JEMS

Emmanuel Breuillard · Nicolas de Saxcé

A subspace theorem for manifolds

Abstract. We prove a theorem that generalizes Schmidt's Subspace Theorem in the context of metric diophantine approximation. To do so we reformulate the Subspace Theorem in the framework of homogeneous dynamics by introducing and studying a slope formalism and the corresponding notion of semistability for diagonal flows.

Keywords. Diophantine approximation, lattices, algebraic numbers

Introduction

In 1972, Wolfgang Schmidt formulated his celebrated *subspace theorem* [36, Lemma 7], a far reaching generalization of results of Thue [41], Siegel [39], and Roth [31] on rational approximations to algebraic numbers. Around the same time, in his work on arithmeticity of lattices in Lie groups, Gregory Margulis [29] used the geometry of numbers to establish the recurrence of unipotent flows on the space of lattices $\mathrm{GL}_d(\mathbb{R})/\mathrm{GL}_d(\mathbb{Z})$. More than two decades later, a quantitative refinement of this fact, the so-called quantitative non-divergence estimate, was used by Kleinbock and Margulis [25] in their solution to the Sprindzuk conjecture regarding the extremality of non-degenerate manifolds in metric diophantine approximation. As it turns out, these two remarkable results – the subspace theorem and the Sprindzuk conjecture – are closely related and can be understood together as statements about diagonal orbits in the space of lattices. In this paper we prove a theorem that generalizes both results at the same time. We also provide several applications. This marriage is possible thanks to a better understanding of the geometry lying behind the subspace theorem, in particular the notion of Harder–Narasimhan filtration for one-parameter diagonal actions, which leads both to a dynamical reformulation of the original subspace theorem and to a further geometric understanding of the family of exceptional subspaces arising in Schmidt's theorem. The proof blends the diophantine

Emmanuel Breuillard: Mathematical Institute, University of Oxford,
Oxford OX2 6GG, United Kingdom; emmanuel.breuillard@maths.ox.ac.uk

Nicolas de Saxcé: Laboratoire Analyse, Géométrie et Applications CNRS,
Université Sorbonne Paris Nord 99, 93430 Villetaneuse, France; desaxce@math.univ-paris13.fr

input of Schmidt's original theorem with the dynamical input arising from the Kleinbock–Margulis approach and refinements recently obtained in [2, 9].

We now formulate the main theorem. Let $M = \phi(U)$ be a connected analytic submanifold of $\mathrm{GL}_d(\mathbb{R})$ parametrized by an analytic map $\phi : U \to \mathrm{GL}_d(\mathbb{R})$, where $U \subset \mathbb{R}^n$ is a connected open set and $n \in \mathbb{N}$. Let μ be the push-forward in M of the Lebesgue measure on U. The Zariski closure of M is said to be *defined over* $\overline{\mathbb{Q}}$ if for $(a_{ij})_{ij} \in \mathrm{GL}_d$ the ideal of polynomial functions in $\mathbb{C}[a_{ij}, \det^{-1}]$ that vanish on M can be generated by polynomials with coefficients in $\overline{\mathbb{Q}}$.

Theorem 1 (Subspace theorem for manifolds). *Assume that the Zariski closure of M in GL_d is defined over $\overline{\mathbb{Q}}$. Then there exists a finite family of proper subspaces $V_1, \ldots, V_r$ of $\mathbb{Q}^d$ with $r = r(d)$ such that, for μ-almost every L in M, for every $\varepsilon > 0$, all but finitely many integer solutions $\mathbf{x} \in \mathbb{Z}^d$ to the inequality*

$$\prod_{i=1}^{d} |L_i(\mathbf{x})| \le \|\mathbf{x}\|^{-\varepsilon} \tag{1}$$

lie in the union $V_1 \cup \cdots \cup V_r$.

Here the L_i are the linear forms on $\mathbb{R}^d$ given by the rows of $L \in \mathrm{GL}_d(\mathbb{R})$, and $\|\cdot\|$ is the canonical Euclidean norm on $\mathbb{R}^d$. Note that we only assume that the Zariski closure of M is defined over $\overline{\mathbb{Q}}$, so that the manifold M itself may well be transcendental. We will in fact prove the theorem under a slightly weaker assumption on M requiring only that what we call the *Plücker closure* of M be defined over $\overline{\mathbb{Q}}$; see §1.8 for the definition of Plücker closure.

Of course, the main motivation to derive the above theorem was the usual subspace theorem and its numerous applications, for which the reader is referred to Schmidt [37], Vojta [42], Edixhoven–Evertse [10], or Bombieri–Gubler [5]. And we recover the original Schmidt subspace theorem as the special case of Theorem 1 when M is a singleton $\{L\}$ (then $n = 0$ and μ is the Dirac mass at L). However, since we shall use Schmidt's result to derive the above theorem, our approach does not yield a new proof of the subspace theorem. The main interest of our statement is when the manifold M is positive-dimensional. The theorem is then non-trivial even in the case where M is defined over $\mathbb{Q}$. Indeed, as we shall see in Section 3.1, it recovers the main result of Kleinbock and Margulis regarding the Sprindzuk conjecture.

The exceptional subspaces V_i are independent of ε, as in Vojta's refinement [38, 42], and they depend only on the *rational* Zariski closure of M. In fact, they are determined by what we call the *rational Schubert closure* of M, that is, the intersection of all rational translates $\mathcal{S}_\sigma g := \overline{B\sigma B}$ containing M, where $g \in \mathrm{GL}_d(\mathbb{Q})$ and $\mathcal{S}_\sigma$ is a standard Schubert variety associated to a permutation σ and a Borel subgroup B containing the diagonal subgroup. Each V_i contains infinitely many solutions to (1), regardless of ε. The number r of exceptional subspaces can be bounded by a number depending only on d (see Lemma 3 and the remark following it).

The proof of Theorem 1 goes via the proof of a stronger result, a parametric subspace theorem for manifolds, Theorem 3 below. The idea behind parametric versions of the subspace theorem, already present in Schmidt's work [35, Theorem 3], is to study the behavior of the successive minima of a lattice along the orbit of a one-parameter diagonal subgroup $(a_t)_{t \in \mathbb{R}}$ in $\mathrm{GL}_d(\mathbb{R})$, i.e. $a_t = e^{tA}$ for some diagonal matrix A. Former versions of the parametric subspace theorem, such as Faltings–Wüstholz [16, Theorem 9.1], Evertse–Schlickewei [15] or Evertse–Ferretti [13, Theorem 2.1], essentially state that short vectors in the lattices from the (a_t)-orbit are all contained in some fixed proper subspace. We refine this statement and show that it is in fact possible to obtain a complete description of the shape of the lattices along the diagonal orbit.

We may summarize the result informally as follows:

Theorem 2 (Parametric version). *For μ-almost every L in M the lattice $a_t L \mathbb{Z}^d$ assumes a fixed asymptotic shape as t tends to $+\infty$.*

By "fixed asymptotic shape" we mean two things. Firstly, the successive minima are asymptotic to $e^{\Lambda_k t}$ for some real numbers Λ_k, Lyapunov exponents of sorts, depending only on M and $a = (a_t)_{t>0}$ (the dependence on a is piecewise linear). In particular, as t varies there can only be oscillations of subexponential size for successive minima. And secondly, the successive minima determine a fixed partial flag in $\mathbb{Z}^d$. In other words, there is a fixed (independent of L) partial rational flag $W_1 \leq \cdots \leq W_d$ in $\mathbb{Q}^d$ such that if $\Lambda_k < \Lambda_{k+1}$, then the k first successive minima of $a_t L \mathbb{Z}^d$ are always realized by vectors from W_k when t is large enough. The constants Λ_k and the flag $\{W_k\}_k$ depend only on a and on the rational Schubert closure of M. As a first example, it is instructive to understand Theorem 2 in dimension 2, where it is equivalent to Roth's theorem.

Example (Roth's theorem and diagonal flows). In dimension 2, taking $a_t = \begin{pmatrix} 1 & 0 \\ 0 & e^t \end{pmatrix}$ and $L = \begin{pmatrix} 1 & 0 \\ -\alpha & 1 \end{pmatrix}$, where α is any real algebraic number, one can observe two different behaviors. If $\alpha \in \mathbb{Q}$, then $\Lambda_1 = 0$, $\Lambda_2 = 1$, and the first minimum is always attained on the line generated by $\begin{pmatrix} 1 \\ \alpha \end{pmatrix}$. If $\alpha \notin \mathbb{Q}$ however, given any non-zero $v \in \mathbb{Z}^2$, $\|a_t L v\|$ grows at the highest possible rate of e^t. Hence Theorem 2 forces $\Lambda_1 = \Lambda_2$ and this common value must be $\frac{1}{2}$ since $\Lambda_1 + \Lambda_2 = 1$. This is another way to formulate Roth's theorem.

In general, this rational flag arises naturally as the *Harder–Narasimhan filtration* associated to a certain submodular function on the rational grassmannian: the maximal expansion rate of the subspace under the flow. We recall in §1.3 that any submodular function on a grassmannian gives rise to a Grayson polygon, a notion of semistability, a Harder–Narasimhan filtration and certain coefficients, the slopes of the polygon, which in our case will correspond to the Lyapunov exponents Λ_k mentioned above. This is the so-called "slope formalism", which arises in particular in the study of Euclidean lattices as first described by Stuhler [40] and Grayson [17], and in many other subjects as well [6, 30]. The family of exceptional subspaces V_i appearing in Theorem 1 is obtained by grouping together the different W_i appearing when one varies the one-parameter subgroup a.

Although we have restricted to the current setting for clarity of exposition in this introduction, the result will be proved for more general measures μ than push-forwards

of the Lebesgue measure by analytic maps; the exceptional subspaces then depend only on the Zariski closure of the support of μ. The right technical framework is that of *good measures*, which are closely related to the friendly measures of [20]; see §1.9.

The paper is organized as follows. In Section 1 we begin by formulating the technical version of Theorem 2 and then proceed to describe the slope formalism on the grassmannian associated to a one-parameter flow and in particular discuss the associated notion of Harder–Narasimhan filtration. The proof of Theorems 1 and 2 is carried out in §1.5 and §1.6 after a discussion of the Kleinbock–Margulis quantitative non-divergence estimates. In Section 2 we formulate and sketch a proof of an extension of Theorem 1 to arbitrary number fields, which is analogous to the classical extension of Schmidt's subspace theorem due to Schlickewei to multiple places and targets [5, 34, 43]. Finally, in Section 3 we prove several applications of the main result.

For the sake of brevity we do not state these applications in the introduction and refer the reader to Section 3 directly instead. Let us only briefly mention that there are five main applications: (i) we explain how to recover the Sprindzuk conjecture (Kleinbock–Margulis theorem) from Theorem 1; (ii) we establish a manifold version of the classical Ridout theorem regarding approximation by rationals whose denominators have prescribed prime factors; (iii) we recover the main results of [2] regarding (weighted) diophantine approximation on submanifolds of matrices showing that they also hold for submanifolds defined over $\bar{\mathbb{Q}}$ (and not only over $\mathbb{Q}$); (iv) we prove an optimal criterion for strong extremality (Corollary 3), which answers in this case a question from [4, 21]; (v) we prove a Roth-type theorem for non-commutative diophantine approximation on nilpotent Lie groups, extending to algebraic points what was done for Lebesgue almost every point in our previous work with Aka and Rosenzweig [1, 2].

Further applications and an extension of some of the results of this paper to other reductive groups and homogeneous varieties can be found in the second author's forthcoming work [33].

Notation. Given two variable quantities A and B in $\mathbb{R}$, we use the Vinogradov notation $A \ll B$ to indicate that there exists a constant $C \geq 0$ such that $A \leq CB$. When we wish to specify the dependence of the constant C involved on some parameter L, we write $A \ll_L B$.

1. The main result

1.1. Dynamical formulation

A *lattice*, that is, a discrete subgroup Δ of rank d in $\mathbb{R}^d$, can be written as

$$\Delta = \mathbb{Z}u_1 \oplus \cdots \oplus \mathbb{Z}u_d,$$

where $(u_i)_{1 \leq i \leq d}$ is a basis of $\mathbb{R}^d$. And the space Ω of lattices can be identified with the homogeneous space

$$\Omega \simeq \mathrm{GL}_d(\mathbb{R})/\mathrm{GL}_d(\mathbb{Z}).$$

The position of a lattice Δ in the space Ω, up to a bounded error, is described by its *successive minima* $\lambda_1(\Delta) \leq \cdots \leq \lambda_d(\Delta)$, defined by

$$\lambda_i(\Delta) = \inf\{\lambda > 0;\, \mathrm{rk}(\Delta \cap B(0, \lambda)) \geq i\},$$

where $\mathrm{rk}(A)$ for $A \subset \Delta$ denotes the rank of the free abelian subgroup of Δ generated by A, and $B(0, \lambda)$ is the Euclidean ball of radius λ centered at the origin in $\mathbb{R}^d$.

Theorem 1 will be deduced from the following description of the asymptotic behavior of the successive minima along a diagonal orbit of the lattice $L\mathbb{Z}^d$, where L is a μ-generic point of M. Here, as in Theorem 1, $M = \phi(U)$ is the image of a connected open set in some Euclidean space $U \subset \mathbb{R}^n$ under an analytic map $\phi : U \to \mathrm{GL}_d(\mathbb{R})$, and μ is the push-forward under ϕ of the Lebesgue measure on U.

Theorem 3 (Strong parametric subspace theorem for manifolds). *Assume that the Zariski closure of M is defined over $\overline{\mathbb{Q}}$. Let $(a_t)_{t \geq 0}$ be a diagonal one-parameter semigroup in $\mathrm{GL}_d(\mathbb{R})$. Then there exist real numbers $\Lambda_1 \leq \cdots \leq \Lambda_d$ such that for μ-almost every $L \in M$ and each $k \in \{1, \ldots, d\}$,*

$$\lim_{t \to +\infty} \frac{1}{t} \log \lambda_k(a_t L\mathbb{Z}^d) = \Lambda_k. \tag{2}$$

Moreover, if $0 = d_0 < d_1 < \cdots < d_h = d$ in $\{0, \ldots, d\}$ are chosen so that

$$\Lambda_1 = \cdots = \Lambda_{d_1} < \Lambda_{d_1+1} = \cdots = \Lambda_{d_2} < \cdots < \Lambda_{d_{h-1}+1} = \cdots = \Lambda_d,$$

then there exist rational subspaces V_ℓ, $\ell = 0, \ldots, h$, in $\mathbb{Q}^d$ such that

- *$\dim V_\ell = d_\ell$ and $\{0\} = V_0 < V_1 < \cdots < V_h = \mathbb{Q}^d$,*
- *for μ-almost every $L \in M$ the first d_ℓ successive minima of $a_t L\mathbb{Z}^d$ are attained in V_ℓ provided t is large enough.*

In other words: for all $\varepsilon > 0$, for μ-almost every $L \in M$, there is $t_{L,\varepsilon} > 0$ such that for $t > t_{L,\varepsilon}$, $\ell = 1, \ldots, h$, and $\mathbf{x} \in \mathbb{Z}^d$,

$$\|a_t L\mathbf{x}\| \leq e^{t(\Lambda_{d_\ell} - \varepsilon)} \implies \mathbf{x} \in V_{\ell-1}. \tag{3}$$

In the proof of this result, we will make crucial use of the special case when M is reduced to a singleton. In that case the theorem appears,[1] albeit in a slightly different but equivalent language, in a paper of Evertse and Ferretti [14, Theorem 16.1]. It is a refinement of the parametric subspace theorem, often attributed to Faltings and Wüstholz [16, Theorem 9.1]. As shown below, the case when M is a singleton can also be obtained directly from Schmidt's subspace theorem in its original form. The two results are really equivalent. For convenience we will give a self-contained derivation of this special case and we note that our argument shares many similarities with that of [14, Sections 15 and 16].

---

[1]We thank the anonymous referee from whom we learned about this reference.

An important point to make is that (2) is a limit and not only a liminf or a limsup. This can be understood as saying that the diagonal orbit of a lattice defined over $\overline{\mathbb{Q}}$ has an *asymptotic shape* at infinity. Of course the λ_k can fluctuate, but only up to a small exponential error. This is of course in sharp contrast with what happens for certain specific values of L. Indeed, it is possible to construct matrices L for which the successive minima have an almost arbitrary behavior along a given diagonal orbit; see [32, Theorem 1.3] and [8, Theorem 2.2].

Corollary 1 (Parametric subspace theorem for manifolds). *Keep the same assumptions as in Theorem* 3. *Let* $a = (a_t)_{t \geq 0}$ *be a one-parameter diagonal semigroup in* $\mathrm{SL}_d(\mathbb{R})$. *There exists a proper subspace* $V(a)$ *of* $\mathbb{Q}^d$ *such that given* $\varepsilon > 0$, *for* μ-*almost every* $L \in M$, *there is* $t_{L,\varepsilon} > 0$ *such that if* $t > t_{L,\varepsilon}$ *and* $\mathbf{x} \in \mathbb{Z}^d$,

$$\|a_t L \mathbf{x}\| \leq e^{-\varepsilon t} \implies \mathbf{x} \in V(a).$$

Proof. Here we have assumed that a_t is unimodular. By Minkowski's theorem the product of all d successive minima of $a_t L \mathbb{Z}^d$ is bounded above and below independently of t. In view of (2), this implies that $\sum_{i=1}^{d} \Lambda_i = 0$. So $\Lambda_d \geq 0$. Hence we can take $V(a) = V_{h-1}$ in the notation of Theorem 3. ∎

The rational subspaces V_i appearing in Theorem 3 depend only on $(a_t)_{t \geq 0}$ and on the rational Zariski closure of M, namely the intersection of the closed algebraic subsets of GL_d defined over $\mathbb{Q}$ and containing M. This will be clear from the proof of Theorem 3 given below, where a more precise description of $V(a)$ and the V_i will be given. As we will see, the filtration $\{V_i\}_i$ is the *Harder–Narasimhan* filtration associated to M and $(a_t)_{t \geq 0}$, and the Λ_i are the slopes of the *Grayson polygon*. The next few subsections contain preparations towards the proof of Theorem 3 given at the end of this section.

1.2. Expansion rate and submodularity

In this subsection M is an arbitrary subset of $\mathrm{GL}_d(\mathbb{R})$ and $a = (a_t)_{t \geq 0}$ a one-parameter diagonal semigroup. We write $a_t = \mathrm{diag}(e^{A_1 t}, \ldots, e^{A_d t})$ for some real numbers $A_1, \ldots, A_d$. For a non-zero subspace $V \leq \mathbb{R}^d$ we define its *expansion rate* with respect to a by

$$\tau(V) := \lim_{t \to +\infty} \frac{1}{t} \log \|a_t \mathbf{v}\| \tag{4}$$

where $\mathbf{v}$ represents V in an exterior power $\bigwedge^k \mathbb{R}^d$. This quantity takes values in the finite set of eigenvalues of $\log a_1$ in exterior powers. More precisely,

$$\tau(V) = \max \{I(a); \ I \subset [1, d], \ |I| = k, \ \mathbf{v}_I \neq 0\} \tag{5}$$

where $I(a) = \sum_{i \in I} A_i$ and $\mathbf{v}_I$ is the coordinate of $\mathbf{v}$ in the basis $e_I = e_{i_1} \wedge \cdots \wedge e_{i_k}$ of $\bigwedge^k \mathbb{R}^d$, where $I = \{i_1, \ldots, i_k\}, k = \dim V$. By convention we will also set $\tau(\{0\}) = 0$. We leave it to the reader to check that if $A_1 \geq \cdots \geq A_d$, then

$$\tau(V) = I_V(a) \quad \text{with} \quad I_V := \{i \in [1, d]; \ V \cap \mathscr{F}_i > V \cap \mathscr{F}_{i+1}\} \tag{6}$$

where $\mathscr{F}_i = \langle e_i, \ldots, e_d \rangle$ and $(e_1, \ldots, e_d)$ is the canonical basis of $\mathbb{R}^d$.

For a subset $M \subset \mathrm{GL}_d(\mathbb{R})$, we set

$$\tau_M(V) := \max_{L \in M} \tau(LV). \qquad (7)$$

Similarly we see readily that

$$\tau_M(V) = \max\{I(a);\ I \subset [1, d],\ |I| = \dim V,\ (L\mathbf{v})_I \neq 0 \text{ for some } L \in M\}. \qquad (8)$$

From this formula it is clear that for all subspaces V,

$$\tau_M(V) = \tau_{\mathrm{Zar}(M)}(V)$$

where $\mathrm{Zar}(M)$ is the Zariski closure.

Lemma 1 (Submodularity of expansion rate). *Let $M \subset \mathrm{GL}_d(\mathbb{R})$ and assume that its Zariski closure is irreducible. Then the map $V \mapsto \tau_M(V)$ is submodular on the grassmannian, i.e. satisfies, for every pair of subspaces W_1, W_2,*

$$\tau_M(W_1) + \tau_M(W_2) \geq \tau_M(W_1 \cap W_2) + \tau_M(W_1 + W_2). \qquad (9)$$

Proof. Given a subspace W in $\mathbb{R}^d$, it is clear from (8) that the Zariski closure of the set $\{L \in M;\ \tau_L(W) < \tau_M(W)\}$ is a proper subset of $\mathrm{Zar}(M)$. By irreducibility, we may choose L in M such that $\tau_M = \tau_L$ on all four subspaces $W_1, W_2, W_1 \cap W_2$ and $W_1 + W_2$. It is therefore enough to prove the lemma for $M = \{L\}$. Now let $\mathbf{u}$ be a vector representing $U = W_1 \cap W_2$ in some exterior power of $\mathbb{R}^d$. Let also $\mathbf{w}_1'$ and $\mathbf{w}_2'$ be such that $\mathbf{u} \wedge \mathbf{w}_1'$ and $\mathbf{u} \wedge \mathbf{w}_2'$ represent W_1 and W_2, respectively. The subspace $W_1 + W_2$ is then represented by $\mathbf{u} \wedge \mathbf{w}_1' \wedge \mathbf{w}_2'$, and moreover, for every t,

$$\|a_t L(\mathbf{u} \wedge \mathbf{w}_1')\| \cdot \|a_t L(\mathbf{u} \wedge \mathbf{w}_2')\| \geq \|a_t L\mathbf{u}\| \cdot \|a_t L(\mathbf{u} \wedge \mathbf{w}_1' \wedge \mathbf{w}_2')\|.$$

Together with formula (4), this shows that τ_L is submodular. Note that this is compatible with the convention $\tau_M(\{0\}) = 0$. ∎

1.3. Harder–Narasimhan filtration

Submodular functions on partially ordered sets give rise to a "slope formalism" as in [6, 16, 17, 30, 40]. This is well known. In this subsection we recall the main facts we need and for the reader's convenience we give short proofs. The key to them is the following submodularity lemma, which in implicit form goes back at least to Stuhler [40] and Grayson [17] in the context of Euclidean sublattices and their covolume and which we rediscovered in [2] in the present context (subspaces and their expansion rate). Let k be a field and $\mathrm{Grass}(k^d)$ the grassmannian of non-zero subspaces of k^d. Let $\phi : \mathrm{Grass}(k^d) \to \mathbb{R}$ be a *submodular function*, that is, one satisfying (9) with ϕ in place of τ_M.

Lemma 2 (Submodularity lemma). *There is a subspace $V_\phi \in \mathrm{Grass}(k^d)$ achieving the infimum of $\frac{\phi(V) - \phi(0)}{\dim V}$, $V \in \mathrm{Grass}(k^d)$, and containing all other such subspaces.*

Proof. Let I be that infimum. Without loss of generality, up to changing ϕ to $\phi - \phi(0)$, we may assume that $\phi(0) = 0$. We begin by observing that ϕ is bounded below: if $(V_n)_n$ is

a sequence of distinct subspaces of maximal dimension with $\phi(V_n) \to -\infty$, pick a fixed line L with $L \not\subset V_n$ for infinitely many V_n; by submodularity $\inf_n \phi(V_n + L) = -\infty$, contradicting the maximality of $\dim V_n$. So I is finite. For $k \geq 1$, we set I_k to be the same infimum restricted to those subspaces V with $\dim V \geq k$. There is a maximal k_0 such that $I = I_{k_0}$. If $k_0 = d$, then we can take $V_\phi = k^d$ and there is nothing to prove. Otherwise let $\varepsilon > 0$ be such that $I_{k_0+1} > I + 2\varepsilon$. If a subspace W satisfies

$$\phi(W) \leq (I + \varepsilon) \dim W, \qquad (*)$$

then $\dim W \leq k_0$. By definition there is such a subspace with $\dim V_\phi \geq k_0$; call it V_ϕ. If Z is another subspace with $(*)$, then

$$\begin{aligned}
\phi(Z + V_\phi) &\leq \phi(Z) + \phi(V_\phi) - \phi(Z \cap V_\phi) \\
&\leq (I + \varepsilon)(\dim Z + \dim V_\phi) - I \dim(Z \cap V_\phi) \\
&\leq I \dim(Z + V_\phi) + 2\varepsilon \dim(Z + V_\phi),
\end{aligned}$$

which forces $\dim(Z + V_\phi) \leq k_0$, and hence $Z \leq V_\phi$, as desired. ∎

Definition 1 (Semistability). We say that k^d is *semistable* with respect to ϕ if $V_\phi = k^d$.

Definition 2 (Grayson polygon). Let $P_\phi : [0, d] \to \mathbb{R}$ be the convex piecewise linear function that is the supremum of all linear functions whose graph in $[0, d] \times \mathbb{R}$ lies below all points $(\dim V, \phi(V))$, $V \in \text{Grass}(k^d) \cup \{0\}$. Its graph is called the *Grayson polygon* of ϕ.
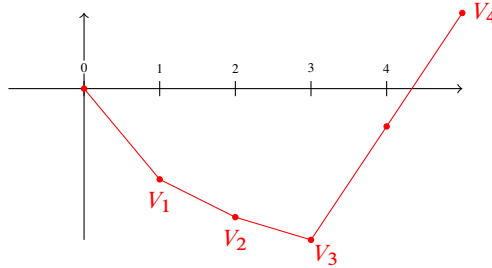

Fig. 1. A Grayson polygon.

Let (d_i, f_i), $i = 0, \ldots, h$, be the vertices of the Grayson polygon with $d_0 = 0$ and $d_h = d$, that is, the angular points, where the slope changes, i.e. for $i = 1, \ldots, h - 1$,

$$s_i < s_{i+1} \quad \text{where} \quad s_i := \frac{f_i - f_{i-1}}{d_i - d_{i-1}}.$$

The main result is the following:

Proposition 1 (Harder–Narasimhan filtration). *For each $i = 0, \ldots, h$, there is a unique k-subspace $V_i \leq k^d$ such that $\dim V_i = d_i$ and $\phi(V_i) = f_i$. Moreover, the subspaces V_i are nested, i.e. $0 = V_0 < V_1 < \cdots < V_h = k^d$, forming the so-called* Harder–Narasimhan filtration *of ϕ.*

In particular, we see that k^d is semistable if and only if its Harder–Narasimhan filtration is the trivial one, $\{0\} < k^d$.

Remark. Note that given any k-subspace $V \leq k^d$ the function $\phi_V(\overline{W}) := \phi(W) - \phi(V)$ defined on the quotient k^d/V, where $\overline{W} = W/V$ for any k-subspace W containing V, is submodular on $\mathrm{Grass}(k^d/V)$. It is clear from the proposition that V_i/V_{i-1} is semistable with respect to $\phi_{V_{i-1}}$ for $i = 1, \ldots, h$, and that $\{V_i/V_1\}_{i \geq 1}$ is the Harder–Narasimhan filtration of V/V_1 with respect to ϕ_{V_1}.

Proof of Proposition 1. The existence and uniqueness of V_1 is exactly what the submodularity lemma tells us. Suppose V_i has been defined. We may apply the submodularity lemma again to ϕ_{V_i} on the quotient k^d/V_i and thus obtain a subspace V_{i+1} containing V_i strictly and such that the function $(\phi(V) - \phi(V_i))/(\dim V - \dim V_i)$ reaches at V_{i+1} its unique minimum ℓ_{i+1} among subspaces V containing V_i. By construction, $\ell_i < \ell_{i+1}$ for each $i \geq 1$.

We now need to show that the Grayson polygon coincides with the polygon P drawn out of the points $(\dim V_i, \phi(V_i))$. In other words, we have to prove that if V is a subspace of k^d and i is such that $\dim V_i \leq \dim V < \dim V_{i+1}$, then

$$\phi(V) - \phi(V_i) \geq \ell_{i+1}(\dim V - \dim V_i). \tag{10}$$

If $V_i \leq V$, this is by definition of V_{i+1}. Otherwise $V \cap V_i < V_i$ and by induction we may assume that $(\dim(V \cap V_i), \phi(V \cap V_i))$ lies above P. So

$$\phi(V_i) - \phi(V \cap V_i) \leq \ell_i(\dim V_i - \dim(V \cap V_i)). \tag{11}$$

Moreover, again by definition of V_{i+1} we have

$$\phi(V_i + V) - \phi(V_i) \geq \ell_{i+1}(\dim(V_i + V) - \dim V_i). \tag{12}$$

On the other hand, ϕ is submodular, so $\phi(V_i + V) + \phi(V_i \cap V) \leq \phi(V_i) + \phi(V)$. Combining this with (11) and (12) we obtain

$$\phi(V) \geq \phi(V_i) + \ell_{i+1}(\dim(V) - \dim(V_i \cap V)) - \ell_i(\dim V_i - \dim(V \cap V_i)). \tag{13}$$

But $\ell_{i+1} \geq \ell_i$. So (10) follows.

This shows the existence of the V_i and the fact that they are nested. To see the uniqueness note that if $\dim V = \dim V_i$ and $\phi(V) = \phi(V_i)$ but $V \neq V_i$, then $\ell_i \geq \ell_{i+1}$ in view of (13), which is a contradiction. ∎

In the following, we apply this general theory by taking $k = \mathbb{Q}$ and $\phi(V)$ the expansion rate $\tau_M(V)$ defined in (7) on the grassmannian of rational subspaces. The above definition of semistability reads:

Definition 3. A non-zero rational subspace V in $\mathbb{R}^d$ will be called *M-semistable* with respect to $a = (a_t)_{t \geq 0}$ if for every rational subspace $W \leq V$,

$$\frac{\tau_M(W)}{\dim W} \geq \frac{\tau_M(V)}{\dim V}.$$

Similarly this yields the notions of Grayson polygon and Harder–Narasimhan filtration of M with respect to a.

Remark (unstable subspace). When $\det a = 1$, a subspace V with $\tau(V) < 0$ corresponds to a point $\mathbf{v}$ in some $\bigwedge^k \mathbb{R}^d$, which is *unstable with respect to a* in the terminology of geometric invariant theory, i.e. its a-orbit contains 0 in its closure. So $\mathbb{R}^d$ is M-semistable if and only if there are no unstable subspaces in the full grassmannian.

Next we make a remark about the dependence of the Harder–Narasimhan filtration on the choice of a one-parameter semigroup. It is easy to see that the Grayson polygon depends continuously on a. This is not so for the filtration, because new nodes can appear under small deformations, but the following lemma shows that the subspaces involved remain confined to a fixed finite family. Let $b(n)$ be the *ordered Bell number*, that is, the number of weak orderings (i.e. orderings with ties) on a set with n elements.

Lemma 3. *Let $M \subset \mathrm{GL}_d(\mathbb{R})$ with irreducible Zariski closure. There is a finite set S_M of rational subspaces of $\mathbb{R}^d$ with $|S_M| \leq b(2^d)$ such that, as $a = (a_t)_{t \geq 0}$ varies among all one-parameter diagonal semigroups of $\mathrm{GL}_d(\mathbb{R})$, the subspaces $V_i(a)$ arising in the Harder–Narasimhan filtration all belong to S_M.*

Proof. For $I \subset [d]$, let $\widehat{I}(a) = \frac{1}{|I|} \sum_{i \in I} A_i$, where $a_t = \mathrm{diag}(e^{A_1 t}, \ldots, e^{A_d t})$. We claim that the entire Harder–Narasimhan filtration of M depends on a only via the ordering of the various $\widehat{I}(a)$ for $I \subset [d]$: namely, if $\widehat{I}(a)$ and $\widehat{I}(a')$ define the same weak ordering on the family of subsets of $[d]$, then the filtrations coincide. To see the claim, note that every slope $(\tau_M(V) - \tau_M(W))/(\dim V - \dim W)$ for $W \leq V$ is equal to $\widehat{I}(a)$ for some I (because $I_W \subseteq I_V$ as follows from (6) and τ_M can be replaced by τ_L for some fixed L as in the proof of Lemma 1), and Proposition 1 tells us that $V_i(a)$ is defined as the unique solution to an extremal problem involving the comparison of slopes. So only their order matters. Since there are at most $b(2^d)$ possible orders, we are done. ∎

We also see from this proof that the slopes Λ_i are continuous and piecewise linear in $\log a$ and actually linear on each one of the cells cut out by the hyperplanes $\widehat{I}(a) = \widehat{I'}(a)$ for $I, I' \subset [1, d]$.

Remark. The ordered Bell number $b(n)$ grows superexponentially with n. This gives a rather poor bound on the number of exceptional subspaces in Theorem 1, especially in view of Schmidt's bound d^{2d^2} from [38]. A more refined argument, which we do not include here and which is based on the study of the set of permutations arising from the Schubert closure of M (see §1.8), allows one to improve this to $(2d)^d$.

1.4. Dynamics of diagonal flows

We now describe the dynamical ingredient of the proof. Using the quantitative non-divergence estimates (see Theorem 5 below), Kleinbock [19] showed the existence of a well-defined almost sure diophantine exponent for analytic manifolds. As described in

our previous work [2] with Menny Aka and Lior Rosenzweig, this holds for more general measures and the exponent actually depends only on the Zariski closure of the support of the measure, a property called *inheritance* in this paper, because the measure inherits its exponent from the Zariski closure of its support. We will need the following version of these results:

Theorem 4 (Inheritance principle). *Let $(a_t)_{t \geq 0}$ be a one-parameter diagonal semigroup in $\mathrm{GL}_d(\mathbb{R})$ and for $L \in \mathrm{GL}_d(\mathbb{R})$, set*

$$\mu_k(L) := \liminf_{t \to +\infty} \sum_{j \leq k} \frac{1}{t} \log \lambda_j(a_t L \mathbb{Z}^d).$$

Let $U \subset \mathbb{R}^n$ be a connected open set, $\phi : U \to \mathrm{GL}_d(\mathbb{R})$ an analytic map, and μ the image of the Lebesgue measure under ϕ. Let $M := \phi(U)$ and let $\mathrm{Zar}(M)$ be its Zariski closure in $\mathrm{GL}_d(\mathbb{R})$. Then for μ-almost every L in M and each $k = 1, \ldots, d$,

$$\mu_k(L) = \sup_{L' \in \mathrm{Zar}(M)} \mu_k(L').$$

Remark. When the Zariski closure of M is defined over $\overline{\mathbb{Q}}$, this inheritance principle, combined with Schmidt's subspace theorem, will allow us to show that the suprema in the above theorem are attained on some algebraic points of $\mathrm{Zar}(M)$. In fact, there exists $L \in \mathrm{Zar}(M) \cap \mathrm{GL}_d(\overline{\mathbb{Q}} \cap \mathbb{R})$ such that $\mu_k(L) = \sup_{L' \in \mathrm{Zar}(M)} \mu_k(L')$ for $k = 1, \ldots, d$, and in some sense, this is true for most elements $L \in \mathrm{Zar}(M) \cap \mathrm{GL}_d(\overline{\mathbb{Q}} \cap \mathbb{R})$. This will be of central importance in the proof of Theorem 3.

Proof of Theorem 4. A lemma of Mahler [28, Theorem 3], which is a simple consequence of Minkowski's second theorem in the geometry of numbers, asserts that $\lambda_1(\bigwedge^k g \mathbb{Z}^d)$ is comparable to $\prod_{j \leq k} \lambda_j(g \mathbb{Z}^d)$ to within multiplicative constants depending only on d. Since the k-th wedge representation $\rho_k : \mathrm{GL}(\mathbb{R}^d) \to \mathrm{GL}(\bigwedge^k \mathbb{R}^d)$ is an embedding of algebraic varieties, it maps $\mathrm{Zar}(M)$ onto $\mathrm{Zar}(\rho_k(M))$. These observations allow us to reduce the proof to the case where $k = 1$, which we now assume. To this end we first recall the quantitative non-divergence estimates in a form established in [18]:

Theorem 5 ([18, Theorem 2.2]). *Let $M = \phi(U)$ and μ be as in Theorem 4. There are $C, \alpha > 0$ such that the following holds. Let $\rho \in (0, 1]$ and $t > 0$, and let $B := B(z, r)$ be an open ball such that $B(z, 3^n r)$ is contained in U. Assume that for any $v_1, \ldots, v_k$ in $\mathbb{Z}^d$ with $\mathbf{w} := v_1 \wedge \cdots \wedge v_k \neq 0$,*

$$\sup_{y \in B} \|a_t \phi(y) \mathbf{w}\| > \rho^k.$$

Then for every $\varepsilon \in (0, \rho]$, we have

$$|\{x \in B; \lambda_1(a_t \phi(x) \mathbb{Z}^d) \leq \varepsilon\}| \leq C(\varepsilon/\rho)^{\alpha} |B|.$$

Given $\beta \in \mathbb{R}$, we say that a subset $S \subset \mathrm{GL}_d(\mathbb{R})$ satisfies *condition* ($\mathcal{C}_\beta$) if

$$\exists c > 0, \ \forall k \in \{1, \ldots, d\}, \ \forall t > 0, \ \forall \mathbf{w} = v_1 \wedge \cdots \wedge v_k \in \bigwedge^k \mathbb{Z}^d \setminus \{0\},$$

$$\sup_{g \in S} \|a_t g \mathbf{w}\| \geq c e^{k\beta t}. \qquad (\mathcal{C}_\beta)$$

And we set

$$\beta(S) := \sup \{\beta \in \mathbb{R}; \ S \text{ satisfies } (\mathcal{C}_\beta)\}. \qquad (14)$$

Note that by construction, if $S \subset S'$, then $\beta(S) \leq \beta(S')$.

Claim 1. *For all $L \in S$, we have $\beta(S) \geq \mu_1(L)$. If $S = \phi(B)$, where ϕ and B are as in Theorem 5, equality holds for μ-almost every $L \in \phi(B)$.*

Proof of claim. If $\beta > \beta(S)$, then ($\mathcal{C}_\beta$) fails. This implies that there exists t arbitrarily large such that $\sup_{g \in S} \|a_t g \mathbf{w}\| \leq e^{k\beta t}$ for some $\mathbf{w} \neq 0$. However by Minkowski's theorem applied to the sublattice represented by $a_t g \mathbf{w}$, this means that $\sup_{g \in S} \lambda_1(a_t g \mathbb{Z}^d) \leq e^{\beta t}$. Hence $\mu_1(L) \leq \beta$.

The opposite inequality for $S = \phi(B)$ will follow from the quantitative non-divergence estimate combined with Borel–Cantelli. Let $\beta < \beta(\phi(B))$. Then ($\mathcal{C}_\beta$) holds and, given $\delta > 0$, Theorem 5 applies with $\rho := c e^{\beta t}$ and $\varepsilon = e^{(\beta - \delta)t}$ so that

$$|\{x \in B; \ \lambda_1(a_t \phi(x) \mathbb{Z}^d) \leq e^{(\beta - \delta)t}\}| \leq C e^{-\alpha \delta t} |B|.$$

Summing this over all $t \in \mathbb{N}$, we obtain by Borel–Cantelli that for almost every $x \in B$, $\lambda_1(a_t \phi(x) \mathbb{Z}^d) \geq e^{(\beta - \delta)t}$ if t is a large enough integer. But this clearly implies that $\lambda_1(a_t \phi(x) \mathbb{Z}^d) \geq e^{(\beta - 2\delta)t}$ for all large enough $t > 0$. Hence $\mu_1(\phi(x)) \geq \beta - 2\delta$ for Lebesgue almost every $x \in B$. Since $\delta > 0$ is arbitrary, this proves the first claim.

Now we make the following key observation. For every bounded set $S \subset \mathrm{GL}_d(\mathbb{R})$ and compact set $K \supset S$,

$$\beta(S) = \beta(\mathrm{Zar}(S) \cap K) = \beta(\mathcal{H}(S) \cap K). \qquad (15)$$

Here $\mathcal{H}(S)$ is the preimage in $\mathrm{GL}_d(\mathbb{R})$ under ρ of the linear span $\mathcal{H}_S$ of all $\rho(g), g \in S$, where ρ is the linear representation with total space $E = \bigoplus_{k=1}^d \bigwedge^k \mathbb{R}^d$. This follows immediately from the following claim:

Claim 2. *There is $C = C(S) > 0$ such that for all $\mathbf{w}$ and t we have*

$$\sup_{g \in \mathcal{H}(S) \cap K} \|a_t g \mathbf{w}\| \leq C \sup_{g \in S} \|a_t g \mathbf{w}\|. \qquad (16)$$

Proof of claim. We note that $\mathcal{H}_S = \mathcal{H}_{\mathrm{Zar}(S)} = \mathcal{H}_{\mathrm{Zar}(S) \cap K}$, because S and $\mathrm{Zar}(S) \cap K$ have the same Zariski closure. Now consider the space $\mathcal{L}(\mathcal{H}_S, E)$ of linear maps from $\mathcal{H}_S$ to E. If $X \subset \mathcal{H}_S$ is a bounded set that spans $\mathcal{H}_S$, then the quantity $L \mapsto \sup_{A \in X} \|L(A)\|$

defines a norm $| \cdot |_X$ on $\mathcal{L}(\mathcal{H}_S, E)$. Therefore for any two such sets X, X' there is a constant $C_{X',X} > 0$ such that for all $L \in \mathcal{L}(\mathcal{H}_S, \bigwedge^k \mathbb{R}^d)$ we have

$$|L|_{X'} \leq C_{X',X} |L|_X.$$

This applies in particular to the sets $X := \rho(S)$ and $X' := \rho(\mathrm{Zar}(S) \cap K)$. Now (16) follows by setting $L : A \mapsto a_t A \mathbf{w}$, an element of $\mathcal{L}(\mathcal{H}_S, E)$. This ends the proof of Claim 2.

We may now finish the proof of Theorem 4. Since M is connected, it follows from Claim 1 that the μ-almost sure value of $\mu_1(L)$ for $L \in M$ is unique and well-defined and equals $\beta(\phi(B))$ for any ball B as in Theorem 5. It is also equal to $\sup_{L \in M} \mu_1(L)$ by the first part of the claim. However, since ϕ is analytic, the Zariski closure of $\phi(B)$ coincides with $\mathrm{Zar}(M)$. So (15) entails $\beta(\phi(B)) = \beta(\mathcal{H}(M) \cap K)$ for any compact $K \supset \phi(B)$. But Claim 1 applied to $S = \mathrm{Zar}(M) \cap K$ implies that $\mu_1(L) \leq \beta(\mathrm{Zar}(M) \cap K)$ for all $L \in \mathrm{Zar}(M) \cap K$. Since K is arbitrary, we get $\sup_{L \in \mathrm{Zar}(M)} \mu_1(L) \leq \beta(\mathrm{Zar}(M) \cap K)$. The right-hand side is the μ-almost sure value of $\mu_1(L)$, so this inequality is an equality. This ends the proof of Theorem 4. ∎

1.5. Proof of Theorem 3

Without loss of generality we may assume that $A_1 \geq \cdots \geq A_d$, where $A = \mathrm{diag}(A_1, \ldots, A_d)$ and $a_t = \exp(tA)$. Let $0 = V_0 < V_1 < \cdots < V_h = \mathbb{Q}^d$ be the Harder–Narasimhan filtration associated to the submodular function τ_M on the grassmannian of $\mathbb{Q}^d$. Let $d_\ell = \dim V_\ell$ and

$$\Lambda_k = \frac{\tau_M(V_\ell) - \tau_M(V_{\ell-1})}{d_\ell - d_{\ell-1}} \quad \text{if } d_{\ell-1} < k \leq d_\ell.$$

We need to show that for each $\ell = 1, \ldots, h$ and for μ-almost every L the limit (2) holds when $d_{\ell-1} < k \leq d_\ell$ and that for t large enough the first $d_{\ell-1}$ successive minima of $a_t L \mathbb{Z}^d$ are attained in $V_{\ell-1}$. Suppose this has been proved for all $\ell < i$ and let us prove it for $\ell = i$.

By Minkowski's second theorem, for every $L \in M$,

$$\limsup_{t \to +\infty} \frac{1}{t} \sum_{k \leq d_i} \log \lambda_k(a_t L \mathbb{Z}^d) \leq \tau_L(V_i) \leq \tau_M(V_i).$$

On the other hand, we already know that for μ-almost every L,

$$\lim_{t \to +\infty} \frac{1}{t} \sum_{k \leq d_{i-1}} \log \lambda_k(a_t L \mathbb{Z}^d) = \sum_{k \leq d_{i-1}} \Lambda_k = \tau_M(V_{i-1}). \tag{17}$$

Hence

$$\limsup_{t \to +\infty} \frac{1}{t} \sum_{d_{i-1} < k \leq d_i} \log \lambda_k(a_t L \mathbb{Z}^d) \leq \tau_M(V_i) - \tau_M(V_{i-1}).$$

Therefore to prove (2) it suffices to show that for μ-almost every L,

$$\liminf_{t\to+\infty} \frac{1}{t} \log \lambda_{d_{i-1}+1}(a_t L \mathbb{Z}^d) \geq \frac{\tau_M(V_i) - \tau_M(V_{i-1})}{d_i - d_{i-1}} = \Lambda_{d_i}. \tag{18}$$

This will also prove (3) for $\ell = i$ as we now explain. By Minkowski's theorem,

$$\lim_{t\to+\infty} \frac{1}{t} \sum_{k \leq d_{i-1}} \log \lambda_k(a_t L V_{i-1}(\mathbb{Z})) = \tau_L(V_{i-1}) \leq \tau_M(V_{i-1}).$$

In view of (17) this quantity is actually equal to $\tau_M(V_{i-1})$. Therefore the d_{i-1} first minima of $a_t L \mathbb{Z}^d$ are attained in V_{i-1}, and (3) follows from (18). We now establish (18) separating two cases.

First case: $M = \{L\}$ with $L \in \mathrm{GL}_d(\bar{\mathbb{Q}})$. Fixing $\varepsilon > 0$, we want to show that for $t > 0$ large enough, any integer solution $v \in \mathbb{Z}^d$ to

$$\|a_t L v\| \leq e^{t(\Lambda_{d_i} - \varepsilon)} \tag{19}$$

satisfies $v \in V_{i-1}$. Argue by contradiction and suppose that for arbitrarily large $t > 0$ there are integer solutions v to (19) with $v \notin V_{i-1}$. Let $V \supset V_{i-1}$ be a subspace of $\mathbb{Q}^d$ of minimal dimension containing such solutions.

Consider the linear forms $L_1, \ldots, L_d$ on $\mathbb{R}^d$ given by the rows of L. They are linearly independent and have coefficients in $\bar{\mathbb{Q}}$. Note that (19) implies

$$\forall k \in \{1, \ldots, d\}, \quad e^{tA_k} |L_k(v)| \leq e^{t(\Lambda_{d_i} - \varepsilon)}. \tag{20}$$

Recall the subsets of indices I_V and $I_{V_{i-1}}$ defined as in (6). For $k \notin I_{V_{i-1}}$ the restriction $L_k|_{V_{i-1}}$ of L_k to V_{i-1} lies in the span of the $L_\ell|_{V_{i-1}}$ for $\ell < k$, $\ell \in I_{V_{i-1}}$. Since the linear forms $\{L_\ell|_{V_{i-1}}\}_{\ell \in I_{V_{i-1}}}$ are linearly independent, there is a unique choice of scalars $\alpha_{k,\ell} \in \bar{\mathbb{Q}} \cap \mathbb{R}$ such that $L_k|_{V_{i-1}} = \sum_{\ell \in I_{V_{i-1}}, \ell < k} \alpha_{k,\ell} L_\ell|_{V_{i-1}}$. We define

$$M_k := L_k - \sum_{\ell \in I_{V_{i-1}}, \ell < k} \alpha_{k,\ell} L_\ell.$$

By construction M_k vanishes on V_{i-1} and induces a linear form $\bar{M}_k$ on the quotient V/V_{i-1}. Also by construction, the linear forms $\{\bar{M}_k\}_{k \in I_V \setminus I_{V_{i-1}}}$ have coefficients in $\bar{\mathbb{Q}} \cap \mathbb{R}$ and are linearly independent on V/V_{i-1}.

It follows from (20) that

$$\forall k \notin I_{V_{i-1}}, \quad e^{tA_k} |\bar{M}_k(\bar{v})| \ll_L e^{t(\Lambda_{d_i} - \varepsilon)}. \tag{21}$$

where $\bar{v} = v \mod V_{i-1}$. But $\sum_{k \in I_V \setminus I_{V_{i-1}}} A_k = \tau_L(V) - \tau_L(V_{i-1})$, so

$$\left| \prod_{k \in I_V \setminus I_{V_{i-1}}} \bar{M}_k(\bar{v}) \right| \ll_L e^{-(\dim V - d_{i-1})t\varepsilon},$$

because by definition of V_i and Λ_{d_i} we know that $\Lambda_{d_i} \leq \frac{\tau_L(V) - \tau_L(V_{i-1})}{\dim V - d_{i-1}}$. On the other

hand, $|\Lambda_{d_i}| \leq |A| := \max |A_k|$, and in view of (19) we have $\|\bar{v}\| \leq \|v\| \ll_L e^{2|A|t}$. So setting $\varepsilon' = \varepsilon/(2d|A|)$, we obtain

$$\left| \prod_{k \in I_V \setminus I_{V_{i-1}}} \bar{M}_k(\bar{v}) \right| \ll_L \|v\|^{-\varepsilon'}. \tag{22}$$

We are thus in a position to apply Schmidt's subspace theorem [37, Theorem 1F] and conclude that all but finitely many integer solutions to this inequality lie in a finite union of proper rational subspaces of V/V_{i-1}. This contradicts the minimality of V.

General case: Note that $\mathrm{Zar}(M)$ is an irreducible algebraic variety, so τ_M is submodular. In view of (5), given any rational subspace V in $\mathbb{R}^d$ the set of elements L in M such that $\tau_L(V) \neq \tau_M(V)$ is a proper closed subvariety of $\mathrm{Zar}(M)$ that is defined over $\mathbb{Q}$ and of bounded degree. By Lemma 4 below (applied to $F = \mathbb{R} \cap \bar{\mathbb{Q}}$) we may choose a point L_0 of $\mathrm{Zar}(M)$ in $\mathrm{GL}_d(\bar{\mathbb{Q}})$ such that $\tau_{L_0}(V) = \tau_M(V)$ for all rational V. Then the Harder–Narasimhan filtrations and the Grayson polygons of M and $\{L_0\}$ coincide. By the first part of the proof,

$$\liminf_{t \to +\infty} \frac{1}{t} \log \lambda_{d_{i-1}+1}(a_t L_0 \mathbb{Z}^d) \geq \Lambda_{d_i}. \tag{23}$$

Now we may invoke Theorem 4. In view of (23) and (17) this implies (18).

Lemma 4 (Countable unions of proper subvarieties). *Let X be an irreducible algebraic variety defined over a number field K. Let F be an algebraic extension of K of infinite degree. Let $k \geq 1$ and suppose $(X_j)_{j \geq 1}$ is a countable family of proper closed subvarieties of X of degree at most k, each defined over a field K_j of degree at most k over K. Then $X(F)$ is not contained in the union of the $X_j(\bar{\mathbb{Q}})$, $j \geq 1$.*

Proof. Looking at a finite cover by affine varieties, we may assume that X is affine. Then by Noether's normalization theorem, there is a finite morphism $f : X \to \mathbb{A}^d$ defined over $\bar{\mathbb{Q}}$, where $d = \dim X$. So again without loss of generality, we may assume that $X = \mathbb{A}^d$, the d-dimensional affine space. But we can of course find elements $x_i \in F$, $i = 1, \ldots, d$, such that x_i has degree $> k^2$ over $K(x_1, \ldots, x_{i-1})$. Then $K_j(x_1, \ldots, x_i)$ has degree $> k$ over $K_j(x_1, \ldots, x_{i-1})$ for all j, so $(x_1, \ldots, x_d)$ will not belong to any X_j. ■

1.6. The subspace theorem

We are now ready for the proof of Theorem 1.

Proof of Theorem 1. Without loss of generality, we may assume that M is a bounded subset of $\mathrm{GL}_d(\mathbb{R})$. In particular, there is $C = C(M) \geq 1$ such that $\|L\mathbf{x}\| \leq C \|\mathbf{x}\|$ for every $\mathbf{x} \in \mathbb{R}^d$ and $L \in M$. We are going to show that there is a finite set P with $|P| \leq (2Cd^2\varepsilon^{-1})^d$ of one-parameter unimodular diagonal semigroups $a = (a_t)_{t \geq 0}$ with the following property. If $L \in M$ and $\mathbf{x} \in \mathbb{Z}^d$ is a solution to (1) such that the integer part t of $\frac{\varepsilon}{4d} \log \|\mathbf{x}\|$ is at least 1, then there is $a \in P$ such that

$$\|a_t L\mathbf{x}\| \leq d e^{-t}. \tag{24}$$

To see this, let $\ell_i = \log |L_i(\mathbf{x})|$. By (1) we have $\ell_1 + \cdots + \ell_d \leq -4dt$. Note that $\ell_i \leq \log \|x\| + \log C \leq (8d/\varepsilon + \log C)t$. Let $\ell_i' = \max \{\ell_i, -Dt\}$, where $D = 5d(\log C + 8d/\varepsilon)$. If there is an index i such that $\ell_i \neq \ell_i'$, then $\ell_1' + \cdots + \ell_d' \leq -Dt + (d-1)(D/5d)t \leq -4dt$. We conclude that $\ell_1' + \cdots + \ell_d' \leq -4dt$ always. Now define $b_i := \frac{1}{d}(\ell_1' + \cdots + \ell_d') - \ell_i'$. Then $b_1 + \cdots + b_d = 0$, and for each i,

$$\ell_i + b_i \leq \ell_i' + b_i \leq -4t.$$

On the other hand, $|\ell_i'| \leq Dt$, so $|b_i| \leq 2Dt$. Let B be set of integers points in $\mathbb{Z}^d$ with coordinates in $[-3D, 3D]$. Choose $(n_1, \ldots, n_d) \in B$ such that $|b_i/t - n_i| \leq 1/2$ for all i. In particular, $|\sum_{n=1}^{d} n_i| \leq d/2$. Changing some n_i to the next or previous integer if needed, we may ensure that $\sum_{n=1}^{d} n_i = 0$ and $|b_i/t - n_i| \leq 3/2$. Then we set $a_t = \mathrm{diag}(e^{n_1 t}, \ldots, e^{n_d t})$ and let P be the finite set of such diagonal semigroups. Note that $|P| \leq (6D)^d$. Clearly $\ell_i + n_i t \leq -t$ and (24) follows.

Now we may apply Corollary 1 and conclude that for μ-almost every $L \in M$, if $\mathbf{x}$ is a large enough solution of (1), it must lie inside $V(a)$ for some a in P. This shows that the number of exceptional rational subspaces is finite. However, by Lemma 3 above the subspace $V(a)$ can take at most $b(2^d)$ possible values as a varies among all unimodular diagonal semigroups. This ends the proof. ∎

Remark. Note that conversely, each $V(a)$, and hence each V_i in Theorem 1, contains infinitely many solutions to (1) for every $\varepsilon > 0$.

Remark. Furthermore, the rational subspaces $V_1 \cup \cdots \cup V_r$ depend only on the rational Zariski closure of M and not on the choice of L. And because they are defined by a simple slope condition, their height is effectively bounded in terms of the height of $\mathrm{Zar}(M)$. On the other hand, the finite set of exceptional solutions lying outside the V_i depends on L and ε and there is no known bound on their height or number; see [12, Proposition 5.1]. When M is a single point it is however possible to group together the finitely many exceptional solutions into another set of proper subspaces whose number, but not height, can be effectively bounded; see [11].

1.7. Varieties defined over $\mathbb{R}$

What happens if we remove the assumption in Theorem 3 that the Zariski closure of M is defined over $\overline{\mathbb{Q}}$? Without this assumption, diagonal flow trajectories may not behave as nicely and typically no limit shape is to be expected. However, we may give a simple upper and lower bound on the almost sure value of $\mu_k(L)$ for $k = 1, \ldots, d$, which exists by Theorem 4, in terms of the rational and real Grayson polygons, as we now discuss. So far we have only considered the rational Harder–Narasimhan filtration $\{V_i^{\mathbb{Q}}\}_{i=0}^{h_{\mathbb{Q}}}$ and its rational polygon $\mathscr{G}^{\mathbb{Q}}$ with slopes $s_i^{\mathbb{Q}}$, because we have restricted ourselves to considering the grassmannian of rational subspaces. But we may also take $k = \mathbb{R}$ in §1.3 with the same submodular function τ_M. This yields a new Harder–Narasimham filtration $\{V_i^{\mathbb{R}}\}_{i=0}^{h_{\mathbb{R}}}$ for the real field and a new Grayson polygon $\mathscr{G}^{\mathbb{R}}$ with slopes $s_i^{\mathbb{R}}$, which obviously lies *below* the rational polygon.

Let as before $M = \phi(U)$ be the image of a connected open set $U \subset \mathbb{R}^n$ under an analytic map $\phi : U \to \mathrm{GL}_d(\mathbb{R})$ and μ the measure on M that is the image of the Lebesgue measure on U. In this subsection we no longer assume that $\mathrm{Zar}(M)$ is defined over $\overline{\mathbb{Q}}$. Let μ_k be the μ-almost sure value of $\mu_k(L)$ as given by Theorem 4 and μ_k^{sup} the supremum over all $L \in M$ of $\mu_k^{\mathrm{sup}}(L)$, where $\mu_k^{\mathrm{sup}}(L)$ is defined by the same formula as $\mu_k(L)$ with a limsup in place of the liminf. Consider the points (k, μ_k) for $k = 1, \ldots, d$ and interpolate linearly between them, so as to form a polygon $\mathscr{G}_\mu$ as in Fig. 1. Similarly form $\mathscr{G}_\mu^{\mathrm{sup}}$ with $(k, \mu_k^{\mathrm{sup}})$. Note that $\mathscr{G}_\mu^{\mathrm{sup}}$ is convex (being a supremum of convex polygons), but $\mathscr{G}_\mu$ may not be.

Proposition 2 ("Sandwich theorem"). *The polygons $\mathscr{G}_\mu, \mathscr{G}_\mu^{\mathrm{sup}}$ lie between the rational Grayson polygon $\mathscr{G}^{\mathbb{Q}}$ and the real Grayson polygon $\mathscr{G}^{\mathbb{R}}$. In other words, for each $k = 1, \ldots, d$,*

$$\sum_{i \le k} s_{m_i}^{\mathbb{R}} \le \mu_k \le \mu_k^{\mathrm{sup}} \le \sum_{i \le k} s_{n_i}^{\mathbb{Q}},$$

where m_i, n_i defined by $\dim V_{m_i-1}^{\mathbb{R}} \le i < \dim V_{m_i}^{\mathbb{R}}$ and $\dim V_{n_i-1}^{\mathbb{Q}} \le i < \dim V_{n_i}^{\mathbb{Q}}$.
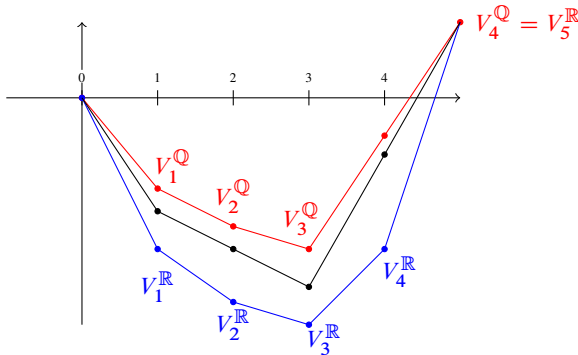

Fig. 2. $\mathscr{G}_\mu$ lies between the rational and the real polygons.

Proof. The upper bound follows from Minkowski's theorem: the first d_i successive minima in $a_t L \mathbb{Z}^d$ are smaller than those attained in $V_i^{\mathbb{Q}}$, so

$$\mu_k(L) \le \limsup_{t \to +\infty} \frac{1}{t} \sum_{k \le d_i} \log \lambda_k(a_t L \mathbb{Z}^d) \le \tau_L(V_i^{\mathbb{Q}}) \le \tau_M(V_i^{\mathbb{Q}}).$$

Since this holds for each $i = 0, \ldots, h_{\mathbb{Q}}$, and the polygons are convex, we conclude that $\mathscr{G}_\mu^{\mathrm{sup}}$ lies below $\mathscr{G}^{\mathbb{Q}}$.

The lower bound follows from a modified version of the proof of Theorem 4 that uses instead the refined quantitative non-divergence estimates for successive minima already mentioned and established in [33, Chapter 6] or [27, Theorem 5.3]. We now explain this briefly and refer to [33, Theorem 7.3.1] for the details of a more general statement. Let B be a small ball around some $x \in U$. Note that $\phi(B)$ is Zariski dense in M and thus

$\tau_M = \tau_{\phi(B)}$. Let β_k be the infimum of all $\tau_M(W)$, where W ranges among real subspaces with dimension k. In view of (8) this implies that there is I with $I(a) \geq \beta_k$ such that $(L\mathbf{w})_I \neq 0$ for some $L \in M$. By compactness of the grassmannian of k-dimensional subspaces there is $c > 0$ such that $\max_{I, I(a) \geq \beta_k} \sup_{L \in M} \|(L\mathbf{w})_I\| \geq c\|\mathbf{w}\|$ for all W. It follows in particular that $\sup_{L \in M} \|a_t L\mathbf{w}\| \gg e^{t\beta_k}$ uniformly in $t > 0$ and in non-zero integer-valued $\mathbf{w}$. Now let $\gamma_k \leq \beta_k$ be the largest such function of k that is convex in k. Then the exact same Borel–Cantelli argument used in the proof of Claim 1 in Theorem 4, using instead the refined quantitative non-divergence in the form of [27, Theorem 5.3], shows that for μ-almost every L in B, $\mu_k(L) \geq \gamma_k$. Since $\beta_{d_i} = \tau_M(V_i^{\mathbb{R}})$, we have shown that $\mathscr{G}_\mu$ lies above $\mathscr{G}^{\mathbb{R}}$.   ∎

This proposition also holds for measures μ on $\mathrm{GL}_d(\mathbb{R})$ which are good in the sense of §1.9 below; see [33, Theorem 7.3.1].

If $\mathrm{Zar}(M)$ is defined over $\mathbb{Q}$, then by uniqueness of the Harder–Narasimhan filtration, we see that the real and rational filtrations coincide. In particular, in this case all three polygons coincide. If $\mathrm{Zar}(M)$ is defined over $\bar{\mathbb{Q}}$, then the filtration over $\mathbb{R}$ is in fact defined over $\bar{\mathbb{Q}}$, and Theorem 3 asserts that $\mathscr{G}_\mu$ coincides with $\mathscr{G}^{\mathbb{Q}}$. However, $\mathscr{G}^{\mathbb{R}}$ may be different.

Similarly:

Corollary 2. *Let M be as in Theorem 1, except we no longer assume that $\mathrm{Zar}(M)$ is defined over $\bar{\mathbb{Q}}$. Let $a = (a_t = \exp(tA))_t$ be a diagonal flow. Assume that $\mathbb{R}^d$ is M-semistable with respect to a (see Def. 3). Then for μ-almost every $L \in M$,*

$$\liminf_{t \to +\infty} \lambda_1(a_t L\mathbb{Z}^d) \geq \frac{1}{d} \operatorname{Tr}(A).$$

1.8. Plücker closure

In this subsection we define the notion of Plücker closure of a subset $M \subset \mathrm{GL}_d(\mathbb{R})$ and we explain why in Theorems 1 and 3 instead of assuming that $\mathrm{Zar}(M)$ is defined over $\bar{\mathbb{Q}}$, it is enough to assume that the Plücker closure of M is defined over $\bar{\mathbb{Q}}$.

Let $M \subset \mathrm{GL}_d(\mathbb{R})$ be a subset and (E, ρ) be the direct sum of the exterior power representations of GL_d, namely $E = \bigoplus_{k=1}^d \bigwedge^k \mathbb{R}^d$. We denote by $\mathcal{H}_M$ the $\mathbb{R}$-linear span in $\mathrm{End}(E)$ of all $\rho(g)$, $g \in M$. We further define the *Plücker closure* $\mathcal{H}(M)$ of M as the inverse image of $\mathcal{H}_M$ under ρ in $\mathrm{GL}_d(\mathbb{R})$. Note that $\mathcal{H}(M)$ contains the Zariski closure $\mathrm{Zar}(M)$ of M. We say that $\mathcal{H}(M)$ is defined over $\bar{\mathbb{Q}}$ if $\mathcal{H}_M$ has a basis with coefficients in $\bar{\mathbb{Q}}$ in the canonical basis of E. An obvious sufficient condition for this to hold is to ask for $\mathrm{Zar}(M)$ to be defined over $\bar{\mathbb{Q}}$.

We also say that M is *Plücker irreducible* if $\rho(M)$ is not contained in a finite union of proper subspaces of $\mathcal{H}_M$ in $\mathrm{End}(E)$. Clearly Zariski irreducibility implies Plücker irreducibility. It is clear that Plücker irreducibility of M is enough to guarantee that τ_M is submodular by the argument of Lemma 1. It is also clear that $\tau_M = \tau_{\mathcal{H}(M)}$.

For simplicity of exposition, we have chosen to state the assumptions in our main theorems in terms of the Zariski closure of M, but in fact Theorems 1 and 3 hold assuming

only that the Plücker closure is defined over $\overline{\mathbb{Q}}$. The proof is verbatim the same, except that in Theorem 4 we get the following slightly stronger statement: for μ-almost every L in M and each k,

$$\mu_k(L) = \sup_{L' \in \mathcal{H}(M)} \mu_k(L').$$

Again the proof of this equality is exactly the one of Theorem 4 when $k = 1$. However, the reduction to the case $k = 1$ via Mahler's lemma no longer works here, because $\mathcal{H}(M)$ may differ from $\mathcal{H}(\rho(M))$. Instead one may use the enhanced version of the quantitative non-divergence estimates already mentioned in the proof of Proposition 2, namely [33, Chapter 6] or [27, Theorem 5.3] in place of Theorem 5, which enables one to run the argument simultaneously for all k. This shows that the almost sure value of each $\mu_k(L)$, and thus the polygon $\mathcal{G}_\mu$ defined in the previous subsection, depend only on the Plücker closure of M.

There is another notion of envelope of M that is also natural to consider, namely the intersection $\mathrm{Sch}(M)$ of all translates $\mathcal{S}_\sigma g$ of Schubert varieties $\mathcal{S}_\sigma = \overline{B\sigma B}$ containing M. Here B is one of the $d!$ Borel subgroups containing the diagonal subgroup and the closure is the Zariski closure. This *Schubert closure* contains the Plücker closure. It is easy to see from (6) that the submodular function τ_M depends only on $\mathrm{Sch}(M)$. Restricting to rational translates $\mathcal{S}_\sigma g$ with $g \in \mathrm{GL}_d(\mathbb{Q})$ one obtains the *rational Schubert closure* $\mathrm{Sch}^{\mathbb{Q}}(M)$. The asymptotic shape and the exceptional subspaces appearing in Theorems 1 and 3 depend only on $\mathrm{Sch}^{\mathbb{Q}}(M)$. A natural question we could not answer is whether or not the main theorem remains valid under the (weaker) assumption that $\mathrm{Sch}(M)$ is defined over $\overline{\mathbb{Q}}$. The answer is clearly yes when $\mathrm{Sch}(M)$ is defined over $\mathbb{Q}$ as follows readily from Proposition 2.

1.9. More general measures

In this subsection, we define a class of measures called *good measures*, which is wider than the family considered so far of push-forwards of the Lebesgue measure under analytic maps, and for which Theorems 1, 3 and 4 continue to hold. This class is very closely related to the so-called friendly measures of [20, 24]: instead of being expressed in terms of an affine span, the non-degeneracy condition is defined in terms of local Plücker closures.

First we need to recall some terminology. We fix a metric on $\mathrm{GL}_d(\mathbb{R})$, say induced from the euclidean metric on the matrices $M_d(\mathbb{R})$. Given two positive parameters C and α, a real-valued function f on the support of μ is called (C, α)-*good with respect to μ* if for any ball B in $\mathrm{GL}_d(\mathbb{R})$ and all $\varepsilon > 0$,

$$\mu(\{x \in B; \ |f(x)| \le \varepsilon \|f\|_{\mu,B}\}) \le C\varepsilon^\alpha \mu(B),$$

where $\|f\|_{\mu,B} = \sup_{x \in B \cap \mathrm{supp}\,\mu} |f(x)|$. The measure μ is *doubling* on a subset $X \subset \mathrm{GL}_d(\mathbb{R})$ if there exists a constant C' such that for every ball $B(x, r) \subset X$,

$$\mu(B(x, 2r)) \le C'\mu(B(x, r)).$$

Then we say that a Borel measure μ is *locally good* at $L \in \mathrm{GL}_d(\mathbb{R})$ if there exists a ball B around L and positive constants C, α such that

(i) the measure μ is doubling on B;

(ii) for every $k \in \{1, \ldots, d\}$, for every pure k-vector $\mathbf{w} = \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k$ in $\bigwedge^k \mathbb{R}^d$, and every $a \in \mathrm{GL}_d(\mathbb{R})$, the map $y \mapsto \|ay \cdot \mathbf{w}\|$ is (C, α)-good on B with respect to μ.

Recall next that given a subset $S \subset \mathrm{GL}_d(\mathbb{R})$ and a point $x \in S$, the *local Plücker closure* $\mathcal{H}_x(S)$ of S at x is the intersection over $r > 0$ of the Plücker closures of $S \cap B(x, r)$:

$$\mathcal{H}_x(S) = \bigcap_{r>0} \mathcal{H}(S \cap B(x, r)).$$

We may now define the class of good measures.

Definition 4. A locally finite Borel measure μ on $\mathrm{GL}_d(\mathbb{R})$ will be called a *good measure* if $\mathrm{supp}\,\mu$ is Plücker irreducible and for μ-almost every L,

(1) the measure μ is locally good at L;

(2) the local Plücker closure of $\mathrm{supp}\,\mu$ at L is equal to that of $\mathrm{supp}\,\mu$.

Of course, the most important example of a good measure is that of the push-forward of the Lebesgue measure under an analytic map.

Proposition 3 (Analytic measures are good). *Let $n \in \mathbb{N}$, let U be a connected open set in $\mathbb{R}^n$, and let $\varphi : U \to \mathrm{GL}_d(\mathbb{R})$ be a real-analytic map. Then the push-forward under φ of the Lebesgue measure on U is a good measure on the Zariski closure M of $\varphi(U)$.*

Proof. Note that the maps of the form $u \mapsto \|a\varphi(u) \cdot \mathbf{w}\|^2$ are linear combinations of products of matrix coefficients of $\varphi(u)$ and hence belong to a finite-dimensional linear subspace of analytic functions on U depending only on φ. So [19, Proposition 2.1] applies. ∎

Theorem 4 holds for all good measures μ on $\mathrm{GL}_d(\mathbb{R})$ with the same proof (suitably modified via the enhanced quantitative non-divergence estimates as mentioned in the previous subsection). In fact, the definition of good measures has been tailored precisely for Theorem 4 to hold. Thus the subspace theorem for manifolds and the parametric subspace theorem, Theorems 1 and 3, hold for good measures μ such that the Plücker closure of $\mathrm{supp}\,\mu$ is defined over $\overline{\mathbb{Q}}$.

2. A generalization to number fields

Schmidt himself observed in [36] that his theorem for $\mathbb{Q}$ implied a more general version for any number field K, and this was generalized shortly after by Schlickewei [34], who gave a statement allowing also finite places. In this section, we formulate a similar generalization of Theorems 1 and 3.

For a place v of a number field K, the completion of K at v is denoted by K_v. As in Bombieri–Gubler [5, §§1.3–1.4] we use the following normalization for the absolute value $| \cdot |_v$ on K_v.

$$|x|_v := N_{K_v/\mathbb{Q}_v}(x)^{1/[K:\mathbb{Q}]},$$

where $\mathbb{Q}_v$ is the completion of $\mathbb{Q}$ at the place v restricted to $\mathbb{Q}$ and $N_{K_v/\mathbb{Q}_v}(x)$ is the norm of x in the extension $K_v/\mathbb{Q}_v$. The product formula then reads $\prod_v |x|_v = 1$ for all $x \in K$. If S is a finite set of places of K containing all archimedean ones, the ring $\mathcal{O}_{K,S} \subset K$ of S-integers is the set of $x \in K$ such that $|x|_v \leq 1$ for all places v lying outside S. Elements of its group of units are called S-units.

Let d be a positive integer. The d-dimensional space K_v^d will be endowed with the supremum norm $\| \cdot \|_v$, i.e. for an element $\mathbf{x} = (x_1^{(v)}, \ldots, x_d^{(v)})$ in K_v^d, $\|\mathbf{x}\|_v = \max_{1 \leq i \leq d} |x_i^{(v)}|_v$. More generally, we let $K_S = \prod_{v \in S} K_v$ be the product of all completions of K at the places of S, and if $\mathbf{x} = (\mathbf{x}^{(v)})_{v \in S}$ is an element of K_S^d, we define its *norm* $\|\mathbf{x}\|$, its *content* $c(\mathbf{x})$ and *height* $H(\mathbf{x})$ by

$$\|\mathbf{x}\| = \max_{v \in S} \|\mathbf{x}^{(v)}\|_v, \quad c(\mathbf{x}) = \prod_{v \in S} \|\mathbf{x}^{(v)}\|_v, \quad H(\mathbf{x}) = \prod_{v \in S} \max\{1, \|\mathbf{x}^{(v)}\|_v\}.$$

It is clear that $c(\mathbf{x}) \leq H(\mathbf{x})$ and $\|\mathbf{x}\| \leq H(\mathbf{x}) \leq \max\{1, \|\mathbf{x}\|^{|S|}\}$. It follows from the product formula that $\|\mathbf{x}\| \geq 1$ if $0 \neq \mathbf{x} \in \mathcal{O}_{K,S}$. The image of $\mathcal{O}_{K,S}$ in K_S under the diagonal embedding is discrete and cocompact in K_S [26, Chapter VII], and norm-closed balls in K_S are compact. Furthermore, it is easily seen from Dirichlet's unit theorem that there is a constant $C = C(K, S) > 0$ such that for all $\mathbf{x} \in K_S^d$ with $c(\mathbf{x}) \neq 0$, there is an S-unit $\alpha \in \mathcal{O}_{K,S}$ such that

$$e^{-C} \|\alpha\mathbf{x}\|^{|S|} \leq c(\mathbf{x}) \leq \|\alpha\mathbf{x}\|^{|S|}. \tag{25}$$

For each $v \in S$, we denote by $\mathrm{GL}_d(K_v)$ the group of invertible $d \times d$ matrices with coefficients in K_v, and we set

$$\mathrm{GL}_d(K_S) = \prod_{v \in S} \mathrm{GL}_d(K_v).$$

A product measure $\mu = \bigotimes_{v \in S} \mu_v$ on $\mathrm{GL}_d(K_S)$ will be called a *good measure* if each μ_v is a good measure on $\mathrm{GL}_d(K_v)$. The definition of a good measure given in §1.9 for $K_v = \mathbb{R}$ extends verbatim to other local fields K_v.

Examples of good measures are provided by push-forwards of Haar measure under strictly analytic maps. Indeed, Proposition 3 continues to hold for analytic maps $\phi : U \to \mathrm{GL}_d(K_v)$ whose coordinates are defined by convergent power series on a ball $U := B(x, r) = \{y \in K_v^n; \max_{i=1}^n |y_i - x_i| \leq r\}$ in K_v^n. This is because, on the one hand the push-forward of Haar measure on K_v under ϕ will be locally good everywhere by [24, Proposition 4.2], and on the other hand the Plücker closure of the image of a ball $B \subset U$ of positive radius is independent of the ball, because convergent power series that vanish on an open ball must vanish everywhere.

Let $E_v := \bigoplus_{k \leq d} \bigwedge^k K_v^d$. We will say that the Plücker closure of μ is defined over $\overline{\mathbb{Q}}$ if for each $v \in S$, the subspace $\mathcal{H}_{\operatorname{supp} \mu_v}$ of $\operatorname{End}(E_v)$ (see §1.8) is defined over $\overline{\mathbb{Q}}$, i.e. is the zero set of a family of linear forms on $\operatorname{End}(E_v)$ with coefficients in $\overline{\mathbb{Q}} \cap K_v$. Clearly this is the case if for each $v \in S$, the Zariski closure of $\operatorname{supp} \mu_v$ in $\operatorname{GL}_d(K_v)$ is defined over $\overline{\mathbb{Q}}$. We will also denote by $\mathcal{H}(\mu)$ the cartesian product of all $\mathcal{H}(\mu_v)$, where $\mathcal{H}(\mu_v) := \{g \in \operatorname{GL}_d(K_v); \rho_v(g) \in \mathcal{H}_{\operatorname{supp} \mu_v}\}$.

We are now ready to state:

Theorem 6 (Subspace theorem for manifolds, S-arithmetic version). *Let K be a number field, S a finite set of places including all archimedean ones, $\mathcal{O}_{K,S}$ its ring of S-integers, and d in $\mathbb{N}$. Let μ be a good measure on $\operatorname{GL}_d(K_S)$ whose Plücker closure is defined over $\overline{\mathbb{Q}}$. Then there are proper subspaces $V_1, \ldots, V_r$ of K^d, with $r = r(d, |S|)$, such that for μ-almost every L and for every $\varepsilon > 0$, the inequality*

$$\prod_{v \in S} \prod_{i=1}^{d} |L_i^{(v)}(\mathbf{x})|_v \leq \frac{1}{c(\mathbf{x})^{\varepsilon}} \tag{26}$$

has only finitely many solutions $\mathbf{x} \in \mathcal{O}_{K,S}^d \setminus (V_1 \cup \cdots \cup V_r)$ up to multiplication by an S-unit.

Here $L = (L^{(v)})_{v \in S} \in \operatorname{GL}_d(K_S)$ and each $L_i^{(v)}$ denotes the i-th row of the matrix $L^{(v)} \in \operatorname{GL}_d(K_v)$. Note that the left- and right-hand sides of (26) are unchanged if $\mathbf{x}$ is changed to $\alpha \mathbf{x}$ for some S-unit α, so we will focus on the equivalence classes of solutions. The bound on the number r of exceptional subspaces depends only on d and $|S|$, and each subspace contains infinitely many (classes of) solutions to (26).

If μ is a Dirac mass at a point $L = (L^{(v)})_{v \in S}$ with $L^{(v)} \in \operatorname{GL}_d(\overline{\mathbb{Q}}) \cap \operatorname{GL}_d(K_v)$, then the theorem is exactly the S-arithmetic Schmidt subspace theorem as stated in [5, §7.2.5].

2.1. Parametric version

Theorem 6 is deduced from Theorem 7 below, which is a parametric version analogous to Theorem 3. To formulate it, we need to define the S-arithmetic analogues of a lattice and its successive minima. A family of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in K_S^d is said to be *linearly independent* if it spans a free K_S-submodule of rank k. Equivalently, the vectors $\mathbf{x}_1^{(v)}, \ldots, \mathbf{x}_k^{(v)}$ are linearly independent over K_v for each place v.

Definition 5 (Lattice in a number field). For any positive integers $k \leq d$, we define a *sublattice in K_S^d* of rank k to be a discrete free $\mathcal{O}_{K,S}$-submodule of rank k in K_S^d. In other words, it is a subgroup $\Delta \leq K_S^d$ that can be written as

$$\Delta = \mathcal{O}_{K,S}\mathbf{x}_1 \oplus \cdots \oplus \mathcal{O}_{K,S}\mathbf{x}_k$$

for some linearly independent elements $\mathbf{x}_i = (\mathbf{x}_i^{(v)})_{v \in S}$ in K_S^d. If $k = d$ we say that Δ is a *lattice in K_S^d*.

If L is any element of $\mathrm{GL}_d(K_S)$, then $\Delta = L\mathcal{O}_{K,S}^k$ is a sublattice of rank k in K_S^d. Conversely, they are all of this form. Note that if $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are vectors from a lattice Δ, they are linearly independent if and only if $\mathbf{x}_1^{(v)}, \ldots, \mathbf{x}_k^{(v)}$ are linearly independent for *some* place v.

Definition 6 (Successive minima). If Δ is a sublattice in K_S^d and $k \in [1, d]$,

$$\lambda_k(\Delta) = \inf\{\lambda > 0; \exists \mathbf{x}_1, \ldots, \mathbf{x}_k \in \Delta \text{ linearly independent with } \forall j, \ c(\mathbf{x}_j) \le \lambda\}$$

is its *k-th successive minimum*, where $c(\mathbf{x}_j)$ is the content defined earlier.

It follows from (25) that we may have defined the successive minima using $\|\mathbf{x}_j\|^{|S|}$ in place of $c(\mathbf{x}_j)$ without much difference. In particular, either definition will suit the theorem below.

The analogue of Theorem 3 now reads as follows. Fix a diagonal element $a = \mathrm{diag}(a_i^{(v)})$ in $\mathrm{GL}_d(K_S)$, and consider the flow $a_t = a^t$ for every $t \in \mathbb{N}$.

Theorem 7 (*S*-arithmetic strong parametric subspace theorem). *Let K be a number field and S a finite set of places containing all archimedean ones. Let $(a_t)_{t \in \mathbb{N}}$ be a diagonal flow in $\mathrm{GL}_d(K_S)$ and μ a good measure on $\mathrm{GL}_d(K_S)$ whose Plücker closure is defined over $\overline{\mathbb{Q}}$. Then there are K-subspaces $0 = V_0 < V_1 < \cdots < V_h = K^d$ and real numbers $s_1 < \ldots < s_h$ such that for each $i = 1, \ldots, h$ and for μ-almost every L,*

(1) *if $\dim V_{i-1} < k \le \dim V_i$, then $\lim_{t \to +\infty} \frac{1}{t} \log \lambda_k(a_t L(\mathcal{O}_{K,S})^d) = s_i$;*

(2) *for all $t > 0$ large enough, the first $\dim V_i$ successive minima of $a_t L(\mathcal{O}_{K,S})^d$ are attained in V_i.*

In other words, for every $\varepsilon > 0$ and μ-almost every L, there is $t_{\varepsilon,L}$ such that if $t > t_{\varepsilon,L}$, $\ell = 1, \ldots, h$, and $\mathbf{x} \in (\mathcal{O}_{K,s})^d$,

$$c(a_t L\mathbf{x}) \le e^{t(s_\ell - \varepsilon)} \implies \mathbf{x} \in V_{\ell-1}. \tag{27}$$

The K-subspaces V_i, $i = 1, \ldots, h$, appearing in Theorem 7 are the terms of the Harder–Narasimhan filtration associated to $M := \mathrm{supp}\,\mu \subset \mathrm{GL}_d(K_S)$ and the quantities s_i are its slopes. We describe this filtration in the next subsection.

2.2. Harder–Narasimhan filtrations for K_S^d

In this subsection we define a submodular function on the grassmannian $\mathrm{Grass}(K^d)$ that generalizes the expansion rate τ_M defined earlier in §1.2. Given a K_S-submodule V in K_S^d we define its *expansion rate* as follows:

$$\tau(V) := \lim_{t \to +\infty} \frac{1}{t} \log c(a_t \mathbf{v})$$

where $\mathbf{v} \in \bigwedge^* K_S^d = \prod_{v \in S} \bigwedge^* K_v$ and $c(\mathbf{v})$ is the content as defined earlier. Here we identify $\bigwedge^k K_v$ with K_v^N, $N = \binom{d}{k}$ and use the standard basis $e_{i_1} \wedge \cdots \wedge e_{i_k}$ to define the norm, and note that $V = \prod_v V_v$, where each V_v is a K_v-vector subspace represented by

an element $\mathbf{v}_v \in \bigwedge^* K_v$. So $\tau(V)$ is just the sum of the expansion rates of V_v in K_v^d over all places $v \in S$. For a subset $M \subset \mathrm{GL}_d(K_S)$ we also define

$$\tau_M(V) = \max_{L \in M} \tau(LV).$$

As in (8) above, we see that if V represented by $\mathbf{v} \in \bigwedge^* K_S^d$, then

$$\tau_M(V) = \max \{I(a); \ I = (I_v)_{v \in S}, \ \exists L \in M, \ \forall v \in S, \ (L^{(v)}\mathbf{v}^{(v)})_{I_v} \neq 0\},$$

where $I_v \subset [1,d]$, $|I_v| = \dim V_v$, $I(a) = \sum_{v \in S} I_v(a)$ and $I_v(a) = \sum_{i \in I_v} \log |a_i^{(v)}|$, and $\mathbf{w}_I$ is defined by the expression $\mathbf{w} = \sum_{|I|=k} \mathbf{w}_I e_I$, where $e_I = e_{i_1} \wedge \cdots \wedge e_{i_k}$ when $I = \{i_1, \ldots, i_k\}$.

We will say that M is *Plücker irreducible* if its projection to $\mathrm{GL}_d(K_v)$ is Plücker irreducible for each $v \in S$. Under this assumption we see by the same argument as in Lemma 1 that τ_M is submodular on the set of all K_S-submodules of K_S^d.

If we restrict τ_M to the set of K-subspaces of K^d, i.e. the grassmannian $\mathrm{Grass}(K^d)$, then we obtain a well-defined notion of Harder–Narasimhan filtration, Grayson polygon and slopes. This gives what we will call the *rational Grayson polygon* $\mathscr{G}^K$. And a K-linear space V will be *M-semistable* for the semigroup (a_t) if for every K-subspace $W < V$,

$$\frac{\tau_M(W)}{\dim W} \geq \frac{\tau_M(V)}{\dim V}.$$

But we may also consider τ_M as a function on the set of all K_S-submodules of K_S^d. Since the dimension of each projection to K_v may not be the same for all v, we use the following definition for the dimension of a submodule $V = \prod_v V_v$:

$$\dim V = \frac{1}{|S|} \sum_{v \in S} \dim_{K_v} V_v.$$

Then $\dim V$ is a modular function on the "full grassmannian", i.e. the set of all K_S-submodules of K_S^d. Thus Proposition 1 and its proof are still valid and we obtain a "full" Harder–Narasimhan filtration and a "full" Grayson polygon $\mathscr{G}^{K_S}$, whose nodes now have x-coordinates in $\frac{1}{|S|}\mathbb{N}$.

2.3. Inheritance principle and proofs

In this section we discuss the proof of Theorems 6 and 7. A basic ingredient is the S-arithmetic version of Minkowski's second theorem, which without paying attention to numerical constants takes the following form:

Theorem 8 (Minkowski's second theorem). *Let Δ be a sublattice in K_S^d as in (5). Then*

$$c(\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_k)^{1/[K:\mathbb{Q}]} \leq \lambda_1(\Delta) \cdot \ldots \cdot \lambda_k(\Delta) \ll c(\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_k)^{1/[K:\mathbb{Q}]},$$

where the constant involved in the Vinogradov notation $\ll$ depends only on K, S and d, but not on Δ.

The content $c(\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_k)$ is proportional to the covolume of Δ in its K_S-span. See [5, Theorem C.2.11, p. 611] for a proof when S has no non-archimedean places and [23] in the general case with the caveat that the normalizations used in the latter paper differ from ours especially at the complex place, leading to a slightly different definition of the successive minima.

The derivation of Theorem 6 from Theorem 7 goes verbatim as that of Theorem 1 from Theorem 3 given in §1.6. One replaces d with $d|S|$ and treats all linear forms $L_i^{(v)}$ on an equal footing. The constant D needs to be increased appropriately, and at non-archimedean places v the exponential used in the definition of the flow will be replaced by a power of a uniformizer π_v of K_v, namely $a_t^{(v)} = \mathrm{diag}(\pi_v^{n_{1,v}t}, \ldots, \pi_v^{n_{d,v}t})$ for integers $n_{i,v}$. One also needs to recall that in view of (25), for each $T > 0$ there are only finitely many classes of $\mathbf{x} \in \mathcal{O}_{K,S}^d$ with $c(\mathbf{x}) \le T$, so we may assume that $c(\mathbf{x})$ is large. The number r of distinct V_i's obtained is bounded by $b(2^{d|S|})$ similarly to Lemma 3.

Now the proof of Theorem 7 is again verbatim as that of Theorem 3, treating all $L_i^{(v)}$ on an equal footing and keeping the argument unchanged. One needs to invoke the S-arithmetic subspace theorem (in the form of Theorem 6 for a single point, or as in [5, Corollary 7.2.5]) in place of the ordinary subspace theorem. For the dynamical ingredient at the end, one applies instead the following generalization of Theorem 4.

Theorem 9 (Inheritance principle). *Let K be a number field, S a finite set of places containing all archimedean places, and d in $\mathbb{N}$. Let (a_t) be a diagonal one-parameter semigroup in $\mathrm{GL}_d(K_S)$, and μ a good measure on $\mathrm{GL}_d(K_S)$ with $\mathcal{H}(\mu) = \prod_{v \in S} \mathcal{H}(\mu_v)$ the Plücker closure of its support. For $k \in [1, d]$ let*

$$\mu_k(L) := \liminf_{t \to \infty} \frac{1}{t} \sum_{i \le k} \log \lambda_i(a_t L \mathcal{O}_{K,S}^d)$$

for each $L \in \mathrm{GL}_d(K_S)$. Then, for μ-almost every L,

$$\mu_k(L) = \sup_{L' \in \mathcal{H}(\mu)} \mu_k(L').$$

Sketch of proof. When $k = 1$ the proof is identical to the proof of Theorem 4 with the following adjustments. Theorem 5 was extended to this context by Kleinbock–Tomanov [24, §8.4] (technically speaking only for $K = \mathbb{Q}$, but the general case is entirely analogous). The norm there and in the definition ($\mathcal{C}_\beta$) of β must be replaced by the content of the corresponding vector. Theorem 8 must be used in place of the original Minkowski theorem to prove the first claim. Also (16) continues to hold with the content in place of the norm for subsets of $\mathrm{GL}_d(K_S)$ that are cartesian products of subsets of $\mathrm{GL}_d(K_v)$ for $v \in S$, which is the case for $\mathrm{supp}(\mu)$ by definition of a good measure. So Claim 2 also holds. The case $k > 1$ is analogous, but as in the proof of Proposition 2, one needs to use the refined quantitative non-divergence estimate proved in [33, Chapter 6] and [27, Theorem 5.3] instead of Theorem 5. ∎

We end this section by stating the analogue of Proposition 2 in the S-arithmetic context. The following describes what is left of Theorem 7 if we remove the assumption that

the Plücker closure of the good measure μ is defined over $\overline{\mathbb{Q}}$. In §2.2 we have defined two Grayson polygons: the rational one $\mathscr{G}^K$ coming from $\mathrm{Grass}(K^d)$ and the full one $\mathscr{G}^{K_S}$ coming from the full grassmannian of all K_S-submodules. And of course we have as before the polygon $\mathscr{G}_\mu$ with nodes (k, μ_k), $k \in [1, d]$, where μ_k is the μ-almost sure value of $\mu_k(L)$ provided by Theorem 9 and the polygon $\mathscr{G}_\mu^{\mathrm{sup}}$ with nodes $(k, \mu_k^{\mathrm{sup}})$, where μ_k^{sup} is the supremum of $\mu_k^{\mathrm{sup}}(L)$ over $L \in M$, and $\mu_k^{\mathrm{sup}}(L)$ is defined as $\mu_k(L)$ with limsup in place of liminf.

Proposition 4 (S-arithmetic sandwich theorem). *The polygons $\mathscr{G}_\mu, \mathscr{G}_\mu^{\mathrm{sup}}$ lie between $\mathscr{G}^K$ and $\mathscr{G}^{K_S}$, i.e.*

$$\mathscr{G}^{K_S} \leq \mathscr{G}_\mu \leq \mathscr{G}_\mu^{\mathrm{sup}} \leq \mathscr{G}^K.$$

Again the proof is mutatis mutandis that of Proposition 2.

3. Examples and applications

In this last section we present a number of applications of the main theorem.

3.1. Sprindzuk conjecture

We begin by the demonstration of how the main result of Kleinbock and Margulis [25, Conjectures H1, H2], namely the Sprindzuk conjecture, can be easily deduced from Theorem 1. Let us recall this result. For $\mathbf{q} \in \mathbb{Z}^d$ we define $\Pi_+(\mathbf{q}) := \prod_{i=1}^d |q_i|_+$, where $|x|_+ := \max\{1, |x|\}$ for all $x \in \mathbb{R}$. A point $y \in \mathbb{R}^n$ is said to be *very well multiplicatively approximable* (or VWMA for short) if for some $\varepsilon > 0$ there are infinitely many $(p, \mathbf{q}) \in \mathbb{Z} \times \mathbb{Z}^d$ such that

$$|p + \mathbf{q} \cdot y| \cdot \Pi_+(\mathbf{q}) \leq \Pi_+(\mathbf{q})^{-\varepsilon}. \tag{28}$$

A manifold $M \subset \mathbb{R}^d$ is said to be *strongly extremal* if Lebesgue almost every point on M is not VWMA.

Theorem 10 (Kleinbock–Margulis [25]). *Let $U \subset \mathbb{R}^n$ be a connected open set and let $f_1, \ldots, f_d : U \to \mathbb{R}$ be real analytic functions, which together with 1, are linearly independent over $\mathbb{R}$. Write $f = (f_1, \ldots, f_d)$. Then $M := \{f(x); x \in U\}$ is strongly extremal.*

As often with applications of the subspace theorem, proofs proceed by induction on dimension. The induction hypothesis will be as follows: if $g = (g_0, \ldots, g_d)$ is a tuple of linearly independent analytic functions on U and $b_1, \ldots, b_d \in \mathbb{Z}^{d+1}$ are linearly independent, then for almost every $x \in U$ and every $\varepsilon > 0$, there are only finitely many solutions $\mathbf{v} \in \mathbb{Z}^{d+1}$ to the inequalities

$$0 < \prod_{i=0}^d |L_i(\mathbf{v})| \leq \frac{1}{\|\mathbf{v}\|^\varepsilon}, \tag{29}$$

where $\|\mathbf{v}\| = \max |v_i|$, $L_0(\mathbf{v}) = g(x) \cdot \mathbf{v}$ and $L_i(\mathbf{v}) = b_i \cdot \mathbf{v}$ for $i \geq 1$. It is straightforward that this statement implies Theorem 10, by letting $\mathbf{v} = (p, \mathbf{q})$, $g = (1, f)$ and $b_1, \ldots, b_d$ the standard basis of $\{0\} \times \mathbb{Z}^d$.

Let $\phi(x)$ be the matrix whose rows are $L_0/g_0, L_1, \ldots, L_d$. The linear independence assumption implies that $\phi(x) \in \mathrm{GL}_{d+1}(\mathbb{R})$ on an open subset $U' \subset U$, and without loss of generality we may assume that $U' = U$. The equations defining the Plücker closure of $\phi(U)$ are linear combinations of $k \times k$ minors of $\phi(x)$. Since those are linear combinations of the g_i/g_0, the Plücker closure of $\phi(U)$ is the Plücker closure of the set of all matrices in $\mathrm{GL}_{d+1}(\mathbb{R})$ whose first row is $(1, y)$ with $y \in \mathbb{R}^d$ arbitrary and whose other rows are $L_1, \ldots, L_d$. So it is defined over $\mathbb{Q}$.

We may thus apply Theorem 1 and conclude that there are a finite number of proper rational hyperplanes V such that for almost every x, the large enough solutions $\mathbf{v}$ to (29) are contained in some V.

Now pick one such V, and consider the restriction of L_0 to V. For $\mathbf{v} \in V$ we can write $\mathbf{v} = \sum_{i=1}^d k_i v_i$ for $\mathbf{k} = (k_1, \ldots, k_d) \in \mathbb{Z}^d$, where $v_1, \ldots, v_d$ is a basis of $V \cap \mathbb{Z}^{d+1}$. Set $L_0'(\mathbf{k}) := L_{0|V}(\mathbf{v}) = \mathbf{v} \cdot g = \mathbf{k} \cdot h$, where $h = (h_1, \ldots, h_d)$ and $h_i = v_i \cdot g$. And for $i \geq 1$, $L_i'(\mathbf{k}) := L_{i|V}(\mathbf{v}) = \mathbf{v} \cdot b_i = \mathbf{k} \cdot c_i$, where $c_i = (b_i \cdot v_1, \ldots, b_i \cdot v_d)$. The L_i are linearly independent, so the $(L_i')_{i=0}^d$ have rank at least d on $\mathbb{R}^d$, and thus $(c_1, \ldots, c_d)$ has rank at least $d - 1$. Up to reordering, we may assume that $c_1, \ldots, c_{d-1}$ are linearly independent. Similarly we see that $h_1, \ldots, h_d$ are linearly independent.

Finally observe that a solution $\mathbf{v} \in V$ to (29) yields a $\mathbf{k} \in \mathbb{Z}^d$ such that

$$0 < \prod_{i=0}^d |L_i'(\mathbf{k})| \ll \frac{1}{\|\mathbf{k}\|^\varepsilon}.$$

But $L_d'(\mathbf{k}) = \mathbf{v} \cdot b_d \in \mathbb{Z} \setminus \{0\}$. Thus $0 < \prod_{i=0}^{d-1} |L_i'(\mathbf{k})| \ll 1/\|\mathbf{k}\|^\varepsilon$ and thus, by induction hypothesis, $\mathbf{k}$ belongs to a finite set of points. Hence so does $\mathbf{v}$. This ends the proof.

3.2. Ridout's theorem for manifolds

Ridout's theorem [5, 43] is an extension of Roth's theorem where p-adic places are allowed. This improves the exponent in Roth's theorem from 2 to 1 in case the rational approximations have denominators with prime factorization in a fixed subset. In this subsection, we present one possible similar variant of the Kleinbock–Margulis theorem (Theorem 10). This will rely on the S-arithmetic subspace theorem for manifolds (Theorem 6).

Theorem 11. *Let S be a finite set of primes. Let $U \subset \mathbb{R}^n$ be a connected open subset and $f_1, \ldots, f_d : U \to \mathbb{R}$ be real analytic functions, which together with 1 are linearly independent over $\mathbb{R}$. Then for every $\varepsilon > 0$ and for Lebesgue almost every $u \in U$, we have*

$$\max_{1 \leq i \leq d} |q f_i(u) - p_i| \geq q^{-\varepsilon} \tag{30}$$

for all $(q, p_1, \ldots, p_d) \in \mathbb{Z}^{d+1}$ with all but finitely many prime factors of q in S.

This is in contrast with the same result [25] without the restriction on the denominator q, where the right-hand side of (30) needs to be replaced by the weaker bound $q^{-1/d-\varepsilon}$. We have chosen simultaneous approximation for a change, but a similar statement with similar proof also holds for linear forms. Moreover, the theorem holds under the weaker assumption that the subspace of $\mathbb{R}^{d+1}$ spanned by all vectors $(1, f_1(u), \ldots, f_d(u))$, $u \in U$, is defined over $\overline{\mathbb{Q}}$.

Proof of Theorem 11. The proof is similar to the one we gave of Theorem 10 in §3.1, but it is convenient to formulate things projectively. Let $x \in \mathbb{P}^{d-1}$ denote the point with homogeneous coordinates $[1 : f_1(u) : \cdots : f_d(u)]$, let $v \in \mathbb{P}^{d-1}(\mathbb{Q})$ be the line containing $\mathbf{v} = (q, p_1, \ldots, p_d)$, and denote by $d(x, v)$ the usual distance on $\mathbb{P}^{d-1}(\mathbb{R})$. We want to show that if x is chosen randomly on an analytic submanifold not contained in any proper rational subspace, and if ϕ is a non-zero integer linear form on $\mathbb{Q}^d$, then almost surely the conditions

$$d(x, v) \le \|\mathbf{v}\|^{-1-\varepsilon} \quad \text{and} \quad \phi(\mathbf{v}) = \prod_{p \in S} p^{m_p}, \quad m_p \in \mathbb{N}^*, \tag{31}$$

have only finitely many solutions $\mathbf{v} \in \mathbb{Z}^d \setminus \{0\}$. Define linear forms

$$L_{u,0}^{(\infty)}(\mathbf{v}) = \phi(\mathbf{v}) \quad \text{and} \quad L_{u,i}^{(\infty)}(\mathbf{v}) = q f_i(u) - p_i \quad \text{if } i > 0$$

and if $p \in S$,

$$L_{u,0}^{(p)}(\mathbf{v}) = \phi(\mathbf{v}) \quad \text{and} \quad L_{u,i}^{(p)}(\mathbf{v}) = \phi_i(\mathbf{v}),$$

where ϕ_i, $i \ge 1$, are chosen arbitrarily so that $(\phi, \phi_1, \ldots, \phi_d)$ is a basis of $\mathrm{Hom}(\mathbb{Z}^{d+1}, \mathbb{Q})$. We note that if $\mathbf{v} := (q, p_1, \ldots, p_d) \in \mathbb{Z}^{d+1}$ satisfies (31) and $\|\mathbf{v}\|$ is large enough then

$$\prod_{0 \le i \le d} \prod_{p \in \{\infty\} \cup S} |L_{u,i}^{(p)}(\mathbf{v})|_p \le \|\mathbf{v}\|^{-d\varepsilon/2},$$

so Theorem 6 applies and if q is large enough, $\mathbf{v}$ belongs to a finite family of proper rational subspaces. But since such a subspace almost surely does not contain x, it can contain only finitely many solutions to (31). ∎

3.3. Submanifolds of matrices: extremality and inheritance

In this subsection, we describe the extension of the Kleinbock–Margulis theorem to the case of submanifolds of matrices and we show how to recover from the subspace theorem for manifolds one of the main results of [2], which is a criterion for extremality in terms of so-called constraining pencils and an explicit computation of the exponent.

In what follows, E and V are two finite-dimensional real vector spaces with a $\mathbb{Q}$-structure. We fix a lattice Δ in V, which defines the rational structure. For x in $\mathrm{Hom}(V, E)$, we define

$$\beta(x) = \inf \{\beta > 0; \exists c > 0, \forall v \in \Delta \setminus \{0\}, \|x(v)\| \ge c\|v\|^{-\beta}\}.$$

Note that $\beta(x)$ only depends on the subspace $\ker x$ in V.

Given a subspace W in V and an integer r, we define the *pencil of endomorphisms* $\mathcal{P}_{W,r}$ by

$$\mathcal{P}_{W,r} = \{x \in \mathrm{Hom}(V, E); \dim x(W) \leq r\}.$$

We say that $\mathcal{P}_{W,r}$ is *constraining* if $\dim W / r < \dim V / \dim E$, and *rational* if W is so. In the case of algebraic sets defined over $\mathbb{Q}$, the following theorem was proved in joint work with Menny Aka and Lior Rosenzweig [2, Theorem 1.2]. The approach taken here yields a different proof of that result, and allows us to generalize it to subsets defined over $\bar{\mathbb{Q}}$.

Theorem 12 (Diophantine exponent for submanifolds of matrices). *Let $U \subset \mathbb{R}^n$ a connected open set and $\phi : U \to \mathrm{Hom}(V, E)$ an analytic map. Assume that the Zariski closure of $M = \phi(U)$ is defined over $\bar{\mathbb{Q}}$. Then, for Lebesgue almost every u in U, setting $x = \phi(u)$,*

$$\beta(x) = \max\left\{\frac{\dim W}{r} - 1; W \text{ a rational subspace such that } M \subset \mathcal{P}_{W,r}\right\}.$$

Remark. For any sublattice $\Delta' \leq \Delta$, one may define

$$r(\Delta') = \max\left\{\dim \mathrm{span}_{\mathbb{R}} x(\Delta'); x \in M\right\}.$$

Then the formula in the above theorem is simply, for almost every x in M,

$$\beta(x) = \beta := \max\left\{\frac{\mathrm{rk}\,\Delta'}{r(\Delta')} - 1; \Delta' \leq \Delta\right\}.$$

Proof of Theorem 12. We prove the theorem by induction on $d = \dim V$, using the subspace theorem.

$d = 1$: The result is clear because, for every x, the subgroup $x(\Delta)$ is a discrete subgroup of E.

$d - 1 \to d$: Suppose the result has been proven for $d - 1 \geq 1$. Let $m = \dim E$ and $d = \dim V$. Fixing bases for E and Δ, we identify $\mathrm{Hom}(V, E)$ with $m \times d$ matrices. Given x in M, we denote by x_i, $1 \leq i \leq d$, its columns, which are vectors in E. The rank of x is almost everywhere constant, and taking a coordinate projection if necessary, we assume that this rank is $m \leq d$ and that for almost every x in M, the vector space E is spanned by the first m columns x_i, $1 \leq i \leq m$. We want to show that, for almost every x in M, for all $\varepsilon > 0$, the set of inequalities

$$|v_1 x_{i1} + v_2 x_{i2} + \cdots + v_d x_{id}| \leq \|v\|^{-\beta-\epsilon}, \quad 1 \leq i \leq m, \tag{32}$$

has only finitely many solutions $\mathbf{v} = (v_1, \ldots, v_d)$ in $\mathbb{Z}^d$. For $1 \leq i \leq m$, define a linear form on $V \simeq \mathbb{R}^d$ by

$$L_i(\mathbf{v}) = v_1 x_{i1} + v_2 x_{i2} + \cdots + v_d x_{id},$$

and for $m < i \leq d$,

$$L_i(\mathbf{v}) = v_i.$$

Since $(x_i)_{1 \le i \le m}$ spans E, the family $(L_i)_{i=1}^d$ of linear forms is linearly independent. Moreover, as $\beta \ge \frac{\text{rk}\,\Delta}{r(\Delta)} - 1 = \frac{d-m}{m}$, condition (32) certainly implies

$$\prod_{i=1}^{d} |L_i(\mathbf{v})| \le \|v\|^{-\epsilon}.$$

We may therefore apply Theorem 1: there exists a finite family $V_1, \ldots, V_h$ of hyperplanes in $\mathbb{Q}^k$ such that, for almost every x in M, all but finitely many integer solutions to (32) lie in the union $V_1 \cup \cdots \cup V_h$. It now suffices to check that in each V_i, there can only be finitely many solutions. This follows from the induction hypothesis applied to $V' = V_i$, $\Delta' = V_i \cap \Delta$ and to the manifold M' that is the image of M under restriction to V'.

   The converse inequality $\beta(x) \ge \beta$ is true for all x in M, as is easily seen by using the classical Dirichlet argument.   ∎

   It is also worth observing the following relation between the notions of extremality and semistability. In [21,25] an analytic submanifold M of $M_{m,n}(\mathbb{R})$ is said to be *extremal* if for almost every $Y \in M$ and every $\varepsilon > 0$ there are only finitely many vectors $q \in \mathbb{Z}^n$ and $p \in \mathbb{Z}^m$ such that $\|Yq - p\| \le \|q\|^{-(n/m+\varepsilon)}$. Now consider the matrix

$$L_Y := \begin{pmatrix} I_m & Y \\ & I_n \end{pmatrix}.$$

The image $\widetilde{M}$ of M under $Y \mapsto L_Y$ defines a submanifold of $\text{GL}_{n+m}(\mathbb{R})$. Using Proposition 2 and Theorem 3, it is then easy to see the following:

Proposition 5 (Extremality vs. semistability). *If $\mathbb{R}^{n+m}$ is $\widetilde{M}$-semistable with respect to the unimodular flow*

$$a_t = (e^{tn}, \ldots, e^{tn}, e^{-tm}, \ldots, e^{-tm}),$$

then M is extremal. If the Zariski (or Plücker) closure of M is defined over $\overline{\mathbb{Q}}$, then the almost sure diophantine exponent β from Theorem 12 can be read off from the rational Grayson polygon of $\widetilde{M}$ by the formula

$$1 + \beta = \frac{n+m}{\gamma+m},$$

where γ is the smallest slope of the rational Grayson polygon. In particular, $\mathbb{Q}^{n+m}$ is $\widetilde{M}$-semistable if and only if M is extremal.

3.4. Multiplicative approximation and strong extremality

Following the suggestion of Baker [3, p. 96] to study the multiplicative diophantine properties of the Mahler curve, Kleinbock and Margulis introduced the notion of *strong extremality* for manifolds in $\mathbb{R}^n$. This was later generalized in [4,21] to the context of diophantine approximation on matrices, but in that generalized setting the optimal criterion for strong extremality remained to be found [4,21]. The method of the present paper can

be used to answer this problem. Below we apply Theorem 1 and give a complete solution when the Zariski closure of the manifold is defined over $\bar{\mathbb{Q}}$.

Let m and n be positive integers, and $M_{m,n}(\mathbb{R})$ the space of $m \times n$ matrices with real entries. Following Kleinbock and Margulis [25], we say that a matrix $Y \in M_{m,n}(\mathbb{R})$ is *very well multiplicatively approximable* (VWMA) if there exists $\varepsilon > 0$ such that the inequality

$$\prod_{i=1}^{m} |Y_i \mathbf{q} - p_i| \leq \prod_{j=1}^{n} |q_j|_+^{-1-\varepsilon}$$

has infinitely many solutions $(\mathbf{p}, \mathbf{q}) \in \mathbb{Z}^m \times \mathbb{Z}^n$. In the above inequality, Y_i denotes the i-th row of Y, for $i = 1, \ldots, m$, and $|q|_+ = \max(|q|, 1)$. More generally, we define as in [9, §1.4] the *multiplicative diophantine exponent* of $Y \in M_{m,n}(\mathbb{R})$ as

$$\omega_\times(Y) = \sup \left\{ \omega > 0; \prod_{i=1}^{m} |Y_i \mathbf{q} - p_i| \leq \prod_{j=1}^{n} |q_j|_+^{-\omega} \text{ for infinitely many } (\mathbf{p}, \mathbf{q}) \right\}.$$

With this definition, we see that a matrix Y is VWMA if and only if $\omega_\times(Y) > 1$. Using the Borel–Cantelli lemma, it is not difficult to check that for the Lebesgue measure, almost every Y in $M_{m,n}(\mathbb{R})$ satisfies $\omega_\times(Y) = 1$. It is therefore natural to ask what other measures μ on $M_{m,n}(\mathbb{R})$ have this property.

We now set up some notation to formulate our criterion for strong extremality. Given a matrix Y in $M_{m,n}(\mathbb{R})$, with rows $Y_1, \ldots, Y_m$, we let

$$L_Y = \begin{pmatrix} -I & Y \\ 0 & I \end{pmatrix},$$

and denote by L_i, $i = 1, \ldots, m + n$, the linear forms on $\mathbb{R}^{m+n}$ given by the rows of the matrix L_Y:

$$L_i(\mathbf{p}, \mathbf{q}) = \begin{cases} Y_i \mathbf{q} - p_i & \text{for } i \in \{1, \ldots, m\}, \\ q_{i-m} & \text{for } i \in \{m+1, \ldots, m+n\}. \end{cases}$$

If W is a linear subspace of $\mathbb{R}^{n+m}$, and I a non-empty subset of $\{1, \ldots, m+n\}$, we let

$$s_{I,W} = \mathrm{rk}(L_i|_W)_{i \in I}.$$

Definition 7 (Multiplicative pencils). Let I, J be proper subsets of $\{1, \ldots, m+n\}$ such that $I \subset \{1, \ldots, m\} \subset J$, and r, s non-negative integers. Given a subspace $W \leq \mathbb{R}^{n+m}$, we define a subvariety of endomorphisms $\mathcal{P}_{I,J,r,s,W} \subset M_{m,n}(\mathbb{R})$ by

$$\mathcal{P}_{I,J,r,s,W} = \{Y \in M_{m,n}(\mathbb{R}); s_{I,W} \leq r \text{ and } s_{J,W} \leq s\}.$$

To justify the relevance of this definition to our problem, we start by an easy proposition, which is a consequence of Minkowski's first theorem or Dirichlet's pigeonhole principle.

Proposition 6 (Dirichlet's principle). *Fix $Y \in M_{m,n}(\mathbb{R})$, and denote by L_i, $i = 1, \ldots,$ $m + n$, the rows of the matrix L_Y. Assume that W is a k-dimensional rational subspace of $\mathbb{R}^{m+n}$ and that $\emptyset \neq I \subset \{1, \ldots, m\} \subset J \subsetneq \{1, \ldots, m+n\}$ are such that*

$$r = \mathrm{rk}(L_i|_W)_{i \in I} \quad \text{and} \quad s = \mathrm{rk}(L_i|_W)_{i \in J}.$$

Then

$$\omega_\times(Y) \geq \frac{(k-s)|I|}{r(n+m-|J|)}.$$

Remark. By convention, if $r = 0$, the ratio is equal to $+\infty$. Note that one can always take $W = \mathbb{R}^{m+n}$, $I = J = \{1, \ldots, m\}$, and $r = s = m$, in which case the ratio is equal to 1.

Proof of Proposition 6. If $r = 0$, then one must have $Y_i \mathbf{q} - p_i = 0$ for some integer vector $(\mathbf{p}, \mathbf{q})$ in $\mathbb{Z}^{m+n}$. It is then clear that $\omega_\times(Y) = \infty$. So we may assume that $r \neq 0$. Since by definition $\omega_\times(Y) \geq 0$, we may also assume that $s < k$, otherwise there is nothing to prove.

Assuming that $Y \in \mathcal{P}_{I,J,r,s,W}$, we shall prove that there exists a constant $C > 0$ depending only on Y and W such that the inequality

$$\prod_{i=1}^m |L_i(v)| \leq C \prod_{j=1}^n |L_{m+j}(v)|^{-\frac{(k-s)|I|}{r(n+m-|J|)}}$$

has infinitely many solutions v in W. This will yield the desired lower bound on $\omega_\times(Y)$. Let $Q > 0$ be a large parameter. Pick $i_1, \ldots, i_r$ in I such that $L_{i_1}|_W, \ldots, L_{i_r}|_W$ are linearly independent, pick $i_{r+1}, \ldots, i_s$ in J such that $L_{i_1}|_W, \ldots, L_{i_s}|_W$ are linearly independent, and complete with $i_{s+1}, \ldots, i_k$ such that $L_{i_1}|_W, \ldots, L_{i_k}|_W$ are linearly independent. The symmetric convex body in W defined by

$$\begin{cases} |L_{i_\ell}(v)| \leq Q^{-(k-s)} & \text{for } 1 \leq \ell \leq r, \\ |L_{i_\ell}(v)| \ll 1 & \text{for } r < \ell \leq s, \\ |L_{i_\ell}(v)| \leq Q^r & \text{for } s < \ell \leq k, \end{cases}$$

has volume $\gg 1$ and therefore, by Minkowski's first theorem, it contains a non-zero point v in $W \cap \mathbb{Z}^{n+m}$. By our choice of the indices i_ℓ, $1 \leq \ell \leq k$, such a point satisfies

$$\begin{cases} |L_i(v)| \leq Q^{-(k-s)} & \text{for } i \in I, \\ |L_i(v)| \ll 1 & \text{for } i \in J \setminus I, \\ |L_i(v)| \leq Q^r & \text{for } i \notin J, \end{cases}$$

and therefore

$$\prod_{i=1}^m |L_i(v)| \ll Q^{-|I|(k-s)} \ll \prod_{j=1}^n |L_{m+j}(v)|^{-\frac{(k-s)|I|}{r(n+m-|J|)}}. \qquad \blacksquare$$

Now, as in the introduction, let $M = \phi(U)$ be a connected analytic submanifold of $M_{m,n}(\mathbb{R})$ endowed with the measure μ equal to the push-forward of the Lebesgue measure under the analytic map $\phi : U \to M_{m,n}(\mathbb{R})$. When the Zariski closure of M is defined over $\overline{\mathbb{Q}}$ – for example when ϕ is given by a polynomial map with coefficients in $\overline{\mathbb{Q}}$ – the above proposition actually provides a formula for $\omega_\times(Y)$ when Y is a μ-generic point of M. In other words:

Theorem 13 (Formula for the multiplicative exponent). *Assume that the Zariski closure of M is defined over $\overline{\mathbb{Q}}$. Then for μ-almost every Y in M,*

$$\omega_\times(Y) = \max_{\substack{M \subset \mathcal{P}_{I,J,r,s,W} \\ W \text{ rational}}} \frac{(\dim W - s)|I|}{r(m + n - |J|)}.$$

We stated the result for analytic submanifolds for convenience, but it holds with the same proof for all good measures μ in the sense of §1.9 provided the Zariski closure of the support of μ is defined over $\overline{\mathbb{Q}}$. We shall say that a multiplicative pencil $\mathcal{P}_{I,J,r,s,W}$ is *constraining* if

$$\frac{(\dim W - s)|I|}{r(m + n - |J|)} > 1.$$

Our criterion[2] for strong extremality immediately follows from the above formula.

Corollary 3 (Criterion for strong extremality). *If M is an analytic submanifold of $M_{m,n}(\mathbb{R})$ whose Zariski closure is defined over $\overline{\mathbb{Q}}$, then M is strongly extremal if and only if it is not contained in any rational constraining pencil.*

The proof of Theorem 13 is inspired by Schmidt's proof of an analogous result on products of linear forms [37, §12, p. 242].

Proof of Theorem 13. Let $\omega > 0$ be such that for Y in a set of positive measure in M, the inequality

$$\prod_{i=1}^{m} |Y_i \mathbf{q} - p_i| \leq \prod_{j=1}^{m} |q_j|_+^{-\omega} \tag{33}$$

has infinitely many solutions $(\mathbf{p}, \mathbf{q})$ in $\mathbb{Z}^{m+n}$. We want to show that M is included in a pencil $\mathcal{P}_{I,J,r,s,W}$ such that

$$\frac{(\dim W - s)|I|}{r(m + n - |J|)} \geq \omega.$$

Let W be a rational subspace of minimal dimension k containing infinitely many solutions to (33). If $k = 1$, then there must exist $i \in \{1, \ldots, m\}$ and $(\mathbf{p}, \mathbf{q}) \in \mathbb{Z}^{m+n}$ such that $Y_i \mathbf{q} - p_i = 0$, and one can take $I = \{i\}$, $J = \{1, \ldots, m\}$, $r = 0$, and $s = 1$. So we assume

---

[2]We were informed by David Simmons (private communication) that he and Tushar Das also obtained a similar criterion for strong extremality of submanifolds defined over $\mathbb{R}$.

$k \geq 2$. Reordering the indices if necessary, we may assume that W contains infinitely many solutions to (33) satisfying

$$0 < |Y_1\mathbf{q} - p_1| \leq \cdots \leq |Y_m\mathbf{q} - p_m| \tag{34}$$

and

$$|q_1| \leq \cdots \leq |q_n|. \tag{35}$$

Define $i_1 < \cdots < i_f \leq m$ inductively so that each i_ℓ is minimal with $\mathrm{rk}(L_{i_1}|_W, \ldots, L_{i_\ell}|_W) = \ell$, and then $m < j_1 < \cdots < j_g \leq m + n$ such that j_ℓ is minimal with $\mathrm{rk}(L_{i_1}|_W, \ldots, L_{i_f}|_W, L_{j_1}|_W, \ldots, L_{j_\ell}|_W) = f + \ell$. Note that in our notation, $f + g = k = \dim W$.

Given a large solution $\mathbf{v} = (\mathbf{p}, \mathbf{q})$ to (33) in W, choose numbers $c_{1,\mathbf{v}}, \ldots, c_{f,\mathbf{v}}$ such that

$$|L_{i_\ell}(\mathbf{v})| = \|\mathbf{v}\|^{-c_{\ell,\mathbf{v}}} \quad \text{for } \ell = 1, \ldots, f$$

and $d_{1,\mathbf{v}}, \ldots, d_{g,\mathbf{v}}$ such that

$$|L_{j_\ell}(\mathbf{v})| = \|\mathbf{v}\|^{d_{\ell,\mathbf{v}}} \quad \text{for } \ell = 1, \ldots, g.$$

Extracting a subsequence of solutions $\mathbf{v}$ if necessary, we may assume that $c_{\ell,\mathbf{v}}$ and $d_{\ell,\mathbf{v}}$ converge to some limits c_ℓ and d_ℓ.

By (34), one has $c_1 \geq \cdots \geq c_f \geq 0$. By (35) and the fact that always $|L_{j_\ell}(\mathbf{v})| \ll \|\mathbf{v}\|$, one finds $0 \leq d_1 \leq \cdots \leq d_g \leq 1$. Moreover, by minimality of W, the subspace theorem applied in W to the set of linear forms

$$L_{i_1}|_W, \ldots, L_{i_f}|_W, L_{j_1}|_W, \ldots, L_{j_g}|_W$$

shows that $-c_1 + \cdots - c_f + d_1 + \cdots + d_g \geq 0$. In conclusion, one sees that the k-tuple $(c_1, \ldots, c_f, d_1, \ldots, d_g)$ belongs to the convex polytope (P) defined by

$$(P) \quad \begin{cases} c_1 \geq \cdots \geq c_f \geq 0, \\ 0 \leq d_1 \leq \cdots \leq d_g \leq 1, \\ d_1 + \cdots + d_g \geq c_1 + \cdots + c_f. \end{cases}$$

Now inequality (33) implies

$$c_1(i_2 - i_1) + \cdots + c_f(m - i_f) - \omega[d_1(j_1 - m) + \cdots + d_g(m + n - j_g)] \geq 0.$$

The linear map

$$f(\mathbf{c}, \mathbf{d}) = c_1(i_2 - i_1) + \cdots + c_f(m - i_f) - \omega[d_1(j_1 - m) + \cdots + d_g(m + n - j_g)]$$

is non-negative at some point in the convex polytope (P), so it must be non-negative at one of its vertices. The polytope (P) has gf vertices, given by

$$p_{a,b} : \begin{cases} 0 = d_1 = \cdots = d_a < d_{a+1} = \cdots = d_g = 1, \\ \frac{g-a}{b} = c_1 = \cdots = c_b > c_{b+1} = \cdots = c_f = 0, \end{cases} \quad \text{where} \quad \begin{cases} a = 0, \ldots, g - 1, \\ b = 1, \ldots, f. \end{cases}$$

Choose (a, b) such that $f(p_{a,b}) \geq 0$, and let

$$\begin{cases} I = \{1, \ldots, i_b\}, \\ J = \{1, \ldots, j_a\}, \\ r = b, \\ s = f + a. \end{cases}$$

We then obtain

$$0 \leq f(p_{a,b}) = \frac{g - a}{b} |I| - \omega(m + n - |J|),$$

whence

$$\frac{(k - s)|I|}{r(m + n - |J|)} \geq \omega. \qquad \blacksquare$$

Remark. Strong extremality also relates to semistability in a similar way to Proposition 5. We leave it to the reader to check that in the setting of Theorem 13, M is strongly extremal if and only if $\mathbb{Q}^{m+n}$ is semistable for $\{L_Y; Y \in M\}$ with respect to all unimodular flows $a_t = (e^{tA_1}, \ldots, e^{tA_{n+m}})$ with

$$A_1 \geq \cdots \geq A_m \geq 0 \geq A_{m+1} \geq \cdots \geq A_{m+n}.$$

See [21] where the relevance of this family of flows for strong extremality was uncovered.

3.5. Roth's theorem for nilpotent Lie groups

Diophantine approximation on nilpotent Lie groups was studied in [1, 2]. In this subsection we continue this study and derive an analogue of Roth's theorem [31] in this context. In doing so we extend Theorem 12 to approximation with quasi-norms.

Diophantine approximation in nilpotent Lie groups. In this subsection G will denote a simply connected nilpotent Lie group, endowed with a left-invariant riemannian metric. We identify it with its Lie algebra $\mathfrak{g}$ under the exponential map, so that its Haar measure is given by the Lebesgue measure on $\mathfrak{g}$. For a finite symmetric subset S of G and $\Gamma = \langle S \rangle$ the subgroup it generates, the *diophantine exponent* of Γ in G is

$$\beta(\Gamma) = \inf\{\beta; \exists c > 0, \forall n \in \mathbb{N}^*, \forall x \in S^n \setminus \{1\}, d(x, 1) \geq cn^{-\beta}\},$$

where S^n denotes the set of elements of Γ that can be written as a product of at most n elements of S. Because Γ is nilpotent, this definition does not depend on the choice of S. The following theorem was proved in [2, Theorem 7.4].

Theorem 14 (Existence of the exponent). *Let G be a connected and simply connected real nilpotent Lie group endowed with a left-invariant riemannian metric d. For each $k \geq 1$, there is $\beta_k \in [0, \infty]$ such that for almost every k-tuple $\mathbf{g} = (g_1, \ldots, g_k) \in G^k$ with respect to the Haar measure, the subgroup $\Gamma_{\mathbf{g}}$ generated by $g_1, \ldots, g_k$ satisfies*

$$\beta(\Gamma_{\mathbf{g}}) = \beta_k.$$

We also showed in [2] that when G is rational, i.e. when its Lie algebra $\mathfrak{g}$ admits a basis with rational structure constants, β_k can be explicitly computed and in particular is rational. In fact, $\beta_k = F(k)$ for some rational function $F \in \mathbb{Q}(X)$ when k is large enough. Using the subspace theorem for manifolds (Theorem 1), we can now show the following.

Theorem 15 (Roth's theorem for nilpotent groups). *Let G be a connected simply connected nilpotent Lie group and k a positive integer. Assume that the Lie algebra $\mathfrak{g}$ of G admits a basis with structure constants in a number field K. Then*

$$\beta_k \in \mathbb{Q}.$$

Moreover, for every k-tuple $\mathbf{g} \in \mathfrak{g}(\overline{\mathbb{Q}})^k$ we have $\beta(\Gamma_{\mathbf{g}}) \in \mathbb{Q}$. Furthermore, there is $C = C(\dim \mathfrak{g}) > 0$ and a countable union $\mathcal{U}$ of proper algebraic subsets of $\mathfrak{g}^k$ defined over K and of degree at most C such that

$$\beta(\Gamma_{\mathbf{g}}) = \beta_k \quad \text{for every } \mathbf{g} \in \mathfrak{g}^k(\overline{\mathbb{Q}}) \setminus \mathcal{U}.$$

The case when $k = 2$ and $G = (\mathbb{R}, +)$ is exactly Roth's theorem. In this case $\beta_2 = 1$ and $\mathcal{U}$ is the family of lines in $\mathbb{R}^2$ with rational slopes.

The basic idea for the proof of Theorem 15, developed in [2], is to reduce the problem to a question of diophantine approximation on submanifolds. For that, we introduce the free Lie algebra $\mathcal{F}_k$ over k generators $\mathsf{x}_1, \ldots, \mathsf{x}_k$. Inside $\mathcal{F}_k$, the ideal of laws $\mathcal{L}_{k,\mathfrak{g}}$ on the Lie algebra $\mathfrak{g}$ of G is the set of elements r in $\mathcal{F}_k$ such that $\mathsf{r}(X_1, \ldots, X_k) = 0$ for every $X_1, \ldots, X_k$ in $\mathfrak{g}$, and the ideal of rational laws $\mathcal{L}_{k,\mathfrak{g},\mathbb{Q}}$ is the real span of the intersection of $\mathcal{L}_{k,\mathfrak{g}}$ with $\mathcal{F}_k(\mathbb{Q})$, the natural $\mathbb{Q}$-structure on $\mathcal{F}_k$. The Lie algebra $\mathcal{F}_{k,\mathfrak{g},\mathbb{Q}} = \mathcal{F}_k / \mathcal{L}_{k,\mathfrak{g},\mathbb{Q}}$ has a graded structure

$$\mathcal{F}_{k,\mathfrak{g},\mathbb{Q}} = \bigoplus_{i=1}^{s} \mathcal{F}_{k,\mathfrak{g},\mathbb{Q}}^{[i]},$$

where $\mathcal{F}_{k,\mathfrak{g},\mathbb{Q}}^{[i]}$ is the homogeneous part of $\mathcal{F}_{k,\mathfrak{g},\mathbb{Q}}$ consisting of brackets of degree i. For $\mathsf{r} = \sum \mathsf{r}_i$ with $\mathsf{r}_i \in \mathcal{F}_{k,\mathfrak{g},\mathbb{Q}}^{[i]}$, we let

$$|\mathsf{r}| := \max_{i=1,\ldots,s} \|\mathsf{r}_i\|^{1/i}, \tag{36}$$

where $\| \cdot \|$ is a fixed norm on $\mathcal{F}_{k,\mathfrak{g},\mathbb{Q}}$. Endowed with this quasi-norm, the Lie algebra $\mathcal{F}_{k,\mathfrak{g},\mathbb{Q}}$ is quasi-isometric to the group of word maps on G, endowed with the word metric [2, Proposition 7.2]. This yields the following characterization for the above diophantine exponent $\beta(\Gamma_{\mathbf{g}})$ proved in [2, Proposition 7.3]. We say that $\Gamma_{\mathbf{g}}$ is *relatively free* in G if the only relations satisfied by $\mathbf{g}$ are the laws of G. This holds for all $\mathbf{g}$ outside a countable union of proper algebraic subvarieties of bounded degree defined over $\mathbb{Q}$, and in particular for Lebesgue almost every $\mathbf{g} \in G^k$.

Proposition 7. *Let G be a simply connected nilpotent Lie group with Lie algebra $\mathfrak{g}$. Let $\mathbf{g} = (e^{X_1}, \ldots, e^{X_k})$ be a k-tuple in G such that $\Gamma_{\mathbf{g}}$ is relatively free in G. Then the exponent $\beta(\Gamma_{\mathbf{g}})$ defined above is also the infimum of all $\beta > 0$ such that*

$$\|\mathsf{r}(X_1, \ldots, X_k)\| \geq |\mathsf{r}|^{-\beta} \tag{37}$$

for all but finitely many $\mathsf{r} \in \mathcal{F}_{k,\mathfrak{g},\mathbb{Q}}(\mathbb{Z})$.

Below, we show how Theorem 1 yields a formula for diophantine exponents defined with quasi-norms, as in the above proposition.

Weighted diophantine approximation. We now generalize the results of §3.3 to diophantine approximation with quasi-norms, also called *weighted diophantine approximation* (see e.g. [22, §1.5] and references therein). Again, E and V are two finite-dimensional real vector spaces, and Δ is a lattice in V, defining a rational structure.

We fix a norm $\|\cdot\|$ on E. On V, we measure the size of vectors using a quasi-norm $|\cdot|$ given by the formula

$$|v| = \max_{1 \le i \le d} |\langle v, u_i^* \rangle|^{1/\alpha_i}, \tag{38}$$

where $\alpha = (\alpha_1, \ldots, \alpha_d)$ is a d-tuple of positive real numbers and $(u_i^*)_{1 \le i \le d}$ a basis of $\mathrm{Hom}(V, \mathbb{R})$. Given x in $\mathrm{Hom}(V, E)$, define the diophantine exponent of x by

$$\beta_\alpha(x) = \inf \{\beta > 0; \exists c > 0, \forall v \in \Delta, \|xv\| \ge c|v|^{-\beta}\}.$$

Definition 8 (Growth rate for a quasi-norm). Let W be a linear subspace in V. The *growth rate* of balls in W for the quasi-norm $|\cdot|$ is given by

$$\alpha(W) = \lim_{R \to \infty} \frac{1}{\log R} \log \mathrm{Vol} \{v \in W; |v| \le R\}.$$

Remark. It is not hard to see that this limit exists and equals $\sum_{i \in I_W} \alpha_i$ for a certain subset $I_W \subset [1, d]$. Indeed, the restriction of $|\cdot|$ to W is itself comparable up to multiplicative constants to a quasi-norm with exponents α_i, $i \in I_W$, where $I_W = \{i_1, \ldots, i_k\}$ is defined as follows. Choose i_1 minimal such that the restriction of $u_{i_1}^*$ to W is nonzero, then inductively choose i_j minimal such that the linear forms $u_{i_1}^*|_W, \ldots, u_{i_j}^*|_W$ are linearly independent.

This gives us a lower bound for the diophantine exponent, using a standard Dirichlet type argument:

Lemma 5. *Let V and $|\cdot|$ be as above. For any x in $\mathrm{Hom}(V, E)$,*

$$\beta_\alpha(x) \ge \max \left\{ \frac{\alpha(W \cap \ker x)}{\dim W - \dim W \cap \ker x} ; W \le V \text{ a rational subspace} \right\}.$$

Proof. Let $R > 0$ be some large parameter. The number of points v in $W \cap \Delta$ such that $|v| \le R$ is roughly $R^{\alpha(W)}$, and their images in $x(W) \simeq W/(W \cap \ker x)$ lie in a distorted ball of volume $O(R^{\alpha(W) - \alpha(W \cap \ker x)})$. Comparing volumes, we find that balls of radius ε around those points cannot be disjoint if $\varepsilon \gg R^{-\frac{\alpha(W \cap \ker x)}{\dim W - \dim W \cap \ker x}}$. ∎

It turns out that this lower bound is in fact attained almost everywhere on analytic submanifolds of $\mathrm{Hom}(V, E)$ whose Zariski closure is defined over $\overline{\mathbb{Q}}$. This is the content of the next theorem. As earlier, we endow $M := \phi(U)$ with the push-forward μ of the Lebesgue measure on the connected open subset $U \subset \mathbb{R}^d$ via the analytic map $\phi : U \to \mathrm{Hom}(V, E)$.

Theorem 16 (Diophantine exponent for quasi-norms). *Assume that V and $|\cdot|$ are as above, and the Zariski closure $\mathrm{Zar}(M)$ of M is defined over $\bar{\mathbb{Q}}$. Then, for almost every x in M,*

$$\beta_\alpha(x) = \max \left\{ \min_{y \in M} \frac{\alpha(W \cap \ker y)}{\dim W - \dim W \cap \ker y}; \ W \le V \text{ a rational subspace} \right\}. \quad (39)$$

This equality is also true for every $\bar{\mathbb{Q}}$-point of $\mathrm{Zar}(M)$ outside a union of proper algebraic subsets of M defined over $\mathbb{Q}$ and of bounded degree.

Remark. Note that α can take only finitely many values, so the maximum and minimum in the above formula are indeed attained.

Remark. The theorem implies that $\beta_\alpha(x) \ge \beta$ for almost every x on M if and only if $\beta_\alpha(y) \ge \beta$ for every y on M, if and only if there exists a rational subspace W such that M is included in the (closed) algebraic subset of all $y \in \mathrm{Hom}(V, E)$ such that $\alpha(W \cap \ker y) \ge \beta(\dim W - \dim W \cap \ker y)$.

Proof of Theorem 16. Since we can always reorder the u_i^*, $1 \le i \le d$, we may assume without loss of generality that $0 < \alpha_1 \le \cdots \le \alpha_d$. Then, for x in $\mathrm{Hom}(V, E)$, let

$$m = \dim x(V) \quad \text{and} \quad n = \dim \ker x.$$

and define inductively a subset $I_x = \{i_{1,x}, \dots, i_{n,x}\}$ in $\{1, \dots, d\}$ by

$$i_{1,x} = \min \{i; \ u_i^* |_{\ker x} \ne 0\},$$
$$i_{j+1,x} = \min \{i; \ (u_{i_{1,x}}^* |_{\ker x}, \dots, u_{i_{j,x}}^* |_{\ker x}, u_i^* |_{\ker x}) \text{ is linearly independent}\}.$$

Now suppose L_x is an element of $\mathrm{GL}(d, \mathbb{R})$ with rows $L_{j,x}$ satisfying

(1) $\ker x = \bigcap_{j=1}^m \ker L_{j,x}$,

(2) $\forall j \in \{1, \dots, n\}$, $L_{m+j,x} = u_{i_{j,x}}^*$.

Then $\beta_\alpha(x)$ is the minimal $\beta > 0$ such that, for all $\varepsilon > 0$, for Q large enough, the set of inequalities

$$\begin{cases} |L_{j,x}(v)| \le Q^{-\beta-\varepsilon} & \text{for } 1 \le i \le m, \\ |L_{m+j,x}(v)| \le Q^{\alpha_{i_j}} & \text{for } 1 \le j \le n, \end{cases} \quad (40)$$

has no non-zero solution v in Δ.

Let β be the right-hand side of (39). We want to show that for μ-almost every x, for all $\varepsilon > 0$ and for $Q > 0$ large enough, the inequalities (40) have no non-zero solution in Δ. We prove this by induction on d, using our subspace theorem (i.e. Theorem 1).

$d = 1$: There is not much to prove in this case, since $\ker x$ is equal to $\{0\}$ or V, for μ-almost every x.

$d - 1 \to d$: Assume the result is already proven for $d - 1$. The map $x \mapsto I_x$ is constant on a dense Zariski open subset of $\mathrm{Zar}(M)$; we denote by $I = \{i_1, \dots, i_n\}$ its value on

this dense open subset, and by $\widetilde{M}$ the set of all L_x, for x in that subset and in M. Since $\beta \geq \alpha(V)/m = \sum_{j=1}^n \alpha_{i_j}/m$, the solutions v to (40) satisfy

$$\prod_{i=1}^d |L_{i,x}(v)| \leq Q^{-m\beta - m\varepsilon + \sum_{j=1}^n \alpha_{i_j}} \leq \|v\|^{-\varepsilon'},$$

and therefore, by Theorem 1, there exist proper subspaces $V_1, \ldots, V_\ell$ in $\mathbb{Q}^d$ containing all but a finite number of solutions. All we need to check is that in each V_i, there can only be finitely many solutions. But this follows from the induction hypothesis, applied to $V' = V_i$, $\Delta' = \Delta \cap V_i$, $|\cdot| = |\cdot|_{V'}$ (the restriction of a quasi-norm to a subspace is comparable to a quasi-norm [2, §4.1]), and to the submanifold $M' = \{x|_{V'}; \ x \in \widetilde{M}\}$.

Indeed, the almost sure diophantine exponent of a point in M' is

$$\beta' := \max \left\{ \min_{y \in M'} \frac{\alpha(W \cap \ker y)}{\dim W - \dim W \cap \ker y}; \ W \leq V' \text{ a rational subspace} \right\},$$

which satisfies $\beta' \leq \beta$. ∎

We can now easily derive our theorem about nilpotent groups.

Proof of Theorem 15. Let

$$V = \mathcal{F}_{k,\mathfrak{g},\mathbb{Q}}, \quad \Delta = \mathcal{F}_{k,\mathfrak{g},\mathbb{Q}}(\mathbb{Z}) \quad \text{and} \quad E = \mathfrak{g}.$$

For each k-tuple $\mathbf{g} = (e^{X_1}, \ldots, e^{X_k})$ in G^k, we obtain an element $x_{\mathbf{g}}$ in $\mathrm{Hom}(V, E)$ given by evaluation at $(X_1, \ldots, X_k)$:

$$x_{\mathbf{g}}(\mathsf{r}) = \mathsf{r}(X_1, \ldots, X_k).$$

The map $G^k \to \mathrm{Hom}(V, E)$, $\mathbf{g} \mapsto x_{\mathbf{g}}$, is a polynomial map with coefficients in K. In particular, the Zariski closure of its image is defined over $\overline{\mathbb{Q}}$. When $\Gamma_{\mathbf{g}}$ is relatively free, Proposition 7 shows that $\beta(\Gamma_{\mathbf{g}}) = \beta(x_{\mathbf{g}})$, where $\beta(x_{\mathbf{g}})$ is the diophantine exponent with respect to the quasi-norm $|\cdot|$ on V defined in (36). By Theorem 16,

$$\beta(x_{\mathbf{g}}) = \max \left\{ \min_{\mathbf{h} \in G^k} \frac{\alpha(W \cap \ker x_{\mathbf{h}})}{\dim W - \dim W \cap \ker x_{\mathbf{h}}}; \ W \leq V \text{ a rational subspace} \right\}$$

for Lebesgue almost every $\mathbf{g} \in G^k$. This shows that β_k is well defined, and since α takes rational values, this formula shows that $\beta_k \in \mathbb{Q}$. For each rational $W \leq V$, the set of all $\mathbf{h} \in G^k$ such that $\alpha(W \cap \ker x_{\mathbf{h}}) > \beta_k \dim(W/(W \cap \ker x_{\mathbf{h}}))$ is a proper algebraic subset of G^k defined by equations of bounded degree with coefficients in K. The union of those subsets forms a proper subset of G^k by Lemma 4, and Theorem 16 implies that $\beta(\Gamma_{\mathbf{g}}) = \beta_k$ for every $\mathbf{g}$ outside this union. ∎

References

[1] Aka, M., Breuillard, E., Rosenzweig, L., de Saxcé, N.: Diophantine properties of nilpotent Lie groups. Compos. Math. **151**, 1157–1188 (2015) Zbl 1332.22011 MR 3357181

[2] Aka, M., Breuillard, E., Rosenzweig, L., de Saxcé, N.: Diophantine approximation on matrices and Lie groups. Geom. Funct. Anal. **28**, 1–57 (2018) Zbl 1440.11134 MR 3777412

[3] Baker, A.: Transcendental number theory. 2nd ed., Cambridge Mathematical Library, Cambridge University Press, Cambridge (1990) Zbl 0715.11032 MR 1074572

[4] Beresnevich, V., Kleinbock, D., Margulis, G.: Non-planarity and metric Diophantine approximation for systems of linear forms. J. Théor. Nombres Bordeaux **27**, 1–31 (2015) Zbl 1375.11053 MR 3346961

[5] Bombieri, E., Gubler, W.: Heights in Diophantine geometry. New Mathematical Monographs 4, Cambridge University Press, Cambridge (2006) Zbl 1115.11034 MR 2216774

[6] Bost, J.-B.: Réseaux euclidiens, séries thêta et pentes. Astérisque **422**, 1–59 (2020) Zbl 1465.11101 MR 4224631

[7] Corvaja, P., Zannier, U.: On a general Thue's equation. Amer. J. Math. **126**, 1033–1055 (2004) Zbl 1125.11022 MR 2089081

[8] Das, T., Fishman, L., Simmons, D., Urbański, M.: A variational principle in the parametric geometry of numbers, with applications to metric Diophantine approximation. C. R. Math. Acad. Sci. Paris **355**, 835–846 (2017) Zbl 1427.11072 MR 3693502

[9] Das, T., Fishman, L., Simmons, D., Urbański, M.: Extremality and dynamically defined measures, part I: Diophantine properties of quasi-decaying measures. Selecta Math. (N.S.) **24**, 2165–2206 (2018) Zbl 1453.11094 MR 3816502

[10] Edixhoven, B., Evertse, J.-H. (eds.): Diophantine approximation and abelian varieties. Lecture Notes in Mathematics 1566, Springer, Berlin (1993) Zbl 0811.14019 MR 1288998

[11] Evertse, J.-H.: An improvement of the quantitative subspace theorem. Compos. Math. **101**, 225–311 (1996) Zbl 0856.11030 MR 1394517

[12] Evertse, J.-H.: On the quantitative subspace theorem. Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **377**, 217–240, 245 (2010) Zbl 1026.11060 MR 2753657

[13] Evertse, J.-H., Ferretti, R. G.: A generalization of the Subspace Theorem with polynomials of higher degree. In: Diophantine approximation, Developments in Mathematics 16, Springer, Wien, 175–198 (2008) Zbl 1153.11032 MR 2487693

[14] Evertse, J.-H., Ferretti, R. G.: A further improvement of the quantitative subspace theorem. Ann. of Math. (2) **177**, 513–590 (2013) Zbl 1352.11060 MR 3010806

[15] Evertse, J.-H., Schlickewei, H. P.: A quantitative version of the absolute subspace theorem. J. Reine Angew. Math. **548**, 21–127 (2002) Zbl 1026.11060 MR 1915209

[16] Faltings, G., Wüstholz, G.: Diophantine approximations on projective spaces. Invent. Math. **116**, 109–138 (1994) Zbl 0805.14011 MR 1253191

[17] Grayson, D. R.: Reduction theory using semistability. Comment. Math. Helv. **59**, 600–634 (1984) Zbl 0564.20027 MR 780079

[18] Kleinbock, D.: An extension of quantitative nondivergence and applications to Diophantine exponents. Trans. Amer. Math. Soc. **360**, 6497–6523 (2008) Zbl 1230.37009 MR 2434296

[19] Kleinbock, D.: An 'almost all versus no' dichotomy in homogeneous dynamics and Diophantine approximation. Geom. Dedicata **149**, 205–218 (2010) Zbl 1223.37005 MR 2737689

[20] Kleinbock, D., Lindenstrauss, E., Weiss, B.: On fractal measures and Diophantine approximation. Selecta Math. (N.S.) **10**, 479–523 (2004) Zbl 1130.11039 MR 2134453

[21] Kleinbock, D., Margulis, G., Wang, J.: Metric Diophantine approximation for systems of linear forms via dynamics. Int. J. Number Theory **6**, 1139–1168 (2010) Zbl 1211.11086 MR 2679461

[22] Kleinbock, D., Moshchevitin, N., Weiss, B.: Singular vectors on manifolds and fractals. Israel J. Math. **245**, 589–613 (2021) Zbl 07513370 MR 4358257

[23] Kleinbock, D., Shi, R., Tomanov, G.: S-adic version of Minkowski's geometry of numbers and Mahler's compactness criterion. J. Number Theory **174**, 150–163 (2017) Zbl 1401.11112 MR 3597385

[24] Kleinbock, D., Tomanov, G.: Flows on S-arithmetic homogeneous spaces and applications to metric Diophantine approximation. Comment. Math. Helv. **82**, 519–581 (2007) Zbl 1135.11037 MR 2314053

[25] Kleinbock, D. Y., Margulis, G. A.: Flows on homogeneous spaces and Diophantine approximation on manifolds. Ann. of Math. (2) **148**, 339–360 (1998) Zbl 0922.11061 MR 1652916

[26] Lang, S.: Algebraic number theory. 2nd ed., Graduate Texts in Mathematics 110, Springer, New York (1994) Zbl 0811.11001 MR 1282723

[27] Lindenstrauss, E., Margulis, G., Mohammadi, A., Shah, N.: Quantitative behavior of unipotent flows and an effective avoidance principle. arXiv:1904.00290 (2019)

[28] Mahler, K.: On compound convex bodies. I. Proc. London Math. Soc. (3) **5**, 358–379 (1955) Zbl 0065.28002 MR 74460

[29] Margulis, G. A.: The action of unipotent groups in a lattice space. Mat. Sb. (N.S.) **86(128)**, 552–556 (1971) (in Russian) Zbl 0257.54038 MR 0291352

[30] Randriambololona, H.: Harder–Narasimhan theory for linear codes (with an appendix on Riemann–Roch theory). J. Pure Appl. Algebra **223**, 2997–3030 (2019) Zbl 1409.14047 MR 3912956

[31] Roth, K. F.: Rational approximations to algebraic numbers. Mathematika **2**, 1–20; corrigendum, 168 (1955) Zbl 0064.28501 MR 72182

[32] Roy, D.: On Schmidt and Summerer parametric geometry of numbers. Ann. of Math. (2) **182**, 739–786 (2015) Zbl 1328.11076 MR 3418530

[33] de Saxcé, N.: Groupes arithmétiques et approximation diophantienne. Mémoire d'habilitation (2020)

[34] Schlickewei, H. P.: The p-adic Thue–Siegel–Roth–Schmidt theorem. Arch. Math. (Basel) **29**, 267–270 (1977) Zbl 0365.10026 MR 491529

[35] Schmidt, W. M.: Linear forms with algebraic coefficients. I. J. Number Theory **3**, 253–277 (1971) Zbl 0221.10034 MR 308061

[36] Schmidt, W. M.: Norm form equations. Ann. of Math. (2) **96**, 526–551 (1972) Zbl 0226.10024 MR 314761

[37] Schmidt, W. M.: Diophantine approximation. Lecture Notes in Mathematics 785, Springer, Berlin (1980) Zbl 0421.10019 MR 568710

[38] Schmidt, W. M.: Vojta's refinement of the subspace theorem. Trans. Amer. Math. Soc. **340**, 705–731 (1993) Zbl 0794.11030 MR 1152325

[39] Siegel, C.: Approximation algebraischer Zahlen. Math. Z. **10**, 173–213 (1921) JFM 48.0163.07 MR 1544471

[40] Stuhler, U.: Eine Bemerkung zur Reduktionstheorie quadratischer Formen. Arch. Math. (Basel) **27**, 604–610 (1976) Zbl 0338.10024 MR 424707

[41] Thue, A.: Über Annäherungswerte algebraischer Zahlen. J. Reine Angew. Math. **135**, 284–305 (1909) JFM 40.0265.01 MR 1580770

[42] Vojta, P.: A refinement of Schmidt's subspace theorem. Amer. J. Math. **111**, 489–518 (1989) Zbl 0662.14002 MR 1002010

[43] Zannier, U.: Some Applications of Diophantine Approximation to Diophantine Equations (with special emphasis on the Schmidt Subspace Theorem). Forum Udine (2003)