J. reine angew. Math. **700** (2015), 93–112
DOI 10.1515/crelle-2013-0002

Journal für die reine und angewandte Mathematik

On the Morse–Sard property and level sets of $W^{n,1}$ Sobolev functions on $\mathbb{R}^n$

By *Jean Bourgain* at Princeton, *Mikhail V. Korobkov* at Novosibirsk and *Jan Kristensen* at Oxford

Abstract. We establish Luzin N and Morse–Sard properties for functions from the Sobolev space $W^{n,1}(\mathbb{R}^n)$. Using these results we prove that almost all level sets are finite disjoint unions of C^1-smooth compact manifolds of dimension $n-1$. These results remain valid also within the larger space of functions of bounded variation $BV_n(\mathbb{R}^n)$. For the proofs we establish and use some new results on Luzin-type approximation of Sobolev and BV-functions by C^k-functions, where the exceptional sets have small Hausdorff content.

Introduction

The starting point of the paper is the following classical result (see also [11] for more general expositions):

Theorem (Morse–Sard 1942, [15, 18]). *Let $f:\mathbb{R}^n \to \mathbb{R}^m$ be a C^k-smooth mapping with $k \geq \max(n-m+1, 1)$. Then*

$$\mathcal{L}^m(f(Z_f)) = 0, \tag{1}$$

where $\mathcal{L}^m$ denotes the m-dimensional Lebesgue measure and Z_f denotes the set of critical points of f, i.e., $Z_f = \{x \in \mathbb{R}^n : \operatorname{rank} \nabla f(x) < m\}$.

The order of smoothness in the assumptions of this theorem is sharp on the scale C^j (see, e.g., [21]). However, some analogs of the Morse–Sard theorem remain valid for functions lacking the required smoothness in the classical theorem. Although (1) may be no longer valid then, Dubovitskiĭ [9] obtained some results on the structure of level sets in the case of reduced smoothness (also see [4]).

The second author was supported by the Russian Foundation for Basic Research (project No. 12-01-00390-a) and by the Integration Project SB-FEB RAS (project No. 56). The third author was supported by the EPSRC Science and Innovation award to the Oxford Centre for Nonlinear PDE (EP/E035027/1).

Another direction of the research was the generalization of the Morse–Sard theorem to functions in more refined scales of spaces, and especially in Hölder and Sobolev spaces (for example, see [2, 4, 6, 7, 12, 16]). In particular, De Pascale ([7], see also [12]) proved that (1) holds when $f \in \mathrm{W}^{k,p}(\mathbb{R}^n, \mathbb{R}^m)$ with $p > n$, $k \geq \max(n - m + 1, 2)$. Note that in this case v is C^1-smooth by virtue of the Sobolev imbedding theorem, and so the critical set is defined as usual.

For a historical review for the plane case $n = 2$, $m = 1$ see for instance [5]. We mention only the paper [17] where it was proved that (1) holds for Lipschitz functions f of class $\mathrm{BV}_2(\mathbb{R}^2)$, where $\mathrm{BV}_2(\mathbb{R}^2)$ is the space of functions $f \in \mathrm{L}^1(\mathbb{R}^2)$ such that all its partial (distributional) derivatives of the second order are $\mathbb{R}$-valued Radon measures on $\mathbb{R}^2$.

In this paper we consider the case of $\mathbb{R}$-valued Sobolev functions $v \in \mathrm{W}^{n,1}(\mathbb{R}^n)$. It is known (see, e.g., [8]) that such a function admits a continuous representative which is (Fréchet-)differentiable $\mathscr{H}^1$-almost everywhere. The critical set Z_v is defined as the set of points x, where v is differentiable with total (Fréchet-)differential $v'(x) = 0$. As our main result we prove that $\mathscr{L}^1(v(Z_v)) = 0$ (see Theorem 4.1).

Also we show that for any $v \in \mathrm{W}^{n,1}(\mathbb{R}^n)$ and $\varepsilon > 0$ there exists $\delta > 0$ such that for all subsets $E \subset \mathbb{R}^n$ with $\mathscr{H}^1_\infty(E) < \delta$ we have $\mathscr{L}^1(v(E)) < \varepsilon$, where $\mathscr{H}^1_\infty$ is the Hausdorff content. In particular, it follows that $\mathscr{L}^1(v(E)) = 0$ whenever $\mathscr{H}^1(E) = 0$ (see Theorem 2.1). So the image of the exceptional "bad" set, where the differential is not defined, has zero Lebesgue measure. This ties nicely with our definition of the critical set and our version of the Morse–Sard result.

Finally, using these results we prove that almost all level sets of $\mathrm{W}^{n,1}$-functions defined on $\mathbb{R}^n$ are finite disjoint unions of C^1-smooth compact manifolds of dimension $n - 1$ without boundary (see Theorem 5.3).

The proof of the last result relies in turn on new Luzin-type approximation results for $\mathrm{W}^{l,1}$ Sobolev functions by C^k-functions, $k \leq l$, where the exceptional sets are of small Hausdorff content (see Theorem 3.1). The L^p analogs of such results are well known when $p > 1$, see, e.g., [3, 20, 23], where Bessel and Riesz capacities are used instead of Hausdorff content. In fact, the exceptional set can be precisely characterized in terms of the Bessel and Riesz capacities when $f \in \mathrm{W}^{l,p}(\mathbb{R}^n)$ and $p > 1$.

We extend our results also to the space $\mathrm{BV}_n(\mathbb{R}^n)$ consisting of functions $v \in \mathrm{L}^1(\mathbb{R}^n)$ such that all its partial (distributional) derivatives of the n-th order are $\mathbb{R}$-valued Radon measures on $\mathbb{R}^n$ (see Section 6).

For the plane case $n = 2$ these results were obtained in [5] by different methods that do not easily extend to the multidimensional case $n > 2$ that is the main focus here.

Our proofs rely on the results of [14] on advanced versions of Sobolev imbedding theorems (see Theorem 1.3), of [1] on Choquet integrals of Hardy–Littlewood maximal functions with respect to Hausdorff content (see Theorem 1.5), and of [22] on the entropy estimate of near-critical values of differentiable functions (see Theorem 1.6). The key step in the proof of the assertion of the Morse–Sard theorem is contained in Lemma 4.2.

1. Preliminaries

By an *n-dimensional interval* we mean a closed cube $I = [a, b]^n \subset \mathbb{R}^n$ with sides parallel to the coordinate axes. Furthermore we write $\ell(I) = b - a$ for its sidelength.

We denote by $\mathcal{L}^n(F)$ the outer Lebesgue measure of a set $F \subset \mathbb{R}^n$. Denote by $\mathcal{H}^k$, $\mathcal{H}^k_\infty$ the k-dimensional Hausdorff measure, Hausdorff content, respectively: for any $F \subset \mathbb{R}^n$,

$$\mathcal{H}^k(F) = \lim_{\alpha \searrow 0} \mathcal{H}^k_\alpha(F) = \sup_{\alpha > 0} \mathcal{H}^k_\alpha(F),$$

where for each $0 < \alpha \le \infty$,

$$\mathcal{H}^k_\alpha(F) = \inf \Big\{ \sum_{i=1}^{\infty} (\operatorname{diam} F_i)^k : \operatorname{diam} F_i \le \alpha,\ F \subset \bigcup_{i=1}^{\infty} F_i \Big\}.$$

It is well known that $\mathcal{H}^n(F) \sim \mathcal{H}^n_\infty(F) \sim \mathcal{L}^n(F)$ for sets $F \subset \mathbb{R}^n$.

To simplify the notation, we write $\|f\|_{\mathrm{L}^1}$ instead of $\|f\|_{\mathrm{L}^1(\mathbb{R}^n)}$, etc.

The space $\mathrm{W}^{k,1}(\mathbb{R}^n)$ is as usual defined as consisting of those functions $f \in \mathrm{L}^1(\Omega)$ whose distributional partial derivatives of order $l \le k$ belong to $\mathrm{L}^1(\mathbb{R}^n)$ (for detailed definitions and differentiability properties of such functions see, e.g., [8, 10, 23]). Denote by $\nabla^k f$ the vector-valued function consisting of all k-th order partial derivatives of f arranged in some fixed order. We use the norm

$$\|f\|_{\mathrm{W}^{k,1}} = \|f\|_{\mathrm{L}^1} + \|\nabla f\|_{\mathrm{L}^1} + \dots + \|\nabla^k f\|_{\mathrm{L}^1}.$$

Working with Sobolev functions we always assume that the precise representatives are chosen. If $w \in L^1_{\mathrm{loc}}(\Omega)$, then the precise representative w^* is defined by

$$w^*(x) = \begin{cases} \lim\limits_{r \to 0} \fint_{B(x,r)} w(z)\,\mathrm{d}z, & \text{if the limit exists and is finite,} \\ 0 & \text{otherwise,} \end{cases}$$

where the dashed integral as usual denotes the integral mean,

$$\fint_{B(x,r)} w(z) dz = \frac{1}{\mathcal{L}^n(B(x,r))} \int_{B(x,r)} w(z)\,\mathrm{d}z,$$

and $B(x,r) = \{y : |y - x| < r\}$ is the open ball of radius r centered at x.

The following well-known assertion follows immediately from the definition of Sobolev spaces.

Lemma 1.1. *Let $f \in \mathrm{W}^{l,1}(\mathbb{R}^n)$. Then for any $\varepsilon > 0$ there exist functions $f_0 \in C_0^\infty(\mathbb{R}^n)$, $f_1 \in \mathrm{W}^{l,1}(\mathbb{R}^n)$, such that $f = f_0 + f_1$ and $\|f_1\|_{\mathrm{W}^{l,1}} < \varepsilon$.*

We need a version of the Sobolev Embedding Theorem that gives inclusions in Lebesgue spaces with respect to suitably general positive measures. Very general and precise statements are known, but here we restrict attention to the following class of measures:

Definition 1.2. Let μ be a positive measure on $\mathbb{R}^n$. We say that μ has property $(* - l)$ for some $l \le n$, if

$$\mu(I) \le \ell(I)^{n-l}$$

for any n-dimensional interval $I \subset \mathbb{R}^n$.

Theorem 1.3 (see [14, §1.4.3]). *If* $f \in \mathrm{W}^{l,1}(\mathbb{R}^n)$ *and* μ *has property* $(*-l)$, *then*

$$\int |f| d\mu \leq C \|\nabla^l f\|_{\mathrm{L}^1}, \tag{2}$$

where C *does not depend on* μ, f.

For a function $u \in \mathrm{L}^1(I)$, $I \subset \mathbb{R}^n$, define the polynomial $P_{I,k}[u]$ of degree at most k by the following rule:

$$\int_I y^\alpha \big(u(y) - P_{I,k}[u](y)\big)\,\mathrm{d}y = 0 \tag{3}$$

for any multi-index $\alpha = (\alpha_1, \dots, \alpha_n)$ of length $|\alpha| = \alpha_1 + \dots + \alpha_n \leq k$.

We will often use the following simple technical assertion.

Lemma 1.4. *Suppose* $v \in \mathrm{W}^{n,1}(\mathbb{R}^n)$. *Then* v *is a continuous function and for any* $k = 0, \dots, n-1$ *and for any* n*-dimensional interval* $I \subset \mathbb{R}^n$ *the estimate*

$$\sup_{y \in I} |v(y) - P_{I,k}[v](y)| \leq C\Big(\frac{\|\nabla^{k+1} v\|_{\mathrm{L}^1(I)}}{\ell(I)^{n-k-1}} + \|\nabla^n v\|_{\mathrm{L}^1(I)}\Big) \tag{4}$$

holds, where C *depends on* n *only. Moreover, the function* $v_{I,k}(y) = v(y) - P_{I,k}[v](y)$, $y \in I$, *can be extended from* I *to the whole of* $\mathbb{R}^n$ *such that* $v_{I,k} \in \mathrm{W}^{n,1}(\mathbb{R}^n)$ *and*

$$\|\nabla^n v_{I,k}\|_{\mathrm{L}^1(\mathbb{R}^n)} \leq C_0 R(I,k), \tag{5}$$

where C_0 *also depends on* n *only and* $R(I,k)$ *denotes the right-hand side of the estimates* (4) *(in brackets).*

Proof. The existence of a continuous representative for v follows from [14, §1.4.5, Remark 2]. Because of coordinate invariance it is sufficient to prove the estimate (4)–(5) for the case when I is a unit cube: $I = [0,1]^n$. By results of [14, §1.1.15] for any $u \in \mathrm{W}^{n,1}(I)$ the estimates

$$\sup_{y \in I} |u(y)| \leq c\|u\|_{\mathrm{W}^{n,1}(I)} \leq c\big(\|P_{I,k}[u]\|_{\mathrm{L}^1(I)} + \|\nabla^{k+1} u\|_{\mathrm{L}^1(I)} + \|\nabla^n u\|_{\mathrm{L}^1(I)}\big) \tag{6}$$

hold, where $c = c(n,k)$ is a constant. Taking $u(y) = v(y) - P_{I,k}[v](y)$, the first term on the right-hand side of (6) vanishes and so the inequality (6) turns to the estimates (4)–(5) (here we used also the following fact: every function $u \in \mathrm{W}^{n,1}(I)$ can be extended to a function $u \in \mathrm{W}^{n,1}(\mathbb{R}^n)$ such that the estimate $\|\nabla^n u\|_{\mathrm{L}^1(\mathbb{R}^n)} \leq c\|u\|_{\mathrm{W}^{n,1}(I)}$ holds, see [14, §1.1.15]). □

The following two results are crucial for our proof.

Theorem 1.5 ([1]). *If* $f \in \mathrm{W}^{k,1}(\mathbb{R}^n)$, *where* $k \in \{1, \dots, n-1\}$, *then*

$$\int_0^\infty \mathcal{H}^{n-k}_\infty(\{x \in \mathbb{R}^n : \mathcal{M}f(x) \geq \lambda\})\,\mathrm{d}\lambda \leq C \int_{\mathbb{R}^n} |\nabla^k f(y)|\,\mathrm{d}y,$$

where C depends on n, k only and

$$\mathcal{M} f(x) = \sup_{r>0} r^{-n} \int_{B(x,r)} |f(y)| \, \mathrm{d}y$$

is the usual Hardy–Littlewood maximal function of f.

Theorem 1.6 ([22]). *For $A \subset \mathbb{R}^m$ and $\varepsilon > 0$ let $\operatorname{Ent}(\varepsilon, A)$ denote the minimal number of balls of radius ε covering A. Then for any polynomial $P : \mathbb{R}^n \to \mathbb{R}$ of degree at most k, for each ball $B \subset \mathbb{R}^n$ of radius $r > 0$, and any number $\varepsilon > 0$ the estimate*

$$\operatorname{Ent}(\varepsilon r, \{P(x) : x \in B, \ |\nabla P(x)| \leq \varepsilon\}) \leq C_*$$

holds, where C_* depends on n, k only.

To apply Theorem 1.5, we need also the following simple estimate and its corollary.

Lemma 1.7 (see [8, Lemma 2]). *Let $u \in \mathrm{W}^{1,1}(\mathbb{R}^n)$. Then for any ball $B(z,r) \subset \mathbb{R}^n$, $B(z,r) \ni x$, the estimate*

$$\left| u(x) - \fint_{B(z,r)} u(y) \, \mathrm{d}y \right| \leq C r (\mathcal{M} \nabla u)(x)$$

holds, where C depends on n only and $\mathcal{M} \nabla u$ is a Hardy–Littlewood maximal function of ∇u.

Corollary 1.8. *Let $u \in \mathrm{W}^{1,1}(\mathbb{R}^n)$. Then for any ball $B \subset \mathbb{R}^n$ of a radius $r > 0$ and for any number $\varepsilon > 0$ the estimate*

$$\operatorname{diam}(\{u(x) : x \in B, \ (\mathcal{M} \nabla u)(x) \leq \varepsilon\}) \leq C_{**} \varepsilon r$$

holds, where C_{**} is a constant depending on n only.

We will use the following k-order analog of Lemma 1.7.

Lemma 1.9 (see [8, Lemma 2]). *Let $u \in \mathrm{W}^{k,1}(\mathbb{R}^n)$, $k \leq n$. Then for any n-dimensional interval $I \subset \mathbb{R}^n$, $x \in I$, and for any $m = 0, 1, \ldots, k-1$ the estimate*

$$\left| \nabla^m u(x) - \nabla^m P_{I,k-1}[u](x) \right| \leq C \ell(I)^{k-m} (\mathcal{M} \nabla^k u)(x)$$

holds, where the constant C depends on n, k only.

2. On images of sets of small Hausdorff contents

The main result of this section is the following Luzin N-property for $\mathrm{W}^{n,1}$-functions:

Theorem 2.1. *Let $v \in \mathrm{W}^{n,1}(\mathbb{R}^n)$. Then for each $\varepsilon > 0$ there exists $\delta > 0$ such that for any set $E \subset \mathbb{R}^n$ if $\mathcal{H}^1_\infty(E) < \delta$, then $\mathcal{H}^1(v(E)) < \varepsilon$. In particular, $\mathcal{H}^1(v(E)) = 0$ whenever $\mathcal{H}^1(E) = 0$.*

For the plane case, $n = 2$, Theorem 2.1 was obtained in the paper [5].

For the remainder of this section we fix a function $v \in \mathrm{W}^{n,1}(\mathbb{R}^n)$. To prove Theorem 2.1, we need some preliminary lemmas that we turn to next.

By *a dyadic interval* we understand an interval of the form $[\frac{k_1}{2^m}, \frac{k_1+1}{2^m}] \times \cdots \times [\frac{k_n}{2^m}, \frac{k_n+1}{2^m}]$, where k_i, m are integers. The following assertion is straightforward, and hence we omit its proof here.

Lemma 2.2. *For any n-dimensional interval $I \subset \mathbb{R}^n$ there exist dyadic intervals $Q_1, \dots, Q_{2^n}$ such that $I \subset Q_1 \cup \dots \cup Q_{2^n}$ and $\ell(Q_1) = \dots = \ell(Q_{2^n}) \le 2\ell(I)$.*

Let $\{I_\alpha\}_{\alpha \in A}$ be a family of n-dimensional dyadic intervals. We say that the family $\{I_\alpha\}$ is k*-regular*, if for any n-dimensional dyadic interval Q the following estimate holds:

$$\ell(Q)^k \ge \sum_{\alpha : I_\alpha \subset Q} \ell(I_\alpha)^k. \tag{7}$$

The next two assertions are the multidimensional analogs of the corresponding plane results from the paper [5].

Lemma 2.3. *Let $k \in \{1, \dots, n\}$ and let I_α be a family of n-dimensional dyadic intervals. Then there exists a k-regular family J_β of n-dimensional dyadic intervals such that $\bigcup_\alpha I_\alpha \subset \bigcup_\beta J_\beta$ and*

$$\sum_\beta \ell(J_\beta)^k \le \sum_\alpha \ell(I_\alpha)^k.$$

Proof. Define

$$\mathcal{F} = \Big\{ J : J \subset \mathbb{R}^n \text{ dyadic interval; } \sum_{I_\alpha \subset J} \ell(I_\alpha)^k \ge \ell(J)^k \Big\}.$$

Thus $I_\alpha \in \mathcal{F}$ for each α. Denote by $\mathcal{F}^* = \{J_\beta\}$ the collection of maximal elements of $\mathcal{F}$. Clearly

$$\bigcup_\alpha I_\alpha \subset \bigcup_\beta J_\beta,$$

and since dyadic intervals are either disjoint or contained in one another, the $\{J_\beta\}$ are mutually disjoint[1)]. It follows that

$$\sum_\beta \ell(J_\beta)^k \le \sum_\beta \sum_{I_\alpha \subset J_\beta} \ell(I_\alpha)^k \le \sum_\alpha \ell(I_\alpha)^k.$$

Observe also that for any dyadic interval $Q \subset \mathbb{R}^n$,

$$\sum_{J_\beta \subset Q} \ell(J_\beta)^k \le \ell(Q)^k.$$

1) By *disjoint dyadic intervals* we mean intervals with disjoint interiors.

Indeed, if $J_\beta \subset Q$ for some β, then clearly either $J_\beta = Q$ or $J_\beta \neq Q$. In the first case the estimate is evident, and in the second case we deduce from maximality of J_β that $Q \notin \mathcal{F}$, and hence that

$$\sum_{J_\beta \subset Q} \ell(J_\beta)^k \leq \sum_{I_\alpha \subset Q} \ell(I_\alpha)^k < \ell(Q)^k. \qquad \square$$

Lemma 2.4. *Let $k = 0, \dots, n-1$. Then for each $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon, v, k) > 0$ such that for any $(k+1)$-regular family $I_\alpha \subset \mathbb{R}^n$ of n-dimensional dyadic intervals we have if $\sum_\alpha \ell(I_\alpha)^{k+1} < \delta$, then $\sum_\alpha R(I_\alpha, k) < \varepsilon$.*

Proof. Fix $\varepsilon > 0$ and let $I_\alpha \subset \mathbb{R}^n$ be a $(k+1)$-regular family of n-dimensional dyadic intervals with $\sum_\alpha \ell(I_\alpha)^{k+1} < \delta$, where $\delta > 0$ will be specified below. By virtue of Lemma 1.1 we can find a decomposition $v = v_0 + v_1$, where $\|\nabla^j v_0\|_{\mathrm{L}^\infty} \leq K = K(\varepsilon, v)$ for all $j = 0, 1, \dots, n$ and

$$\|\nabla^n v_1\|_{\mathrm{L}^1} < \varepsilon. \tag{8}$$

Assume that

$$\sum_\alpha \ell(I_\alpha)^{k+1} < \delta < \tfrac{1}{K+1}\varepsilon. \tag{9}$$

Define the measure μ by

$$\mu = \Big(\sum_\alpha \frac{1}{\ell(I_\alpha)^{n-k-1}} 1_{I_\alpha}\Big)\mathscr{L}^n,$$

where 1_{I_α} denotes the indicator function of the set I_α.

Claim. $\frac{1}{2^{n+k+2}}\mu$ has property $(* - (k+1))$.

Indeed, write for a dyadic interval Q

$$\mu(Q) = \sum_{I_\alpha \subset Q} \ell(I_\alpha)^{k+1} + \sum_{Q \subset I_\alpha} \frac{\ell(Q)^n}{\ell(I_\alpha)^{n-k-1}} \leq 2\ell(Q)^{k+1},$$

where we invoked (7) and the fact that $Q \subset I_\alpha$ for at most one α. Then for any interval I we have the estimate $\mu(I) \leq 2^{n+k+2}\ell(I)^{k+1}$ (see Lemma 2.2). This proves the claim.

Now return to the estimate of $\sum_\alpha R(I_\alpha, k)$. In addition to (9) we now decrease $\delta > 0$ further such that

$$\sum_\alpha \|\nabla^n v\|_{\mathrm{L}^1(I_\alpha)} < \tfrac{\varepsilon}{2}.$$

By definition of $R(I, k)$ (see Lemma 1.4) and properties (8) and (2) (applied to $f = \nabla^{k+1} v_1$, $l = n-k-1$), we have

$$\begin{aligned}
\sum_\alpha R(I_\alpha, k) &= \sum_\alpha \|\nabla^n v\|_{\mathrm{L}^1(I_\alpha)} + \sum_\alpha \frac{1}{\ell(I_\alpha)^{n-k-1}} \int_{I_\alpha} |\nabla^{k+1} v| \\
&\leq \tfrac{\varepsilon}{2} + \tfrac{K}{K+1}\varepsilon + \sum_\alpha \frac{1}{\ell(I_\alpha)^{n-k-1}} \int_{I_\alpha} |\nabla^{k+1} v_1| \\
&= C'\varepsilon + C \int |\nabla^{k+1} v_1| \,\mathrm{d}\mu \leq C''\varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, the proof of Lemma 2.4 is complete. $\square$

Proof of Theorem 2.1. We have an obvious estimate $\operatorname{diam} v(I) \le CR(I,0)$ for any n-dimensional interval $I \subset \mathbb{R}^n$ (see Lemma 1.4). Fix $\varepsilon > 0$ and take $\delta = \delta(\varepsilon)$ from Theorem 2.4 for $k = 0$, that is, for any 1-regular family $I_\alpha \subset \mathbb{R}^n$ of n-dimensional dyadic intervals, if $\sum_\alpha \ell(I_\alpha) < \delta$, then $\sum_\alpha R(I_\alpha, 0) < \varepsilon$, consequently, $\sum_\alpha \operatorname{diam} v(I_\alpha) < C\varepsilon$. Now the assertion of Theorem 2.1 follows from Lemmas 2.2 and 2.3 (by these lemmas, there exists $\delta_1 > 0$ such that if $\mathcal{H}^1_\infty(E) < \delta_1$, then E can be covered by a 1-regular family $I_\alpha \subset \mathbb{R}^n$ of n-dimensional dyadic intervals with $\sum_\alpha \ell(I_\alpha) < \delta$). □

3. On approximation of $\mathrm{W}^{l,1}$ Sobolev functions

Theorem 3.1. *Let $k, l \in \{1, \dots, n\}$, $k \le l$. Then for any $f \in \mathrm{W}^{l,1}(\mathbb{R}^n)$ and for each $\varepsilon > 0$ there exist an open set $U \subset \mathbb{R}^n$ and a function $g \in C^k(\mathbb{R}^n)$ such that $\mathcal{H}^{n-l+k}_\infty(U) < \varepsilon$ and $f \equiv g$, $\nabla^m f \equiv \nabla^m g$ on $\mathbb{R}^n \setminus U$ for $m = 1, \dots, k$.*

The proof of Theorem 3.1 is based on the results of [1, 8] and on the classical Whitney Extension Theorem:

Theorem 3.2. *Let $k \in \mathbb{N}$ and let $f = f_0, f_\alpha$ be a finite family of functions defined on the closed set $E \subset \mathbb{R}^n$, where α ranges over all multi-indices $\alpha = (\alpha_1, \dots, \alpha_n)$ with $|\alpha| = \alpha_1 + \dots + \alpha_n \le k$. For $x, y \in E$ and a multi-index α, $|\alpha| \le k$, put*

$$T_\alpha(x; y) = \sum_{|\beta| \le k - |\alpha|} \frac{1}{\beta!} f_{\alpha+\beta}(x) \cdot (y - x)^\beta, \quad R_\alpha(x; y) = f_\alpha(y) - T_\alpha(x; y).$$

Suppose that there exists a function $\omega\colon [0, +\infty) \to [0, +\infty)$ such that $\omega(t) \to 0$ as $t \searrow 0$ and for each multi-index α, $|\alpha| \le k$, and for all $x, y \in E$ the estimate

$$|R_\alpha(x; y)| \le \omega(|x - y|)|x - y|^{k-|\alpha|}$$

holds. Then there exists a function $g \in C^k(\mathbb{R}^n)$ such that $f \equiv g$, $f_\alpha \equiv \partial^\alpha g$ on E for $|\alpha| \le k$.

Proof of Theorem 3.1. Let the assumptions of Theorem 3.1 be fulfilled. For the case $k = l$ the assertion of the theorem is well known (see, e.g., [3, 13, 23]).

Now fix $k < l$. Then the gradients $\nabla^m f(x)$, $m \le k$, are well-defined for all $x \in \mathbb{R}^n \setminus A_k$, where $\mathcal{H}^{n-l+k}(A_k) = 0$ (see [8]). For a multi-index α with $|\alpha| \le k$ denote by $T_\alpha(f, x; y)$ the Taylor polynomial of order at most $k - |\alpha|$ for the function $\partial^\alpha f$ with the center at x:

$$T_\alpha(f, x; y) = \sum_{|\beta| \le k - |\alpha|} \frac{1}{\beta!} \partial^{\alpha+\beta} f(x) \cdot (y - x)^\beta.$$

By virtue of the Whitney Extension Theorem 3.2, we finish the proof of Theorem 3.1 by checking that for each multi-index α with $|\alpha| \le k$ the corresponding Taylor remainder term satisfies the estimate $\partial^\alpha f(y) - T_\alpha(f, x; y) = o(|x - y|^{k-|\alpha|})$ uniformly for $x, y \in \mathbb{R}^n \setminus U$, where $\mathcal{H}^{n-l+k}_\infty(U)$ is small.

Take a sequence $f_i \in C_0^\infty(\mathbb{R}^n)$ such that

$$\|\nabla^l f_i - \nabla^l f\|_{\mathrm{L}^1} < 4^{-i}.$$

Denote $\tilde{f}_i = f - f_i$. Put

$$B_i = \{x \in \mathbb{R}^n : (\mathcal{M}\nabla^k \tilde{f}_i)(x) > 2^{-i}\}, \quad G_i = A_k \cup \Big(\bigcup_{j=i}^{\infty} B_j\Big).$$

Then by Theorem 1.5 we have

$$\mathcal{H}_{\infty}^{n-l+k}(B_i) \le c\, 2^{-i},$$

and consequently,

$$\mathcal{H}_{\infty}^{n-l+k}(G_i) < C\, 2^{-i}.$$

By construction,

$$|\nabla^k \tilde{f}_j(x)| \le 2^{-j} \tag{10}$$

for all $x \in \mathbb{R}^n \setminus G_i$ and all $j \ge i$. For a multi-index α with $|\alpha| \le k-1$ denote by $T_{\alpha,k-1}(f,x;y)$ the Taylor polynomial of order $k-1-|\alpha|$ for the function $\partial^\alpha f$ with the center at x:

$$T_{\alpha,k-1}(f,x;y) = \sum_{|\beta|\le k-1-|\alpha|} \frac{1}{\beta!}\partial^{\alpha+\beta} f(x)\cdot(y-x)^\beta.$$

In our notation,

$$T_\alpha(f,x;y) = T_{\alpha,k-1}(f,x;y) + \sum_{|\beta|=k-|\alpha|} \frac{1}{\beta!}\partial^{\alpha+\beta} f(x)\cdot(y-x)^\beta.$$

We start by estimating the remainder term $\partial^\alpha \tilde{f}_j(y) - T_{\alpha,k-1}(\tilde{f}_j,x;y)$ for a multi-index α with $|\alpha| \le k-1$. Fix $x, y \in \mathbb{R}^n \setminus G_i$, $j \ge i$, and an n-dimensional interval I such that $x, y \in I$, $|x-y| \sim \ell(I)$. By construction and Lemma 1.9,

$$\big|\partial^\alpha \tilde{f}_j(y) - \partial^\alpha P_{I,k-1}[\tilde{f}_j](y)\big| \le C\,\ell(I)^{k-|\alpha|}(\mathcal{M}\nabla^k \tilde{f}_j)(y) \le C|x-y|^{k-|\alpha|}2^{-j}.$$

For the same reasons we find for any multi-index β with $|\beta| \le k-1-|\alpha|$ that

$$\begin{aligned}\big|\partial^{\alpha+\beta} \tilde{f}_j(x) - \partial^{\alpha+\beta} P_{I,k-1}[\tilde{f}_j](x)\big| &\le C\,\ell(I)^{k-|\alpha|-|\beta|}(\mathcal{M}\nabla^k \tilde{f}_j)(x)\\ &\le C|x-y|^{k-|\alpha|-|\beta|}2^{-j}.\end{aligned}$$

Consequently,

$$\begin{aligned}&\big|\partial^\alpha \tilde{f}_j(y) - T_{\alpha,k-1}(\tilde{f}_j,x;y)\big|\\ &\quad\le \big|\partial^\alpha \tilde{f}_j(y) - \partial^\alpha P_{I,k-1}[\tilde{f}_j](y)\big| + \big|\partial^\alpha P_{I,k-1}[\tilde{f}_j](y) - T_{\alpha,k-1}(\tilde{f}_j,x;y)\big|\\ &\quad\le C|x-y|^{k-|\alpha|}2^{-j}\\ &\qquad + \sum_{|\beta|\le k-1-|\alpha|}\frac{1}{\beta!}\big|\big(\partial^{\alpha+\beta}\tilde{f}_j(x) - \partial^{\alpha+\beta}P_{I,k-1}[\tilde{f}_j](x)\big)\cdot(y-x)^\beta\big|\\ &\quad\le C_1|x-y|^{k-|\alpha|}2^{-j}.\end{aligned}$$

Finally from the last estimate and (10) we have

$$\begin{aligned}\left|\partial^\alpha f(y) - T_\alpha(f, x; y)\right| &\le \left|\partial^\alpha \tilde{f}_j(y) - T_\alpha(\tilde{f}_j, x; y)\right| + \left|\partial^\alpha f_j(y) - T_\alpha(f_j, x; y)\right| \\ &\le \left|\partial^\alpha \tilde{f}_j(y) - T_{\alpha,k-1}(\tilde{f}_j, x; y)\right| + |\nabla^k \tilde{f}_j(x)| \cdot |x-y|^{k-|\alpha|} \\ &\qquad + \omega_{f_j}(|x-y|) \cdot |x-y|^{k-|\alpha|} \\ &\le C'|x-y|^{k-|\alpha|} 2^{-j} + \omega_{f_j}(|x-y|) \cdot |x-y|^{k-|\alpha|} \\ &= \left(C' 2^{-j} + \omega_{f_j}(|x-y|)\right) \cdot |x-y|^{k-|\alpha|},\end{aligned}$$

where $\omega_{f_j}(r) \to 0$ as $r \to 0$ (the latter holds since $f_j \in C_0^\infty(\mathbb{R}^n)$). We emphasize that the last inequality is valid for all $j \ge i$ and $x, y \in \mathbb{R}^n \setminus G_i$. Take an open set $U_i \supset G_i$ such that

$$\mathcal{H}^{n-l+k}_\infty(U_i) < C 2^{-i}.$$

Put $E_i = \mathbb{R}^n \setminus U_i$. Then by construction

$$\left|\partial^\alpha f(y) - T_\alpha(f, x; y)\right| \le \left(C' 2^{-j} + \omega_{f_j}(|x-y|)\right) \cdot |x-y|^{k-|\alpha|}$$

for all $j \ge i$, $|\alpha| \le k$, and $x, y \in E_i$. Then the assumptions of the Whitney Extension Theorem 3.2 are fulfilled, and hence the proof of Theorem 3.1 is complete. □

Remark 3.3. Using the extension formula and the methods from the proof of Theorem 6.2 (see Section 6 below; this approach was originally introduced in [3]), one can prove that for $k < l$ the function g from the assertion of Theorem 3.1 can be constructed such that in addition the estimate $\|f - g\|_{\mathrm{W}^{k+1,1}} < \varepsilon$ holds.

4. Morse–Sard theorem in $\mathrm{W}^{n,1}(\mathbb{R}^n)$

Recall that if $v \in \mathrm{W}^{n,1}(\mathbb{R}^n)$ and $k = 1, \dots, n$, then $\nabla^k v(x)$ is well-defined for $\mathcal{H}^k$-almost all $x \in \mathbb{R}^n$ (see [8]). In particular, v is differentiable (in the classical Fréchet sense) and the classical derivative coincides with $\nabla v(x) = \lim_{r\to 0} ⨍_{B(x,r)} \nabla v(z)\,\mathrm{d}z$ at all points $x \in \mathbb{R}^n \setminus A_v$, where $\mathcal{H}^1(A_v) = 0$. Consequently, in view of Theorem 2.1, $\mathcal{H}^1(v(A_v)) = 0$.

Denote $Z_v = \{x \in \mathbb{R}^n \setminus A_v : \nabla v(x) = 0\}$. The main result of the section is as follows:

Theorem 4.1. *If $v \in \mathrm{W}^{n,1}(\mathbb{R}^n)$, then $\mathcal{H}^1(v(Z_v)) = 0$.*

For the remainder of the section we fix a function $v \in \mathrm{W}^{n,1}(\mathbb{R}^n)$.

The key point of the proof is contained in the following lemma.

Lemma 4.2. *For any n-dimensional dyadic interval $I \subset \mathbb{R}^n$ the estimate*

$$\mathcal{H}^1(v(Z_v \cap I)) \le C \|\nabla^n v\|_{\mathrm{L}^1(I)}$$

holds, where C depends on n only.

Proof. Note that by formula (5) it is sufficient to prove the estimate

$$\mathcal{H}^1(v(Z_v \cap I)) \le C \|\nabla^n v_{I,n-1}\|_{\mathrm{L}^1(\mathbb{R}^n)},$$

where the function $v_{I,n-1}$ was defined in Lemma 1.4.

Fix an n-dimensional dyadic interval $I \subset \mathbb{R}^n$. To simplify the notation, we will write v_I and P_I instead of $v_{I,n-1}$ and $P_{I,n-1}[v]$ respectively. In particular, $v_I(x) = v(x) - P_I(x)$ for all $x \in I$. Denote

$$\sigma = \|\nabla^n v_I\|_{\mathrm{L}^1(\mathbb{R}^n)}, \quad E_j = \{x \in \mathbb{R}^n : (\mathcal{M}\nabla v_I)(x) \in (2^{j-1}, 2^j]\}, \quad j \in \mathbb{Z}.$$

Denote also $\delta_j = \mathcal{H}^1_\infty(E_j)$. Then by Theorem 1.5,

$$\sum_{j=-\infty}^{\infty} \delta_j 2^j \le C_1 \sigma,$$

where C_1 depends on n only. By construction, for each $j \in \mathbb{Z}$ there exists a family of balls $B_{ij} \subset \mathbb{R}^n$ of radii r_{ij} such that

$$E_j \subset \bigcup_{i=1}^{\infty} B_{ij} \quad \text{and} \quad \sum_{i=1}^{\infty} r_{ij} \le 3\delta_j.$$

Denote

$$Z_{ij} = Z_v \cap I \cap E_j \cap B_{ij} \quad \text{and} \quad Z_\infty = Z_v \cap I \setminus \Big(\bigcup_{i,j} Z_{ij}\Big).$$

By construction $Z_\infty \subset \{x \in \mathbb{R}^n : (\mathcal{M}\nabla v_I)(x) = \infty\}$, so by Theorem 1.5, $\mathcal{H}^1(Z_\infty) = 0$ and hence by Theorem 2.1, $\mathcal{H}^1(v(Z_\infty)) = 0$. Thus it is sufficient to estimate $\mathcal{H}^1(v(Z_{ij}))$.

Since $\nabla P_I(x) = -\nabla v_I(x)$ at each point $x \in Z_v \cap I$, we have by construction for all i, j:

$$Z_{ij} \subset \{x \in B_{ij} : |\nabla P_I(x)| = |\nabla v_I(x)| \le (\mathcal{M}\nabla v_I)(x) \le 2^j\}.$$

Applying Theorem 1.6 and Corollary 1.8 to functions P_I, v_I, respectively, with $B = B_{ij}$ and $\varepsilon = 2^j$, we find a finite family of balls $T_k \subset \mathbb{R}$ each of radius $(1 + C_{**})2^j r_{ij}$, $k = 1, \dots, C_*$, such that

$$\bigcup_{k=1}^{C_*} T_k \supset v(Z_{ij}).$$

Therefore

$$\mathcal{H}^1(v(Z_{ij})) \le 2C_*(1 + C_{**})2^j r_{ij},$$

and consequently,

$$\mathcal{H}^1(v(Z_v \cap I)) \le \sum_{j=-\infty}^{\infty} \sum_{i=1}^{\infty} 2C_*(1 + C_{**})2^j r_{ij} \le 6C_*(1 + C_{**}) \sum_{j=-\infty}^{\infty} 2^j \delta_j \le C'\sigma.$$

The last estimate finishes the proof of the lemma. □

From the last result and the absolute continuity of the Lebesgue integral we infer

Corollary 4.3. *For any $\varepsilon > 0$ there exists $\delta > 0$ such that for any set $E \subset \mathbb{R}^n$ if $\mathcal{H}^n_\infty(E) \le \delta$, then $\mathcal{H}^1(v(Z_v \cap E)) \le \varepsilon$. In particular, $\mathcal{H}^1(v(Z_v \cap E)) = 0$ for any $E \subset \mathbb{R}^n$ with $\mathcal{H}^n_\infty(E) = 0$.*

Because of the classical Morse–Sard theorem for $g \in C^n(\mathbb{R}^n)$, Theorem 3.1 (applied to the case $k = n$) implies

Corollary 4.4. *There exists a set $Z_{0,v}$ of n-dimensional Lebesgue measure zero such that $\mathcal{H}^1(v(Z_v \setminus Z_{0,v})) = 0$. In particular, $\mathcal{H}^1(v(Z_v)) = \mathcal{H}^1(v(Z_{0,v}))$.*

From Corollaries 4.4 and 4.3 we conclude the proof of Theorem 4.1.

5. Application to the level sets of $\mathrm{W}^{n,1}$ functions

Theorem 3.1 for the case $k = 1$ implies

Theorem 5.1. *Let $v \in \mathrm{W}^{n,1}(\mathbb{R}^n)$. Then for any $\varepsilon > 0$ there exist an open set $U \subset \mathbb{R}^n$ and a function $g \in \mathrm{C}^1(\mathbb{R}^n)$ such that $\mathcal{H}^1_\infty(U) < \varepsilon$ and $v \equiv g$, $\nabla v \equiv \nabla g$ on $\mathbb{R}^n \setminus U$.*

If we apply Theorems 2.1 and 4.1 to the last assertion, we obtain

Corollary 5.2. *Let $v \in \mathrm{W}^{n,1}(\mathbb{R}^n)$. Then for any $\varepsilon > 0$ there exist an open set $V \subset \mathbb{R}$ and a function $g \in \mathrm{C}^1(\mathbb{R}^n)$ such that $\mathcal{H}^1(V) < \varepsilon$, $v(A_v) \subset V$ and $v|_{v^{-1}(\mathbb{R}\setminus V)} = g|_{v^{-1}(\mathbb{R}\setminus V)}$, $\nabla v|_{v^{-1}(\mathbb{R}\setminus V)} = \nabla g|_{v^{-1}(\mathbb{R}\setminus V)} \neq 0$.*

Finally we have

Theorem 5.3. *Let $v \in \mathrm{W}^{n,1}(\mathbb{R}^n)$. Then for almost all $y \in \mathbb{R}$ the preimage $v^{-1}(y)$ is a finite disjoint family of $(n-1)$-dimensional C^1-smooth compact manifolds (without boundary) S_j, $j = 1, \dots, N(y)$.*

Proof. The inclusion $v \in \mathrm{W}^{n,1}(\mathbb{R}^n)$ and Lemma 1.4 easily imply the following statement:

(i) For any $\varepsilon > 0$ there exists $R_\varepsilon \in (0, +\infty)$ such that $|v(x)| < \varepsilon$ for all $x \in \mathbb{R}^n \setminus B(0, R_\varepsilon)$.

Fix arbitrary $\varepsilon > 0$. Take the corresponding set V and function $g \in C^1(\mathbb{R}^n)$ from Corollary 5.2. Let $0 \neq y \in v(\mathbb{R}^n) \setminus V$. Denote $F_v = v^{-1}(y)$, $F_g = g^{-1}(y)$. We assert the following properties of these sets.

(ii) F_v is a compact set.

(iii) $F_v \subset F_g$.

(iv) $\nabla v = \nabla g \neq 0$ on F_v.

(v) The function v is differentiable (in the classical sense) at each $x \in F_v$, and the classical derivative coincides with $\nabla v(x) = \lim_{r\to 0} ⨍_{B(x,r)} \nabla v(z)\, dz$.

Indeed, property (ii) follows from (i), properties (iii)–(iv) follow from Corollary 5.2, and property (v) follows from the condition $v(A_v) \subset V$ of Corollary 5.2.

We require one more property of these sets:

(vi) For any $x_0 \in F_v$ there exists $r > 0$ such that $F_v \cap B(x_0, r) = F_g \cap B(x_0, r)$.

Indeed, take any point $x_0 \in F_v$ and suppose the claim (vi) is false. Then there exists a sequence of points $F_g \setminus F_v \ni x_i \to x_0$. Denote by I_x the straight line segment of length r with the center at x parallel to the vector $\nabla v(x_0) = \nabla g(x_0)$. Evidently, for sufficiently small $r > 0$ the equality $I_x \cap F_g = \{x\}$ holds for any $x \in F_g \cap B(x_0, r)$. Then by construction $I_{x_i} \cap F_v = \emptyset$ for sufficiently large i. Hence for sufficiently large i either $v > y$ on I_{x_i} or $v < y$ on I_{x_i}. For definiteness, suppose $v > y$ on I_{x_i} for all $i \in \mathbb{N}$. In the limit we obtain the inequality $v \geq y = v(x_0)$ on I_{x_0}. But the last assertion contradicts (iv)–(v). This contradiction finishes the proof of (vi).

Obviously, (ii)–(vi) imply that each connected component of the set $F_v = v^{-1}(y)$ is a compact $(n-1)$-dimensional C^1-smooth manifold (without boundary). □

6. On the case of BV$_n$ functions

For signed or vector-valued Radon measures μ we denote by $\|\mu\|$ the total variation measure of μ. The space $\mathrm{BV}_k(\mathbb{R}^n)$ is as usual defined as consisting of those functions $f \in \mathrm{W}^{k-1,1}(\mathbb{R}^n)$ whose distributional partial derivatives of order k are Radon measures with $\|D^k f\|(\mathbb{R}^n) < \infty$, where we denote by $D^k f$ the vector-valued measure consisting of all k-order partial derivatives of f (for detailed definitions and differentiability properties of such functions see, e.g., [8, 10, 23]). In particular, the following fact is well known.

Lemma 6.1. *Let $f \in \mathrm{BV}_k(\mathbb{R}^n)$. Then there exists a sequence $f_i \in \mathrm{C}_0^\infty(\mathbb{R}^n)$ such that*

$$\|f_i - f\|_{\mathrm{W}^{k-1,1}} \to 0, \quad \|\nabla^k f_i\|_{\mathrm{L}^1} \leq C, \quad \textit{and} \quad \|\nabla^k f_i\|_{\mathrm{L}^1(U)} \to \|D^k f\|(U)$$

for any open subset $U \subset \mathbb{R}^n$ with $\|D^k f\|(\partial U) = 0$.

The results obtained in the previous sections were established for functions of class $\mathrm{BV}_2(\mathbb{R}^2)$ in [5], hence in the present section we consider only functions of class $\mathrm{BV}_n(\mathbb{R}^n)$ for $n \geq 3$. Recall that in this case $\nabla^k v(x)$ is well-defined for $\mathcal{H}^k$-almost all $x \in \mathbb{R}^n$, $k = 1, \dots, n-2$ (see [8]). In particular, v is differentiable (in the classical Fréchet sense) at all points $x \in \mathbb{R}^n \setminus A_v$, where $\mathcal{H}^1(A_v) = 0$. Denote $Z_v = \{x \in \mathbb{R}^n \setminus A_v : \nabla v(x) = 0\}$. Most of the results from the previous sections remain valid for functions $v \in \mathrm{BV}_n(\mathbb{R}^n)$. More precisely, Theorem 2.1, Lemma 2.4 for $k \leq n-2$, and Theorems 4.1, 5.1, 5.3 are also true in this more general BV context. Except for Theorem 4.1, whose proof we discuss below, the proofs are entirely analogous. Also, the assertion of Approximation Theorem 3.1 remains valid for $f \in \mathrm{BV}_l(\mathbb{R}^n)$, $k, l \in \{1, \dots, n\}$, $k \leq l$, $k \neq l-1$ (for the case $k = l$ it follows immediately from the results of [8] and [13]; the proof for $k \leq l-2$ will be discussed below).

On the other hand, the assertion of Lemma 2.4 for $k = n-1$ is not valid for a general $v \in \mathrm{BV}_n(\mathbb{R}^n)$. Also the assertion of the Approximation Theorem 3.1 is not valid for $f \in \mathrm{BV}_l(\mathbb{R}^n)$ when $k = l-1$.

In order to prove the assertion of the Approximation Theorem 3.1 for $f \in \mathrm{BV}_l(\mathbb{R}^n)$, $k, l \in \{1, \dots, n\}$ when $k \le l-2$, one can repeat the arguments from the proof for the Sobolev case (see Section 3). Proceeding in this manner, one notices that in the first step it is necessary to have a sequence of functions $f_i \in \mathrm{C}^k(\mathbb{R}^n)$ with $\|f - f_i\|_{\mathrm{BV}_l} \to 0$. Such a sequence exists because of the following result.

Theorem 6.2. *Let $f \in \mathrm{BV}_l(\mathbb{R}^n)$, $l \le n$. Then for any $\varepsilon > 0$ there exists a function $g \in \mathrm{BV}_l(\mathbb{R}^n)$ such that*

(i) $\|f - g\|_{\mathrm{BV}_l} < \varepsilon$*;*

(ii) $g \in \mathrm{C}^{l-2,1}(\mathbb{R}^n)$*, i.e., $g \in \mathrm{C}^{l-2}(\mathbb{R}^n)$ and $\nabla^{l-2} g$ is a Lipschitz function;*

(iii) *there exists an open set $U \subset \mathbb{R}^n$ such that $\mathcal{H}^{n-1}_\infty(U) < \varepsilon$ and $f \equiv g$, $\nabla^m f \equiv \nabla^m g$ on $\mathbb{R}^n \setminus U$ for $m = 1, \dots, l-2$.*

Very similar results were proved in [3] for the case of Sobolev functions $f \in \mathrm{W}^{l,p}(\mathbb{R}^n)$ with $p > 1$, and our proof follows the ideas from [3].

To prove Theorem 6.2, we need some preliminary results.

Lemma 6.3. *Let $f \in \mathrm{BV}_l(\mathbb{R}^n)$, $l \le n$. Then for each $\varepsilon > 0$ there exists $\delta > 0$ such that for any open set $U \subset \mathbb{R}^n$ we have that if $\mathcal{H}^{n-1}_\infty(U) < \delta$ then $\|D^l f\|(U) < \varepsilon$.*

Proof. It is an easy consequence of the Coarea Formula, and we leave details to the interested reader (or see [5, Lemma 2.4]). □

Remark 6.4. Using the methods of the proof of [8, Lemma 2], one can prove the following result. Let $u \in \mathrm{BV}_{k+1}(\mathbb{R}^n)$, $k+1 \le n$. Then for any n-dimensional interval $Q \subset \mathbb{R}^n$ and any point $x \in \mathbb{R}^n$ with $\operatorname{dist}(x, Q) \le 9n\ell(Q)$ the estimates

$$\left|\nabla^k P_{Q,k}[u](x)\right| \le C(\mathcal{M}\nabla^k u)(x),$$
$$\left|\nabla^m u(x) - \nabla^m P_{Q,k}[u](x)\right| \le C\ell(Q)^{k-m}(\mathcal{M}\nabla^k u)(x)$$

hold for each $m \in \{0, \dots, k-1\}$, where the constant C depends on n only.

Proof of Theorem 6.2. Fix $\varepsilon \in (0,1)$. Let U be an open set such that

$$\mathcal{H}^{n-1}_\infty(U) < \varepsilon,$$
$$\|D^l f\|(U) < \varepsilon, \tag{11}$$
$$(\mathcal{M}\nabla^{l-1} f)(x) \le C_\varepsilon \quad \text{for all } x \in \mathbb{R}^n \setminus U. \tag{12}$$

The existence of U follows from Lemma 6.3 and Theorem 1.5, that remains valid for $f \in \mathrm{BV}_l(\mathbb{R}^n)$ provided the L^1 norm is replaced by the total variation norm (see [1]). Denote $F = \mathbb{R}^n \setminus U$. Take a Whitney cube decomposition $U = \bigcup_{j=1}^\infty Q_j$, where all cubes Q_j are dyadic, and select an associated smooth partition of unity $\{\varphi_j\}_{j \in \mathbb{N}}$. Recall the standard properties of Q_j, φ_j (see [19, Chapter VI]):

(i) $\operatorname{diam}(Q_j) \le \operatorname{dist}(Q_j, F) \le 4\operatorname{diam}(Q_j) < 1$.

(ii) Every point $x \in U$ is covered by at most $N = (12)^n$ different cubes Q_j^*, where the cube Q_j^* has the same center as Q_j and $\ell(Q_j^*) = \frac{9}{8}\ell(Q_j)$.

(iii) For each $j \in \mathbb{N}$, we have $\operatorname{supp}\varphi_j \subset Q_j^* \subset U$, moreover, $|\nabla^m \varphi_j| \le C_m \ell(Q_j)^{-m}$ for all $m \in \mathbb{N}$.

(iv) All $\varphi_j \ge 0$ and $\sum_{j=1}^\infty \varphi_j(x) \equiv 1$ on U.

Now we define the function $g\colon \mathbb{R}^n \to \mathbb{R}$ by

$$g(x) = \begin{cases} f(x), & \text{if } x \in F, \\ \sum_{j=1}^\infty \varphi_j(x) P_{Q_j^*,l-1}[f](x), & \text{if } x \in U. \end{cases}$$

Recall the following properties of the polynomials $P_{Q_j^*,l-1}[f](x)$ (see [8, p. 1034]):

$$\int_{Q_j^*} \big|\nabla^m f(z) - \nabla^m P_{Q_j^*,l-1}[f](z)\big| \, \mathrm{d}z \le C\ell(Q_j^*)^{l-m} \|D^l f\|(Q_j^*);$$

$$\big\|D^l(f - P_{Q_j^*,l-1}[f])\big\|(Q_j^*) = \|D^l f\|(Q_j^*),$$

where $m \in \{0, 1, \dots, l-1\}$. From these properties and assumption (11) we get by direct calculation for each $m = 0, \dots, l-1$ the estimates

$$\sum_{j=1}^\infty \big\|\nabla^m\big(\varphi_j(f - P_{Q_j^*,l-1}[f])\big)\big\|_{\mathrm{L}^1(Q_j^*)} \le C\|D^l f\|(U) < C\varepsilon.$$

Analogously,

$$\sum_{j=1}^\infty \|D^l\big(\varphi_j(f - P_{Q_j^*,l-1}[f])\big)\|(Q_j^*) \le C\|D^l f\|(U) < C\varepsilon.$$

From the convergence of the above series and from the completeness of the space $\mathrm{BV}_l(\mathbb{R}^n)$ it follows readily that $f - g = \sum_{j=1}^\infty \varphi_j(f - P_{Q_j^*,l-1}[f]) \in \mathrm{BV}_l(\mathbb{R}^n)$. Consequently,

$$g \in \mathrm{BV}_l(\mathbb{R}^n) \quad \text{and} \quad \|f - g\|_{\mathrm{BV}_l} < C\varepsilon.$$

Thus to finish the proof of the theorem, it is sufficient to check that

$$\|\nabla^{l-1} g\|_{\mathrm{L}^\infty} < \infty. \tag{13}$$

From (12) by construction it follows that

$$\operatorname{ess\,sup}_F |\nabla^{l-1} g| = \operatorname{ess\,sup}_F |\nabla^{l-1} f| \le C_\varepsilon.$$

Now estimate $|\nabla^{l-1} g(y)|$ for $y \in U$. Let $y \in Q_{j_0}$. Take $x \in F$ such that

$$\operatorname{dist}(x, Q_{j_0}) = \operatorname{dist}(F, Q_{j_0}).$$

Then $C_0\ell(Q_j^*) \le |y - x| \le C_1\ell(Q_j^*)$ for each $Q_j^* \ni y$. Consider the $(l-2)$-order Taylor polynomial

$$T(f, x; y) = \sum_{|\beta| \le l-2} \frac{1}{\beta!} \partial^\beta f(x) \cdot (y - x)^\beta.$$

From assumption (12) and Remark 6.4 (with $k = l-1$) it follows that for arbitrary multi-index α with $|\alpha| \le l-1$

$$\begin{aligned}
&\left|\partial^\alpha (P_{Q_j^*,l-1}[f](y) - T(f,x;y))\right| \\
&\quad \le \left|\nabla^{l-1} P_{Q_j^*,l-1}[f](x)\right| \cdot |x-y|^{l-1-\alpha} \\
&\qquad + \sum_{|\beta|\le l-2-|\alpha|} \frac{1}{\beta!}\left|\left(\partial^{\alpha+\beta} P_{Q_j^*,l-1}[f](x) - \partial^{\alpha+\beta} f(x)\right)\cdot (y-x)^\beta\right| \\
&\quad \le C_2|x-y|^{l-1-|\alpha|} \le C_3 \ell(Q_j^*)^{l-1-|\alpha|}.
\end{aligned}$$

From the last estimate we have

$$\begin{aligned}
|\nabla^{l-1} g(y)| &= \left|\nabla^{l-1}(g(y) - T(f,x;y))\right| \\
&= \Big|\sum_{j:Q_j^*\ni y} \nabla^{l-1}\big(\varphi_j(y)(P_{Q_j^*,l-1}[f](y) - T(f,x;y))\big)\Big| \\
&\le \sum_{j:Q_j^*\ni y}\sum_{m=0}^{l-1} |\nabla^{l-1-m}\varphi_j(y)| \cdot \left|\nabla^m(P_{Q_j^*,l-1}[f](y) - T(f,x;y))\right| \le C_4,
\end{aligned}$$

where the constant C_4 does not depend on $y \in U$. The last estimate finishes the proof of the target assertion (13). □

Now we discuss the proof of Theorem 4.1 in the BV-case, which is more delicate than the Sobolev case.

The assertion of the key Lemma 4.2 remains valid for $v \in \mathrm{BV}_n(\mathbb{R}^n)$ with identical proof if we replace in its formulation $\|\nabla^n v\|_{\mathrm{L}^1(I)}$ by $\|D^n v\|(I) \sim \|D^n v_{I,n-1}\|(\mathbb{R}^n)$. From the last fact using the standard covering lemmas one can easily deduce the following:

Lemma 6.5. *Let $v \in \mathrm{BV}_n(\mathbb{R}^n)$. Then for each $\varepsilon > 0$ there exists $\delta > 0$ such that for any Borel set $E \subset \mathbb{R}^n$ the estimate $\mathcal{H}^1(v(Z_v \cap E)) \le C\|D^n v\|(E)$ holds, where C does not depend on E, v.*

From this lemma and from Lemma 6.3 we infer easily

Corollary 6.6. *Let $v \in \mathrm{BV}_n(\mathbb{R}^n)$. Then for each $\varepsilon > 0$ there exists $\delta > 0$ such that for any Borel set $E \subset \mathbb{R}^n$ we have that if $\mathcal{H}^{n-1}_\infty(E) < \delta$ then $\mathcal{H}^1(v(Z_v \cap E)) \le \varepsilon$. In particular, $\mathcal{H}^1(v(Z_v \cap E)) = 0$ whenever $\mathcal{H}^{n-1}(E) = 0$.*

We need a more refined version of Lemma 1.4 in the BV case.

Lemma 6.7. *Suppose $v \in \mathrm{BV}_n(\mathbb{R}^n)$ and $S \subset \mathbb{R}^n$ is an $(n-1)$-dimensional C^1-smooth compact manifold (without boundary). Then there exists $\delta = \delta(S) > 0$ such that for any ball $B = B(z,r)$ with $z \in S$ and $r < \delta$ the estimates*

$$\sup_{y\in \bar{B}_+} \left|v(y) - P_{B_+,n-1}[v](y)\right| \le C\|D^n v\|(B_+),$$
$$\sup_{y\in \bar{B}_-} \left|v(y) - P_{B_-,n-1}[v](y)\right| \le C\|D^n v\|(B_-)$$

hold, where C depends on n only, B_+, B_- are the connected components of the open set $B \setminus S$, and the polynomials $P_{B_\pm,n-1}[v]$ are defined by formula (3) *with I replaced by $B_\pm$, respectively. Moreover, each function $v_{B_\pm}(y) = v(y) - P_{B_\pm,n-1}[v](y)$, $y \in B_\pm$, can be extended from $\bar{B}_\pm$ to the whole of $\mathbb{R}^n$ such that $v_{B_\pm} \in \mathrm{BV}_n(\mathbb{R}^n)$ and*

$$\|D^n v_{B_\pm}\|(\mathbb{R}^n) \le C_0 \|D^n v\|(B_\pm),$$

where C_0 also depends on n only.

The proof of this lemma is similar to the proof of Lemma 6.7 with the following addition: we must apply the advanced version of the Sobolev Extension Theorem from bounded Lipschitz domains to the whole of $\mathbb{R}^n$ with the estimate of the norm of the extension operator depending on n and on the Lipschitz constant of the domain only (see [19, Chapter VI, §3.2, Theorem 5']).

From Lemmas 6.7 and 4.2 (more precisely, from its proof), we have

Corollary 6.8. *Suppose $v \in \mathrm{BV}_n(\mathbb{R}^n)$ and $S \subset \mathbb{R}^n$ is an $(n-1)$-dimensional C^1-smooth compact manifold (without boundary). Then there exists $\delta = \delta(S) > 0$ such that for any ball $B = B(z,r)$ with $z \in S$ and $r < \delta$ the estimate*

$$\mathcal{H}^1(v(Z_v \cap B \cap S)) \le C \|D^n v\|(B_+)$$

holds, where C depends on n only and B_+ is a connected component of the open set $B \setminus S$.

The next lemma follows from the elementary observation that for any finite measure μ we have that $\mu(\{x \in \mathbb{R}^n : 0 < \operatorname{dist}(x,S) < \varepsilon\}) \to 0$ as $\varepsilon \searrow 0$.

Lemma 6.9. *Suppose $v \in \mathrm{BV}_n(\mathbb{R}^n)$ and $S \subset \mathbb{R}^n$ is an $(n-1)$-dimensional C^1-smooth compact manifold (without boundary). Then for any $\varepsilon > 0$ there exists a finite family of balls $B^j = B(z_j, r_j)$, $j = 1, \dots, N$, such that $z_j \in S$, $r_j < \varepsilon$, and*

$$S \subset \bigcup_{j=1}^N B^j, \quad \sum_{j=1}^N \|D^n v\|(B^j_+) < \varepsilon.$$

Combining these results we find

Corollary 6.10. *Suppose $v \in \mathrm{BV}_n(\mathbb{R}^n)$ and $S \subset \mathbb{R}^n$ is an $(n-1)$-dimensional C^1-smooth compact manifold. Then $\mathcal{H}^1(v(Z_v \cap S)) = 0$.*

Recall that a set $K \subset \mathbb{R}^n$ is called $(n-1)$-rectifiable, if there exists an at most countable family of C^1-surfaces $S_i \subset \mathbb{R}^n$ of dimension $(n-1)$ such that $\mathcal{H}^{n-1}(K \setminus \bigcup_i S_i) = 0$.

We can therefore reformulate Corollaries 6.10 and 6.6 in the following form.

Corollary 6.11. *Suppose $v \in \mathrm{BV}_n(\mathbb{R}^n)$ and $K \subset \mathbb{R}^n$ is an $(n-1)$-rectifiable set. Then $\mathcal{H}^1(v(Z_v \cap K)) = 0$.*

The following fact is well known.

Theorem 6.12 (see [8, Theorems B and 1]). *Suppose that $v \in \mathrm{BV}_n(\mathbb{R}^n)$. Then there exists a decomposition $\mathbb{R}^n = K_v \cup G_v$ with the following properties:*

(i) *K_v is $(n-1)$-rectifiable.*

(ii) *Each $x \in G_v$ is a Lebesgue point for $\nabla^{n-1}v$, moreover, $\nabla^{n-2}v$ is differentiable at x in the following integral sense:*

$$\fint_{B(x,r)} \left|\nabla^{n-2}v(y) - \nabla^{n-2}v(x) - \nabla^{n-1}v(x)\cdot(y-x)\right| \mathrm{d}y = o(r) \quad \text{as } r \searrow 0. \tag{14}$$

Now we are able to prove the following main result:

Theorem 6.13. *Suppose $v \in \mathrm{BV}_n(\mathbb{R}^n)$. Then $\mathcal{H}^1(v(Z_v)) = 0$.*

Proof. In view of Corollary 6.6 and Theorem 6.2 it is sufficient to prove the target equality $\mathcal{H}^1(v(Z_v)) = 0$ only for a case when $v \in \mathrm{BV}_n(\mathbb{R}^n) \cap \mathrm{C}^{n-2,1}(\mathbb{R}^n)$, i.e., $v \in \mathrm{C}^{n-2}(\mathbb{R}^n)$ and $\nabla^{n-2}v$ satisfies the Lipschitz condition

$$\left|\nabla^{n-2}v(y) - \nabla^{n-2}v(x)\right| \le L|y-x| \quad \text{for all } x, y \in \mathbb{R}^n \tag{15}$$

and for some constant $L > 0$. Consider the sets K_v, G_v from Theorem 6.12. In view of Corollary 6.11 we have

$$\mathcal{H}^1(v(Z_v \cap K_v)) = 0.$$

So we need only to prove that $\mathcal{H}^1(v(Z_v \cap G_v)) = 0$.

Take the decomposition (nondisjoint in general) $G_v = G_1 \cup G_2 \cup G_3$, where

$$\begin{aligned}
G_1 &= \left\{x \in G_v : \exists\, m = 2, \dots, n-2 : \nabla^m v(x) \neq 0\right\},\\
G_2 &= \left\{x \in G_v : \nabla^{n-1}v(x) = 0\right\},\\
G_3 &= \left\{x \in G_v : \nabla v^{n-2}(x) = 0,\ \nabla v^{n-1}(x) \neq 0\right\}.
\end{aligned}$$

Because of Corollary 6.10 and the Implicit Function Theorem for smooth functions we have

$$\begin{aligned}
&\mathcal{H}^1(v(Z_v \cap G_1))\\
&\quad \le \sum_{m=2}^{n-2} \mathcal{H}^1\Big(v\Big(\big\{x \in G_v : \nabla v(x) = \dots = \nabla^{m-1}v(x) = 0,\ \nabla^m v(x) \neq 0\big\}\Big)\Big) = 0.
\end{aligned}$$

On the other hand, by the Coarea Formula (see [10]) $\|D^n v\|(G_2) = 0$, hence by Lemma 6.5,

$$\mathcal{H}^1(v(Z_v \cap G_2)) = 0.$$

Now estimate $v(Z_v \cap G_3)$. From the integral differentiability (14) and the Lipschitz condition (15) it follows that $\nabla^{n-2}v$ is differentiable in the classical sense for each $x \in G_v$, i.e., we have for all $x \in G_v$

$$\left|\nabla^{n-2}v(y) - \nabla^{n-2}v(x) - \nabla^{n-1}v(x)\cdot(y-x)\right| = o(r) \quad \text{as } r \searrow 0.$$

Let $\mathbf{e}_i$, $i = 1, \dots, n$, be the unit coordinate vectors of $\mathbb{R}^n$. Denote

$$E_{i,j,k} = \left\{x \in G_3 : |\nabla v^{n-2}(x + t\mathbf{e}_i)| \ge \tfrac{1}{j}|t| \text{ for all } t \in [-\tfrac{1}{k}, \tfrac{1}{k}]\right\}.$$

By construction,

$$G_3 = \bigcup_{j,k \in \mathbb{N},\, i=1,\dots,n} E_{ijk}.$$

It is easy to see (using the Lipschitz condition (15)) that locally each set E_{ijk} is a graph of some Lipschitz function of $(n-1)$ variables $(x_1, \dots, \widehat{x}_i, \dots, x_n)$, i.e., E_{ijk} is $(n-1)$-rectifiable. Then by Corollary 6.11, $\mathcal{H}^1(v(Z_v \cap E_{ijk})) = 0$. $\square$

References

[1] *D. R. Adams*, A note on Choquet integrals with respect to Hausdorff capacity, in: Function spaces and applications (Lund 1986), Lecture Notes in Math. **1302**, Springer (1988), 115–124.

[2] *S. M. Bates*, Toward a precise smoothness hypothesis in Sard's theorem, Proc. Amer. Math. Soc. **117** (1993), no. 1, 279–283.

[3] *B. Bojarski*, *P. Hajlasz* and *P. Strzelecki*, Improved $C^{k,\lambda}$ approximation of higher order Sobolev functions in norm and capacity, Indiana Univ. Math. J. **51** (2002), no. 3, 507–540.

[4] *B. Bojarski*, *P. Hajlasz* and *P. Strzelecki*, Sard's theorem for mappings in Hölder and Sobolev spaces, Manuscripta Math. **118** (2005), 383–397.

[5] *J. Bourgain*, *M. V. Korobkov* and *J. Kristensen*, On the Morse–Sard property and level sets of Sobolev and BV functions, Rev. Mat. Iberoam. **29** (2013), no. 1, 1–23.

[6] *D. Bucur*, *A. Giacomini* and *P. Trebeschi*, Whitney property in two dimensional Sobolev spaces, Proc. Amer. Math. Soc. **136** (2008), no. 7, 2535–2545.

[7] *L. De Pascale*, The Morse–Sard theorem in Sobolev spaces, Indiana Univ. Math. J. **50** (2001), 1371–1386.

[8] *J. R. Dorronsoro*, Differentiability properties of functions with bounded variation, Indiana Univ. Math. J. **38** (1989), no. 4, 1027–1045.

[9] *A. Y. Dubovitskiĭ*, Structure of level sets for differentiable mappings of an n-dimensional cube into a k-dimensional cube (Russian), Izv. Akad. Nauk SSSR Ser. Mat. **21** (1957), no. 3, 371–408.

[10] *L. C. Evans* and *R. F. Gariepy*, Measure theory and fine properties of functions, Stud. Adv. Math., CRC Press, Boca Raton 1992.

[11] *H. Federer*, Geometric measure theory, Springer, New York 1969.

[12] *A. Figalli*, A simple proof of the Morse–Sard theorem in Sobolev spaces, Proc. Amer. Math. Soc. **136** (2008), 3675–3681.

[13] *F. C. Liu* and *W. S. Tai*, Approximate Taylor polynomials and differentiation of functions, Topol. Methods Nonlinear Anal. **3** (1994), no. 1, 189–196.

[14] *V. G. Maz'ya*, Sobolev spaces, Springer, Berlin 1985.

[15] *A. P. Morse*, The behavior of a function on its critical set, Ann. of Math. **40** (1939), 62–70.

[16] *A. Norton*, A critical set with nonnull image has large Hausdorff dimension, Trans. Amer. Math. Soc. **296** (1986), 367–376.

[17] *D. Pavlica* and *L. Zajíček*, Morse–Sard theorem for d.c. functions and mappings on $\mathbb{R}^2$, Indiana Univ. Math. J. **55** (2006), no. 3, 1195–1207.

[18] *A. Sard*, The measure of the critical values of differentiable maps, Bull. Amer. Math. Soc. **48** (1942), 883–890.

[19] *E. M. Stein*, Singular integrals and differentiability properties of functions, Princeton Math. Ser. **30**, Princeton University Press, Princeton 1970.

[20] *W.-S. Tai*, A capacitary weak type inequality for Sobolev functions and its applications, Proc. Amer. Math. Soc. **129** (2000), no. 3, 699–711.

[21] *H. Whitney*, A function not constant on a connected set of critical points, Duke Math. J. **1** (1935), 514–517.

[22] *Y. Yomdin*, The geometry of critical and near-critical values of differentiable mappings, Math. Ann. **264** (1983), no. 4, 495–515.

[23] *W. P. Ziemer*, Weakly differentiable functions. Sobolev spaces and functions of bounded variation, Graduate Texts in Math. **120**, Springer, New York 1989.

Jean Bourgain, School of Mathematics, Institute for Advanced Study,
Einstein Drive, Princeton, NJ 08540, USA
e-mail: bourgain@ias.edu

Mikhail V. Korobkov, Sobolev Institute of Mathematics, Acad. Koptyuga pr. 4;
and Novosibirsk State University, Pirogova Str. 2, 630090 Novosibirsk, Russia
e-mail: korob@math.nsc.ru

Jan Kristensen, Mathematical Institute, University of Oxford,
24–29 St. Giles', Oxford OX1 4AU, England
e-mail: kristens@maths.ox.ac.uk

Eingegangen 17. Januar 2012, in revidierter Fassung 22. Dezember 2012