Khinchine-Type Theorems on Manifolds:
The Convergence Case for Standard and Multiplicative Versions

V. Bernik, D. Kleinbock, and G. A. Margulis

1 Introduction

The goal of this paper is to prove the convergence part of the Khinchine-Groshev theorem, as well as its multiplicative version, for nondegenerate smooth submanifolds in \( \mathbb{R}^n \). The proof combines methods from metric number theory with a new approach involving the geometry of lattices in Euclidean spaces.

Notation. The main objects of this paper are \( n \)-tuples \( y = (y_1, \ldots, y_n) \) of real numbers viewed as linear forms, that is, as row vectors. In what follows, \( y \) always means a row vector, and we are interested in values of a linear form given by \( y \) at integer points \( q = (q_1, \ldots, q_n)^T \), the latter being a column vector. Thus \( yq \) stands for \( y_1 q_1 + \cdots + y_n q_n \). Hopefully, it will cause no confusion.

We study differentiable maps \( f = (f_1, \ldots, f_n) \) from open subsets \( U \) of \( \mathbb{R}^d \) to \( \mathbb{R}^n \); again, \( f \) is interpreted as a row vector, so that \( f(x)q \) stands for \( q_1 f_1(x) + \cdots + q_n f_n(x) \). In contrast, the elements of the parameter set \( U \) are denoted by \( x = (x_1, \ldots, x_d) \) without boldfacing, since the linear structure of the parameter space is not significant.

For \( f \) as above we denote by \( \partial_i f : U \to \mathbb{R}^n, \ i = 1, \ldots, d, \) its partial derivative (also a row vector) with respect to \( x_i \). If \( F \) is a scalar function on \( U \), we denote by \( \nabla F \) the column vector consisting of partial derivatives of \( F \). With some abuse of notation, we treat vector functions \( f \) the same way; namely, \( \nabla f \) stands for the matrix function \( U \mapsto M_{d \times n}(\mathbb{R}) \) with

Received 3 August 2000.
rows given by partial derivatives $\partial_i f$. We also need higher-order differentiation: for a multi-index $\beta = (i_1, \ldots, i_d)$, $i_j \in \mathbb{Z}_+$, we let $|\beta| = \sum_{j=1}^d i_j$ and $\partial^\beta = \partial_1^{i_1} \circ \cdots \circ \partial_d^{i_d}$.

Unless otherwise indicated, the norm $\|x\|$ of a vector $x \in \mathbb{R}^k$ (either row or column vector) stands for $\|x\| = \max_{1 \leq i \leq k} |x_i|$. In some cases, however, we work with the Euclidean norm $\|x\| = \|x\|_e = \sqrt{\sum_{i=1}^k x_i^2}$, keeping the same notation. This distinction is clearly emphasized to avoid confusion. We denote by $\mathbb{R}_+^k$ the set of unit vectors in $\mathbb{R}^k$ (with respect to the Euclidean norm).

We use the notation $|\langle x \rangle|$ for the distance between $x \in \mathbb{R}$ and the closest integer, $|\langle x \rangle| \overset{def}{=} \min_{k \in \mathbb{Z}} |x - k|$. (It is quite customary to use $\|x\|$ instead, but we do not do this in order to save the latter notation for norms in vector spaces.) If $B \subset \mathbb{R}^k$, we let $|B|$ stand for the Lebesgue measure of $B$.

### Basics on Diophantine approximation

In what follows, we let $\Psi$ be a positive function defined on $\mathbb{Z}^n \setminus \{0\}$, and we consider the set

$$W(\Psi) \overset{def}{=} \{ y \in \mathbb{R}^n \mid |\langle yq \rangle| \leq \Psi(q) \text{ for infinitely many } q \}.$$ 

Clearly, the faster $\Psi$ decays at infinity, the smaller the set $W(\Psi)$ is. In particular, the Borel-Cantelli lemma gives a sufficient condition for this set to have measure zero: $|W(\Psi)| = 0$ if

$$\sum_{q \in \mathbb{Z}^n \setminus \{0\}} \Psi(q) < \infty. \quad (1.1)$$

It is customary to refer to statement (1.1), as well as to its various analogues, as the convergence case of a Khintchine-type theorem, since the fact that (under some regularity restrictions on $\Psi$) condition (1.1) is also necessary was first proved by A. Khintchine for $n = 1$ and was later generalized by A. Groshev and W. Schmidt. See Section 8.5 for more detail.

The standard class of examples is given by functions that depend only on the norm of $q$. If $\psi$ is a positive function defined on positive integers, let us say, following [KM2], that $y$ is $\psi$-approximable, abbreviated as $\psi$-$A$, if it belongs to $W(\Psi)$, where

$$\Psi(q) = \psi(\|q\|^n). \quad (1.2s)$$

If $\psi$ is nonincreasing (which is our standing assumption), (1.1) is satisfied if and only if

---

1We are grateful to M. M. Dodson for permission to modify his terminology used in [Do] and [BD]. In our opinion, parametrization (1.2s) instead of the traditional $\psi(\|q\|)$ makes the structure more transparent and less dimension-dependent (see [KM1], [KM2] for justification).
\[ \sum_{k=1}^{\infty} \psi(k) < \infty. \] \hspace{1cm} (1.1s)

For example, almost all \( y \in \mathbb{R}^n \) are not very well approximable (VWA) (see [S2, Chapter IV, Section 5]). The latter term is defined as \( \psi_\epsilon \)-approximable for some positive \( \epsilon \) with \( \psi_\epsilon(k) \overset{\text{def}}{=} k^{-(1+\epsilon)} \).

Another important special case is given by

\[ \Psi(q) = \psi(\Pi_+(q)), \] \hspace{1cm} (1.2m)

where \( \Pi_+(q) \) is defined as \( \prod_{i=1}^{n} \max(|q_i|, 1) \), that is, the absolute value of the product of all the nonzero coordinates of \( q \). We say that \( y \in \mathbb{R}^n \) is \( \psi \)-\textit{multiplicatively approximable} (\( \psi \)-MA) if it belongs to \( \mathcal{W}(\Psi) \) with \( \Psi \) as in (1.2m). In this case, again assuming the monotonicity of \( \psi \), (1.1) is satisfied if and only if

\[ \sum_{k=1}^{\infty} (\log k)^{n-1} \psi(k) < \infty. \] \hspace{1cm} (1.1m)

Also, since \( \Pi_+(q) \) is not greater than \( \|q\|^n \), any \( \psi \)-approximable \( y \) is automatically \( \psi \)-MA. For example, one can define \textit{very well multiplicatively approximable} (VWMA) points as \( \psi_\epsilon \)-\textit{multiplicatively approximable} for some positive \( \epsilon \) with \( \psi_\epsilon \) as above. It follows that almost all \( y \in \mathbb{R}^n \) are not VWMA.

\textbf{Diophantine approximation on manifolds}

Far more intricate questions arise if one restricts \( y \) to lie on a submanifold \( M \) of \( \mathbb{R}^n \). In 1932 K. Mahler [M] conjectured that almost all points on the curve

\[ \{(x, x^2, \ldots, x^n) \mid x \in \mathbb{R}\} \] \hspace{1cm} (1.3)

are not VWA. V. Sprindžuk’s proof of this conjecture [Sp1], [Sp2] has eventually led to the development of a new branch of metric number theory, usually referred to as “Diophantine approximation with dependent quantities” or “Diophantine approximation on manifolds.” In particular, Sprindžuk’s result was improved by A. Baker [B1] in 1966. He showed that if \( \psi \) is a positive nonincreasing function such that

\[ \sum_{k=1}^{\infty} \frac{\psi(k)^{1/n}}{k^{1-1/n}} < \infty, \] \hspace{1cm} (1.4)

then almost all points on the curve (1.3) are not \( \psi \)-approximable. Baker conjectured that (1.4) could be replaced by the optimal condition (1.1s); this conjecture was proved
later by V. Bernik [Bern]. As for the multiplicative approximation, it was conjectured by Baker in [B2] that almost all points on the curve (1.3) are not VWMA; the validity of this conjecture for \( n \leq 4 \) was verified in 1997 by Bernik and V. Borbat [BB].

Since the mid-1960s, many efforts have been directed toward obtaining similar results for larger classes of smooth submanifolds of \( \mathbb{R}^n \). A new method, based on combinatorics of the space of lattices, was developed in 1998 by D. Kleinbock and G. Margulis in [KM1]. Let us employ the following definition: If \( U \) is an open subset of \( \mathbb{R}^d \) and \( l \leq m \in \mathbb{N} \), we say that an \( n \)-tuple \( f = (f_1, \ldots, f_n) \) of \( C^m \) functions \( U \to \mathbb{R} \) is \( l \)-nondegenerate at \( x \in U \) if the space \( \mathbb{R}^n \) is spanned by partial derivatives of \( f \) at \( x \) of order up to \( l \). We say that \( f \) is nondegenerate at \( x \) if it is \( l \)-nondegenerate for some \( l \). If \( M \subset \mathbb{R}^n \) is a \( d \)-dimensional \( C^m \) submanifold, we say that \( M \) is nondegenerate at \( y \in M \) if any (equivalently, some) diffeomorphism \( f \) between an open subset \( U \) of \( \mathbb{R}^d \) and a neighborhood of \( y \) in \( M \) is nondegenerate at \( f^{-1}(y) \). We say that \( f : U \to \mathbb{R}^n \) (resp., \( M \subset \mathbb{R}^n \)) is nondegenerate if it is nondegenerate at almost every point of \( U \) (resp., \( M \), in the sense of the natural measure class on \( M \)).

**Theorem A** [KM1]. Let \( M \) be a nondegenerate \( C^m \) submanifold of \( \mathbb{R}^n \). Then almost all points of \( M \) are not VWMA (hence not VWA as well). \( \square \)

In particular, the aforementioned multiplicative conjecture of Baker follows from this theorem. Note also that if the functions \( f_1, \ldots, f_n \) are analytic and \( U \) is connected, the nondegeneracy of \( f \) is equivalent to the linear independence of \( 1, f_1, \ldots, f_n \) over \( \mathbb{R} \); in this setting the conclusion of Theorem A was conjectured by Sprindžuk [Sp4, Conjectures H1, H2].

**Main results and structure of the paper**

The primary goal of this paper is to obtain a Khintchine-type generalization of Theorem A. More precisely, we prove the following theorem.

**Theorem 1.1.** Let \( U \subset \mathbb{R}^d \) be an open set, and let \( f : U \to \mathbb{R}^n \) be a nondegenerate \( n \)-tuple of \( C^m \) functions on \( U \). Also, let \( \Psi : \mathbb{Z}^n \setminus \{0\} \to (0, \infty) \) be a function satisfying (1.1) and such that for \( i = 1, \ldots, n \) one has

\[
\Psi(q_1, \ldots, q_i, \ldots, q_n) \geq \Psi(q_1, \ldots, q'_i, \ldots, q_n) \quad \text{whenever} \quad |q_i| \leq |q'_i| \quad \text{and} \quad q_i q'_i > 0.
\]

(That is, \( \Psi \) is nonincreasing with respect to the absolute value of any coordinate in any orthant of \( \mathbb{R}^n \).) Then \( \|x \in U \mid f(x) \in \mathcal{W}(\Psi)\| = 0. \) \( \square \)
In particular, if $\Psi$ is of the form (1.2s) or (1.2m) for a nonincreasing function $\psi : \mathbb{N} \mapsto \mathbb{R}_+$, condition (1.5) is clearly satisfied. Thus one has the following corollary.

**Corollary 1.2.** Let $f : U \rightarrow \mathbb{R}^n$ be as in Theorem 1.1, and let $\psi : \mathbb{N} \mapsto (0, \infty)$ be a nonincreasing function. Then

- (S) assuming (1.1s), for almost all $x \in U$, the points $f(x)$ are not $\psi$-A;
- (M) assuming (1.1m), for almost all $x \in U$, the points $f(x)$ are not $\psi$-MA. \qed

It is worth mentioning that Corollary 1.2(S) was recently proved by V. Beresnevich in [Be5] using a refinement of Sprindžuk’s method of “essential and inessential domains.” Earlier, several special cases were treated in [DRV1], [BDD], and [Be2]. A preliminary version (see [BKM]) of the present paper, where the two statements of Corollary 1.2 were proved for the case $d = 1$, appeared in 1999 as a preprint of the University of Bielefeld.

Our proof of Theorem 1.1 is based on carefully measuring sets of solutions of certain systems of Diophantine inequalities. Specifically, we fix a ball $B \subset \mathbb{R}^d$, and we look at the set of all $x \in B$ for which there exists an integer vector $q$ in a certain range such that the value of the function $F(x) = f(x)q$ is close to an integer. Our estimates require considering two special cases: when the norm of the gradient $\nabla F(x) = \nabla f(x)q$ is large and when it is not very large. We show in Section 8.1 that, by means of straightforward measure computations, Theorem 1.1 reduces to the following two theorems.

**Theorem 1.3.** Let $B \subset \mathbb{R}^d$ be a ball of radius $r$, let $\bar{B}$ stand for the ball with the same center as $B$ and of radius $2r$, and let functions $f = (f_1, \ldots, f_n) \in C^2(\bar{B})$ be given. Fix $\delta > 0$, and define

$$L = \max_{\beta : |\beta| = 2, x \in B} \|\partial^\beta f(x)\|. \tag{1.6a}$$

Then, for every $q \in \mathbb{Z}^n$ such that

$$\|q\| \geq \frac{1}{4nLr^2}, \tag{1.6b}$$

the set of solutions $x \in B$ of the inequalities

$$|\langle f(x)q \rangle| < \delta \tag{1.6c}$$

and

$$\|\nabla f(x)q\| \geq \sqrt{n}dL\|q\| \tag{1.6d}$$

has measure at most $C_d \delta |B|$, where $C_d$ is a constant dependent only on $d$. \qed

**Theorem 1.4.** Let $U \subset \mathbb{R}^d$ be an open set, $x_0 \in U$, and let $f = (f_1, \ldots, f_n)$ be an $n$-tuple of
smooth functions on $U$ which is $l$-nondegenerate at $x_0$. Then there exists a neighborhood $V \subset U$ of $x_0$ with the following property: for any ball $B \subset V$ there exists $E > 0$ such that, for any choice of

$$0 < \delta \leq 1, \quad T_1, \ldots, T_n \geq 1, \quad \text{and} \quad K > 0 \quad \text{with} \quad \frac{\delta K T_1 \cdot \ldots \cdot T_n}{\max_i T_i} \leq 1,$$

the set

$$\left\{ x \in B \mid \exists \mathbf{q} \in \mathbb{Z}^n \setminus \{0\} \text{ such that } \begin{cases} |\langle f(x) \mathbf{q} \rangle| < \delta, \\ \|\nabla f(x)\mathbf{q}\| < K, \\ |q_i| < T_i, \quad i = 1, \ldots, n \end{cases} \right\}$$

has measure at most $E \varepsilon^{1/(d(2l-1))} |B|$, where one defines

$$\varepsilon \overset{\text{def}}{=} \max \left( \delta, \left( \frac{\delta K T_1 \cdot \ldots \cdot T_n}{\max_i T_i} \right)^{1/(n+1)} \right). \tag{1.7c}$$

Theorem 1.3, roughly speaking, says that a function with a large gradient and not very large second-order partial derivatives cannot have values very close to integers on a set of large measure. It is proved in Section 2 using an argument that is apparently originally due to Bernik and is (in the case $d = 1$) implicitly contained in one of the steps of [BDD].

As for Theorem 1.4, it is proved by a modification of a method from [KM1] involving the geometry of lattices in Euclidean spaces. The connection with lattices is discussed in Section 5, where Theorem 1.4 is translated into the language of lattices (see Theorem 5.1). To prove the latter, we rely on the notion of $(C, \alpha)$-good functions introduced in [KM1]. This concept is reviewed in Section 3, as well as in Section 4, where we prove that certain functions arising in the proof of Theorem 5.1 are $(C, \alpha)$-good for suitable $C, \alpha$. We prove Theorem 5.1 in Section 7 after doing some preparatory work in the preceding section. Section 8 is devoted to reducing Theorem 1.1 to Theorems 1.3 and 1.4, as well as to several concluding remarks, applications, and some open questions. In particular, there we discuss the complementary divergence case of Theorem 1.1 and also present applications involving the approximation of zero by values of functions and their derivatives.

## 2 Proof of Theorem 1.3

We first state a covering result that is well known. (It is one of the ingredients of the main estimate from [KM1].)
Theorem 2.1 (Besicovitch's covering theorem [Mat, Theorem 2.7]). There is an integer $N_d$ depending only on $d$ with the following property: let $S$ 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 $x \in S$ is the center of some ball of $\mathcal{B}$; then there exists a finite or countable subfamily $\{U_i\}$ of $\mathcal{B}$ with $1_S \leq \sum_i 1_{U_i} \leq N_d$. (That is, $S \subset \bigcup_i U_i$ and the multiplicity of that subcovering is at most $N_d$.)

This theorem and the constant $N_d$ are used repeatedly in this paper.

Lemma 2.2. Let $B \subset \mathbb{R}^d$ be a ball of radius $r$, and let the numbers

$$M \geq \frac{1}{4r^2} \quad (2.1a)$$

and $\delta > 0$ be given. Denote by $\tilde{B}$ the ball with the same center as $B$ and of radius $2r$. Take a function $F \in C^2(\tilde{B})$ such that

$$\sup_{|\beta|=2, x \in \tilde{B}} |\partial_\beta F(x)| \leq M, \quad (2.1b)$$

and denote by $S$ the set of all $x \in B$ for which the inequalities

$$|\langle F(x)\rangle| < \delta \quad (2.1c)$$

and

$$\|\nabla F(x)\| \geq \sqrt{dM} \quad (2.1d)$$

hold. Then $|S| \leq C_d \delta |B|$, where $C_d$ is a constant dependent only on $d$. □

Proof. Clearly $|S| \leq 16\delta|B|$ when $\delta \geq 1/16$, so without loss of generality we can assume that $\delta$ is less than $1/16$. Also, given $x \in S$, without loss of generality we can assume that the maximal value of $|\partial_j F(x)|$, $j = 1, \ldots, d$, occurs when $j = 1$. Denote $1/(2|\partial_1 F(x)|)$ by $\rho$, and note that $\rho \sqrt{d} \leq 1/(2\sqrt{M}) \leq r$ due to (2.1a) and (2.1d); therefore the ball $B(x, \rho \sqrt{d})$ is contained in $\tilde{B}$. Also, let us denote by $U(x)$ the maximal ball centered in $x$ such that $|\langle F(y)\rangle| < 1/4$ for all $y \in U(x)$. It is clear that there exists a unique $p \in \mathbb{Z}$ such that $|F(y) + p| < 1/4$ for all $y \in U(x)$. We claim that the radius of $U(x)$ is not larger than $\rho$. Indeed, one has

$$F(x_1 \pm \rho, x_2, \ldots, x_d) + p = F(x) + p \pm \partial_1 F(x)\rho + \frac{\partial^2_1 F(z)}{2} \rho^2$$

for some $z$ between $x$ and $(x_1 \pm \rho, x_2, \ldots, x_d)$. Thus

$$|F(x_1 \pm \rho, x_2, \ldots, x_d) + p| \geq \delta + \frac{1}{2} - \frac{M\rho^2}{2}$$

(2.1c)
and the claim is proved. In particular, \( U(x) \subset \tilde{B} \); moreover, if one denotes by \( \bar{U}(x) \) the cube circumscribed around \( U(x) \) with sides parallel to the coordinate axes, then \( \bar{U}(x) \subset \tilde{B} \) as well.

On the other hand, the radius of \( U(x) \) cannot be too small. If \( y \in B(x, \rho/(4\sqrt{d})) \), one has

\[
|F(y) + p| \leq |F(x) + p| + \left| \frac{\partial F}{\partial u}(x) \right| \frac{\rho}{4\sqrt{d}} + \frac{1}{2} \frac{\partial^2 F}{\partial u^2}(z) \frac{\rho^2}{16d},
\]

where \( u \) is the unit vector parallel to \( y - x \), and \( z \) is between \( x \) and \( y \). Note that it follows from our ordering of coordinates that \( |(\partial^2 F/\partial u)(x)| \leq \sqrt{d} |\partial_1 F(x)| = \sqrt{d}/(2\rho) \), and it follows from (2.1b) that \( |(\partial^2 F/\partial u^2)(z)| \leq dM \leq d/4\rho^2 \). Therefore \( |F(y) + p| \leq \delta + 1/8 + 1/128 < 1/4 \), which shows that \( U(x) \subset B(x, \rho/(4\sqrt{d})) \) and that, in particular,

\[
|U(x)| \geq C_d \rho^d. 
\]  
(2.2)

(The values of constants \( C_d' \), and also those of \( C_d'', C_d''' \), which are introduced later, depend only on \( d \).)

Also, one can observe that \( \partial_1 F(y) \) does not oscillate too much when \( y \in \bar{U}(x) \): for some \( z \) between \( x \) and \( y \), one gets

\[
\left| \partial_1 F(y) - \partial_1 F(x) \right| \leq \left| \frac{\partial}{\partial u} \partial_1 F(z) \right| \rho \sqrt{d} \leq \frac{M \rho \sqrt{d}}{(2.1b)} \leq \frac{1}{4\rho} \frac{\rho^2}{2} = \frac{\partial_1 F(x)}{2}.
\]

(Here again \( u \) is the unit vector parallel to \( y - x \).) This implies that the absolute value of \( \partial_1 F(y), y \in \bar{U}(x) \), is not less than \( 1/2|\partial_1 F(x)| \); in particular, for every \( y_2, \ldots, y_d \) such that \( |y_i - x_i| < \rho, i = 2, \ldots, d \), the function \( F(\cdot, y_2, \ldots, y_d) \) is monotonic on \((x_1 - \rho, x_1 + \rho)\), and therefore

\[
\left| \left\{ y_1 \in (x_1 - \rho, x_1 + \rho) \mid |F(y_1, \ldots, y_d) + p| < \delta \right\} \right| \leq 2\delta \frac{2}{|\partial_1 F(x)|} = 8\rho\delta.
\]

Now we can estimate \( \|x \in U(x) \mid |F(x) + p| < \delta\| \) from above by

\[
\left| \left\{ x \in \bar{U}(x) \mid |F(x) + p| < \delta \right\} \right| \leq 8\rho\delta 2^{d-1} \rho^{d-1} = C_d'' \rho^d \leq \frac{C_d''' |U(x)|}{(2.2)}.
\]

The set \( S \) is covered by all the balls \( U(x), x \in S \), and, using Theorem 2.1, one can choose a subcovering \( \{U_i\} \) of multiplicity at most \( N_d \). Then one has
\[ |S| \leq \sum_i C_i'' \delta |U_i| \leq C'' |\tilde{B}| = C \delta |B|, \]

which finishes the proof. \(\square\)

Now we need very little to complete the following proof.

Proof of Theorem 1.3. Given the balls \( B \subset \tilde{B} \subset \mathbb{R}^d \), an \( n \)-tuple of \( C^2 \) functions \( f \) on \( \tilde{B} \), a positive \( \delta \), and \( q \in \mathbb{Z}^n \) satisfying (1.6b) with \( L \) as in (1.6a), denote \( F(x) \overset{\text{def}}{=} f(x)q \) and \( M \overset{\text{def}}{=} nL\|q\|. \) Then inequalities (1.6a), (1.6b), (1.6c), and (1.6d) can be rewritten as (2.1b), (2.1a), (2.1c), and (2.1d), and the theorem follows. \(\square\)

3 \( (C, \alpha) \)-Good functions

Let us recall the definition introduced in [KM1]. If \( C \) and \( \alpha \) are positive numbers and \( V \) is a subset of \( \mathbb{R}^d \), let us say that a function \( f: V \mapsto \mathbb{R} \) is \((C, \alpha)\)-good on \( V \) if for any open ball \( B \subset V \) and any \( \epsilon > 0 \) one has

\[ \left| \left\{ x \in B \mid |f(x)| < \epsilon \cdot \sup_{x \in B} |f(x)| \right\} \right| \leq C \epsilon^\alpha |B|. \]  (3.1)

Several elementary facts about \((C, \alpha)\)-good functions are as follows.

**Lemma 3.1.** We have that
(a) \( f \) is \((C, \alpha)\)-good on \( V \) \( \Rightarrow \) so is \( \lambda f \forall \lambda \in \mathbb{R} \);
(b) \( f_i, \ i \in I, \) are \((C, \alpha)\)-good on \( V \) \( \Rightarrow \) so is \( \sup_{i \in I} |f_i| \);
(c) if \( f \) is \((C, \alpha)\)-good on \( V \) and \( c_1 \leq |f(x)|/|g(x)| \leq c_2 \) for all \( x \in V \), then \( g \) is \((C(c_2/c_1)^\alpha, \alpha)\)-good on \( V \);
(d) \( f \) is \((C, \alpha)\)-good on \( V \) \( \Rightarrow \) it is \((C', \alpha')\)-good on \( V' \) for every \( C' \geq C, \alpha' \leq \alpha, \) and \( V' \subset V. \) \(\square\)

Note that it follows from Lemma 3.1(b) that the (supremum) norm of a vector function \( f \) is \((C, \alpha)\)-good whenever every component of \( f \) is \((C, \alpha)\)-good. Also, Lemma 3.1(c) shows that one is allowed to replace the norm by an equivalent one, affecting \( C \) only, not \( \alpha. \)

**Lemma 3.2.** Any polynomial \( f \in \mathbb{R}[x_1, \ldots, x_d] \) of degree not greater than \( l \) is \((C_{d,l}, 1/dl)\)-good on \( \mathbb{R}^d \), where \( C_{d,l} = (2^{d+1} dl(l+1)^{1/l})/v_d. \) (Here and in Lemma 3.3 \( v_d \) stands for the volume of the unit ball in \( \mathbb{R}^d. \)) \(\square\)

Proof. The case \( d = 1 \) is proved in [KM1, Proposition 3.2]. By induction on \( d \), as in the proof of [KM1, Lemma 3.3], one can show that for any \( d \)-dimensional cube \( B \) and for any
\( \varepsilon > 0 \) one has
\[
\left| \left\{ x \in B \mid |f(x)| < \varepsilon \cdot \sup_{x \in B} |f(x)| \right\} \right| \leq 2 d l (l + 1)^{1/l} \varepsilon^{1/d} |B|,
\]
and the claim follows by circumscribing a cube around any ball in \( \mathbb{R}^d \).

The following lemma is a direct consequence of [KM1, Lemma 3.3].

**Lemma 3.3.** Let \( U \) 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
\[
|\partial_\beta f(x)| \leq A_1, \quad \forall \beta \text{ with } |\beta| \leq k,
\]
and
\[
|\partial_i^k f(x)| \geq A_2, \quad \forall i = 1, \ldots, d,
\]
for all \( x \in U \). Also, let \( V \) be a subset of \( U \) such that whenever a ball \( B \) lies in \( V \) any cube circumscribed around \( B \) is contained in \( U \). Then \( f \) is \((C,1/dl)\)-good on \( V \), where
\[
C = \frac{2^d}{v_d} d k (k + 1) \left( \frac{A_1}{A_2} (k + 1) (2k^k + 1) \right)^{1/k}.
\]

The following proposition describes the \((C, \alpha)\)-good property of functions chosen from certain compact families defined by the nonvanishing of partial derivatives. The argument used is similar to that of [KM1, Proposition 3.4].

**Proposition 3.4.** Let \( U \) be an open subset of \( \mathbb{R}^d \), and let \( F \subset C^l(U) \) be a family of functions \( f : U \to \mathbb{R} \) such that
\[
\{ \nabla f \mid f \in F \} \text{ is compact in } C^{1-1}(U).
\]
Also, assume that
\[
\inf_{f \in F} \sup_{|\beta| \leq l} |\partial_\beta f(x_0)| > 0.
\]
(In other words, the derivatives of \( f \) at \( x_0 \) uniformly (in \( f \in F \)) generate \( \mathbb{R} \).) Then there exist a neighborhood \( V \subset U \) of \( x_0 \) and a positive \( C = C(F) \) such that the following hold for all \( f \in F \):

(a) \( f \) is \((C,1/dl)\)-good on \( V \);

(b) \( \|\nabla f\| \) is \((C,1/(d(l-1)))\)-good on \( V \).

**Proof.** Assumption (3.4) says that there exists a constant \( C_1 > 0 \) such that for any \( f \in F \)
one can find a multi-index $\beta$ with $|\beta| = k \leq l$ and with

$$|\partial_\beta f(x_0)| \geq C_1.$$  

By an appropriate rotation of the coordinate system, one can guarantee that $|\partial^k_i f(x_0)| \geq C_2$ for all $i = 1, \ldots, d$ and some positive $C_2$ independent of $f$. Then one uses the continuity of the derivatives of $f$ and compactness of $\mathcal{F}$ to choose a neighborhood $V' \subset U$ of $x_0$ and positive $A_1, A_2$ (again independently of $f \in \mathcal{F}$) such that the inequalities in (3.2) hold for all $x \in V'$. (Note that in these inequalities both $k \in \{1, \ldots, l\}$ and the coordinate system depend on $f$.) Finally, we let $V$ be a smaller neighborhood of $x_0$ such that whenever a ball $B$ lies in $V$ any cube $\tilde{B}$ circumscribed around $B$ is contained in $V'$.

Now Proposition 3.4(a) immediately follows from Lemma 3.3. As for Proposition 3.4(b), let $\mathcal{F}'$ be the family of functions in $\mathcal{F}$ such that (3.2$\geq$) holds with $k \geq 2$. Then the family $\{\partial_i f \mid f \in \mathcal{F}', i = 1, \ldots, d\}$ satisfies (3.3) and (3.4) with $l - 1$ in place of $l$; hence, by Proposition 3.4(a), all functions from this family are $(C', 1/(dl(l - 1)))$-good on some neighborhood of $x_0$ with some uniform constant $C'$. Thus, by virtue of Lemma 3.1(b) and (c), the norm of $\nabla f$ is $(C, 1/(dl(l - 1)))$-good with perhaps a different constant $C$.

It remains to consider the case when $k$ as in (3.2$\geq$) is equal to 1. Then $A_1 \leq \|\nabla f(x)\| \leq A_2$ for $x \in B$; therefore for any positive $\varepsilon$ and any $B \subset V$ one has

$$\left| \left\{ x \in B \mid \|\nabla f(x)\| < \varepsilon \cdot \sup_{x \in B} \|\nabla f(x)\| \right\} \right| \leq \left( \frac{A_1}{A_2} \right)^{1/(n-1)} \varepsilon^{1/(n-1)} |B|. \quad \Box$$

**Corollary 3.5.** Let $U$ be an open subset of $\mathbb{R}^d$, $x_0 \in U$, and let $f = (f_1, \ldots, f_n) : U \rightarrow \mathbb{R}^n$ be an $n$-tuple of smooth functions which is $l$-nondegenerate at $x_0$. Then there exist a neighborhood $V \subset U$ of $x_0$ and a positive $C$ such that

(a) any linear combination of $1, f_1, \ldots, f_n$ is $(C, 1/dl)$-good on $V$;

(b) the norm of any linear combination of $\nabla f_1, \ldots, \nabla f_n$ is $(C, 1/(nl(n - 1)))$-good on $V$.

□

**Proof.** Take $f = c_0 + \sum_{i=1}^n c_i f_i$; in view of Lemma 3.1(a), one can without loss of generality assume that the norm of $(c_1, \ldots, c_n)$ is equal to 1. All such functions $f$ belong to a family satisfying (3.3) and (due to the nondegeneracy of $f$ at $x_0$) (3.4); thus Proposition 3.4 applies. \quad \Box

We close the section with the following two auxiliary lemmas that are used below to prove that certain functions are $(C, \alpha)$-good.

**Lemma 3.6.** Let $B$ be a ball in $\mathbb{R}^d$ of radius $r$, let $f \in C^l(B)$, and let $c > 0$ be such that for some unit vector $u$ in $\mathbb{R}^d$, some $k \leq l$, and all $x \in B$ one has $|\partial^k f/\partial u^k(x)| \geq c$. Then
\[
\sup_{x, y \in B} |f(x) - f(y)| \geq \frac{c}{k^k(k + 1)!} (2r)^k.
\] (3.5)

Proof. If \(x_0\) is the center of \(B\), consider the function \(g(t) = f(x_0 + tu)\) defined on \(I \equiv [-r, r]\). Denote \(\sup_{s, t \in I} |f(s) - f(t)|\) by \(\sigma\). We claim that
\[
\sigma \geq \frac{a}{k^k(k + 1)!} (2r)^k;
\] (3.6)

this clearly implies (3.5). To prove (3.6), take any \(s \in I\), divide \(I\) into \(k\) equal segments, and let \(p(t)\) be the Lagrange polynomial of degree \(k\) formed by using values of \(g(s) - g(t)\) at the boundary points of these segments. Then there exists \(t \in I\) such that \(p^{(k)}(t) = g^{(k)}(t)\); hence, by the assumption, \(|p^{(k)}(t)| \geq c\). On the other hand, after differentiating \(p(t)\) \(k\) times (see [KM1, (3.3a)]), one gets \(|p^{(k)}(t)| \leq (k + 1)(\sigma k!)/(2r/k)^k\). Combining the last two inequalities, one easily gets the desired estimate. \(\Box\)

**Lemma 3.7.** Let \(V \subset \mathbb{R}^d\) be an open ball, and let \(\tilde{V}\) be the ball with the same center as \(V\) and double the radius. Let \(f\) be a continuous function on \(\tilde{V}\), and suppose \(C, \alpha > 0\) and \(0 < \delta < 1\) are such that (3.1) holds for any ball \(B \subset V\) and any \(\epsilon \geq \delta\). Then \(f\) is \((C, \alpha')\)-good on \(V\) whenever \(0 < \alpha' < \alpha\) is such that \(CN_d \delta^{\alpha - \alpha'} \leq 1\). (Here \(N_d\) is the constant from Theorem 2.1.)

Proof. Take \(\epsilon > 0\) and a ball \(B \subset V\), and denote
\[
S_{B, \epsilon} \equiv \left\{ x \in \tilde{V} \mid |f(x)| < \epsilon \cdot \sup_{x \in B} |f(x)| \right\}.
\]

The goal is to prove that the measure of \(B \cap S_{B, \epsilon}\) is not greater than \(C \epsilon^\alpha |B|\).

Obviously it suffices to consider \(\epsilon < 1\). Choose \(m \in \mathbb{Z}_+\) such that \(\delta^{m+1} \leq \epsilon < \delta^m\). We show by induction on \(m\) that
\[
|B \cap S_{B, \epsilon}| \leq C^{m+1} N_d^m \epsilon^\alpha |B|. \tag{3.7}
\]

Indeed, the case \(m = 0\) follows from the assumption. Assume that (3.7) holds for some \(m\), and for every \(y \in B \cap S_{B, \epsilon}\) let \(B(y)\) be the maximal ball centered in \(y\) and contained in \(S_{B, \epsilon}\). Observe that, by the continuity of \(f\), one has \(\sup_{x \in B(y)} |f(x)| = \epsilon \cdot \sup_{x \in B} |f(x)|\) for every \(y \in B\). Clearly, the set \(B \cap S_{B, \epsilon}\) is covered by all the balls \(B(y)\), and, using Theorem 2.1, one can choose a subcovering \(\{B_i\}\) of multiplicity at most \(N_d\). Therefore one has
\[
|B \cap S_{B, \epsilon}| \leq \sum_i \left| \left\{ x \in B_i \mid |f(x)| < \delta \sup_{x \in B_i} |f(x)| \right\} \right| \leq \sum_i C \delta^\alpha |B_i| \leq C N_d \cdot C^{m+1} N_d^m (\delta \epsilon)^\alpha |B|,
\]
which proves (3.7) with $\delta \varepsilon$ in place of $\varepsilon$.

It remains to write
\[
C^{m+1}N^d \varepsilon^\alpha = C \cdot (CN_d)^m (\varepsilon)^{\alpha - \alpha'} \varepsilon^{\alpha'} \\
< C \cdot (CN_d)^m (\delta^m)^{\alpha - \alpha'} \varepsilon^{\alpha'} \\
= C \cdot (CN_d \delta^{\alpha - \alpha'})^m \varepsilon^{\alpha'},
\]
which implies that $f$ is $(C, \alpha')$-good on $B$ provided $CN_d \delta^{\alpha - \alpha'} \leq 1$. ■

4 Skew gradients

In this section we define and study the following construction. The main object is a pair of real-valued differentiable functions $g_1, g_2$ defined on an open subset $V$ of $\mathbb{R}^d$, that is, a map $g : V \mapsto \mathbb{R}^2$. For such a pair, let us define its skew gradient $\tilde{\nabla} g : V \mapsto \mathbb{R}^d$ by
\[
\tilde{\nabla} g(x) \overset{\text{def}}{=} g_1(x) \nabla g_2(x) - g_2(x) \nabla g_1(x).
\]
Equivalently, the $i$th component of $\tilde{\nabla} g$ at $x$ is equal to $\left| \begin{array}{cc} g_1(x) & g_2(x) \\ \delta_i g_1(x) & \delta_i g_2(x) \end{array} \right|$, that is, to the signed area of the parallelogram spanned by $g(x)$ and $\delta_i g(x)$. There is another interpretation: if one represents $g(x)$ in polar coordinates, that is, via functions $\rho(x)$ and $\theta(x)$, it is straightforward to verify that $\tilde{\nabla} g(x)$ can be written as $\rho^2(x) \nabla \theta(x)$.

Loosely speaking, the skew gradient measures how different the two functions are from being proportional to each other; it is easy to see that $\tilde{\nabla} g$ is identically equal to zero on an open set if and only if $g_1$ and $g_2$ are proportional (with a locally constant coefficient). Therefore if the image $g(V) \subset \mathbb{R}^2$ does not look like a part of a straight line passing through the origin, one should expect the values of $\tilde{\nabla} g$ to be not very small. Moreover, if the map $g$ is a polynomial map of degree $\leq k$, then $\tilde{\nabla} g$ is a polynomial map of degree less than or equal to $2k - 2$; in particular, its norm is $(C, \alpha)$-good for some $C, \alpha$.

The results of Section 3 suggest that the latter property should be shared by maps that are "close to polynomial" in the sense of Lemma 3.4 (i.e., for families of functions with some uniformly nonvanishing partial derivatives).

The goal of this section is to prove the following result.

**Proposition 4.1.** Let $U$ be an open subset of $\mathbb{R}^d$, $x_0 \in U$, and let $\mathcal{G} \subset C^1(U)$ be a family of maps $g : U \mapsto \mathbb{R}^2$ such that
\[
\text{the family } \{ \nabla g_i \mid g = (g_1, g_2) \in \mathcal{G}, \ i = 1, 2 \} \text{ is compact in } C^{1-1}(U). \quad (4.1)
\]
Also, assume that
\[
\inf_{g \in \mathcal{G}} \sup_{|\beta| \leq 1} |v \cdot \nabla g(x_0)| > 0. \tag{4.2}
\]

(In other words, assume that the partial derivatives of \(g\) at \(x_0\) of order up to \(l\) uniformly in \(g \in \mathcal{G}\) generate \(\mathbb{R}^2\).) Then there exists a neighborhood \(V \subset \mathbb{R}^2\) of \(x_0\) such that

(a) \(\|\nabla g\| \geq (2C_{d,1}, 1/(d(2l - 1)))\)-good on \(V\) for every \(g \in \mathcal{G}\) (here \(C_{d,1}\) as in Lemma 3.2);

(b) for every neighborhood \(B \subset V\) of \(x_0\), there exists \(\rho = \rho(\mathcal{G}, B)\) such that

\[
\sup_{x \in B} \|\nabla g(x)\| \geq \rho \quad \text{for every } g \in \mathcal{G}. \tag*{\square}
\]

To prove Proposition 4.1, we use Lemmas 4.2 and 4.3. Note that in this section for convenience we switch to the Euclidean norm \(\|x\| = \|x\|_e\).

**Lemma 4.2.** Let \(B \subset \mathbb{R}^d\) be a ball of radius \(r\), and let \(g\) be a \(C^1\) map \(B \mapsto \mathbb{R}^2\). Take \(x_0 \in B\) such that \(a = g(x_0) \neq 0\), denote the line connecting \(g(x_0)\) and the origin by \(\mathcal{L}\), let \(\delta = \sup_{x \in B} \|g(x) - g(x_0)\|\), and let \(\omega = \sup_{x \in B} \text{dist}(g(x), \mathcal{L})\). Then

\[
\sup_{x \in B} \|\nabla g(x)\| \geq \frac{\omega(a - \delta)^2}{2r\sqrt{\omega^2 + (a + \delta)^2}}. \tag*{\square}
\]

**Proof.** Let us use polar coordinates, choosing \(\mathcal{L}\) to be the polar axis. Take \(x_1 \in B\) such that \(\text{dist}(g(x_0), \mathcal{L}) = \omega\); then one has

\[
\theta(x_1) \geq \sin \theta(x_1) = \frac{\text{dist}(g(x_0), \mathcal{L})}{\|g(x_1)\|} \geq \frac{\omega}{\sqrt{\omega^2 + (a + \delta)^2}}.
\]

Denote by \(J\) the straight line segment \([x_0, x_1] \subset B\), and denote by \(u\) the unit vector proportional to \(x_1 - x_0\). Restricting \(g\) to \(J\) and using Lagrange’s theorem, one can find \(y\) between \(x_0\) and \(x_1\) such that \(\theta(x_1) = (\partial \theta / \partial u)(y)\). Then one has \(|u \cdot \nabla g(y)| = \rho^2(y)(\partial \theta / \partial u)(y) \geq (a - \delta)^2(\theta(x_1))/\|J\|\), which completes the proof. \(\blacksquare\)

**Lemma 4.3.** Let \(B \subset \mathbb{R}^d\) be a ball of radius \(1\), and let \(p = (p_1, p_2) : B \mapsto \mathbb{R}^2\) be a polynomial map of degree less than or equal to \(l\) such that

\[
\sup_{x, y \in B} \|p(x) - p(y)\| \leq 2 \tag{4.3a}
\]

(the diameter of the image of \(p\) is bounded from above) and

\[
\sup_{x \in B} \text{dist} (\mathcal{L}, p(x)) \geq \frac{1}{8} \quad \text{for any straight line } \mathcal{L} \subset \mathbb{R}^2 \tag{4.3b}
\]

(i.e., the width of \(p(B)\) in any direction is bounded from below). Then there exist

(a) a constant \(0 < \gamma < 1\) (dependent only on \(d\) and \(l\)) such that
\[
\sup_{x \in B} \|\nabla p(x)\| \geq \gamma \left(1 + \sup_{x \in B} \|p(x)\|\right);
\]  
\[\text{(4.4a)}\]

(b) \(M \geq 1\) (dependent only on \(d\) and \(l\)) such that
\[
\sup_{x \in B, i=1,2} \|\nabla p_i(x)\| \leq M.
\]  
\[\text{(4.4b)}\]

Proof. Let \(\mathcal{P}\) be the set of polynomial maps \(p : B \rightarrow \mathbb{R}^2\) of degree \(\leq l\) satisfying (4.3a), (4.3b) and such that \(\sup_{x \in B} \|p(x)\| \leq 6\). We first prove that there exists \(\gamma > 0\) such that (4.4a) holds for any \(p \in \mathcal{P}\). Indeed, otherwise from the compactness of \(\mathcal{P}\) it follows that there exists \(p \in \mathcal{P}\) such that \(\nabla p(x)\) is identically equal to zero. Clearly, this can happen only when all coefficients of \(p\) are proportional to each other, which contradicts (4.3b).

Now assume that \(p\) satisfies (4.3a), (4.3b), and assume that \(a = \|p(y)\| > 6\) for some \(y \in B\). Then one can apply Lemma 4.2 to the map \(p : B \rightarrow \mathbb{R}^2\) to get
\[
\sup_{x \in B} \|\nabla p(x)\| \geq \frac{1}{2\sqrt{\frac{1}{64} + (a + 2)^2}} \geq \frac{1}{16(a - 2)} \geq \frac{1}{32} (a - 2) \geq \frac{1}{64} (a + 1),
\]
which finishes the proof of Lemma 4.3(a). It remains to observe that Lemma 4.3(b) trivially follows from the compactness of the set of polynomials of the form \(\nabla p(x)\) where \(p(x)\) satisfies (4.3a) and has degree less than or equal to \(l\). \[\square\]

Proof of Proposition 4.1. Choose \(0 < \delta < 1/8\) such that
\[
2C_{d,1}N_d \delta^{1/(d(2l - 1)(2l - 2))] \leq 1.
\]
From (4.1) and (4.2) it follows that there exist a neighborhood \(V\) of \(x_0\) and a positive \(c\) such that for every \(g \in \mathcal{S}\) one has
\[
\forall v \in \mathbb{R}^2, \ \exists u \in \mathbb{R}^d, \ \text{and} \ k \leq l \ such \ that \ \inf_{x \in V} |v \cdot \frac{\partial^k g}{\partial u^k}(x)| \geq c \]  
\[\text{(4.5a)}\]
and
\[
\sup_{x, y \in V} \|\beta \partial g(x) - \beta \partial g(y)\| \leq \frac{\delta c v}{8M! (l + 1)!} \text{ for all multi-indices } \beta \text{ with } |eta| = l.
\]  
\[\text{(4.5b)}\]

In view of Lemma 3.7, to show (4.5a) it suffices to prove the following: given any ball \(B = B(x_0, r) \subset V\) and a \(C^1\) map \(g : B \rightarrow \mathbb{R}^2\) such that inequalities (4.5a), (4.5b) hold for all \(x, y \in B\), one has
\[ \left\{ x \in B \mid \| \tilde{\nabla} g(x) \| < \varepsilon \cdot \sup_{x \in B} \| \nabla g(x) \| \right\} \leq 2C_d, \varepsilon^{1/(d(2l-2))} |B| \quad \text{whenever } \varepsilon \geq \delta. \]  

(4.6)

We do this in several steps.

Step 1. Note that conditions (4.5a), (4.5b), as well as the function $\tilde{\nabla}g$, do not change if one replaces $g$ by $Lg$, where $L$ is any rotation of the plane $(g_1, g_2)$. Thus one can choose the $g_1$-axis in such a way that it is parallel to the line connecting the two most distant points of $g(B)$. By (4.5a) there exist $1 \leq k_1, k_2 \leq n$, and $u_1, u_2 \in \mathbb{R}^d$ such that $|\partial^k g_i / \partial u_1^i|(x) \geq c$ for $i = 1, 2$ and all $x \in B$. If $s_i$ stands for $\sup_{x,y \in B} |g_i(x) - g_i(y)|$, $i = 1, 2$, then it follows from Lemma 3.6 that

\[ s_i \geq \frac{c}{k_i^k (k_i + 1)!} (2r)^k \geq \frac{c}{l!(l+1)!} (2r)^l. \]  

(4.7)

Step 2. Here we replace the functions $g_i(x)$ by $1/s_i g_i(x_0 + rx)$, and we replace the ball $B$ by the unit ball $B(0,1)$. This way the function $\tilde{\nabla} g$ is multiplied by a constant, and the statement (4.6) that we need to prove is left unchanged. However, the partial derivatives of order $l$ of the functions $g_i(x)$ are multiplied by factors $r^l/s_i$. In view of (4.7), inequality (4.5b) then implies

\[ \sup_{x,y \in B} \| \partial^\beta g(x) - \partial^\beta g(y) \| \leq \frac{\delta \gamma}{8M} \]  

(4.8)

for all multi-indices $\beta$ with $|\beta| = l$. (Here and until the end of the proof, $B$ stands for $B(0,1)$.) Also, note that it follows from the construction that $g(B)$ is contained in a translate of the square $[-1/2, 1/2]^2$ and that $\sup_{x \in B} \text{dist}(L, g(x)) \geq 1/2\sqrt{2}$ for any straight line $L \subset \mathbb{R}^2$.

Step 3. Here we introduce the $l$th degree Taylor polynomial $p(x)$ of $g(x)$ at zero. Using (4.8), one can show that $p$ is $(\delta \gamma / 8M)$-close to $g$ in the $C^l$ topology; that is,

\[ \sup_{x \in B} \| g(x) - p(x) \| \leq \frac{\delta \gamma}{8M} \quad \text{and} \quad \sup_{x \in B} \| \nabla g_i(x) - \nabla p_i(x) \| \leq \frac{\delta \gamma}{8M}, \quad i = 1, 2. \]

It follows that conditions (4.3a), (4.3b) are satisfied by $p$, and therefore, by Lemma 4.3, inequalities (4.4a), (4.4b) hold.

Step 4. Now let us compare the functions $\tilde{\nabla} g$ and $\tilde{\nabla} p$. One has

\[ \tilde{\nabla} g - \tilde{\nabla} p = (g_1 (\nabla g_2 - \nabla p_2) - (g_2 - p_2) \nabla g_1) - ((g_1 - p_1) \nabla p_2 - p_2 (\nabla g_1 - \nabla p_1)); \]

therefore
\[ \| \tilde{\nabla} g(x) - \tilde{\nabla} p(x) \| \leq \frac{\delta \gamma}{8M} \left( \sup_{x \in B} |g_1(x)| + \sup_{x \in B} \| \nabla g_1(x) \| + \sup_{x \in B} |p_2(x)| + \sup_{x \in B} \| \nabla p_2(x) \| \right) \]
\[ \leq \frac{\delta \gamma}{4M} \left( \left( \sup_{x \in B} \| p(x) \| + \sup_{x \in B} \| \nabla p_2(x) \| \right) + \frac{\delta \gamma}{8M} \right) \]
\[ \leq \frac{3\delta \gamma}{8} \left( 1 + \sup_{x \in B} \| p(x) \| \right) \leq \frac{3\delta}{8} \sup_{x \in B} \| \nabla p(x) \|. \]

Step 5. Finally, we are ready to prove (4.6). Take \( \varepsilon \) between \( \delta \) and 1, put \( s \overset{\text{def}}{=} \sup_{x \in B} \| \nabla p(x) \| \), and observe that, in view of Step 4, the set in the left-hand side of (4.6) is contained in
\[ \left\{ x \in B \mid \| \nabla p(x) \| < \frac{3\delta}{8} s < \varepsilon \left( 1 + \frac{3\delta}{8} \right) s \right\} = \left\{ x \in B \mid \| \nabla p(x) \| < \left( \varepsilon + \frac{3\delta}{8} (1 + \varepsilon) \right) s \right\}. \]

Since \( \varepsilon + 3\delta/8 (1 + \varepsilon) \leq \varepsilon + 3\delta/4 \leq 2\varepsilon \), and since \( \nabla p \) is a polynomial of degree not greater than \( 2l - 2 \), one can apply Lemma 3.2 and conclude that the left-hand side of (4.6) is not greater than
\[ \left| \left\{ x \in B \mid \| \nabla p(x) \| < 2\varepsilon s \right\} \right| \leq C_{d,1} (2\varepsilon)^{1/(d(2l-2))} |B| \leq 2C_{d,1} \varepsilon^{1/(d(2l-2))} |B|, \]
which finishes the proof of Lemma 4.3(a).

As for Lemma 4.3(b), take a ball \( B \subset V \) of the same center and a radius twice as small. It is clear that there exists \( \tau > 0 \) such that for any \( g \in \mathcal{S} \) one can choose \( y \in \widehat{B} \) with \( \| g(y) \| \geq \tau \). (Otherwise, by a compactness argument similar to that of Lemma 4.3, one would get that \( 0 \|_{\widehat{B}} \in \mathcal{S} \), contradicting (4.2).) Also, take \( K \geq \tau/r \) such that
\[ \sup_{g \in \mathcal{S}, x \in B, u \in \mathbb{R}^d} \left\| \frac{\partial g}{\partial u}(x) \right\| \leq K. \tag{4.9} \]

Now let \( B' \subset B \) be a ball of radius \( \tau/2K \leq \tau/2 \) centered at \( y \). Take \( v \in \mathbb{R}^2 \) orthogonal to \( g(y) \). Applying Lemma 3.6 to \( B' \) and the function \( v \cdot g(x) \), one gets
\[ \sup_{x \in B'} |v \cdot g(x)| \geq \frac{c}{k^k(k+1)!} \left( \frac{\tau}{K} \right)^k \geq \frac{c}{l!(l+1)!} \left( \frac{\tau}{K} \right)^l. \]

On the other hand, (4.9) shows that \( \sup_{x \in B'} \| g(x) - g(y) \| \) is not greater than \( \tau/2 \). Now one can apply Lemma 4.2 to the map \( g : B' \rightarrow \mathbb{R}^2 \) to get
\[ \sup_{x \in B'} \| \tilde{\nabla} g(x) \| \geq \frac{c}{l!(l+1)!} \left( \frac{\tau}{K} \right)^{l-1} \left( \frac{\tau}{2} \right)^2 \left( \frac{3\tau}{2} \right)^2. \]
giving a uniform lower bound for $\sup_{x \in B} \| \nabla g(x) \|$. ■

## 5 Theorem 1.4 and lattices

Roughly speaking, the method of lattices simply allows one to write down the system of inequalities (1.7b) from Theorem 1.4 in an intelligent way. In what follows, we let $m$ stand for $n + d + 1$. Denote the standard basis of $\mathbb{R}^m$ by $\{e_0, e_1^*, \ldots, e_d^*, e_1, \ldots, e_n\}$. Also, denote by $\Lambda$ the intersection of $\mathbb{Z}^m$ with the span of $e_0, e_1, \ldots, e_n$; that is,

$$\Lambda = \left\{ \begin{pmatrix} p \\ 0 \\ q \end{pmatrix} \bigg| p, q \in \mathbb{Z}, q \in \mathbb{Z}^n \right\}. \tag{5.1}$$

Take $f : U \mapsto \mathbb{R}^n$ as in Theorem 1.4, and let $U_x$ stand for the matrix

$$U_x \overset{\text{def}}{=} \begin{pmatrix} 1 & 0 & f(x) \\ 0 & I_d & \nabla f(x) \\ 0 & 0 & I_n \end{pmatrix} \in \text{SL}_m(\mathbb{R}). \tag{5.2}$$

Note that $U_x \begin{pmatrix} p \\ 0 \\ q \end{pmatrix} = \begin{pmatrix} f(x)q + p \\ q \end{pmatrix}$ is the vector whose components appear in the right-hand sides of the inequalities in (1.7b). Therefore the fact that there exists $q \in \mathbb{Z}^n \setminus \{0\}$ satisfying (1.7b) implies the existence of a nonzero element of $U_x \Lambda$ which belongs to a certain parallelepiped in $\mathbb{R}^m$. Our strategy is as follows. We find a diagonal matrix $D \in \text{GL}_m(\mathbb{R})$ that transforms the above parallelepiped into a small cube; then the solvability of the above system of inequalities forces the lattice $DU_x \Lambda$ to have a small nonzero vector, and we use a theorem proved by methods from [KM1] (see Theorem 6.2) to estimate the measure of the set of $x \in B$ for which it can happen.

Specifically, take $\delta, K, T_1, \ldots, T_n$ as in Theorem 1.4, fix $\varepsilon > 0$, and denote

$$D = \text{diag} \left( a_0^{-1}, a_1^{-1}, \ldots, a_1^{-1}, a_1^{-1}, \ldots, a_n^{-1} \right), \tag{5.3}$$

where

$$a_0 = \frac{\delta}{\varepsilon}, \quad a_1 = \frac{K}{\varepsilon}, \quad a_i = \frac{T_i}{\varepsilon}, \quad i = 1, \ldots, n. \tag{5.4}$$

It is easily seen that set (1.7b) is exactly equal to

$$\{ x \in B \mid \| DU_x v \| < \varepsilon \text{ for some } v \in \Lambda \setminus \{0\} \}, \tag{5.5}$$

where $\| \cdot \|$ stands for the supremum norm. However, from this point on it is more con-
venient to use the Euclidean norm $\| \cdot \|_e$ on $\mathbb{R}^m$. Let us now state a theorem from which Theorem 1.4 can be easily derived.

**Theorem 5.1.** Let $U, x_0, d, l, n,$ and $f$ be as in Theorem 1.4. Take $\Lambda$ as in (5.1) and $U_\alpha$ as in (5.2). Then there exists a neighborhood $V \subset U$ of $x_0$ with the following property: for any ball $B \subset V$ there exists $E > 0$ such that for any diagonal matrix $D$ as in (5.3) with

\[ 0 < a_0 \leq 1, \quad a_n \geq \cdots \geq a_1 \geq 1, \quad \text{and} \quad 0 < a_* \leq \left( a_0 a_1 \cdots a_{n-1} \right)^{-1} \]  

(5.6)

and for any positive $\epsilon$, one has

\[ \left| \left\{ x \in B \mid \| D U_\alpha v \|_e < \epsilon \text{ for some } v \in \Lambda \setminus \{0\} \right\} \right| \leq E \epsilon^{1/(d(2l-1))} |B|. \]  

(5.7)

**Proof of Theorem 1.4 modulo Theorem 5.1.** Take $V \subset U$, and, given any ball $B \subset V$, choose $E$ as in Theorem 5.1. Then take $\delta, T_1, \ldots, T_n$, and $K$ satisfying (1.7a). Observe that without loss of generality one can assume that $T_1 \leq \cdots \leq T_n$. (Otherwise, one can replace $T_1, \ldots, T_n$ by a permutation $T_{i_1}, \ldots, T_{i_n}$ with $T_{i_1} \leq \cdots \leq T_{i_n}$ and consider the $n$-tuple $(f_{i_1}, \ldots, f_{i_n})$ instead of the original one, which is still nondegenerate at $x_0$.)

Now take $\epsilon$ as in (1.7c), and define $a_0, a_*, a_1, \ldots, a_n$ as in (5.4). Then all the constraints in (5.6) easily follow. (Indeed, (1.7c) shows that $a_0 \leq 1$ and $a_0 a_1, \ldots, a_{n-1} \leq 1$, while (1.7a) implies that $\epsilon \leq 1$; hence $a_i \geq 1$.) Thus Theorem 5.1 applies, and, to complete the proof, it remains to observe that set (1.7b) $= (5.5)$ is contained in the set

\[ \left\{ x \in B \mid \| D U_\alpha v \|_e < \epsilon \sqrt{m} \text{ for some } v \in \Lambda \setminus \{0\} \right\}; \]

hence its measure is not greater than $E \epsilon^{1/(d(2l-1))} 1/(\epsilon^{d(2l-1)}) |B|$.  

\[ \square \]

6 **Lattices and posets**

The proof of Theorem 5.1 depends on a result from [KM1] involving mappings of partially ordered sets into spaces of $(C, \alpha)$-good functions. Let us recall some terminology from [KM1, §4]. For $d \in \mathbb{N}$, $k \in \mathbb{Z}_+$, and $C, \alpha, \rho > 0$, define $A(d, k, C, \alpha, \rho)$ to be the set of triples $(S, \varphi, B)$ where $S$ is a partially ordered set (poset), $B = B(x_0, r_0)$, where $x_0 \in \mathbb{R}^d$ and $r_0 > 0$, and $\varphi$ is a mapping from $S$ to the space of continuous functions on $B \overset{\text{def}}{=} B(x_0, 3^kr_0)$ (this mapping is denoted by $s \rightarrow \varphi_s$) such that the following hold:

(A0) the length of $S$ is not greater than $k$;

(A1) $\forall s \in S$, $\varphi_s$ is $(C, \alpha)$-good on $B$;

(A2) $\forall s \in S$, $\| \varphi_s \|_B \geq \rho$;

(A3) $\forall x \in B$, $\# \{ s \in S \mid |\varphi_s(x)| < \rho \} < \infty$. 

We need this exterior power representation mainly to be able to measure the “size” of \( (\varepsilon, S, \varphi) \)-marked if there exists a linearly ordered subset \( \Sigma_x \) of \( S \) such that

(\( M1 \)) \( \varepsilon \leq |\varphi_x(x)| \leq \rho, \forall s \in \Sigma_x; \)

(\( M2 \)) \( |\varphi_x(x)| \geq \rho, \forall s \in S \setminus \Sigma_x \) comparable with any element of \( \Sigma_x \).

We denote by \( \Phi(\varepsilon, S, B) \) the set of all the \( (\varepsilon, S, \varphi) \)-marked points \( x \in B \).

**Theorem 6.1 [KM1, Theorem 4.1].** Let \( d \in \mathbb{N} \), let \( k \in \mathbb{Z}_+ \), and let \( C, \alpha, \rho > 0 \) be given. Then for all \( (S, \varphi, B) \in A(d, k, C, \alpha, \rho) \) and \( 0 < \varepsilon \leq \rho \) one has

\[
|B \setminus \Phi(\varepsilon, S, \varphi, B)| \leq kC(3^d N_d)^k \left( \frac{\varepsilon}{\rho} \right)^\alpha |B|.
\]

We apply Theorem 6.1 to the poset of subgroups of the group of integer points of a finite-dimensional real vector space \( W \). For a discrete subgroup \( \Gamma \) of \( W \), we denote by \( \Gamma_\mathbb{R} \) the minimal linear subspace of \( W \) containing \( \Gamma \). Let \( k = \dim(\Gamma_\mathbb{R}) \) be the rank of \( \Gamma \); we say that \( w \in \wedge^k(W) \) represents \( \Gamma \) if

\[
w = \begin{cases} 
1 & \text{if } k = 0, \\
v_1 \land \cdots \land v_k & \text{if } k > 0 \text{ and } v_1, \ldots, v_k \text{ is a basis of } \Gamma.
\end{cases}
\]

We need this exterior power representation mainly to be able to measure the “size” of discrete subgroups. Namely, these “sizes” are given by suitable “norm-like” functions \( \varphi \) on the exterior algebra \( \wedge(W) \) of \( W \), and we set

\[
\varphi(\Gamma) = \varphi(w) \quad \text{if } w \text{ represents } \Gamma.
\]

More precisely, let us say that a function \( \varphi : \wedge(W) \to \mathbb{R}_+ \) is submultiplicative if

1. \( \varphi \) is continuous (in the natural topology);
2. it is homogeneous; that is, \( \varphi(tw) = |t|\varphi(w) \) for all \( t \in \mathbb{R} \) and \( w \in \wedge(W) \);
3. \( \varphi(u \land w) \leq \varphi(u) \varphi(w) \) for all \( u, w \in \wedge(W) \).

Note that, in view of condition (2), (6.1) is a correct definition of \( \varphi(\Gamma) \).

Examples. If \( W \) is a Euclidean space, one can extend the Euclidean structure to \( \wedge(W) \) (by making \( \wedge^i(W) \) and \( \wedge^j(W) \) orthogonal for \( i \neq j \)); clearly, then the Euclidean norm \( \varphi(w) = ||w|| \) is submultiplicative. In this case the restriction of \( \varphi \) to \( W \) coincides with the usual (Euclidean) norm on \( W \). Also, if \( W \subset \wedge(W) \) is an ideal, one can define \( \varphi(w) \) to be the norm of the projection of \( w \) orthogonal to \( W \). If this ideal is orthogonal to \( W \subset \wedge(W) \), again the function \( \varphi \) coincides with the norm when restricted to \( W \).

We also need the notion of primitivity of a discrete subgroup. If \( \Lambda \) is a discrete subgroup of \( W \), we say that a subgroup \( \Gamma \) of \( \Lambda \) is primitive (in \( \Lambda \)) if \( \Gamma = \Gamma_\mathbb{R} \cap \Lambda \), and we denote by \( \mathcal{L}(\Lambda) \) the set of all nonzero primitive subgroups of \( \Lambda \). For example, a cyclic...
subgroup of $\mathbb{Z}^1$ is primitive in $\mathbb{Z}^1$ if and only if it is generated by a primitive vector, that is, by a vector that is not equal to a nontrivial multiple of another element of $\mathbb{Z}^1$. Note that the inclusion relation makes $\mathcal{L}(\Lambda)$ a poset, its length being equal to the rank of $\Lambda$.

The following result, which we derive from Theorem 6.1, is a generalization of [KM1, Theorem 5.2].

**Theorem 6.2.** Let $W$ be a finite-dimensional real vector space, let $\Lambda$ be a discrete subgroup of $W$ of rank $k$, and let a ball $B = B(x_0, r_0) \subset \mathbb{R}^d$ and a map $H : \overline{B} \to \text{GL}(W)$ be given, where $\overline{B}$ stands for $B(x_0, 3^k r_0)$. Take $C, \alpha > 0$, $0 < \rho \leq 1$, and let $\nu$ be a submultiplicative function on $\Lambda(W)$. Assume that, for any $\Gamma \in \mathcal{L}(\Lambda)$,

(i) the function $x \mapsto \nu(H(x)\Gamma)$ is $(C, \alpha)$-good on $\overline{B}$; and

(ii) $\exists x \in B$ such that $\nu(H(x)\Gamma) \geq \rho$.

Also, assume that

(iii) $\forall x \in \overline{B}$, $\# \{|\Gamma \in \mathcal{L}(\Lambda) \mid \nu(H(x)\Gamma) < \rho\} < \infty$.

Then for any positive $\epsilon \leq \rho$ one has

$$\left|\{ x \in B \mid \nu(H(x)\nu) < \epsilon \text{ for some } \nu \in \Lambda \setminus \{0\}\right| \leq k(3^d N_d)^k \cdot C \left(\frac{\epsilon}{\rho}\right)^\alpha |B|. \quad (6.2)$$

Note that when $\nu|_W$ agrees with the Euclidean norm, (6.2) estimates the measure of $x \in B$ for which the subgroup $H(x)\Lambda$ has a nonzero vector with length less than $\epsilon$.

**Proof.** We apply Theorem 6.1 to the triple $(S, \varphi, B)$, where $S = \mathcal{L}(\Lambda)$ and $\varphi$ is defined by $\varphi_{\Gamma}(x) \overset{\text{def}}{=} \nu(H(x)\Gamma)$. It is easy to verify that $(S, \varphi, B) \in \mathcal{A}(d, k, C, \alpha, \rho)$. Indeed, the functions $\varphi_{\Gamma}$ are continuous since so is $H$ and $\nu$, property (A0) is clear, (A1) is given by condition (i), (A2) by condition (ii), and (A3) by condition (iii).

In view of Theorem 6.1, it remains to prove that

$$\Phi(\epsilon, S, \varphi, B) \subset \{ x \in B \mid \nu(H(x)\nu) \geq \epsilon \text{ for all } \nu \in \Lambda \setminus \{0\}\}. \quad (6.3)$$

Take an $(\epsilon, S, \varphi)$-marked point $x \in B$, and let $\{0\} = \Gamma_0 \subsetneq \Gamma_1 \subsetneq \cdots \subsetneq \Gamma_m = \Lambda$ be all the elements of $\Sigma \setminus \{0, \Lambda\}$. Take any $\nu \in \Lambda \setminus \{0\}$. Then there exists $i$, $1 \leq i \leq m$, such that $\nu \in \Gamma_i \setminus \Gamma_{i-1}$. Denote $(\Gamma_{i-1} + \mathbb{R} \nu) \cap \Lambda$ by $\Delta$. Clearly, $\Delta$ is a primitive subgroup of $\Gamma$ contained in $\Gamma_i$; therefore it is comparable to any element of $\Sigma$. By submultiplicativity of $\nu$, one has $\nu(H(x)\Delta) \leq \nu(H(x)\Gamma_{i-1})\nu(H(x)\nu)$. Now one can use properties (M1) and (M2) to deduce that $|\varphi_\Delta(\epsilon)| = \nu(H(x)\Delta) \geq \min(\epsilon, \rho) = \epsilon$ and then conclude that

$$\nu(H(x)\nu) \geq \frac{\nu(H(x)\Delta)}{\nu(H(x)\Gamma_{i-1})} \geq \frac{\epsilon}{\rho} \geq \epsilon.$$
This shows (6.3) and completes the proof of the theorem. ☐

7 Proof of Theorem 5.1

Here we take $U$, $x_0$, $d$, $l$, $n$, and $f$ as in Theorem 1.4, we set $m = n + d + 1$ and $W = \mathbb{R}^m$, and we use the notation introduced in Section 5. Denote by $W^\perp$ the $d$-dimensional subspace spanned by $e_1^*, \ldots, e_d^*$, so that $\Lambda$ as in (5.1) is equal to the intersection of $\mathbb{Z}^m$ and $(W^\perp)^\perp$. Also, let $\mathcal{H}$ be the family of functions $H : \mathbb{U} \to \text{GL}_m(\mathbb{R})$ given by $H(x) = DU_x$, where $U_x$ is as in (5.2) and where $D$ is as in (5.3) with coefficients satisfying (5.6).

In order to use Theorem 6.2, we also need to choose the submultiplicative function $\nu$ on $W$ in a special way. Namely, we let $W \subset \wedge(W)$ be the ideal generated by $\wedge^2(W^*)$, we denote by $\pi$ the orthogonal projection with kernel $W$, and we take $\nu(w)$ to be the Euclidean norm of $\pi(w)$. In other words, if $w$ is written as a sum of exterior products of base vectors $e_i$ and $e_i^*$, to compute $\nu(w)$ one should ignore components containing $e_i^* \wedge e_j^*$, $1 \leq i \neq j \leq d$, and take the norm of the sum of the remaining components.

Since $\nu|_W$ agrees with the Euclidean norm, to derive Theorem 5.1 from Theorem 6.2, it suffices to find a neighborhood $\tilde{V} \ni x_0$ such that

1. there exists $C > 0$ such that all the functions $x \mapsto \nu(H(x)\Gamma)$, where $H \in \mathcal{H}$ and $\Gamma \in \mathcal{L}(\Lambda)$ are $(C, 1/(d(2l + 1)))$-good on $\tilde{V}$;

2. for every ball $B \subset \tilde{V}$, there exists $\rho > 0$ such that $\sup_{x \in B} \nu(H(x)\Gamma) \geq \rho$ for all $H \in \mathcal{H}$ and $\Gamma \in \mathcal{L}(\Lambda)$; and

3. for all $x \in \tilde{V}$ and $H \in \mathcal{H}$, one has $\#(\Gamma \in \mathcal{L}(\Lambda) \mid \nu(H(x)\Gamma) \leq 1) < \infty$.

Indeed, then one can take a smaller neighborhood $V$ of $x_0$ such that whenever $B = B(x, r)$ lies in $V$ its dilate $\tilde{B} = B(x, 3^{n+1}r)$ is contained in $\tilde{V}$. This way it would follow from Theorem 6.2 that for any $B \subset V$ the measure of the set in (5.7) is not greater than $C(n + 1)(3dN_d)^{n+1}(\varepsilon/\rho)^{1/d(2l+1)} |B|$ for any $\varepsilon \leq \rho$; therefore it is not greater than

$$\max \left( C(n + 1)(3dN_d)^{n+1}, 1 \right) \rho^{-1/d(2l+1)} \varepsilon^{1/d(2l+1)} |B|$$

for any positive $\varepsilon$.

Thus we are led to explicitly computing the functions $\nu(H(x)\Gamma)$ for arbitrary choices of subgroups $\Gamma \subset \Lambda$ and positive numbers $a_i$, $i = 0, *, 1, \ldots, n$. In fact, we do this in two different ways, which are relevant for checking conditions 1 (along with 3) and 2 respectively.

Let $k$ be the rank of $\Gamma$. The claims are trivial for $\Gamma = \{0\}$; thus we can set $1 \leq k \leq n + 1$. Since $D\Gamma$ is a $k$-dimensional subspace of $(W^*)^\perp = \mathbb{R}e_0 \oplus \mathbb{R}e_1 \oplus \cdots \oplus \mathbb{R}e_n$, it is possible to choose an orthonormal set $v_1, \ldots, v_{k-1} \in D\Gamma$ such that each $v_i, i = 1, \ldots, k-1$, is orthogonal to $e_0$. Now let us consider two cases.
Case 1: $D\Gamma_\mathbb{R}$ contains $e_0$. Then $\{e_0, v_1, \ldots, v_{k-1}\}$ is a basis of $\mathbb{R}e_0 \oplus D\Gamma_\mathbb{R}$. Thus one can find $w \in \bigwedge^k(\mathbb{R}^m)$ representing $\Gamma$ such that $Dw$ can be written as $ae_0 \wedge v_1 \cdots \wedge v_{k-1}$ for some $a > 0$.

Case 2: $D\Gamma_\mathbb{R}$ does not contain $e_0$. Then it is possible to choose $v_0 \in \mathbb{R}e_0 \oplus D\Gamma_\mathbb{R}$ such that $\{e_0, v_0, v_1, \ldots, v_{k-1}\}$ is an orthonormal basis of $\mathbb{R}e_0 \oplus D\Gamma_\mathbb{R}$. In this case, one can represent $\Gamma$ by $w$ such that

$$
Dw = (ae_0 + bv_0) \wedge v_1 \cdots \wedge v_{k-1}
= ae_0 \wedge v_1 \cdots \wedge v_{k-1} + bv_0 \wedge v_1 \cdots \wedge v_{k-1}
$$

for some $a, b \in \mathbb{R} \setminus \{0\}$. In fact, (7.1) is valid in Case 1 as well; one simply has to put $b$ equal to zero and bear in mind that the vector $v_0$ is not defined.

Now write

$$
H(x)w = DU_x D^{-1}(Dw),
$$
and introduce the $m$-tuple (interpreted as a row vector) of functions

$$
\hat{f}(x) = \left(1, 0, \ldots, 0, \frac{a_1}{a_0} f_1(x), \ldots, \frac{a_n}{a_0} f_n(x)\right).
$$

It is convenient to identify $\mathbb{R}^d$ with $W^*$ and introduce the $W^*$-valued gradient $\nabla^* = \sum_{i=1}^d e_i^* \partial_i$ of a scalar function on $U \subset \mathbb{R}^d$, so that $\nabla^* f(x) = \sum_{i=1}^d \partial_i f(x) e_i^*$. As a step further, we let the skew gradients discussed in Section 4 take values in $W^*$ as well. That is, for a map $g = (g_1, g_2) : \mathbb{R}^d \to \mathbb{R}^2$, we define $\nabla^* g : \mathbb{R}^d \to W^*$ by

$$
\nabla^* g(x) \overset{\text{def}}{=} g_1(x) \nabla^* g_2(x) - g_2(x) \nabla^* g_1(x).
$$

Then it becomes straightforward to verify that $DU_x D^{-1} e_0 = e_0$ and

$$
DU_x D^{-1} v = v + (\hat{f}(x) v) e_0 + \frac{a_0}{a_x} \nabla^* (\hat{f}(x) v)
$$

whenever $v$ is orthogonal to $e_0$ and $W^*$. Therefore the space $\mathbb{R}e_0 \oplus W^* \oplus D\Gamma_\mathbb{R}$ is invariant under $DU_x D^{-1}$; hence we can restrict ourselves to the coordinates of the $DU_x D^{-1}$-image of $Dw$ with respect to the basis chosen above, and we write

$$
DU_x D^{-1} (e_0 \wedge v_1 \cdots \wedge v_{k-1}) = e_0 \wedge v_1 \wedge \cdots \wedge v_{k-1}
+ \frac{a_0}{a_x} \sum_{i=1}^{k-1} \pm \nabla^* (\hat{f}(x) v_i) e_0 \wedge \bigwedge_{s \neq i} v_s + w_i^*
$$

and
Indeed, in view of Lemma 3.1(b), looking at the first line of (7.3), one follows from the nondegeneracy of \( \nabla g \) at \( x_0 \) that condition (4.2) is satisfied.

One can also use (7.3) to prove \( \mathcal{G} \). Indeed, looking at the first line of (7.3), one
sees that
\[ \nu(H(x)\Gamma) \leq \max \left( |a + b\hat{f}(x)v_0|, |b| \right); \]
thus \( \nu(H(x)\Gamma) \leq 1 \) implies \( |b| \leq 1 \) and \( |a| \leq 1 + \|\hat{f}(x)\| \). On the other hand, in view of (7.1), the norm of \( Dw \) is equal to \( \sqrt{a^2 + b^2} \), and from the discreteness of \( \Lambda(\Lambda) \) in \( \Lambda(\mathbb{R}^m) \) it follows that for any \( R > 0 \) the set of \( \Gamma \in \mathcal{L}(\Lambda) \) such that both \( |a| \) and \( |b| \) are bounded from above by \( R \) is finite.

We now turn to condition 2. Take any neighborhood \( B \) of \( x_0 \). It follows from the linear independence of the functions \( 1, f_1, \ldots, f_n \) and from the linear independence of their gradients that there exists \( \rho_1 > 0 \) such that
\[ \forall v \in \mathbb{R}^n, \forall v_0 \in \mathbb{R}, \sup_{x \in B} |f(x)v + v_0| \geq \rho_1 \quad \text{and} \quad \sup_{x \in B} |\nabla f(x)v| \geq \rho_1. \quad (7.5) \]
Also, let \( \rho_2 = \rho(G, B) \), where \( G \) is the class of 2-tuples of functions defined in (7.4) and where \( \rho(G, B) \) is as in Proposition 4.1(b). Consider
\[ M \overset{\text{def}}{=} \max \left( \sup_{x \in B} \|f(x)\|, \sup_{x \in B} \|\nabla f(x)\| \right); \]
we show that Theorem 6.2(ii) holds for any nonzero subgroup \( \Gamma \) of \( \Lambda \) if we choose
\[ \rho = \frac{\rho_1 \rho_2}{\sqrt{\rho_1^2 + (\rho_2 + 2M^2)^2}}. \]

First let us consider the case \( k = \dim(\Gamma_k) = 1 \). Then \( \Gamma \) can be represented by a vector \( v = (v_0, 0, \ldots, v_1, \ldots, v_n)^T \) with integer coordinates, and it is straightforward to verify that the first coordinate of \( H(x)v \) is equal to \( 1/a_0(v_0 + v_1 f_1(x) + \cdots + v_n f_n(x)) \), which deviates from zero by not less than \( \rho \) at some point of \( B \) due to (7.5) and since \( \rho \leq \rho_1 \) and \( a_0 \leq 1 \).

Now let \( k \) be greater than 1. Our method is similar to that of the proof of condition 1: given \( w \in \Lambda^k(\mathbb{R}^m) \) representing \( \Gamma \), we choose a suitable orthogonal decomposition of \( \Lambda^k(\mathbb{R}^m) \), and then we show that the norm of the projection of \( H(x)w \) to some subspace is not less than \( \rho \) for some \( x \in B \).

In order to prove the desired estimate, it is important to pay special attention to the vector \( e_n \), which is the eigenvector of \( D \) with the smallest eigenvalue. We do this by first choosing an orthonormal set \( v_1, \ldots, v_{k-2} \in \Gamma_R \) such that each \( v_i, i = 1, \ldots, k-2 \), is orthogonal to both \( e_0 \) and \( e_n \). Then choose \( v_{k-1} \) orthogonal to \( v_i, i = 1, \ldots, k-2 \), and to \( e_0 \) (but, in general, not to \( e_n \)). Now, if necessary (see the remark after (7.1)), choose
Indeed, since \( v_0 \) to complete \( \{e_0, v_1, \ldots, v_{k-1}\} \) to an orthonormal basis of \( \mathbb{R}e_0 \oplus \Gamma \). This way, similarly to (7.1), we represent \( \Gamma \) by \( w \) of the form

\[
w = (ae_0 + bv_0) \wedge v_1 \cdots \wedge v_{k-1} = ae_0 \wedge v_1 \cdots \wedge v_{k-1} + bv_0 \wedge \cdots \wedge v_{k-1} \tag{7.6}
\]

for some \( a, b \in \mathbb{R} \) with \( a^2 + b^2 \geq 1 \). As before, we use (7.6) even when \( v_0 \) is not defined; in this case the coefficient \( b \) vanishes.

Now, similarly to the proof of condition (i), introduce the \( m \)-tuple of functions

\[
\hat{f}(x) = (1, 0, \ldots, 0, f_1(x), \ldots, f_n(x)),
\]

and observe that \( U_x e_0 = e_0 \) and

\[
U_x v = v + (\hat{f}(x)v) e_0 + \nabla^* (\hat{f}(x)v)
\]

whenever \( v \) is orthogonal to \( e_0 \) and \( e_\ast \). Using this and (7.6), one can obtain an expression analogous to (7.3). This time, however, we are interested only in terms of the form \( e_0 \wedge e_*^\perp \wedge w' \), where \( w' \) is orthogonal to \( \Lambda^{k-2}(\mathbb{R}e_0 \oplus W^*)^\perp \). (Note that these terms are present only if \( k \geq 2 \).) Namely, let us write

\[
\pi(U_x w) = (a + b(\hat{f}(x)v_0)) e_0 \wedge v_1 \wedge \cdots \wedge v_{k-1} + bv_0 \wedge \cdots \wedge v_{k-1} + b \sum_{i=1}^{k-1} \pm (\hat{f}(x)v_i) e_0 \wedge \bigwedge_{s \neq i} v_s + b \sum_{i=0}^{k-1} \pm \nabla^* (\hat{f}(x)v_i) \wedge \bigwedge_{s \neq i} v_s + e_0 \wedge \hat{w}(x),
\]

where

\[
\hat{w}(x) \overset{\text{def}}{=} \sum_{i=1}^{k-1} \pm \nabla^* (\hat{f}(x)v_i) e_0 \wedge \bigwedge_{s \neq i} v_s + b \sum_{i,j=1}^{k-1} \pm \nabla^* (\hat{f}(x)v_i) \hat{f}(x)v_j) \wedge \bigwedge_{s \neq i, j} v_s. \tag{7.7}
\]

Note that \( e_0 \wedge \hat{w}(x) \) lies in the space \( e_0 \wedge W^* \wedge \Lambda^{k-2}(W^*)^\perp \), while \( \pi(U_x w) - e_0 \wedge \hat{w}(x) \) belongs to its orthogonal complement. Since both spaces are \( D \)-invariant, to prove that \( \psi(H(x)w) = \|\pi(DU_x w)\| = \|D\pi(U_x w)\| \) is not less than \( \rho \) for some \( x \in B \), it will suffice to show that \( \sup_{x \in B} \|D\hat{w}(x)\| \) is not less than \( a_0 \rho \).

Now consider the product \( e_n \wedge \hat{w}(x) \). We claim that it is enough to show that

\[
\|e_n \wedge \hat{w}(x)\| \geq \rho \quad \text{for some} \ x \in B. \tag{7.8}
\]

Indeed, since \( e_n \) is an eigenvector of \( D \) with eigenvalue \( a_n^{-1} \), for any \( x \in B \) the norm of
D(\(e_n \land \hat{w}(x)\)) is not greater than \(a_n^{-1} \|D\hat{w}(x)\|\). Therefore, since the smallest eigenvalue of D on \(W^* \land \bigwedge^{k-2}(W^*)^\perp\) is equal to \((a_0a_{n-k+1} \cdots a_n)^{-1}\), the norm of \(D\hat{w}(x)\) is not less than
\[
a_n \|D(e_n \land \hat{w}(x))\| \geq \frac{a_n}{a_0a_{n-k+1} \cdots a_n} \|e_n \land \hat{w}(x)\| \geq \frac{\rho}{a_0a_1 \cdots a_{n-1}} \geq \frac{a_0\rho}{\rho_1 a - 2M^2b} \geq a_0\rho.
\]

Thus it remains to prove (7.8). For this let us select the term containing \(v_1 \land \cdots \land v_{k-2}\), and let us multiply (7.7) by \(e_n\) as follows:
\[
e_n \land \hat{w}(x) = \pm v^*(x) \land e_n \land v_1 \land \cdots \land v_{k-2} + \text{other terms where one or two of } v_i, \ i = 1, \ldots, k-2, \text{ are missing,}
\]
where
\[
v^*(x) \overset{\text{def}}{=} \nabla^* (\hat{f}(x)v_{k-1}, a + b\hat{f}(x)v_0) = b\nabla^* (\hat{f}(x)v_{k-1}, \hat{f}(x)v_0) - \nabla^* (\hat{f}(x)v_{k-1}).
\]

Because of the orthogonality of the two summands in (7.9), and also because \(e_n\) is orthogonal to \(v_i, \ i = 1, \ldots, k-2\), it follows that \(\|e_n \land \hat{w}(x)\|\) is not less than \(\|v^*(x)\|\). It follows from the first expression for \(v^*(x)\) that \(\sup_{x \in B} \|v^*(x)\| \geq \rho_2 b\), and it follows from the second expression that \(\sup_{x \in B} \|v^*(x)\| \geq \rho_1 a - 2M^2b\). An elementary computation shows that \(\rho\) as defined by (7.5) is not greater than \(\min_{a^2 + b^2 \geq 1} \max(\rho_2 b, \rho_1 a - 2M^2b)\). This completes the proof of (7.8) and hence of Theorem 5.1.

8 Completion of the proof and concluding remarks

8.1

First let us finish the proof of Theorem 1.1 by writing the following reduction.

Reduction of Theorem 1.1 to Theorems 1.3 and 1.4. Recall that we are given an open subset \(U\) of \(\mathbb{R}^d\), an \(n\)-tuple \(f = (f_1, \ldots, f_n)\) of \(C^m\) functions on \(U\), and a function \(\Psi : \mathbb{Z}^n \setminus \{0\} \rightarrow (0, \infty)\) satisfying (1.1) and (1.5). Take \(x_0 \in U\) such that \(f\) is 1-nondegenerate at \(x_0\) for some \(l \leq m\), choose \(V \subset U\) as in Theorem 1.4, and pick a ball \(B \subset V\) containing \(x_0\) such that its dilate \(\tilde{B}\) (the ball with the same center as \(B\) and double the radius) is contained in \(U\). We prove that for almost every \(x \in B\) one has \(f(x) \in W(\Psi)\). In other words, define \(A(q)\) to be the set of \(x \in B\) satisfying \(|\langle f(x)q \rangle| < \Psi(q)\); we show that points \(x\)
that belong to infinitely many sets \( A(q) \) form a set of measure zero.

We proceed by induction on \( n \). If \( n \geq 2 \), let us assume that the claim is proven for any nondegenerate \((n-1)\)-tuple of functions. Because of the induction assumption and the fact that projections of a nondegenerate manifold are nondegenerate, we know that almost every \( x \in B \) belongs to at most finitely many sets \( A(q) \) such that \( q_i = 0 \) for some \( i = 1, \ldots, n \). It remains to show that the same is true if one includes integer vectors \( q \) with all coordinates different from zero. (If \( n = 1 \), there is no difference, so the argument below provides both the base and the induction step.)

Take \( L \) as in (1.6a), denote by \( A_{\geq}(q) \) the set of \( x \in A(q) \) satisfying (1.6d), and set \( A_{<}(q) \) to be \( A(q) \setminus A_{\geq}(q) \). Theorem 1.3 guarantees that the measure of \( A_{\geq}(q) \) is not greater than \( C_d \Psi(q)|B| \) whenever \( q \) is far enough from the origin. Because of (1.1), the sum of measures of the sets \( A_{\geq}(q) \) is finite; hence, by the Borel-Cantelli lemma, almost every \( x \in B \) is contained in at most finitely many sets \( A_{\geq}(q) \).

Our next task is to use Theorem 1.4 to estimate the measure of the union

\[
\bigcup_{q \in \mathbb{Q}, 2^{t_1} \leq |q_i| < 2^{t_i+1}} A_{<}(q) \tag{8.1}
\]

for any \( n \)-tuple \( t = (t_1, \ldots, t_n) \in \mathbb{Z}^n_+ \) with large enough \( \|t\| = \max_i t_i \). Observe that conditions (1.1) and (1.5) imply that \( \Psi(q) \leq (\prod |q_i|)^{-1} \) whenever \( q \) is far enough from the origin. It follows that \( \Psi(q) \leq 2^{-\sum t_i} \) whenever \( q \) satisfies the restrictions of (8.1) with \( t \) far enough from the origin. Therefore for such \( t \) set (8.1) is contained in set (1.7b) where one puts \( \delta = 2^{-\sum t_i}, K = \sqrt{n dL2^{\|t\|}}/2 \), and \( T_i = 2^{t_i+1} \). It is straightforward to verify that inequalities (1.7a) are satisfied whenever \( \|t\| \) is large enough; in fact, one has

\[
\frac{\delta K T_1 \cdots T_n}{\max_i T_i} = 2^{-\sum t_i} \sqrt{n dL2^{\|t\|}/2} 2^{n+\sum t_i} = \sqrt{n dL2^{n-1-\|t\|}/2},
\]

which for large \( \|t\| \) is less than 1 but larger than \( \delta^{n+1} \). Therefore \( \varepsilon \) as in (1.7c) is equal to \( \sqrt{n dL2^{n-1}2^{-1/(2(n+1))\|t\|}} \), so, by Theorem 1.4, the measure of set (8.1) is at most

\[
E\left(\sqrt{n dL2^{n-1}}\right)^{-1/(d(2l-1))} 2^{-1/(2d(2l-1)(n+1))\|t\|}.
\]

Hence the sum of measures of sets (8.1) over all \( t \in \mathbb{Z}^n_+ \) is finite, which implies that almost every \( x \in B \) is contained in at most finitely many such sets. To finish the proof, it remains to observe that parallelepipeds \( \{2^{t_1} \leq |q_i| < 2^{t_i+1}\} \) cover all the integer vectors \( q \) with each of the coordinates different from zero.

\[\blacksquare\]
8.2

Here is one more example of functions $\Psi$ one can consider. For an $n$-tuple $s = (s_1, \ldots, s_n)$ with positive components, define the $s$-quasinorm $\| \cdot \|_s$ on $\mathbb{R}^n$ by $\|x\|_s \overset{\text{def}}{=} \max_{1 \leq i \leq n} |x_i|^{1/s_i}$. Then, following [Kl], we say that $y \in \mathbb{R}^n$ is $s$-$\psi$-approximable if it belongs to $W(\Psi)$, where

$$\Psi(q) = \psi(\|q\|_s).$$

We normalize $s$ so that $\sum_i s_i = 1$. (This way, e.g., one can see that, for a nonincreasing $\psi$, any $s$-$\psi$-approximable $y$ is $\psi$-MA.) The choice $s = (1/n, \ldots, 1/n)$ gives the standard definition of $\psi$-approximability. One can also show that (1.1) holds if and only if $\sum_{k=1}^{\infty} \psi(k) < \infty$. Thus one has the following generalization of Corollary 1.2 (S).

**Corollary.** Let $f : U \to \mathbb{R}^n$ be as in Theorem 1.1, let $\psi : \mathbb{N} \to (0, \infty)$ be a nonincreasing function, and take any $s = (s_1, \ldots, s_n)$ with $s_i > 0$ and $\sum_i s_i = 1$. Then, assuming (1.2s), for almost all $x \in U$ the points $f(x)$ are not $s$-$\psi$-approximable.

□

8.3

The idea of studying the set of points $x$ such that $F(x) \overset{\text{def}}{=} f(x)q$ is close to an integer by looking at the values of the gradient $\nabla F(x) = f(x)q$ of $F$ has a long history. It was extensively used by Sprindžuk [Sp2], [Sp4] in his proof of Mahler’s conjecture, that is, when $d = 1$ and $f_i(x) = x^i$. Also, from a paper of Baker and Schmidt [BS] it follows that, for some $\gamma, \varepsilon > 0$, on a set of positive measure, the system

$$\begin{aligned}
|P(x)| &< \|q\|^{-n+\gamma}, \\
|P'(x)| &< \|q\|^{-1-\gamma-\varepsilon}
\end{aligned} \quad (8.2)$$

has at most finitely many solutions $p \in \mathbb{Z}$, $q \in \mathbb{Z}^n$. (Here $P(x)$ is the polynomial $p + q_1 x + \cdots + q_n x^n$.) This was used to construct a certain regular system of real numbers and to obtain the sharp lower estimate for the Hausdorff dimension of the set

$$\left\{ x \in \mathbb{R} \mid \langle q_1 x + \cdots + q_n x^n \rangle < \|q\|^{-\lambda} \text{ for infinitely many } q \in \mathbb{Z}^n \right\}$$

for $\lambda > n$. Also, note that system (8.2) is related to the distribution of values of discriminants of integer polynomials (see [D], [Sp2], [Bern]).

In 1995 Borbat [Bo] proved that, given any $\varepsilon > 0$ and $0 < \gamma < 1$, for almost all $x$ there are at most finitely many solutions of (8.2). Now we can use Theorem 1.4 to relax the restriction $\gamma < 1$. More precisely, we derive the following generalization and strengthening of the aforementioned result of Borbat.
Theorem. Let $U \subset \mathbb{R}^d$ be an open subset, and let $f = (f_1, \ldots, f_n)$ be a nondegenerate $n$-tuple of $C^m$ functions on $U$. Take $\varepsilon > 0$ and $0 < \gamma < n$. Then for almost all $x \in V$ there exist at most finitely many solutions $q \in \mathbb{Z}^n$ of the system

$$
\begin{align*}
\| f(x) q \| &< \Pi_+(q)^{-1+\gamma/n}, \\
\| \nabla f(x) q \| &< \| q \|^{1-\gamma-\varepsilon}.
\end{align*}
$$

(8.3)

□

Proof. As in the proof of Theorem 1.1, one can use induction to be left with integer vectors $q$ with all coordinates different from zero. Then one estimates the measure of the union of the sets of solutions of (8.3) over all $q \in \mathbb{Z}^n$ with $2^{t_1} \leq |q_1| < 2^{t_1+1}$ for any $n$-tuple $t = (t_1, \ldots, t_n) \in \mathbb{Z}_+^n$ by using Theorem 1.4 with

$$\delta = 2^{(-1+\gamma/n) \Sigma_i t_i}, \ K = \sqrt{n} d \Pi_+ 2^{-(1-\gamma-\varepsilon) t_1}, \ T_i = 2^{t_1+1}.
$$

Inequalities (1.7a) are clearly satisfied; in particular, one has

$$\frac{\delta K T_1 \cdots T_n}{\max_i T_i} = 2^{\gamma/n(\Sigma_i t_i - n \| t \|)} 2^{-\varepsilon \| t \|} \leq 2^{-\varepsilon \| t \|} \leq 1.
$$

Therefore, by Theorem 1.4, the measure of the above union is at most

$$E \max \left( 2^{(-1+\gamma/n)/\Pi_+(\Sigma_i \max_j t_j)} 2^{-\varepsilon \Pi_+(\Sigma_i t_i)} \right).
$$

Obviously, the sums of both functions in the right-hand side over all $t \in \mathbb{Z}_+^n$ are finite, which completes the proof. ■

The above theorem naturally invites one to think about the possibility of Khintchine-type results involving derivative estimates, that is, replacing (8.3) by, say,

$$
\begin{align*}
\| f(x) q \| &< \Psi_1(q), \\
\| \nabla f(x) q \| &< \Psi_2(q)
\end{align*}
$$

(8.4)

and finding optimal conditions on $\Psi_1$ implying at most finitely many solutions of (8.4) for almost all $x$.

8.4

The main result of this paper (Theorem 1.1) was proved in the summer of 1998 but only when the functions $f_1, \ldots, f_n$ are analytic. More precisely, the analytic setup was reduced to the case $d = 1$ (see [Sp4, Section 3] or [P] for a related “slicing” technique). In the latter case, in addition to the nondegeneracy of $f$, we had to assume that there exist positive constants $C$ and $\alpha$ such that for almost all $x \in U \subset \mathbb{R}$ one can find a subinterval $B$ of $U$ containing $x$ such that
Span \( \left( f_1', \ldots, f_n', \frac{f_i}{f'_i}, \frac{f_j}{f'_j} \right)_{1 \leq i < j \leq n} \) consists of functions \((C, \alpha)\)-good on \(B\).

For analytic functions \(f_1, \ldots, f_n\), this condition can be easily verified by applying Corollary 3.5(a) to the basis of the above function space.

In our original approach for \(d = 1\), we considered sets more general than (1.7b), namely, the sets

\[
\left\{ x \in B \mid \exists \, q \in \mathbb{Z}^n \setminus \{0\} \text{ such that } \begin{cases} |f(x)q| < \delta, \\ |g(x)q| < K, \\ |q_i| < T_i, & i = 1, \ldots, n \end{cases} \right\}
\]

with \(f\) as in Theorem 1.4 and with \(g\) another nondegenerate \(n\)-tuple of functions on \(U\).

Instead of \(U\), as in (5.2), we considered more general matrices

\[
U_{f,g}^x := \begin{pmatrix} 1 & 0 & f(x) \\ 0 & 1 & g(x) \\ 0 & 0 & I_n \end{pmatrix}.
\]

To prove an analogue of Theorem 6.2(i), or, more precisely, to prove the statement that for some positive \(C, \alpha\) and any subgroup \(\Gamma\) of \(\Lambda\) the function \(x \mapsto \|DU_{f,g}^x\|\) is \((C, \alpha)\)-good on some neighborhood \(B\) of \(x_0\), it was enough to consider the standard basis \(\{e_0, e_*e_1, \ldots, e_n\}\) of \(\mathbb{R}^{n+1}\) and the corresponding basis

\[
\{ e_I := e_{i_1} \wedge \cdots \wedge e_{i_k} \mid I = \{i_1, \ldots, i_k\} \subset \{0, *, 1, \ldots, n\} \}
\]

of \(\wedge^k(\mathbb{R}^{n+1})\), to decompose an element \(w\) representing \(\Gamma\) as \(w = \sum w_I e_I\), to write an expansion similar to (7.3), and to use (8.6).

To prove an analogue of Theorem 6.2(ii), that is, the statement that for any neighborhood \(B\) of \(x_0\) there exists \(\rho > 0\) such that \(\sup_{x \in B} \|DU_{f,g}^x\| \geq \rho\) for every \(\Gamma \subset \Lambda\), we used the fact that the coefficients \(w_I\) are integers and considered the following two cases:

1. \(w_I \neq 0\) for some \(I \subset \{1, \ldots, n\}\) and
2. \(w_I = 0\) for all \(I \subset \{1, \ldots, n\}\) in this approach.
it was important that $K \geq 1$. This was enough for the proof of the case $d = 1$ in Theorem 1.1; however, as we saw in Section 8.3, the stronger version, allowing arbitrarily small positive values of $K$, is important for other applications.

### 8.5

For completeness let us discuss the complementary divergence case of Khintchine-type theorems mentioned in this paper. It was proved by Khintchine [Kh] in 1924 (resp., by Groshev [G] in 1938) that almost everywhere $y \in \mathbb{R}$ (resp., $\mathbb{R}^n$) is $\psi$-approximable whenever $\psi$ is a nonincreasing function that does not satisfy (1.1s). In 1960 Schmidt [S1] showed that almost everywhere $y \in \mathbb{R}^n$ belongs to $W(\psi)$ whenever the series in (1.1) diverges. (Note that there are no monotonicity restrictions on $\psi$ unless $n = 1$.) It seems plausible to conjecture the divergence counterpart of Theorem 1.1, namely, that, for $f : U \rightarrow \mathbb{R}^n$ as in Theorem 1.1 and $\Psi$ satisfying (1.5) but not (1.1), the set $\{x \in U \mid f(x) \in W(\Psi)\}$ has full measure. For functions $\Psi$ of the form (1.2s), this can be done using Theorem 1.4 and the method of regular systems, which dates back to [BS] and has been extensively used in the existing proofs of divergence Khintchine-type results for special classes of manifolds (see [DRV2], [DRV3], [BBDD], [Be1], [Be3], [Be4]).

### Acknowledgments

A substantial part of this work was done during the authors’ stays at the University of Bielefeld in 1998 and 1999. These stays were supported by Sonderforschungsbereich-343 and the Humboldt Foundation. The paper was completed during the Spring 2000 programme on Ergodic Theory, Geometric Rigidity and Number Theory at the Isaac Newton Institute for Mathematical Sciences. Thanks are also due to V. Beresnevich and M. Dodson for useful remarks. D. Kleinbock’s work was supported in part by National Science Foundation grant number DMS-9704489, and G. A. Margulis’s work was supported by National Science Foundation grant number DMS-9800607.

### References

Khintchine-Type Theorems on Manifolds


Bernik, Kleinbock, and Margulis


Bernik: Institute of Mathematics, Belarus Academy of Sciences, Minsk 220072, Belarus; bernik@im.bas-net.by

Kleinbock: Department of Mathematics, Rutgers University, New Brunswick, New Jersey 08903, USA; kleinboe@brandeis.edu

Margulis: Department of Mathematics, Yale University, New Haven, Connecticut 06520, USA; margulis@math.yale.edu