FLOWS ON HOMOGENEOUS SPACES AND DIOPHANTINE APPROXIMATION ON MANIFOLDS

D. Y. Kleinbock
Institute for Advanced Study

and G. A. Margulis
Yale University

January 8, 1997

Abstract. We present a new approach to metric Diophantine approximation on manifolds based on the correspondence between approximation properties of numbers and orbit properties of certain flows on homogeneous spaces. This approach yields a new proof of a conjecture of Mahler, originally settled by V. G. Sprindžuk in 1964. We also prove several related hypotheses of Baker and Sprindžuk formulated in 1970s. The core of the proof is a theorem which generalizes and sharpens earlier results on non-divergence of unipotent flows on the space of lattices.

1. Introduction

We start by recalling several basic facts from the theory of simultaneous Diophantine approximation. For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ we let

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i, \quad \|\mathbf{x}\| = \max_{1 \le i \le n} |x_i|,$$

$$\Pi(\mathbf{x}) = \prod_{i=1}^{n} |x_i| \quad \text{and} \quad \Pi_+(\mathbf{x}) = \prod_{i=1}^{n} |x_i|_+,$$

where $|x|_+$ stands for $\max(|x|, 1)$. One says that a vector $\mathbf{y} \in \mathbb{R}^n$ is *very well approximable* (cf. [S2]), to be abbreviated as VWA, if the following two equivalent conditions are satisfied:

(i) for some $\varepsilon > 0$ there are infinitely many $\mathbf{q} \in \mathbb{Z}^n$ such that

$$|\mathbf{q} \cdot \mathbf{y} + p| \cdot \|\mathbf{q}\|^n \le \|\mathbf{q}\|^{-n\varepsilon} \tag{1.1}$$

for some $p \in \mathbb{Z}$;

1991 *Mathematics Subject Classification.* Primary 11J13, 11J83; Secondary 22E99, 57S25.

The work of the first author was supported in part by NSF Grant DMS-9304580 and that of the second author by NSF Grant DMS-9424613.

(ii) for some $\varepsilon > 0$ there are infinitely many $q \in \mathbb{Z}$ such that

$$\|q\mathbf{y} + \mathbf{p}\|^n \cdot |q| \le |q|^{-\varepsilon} \tag{1.2}$$

for some $\mathbf{p} \in \mathbb{Z}^n$.

The equivalence of (1.1) and (1.2) follows from Khintchine's Transference Principle, see e.g. [C, Chapter V]. It is well known (see [C, Chapter VII]) that almost every $\mathbf{y} \in \mathbb{R}^n$ is not VWA. A more difficult question arises if one considers almost all points $\mathbf{y}$ on a submanifold M of $\mathbb{R}^n$ (in the sense of the natural measure on M). In 1932 K. Mahler [Mah] conjectured that almost all points on the curve

$$M = \{(x, x^2, \dots, x^n) \mid x \in \mathbb{R}\} \tag{1.3}$$

are not very well approximable. V. Sprindžuk's proof of this conjecture (see [Sp1, Sp2]) has eventually led to the development of a new branch in approximation theory, usually referred to as "Diophantine approximation with dependent quantities" or "Diophantine approximation on manifolds". According to Sprindžuk's terminology, a submanifold $M \subset \mathbb{R}^n$ is called *extremal* if almost all $\mathbf{y} \in M$ are not VWA. Since 1964, extensive classes of extremal manifolds have been found by Sprindžuk and his followers, see [Sp3, Sp4] for a review. However, the following conjecture made by Sprindžuk in 1980 has remained unproved:

Conjecture $\mathbf{H_1}$ [Sp4]. *Let $f_1, \dots, f_n$ be real analytic functions in $\mathbf{x} \in U$, U a domain in $\mathbb{R}^d$, which together with 1 are linearly independent over $\mathbb{R}$. Then the manifold $M = \{\mathbf{f}(\mathbf{x}) \mid \mathbf{x} \in U\}$ is extremal.*

The case $n = 2$ was settled by Schmidt in 1964 [S1], and Schmidt's method was recently extended to the case $n = 3$ by V. Beresnevich and V. Bernik [BB].

Note that the validity of the above conjecture for polynomials of one variable follows from Sprindžuk's original proof of Mahler's Conjecture. However, there exists a strengthening of Conjecture H_1 which has not been proved even for the curve (1.3). Namely, one says that $\mathbf{y} \in \mathbb{R}^n$ is *very well multiplicatively approximable* (to be abbreviated as VWMA) if the following two equivalent conditions are satisfied:

(i) for some $\varepsilon > 0$ there are infinitely many $\mathbf{q} \in \mathbb{Z}^n$ such that

$$|\mathbf{q} \cdot \mathbf{y} + p| \cdot \Pi_+(\mathbf{q}) \le \Pi_+(\mathbf{q})^{-\varepsilon} \tag{1.1M}$$

for some $p \in \mathbb{Z}$;

(ii) for some $\varepsilon > 0$ there are infinitely many $q \in \mathbb{Z}$ such that

$$\Pi(q\mathbf{y} + \mathbf{p}) \cdot |q| \le |q|^{-\varepsilon} \tag{1.2M}$$

for some $\mathbf{p} \in \mathbb{Z}^n$.

In other words, instead of taking the nth power of the maximum of the "differences" $|qy_i + p_i|$ in (1.2) one simply multiplies them, and instead of taking the nth power of the maximum of the numbers $|q_i|$ in (1.1) one multiplies those of them which are different from zero. The equivalence of (1.1M) and (1.2M) can be shown by modification of the argument needed to prove the standard form of Khintchine's Transference Principle, see a remark in [Sp3, p. 69]. A manifold $M \subset \mathbb{R}^n$ is said to be *strongly extremal* if almost all $\mathbf{y} \in M$ are not VWMA. Clearly $(1.i)$ implies $(1.i\mathrm{M})$, $i = 1, 2$; therefore any strongly extremal manifold is extremal. In his book [B], A. Baker raised the question of confirming the strong extremality of the curve (1.3). Later, in his review [Sp4], Sprindžuk also stated

Conjecture H_2. *Any manifold $M \subset \mathbb{R}^n$ satisfying the assumptions of Conjecture* H_1 *is strongly extremal.*

We remark that the validity of the above conjecture for polynomials with $n \le 4$ is the recent result of V. Bernik and V. Borbat (Bernik, private communication). See also [VC] and references therein for other special cases.

In this paper we present a proof of the above conjectures as well as some of their generalizations. More precisely, consider a d-dimensional submanifold $M = \{\mathbf{f}(\mathbf{x}) \mid \mathbf{x} \in U\}$ of $\mathbb{R}^n$, where U is an open subset of $\mathbb{R}^d$ and $\mathbf{f} = (f_1, \dots, f_n)$ is a C^m map from U to $\mathbb{R}^n$. For $l \le m$, say that $\mathbf{y} = \mathbf{f}(\mathbf{x})$ is an *l-nondegenerate point* of M if the space $\mathbb{R}^n$ is spanned by partial derivatives of $\mathbf{f}$ at $\mathbf{x}$ of order up to l. We will say that $\mathbf{y}$ is *nondegenerate* if it is l-nondegenerate for some l. One can view this condition as an infinitesimal version of not lying in any proper affine hyperplane, i.e. of the linear independence of $1, f_1, \dots, f_n$ over $\mathbb{R}$. Indeed, if the functions f_i are analytic, it is easy to see that the linear independence of $1, f_1, \dots, f_n$ over $\mathbb{R}$ in a domain U is equivalent to all points of $M = \mathbf{f}(U)$ being nondegenerate. Thus Conjecture H_2 would follow from

Theorem A. *Let $f_1, \dots, f_n \in C^m(U)$, U an open subset of $\mathbb{R}^d$, be such that a.e. point of $M = \{\mathbf{f}(\mathbf{x}) \mid \mathbf{x} \in U\}$ is nondegenerate. Then M is strongly extremal.*

Note that if $n = 2$ and $d = 1$, the above assumption says that the curve $\big\{\big(f_1(x), f_2(x)\big) \mid x \in U\big\}$, $U \subset \mathbb{R}$, has nonzero curvature almost everywhere. This was exactly the assumption used by Schmidt [S1], and the strong extremality of such a curve is the subject of Theorem 14 from [Sp3, Chapter 2].

Our proof of Theorem A is based on the correspondence (cf. [D1, K1, K2]) between approximation properties of numbers $\mathbf{y} \in \mathbb{R}^n$ and behavior of certain orbits in the space of unimodular lattices in $\mathbb{R}^{n+1}$. More precisely, one has to act on the lattice

$$\Lambda_{\mathbf{y}} = \begin{pmatrix} 1 & \mathbf{y}^T \\ 0 & I_n \end{pmatrix} \mathbb{Z}^{n+1} \tag{1.4}$$

by semisimple elements of the form

$$g_{\mathbf{t}} = \operatorname{diag}(e^t, e^{-t_1}, \dots, e^{-t_n}),\ \mathbf{t} = (t_1, \dots, t_n),\ t_i > 0,\ t = \sum_{i=1}^n t_i \tag{1.5}$$

(the latter notation will be used throughout the paper, so that whenever t and $\mathbf{t}$ appear in the same formula, t will stand for $\sum_{i=1}^n t_i$).

Define a function δ on the space of lattices by

$$\delta(\Lambda) \stackrel{\text{def}}{=} \inf_{\mathbf{v} \in \Lambda \smallsetminus \{0\}} \|\mathbf{v}\| \tag{1.6}$$

(note that the ratio of $1 + \log\big(1/\delta(\Lambda)\big)$ and $1 + \operatorname{dist}(\Lambda, \mathbb{Z}^{n+1})$ is bounded between two positive constants for any right invariant Riemannian metric "dist" on the space of lattices). We prove in §2 that for any very well multiplicatively approximable $\mathbf{y} \in \mathbb{R}^n$ there exists $\gamma > 0$ and infinitely many $\mathbf{t} \in \mathbb{Z}_+^n$ such that

$$\delta(g_{\mathbf{t}}\Lambda_{\mathbf{y}}) \le e^{-\gamma t}; \tag{1.7}$$

in other words, if $\mathrm{dist}(g_{\mathbf{t}}\Lambda_{\mathbf{y}}, \mathbb{Z}^{n+1})$ grows sublinearly as a function of $\mathbf{t}$, then $\mathbf{y}$ is not VWMA. Thus to prove the strong extremality of M as in Theorem A it is enough to show that for any nondegenerate point $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0)$ there is a neighborhood B of $\mathbf{x}_0$ in U such that for almost all $\mathbf{y} \in \mathbf{f}(B)$ and any $\gamma > 0$, there are at most finitely many $\mathbf{t} \in \mathbb{Z}_+^n$ such that (1.7) holds.

In view of Borel-Cantelli, the latter can be done by estimating the measure of sets

$$E_{\mathbf{t}} \stackrel{\text{def}}{=} \{\mathbf{x} \in B \mid \delta(g_{\mathbf{t}}\Lambda_{\mathbf{f}(\mathbf{x})}) \le e^{-\gamma t}\} \tag{1.8}$$

for any fixed $\mathbf{t} \in \mathbb{Z}_+^n$, so that

$$\sum_{\mathbf{t} \in \mathbb{Z}_+^n} |E_{\mathbf{t}}| < \infty \tag{1.9}$$

(here and hereafter $|\cdot|$ stands for the Lebesgue measure). Such estimates are obtained in §5 by modifying proofs of earlier results on non-divergence of unipotent flows in the space of lattices. According to the theorem of Dani [D2] strengthening earlier result of Margulis [Mar], for any $c > 0$ and any lattice Λ in $\mathbb{R}^k$ there exists $\varepsilon > 0$ such that for any unipotent subgroup $\{u_x | x \in \mathbb{R}\}$ of $SL_k(\mathbb{R})$ one has

$$\Big|\{x \in [0,T] \mid \delta(u_x\Lambda) < \varepsilon\}\Big| \le cT\,. \tag{1.10}$$

Similar estimates are known for any polynomial map from $\mathbb{R}^d$ to $GL_k(\mathbb{R})$ instead of $x \to u_x$, cf. [Sh] (or [EMS] for a bigger class of maps). In §3 (see Proposition 3.4 for a precise statement) we show that if $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0)$ is nondegenerate, $\mathbf{x}_0$ has a neighborhood on which linear combinations of $1, f_1, \dots, f_n$ behave like polynomials of uniformly bounded degree. Then in §§ 4 and 5 we modify the argument of Margulis and Dani in order to get a quantitative relation between c and ε in the analogue of (1.10) (see Proposition 2.3) which will guarantee convergence in (1.9). The last section of the paper deals with several possible extensions of the main result, as well as some open questions.

2. Reduction to a statement about lattices

Given a vector $\mathbf{y} \in \mathbb{R}^n$, consider a lattice $\Lambda_{\mathbf{y}}$ in $\mathbb{R}^{n+1}$ defined as in (1.4). Note that elements of $\Lambda_{\mathbf{y}}$ are of the form

$$(\mathbf{q} \cdot \mathbf{y} + p,\ q_1, \dots, q_n)\,, \tag{2.1}$$

where $p \in \mathbb{Z}$ and $\mathbf{q} = (q_1, \dots, q_n) \in \mathbb{Z}^n$. We are going to consider the action of the semigroup $\{g_{\mathbf{t}}\}$ as in (1.5) on the space of unimodular lattices in $\mathbb{R}^{n+1}$, i.e. the homogeneous space $SL_{n+1}(\mathbb{R})/SL_{n+1}(\mathbb{Z})$. The function δ defined in (1.6) will be used to "describe the structure of the space of lattices at ∞": by Mahler's Compactness Criterion (see [R, Corollary 10.9]) a sequence Λ_k tends to infinity in $SL_{n+1}(\mathbb{R})/SL_{n+1}(\mathbb{Z})$ iff $\delta(\Lambda_k) \to 0$ as $k \to \infty$.

The next lemma helps one to reduce problems involving approximation properties of $\mathbf{y}$ to studying the rate, calculated in terms of the function δ, with which orbit points $\{g_{\mathbf{t}}\Lambda_{\mathbf{y}}\}$ may go to infinity.

Lemma 2.1. *Let $\varepsilon > 0$, $\mathbf{y} \in \mathbb{R}^n$ and $(p, \mathbf{q}) \in \mathbb{Z}^{n+1}$ be such that* (1.1M) *holds. Put*

$$r = \Pi_+(\mathbf{q})^{-\frac{\varepsilon}{n+1}}\,, \tag{2.2a}$$

and for $i = 1, \dots, n$ define $t_i > 0$ by

$$|q_i|_+ = re^{t_i}\,. \tag{2.2b}$$

Then $\delta(g_{\mathbf{t}}\Lambda_{\mathbf{y}}) \le r$.

Proof. In view of (2.1) and (1.5), the inequality to be proved is equivalent to

$$e^t|\mathbf{q}\cdot\mathbf{y} + p| \le r \tag{2.3a}$$

and

$$e^{-t_i}|q_i| \le r,\ \ i = 1, \dots, n\,. \tag{2.3b}$$

One immediately gets (2.3b) from (2.2b) and $|q_i| \le |q_i|_+$, and

$$\Pi_+(\mathbf{q}) = r^n e^t \tag{2.4}$$

as a product of the equalities (2.2b). Thus (1.1M) can be written in the form

$$e^t|\mathbf{q}\cdot\mathbf{y} + p| \le r^{-n}\Pi_+(\mathbf{q})^{-\varepsilon} \underset{\text{by (2.2a)}}{=} r^{-n}\cdot r^{n+1}\,,$$

which proves (2.3a). □

An elementary computation shows that (2.2a) and (2.4) can be combined to yield $r = e^{-\gamma t}$, where

$$\gamma = \frac{\varepsilon}{n + 1 + n\varepsilon}\,. \tag{2.5}$$

Therefore one has

Corollary 2.2. *Assume that $\mathbf{y} \in \mathbb{R}^n$ is VWMA. Then for some $\gamma > 0$ there are infinitely many $\mathbf{t} \in \mathbb{Z}^n_+$ such that* (1.7) *holds.*

Proof. By definition, for some $\varepsilon > 0$ there are infinitely many solutions $\mathbf{q} \in \mathbb{Z}^n$ of (1.1M). Therefore, by the above lemma and with γ as in (2.5), there exists a sequence $\mathbf{t}^{(k)} \in \mathbb{R}^n_+$ with $t^{(k)} = \sum_{i=1}^n t_i^{(k)} \to +\infty$ such that $\delta(g_{\mathbf{t}^{(k)}}\Lambda_{\mathbf{y}}) \le e^{-\gamma t^{(k)}}$. Denote by $[\mathbf{t}^{(k)}]$ the vector consisting of integer parts of $t_i^{(k)}$, then clearly the ratio of $\delta(g_{\mathbf{t}^{(k)}}\Lambda_{\mathbf{y}})$ and $\delta(g_{[\mathbf{t}^{(k)}]}\Lambda_{\mathbf{y}})$ is bounded from above by

$$\|g_{\mathbf{t}^{(k)}}g^{-1}_{[\mathbf{t}^{(k)}]}\| = \|g_{\mathbf{t}^{(k)}-[\mathbf{t}^{(k)}]}\| \le e^n\,.$$

Thus one gets infinitely many solutions $\mathbf{t} \in \mathbb{Z}^n_+$ of $\delta(g_{\mathbf{t}}\Lambda_{\mathbf{y}}) \le e^n e^{-\gamma t}$, and a slight variation of γ yields infinitely many solutions of (1.7). □

Now we are ready to start the

Proof of Theorem A. We are given an open subset U of $\mathbb{R}^d$ and a smooth manifold M parametrized by $\mathbf{f} : U \to \mathbb{R}^n$. It is clearly enough to prove that almost every point of M has a neighborhood $W \subset M$ such that the set $\{\mathbf{y} \in W \mid \mathbf{y}$ is not VWMA$\}$ has full measure. By virtue of Corollary 2.2, it suffices to show that for a.e. $\mathbf{y}_0 \in M$ there exists a neighborhood $W \subset M$ of $\mathbf{y}_0$ such that for any $\gamma > 0$ and a.e. $\mathbf{y} \in W$, the inequality (1.7) holds for at most finitely many $\mathbf{t} \in \mathbb{Z}^n_+$.

We now state the proposition, to be proved in §5, which will easily imply the above statement.

Proposition 2.3. *Let $\mathbf{f}$ be a C^l map from an open subset U of $\mathbb{R}^d$ to $\mathbb{R}^n$, and let $\mathbf{x}_0 \in U$ be such that $\mathbb{R}^n$ is spanned by partial derivatives of $\mathbf{f}$ at $\mathbf{x}_0$ of order up to l. Then there exists a ball $B \subset U$ centered in $\mathbf{x}_0$ and positive constants D and ρ such that for any $t_1, \dots, t_n \geq 0$ and $0 < \varepsilon \leq \rho$ one has*

$$\big|\{\mathbf{x} \in B \mid \delta\big(g_{\mathbf{t}}\Lambda_{\mathbf{f}(\mathbf{x})}\big) < \varepsilon\}\big| \leq D\left(\frac{\varepsilon}{\rho}\right)^{1/dl}|B|\,. \tag{2.6}$$

Indeed, by the assumptions on M, almost every $\mathbf{y}_0 \in M$ is nondegenerate, hence has a neighborhood $W = \mathbf{f}(B)$ with B as in the above proposition. For $\gamma > 0$ and $\mathbf{t} \in \mathbb{Z}_+^n$ consider the set $E_{\mathbf{t}}$ defined in (1.8). Then there is a constant $\tilde{D}$ independent on $\mathbf{t}$ such that for large enough $\|\mathbf{t}\|$ one has $|E_{\mathbf{t}}| \leq \tilde{D}e^{-\gamma t/dl}$. Therefore the series $\sum_{\mathbf{t}\in\mathbb{Z}_+^n} |E_{\mathbf{t}}|$ converges, hence, by the Borel-Cantelli Lemma, for a.e. $\mathbf{x} \in B$ the inequality (1.7), with $\mathbf{y} = \mathbf{f}(\mathbf{x})$, holds for at most finitely many $\mathbf{t} \in \mathbb{Z}_+^n$. Since the measure on M is obtained from the Lebesgue measure on U by a smooth map, this finishes the proof of Theorem A modulo Proposition 2.3. $\square$

3. Good functions and nondegenerate points

Let V be a subset of $\mathbb{R}^d$ and f a continuous function on V. In what follows, we will let $\|f\|_B \overset{\text{def}}{=} \sup_{\mathbf{x}\in B} |f(\mathbf{x})|$ for a subset B of V. For positive numbers C and α, say that f is *(C, α)-good on V* if for any open ball $B \subset V$ one has[1]

$$\forall\, \varepsilon > 0 \quad \big|\{\mathbf{x} \in B \;\big|\; |f(\mathbf{x})| < \varepsilon\}\big| \leq C \cdot \left(\frac{\varepsilon}{\|f\|_B}\right)^{\alpha} \cdot |B|\,. \tag{3.1}$$

Cf. [EMS, Corollary 2.10] for a similar condition. The properties listed below follow immediately from the definition.

Lemma 3.1. *Let $V \subset \mathbb{R}^d$ and $C, \alpha > 0$ be given.*
(a) *f is (C, α)-good on V $\Leftrightarrow$ so is $|f|$;*
(b) *f is (C, α)-good on V $\Rightarrow$ so is λf $\forall\, \lambda \in \mathbb{R}$;*
(c) *f_i, $i \in I$, are (C, α)-good on V $\Rightarrow$ so is $\sup_{i\in I} |f_i|$;*

We now consider what can be called a model example of good functions.

Proposition 3.2 (cf. [DM, Lemma 4.1]). *For any $k \in \mathbb{N}$, any polynomial $f \in \mathbb{R}[x]$ of degree not greater than k is $\big(2k(k+1)^{1/k}, 1/k\big)$-good on $\mathbb{R}$.*

Proof. Fix an open interval $B \subset \mathbb{R}$, a polynomial $f \in \mathbb{R}[x]$ of degree not exceeding k and an $\varepsilon > 0$. We need to show that

$$\big|\{x \in B \;\big|\; |f(x)| < \varepsilon\}\big| \leq 2k(k+1)^{1/k}\left(\frac{\varepsilon}{\|f\|_B}\right)^{1/k}|B|\,. \tag{3.2}$$

Denote $|B|$ by b and $\dfrac{1}{b}\big|\{x \in B \;\big|\; |f(x)| < \varepsilon\}\big|$ by σ. Then there exist $x_1, \dots, x_{k+1} \in B$ such that $|f(x_i)| \leq \varepsilon$, $1 \leq i \leq k+1$, and $|x_i - x_j| \geq \sigma b/2k$, $1 \leq i < j \leq k+1$.

[1] Here we adopt the convention $\frac{1}{0} = \infty$, so that (3.1) holds if $f|_B \equiv 0$.

Using Lagrange's interpolation formula, one can write down the exact expression for f:

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

and conclude that

$$\|f\|_B \le (k+1)\varepsilon \frac{b^k}{(\sigma b/2k)^k}\,, \tag{3.3b}$$

which immediately implies (3.2). $\square$

The goal of this section is to prove that if the point $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0)$ of the manifold $M = \mathbf{f}(U) \subset \mathbb{R}^n$, $U \subset \mathbb{R}^d$, is nondegenerate, then there exist a neighborhood V of $\mathbf{x}_0$ and positive constants C and α such that all linear combinations of $1, f_1, \dots, f_n$ are (C,α)-good on V. Our argument is based on the following standard but quite technical lemma. For $i = 1, \dots, d$, we let ∂_i be the operator of partial differentiation with respect to x_i, and for a *multiindex* $\beta = (i_1, \dots, i_d)$, $i_j \in \mathbb{Z}_+$, we let $|\beta| = \sum_{j=1}^d i_j$ and $\partial_\beta = \partial_1^{i_1} \circ \cdots \circ \partial_d^{i_d}$.

Lemma 3.3. *Let V be an open subset of $\mathbb{R}^d$, and let $f \in C^k(V)$ be such that for some constants $A_1, A_2 > 0$ one has*

$$\|\partial_\beta f\|_V \le A_1 \quad \forall\, \beta \text{ with } |\beta| \le k\,, \tag{3.4$\le$}$$

and

$$|\partial_i^k f(x)| \ge A_2 \quad \forall\, x \in V,\ i = 1, \dots, d\,. \tag{3.4$\ge$}$$

Then for any d-dimensional cube $B = B_1 \times \cdots \times B_d \subset V$, where B_i are open intervals of the same length, and for any $\varepsilon > 0$ one has

$$\big|\{\mathbf{x} \in B \mid |f(\mathbf{x})| < \varepsilon\}\big| \le d C_{k,A_1,A_2} \left(\frac{\varepsilon}{\|f\|_B}\right)^{1/dk} |B|\,, \tag{3.5}$$

with

$$C_{k,A_1,A_2} = k(k+1)\left(\frac{A_1}{A_2}(k+1)(2k^k+1)\right)^{1/k}.$$

Proof. For B, f as above and $\varepsilon > 0$, we will denote by b the sidelength of B and by s the supremum of $|f|$ over B. First consider the case $d = 1$. Take $\varepsilon > 0$; since, by (3.4$\ge$), the kth derivative of f does not vanish on B, the set $\{x \in B \mid |f(x)| < \varepsilon\}$ consists of at most $k+1$ intervals. Let I be the maximal of those, then

$$\big|\{x \in B \mid |f(x)| < \varepsilon\}\big| \le (k+1)|I|\,, \tag{3.6}$$

so it suffices to estimate $|I|$ from above. Denote by r the length of I and also let

$$\tau = k!s/b^k\,. \tag{3.7}$$

The argument will be based on the following two inequalities:

Sublemma 3.3.1. (a) $r \le \varphi_1(\tau)\left(\frac{\varepsilon}{s}\right)^{1/k} b$, *where* $\varphi_1(\tau) = k\left(\frac{k+1}{A_2}\right)^{1/k}\tau^{1/k}$;

(b) $r \le \varphi_2(\tau)\left(\frac{\varepsilon}{s}\right)^{1/k} b$, *where* $\varphi_2(\tau) = k^2\left(\frac{\frac{2A_1}{A_2}(k+1)}{1-A_1/\tau}\right)^{1/k}$.

Proof. Divide I into k equal parts by points $x_1,\dots,x_{k+1}$, and let P be the Lagrange polynomial of degree k formed by using values of f at these points, i.e. given by the expression in the r.h.s. of (3.3a). Then there exists $x \in I$ such that $P^{(k)}(x) = f^{(k)}(x)$, hence, by (3.4$\ge$), $|P^{(k)}(x)| \ge A_2$. On the other hand, one can differentiate the r.h.s. of (3.3a) k times to get $|P^{(k)}(x)| \le (k+1)\frac{\varepsilon k!}{(r/k)^k}$. Combining the last two inequalities, one obtains

$$r \le k\left(\frac{(k+1)!}{A_2}\varepsilon\right)^{1/k} = k\left(\frac{(k+1)!}{A_2}\frac{s}{b^k}\right)^{1/k}\left(\frac{\varepsilon}{s}\right)^{1/k} b\,, \tag{3.8}$$

which reduces to the inequality (a) by the substitution (3.7).

Next, let Q be the Taylor polynomial of f at x_1 of degree $k-1$. By Taylor's formula,

$$\|f-Q\|_I \le \|f^{(k)}\|_I\frac{r^k}{k!} \underset{\text{by } (3.4\le)}{\le} \frac{A_1 r^k}{k!} \underset{\text{by } (3.8)}{\le} \frac{A_1}{A_2}k^k(k+1)\varepsilon\,,$$

therefore $\|Q\|_I \le 2\frac{A_1}{A_2}k^k(k+1)\varepsilon$.

We now apply Lagrange's formula to reconstruct Q on B by its values at $x_1,\dots,x_k$. Similarly to (3.3b), we get

$$\|Q\|_B \le k\cdot 2\frac{A_1}{A_2}k^k(k+1)\varepsilon\frac{b^{k-1}}{(r/k)^{k-1}} \le 2\frac{A_1}{A_2}k^{2k}(k+1)\varepsilon\frac{b^k}{r^k}\,. \tag{3.9}$$

Finally, the difference between f and Q on B is, again by (3.4$\le$), bounded from above by $A_1 b^k/k!$, so from (3.9) one deduces that

$$s \le \left(2\frac{A_1}{A_2}k^{2k}(k+1)\frac{\varepsilon}{r^k} + \frac{A_1}{k!}\right)b^k\,,$$

which is equivalent to the inequality (b) modulo (3.7). $\square$

To use the above estimates, note that φ_1 (resp. φ_2) is an increasing (resp. decreasing) function, therefore one has $r \le \varphi_1(\tau_0)\left(\frac{\varepsilon}{s}\right)^{1/k} b$, where τ_0 is the root of $\varphi_1(\tau) = \varphi_2(\tau)$. An elementary computation yields $\tau_0 = A_1(2k^k+1)$, and the validity of (3.5) for $d=1$ follows immediately from (3.6).

Next we argue by induction on d, assuming that (3.5) is established for all the lower values of d. We represent $\mathbf{x} \in \mathbb{R}^d$ as $(x_1,\mathbf{x}')$ and let B' be the product of the intervals $B_2,\dots,B_d$. Denote by σ the value $\frac{1}{b^d}\left|\{\mathbf{x}\in B \mid |f(\mathbf{x})| < \varepsilon\}\right|$. Choose a positive number $\lambda < \sigma$, and let

$$B_1(\lambda) \overset{\text{def}}{=} \left\{x_1 \in B_1 \,\middle|\, \frac{1}{b^{d-1}}\left|\{\mathbf{x}'\in B' \mid f(x_1,\mathbf{x}')| < \varepsilon\}\right| \ge \lambda\right\}.$$

The induction step will be based on the following

Sublemma 3.3.2. $\frac{1}{b}|B_1(\lambda)| \geq \sigma - \lambda.$

Proof. Write

$$\{\mathbf{x} \in B \mid |f(\mathbf{x})| < \varepsilon\} \subset B_1(\lambda) \times B' \ \cup \ \left\{\mathbf{x} \in \left(B_1 \smallsetminus B_1(\lambda)\right) \times B' \mid |f(\mathbf{x})| < \varepsilon\right\},$$

and put $\mu = \frac{1}{b}|B_1(\lambda)|$. Then, by Fubini, $b^d\sigma = \left|\{\mathbf{x} \in B \mid |f(\mathbf{x})| < \varepsilon\}\right|$ is not greater than $b\mu \cdot b^{d-1} + b(1-\mu)\cdot b^{d-1}\lambda \leq b^d(\lambda + \mu)$, hence the claim. $\square$

For $x_1 \in B_1$, denote by f_{x_1} the function on B' given by $f_{x_1}(\mathbf{x}') \stackrel{\text{def}}{=} f(x_1, \mathbf{x}')$. Clearly such functions satisfy assumptions (3.4) on derivatives with $d-1$ in place of d, therefore, by the induction assumption, for any $x_1 \in B_1(\lambda)$ one has

$$\lambda \leq \frac{1}{b^{d-1}}\left|\{\mathbf{x}' \in B' \mid f_{x_1}(\mathbf{x}')| < \varepsilon\}\right| \leq (d-1)C_{k,A_1,A_2}\left(\frac{\varepsilon}{\|f_{x_1}\|_{B'}}\right)^{1/(d-1)k},$$

or

$$\|f_{x_1}\|_{B'} \leq \left(\frac{(d-1)C_{k,A_1,A_2}}{\lambda}\right)^{(d-1)k}\varepsilon\,.$$

On the other hand, for any $\mathbf{x}' \in B'$ the functions $x_1 \to f(x_1, \mathbf{x}')$ also satisfy the assumptions on derivatives (with 1 in place of d), hence are $(C_{k,A_1,A_2}, 1/k)$-good on B_1. By Lemma 3.1(c), the function $x_1 \to \|f_{x_1}\|_{B'}$ is also $(C_{k,A_1,A_2}, 1/k)$-good on B_1, thus, by the above sublemma,

$$\begin{aligned}\sigma - \lambda &\leq \frac{1}{b}\left|\left\{x_1 \in B_1 \,\middle|\, \|f_{x_1}\|_{B'} \leq \left((d-1)C_{k,A_1,A_2}/\lambda\right)^{(d-1)k}\varepsilon\right\}\right| \\ &\leq C_{k,A_1,A_2}\left(\frac{\left((d-1)C_{k,A_1,A_2}/\lambda\right)^{(d-1)k}\varepsilon}{s}\right)^{1/k},\end{aligned}$$

or equivalently,

$$\lambda^{d-1}(\sigma - \lambda) \leq C^d_{k,A_1,A_2}(d-1)^{d-1}\left(\frac{\varepsilon}{s}\right)^{1/k}. \tag{3.10}$$

The function in the l.h.s. of (3.10) attains its maximum, $\sigma^d(d-1)^{d-1}/d^d$, when $\lambda = \sigma(d-1)/d$; substituting it into (3.10), one easily obtains (3.5). $\square$

We now come to the main result of the section.

Proposition 3.4. *Let $\mathbf{f} = (f_1, \dots, f_n)$ be a C^l map from an open subset U of $\mathbb{R}^d$ to $\mathbb{R}^n$, and let $\mathbf{x}_0 \in U$ be such that $\mathbb{R}^n$ is spanned by partial derivatives of $\mathbf{f}$ at $\mathbf{x}_0$ of order up to l. Then there exists a neighborhood $V \subset U$ of $\mathbf{x}_0$ and positive C such that any linear combination of $1, f_1, \dots, f_n$ is $(C, 1/dl)$-good on V.*

Proof. Take $f = c_0 + \sum_{i=1}^n c_i f_i$; in view of Lemma 3.1(b), one can without loss of generality assume that the norm of $\mathbf{c} = (c_0, \dots, c_n)$ is equal to 1. ¿From the nondegeneracy assumption it follows that there exists a constant $C_1 > 0$ such that for any $\mathbf{c}$ with $\|\mathbf{c}\| = 1$ one can find a multiindex β with $|\beta| = k \le l$ and

$$\Big|\sum_{i=1}^{n} c_i \partial_\beta f_i(\mathbf{x}_0)\Big| = |\partial_\beta f(\mathbf{x}_0)| \ge C_1 \,.$$

By an appropriate rotation of the coordinate system around $\mathbf{x}_0$ one can guarantee that $|\partial_i^k f(\mathbf{x}_0)| \ge C_2$ for all $i = 1, \dots, d$ and some positive C_2 independent of $\mathbf{c}$. Then one uses the continuity of the derivatives of $f_1, \dots, f_n$ to choose a neighborhood $V' \subset U$ of $\mathbf{x}_0$ and positive A_1, A_2 (again independently of $\mathbf{c}$) such that the inequalities (3.4) hold. Now let V be a smaller neighborhood of $\mathbf{x}_0$ such that whenever a ball B lies in V, any cube $\hat{B}$ circumscribed around B is contained in V'. Then for any $\varepsilon > 0$ one has

$$\begin{aligned}\big|\{\mathbf{x}\in B \,\big|\, |f(\mathbf{x})| < \varepsilon\}\big| &\le \Big|\{\mathbf{x}\in \hat{B} \,\big|\, |f(\mathbf{x})| < \varepsilon\}\Big| \\ \text{(by Lemma 3.3)} \quad &\le dC_{k,A_1,A_2}\left(\frac{\varepsilon}{\|f\|_{\hat{B}}}\right)^{1/dk} |\hat{B}| \le \frac{2^d}{v_d} dC_{k,A_1,A_2}\left(\frac{\varepsilon}{\|f\|_{B}}\right)^{1/dk} |B|\end{aligned}$$

(here v_d stands for the volume of the unit ball in $\mathbb{R}^d$), which implies that f is $\left(\frac{2^d}{v_d} dC_{l,A_1,A_2}, 1/dl\right)$-good on V. $\square$

4. Maps of posets into spaces of good functions

The purpose of §§ 4 and 5 is to generalize and strengthen results on nondivergence of unipotent flows on spaces of lattices obtained in [Mar, D2]. The base objects will be mappings of partially ordered sets (*posets*) into spaces of functions on $\mathbb{R}^d$. In what follows, $B(\mathbf{x}, r)$, where $\mathbf{x} \in \mathbb{R}^d$ and $r > 0$, will stand for the open ball of radius r centered in $\mathbf{x}$.

For a partially ordered set (*poset*) S, we will denote by $l(S)$ the *length* of S (i.e. the number of elements in a maximal linearly ordered subset of S). If T is a subset of S, we let $S(T)$ be the set of elements of $S \smallsetminus T$ comparable with any element of T. Note that one always has

$$l\big(S(T)\big) \le l(S) - l(T)\,. \tag{4.1}$$

For $d \in \mathbb{N}$, $k \in \mathbb{Z}_+$ and $C, \alpha, \rho > 0$, define $\mathcal{A}(d, k, C, \alpha, \rho)$ to be the set of triples (S, φ, B) where S is a poset, $B = B(\mathbf{x}_0, r_0)$, where $\mathbf{x}_0 \in \mathbb{R}^d$ and $r_0 > 0$, and φ is a mapping from S to the space of continuous functions on $\tilde{B} \stackrel{\text{def}}{=} B\big(\mathbf{x}_0, 3^k r_0\big)$ (this mapping will be denoted by $s \to \psi_s$) such that the following holds:

(A0) $l(S) \le k$;
(A1) $\forall\, s \in S\,, \quad \psi_s$ is (C, α)-good on $\tilde{B}$;
(A2) $\forall\, s \in S\,, \quad \|\psi_s\|_B \ge \rho$;
(A3) $\forall\, \mathbf{x} \in \tilde{B}, \quad \#\{s \in S \,\big|\, |\psi_s(\mathbf{x})| < \rho\} < \infty$.

Then, given $(S, \varphi, B) \in \mathcal{A}(d, k, C, \alpha, \rho)$ and $\varepsilon > 0$, say that a point $\mathbf{z} \in B$ is $(\varepsilon, S, \varphi)$-*marked* if there exists a linearly ordered subset $\Sigma_{\mathbf{z}}$ of S such that

(M1) $\varepsilon \le |\psi_s(\mathbf{z})| \le \rho \quad \forall\, s \in \Sigma_{\mathbf{z}}$;
(M2) $|\psi_s(\mathbf{z})| \ge \rho \quad \forall\, s \in S(\Sigma_{\mathbf{z}})$.

We will denote by $\Phi(\varepsilon, S, \varphi, B)$ the set of all the $(\varepsilon, S, \varphi)$-marked points $\mathbf{z} \in B$. We will also need to use

Besicovitch's Covering Theorem (see [Mat, Theorem 2.7]). *There is an integer N_d depending only on d with the following property. Let A be a bounded subset of $\mathbb{R}^d$ and let $\mathcal{B}$ be a family of nonempty open balls in $\mathbb{R}^d$ such that each $\mathbf{x} \in A$ is the center of some ball of $\mathcal{B}$. Then there exists a finite or countable subfamily $\{B_i\}$ of $\mathcal{B}$ with $1_A \le \sum_i 1_{B_i} \le N_d$ (i.e. $A \subset \bigcup_i B_i$ and the multiplicity of that subcovering is at most N_d).*

The goal of the section is to prove the following

Theorem 4.1 (cf. [Mar, Main Lemma] or [D2, Proposition 2.7]). *Let $d \in \mathbb{N}$, $k \in \mathbb{Z}_+$ and $C, \alpha, \rho > 0$ be given. Then for all $(S, \varphi, B) \in \mathcal{A}(d, k, C, \alpha, \rho)$ and $\varepsilon > 0$ one has*

$$\left|\left(B \smallsetminus \Phi(\varepsilon, S, \varphi, B)\right)\right| \le kC\left(3^d N_d\right)^k \left(\frac{\varepsilon}{\rho}\right)^\alpha |B|\,.$$

The proof will be built up from Lemmas 4.2–4.6 below. In these lemmas, d, k, C, α and ρ will be as in the above theorem, and $(S, \varphi, B) \in \mathcal{A}(d, k, C, \alpha, \rho)$ will be fixed. We also define

$$H(\mathbf{x}) \stackrel{\text{def}}{=} \{s \in S \mid |\psi_s(\mathbf{x})| < \rho\}$$

for any $\mathbf{x} \in B$ (this is a finite subset of S in view of (A3)), and let

$$E \stackrel{\text{def}}{=} \{\mathbf{x} \in B \mid H(\mathbf{x}) \ne \varnothing\} = \{\mathbf{x} \in B \mid \exists\, s \in S \text{ with } |\psi_s(\mathbf{x})| < \rho\}\,.$$

Lemma 4.2. *Any point of B which does not belong to E is $(\varepsilon, S, \varphi)$-marked for any positive ε. In other words, $B \smallsetminus E \subset \Phi(\varepsilon, S, \varphi, B)$, and therefore*

$$B \smallsetminus \Phi(\varepsilon, S, \varphi, B) \subset E \smallsetminus \Phi(\varepsilon, S, \varphi, B)\,.$$

Proof. Take $\mathbf{z} \in B \smallsetminus E$; by the definition of E, one has $|\psi_s(\mathbf{z})| \ge \rho$ for all $s \in S$. Then one can take $\Sigma_{\mathbf{z}}$ to be the empty set and check that (M1) and (M2) are satisfied. $\square$

The next four lemmas deal with some properties of the set E. Note that if $l(S) = 0$ (which means $S = \varnothing$), one has $H(\mathbf{x}) = \varnothing$ for all $\mathbf{x} \in B$, therefore $E = \varnothing$. Thus in the argument below we will tacitly assume k to be not less than 1.

Take $\mathbf{x} \in E$ and $s \in H(\mathbf{x})$, and define

$$r_{s,\mathbf{x}} \stackrel{\text{def}}{=} \sup\{0 < r \le 2r_0 \mid \|\psi_s\|_{B(\mathbf{x},r)} \le \rho\}\,.$$

It follows from the continuity of functions ψ_s that $r_{s,\mathbf{x}} > 0$. We also let $B_{s,\mathbf{x}} \stackrel{\text{def}}{=} B(\mathbf{x}, r_{s,\mathbf{x}})$.

Lemma 4.3. *For any $\mathbf{x} \in E$ and $s \in H(\mathbf{x})$, one has $\|\psi_s\|_{B_{s,\mathbf{x}}} \geq \rho$.*

Proof. By the definition of $B_{s,\mathbf{x}}$, one has either $r_{s,\mathbf{x}} = 2r_0$ (then $B_{s,\mathbf{x}} \supset B$ and the claim follows from (A2)), or $\|\psi_s\|_{B(\mathbf{x},r)} > \rho \quad \forall\, r > r_{s,\mathbf{x}}$, in which case we are done by the continuity of ψ_s. $\square$

For any $\mathbf{x} \in E$ choose an element $s_\mathbf{x}$ of $H(\mathbf{x})$ such that $r_{s_\mathbf{x},\mathbf{x}} \geq r_{s,\mathbf{x}}$ for all $s \in H(\mathbf{x})$ (this can be done since $H(\mathbf{x})$ is finite). For brevity we will denote $r_{s_\mathbf{x},\mathbf{x}}$ by $r_\mathbf{x}$ and $B_{s_\mathbf{x},\mathbf{x}}$ by $B_\mathbf{x}$. Note that in fact $B_\mathbf{x} = \bigcup_{s\in H(\mathbf{x})} B_{s,\mathbf{x}}$ and $r_\mathbf{x} = \max_{s\in H(\mathbf{x})} r_{s,\mathbf{x}}$; in particular, $r_\mathbf{x} \leq 2r_0$ for any $\mathbf{x} \in E$.

Lemma 4.4. *For any $\mathbf{x} \in E$ and $s \in S$, one has $\|\psi_s\|_{B_\mathbf{x}} \geq \rho$.*

Proof. Assume that $\|\psi_s\|_{B_\mathbf{x}} < \rho$. Then one necessarily has $|\psi_s(\mathbf{x})| < \rho$, therefore $s \in H(\mathbf{x})$ and $B_{s,\mathbf{x}}$ is defined. But $B_{s,\mathbf{x}}$ is contained in $B_\mathbf{x}$, so the claim follows from Lemma 4.3. $\square$

For any $\mathbf{x} \in E$ we let $S_\mathbf{x} \stackrel{\text{def}}{=} S(\{s_\mathbf{x}\})$. The induction procedure of the proof of Theorem 4.1 will be based on the following

Lemma 4.5. *For any $\mathbf{x} \in E$, one has*

$$(S_\mathbf{x}, \varphi|_{S_\mathbf{x}}, B_\mathbf{x}) \in \mathcal{A}(d, k-1, C, \alpha, \rho)\,.$$

Proof. The properties (A0) and (A2) for $(S_\mathbf{x}, \varphi|_{S_\mathbf{x}}, B_\mathbf{x})$ follow from (4.1) and Lemma 4.4 respectively. To prove (A1) and (A3) it suffices to notice that

$$\begin{aligned}\tilde{B}_\mathbf{x} \stackrel{\text{def}}{=} B(\mathbf{x}, 3^{k-1}r_\mathbf{x}) &\subset B(\mathbf{x}_0, 3^{k-1}r_\mathbf{x} + r_0)\\ &\subset B\big(\mathbf{x}_0, (2\cdot 3^{k-1}+1)r_0\big) \subset B(\mathbf{x}_0, 3^k r_0) = \tilde{B}\,. \quad \square\end{aligned}$$

The next lemma gives one a way to prove that a point $\mathbf{z} \in B$ is $(\varepsilon, S, \varphi)$-marked provided it is $(\varepsilon, S_\mathbf{x}, \varphi|_{S_\mathbf{x}})$-marked for some $\mathbf{x} \in E$.

Lemma 4.6. *For $\varepsilon > 0$ and $\mathbf{x} \in E$, let $\mathbf{z} \in B \cap \Phi(\varepsilon, S_\mathbf{x}, \varphi|_{S_\mathbf{x}}, B_\mathbf{x})$ be such that $|\psi_{s_\mathbf{x}}(\mathbf{z})| \geq \varepsilon$. Then $\mathbf{z} \in \Phi(\varepsilon, S, \varphi, B)$. Equivalently,*

$$B \cap \big(B_\mathbf{x} \smallsetminus \Phi(\varepsilon, S, \varphi, B)\big) \subset B_\mathbf{x} \smallsetminus \Phi(\varepsilon, S_\mathbf{x}, \varphi|_{S_\mathbf{x}}, B_\mathbf{x}) \cup \{\mathbf{z} \in B_\mathbf{x} \mid |\psi_{s_\mathbf{x}}(\mathbf{z})| < \varepsilon\}\,.$$

Proof. By the definition of $\Phi(\varepsilon, S_\mathbf{x}, \varphi|_{S_\mathbf{x}}, B_\mathbf{x})$, there exists a linearly ordered subset $\Sigma_{\mathbf{x},\mathbf{z}}$ of $S_\mathbf{x}$ such that

$$\varepsilon \leq |\psi_s(\mathbf{z})| \leq \rho \quad \forall\, s \in \Sigma_{\mathbf{x},\mathbf{z}} \tag{4.2}$$

and

$$|\psi_s(\mathbf{z})| \geq \rho \quad \forall\, s \in S_\mathbf{x}(\Sigma_{\mathbf{x},\mathbf{z}})\,. \tag{4.3}$$

Put $\Sigma_\mathbf{z} \stackrel{\text{def}}{=} \Sigma_{\mathbf{x},\mathbf{z}} \cup \{s_\mathbf{x}\}$. Then $S(\Sigma_\mathbf{z}) = S_\mathbf{x}(\Sigma_{\mathbf{x},\mathbf{z}})$; therefore (M2) immediately follows from (4.3), and, in view of (4.2), it remains to check (M1) for $s = s_\mathbf{x}$. The latter is straightforward: $|\psi_{s_{\mathbf{x}}}(\mathbf{z})|$ is not less than ε by the assumption and is not greater than ρ since $\mathbf{z} \in B_{\mathbf{x}}$. $\square$

We are now ready to give a

Proof of Theorem 4.1. We proceed by induction on k. First take $(S, \varphi, B) \in \mathcal{A}(d, 0, C, \alpha, \rho)$; since the poset S is empty, all points of B are $(\varepsilon, S, \varphi)$-marked for any $\varepsilon > 0$, which means that in the case $k = 0$ the claim is trivial. Now take $j \geq 1$ and suppose that the theorem is proved for $k = j - 1$; put $k = j$ and take $(S, \varphi, B) \in \mathcal{A}(d, k, C, \alpha, \rho)$ and a positive ε.

In view of Lemma 4.2, it suffices to estimate the measure of $E \smallsetminus \Phi(\varepsilon, S, \varphi, B)$. On the other hand, from Lemma 4.6 one deduces that for any $\mathbf{x} \in E$, the measure of the intersection of $B_{\mathbf{x}} \smallsetminus \Phi(\varepsilon, S, \varphi, B)$ with B is not greater than

$$|B_{\mathbf{x}} \smallsetminus \Phi(\varepsilon, S_{\mathbf{x}}, \varphi|_{S_{\mathbf{x}}}, B_{\mathbf{x}})| + \left|\{\mathbf{z} \in B_{\mathbf{x}} \;\middle|\; |\psi_{s_{\mathbf{x}}}(\mathbf{z})| < \varepsilon\}\right| .$$

The first summand is not greater than $(k-1)C\left(3^d N_d\right)^{k-1} \left(\frac{\varepsilon}{\rho}\right)^{\alpha} |B_{\mathbf{x}}|$ by Lemma 4.5 and the induction assumption. The second one, in view of $\psi_{s_{\mathbf{x}}}$ being (C, α)-good on $\tilde{B} \supset B_{\mathbf{x}}$, is not greater than

$$C\left(\frac{\varepsilon}{\|\psi_{s_{\mathbf{x}}}\|_{B_{\mathbf{x}}}}\right)^{\alpha} |B_{\mathbf{x}}| \underset{\text{Lemma 4.4}}{\leq} C\left(\frac{\varepsilon}{\rho}\right)^{\alpha} |B_{\mathbf{x}}| .$$

Consequently, one gets

$$\begin{aligned} \left|B \cap \left(B_{\mathbf{x}} \smallsetminus \Phi(\varepsilon, S, \varphi, B)\right)\right| &\leq C\left((k-1)\left(3^d N_d\right)^{k-1} + 1\right)\left(\frac{\varepsilon}{\rho}\right)^{\alpha} |B_{\mathbf{x}}| \\ &\leq kC\left(3^d N_d\right)^{k-1}\left(\frac{\varepsilon}{\rho}\right)^{\alpha} |B_{\mathbf{x}}| . \end{aligned} \tag{4.4}$$

Now consider the covering $\{B_{\mathbf{x}} \mid \mathbf{x} \in E\}$ of E and, using Besicovitch's Covering Theorem, choose a subcovering $\{B_i\}$ of multiplicity at most N_d. Then

$$\sum_i |B_i| \leq N_d \left|\textstyle\bigcup_i B_i\right| \leq N_d \left|B(\mathbf{x}_0, 3r_0)\right| \leq N_d 3^d |B| . \tag{4.5}$$

The sets $B \cap \left(B_i \smallsetminus \Phi(\varepsilon, S, \varphi, B)\right)$ cover $E \smallsetminus \Phi(\varepsilon, S, \varphi, B)$, therefore

$$\begin{aligned} \left|\left(E \smallsetminus \Phi(\varepsilon, S, \varphi, B)\right| \right. &\leq \sum_i \left|B \cap \left(B_i \smallsetminus \Phi(\varepsilon, S, \varphi, B)\right)\right| \\ \text{(by (4.4))} \quad &\leq kC\left(3^d N_d\right)^{k-1}\left(\frac{\varepsilon}{\rho}\right)^{\alpha} \sum_i |B_i| \\ \text{(by (4.5))} \quad &\leq kC\left(3^d N_d\right)^{k-1}\left(\frac{\varepsilon}{\rho}\right)^{\alpha} N_d 3^d |B| \leq kC\left(3^d N_d\right)^{k}\left(\frac{\varepsilon}{\rho}\right)^{\alpha} |B| , \end{aligned}$$

and the theorem is proven. $\square$

5. Quantitative non-divergence in the space of lattices

We now apply Theorem 4.1 to the poset of discrete subgroups of $\mathbb{Z}^k$, $k \in \mathbb{N}$. We fix a basis $\mathbf{e}_1, \dots, \mathbf{e}_k$ of $\mathbb{R}^k$, and for $I = \{i_1, \dots, i_j\} \subset \{1, \dots, k\}$, $i_1 < i_2 < \cdots < i_j$, we let $\mathbf{e}_I \stackrel{\text{def}}{=} \mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_j} \subset \bigwedge^j(\mathbb{R}^k)$, with the convention $\mathbf{e}_\varnothing = 1$. We extend the norm $\|\cdot\|$ from $\mathbb{R}^k$ to the exterior algebra $\bigwedge(\mathbb{R}^k)$ by $\|\sum_{I\subset\{1,\dots,k\}} w_I \mathbf{e}_I\| = \max_{I\subset\{1,\dots,k\}} |w_I|$.

For a discrete subgroup Γ of $\mathbb{R}^k$, we denote by $\Gamma_\mathbb{R}$ the minimal linear subspace of $\mathbb{R}^k$ containing Γ. If $j = \dim(\Gamma_\mathbb{R})$ is positive, we say that $\mathbf{w} \in \bigwedge^j(\mathbb{R}^k)$ *represents* Γ if $\mathbf{w} = \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_j$, where $\mathbf{v}_1, \dots, \mathbf{v}_j$ is a basis of Γ. Clearly the element representing Γ is defined up to a sign. Therefore it makes sense to define the norm of Γ in the following way:

$$\|\Gamma\| \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } \Gamma = \{0\} \\ \|\mathbf{w}\| & \text{if } \mathbf{w} \text{ represents } \Gamma\,. \end{cases}$$

Note that if Γ is a lattice, the ratio of $\|\Gamma\|$ and the volume of the quotient space $\mathbb{R}^k/\Gamma$ is uniformly bounded between two positive constants.

The next lemma easily follows from the definitions:

Lemma 5.1. *Let Γ be a discrete subgroup of $\mathbb{R}^k$, $\mathbf{v} \in \mathbb{R}^k \smallsetminus \Gamma_\mathbb{R}$, and let $\Lambda \subset \mathbb{R}^k$ be a discrete subgroup containing both Γ and $\mathbf{v}$. Then $\|\Lambda\| \le \|\Gamma\| \cdot \|\mathbf{v}\|$, or, equivalently, $\|\mathbf{v}\| \ge \|\Lambda\|/\|\Gamma\|$.*

Proof. The claim is trivial if $\Gamma = \{0\}$. Otherwise, let $\mathbf{w}$ represent Γ; it suffices to show that $\|\mathbf{w} \wedge \mathbf{v}\| \le \|\mathbf{w}\| \cdot \|\mathbf{v}\|$. Write $\mathbf{w} = \sum_{I\subset\{1,\dots,k\}} w_I \mathbf{e}_I$ and $\mathbf{v} = \sum_{i=1}^k v_i \mathbf{e}_i$, then clearly

$$\begin{aligned} \Big\| \sum_{I\subset\{1,\dots,k\}} w_I \mathbf{e}_I \wedge \sum_{i=1}^k v_i \mathbf{e}_i \Big\| &\le \max_{\substack{1\le i\le k \\ I\subset\{1,\dots,k\}}} |w_I v_i| \\ &\le \max_{I\subset\{1,\dots,k\}} |w_I| \cdot \max_{1\le i\le k} |v_i| = \|\mathbf{w}\| \cdot \|\mathbf{v}\|\,. \quad \square \end{aligned}$$

Let Λ be a discrete subgroup of $\mathbb{R}^k$. We say that a subgroup Γ of Λ is *complete* (in Λ) if $\Gamma = \Gamma_\mathbb{R} \cap \Lambda$, and denote by $\mathcal{L}(\Lambda)$ the set of all nonzero complete subgroups of Λ. The inclusion relation makes $\mathcal{L}(\Lambda)$ a poset, its length being equal to the dimension of $\Lambda_\mathbb{R}$.

Theorem 5.2. *Let $d, k \in \mathbb{N}$, $C, \alpha > 0$, $0 < \rho \le 1$, and let a ball $B = B(\mathbf{x}_0, r_0) \subset \mathbb{R}^d$ and a map $h : \tilde{B} \to GL_k(\mathbb{R})$ be given, where $\tilde{B}$ stands for $B\big(\mathbf{x}_0, 3^k r_0\big)$. For any $\Gamma \in \mathcal{L}(\mathbb{Z}^k)$, denote by ψ_Γ the function $\psi_\Gamma(\mathbf{x}) \stackrel{\text{def}}{=} \|h(\mathbf{x})\Gamma\|$, $\mathbf{x} \in \tilde{B}$. Assume that for any $\Gamma \in \mathcal{L}(\mathbb{Z}^k)$,*

(i) *ψ_Γ is (C, α)-good on $\tilde{B}$;*
(ii) *$\|\psi_\Gamma\|_B \ge \rho$.*

Then for any positive $\varepsilon \le \rho$ one has

$$\big|\{\mathbf{x} \in B \mid \delta\big(h(\mathbf{x})\mathbb{Z}^k\big) < \varepsilon\}\big| \le kC\big(3^d N_d\big)^k \left(\frac{\varepsilon}{\rho}\right)^\alpha |B|\,. \tag{5.1}$$

Proof. We let $S = \mathcal{L}(\mathbb{Z}^k)$ and denote by φ the map $\Gamma \to \psi_\Gamma$. It is easy to verify that $(S, \varphi, B) \in \mathcal{A}(d, k, C, \alpha, \rho)$. Indeed, the property (A0) is clear, (A1) is given by (i), (A2) by (ii), and (A3) follows from the discreteness of $\bigwedge(\mathbb{Z}^k)$ in $\bigwedge(\mathbb{R}^k)$.

In view of Theorem 4.1, it remains to prove that a point $\mathbf{x} \in B$ with $\delta\big(h(\mathbf{x})\mathbb{Z}^k\big) < \varepsilon$ can not be $(\varepsilon, S, \varphi)$-marked. In other words,

$$\Phi(\varepsilon, S, \varphi, B) \subset \{\mathbf{x} \in B \mid \delta\big(h(\mathbf{x})\mathbb{Z}^k\big) \geq \varepsilon\}\,. \tag{5.2}$$

Take an $(\varepsilon, S, \varphi)$-marked point $\mathbf{x} \in B$, and let $\{0\} = \Gamma_0 \subsetneq \Gamma_1 \subsetneq \cdots \subsetneq \Gamma_l = \mathbb{Z}^k$ be all the elements of $\Sigma_{\mathbf{x}} \cup \big\{\{0\}, \mathbb{Z}^k\big\}$. Take any $\mathbf{v} \in \mathbb{Z}^k \smallsetminus \{0\}$. Then there exists i, $1 \leq i \leq l$, such that $\mathbf{v} \in \Gamma_i \smallsetminus \Gamma_{i-1}$. Denote $(\Gamma_{i-1} + \mathbb{R}\mathbf{v}) \cap \mathbb{Z}^k$ by Λ. Clearly Λ is a complete subgroup of $\mathbb{Z}^k$ contained in Γ_i, therefore $\Lambda \in \Sigma_{\mathbf{x}} \cup S(\Sigma_{\mathbf{x}})$. Now one can use properties (M1) and (M2) to deduce that

$$|\psi_\Lambda(\mathbf{x})| = \|h(\mathbf{x})\Lambda\| \geq \min(\varepsilon, \rho) = \varepsilon\,,$$

and then apply Lemma 5.1 to conclude that

$$\|h(\mathbf{x})\mathbf{v}\| \geq \frac{\|h(\mathbf{x})\Lambda\|}{\|h(\mathbf{x})\Gamma_{i-1}\|} \geq \varepsilon/\rho \geq \varepsilon\,.$$

This shows (5.2) and completes the proof of the theorem. $\square$

As was mentioned in the introduction, our method of proof is based, with some technical changes, on the argument from [Mar] and its modification in [D2]. As an illustration, let us show how one can use the above theorem to get a quantitative strengthening of Theorem 2.1 from [D2].

Theorem 5.3. *For any lattice Λ in $\mathbb{R}^k$ there exists a constant $\rho = \rho(\Lambda) > 0$ such that for any one-parameter unipotent subgroup $\{u_x\}_{x\in\mathbb{R}}$ of $SL_k(\mathbb{R})$, for any $T > 0$ and any $\varepsilon \leq \rho$, one has*

$$\big|\{0 < x < T \mid \delta\big(u_t\Lambda\big) < \varepsilon\}\big| \leq 2k^3 6^k (k^2+1)^{1/k^2} \left(\frac{\varepsilon}{\rho}\right)^{1/k^2} T\,. \tag{5.3}$$

Proof. Write Λ in the form $g\mathbb{Z}^k$ with $g \in GL_k(\mathbb{R})$, and denote by h the function $h(x) = u_x g$. For any $\Gamma \in \mathcal{L}(\mathbb{Z}^k)$ with basis $\mathbf{v}_1, \ldots, \mathbf{v}_j$, the coordinates of $h(x)(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_j)$ will be polynomials in x of degree not exceeding k^2, hence $\big(2k^2(k^2+1)^{1/k^2}, 1/k^2\big)$-good on $\mathbb{R}$ by Proposition 3.2. In view of Lemma 3.1(c) and the definition of the norm of a lattice, the functions $\psi_\Gamma(x) \stackrel{\text{def}}{=} \|h(x)\Gamma\|$ will also be $\big(2k^2(k^2+1)^{1/k^2}, 1/k^2\big)$-good on $\mathbb{R}$.

Now let $\rho \stackrel{\text{def}}{=} \inf_{\Gamma\in\mathcal{L}(\mathbb{Z}^k)} \|g\Gamma\|$, positive by the discreteness of Λ in $\mathbb{R}^k$. Then $\psi_\Gamma(0) \geq \rho$ for any $\Gamma \in \mathcal{L}(\mathbb{Z}^k)$, therefore $\|\psi_\Gamma\|_{(0,T)} \geq \rho$ by the continuity of h.

We see now that with the the substitutions $B = (0, T)$, $C = (2k^2(k^2+1)^{1/k^2}$, $\alpha = 1/k^2$ and $d = 1$ (note that it is an elementary fact that $N_1 = 2$), assumptions (i) and (ii) of Theorem 5.2 are satisfied, and one immediately gets (5.3) from (5.1). $\square$

We now derive another corollary from Theorem 5.2 which will immediately imply Proposition 2.3. In what follows, we put $k = n+1$ and for $\mathbf{y} \in \mathbb{R}^n$ define $u_{\mathbf{y}} \stackrel{\text{def}}{=} \begin{pmatrix} 1 & \mathbf{y}^T \\ 0 & I_n \end{pmatrix}$, so that the lattice $\Lambda_{\mathbf{y}}$ (see (1.4)) is given by $u_{\mathbf{y}}\mathbb{Z}^{n+1}$. We also let $\{\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be the standard basis of $\mathbb{R}^{n+1}$.

Theorem 5.4. *Let d, C, α, ρ and $B = B(\mathbf{x}_0, r_0)$ be as in Theorem 5.2. For $n \in \mathbb{N}$, let $\mathbf{f} = (f_1, \dots, f_n)$ be a continuous map from $\tilde B \stackrel{\text{def}}{=} B\big(\mathbf{x}_0, 3^{n+1}r\big)$ to $\mathbb{R}^n$ such that*

(i) *for any $\mathbf{c} = (c_0, c_1, \dots, c_n) \in \mathbb{R}^{n+1}$, $c_0 + \sum_{i=1}^n c_i f_i$ is (C,α)-good on $\tilde B$;*

(ii) *for any $\mathbf{c} \in \mathbb{R}^{n+1}$ with $\|\mathbf{c}\| \ge 1$, $\|c_0 + \sum_{i=1}^n c_i f_i\|_B \ge \rho$.*

Fix nonnegative numbers $t_1, \dots, t_n$, and for $\mathbf{x} \in \tilde B$ let

$$h(\mathbf{x}) \stackrel{\text{def}}{=} g_{\mathbf{t}} u_{\mathbf{f}(\mathbf{x})}\,, \tag{5.4}$$

where $g_{\mathbf{t}}$ is given by (1.5). Then for any positive $\varepsilon \le \rho$ one has

$$\big|\{\mathbf{x} \in B \mid \delta\big(h(\mathbf{x})\mathbb{Z}^{n+1}\big) < \varepsilon\}\big| \le (n+1)C\big(3^d N_d\big)^{n+1}\left(\frac{\varepsilon}{\rho}\right)^{\alpha}|B|\,. \tag{5.5}$$

Proof. In view of Theorem 5.2, it suffices to show that (i) and (ii) above will remain true with $c_0 + \sum_{i=1}^n c_i f_i$ replaced with $\psi_\Gamma \stackrel{\text{def}}{=} \|h(\cdot)\Gamma\|$, where Γ is any element of $\mathcal{L}(\mathbb{Z}^{n+1})$. Take $\mathbf{w} = \sum_{I \subset \{0,\dots,n\}} w_I \mathbf{e}_I$ representing Γ. To see how the coordinates of $\mathbf{w}$ change under the action of $h(\mathbf{x})$ of the form (5.4), first note that the action of $u_{\mathbf{f}(\mathbf{x})}$ leaves $\mathbf{e}_0$ invariant and sends $\mathbf{e}_i$ to $\mathbf{e}_i + f_i(\mathbf{x})\mathbf{e}_0$, $i = 1, \dots, n$. Therefore

$$u_{\mathbf{f}(\mathbf{x})}\mathbf{e}_I = \begin{cases} \mathbf{e}_I \text{ if } 0 \in I \\ \mathbf{e}_I + \sum_{i\in I} \pm f_i(\mathbf{x})\mathbf{e}_{I\cup\{0\}\smallsetminus\{i\}} \text{ otherwise}\,, \end{cases}$$

which shows that

$$u_{\mathbf{f}(\mathbf{x})}\mathbf{w} = \sum_{0\notin I} w_I\mathbf{e}_I + \sum_{0\in I}\left(w_I + \sum_{i\notin I} \pm w_{I\cup\{i\}\smallsetminus\{0\}} f_i(\mathbf{x})\right)\mathbf{e}_I\,.$$

Applying $g_{\mathbf{t}}$ to both sides of the above formula, one gets $h(\mathbf{x})\mathbf{w} = \sum_{I\subset\{0,\dots,n\}} h_I(\mathbf{x})\mathbf{e}_I$, where

$$h_I(\mathbf{x}) = \begin{cases} e^{-\sum_{i\in I} t_i} w_I \text{ if } 0 \notin I \\ e^{\sum_{i\notin I} t_i}\big(w_I + \sum_{i\in I} \pm w_{I\cup\{i\}\smallsetminus\{0\}} f_i(\mathbf{x})\big) \text{ otherwise}. \end{cases}$$

It follows that all the coordinates $h_I(\mathbf{x})$ of $h(\mathbf{x})\mathbf{w}$ are of the form $c_0 + \sum_{i=1}^n c_i f_i(\mathbf{x})$ for some $\mathbf{c} \in \mathbb{R}^{n+1}$. Thus, by virtue of assumption (i) of the Theorem, they are (C,α)-good on $\tilde B$. Then Lemma 3.1(c) applies and one can conclude that $\psi_\Gamma = \sup_{I\subset\{0,\dots,n\}} |h_I|$ is (C,α)-good on $\tilde B$. Furthermore, since the coordinates w_I of $\mathbf{w}$ are integers and at least one of them is nonzero, one can conclude that for some $I \ni 0$, the function $h_I(\mathbf{x})$ is of the form $c_0 + \sum_{i=1}^n c_i f_i$ with $\|\mathbf{c}\| \ge 1$. Hence $\|h_I\|_B \ge \rho$ for this I. Therefore $\|\psi_\Gamma\|_B \ge \rho$, which is all one needs to apply Theorem 5.2. $\square$

It is now easy to write down the

Proof of Proposition 2.3. Take $U \subset \mathbb{R}^d$, $\mathbf{f} : U \to \mathbb{R}^n$, $\mathbf{x}_0 \in U$ and l as in the statement of the proposition. Using Proposition 3.4, find a neighborhood $V \subset U$ of $\mathbf{x}_0$ and $C > 0$ such that any linear combination of $1, f_1, \dots, f_n$ is $(C, 1/dl)$-good on V. Choose a ball $\tilde B = B(\mathbf{x}_0, \tilde r)$ contained in V and let $B = B\big(\mathbf{x}_0, 3^{-(n+1)}\tilde r\big)$. Then condition (i) of Theorem 5.4 is satisfied with $\alpha = 1/dl$, while the existence of ρ satisfying (ii) follows from the linear independence of $1, f_1, \dots, f_n$ on B over $\mathbb{R}$. The validity of (2.6), with $D = (n+1)C\big(3^d N_d\big)^{n+1}$, is now an immediate consequence of (5.5). $\square$

6. Concluding remarks and generalizations

6.1. Observe that a notion of (C,α)-good functions can be defined for any metric space X with a Borel measure μ in place of $(\mathbb{R}^d, |\cdot|)$. Moreover, the argument of §4 works for arbitrary (X,μ) provided two additional conditions are satisfied:

- Besicovitch's Covering Theorem holds, the constant N_d being replaced by some positive number N_X;
- (cf. (4.5)) $\sup_{x\in X,\, r>0} \dfrac{\mu\big(B(x,3r)\big)}{\mu\big(B(x,r)\big)} < \infty$.

These conditions, in particular, are satisfied for $X = \mathbb{Q}_p$ and μ a Haar measure on $\mathbb{Q}_p$. This way one can attempt to apply the methods of the present paper to describe extremal and strongly extremal manifolds in p-adic spaces, generalizing the results from [Sp2, Chapter 2]. This work is currently in progress.

6.2. One may ask whether it is possible to develop a similar proof starting from the inequality (1.2M) instead of (1.1M). The answer is yes, although the proof turns out to be slightly more complicated. The recipe is simple: one should consider lattices $\begin{pmatrix} I_n & \mathbf{y} \\ 0 & 1 \end{pmatrix}\mathbb{Z}^{n+1}$ instead of (1.4) and act on them by elements of the form $\mathrm{diag}(e^{t_1},\dots,e^{t_n},e^{-t})$ instead of (1.5).

More generally, one can unify these two approaches by saying that a matrix $Y \in \mathrm{Mat}_{m,n}(\mathbb{R})$, interpreted as a system of m linear forms in n variables, is VWA (resp. VWMA) if for some $\varepsilon > 0$ there are infinitely many $\mathbf{q}\in\mathbb{Z}^n$ such that

$$\|Y\mathbf{q}+\mathbf{p}\|^m \cdot \|\mathbf{q}\|^n \le \|\mathbf{q}\|^{-n\varepsilon}, \tag{6.1}$$

or, respectively,

$$\Pi(Y\mathbf{q}+\mathbf{p})\cdot \Pi_+(\mathbf{q}) \le \Pi_+(\mathbf{q})^{-\varepsilon}, \tag{6.1M}$$

for some $\mathbf{p}\in\mathbb{Z}^m$. Then one faces a problem of describing (strongly) extremal submanifolds of $\mathrm{Mat}_{m,n}(\mathbb{R})$[2]. It turns out that one can apply Theorem 5.2 to get a general result of which Theorem A is a special case. This is going to be a topic of a forthcoming paper.

6.3. It is instructive to compare Theorem A with results recently obtained by M. Dodson, B. Rynne and J. Vickers. Following [DRV2], say that the manifold M satisfies condition $K1$ at $\mathbf{y}\in M$ if for any $\mathbf{v}\in T_{\mathbf{y}}M^{\perp}$, at least two of the principal curvatures of M at $\mathbf{y}$ with respect to $\mathbf{v}$ (see [DRV1] for further details) are nonzero and have the same sign. It is proved in [DRV1, DRV2] that a C^3 manifold M is extremal provided $K1$ holds for almost all $\mathbf{y}\in M$. Now say that M satisfies condition $K0$ at $\mathbf{y}\in M$ if for any $\mathbf{v}\in T_{\mathbf{y}}M^{\perp}$, at least one principal curvature of M at $\mathbf{y}$ with respect to $\mathbf{v}$ is nonzero. Then from Theorem A it immediately follows that a C^2 submanifold M of $\mathbb{R}^n$ is strongly extremal whenever condition $K0$ holds for almost every its point. Indeed, using the analytic formulation of curvature conditions as in [DRV1], it is easy to see that M satisfies condition $K0$ at $\mathbf{y}$ if and only if $\mathbf{y}$ is a 2-nondegenerate point of M.

[2] See [Ko1, Ko2] where similar objects are called "systems of (strongly) jointly extremal manifolds".

On the other hand, Dodson, Rynne and Vickers were able to use stronger curvature conditions to derive several Khintchine-type theorems and asymptotic formulae, see [DRV2, DRV3]. It would be interesting to know whether it is possible to obtain similar results assuming only the mild condition $K0$. In particular, it is proven in [DRV2] that if $K1$ holds for almost all $\mathbf{y} \in M$ and $\varphi : \mathbb{N} \to \mathbb{R}_+$ is a decreasing function such that

$$\sum_{q=1}^{\infty} \frac{\varphi(q)}{q} < \infty\,, \tag{6.2}$$

then for a.e. $\mathbf{y} \in M$ the inequality

$$|\mathbf{q} \cdot \mathbf{y} + p| \cdot \|\mathbf{q}\|^n \le \varphi\big(\|\mathbf{q}\|^n\big) \tag{6.3}$$

has at most finitely many solutions $\mathbf{q} \in \mathbb{Z}^n$, $p \in \mathbb{Z}$. On the other hand, one can modify Lemma 2.1 in the spirit of [K1, Theorem 8.3], and then apply Proposition 2.3 with $\mathbf{t} = (t/n, \dots, t/n)$ to obtain the following

Theorem B. *Let $f_1, \dots, f_n \in C^l(U)$, U an open subset of $\mathbb{R}^d$, be such that a.e. point of $M = \{\mathbf{f}(\mathbf{x}) \mid \mathbf{x} \in U\}$ is l-nondegenerate, and let $\varphi : \mathbb{N} \to \mathbb{R}_+$ be a decreasing function such that*

$$\sum_{q=1}^{\infty} \frac{\varphi(q)^{1/dl(n+1)}}{q} < \infty\,. \tag{6.4}$$

Then for a.e. $\mathbf{y} \in M$ the inequality (6.3) *has at most finitely many solutions.*

The details will appear elsewhere; whether it is possible to replace (6.4) by (6.2) in the above theorem is an open question. We remark that for M of the form (1.3) and φ satisfying (6.2), the conclusion of Theorem B was conjectured by A. Baker in 1966 and proved by V. Bernik [Be] in 1984.

Acknowledgements

A substantial part of this work was done during the authors' stay at the University of Bielefeld in June–July 1996. This stay was supported by SFB-343 and Humboldt Foundation. Thanks are also due to D. Dickinson, M. Dodson, T. Khovanova and N. Shah for useful comments.

References

[B] A. Baker, *Transcendental number theory*, Cambridge Univ. Press, Cambridge, 1975.

[Be] V. Bernik, *A proof of Baker's conjecture in the metric theory of transcendental numbers*, Doklady Akad. Nauk SSSR **277** (1984), 1036–1039. (Russian)

[BB] V. Beresnevich and V. Bernik, *On a metrical theorem of W. Schmidt*, Acta Arith. **75** (1996), 219–233.

[C] J. W. S. Cassels, *An introduction to Diophantine approximation*, Cambridge Tracts in Math., vol. 45, Cambridge Univ. Press, Cambridge, 1957.

[D1] S. G. Dani, *Divergent trajectories of flows on homogeneous spaces and Diophantine approximation*, J. Reine Angew. Math. **359** (1985), 55–89.

[D2] ———, *On orbits of unipotent flows on homogeneous spaces*, II, Ergodic Theory Dynamical Systems **6** (1986), 167–182.

[DM] S. G. Dani and G. A. Margulis, *Limit distributions of orbits of unipotent flows and values of quadratic forms*, Adv. in Soviet Math., vol. 16, Amer. Math. Soc., Providence, RI, 1993, pp. 91–137.

[DRV1] M. M. Dodson, B. P. Rynne and J. A. G. Vickers, *Metric Diophantine approximation and Hausdorff dimension on manifolds*, Math. Proc. Cambridge Philos. Soc. **105** (1989), 547–558.

[DRV2] ———, *Khintchine-type theorems on manifolds*, Acta Arith. **57** (1991), 115–130.

[DRV3] ———, *Simultaneous Diophantine approximation and asymptotic formulae on manifolds*, J. Number Theory **58** (1996), 298–316.

[EMS] A. Eskin, S. Mozes and N. Shah, *Non-divergence of translates of certain algebraic measures*, Geom. Funct. Anal. (to appear).

[K1] D. Kleinbock, *Flows on homogeneous spaces and Diophantine properties of matrices*, Duke Math. J. (to appear).

[K2] ———, *Nondense orbits of nonquasiunipotent flows and applications to Diophantine approximation*, Ph.D. Thesis, Yale University, 1996.

[Ko1] E. Kovalevskaya, *Simultaneously extremal manifolds*, Dokl. Akad. Nauk BSSR **31** (1987), 405–408. (Russian)

[Ko2] ———, *Strongly jointly extremal manifolds*, Vestsī Akad. Navuk BSSR Ser. Fīz.-Mat. Navuk (1987), no. 6, 16–19. (Russian)

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

[Mar] G. A. Margulis, *On the action of unipotent group in the space of lattices*, Proceedings of the Summer School on group representations, (Budapest 1971), Académiai Kiado, Budapest, 1975, pp. 365–370.

[Mat] P. Mattila, *Geometry of sets and measures in Euclidean spaces. Fractals and rectifiability*, Cambridge Studies in Advanced Mathematics, 44, Cambridge University Press, Cambridge, 1995.

[R] M. S. Raghunathan, *Discrete subgroups of Lie groups*, Springer-Verlag, Berlin and New York, 1972.

[S1] W. Schmidt, *Metrische Sätze über simultane Approximation abhänginger Grössen*, Monatsch. Math. **68** (1964), 154–166.

[S2] ———, *Diophantine approximation*, Springer-Verlag, Berlin and New York, 1980.

[Sh] N. Shah, *Limit distribution of polynomial trajectories on homogeneous spaces*, Duke Math. J. **75** (1994), 711–732.

[Sp1] V. Sprindžuk, *More on Mahler's conjecture*, Doklady Akad. Nauk SSSR **155** (1964), 54–56 (Russian); English transl. in Soviet Math. Dokl **5** (1964), 361–363.

[Sp2] ———, *Mahler's problem in metric number theory*, Translations of Mathematical Monographs, vol. 25, Amer. Math. Soc., Providence, RI, 1969.

[Sp3] ———, *Metric theory of Diophantine approximations*, John Wiley & Sons, New York-Toronto-London, 1979.

[Sp4] ———, *Achievements and problems in Diophantine approximation theory*, Russian Math. Surveys **35** (1980), 1–80.

[VC] A. I. Vinogradov and G. V. Chudnovsky, *The proof of extremality of certain manifolds*, Contributions to the theory of transcendental numbers, Math. Surveys Monographs, 19, Amer. Math. Soc., Providence, R.I., 1984, pp. 421–447.

Dmitry Y. Kleinbock, School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540; *e-mail address:* dklein@math.ias.edu

G. A. Margulis, Department of Mathematics, Yale University, New Haven, CT 06520; *e-mail address:* margulis@math.yale.edu