ESAIM: PROCEEDINGS, April 2007, Vol.17, 80-95
Alain Piétrus & Michel H. Geoffroy, Editors

EXACT CALCULUS FOR PROXIMAL SUBGRADIENTS WITH APPLICATIONS TO OPTIMIZATION *

BORIS S. MORDUKHOVICH[1] AND NGUYEN MAU NAM[1]

Abstract. The paper contains new exact calculus rules for proximal subgradients of extended-real-valued functions defined on arbitrary real Banach spaces. We also develop efficient formulas for evaluating proximal subgradients of marginal/value functions in various problems of parametric optimization. The results obtained are employed to derive new necessary optimality conditions in unconstrained nondifferentiable programming.

Introduction

This paper concerns some aspects of variational analysis related to generalized differentiation and its applications to marginal/value functions and optimality conditions in problems of nonsmooth optimization in general Banach spaces. The main generalized differential objects of our study are the so-called *proximal subdifferential* (the collection of proximal subgradients) of extended-real-valued functions and associated notions of normals and coderivatives for sets and set-valued mappings.

Let X be an arbitrary real Banach space, and let $\varphi\colon X \to \overline{R} := [-\infty, \infty]$ be an extended-real-valued function finite at $\bar{x}$. Following Rockafellar [14], we say that $x^* \in X^*$ is a *proximal subgradient* of φ at $\bar{x}$ if there exist a neighborhood U of $\bar{x}$ and a number $\sigma > 0$ such that

$$\varphi(x) \geq \varphi(\bar{x}) + \langle x^*, x - \bar{x}\rangle - \sigma\|x - \bar{x}\|^2 \quad \text{whenever} \quad x \in U. \tag{0.1}$$

The set of all such x^* is called the proximal subdifferential of φ at $\bar{x}$ and is denoted by $\partial_P\varphi(\bar{x})$. In finite dimensions, the proximal subdifferential is a functional counterpart of proximal normals to closed sets originally considered by Clarke; see the book [4] for a Hilbert space version of *proximal analysis* with references and applications. Note that the proximal subdifferential may be *empty* at some points of *smooth* functions and that the corresponding proximal normals may not exist at boundary points of sets defined by smooth inequalities even in finite dimensions (see, e,g., [15, p. 213]). However, these objects are known to be *nontrivial*, for all lower semicontinuous (l.s.c.) functions and closed sets, at *dense* subsets of their domains and boundaries in any *Hilbert* spaces; cf. the density results in [4, Section 1.3]. Furthermore, in Hilbert spaces proximal subgradients satisfy the so-called "fuzzy calculus" originated by Ioffe [6]; see [4, Chapter 1] for a systematic theory. The main difference of fuzzy calculus rules from their "exact" counterpart is that the former provide evaluations of proximal subgradients for various compositions via those for separate components not exactly at the points in question but at some points nearby in a certain approximate/fuzzy way. Of course, the exact calculus is more

* *This research was partially supported by the National Science Foundation under grant DMS-0304989 and by the Australian Research Council under grant DP-0451168.*

[1] Department of Mathematics, Wayne State University, Detroit, MI 48202, U.S.A.

appreciated if available, which is unfortunately not the case for proximal subgradients even in rather simple finite-dimensional settings.

The primary goal of this paper is to find new classes of functions and mappings for which proximal subgradients and associated proximal normals and coderivatives satisfy valuable *exact calculus rules*. Surprisingly, we succeed to achieve this goal in the general *Banach space* framework employing, together with proximal subgradients, their *upper counterparts* (known also as "proximal supergradients") for some components involved in compositions. Certain results in this direction require *modified* definitions of proximal normals (and hence proximal coderivatives while not proximal subgradients (0.1)) in Banach spaces that go back to the standard ones in the Hilbert space setting. The results obtained here for proximal constructions are largely similar to those derived for Fréchet subgradients in our related paper [10] by using somewhat different techniques.

Besides developing calculus rules for proximal constructions, we derive new calculus results for their *sequential limiting*counterparts, which go back to those introduced by Mordukhovich [7] in finite dimensions and agree with the basic constructions of the books [8, 9] in the case of Hilbert spaces; see also the references therein and the books [4, 15] for finite dimensional and Hilbert spaces theories, respectively. Note that the limiting constructions of this paper are generally different from the ones used in [8, 9] out of Hilbert spaces.

Together with the mentioned calculus rules, we derive new results on evaluating proximal subgradients of general *marginal functions* and their specifications (known as *value functions*) for parametric problems of nonlinear programming. Furthermore, we develop applications of the results obtained to *necessary optimality conditions* in problems of the so-called *nondifferentiable difference programming* including new conditions for *sharp minimizers* in the sense of Polyak [13].

The rest of the paper is organized as follows. In Section 2 we present basic definitions and preliminaries, which are widely used in the sequel. Section 3 contains new calculus rules for proximal subgradients of extended-real-valued functions and proximal coderivatives of Lipschitzian mappings. In Section 4 we establish upper estimates for proximal subgradients of marginal/value functions and their specifications. The final Section 5 is devoted to applications of the calculus results to necessary optimality conditions in nondifferentiable difference programming.

1. Basic Definitions and Preliminaries

Throughout this paper all spaces are supposed to be *real Banach* unless otherwise stated. Let X be a Banach space equipped with the norm $\|\cdot\|$ and the canonical pairing $\langle\cdot,\cdot\rangle$ between X and its topologically dual space X^*; obviously $\langle\cdot,\cdot\rangle$ reduces to the standard inner product in the case of Hilbert spaces with $X^* = X$.

The definition of proximal subgradients and the proximal subdifferential $\partial_P\varphi(\bar{x})$, given in (0.1) for an arbitrary functions $\varphi\colon X \to \overline{R}$ finite at $\bar{x}$, holds in any Banach space. Considering a nonempty set $\Omega \subset X$ and its *indicator function* $\delta(\cdot;\Omega)$ equal 0 if $x \in \Omega$ and ∞ otherwise, we define the *proximal normal cone* to Ω at $\bar{x} \in \Omega$ by

$$N_P(\bar{x};\Omega) := \partial_P\delta(\bar{x};\Omega). \tag{1.1}$$

If X is a Hilbert space (particularly a finite-dimensional space with the Euclidean norm) and if $\Omega \subset X$ is closed, proximal normals (1.1) can be equivalently described via the *Euclidean projection* to Ω; see, e.g., [4]. It does not generally hold out of Hilbert spaces, while the following description of proximal normals directly follows from (0.1) and (1.1) in the arbitrary Banach space setting: $x^* \in N_P(\bar{x};\Omega)$ *if and only if* there exists $\gamma > 0$ and $\sigma = \sigma(x^*,\bar{x}) > 0$ such that

$$\langle x^*, x - \bar{x}\rangle \le \sigma\|x - \bar{x}\|^2 \text{ for any } x \in \Omega \cap B_\gamma(\bar{x}), \tag{1.2}$$

where $B_\gamma(\bar{x})$ stands for the ball centered at $\bar{x}$ with radius γ. On the other hand, one can observe based on (1.2) that the representation

$$\partial_P\varphi(\bar{x}) = \big\{x^* \in X^* \big|\ (x^*,-1) \in N_P((\bar{x},\varphi(\bar{x})); \operatorname{epi}\varphi)\big\} \tag{1.3}$$

holds for every function $\varphi\colon X \to \overline{I\!R}$ finite at $\bar{x}$ with its *epigraph* defined by

$$\operatorname{epi}\varphi := \big\{(x,\alpha) \in X \times I\!R \big|\ \alpha \geq \varphi(x)\big\};$$

cf. the proof of [8, Theorem 1.86]. Directly from definition (0.1) we get that $\partial_P\varphi(\bar{x}) \neq \emptyset$ for local minimizers $\bar{x}$ of φ, and moreover

$$0 \in \partial_P\varphi(\bar{x}). \tag{1.4}$$

This is a *proximal Fermat rule*, which may not reduce to the classical one. Indeed, it has been mentioned that the proximal subdifferential $\partial_P\varphi(\bar{x})$ does not always agree with the classical Fréchet derivative $\nabla\varphi(\bar{x})$ of φ at $\bar{x}$ for smooth functions, since $\partial_P\varphi(\bar{x})$ may be empty even for smooth (C^1) functions in finite dimensions. It does not happen for *twice continuously differentiable* functions:

$$\partial_P\varphi(\bar{x}) = \big\{\nabla\varphi(\bar{x})\big\} \ \text{ whenever } \ \varphi \in C^2 \ \text{ around } \ \bar{x}. \tag{1.5}$$

The latter is proved in [4, Corollary 1.2.6] for Hilbert spaces, while the proof therein is based on (0.1) and holds in an arbitrary Banach space setting.

Given a set-valued mapping $F\colon X \rightrightarrows Y$ between Banach spaces, we consider its graph

$$\operatorname{gph} F := \big\{(x,y) \in X \times Y \big|\ y \in F(x)\big\}$$

and define the *proximal coderivative* of F at $(\bar{x},\bar{y}) \in \operatorname{gph} F$ by

$$D_P^*F(\bar{x},\bar{y})(y^*) := \big\{x^* \in Y \ \big|\ (x^*,-y^*) \in N_P((\bar{x};\bar{y});\operatorname{gph} F)\big\}, \tag{1.6}$$

where $\bar{y}$ is omitted if $F = f\colon X \to Y$ is single-valued. Note that by (1.3) the proximal subdifferential $\partial_P\varphi(\bar{x})$ of $\varphi\colon X \to \overline{I\!R}$ can be considered as a special case of the proximal coderivative

$$\partial_P\varphi(\bar{x}) = D_P^*E_\varphi(\bar{x},\varphi(\bar{x}))(1)$$

applied to the epigraphical multifunction $E_\varphi(x) := \{\alpha \in I\!R |\ \alpha \geq \varphi(x)\}$.

We also consider the *limiting subdifferential* of $\varphi\colon X \to \overline{I\!R}$ at $\bar{x}$ with $|\varphi(\bar{x})| < \infty$ defined by the *sequential* limits

$$\partial\varphi(\bar{x}) := \big\{x^* \in X^* |\ x_k^* \overset{w^*}{\to} x^*, x_k^* \in \partial_P\varphi(x_k),\ x_k \overset{\varphi}{\to} \bar{x}\big\}, \tag{1.7}$$

where w^* signifies the weak* topology of X^*, and where $x \overset{\varphi}{\to} \bar{x}$ means that $x \to \bar{x}$ with $\varphi(x) \to \varphi(\bar{x})$. Similarly to (1.6), define the *limiting coderivative* of $F\colon X \rightrightarrows Y$ at $(\bar{x},\bar{y}) \in \operatorname{gph} F$ by

$$D^*F(\bar{x},\bar{y})(y^*) := \big\{x^* \in Y \ \big|\ (x^*,-y^*) \in N((\bar{x};\bar{y});\operatorname{gph} F)\big\}, \tag{1.8}$$

where the corresponding *limiting normal cone* is $N(\cdot;\Omega) := \partial\delta(\cdot;\Omega)$.

We get directly from (1.7) that $\partial_P\varphi(\bar{x}) \subset \partial\varphi(\bar{x})$. A function $\varphi\colon X \to \overline{I\!R}$ finite at $\bar{x}$ is called *proximally regular* at $\bar{x}$ if

$$\partial_P\varphi(\bar{x}) = \partial\varphi(\bar{x}). \tag{1.9}$$

Besides C^2 and convex functions, the proximal regularity relation (1.9) holds for a number of important classes of functions frequently encountered in variational analysis and optimization; see [2] for more details, discussions, and references. In particular, a set $\Omega \subset X$ is proximally regular at $\bar{x} \in \Omega$ if (1.9) holds for the indicator function $\varphi(x) = \delta(x;\Omega)$.

2. Calculus of Proximal Subgradients and Coderivatives

In this section we develop *exact* calculus rules for proximal subgradients of extended-real-valued functions in Banach spaces and for the associate coderivative construction. Let us start with a simple while important *difference rule* for proximal subgradients, which holds in a rather general setting, in contrast to its sum rule counterpart.

Theorem 2.1. (difference rule for proximal subgradients). *Let $\varphi_i\colon X \to \overline{R}$ be finite at $\bar{x}$ for $i = 1, 2$. Assume that $\partial_P\varphi_2(\bar{x}) \neq \emptyset$. Then*

$$\partial_P(\varphi_1 - \varphi_2)(\bar{x}) \subset \bigcap_{x^*\in\partial_P\varphi_2(\bar{x})} \big[\partial_P\varphi_1(\bar{x}) - x^*\big] \subset \partial_P\varphi_1(\bar{x}) - \partial_P\varphi_2(\bar{x}). \tag{2.1}$$

Proof. Fix arbitrary subgradients $x^* \in \partial_P(\varphi_1 - \varphi_2)(\bar{x})$ and $x_2^* \in \partial_P\varphi_2(\bar{x})$. Then we find positive numbers σ_1, σ_2, and γ such that

$$\langle x^*, x - \bar{x}\rangle \le \varphi_1(x) - \varphi_2(x) - (\varphi_1(\bar{x}) - \varphi_2(\bar{x})) + \sigma_1\|x - \bar{x}\|^2 \text{ and}$$

$$\langle x_2^*, x - \bar{x}\rangle \le \varphi_2(x) - \varphi_2(\bar{x}) + \sigma_2\|x - \bar{x}\|^2 \text{ whenever } \|x - \bar{x}\| < \gamma.$$

The above inequalities directly imply that

$$\langle x^* + x_2^*, x - \bar{x}\rangle \le \varphi_1(x) - \varphi_1(\bar{x}) + (\sigma_1 + \sigma_2)\|x - \bar{x}\|^2 \text{ for all } \|x - \bar{x}\| < \gamma,$$

which justifies $x^* + x_2^* \in \partial_P\varphi_1(\bar{x})$ and hence (2.1). □

Remark 2.2. Let us establish a *stronger* version of Theorem 2.1. Let X be a Banach space and let $\varphi : X \to [-\infty, \infty]$ be finite at $\bar{x}$. Then $x^* \in X^*$ is an *$s-$Holder subgradient* (see [1]) of φ at $\bar{x}$ ($s \in (0, 1]$) if there exists positive constants $\delta_{\bar{x}}$ and $C_{\bar{x}}$ such that

$$\varphi(x) - \varphi(\bar{x}) - \langle x^*, x - \bar{x}\rangle \ge -C_{\bar{x}}||x - \bar{x}||^{1+s}$$

whenever $||x - \bar{x}|| < \delta_{\bar{x}}$. The set of all such x^* is called the *$s-$Holder subdifferential* of φ at $\bar{x}$ and is denoted by $\partial_{H(s)}\varphi(\bar{x})$.

We can easily prove the following: if $\varphi_1(\bar{x})$ and $\varphi_2(\bar{x})$ are finite, then

$$\partial_{H(s)}(\varphi_1 - \varphi_2)(\bar{x}) \subset \bigcap_{x^*\in\partial_{H(s)}\varphi_2(\bar{x})} [\partial_{H(s)}\varphi_1(\bar{x}) - x^*] \subset \partial_{H(s)}\varphi_1(\bar{x}) - \partial_{H(s)}\varphi_2(\bar{x}),$$

provided that $\partial_{H(s)}\varphi_2(\bar{x})) \neq \emptyset$.

All of the results of the paper hold similarly for Holder subgradients. However, we just formulate for proximal subgradients for simplicity.

As a consequence of Theorem 2.1, we derive a new difference rule for limiting subgradients (1.7) that holds in any Banach space whose unit dual ball is weak* *sequentially* compact. The latter class is known to be sufficiently large including all spaces with a Gâteaux differentiable renorm, all Asplund spaces, etc.; see [11] for more details.

Corollary 2.3. (difference rule for limiting subgradients). *Let X be a Banach space whose unit dual ball is weak* sequentially compact in X^*, let $\varphi_1\colon X \to \overline{R}$ be finite around $\bar{x}$, and let $\varphi_2\colon X \to \overline{R}$ be Lipschitz continuous and such that $\partial_P\varphi_2(x) \neq \emptyset$ around $\bar{x}$; all the assumptions on φ_2 are automatic when φ_2 is convex and continuous around this point. Then*

$$\partial(\varphi_1 - \varphi_2)(\bar{x}) \subset \partial\varphi_1(\bar{x}) - \partial\varphi_2(\bar{x}).$$

Proof. Pick any $x^* \in \partial(\varphi_1 - \varphi_2)(\bar{x})$ and, by the definition of limiting subgradients in (1.7), find sequences $x_k \to \bar{x}$ and $x_k^* \xrightarrow{w^*} x^*$ satisfying $\varphi_1(x_k) - \varphi_2(x_k) \to \varphi_1(\bar{x}) - \varphi_2(\bar{x})$ and

$$x_k^* \in \partial_P(\varphi_1 - \varphi_2)(x_k) \text{ for all } k \in I\!N := \{1, 2, \ldots\}.$$

Applying Theorem 2.1, we select sequences $x_{ik}^* \in \partial_P \varphi_i(x_k)$ as $i = 1, 2$ such that

$$x_k^* = x_{1k}^* - x_{2k}^* \text{ for all } k \in I\!N. \tag{2.2}$$

Since φ_2 is Lipschitz continuous around $\bar{x}$, it is easy to observe the sets $\partial_P \varphi_2(x)$ are uniformly bounded in X^*. By the sequential weak* compactness imposed on the dual unit ball, we suppose without loss of generality that the sequence $\{x_{2k}^*\}$ converges weak* to some $x_2^* \in X^*$. The continuity of φ_2 gives $\varphi_2(x_k) \to \varphi_2(\bar{x})$, and hence $\varphi_1(x_k) \to \varphi_1(\bar{x})$ as $k \to \infty$ by the above choice of $\{x_k\}$. By definition (1.7) of the limiting subdifferential, we get that $x_2^* \in \partial\varphi_2(\bar{x})$. Furthermore, it follows from (2.2) that the sequence $\{x_{1k}^*\}$ converges weak* to some $x_1^* := x^* + x_2^*$, which must belong to $\partial\varphi_1(\bar{x})$ by the discussions above. We conclude that $x^* \in \partial\varphi_1(\bar{x}) - \partial\varphi_2(\bar{x})$ and complete the proof. □

The next result establishes the *scalarization formula* that relates the proximal coderivative of an arbitrary locally Lipschitzian mapping $f\colon X \to Y$ between *Banach spaces* with the proximal subdifferential of its scalarization

$$\langle y^*, f\rangle(x) := \langle y^*, f(x)\rangle, \quad x \in X, \quad y^* \in Y^*.$$

Theorem 2.4. (scalarization formula). *Let $f\colon X \to Y$ be Lipschitz continuous around $\bar{x}$. Then we have the equality*

$$D_P^* f(\bar{x})(y^*) = \partial_P\langle y^*, f\rangle(\bar{x}) \text{ for all } y^* \in Y. \tag{2.3}$$

Proof. Fix any $x^* \in \partial_P\langle y^*, f\rangle(\bar{x})$. By definition (0.1) of proximal subgradients, find positive numbers σ and γ such that

$$\begin{aligned}\langle x^*, x - \bar{x}\rangle &\le \langle y^*, f\rangle(x) - \langle y^*, f\rangle(\bar{x}) + \sigma\|x - \bar{x}\|^2\\ &= \langle y^*, f(x)\rangle - \langle y^*, f(\bar{x})\rangle + \sigma\|x - \bar{x}\|^2\end{aligned}$$

whenever $\|x - \bar{x}\| < \gamma$. For any such x we have

$$\langle x^*, x - \bar{x}\rangle - \langle y^*, f(x) - f(\bar{x})\rangle \le \sigma\|x - \bar{x}\|^2 \le \sigma(\|x - \bar{x}\| + \|f(x) - f(\bar{x})\|)^2,$$

which implies by (1.2) that $(x^*, -y^*) \in N_P((\bar{x}, f(\bar{x})); \operatorname{gph} f)$ and hence $x^* \in D_P^* f(\bar{x})(y^*)$.

Let us prove that the opposite inclusion holds if f is Lipschitz continuous around $\bar{x}$ with modulus $\ell \ge 0$. Taking arbitrary $x^* \in D_P^* f(\bar{x})(y^*)$, find $\sigma, \gamma > 0$ such that

$$\langle x^*, x - \bar{x}\rangle - \langle y^*, f(x) - f(\bar{x})\rangle \le \sigma(\|x - \bar{x}\| + \|f(x) - f(\bar{x})\|)^2 \le (\ell + 1)^2\sigma\|x - \bar{x}\|^2$$

whenever $\|x - \bar{x}\| < \gamma$. This gives

$$\langle x^*, x - \bar{x}\rangle \le \langle y^*, f\rangle(x) - \langle y^*, f\rangle(\bar{x}) + (\ell + 1)^2\sigma\|x - \bar{x}\|^2 \text{ as } \|x - \bar{x}\| < \gamma,$$

which yields $x^* \in \partial_P\langle y^*, f\rangle(\bar{x})$ and thus justifies (2.3). □

Next, combining the results from Theorem 2.4 and Theorem 2.1, we obtain the following upper estimate for the proximal coderivative of the difference of single-valued Lipschitzian mappings.

Proposition 2.5. (difference rule for proximal coderivatives). *Let the mappings $f_i\colon X \to Y$, $i = 1, 2$, be Lipschitz continuous around $\bar{x}$, and let $y^* \in Y^*$ be such that $D^*_P f_2(\bar{x})(y^*) \neq \emptyset$. Then*

$$D^*_P(f_1 - f_2)(\bar{x})(y^*) \subset \bigcap_{x^* \in D^*_P f_2(\bar{x})(y^*)} \big[D^*_P f_1(\bar{x})(y^*) - x^*\big]. \tag{2.4}$$

Proof. For any $y^* \in Y$ and $x \in X$ we obviously have

$$\langle y^*, f_1 - f_2\rangle(x) = \langle y^*, f_1\rangle(x) - \langle y^*, f_2\rangle(x).$$

Theorem 2.4 and the assumption of this proposition ensure that

$$\partial_P \langle y^*, f_2\rangle(\bar{x}) = D^*_P f_2(\bar{x})(y^*) \neq \emptyset.$$

Employing now the subdifferential difference rule from Theorem 2.1, we get

$$\begin{aligned} D^*_P(f_1 - f_2)(\bar{x})(y^*) &= \partial_P\langle y^*, f_1 - f_2\rangle(\bar{x}) \\ &\subset \bigcap_{x^* \in \partial_P\langle y^*, f_2\rangle(\bar{x})} \big[\partial_P\langle y^*, f_1\rangle(\bar{x}) - x^*\big], \\ &= \bigcap_{x^* \in D^*_P f_2(\bar{x})(y^*)} \big[D^*_P f_1(\bar{x})(y^*) - x^*\big], \end{aligned}$$

which thus yields (2.4) and completes the proof of the proposition. □

Our next topic concerns *exact chain rules* for proximal subgradients. Consider first the *general composition* of the type

$$(\varphi \circ f)(x) := \varphi(x, f(x)) \tag{2.5}$$

involving an extended-real-valued function $\varphi\colon X \times Y \to \overline{R}$ and a single-valued mapping $f\colon X \to Y$ between Banach spaces. To formulate the following and some subsequent results, we need to define the symmetric proximal construction

$$\partial^+_P \varphi(\bar{x}) := -\partial_P(-\varphi)(\bar{x}) \tag{2.6}$$

that is called, in accordance with [8,15], the *proximal upper subdifferential* of $\varphi\colon X \to \overline{R}$ at $\bar{x}$ with $|\varphi(\bar{x})| < \infty$. The set (2.6) is also known as the "proximal superdifferential" of φ at $\bar{x}$; see, e.g., [4].

Theorem 2.6. (proximal chain rules for general compositions). *Given the general composition* (2.5)*, suppose that f is Lipschitz continuous around $\bar{x}$ and that φ is finite at $(\bar{x}, \bar{y})$ with $\bar{y} := f(\bar{x})$. Furthermore, we assume that $\partial^+_P \varphi(\bar{x}, \bar{y}) \neq \emptyset$. Then*

$$\partial_P(\varphi \circ f)(\bar{x}) \subset \bigcap_{(x^*, y^*) \in \partial^+_P \varphi(\bar{x}, \bar{y})} \big[x^* + D^*_P f(\bar{x})(y^*)\big] = \bigcap_{(x^*, y^*) \in \partial^+_P \varphi(\bar{x}, \bar{y})} \big[x^* + \partial_P\langle y^*, f\rangle(\bar{x})\big]. \tag{2.7}$$

If in addition $\varphi \in C^2$ around $(\bar{x}, \bar{y})$, we have the equality

$$\partial_P(\varphi \circ f)(\bar{x}) = \nabla_x \varphi(\bar{x}, \bar{y}) + D^*_P f(\bar{x})(\nabla_y \varphi(\bar{x}, \bar{y})) = \nabla_x \varphi(\bar{x}, \bar{y}) + \partial_P\langle \nabla_y \varphi(\bar{x}, \bar{y}), f\rangle(\bar{x}). \tag{2.8}$$

Proof. Take arbitrary $u^* \in \partial_P(\varphi \circ f)(\bar{x})$ and $(x^*, y^*) \in \partial^+_P \varphi(\bar{x}, \bar{y})$. By definition (0.1), we find $\sigma_1 > 0, \gamma_1 > 0$ such that

$$\langle u^*, x - \bar{x}\rangle \quad \leq \varphi(x, f(x)) - \varphi(\bar{x}, f(\bar{x})) + \sigma_1\|x - \bar{x}\|^2$$

whenever $x \in B_{\gamma_1}(\bar{x})$. Since $(-y^*, -x^*) \in \partial_P(-\varphi)(\bar{x}, \bar{y})$, we employ again definition (0.1) and find $\sigma_2 > 0, \gamma_2 > 0$ such that

$$\varphi(x, y) - \varphi(\bar{x}, \bar{y}) \le \langle x^*, x - \bar{x}\rangle + \langle y^*, y - \bar{y}\rangle + \sigma_2(\|x - \bar{x}\| + \|y - \bar{y}\|)^2$$

whenever $\|x - \bar{x}\| + \|y - \bar{y}\| < \gamma_2$. Choosing

$$\sigma := \sigma_1 + \sigma_2 \ \text{ and } \ \gamma := \min\Big\{\frac{\gamma_1}{2}, \frac{\gamma_2}{2(\ell + 1)}\Big\},$$

where $\ell \ge 0$ is the Lipschitzian modulus of f around $\bar{x}$, we have

$$\begin{aligned}\langle u^*, x - \bar{x}\rangle &\le \varphi(x, f(x)) - \varphi(\bar{x}, f(\bar{x})) + \sigma_1\|x - \bar{x}\|^2\\ &\le \langle x^*, x - \bar{x}\rangle + \langle y^*, f(x) - f(\bar{x})\rangle + \sigma_1\|x - \bar{x}\|^2 + \sigma_2(\|x - \bar{x}\| + \|f(x) - f(\bar{x})\|)^2\\ &\le \langle x^*, x - \bar{x}\rangle + \langle y^*, f(x) - f(\bar{x})\rangle + \sigma(\|x - \bar{x}\| + \|f(x) - f(\bar{x})\|)^2\end{aligned}$$

whenever $\|x - \bar{x}\| < \gamma$. This implies that

$$\langle u^* - x^*, x - \bar{x}\rangle - \langle y^*, f(x) - f(\bar{x})\rangle \le \sigma(\|x - \bar{x}\| + \|f(x) - \bar{y}\|)^2 \ \text{ as } \ \|x - \bar{x}\| < \gamma,$$

which yields $(u^* - x^*, -y^*) \in N_P((\bar{x}, \bar{y}); \operatorname{gph} f)$. Using the coderivative definition (1.6) and the scalarization formula (2.3) from Theorem 2.4, we get

$$u^* - x^* \in D_P^* f(\bar{x})(y^*) = \partial_P\langle y^*, f\rangle(\bar{x})$$

and arrive at both chain rules (2.7) for general compositions.

To justify equality (2.8) under the assumption on $\varphi \in C^2$ around $(\bar{x}, \bar{y})$, we take an arbitrary element $u^* \in \nabla_x\varphi(\bar{x}, \bar{y}) + D_P^* f(\bar{x})(\nabla_y(\bar{x}, \bar{y}))$ and get by the coderivative definition (1.6) that

$$(u^* - \nabla_x\varphi(\bar{x}, \bar{y}), -\nabla_y\varphi(\bar{x}, \bar{y})) \in N_P((\bar{x}, \bar{y}); \operatorname{gph} f).$$

Employing the proximal normal description (1.2) and the Lipschitz continuity of f around $\bar{x}$, we find $\sigma_1, \gamma_1 > 0$ such that

$$\begin{aligned}\langle u^* - \nabla_x\varphi(\bar{x}, \bar{y}), x - \bar{x}\rangle - \langle \nabla_y\varphi(\bar{x}, \bar{y}), f(x) - f(\bar{x})\rangle &\le \sigma_1(\|x - \bar{x}\| + \|f(x) - f(\bar{x})\|)^2\\ &\le \sigma_1(\ell + 1)^2\|x - \bar{x}\|^2\end{aligned}$$

whenever $\|x - \bar{x}\| < \gamma_1$. Furthermore, the assumption on $\varphi \in C^2$ around $(\bar{x}, \bar{y})$ yields the existence of $\sigma_2, \gamma_2 > 0$ such that

$$\langle \nabla_x\varphi(\bar{x}, \bar{y}), x - \bar{x}\rangle + \langle \nabla_y\varphi(\bar{x}, \bar{y}), y - \bar{y}\rangle \le \varphi(x, y) - \varphi(\bar{x}, \bar{y}) + \sigma_2(\|x - \bar{x}\| + \|y - \bar{y}\|)^2$$

whenever $\|x - \bar{x}\| + \|y - \bar{y}\| < \gamma_2$. Denoting

$$\gamma := \min\Big\{\frac{\gamma_1}{2}, \frac{\gamma_2}{2(\ell + 1)}\Big\} \ \text{ and } \ \sigma := 2\max\big\{\sigma_1(\ell + 1)^2, \sigma_2(\ell + 1)^2\big\},$$

we obtain the estimate

$$\langle u^*, x - \bar{x}\rangle \le \varphi(x, f(x)) - \varphi(\bar{x}, f(\bar{x})) + \sigma\|x - \bar{x}\|^2 \ \text{ as } \ \|x - \bar{x}\| < \gamma,$$

which implies (2.8) and thus completes the proof of the theorem. □

When $\varphi = \varphi(y)$, the general composition (2.6) reduces to the *standard composition* $(\varphi \circ f)(x) := \varphi(f(x))$, and we get the following consequences of Theorem 2.6.

Corollary 2.7. (proximal chain rules for standard compositions). *Let $f\colon X \to Y$ be Lipschitz continuous around $\bar{x}$, and let $\varphi\colon Y \to \overline{R}$ be finite at $\bar{y} := f(\bar{x})$. Then*

$$\partial_P(\varphi \circ f)(\bar{x}) \subset \bigcap_{y^* \in \partial_P^+ \varphi(\bar{y})} D_P^* f(\bar{x})(y^*) = \bigcap_{y^* \in \partial_P^+ \varphi(\bar{y})} \partial_P \langle y^*, f\rangle(\bar{x})$$

provided that $\partial_P^+ \varphi(\bar{y}) \neq \emptyset$. If an addition $\varphi \in C^2$ around $\bar{y}$, then

$$\partial_P(\varphi \circ f)(\bar{x}) = D_P^* f(\bar{x})(y^*) = D_P^* f(\bar{x})(\nabla \varphi(\bar{y})) = \partial_P \langle \nabla \varphi(\bar{y}), f\rangle(\bar{x}).$$

The next corollary of Theorem 2.6 gives new chain rules for limiting subgradients. Along with the limiting subdifferential (1.7), we employ its upper counterpart, the *limiting upper subdifferential* of φ at $\bar{x}$ with $|\varphi(\bar{x})| < \infty$, defined by

$$\partial^+ \varphi(\bar{x}) := -\partial(-\varphi)(\bar{x}). \tag{2.9}$$

According to [8, Subsection 2.5.2D], construction (2.9) agrees with the *basic* (limiting Fréchet) upper subdifferential of φ at $\bar{x}$ if X is Hilbert (and also in some other cases of reflexive Banach spaces), and hence $\partial^+ \varphi(\bar{x}) \neq \emptyset$ in such cases provided that the function φ is Lipschitz continuous around $\bar{x}$; see [8, Corollary 2.25].

Corollary 2.8. (chain rules for limiting subgradients). *Let $f\colon X \to Y$ be Lipschitz continuous around $\bar{x}$, where the unit dual balls of both spaces X and Y are weak* sequentially compact in X^* and Y^*, respectively, and where $\bar{y} := f(\bar{x})$. Assume also that φ is Lipschitz continuous around $(\bar{x}, \bar{y})$ and that $\partial_P^+ \varphi(x, y) \neq \emptyset$ for all (x, y) around $(\bar{x}, \bar{y})$. Then*

$$\partial(\varphi \circ f)(\bar{x}) \subset \bigcup_{(x^*, y^*) \in \partial^+ \varphi(\bar{x}, \bar{y})} \big[x^* + D^* f(\bar{x})(y^*)\big].$$

In particular, for $\varphi = \varphi(y)$ we have

$$\partial(\varphi \circ f)(\bar{x}) \subset \bigcup_{y^* \in \partial^+ \varphi(\bar{y})} D^* f(\bar{x})(y^*).$$

Proof. To justify the limiting chain rules, we employ the procedure similar to the proof of Corollary 2.3 with using the proximal chain rule of Theorem 2.6 and taking into account that the limiting upper subdifferential (2.9) is bounded for locally Lipschitzian functions. □

Employing the major calculus results of Theorems 2.1 and 2.6, we derive now some other calculus rules for proximal subgradients in Banach spaces. The next theorem gives a general *product rule* involving Lipschitzian functions.

Theorem 2.9. (product rule for proximal subgradients). *Let the functions $\varphi\colon X \to \overline{R}$, $i = 1, 2$, be Lipschitz continuous around $\bar{x}$. Assume that $\partial_P\big(-\varphi_1(\bar{x})\varphi_2\big)(\bar{x}) \neq \emptyset$. Then*

$$\partial_P\big(\varphi_1 \cdot \varphi_2\big)(\bar{x}) \subset \bigcap_{x^* \in \partial_P(-\varphi_1(\bar{x})\varphi_2)(\bar{x})} \big[\partial_P\big(\varphi_2(\bar{x})\varphi_1\big)(\bar{x}) - x^*\big]. \tag{2.10}$$

Proof. Define $f\colon X \to I\!R^2$ and $\psi\colon I\!R^2 \to I\!R$ by

$$f(x) := \big(\varphi_1(x), \varphi_2(x)\big) \quad \text{and} \quad \psi(y_1, y_2) := y_1 \cdot y_2.$$

Note that in this case $\partial_P\psi(\bar{y}_1,\bar{y}_2)=\nabla\psi(\bar{y}_1,\bar{y}_2)=(\bar{y}_2,\bar{y}_1)$. Then $\varphi_1\cdot\varphi_2=\psi\circ f$, and we use the chain rule from Corollary 2.7 to obtain

$$\partial_P\big(\varphi_1\cdot\varphi_2\big)(\bar{x})=D_P^*f(\bar{x})\big(\nabla\psi(f(\bar{x})\big)=D_P^*f(\bar{x})\big(\varphi_2(\bar{x}),\varphi_1(\bar{x})\big). \tag{2.11}$$

Since $f(x)=f_1(x)-f_2(x)$ with $f_1(x):=(\varphi_1(x),0)$ and $f_2(x):=(0,-\varphi_2(x))$, we derive from the coderivative difference rule of Corollary 2.5 that

$$D_P^*f(\bar{x})(u_1,u_2)\subset\bigcap_{x^*\in D_P^*f_2(\bar{x})(u_1,u_2)}\big[D_P^*f_1(\bar{x})(u_1,u_2)-x^*\big]. \tag{2.12}$$

Using now the scalarization formula (2.3) from Theorem 2.4 and the obvious representation

$$N_P((\bar{x}_1,\bar{x}_2);\Omega_1\times\Omega_2)=N_P(\bar{x}_1;\Omega_1)\times N_P(\bar{x}_2;\Omega_2)$$

of proximal normals in product spaces, we have

$$D_P^*f_1(\bar{x})\big(\varphi_2(\bar{x}),\varphi_1(\bar{x})\big)=D_P^*\varphi_1(\bar{x})\big(\varphi_2(\bar{x})\big)=\partial_P\big(\varphi_2(\bar{x})\varphi_1\big)(\bar{x})\quad\text{and}$$

$$D_P^*f_2(\bar{x})\big(\varphi_2(\bar{x}),\varphi_1(\bar{x})\big)=D_P^*\big(-\varphi_2\big)(\bar{x})\big(\varphi_1(\bar{x})\big)=\partial_P\big(-\varphi_1(\bar{x})\varphi_2\big)(\bar{x}),$$

and thus inclusion (2.10) follows from (2.11) and (2.12). □

To continue, we present the following *reciprocal rule* for proximal subgradients, which is then used in the proof of the proximal quotient rule.

Proposition 2.10. (reciprocal rule for proximal subgradients). *Let $\varphi\colon X\to I\!R$ be Lipschitz continuous around $\bar{x}$. Assume that $\varphi(\bar{x})\neq 0$. Then*

$$\partial_P\Big(\frac{1}{\varphi}\Big)(\bar{x})=\frac{\partial_P(-\varphi)(\bar{x})}{(\varphi(\bar{x}))^2}.$$

Proof. Consider $\phi\colon I\!R\to I\!R$ defined by $\phi(x):=\dfrac{1}{x}$. Then Corollary 2.7 and Theorem 2.4 give us

$$\partial_P\Big(\frac{1}{\varphi}\Big)(\bar{x})=\partial_P(\phi\circ\varphi)(\bar{x})=D_P^*\varphi(\bar{x})\big(\nabla\phi(\varphi(\bar{x})\big).$$

Since $\nabla\phi(\varphi(\bar{x}))=-\dfrac{1}{(\varphi(\bar{x}))^2}$, we have

$$D_P^*\varphi(\bar{x})\big(\nabla\phi(\varphi(\bar{x})\big)=D_P^*\varphi(\bar{x})\Big(-\frac{1}{(\varphi(\bar{x}))^2}\Big)=\partial_P\Big(-\frac{1}{(\varphi(\bar{x}))^2}\varphi\Big)(\bar{x})=\frac{1}{(\varphi(\bar{x}))^2}\partial_P(-\varphi)(\bar{x}),$$

which completes the proof. □

Theorem 2.11. (quotient rule for proximal subgradients). *Let $\varphi_i\colon X\to I\!R$, $i=1,2$, be Lipschitz continuous around $\bar{x}$. Assume that $\partial_P(\varphi_1(\bar{x})\varphi_2)(\bar{x})\neq\emptyset$ and that $\varphi_2(\bar{x})\neq 0$. Then*

$$\partial_P\Big(\frac{\varphi_1}{\varphi_2}\Big)(\bar{x})\subset\bigcap_{x^*\in\partial_P(\varphi_1(\bar{x})\varphi_2)(\bar{x})}\frac{\big[\partial_P\big(\varphi_2(\bar{x})\varphi_1\big)(\bar{x})-x^*\big]}{[\varphi_2(\bar{x})]^2}$$

Proof. We justify this similarly to the proof of Theorem 2.9 using the representation $\Big(\dfrac{\varphi_1}{\varphi_2}\Big) = \phi\circ f$, where $f\colon X\to I\!R^2$ and $\phi\colon I\!R^2\to I\!R$ are defined by

$$f(x) := \big(\varphi_1(x),\varphi_2(x)\big)\ \text{ and }\ \phi(y_1,y_2) := y_1/y_2.$$

By the classical chain rule we have

$$\nabla\phi\big(f(\bar{x})\big) = \Big(\frac{1}{\varphi_2(\bar{x})}, -\frac{\varphi_1(\bar{x})}{(\varphi_2(\bar{x}))^2}\Big).$$

Let us represent f in the difference form $f = f_1 - f_2$ with $f_1(x) := (\varphi_1(x), 0)$ and $f_2(x) := (0, -\varphi_2(x))$. The assumptions made in the theorem ensure that

$$D^*_P f_2(\bar{x})\Big(\frac{1}{\varphi_2(\bar{x})}, -\frac{\varphi_1(\bar{x})}{(\varphi_2(\bar{x}))^2}\Big) = \frac{1}{(\varphi_2(\bar{x}))^2}\partial_P\big(\varphi_1(\bar{x})\varphi_2\big)(\bar{x}) \neq \emptyset \quad\text{and}$$

$$D^*_P f_1(\bar{x})\Big(\frac{1}{\varphi_2(\bar{x})}, -\frac{\varphi_1(\bar{x})}{(\varphi_2(\bar{x}))^2}\Big) = \partial_P\Big(\frac{1}{\varphi_2(\bar{x})}\varphi_1\Big)(\bar{x}).$$

Employing Theorem 2.1, we have therefore that

$$\begin{aligned}\partial_P\Big(\frac{\varphi_1}{\varphi_2}\Big)(\bar{x}) &\subset \bigcap_{x^*\in\frac{1}{(\varphi_2(\bar{x}))^2}\partial_P(\varphi_1(\bar{x})\varphi_2)(\bar{x})}\Big[\partial_P\big(\frac{1}{\varphi_2(\bar{x})}\varphi_1\big)(\bar{x}) - x^*\Big]\\ &= \bigcap_{x^*\in\partial_P(\varphi_1(\bar{x})\varphi_2)(\bar{x})}\frac{\partial_P\big(\varphi_2(\bar{x})\varphi_1\big)(\bar{x}) - x^*}{(\varphi_2(\bar{x}))^2},\end{aligned}$$

which completes the proof. □

Finally in this section, we present a useful result allowing us to evaluate proximal subgradients of the *minimum function*

$$\big(\wedge\varphi_j\big)(x) := \min\big\{\varphi_j(x)\big|\ j=1,\ldots,n\big\},\ n\geq 2, \tag{2.13}$$

for the extended-real-valued functions $\varphi_j\colon X\to\overline{R}$ in Banach spaces. Consider the *active index set*

$$J(x) := \big\{j\in\{1,\ldots,n\}\big|\ \varphi_j(x) = (\wedge\varphi_j)(x)\big\}$$

needed in the following *minimum rule* for proximal subgradients.

Proposition 2.12. (proximal subgradients of minimum functions). *Given the minimum function* (2.13) *finite at some point $\bar{x}$, we have*

$$\partial_P\big(\wedge\varphi_j\big)(\bar{x}) \subset \bigcap_{j\in J(\bar{x})}\partial_P\varphi_j(\bar{x}).$$

Proof. Pick an arbitrary subgradient $x^*\in\partial_P(\wedge\varphi_j)(\bar{x})$. By definition (0.1), find $\sigma,\gamma>0$ such that

$$\langle x^*, x-\bar{x}\rangle \leq \big(\wedge\varphi_j\big)(x) - \big(\wedge\varphi_j\big)(\bar{x}) + \sigma\|x-\bar{x}\|^2$$

whenever $\|x-\bar{x}\|<\gamma$. Taking any $j\in J(\bar{x})$, for such x we have the relations

$$\begin{aligned}\langle x^*, x-\bar{x}\rangle &\leq \big(\wedge\varphi_j\big)(x) - \big(\wedge\varphi_j\big)(\bar{x}) + \sigma\|x-\bar{x}\|^2\\ &= \big(\wedge\varphi_j\big)(x) - \varphi_j(\bar{x}) + \sigma\|x-\bar{x}\|^2\\ &\leq \varphi_j(x) - \varphi_j(\bar{x}) + \sigma\|x-\bar{x}\|^2,\end{aligned}$$

which implies that $x^* \in \partial_P \varphi_j(\bar{x})$ and thus completes the proof of the proposition. □

3. Proximal Subgradients of Marginal Functions

In this section we establish new results on evaluating proximal subgradients of *marginal functions*, which can be interpreted as *value functions* for general problems of *parametric optimization*. Then we provide specifications of general results for the case of parametric problems of nonlinear programming with equality and inequality constraints given by differentiable functions.

Given Banach spaces X and Y, consider the parametric optimization problem

$$\text{minimize } \varphi(x,y) \quad \text{subject to } y \in F(x) \tag{3.1}$$

with the *cost function* $\varphi\colon X \times Y \to I\!\!R$ and *constraint mapping* $F\colon X \rightrightarrows Y$. In what follows we always assume that F has a closed graph around reference points. The corresponding *marginal/value function* in (3.1) is given by

$$\mu(x) := \inf\big\{\varphi(x,y)\big|\ y \in F(x)\big\}. \tag{3.2}$$

The *solution/argminimum map* of the problem is defined by

$$M(x) := \big\{y \in F(x)\big|\ \mu(x) = \varphi(x,y)\big\}. \tag{3.3}$$

The following theorem is the main result of this section. It gives a general upper estimate and its C^2-cost specification for proximal subgradients of the marginal function (3.2) via the proximal coderivative of the constraint mapping F and the proximal upper subdifferential (reduced to the classical Fréchet gradient in the smooth case) of the cost function φ.

Theorem 3.1. (upper estimates for proximal subgradients of marginal functions). *Let μ be the marginal function in* (3.2) *finite at $\bar{x}$, and let $\bar{y} \in M(\bar{x})$ be an element of the solution set* (3.3) *such that $\partial_P^+\varphi(\bar{x},\bar{y}) \neq \emptyset$. Then*

$$\partial_P\mu(\bar{x}) \subset \bigcap_{(x^*,y^*)\in\partial_P^+\varphi(\bar{x},\bar{y})} \Big[x^* + D_P^*F(\bar{x},\bar{y})(y^*)\Big]. \tag{3.4}$$

Furthermore, for $\varphi \in C^2$ around $(\bar{x},\bar{y})$ we have

$$\partial_P\mu(\bar{x}) \subset \Big[\nabla_x\varphi(\bar{x},\bar{y}) + D_P^*F(\bar{x},\bar{y})\big(\nabla_y\varphi(\bar{x},\bar{y})\big)\Big]. \tag{3.5}$$

Proof. To prove (3.4), pick any $u^* \in \partial_P\mu(\bar{x})$. Then by definition (0.1) we find numbers $\sigma_1 > 0$ and $\gamma_1 > 0$ such that

$$\langle u^*, x-\bar{x}\rangle \le \mu(x) - \mu(\bar{x}) + \sigma_1\|x-\bar{x}\|^2$$

whenever $x \in B_{\gamma_1}(\bar{x})$. Since $\bar{y} \in M(\bar{x})$, we have

$$\langle u^*, x-\bar{x}\rangle \quad \le \mu(x) - \varphi(\bar{x},\bar{y}) + \sigma_1\|x-\bar{x}\|^2$$

for all $x \in B_{\gamma_1}(\bar{x})$. Now fix any $(x^*,y^*) \in \partial_P^+\varphi(\bar{x},\bar{y})$. Using

$$\partial_P^+\varphi(\bar{x},\bar{y}) = -\partial_P(-\varphi)(\bar{x},\bar{y}),$$

find by (0.1) positive numbers σ_2 and γ_2 such that

$$\varphi(x,y) - \varphi(\bar{x},\bar{y}) \le \langle x^*, x-\bar{x}\rangle + \langle y^*, y-\bar{y}\rangle + \sigma_2(\|x-\bar{x}\| + \|y-\bar{y}\|)^2$$

whenever $\|x-\bar{x}\|+\|y-\bar{y}\|<\gamma_2$. Denote

$$\gamma := \min\{\gamma_1,\gamma_2\} \quad \text{and} \quad \sigma := 2\max\{\sigma_1,\sigma_2\}.$$

For any $y\in F(x)$ with $\|x-\bar{x}\|+\|y-\bar{y}\|<\gamma$ we have $\mu(x)\le\varphi(x,y)$, and hence

$$\begin{aligned}\langle u^*,x-\bar{x}\rangle &\le \varphi(x,y)-\varphi(\bar{x},\bar{y})+\sigma_1\|x-\bar{x}\|^2\\ &\le \langle x^*,x-\bar{x}\rangle+\langle y^*,y-\bar{y}\rangle+\sigma_1\|x-\bar{x}\|^2+\sigma_2(\|x-\bar{x}\|+\|y-\bar{y}\|)^2\\ &\le \langle x^*,x-\bar{x}\rangle+\langle y^*,y-\bar{y}\rangle+\sigma(\|x-\bar{x}\|+\|y-\bar{y}\|)^2.\end{aligned}$$

This implies the estimate

$$\langle u^*-x^*,x-\bar{x}\rangle-\langle y^*,y-\bar{y}\rangle\le\sigma(\|x-\bar{x}\|+\|y-\bar{y}\|)^2$$

for all $(x,y)\in\operatorname{gph}F$ with $\|x-\bar{x}\|+\|y-\bar{y}\|<\gamma$. Therefore $u^*-x^*\in D^*_PF(\bar{x},\bar{y})(y^*)$, i.e., $u^*\in x^*+D^*_PF(\bar{x},\bar{y})(y^*)$, which justifies (3.4). When $\varphi\in C^2$ around $(\bar{x},\bar{y})$, we arrive at (3.5) due to (1.5) and thus complete the proof of the theorem. □

Observe that the proximal subdifferential inclusion (3.4) may hold as *equality* even for *nonsmooth* functions describing the cost and constraint in (3.1). To illustrate, we consider the following *example*:

Let $X=Y=I\!R$, $\varphi(x,y)=-|x|+2y$, and $F(x):=\{y\in I\!R\,|\,y\ge|x|\}$ in (3.1). We easily have $\mu(x)=|x|$ and $0\in S(0)$. Clearly

$$\partial_P\mu(0)=[-1,1],\quad \partial^+_P\varphi(0,0)=[-1,1]\times\{2\},\quad D^*_PF(0,0)(2)=[-2,2].$$

Thus inclusion (3.4) holds as equality, since

$$\bigcap_{(x^*,y^*)\in\partial^+_P\varphi(0,0)}\{x^*+D^*F(0,0)(y^*)\}=\bigcap_{x^*\in[-1,1])}\{x^*+[-2,2]\}=[-1,1].$$

When $\varphi=\varphi(y)$ in (3.1), we have the following simplifications of estimates (3.4) and (3.5).

Corollary 3.2. (marginal functions with parameter-independent costs). *Let $\varphi(x,y)=\varphi(y)$ in (3.1), and let the marginal function μ be finite at $\bar{x}$. Given $\bar{y}\in M(\bar{x})$ with $\partial^+_P\varphi(\bar{y})\ne\emptyset$, we have*

$$\partial_P\mu(\bar{x})\subset\bigcap_{y^*\in\partial^+_P\varphi(\bar{y})}D^*_PF(\bar{x},\bar{y})(y^*).$$

If in particular $\varphi\in C^2$ around $\bar{y}$, then

$$\partial_P\mu(\bar{x})\subset D^*_PF(\bar{x},\bar{y})\big(\nabla\varphi(\bar{y})\big).$$

Consider next the case of classical parametric constraints parametric in (3.1) when the constraint mapping $F(\cdot)$ describes the sets of *feasible solutions* to parametric problems of *nonlinear programming* with smooth data, i.e.,

$$F(x):=\Big\{y\in Y\Big|\quad \begin{array}{l}\varphi_i(x,y)\le0,\ \ i=1,\dots,m,\\ \varphi_i(x,y)=0,\ \ i=m+1,\dots,m+r\Big\},\end{array}\tag{3.6}$$

where $\varphi_i\colon X\times Y\to I\!R$, $i=1,\ldots,m+r$, are strictly differentiable at the point in question.

Define the standard *Lagrangian*

$$L(\bar x,\bar y,\lambda):=\varphi(\bar x,\bar y)+\lambda_1\varphi_1(\bar x,\bar y)+\ldots+\lambda_{m+r}\varphi_{m+r}(\bar x,\bar y)$$

and consider the set of *Lagrange multipliers* $\lambda=(\lambda_1,\ldots,\lambda_{m+r})\in I\!R^{m+r}$ corresponding to the decision variable y in the parametric problem (3.1), (3.6) at the reference point $(\bar x,\bar y)$ as

$$\Lambda(\bar x,\bar y):=\Big\{\lambda\in I\!R^{m+r}\Big|\ \nabla_y L(\bar x,\bar y,\lambda)=\nabla_y\varphi(\bar x,\bar y)+\sum_{i=1}^{m+r}\lambda_i\nabla_y\varphi_i(\bar x,\bar y)=0\ \text{ with }\ \lambda_1\ge 0,\ldots,\lambda_m\ge 0\ \text{ and }\ \lambda_i\varphi_i(\bar x,\bar y)=0\ \text{ if }\ i=1,\ldots,m\Big\}.$$

For convenience we consider also the *constraint Lagrangian*

$$L_0(\bar x,\bar y,\lambda):=\lambda_1\varphi_1(\bar x,\bar y)+\ldots+\lambda_{m+r}\varphi_{m+r}(\bar x,\bar y)$$

and the corresponding set of Lagrange multipliers depending on a given element $y^*\in Y^*$:

$$\Lambda_0(\bar x,\bar y;y^*):=\Big\{\lambda\in I\!R^{m+r}\Big|\ y^*+\nabla_y L_0(\bar x,\bar y,\lambda)=y^*+\sum_{i=1}^{m+r}\lambda_i\nabla_y\varphi_i(\bar x,\bar y)=0\ \text{ with }\ \lambda_1\ge 0,\ldots,\lambda_m\ge 0\ \text{ and }\ \lambda_i\varphi_i(\bar x,\bar y)=0\ \text{ if }\ i=1,\ldots,m\Big\}.$$

Theorem 3.3. (proximal subgradients of value functions in parametric nonlinear programming). *Let the marginal/value function μ from* (3.2) *be finite at $\bar x$, and let the constraint mapping F be defined by* (3.6)*, where the spaces X and Y are Asplund. Given $\bar y\in M(\bar x)$ from the corresponding argminimum set* (3.3)*, suppose that $\partial_P^+\varphi(\bar x,\bar y)\neq\emptyset$, that the constraint functions φ_i are strictly differentiable at $(\bar x,\bar y)$, and that the following parametric Mangasarian-Fromovitz constraint qualification holds:*

$$\begin{array}{l}\text{the gradients } \nabla\varphi_{m+1}(\bar x,\bar y),\ldots,\nabla\varphi_{m+r}(\bar x,\bar y) \text{ are linearly independent in } X^*\times Y^*;\\ \text{there is } w\in X\times Y \text{ such that } \langle\nabla\varphi_i(\bar x,\bar y),w\rangle=0 \text{ for } i=m+1,\ldots,m+r,\\ \text{and that } \langle\nabla\varphi_i(\bar x,\bar y),w\rangle<0 \text{ whenever } i=1,\ldots,m \text{ with } \varphi_i(\bar x,\bar y)=0.\end{array}\tag{3.7}$$

Then we have the inclusion

$$\partial_P\mu(\bar x)\subset\bigcap_{(x^*,y^*)\in\partial_P^+\varphi(\bar x,\bar y)}\ \bigcup_{\lambda\in\Lambda_0(\bar x,\bar y;y^*)}\Big[x^*+\nabla_x L_0(\bar x,\bar y,\lambda)\Big].\tag{3.8}$$

If in addition $\varphi\in C^2$ around $(\bar x,\bar y)$ for all $\bar y\in M(\bar x)$, then

$$\partial_P\mu(\bar x)\subset\bigcap_{\bar y\in M(\bar x)}\Big\{\nabla_x L(\bar x,\bar y,\lambda)\ \Big|\ \lambda\in\Lambda(\bar x,\bar y)\Big\}.\tag{3.9}$$

Proof. Pick $u^*\in\partial_P\mu(\bar x)$. By Theorem 3.1 we have, for any $(x^*,y^*)\in\partial_P^+\varphi(\bar x,\bar y)$, that

$$u^*-x^*\in D_P^*F(\bar x,\bar y)(y^*).$$

Since $D_P^*F(\bar x,\bar y)(y^*)\subset D^*F(\bar x,\bar y)$, we apply the result of [8, Corollary 4.35] for computing the limiting coderivative (1.8) of the constraint mapping (3.6) between Asplund spaces under the constraint qualification (3.7). It gives

$$D_P^* F(\bar{x},\bar{y})(v^*) \subset \Big\{ u^* \in X^* \Big| \quad (u^*, -v^*) = \sum_{i=1}^{m+r} \lambda_i \nabla \varphi_i(\bar{x},\bar{y}),\ (\lambda_1,\ldots,\lambda_{m+r} \in I\!R^{m+r} \text{ with } \lambda_i \ge 0 \text{ and } \lambda_i \varphi_i(\bar{x},\bar{y}) = 0 \text{ as } i = 1,\ldots,m \Big\}.$$

This implies (3.8) due to (3.4). To get (3.9), we use (3.5) in the same way together with the representation (1.5) for C^2 functions. □

4. Necessary Conditions in Nondifferentiable Programming

In the concluding section of the paper we present some applications of the obtained proximal calculus results to necessary optimality conditions in nonsmooth problems of *difference optimization*, where cost functions are represented as differences of two functions; cf. [5] for the convex case and [10] for differences of general nonsmooth functions, where Fréchet subgradients are applied.

Let us start with simple conditions concerning *unconstrained* problems of minimizing difference functions $\varphi = \varphi_1 - \varphi_2$ on Banach spaces.

Proposition 4.1. (necessary conditions for unconstrained minimization of difference functions). *Let $\bar{x}$ be a local minimizer for the function $\varphi = \varphi_1 - \varphi_2$, where both $\varphi_i \colon X \to \overline{R}$ are finite at $\bar{x}$. Then one has the following inclusion for any $s \in (0,1]$*

$$\partial_{H(s)} \varphi_2(\bar{x}) \subset \partial_{H(s)} \varphi_1(\bar{x}). \tag{4.1}$$

In particular, one has

$$\partial_P \varphi_2(\bar{x}) \subset \partial_P \varphi_1(\bar{x}). \tag{4.2}$$

If in addition φ_2 is proximally regular at $\bar{x}$, then

$$\partial \varphi_2(\bar{x}) \subset \partial \varphi_1(\bar{x}). \tag{4.3}$$

Proof. Inclusion (4.1) immediately follows from the difference rule in the Remark 2.2 combined with the proximal Fermat rule (1.4). The second one (4.3) follows from (4.2) due to the proximal regularity definition (1.9) and the inclusion $\partial_P \varphi_1(\bar{x}) \subset \partial \varphi_1(\bar{x})$. □

To continue, we establish optimality conditions of minimizing functions represented as compositions.

Proposition 4.2. (necessary conditions for unconstrained minimization of compositions). *Let $\bar{x}$ be a local minimizer for the function $\psi = f \circ \varphi$, where $f : X \to Y$ is locally Lipschitzian around $\bar{x}$ and $\varphi_i \colon X \times Y \to \overline{R}$ is finite at $(\bar{x},\bar{y})$ with $\bar{y} := f(\bar{x})$. Then one has the following inclusion*

$$-\partial_P^+ \varphi(\bar{x},\bar{y}) \subset N_P((\bar{x},\bar{y}); \operatorname{gph} f). \tag{4.4}$$

Proof. The inclusion (4.4) follows from Fermat rule (1.4) and the chain rule from Theorem 2.6. □

Let us apply the optimality conditions of Proposition 4.1 to establish new necessary conditions for the so-called "weak sharp minima" in terms of proximal subgradients. Some optimality conditions for weak sharp minima, are established in [3] by using Clarke subgradients and in [10] by using Fréchet subgradients.

Given a proper function $\varphi \colon X \to \overline{R}$ and a nonempty subset $\Omega \subset X$ of a Banach space, recall that $S \subset \Omega$ is a set of *weak sharp minima* for φ relative to $\Omega \subset X$ with modulus $\alpha > 0$ if

$$\varphi(x) \ge \varphi(y) + \alpha \operatorname{dist}(x;S) \quad \text{for all} \quad x \in \Omega \quad \text{and} \quad y \in S,$$

where $\operatorname{dist}(x;S)$ stands for the distance function of the set S.

Corollary 4.3. (necessary conditions for unconstrained weak sharp minima). *Let $S \subset X$ be the set of weak sharp minima for the function φ relative to the whole space X with modulus α. Then for every $\bar{x} \in S$ we have*

$$\alpha I\!B^* \cap N_P(\bar{x}; S) \subset \partial_P \varphi(\bar{x}), \tag{4.5}$$

where $I\!B^* \subset X$ is the closed unit ball of X^*. If in addition S is proximally regular at $\bar{x}$ and X is reflexive, then

$$\alpha I\!B^* \cap N(\bar{x}; S) \subset \partial \varphi(\bar{x}). \tag{4.6}$$

Proof. By definition of weak sharp minimizers we have

$$\varphi(x) \geq \varphi(y) + \alpha \, \mathrm{dist}(x; S) \ \text{ for all } \ x \in X \ \text{ and } \ y \in S.$$

Thus every $y \in S$ is an optimal solution to the unconstrained problem of minimizing the *difference function* $\psi(x) := \varphi(x) - \alpha \, \mathrm{dist}(x; S)$. Employing Theorem 4.1, we get

$$\alpha \, \partial_P \mathrm{dist}(y; S) \subset \partial_P \varphi(y). \tag{4.7}$$

It is well-known (see, e.g., [2, Theorem 4.1]) that

$$\partial_P \mathrm{dist}(\bar{x}; S) = N_P(\bar{x}; S) \cap I\!B^*$$

in any Banach space. Substituting it into (4.7), we arrive at (4.5). Inclusion (4.6) in the proximal regularity and reflexivity assumptions follows from (4.3) in Proposition 4.1 due to the fact that the proximal regularity of a set in a reflexive Banach space agree with the proximal regularity of its distance function at any in-set point; see [2, Theorem 4.2]. □

Next we consider difference optimization problems under general *geometric* constraints given by

$$\text{minimize } \ \varphi(x) \ \text{ subject to } \ x \in \Omega, \tag{4.8}$$

where the cost function φ is represented as $\varphi = \varphi_1 - \varphi_2$.

Theorem 4.4. (necessary conditions for difference problems with geometric constraints). *Let $\bar{x}$ be a local solution to problem (4.8) in an Asplund space X, where φ is represented as $\varphi_1 - \varphi_2$. Assume that φ_1 is Lipschitz continuous around $\bar{x}$ and that Ω is locally closed around this point. Then one has the inclusion*

$$\partial_P \varphi_2(\bar{x}) \subset \partial \varphi_1(\bar{x}) + N(\bar{x}; \Omega). \tag{4.9}$$

Proof. Problem (4.8) under consideration can obviously be reformulated in the unconstrained difference form:

$$\text{minimize } \ \big[\varphi_1(x) + \delta(x; \Omega)\big] - \varphi_2(x) \ \text{ subject to } \ x \in \Omega.$$

By Proposition 4.1 we have

$$\partial_P \varphi_2(\bar{x}) \subset \partial_P \big(\varphi_1 + \delta(\cdot; \Omega)\big)(\bar{x})$$

for a given local minimizer $\bar{x}$ in (4.8). To justify (4.9), observe that

$$\partial_P \big(\varphi_1 + \delta(\cdot; \Omega)\big)(\bar{x}) \subset \partial(\varphi_1 + \delta(\cdot; \Omega)(\bar{x}) \subset \partial \varphi_1(\bar{x}) + N(\bar{x}; \Omega),$$

where the latter inclusion holds due to the sum rule for limiting subgradients in [8, Theorem 2.32]. This completes the proof of the theorem. □

As a useful corollary of Theorem 4.4, we get the following necessary conditions for weak sharp minima under general geometric constraints.

Corollary 4.5. (necessary conditions for weak sharp minima under geometric constraints). *Let S be the set of weak sharp minima for $\varphi\colon X \to \overline{R}$ relative to $\Omega \subset X$ with modulus $\alpha > 0$. Assume that X is Asplund, that φ is Lipschitz continuous around $\bar{x}$, and that Ω is locally closed around this point. Then*

$$\alpha I\!B^* \cap N_P(\bar{x}; S) \subset \partial\varphi(\bar{x}) + N(\bar{x}; \Omega).$$

Furthermore, we have

$$\alpha I\!B^* \cap N(\bar{x}; S) \subset \partial\varphi(\bar{x}) + N(\bar{x}; \Omega)$$

if in addition S is proximally regular at $\bar{x}$ and X is reflexive.

Proof. This follows from Theorem 4.4 by reducing weak sharp minima to constrained minimization of difference functions and employing the arguments from the proof of Corollary 4.3. □

References

[1] J. M. Borwein and D. Preiss (1987). A smooth variational principle with applications to subdifferentiability and to differentiability of convex functions. *Trans. Amer. Math. Soc.* **303**, 517–527.

[2] M. Bounkhel and L. Thibault (2002). On various notions of regularity of sets in nonsmooth analysis. *Nonlinear Analysis* **48**, 223–246.

[3] J. V. Burke and M. C. Ferris (1993). Weak sharp minima in mathematical programming. *SIAM Journal on Control and Optimization* **31**, 1340–1359.

[4] F. H. Clarke, Y. S. Ledyaev, R. J. Stern and P. R. Wolenski (1998). *Nonsmooth Analysis and Control Theory.* Springer-Verlag, New York.

[5] J.-B. Hiriart-Urruty (1985). Generalized differentiability, duality and optimization for problems dealing with differences of convex functions. *Lecture Notes in Economics and Mathematical Sciences* **256**, pp. 219–239. Springer-Verlag, Berlin.

[6] A. D. Ioffe (1983). On subdifferentiability spaces. *Ann. New York Acad. Sci.* **410**, 107–119.

[7] B. S. Mordukhovich (1976). Maximum principle in problems of time optimal control with nonsmooth constraints *J. Appl. Math. Mech.* **40**, 960–969.

[8] B. S. Mordukhovich (2006). *Variational Analysis and Generalized Differentiation, I: Basic Theory.* Springer-Verlag, Berlin.

[9] B. S. Mordukhovich (2006). *Variational Analysis and Generalized Differentiation, II: Applications.* Springer-Verlag, Berlin.

[10] B. S. Mordukhovich, N. M. Nam and N. D. Yen (2006). Fréchet subdifferential calculus and optimality conditions in nondifferentiable programming. *Optimization*, to appear.

[11] R. R. Phelps (1993). *Convex Functions, Monotone Operators and Differentiability*, 2nd edition. Springer-Verlag , Berlin.

[12] J.-P. Penot (1998). On the minimization of difference functions. *Journal of Global Optimization*, **12**, 373–382.

[13] B. T. Polyak (1987). *Introduction to Optimization.* Optimization Software, New York.

[14] R. T. Rockafellar (1981). Proximal subgradients, marginal values and augmented Lagrangians in nonconvex optimization. *Math. Oper. Res.* **6**, 424–436.

[15] R. T. Rockafellar and R. J-B. Wets (1998). *Variational Analysis.* Springer-Verlag, Berlin.