Journal of Mathematical Sciences, Vol. 116, No. 3, 2003

ON FRÉCHET SUBDIFFERENTIALS

A. Ya. Kruger UDC 517.98

Introduction

This survey is devoted to some aspects of the theory of *Fréchet subdifferentiation*. The selection of the material reflects the interests of the author and is far from being complete. The paper contains definitions and statements of some important results in the field with very few proofs. The author hopes that reading the paper will not be difficult even for those mathematicians whose main scientific interests are not in the field of nonsmooth analysis.

The variety of different subdifferentials known by now can be divided into two large groups: "simple" subdifferentials and "strict" subdifferentials. A *simple subdifferential* is defined at a given point and it does not take into account "differential" properties of a function in its neighborhood. Usually, such subdifferentials generalize some classical differentiability notions (*Fréchet*, *Gâteaux*, *Dini*, etc.). They are not widely used directly because of rather poor calculus.

Contrary to simple subdifferentials, the definitions of strict subdifferentials incorporate differential properties of a function near a given point. Usually, *strict subdifferentials* can be represented as (some kinds of) limits of simple ones. This procedure makes them generalizations of the notion of a *strict derivative* [14], enriches their properties, and allows constructing satisfactory calculus. The examples of *limiting subdifferentials* are the *generalized differential* (the *limiting Fréchet subdifferential*) [49, 53, 63, 66, 67] and the *approximate subdifferential* (the *limiting Dini subdifferential*) [38, 39, 41, 43]. The famous *generalized gradient* of Clarke [15, 17] can also be considered as being a strict subdifferential. The Warga's *derivate container* [98, 101] also belongs to this class.

The limiting subdifferentials proved to be very efficient in nonsmooth analysis and optimization (see [17, 38, 41–44, 49, 50, 52, 53, 62, 63, 67, 69, 74, 75, 93, 95, 97–102]), especially in finite dimensions. When applying limiting subdifferentials in infinite-dimensional spaces, one must be careful about nontriviality of limits in the weak* topology. Additional regularity conditions are needed (*compact epi-Lipschitzness* [7], *sequential normal compactness*, *partial sequential normal compactness* [74, 75, 77], etc.).

On the other hand, it is possible to formulate the results without taking limits and thus avoid the above-mentioned difficulties. Such statements are formulated (without additional regularity conditions) in terms of simple subdifferentials calculated at some points arbitrarily close to the point under consideration. They are usually referred to as *fuzzy* results [10, 12, 27–29, 40, 45, 62, 76, 78, 104]. In general, such results are stronger than the corresponding statements in terms of limiting subdifferentials. In this paper, we discuss only fuzzy results.

The paper consists of three sections. Section 1 is devoted to definitions and elementary properties of Fréchet subdifferentials, normal cones, and coderivatives. It partly follows the earlier papers [48, 51], some parts of which have never been published. The main fuzzy results (from the author's standpoint) in terms of Fréchet subdifferentials are presented in Sec. 2. Some of them are formulated by using strict δ-subdifferentials [55, 61]. The extended extremality notions [61] are discussed in Sec. 3. Being weaker than the traditional definitions, they describe some "almost extremal" points for which the known dual necessary conditions in terms of Fréchet subdifferentials become sufficient. Adopting these extended extremality notions leads to a form of duality in nonsmooth nonconvex optimization.

Some constants are defined in the paper which simplify the definitions and statements of the results.

Translated from Itogi Nauki i Tekhniki, Seriya Sovremennaya Matematika i Ee Prilozheniya. Tematicheskie Obzory. Vol. 94, Optimization and Related Topics–3, 2001.

1072–3374/03/1163–3325 $ 25.00

Mostly standard notation is used throughout the paper. X and Y denote Banach spaces and X^* and Y^* denote their topological duals. $\langle \cdot, \cdot \rangle$ is a bilinear form defining a canonical pairing between a space and its dual. $B_\rho(x)$ stands for a closed ball with center x and radius ρ. We write B_ρ instead of $B_\rho(0)$. The norms in the primal and the dual spaces will be denoted respectively by $\|\cdot\|$ and $\|\cdot\|_*$.

1. Fréchet Subdifferentials

Fréchet subdifferentials have been known for more than a quarter of a century. They were probably first introduced in finite dimensions in [3] (under the name "lower semidifferentials"). Some of their properties in an infinite-dimensional setting were investigated in [48, 51].

1.1. Definitions and elementary properties. Let X be a real Banach space and f be a function from X into an extended real line $\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$, finite at x.

A set

$$\partial f(x) = \left\{ x^* \in X^* : \liminf_{u \to x} \frac{f(u) - f(x) - \langle x^*, u - x \rangle}{\|u - x\|} \geq 0 \right\} \tag{1.1}$$

is called a (*Fréchet*) *subdifferential* of f at x. Its elements are sometimes referred to as (*Fréchet*) *subgradients* (*regular subgradients* [93]).

Set (1.1) is closed and convex. The following two propositions show that it generalizes the notions of a Fréchet derivative and a subdifferential of convex analysis.

Proposition 1.1. *If f is Fréchet differentiable at x with a derivative $\nabla f(x)$, then $\partial f(x) = \{\nabla f(x)\}$.*

Proposition 1.2. *If f is convex, then*

$$\partial f(x) = \{x^* \in X^* : f(u) - f(x) \geq \langle x^*, u - x \rangle \ \forall u \in X\}. \tag{1.2}$$

Let us also note that set (1.1) does not depend on the specific (equivalent) norm in X.

Example 1.1. Set (1.1) can be empty. Take $f : \mathbb{R} \to \mathbb{R} : f(u) = -|u|$, $u \in \mathbb{R}$. Obviously, $\partial f(0) = \emptyset$.

If $\partial f(x) \neq \emptyset$, we say that f is (*Fréchet*) *subdifferentiable* at x.

In addition to subdifferential (1.1), one can consider a (*Fréchet*) *superdifferential*

$$\partial^+ f(x) = \left\{ x^* \in X^* : \limsup_{u \to x} \frac{f(u) - f(x) - \langle x^*, u - x \rangle}{\|u - x\|} \leq 0 \right\}. \tag{1.3}$$

It is also closed and convex. If $\partial^+ f(x) \neq \emptyset$, we say that f is (*Fréchet*) *superdifferentiable* at x.

While set (1.1) consists of linear continuous functionals "supporting" f from below, functionals from (1.3) "support" f from above. Contrary to the classical case, the existence of two different derivative-like objects is quite natural for nonsmooth analysis: "differential" properties of a function "from below" and "from above" could be essentially different.

Surely in the nondifferential case at least one of the sets (1.1) and (1.3) must be empty.

Proposition 1.3. *Both sets* (1.1) *and* (1.3) *are simultaneously nonempty if and only if f is Fréchet differentiable at x. In this case, one has*

$$\partial f(x) = \partial^+ f(x) = \{\nabla f(x)\}.$$

In general, the following relation holds:

$$\partial(-f)(x) = -\partial^+ f(x).$$

Example 1.2. Both sets (1.1) and (1.3) can be simultaneously empty. Take $f : \mathbb{R} \to \mathbb{R} : f(u) = u \sin(1/u)$ for $u \neq 0$ and $f(0) = 0$. Then $\partial f(0) = \partial f^+(0) = \emptyset$.

Example 1.3. The fact that set (1.1) is a singleton does not imply differentiability. Take $f : \mathbb{R} \to \mathbb{R} : f(u) = \max(u \sin(1/u), 0)$ for $u \neq 0$ and $f(0) = 0$. Then f is nondifferentiable at 0, although we obviously have $\partial f(0) = \{0\}$.

Example 1.4. Fréchet differentiability is essential in Proposition 1.1. Gâteaux differentiable functions could be nonsubdifferentiable in the Fréchet sense. Take $f : \mathbb{R}^2 \to \mathbb{R} : f(u_1, u_2) = -\sqrt{|u_1|^2 + |u_2|^2}$ for $u_2 = u_1^2$ and $f(u_1, u_2) = 0$ otherwise. Obviously, f is Gâteaux differentiable at 0 (with the derivative equal to 0) while $\partial f(0) = \emptyset$.

Proposition 1.4. *If f is Gâteaux differentiable and Fréchet subdifferentiable at x with a (Gâteaux) derivative $\nabla f(x)$, then $\partial f(x) = \{\nabla f(x)\}$.*

Example 1.5. Under the conditions of Proposition 1.4, f can still be nondifferentiable in the sense of Fréchet. Take $f : \mathbb{R}^2 \to \mathbb{R} : f(u_1, u_2) = \sqrt{|u_1|^2 + |u_2|^2}$ for $u_2 = u_1^2$ and $f(u_1, u_2) = 0$ otherwise.

Remark 1.1. It is possible to define a *Gâteaux subdifferential* based on the notion of Gateaux differentiability. For this subdifferential, analogues of Propositions 1.1–1.3 and some other results hold. Considering Gâteaux (and other types of) subdifferentials can be useful in some applications. In general, a Gateaux subdifferential is a larger set than a Fréchet subdifferential.

Definition (1.1) of the Fréchet subdifferential can be reformulated in the following way.

Proposition 1.5. *$x^* \in \partial f(x)$ if and only if there exists a function $g : X \to \mathbb{R}$ such that*

(a) *$g(u) \leq f(u)$ for any $u \in X$ and $g(x) = f(x)$,*
(b) *g is Fréchet differentiable at x and $\nabla g(x) = x^*$.*

Condition (a) in Proposition 1.5 means that g "supports" f from below.

The sufficient part of Proposition 1.5 follows directly from definition (1.1). To prove the necessity, one can set $g(u) = \min(f(u), f(x) + \langle x^*, u - x\rangle)$, $u \in X$.

One more differentiability notion must be mentioned here. It is the so-called strict differentiability. Recall that f is called *strictly differentiable* [14] at x (with a strict derivative $\nabla f(x)$) if

$$\lim_{\substack{u \to x \\ u' \to x}} \frac{f(u') - f(u) - \langle \nabla f(x), u' - u\rangle}{\|u' - u\|} = 0. \tag{1.4}$$

Clearly, (1.4) is a more restrictive condition than simple Fréchet differentiability, although it is less restrictive than continuous differentiability. It is exactly the property of strict differentiability which is actually needed for such classical analysis results as the inverse function theorem or the implicit function theorem to hold.

Proposition 1.6. *If f is strictly differentiable at x with a derivative $\nabla f(x)$, then for any $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\partial f(u) \cup \partial^+ f(u) \subset \nabla f(x) + \varepsilon B^*$$

for all $u \in B_\delta(x)$.

Proof. It follows from (1.4) that for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|f(u') - f(u) - \langle \nabla f(x), u' - u\rangle| \leq \frac{\varepsilon}{2}\|u' - u\| \quad \forall u, u' \in B_{2\delta}(x). \tag{1.5}$$

Let $u \in B_\delta(x)$ and $x^* \in \partial f(u)$. Then it follows from (1.1) that there exists positive $\delta' \leq \delta$ such that

$$f(u') - f(u) - \langle x^*, u' - u\rangle \geq -\frac{\varepsilon}{2}\|u' - u\| \quad \forall u' \in B_{\delta'}(u). \tag{1.6}$$

Inequalities (1.5) and (1.6) yield

$$\langle x^* - \nabla f(x), u' - u\rangle \leq \varepsilon\|u' - u\| \quad \forall u' \in B_{\delta'}(u)$$

and consequently $\|x^* - \nabla f(x)\|_* \leq \varepsilon$. The case $x^* \in \partial^+ f(u)$ can be treated similarly. ∎

$\partial f(x)$ characterizes local properties of f near x, e.g., subdifferentiability implies lower semicontinuity.

Proposition 1.7. *If $\partial f(x) \neq \emptyset$, then f is lower semicontinuous at x.*

Proposition 1.8. *If f is lower semicontinuous at x, then $\partial f(x) = \partial(\operatorname{cl} f)(x)$, where $\operatorname{cl} f$ is a lower semicontinuous envelope of f.*

Comparing (1.1) and (1.2), we can see that in the convex case, the definition of a subdifferential is significantly simplified. Another example of such a simplification is given by positively homogeneous functions. Recall that f is *positively homogeneous* if $f(\lambda u) = \lambda f(u)$ for any $u \in X$ and any $\lambda > 0$.

Proposition 1.9. *Let f be positively homogeneous.*

(a) *If $f(0) = 0$, then*
$$\partial f(0) = \{x^* \in X^* : \ f(u) \geq \langle x^*, u\rangle, \ \forall u \in X\}.$$
(b) *If f is finite at x, then $\partial f(\lambda x) = \partial f(x)$ for any $\lambda > 0$.*

1.2. Simple calculus. The propositions below present some simple calculus results for Fréchet subdifferentials deduced directly from the definitions. More advanced statements of fuzzy calculus will be presented in Sec. 2.

Proposition 1.10. *If f attains a local minimum at x, then $0 \in \partial f(x)$.*

Proposition 1.11. *$\partial(\lambda f)(x) = \lambda \partial f(x)$ for any $\lambda > 0$.*

Proposition 1.12. *Let $f_1 : X \to \bar{\mathbb{R}}$ and $f_2 : X \to \bar{\mathbb{R}}$ be subdifferentiable at x. Then $f_1 + f_2$ is subdifferentiable at x and*
$$\partial(f_1 + f_2)(x) \supset \partial f_1(x) + \partial f_2(x). \tag{1.7}$$

Proposition 1.12 is an example of a *sum rule*. Usually, the sum rule is the central result of any subdifferential calculus. Unfortunately, inclusion (1.7) is almost useless: it does not allow one to decompose elements of the subdifferential of the sum of functions in terms of elements of subdifferentials of initial functions.

Corollary 1.12.1. *Let $f_1 : X \to \bar{\mathbb{R}}$ and $f_2 : X \to \bar{\mathbb{R}}$ be finite at x and $f_1 + f_2$ and $-f_1$ be subdifferentiable at x. Then f_2 is subdifferentiable at x and*
$$\partial f_2(x) \supset \partial(f_1 + f_2)(x) - \partial^+ f_1(x).$$

Combining Proposition 1.12, Corollary 1.12.1, and Proposition 1.3, we obtain the following result.

Corollary 1.12.2. *Let $f_1 : X \to \bar{\mathbb{R}}$ and $f_2 : X \to \bar{\mathbb{R}}$ be finite at x and f_1 be Fréchet differentiable at x. Then*
$$\partial(f_1 + f_2)(x) = \nabla f_1(x) + \partial f_2(x). \tag{1.8}$$

Corollary 1.12.2 gives an important case where equality holds in (1.7). It is interesting to note that (1.8) follows from (1.7).

Corollary 1.12.3. *Let $f_1 : X \to \bar{\mathbb{R}}$ and $f_2 : X \to \bar{\mathbb{R}}$ be finite at x and f_1 be Fréchet differentiable at x. If $f_1 + f_2$ attains a local minimum at x, then $-\nabla f_1(x) \subset \partial f_2(x)$.*

Now let us come to *chain rules*. Let h be a function on X taking values in another real Banach space Y. We assume that it satisfies at x the following *calmness* condition (cf. [93]):
$$\|f(u) - f(x)\| \leq l\|u - x\|$$
for some $l > 0$ and for all u in some neighborhood of x.

For any $y^* \in Y^*$, we consider a scalar function $\langle y^*, h\rangle$ defined by the equality

$$\langle y^*, h\rangle(u) = \langle y^*, h(u)\rangle.$$

Let $g : Y \to \bar{\mathbb{R}}$ be finite at $y = h(x)$. We consider a composition $f(u) = g(h(u))$, $u \in X$.

Proposition 1.13. *Let g be subdifferentiable at y and $\langle y^*, h\rangle$ be subdifferentiable at x for some $y^* \in \partial g(y)$. Then f is subdifferentiable at x and*

$$\partial f(x) \supset \partial\langle y^*, h\rangle(x).$$

The conclusion of Proposition 1.13 can be rewritten in the following form:

$$\partial f(x) \supset \bigcup\{\partial\langle y^*, h\rangle(x) : y^* \in \partial g(y)\}.$$

Corollary 1.13.1. *Let h be Fréchet differentiable at x. Then*

$$\partial f(x) \supset (\nabla h(x))^* \partial g(h(x)), \tag{1.9}$$

where $(\nabla h(x))^* : Y^* \to X^*$ is the adjoint operator to $\nabla h(x)$.

Taking into account the inverse function theorem [83], it is possible to deduce from Corollary 1.13.1 the following result giving conditions guaranteeing equality in (1.9).

Corollary 1.13.2. *Let h be strictly differentiable at x and $\nabla h(x)$ be invertible. Then in* (1.9) *the equality holds.*

Corollary 1.13.3. *Let $f(u) = g(au + b)$, $u \in X$. Then*

$$\partial f(x) = a\partial g(ax + b).$$

Proposition 1.14. *Let f be subdifferentiable at x and g be superdifferentiable at y. Then $\langle y^*, h\rangle$ is subdifferentiable at x for any $y^* \in \partial^+ g(y)$ and*

$$\partial f(x) \subset \partial\langle y^*, h\rangle(x).$$

Combining Propositions 1.13 and 1.14, we obtain the following corollary.

Corollary 1.14.1. *Let g be Fréchet differentiable at y. Then*

$$\partial f(x) = \partial\langle \nabla g(y), h\rangle(x).$$

As an easy consequence of Corollary 1.14.1, one can deduce formulas for subdifferentials of the product and the quotient of two scalar functions.

Corollary 1.14.2. *Let f_1 and f_2 be finite at x and satisfy the calmness condition at x. Let us denote $\alpha_i = f_i(x)$, $i = 1, 2$. Then*

$$\partial(f_1 \cdot f_2)(x) = \partial(\alpha_2 f_1 + \alpha_1 f_2)(x).$$

If $\alpha_2 \neq 0$, then

$$\partial\left(\frac{f_1}{f_2}\right)(x) = \frac{\partial(\alpha_2 f_1 - \alpha_1 f_2)(x)}{\alpha_2^2}.$$

Proposition 1.15. *Let $f(u) = \sup\limits_{i \in I} f_i(u)$, $u \in X$, where I is a nonempty set of indices and all the functions f_i, $i \in I$, and f are finite at x. Then*

$$\partial f(x) \supset \operatorname{cl}\operatorname{co} \bigcup_{i \in I_0(x)} \partial f_i(x),$$

where $I_0(x) = \{i \in I : f_i(x) = f(x)\}$ and cl co *denotes a convex closure.*

1.3. Fréchet subdifferentials and directional derivatives. Subdifferentials are dual-space objects. The Fréchet subdifferential was defined above (see (1.1)) directly, without invoking any local approximations of a function. Another approach to investigating nonsmooth functions consists of considering first some kind of directional derivative at a given point.

Let us define for some $z \in X$ (possibly infinite) limits

$$\begin{aligned} df(x)(z) &= \liminf_{\substack{t \to +0 \\ y \to z}} \frac{f(x+ty) - f(x)}{t}, \\ d_w f(x)(z) &= \liminf_{\substack{t \to +0 \\ y \xrightarrow{w} z}} \frac{f(x+ty) - f(x)}{t}, \end{aligned} \tag{1.10}$$

where $y \xrightarrow{w} z$ means that y tends to z in the *weak topology* of X. They are called respectively a *subderivative* (see [3, 38, 42, 86, 89, 93]) and a *weak subderivative* (see [48, 94]) of f at x in the direction z.

$df(x)(\cdot)$ and $d_w f(x)(\cdot)$ are positively homogeneous functions from X into $\mathbb{R} \cup \{\pm\infty\}$, lower semicontinuous in the norm and the weak topology of X respectively. The inequality

$$d_w f(x)(z) \le df(x)(z)$$

holds for any $z \in X$. If $\dim X < \infty$, then both subderivatives coincide. In general, the functions are different and they can differ from the usual directional derivative even if the latter exists. If f is *uniformly differentiable* [24, 47, 81] at x in the direction z, then $df(x)(z)$ reduces to the usual directional derivative. In the Lipschitz case, definition (1.10) can be simplified.

Proposition 1.16. *If f is Lipschitz continuous near x, then*

$$df(x)(z) = \liminf_{t \to +0} \frac{f(x+tz) - f(x)}{t}.$$

Surely, subderivatives can be used for characterizing local properties of f near x, e.g., the equality $df(x)(0) = 0$ implies the lower semicontinuity of f at x.

$d_w f(x)(\cdot)$ is in a sense the lowest possible directional derivative. It is closely related to subdifferential (1.1).

Proposition 1.17. *The following inclusion holds*:

$$\partial f(x) \subset \{x^* \in X^* : d_w f(x)(z) \ge \langle x^*, z\rangle \ \forall z \in X\}. \tag{1.11}$$

If X is reflexive, then in (1.11) *the equality holds.*

The first assertion of Proposition 1.17 follows directly from the definitions. The second assertion is a consequence of the fact that a unit ball in a reflexive space is weakly compact [2].

The set in the right-hand side of (1.11) can be taken as a definition of a subdifferential. It agrees with (1.1) in reflexive spaces, but, in general, this set is larger than (1.1).

1.4. Fréchet normals. Now let Ω be a nonempty set in X and let $x \in \Omega$.

Similarly to definition (1.1) of a (Fréchet) subdifferential, one can define a geometrical object, a (*Fréchet*) *normal cone*

$$N(x|\Omega) = \left\{ x^* \in X^* : \limsup_{u \xrightarrow{\Omega} x} \frac{\langle x^*, u - x\rangle}{\|u - x\|} \le 0 \right\} \tag{1.12}$$

to Ω at x, where $u \xrightarrow{\Omega} x$ means that $u \to x$ with $u \in \Omega$.

It is really a closed and convex cone closely related to the subdifferential defined above.

Let us consider an indicator function δ_Ω of Ω: $\delta_\Omega(u) = 0$ for $u \in \Omega$ and $\delta_\Omega(u) = \infty$ otherwise.

Proposition 1.18. $N(x|\Omega) = \partial \delta_\Omega(x)$.

Due to Proposition 1.18, one can deduce some properties of normal cones from the corresponding statements about subdifferentials. Thus, it follows from Proposition 1.2 that the normal cone (1.12) generalizes the corresponding notion of convex analysis.

Proposition 1.19. *If Ω is convex, then*

$$N(x|\Omega) = \{x^* \in X^* : \langle x^*, u - x\rangle \le 0 \ \forall u \in \Omega\}.$$

Proposition 1.20. $N(x|\Omega) = N(x|\operatorname{cl}\Omega)$.

Remark 1.2. A normal cone can be defined by (1.12) for any $x \in \operatorname{cl}\Omega$.

Proposition 1.21. *Let Ω be a cone. Then $N(\lambda x|\Omega) = N(x|\Omega)$ for any $\lambda > 0$ and*

$$N(0|\Omega) = \{x^* \in X^* : \langle x^*, u\rangle \le 0 \ \forall u \in \Omega\}.$$

Proposition 1.22. *Let $\Omega = \Omega_1 \cap \Omega_2$. Then*

$$N(x|\Omega) \supset N(x|\Omega_1) + N(x|\Omega_2).$$

Proposition 1.23. *Let f be Fréchet differentiable at x. If f attains at x a local minimum relative to Ω, then $-\nabla f(x) \in N(x|\Omega)$.*

Let us define for Ω a primal space local approximation, a *tangent cone*

$$T(x|\Omega) = \{z \in X : \exists\{x_k\} \in \Omega,\ \{\alpha_k\} \in \mathbb{R}_+,\ x_k \to x,\ \alpha_k(x_k - x) \to z\}$$

and a *weak tangent cone* (to Ω at x)

$$T_w(x|\Omega) = \{z \in X : \exists\{x_k\} \in \Omega,\ \{\alpha_k\} \in \mathbb{R}_+,\ x_k \to x,\ \alpha_k(x_k - x) \xrightarrow{w} z\}.$$

These are nonconvex cones, closed in the norm and weak topologies of X, respectively. They are widely used in optimization theory (see, e.g., [3, 4, 13, 24, 35, 94, 96]).

Surely, the inclusion $T(x|\Omega) \subset T_w(x|\Omega)$ holds and it can be strict [35]. The two cones coincide if $\dim X < \infty$ or Ω is convex.

It is easy to verify that the indicator functions of $T(x|\Omega)$ and $T_w(x|\Omega)$ coincide with the subderivative and the weak subderivative of δ_Ω at x, respectively.

Proposition 1.24. *The following inclusion holds*:

$$N(x|\Omega) \subset \{x^* \in X^* : \langle x^*, z\rangle \le 0 \ \forall z \in T_w(x|\Omega)\}. \tag{1.13}$$

If X is reflexive, then in (1.13) *the equality holds.*

The set in the right-hand side of (1.13), a *polar cone* [47] of $T_w(x|\Omega)$, can be taken as a definition of a normal cone. It agrees with (1.12) in reflexive spaces, but, in general, this set is larger than (1.12).

Proposition 1.25. *$x^* \in N(x|\Omega)$ if and only if there exists a function $g : X \to \mathbb{R}$ such that*
(a) *$g(u) \le 0$ for any $u \in \Omega$ and $g(x) = 0$;*
(b) *g is Fréchet differentiable at x and $\nabla g(x) = x^*$.*

Finally, we state four more simple statements about normal cones, which easily follow from the definition.

Proposition 1.26. *If $\Omega' \supset \Omega$, then $N(x|\Omega') \subset N(x|\Omega)$.*

Proposition 1.27. *Let $\Omega = \Omega_1 + \Omega_2$, $x = x_1 + x_2$, $x_i \in \Omega_i$, $i = 1, 2$. Then*

$$N(x|\Omega) \subset N(x_1|\Omega_1) \cap N(x_2|\Omega_2).$$

Proposition 1.28. *Let $\tilde\Omega = \{(\omega, \omega) : \omega \in \Omega\}$, $\tilde x = (x, x)$. Then*

$$N(\tilde x|\tilde\Omega) = \{(x_1^*, x_2^*) \in X^* \times X^* : x_1^* + x_2^* \in N(x|\Omega)\}.$$

Proposition 1.29. *Let $X = X_1 \times X_2$, $\Omega = \Omega_1 \times \Omega_2$, $x = (x_1, x_2)$, $x_i \in \Omega_i \subset X_i$, $i = 1, 2$. Then*

$$N(x|\Omega) = N(x_1|\Omega_1) \times N(x_2|\Omega_2).$$

1.5. Normal cones and subdifferentials. Another approach to defining the normal cone is based on considering first the subdifferential of the distance function. Recall that the distance function (to Ω) is defined by the formula

$$d_\Omega(u) = \inf_{\omega \in \Omega} \|u - \omega\|.$$

Proposition 1.30. $\partial d_\Omega(x) = \{x^* \in N(x|\Omega) : \|x^*\| \le 1\}$.

This statement was proved in [48]. Here we present an improved version of the proof.

Proof. Let $x^* \in \partial d_\Omega(x)$. Thus

$$\liminf_{u \to x} \frac{d_\Omega(u) - \langle x^*, u - x\rangle}{\|u - x\|} \ge 0 \tag{1.14}$$

and consequently

$$\limsup_{u \xrightarrow{\Omega} x} \frac{\langle x^*, u - x\rangle}{\|u - x\|} \le 0.$$

The last inequality means that $x^* \in N(x|\Omega)$. It also follows from (1.14) that for any $z \in X$, $z \neq 0$, we have

$$\liminf_{t \to +0} \frac{d_\Omega(x + tz) - t\langle x^*, z\rangle}{t} \ge 0$$

and consequently $\langle x^*, z\rangle \le \|z\|$. This yields $\|x^*\| \le 1$.

Now let $x^* \notin \partial d_\Omega(x)$ and $\|x^*\| \le 1$. We prove that $x^* \notin N(x|\Omega)$. According to the definition of the subdifferential, there exist a sequence $\{x_k\} \in X$ and a positive number ε_0 such that $x_k \to x$ and

$$d_\Omega(x_k) - \langle x^*, x_k - x\rangle + \varepsilon_0\|x_k - x\| < 0.$$

In order to achieve the goal, we must replace x_k by some point $\omega_k \in \Omega$. Let ω_k be a point in Ω such that

$$\|x_k - \omega_k\| \le d_\Omega(x_k) + \frac{\varepsilon_0}{2}\|x_k - x\|. \tag{1.15}$$

Adding the last two inequalities, we obtain

$$\langle x^*, x_k - x\rangle > \|x_k - \omega_k\| + \frac{\varepsilon_0}{2}\|x_k - x\|.$$

This yields

$$\langle x^*, \omega_k - x\rangle > \frac{\varepsilon_0}{2}\|x_k - x\|. \tag{1.16}$$

To complete the proof, we need some lower estimate for $\|x_k - x\|$ in terms of $\|\omega_k - x\|$. It follows from (1.15) that

$$\|x_k - \omega_k\| \le (1 + \varepsilon_0/2)\|x_k - x\|$$

and consequently

$$\|\omega_k - x\| \le (2 + \varepsilon_0/2)\|x_k - x\|.$$

Combining the last inequality with (1.16), we obtain

$$\langle x^*, \omega_k - x\rangle > \frac{\varepsilon_0}{\varepsilon_0 + 4}\|\omega_k - x\|.$$

Consequently $x^* \notin N(x|\Omega)$. □

The following corollary gives an equivalent definition of the normal cone. Contrary to the indicator function whose subdifferential can be used for defining the normal cone (see Proposition 1.18), the distance function is Lipschitz continuous. This makes it more convenient in some situations.

Corollary 1.30.1. $N(x|\Omega) = \{\lambda x^* : \lambda > 0,\ x^* \in \partial d_\Omega(x)\}$.

It follows from Proposition 1.18 that a normal cone is a particular case of a subdifferential. The converse is also true: the subdifferential of an arbitrary function can be equivalently defined through the normal cone to its epigraph. Recall that the *epigraph* of f is the set

$$\operatorname{epi} f = \{(u, \mu) \in X \times \mathbb{R} : f(u) \leq \mu\}.$$

Proposition 1.31. *The following assertions hold*:

(a) *if* $x^* \in \partial f(x)$, *then* $(x^*, -1) \in N(x, f(x)|\operatorname{epi} f)$;
(b) *if* $\mu \geq f(x)$ *and* $(x^*, \lambda) \in N(x, \mu|\operatorname{epi} f)$, *then* $\lambda \leq 0$;
(c) *if* $\lambda \neq 0$ *in* (b), *then* $\mu = f(x)$ *and* $-x^*/\lambda \in \partial f(x)$.

Corollary 1.31.1. $\partial f(x) = \{x^* \in X^* : (x^*, -1) \in N(x, f(x)|\operatorname{epi} f)\}$.

The case $\lambda = 0$ in Proposition 1.31(b) (the case of "horizontal normals" to the epigraph) can be important.

Let us define a set

$$\partial^\infty f(x) = \{x^* \in X^* : (x^*, 0) \in N(x, f(x)|\operatorname{epi} f)\}.$$

It is a convex cone, which is usually referred to as a *singular subdifferential*. Then the normal cone $N(x, f(x)|\operatorname{epi} f)$ is completely defined by the sets $\partial f(x)$ and $\partial^\infty f(x)$.

Corollary 1.31.2. *We have*

$$N(x, f(x)|\operatorname{epi} f) = \bigcup_{\lambda \geq 0} \lambda(\partial f(x), -1) \cup (\partial^\infty f(x), 0).$$

If f satisfies a calmness condition at x, then

$$N(x, f(x)|\operatorname{epi} f) = \bigcup_{\lambda \geq 0} \lambda(\partial f(x), -1).$$

Under the calmness condition, one has $\partial^\infty f(x) = \{0\}$. This remark proves the last assertion in Corollary 1.31.2.

One can also consider normals to the graph

$$\operatorname{gph} f = \{(u, \mu) \in X \times \mathbb{R} : f(u) = \mu\}$$

of f. It is a subset of $\operatorname{epi} f$.

Corollary 1.31.3. *The following inclusion holds*:

$$N(x, f(x)|\operatorname{gph} f) \supset (\partial f(x), -1) \cup (-\partial^+ f(x), 1).$$

It is possible to formulate the exact formula as in Corollary 1.31.2. To do this, one must use, in addition to the singular subdifferential, also the *singular superdifferential* (the definition is obvious) or assume the calmness condition.

1.6. Fréchet coderivatives. Starting from the definition of a normal cone, it is possible to define a derivative-like object for a set-valued mapping (multifunction) $F : X \rightrightarrows Y$ from X into another Banach space Y. To do this, one must consider the graph $\operatorname{gph} F = \{(u, v) \in X \times Y : v \in F(u)\}$ of F and the normal cone to the graph at some point $(x, y) \in \operatorname{gph} F$.

The multifunction $\partial F(x, y) : Y^* \rightrightarrows X^*$ defined by the equality

$$\partial F(x, y)(y^*) = \{x^* \in X^* :\ (x^*, -y^*) \in N(x, y|\operatorname{gph} F)\}$$

is called the (*Fréchet*) *coderivative* of F at (x, y).

If $F(u) = f(u) + R_+$, $u \in X$, for some function $f : X \to \bar{\mathbb{R}}$, then $\operatorname{gph} F = \operatorname{epi} f$ and it follows from Corollary 1.31.1 that $\partial F(x, f(x))(1) = \partial f(x)$.

If F is single-valued at x, we write $\partial F(x)$ instead of $\partial F(x, F(x))$.

If $F(u) = \{f(u)\}$ in a neighborhood of x for some (single-valued) function f, then (under the calmness condition) the coderivative reduces to the subdifferential of the scalar function $\langle y^*, f\rangle$ defined by the equality $\langle y^*, f\rangle(u) = \langle y^*, f(u)\rangle$, $u \in X$.

Proposition 1.32. *If f satisfies a calmness condition at x, then $\partial f(x)(y^*) = \partial\langle y^*, f\rangle(x)$.*

1.7. Proximal subdifferentials. In many cases, especially in finite dimensions and Hilbert spaces, the following subset of the Fréchet subdifferential could be of importance:

$$\begin{aligned} \partial_P f(x) = \{x^* \in X^* :\ &\exists \gamma > 0,\ \rho > 0 \text{ such that} \\ f(u) - f(x) - \langle x^*, u - x\rangle &+ \gamma\|u - x\|^2 \geq 0\ \forall u \in B_\rho(x)\}. \end{aligned} \tag{1.17}$$

The elements in $\partial_P f(x)$ supporting f at x from below up to an infinitely small (in comparison with $\|u - x\|$) term which is contrary to (1.1) is given in (1.17) explicitly: its degree equals to 2.

$\partial_P f(x)$ is called a *proximal subdifferential* of f at x [92, 93]. It is a convex set, which, in general, is not closed. $\partial_P f(x)$ reduces to the subdifferential if f is convex, but in a nonconvex case, it can be empty even if f is Fréchet differentiable at x.

Let us also note that (1.17) depends on the specific (equivalent) norm in X. In finite dimensions, it is usually used with the Euclidean norm.

(1.17) can be rewritten in the following equivalent form:

$$\partial_P f(x) = \left\{ x^* \in X^* : \liminf_{u \to x} \frac{f(u) - f(x) - \langle x^*, u - x\rangle}{\|u - x\|^2} > -\infty \right\}.$$

The geometrical counterpart of (1.17) is defined in a similar way:

$$N_P(x|\Omega) = \{x^* \in X^* : \exists \gamma > 0 \text{ such that } \langle x^*, u - x\rangle \leq \gamma\|u - x\|^2\ \forall u \in \Omega\},$$

or, equivalently,

$$N_P(x|\Omega) = \left\{ x^* \in X^* : \limsup_{u \xrightarrow{\Omega} x} \frac{\langle x^*, u - x\rangle}{\|u - x\|^2} < \infty \right\}.$$

It is a convex cone called a *proximal normal cone* to Ω at x [92, 93].

If x is a Hilbert space and $\langle \cdot, \cdot\rangle$ is an inner product (X^* can be identified with X under these assumptions), then one can use the following equivalent representation of a proximal normal cone: $x^* \in N_P(x|\Omega)$ if and only if x^* is *perpendicular* to Ω at x: $x^* = \alpha(u - x)$ for some $\alpha > 0$ and $u \in X$ such that x is the closest to u point in Ω. In other words, x belongs to the *metric projection*

$$\mathrm{Pr}_\Omega(u) = \{\omega \in \Omega : \|u - \omega\| = d(u, \Omega)\}$$

of u onto Ω.

Proximal normals were used in [63, 66, 67] when defining generalized normals. There also exist relations between proximal normals and generalized gradients of Clarke and approximate subdifferentials (see [8, 17, 44, 92]).

1.8. Strict Fréchet δ-subdifferentials. As was mentioned in the introduction, "simple" subdifferentials have poor calculus and their direct application is rather limited. There exists a way of enriching the properties of subdifferentials. It consists of considering differential properties of a function not only at a given point but also at points nearby.

Let us introduce a new derivative-like object based on (Fréchet) subdifferential (1.1):

$$\hat{\partial}_\delta f(x) = \bigcup_{\substack{u \in B_\delta(x) \\ |\operatorname{cl} f(u) - f(x)| \leq \delta}} \partial(\operatorname{cl} f)(u). \tag{1.18}$$

It depends on some positive δ; $\operatorname{cl} f$ denotes the lower semicontinuous envelope of f. Contrary to (1.1), set (1.18) can be nonconvex. We call it a *strict* (*Fréchet*) *δ-subdifferential* of f at x.

A *strict δ-superdifferential* of f at x can be defined similarly:

$$\hat{\partial}^+_\delta f(x) = \bigcup_{\substack{u \in B_\delta(x) \\ |\operatorname{cl}^\uparrow f(u) - f(x)| \le \delta}} \partial^+ (\operatorname{cl}^\uparrow f)(u), \tag{1.19}$$

where $\operatorname{cl}^\uparrow f$ is the upper semicontinuous envelope of f. The equality

$$\hat{\partial}^+_\delta f(x) = -\hat{\partial}^-_\delta (-f)(x)$$

holds.

Let us note that strict sub- and superdifferentials can be nonempty simultaneously and can be essentially different as in the nonsmooth case "differential" properties of a function "from below" and "from above" can differ significantly.

The set

$$\hat{\partial}^0_\delta \varphi(x) = \hat{\partial}_\delta f(x) \cup \hat{\partial}^+_\delta f(x) \tag{1.20}$$

is called the *strict δ-differential* of f at x.

Definitions (1.18)–(1.20) are modifications of the definitions of strict ε-semidifferentials introduced in [55]. Strict δ-subdifferentials were used in [59].

All "strict" sets (1.18)–(1.20) do have some properties of a strict derivative.

The corresponding geometrical objects are defined similarly: a *strict δ-normal cone* to a set,

$$\hat{N}_\delta(x|\Omega) = \bigcup \left[N(u|\operatorname{cl}\Omega) : u \in \operatorname{cl}\Omega \cap B_\delta(x) \right],$$

and a *strict δ-coderivative* [59] for a multifunction,

$$\hat{\partial}_\delta F(x,y)(y^*) = \{x^* \in X^* : (x^*, -y^*) \in \hat{N}_\delta(x, y|\operatorname{gph} F)\}.$$

They are closely related to strict δ-subdifferentials.

Let us note that the purpose of introducing strict δ-subdifferentials is mainly notational. They are convenient for formulating "fuzzy" results, but all of them can certainly be formulated in terms of ordinary subdifferentials.

1.9. Limiting subdifferentials. The limiting Fréchet subdifferentials are defined in [53, 63, 66, 67] as limits of "simple" subdifferentials. To simplify the definitions, we assume in this section that $f : X \to \bar{\mathbb{R}}$ is lower semicontinuous in a neighborhood of x.

A set

$$\begin{aligned} \bar{\partial} f(x) = \{x^* \in X^* : \exists \text{ sequences } \{x_k\} \subset X,\ \{x_k^*\} \subset X^* \text{ such that} \\ x_k \xrightarrow{f} x,\ x_k^* \xrightarrow{w^*} x^*, \text{ and } x_k^* \in \partial f(x_k),\ k = 1, 2, \dots\} \end{aligned} \tag{1.21}$$

is called a *limiting Fréchet subdifferential* of f at x.

The denotations $x_k \xrightarrow{f} x$ and $x_k^* \xrightarrow{w^*} x^*$ in (1.21) mean respectively that $x_k \to x$ with $f(x_k) \to f(x)$ (*f-attentive convergence* [93]) and x_k^* converges to x^* in the *weak* topology* of X^*. In [93], the elements of (1.21) are referred to as *general subgradients*.

Obviously, $\bar{\partial} f(x)$ is a weakly* sequentially closed set in X^*. In general, it is nonconvex. If f is strictly differentiable at x, then set (1.21) reduces to the derivative.

Using strict δ-subdifferentials, one can rewrite definition (1.21) in the following way:

$$\bar{\partial} f(x) = \bigcap_{\delta > 0} \operatorname{cl}^*\ \hat{\partial}_\delta f(x), \tag{1.22}$$

where cl^* denotes the weak* sequential closure.

The following formula holds:

$$\bar{\partial} f(x) = \limsup_{u \xrightarrow{f} x} \partial f(u),$$

where lim sup denotes the *sequential Kuratowski–Painlevé upper limit* of the multifunction $\partial f(\cdot)$ with respect to the norm topology in X and the weak* topology in X^*.

Other limiting objects (the *limiting superdifferential*, the *limiting differential*, the *limiting normal cone*, the *singular limiting subdifferential*, and the *limiting coderivative*) can be defined similarly.

Thus, the limiting normal cone to a closed set Ω is defined by the equality

$$\bar{N}(x|\Omega) = \bigcap_{\delta>0} \mathrm{cl}^* \ \hat{N}_\delta(x|\Omega). \tag{1.23}$$

It coincides with the limiting subdifferential of the indicator function δ_Ω of Ω. An analogue of Corollary 1.31.1 is also valid:

$$\bar{\partial} f(x) = \{x^* \in X^* : (x^*, -1) \in \bar{N}(x, f(x) | \operatorname{epi} f)\}. \tag{1.24}$$

If $\dim X < \infty$, then limiting normal cone (1.23) coincides with the conjugate cone defined in [66] as a set of limits of proximal normals. Due to (1.24), the limiting subdifferential (1.22) coincides in this case with the generalized derivative in [66].

Limiting objects (1.22) and (1.23) have been well investigated. They possess good calculus (see [63,67,69] for the properties of these objects in finite dimensions and [49,53,70,75] for infinite-dimensional generalization). Some examples of calculating limiting subdifferentials can be found in [63, 67].

The limiting subdifferentials and normal cones proved to be very efficient for formulating optimality conditions in nonsmooth optimization (see [50, 52, 62, 63, 67, 69, 70, 75]), especially in finite dimensions. When applying limiting subdifferentials in infinite-dimensional spaces, one must be careful about nontriviality of limits in the weak* topology. Additional regularity conditions are needed (*compact epi-Lipschitzness* [7], *sequential normal compactness*, *partial sequential normal compactness* [74, 75, 77], etc.).

Many nice finite-dimensional results in terms of limiting objects cannot be extended to infinite dimensions in full generality (see examples in [11, 12]). Such results can be formulated in infinite-dimensional spaces in a fuzzy form (see Sec. 2).

1.10. Fréchet ε-subdifferentials. In some cases, it is convenient to use the following modifications of subdifferentials depending on a parameter $\varepsilon \geq 0$:

$$\partial_\varepsilon f(x) = \left\{ x^* \in X^* : \liminf_{u \to x} \frac{f(u) - f(x) - \langle x^*, u - x \rangle}{\|u - x\|} \geq -\varepsilon \right\}, \tag{1.25}$$

$$\partial_\varepsilon^+ f(x) = \left\{ x^* \in X^* : \limsup_{u \to x} \frac{f(u) - f(x) - \langle x^*, u - x \rangle}{\|u - x\|} \leq \varepsilon \right\}. \tag{1.26}$$

They are called respectively a (*Fréchet*) *ε-subdifferential* and a (*Fréchet*) *ε-superdifferential* of f at x [51, 53] (see also [95]).

If $\varepsilon = 0$, the sets above coincide with the sub- and superdifferential defined by (1.1) and (1.3). Contrary to sets (1.1) and (1.3), the ε-sub- and ε-superdifferential (1.25) and (1.26) (for $\varepsilon > 0$) depend on the specific norm in X.

Proposition 1.33. *The following relation holds*:

$$\partial_\varepsilon f(x) = \bigcap_{\alpha > \varepsilon} \partial_\alpha f(x).$$

The following three propositions extend Propositions 1.2, 1.3, and 1.6, respectively.

Proposition 1.34. *If f is convex, then*

$$\begin{aligned}\partial_\varepsilon f(x) &= \partial f(x) + \varepsilon B^* \\ &= \{x^* \in X^* : f(u) - f(x) \geq \langle x^*, u - x\rangle - \varepsilon\|u - x\| \ \forall u \in X\}.\end{aligned} \tag{1.27}$$

Remark 1.3. Set (1.27) differs from the ε-subdifferential in the sense of convex analysis which is usually defined [91] as the set of $x^* \in X^*$ such that $f(u) - f(x) \geq \langle x^*, u - x\rangle - \varepsilon$ for all $u \in X$.

Proposition 1.35. *If $x_1^* \in \partial_{\varepsilon_1} f(x)$, $x_2^* \in \partial^+_{\varepsilon_2} f(x)$, $\varepsilon_1 \geq 0$, and $\varepsilon_2 \geq 0$, then $\|x_1^* - x_2^*\|_* \leq \varepsilon_1 + \varepsilon_2$.*

Proposition 1.36. *If f is strictly differentiable at x with a derivative $\nabla f(x)$, then for any $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\nabla f(x) \in \partial_\varepsilon f(u) \cap \partial^+_\varepsilon f(u)$$

for all $u \in B_\delta(x)$.

Using the scheme described above, one can define a *set of ε-normals* [51, 53, 62]

$$N_\varepsilon(x|\Omega) = \left\{x^* \in X^* : \limsup_{u \xrightarrow{\Omega} x} \frac{\langle x^*, u - x\rangle}{\|u - x\|} \leq \varepsilon\right\}$$

to Ω (it is not a cone for $\varepsilon > 0$) and an *ε-coderivative*

$$\partial_\varepsilon F(x, y)(y^*) = \{x^* \in X^* : (x^*, -y^*) \in N_\varepsilon(x, y|\operatorname{gph} F)\}$$

to a multifunction F and extend to the case $\varepsilon > 0$ the corresponding statements.

The following proposition generalizing Proposition 1.31 describes ε-normals to the epigraph.

Proposition 1.37. *The following assertions hold*:

(a) *if $x^* \in \partial_\varepsilon f(x)$, then $(x^*, -1) \in N_\varepsilon(x, f(x)|\operatorname{epi} f)$;*
(b) *if $\mu \geq f(x)$ and $(x^*, \lambda) \in N_\varepsilon(x, \mu|\operatorname{epi} f)$, then $\lambda \leq \varepsilon$;*
(c) *if $\lambda < -\varepsilon$ in* (b)*, then $\mu = f(x)$ and $-x^*/\lambda \in \partial_{\hat\varepsilon} f(x)$, where $\hat\varepsilon = \varepsilon(1 + |\lambda|^{-1}\|x^*\|_*)/(|\lambda| - \varepsilon)$.*

ε-Subdifferentials (1.25) can be used for defining a modified version of the strict δ-subdifferential:

$$\hat\partial_{\varepsilon,\delta} f(x) = \bigcup_{\substack{u \in B_\delta(x) \\ |\operatorname{cl} f(u) - f(x)| \leq \delta}} \partial_\varepsilon(\operatorname{cl} f)(u). \tag{1.28}$$

Set (1.28) is called a *strict (ε, δ)-subdifferential* of f at x. A *strict (ε, δ)-superdifferential*, a *strict set of (ε, δ)-normals*, and a *strict (ε, δ)-coderivative* can be defined similarly (see [60, 61, 63]).

In turn, strict (ε, δ)-subdifferentials (1.28) can be used for defining a kind of a limiting subdifferential

$$\tilde\partial f(x) = \bigcap_{\substack{\varepsilon > 0 \\ \delta > 0}} \operatorname{cl}^* \hat\partial_{\varepsilon,\delta} f(x). \tag{1.29}$$

It follows from [75] that sets (1.29) and (1.22) coincide on a broad class of Banach spaces, namely, on *Asplund* spaces.

1.11. Other subdifferentials.

1.11.1. Subdifferentials based on directional derivatives. Let f be directionally differentiable at x, i.e., the limit

$$f'(x)(z) = \lim_{t \to +0} \frac{f(x + tz) - f(x)}{t}$$

(possibly infinite) exists for any $z \in X$. The function $f'(x)(\cdot)$ is positively homogeneous. Its subdifferential at 0, i.e., the set

$$\underline{\partial} f(x) = \{x^* \in X^* : f'(x)(z) \geq \langle x^*, z\rangle \ \forall z \in X\},$$

is sometimes taken as a subdifferential of f at x [47].

If $f'(x)(\cdot)$ is convex, then f is called *locally convex* at x [47]. In this case, the application of convex analysis allows deriving some calculus for such subdifferentials. If $f'(x)(\cdot)$ is also proper and closed, then

$$f'(x)(z) = \sup\{\langle x^*, z\rangle : x^* \in \underline{\partial} f(x)\} \tag{1.30}$$

for any $z \in X$ and f is called *quasidifferentiable* at x [90].

Proposition 1.38. *We have $\partial f(x) \subset \underline{\partial} f(x)$. If $\dim X < \infty$ and f is Lipschitz continuous near x, then the equality holds in the inclusion.*

Proposition 1.38 follows easily from Propositions 1.17 and 1.16.

The class of functions admitting representation (1.30) is a subclass of a more general class of functions admitting the following representation [18, 19]:

$$f'(x)(z) = \sup\{\langle x^*, z\rangle : x^* \in \underline{\partial} f(x)\} + \inf\{\langle x^*, z\rangle : x^* \in \overline{\partial} f(x)\} \tag{1.31}$$

for any $z \in X$ with some pair of closed convex sets $\underline{\partial} f(x)$ and $\overline{\partial} f(x)$. This pair, although not uniquely defined, plays the role of a derivative for such functions.

Proposition 1.39. *If f admits the representation* (1.31), *then*

$$\partial f(x) - \overline{\partial} f(x) \subset \underline{\partial} f(x), \qquad \partial^+ f(x) - \underline{\partial} f(x) \subset \overline{\partial} f(x).$$

1.11.2. Weakly convex functions. A continuous function f defined on a finite-dimensional space X is called *weakly convex* [84, 85] if there exists a function $r : X \times X \to \mathbb{R}$ such that $r(x, u)/\|x - u\| \to 0$ as $u \to x$ uniformly relative to x in any closed bounded subset of X and the set

$$G(x) = \{x^* \in X : f(u) - f(x) - \langle x^*, u - x\rangle + r(x, u) \geq 0 \ \forall u \in X\}$$

is nonempty for any $x \in X$.

The class of weakly convex functions includes smooth and convex functions and functions of maximum type. Under the assumptions made, the set $G(x)$ is obviously closed, convex, and bounded. It was proved in [84, 85] that the multifunction $G(\cdot)$ is upper semicontinuous, f is locally Lipschitz, everywhere directionally differentiable, and quasidifferentiable:

$$f'(x)(z) = \max\{\langle x^*, z\rangle : x^* \in G(x)\} \quad \forall x, z \in X.$$

Proposition 1.40. *If f is weakly convex, then $\partial f(x) = G(x)$ for all $x \in X$.*

1.11.3. ε-Support functionals. An element $x^* \in X^*$ is called [26] an *ε-support functional* for f at x if there exists $\delta > 0$ such that

$$f(u) - f(x) \geq \langle x^*, u - x\rangle - \varepsilon\|u - x\| \quad \forall u \in B_\delta(x).$$

The set of all such elements is denoted by $S_\varepsilon f(x)$ and is called an *ε-support* for f at x. This set has properties very similar to those of the ε-subdifferential, but it may be nonclosed.

Proposition 1.41. *The following relation holds*:

$$\partial_\varepsilon f(x) = \bigcap_{\alpha > \varepsilon} S_\alpha f(x).$$

1.11.4. Screens and derivate containers. Let a set $\mathfrak{U} f(x) \subset X^*$ have the following property: for any $\varepsilon > 0$ and $\alpha > 0$, there exist $\delta \in (0, \alpha]$ and a continuously finitely differentiable function $g : X \to \mathbb{R}$ such that $|f(u) - g(u)| \leq \varepsilon\delta$ and $\nabla g(u) \in \mathfrak{U} f(x)$ for all $u \in B_\delta(x)$.

Recall that a function g is called *finitely differentiable* [33, 34, 82] at x (with a derivative $\nabla g(x)$) if for any finite-dimensional subspace $Z \subset X$, the function $z \to f(x + z) : Z \to \mathbb{R}$ is differentiable at 0 and its derivative coincides with the restriction of $\nabla g(x)$ to Z.

The set $\mathfrak{U}f(x)$ is a derivative-like object. It is not uniquely defined. If f is continuous and can be represented as a uniform limit of a sequence of continuously finitely differentiable functions f_i, $i = 1, 2, \dots$, then for any $\delta > 0$ and $j > 0$, one can take

$$\mathfrak{U}f(x) = \bigcup_{\substack{u \in B_\delta(x)\\ i \geq j}} \{\nabla f_i(u)\}.$$

Theorem 1.42. *Let $\mathfrak{U}f(x) \neq \emptyset$. Then $\partial f(x) \subset \operatorname{cl} \mathfrak{U}f(x)$.*

Proof. Let $x^* \notin \operatorname{cl} \mathfrak{U}f(x)$. Then there exists $\eta > 0$ such that

$$\|x^* - \mathfrak{u}\|_* > \eta \quad \forall \mathfrak{u} \in \mathfrak{U}f(x). \tag{1.32}$$

Let us denote $\varepsilon_0 = \eta/4$ and select a number δ_k and a function g_k in accordance with the definition of the set $\mathfrak{U}f(x)$ for $\varepsilon = \varepsilon_0/4$ and $\alpha = 1/k$.

Let us define, for some positive interger N_k, a finite set of points $x_i \in X$, $i = 0, 1, \dots, N_k$, satisfying the following conditions:

(a) $x_0 = x$, $x_{i+1} = x_i + hz_i$, $i = 0, 1, \dots, N_k - 1$;
(b) $\|z_i\| = 1$, $i = 0, 1, \dots, N_k - 1$;
(c) $h = \delta_k/(2N_k)$;
(d) $\langle x^* - \nabla g_k(x_i), z_i\rangle > \eta$, $i = 0, 1, \dots, N_k - 1$.

It is possible to find z_i satisfying (d), since, due to (a)–(c), one has

$$\|x_i - x\| \leq N_k h = \frac{\delta_k}{2}, \quad i = 1, 2, \dots, N_k, \tag{1.33}$$

g is finitely differentiable at x_i, $\nabla g(x_i) \in \mathfrak{U}f(x)$, and (1.32) holds.

The following estimate is valid for sufficiently large N_k:

$$\begin{aligned} g_k(x_{N_k}) - g_k(x) - \langle x^*, x_{N_k} - x\rangle &= \sum_{i=0}^{N_k-1} \left(\int_0^h \langle \nabla g_k(x_i + tz_i), z_i\rangle dt - h\langle x^*, z_i\rangle \right) \\ &\leq h \sum_{i=0}^{N_k-1} \langle \nabla g_k(x_i) - x^*, z_i\rangle + \frac{\eta\delta_k}{4}. \end{aligned}$$

Taking (d) and (c) into account, we have

$$g_k(x_{N_k}) - g_k(x) - \langle x^*, x_{N_k} - x\rangle < -\frac{\eta\delta_k}{4} = -\varepsilon_0\delta_k. \tag{1.34}$$

Recall that g_k is the approximation of the initial function f,

$$|f(u) - g_k(u)| \leq \frac{\varepsilon_0}{4}\delta_k \quad \forall u \in B_{\delta_k}(x). \tag{1.35}$$

It follows from (1.34), (1.35), and (1.33) that

$$f(x_{N_k}) - f(x) - \langle x^*, x_{N_k} - x\rangle < -\varepsilon_0\frac{\delta_k}{2} \leq -\varepsilon_0\|x_{N_k} - x\|. \tag{1.36}$$

Since $x_{N_k} \to x$ as $k \to \infty$, condition (1.36) means that $x^* \notin \partial f(x)$. □

Theorem 1.42 characterizes subdifferentials of functions that can be approximated by smooth functions near the point under consideration. It is easy to deduce from it the relation between the (Fréchet) subdifferential and the *screen* of Halkin [36, 37].

Let f be a function defined on an open set U in $\mathbb{R}^n$ and taking values in $\mathbb{R}^m$. A set $\mathfrak{U} \subset \mathbb{R}^{mn}$ is called a *screen* of f at $x \in U$ if for any $\varepsilon > 0$ and $\alpha > 0$, there exist $\delta \in (0, \alpha]$ and a continuously differentiable function $g : B^n_\delta(x) \to \mathbb{R}^m$ such that $B^n_\delta(x) \subset U$, $|f(u) - g(u)| \leq \varepsilon\delta$, and $\nabla g(u) \in \mathfrak{U} + \varepsilon B^{mn}$ for all $u \in B^n_\delta(x)$.

Let y^* be an arbitrary vector in $\mathbb{R}^m$. It is easy to see that if $\mathfrak{U}$ is a screen of f at x, then the set $y^*\mathfrak{U} = \{y^*\mathfrak{u} : \mathfrak{u} \in \mathfrak{U}\}$ satisfies all the properties of the above-introduced derivative-like object for the function $\langle y^*, f\rangle$.

Corollary 1.42.1. *If $\mathfrak{U}$ is a screen of f at x, then $\partial\langle y^*, f\rangle(x) \subset \mathrm{cl}(y^*\mathfrak{U})$ for any $y^* \in \mathbb{R}^m$.*

A screen of a function is defined not uniquely. As was noted in [36], the examples of screens are the *generalized gradient* [15] (and the *generalized Jacobian* [16]) of Clarke (see [17]) and the *derivate container* of Warga [97, 98].

In [99], Warga presented a modified definition of the *directional-derivate container* $\{\Lambda^\varepsilon f(x) : \varepsilon > 0\}$ for a function $f : \Omega \to Y$, where Ω is a convex compact set in X and Y is a Banach space. Application of Theorem 1.42 makes it possible to derive the following result.

Corollary 1.42.2. *If $\{\Lambda^\varepsilon f(x) : \varepsilon > 0\}$ is a directional-derivate container of f at x and $x \in \mathrm{int}\,\Omega$, then for any $y^* \in Y^*$, $\varepsilon > 0$, and $\eta > 0$, there exists $\delta > 0$ such that*

$$\partial\langle y^*, f\rangle(u) \subset \{A^*y^* : A \in \Lambda^\varepsilon f(x)\} + \eta B^*$$

for any $u \in B_\delta(x)$.

Remark 1.4. The assumption $x \in \mathrm{int}\,\Omega$ in the statement of Corollary 1.42.2 is essential. Let us consider a function $f : [0, 1] \to \mathbb{R} : f \equiv 0$. If we define f outside of $[0, 1]$ by setting $f(u) = \infty$, then obviously $\partial f(1) = [0, \infty)$. But at the same time, the singleton $\{0\}$ is a directional-derivate container of f at 1.

2. Variational Principles and Fuzzy Calculus

2.1. Variational principles. The *variational analysis* is based on some cornerstone results named variational principles (see [88, 93]). At present, several of them are known. The first and probably the most important one was certainly the *variational principle* of Ekeland.

Theorem 2.1 (Ekeland [25]). *Let $f : X \to \bar{\mathbb{R}}$ be a lower semicontinuous and bounded below function, $\varepsilon > 0$, and $\lambda > 0$. Suppose that*

$$f(v) < \inf f + \varepsilon.$$

Then there exists a point $x \in X$ such that

(a) $\|x - v\| < \lambda$;
(b) $f(x) \le f(v)$;
(c) *the function $u \to f(u) + (\varepsilon/\lambda)\|u - x\|$ attains a local minimum at x.*

This theorem was proved in [25] for a more general setting of an arbitrary complete metric space. Since then it has been widely used in variational analysis and proved to be a very powerful tool in investigating extremal problems. It makes it possible to substitute an "almost minimal" point (up to ε) by another point, arbitrarily close to the initial point, which is a local minimizer for a slightly perturbed (by adding a small norm-type term) function.

The only disadvantage of the conclusion of Theorem 2.1 for some applications is that the perturbation term in (c) is nonsmooth even if the norm in X is differentiable on $X\backslash\{0\}$. This disadvantage was eliminated in the *smooth variational principle* of Borwein and Preiss at the cost of narrowing the class of spaces where it is valid.

Let us say that X is *Fréchet smooth* if there exisits an equivalent norm in X which is Fréchet differentiable away from 0.

Theorem 2.2 (Borwein and Preiss [6]). *Let X be a Fréchet smooth space, $f : X \to \bar{\mathbb{R}}$ be a lower semicontinuous and bounded below function, $\varepsilon > 0$, and $\lambda > 0$. Suppose that*

$$f(v) < \inf f + \varepsilon.$$

Then there exist a convex C^1 function g on X and a point $x \in X$ such that

(a) $\|x - v\| < \lambda$;
(b) $f(x) \le f(v)$;
(c) *the function* $u \to f(u) + g(u)$ *attains a minimum at* x;
(d) $\|\nabla g(x)\|_* < \varepsilon/\lambda$.

Theorem 2.2 was proved in [6] for a more general setting of an arbitrary Banach space. But the differentiability of the perturbation term g and the estimate (d) is guaranteed only for Fréchet smooth spaces.

Both Theorems 2.1 and 2.2 can be considered as examples of "fuzzy" results. All other "fuzzy" results are based on Theorems 2.1 and 2.2 and their modifications. The traditional approach of variational analysis consists of applying some necessary optimality conditions to the "perturbed" function and formulating some ε-optimality conditions for the initial problem at a point close to the initial one.

Let us mention such important "fuzzy" results which follow from Theorem 2.2 as the *extremal principle* and the *fuzzy sum rule* (see below). It was proved in [5] that they are actually equivalent to Theorem 2.2 (under the assumption that the space is Fréchet smooth).

Other useful variational principles can be found in later publications [9, 20, 21, 30, 80, 88]. There exist strong relations between differential properties of the perturbation (or supporting) function in the corresponding variational principles and the differential properties of the norm (or the "bump" function) in the space under consideration (see [31]).

It is possible to obtain an estimate similar to (d) in Theorem 2.2 in more general spaces, this time at the cost of eliminating any mention of the perturbation function from the statement. The corresponding result is called the *subdifferential variational principle.*

Theorem 2.3 (Mordukhovich and Wang [80])**.** *Let X be an Asplund space, $f : X \to \bar{\mathbb{R}}$ be a lower semicontinuous and bounded below function, $\varepsilon > 0$, and $\lambda > 0$. Suppose that*

$$f(v) < \inf f + \varepsilon.$$

Then there exist points $x \in X$ and $x^* \in \partial f(x)$ such that

(a) $\|x - v\| < \lambda$;
(b) $f(x) \le f(v)$;
(c) $\|x^*\|_* < \varepsilon/\lambda$.

Recall that a Banach space is called *Asplund* [1] (see [88]) if any continuous convex function on it is Fréchet differentiable on a dense set of points. Asplund spaces form a rather broad subclass of Banach spaces (for various properties and characterizations of Asplund spaces, see [21, 88]). This class includes, for example, all spaces that admit Fréchet differentiable bump functions (in particular, Fréchet smooth spaces). Reflexive spaces are examples of Fréchet smooth spaces.

Asplund spaces have proved to be very convenient for investigating different properties of nonsmooth functions. Actually, the Asplund property of Banach spaces is not only a sufficient but also a necessary condition for the fulfillment of some basic results in nonsmooth analysis involving Fréchet normals and subdifferentials (see [31, 32, 73, 79] and the statements below).

If X is a Fréchet smooth space, the statement above follows immediately from Theorem 2.2 due to Corollary 1.12.3. It was actually contaned in the statement of the main result in [6].

It was also proved in [80] that the subdifferential variational principle is equivalent to the extremal principle and cannot be extended to non-Asplund spaces.

2.2. Fréchet subdifferentials in differentiability spaces. The equivalent representation of the Fréchet subdifferential given by Proposition 1.5 was presented in a general Banach space setting. Under an additional assumption on the space X, this statement can be strengthened.

Theorem 2.4. *Let X be a Fréchet smooth space. Then $x^* \in \partial f(x)$ if and only if there exists a function $g : X \to \mathbb{R}$ such that*

(a) $g(u) \le f(u)$ *for any* $u \in X$, *and* $g(x) = f(x)$;
(b) g *is continuously Fréchet differentiable on* X *and* $\nabla g(x) = x^*$;
(c) g *is concave.*

Theorem 2.4 (without condition (c)) was proved in [21]. The fact that g can be chosen concave was added in [12]. Stronger versions of Theorem 2.4 were obtained in [31], where the necessity of smooth renorms (bump functions) for the validity of variational principles was also proved.

Actually, Theorem 2.4 establishes the equivalence (for the Fréchet smooth case) between the Fréchet subdifferential and the *viscosity Fréchet subdifferential* (see [10, 12]).

Corollary 2.4.1. *Let* X *be a Fréchet smooth space. Then* $x^* \in N(x|\Omega)$ *if and only if there exists a function* $g : X \to \mathbb{R}$ *such that*

(a) $g(u) \le 0$ *for any* $u \in \Omega$ *and* $g(x) = 0$;
(b) g *is continuously Fréchet differentiable on* X *and* $\nabla g(x) = x^*$;
(c) g *is concave.*

Simple examples show that subdifferential (1.1) can be empty (see Examples 1.1, 1.2, and 1.4 above). Given an arbitrary lower semicontinuous function, it is important to know how large the set of points of subdifferentiability is. To make this set sufficiently rich, one must again impose additional assumptions on the space X.

A Banach space is called a *subdifferentiability space* [40] (for some kind of a subdifferential) if any lower semicontinuous function on it is subdifferentiable on a dense subset of its domain $\operatorname{dom} f = \{u \in X : f(u) < \infty\}$.

The following theorem states that for the Fréchet subdifferential, the class of subdifferentiability spaces coincides with Asplund spaces. It even says more: in a non-Asplund space, there exists a lower semicontinuous function which is nowhere Fréchet subdifferentiable.

Theorem 2.5. *The following assertions are equivalent*:

(a) X *is an Asplund space*;
(b) *for any lower semicontinuous function* $f : X \to \bar{\mathbb{R}}$, *the set* $\{u \in X : \partial f(u) \ne \emptyset\}$ *is dense in* $\operatorname{dom} f$;
(c) *for any lower semicontinuous function* $f : X \to \bar{\mathbb{R}}$, *there exists* $x \in \operatorname{dom} f$ *such that* $\partial f(x) \ne \emptyset$.

Some parts of Theorem 2.5 can be found in [26, 29, 31, 40]. It actually shows that Fréchet subdifferentials can be considered appropriate for Asplund spaces and that they are not very good in general Banach spaces.

Let us note that the implication (a)$\Rightarrow$(b) in Theorem 2.5 follows immediately from Theorem 2.3.

Condition (b) in Theorem 2.5 can be strengthened: the set $\{(u, f(u)) \in X \times \mathbb{R} : \partial f(u) \ne \emptyset\}$ is dense in $\operatorname{gph} f$.

Theorem 2.5 guarantees that for a lower semicontinuous function on an Asplund space, there exists a point of subdifferentiability in any neighborhood of a given point in its domain. Other fuzzy results in Asplund spaces can be found below in the rest of the section.

Using the notion of a strict δ-subdifferential, one can formulate the following corollary of Theorem 2.5.

Proposition 2.6. *The following assertions are equivalent*:

(a) X *is an Asplund space*;
(b) $\partial_\delta f(x) \ne \emptyset$ *for any lower semicontinuous at* x *function* $f : X \to \bar{\mathbb{R}}$ *and any* $\delta > 0$.

2.3. Sum rules. As was mentioned in the introduction, the direct calculus of Fréchet and other "simple" subdifferentials is rather poor because their definitions do not take into account "differential" properties of a function in a neighborhood of a given point. Nevertheless, there exists a way of developing the calculus for them either in the limiting (see [38,49,53,67,69,75]) or in the "fuzzy" form (see [12,27–29,39,40,42,55,104]).

The central point of any subdifferential calculus is certainly the *sum rule*, which allows one to express elements of a subdifferential of the sum of functions in terms of subdifferentials of initial functions.

After the sum rule was first established in the limiting form in [49] (see [53, 67]), most efforts were devoted to deriving fuzzy sum rules (see [10, 12, 22, 28, 29, 40, 44, 76]). Now two main versions of the fuzzy sum rule are known. For simplicity, they are formulated below in terms of strict δ-subdifferentials.

Rule 2.1 (weak fuzzy sum rule)**.** *Let $f_1, f_2, \dots, f_n : X \to \mathbb{R}$ be lower semicontinuous in a neighborhood of x. Then*

$$\partial\left(\sum_{i=1}^{n} f_i\right)(x) \subset \sum_{i=1}^{n} \hat{\partial}_\delta f_i(x) + U^*$$

for any $\delta > 0$ and any weak neighborhood U^* of 0 in X^*.*

Rule 2.2 (strong fuzzy sum rule)**.** *Let $f_1, f_2, \dots, f_n : X \to \mathbb{R}$ be lower semicontinuous in a neighborhood of x. Suppose that all f_i but at most one of them are Lipschitz in a neighborhood of x. Then*

$$\partial\left(\sum_{i=1}^{n} f_i\right)(x) \subset \sum_{i=1}^{n} \hat{\partial}_\delta f_i(x) + \delta B^*$$

for any $\delta > 0$.

Unfortunately, the above sum rules can fail in infinite dimensions. To improve the situation, one must again narrow the class of spaces. A Banach space is called a *trustworthy space* [40] (for some kind of subdifferential) if Rule 2.1 is valid in it.

The following theorem proved by Fabian [29] states that for the Fréchet subdifferential, the class of trustworthy spaces coincides with Asplund spaces.

Theorem 2.7. *The following assertions are equivalent*:

(a) *X is an Asplund space*;
(b) *weak fuzzy sum rule* 2.1 *is valid in X*;
(c) *strong fuzzy sum rule* 2.2 *is valid in X.*

Both Rules 2.1 and 2.2 are corollaries of the following *basic* (or *null*) *sum rule*, which is also valid only in Asplund spaces.

Rule 2.3 (basic fuzzy sum rule)**.** *Let $f_1, f_2, \dots, f_n : X \to \mathbb{R}$ be locally uniformly lower semicontinuous at x. Suppose that $\sum\limits_{i=1}^{n} f_i$ attains a local minimum at x. Then*

$$0 \in \sum_{i=1}^{n} \hat{\partial}_\delta f_i(x) + \delta B^*$$

for any $\delta > 0$.

Recall that lower semicontinuous in a neighborhood of x functions $f_1, f_2, \dots, f_n : X \to \mathbb{R}$ are called *locally uniformly lower semicontinuous* [10, 12] at x if

$$\inf_{u \in B_\delta(x)} \sum_{i=1}^{n} f_i(u) \le \lim_{\eta \to +0} \inf_{\substack{\|u_i - u_j\| \le \eta \\ u_i, u_j \in B_\delta(x) \\ i,j = 1,2,\dots,n}} \sum_{i=1}^{n} f_i(u_i)$$

for some $\delta > 0$.

The following proposition gives two important sufficient conditions for the local uniform lower semicontinuity property. It explains the way in which Rules 2.1 and 2.2 follow from Rule 2.3.

Proposition 2.8. *The functions $f_1, f_2, \dots, f_n : X \to \mathbb{R}$ are locally uniformly lower semicontinuous at x if one of the following conditions holds*:

(a) *all f_i except for at most one of them are uniformly continuous in a neighborhood of x*;
(b) *at least one of the f_i has compact level sets in a neighborhood of x.*

The main argument used when deducing Rules 2.1 and 2.2 from Rule 2.3 is the following: if $x^* \in \partial\left(\sum\limits_{i=1}^{n} f_i\right)(x)$, then for any $\varepsilon > 0$, the sum of $n+2$ functions $f_1, f_2, \dots, f_{n+2}$ attains a local minimum at x, where $f_{n+1}(u) = -\langle x^*, u\rangle$ and $f_{n+2}(u) = \varepsilon\|u - x\|$.

In the case of Rule 2.2, the functions $f_1, f_2, \dots, f_n$ except for at most one of them are uniformly continuous in a neighborhood of x; so are the functions $f_1, f_2, \dots, f_{n+2}$. Local uniform lower semicontinuity follows from Proposition 2.8(a).

In case of Rule 2.1, to make the functions locally uniformly lower semicontinuous at x, one must add one more function δ_L, an indicator function of some finite-dimensional subspace L of X containing x. Obviously, it has compact level sets in a neighborhood of x and makes the whole collection of functions locally uniformly lower semicontinuous due to Proposition 2.8(b). The presence of this last function in the system explains the necessity of considering a weak* neighborhood of 0 in the statement of Rule 2.1.

To prove Rule 2.3, one must first allow somehow each function in $\sum\limits_{i=1}^{n} f_i$ to have its own argument. In the Fréchet smooth case, this is usually done by considering the following sequence of penalized functions on X^n:

$$v_k(u_1, u_2, \dots, u_n) = \sum_{i=1}^{n} f_i(u_i) + k \sum_{i,j=1}^{n} \|u_i - u_j\|^2 + \|u_n - x\|^2.$$

The penalty terms are differentiable. Application of the smooth variational principle (Theorem 2.2) at the point $(x, x, \dots, x)$ gives all the desired estimates.

To prove Rule 2.3 in the Asplund case, the separable reduction technique is used (see [27, 29]).

The weak fuzzy sum rule yields the following representation of Fréchet normals to the intersection of closed sets.

Proposition 2.9. *Let X be an Asplund space and Ω_i, $i = 1, 2, \dots, n$, be closed sets in X. Then*

$$N\left(x, \bigcap_{i=1}^{n} \Omega_i\right) \subset \sum_{i=1}^{n} \hat{N}_\delta(x, \Omega_i) + U^*$$

for any $\delta > 0$ and any weak neighborhood U^* of* 0 *in X^*.*

All the sum rules formulated above could be called *local fuzzy sum rules*: they are related to some point x. There exists a nonlocal version of the sum rule [103]. It is not related to any point, and strict δ-subdifferentials are not appropriate for the formulation. The rule is formulated below in terms of "simple" subdifferentials.

Let us define a constant

$$\mu_0 = \lim_{\eta \to 0} \inf_{\operatorname{diam}(u_1, \dots, u_n) < \eta} \sum_{i=1}^{n} f_i(u_i),$$

some extended minimal value for the sum of functions (we use the denotation $\operatorname{diam}(\Omega) = \sup\{\|\omega_1 - \omega_2\| \colon \omega_1, \omega_2 \in \Omega\}$).

Rule 2.4 (nonlocal fuzzy sum rule). *Let $f_1, f_2, \dots, f_n : X \to \mathbb{R}$ be lower semicontinuous in a neighborhood of x. Suppose that $\mu_0 < \infty$. Then for any $\delta > 0$, there exist $x_i \in X$, $i = 1, 2, \dots, n$, such that*

$\operatorname{diam}(x_1, \ldots, x_n) < \delta$, $\sum\limits_{i=1}^{n} f_i(x_i) < \mu_0 + \delta$, *and*

$$0 \in \sum_{i=1}^{n} \partial f_i(x_i) + \delta B^*.$$

It was proved in [104] that the nonlocal fuzzy sum rule is equivalent to the local one (any one of them) and it is one more characterization of Asplund spaces.

Other fuzzy calculus results (chain rules, formulas for maximum-type functions, mean-value theorems, etc.) for functions and multifunctions can be deduced from (some form of) the sum rule (see [12, 42, 46, 55, 56, 70, 76, 103]).

2.4. Extremal principle. Now we state one more fuzzy result, the *extremal principle*. It continues the line of variational principles (see Sec. 2.1) and is in a sense equivalent to them and to the sum rules in Sec. 2.3.

Let Ω_1 and Ω_2 be closed sets in X.

Definition 2.1. A system of sets Ω_1 and Ω_2 is said to be extremal if $\Omega_1 \cap \Omega_2 \neq \emptyset$ and there exist sequences $\{a_{ik}\} \in X$, $i = 1, 2$, such that $a_{ik} \to 0$ as $k \to \infty$ and

$$(\Omega_1 - a_{1k}) \cap (\Omega_2 - a_{2k}) = \emptyset, \quad k = 1, 2, \ldots . \tag{2.1}$$

This definition means that (1) the two sets have nonempty intersection and (2) their intersection can be made empty by an arbitrarily small shift of the sets. Both sets are shifted in the definition above. It is easy to show that it can be reformulated equivalently with a single sequence and only one set being shifted.

Definition 2.2. A system of sets Ω_1 and Ω_2 is said to be locally extremal near $x \in \Omega_1 \cap \Omega_2$ if there exists a neighborhood U of x such that the system of sets $\Omega_1 \cap U$ and $\Omega_2 \cap U$ is extremal.

This is equivalent to replacing condition (2.1) in the original definition by the following condition:

$$(\Omega_1 - a_{1k}) \cap (\Omega_2 - a_{2k}) \cap U = \emptyset \quad k = 1, 2, \ldots .$$

The notion of an extremal set system was introduced in [62, 63] for the case of n sets (which can be easily reduced to the case of two sets). It characterizes the mutual arrangement of sets in space and represents a rather general notion of extremality: some (locally) extremal system corresponds to a (local) solution of any optimization problem (see various examples in [52, 62, 67] and the recent survey [71]). A simple example of an extremal system is provided by the pair $\{x\}, \Omega$, where x is a boundary point of Ω.

Following [57], we introduce a constant

$$\theta(\Omega_1, \Omega_2) = \sup\{r \geq 0 : B_r \subset \Omega_1 - \Omega_2\}, \tag{2.2}$$

describing the rate of "overlapping" of Ω_1 and Ω_2. The difference of sets in (2.2) is understood in the algebraic sense:

$$\Omega_1 - \Omega_2 = \{\omega_1 - \omega_2 : \omega_1 \in \Omega_1,\ \omega_2 \in \Omega_2\}.$$

If $\Omega_1 \cap \Omega_2 = \emptyset$, we suppose that $\theta(\Omega_1, \Omega_2) = -\infty$.

Using this constant makes the definition of an extremal system simpler.

Proposition 2.10. *A system of sets Ω_1 and Ω_2 is extremal if and only if $\theta(\Omega_1, \Omega_2) = 0$.*

For $\omega_1 \in \Omega_1$, $\omega_2 \in \Omega_2$, and $\rho > 0$, we define one more constant based on (2.2):

$$\tilde{\theta}_{\Omega_1,\Omega_2}(\omega_1, \omega_2, \rho) = \theta([\Omega_1 - \omega_1] \cap B_\rho, [\Omega_2 - \omega_2] \cap B_\rho). \tag{2.3}$$

It is a local constant related to some $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$. It is nondecreasing as a function of ρ. If $\omega_1 = \omega_2 = x$, one obviously has

$$\tilde{\theta}_{\Omega_1,\Omega_2}(x, x, \rho) = \theta(\Omega_1 \cap B_\rho(x), \Omega_2 \cap B_\rho(x)),$$

and we immediately come to the following equivalent definition of the locally extremal system.

Proposition 2.11. *A system of sets Ω_1 and Ω_2 is locally extremal near $x \in \Omega_1 \cap \Omega_2$ if and only if there exists $\rho > 0$ such that $\tilde{\theta}_{\Omega_1,\Omega_2}(x, x, \rho) = 0$.*

Constant (2.3) will be used in Sec. 3. It is important that the points ω_1 and ω_2 be allowed to be different.

Both Definitions 2.1 and 2.2 (and their equivalent representations given by Propositions 2.10 and 2.11) are primal-space conditions. Now let us consider some dual-space conditions expressed in terms of Fréchet normals.

Definition 2.3. Let us say that the generalized Euler equation holds at $x \in \Omega_1 \cap \Omega_2$ if for any $\delta > 0$, there exist elements $x_1^* \in \hat{N}_\delta(x|\Omega_1)$ and $x_2^* \in \hat{N}_\delta(x|\Omega_2)$ such that

$$\|x_1^* + x_2^*\|_* < \delta, \qquad \|x_1^*\|_* + \|x_2^*\|_* = 1.$$

Definition 2.3 describes a fuzzy form of the *separation* property. In finite dimensions, it can be equivalently reformulated in terms of limiting normals.

As the following *extremal principle* says, the generalized Euler equation is closely related to the notion of the extremal set system described above.

Extremal principle. *If a system of sets Ω_1 and Ω_2 is locally extremal near $x \in \Omega_1 \cap \Omega_2$, then the generalized Euler equation holds at x.*

The extremal principle was first proved in [62] (see also [52, 63, 67]) for the case of n sets in a Fréchet smooth space (and in terms of ε-normals). The following theorem proved in [73] says that it is valid in an arbitrary Asplund space and can be considered as an extremal characterization of Asplund spaces.

Theorem 2.12. *The following assertions are equivalent*:
(a) *X is an Asplund space*;
(b) *the extremal principle is valid in X.*

Due to Theorems 2.7 and 2.12, the extremal principle is equivalent to the sum rules. It is also equivalent to some other basic results of nonsmooth analysis (see [12, 104]).

The extremal principle can be considered as a certain generalization of the classical separation theorem for convex sets with no interiority-like assumptions. The generalized Euler equation characterizes extremal properties of set systems. It was used in [52, 62, 63, 67, 75] as a main tool for deducing calculus formulas and necessary optimality conditions.

The stronger version of the extremal principle (for extended extremal systems) will be proved in Sec. 3.

As was noted in [73], considering the extremal system provided by the pair $\{x\}, \Omega$, where x is a boundary point of a closed set Ω, makes it possible to deduce from Theorem 2.12 the following nonconvex generalization of the well-known *Bishop–Phelps theorem* (see [88]).

Corollary 2.12.1. *Let X be an Asplund space, Ω be closed, and $x \in \mathrm{bd}\,\Omega$. Then for any $\delta > 0$, there exists $x^* \in \hat{N}_\delta(x|\Omega)$ such that $\|x^*\|_* = 1$.*

3. Extended Extremality

This section is devoted to extending traditional primal-space extremality notions. The goal is to formulate the weakest possible conditions (definitions of extremality) for which known dual space necessary conditions expressed in a fuzzy or limiting form remain valid. In this way, we come to, in a sense, "fuzzy" primal-space conditions, and dual-space necessary conditions become also sufficient.

Several results of this kind are discussed below. We start with the definition of *covering* for a multifunction, which was from the very beginning defined in a fuzzy form (see [23]).

3.1. Covering (metric regularity). Let us consider a multifunction $F : X \rightrightarrows Y$ from X into another Banach space Y with a closed graph $\operatorname{gph} F$. Let $(x, y) \in \operatorname{gph} F$.

Definition 3.1. F covers near (x, y) if there exist $a > 0$ and neighborhoods U of x and V of y such that

$$B_{a\rho}(F(u) \cap V) \subset F(B_\rho(u))$$

for any $u \in U$, $\rho > 0$ with $B_\rho(u) \subset U$.

This property is sometimes referred to as *covering* or *openness at a linear rate*. It is equivalent to the following *metric* or *pseudoregularity* property.

Definition 3.2. F is metrically regular near (x, y) if there exist $c > 0$ and neighborhoods U of x and V of y such that

$$\operatorname{dist}(u, F^{-1}(v) \le c \operatorname{dist}(v, F(u))$$

for any $u \in U$ and $v \in V$.

Both covering and metric regularity are equivalent to the *pseudo-Lipschitzness* of the inverse mapping F^{-1}.

Definition 3.3. F is pseudo-Lipschitzian near (x, y) if there exist $l > 0$ and neighborhoods U of x and V of y such that

$$F(u_1) \cap V \subset F(u_2) + \|u_1 - u_2\|$$

for any $u_1, u_2 \in U$.

The three properties defined above play a very important role in nonsmooth analysis (see [23, 45, 64, 65, 68, 72, 77, 87]).

Theorem 3.1. *The following assertions are equivalent*:

(a) *F covers near (x, y)*;
(b) *F is metrically regular near (x, y)*;
(c) *F^{-1} is pseudo-Lipschitzian near (y, x).*

Let us introduce some constants describing the covering property:

$$\theta_F(x, y, \rho) = \sup\{r \ge 0 : B_r(y) \subset F(B_\rho(x))\}, \tag{3.1}$$

$$\hat{\theta}_F(x, y) = \liminf_{\substack{(u,v) \overset{\operatorname{gph} F}{\to} (x,y) \\ \rho \to +0}} \frac{\theta_F(u, v, \rho)}{\rho}. \tag{3.2}$$

Definition 3.1 can now be reformulated equivalently in the following way.

Proposition 3.2. *F covers near (x, y) if and only if $\hat{\theta}_F(x, y) > 0$.*

Let us note that the constant $\hat{\theta}_F(x, y)$ is defined by (3.2) "in a fuzzy way": it incorporates other constants calculated in nearby points.

The absence of the covering property, i.e., the case $\hat{\theta}_F(x, y) = 0$, corresponds to, in a sense, extremal (singular) behavior of F. Optimality in extremal problems can be treated as extremality (noncovering) for some multifunctions, and the covering theorem (Theorem 3.3 below) can serve as a tool for deducing optimality conditions. For example, the extremal principle (Sec. 2) can be deduced from Theorem 3.3.

In general, the above definition of covering transforms into definitions of "extended extremality" and the covering theorem leads to necessary and sufficient extremality conditions. This explains why this property was selected to start the current section.

The following constant defined in terms of dual-space elements is used for characterizing the covering property:

$$b_F(x, y) = \sup_{\delta > 0} \inf\{\|x^*\|_* : x^* \in \hat{\partial}_\delta F(x, y)(y^*),\ \|y^*\|_* = 1\}. \tag{3.3}$$

Theorem 3.3 (covering theorem)**.** *Let X and Y be Asplund spaces. F covers near (x, y) if and only if $b_F(x, y) > 0$.*

An analogue of Theorem 3.3 was proved in [54] for the case of a Fréchet smooth space Y and with no assumptions on X. It was formulated in terms of ε-coderivatives.

Using the result of Fabian [29] (see Theorem 2.7) instead of applying the Ekeland variational principle [25] (Theorem 2.1) makes it possible to derive the covering theorem as it is stated here. Below, we prove the following statement, which yields Theorem 3.3.

Proposition 3.4. $\hat{\theta}_F(x, y) \le b_F(x, y)$. *If $b_F(x, y) > 0$ and X and Y are Asplund spaces, then $\hat{\theta}_F(x, y) > 0$.*

Proof. For simplicity, we suppose that the maximum-type norm is used in $X \times Y$: $\|u, v\| = \max(\|u\|, \|v\|)$. Let $x^* \in \partial F(u, v)(y^*)$ and $\|y^*\|_* = 1$. Then, due to the definition of the coderivative, one has

$$\lim_{\rho \to +0} \sup_{\substack{(u',v') \in \operatorname{gph} F \\ \|(u',v')-(u,v)\| \le \rho}} \frac{\langle x^*, u' - u\rangle - \langle y^*, v' - v\rangle}{\|(u', v') - (u, v)\|} \le 0$$

and it follows from (3.1) that

$$\|x^*\|_* \ge \liminf_{\rho \to +0} \frac{\theta_F(u, v, \rho)}{\rho}.$$

Taking definitions (3.2) and (3.3) into account, we conclude that $b_F(x, y) \ge \hat{\theta}_F(x, y)$.

Now let $\hat{\theta}_F(x, y) = 0$ and

$$\hat{\theta}_F(x, y) = \lim_{k \to \infty} \theta_F(u_k, v_k, \rho_k)/\rho_k$$

for some sequences $(u_k, v_k) \overset{\operatorname{gph} F}{\to} (x, y)$ and $\rho_k \to +0$. We can suppose that $\theta_F(u_k, v_k, \rho_k) < \infty$. Then there exists $w_k \in Y$ such that

$$\theta_F(u_k, v_k, \rho_k) < \|w_k - v_k\| < \theta_F(u_k, v_k, \rho_k) + \rho_k^2$$

and $w_k \notin F(B_{\rho_k}(u_k))$. Thus, $\|v - w_k\| > 0$ for any $(u, v) \in \operatorname{gph} F$ such that $\|u - u_k\| \le \rho_k$ and it follows from the Ekeland variational principle (Theorem 2.1) that there exists $(x_k, y_k) \in \operatorname{gph} F$ such that $\|(x_k, y_k) - (u_k, v_k)\| \le \rho_k/(1 + \rho_k)$ and the function

$$(u, v) \to \|v - w_k\| + \rho_k^{-1}(1 + \rho_k)\|w_k - v_k\| \cdot \|(u, v) - (x_k, y_k)\|$$

attains at (x_k, y_k) a local minimum on $\operatorname{gph} F$.

Therefore, (x_k, y_k) is a local minimizer for the sum of three functions

$$f_1(u, v) = \|v - w_k\|, \quad f_2(u, v) = \rho_k^{-1}(1 + \rho_k)\|w_k - v_k\| \cdot \|(u, v) - (x_k, y_k)\|, \quad f_3(u, v) = \delta_{\operatorname{gph} F}(u, v)$$

on $X \times Y$. The first two functions are convex and Lipschitz and the third function is lower semicontinuous. One can apply strong fuzzy sum rule 2.2. There exist $(x_{ik}, y_{ik}) \in X \times Y$ and $(x^*_{ik}, y^*_{ik}) \in \partial f_i(x_{ik}, y_{ik})$, $i = 1, 2, 3$, such that

$$\|(x_{ik}, y_{ik}) - (x_k, y_k)\| \le \rho_k, \quad (x_{3k}, y_{3k}) \in \operatorname{gph} F,$$
$$\left\| \sum_{i=1}^{3} (x^*_{ik}, y^*_{ik}) \right\|_* \le \rho_k.$$

Obviously, $x^*_{1k} = 0$ and $\|(x^*_{2k}, y^*_{2k})\|_* \le \rho_k^{-1}(1 + \rho_k)\|w_k - v_k\|$. Without loss of generality, we can assume that $\|y_{1k} - y_k\| < \|w_k - y_k\|$. Hence $\|y_{1k} - w_k\| > 0$ and $\|y^*_{1k}\|_* = 1$. Consequently,

$$\|x^*_{3k}\|_* \le \rho_k^{-1}(1 + \rho_k)\|w_k - v_k\| + \rho_k < (1 + \rho_k)(\theta_F(u_k, v_k, \rho_k)/\rho_k + \rho_k) + \rho_k,$$
$$\|y^*_{3k}\|_* \ge 1 - \rho_k^{-1}(1 + \rho_k)\|w_k - v_k\| - \rho_k > 1 - (1 + \rho_k)(\theta_F(u_k, v_k, \rho_k)/\rho_k + \rho_k) - \rho_k > 0$$

if k is sufficiently large.

Let us denote $x_k^* = x_{3k}^*/\|y_{3k}^*\|_*$ and $y_k^* = -y_{3k}^*/\|y_{3k}^*\|_*$. Then $\|y_k^*\|_* = 1$, $x_k^* \in \partial F(u_{3k}, v_{3k})(y_k^*)$, and $\|x_k^*\|_* \to 0$ as $k \to \infty$. This means that $b_F(x, y) = 0$. $\square$

In the case of an ordinary (single-valued) continuous mapping $f : X \to Y$, the following constants are used in the definition of the covering property instead of (3.1) and (3.2):

$$\theta_f(x, \rho) = \sup\{r \geq 0 : B_r(f(x)) \subset f(B_\rho(x))\},$$

$$\hat{\theta}_f(x) = \liminf_{\substack{u \to x \\ \rho \to +0}} \frac{\theta_f(u, \rho)}{\rho},$$

and under the additional Lipschitzness assumption in the dual covering criterion instead of (3.3), one can use the constant

$$b_f(x) = \sup_{\delta > 0} \inf\{\|x^*\|_* : x^* \in \hat{\partial}_\delta^- \langle y^*, f\rangle(x),\ \|y^*\|_* = 1\}.$$

Theorem 3.5. *Let X and Y be Asplund spaces and f be Lipschitz continuous near x. f covers near x if and only if $b_f(x) > 0$.*

Various fuzzy criteria and constants for covering (metric regularity) properties were obtained in [72] in Asplund (sometimes in Banach) spaces.

3.2. Extended extremal principle. The extremal principle (Sec. 2) says that the generalized Euler equation is a necessary condition for (local) extremality of a set system. At the same time, the generalized Euler equation can hold for sets not necessarily satisfying Definition 2.2. There exists a way of extending the definition in such a way that the necessary condition remains valid (it even becomes sufficient). It is more convenient to do so starting from the equivalent definition of extremality given by Proposition 2.11 (see [58, 60, 61]).

Let us introduce one more local constant based on (2.3) for closed sets Ω_1 and Ω_2 in X and $x \in \Omega_1 \cap \Omega_2$:

$$\hat{\theta}_{\Omega_1,\Omega_2}(x) = \liminf_{\substack{\omega_1 \xrightarrow{\Omega_1} x \\ \omega_2 \xrightarrow{\Omega_2} x \\ \rho \to +0}} \frac{\tilde{\theta}_{\Omega_1,\Omega_2}(\omega_1, \omega_2, \rho)}{\rho}. \tag{3.4}$$

Definition (3.4) is very similar to (3.2) and, as is shown below, the notion of extended extremality of sets corresponds exactly to the notion of covering of multifunctions.

Definition 3.4. A system of sets Ω_1 and Ω_2 is extended extremal (e-extremal) near $x \in \Omega_1 \cap \Omega_2$ if $\hat{\theta}_{\Omega_1,\Omega_2}(x) = 0$.

The condition $\hat{\theta}_{\Omega_1,\Omega_2}(x) = 0$ is weaker than the condition used in Proposition 2.11 for characterizing local extremality: if $\tilde{\theta}_{\Omega_1,\Omega_2}(x, x, \rho) = 0$ for some $\rho > 0$, then, of course, $\hat{\theta}_{\Omega_1,\Omega_2}(x) = 0$.

Proposition 3.6. *If a system of sets Ω_1, Ω_2 is locally extremal near x, then it is e-extremal near x.*

Contrary to the condition in Proposition 2.11, Definition 3.4 does not impose strict "nonoverlapping" of sets but up to an arbitrarily small deformation. Second, the sets do not need to "nonoverlap" in x: it is sufficient that in any of its neighborhoods, there exist points $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$ such that the sets $\Omega_1 - \omega_1$ and $\Omega_2 - \omega_2$ "almost nonoverlap."

Definition 3.4 leads to the extended extremal principle.

Extended extremal principle. *A system of sets Ω_1 and Ω_2 is e-extremal near $x \in \Omega_1 \cap \Omega_2$ if and only if the generalized Euler equation holds at x.*

The next theorem extends Theorem 2.12.

Theorem 3.7. *The following assertions are equivalent*:

(a) *X is an Asplund space*;
(b) *the extremal principle is valid in X*;
(c) *the extended extremal principle is valid in X.*

Due to Proposition 3.6, the only implication which needs to be proved is (a) $\Rightarrow$ (c). The direct proof of this statement can be found in [61]. Below, this implication is deduced from the covering theorem (Theorem 3.3).

Proof of (a)$\Rightarrow$(c). Let us consider a function $F : X \times X \to X$ defined by the relation

$$F(\omega_1, \omega_2) = \begin{cases} \omega_1 - \omega_2 & \text{if } \omega_1 \in \Omega_1,\ \omega_2 \in \Omega_2, \\ \emptyset & \text{otherwise.} \end{cases}$$

It is continuous on $\Omega_1 \times \Omega_2$ and its graph $\operatorname{gph} F$ is a closed set in $X \times X \times X$. If $\omega_1 \in \Omega_1$, $\omega_2 \in \Omega_2$, and $\rho > 0$, then

$$\begin{aligned} \theta_F(\omega_1, \omega_2, \omega_1 - \omega_2, \rho) &= \sup\{r \geq 0 : B_r(\omega_1 - \omega_2) \subset \Omega_1 \cap B_\rho(\omega_1) - \Omega_2 \cap B_\rho(\omega_2)\} \\ &= \sup\{r \geq 0 : B_r \subset [\Omega_1 - \omega_1] \cap B_\rho - [\Omega_2 - \omega_2] \cap B_r\} = \tilde{\theta}_{\Omega_1,\Omega_2}(\omega_1, \omega_2, \rho). \end{aligned}$$

Hence, $\hat{\theta}_F(x, x, 0) = \hat{\theta}_{\Omega_1,\Omega_2}(x)$ and the extended extremality of Ω_1, Ω_2 near x is equivalent to the absence of the covering property for F near $(x, x, 0)$.

Let Ω_1 and Ω_2 be extended extremal near x. Due to Theorem 3.3, this yields $b_F(x, x, 0) = 0$, i.e., for any $\delta > 0$, there exist $\omega_1 \in \Omega_1$, $\omega_2 \in \Omega_2$, $v_1^*, v_2^*, y^* \in X^*$ such that

$$\|\omega_1 - x\| \leq \delta, \quad \|\omega_2 - x\| \leq \delta, \quad \|y^*\|_* = 1, \quad \|(v_1^*, v_2^*)\|_* \leq \delta,$$
$$(v_1^*, v_2^*) \in \partial F(\omega_1, \omega_2, \omega_1 - \omega_2)(y^*).$$

The last condition yields two inclusions:

$$v_1^* - y^* \in N(\omega_1|\Omega_1), \quad v_2^* + y^* \in N(\omega_2|\Omega_2).$$

Without loss of generality, we can suppose that $\delta \leq 1/2$. Then

$$\alpha = \|v_1^* - y^*\|_* + \|v_2^* + y^*\|_* \geq 1.$$

Let us denote

$$x_1^* = (v_1^* - y^*)/\alpha, \quad x_2^* = (v_2^* + y^*)/\alpha.$$

Then one has

$$x_1^* \in N(\omega_1|\Omega_1), \quad x_2^* \in N(\omega_2|\Omega_2),$$
$$\|x_1^* + x_2^*\|_* \leq \delta, \quad \|x_1^*\|_* + \|x_2^*\|_* = 1,$$

i.e., the generalized Euler equation is true.

Conversely, let the generalized Euler equation hold at x: for any $\delta > 0$, there exist $x_1 \in \Omega_1$, $x_2 \in \Omega_2$, $x_1^* \in N(x_1|\Omega_1)$, and $x_2^* \in N(x_2|\Omega_2)$ such that

$$\|x_1 - x\| \leq \delta, \quad \|x_2 - x\| \leq \delta, \quad \|x_1^* + x_2^*\|_* \leq \delta, \quad \|x_1^*\|_* + \|x_2^*\|_* = 2.$$

Then the norm of one of the elements x_1^* and x_2^*, say x_1^*, is not less than 1. Let us denote $y^* = x_1^*/\|x_1^*\|_*$ and $v^* = x_2^*/\|x_1^*\|_*$. Then

$$\|y^*\|_* = 1, \quad \|y^* + v^*\|_* \leq \delta, \quad y^* \in N(x_1|\Omega_1), \quad y^* \in N_\delta(-x_2| - \Omega_2).$$

Hence

$$\limsup_{\substack{\omega_1 \xrightarrow{\Omega_1} x_1 \\ \omega_2 \xrightarrow{\Omega_2} x_2}} \frac{\langle y^*, (\omega_1 - \omega_2) - (x_1 - x_2)\rangle}{\|(\omega_1, \omega_2) - (x_1, x_2)\|} \leq \delta \tag{3.5}$$

and F does not cover near $(x, x, 0)$, since otherwise the upper limit in the left-hand side of (3.5) is greater than some fixed positive a if δ is sufficiently small. The system Ω_1 and Ω_2 is extended extremal near x and, consequently, the extended extremal principle is valid in X. $\square$

Let us note that the sufficient part of the extended extremal principle was proved above without using the asplundity assumption. It is valid in an arbitrary Banach space.

Similarly to the initial definition of an extremal system, the notion of an extended extremal system can be expanded for the case of n sets.

Definition 3.5. A system of n closed sets Ω_i, $i = 1, 2, \ldots, n$, is e-extremal near x if the system of two sets $\tilde{\Omega}_1 = \prod_{i=1}^{n} \Omega_i$ and $\tilde{\Omega}_2 = \{(\omega, \omega, \ldots, \omega) \in X^n\}$ is e-extremal near $(x, x, \ldots, x) \in X^n$.

Proposition 3.8. *Let $I = \{1, 2, \ldots, n\}$, $j \in I$. A system of sets Ω_i, $i \in I$, is e-extremal near x if the system of sets $\tilde{\Omega}_1 = \prod_{i \in I \setminus \{j\}} \Omega_i$ and $\tilde{\Omega}_2 = \{(\omega, \omega, \ldots, \omega) \in X^{n-1} : \omega \in \Omega_j\}$ is e-extremal near $(x, x, \ldots, x) \in X^{n-1}$.*

The following theorem gives a dual criterion of e-extremality (the generalized Euler equation).

Theorem 3.9. *Let X be an Asplund space. A system of sets Ω_i, $i = 1, 2, \ldots, n$, is e-extremal near x if and only if for any $\delta > 0$, there exist elements $x_i^* \in \hat{N}_\delta(x|\Omega_i)$, $i = 1, 2, \ldots, n$, such that*

$$\|x_1^* + x_2^* + \cdots + x_n^*\|_* < \delta,$$
$$\|x_1^*\|_* + \|x_2^*\|_* + \cdots + \|x_n^*\|_* = 1.$$

The proof of Theorem 3.9 reduces to calculating the cones $N(\tilde{x}|\tilde{\Omega}_1)$ and $N(\tilde{x}|\tilde{\Omega}_2)$.

3.3. Extended (f, Ω, M)-extremality. One more abstract scheme of deducing (extended) extremality conditions is developed below. It is in a sense a counterpart of the extended extremal principle and is equivalent to it.

Let Ω and M be closed sets in Banach spaces X and Y respectively and f be a function from Ω into Y. Let $x \in \Omega$ and $f(x) \in M$.

We suppose that f is M-closed, i.e., the graph $\operatorname{gph} F$ of the multifunction

$$F(u) = f(u) - M, \quad u \in \Omega, \tag{3.6}$$

is closed in $X \times Y$.

If $M = \{0\}$, then the last condition means that f is continuous (on Ω). If $Y = \mathbb{R}$ and $M = \mathbb{R}_-$ ($M = \mathbb{R}_+$) then f is lower (upper) semicontinuous.

Ω, M, and f are treated here as elements of an abstract extremal problem characterized by multifunction (3.6), (the absence of) the covering property of the latter multifunction playing the crucial role in analysis of the problem.

Constants (3.1) and (3.2) in the case of (3.6) take the following form:

$$\theta_{\Omega,M,f}(x, y, \rho) = \theta(f(\Omega \cap B_\rho(u)) - y, M), \tag{3.7}$$

$$\hat{\theta}_{\Omega,M,f}(x) = \liminf_{\substack{(u,v) \xrightarrow{\operatorname{gph} F} (x,0) \\ \rho \to +0}} \frac{\theta_F(u, v, \rho)}{\rho}. \tag{3.8}$$

It is supposed in (3.7) that $y \in F(x)$ and (3.8) corresponds to the case $y = 0$. Surely, $0 \in F(x)$ due to the assumption that $f(x) \in M$.

Definition 3.6. x is extended (f, Ω, M)-*extremal* if $\hat{\theta}_{\Omega,M,f}(x) = 0$.

The above definition is a modification of the corresponding definitions in [52, 54, 61]. Contrary to the known abstract notions of extremality, there are no assumptions on the set M in Definition 3.6: it does not need to be convex and/or to have nonempty interior.

Definition 3.6 gives a rather general notion of (extended) extremality, embracing different optimality notions in optimization problems. For example, if x is a local solution of the nonlinear programming problem

$$\begin{array}{ll} \text{minimize} & f_0(u) \\ \text{subject to} & f_i(u) \le 0,\ i = 1, 2, \dots, m, \\ & f_i(u) = 0,\ i = m+1, \dots, n, \\ & u \in \Omega, \end{array}$$

where Ω is a closed set, m and n are nonnegative integers, $m \le n$, and functions f_i are lower semicontinuous for $i = 0, 1, \dots, m$ and continuous for $i = m+1, \dots, n$, then x is extended (f, Ω, M)-extremal if one takes

$$f = (f_0 - f_0(x), f_1, f_2, \dots, f_n) : X \to R^{n+1}, \quad M = R_-^{m+1} \times 0.$$

Application of the covering theorem (Theorem 3.3) to multifunction (3.6) leads to the following statement.

Theorem 3.10. *Let X and Y be Asplund spaces. x is extended (f, Ω, M)-extremal if and only if for any $\delta > 0$, there exists an element $y^* \in Y^*$ such that $\|y^*\|_* = 1$ and*

$$0 \in \hat{\partial}_\delta F(x, 0)(y^*). \tag{3.9}$$

The above theorem gives general extremality conditions. The element y^* can be considered as an analogue of the Lagrange multiplier vector in the classical problem of nonlinear programming. Inclusion (3.9) yields the inclusion

$$y^* \in \hat{N}_\delta(f(x)|M) \tag{3.10}$$

generalizing conditions on signs of multipliers and complementarity slackness conditions.

The following statement is a corollary of Theorem 3.10 under the additional Lipschitzness assumption.

Theorem 3.11. *Let X and Y be Asplund spaces and f be Lipschitz continuous near x. x is extended (f, Ω, M)-extremal if and only if for any $\delta > 0$, there exists an element $y^* \in Y^*$ such that $\|y^*\|_* = 1$,*

$$0 \in \hat{\partial}_\delta^- \langle y^*, f \rangle (x), \tag{3.11}$$

and inclusion (3.10) *holds.*

Inclusion (3.11) generalizes the classical Lagrange multipliers rule.

Some examples of necessary optimality conditions derived from Theorem 3.11 can be found in [57].

3.4. Extended minimality. This section is devoted to an extended notion of minimality of a real-valued function introduced in [61].

Let $\varphi : X \to \bar{\mathbb{R}}$ be lower semicontinuous and $\varphi(x) < \infty$. Let us denote

$$\theta_\varphi(x, \rho) = \inf_{u \in B_\rho(x)} \varphi(u) - \varphi(x),$$

$$\hat{\theta}_\varphi(x) = \limsup_{\substack{u \xrightarrow{\varphi} x \\ \rho \to +0}} \frac{\theta_\varphi(u, \rho)}{\rho}.$$

Surely, both constants are nonpositive.

Definition 3.7. x is a point of extended minimum (e-minimum) of φ if $\hat{\theta}_\varphi(x) = 0$.

This is a particular case of Definition 3.6: one can take $f(u) = \varphi(u) - \varphi(x)$, $\Omega = \operatorname{dom} \varphi$, and $M = R_-$.

The following statement is an easy consequence of the definition: it corresponds to the case $\theta_\varphi(x, \rho) = 0$ for some $\rho > 0$.

Proposition 3.12. *If x is a point of local minimum of φ, then it is a point of e-minimum of φ.*

Application of Theorem 3.10 leads to the following result.

Proposition 3.13. *Let X be an Asplund space. x is a point of e-minimum for φ if and only if $0 \in \hat{\partial}_{\delta}^{-}\varphi(x)$ for any $\delta > 0$.*

In the smooth case, a set of e-minimal points coincides with a set of stationary points (in Asplund spaces, this statement follows from Proposition 3.13).

Proposition 3.14. *Let φ be strictly differentiable at x. x is a point of e-minimum for φ if and only if $\nabla\varphi(x) = 0$.*

In general, the notion of extended minimality is closely related to some extended notion of stationarity introduced by Kummer [65].

Definition 3.8. x is a stationary point of φ (with respect to minimization) if it is a limit of local ε-Ekeland points x_ε of φ for $\varepsilon \to +0$ ($\varphi(u) + \varepsilon\|u - x_\varepsilon\| \geq \varphi(x_\varepsilon)$ for all u in some neighborhood of x_ε) with $\varphi(x_\varepsilon) \to \varphi(x)$.

Following the approach of the current paper, this definition can be rewritten equivalently by using constants

$$\tau_\varphi(x,\rho) = \inf_{u \in B_\rho(x)\setminus\{x\}} \min\left(\frac{\varphi(u) - \varphi(x)}{\|u - x\|}, 0\right), \tag{3.12}$$

$$\hat{\tau}_\varphi(x) = \limsup_{\substack{u \xrightarrow{\varphi} x \\ \rho \to +0}} \tau_\varphi(u,\rho). \tag{3.13}$$

Proposition 3.15. *x is a stationary point of φ (with respect to minimization) if and only if $\hat{\tau}_\varphi(x) = 0$.*

The following theorem says that both Definitions 3.7 and 3.8 are actually equivalent.

Theorem 3.16. *x is a point of extended minimum of φ if and only if x is a stationary point of φ (with respect to minimization).*

Proof. The sufficient part is obvious since $\tau_\varphi(x,\rho) \leq \theta_\varphi(x,\rho)/\rho \leq 0$ for any $\rho > 0$. The necessity was proved by Kummer[1] by using the Ekeland variational principle. If x is a point of extended minimum of φ, then $\hat{\theta}_\varphi(x) = 0$ and, consequently, there exist sequences $\{u_k\} \subset X$ and $\{\rho_k\} \subset \mathbb{R}_+$ such that $u_k \xrightarrow{\varphi} x$, $\rho_k \to 0$, and $\varphi(u) - \varphi(u_k) \geq -\rho_k^2$ for all $u \in B_{2\rho_k}(u_k)$. Then it follows from Theorem 2.1 that there exists $x_k \in B_{\rho_k}(u_k)$ such that $\varphi(x_k) \leq \varphi(u_k)$ and $\varphi(u) - \varphi(x_k) \geq -\rho_k\|u - x_k\|$ for all u near x_k. Surely, $x_k \xrightarrow{\varphi} x$ and $\tau_\varphi(x_k,\rho) \geq -\rho_k$ if ρ is sufficiently small, and $\hat{\tau}_\varphi(x) = 0$. $\square$

It is worth noting one more relation of the extended minimality notion. If one takes the limit in (3.12) as $\rho \to +0$, then one more constant appears:

$$\tilde{\tau}_\varphi(x) = \liminf_{u \to x} \min\left(\frac{\varphi(u) - \varphi(x)}{\|u - x\|}, 0\right).$$

It coincides up to the sign with the *slope* $|\nabla\varphi|(x)$ of φ at x, which was used in [45] for characterizing metric regularity properties of multifunctions.

Taking (3.13) into account, one has

$$\hat{\tau}_\varphi(x) = \limsup_{u \xrightarrow{\varphi} x} \tilde{\tau}_\varphi(u) = -\liminf_{u \xrightarrow{\varphi} x} |\nabla\varphi|(u). \tag{3.14}$$

The lower limit in the right-hand side of (3.14) could be called a *strict slope* of φ at x. It is this constant which (without being defined explicitly) actually works in [45].

[1]Private communication.

The last statement of the section shows that the definition of extended minimality is stable relative to small deformations of the data.

Proposition 3.17. *Let ψ be strictly differentiable at x with $\nabla\psi(x) = 0$. If x is a point of e-minimum for φ, then it is a point of e-minimum for $\varphi + \psi$.*

Let us take, for example, a problem of unconditional minimization of a real-valued function $\varphi(u) = u^2$ defined on the real line. $u = 0$ is obviously a point of minimum. If we add to φ an indefinitely small (in a neighborhood of $u = 0$) function $\psi(u) = -|u|^{3/2}$, then $u = 0$ is no longer a point of minimum (actually, it is a point of local maximum of $\varphi + \psi$). When using the extended definition of minimality, $u = 0$ remains a point of minimum for $\varphi + \psi$ since it is a point of minimum for φ and $\nabla\psi(0) = 0$.

Some other examples of extended minimal points can be found in [61].

Acknowledgments. This research was supported by the INTAS (grant No. 97-1050). The author would like to express his gratitude to B. Mordukhovich and the referee for many helpful remarks, especially regarding the references.

REFERENCES

1. E. Asplund, "Fréchet differentiability of convex functions," *Acta Math.,* **121**, 31–47 (1968).
2. J.-P. Aubin and I. Ekeland, *Applied Nonlinear Analysis*, Wiley, New York (1984).
3. M. S. Bazaraa, J. J. Goode, and Z. Z. Nashed, "On the cones of tangents with applications to mathematical programming," *J. Optimization Theory Appl.,* **13**, 389–426 (1974).
4. J. M. Borwein, "Weak tangent cones and optimization in a Banach space," *SIAM J. Control Optimization,* **16**, 512–522 (1978).
5. J. M. Borwein, B. S. Mordukhovich, and Y. Shao, "On the equivalence of some basic principles in variational analysis," *J. Math. Anal. Appl.,* **229**, 228–257 (1999).
6. J. M. Borwein and D. Preiss, "A smooth variational principle with applications to subdifferentiability and differentiability of convex functions," *Trans. Amer. Math. Soc.,* **303**, 517–527 (1987).
7. J. M. Borwein and H. M. Strojwas, "Tangential approximations," *Nonlinear Anal., Theory Methods Appl.,* **9**, 1347–1366 (1985).
8. J. M. Borwein and H. M. Strojwas, "Proximal analysis and boundaries of closed sets in Banach space, I. Theory," *Can. J. Math.,* **38**, 431–452 (1986).
9. J. M. Borwein, J. S. Treiman, and Q. J. Zhu, "Partially smooth variational principles and applications," *Nonlinear Anal., Theory Methods Appl.,* **35**, 1031–1059 (1999).
10. J. M. Borwein and Q. J. Zhu, "Viscosity solutions and viscosity subderivatives in smooth Banach spaces with applications to metric regularity," *SIAM J. Control Optimization,* **34**, 1568–1591 (1996).
11. J. M. Borwein and Q. J. Zhu, "Limiting convex examples for nonconvex subdifferential calculus," *J. Convex Anal.,* **5**, 221–235 (1998).
12. J. M. Borwein and Q. J. Zhu, "A survey of subdifferential calculus with applications," *Nonlinear Anal.,* **38**, 687–773 (1999).
13. G. Bouligand, *Introduction à la Géométrie Infinitésimale Directe*, Gauthier-Villars, Paris (1932).
14. N. Bourbaki, *Variétés Différentielles et Analytiques*, Hermann, Paris (1967).
15. F. H. Clarke, "Generalized gradients and applications," *Trans. Amer. Math. Soc.,* **204**, 247–262 (1975).
16. F. H. Clarke, "On the inverse function theorem," *Pac. J. Math.,* **64**, 97–102 (1976).
17. F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York (1983).
18. V. F. Demianov and A. M. Rubinov, "On quasidifferentiable functionals," *Sov. Math. Dokl.,* **21**, 14–17 (1980).
19. V. F. Demianov and L. V. Vasiliev, *Nonsmooth Optimization* [in Russian], Nauka, Moscow (1981).
20. R. Deville, G. Godefroy, and V. Zizler, "A smooth variational principle with applications to Hamilton–Jacobi equations in infinite dimensions," *J. Funct. Anal.,* **111**, 197–212 (1993).

21. R. Deville, G. Godefroy, and V. Zizler, *Smoothness and Renorming in Banach Spaces*, Pitman Monogr. Surv. Pure Appl. Math., Vol. 64, Longman, Harlow, UK (1993).
22. R. Deville and E. M. E. Haddad, "The subdifferential of the sum of two functions in Banach spaces, I, First-order case," *J. Convex Anal.,* **3**, 295–308 (1996).
23. A. V. Dmitruk, A. A. Milyutin, and N. P. Osmolovsky, "Lyusternik's theorem and the theory of extrema," *Russ. Math. Surv.,* **35**, 11–51 (1980).
24. A. Y. Dubovitskii and A. A. Milyutin, "Extremum problems in the presence of restrictions," *Zh. Vychisl. Mat. Mat. Fiz.,* **5**, 1–80 (1965).
25. I. Ekeland, "On the variational principle," *J. Math. Anal. Appl.,* **47**, 324–353 (1974).
26. I. Ekeland and G. Lebourg, "Generic Fréchet differentiability and perturbed optimization problems in Banach spaces," *Trans. Amer. Math. Soc.,* **224**, 193–216 (1976).
27. M. Fabian, "Subdifferentials, local ε-supports, and Asplund spaces," *J. London Math. Soc.,* **34**, 568–576 (1986).
28. M. Fabian, "On classes of subdifferentiability spaces of Ioffe," *Nonlinear Anal.,* **12**, 63–74 (1988).
29. M. Fabian, "Subdifferentiability and trustworthiness in the light of a new variational principle of Borwein and Preiss," *Acta Univ. Carolinae,* **30**, 51–56 (1989).
30. M. Fabian, P. Hájek, and J. Vanderwerff, "On smooth variational principles in Banach spaces," *J. Math. Anal. Appl.,* **197**, 153–172 (1996).
31. M. Fabian and B. S. Mordukhovich, "Nonsmooth characterizations of Asplund spaces and smooth variational principles," *Set-Valued Anal.,* **6**, 381–406 (1998).
32. M. Fabian and B. S. Mordukhovich, "Separable reduction and supporting properties of Fréchet-like normals in Banach spaces," *Can. J. Math.,* **51**, 26–48 (1999).
33. R. V. Gamkrelidze, "First-order necessary conditions and axiomatics of extremal problems," *Tr. MIAN,* **112**, 152–180 (1971).
34. R. V. Gamkrelidze and G. L. Kharatishvili, "Extremal problems in linear topological spaces," *Izv. Akad. Nauk SSSR, Ser. Mat.,* **33**, 781–839 (1969).
35. F. J. Gould and J. W. Tolle, "Optimality conditions and constraint qualifications in Banach space," *J. Optimization Theory Appl.,* **15**, 667–684 (1975).
36. H. Halkin, "Interior mapping theorem with set-valued derivatives," *J. Anal. Math.,* **30**, 200–207 (1976).
37. H. Halkin, "Mathematical programming without differentiability," In: *Calculus of Variations and Control Theory* (D. L. Russell, Ed.), Academic Press, New York (1976), pp. 279–287.
38. A. D. Ioffe, "Nonsmooth analysis: Differential calculus of nondifferentiable mappings," *Trans. Amer. Math. Soc.,* **266**, 1–56 (1981).
39. A. D. Ioffe, "Sous-différentielles approchees de fonctions numériques," *C. R. Acad. Sci. Paris, Ser. 1,* **292**, 675–678 (1981).
40. A. D. Ioffe, "On subdifferentiability spaces," *Ann. New York Acad. Sci.,* **410**, 107–119 (1983).
41. A. D. Ioffe, "Approximate subdifferentials and applications. I. The finite-dimensional theory," *Trans. Amer. Math. Soc.,* **281**, 389–416 (1984).
42. A. D. Ioffe, "Calculus of Dini subdifferentials of functions and contingent coderivatives of set-valued maps," *Nonlinear Anal., Theory Methods Appl.,* **8**, 517–539 (1984).
43. A. D. Ioffe, "Approximate subdifferentials and applications. II," *Mathematika,* **33**, 111–128 (1986).
44. A. D. Ioffe, "Proximal analysis and approximate subdifferentials," *J. London Math. Soc.,* **41**, 175–192 (1990).
45. A. D. Ioffe, "Metric regularity and subdifferential calculus," *Russ. Math. Surv.,* **55**, 501–558 (2000).
46. A. D. Ioffe and J.-P. Penot, "Subdifferentials of performance functions and calculus of coderivatives of set-valued mappings," *Serdica Math. J.,* **22**, 257–282 (1996).
47. A. D. Ioffe and V. M. Tikhomirov, *Theory of Extremal Problems*, North Holland, Amsterdam (1979).

48. A. Y. Kruger, “Subdifferentials of nonconvex functions and generalized directional derivatives” [in Russian], Deposited at VINITI, No. 2661-77, Minsk (1977).
49. A. Y. Kruger, “Generalized differentials of nonsmooth functions” [in Russian], Deposited at VINITI, No. 1332-81, Minsk (1981).
50. A. Y. Kruger, “Necessary conditions of an extremum in problems of nonsmooth optimization” [in Russian], Deposited at VINITI, No. 1333-81, Minsk (1981).
51. A. Y. Kruger, “ε-Semidifferentials and ε-normal elements” [in Russian], Deposited at VINITI, No. 1331-81, Minsk (1981).
52. A. Y. Kruger, “Generalized differentials of nonsmooth functions and necessary conditions for an extremum,” *Sib. Math. J.,* **26**, 370–379 (1985).
53. A. Y. Kruger, “Properties of generalized differentials,” *Sib. Math. J.,* **26**, 822–832 (1985).
54. A. Y. Kruger, “Covering theorem for set-valued mappings,” *Optimization,* **19**, 763–780 (1988).
55. A. Y. Kruger, “On calculus of strict ε-semidifferentials,” *Dokl. Akad. Nauk Belarusi,* **40**, No. 4, 34–39 (1996).
56. A. Y. Kruger, “Strict ε-semidifferentials and differentiating set-valued mappings,” *Dokl. Akad. Nauk Belarusi,* **40**, No. 6, 38–43 (1996).
57. A. Y. Kruger, “Strict ε-semidifferentials and extremality conditions,” *Dokl. Akad. Nauk Belarusi,* **41**, No. 3, 21–26 (1997).
58. A. Y. Kruger, “On extremality of sets systems,” *Dokl. Nat. Akad. Nauk Belarusi,* **42**, No. 1, 24–28 (1998).
59. A. Y. Kruger, “On necessary and sufficient extremality conditions,” In: *Dynamical Systems: Stability, Control, Optimization, Abstr.*, Vol. 2, Minsk (1998), pp. 165–168.
60. A. Y. Kruger, “Strict (ε, δ)-semidifferentials and extremality of sets and functions,” *Dokl. Nat. Akad. Nauk Belarusi,* **44**, No. 4, 21–24 (2000).
61. A. Y. Kruger, “Strict (ε, δ)-semidifferentials and extremality conditions,” *Optimization* (to appear).
62. A. Y. Kruger and B. S. Mordukhovich, “Extremal points and the Euler equation in nonsmooth optimization,” *Dokl. Akad. Nauk BSSR,* **24**, No. 8, 684–687 (1980).
63. A. Y. Kruger and B. S. Mordukhovich, “Generalized normals and derivatives and necessary conditions of an extremum in problems of nondifferentiable programming” [in Russian], Deposited at VINITI, I: No. 408-80, II: No. 494-80, Minsk (1980).
64. B. Kummer, “Metric regularity: Characterizations, nonsmooth variations and successive approximation,” *Optimization,* **46**, 247–281 (1999).
65. B. Kummer, “Inverse functions of pseudo regular mappings and regularity conditions,” *Math. Program., Ser. B,* **88**, 313–339 (2000).
66. B. S. Mordukhovich, “Maximum principle in problems of time optimal control with nonsmooth constraints,” *J. Appl. Math. Mech.,* **40**, 960–969 (1976).
67. B. S. Mordukhovich, *Approximation Methods in Problems of Optimization and Control* [in Russian], Nauka, Moscow (1988).
68. B. S. Mordukhovich, “Complete characterization of openness, metric regularity and Lipschitzian properties of multifunctions,” *Trans. Amer. Math. Soc.,* **340**, 1–35 (1993).
69. B. S. Mordukhovich, “Generalized differential calculus for nonsmooth and set-valued mappings,” *J. Math. Anal. Appl.,* **183**, 250–288 (1994).
70. B. S. Mordukhovich, “Coderivatives of set-valued mappings: calculus and applications,” *Nonlinear Anal.,* **30**, 3059–3070 (1997).
71. B. S. Mordukhovich, “The extremal principle and its applications to optimization and economics,” In: *Optimization and Related Topics* (A. Rubinov and B. Glover, Eds.), Appl. Optimization, Vol. 47, Kluwer (2001), pp. 343–369.

72. B. S. Mordukhovich and Y. Shao, “Differential characterizations of covering, metric regularity, and Lipschitzian properties of multifunctions between Banach spaces,” *Nonlinear Anal.,* **25**, 1401–1424 (1995).
73. B. S. Mordukhovich and Y. Shao, “Extremal characterizations of Asplund spaces,” *Proc. Amer. Math. Soc.,* **124**, 197–205 (1996).
74. B. S. Mordukhovich and Y. Shao, “Nonconvex differential calculus for infinite dimensional multifunctions,” *Set-Valued Anal.,* **4**, 205–236 (1996).
75. B. S. Mordukhovich and Y. Shao, “Nonsmooth sequential analysis in Asplund spaces,” *Trans. Amer. Math. Soc.,* **348**, 1235–1280 (1996).
76. B. S. Mordukhovich and Y. Shao, “Fuzzy calculus for coderivatives of multifunctions,” *Nonlinear Anal., Theory, Methods Appl.,* **29**, 605–626 (1997).
77. B. S. Mordukhovich and Y. Shao, “Stability of set-valued mappings in infinite dimensions: point criteria and applications,” *SIAM J. Control Optimization,* **35**, 285–314 (1997).
78. B. S. Mordukhovich, Y. Shao, and Q. J. Zhu, “Viscosity coderivatives and their limiting behavior in smooth Banach spaces,” *Positivity,* **4**, 1–39 (2000).
79. B. S. Mordukhovich and B. Wang, “On variational characterizations of Asplund spaces,” In: *Constructive, Experimental, and Nonlinear Analysis* (M. Thera, Ed.), Can. Math. Soc. Conf. Proc. Vol. 27 (2000), pp. 245–254.
80. B. S. Mordukhovich and B. Wang, “Necessary suboptimality and optimality conditions via variational principles,” *SIAM J. Control Optimization* (to appear).
81. L. W. Neustadt, “A general theory of extremals,” *J. Comput. Syst. Sci.,* **3**, 57–92 (1969).
82. L. W. Neustadt, *Optimization: A Theory of Necessary Conditions*, Princeton Univ. Press, Princeton, New Jersey (1976).
83. A. Nijenhuis, “Strong derivatives and inverse mappings,” *Amer. Math. Monthly,* **81**, 969–980 (1974).
84. E. A. Nurminsky, “On properties of a class of functions,” In: *Theory of Optimal Decisions* [in Russian], Kiev (1972), pp. 92–96.
85. E. A. Nurminsky, *Numerical Methods for Solving Deterministic and Stochastic Minimax Problems* [in Russian], Naukova Dumka, Kiev (1979).
86. J.-P. Penot, “Calcul sous-differentiel et optimisation,” *J. Funct. Anal.,* **27**, 248–276 (1978).
87. J.-P. Penot, “Metric regularity, openness and Lipschitz behavior of maps,” *Nonlinear Anal. Theory, Methods Appl.,* **13**, 629–643 (1989).
88. R. R. Phelps, *Convex Functions, Monotone Operators and Differentiability*, Lect. Notes Math., Vol. 1364, Springer-Verlag, New York (1993).
89. B. N. Pshenichny, “On necessary conditions for an extremum of nonsmooth functions,” *Kibernetika,* **6**, 92–96 (1977).
90. B. N. Pshenichnyi, *Necessary Conditions for an Extremum*, Marcel Dekker, New York (1971).
91. R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, New Jersey (1970).
92. R. T. Rockafellar, “Proximal subgradients, marginal values, and augmented Lagrangians in nonconvex optimization,” *Math. Oper. Res.,* **6**, 424–436 (1981).
93. R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, New York (1998).
94. E. Sachs, “Differentiability in optimization theory,” *Math. Operationsforsch. Statist., Ser. Optimization,* **9**, 497–513 (1978).
95. J. S. Treiman, “Clarke’s gradients and ε-subgradients in Banach spaces,” *Trans. Amer. Math. Soc.,* **294**, 65–78 (1986).
96. P. Varaiya, “Nonlinear programming in Banach space,” *SIAM J. Appl. Math.,* **15**, 284–293 (1967).
97. J. Warga, “Necessary conditions without differentiability assumptions in optimal control,” *J. Diff. Equations,* **15**, 13–46 (1975).
98. J. Warga, “Derivate containers, inverse functions and controllability,” In: *Calculus of Variations and Control Theory* (D. L. Russell, Ed.), Academic Press, New York (1976), pp. 13–46.

99. J. Warga, "Controllability and a multiplier rule for nondifferentiable optimization problems," *SIAM J. Control Optimization,* **16**, 803–812 (1978).
100. J. Warga, "An implicit function theorem without differentiability," *J. Diff. Equations,* **15**, 65–69 (1978).
101. J. Warga, "Fat homeomorphisms and unbounded derivate containers," *J. Math. Anal. Appl.,* **81**, 545–560 (1981).
102. J. Warga, "Optimization and controllability without differentiability assumptions," *SIAM J. Control Optimization,* **81**, 837–855 (1983).
103. Q. J. Zhu, "Clarke–Ledyaev mean value inequality in smooth Banach spaces," *Nonlinear Anal., Theory, Methods Appl.,* **32**, 315–324 (1996).
104. Q. J. Zhu, "The equivalence of several basic theorems for subdifferentials," *Set-Valued Anal.,* **81**, 171–185 (1998).