On the theory of subdifferentials

Alexander D. Ioffe

Abstract. The theory presented in the paper consists of two parts. The first is devoted to basic concepts and principles such as the very concept of a subdifferential, trustworthiness and its characterizations, geometric consistence, fuzzy principles and calculus rules, methods of creation of new subdifferentials etc. In the second part we study certain specific subdifferentials, namely, subdifferentials associated with bornologies, their limiting versions and metric modifications. For each subdifferential we verify that the basic properties discussed in the first part are satisfied and prove calculus rules for two main operations: summation and partial minimization. Separate sections are devoted to the Fréchet and limiting Fréchet subdifferentials and to the approximate $G$-subdifferential. For the last two new definitions are given which lead to a certain unification and simplification of analysis. The sum rule for these two subdifferentials is proved with the so-called “linear metric qualification condition”, so far the most general. In the last section we briefly discuss how other operations reduce to the two mentioned basic operations and give the corresponding calculus rules (with suitable versions of the metric qualification conditions).

Keywords. Banach space, subdifferential calculus, limiting subdifferentials, metric modification, $G$-subdifferential, normal cone, coderivative, trustworthiness, separable space, separable reduction.

2010 Mathematics Subject Classification. 49J52, 49J53, 26E15, 58C06, 58C20.

1 Introduction

This paper is an attempt to present the theory of subdifferentials of nonsmooth and nonconvex functions on infinite dimensional Banach spaces in a unified and systematic way. The necessity to study such “bad” functions and mappings was mainly determined by the needs of optimization and optimal control theories. The first subdifferentials for such functions were introduced in mid-70s ([14, 51, 57]), more appeared shortly afterwards ([19, 20, 31, 45, 47, 50]) and since then nonconvex subdifferentials have developed into a powerful instrument of analysis with a wide range of applications. The crucial role of smooth variational principles of Borwein–Preiss and Deville–Godefroy–Zizler in these developments should be emphasized.
Multiplicity of subdifferentials used now in variational analysis is necessitated by complexity of objects, on the one hand, and differences in geometries of different classes of Banach spaces, on the other, so that certain constructions that work well in some spaces are not very useful in others. The idea to find a common viewpoint for the variety of available and utilizable subdifferentials is by no means new. We can mention e.g. [2, 5, 18, 34, 41, 54, 64] where one or another generalizing concept was offered, each time, however, to embrace properties relating to some specific topic or group of results. Interconnections between various types of subdifferentials have been thoroughly studied in many publications since the very beginning, but no real attempt has been made to develop a sufficiently comprehensive theory and/or unified techniques that could be equally applied to various subdifferentials.

The reader now has to turn to different sources to learn about one or another subdifferential. A sufficiently complete account of the finite dimensional theory is contained in [63]. The calculus of Clarke’s generalized gradients is presented in [15, 17]. For acquaintance with Fréchet subdifferential one can go to [13], a fairly thorough study of limiting Fréchet subdifferentials can be found in [54], information about approximate subdifferentials is scattered in many non-monographic publications.

The two main new elements that have made the unification suggested in this paper possible are a (rather simple) observation that a more accurate fuzzy calculus of $\tilde{\beta}$-subdifferential (associated with bornologies) is available with the level of fuzziness gradually varying from fuzziness up to a norm neighborhood of zero for the Fréchet subdifferential to fuzziness up to a weak*-neighborhood for the subdifferential associated with the simple (Gâteaux) bornology, and, on the other hand, a new series of subdifferentials, called here “metric modifications” that includes the limiting Fréchet subdifferential and the approximate $G^*$-subdifferential at its opposite ends.

The paper consists of two parts. The first, “General theory” is devoted to principles and results that are equally applied to all subdifferentials. The second part, “Special theory” deals with some basic specific subdifferentials that are studied on the basis of the theory developed in the first part but with peculiarities of specific constructions taken into account. Before we pass to the description of the content of the paper, I would like to mention three basic technical ideas that were behind the creation of new subdifferentials in the course of the developments.

The first idea, due to Clarke [14], was to invoke the distance functions to define normal cones. The idea is instrumental in extending subdifferentials from Lipschitz to arbitrary lower semicontinuous functions. Introduced first in the purely convex context of generalized gradients, it later played a crucial role at the final step of construction of the approximate $G^*$-subdifferential (which as we shall see is
the only subdifferential having a certain collection of useful properties and at the same time can be trusted on all Banach spaces). It is also the basis for the definition of metric modifications in this paper.

The second idea originated in Kruger’s mimeographed paper [45] was to consider sequential weak*-closures rather than closures in the weak*-topology of the dual space\(^1\). Its role is mainly in providing reasonable balance between contradictory requirements of robustness (that is, a certain level of stability of the subdifferential with respect to small variations of the point of subdifferentiation) and quality of approximation. The first makes limiting (topological or sequential) procedures inevitable while the second makes subdifferentials with smaller values more desirable.

The third was the idea of “lifting” subdifferentials from simple spaces (e.g. finite dimensional or smooth) where certain subdifferentials can be simply and naturally defined into spaces with more complex geometry. This idea was first used in my papers [29,31,35] to define various approximate subdifferentials. A very different and totally independent implementation of this idea was the deep separable reduction theorem of Fabian and Zhivkov [28] and its modification by Fabian [25] which was the basis for the subsequent developments of the calculus of the Fréchet (Fabian [25]) and limiting Fréchet subdifferentials (Mordukhovich and Shao [55, 56]) in Asplund spaces.

Here is the table of content of the paper.

§ 1. Introduction.

Part 1  General theory

§ 2. Basic definitions.

§ 3. Trustworthiness on classes of spaces.

§ 4. Tightness and weak fuzzy calculus.

§ 5. Sequential and topological closures.


Part 2  Special theory

§ 7. Elementary subdifferentials

§ 8. Limiting and modified $\beta$-subdifferentials.

\(^1\) I believe very few are aware of the existence of this paper (as I also was till very recently). The reason, probably, is that the only reference to this paper available in the literature (I have been able to find) is in Kruger’s [46] and the very reference was too modest to suggest that the results of [46] were already contained in [45]. See also the footnote at p. 113.

§ 10. $G$-subdifferential.

§ 11. Some other operations.

§ 2 contains a definition of a subdifferential in which we axiomatize basic properties shared by all interesting subdifferentials. We do not know whether the axioms are independent, but the examples of constructions that fail to satisfy even only one of the axioms (not presented in the paper) all produce something highly unpleasant. We further introduce the concept of trustworthiness, one of the central in the theory, which determines which subdifferential can be productively used on a given space. Its role is demonstrated by proofs of an approximate mean value theorem and a dense subdifferentiability theorem with trusted subdifferentials. In § 3 we prove (necessary parts of) the equivalence theorem with several important characterization of trustworthiness. In § 4 we introduce the concept of tightness, another fundamental property, characterizing a reasonable level of lower approximation of a function by elements of the subdifferential. Unlike trustworthiness, tightness is a property of a subdifferential alone, not connected with the space. But for trusted subdifferentials, tightness is a key to a “weak fuzzy calculus” with calculus rules valid up to an arbitrary weak*-neighborhood of the origin. In § 5 we consider operations that allow to get new subdifferentials from the given ones. The main problem is to show that the new objects are subdifferentials in the sense of § 2 and to check whether other essential properties of subdifferentials are satisfied. In § 6 we study the mentioned operation of metric modification that allows to get a subdifferential on the class of lower semicontinuous functions, given a subdifferential on the class of locally Lipschitz functions.

The titles of sections of the second part of the paper speak for themselves. We study here certain specific constructions and build a sort of a pyramid containing main subdifferentials used in variational analysis. At the lower level of the pyramid there are “elementary” $\beta$-subdifferentials, the second level is occupied by their limiting versions and metric modifications which are generally smaller and better behaving than the limiting subdifferentials. At the upper level of the pyramid (to which we climb using one or another lifting techniques) there are three main “robust” subdifferentials: the limiting Fréchet subdifferential (on Asplund spaces), the approximate $G$-subdifferential and the generalized gradient. A characteristic property of the Fréchet subdifferential (among other $\beta$-subdifferentials) is that its limiting version coincides with the metric modification. We also give a new definition of the $G$-subdifferential that allows to make proofs of certain facts for the $G$-subdifferential and for the limiting Fréchet subdifferentials almost identical (and actually easier than available in the literature). In each section we prove calculus rules for two basic functional operations: summation and minimization with
respect to one of the variables (marginal function), of course for the subdifference studied in the section. As is well known, other operations involving functions, sets and set-valued mappings can be reduced to the two mentioned basic operations. In the last section we do this for some of them and state corresponding calculus rules for these operations, for brevity only for the limiting Fréchet and \( G \)-subdifferentials. We systematically use so-called “linear metric qualification conditions” in the calculus, so far the most general and stated in terms of the original objects (rather than subdifferentials themselves).

In the paper we concentrate on two main topics: how new subdifferentials are born (including verification that they satisfy the basic properties (axioms) of the given list) and the calculus proper – how a subdifferential of a composite function is connected with the same subdifferentials of the component functions.

Frameworks of a journal paper do not allow to touch upon many important and interesting issues. We talk very little or do not talk at all about Clarke’s generalized gradients, well studied and presented in the literature (although the definition of the generalized gradient through convexification of smaller subdifferentials or normal cones allows to get some of its basic properties with extremely little effort), proximal subdifferentials, probably the best in terms of their approximating power among “elementary” subdifferentials, but with very restricted domain of trust, Michel–Penot “moderate” subdifferential which shares some valuable properties of elementary subdifferentials and generalized gradient, small Treiman’s \( B \)-subdifferential and so-called controlled limits. We also do not discuss such topics as connection with tangency, integration of subdifferentials, compactness properties and their connection with tightness and qualification conditions, scalarization of coderivatives, although there are reasons to believe that they may find a place in the frameworks of the general theory.

**Notation.** All spaces are real and normed, typically Banach. As usual we denote by \( \langle x^*, x \rangle \) the action of an \( x^* \in X^* \) on \( x \in X \). By \( B \) we denote the unit ball, often with a subscript indicating the space. By the standard sum norm in a product space \( X \times Y \) we mean \( \|(x, y)\| = \|x\| + \|y\| \), but occasionally we consider other norms as well. Functions are usually assumed extended-real-valued with the standard notation

\[
\text{dom } f = \{x : |f(x)| < \infty\}
\]

for the domain of \( f \) and

\[
\text{epi } f = \{(x, \alpha) : \alpha \geq f(x)\}
\]

for its epigraph. By \( d_Q \) we denote the distance function (to \( Q \)) and by \( \text{Ind}_Q \) the indicator of \( Q \) which is the function equal to zero on \( Q \) and \( +\infty \) outside of \( Q \). If \( f \) is a function, then \( f \mid_Q \) is the restriction of \( f \) to \( Q \).
Part 1 General theory

2 Basic definitions

We start with a general definition of a subdifferential.

Definition 2.1. By a subdifferential we mean a correspondence which associates with any triple \((X, f, x)\), where \(X\) is a normed space, \(x \in X\) and \(f\) is a function on \(X\), a set \(\partial f(x) \subseteq X^*\) in such a way that the following properties are satisfied:

- **substantiality**, i.e.
  \[(S1) \quad \partial f(x) = \emptyset \text{ if } x \notin \text{dom } f.\]

- **localizability**, i.e.
  \[(S2) \quad \text{if } f \text{ and } g \text{ coincide in a neighborhood of } x, \text{ then } \partial f(x) = \partial g(x).\]

- **contiguity**, i.e.
  \[(S3) \quad \text{if } f \text{ is convex, then } \partial f(x) \text{ coincides with the subdifferential of } f \text{ in the sense of convex analysis, that is } \partial f(x) = \{x^* : \langle x^*, w \rangle \leq f(x + w) - f(x), \forall w \in X\}.\]

- **optimality**, i.e.
  \[(S4) \quad \text{if } f \text{ attains a local minimum at } x \in \text{dom } f, \text{ then } 0 \in \partial f(x).\]

- **calculability**, i.e.
  \[(S5a) \quad \text{if } f \text{ is of the form } f(x) = \lambda g(Ax + b) + \langle \ell, x \rangle + \alpha, \text{ where } X, Y \in \mathcal{X}, \lambda > 0, A \in L(X, Y) \text{ with Im } A = Y, b \in Y, \ell \in X^*, \alpha \in \mathbb{R}, \text{ then } \partial f(x) = \lambda A^* \partial g(Ax + b) + \ell,\]
  \[(S5b) \quad \text{if } f(x, y) = g(x) + h(y), \text{ then } \partial f(x, y) \subseteq \partial g(x) \times \partial h(y).\]

- **boundedness**, i.e.
  \[(S6) \quad \text{if for a given norm in } X, f \text{ satisfies the Lipschitz condition with rate } r \text{ near } x, \text{ then } \|x^*\| \leq r \text{ for any } x^* \in \partial f(x).\]

Remark 2.2. It is the property (S5a) and, specifically, a reference to an operator with the image covering the entire range space that makes us speak about Banach (rather than normed) spaces.

Remark 2.3. Strictly speaking, in a definition of a subdifferential we have to mention the class of functions and spaces on which it is defined or considered. In many cases however definitions make sense for all functions on all Banach space. Therefore we mention the classes only when it is necessary or essential for the future discussion.
The list of the properties in the definition does not include any reference to inter-relations between subdifferentials and derivatives. A partial reason is the following fact which is a consequence of (S6) (and (S5a)).

**Proposition 2.4.** If \( f \) is strictly Fréchet differentiable at \( x \), then \( \partial f(x) \subset \{ f'(x) \} \).

**Proof.** Set \( \ell = f'(x) \) and \( \varphi(h) = f(x + h) - f(x) - \langle \ell, h \rangle \). As \( f \) is strictly differentiable at \( x \), for any \( \varepsilon > 0 \) we may find a neighborhood of zero such that the Lipschitz constant of \( \varphi \) is smaller than \( \varepsilon \) in the neighborhood. Thus by (S6) the norm of any element of \( \partial \varphi(0) \) is not greater than \( \varepsilon \), hence zero may be the only element of \( \partial \varphi(0) \). On the other hand if \( x^* \in \partial f(x) \), then \( x^* - \ell \in \partial \varphi(0) \) by (S5a), that is, \( \ell \) may be the only element of \( \partial f(x) \).

**Definition 2.5.** A subdifferential is *elementary* if

\[ \partial f(x) + \partial g(x) \subset \partial(f + g)(x) \]

whenever \( f \) and \( g \) are finite at \( x \).

An elementary subdifferential has the following monotonicity property.

**Proposition 2.6** (Monotonicity of elementary subdifferentials). Let \( \partial \) be an elementary subdifferential, and let \( f, g \) be two functions on \( X \) such that \( f(x) = g(x) \) and \( f(x) \leq g(x) \) in a neighborhood of \( x \). Then \( \partial f(x) \subset \partial g(x) \).

**Proof.** We have \( g(x) = f(x) + \varphi(x) \), where \( \varphi \) attains a local minimum at \( x \) so that \( 0 \in \partial \varphi(x) \) by the property (b). Hence \( \partial f(x) \subset \partial f(x) + \partial \varphi(x) \subset \partial g(x) \).

**Proposition 2.7.** If \( \partial \) is an elementary subdifferential, then (S6) follows from (S3).

**Proof.** Indeed, if \( f \) is Lipschitz of rank \( K \) near \( x \), then

\[ f(u) \leq f(x) + K\|u - x\| = \varphi(u). \]

As \( \varphi \) is a convex function, we get \( \partial \varphi(x) = KBX^* \) by (S3) and since \( \partial \) is elementary, \( \partial f(x) \subset \partial \varphi(u) \).

We now proceed with the introduction of related geometric concepts.

**Definition 2.8.** The set (which is a cone by (S5))

\[ N(S, x) = \partial(\text{Ind}_S)(x) \]

is called the *normal cone to \( S \) at \( x \) (associated with \( \partial \)).
If $F : X \rightrightarrows Y$ is a set-valued mapping and $y \in F(x)$, then the (set-valued) mapping $Y^* \to X^*$ defined by

$$y^* \mapsto D^* F(x, y)(y^*) = \{x^* : (x^*, -y^*) \in N(\text{Graph } F, (x, y))\}$$

is called the coderivative of $F$ at $(x, y)$ (associated with $\partial$).

**Definition 2.9.** The singular subdifferential of $f$ at $x$ is

$$\partial f^\infty(x) = \{x^* : (x^*, \gamma) \in N_{\text{epi } f}(x, f(x)) \text{ for some } \gamma \geq 0\}.$$

**Definition 2.10.** The subdifferential $\partial$ is said to be geometrically consistent if

(a) $\gamma \leq 0$ whenever $(x^*, \gamma) \in N(\text{epi } f, (x, \alpha))$,

(b) $\partial f^\infty(x) = \{0\}$ if $f$ is Lipschitz near $x$

and

(c) the following relations are equivalent:

$$(x^*, -1) \in N(\text{epi } f, (x, \alpha)) \quad \text{and} \quad \alpha = f(x) \& x^* \in \partial f(x).$$

For a geometrically consistent subdifferential, the concept of coderivative turns out to be a direct extension of the concept of subdifferential to set-valued mappings. Indeed, let $[\text{epi } f]$ stand for the set-valued mapping $x \mapsto f(x) + \mathbb{R}_+$ from $X$ into $\mathbb{R}$.

**Proposition 2.11.** If $\partial$ is geometrically consistent, then

$$\partial f(x) = D^*[\text{epi } f](x, f(x))(1).$$

The following concept plays a crucial role in the calculus: it determines the type of subdifferentials that can be used for local variational analysis on a given class of spaces or, alternatively, the class of spaces on which a given subdifferential works.

**Definition 2.12.** It is said that a subdifferential $\partial$ is trusted on $X$, or that $\partial$ is a trustworthy subdifferential on $X$, or that $X$ is trustworthy space for $\partial$ if the following fuzzy minimization rule holds: let $f$ be a lower semicontinuous function on $X$ finite at $\overline{x} \in X$, and let $g$ be Lipschitz continuous on $X$. Assume that $f + g$ attains a local minimum at $\overline{x}$. Then for any $\varepsilon > 0$ there are $x, u \in X$ and $x^* \in \partial f(x)$, $u^* \in \partial g(u)$ such that

$$\|x - \overline{x}\| < \varepsilon, \quad \|u - \overline{x}\| < \varepsilon, \quad |f(x) - f(\overline{x})| < \varepsilon, \quad \text{and} \quad \|x^* + u^*\| < \varepsilon.$$

We say that $\partial$ is trusted on a class of spaces if it is trusted on each space of the class.
Proofs of trustworthiness of specific subdifferentials on specific classes of spaces are typically based on the use of one or another variational principle which require completeness of the space. It is another reason why Banach spaces are in the focus of our attention. Many interesting things can be said about trusted subdifferentials. Here is one of them.

**Theorem 2.13** (Mean value theorem). Let $X$ be a Banach space, let $\partial$ be a subdifferential trusted on $X$, let $f$ be a function on $X$ which is defined, lower semicontinuous and bounded below on an open convex set $U$ containing the line segment $[\bar{x}, \bar{y}]$ that joins two given points $\bar{x}$ and $\bar{y}$ at which $f$ is finite. Then for any $\varepsilon > 0$ there exists a pair $(u, u^*)$ such that the distance from $u$ to $[\bar{x}, \bar{y}]$ is less than $\varepsilon$, $u^* \in \partial f(u) + \varepsilon B$ and either $\langle u^*, x - y \rangle = f(\bar{x}) - f(\bar{y})$ or $u \in B(\bar{x}, \varepsilon)$ and $\langle u^*, \bar{y} - \bar{x} \rangle \geq f(\bar{y}) - f(\bar{x})$.

**Proof.** We can assume without loss of generality that $\|\bar{y} - \bar{x}\| = 1$ and replace the original norm in $X$ by an equivalent norm more convenient for our calculation. Namely, let $L$ be the one-dimensional subspace spanned by $\bar{y} - \bar{x}$, and let $Z$ be a complementary subspace of codimension one. Then for any $x \in X$ we have a unique representation $x = z + \lambda(\bar{y} - \bar{x})$ and set $\|x\| = (\|z\|^2 + \lambda^2)^{1/2}$.

Consider first the case when $f(\bar{x}) = f(\bar{y})$. Set

$$\rho = \min \{ f(x) : x = t\bar{x} + (1-t)\bar{y}, \ 0 \leq t \leq 1 \},$$

and let $\bar{u}$ be the point at which the minimum is attained. Then either $\bar{u}$ is in the interior of the segment or we can set $\bar{u} = \bar{x}$. Choose $\varepsilon \in (0, 1)$ so small that $f(x) > \rho - 1$ if $d(x, [\bar{x}, \bar{y}]) < 2\varepsilon$ and, in case when $\bar{u}$ belongs to the interior of the segment $[\bar{x}, \bar{y}]$, the distance from $\bar{u}$ to its ends is greater than $2\varepsilon$. Consider the function

$$g(x) = f(x) + Kd(x, [\bar{x}, \bar{y}])$$

with $K > 1/\varepsilon$. Then $g(x) > \rho$ if $2\varepsilon \geq \|x - \bar{u}\| \geq \varepsilon$ and on the other hand the lower bound of $g$ on the $2\varepsilon$-ball around $\bar{u}$ is not greater than $g(\bar{u}) = \rho$. In other words, the points with values of $g$ close to its lower bound on $B(\bar{u}, 2\varepsilon)$ belong to the interior $B(\bar{u}, \varepsilon)$.

Take a $\delta < \varepsilon^2/2$. By Ekeland’s principle there is a point $x$ such that

$$\|x - \bar{u}\| < 2\varepsilon, \quad f(x) \leq g(x) \leq g(\bar{u}) = \rho$$

and the function $u \mapsto g(u) + (\delta/\varepsilon)\|u - x\|$ attains a local minimum at $x$. Hence zero belongs to the subdifferential of the function at $x$. This function is a sum of $f$ and a convex continuous function $(\delta/\varepsilon)\|\cdot - x\| + Kd(\cdot, [\bar{x}, \bar{y}])$. We therefore can apply the fuzzy minimization rule and find two pairs $(u_i, u^*_i)$ such that

$$\|u_i - x\| < 2\varepsilon, \quad f(u_1) \leq f(x) + \varepsilon < \rho + \varepsilon$$
and 
\[ u_1^* \in \partial f(u_1), \quad u_2^* \in K d(\cdot, [\bar{x}, \bar{y}]) (u_2), \quad \|u_1^* + u_2^*\| < \varepsilon. \]

Then \( u_2^* \in d(\cdot, [\bar{x}, \bar{y}]) (w) \) if \( w \) is the nearest to \( u_2 \) point of \([\bar{x}, \bar{y}]\). By the choice of the norm in \( X \), we can be sure that \( \|w-u\| \leq \|u_2-\bar{x}\| < 2\varepsilon \).

Set \( u = u_1, \quad u^* = -u_2^* \). Then \( u^* \in \partial f(u) + \varepsilon B \). Furthermore, if \( \bar{u} \in (\bar{x}, \bar{y}) \), then \( w \in (\bar{x}, \bar{y}) \) as well and therefore \( \langle u^*, y-\bar{x} \rangle = 0 = f(\bar{y}) - f(\bar{x}) \). Otherwise \( w \) must coincide with \( \bar{x} \) so that \( \langle u^*, \bar{y} - \bar{x} \rangle \geq 0 \).

This proves the theorem when \( f(\bar{x}) = f(\bar{y}) \). Otherwise, take an \( x^* \in X^* \) such that \( \|x^*\| = 1 \) and \( \langle x^*, \bar{y} - \bar{x} \rangle = 1 \) and apply the already proven result to the function \( \varphi(x) = f(x) - (f(\bar{y}) - f(\bar{x})) \langle x^*, x-\bar{x} \rangle \).

**Remark 2.14.** If the function \( f \) is Lipschitz on \([\bar{x}, \bar{y}]\), the proof is much simpler: take \( g = f + \text{Ind}_{[\bar{x}, \bar{y}]} \) and apply directly Definition 2.12 immediately after the choice of \( \bar{u} \).

We conclude the section with one of the most principal consequences of trustworthiness.

**Theorem 2.15** (Dense subdifferentiability theorem). Let \( \partial \) be a subdifferential trusted on \( X \). Then:

(a) If \( f \) is a lower semicontinuous extended-real-valued function on \( X \), then \( \partial f(x) \neq \emptyset \) for all \( x \) of a dense subset of \( \text{dom } f \).

(b) If \( Q \subset X \) is a closed set, then the collection of \( x \in Q \) at which \( N(S, x) \) contains a nonzero vector is dense in the boundary of \( Q \).

**Proof.** (a) Take an \( \bar{x} \in \text{dom } f \) and an \( \varepsilon > 0 \). We may assume \( \varepsilon \) small enough to guarantee for example that \( f(x) \geq f(\bar{x}) - 1 \) if \( \|x-\bar{x}\| \leq \varepsilon \). By Ekeland’s principle there is a \( w \) such that \( \|w-\bar{x}\| \leq \varepsilon/2 \), and \( f(x) + (2/\varepsilon) \|x-w\| \) attains a local minimum at \( w \). This function is a sum of an l.s.c. and a Lipschitz functions. As \( \partial \) is trusted on \( X \), there is an \( x \) such that \( \|x-w\| < \varepsilon/2 \), hence \( \|x-\bar{x}\| < \varepsilon \), and \( \partial f(x) \neq \emptyset \).

(b) Let \( \bar{x} \) be a boundary point of \( Q \). Take an \( \varepsilon > 0 \) and find a \( w \notin Q \) such that \( \|w-\bar{x}\| < \varepsilon^2/2 \). By Ekeland’s principle there is a \( \bar{w} \in Q \) with \( \|\bar{w}-\bar{x}\| < \varepsilon/2 \) such that 
\[ g(x) = \|w-x\| + \text{Ind}_Q(x) + \varepsilon \|x-\bar{w}\| \]
attains its minimum at \( \bar{w} \). Again we have a sum of an l.s.c. and a Lipschitz function, so there are \( x \in Q \) and \( u \) such that \( \|u-\bar{w}\| \) is strictly smaller that the distance from \( w \) to \( Q \), \( \|\bar{x}-\bar{w}\| < \varepsilon/2 \) and for some \( x^* \) in the subdifferential of \( \|w-\cdot\| + \varepsilon \|\cdot-\bar{w}\| \) at \( u \) we have \( x^* \in N(Q, x) + \varepsilon B \). As \( u \neq w \), we conclude that \( \|x^*\| > 1 - \varepsilon \). \( \square \)
3 Trustworthiness on classes of spaces

Although in our analysis we often deal mainly with individual spaces (as say, in the last theorem), it is trustworthiness on a class of Banach spaces where by a class of Banach (or normed) spaces we mean any collection of spaces which contains Cartesian products of its elements. The most important (for variational analysis) are the classes of finite dimensional, separable, weakly compactly generated (WCG), Fréchet and Gâteaux smooth spaces, Asplund spaces, and of course the class of all Banach spaces.

There are many tests to verify trustworthiness on classes of spaces. Here we state and partly (only those parts that we need in the sequel) prove the fundamental equivalence theorem which can be considered the backbone of the general theory of subdifferentials. It should be emphasized that the formulations of trustworthiness and all equivalent properties display a certain amount of “fuzziness”: locations of all vectors, values of functions and elements of subdifferentials are defined only up to an arbitrarily small \( \epsilon \). We start the section with a brief discussion of a kind of robust minimization in normed spaces.

3.1 Robust minima

Formally the concept of a robust minimum is introduced as follows: consider the problem

\[
(P) \quad \text{minimize } f(x) \quad \text{s.t. } F(x) = 0, \ x \in S
\]

where \( F \) is a mapping from \( X \) into another space \( Y \). We make distinction between the “non-functional constraint” \( x \in S \) which cannot be violated and the equality type “functional constraint” \( F(x) = 0 \).

Along with \((P)\) we shall consider the \( \epsilon \)-regularization of the problem

\[
(P(\epsilon)) \quad \text{minimize } f(x) \quad \text{s.t. } \|F(x)\| \leq \epsilon, \ x \in S
\]

with small positive \( \epsilon \). Denote by \( v(\epsilon) \) the value of \((P(\epsilon))\), that is, the infimum of \( f \) subjects to constraints of the problem.

**Definition 3.1.** Vector \( \overline{x} \), admissible in \((P)\), is a robust minimizer in \((P)\) if

\[f(\overline{x}) = \lim_{\epsilon \to 0} v(\epsilon).\]

We also say that \( \overline{x} \) is a local robust minimum in \((P)\) if there is an \( r > 0 \) such that \( \overline{x} \) is a robust minimum in the problem in which \( S \) is replaced by \( B(\overline{x}, r) \cap S \).

Problem \((P)\) contains formulations of a huge variety of optimization problems. For the purpose of the subdifferential calculus one of the most specific instances
occurs when we need to minimize a sum of several functions

\[ f(x) = f_1(x) + \cdots + f_k(x) \]

on a certain closed set \( S \). The reduction of minimization of \( f \) on \( S \) to the form of (P) is based on decoupling of variables as follows. We can rewrite the minimization problem for \( f \) on \( S \) in the form

\[
\begin{aligned}
\text{minimize } & f_1(x_1) + \cdots + f_k(x_k) \\
\text{s.t. } & x_i - x_j = 0, \ x_i \in S, \ i, j = 1, \ldots, k.
\end{aligned}
\]

Thus the \( \varepsilon \)-regularization of the problem can be stated as

\[
\begin{aligned}
\text{minimize } & \sum_{i=1}^{k} f_i(x_i) \\
\text{s.t. } & \|x_i - x_j\| \leq \varepsilon, \ x_i \in S, \ i, j = 1, \ldots, k.
\end{aligned}
\]

In other words, \( \overline{x} \in S \) is a robust minimum of \( f \) on \( S \) if

\[
f(\overline{x}) = \lim_{\varepsilon \to 0} \inf \left\{ \sum_{i=1}^{k} f_i(x_i) : \|x_i - x_j\| \leq \varepsilon, \ x_i \in S, \ i, j = 1, \ldots, k \right\}. \quad (3.1)
\]

If \( S \) is a closed ball around \( \overline{x} \), we speak about a local robust minimum of \( f \).

The following simple proposition gives sufficient condition for a local minimum of a sum to be robust.

Proposition 3.2 (e.g. [13, Proposition 3.3.2]). Let \( f_1, \ldots, f_k \) be functions on \( X \) which are finite at \( \overline{x} \), and let \( S \subset X \) be a closed set containing \( \overline{x} \). Assume that \( f = f_1 + \cdots + f_k \) attains minimum on \( S \) at \( \overline{x} \). Then each of the following three conditions is sufficient for \( \overline{x} \) to be a robust minimum of \( f_1 + \cdots + f_k \) on \( S \):

(i) all functions but for at most one of them satisfy the Lipschitz condition on \( S \),

(ii) all functions but for at most one of them are uniformly continuous on \( S \),

(iii) all functions are bounded below and lower semicontinuous on \( S \) and one of the functions has compact intersections of its sublevel sets with \( S \).

3.2 Penalization

Penalty functions is one of the oldest and widely used technical devices both in theoretical and numerical optimization. Penalization techniques offer also one of the main tools in variational analysis. We start with a simple fact to be often used in the sequel.

Proposition 3.3. Let \( (X, d) \) be a metric space, let \( S \subset X \), and let \( f \) be a function on \( X \) satisfying the Lipschitz condition with constant \( K \) in a neighborhood of a certain \( \overline{x} \in S \). If \( f \) has a local minimum on \( S \) at \( \overline{x} \), then the function \( g \) defined by

\[
g(x) = f(x) + Kd(x, S)
\]

attains an unconditional local minimum at \( \overline{x} \).
Proof. We only need to verify that \( g(x) \geq g(\overline{x}) \) if \( x \not\in S \) is sufficiently close to \( \overline{x} \). Take an \( \varepsilon > 0 \) and a \( u \in S \) such that \( d(u, x) \leq d(x, S) + \varepsilon \). Then

\[
g(x) = f(x) + Kd(x, S) \geq f(u) + K(d(x, S) - d(x, u))
\]

\[
\geq f(\overline{x}) - K\varepsilon = g(\overline{x}) - K\varepsilon
\]

and the result follows as \( \varepsilon \) can be arbitrarily small. \( \square \)

This result basically says that constraint minimization of a Lipschitz function admits exact penalization. However we shall often work with lower semicontinuous functions in which case finding a universal exact penalties does not seem to be feasible.

**Definition 3.4.** An extended real-valued function \( \varphi \) on \( X \) is called a forcing function if it is nonnegative, \( \varphi(0) = 0 \) and \( x \to 0 \) whenever \( \varphi(x) \to 0 \).

Clearly a forcing function must be positive outside of zero. Below we shall consider several important examples of forcing functions and now we are going to proof the main penalization result relating to minimization of sums of lower semicontinuous functions.

Given functions \( f_1, \ldots, f_k \), all finite at \( \overline{x} \), we consider for any \( r > 0 \) the function

\[
p_r(x_1, \ldots, x_k) = \sum_{i=1}^{k} f_i(x_i) + r \sum_{i,j=1}^{k} \varphi(x_i - x_j) + \varphi(x_1 - \overline{x}).
\]

**Lemma 3.5** (Decoupling penalization lemma). Let \( f_1, \ldots, f_k \) be lower semicontinuous and bounded from below on a closed set \( S \subset X \), and let \( \varphi \) be a forcing function. Assume that \( \overline{x} \in S \) is a robust minimum of \( f = f_1 + \cdots + f_k \) on \( S \). Then the following holds true: if \( (\delta_n) \) is a sequence of positive numbers going to zero and \( (u_{1n}, \ldots, u_{kn}) \in S^k \) is a \( \delta_n \)-minimizer of \( p_n \) on \( S^k \), then \( u_{in} \to \overline{x} \) as \( n \to \infty \) for all \( i = 1, \ldots, k \).

**Proof.** Take an \( N > 0 \) such that \( \sum f_i(x_i) \geq -N \) for \( x_i \in S \). Set \( a_n = \inf p_n \). Then by the assumption

\[
a_n + \delta_n \geq \sum_{i=1}^{k} f_i(u_{in}) + n \sum_{i,j=1}^{k} \varphi(u_{in} - u_{jn}) + \varphi(u_{1n} - \overline{x})
\]

\[
\geq n \sum_{i,j=1}^{k} \varphi(u_{in} - u_{jn}) - N.
\]
It follows that \( \sum \varphi(u_{in} - u_{jn}) \leq n^{-1} (N + a_n + \delta_n) \to 0 \) as \( n \to \infty \). As \( \varphi \) is a forcing function, we conclude that \( u_{in} - u_{jn} \to 0 \) for all \( i, j \). Thus for any \( \varepsilon > 0 \) there is an \( n(\varepsilon) \) such that \( \|u_{in} - u_{jn}\| < \varepsilon \) if \( n \geq n(\varepsilon) \). Set

\[
v(\varepsilon) = \inf \left\{ \sum_{i=1}^{k} f_i(x_i) : \|x_i - x_j\| \leq \varepsilon, x_i \in S \right\}.
\]

Then for \( n \geq n(\varepsilon) \)

\[
\sum_{i=1}^{k} f_i(u_{in}) \geq v(\varepsilon) \to \sum_{i=1}^{k} f_i(\bar{x}) = f(\bar{x})
\]
as \( \varepsilon \to 0 \) since \( \bar{x} \) is a robust minimum of \( f \) on \( S \) and \( \gamma(\varepsilon) = v(\varepsilon) - f(\bar{x}) \to 0 \) when \( \varepsilon \to 0 \). It follows that for \( n \geq n(\varepsilon) \)

\[
a_n + \delta_n \geq p_n(u_{1n}, \ldots, u_{kn}) \geq v(\varepsilon) + \varphi(u_{1n} - \bar{x})
\]
\[
\geq f(\bar{x}) + \varphi(u_{1n} - \bar{x}) - \gamma(\varepsilon).
\]

Since \( f(\bar{x}) \geq a_n \), the latter implies that

\[
\varphi(u_{1n} - \bar{x}) \leq \delta_n + \gamma(\varepsilon)
\]
and therefore \( \varphi(u_{1n}) \to 0 \) as \( n \to \infty \), hence \( u_{1n} \to \bar{x} \).

The last question to discuss concerns the types of forcing functions available in Banach spaces. The first and obvious series of such functions is \( \varphi_s(x) = \|x\|^s \), where \( s > 0 \). Another useful class of forcing functions is associated with so-called bump functions.

**Definition 3.6.** A continuous function \( \varphi \) on a Banach space \( X \) is called a *bump function* or just *bump* if it is nonnegative, not identical zero, and equal to zero outside of the unit ball. We shall call \( \varphi \) a *strict bump* if in addition \( \varphi(x) < \varphi(0) \) for \( x \neq 0 \).

The existence of a bump with certain analytic properties carries valuable information about geometry of the space. For instance twice continuously Fréchet differentiable bump may exist only on a Hilbert space ([21, 27]). Likewise, a Lipschitz Fréchet differentiable bump functions may exist only on an Asplund space ([21]). For us the most important fact is that, given a strict bump \( \varphi \) (we can always assume that \( \varphi(0) = 1 \)), the functions

\[
\varphi_1 = 1 - \varphi \quad \text{and} \quad \varphi_2 = \varphi^{-1} : \varphi_2(x) = \begin{cases} 1, & \text{if } \varphi(x) > 0, \\ \frac{1}{\varphi(x)}, & \text{if } \varphi(x) = 0, \\ \infty, & \text{if } \varphi(x) = 0. 
\end{cases}
\]
are forcing functions. The concluding lemma proved below shows that every bump can be transformed into a strict bump without losing any of its analytic properties.

**Lemma 3.7.** If there is on $X$ a continuous bump function $\psi$, then there is a continuous strict bump. Moreover if $\psi$ is either Lipschitz or (continuously) Gâteaux or Fréchet differentiable, then there is a strict bump with the same properties.

**Proof.** First note that we can assume that $\psi(0) = 1$. If this is not the case, take any $a$ with $\psi(a) > (1/2) \sup \psi$ and replace $\psi$ by $\psi_1(x) = \psi(a)^{-1} \psi(a + 2x)$. (This is also a bump as $\|a\| < 1$.) differentiable or Lipschitz if so is the original bump).

We can also assume that $\psi(x) \leq 1 = \psi(0)$. If this is not the case, take a continuously differentiable $\theta(t)$ on $\mathbb{R}$ such satisfying $\theta(t) = t$ for $t \leq \xi$ and $\theta(t) = 1$ for $t \geq \psi(0)$ and replace $\psi$ by $\theta \circ \psi$. Again, the differentiability or Lipschitz properties are preserved under the transformation.

Now given a $q \in (0, 1)$, we set

$$\varphi(x) = (1 - q) \sum_{k=0}^{\infty} q^k \psi(q^{-k} x).$$

Then $\varphi(x) = 0$ if $\|x\| \geq 1$ and $\varphi(0) = \psi(0) = 1$. For $x \neq 0$ we have

$$\varphi(x) = (1 - q) \sum_{k=0}^{k(x)} q^{-k} \psi(q^k x) \leq (1 - q) \sum_{k=0}^{k(x)} q^{-k} < 1 = \varphi(0)$$

where $k(x)$ is the maximal integer for which $q^k \|x\| < 1$. If $\psi$ is Lipschitz or (continuously) differentiable in a certain sense, the last formula shows that $\varphi$ has the same properties.

\[\square\]

### 3.3 The equivalence theorem

**Definition 3.8.** Let $X$ be a Banach spaces, and let $\partial$ be a subdifferential. We say that the local fuzzy minimization principle holds for $\partial$ on $X$ if whenever $f_1, \ldots, f_k$ are lower semicontinuous functions on $X$ which are finite at $x$ and such that the sum $f = f_1 + \cdots + f_k$ attains for some $r > 0$ a robust local minimum at $\overline{x}$ on $B(\overline{x}, r)$, for any $\varepsilon > 0$ there are $x_1, \ldots, x_k, x_1^*, \ldots, x_k^*$ such that

$$\|x_i - \overline{x}\| < \varepsilon, \quad |f_i(x_i) - f_i(\overline{x})| < \varepsilon,$$

$$x_i^* \in \partial f_i(x_i), \quad \|x_1^* + \cdots + x_k^*\| < \varepsilon. \quad (3.3)$$

Given functions $f_1, \ldots, f_k$ on $X$, we set

$$a(f_1, \ldots, f_k, \eta) := \inf \left\{ \sum_{i=1}^{k} f_i(x_i) : \|x_i - x_j\| < \eta, i, j = 1, \ldots, k \right\}.$$
Definition 3.9. We say that the nonlocal fuzzy minimization principle holds for $\partial$ on $X$ if for any finite collection $(f_1, \ldots, f_k)$ of lower semicontinuous functions such that
\[
\liminf_{\eta \to 0} a(f_1, \ldots, f_k, \eta) > -\infty
\] (3.4)
and any $\varepsilon > 0$ and any $\eta > 0$ there are vectors $x_i, x_i^*, i = 1, \ldots, k$ such that for all $i, j = 1, \ldots, k$
\[
\|x_i - x_j\| < \varepsilon, \quad \sum_{i=1}^k f_i(x_i) < a(f_1, \ldots, f_k, \eta) + \varepsilon, \\
x_i^* \in \partial f_i(x_i), \quad \|x_1^* + \cdots + x_k^*\| < \varepsilon.
\]

Definition 3.10. Let $S_1, \ldots, S_k$ be closed subsets of $X$ with $\bar{x} \in S = \bigcap S_i$. It is said that $\bar{x}$ is an extremal point of $(S_1, \ldots, S_k)$ if there is an $r > 0$ such that for any $\varepsilon > 0$ there are vectors $a_i, i = 1, \ldots, k$, such that $\|a_i\| < \varepsilon$ and
\[
\bigcap_{i=1}^k \left( (S_i - a_i) \cap B(\bar{x}, r) \right) = \emptyset.
\]

Definition 3.11. It is said that the fuzzy extremal principle is valid for $\partial$ on $X$ if the following property holds for any collection $(S_1, \ldots, S_k)$ of closed subsets of $X$, and any extremal point $\bar{x}$ of $(S_1, \ldots, S_k)$: for any $\varepsilon > 0$ there are vectors $x_i \in S_i$, $i = 1, \ldots, k$, and $x_i^* \in N(S_i, x_i)$ such that
\[
\|x_i - \bar{x}\| < \varepsilon, \quad \max\{\|x_1^*\|, \ldots, \|x_k^*\|\} \geq 1, \quad \|x_1^* + \cdots + x_k^*\| < \varepsilon. \quad (3.5)
\]

Definition 3.12. It is said that the multi-directional mean value inequality holds on $X$ if for any convex closed set of $Q \subset X$, any vector $x \in X$, any l.s.c. function $f$ on $X$ bounded below on the set $[x, Q] = \text{conv}\{x\} \cup Q$, any $\varepsilon > 0$ and any $r < \lim_{\eta \to 0} \inf_{u \in Q + \eta B} f(u) - f(x)$ there is a $z \in [x, Q] + \varepsilon B$ and $z^* \in \partial f(z)$ such that
\[
\begin{align*}
& r < \langle z^*, u - x \rangle + \varepsilon\|u - x\|, \quad \forall u \in Q, \\
& f(z) < \lim_{\eta \to 0} \inf_{u \in [x, Q] + \eta B} f(u) + |r| + \varepsilon. \quad (3.6)
\end{align*}
\]

Definition 3.13. We say that the linear set minimization rule holds on $X$ if for any l.s.c. $f$ on $X$, any closed linear subspace $L \subset X$ and any sequence $(x_n) \subset X$ such that $d(x_n, L) \to 0$ and
\[
\lim_{n \to \infty} f(x_n) = \liminf_{\eta \to 0} a(f, \text{Ind}_L, \eta)
\]
there is a sequence of pairs \((u_n, u_n^*) \in X \times X^*\) such that
\[
\|x_n - u_n\| \to 0, \quad u_n^*|L \to 0, \quad d(x_n, L)\|u_n^*\| \to 0, \quad u_n^* \in \partial f(x_n).
\]

We shall say that one or another of the properties is valid for \(\partial\) on a class of normed spaces if it holds on each space of the class.

**Theorem 3.14.** For any subdifferential \(\partial\) and any class \(\mathcal{X}\) of Banach spaces the following properties are equivalent:

(a) \(\partial\) can be trusted on \(\mathcal{X}\),

(b) the local fuzzy minimization principle holds for \(\partial\) on \(\mathcal{X}\),

(c) the nonlocal fuzzy minimization principle holds for \(\partial\) on \(\mathcal{X}\),

(d) the multi-directional mean value inequality holds for \(\partial\) on \(\mathcal{X}\),

(e) the linear set minimization principle holds for \(\partial\) on \(\mathcal{X}\),

and each of them implies that

(f) the fuzzy extremal principle holds for \(\partial\) on \(\mathcal{X}\).

The opposite implication also holds for geometrically consistent subdifferentials.

**Proof.** We shall consider only (a), (b) and (f) needed in the sequel. For proofs of other equivalences see [13, 37, 49, 68].

(a) \(\Leftrightarrow\) (b). Implication (b) \(\Rightarrow\) (a) is immediate in view of Proposition 3.2. Let us prove the opposite implication. Let an \(X \in \mathcal{X}\), an \(x \in X\) and lower semicontinuous functions \(f_1, \ldots, f_k\) having a robust local minimum at \(x\) be given. Consider the function
\[
p_n(x_1, \ldots, x_k) = \sum_{i=1}^{k} f_i(x_i) + n \sum_{i=1}^{k} \|x_i - x_j\|^2 + \|x_1 - \bar{x}\|^2 \quad (3.7)
\]
and find and \(r > 0\) and sequences \((u_{in})\), \(i = 1, \ldots k\), such that
\[
p_n(u_{1n}, \ldots, u_{kn}) \leq \inf\{p_n(x_1, \ldots, x_k) : \|x_i - x_j\| < 1/n, \|x_i - \bar{x}\| \leq r\} + 1/n^2.
\]
By the Penalization Lemma all \(u_{in}\) converge to \(\bar{x}\).

Applying Ekeland’s variational principle with \(\lambda = 1/n\), we find \(w_{in}\) such that
\[
\|u_{in} - w_{in}\| \leq 1/n \text{ for all } i = 1, \ldots, k \text{ and for any } n \text{ the function}
\]
\[
(x_1, \ldots, x_k) \mapsto p_n(x_1, \ldots, x_k) + (1/n) \sum_{i=1}^{k} \|x_i - w_{in}\|
\]
attains a local minimum at \(w = (w_{1n}, \ldots, w_{kn})\).
Consider the space $Z = X^k$ with e.g. the max-norm

$$\|z\| = \max\{\|x_1\|, \ldots, \|x_k\|\} \quad \text{for } z = (x_1, \ldots, x_k)$$

and, for any $n$, the following two functions on $Z$:

$$\varphi_n(z) = \sum_{i=1}^k f_i(x_i),$$

$$\psi_n(z) = n \sum_{i=1}^k \|x_i - x_j\|^2 + (1/n) \sum_{i=1}^k \|x_i - w_{in}\| + \|x_1 - \bar{x}\|^2.$$

Then $\varphi_n$ is lower semicontinuous, $\psi_n$ is convex continuous and $\varphi_n + \psi_n$ attains a local minimum at $(w_{1n}, \ldots, w_{kn})$. By (a) we can find

$$z_n = (x_{1n}, \ldots, x_{kn}), \quad v_n = (y_{1n}, \ldots, y_{kn})$$

and

$$z_n^* = (x_{1n}^*, \ldots, x_{kn}^*) \in \partial \varphi(z_n), \quad v_n^* = (y_{1n}^*, \ldots, y_{kn}^*) \in \partial \psi(v_n)$$

such that $\|x_{in} - w_{in}\| < 1/n$, $\|y_{in} - w_{in}\| < 1/n$ and $\|z_n^* + v_n^*\| < 1/n$.

By (S5b), $x_{in}^* \in \partial f_i(x_{in})$. We have to show that $\|y_{1n}^* + \cdots + y_{kn}^*\| \to 0$ and for that we only need to verify that $\|y_{1n}^* + \cdots + y_{kn}^*\| \to 0$ as

$$\left\| \sum_{i=1}^k x_{in}^* - \sum_{i=1}^k y_{in}^* \right\| \leq \sum_{i=1}^k \|x_{in}^* + y_{in}^*\| = \|z_n^* + v_n^*\| \to 0.$$

By the standard rules of convex analysis,

$$y_{in}^* = \sum_{j \neq i} (p_{ijn}^* - p_{jin}^*) + q_{in}^* + s_{in}^*,\quad \text{where}$$

$$p_{ijn}^* \in \partial \|\|^2(y_{in} - y_{jn}), \quad q_{in}^* \in (1/n)\partial \|(y_{in} - w_{in})\|, \quad s_{in}^* \in \partial \|(y_{1n} - \bar{x})\|$$

(and $\delta_{ij}$ is the Kronecker’s factor, that is, 1 if $i = j$ and zero otherwise). Thus

$$\|y_{1n}^* + \cdots + y_{kn}^*\| \leq \|q_{1n}^*\| + \cdots + \|q_{kn}^*\| + \|s_n^*\|,\quad \text{and every term in the right-hand side obviously goes to zero.}$$
(a) ⇒ (f). Let $\bar{x}$ be an extremal point of $(S_1, \ldots, S_k)$, and let an $\varepsilon > 0$ be given. Take a positive $\delta < r$ and choose $a_i$ as in Definition 3.10 but with $\delta^2/2k$ instead of $\varepsilon$. We have $0 \not\in \bigcap_i (S_i - (x + a_i))$ for $x$ close to $\bar{x}$ and

$$\sum_{i=1}^k d_{S_i}(\bar{x} + a_i) \leq \delta^2/2.$$ 

Find next $\bar{x}_i \in S_i$ such that $\|\bar{x}_i - \bar{x}\| < \delta^2/k$ and consider the following function on $X^{k+1}$:

$$f(x, x_1, \ldots, x_k) = \sum_{i=1}^k \|x_i - (x + a_i)\| + \sum_{i=1}^k \text{Ind}_{S_i}(x_i).$$

This function is nonnegative and $f(\bar{x}, \bar{x}_1, \ldots, \bar{x}_k) \leq \delta^2$. By Ekeland’s principle we can find $\bar{x}, \bar{x}_1, \ldots, \bar{x}_k$, $\bar{x}_i \in S_i$, such that $f(\bar{x}, \bar{x}_1, \ldots, \bar{x}_k) \leq \delta^2$ and

$$g(x, x_1, \ldots, x_k) = f(x, x_1, \ldots, x_k) + \delta \left(\|x - \bar{x}\| + \sum_{i=1}^k \|x_i - \bar{x}_i\|\right)$$

attains its minimum at $(\bar{x}, \bar{x}_1, \ldots, \bar{x}_k)$.

We can view $g$ as a sum of two functions on $X^{k+1}$: a lower semicontinuous function

$$\varphi(x_1, \ldots, x_k) = \sum \text{Ind}_{S_i}(x_i)$$

and a convex continuous function

$$\psi(x, x_1, \ldots, x_k) = \sum \|x_i - (x + a_i)\| + \delta \left(\|x - \bar{x}\| + \sum \|x_i - \bar{x}_i\|\right).$$

Therefore (a) guarantees the existence of $w \in X, x_i \in S_i, w^*, w_1^*, \ldots, w_k^*, x_1^*, \ldots, x_k^*$ such that $\|w - \bar{x}\| < \delta, \|w_i - \bar{x}_i\| < \delta, \|x_i - \bar{x}_i\| < \delta$,

$$(w^*, w_1^*, \ldots, w_k^*) \in \partial \psi(w, w_1, \ldots, w_k), \quad (x_1^*, \ldots, x_k^*) \in \partial \varphi(x_1, \ldots, x_k)$$

and $\|(w^*, w_1^*, \ldots, w_k^*) - (0, x_1^*, \ldots, x_k^*)\| < \delta$.

Applying (S5b) to the subdifferential of $\varphi$, we conclude that $x_i^* \in N(S_i, x_i)$. On the other hand, by (S3) there are $u_i^* \in \partial \|\|(x_i - (x + a_i))\|$ such that $\max_i \|u_i^*\| = 1$ (since $x_i - (x + a_i) \neq 0$ for at least one $i$), $w_i^* \in u_i^* + 2\delta B$, $i = 1, \ldots, k$, and $-w^* \in \sum u_i^* + 2\delta B$. Compare this with the inequality at the end of the previous paragraph, we see that $\|\sum u_i^*\| < \delta$ and $\|u_i^* + x_i^*\| < 2\delta$. Thus

$$\|x_1^* + \cdots + x_k^*\| < (2k + 1)\delta$$

and it remains to take $\delta < (2k + 1)^{-1} \varepsilon$ to complete the proof of the implication.
Finally, let us assume that \( \partial \) is geometrically consistent and prove that \((f) \Rightarrow (a)\).

So let \( f_1 \) be a lower semicontinuous function and let \( f_2 \) be Lipschitz near \( \bar{x} \). Let \( \bar{x} \) be a local minimum of \( f = f_1 + f_2 \), and let an \( \varepsilon > 0 \) be given. Take a positive \( \delta < (1 + K)^{-1}\varepsilon \), where \( K \) is the Lipschitz constant of \( f_2 \) in a neighborhood of \( \bar{x} \) and find an \( r > 0 \) such that \( f(x) \geq f(\bar{x}) - \delta/2 \) if \( \|x - \bar{x}\| \leq r \). Consider the sets

\[
S_1 = \text{epi } f_1, \quad S_2 = \{(x, \alpha) \in X \times \mathbb{R} : \alpha \leq f(\bar{x}) - f_2(x)\}.
\]

Then \( (\bar{x}, f_1(\bar{x})) \in S_1 \cap S_2 \) but \( S_1 \cap (S_2 - (0, \delta)) \cap (B(\bar{x}, r) \times \mathbb{R}) = \emptyset \). Indeed if \( (x, \alpha) \) belongs to the intersection, then

\[
\alpha \geq f_1(x) \geq f(\bar{x}) - f_2(x) - \frac{\delta}{2} > f_1(x) \geq f(\bar{x}) - f_2(x) - \delta \geq \alpha.
\]

Thus, \((\bar{x}, f_1(\bar{x}))\) is an extremal point of \((S_1, S_2)\).

By the extremal principle there are \( (x_i, \alpha_i) \in S_i \) and \( (x_i^*, \beta_i) \in N(S_i, (x_i, \alpha_i)) \) such that \( \|x_i - \bar{x}\| < \delta \), \( |\alpha_i - f_1(\bar{x})| < \delta \) and

\[
\max\{\|x_i^*\| + |\beta_1|, \|x_2^*\| + |\beta_2|\} \geq 1 \quad \text{and} \quad \|x_1^* + x_2^*\| + |\beta_1 + \beta_2| < \delta. \quad (3.8)
\]

Note that \( (u, \gamma) \in S_2 \) if and only if \( (u, f(\bar{x}) - \gamma) \in \text{epi } f_2 \). Thus, as \( (x_2, \alpha_2) \in S_2 \), we have \( f_2(x) \leq \alpha := f(\bar{x}) - \alpha_2 \) and if we consider the operator \( A \) from \( X \times \mathbb{R}^2 \) into itself, \( A(x, \alpha) = (x, -\alpha) \), then

\[
\text{Ind}_{S_2}(x, \gamma) = \text{Ind}_{\text{epi } f_2}(A(x, \gamma)) + (0, f(\bar{x})).
\]

Thus, by \((S5a)\), \( (x_2^*, -\beta_2) \in N_{\text{epi } f_2}(x_2, \alpha) \). As \( (x_i^*, \beta_i) \neq 0 \) and \( f_2 \) is Lipschitz near \( \bar{x} \), it follows from geometric consistence of \( \partial \) that \( \beta_2 > 0 \). This means that \( \alpha = f(x_2) \) and \( u_2^* = x_2^*/|\beta_2| \) belongs to \( \partial f_2(x_2) \). We have, by \((S6)\), \( \|u_2^*\| \leq K \) which together with the first inequality in \((3.8)\) gives \( \beta_2 \geq (K + 1)^{-1} \). It follows from the last inequality in \((3.8)\) that \( \beta_1 < 0 \) and therefore by geometric consistence \( u_1^* = x_1^*/|\beta_1| \) belongs to \( \partial f_1(x_1) \). Now the simple calculation

\[
\|u_1^* + u_2^*\| \leq \frac{\|x_1^* + x_2^*\|}{|\beta_1|} + \frac{\|x_2^*\|}{|\beta_1|} \frac{\beta_1 + \beta_2}{|\beta_2|} = \frac{1 - \varepsilon}{K + 1},
\]

gives \( \|u_1^* + u_2^*\| \leq (1 - \varepsilon)^{-1}\varepsilon. \)

\[\Box\]

**Remark.** It follows from the proof that all quantities \( n \|x_{1n} - x_{jn}\|^2 \) are uniformly bounded. Therefore, we can include in the definition of the fuzzy minimization rule also the requirement that

\[
\text{diam}\{x_1, \ldots, x_k\} \cdot \max\{1, \|x_1^*\|, \ldots, \|x_k^*\|\} < \varepsilon. \quad (3.9)
\]
4 Tightness and weak fuzzy calculus

In this section we study the property of tightness which roughly speaking means that any element of the subdifferential at a point provides good lower approximation of the function along finite dimensional linear manifolds passing through some arbitrarily close points. Combined with trustworthiness, tightness provides for a mechanism to get estimates (up to arbitrary weak*-neighborhood of the origin) of subdifferentials of various functions obtained as results of one or another operation. We shall mainly concentrate on two basic operations: summation and minimization with respect to one of the variables. It is well known that other operations can be reduced to these two. We shall briefly discuss that at the end of the paper, prior to comments.

Another principal result of this section is the metric projection theorem that describes the connection between the subdifferentials of the distance function at outside points and normal cones to the set.

Definition 4.1. We shall say that a subdifferential is tight if for any \( x^* \in \partial f(x) \), any \( \varepsilon > 0 \) and any finite dimensional subspace \( L \subset X \) there exists a \( w \) such that \( \| w - x \| < \varepsilon, |f(w) - f(x)| < \varepsilon \) and the function

\[
\varphi(h) = f(w + h) + \varepsilon\|h\| - \langle x^*, h \rangle
\]

attains a local minimum on \( L \) at zero.

Tightness is a property of a subdifferential itself. Majority of subdifferentials used in variational analysis have this property; some (e.g. the generalized gradient) do not. We also note that the definition of tightness contains the same amount of fuzziness as the definition of trustworthiness: it suggests that we do not have an a priori information about specific location of \( w \) beyond the fact that it can be arbitrarily close to \( x \).

We start by looking at the subdifferential of a sum of finitely many functions.

Theorem 4.2. Let \( X \) and \( Y \) be Banach spaces, and let \( \partial \) be a tight subdifferential which is trusted on a class of Banach spaces containing \( X \). Assume that the functions \( f_1, \ldots, f_k \) and \( x^* \in \partial f(x) \) are finite and lower semicontinuous in a neighborhood of a certain \( \overline{x} \in X \). Set \( f = f_1 + \cdots + f_k \) and assume that \( x^* \in \partial f(\overline{x}) \). Then for any \( \varepsilon > 0 \) and any weak*-neighborhood \( V \) of the origin in \( X^* \) there are pairs \( (x_i, x_i^*) \), \( i = 1, \ldots, k \), such that

\[
\|x_i - \overline{x}\| < \varepsilon, \quad |f_i(x_i) - f_i(\overline{x})| < \varepsilon, \quad x_i^* \in \partial f_i(x_i), \quad x^* - (x_1^* + \cdots + x_k^*) \in V.
\]

(4.1)
Proof. Take a finite dimensional subspace $L \subset X$ and, if necessary, a smaller $\varepsilon$ to guarantee that $L^\perp + \varepsilon B_{X^*} \subset V$. As $f_i$ are l.s.c., we can choose $\delta > 0$ such that $|f(x) - f_i(x)| < \delta$ would imply $|f_i(x) - f_i(\bar{x})| < \varepsilon/2$ for all $i = 1, \ldots, k$. As $\partial$ is tight, there is a $w$ such that $\|w - \bar{x}\| < \varepsilon/2$, $|f(w) - f(\bar{x})| < \delta$ and the function

$$x \mapsto f(x) - \langle x^*, x - \bar{x} \rangle + \delta\|x - \bar{x}\|$$

attains a local minimum on $w + L$ at $w$, that is,

$$f_1(x) + \cdots + f_k(x) - \langle x^*, x - w \rangle + \delta\|x - w\| + \text{Ind}_L(x - w) \geq f_1(w) + \cdots + f_k(w).$$

As $L$ is finite dimensional, the function $\delta\|\cdot\| + \text{Ind}_L$ has compact sublevel sets, so this is a robust minimum by Proposition 3.2 and we may use the fact that $\partial$ is trusted on $X$ and Theorem 3.14. It follows that there are a $x_i \in \partial f_i(x_i)$ and $u^* \in L^\perp$ such that $\|x^*_i + \cdots + x - x^* - u^*\| \leq \varepsilon$ which implies the inclusion in (4.1).

Theorem 4.3. Assume that $\partial$ is a tight subdifferential which is trusted on $X \times Y$. Let $\varphi$ be a lower semicontinuous function on $X \times Y$. Set $f(x) = \inf_y \varphi(x, y)$. Let finally $\bar{x}^* \in \partial f(\bar{x})$. Then for any $\varepsilon > 0$ and any weak*-neighborhood $V$ of the origin in $X^*$ there is a pair $(x, y) \in \text{Graph } F$ such that $x$ is $\varepsilon$-close to $\bar{x}$,

$$\varphi(x, y) < f(\bar{x}) + \varepsilon \quad \text{and} \quad (\bar{x}^*, 0) \in \partial \varphi(x, y) + V \times (\varepsilon B_{Y^*}).$$

Proof. Fix as above $\varepsilon > 0$ and a finite dimensional $L \subset X$. Let $\delta \in (0, \varepsilon/2)$ and $w$ be such that

$$\|w - \bar{x}\| < \varepsilon/2, \quad |f(w) - f(\bar{x})| < \varepsilon/2$$

and $f - \langle \bar{x}^*, \cdot - w \rangle + (\varepsilon/2)\|\cdot - w\|$ attains at $w$ its minimum on the intersection of $w + L$ with the $\delta$-ball around $w$. Choose next a $v$ such that $\varphi(w, v) \leq f(v) + \delta^2$. Applying Ekeland’s principle, find a pair $(\bar{x}, \bar{y})$ such that

$$\|\bar{x} - w\| < \delta, \quad \|\bar{y} - v\| < \delta, \quad \varphi(\bar{x}, \bar{y}) < f(v) + \delta$$

and the function

$$g(x, y) = \varphi(x, y) - \langle \bar{x}^*, x \rangle + (\varepsilon/2)\|x - w\| + \delta(\|x - \bar{x}\| + \|y - \bar{y}\|) + \text{Ind}_L(x - w)$$

attains local minimum at $(\bar{x}, \bar{y})$. (Of course, if there exists a $v$ such that $f(w) = \varphi(w, v)$, we do not need to use Ekeland’s principle and can take $(\bar{x}, \bar{y}) = (w, v)$.) Again, this is a robust minimum as $L$ is finite dimensional and the application of trustworthiness (through the local fuzzy minimization principle) completes the proof (as in the prove of the previous theorem).
We conclude this section with an important result that establishes a connection between subgradients of the distance function to a set at a point \( x \) not belonging to the set and normals to the set at points almost realizing the distance to the set from \( x \).

**Theorem 4.4** (Approximate projection theorem). Let \( \partial \) be a tight subdifferential which is trusted on a class of spaces containing \( X \). Let \( S \subset X \) be a closed set, and let \( u^* \in \partial d_S(u) \). Then for any \( \varepsilon > 0 \) and any weak*-neighborhood \( V \) of the origin in \( X^* \) there are \( x \in S \) and \( x^* \in N(S, x) \) such that \( \|x - u\| < d_S(u) + \varepsilon \), \( x^* - u^* \in V \) and \( \|x^* - u^*\| \leq 3 \).

**Proof.** Take a finite dimensional subspace \( L \subset X \) and a (smaller if necessary) \( \varepsilon > 0 \) to make sure that \( L^\perp + \varepsilon B \subset V \). By tightness there is a \( w \) arbitrarily close to \( u \) such that

\[
d_S(w + h) - \langle u^*, h \rangle + \frac{\varepsilon}{4}\|h\| \geq d_S(w)
\]

if \( h \in L \) satisfies \( \|h\| < \delta \) for some \( \delta > 0 \). We can choose \( \delta \) arbitrarily small. A necessary estimate will be specified later in the proof. If we set

\[
g(x, h) = \|x - (w + h)\| - \langle u^*, h \rangle + \frac{\varepsilon}{4}\|h\| + Kd_L(h)
\]

with \( K > 2 + (\varepsilon/4) \), the inequality implies by Proposition 3.3 that

\[
g(x, h) \geq d_S(w) \quad \text{for all } x \in S \text{ and } h \in \delta B.
\]

Choose an \( \overline{x} \in S \) such that

\[
g(\overline{x}, 0) = \|\overline{x} - w\| < d_S(w) + (\varepsilon\delta/8).
\]

By Ekeland’s principle there are \( \overline{w} \in S \) and \( \overline{h} \in X \) such that

\[
\|\overline{w} - \overline{x}\| < \frac{\delta}{2}, \quad \|\overline{h}\| < \frac{\delta}{2}, \quad g(\overline{w}, e) \leq g(\overline{x}, 0)
\]

and

\[
g(x, h) + \frac{\varepsilon}{4}(\|x - \overline{w}\| + \|h - \overline{h}\|) \geq g(\overline{w}, \overline{h}), \quad \forall x \in S, \|h\| < \delta/2 \quad (4.3)
\]

or, equivalently,

\[
g(x, h) + \frac{\varepsilon}{4}(\|x - \overline{w}\| + \|h\|) + \text{Ind}_S(x) \geq g(\overline{w}, \overline{h})
\]

if \( x \in X \), \( \|h\| < \delta/2 \).
The function on the left is the sum of a lower semicontinuous function \( \text{Ind}_S \) and a convex continuous function
\[
(x, h) \mapsto \varphi(x, h) = \|x - (w + h)\| - \langle u^*, h \rangle + \frac{\epsilon}{4}\|h\| + K d_L(h)
\]
\[
+ \frac{\epsilon}{4}(\|x - w\| + \|h - \bar{h}\|).
\]

As \( \partial \) is trusted on \( X \), for any \( \eta > 0 \), there are \( x \in S, x^* \in N(S, x) \) and a pair \( (w^*, h^*) \) in the subdifferential of \( \varphi \) at a certain \( (x', h') \) such that \( \|x^* + w^*\| < \eta \) and \( \|h^*\| < \eta \). On the other hand, by the standard rules of convex analysis, there is a \( z^* \) in the subdifferential of \( \| \cdot \| \) at \( x' - (w - h') \) such that
\[
\|w^* - z^*\| \leq \varepsilon/2, \quad h^* \in -z^* - u^* + L^\perp \cap KB + (\varepsilon/2)B,
\]
and we see that
\[
\|x^* + z^*\| \leq \frac{\varepsilon}{2} + \eta, \quad z^* + u^* \in L^\perp \left(\frac{\varepsilon}{2} + \eta\right)B, \quad \|z^* + u^*\| \leq K + \frac{\varepsilon}{2} + \eta.
\]
Taking \( \eta < \varepsilon/2 \), we conclude the proof, provided \( \varepsilon \) has been chosen sufficiently small and keeping in mind that \( w \) can be arbitrarily close to \( u \).

\[\square\]

5 Sequential and topological closures

In this and in the following sections we shall consider some operations that allow to get new subdifferentials from given ones. We shall assume all functions lower semicontinuous, at least near the points of an interest.

Let \( \partial \) be a subdifferential. The set
\[
[\partial f] = \{(x, \alpha, x^*) \in X \times \mathbb{R} \times X^* : \alpha = f(x), x^* \in \partial f(x)\}
\]
is called the \( \partial \)-jet of \( f \). Dealing with jets, we shall always consider \( X \times \mathbb{R} \times X^* \) together with the product of the metric topology in \( X \), standard topology in \( \mathbb{R} \) and the weak* topology in \( X^* \).

We denote by \( \text{cl}[\partial f] \) the closure of \( [\partial f] \) in the product of the norm topology of \( X \times \mathbb{R} \) and the weak* topology of \( X^* \) and by \( \text{cl}_{s}[\partial f] \) the sequential closure of \( [\partial f] \), that is, \( \text{cl}_{s}[\partial f] \) is the collection of triples \( (x, \alpha, x^*) \) with the property that for each of them there is a sequence \( (x_n, x_n^*) \) such that \( x_n \to x, f(x_n) \to \alpha \) and \( x_n^* \) weak*-converge to \( x^* \).

**Definition 5.1.** Given a subdifferential \( \partial \). The closure of \( \partial \) associates with every \( (X, f, x) \) the set \( \overline{\partial f}(x) = \{x^* : (x, f(x), x^*) \in \text{cl}[\partial f]\} \) if \( x \in \text{dom} f \) and \( \emptyset \) otherwise. Replacing in this definition the topological closure of \( [\partial f] \) by its sequential closure \( \text{cl}_{s}[\partial f] \) we get the definition of the limiting \( \partial \)-subdifferential \( \partial_L \). We also denote by \( \partial^*_L f(x) \) the weak*-closure of the set \( \partial_L f(x) \).
Clearly, we always have
\[ \partial f(x) \subset \partial_L f(x) \subset \partial_L^* f(x) \subset \overline{\partial} f(x). \]  
(5.1)

**Theorem 5.2.** Let \( \partial \) be a subdifferential. Then \( \partial_L, \partial_L^* \) and \( \overline{\partial} \) are subdifferentials on the class of lower semicontinuous functions. Moreover, the properties of being trustworthy or tight are inherited by either of the three.

**Proof.** (S1) is a part of the definition, (S2) is also obvious from the definition, (S3) follows from the stability of the graph of a convex subdifferential with respect to the convergence of jets: the inequality
\[ \alpha - f(x) \geq \langle u^*, u - x \rangle \]
is not affected by the norm convergence of \((u, \alpha)\) and weak*-convergence (no matter sequential or topological) of \(u^*\). (S4) is immediate from (5.1). (S5b) and (S6) are also obvious, the latter because the norm is a weak*-l.s.c. function.

It remains to verify (S5a). We shall do this only for the limiting subdifferential. The passage from \( \partial_L \) to \( \partial_L^* \) is trivial and the proof for \( \overline{\partial} \) follows the same lines with sequences replaced by nets. Let \( A : X \to Y \) be a bounded linear operator with \( \text{Im} A = Y \), let \( \ell \in X^*, v \in Y, \beta \in \mathbb{R}, \) and let \( g(y) \) be a function on \( Y \). Set \( f(x) = \lambda g(Ax + v) + \langle \ell, x \rangle + \beta \). As \( A \) is onto, there is a positive \( K \) such that \( \|y^*\| \leq K \|A^*y^*\| \) for all \( y^* \) and for any \( y \in Y \) there is an \( x \in X \) such that \( Ax = y \) and \( \|x\| \leq K \|y\| \) (by the Banach open mapping theorem).

If \( y^* \in \partial g(Ax + v), y_n \to Ax, g(y_n + v) \to g(Ax + v) \) and \( y_n^* \in \partial g(y_n + v) \) weak*-converge to \( y^* \), then taking \( x_n \) such that
\[ Ax_n = y_n \quad \text{and} \quad \|x_n - x\| \leq K \|y_n - Ax\| \]
we see that \( f(x_n) \to f(x) \) and get from (S5a) that \( x_n^* = A^*y_n^* + \ell \in \partial f(x_n) \). As the adjoint operator is weak*-to-weak*-continuous, it follows that \( x_n^* \) weak*-converge to \( x^* = A^*y^* + \ell \). Thus \( \partial_L g(Ax + v) + \ell \subset \partial_L f(x) \).

Conversely, let \( x^* \in \partial_L f(x) \) and \( x_n^* \in \partial f(x_n) \) with \( x_n \to x, f(x_n) \to f(x) \) and \( x_n^* \) weak*-converging to \( x^* \). Set \( y_n = Ax_n + v \). By property (S5a) there are \( y_n^* \in \partial g(y_n) \) such that \( x_n^* - \ell = A^*y_n^* \). Take a \( z \in Y \) and a \( u \in X \) such that \( Au = z \) and \( \|u\| \leq K\|z\| \). Then
\[ \langle y_{n+m}^* - y_n^*, z \rangle = \langle x_{n+m}^* - x_n^*, u \rangle. \]

It follows that \( \langle y_n^*, z \rangle \) is a Cauchy sequence converging to \( \langle x^* - \ell, u \rangle \). Thus the limit \( \varphi(z) = \lim_{n \to \infty} \langle y_n^*, z \rangle \) exists for all \( z \in Y \). Clearly, \( \varphi \) is a linear function and we have \( |\varphi(z)| \leq |\langle x^* - \ell, u \rangle| \leq (K + \|\ell\|)\|z\| \) which immediately implies that \( \varphi(z) = \langle y^*, z \rangle \) for some \( y^* \) such that \( A^*y^* = x^* - \ell \). This proves the opposite inclusion and (S5a).
The fact that $\partial_L$, $\partial^*_L$ and $\overline{\partial}$ are trusted on $X$ if so is $\partial$ again follows immediately from (5.1). To check that the same is true for tightness, take an $x^* \in \partial_L f(x)$, an $\varepsilon > 0$ and a finite dimensional subspace $L \subset X$. By definition, there is a sequence $(x_n, x^*_n)$ such that $x_n \to x$, $f(x_n) \to f(x)$, $x^*_n \in \partial f(x_n)$ and
\[ |\langle x^*_n - x^*, h \rangle| \leq \lambda_n \|h\| \quad \text{for all } h \in L, \]
where $\lambda_n \to 0$. Take an index $n$ such that $\|x_n - x\| < \varepsilon/2$, $|f(x_n) - f(x)| < \varepsilon/2$, $\lambda_n < \varepsilon/2$ and find (as $\partial$ is tight) a $u$ with $\|x_n - u\| < \varepsilon/2$ such that the function $h \mapsto f(u + h) - \langle x^*_n, h \rangle + (\varepsilon/2)\|h\|$ attains at zero a local minimum on $L$. But then $\|u - x\| < \varepsilon$ and zero is a local minimizer of $h \mapsto f(u + h) - \langle x^*, h \rangle + \varepsilon\|h\|$. Thus $\partial_L L$ is tight. Again the passage to $\partial^*_L$ is trivial and replacing a sequence by a net in the argument shows tightness of $\overline{\partial}$. \hfill \Box

The fact deserving a mention is that a stronger form of the characteristic trustworthiness property holds for the limiting subdifferential (and also for $\partial^*_L$ and $\overline{\partial}$) if the unit ball in $X^*$ is weak*-sequentially compact.

**Proposition 5.3.** Let $\partial$ be trusted on a class of Banach spaces containing $X$. Assume that the unit ball in $X^*$ is weak*-sequentially compact. Let finally the functions $f$ and $g$ on $X$ be lower semicontinuous and one of them satisfy the Lipschitz condition near a point $x$ at which the sum $f + g$ attains a local minimum. Then
\[ 0 \in \partial_L f(x) + \partial_L g(x). \]
In particular, $\partial_L f(x) \neq \emptyset$ if $f$ is Lipschitz near $x$.

**Proof.** As $\partial$ is trusted, there is a sequence of quadruples $(x_n, u_n, x^*_n, u^*_n)$ such that
\[ \|x_n - x\| < \frac{1}{n}, \quad \|x - u_n\| < \frac{1}{n}, \quad x^*_n \in \partial f(x_n), \quad u^*_n \in \partial g(u_n), \quad \|x^*_n + u^*_n\| < \frac{1}{n}. \]
As one of the function, say $f$, is Lipschitz near $x$, it follows that the sequence $(x^*_n)$ is norm bounded and contains a weak*-converging subsequence. This proves the first statement.

To prove the second statement, take $g$ being an indicator of $\{x\}$ (which is a convex function with $\partial g(x) = X^*$) and apply the first part of the proposition. \hfill \Box

It is to be said that for a non-Lipschitz function the closure of a subdifferential can be unacceptably large (example of that sort for the closure of the Dini–Hadamard subdifferential in a separable space called approximate $A$-subdifferential in [29,31]) was given by Treiman in [65]). But for Lipschitz functions all three subdifferentials are workable instruments. This is underscored by what may be viewed as a consequence of this theorem.
Theorem 5.4. Let $X$ be a Banach space, and let $\partial$ and $\partial'$ be two trusted tight subdifferentials on $X$. Let finally $f$ be a function on $X$ defined and Lipschitz near $x$. Then:

(a) $\partial f(x) = \partial' f(x)$,

(b) if the unit ball in $X^*$ is weak*-sequentially compact, then $\partial_L^* f(x) = \partial f(x)$.

Proof. Let $x^* \in \partial f(x)$. Fix an $\epsilon > 0$ and a finite dimensional subspace $L \subset X$, and let $V(\epsilon, L) = L^\perp + \epsilon B$. The sets $V(\epsilon, L)$ form a basis of weak*-neighborhoods around the origin in $X^*$. By definition for any $\delta > 0$ there is a pair $(w, w^*)$ such that $\|x - w\| < \delta/3$, $w^* \in \partial f(w)$ and $w^* - x^* \in V(\epsilon/3, L)$, that is, we have $|\langle w^* - x^*, h \rangle| < (\epsilon/3)\|h\|$ for all $h \in L$. (As $f$ is Lipschitz near $x$, we need not care about values of $f$.)

Let us prove (b) first. As follows from what has been said in the previous paragraph, there is a sequence of pairs $(w_n, w_n^*)$ such that $w_n \to x$, $w_n^* \in \partial f(w_n)$ and $x^* \in w_n^* + V(\epsilon, L)$. The sequence $(w_n^*)$ is norm bounded as $f$ is Lipschitz, hence it contains a weak*-converging subsequence. This means that $x^* \in \partial_L^* f(x) + V(\epsilon, L) \subset \partial_L^* f(x) + V(\epsilon, L)$. This is true for all $\epsilon$ and $L$, so as $\partial_L^* f(x)$ is a weak*-compact set (and as has been mentioned, $V(\epsilon, L)$ form a basis of weak*-neighborhoods of the origin), it follows that $x^* \in \partial_L^* f(x)$ and therefore $\partial f(x) \subset \partial_L^* f(x)$. The opposite inclusion is obvious.

Let us pass to the proof of (a). As $\partial$ is tight, there is a $u$ such that $\|u - w\| < \delta/3$ and the function $h \mapsto f(u + h) - \langle w^*, h \rangle + (\epsilon/3)\|h\| + \text{Ind}_L(h)$ attains a local minimum at zero. By trustworthiness there is a pair $(x, x^*)$ such that $\|x - u\| < \epsilon/3$, $x^* \in \partial f(x)$ and $x^* - w^* \in V(\epsilon/3, L)$, that is,

$$\|x - x^*\| < \epsilon \quad \text{and} \quad x^* \in \partial f(x) + V(2\epsilon/3, L).$$

If we let $\delta$ go to zero, we conclude that for any $\epsilon > 0$ and any finite dimensional $L$, $x^* \in \partial f(x) + V(\epsilon)$ and, as $x^*$ is an arbitrary element of $\partial f(x)$, that

$$\partial f(x) \subset \partial f(x) + V(\epsilon).$$

Now the inclusion $\partial f(x) \subset \partial f(x)$ follows as in the proof of (b). To complete the proof of (a), we have to change the roles of $\partial$ and $\partial'$ in the above argument. 

Thus in every Banach space there may be at most one tight and trustworthy subdifferential on the class of Lipschitz functions whose graph is closed in the product of the norm and the weak*-topologies. Likewise in spaces with weak*-sequentially compact adjoint balls limiting versions of tight and trustworthy subdifferentials coincide up to the closure of their values and the latter must be equal to the values of the closure of any of the subdifferentials.
6 Metric modification

Definition 6.1. Let $\partial$ be a subdifferential on the class of locally Lipschitz functions. We shall say, following Penot, that $\partial$ is distance-homotone if the following property holds: if $A : X \to Y$ is a bounded linear operator onto, $S \subseteq Y$ is closed and $f$ is a locally Lipschitz function on $X$ such that $f(x) \geq d_S(Ax)$ for all $x$ of a neighborhood of an $\overline{x} \in S$ and $f(x) = 0$ on the intersection of $S$ with the neighborhood, then $\partial(d_S \circ A)(\overline{x}) \subseteq \partial f(\overline{x})$.

Clearly, such is any elementary subdifferential.

Definition 6.2 (Metric modification). Given a lower semicontinuous function $f$ on $X$ and an $x \in \text{dom } f$, denote by $\mathcal{L}_\partial(f, x)$ the collection of all weak*-limits of sequences $(x_n^*, \gamma_n)$ such that $(x_n^*, \gamma_n) \in \partial \text{epi } f(x_n, f(x_n))$ for some $x_n$ such that $(x_n, f(x_n)) \to (x, f(x))$:

$$\mathcal{L}_\partial(f, x) = \{ \lim (x_n^*, \gamma_n) : (x_n^*, \gamma_n) \in \partial \text{epi } f(x_n, f(x_n)), (x_n, f(x_n)) \to (x, f(x)) \}.$$ 

Set

$$\partial_{\text{mod}} f(x) = \left\{ x^* : (x^*, -1) \in \bigcup_{\lambda > 0} \lambda \mathcal{L}_\partial(f, x) \right\}.$$ 

We shall call $\partial_{\text{mod}}$ the metric modification of $\partial$.

We emphasize that the property of being distance homotone is a property of the subdifferential, not connected with any space. Therefore if it holds, it holds with any distance associated with an equivalent norm.

Proposition 6.3. If $\partial$ is distance-homotone, the definition of metric modification is correct in the sense that the result does not depend on the choice of an equivalent norm in $X \times \mathbb{R}$.

Proof. Let $\|(x, \alpha)\|$ and $\|(x, \alpha)\|'$ be two equivalent norms in $X \times \mathbb{R}$. Then there are $K \geq k > 0$ such that

$$kd_{\text{epi }}f(u, \alpha) \leq d'_{\text{epi }}f(u, \alpha) \leq Kd_{\text{epi }}f(u, \alpha)$$

for all $(x, \alpha)$. By the distance homotonicity

$$k\partial d_{\text{epi }}f(x, \alpha) \subset \partial d'_{\text{epi }}f(x, \alpha) \subset K\partial d_{\text{epi }}f(x, \alpha)$$

if $(x, \alpha) \in \text{epi } f$. It follows that $\partial_{\text{mod}}$ does not depend on the choice of the norm. \qed
Theorem 6.4. If $\partial$ is distance-homotone, then $\partial_{\text{mod}}$ is a subdifferential on the class of lower semi-continuous functions.

Proof. The properties (S1) and (S2) are immediate. To prove (S3) for $\partial_{\text{mod}}$, we note first that $d_{\text{epi}} f$ is a convex function if so is $f$ and the fact that the standard subdifferential of convex analysis is geometrically consistent on the class of the convex functions is well known. It is also well known that the subdifferential of a convex lower semicontinuous function is upper semicontinuous in the sense that $x^* \in \partial f(x)$ if $x^*$ is a weak*-limit point of a sequence $(x^n_*)$ such that $x_n \to x$, $f(x_n) \to f(x)$ and $x^n_* \in \partial f(x_n)$. Thus, $\partial_{\text{mod}}$ satisfies (S3) if so does $\partial$.

The verification of (S4) for $\partial_{\text{mod}}$ is also sufficiently easy. Let $f$ have a local minimum at $x$. Then an intersection of epi $f$ with a neighborhood of $(x, f(x))$ lies in $\{ (x, \alpha) : \alpha \geq f(x) \} = X \times [f(x), \infty)$, so that for all $x$ close to $x$ we have

$$d_{\text{epi}} f(x, \alpha) \geq (f(x) - \alpha)^+.$$ 

It follows that $g(x, \alpha) = d_{\text{epi}} f(x, \alpha) + \alpha - f(x) \geq 0$ for $(x, \alpha)$ close to $(x, f(x))$ and $g(x, f(x)) = 0$. Thus by (S4) and (S5) (recall that $\partial$ is a subdifferential)

$$(0, 0) \in \partial d_{\text{epi}} f(x, f(x)) + (0, 1),$$

that is, $(0, -1) \in \partial d_{\text{epi}} f(x, f(x)) \subset \mathcal{L}_\partial (f, x)$ and therefore $0 \in \partial_{\text{mod}} f(x)$.

The proof of property (S5) is the longest. First consider the case when $f$ and $g$ satisfy $f(x) = g(x + x_0) + \alpha_0$. In this case $d_{\text{epi}} f(x, \alpha) = d_{\text{epi}} g(x + x_0, \alpha - \alpha_0)$, so that by (S5) (applied to $\partial$) $\partial d_{\text{epi}} f(x, f(x)) = \partial d_{\text{epi}} g(x + x_0, g(x + x_0))$ and therefore $\mathcal{L}_\partial(f, x) = \mathcal{L}_\partial(g, x + x_0)$.

Next we consider $f = \lambda g$ with $\lambda > 0$. In this case, along with the given norm $\|(x, \alpha)\|$ in $X \times \mathbb{R}$, we consider also the norm $\|(x, \lambda \alpha)\| = \|(x, \alpha)\|^\lambda$. Then $d_{\text{epi}} f(x, \alpha) = d_{\text{epi}} ^\lambda g(x, \alpha / \lambda)$. Consider the operator $T : X \times \mathbb{R} \to Y \times \mathbb{R}$ defined by $T(x, \alpha) = (x, \alpha / \lambda)$ so that $d_{\text{epi}} f(x, \alpha) = d_{\text{epi}} ^\lambda g(T(x, \alpha))$. Clearly, $T$ is an operator onto, so that

$$\partial d_{\text{epi}} f(x, f(x)) = T^* (\partial d_{\text{epi}} g(x, g(x))).$$

We have $T^*(x^*, \gamma) = (x^*, \gamma / \lambda)$. This is a one-to-one and weak*-continuous operator, so $\mathcal{L}_\partial(f, x) = T^* (\mathcal{L}_\partial ^\lambda (g, x))$ and therefore $(x^*, -1) \in \mathcal{L}_\partial(x, f(x))$ if and only if $(x^*/\lambda, -1) \in \mathcal{L}_\partial ^\lambda (g, x)$ which is the same as $(x^*/\lambda, -1) \in \lambda^{-1} \mathcal{L}_\partial ^\lambda (g, x)$. The latter by definition means that $(x^*/\lambda) \in \partial_{\text{mod}} g(x)$ or $x^* \in \lambda x \partial_{\text{mod}} g(x)$.

At the third step of verification of (S5a) we consider the following function:

$$f(x) = g(x) + \langle x^*, x \rangle.$$ 

Again, along with the given norm $\| \cdot \|$ in $X \times \mathbb{R}$, we consider another equivalent norm

$$\|(x, \alpha)\|' = \|(x, \alpha)\| + \|(0, \langle x^*, x \rangle)\|.$$
Clearly, \( \| (x, \alpha) \| \leq \| (x, \alpha') \|' \) for all \( (x, \alpha) \) and therefore

\[
d_{\text{epi}} f(x, \alpha) = \inf \{ \| (x - u, \alpha - \gamma) \| : \gamma \geq f(u) \} \\
= \inf \{ \| (x - u, \alpha - \langle x^*, u \rangle - \xi) \| : \xi \geq g(u) \} \\
= \inf \{ \| (x - u, \alpha - \langle x^*, x \rangle - \xi + \langle x^*, x - u \rangle) \| : \xi \geq g(u) \} \\
\leq \inf \{ \| (x - u, \alpha - \langle x^*, x \rangle - \xi) \|' : \xi \geq g(u) \} \\
= d'_{\text{epi}} g(x, \alpha - \langle x^*, x \rangle).
\] (6.1)

Consider the operator \( T : X \times \mathbb{R} \to X \times \mathbb{R} \) defined by \( T(x, \alpha) = (x, \alpha - \langle x^*, x \rangle) \). Then, as \( T(x, f(x)) = (x, g(x)) \) and \( \partial \) is distance-homotone, (6.1) implies that

\[
\partial d_{\text{epi}} f(x, f(x)) \subset T^* (\partial d'_{\text{epi}} g(x, g(x))).
\] (6.2)

Again it is easy to see that (6.2) carries over to the corresponding \( L^0 \) sets,

\[
L^0(f, x) \subset T^* (L^0'(g, x)),
\]

and therefore for a \( (u^*, -1) \in \lambda L^0'(f, x) \) we have a representation

\[
(u^*, -1) = T^* (v^*, -\gamma) \quad \text{for some} \quad (v^*, -\gamma) \in \lambda L^0'(g, x).
\]

We have \( T^* (v^*, \eta) = (v^* - \eta x^*, \eta) \), so \( \gamma = 1 \) and \( v^* = u^* + x^* \). This means that \( \partial_{\text{mod}} f(x) \subset \partial_{\text{mod}} g(x) + x^* \). Changing the roles of \( f \) and \( g \), we get the opposite inclusion.

It remains to consider the case \( f(x) = g(Ax) \) to finish with (S5a). Since \( A \) is onto, there is a \( K > 0 \) such that for any \( y \in Y \) we can find an \( x \) such that \( Ax = y \) and \( \|x\| \leq K \|y\| \). It will be convenient for us to work now with the standard sum norm \( \|(x, \alpha)\| = \|x\| + \|\alpha\| \) in \( X \) and the two norms \( \|(y, \alpha)\| = K \|y\| + \|\alpha\| \) and \( \|(y, \alpha)\|' = \|A\|^{-1} \|y\| + \|\alpha\| \) in \( Y \). Fix an \( x \in X \). We have

\[
d_{\text{epi}} f(x, \alpha) = \inf \{ \|x - u\| + |\alpha - \gamma| : \gamma \geq f(u) \} \\
\leq \inf \{ K \|Ax - v\| + |\alpha - \gamma| : \gamma \geq g(v) \} = d_{\text{epi}} g(Ax, \alpha)
\]

and

\[
d_{\text{epi}} f(x, \alpha) = \inf \{ \|x - u\| + |\alpha - \gamma| : \gamma \geq f(u) \} \\
\geq \inf \{ \|A\|^{-1} \|Ax - v\| + |\alpha - \gamma| : \gamma \geq g(v) \} = d'_{\text{epi}} g(Ax, \alpha).
\]

Now we set \( T(x, \alpha) = (Ax, \alpha) \) and using the same arguments as above find out that

\[
T^* (L^0'(g, Ax)) \subset L^0(f, x) \subset T^* (L^0(g, Ax))
\]

and consequently (as \( T^* (v^*, \xi) = (A^* v^*, \xi) \)) that \( \partial_{\text{mod}} f(x) = \partial_{\text{mod}} g(Ax) \).
Let us check (S5b). Let $X$ and $Y$ be Banach spaces, and let $g$ and $h$ be lower semicontinuous functions on $X$ and $Y$ respectively. Set $f(x, y) = g(x) + h(y)$. An easy calculation gives (if e.g. we take the standard sum norms in all spaces)
\[
\varphi((x, \alpha), (y, \gamma)) = d_{\text{epi}} f((x, y), \alpha + \gamma) \leq d_{\text{epi}} g(x, \alpha) + d_{\text{epi}} h(y, \gamma).
\] (6.3)
Consider the operator $T : (X \times \mathbb{R}) \times (Y \times \mathbb{R})$ defined by
\[
T(x, \alpha, y, \gamma) = (x, y, \alpha + \gamma).
\]
Then the left side of the inequality can be rewritten as $d_{\text{epi}} f = \varphi \circ T$. Therefore
\[
\partial d_{\text{epi}} f((x, y), f(x, y)) = T^*(\partial \varphi((x, g(x)), (y, h(y)))) \\
\subseteq T^*(\partial d_{\text{epi}} g(x, g(x)) \times \partial d_{\text{epi}} h(y, h(y))).
\]
We have $T^*((x, y), \xi) = ((x^*, \xi), (y^*, \xi))$. The desired conclusion now follows from the already standard arguments.

The question about trustworthiness and tightness of the modified subdifferential should be verified in each specific case.

**Part 2** Special theory

7 Elementary subdifferentials

In this section we consider a family of elementary subdifferentials whose definitions combine in a way some basic features of derivatives and convex subdifferentials.

7.1 Bornologies and derivatives

Recall that a bornology in a normed space $X$ is a collection $\beta(X)$ of bounded subsets of $X$ such that the union of elements of $\beta$ is the whole of $X$ and for any two elements $S_1$, $S_2$ of $\beta$ there is an $S \in \beta(X)$ containing both $S_1$ and $S_2$. A bornology is convex if all its elements are convex sets and symmetric if $S = -S$ for every $S \in \beta$. A subset $\mathcal{B}$ of $\beta(X)$ is a basis of $\beta(X)$ if any element of $\beta(X)$ is contained in some element of $\mathcal{B}$.

A bornology in the category of Banach spaces is a correspondence $\beta$ which associates with every Banach space $X$ a bornology $\beta(X)$ in $X$ in such a way that the set $\{A(S) : S \in \beta(X)\}$ is a basis of $\beta(Y)$ whenever $A : X \to Y$ is a bounded

---

2 Strictly speaking, what we have just defined is just a net of bounded sets that should be considered a basis of a bornology rather then a bornology itself. For a definition of a “full” bornology see e.g. H. H. Shaeffer, *Topological Vector Spaces*, Springer-Verlag, New York, 1970.
linear operator onto $Y$. In what follows we shall deal only with convex symmetric bornologies in the category of Banach spaces. But for the sake of simplicity we shall somewhat abuse the language and speak just of bornologies, without an indication of whether it is a specific bornology in a given Banach space or a bornology in the category of Banach space, and often write $\beta$ instead of $\beta(X)$ when $X$ is clear from the context.

The two most important bornologies are formed by all symmetric convex bounded sets (bounded, or Fréchet bornology) and symmetric convex compact sets (compact, or Hadamard bornology). We also mention the weakest finite bornology consisting of convex hulls of all symmetric finite sets and the weak compact bornology which is intermediate between the Hadamard and Fréchet bornologies and consists of all weakly compact symmetric subsets. For $X = \mathbb{R}^n$ all four mentioned bornologies coincide and in a reflexive space the weak compact and the Fréchet bornologies are identical. For any bornology $\beta$ there is a canonical way to define a basis in $\beta(X \times Y)$, given two bases $\mathcal{B}(X)$ of $\beta(X)$ and $\mathcal{B}(Y)$ of $\beta(Y)$, we take all sets $S \times Q$, $S \in \mathcal{B}(X)$, $Q \in \mathcal{B}(Y)$. In particular, if $Y = \mathbb{R}^n$, we get a basis formed by all sets $S \times \lambda B$ ($\lambda > 0$ and $B$ being the unit ball in $\mathbb{R}^n$).

Every bornology generates a certain concept of derivative (see e.g. [3]): given a mapping $F$ from a neighborhood of an $x \in X$ into another space $Y$ and a bornology $\beta$ in $X$, a linear bounded operator $A : X \to Y$ is the $\beta$-derivative of $F$ at $x$ if for any $S \in \beta$

$$
\lim_{t \to 0} t^{-1} \sup_{h \in S} \|F(x + th) - F(x) - tAh\| = 0.
$$

The derivative, if it exists, is uniquely defined. Let us denote it for a time being by $F'_\beta(x)$. It is usually said that $F$ is $\beta$-differentiable if $F'_\beta$ exists. If $\beta$ is the Fréchet bornology, we speak about Fréchet differentiability, and in case of the finite bornology – about Gâteaux differentiability. The other two mentioned bornologies correspond to Hadamard and weak Hadamard differentiability.

Given two bornologies $\beta$ and $\gamma$, it is said that $\beta$ is stronger and $\gamma$ is weaker if a certain basis of $\gamma(X)$ is a subset of $\beta(X)$. As follows from the definition, the Fréchet bornology is the strongest and the finite bornology is the weakest. Clearly $\beta$-differentiability implies $\gamma$-differentiability if $\beta$ is stronger then $\gamma$. Thus a Fréchet differentiable function is $\beta$-differentiable for any $\beta$ and any $\beta$-differentiability implies Gâteaux differentiability and, by uniqueness, the $\beta$-derivative coincides with the Gâteaux derivative. Therefore it is possible to use just the symbol $F'(x)$ and to specify suitable bornologies if necessary.

Of course the $\beta$-differentiation is a linear operation in the sense that

$$(\lambda F + \mu G)' = \lambda F'(x) + \mu G'(x)$$

for any $\lambda$ and $\mu$. The situation is a little bit more complicated with compositions. If
$F : X \to Y$ and $G : Y \to Z$ with $F$ $\beta$-differentiable at $x$ and $G$ $\beta$-differentiable at $y = F(x)$, then $(G \circ F)'(x) = G'(y) \circ F'(x)$ if $\beta$ is the Fréchet bornology, but the equality may fail for other bornologies.

### 7.2 $\beta$-subdifferentials

With every bornology $\beta$ we can associate two subdifferentials, one following the idea of linear approximation borrowed from the differential calculus, but this time one sided – from below, and the other based on the idea of support borrowed from convex analysis, but this time by a $\beta$-differentiable function rather than by a linear functional.

So let $X$ be a normed space, let $f$ be a function on $X$ finite at $x$, and let a certain bornology $\beta$ be fixed.

**Definition 7.1 (Canonical $\beta$-subdifferential).** We say that $x^*$ belongs to the canonical $\beta$-subdifferential of $f$ at $x$ if for any $S \in \beta$

$$\liminf_{t \to +0} \inf_{h \in S} (f(x + th) - f(x) - t \langle x^*, h \rangle \geq 0. \quad (7.1)$$

We denote the collection of all such $x^*$ by $\partial^C_{\beta} f(x)$.

With a slight deviation of this notation and terminology, we shall denote the canonical subdifferentials corresponding to the Fréchet and Hadamard bornologies just by $\partial_F f(x)$ and $\partial_{DH} f(x)$ respectively and call them the (canonical) Fréchet and Dini–Hadamard subdifferentials.

As elements of $\beta$ are bounded, $x^* \in \partial^C_{\beta} f(x)$ if and only if for any $\varepsilon > 0$ and any $S \in \beta$ there is a $\delta > 0$ such that

$$f(x + h) - f(x) - \langle x^*, h \rangle + \varepsilon\|h\| \geq 0 \quad (7.2)$$

for all $h \in S$ with $\|h\| < \delta$. It follows that

**Proposition 7.2.** Any canonical $\beta$-subdifferential is tight.

Indeed, the intersection of any finite dimensional subspace with a small ball around zero must lie inside of a certain element in any bornology.

For lower semicontinuous functions the definition can be further specialized.

**Proposition 7.3.** If $f$ is lower semicontinuous at $x$, then $x^* \in \partial^C_{\beta} f(x)$ if and only if for any $\varepsilon > 0$ and any $S \in \beta$ there is a $\lambda > 0$ such that

$$f(x + h) - f(x) - \langle x^*, h \rangle + \varepsilon\|h\| \geq 0 \quad (7.3)$$

for all $h \in S$ with $\|h\| < \lambda$ and $|f(x + h) - f(x)| < \lambda$. 

---

---
Proof. Clearly, inequality (7.3) holds if \( x^* \in \partial^c_\beta f(x) \). So suppose (7.3) holds for some \( x^* \) and an \( \varepsilon > 0 \) and \( S \in \beta \) be given. Take a positive \( \delta \) to make sure that \( f(x + h) > f(x) - \lambda \) if \( \|h\| < \delta \) (which is possible as \( f \) is l.s.c.) and \( \delta\|x^*\| < \lambda \). Take an \( h \in S, \|h\| < \delta \). If \( |f(x + h) - f(x)| < \lambda \), then (7.2) follows from (7.3). If on the other hand \( f(x + h) - f(x) \geq \lambda \), then

\[
f(x + h) - f(x) - \langle x^*, h \rangle + \varepsilon\|h\| \geq \lambda - \delta\|x^*\| + \varepsilon\|h\| \geq 0,
\]

which means that (7.2) holds for such \( h \) as well. \( \square \)

**Definition 7.4** (Viscosity \( \beta \)-subdifferential). We say that \( x^* \) belongs to the viscosity \( \beta \)-subdifferential of \( f \) at \( x \) if there is a locally Lipschitz \( \beta \)-differentiable function \( \varphi \) such that \( f - \varphi \) attains a local minimum at \( x \) and \( \varphi'(x) = x^* \). The collection of all such \( x^* \) we denote \( \partial^v_\beta f(x) \).

**Proposition 7.5.** Both \( \partial^c_\beta \) and \( \partial^v_\beta \) are elementary subdifferentials. We always have

\[
\partial_\beta f(x) \subset \partial_\gamma f(x)
\]

if \( \beta \) is stronger than \( \gamma \), and

\[
\partial^v_\beta f(x) \subset \partial^c_\beta f(x)
\]

with the equality valid if \( \beta \) is the Fréchet bornology and \( X \) is Fréchet smooth (there is an equivalent norm in \( X \) which is Fréchet differentiable off the origin).

The general statements, relating to all \( \beta \)-subdifferentials, follow immediately from the definitions. A proof of the equality of the two Fréchet subdifferentials can be found in [21].

Note further that viscosity subdifferentials are geometrically consistent which is easy to check. On the other hand, the definition of the canonical subdifferential offers a way to verify whether a given element of \( X^* \) is an element of the subdifferential of a given function at a given point. For the canonical Dini–Hadamard subdifferential there is a more constructive (and well-known) representation through a directional derivative, similar to what we have in convex analysis, which allows to actually find elements of the subdifferential. Namely, let

\[
d^- f(x; h) = \liminf_{(t,u) \to (+0,h)} t^{-1}(f(x + tu) - f(x))
\]

be the Hadamard or contingent directional derivative of \( f \) at \( x \).

**Proposition 7.6.** We have

\[
\partial_{DH} f(x) = \{x^*: d^- f(x; h) \geq \langle x^*, h \rangle, \forall h \in X\}.
\]
It follows in particular that the Dini–Hadamard normal cone to (that is the Dini–Hadamard subdifferential of the indicator function of) a set $S \subset X$ at $x \in S$ is the polar of the contingent cone $T(S, x)$ to $S$ (see [1]):

$$T(S, x) = \limsup_{t \to +0} t^{-1} (S - x),$$

where the limsup is understood in the standard Painlevé–Kuratowski sense as the collection of all limits $\lim h_n$ such that $x + t_n h_n \in S$ for some $t_n \to +0$.

Another consequence of the proposition is that for a Lipschitz function the Dini–Hadamard subdifferential coincides with the subdifferential associated with the weakest finite bornology.\(^3\)

We can now pass to the calculus of $\beta$-subdifferentials and state, to begin with, a theorem containing a characterization of a class of Banach space on which $\beta$-subdifferentials can be trusted.

**Theorem 7.7 ([38]).** Assume that $X$ is a Banach space and there is a $\beta$-differentiable Lipschitz bump on $X$. Then both the canonical and the viscosity $\beta$-subdifferentials are trusted on $X$.

A Lipschitz $\beta$-differentiable bump always exists in a $\beta$-smooth space, which is a space with an equivalent $\beta$-differentiable norm.

**Theorem 7.8.** Assume that $\partial_{\beta}$ is trusted on a class of Banach spaces containing $X$. Let $f_1, \ldots, f_k$ be functions on $X$ defined and Lipschitz in a neighborhood of a certain $\overline{x} \in X$. Set $f = f_1 + \cdots + f_k$ and assume that $x^* \in \partial_{\beta} f(\overline{x})$, where $\partial_{\beta}$ stands either for the canonical or for the viscosity $\beta$-subdifferential. Then for any $\varepsilon > 0$ and any $S \in \beta$ there are $x_0 \in X$ and $(x_i, x_i^*) \in X \times X^*$, $i = 1, \ldots, k$, such that

$$\|x_i - \overline{x}\| < \varepsilon, \quad x_i^* \in \partial_{\beta} f_i(x_i), \quad x^* - (x_1^* + \cdots + x_k^*) \in K \partial_{\beta} d_S(x_0 - \overline{x}) + \varepsilon B_{X^*},$$

where $K$ is not smaller than the sum of Lipschitz constants of $f_i$ and $\|x^*\| + \varepsilon$.

**Proof.** The proof basically follows the lines of the proof of Theorem 4.2 with $L$ replaced by $S$. By definition $f(x) - (x^*, x - \overline{x}) + \varepsilon \|x - \overline{x}\|$ attains a local minimum on $\overline{x} + S$ at $\overline{x}$. As this function is Lipschitz, we can be sure, after adding $K d_S(x - \overline{x})$, that the resulting function

$$x \mapsto f_1(x) + \cdots + f_k(x) - (x^*, x - \overline{x}) + \varepsilon \|x - \overline{x}\| + K d_S(x - \overline{x}) \quad (7.4)$$

\(^3\) This explains why we choose the name “Dini–Hadamard” rather than “Hadamard”: for a Lipschitz function the contingent directional derivative coincides with the function whose values are directional lower Dini numbers.
attains an unconditional local minimum at $\bar{x}$. This function is a sum of Lipschitz functions, so we can use trustworthiness directly with the subsequent application of the convex subdifferential calculus to get the desired result exactly as in the proof of Theorem 4.2.

Observe that, unlike in Theorem 4.2, here we assume $f_i$ Lipschitz near $\bar{x}$. The reason is that without that we would be unable to check that the minimum of the displayed functions is robust (unless we assume this explicitly or add some new conditions). The same applies to the next theorem as well.

**Theorem 7.9.** Assume that $\partial \beta$ can be trusted on a class of Banach spaces containing $X$ and $Y$. Let $\varphi$ be a function on $X \times Y$ which is locally Lipschitz on its domain. Set $f(x) = \inf_y \varphi(x, y)$. Let finally $\bar{x}^* \in \partial \beta f(\bar{x})$. Then for any $\varepsilon > 0$ and any $S \in \beta(X)$ there are a pair $(x, y)$ and a $w \in x + S$ such that $x$ and $w$ are $\varepsilon$-close to $\bar{x}$, $\varphi(x, y) < f(\bar{x}) + \varepsilon$ and

$$(\bar{x}^*, 0) \in \partial \beta \varphi(u, v) + N(S, w - x) \times \{0\} + \varepsilon(B_{X^*} \times \varepsilon B_{Y^*}).$$

If moreover, $\bar{y} \in F(\bar{x})$ is such that $f(\bar{x}) = \varphi(\bar{x}, \bar{y})$, then $y$ can be chosen in the $\varepsilon$-neighborhood of $\bar{y}$.

**Proof.** The proof is identical to the proof of Theorem 4.3 with $S$ instead of $L$ and just one difference: the minimum in (4.2) is now robust because there is only one non-Lipschitz term in the function attaining minimum at $(\bar{x}, \bar{y})$.

The main improvement of the two theorems over their general counterparts, Theorems 4.2 and 4.3, is that here we know subgradients not up to weak*-neighborhoods of the origin but up to smaller sets (sums of small balls and normal cones to elements of the bornology).

The improvement of the approximate projection theorem for $\beta$-subdifferentials is even more substantial, at least in case of $\beta$-smooth spaces.

**Theorem 7.10.** Let $X$ be a space with a $\beta$-differentiable norm. Let $P \subset X$ be a closed set, and let $u^* \in \partial \beta d_P(u)$. Then for any $\varepsilon > 0$ and any $S \in \beta$ there are $x \in P$, $x^* \in (1 + \varepsilon)\partial \beta d_P(x)$ and $e \in S$ such that $\|x - u\| < d_S(u) + \varepsilon$, $\|e\| < \varepsilon$, $x^* - u^* \in N(S, e) + \varepsilon B$.

**Proof.** The conclusion of the theorem is trivial if $u$ itself belongs to $P$, so we assume in the proof that $u \not\in P$. Fix an $\varepsilon > 0$ and an $S \in \beta$. We have

$$d_P(u + h) - d_P(u) - \langle u^*, h \rangle + \frac{\varepsilon}{4} \|h\| \geq 0$$

for all $h \in S$ of a neighborhood of zero. Set

$$g(v, h) = \|v - (u + h)\| - \langle u^*, h \rangle + \frac{\varepsilon}{4} \|h\|.$$
Then there is a $\delta > 0$ such that
\[
\inf \{ g(v, h) : x \in P, \, h \in S, \, \| h \| < \delta \} = \inf_{v \in P} g(v, 0) = d_P(u).
\]
We may of course assume that $\delta < d(u, S)$ and $\delta < \varepsilon/4$.

Let $F$ be the space of functions on $X \times X$ of the form $\varphi(v, h) = \xi(v) + \eta(h)$ with both $\xi$, $\eta$ (globally) bounded and Lipschitz and everywhere $\beta$-differentiable. This space is nontrivial: take for instance $\xi(v) = \eta(v) = (\max\{0, 1 - \|v\|^2\})^2$.

We shall consider this space with the norm (lip $\xi$ is the Lipschitz constant of $\xi(\cdot)$)
\[
\| \varphi \| = \sup_{v \in X} |\xi(v)| + \sup_{h \in X} |\eta(h)| + \text{lip} \xi + \text{lip} \eta.
\]

By the variational principle of Deville–Godefroy–Zizler there is a $\varphi \in F$ with $\| \varphi \| < \delta$ such that $g - \varphi$ attains its minimum on $P \times (S \cap \delta B)$ at a certain $(x, e)$. We have
\[
\delta \geq \sup |\varphi| = \sup |\varphi| + \inf g - d_P(u) \geq \inf (g - \varphi) - d_P(u)
\]
\[
\geq g(x, e) - d_P(u) - \delta,
\]
that is (as $\|u^*\| \leq 1$),
\[
2\delta > \| x + e - u \| - d_P(u) - \langle u^*, e \rangle + \frac{\varepsilon}{4}\|e\|
\]
\[
\geq \| x - u \| - d_P(u) - 2\|e\|.
\]
Thus $\| x - u \| < d_P(u) + 4\delta < d_P(u) + \varepsilon$.

The function $g - \varphi$ satisfies the Lipschitz condition with constant smaller than $(1 + \delta)$ as a function of $v$ and with constant smaller than $2 + (\varepsilon/2)$ as a function of $h$. Therefore
\[
G(v, h) = g(v, h) - \xi(v) - \eta(h) + (1 + \delta)d_P(v) + \left(2 + \frac{\varepsilon}{2}\right)d_S(h)
\]
attains an unconditional minimum at $(x, e)$.

As $x - (u + e) \neq 0$, the norm $\cdot$ is $\beta$-differentiable at $x - (u + e)$. Let $-z^*$ stand for the $\beta$-derivative of the norm at this point. We have (as $G(v, e) - G(x, e) \geq 0$)
\[
(1 + \delta)(d_P(v) - d_P(x)) \geq \xi(v) - \xi(x) - (\|v - (u + e)\| - \|x - (u + e)\|).
\]

The function in the right-hand side of the inequality is $\beta$-differentiable at $x$ and both sides equal to zero at the point. Therefore
\[
x^* := z^* + \xi'(x) \in (1 + \delta)\partial_\beta d_P(x).
\]
A. D. Ioffe

(We may write even the viscosity subdifferential on the right.) Likewise,

\[ 0 \leq g(x, h) - g(x, e) + \left( 2 + \frac{\varepsilon}{2} \right) d_S(h) \]

\[ = \|x - u - h\| - \|x - u - e\| - \langle u^*, h - e \rangle + \left( 2 + \frac{\varepsilon}{2} \right) d_S(h) + \eta(h) - \eta(e) + \frac{\varepsilon}{4} (\|h\| - \|e\|) + \left( 2 + \frac{\varepsilon}{2} \right) d_S(h) \]

from which we get that

\[ u^* \in z^* + \eta'(e) + \frac{\varepsilon}{4} B + \left( 2 + \frac{\varepsilon}{2} \right) d_S(e) \subset x^* - \xi'(x) + \eta'(e) + \frac{\varepsilon}{4} B + N(S, e). \]

As the Lipschitz constant of \( \varphi \) is smaller than \( \delta \), the same is true for the norms of \( \xi' \) and \( \eta' \). So the right inclusion implies that \( u^* - x^* \in \varepsilon B + N(S, e) \). \( \Box \)

8 Limiting and modified \( \beta \)-subdifferentials

As in Sections 5 and 6, we assume here all functions lower semicontinuous.

We shall speak in this section only about canonical subdifferentials and denote by \( \partial \lambda \beta \) the limiting canonical \( \beta \)-subdifferential. As follows from Theorem 5.2, this is a tight subdifferential which is trusted on spaces with Lipschitz \( \beta \)-differentiable bump.

We shall mainly work in this section with the distance functions to epigraphs of functions. It is immediate from the definition of the \( \beta \)-subdifferentials that

\[ (x, \alpha) \in \text{epi } f, \ (x^*, \gamma) \in \partial \beta d_{\text{epi } f}(x, \alpha), \ \gamma < 0 \quad \Rightarrow \quad \alpha = f(x). \quad (8.1) \]

The starting point of our discussion is the following lemma describing the connection between the \( \beta \)-subdifferentials of a function, the \( \beta \)-normal cone to its epigraph and the \( \beta \)-subdifferential of the distance function to the epigraph.

**Lemma 8.1.** Let \( X \) be a Banach space. Assume that the \( \beta \)-subdifferential is trusted on \( X \). Then for any lower semicontinuous function \( f \) on \( X \), any \( x \in \text{dom } f \) and any \( \lambda > 0 \)

\[ \{ x^* : (x^*, -1) \in \lambda \partial \beta d_{\text{epi } f}(x, f(x)) \} \subset \partial \beta f(x) \]

\[ \subset \{ x^* : (x^*, -1) \in N_\beta(\text{epi } f, (x, f(x))) \}. \]

If moreover, \( f \) is Lipschitz near \( x \) and the distance in \( X \times \mathbb{R} \) corresponds to the norm \( \|(x, \alpha)\| = K \|x\| + |\alpha| \) with \( K \) greater than the Lipschitz constant of \( f \) at \( x \), then

\[ \{ x^* : (x^*, -1) \in \partial \beta d_{\text{epi } f}(x, f(x)) \} = \partial \beta f(x). \]
Proof. We shall start with the right inclusion. Let \( x^* \in \partial f(x) \), and let \( S \in \beta \). Then
\[
f(x + h) - f(x) - \langle x^*, h \rangle + \varepsilon \|h\| \geq 0
\]
for all \( h \in S \) sufficiently close to zero. Therefore whenever \( f(x) + \alpha \geq f(x + h) \), we have
\[
\langle x^*, h \rangle - \alpha \leq \varepsilon (\|h\| + |\alpha|).
\]
The inequality therefore holds for any \((h, \alpha) \in S \times [-1, 1]\) sufficiently close to zero and such that \((x, f(x)) + (h, \alpha) \in \text{epi } f \) whenever \( S \in \beta \) which means that \((x^*, -1) \in N_\beta(x, f(x))\).

Let us prove the left inclusion. Choose an \( r > 0 \) such that \( r \varepsilon > 0 \) and \( \gamma > 0 \) to make sure that \( \mu(1 + (1 - \mu)^{-1}\|x^*\|) \leq \varepsilon \). Then as follows from (8.2) there is a \( \delta > 0 \) such that
\[
\lambda f(x + h) - (f(x) + \xi) + \mu(\|h\| + |\xi|) \geq 0
\]
if \( h \in S, \|h\| \leq \delta, |\xi| < \delta \). Taking \( \xi = f(x + h) - f(x) \) and setting
\[
\sigma(h) = \text{sign}(f(x + h) - f(x)),
\]
we get from (8.3)
\[
(1 + \mu \sigma(h))(f(x + h) - f(x)) - \langle x^*, h \rangle + \mu \|h\| \geq 0
\]
if \( h \in S, \|h\| < \delta, |f(x + h) - f(x)| < \delta \), and dividing the above inequality by \((1 + \mu \sigma(h))\) we see that for such \( h \)
\[
f(x + h) - f(x) - \langle x^*, h \rangle + \varepsilon \|h\|
\]
\[
\geq f(x + h) - f(x) - \langle x^*, h \rangle + \mu(1 + (1 - \mu)^{-1}\|x^*\|)\|h\| \geq 0.
\]
The proof of the first part of the lemma is now completed by a reference to Proposition 7.3.

Assume now that \( f \) is Lipschitz near \( x \) and the Lipschitz constant of \( f \) is smaller than \( K \). Simple and well-known arguments show that in this case we have \( d_{\text{epi } f}(u, \alpha) = (f(u) - \alpha)^+ \) for all \((u, \alpha)\) of a neighborhood of \((x, f(x))\). Let now \( x^* \in \partial f(x), \varepsilon > 0 \) and \( S \in \beta \) be given. Then for some \( \delta > 0 \) and all \( h \in S \) with \( \|h\| < \delta \) and \( \xi \in \mathbb{R} \) with \( |\xi| < \delta \) we have
\[
d_{\text{epi } f}(x + h, f(x) + \xi) = (f(x + h) - (f(x) + \xi))^+
\]
\[
\geq f(x + h) - f(x) - \xi
\]
\[
\geq \langle x^*, h \rangle - \xi - \varepsilon \|h\|.
\]
that is,
\[ d_{\text{epi}} f(x + h, f(x) + \xi) - ((x^*, h) - \xi) + \varepsilon(\|h\| + |\xi|) \geq 0 \]
which means that \((x^* , -1) \in \partial_{\text{epi}} f(x, f(x))\). \qedhere

We shall next turn to the modified \(\beta\)-subdifferential. For notational simplicity we shall denote it by \(\partial_{M\beta}\). As follows from Definition 6.2, for any lower semicontinuous function \(f\)
\[ \partial_{M\beta} f(x) = \bigcup_{\lambda > 0} \{x^* : (x^*, -1) \in \lambda \mathcal{L}_\beta(f, x)\}, \]
where
\[ \mathcal{L}_\beta(f, x) = \{\lim(x_n^*, \gamma_n) : (x_n^*, \gamma_n) \in \partial_\beta d_{\text{epi}} f(x_n, f(x_n)), (x_n, f(x_n)) \to (x, f(x))\}. \]

Since \(\partial_\beta\) is an elementary subdifferential, it is distance-homotone and therefore, by Theorem 6.4, \(\partial_{M\beta}\) is a subdifferential on the class of lower semicontinuous functions.

**Proposition 8.2.** If the \(\beta\)-subdifferential is trusted on \(X\), then
\[ \partial_{M\beta} f(x) \subset \partial_{L\beta} f(x) \]
for any \(x\) if \(f\) is lower semicontinuous and \(\partial_{M\beta} f(x) = \partial_{L\beta} f(x)\) if \(f\) is Lipschitz near \(x\).

**Proof.** This is a direct consequence of Lemma 8.1. Indeed, the first part of the lemma shows that \(\{x^* : (x^*, -1) \in \mathcal{L}_\beta(f, x)\} \subset \partial_{L\beta} f(x)\). On the other hand, by Proposition 6.3 we are free to choose an equivalent norm in \(X \times \mathbb{R}\) for the calculation of \(\partial_{M\beta}\). So if \(f\) is Lipschitz near \(f\) we can take \(\|u, \alpha\| = K\|u\| + |\alpha|\) with \(K\) greater that the Lipschitz constant of \(f\) and the second part of the lemma guarantees the equality. \qedhere

We now can define the modified \(\beta\)-normal cone in the standard way.

**Definition 8.3.** Let \(Q \subset X\) be a closed set. Then the modified \(\beta\)-normal cone to \(Q\) at \(x \in Q\) is
\[ N_{M\beta}(Q, x) = \partial_{M\beta} \text{Ind}_Q(x). \]
The modified \(\beta\)-coderivative is defined in the standard way: if \(F : X \rightrightarrows Y\) and \(y \in F(x)\), then
\[ D^*_M F(x, y)(y^*) = \{x^* : (x^*, -y^*) \in N_{M\beta} (\text{Graph } F, (x, y))\}. \]
Proposition 8.4. For a closed $Q \subset X$ set
\[
\mathcal{L}_\beta(Q, x) = \{x^* = w^* - \lim x_n^* : x_n^* \in \partial \beta d_Q(x_n), \; x_n \in Q\; x_n \to x\}.
\]
Then
\[
N_{M\beta}(Q, x) = \bigcup_{\lambda > 0} \lambda \mathcal{L}_\beta(Q, x).
\]

Proof. Set for brevity $f = \text{Ind}_Q$. Then $\text{epi } f = Q \times \mathbb{R}_+$ and therefore
\[
d_{\text{epi } f}(x, \alpha) = d(x, Q) + \alpha^-
\]
(if we consider the sum norm $\| (x, \alpha) \| = \| x \| + |\alpha| \text{ in } X \times \mathbb{R}$, where $\alpha^-$ is defined by $\alpha^- = \max\{0, -\alpha\}$. Apply Definition 6.2.

We can now pass to the study of the calculus rules for $\partial_{M\beta}$ and the two basic operations.

Definition 8.5. We say that $f = f_1 + \cdots + f_k$ satisfies the (linear) metric qualification condition near $\overline{x} \in \text{dom } f$ if there are $K > 0$ and $\varepsilon > 0$ such that
\[
d_{\text{epi } f}(x, \alpha) \leq K (d_{\text{epi } f_1}(x, \alpha_1) + \cdots + d_{\text{epi } f_k}(x, \alpha_k))
\]
(8.5)
if $\| x - \overline{x} \| < \varepsilon$, $\alpha = \alpha_1 + \cdots + \alpha_k$ and $|\alpha_i - f_i(\overline{x})| < \varepsilon$.

Theorem 8.6. Let $X$ be a $\beta$-smooth Banach space. Let $f_1, \ldots, f_k$ be lower semi-continuous in a neighborhood of $\overline{x}$ and finite at $\overline{x}$. If $f = f_1 + \cdots + f_k$ satisfies the metric qualification condition near $\overline{x}$, then for any $S \in \beta(X)$
\[
\partial_{M\beta} f(x) \subset \partial_{M\beta} f_1(\overline{x}) + \cdots + \partial_{M\beta} f_k(\overline{x}) + N(S, 0).
\]

Note that $N(S, 0) = \{x^* : \langle x^*, x \rangle \geq 0, \; \forall x \in S\}$ is a subspace as $S$ is convex and $S = -S$.

Proof. So let $x^* \in \partial_{M\beta} f(x)$. This means that there is a positive $\lambda$ and a sequence of pairs $(x_n, x_n^*)$ such that $x_n \to x$, $f(x_n) \to f(\overline{x})$, $x_n^*$ weak*-converge to $x^*$ and $(x_n^*, -1) \in \lambda \partial \beta d_{\text{epi } f}(x_n, f(x_n))$. Fix an $S \in \beta$ and let $(\varepsilon_n)$ be a sequence of positive numbers going to zero. By definition for any $n$ the function
\[
(h, \xi) \mapsto \lambda d_{\text{epi } f}(x_n + h, f(x_n) + \xi) - \langle x_n^*, h \rangle + \xi + \varepsilon_n(\| h \| + |\xi|)
\]
attains a local minimum on $S \times [-1, 1]$ at zero. As the metric qualification condition is satisfied near $\overline{x}$, we can be sure that there is an $r > 0$ such that for any $n$
the function

\[ (h, \xi_1, \ldots, \xi_k) \mapsto r \sum_{i=1}^{k} d_{\text{epi} f_i} (x_n + h, f_i(x_n) + \xi_i) - \langle x_n^*, h \rangle - \sum_{i=1}^{k} \xi_i + \varepsilon_n (\|h\| + |\xi_1 + \cdots + \xi_k|) + \text{Ind}_S(h) \]

attains a local minimum on at zero. As \( X \times \mathbb{R}^k \) is a \( \beta \)-smooth space, we can find vectors \( h_n \to 0 \) (as \( n \to \infty \)), numbers \( \xi_n \to 0 \), pairs

\[ (u_{i,n}, \eta_{i,n}) \in r \partial \beta d_{\text{epi} f_i} (x_n + h_n, f_i(x_n) + \xi_n). \]

\( u_n \in S \) also converging to zero and \( u_n^* \in N(S, u_n) \) such that

\[ \left\| \sum_{i=1}^{k} u_{i,n}^* - x^* + u_n^* \right\| < 2 \varepsilon_n, \quad |\eta_{i,n} + 1| < \varepsilon_n. \]

Set for brevity \( x_{i,n} = x_n + h_n \).

Using Theorem 7.10, we can guarantee that \( (x_{i,n}, f_i(x_n) + \xi_{i,n}) \in \text{epi} f_i \) and, moreover, in view of (8.1) that \( f_i(x_n) + \xi_{i,n} = f(x_{i,n}) \) (as \( \eta_{i,n} \to -1 \)). Furthermore, the norms of \( x_{i,n}^* = u_{i,n}^* / |\eta_{i,n}| \) are uniformly bounded. Therefore each sequence \( (x_{i,n}^*) \) contains a subsequence weak*-converging to some \( x_i^* \) and we have \( (x_i^*, -1) \in r \mathcal{L}_\beta (f_i, \overline{x}) \), that is, \( x_i^* \in \partial \mathcal{M}_{\beta} f_i(\overline{x}) \). It remains to observe that \( u_n^* \) weak*-converge to \( u^* = x^* - \sum x_i^* \in N(S, 0) \). \( \square \)

**Theorem 8.7.** Let \( X \) and \( Y \) be \( \beta \)-smooth spaces, and let \( \varphi \) be a lower semicontinuous function on \( X \times Y \). We set \( f(x) = \inf_y \varphi(x, y) \) and denote by \( M(x) \) the collection of \( y \in Y \) such that \( \varphi(x, y) = f(x) \). Let \( \overline{x} \in \text{dom } f \). We assume that \( M(x) \neq \emptyset \) for all \( x \in \text{dom } f \) sufficiently close to \( \overline{x} \) and, moreover, the set-valued mapping \( x \mapsto M(x) \) is semicompact at \( \overline{x} \) in the following sense: if \( x_n \to \overline{x} \) and \( y_n \in M(x_n) \), then a subsequence of \( (y_n) \) converge to a certain \( \overline{y} \in M(\overline{x}) \). Let finally \( x^* \in \partial \mathcal{M}_{\beta} f(\overline{x}) \). Then for any \( S \in \beta(X) \) and any \( Q \in \beta(Y) \)

\[ (x^*, 0) \in \bigcup_{\overline{y} \in M(\overline{x})} \partial \mathcal{M}_{\beta} \varphi(\overline{x}, \overline{y}) + N(S, 0) \times N(Q, 0). \]

**Proof.** Observe that under the assumption, \( f \) is necessarily lower semicontinuous near \( \overline{x} \). As in the proof of the previous theorem we see that the function in (8.6) attains a local minimum on \( S \times [-1, 1] \) at \((0, 0)\). Let us fix the standard sum norms in \( X \times \mathbb{R} \) and \( X \times Y \times \mathbb{R} \): \( \| (x, \alpha) \| = \| x \| + |\alpha| \) and \( \| (x, y, \alpha) \| = \| x \| + \| y \| + |\alpha| \). Then of course \( d_{\text{epi} f} (x, \alpha) \leq d_{\text{epi} \varphi} (x, y, \alpha) \) and, on the other hand, if \( y \in M(x) \),
then

\[ 0 = d_{\text{epi}} f(x, f(x)) = d_{\text{epi}} \varphi(x, y, \varphi(x, y)) \]

It follows that taking a \( \tilde{y}_n \in M(x_n) \) (which is possible by our assumption, at least for large \( n \)), we conclude from (8.4) that the function

\[(h, v, \xi) \mapsto \lambda d_{\text{epi}} \varphi(x_n + h, \tilde{y}_n + v, \varphi(x_n, y_n) + \xi) - \langle x_n^*, h \rangle + \xi + \varepsilon_n (\|h\| + |\xi|) \]

attains a local minimum on \( S \times Y \) at zero.

The rest of the proof is also similar to the proof of the preceding theorem. As follows from the semicompactness property of \( M(\cdot) \), we may assume that \( (\tilde{y}_n) \) converges to some \( \tilde{y} \in M(\bar{x}) \). By trustworthiness we can find \( (h_n, v_n, \xi_n) \) converging to zero, \( (h_n^*, v_n^*, y_n) \in r \partial d_{\text{epi}} \varphi(x_n + h_n, \tilde{y}_n + v_n, \varphi(x_n, \tilde{y}_n) + \xi_n) \), \( \xi_n \to 0 \), \( u_n \in S \) converging to zero and \( u_n^* \in N(S, u_n) \) such that

\[ \|h_n^* - x^* + u_n^*\| < 2\varepsilon_n, \quad \|v_n^*\| < \varepsilon_n, \quad |y_n + 1| < \varepsilon_n \]

etc. If \( \varphi(x_n, \tilde{y}_n) + \xi_n \neq \varphi(x_n + h_n, \tilde{y}_n + v_n) \), we, as above, use Theorem 7.10 and then pass to the weak*-limit using norm boundedness of all sequences of dual vectors.

9 Fréchet and limiting Fréchet subdifferential

The calculus of subdifferentials associated with the Fréchet bornology can be further advanced thanks to three remarkable properties of the Fréchet bornology and the Fréchet subdifferential. The first and the most elementary is that, as the unit ball is an element of the bornology, the normal cone to it at any interior point reduces to zero and therefore in all Theorems 7.9, 7.8, 8.6, 8.7 the terms containing normal cones to elements of the bornology should be dropped. For instance the Fréchet version of the last Theorem 8.7 states that

\[ \partial_{MF} f(\bar{x}) \subset \left\{ x^* : (x^*, 0) \in \bigcup_{\gamma \in M(x)} \partial_{MF} \varphi(\bar{x}, \gamma) \right\} \]

(see Theorem 9.11 below). The following is a useful consequence of this property.

**Proposition 9.1.** Let \( X \) be a Banach space such that \( \partial_F \) is trusted on a class of spaces containing \( X \) and the ball \( B_{X^*} \) is weak*-sequentially compact. Then for any two functions on \( X \) defined and Lipschitz near \( \bar{x} \)

\[ \partial_{LF} (f + g)(\bar{x}) \subset \partial_{LF} f(\bar{x}) + \partial_{LF} g(\bar{x}) \]

(9.1)

(We shall see later that \( \partial_F \) is trusted only on Asplund spaces so that the dual ball is necessarily weak*-sequentially compact.)
Proof. Set \( \varphi = f + g \). If \( x^* \in \partial_F \varphi(x) \), then taking \( S = B_x \) we can find, for a given \( \varepsilon > 0 \) a \( \delta > 0 \) such that \( \varphi(x + h) - \langle x^*, h \rangle + \varepsilon \|h\| \geq \varphi(x) \) if \( \|h\| \leq \delta \). Let now \( x^* \in \partial_{LF} \varphi(x) \), let \( x_n \to x \) and \( x^*_n \in \partial F \varphi(x_n) \) weak*-converge to \( x^* \). Then for any \( x \), \( \psi_n(h) = f(x_n + h) + g(x_n + h) - \langle x^*_n, h \rangle + (1/n) \|h\| \) attains a local minimum at zero. This is a robust minimum as all terms of the function are Lipschitz. Therefore there are \( \{u_n, u^*_n, v_n, v^*_n \} \) such that \( u_n \) and \( v_n \) converge to \( \overline{x} \), \( u^*_n \in \partial F f(u_n), v^*_n \in \partial F g(v_n) \) and \( \|u_n + v_n^* - x^*\| \leq 2/n \). The sequences \( \{u^*_n\} \) and \( \{v^*_n\} \), being bounded, have subsequences converging to some \( u^* \) and \( v^* \). Then \( u^* \in \partial_{LF} f(\overline{x}), v^* \in \partial_{LF} g(\overline{x}) \) and \( u^* + v^* = x^* \).

\( \boxdot \)

The second remarkable property of the Fréchet subdifferential is revealed by the following proposition.

**Proposition 9.2.** Let \( X \) be any Banach space and \( S \subset X \) be a closed subset of \( X \). Then for any \( x \in S \)

\[ \partial_F d(\cdot, S)(x) = B \cap N_F(S, x). \]

**Proof.** We only need to prove the “\( \supset \)” part of the equality: the opposite inclusion is trivial. So let \( x^* \in \partial F \text{Ind}_S(x) \). This means that \( |\langle x^*, h \rangle| \leq o(\|h\|) \) if \( x + h \in S \). We have to show that \( x^* \in \partial_F d(\cdot, S)(x) \) if in addition \( \|x^*\| \leq 1 \), that is, that

\[
\liminf_{\|h\| \to 0} \frac{d(x + h, S) - \langle x^*, h \rangle}{\|h\|} \geq 0. \tag{9.2}
\]

Let \( (h_n), h_n \neq 0 \), be a sequence that realizes the liminf in (9.2). If

\[
\lim \|h_n\|^{-1} \langle x^*, h_n \rangle \leq 0,
\]

then (9.2) obviously holds, so we assume that there is a positive \( \alpha \) such that

\[
\langle x^*, h_n \rangle \geq \alpha \|h_n\|.
\]

Then \( x + h_n \notin S \). Take a \( w_n \in S \) such that

\[
\|w_n - (x + h_n)\| \leq d(x + h_n, S) + \|h_n\|^2 \leq 2\|h_n\|. \tag{9.3}
\]

Set \( v_n = w_n - x \). Then \( \|v_n\| \leq 3\|h_n\| \to 0 \). We have \( \langle x^*, v_n \rangle \leq o(\|v_n\|) \) as \( x + v_n \in S \). Therefore by (9.3) (as \( v_n - h_n = w_n - (x + h_n) \))

\[
\frac{d(x + h_n, S) - \langle x^*, h_n \rangle}{\|h_n\|} = \frac{d(x + h_n, S)}{\|h_n\|} - \frac{\langle x^*, v_n \rangle}{\|h_n\|} + \frac{\langle x^*, v_n - h_n \rangle}{\|h_n\|} \]

\[
\geq \frac{\|v_n - h_n\|}{\|h_n\|} + \frac{\langle x^*, v_n - h_n \rangle}{\|h_n\|} - \frac{\langle x^*, v_n \rangle}{\|h_n\|} - \|h_n\|. \]

90

A. D. Ioffe
As \( \|x^*\| \leq 1 \), the sum of the first two terms on the right-hand side of the inequality is non-negative. We also have \( \|h_n\| \to 0 \) and

\[
-\frac{\langle x^*, v_n \rangle}{\|h_n\|} \geq \frac{o(\|v_n\|)}{\|h_n\|} = \frac{o(\|h_n\|)}{\|h_n\|} \to 0,
\]

and the proof is completed. \( \square \)

Together with Lemma 8.1 this immediately leads to

**Proposition 9.3.** Fréchet subdifferential is geometrically consistent on the class of lower semicontinuous functions.

Indeed, it follows from Proposition 9.2 that the left and the right part of the chain of inclusions in Lemma 8.1 coincide. Note also that geometric consistence of the Fréchet subdifferential follows from the fact that the Fréchet subdifferential is also a viscosity subdifferential (Proposition 7.5).

Another consequence of a joint application of Lemma 8.1 and Proposition 9.3 is

**Theorem 9.4.** The equality

\[
\partial_{MF} f(x) = \partial_{LF} f(x)
\]

holds for any lower semicontinuous function on \( X \), provided the Fréchet subdifferential is trusted on \( X \).

**Proof.** Let \( f \) be an l.s.c. function on \( X \), and let \( x^* \in \partial_{LF} f(x) \). This means that there is a sequence of pairs \( (x_n, x_n^*) \) such that \( x_n \to x \), \( x_n^* \) weak*-converge to \( x^* \) and \( x_n^* \in \partial_F f(x_n) \). As \( \partial_F \) is trusted on \( X \), it follows from Lemma 8.1 that

\[
(x_n^*, -1) \in N_F (\text{epi} f; (x_n, f(x_n))).
\]

We have \( \|x_n^*\| \leq K \) for some \( K > 0 \), so taking the standard sum norm in \( X \times \mathbb{R} \) and applying Proposition 9.2, we conclude that

\[
(x_n^*, -1) \in (K + 1)\partial_F d_{\text{epi} f} (x_n, f(x_n)).
\]

This means that \( (x^*, -1) \in \mathcal{L}_F (f, x) \), that is, \( x^* \in \partial_{MF} f(x) \) and consequently,

\[
\partial_{LF} f(x) \subseteq \partial_{MF} f(x).
\]

The opposite inclusion has been obtained in Proposition 8.2. \( \square \)

This seems to be a characteristic property of the Fréchet subdifferential among other \( \beta \)-subdifferentials. A better approximate projection theorem available for the Fréchet subdifferential is also connected with the two mentioned remarkable properties.
Theorem 9.5. Let $X$ be a Banach space, and let $\partial F$ be trusted on the class of spaces containing $X$. Let $P \subset X$ be a closed set, and let $u^* \in \partial_F d_P(u)$, $u \not\in P$. Then for any $\varepsilon > 0$ there are $x \in P$ and $x^* \in (1 + \varepsilon)\partial_F d_P(x)$ such that

$$\|x - u\| < d_P(u) + \varepsilon, \quad \|x^* - u^*\| < \varepsilon.$$ 

Proof. The proof begins as the proof of Theorem 7.10. As there we set

$$g(v, h) = \|v - (u + h)\| - \langle u^*, h \rangle + \frac{\varepsilon}{4}\|h\|$$

and find a positive $\delta < \min\{d_P(u), \varepsilon\}/4$ such that

$$\inf\{g(v, h) : v \in P, \|h\| < \delta\} = \inf_{v \in P} g(v, 0) = d_P(u).$$

Arguing in a standard way, we get from here with the help of Ekeland’s variational principle that for some $\overline{v} \in P$ and $\overline{h}$ with $\|\overline{h}\| < \delta/2$ with $\|\overline{v} - u\| < d_P(u) + \varepsilon/4$ the function

$$(v, h) \mapsto g(v, h) + \text{Ind}_P(v) + \delta(\|v - \overline{v}\| + \|h - \overline{h}\|)$$

attains at $(\overline{v}, \overline{h})$ an unconditional local minimum. This function is a sum of a lower semicontinuous function $\text{Ind}_P$ and a convex continuous function. So by trustworthiness there are $x, x^*, w, h,$ and $w^*$ such that $x \in P$, $\|x - u\| \leq d_P(u) + \varepsilon/2$, $x^* \in N(P, x)$, $w^*$ belongs to the subdifferential of $\|\cdot\|$ at $w - (u + h) \neq 0$ and satisfies $\|w^* - x^*\| < \varepsilon/2$ and $\|w^* + u^*\| < \varepsilon/2$, whence $\|x^* - u^*\| < \varepsilon$. We note finally that $\|w^*\| = 1$ as $w \neq u + h$, so that $\|x^*\| \leq 1 + \varepsilon$ and Proposition 9.2 implies that $x^* \in (1 + \varepsilon)\partial d_P(x)$. □

As an easy corollary we get

Proposition 9.6. $\partial LF$ is distance homotone on the class of spaces on which $\partial F$ is trusted. Specifically, let $\partial F$ be trusted on the class of spaces containing $X$ and $Y$, let $S \subset Y$ and let $A : X \to Y$ be a linear bounded operator onto $Y$. Set $\varphi(x) = d_S(Ax)$, let $\overline{\varphi} = A\overline{x} \in S$, and let $f$ be a function on $X$ defined and l.s.c. in a neighborhood of $\overline{x}$ and such that $f(x) \geq \varphi(x)$ in the neighborhood and $f(x) = \varphi(x) = 0$ for all $x$ for which $Ax \in S$. Then $\partial_L F \varphi(\overline{x}) \subset \partial_L F f(\overline{x})$.

Proof. Take an $x^* \in \partial_L F \varphi(\overline{x})$, and let $(x_n, x^*_n)$ be given with $x_n \to \overline{x}$, $x^*_n$ weak*-converging to $x^*$ and $x^*_n \in \partial_F \varphi(Ax_n)$. Set $y_n = Ax_n$. Then $x^*_n = A^* y^*_n$ for some $y^*_n \in \partial_F d_S(y_n)$. By Theorem 9.5, given a sequence $(\varepsilon_n)$ of positive numbers converging to zero and such that $d_S(y_n) \leq \varepsilon_n$, there are $v_n \in S$ and $v^*_n \in \partial_F d_S(v_n)$ such that $\|v_n - y_n\| < 2\varepsilon_n$ and $\|v^*_n - y^*_n\| < \varepsilon_n$. As $A$ is onto, there is a $K > 0$ and for any $n$ a $u_n \in X$ such that $\|u_n - x_n\| \leq K\|v_n - y_n\| < 2K\varepsilon_n$. We have $u^*_n = A^* v^*_n \in \partial_F \varphi(u_n)$ (by (S5)) weak*-converge to $\overline{x}$. As $u_n \in S$, we obtain
0 = \varphi(u_n) = f(u_n), so as \partial_F is an elementary subdifferential, \( u_n^* \in \partial_F f(u_n) \) and hence \( x^* \in \partial_{LF} f(x) \). \( \Box \)

The third and the deepest and most fundamental property of Fréchet subdifferentiable is the possibility of separable reduction that allows to complete the theory of Fréchet associated subdifferentials and in particular to completely describe the class of spaces on which they can be trusted. The first version of the separable reduction theorem for Fréchet subdifferentials is due to Fabian and Zhivkov [28]. By now there are several more versions of the theorem [40, 54, 58]. The statement below is a slightly modified version of the main result of [40] which is the most convenient in the context of this paper.

**Definition 9.7.** Let \( X \) be a Banach space, and let \( \mathcal{S} \) be a family of separable subspaces of \( X \). We say that \( \mathcal{S} \) is a complete family of separable subspaces of \( X \) if

(a) every separable subspace of \( X \) is a subspace of some element of \( \mathcal{S} \),

(b) if \( L \in \mathcal{S} \), then the closure of \( L \) also belongs to \( \mathcal{S} \),

(c) if \( L_1 \subseteq L_2 \subseteq \cdots \subseteq L_n \subseteq \cdots \) is a sequence of elements of \( \mathcal{S} \), then \( \bigcup L_n \in \mathcal{S} \).

**Theorem 9.8** (Separable reduction of Fréchet subdifferentiability). Let \( X \) and \( Y \) be Banach spaces, let \( f \) be a function on \( Y \) locally bounded from below on \( \text{dom} \, f \), and let \( A : X \mapsto Y \) be a bounded linear operator. Let finally \( \mathcal{S}_X \) and \( \mathcal{S}_Y \) be complete families of separable subspaces of \( X \) and \( Y \) respectively. Then for any separable subspaces \( V_0 \subseteq X \), \( W_0 \subseteq Y \) and any \( \varepsilon \geq 0 \) there are subspaces \( W \in \mathcal{S}_Y \) containing \( W_0 \) and \( V \in \mathcal{S}_X \) containing \( V_0 \) such that \( A(V) \subseteq W \) and for any \( y \in W \) the relation \( A^* (\partial_F f(y)) \cap (\varepsilon B_{X^*}) \neq \emptyset \) holds if and only if

\[
(A|_V)^* (\partial_F (f|_W(y))) \cap (\varepsilon B_{X^*}) \neq \emptyset.
\]

The theorem paves a way to a complete characterization of spaces on which the Fréchet subdifferential is trusted.

**Theorem 9.9.** The Fréchet subdifferential is trusted on a Banach space \( X \) if and only if \( X \) is Asplund.

**Proof.** The fact that \( X \) is Asplund if \( \partial_F \) is trusted on \( X \) does not need separable reduction: it is actually an easy consequence of [22]. For the first time this implication was mentioned in [32] with a short proof. The opposite implication based on separable reduction was proved by Fabian [24, 25]. A short proof based on Theorem 9.8 follows.

---

4 Note that the concept of trustworthiness in the quoted papers including [32] differs from (and is seemingly weaker than) what has been defined here – see the bibliographic comments at the end of the paper.
Let $X$ be an Asplund space, and let $f$ and $g$ be two functions on $X$, the first lower semicontinuous and, generally speaking, extended real-valued and the second Lipschitz continuous. We have to prove the following: if $f + g$ attains a local minimum at $\bar{x}$, then for any $\varepsilon > 0$ we can find a quadruple $(x, u, x^*, u^*)$ such that

$$
\|x - \bar{x}\| < \varepsilon, \quad \|u - \bar{x}\| < \varepsilon, \quad |f(x) - f(\bar{x})| < \varepsilon, \quad x^* \in \partial_F f(x), \quad u^* \in \partial_F g(u), \quad \|x^* + u^*\| < \varepsilon.
$$

To this end we apply Theorem 9.8 with $Y = X \times X$ and $A$ being the diagonal operator $X \to Y$, that is, $Ax = (x, x)$, $\mathcal{S}_X$ being the collection of all separable subspaces of $X$ and $\mathcal{S}_Y$ consisting of subspaces of the form $V \times V$, where $V \in \mathcal{S}_X$. Then

$$
A^*(x^*, u^*) = x^* + u^*.
$$

It is also an easy matter to see that $\mathcal{S}_Y$ is indeed a complete family.

Then by Theorem 9.8 there is a separable subspace $V \subset X$ containing $\bar{x}$ such that for any $x, u \in V$ the relation $(\partial f(x) + \partial g(u)) \cap (\varepsilon B) \neq \emptyset$ holds if and only if the same is true for the restrictions of $f$ and $g$ to $V$. As a separable subspace of an Asplund space, $V$ has a Fréchet smooth norm and so does $V \times V$. Thus the Fréchet subdifferential is trusted on $V \times V$ and, as $f|_V + g|_V$ attains a local minimum at $\bar{x}$, there are $x, u \in V$ and $x^* \in \partial_F f|_V (x), u^* \in \partial_F g|_V (u)$ such that $\|x - \bar{x}\| < \varepsilon, \|u - \bar{x}\| < \varepsilon$ and $\|x^* + u^*\| < \varepsilon$. By the choice of $V$ the same is true for $f$ and $g$ themselves at the same points $x$ and $u$.

We conclude the section with the calculus rules for Fréchet and limiting Fréchet subdifferentials of results of the two basic operations.

**Theorem 9.10.** Let $X$ be an Asplund space, and let $f_1, \ldots, f_k$ be functions on $X$ defined and lower semicontinuous in a neighborhood of some $\bar{x} \in X$. Set

$$
f = f_1 + \cdots + f_k.
$$

Then the following two statements hold true.

(a) Let $x^* \in \partial_F f(\bar{x})$. If all $f_i$ but for at most one of them are Lipschitz near $\bar{x}$, then for any $\varepsilon > 0$, there are pairs $(x_i, x_i^*) \in X \times X^*$, $i = 1, \ldots, k$, such that $\|x_i - \bar{x}\| < \varepsilon, x_i^* \in \partial_F f_i(x_i)$ and $\|x^* - (x_1^* + \cdots + x_k^*)\| < \varepsilon$.

(b) Assume that $f_1, \ldots, f_k$ are lower semicontinuous near $\bar{x}$ and finite at $\bar{x}$ and the metric qualification condition (8.5) is satisfied near $\bar{x}$. Then

$$
\partial_{LF} f(\bar{x}) \subset \partial_{LF} f_1(\bar{x}) + \cdots + \partial_{LF} f_k(\bar{x}).
$$

**Proof.** (a) In view of Theorem 9.9, this is a direct consequence of Theorem 7.8: just take $S = B_X$. But for the proof of (b) a similar argument does not work (since an Asplund space need not be Fréchet smooth).
As $\partial F$ is geometrically consistent (Proposition 9.3), we can work with subdifferentials of the distance functions to epigraphs. Set
\[ g_i(x, \alpha_1, \ldots, \alpha_k) = d_{\text{epi}} f_i(x, \alpha_i), \quad g(x, \alpha_1, \ldots, \alpha_k) = d_{\text{epi}} f(x, \alpha_1 + \cdots + \alpha_k). \]

The formula
\[ T(x, \alpha_1, \ldots, \alpha_k) = (x, \alpha_1 + \cdots + \alpha_k) \]
defines an operator $T : X \times \mathbb{R}^k \to X \times \mathbb{R}$. Then $g = d_{\text{epi}} f \circ T$. Clearly $T$ is an operator onto, so by (S4)
\[ \partial_{\text{LF}} g(x, \alpha_1, \ldots, \alpha_k) = T^* (\partial_{\text{LF}} d_{\text{epi}} f(x, \alpha_1 + \cdots + \alpha_k)). \quad (9.4) \]

On the other hand from the metric qualification condition and the fact that $\partial_{\text{LF}}$ is distance homotone (Proposition 9.6), we get
\[ \partial_{\text{LF}} g(\bar{x}, f_1(\bar{x}), \ldots, f_k(\bar{x})) \subset \partial_{\text{LF}} \left( \sum_{i=1}^k g_i(\bar{x}, f_1(\bar{x}), \ldots, f_k(\bar{x})) \right). \quad (9.5) \]

All $g_i$ are Lipschitz functions, so by (9.1)
\[ \partial_{\text{LF}} g(x, \alpha_1, \ldots, \alpha_k) \subset \sum_{i=1}^k \partial_{\text{LF}} g_i(x, \alpha_1, \ldots, \alpha_k). \quad (9.6) \]

Observe that $g_i$ does not depend on $\alpha_j$ with $j \neq i$. Therefore if
\[ (x^*, \eta_1, \ldots, \eta_k) \in \partial_{\text{LF}} g_i(x, \alpha_1, \ldots, \alpha_k), \]
then $\eta_j = 0$ for $j \neq i$. We have further $T^*(x^*, \eta) = (x^*, \eta, \ldots, \eta)$.

Let now $x^* \in \partial_{\text{LF}} f(\bar{x})$, that is, $(x^*, -1) \in \lambda \partial_{\text{LF}} d_{\text{epi}} f(\bar{x}, f(\bar{x}))$ for some $\lambda > 0$. By (9.4) this means that
\[ (x^*, -1, \ldots, -1) \in \lambda \partial_{\text{LF}} g(\bar{x}, f_1(\bar{x}), \ldots, f_k(\bar{x})) \]
and therefore by (9.6)
\[ (x^*, -1, \ldots, -1) = \sum_{i=1}^k (x_i^*, \eta_i, \ldots, \eta_k) \]
for some $(x_i^*, \eta_i, \ldots, \eta_k) \in \lambda \partial_{\text{LF}} g_i(\bar{x}, f_1(\bar{x}), \ldots, f_k(\bar{x}))$. But as we have seen $\eta_{ij} = 0$ if $j \neq i$. It follows that $\eta_{ii} = -1$ and as follows from the definition of $g_i$, $(x_i^*, -1) \in \lambda \partial_{\text{LF}} d_{\text{epi}} f_i(\bar{x}, f_i(\bar{x}))$, that is, $x_i^* \in \partial f_i(\bar{x})$. \qed
**Theorem 9.11.** Let $X$ and $Y$ be Asplund spaces. Let $\varphi$ be a lower semicontinuous function on $X \times Y$. Set $f(x) = \inf_y \varphi(x, y)$.

(a) Let $\overline{x}^* \in \partial F f(\overline{x})$. Then for any $\varepsilon > 0$ there are $(u, v) \in X \times Y$ such that $u$ is $\varepsilon$-close to $\overline{x}$, $\varphi(u, v) < f(\overline{x}) + \varepsilon$ and

$$(\overline{x}^*, 0) \in \partial F \varphi(u, v) + \varepsilon(B_{X^*} \times B_{Y^*}).$$

If moreover, $\overline{y} \in M(\overline{x}) = \{y : f(\overline{x}) = \varphi(\overline{x}, y)\}$, then $v$ can be chosen in the $\varepsilon$-neighborhood of $\overline{y}$.

(b) Assume that $M(x) \neq \emptyset$ for all $x$ in a neighborhood of $\overline{x}$ and the set-valued mapping $x \mapsto M(x)$ is semicompact (see the statement of Theorem 8.7). Then

$$\partial_{LF} f(\overline{x}) \times \{0\} \subset \bigcup_{\overline{y} \in M(\overline{x})} \partial_{LF} \varphi(\overline{x}, \overline{y}).$$

**Proof.** Again, thanks to Theorem 9.9, the first statement is a direct consequence of Theorem 7.9. However, unlike in Theorem 7.9, we do not need to assume that $\varphi$ is Lipschitz. Indeed, let $S$ be a ball around zero. The indicator of $S$ is constant on $S$, hence Lipschitz near zero and therefore in (4.2), with $L$ replaced by $S$, the only non-Lipschitz term is $\varphi$ itself. So the desired result follows directly from the fact that $\partial F$ is trusted on the class of Asplund spaces.

The proof of the second statement is also simple. Let $x^* \in \partial_{LF} f(x)$. Then there is a sequence $(u_n, u^*_n)$ such that $u_n \to x$, $f(u_n) \to f(x)$, $u^*_n \in \partial F f(u_n)$ and $u^*_n$ weak*-converge to $x^*$. By (a) there are $(x_n, x^*_n, y_n, y^*_n)$ such that

$$\|x_n - u_n\| \to 0, \quad \|x^*_n - u^*_n\| \to 0, \quad d_M(x_n)(y_n) \to 0, \quad \|y^*_n\| \to 0$$

and $(x^*_n, y^*_n) \in \partial F \varphi(x_n, y_n) + (1/n)(B_{X^*} \times B_{Y^*})$. By semicompactness we may assume, taking a subsequence if necessary, that $y_n$ converge to some $\overline{y} \in M(\overline{x})$, hence $(x^*, 0) \in \partial_{LF} \varphi(\overline{x}, \overline{y})$. 

\[\square\]

10 \textbf{G-subdifferential}

As it is made clear by Theorem 5.4, in every Banach space there may be at most one subdifferential on the class of locally Lipschitz functions which is tight, trusted on the space and coincides with its closure. The purpose of this section is to prove the existence of such a subdifferential and to produce an explicit construction. Its metric extension, which is called the (approximate) G-subdifferential turns out to be a unique subdifferential on the class of lower semicontinuous functions which is tight, trusted on every Banach space, geometrically consistent and coinciding with its closure on the class of locally Lipschitz functions.
10.1 The Lipschitz case

Denote by $\mathcal{S}(X)$ the collection of all separable subspaces of $X$. Let $E \subset X$ be a subspace. By $\pi_E$ we denote the natural projection of $X^*$ onto $E^* : u^* = \pi_E(x^*)$ means that $\langle u^*, x \rangle = \langle x^*, x \rangle$ for $x \in E$. (Strictly speaking, $\pi_E$ is the adjoint to the natural embedding $E \to X$.) This is a weak*-continuous linear operator of norm one.

If $X$ is a separable space, and $Q \subset X^*$ is a bounded set, then the sequential weak*-closure of $Q$ coincides with its topological weak*-closure and therefore by Theorem 5.4 the limiting versions of all tight trusted subdifferentials on $X$ coincide on the class of locally Lipschitz functions. On the other hand, every separable space admits an equivalent Gâteaux differentiable norm and therefore the Dini–Hadamard subdifferential is trusted on every separable space. It is obviously tight, as every $\beta$-subdifferential. Therefore it is sufficient to consider only the Dini–Hadamard and limiting Dini–Hadamard subdifferential when dealing with separable Banach spaces.

Now let $X$ be an arbitrary Banach space and let $f$ be a function on $X$ which is Lipschitz in a neighborhood of a certain $x \in X$. Let $E \in \mathcal{S}(X)$. We set

$$\partial^E_{DH} f(x) = \{ u^* \in E^* : \langle u^*, h \rangle \leq f'(x; h), \forall h \in E \},$$

where, as before, $f'(x; \cdot)$ stands for the lower Dini derivative of $f$ at $x$:

$$f'(x; h) = \liminf_{t \to +0} t^{-1} (f(x + th) - f(x)).$$

In other words, $\partial^E_{DH} f(x)$ is the Dini–Hadamard subdifferential at zero of the restriction of the function $h \mapsto f(x + h)$ to $E$: $\partial^E_{DH} f(x) = \partial_{DH} f(x + E)$. Furthermore, we denote by $\partial^E_{LDH} f(x)$ the collection of all weak*-limits of sequences $(u^*_n)$ such that $u^*_n \in \partial^E_{DH} f(x_n)$ for some $x_n$ converging to $x$. As $f$ is Lipschitz near $x$, $\partial^E_{LDH} f(x)$ is a weak*-closed subset of any $K B_{X^*}$ with $K$ greater than the Lipschitz constant of $f$ at $x$. We also observe the obvious relation

$$E, E' \in \mathcal{S}(X), \ E' \subset E \quad \Rightarrow \quad \partial^E_{LDH} f(x) \subset \partial^{E'}_{LDH} f(x). \quad (10.1)$$

**Definition 10.1.** Let $X$ be a Banach space, and let $f$ be a function on $X$ which is Lipschitz in a neighborhood of a certain $x \in X$. The set

$$\partial_G f(x) = \bigcap_{E \in \mathcal{S}(X)} \partial^E_{LDH} f(x)$$

will be called the (approximate) $G$-subdifferential of $f$ at $x$.

As follows from (10.1), it is sufficient to take the intersection over some cofinal subset of $\mathcal{S}(X)$ that is over any collection of separable subspaces such that any
element of $\mathcal{S}(X)$ is a subspace of a certain element of the collection, in particular over any complete family of separable subspaces of $X$. We note also an obvious fact that if $X$ is separable itself, then $\partial_G f(x)$ coincides with $\partial_{LDH} f(x)$. The three propositions to follow reveal some important properties of the $G$-subdifferential.

**Proposition 10.2.** Let $X$ be a Banach space, and let $f$ be a function on $X$ defined and Lipschitz in a neighborhood of $\overline{x} \in X$. Then $\partial_G f(\overline{x})$ is a nonempty weak*-compact set and $\|x^*\| \leq K$ for all $x^* \in \partial_G f(\overline{x})$ if $K$ is equal to or greater than the Lipschitz constant of $f$ at $\overline{x}$. Moreover, the set-valued mapping $x \mapsto \partial_G f(x)$ is norm-to-weak* upper semi-continuous at $\overline{x}$, that is, for any weak*-neighborhood $V$ of the origin in $X^*$ there is an $\varepsilon > 0$ such that $\partial_G f(u) \subset \partial_G f(x) + V$ if $\|u - x\| \leq \varepsilon$.

*Proof.* As easily follows from Proposition 5.3,

$$\partial_{LDH}^{E} f(x) \neq \emptyset \quad \text{whenever } E \in \mathcal{S}(X).$$

It is also clear that $\|u^*\| \leq K$ whenever $u^* \in \partial_{LDH}^{E} f(\overline{x})$ and $K$ is greater than the Lipschitz constant of $f$ at $\overline{x}$. It follows that $\|\pi_E(x^*)\| \leq K$ for all $E \in \mathcal{S}(X)$ and every $x^* \in \partial_G f(x)$ and therefore the norm of $x^*$ itself cannot be greater than $K$. This means that

$$\partial_G f(\overline{x}) = \bigcap_{E \in \mathcal{S}(X)} \left( \pi_E^{-1}(\partial_{LDH}^{E} f(\overline{x})) \cap K B \right).$$

As $\pi_E$ is weak*-continuous, the $\pi_E$-preimage of any weak*-closed set is weak*-closed. Hence the above equality shows together with (10.1) that $\partial_G f(x)$ is an intersection of a nested family of nonempty weak*-compact sets, hence $\partial_G f(\overline{x})$ is nonempty.

It remains to show that the graph of $\partial_G f$ is closed at $\overline{x}$ in the product of the norm topology of $X$ and the weak*-topology of $X^*$. To this end we first note that Graph $\partial_{LDH}^{E} f$ is closed in the norm-to-weak*-topology of $X \times E^*$. This is immediate from the definition and the fact that the unit ball in $X^*$ is weak*-sequentially compact. The passage to $\partial_G f$ is now straightforward by way of compactness. \qed

**Proposition 10.3.** Let $A$ be a linear bounded operator from $X$ onto $Y$, and let $f$ be a function on $Y$ which is defined and Lipschitz in a neighborhood of $\overline{y} = A\overline{x}$. Set $g(x) = (f \circ A)(x)$. Then

$$\partial_G g(\overline{x}) = A^* \partial_G f(A\overline{x}).$$

*Proof.* First we note that for any $L \in \mathcal{S}(Y)$ and $M \in \mathcal{S}(X)$ with $A(M) \subset L$ there is a $E \in \mathcal{S}(X)$ containing $M$ such that $A(E) = L$. Indeed, as $A$ is onto, there is a
constant \( K > 0 \) such that for any \( y \in Y \) we can find an \( x \in X \) such that \( Ax = y \) and \( \|x\| \leq K\|y\| \). Let now \((y_n)\) be a dense countable subset of \( L \) containing zero, say \( y_0 = 0 \). For any pair of integers \( m, n \) we find an \( x_{mn} \in X \) such that

\[
Ax_{mn} = y_n - y_m \quad \text{and} \quad \|x_{mn}\| \leq K\|y_n - y_m\|.
\]

Let finally \( E \) denote the closed linear subspace containing \( M \) and all \( x_{mn}, m, n = 0, 1, \ldots \). Clearly \( E \) is a separable subspace of \( X \) and \( A(E) = L \). To see that the latter indeed holds, take a \( y \in L \) and a sequence \((y_{nk})\) converging to \( y \) and such that \( \sum_k \|y_{nk+1} - y_{nk}\| < \infty \). Then the series \( \sum_k x_{nk+1}n_k \) is absolutely converging to some \( x \) for which we obviously have \( Ax = y \).

Let us say for brevity that a \( E \in \mathcal{S}(X) \) is an element of \( \mathcal{S}(X) \) corresponding to \( L \). Such a subspace of course is not unique. It is also clear from the above arguments that the collection of separable subspaces of \( X \) corresponding to elements of \( \mathcal{S}(Y) \) is cofinal with \( \mathcal{S}(X) \). Now, given a \( L \in \mathcal{S}(Y) \), an \( x \in X \) and an \( E \in \mathcal{S}(X) \) corresponding to \( L \), we have

\[
\partial^E_{DH} g(x) = A^* \partial^L_{DH} f(Ax).
\]

(Indeed, as we have mentioned \( \partial^E_{DH} g(x) = \partial_{DH} g|_{x+E}(0) \), the function on \( E \) equal to the restriction of \( g \) to \( x + E \) and, likewise \( \partial^L_{DH} f(y) = \partial_{DH} f|_{y+L}(0) \). On the other hand, \( g|_{x+E} = f|_{y+L} \circ A|_E \) and the image of \( A|_E \) is the whole of \( L \). The equality therefore follows from the fact that \( \partial_{DH} \) is a subdifferential both on \( L \) and \( E \).)

Elementary arguments involving weak*-compactness lead to the conclusion that \( \partial^E_{LDH} g(x) = A^* \partial^L_{LDH} f(Ax) \) and a reference to the definition of \( G \)-subdifferential and Proposition 10.2 completes the proof.

\[ \square \]

**Proposition 10.4.** \( \partial_G \) is a tight subdifferential on the on the class of locally Lipschitz functions.

**Proof.** As usual (S1) and (S2) are trivial. To show that (S3) holds it is enough to note that if \( f \) is convex, then for any \( x^* \in \partial_G f(x) \), any \( E \in \mathcal{S}(X) \) and any \( h \in E \) we have

\[
(x^*, h) = (\pi_E x^*, h) \leq f(x + h) - f(x).
\]

(S4) is a part of Proposition 10.2, (S5a) follows from Proposition 10.3 and the fact that if \( f(x) = \lambda g(x + x_0) + (x^*, x) + \alpha \) then \( \partial_G f(x) = \partial_G g(x + x_0) + x^* \) which easily follows from the definition. Finally, to verify property (S5b), we note that \( \mathcal{S}(X) \times \mathcal{S}(Y) \) is cofinal with \( \mathcal{S}(X \times Y) \) and \( \partial^L_{DH} \varphi(x, y) = \partial^L_{DH} f(x) \times \partial^M_{DH} g(y) \) if \( \varphi(x, y) = f(x) + g(y) \) (with both \( f \) and \( g \) Lipschitz near the points of interest), again by the interpretation of \( \partial^E_{DH} \) as the Dini–Hadamard subdifferential of a certain function on \( E \).

To check tightness, take an \( x \), an \( x^* \in \partial_G f(x) \) and a finite dimensional subspace \( L \subset X \). Let \( E \in \mathcal{S}(X) \) contain \( L \), and let \( u^*_n \in \partial^E_{DH} f(x_n) \) for some se-
quence \((x_n) \to x\) weak*-converge to \(\pi_E(x^*)\). Given an \(\varepsilon > 0\), we can be sure that 
\[ |\langle x^*, h \rangle - \langle u^*_n, h \rangle| < (\varepsilon/2)\|h\| \]
for large \(n\) and all \(h \in L\). On the other hand,
\[ f(x_n + h) - f(x_n) - \langle u^*_n, h \rangle + \frac{\varepsilon}{2} \|h\| \geq 0 \]
for all \(h \in L\) sufficiently close to zero (because \(u^*_n\) belongs to the Dini–Hadamard subdifferential at zero of the restriction of \(h \mapsto f(x_n + h)\) to \(E\)). Thus
\[ f(x_n + h) - f(x_n) - \langle x^*, h \rangle + \varepsilon \|h\| \geq 0 \]
for all \(h \in L\) close to zero.

Our next goal is to show that \(\partial G\) is distance homotone. This will follow from certain auxiliary results, some of which are of independent interest.

**Proposition 10.5.** Let \(X\) be a separable Banach space, let \(S \subset X\) be a closed set, and let \(x \in S\). Then \(x^* \in \partial_{LDH} d_S(x)\) if and only if there is a sequence of pairs \((x_n, x_n^*)\) such that \(x_n \in S\) norm converge to \(x\) and \(x_n^* \in \partial DH d_S(x_n)\) weak*-converge to \(x^*\).

**Proof.** This is immediate from Theorem 4.4.

**Lemma 10.6.** Let \(X\) and \(Y\) be separable Banach spaces, let \(A : X \to Y\) be a linear continuous operator onto \(Y\), and let \(\overline{x} \in X\), \(\overline{y} = Ax\).

(a) Let \(S \subset Y\) be a closed set and \(\overline{y} \in S\). Set \(\varphi(x) = d_S(Ax)\). If \(f(x) \geq \varphi(x)\) for all \(x\) of a neighborhood of \(\overline{x}\) and \(f(x) = \varphi(x) = 0\) in the intersection of the neighborhood with \(Q = A^{-1}(S)\), then \(\partial_{LDH} \varphi(\overline{x}) \subset \partial_{LDH} f(\overline{x})\).

(b) Let \(\varphi\) be as in (a). If \(x^* \in \partial_{LDH} \varphi(\overline{x})\), then the entire segment 
\[ [0, x^*] = \{\lambda x^* : 0 \leq \lambda \leq 1\} \]
lies in \(\partial_{LDH} \varphi(\overline{x})\).

**Proof.** The second statement is immediate from convexity of \(\partial DH f(x)\) and Theorem 7.10. Let us prove the first. If \(x^* \in \partial_{LDH} \varphi(x)\), then (Proposition 10.3) \(x^* = A^* y^*\) for some \(y^* \in \partial_{LDH} d_S(Ax)\). If \(x \in Q\), then \(y \in S\). By Proposition 10.5 there is a sequence \((y_n) \subset S\) and for any \(n\) a \(y^*_n \in \partial_{DH} d_S(y_n)\) weak*-converging to \(y^*\). As \(A\) is onto, we can choose \(x_n\) converging to \(x\) and such that \(A x_n = y_n\). Then \(x_n \in Q\) (that is, \(\varphi(x_n) = f(x_n) = 0\)) and 
\[ \partial_{DH} \varphi(x_n) = A^* \partial_{DH} d_S(y_n) \]
by (S5). It follows that \(x^*_n = A^* y^*_n \in \partial_{DH} \varphi(x_n)\). Also \(x^*_n\) converge to \(x^*\) in the weak*-topology. By Proposition 2.6 \(x^*_n \in \partial_{DH} f(x_n)\) and therefore we get \(x^* \in \partial_{LDH} f(x)\). \(\square\)
Note that the first statement means that $\partial G$ is distance homotone when restricted to separable spaces.

**Lemma 10.7.** Let $X$ be a separable Banach space, let $f$ be a function on $X$ which is Lipschitz near an $x \in X$, and let $\{h_1, h_2, \ldots\}$ be a dense countable subset of the unit sphere of $X$. Set $L_n = \text{span}\{x_1, \ldots, x_n\}$. Then $x^* \in \partial_{LDH} f(x)$ if and only if there is a sequence $(x_n, x_n^*)$ such that $x_n \to x$, $x_n^*$ weak*-converge to $x^*$ and $\langle x_n^*, h \rangle \leq f'(x_n; h)$ for all $h \in L_n$.

**Proof.** Only the if part needs a proof. So let a sequence $(x_n, x_n^*)$ be given satisfying the indicated property. It means that $\pi_{L_n}(x_n^*)$ belongs to $\partial_{DH} f|_{x_n + L_n(0)}$. The latter implies that $f(x_n + h) - \langle x_n^*, h \rangle + (1/n)\|h\| \geq f(x_n)$ for all $h \in L_n$ sufficiently close to zero. As the function on the left is Lipschitz in a neighborhood of zero, there is a $K > 0$, not depending on $n$ (as the norms of $x_n^*$ are uniformly bounded) such that the function

$$h \mapsto f(x_n + h) - \langle x_n^*, h \rangle + \frac{1}{n}\|h\| + Kd_{x_n + L_n}(h)$$

has an unconditional local minimum at zero. It follows (as $\partial_{DH}$ is trusted on $X$) that there are $(u_n, u_n^*)$ such that $\|x_n - u_n\| \to 0$, $u_n^* \in \partial_{DH} f(u_n)$, $\|u_n^* - x_n^*\| \leq K + 1$ and $u_n^* - x_n^* \in (1/n)B_{X^*} + L_n^1$. The last two relations show that $u_n^* - x_n^*$ weak*-converge to zero, so that $u_n^*$ weak*-converge to $x^*$.

**Lemma 10.8.** Let $X$ be a Banach space, and let $E_0$ be a separable subspace of $X$.

(a) If $S_1, S_2, \ldots$ is a countable collection of subsets of $X$, then there is a subspace $E \in \mathcal{S}(X)$ containing $E_0$ and such that

$$d_{S_i}(x) = d_{S_i \cap E}(x), \quad \forall x \in E, \quad \forall i = 1, 2, \ldots$$

(b) Let $S \subseteq X$, and let $x^* \in \partial G d_S(\bar{x})$. Then there is an $E \in \mathcal{S}(X)$ containing $\bar{x}$ and $E_0$ and such that $d_S(x) = d_{S \cap E}(x)$ for all $x \in E$ and there exists a sequence of pairs $(x_k, x_k^*) \in E \times E^*$ such that $x_k \to \bar{x}$, $x_k^* \in \partial_{DH} d_{S \cap E}(x_k)$ and $x_k^*$ weak*-converge in $E^*$ to $x^*|E$. Moreover, if $\bar{x} \in S$, we may assume that $x_k \in S$ as well.

Note that the equality $\partial (d_S)|E(x) = \partial d_{S \cap E}(x)$ holds for $x \in E$, provided that $d_S(x) = d_{S \cap E}(x)$ on $E$. Here of course $x \mapsto d_{S \cap E}(x)$ is a function on $E$.

**Proof.** To prove (a) it is sufficient to construct, starting with $E_0$, an increasing system of separable subspaces $E_n$ such that $d_{S_i \cap E_{n+1}}(x) = d_{S_i}(x)$ for all $x \in E_n$ and all $i$, and then define $E$ as the closure of $\bigcup E_n$. If we have already $E_n$, then we take a dense countable subset $C_n = \{x_1, x_2, \ldots\} \subset E_n$, for every $x_k$ choose a
sequence \((u_{ikm}) \subset S_i\) such that \(d_{S_i}(x_k) \geq \|x_k - u_{ikm}\| - (km)^{-1}\) and define \(E_{n+1}\) as the space spanned by the union of \(E_n\) and all \(u_{ikm}, i, k, m = 1, 2, \ldots\). If now \(x \in E_n\) and \(x_{kr} \in C_n\) converge to \(x\), then

\[
d_{S_i}(x) = \lim_{r \to \infty} d_{S_i}(x_{kr}) = \lim_{r \to \infty} \lim_{m \to \infty} \|x_{kr} - u_{ikr,m}\|
\]

\[
= \lim_{r \to \infty} d_{S_i \cap E_{n+1}}(x_{kr})
\]

\[
= d_{S_i \cap E_{n+1}}(x)
\]

and (a) follows.

Let us prove (b). Take a \(E_1 \in \mathcal{S}(X)\) containing \(E_0\) and \(\overline{x}\) and such that

\[
d_S(u) = d_{S \cap E_1}(u) \quad \text{for all } u \in E_1.
\]

Suppose we have already \(E_k\). As \(x^*|E_k \in \partial_{LDH}^E d_S(\overline{x})\), there is a sequence of pairs \((u_{kn}, u_{kn}^*) \subset X \times E_k^*\) such that \(u_{kn}\) norm converge to \(\overline{x}\), \(u_{kn}^* \in \partial_{DH} d_{S \cap E_k}(u_{kn})\) and \(u_{kn}^*\) weak* converge in \(E_k^*\) to \(x^*|E_k\). Let \(E_{k+1} \in \mathcal{S}(X)\) be any subspace containing \(E_k\), all \(u_{nk}\) and such that \(d_S(u) = d_{S \cap E_{k+1}}(u)\) for all \(u \in E_{k+1}\). As before, let \(E\) be the closed subspace spanned by \(\bigcup E_k\).

Clearly, the norms of all \(u_{kn}^*\) do not exceed the Lipschitz constant of \(f\) in a neighborhood of \(\overline{x}\). Next we consider norm preserving extensions of every \(u_{kn}^*\) to the whole of \(E\) which we still denote by \(u_{kn}^*\). Then \(\pi_E(x^*)\) belongs to the weak* closure of \(\{u_{kn}^* : k, n, = 1, 2, \ldots\}\), and we can choose \(n_k\) for any \(k\) such that \(u_{kn_k}\) weak*-converge to \(\pi_E(x^*)\). By Proposition 10.7 (as the union of \(E_k\) is dense in \(E\), \(\pi_E(x^*) \in \partial_{LDH}^E d_S(\overline{x})\)) which is tantamount to the existence of the desired sequence \((x_k, x_k^*)\). The last statement follows from Proposition 10.5. \(\square\)

**Proposition 10.9.** \(\partial_G\) is distance homotone.

*Proof.* Let \(X, Y\) be Banach spaces, let \(A : X \to Y\) be a linear bounded operator onto \(Y\), let \(S \subset Y\) be a closed set, and let \(\overline{y} = Ax \in S\). Set \(\varphi = d_S \circ A\), that is, \(\varphi(x) = d_S(Ax)\). Let \(f\) be a function on \(X\) defined and Lipschitz in a neighborhood of \(\overline{x}\) and such that \(f(x) \geq \varphi(x)\) in a neighborhood of \(\overline{x}\) and \(f(x) = \varphi(x) = 0\) in the intersection of the neighborhood with \(Q = A^{-1}(S)\). We have to show that

\[
\partial_G \varphi(\overline{x}) \subset \partial_G f(\overline{x}).
\]

If \(X\) itself is a separable space, then (so is \(Y\) and) the result follows from Lemma 10.6. So let \(X\) be an arbitrary space and \(x^* \in \partial_G \varphi(\overline{x})\). By Proposition 10.3, \(x^* = A^* y^*\) for some \(y^* \in \partial_G d_S(A\overline{x})\). Take an \(E_0 \in \mathcal{S}(X)\), set \(L_0 = A(E_0)\), and let \(L \in \mathcal{S}(Y)\) satisfy property (b) of Lemma 10.8 (with \(X, E_0\) and \(x^*\) replaced by \(Y, L_0\) and \(y^*\)). Let finally \(E \in \mathcal{S}(X)\) be such that \(L = A(E)\) and \(E_0 \subset E\).
Then
\[ \varphi|_E(x) = (d_S)|_L(A|_E x) = d_{S \cap L}(Ax) \]
(for \( x \in E \)) and
\[ \pi_L(y^*) \in \partial_{LDH} d_{S \cap L}(\overline{y}), \]
where \( \pi_L \) is the natural projection of \( Y^* \) onto \( L^* \). We also observe that the equality \( \pi_E(x^*) = (A|_E)^* \pi_L(y^*) \) holds, where \( A|_E \) is the restriction of \( A \) to \( E \). Indeed, setting for brevity \( u^* = \pi_E(x^*) \) and \( v^* = \pi_L(y^*) \), we have for any \( h \in E \)
\[ ((A|_E)^* v^*, h) = \langle v^*, A|_E h \rangle = \langle v^*, Ah \rangle = \langle y^*, Ah \rangle = \langle x^*, h \rangle = \langle u^*, h \rangle, \]
whence the equality. Therefore
\[ \pi_E(x^*) = (A|_E)^*(\pi_L(y^*)) \subset \partial_{LDH} \varphi|_E(\overline{x}) \subset \partial_{LDH} f|_E(\overline{x}) \subset \partial_{E}^{F} d_{\text{epi} f}(\overline{x}). \]
When \( E_0 \) runs through \( S(X) \), the collection of the corresponding \( E \) is obviously cofinal with \( S(X) \). Therefore the last inclusion finally implies that \( x^* \in \partial_G f(\overline{x}) \) as claimed.

\[ \square \]

10.2 The general case

Since \( \partial_G \) is distance homotone, its metric modification \( \partial_{MG} \) is a subdifferential on the class of lower semicontinuous functions (Theorem 6.4). A remarkable property of \( G \)-subdifferential is that for a Lipschitz function \( \partial_G f(x) \) and \( \partial_{MG} f(x) \) coincide.

**Proposition 10.10.** The equality
\[ \mathcal{L}_G(f, \overline{x}) = \partial_G d_{\text{epi} f}(\overline{x}, f(\overline{x})) \]
holds for any l.s.c. \( f \). If, moreover, \( f \) is Lipschitz near \( \overline{x} \), then
\[ \partial_G f(\overline{x}) = \bigcup_{\lambda > 0} \{ x^* : (x^*, -1) \in \lambda \partial_G d_{\text{epi} f}(\overline{x}, f(\overline{x})) \}. \tag{10.2} \]

Thus \( \partial_G f(\overline{x}) = \partial_{MG} f(\overline{x}) \) if \( f \) is Lipschitz near \( \overline{x} \).

**Proof.** The equality \( \mathcal{L}_G(f, \overline{x}) = \partial_G d_{\text{epi} f}(\overline{x}, f(\overline{x})) \) is immediate from Proposition 10.2 (specifically, the fact that the graph of \( \partial_G d_{\text{epi} f} \) is closed in the product of the norm and the weak*-topologies.)

Let us prove the second statement. With every \( E \in S(X) \) we shall associate the subspace \( E \times \mathbb{R} \) in \( X \times \mathbb{R} \). The collection of such subspaces of \( X \times \mathbb{R} \) is obviously cofinal with \( S(X \times \mathbb{R}) \), which allows us not to consider other separable subspaces
of $X \times \mathbb{R}$. Let $K$ be greater than the Lipschitz constant of $f$ in a neighborhood of $\bar{x}$. We consider $X \times \mathbb{R}$ with the norm $K \|x\| + |\alpha|$. Then by Lemma 8.1 for any separable $E \subset X$ and any $x$ close to $\bar{x}$

$$
\partial_{\text{DH}} f \big|_{x+E} (x) = \{ x^* : (x^*, -1) \in \partial_{\text{DH}} d_{\text{epi} f \mid_{x+E}} (x, f(x)) \}.
$$

On the other hand, with the chosen norm in $X \times \mathbb{R}$

$$
d_{\text{epi} f \mid_{x+E}} (x, f(x)) = (d_{\text{epi} f})_{(x(f(x)) + E \times \mathbb{R}} (x, f(x)).
$$

Therefore

$$
\partial_{\text{LDH}} E f (\bar{x}) = \{ x^* : (x^*, -1) \in \partial_{\text{LDH} E} d_{\text{epi} f} (\bar{x}, f(\bar{x})) \}
$$

and consequently,

$$
\partial_G f (\bar{x}) = \{ x^* : (x^*, -1) \in \partial_G d_{\text{epi} f} (\bar{x}, f(\bar{x})) \}.
$$

The same argument (based on Lemma 8.1) shows that

$$
\bigcup_{\lambda > 0} \{ x^* : (x^*, -1) \in \lambda \partial_G d_{\text{epi} f} (\bar{x}, f(\bar{x})) \} \subset \partial_G f (\bar{x})
$$

which completes the proof of (10.2) and the proposition. $\square$

The proposition justifies the following definition.

**Definition 10.11.** Let $X$ be a Banach space, let $S$ be a closed subset of $X$, and let $\bar{x} \in S$. The $G$-normal cone to $S$ at $\bar{x}$ is

$$
N_G (S, \bar{x}) = \bigcup_{\lambda > 0} \lambda \partial_G d_S (x).
$$

Let now $f$ be a lower semicontinuous function on $X$. Let $x \in \text{dom } f$. We define the approximate $G$-subdifferential of $f$ at $x$ by

$$
\partial_G f (x) = \{ x^* : (x^*, -1) \in N_G (\text{epi } f, (x, f(x))) \}.
$$

In view of Proposition 10.10 the definition just means that we write $\partial_G f (x)$ instead of $\partial_{\text{MG} G} f (x)$ also for lower semicontinuous functions.

**Theorem 10.12.** The $G$-subdifferential is a tight geometrically consistent subdifferential on the class of lower semicontinuous functions which is trusted on all Banach spaces. Moreover, let $X$ be a Banach space, and let $f$ and $g$ be two functions on $X$, the first lower semi-continuous and the second Lipschitz continuous near some $x \in X$. Then

$$
\partial_G (f + g) (x) \subset \partial_G f (x) + \partial_G g (x).
$$

(10.3)
Proof. The fact that $\partial G$ is a subdifferential on the class of lower semicontinuous functions follows from Theorem 6.4. As to geometric consistence, it is embedded in the definition of the $G$-subdifferential above (and justified for Lipschitz functions in Proposition 10.10).

To prove tightness, take an $(x^*, -1) \in \partial_G d_{\text{epi}} f(x, f(x))$. By Lemma 10.8 for any finite dimensional $L \subset X$ there exists an $E \in \mathcal{S}(X)$ containing both $x$ and $L$ and such that $d_{\text{epi}} f(x, \alpha) = d_{\text{epi}} f \cap E \times \mathbb{R}(x, \alpha)$ for all $(x, \alpha) \in E \times \mathbb{R}$ and there is a sequence $(x_n, \alpha_n, x^*_n, \gamma_n) \in E \times E^*$ such that $(x_n, \alpha_n) \in \text{epi} f$, $x_n \to x$, $x^*_n$ weak*-converge to $\pi_E(x^*)$, $\gamma_n \to -1$ and $(x^*_n, \gamma_n) \in \partial DH(d_{\text{epi}} f)|_{E \times \mathbb{R}} (x_n, \alpha_n)$. As $\gamma_n < 0$ and $(x_n, \alpha_n) \in \text{epi} f$, we get $\alpha_n = f(x_n)$.

Thus, for any $\varepsilon > 0$

$$d_{\text{epi}} f(x_n + h, f(x_n) + \xi) - (\langle x^*_n, h \rangle + \gamma_n \xi) + \varepsilon(\|h\| + \|\xi\|) \geq 0 \quad (10.4)$$

for all $(h, \xi) \in L \times \mathbb{R}$ sufficiently close to zero, say if $\|h\| < \delta$ and $|\xi| < \delta$. Let $K > \|x^*_n\|$. (The sequence $(x_n^*)$ weak*-converges, hence is norm bounded.) Take $\delta' > 0$ so small that $f(x_n + h) > f(x_n) - \delta$ if $\|h\| < \delta'$ and $K\delta' < \delta$. Take an $h$ with $\|h\| < \delta'$. If $f(x_n + h) - f(x_n) \geq \delta$, then $f(x_n + h) - f(x_n) - \langle x^*_n, h \rangle \geq 0$. Otherwise, take $\xi = f(x_n + h) - f(x_n)$. Then by (10.4)

$$-(\langle x^*_n, h \rangle + \gamma_n (f(x_n + h) - f(x_n))) + \varepsilon(\|h\| + |f(x_n + h) - f(x_n)|) \geq 0.$$

Setting $\rho_n = K[1 + |\gamma_n|] / (1 - \varepsilon)$, we get from here

$$f(x_n + h) - f(x_n) - \langle x^*_n, h \rangle + (\varepsilon + \rho_n)\|h\| \geq 0.$$

This proves tightness as $\rho_n \to 0$.

The proof of trustworthiness is a bit more tricky. Let two functions $f$ and $g$ on $X$ be given, the first lower semicontinuous and the second Lipschitz near $\bar{x}$. There will be no loss of generality in assuming that $g$ is Lipschitz on its entire domain.

First we shall show that (10.3) holds if both functions are Lipschitz near $x$. To this end, it is sufficient to verify that for any $E \in \mathcal{S}(X)$

$$\partial^E_{\text{LDH}} (f + g)(x) \subset \partial^E_{\text{LDH}} f(x) + \partial^E_{\text{LDH}} g(x). \quad (10.5)$$

Let $x^* \in \partial^E_{\text{LDH}} (f + g)(x)$. Then there are sequences $(x_n), (x^*_n)$ such that $x_n \to x$, $x^*_n$ weak*-converge to $x^*$ and $x^*_n \in \partial^E_{\text{DH}} (f + g)(x_n)$. As the Dini–Hadamard subdifferential is trusted on a separable space, we can, given a weak*-neighborhood $V$ of zero in $X$, find $u_n, w_n \in x_n + E$ converging to $x$ and such that

$$x^*_n \in \partial^E_{\text{DH}} f(u_n) + \partial^E_{\text{DH}} g(w_n) + V.$$

It follows (because of the uniform boundedness of $\partial^E_{\text{DH}} f(u_n)$ and $\partial^E_{\text{DH}} g(w_n)$) that $x^* \in \partial^E_{\text{LDH}} g(x) + \partial^E_{\text{LDH}} h(x) + V$. This is true for any $V$ and we can conclude that $x^* \in \partial_G g(x) + \partial_G h(x)$ as the set is weak*-compact. This proves (10.3) if both functions are Lipschitz.
Assume now that one of the functions, say $f$, may not be Lipschitz. To be able to deal with this case we need to know the connection between the distance to the epigraph of $f + g$ and the distances to the epigraphs of the component functions. To this end we consider the space $Y = X \times \mathbb{R} \times \mathbb{R}$ and the set

$$Q = \{(x, \alpha, \beta) \in X \times \mathbb{R} \times \mathbb{R} : \alpha + \beta \geq f(x) + g(x)\}.$$ 

Let $K$ be the Lipschitz constant of $g$ in a neighborhood of $x$. We shall consider $Y$ with the norm $\|(x, \alpha, \beta)\| = K\|x\| + |\alpha| + |\beta|$ and $X \times \mathbb{R}$ with the norm $\|(x, \alpha)\| = K\|x\| + |\alpha|$. We claim that with these norms

$$d_{\text{epi}}(f+g)(x, \alpha + \beta) \leq 2(d_{\text{epi}} f(x, \alpha) + d_{\text{epi}} g(x, \beta)).$$  

(10.6)

This will be sufficient to complete the proof. Indeed, set

$$\varphi(x, \alpha, \beta) = d_Q(x, \alpha + \beta),$$

and let $T(x, \alpha, \beta) = (x, \alpha + \beta)$. Then $\varphi = (d_{\text{epi}}(f+g) \circ T)(x, \alpha, \beta)$ and by Proposition 10.3 (as $T$ is obviously an operator onto and $T(x^*, \gamma) = (x^*, \gamma, \gamma)$)

$$\partial_G \varphi(x, \alpha, \beta) = \{(x^*, \gamma, \gamma) : (x^*, \gamma, \gamma) \in \partial_G d_{\text{epi}}(f+g)(x, \alpha + \beta)\}.$$ 

On the other hand, if we set

$$\varphi_1(x, \alpha, \beta) = d_{\text{epi}} f(x, \alpha) \quad \text{and} \quad \varphi_2(x, \alpha, \beta) = d_{\text{epi}} g(x, \beta),$$

then (10.6) implies, through Proposition 10.9, that

$$\partial_G \varphi(x, \alpha, \beta) \subset 2(\partial_G \varphi_1(x, \alpha, \beta) + \partial_G \varphi_2(x, \alpha, \beta))$$

and the desired conclusion follows from the definition in view of the fact that the $\beta$-component of $\partial_G \varphi_1$ and the $\alpha$-component of $\partial_G \varphi_2$ are equal to zero.

To prove (10.6), we note the obvious equality

$$\inf\{|\alpha - \xi| : \xi \geq \eta\} = (\eta - \alpha)^+$$

which, along with Lemma (8.2) justifies the following calculation:

$$d_Q(x, \alpha, \beta) = \inf\{K\|x - x'\| + |\alpha - \alpha'| + |\beta - \beta'| : \alpha' + \beta' \geq f(x') + g(x')\}$$

$$\leq \inf\{K\|x - x'\| + |\alpha - \alpha'| + |\beta - \beta'| : \alpha' \geq f(x'), \beta' \geq g(x')\}$$

$$= \inf\{K\|x - x'\| + |\alpha - \alpha'| + (g(x') - \beta)^+ : (x', \alpha') \in X \times \mathbb{R}, \alpha' \geq f(x')\}$$

$$\leq \inf\{2K\|x - x'\| + |\alpha - \alpha'| : (x', \alpha') \in X \times \mathbb{R}, \alpha' \geq f(x') + (g(x) - \beta)^+\}$$

$$\leq 2(d_{\text{epi}} f(x, \alpha) + d_{\text{epi}} g(x, \beta)).$$

This proves (10.6) and the theorem. \qed
The calculus rules for the $G$-subdifferential are now easy consequences of the general theorems of the first part.

**Theorem 10.13.** Let $X$ be a Banach space, and let the functions $f_1, \ldots, f_k$ be lower semicontinuous near $\bar{x}$ and finite at $\bar{x}$. Assume that the metric qualification condition (8.5) is satisfied near $\bar{x}$. Then

$$\partial_G f(\bar{x}) \subset \partial_G f_1(\bar{x}) + \cdots + \partial_G f_k(\bar{x}).$$

**Proof.** The proof repeats word for word the proof of part (b) of Theorem 9.10 with two small changes: instead of references to Proposition 9.6 and (9.1) we should refer to Proposition 10.9 and (10.3).

**Theorem 10.14.** Let $X, Y$ be Banach spaces and let $\varphi$ be a lower semicontinuous function on $X \times Y$. Set $f(x) = \inf_y \varphi(x, y)$, and $M(x) = \{y : \varphi(x, y) = f(x)\}$. Assume that $M(x) \neq \emptyset$ for all $x$ in a neighborhood of $\bar{x}$ and the set-valued mapping $x \mapsto M(x)$ is semicompact on a neighborhood of $\bar{x}$ (see the statement of Theorem 8.7). Then

$$\partial_G f(\bar{x}) \times \{0\} \subset \bigcup_{\tilde{y} \in M(\bar{x})} \partial_G \varphi(\tilde{x}, \tilde{y}).$$

**Remark 10.15.** The union in the last formula can be deleted if for some $\tilde{y} \in M(\bar{x})$ the set-valued mapping $M$ is lower semicontinuous at $(\tilde{x}, \tilde{y})$ in the sense that for any sequence $(x_n)$ converging to $\tilde{x}$ there are $y_n \in M(x_n)$ converging to $\tilde{y}$.

**Proof.** Consider $X \times \mathbb{R}$ and $X \times Y \times \mathbb{R}$ with the standard sum norms

$$\|(x, \alpha)\| = \|x\| + |\alpha| \quad \text{and} \quad \|(x, y, \alpha)\| = \|x\| + \|y\| + |\alpha|.$$

Take a $y \in Y$. Then

$$d_{epi} f(x, \alpha) = \inf \{\|x - u\| + |\alpha - \beta| : \beta \geq f(u)\}$$

$$= \inf \{\|x - u\| + |\alpha - \beta| : \beta \geq \inf_v \varphi(u, v)\}$$

$$= \inf \{\|x - u\| + |\alpha - \beta| : v \in Y, \beta \geq \varphi(u, v)\}$$

$$\leq \inf \{\|x - u\| + \|y - v\| + |\alpha - \beta| : \beta \geq \varphi(u, v)\}$$

$$= d_{epi} \varphi(x, y, \alpha).$$

If $\alpha \leq f(x)$, it easily follows from this calculation that

$$d_{epi} f(x, \alpha) = d_{epi} \varphi(x, y, \alpha)$$
if and only if $y = M(x)$. Set $g(x, \alpha) = \inf_y d_{\text{epi} \varphi}(x, y, \alpha)$, and let

$$M_g(x, \alpha) = \{y : g(x, \alpha) = d_{\text{epi} \varphi}(x, y, \alpha)\}.$$  

Thus $M_g(x, \alpha) = M(x)$ if $\alpha \leq f(x)$. We have furthermore

$$d_{\text{epi} f}(x, \alpha) \leq g(x, \alpha) \quad \text{and} \quad 0 = d_{\text{epi} f}(x, \alpha) = g(x, \alpha)$$

if $\alpha \geq f(x)$. As $\partial G$ is distance homotone, it follows that

$$\partial G d_{\text{epi} f}(\bar{x}, f(\bar{x})) \subset \partial G g(\bar{x}, f(\bar{x})).$$

Let $x^* \in \partial G f(\bar{x})$, that is, $(x^*, -1) \in \lambda \partial G d_{\text{epi} f}(\bar{x}, f(\bar{x})) \subset \lambda \partial G g(\bar{x}, f(\bar{x}))$ for some $\lambda > 0$. By Theorem 4.3 (or rather from (4.2) as we can take $v \in M(x)$) and Theorem 10.12, for any $\varepsilon > 0$, $\delta > 0$ and any weak*-neighborhood $V$ of zero in $X^*$ there are $x$ and $y$ such that $\|x - \bar{x}\| < \delta$, $|\alpha - f(\bar{x})| < \delta$, $y \in M_g(x, \alpha)$ and $(x^*, 0, -1) \in \partial G d_{\text{epi} \varphi}(x, y, \alpha)) + V \times (\varepsilon B_Y \star) \times [-\varepsilon, \varepsilon]$. But $\alpha$ cannot be greater than $\varphi(x, y)$ for otherwise the third component of any element of $\partial G d_{\text{epi} \varphi}(x, y, \alpha))$ is zero. Thus $y \in M(x)$. Passing to the limit as $\delta \rightarrow 0$, we get

$$(x^*, 0, -1) \in \bigcup_{\bar{y} \in M(\bar{x})} \lambda \partial G d_{\text{epi} \varphi}(\bar{x}, \bar{y}, \varphi(\bar{x}, \bar{y})) + V \times (\varepsilon B_Y \star) \times [-\varepsilon, \varepsilon].$$

Semicompactness of $M$ obviously implies norm compactness of $M(\bar{x})$. Together with Proposition 10.2 this implies that the union of $\partial G d_{\text{epi} \varphi}(\bar{x}, \bar{y}, \varphi(\bar{x}, \bar{y}))$ over $\bar{y} \in M(\bar{x})$ is weak*-compact and therefore, taking the intersection of the right-hand side of the inclusion over all $V$ and $\varepsilon$, we get

$$(x^*, 0, -1) \in \bigcup_{\bar{y} \in M(\bar{x})} \lambda \partial G d_{\text{epi} \varphi}(\bar{x}, \bar{y}, \varphi(\bar{x}, \bar{y}))$$

which completes the proof.

\[\square\]

11 Some other operations

In this section we mainly show how some other operations with functions, sets and set-valued mappings can be reduced to the two basic operations considered in the main body of the paper. We state corresponding calculus rules for subdifferentials, normal cones and coderivatives. In all cases proofs are very simple (and many relevant arguments can be found in the literature), so we give only very brief comments. We also emphasize every time the form of the corresponding metric qualification condition. Everywhere, if no special mention is made, the functions are lower semicontinuous, single-valued mappings are continuous, sets are closed as well as the graphs of set-valued mappings.
11.1 Operations with functions

11.1.1 Composition with a nonlinear mapping

Let $F : X \rightarrow Y$ (single valued), and let $\varphi$ be a function on $Y$. Set $f = \varphi \circ F$, that is, $f(x) = \varphi(F(x))$. We can represent $f$ as follows:

$$f(x) = \inf_y g(x, y), \quad \text{where } g(x, y) = \varphi(y) + \text{Ind}_{\text{Graph} F}(x, y).$$

This effectively reduces subdifferentiation of $f$ to the case of a sum. The marginal function part is trivial as in this case $M(x) = F(x)$. The representation can be directly used if we are interested in fuzzy calculus rules for elementary subdifferentials of $f$. For exact estimates involving limiting or approximate subdifferential we need a metric qualification condition for $g$ which reduces to

$$d_{\text{epi} f}(x, \alpha) \leq K d_{\text{epi} \varphi}(F(x), \alpha). \quad (11.1)$$

It is an easy matter to see that under this qualification condition we obtain the following estimate (for $\partial$ being either $\partial_{LF}$, if both $X$ and $Y$ are Asplund, or $\partial_G$):

$$\partial f(\bar{x}) \subset \bigcup_{y^* \in \partial \varphi(\bar{y})} D^* F(\bar{x}, \bar{y})(y^*) \quad (\bar{y} = F(\bar{x})). \quad (11.2)$$

11.1.2 Value function (marginal function with constraints)

By that we mean

$$f(x) = \inf \{\varphi(x, y) : y \in F(x)\},$$

where $F : X \Rightarrow Y$. We can rewrite this function as

$$f(x) = \inf_y (\varphi(x, y) + \text{Ind}_{\text{Graph} F}(x, y)).$$

Thus subdifferentiation of $f$ reduces to applications of calculus rules for marginal functions and sums. If we are interested in $\partial_{LF}$ or $\partial_G$, the latter needs a qualification condition, which obviously is the same as (11.1) (up to the additional dependence of $\varphi$ on $x$), and a semicompactness condition stated exactly as in Theorem 10.14 with $M(x) = \{y \in F(x) : f(x) = \varphi(x, y)\}$. The estimate for the subdifferential of $f$ we get in this case is

$$\partial f(\bar{x}) \subset \bigcup_{\bar{y} \in M(\bar{x})} \bigcup_{(x^*, y^*) \in \partial \varphi(\bar{x}, \bar{y})} (x^* + D^* F(\bar{x}, \bar{y}))(y^*). \quad (11.3)$$

(The first union can be deleted under a stronger assumption that $M$ is lower semicontinuous at some $(\bar{x}, \bar{y}), \bar{y} \in M(\bar{x})$.)
11.2 Operations with sets

11.2.1 Intersection of sets

Suppose $Q_1, \ldots, Q_k$ are closed subsets of $X$ and we are interested in the normal cone to their intersection in a certain $\overline{x} \in Q = \bigcap Q_i$. This can easily be found as $\text{Ind}_Q$ is the sum of indicators of $Q_i$, so all we need is to apply Theorems 4.2, 7.8, 8.6, 9.10 or 10.13 depending on what kind of normal cone we are talking about. In case of $\partial_L F$ or $\partial_G$ a suitable metric qualification condition is

$$d_Q(x) \leq K(d_{Q_1}(x) + \cdots + d_{Q_k}(x))$$

and the corresponding estimate for the normal cone to $Q$ is

$$N(Q, \overline{x}) \subseteq N(Q_1, x) + \cdots + N(Q_k, \overline{x}).$$

This inclusion is a basis for calculation of subdifferentials of maxima of finitely many functions as the epigraph of a maximum function is the intersection of epigraphs of the component functions.

11.2.2 Inverse image

Here we mean the set $Q = F^{-1}(S)$ where $F : X \Rightarrow Y$ and $S \subseteq Y$. The question is about the normal cone to $Q$ at $\overline{x} \in Q$. We have

$$\text{Ind}_Q(x) = \inf_y (\text{Ind}_S(y) + \text{Ind}_{\text{Graph} F}(x, y)),$$

so we again get a combination of a marginal function and a sum, as in the case of value function but with $\varphi$ not depending on $x$. Thus, we have (11.1) as an appropriate metric qualification condition and assuming semicompactness of the mapping $x \mapsto M(x) = F(x) \cap S$, get from (11.3)

$$N(F^{-1}(S), \overline{x}) \subseteq \bigcup_{y \in F(\overline{x})} \bigcup_{y^* \in N(S, y)} D^* F(\overline{x}, \bar{y})(y^*).$$

11.3 Operations with set-valued mappings

11.3.1 Composition

Given two set-valued mappings $F : X \Rightarrow Y$ and $G : Y \Rightarrow Z$, the value at $x$ of the composition $G \circ F : X \Rightarrow Z$ is defined as the union of all $G(y)$ corresponding to $y \in F(x)$. It follows that

$$\text{Ind}_{\text{Graph}} G \circ F(x, z) = \inf_y (\text{Ind}_{\text{Graph}} F(x, y) + \text{Ind}_{\text{Graph}} G(y, z)),$$
and again we have a combination of marginal function and sum. We have

\[ M(x, z) = F(x) \cap G^{-1}(z), \]

and the metric qualification condition (if we are interested in the coderivative of \( G \circ F \) at \((\overline{x}, \overline{z})\)) should be written as

\[ d_{\text{Graph}} G \circ F(x, z) \leq K \left( d_{\text{Graph}} F(x, y) + d_{\text{Graph}} G(y, z) \right) \quad (11.4) \]

for \( x, z \) close to \( \overline{x}, \overline{z} \) and all \( y \). If \( M \) is semicompact at \( \overline{x} \), then for the coderivative associated with either \( \partial_{LF} \) or \( \partial_G \) we have

\[ D^* (G \circ F)(\overline{x}, \overline{y})(z^*) \subset \bigcup_{\overline{y} \in M(\overline{x})} \bigcup_{y^* \in D^* G(\overline{y}, \overline{z})(z^*)} D^* F(\overline{x}, \overline{y})(y^*). \quad (11.5) \]

Again, the first union can be omitted if \( M \) is lower semicontinuous at \((\overline{x}, \overline{y})\).

### 11.3.2 Cartesian product

Given two set-valued mappings \( F, G : X \Rightarrow Y \), we consider \( \Phi : X \Rightarrow Y \times Y \) defined by

\[ \Phi(x) = F(x) \times G(x). \]

We can consider it as a composition of a \( \Psi(x, u) = F(x) \times G(u) \) and a linear mapping \( A : X \to X \times X \) defined by \( Ax = (x, x) \). For the latter we have

\[ D^* A(x^*, u^*) = A^*(x^*, u^*) = x^* + u^*. \]

It also clear that in this case \( M(x) = (x, x) \).

We therefore have

\[ D^* \Psi((x, u), (y, v))(y^*, v^*) = D^* F(x, y)(y^*) \times D^* G(u, v)(v^*). \]

Combining this with (11.5), we get

\[ D^* \Phi(x, (y, v))(y^*, v^*) = D^* F(x, y)(y^*) + D^* G(x, v)(v^*), \]

provided the metric qualification condition

\[ d_{\text{Graph}} \Phi(x, (y, z)) \leq K \left( d_{\text{Graph}} F(x, y) + d_{\text{Graph}} G(x, z) \right) \]

holds (easily following from (11.4)).
11.3.3 Sum of multifunctions

Let $F, G : X \Rightarrow Y$, set

$$\Phi(x) = F(x) + G(x).$$

We can view $\Phi$ as a composition $T \circ (F \times G)$, where $T : X \times X \rightarrow X$ is summation $T(y, v) = y + v$. Clearly

$$T^* y^* = (y^*, y^*).$$

An appropriate metric qualification condition (also following from (11.4)) can be

$$d_{\text{Graph}(F+G)}(x, y) \leq K(d_{\text{Graph}_F}(x, v) + \text{Graph}_G(x, y-v)).$$

If we assume that the mapping

$$M(x, y) = \{(v, w) : v \in F(x), w \in G(x), v + w = y\}$$

is semicompact, it follows that

$$D^*(F + G)(x, y)(y^*) = \bigcup_{(v, w) \in M(x, y)} (D^* F(x, v)(y^*) + D^* G(x, w)(y^*)).$$

12 Comments

1. Here is a brief chronology of events at the initial stage of developments.

1974. Dini–Hadamard subdifferential is introduced by Penot [57] and Bazaraa, Goode and Nashed [6].

1975. Clarke’s first publication [14] based on his 1973 dissertation. First calculus rules for generalized gradient of Lipschitz functions are established. In finite dimension, generalized normals appear as elements of convex closure of the proximal normal cone. (The latter however is not introduced as a separate entity.)

1976. (a) In [51] Mordukhovich explicitly introduces non-convexified limiting proximal normal cone (using rather topological, not sequential language) and he proves necessary conditions for an optimal control problem with transversality conditions involving elements of this cone.

    (b) Ekeland and Lebourg [22] introduce Fréchet $\varepsilon$-subdifferentials ($\varepsilon$-supports), in their proof that a space with a Fréchet smooth bump is Asplund, and prove dense Fréchet $\varepsilon$-subdifferentiability of l.s.c. function on such spaces.

1980. Kruger and Mordukhovich publish a note [47] containing the statement of
the extremal principle for Fréchet $\varepsilon$-subdifferentials and their topological closures
in Fréchet smooth spaces\(^5\).

1981. (a) Kruger deposits preprints of three papers corresponding to his dissertation. One of them [45], already mentioned in the introduction, contains calculus
of limiting Fréchet subdifferentials for l.s.c. functions on Fréchet smooth spaces.

(b) Two preprints of Ioffe [29, 30], one with a fuzzy calculus of the Dini–Hadamard $\varepsilon$-subdifferential for functions on Gâteaux smooth spaces and the other
containing definitions and calculus rules for several types of approximate subdifferen-
tials for functions on arbitrary Banach spaces. The summary of the results (mainly
in the Lipschitz setting) is published in [31].

It should be noted that the concept of approximate subdifferential underwent a
substantial evolution till the final definition presented here was worked out (see
[7, 35]). Originally, the starting point was subdifferentiation on finite dimensional
subspaces. Proposition 10.7 implies that both approaches (through finite dimen-
sional and separable subspaces) are in principle equivalent. The main difference
of the present definition of the $G$-subdifferential (coinciding with the “nucleus”
introduces in [35] and for the first time used as the “right” definition in [7]) with
the original definition of the $G$-subdifferential is the absence of the weak*-closure
operation at the concluding stage of the definition.

Another clarifying remark is that at the initial stage the main instrument of anal-
ysis was Ekeland’s variational principle which made inevitable the use of $\varepsilon$-sub-
differentials at the lower level. The appearance of the smooth variational principle
of Borwein-Preiss [10] and Deville–Godefroy–Zizler radically changed the situ-
aton as it became possible to work with “pure” subdifferentials [25, 36] without
any loss.

2. The concept of trustworthiness in a somewhat different form first appeared in
[31] and was more thoroughly studied in [32]. In fact, there were two different
types of trustworthiness in the original definition, one for the Fréchet and the other

\(^5\) In many publications this paper and more rarely [48], both existing only in Russian and difficult
to obtain, are cited as where the limiting Fréchet subdifferential in Banach spaces has been
originated. This is a mistake. The construction introduced in [47] is the limit as $\varepsilon \rightarrow 0$ of the
closure of the graph of the Fréchet $\varepsilon$-subdifferential in the product of the norm topology in $X$
and the weak*-topology in $X^*$, while [48] is completely finite dimensional and even there the
definition is given in topological, rather than sequential terms, precisely as in [47, 51]. In fact,
the subdifferential defined in [47] coincides on Fréchet smooth spaces with the approximate
$A$-subdifferential of [31]. (In particular, Treiman’s counterexample of [65] is applicable to it as
well.) As we have mentioned, the sequentially limiting Fréchet subdifferential was introduced
in [45].
for the Dini–Hadamard \( \varepsilon \)-subdifferential with proofs of dense subdifferentiability for both. Shortly afterwards Fabian [23] produced an elegant proof that those concepts were actually equivalent to dense subdifferentiability of every lower semi-continuous function. Along with the proof by Borwein and Preiss [10] that an l.s.c. function is densely \( \beta \)-subdifferentiable on a space with a \( \beta \)-differentiable norm, these results suggested the necessity of a stronger unified concept as presented here.

An approximate mean value theorem was proved by Zagrodny [66] for Clarke’s generalized gradients. Later Thibault and Zagrodny [64] demonstrated that the theorem (and actually its original proof) can be easily extended to a number of other subdifferentials. Many interesting applications of the mean value theorem is discussed in [13].

An interesting approximate version of Rolle’s theorem for the proximal subdifferential on a Hilbert space was proved by Azagra, Ferrera and Lopez-Mesas [4]: If \( f \) is a continuous function on the ball of radius \( R \) in a Hilbert space such that \( |f(x)| \leq \varepsilon \) on the sphere, then for any \( \alpha > 0 \) there is a point \( x \in \text{int } B \) and an \( x^* \in \partial p(x) \) with \( \|x^*\| < (\varepsilon/R) + 1 \). To emphasize nontriviality of this results, it is shown that “exact” versions of Rolle’s theorem does not hold even for the generalized gradients (see also [8]).

In the context of variational analysis the decoupling penalization techniques was first applied in [30, 32]. But I am sure it has been somewhere used much earlier for one or another purpose.

Four of the five mentioned properties in the equivalence theorem were originally established for proximal, Fréchet or Dini–Hadamard subdifferentials on Fréchet smooth or Gâteaux smooth spaces: local fuzzy minimization principle in [29, 32], global fuzzy minimization principle in [67], multi-directional mean value inequality in [16] and the extremal principle was stated in [47] and proved in [45]. The linear set minimization rule was introduced in [49] for a generally defined subdifferential.

The equivalence theorem was proved by Zhu [68] for viscosity \( \beta \)-subdifferentials (see also [9] for equivalence of trustworthiness and extremal principle for “controlled” \( \beta \)-subdifferentials), then Ioffe [37] showed that the theorem can be extended to all subdifferentials satisfying the properties close to (S1)–(S6) and Lassonde [49] added several more equivalent properties and emphasized that only (S3), (S4) and (S6) are really needed to prove the equivalence theorem (not including the extremal principle, we add – to include the latter we need geometric consistence). The interpretation of the “uniform lower semicontinuity” property of [7] in terms of robust optimization was de facto also suggested in this Lassond’s paper. We refer to Borwein–Zhu monograph [13] for extensive discussions on four equivalent properties (local and global fuzzy minimization rules, multi-directional
mean value inequality and extremal principle) for the Fréchet subdifferential on Fréchet smooth spaces.

It is easy to see from the proof of the equivalence theorem that in the definition of trustworthiness the Lipschitz function can be assumed even convex continuous. This leads to still another (and probably the most elegant) equivalent characterization of trustworthiness for a geometrically consistent subdifferential (due to Penot – see e.g. [59]): if a convex continuous function \( \varphi \) attains a local minimum on a closed set \( Q \) at \( \bar{x} \), then for any \( \varepsilon > 0 \) there are \( x \in B(\bar{x}, \varepsilon), \ u \in Q \cap B(\bar{x}, \varepsilon), \ x^* \in \partial f(x), \ u^* \in N(Q, u) \) such that \( \|x^* + u^*\| < \varepsilon \).

Tightness is a relatively new concept: it was defined in [39]. Among the major subdifferentials only the generalized gradient is not tight. This explains why it does not have some good properties (e.g. the scalarization formula, not discussed here). Tightness is crucial for the uniqueness property of the \( G \)-subdifferential. Without tightness, we can at best state that the \( G \)-subdifferential is minimal among the subdifferential having other properties mentioned in the beginning of § 10.

3. The progress of the calculus of limiting Fréchet and approximate subdifferentials for lower semicontinuous functions was to a large extent associated with the developments of qualification conditions. First versions of calculus rules for non-convex “robust” subdifferentials (including formulas for sums and composition with a single valued mapping), appeared in 1981: for approximate subdifferentials in general Banach spaces in [29, 31] and for limiting Fréchet subdifferentials on Fréchet smooth spaces in [45]. In [29,31] all functions but possibly one were Lipschitz; in [45] (and later in [46] based on [45]) the assumptions on the functions were the same as in Rockafellar’s treatment of generalized gradients in [60]: all functions but possibly one are directionally Lipschitz and

\[
\text{dom} \ f^\uparrow_1 \cap \text{int} (\text{dom} \ f^\uparrow_2) \neq \emptyset,
\]

where \( f^\uparrow \) stands for Rockafellar’s directional derivative.

In [33] the condition (12.1) for l.s.c. functions on \( \mathbb{R}^n \) was weakened to

\[
\text{dom} \ f^\uparrow_1 - \text{dom} \ f^\uparrow_2 = X
\]

Then, shortly afterwards, Mordukhovich [52] and Rockafellar [62] suggested a still weaker qualification condition for functions on \( \mathbb{R}^n \):

\[
(x_1^*, 0) \in N(\text{epi} \ f, (\bar{x}, f(\bar{x}))), \ x_1^* + x_2^* = 0 \quad \Rightarrow \quad x_1^* = x_2^* = 0.
\]

Later in the 90s Mordukhovich and Shao [54–56] systematically used this condition combined with one or another compactness property, replacing directional Lipschitzness, for the limiting Fréchet subdifferential on Asplund spaces.
This combination works well also for the $G$-subdifferential. However, condition (12.2) has one serious, partly aesthetic, drawback: it is stated in the same terms as the result. Indeed, if we already know normal cones or subdifferentials why should we verify the qualification condition instead of checking directly whether the formula we are interested in holds?

The metric qualification condition is stated in terms of the original data and is free from this defect. This condition appeared first in [35] and was instrumental in the development of the calculus of $G$-subdifferential. It was shown in particular in [35], for the case of sum of two l.s.c. functions on an arbitrary Banach space, that if one of the functions is directionally Lipschitz and (12.2) holds, then the linear metric qualification condition is satisfied. The condition that a function is directionally Lipschitz is actually a sort of a compactness condition. In the course of further developments the same was proved for the entire spectrum of compactness conditions [41, 43, 44] (see also [59]). Versions of the metric qualification conditions were systematically used by Jourani and Thibault [42–44] to prove chain rules and calculus rule for coderivatives of set-valued mappings.

In fact, the metric qualification condition is substantially weaker than a combination of (12.2) and one or another compactness property as can be seen from the following elementary example in $\ell_2$: let

$$Q_1 = \{ x = (\xi_1, 0, 0, \ldots) : \xi_1 \geq 0 \} \quad \text{and} \quad Q_2 = \{ x = (\xi_1, 0, 0, \ldots) : \xi_1 \leq 0 \}.$$

Then neither (12.2) nor any compactness condition for normal cones to $Q_i$ at zero is satisfied, but the metric qualification condition is trivially valid.

The example shows that even for convex objects the metric qualification condition can be applied when all standard conditions of convex analysis (even finite dimensional) fail (this fact was mentioned already in [41]). Thus the linear metric qualification condition is the weakest among the available qualification conditions, even for convex objects in case of a finite dimensional $X$ (when some more advanced compactness properties are automatic).

Acknowledgments. I am thankful to A. Kruger for getting me acquainted with his 1981 preprint [45].

Bibliography


Received May 4, 2011; accepted November 21, 2011.

Author information
Alexander D. Ioffe, Department of Mathematics, Technion, Haifa 32000, Israel.
E-mail: ioffe@math.technion.ac.il