# KHINTCHINE-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 Khintchine-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 $\mathbf{y} = (y_1, \dots, y_n)$ of real numbers viewed as *linear forms*, i.e. as row vectors. In what follows, $\mathbf{y}$ will always mean a row vector, and we will be interested in values of a linear form given by $\mathbf{y}$ at integer points $\mathbf{q} = (q_1, \dots, q_n)^T$, the latter being a column vector. Thus $\mathbf{yq}$ will stand for $y_1q_1 + \cdots + y_nq_n$. Hopefully it will cause no confusion.

We will study differentiable maps $\mathbf{f} = (f_1, \dots, f_n)$ from open subsets $U$ of $\mathbb{R}^d$ to $\mathbb{R}^n$; again, $\mathbf{f}$ will be interpreted as a row vector, so that $\mathbf{f}(x)\mathbf{q}$ stands for $q_1f_1(x) + \cdots + q_nf_n(x)$. In contrast, the elements of the "parameter set" $U$ will be denoted by $x = (x_1, \dots, x_d)$ without boldfacing, since the linear structure of the parameter space is not significant.

For $\mathbf{f}$ as above we will denote by $\partial_i\mathbf{f} : U \mapsto \mathbb{R}^n$, $i = 1, \dots, d$, its partial derivative (also a row vector) with respect to $x_i$. If $F$ is a scalar function on $U$, we will denote by $\nabla F$ the column vector consisting of partial derivatives of $F$. With some abuse of notation, the same way we will treat vector functions $\mathbf{f}$: namely, $\nabla\mathbf{f}$ will stand for the matrix function $U \mapsto M_{d\times n}(\mathbb{R})$ with rows given by partial derivatives $\partial_i\mathbf{f}$. We will also need higher order differentiation: for a *multiindex* $\beta = (i_1, \dots, i_d)$, $i_j \in \mathbb{Z}_+$, we let $|\beta| = \sum_{j=1}^d i_j$ and $\partial_\beta = \partial_1^{i_1} \circ \cdots \circ \partial_d^{i_d}$.

Unless otherwise indicated, the norm $\|\mathbf{x}\|$ of a vector $\mathbf{x} \in \mathbb{R}^k$ (either row or column vector) will stand for $\|\mathbf{x}\| = \max_{1\le i\le k} |x_i|$. In some cases however we will work with the Euclidean norm $\|\mathbf{x}\| = \|\mathbf{x}\|_e = \sqrt{\sum_{i=1}^k x_i^2}$, keeping the same notation. This distinction will be clearly emphasized to avoid confusion. We will denote by $\mathbb{R}_1^k$ the set of unit vectors in $\mathbb{R}^k$ (with respect to the Euclidean norm).

We will use the notation $|\langle x\rangle|$ for the distance between $x \in \mathbb{R}$ and the closest integer, $|\langle x\rangle| \stackrel{\text{def}}{=} \min_{k\in\mathbb{Z}} |x - k|$. (It is quite customary to use $\|x\|$ instead, but we are not going to 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 $B$.

The work of the second named author was supported in part by NSF Grant DMS-9704489, and that of the third named author by NSF Grant DMS-9800607.

**Basics on Diophantine approximation.** In what follows, we let $\Psi$ be a positive function defined on $\mathbb{Z}^n \smallsetminus \{0\}$, and consider the set

$$\mathcal{W}(\Psi) \stackrel{\text{def}}{=} \left\{\mathbf{y} \in \mathbb{R}^n \;\middle|\; |\langle \mathbf{y}\mathbf{q}\rangle| \le \Psi(\mathbf{q}) \text{ for infinitely many } \mathbf{q}\right\}.$$

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

$$\sum_{\mathbf{q}\in\mathbb{Z}^n \smallsetminus \{0\}} \Psi(\mathbf{q}) < \infty\,. \tag{1.1}$$

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

The standard class of examples is given by functions which depend only on the norm of $\mathbf{q}$. If $\psi$ is a positive function defined on positive integers, let us say, following [KM2], that $\mathbf{y}$ is *$\psi$-approximable*, to be abbreviated as *$\psi$-A*, if it belongs to $\mathcal{W}(\Psi)$ where[1]

$$\Psi(\mathbf{q}) = \psi(\|\mathbf{q}\|^n)\,. \tag{1.2s}$$

If $\psi$ is non-increasing (which will be our standing assumption), (1.1) is satisfied if and only if

$$\sum_{k=1}^{\infty} \psi(k) < \infty\,. \tag{1.1s}$$

An example: almost all $\mathbf{y} \in \mathbb{R}^n$ are not VWA (*very well approximable*, see [S2, Chapter IV, §5]); the latter is defined to be $\psi_\varepsilon$-approximable for some positive $\varepsilon$, with $\psi_\varepsilon(k) \stackrel{\text{def}}{=} k^{-(1+\varepsilon)}$.

Another important special case is given by

$$\Psi(\mathbf{q}) = \psi\big(\Pi_+(\mathbf{q})\big)\,. \tag{1.2m}$$

where $\Pi_+(\mathbf{q})$ is defined as $\prod_{i=1}^n \max(|q_i|, 1)$, i.e. the absolute value of the product of all the nonzero coordinates of $\mathbf{q}$. We will say that $\mathbf{y} \in \mathbb{R}^n$ is *$\psi$-MA* (*$\psi$-multiplicatively approximable*) 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

$$\sum_{k=1}^{\infty} (\log k)^{n-1} \psi(k) < \infty\,. \tag{1.1m}$$

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

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

**Diophantine approximation on manifolds.** Much more intricate questions arise if one restricts $\mathbf{y}$ to lie on a submanifold $M$ of $\mathbb{R}^n$. In 1932 K. Mahler [M] conjectured that almost all points on the curve

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

are not VWA. V. Sprindžuk's proof of this conjecture (see [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 non-increasing function such that

$$\sum_{k=1}^{\infty} \frac{\psi(k)^{1/n}}{k^{1-1/n}} < \infty\,, \tag{1.4}$$

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 A. Baker in his book [B2] that almost all points on the curve (1.3) are not VWMA; the validity of this conjecture for $n \le 4$ was verified in 1997 by V. Bernik and V. Borbat [BB].

Since the mid-sixties, a lot of efforts have been directed to 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 in the paper [KM1] by Kleinbock and Margulis. Let us employ the following definition: if $U$ is an open subset of $\mathbb{R}^d$ and $l \le m \in \mathbb{N}$, say that an $n$-tuple $\mathbf{f} = (f_1, \dots, f_n)$ of $C^m$ functions $U \mapsto \mathbb{R}$ is *$l$-nondegenerate at $x \in U$* if the space $\mathbb{R}^n$ is spanned by partial derivatives of $\mathbf{f}$ at $x$ of order up to $l$. We will say that $\mathbf{f}$ 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 will say that $M$ is *nondegenerate at* $\mathbf{y} \in M$ if any (equivalently, some) diffeomorphism $\mathbf{f}$ between an open subset $U$ of $\mathbb{R}^d$ and a neighborhood of $\mathbf{y}$ in $M$ is nondegenerate at $\mathbf{f}^{-1}(\mathbf{y})$. We will say that $\mathbf{f} : U \to \mathbb{R}^n$ (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).*

In particular, the aforementioned multiplicative conjecture of Baker follows from this theorem. Note also that if the functions $f_1, \dots, f_n$ are analytic and $U$ is connected, the nondegeneracy of $\mathbf{f}$ is equivalent to the linear independence of $1, f_1, \dots, f_n$ over $\mathbb{R}$; in this setting the above statement was conjectured by Sprindžuk [Sp4, Conjectures $H_1$, $H_2$].

**Main results and the structure of the paper.** The primary goal of the present paper is to obtain a Khintchine-type generalization of Theorem A. More precisely, we prove the following

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

$$\Psi(q_1, \dots, q_i, \dots, q_n) \ge \Psi(q_1, \dots, q_i', \dots, q_n) \quad \textit{whenever} \quad |q_i| \le |q_i'| \textit{ and } q_i q_i' > 0 \tag{1.5}$$

*(i.e., $\Psi$ is non-increasing with respect to the absolute value of any coordinate in any orthant of $\mathbb{R}^n$). Then* $|\{x \in U \mid \mathbf{f}(x) \in \mathcal{W}(\Psi)\}| = 0$.

In particular, if $\Psi$ is of the form (1.2s) or (1.2m) for a non-increasing function $\psi : \mathbb{N} \mapsto \mathbb{R}_+$, condition (1.5) is clearly satisfied. Thus one has

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

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

It is worth mentioning that the statement (S) was recently proved in a paper [Be5] of V. Beresnevich using a refinement of Sprindžuk's method of "essential and inessential domains". Earlier several special cases were treated in [DRV1, BDD, Be2]. A preliminary version [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 look at the set of all $x \in B$ for which there exists an integer vector $\mathbf{q}$ in a certain range such that the value of the function $F(x) = \mathbf{f}(x)\mathbf{q}$ is close to an integer. Our estimates will require considering two special cases: when the norm of the gradient $\nabla F(x) = \nabla \mathbf{f}(x)\mathbf{q}$ is big, or respectively, not very big. We will show in §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 $\tilde{B}$ stand for the ball with the same center as $B$ and of radius $2r$, and let functions $\mathbf{f} = (f_1, \dots, f_n) \in C^2(\tilde{B})$ be given. Fix $\delta > 0$ and define*

(1.6a)
$$L = \max_{|\beta|=2,\, x \in \tilde{B}} \|\partial_\beta \mathbf{f}(x)\| .$$

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

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

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

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

*and*

(1.6d)
$$\|\nabla \mathbf{f}(x)\mathbf{q}\| \geq \sqrt{ndL\|\mathbf{q}\|}$$

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

**Theorem 1.4.** *Let $U \subset \mathbb{R}^d$ be an open set, $x_0 \in U$, and let $\mathbf{f} = (f_1, \dots, f_n)$ 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 exist $E > 0$ such that for any choice of*

$$0 < \delta \le 1, \quad T_1, \dots, T_n \ge 1 \quad \text{and} \quad K > 0 \text{ with } \frac{\delta K T_1 \cdot \ldots \cdot T_n}{\max_i T_i} \le 1\,, \tag{1.7a}$$

*the set*

$$\left\{ x \in B \;\middle|\; \exists\, \mathbf{q} \in \mathbb{Z}^n \smallsetminus \{0\} \text{ such that } \begin{cases} |\langle \mathbf{f}(x)\mathbf{q}\rangle| < \delta \\ \|\nabla \mathbf{f}(x)\mathbf{q}\| < K \\ |q_i| < T_i, \quad i = 1, \dots, n \end{cases} \right\} \tag{1.7b}$$

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

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

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

As for Theorem 1.4, it is done by a modification of a method from [KM1] involving the geometry of lattices in Euclidean spaces. The connection with lattices is discussed in §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 §3, where, as well as in §4, 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 §7, after doing some preparatory work in the preceding section. The last section of the paper is devoted to the reduction of 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 approximation of zero by values of functions and their derivatives.

## 2. Proof of Theorem 1.3

We first state a covering result which is well known (and is one of the ingredients of the main estimate from [KM1]).

**Theorem 2.1 (Besicovitch's Covering Theorem**, see [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 $1_S \le \sum_i 1_{U_i} \le N_d$ (i.e., $S \subset \bigcup_i U_i$ and the multiplicity of that subcovering is at most $N_d$).*

This theorem, and the constant $N_d$, will be used repeatedly in the paper.

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

$$M \geq 1/4r^2 \tag{2.1a}$$

*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*

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

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

$$|\langle F(x)\rangle| < \delta \tag{2.1c}$$

*and*

$$\|\nabla F(x)\| \geq \sqrt{d M} \tag{2.1d}$$

*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 $\frac{1}{2|\partial_1 F(x)|}$ by $\rho$; note that $\rho\sqrt{d} \leq \frac{1}{2\sqrt{M}} \leq r$ 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 bigger than $\rho$. Indeed, one has

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

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

$$|F(x_1 \pm \rho, x_2, \ldots, x_d) + p| \underset{(2.1\text{bc})}{\geq} -\delta + 1/2 - M\rho^2/2 \underset{(2.1\text{d})}{\geq} 1/2 - \tfrac{1}{8\sqrt{d}} - \delta \underset{(\delta<1/16)}{>} 1/4\,,$$

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 $y \in B(x, \frac{\rho}{4\sqrt{d}})$, one has

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

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 $\left|\frac{\partial F}{\partial u}(x)\right| \leq \sqrt{d}|\partial_1 F(x)| = \frac{\sqrt{d}}{2\rho}$, and from (2.1b) that $\left|\frac{\partial^2 F}{\partial u^2}(z)\right| \leq dM \leq \frac{d}{4\rho^2}$. Therefore $|F(y) + p| \leq \delta + 1/8 + 1/128 < 1/4$, which shows that $U(x) \subset B(x, \frac{\rho}{4\sqrt{d}})$, and, in particular,

$$|U(x)| \geq C_d' \rho^d \tag{2.2}$$

(the values of constants $C'_d$, and also of $C''_d$, $C'''_d$ to be 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

$$|\partial_1 F(y) - \partial_1 F(x)| \le |\tfrac{\partial}{\partial u}\partial_1 F(z)|\rho\sqrt{d} \underset{(2.1b)}{\le} M\rho\sqrt{d} \underset{(2.1d)}{\le} 1/4\rho = \tfrac{|\partial_1 F(x)|}{2}\,,$$

(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 $\frac{1}{2}|\partial_1 F(x)|$; in particular, for every $y_2, \dots, y_d$ such that $|y_i - x_i| < \rho$, $i = 2, \dots, d$, the function $F(\cdot, y_2, \dots, y_d)$ is monotonic on $(x_1 - \rho, x_1 + \rho)$, and therefore

$$\left|\{y_1 \in (x_1 - \rho, x_1 + \rho) \mid |F(y_1, \dots, y_d) + p| < \delta\}\right| \le 2\delta\tfrac{2}{|\partial_1 F(x)|} = 8\rho\delta\,.$$

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

$$\left|\{x \in \bar{U}(x) \mid |F(x) + p| < \delta\}\right| \le 8\rho\delta 2^{d-1}\rho^{d-1} = C''_d \delta\rho^d \underset{(2.2)}{\le} C'''_d |U(x)|\,.$$

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

$$|S| \le \sum_i C'''_d \delta |U_i| \le C'''_d N_d \delta |\tilde{B}| = C_d \delta |B|\,,$$

which finishes the proof. □

Now it takes very little to complete the

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

## 3. $(C, \alpha)$-GOOD FUNCTIONS

Let us recall the definition introduced in [KM1]. If $C$ and $\alpha$ are positive numbers and $V$ 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 $\varepsilon > 0$ one has

$$\left|\{x \in B \mid |f(x)| < \varepsilon \cdot \sup_{x \in B} |f(x)|\}\right| \le C\varepsilon^\alpha |B|\,. \tag{3.1}$$

Several elementary facts about $(C, \alpha)$-good functions are listed below:

**Lemma 3.1.** (a) *$f$ is $(C, \alpha)$-good on $V$ ⇒ so is $\lambda f$ $\forall \lambda \in \mathbb{R}$;*

(b) *$f_i$, $i \in I$, are $(C, \alpha)$-good on $V$ ⇒ so is* $\sup_{i \in I} |f_i|$;

(c) *If $f$ is $(C, \alpha)$-good on $V$ and $c_1 \le \frac{|f(x)|}{|g(x)|} \le c_2$ for all $x \in V$, then $g$ is $\big(C(c_2/c_1)^\alpha, \alpha\big)$-good on $V$;*

(d) *$f$ is $(C, \alpha)$-good on $V$ ⇒ it is $(C', \alpha')$-good on $V'$ for every $C' \ge C$, $\alpha' \le \alpha$ and $V' \subset V$.*

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

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

*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

$$\left|\left\{x \in B \;\middle|\; |f(x)| < \varepsilon \cdot \sup_{x \in B} |f(x)|\right\}\right| \le 2dl(l+1)^{1/l} \varepsilon^{1/dl} |B|,$$

and the claim follows by circumscribing a cube around any ball in $\mathbb{R}^d$. □

The next 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*

$$|\partial_\beta f(x)| \le A_1 \quad \forall\, \beta \text{ with } |\beta| \le k, \tag{3.2$\le$}$$

*and*

$$|\partial_i^k f(x)| \ge A_2 \quad \forall\, i = 1, \dots, d \tag{3.2$\ge$}$$

*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 $(C, \frac{1}{dl})$-good on $V$, where*
$C = \frac{2^d}{v_d} dk(k+1) \left(\frac{A_1}{A_2}(k+1)(2k^k+1)\right)^{1/k}$.

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

**Proposition 3.4.** *Let $U$ be an open subset of $\mathbb{R}^d$, and let $\mathcal{F} \subset C^l(U)$ be a family of functions $f : U \mapsto \mathbb{R}$ such that*

$$\{\nabla f \mid f \in \mathcal{F}\} \text{ is compact in } C^{l-1}(U). \tag{3.3}$$

*Assume also that*

$$\inf_{f \in \mathcal{F}} \sup_{|\beta| \le l} |\partial_\beta f(x_0)| > 0 \tag{3.4}$$

*(in other words, the derivatives of $f$ at $x_0$ uniformly (in $f \in \mathcal{F}$) generate $\mathbb{R}$). Then there exists a neighborhood $V \subset U$ of $x_0$ and a positive $C = C(\mathcal{F})$ such that the following holds for all $f \in \mathcal{F}$:*

(a) *$f$ is $(C, \frac{1}{dl})$-good on $V$;*

(b) *$\|\nabla f\|$ is $(C, \frac{1}{d(l-1)})$-good on $V$.*

*Proof.* Assumption (3.4) says that there exists a constant $C_1 > 0$ such that for any $f \in \mathcal{F}$ one can find a multiindex $\beta$ with $|\beta| = k \le l$ with

$$\left|\partial_\beta f(x_0)\right| \ge C_1.$$

By an appropriate rotation of the coordinate system one can guarantee that $|\partial_i^k f(x_0)| \ge C_2$ for all $i = 1, \dots, d$ 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 (3.2) hold for all $x \in V'$ (note that in the above inequalities both $k \in \{1, \dots, l\}$ 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 $\hat{B}$ circumscribed around $B$ is contained in $V'$.

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

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

$$\left|\{x \in B \mid \|\nabla f(x)\| < \varepsilon \cdot \sup_{x \in B} \|\nabla f(x)\|\}\right| \le \left(\frac{A_1}{A_2}\right)^{\frac{1}{n-1}} \varepsilon^{\frac{1}{n-1}} |B| . \qquad \square$$

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

(a) *any linear combination of $1, f_1, \dots, f_n$ is $(C, \frac{1}{dl})$-good on $V$;*

(b) *the norm of any linear combination of $\nabla f_1, \dots, \nabla f_n$ is $(C, \frac{1}{n-1})$-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 $(c_1, \dots, c_n)$ is equal to 1. All such functions $f$ belong to a family satisfying (3.3) and (due to the nondegeneracy of $\mathbf{f}$ at $x_0$) (3.4), thus the above proposition applies. $\square$

We close the section with two auxiliary lemmas which will be 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 $r$, and let $f \in C^l(B)$ and $c > 0$ be such that for some unit vector $u$ in $\mathbb{R}^d$, some $k \le l$ and all $x \in B$ one has $\left|\frac{\partial^k f}{\partial u^k}(x)\right| \ge c$. Then*

$$\sup_{x,y \in B} |f(x) - f(y)| \ge \frac{c}{k^k (k+1)!} (2r)^k . \tag{3.5}$$

*Proof.* If $x_0$ is the center of $B$, consider the function $g(t) = f(x_0 + tu)$ defined on $I \stackrel{\text{def}}{=} [-r, r]$. Denote $\sup_{s,t \in I} |f(s) - f(t)|$ by $\sigma$. We claim that

$$\sigma \ge \frac{a}{k^k (k+1)!} (2r)^k ; \tag{3.6}$$

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 $|p^{(k)}(t)| \leq (k+1)\frac{\sigma k!}{(2r/k)^k}$. Combining the last two inequalities, one easily gets the desired estimate.

**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 twice bigger 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 $B \subset \tilde{V}$ and any $\varepsilon \geq \delta$. 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 $\varepsilon > 0$ and a ball $B \subset V$, and denote

$$S_{B,\varepsilon} \stackrel{\text{def}}{=} \{x \in \tilde{V} \mid |f(x)| < \varepsilon \cdot \sup_{x \in B} |f(x)|\} .$$

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

Obviously it suffices to consider $\varepsilon < 1$. Choose $m \in \mathbb{Z}_+$ such that $\delta^{m+1} \leq \varepsilon < \delta^m$. We will show by induction on $m$ that

$$|B \cap S_{B,\varepsilon}| \leq C^{m+1} N_d^m \varepsilon^\alpha |B| . \tag{3.7}$$

Indeed, the case $m = 0$ follows from the assumption. Assume that (3.7) holds for some $m$, and for every $y \in B \cap S_{B,\varepsilon}$ let $B(y)$ be the maximal ball centered in $y$ and contained in $S_{B,\varepsilon}$. Observe that, by the continuity of $f$, one has $\sup_{x \in B(y)} |f(x)| = \varepsilon \cdot \sup_{x \in B} |f(x)|$ for every $y \in B$. Clearly the set $B \cap S_{B,\varepsilon}$ 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

$$\begin{aligned} |B \cap S_{B,\delta\varepsilon}| &\leq \sum_i \left|\{x \in B_i \mid |f(x)| < \delta \sup_{x \in B_i} |f(x)|\}\right| \leq \sum_i C\delta^\alpha |B_i| \\ &= C\delta^\alpha N_d |B \cap S_{B,\varepsilon}| \leq CN_d \cdot C^{m+1} N_d^m (\delta\varepsilon)^\alpha |B| , \end{aligned}$$

which proves (3.7) with $\delta\varepsilon$ in place of $\varepsilon$.

It remains to write

$$C^{m+1} N_d^m \varepsilon^\alpha = C \cdot (CN_d)^m (\varepsilon)^{\alpha-\alpha'} \varepsilon^{\alpha'} < C \cdot (CN_d)^m (\delta^m)^{\alpha-\alpha'} \varepsilon^{\alpha'} = C \cdot (CN_d \delta^{\alpha-\alpha'})^m \varepsilon^{\alpha'} ,$$

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 will define and study the following construction. The main object will be 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 $\mathbf{g} : V \mapsto \mathbb{R}^2$. For such a pair, let us define its *skew-gradient* $\tilde{\nabla}\mathbf{g} : V \mapsto \mathbb{R}^d$ by

$$\tilde{\nabla}\mathbf{g}(x) \stackrel{\text{def}}{=} g_1(x)\nabla g_2(x) - g_2(x)\nabla g_1(x) .$$

Equivalently, the $i$th component of $\tilde{\nabla}\mathbf{g}$ at $x$ is equal to $\begin{vmatrix} g_1(x) & g_2(x) \\ \partial_i g_1(x) & \partial_i g_2(x) \end{vmatrix}$, that is, to the signed area of the parallelogram spanned by $\mathbf{g}(x)$ and $\partial_i \mathbf{g}(x)$. Another interpretation: if one represents $\mathbf{g}(x)$ in polar coordinates, i.e. via functions $\rho(x)$ and $\theta(x)$, it is straightforward to verify that $\tilde{\nabla}\mathbf{g}(x)$ can be written as $\rho^2(x)\nabla\theta(x)$.

Loosely speaking, the skew-gradient measures how different are the two functions from being proportional to each other: it is easy to see that $\tilde{\nabla}\mathbf{g}$ is identically equal to zero on an open set iff $g_1$ and $g_2$ are proportional (with a locally constant coefficient). Therefore if the image $\mathbf{g}(V) \subset \mathbb{R}^2$ does not look like a part of a straight line passing through the origin, one should expect the values of $\tilde{\nabla}\mathbf{g}$ to be not very small. Moreover, if the map $\mathbf{g}$ is polynomial of degree $\le k$, then $\tilde{\nabla}\mathbf{g}$ is a polynomial map of degree $\le 2k-2$; in particular, its norm is $(C,\alpha)$-good for some $C,\alpha$. The results of the previous section suggest that the latter property should be shared by maps which are "close to polynomial" in the sense of Lemma 3.4 (that is, for families of functions with some uniformly non-vanishing partial derivatives).

The goal of the 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 $\mathcal{G} \subset C^l(U)$ be a family of maps $\mathbf{g} : U \mapsto \mathbb{R}^2$ such that*

$$\text{(4.1)} \qquad \textit{the family } \{\nabla g_i \mid \mathbf{g} = (g_1, g_2) \in \mathcal{G},\ i = 1,2\} \textit{ is compact in } C^{l-1}(U)\,.$$

*Assume also that*

$$\text{(4.2)} \qquad \inf_{\substack{\mathbf{g}\in\mathcal{G} \\ \mathbf{v}\in\mathbb{R}^2_1}} \sup_{|\beta|\le l} |\mathbf{v}\cdot\partial_\beta \mathbf{g}(x_0)| > 0$$

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

(a) *$\|\tilde{\nabla}\mathbf{g}\|$ is $(2C_{d,l}, \frac{1}{d(2l-1)})$-good on $V$ for every $\mathbf{g}\in\mathcal{G}$ (here $C_{d,l}$ is as in Lemma 3.2);*

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

$$\sup_{x\in B} \|\tilde{\nabla}\mathbf{g}(x)\| \ge \rho \qquad \textit{for every } \mathbf{g}\in\mathcal{G}\,.$$

To prove this proposition, we will use two lemmas below. Note that in this section for convenience we will switch to the Euclidean norm $\|\mathbf{x}\| = \|\mathbf{x}\|_e$.

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

$$\sup_{x\in B} \|\tilde{\nabla}\mathbf{g}(x)\| \ge \frac{w(a-\delta)^2}{2r\sqrt{w^2 + (a+\delta)^2}}\,.$$

*Proof.* Let us use polar coordinates, choosing $\mathcal{L}$ to be the polar axis. Take $x_1 \in \overline{B}$ such that $\operatorname{dist}\big(\mathbf{g}(x), \mathcal{L}\big) = w$; then one has

$$\theta(x_1) \ge \sin\theta(x_1) = \frac{\operatorname{dist}\big(\mathbf{g}(x), \mathcal{L}\big)}{\|\mathbf{g}(x_1)\|} \ge \frac{w}{\sqrt{w^2 + (a+\delta)^2}}\,.$$

Denote by $J$ the straight line segment $[x_0, x_1] \subset B$, and by $u$ the unit vector proportional to $x_1 - x_0$. Restricting $\mathbf{g}$ to $J$ and using Lagrange's Theorem, one can find $y$ between $x_0$ and $x_1$ such that $\theta(x_1) = \frac{\partial\theta}{\partial u}(y)|J|$. Then one has $|u \cdot \tilde{\nabla}\mathbf{g}(y)| = \rho^2(y)|\frac{\partial\theta}{\partial u}(y)| \geq (a-\delta)^2 \frac{\theta(x_1)}{|J|}$ which completes the proof. $\square$

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

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

*(the diameter of the image of $\mathbf{p}$ is bounded from above), and*

$$\sup_{x\in B} \operatorname{dist}\big(\mathcal{L}, \mathbf{p}(x)\big) \geq 1/8 \text{ for any straight line } \mathcal{L} \subset \mathbb{R}^2 \tag{4.3b}$$

*(that is, the "width" of $\mathbf{p}(B)$ in any direction is bounded from below). Then:*

(a) *there exists a constant $0 < \gamma < 1$ (dependent only on $d$ and $l$) such that*

$$\sup_{x\in B} \|\tilde{\nabla}\mathbf{p}(x)\| \geq \gamma\big(1 + \sup_{x\in B} \|\mathbf{p}(x)\|\big)\,; \tag{4.4a}$$

(b) *there exists $M \geq 1$ (dependent only on $d$ and $l$) such that*

$$\sup_{x\in B,\, i=1,2} \|\nabla p_i(x)\| \leq M\,. \tag{4.4b}$$

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

Now assume that $\mathbf{p}$ satisfies (4.3ab) and $a = \|\mathbf{p}(y)\| > 6$ for some $y \in B$. Then one can apply the previous lemma to the map $\mathbf{p} : B \mapsto \mathbb{R}^2$ to get $\sup_{x\in B} \|\tilde{\nabla}\mathbf{p}(x)\| \geq \frac{\frac{1}{8}(a-2)^2}{2\sqrt{\frac{1}{64}+(a+2)^2}} \geq \frac{\frac{1}{16}(a-2)^2}{\sqrt{4(a-2)^2}} \geq \frac{1}{32}(a-2) \geq \frac{1}{64}(a+1)$, which finishes the proof of part (a). It remains to observe that part (b) trivially follows from the compactness of the set of polynomials of the form $\nabla\mathbf{p}(x)$ where $\mathbf{p}(x)$ satisfies (4.3a) and has degree $\leq l$. $\square$

*Proof of Proposition 4.1.* Choose $0 < \delta < 1/8$ such that

$$2C_{d,l}N_d\delta^{\frac{1}{d(2l-1)(2l-2)}} \leq 1\,.$$

From (4.1) and (4.2) it follows that there exists a neighborhood $V$ of $x_0$ and a positive $c$ such that for every $\mathbf{g} \in \mathcal{G}$ one has

$$\forall\, \mathbf{v} \in \mathbb{R}^2_1 \quad \exists\, u \in \mathbb{R}^d_1 \text{ and } k \leq l \text{ such that } \inf_{x\in V} \Big|\mathbf{v} \cdot \frac{\partial^k \mathbf{g}}{\partial u^k}(x)\Big| \geq c\,, \tag{4.5a}$$

and

$$\text{(4.5b)} \qquad \sup_{x,y\in V} \|\partial_\beta \mathbf{g}(x) - \partial_\beta \mathbf{g}(y)\| \le \frac{\delta c\gamma}{8Ml^l(l+1)!} \text{ for all multiindices } \beta \text{ with } |\beta| = l\,.$$

In view of Lemma 3.7, to show (a) it suffices to prove the following: given any ball $B = B(x_0, r) \subset V$ and a $C^l$ map $\mathbf{g} : B \to \mathbb{R}^2$ such that inequalities (4.5ab) hold for all $x, y \in B$, one has

$$\text{(4.6)} \qquad \left|\{x \in B \;\middle|\; \|\tilde\nabla \mathbf{g}(x)\| < \varepsilon \cdot \sup_{x\in B} \|\tilde\nabla \mathbf{g}(x)\|\}\right| \le 2C_{d,l}\varepsilon^{\frac{1}{d(2l-2)}}|B| \quad \text{whenever } \varepsilon \ge \delta\,.$$

We will do this in several steps.

**Step 1.** Note that conditions (4.5ab), as well as the function $\tilde\nabla \mathbf{g}$, will not change if one replaces $\mathbf{g}$ by $L\mathbf{g}$ 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 two most distant points of $\mathbf{g}(B)$. By (4.5a) there exist $1 \le k_1, k_2 \le n$ and $u_1, u_2 \in \mathbb{R}^d_1$ such that $|\frac{\partial^k g_i}{\partial u_i^k}(x)| \ge c$ 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

$$\text{(4.7)} \qquad s_i \ge \frac{c}{{k_i}^{k_i}(k_i+1)!}(2r)^{k_i} \ge \frac{c}{l^l(l+1)!}(2r)^l\,.$$

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

$$\text{(4.8)} \qquad \sup_{x,y\in B} \|\partial_\beta \mathbf{g}(x) - \partial_\beta \mathbf{g}(y)\| \le \delta\gamma/8M$$

for all multiindices $\beta$ with $|\beta| = l$ (here and until the end of the proof, $B$ stands for $B(0, 1)$). Note also that it follows from the construction that $\mathbf{g}(B)$ is contained in a translate of the square $[-\frac{1}{2}, \frac{1}{2}]^2$, and that $\sup_{x\in B} \operatorname{dist}\big(\mathcal{L}, \mathbf{g}(x)\big) \ge 1/2\sqrt{2}$ for any straight line $\mathcal{L} \subset \mathbb{R}^2$.

**Step 3.** Here we introduce the $l$-th degree Taylor polynomial $\mathbf{p}(x)$ of $\mathbf{g}(x)$ at 0. Using (4.8) one can show that $\mathbf{p}$ is $\frac{\delta\gamma}{8M}$-close to $\mathbf{g}$ in the $C^1$ topology, that is,

$$\sup_{x\in B} \|\mathbf{g}(x) - \mathbf{p}(x)\| \le \delta\gamma/8M \quad \text{and} \quad \sup_{x\in B} \|\nabla g_i(x) - \nabla p_i(x)\| \le \delta\gamma/8M,\ i = 1, 2\,.$$

It follows that conditions (4.3ab) are satisfied by $\mathbf{p}$, and therefore, by Lemma 4.3, the inequalities (4.4ab) hold.

**Step 4.** Now let us compare the functions $\tilde\nabla \mathbf{g}$ and $\tilde\nabla \mathbf{p}$: one has

$$\tilde\nabla \mathbf{g} - \tilde\nabla \mathbf{p} = \big(g_1(\nabla g_2 - \nabla p_2) - (g_2 - p_2)\nabla g_1\big) - \big((g_1 - p_1)\nabla p_2 - p_2(\nabla g_1 - \nabla p_1)\big)$$

therefore

$$
\begin{aligned}
\|\tilde\nabla \mathbf{g}(x) - \tilde\nabla \mathbf{p}(x)\| &\le \frac{\delta\gamma}{8M}\big(\sup_{x\in B}|g_1(x)| + \sup_{x\in B}\|\nabla g_1(x)\| + \sup_{x\in B}|p_2(x)| + \sup_{x\in B}\|\nabla p_2(x)\|\big) \\
&\le \frac{\delta\gamma}{4M}\big(\big(\sup_{x\in B}\|\mathbf{p}(x)\| + \sup_{x\in B}\|\nabla p_2(x)\|\big) + \frac{\delta\gamma}{8M}\big) \\
&\underset{(4.4\text{b})}{\le} \frac{3\delta\gamma}{8}\big(1+\sup_{x\in B}\|\mathbf{p}(x)\|\big) \underset{(4.4\text{a})}{\le} \frac{3\delta}{8}\sup_{x\in B}\|\tilde\nabla \mathbf{p}(x)\|\,.
\end{aligned}
$$

**Step 5.** Finally we are ready to prove (4.6): take $\varepsilon$ between $\delta$ and 1, put $s \overset{\text{def}}{=} \sup_{x\in B}\|\tilde\nabla \mathbf{p}(x)\|$ and observe that, in view of Step 4, the set in the left hand side of (4.6) is contained in

$$
\Big\{x\in B \;\Big|\; \|\tilde\nabla \mathbf{p}(x)\| - \tfrac{3\delta}{8}s < \varepsilon(1+\tfrac{3\delta}{8})s\Big\} = \Big\{x\in B \;\Big|\; \|\tilde\nabla \mathbf{p}(x)\| < \big(\varepsilon + \tfrac{3\delta}{8}(1+\varepsilon)\big)\,s\Big\}\,.
$$

Since $\varepsilon + \frac{3\delta}{8}(1+\varepsilon) \le \varepsilon + \frac{3\delta}{4} \le 2\varepsilon$, and since $\tilde\nabla\mathbf{p}$ 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

$$
\Big|\{x\in B \mid \|\tilde\nabla \mathbf{p}(x)\| < 2\varepsilon s\}\Big| \le C_{d,l}(2\varepsilon)^{\frac{1}{d(2l-2)}}|B| \le 2C_{d,l}\varepsilon^{\frac{1}{d(2l-2)}}|B|\,,
$$

which finishes the proof of part (a).

As for part (b), take a ball $B\subset V$ of radius $r$, and denote by $\hat B$ the ball with the same center and twice smaller radius. It is clear that there exists $\tau>0$ such that for any $\mathbf{g}\in\mathcal{G}$ one can choose $y\in\hat B$ with $\|\mathbf{g}(y)\|\ge\tau$ (otherwise, by a compactness argument similar to that of Lemma 4.3, one would get that $0|_{\hat B}\in\mathcal{G}$, contradicting (4.2)). Also take $K\ge\tau/r$ such that

$$
\sup_{\mathbf{g}\in\mathcal{G},\, x\in B,\, u\in\mathbb{R}^d_1} \|\frac{\partial\mathbf{g}}{\partial u}(x)\| \le K\,. \tag{4.9}
$$

Now let $B'\subset B$ be a ball of radius $\tau/2K \le r/2$ centered at $y$. Take $\mathbf{v}\in\mathbb{R}^2_1$ orthogonal to $\mathbf{g}(y)$. Applying Lemma 3.6 to $B'$ and the function $\mathbf{v}\cdot\mathbf{g}(x)$ one gets $\sup_{x\in B'}|\mathbf{v}\cdot\mathbf{g}(x)| \ge \frac{c}{k^k(k+1)!}(\tau/K)^k \ge \frac{c}{l^l(l+1)!}(\tau/K)^l$. On the other hand (4.9) shows that $\sup_{x\in B'}\|\mathbf{g}(x)-\mathbf{g}(y)\|$ is not greater than $\tau/2$. Now one can apply Lemma 4.2 to the map $\mathbf{g}: B'\mapsto\mathbb{R}^2$ to get

$$
\sup_{x\in B'}\|\tilde\nabla\mathbf{g}(x)\| \ge \frac{c}{l^l(l+1)!}\Big(\frac{\tau}{K}\Big)^{l-1}\frac{(\tau/2)^2}{\sqrt{\big(\frac{c}{l^l(l+1)!}(\tau/K)^l\big)^2 + (3\tau/2)^2}}\,,
$$

giving a uniform lower bound for $\sup_{x\in B}\|\tilde\nabla\mathbf{g}(x)\|$. □

## 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 $\{\mathbf{e}_0, \mathbf{e}^*_1, \dots, \mathbf{e}^*_d, \mathbf{e}_1, \dots, \mathbf{e}_n\}$. Also denote by $\Lambda$ the intersection of $\mathbb{Z}^m$ with the span of $\mathbf{e}_0, \mathbf{e}_1, \dots, \mathbf{e}_n$, that is,

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

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

$$U_x \stackrel{\text{def}}{=} \begin{pmatrix} 1 & 0 & \mathbf{f}(x) \\ 0 & I_d & \nabla\mathbf{f}(x) \\ 0 & 0 & I_n \end{pmatrix} \in SL_m(\mathbb{R})\,. \tag{5.2}$$

Note that $U_x \begin{pmatrix} p \\ 0 \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} \mathbf{f}(x)\mathbf{q} + p \\ \nabla\mathbf{f}(x)\mathbf{q} \\ \mathbf{q} \end{pmatrix}$ is the vector whose components appear in the right hand sides of the inequalities (1.7b). Therefore the fact that there exists $\mathbf{q} \in \mathbb{Z}^n \smallsetminus \{0\}$ 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 will be as follows: we will find a diagonal matrix $D \in GL_m(\mathbb{R})$ which transforms the above parallelepiped into a small cube; then the solvability of the above system of inequalities will force the lattice $DU_x\Lambda$ to have a small nonzero vector, and we will use a theorem proved by methods from [KM1] (see Theorem 6.2 below) to estimate the measure of the set of $x \in B$ for which it can happen.

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

$$D = \operatorname{diag}(a_0^{-1}, a_*^{-1}, \dots, a_*^{-1}, a_1^{-1}, \dots, a_n^{-1})\,, \tag{5.3}$$

where

$$a_0 = \frac{\delta}{\varepsilon},\ a_* = \frac{K}{\varepsilon},\ a_i = \frac{T_i}{\varepsilon},\ i = 1, \dots, n\,. \tag{5.4}$$

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

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

where $\|\cdot\|$ stands for the supremum norm. However from this point on it will be more convenient 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 $\mathbf{f}$ be as in Theorem 1.4. Take $\Lambda$ as in* (5.1) *and $U_x$ 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*

$$0 < a_0 \le 1, \quad a_n \ge \dots \ge a_1 \ge 1 \quad \text{and} \quad 0 < a_* \le (a_0 a_1 \dots a_{n-1})^{-1} \tag{5.6}$$

*and any positive $\varepsilon$, one has*

$$\Big|\big\{x \in B \mid \|DU_x\mathbf{v}\|_e < \varepsilon \text{ for some } \mathbf{v} \in \Lambda \smallsetminus \{0\}\big\}\Big| \le E\varepsilon^{\frac{1}{d(2l-1)}}|B|\,. \tag{5.7}$$

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

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

$$\big\{x \in B \;\big|\; \|DU_x\mathbf{v}\|_e < \varepsilon\sqrt{m} \text{ for some } \mathbf{v} \in \Lambda \smallsetminus \{0\}\big\}\,,$$

hence its measure is not greater than $Em^{\frac{1}{2d(2l-1)}}\varepsilon^{\frac{1}{d(2l-1)}}|B|$. □

## 6. Lattices and posets

The proof of Theorem 5.1 will depend 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 $\mathcal{A}(d,k,C,\alpha,\rho)$ to be the set of triples $(S,\varphi,B)$ where $S$ is a partially ordered set (*poset*), $B=B(\mathbf{x}_0,r_0)$, where $\mathbf{x}_0\in\mathbb{R}^d$ and $r_0>0$, and $\varphi$ is a mapping from $S$ to the space of continuous functions on $\tilde B\stackrel{\text{def}}{=}B\big(\mathbf{x}_0,3^kr_0\big)$ (this mapping will be denoted by $s\to\varphi_s$) such that the following holds:

(A0) the length of $S$ is not greater than $k$;
(A1) $\forall\, s\in S\,,\quad \varphi_s$ is $(C,\alpha)$-good on $B$;
(A2) $\forall\, s\in S\,,\quad \|\varphi_s\|_B\ge\rho$.
(A3) $\forall\, x\in B,\quad \#\{s\in S\mid |\varphi_s(x)|<\rho\}<\infty$.

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

(M1) $\varepsilon\le|\varphi_s(x)|\le\rho\quad\forall\, s\in\Sigma_x$;
(M2) $|\varphi_s(x)|\ge\rho\quad\forall\, s\in S\smallsetminus\Sigma_x$ comparable with any element of $\Sigma_x$.

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

**Theorem 6.1** (cf. [KM1, Theorem 4.1]). *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 $0<\varepsilon\le\rho$ one has*

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

We will 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 will denote by $\Gamma_\mathbb{R}$ the minimal linear subspace of $W$ containing $\Gamma$. Let $k=\dim(\Gamma_\mathbb{R})$ be the *rank* of $\Gamma$; say that $\mathbf{w}\in\bigwedge^k(W)$ *represents* $\Gamma$ if

$$\mathbf{w}=\begin{cases}1 & \text{if } k=0\\ \mathbf{v}_1\wedge\cdots\wedge\mathbf{v}_k & \text{if } k>0 \text{ and } \mathbf{v}_1,\dots,\mathbf{v}_k \text{ is a basis of } \Gamma\,.\end{cases}$$

We will need this exterior power representation mainly to be able to measure the "size" of discrete subgroups. Namely, these "sizes" will be given by suitable "norm-like" functions $\nu$ on the exterior algebra $\bigwedge(W)$ of $W$, and we will set

$$\nu(\Gamma)=\nu(\mathbf{w})\quad\text{if } \mathbf{w} \text{ represents } \Gamma\,. \tag{6.1}$$

More precisely, let us say that a function $\nu:\bigwedge(W)\mapsto\mathbb{R}_+$ is *submultiplicative* if

(i) $\nu$ is continuous (in the natural topology);
(ii) it is homogeneous, that is, $\nu(t\mathbf{w})=|t|\nu(\mathbf{w})$ for all $t\in\mathbb{R}$ and $\mathbf{w}\in\bigwedge(W)$;
(iii) $\nu(\mathbf{u}\wedge\mathbf{w})\le\nu(\mathbf{u})\nu(\mathbf{w})$ for all $\mathbf{u},\mathbf{w}\in\bigwedge(W)$.

Note that in view of (ii), (6.1) is a correct definition of $\nu(\Gamma)$.

Examples: if $W$ is a Euclidean space, one can extend the Euclidean structure to $\bigwedge(W)$ (by making $\bigwedge^i(W)$ and $\bigwedge^j(W)$ orthogonal for $i\ne j$); clearly then the Euclidean norm $\nu(\mathbf{w})=\|\mathbf{w}\|$ is submultiplicative. In this case the restriction of $\nu$ to $W$ coincides with the usual (Euclidean) norm on $W$. Also if $\mathcal{W} \subset \bigwedge(W)$ is an ideal, one can define $\nu(\mathbf{w})$ to be the norm of the projection of $\mathbf{w}$ orthogonal to $\mathcal{W}$. If this ideal is orthogonal to $W \subset \bigwedge(W)$, again the function $\nu$ will coincide with the norm when restricted to $W$.

We will also need the notion of *primitivity* of a discrete subgroup. If $\Lambda$ is a discrete subgroup of $W$, say that a subgroup $\Gamma$ of $\Lambda$ is *primitive* (in $\Lambda$) if $\Gamma = \Gamma_{\mathbb{R}} \cap \Lambda$, and denote by $\mathcal{L}(\Lambda)$ the set of all nonzero primitive subgroups of $\Lambda$. Example: a cyclic subgroup of $\mathbb{Z}^l$ is primitive in $\mathbb{Z}^l$ iff it is generated by a *primitive* vector (that is, a vector which is not equal to a nontrivial multiple of another element of $\mathbb{Z}^l$). 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 will derive from Theorem 6.1, is a generalization of Theorem 5.2 from [KM1].

**Theorem 6.2.** *Let $W$ be a finite-dimensional real vector space, $\Lambda$ 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 $H : \tilde{B} \to GL(W)$ be given, where $\tilde{B}$ stands for $B(x_0, 3^k r_0)$. Take $C, \alpha > 0$, $0 < \rho \le 1$, and let $\nu$ be a submultiplicative function on $\bigwedge(W)$. Assume that for any $\Gamma \in \mathcal{L}(\Lambda)$,*

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

(ii) *$\exists\, x \in B$ such that $\nu\big(H(x)\Gamma\big) \ge \rho$.*

*Also assume that*

(iii) *$\forall\, x \in \tilde{B}, \quad \#\big\{\Gamma \in \mathcal{L}(\Lambda) \bigm| \nu\big(H(x)\Gamma\big) < \rho\big\} < \infty.$*

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

$$\Big|\big\{x \in B \bigm| \nu\big(H(x)\mathbf{v}\big) < \varepsilon \text{ for some } \mathbf{v} \in \Lambda \smallsetminus \{0\}\big\}\Big| \le k(3^d N_d)^k \cdot C \left(\frac{\varepsilon}{\rho}\right)^{\alpha} |B| \,. \tag{6.2}$$

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 $\varepsilon$.

*Proof.* We will apply Theorem 6.1 to the triple $(S, \varphi, B)$, where $S = \mathcal{L}(\Lambda)$ and $\varphi$ is defined by $\varphi_\Gamma(x) \stackrel{\text{def}}{=} \nu\big(H(x)\Gamma\big)$. 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 (i), (A2) by (ii) and (A3) by (iii).

In view of Theorem 6.1, it remains to prove that

$$\Phi(\varepsilon, S, \varphi, B) \subset \big\{x \in B \mid \nu\big(H(x)\mathbf{v}\big) \ge \varepsilon \text{ for all } \mathbf{v} \in \Lambda \smallsetminus \{0\}\big\} \,. \tag{6.3}$$

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

$$\nu\big(H(x)\mathbf{v}\big) \ge \nu\big(H(x)\Delta\big)/\nu\big(H(x)\Gamma_{i-1}\big) \ge \varepsilon/\rho \ge \varepsilon \,.$$

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 $\mathbf{f}$ as in Theorem 1.4, set $m = n + d + 1$ and $W = \mathbb{R}^m$, and use the notation introduced in §5. Denote by $W^*$ the $d$-dimensional subspace spanned by $\mathbf{e}_1^*, \ldots, \mathbf{e}_d^*$, so that $\Lambda$ as in (5.1) is equal to the intersection of $\mathbb{Z}^m$ and $(W^*)^\perp$. Also let $\mathcal{H}$ be the family of functions $H : U \mapsto GL_m(\mathbb{R})$ given by $H(x) = DU_x$, where $U_x$ is as in (5.2), $D$ 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 $\mathcal{W} \subset \bigwedge(W)$ be the ideal generated by $\bigwedge^2(W^*)$, denote by $\pi$ the orthogonal projection with kernel $\mathcal{W}$, and take $\nu(\mathbf{w})$ to be the Euclidean norm of $\pi(\mathbf{w})$. In other words, if $\mathbf{w}$ is written as a sum of exterior products of base vectors $\mathbf{e}_i$ and $\mathbf{e}_i^*$, to compute $\nu(\mathbf{w})$ one should ignore components containing $\mathbf{e}_i^* \wedge \mathbf{e}_j^*$, $1 \le i \ne j \le d$, 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

- $\boxed{1}$ there exists $C > 0$ such that all the functions $x \mapsto \nu\big(H(x)\Gamma\big)$, where $H \in \mathcal{H}$ and $\Gamma \in \mathcal{L}(\Lambda)$, are $(C, \frac{1}{d(2l-1)})$-good on $\tilde{V}$,
- $\boxed{2}$ for every ball $B \subset \tilde{V}$ there exists $\rho > 0$ such that $\sup_{x \in B} \nu\big(H(x)\Gamma\big) \ge \rho$ for all $H \in \mathcal{H}$ and $\Gamma \in \mathcal{L}(\Lambda)$, and
- $\boxed{3}$ for all $x \in \tilde{V}$ and $H \in \mathcal{H}$ one has $\#\big\{\Gamma \in \mathcal{L}(\Lambda) \mid \nu\big(H(x)\Gamma\big) \le 1\big\} < \infty$.

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 $C(n+1)(3^d N_d)^{n+1} (\varepsilon/\rho)^{1/d(2l-1)} |B|$ for any $\varepsilon \le \rho$, therefore not greater than

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

for any positive $\varepsilon$.

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

Let $k$ be the rank of $\Gamma$. The claims are trivial for $\Gamma = \{0\}$, thus we can set $1 \le k \le n+1$. Since $D\Gamma_{\mathbb{R}}$ is a $k$-dimensional subspace of $(W^*)^\perp = \mathbb{R}\mathbf{e}_0 \oplus \mathbb{R}\mathbf{e}_1 \oplus \cdots \oplus \mathbb{R}\mathbf{e}_n$, it is possible to choose an orthonormal set $\mathbf{v}_1, \ldots, \mathbf{v}_{k-1} \in D\Gamma_{\mathbb{R}}$ such that each $\mathbf{v}_i$, $i = 1, \ldots, k-1$, is orthogonal to $\mathbf{e}_0$. Now let us consider two cases:

**Case 1.** $D\Gamma_{\mathbb{R}}$ contains $\mathbf{e}_0$; then $\{\mathbf{e}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{k-1}\}$ is a basis of $\mathbb{R}\mathbf{e}_0 \oplus D\Gamma_{\mathbb{R}}$. Thus one can find $\mathbf{w} \in \bigwedge^k(\mathbb{R}^m)$ representing $\Gamma$ such that $D\mathbf{w}$ can be written as $a\mathbf{e}_0 \wedge \mathbf{v}_1 \cdots \wedge \mathbf{v}_{k-1}$ for some $a > 0$.

**Case 2.** $D\Gamma_{\mathbb{R}}$ does not contain $\mathbf{e}_0$; then it is possible to choose $\mathbf{v}_0 \in \mathbb{R}\mathbf{e}_0 \oplus D\Gamma_{\mathbb{R}}$ such that $\{\mathbf{e}_0, \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{k-1}\}$ is an orthonormal basis of $\mathbb{R}\mathbf{e}_0 \oplus D\Gamma_{\mathbb{R}}$. In this case, one can represent $\Gamma$ by $\mathbf{w}$ such that

$$D\mathbf{w} = (a\mathbf{e}_0 + b\mathbf{v}_0) \wedge \mathbf{v}_1 \cdots \wedge \mathbf{v}_{k-1} = a\mathbf{e}_0 \wedge \mathbf{v}_1 \cdots \wedge \mathbf{v}_{k-1} + b\mathbf{v}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{k-1} \tag{7.1}$$

for some $a, b \in \mathbb{R} \smallsetminus \{0\}$. 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 $\mathbf{v}_0$ is not defined.

Now write

$$H(x)\mathbf{w} = DU_xD^{-1}(D\mathbf{w}) \,, \tag{7.2}$$

and introduce the $m$-tuple (interpreted as a row vector) of functions

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

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

$$\tilde{\nabla}^* \mathbf{g}(x) \stackrel{\text{def}}{=} g_1(x) \nabla^* g_2(x) - g_2(x) \nabla^* g_1(x) \,.$$

Then it becomes straightforward to verify that $DU_xD^{-1}\mathbf{e}_0 = \mathbf{e}_0$ and

$$DU_xD^{-1}\mathbf{v} = \mathbf{v} + \big(\hat{\mathbf{f}}(x)\mathbf{v}\big)\mathbf{e}_0 + \frac{a_0}{a_*} \nabla^* \big(\hat{\mathbf{f}}(x)\mathbf{v}\big)$$

whenever $\mathbf{v}$ is orthogonal to $\mathbf{e}_0$ and $W^*$. Therefore the space $\mathbb{R}\mathbf{e}_0 \oplus W^* \oplus D\Gamma_{\mathbb{R}}$ is invariant under $DU_xD^{-1}$; hence we can restrict ourselves to the coordinates of $DU_xD^{-1}$-image of $D\mathbf{w}$ with respect to the basis chosen above, and write

$$DU_xD^{-1}(\mathbf{e}_0 \wedge \mathbf{v}_1 \cdots \wedge \mathbf{v}_{k-1}) = \mathbf{e}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{k-1} + \frac{a_0}{a_*} \sum_{i=1}^{k-1} \pm \nabla^* \big(\hat{\mathbf{f}}(x)\mathbf{v}_i\big) \, \mathbf{e}_0 \wedge \bigwedge_{s \neq i} \mathbf{v}_s + \mathbf{w}_1^*$$

and

$$\begin{aligned} DU_xD^{-1}(\mathbf{v}_0 \wedge \cdots \wedge \mathbf{v}_{k-1}) \;&=\; \mathbf{v}_0 \wedge \cdots \wedge \mathbf{v}_{k-1} \;+\; \sum_{i=0}^{k-1} \pm \big(\hat{\mathbf{f}}(x)\mathbf{v}_i\big) \, \mathbf{e}_0 \wedge \bigwedge_{s \neq i} \mathbf{v}_s \\ &+ \frac{a_0}{a_*} \sum_{i=0}^{k-1} \pm \nabla^* \big(\hat{\mathbf{f}}(x)\mathbf{v}_i\big) \wedge \bigwedge_{s \neq i} \mathbf{v}_s \;+\; \frac{a_0}{a_*} \sum_{\substack{i=0 \\ j>i}}^{k-1} \pm \tilde{\nabla}^* \Big(\hat{\mathbf{f}}(x)\mathbf{v}_i, \hat{\mathbf{f}}(x)\mathbf{v}_j\Big) \wedge \mathbf{e}_0 \wedge \bigwedge_{s \neq i,j} \mathbf{v}_s + \mathbf{w}_2^* \,, \end{aligned}$$

where $\mathbf{w}_1^*$ and $\mathbf{w}_2^*$ belong to $\bigwedge^2(W^*) \wedge \bigwedge(W)$.

Collecting terms and using (7.1) and (7.2), one finds that

$$\begin{aligned} \pi\big(H(x)\mathbf{w}\big) \;&=\; \big(a + b\hat{\mathbf{f}}(x)\mathbf{v}_0\big) \, \mathbf{e}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{k-1} \;+\; b \, \mathbf{v}_0 \wedge \cdots \wedge \mathbf{v}_{k-1} \\ &+\; b \sum_{i=1}^{k-1} \pm \big(\hat{\mathbf{f}}(x)\mathbf{v}_i\big) \, \mathbf{e}_0 \wedge \bigwedge_{s \neq i} \mathbf{v}_s \;+\; b \frac{a_0}{a_*} \sum_{i=0}^{k-1} \pm \nabla^* \big(\hat{\mathbf{f}}(x)\mathbf{v}_i\big) \wedge \bigwedge_{s \neq i} \mathbf{v}_s \\ &+\; \frac{a_0}{a_*} \sum_{i=1}^{k-1} \pm \tilde{\nabla}^* \big(\hat{\mathbf{f}}(x)\mathbf{v}_i, a + b\hat{\mathbf{f}}(x)\mathbf{v}_0\big) \wedge \mathbf{e}_0 \wedge \bigwedge_{s \neq 0,i} \mathbf{v}_s \\ &+\; b \frac{a_0}{a_*} \sum_{\substack{i,j=1 \\ j>i}}^{k-1} \pm \tilde{\nabla}^* \big(\hat{\mathbf{f}}(x)\mathbf{v}_i, \hat{\mathbf{f}}(x)\mathbf{v}_j\big) \wedge \mathbf{e}_0 \wedge \bigwedge_{s \neq i,j} \mathbf{v}_s \end{aligned} \tag{7.3}$$

Now the stage is set to prove condition $\boxed{1}$. Indeed, in view of Lemma 3.1(bc), it suffices to show that the norms of each of the summands in (7.3) are $(C,\alpha)$-good functions. Note that the elements appearing in the first two lines are linear combinations of either $1, f_1, \dots, f_n$ or $\nabla^* f_1, \dots, \nabla^* f_n$, hence one can, using Corollary 3.5, find $\tilde{V}_1 \ni x_0$ and $C_1 > 0$ such that all these norms are $(C_1, 1/n)$-good on $\tilde{V}_1$. The same can be said about the rest of the components when $b = 0$. Otherwise one can observe that the summands in the last two lines of (7.3) are of the form $\pm\tilde{\nabla}^*(L\mathbf{g})$ where $L$ is some linear transformation of $\mathbb{R}^2$ and $\mathbf{g}$ belongs to

$$\mathcal{G} \stackrel{\text{def}}{=} \left\{ \big(\mathbf{f}(x)\mathbf{u}_1, \mathbf{f}(x)\mathbf{u}_2 + u_0\big) \mid u_0 \in \mathbb{R}, \mathbf{u}_1 \perp \mathbf{u}_2 \in \mathbb{R}^n_1 \right\}. \tag{7.4}$$

Since $\tilde{\nabla}^*(L\mathbf{g})$ is proportional to $\tilde{\nabla}^*\mathbf{g}$, it suffices, in view of Lemma 3.1(a), to prove the existence of $C_2 \geq C_1$ and a neighborhood $\tilde{V}_2 \subset \tilde{V}_1$ of $x_0$ such that the norms of $\tilde{\nabla}^*\mathbf{g}$ are $(C_2, \frac{1}{d(2l-1)})$-good on $\tilde{V}_2$ for any $\mathbf{g} \in \mathcal{G}$. The latter is a direct consequence of Proposition 4.1(a), since the family $\{\nabla g_i \mid \mathbf{g} = (g_1, g_2) \in \mathcal{G}\}$ is compact in $C^{l-1}(U)$, and it follows from the nondegeneracy of $\mathbf{f}$ at $x_0$ that condition (4.2) is satisfied.

One can also use (7.3) to prove $\boxed{3}$. Indeed, looking at the first line of (7.3) one sees that

$$\nu\big(H(x)\Gamma\big) \leq \max\big(|a + b\hat{\mathbf{f}}(x)\mathbf{v}_0|, |b|\big),$$

thus $\nu\big(H(x)\Gamma\big) \leq 1$ would imply $|b| \leq 1$ and $|a| \leq 1 + \|\hat{\mathbf{f}}(x)\|$. On the other hand, in view of (7.1), the norm of $D\mathbf{w}$ is equal to $\sqrt{a^2 + b^2}$, and from the discreteness of $\bigwedge(D\Lambda)$ in $\bigwedge(\mathbb{R}^m)$ 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 $\boxed{2}$. Take any neighborhood $B$ of $x_0$. It follows from the linear independence of the functions $1, f_1, \dots, f_n$ and from the linear independence of their gradients that there exists $\rho_1 > 0$ such that

$$\forall\, \mathbf{v} \in \mathbb{R}^n_1 \ \forall\, v_0 \in \mathbb{R} \text{ one has } \sup_{x \in B} |\mathbf{f}(x)\mathbf{v} + v_0| \geq \rho_1 \text{ and } \sup_{x \in B} |\nabla\big(\mathbf{f}(x)\mathbf{v}\big)| \geq \rho_1\,. \tag{7.5}$$

Also let $\rho_2 = \rho(\mathcal{G}, B)$ where $\mathcal{G}$ is the class of 2-tuples of functions defined in (7.4), and $\rho(\mathcal{G}, B)$ is as in Proposition 4.1(b). Consider

$$M \stackrel{\text{def}}{=} \max\big(\sup_{x \in B} \|\mathbf{f}(x)\|, \sup_{x \in B} \|\nabla\mathbf{f}(x)\|\big);$$

we will show that (ii) will hold 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 $k = \dim(\Gamma_{\mathbb{R}}) = 1$. Then $\Gamma$ can be represented by a vector $\mathbf{v} = (v_0, 0, \dots, 0, v_1, \dots, v_n)^T$ with integer coordinates, and it is straightforward to verify that the first coordinate of $H(x)\mathbf{v}$ will be equal to $\frac{1}{a_0}\big(v_0 + v_1 f_1(x) + \cdots + v_n f_n(x)\big)$, which will deviate 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 will be similar to that of the proof of the previous lemma: given $\mathbf{w} \in \bigwedge^k(\mathbb{R}^m)$ representing $\Gamma$, we will choose a suitable orthogonal decomposition of $\bigwedge^k(\mathbb{R}^m)$ and then show that the norm of the projection of $H(x)\mathbf{w}$ to some subspace will be 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 $\mathbf{e}_n$, which is the eigenvector of $D$ with the smallest eigenvalue. We do it by first choosing an orthonormal set $\mathbf{v}_1, \dots, \mathbf{v}_{k-2} \in \Gamma_{\mathbb{R}}$ such that each $\mathbf{v}_i$, $i = 1, \dots, k-2$, is orthogonal to both $\mathbf{e}_0$ and $\mathbf{e}_n$. Then choose $\mathbf{v}_{k-1}$ orthogonal to $\mathbf{v}_i$, $i = 1, \dots, k-2$, and to $\mathbf{e}_0$ (but in general not to $\mathbf{e}_n$). Now, if necessary (see the remark after (7.1)), choose a vector $\mathbf{v}_0$ to complete $\{\mathbf{e}_0, \mathbf{v}_1, \dots, \mathbf{v}_{k-1}\}$ to an orthonormal basis of $\mathbb{R}\mathbf{e}_0 \oplus \Gamma_{\mathbb{R}}$. This way, similarly to (7.1), we will represent $\Gamma$ by $\mathbf{w}$ of the form

$$\mathbf{w} = (a\mathbf{e}_0 + b\mathbf{v}_0) \wedge \mathbf{v}_1 \cdots \wedge \mathbf{v}_{k-1} = a\mathbf{e}_0 \wedge \mathbf{v}_1 \cdots \wedge \mathbf{v}_{k-1} + b\mathbf{v}_0 \wedge \cdots \wedge \mathbf{v}_{k-1} \tag{7.6}$$

for some $a, b \in \mathbb{R}$ with $a^2 + b^2 \geq 1$. As before, we will use (7.6) even when $\mathbf{v}_0$ is not defined, in this case the coefficient $b$ will vanish.

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

$$\check{\mathbf{f}}(x) = \big(1, 0, \dots, 0, f_1(x), \dots, f_n(x)\big)\,,$$

and observe that $U_x\mathbf{e}_0 = \mathbf{e}_0$ and

$$U_x\mathbf{v} = \mathbf{v} + \big(\check{\mathbf{f}}(x)\mathbf{v}\big)\mathbf{e}_0 + \nabla^*\big(\check{\mathbf{f}}(x)\mathbf{v}\big)$$

whenever $\mathbf{v}$ is orthogonal to $\mathbf{e}_0$ and $\mathbf{e}_*$. Using this and (7.6), one can obtain an expression analogous to (7.3). This time however we are interested only in the terms of the form $\mathbf{e}_0 \wedge \mathbf{e}_i^* \wedge \mathbf{w}'$, where $\mathbf{w}'$ is orthogonal to $\bigwedge^{k-2}\big((\mathbb{R}\mathbf{e}_0 \oplus W^*)^\perp\big)$ (note that these terms are present only if $k \geq 2$). Namely, let us write

$$\begin{aligned}\pi(U_x\mathbf{w}) \;=\; & \big(a + b\big(\check{\mathbf{f}}(x)\mathbf{v}_0\big)\big)\, \mathbf{e}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{k-1} \;+\; b\,\mathbf{v}_0 \wedge \cdots \wedge \mathbf{v}_{k-1} \\ & +\; b\sum_{i=1}^{k-1} \pm\big(\check{\mathbf{f}}(x)\mathbf{v}_i\big)\,\mathbf{e}_0 \wedge \bigwedge_{s\neq i}\mathbf{v}_s \;+\; b\sum_{i=0}^{k-1} \pm\nabla^*\big(\check{\mathbf{f}}(x)\mathbf{v}_i\big) \wedge \bigwedge_{s\neq i}\mathbf{v}_s \;+\; \mathbf{e}_0 \wedge \check{\mathbf{w}}(x)\,,\end{aligned}$$

where

$$\check{\mathbf{w}}(x) \stackrel{\text{def}}{=} \sum_{i=1}^{k-1} \pm\tilde{\nabla}^*\big(\check{\mathbf{f}}(x)\mathbf{v}_i, a+b\check{\mathbf{f}}(x)\mathbf{v}_0\big) \wedge \bigwedge_{s\neq 0,i}\mathbf{v}_s + b\sum_{\substack{i,j=1\\ j>i}}^{k-1} \pm\tilde{\nabla}^*\big(\check{\mathbf{f}}(x)\mathbf{v}_i, \check{\mathbf{f}}(x)\mathbf{v}_j\big) \wedge \bigwedge_{s\neq i,j}\mathbf{v}_s\,. \tag{7.7}$$

Note that $\mathbf{e}_0 \wedge \check{\mathbf{w}}(x)$ lies in the space $\mathbf{e}_0 \wedge W^* \wedge \bigwedge^{k-2}\big((W^*)^\perp\big)$, while $\pi(U_x\mathbf{w}) - \mathbf{e}_0 \wedge \check{\mathbf{w}}(x)$ belongs to its orthogonal complement. Since both spaces are $D$-invariant, to prove that $\nu\big(H(x)\mathbf{w}\big) = \|\pi(DU_x\mathbf{w})\| = \|D\pi(U_x\mathbf{w})\|$ is not less than $\rho$ for some $x \in B$ it will suffice to show that $\sup_{x\in B}\|D\check{\mathbf{w}}(x)\|$ is not less than $a_0\rho$.

Now consider the product $\mathbf{e}_n \wedge \check{\mathbf{w}}(x)$. We claim that it is enough to show that

$$\|\mathbf{e}_n \wedge \check{\mathbf{w}}(x)\| \geq \rho \text{ for some } x \in B\,. \tag{7.8}$$

Indeed, since $\mathbf{e}_n$ is an eigenvector of $D$ with eigenvalue $a_n^{-1}$, for any $x \in B$ the norm of $D\big(\mathbf{e}_n \wedge \check{\mathbf{w}}(x)\big)$ is not greater than $a_n^{-1}\|D\check{\mathbf{w}}(x)\|$. Therefore, since the smallest eigenvalue of $D$ on $W^* \wedge \bigwedge^{k-2}\big((W^*)^\perp\big)$ is equal to $(a_* a_{n-k+1} \cdot \dots \cdot a_n)^{-1}$, the norm of $D\check{\mathbf{w}}(x)$ will be not less than

$$a_n\|D\big(\mathbf{e}_n \wedge \check{\mathbf{w}}(x)\big)\| \geq \frac{a_n}{a_* a_{n-k+1} \cdot \dots \cdot a_n}\|\mathbf{e}_n \wedge \check{\mathbf{w}}(x)\|$$
$$\underset{\text{for some } x\in B, \text{ by (7.8)}}{\geq} \frac{\rho}{a_* a_{n-k+1} \cdot \dots \cdot a_{n-1}} \underset{\text{since } a_i \geq 1}{\geq} \frac{a_0\rho}{a_0 a_* a_1 \cdot \dots \cdot a_{n-1}} \underset{\text{by (5.6)}}{\geq} a_0\rho\,.$$

Thus it remains to prove (7.8). For this let us select the term containing $\mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_{k-2}$, and multiply (7.7) by $\mathbf{e}_n$ as follows:

$$\begin{aligned} \mathbf{e}_n \wedge \check{\mathbf{w}}(x) &= \pm \mathbf{v}^*(x) \wedge \mathbf{e}_n \wedge \mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_{k-2} \\ &+ \text{other terms where one or two of } \mathbf{v}_i,\ i = 1, \dots, k-2, \text{ are missing}, \end{aligned} \tag{7.9}$$

where

$$\mathbf{v}^*(x) \stackrel{\text{def}}{=} \tilde{\nabla}^*\big(\check{\mathbf{f}}(x)\mathbf{v}_{k-1}, a + b\check{\mathbf{f}}(x)\mathbf{v}_0\big) = b\tilde{\nabla}^*\big(\check{\mathbf{f}}(x)\mathbf{v}_{k-1}, \check{\mathbf{f}}(x)\mathbf{v}_0\big) - \nabla^*\big(\check{\mathbf{f}}(x)\mathbf{v}_{k-1}\big)\,.$$

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

## 8. Completion of the proof and concluding remarks

**8.1.** First let us finish the proof of Theorem 1.1 by writing down the

*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 $\mathbf{f} = (f_1, \dots, f_n)$ of $C^m$ functions on $U$ and a function $\Psi : \mathbb{Z}^n \smallsetminus \{0\} \mapsto (0, \infty)$ satisfying (1.1) and(1.5). Take $x_0 \in U$ such that $\mathbf{f}$ is $l$-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 twice bigger radius) is contained in $U$. We are going to prove that for a.e. $x \in B$ one has $\mathbf{f}(x) \in \mathcal{W}(\Psi)$. In other words, define $A(\mathbf{q})$ to be the set of $x \in B$ satisfying $|\langle\mathbf{f}(x)\mathbf{q}\rangle| < \Psi(\mathbf{q})$; we need to show that points $x$ which belong to infinitely many sets $A(\mathbf{q})$ 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 $A(\mathbf{q})$ such that $q_i = 0$ for some $i = 1, \dots, n$. It remains to show that the same is true if one includes integer vectors $\mathbf{q}$ 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 $A_{\geq}(\mathbf{q})$ the set of $x \in A(\mathbf{q})$ satisfying (1.6d), and set $A_{<}(\mathbf{q}) \stackrel{\text{def}}{=} A(\mathbf{q}) \smallsetminus A_{\geq}(\mathbf{q})$. Theorem 1.3 guarantees that the measure of $A_{\geq}(\mathbf{q})$ is not greater than $C_d\Psi(\mathbf{q})|B|$ whenever $\mathbf{q}$ is far enough from the origin. Because of (1.1), the sum of measures of the sets $A_{\geq}(\mathbf{q})$ is finite, hence, by the Borel-Cantelli Lemma, almost every $x \in B$ is contained in at most finitely many sets $A_{\geq}(\mathbf{q})$.

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

$$\bigcup_{\mathbf{q}\in Q,\, 2^{t_i}\leq |q_i| < 2^{t_i+1}} A_{<}(\mathbf{q}) \tag{8.1}$$

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

$$\frac{\delta K T_1 \cdot \ldots \cdot T_n}{\max_i T_i} = \frac{2^{-\sum_i t_i}\sqrt{ndL}2^{\|\mathbf{t}\|/2}2^{n+\sum_i t_i}}{2^{\|\mathbf{t}\|+1}} = \sqrt{ndL}2^{n-1-\|\mathbf{t}\|/2},$$

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

$$E\left(\sqrt{ndL}2^{n-1}\right)^{-\frac{1}{d(2l-1)}} 2^{-\frac{1}{2d(2l-1)(n+1)}\|\mathbf{t}\|}.$$

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

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

$$\Psi(\mathbf{q}) = \psi(\|\mathbf{q}\|_{\mathbf{s}}).$$

We will normalize $\mathbf{s}$ so that $\sum_i s_i = 1$ (this way, for example, one can see that, for a non-increasing $\psi$, any $\mathbf{s}$-$\psi$-approximable $\mathbf{y}$ is $\psi$-MA). The choice $\mathbf{s} = (1/n, \dots, 1/n)$ 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 part (S) of Corollary 1.2:

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

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

$$\begin{cases} |P(x)| < \|\mathbf{q}\|^{-n+\gamma} \\ |P'(x)| < \|\mathbf{q}\|^{1-\gamma-\varepsilon} \end{cases} \tag{8.2}$$

(here $P(x)$ is the polynomial $p+q_1x+\cdots+q_nx^n$) has at most finitely many solutions $p\in\mathbb{Z}$, $\mathbf{q}\in\mathbb{Z}^n$. This was used to construct a certain regular system of real numbers and obtain the sharp lower estimate for the Hausdorff dimension of the set

$$\left\{x\in\mathbb{R}\;\middle|\;|\langle q_1x+\cdots+q_nx^n\rangle|<\|\mathbf{q}\|^{-\lambda}\text{ for infinitely many }\mathbf{q}\in\mathbb{Z}^n\right\}$$

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

In 1995 V. 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 $\mathbf{f}=(f_1,\ldots,f_n)$ 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 $\mathbf{q}\in\mathbb{Z}^n$ of the system*

$$\left\{\begin{array}{l}|\langle\mathbf{f}(x)\mathbf{q}\rangle|<\Pi_+(\mathbf{q})^{-1+\gamma/n}\\ \|\nabla\mathbf{f}(x)\mathbf{q}\|<\|\mathbf{q}\|^{1-\gamma-\varepsilon}\end{array}\right. \tag{8.3}$$

*Proof.* As in the proof of Theorem 1.1, one can use induction to be left with integer vectors $\mathbf{q}$ with all coordinates different from zero. Then one estimates the measure of the union of the sets of solutions of (8.3) over all $\mathbf{q}\in\mathbb{Z}^n$ with $2^{t_i}\le|q_i|<2^{t_i+1}$ for any $n$-tuple $\mathbf{t}=(t_1,\ldots,t_n)\in\mathbb{Z}^n_+$ by using Theorem 1.4 with $\delta=2^{(-1+\gamma/n)\sum_i t_i}$, $K=\sqrt{ndL}2^{-(1-\gamma-\varepsilon)\|\mathbf{t}\|}$ and $T_i=2^{t_i+1}$. Inequalities (1.7a) are clearly satisfied; in particular one has

$$\frac{\delta KT_1\cdot\ldots\cdot T_n}{\max_i T_i}=2^{\frac{\gamma}{n}(\sum_i t_i-n\|\mathbf{t}\|)}2^{-\varepsilon\|\mathbf{t}\|}\le 2^{-\varepsilon\|\mathbf{t}\|}\le 1\,.$$

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

$$E\max\left(2^{\frac{-1+\gamma/n}{d(2l-1)}\sum_i t_i},2^{-\frac{\varepsilon}{d(2l-1)(n+1)}\|\mathbf{t}\|}\right).$$

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

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

$$\left\{\begin{array}{l}|\langle\mathbf{f}(x)\mathbf{q}\rangle|<\Psi_1(\mathbf{q})\\ \|\nabla\mathbf{f}(x)\mathbf{q}\|<\Psi_2(\mathbf{q})\end{array}\right. \tag{8.4}$$

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

**8.4.** The main result of the paper (Theorem 1.1) was proved already in the summer of 1998, but only in the case when the functions $f_1, \dots, f_n$ are analytic. More precisely, the analytic set-up was reduced to the case $d = 1$ (see [Sp4, §3] or [P] for a related "slicing" technique). In the latter case, in addition to the nondegeneracy of $\mathbf{f}$, 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

$$\text{Span}\left(f_1', \dots, f_n'; \begin{vmatrix} f_i & f_j \\ f_i' & f_j' \end{vmatrix}, 1 \le i < j \le n\right) \text{ consists of functions } (C,\alpha)\text{-good on } B\,.$$

For analytic functions $f_1, \dots, f_n$ 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

$$\left\{x \in B \;\middle|\; \exists\, \mathbf{q} \in \mathbb{Z}^n \smallsetminus \{0\} \text{ such that } \begin{cases} |\langle \mathbf{f}(x)\mathbf{q}\rangle| < \delta \\ |\mathbf{g}(x)\mathbf{q}| < K \\ |q_i| < T_i, \quad i = 1, \dots, n \end{cases} \right\}, \tag{8.5}$$

with $\mathbf{f}$ as in Theorem 1.4 and $\mathbf{g}$ another nondegenerate $n$-tuple of functions on $U$. We were able to prove an analogue of Theorem 1.4 for sets (8.5) but only for $K \ge 1$, and for $n$-tuples $\mathbf{f}$ and $\mathbf{g}$ with an additional assumption that

$$\text{Span}\left(g_1, \dots, g_n; \begin{vmatrix} f_i & f_j \\ g_i & g_j \end{vmatrix}, 1 \le i < j \le n\right) \text{ consists of } (C,\alpha)\text{-good functions}\,. \tag{8.6}$$

Instead of $U_x$ as in (5.2) we considered more general matrices

$$U_x^{\mathbf{f},\mathbf{g}} \stackrel{\text{def}}{=} \begin{pmatrix} 1 & 0 & \mathbf{f}(x) \\ 0 & 1 & \mathbf{g}(x) \\ 0 & 0 & I_n \end{pmatrix}.$$

. To prove an analogue of condition (i) of Theorem 6.2, or, more precisely, the statement that for some positive $C, \alpha$ and any subgroup $\Gamma$ of $\Lambda$ the function $x \mapsto \|DU_x^{\mathbf{f},\mathbf{g}}\Gamma\|$ is $(C,\alpha)$-good on some neighborhood $B$ of $x_0$, it was enough to consider the standard basis $\{\mathbf{e}_0, \mathbf{e}_*\mathbf{e}_1, \dots, \mathbf{e}_n\}$ of $\mathbb{R}^{n+1}$ and the corresponding basis

$$\left\{\mathbf{e}_I \stackrel{\text{def}}{=} \mathbf{e}_{i_1} \wedge \dots \wedge \mathbf{e}_{i_k} \mid I = \{i_1, \dots, i_k\} \subset \{0, *, 1, \dots, n\}\right\}$$

of $\bigwedge^k(\mathbb{R}^{n+1})$, decompose an element $\mathbf{w}$ representing $\Gamma$ as $\mathbf{w} = \sum_I w_I \mathbf{e}_I$, write an expansion similar to (7.3) and use (8.6).

To prove an analogue of condition (ii), that is the statement that for any neighborhood $B$ of $x_0$ there exists $\rho > 0$ such that $\sup_{x \in B} \|DU_x^{\mathbf{f},\mathbf{g}}\Gamma\| \ge \rho$ for every $\Gamma \subset \Lambda$, we used the fact that the coefficients $w_I$ are integers and considered the following two cases: 1) $w_I \ne 0$ for some $I \subset \{1, \dots, n\}$, and 2) $w_I = 0$ for all $I \subset \{1, \dots, n\}$. In the just described approach it was important that $K \ge 1$. This was enough for the proof of the "d = 1"-case of Theorem 1.1; however, as we saw in §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 the paper. It was proved by A. Khintchine in 1924 [Kh] (resp. by A. Groshev in 1938 [G]) that a.e. $\mathbf{y} \in \mathbb{R}$ (resp. $\mathbb{R}^n$) is $\psi$-approximable whenever $\psi$ is a non-increasing function which does not satisfy (1.1s). In 1960 W. Schmidt [S1] showed that a.e. $\mathbf{y} \in \mathbb{R}^n$ belongs to $\mathcal{W}(\Psi)$ 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 $\mathbf{f} : U \mapsto \mathbb{R}^n$ as in Theorem 1.1 and $\Psi$ satisfying (1.5) but not (1.1), the set $\{x \in U \mid \mathbf{f}(x) \in \mathcal{W}(\Psi)\}$ 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 [DRV2, DRV3, BBDD, Be1, Be3, Be4].

## Acknowledgements

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 SFB-343 and Humboldt Foundation. The paper was completed during the Spring 2000 programme on Ergodic Theory, Geometric Rigidity and Number Theory at the Isaac Newton Institute of Mathematical Sciences. Thanks are also due to V. Beresnevich and M. Dodson for useful remarks.

## References

[B1] A. Baker, *On a theorem of Sprindzhuk*, Proc. Roy. Soc. London **A 292** (1966), 92–104.

[B2] ———, *Transcendental number theory*, Cambridge Univ. Press, Cambridge, 1975.

[Be1] V. Beresnevich, *On approximation of real numbers by real algebraic numbers*, Acta Arith. **90** (1999)), 97–112.

[Be2] ———, *Optimal approximation order of points of smooth curves in 3-dimensional Euclidean space*, Dokladi NAN Belarusi **43** (1999), no. 4, 9–13.

[Be3] ———, *Application of the concept of a regular system of points in metric number theory*, Vestsī Akad. Navuk Belarusī Ser. Fīz.-Mat. Navuk (2000), no. 1, 35–39.

[Be4] ———, *On proof of Khintchine type theorem for curves*, Vestsī Akad. Navuk Belarusī Ser. Fīz.-Mat. Navuk (to appear).

[Be5] ———, *A Groshev type theorem for convergence on manifolds*, Acta Math. Hungar. (to appear).

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

[Bo] V. Borbat, *Joint zero approximation by the values of integral polynomials and their derivatives*, Vestsī Akad. Navuk Belarusī Ser. Fīz.-Mat. Navuk (1995), no. 1, 9–16.

[BB] V. Bernik and V. Borbat, *Polynomials with differences in values of coefficients and a conjecture of A. Baker*, Vestsī Akad. Navuk Belarusī Ser. Fīz.-Mat. Navuk (1997), no. 3, 5–8. (Russian)

[BBDD] V. Beresnevich, V. Bernik, H. Dickinson and M. Dodson, *The Khintchine-Groshev theorem for planar curves*, Proc. Roy. Soc. London **A 455** (1999), 3053–3063.

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

[BDD] V. Bernik, H. Dickinson and M. Dodson, *A Khintchine-type version of Schmidt's theorem for planar curves*, Proc. Roy. Soc. London **A 454** (1998), 179–185.

[BKM] V. Bernik, D. Kleinbock and G. A. Margulis, *Khintchine-type theorems on manifolds: convergence case for standard and multiplicative versions*, Preprint 99 − 092, Universität Bielefeld, SFB 343 "Diskrete Strukturen in der Mathematik" (1999).

[BS] A. Baker and W. Schmidt, *Diophantine approximation and Hausdorff dimension*, Proc. Lond. Math. Soc. **21** (1970), 1–11.

[D] H. Davenport, *A note on binary cubic forms*, Mathematika **8** (1961), 58–62.

[Do] M. M. Dodson, *Geometric and probabilistic ideas in metric Diophantine approximation*, Russian Math. Surveys **48** (1993), 73–102.

[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.

[G] A. V. Groshev, *Une théorème sur les systèmes des formes linéaires*, Dokl. Akad. Nauk SSSR **9** (1938), 151–152.

[Kh] A. Khintchine, *Einige Sätze über Kettenbrüche, mit Anwendungen auf die Theorie der Diophantischen Approximationen*, Math. Ann. **92** (1924), 115–125.

[Kl] D. Kleinbock, *Flows on homogeneous spaces and Diophantine properties of matrices*, Duke Math. J. **95** (1998), 107–124.

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

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

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

[P] A. S. Pyartli, *Diophantine approximations on submanifolds of Euclidean space*, Functional Anal. Appl. **3** (1969), 303–306.

[S1] W. Schmidt, *A metrical theorem in Diophantine approximation*, Canadian J. Math. **12** (1960), 619–631.

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

[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.

V. I. Bernik, Institute of Mathematics, Belarus Academy of Sciences, Minsk, 220072, Belarus
*E-mail address*: bernik@im.bas-net.by

D. Kleinbock, Department of Mathematics, Brandeis University, Waltham, MA 02454-9110
*E-mail address*: kleinboc@brandeis.edu

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