened by Furstenberg [56] in 1972 and by Dani [16] in 1978, who showed the following theorems.

**Theorem 5.6 (Unique ergodicity of horocycle flow).** *If $\Gamma$ is a uniform lattice in $\mathrm{SL}_2(\mathbb{R})$, then the horocycle flow (that is, the action of the subgroup $U$) is uniquely ergodic on the quotient $X$ of $\mathrm{SL}_2(\mathbb{R})$ by $\Gamma$.*

**Theorem 5.7 (Almost unique ergodicity of horocycle flow).** *If $X = \mathsf{X}_2$ is the quotient of $\mathrm{SL}_2(\mathbb{R})$ defined by $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ (or another non-uniform lattice) then a probability measure $m$ on $X$ that is invariant and ergodic for the action of $U$ is either*

- *the Haar measure $m_X$ on $X$ (inherited from the Haar measure $m_{\mathrm{SL}_2(\mathbb{R})}$) or*
- *a one-dimensional Lebesgue measure supported on a periodic orbit of the action for $U$.*

*Moreover, both types of invariant measure indeed exist.*

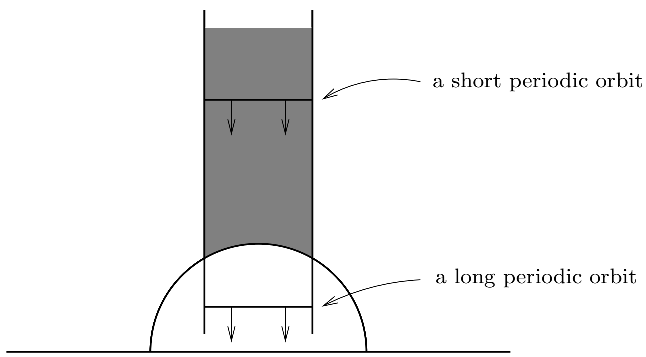


**Fig. 5.2:** In the standard fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$, the observed speed of a periodic horocycle orbit increases with the height, so the two different periodic orbits shown are of different lengths. The longer periodic orbit could also be drawn in the fundamental domain, but it would look very complicated.

Moreover, the tool discussed in the next section also gives the following theorem[†] of Sarnak [137] as well as Theorem 1.16 concerning expanding circles.

**Theorem 5.8 (Equidistribution of long periodic horocycles).** *Let $X$ be a quotient of $\mathrm{SL}_2(\mathbb{R})$ by a non-uniform lattice and let $A$ be the diagonal subgroup as in (5.1). Let $x \in X$ be a periodic orbit for the horocycle flow $U = G_{a_1}^+$ and let $\mu$ be the normalized Lebesgue measure on the one-dimensional orbit $U \bullet x$. Then the periodic orbit measures $(a_t)_* \mu$*

[†] Sarnak also gives an error rate in this equidistribution result—obtaining this (or even any) error estimate requires more sophisticated methods than those we will discuss here.

- *diverge for $t \to -\infty$ to infinity (in which case the periodic orbit $a_t U \bullet x$ becomes shorter and shorter) and*
- *equidistribute for $t \to \infty$ with respect to the Haar measure $m_X$ (in which case the periodic orbit $a_t U \bullet x$ become longer and longer).*

We will prove Theorem 5.6 in Section 5.2.1 and Theorems 5.7 and 5.8 in Section 5.3.1.

## 5.2 The Banana Mixing Trick and Unique Ergodicity

We suppose in the following that $G$ is a closed linear group and that the element $a \in G \leqslant \mathrm{SL}_d(\mathbb{R})$ only has real and positive eigenvalues. Let

$$G_a^+ = \left\{ g \in G \;\middle|\; a^n g a^{-n} \longrightarrow I \text{ as } n \longrightarrow -\infty \right\}$$

be the unstable horospherical subgroup of $a$. The general method discussed below gives a way to classify the $G_a^+$-invariant ergodic probability measures on $X$. The method goes back to the PhD thesis of Margulis, who refers to this as the banana argument due to the shape of the sets involved.

**Theorem 5.9 (Banana mixing argument for $G_a^+$).** *Let $X = G \bullet x_0 \subseteq \mathsf{X}_d$ be a finite volume orbit for a closed connected subgroup $G \leqslant \mathrm{SL}_d(\mathbb{R})$. Let $a \in G$ only have real and positive eigenvalues, and suppose that $a$ acts as a mixing transformation on $X$ with respect to $m_X$. Let $G_a^+$ be the unstable horospherical subgroup for $a$, and let $B_0$ be a neighbourhood of $I \in G_a^+$ with compact closure and a boundary of zero Haar measure. Let $f \in C_c(X)$ and $\varepsilon > 0$. Finally suppose that $K \subseteq X$ is a compact set such that $B_0 \ni u \mapsto u \bullet x$ is injective for any $x \in K$. Then there exists an integer $N$ such that*

$$\left| \frac{1}{m_{G_a^+}(a^n B_0 a^{-n})} \int_{a^n B_0 a^{-n}} f(u \bullet x) \, \mathrm{d}m_{G_a^+}(u) - \frac{1}{m_X(X)} \int_X f \, \mathrm{d}m_X \right| < \varepsilon$$

*for all $n \geqslant N$ whenever $a^{-n} \bullet x \in K$.*

We will prove the theorem in Section 5.2.2.

### 5.2.1 Unique Ergodicity on Compact Quotients

The following consequence of Theorem 5.9 is a generalization of Theorem 5.6.

**Theorem 5.10 (Unique ergodicity of horospherical actions[(28)]).** *Let $G$ be a linear Lie group, $\Gamma < G$ be a uniform lattice, and let $a \in G$ have only*

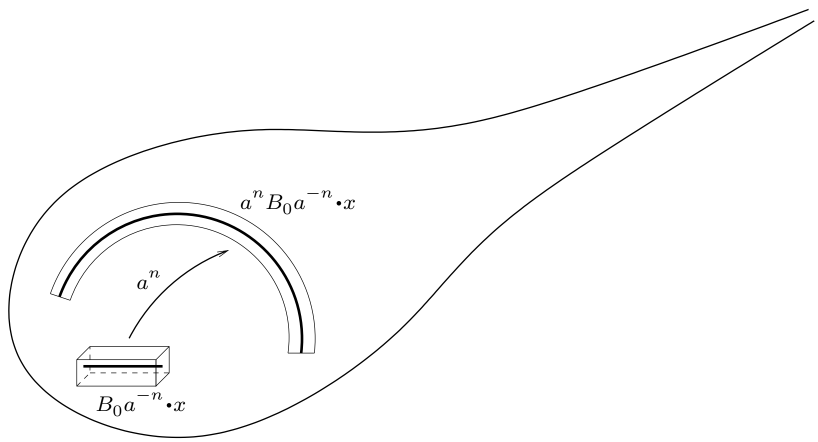


**Fig. 5.3:** A 'box' containing $B_0 a^{-n}\text{•}x$ is mapped by $a^n$ to a 'banana' that contains the much bigger set $a^n B_0 a^{-n}\text{•}x$ in the direction of $G_a^+$, is about as thick as the original box in the direction of $C_G(a)$, but is much thinner in the direction of $G_a^-$ (not drawn).

*real and positive eigenvalues. Suppose a acts mixingly on $X = G/\Gamma$. Then the action of $G_a^+$ on $X$ is* uniquely ergodic*: $m_X$ is the only $G_a^+$-invariant probability measure on $X$ and every point $x \in X$ is generic for $G_a^+$ and $m_X$.*

Proof of Theorems 5.6 and 5.10. We note that compactness of $X$ implies that $G_a^+ \ni u \mapsto u\text{•}x \in X$ is injective for any $x \in X$. Indeed, if $u\text{•}x = x$ for some $u \in G_a^+ \smallsetminus \{I\}$ and $x \in X$ then the injectivity radius at

$$a^{-n}\text{•}x = (a^{-n} u a^n) a^{-n}\text{•}x$$

would go to 0 for $n \to \infty$ and as a result contradict Lemma 1.17. Let $B_0 \subseteq G_a^+$ be as in Theorem 5.9,[†] set $K = X$ and $B_n = a^n B_0 a^{-n}$ for $n \in \mathbb{N}$. By Theorem 5.9 we have

$$\frac{1}{m_{G_a^+}(B_n)} \int_{B_n} f(u\text{•}x)\,\mathrm{d}m_{G_a^+}(u) \longrightarrow \int_X f\,\mathrm{d}m_X$$

as $n \to \infty$ for any $f \in C(X)$ and any $x \in X$ (as the constraint $a^{-n}\text{•}x \in K$ is meaningless for $K = X$).

Now let $\mu$ be a $G_a^+$-invariant probability measure. Then

$$\int_X f\,\mathrm{d}\mu = \int_X \frac{1}{m_{G_a^+}(B_n)} \int_{B_n} f(u\text{•}x)\,\mathrm{d}m_{G_a^+}(u)\,\mathrm{d}\mu(x) \longrightarrow \int_X f\,\mathrm{d}m_X$$

as $n \to \infty$ by Fubini's theorem and dominated convergence. As this holds for any $f \in C(X)$ we deduce that $\mu = m_X$, as claimed. □

[†] For example, $B_0 = B_r^{G_a^+}$ for $r > 0$ with $m_{G_a^+}\big(\{u \in G_a^+ \mid \mathsf{d}_G(u, I) = r\}\big) = 0$

Notice that once unique ergodicity is proved then the pointwise everywhere convergence of the ergodic averages also follows for other Følner sets (see Exercises 5.11–5.12).

**Exercise 5.11.** Let $B_n = a^n B_0 a^{-n}$ be as in the proof of Theorem 5.10 with $m_{G_a^+}(\partial B_0) = 0$ for $n \geqslant 1$. Show that $(B_n)$ is a Følner sequence in $G_a^+$, that is a sequence satisfying

$$\frac{m_{G_a^+}(B_n \triangle (K B_n))}{m_{G_a^+}(B_n)} \longrightarrow 0 \tag{5.2}$$

as $n \to \infty$ for every compact subset $K \subseteq G_a^+$.

**Exercise 5.12.** Let $a$ and $X$ be as in Theorem 5.10. Let $(F_n)$ be any Følner sequence in $G_a^+$ satisfying (5.2) and show that

$$\frac{1}{m_{G_a^+}(F_n)} \int_{F_n} f(u \raisebox{0.1ex}{\scriptsize$\bullet$} x)\, \mathrm{d}m_{G_a^+}(u) \longrightarrow \int_X f \, \mathrm{d}m_X$$

as $n \to \infty$, for any $f \in C(X)$ and any $x \in X$.

### 5.2.2 Proving the Banana Trick

For the proof of Theorem 5.9 we will use the 'stable parabolic subgroup'

$$P_a^- = \bigl\{ g \in G \mid a^n g a^{-n} \text{ stays bounded as } n \to \infty \bigr\}$$

together with the unstable horospherical subgroup $G_a^+$.

**Lemma 5.13 (A coordinate system).** *Let $G \leqslant \mathrm{SL}_d(\mathbb{R})$ be a closed linear group and let $a \in G$ only have real and positive eigenvalues. Then $P_a^-$ and $G_a^+$ are closed subgroups that together define a coordinate system in the following sense. The set $P_a^- G_a^+$ is open in $G$, the map $P_a^- \times G_a^+ \ni (h, u) \mapsto hu \in P_a^- G_a^+$ is a homeomorphism, and the Haar measure $m_G$ restricted to $P_a^- G_a^+$ is proportional to the push-forward of the product of the Haar measures of $P_a^-$ and $G_a^+$.*

Proof. By conjugating $a$, $G$, and its subgroups $P_a^-$ and $G_a^+$ we may assume that $a = \mathrm{diag}(a_1, \ldots, a_d)$ is diagonal. With this $P_a^-$ can be defined as a subgroup of the set of all $g = (g_{i,j}) \in G$ with $g_{i,j} = 0$ for all indices $i, j$ with $a_i > a_j$. This shows that $P_a^-$ is closed subgroup. Similarly $G_a^+$ is also a closed subgroup. Moreover, from the definition it follows that the Lie algebra $\mathfrak{p}_a^-$ of $P_a^-$ (respectively $\mathfrak{g}_a^+$ of $G_a^+$) is the direct sum of all eigenspaces in $\mathfrak{g}$ of $\mathrm{Ad}_a$ with eigenvalue less than or equal to 1 (respectively bigger than 1). This shows that the derivative of the map $\phi$: $P_a^- \times G_a^+ \ni (h, u) \mapsto hu \in P_a^- G_a^+$ at $(I, I)$ is the linear isomorphism $\mathfrak{p}_a^- \times \mathfrak{g}_a^+ \ni (x, y) \mapsto x + y \in \mathfrak{g}$. By the inverse mapping theorem $\phi$ is locally a diffeomorphism.

To see that $\phi$ is injective let $(h_1, u_1), (h_2, u_2) \in P_a^- \times G_a^+$ satisfy $h_1 u_1 = h_2 u_2$. Then $g = h_2^{-1} h_1 = u_2 u_1^{-1} \in P_a^- \cap G_a^+$ has the property that $a^n g a^{-n}$ remains bounded for $n \geqslant 0$ and converges to $I$ as $n \to -\infty$. The latter implies that the matrix $g - I$ is zero or a sum of eigenvectors for conjugation by $a$ with eigenvalues bigger than 1. Together with the behaviour for $n \geqslant 0$ this implies that $g = I$ and hence $(h_1, u_1) = (h_2, u_2)$.

Now let $O \subseteq P_a^- \times G_a^+$ be open and $(h, u) \in O$. Then $(h, I)^{-1} O (I, u)^{-1}$ is also open in $P_a^- \times G_a^+$. By the behaviour of $\phi$ near $(I, I)$ obtained above we deduce that $\phi\bigl((h, I)^{-1} O (I, u)^{-1}\bigr) = h^{-1}\phi(O)u^{-1}$ is a neighbourhood of $I$, which shows that $\phi(O)$ is a neighbourhood of $hu = \phi\bigl((h, u)\bigr)$. It follows that $\phi(O) \subseteq G$ is open for any open subset $O \subseteq P_a^- \times G_a^+$, that $P_a^- G_a^+ \subseteq G$ is open, and that $\phi\colon P_a^- \times G_a^+ \to P_a^- G_a^+$ is a homeomorphism.

With this we may apply Lemma 1.58 and obtain that $m_G$ restricted to $P_a^- G_a^+$ is proportional to the push-forward of the product of the (left) Haar measure on $P_a^-$ and the (right) Haar measure on $G_a^+$. As $G_a^+$ is unipotent it is also unimodular and the lemma follows. □

The following upgrade (a fairly standard compactness argument) to the injectivity assumption for $B_0$ and $K$ in Theorem 5.9 will be useful in the proof.

**Lemma 5.14 (Upgrade to injectivity).** *Let $K \subseteq X$ and $B_0 \subseteq G_a^+$ be compact sets for which $B_0 \ni u \mapsto u \cdot x$ is injective for all $x \in K$. Then there exists some $\delta = \delta(K, B_0) > 0$ such that*

$$B_\delta^{P_a^-} \times B_0 \ni (h, u) \longmapsto hu \cdot x$$

*is injective for all $x \in K$.*

Proof. If the conclusion of the lemma does not hold then there exist sequences $h_n \to I$ and $h_n' \to I$ as $n \to \infty$, $(u_n), (u_n')$ in $B_0$, and $(x_n)$ in $K$ with $(h_n, u_n) \neq (h_n', u_n')$ but $h_n u_n \cdot x_n = h_n' u_n' \cdot x_n$ for all $n \in \mathbb{N}$. As $K$ and $B_0$ are assumed to be compact we may assume without loss of generality that $x_n \to x \in K$, $u_n \to u \in B_0$, and $u_n' \to u' \in B_0$ as $n \to \infty$. Together we obtain $u \cdot x = u' \cdot x$ which gives $u = u'$ by our assumption. Moreover, using

$$(\underbrace{h_n u_n u^{-1}}_{\to I}) u \cdot x_n = h_n u_n \cdot x_n = h_n' u_n' \cdot x_n = \underbrace{h_n' u_n' u^{-1}}_{\to I} (u \cdot x_n)$$

as $n \to \infty$ and the injectivity radius at $u \cdot x_n \in B_0 \cdot K$ we obtain

$$h_n u_n u = h_n' u_n' u^{-1}$$

for all sufficiently large $n$. Together with the properties of the local coordinate system $P_a^- G_a^+$ in Lemma 5.13 we deduce that $(h_n, u_n) = (h_n', u_n')$ for all sufficiently large $n$. This contradicts our choice of the sequences and proves the lemma. □

Proof of Theorem 5.9. Let us assume compatibility of the Haar measures in the sense that $m_X(\pi(B)) = m_G(B)$ for any injective Borel subset $B \subseteq G$ and that $m_G$ restricted to $P_a^- G_a^+$ is equal to the product of the Haar measures $m_{P_a^-}$ and $m_{G_a^+}$.

We let $B_0 \subseteq G_a^+$ be a neighbourhood of the identity as in the theorem and define $B_n = a^{-n} B_0 a^n$ for $n \geqslant 1$. We suppose for now in addition that $B_0$ is compact and let $\delta(K, B_0)$ be as in Lemma 5.14.

Using continuity. Now fix a function $f \in C_c(X) \smallsetminus \{0\}$. By compactness of the support, $f$ is uniformly continuous. So for $\varepsilon > 0$ there is a $\delta \in \bigl(0, \delta(K, B_0)\bigr)$ for which

$$\mathsf{d}_G(g, I) < \delta \implies |f(g \centerdot y) - f(y)| < \varepsilon \tag{5.3}$$

for all $g \in G$ and $y \in X$, where $\mathsf{d}_G$ is a right-invariant metric on $G$ (giving rise to the metric $\mathsf{d}$ on $X$). We choose a compact neighbourhood $V \subseteq B_\delta^{P_a^-}$ of the identity whose boundary has measure zero with

$$\mathsf{d}_G(a^n h a^{-n}, I) < \delta$$

for $h \in V$ and $n \geqslant 0$.

The banana trick. We now come to the heart of the argument involving the 'box' $V B_0 a^{-n} \centerdot x$ and the 'banana' $a^n V B_0 a^{-n} \centerdot x$ illustrated in Figure 5.3. Indeed

$$\frac{1}{m_{G_a^+}(B_n)} \int_{B_n} f(u \centerdot x) \,\mathrm{d}m_{G_a^+}(u)$$

is within $\varepsilon$ of

$$\frac{1}{m_{P_a^-}(a^n V a^{-n}) m_{G_a^+}(B_n)} \int_{a^n V a^{-n}} \int_{B_n} f(gu \centerdot x) \,\mathrm{d}m_{P_a^-}(g) \,\mathrm{d}m_{G_a^+}(u)$$

because of (5.3) applied for $y = u \centerdot x$ and $g = a^n h a^{-n}$ with $h \in V$. Using the definition $B_n = a^n B_0 a^{-n} \subseteq G_a^+$ and Lemma 5.13, the latter may in turn be written as

$$\frac{1}{m_G(V B_0)} \int_{V B_0} f(a^n g a^{-n} \centerdot x) \,\mathrm{d}m_G(g), \tag{5.4}$$

since $m_G$ is bi-invariant and on $P_a^- G_a^+$ the product of $m_{P_a^-}$ and $m_{G_a^-}$. Moreover, using the injectivity in Lemma 5.14 at $a^{-n} \centerdot x \in K$ and the notation $y = g a^{-n} \centerdot x$ with $g \in V B_0$ we see that (5.4) can also be written as

$$\frac{1}{m_G(V B_0)} \int_X f(a^n \centerdot y) \mathbb{1}_{V B_0 a^{-n} \centerdot x}(y) \,\mathrm{d}m_X(y). \tag{5.5}$$

Using mixing. The expression in (5.5) is an inner product of $f$ composed with $a^n$ and a normalized characteristic function. Hence we would like to apply mixing of $a$ to conclude that (5.5) is $\varepsilon$-close to $\int f\,\mathrm{d}m_X$ if $n$ is large enough. However the characteristic function also depends on $n$, which in general would be an issue. Fortunately in our case $z = a^{-n}\cdot x \in K$, the map

$$K \ni z \longmapsto F_z = \mathbb{1}_{VB_0\cdot z} \in L^2_{m_X}(X)$$

is continuous, and the mixing property for $f$ and these characteristic functions holds uniformly on the compact image (see below). So we indeed obtain for $n$ large enough with $a^{-n}\cdot K$ that (5.5) is within $\mathrm{O}(\varepsilon)$ of $\int f\,\mathrm{d}m_X$.

Compactness in $L^2(X)$. Let $\eta > 0$ and recall that $m_G(\partial(VB_0)) = 0$. Using the fact that $(VB_0)^o$ is $\sigma$-compact there exists a compact subset $C \subseteq (VB_0)^o$ with $m_G(VB_0 \smallsetminus C) < \eta$. Now let $z' = g\cdot z, z \in K$ for $g$ sufficiently close to $I$ so that $Cg \subseteq VB_0$. With this we obtain

$$\begin{aligned}\|F_{z'} - F_z\|^2 &= \|F_{z'}\|^2 - 2\Re\langle F_z, F_{z'}\rangle + \|F_z\|^2\\ &= m_X(VB_0\cdot z') - 2m_X\bigl(VB_0 g\cdot z \cap VB_0\cdot z\bigr) + m_X(VB_0\cdot z)\\ &\leqslant \bigl(2m_G(VB_0) - 2m_G(C)\bigr) < 2\eta.\end{aligned}$$

As $\eta > 0$ was arbitrary, this shows the continuity of $K \ni z \mapsto F_z \in L^2_{m_X}(X)$ claimed earlier. In particular, $\mathscr{F} = \{F_z \mid z \in K\} \subseteq L^2_{m_X}(X)$ is compact.

Uniform mixing. To prove the uniform mixing we use compactness of $\mathscr{F}$ and find a finite collection $z_1, \ldots, z_J \in K$ so that for every $z \in K$ there exists some $j \in \{1, \ldots, J\}$ with

$$\|F_z - F_{z_j}\| < \frac{\varepsilon m_G(VB_0)}{\|f\|_2}. \tag{5.6}$$

Applying mixing to $f$ and $F_{z_j}$ for $j = 1, \ldots, J$ we may find $N$ so that for $n \geqslant N$ we have

$$\left|\frac{1}{m_G(VB_0)}\langle f\circ a^n, F_{z_j}\rangle - \int f\,\mathrm{d}m_X\right| < \varepsilon. \tag{5.7}$$

However, this now implies for $z \in K$ and $j \in \{1, \ldots, J\}$ satisfying (5.6) that

$$\begin{aligned}\left|\frac{1}{m_G(VB_0)}\langle f\circ a^n, F_z\rangle - \int f\,\mathrm{d}m_X\right| &\leqslant \left|\langle f\circ a^n, \frac{1}{m_G(VB_0)}(F_z - F_{z_j})\rangle\right|\\ &\quad + \left|\langle f\circ a^n, \frac{1}{m_G(VB_0)}F_{z_j}\rangle - \int f\,\mathrm{d}m_X\right|\\ &< 2\varepsilon\end{aligned}$$

by Cauchy–Schwarz, (5.6), and (5.7). This concludes the proof for compact sets $B_0 \subseteq G_a^+$.

Finding a compact $B_0$. If $B_0 \subseteq G_a^+$ is not compact, then we claim that there exists for any $\eta \in (0,1)$ a compact set $B_0' \subseteq B_0$ with $m_{G_a^+}(B_0 \smallsetminus B_o') < \eta m_{G_a^+}(B_0)$

and $m_{G_a^+}(\partial B_0')=0$. The difference between the averages obtained using the sets $B_n=a^nB_0a^{-n}$ and $B_n'=a^nB_0'a^-$ is then easily estimated. Indeed

$$\left|\frac{1}{m_{G_a^+}(B_n)}\int_{B_n} f(u\centerdot x)\,\mathrm{d}m_{G_a^+}(u)-\frac{1}{m_{G_a^+}(B_n)}\int_{B_n'} f(u\centerdot x)\,\mathrm{d}m_{G_a^+}(u)\right|$$
$$\leqslant\frac{m_{G_a^+}(B_n\smallsetminus B_n')}{m_{G_a^+}(B_n)}\|f\|_\infty=\frac{m_{G_a^+}(B_0\smallsetminus B_0')}{m_{G_a^+}(B_0)}\|f\|_\infty<\eta\|f\|_\infty$$

and

$$\left|\frac{1}{m_{G_a^+}(B_n)}\int_{B_n'} f(u\centerdot x)\,\mathrm{d}m_{G_a^+}(u)-\frac{1}{m_{G_a^+}(B_n')}\int_{B_n'} f(u\centerdot x)\,\mathrm{d}m_{G_a^+}(u)\right|$$
$$\leqslant\left|\frac{1}{m_{G_a^+}(B_n)}-\frac{1}{m_{G_a^+}(B_n')}\right|m_{G_a^+}(B_n')\|f\|_\infty$$
$$=\frac{m_{G_a^+}(B_0\smallsetminus B_0')}{m_{G_a^+}(B_0)}\|f\|_\infty<\eta\|f\|_\infty$$

shows that the two averages differ by at most $2\eta\|f\|_\infty$. Using our discussion above for $B_0$ and setting $\eta=\frac{\varepsilon}{2\|f\|_\infty}$ gives the desired conclusion for $B_0$.

To prove the claim we recall that $m_{G_a^+}(\partial B_0)=0$ and so we may assume that $B_0$ is open. It follows that $B_0$ is $\sigma$-compact and we can find a compact subset $C\subseteq B_0$ with $m_{G_a^+}(B_0\smallsetminus C)<\eta m_{G_a^+}(B_0)$. It remains to ensure that the boundary is a null set. For this we note that for any $u_0\in C$ and all but at most countably many radii $r>0$ the boundary

$$\partial B_r(u_0)\subseteq\{u\in G_a^+\mid \mathsf{d}(u,u_0)=r\}$$

is a null set. We choose $r(u_0)>0$ small enough to ensure that in addition $\overline{B_{r(u_0)}^{G_a^+}(u_0)}\subseteq B_0$. The set $B_0'\subseteq B_0$ is then obtained as the closure of the union of a finite cover

$$B_{r(u_1)}^{G_a^+}(u_1)\cup\cdots\cup B_{r(u_k)}^{G_a^+}(u_k)\subseteq B_0$$

of $C$, completing the proof of the theorem. □

## 5.3 Almost Unique Ergodicity on Non-Compact Quotients with Finite Volume

We now explain, guided by examples, how the presence of a cusp (that is, the lack of compactness of the quotient) and the presence of horospherical invariant measures other than the Haar measure are related to each other.

### 5.3.1 Horocycle Action on Non-Compact Quotients

The following result is important for the study of the horocycle flow on quotients of $\mathrm{SL}_2(\mathbb{R})$ and holds much more generally (see also Exercise 1.39). We also allow $\mathrm{SL}_2(\mathbb{C})$ as this makes no difference to the argument.

**Proposition 5.15 (Non-uniform lattices and unipotents).** *Let $G = \mathrm{SL}_2(\mathbb{R})$ or $G = \mathrm{SL}_2(\mathbb{C})$. A lattice $\Gamma < G$ is non-uniform if and only if $\Gamma$ contains non-trivial unipotent elements.*

For the proof we will need the following lemma which will help us to understand the small 'loops' for points in $X = G/\Gamma$. Here we say that $g \in G$ is a *loop* at $x \in X$ if $g\centerdot x = x$.[(29)]

**Lemma 5.16 (Zassenhaus neighbourhoods).** *There exists a norm $\|\cdot\|$ on $\mathrm{Mat}_d(\mathbb{C})$ such that for $\mathcal{N} = \{g \in \mathrm{GL}_d(\mathbb{C}) \mid \|g - I\| < 1\}$ all nontrivial discrete subgroups $\Gamma < \mathrm{GL}_d(\mathbb{C})$ generated by $\Gamma \cap \mathcal{N}$ have nontrivial centre. Moreover, for $g, h \in \mathcal{N}$ we have*

$$\big\|[g,h] - I\big\| \leqslant \|g - I\|\|h - I\|. \tag{5.8}$$

Proof. Let $\|\cdot\|_{\mathrm{op}}$ be the operator norm on $\mathrm{Mat}_d(\mathbb{C})$ satisfying

$$\|uv\|_{\mathrm{op}} \leqslant \|u\|_{\mathrm{op}}\|v\|_{\mathrm{op}}$$

for any $u, v \in \mathrm{Mat}_d(\mathbb{C})$. Let $g, h \in \mathrm{Mat}_d(\mathbb{C})$ and define $u = g - I$ and $v = h - I$. We suppose that $\|u\|_{\mathrm{op}} < \frac{1}{2}$ and $\|v\|_{\mathrm{op}} < \frac{1}{2}$. Notice that this implies that the geometric series giving $g^{-1} = (I + u)^{-1}$ and $h^{-1} = (I + v)^{-1}$ converge and that $\|g^{-1}\|_{\mathrm{op}}, \|h^{-1}\|_{\mathrm{op}} < 2$. For the commutator $[g, h] = g^{-1}h^{-1}gh$ of $g, h$ we then obtain

$$\begin{aligned}
[g,h] &= (I+u)^{-1}(I+v)^{-1}(I+u)(I+v)\\
&= (I+u)^{-1}(I+v)^{-1}(I+u+v) + \mathrm{O}(\|u\|_{\mathrm{op}}\|v\|_{\mathrm{op}})\\
&= (I+u)^{-1}\big((I+v)^{-1}(I+v) + (I+v)^{-1}u\big) + \mathrm{O}(\|u\|_{\mathrm{op}}\|v\|_{\mathrm{op}})\\
&= (I+u)^{-1}\big(I + u + \mathrm{O}(\|v\|_{\mathrm{op}})u\big) + \mathrm{O}(\|u\|_{\mathrm{op}}\|v\|_{\mathrm{op}})\\
&= I + \mathrm{O}(\|u\|_{\mathrm{op}}\|v\|_{\mathrm{op}}).
\end{aligned}$$

To summarize, we have shown that there exists a constant $c \geqslant 1$ such that

$$\|[g,h]-I\|_{\mathrm{op}} \leqslant c\|g-I\|_{\mathrm{op}}\|h-I\|_{\mathrm{op}} \tag{5.9}$$

for all $g,h$ with $\|g-I\|_{\mathrm{op}}, \|h-I\|_{\mathrm{op}} < \frac{1}{2}$. We assume $c \geqslant 2$, define $\|g\| = c\|g\|_{\mathrm{op}}$, multiply (5.9) by $c$ and obtain (5.8) for all $g,h$ with $\|g-I\|, \|h-I\| < \frac{c}{2}$. We define $\mathcal{N}$ as in the lemma, which in particular ensures that $[g,h] \in \mathcal{N}$ for all $g,h \in \mathcal{N}$.

Now let $\Gamma$ be a nontrivial discrete subgroup generated by $\Gamma \cap \mathcal{N}$. Let

$$g \in \Gamma \cap \mathcal{N} \smallsetminus \{I\}$$

have the minimal distance to $I$ with respect to the above norm. Then (5.8) shows that $\|[g,h]-I\| < \|g-I\|$ for all $h \in \Gamma \cap \mathcal{N}$. This forces $[g,h] = I$ and hence $g$ belongs to the centre of $\langle \Gamma \cap \mathcal{N} \rangle = \Gamma$. □

The following special feature of $\mathrm{SL}_2$ is particularly useful.

**Lemma 5.17 (Centralizers).** *Let $G = \mathrm{SL}_2(\mathbb{K})$ for $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. Any two $g,h \in G$ that commute but do not belong to the centre define the same (automatically abelian) centralizers. If $g,h \in \mathrm{SL}_2(\mathbb{K})$ are close to the identity, then $g$ and $h$ commute if and only if $\log g$ and $\log h$ are linearly dependent over $\mathbb{K}$.*

Proof. It is sufficient to study the case $\mathbb{K} = \mathbb{C}$ and we will prove a version of the lemma for $\mathrm{Mat}_2(\mathbb{C})$. Assume first that $g = \mathrm{diag}(\alpha, \beta)$ for some $\alpha \neq \beta \in \mathbb{C}$. A simple calculation shows that $g$ and some $h \in \mathrm{Mat}_2(\mathbb{C})$ commute if and only if $h$ is also diagonal. By conjugation the first claim of the lemma follows in a more general form within $\mathrm{Mat}_2(\mathbb{C})$ if one of $g$ or $h$ is diagonalizable.

If $g$ is not diagonalizable we may assume that $g = \lambda I + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. For $h = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ we then have $gh = \begin{pmatrix} \lambda a + c & \lambda b + d \\ \lambda c & \lambda d \end{pmatrix}$ and $hg = \begin{pmatrix} \lambda a & \lambda b + a \\ \lambda c & \lambda d + c \end{pmatrix}$. If $h$ commutes with $g$ this gives $h = \begin{pmatrix} a & b \\ 0 & a \end{pmatrix}$. If $h$ is not a scalar multiple of $I$ we obtain that $h$ once again has the same structure as $g$. Together with the above, this gives the first claim in the lemma for $\mathrm{Mat}_2(\mathbb{C})$.

For the second claim suppose that $g,h \in \mathrm{SL}_2(\mathbb{K}) \smallsetminus \{I\}$ commute and are close to the identity so that $u = \log g$ and $v = \log h$ are well-defined. Then $g = \exp(u)$ commutes with $\exp(su)$ for $s \in \mathbb{C}$ and $h = \exp(v)$ commutes with $\exp(tv)$ for $t$ in $\mathbb{C}$. By the first part of the lemma the two one-parameter subgroups defined by $s \mapsto \exp(su)$ and $t \mapsto \exp(tv)$ commute. Moreover, this implies that $[u,v] = 0$. Now the first part of the proof implies that the traceless matrices $u,v \in \mathfrak{sl}_2(\mathbb{C})$ are multiples of each other. □

Proof of Proposition 5.15. Let $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$ and let $\Gamma < G = \mathrm{SL}_2(\mathbb{K})$ be a lattice. Let $a_t = \mathrm{diag}\big(\mathrm{e}^{-\frac{t}{2}}, \mathrm{e}^{\frac{t}{2}}\big)$ as before and let $U = G_{a_1}^+ < G$ the lower unipotent subgroup. If $\Gamma$ contains a nontrivial unipotent element $\gamma$ then there exists some $g \in G$ so that $u = g\gamma g^{-1} \in U$. However, this implies that the point $g\Gamma$ satisfies $u\cdot g\Gamma = g\Gamma$. Applying $a_{-t}$ gives $a_{-t}ua_t\cdot a_{-t}g\Gamma = a_{-t}g\Gamma$. As $u \neq I$ and $a_{-t}ua_t \to I$ as $t \to \infty$ it follows that $a_{-t}g\Gamma \to \infty$ as $t \to \infty$ by the divergence criterion in Proposition 1.35. Hence $X$ is non-compact.

The converse is the more difficult direction. So suppose that $\Gamma$ contains no nontrivial unipotent elements. Let $\mathcal{N}_0 \subseteq \mathrm{SL}_2(\mathbb{R})$ be an open neighbourhood of $\Gamma$ so that the conclusions of Lemmas 5.16 and 5.17 hold on $\mathcal{N}_0$. Let $\mathcal{N}$ be an open neighbourhood with $\overline{\mathcal{N}} \subseteq \mathcal{N}_0$. By the divergence criterion in Proposition 1.35 the set

$$K = \{x \in X \mid \mathcal{N} \ni g \longmapsto g\cdot x \text{ is injective}\}$$

is compact. For $x_0 \in K$ there might be loops $g \in \mathcal{N}_0 \smallsetminus \mathcal{N}$ with $g\cdot x_0 = x_0$. Note that if $x_0 = g_0\Gamma$ then there exists some $\gamma \in \Gamma$ with $gg_0 = g_0\gamma$, which shows that the characteristic polynomial of the loop $g$ is also the characteristic polynomial of $\gamma$ (corresponding to the loop $g$ at $x_0$). Similarly replacing $x_0$ by $hx_0$ for some $h \in G$ creates a loop $hgh^{-1}$ at $hx_0$ with the same characteristic polynomial. A simple compactness argument now shows that varying $x_0 \in K$ and $g \in \mathcal{N}_0$ gives only a finite set $\mathscr{F} \subseteq \mathbb{K}[T]$ of characteristic polynomials of loops. As every such polynomial is also a characteristic polynomial of an element of $\Gamma$ and we assume that $\Gamma$ contains no nontrivial unipotent elements we deduce that $\mathscr{F}$ does not contain $(T-1)^2$. Hence there exists a neighbourhood $O$ of $I \in \mathrm{SL}_2(\mathbb{K})$ so that $g \in O$ implies that the characteristic polynomial of $g$ does not belong to $\mathscr{F}$.

We will show that $g \in O \smallsetminus \{I\}$ cannot appear as a loop of any $x_0 \in X$. By the divergence criterion in Proposition 1.35 this then shows that $X$ must be compact. So suppose for the purposes of a contradiction that $g \in O \smallsetminus \{I\}$ is a loop at some $x_0 = g_0\Gamma$. Then $g_0^{-1}gg_0 \in \Gamma$ or equivalently $g \in \Lambda = g_0\Gamma g_0^{-1}$. As $\Lambda$ is discrete we may apply Lemmas 5.16 and 5.17 and obtain that $\Lambda \cap \mathcal{N} \subseteq \exp(\mathbb{K}v)$ for some unit vector $v \in \mathrm{SL}_2(\mathbb{K})$. We may assume that $g = \exp(su) \in \Lambda \cap \mathcal{N} \smallsetminus \{I\}$ is a smallest element with $|s| < 1$.

If $v = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}$ has $c \neq 0$, we apply $a_t$ with $t > 0$ to $x_0$. Simultaneously we conjugate the elements of $\Lambda \cap \mathcal{N}$ and the smallest loop $a_tga_{-t} = \exp(s\,\mathrm{Ad}_{a_t} u)$. Using the fact that $c$ is expanded and continuity we find $t \in \mathbb{R}$ so that the loop $a_tga_{-t}$ at $a_t\cdot x_0$ belongs to $\mathcal{N}_0 \smallsetminus \overline{\mathcal{N}}$. As this was the smallest loop, Lemmas 5.16 and 5.17 imply that $a_t\cdot x_0 \in K$. By our choice of $O$ this gives a contradiction. If $c = 0$ but $b \neq 0$ we similarly apply $a_t$ with $t < 0$. If $v = \begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix}$ we first apply $\begin{pmatrix} 1 & \\ 1 & 1 \end{pmatrix}$ to $x_0$, which replaces $v$ by $\begin{pmatrix} a & \\ 2a & -a \end{pmatrix}$. After this we apply $a_t$ for $t > 0$ and argue as above. □

Proof of Theorem 5.7. Suppose that $\Gamma < \mathrm{SL}_2(\mathbb{R})$ is a non-uniform lattice. By Proposition 5.15 $\Gamma$ contains a nontrivial unipotent element $\gamma$ which is conjugated to an element $g\gamma g^{-1} \in U$ for some $g \in \mathrm{SL}_2(\mathbb{R})$. However, this implies that $Ug\Gamma$ is a periodic orbit supporting a one-dimensional $U$-invariant Lebesgue measure.

We still have to show that $m_X$ and the one-dimensional periodic orbit measure are the only $U$-invariant and ergodic probability measures on $X$. So assume that $\mu$ is a $U$-invariant and ergodic probability measure on $X$ and let $x \in X$ be a generic point for $\mu$. We consider the orbit $a_{-t}\bullet x$ for $t \geqslant 0$.

If $x$ is periodic under $U$, then $\mu$ is the one-dimensional Lebesgue measure on $U\bullet x$ and $a_{-t}\bullet x \to \infty$ for $t \to \infty$. So suppose now that $x$ is not periodic under $U$, let $t_0 > 0$ be arbitrary, $x' = a_{-t_0}\bullet x$, and $\mathcal{N} \subseteq \mathcal{N}_0$ be the neighbourhoods of $I$ as in the proof of Proposition 5.15. Suppose $x'$ has a loop in $\mathcal{N}$. As $x$ is not periodic under $U$ and $a_{t_0}$ normalizes $U$ we see that $x'$ is also not periodic and so the loop cannot belong to $U$. We now argue along the lines of the proof of Proposition 5.15: Let $g = \exp(v)$ be a smallest loop at $x$ with $v = \begin{pmatrix} a & b \\ c & -a \end{pmatrix} \neq 0$. If $b \neq 0$ then the smallest loop at $a_{-t}\bullet x$ eventually grows for $t > t_0$. If $b = 0$ and $a \neq 0$, then the smallest loop at $a_{-t}\bullet x$ will not grow but also will not go to zero. Finally, $b = 0$, $a = 0$, and $c \neq 0$ would mean that $x'$ and hence also $x$ are periodic for $U$. It follows that for any $t_0$ there exists $t \geqslant t_0$ so that $a_{-t}$ belongs to a fixed compact subset $K \subseteq X$. This allows us to find a subsequence $t_n \to \infty$ with $a_{-t_n}\bullet x \in K$, apply Theorem 5.9, and obtain a subsequence of times for which the time average along $U\bullet x$ converges to $\int f \,\mathrm{d}m_X$ for all $f \in C_c(X)$. As the time average converges to $\int f \,\mathrm{d}\mu$ by the choice of $x$ we obtain $\mu = m_X$. □

Proof of Theorem 5.8. Let $x$ be periodic for $U$. Then the injectivity radius at any point in $a_t\bullet Ux$ goes to 0 for $t \to -\infty$, which shows that $a_t\bullet(Ux)$ diverges for $t \to -\infty$. For $t \to \infty$ we let $K = \{x\}$, let $B_0 \subseteq U$ be an interval so that $B_0 \ni s \mapsto u_s\bullet x \in Ux$ is bijective, and apply Theorem 5.9 to see that the normalized Lebesgue measure on the expanding orbits $a_t\bullet Ux$ equidistributes for $t \to \infty$ to the Haar measure $m_X$ on $X$. □

### 5.3.2 Equidistribution for Non-Compact Quotients

Dani and Smillie showed in [25] that even for non-compact quotients of $\mathrm{SL}_2(\mathbb{R})$ a rather strong equidistribution theorem holds: A horocycle orbit is either periodic or it equidistributes with respect to the uniform measure $m_X$.

For higher dimensional non-compact quotients $X = G/\Gamma$ and their horospherical actions other possibilities can occur. For the following characterization of whether or not a horospherical orbit equidistributes we specialize to the case where the horospherical subgroup is abelian.

**Theorem 5.18.** *Let $G\bullet x_0 \subseteq \mathsf{X}_d = \mathrm{SL}_d(\mathbb{R})/\mathrm{SL}_d(\mathbb{Z})$ be a finite volume orbit for some closed connected subgroup $G \leqslant \mathrm{SL}_d(\mathbb{R})$ and some point $x_0 \in \mathsf{X}_d$. Let $a \in G$ have only real and positive eigenvalues so that the action of $a$ is mixing with respect to $m_{G\bullet x_0}$. Let $U = G_a^+$ be the unstable horospherical subgroup of $a$ and* *suppose that it is abelian. Let $(F_n)$ be a Følner sequence in $U$ containing $I$ consisting of blocks whose sides are parallel to some fixed coordinate system spanned by eigenvectors for the conjugation map by $a$. Then for every $x \in G\boldsymbol{\cdot}x_0$ the following are equivalent:*

(1) *The $U$-orbit through $x$ is equidistributed, meaning that*

$$\frac{1}{m_U(F_n)} \int_{F_n} f(u)\,\mathrm{d}m_U(u) \to \int_{\mathsf{X}_d} f\,\mathrm{d}m_{\mathsf{X}_d}$$

*as $n \to \infty$ for any $f \in C_c(\mathsf{X}_d)$.*

(2) *The orbit $U\boldsymbol{\cdot}x$ is not contained in a closed orbit $L\boldsymbol{\cdot}x$ for any proper connected subgroup $L < G$.*

*If, in addition, $G = \mathrm{SL}_d(\mathbb{R})$ and $x = g\,\mathrm{SL}_d(\mathbb{Z})$ for some $g \in \mathrm{SL}_d(\mathbb{R})$ then we also have the equivalence to the next property.*

(3) *There is no rational subspace $V \subseteq \mathbb{R}^d$ for which $gV$ is fixed by $U$ and expanded by $a$.*

Proof. We let $x \in G\boldsymbol{\cdot}x_0$ be as in the theorem. If the $U$-orbit of $x$ is contained in a closed orbit of a proper connected subgroup $L < G$ as in (2), then clearly we cannot have equidistribution of the $U$-orbit as in (1). This shows that (1) implies (2).

Assume now that the $U$-orbit $U\boldsymbol{\cdot}x$ is not contained in a closed orbit $L\boldsymbol{\cdot}x$ for any proper closed subgroup $L < G$ as in (2). Fix some $f \in C_c(\mathsf{X}_d)$ and $\varepsilon > 0$. We let $x_0 = g_0\,\mathrm{SL}_d(\mathbb{Z})$, let $\Lambda_0 = g_0\mathbb{Z}^d$ be the lattice corresponding to $x_0$, and define

$$\eta = \min\left\{\mathrm{covol}(\Lambda_0 \cap V, V)^{1/\dim V} \;\middle|\; V \text{ is } \Lambda_0\text{-rational}\right\}. \tag{5.10}$$

By quantitative non-divergence for the action of $U = G_a^+$ there exists some compact set $K \subseteq \mathsf{X}_d$ with the property as in Proposition 4.13 with $\delta = \varepsilon$. We let $B_0$ be the symmetric cube (that is, centred at the origin) in

$$U = G_a^+ \cong \mathbb{R}^\ell$$

satisfying the injectivity requirement of Theorem 5.9 on $K$. Applying that theorem to $f$, $B_0K$, and $\varepsilon$, we find some $k \geqslant 1$ such that

$$\left|\frac{1}{m_{G_a^+}(a^kB_0a^{-k})} \int\limits_{a^kB_0a^{-k}} f(u\boldsymbol{\cdot}y)\,\mathrm{d}m_{G_a^+}(u) - \int\limits_X f\,\mathrm{d}m_X\right| < \varepsilon \tag{5.11}$$

whenever $B_0a^{-k}\boldsymbol{\cdot}y$ intersects $K$ non-trivially.

Now let $x' = a^{-k}\boldsymbol{\cdot}x$ and notice that it may not belong to $K$. Since $(F_n)$ is chosen to be a Følner sequence consisting of blocks, the same is true for $a^{-k}F_na^k$. If $U = G_a^+$ fixes a $\Lambda_{x'}$-rational subspace $V$ of covolume $< \eta^{\dim V}$, then we can define the subgroup

$$L = \mathrm{Stab}_G^1(V) = \{g \in G \mid gV = V \text{ and } g|_V \text{ has determinant } 1\} \lneq G$$

that does not contain $a$ (by our definition of $\eta$ in (5.10)). Exercise 3.12 shows that $Lx'$ is closed, which shows that

$$a^k L \cdot x' = a^k L a^{-k} \cdot x$$

is a closed orbit of a proper subgroup which contradicts our assumption in (2).

It follows that $U$ does not fix any $\Lambda_x$-rational subspaces $V$ with covolume less than $\eta^{\dim V}$. Applying Proposition 4.13 we see now that for large enough $n$ we have

$$\frac{1}{m_{G_a^+}(a^{-k}F_n a^k)} m_{G_a^+}\Big(\big\{u \in a^{-k}F_n a^k \mid u \cdot x' \notin K\big\}\Big) < \varepsilon. \tag{5.12}$$

We now split $a^{-k}F_n a^k$ into translates $B_0 u_\ell$ for $\ell = 1, \ldots, L$ of the cube $B_0$. Ignoring the effects of the boundary which contribute no more than $\mathrm{o}_f(1)$ to the ergodic average as $n \to \infty$, we now have

$$\begin{aligned}\frac{1}{m_{G_a^+}(F_n)} \int_{F_n} & f(u \cdot x)\,\mathrm{d}m_{G_a^+} \\ &= \frac{1}{L}\sum_{\ell=1}^{L} \frac{1}{m_{G_a^+}(a^k B_0 a^{-k})} \int_{a^k B_0 a^{-k}} f\left(u a^k u_\ell a^{-k} \cdot x\right)\,\mathrm{d}m_{G_a^+} + \mathrm{o}_f(1).\end{aligned}$$

For all those $\ell$ for which $B_0 u_\ell a^{-k} \cdot x$ intersects $K$ the corresponding average is $\varepsilon$-close to $\int_X f\,\mathrm{d}m_X$ by (5.11). However, the number of boxes $B_0 u_\ell \cdot x'$ that do not intersect $K$ is controlled by (5.12), and gives

$$\frac{1}{m_{G_a^+}(F_n)} \int_{F_n} f(u \cdot x)\,\mathrm{d}m_{G_a^+}(u) = \int_X f\,\mathrm{d}m_X + \mathrm{o}_f(1) + \mathrm{O}_f(\varepsilon)$$

for $n \to \infty$. As $\varepsilon > 0$ and $f \in C_c(X)$ were arbitrary, this shows (1).

Now suppose that $G = \mathrm{SL}_d(\mathbb{R})$. We note that (2) implies (3) by Exercise 3.12. It remains to show that (3) implies (1). For this let

$$a = \begin{pmatrix} \lambda^{-n} I_m & \\ & \lambda^m I_n \end{pmatrix} \in \mathrm{SL}_d(\mathbb{R})$$

for some $\lambda > 1$ so that

$$G_a^+ = \left\{ \begin{pmatrix} I_m & \\ * & I_n \end{pmatrix} \right\}$$

is indeed abelian (up to conjugation and the choice of $m$ and $n$ this is the only choice of $a$ for which $G_a^+$ is abelian). Suppose now that $V \subseteq \mathbb{R}^d$ is a proper $G_a^+$-invariant subspace. Then either $V \subseteq \{0\}^m \times \mathbb{R}^n$ or $V$ contains some $v = (v_m, v_n)$

with $v_m \in \mathbb{R}^m \smallsetminus \{0\}$ and $v_n \in \mathbb{R}^n$, which implies that $\{0\}^m \times \mathbb{R}^n \subseteq V$. In both cases $aV = V$ and the restriction of $a$ to $V$ has determinant bigger than 1. It follows that (3) is equivalent to the assumption that $U$ does not fix any proper $g\mathbb{Z}^d$-rational subspace. By setting $\eta = 1$ the above argument now proves (1). □

In the exercises we outline how one can remove the assumptions on commutativity of $U$.

**Exercise 5.19.** Let $G \cdot x_0 \subseteq \mathsf{X}_d = \mathrm{SL}_d(\mathbb{Z})\backslash \mathrm{SL}_d(\mathbb{R})$ be a finite volume orbit for some closed connected subgroup $G \leqslant \mathrm{SL}_d(\mathbb{R})$ and some point $x_0 \in \mathsf{X}_d$. Let $a \in G$ have only real and positive eigenvalues so that the action of $a$ is mixing with respect to $m_{G \cdot x_0}$. Let $U = G_a^+$ be the unstable horospherical subgroup of $a$ and let $F_n$ be as in Exercise 4.14. Let $x \in G \cdot x_0$. Suppose that $U$ does not fix any $\Lambda_x$-rational subspace which is not also fixed by $G$. Show that $F_n \cdot x$ equidistributes in $X = G \cdot x_0$ in the sense that

$$\frac{1}{m_U(F_n)} \int_{F_n} f(u \cdot x) dm_U(u) \longrightarrow \int_X f dm_X$$

as $n \to \infty$ for any $f \in C_c(X)$.

## 5.4 The Counting Method of Duke–Rudnick–Sarnak and Eskin–McMullen

We return to the topic of Sections 1.1 and 1.2.6. In fact we wish to explain the work of Eskin and McMullen [51] who use mixing to establish asymptotic counting results in a more general context. For this (and in preparation for other special cases to be considered later) we describe in this section the general set-up for the work of Duke, Rudnick and Sarnak [37] (which is also used in the work of Eskin and McMullen) on how to relate a counting problem for points in $\Gamma$-orbits on $V = G/H$ to the equidistribution problem for 'translated' $H$-orbits of the form

$$gH\Gamma \subseteq X = G/\Gamma.$$

In many cases (for example, in the context of affine symmetric spaces), the methods of this chapter can be used to give the asymptotic of the counting for the number of integer points on varieties. In fact, suppose $G$ and $H$ consist of the $\mathbb{R}$-points of algebraic groups defined over $\mathbb{Q}$, $V = G/H$ can be identified with an affine variety defined over $\mathbb{Q}$, and $V(\mathbb{Z})$ is non-empty. Then we get that $V(\mathbb{Z})$ is a disjoint union

$$V(\mathbb{Z}) = \bigsqcup_i G(\mathbb{Z}) v_i$$

of different $\Gamma = G(\mathbb{Z})$-orbits. Frequently this is a finite union, and then one gets the asymptotic for $|V(\mathbb{Z}) \cap B_t|$ by assembling the results for the individual counts $|G(\mathbb{Z}) v_i \cap B_t|$. We will discuss the details of such integer point counting problems in special cases in the remaining sections of this chapter, and we refer to the papers of Duke, Rudnick and Sarnak and of Eskin and McMullen [37, 51] for a detailed discussion of the general problem of counting lattice points in affine symmetric spaces.

### 5.4.1 Compatibility of all Haar Measures Involved

In order to state the result, we have to briefly describe the necessary compatibility of all the Haar measures involved. Let $m_G$ be a Haar measure on a unimodular group $G$, and let $\mathit{\Gamma} < G$ be a lattice, on which we choose counting measure as the Haar measure. As we know $m_G$ induces in a natural way a Haar measure $m_X$ on $X = G/\mathit{\Gamma}$, giving total mass $m_X(X) = m_G(F)$ where $F \subseteq G$ is a Borel fundamental domain for (the right action of) $\mathit{\Gamma}$.

Assume that $H < G$ is a closed unimodular subgroup with Haar measure $m_H$. Then (see Section A.2) we may define a locally finite measure $m_{G/H}$ with the following compatibility property, which is analogous to Fubini's theorem if $G$ is thought of measurably as a product of $H$ and $G/H$. If $f \in L^1_{m_G}(G)$ then the function $F$ defined by the relation

$$F(gH) = \int_H f(gh)\,\mathrm{d}m_H(h) \tag{5.13}$$

exists for almost every $g \in G$, and the measure $m_{G/H}$ satisfies

$$\int_{G/H} F(gH)\,\mathrm{d}m_{G/H} = \int_G f\,\mathrm{d}m_G. \tag{5.14}$$

### 5.4.2 First step: Equidistribution gives an Averaged Counting Result

Let $\mathit{\Gamma} < G$ be a lattice, and assume that $H < G$ is a closed subgroup with the property that $\mathit{\Gamma} \cap H < H$ is also a lattice. Let $Y = H/\mathit{\Gamma} \cap H$ identified with the closed orbit $H \centerdot \mathit{\Gamma} \subseteq X$, and let $m_Y$ be the Haar measure on $Y$ induced by the Haar measure $m_H$ on $H$. We make the following[†] *equidistribution assumption:*

$$\text{the translated } H\text{-orbits } gH\centerdot\mathit{\Gamma} \text{ equidistribute in } X = G/\mathit{\Gamma} \tag{5.15}$$

as $gH \to \infty$ in $G/H$. In other words the push-forward of $\frac{1}{m_Y(Y)} m_Y$ under $g$ should converge to $\frac{1}{m_X(X)} m_X$ in the weak* topology as $gH \to \infty$.

The assumptions above already imply a weak* version of our desired counting result in the following sense. We let $\{B_t \mid t \in \mathbb{R}\}$ be a collection of subsets of $G/H$ each with finite Haar measure, and define a modified orbit-counting

[†] Alternatively, we may just assume some equidistribution on average—in a sense to be made clear in the proof.

function $F_t\colon X \to \mathbb{R}_{\geqslant 0}$ by

$$F_t(g\Gamma) = \frac{1}{m_{G/H}(B_t)} \left| g\Gamma{\boldsymbol\cdot}H \cap B_t \right|, \tag{5.16}$$

which counts elements in $B_t$ within the $\Gamma$-orbit of $H \in G/H$ translated by $g$.

**Proposition 5.20 (Weak* Counting Result).** *If $m_{G/H}(B_t) \to \infty$ as $t \to \infty$, then* (5.15) *implies the* weak*-convergence

$$F_t \, \mathrm{d}m_X \longrightarrow \frac{m_Y(Y)}{m_X(X)} \, \mathrm{d}m_X \tag{5.17}$$

*as $t \to \infty$, where $Y = H/\Gamma \cap H$.*

### 5.4.3 Second step: Additional Geometric Assumptions imply the Counting Result

In order to be able to obtain the desired counting result from the averaged weak counting result above, we need to assume that the sets $B_t$ are well behaved in a geometric manner.

**Definition 5.21 (Geometric Assumption).** A monotonically increasing family $\{B_t \mid t \in \mathbb{R}\}$ of subsets of $G/H$ is *well-rounded* if $m_{G/H}(B_t) \to \infty$ as $t \to \infty$, for every $\delta > 0$ there exists a neighbourhood $U$ of $I \in G$ with

$$B_{t-\delta} \subseteq \bigcap_{g \in U} gB_t \subseteq B_t \subseteq \bigcup_{g \in U} gB_t \subseteq B_{t+\delta},$$

and furthermore for every $\varepsilon > 0$ there exists $\delta > 0$ with

$$\frac{m_{G/H}(B_{t+\delta})}{m_{G/H}(B_t)} < 1 + \varepsilon$$

for all $t \geqslant 0$.

**Theorem 5.22 (Asymptotic Counting).** *If the translated $H$-orbits equidistribute as assumed in* (5.15)*, and the family of sets $\{B_t\}$ is well-rounded as above, then we have the asymptotic*

$$\lim_{t \to \infty} \frac{1}{m_{G/H}(B_t)} \left| \Gamma H \cap B_t \right| = \frac{m_{H/\Gamma \cap H}(H/\Gamma \cap H)}{m_X(X)} \tag{5.18}$$

*for the orbit-point counting problem.*

We note that Selberg's Theorem 1.15 turns out to be a very special case of this setup.

### 5.4.4 Proofs

We now turn to considering the components of the outlined argument in greater detail and start by proving the equidistribution of expanding circles in Theorem 1.16 (leading to the instance of (5.15) needed for Selberg's theorem).

Proof of Theorem 1.16. The argument is similar to the banana mixing argument from Section 5.2 but easier. Let $\Gamma < \mathrm{SL}_2(\mathbb{R}) = G$ be a lattice, let $K = \mathrm{SO}_2(\mathbb{R})$, let $A = \{a_t \mid t \in \mathbb{R}\}$ be the diagonal subgroup, and let $N = G_a^-$ be the stable horocycle subgroup. By the Iwasawa decomposition in Proposition 1.55 we have $G = NAK$ and uniqueness of the corresponding decomposition. Moreover, by Lemma 1.58 the Haar measure on $G$ is the direct product of the Haar measure $m_{NA}$ on $B = NA$ and the Haar measure on $K$.

Let $x_0 \in X$, $f \in C_c(X)$, and $\varepsilon > 0$. Using a compactness argument as in Lemma 5.14 there exists a $\delta > 0$ such that the map

$$B_\delta^{NA} \times K\raisebox{0.1ex}{\scriptsize$\bullet$}x_0 \ni (h, k\raisebox{0.1ex}{\scriptsize$\bullet$}x_0) \longmapsto hkx_0$$

is injective. By shrinking $\delta$ we may assume that it satisfies the uniform continuity claim in (5.3) for $f$ and $\varepsilon$. Using this and applying mixing for $f$ and $\frac{1}{m_G(B_\delta^{NA}K\bullet x)}\mathbb{1}_{B_\delta^{NA}K\bullet x}$ leads to the desired estimate. □

Proof of the instance of (5.15) for Theorem 1.15. Let $H = K = \mathrm{SO}_2(\mathbb{R})$ and $(g_n)$ in $G = \mathrm{SL}_2(\mathbb{R})$ so that $g_nK \to \infty$ as $n \to \infty$. Applying the Cartan decomposition to $g_n$ we find $k_n \in \mathrm{SO}_2(\mathbb{R})$ and diagonal matrices $a_{t_n}$ with $t_n \geqslant 0$ so that $g_nK = k_n a_{t_n} K$ for all $n \geqslant 1$. By choosing a subsequence we may assume that $k_n \to k$ as $n \to \infty$ for some $k \in K$.

Now fix $f \in C_c(X)$ and $\varepsilon > 0$. By uniform continuity we have

$$|f(k_n\bullet x) - f(k\bullet x)| < \varepsilon$$

for $x \in X$ and all sufficiently large $n$. Applying in addition the equidistribution of expanding circles in Theorem 1.16 to $f\bullet k$ we have that

$$\left| \int_Y f(g_n\bullet y)\,\mathrm{d}m_Y(y) - \int_X f\,\mathrm{d}m_X \right| \leqslant \varepsilon + \left| \int_Y f(ka_{t_n}\bullet y)\,\mathrm{d}m_Y(y) - \int_Y f\,\mathrm{d}m_X \right| \leqslant 2\varepsilon$$

for all sufficiently large $n$. □

We return now to the general setup considered in Sections 5.4.1–5.4.3.

Proof of weak*-convergence in Proposition 5.20. We assume (5.15), or more precisely that the normalized translation

$$\frac{1}{m_Y(Y)} g_* m_Y$$

of the Haar measure $m_Y$ on

$$Y = {}^{H}\!/\Gamma \cap H \subseteq X = {}^{G}\!/\Gamma$$

translated by $gH \in G/H$ converges to the normalized Haar measure

$$\frac{1}{m_X(X)} m_X$$

in the following averaged sense. For a test function $\alpha \in C_c(X)$ we require that

$$\frac{1}{m_Y(Y) m_{G/H}(B_t)} \int\limits_{B_t}\int\limits_{Y} \alpha\,(gh\Gamma)\,\mathrm{d}m_Y(h\Gamma)\,\mathrm{d}m_{G/H}(gH)$$
$$\longrightarrow \frac{1}{m_X(X)} \int \alpha\,\mathrm{d}m_X \qquad (5.19)$$

as $t \to \infty$. As $m_{G/H}$ is locally finite this is certainly satisfied if both

$$\frac{1}{m_Y(Y)} \int\limits_{Y} \alpha\,(gh\Gamma)\,\mathrm{d}m_Y \longrightarrow \frac{1}{m_X(X)} \int \alpha\,\mathrm{d}m_X$$

as $Hg \to \infty$ in $G/H$ and

$$m_{G/H}(B_t) \longrightarrow \infty$$

as $t \to \infty$, but (5.19) is a weaker requirement because of the additional averaging.

We wish to deduce from this assumption that

$$\int_X F_t(x)\alpha(x)\,\mathrm{d}m_X \longrightarrow \frac{m_Y(Y)}{m_X(X)} \int_X \alpha\,\mathrm{d}m_X$$

as $t \to \infty$.

The proof is relatively short, and consists of an application of the folding/unfolding trick (see also Proposition 1.31) using the spaces

$$\begin{array}{ccc} & G/\Gamma \cap H & \\ \swarrow & & \searrow \\ G/\Gamma & & G/H. \end{array}$$

By the definition of $F_t$ in (5.16) we have

$$\begin{aligned} A_t^\alpha &= \int_X F_t(x)\alpha(x)\,\mathrm{d}m_X \\ &= \frac{1}{m_{G/H}(B_t)} \int_{G/\Gamma} \sum_{\gamma \in \Gamma/\Gamma\cap H} \mathbb{1}_{B_t}(g\gamma H)\alpha(g\Gamma)\,\mathrm{d}m_X(g\Gamma), \end{aligned}$$

in which the sum over $\gamma \in \Gamma/\Gamma \cap H$ denotes the sum over a list of representatives of the cosets of $\Gamma \cap H$ in $\Gamma$. Thus by using the compatibility of the Haar measures in (5.13)–(5.14) we get

$$\begin{aligned}
A_t^\alpha &= \frac{1}{m_{G/H}(B_t)} \int_{G/\Gamma\cap H} \mathbb{1}_{B_t}(gH)\alpha(g\Gamma)\,\mathrm{d}m_{G/\Gamma\cap H}(g(\Gamma\cap H)) \\
&= \frac{1}{m_{G/H}(B_t)} \int_{G/H} \mathbb{1}_{B_t}(gH) \int_{H/\Gamma\cap H} \alpha(gh\Gamma)\,\mathrm{d}m_Y(h\Gamma)\,\mathrm{d}m_{G/H}(gH) \\
&= \frac{1}{m_{G/H}(B_t)} \int_{B_t} \int_Y \alpha(gh\Gamma)\,\mathrm{d}m_Y(h\Gamma)\,\mathrm{d}m_{G/H}(gH).
\end{aligned}$$

For the first unfolding step note that if $F \subseteq G$ is a fundamental domain for $\Gamma$ then

$$\bigcup_{\gamma \in \Gamma/\Gamma\cap H} F\gamma$$

is a fundamental domain for $\Gamma \cap H$. For the second, note that a fundamental domain $F$ for $\Gamma \cap H$ intersects any coset $gH$ in the $g$-translate of a fundamental domain for $\Gamma \cap H$ in $H$.

Finally note that the last expression for $A_t^\alpha$ converges by our assumption in (5.19) to

$$\frac{m_Y(Y)}{m_X(X)} \int_X \alpha\,\mathrm{d}m_X$$

as $t \to \infty$. □

Proof of the pointwise count in Theorem 5.22. We now suppose that the weak*-convergence discussed above holds, and that the family of sets $B_t$ is well-rounded as in Definition 5.21. From this we wish to derive the asymptotic

$$\frac{1}{m_{G/H}(B_t)} \left|(\Gamma H) \cap B_t\right| \longrightarrow \frac{m_Y(Y)}{m_X(X)}$$

as $t \to \infty$.

Let $\varepsilon > 0$ be arbitrary, and choose $\delta > 0$ so that

$$\frac{m_{G/H}(B_{t+\delta})}{m_{G/H}(B_t)} < 1 + \varepsilon$$

for all $t$, and choose a symmetric neighbourhood $U = U^{-1} \subseteq G$ of $I \in G$ with

$$UB_t \subseteq B_{t+\delta}$$

for all $t$. Further let $\alpha \in C_c(X)$ be an approximate identity at the identity coset, in the sense that $\alpha \geqslant 0$

$$\int_X \alpha\,\mathrm{d}m_X = 1,$$

and $\operatorname{supp}(\alpha) \subseteq U\mathit{\Gamma}$. Then we have for any $g \in U$ that

$$\begin{aligned} F_{t+\delta}(g) &= \frac{1}{m_{G/H}(B_{t+\delta})} \left| g\mathit{\Gamma} H \cap B_{t+\delta} \right| \\ &= \frac{1}{m_{G/H}(B_{t+\delta})} \Bigg| \mathit{\Gamma} H \cap \underbrace{g^{-1} B_{t+\delta}}_{\supseteq B_t} \Bigg| \\ &\geqslant \frac{m_{G/H}(B_t)}{m_{G/H}(B_{t+\delta})} \frac{1}{m_{G/H}(B_t)} \left| \mathit{\Gamma} H \cap B_t \right| \\ &\geqslant \frac{1}{1+\varepsilon} \frac{1}{m_{G/H}(B_t)} \left| \mathit{\Gamma} H \cap B_t \right|. \end{aligned}$$

Multiplying by $\alpha$, integrating with respect to $m_X$ and letting $t \to \infty$ gives

$$\limsup_{t\to\infty} \frac{1}{m_{G/H}(B_t)} \left| \mathit{\Gamma} H \cap B_t \right| \leqslant (1+\varepsilon) \frac{m_Y(Y)}{m_X(X)}.$$

The second inequality is derived in the same way. □

**Exercise 5.23.** Give a detailed argument to show that

$$\liminf_{t\to\infty} \frac{1}{m_{G/H}(B_t)} \left| \mathit{\Gamma} H \cap B_t \right| \geqslant (1+\varepsilon)^{-1} \frac{m_Y(Y)}{m_X(X)}.$$