J. Eur. Math. Soc. 9, 219–252

Giovanni Leoni · Massimiliano Morini

Necessary and sufficient conditions for the chain rule in $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N;\mathbb{R}^d)$ and $BV_{\mathrm{loc}}(\mathbb{R}^N;\mathbb{R}^d)$

Dedicated to the memory of Vic Mizel

Received November 5, 2004, and in revised form July 25, 2005

Abstract. We prove necessary and sufficient conditions for the validity of the classical chain rule in the Sobolev space $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N;\mathbb{R}^d)$ and in the space $BV_{\mathrm{loc}}(\mathbb{R}^N;\mathbb{R}^d)$ of functions of bounded variation.

Keywords. Chain rule, Sobolev functions, functions of bounded variation

1. Introduction

The purpose of this paper is to settle a classical problem in the theory of Sobolev spaces, namely the validity of the chain rule in $W^{1,1}_{\mathrm{loc}}(\Omega;\mathbb{R}^d)$ in the vectorial case $d > 1$. Since the problem is local, in the rest of the paper we assume, without loss of generality, that $\Omega = \mathbb{R}^N$.

In 1979 Marcus and Mizel [17] proved that given a Borel function $f : \mathbb{R}^d \to \mathbb{R}$, the superposition operator

$$u \mapsto f \circ u$$

maps $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N;\mathbb{R}^d)$ into $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N)$ if and only if f is Lipschitz continuous (resp. locally Lipschitz if $N = 1$). Since $f \circ u \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N)$ the next step is to find a formula for the partial derivatives of $f \circ u$.

In the scalar case, that is, when $d = 1$, the problem has been completely solved in $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N)$ by Serrin [21] in an unpublished paper (see also [23], [7] and [15]), where he showed that if $f : \mathbb{R} \to \mathbb{R}$ is a Lipschitz continuous function, then for every function $u \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N)$,

$$\nabla(f \circ u)(x) = f'(u(x))\nabla u(x) \quad \text{for } \mathcal{L}^N\text{-a.e. } x \in \Omega, \tag{1.1}$$

G. Leoni: Department of Mathematical Sciences, Carnegie-Mellon University, Pittsburgh, PA, USA; e-mail: giovanni@andrew.cmu.edu

M. Morini: Department of Mathematical Sciences, Carnegie-Mellon University, Pittsburgh, PA, USA; current address: SISSA, Trieste, Italy; e-mail: morini@sissa.it

Mathematics Subject Classification (2000): Primary 46E35; Secondary 26B05, 26B30, 26B40, 28A75

where the right side of (1.1) is always well defined provided $f'(u(x))\nabla u(x)$ is interpreted to be zero whenever $\nabla u(x) = 0$, irrespective of whether $f'(u(x))$ is defined. The validity of (1.1) relies on the fact that by Rademacher's theorem the set

$$\Sigma^f := \{u \in \mathbb{R} : f'(u) \text{ does not exist}\}$$

is $\mathcal{L}^1$-null and hence, by a result of Serrin and Varberg [22],

$$\nabla u(x) = 0 \quad \text{for } \mathcal{L}^N\text{-a.e. } x \in u^{-1}(\Sigma^f), \tag{1.2}$$

for every $u \in W^{1,1}_{\text{loc}}(\mathbb{R}^N)$.

The situation is significantly more complicated in the vectorial case, namely when $f : \mathbb{R}^d \to \mathbb{R}$ is a Lipschitz continuous function with $d > 1$. In this case, if we fix a basis $\{e_1, \ldots, e_d\}$ in $\mathbb{R}^d$ (not necessarily orthonormal) then the analog of (1.1) becomes[1]

$$\frac{\partial}{\partial x_j}(f \circ u)(x) = \sum_{i=1}^{d} \frac{\partial f}{\partial e_i}(u(x)) \frac{\partial u_i}{\partial x_j}(x), \tag{1.3}$$

where $\frac{\partial f}{\partial e_i}(u(x)) \frac{\partial u_i}{\partial x_j}(x)$ is interpreted to be zero whenever $\frac{\partial u_i}{\partial x_j}(x) = 0$.

By Rademacher's theorem the set

$$\Sigma^f := \{u \in \mathbb{R}^d : f \text{ is not differentiable at } u\} \tag{1.4}$$

is $\mathcal{L}^d$*-null*, but *the analog of* (1.2) *is false in general*. Hence the right hand side of (1.3) *may be nowhere defined*. Indeed, let $d = 2,\ N = 1$, and consider the functions (cf. [15]) $f(u) := \max\{u_1, u_2\}$ and $u(x) := (x, x)$ for $x \in \mathbb{R}$. Then $v(x) := (f \circ u)(x) = x$ so that $v'(x) = 1$ while the right hand side of (1.3) is nowhere defined since $u'(x) = (1, 1)$.

Nevertheless, as shown by Ambrosio and Dal Maso [2], the following weaker form of the chain rule holds for any Lipschitz continuous function $f : \mathbb{R}^d \to \mathbb{R}$:[2]

Theorem 1.1. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz continuous function. Then for every function $u \in W^{1,1}_{\text{loc}}(\mathbb{R}^N; \mathbb{R}^d)$ the composite function $v = f \circ u$ belongs to $W^{1,1}_{\text{loc}}(\mathbb{R}^N)$ and for $\mathcal{L}^N$-a.e. $x \in \mathbb{R}^N$ the restriction of the function f to the affine space*

$$T^u_x := \{w \in \mathbb{R}^d : w = u(x) + \nabla u(x)z \text{ for some } z \in \mathbb{R}^N\}$$

is differentiable at $u(x)$ and

$$\nabla(f \circ u)(x) = \nabla_u(f|_{T^u_x})(u(x))\nabla u(x). \tag{1.5}$$

[1] Here for every $u \in \mathbb{R}^d$ we write $u = u_1 e_1 + \cdots + u_d e_d$.

[2] Theorem 1.1 follows from a more general version for functions of bounded variation. We refer to [2] for the precise statement.

An alternative proof of the previous result when f is a "piecewise" C^1 function has been given in [19], where using the special structure of f it is possible to give an explicit formula for the right hand side of (1.5).

Theorem 1.1 leaves us with an important open problem: to establish under which additional conditions on the function f the right side of (1.5) coincides with the right side of (1.3), in other words, to *find necessary and sufficient conditions on f for the classical chain rule* (1.3) *to hold.* To be precise, when we say that (1.3) holds for a function $u \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N;\mathbb{R}^d)$, we also mean that the partial derivative $\partial f/\partial e_i$ exists at $u(x)$ for $\mathcal{L}^N$-a.e. x in the set where $\partial u_i/\partial x_j$ does not vanish.

The main purpose of this paper is to thoroughly investigate the relation between the validity of the chain rule (1.3) and the structure of the singular set Σ^f (defined in (1.4)) of the Lipschitz continuous function f.

It is clear from the definition that (1.3) depends on the choice of basis in $\mathbb{R}^d$. To illustrate this, we begin by considering the special case $d = 2$. Fix a basis $\{e_1, e_2\}$ in $\mathbb{R}^2$ not necessarily orthonormal and let $f : \mathbb{R}^2 \to \mathbb{R}$ be a Lipschitz continuous function.

It turns out that the classical chain rule (1.3) holds if and only if the singular set Σ^f has a one-dimensional "rectifiable" part only in the directions $\{e_1, e_2\}$. Precisely, we prove the following result:

Theorem 1.2. *The classical chain rule* (1.3) *holds in $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N;\mathbb{R}^2)$ with respect to the coordinate system $\{e_1, e_2\}$ if and only if for every $\mathcal{H}^1$-rectifiable set $E \subset \Sigma^f$ and for $\mathcal{H}^1$-a.e. $u \in E$ either*

$$\mathrm{Tan}^1(E,u) = \mathrm{span}\{e_1\} \quad \textit{or} \quad \mathrm{Tan}^1(E,u) = \mathrm{span}\{e_2\}. \tag{1.6}$$

Here $\mathrm{Tan}^1(E,u)$ is the approximate tangent space to the set E at the point u. The deeper part of the result is the necessary condition, whose proof relies on some new differentiability results for Lipschitz functions (see Theorems 3.1 and 3.3 below), inspired by recent work of Bessis and Clarke [6]. We recall that an $\mathcal{H}^k$-measurable set $E \subset \mathbb{R}^d$ is called *(countably) $\mathcal{H}^k$-rectifiable,* $0 \le k \le d$, if there exists a sequence of Lipschitz functions $w_n : \mathbb{R}^k \to \mathbb{R}^d$ such that

$$\mathcal{H}^k\Big(E \setminus \bigcup_{n=1}^{\infty} w_n(\mathbb{R}^k)\Big) = 0.$$

The analog of condition (1.6) is still sufficient when $d \ge 3$. Indeed, we show:

Theorem 1.3. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz continuous function, and let $\{e_1, \dots, e_d\}$ be a basis in $\mathbb{R}^d$. Assume that for every countably $\mathcal{H}^1$-rectifiable set $E \subset \Sigma^f$ and for $\mathcal{H}^1$-a.e. $u \in E$, there exists $i = 1, \dots, d$ depending on u such that*

$$\mathrm{Tan}^1(E,u) = \mathrm{span}\{e_i\}. \tag{1.7}$$

Then the classical chain rule (1.3) *holds in $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N;\mathbb{R}^d)$ with respect to the coordinate system $\{e_1, \dots, e_d\}$.*

Condition (1.7) is no longer necessary when $d \geq 3$ since in this case the chain rule may hold for Lipschitz functions whose singular set is $\mathcal{H}^k$-rectifiable with $1 \leq k \leq d-2$.

Indeed, we can prove the following result:

Theorem 1.4. *Let $\{e_1, \dots, e_d\}$ be an orthonormal basis and let $E \subset \mathbb{R}^d$, $d \geq 3$, be an $\mathcal{H}^{d-2}$-rectifiable Borel set. Then there exists a Lipschitz continuous function $f : \mathbb{R}^d \to \mathbb{R}$ such that $\Sigma^f \subset E$ and $\mathcal{H}^{d-2}(E \setminus \Sigma^f) = 0$, and for which the chain rule holds in $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N; \mathbb{R}^d)$ for any $N \in \mathbb{N}$ with respect to the coordinate system $\{e_1, \dots, e_d\}$.*

It is actually possible to construct f in such a way that the chain rule holds in $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N; \mathbb{R}^d)$ with respect to *any finite family of bases* in $\mathbb{R}^d$.

Note that the case $k = d-2$ represents the worst possible situation. Indeed, we can prove that a necessary (but not sufficient) condition for the validity of the chain rule in $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N; \mathbb{R}^d)$ is that for every $\mathcal{H}^{d-1}$-rectifiable set $E \subset \Sigma^f$ and for $\mathcal{H}^{d-1}$-a.e. $u \in E$, there exists $i = 1, \dots, d-1$ depending on u such that

$$\mathrm{Tan}^{d-1}(E, u) = \mathrm{span}\{\{e_1, \dots, e_d\} \setminus \{e_i\}\}.$$

Nevertheless, we show that this condition becomes necessary and sufficient for the validity of the chain rule in the smaller class $\mathcal{A}_{d-1}(\mathbb{R}^N; \mathbb{R}^d)$ of all functions $u \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N; \mathbb{R}^d)$ such that $\mathrm{rank}(\nabla u(x))$ is either zero or greater than or equal to $d-1$ for $\mathcal{L}^N$-a.e. $x \in \mathbb{R}^N$ (see Theorem 4.1 below). Note that $\mathcal{A}_1(\mathbb{R}^N; \mathbb{R}^2) = W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N; \mathbb{R}^2)$ so that in particular we recover Theorem 1.2.

It is important to remark again that all the results presented so far depend on the particular choice of the coordinate system $\{e_1, \dots, e_d\}$. We next address the case where *the chain rule holds with respect to every coordinate system*. In this case Theorem 1.2 clearly indicates that a necessary condition is that the singular set has no $\mathcal{H}^1$-rectifiable part, that is, it is *purely $\mathcal{H}^1$-unrectifiable*.

Indeed, the second main result of the paper is given by the following theorem:

Theorem 1.5. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz continuous function. Then a necessary and sufficient condition for the chain rule* (1.3) *to hold in $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N; \mathbb{R}^d)$ with respect to every coordinate system in $\mathbb{R}^d$ is that Σ^f is purely $\mathcal{H}^1$-unrectifiable.*

We remark that the sufficiency part of the theorem was already known (see [15]–[17]), while the necessity part, which in our opinion is the most interesting, is completely new. Note that in the two-dimensional case the conclusion of the theorem follows directly from Theorem 1.2, whereas in the higher dimensional case the proof is significantly more involved.

A similar result holds in the class $\mathcal{A}_k(\mathbb{R}^N; \mathbb{R}^d)$ of all functions u in the space $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N; \mathbb{R}^d)$ such that $\mathrm{rank}(\nabla u(x))$ is either zero or greater than or equal to k for $\mathcal{L}^N$-a.e. $x \in \mathbb{R}^N$. In this case the appropriate necessary and sufficient condition is the pure $\mathcal{H}^k$-unrectifiability of the singular set Σ^f. More precisely, we can show the following:

Theorem 1.6. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz continuous function, let $1 \leq k \leq \min\{N, d\}$. Then a necessary and sufficient condition for the chain rule* (1.3) *to hold in $\mathcal{A}_k(\mathbb{R}^N; \mathbb{R}^d)$ with respect to every coordinate system is that Σ^f is purely $\mathcal{H}^k$-unrectifiable.*

The final part of the paper is devoted to the extension of some of the results presented above to the space of functions of bounded variation. More precisely, we prove necessary and sufficient conditions for the validity of the classical chain rule in the space of functions of bounded variation $BV_{\text{loc}}(\mathbb{R}^N; \mathbb{R}^d)$.

Besides the intrinsic interest of these results, we hope that the techniques introduced in this paper will be useful in the study of transport equations and hyperbolic systems of conservation laws in several space dimensions, where one is often led to the problem of justifying some kind of chain rule for functions with low regularity, and for which there has been a remarkable and renewed interest in the last few years (see e.g. [3] and [8]).

2. Preliminaries

In this section we collect some preliminary results which will be used in the sequel. We start with some notation. Here $\mathcal{L}^k$ and $\mathcal{H}^k$ are, respectively, the k-dimensional Lebesgue measure and the k-dimensional Hausdorff measure in Euclidean spaces. We denote by S^{d-1} the unit sphere in $\mathbb{R}^d$. Given $f : \mathbb{R}^d \to \mathbb{R}$, for every $u, v \in \mathbb{R}^d$ the *directional derivative* $\frac{\partial f}{\partial v}(u)$ is defined by

$$\frac{\partial f}{\partial v}(u) := \lim_{t \to 0} \frac{f(u + tv) - f(u)}{t}.$$

Given a basis $\{e_1, \ldots, e_d\}$ in $\mathbb{R}^d$ we denote by $(u_1, \ldots, u_d)$ the components of a given $u \in \mathbb{R}^d$, that is,

$$u = u_1 e_1 + \cdots + u_d e_d.$$

The directional derivatives in the direction e_i are also denoted $\frac{\partial f}{\partial u_i}(u)$. If all the derivatives $\frac{\partial f}{\partial u_i}(u)$ exist at $u \in \mathbb{R}^d$, we define the vector $\nabla f(u) \in \mathbb{R}^d$ by

$$\nabla f(u) := \left(\frac{\partial f}{\partial u_1}(u), \ldots, \frac{\partial f}{\partial u_d}(u) \right).$$

Of course, the existence of $\nabla f(u)$ does not imply the differentiability of f at u when $d > 1$.

For $1 \leq k < d$ we shall use the standard notation for ordered multi-indices:

$$I(k, d) := \{\alpha = (\alpha_1, \ldots, \alpha_k) \in \mathbb{N}^k : 1 \leq \alpha_1 < \cdots < \alpha_k \leq d\}.$$

If $\alpha \in I(k, d)$ we denote by $\bar{\alpha} \in I(d - k, d)$ the multi-index which complements α in $\{1, \ldots, d\}$ in the natural increasing order.

With an abuse of notation we write

$$u = (u_\alpha, u_{\bar\alpha}) \in \mathbb{R}^k \times \mathbb{R}^{d-k},$$

where $u_\alpha := (u_{\alpha_1}, \dots, u_{\alpha_k})$ and $u_{\bar\alpha} := (u_{\bar\alpha_1}, \dots, u_{\bar\alpha_{d-k}})$.

Given $f : \mathbb{R}^d \to \mathbb{R}$, we define

$$\nabla_\alpha f := \left(\frac{\partial f}{\partial u_{\alpha_1}}, \dots, \frac{\partial f}{\partial u_{\alpha_k}}\right), \qquad \nabla_{\bar\alpha} f := \left(\frac{\partial f}{\partial u_{\bar\alpha_1}}, \dots, \frac{\partial f}{\partial u_{\bar\alpha_{d-k}}}\right).$$

Next we introduce some basic ingredients in geometric measure theory that will be useful in the rest of the paper. We refer to [4], [14] and [18] for more details.

An $\mathcal{H}^k$-measurable set $E \subset \mathbb{R}^d$ is called *(countably) $\mathcal{H}^k$-rectifiable,* $0 \le k \le d$, if there exists a sequence of Lipschitz functions $w_n : \mathbb{R}^k \to \mathbb{R}^d$ such that

$$\mathcal{H}^k\Big(E \setminus \bigcup_{n=1}^\infty w_n(\mathbb{R}^k)\Big) = 0.$$

It can be shown that E is $\mathcal{H}^k$-rectifiable if and only there exists a sequence $\{M_n\}$ of k-dimensional C^1 manifolds such that

$$\mathcal{H}^k\Big(E \setminus \bigcup_{n=1}^\infty M_n\Big) = 0. \tag{2.1}$$

Moreover, if E is $\mathcal{H}^k$-rectifiable then it admits an approximate tangent space (see Def. 2.86 in [4]), which we denote by $\mathrm{Tan}^k(E, u)$, for $\mathcal{H}^k$-a.e. $u \in E$, and it can be shown that for $\mathcal{H}^k$-a.e. $u \in E \cap M_n$ the approximate tangent space to E at u coincides with the tangent space to the manifold M_n at u, that is,

$$\mathrm{Tan}^k(E, u) = \mathrm{Tan}^k(M_n, u). \tag{2.2}$$

We refer to [4] for more details.

Let $f : \mathbb{R}^d \to \mathbb{R}$ be Lipschitz and let $M \subset \mathbb{R}^d$ be a k-manifold of class C^1. We say that f is *tangentially differentiable* at $u \in M$ if the restriction to the affine space $u + \mathrm{Tan}^k(M, u)$ is differentiable at u. The tangential differential, denoted by $d^M f(u)$, is a linear map between the space $\mathrm{Tan}^k(M, u)$ and $\mathbb{R}$.

Remark 2.1. Let $f : \mathbb{R}^d \to \mathbb{R}$ be Lipschitz and let $M \subset \mathbb{R}^d$ be a k-manifold of class C^1. If f is tangentially differentiable at $u \in M$ then for every curve[3] $\gamma : (-\delta, \delta) \to \mathbb{R}^d$ with $\gamma(0) = u$ and $\gamma'(0) \in \mathrm{Tan}^k(M, u)$ we have

$$(f \circ \gamma)'(0) = d^M f(u)[\gamma'(0)].$$

Indeed, by taking δ smaller if necessary we can write

$$\gamma(t) = \gamma_1(t) + o(t)$$

where $\gamma_1 : (-\delta, \delta) \to M$. Since f is Lipschitz we have $(f \circ \gamma)(t) = (f \circ \gamma_1)(t) + o(t)$, from which the conclusion follows.

[3] Note that the support of γ is not contained in M.

Theorem 2.2. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be Lipschitz and let $M \subset \mathbb{R}^d$ be a k-manifold of class C^1. Then f is tangentially differentiable at $\mathcal{H}^k$-a.e. $u \in M$.*

Proof. For every $u \in M$ consider a local parametrization $\psi : D \to M$ of M, with $\psi(0) = u$, where $D \subset \mathbb{R}^k$ is an open neighborhood of the origin. By Rademacher's theorem the function $f \circ \psi$ is differentiable $\mathcal{L}^k$-almost everywhere in D. Using the Lipschitz continuity, it is easy to see that if $f \circ \psi$ is differentiable at $v \in D$, then f is tangentially differentiable at $\psi(v) \in M$. Hence f is tangentially differentiable $\mathcal{H}^k$-almost everywhere in $\psi(D)$. □

In a similar way we may define tangential differentiability of a Lipschitz function f at points $u \in E$ where $E \subset \mathbb{R}^d$ is an $\mathcal{H}^k$-rectifiable set, $1 \le k < d$. In this case, the tangential differential, denoted by $d^E f(u)$, is a linear map between the space $\mathrm{Tan}^k(E, u)$ and $\mathbb{R}$. It can be shown that $d^E f(u)$ exists for $\mathcal{H}^k$-a.e. $u \in E$. Moreover if $f : \mathbb{R}^d \to \mathbb{R}^m$ with $m \ge k$ then the following Generalized Area Formula holds:

$$\int_E J_k^E f(u)\, d\mathcal{H}^k(u) = \int_{f(E)} \mathcal{H}^0(f^{-1}(v) \cap E)\, d\mathcal{H}^k(v), \tag{2.3}$$

where

$$J_k^E f(u) := \sqrt{\det((d^E f(u))^* \circ d^E f(u))},$$

with $(d^E f(u))^*$ the adjoint of $d^E f(u)$.

An $\mathcal{H}^k$-measurable set $E \subset \mathbb{R}^d$ is *purely $\mathcal{H}^k$-unrectifiable* if

$$\mathcal{H}^k(E \cap w(\mathbb{R}^k)) = 0$$

for any Lipschitz function $w : \mathbb{R}^k \to \mathbb{R}^d$.

Theorem 2.3. *Consider a set $E \subset \mathbb{R}^d$ of finite $\mathcal{H}^k$ measure. Then E can be decomposed into the disjoint union of a Borel $\mathcal{H}^k$-rectifiable set $E^{k\text{-}\mathrm{rect}}$ and of a purely $\mathcal{H}^k$-unrectifiable set $E^{k\text{-}\mathrm{unrect}}$. The decomposition is unique, up to sets of $\mathcal{H}^k$ measure zero.*

Purely $\mathcal{H}^k$-unrectifiable sets with finite (or σ-finite) $\mathcal{H}^k$ measure may be characterized in a simple way by virtue of the Structure Theorem of Besicovitch–Federer (see [18]). In what follows for $0 < k < d$ we denote by $\gamma_{d,k}$ the Haar measure defined on the Grassmannian manifold $G(d, k)$ of all k-dimensional planes in $\mathbb{R}^d$ (see [18]). We identify each element $L \in G(d, k)$ with the orthogonal projection $\pi_L : \mathbb{R}^d \to L$.

Theorem 2.4 (Structure Theorem). *Let $E \subset \mathbb{R}^d$ be an $\mathcal{H}^k$-measurable set with $\mathcal{H}^k(E) < \infty$. Then*

(i) *E is $\mathcal{H}^k$-rectifiable if and only if*

$$\mathcal{H}^k(\pi_L E_1) > 0 \quad \text{for } \gamma_{d,k}\text{-a.e. } L \in G(d, k),$$

for all $\mathcal{H}^k$-measurable subsets E_1 of E with $\mathcal{H}^k(E_1) > 0$.

(ii) *E is purely $\mathcal{H}^k$-unrectifiable if and only if*

$$\mathcal{H}^k(\pi_L E) = 0 \quad \text{for } \gamma_{d,k}\text{-a.e. } L \in G(d, k).$$

Remark 2.5. If Γ is a rectifiable curve on the plane $\mathbb{R}^2$ then it may be proved that any $\mathcal{H}^1$-measurable subset of Γ with positive $\mathcal{H}^1$ measure can project into a set of length zero in at most one direction. Hence if $E \subset \mathbb{R}^2$ is $\mathcal{H}^1$-measurable with $\mathcal{H}^1(E) < \infty$ and if there exist two lines $L, L_1 \in G(2,1)$ such that $\mathcal{H}^1(\pi_L E) = \mathcal{H}^1(\pi_{L_1} E) = 0$ then E is purely $\mathcal{H}^1$-unrectifiable. From the previous theorem we then deduce that $\mathcal{H}^1(\pi_L E) = 0$ for $\gamma_{2,1}$-a.e. $L \in G(2,1)$.

Next we present some simple properties of the differentiability of Lipschitz functions. It is well known that if a function f is differentiable at some point, say the origin, then necessarily

- f is continuous at 0;
- the directional derivatives $\frac{\partial f}{\partial v}(0)$ exist for every $v \in S^{d-1}$;
- for every $v \in S^{d-1}$,

$$\frac{\partial f}{\partial v}(0) = \sum_{i=1}^{d} \frac{\partial f}{\partial e_i}(0) v_i . \tag{2.4}$$

These properties are in general not sufficient to guarantee differentiability at 0. Indeed the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(v_1, v_1) := \begin{cases} v_1 & \text{if } v_2 = (v_1)^2, \\ 0 & \text{otherwise,} \end{cases}$$

is clearly continuous at 0, $\frac{\partial f}{\partial v}(0) = 0$ for every $v \in S^{d-1}$, but f is not differentiable at the origin.

The situation is quite different if the function f is Lipschitz continuous, as in this case it is easy to verify that if (2.4) holds for every v in a dense subset of S^{d-1} then f *is differentiable at the origin*. More precisely we have:

Proposition 2.6. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz function. Then the following conditions are equivalent:*

(1) *f is differentiable at* 0;
(2) *there exists a linear operator $L : \mathbb{R}^d \to \mathbb{R}$ such that the limit*

$$\lim_{h \to 0^+} \frac{f(h\nu) - f(0)}{h} = L(\nu)$$

exists for all ν in a countable dense subset of $\mathbb{R}^d$;
(3) *there exists a countable dense family $\mathcal{B}$ of orthonormal bases such that*

$$\sum_{i=1}^{d} \frac{\partial f}{\partial e_i}(0) e_i = \sum_{i=1}^{d} \frac{\partial f}{\partial \epsilon_i}(0) \epsilon_i$$

for any two bases $\{e_1, \ldots, e_d\}$ and $\{\epsilon_1, \ldots, \epsilon_d\}$ in $\mathcal{B}$.

Finally, the following Lusin-type theorem holds (see [14]):

Theorem 2.7. *Let $u \in W^{1,1}(\mathbb{R}^N)$. Given $\lambda > 0$, there exist a closed set C_λ and a function $v_\lambda \in C^1(\mathbb{R}^N)$ such that $u = v_\lambda$ and $\nabla u = \nabla v_\lambda$ on C_λ, and*

$$|\mathbb{R}^N \setminus C_\lambda| \leq \frac{C(N,d)}{\lambda}\|u\|_{W^{1,1}(\mathbb{R}^N)}, \quad \|v_\lambda\|_{W^{1,\infty}} \leq \lambda.$$

Moreover,

$$\|v_\lambda\|_{W^{1,\infty}(\mathbb{R}^N)}|\mathbb{R}^N \setminus C_\lambda| \to 0, \quad \|u - g_\lambda\|_{W^{1,1}} \to 0 \quad \text{as } \lambda \to \infty.$$

3. Differentiability criteria for Lipschitz functions

In this section we prove some differentiability criteria for Lipschitz functions.

Theorem 3.1. *Let $f : \mathbb{R}^d \to \mathbb{R}$, $d > 1$, be a Lipschitz function, let $\{e_1, \ldots, e_d\}$ be a basis in $\mathbb{R}^d$ and let $1 \leq k < d$. Then the following two conditions are equivalent:*

(1) *for every $\alpha \in I(k,d)$ the set*

$$\Sigma_\alpha^f := \{u = (u_\alpha, u_{\bar\alpha}) \in \mathbb{R}^k \times \mathbb{R}^{d-k} : f(u_\alpha, \cdot) \text{ is not differentiable at } u_{\bar\alpha}\} \tag{3.1}$$

is purely $\mathcal{H}^k$-unrectifiable;

(2) *the singular set*

$$\Sigma^f := \{u \in \mathbb{R}^d : f \text{ is not differentiable at } u\} \tag{3.2}$$

is purely $\mathcal{H}^k$-unrectifiable.

Proof. The implication (2)⇒(1) is trivial. To prove the converse, let $M \subset \mathbb{R}^d$ be a k-manifold of class C^1. We claim that

$$\mathcal{H}^k(M \cap \Sigma^f) = 0.$$

Fix $u_0 \in M \cap \Sigma^f$. By the implicit function theorem, we can find an open neighborhood U of u_0 such that

$$M \cap U \subset \{u = (u_\alpha, u_{\bar\alpha}) \in \mathbb{R}^k \times \mathbb{R}^{d-k} : u_{\bar\alpha} = \varphi(u_\alpha)\}$$

for some $\varphi : \mathbb{R}^k \to \mathbb{R}^{d-k}$ of class C^1 and for some $\alpha \in I(k,d)$. To prove the claim it is enough to show that

$$\mathcal{H}^k(M \cap \Sigma^f \cap U) = 0.$$

By hypothesis (1), for $\mathcal{H}^k$-a.e. $(u_\alpha, \varphi(u_\alpha)) \in \mathbb{R}^k \times \mathbb{R}^{d-k}$, the function $f(u_\alpha, \cdot)$ is differentiable at $\varphi(u_\alpha)$. Since the projection is Lipschitz continuous, it follows that for $\mathcal{L}^k$-a.e. $u_\alpha \in \mathbb{R}^k$, the function $f(u_\alpha, \cdot)$ is differentiable at $\varphi(u_\alpha)$. Consider the change of variables $h : \mathbb{R}^k \times \mathbb{R}^{d-k} \to \mathbb{R}^k \times \mathbb{R}^{d-k}$ given by

$$(u_\alpha, u_{\bar\alpha}) \mapsto (w_\alpha, w_{\bar\alpha}) := (u_\alpha, u_{\bar\alpha} - \varphi(u_\alpha)),$$

define the function $g : \mathbb{R}^k \times \mathbb{R}^{d-k} \to \mathbb{R}$ by

$$g(w_\alpha, w_{\bar\alpha}) := f(w_\alpha, \varphi(w_\alpha) + w_{\bar\alpha}),$$

and let $L > 0$ denote its Lipschitz constant. It is clear that

$$\mathcal{L}^k(\{w_\alpha \in \mathbb{R}^k : \text{the function } g(w_\alpha, \cdot) \text{ is not differentiable at } 0\}) = 0. \tag{3.3}$$

We claim that

$$\mathcal{L}^k(\{w_\alpha \in \mathbb{R}^k : \text{the function } g \text{ is not differentiable at } (w_\alpha, 0)\}) = 0. \tag{3.4}$$

Fix $\nu = (\nu_\alpha, \nu_{\bar\alpha}) \in S^{d-1}$, $r \in \mathbb{R}$, and $n \in \mathbb{N}$. Following [6] we define the set

$$C(\nu, r, n) := \Big\{ w_\alpha \in \mathbb{R}^k : \frac{g(w_\alpha, t\nu_{\bar\alpha}) - g(w_\alpha, 0)}{t} > r > \underline{\partial}_\nu g(w_\alpha, 0) - d_\alpha g(w_\alpha, 0)(\nu_\alpha) \text{ for all } t \in (0, 1/n) \Big\},$$

where

$$\underline{\partial}_\nu g(w_\alpha, 0) := \liminf_{s \to 0^+} \frac{g(w_\alpha + s\nu_\alpha, s\nu_{\bar\alpha}) - g(w_\alpha, 0)}{s},$$

and $d_\alpha g(w_\alpha, 0)$ denotes the differential of the function $g(\cdot, 0)$ at the point w_α, which exists for $\mathcal{L}^k$-a.e. $w_\alpha \in \mathbb{R}^k$, by Rademacher's theorem and the fact that g is Lipschitz.

We claim that for every $w_\alpha \in C(\nu, r, n)$ there exist a constant $\lambda \in (0, 1)$ and a sequence $t_j \searrow 0$ such that

$$B_k(w_\alpha + t_j\nu_\alpha, \lambda t_j) \cap C(\nu, r, n) = \emptyset, \tag{3.5}$$

where $B_k(w_\alpha, \rho)$ denotes the open ball in $\mathbb{R}^k$ with center w_α and radius ρ. The proof of the claim follows closely the argument of Bessis and Clarke (see [6]). We present it here for the convenience of the reader. Fix $w_\alpha \in C(\nu, r, n)$ and let $t_j \searrow 0$ be such that

$$\underline{\partial}_\nu g(w_\alpha, 0) = \lim_{j \to \infty} \frac{g(w_\alpha + t_j\nu_\alpha, t_j\nu_{\bar\alpha}) - g(w_\alpha, 0)}{t_j}.$$

By definition of $C(\nu, r, n)$ we can find $0 < \delta < 2L$ such that for all j sufficiently large,

$$\begin{aligned} r &> 2\delta + \underline{\partial}_\nu g(w_\alpha, 0) - d_\alpha g(w_\alpha, 0)(\nu_\alpha) \\ &> \delta + \frac{g(w_\alpha + t_j\nu_\alpha, t_j\nu_{\bar\alpha}) - g(w_\alpha, 0)}{t_j} - \frac{g(w_\alpha + t_j\nu_\alpha, 0) - g(w_\alpha, 0)}{t_j} \\ &= \delta + \frac{g(w_\alpha + t_j\nu_\alpha, t_j\nu_{\bar\alpha}) - g(w_\alpha + t_j\nu_\alpha, 0)}{t_j}, \end{aligned} \tag{3.6}$$

while for $z_\alpha \in C(\nu, r, n)$ we have

$$\frac{g(z_\alpha, t_j\nu_{\bar\alpha}) - g(z_\alpha, 0)}{t_j} > r$$

for all $t_j < 1/n$. Combining this inequality with (3.6) we obtain

$$
\begin{aligned}
\delta &< \frac{g(z_\alpha, t_j\nu_{\bar\alpha}) - g(w_\alpha + t_j\nu_\alpha, t_j\nu_{\bar\alpha})}{t_j} + \frac{g(w_\alpha + t_j\nu_\alpha, 0) - g(z_\alpha, 0)}{t_j} \\
&\le \frac{2L}{t_j}|z_\alpha - (w_\alpha + t_j\nu_\alpha)|,
\end{aligned}
$$

which gives (3.5) with $\lambda := \delta/(2L)$.

Let now $w_\alpha \in C(\nu, r, n)$ be a Lebesgue point for the characteristic function $\chi_{C(\nu,r,n)}$. Since

$$B_k(w_\alpha + t_j\nu_\alpha, \lambda t_j) \subset B_k(w_\alpha, 2t_j),$$

by (3.5) we have

$$
\begin{aligned}
0 = 1 - \chi_{C(\nu,r,n)}(w_\alpha) &= \lim_{j\to\infty}\left(1 - \frac{\mathcal{L}^k(B_k(w_\alpha, 2t_j) \cap C(\nu, r, n))}{\mathcal{L}^k(B_k(w_\alpha, 2t_j))}\right) \\
&= \lim_{j\to\infty} \frac{\mathcal{L}^k(B_k(w_\alpha, 2t_j) \setminus C(\nu, r, n))}{\mathcal{L}^k(B_k(w_\alpha, 2t_j))} \\
&\ge \limsup_{j\to\infty} \frac{\mathcal{L}^k(B_k(w_\alpha + t_j\nu_\alpha, \lambda t_j) \setminus C(\nu, r, n))}{\mathcal{L}^k(B_k(w_\alpha, 2t_j))} \\
&= \limsup_{j\to\infty} \frac{\mathcal{L}^k(B_k(w_\alpha + t_j\mu_\alpha, \lambda t_j))}{\mathcal{L}^k(B_k(w_\alpha, 2t_j))} = \left(\frac{\lambda}{2}\right)^k,
\end{aligned}
$$

which is clearly a contradiction. Hence

$$\mathcal{L}^k(C(\nu, r, n)) = 0. \tag{3.7}$$

Let $E \subset S^{d-1}$ be a countable dense set. In view of (3.7) the set

$$\{w_\alpha \in \mathbb{R}^k : \nabla g(w_\alpha, 0) \cdot \nu > \underline{\partial}_\nu g(w_\alpha, 0) \text{ for all } \nu \in E\} \subseteq \bigcup_{\nu\in E}\bigcup_{r\in\mathbb{Q}}\bigcup_{n\in\mathbb{N}} C(\nu, r, n)$$

has zero $\mathcal{L}^k$ measure. By applying the same argument to the function $-g$ and taking into account (3.3), for $\mathcal{L}^k$-a.e. $w_\alpha \in \mathbb{R}^k$ we obtain

$$\nabla g(w_\alpha, 0) \cdot \nu = \partial_\nu g(w_\alpha, 0) \quad \text{for all } \nu \in E.$$

Using now the fact that g is Lipschitz and Proposition 2.6 yields (3.4). Since

$$f(u_\alpha, u_{\bar\alpha}) = (g \circ h)(u_\alpha, u_{\bar\alpha}),$$

and φ is of class C^1 we have

$$
\begin{aligned}
\mathcal{H}^k(M \cap \Sigma^f \cap U) &\le \mathcal{H}^k(h^{-1}\{(w_\alpha, 0) : w_\alpha \in \mathbb{R}^k\} \cap \Sigma^f) \\
&\le \mathrm{Lip}(h^{-1})\mathcal{H}^k(\{(w_\alpha, 0) : w_\alpha \in \mathbb{R}^k\} \cap \Sigma^g) \\
&= \mathrm{Lip}(h^{-1})\mathcal{L}^k(\{w_\alpha \in \mathbb{R}^k : (w_\alpha, 0) \in \Sigma^g\}) = 0,
\end{aligned}
$$

where

$$\Sigma^g := \{w \in \mathbb{R}^d : g \text{ is not differentiable at } w\}.$$

This concludes the proof. □

Remark 3.2. From the proof of the previous theorem it is clear that if $M \subset \mathbb{R}^d$ is a k-manifold of class C^1 and Π is a $(d-k)$-plane such that f restricted to $u + \Pi$ is differentiable at u for $\mathcal{H}^k$-a.e. $u \in M$ and

$$\mathrm{Tan}^k(M, u) + \Pi = \mathbb{R}^d$$

for every $u \in M$, then f is differentiable at $\mathcal{H}^k$-a.e. $u \in M$.

For the applications to the chain rule in Sobolev spaces we will need the following variant of the previous theorem.

Theorem 3.3. *Let $f : \mathbb{R}^d \to \mathbb{R}$, $d > 1$, be a Lipschitz function, let $\{e_1, \dots, e_d\}$ be a basis in $\mathbb{R}^d$ and let $1 \le k < d$. Assume that for every $\alpha \in I(k, d)$ and for every $\mathcal{H}^k$-rectifiable set $E \subset \Sigma^f_\alpha$, where Σ^f_α is the set defined in* (3.1), *we have*

$$\mathrm{Tan}^k(E, u) = \mathrm{span}\{e_{\alpha_1}, \dots, e_{\alpha_k}\} \tag{3.8}$$

for $\mathcal{H}^k$-a.e. $u \in E$. Then for every $\mathcal{H}^k$-rectifiable set $E \subset \Sigma^f$ and for $\mathcal{H}^k$-a.e. $u \in E$, there exists $\alpha \in I(k, d)$ depending on u such that (3.8) *holds.*

Proof. Fix an $\mathcal{H}^k$-rectifiable set $E \subset \Sigma^f$. By (2.1) and (2.2), we may assume, without loss of generality, that $E \subset M$, where M is a k-manifold of class C^1, and that

$$\mathrm{Tan}^k(E, u) = \mathrm{Tan}^k(M, u)$$

for all $u \in E$. Thus to prove the theorem it suffices to show that the set

$$E_1 := \{u \in E : \mathrm{Tan}^k(M, u) \neq \mathrm{span}\{e_{\alpha_1}, \dots, e_{\alpha_k}\} \text{ for every } \alpha \in I(k, d)\}$$

has $\mathcal{H}^k$ measure zero. Fix $u \in E_1$. Since M is of class C^1, we may find an open neighborhood U of u such that

$$\mathrm{Tan}^k(M, z) \neq \mathrm{span}\{e_{\alpha_1}, \dots, e_{\alpha_k}\} \tag{3.9}$$

for all $z \in M \cap U$ and for every $\alpha \in I(k, d)$. It is enough to show that

$$\mathcal{H}^k(\Sigma^f \cap M \cap U) = 0.$$

By hypothesis, it follows from (3.9) that

$$\mathcal{H}^k(\Sigma^f_\alpha \cap M \cap U) = 0$$

for every $\alpha \in I(k, d)$. We can now continue as in the proof of Theorem 3.1. □

Remark 3.4. We remark that if $k = 1$ then condition (3.8) is equivalent to the following one: for every $\alpha \in I(1, d)$ and for every $\mathcal{H}^1$-rectifiable set $E \subset \Sigma_\alpha^f$ we have

$$\mathcal{H}^1(\Pi_{\bar{\alpha}}(E)) = 0, \tag{3.10}$$

where $\Pi_{\bar{\alpha}} : \mathbb{R}^d \to \mathbb{R}^{d-1}$ is the projection defined by

$$u = (u_\alpha, u_{\bar{\alpha}}) \mapsto u_{\bar{\alpha}}.$$

This follows from the Generalized Area Formula (2.3)

$$\int_E J_1 \Pi_{\bar{\alpha}}(u)\, d\mathcal{H}^1(u) = \int_{\Pi_\alpha(E)} \mathcal{H}^0(\Pi_{\bar{\alpha}}^{-1}(v) \cap E)\, d\mathcal{H}^1(v). \tag{3.11}$$

A simple calculation shows that

$$J_1 \Pi_{\bar{\alpha}}(u) = |(\tau(u))_{\bar{\alpha}}|, \tag{3.12}$$

where $\tau(u)$ is the tangent unit vector to E at u.

Assume now that (3.10) holds. Then from (3.11) and (3.12) we have

$$\int_E |(\tau(u))_{\bar{\alpha}}|\, d\mathcal{H}^1(u) = 0,$$

which implies that $(\tau(u))_{\bar{\alpha}} = 0$ for $\mathcal{H}^1$-a.e. $u \in E$, that is, (3.8). Conversely, if (3.8) is satisfied then $(\tau(u))_{\bar{\alpha}} = 0$ for $\mathcal{H}^1$-a.e. $u \in E$, and therefore

$$\int_{\Pi_{\bar{\alpha}}(E)} \mathcal{H}^0(\Pi_{\bar{\alpha}}^{-1}(v) \cap E)\, d\mathcal{H}^1(v) = 0.$$

Since $\mathcal{H}^0(\Pi_{\bar{\alpha}}^{-1}(v) \cap E) \geq 1$ for every $v \in \Pi_{\bar{\alpha}}(E)$, we deduce that (3.10) holds.

4. Chain rule in $W^{1,1}_{\text{loc}}(\mathbb{R}^N; \mathbb{R}^d)$

In this section we prove the main results of the paper, namely we give sufficient and necessary conditions for the validity of the chain rule in $W^{1,1}_{\text{loc}}(\mathbb{R}^N; \mathbb{R}^d)$. We begin by studying the validity of the chain rule with respect to a fixed basis $\{e_1, \ldots, e_d\}$ in $\mathbb{R}^d$. We recall that $\mathcal{A}_k(\mathbb{R}^N; \mathbb{R}^d)$ is the class of all functions u in the space $W^{1,1}_{\text{loc}}(\mathbb{R}^N; \mathbb{R}^d)$ such that $\text{rank}(\nabla u(x))$ is either zero or greater than or equal to k for $\mathcal{L}^N$-a.e. $x \in \mathbb{R}^N$.

Theorem 4.1. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz continuous function, let $\{e_1, \ldots, e_d\}$ be a basis in $\mathbb{R}^d$, and let $1 \leq k \leq \min\{N, d\}$. Assume that for every $\mathcal{H}^k$-rectifiable set $E \subset \Sigma^f$ and for $\mathcal{H}^k$-a.e. $u \in E$ there exists $\alpha \in I(k, d)$ depending on u such that*

$$\text{Tan}^k(E, u) = \text{span}\{e_{\alpha_1}, \ldots, e_{\alpha_k}\}. \tag{4.1}$$

Then the classical chain rule (1.3) *holds in the class $\mathcal{A}_k(\mathbb{R}^N; \mathbb{R}^d)$ with respect to the coordinate system $\{e_1, \ldots, e_d\}$.*

Moreover, if $k = d - 1$, then condition (4.1) *is also necessary for the validity of* (1.3) *in the class $\mathcal{A}_{d-1}(\mathbb{R}^N; \mathbb{R}^d)$.*

Remark 4.2. (i) Since $\mathcal{H}^d = \mathcal{L}^d$ it follows by Rademacher's theorem that $\mathcal{H}^d(\Sigma^f) = 0$ and so condition (4.1) is automatically satisfied for every Lipschitz function f. Hence Theorem 4.1 implies in particular that for any Lipschitz function f the classical chain rule always holds in $\mathcal{A}_d(\mathbb{R}^N;\mathbb{R}^d)$.

(ii) Note that if Σ^f has σ-finite $\mathcal{H}^k$ measure then in view of Theorem 2.3 it suffices to verify condition (4.1) for the $\mathcal{H}^k$-rectifiable part of Σ^f, that is, for $(\Sigma^f)^{k\text{-}\mathrm{rect}}$.

We begin with some preliminary lemmas.

Lemma 4.3. *Under the hypotheses of Theorem* 4.1 *the chain rule holds in the class* $\mathcal{A}_k(\mathbb{R}^N;\mathbb{R}^d)$ *if and only if for every function* $u \in C^1(\mathbb{R}^N;\mathbb{R}^d)$ *it holds* $\mathcal{L}^N$*-a.e. in the set*

$$\{x \in \mathbb{R}^N : \textit{either } \operatorname{rank}(\nabla u(x)) \geq k \textit{ or } \nabla u(x) = 0\}. \tag{4.2}$$

Proof. Assume that the chain rule holds in the class $\mathcal{A}_k(\mathbb{R}^N;\mathbb{R}^d)$ and let $u \in C^1(\mathbb{R}^N;\mathbb{R}^d)$. Since f is Lipschitz, for every $x \in \mathbb{R}^N$ such that $\nabla u(x) = 0$ it is clear that $\nabla(f \circ u)(x) = 0$, so that the chain rule always holds on the set $\{x \in \mathbb{R}^N : \nabla u(x) = 0\}$. To prove it in the set $A := \{x \in \mathbb{R}^N : \operatorname{rank}(\nabla u(x)) \geq k\}$ fix $x_0 \in A$ and let $m := \operatorname{rank}(\nabla u(x_0))$. We claim that there exists $B(x_0, r) \subset\subset A$ and a function $v \in \mathcal{A}_k(\mathbb{R}^N;\mathbb{R}^d)$ such that $u \equiv v$ on $B(x_0, r)$ and $\operatorname{rank}(\nabla v) \geq m$ in $\mathbb{R}^N$. Indeed, it is enough to take

$$v(x) := \varphi(x)u(x) + (1 - \varphi(x))(u(x_0) + \nabla u(x_0)(x - x_0))$$

where $\varphi \in C^1_c(B(x_0, 2r))$ is such that $\varphi \equiv 1$ on $B(x_0, r)$ and $\|\nabla\varphi\|_\infty \leq C/r$. Then

$$\begin{aligned}\nabla v(x) = \nabla u(x_0) &+ \varphi(x)(\nabla u(x) - \nabla u(x_0)) \\ &+ \nabla\varphi(x) \otimes (u(x) - u(x_0) - \nabla u(x_0)(x - x_0)).\end{aligned} \tag{4.3}$$

Clearly $\operatorname{rank}(\nabla v(x)) = m$ for all $x \in \mathbb{R}^N \setminus B(x_0, 2r)$. Since u is of class C^1 for every $\varepsilon > 0$ we may find $r > 0$ so small that

$$|\nabla u(x) - \nabla u(x_0)| \leq \varepsilon, \quad |u(x) - u(x_0) - \nabla u(x_0)(x - x_0)| \leq \varepsilon|x - x_0|$$

for all $x \in B(x_0, 2r)$. Hence from (4.3) we obtain

$$|\nabla v(x) - \nabla u(x_0)| \leq C\varepsilon \tag{4.4}$$

for all $x \in B(x_0, 2r)$.

Find $\alpha \in I(m, N)$ and $\beta \in I(m, d)$ such that, with the usual notation $x = (x_\alpha, x_{\bar\alpha}) \in \mathbb{R}^m \times \mathbb{R}^{N-m}$ and $u = (u_\beta, u_{\bar\beta}) \in \mathbb{R}^m \times \mathbb{R}^{d-m}$,

$$\det \nabla_\alpha u_\beta(x_0) \neq 0.$$

Using the inequality

$$|\det A - \det B| \leq C(m)|A - B|(|A|^{m-1} + |B|^{m-1}),$$

which holds for $A, B \in \mathbb{R}^{m\times m}$, from (4.3) and (4.4) we obtain

$$|\det \nabla_\alpha v_\beta(x) - \det \nabla_\alpha u_\beta(x_0)| \le C\varepsilon$$

for all $x \in B(x_0, 2r)$. By taking ε (and in turn r) sufficiently small, we conclude that $\operatorname{rank}(\nabla v(x)) \ge m$ for all $x \in \mathbb{R}^N$.

To prove the opposite implication assume by contradiction that there exists a function $u \in \mathcal{A}_k(\mathbb{R}^N;\mathbb{R}^d)$ for which the chain rule fails in a set $F \subset \mathbb{R}^N$ of positive measure. By Theorem 2.7 there exists a function $v \in C^1(\mathbb{R}^N;\mathbb{R}^d)$ which coincides with u in a subset of F of positive measure. This is clearly a contradiction. □

Lemma 4.4. *Under the hypotheses of Theorem* 4.1 *if* Σ^f *contains an* $\mathcal{H}^m$*-rectifiable subset* E *with* $\mathcal{H}^m(E) > 0$ *then, necessarily,* $m \le k$.

Proof. Indeed, let E be as above and assume by contradiction that $d > m > k$. By (2.1) and (2.2), we may assume, without loss of generality, that $E \subset M$, where M is an m-dimensional manifold of class C^1, and that

$$\operatorname{Tan}^m(E, u) = \operatorname{Tan}^m(M, u)$$

for all $u \in E$. Clearly,

$$\mathcal{H}^m(M \cap \Sigma^f) > 0. \tag{4.5}$$

After a translation and a rotation, and by taking M smaller if necessary we may assume that 0 is a point of $\mathcal{H}^m$ density 1 in $M \cap \Sigma^f$ and that

$$M \subset \operatorname{Graph} g, \tag{4.6}$$

where

$$\operatorname{Graph} g := \{v = (v_\beta, v_{\bar\beta}) \in \mathbb{R}^m \times \mathbb{R}^{d-m} : v_{\bar\beta} = g(v_\beta)\},$$

and $g : \mathbb{R}^m \to \mathbb{R}^{d-m}$ is a function of class C^1 with $g(0) = 0$, $\nabla g(0) = 0$, and $\beta = (1, \dots, m)$. It is also clear that without loss of generality we may assume that the k-plane

$$L_0 := \{v = (v_1, \dots, v_k, 0, \dots, 0) \in \mathbb{R}^d : v_1, \dots, v_k \in \mathbb{R}\}$$

is not a coordinate plane with respect to the old coordinate system $\{e_1, \dots, e_d\}$. For $w = (w_{k+1}, \dots, w_m) \in \mathbb{R}^{m-k}$ let L_w and M_w denote respectively the k-plane

$$L_w := \{v = (v_1, \dots, v_k, w_{k+1}, \dots, w_m, 0, \dots, 0) \in \mathbb{R}^d : v_1, \dots, v_k \in \mathbb{R}\}$$

and the k-manifold

$$M_w := M \cap \operatorname{Graph}(g|_{L_w}).$$

Since $\operatorname{Tan}^k(M_0, 0) = L_0$, by continuity we can assume that $\operatorname{Tan}^k(M_w, u)$ is not a coordinate plane with respect to the old coordinate system $\{e_1, \dots, e_d\}$ for all $w \in \mathbb{R}^{m-k}$ and all $u \in M_w$ with $|u|, |w| < \varepsilon_0$ for some $\varepsilon_0 > 0$. To conclude the proof it suffices to show that there is $w \in \mathbb{R}^{m-k}$ with $|w| < \varepsilon_0$ such that $\mathcal{H}^k(M_w \cap \Sigma^f \cap B(0, \varepsilon_0)) > 0$. Indeed, this would contradict (4.1).

By condition (4.5) and by the fact that 0 is a point of $\mathcal{H}^m$ density 1 in $M \cap \Sigma^f$ it follows that the projection P of $M \cap \Sigma^f$ on the m-plane $v_{\bar{\beta}} = 0$ has positive $\mathcal{L}^m$ measure and that $0 \in \mathbb{R}^m$ is a point of $\mathcal{L}^m$ density 1 for P. By Fubini's theorem there exists $w = (w_{k+1}, \dots, w_m) \in \mathbb{R}^{m-k}$ with $|w| < \varepsilon_0$ such that $\mathcal{H}^k(L_w \cap P \cap B(0, \varepsilon_0)) > 0$. Hence $\mathcal{H}^k(M_w \cap \Sigma^f \cap B(0, \varepsilon_0)) > 0$ and the proof is concluded. □

Lemma 4.5. *Let $u \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N; \mathbb{R}^d)$ and assume that there exists a measurable set F of positive measure such that*

$$\operatorname{rank}(\nabla u(x)) \geq k$$

for all $x \in F$, for some $1 \leq k \leq \min\{N, d\}$. Then there exists a k-dimensional manifold $M \subset \mathbb{R}^d$ of class C^1 such that

$$\mathcal{H}^k(M \cap u(F)) > 0.$$

Proof. As in Lemma 4.3 we may assume without loss of generality that $u \in C^1(\mathbb{R}^N; \mathbb{R}^d)$. Let $\bar{x} \in F$ be a Lebesgue point for the characteristic function χ_F. Since $\operatorname{rank}(\nabla u(\bar{x})) \geq k$ we may find $\alpha \in I(k, N)$ and $\varepsilon > 0$ such that, with the usual notation $x = (x_\alpha, x_{\bar{\alpha}}) \in \mathbb{R}^k \times \mathbb{R}^{N-k}$,

$$\operatorname{rank}(\nabla_\alpha u(x_\alpha, x_{\bar{\alpha}})) = k \tag{4.7}$$

for every $x_\alpha \in B_k(\bar{x}_\alpha, \varepsilon)$ and $x_{\bar{\alpha}} \in B_{N-k}(\bar{x}_{\bar{\alpha}}, \varepsilon)$. Set

$$A := B_k(\bar{x}_\alpha, \varepsilon) \times B_{N-k}(\bar{x}_{\bar{\alpha}}, \varepsilon).$$

Note that since $\bar{x}$ is a Lebesgue point we have

$$\mathcal{L}^N(A \cap F) > 0. \tag{4.8}$$

By Fubini's theorem and (4.8) it is easy to see that there exists $\hat{x}_{\bar{\alpha}} \in B_{N-k}(\bar{x}_{\bar{\alpha}}, \varepsilon)$ such that

$$\mathcal{L}^k(\{x_\alpha \in B_k(\bar{x}_\alpha, \varepsilon) : (x_\alpha, \hat{x}_{\bar{\alpha}}) \in F\}) > 0. \tag{4.9}$$

It is clear that

$$M := \{u(x_\alpha, \hat{x}_{\bar{\alpha}}) : x_\alpha \in B_k(\bar{x}_\alpha, \varepsilon)\}$$

is a k-dimensional manifold such that

$$\mathcal{H}^k(M \cap u(F)) > 0.$$ □

As a corollary of the previous lemma we obtain the following characterization of purely $\mathcal{H}^k$-unrectifiable Borel sets, which can be considered as an extension of a classical result of Serrin and Varberg [22].

Corollary 4.6. *A Borel set $E \subset \mathbb{R}^d$ is purely $\mathcal{H}^k$-unrectifiable, $1 \leq k \leq d$, if and only if for every $N \geq k$,*

$$\nabla u = 0 \quad \mathcal{L}^N\text{-a.e. in } u^{-1}(E)$$

for every $u \in \mathcal{A}_k(\mathbb{R}^N; \mathbb{R}^d)$.

Proof. Assume that $E \subset \mathbb{R}^d$ is purely $\mathcal{H}^k$-unrectifiable. We claim that

$$\mathcal{L}^N(u^{-1}(E) \cap \{\nabla u \neq 0\}) = 0 \tag{4.10}$$

for every $u \in \mathcal{A}_k(\mathbb{R}^N; \mathbb{R}^d)$. Indeed, if $\mathcal{L}^N(u^{-1}(E) \cap \{\nabla u \neq 0\}) > 0$ for some $u \in \mathcal{A}_k(\mathbb{R}^N; \mathbb{R}^d)$ then, since $\operatorname{rank}(\nabla u) \geq k$ $\mathcal{L}^N$-a.e. in $u^{-1}(E) \cap \{\nabla u \neq 0\}$, by the previous lemma, we may find a k-dimensional manifold $M \subset \mathbb{R}^d$ such that

$$\mathcal{H}^k(M \cap E) > 0,$$

which contradicts the fact that E is purely $\mathcal{H}^k$-unrectifiable.

Conversely, assume that (4.10) holds and let $M \subset \mathbb{R}^d$ be a k-dimensional manifold. We claim that

$$\mathcal{H}^k(M \cap E) = 0.$$

If not then we can find a local parametrization $\psi : D \subset \mathbb{R}^k \to M$ of class C^1 such that

$$\mathcal{H}^k(\psi(D) \cap E) > 0 \quad \text{and} \quad \operatorname{rank}(\nabla u) = k.$$

This implies that $\mathcal{L}^k(D \cap \psi^{-1}(E)) > 0$. Reasoning as in the first part of the proof of Lemma 4.3, without loss of generality we may assume that $\psi \in \mathcal{A}_k(\mathbb{R}^k; \mathbb{R}^d)$ and thus we have a contradiction to (4.10). □

We are now ready to prove Theorem 4.1.

Proof of Theorem 4.1. By Lemma 4.3 it suffices to prove that for every $u \in C^1(\mathbb{R}^N; \mathbb{R}^d)$ the chain rule holds $\mathcal{L}^N$-a.e. in the set

$$\{x \in \mathbb{R}^N : \text{either } \operatorname{rank}(\nabla u(x)) \geq k \text{ or } \nabla u(x) = 0\}.$$

Since f is Lipschitz it is clear that

$$\frac{\partial}{\partial x_j}(f \circ u)(x) = 0 \quad \text{whenever} \quad \nabla u(x) = 0.$$

Moreover, if $u(x) \notin \Sigma^f$ then there is nothing to prove. Hence it remains to show the chain rule in the set

$$R_k := u^{-1}(\Sigma^f) \cap \{x \in \mathbb{R}^N : \operatorname{rank}(\nabla u(x)) \geq k\}. \tag{4.11}$$

We claim that for $\mathcal{L}^N$-a.e. $x \in R_k$ the rank of $\nabla u(x)$ is k. Indeed, if this is not the case then there exist a set $F \subset R_k$ with $\mathcal{L}^N(F) > 0$ and $m > k$ such that $\operatorname{rank}(\nabla u(x)) = m$ in F. Then by Lemma 4.5 there exists an m-dimensional manifold $M \subset \mathbb{R}^d$ of class C^1 such that

$$\mathcal{H}^m(M \cap \Sigma^f) > 0,$$

which contradicts Lemma 4.4.

Next we prove that for $\mathcal{L}^N$-a.e. $x \in R_k$ the affine space

$$T_x^u := \{w \in \mathbb{R}^d : w = u(x) + \nabla u(x)v \text{ for some } v \in \mathbb{R}^N\}$$

is parallel to a coordinate k-plane. For every $\bar{x} \in R_k$ there exists $\beta \in I(k, N)$ such that $\operatorname{rank}(\nabla_\beta u(\bar{x})) = k$. Since u is of class C^1 it follows that $\operatorname{rank}(\nabla_\beta u(x)) = k$ for all $x \in A := B_k(\bar{x}_\beta, \varepsilon) \times B_{N-k}(\bar{x}_{\bar{\beta}}, \varepsilon)$ for some $\varepsilon > 0$. By the previous claim it follows that for $\mathcal{L}^N$-a.e. $x \in R_k \cap A$,

$$T_x^u = \{w \in \mathbb{R}^d : w = u(x) + \nabla_\beta u(x) v \text{ for some } v \in \mathbb{R}^k\}.$$

Therefore by Fubini's theorem, for $\mathcal{L}^{N-k}$-a.e. $z \in B_{N-k}(\bar{x}_{\bar{\beta}}, \varepsilon)$,

$$u(y, z) + \operatorname{Tan}^k(M_z, u(y, z)) = T^u_{(y,z)} \tag{4.12}$$

for $\mathcal{L}^k$-a.e. $y \in B_k(\bar{x}_\beta, \varepsilon)$ such that $(y, z) \in R_k$, where M_z is the k-dimensional manifold

$$M_z := \{u(y, z) : y \in B_k(\bar{x}_\beta, \varepsilon)\}.$$

Fix $z \in B_{N-k}(\bar{x}_{\bar{\beta}}, \varepsilon)$ for which (4.12) holds. By the assumption (4.1) it follows that for $\mathcal{L}^k$-a.e. $y \in B_k(\bar{x}_\beta, \varepsilon)$ with $(y, z) \in R_k$ there exists $\alpha \in I(k, d)$ such that

$$T^u_{(y,z)} = u(y, z) + \operatorname{span}\{e_{\alpha_1}, \dots, e_{\alpha_k}\}. \tag{4.13}$$

Moreover by Theorem 2.2 we may assume that for the same set of y's there exists the tangential differential $d^{M_z} f(u(y, z))$. Hence, by Remark 2.1 applied to the curve $\gamma(t) := u(x + te_j)$, by (4.12), and (4.13), for all such points $x = (y, z)$ we have

$$\frac{\partial}{\partial x_j}(f \circ u)(x) = d^{M_z} f(u(x))\left[\frac{\partial u}{\partial x_j}(x)\right] = \sum_{i=1}^k \frac{\partial f}{\partial u_{\alpha_i}}(u(x)) \frac{\partial u_{\alpha_i}}{\partial x_j}(x),$$

which, since by (4.13)

$$\frac{\partial u_l}{\partial x_j}(x) = 0 \qquad \text{for all } l \notin \{\alpha_1, \dots, \alpha_k\},$$

implies that the chain rule holds for $\mathcal{L}^k$-a.e. $y \in B_k(\bar{x}_\beta, \varepsilon)$ with $x = (y, z) \in R_k$. As this is true for $\mathcal{L}^{N-k}$-a.e. $z \in B_{N-k}(\bar{x}_{\bar{\beta}}, \varepsilon)$, the proof of the first part of the theorem follows from Fubini's theorem.

Finally, we show that if $k = d - 1$, then condition (4.1) is also necessary for the classical chain rule to hold in the class $\mathcal{A}_{d-1}(\mathbb{R}^N; \mathbb{R}^d)$. By Theorem 3.3 it is enough to show that for every $\alpha \in I(d-1, d)$ and for every $\mathcal{H}^{d-1}$-rectifiable set $F \subset \Sigma^f_\alpha$, where

$$\Sigma^f_\alpha := \{u \in \mathbb{R}^d : \partial f / \partial u_{\bar{\alpha}} \text{ does not exist at } u\},$$

we have

$$\operatorname{Tan}^{d-1}(F, u) = \operatorname{span}\{e_{\alpha_1}, \dots, e_{\alpha_{d-1}}\} \tag{4.14}$$

for $\mathcal{H}^{d-1}$-a.e. $u \in F$.

Assume by contradiction that there exist $\alpha \in I(d-1, d)$ and a $(d-1)$-dimensional manifold $M \subset \mathbb{R}^d$ such that

$$\mathcal{H}^{d-1}(M \cap \Sigma^f_\alpha) > 0$$

and (4.14) fails on a subset $E \subset M \cap \Sigma_\alpha^f$ with $\mathcal{H}^{d-1}(E) > 0$. Consider a local parametrization $\psi : \mathbb{R}^{d-1} \supset D \to M$ of class C^1 such that

$$\mathcal{H}^{d-1}(\psi(D) \cap E) > 0 \tag{4.15}$$

and

$$\operatorname{rank}(\nabla \psi) = d - 1 \tag{4.16}$$

in D. Since (4.14) fails in E we have

$$\nabla \psi_{\bar{\alpha}} \neq 0 \quad \text{in } \psi^{-1}(E). \tag{4.17}$$

Reasoning as in the first part of the proof of Lemma 4.3, without loss of generality we may assume that $\psi \in \mathcal{A}_{d-1}(\mathbb{R}^{d-1}; \mathbb{R}^d)$. Let

$$u(x) := \psi(x_1, \ldots, x_{d-1}), \qquad x \in \mathbb{R}^N. \tag{4.18}$$

Then $u \in \mathcal{A}_{d-1}(\mathbb{R}^N; \mathbb{R}^d)$ and by (4.17),

$$\nabla u_{\bar{\alpha}} \neq 0 \quad \text{in } \psi^{-1}(E) \times \mathbb{R}^{N-d+1},$$

while by (4.15), (4.16) we have $\mathcal{L}^N(\psi^{-1}(E) \times \mathbb{R}^{N-d+1}) > 0$. Hence in this set the chain rule fails since by definition of Σ_α^f the partial derivative $\partial f / \partial u_{\bar{\alpha}}$ does not exist at $u(x)$.

□

The previous proof implies:

Corollary 4.7. *Let F be a Borel subset of $\mathbb{R}^d$ and let $\alpha \in I(k, d)$. Then the following two conditions are equivalent:*

(i) *for every $\mathcal{H}^k$-rectifiable subset $E \subset F$ and for $\mathcal{H}^k$-a.e. $u \in E$ we have*

$$\operatorname{Tan}^k(E, u) = \operatorname{span}\{e_{\alpha_1}, \ldots, e_{\alpha_k}\}; \tag{4.19}$$

(ii) *for every $u \in \mathcal{A}_k(\mathbb{R}^N; \mathbb{R}^d)$ and for $\mathcal{L}^N$-a.e. $x \in u^{-1}(F)$ we have $\nabla u_{\bar{\alpha}}(x) = 0$.*

Corollary 4.8. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a Lipschitz continuous function and let $\{e_1, e_2\}$ be a basis in $\mathbb{R}^2$. Then the classical chain rule holds in $W^{1,1}_{\text{loc}}(\mathbb{R}^N; \mathbb{R}^2)$ with respect to the coordinate system $\{e_1, e_2\}$ if and only if for every $\mathcal{H}^1$-rectifiable set $E \subset \Sigma^f$ and for $\mathcal{H}^1$-a.e. $u \in E$ either*

$$\operatorname{Tan}^1(E, u) = \operatorname{span}\{e_1\} \quad \textit{or} \quad \operatorname{Tan}^1(E, u) = \operatorname{span}\{e_2\}. \tag{4.20}$$

Remark 4.9. By Remark 3.4 condition (4.20) is equivalent to requiring that for all $\mathcal{H}^1$-rectifiable sets $E_i \subset \Sigma_i^f$, $i = 1, 2$,

$$\mathcal{H}^1(\Pi_1(E_2)) = 0 \quad \text{and} \quad \mathcal{H}^1(\Pi_2(E_1)) = 0$$

where $\Pi_i : \mathbb{R}^2 \to \mathbb{R}$ is the projection $u = (u_1, u_2) \mapsto u_i$, and we recall that

$$\Sigma_i^f := \{u \in \mathbb{R}^2 : \partial f / \partial e_i \text{ does not exist at } u\}.$$

If $k < d - 1$, then condition (4.1) is not necessary for the validity of the chain rule, as the following theorem shows:

Theorem 4.10. *Let $\{e_1, \dots, e_d\}$ be a basis and let $E \subset \mathbb{R}^d$, $d \geq 3$, be an $\mathcal{H}^{d-2}$-rectifiable Borel set. Then there is a bounded Lipschitz continuous function $f : \mathbb{R}^d \to \mathbb{R}$ such that $\Sigma^f \subset E$ and $\mathcal{H}^{d-2}(E \setminus \Sigma^f) = 0$, and for which the chain rule holds in $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N; \mathbb{R}^d)$ for any $N \in \mathbb{N}$ with respect to the coordinate system $\{e_1, \dots, e_d\}$.*

Proof. For simplicity we assume that $\{e_1, \dots, e_d\}$ is an orthonormal basis.

Step 1. Assume first that E is a compact set contained in

$$\{u \in \mathbb{R}^d : \varphi(u) = 0\},$$

where $\varphi : \mathbb{R}^d \to \mathbb{R}^2$ is a function of class C^1 with $\operatorname{rank}(\nabla\varphi) = 2$ in E. We show that for every $u_0 \in E$ it is possible to construct a bounded Lipschitz continuous function f such that $\Sigma^f = E \cap \overline{B(u_0, r)}$ for some small r and for which the chain rule holds. For simplicity we also assume that the vectors

$$\frac{\frac{\partial\varphi}{\partial u_i}}{\left|\frac{\partial\varphi}{\partial u_i}\right|}, \qquad i = 1, \dots, d,$$

are all distinct. The general case can be treated similarly.

Consider a bounded Lipschitz continuous function $g : \mathbb{R}^3 \to \mathbb{R}$ such that $g \in C^1(\mathbb{R}^3 \setminus \{0\})$, $g(\cdot, \cdot, 0)$ is not differentiable at the origin but admits all directional derivatives,

$$\frac{\partial g}{\partial v}(0) = \nabla g(0) \cdot v \tag{4.21}$$

for all $v \in \left(\bigcup_{i=1}^d \Lambda_i\right) \times \{0\}$ where

$$\Lambda_i := \{w = (\cos\theta, \sin\theta) : |\theta - \theta_i| < \varepsilon,\ |\theta + \theta_i| < \varepsilon\},$$

θ_i are the angles corresponding to the vectors $\frac{\partial\varphi}{\partial u_i}(u_0)$, and ε is chosen so small that the sets Λ_i are pairwise disjoint.[4] Since φ is of class C^1, for r sufficiently small we have

$$\frac{\frac{\partial\varphi}{\partial u_i}(u)}{\left|\frac{\partial\varphi}{\partial u_i}(u)\right|} \in \Lambda_i \tag{4.22}$$

for all $u \in \overline{B(u_0, r)}$.

[4] An example of a function satisfying all the desired properties is given by

$$g(x, y, z) := \begin{cases} \frac{x^2 y^3 \prod_{i=1}^d (y\cos(\theta_i+\varepsilon) - x\sin(\theta_i+\varepsilon))^2 (y\cos(\theta_i-\varepsilon) - x\sin(\theta_i-\varepsilon))^2}{x^{4+4d} + y^{4+4d} + z^2} & \text{if } (x, y) \notin D, \\ 0 & \text{otherwise,} \end{cases}$$

where D is the set of all points of $\mathbb{R}^2$ whose angle belongs to Λ_i for some i.

For $u \in \mathbb{R}^d$ define

$$f(u) := g(\varphi(u), \delta^2(u)),$$

where $\delta(\cdot)$ is the regularized distance from the set $K := E \cap \overline{B(u_0, r)}$ (see [24]). It is well known that δ is a Lipschitz continuous function with $\delta \in C^\infty(\mathbb{R}^d \setminus K)$,

$$\frac{1}{C}\,\mathrm{dist}(u, K) \le \delta(u) \le C\,\mathrm{dist}(u, K)$$

for all $u \in \mathbb{R}^d \setminus K$.

We claim that $\Sigma^f = K$. It is clear that $\Sigma^f \subset K$. To prove the opposite inclusion fix $\bar{u} \in K$ and let $\xi \in S^{d-1}$. Then

$$\begin{aligned}\frac{f(\bar{u}+t\xi) - f(\bar{u})}{t} &= \frac{g(\varphi(\bar{u}) + t\frac{\partial\varphi}{\partial\xi}(\bar{u}) + o(t), o(t)) - g(\varphi(\bar{u}), 0)}{t}\\ &= \frac{g(\varphi(\bar{u}) + t\frac{\partial\varphi}{\partial\xi}(\bar{u}), 0) - g(\varphi(\bar{u}), 0)}{t} + o(1)\end{aligned} \tag{4.23}$$

where we have used the facts that g is Lipschitz, φ is of class C^1, and $\delta^2 \in C^1(\mathbb{R}^d)$ with $\nabla\delta^2 = 0$ on K. Hence

$$\frac{\partial f}{\partial\xi}(\bar{u}) = \frac{\partial g}{\partial\nu_\varphi}(\varphi(\bar{u}), 0)\left|\frac{\partial\varphi}{\partial\xi}(\bar{u})\right|, \tag{4.24}$$

where

$$\nu_\varphi := \left(\frac{\frac{\partial\varphi}{\partial\xi}(\bar{u})}{|\frac{\partial\varphi}{\partial\xi}(\bar{u})|}, 0\right).$$

Since $g(\cdot, \cdot, 0)$ is not differentiable at the origin, by Proposition 2.6 we may find a direction $\nu = (w, 0) \in S^2 \setminus \bigcup_{i=1}^d (\Lambda_i \times \{0\})$ such that

$$\frac{\partial g}{\partial\nu}(0) \neq \nabla g(0) \cdot \nu.$$

Using the fact that $\mathrm{rank}(\nabla\varphi)(\bar{u}) = 2$ we may find a direction $\xi_0 \in S^{d-1}$ such that $\frac{\partial\varphi}{\partial\xi_0}(\bar{u}) = w|\frac{\partial\varphi}{\partial\xi_0}(\bar{u})| \neq 0$. Moreover by (4.21), (4.22), and (4.24),

$$\begin{aligned}\nabla f(\bar{u}) \cdot \xi_0 &= \nabla g(0) \cdot \left(\frac{\partial\varphi}{\partial\xi_0}(\bar{u}), 0\right) = \nabla g(0) \cdot \nu\left|\frac{\partial\varphi}{\partial\xi_0}(\bar{u})\right|\\ &\neq \frac{\partial g}{\partial\nu}(0)\left|\frac{\partial\varphi}{\partial\xi_0}(\bar{u})\right| = \frac{\partial f}{\partial\xi_0}(\bar{u})\end{aligned}$$

and so f is not differentiable at $\bar{u}$. This shows that $\Sigma^f = K$.

Next we prove that the chain rule holds for f. As in the proof of Theorem 4.1, to show the validity of the chain rule in $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N; \mathbb{R}^d)$ it is enough to prove it for any $u \in C^1(\mathbb{R}^N; \mathbb{R}^d)$. Moreover it is clear that we can assume, without loss of generality, that

$N = 1$. If $u(x) \notin \Sigma^f$ then there is nothing to prove, hence assume that $u(x) \in \Sigma^f$. As in (4.23) we have

$$\frac{f(u(x+t)) - f(u(x))}{t} = \frac{g((\varphi \circ u)(x) + t(\varphi \circ u)'(x), 0) - g((\varphi \circ u)(x), 0)}{t} + o(1)$$
$$= \frac{g((\varphi \circ u)(x+t), 0) - g((\varphi \circ u)(x), 0)}{t} + o(1).$$

Hence $(f \circ u)'(x) = (g(\cdot, \cdot, 0) \circ (\varphi \circ u))'(x)$ for every $x \in u^{-1}(\Sigma^f)$. Since the chain rule is valid for $g(\cdot, \cdot, 0)$ by Theorem 4.1, and recalling that $(\varphi \circ u)(x) = 0$, it follows that for $\mathcal{L}^1$-a.e. $x \in u^{-1}(\Sigma^f)$,

$$(f \circ u)'(x) = \nabla_\alpha g(0) \cdot (\varphi \circ u)'(x) = (\nabla_\alpha g(0) \nabla \varphi(u(x))) u'(x)$$
$$= \nabla f(u(x)) \cdot u'(x)$$

where $\alpha = (1, 2)$ and in the last equality we have used the fact that

$$\frac{\partial f}{\partial u_i}(u(x)) = \nabla_\alpha g(0) \frac{\partial \varphi}{\partial u_i}(u(x), 0),$$

which follows from (4.21), (4.22), and (4.24). Therefore the chain rule holds.

Step 2. Assume now that $E \subset \mathbb{R}^d$ is an $\mathcal{H}^{d-2}$-rectifiable set, that is,

$$E = \bigcup_{j=1}^{\infty} K_j \cup \mathcal{N}$$

where $\mathcal{H}^{d-2}(\mathcal{N}) = 0$ and the sets K_j are disjoint compact subsets of $(d-2)$-manifolds of class C^1.

Fix $j \in \mathbb{N}$. By Step 1 it is clear that for each $u \in K_j$ we can find $r_u > 0$ such that for all $0 < r < r_u$ there exists a bounded Lipschitz continuous function $f_{u,r}$ with $\Sigma^{f_{u,r}} = K_j \cap \overline{B(u, r)}$ and for which the chain rule holds. The union of all such balls for $u \in K_j$ is a fine cover for K_j and hence, by the Vitali–Besicovitch covering theorem (see e.g. Thm. 2.19 in [4]) we can find a countable sequence of disjoint closed balls $\overline{B(u_n, r_n)}$ such that

$$\mathcal{H}^{d-2}\Big(K_j \setminus \bigcup_{n=1}^{\infty} \overline{B(u_n, r_n)}\Big) = 0.$$

By repeating the same procedure for each K_j it is clear that we can find a sequence of bounded Lipschitz continuous functions f_n for which the chain rule holds, and such that the sets Σ^{f_n} are pairwise disjoint compact subsets of E with

$$\mathcal{H}^{d-2}\Big(E \setminus \bigcup_{n=1}^{\infty} \Sigma^{f_n}\Big) = 0.$$

Moreover we can assume that the partial derivatives of every f_n exist everywhere in $\mathbb{R}^d$ (see Step 1).

We may now define

$$f(u) := \sum_{n=1}^{\infty} \frac{1}{2^n L_n} f_n(u), \qquad L_n := \|f_n\|_{W^{1,\infty}(\mathbb{R}^d)}.$$

It is clear that f is bounded and Lipschitz continuous. We claim that

$$\Sigma^f = \bigcup_{n=1}^{\infty} \Sigma^{f_n}. \tag{4.25}$$

Indeed, if $u \notin \bigcup_{n=1}^{\infty} \Sigma^{f_n}$ then for any $\xi \in S^{d-1}$,

$$\frac{\partial f}{\partial \xi}(u) = \sum_{n=1}^{\infty} \frac{1}{2^n L_n} \frac{\partial f_n}{\partial \xi}(u) = \sum_{n=1}^{\infty} \frac{1}{2^n L_n} \nabla f_n(u) \cdot \xi = \nabla f(u) \cdot \xi, \tag{4.26}$$

where we repeatedly used the fact that we can differentiate term-by-term. By Proposition 2.6 and since f is Lipschitz continuous it follows that $u \notin \Sigma^f$. Conversely, if $u \in \bigcup_{n=1}^{\infty} \Sigma^{f_n}$ then there exists a unique $n_0 \in \mathbb{N}$ such that $u \in \Sigma^{f_{n_0}}$. Write

$$f = f_{n_0} + \sum_{n \neq n_0} \frac{1}{2^n L_n} f_n.$$

Arguing as in (4.26) we deduce that the function $\sum_{n \neq n_0} \frac{1}{2^n L_n} f_n$ is differentiable at u. Since $u \in \Sigma^{f_{n_0}}$ it follows that $u \in \Sigma^f$. Hence (4.25) holds.

Finally, to show that the chain rule holds for f let $u \in C^1(\mathbb{R}; \mathbb{R}^d)$. Since for $\mathcal{L}^1$-a.e. $x \in \mathbb{R}$ the functions $f_n \circ u$ are differentiable at x and

$$(f_n \circ u)'(x) = \nabla f_n(u(x)) \cdot u'(x)$$

for all n, using once more term-by-term differentiation we conclude that

$$(f \circ u)'(x) = \nabla f(u(x)) \cdot u'(x)$$

for $\mathcal{L}^1$-a.e. $x \in \mathbb{R}$. This concludes the proof. □

Remark 4.11. (i) It is clear from the previous proof that given any finite family $\mathcal{E}$ of bases in $\mathbb{R}^d$ and any $\mathcal{H}^{d-2}$-rectifiable set $E \subset \mathbb{R}^d$, $d \geq 3$, one can construct a Lipschitz continuous function $f : \mathbb{R}^d \to \mathbb{R}$ such that $\Sigma^f \subset E$ and $\mathcal{H}^{d-2}(E \setminus \Sigma^f) = 0$, and for which the chain rule holds in $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N; \mathbb{R}^d)$ with respect to every coordinate system $\{e_1, \ldots, e_d\}$ in $\mathcal{E}$.

(ii) It is clear that the above construction can be carried out for every $1 \leq k \leq d-2$. We considered the case $k = d-2$ because in a sense it represents the worst possible situation.

All the results presented so far depend on the particular choice of the coordinate system $\{e_1, \ldots, e_d\}$. We next address the case where the chain rule holds with respect to every coordinate system.

Theorem 4.12. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz continuous function, and let $1 \le k \le \min\{N, d\}$. Assume that the set Σ^f is not purely $\mathcal{H}^k$-unrectifiable. Then there exists a coordinate system in $\mathbb{R}^d$ for which the classical chain rule* (1.3) *fails in the class $\mathcal{A}_k(\mathbb{R}^N; \mathbb{R}^d)$. Hence a necessary and sufficient condition for the chain rule* (1.3) *to hold in $\mathcal{A}_k(\mathbb{R}^N; \mathbb{R}^d)$ with respect to every coordinate system is that Σ^f is purely $\mathcal{H}^k$-unrectifiable.*

Remark 4.13. Note that if $k = d-1$ then the result follows immediately from Theorem 4.1. Indeed, since (4.1) must hold for every coordinate system, it is clear that Σ^f must be purely $\mathcal{H}^{d-1}$-unrectifiable.

Proof of Theorem 4.12. The sufficiency part of the statement follows immediately from Theorem 4.1. The rest of the proof is devoted to showing the necessity part.

Step 1. We consider the case $k = 1$. Let $\gamma : [0,1] \to \mathbb{R}^d$ be a C^1 regular curve. For every $t \in [0,1]$ denote the unit tangent vector by $\tau(t) := \gamma'(t)/|\gamma'(t)|$. Fix $t_0 \in [0,1]$ and a unit vector e different from $\pm\tau(t_0)$ and not orthogonal to $\tau(t_0)$. Let $e^\perp$ be the hyperplane orthogonal to e and let $\sigma(t)$ be the unit vector obtained by projecting $\tau(t)$ on $e^\perp$ and normalizing.

We claim that for $\mathcal{L}^1$-a.e. t in a neighborhood I of t_0 the function f restricted to the plane $\gamma(t) + \operatorname{span}\{e, \tau(t)\}$ is differentiable at $\gamma(t)$.

Indeed, considering an orthonormal basis $\{e_1, \dots, e_{d-1}, e\}$, from the chain rule and the fact that $e \cdot \tau(t_0) \neq 0$ it follows that the partial derivative $\frac{\partial f}{\partial e}(\gamma(t))$ exists for $\mathcal{L}^1$-a.e. t near t_0. If $\tau(t) \equiv \tau(t_0)$ for all t near t_0 then since by Rademacher's theorem $\frac{\partial f}{\partial \tau(t_0)}(\gamma(t))$ exists for $\mathcal{L}^1$-a.e. t near t_0 the conclusion follows from Remark 3.2 applied to f restricted to the plane

$$\gamma(t) + \operatorname{span}\{e, \tau(t)\} = \gamma(t_0) + \operatorname{span}\{e, \tau(t_0)\}.$$

If $\tau(t)$ is not constant near t_0 then as in the proof of Theorem 3.1 we can reduce to the previous case by using a local diffeomorphism. We omit the details.

Hence the claim holds and therefore for $\mathcal{L}^1$-a.e. $t \in I$ we have

$$\frac{\partial f}{\partial \tau(t)}(\gamma(t)) = \frac{\partial f}{\partial e}(\gamma(t))(\tau(t)\cdot e) + \frac{\partial f}{\partial \sigma(t)}(\gamma(t))(\tau(t)\cdot\sigma(t)). \tag{4.27}$$

Moreover by the chain rule for every orthonormal basis $\{e_1, \dots, e_{d-1}, e\}$ and for $\mathcal{L}^1$-a.e. $t \in I$,

$$\begin{aligned}(f\circ\gamma)'(t) &= \frac{\partial f}{\partial e}(\gamma(t))(\gamma'(t)\cdot e) + \sum_{i=1}^{d-1}\frac{\partial f}{\partial e_i}(\gamma(t))(\gamma'(t)\cdot e_i)\\ &= \frac{\partial f}{\partial e}(\gamma(t))(\gamma'(t)\cdot e) + \sum_{i=1}^{d-1}\frac{\partial f}{\partial e_i}(\gamma(t))(\sigma(t)\cdot e_i)(\gamma'(t)\cdot\sigma(t)),\end{aligned}$$

or equivalently,

$$\frac{\partial f}{\partial \tau(t)}(\gamma(t)) = \frac{\partial f}{\partial e}(\gamma(t))(\tau(t)\cdot e) + \sum_{i=1}^{d-1}\frac{\partial f}{\partial e_i}(\gamma(t))(\sigma(t)\cdot e_i)(\tau(t)\cdot\sigma(t)).$$

Hence also by (4.27) for every orthonormal basis $\{e_1, \dots, e_{d-1}, e\}$ and for $\mathcal{L}^1$-a.e. $t \in I$ we have

$$\frac{\partial f}{\partial \sigma(t)}(\gamma(t)) = \sum_{i=1}^{d-1} \frac{\partial f}{\partial e_i}(\gamma(t))(\sigma(t) \cdot e_i). \tag{4.28}$$

Define the invertible linear transformation $L : \mathbb{R}^d \to \mathbb{R}^d$ by

$$L(u) := u + (u \cdot e)w, \tag{4.29}$$

where $w \in e^\perp$ is a unit vector different from $\pm\sigma(t_0)$ and let $f_L := f \circ L$, $\gamma_L := L^{-1} \circ \gamma$. Since the chain rule (1.3) holds in $W^{1,1}_{\text{loc}}(\mathbb{R}^N; \mathbb{R}^d)$ with respect to every coordinate system it is clear that it also holds for f_L with respect to every coordinate system. Therefore we can also assume that for every orthonormal basis $\{e_1, \dots, e_{d-1}, e\}$ and for $\mathcal{L}^1$-a.e. $t \in [0, 1]$,

$$\frac{\partial f_L}{\partial \sigma_L(t)}(\gamma_L(t)) = \sum_{i=1}^{d-1} \frac{\partial f_L}{\partial e_i}(\gamma_L(t))(\sigma_L(t) \cdot e_i),$$

where as before $\sigma_L(t)$ is the unit vector obtained by projecting the tangent unit vector $\tau_L(t)$ on $e^\perp$ and normalizing. Since by (4.29) for any $\nu \in e^\perp \cap S^{d-1}$ we have

$$\frac{\partial f_L}{\partial \nu}(\gamma_L(t)) = \frac{\partial f}{\partial \nu}(\gamma(t)),$$

the above identity can be rewritten as

$$\frac{\partial f}{\partial \sigma_L(t)}(\gamma(t)) = \sum_{i=1}^{d-1} \frac{\partial f}{\partial e_i}(\gamma(t))(\sigma_L(t) \cdot e_i). \tag{4.30}$$

Let $\mathcal{B}$ be a countable dense family of orthonormal bases in $e^\perp$, and, for every $t \in I$, let $\mathcal{B}(t)$ be the dense subfamily formed by all bases $\{e_1, \dots, e_{d-1}\} \in \mathcal{B}$ such that $\sigma(t) \cdot e_i \neq 0$ for $i = 1, \dots, d-1$. From (4.28) and from our conventions on the validity of the chain rule it follows that $(\partial f/\partial e_i)(\gamma(t))$ exists for $\mathcal{L}^1$-a.e. $t \in I$, for every $\{e_1, \dots, e_{d-1}\} \in \mathcal{B}(t)$, and for every $i = 1, \dots, d-1$. By (4.28) and (4.30) there exists a set $\mathcal{N} \subset I$ with $\mathcal{L}^1(\mathcal{N}) = 0$ such that for all $t \in I \setminus \mathcal{N}$ and for any two bases $\{e_1, \dots, e_{d-1}\}$ and $\{\epsilon_1, \dots, \epsilon_{d-1}\}$ in $\mathcal{B}(t)$ we have

$$\sum_{i=1}^{d-1} \frac{\partial f}{\partial e_i}(\gamma(t))(\sigma(t) \cdot e_i) = \sum_{i=1}^{d-1} \frac{\partial f}{\partial \epsilon_i}(\gamma(t))(\sigma(t) \cdot \epsilon_i),$$
$$\sum_{i=1}^{d-1} \frac{\partial f}{\partial e_i}(\gamma(t))(\sigma_L(t) \cdot e_i) = \sum_{i=1}^{d-1} \frac{\partial f}{\partial \epsilon_i}(\gamma(t))(\sigma_L(t) \cdot \epsilon_i).$$

Since for every t near t_0 the space $e^\perp$ is generated by vectors of the form $\sigma_L(t) - \sigma(t)$ for a suitable choice of linear maps L satisfying (4.29), it follows that

$$\sum_{i=1}^{d-1} \frac{\partial f}{\partial e_i}(\gamma(t))e_i = \sum_{i=1}^{d-1} \frac{\partial f}{\partial \epsilon_i}(\gamma(t))\epsilon_i$$

for all $t \in I \setminus \mathcal{N}$. As this holds for any pair of orthonormal bases in $\mathcal{B}(t)$, Proposition 2.6 shows that f restricted to the hyperplane $\gamma(t) + e^{\perp}$ is differentiable for all $t \in I \setminus \mathcal{N}$. By Remark 3.2 this implies that f is differentiable at $\gamma(t)$ for $\mathcal{L}^1$-a.e. t near t_0. Since this is true for any $t_0 \in [0, 1]$ a compactness argument allows us to conclude that f is differentiable at $\mathcal{H}^1$-a.e. point of γ. Given the arbitrariness of the curve γ we may conclude that Σ^f is purely $\mathcal{H}^1$-unrectifiable.

Step 2. If $1 < k < d - 1$, let $M \subset \mathbb{R}^d$ be a k-dimensional manifold such that

$$M \cap \Sigma^f \neq \emptyset$$

and let $\bar{u} \in M \cap \Sigma^f$. We claim that there exists $\varepsilon > 0$ such that f is differentiable for $\mathcal{H}^k$-a.e. $u \in B(\bar{u}; \varepsilon) \cap M$. Fix any $\bar{u} \in M \cap \Sigma^f$ and consider a local regular parametrization $\psi : D \subset \mathbb{R}^k \to M$ of class C^1 such that $\bar{u} \in \psi(D)$, where D is an open set and $\psi(0) = \bar{u}$. Let $e \in S^{d-1}$ be a vector transversal to $\mathrm{Tan}^k(M, \bar{u})$, that is, e is not orthogonal and it does not belong to $\mathrm{Tan}^k(M, \bar{u})$. Taking a smaller D if necessary, we may assume that $\frac{\partial \psi}{\partial x_1}(x)$ is not parallel to e for every $x \in D$. Let $e^{\perp}$ be the hyperplane orthogonal to e and let $\mathcal{B}$ be a countable dense family of orthonormal bases in $e^{\perp}$, and, for every $x \in D$, let $\mathcal{B}(x)$ be the dense subfamily composed of all bases $\{e_1, \ldots, e_{d-1}\} \in \mathcal{B}$ such that $\frac{\partial \psi}{\partial x_1}(x) \cdot e_i \neq 0$ for $i = 1, \ldots, d - 1$. Let $\sigma(x)$ be the vector obtained by projecting $\frac{\partial \psi}{\partial x_1}(x)$ on $e^{\perp}$ and normalizing. Define the invertible linear transformation $L : \mathbb{R}^d \to \mathbb{R}^d$ as

$$L(u) := u + (u \cdot e)w,$$

where $w \in e^{\perp}$ is a unit vector different from $\pm\sigma(0)$ and let $f_L := f \circ L$ and $\psi_L := L^{-1} \circ \psi$. Without loss of generality assume that

$$D = B_1(0; r) \times B_{k-1}(0; r),$$

where $r > 0$ is so small that $w \neq \pm\sigma(x)$ for every $x \in D$ and write $x = (x_1, x') \in \mathbb{R} \times \mathbb{R}^{k\text{-}1}$. Since f and f_L satisfy the chain rule in the class $\mathcal{A}_k(\mathbb{R}^k; \mathbb{R}^d)$ with respect to every coordinate system of the form $\{e_1, \ldots, e_{d-1}, e_d\}$ where $e_d := e$ and $\{e_1, \ldots, e_{d-1}\}$ is in $\mathcal{B}$, we have

$$\frac{\partial}{\partial x_1}(f \circ \psi)(x) = \sum_{i=1}^{d} \frac{\partial f}{\partial e_i}(\psi(x)) \frac{\partial \psi_i}{\partial x_1}(x),$$
$$\frac{\partial}{\partial x_1}(f_L \circ \psi)(x) = \sum_{i=1}^{d} \frac{\partial f_L}{\partial e_i}(\psi(x)) \frac{\partial \psi_i}{\partial x_1}(x)$$

for $\mathcal{L}^k$-a.e. $x \in D$, where $\frac{\partial f}{\partial e_i}(\psi(x))$ exist near $\bar{x} := \psi^{-1}(\bar{u})$ whenever $\{e_1, \ldots, e_{d-1}\} \in \mathcal{B}(x)$, as $\frac{\partial \psi_i}{\partial x_1}(x) \neq 0$ by the definition of $\mathcal{B}(x)$. By Fubini's theorem and using the fact that $\mathcal{B}$ is countable, there exists a set $\mathcal{M} \subset B_{k-1}(0; r)$ with $\mathcal{L}^{k\text{-}1}(\mathcal{M}) = 0$ such that for all $x' \in B_{k-1}(0; r) \setminus \mathcal{M}$ and for every coordinate system of the form $\{e_1, \ldots, e_{d-1}, e_d\}$, where $e_d := e$ and $\{e_1, \ldots, e_{d-1}\}$ is in $\mathcal{B}$, we have

$$\frac{\partial}{\partial x_1}(f \circ \psi)(x_1, x') = \sum_{i=1}^{d} \frac{\partial f}{\partial e_i}(\psi(x_1, x'))\frac{\partial \psi_i}{\partial x_1}(x_1, x'),$$
$$\frac{\partial}{\partial x_1}(f_L \circ \psi)(x_1, x') = \sum_{i=1}^{d} \frac{\partial f_L}{\partial e_i}(\psi(x_1, x'))\frac{\partial \psi_i}{\partial x_1}(x_1, x'),$$

for $\mathcal{L}^1$-a.e. $x_1 \in B_1(0; r)$. Fix $x' \in B_{k-1}(0; r) \setminus \mathcal{M}$ and consider the curve $\psi(\cdot, x')$. Reasoning as in the previous step we deduce from the above identities that for $\mathcal{L}^1$-a.e. $x_1 \in B_1(0; r)$ and for every coordinate system $\{e_1, \ldots, e_{d-1}\}$ in $\mathcal{B}$ we have

$$\frac{\partial f}{\partial \sigma(x_1, x')}(\psi(x_1, x')) = \sum_{i=1}^{d-1} \frac{\partial f}{\partial e_i}(\psi(x_1, x'))(\sigma(x_1, x') \cdot e_i),$$
$$\frac{\partial f_L}{\partial \sigma(x_1, x')}(\psi_L(x_1, x')) = \sum_{i=1}^{d-1} \frac{\partial f_L}{\partial e_i}(\psi_L(x_1, x'))(\sigma_L(x_1, x') \cdot e_i).$$

We may continue as in the previous step to conclude that for $\mathcal{L}^1$-a.e. $x_1 \in B_1(0; r)$ the function f is differentiable at $\psi(x_1, x')$. Since this is true for all $x' \in B_{k-1}(0; r) \setminus \mathcal{M}$, Fubini's theorem implies that f is differentiable at $\psi(x)$ for $\mathcal{L}^k$-a.e. $x \in D$. Hence f is differentiable $\mathcal{H}^k$-almost everywhere in $\psi(D)$. This concludes the proof. □

Remark 4.14. It is clear from the previous proof that for Σ^f to be purely $\mathcal{H}^k$-unrectifiable it is enough to assume that the chain rule (1.3) holds in $\mathcal{A}_k(\mathbb{R}^N; \mathbb{R}^d)$ with respect to a dense set of coordinate systems.

5. Chain rule in $BV(\Omega; \mathbb{R}^d)$

In this section we extend the results of the previous section to the space of functions of bounded variation. We refer to [4] for the definition and main properties. As already mentioned in the introduction a weak form of the chain rule in $BV_{\text{loc}}(\mathbb{R}^N; \mathbb{R}^d)$ was established by Ambrosio and Dal Maso in [2] for any Lipschitz continuous function $f : \mathbb{R}^d \to \mathbb{R}$ (see also [10] for a different proof in the scalar case $d = 1$).

We study here the classical chain rule. Since by a result of Alberti [1] the Cantor part of the distributional derivative of a function of bounded variation has rank one, to extend the results of the previous section we can only consider the case $k = 1$.

Theorem 5.1. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz continuous function, let $\{e_1, \ldots, e_d\}$ be a basis in $\mathbb{R}^d$, and assume that for every $\mathcal{H}^1$-rectifiable set $E \subset \Sigma^f$ and for $\mathcal{H}^1$-a.e. $u \in E$, there exists $i \in \{1, \ldots, d\}$ depending on u such that*

$$\text{Tan}^1(E, u) = \text{span}\{e_i\}. \tag{5.1}$$

Then for every $u \in BV_{\mathrm{loc}}(\mathbb{R}^N;\mathbb{R}^d)$ *and* $j = 1, \dots, N$ *we have*

$$D_j(f \circ u)\lfloor(\mathbb{R}^N \setminus S(u)) = \sum_{i=1}^{d} \frac{\partial f}{\partial e_i}(u_*) D_j u_i \lfloor(\mathbb{R}^N \setminus S(u)), \tag{5.2}$$

where $\frac{\partial f}{\partial e_i}(u_*) D_j u_i$ *is interpreted to vanish on sets where* $|D_j u_i|$ *vanishes, and*

$$D(f \circ u)\lfloor S(u) = (f(u^+) - f(u^-))\nu\lfloor S(u). \tag{5.3}$$

We present some preliminary results which extend to functions of bounded variation the lemmas of the previous section. The main difficulty is the treatment of the Cantor part of the distributional derivative.

The next two results are well known. We give their proofs for the convenience of the reader.

Lemma 5.2. *Let* $u : (a,b) \to \mathbb{R}^d$ *be a function of bounded variation. Then there exists a continuous function of bounded variation* $v : (a, b+1) \to \mathbb{R}^d$ *such that*

$$u(a,b) \subset v(a,b+1).$$

Proof. **Step 1.** Assume first that u is scalar-valued and monotone and let $I \subset (a,b)$ be any countable set such that

$$S(u) \subset I.$$

Write I as $I = \{t_n\}$ and define

$$s(t) := t + \sum_{t_n < t} \frac{1}{2^n}, \qquad t \in (a,b).$$

Then $s : (a,b) \to (a,b+1)$ is a one-to-one function whose discontinuity set is I. Let

$$I_n := [a_n, b_n], \qquad n \in \mathbb{N},$$

and

$$a_n := t_n + \sum_{t_j < t_n} \frac{1}{2^j}, \qquad b_n := t_n + \sum_{t_j \le t_n} \frac{1}{2^j},$$

and let $t = t(s)$ be the inverse function of s. Define

$$v(s) := \begin{cases} u(t(s)), & s \in (a,b+1) \setminus \bigcup I_n, \\ 2^n(u(t_n^+) - u(t_n^-))(s - b_n) + u(t_n^+), & s \in I_n = [a_n, b_n]. \end{cases}$$

Clearly

$$u(a,b) \subset v(a,b+1).$$

The function v is strictly monotone and hence of bounded variation. Moreover, since u is continuous on $(a,b) \setminus I$ and $t(s)$ is continuous on $(a,b+1) \setminus \bigcup I_n$ it follows that v is continuous on $(a,b+1) \setminus \bigcup I_n$. The function v is also continuous in the interior of each

I_n and thus it remains to check continuity at the endpoints of each I_n, but this follows immediately since

$$\lim_{s\to a_n^+} v(s) = \lim_{s\to a_n^+} 2^n(u(t_n^+) - u(t_n^-))(s - b_n) + u(t_n^+) = u(t_n^-),$$
$$\lim_{s\to a_n^-} v(s) = \lim_{s\to a_n^-} u(t(s)) = u(t_n^-)$$

and similarly for b_n. Thus v is continuous.

Step 2. In the general case where $u : (a, b) \to \mathbb{R}^d$ let $u = (u_1, \dots, u_d)$. For each $i = 1, \dots, d$ we may write u_i as $u_i = w_i - z_i$, where w_i, z_i are monotone functions. Let I be the union of points of discontinuity of w_i, z_i for all $i = 1, \dots, d$. Clearly $S(u) \subset I$. By Step 1 we may construct functions $\tilde{w}_i, \tilde{z}_i : (a, b+1) \to \mathbb{R}^d$, continuous and monotone, such that

$$w_i(a, b) \subset \tilde{w}_i(a, b+1), \qquad z_i(a, b) \subset \tilde{z}_i(a, b+1)$$

for all $i = 1, \dots, d$. The function $v := (\tilde{w}_1 - \tilde{z}_1, \dots, \tilde{w}_d - \tilde{z}_d)$ has all the desired properties. ◻

Corollary 5.3. *Let $u : I \to \mathbb{R}^d$ be a function of bounded variation, where I is an interval. Then $u(I)$ is $\mathcal{H}^1$-rectifiable.*

Proof. By the previous lemma there exists a continuous function of bounded variation $v : J \to \mathbb{R}^d$ such that $u(I) \subset v(J)$ for some interval J. The result now follows from Theorem 16 in [9]. ◻

As a consequence of the previous lemma we deduce the following result, which although not needed in the remainder of this section, is of interest in its own right since it completes Proposition 3.92(c) of [4].

Proposition 5.4. *Let E be a subset of $\mathbb{R}^d$. Then the following properties are equivalent:*

(i) *E is purely $\mathcal{H}^1$-unrectifiable;*
(ii) *$\mathcal{H}^1(E \cap w(\mathbb{R})) = 0$ for any $w \in BV_{\text{loc}}(\mathbb{R}; \mathbb{R}^d)$;*
(iii) *for all $N \in \mathbb{N}$ and for any $u \in BV_{\text{loc}}(\mathbb{R}^N; \mathbb{R}^d)$ the measure $|Du|$ vanishes on the set $u_*^{-1}(E) \cap (\mathbb{R}^N \setminus S(u))$, where $|Du|$, u_* and $S(u)$ are, respectively, the total variation of the distributional derivative Du, a specific representative and the jump set of u.*

Proof. We begin by showing that (i)⇒(ii). Let $w \in BV_{\text{loc}}(\mathbb{R}; \mathbb{R}^d)$ and consider $(a, b) \subset \mathbb{R}$. By the previous lemma there exists a continuous function of bounded variation $v : (a, b+1) \to \mathbb{R}^d$ such that

$$u(a, b) \subset v(a, b+1).$$

Since $\mathcal{H}^1(E \cap v(a, b+1)) = 0$ (see [9], [11]) the result follows.

To prove the implication (i)⇒(iii) let E be purely $\mathcal{H}^1$-unrectifiable and consider $u \in BV_{\mathrm{loc}}(\mathbb{R}^N;\mathbb{R}^d)$. We follow the proof of Lemma 2.1 in [15]. Fix $i \in \{1,\dots,N\}$. For every $x = (x_1,\dots,x_N) \in \mathbb{R}^N$ we denote by $x' \in \mathbb{R}^{N-1}$ the vector

$$x' = \begin{cases} (x_1,\dots,x_{i-1},x_{i+1},\dots,x_N) & \text{if } 1 < i < N, \\ (x_1,\dots,x_{N-1}) & \text{if } i = N, \\ (x_2,\dots,x_N) & \text{if } i = 1, \end{cases}$$

and we write, with an abuse of notation,

$$x = (x', x_i).$$

By Theorem 3.108 in [4] for $\mathcal{L}^{N-1}$-a.e. $x' \in \mathbb{R}^{N-1}$ the function

$$u^{x'}(t) := u(x',t), \qquad t \in \mathbb{R},$$

belongs to $BV_{\mathrm{loc}}(\mathbb{R};\mathbb{R}^d)$, and

$$v(t) := (u_*)^{x'}(t) = u_*(x',t), \qquad t \in \mathbb{R},$$

is a good representative for $u^{x'}$. By (ii),

$$\mathcal{H}^1(E \cap v(\mathbb{R})) = 0.$$

By Proposition 3.92(c) in [4],

$$|Dv|(M \cap (\mathbb{R} \setminus S(v))) = 0,$$

where $M := v^{-1}(E \cap v(\mathbb{R}))$ and so

$$|D(u_*)^{x'}|(M \cap (\mathbb{R} \setminus S((u_*)^{x'}))) = 0$$

for $\mathcal{L}^{N-1}$-a.e. $x' \in \mathbb{R}^{N-1}$. Since this is true for all $i = 1,\dots,N$ by Theorems 3.107 and 3.108 in [4] we have

$$|Du|(u_*^{-1}(E) \cap (\mathbb{R}^N \setminus S(u))) = 0. \qquad \square$$

Lemma 5.5. *Let $E \subset \mathbb{R}^d$ be $\mathcal{H}^1$-rectifiable and let $\alpha = 1,\dots,d$. Then the following two conditions are equivalent:*

(i) *for $\mathcal{H}^1$-a.e. $u \in E$,*

$$\mathrm{Tan}^1(E,u) = \mathrm{span}\{e_\alpha\}; \tag{5.4}$$

(ii) *for every $u \in BV_{\mathrm{loc}}(\mathbb{R}^N;\mathbb{R}^d)$,*

$$|Du_{\tilde{\alpha}}|(u_*^{-1}(E) \cap (\mathbb{R}^N \setminus S(u))) = 0.$$

Proof. By Corollary 4.7 we only need to show that (i) implies (ii). Assume (i), fix $u \in BV_{\mathrm{loc}}(\mathbb{R}^N;\mathbb{R}^d)$ and let E_1 be the set of $u \in E$ for which (5.4) does not hold. Then $\mathcal{H}^1(E_1) = 0$ and so by Proposition 3.92(c) in [4],

$$|Du|(u_*^{-1}(E_1) \cap (\mathbb{R}^N \setminus S(u))) = 0.$$

Let $\Gamma \subset \mathbb{R}^d$ be a C^1 curve such that

$$\mathcal{H}^1(\Gamma \cap (E \setminus E_1)) > 0.$$

By taking Γ small enough we may assume that there exists an open set $D \subset \mathbb{R}^d$ and a function $\Phi : D \to \mathbb{R}^{d-1}$ of class C^1 such that $D\Phi$ has rank $d-1$ for every $x \in D$ and

$$\Gamma = \{w \in D : \Phi(w) = 0\}.$$

By Proposition 3.92 in [4] the measure $D(\Phi \circ u)$ vanishes on all subsets of $u_*^{-1}(\Gamma) \cap (\mathbb{R}^N \setminus S(u)) \subset (\Phi \circ u)_*^{-1}(0) \cap (\mathbb{R}^N \setminus S(\Phi \circ u))$. By the classical chain rule in BV (see Theorem 3.96 in [4]) we have $D(\Phi \circ u) = \nabla\Phi(u_*)Du$, hence

$$D\Phi(u_*(x))\frac{dDu}{d|Du|}(x) = 0 \quad \text{for } |Du|\text{-a.e. } x \in u_*^{-1}(\Gamma) \cap (\mathbb{R}^N \setminus S(u)).$$

Hence

$$\frac{dDu}{d|Du|}(x) \in \mathrm{Tan}^1(\Gamma, u_*(x)) \quad \text{for } |Du|\text{-a.e. } x \in u_*^{-1}(\Gamma) \cap (\mathbb{R}^N \setminus S(u))$$

and by (i) we have

$$\frac{dDu_{\tilde{\alpha}}}{d|Du|}(x) = 0 \quad \text{for } |Du|\text{-a.e. } x \in u_*^{-1}(E) \cap (\mathbb{R}^N \setminus S(u)).$$

This shows (ii). □

Proof of Theorem 5.1. For the proof of (5.3) we refer to Step 2 of the proof of Theorem 3.96 in [4].

Step 1. Consider first the case $N = 1$. By Theorem 2.1 in [2] for $|Du|$-a.e. $x \in \mathbb{R} \setminus S(u)$ the restriction of the function f to the affine space

$$T_x^u := \left\{ y \in \mathbb{R}^d : y = u_*(x) + z\frac{dDu}{d|Du|}(x) \text{ for some } z \in \mathbb{R} \right\} \tag{5.5}$$

is differentiable at $u_*(x)$ and

$$D(f \circ u) = \nabla(f|_{T_x^u})(u_*)Du \quad \text{as measures on } \mathbb{R} \setminus S(u). \tag{5.6}$$

On $u_*^{-1}(\mathbb{R}^d \setminus \Sigma^f)$ the right hand side of (5.6) coincides with $\nabla f(u_*)Du$. Let

$$\Sigma_u := u_*(\mathbb{R}) \cap \Sigma^f.$$

By Corollary 5.3 the set Σ_u is $\mathcal{H}^1$-rectifiable. By (5.1) we may decompose it as

$$\Sigma_u = \bigcup_{i=1}^{d} \Sigma_i \cup \mathcal{N},$$

where

$$\mathrm{Tan}^1(\Sigma_u, u) = \mathrm{span}\{e_i\} \quad \text{for all } u \in \Sigma_i \quad \text{and} \quad \mathcal{H}^1(\mathcal{N}) = 0.$$

By Proposition 3.92 in [4] the measure Du vanishes on all subsets of $u_*^{-1}(\mathcal{N}) \cap (\mathbb{R} \setminus S(u))$ and so it is enough to show that for every fixed $i = 1, \ldots, d$ the chain rule holds on $u_*^{-1}(\Sigma_i) \cap (\mathbb{R} \setminus S(u))$. Since by Lemma 5.5,

$$\frac{dDu}{d|Du|}(x) \in \mathrm{Tan}^1(\Sigma_i, u_*(x)) = \mathrm{span}\{e_i\}$$

for $|Du|$-a.e. $x \in u_*^{-1}(\Sigma_i) \cap (\mathbb{R} \setminus S(u))$ it follows that for all $j \neq i$,

$$\frac{dDu_j}{d|Du|}(x) = 0 \tag{5.7}$$

for $|Du|$-a.e. $x \in u_*^{-1}(\Sigma_i) \cap (\mathbb{R} \setminus S(u))$, and, from (5.1), that

$$\nabla(f|_{T_x^u})(u_*(x)) = \frac{\partial f}{\partial e_i}(u_*(x))$$

for $|Du|$-a.e. $x \in u_*^{-1}(\Sigma_i) \cap (\mathbb{R} \setminus S(u))$, and hence from (5.6) we conclude that

$$D(f \circ u) = \frac{\partial f}{\partial e_i}(u_*)Du_i = \sum_{j=1}^{d} \frac{\partial f}{\partial e_j}(u_*)Du_j$$

on $u_*^{-1}(\Sigma_i) \cap (\mathbb{R} \setminus S(u))$, since $\frac{\partial f}{\partial e_j}(u_*)Du_j$ is interpreted to be zero whenever Du_j vanishes.

Step 2. The general case follows exactly from the previous step by a slicing argument entirely similar to Step 2 of the proof of Theorem 2.1 in [2]; we omit the details. □

From the previous theorem and Theorem 4.12 we deduce the following result:

Theorem 5.6. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz continuous function. Assume that the set Σ^f is not purely $\mathcal{H}^1$-unrectifiable. Then there exists a coordinate system in $\mathbb{R}^d$ for which the classical chain rule* (5.2) *fails in $BV_{\mathrm{loc}}(\mathbb{R}^N; \mathbb{R}^d)$. Hence a necessary and sufficient condition for the chain rule* (5.2) *to hold in $BV_{\mathrm{loc}}(\mathbb{R}^N; \mathbb{R}^d)$ with respect to every coordinate system is that Σ^f is purely $\mathcal{H}^1$-unrectifiable.*

Acknowledgments. The authors are grateful to an anonymous referee for many useful suggestions that significantly improved the paper. The research of G. Leoni was partially supported by the National Science Foundation under Grant No. DMS-0405423. The authors thank the Center for Nonlinear Analysis (NSF Grant No. DMS-9803791) for its support during the preparation of this paper.

References

[1] Alberti, G.: Rank one property for derivatives of functions with bounded variation. Proc. Roy. Soc. Edinburgh Sect. A **123**, 239–274 (1993) Zbl 0791.26008 MR 1215412

[2] Ambrosio, L., Dal Maso, G.: A general chain rule for distributional derivatives. Proc. Amer. Math. Soc. **108**, 691–702 (1990) Zbl 0685.49027 MR 0969514

[3] Ambrosio, L., De Lellis, C.: Existence of solutions for a class of hyperbolic systems of conservation laws in several space dimensions. Int. Math. Res. Not. Soc. no. 41 **2003**, 2205–2220 Zbl 1061.35048 MR 2000967

[4] Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Oxford Univ. Press, New York (2000) Zbl 0957.49001 MR 1857292

[5] Besicovitch, A. S.: On the fundamental geometrical properties of linearly measurable plane sets of points I. Math. Ann. **98**, 422–464 (1928) JFM 53.0175.04 MR 1512414, II, Math. Ann. **115**, 296–329 (1938) Zbl 0018.11302 MR 1513189, III, Math. Ann. **116**, 349–357 (1939) JFM 65.0197.04 MR 1513231

[6] Bessis, D. N., Clarke, F. H.: Partial subdifferentials, derivates and Rademacher's theorem. Trans. Amer. Math. Soc. **351**, 2899–2926 (1999) Zbl 0924.49013 MR 1475676

[7] Boccardo, L., Murat, F.: Remarques sur l'homogénéisation de certains problèmes quasilinéaires. Portugal. Math. **41**, 535–562 (1982) Zbl 0524.35042 MR 0766874

[8] Bouchut, F.: Renormalized solutions to the Vlasov equation with coefficients of bounded variation. Arch. Ration. Mech. Anal. **157**, 75–90 (2001) Zbl 0979.35032 MR 1822415

[9] Choquet, G.: Application des propriétés descriptives de la fonction contingent à la théorie des fonctions de variable réelle et à la géométrie différentielle des variétés cartésiennes. J. Math. Pures Appl. **26**, 115–226 (1947) Zbl 0035.24201 MR 0023897

[10] Dal Maso, G., Lefloch, P. G., Murat, F.: Definition and weak stability of nonconservative products. J. Math. Pures Appl. **74**, 483–548 (1995) Zbl 0853.35068 MR 1365258

[11] De Cicco, V., Leoni, G.: A chain rule in $L^1(\mathrm{div};\Omega)$ and its applications to lower semicontinuity. Calc. Var. Partial Differential Equations **19**, 23–51 (2003) Zbl 1056.49019 MR 2027846

[12] Federer, H.: The (ϕ, k)-rectifiable subsets of n space. Trans. Amer. Math. Soc. **62**, 114–192 (1947) Zbl 0032.14902 MR 0022594

[13] Federer, H.: Geometric Measure Theory. Springer (1969) Zbl 0176.00801 MR 0257325

[14] Giaquinta, M., Modica, G., Souček, J.: Cartesian Currents in the Calculus of Variations I. Cartesian currents. Ergeb. Math. Grenzgeb. 37, Springer, Berlin (1998) Zbl 0914.49001 MR 1645086

[15] Marcus, M., Mizel, V. J.: Absolute continuity on tracks and mappings of Sobolev spaces. Arch. Ration. Mech. Anal. **45**, 294–320 (1972) Zbl 0236.46033 MR 0338765

[16] Marcus, M., Mizel, V. J.: Continuity of certain Nemitsky operators on Sobolev spaces and the chain rule. J. Anal. Math. **28**, 303–334 (1975) Zbl 0328.46028 MR 0482444

[17] Marcus, M., Mizel, V. J.: Complete characterization of functions which act, via superposition, on Sobolev spaces. Trans. Amer. Math. Soc. **251**, 187–218 (1979) Zbl 0417.46035 MR 0531975

[18] Mattila, P.: Geometry of Sets and Measures in Euclidean Spaces. Fractals and Rectifiability. Cambridge Stud. Adv. Math. 44, Cambridge Univ. Press, Cambridge (1995) Zbl 0819.28004 MR 1333890

[19] Murat, F., Trombetti, C.: A chain rule formula for the composition of a vector-valued function by a piecewise smooth function. Boll. Unione Mat. Ital. Sez. B Artic. Ric. Mat. (8) **6**, 581–595 (2003) MR 2014820

[20] Saks, S.: Theory of the Integral. 2nd rev. ed., English transl. by L. C. Young, G. E. Stechert & Co., New York (1937) Zbl 0017.30004 MR 0167578

[21] Serrin, J.: unpublished

[22] Serrin, J., Varberg, D.: A general chain rule for derivatives and the change of variables formula for the Lebesgue integral. Amer. Math. Monthly **76**, 514–520 (1969) Zbl 0175.34401 MR 0247011

[23] Stampacchia, G.: Le problème de Dirichlet pour les équations elliptiques du second ordre à coefficients discontinus. Ann. Inst. Fourier (Grenoble) **15**, 189–258 (1965) Zbl 0151.15401 MR 0192177

[24] Stein, E. M.: Singular Integrals and Differentiability Properties of Functions. Princeton Math. Ser. 30, Princeton Univ. Press, Princeton, NJ (1970) Zbl 0207.13501 MR 0290095

[25] White, B.: A new proof of Federer's Structure Theorem for k-dimensional subsets of $\mathbb{R}^N$. J. Amer. Math. Soc. **11**, 693–701 (1998) Zbl 0904.28004 MR 1603866