QUANTITATIVE NONDIVERGENCE AND ITS DIOPHANTINE APPLICATIONS

DMITRY KLEINBOCK

December 2009

Abstract. The main goal of these notes is to describe a proof of quantitative nondivergence estimates for quasi-polynomial trajectories on the space of lattices, and show how estimates of this kind are applied to some problems in metric Diophantine approximation.

1. Introduction

These lecture notes constitute part of a course taught together with Alex Eskin at the Clay Mathematics Institute Summer School at Centro de Giorgi, Pisa, in June 2007. The exposition below is a continuation of [E]; the reader is referred there, as well as to books [BM, Mor, St] and the article [KSS] from the Handbook of Dynamical Systems, for background information on homogeneous spaces and unipotent flows.

In what follows, most of the work will be done on the space $\mathcal{L}_n$ of unimodular lattices in $\mathbb{R}^n$. We recall that $G = \mathrm{SL}(n,\mathbb{R})$ acts transitively on $\mathcal{L}_n$ (if $g \in G$ and $\Lambda \in \mathcal{L}_n$ is the $\mathbb{Z}$-span of the vectors $\mathbf{v}_1, \dots \mathbf{v}_n$, then $g\Lambda$ is the $\mathbb{Z}$-span of $\{g\mathbf{v}_1, \dots, g\mathbf{v}_n\}$), and the stabilizer of the standard lattice $\mathbb{Z}^n$ is $\Gamma = \mathrm{SL}(n,\mathbb{Z})$. This gives an identification of $\mathcal{L}_n$ with G/Γ. We choose a right-invariant metric on G; then this metric descends to G/Γ. Equivalently, one can define topology on $\mathcal{L}_n$ by saying that two lattices are close to each other if so are their generating sets.

For $\varepsilon > 0$ we will denote by $\mathcal{L}_n(\varepsilon) \subset \mathcal{L}_n$ the set of lattices whose shortest non-zero vector has norm at least ε. It is clear from the above description of the topology on $\mathcal{L}_n$ that any compact subset of $\mathcal{L}_n$ is contained in $\mathcal{L}_n(\varepsilon)$ for some positive ε. Conversely, one has

Theorem 1.1 (Mahler Compactness Criterion). *For any $\varepsilon > 0$ the set $\mathcal{L}_n(\varepsilon)$ is compact.*

See [Cas] or [BM] for a proof. We note that the set $\mathcal{L}_n(\varepsilon)$ depends on the choice of the norm on $\mathbb{R}^n$, but in a rather mild way: change of one norm for another would result in multiplication/division of ε by at most a fixed positive constant.

Recall that an element g of G is unipotent if all its eigenvalues are equal to 1. If $n = 2$, every one-parameter unipotent subgroup of $G = \mathrm{SL}(2, \mathbb{R})$ is conjugate to

$$U = \{u_x : x \in \mathbb{R}\} \text{ where } u_x = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} . \tag{1.1}$$

In general, a crucial property of an arbitrary unipotent subgroup $\{u_x\}$ of $\mathrm{SL}(n, \mathbb{R})$ is that the map $x \mapsto u_x$ is polynomial of degree depending only on n. This observation was instrumental in the proof due to Margulis that one-parameter unipotent trajectories on $\mathcal{L}_n$ are never divergent. Namely the following theorem was conjectured by Piatetski-Shapiro in the late 1960s and showed in 1971 by Margulis [Mar] as part of the program aimed at proving arithmeticity of lattices in higher rank algebraic groups:

Theorem 1.2. *Let $\{u_x\}$ be a one-parameter unipotent subgroup of $\mathrm{SL}(n, \mathbb{R})$. Then for any $\Lambda \in \mathcal{L}_n$, $u_x\Lambda$ does not tend to ∞ as $x \to \infty$. Equivalently, there exists $\varepsilon > 0$ such that the set $\{x \in \mathbb{R}_+ : u_x\Lambda \in \mathcal{L}_n(\varepsilon)\}$ is unbounded.*

In fact, ε in the above theorem can be chosen independent on the choice of $\{u_x\}$, although it does depend on Λ, see §3 for more detail. The above statement is very easy to prove when $n = 2$, but much more difficult for bigger n. In this exposition we first discuss the easy special case, then the general strategy of Margulis in various modifications, and then some applications and further extensions of the general result.

Acknowledgements: The author is grateful to the Clay Mathematics Institute for a wonderful opportunity to participate in the 2007 Summer School, to the co-organizers of the event for their help and encouragement, and to the staff at Centro di Georgi for being so very helpful and attentive. Special thanks are due to Alex Eskin, the co-lecturer of this course, to Elon Lindenstrauss for careful reading of a preliminary version of these notes, and to many participants of the summer school for their patience and valuable comments during and after the lectures. The work on the manuscript was partially supported by NSF grants DMS-0239463 and DMS-0801064.

2. Non-divergence of unipotent flows: the case of $\mathrm{SL}(2, \mathbb{R})$.

2.1. **Geometry of lattices in $\mathbb{R}^2$.** Recall the following lemma from [E]:

Lemma 2.1. *There exists $\varepsilon_0 > 0$ (depending on the choice of the norm on $\mathbb{R}^2$) such that no $\Lambda \in \mathcal{L}_2$ contains two linearly independent vectors each of norm less than ε_0.*

Let us now use this lemma to prove a nondivergence result for the U-action on $\mathcal{L}_2$, where U is as in (1.1):

Proposition 2.2. *For any $\Lambda \in \mathcal{L}_2$, $u_x\Lambda$ does not tend to ∞ as $x \to \infty$.*

In other words, for any $\Lambda \in \mathcal{L}_2$ there exists a compact subset K of $\mathcal{L}_2$ such that the set $\{x > 0 : u_x\Lambda \in K\}$ is unbounded.

Proof. Assume the contrary; in view of Theorem 1.1, this would amount to assuming that the norm of the shortest nonzero vector of $u_x\Lambda$ tends to zero as $x \to \infty$. Note that an obvious example of a divergent orbit would be constructed if one could find a vector $\mathbf{v} \in \Lambda \smallsetminus \{0\}$ such that $u_x\mathbf{v} \to 0$. But this is impossible: either $\mathbf{v}$ is horizontal and thus fixed by U, or its y-component is nonzero and does not change under the action. Thus the only allowed scenario for a divergent U-trajectory would be the following: for some $\mathbf{v} \in \Lambda \smallsetminus \{0\}$, $u_x\mathbf{v}$ gets very small, say shorter than ε, then starts growing but before it grows too big (longer than ε), another vector in $\Lambda \smallsetminus \{0\}$ not proportional to $\mathbf{v}$ gets shrunk by u_x to the length less than ε. This however is prohibited by Lemma 2.1. $\square$

Remark 2.3. Note that the analogue of this proposition is false if U is replaced by

$$A = \{a_t : t \in \mathbb{R}\} \text{ where } a_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}, \tag{2.1}$$

since a_t can contract nonzero vectors. However the same argument as above shows that for any continuous function $h : \mathbb{R}_+ \to \mathrm{SL}(2,\mathbb{R})$ and any $\Lambda \in \mathcal{L}_2$ such that $h(x)\Lambda$ diverges, it must do so in a degenerate way (a terminology suggested by Dani, see [D3]), that is, shrinking some nonzero vector $\mathbf{v} \in \Lambda$. This phenomenon is specific to dimension 2: if $n > 2$, as shown in [D3], one can construct divergent trajectories $\{a_t\Lambda\} \subset \mathcal{L}_n$ of diagonal one-parameter semigroups $\{a_t\} \subset \mathrm{SL}(n,\mathbb{R})$ in $\mathcal{L}_n$ which diverge in a non-degenerate way (without shrinking any subpace of $\mathbb{R}^n$).

Despite the above remark, Theorem 1.2, which is an analogue of Proposition 2.2, holds for $n > 2$ as well. An attempt to replicate the proof of Proposition 2.2 verbatim fails miserably: there are no obstructions to having many short linear independent vectors. We will prove Theorem 1.2 in the next section in a much stronger (quantitative) form, which also happens to have important applications to problems arising in Diophantine approximation theory. But first, following the methodology of [E] where the exposition of Ratner's theorem begins with an extensive discussion of the U-action on $\mathcal{L}_2$, we explain how one can easily establish a stronger form of Proposition 2.2, just for $n = 2$.

2.2. **Quantitative nondivergence in $\mathcal{L}_2$.** We are going to fix an interval $B \subset \mathbb{R}$ and $\Lambda \in \mathcal{L}_2$, and will look at the piece of trajectory $\{u_x\Lambda : x \in B\}$. Applying the philosophy of the proof of Proposition 2.2, one can see that one of the following two alternatives can take place:

Case 1. There exists a vector $\mathbf{v} \in \Lambda \smallsetminus \{0\}$ such that $\|u_x\mathbf{v}\|$ is small, say less than ε_0, for all $x \in B$. (For example this $\mathbf{v}$ may be fixed by U.) This case is not so interesting: again by Lemma 2.1, we know that this vector $\mathbf{v}$ is "the only source of trouble", namely no other vector can get small at the same time.

Case 2. The contrary, i.e.

$$\forall\, \mathbf{v} \in \Lambda \smallsetminus \{0\} \quad \sup_{x\in B} \|u_x\mathbf{v}\| \ge \rho\,. \tag{2.2}$$

In other words, every nonzero vector grows big enough at least at some point $x \in B$. This assumption turns out to be enough to conclude that for small ε the trajectory $\{u_x\Lambda : x \in B\}$ spends relatively small proportion of time, in terms of Lebesgue measure λ on $\mathbb{R}$, outside of $\mathcal{L}_2(\varepsilon)$.

Theorem 2.4. *Suppose an interval $B \subset \mathbb{R}$, $\Lambda \in \mathcal{L}_2$ and $0 < \rho < \varepsilon_0$ are such that (2.2) holds. Then for any $\varepsilon > 0$,*

$$\lambda(\{x \in B : u_x\Lambda \notin \mathcal{L}_2(\varepsilon)\}) \le 2\frac{\varepsilon}{\rho}\lambda(B)\,.$$

Thus, if one studies the curve $\{u_x\Lambda\}$ where x ranges from 0 to T, it suffices to look at the starting point Λ of the trajectory, find its shortest vector $\mathbf{v}$, choose $\rho < \min(\varepsilon_0, \|\mathbf{v}\|)$, and apply the theorem to get a quantitative statement concerning the behavior of $\{u_x\Lambda : 0 \le x \le T\}$ for any T. Note that it is meaningful, and requires proof, only when ε is small enough (not greater than $\rho/2$).

Proof. Denote by $P(\Lambda)$ the set of primitive vectors in Λ ($\mathbf{v}$ is said to be *primitive* in Λ if $\mathbb{R}\mathbf{v} \cap \Lambda$ is generated by $\mathbf{v}$ as a $\mathbb{Z}$-module). Clearly in all the argument it will suffice to work with primitive vectors.

Now for each $\mathbf{v} \in P(\Lambda)$ consider

$$B_{\mathbf{v}}(\varepsilon) \stackrel{\text{def}}{=} \{x \in B : \|u_x\mathbf{v}\| < \varepsilon\} \quad \text{and} \quad B_{\mathbf{v}}(\rho) \stackrel{\text{def}}{=} \{x \in B : \|u_x\mathbf{v}\| < \rho\}\,,$$

where $\|\cdot\|$ is the supremum norm. Let $\mathbf{v} = \begin{pmatrix} a \\ b \end{pmatrix} \in P(\Lambda)$ be such that $B_{\mathbf{v}}(\varepsilon) \ne \varnothing$. Then, since $u_x\mathbf{v} = \begin{pmatrix} a + bx \\ b \end{pmatrix}$, it follows that $|b| < \varepsilon$, and (2.2) implies that b is nonzero. Therefore, if we denote $f(x) = a + bx$, we have

$$B_{\mathbf{v}}(\varepsilon) = \{x \in B : |f(x)| < \varepsilon\} \quad \text{and} \quad B_{\mathbf{v}}(\rho) = \{x \in B : |f(x)| < \rho\}\,.$$

Clearly the ratio of lengths of intervals $B_{\mathbf{v}}(\varepsilon)$ and $B_{\mathbf{v}}(\rho)$ is bounded from above by $2\varepsilon/\rho$ (by looking at the worst case when $B_{\mathbf{v}}(\varepsilon)$ is close to one of the endpoints of B). Lemma 2.1 guarantees that the sets $B_{\mathbf{v}}(\rho)$ are disjoint for different $\mathbf{v} \in P(\Lambda)$, and also that $u_x\Lambda \notin \mathcal{L}_2(\varepsilon)$ whenever $x \in B_{\mathbf{v}}(\rho) \smallsetminus B_{\mathbf{v}}(\varepsilon)$ for some $\mathbf{v} \in P(\Lambda)$. Thus we conclude that

$$\lambda(\{x \in B : u_x\Lambda \notin \mathcal{L}_2(\varepsilon)\}) \le \sum_{\mathbf{v}} \lambda\big(B_{\mathbf{v}}(\varepsilon)\big) \le 2\frac{\varepsilon}{\rho}\sum_{\mathbf{v}} \lambda\big(B_{\mathbf{v}}(\rho)\big) \le 2\frac{\varepsilon}{\rho}\lambda(B)\,. \quad \square$$

Remark 2.5. Before proceeding to the more general case, let us summarize the main features of the argument. Each primitive vector $\mathbf{v}$ came with a function, $x \mapsto \|u_x\mathbf{v}\|$, which

[2.5-i] allowed to compare measure of the subsets of B where this function is less than ε and ρ respectively, and

[2.5-ii] attained value at least ρ on B.

Let us say that a point $x \in B$ is (ε/ρ)-*protected* if $x \in \overline{B_{\mathbf{v}}(\rho)} \smallsetminus B_{\mathbf{v}}(\varepsilon)$ for some $\mathbf{v} \in P(\Lambda)$. [2.5-i] and [2.5-ii] imply that for each $\mathbf{v}$, the relative measure of protected points inside $B_v(\rho)$ is big. Then Lemma 2.1 shows that protected points are safe (no other vector can cause trouble), i.e. brings us to the realm of Case 1 when restricted to $B_{\mathbf{v}}(\rho)$.

In the analog of the argument for $n > 2$, properties [2.5-i] and [2.5-ii] of certain functions will play an important role. However it will be more difficult to protect points from small vectors, and the final step, that is, an application of Lemma 2.1, will be replaced by an inductive procedure, described in the next section.

3. Quantitative non-divergence in $\mathcal{L}_n$.

3.1. **The main concepts needed for the proof.** The crucial idea that serves as a substitute for the absence of Lemma 2.1 in dimensions 3 and up is an observation that whenever a lattice Λ in $\mathbb{R}^n$ contains two linearly independent short vectors, one can consider a subgroup of rank two generated by them, and this subgroup will be "small", which should eventually contribute to preventing other small vectors from showing up. (Here and hereafter by the *rank* $\operatorname{rk}(\Delta)$ of a discrete subgroup Δ of $\mathbb{R}^n$ we mean its rank as a free $\mathbb{Z}$-module, or, equivalently, the dimension of the real vector space spanned by its elements.) Thus we are led to consider all subgroups of Λ, not just of rank one. In fact, similarly to the $n = 2$ case, it suffices to work with *primitive* subgroups. Namely, a subgroup Δ of Λ is called *primitive* in Λ if $\Delta = \mathbb{R}\Delta \cap \Lambda$; equivalently, if Δ admits a generating set which can be completed to a generating set of Λ. The inclusion relation makes the set $P(\Lambda)$ of all nonzero primitive subgroups of Λ a partially ordered set of length equal to $\operatorname{rk}(\Lambda)$ (any two primitive subgroups properly included in one another must have different ranks). This partial order turns out to be instrumental in creating a substitute for Lemma 2.1.

We also need a way to measure the size of a discrete subgroup Δ of $\mathbb{R}^n$. The best solution seems to be to use Euclidean norm $\|\cdot\|$ and extend it by letting $\|\Delta\|$ to be the volume of the quotient space $\mathbb{R}\Delta/\Delta$. This is clearly consistent with the one-dimensional picture, since $\|\mathbb{Z}\mathbf{v}\| = \|\mathbf{v}\|$. This is also consistent with the induced Euclidean structure on the exterior algebra of $\mathbb{R}^n$: if Δ is generated by $\mathbf{v}_1, \dots, \mathbf{v}_k$, then $\|\Delta\| = \|\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k\|$.

Our goal is to understand the trajectories $u_x\Lambda$ as in Theorem 1.2. However, observe that the group structure of U was not used at all in the proof in the previous section. Thus we are going to consider "trajectories" of a more general type. Namely, we will work with continuous functions h from an interval $B \subset \mathbb{R}$ into $\operatorname{SL}(n, \mathbb{R})$, and replace the map $x \mapsto u_x\Lambda$ with $x \mapsto h(x)\mathbb{Z}^n$ (then in the case of Theorem 1.2 we are going to have $h(x) = u_x g$ where $\Lambda = g\mathbb{Z}^n$).

Among the assumptions to be imposed on h, the central role is played by an analogue of [2.5-i]. This is taken care of by introducing a certain class of functions and then demanding that all functions of the form $x \mapsto \|h(x)\Delta\|$ where $\Delta \in P(\mathbb{Z}^n)$, belong to this class.

If C and α are positive numbers and B a subset of $\mathbb{R}$, let us say that a function $f : B \mapsto \mathbb{R}$ is (C, α)-*good on* B if for any open interval $J \subset B$ and any $\varepsilon > 0$ one has

$$\lambda\big(\{x \in J : |f(x)| < \varepsilon\}\big) \le C \left(\frac{\varepsilon}{\sup_{x\in J}|f(x)|}\right)^{\alpha} \lambda(J)\,. \tag{3.1}$$

Informally speaking, graphs of good functions are not allowed to spend a big proportion of "time" near the x-axis and then suddenly jump up. Several elementary facts about (C, α)-good functions are listed below:

Lemma 3.1. (a) *f is (C, α)-good on B $\Leftrightarrow$ so is $|f|$ $\Rightarrow$ so is cf $\forall\, c \in \mathbb{R}$;*
(b) *f_i, $i = 1, \dots, k$, are (C, α)-good on B $\Rightarrow$ so is $\sup_i |f_i|$;*
(c) *If f is (C, α)-good on B and $c_1 \le \left|\frac{f(x)}{g(x)}\right| \le c_2$ for all $x \in B$, then g is $\big(C(c_2/c_1)^{\alpha}, \alpha\big)$-good on B;*

The proofs are left as exercises. Another exercise is to construct a C^∞ function which is not good on (a) some interval (b) any interval.

The notion of (C, α)-good functions was introduced in [KM1] in 1998, but the importance of (3.1) for measure estimates on the space of lattices was observed earlier. For instance, the next proposition, which describes what can be called a model example of good functions, can be traced to [DM2, Lemma 4.1]. We will prove a slightly stronger version paying more atention to the constant C (which will not really matter for the main results).

Proposition 3.2. *For any $k \in \mathbb{N}$, any polynomial of degree not greater than k is $\big(k(k+1)^{1/k}, 1/k\big)$-good on $\mathbb{R}$.*

Proof. Fix an open interval $J \subset \mathbb{R}$, a polynomial f of degree not exceeding k, and a positive ε. We need to show that

$$\lambda\big(\{x \in J : |f(x)| < \varepsilon\}\big) \le k(k+1)^{1/k} \left(\frac{\varepsilon}{\sup_{x\in J}|f(x)|}\right)^{1/k} \lambda(J)\,. \tag{3.2}$$

Suppose that the left hand side of (3.2) is strictly bigger than some number m. Then it is possible to choose $x_1, \dots, x_{k+1} \in \{x \in J : |f(x)| < \varepsilon\}$ with $|x_i - x_j| \ge m/k$ for each $1 \le i \ne j \le k+1$. (Exercise.) Using Lagrange's interpolation formula one can write down the exact expression for f:

$$f(x) = \sum_{i=1}^{k+1} f(x_i) \frac{\prod_{j=1,\, j\ne i}^{k+1}(x - x_j)}{\prod_{j=1,\, j\ne i}^{k+1}(x_i - x_j)}\,. \tag{3.3}$$

Note that $|f(x_i)| < \varepsilon$ for each i, $|x - x_j| < \lambda(J)$ for each j and $x \in J$, and also $|x_i - x_j| \geq m/k$. Therefore

$$\sup_{x\in J} |f(x)| < (k+1)\varepsilon \frac{\lambda(J)^k}{(m/k)^k}\,.$$

which can be rewritten as

$$m < k(k+1)^{1/k} \left(\frac{\varepsilon}{\sup_{x\in J} |f(x)|}\right)^{1/k} \lambda(J)\,,$$

proving (3.2). □

Observe that in the course of the proof of Theorem 2.4 it was basically shown that linear functions are $(2,1)$-good on $\mathbb{R}$. The relevance of the above proposition for the nondivergence of unipotent flows on $\mathcal{L}_n$ is highlighted by

Corollary 3.3. *For any $n \in \mathbb{N}$ there exist (explicitly computable) $C = C(n)$, $\alpha = \alpha(n)$ such that for any one-parameter unipotent subgroup $\{u_x\}$ of $\mathrm{SL}(n,\mathbb{R})$, any $\Lambda \in \mathcal{L}_n$ and any subgroup Δ of Λ, the function $x \mapsto \|u_x\Delta\|$ is (C,α)-good.*

Proof. Represent Δ by a vector $w \in \bigwedge^k(\mathbb{R}^n)$ where k is the rank of Δ; the action of u_x on $\bigwedge^k(\mathbb{R}^n)$ is also unipotent, therefore every component of $u_x w$ (with respect to some basis) is a polynomial in x of degree uniformly bounded in terms of n. Thus the claim follows from Proposition 3.2, Lemma 3.1(b) for the supremum norm, and then Lemma 3.1(c) for the Euclidean norm. □

3.2. **The main nondivergence result and its history.** Let us now state a generalization of Theorem 2.4 to the case of arbitrary n.

Theorem 3.4. *Suppose an interval $B \subset \mathbb{R}$, $C, \alpha > 0$, $0 < \rho < 1$ and a continuous map $h : B \to \mathrm{SL}(n,\mathbb{R})$ are given. Assume that for any $\Delta \in P(\mathbb{Z}^n)$,*
[3.4-i] *the function $x \mapsto \|h(x)\Delta\|$ is (C,α)-good on B, and*
[3.4-ii] $\sup_{x\in B} \|h(x)\Delta\| \geq \rho^{\mathrm{rk}(\Delta)}$.
Then for any $\varepsilon < \rho$,

$$\lambda(\{x \in B : h(x)\mathbb{Z}^n \notin \mathcal{L}_n(\varepsilon)\}) \leq n2^n C \left(\frac{\varepsilon}{\rho}\right)^{\alpha} \lambda(B)\,. \tag{3.4}$$

This is a simplified version of a theorem from [Kl5], which sharpens the one proved in [KM1]. The latter had a slightly stronger assumptions, with ρ in place of $\rho^{\mathrm{rk}(\Delta)}$ in [3.4-ii]. In most of the applications this improvement is not needed – but there are some situations in metric Diophantine approximation, described later in the notes, where it becomes important. Anyway, the scheme of the proof, see §3.3, is the same for both original and new versions, and also there are some reasons why the sharpening appears to be more natural, as will be seen below. See [KLW] for another exposition of the proof.

It is straightforward to verify that Theorem 1.2 follows from Theorem 3.4: take $B = [0,T]$ and $h(x) = u_x g$ where $\Lambda = g\mathbb{Z}^n$. Condition [3.4-i] has already been established in Corollary 3.3, and [3.4-ii] clearly holds with some ρ dependent of Λ: just put $x = 0$ and

$$\rho = \rho(\Lambda) = \inf_{\Delta\in P(\Lambda)} \|\Delta\|^{1/\operatorname{rk}(\Delta)}\,, \tag{3.5}$$

positive since Λ is discrete. Furthermore, Theorem 3.4 implies the following

Corollary 3.5. *For any $\Lambda \in \mathcal{L}_n$ and any positive δ there exists a compact subset K of $\mathcal{L}_n$ such that for any unipotent one-parameter $\{u_x\} \subset \mathrm{SL}(n,\mathbb{R})$ and any positive T one has*

$$\frac{1}{T}\lambda(\{0 \le x \le T : u_x\Lambda \notin K\}) \le \delta\,. \tag{3.6}$$

This was proved by Dani in 1979 [D1]. For the proof using Theorem 3.4, just take $K = \mathcal{L}_n(\varepsilon)$ where ε is such that

$$n2^n C(n)\,(\varepsilon/\rho)^{\alpha(n)} < \delta\,, \tag{3.7}$$

$C(n), \alpha(n)$ are as in Corollary 3.3 and $\rho(\Lambda)$ as defined in (3.5). Thus, on top of Dani's result, one can recover an expression for the "size" of K in terms of δ.

But this is not the end of the story – one can conclude much more. It immediately follows from Minkowski's Lemma that if $\operatorname{rk}(\Delta)$ is, say, k, then the intersection of Δ with any compact convex subset of $\mathbb{R}\Delta$ of volume $2^k\|\Delta\|$ contains a nonzero vector. Thus such a Δ must contain a nonzero vector of length $\le 2\|\Delta\|/\nu_k^{1/k}$, where ν_k is the volume of the unit ball in $\mathbb{R}^k$. Consequently, if we know that $\Lambda \in \mathcal{L}_n(\rho')$ for some positive ρ', then $\rho(\Lambda)$ as defined in (3.5) is at least $c'\rho'$ where $c' = c'(n)$ depends only on n. Thus we have derived (modulo elementary computations left as an exercise) the following statement:

Corollary 3.6. *For any $\delta > 0$ there exists (explicitly computable) $c = c(n,\delta)$ such that whenever $\{u_x\Lambda : 0 \le x \le T\} \subset \mathcal{L}_n$ is a unipotent trajectory nontrivially intersecting $\mathcal{L}_n(\rho)$ for some $\rho > 0$, (3.6) holds with $K = \mathcal{L}_n(c\rho)$.*

In order to appreciate a geometric meaning of the above corollary and other related results, it will be convenient to choose a right-invariant Riemannian metric on $\mathrm{SL}_n(\mathbb{R})$ and use it to induce a Riemannian metric on $\mathcal{L}_n$. Then it is not hard to see that the distance between $\mathcal{L}_n(\rho)$ and the complement of $\mathcal{L}_n(c\rho)$ is uniformly bounded from above by a constant depending only on c, not on ρ. Thus Corollary 3.6 guarantees that, regardless of the size of the compact set where a unipotent trajectory begins, one only needs to increase the set by a bounded distance to make sure that the trajectory spends, say, at least half the time in the bigger set. Note that for the last conclusion it is important to have $\rho^{\operatorname{rk}(\Delta)}$ and not ρ in the right hand side of [3.4-ii]; previously available non-divergence estimates forced a much more significant expansion of $\mathcal{L}_n(\rho)$.

Let us now turn our attention to another non-divergence theorem, proved by Dani in 1986 [D4], and later generalized by Eskin, Mozes and Shah [EMS]:

Corollary 3.7. *For any $\delta > 0$ there exists a compact subset $K \subset \mathcal{L}_n$ such that for any unipotent one-parameter subgroup $\{u_x\} \subset \mathrm{SL}(n,\mathbb{R})$ and any $\Lambda = g\mathbb{Z}^n \in \mathcal{L}_n$, either* (3.6) *holds for all large T, or there exists a $(g^{-1}u_xg)$-invariant proper subspace of $\mathbb{R}^n$ defined over $\mathbb{Q}$.*

Proof. Apply Theorem 3.4 with an arbitrary $\rho < 1$ and ε as in (3.7), as before choosing K to be equal to $\mathcal{L}_n(\varepsilon)$. Assume that the first alternative in the statement of the corollary is not satisfied for some $\{u_x\}$, Λ and this K. This means that there exists an unbounded sequence T_k such that for each k, the conclusion of Theorem 3.4 with $\rho = 1$, ε chosen as above and $h(x) = u_xg$, does not hold for $B = [0, T_k]$. Since assumption [3.4-i] is always true, [3.4-ii] must go wrong, i.e. for each k there must exist $\Delta_k \in P(\mathbb{Z}^n)$ such that $\|u_xg\Delta_k\| < 1$ for all $0 \le x \le T_k$. However, by the discreteness of $\bigwedge(g\mathbb{Z}^n)$ in $\bigwedge(\mathbb{R}^n)$, there are only finitely many choices for such subgroups; hence one of them, Δ, works for infinitely many k. But $\|u_xg\Delta\|^2$ is a polynomial, therefore it must be constant, which implies that u_x fixes $g(\mathbb{R}\Delta) \Leftrightarrow g^{-1}u_xg$ fixes the proper rational subspace $\mathbb{R}\Delta$. □

3.3. **The proof.** In order to prove Theorem 3.4, we are going to create a substitute for the procedure of marking points by vectors (and thereby declaring them safe from any other small vectors) used in the proof of Theorem 2.4. However now vectors will not be sufficient for our purposes, we will need to replace it with *flags*, that is, linearly ordered subsets of the partially ordered set (poset) $P(\Lambda)$, $\Lambda \in \mathcal{L}_n$. Furthermore, to set up the induction we will need to prove a version of the theorem with $P(\mathbb{Z}^n)$ repalced by its subsets (more precisely, sub-posets) P. The induction will be on the *length* of P, i.e. the number of elements in its maximal flag. In this more general theorem we will also get rid of the expressions $\rho^{\mathrm{rk}(\Delta)}$ in the right hand side of [3.4-ii], replacing them with $\eta(\Delta)$, where η is an arbitrary function $P \to (0,1]$ (to be called the *weight function*).

Now let us fix an interval $B \subset \mathbb{R}$, a sub-poset $P \subset P(\mathbb{Z}^n)$, a weight function η and a map $h : B \to \mathrm{SL}(n,\mathbb{R})$. Then say that, given $\varepsilon > 0$, a point $x \in B$ is *ε-protected relative to P* if there exists a flag $F \subset P$ with the following properties:

(M1) $\varepsilon\eta(\Delta) \le \|h(x)\Delta\| \le \eta(\Delta) \quad \forall\, \Delta \in F$;

(M2) $\|h(x)\Delta\| \ge \eta(\Delta) \quad \forall\, \Delta \in P \smallsetminus F$ comparable with every element of F.

We are going to show that with the choice $\eta(\Delta) = \rho^{\mathrm{rk}(\Delta)}$ and $P = P(\mathbb{Z}^n)$, any (ε/ρ)-protected point $x \in B$ is indeed protected from vectors in $h(x)\mathbb{Z}^n$ of length less than ε. But first let us check that the above definition reduces to the one used for the proof of Theorem 2.4 when $P = P(\mathbb{Z}^2)$. Indeed, for $h(x) = u_xg$, $\Delta = \mathbb{Z}\mathbf{v}$ of rank 1, $\eta(\Delta) = \rho$ and ε substituted with ε/ρ, (M1) reduces to $\varepsilon \le \|u_xg\mathbf{v}\| \le \rho$, which was exactly the condition satisfied by some vector $\mathbf{v} \in \mathbb{Z}^2$ for $x \in \overline{B_{g\mathbf{v}}(\rho)} \smallsetminus B_{g\mathbf{v}}(\varepsilon)$. Further, (M2) in that case holds trivially, since the only element of $P(\mathbb{Z}^2) \smallsetminus \{\Delta\}$

comparable with Δ is $\mathbb{Z}^2$ itself, and $\|g\mathbb{Z}^2\| = 1 > \rho^2$. And the conclusion was that the existence of such $\mathbf{v}$ forces $u_x g\mathbb{Z}^2$ to belong to $\mathcal{L}_2(\varepsilon)$.

Here is a generalization:

Proposition 3.8. *Let η be given by $\eta(\Delta) = \rho^{\mathrm{rk}(\Delta)}$ for some $0 < \rho < 1$. Then for any $\varepsilon < \rho$ and any $x \in B$ which is (ε/ρ)-protected relative to $P(\mathbb{Z}^n)$, one has $h(x)\mathbb{Z}^n \in \mathcal{L}_n(\varepsilon)$.*

Proof. For x as above, let $\{0\} = \Delta_0 \subsetneq \Delta_1 \subsetneq \cdots \subsetneq \Delta_\ell = \mathbb{Z}^n$ be all the elements of $F \cup \big\{\{0\}, \mathbb{Z}^n\big\}$. Properties (M1) and (M2) translate into:

(M1) $\frac{\varepsilon}{\rho} \cdot \rho^{\mathrm{rk}(\Delta_i)} \le \|h(x)\Delta_i\| \le \rho^{\mathrm{rk}(\Delta_i)} \quad \forall\, i = 0, \dots, \ell - 1$;

(M2) $\|h(x)\Delta\| \ge \rho^{\mathrm{rk}(\Delta)} \quad \forall\, \Delta \in P(\mathbb{Z}^n) \smallsetminus F$ comparable with every Δ_i.

(Even though $\Delta_0 = \{0\}$ is not in $P(\mathbb{Z}^n)$, it would also satisfy (M1) with the convention $\|\{0\}\| = 1$.)

Take any $\mathbf{v} \in \mathbb{Z}^n \smallsetminus \{0\}$. Then there exists j, $1 \le j \le \ell$, such that $\mathbf{v} \in \Delta_j \smallsetminus \Delta_{j-1}$. Denote $\mathbb{R}(\Delta_{j-1} + \mathbb{Z}\mathbf{v}) \cap \Lambda$ by Δ. Clearly it is a primitive subgroup of Λ satisfying $\Delta_{j-1} \subset \Delta \subset \Delta_j$, therefore Δ is comparable with Δ_i for every i (and may or may not coincide with one of the Δ_is). Now one can use properties (M1) and (M2) to deduce that

$$\|h(x)\Delta\| \ge \min\left(\frac{\varepsilon}{\rho} \cdot \rho^{\mathrm{rk}(\Delta)},\ \rho^{\mathrm{rk}(\Delta)}\right) = \varepsilon\rho^{\mathrm{rk}(\Delta)-1} = \varepsilon\rho^{\mathrm{rk}(\Delta_{i-1})}\,. \tag{3.8}$$

On the other hand, from the submultiplicativity of the covolume it follows that $\|h(x)\Delta\|$ is not greater than $\|h(x)\Delta_{i-1}\| \cdot \|\mathbf{v}\|$ (recall a similar step in the proof of Lemma 2.1). Thus

$$\|h(x)\mathbf{v}\| \ge \frac{\|h(x)\Delta\|}{\|h(x)\Delta_{i-1}\|} \underset{\text{by (M1) and (3.8)}}{\ge} \frac{\varepsilon\rho^{\mathrm{rk}(\Delta_{i-1})}}{\rho^{\mathrm{rk}(\Delta_{i-1})}} = \varepsilon\,.$$

Hence $\Lambda \in \mathcal{L}_n(\varepsilon)$ and the proof is finished. □

This is perhaps the crucial point in the proof: we showed that a flag with certain properties does exactly what a single vector was doing in the case of $\mathrm{SL}(2, \mathbb{R})$; namely, it guarantees that in the lattices corresponding to protected points, no vector can be shorter than ε.

Now that the above proposition is established, we will forget about the specific form of the weight function and work with an arbitrary η. Here is a more general theorem:

Theorem 3.9. *Fix $0 \le k \le n$, and suppose an interval $B \subset \mathbb{R}$, $C, \alpha > 0$, a continuous map $h : B \to \mathrm{SL}(n, \mathbb{R})$, a poset $P \subset P(\mathbb{Z}^n)$ of length k and a weight function $\eta : P \to (0, 1]$ are given. Assume that for any $\Delta \in P$*

[3.9-i] *the function $x \mapsto \|h(x)\Delta\|$ is (C, α)-good on B, and*

[3.9-ii] $\sup_{x \in B} \|h(x)\Delta\| \ge \eta(\Delta)$.

Then for any $0 < \varepsilon < 1$,

$$\lambda(\{x \in B : x \textit{ is not } \varepsilon\textit{-protected relative to } P\}) \le k2^k C\varepsilon^\alpha \lambda(B)\,.$$

We remark that the use of an arbitrary P in place of $P(\mathbb{Z}^n)$ is justified not only by a possibility to prove the theorem by induction, but also by some applications to Diophantine approximation, see e.g. [BKM, Kl3, G1], where proper sub-posets of $P(\mathbb{Z}^n)$ arise naturally.

Proof. We will break the argument into several steps.

Step 0. First let us see what happens when $k = 0$, the base case of the induction. In this case P is empty, and the flag $F = \varnothing$ will satisfy both (M1) and (M2). Thus all points of B are ε-protected relative to P for any ε, which means that in the case $k = 0$ the claim is trivial. So we can take $k \ge 1$ and suppose that the theorem is proved for all the smaller lengths of P.

Step 1. For any $y \in B$ let us define

$$S(y) \stackrel{\text{def}}{=} \{\Delta \in P : \|h(y)\Delta\| < \eta(\Delta)\}\,.$$

Roughly speaking, $S(y)$ is the set of Δs which gets small enough at y, i.e. potentially could bring trouble. By the discreteness of $h(y)\mathbb{Z}^n$ in $\mathbb{R}^n$, this is a finite subset of P. Note that if this set happens to be empty, then $\|h(y)\Delta\| \ge \eta(\Delta)$ for all $\Delta \in P$, which means that $F = \varnothing$ can be used to ε-protect y for any ε. So let us define

$$E \stackrel{\text{def}}{=} \{y \in B : S(y) \ne \varnothing\} = \{y \in B : \exists\, \Delta \in P \text{ with } \|h(y)\Delta\| < \eta(\Delta)\}\,;$$

then to prove the theorem it suffices to estimate the measure of the set of points $x \in E$ which are not ε-protected relative to P. A flashback to the proof for $n = 2$: there $S(y)$ consisted of primitive vectors $\mathbf{v}$ for which $\|u_y\mathbf{v}\|$ was less than ρ, not more than one such vector was allowed, and nonexistence of such vectors automatically placed the lattice in $\mathcal{L}_n(\varepsilon)$.

Step 2. Take $y \in E$ and $\Delta \in S(y)$, and define $B_{\Delta,y}$ to be the maximal interval of the form $B \cap (y - r, y + r)$ on which the absolute value of $\|h(\cdot)\Delta\|$ is not greater than $\eta(\Delta)$. From the definition of $S(y)$ and the continuity of functions $\|h(\cdot)\Delta\|$ it follows that $B_{\Delta,y}$ contains some neighborhood of y. Further, the maximality property of $B_{\Delta,y}$ implies that

$$\sup_{x \in B_{\Delta,y}} \|h(x)\Delta\| = \eta(\Delta)\,. \tag{3.9}$$

Indeed, either $B_{s,y} = B$, in which case the claim follows from [3.9-ii], or at one of the endpoints of $B_{\Delta,y}$, the function $\|h(\cdot)\Delta\|$ must attain the value $\eta(\Delta)$ – otherwise one can enlarge the interval and still have $\|h(\cdot)\Delta\|$ not greater than $\eta(\Delta)$ for all its points. (Another flashback: intervals $B_{\Delta,y}$ are analogues of $B_{\mathbf{v}}(\rho)$ from the proof of Theorem 2.4 – but this time there is no disjointness, since many Δs can get small simultaneously.)

Step 3. For any $y \in E$ let us choose an element Δ_y of $S(y)$ such that $B_{\Delta_y,y} = \bigcup_{\Delta\in S(y)} B_{\Delta,y}$ (this can be done since $S(y)$ is finite). In other words, $B_{\Delta_y,y}$ is maximal among all $B_{\Delta,y}$. For brevity we will denote $B_{\Delta_y,y}$ by B_y. We now claim that

$$\sup_{x\in B_y} \|h(x)\Delta\| \geq \eta(\Delta) \text{ for any } y \in E \text{ and } \Delta \in P\,. \tag{3.10}$$

Indeed, if not, then $\|h(x)\Delta\| < \eta(\Delta)$ for all $x \in B_y$, in particular one necessarily has $\|h(y)\Delta\| < \eta(\Delta)$, hence $\Delta \in S(y)$ and $B_{\Delta,y}$ is defined. But $B_{\Delta,y}$ is contained in B_y, so (3.10) follows from (3.9). This step allows one to replace the covering $\{B_{\Delta,y} : \Delta \in S(y),\ y \in E\}$ of E by a more efficient covering $\{B_y :\ y \in E\}$; informally speaking, this is achieved by selecting $\Delta = \Delta_y$ which works best for every given y.

Step 4. Now we are ready to perform the induction step. For any $y \in E$ define

$$P_y \stackrel{\text{def}}{=} \{\Delta \in P \smallsetminus \{\Delta_y\} : \Delta \text{ is comparable with } \Delta_y\}\,.$$

We claim that P_y (a poset of length $k-1$) in place of P and B_y in place of B satisfy all the conditions of the theorem. Indeed, [3.9-i] is clear since B_y is a subset of B, and [3.9-ii] follows from (3.10). Therefore, by induction,

$$\lambda(\{x \in B_y : x \text{ is not } \varepsilon\text{-protected relative to } P_y\}) \leq (k-1)2^{k-1}C\varepsilon^\alpha\lambda(B_y)\,. \tag{3.11}$$

Step 5. Does the previous step help us, and how? let us take x outside of this set of relatively small measure, that is, assume that x is ε-protected relative to P_y, and try to use this protection. By definition, there exists a flag F' inside P_y such that

$$\varepsilon\eta(\Delta) \leq \|h(x)\Delta\| \leq \eta(\Delta) \quad \forall\,\Delta \in\ F' \tag{3.12}$$

and

$$\|h(x)\Delta\| \geq \eta(\Delta) \quad \forall\,\Delta \in P_y \smallsetminus F' \text{ comparable with every element of } F'\,. \tag{3.13}$$

However this F' will NOT protect x relative to the bigger poset P, because Δ_y, comparable with every element of F', would not satisfy (M2) – on the contrary, recall that it was chosen so that the reverse inequality, $\|h(x)\Delta_y\| \leq \eta(\Delta_y)$, holds for all $x \in B_y$, see (3.10)! Thus our only choice seems to be to add Δ_y to F', for extra protection, and put $F \stackrel{\text{def}}{=} F' \cup \{\Delta_y\}$. Then $\Delta \in P \smallsetminus F$ is comparable with every element of F if and only if Δ is in $P_y \smallsetminus F'$, and is comparable with every element of F'. Because of that, (M2) immediately follows from (3.13). As for (M1), we already know it for for $\Delta \neq \Delta_y$ by (3.12), so it remains to put $\Delta = \Delta_y$. The upper estimate in (M1) is immediate from (3.10). The lower estimate, on the other hand, can fail – but only on a set of relatively small measure, because of assumption [3.9-i] which, by the way, has not been used so far at all:

$$\begin{aligned}\lambda(\{x \in B_y : \|h(x)\Delta_y\| < \varepsilon\eta(\Delta_y)\}) &\le C\left(\frac{\varepsilon\eta(\Delta_y)}{\sup_{x\in B_y}\|h(x)\Delta_y\|}\right)^\alpha \lambda(B_y)\\ &\underset{(3.9)}{\le} C(\varepsilon)^\alpha\lambda(B_y)\,.\end{aligned} \tag{3.14}$$

The union of the two sets above, in the left hand sides of (3.11) and (3.3), has measure at most $k2^{k-1}C\varepsilon^\alpha\lambda(B_y)$. We have just shown that this union exhausts all the unprotected points as long as we are restricted to B_y. Thus we have achieved an analogue of what was extremely easy for $n = 2$: bounded the measure of the set of points where things can go wrong on each of the intervals $B_{\mathbf{v}}(\rho)$.

Step 6. It remains to produce a substitute for the disjointness of the intervals, that is, put together all the B_ys. For that, consider the covering $\{B_y : y \in E\}$ of E and choose a subcovering $\{B_i\}$ of multiplicity at most 2. (Exercise: this is always possible.) Then the measure of $\{x \in E : x \text{ is not } \varepsilon\text{-protected relative to } P\}$ is not greater than

$$\begin{aligned}\sum_i \lambda(\{x \in B_i : x \text{ is not } \varepsilon\text{-protected relative to } P\}) &\le k2^{k-1}C\varepsilon^\alpha\sum_i\lambda(B_i)\\ &\le k2^kCe^\alpha\lambda(B)\,,\end{aligned}$$

and the theorem is proven. □

4. Applications of non-divergence to metric Diophantine approximation

Here we present applications of Theorem 3.4 to number theory which reach beyond the unipotent, or even polynomial, case.

4.1. **Inheritance of sublinear growth.** Our main object of studying will be a fixed parametrized curve $B \to \mathcal{L}_n$, where $B \subset \mathbb{R}$ is an interval. Given such a curve, we will consider a family of curves which are translations of the initial one by some group elements a_t. That is, put

$$h(x) = h_t(x) = a_t h_0(x) \tag{4.1}$$

in Theorem 3.4, where h_0 is a fixed map from B to $\mathrm{SL}(n,\mathbb{R})$. We would like to investigate the following two questions:

(1) What are interesting examples of h_0 and a_t for which one can establish conditions [3.4-i] and [3.4-ii] uniformly for all $t > 0$?
(2) What would be consequences of that for the initial curve $h_0(x)\mathbb{Z}^n$?

Let us start with the second question, since it is easier. That is, suppose we are given an interval $B \subset \mathbb{R}$, $C, \alpha > 0$, $0 < \rho < 1$, a continuous map $h_0 : B \to \mathrm{SL}(n,\mathbb{R})$ and $h = h_t$ as in (4.1). Also let us assume that for any $\Delta \in P(\mathbb{Z}^n)$ and any $t > 0$, conditions [3.4-i] and [3.4-ii] are satisfied. The trick is now to choose $\varepsilon = e^{-\gamma t}$ for some positive γ. From Theorem 3.4 it follows that there exists a constant $\tilde{C}$ (depending on n, C, ρ, B) such that for any t,

$$\lambda\big(\{x \in B : a_t h_0(x)\mathbb{Z}^n \notin \mathcal{L}_n(e^{-\gamma t})\}\big) \le \tilde{C} e^{-\alpha\gamma t}\,.$$

The sum of the right hand sides of the above equation will converge if added up say for $t \in \mathbb{N}$. This immediately calls for an application of the following standard principle from elementary probability theory (the proof is left as an exercise):

Lemma 4.1 (Borel-Cantelli Lemma)**.** *If μ is a measure on a space X and $\{A_i\}$ is a countable collection of measurable subsets of X with $\sum_i \mu(A_i) < \infty$, then μ-almost every $x \in X$ is contained in at most finitely many sets A_i.*

The conclusion from this is: given an arbitrary $\gamma > 0$, for λ-almost every $x \in B$ we have $a_t h_0(x)\mathbb{Z}^n \in \mathcal{L}_n(e^{-\gamma t})$ if $t \in \mathbb{N}$ is sufficiently large. In fact, by changing γ just a little bit it is easily seen that $t \in \mathbb{N}$ in the last statement can be replaced by $t > 0$. (Exercise.) That is, for all $\gamma > 0$ we have

(4.2) $\{a_t\Lambda \;:\; t > 0\}$ eventually grows slower than the family $\mathcal{L}_n(e^{-\gamma t})$

for (Lebesgue) almost every Λ of the form $h_0(x)\mathbb{Z}^n$.

To put this conclusion in an appropriate context, we need to describe the family of sets $\mathcal{L}_n(\varepsilon)$ in a more detailed way. It is not hard to see, using reduction theory for $\mathrm{SL}(n,\mathbb{R})/\mathrm{SL}(n,\mathbb{Z})$, that minus logarithm of the biggest ε such that $\Lambda \in \mathcal{L}_n(\varepsilon)$ is (asymptotically for far away Λ) roughly the same as the distance[1] from Λ to $\mathbb{Z}^n$ or some other base point. Thus the validity of (4.2) for any $\gamma > 0$ can, and will, be referred to as the *sublinear growth* of $\{a_t\Lambda\}$. More generally, we will say, for fixed $\gamma_0 \ge 0$, that $\{a_t\Lambda\}$ has *growth rate* $\le \gamma_0$ if (4.2) holds for any $\gamma > \gamma_0$.

Now denote by ν the Haar probablilty measure on $\mathcal{L}_n$. One can show using Siegel's Formula (see [E] for more detail) that $\nu\big(\mathcal{L}_n \smallsetminus \mathcal{L}_n(\varepsilon)\big) \le \mathrm{const}_n\, \varepsilon^n$. (Exercise: compute this constant; a more difficult exercise: prove that the right hand side captures the asymptotics of $\nu\big(\mathcal{L}_n \smallsetminus \mathcal{L}_n(\varepsilon)\big)$ as $\varepsilon \to 0$; this is done in [KM2].) Since a_t preserves ν, for any $\gamma > 0$ and any t we have

$$\nu\big(\{\Lambda \in \mathcal{L}_n : a_t\Lambda \notin \mathcal{L}_n(e^{-\gamma t})\}\big) \le \mathrm{const}_n\, e^{-n\gamma t}\,;$$

therefore for the same (Borel-Cantelli) reason as above, for any positive γ (4.2) is satisfied by ν-a.e. $\Lambda \in \mathcal{L}_n$. Thus we have proved that, assuming all the functions of the form (4.1) satisfy [3.4-i] and [3.4-ii], certain dynamical behavior (sublinear growth of trajectories) of generic points of the phase space is inherited by generic points on the curve $\{h_0(x)\mathbb{Z}^n\}$.

We note that problems of this type, i.e. studying rates of growth of trajectories, or rates with which dense trajectories approximate points, are sometimes referred to

[1] See a remark after Corollary 3.6 for a description of a metric on $\mathcal{L}_n$; note also that $\mathrm{dist}(\Lambda, \mathbb{Z}^n)$ is also roughly the same as minus logarithm of $\rho(\Lambda)$ defined in (3.5)

as *shrinking target problems*. Indeed, the family of complements of the sets $\mathcal{L}_n(e^{-\gamma t})$ can be thought of as a shrinking target zooming at the cusp of $\mathcal{L}_n$, and to hit this target means to get into those "neighborhoods of infinity" infinitely many times. See [KM2] for a detailed discussion.

Also, observe that we haven't really used the full strength of Theorem 3.4, with $\rho^{\mathrm{rk}(\Delta)}$ in place of ρ, and it was promised that it is supposed to be important for applications. The next theorem summarizes the above discussion and strengthens its conclusions:

Theorem 4.2 ([Kl5])**.** *Suppose an interval $B \subset \mathbb{R}$, $C, \alpha, \gamma_0 > 0$, a continuous map $h_0 : B \to \mathrm{SL}(n, \mathbb{R})$ and a subgroup $\{a_t\} \subset \mathrm{SL}(n, \mathbb{R})$ are given.*

(a) *Assume that:*

[4.2-i] *for all $\Delta \in P(\mathbb{Z}^n)$ and $t > 0$, functions $x \mapsto \|a_t h_0(x)\Delta\|$ are (C, α)-good on B, and*

[4.2-ii] *for any $\beta > \gamma_0$ there exists T such that $\sup_{x\in B} \|a_t h_0(x)\Delta\| \geq (e^{-\beta t})^{\mathrm{rk}(\Delta)}$ for all $\Delta \in P(\mathbb{Z}^n)$ and $t > T$.*

Then for λ-a.e. $x \in B$, $\{a_t h_0(x)\mathbb{Z}^n\}$ has growth rate $\leq \gamma_0$.

(b) *Suppose that* [4.2-ii] *does not hold; then $\{a_t h_0(x)\mathbb{Z}^n\}$ has growth rate $> \gamma_0$ for all $x \in B$.*

Proof. Part (a) follows from a minor modification of the argument preceding the theorem: for any $\gamma > \gamma_0$ choose β between γ and γ_0, and apply Theorem 3.4 with $\rho = e^{-\beta t}$, and then the Borel-Cantelli Lemma. For part (b), if for some $\beta > \gamma_0$ there exist $t_k \to \infty$ and $\Delta_k \in P(\mathbb{Z}^n)$ such that $\|a_{t_k} h_0(x)\Delta_k\| < (e^{-\beta t_k})^{\mathrm{rk}(\Delta_k)}$ for all $x \in B$, then for each x, using Minkowski Lemma, one can choose a nonzero vector $v_k \in \Delta_k$ such that $\|a_{t_k} h_0(x)v_k\| < e^{-\beta t_k}$, which implies that $a_{t_k} h_0(x)\mathbb{Z}^n \notin \mathcal{L}_n(e^{-\beta t_k})$. □

We have therefore established a remarkable dichotomy: for curves satisfying [4.2-i], either almost all trajectories grow slowly, or all trajectories grow fast. See [Kl6] for a further exploration of this theme.

4.2. **Checking [3.4-i] and [3.4-ii].** Of course there would be no point in the argument of the previous section if we didn't know that there exist examples, and moreover very naturally arising in number theory, of functions h_t as in (4.1) satisfying the assumptions of Theorem 3.4 uniformly in t. We are going to describe a special case which is very useful for applications.

For this, it will be convenient to upgrade the dimension of the space where all the lattices live from n to $n + 1$. Then choose

$$a_t = \mathrm{diag}(e^{nt}, e^{-t}, \ldots, e^{-t}),$$

that is, consider a generalization of $\{a_t\} \subset \mathrm{SL}(2, \mathbb{R})$ as in (2.1). One can easily see that the unstable leaves of the action of a_t, $t > 0$, on $\mathcal{L}_{n+1}$ are given by the orbits of the group

$$\left\{ u_{\mathbf{y}} \stackrel{\text{def}}{=} \begin{pmatrix} 1 & \mathbf{y}^T \\ 0 & I_n \end{pmatrix} \;:\; \mathbf{y} \in \mathbb{R}^n \right\},$$

a higher-dimensional analogue of $U \subset \mathrm{SL}(2,\mathbb{R})$ (This group is denoted by $G^+_{a_1}$ in the notation of [EL] and is also known as the *expanding horospherical subgroup* corresponding to a_1). We are going to put our initial curve $\{h_0(x)\}$ inside this group; that is, consider

$$h_0(x) = \begin{pmatrix} 1 & \mathbf{f}(x)^T \\ 0 & I_n \end{pmatrix},$$

where $\mathbf{f}$ is a map $B \to \mathbb{R}^n$. The question now becomes: under what conditions on $\mathbf{f}$ can we verify the assumptions of Theorem 3.4 with $h_t(x) = a_t u_{\mathbf{f}(x)}$ uniformly in t.

In order to do that, we need to understand the action of the elements $u_{\mathbf{y}}$ on the exterior powers of $\mathbb{R}^{n+1}$. Choose the standard basis $\mathbf{e}_0, \mathbf{e}_1, \dots, \mathbf{e}_n$ of $\mathbb{R}^{n+1}$, and denote by V the space spanned by $\mathbf{e}_1, \dots, \mathbf{e}_n$. It will be convenient to identify $\mathbf{y} \in \mathbb{R}^n$ with $y_1\mathbf{e}_1 + \cdots + y_n\mathbf{e}_n$. Note that $\mathbf{e}_0$ is expanded by a_t (eigenvalue e^{nt}) and V is the contracting subspace (eigenvalue e^{-t}). Similarly for any $k \le n$, the k-th exterior power of $\mathbb{R}^{n+1}$ splits into the expanding (subspaces containing $\mathbf{e}_0$) and contracting (contained in V) parts.

Observe that $u_{\mathbf{y}}$ leaves $\mathbf{e}_0$ fixed and sends vectors $\mathbf{v} \in V$ to $\mathbf{v} + (\mathbf{y}\cdot\mathbf{v})\mathbf{e}_0$. From this is is easy to conclude how $u_{\mathbf{y}}$ acts on $\bigwedge^k(\mathbb{R}^{n+1})$: elements of the form $\mathbf{e}_0 \wedge \mathbf{w}$ are fixed, and

$$\begin{aligned} \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k &\xrightarrow{u_{\mathbf{y}}} \big(\mathbf{v}_1 + (\mathbf{y}\cdot\mathbf{v}_1)\mathbf{e}_0\big) \wedge \cdots \wedge \big(\mathbf{v}_k + (\mathbf{y}\cdot\mathbf{v}_k)\mathbf{e}_0\big) \\ &= \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k + \mathbf{e}_0 \wedge \left(\sum_{i=1}^{k} \pm(\mathbf{y}\cdot\mathbf{v}_i) \bigwedge_{j\ne i} \mathbf{v}_j \right). \end{aligned} \tag{4.3}$$

Now let us see what conditions on $\mathbf{f}$ are sufficient to establish [3.4-i] and [3.4-ii]. Take $\Delta \in P(\mathbb{Z}^{n+1})$ and represent it (up to $\pm$) by the exterior product of generators of Δ, let us call it $\mathbf{w}$. First of all it follows from the above formula that for any $\mathbf{w} \in \bigwedge(\mathbb{R}^{n+1})$, all the coordinates of $u_{\mathbf{y}}\mathbf{w}$, and hence of $a_t u_{\mathbf{y}}\mathbf{w}$ for any t, are linear combinations of $1, y_1, \dots, y_n$ (coefficients in these linear combinations depend on t). Thus property [3.4-i] uniformly over all t would follow if we could find C, α such that all the linear combinations of $1, f_1, \dots, f_n$ are (C,α)-good on B.

It turns out that condition [3.4-ii], that is, $\sup_{x\in B} \|a_t h_0(x)\Delta\| \ge \rho^{\mathrm{rk}(\Delta)}$ for all Δ and all (large enough) t is also easy to check:

Lemma 4.3. *Suppose that $\mathbf{f}(B)$ is not contained in any affine hyperplane (equivalently, the restrictions of $1, f_1, \dots, f_n$ to B are linearly independent over $\mathbb{R}$). Then:*

(a) *there exists $\rho > 0$ such that [3.4-ii] holds for any $\Delta \in P(\mathbb{Z}^{n+1})$ and any $t > 0$;*
(b) *$\exists\, t_0 > 0$ such that [3.4-ii] holds with $\rho = 1$ for all $\Delta \in P(\mathbb{Z}^{n+1})$ and $t > t_0$.*

Proof. Both claims are trivial if $\mathbb{R}\Delta$ contains $\mathbf{e}_0$: indeed, from the previous discussion it follows that $\mathbf{w}$ representing Δ is then fixed by $u_{\mathbf{f}(x)}$ and expanded by a_t. If not, suppose that $\operatorname{rk}(\Delta) = k$; then $\dim(\mathbb{R}\Delta \oplus \mathbb{R}\mathbf{e}_0) = k+1$. One can choose an orthonormal set $\{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}\} \subset \mathbb{R}\Delta \cap V$ and complete it to an orthonormal basis $\{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{e}_0\}$ of $\mathbb{R}\Delta \oplus \mathbb{R}\mathbf{e}_0$. Then

$$\mathbf{w} = a\mathbf{e}_0 \wedge \mathbf{v}_1 \cdots \wedge \mathbf{v}_{k-1} + b\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k \,,$$

where $a^2 + b^2 \geq 1$. (Note: a, b do not have to be integers, and vectors $\mathbf{v}_i$ are not necessarily with integer coordinates; however orthonormality is important.) Now we can, using (4.3), simply look at the projection of $u_{\mathbf{f}(x)}\mathbf{w}$ onto $\mathbf{e}_0 \wedge \mathbf{v}_1 \cdots \wedge \mathbf{v}_{k-1}$:

$$u_{\mathbf{f}(x)}\mathbf{w} = \big(a + b(\mathbf{f}(x) \cdot \mathbf{v}_k)\big)\mathbf{e}_0 \wedge \mathbf{v}_1 \cdots \wedge \mathbf{v}_{k-1} + \dots \,.$$

Regardless of the choice of a, b, the coefficient in front of $\mathbf{e}_0 \wedge \mathbf{v}_1 \cdots \wedge \mathbf{v}_{k-1}$ is of the form $c_0 + c_1 f_1 + \cdots + c_n f_n$ with $\sum |c_i|^2 \geq 1$. In view of the linear independence assumption and the compactness of the unit sphere in $\mathbb{R}^{n+1}$, there exists $\rho = \rho(B) > 0$ such that the supremum of the absolute value of every such function, and hence $\sup_{x \in B} \|u_{\mathbf{f}(x)}\Delta\|$ is at least ρ. But $\mathbf{e}_0 \wedge \mathbf{v}_1 \cdots \wedge \mathbf{v}_{k-1}$ is expanded by a_t with a rate at least e^t, and both conclusions follow. □

Now, abusing terminology for some more, let us introduce the following definitions. Say that a map $\mathbf{f}$ from a subset U of $\mathbb{R}$ to $\mathbb{R}^n$ is *good* if for λ-a.e. $x \in U$ there exists a neighborhood $B \subset U$ of x and $C, \alpha > 0$ such that any linear combination of $1, f_1, \dots, f_n$ is (C, α)-good on B. We will also say that $\mathbf{f}$ is (C, α)-good if C and α can be chosen uniformly for all x as above. Polynomial maps form a basic example. Later we will explain how one can prove that real analytic maps also have this property.

Also, say that $\mathbf{f}$ is *nonplanar* if for any nonempty interval $B \subset U$, the restrictions of $1, f_1, \dots, f_n$ to B are linearly independent over $\mathbb{R}$; in other words, no nonempty relatively open piece of $\mathbf{f}(U)$ is contained in a proper affine subspace of $\mathbb{R}^n$. The above discussion can be thus summarized in the following way:

Theorem 4.4. *Let U be a subset of $\mathbb{R}$ and let $\mathbf{f} : U \to \mathbb{R}^n$ be a continuous good nonplanar map. Then for λ-a.e. $x \in U$, the a_t-trajectory of $u_{\mathbf{f}(x)}\mathbb{Z}^{n+1}$ has sublinear growth.*

Note: it follows from remarks made at the end of the previous section and a "flowbox" argument (see [E]) that the a_t-trajectory of $u_{\mathbf{y}}\mathbb{Z}^n$ has sublinear growth for λ-a.e. $\mathbf{y} \in \mathbb{R}^n$. Thus the above theorem describes examples of curves in $\mathbb{R}^n$ whose generic points inherit certain property of generic points of $\mathbb{R}^n$.

4.3. **Inheritance of Diophantine properties.** Of course a reasonable question concerning all the argument above would be – why would anybody at all care about orbit growth properties of typical poitns on some curves. The answer is – that all along, like monsieur Jourdain speaking in prose, we were actively involved in proving theorems in Diophantine approximation without knowing it.

Indeed, let us see how $\mathbf{y} \in \mathbb{R}^n$ is characterized by the fact that $\{a_t u_{\mathbf{y}} \mathbb{Z}^{n+1}\}$ has growth $\le \gamma_0$. Suppose that for any $\gamma > \gamma_0$ there exists $T > 0$ such that for any $t > T$ and any nonzero $(p, \mathbf{q}) \in \mathbb{Z} \times \mathbb{Z}^n$ one has

$$\left\| a_t u_{\mathbf{y}} \begin{pmatrix} p \\ \mathbf{q} \end{pmatrix} \right\| = \max\left(e^{nt}|p + \mathbf{y}\cdot\mathbf{q}|, e^{-t}\|\mathbf{q}\|\right) \ge e^{-\gamma t}\,. \tag{4.4}$$

For such γ, given $\mathbf{q} \in \mathbb{Z}^n$, choose t such that $e^{-t}\|\mathbf{q}\| = e^{-\gamma t} \Leftrightarrow e^t = \|\mathbf{q}\|^{\frac{1}{1-\gamma}}$. For large enough $\|\mathbf{q}\|$ this t will be greater than T. In view of (4.4), $e^{nt}|p + \mathbf{y}\cdot\mathbf{q}|$ must be at least $e^{-\gamma t}$, which translates into

$$|p + \mathbf{y}\cdot\mathbf{q}| \ge e^{-(n+\gamma)t} = \|\mathbf{q}\|^{-\frac{n+\gamma}{1-\gamma}}\,.$$

We proved that $\{a_t u_{\mathbf{y}} \mathbb{Z}^{n+1}\}$ having growth rate $\le \gamma_0$ implies that $\mathbf{y}$ is Diophantine[2] of order v for any $v > \frac{n+\gamma_0}{1-\gamma_0}$. In fact, converse implication is also true, and is left as an exercise; see [KM2, Kl2]. Consequently, sublinear growth of $\{a_t u_{\mathbf{y}} \mathbb{Z}^{n+1}\}$ is equivalent to $\mathbf{y}$ being Diophantine of all orders $> n$; those $\mathbf{y}$ are called *not very well approximable*, to be abbreviated as not VWA. It is an elementary fact, immediately implied by the Borel-Cantelli Lemma, that λ-a.e. $\mathbf{y} \in \mathbb{R}^n$ is not VWA. Thus we can reformulate the theorem proved in the previous section as follows:

Theorem 4.5. *Let U be an open subset of $\mathbb{R}$ and let $\mathbf{f} : U \to \mathbb{R}^n$ be a continuous good nonplanar map. Then for λ-a.e. $x \in U$, $\mathbf{f}(x)$ is not VWA.*

Results of this type have a long history, see [BD] and surveys [Sp3, Kl1, Kl4]. The above statement was conjectured by Mahler [Mah] in 1932 for

$$\mathbf{f}(x) = (x, x^2, \dots, x^n)\,. \tag{4.5}$$

This curve is indeed somewhat special: for any $x \in \mathbb{R}$, $(x, x^2, \dots, x^n)$ is VWA if and only if for some $v > n$ there are infinitely many integer polynomials P of degree $\le n$ such that $|P(x)| < (\text{height of } P)^{-v}$. Thus Mahler's Conjecture asserts, roughly speaking, that almost all transcendental numbers are "not very algebraic". Mahler himself proved a bound with a weaker exponent, and the full strength of the conjecture was established in 1964 by Sprindžuk. [Sp1, Sp2]. Then Sprindžuk in 1980 [Sp3] made the following

Conjecture 4.6 (now a theorem)**.** *For open $U \subset \mathbb{R}$, let $\mathbf{f} : U \to \mathbb{R}^n$ be nonplanar and real analytic. Then for λ-a.e. $x \in U$, $\mathbf{f}(x)$ is not VWA.*

This was proved in [KM1] via deducing it from a more general Theorem 4.5. Even for general polynomial maps, not of the form (4.5), this was new.

[2]Definition: $\mathbf{y}$ is *Diophantine of order* v if $|\mathbf{q}\cdot\mathbf{y} + p| \ge \text{const}\,\|\mathbf{q}\|^{-v}$ for all large enough $\mathbf{q}$ and all p. Note that here we interpret $\mathbf{y}$ as a linear form, but the method is equally well applicable to treating $\mathbf{y}$ as a vector, that is, looking at inequalities of type $\|q\mathbf{y} + \mathbf{p}\| \ge \text{const}\,|q|^{-v}$, where $q \in \mathbb{Z}$ and $\mathbf{p} \in \mathbb{Z}^n$. The book [Sch] by Schmidt is an excellent reference.

At this point the only missing part for us is to understand why real analytic implies good. The explanation involves passing from C^∞ to C^k class. The next lemma produces a wide variety of examples of good functions:

Lemma 4.7. *For any $k \in \mathbb{N}$ there exists $C_k > 0$ such that whenever an interval $B \subset \mathbb{R}$ and $f \in C^k(B)$, $k \in \mathbb{N}$, are such that for some $0 < a \le A$ one has*

$$a \le |f^{(k)}(x)| \le A \quad \forall\, x \in B\,, \tag{4.6}$$

then f is $\big(C_k(A/a)^{1/k}, 1/k\big)$-good on B.

This can be seen as a generalization of Proposition 3.2: indeed, polynomials of degree k satisfy the above assumptions with $A = a$.

Proof. We outline the argument in the case $k = 2$, with $C_2 = 2\sqrt{22}$; an extension to arbitrary degree of smoothness is straightforward and is left as an exercise, see [KM1] for hints. (However it is interesting that letting $A = a$ in the above lemma produces a constant which is not as good as the one for polynomials.)

Fix a subinterval J of B and denote by d the length of J and by s the supremum of $|f|$ on J. Take $\varepsilon > 0$; since, by the lower estimate in (4.6), the second derivative of f does not vanish on J, the set $\{x \in J : |f(x)| < \varepsilon\}$ consists of at most 2 intervals. Let I be the maximal of those, and denote its length by r. Then

$$\lambda(\{x \in J : |f(x)| < \varepsilon\}) \le 2r\,, \tag{4.7}$$

so it suffices to estimate r from above.

Sublemma 4.8. *$r \le 2\sqrt{6\varepsilon/a}$.*

Proof. Let x_1, x_2, x_3 be the left endpoint, midpoint and right endpoint of I respectively, and let P be the Lagrange polynomial of degree 2 formed by using values of f at these points, i.e. given by the expression in the right hand side of (3.3) with $k = 2$. Then there exists $x \in I$ such that $P''(x) = f''(x)$. Hence, by the lower estimate in (4.6), $|P''(x)| \ge a$. On the other hand, one can differentiate the right hand side of (3.3) twice to get $|P''(x)| \le 3\varepsilon\dfrac{2}{(r/2)^2} = 24\varepsilon/r^2$. Combining the last two inequalities yields the desired estimate. □

Now recall that, since we are after the (C,α)-good property, we would like to have an upper estimate for r in the form $r \le C(\varepsilon/s)^\alpha d$. Thus let us rewrite the conclusion of the lemma as

$$r \le 2\sqrt{\frac{6s}{ad^2}}\left(\frac{\varepsilon}{s}\right)^{1/2} d = 2\sqrt{\frac{6t}{a}}\left(\frac{\varepsilon}{s}\right)^{1/2} d\,, \tag{4.8}$$

where we introduced a parameter $t \stackrel{\text{def}}{=} s/d^2$. We see that the above estimate is useful when t is small, and to finish the proof it suffices to produce an estimate improving (4.8) for large values of t. Here it goes:

Sublemma 4.9. $r \le \sqrt{\frac{10A/a}{1-A/2t}} \cdot \left(\frac{\varepsilon}{s}\right)^{1/2} d.$

Proof. Let Q be the Taylor polynomial of f of degree 1 at x_1. By Taylor's formula,

$$|f(x_2) - Q(x_2)| \le \sup_{x\in I} |f''(x)| \frac{(r/2)^2}{2} \underset{(4.6)}{\le} \frac{Ar^2}{8} \underset{\text{Lemma 4.8}}{\le} \frac{A}{8}\frac{24\varepsilon}{a} = 3\frac{A}{a}\varepsilon\,.$$

But also $|f(x_2)| \le \varepsilon$, therefore

$$|Q(x_2)| \le \left(3\frac{A}{a} + 1\right)\varepsilon \underset{\text{to simplify computations}}{\le} 4\frac{A}{a}\varepsilon\,.$$

We now apply Lagrange's formula to reconstruct Q on B by its values at x_1, x_2. As in the proof of Proposition 3.2, we get

$$\|Q\|_B \le \left(4\frac{A}{a}\varepsilon + \varepsilon\right)\frac{d}{r/2} \underset{\text{to simplify computations}}{\le} 10\frac{A}{a}\cdot\varepsilon\frac{d}{r}\,. \tag{4.9}$$

Finally, the difference between f and Q on B is, again by the upper estimate in (4.6), bounded from above by $Ad^2/2$, so from (4.9) one deduces that

$$s \le 10\frac{A}{a}\cdot\varepsilon\frac{d}{r} + Ad^2/2 \underset{\text{to simplify computations}}{\le} 10\frac{A}{a}\cdot\varepsilon\frac{d^2}{r^2} + Ad^2/2\,,$$

or, equivalently,

$$r \le \sqrt{\frac{10\frac{A}{a}\cdot\varepsilon d^2}{s - Ad^2/2}} = \sqrt{\frac{10\frac{A}{a}}{1 - Ad^2/2s}}\cdot\left(\frac{\varepsilon}{s}\right)^{1/2} d\,.$$

which is what we wanted to prove. □

It remains to observe (Exercise) that the right hand sides of the two inequalities in Sublemmas 4.8 and 4.9 are equal to each other when $t = 11A/12$, and substitute $t = 11A/12$ in (4.8) to obtain $r \le 2\sqrt{\frac{11A}{2a}}\left(\frac{\varepsilon}{s}\right)^{1/2} d$, which, in view of (4.7), gives the conclusion of Lemma 4.7. □

Here is another important definition. Say that $\mathbf{f}$ is *ℓ-nondegenerate at x* if $\mathbb{R}^n$ is spanned by $\mathbf{f}'(x), \mathbf{f}''(x), \dots, \mathbf{f}^{(\ell)}(x)$. We will say that $\mathbf{f} : U \to \mathbb{R}^n$ is *ℓ-nondegenerate* if it is ℓ-nondegenerate at almost every point. It is clear that nondegeneracy implies nonplanarity (if $f(B)$ belongs to a proper affine hyperplane for some interval B, derivatives of all orders at any point of B won't generate anything more than the tangent space to this hyperplane). On top of this, we also have

Proposition 4.10. *Nondegenerate maps are good. More precisely, if $\mathbf{f}$ is ℓ-nondegenerate at x_0, then there exists a neighborhood B of x_0 and positive C such that any linear combination of $1, f_1, \dots, f_n$ is $(C, 1/\ell)$-good on B; and with a little more work this C can be chosen uniformly if we are given an ℓ-nondegenerate $\mathbf{f} : U \to \mathbb{R}^n$.*

Proof. The ℓ-nondegeneracy of $\mathbf{f}$ at x_0 implies that for any $\mathbf{c} = (c_1, \dots, c_n) \neq 0$ there exists $1 \leq k \leq \ell$ such that $\mathbf{c} \cdot \mathbf{f}^{(k)}(x_0) \neq 0$, and in fact for this k depending on $\mathbf{c}$ there is a uniform lower bound on $|\mathbf{c} \cdot \mathbf{f}^{(k)}(x)|$ over all $\mathbf{c}$ on (or outside) the unit sphere and x in some neighborhood B of x_0. But $\mathbf{c} \cdot \mathbf{f}^{(k)} = f^{(k)}$ where $f = c_0 + \sum_{i=1}^n c_i f_i$; this produces an a as in Lemma 4.7, and using it with some upper bound A (which can be made closer to a by making B smaller) one concludes that f is $(C, 1/k)$-good (and therefore $(C, 1/\ell)$-good) on B. □

Proof of Conjecture 4.6. It remains to take a nonplanar analytic $\mathbf{f} : U \to \mathbb{R}$, $U \subset \mathbb{R}$ a bounded interval, and verify that it must be ℓ-nondegenerate with some uniform ℓ. This is an easy exercise. (Hint: if derivatives of $\mathbf{f}$ at x_k of order up to k are contained in a hyperplane L_k, then all derivatives of $\mathbf{f}$ at $\lim x_k$ will be contained in $\lim L_k$.) □

We remark that as long as the nonplanarity of $\mathbf{f}$ is assumed, we are guaranteed to have condition [3.4-ii] with some ρ uniform in t, and do not really care to distinguish between ρ and $\rho^{\mathrm{rk}(\Delta)}$. However this distinction becomes important when $\mathbf{f}(B)$ belongs to a proper affine subspace $L \subset \mathbb{R}^n$. Then it matters how fast L can be approximated by rational subspaces. It is possible to use Theorem 4.2 to write a condition on L equivalent to almost all ($\Leftrightarrow$ at least one of) its points being not VWA, or more generally, Diophantine of order v for all $v > v_0$, and also to prove that this condition is inherited by $\mathbf{f}(B)$ whenever $\mathbf{f}$ is good and "nonplanar in L", in particular, smooth and "nondegenerate in L" [Kl5]. (Exercises: give definitions of the terms in quotation marks, show that nondegenerate in $L \Rightarrow$ good and nonplanar in L, and real analytic $\Rightarrow$ nondegenerate in some L.)

4.4. **More about the correspondence between approximation and dynamics.** The principle that was used to connect growth rate of $\{a_t u_{\mathbf{y}} \mathbb{Z}^{n+1}\}$ and approximation properties of $\mathbf{y}$ has various manifestations and has been extensively used to relate Diophantine approximation to dynamics. In a nutshell, a very good approximation for $\mathbf{y}$ amounts to a very small value of the function $(p, \mathbf{q}) \mapsto \|\mathbf{q}\|^n |p + \mathbf{y} \cdot \mathbf{q}|$ at a nonzero integer point $\Leftrightarrow$ a small value of the function $\mathbf{v} = (v_0, v_1, \dots, v_n) \mapsto |v_0| \cdot \max(|v_1|, \dots, |v_n|)^n$ at a nonzero vector $\mathbf{v}$ from the latice $u_{\mathbf{y}} \mathbb{Z}^{n+1}$. And the reason for the n-th power is precisely to make the latter function invariant by $a_t \in \mathrm{SL}(n+1, \mathbb{R})$ and use the action to produce a very small nonzero vector in the lattice $a_t u_{\mathbf{y}} \mathbb{Z}^{n+1}$ ($\Leftrightarrow$ a very deep excursion into the cusp).

The same principle was involved in the reduction of the Oppenheim conjecture to dynamics of the stabilizer of a quadratic form on the space of lattices, described in detail in [E]. For the same reasons, the trajectory $a_t u_{\mathbf{y}} \mathbb{Z}^{n+1}$ is bounded in $\mathcal{L}_{n+1}$ if and only if $\mathbf{y}$ is *badly approximable*[3]; this is a theorem of Dani [D3].

As another application of this principle, consider the product of $n+1$ linear forms, $\mathbf{v} \mapsto \prod_{i=0}^n |v_i|$; its stabilizer is the full diagonal subgroup of $\mathrm{SL}(n+1, \mathbb{R})$. It is not

[3] that is, if and only if $\inf_{p \in \mathbb{Z}, \mathbf{q} \in \mathbb{Z}^n \smallsetminus \{0\}} \|\mathbf{q}\|^n |p + \mathbf{y} \cdot \mathbf{q}| > 0$

hard to show that the orbit of the lattice $u_{\mathbf{y}}\mathbb{Z}^{n+1}$ under the action of the semigroup

$$a_{\mathbf{t}} = \operatorname{diag}(e^{t_1+\cdots+t_n}, e^{-t_1}, \dots, e^{-t_n}), \quad t_i > 0$$

is bounded in $\mathcal{L}_{n+1}$ if and only if $\mathbf{y}$ is an exception to the (n-dimensional version of the) Littlewood's Conjecture. More about it can be found in [EL]

In fact, it is worthwhile to mention that all the results discussed in this section have their multi-parameter analogues; if $h_{\mathbf{t}}(x) = a_{\mathbf{t}}h_0(x)$ is used instead of h_t, it is often possible to establish conditions [3.4-ii] and [3.4-ii] uniformly over $\mathbf{t} \in \mathbb{R}^n_+$. This yields the proof of a *multiplicative* version of Sprindžuk's Conjecture 4.6 (a special case for the curve (4.5) was conjectured earlier by A. Baker) and many of its generalizations.

Also note that the correspondence described above already made an appearance in two more lecture courses in this volume: by Svetlana Katok in the case $n = 1$ [K], where it was shown that the diagonal action on $\mathcal{L}_2$ is a suspension of the Gauss map, and by Jean-Christophe Yoccoz [Y], who treated the "$n = 1$" case as a platform for a generalization of the aforementioned suspension to the moduli space of translation surface structures in higher genus. Here we are talking about a generalization of a different kind. In fact one can also treat the a_t-action on $\mathcal{L}_{n+1}$ as a suspension of certain first return map, thus obtaining a higher-dimensional version of continued fractions. However in order to obtain results in Diophantine approximation it is often efficient to simply work with the suspension itself, as was demonstrated during these lectures.

Let us describe one more example of the use of Theorem 3.4 in a slightly different context. Suppose we fix $\varepsilon > 0$ and want to look at the a_t-trajectories which, starting from some t_0, decide to leave the compact set $\mathcal{L}_{n+1}(\varepsilon)$ and never come back again. This is possible, for example the a_t-trajectory can be divergent; but of course this can only happen on a null set by ergodicity, i.e. for any ε, the trajectory $\{a_t u_{\mathbf{y}}\mathbb{Z}^{n+1}\}$ will do it for almost no $\mathbf{y}$. Now what about $\mathbf{y}$ of the specific form $\mathbf{f}(x)$, for $\mathbf{f}$ as above, for Lebesgue-generic x?

Before answering this question, let me restate it in Diophantine language: such behavior amounts to existence of t_0 such that for any $t > t_0$ the system

$$|\mathbf{y}\cdot\mathbf{q} + p| < \varepsilon e^{-tn} \quad \text{and} \quad \|\mathbf{q}\| < \varepsilon e^t$$

has a nontrivial integer solution $(p, \mathbf{q})$. Number theorists say in this situation that *Dirichlet's Theorem can be ε-improved* for $\mathbf{y}$, see [DS, **?**]. Partial results, such as for $\mathbf{f}$ of the form (4.5), has been known due to Davenport–Schmidt [DS], Baker [Ba1, Ba2] and Bugeaud [Bu].

Let us now see what one can do with the dynamical method. Assume that $\mathbf{f} : U \to \mathbb{R}$ is (C,α)-good; then condition [3.4-i] will hold with uniform C, α for all $B \subset U$ except for those touching a null set. If $\mathbf{f}$ is also nonplanar, we can use Lemnma 4.3(b) and for any B find t_0 such that [3.4-ii] holds with $\rho = 1$. Then take ε such that $n2^nC\varepsilon^\alpha < 1$. It follows that for almost every $x_0 \in U$, any interval $B \subset U$ centered at

x_0, the intersection of B with

$$\{x \in U : a_t u_{\mathbf{y}} \mathbb{Z}^{n+1} \notin \mathcal{L}_{n+1}(\varepsilon) \text{ for large enough } t\} \tag{4.10}$$

has relative measure in B strictly less than one. By the Lebesgue Density Theorem, the set (4.10) must have measure zero. This conclusion can be phrased as follows:

Theorem 4.11. *For any n, C, α there exists $\varepsilon_0 > 0$ with the following property: let U be an open subset of $\mathbb{R}$ and let $\mathbf{f} : U \to \mathbb{R}^n$ be a continuous (C, α)-good nonplanar map. Then for any $\varepsilon < \varepsilon_0$ and for λ-a.e. $x \in U$, Dirichlet's Theorem cannot be ε-improved for $\mathbf{f}(x)$.*

This was done several years ago in [KW2] in a much more general, multiplicative, context. Then, in the special case when $\mathbf{f}$ is real analytic and nonplanar, the above result was sharpened by Shah [Sh1, Sh2], who established that the conclusion of Theorem 4.11 holds for arbitrary $\varepsilon_0 < 1$ using approximation by unipotent trajectories and the linearization technique described in [E].

5. Concluding remarks

5.1. **Generalizations.** Quantitative nondivergence results can be proved, and have applications, in a much more general situations then described above. In particular, intervals $B \subset \mathbb{R}$ can be replaced by balls in a metric space satisfying certain (Besicovitch) covering property; other measures, including those supported on fractals, can be used instead of Lebesgue measure λ. These generalizations are based on the work done by [KLW, KT], but most of the main ideas are contained in the proof presented in §3.3. Among applications to number theory was the proof of so-called Khintchine-type theorems on nondegenerate manifolds, both convergence and divergence cases (with Beresnevich, Bernik and Margulis, see [BKM, BBKM]). More recent developments include studying Diophantine properties of points on fractal subsets of $\mathbb{R}^n$. For example, a repeated application of Theorem 3.4 to measures supported on certain self-similar fractals (or, more generally, satisfying certain decay conditions) allows to construct many bounded orbits (read: badly approximable vectors) in the supports of those measures. This was recently done in [KW1]. Other applications involve analogues of Diophantine approximation results in the S-arithmetic [KT] and positive characteristic [G2] setting.

5.2. **What else is in the proof.** It is fair to say that the approach to metric Diophantine approximation described above is definitely not the only one available. Methods originally developed by Sprindžuk have produced many important results, including an independent proof of Conjecture 4.6 by Beresnevich [Ber], and also including many theorems which do not seem to be attainable by dynamics on the space of lattices. However in some cases one can see how the main constructions used in the the proof of Theorem 3.4 (primitive subgroups of various ranks) show up in the other proofs in various disguises, and the passage to the space of latices seems to work as a tool to somehow change variables and compress all the induction argument into one scheme based on the partial order of subgroups of $\mathbb{Z}^n$.

Another classical argument somewhat similar to that proof is the reduction algorithm of Minkowski (mentioned in [K]) and Siegel (its generalization to $\mathcal{L}_n$). In fact, flags naturally appear in the construction of Siegel sets (fundamental sets for $\mathrm{SL}(n,\mathbb{R})/\mathrm{SL}(n,\mathbb{Z})$), hence in the proof of the finiteness of Haar measure on $\mathcal{L}_n$. Another feature of the proof of Theorem 3.4 is the fact that it does not use the fact that $\mathcal{L}_n$ has finite volume. And indeed, it can be perhaps thought of as an alternative approach to reduction theory. Using Corollary 3.5 (high frequency of visits of unipotent trajectories to compact sets), one can construct an everywhere positive u_x-invariant integrable function on $\mathcal{L}_n$, which would contradict to Moore's Ergodicity Theorem unless $\nu(\mathcal{L}_n)$ is finite, thus providing an alternative proof of a theorem of Borel and Harish-Chandra. Details are left as an exercise. A more general version of this exercise is a theorem of Dani [D1], which he derived from Corollary 3.5, that any locally finite ergodic u_x-invariant measure on $\mathcal{L}_n$ is finite.

5.3. **From $\mathcal{L}_n$ to G/Γ and beyond.** The reader is invited to look at Dani's papers from late 1970s and early 1980s [D1, D2, D4], see also [DM1], where it is shown how Corollary 3.5 and other quantitative nondivergence results can be extended to an arbitrary homogeneous space, using the Margulis Arithmeticity Theorem and some standard facts from the theory of algebraic groups. See also [KT] and [GO] where a similar reduction is carried out in the S-arithmetic case. Finally, it is worthwhile to mention another important theme of the summer school, the analogy between homogeneous space of Lie groups and moduli spaces of translation surface structures. The scheme of proof presented in §3 was used in [MW] to establish quantitative nondivergence of horocyclic flows on the moduli space of quadratic differentials, see also [LM, Appendix].

References

[Ba1] R.C. Baker, *Metric diophantine approximation on manifolds*, J. Lond. Math. Soc. (2) **14** (1976), 43–48.

[Ba2] ______, *Dirichlet's theorem on diophantine approximation*, Math. Proc. Cambridge Phil. Soc. **83** (1978), 37–59.

[BM] M. Bekka and M. Mayer, **Ergodic theory and topological dynamics of group actions on homogeneous spaces**, London Math. Soc. Lecture Note Series, Vol. 269, Cambridge University Press, Cambridge, 2000.

[Ber] V. Beresnevich, *A Groshev type theorem for convergence on manifolds*, Acta Mathematica Hungarica **94** (2002), 99–130.

[BBKM] V. Beresnevich, V.I. Bernik, D. Kleinbock and G.A. Margulis, *Metric Diophantine approximation: the Khintchine-Groshev theorem for nondegenerate manifolds*, Moscow Math. J. **2** (2002), 203–225.

[BD] V.I. Bernik and M.M. Dodson, **Metric Diophantine approximation on manifolds**, Cambridge University Press, Cambridge, 1999.

[BKM] V.I. Bernik, D. Kleinbock and G.A. Margulis, *Khintchine-type theorems on manifolds: convergence case for standard and multiplicative versions*, Internat. Math. Res. Notices **2001**, 453–486.

[Bu] Y. Bugeaud, *Approximation by algebraic integers and Hausdorff dimension*, J. London Math. Soc. (2) **65** (2002), no. 3, 547–559.

[Cas] J.W.S. Cassels, **An introduction to the geometry of numbers**, Springer, Berlin, 1997.

[D1] S.G. Dani, *On invariant measures, minimal sets and a lemma of Margulis*, Invent. Math. **51** (1979), 239–260.

[D2] ———, *On orbits of unipotent flows on homogeneous spaces*, Ergod. Theor. Dynam. Syst. **4** (1984), 25–34.

[D3] ———, *Divergent trajectories of flows on homogeneous spaces and diophantine approximation*, J. Reine Angew. Math. **359** (1985), 55–89.

[D4] ———, *On orbits of unipotent flows on homogenous spaces II*, Ergod. Theor. Dynam. Syst. **6** (1986), 167–182.

[DM1] S.G. Dani and G.A. Margulis. *Asymptotic behaviour of trajectories of unipotent flows on homogeneous spaces*, Indian. Acad. Sci. J. **101** (1991), 1–17.

[DM2] ———, *Limit distributions of orbits of unipotent flows and values of quadratic forms*, in: **I.M. Gelfand Seminar**, Amer. Math. Soc., Providence, RI, 1993, pp. 91–137.

[DS1] H. Davenport and W. M. Schmidt, *Dirichlet's theorem on diophantine approximation*, in: Symposia Mathematica, Vol. IV (INDAM, Rome, 1968/69), pp. 113–132, 1970.

[DS2] ———, *Dirichlet's theorem on diophantine approximation. II*, Acta Arith. **16** (1969/1970) 413–424.

[E] A. Eskin, *Unipotent Flows and Applications*, lecture notes in this volume.

[EL] M. Einsiedler and E. Lindenstrauss, *Diagonalizable actions and arithmetic applications*, lecture notes in this volume.

[EMS] A. Eskin, S. Mozes and N. Shah, *Unipotent flows and counting lattice points on homogeneous varieties*, Ann. Math. **143** (1996), 253–299.

[G1] A. Ghosh, *A Khintchine-type theorem for hyperplanes*, J. London Math. Soc. (2) **72** (2005), no. 2, 293–304.

[G2] ———, *Metric Diophantine approximation over a local field of positive characteristic*, J. Number Theory **124** (2007), 454–469.

[GO] A. Gorodnik and H. Oh, *Rational points on homogeneous varieties and equidistribution of adelic periods* (with an appendix by M. Borovoi), preprint, `arXiv:0803.1996v2`.

[K] S. Katok, *Fuchsian groups, geodesic flows on surfaces of constant negative curvature and symbolic coding of geodesics*, lecture notes in this volume.

[Kl1] D. Kleinbock, *Some applications of homogeneous dynamics to number theory*, in: **Smooth ergodic theory and its applications**, Proc. Sympos. Pure Math., Vol. 69, Amer. Math. Soc., Providence, RI, 2001, pp. 639–660.

[Kl2] ———, *Extremal subspaces and their submanifolds*, Geom. Funct. Anal. **13** (2003), 437–466.

[Kl3] ———, *Baker-Sprindžuk conjectures for complex analytic manifolds*, in: **Algebraic groups and Arithmetic**, Tata Inst. Fund. Res., Mumbai, 2004, 539–553.

[Kl4] ———, *Diophantine properties of measures and homogeneous dynamics*, Pure Appl. Math. Quarterly **4** (2008), 81–97.

[Kl5] ———, *An extension of quantitative nondivergence and applications to Diophantine exponents*, Trans. Amer. Math. Soc. **360** (2008), 6497–6523.

[Kl6] ______, *An 'almost all versus no' dichotomy in homogeneous dynamics and Diophantine approximation*, arXiv:0904.1614, Geom. Dedicata, to appear.

[KLW] D. Kleinbock, E. Lindenstrauss and B. Weiss, *On fractal measures and diophantine approximation*, Selecta Math. **10** (2004), 479–523.

[KM1] D. Kleinbock and G.A. Margulis, *Flows on homogeneous spaces and Diophantine approximation on manifolds*, Ann. Math. **148** (1998), 339–360.

[KM2] ______, *Logarithm laws for flows on homogeneous spaces*, Invent. Math. **138** (1999), 451–494.

[KSS] D. Kleinbock, N. Shah and A. Starkov, *Dynamics of subgroup actions on homogeneous spaces of Lie groups and applications to number theory*, in: **Handbook on Dynamical Systems**, Volume 1A, Elsevier Science, North Holland, 2002, pp. 813–930.

[KT] D. Kleinbock and G. Tomanov, *Flows on S-arithmetic homogeneous spaces and applications to metric Diophantine approximation*, Comm. Math. Helv. **82** (2007), 519–581.

[KW1] D. Kleinbock and B. Weiss, *Badly approximable vectors on fractals*, Israel J. Math. **149** (2005), 137–170.

[KW2] ______, *Dirichlet's theorem on diophantine approximation and homogeneous flows*, J. Mod. Dyn. **2** (2008), 43–62.

[LM] E. Lindenstrauss and M. Mirzakhani, *Ergodic theory of the space of measured laminations*, Internat. Math. Res. Notices **2008**, no. 4, doi:10.1093/imrn/rnm126.

[Mah] K. Mahler, *Über das Mass der Menge aller S-Zahlen*, Math. Ann. **106** (1932), 131–139.

[Mar] G.A. Margulis, *On the action of unipotent groups in the space of lattices*, In **Lie Groups and their representations**, Proc. of Summer School in Group Representations, Bolyai Janos Math. Soc., Akademai Kiado, Budapest, 1971, pp. 365–370, Halsted, New York, 1975.

[MW] Y. Minsky and B. Weiss, *Nondivergence of horocyclic flows on moduli space*, J. Reine Angew. Math. **552** (2002), 131–177.

[Mor] D. Morris, **Ratner's theorems on unipotent flows**, Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 2005.

[Sch] W. Schmidt, **Diophantine approximation**, Springer-Verlag, Berlin, 1980.

[Sh1] N. Shah, *Equidistribution of expanding translates of curves and Dirichlet's theorem on Diophantine approximation*, Invent. Math. **177** (2009), no. 3, 509–532.

[Sh2] ______, *Expanding translates of curves and Dirichlet-Minkowski theorem on linear forms*, Preprint, arXiv:0804.1424.

[Sp1] V.G. Sprindžuk, *More on Mahler's conjecture*, Dokl. Akad. Nauk SSSR, **155** (1964), 54–56.

[Sp2] ______, **Mahler's problem in metric number theory**, Translations of Mathematical Monographs, Amer. Math. Soc., Providence, R.I., 1969.

[Sp3] ______, *Achievements and problems in Diophantine approximation theory*, Russian Math. Surveys **35** (1980), 1–80.

[St] A. Starkov, **Dynamical systems on homogeneous spaces**, Translations of Mathematical Monographs, Amer. Math. Soc., Providence, RI, 2000.

[Y] J-C. Yoccoz, *Interval exchange maps and translation surfaces*, lecture notes in this volume.

Brandeis University, Waltham MA 02454-9110 kleinboc@brandeis.edu