Chapter 4
Quantitative Non-Divergence

In this chapter we will show that a unipotent trajectory cannot diverge to infinity in $\mathrm{SL}_d(\mathbb{R})/\mathrm{SL}_d(\mathbb{Z})$. In fact we will show that unipotent orbits have no 'escape of mass', which is also called 'quantitative non-divergence'. The topological claim is due to Margulis and the quantitative refinement is due to Dani [19, 21]. About 20 years later, the argument was further refined by Kleinbock and Margulis [82] and Kleinbock [83], and applied to various Diophantine problems. As a corollary we will also obtain a special case of the Borel–Harish-Chandra theorem [8]: $\mathbb{G}(\mathbb{Z})$ is a lattice in $\mathbb{G}(\mathbb{R})$ if $\mathbb{G}$ is a semisimple algebraic group defined over $\mathbb{Q}$.

4.1 The Case of the Modular Surface

We first describe a case that is both easy and familiar: Horocycle orbits on the unit tangent bundle X_2 of the modular surface. We refer to Section 1.2 or [45, Ch. 9] for the background and to [45, Ch. 11] for a more detailed proof.

4.1.1 A Topological Claim

In the hyperbolic description of X_2, the topological non-divergence claim is particularly easy to see.

Lemma 4.1 (Non-divergence for X_2). *For any $x \in \mathsf{X}_2$, the horocycle orbit $u_t{\scriptstyle\bullet}x$ does not go to infinity as $t \to \infty$, nor as $t \to -\infty$.*

Proof. Every $x \in \mathsf{X}_2$ corresponds to a point $(z, v) \in \mathrm{T}^1(\mathbb{H})$ with z chosen in the usual fundamental domain, which we denote by F, for $\mathrm{SL}_2(\mathbb{Z})$ in $\mathbb{H}$ (see Section 1.2). To prove the lemma we find for a given x a compact set K and a sequence $t_n \to \infty$ with $u_{t_n}{\scriptstyle\bullet}x \in K$ for all $n \geqslant 1$. If x is periodic under the action of $\{u_t \mid t \in \mathbb{R}\}$ then the orbit is compact and we may take

$$K = \{u_t \cdot x \mid t \in \mathbb{R}\}$$

and obtain the claim trivially. Otherwise, we may take

$$K = \{(z, v) \mid z \in F, \Im(z) \leqslant 1\}.$$

Then it is easy to see (from the geometric picture of the horocycle flow) that there exists some $t_1 \geqslant 0$ with $u_{t_1} \cdot x \in K$, as illustrated in Figure 4.1. In fact, the

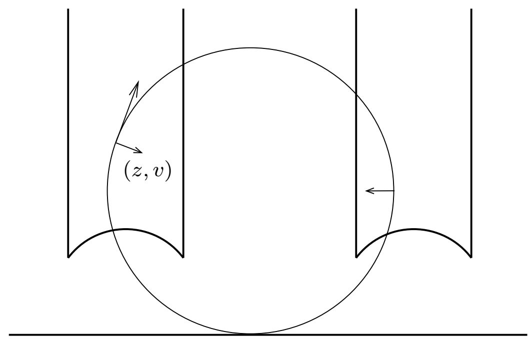

Fig. 4.1: A horocycle orbit returns to K.

horocycle orbit is a circle touching $\mathbb{R}$. Hence it moves up and then down again, returning to K. Now consider the point $u_{t_1+1} \cdot x$, and apply the same argument to find some $t_2 \geqslant t_1 + 1$ with $u_{t_2} \cdot x \in K$. Repeating the argument proves the lemma by induction. □

4.1.2 Non-escape of Mass

While the topological statement in Lemma 4.1 above was easy to derive from the hyperbolic geometry of horocycle orbits, the quantitative claim is more difficult to see from this geometric picture. Hence we will switch the description and think of X_2 as the space of unimodular lattices in $\mathbb{R}^2$.

Proposition 4.2 (Quantitative non-divergence for X_2). *A point $x \in \mathsf{X}_2$ is either periodic for the horocycle flow or*† *has the property that there exists some $T_x \geqslant 0$ such that for all $\varepsilon > 0$ and all $T \geqslant T_x$ we have*

$$\frac{1}{T} \left|\{t \in [0, T] \mid u_t \cdot x \notin \mathsf{X}_2(\varepsilon)\}\right| \ll \varepsilon. \tag{4.1}$$

† Note that the distinction of the two cases is absolutely necessary here: If $U \cdot x$ is a periodic orbit that is stuck high up in the cusp (equivalently a periodic orbit of short period), then the estimate (4.1) cannot hold uniformly for all $\varepsilon \leqslant 1$.

Here we are using $|A|$ as a shorthand for the Lebesgue measure of a measurable subset $A \subseteq \mathbb{R}$, and the notation

$$\mathsf{X}_2(\varepsilon) = \{x \in \mathsf{X}_2 \mid \lambda_1(x) \geqslant \varepsilon\}$$

introduced in Section 1.4.3.

Proof of Proposition 4.2. Suppose that x is not periodic, fix $T > 0$, and define, for every vector $v \in \Lambda_x \smallsetminus \{0\}$, a 'protecting' interval

$$P_v = \{t \in [0,T] \mid \|u_t v\| < 1\}.$$

Notice that if $v = (v_1, v_2)^{\mathrm{t}}$, then

$$\|u_t v\| = \left\| \begin{pmatrix} v_1 + t v_2 \\ v_2 \end{pmatrix} \right\| = \sqrt{(v_1 + t v_2)^2 + v_2^2},$$

and so P_v is a subinterval of $[0,T]$. If $v \in \Lambda_x \smallsetminus \{0\}$ is large enough (how large depends on T), then P_v is trivial. Hence there are only finitely many nontrivial intervals. Note that the unimodular lattice $u_t \Lambda_x$ for $t \in \mathbb{R}$ cannot contain two linearly independent vectors of length strictly less than 1. Hence these intervals can only intersect if they are associated to linearly dependent vectors. To rule even this out, we may choose within every Λ_x-rational line (that is, every line $\mathbb{R}v$ with $v \in \Lambda_x \smallsetminus \{0\}$) one and only one primitive vector in the lattice (that is, a vector $v \in \Lambda_x \smallsetminus \{0\}$ with $\mathbb{R}v \cap \Lambda_x = \mathbb{Z}v$). If $v \in \Lambda_x \smallsetminus \{0\}$ is primitive with $\|v\| < 1$ then $v_2 \neq 0$ (for otherwise x would be periodic) and we suppose that T is large enough to ensure that $\|u_T v\| > 1$.

Let $v^{(1)}, \dots, v^{(n)}$ be the resulting list of pairwise linearly independent primitive vectors, so that $P_i = P_{v^{(i)}}$ and

$$P_1 \sqcup \cdots \sqcup P_n = \{t \in [0,T] \mid \lambda_1(u_t \cdot x) < 1\}. \tag{4.2}$$

Now let $\varepsilon \geqslant 0$ and define the 'bad' set

$$B_i^{\varepsilon} = \{t \in [0,T] \mid \|u_t v^{(i)}\| \leqslant \varepsilon\}$$

for $i = 1, \dots, n$. We see that

$$B_1^{\varepsilon} \sqcup \cdots \sqcup B_n^{\varepsilon} = \{t \in [0,T] \mid \lambda_1(u_t \cdot x) \leqslant \varepsilon\}$$

is precisely the set whose measure we wish to estimate. For this, we claim that

$$|B_i^{\varepsilon}| \ll \varepsilon \, |P_i| \tag{4.3}$$

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

Summing this up, and using the disjointness in (4.2), the estimate (4.1) follows at once (for the case at hand, $\lambda_1(x) \geqslant 1$, and with $T_x = 0$).

To see the claim (4.3) we estimate both $|B_i^\varepsilon|$ and $|P_i|$ in terms of $|v_2^{(i)}|$. Notice first that we may assume $\varepsilon \leqslant \frac{1}{2}$ (for otherwise (4.3) is trivial) and hence $|v_2^{(i)}| \leqslant \frac{1}{2}$ (for otherwise B_i^ε is empty and (4.3) is trivial). With this it follows in a quite elementary way that $|P_i| \gg |v_2^{(i)}|^{-1}$ and $|B_i^\varepsilon| \ll \varepsilon |v_2^{(i)}|^{-1}$, see Figure 4.2 and Exercise 4.4.

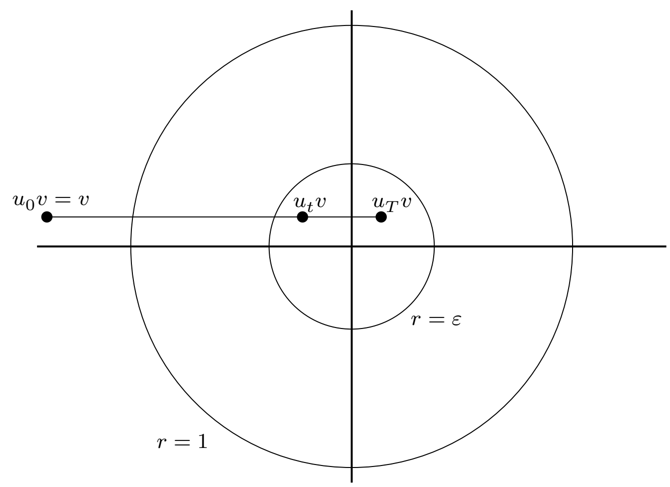

Fig. 4.2: The u_t-orbits of points $v \in \mathbb{R}^2$ travel at linear speed (determined by the second coordinate of v). Thus the set B of bad times where $\|u_t v\| \leqslant \varepsilon$ is a $\ll \varepsilon$-fraction of the protecting set P where $\|u_t v\| \leqslant 1$ if only $\max\{\|v\|, \|u_T v\|\} \geqslant 1$.

□

Corollary 4.3 (Non-escape of mass for X_2). *If $x \in \mathsf{X}_2$, then every weak*-limit of the collection of measures*

$$\Big\{\frac{1}{T}\int_0^T (u_t)_*\delta_x \,\mathrm{d}t \mid T \geqslant 0\Big\}$$

is a probability measure on X_2.

Exercise 4.4. Give a more detailed proof of the claim in (4.3). Note that two intervals may need special attention as $\lambda_1(x)$ or $\lambda_1(u_T \bullet x)$ may be less than 1.

Exercise 4.5. Prove Corollary 4.3.

4.2 The Case of $\mathsf{X}_3 = \mathrm{SL}_3(\mathbb{R})/\,\mathrm{SL}_3(\mathbb{Z})$

The proof for the generalizations of Proposition 4.2 and Corollary 4.3 becomes significantly more involved for X_d with $d \geqslant 3$. We start with the case $d = 3$ because it is easier to envision and because it already contains all the main ingredients of the general case. Throughout we will again write $|B|$ for the Lebesgue measure of a subset $B \subseteq \mathbb{R}$.

4.2.1 Non-Escape of Mass for Polynomial Trajectories

Even though we are primarily interested in unipotent trajectories, we will prove a more general claim allowing for general *polynomial trajectories* of the shape

$$p(t)\mathbb{Z}^3 \in \mathsf{X}_3$$

for $t \geqslant 0$ or for $t \in [0, T]$ for some $T \geqslant 0$, where

$$p\colon \mathbb{R} \longrightarrow \mathrm{SL}_3(\mathbb{R}) \subseteq \mathrm{Mat}_3(\mathbb{R})$$

is a polynomial map taking values in $\mathrm{SL}_3(\mathbb{R})$. We say that p has degree no more than D if each matrix entry is a polynomial of degree no more than D.

Notice that if $\{u_t = \exp(tv) \mid t \in \mathbb{R}\}$ is a one-parameter unipotent subgroup (of which there are precisely two up to conjugation in $\mathrm{SL}_3(\mathbb{R})$) with Lie algebra $\mathbb{R}v$ then

$$p(t) = u_t g = \exp(tv)g$$

is a polynomial in t for any $g \in \mathrm{SL}_3(\mathbb{R})$. Hence a unipotent trajectory is also a polynomial trajectory. The generalization comes more or less for free in the sense that it does not complicate the proof, while the generalization does have interesting consequences (see Section 4.4).

Much like a short periodic orbit for the horocycle flow on X_2, there is always the possibility that there are ‘rational reasons’ for a polynomial trajectory to remain stuck in the cusp in the following sense. The polynomial trajectory $p(t)\mathbb{Z}^3$ for $t \in [0, T]$ would surely be entirely far out if $\varepsilon \in (0, 1]$ is small and there was a vector $v \in \mathbb{Z}^3$ with

$$\|p(t)v\| \leqslant \varepsilon$$

for all $t \in [0, T]$, or if there is a rational plane $V \subseteq \mathbb{R}^3$ for which

$$\mathrm{vol}\Big(p(t)V/p(t)(V \cap \mathbb{Z}^3)\Big) \leqslant \varepsilon^2$$

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

The last volume expression looks quite complicated but expresses the simple concept that we are studying the volume of the deformed plane with respect to the deformed lattice inside it. We now define some abbreviations for such expressions. For any $d \geqslant 2$ and any given discrete subgroup $\Lambda \leqslant \mathbb{R}^d$ spanning a subspace W (possibly of smaller rank) we write $\operatorname{covol}(\Lambda)$ as shorthand for the volume of W/Λ (taken with respect to the volume induced by the standard inner product on $\mathbb{R}^d$). Also if a polynomial p with values in $\mathrm{SL}_d(\mathbb{R})$ is given, we define for the study of the polynomial trajectory $p(t)\mathbb{Z}^d$ the expression

$$\operatorname{covol}(V,t) = \operatorname{covol}\Bigl(p(t)(V \cap \mathbb{Z}^d)\Bigr) = \operatorname{vol}\Bigl(p(t)V/p(t)(V \cap \mathbb{Z}^d)\Bigr)$$

for any rational subspace $V \subseteq \mathbb{R}^d$ and $t \in \mathbb{R}$.

To avoid the above mentioned 'rational constraints' for $d = 3$ we assume that $\mathcal{I} \subseteq \mathbb{R}$ is a compact interval and that there is some $\eta \in (0,1]$ such that

$$\sup_{t \in \mathcal{I}} \|p(t)v\| \geqslant \eta \tag{4.4}$$

for all $v \in \mathbb{Z}^3 \smallsetminus \{0\}$, and

$$\sup_{t \in \mathcal{I}} \operatorname{covol}\Bigl(p(t)(V \cap \mathbb{Z}^3)\Bigr) \geqslant \eta^2 \tag{4.5}$$

for all rational planes $V \subseteq \mathbb{R}^3$. Using our abbreviation we could combine these two estimates into the assumption that

$$\sup_{t \in \mathcal{I}} \operatorname{covol}(V,t) \geqslant \eta^{\dim V}$$

for any rational subspace $V \subseteq \mathbb{R}^3$. This unified treatment of all intermediate subspaces will be our view point in the general case (see Section 4.3), but will already play a role in the proof of the following theorem.[(22)]

Theorem 4.6 (Quantitative non-divergence for X_3). *Let $D \in \mathbb{N}$ and a compact interval $\mathcal{I} \subseteq \mathbb{R}$ be given. Suppose that p is a polynomial map with values in $\mathrm{SL}_3(\mathbb{R})$ so that $\operatorname{covol}(V,t)^2$ is a polynomial of degree at most $2D$ for any rational subspace $V \subseteq \mathbb{R}^3$. Suppose furthermore that the piece $p(t)\mathbb{Z}^3$, $t \in \mathcal{I}$ of the polynomial trajectory satisfies* (4.4) *and* (4.5) *for some $\eta \leqslant 1$. Then, for $\varepsilon \in (0,\eta]$,*

$$\Bigl|\Bigl\{t \in \mathcal{I} \;\Big|\; p(t)\mathbb{Z}^3 \notin \mathsf{X}_3(\varepsilon)\Bigr\}\Bigr| \ll_D \left(\frac{\varepsilon}{\eta}\right)^{1/D} |\mathcal{I}|.$$

We note that the alternating tensor product $\bigwedge^2(\mathbb{R}^3)$ may be identified with $\mathbb{R}^3$ by choosing (for example) the basis $e_2 \wedge e_3$, $e_3 \wedge e_1$ and $e_1 \wedge e_2$ where as usual e_1, e_2, e_3 is the standard basis of $\mathbb{R}^3$. In this way the map

$$(v,w) \in \mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow v \wedge w \in \textstyle\bigwedge^2 \mathbb{R}^3$$

is identified with the exterior product

$$(v, w) \in \mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow v \times w \in \mathbb{R}^3.$$

For a polynomial p with values in $\mathrm{SL}_3(\mathbb{R})$ and $t \in \mathbb{R}$ the linear map

$$\bigwedge^2 p(t)\colon \bigwedge^2(\mathbb{R}^3) \longrightarrow \bigwedge^2(\mathbb{R}^3)$$

is then the linear map with

$$e_i \wedge e_j \longmapsto (p(t)e_i) \wedge \bigl(p(t)e_j\bigr)$$

for $1 \leqslant i, j \leqslant 3$. This again defines a polynomial (of at most doubled degree) in t with values in $\mathrm{SL}\left(\bigwedge^2(\mathbb{R}^3)\right)$. Moreover, note that the covolume of $\mathbb{Z}v_1 + \mathbb{Z}v_2$ is equal to the area of the parallelogram spanned by v_1 and v_2, or equivalently the length of $v_1 \wedge v_2$ (identified with the exterior product $v_1 \times v_2$).

We also note that in the case of the orbit of the one-parameter unipotent subgroup given by

$$p(t) = \begin{pmatrix} 1 & t & \\ & 1 & \\ & & 1 \end{pmatrix} g, \tag{4.6}$$

we may take $D = 1$, while for that defined by

$$p(t) = \begin{pmatrix} 1 & t & \frac{1}{2}t^2 \\ & 1 & t \\ & & 1 \end{pmatrix} g \tag{4.7}$$

we may take $D = 2$ in Theorem 4.6.

There are two ways in which one can establish the assumptions (4.4) and (4.5), and both are important in applying Theorem 4.6.

Given p and an interval $\mathcal{I}$ containing 0, one can find $\eta > 0$ with the desired property, for example by taking

$$\eta = \min\left\{\lambda_1(\mathbb{Z}^3 p(0)), \sqrt{\alpha_2(\mathbb{Z}^3 p(0))}\right\}.$$

With this η our assumptions (4.4) and (4.5) hold trivially. The conclusion can be viewed as a weak but uniform form of recurrence. Starting in $\mathsf{X}_d(\eta)$ the polynomial trajectory spends most of its time in $\mathsf{X}_d(\varepsilon)$ if ε is chosen sufficiently small in terms of η.

Given a polynomial p with the property that $p(t)v$ is non-constant for any vector $v \in \mathbb{Z}^3 \smallsetminus \{0\}$ and $(p(t)v_1) \wedge (p(t)v_2)$ is non-constant for any linearly independent vectors $v_1, v_2 \in \mathbb{Z}^3$, one can find some $T_0 > 0$ such that for $T \geqslant T_0$ and $\mathcal{I} = [0, T]$ we can use $\eta = 1$. In fact, there are only finitely many vectors $v \in \mathbb{Z}^3$ with $\|p(0)v\| \leqslant 1$, and for each of them $p(t)v$ is non-constant and hence there must be some T_0 such that (4.4) holds for $T \geqslant T_0$ and $\eta = 1$. The argument to establish (4.5) is similar. In this case the conclusion is, for $T \geqslant T_0$, even stronger. Even if x is far out in the non-compact X_d, the polynomial trajectory will visit and spend most of its time in $\mathsf{X}_d(\varepsilon_0)$ for some $\varepsilon_0 > 0$ independent of x.

Exercise 4.7. Calculate $\bigwedge^2 p(t)$ for the two polynomial maps in (4.6) and (4.7) to verify the claims made concerning D.

4.2.2 A Lemma About Polynomials

We now prove a lemma which replaces the argument involving the linear function $v_1 + tv_2$ in the proof of Proposition 4.2 (see in particular Figure 4.1).

Lemma 4.8 (Small values of polynomials). *Let $p \in \mathbb{R}[t]$ be a polynomial of degree L, and fix $T > 0$. Then, for every $\varepsilon > 0$ and compact interval $\mathcal{I} \subseteq \mathbb{R}$,*

$$\left|\left\{t \in \mathcal{I} \mid |p(t)| < \varepsilon\|p\|_{\mathcal{I},\infty}\right\}\right| \ll_L \varepsilon^{1/L}|\mathcal{I}|, \tag{4.8}$$

where

$$\|p\|_{\mathcal{I},\infty} = \sup_{t\in\mathcal{I}} |p(t)|.$$

The situation is illustrated in Figure 4.3 for the polynomial $p(t) = t^4$.

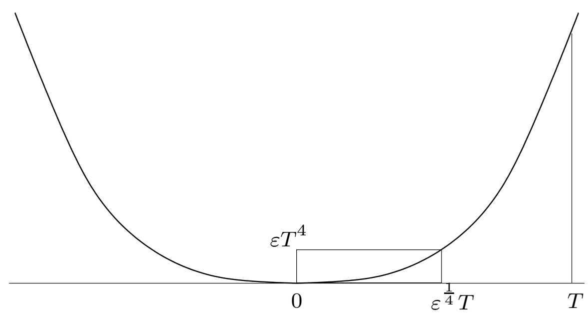

Fig. 4.3: The graph of $p(t) = t^L$ for $L = 4$ shows that the left-hand side of (4.8) can indeed be of the size $\varepsilon^{1/L}$.

The main property of polynomials that will be used in the proof of Theorem 4.6 is Lemma 4.8. A function or family of functions $p\colon \mathcal{I} \to \mathbb{R}$ is[†] *polynomial-like of degree no more than L*, or simply is *of degree no more than L* if p satisfies the conclusion of Lemma 4.8, and the implied constant does not depend on the particular function p if a whole family of such functions is being considered. We

[†] The more common, but less informative, terminology is (C, α)-good, where $\alpha = \frac{1}{L}$ and C is the implied constant.

will not pursue this generality here, and instead refer to the papers of Kleinbock and Margulis [82] and of Kleinbock [83].

PROOF OF LEMMA 4.8. Let $\delta > 0$ to be determined later, and suppose that there are $d+1$ points $t_0, \dots, t_d \in \mathcal{I} = [a,b]$ with $|t_i - t_j| \geqslant \delta$ for all $i \neq j$ and with $|p(t_i)| \leqslant \varepsilon \|p\|_{\mathcal{I},\infty}$ for all i. Since a polynomial of degree at most D points is determined by $D+1$ of its values we have

$$p(t) = \sum_{i=0}^{D} p(t_i) \prod_{j\neq i} \frac{t-t_j}{t_i - t_j}.$$

For $t \in \mathcal{I}$ this gives the estimate

$$|p(t)| \leqslant (D+1)\varepsilon \|p\|_{\mathcal{I},\infty} \frac{(b-a)^D}{\delta^D}$$

and so

$$\|p\|_{\mathcal{I},\infty} \leqslant (D+1)\varepsilon \|p\|_{\mathcal{I},\infty} \frac{(b-a)^D}{\delta^D}.$$

Solving for δ gives

$$\delta \leqslant \sqrt[D]{D+1}\,\sqrt[D]{\varepsilon}(b-a).$$

We set $\delta = 2\sqrt[D]{D+1}\,\sqrt[D]{\varepsilon}(b-a)$ and obtain that the set

$$\left\{t \in \mathcal{I} \mid |p(t)| \leqslant \varepsilon \|p\|_{\mathcal{I},\infty}\right\}$$

is contained in at most D subintervals of length 2δ. This implies the lemma. $\square$

4.2.3 Protection Arising From a Flag

The most important feature that makes the proof of Proposition 4.2 easier than the case of $\mathrm{SL}_3(\mathbb{R})$ considered here is the fact that a unimodular lattice $\Lambda \leqslant \mathbb{R}^2$ cannot have two linearly independent vectors of length less than one. This gave automatic 'protection' from short vectors: If there is a Λ-primitive vector of length less than one, and this vector is not tiny, then no tiny non-zero vector can exist in Λ. Using this we defined protecting intervals which were automatically disjoint.

This property of only one short vector is manifestly false for unimodular lattices in $\mathbb{R}^3$. For example, the lattice $\Lambda_n = \frac{1}{n}\mathbb{Z}e_1 + \frac{1}{n}\mathbb{Z}e_2 + n^2\mathbb{Z}e_3$ is unimodular for any $n \geqslant 1$, and contains two linearly independent vectors of length $\frac{1}{n}$. What we need to discuss in order to get a similar protection phenomenon in $\mathbb{R}^3$ are flags.

A *complete flag* in $\mathbb{R}^d$ is a collection

$$\mathscr{F} = \{V_0 = \{0\} \subsetneq V_1 \subsetneq V_2 \subsetneq \cdots \subsetneq V_d = \mathbb{R}^d\}$$

of nested subspaces, with one in each dimension from 0 to d. A *partial flag* is a collection of nested subspaces

$$\mathscr{F} = \{V_0 = \{0\} \subsetneq V_1 \subsetneq V_2 \subsetneq \cdots \subsetneq V_k = \mathbb{R}^d\}$$

with $k < d$. A subspace $V \subseteq \mathbb{R}^d$ is said to be *compatible* with a partial flag $\mathscr{F}$ if $V \notin \mathscr{F}$ and $\mathscr{F} \cup \{V\}$ is a (partial or complete) flag.

Lemma 4.9 (Protection coming from flags). *Let $\Lambda \leqslant \mathbb{R}^d$ be a unimodular lattice, and let*

$$V_0 = \{0\} \subseteq V_1 \subseteq V_2 \subseteq \cdots \subseteq V_d = \mathbb{R}^d$$

be a complete flag of Λ-rational subspaces. Then

$$\lambda_1(\Lambda) \geqslant \min_{i=1,\dots,d} \frac{\operatorname{covol}(\Lambda \cap V_i)}{\operatorname{covol}(\Lambda \cap V_{i-1})},$$

where $\operatorname{covol}(\{0\}) = \operatorname{covol}(\Lambda) = 1$.

This gives the desired protection in the following sense, illustrated for the case $d = 3$: If $\operatorname{covol}(\Lambda \cap V_1) \geqslant \varepsilon$ and $\operatorname{covol}(\Lambda \cap V_2) \geqslant \varepsilon^2$, then Λ does not contain vectors that are shorter than ε.

Proof of Lemma 4.9. Let $v \in \Lambda$ be chosen with norm $\|v\| = \lambda_1(\Lambda)$. If v does not lie in V_{d-1}, then the covolume of

$$\Lambda \cap V_{d-1} + \mathbb{Z}v \subseteq \mathbb{R}^d$$

is equal to $\operatorname{covol}(\Lambda \cap V_{d-1}) \cdot \|\pi(v)\|$, where $\pi\colon \mathbb{R}^d \to V_{d-1}^{\perp}$ is the orthogonal projection. In particular, since the covolume of $\Lambda \supseteq \Lambda \cap V_{d-1} + \mathbb{Z}v$ is 1, we have

$$\begin{aligned} 1 = \operatorname{covol}(\Lambda) &\leqslant \operatorname{covol}(\Lambda \cap V_{d-1} + \mathbb{Z}v) \\ &= \operatorname{covol}(\Lambda \cap V_{d-1})\|\pi(v)\| \leqslant \operatorname{covol}(\Lambda \cap V_{d-1})\|v\|, \end{aligned}$$

which implies the lemma in the case $v \notin V_{d-1}$.

Suppose now more generally that $v \in \Lambda \cap V_i$ but $v \notin V_{i-1}$ for some i in $\{1, \dots, d\}$. As before,

$$\begin{aligned} \operatorname{covol}(\Lambda \cap V_i) &\leqslant \operatorname{covol}(\Lambda \cap V_{i-1} + \mathbb{Z}v) \\ &= \operatorname{covol}(\Lambda \cap V_{i-1})\|\pi(v)\| \leqslant \operatorname{covol}(\Lambda \cap V_{i-1})\|v\|, \end{aligned}$$

where π is the appropriate projection. The lemma follows at once. □

To handle the lack of disjointness of the protecting intervals for individual vectors or subspaces we are also going to use a simple *covering lemma.*[(23)]

Lemma 4.10 (A covering lemma on intervals). *Let $\mathcal{I} \subseteq \mathbb{R}$ be a compact interval, and let $P_1, \dots, P_N \subseteq \mathcal{I}$ be a finite collection of compact sub-intervals. Then there exists a subcollection of these intervals $P_{j(1)}, \dots, P_{j(K)}$ which are* nearly disjoint *in the sense that*

$$\sum_{k=1}^{K} \mathbb{1}_{P_{j(k)}} \leqslant 2 \tag{4.9}$$

while still having the same union

$$\bigcup_{k=1}^{K} P_{j(k)} = \bigcup_{n=1}^{N} P_n. \tag{4.10}$$

Moreover, none of the selected intervals $P_{j(k)}$ is strictly contained in any of the original intervals $P_1, \ldots, P_N$.

Proof. We use induction on N, using what amounts to a simple greedy algorithm. If $N = 1$ (or the intervals are disjoint) there is nothing to prove. Moreover, we may assume that no interval is (properly or otherwise) contained in another. For otherwise we may simply remove the former interval without affecting their union

$$U = \bigcup_{n=1}^{N} P_n$$

and apply the inductive hypothesis. Let $c = \min U$ and let $j(1) \in \{1, \ldots, N\}$ be such that $P_{j(1)} = [c, d]$. Note that the assumed non-containment implies that $j(1)$ is uniquely determined.

We wish to apply the inductive hypothesis on $\mathcal{I}' = \mathcal{I} \smallsetminus [c, d]$. In fact if d is not an interior point of U we simply remove $P_{j(1)}$ from our list and apply the inductive hypothesis to $\mathcal{I}'$ and the remaining intervals to find $j(2), \ldots, j(K)$ so that $P_{j(1)}, \ldots, P_{j(K)}$ satisfy the lemma.

So assume now that d is an interior point. We again remove $P_{j(1)}$ from our list and define $P'_n = P_n \cap \mathcal{I}'$ for the remaining intervals. If several of these contain d we remove all but the longest of these. By the inductive assumption we find $j(2), \ldots, j(K)$ so that $P'_{j(2)}, \ldots, P'_{j(K)}$ are nearly disjoint and cover $U \cap \mathcal{I}'$. As only one, say $P'_{j(2)}$, contains d, we see that

$$P_{j(k)} = P'_{j(k)} \subseteq (d, \infty)$$

for $k = 3, \ldots, K$. Therefore $P_{j(1)}, P_{j(2)}, \ldots, P_{j(K)}$ are nearly disjoint with union U. This concludes the inductive step and so also the proof. □

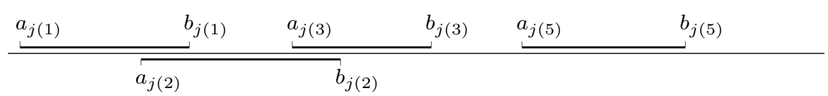

Fig. 4.4: By construction the selected intervals have the same union U and can be drawn without overlaps above or below the real line.

4.2.4 Non-Divergence for X_3—Obtaining Protecting Flags

In the course of the proof we will treat 1- and 2-dimensional subspaces on the same footing, so we will use the notation V uniformly for both from now on.

Proof of Theorem 4.6. Assume that $p\colon [0,T] \to \mathrm{SL}_3(\mathbb{R})$ has the property that

$$\mathrm{covol}\,(V,t)^2$$

is polynomial of degree no more than $2D$ for every rational subspace $V \subseteq \mathbb{R}^3$. Furthermore, let $\eta \leqslant 1$ satisfy (4.4) and (4.5), and fix $\varepsilon \in (0,\eta]$. We may assume $\varepsilon < 1$ as otherwise the conclusion of the theorem is trivial.

First stage protection intervals. Notice that there are only finitely many rational subspaces $V \subseteq \mathbb{R}^3$ for which

$$\mathrm{covol}\,(V,t) \leqslant \eta^{\dim V}$$

for some $t \in \mathcal{I}$. For each of those subspaces V we define the intervals $P_{V,\ell}$ for $\ell = 1, \dots, L_V$ to be the set of maximal subintervals[†] of

$$P_V = \{t \in \mathcal{I} \mid \mathrm{covol}\,(V,t) \leqslant \eta^{\dim V}\}.$$

Notice that by maximality of the subintervals and the assumptions (4.4) and (4.5) we have $\mathrm{covol}(V,t) = \eta^{\dim V}$ for at least one of the endpoints t of each of the intervals $P_{V,\ell}$. In particular

$$\sup_{t \in P_{V,\ell}} \mathrm{covol}\,(V,t) = \eta^{\dim V}. \tag{4.11}$$

This defines a collection of closed intervals $P_{V,\ell}$ where we vary both V (among lines and planes) and ℓ. Applying Lemma 4.10 to this collection and the interval $\mathcal{I}$, we obtain a nearly disjoint subcollection

$$P_1, \dots, P_m.$$

More precisely we have

$$\bigcup_V P_V = \bigcup_V \bigcup_{\ell=1}^{L_V} P_{V,\ell} = \bigcup_{i=1}^m P_i$$

and

$$\sum_{i=1}^m \mathbb{1}_{P_i} \leqslant 2.$$

[†] Each subinterval accounts for two roots of the polynomial equation $\mathrm{covol}\,(V,t)^2 = \eta^{2\dim V}$ so there can be at most D such intervals.

We write V_i for the subspace that gave rise to the interval P_i so that P_i is a maximal subinterval of $P_{V,i}$. As this subspace alone does not give protection (since Lemma 4.9 needs a complete flag and we only have one subspace), we need to do another search for a compatible subspace as follows.

Second stage protection intervals. Suppose first that V_i for $1 \leqslant i \leqslant m$ is a line. In this case we consider the intervals

$$P_{V,\ell} \cap P_i$$

for $\ell = 1, \ldots, L_V$ and all rational planes $V \subseteq \mathbb{R}^3$ that are compatible with V_i, in the sense that $V_i \subseteq V$. We apply the covering lemma on P_i to this collection to obtain nearly disjoint subintervals

$$P_{i,1}, \ldots, P_{i,n(i)} \subseteq P_i \tag{4.12}$$

with

$$\bigcup_{\substack{V_i \subseteq V,\\ V \text{ a plane}}} \bigcup_{\ell=1}^{L_V} P_{V,\ell} \cap P_i = \bigcup_{j=1}^{n(i)} P_{i,j} \subseteq P_i.$$

Similarly, if V_i for $i \in \{1, \ldots, m\}$ is a plane, then we obtain nearly disjoint subintervals as in (4.12) defined by compatible rational lines $V \subseteq V_i$ with

$$\bigcup_{\substack{V \subseteq V_i,\\ V \text{ a line}}} \bigcup_{\ell=1}^{L_V} P_{V,\ell} \cap P_i = \bigcup_{j=1}^{n(i)} P_{i,j} \subseteq P_i.$$

In both cases $n(i) = 0$ is possible.

Just as we denote by V_i the subspace that gave rise to the interval P_i, we also write $V_{i,j}$ for the subspace giving rise to $P_{i,j}$.

By construction V_i and $V_{i,j}$ are compatible (that is, they define a complete flag in $\mathbb{R}^3$) for all $i = 1, \ldots, m$ and $j = 1, \ldots, n(i)$. We will show that the intervals

$$P_1, \ldots, P_m, P_{1,1}, \ldots, P_{1,n(1)}, \ldots, P_{m,1}, \ldots, P_{m,n(m)}$$

together give the 'desired protection'.

Bad subsets. We now define for $\varepsilon > 0$ the associated bad subsets of the intervals above:

$$\mathrm{Bad}\,(i, \varepsilon) = \left\{ t \in P_i \;\middle|\; \mathrm{covol}\,(V_i, t) \leqslant \varepsilon \eta^{\dim V_i - 1} \right\},$$
$$\mathrm{Bad}\,(i, j, \varepsilon) = \left\{ t \in P_{i,j} \;\middle|\; \mathrm{covol}\left(V_{i,j}, t\right) \leqslant \varepsilon \eta^{\dim V_{i,j} - 1} \right\}$$

and the union

$$\mathrm{Bad}(\varepsilon)=\bigcup_{i=1}^{m}\left(\mathrm{Bad}\,(i,\varepsilon)\cup\bigcup_{j=1}^{n(i)}\mathrm{Bad}\,(i,j,\varepsilon)\right).$$

Estimate of bad subset. We now apply Lemma 4.8 to the polynomial

$$\mathrm{covol}\,(V_i,t)^2$$

of degree no larger than $2D$ on the interval P_i, with supremum norm $\eta^{2\dim V_i}$ by (4.11). This gives

$$|\mathrm{Bad}\,(i,\varepsilon)|\ll\left(\frac{\varepsilon}{\eta}\right)^{\frac{1}{D}}|P_i|\,, \tag{4.13}$$

by definition of $\mathrm{Bad}\,(i,\varepsilon)$.

To prove the same for $\mathrm{Bad}(i,j,\varepsilon)$ we need to show an analogue of (4.11) for the interval $P_{i,j}$. Recall that $P_{i,j}=P_{V_{i,j},\ell}\cap P_i$ for some $\ell\in\{1,\dots,L_{V_{i,j}}\}$. Now notice that by Lemma 4.10 (from the first application that gave rise to $P_1,\dots,P_i,\dots,P_m$) the intervals $P_{V_{i,j},\ell}$ cannot contain P_i properly—let us refer to this as the *non-containment.*

If both end points t of $P_{V_{i,j},\ell}$ satisfy $\mathrm{covol}\,(V_{i,j},t)=\eta^{\dim V_{i,j}}$ (because they are in $(0,T)$, for example) then (due to the non-containment) one of them must be in P_i, and so

$$\sup_{t\in P_{i,j}}\mathrm{covol}\,(V_{i,j},t)=\eta^{\dim V_{i,j}}. \tag{4.14}$$

If, on the other hand, we have $\mathrm{covol}\,(V_{i,j},t)<\eta^{\dim V_{i,j}}$ for one of the endpoints of $P_{V_{i,j},\ell}$ (this endpoint would have to be 0 or T), then the other will have to be in P_i (due to the non-containment) and we again get (4.14). Therefore, using Lemma 4.8 together with (4.14) as a replacement for (4.11) gives as before

$$|\mathrm{Bad}\,(i,j,\varepsilon)|\ll\left(\frac{\varepsilon}{\eta}\right)^{\frac{1}{D}}\left|P_{i,j}\right|. \tag{4.15}$$

Since the intervals $P_{i,j}\subseteq P_i$ are all nearly disjoint we get

$$\sum_{j=1}^{n(i)}\left|P_{i,j}\right|\leqslant 2\,|P_i|\,. \tag{4.16}$$

Thus we may take the union and use (4.13), (4.15) and (4.16) to obtain the estimate

$$
\begin{aligned}
|\mathrm{Bad}(\varepsilon)| &\leqslant \sum_{i=1}^{m} \left(|\mathrm{Bad}\,(i,\varepsilon)| + \sum_{j=1}^{n(i)} |\mathrm{Bad}(i,j,\varepsilon)| \right) \\
&\ll \left(\frac{\varepsilon}{\eta}\right)^{\frac{1}{D}} \sum_{i=1}^{m} |P_i| \leqslant 2 \left(\frac{\varepsilon}{\eta}\right)^{\frac{1}{D}} |\mathcal{I}|,
\end{aligned}
$$

since also the intervals $P_1, \dots, P_m \subseteq \mathcal{I}$ are nearly disjoint.

Protection. We now show that

$$
\left\{ t \in \mathcal{I} \;\middle|\; p(t)\mathbb{Z}^3 \notin \mathsf{X}_3(\varepsilon) \right\} \subseteq \mathrm{Bad}(\varepsilon), \tag{4.17}
$$

for all $\varepsilon \leqslant \eta$, so that the estimate above then implies the theorem.

Suppose therefore that $t \in \mathcal{I}$ has the property that $p(t)\mathbb{Z}^3$ contains an ε-short vector $p(t)v$. Since $\varepsilon \leqslant \eta$, this shows that t belongs to one of the protecting intervals defined by $V = \mathbb{R}v$. Hence we must have $t \in P_i$ for some $i \in \{1, \dots, m\}$ by choice of these intervals.

If $V = V_i$ then we have $t \in \mathrm{Bad}\,(i,\varepsilon) \subseteq \mathrm{Bad}(\varepsilon)$. If V_i is a line but $V \neq V_i$, then $V + V_i$ is a subspace compatible with V_i and

$$
\mathrm{covol}\,(V + V_i, t) \leqslant \mathrm{covol}\,(V,t)\,\mathrm{covol}\,(V_i,t) \leqslant \varepsilon\eta \leqslant \eta^2.
$$

Therefore $t \in P_i \cap P_{V+V_i,\ell}$ (for some ℓ) and so $t \in P_{i,j}$ for some $j \in \{1, \dots, n(i)\}$, by construction. We have obtained a complete flag: $V_i \subseteq V_{i,j}$ with

$$
t \in P_i \cap P_{i,j}.
$$

Suppose now that V_i is a plane and recall that

$$
\mathrm{covol}(V,t) \leqslant \varepsilon
$$

and

$$
\mathrm{covol}(V_i,t) \leqslant \eta^2.
$$

We may assume that $V \subseteq V_i$. For if $V + V_i = \mathbb{R}^3$, $\eta \leqslant 1$ and $\varepsilon < 1$ (which we may assume), we get a contradiction to the unimodularity of the three-dimensional lattice. Therefore, $t \in P_i \cap P_{V,\ell}$ for some ℓ and so there must exist some $j \in \{1, \dots, n(i)\}$ with $t \in P_{i,j}$. Once more we have obtained a complete flag: $V_{i,j} \subseteq V_i$ with $t \in P_i \cap P_{i,j}$.

Hence it remains to consider the case $t \in P_i \cap P_{i,j}$ for some i and j. Let us also assume, for the purposes of a contradiction, that

$$
t \notin \mathrm{Bad}\,(i,\varepsilon) \cup \mathrm{Bad}\,(i,j,\varepsilon)\,.
$$

Hence

$$
\varepsilon\eta^{\dim V_i - 1} \leqslant \mathrm{covol}(V_i,t) \leqslant \eta^{\dim V_i}
$$

and
$$\varepsilon\eta^{\dim V_{i,j}-1} \leqslant \operatorname{covol}(V_{i,j},t) \leqslant \eta^{\dim V_{i,j}},$$
and together V_i and $V_{i,j}$ define a flag in $\mathbb{R}^3$. Lemma 4.9 may now be applied to show that
$$\lambda_1\left(\mathbb{Z}^3 p(t)\right) \geqslant \min\left(\varepsilon, \varepsilon, \tfrac{1}{\eta^2}\right) = \varepsilon,$$
in contradiction to the assumption on t. This proves the claim (4.17), and hence the theorem. □

4.3 The General Case of $\mathsf{X}_d = \mathrm{SL}_d(\mathbb{R})/\mathrm{SL}_d(\mathbb{Z})$

Let us now state and prove the general version of the non-divergence theorem (using the abbreviations and tools introduced in the last section).

Theorem 4.11 (Quantitative non-divergence for X_d by Margulis, Dani and Kleinbock[(24)]). *Let*
$$p\colon \mathbb{R} \longrightarrow \mathrm{SL}_d(\mathbb{R})$$
be a polynomial, $\mathcal{I} \subseteq \mathbb{R}$ be a compact interval, and let $\eta \in (0,1]$. Assume that
$$\sup_{t\in\mathcal{I}} \operatorname{covol}(V,t) \geqslant \eta^{\dim V} \tag{4.18}$$
for all rational subspaces $V \subseteq \mathbb{R}^d$ and that $2D$ is an upper bound for the degrees of $\operatorname{covol}(V,t)^2$ for all rational subspaces $V \subseteq \mathbb{R}^d$. Then, for $\varepsilon \in (0,\eta]$,
$$\left|\left\{t \in \mathcal{I} \;\middle|\; p(t)\mathbb{Z}^d \notin \mathsf{X}_d(\varepsilon)\right\}\right| \ll_{d,D} \left(\frac{\varepsilon}{\eta}\right)^{\frac{1}{D}} |\mathcal{I}|. \tag{4.19}$$

Proof. The proof comprises the following steps:

- Iterated construction of intervals and flags with desired properties;
- Definition and estimate of the size of the bad subsets.
- Reaching the conclusion by combining the established properties.

Inductive step to construct the intervals and flags. Suppose we are given an interval $I_{\mathscr{F}} \subseteq \mathbb{R}$ and a partial flag
$$\mathscr{F} = \left\{\{0\} \subsetneq V_1 \subsetneq V_2 \subsetneq \cdots \subsetneq V_k \subsetneq \mathbb{R}^d\right\}$$
of rational subspaces of $\mathbb{R}^d$ with $0 \leqslant k < d-1$ such that
$$\sup_{t\in I_{\mathscr{F}}} \operatorname{covol}(V,t) \geqslant \eta^{\dim V} \tag{4.20}$$

for any rational subspace $V \leqslant \mathbb{R}^d$ that is compatible with $\mathscr{F}$.

We consider all rational subspaces $V \leqslant \mathbb{R}^d$ that are compatible with the partial flag $\mathscr{F}$. For each such subspace split

$$P_V = \left\{t \in I_{\mathscr{F}} \mid \operatorname{covol}(V,t) \leqslant \eta^{\dim V}\right\} \tag{4.21}$$

into its connected components, giving rise to subintervals

$$P_{V,1}, \dots, P_{V,L_V}.$$

Varying both V and the second index, we may apply Lemma 4.10 to obtain a nearly disjoint subcollection

$$P_1, \dots, P_m$$

of these intervals with the same union. That is,

$$\bigcup_{\substack{V \text{ compatible} \\ \text{with } \mathscr{F}}} \left\{t \in I_{\mathscr{F}} \mid \operatorname{covol}(V,t) \leqslant \eta^{\dim V}\right\} = \bigcup_{i=1}^m P_i$$

and

$$\sum_{i=1}^m |P_i| \leqslant 2|I_{\mathscr{F}}|.$$

Let us write V_i for the subspace that gave rise to the connected component P_i of P_{V_i}.

On each of those sub-intervals P_i we have the new (partial or complete) flag

$$\mathscr{F} \cup \{V_i\}.$$

Now let V be either V_i or a rational subspace that is compatible with $\mathscr{F} \cup \{V_i\}$. In particular, V is compatible with $\mathscr{F}$. Suppose for the moment that

$$\sup_{t \in P_i} \operatorname{covol}(V,t) < \eta^{\dim V}.$$

By (4.20) this shows that P_i is strictly contained in I. By continuity of the map $t \mapsto \operatorname{covol}(V,t)$ this implies that P_i is strictly contained in one of the connected components of P_V. However, that interval was considered in the construction of $P_1, \dots, P_m \subseteq I$ and we obtain a contradiction to Lemma 4.10. We therefore have

$$\sup_{t \in P_i} \operatorname{covol}(V,t) \geqslant \eta^{\dim V}$$

for $V = V_i$ and all V compatible with $\mathscr{F} \cup \{V_i\}$, giving (4.20) for the new flag. Moreover,

$$\sup_{t \in P_i} \operatorname{covol}(V_i,t) = \eta^{\dim V_i} \tag{4.22}$$

by combining the above with $P_i \subseteq P_{V_i}$ and the definition of P_{V_i} in (4.21).

Iterating the construction for constructing a finite tree. Initially we have

$$\mathscr{F}_{\varnothing}=\big\{\{0\}\subsetneq\mathbb{R}^d\big\},\ I_{\varnothing}=\mathcal{I},\ k=0.$$

In this case (4.20) is precisely the assumption (4.18) in Theorem 4.11. Applying the inductive step above defines intervals

$$P_1,\dots,P_m$$

and subspaces

$$V_1,\dots,V_m.$$

On each of the intervals P_{i_1} for $i_1=1,\dots,m$ the partial flag

$$\mathscr{F}_{i_1}=\{\{0\}\subsetneq V_{i_1}\subsetneq\mathbb{R}^d\}$$

satisfies the inductive hypothesis so that the inductive step can be repeated, giving rise to intervals

$$P_{i_1,i_2}\subseteq P_{i_1}$$

and partial flags

$$\mathscr{F}_{i_1,i_2}=\mathscr{F}_{i_1}\cup\{V_{i_1,i_2}\}$$

for a subspace V_{i_1,i_2} compatible with $\mathscr{F}_{i_1}$.

In general, let us write

$$\bar\imath=(i_1,\dots,i_k)$$

for the multi-index arising,

$$P_{\bar\imath}=P_{i_1,\dots,i_k}$$

for the intervals arising, and

$$\mathscr{F}_{\bar\imath}=\mathscr{F}_{i_1,\dots,i_k}$$

for the flags arising. The construction stops when, for a given interval $P_{\bar\imath}$ and flag $\mathscr{F}_{\bar\imath}$ there is no compatible rational subspace V for which

$$\big\{t\in P_{\bar\imath}\mid\operatorname{covol}(V,t)\leqslant\eta^{\dim V}\big\}$$

is non-empty. In particular the construction certainly stops if $\mathscr{F}_{\bar\imath}$ is a complete flag (with $k=d-1$). This may be thought of as a finite graded tree labeled by the intervals and the flags, as illustrated in Figure 4.5.

Let us highlight in this notation an important feature of the inductive construction. The last subspace $V_{\bar\imath}\in\mathscr{F}_{\bar\imath}$ found jointly with $P_{\bar\imath}$ satisfies

$$\sup_{t\in P_{\bar\imath}}\operatorname{covol}(V_{\bar\imath},t)=\eta^{\dim V_{\bar\imath}}\tag{4.23}$$

by (4.22).

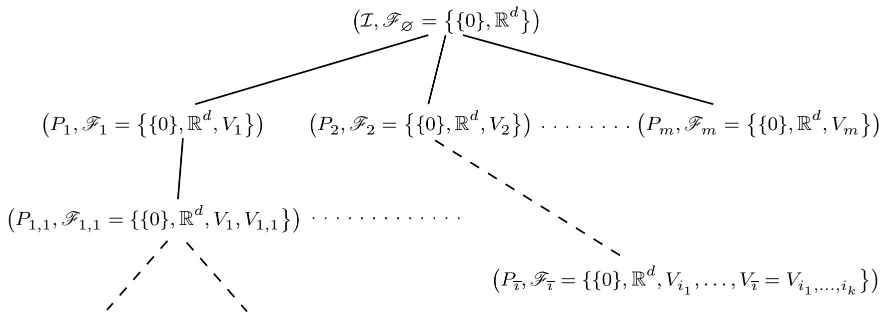

Fig. 4.5: Inductive construction of the intervals and flags.

Definition of Bad subsets. For any $(P_{\overline{\imath}}, \mathscr{F}_{\overline{\imath}})$ as constructed above, we define the following bad subset

$$\mathrm{Bad}(\overline{\imath}, \varepsilon) = \big\{t \in P_{\overline{\imath}} \mid \mathrm{covol}(V_{\overline{\imath}}, t) \leqslant \varepsilon \eta^{\dim V_{\overline{\imath}} - 1}\big\}.$$

Taking the union we define

$$\mathrm{Bad}(\varepsilon) = \bigcup_{\substack{\overline{\imath} = (i_1, \ldots, i_k), \\ k \geqslant 1}} \mathrm{Bad}(\overline{\imath}, \varepsilon)$$

Estimate for Bad subset. Applying Lemma 4.8 to the interval $P_{\overline{\imath}}$ and the polynomial $\mathrm{covol}(V_{\overline{\imath}}, t)^2$ (using (4.23) and the definition of $\mathrm{Bad}(\overline{\imath}, \varepsilon)$), we get

$$|\mathrm{Bad}(\overline{\imath}, \varepsilon)| \ll_D \left(\frac{\varepsilon}{\eta}\right)^{\frac{1}{D}} |P_{\overline{\imath}}| \,. \tag{4.24}$$

We now have to induct backwards to obtain the desired estimate for $\mathrm{Bad}(\varepsilon)$. In fact we claim that

$$\left| \bigcup_{(j_1, \ldots, j_s)} \mathrm{Bad}\left((\overline{\imath}, j_1, \ldots, j_s), \varepsilon\right) \right| \ll_{d,D} \left(\frac{\varepsilon}{\eta}\right)^{\frac{1}{D}} |P_{\overline{\imath}}| \,. \tag{4.25}$$

If $\{P_{\overline{\imath}}, \mathscr{F}_{\overline{\imath}}\}$ is a bottom leaf of the tree in Figure 4.5, then this is the same bound as (4.24). If, on the other hand, it is not then we may assume that (4.25) already holds for $(\overline{\imath}, j_1)$ for all $j_1 = 1, 2, \ldots$. Therefore

$$\left|\bigcup_{(j_1,\dots,j_s)} \mathrm{Bad}\left((\overline{\imath}, j_1, \dots, j_s), \varepsilon\right)\right| \leqslant |\mathrm{Bad}(\overline{\imath}, \varepsilon)| + \sum_{j_1} \left|\bigcup_{(j_2,\dots,j_s)} \mathrm{Bad}\left((\overline{\imath}, j_1, \dots, j_s), \varepsilon\right)\right|$$
$$\ll_{d,D} \left(\frac{\varepsilon}{\eta}\right)^{\frac{1}{D}} |P_{\overline{\imath}}| + \left(\frac{\varepsilon}{\eta}\right)^{\frac{1}{D}} \sum_{j_1} |P_{\overline{\imath},j_1}|$$

by (4.24) for $\mathrm{Bad}(\overline{\imath}, \varepsilon)$ and the inductive hypothesis. Since the intervals

$$P_{\overline{\imath},1}, \dots, P_{\overline{\imath},m} \subseteq P_{\overline{\imath}}$$

are nearly disjoint we also have

$$\sum_{j_1} |P_{\overline{\imath},j_1}| \leqslant 2|P_{\overline{\imath}}|,$$

which concludes the inductive step. For $\overline{\imath} = \varnothing$ (the root at the top of the graded tree) this shows

$$|\mathrm{Bad}(\varepsilon)| \ll_{d,D} \left(\frac{\varepsilon}{\eta}\right)^{\frac{1}{D}} |\mathcal{I}|. \tag{4.26}$$

Conclusion of the argument. It remains to show that

$$\{t \in \mathcal{I} \mid p(t)\mathbb{Z}^d \notin \mathsf{X}_d(\varepsilon)\} \subseteq \mathrm{Bad}(\varepsilon), \tag{4.27}$$

since (4.26) then proves the theorem. Suppose therefore that $t \in \mathcal{I}$ and

$$p(t)\mathbb{Z}^d \notin \mathsf{X}_d(\varepsilon),$$

or equivalently that there exists some vector $w \in \mathbb{Z}^d \smallsetminus \{0\}$ with $\|p(t)w\| < \varepsilon$. Since $\varepsilon \leqslant \eta$ we have $t \in P_{W,j}$ for $W = \mathbb{R}w$ and some j. Hence t lies in P_{i_1} for some i_1. If $t \in \mathrm{Bad}(i_1, \varepsilon) \subseteq \mathrm{Bad}(\varepsilon)$ then we have shown (4.27) for this value of t. So we may assume that $t \notin \mathrm{Bad}(i_1, \varepsilon)$. For the sake of the induction to come we continue the argument in greater generality.

Suppose we have reduced the problem to the case $t \in P_{\overline{\imath}}$ but

$$\varepsilon\eta^{\dim V - 1} < \mathrm{covol}(V, t) \leqslant \eta^{\dim V}$$

for all $V \in \mathscr{F}_{\overline{\imath}}$. Write

$$\mathscr{F}_{\overline{\imath}} = \left\{V_0 = \{0\} \subsetneq V_1 \subsetneq V_2 \subsetneq \cdots \subsetneq V_k \subsetneq \mathbb{R}^d\right\}$$

and assume that $a \in \{1, \dots, k\}$ is maximal with respect to the property

$$W = \mathbb{R}w \not\subseteq V_a.$$

This implies that

$$\operatorname{covol}(V_a + W, t) \leqslant \eta^{\dim V_a}\varepsilon \leqslant \eta^{\dim V_a + 1}$$

and so $V_a + W \notin \mathscr{F}_{\overline{\imath}}$ and $V_a + W$ is compatible with $\mathscr{F}_{\overline{\imath}}$ (since it contains V_a and is contained in V_{a+1}). In other words, $\mathscr{F}_{\overline{\imath}}$ is not a complete flag and t belongs to one of the intervals defined by $V_a + W$, so that $t \in P_{(\overline{\imath}, i_{k+1})}$ for some i_{k+1}. If $t \in \mathrm{Bad}(\overline{\imath}, i_{k+1}, \varepsilon)$ then we are again done. That is, we have the same situation as before and can repeat the argument.

The iterative argument above will only stop when t lies in $\mathrm{Bad}(\varepsilon)$. Since every time the argument repeats we know that we only had a partial flag, it can take at most d iterations to reach the conclusion. □

4.3.1 The Inner Product on the Alternating Tensor

We explain briefly how the natural inner product on $\bigwedge^k \mathbb{R}^d$ for $1 \leqslant k \leqslant d$ is related to k-dimensional volumes of parallelotopes in $\mathbb{R}^d$.

The inner product on $\bigwedge^k \mathbb{R}^d$ can be obbtained from the functorial property of $\bigwedge^k \mathbb{R}^d$. Indeed for $v_1, \dots, v_k, w_1, \dots, w_k \in \mathbb{R}^d$ the expression

$$\det\bigl((\langle v_i, w_j\rangle)_{i,j=1,\dots,k}\bigr)$$

depends multilinearly and alternatingly on $v_1, \dots, v_k$ as well as multilinearly and alternatingly on $w_1, \dots, w_k$. From this one obtains a bilinear map

$$\langle\cdot,\cdot\rangle\colon \bigwedge^k \mathbb{R}^d \to \mathbb{R}$$

satisfying

$$\langle v_1 \wedge \cdots \wedge v_k, w_1 \wedge \cdots \wedge w_k\rangle = \det\bigl((\langle v_i, w_j\rangle)_{i,j=1,\dots,k}\bigr)$$

for $v_1, \dots, v_k, w_1, \dots, w_k \in \mathbb{R}^d$. It is straightforward to verify that the standard basis $e_{i_1} \wedge \cdots \wedge e_{i_k}$ with $1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant d$ is an orthonormal basis and that $\langle\cdot,\cdot\rangle$ is an inner product.

Next note that for $v_1, \dots, v_k \in \mathbb{R}^k$ we have that

$$\begin{aligned}
\operatorname{vol}\bigl([0,1]v_1 + \cdots + [0,1]v_k\bigr) &= |\det(v_1, \dots, v_k)| \\
&= \sqrt{\det\bigl((v_1, \dots, v_k)^{\mathrm{t}}(v_1, \dots, v_k)\bigr)} \\
&= \sqrt{\det\bigl(\langle v_i, v_j\rangle_{i,j}\bigr)}.
\end{aligned}$$

Together with the above this means that

$$\|v_1 \wedge \cdots \wedge v_k\| = \sqrt{\langle v_1 \wedge \cdots \wedge v_k, v_1 \wedge \cdots \wedge v_k\rangle} = \sqrt{\det\bigl(\langle v_i, v_j\rangle_{i,j}\bigr)}$$

expresses the natural volume of $[0,1]v_1+\cdots+[0,1]v_k$ for $v_1,\ldots,v_k\in\mathbb{R}^d$.

Finally we note that for a given polynomial map p with values in $\mathrm{SL}_d(\mathbb{R})$ the linear maps $\bigwedge^k p(t)$ satisfying

$$\bigl(\textstyle\bigwedge^k p(t)\bigr)(v_1\wedge\cdots\wedge v_k)=\bigl(p(t)v_1\bigr)\wedge\cdots\wedge\bigl(p(t)v_k\bigr)$$

again defines a polynomial map $\bigwedge^k p$ with values in $\mathrm{SL}(\bigwedge^k\mathbb{R}^d)$ for $1\leqslant k\leqslant d-1$. For a rational subspace $V\subseteq\mathbb{R}^d$ and vectors $v_1,\ldots,v_k$ that form a $\mathbb{Z}$-basis of $V\cap\mathbb{Z}^d$ this implies that

$$\mathrm{covol}(V,t)^2=\bigl\|\bigl(p(t)v_1\bigr)\wedge\cdots\wedge\bigl(p(t)v_k\bigr)\bigr\|^2=\det\bigl(\langle p(t)v_i,p(t)v_j\rangle_{i,j}\bigr)$$

is a polynomial (with degree depending only on p).

4.3.2 Obtaining Invariant Probability Measures

Corollary 4.12 (Non-escape of mass for X_d). *If $x\in\mathsf{X}_d$ and*

$$U=\{u_t\mid t\in\mathbb{R}\}<\mathrm{SL}_d(\mathbb{R})$$

is a one-parameter unipotent subgroup, then every weak-limit of*

$$\frac{1}{T}\int_0^T(u_t)_*\delta_x\,\mathrm{d}t$$

for $T\to\infty$ is a U-invariant probability measure on X_d.

Proof. The invariance follows from the 'almost translation invariance' of $[0,T]$ for large T. Let $x=g\mathbb{Z}^d\in\mathsf{X}_d$, and define

$$\eta=\min\left\{\sqrt[k]{\alpha_k(g\mathbb{Z}^d)}\;\middle|\;1\leqslant k\leqslant d\right\}.$$

Fix an arbitrary $\varepsilon\in(0,\eta]$ and choose some $f\in C_c(\mathsf{X}_d)$ with

$$\mathbb{1}_{\mathsf{X}_d(\varepsilon)}\leqslant f\leqslant\mathbb{1}=\mathbb{1}_{\mathsf{X}_d}.$$

By choice of η and our discussion in Section 4.3.1 the assumptions for Theorem 4.11 are satisfied for some $D\in\mathbb{N}$. Therefore we have

$$1-c\left(\frac{\varepsilon}{\eta}\right)^{\frac{1}{D}}\leqslant\frac{1}{T}\int_0^T f\left(u_t\boldsymbol{\cdot}x\right)\,\mathrm{d}t\leqslant 1$$

for some constant $c=c_{d,D}$. Now choose a weak*-convergent subsequence of the measures

$$\frac{1}{T}\int_0^T (u_t)_*\,\delta_x\,\mathrm{d}t$$

to obtain the bound

$$1 - c\left(\frac{\varepsilon}{\eta}\right)^{\frac{1}{D}} \leqslant \int_{\mathsf{X}_d} f\,\mathrm{d}\mu$$

for the limit measure μ. Since $f \leqslant 1$ this shows that

$$\mu\left(\mathsf{X}_d\right) \geqslant 1 - c\left(\frac{\varepsilon}{\eta}\right)^{\frac{1}{D}}.$$

As $\varepsilon \in (0,\eta]$ was arbitrary, the corollary follows. □

Proposition 4.13 (Quantitative non-divergence for blocks). *Let U be an abelian unipotent subgroup of $\mathrm{SL}_d(\mathbb{R})$, and fix some coordinate system identifying U with $\mathbb{R}^\ell$ and with respect to which we can describe 'blocks' whose edges are parallel to the coordinate axes. Then for every $\delta > 0$ there exists a compact subset $K \subseteq \mathsf{X}_d$ with the property that for any $x \in \mathsf{X}_d$ either*

- *there is a Λ_x-rational U-invariant subspace V with $\mathrm{covol}(V) < 1$, or*
- *for any block $F \subseteq U$ containing I and with sufficiently large width we have*

$$\frac{1}{m_U(F)} m_U\left(\{u \in F \mid u\raisebox{0.2ex}{\tiny$\bullet$}x \in K\}\right) > 1 - \delta.$$

Proof. The corollary follows quite directly from the quantitative non-divergence in Theorem 4.11. Let D be chosen so that we may apply Theorem 4.11 for any one-parameter subgroup of U, and let $x \in \mathsf{X}_d$ be arbitrary.

If U fixes a Λ_x-rational subspace $V \subseteq \mathbb{R}^d$ with $\mathrm{covol}(V) < 1$, then there is nothing to prove. So suppose that this is not the case. As there are only finitely many Λ_x-rational subspaces with covolume less than 1, and for each such subspace the subgroup of U that fixes V is of codimension at least 1, there exists a one-parameter subgroup

$$U' = \{u'(t) \mid t \in \mathbb{R}\} \subseteq U$$

that does not fix any of these subspaces. Applying Theorem 4.11 to $p(t) = u'(t)g$ with $x = g\,\mathrm{SL}_d(\mathbb{Z})$, $\eta = 1$, some $\varepsilon_0 > 0$ (depending on the implicit constant in the non-divergence estimate (4.19) only), and some possibly very large T (depending on g), it follows that there exists at least one $t \in \mathbb{R}$ with

$$x' = u'(t)\raisebox{0.2ex}{\tiny$\bullet$}x \in \mathsf{X}_d(\varepsilon_0).$$

Now let $\{u_j(t_j) \mid t_j \in \mathbb{R}\}$ for $j = 1,\dots,\ell$ be the one-parameter subgroups of U corresponding to the chosen coordinate system in $U \cong \mathbb{R}^\ell$. We let $\mathcal{I}_1,\dots,\mathcal{I}_\ell \subseteq \mathbb{R}$ be compact intervals with

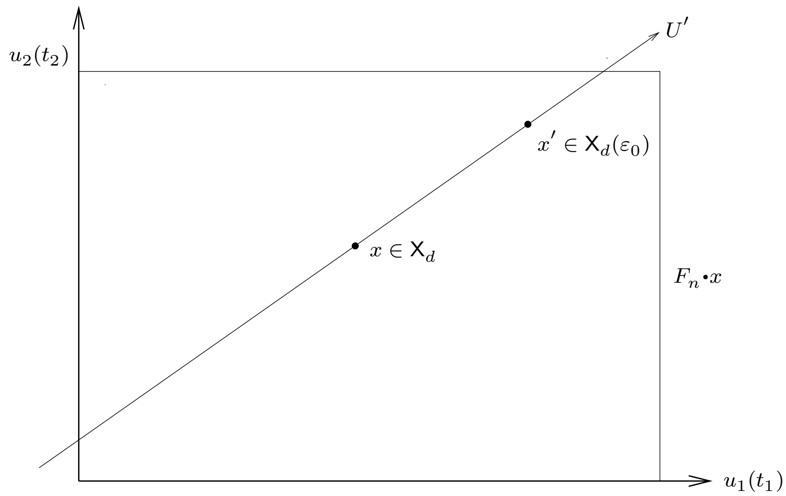

Fig. 4.6: The symmetric box $F{\boldsymbol{\cdot}}x$ inside the U-orbit has to contain $x' \in \mathsf{X}_d(\varepsilon)$ if the width of F is sufficiently large.

$$u(F){\boldsymbol{\cdot}}x = \bigl\{u(t_1)\cdots u_{t_\ell}{\boldsymbol{\cdot}}x' \mid t_1 \in \mathcal{I}_1, \dots, t_\ell \in \mathcal{I}_\ell\bigr\}.$$

If $F = F^{-1}$ is symmetric and the width of F is sufficiently big then we have $x' \in u(F){\boldsymbol{\cdot}}x$ and $0 \in \mathcal{I}_j$ for $j = 1, \dots, \ell$. This allows us to successively choose $\varepsilon_1, \dots, \varepsilon_\ell$ (depending only on δ) such that

$$\frac{1}{|\mathcal{I}_1|}\bigl|\bigl\{t_1 \in \mathcal{I}_1 \bigm| u_1(t_1){\boldsymbol{\cdot}}x' \notin \mathsf{X}_d(\varepsilon_1)\bigr\}\bigr| < \frac{\delta}{\ell},$$

and, if $u_1(t_1){\boldsymbol{\cdot}}x' \in \mathsf{X}_d(\varepsilon_1)$,

$$\frac{1}{|\mathcal{I}_2|}\bigl|\bigl\{t_2 \in \mathcal{I}_1 \bigm| u_1(t_1)u_2(t_2){\boldsymbol{\cdot}}x' \notin \mathsf{X}_d(\varepsilon_2)\bigr\}\bigr| < \frac{\delta}{\ell},$$

and so on, ending with

$$\frac{1}{|\mathcal{I}_\ell|}\bigl|\bigl\{t_k \in \mathcal{I}_\ell \bigm| u_1(t_1)\cdots u_k(t_\ell){\boldsymbol{\cdot}}x' \notin \mathsf{X}_d(\varepsilon_\ell)\bigr\}\bigr| < \frac{\delta}{\ell}$$

if $u_1(t_1)\cdots u_{\ell-1}(t_{\ell-1}) \in \mathsf{X}_d(\varepsilon_{\ell-1})$. We set $K = \mathsf{X}_d(\varepsilon_\ell)$, and the corollary follows for symmetric boxes. If the box F is not symmetric but contains I, then there is a symmetric box $F' \supseteq F$ with $|F'| \leqslant 2^\ell|F|$ and the proposition for F' and $\frac{\delta}{2^\ell}$ implies the proposition for F and δ. $\square$

Exercise 4.14. Let $G{\boldsymbol{\cdot}}x_0 \subseteq \mathsf{X}_d = \mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R})$ 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 U be a unipotent subgroup and let $U_1, \ldots, U_\ell < U$ be one-parameter subgroups such that $U = U_1 \cdots U_\ell$. Suppose $F_n = F_{1,n} \cdots F_{\ell,n}$ is a Følner sequence with respect to left and right translation where $F_{i,n} \subseteq U_i$ corresponds to an interval in U_i.

Prove that

$$\frac{1}{m_U(F_n)} m_U(\{u \in F_n \mid u{\bullet}x \notin \mathsf{X}_d(\delta)\}) \ll \delta^\kappa + o(1)$$

for $n \to \infty$ and some $\kappa > 0$ (depending on $U_1, \ldots, U_\ell$) in the following two cases:

(a) x belongs to a fixed compact subset and the implicit constant is allowed to depend on the compact subset, and
(b) x is arbitrary but U does not fix any Λ_x-rational subspace V of covolume $\eta^{\dim V}$ with $\eta \in (0,1)$ and the implicit constant is allowed to depend on η.

4.4 Inheritance in Diophantine Approximation

A general theme in the theory of Diophantine approximation is to try and show inheritance of Diophantine properties on $\mathbb{R}^d$ to submanifolds, or even more generally to fractals.[(25)] More precisely, for $S \subseteq \mathbb{R}^d$ equipped with a natural measure μ one may ask whether typical vectors in S with respect to μ have the same properties with respect to Diophantine approximations as Lebesgue almost every vector in $\mathbb{R}^d$. We will simply prove a few sample results in this direction without attempting to be comprehensive. For this we set $S = \{v(s) \mid s \in \mathbb{R}\}$ for the moment curve

$$v(s) = \begin{pmatrix} s \\ s^2 \\ \vdots \\ s^d \end{pmatrix} \in \mathbb{R}^d. \tag{4.28}$$

The following are corollaries of the quantitative non-divergence result Theorem 4.11, and special cases of results due to Kleinbock and Weiss and Kleinbock and Margulis.

Corollary 4.15. *For $d \geqslant 2$ there exists some $\lambda_0 \in (0,1)$ such that for almost all $s \in \mathbb{R}$ the vector $v(s)$ as in* (4.28) *is not λ_0-Dirichlet improvable.*

Corollary 4.16. *Let $d \geqslant 2$. Then for almost every $s \in \mathbb{R}$ the vector $v(s)$ as in* (4.28) *is not very well approximable.*

We start our proofs with the second corollary.

Proof of Corollary 4.16. It is sufficient to restrict to a compact interval $\mathcal{I}$ and to consider a fixed $\kappa > 1 + \frac{1}{d}$ in (2.13). By the dynamical interpretation of very well approximability in Proposition 2.37 there exist constants $c, \alpha > 0$ so that if $v(s)$ is very well approximable with exponent κ in (2.13), then $\lambda_1(a_n \Lambda_{v(s)}) \leqslant c\mathrm{e}^{-\alpha n}$ for infinitely many $n \in \mathbb{N}$. For a fixed n we apply Theorem 4.11 for the polynomial

$$p_n(s) = a_n u_v(s) = \begin{pmatrix} \mathrm{e}^{-dn} & \\ & \mathrm{e}^n I_d \end{pmatrix} \begin{pmatrix} 1 & \\ v(s) & I_d \end{pmatrix}.$$

We claim that for all sufficiently large n (depending on $\mathcal{I}$) we may set $\eta = 1$ (see Lemma 4.17 below). For an appropriate $D \geqslant d$ depending only on d we therefore have

$$\left|\{s \in \mathcal{I} \mid \lambda_1(a_n \Lambda_{v(s)}) \leqslant c\mathrm{e}^{-\alpha n}\}\right| \ll_d \mathrm{e}^{-\frac{\alpha}{D} n} |\mathcal{I}|$$

for all sufficiently large n. Hence Borel–Cantelli may be applied to give the corollary. □

Lemma 4.17. *For any compact interval $\mathcal{I} \subseteq \mathbb{R}$ with $|\mathcal{I}| > 0$ there exists some T_0 such that for $t \geqslant T_0$ the polynomial map $p_t(s) = a_t u_{v(s)}$ satisfies the assumption* (4.18) *of Theorem* 4.11 *with $\eta = 1$.*

Proof. Let $t \geqslant 0$ and suppose that $V \subseteq \mathbb{R}^d$ is a rational subspace satisfying

$$\sup_{s \in \mathcal{I}} \operatorname{covol}(V, s) < 1$$

and let $k = \dim V$. We fix some $s \in \mathcal{I}$ so that

$$\operatorname{covol}(V, s) = \operatorname{vol}\Big(a_t u_{v(s)} V / a_t u_{v(s)} \big(V \cap \mathbb{Z}^{d+1}\big)\Big) < 1. \tag{4.29}$$

Our goal is to derive from this the geometric information that $u_{v(s)} V$ almost contains the first basis vector e_1 with an error that is exponentially small in t. Applying $u_{v(s)}^{-1}$ we then obtain that V almost contains $u_{v(s)}^{-1} e_1$. Varying $s \in \mathcal{I}$ will lead to a contradiction if t is sufficiently large. Along the way $\mathcal{I}$ will affect multiplicative constants, which will be dominated by an exponentially decaying function in t.

A geometric estimate. To simplify the notation we set $W = u_{v(s)} V$ and let $v_1, \dots, v_k \in V \cap \mathbb{Z}^{d+1}$ be a $\mathbb{Z}$-basis so that $\|v_1 \wedge \cdots \wedge v_k\| \geqslant 1$ (simply because $\|v_1 \wedge \cdots \wedge v_k\|^2 \in \mathbb{N}$). Applying $u_{v(s)}$ gives

$$c = \left\|u_{v(s)} v_1 \wedge \cdots \wedge u_{v(s)} v_k\right\| \gg_{\mathcal{I}} 1, \tag{4.30}$$

where the multiplicative constant depends on $v(\cdot)$ restricted to $\mathcal{I}$. For the following geometric calculation in $\bigwedge^k \mathbb{R}^d$ it is useful to pick an orthonormal basis $w_1, \dots, w_k$ of W, where we may assume that $w_2, \dots, w_k$ are orthogonal to e_1. We also write $w_1 = b e_1 + w_1^\perp$ for some $b \in \mathbb{R}$ and $w_1^\perp \in \mathbb{R}^d$ orthogonal to e_1. We may assume that $b \geqslant 0$. It follows that $w_1 \wedge \cdots \wedge w_k$ has norm 1 and is a multiple of $u_{v(s)} v_1 \wedge \cdots \wedge u_{v(s)} v_k$. Comparing with (4.30) gives

$$u_{v(s)} v_1 \wedge \cdots \wedge u_{v(s)} v_k = \pm c\big(w_1 \wedge \cdots \wedge w_k\big).$$

Moreover

$$w_1 \wedge \cdots \wedge w_k = be_1 \wedge w_2 \wedge \cdots \wedge w_k + w_1^\perp \wedge w_2 \cdots \wedge w_k$$

is an orthonormal sum expanding $w_1 \wedge \cdots \wedge w_k$ into two eigenfunctions for $\bigwedge^k a_t$. We apply a_t and combine the above to obtain

$$\begin{aligned} 1 > \operatorname{covol}(V, s) &= \left\| a_t u_{v(s)} v_1 \wedge \cdots \wedge a_t u_{v(s)} v_k \right\| \\ &= c \left\| a_t w_1 \wedge \cdots \wedge a_t w_k \right\| \\ &\gg_{\mathcal{I}} \left\| a_t w_1^\perp \wedge a_t w_2 \wedge \cdots \wedge a_t w_k \right\| = \mathrm{e}^{kt} \left\| w_1^\perp \right\| \end{aligned}$$

or, equivalently,

$$\left\| w_1^\perp \right\| \ll_{\mathcal{I}} \mathrm{e}^{-kt}.$$

This implies that e_1 is exponentially close to w_1. Indeed, as $w_1 \in W$ is a unit vector and $b \geqslant 0$ we obtain

$$0 \leqslant 1 - b^2 = \left\| w_1^\perp \right\|^2 \ll_{\mathcal{I}} \mathrm{e}^{-2kt}$$

and with

$$\|e_1 - w_1\|^2 = 2 - 2\langle e_1, w_1 \rangle = 2(1-b)\frac{1+b}{1+b} \leqslant 2(1-b^2) \ll_{\mathcal{I}} \mathrm{e}^{-2kt},$$

also $\|e_1 - w_1\| \ll_{\mathcal{I}} \mathrm{e}^{-kt}$.

Varying $s \in \mathcal{I}$. Applying $u_{v(s)}^{-1}$ it follows that the distance of $u_{v(s)}^{-1} e_1$ to the rational subspace V is bounded by $\ll_{\mathcal{I}} \mathrm{e}^{-kt}$ for all $s \in \mathcal{I}$. We fix pairwise different $s_1, \ldots, s_{d+1} \in \mathcal{I}$ and obtain from this that for $j = 1, \ldots, d+1$ there exists a vector $u_j \in V$ with

$$\left\| u_j - \begin{pmatrix} 1 \\ -s_j \\ \vdots \\ -s_j^d \end{pmatrix} \right\| \ll_{\mathcal{I}} \mathrm{e}^{-kt}. \tag{4.31}$$

The so-called Vandermonde determinant formula then implies that

$$\begin{pmatrix} 1 & \cdots & 1 \\ -s_1 & & -s_{d+1} \\ \vdots & & \vdots \\ -s_1^d & & -s_{d+1}^d \end{pmatrix}$$

is invertible. Recall that $\mathrm{GL}_{d+1}(\mathbb{R}) \subseteq \mathrm{Mat}_{d+1}(\mathbb{R})$ is open. Hence if t is sufficiently large (depending on $\mathcal{I}$) then the estimates in (4.31) imply that $(u_1, \ldots, u_{d+1})$ is also invertible. However, as $u_1, \ldots, u_{d+1} \in V$ this implies that $V = \mathbb{R}^{d+1}$ and gives a contradiction to our assumption in (4.29). □

Proof of Corollary 4.15. For $\varepsilon > 0$ we define

$$O_\varepsilon = \left\{\varLambda \in \mathsf{X}_{d+1} \mid \varLambda \cap [-\varepsilon, \varepsilon]^{d+1} = \{0\}\right\}$$

as in the proof of Corollary 2.29. We wish to apply Theorem 4.11 on X_{d+1} for the polynomial map $s \mapsto p_t(s)$ as in Lemma 4.17. We will fix $\mathcal{I}$ below but assume for now that t is sufficiently large (depending on $\mathcal{I}$) so that we may indeed set $\eta = 1$. By Theorem 4.11, and the equivalence of the norms $\|\cdot\|$ and $\|\cdot\|_\infty$ on $\mathbb{R}^{d+1}$, there exists some $\varepsilon > 0$ (only depending on d) so that

$$\frac{1}{|\mathcal{I}|}\left|\left\{s \in \mathcal{I} \,\middle|\, p_t(s)\mathbb{Z}^{d+1} \notin O_\varepsilon\right\}\right| \leqslant \frac{1}{4}. \tag{4.32}$$

Now assume that the corollary is false for $\lambda_0 = \varepsilon$. Then

$$DT_\varepsilon = \{s \in \mathbb{R} \mid v(s) \text{ is } \varepsilon\text{-Dirichlet improvable}\}$$

must have a Lebesgue density point. In particular, it follows that there exists an interval $\mathcal{I} = [\alpha, \beta] \subseteq \mathbb{R}$ such that

$$\frac{1}{|\mathcal{I}|}\left|\left\{t \in \mathcal{I} \,\middle|\, v(s) \text{ is } \varepsilon\text{-Dirichlet improvable}\right\}\right| \geqslant \frac{3}{4}.$$

Using the definition of ε-Dirichlet improvable (and the basic property of measures), we find some N_0 such that

$$\frac{1}{|\mathcal{I}|}\left|\left\{s \in \mathcal{I} \,\middle|\, v(s) \text{ satisfies (2.9) with } \lambda = \varepsilon \text{ for every } N \geqslant N_0\right\}\right| \geqslant \frac{1}{2}.$$

Using Dani's correspondence (Proposition 2.28), we can also phrase this as

$$\frac{1}{|\mathcal{I}|}\left|\left\{s \in \mathcal{I} \,\middle|\, p_{\log N}(s)\mathbb{Z}^{d+1} \notin Q_\varepsilon \text{ for every } N \geqslant N_0\right\}\right| \geqslant \frac{1}{2}. \tag{4.33}$$

However, now that $\mathcal{I}$ is defined we have (4.32) for all $t \geqslant T_0$ with T_0 depending on $\mathcal{I}$. If we choose $N \geqslant \max\{N_0, \mathrm{e}^{T_0}\}$ and $t = \log N$ then the estimates (4.32) and (4.33) are incompatible. This contradiction to the existence of a Lebesgue density point proves the corollary. □

We will return to the topic of (inheritance in) Diophantine approximation one more time in Chapter 6.

4.5 Closed Orbits (often) Have Finite Volume

In this section we return to the discussion of orbits $H\boldsymbol{\cdot}x$ for a connected subgroup $H < \mathrm{SL}_d(\mathbb{R})$ and point $x \in \mathsf{X}_d = \mathrm{SL}_d(\mathbb{R})/\mathrm{SL}_d(\mathbb{Z})$. Recall that H is called *semisimple* if its Lie algebra is semisimple, and that this implies that H is an almost direct product of normal simple subgroups (which may be compact or non-compact; see Section 2.2). We say that the subgroup H is *unipotent* if H can be conjugated into the strict upper-triangular subgroup N (see (3.15) in Section 3.3). We note that semisimple subgroups without compact factors and connected unipotent subgroups are both *unipotently generated*, meaning that it is generated by finitely many one-parameter unipotent subgrous. For these subgroups we can give another connection between the property of having a closed orbit and the property of having an orbit of finite volume. In fact we will prove a partial converse to Corollary 1.36.

Theorem 4.18 (Borel–Harish-Chandra theorem, Part I). *Let $x \in \mathsf{X}_d$, and let $H < \mathrm{SL}_d(\mathbb{R})$ be a connected subgroup which is semisimple or unipotently generated. If the orbit $H{\scriptstyle\bullet}x$ is closed, then it has finite volume. In the case H is a connected unipotent subgroup, the orbit is compact.*

We refer to Exercise 4.20 for an immediate corollary (which is the standard way of phrasing the Borel–Harish-Chandra theorem) and to Section 9.4 for the general case of the theorem (which requires a few more definitions from the theory of algebraic groups).

Proof of Theorem 4.18 for semisimple subgroups. Let

$$H = H_1 \cdots H_\ell \cdot H_{\mathrm{compact}}$$

be the almost direct product of simple non-compact normal factors $H_1, \dots, H_\ell$ and a compact normal semisimple subgroup H_{compact} as in Section 2.2. Now choose, for each H_i, a nontrivial unipotent one-parameter subgroup

$$U_i = \{u_i(t) \mid t \in \mathbb{R}\}$$

and define the diagonally embedded unipotent subgroup

$$U = \{u_1(t)u_2(t)\cdots u_n(t) \mid t \in \mathbb{R}\}.$$

By Proposition 2.25, this subgroup $U \leqslant H$ satisfies the following form of the Mautner phenomenon: If H acts unitarily on a Hilbert space[†] $\mathcal{H}$ and a vector is fixed by U, then the same vector is fixed by $H_1 \cdots H_\ell$.

Now choose a compact set $K \subseteq H{\scriptstyle\bullet}x$ of positive volume with respect to the H-invariant Haar measure $m_{H{\scriptstyle\bullet}x}$ on the orbit $H{\scriptstyle\bullet}x \subseteq \mathsf{X}_d$ (as in Proposition 1.31 applied to $H/\operatorname{Stab}_H(x)$). Since $K \subseteq \mathsf{X}_d$ is compact, we can find some $\eta \in (0,1]$ such that

$$\alpha_k(\Lambda_x) \geqslant \eta^k$$

for $k = 1, \dots, d$ and any $x \in K$. Now apply Theorem 4.11 to find some $\varepsilon \in (0, \eta]$ with

$$\frac{1}{T}\left|\left\{t \in [0,T] \;\middle|\; u_t{\scriptstyle\bullet}x \notin \mathsf{X}_d(\varepsilon)\right\}\right| < \frac{1}{2} \tag{4.34}$$

for all $T > 0$. Since $H_{\mathrm{compact}} \subseteq H$ is compact and $H{\scriptstyle\bullet}x$ is closed, we have that (see Proposition 1.37)

[†] In this instance, the Hilbert space will be $L^2(H{\scriptstyle\bullet}x, m_{H{\scriptstyle\bullet}x})$.

$$H_{\mathrm{compact}}\mathsf{X}_d(\varepsilon)\cap H\boldsymbol{\cdot}x$$

is a compact subset of the orbit $H\boldsymbol{\cdot}x$, and so

$$f=\mathbb{1}_{H_{\mathrm{compact}}\mathsf{X}_d(\varepsilon)}\in L^2(H\boldsymbol{\cdot}x,m_{H\boldsymbol{\cdot}x})$$

is square-integrable with respect to $m_{H\boldsymbol{\cdot}x}$. We define

$$\underline{f}(y)=\liminf_{n\to\infty}\frac{1}{n}\int_0^n f(u_t\boldsymbol{\cdot}y)\,\mathrm{d}t.$$

As f is bounded, $\underline{f}$ is strictly U-invariant for $y\in H\boldsymbol{\cdot}x$ in the sense that

$$\underline{f}(u_t\boldsymbol{\cdot}y)=\underline{f}(y)$$

for all $t\in\mathbb{R}$ and $y\in H\boldsymbol{\cdot}x$. Notice that

$$\begin{aligned}\left\|\frac{1}{n}\int_0^n f(u_t\boldsymbol{\cdot}x)\,\mathrm{d}t\right\|^2_{L^2(m_{H\boldsymbol{\cdot}x})}&=\frac{1}{n^2}\int_0^n\int_0^n\underbrace{\int_{H\boldsymbol{\cdot}x}f(u_{t_1}\boldsymbol{\cdot}y)f(u_{t_2}\boldsymbol{\cdot}y)\,\mathrm{d}m_{H\boldsymbol{\cdot}x}(y)}_{\leqslant\|f\|^2_{L^2(m_{H\boldsymbol{\cdot}x})}}\\&\leqslant\|f\|^2_{L^2(m_{H\boldsymbol{\cdot}x})}.\end{aligned}$$

Hence, by Fatou's lemma, we get

$$\begin{aligned}\|\underline{f}\|^2_{L^2(m_{H\boldsymbol{\cdot}x})}&=\int_{H\boldsymbol{\cdot}x}\liminf_{n\to\infty}\left(\frac{1}{n}\int_0^n f(u_t\boldsymbol{\cdot}y)\,\mathrm{d}t\right)^2\,\mathrm{d}m_{H\boldsymbol{\cdot}x}(y)\\&\leqslant\liminf_{n\to\infty}\int_{H\boldsymbol{\cdot}x}\left(\frac{1}{n}\int_0^n f(u_t\boldsymbol{\cdot}y)\,\mathrm{d}t\right)^2\,\mathrm{d}m_{H\boldsymbol{\cdot}x}(y)\\&\leqslant\|f\|^2_{L^2(m_{H\boldsymbol{\cdot}x})}<\infty,\end{aligned}$$

or equivalently $\underline{f}\in L^2(m_{H\boldsymbol{\cdot}x})$.

We can now finish the proof quite quickly. Since $f\in L^2(m_{H\boldsymbol{\cdot}x})$ is u_t-invariant for all $t\in\mathbb{R}$ by construction, it is also $H_1\cdots H_\ell$-invariant[†] by the Mautner phenomenon (Proposition 2.25). Furthermore, $f=\mathbb{1}_{H_{\mathrm{compact}}\mathsf{X}_d(\varepsilon)}$ is invariant under H_{compact} by definition. Since u_t commutes with H_{compact}, it follows that $\underline{f}$ is also invariant under H_{compact}. Since $H=H_1\cdots H_\ell H_{\mathrm{compact}}$ and $\underline{f}\in L^2(m_{H\boldsymbol{\cdot}x})$ we see that $\underline{f}\equiv c$ is equal $m_{H\boldsymbol{\cdot}x}$-almost everywhere to some constant c. By definition and (4.34) we have $c\geqslant\frac{1}{2}$ and so

$$c^2 m_{H\boldsymbol{\cdot}x}(H\boldsymbol{\cdot}x)=\|\underline{f}\|^2_{L^2(m_{H\boldsymbol{\cdot}x})}<\infty$$

implies that $H\boldsymbol{\cdot}x$ has finite volume. □

[†] *A priori* this is invariance in L^2. However, by Exercise 2.6 we can choose a representative that is strictly $H_1\cdots H_\ell$-invariant.

Proof of Theorem 4.18 for unipotently generated subgroups. In the proof of Theorem 4.18 for the semisimple case it was convenient that we could find one one-parameter unipotent subgroup that satisfied the hypothesis of the Mautner phenomenon for 'most' of H. In the general case, we have instead to use finitely many one-parameter unipotent subgroups $U_j = \{u_j(t) \mid t \in \mathbb{R}\}$ for $j = 1, \dots n$ that together generate H.

Let $K \subseteq H\bullet x$ be a compact set. Then, finding first $\eta > 0$ and then $\varepsilon \in (0, \eta]$ as above, there exists a compact subset $L \subseteq H\bullet x$ (where $L = \mathsf{X}_d(\varepsilon) \cap H\bullet x$, relying on the assumption that $H\bullet x$ is closed) such that

$$\frac{1}{T}\left|\left\{t \in [0,T] \mid u_1(t)\bullet y \notin L\right\}\right| < \frac{1}{2} \tag{4.35}$$

for all $y \in K$. Now let $f = \mathbb{1}_L \in L^2(m_{H\bullet x})$ and

$$f_1(y) = \underline{f}(y) = \liminf_{n\to\infty} \frac{1}{n}\int_0^n f(u_1(t)\bullet y)\,\mathrm{d}t$$

so that $f_1 \in L^2(m_{H\bullet x})$, f_1 is U_1-invariant, and $f_1(y) \geqslant \frac{1}{2}$ for all $y \in K$.

Suppose now that for $j \leqslant n$ we have already shown that for any compact set $K \subseteq H\bullet x$ there exists some $f_j \in L^2(m_{H\bullet x})$ which is U_1-invariant, U_2-invariant, and so on up to U_j-invariant, and satisfies $f_j(y) \geqslant (\frac{1}{2})^j$ for all y in K. If $j = n$, then the function is H-invariant and the theorem follows as before.

So suppose that $j < n$ and let $K \subseteq H\bullet x$ be a compact subset. Now choose $L \subseteq H\bullet x$ as in (4.35) but for $u_j(t)$ instead of $u_1(t)$. Next apply the inductive hypothesis to L to find a function $f_j \in L^2(m_{H\bullet x})$ which is invariant under $U_1, U_2, \dots, U_j$ and satisfies $f_j(y) \geqslant (\frac{1}{2})^j$ for all $y \in L$. We define

$$f_{j+1}(y) = \underline{f_j}(y) = \liminf_{n\to\infty} \frac{1}{n}\int_0^n f_j(u_{j+1}(t)\bullet y)\,\mathrm{d}t.$$

By construction of f_j, L, and f_{j+1} we know that $f_{j+1} \in L^2(m_{H\bullet x})$, that f_{j+1} is U_{j+1}-invariant, and that $f_{j+1}(y) \geqslant (\frac{1}{2})^{j+1}$ for all $y \in K$. However, at first sight it may not be clear why f_{j+1} is still invariant under U_i for $i = 1, \dots, j$ (since U_i may not commute with U_{j+1}). Here the Mautner phenomenon comes to the rescue. In fact, by Theorem 2.55, f_j is actually invariant under a normal subgroup $N \lhd H$ containing $U_1, \dots, U_j$. By Exercise 2.6 we may once again replace f_j by an equivalent function that is strictly N-invariant. Therefore for all $i = 1, \dots, j$ and $s, t \in \mathbb{R}$ there exists some $n \in N$ with

$$u_{j+1}(t)u_i(s) = nu_{j+1}(t),$$

which shows that

$$f_j(u_{j+1}(t)u_i(s)\bullet y) = f_j(nu_{j+1}(t)\bullet y) = f_j(u_{j+1}(t)\bullet y)$$

for $y \in H\cdot x$. Integrating over $t \in [0, n]$ and taking the limit infimum as in the definition of f_{j+1}, we get

$$f_{j+1}(u_i(s)\cdot y) = f_{j+1}(y).$$

This concludes the induction and so also the proof of the first statement Theorem 4.18 for unipotently generated subgroups.

It remains to show that $H\cdot x$ is compact if H is unipotent. Note that by assumption, H can be conjugated into the upper triangular unipotent subgroup. On the upper triangular unipotent subgroup, the logarithm map is a polynomial with a polynomial inverse. This implies that H consists of the image of the Lie algebra of H. Hence we see that a unipotent connected subgroup H consists of the $\mathbb{R}$-points $H = \mathbb{H}(\mathbb{R})$ of an algebraic subgroup $\mathbb{H}$ over $\mathbb{R}$. If $x = g\,\mathrm{SL}_d(\mathbb{Z})$, then we may conjugate H by $g^{-1} \in G$ and assume without loss of generality that $x = \mathrm{SL}_d(\mathbb{Z})$. Then the Borel density theorem (Theorem 3.50) implies that the intersection $H \cap \mathrm{SL}_d(\mathbb{Z})$ is Zariski dense in $\mathbb{H}$, which in turn implies that $\mathbb{H}$ is an algebraic group over $\mathbb{Q}$. Hence the Lie algebra of H is a rational subspace of $\mathfrak{sl}_d(\mathbb{R})$, and by Theorem 3.20 we see that $H\cdot \mathrm{SL}_d(\mathbb{Z})$ is compact. □

Exercise 4.19. Let Q be a real non-degenerate quadratic form of signature (p, q) in $d \geqslant 3$ variables with $p \geqslant q \geqslant 1$. Suppose that the orbit $\mathrm{SO}_Q(\mathbb{R})\,\mathrm{SL}_d(\mathbb{Z})$ is closed. Show that a multiple of Q has integer coefficients.

Exercise 4.20. [(26)] Let $\mathbb{G} < \mathrm{SL}_d$ be a semisimple or unipotent algebraic group defined over $\mathbb{Q}$. Show that $\mathbb{G}(\mathbb{Z}) = \mathbb{G}(\mathbb{R}) \cap \mathrm{SL}_d(\mathbb{Z})$ is a lattice in $\mathbb{G}(\mathbb{R})$.

Notes to Chapter 4

[(22)](Page 150) This result, or rather its higher-dimensional counterpart in Section 4.3, has a long history; see Margulis [104], [105]; Dani [19], [21]; Kleinbock and Margulis [82]; Kleinbock [83].

[(23)](Page 154) This is a simple special case of the Besicovitch covering lemma (see [4]).

[(24)](Page 160) As mentioned before, this result has a long history; see Margulis [104], [105]; Dani [19], [21]; Kleinbock and Margulis [82]; Kleinbock [83].

[(25)](Page 169) We will not discuss this more general framework concerning the inheritance of Diophantine properties to 'sufficiently curved smooth manifolds' and simply mention here some of the key developments. Davenport and Schmidt [27] showed that almost every point of $\mathbb{R}^d$ is not Dirichlet-improvable and later showed in [28] that almost every point on the curve (t, t^2) is not (1/4)-improvable. Baker [3] extended this to the same statement for almost every point on a sufficiently smooth curve in $\mathbb{R}^2$, and to almost every point on a sufficiently smooth curved manifold by Dodson, Rynne, and Vickers [34]. Bugeaud [10] extended the result to the specific curve $(t, t^2, \ldots, t^d)$. Kleinbock and Weiss [86] used the correspondence introduced by Dani [20] and the machinery of Kleinbock and Margulis [82] to formulate some of these questions in homogeneous dynamics, and the argument used for the proof of Corollary 4.15 is the argument used in [86]. We refer to a paper of Shah [148] for more details on the background and for another direction of similar results for curves that do not lie in translates of proper subspaces.

[(26)](Page 176) This is a special case of the Borel–Harish-Chandra theorem [8]. We will return to it in Chapter 9.