Journal of Convex Analysis
Volume 14 (2007), No. 2, 433–454

Infinite Dimensional Clarke Generalized Jacobian

Zsolt Páles*
Institute of Mathematics, University of Debrecen,
4010 Debrecen, Pf. 12, Hungary
pales@math.klte.hu

Vera Zeidan†
Department of Mathematics, Michigan State University,
East Lansing, MI 48824, USA
zeidan@math.msu.edu

Received: November 22, 2005
Revised manuscript received: January 11, 2006

In this paper we extend for a locally Lipschitz function the notion of Clarke's generalized Jacobian to the setting where the domain lies in an infinite dimensional normed space. When the function is real-valued this notion reduces to the Clarke's generalized gradient. Using this extension, we obtain an exact smooth-nonsmooth chain rule from which the sum rule and the product rule follow. Also an exact formula for the generalized Jacobian of piecewise differentiable functions will be provided.

Keywords: Generalized Jacobian, chain rule, sum rule, piecewise smooth functions

2000 Mathematics Subject Classification: 49A52, 58C20

1. Introduction

The field of nonsmooth analysis eminated from the theory of convex analysis that was developed by Rockafellar in [30].

In his pioneering work [8] Clarke introduced the *generalized gradient* for a locally Lipschitz function $f : \mathcal{D} \to \mathbb{R}$ defined on an open subset $\mathcal{D}$ of a finite dimensional space X as

$$\partial^c f(p) := \operatorname{co}\big\{x^* \in X^* \mid \exists\, (x_i)_{i\in\mathbb{N}} \text{ in } \Omega(f) : \lim_{i\to\infty} x_i = p \text{ and } \lim_{i\to\infty} Df(x_i) = x^*\big\}, \tag{1}$$

where $\Omega(f) \subseteq \mathcal{D}$ is the set of points where f is differentiable, which by Rademacher's celebrated theorem, is of full measure in $\mathcal{D}$. Note that since f is locally Lipschitz over a finite dimensional domain, then the notions of Fréchet- and Gâteaux-differentiability are equivalent.

In the same paper, the following characterization was obtained

$$\partial^c f(p) := \big\{x^* \in X^* \mid f^\circ(p,h) \geq \langle x^*, h\rangle,\ \forall h \in X\big\}, \tag{2}$$

where

$$f^\circ(p,h) := \limsup_{x\to p,t\downarrow 0} \frac{f(x+th)-f(x)}{t}. \tag{3}$$

*Research of the first author is supported by the Hungarian Scientific Research Fund (OKTA) under grant K-62316.
†Research of the second author is supported by the National Science Foundation under grant DMS-0306260.

ISSN 0944-6532 / $ 2.50

It was then observed in [10] that equation (2) can be used to introduce the generalized gradient when X is an infinite dimensional space.

On the other hand, when X and Y are finite dimensional normed spaces and $f : \mathcal{D} \to Y$ is a *vector-valued* locally Lipschitz function Clarke extended in [9], [11] the notion defined by (1) and introduced a *generalized Jacobian* as

$$\partial^c f(p) := \mathrm{co}\big\{A \in \mathcal{L}(X,Y) \mid \exists\,(x_i)_{i\in\mathbb{N}} \text{ in } \Omega(f) : \lim_{i\to\infty} x_i = p \text{ and } \lim_{i\to\infty} Df(x_i) = A\big\}. \quad (4)$$

Based on the Rademacher's theorem another but related approach was proposed by Pourciau in [27].

When X is infinite dimensional, derivative-like objects have been the focus of research since the late seventies. For instance, Warga introduced in [37], [38] the notion of derivate containers, Halkin defined the concepts of screens and "fans" in [15], [14]. Sweetser [33], [34] also considered the concept of shields. Thibault employed the limit points of directional difference quotients in [35] (see (35) below).

The idea of defining derivative-like objects via various tangent cones to the graph of the function goes back to Aubin in [2], a summary of this approach can be found in [3]. A related notion, the *fan derivative* $D^\circ f(p) : X \to 2^Y$, was introduced by Ioffe in [16] as

$$D^\circ f(p)(h) := \big\{y \in Y \mid (y^* \circ f)^\circ(p,h) \geq \langle y^*, y\rangle,\ \forall y^* \in Y^*\big\}, \quad (h \in X), \quad (5)$$

where $(y^* \circ f)^\circ$ is the Clarke generalized directional derivative of the *real-valued* map $y^* \circ f$ defined in (3). It was also shown in [16] that if X is finite dimensional then, for all $h \in X$,

$$D^\circ f(p)(h) = \partial^c f(p)(h),$$

i.e., the fan derivative $D^\circ f(p)$ is generated by the elements of the Jacobian $\partial^c f(p)$.

On the other hand, in 1973 Clarke showed in his thesis [7] and later in [8] that the generalized gradient can be defined via normal cones. This idea, namely, to involve normal cones as opposed to tangent cones, led to the notion of *coderivatives* introduced by Mordukhovich in [21], [22] for the finite dimensional setting and in [23] when both the domain and the range spaces are infinite dimensional. For $f : X \to Y$ the coderivative, $D^* f(p) : Y^* \to X^*$, of f at p is defined as

$$D^* f(p)(y^*) := \big\{x^* \in X^* \mid (x^*, -y^*) \in N\big((p, f(p)); \mathrm{graph}\, f\big)\big\}, \quad (y^* \in Y^*), \quad (6)$$

where $N\big((p, f(p)); \mathrm{graph}\, f\big)$ denotes Mordukhovich's normal cone to the graph of f at $(p, f(p))$.

When X is finite dimensional, it was shown in [23] that, for all $y^* \in Y^*$,

$$\mathrm{co}\, D^* f(p)(y^*) = y^* \circ \partial^c f(p).$$

A relatively recent survey on the different subdifferentials and their properties is given in [5] where also an extended list of references could be found.

Except Halkin's screen and Sweetser's shields, these derivative-like objects are not given in terms of a relevant set of linear operators. However, as clearly stated by Ioffe in

1981, [16]: "One can expect that the derivatives defined by way of operators have some additional good properties. Therefore the question ... seems to be important." Such good properties should definitely be translated as having "tight" calculus rules in terms of *linear operators* and computational utility in the applications. To our knowledge, this question has not been answered in a satisfactory manner.

A possible approach to answer Ioffe's question could be to define, for any normed space X, a generalized Jacobian as the set of linear operators

$$\begin{aligned}\partial^G f(p) := \overline{\mathrm{co}}\, \big\{ \Phi \in \mathcal{L}(X,Y) \,|\, \exists\, (x_i)_{i\in\mathbb{N}} \text{ in } \mathcal{D} \text{ such that } f \text{ is Gâteaux differentiable at } x_i, \\ \lim_{i\to\infty} x_i = p, \text{ and } \Phi \text{ is a } w^*\text{-cluster point of } \big(Df(x_i)\big)_{i\in\mathbb{N}} \big\}.\end{aligned} \tag{7}$$

In general, $\partial^G f(p)$ could be empty. However, if X is a *separable Banach* space, then, by the results of Aronszajn [1], Christensen [6], Mankiewicz [20], and Phelps [26] (see the book [4] for a summary of further results), a locally Lipschitz function f is differentiable in the sense of Gâteaux on a dense subset $\Omega_X(f)$ of $\mathcal{D}$ (whose complement is a Gauss-null set). Therefore, in this special case, $\partial^G f(p)$ is nonempty. Hence, by the weak*-compactness of the ball $\overline{B}_{\mathcal{L}(X,Y)}$, it follows that $\partial^G f(p)$ is a weak*-compact and nonempty subset of the space $\mathcal{L}(X,Y)$. A similar approach was followed by Thibault in [36] where he introduced generalized Jacobians that also depend on a Haar-null set of $\mathcal{D} \subseteq X$, where X is a separable Banach space. Furthermore, a smooth-nonsmooth chain rule for Thibault's generalized Jacobians was obtained in [36].

It is an open problem whether the approach using $\partial^G f(p)$ would obtain results parallel to those in this paper. The main obstacle resides in the absence of properties that lead to a useful chain rule and further calculus rules.

The aim of this paper is to answer the question posed by Ioffe in the affirmative. We provide a satisfactory extension of Clarke's approach of generalized Jacobian (4) to locally Lipschitz functions from *any normed* space X into a *finite* dimensional normed space Y. Our generalized Jacobian, $\partial f(p)$, is defined to be a *set of linear operators* from X to Y. When X is finite dimensional our generalized Jacobian *coincides* with the Clarke's generalized Jacobian (4). On the other hand, when the domain is infinite dimensional and the image space is $\mathbb{R}$, $\partial f(p)$ turns out to be exactly equal to Clarke's generalized gradient (2). Consequently, the fan generated by our generalized Jacobian is Ioffe's fan derivative, and the fan generated by the adjoint of $\partial f(p)$ is the closed convex hull of Mordukhovich's coderivative.

Throughout this paper, let X and Y denote normed spaces and always assume that Y and all the image spaces are *finite* dimensional. As usual, let X^* and Y^* stand for their topological dual spaces, respectively. The open and closed unit balls in any normed space Z in the paper will be denoted by B_Z and $\overline{B}_Z$, respectively. If Y is of dimension n, then the space of continuous linear maps $\mathcal{L}(X,Y)$ is isomorphic to the product space $(X^*)^n$, therefore, when speaking about the weak*-topology in $\mathcal{L}(X,Y)$, we refer to the product weak*-topology in $(X^*)^n$.

Let $\mathcal{D}$ be a nonempty open subset of X, p be an arbitrary point in $\mathcal{D}$ and let $f : \mathcal{D} \to Y$ be a locally Lipschitz function. Given a linear subspace $L \subseteq X$, the function f is called *L-(Gâteaux)-differentiable* at a point p if there exists a continuous linear map $D_L f(p)$:

$L \to Y$ such that, for all $h \in L$,

$$D_L f(p)(h) = f'(p,h) := \lim_{t\to 0} \frac{f(p+th)-f(p)}{t}. \tag{8}$$

We denote by $\Omega_L(f)$ the set of those points p in $\mathcal{D}$ where f is L-differentiable.

Clearly, the X-differentiability is equivalent to the standard Gâteaux-differentiability. In this case, the subscript X will be omitted from the notation, i.e., $D_X f(p)$ will simply be denoted by $Df(p)$. On the other hand, if $L = \operatorname{lin}\{h\} = \langle h\rangle$ is the linear span of a nonzero vector $h \in X$, then the L-differentiability of f means that the two-sided directional derivative, $f'(p,h)$, of f at p in the direction h exists and

$$D_{\langle h\rangle} f(p)(h) = f'(p,h). \tag{9}$$

It is also immediate to see that if f is L-differentiable at $p \in \mathcal{D}$, then

$$\|D_L f(p)\| \le \ell_f(p), \tag{10}$$

where the Lipschitz modulus $\ell_f(p)$ of f at p is defined as

$$\ell_f(p) := \inf_{t>0} \sup \left\{ \frac{\|f(x)-f(y)\|}{\|x-y\|} \;\middle|\; x,y \in p+tB_X,\ x \ne y \right\}. \tag{11}$$

Another important observation is that if K, L are linear subspaces of X with $K \subseteq L$ and f is L-differentiable at p then f is also K-differentiable at p and

$$D_L f(p)|_K = D_K f(p). \tag{12}$$

In the sequel, denote by $\Lambda(X)$ the collection of all finite dimensional subspaces of X. If $L \in \Lambda(X)$ then the function $g : L \cap (\mathcal{D}-p) \to Y$ defined by $g(u) := f(p+u)$ is locally Lipschitz at $u = 0$, therefore, by Rademacher's theorem, g is almost everywhere differentiable in a neighborhood of $u = 0$. Thus, in view of the estimate (10), there exist a sequence $u_i \in L$ such that $u_i \to 0$ and the sequence $Dg(u_i) = D_L f(p+u_i)$ converges. Based on this observation, we introduce the *pre L-Jacobian* and the *L-Jacobian* of f via the following formulae:

$$\begin{aligned} &\Delta_L f(p) := \bigl\{A \in \mathcal{L}(L,Y) \mid \exists\, (x_i)_{i\in\mathbb{N}} \text{ in } \Omega_L(f) : \lim_{i\to\infty} x_i = p \text{ and } \lim_{i\to\infty} D_L f(x_i) = A\bigr\}, \\ &\partial_L f(p) := \operatorname{co} \Delta_L f(p). \end{aligned} \tag{13}$$

Note that here the sequence (x_i) is not necessarily contained in the affine subspace $p+L$, and hence, $\partial_L f(p)$ could be larger than Clarke's generalized Jacobian of the restricted function $f|_{p+L}$ at p, which is $\partial^c g(0)$, where g is the function defined above.

Clearly, $\Delta_L f(p)$ is a nonempty compact and $\partial_L f(p)$ is a nonempty compact convex set of the finite dimensional space $\mathcal{L}(L,Y)$. Now we are able to introduce the *generalized Jacobian* which is the main object of this paper:

$$\partial f(p) := \bigl\{\Phi \in \mathcal{L}(X,Y) : \Phi|_L \in \partial_L f(p),\ \forall L \in \Lambda(X)\bigr\}. \tag{14}$$

As a technical tool, we shall also need the so called *intermediate* L-Jacobian defined by

$$\widehat{\partial}_L f(p) := \bigl\{\Phi \in \mathcal{L}(X,Y) : \Phi|_L \in \partial_L f(p)\bigr\}. \tag{15}$$

The importance of the intermediate L-Jacobians lies in the fact that our generalized Jacobian is the intersection of all the intermediate L-Jacobians.

$$\partial f(p) = \bigcap_{L \in \Lambda(X)} \widehat{\partial}_L f(p). \tag{16}$$

If X is finite dimensional then, by (12), we have that $\Delta_X f(p)\big|_L \subseteq \Delta_L f(p)$. This yields $\partial_X f(p)\big|_L \subseteq \partial_L f(p)$, whence

$$\widehat{\partial}_X f(p) \subseteq \widehat{\partial}_L f(p).$$

Thus, by (16),

$$\partial f(p) = \bigcap_{L \in \Lambda(X)} \widehat{\partial}_L f(p) = \widehat{\partial}_X f(p) = \partial_X f(p). \tag{17}$$

However $\partial_X f(p)$ coincides with Clarke's generalized Jacobian. Therefore, the generalized Jacobian defined above extends Clarke's generalized Jacobian to the general normed space setting.

The structure of this paper is organized as follows. In Section 2, the most basic properties of our generalized Jacobian are obtained. A smooth-nonsmooth exact chain rule and its consequences, the sum rule, the product rule, are derived in Section 3. We provide a comparison between our generalized Jacobian and each of Clarke's generalized gradient, (in the real-valued case), Ioffe's fan derivative, Mordukhovich's coderivative, and Thibault's limit points in Section 4. In the last section, we develop an equation for the generalized Jacobian of "piecewise differentiable" functions which is a generalization of the result obtained for the finite dimensional setting by Scholtes in [32] and by Kuntz and Scholtes in [19].

2. Properties of the Generalized Jacobian

Our first result summarizes the most basic properties of our generalized Jacobian.

Theorem 2.1. *Let $f : \mathcal{D} \to Y$ be a locally Lipschitz function. Then we have the following properties for $\partial f(p)$ defined in* (14).

- (*i*) *For all $p \in \mathcal{D}$, the generalized Jacobian $\partial f(p)$ is nonempty, w^*-compact, and convex.*
- (*ii*) *For all $p \in \mathcal{D}$, $\partial f(p)$ is a subset of the ball $\ell_f(p)\overline{B}_{\mathcal{L}(X,Y)}$, where $\ell_f(p)$ is defined by* (11).
- (*iii*) *If f is Gâteaux-differentiable at $p \in \mathcal{D}$, then its Gâteaux-derivative $Df(p)$ is in $\partial f(p)$.*

Proof. Let ρ be any number bigger than $r := \ell_f(p)$. Then, by (11) there exists $t > 0$, such that f is Lipschitz continuous on the ball $p + tB_X$ with ρ as a Lipschitz constant.

Let $h \in X$ be fixed and let $L = \langle h \rangle$. Then, by (10), $\|D_{\langle h \rangle} f(x)\| \leq \ell_f(x) \leq \rho$ holds for all $x \in \Omega_{\langle h \rangle}(f) \cap (p + tB_X)$.

Let $\Phi \in \partial f(p)$ be arbitrary. Then $\Phi(h) \in \partial_{\langle h \rangle} f(p)(h)$, whence

$$\Phi(h) \in \operatorname{co}\Big\{ \lim_{i \to \infty} D_{\langle h \rangle} f(x_i)(h) \,\Big|\, \exists\, (x_i)_{i \in \mathbb{N}} \text{ in } \Omega_{\langle h \rangle}(f) \cap (p + tB_X) \,:\, \lim_{i \to \infty} x_i = p \Big\}.$$

Thus $\|\Phi(h)\| \leq \rho\|h\|$ follows for all $h \in X$. Since $\rho > r$ was arbitrary, we get $\|\Phi\| \leq r$. This shows that $\partial f(p)$ is bounded and is contained in the ball $r\overline{B}_{\mathcal{L}(X,Y)}$. Thus, (ii) holds.

Let $y_1, \ldots, y_n$ be a basis for Y and, for $y \in Y$, denote by $\alpha_1(y), \ldots, \alpha_n(y)$ the coordinates of y with respect to this basis. Then $\alpha := (\alpha_1, \ldots, \alpha_n) : Y \to \mathbb{R}^n$ is a continuous linear isomorphism. Then, there exist positive constants $0 < \beta \leq \gamma$ such that

$$\beta\|y\| \leq \max\big(|\alpha_1(y)|, \ldots, |\alpha_n(y)|\big) \leq \gamma\|y\|, \quad (y \in Y).$$

Next, we show that, for any $L \in \Lambda(X)$, the intersection $\widehat{\partial}_L f(p) \cap \rho\overline{B}_{\mathcal{L}(X,Y)}$ is nonempty, convex and weak*-compact, where $\rho := \frac{\gamma r}{\beta} \geq r = \ell_f(p)$.

To verify the nonemptiness of $\widehat{\partial}_L f(p) \cap \rho\overline{B}_{\mathcal{L}(X,Y)}$, let $A : L \to Y$ be an arbitrary element of $\partial_L f(p)$. Then $\alpha_1 \circ A, \ldots, \alpha_n \circ A$ are linear functionals on L with norm not greater than $\gamma\|A\| \leq \gamma r$. By the Hahn–Banach extension theorem, there exist linear functionals $x_1^*, \ldots, x_n^* \in X^*$ such that

$$x_k^*|_L = \alpha_k \circ A, \quad \|x_k^*\| \leq \gamma r \quad (k = 1, \ldots, n).$$

Then the linear map Φ defined by $\Phi = \alpha^{-1} \circ (x_1^*, \ldots, x_n^*)$ is obviously a continuous linear extension of A, hence it is an element of $\widehat{\partial}_L f(p)$. On the other hand, for all $h \in X$,

$$\beta\|\Phi(h)\| \leq \max\big(|\alpha_1(\Phi(h))|, \ldots, |\alpha_n(\Phi(h))|\big) = \max\big(|x_1^*(h)|, \ldots, |x_n^*(h)|\big) \leq \gamma r\|h\|,$$

hence $\|\Phi\| \leq \frac{\gamma r}{\beta} = \rho$. Thus $\Phi \in \rho\overline{B}_{\mathcal{L}(X,Y)}$ also holds.

The convexity of $\widehat{\partial}_L f(p) \cap \rho\overline{B}_{\mathcal{L}(X,Y)}$ is obvious. To see that $\widehat{\partial}_L f(p)$ is weak*-closed, let Φ be the weak*-cluster point of a net Φ_ι in $\widehat{\partial}_L f(p)$. Then $\Phi|_L$ is also a cluster point of the net $\Phi_\iota|_L \in \partial_L f(p)$. The set $\partial_L f(p)$ being compact, we get that $\Phi|_L \in \partial_L f(p)$ whence $\Phi \in \widehat{\partial}_L f(p)$ follows. Thus, by the Banach–Alaoglu theorem, the intersection $\widehat{\partial}_L f(p) \cap \rho\overline{B}_{\mathcal{L}(X,Y)}$ is also weak*-compact.

Now we prove that the family $\big\{\widehat{\partial}_L f(p) \cap \rho\overline{B}_{\mathcal{L}(X,Y)} : L \in \Lambda(X)\big\}$ satisfies the finite intersection property.

If $K, L \in \Lambda(X)$ with $K \subseteq L$, then by (12) we get that $\Delta_L f(p)\big|_K \subseteq \Delta_K f(p)$ holds. This yields $\partial_L f(p)\big|_K \subseteq \partial_K f(p)$, whence

$$\widehat{\partial}_L f(p) \subseteq \widehat{\partial}_K f(p) \tag{18}$$

follows for all $K, L \in \Lambda(X)$ with $K \subseteq L$.

If $L_1, \ldots, L_k \in \Lambda(X)$ are arbitrary subspaces then, it follows from the inclusions $L_i \subseteq L_1 + \cdots + L_k$ that

$$\emptyset \neq \widehat{\partial}_{L_1+\cdots+L_k} f(p) \cap \rho\overline{B}_{\mathcal{L}(X,Y)} \subseteq \bigcap_{i=1}^{k} \Big(\widehat{\partial}_{L_i} f(p) \cap \rho\overline{B}_{\mathcal{L}(X,Y)}\Big),$$

Thus, indeed, the family $\big\{\widehat{\partial}_L f(p) \cap \rho\overline{B}_{\mathcal{L}(X,Y)} : L \in \Lambda(X)\big\}$ satisfies the finite intersection property. Hence

$$\emptyset \neq \bigcap_{L \in \Lambda(X)} \Big(\widehat{\partial}_L f(p) \cap \rho\overline{B}_{\mathcal{L}(X,Y)}\Big).$$

To complete the proof of the nonemptiness of $\partial f(p)$ and condition (i), it now suffices to observe that

$$\bigcap_{L\in\Lambda(X)} \Big(\widehat{\partial}_L f(p) \cap \rho\overline{B}_{\mathcal{L}(X,Y)}\Big) = \partial f(p) \cap \rho\overline{B}_{\mathcal{L}(X,Y)} = \partial f(p)$$

and that the left hand side is w^*-compact, convex.

Finally, assume that f is Gâteaux-differentiable at p. Then, due to the local Lipschitz property of f, for all $L \in \Lambda(X)$, the restriction $Df(p)|_L$ is the L-Gâteaux-derivative of f at p, whence $Df(p)|_L \in \Delta_L f(p)$. Thus $Df(p) \in \partial f(p)$ follows. $\square$

Theorem 2.2. *Let $f : \mathcal{D} \to Y$ be a locally Lipschitz function. Then, the map $p \mapsto \partial f(p)$ is upper semicontinuous on $\mathcal{D}$ in the following sense: If a sequence $(p_i)_{i\in\mathbb{N}}$ in $\mathcal{D}$ tends to p, an element $\Phi \in \mathcal{L}(X,Y)$ is a weak*-cluster point of a sequence $(\Phi_i)_{i\in\mathbb{N}}$ and $\Phi_i \in \partial f(p_i)$ holds for all i, then $\Phi \in \partial f(p)$.*

Proof. Let $(p_i)_{i\in\mathbb{N}}$ be a sequence of points, $\Phi_i \in \partial f(p_i)$ be a sequence of linear maps and let $\Phi \in \mathcal{L}(X,Y)$ a weak*-cluster point of the sequence $(\Phi_i)_{i\in\mathbb{N}}$.

Let $L \in \Lambda(X)$ be arbitrarily fixed. Then there is a subsequence (Φ_{i_k}) such that $\big(\Phi_{i_k}|_L\big)$ tends to $\Phi|_L$ in the (finite dimensional) space $\mathcal{L}(L,Y)$ as $k \to \infty$. On the other hand, for all $k \in \mathbb{N}$, we have that $\Phi_{i_k}|_L \in \partial_L f(p_{i_k})$. By Carathéodory's theorem and by the definition of the L-Jacobian $\partial_L f(p_{i_k})$, for all $k \in \mathbb{N}$, there exist nonnegative numbers $\lambda_{k0}, \dots, \lambda_{km}$ with $\lambda_{k0} + \cdots + \lambda_{km} = 1$ and points $u_{k0}, \dots, u_{km} \in \Omega_L(f) \cap \big(p_{i_k} + (1/k)B_X\big)$ (where m stands for the dimension of the space $\mathcal{L}(L,Y)$) such that

$$\Big\|\Phi_{i_k}|_L - \sum_{j=0}^{m} \lambda_{kj} D_L f(u_{kj})\Big\| < \frac{1}{k}. \tag{19}$$

By taking a subsequence if necessary, we may assume that the sequences

$$\big(D_L f(u_{k0}), \dots, D_L f(u_{km})\big) \in \mathcal{L}(L,Y) \quad \text{and} \quad (\lambda_{k0}, \dots, \lambda_{km}) \in \mathbb{R}^{m+1}$$

converge and the latter tends to $(\lambda_0, \dots, \lambda_m)$. Observe also that the sequences u_{kj} tend to p for all $j \in \{0, \dots, m\}$. Thus, taking the limit $k \to \infty$ in (19), we get that

$$\Phi|_L = \lim_{k\to\infty} \Phi_{i_k}|_L = \sum_{j=0}^{m} \lambda_j \lim_{k\to\infty} D_L f(u_{kj}) \in \sum_{j=0}^{m} \lambda_j \Delta_L f(p) \subseteq \partial_L f(p)$$

holds for all $L \in \Lambda(X)$. Hence, $\Phi \in \partial f(p)$ and thus the theorem is proved. $\square$

Remark 2.3. From the proof of Theorem 2.2 the following useful fact about $\partial_L f$ is readily obtained: *If $p_i \to p$, $A_i \in \partial_L f(p_i)$, and $A_i \to A$ in $\mathcal{L}(L,Y)$ then, $A \in \partial_L f(p)$.*

As consequence of Theorem 2.1 and Theorem 2.2, we have the following result.

Corollary 2.4. *Let $f : \mathcal{D} \to Y$ be a locally Lipschitz function. Then, for all $p \in \mathcal{D}$,*

$$\partial^G f(p) \subseteq \partial f(p) \tag{20}$$

where $\partial^G f(p)$ is defined in (7).

Proof. Let $\Phi \in \mathcal{L}(X,Y)$ such that there exists a sequence $(x_i)_{i\in\mathbb{N}}$ in $\Omega_X(f)$ converging to p and Φ is weak*-cluster point of the sequence $(Df(x_i))$. We have that $Df(x_i) \in \partial f(x_i)$ for all i. Therefore, by Theorem 2.2, Φ must belong to $\partial f(p)$. Now, using the convexity and the weak*-closedness of $\partial f(p)$, the statement follows. □

Recall that when X is a separable Banach space, then $\partial^G f(p)$ is nonempty. It is undecided whether, in this case, the reversed inclusion in (20) is valid. However, in the case of piecewise smooth functions, we shall show in Theorem 5.5 that the equality in (20) holds.

3. A Chain Rule and Its Consequences

The main result of this section offers a chain rule for a *smooth-nonsmooth* composition. In this section Z, and $Y_1, \cdots, Y_k$ also denote finite dimensional normed spaces. The proof of this chain rule is based on the following two lemmas.

Lemma 3.1. *Let $f : \mathcal{D} \to Y$ be a locally Lipschitz function and let $L \in \Lambda(X)$ and $x \in \Omega_L(f)$. Then, for any set $N \subseteq L$ of Lebesgue measure zero, and for any $\varepsilon > 0$, there exist elements $y_0, \dots, y_m \in \Omega_L(f) \cap \big(x + (L \setminus N)\big)$ and nonnegative numbers $\lambda_0, \dots, \lambda_m \geq 0$ with $\lambda_0 + \cdots + \lambda_m = 1$ such that, for all $j \in \{0, \dots, m\}$, we have the inequalities $\|y_j - x\| < \varepsilon$ and*

$$\Big\| D_L f(x) - \sum_{j=0}^{m} \lambda_j D_L f(y_j) \Big\| < \varepsilon, \tag{21}$$

where m is the dimension of the space $\mathcal{L}(L, Y)$.

Proof. Consider the function $g(u) : (\mathcal{D} - x) \cap L \to Y$ defined by $g(u) := f(x + u)$. Then, since $x \in \Omega_L(f)$, the function g is Gâteaux-differentiable at $u = 0$. Thus $Dg(0) \in \partial^c g(0)$, where now $\partial^c g$ stands for the standard Clarke generalized Jacobian of g as in (4). On the other hand, by the results of Warga [38] and Fabian & Preiss [12],

$$\partial^c g(0) = \operatorname{co}\big\{A \in \mathcal{L}(L,Y) \mid \exists\, (v_i)_{i\in\mathbb{N}} \text{ in } \Omega(g) \setminus N : \lim_{i\to\infty} v_i = 0 \text{ and } \lim_{i\to\infty} Dg(v_i) = A\big\}.$$

By Carathéodory's theorem, there exist $m + 1$ elements $A_0, \dots, A_m$ of the set

$$\big\{A \in \mathcal{L}(L,Y) \mid \exists\, (v_i)_{i\in\mathbb{N}} \text{ in } \Omega(g) \setminus N : \lim_{i\to\infty} v_i = 0 \text{ and } \lim_{i\to\infty} Dg(v_i) = A\big\}.$$

and $\lambda_0, \dots, \lambda_m \geq 0$ with $\lambda_0 + \cdots + \lambda_m = 1$ such that

$$Dg(0) = \lambda_0 A_0 + \cdots + \lambda_m A_m. \tag{22}$$

Now, we can find points $u_0, \dots, u_m$ in $\big(\Omega(g) \setminus N\big) \cap \varepsilon B_L$ such that

$$\big\|A_j - Dg(u_j)\big\| < \varepsilon \quad \big(j \in \{0, \dots, m\}\big). \tag{23}$$

From (22)–(23), it follows that

$$\Big\| Dg(0) - \sum_{j=0}^{m} \lambda_j Dg(u_j) \Big\| < \varepsilon,$$

This inequality, with the notation $y_j := x + u_j$ directly yields (21). □

Applying Lemma 3.1, we first prove the chain rule for L-Jacobians. Recall that Y and Z are of finite dimension.

Lemma 3.2. *Let $f : \mathcal{D} \to Y$ be a locally Lipschitz function near p and $g : \mathcal{O} \to Z$ be continuously differentiable at $f(p)$, where $\mathcal{O} \subseteq Y$ is an open set containing $f(p)$. Then, for all subspaces $L \in \Lambda(X)$,*

$$\partial_L(g \circ f)(p) = Dg\big(f(p)\big) \circ \partial_L f(p). \tag{24}$$

Proof. In view of the definition of the pre L-Jacobian and of the convexity and compactness of each of the sets $\partial_L(g \circ f)(p)$ and $Dg\big(f(p)\big) \circ \partial_L f(p)$, to prove (24), it suffices to show that

$$\Delta_L(g \circ f)(p) \subseteq Dg\big(f(p)\big) \circ \partial_L f(p) \quad \text{and} \quad Dg\big(f(p)\big) \circ \Delta_L f(p) \subseteq \partial_L(g \circ f)(p). \tag{25}$$

For the first inclusion in (25), let A be an arbitrary element of the set $\Delta_L(g \circ f)(p)$. Then, there exist elements $(x_i)_{i\in\mathbb{N}}$ in $\Omega_L(g \circ f)$ such that the sequences x_i and $D_L(g \circ f)(x_i)$ tend to p and A, respectively. By taking $x := x_i$, $\varepsilon := 1/i$, and

$$N := \big[\big(\Omega_L(g \circ f) \setminus \Omega_L(f)\big) \cap (x_i + L)\big] - x_i \subseteq L$$

in Lemma 3.1, we obtain that, for all $i \in \mathbb{N}$, there exist elements $y_{i0}, \dots, y_{im} \in \Omega_L(g \circ f) \cap \Omega_L(f) \cap (x_i + L)$ and nonnegative numbers $\lambda_{i0}, \dots, \lambda_{im} \ge 0$ with $\lambda_{i0} + \cdots + \lambda_{im} = 1$ such that, for all $i \in \mathbb{N}$ and for all $j \in \{0, \dots, m\}$, we have the inequalities $\|y_{ij} - x_i\| < 1/i$ and

$$\Big\| D_L(g \circ f)(x_i) - \sum_{j=0}^{m} \lambda_{ij} D_L(g \circ f)(y_{ij}) \Big\| < \frac{1}{i}. \tag{26}$$

Without loss of generality, we may assume that $f(y_{ij}) \in \mathcal{O}$, for all $i \in \mathbb{N}$ and $j \in \{0, \dots, m\}$. Then, since f is L-differentiable at y_{ij} and g is differentiable at $f(y_{ij})$, the standard chain rule implies

$$D_L(g \circ f)(y_{ij}) = Dg\big(f(y_{ij})\big) \circ D_L f(y_{ij}). \tag{27}$$

Taking a subsequence if necessary, we may also assume that the sequences

$$\big(D_L f(y_{i0}), \dots, D_L f(y_{im})\big) \in \mathcal{L}(L, Y)^{m+1} \quad \text{and} \quad (\lambda_{i0}, \dots, \lambda_{im}) \in \mathbb{R}^{m+1}$$

converge and the latter tends to $(\lambda_0, \dots, \lambda_m)$. Upon taking the limit as $i \to \infty$ in (26) and in (27), the continuous differentiability of g at $f(p)$ and the fact that, for all $j \in \{0, \dots, m\}$, the sequences y_{ij} tend to p, yield

$$\begin{aligned} A = \lim_{i\to\infty} D_L(g \circ f)(x_i) &= Dg\big(f(p)\big) \circ \sum_{j=0}^{m} \lambda_j \lim_{i\to\infty} D_L f(y_{ij}) \\ &\in Dg\big(f(p)\big) \circ \sum_{j=0}^{m} \lambda_j \Delta_L f(p) \subseteq Dg\big(f(p)\big) \circ \partial_L f(p), \end{aligned}$$

which proves the first part of (25).

The proof of the second inclusion in (25) is elementary. Indeed, let A be an arbitrary element of $Dg\big(f(p)\big) \circ \Delta_L f(p)$. Then there exist $M \in \Delta_L f(p)$ and a sequence $(x_i)_{i\in\mathbb{N}}$

in $\Omega_L(f)$ such that $A = Dg\big(f(p)\big) \circ M$, and the sequences x_i and $D_L f(x_i)$ converge, respectively, to p and to M as $i \to \infty$. Then, by the standard chain rule, $x_i \in \Omega_L(g \circ f)$ and, for all $i \in \mathbb{N}$,

$$D_L(g \circ f)(x_i) = Dg\big(f(x_i)\big) \circ D_L f(x_i).$$

Taking the limit $i \to \infty$, we obtain that the sequence $D(g \circ f)(x_i)$ converges and

$$\lim_{i\to\infty} D_L(g \circ f)(x_i) = Dg\big(f(p)\big) \circ M = A.$$

This shows that $A \in \Delta_L(g \circ f) \subseteq \partial_L(g \circ f)$ and proves the second inclusion of (25). $\square$

The main result of this section is contained in the next theorem called the *smooth-nonsmooth chain rule.*

Theorem 3.3. *Let $f : \mathcal{D} \to Y$ be a locally Lipschitz function near $p \in \mathcal{D}$ and $g : \mathcal{O} \to Z$ be continuously differentiable at $f(p)$, where $\mathcal{O} \subseteq Y$ is an open set containing $f(p)$. Then,*

$$\partial(g \circ f)(p) = Dg\big(f(p)\big) \circ \partial f(p). \tag{28}$$

Proof. For brevity, denote the linear map $Dg\big(f(p)\big)$ by M and its range $M(Y)$ by W. Since M is surjective onto W, there exists a continuous linear map $M^+ : W \to Y$ which is the right inverse of M, i.e., the identity $M \circ M^+(w) = w$ is satisfied for all $w \in W$.

To show the inclusion "$\subseteq$" in (28), let $\Psi \in \partial(g \circ f)(p)$. Then, by Lemma 3.2, for all $L \in \Lambda(X)$, we have

$$\Psi|_L \in \partial_L(g \circ f)(p) = Dg\big(f(p)\big) \circ \partial_L f(p) = M \circ \partial_L f(p).$$

Taking L to be the one dimensional subspace spanned by an element $h \in X$, it follows from this inclusion that $\Psi(h)$ is an element of the space W for all $h \in X$.

Given that $\Psi \in \partial(g \circ f)(p)$, by Theorem 2.1, we have that $\|\Psi\| \le \ell_{g\circ f}(p)$.

Let $L \in \Lambda(X)$ be fixed and $A \in \partial_L f(p)$ such that $\Psi|_L = M \circ A$. Then $\|A\| \le \ell_f(p)$, and repeating the argument of the proof of Theorem 2.1, it follows that there exists a continuous linear extension $\Phi : X \to Y$ of A such that $\|\Phi\| \le \frac{\gamma}{\beta}\ell_f(p)$. Define

$$\widehat{\Phi} := M^+ \circ \Psi + \Phi - M^+ \circ M \circ \Phi.$$

Then,

$$\begin{aligned}\|\widehat{\Phi}\| &\le \|M^+\|\|\Psi\| + \|\Phi\| + \|M^+ \circ M\|\|\Phi\| \\ &\le \|M^+\|\ell_{g\circ f}(p) + \frac{\gamma}{\beta}\ell_f(p) + \|M^+ \circ M\|\frac{\gamma}{\beta}\ell_f(p) =: \rho,\end{aligned}$$

furthermore, by the right inverse property of M^+, and by using that the image of Ψ is in W,

$$M \circ \widehat{\Phi} = M \circ M^+ \circ \Psi + M \circ \Phi - M \circ M^+ \circ M \circ \Phi = \Psi$$

and

$$\begin{aligned}\widehat{\Phi}|_L &= M^+ \circ \Psi|_L + \Phi|_L - M^+ \circ M \circ \Phi|_L \\ &= M^+ \circ M \circ A + A - M^+ \circ M \circ A = A \in \partial_L f(p).\end{aligned}$$

Therefore $\widehat{\Phi}$ is in $\widehat{\partial}_L f(p) \cap (\rho \overline{B}_{\mathcal{L}(X,Y)})$ and hence $\Psi = M \circ \widehat{\Phi} \in M \circ \big(\widehat{\partial}_L f(p) \cap (\rho \overline{B}_{\mathcal{L}(X,Y)})\big)$. Since L was arbitrary, we obtain that

$$\Psi \in \bigcap_{L \in \Lambda(X)} \Big(M \circ \big(\widehat{\partial}_L f(p) \cap (\rho \overline{B}_{\mathcal{L}(X,Y)})\big)\Big). \tag{29}$$

We show that this inclusion implies $\Psi \in M \circ \Big(\bigcap_{L \in \Lambda(X)} \widehat{\partial}_L f(p) \cap (\rho \overline{B}_{\mathcal{L}(X,Y)})\Big)$ yielding

$$\Psi \in M \circ \Big(\bigcap_{L \in \Lambda(X)} \widehat{\partial}_L f(p) \cap (\rho \overline{B}_{\mathcal{L}(X,Y)})\Big) = M \circ \big(\partial f(p) \cap (\rho \overline{B}_{\mathcal{L}(X,Y)})\big) = M \circ \partial f(p),$$

which is to be proved.

For, define the set $\mathcal{F} \subseteq \mathcal{L}(X,Y)$ by

$$\mathcal{F} := \big\{\Phi \in \mathcal{L}(X,Y) \mid M \circ \Phi = \Psi\big\}.$$

Observe that if $\Psi \notin M \circ \big(\bigcap_{L \in \Lambda(X)} \widehat{\partial}_L f(p) \cap \rho \overline{B}_{\mathcal{L}(X,Y)}\big)$, then

$$\mathcal{F} \cap \big(\rho \overline{B}_{\mathcal{L}(X,Y)}\big) \cap \Big(\bigcap_{L \in \Lambda(X)} \widehat{\partial}_L f(p)\Big) = \emptyset.$$

Note that $\mathcal{F}$ is weak* closed, $\rho \overline{B}_{\mathcal{L}(X,Y)}$ is weak*-compact. Therefore, there is a finite system $\{L_1, \dots, L_k\}$ of subspaces such that

$$\emptyset = \mathcal{F} \cap \big(\rho \overline{B}_{\mathcal{L}(X,Y)}\big) \cap \widehat{\partial}_{L_1} f(p) \cap \cdots \cap \widehat{\partial}_{L_k} f(p) \supseteq \mathcal{F} \cap \big(\rho \overline{B}_{\mathcal{L}(X,Y)}\big) \cap \widehat{\partial}_{L_1 + \cdots + L_k} f(p),$$

that is,

$$\Psi \notin M \circ \widehat{\partial}_{L_1 + \cdots + L_k} f(p) \cap \big(\rho \overline{B}_{\mathcal{L}(X,Y)}\big)$$

which, by (29), produces a contradiction. Hence, the inclusion "$\subseteq$" holds.

To see that the reversed inclusion "$\supseteq$" is valid in (28), let $\Psi \in M \circ \partial f(p)$. Then $\Psi = M \circ \Phi$ for some element $\Phi \in \partial f(p)$. Therefore, using Lemma 3.2, for all $L \in \Lambda(X)$, we get

$$\Psi|_L = (M \circ \Phi)|_L = M \circ (\Phi|_L) \in M \circ \partial_L f(p) = \partial_L (g \circ f)(p).$$

This proves that $\Psi \in \partial(g \circ f)(p)$ and thus validates the inclusion "$\supseteq$" in (28). $\square$

The chain rule regarding the nonsmooth-nonsmooth composition is being developed in [24].

Using Theorem 3.3, the following more general chain rule follows. Here, however, the equality is not valid in general.

Corollary 3.4. *Let $f_1 : \mathcal{D} \to Y_1, \dots, f_k : \mathcal{D} \to Y_k$ be locally Lipschitz functions near $p \in \mathcal{D}$ and let $g : \mathcal{O} \to Z$ be continuously differentiable at $(f_1, \dots, f_k)(p)$, where $\mathcal{O} \subseteq Y_1 \times \cdots \times Y_k$ is an open set containing $(f_1, \dots, f_k)(p)$. Then,*

$$\partial\big(g \circ (f_1, \dots, f_k)\big)(p) \subseteq \sum_{j=1}^{k} D_j g \circ (f_1, \dots, f_k)(p) \circ \partial f_j(p), \tag{30}$$

where $D_j g$ stands for the jth partial Jacobian of g.

Proof. The statement follows if Theorem 3.3 is applied to the functions g and $f = (f_1, \dots, f_k)$ and by using the easy-to-obtain inclusion

$$\partial(f_1, \dots, f_k)(p) \subseteq \partial f_1(p) \times \cdots \times \partial f_k(p). \tag{31}$$

□

Note that if $\partial f_j(p)$ is a singleton for all but one j then (31) and also (30) hold with equality.

Now we list a number of important direct special cases of Corollary 3.4.

Taking the function $g(x, y) := x + y$ where $x, y \in Y$, Corollary 3.4 reduces to the so-called *sum rule.*

Corollary 3.5. *Let $f, g : \mathcal{D} \to Y$ be locally Lipschitz functions. Then, for all $p \in \mathcal{D}$,*

$$\partial(f + g)(p) \subseteq \partial f(p) + \partial g(p).$$

If $g(x, y) := x \cdot y$, where x is an $n \times m-$ and y is an $m \times k-$matrix and "$\cdot$" is the matrix multiplication, then we get a *product rule.*

Corollary 3.6. *Let $f : \mathcal{D} \to \mathbb{R}^{n\times m}$ and $g : \mathcal{D} \to \mathbb{R}^{m\times k}$ be locally Lipschitz functions. Then, for all $p \in \mathcal{D}$,*

$$\partial(f \cdot g)(p) \subseteq \partial f(p) \cdot g(p) + f(p) \cdot \partial g(p).$$

4. Comparison to Known Notions

In this section we establish how our generalized Jacobian relates to Clarke's generalized gradients, Mordukhovich's coderivative, Tibault's limit set, and Ioffe's fan derivative.

The first result states that, in the case of a real-valued function, the generalized Jacobian and Clarke's generalized gradient (defined via (2)) coincide.

Theorem 4.1. *Let $f : \mathcal{D} \to \mathbb{R}$ be a locally Lipschitz function. Then, for all $p \in \mathcal{D}$,*

$$\partial f(p) = \partial^c f(p). \tag{32}$$

An immediate consequence of Theorem 4.1 (whose proof is given later) is the following connection between our generalized Jacobian and Mordukhovich's coderivative $D^* f(p)(y^*)$ defined in (6).

Corollary 4.2. *Assume that X is an Asplund space and let $f : \mathcal{D} \to Y$ be a locally Lipschitz function near $p \in \mathcal{D}$. Then, for all $y^* \in Y^*$,*

$$y^* \circ \partial f(p) = \overline{\mathrm{co}}^{w^*} D^* f(p)(y^*). \tag{33}$$

Proof. Since X is Asplund, it results from [23, Theorems 5.2 and 8.11] that, for all $y^* \in Y^*$, $\overline{\mathrm{co}}^{w^*} D^* f(p)(y^*) = \partial^c(y^* \circ f)(p)$. On the other hand, by Theorem 4.1 and the smooth-nonsmooth chain rule Theorem 3.3,

$$\partial^c(y^* \circ f)(p) = \partial(y^* \circ f)(p) = y^* \circ \partial f(p), \tag{34}$$

which completes the proof. □

The next result involves the notion of limit points of directional difference quotients introduced by Thibault [35] regarding a *vector-valued* function. For $f : X \to Y$, Thibault's limit set is defined by

$$\delta f(p,h) := \Big\{ y \in Y \,\Big|\, \exists\, (x_i)_{i\in\mathbb{N}} \text{ in } \mathcal{D},\ \exists\, (t_i)_{i\in\mathbb{N}} \text{ in } \mathbb{R}_+ \text{ such that } \lim_{i\to\infty}(x_i,t_i) = (p,0) \text{ and } \lim_{i\to\infty} \frac{f(x_i+t_ih)-f(x_i)}{t_i} = y \Big\}. \tag{35}$$

The following theorem states that the fan generated by our generalized Jacobian is the same as the convex hull of Thibault's limit set and it also coincides with Ioffe's fan derivative (defined in (5)).

Theorem 4.3. *Let $f : \mathcal{D} \to Y$ be a locally Lipschitz function. Then, for all $p \in \mathcal{D}$ and for all $h \in X$,*

$$\partial f(p)(h) = \operatorname{co} \delta f(p,h) = D^\circ f(p)(h). \tag{36}$$

An immediate consequence of Theorem 4.3 we get

Corollary 4.4. *A locally Lipschitz function $f : \mathcal{D} \to Y$ is* strictly *Gâteaux-differentiable at $p \in \mathcal{D}$ if and only if its generalized Jacobian $\partial f(p)$ reduces to the singleton $\{Df(p)\}$.*

In order to prove both Theorem 4.1 and Theorem 4.3, we shall first establish the following result which, in fact, is needed for the proof of both theorems.

Lemma 4.5. *Let $f : \mathcal{D} \to Y$ be a locally Lipschitz function. Then, for all $p \in \mathcal{D}$ and for all $h \in X$,*

$$\partial f(p)(h) \subseteq \operatorname{co} \delta f(p,h), \tag{37}$$

Proof. Let $h \in X$ be fixed. Taking L to be the linear span $\operatorname{lin}\{h\} = \langle h\rangle$, we have that $\partial f(p)|_{\langle h\rangle} \subseteq \partial_{\langle h\rangle} f(p)$. Thus, it suffices to show that,

$$\partial_{\langle h\rangle} f(p)(h) \subseteq \operatorname{co} \delta f(p,h). \tag{38}$$

This inclusion will follow if we show that $\Delta_{\langle h\rangle} f(p)(h) \subseteq \delta f(p,h)$. For, let $A \in \Delta_{\langle h\rangle} f(p)$ be arbitrary. Then there exist a sequence (x_i) in $\Omega_{\langle h\rangle}(f)$ that tends to p and for which $D_{\langle h\rangle} f(x_i)$ converges to A as $i \to \infty$. Then $D_{\langle h\rangle} f(x_i) = f'(x_i,h)$ tends to $A(h)$. By the $\langle h\rangle$-differentiability of f at x_i, there exists $0 < t_i < 1/i$ such that

$$\left\| \frac{f(x_i+t_ih)-f(x_i)}{t_i} - f'(x_i,h) \right\| < \frac{1}{i}.$$

Thus

$$A(h) = \lim_{i\to\infty} f'(x_i,h) = \lim_{i\to\infty} \frac{f(x_i+t_ih)-f(x_i)}{t_i} \in \delta f(p,h),$$

which yields the inclusion (38). □

Now we proceed to prove Theorem 4.1.

Proof of Theorem 4.1. First, we show that $\partial f(p) \subseteq \partial^c f(p)$. Since f is real-valued, it follows from (37) that, for all $h \in X$,

$$\begin{aligned}\max\big(\partial f(p)(h)\big) \le \max\big(\delta f(p,h)\big) &= \limsup_{(x,t)\to(p,0^+)} \frac{f(x+th)-f(x)}{t}\\ &= f^\circ(p;h) = \max\big(\partial^c f(p)(h)\big).\end{aligned}$$

Since both $\partial f(p)$ and $\partial^c f(p)$ are weak*-compact convex sets in X^*, the above inequality yields $\partial f(p) \subseteq \partial^c f(p)$.

To show the converse, namely, $\partial^c f(p) \subseteq \partial f(p)$, we will use the approximate subdifferential $\partial_A f(p)$ that was introduced by Ioffe (see e.g., [17]) which is defined as

$$\partial_A f(p) := \bigcap_{L\in\Lambda(X)} \limsup_{x\to p} \partial^- f_{x+L}(x), \tag{39}$$

where

$$f_{x+L}(u) := \begin{cases} f(u) & \text{if } u \in \mathcal{D}\cap(x+L)\\ \infty & \text{if } u\in\mathcal{D}\setminus(x+L),\end{cases}$$

$$\partial^- f_{x+L}(x) := \big\{x^*\in X^* \mid d^- f_{x+L}(x,h) \ge \langle x^*,h\rangle,\ \forall\, h\in X\big\},$$

and $d^-F(x,h)$ is the lower Dini directional derivative of the function $F = f_{x+L}$ at a point $x \in \operatorname{dom} F$ defined as

$$d^-F(x,h) := \liminf_{(t,u)\to(0^+,h)} \frac{F(x+tu)-F(x)}{t}.$$

It is known in [17, Proposition 3.3] that

$$\partial^c f(p) = \overline{\mathrm{co}}^{w^*}\, \partial_A f(p).$$

Hence, the convexity and the w^*-closedness of $\partial f(p)$ yield that the inclusion $\partial^c f(p) \subseteq \partial f(p)$ would hold once we show that $\partial_A f(p) \subseteq \partial f(p)$. For, let $x^* \in \partial_A f(p)$, then, for all $L \in \Lambda(X)$, we have $x^* \in \limsup_{x\to p} \partial^- f_{x+L}(x)$.

Fix $L \in \Lambda(X)$. Then, by (39), we have that there exist two sequences $(x_i)_{i\in\mathbb{N}}$ and $(x_i^*)_{i\in\mathbb{N}}$ such that, for all i, $x_i^* \in \partial^- f_{x_i+L}(x_i)$, the sequence x_i converges to p and x^* is a w^*-cluster point of $(x_i^*)_{i\in\mathbb{N}}$. Since f is Lipschitz near p then, for x near p, we have $\|\partial^- f_{x+L}(x)\| \le \ell_f(p)+1$ and thus, a subsequence of (x_i^*) (which we do not relabel) satisfies

$$x_i^*|_L \to x^*|_L.$$

By the definition of $\partial^- f_{x_i+L}(x_i)$, we have, for all $i\in\mathbb{N}$,

$$\liminf_{(t,u)\to(0^+,h)} \frac{f_{x_i+L}(x_i+tu)-f_{x_i+L}(x_i)}{t} \ge \langle x_i^*,h\rangle, \quad (h\in X).$$

By restricting h in the above inequality to L and by using the definition of f_{x_i+L}, we get

$$\begin{aligned} g_i^\circ(0,h) &\ge \liminf_{t\downarrow 0} \frac{f(x_i+th)-f(x_i)}{t}\\ &= \liminf_{t\downarrow 0} \frac{f_{x_i+L}(x_i+th)-f_{x_i+L}(x_i)}{t} \ge \langle x_i^*,h\rangle, \quad (h\in L),\end{aligned}$$

where g_i is defined by $g_i(u) := f(x_i + u)$ for $u \in (\mathcal{D} - x_i) \cap L$.

Thus, by using the note after (13), it results that

$$x_i^*|_L \in \partial^c g_i(0) \subseteq \partial_L f(x_i), \quad (i \in \mathbb{N}).$$

From Remark 2.3, it follows that

$$x^*|_L \in \partial_L f(p).$$

Since L was arbitrarily fixed in $\Lambda(X)$, it results from the last inclusion and from the definition of $\partial f(p)$ that $x^* \in \partial f(p)$, showing that $\partial_A f(p) \subseteq \partial f(p)$, which completes the proof. □

In the rest of this section we shall prove Theorem 4.3. This proof employs Lemma 4.5 and a simple particular case of the chain rule displayed in Theorem 3.3.

Proof of Theorem 4.3. It suffices to show the following inclusions:

(i) $\partial f(p)(h) \subseteq \operatorname{co} \delta f(p,h)$,
(ii) $\delta f(p,h) \subseteq D^\circ f(p)(h)$,
(iii) $D^\circ f(p)(h) \subseteq \partial f(p)(h)$.

Inclusion (i) is already obtained in Lemma 4.5.

For inclusion (ii), let $v \in \delta f(p,h)$, then $v = \lim_{i\to\infty} \frac{f(x_i+t_ih)-f(x_i)}{t_i}$ for some sequences $x_i \in \mathcal{D}$, $t_i > 0$ with $(x_i, t_i) \to (p, 0^+)$. Hence, for any $y^* \in Y^*$, it follows that

$$\begin{aligned}\langle y^*, v\rangle &= \lim_{i\to\infty} \frac{(y^* \circ f)(x_i + t_i h) - (y^* \circ f)(x_i)}{t_i} \\ &\le \limsup_{(t,x)\to(0^+,p)} \frac{(y^* \circ f)(x + th) - (y^* \circ f)(x)}{t},\end{aligned}$$

that is, $v \in D^\circ f(p)(h)$.

For inclusion (iii), we shall show that we have in fact an equality, namely, for all $y^* \in Y^*$,

$$\big\langle y^*, D^\circ f(p)(h)\big\rangle = \big\langle y^*, \partial f(p)(h)\big\rangle.$$

From the definition of Ioffe's fan derivative in (5) and by using (34), we have

$$\begin{aligned}\big\langle y^*, D^\circ f(p)(h)\big\rangle &= \big\langle \big(D^\circ f(p)\big)^* y^*, h\big\rangle = \big\langle \partial^c(y^* \circ f)(p), h\big\rangle \\ &= \big\langle y^* \circ \partial f(p), h\big\rangle = \big\langle y^*, \partial f(p)(h)\big\rangle.\end{aligned}$$

Therefore both identities in (36) are validated. □

5. Generalized Jacobian of Piecewise Differentiable Functions

"Piecewise smooth" functions have received considerable attention in the last few years because of applications to solution methodology in optimization. See [18], [19], [25], [31], [28], [29], [32]. Intuitively, a piecewise smooth function is a function whose domain can be partitioned into finitely many "pieces" relative on which smoothness holds and continuity holds across the joins of the pieces. More formally, given some continuous functions

$g_1, \dots, g_k : \mathcal{D} \to Y$, a function $f : \mathcal{D} \to Y$ is called a *continuous selection* of $\{g_1, \dots, g_k\}$ if it is continuous on $\mathcal{D}$ and, for all $x \in \mathcal{D}$, it satisfies

$$f(x) \in \{g_1(x), \dots, g_k(x)\}.$$

When the functions $g_1, \dots, g_k$ are differentiable (resp. C^1) on $\mathcal{D}$, then we say that f is *piecewise differentiable* (resp. *piecewise smooth*).

The main result of this section is Theorem 5.5 below which offers an explicit formula for the generalized Jacobian of a continuous selection of a finite family of differentiable functions in terms the convex hull of the derivatives of the functions that are "active" at the point considered. For its proof, we need the auxiliary results contained in the subsequent lemmas.

First we recall the notion of the *contingent cone*. Given a set $Q \subseteq X$ and a point x in the closure of Q, the contingent cone $T(Q|x)$ of Q at x is defined by

$$T(Q|x) := \big\{h \in X \mid \exists\, (h_i)_{i\in\mathbb{N}} \text{ in } X,\ \exists\, (t_i)_{i\in\mathbb{N}} \text{ in } \mathbb{R}_+ : h_i \to h,\ t_i \downarrow 0, \text{ and } x + t_i h_i \in Q \text{ for all } i\big\}.$$

It is obvious that the mapping $Q \mapsto T(Q|x)$ is monotone, i.e., if $Q_1 \subseteq Q_2$ and $x \in \overline{Q}_1$, then $T(Q_1|x) \subseteq T(Q_2|x)$. The following lemma describes an additivity property of this map.

Lemma 5.1. *Let $Q_1, \dots, Q_k \subseteq X$ and $x \in \overline{Q}_1 \cap \dots \cap \overline{Q}_k$. Then*

$$T(Q_1 \cup \dots \cup Q_k|x) = T(Q_1|x) \cup \dots \cup T(Q_k|x). \tag{40}$$

Proof. The inclusion "$\supseteq$" easily follows from the monotonicity property of the contingent cone.

To show the reversed inclusion "$\subseteq$", let $h \in T(Q_1 \cup \dots \cup Q_k|x)$ be arbitrary. Then there exist sequences $h_i \to h$ and $t_i \downarrow 0$ such that $x + t_i h_i \in Q_1 \cup \dots \cup Q_k$ for all i. Obviously, one of the sets $Q_1, \dots, Q_k$, say Q_j, contains infinitely many members of the sequence $x + t_i h_i$. Thus, there exists $i_1 < i_2 < \cdots$ such that $x + t_{i_\ell} h_{i_\ell} \in Q_j$ for all ℓ. Consequently, $h \in T(Q_j|x)$, which proves the inclusion "$\subseteq$" in (40). □

We recall that the union of finitely many nowhere dense sets is again nowhere dense. Indeed, let A and B be arbitrary nowhere dense sets. We may assume that they are closed. To show that $A \cup B$ does not contain an open ball, let x be an arbitrary point of $A \cup B$ and let $r > 0$. Then, since A does not contain interior points, the open set $(x + rB_X) \setminus A$ is nonempty. Thus, there exist $y \in x + rB_X$ and $s > 0$ such that $y + sB_X \subseteq (x + rB_X) \setminus A$. The ball $y + sB_X$ is not covered by B since B is also nowhere dense, therefore, $A \cup B$ cannot contain the ball $x + rB_X$.

Then next lemma shows that if a point is surrounded by finitely many sets then at least one of the sets has a large contingent cone at that point.

Lemma 5.2. *Let $Q_1, \dots, Q_k \subseteq X$, $x \in \overline{Q}_1 \cap \dots \cap \overline{Q}_k$ and assume that x is an interior point of $Q_1 \cup \dots \cup Q_k$. Then there exists an index $j \in \{1, \dots, k\}$ such that the closed linear hull $\overline{\operatorname{lin}}\, T(Q_j|x)$ is the entire space X.*

Proof. If x is an interior point of $Q_1 \cup \cdots \cup Q_k$, then $T(Q_1 \cup \cdots \cup Q_k|x) = X$. Thus, by Lemma 5.1, we have that $T(Q_1|x) \cup \cdots \cup T(Q_k|x) = X$. Hence

$$\overline{\mathrm{lin}}\, T(Q_1|x) \cup \cdots \cup \overline{\mathrm{lin}}\, T(Q_k|x) = X. \tag{41}$$

If all the closed linear subspaces $\overline{\mathrm{lin}}\, T(Q_1|x), \dots, \overline{\mathrm{lin}}\, T(Q_k|x)$ are proper subspaces of X then they do not have interior points. Hence they are nowhere dense. Thus, by the property of nowhere dense sets, the left hand side of equation (41) is also nowhere dense, which yields an obvious contradiction with (41). □

Lemma 5.3. *Let $\mathcal{D} \subseteq X$ be an open subset and assume that $\mathcal{D}_1, \dots, \mathcal{D}_k \subseteq \mathcal{D}$ are closed in the relative topology of $\mathcal{D}$ and that $\mathcal{D} = \mathcal{D}_1 \cup \cdots \cup \mathcal{D}_k$. Then $\mathcal{D}$ is covered by the sets $\overline{\mathcal{D}_1^\circ}, \dots, \overline{\mathcal{D}_k^\circ}$.*

Proof. Assume, on the contrary that $\mathcal{D}$ is not covered by the sets $\overline{\mathcal{D}_1^\circ}, \dots, \overline{\mathcal{D}_k^\circ}$. Then let x be arbitrary point in $H := \mathcal{D} \setminus \big(\overline{\mathcal{D}_1^\circ} \cup \cdots \cup \overline{\mathcal{D}_k^\circ}\big)$. Due to the openness of H, there exists $r > 0$ such that the closed ball $x + r\overline{B}_X$ is contained in H. Hence

$$x + r\overline{B}_X \subset \mathcal{D} \quad \text{and} \quad (x + r\overline{B}_X) \cap \mathcal{D}_j^\circ = \emptyset \quad (j \in \{1, \dots, k\}).$$

The emptiness of the intersection $(x + r\overline{B}_X) \cap \mathcal{D}_j^\circ$ means that $(x + r\overline{B}_X) \cap \mathcal{D}_j$ is nowhere dense. On the other hand, by the assumption of the lemma, the sets $(x + r\overline{B}_X) \cap \mathcal{D}_1, \dots, (x + r\overline{B}_X) \cap \mathcal{D}_k$ form a covering for $x + r\overline{B}_X$, hence $x + r\overline{B}_X$ is also nowhere dense. The contradiction obtained proves the statement. □

Lemma 5.4. *Let $\Phi, \Phi_1, \dots, \Phi_k \in \mathcal{L}(X, Y)$ be arbitrary linear maps such that, for all $L \in \Lambda(X)$,*

$$\Phi|_L \in \mathrm{co}\,\big\{\Phi_1|_L, \dots, \Phi_k|_L\big\}. \tag{42}$$

Then

$$\Phi \in \mathrm{co}\,\big\{\Phi_1, \dots, \Phi_k\big\}. \tag{43}$$

Proof. Let, for $L \in \Lambda(X)$ the set $\Gamma_L \in \mathbb{R}^k$ be defined as

$$\Gamma_L := \big\{(\gamma_1, \dots, \gamma_k) \mid \gamma_1, \dots, \gamma_k \geq 0,\ \gamma_1 + \cdots + \gamma_k = 1,\ \gamma_1\Phi_1|_L + \cdots + \gamma_k\Phi_k|_L = \Phi|_L\big\}.$$

Then, Γ_L is a compact convex subset of $\mathbb{R}^k$. Its nonemptiness follows from the assumption (42). It is also obvious that if $L_1 \subseteq L_2$, then $\Gamma_{L_1} \supseteq \Gamma_{L_2}$. Thus, for an arbitrary finite collection of subspaces $L_1, \dots, L_m \in \Lambda(X)$,

$$\Gamma_{L_1} \cap \cdots \cap \Gamma_{L_m} \supseteq \Gamma_{L_1 + \cdots + L_m} \neq \emptyset,$$

which shows that the family $\big\{\Gamma_L \mid L \in \Lambda(X)\big\}$ satisfies the finite intersection property. Due to the compactness, we obtain that

$$\bigcap_{L \in \Lambda(X)} \Gamma_L \neq \emptyset.$$

By taking $(\gamma_1, \dots, \gamma_k) \in \bigcap_{L \in \Lambda(X)} \Gamma_L$ and $L = \mathrm{lin}\{h\}$, it follows that, for all $h \in X$,

$$\Phi(h) = \gamma_1\Phi_1(h) + \cdots + \gamma_k\Phi_k(h).$$

Thus $\Phi = \gamma_1\Phi_1 + \cdots + \gamma_k\Phi_k \in \mathrm{co}\,\big\{\Phi_1, \dots, \Phi_k\big\}$, which completes the proof. □

The main result of this section is contained in the next theorem. It describes the generalized Jacobian of a piecewise differentiable function in terms of the convex hull of the derivatives of the relevant differentiable functions. As an immediate consequence, we also get that our generalized Jacobian and the Gâteaux-Jacobian defined in (7) coincide in this case. When X is finite dimensional analogous results were obtained by Scholtes [32] and by Kuntz & Scholtes [19].

Theorem 5.5. *Let $f : \mathcal{D} \to Y$ be a continuous function and assume that there exist functions $g_1, \dots, g_k : \mathcal{D} \to Y$ that are continuously differentiable at p such that the sets*

$$\mathcal{D}_j := \big\{x \in \mathcal{D} \mid f(x) = g_j(x)\big\} \quad \big(j \in \{1, \dots, k\}\big) \tag{44}$$

form a covering of $\mathcal{D}$ (i.e., $\mathcal{D} = \mathcal{D}_1 \cup \dots \cup \mathcal{D}_k$). Then f is Lipschitz near p and

$$\partial f(p) = \mathrm{co}\big\{Dg_j(p) : p \in \overline{\mathcal{D}_j^\circ}\big\} = \partial^G f(p), \tag{45}$$

where $\partial^G f(p)$ is the Gâteaux-Jacobian of f at p defined in (7)*.*

Proof. The covering property of the sets $\mathcal{D}_1, \dots, \mathcal{D}_k$ assures that f is a continuous selection of $\{g_1, \dots, g_k\}$. On the other hand, the functions g_i are Lipschitz near p, therefore, by a theorem of Hager [13], f is Lipschitz near p.

In view of Corollary 2.4, we have that $\partial^G f(p) \subseteq \partial f(p)$. Hence, it suffices to show that

$$\text{(i) } \partial f(p) \subseteq \mathrm{co}\big\{Dg_j(p) : p \in \overline{\mathcal{D}_j^\circ}\big\} \quad \text{and} \quad \text{(ii) } \mathrm{co}\big\{Dg_j(p) : p \in \overline{\mathcal{D}_j^\circ}\big\} \subseteq \partial^G f(p). \tag{46}$$

The inclusion (i) in (46) will be proved in three steps. By their continuous differentiablity at p, the functions $g_1, \dots, g_k$ are also locally Lipschitz near p. Without loss of generality, we may assume that $g_1, \dots, g_k$ are differentiable on $\mathcal{D}$ and that $f, g_1, \dots, g_k$ are also Lipschitz on $\mathcal{D}$.

Step 1: For all $L \in \Lambda(X)$ and for all $x \in \Omega_L(f)$, there exists an index j such that $x \in \overline{\mathcal{D}_j^\circ}$ and

$$D_L f(x) = Dg_j(x)|_L. \tag{47}$$

The sets $\mathcal{D}_1, \dots, \mathcal{D}_k$ (that are closed in $\mathcal{D}$) form a covering of $\mathcal{D}$, therefore, by Lemma 5.3, the system $\{\overline{\mathcal{D}_1^\circ}, \dots, \overline{\mathcal{D}_k^\circ}\}$ also covers $\mathcal{D}$. Define, for $x \in \mathcal{D}$, the index set $I(x)$ by

$$I(x) := \big\{j \in \{1, \dots, k\} \mid x \in \overline{\mathcal{D}_j^\circ}\big\}.$$

First, we prove that $\bigcup_{j \in I(x)} \overline{\mathcal{D}_j^\circ} \cap \mathcal{D}$ is a neighborhood of the point x. Assuming the contrary, there exists a sequence (x_i) in $\mathcal{D}$, converging to x such that $x_i \notin \bigcup_{j \in I(x)} \overline{\mathcal{D}_j^\circ} \cap \mathcal{D}$ for all i. Hence, due to the covering property of the system $\{\overline{\mathcal{D}_1^\circ}, \dots, \overline{\mathcal{D}_k^\circ}\}$, infinitely many members of the sequence (x_i) are contained in $\overline{\mathcal{D}_j^\circ} \cap \mathcal{D}$ for some $j \notin I(x)$. Then, by the closedness of $\overline{\mathcal{D}_j^\circ}$, the point x belongs to $\overline{\mathcal{D}_j^\circ}$, which contradicts the definition of $I(x)$.

Set $Q_j := (\overline{\mathcal{D}_j^\circ} \cap \mathcal{D} - x) \cap L$ for $j \in I(x)$. Then, for all $j \in I(x)$, 0 is in $\overline{Q_j}$. On the other hand

$$\bigcup_{j \in I(x)} Q_j = \Big(\Big(\bigcup_{j \in I(x)} \overline{\mathcal{D}_j^\circ} \cap \mathcal{D}\Big) - x\Big) \cap L$$

is a neighborhood of 0 in the subspace L. In view Lemma 5.2, it follows that there exists $j \in I(x)$ such that

$$\overline{\operatorname{lin}}\, T(Q_j|0) = L \tag{48}$$

(where the contingent cone is taken only in the subspace L). To complete the proof of this step, we are going to show that (47) is satisfied by this index j.

Let $h \in T(Q_j|0)$ be a nonzero element. Then there exist sequences $h_i \to h$ and $t_i \downarrow 0$ such that, for all i, $t_i h_i \in (\overline{\mathcal{D}_j^\circ} \cap \mathcal{D} - x) \cap L$. Then, $h_i \in L$ and $x + t_i h_i \in \overline{\mathcal{D}_j^\circ} \cap \mathcal{D} \subseteq \mathcal{D}_j$ for all i. Thus, $f(x + t_i h_i) = g_j(x + t_i h_i)$ holds for all i and also $f(x) = g_j(x)$. Hence, for all i,

$$\frac{g_j(x + t_i h_i) - g_j(x)}{t_i} = \frac{f(x + t_i h_i) - f(x)}{t_i}.$$

Upon taking the limit $i \to \infty$, using the local Lipschitz property of f and g_j, the L-differentiability of f at x, and the differentiability of g_j at x, we get that

$$Dg_j(x)(h) = \lim_{i\to\infty} \frac{g_j(x + t_i h_i) - g_j(x)}{t_i} = \lim_{i\to\infty} \frac{f(x + t_i h_i) - f(x)}{t_i} = D_L f(x)(h).$$

Hence, for all elements h of $T(Q_j|0)$,

$$Dg_j(x)(h) = D_L f(x)(h). \tag{49}$$

By the continuity and linearity, this equation is also valid for elements h of the closed linear hull of $T(Q_j|0)$, which is equal to L by (48). Thus we obtain that (49) holds for all $h \in L$, which completes the proof of Step 1.

Step 2: For all $L \in \Lambda(X)$ and for all $p \in \mathcal{D}$,

$$\Delta_L f(p) \subseteq \big\{ Dg_j(p)|_L : j \in I(p) \big\}. \tag{50}$$

Let $A \in \Delta_L f(p)$. Then there exists a sequence $(x_i)_{i\in\mathbb{N}}$ in $\Omega_L(f)$ converging to p such that $D_L f(x_i)$ tends to A as $i \to \infty$. Then, as we have seen in Step 1, for each i, there exists an index j_i such that $x_i \in \overline{\mathcal{D}_{j_i}^\circ}$ and $D_L f(x_i) = Dg_{j_i}(x_i)|_L$. The sequence $j_1, j_2, \dots$ takes only finitely many values, therefore, by taking a subsequence if necessary, we can assume that it is constant, say $j_1 = j_2 = \cdots = j$. Thus, for all i,

$$D_L f(x_i) = Dg_j(x_i)|_L \quad \text{and} \quad x_i \in \overline{\mathcal{D}_j^\circ}.$$

Taking the limit $i \to \infty$, using the continuous differentiability of g_j and the closedness of $\overline{\mathcal{D}_j^\circ}$, we get

$$A = Dg_j(p)|_L \quad \text{and} \quad p \in \overline{\mathcal{D}_j^\circ}.$$

This proves that A is a member of the set on right hand side of (50).

By the convexity of the right hand side in (50), it immediately follows that, for all $L \in \Lambda(X)$

$$\partial_L f(p) \subseteq \operatorname{co} \big\{ Dg_j(p)|_L : j \in I(p) \big\}. \tag{51}$$

Step 3: For all $p \in \mathcal{D}$,

$$\partial f(p) \subseteq \operatorname{co} \big\{ Dg_j(p) : j \in I(p) \big\}. \tag{52}$$

Let $\Phi \in \mathcal{L}(X,Y)$ be an arbitrary element of $\partial f(p)$. Then, in view of (51), for all $L \in \Lambda(X)$,

$$\Phi|_L \in \operatorname{co}\big\{Dg_j(p)|_L : j \in I(p)\big\}.$$

Using Lemma 5.4, it follows that

$$\Phi \in \operatorname{co}\big\{Dg_j(p) : j \in I(p)\big\},$$

which proves (52) and completes the proof of the theorem.

For inclusion (ii) in (46), let, for some j, the point p be in $\overline{\mathcal{D}_j^\circ}$. Then there exists a sequence $(x_i)_{i\in\mathbb{N}}$ in $\mathcal{D}_j^\circ$ such that $x_i \to p$ as $i \to \infty$. Since, for all i, $x_i \in \mathcal{D}_j^\circ$, therefore f and g_j coincide in a neighborhood of x_i which yields that f is Gâteaux-differentiable at x_i and $Df(x_i) = Dg_j(x_i)$ for all i. By the continuity of Dg_j at p, we have that $Df(x_i) \to Dg_j(p)$ as $i \to \infty$, whence $Dg_j(p) \in \partial^G f(p)$ follows. □

The following corollary is parallel to a well-known result for the case where $Y = \mathbb{R}$, [11, Proposition 2.3.12], it immediately follows from Theorem 5.5.

Corollary 5.6. *Let $f := \max(g_1,\dots,g_k)$, where $g_1,\dots,g_k : \mathcal{D} \to \mathbb{R}$ are continuously differentiable functions. Then, for all $p \in \mathcal{D}$,*

$$\partial f(p) = \operatorname{co}\big\{Dg_j(p) : p \in \overline{\mathcal{D}_j^\circ}\big\},$$

where $\mathcal{D}_j := \{x \in \mathcal{D} \mid g_j(x) = f(x)\}$ for $j \in \{1,\dots,k\}$.

Finally, in the example below, we compute the generalized Jacobian of a vector-valued function.

Example 5.7. Let X be an at least two dimensional normed space and let φ and ψ be two linearly independent linear functionals over X. Define $f : X \to \mathbb{R}^3$ by

$$f(x) := \big(|\varphi(x)|, |\psi(x)|, |\varphi(x)+\psi(x)|\big).$$

Then, it is easy to recognize that f can be obtained in terms of the following 6 smooth functions:

$$\begin{aligned} g_1 &= (\varphi,\psi,\varphi+\psi), & g_2 &= (-\varphi,\psi,\varphi+\psi), & g_3 &= (\varphi,-\psi,\varphi+\psi),\\ g_4 &= (-\varphi,-\psi,-\varphi-\psi), & g_5 &= (-\varphi,\psi,-\varphi-\psi), & g_6 &= (\varphi,-\psi,-\varphi-\psi). \end{aligned}$$

It is also easy to see that $0 \in \overline{\mathcal{D}_j^\circ}$ for all j, where $\mathcal{D}_j = \{x \in X : f(x) = g_j(x)\}$. Therefore,

$$\partial f(0) = \operatorname{co}\{g_1,\dots,g_6\}.$$

Acknowledgements. The authors wish to thank Prof. F. H. Clarke for a citation correction, Prof. A. D. Ioffe for his suggestion to use his theory of the approximate subdifferentials in the proof of Theorem 4.1, and an anonymous referee for providing us with relevant references.

References

[1] N. Aronszajn: Differentiability of Lipschitzian mappings between Banach spaces, Studia Math. 57(2) (1976) 147–190.

[2] J.-P. Aubin: Contingent derivatives of set-valued maps and existence of solutions to nonlinear inclusions and differential inclusions, in: Mathematical Analysis and Applications, Part A, L. Nachbin (ed.), Adv. Math., Suppl. Stud. 7A, Academic Press, New York (1981) 159–229.

[3] J.-P. Aubin, H. Frankowska: Set-Valued Analysis, Systems and Control: Foundations and Applications 2, Birkhäuser, Boston (1990).

[4] Y. Benyamini, J. Lindenstrauss: Geometric Nonlinear Functional Analysis, Vol. 1, American Mathematical Society Colloquium Publications 48, American Mathematical Society, Providence (2000).

[5] J. M. Borwein, Q. J. Zhu: A survey of subdifferential calculus with applications, Nonlinear Anal., Theory Methods Appl. 38A(6) (1999) 687–773.

[6] J. P. R. Christensen: Measure theoretic zero sets in infinite dimensional spaces and applications to differentiability of Lipschitz mappings, Coll. Anal. Fonct. (Univ. Bordeaux, 1973), Publ. Dép. Math., Lyon 10(2) (1973) 29–39.

[7] F. H. Clarke: Necessary Conditions for Nonsmooth Problems in Optimal Control and the Calculus of Variations, Ph.D. Thesis, University of Washington, Washington (1973).

[8] F. H. Clarke: Generalized gradients and applications, Trans. Amer. Math. Soc. 205 (1975) 247–262.

[9] F. H. Clarke: On the inverse function theorem, Pacific J. Math. 64(1) (1976) 97–102.

[10] F. H. Clarke: Generalized gradients of Lipschitz functionals, Adv. Math. 40(1) (1981) 52–67.

[11] F. H. Clarke: Optimization and Nonsmooth Analysis, John Wiley & Sons, New York (1983).

[12] M. Fabian, D. Preiss: On the Clarke's generalized Jacobian, Abstract Analysis, Proc. 14th Winter Sch. (Srní, 1986), Suppl. Rend. Circ. Mat. Palermo, II. Ser. 14 (1987) 305–307.

[13] W. W. Hager: Lipschitz continuity for constrained processes, SIAM J. Control Optim. 17(3) (1979) 321–338.

[14] H. Halkin: Interior mapping theorem with set-valued derivatives, J. Anal. Math. 30 (1976) 200–207.

[15] H. Halkin: Mathematical programming without differentiability, in: Calculus of Variations and Control Theory, D. L. Russell (ed.), Academic Press, New York (1976) 279–287.

[16] A. D. Ioffe: Nonsmooth analysis: Differential calculus of nondifferentiable mappings, Trans. Amer. Math. Soc. 266(1) (1981) 1–56.

[17] A. D. Ioffe: Approximate subdifferentials and applications. II, Mathematika 33(1) (1986) 111–128.

[18] H. Th. Jongen, D. Pallaschke: On linearization and continuous selections of functions, Optimization 19(3) (1988) 343–353.

[19] L. Kuntz, S. Scholtes: Qualitative aspects of the local approximation of a piecewise differentiable function, Nonlinear Anal., Theory Methods Appl. 25(2) (1995) 197–215.

[20] P. Mankiewicz: On the differentiability of Lipschitz mappings in Fréchet spaces, Studia Math. 45 (1973) 15–29.

[21] B. S. Mordukhovich: Metric approximations and necessary conditions for optimality for general classes of nonsmooth extremal problems, Dokl. Akad. Nauk SSSR 254(5) (1980) 1072–1076.

[22] B. S. Mordukhovich: Generalized differential calculus for nonsmooth and set-valued mappings, J. Math. Anal. Appl. 183(1) (1994) 250–288.

[23] B. S. Mordukhovich, Y. H. Shao: Nonsmooth sequential analysis in Asplund spaces, Trans. Amer. Math. Soc. 348(4) (1996) 1235–1280.

[24] Zs. Páles, V. Zeidan: Infinite dimensional generalized Jacobian: Properties and calculus rules, preprint.

[25] J.-S. Pang, D. Ralph: Piecewise smoothness, local invertibility, and parametric analysis of normal maps, Math. Oper. Res. 21(2) (1996) 401–426.

[26] R. R. Phelps: Gaussian null sets and differentiability of Lipschitz map on Banach spaces, Pacific J. Math. 77(2) (1978) 523–531.

[27] B. H. Pourciau: Analysis and optimization of Lipschitz continuous mappings, J. Optimization Theory Appl. 22(3) (1977) 311–351.

[28] D. Ralph, S. Scholtes: Sensitivity analysis of composite piecewise smooth equations, Math. Program. 76(3B) (1997) 593–612.

[29] D. Ralph, H. Xu: Implicit smoothing and its application to optimization with piecewise smooth equality constraints, J. Optimization Theory Appl. 124(3) (2005) 673–699.

[30] R. T. Rockafellar: Convex Analysis, Princeton University Press, Princeton (1970).

[31] R. T. Rockafellar: A property of piecewise smooth functions, Comput. Optim. Appl. 25(1-3) (2003) 247–250.

[32] S. Scholtes: Introduction to Piecewise Differentiable Equations, Habilitation Thesis, University of Karlsruhe, Karlsruhe (1994).

[33] T. H. Sweetser: A minimal set-valued strong derivative for vector-valued Lipschitz functions, J. Optimization Theory Appl. 23(4) (1977) 549–562.

[34] T. H. Sweetser: A Set-Valued Strong Derivative in Infinite Dimensional Spaces, with Applications in Hilbert Spaces, Ph.D. Thesis, University of California, San Diego (1979).

[35] L. Thibault: Subdifferentials of compactly Lipschitzian vector-valued functions, Ann. Mat. Pura Appl., IV. Ser. 125 (1980) 157–192.

[36] L. Thibault: On generalized differentials and subdifferentials of Lipschitz vector-valued functions, Nonlinear Anal., Theory Methods Appl. 6(10) (1982) 1037–1053.

[37] J. Warga: Derivative containers, inverse functions, and controllability, in: Calculus of Variations and Control Theory, D. L. Russell (ed.), Academic Press, New York (1976) 13–45; Errata: p. 46.

[38] J. Warga: Fat homeomorphisms and unbounded derivate containers, J. Math. Anal. Appl. 81(2) (1981) 545–560; Errata: pp. 582–583.