

# Necessary Conditions for Nonsmooth Optimization Problems with Operator Constraints in Metric Spaces

**Boris S. Mordukhovich\***

*Department of Mathematics, Wayne State University,  
Detroit, Michigan 48202, USA  
boris@math.wayne.edu*

**Libin Mou**

*Department of Mathematics, Bradley University,  
Peoria, Illinois 61625, USA  
mou@bradley.edu*

*Dedicated to Stephen Simons on the occasion of his 70th birthday.*

Received: June 9, 2008

Revised manuscript received: October 25, 2008

This paper concerns nonsmooth optimization problems involving operator constraints given by mappings on complete metric spaces with values in nonconvex subsets of Banach spaces. We derive general first-order necessary optimality conditions for such problems expressed via certain constructions of generalized derivatives for mappings on metric spaces and axiomatically defined subdifferentials for the distance function to nonconvex sets in Banach spaces. Our proofs are based on variational principles and perturbation/approximation techniques of modern variational analysis. The general necessary conditions obtained are specified in the case of optimization problems with operator constraints described by mappings taking values in approximately convex subsets of Banach spaces, which admit uniformly Gâteaux differentiable renorms (in particular, in any separable spaces).

*Keywords:* Variational analysis, generalized differentiation, optimization in metric spaces, necessary optimality conditions, approximately convex functions and sets

*2000 Mathematics Subject Classification:* 49J53, 49J52, 49K27, 90C48

## 1. Introduction

A vast majority of problems considered in optimization theory are described in Banach (if not finite-dimensional) spaces, where the linear structure is crucial to employ conventional tools of variational analysis and (generalized) differentiation for deriving necessary optimality conditions and subsequently developing numerical algorithms. On the other hand, there is a number of remarkable classes of problems particularly important for optimization, control, and their various applications that admit adequate descriptions

\*The research of the first author was partially supported by the USA National Science Foundation under grants DMS-0304989 and DMS-0603846 and by the Australian Research Council under grant DP-04511668.

in spaces with *no linear structures*; see, e.g., [5, 15, 17, 18, 20, 21] and the references therein.

In this paper we pay the main attention to deriving first-order necessary optimality conditions for a general class of optimization problems with *operator constraints* in *complete metric spaces*. The basic problem is described as follows:

$$\begin{cases} \text{minimize} & \varphi(w) \text{ with } w \in W \\ \text{subject to} & f(w) \in \Theta, \end{cases} \quad (1)$$

where  $(W, \rho)$  is a complete metric space with the metric  $\rho$ , where  $\varphi: W \rightarrow \overline{\mathbb{R}} := (-\infty, \infty]$  is a lower semicontinuous (l.s.c.) extended-real-valued cost function, and where  $f: W \rightarrow X$  is a continuous operator constraint mapping taking values in a closed subset  $\Theta$  of a Banach space  $(X, \|\cdot\|)$  equipped with the norm  $\|\cdot\|$ . Note that the case of additional geometric constraints  $w \in \Omega$  given by a closed subset  $\Omega \subset W$  can be easily reduced to the basic form (1) in the complete metric space  $(\Omega, \rho)$ .

Recently problem (1) has been considered in [17] in the case when  $\varphi$  is a *continuous* function and when  $\Theta$  is a *convex* subset of a Banach space  $X$  whose topological dual space  $X^*$  is *strictly convex* (or *rotund* in the *norm* topology; see, e.g., [7]). A version of the abstract multiplier rule obtained in [17] has been applied in [18] to derive a maximum principle for a general deterministic optimal control problem with state constraints.

Let us particularly emphasize that the convexity assumption on the constraint set  $\Theta$  imposed in [17] is clearly a restriction from both viewpoints of optimization theory and applications. The primary goal of this paper is to establish necessary conditions for local optimal solutions to problem (1) with *no convexity* requirements imposed on the constraint set  $\Theta$  and/or continuity assumptions on the cost function  $\varphi$ . We derive such optimality conditions in the general case of complete metric spaces, *lower semicontinuous* cost functions  $\varphi: W \rightarrow \overline{\mathbb{R}}$ , and continuous mappings  $f: W \rightarrow X$  taking values in closed subsets  $\Theta$  of arbitrary Banach spaces. Furthermore, we obtain efficient specifications of our general necessary optimality conditions in the case of *approximately convex* subsets  $\Theta$  of Banach spaces  $X$  admitting *uniformly Gâteaux differentiable renorms* (equivalent to the *weak\* uniform rotundity* ( $W^*UR$ ) of  $X^*$  [7]) that encompass, in particular, every *separable* Banach space. The latter result essentially improves the multiplier rule derived in [17] for problems with convex constraint sets considered therein in the more restrictive setting.

To establish necessary optimality conditions for the general problem (1), we employ the notions of *subderivates* for functions and mappings on *metric spaces* and also of the (topological and sequential) *outer regular subdifferentials* introduced and applied below for the *distance functions* of closed subsets in Banach spaces. The latter abstract subdifferential notions are defined *axiomatically* via several required properties that hold in natural settings for major subdifferential constructions encountered in variational analysis and optimization.

The rest of the paper is organized as follows. In Section 2 we define and discuss the notions of *approximate (sub)derivates* and *strict (sub)derivates* for generally nonsmooth mappings and extended-real-valued functions on metric spaces as well as of abstract *outer regular subdifferentials* for the distance functions in Banach spaces.

Section 3 presents the main result of the paper establishing *first-order necessary optimality conditions* for the general problem (1) with operator constraints. The result obtained is expressed in terms of the strict subderivates of  $\varphi$  and  $f$  in (1) and of the outer subdifferentials for the distance function  $d_\Theta$  of the constraint set  $\Theta$  defined in Section 2. The proof is based on employing the Ekeland variational principle and advanced perturbation techniques of variational analysis via the strict derivate construction and the appropriate properties of outer subgradients postulated and justified in the previous section.

Section 4 is devoted to the description and certain useful properties of extended-real valued *approximately convex functions* in Banach spaces introduced in [23]. These constructions are closely related to some other remarkable notions of generalized convexity, which play an important role in variational analysis and optimization. We establish new properties of approximately convex functions and sets in terms of generalized differential constructions of variational analysis paying the main attention to a modified version of approximate convexity *around* the reference points.

In the concluding Section 5 we apply the general necessary optimality conditions established in Section 3 and the properties of approximately convex functions and sets from Section 4 to derive efficient specifications of the general result in the case of problem (1) with *approximately convex constraint sets*  $\Theta$  in Banach spaces  $X$  admitting *uniformly Gâteaux differentiable renorms*. As mentioned, this class of spaces contains all *separable* Banach spaces particularly important for variational analysis and its applications to optimization and related topics. We show that the *major subdifferential constructions* in variational analysis – that are known to be the same for the distance functions of approximately convex sets – enjoy the required properties of the topological and sequential outer subdifferentials, which agree in the Banach spaces under consideration and allow us to efficiently apply the main result of Section 3. Furthermore, the latter result is constructively specified for approximately convex sets and expressed in the form similar to the case of (full) convexity developed in [17]. We also discuss various modifications and extensions of the proofs and results developed in Sections 4 and 5.

Throughout the paper we mainly use standard notation of variational analysis; see, e.g., [20, 26]. Recall that  $\mathbb{N} = \{1, 2, \dots\}$ , that  $\mathbb{B}$  and  $\mathbb{B}^*$  stand for the closed unit ball in the Banach space in question and its topological dual, that  $B(\bar{x}; r)$  is the closed ball centered at  $\bar{x}$  with radius  $r > 0$ , and that  $x \xrightarrow{w^*} x^*$  signifies the weak\* convergence in the dual  $X^*$  to a Banach space  $X$  with the canonical paring  $\langle \cdot, \cdot \rangle$  between the primal and dual spaces. We use the notation  $F: X \rightrightarrows Y$  for *set-valued mappings* with the *graph*

$$\text{gph } F := \{(x, y) \in X \times Y \mid y \in F(x)\}$$

to distinguish them from single-valued mappings denoted as usual by  $f: X \rightarrow Y$ . Given a set-valued mapping  $F: X \rightrightarrows X^*$  between a Banach space and its dual, the symbol

$$\begin{aligned} \text{Lim sup}_{x \rightarrow \bar{x}} F(x) := \left\{ x^* \in X^* \mid \exists \text{ a bounded net } (x_\nu, x_\nu^*) \in \text{gph } F \right. \\ \left. \text{with } x_\nu \rightarrow \bar{x} \text{ and } x_\nu^* \xrightarrow{w^*} x^* \right\}. \end{aligned} \tag{2}$$

signifies the *topological Painlevé-Kuratowski outer limit* of  $F$  as  $x \rightarrow \bar{x}$ . If the nets in (2) are replaced by sequences, we call (2) the *sequential Painlevé-Kuratowski outer limit* of

$F$  as  $x \rightarrow \bar{x}$  and use the same notation while indicating each time what kind of the limit is under consideration in the specific situation.

Given further a nonempty subset  $\Theta \subset X$  of a Banach space  $X$ , denote by  $\text{cl } \Theta$  its *closure*, by  $\text{bd } \Theta$  its *boundary*, by  $\text{cone } \Theta := \{\alpha x \mid \alpha \geq 0, x \in \Theta\}$  its *conic hull*, and by

$$d_{\Theta}(x) := \inf \{\|x - y\| \mid y \in \Theta\} \quad (3)$$

the *distance function* associated with  $\Theta$ . We use the symbol  $\Theta' := X \setminus \Theta$  to signify the *complement* of  $\Theta$  in  $X$  and the symbol  $x \xrightarrow{\Theta} \bar{x}$  to indicate that  $x \rightarrow \bar{x}$  with  $x \in \Theta$ . By convention, let  $\alpha \emptyset := \emptyset$  for  $\alpha \in \mathbb{R}$  with  $\alpha \neq 0$  and  $0 \cdot \emptyset := 0$ . We always suppose that all the extended-real-valued functions  $\psi: W \rightarrow \overline{\mathbb{R}}$  under consideration are *proper*, i.e.,

$$\text{dom } \psi := \{w \in W \mid \psi(w) < \infty\} \neq \emptyset.$$

## 2. Subderivates and Subdifferentials

In this section we introduce and discuss the major notions of *generalized differentiation* used in this paper: the *approximate (sub)derivates* and *strict (sub)derivates* for nonsmooth mappings and extended-real-valued functions on metric spaces as well as of the axiomatically defined *outer regular subdifferentials* for the distance functions in normed spaces.

Let us start with the constructions of subderivates and derivates and define them for mappings  $f: W \rightarrow X$  on metric spaces  $(W, \rho)$  with values in normed spaces  $(X, \|\cdot\|)$ . Although the definitions below do not use the completeness of the domain and image spaces, these properties are essential in the proofs of our main results. Thus we always assume that the underlying *domain* metric space  $W$  is *complete* and the *image* space  $X$  is *Banach*. Furthermore, the presented subderivate/derivate definitions are automatically applied to *extended-real-valued* functions  $\varphi: W \rightarrow \overline{\mathbb{R}}$  finite at the reference points.

Given  $f: W \rightarrow X$  and  $\bar{w} \in W$ , denote by  $\mathcal{S}(\bar{w})$  the sets of *sequences*  $(w^i, t^i)_{\mathbb{N}}$  such that  $w^i \in W$ ,  $t^i \in (0, \infty)$ , and  $\rho(w^i, \bar{w}) \leq t^i \downarrow 0$  as  $i \rightarrow \infty$ .

### Definition 2.1 (Subderivates and derivates of mappings on metric spaces).

Let  $f: W \rightarrow X$ ,  $\bar{w} \in W$ , and  $\mathcal{S}(\bar{w})$  be as described above. Then:

- (i) Given  $\varepsilon \geq 0$ , we say that  $v \in X$  is an  $\varepsilon$ -*subderivate* of  $f$  at  $\bar{w}$  if there is a sequence  $(w^i, t^i) \in \mathcal{S}(\bar{w})$  such that

$$\limsup_{i \rightarrow \infty} \left\| \frac{f(w^i) - f(\bar{w})}{t^i} - v \right\| \leq \varepsilon. \quad (4)$$

We call  $v$  a *subderivate* of  $f$  at  $\bar{w}$  if  $\varepsilon = 0$  and *approximate subderivate* of  $f$  at  $\bar{w}$  if  $\varepsilon > 0$ . The collection of  $\varepsilon$ -subderivates of  $f$  at  $\bar{w}$  is called the  $\varepsilon$ -*derivate* (*derivate* and *approximate derivate*, respectively) of  $f$  at this point and is denoted by  $D_{\varepsilon}f(\bar{w})$ .

- (ii) We say that  $v \in X$  is a *strict subderivate* of  $f$  at  $\bar{w}$  if for every sequence  $w_k \rightarrow \bar{w}$  there is a sequence  $\varepsilon_k \downarrow 0$  as  $k \rightarrow \infty$  such that  $v \in D_{\varepsilon_k}f(w_k)$  for all  $k \in \mathbb{N}$ . The collection of strict subderivates of  $f$  at  $\bar{w}$  is called the *strict derivate* of  $f$  at this point and is denoted by  $D_s f(\bar{w})$ .

The above construction of strict derivate slightly extends the one from [17], where the sequence  $\varepsilon_k$  is replaced by a positive function  $\varepsilon(w) \downarrow 0$  as  $w \rightarrow \bar{w}$ . Note that the derivate and strict derivate have certain similarities with the classical derivative and strict derivative of mappings between Banach spaces, while they are different even for smooth real-valued functions  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  in which case

$$D_\varepsilon\varphi(\bar{w}) = [-|\varphi'(\bar{w})| - \varepsilon, |\varphi'(\bar{w})| + \varepsilon] \quad \text{as } \varepsilon \geq 0$$

and  $D_s\varphi(\bar{w}) = [-|\varphi'(\bar{w})|, |\varphi'(\bar{w})|]$ .

On the other hand, the derivate constructions from Definition 2.1 make sense for heavily discontinuous mappings and extended-real-valued functions. We have, e.g.,

$$D_\varepsilon\varphi(w) = [-1 - \varepsilon, 1 + \varepsilon] \quad \text{and} \quad D_s\varphi(w) = [-1, 1] \quad \text{as } w \in \mathbb{R}, \varepsilon \geq 0$$

for the function  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  equal to  $w$  at rational numbers and to  $1 + w$  otherwise.

It is worth mentioning that there is a number of pointwise *calculus rules* available for the strict derivate of mappings between both finite-dimensional and infinite-dimensional spaces. They are not needed in this paper and will be presented in subsequent publications.

Let us next introduce the notions of (topological and sequential) *outer regular subdifferentials* for the class of *distance functions*  $\varphi = d_\Theta: X \rightarrow \mathbb{R}$  defined in (3), where  $\Theta \subset X$  is a closed subset of a Banach space; in fact, we apply these subdifferential constructions just to the distance function of *the constraint set*  $\Theta$  in the original problem. Note that the (Lipschitz continuous) distance functions play a fundamental role in subdifferential theory and variational analysis generating subdifferentials of extended-real-valued functions, which are not needed in this paper; see, e.g., [6, 12, 20, 29] for more details and references.

Given  $\Theta \subset X$  and  $\bar{x} \in \Theta$ , an *outer neighborhood*  $U$  of  $\bar{x}$  relative to  $\Theta$  is defined by

$$U = \Theta' \cap \{x \in X \mid \|x - \bar{x}\| < r\} \quad \text{for some } r > 0,$$

where  $\Theta'$  is the complement to  $\Theta$  in  $X$ . By an *abstract outer regular subdifferential* of the distance function  $d_\Theta: X \rightarrow \mathbb{R}$  around a given point  $\bar{x} \in \Theta$  we understand a set-valued mapping  $\mathcal{D}d_\Theta: U \rightrightarrows X^*$  defined at  $\bar{x}$  and on some outer neighborhood  $U$  of  $\bar{x}$  relative to  $\Theta'$  that satisfies several properties formulated and discussed below including the major outer regularity requirement. We present two generally different versions of the required properties, *topological* and *sequential*, which depend on the (topological or sequential) type of the *weak\** convergence in the dual space  $X^*$  and generate the corresponding notions of topological and sequential outer regular subdifferentials.

Observe that, for a given subdifferential  $\mathcal{D}d_\Theta$  on a Banach space  $X$ , the topological and sequential properties defined below are *equivalent* provided that the dual unit ball  $B^*$  is *sequentially weak\** compact in  $X^*$ . This is the case of all Banach spaces admitting a *Gâteaux differentiable renorm* at nonzero points as well as all *Asplund generated* spaces; the latter class includes every *Asplund* space and every *weakly compactly generated* (WCG) space and thus all reflexive and all separable Banach spaces. We refer the reader to the classical texts [7, 8] and to the paper [11], where similar relations between topological and sequential properties are considered in detail in the framework of variational analysis.

**Definition 2.2 (Outer robustness).** Given  $\bar{x} \in \Theta$ , we say that  $\mathcal{D}d_\Theta$  is *topologically outer robust* around  $\bar{x}$  if there exists an outer neighborhood  $U$  of  $\bar{x}$  relative to  $\Theta$  such that for every  $x \in U$  we have the inclusion

$$\mathcal{D}'d_\Theta(x) := \text{Lim sup}_{u \xrightarrow{\Theta'} x} \mathcal{D}d_\Theta(u) \subset \mathcal{D}d_\Theta(x), \quad (5)$$

where  $\text{Lim sup}$  stands for the topological outer limit (2) relative to  $\Theta'$ . If (5) holds with the replacement of the topological outer limit by the sequential one, we say that  $\mathcal{D}d_\Theta$  is *sequentially outer robust* around  $\bar{x}$ .

Note that the topological outer robustness property implies the sequential one but not vice versa. It is also obvious that these properties are always satisfied around interior points of  $\Theta$ , since the left-hand side set in (5) is empty in this case. For boundary points of *any* closed sets, the outer robustness (both topological and sequential versions) holds for the *generalized gradient* by Clarke [6] and for the “*approximate*” *G-subdifferential* by Ioffe [12] in arbitrary *Banach spaces* as well as for the *basic/limiting subdifferential* by Mordukhovich [20] in *WCG Banach spaces* (not necessarily Asplund); see Theorem 3.60 and the discussions after its proof in [20, pp. 323–326]. We can similarly justify the outer robustness in *WCG Banach spaces* for certain modifications of the limiting subdifferential: namely, for the right-sided subdifferential introduced in [22] (see also [20, Subsection 1.3.3]) and the closely related outer subdifferential of [13], and also for the sequential limiting subdifferential developed in [10] in the case of Asplund generated spaces.

The next required properties (topological and sequential) of  $\mathcal{D}d_\Theta$  are more selective than the corresponding outer robustness and depend, for specific subdifferentials, on the set  $\Theta \subset X$  and the point  $\bar{x} \in \Theta$  under consideration.

**Definition 2.3 (Outer regularity).** Given  $\bar{x} \in \Theta$ , we say that  $\mathcal{D}d_\Theta$  is *topologically outer regular* at  $\bar{x}$  if every sequence  $x_k \xrightarrow{\Theta'} \bar{x}$  as  $k \rightarrow \infty$  has a subnet  $S$  such that the topological Painlevé-Kuratowski outer limit

$$\text{Lim sup}_{x \xrightarrow{S} \bar{x}} \mathcal{D}d_\Theta(x) \text{ is a singleton in } X^*. \quad (6)$$

We say that  $\mathcal{D}d_\Theta$  is *sequentially outer regular* at  $\bar{x}$  if  $S$  can be selected as a subsequence and the topological outer limit in (6) is replaced by a sequential one.

Note that the singleton in (6) generally *depends* on the chosen subnet/subsequence  $S$ . Similarly to the case of outer robustness, observe that the topological outer regularity property implies its sequential counterpart but not vice versa and that these properties obviously hold for interior points  $\bar{x}$  of any set  $\Theta$ .

If  $\Theta$  is “smooth” around  $\bar{x} \in \text{bd } \Theta$  (in the sense that  $d_\Theta$  is smooth around this point), then the outer regularity properties obviously hold for any natural subdifferentials  $\mathcal{D}d_\Theta$  on Banach spaces such that  $\mathcal{D}$  reduces to the classical derivative for smooth functions. We show in Section 5 that all the *major subdifferentials* in variational analysis are outer regular at any points of *approximately convex* sets in Banach spaces admitting *Gâteaux*

*differentiable* renorms. This implies, in particular, the outer regularity of the classical *subdifferential of convex analysis* in the case of convex sets in Definition 2.3.

Another major property required for the abstract subdifferentials considered in this paper is the *Extended Mean Value Inequality* (EMVI), which is a weak extended form of the mean value theorem in generalized differentiation.

**Definition 2.4 (Extended mean value inequality).** We say that the *Extended Mean Value Inequality* (EMVI) holds for  $\mathcal{D}d_\Theta$  around  $\bar{x} \in \Theta$  if there exist an outer neighborhood  $U$  of  $\bar{x}$  relative to  $\Theta$ , a function  $\omega: U \times [0, 1) \rightarrow [0, \infty)$  with  $\omega(x, \tau) \downarrow 0$  as  $(x, \tau) \rightarrow (\bar{x}, 0^+)$ , and a dense subset  $S \subset U$  such that for any  $x, u \in S$  we can find  $v \in (x + \|u - x\|\mathbb{B}) \cap U$  and  $x^* \in \mathcal{D}d_\Theta(v)$  satisfying

$$d_\Theta(u) - d_\Theta(x) \leq \langle x^*, u - x \rangle + \|u - x\| \omega(x, \|u - x\|). \tag{7}$$

The case of  $\omega \equiv 0$  in (7) corresponds to the conventional *Mean Value Inequality* (MVI) and holds for the majority of known subdifferentials of Lipschitz continuous functions useful in applications; see, e.g., [1, 5, 6, 10, 20, 26, 27, 28] and the references therein. Considering a *dense* subset  $S \subset X$  in Definition 2.4 allows us to cover the *sequential limiting subdifferential* on *Asplund generated spaces* in [10] for which the MVI is proved relative to an Asplund space sitting densely in  $X$ . Thus the extended inequality (7) is a natural subdifferential property, which does not impose any restrictions on the class of subdifferentials used in what follows. Observe that the EMVI property from Definition 2.4 is not a limiting one and hence does not have topological and sequential versions as those from Definition 2.2 and Definition 2.3.

Combining the above requirements on  $\mathcal{D}d_\Theta$  with another property that must be always fulfilled, we arrive at the following definition of the topological and sequential *abstract outer regular subdifferentials* for the class of distance functions under consideration.

**Definition 2.5 (Abstract outer regular subdifferentials of distance functions).** Given a nonempty set  $\Theta \subset X$  and a point  $\bar{x} \in \Theta$ , we say that  $\mathcal{D}d_\Theta$  is a *topological outer regular subdifferential* of the distance function  $d_\Theta$  around  $\bar{x}$  if the sets  $\mathcal{D}d_\Theta(x) \subset X^*$  are defined at least at  $\bar{x}$  and on some outer neighborhood  $U$  of this point relative to  $\Theta$  and the following properties are satisfied:

- (P1)  $\mathcal{D}d_\Theta(x) \subset \mathbb{B}^*$  for all  $x \in U$ ;
- (P2)  $\mathcal{D}d_\Theta$  is topologically outer robust around  $\bar{x}$ ;
- (P3)  $\mathcal{D}d_\Theta$  is topologically outer regular at  $\bar{x}$ ;
- (P4) The extended mean value inequality holds for  $\mathcal{D}d_\Theta$  around  $\bar{x}$ .

We say that  $\mathcal{D}d_\Theta$  is a *sequential outer regular subdifferential* of  $d_\Theta$  around  $\bar{x}$  if it satisfies properties (P1), (P4) and the sequential versions of properties (P2) and (P3) from Definition 2.2 and Definition 2.3, respectively.

Note that there are several versions of axiomatically defined abstract subdifferentials in nonsmooth analysis; see, e.g., [1, 12, 14, 19, 20, 28]. Both topological and sequential outer regular subdifferentials of Definition 2.5 are essentially different from all the known constructions. The major differences consist of considering *sets* (via their distance functions in contrast to arbitrary functions) and paying the main attention to *outer* properties of subdifferentials that deal with *out-of-set* points. In this approach

the validity of the imposed subdifferential requirements and their realization for specific subdifferentials depend on the set and its boundary point in question; see the discussions and examples presented above.

### 3. Necessary Optimality Conditions for General Problems

In this section we establish the main result of the paper providing first-order necessary optimality conditions for the general problem (1) via the *strict derivate* and *outer regular subdifferential* constructions introduced and discussed in Section 2.

**Theorem 3.1 (Necessary conditions for constrained optimization in metric spaces).** *Let  $\bar{w}$  be a local minimizer for problem (1), where  $(W, \rho)$  is a complete metric space and  $(X, \|\cdot\|)$  is a Banach space,  $\varphi: W \rightarrow \overline{\mathbb{R}}$  is finite at  $\bar{w}$  and l.s.c. around this point while  $f: W \rightarrow X$  is continuous around  $\bar{w}$  with  $\bar{x} := f(\bar{w})$ , and  $\Theta \subset X$  is locally closed around  $\bar{x}$ . Let further  $D_s(\varphi, f)(\bar{w})$  be the strict derivate of the mapping  $(\varphi, f): W \rightarrow (\overline{\mathbb{R}}, X)$  at  $\bar{w}$  and  $\mathcal{D}d_\Theta$  be a topological outer regular subdifferential of  $d_\Theta$  around  $\bar{x}$ . Assume further that*

$$0 \notin \mathcal{D}'d_\Theta(\bar{x}) \tag{8}$$

*via the topological outer limit of  $\mathcal{D}d_\Theta$  relative to  $\Theta'$  defined in (5). Then there are elements  $(\lambda, x^*) \in [0, 1] \times X^*$  such that*

$$(\lambda, x^*) \neq (0, 0), \quad x^* \in \sqrt{1 - \lambda^2} \mathcal{D}'d_\Theta(\bar{x}), \quad \text{and} \tag{9}$$

$$\lambda \vartheta + \langle x^*, v \rangle \geq 0 \quad \text{for all } (\vartheta, v) \in D_s(\varphi, f)(\bar{w}). \tag{10}$$

*If in addition the dual unit ball  $\mathbb{B}^* \subset X^*$  is weak\* sequentially compact in  $X^*$ , then the topological outer regular subdifferential  $\mathcal{D}d_\Theta$  and its topological outer limit  $\mathcal{D}'d_\Theta$  can be replaced by their sequential counterparts in the relations above.*

**Proof.** The proof of the theorem is rather long but not difficult to follow. We split it into *seven steps*. Observe first that the *interior case* of  $\bar{x} = f(\bar{w}) \in \text{int } \Theta$  is trivial, since  $\mathcal{D}'d_\Theta(\bar{x}) = \emptyset$  in this case by construction (5) and therefore the theorem holds with  $x^* = 0$  and  $\lambda = 1$  by our convention at the end of Section 1 that  $\alpha\emptyset \neq \emptyset$  if and only if  $\alpha = 0$ . Thus we consider the *boundary case*  $\bar{x} \in \text{bd } \Theta$  in what follows. In Steps 1–6, which are devoted to the proof of the “topological” optimality conditions via the topological outer regular subdifferential in (8)–(10), the space  $X$  is assumed to be *arbitrary Banach*.

*Step 1: approximation by unconstrained minimization problems.* The first step of the proof is to construct a sequence of *unconstrained minimization problems* approximating the given minimizer  $\bar{w}$  for the original problem (1) with operator constraints. We proceed by using the *Ekeland variational principle*; see, e.g., [20, Theorem 2,26].

Assume without loss of generality that  $\varphi(\bar{w}) = 0$ , take an arbitrary sequence  $\varepsilon_k \downarrow 0$  as  $k \rightarrow \infty$ , and build the *penalized function*  $\varphi_k: W \rightarrow \overline{\mathbb{R}}$  by

$$\varphi_k(w) := \sqrt{[(\varphi(w) + \varepsilon_k)^+]^2 + d_\Theta(f(w))^2}, \tag{11}$$

where  $\phi^+(w) := \max\{\phi(w), 0\}$  as usual. It is easy to see that for each  $k \in \mathbb{N}$  the function  $\varphi_k$  is lower semicontinuous (l.s.c.) and bounded from below. Applying the Ekeland



variational principle to (11) for each  $k \in \mathbb{N}$ , find  $w_k \in W$  satisfying the relations

$$\varphi_k(w_k) + \sqrt{\varepsilon_k} \rho(w_k, \bar{w}) \leq \varphi_k(\bar{w}) = \varepsilon_k \quad \text{and} \quad (12)$$

$$\varphi_k(w_k) < \varphi_k(w) + \sqrt{\varepsilon_k} \rho(w, w_k) \quad \text{for all } w \in W \setminus \{w_k\}. \quad (13)$$

It follows from (12) that  $\rho(w_k, \bar{w}) \leq \sqrt{\varepsilon_k} \downarrow 0$ , while (13) shows that  $w_k$  is a *global minimizer* for the function  $\varphi_k(w) + \sqrt{\varepsilon_k} \rho(w, w_k)$  and an *approximate minimizer* for the functions  $\varphi_k(w)$  from (11). Since the constraint function  $f: W \rightarrow X$  in (1) is assumed to be continuous around  $\bar{w}$  with  $f(\bar{w}) = \bar{x}$ , for any fixed  $r > 0$  we have

$$\|f(w_k) - \bar{x}\| < r \quad \text{whenever } k \in \mathbb{N} \text{ is sufficiently large.} \quad (14)$$

*Step 2: approximation of strict subderivates.* Intending further to justify the necessary condition (10) of the theorem, take an arbitrary strict subderivate  $(\vartheta, v) \in D_s(\varphi, f)(\bar{w})$  and, by Definition 2.1(ii) along the sequence  $w_k \rightarrow \bar{w}$  built in Step 1, find a numerical sequence  $\gamma_k \downarrow 0$  as  $k \rightarrow \infty$  such that

$$(\vartheta, v) \in D_{\gamma_k}(\varphi, f)(w_k) \quad \text{for all } k \in \mathbb{N} \quad (15)$$

via the approximate subderivates from Definition 2.1(i). Taking into account that  $(\vartheta, v)$  is a  $\gamma_k$ -subderivate (15) of the pair  $(\varphi, f)$  at  $w_k$  and using (4), for each  $k \in \mathbb{N}$  we get a sequence  $(w_k^i, t_k^i)_{i \in \mathbb{N}} \in \mathcal{S}_{w_k}$ , such that

$$\begin{cases} \limsup_{i \rightarrow \infty} |E_\varphi(i, k)| := \limsup_{i \rightarrow \infty} \left| \frac{\varphi(w_k^i) - \varphi(w_k)}{t_k^i} - \vartheta \right| \leq \gamma_k, \\ \limsup_{i \rightarrow \infty} \|E_f(i, k)\| := \limsup_{i \rightarrow \infty} \left\| \frac{f(w_k^i) - f(w_k)}{t_k^i} - v \right\| \leq \gamma_k, \end{cases} \quad (16)$$

where  $E_\varphi(i, k)$  and  $E_f(i, k)$  inside of  $|\cdot|$  and  $\|\cdot\|$  in (16) are the corresponding *relative errors* in approximating the subderivate  $(\vartheta, v)$  of  $\varphi$  and  $f$ . It follows from construction (11) of the penalized functions  $\varphi_k$  that the difference  $\varphi_k(w_k^i) - \varphi_k(w_k)$  can be written as

$$\begin{aligned} & \varphi_k(w_k^i) - \varphi_k(w_k) \\ &= \lambda_k^i \left\{ [\varphi(w_k^i) + \varepsilon_k]^+ - [\varphi(w_k) + \varepsilon_k]^+ \right\} + \alpha_k^i \left\{ d_\Theta(f(w_k^i)) - d_\Theta(f(w_k)) \right\}, \end{aligned} \quad (17)$$

where the coefficients  $\lambda_k^i$  and  $\alpha_k^i$  are defined by

$$\begin{cases} \lambda_k^i := \frac{[\varphi(w_k^i) + \varepsilon_k]^+ + [\varphi(w_k) + \varepsilon_k]^+}{\varphi_k(w_k^i) + \varphi_k(w_k)} \in [0, 1], \\ \alpha_k^i := \frac{d_\Theta(f(w_k^i)) + d_\Theta(f(w_k))}{\varphi_k(w_k^i) + \varphi_k(w_k)} \in [0, 1]. \end{cases} \quad (18)$$

Fixed a natural number  $k \in \mathbb{N}$ , we consider the following *three cases*, which completely cover the situation. For simplicity and with no loss of generality, assume that each of the listed cases hold for all  $k \in \mathbb{N}$ .

(A) The *typical case*: we have

$$\varphi(w_k) + \varepsilon_k > 0, \quad d_{\Theta}(f(w_k)) > 0, \quad k \in \mathbb{N}. \tag{19}$$

(B) The *mixed sign case*: there is a subsequence of  $\{\varepsilon_k\}$ , still denoted by  $\{\varepsilon_k\}$ , such that

$$\varphi(w_k) + \varepsilon_k \leq 0, \quad d_{\Theta}(f(w_k)) > 0, \quad k \in \mathbb{N}. \tag{20}$$

(C) The *zero case*: there is a subsequence of  $\{\varepsilon_k\}$ , still denoted by  $\{\varepsilon_k\}$ , such that

$$d_{\Theta}(f(w_k)) = 0, \quad k \in \mathbb{N}. \tag{21}$$

Next we analyze each case above separately paying the main attention to the typical case (A) and indicating the necessary changes needed in the other case (B) and (C).

*Step 3: relating the subderivates of  $(\varphi, f)$  with the topological outer regular subdifferential  $\mathcal{D}d_{\Theta}(f(w_k))$  in the typical case (A).* Fix an *outer neighborhood*  $U$  of  $\bar{x}$  relative to  $\Theta'$  on which the imposed properties of outer robustness from Definition 2.2 and the extended mean value inequality (EMVI) from Definition 2.4 are satisfied. Using further estimate (14) with  $r > 0$  small enough and employing the lower semicontinuity property of  $\varphi$  around  $w_k$  as well as the continuity property of  $f$  around this point for each fixed  $k \in \mathbb{N}$ , we have the relationships

$$\varphi(w_k^i) + \varepsilon_k > 0, \quad d_{\Theta}(f(w_k^i)) > 0, \quad f(w_k^i) \in U \tag{22}$$

whenever  $i \in \mathbb{N}$  is sufficiently large. By passing to a subsequence of  $\{i\}$  as necessary, find  $\beta_k \in [0, \infty]$  such that  $\varphi(w_k^i) + \varepsilon_k \rightarrow \beta_k$  as  $i \rightarrow \infty$ . Since  $\varphi$  is l.s.c. around  $\bar{w}$ , we get

$$\beta_k \geq \varphi(w_k) + \varepsilon_k > 0 \quad \text{for all } k \in \mathbb{N}.$$

Denote further  $\eta_k := \varphi(w_k) + \varepsilon_k$  and  $\nu_k := d_{\Theta}(f(w_k))$  and observe that the limit

$$(\lambda_k, \alpha_k) := \lim_{i \rightarrow \infty} (\lambda_k^i, \alpha_k^i)$$

of the sequences in (18) exists (along a selected subsequence of  $\{i\}$ ) and is computed by

$$(\lambda_k, \alpha_k) = \frac{1}{\sqrt{\beta_k^2 + \nu_k^2} + \sqrt{\eta_k^2 + \nu_k^2}} (\beta_k + \eta_k, 2\nu_k)$$

due to the strict inequalities in (19), with  $(\lambda_k, \alpha_k) = (1, 0)$  if  $\beta_k = \infty$ . In the general case under consideration we have

$$(\lambda_k, \alpha_k) \in (0, 1) \times (0, 1) \quad \text{and} \quad \lambda_k^2 + \alpha_k^2 \geq 1/2, \tag{23}$$

where the last inequality is verified by

$$\lambda_k^2 + \alpha_k^2 = \frac{(\beta_k + \eta_k)^2 + 4\nu_k^2}{\left(\sqrt{\beta_k^2 + \nu_k^2} + \sqrt{\eta_k^2 + \nu_k^2}\right)^2} \geq \frac{\beta_k^2 + \eta_k^2 + 4\nu_k^2}{2\beta_k^2 + 2\eta_k^2 + 4\nu_k^2} \geq \frac{1}{2}.$$

It follows furthermore that

$$[\varphi(w_k^i) + \varepsilon]^+ - [\varphi(w_k) + \varepsilon]^+ = \varphi(w_k^i) - \varphi(w_k) \quad \text{for all large } i \in \mathbb{N}. \quad (24)$$

Let us handle the  $d_\Theta$  term in (17) by using the *EMVI property* (P4) of the outer subdifferential  $\mathcal{D}d_\Theta$  on the dense subset  $S$  of the outer neighborhood  $U$ . It follows from the density of  $S$  in  $U$  that there are elements  $a_k^i, b_k^i \in S$  satisfying

$$\|a_k^i - f(w_k)\| + \|b_k^i - f(w_k)\| \leq (t_k^i)^2 \quad \text{for all } i \in \mathbb{N}, \quad (25)$$

where the numbers  $t_k^i$  are taken from (16). Since  $d_\Theta$  is Lipschitz continuous with modulus  $\ell = 1$ , the last inequality implies that

$$d_\Theta(f(w_k^i)) - d_\Theta(f(w_k)) \leq d_\Theta(a_k^i) - d_\Theta(b_k^i) + (t_k^i)^2. \quad (26)$$

Employing now the extended mean value inequality (7) on the dense set  $S$ , we find elements  $c_k^i \in S \cap B(a_k^i; \|a_k^i - b_k^i\|)$  and  $u_{ik}^* \in \mathcal{D}d_\Theta(c_k^i)$  such that

$$d_\Theta(a_k^i) - d_\Theta(b_k^i) \leq \langle u_{ik}^*, a_k^i - b_k^i \rangle + \|a_k^i - b_k^i\| \omega(a_k^i, \|a_k^i - b_k^i\|). \quad (27)$$

Combine (26) and (27) to get the inequality

$$d_\Theta(f(w_k^i)) - d_\Theta(f(w_k)) \leq \langle u_{ik}^*, a_k^i - b_k^i \rangle + \|a_k^i - b_k^i\| \omega(a_k^i, \|a_k^i - b_k^i\|) + (t_k^i)^2. \quad (28)$$

Substituting expressions (24) and (28) into (17) and dividing the latter by  $t_k^i$ , we arrive at the upper estimate of the finite difference

$$\begin{aligned} & \frac{\varphi_k(w_k^i) - \varphi_k(w_k)}{t_k^i} \\ & \leq \lambda_k^i \left[ \frac{\varphi(w_k^i) - \varphi(w_k)}{t_k^i} \right] + \alpha_k^i \left\{ \left\langle u_{ik}^*, \frac{a_k^i - b_k^i}{t_k^i} \right\rangle + \left\| \frac{a_k^i - b_k^i}{t_k^i} \right\| \omega(a_k^i, \|a_k^i - b_k^i\|) + t_k^i \right\} \end{aligned} \quad (29)$$

held for all indices  $i \in \mathbb{N}$  that are sufficiently large. Let further  $\Delta f_k^i := f(w_k^i) - f(w_k)$  and observe by (16) that  $\Delta f_k^i = t_k^i [v + E_f(i, k)]$ . It follows from (16) and (25) that

$$\left\{ \begin{aligned} \limsup_{i \rightarrow \infty} \left\| \frac{a_k^i - b_k^i}{t_k^i} - v \right\| & \leq \limsup_{i \rightarrow \infty} \frac{1}{t_k^i} [\|a_k^i - b_k^i - \Delta f_k^i\| + \|\Delta f_k^i - t_k^i v\|] \\ & \leq \limsup_{i \rightarrow \infty} [t_k^i + \|E_f(i, k)\|] \leq \gamma_k \quad \text{and} \\ \limsup_{i \rightarrow \infty} \frac{\varphi(w_k^i) - \varphi(w_k)}{t_k^i} & = \limsup_{i \rightarrow \infty} [\vartheta + E_\varphi(i, k)] \leq \vartheta + \gamma_k, \end{aligned} \right.$$

which imply, in particular, that

$$\limsup_{i \rightarrow \infty} \left\| \frac{a_k^i - b_k^i}{t_k^i} \right\| \leq \|v\| + \gamma_k \quad \text{and} \quad \limsup_{i \rightarrow \infty} \|a_k^i - b_k^i\| = 0. \quad (30)$$

Now we intend to pass to the limit in the finite difference estimate (29) as  $i \rightarrow \infty$  for each fixed  $k \in \mathbb{N}$ . To proceed, we need to take care of an appropriate convergence of the

dual elements  $u_{i_k}^* \in X^*$ . Since the sequence of subgradients  $(u_{i_k}^*)_{i \in \mathbb{N}}$  in (29) is *uniformly bounded* for any  $k \in \mathbb{N}$  by the outer subdifferential property (P1) from Definition 2.5, the classical *Alaoglu-Bourbaki theorem* allows us to conclude that the sequence  $(u_{i_k}^*)_{i \in \mathbb{N}}$  contains a *subnet*  $\{u_{\nu_k}^*\}$  converging to some element  $u_k^*$  in the weak\* topology of  $X^*$ . Passing to the limit in (29) along this subnet (while keeping the notation  $\limsup_{i \rightarrow \infty}$  for the limit) and using (30) as well as the convergence  $\omega(x, \tau) \downarrow 0$  as  $x \rightarrow \bar{x}$  and  $\tau \downarrow 0$ , we get the estimate

$$\limsup_{i \rightarrow \infty} \frac{\varphi_k(w_k^i) - \varphi_k(w_k)}{t_k^i} \leq \lambda_k \vartheta + \alpha_k \langle u_k^*, v \rangle + \sigma_k, \tag{31}$$

where the remainder  $\sigma_k$  is given by

$$\sigma_k = \lambda_k \gamma_k + \alpha_k \{ \gamma_k + (\|v\| + \gamma_k) \omega(f(w_k), 0) \}, \quad k \in \mathbb{N}. \tag{32}$$

Further, it follows from (25) that

$$a_k^i \rightarrow f(w_k) \quad \text{and} \quad \|a_k^i - b_k^i\| \leq \|\Delta f_k^i\| + (t_k^i)^2 \rightarrow 0 \quad \text{as } i \rightarrow \infty.$$

Therefore we have the convergence  $c_k^i \rightarrow f(w_k)$  as  $i \rightarrow \infty$  for the intermediate points  $c_k^i \in S \cap B(a_k^i, \|a_k^i - b_k^i\|)$  defined above via the mean value property  $u_{i_k}^* \in \mathcal{D}d_\Theta(c_k^i)$ . Then the *topological outer robustness* property (P2) of the subdifferential  $\mathcal{D}d_\Theta$  gives

$$u_k^* \in \mathcal{D}'d_\Theta(f(w_k)) \subset \mathcal{D}d_\Theta(f(w_k)), \quad k \in \mathbb{N},$$

for the weak\* limit  $u_k^*$  of  $(u_{i_k}^*)_{i \in \mathbb{N}}$  whenever  $k \in \mathbb{N}$ .

Observe that the left-hand side of (31) is bounded below by  $-\sqrt{\varepsilon_k}$ . This follows from relation (13) with  $w = w_k^i$  in the variational principle and from the estimate  $\rho(w_k^i, w_k) \leq t_k^i$  in the derivate definition. Thus (31) implies that

$$-(\sigma_k + \sqrt{\varepsilon_k}) \leq \lambda_k \vartheta + \alpha_k \langle u_k^*, v \rangle, \quad k \in \mathbb{N}. \tag{33}$$

*Step 4: completing the proof of the topological optimality conditions in the typical case (A).* As justified above in “typical” case (A), inequality (33) holds with some  $u_k^* \in \mathcal{D}d_\Theta(f(w_k))$  for all  $k \in \mathbb{N}$ . Observe that  $f(w_k) \rightarrow f(\bar{w})$  as  $k \rightarrow \infty$  for the sequence of approximate minimizers  $w_k$  from (12) and (13) and that  $f(w_k) \notin \Theta$  for all  $k \in \mathbb{N}$  in this case due to (19). Note also that  $\{w_k\}$  is *independent* of the particular strict subderivate  $(\vartheta, v) \in D_s(\varphi, f)(\bar{w})$  and the selected outer subgradients  $u_k^*$  of  $d_\Theta(f(w_k))$  under consideration. Employing the *topological outer regularity* property (P3) of the subdifferential  $\mathcal{D}d_\Theta$  along the sequence  $\{f(w_k)\}$ , we find by Definition 2.3 a subnet  $f^{-1}(S)$  of  $\{w_k\}$  generated by the chosen subnet  $S$  of  $\{f(w_k)\}$  from the construction in (6) and a dual element  $u^* \in X^*$  *independent* of  $(\vartheta, v)$  such that

$$\text{Lim sup}_{w \xrightarrow{f^{-1}(S)} \bar{w}} \mathcal{D}d_\Theta(f(w)) = \{u^*\} \tag{34}$$

via the topological Painlevé-Kuratowski outer limit (2). It follows from the *topological outer robustness* property (P2) of  $\mathcal{D}d_\Theta$  and the continuity of  $f$  that  $u^* \in \mathcal{D}d_\Theta(f(\bar{w}))$ . Since the sequence of  $u_k^* \in \mathcal{D}d_\Theta(f(w_k))$ ,  $k \in \mathbb{N}$ , is *uniformly bounded* by (P1), it contains

– by the Alaoglu-Bourbaki theorem – a weak\* convergent *subnet* in  $X^*$ . By (34) and definition (2) of the topological Painlevé-Kuratowski outer limit, each subnet of this type generated by *any* strict subderivate  $(\vartheta, v) \in D_s(\varphi, f)(\bar{w})$  weak\* converges to  $u^*$ .

Since  $(\lambda_k, \alpha_k) \in [0, 1]^2$  in (23), assume with no loss of generality that the whole sequence of  $(\lambda_k, \alpha_k)$  converges to some  $(\lambda, \alpha) \in [0, 1]^2$  as  $k \rightarrow \infty$ . Thus, from the second relationship in (23), we have

$$\lambda^2 + \alpha^2 \geq 1/2. \tag{35}$$

Passing now to the limit in (33) as  $k \rightarrow \infty$  along a weak\* convergent *subnet* of  $\{u_k^*\}$  from the discussions above and taking into account that  $\sigma_k \downarrow 0$  as  $k \rightarrow \infty$  by definition (32), we arrive at the inequality

$$\lambda\vartheta + \alpha\langle u^*, v \rangle \geq 0 \text{ for all } (\vartheta, v) \in D_s(\varphi, f)(\bar{w}). \tag{36}$$

It follows from the construction of  $u^*$  in (34) in the case (A) under consideration that  $u^* \in \mathcal{D}'d_\Theta(\bar{x})$  for the outer limit  $\mathcal{D}'d_\Theta$  defined in (5). Thus  $u^* \neq 0$  due to assumption (8) of the theorem. This implies that  $(\lambda, \alpha u^*) \neq (0, 0)$  by (35). Dividing both sides of the inequality in (36) by  $\sqrt{\lambda^2 + \alpha^2}$  allows us to get  $\lambda^2 + \alpha^2 = 1$  with no loss of generality, i.e.,  $\alpha = \sqrt{1 - \lambda^2}$ . Denoting finally

$$x^* := \alpha u^* = \sqrt{1 - \lambda^2} u^*,$$

we arrive at conditions (9) and (10) and thus complete the proof of the “topological” part of the theorem in the typical case (A).

*Step 5: completing the proof of the topological optimality conditions in the mixed case (B).* In this case we have

$$\varphi_k(w_k) = d_\Theta(f(w_k)), \quad k \in \mathbb{N},$$

for the penalized function (11) by (20). Furthermore, formula (23) continues to hold in case (20) with  $(\lambda_k, \alpha_k) = (0, 1)$  for all  $k \in \mathbb{N}$ . Since the function  $x^+ := \max\{x, 0\}$  is obviously Lipschitz continuous, we get the estimate and convergence

$$\begin{aligned} & \frac{\lambda_k^i}{t_k^i} \left| [\varphi(w_k^i) + \varepsilon_k]^+ - [\varphi(w_k) + \varepsilon_k]^+ \right| \\ & \leq \frac{\lambda_k^i}{t_k^i} \left| \varphi(w_k^i) - \varphi(w_k) \right| \leq \lambda_k^i [ |E_\varphi(i, k)| + |\vartheta| ] \rightarrow 0 \text{ as } k \rightarrow \infty, \quad k \in \mathbb{N}, \end{aligned}$$

with  $\lambda_k^i$  and  $E_\varphi(i, k)$  defined in (18) and (16), respectively. Taking into account that  $f(w_k) \notin \Theta$  for all  $k \in \mathbb{N}$  in case (B), we repeat the arguments of case (A) to arrive at all the “topological” conclusions of the theorem with  $(\lambda, \alpha) = (0, 1)$  in the mixed sign case (B).

*Step 6: completing the proof of the topological optimality conditions in the zero case (C).* Considering the case (C), we observe that  $f(w_k) \in \Theta$  for all  $k \in \mathbb{N}$  sufficiently large in (21), since the set  $\Theta$  is assumed to be locally closed around  $\bar{x} = f(\bar{w})$  and since  $f(w_k) \rightarrow \bar{x}$  as  $k \rightarrow \infty$ . Without loss of generality, conclude that  $w_k$  is a feasible solution

to (1) for all  $k \in \mathbb{N}$ , and hence  $\varphi_k(w_k) \geq \varphi(\bar{w})$  as  $k \in \mathbb{N}$  due the local optimality of  $\bar{w}$  in the original constrained problem. Thus

$$\varphi_k(w_k) = \varphi(w_k) + \varepsilon_k \geq \varepsilon_k, \quad k \in \mathbb{N},$$

for the perturbed function (11) in this case, and we have counterparts of relations (23) and (33) with  $(\lambda_k, \alpha_k) = (1, 0)$  for all  $k \in \mathbb{N}$ . Repeating further the arguments of case (A) with *no actual use* of the subdifferential properties of  $d_\Theta$ , we arrive at the necessary optimality conditions (9) and (10) with  $(\lambda, x^*) = (1, 0)$ .

*Step 7: proof of the necessary optimality conditions for the sequential outer regular subdifferential.* It remains to show that the necessary optimality conditions of the theorem hold with the replacement of the topological outer regular subdifferential and its outer limit in (8)–(10) by their *sequential* counterparts from Definition 2.5 and Definition 2.2 *provided that* dual unit ball  $B^* \subset X^*$  is *sequentially weak\* compact*. This follows directly from the arguments above, where the latter assumption and property (P1) allow us to use *subsequences instead of subnets* in the corresponding limiting procedures. Thus we complete the proof of the theorem.  $\square$

It is not hard to show that the necessary optimality conditions obtained in Theorem 3.1 imply the classical *Lagrange multiplier rule* in the case of problems with finitely many equality and inequality constraints given by *strictly differentiable* functions on *Banach spaces*  $W$ . They are also consistent with some extended versions of multiplier rules for problems with *nonsmooth data* on Banach spaces obtained in terms of the aforementioned specific subdifferentials; cf. [5, 6, 21, 26, 27] and the references therein.

In the next section we consider a remarkable class of generally nonconvex constraint sets  $\Theta$  in Banach spaces for which the necessary optimality conditions of Theorem 3.1 can be constructively expressed via the *major subdifferential constructions* of variational analysis that agree with each other and satisfy all the requirements imposed in Theorem 3.1.

#### 4. Approximately Convex Functions and Sets

The main notion studied in this section is *approximate convexity* for extended-real-valued functions on Banach spaces introduced by Ngai, Luc and Théra in [23] and its realization for the case of sets via the distance functions, which is needed in what follows. The concept of approximate convexity has been proved to be very useful for many aspects of variational analysis and optimization being closely related to (while generally different from) other important notions of generalized convexity for functions and sets. We refer the reader to [2, 21, 23, 24, 26, 30] and the bibliographies therein for various properties of approximately convex functions and sets, their relations with other notions of generalized convexity, and a number of applications to variational analysis and generalized differentiation.

In this section we recall some facts on approximate convexity and derive several properties of approximately convex functions and sets needed for the implementation in Section 5 of our general necessary optimality conditions from Theorem 3.1 in the case of approximately convex constraint sets. Together with the approximate convexity of functions and sets *at* the reference point as in [23], we define and study in this section and then apply in Section 5 a version of approximate convexity *around* the reference point involving all

the points in a *neighborhood* of the reference one. Note that the latter modification is generally different from the original one in [23] as well as from the *uniform* approximate convexity introduced recently in [24]. Let us start with the basic definitions.

**Definition 4.1 (Approximately convex functions and sets).** Let  $\psi: X \rightarrow \overline{\mathbb{R}}$  be a proper extended-real-valued function on a Banach space  $X$ , and let  $\Theta \subset X$  be a nonempty subset of  $X$ . Then:

- (i) The function  $\psi$  is *approximately convex at*  $\bar{x} \in \text{dom } \psi$  if for each number  $\gamma > 0$  there is  $\eta > 0$  such that for all  $x, y \in B(\bar{x}; \eta)$  and  $t \in (0, 1)$  we have

$$\psi((1 - t)x + ty) \leq (1 - t)\psi(x) + t\psi(y) + \gamma t(1 - t)\|x - y\|. \tag{37}$$

- (ii) The function  $\psi$  is *approximately convex around*  $\bar{x} \in \text{dom } \psi$  if there is a neighborhood of  $\bar{x}$  such that  $\psi$  is approximately convex at every point of this neighborhood.
- (iii) The set  $\Theta$  is *approximately convex at* (respectively, *around*)  $\bar{x}$  if the distance function  $d_\Theta: X \rightarrow \mathbb{R}$  is approximately convex at (respectively, around) this point.

If  $X$  is a uniformly convex space with a uniformly smooth and uniformly convex norm, then the approximate convexity of the set  $\Theta$  around  $\bar{x}$  implies its *prox-regularity* around this point. The latter fact follows from [3, Theorem 4.9] and extended the previous result of [25] proved in Hilbert spaces. Prox-regular functions and sets and their remarkable properties play a significant role in many aspects of variational analysis and optimization; see, e.g., [3, 25, 26] and the references therein.

Observe further that the approximate convexity around  $\bar{x}$  from Definition 4.1(ii) is generally a *weaker* assumption in comparison with the “uniform approximate convexity” around the reference point defined in [24], where (37) is required to hold for all points  $(x, y)$  close to each other *uniformly* in a fixed neighborhood of  $\bar{x}$ . In finite dimensions, the approximate convexity around  $\bar{x}$  from Definition 4.1(ii) is *equivalent* to the uniform convexity due to the compactness of the unit ball; it is easy to show this by standard compactness arguments. Note also that the approximate convexity *at* the point in question does *not* imply the one *around* this point even for *strictly differentiable* functions on the real line as in the following case taken from [20, p. 19].

**Example 4.2 (Difference between approximate convexity at and around the point).** Consider the function  $\psi: \mathbb{R} \rightarrow \mathbb{R}$  given by

$$\psi(x) := \begin{cases} -x^2 & \text{if } x = 1/k, \quad k \in \mathbb{N}, \\ 0 & \text{if } x = 0, \\ \text{linear} & \text{otherwise.} \end{cases} \tag{38}$$

It is easy to check that this function is *strictly differentiable* at  $\bar{x} = 0$  (although it is *not* Fréchet differentiable at points nearby) and that *strict differentiability always implies approximate convexity at* the point in question. However, this function is *not approximately convex around*  $\bar{x}$ . Indeed, we get directly from the above construction (38) that the function  $\psi$  admits the following representation on  $(0, 1)$ :

$$\psi(x) = \begin{cases} -\frac{1}{k^2} + m_1\left(x - \frac{1}{k}\right) & \text{if } \frac{1}{k+1} < x < \frac{1}{k}, \\ -\frac{1}{k^2} + m_2\left(x - \frac{1}{k}\right) & \text{if } \frac{1}{k} < x < \frac{1}{k-1}, \end{cases} \quad k \in \mathbb{N},$$

where  $m_2 < m_1 < 0$  are the corresponding slopes to the graph of  $\psi$ . Pick  $z_k \in (0, \frac{1}{k(k+1)})$  and let  $x_k := \frac{1}{k} - z_k$  and  $y_k := \frac{1}{k} + z_k$ . Then

$$\psi(x_k) = -\frac{1}{k^2} - m_1 z_k \quad \text{and} \quad \psi(y_k) = -\frac{1}{k^2} + m_2 z_k, \quad k \in \mathbb{N},$$

which implies the following equalities for all  $k \in \mathbb{N}$ :

$$\psi(k^{-1}) - \frac{\psi(x_k) + \psi(y_k)}{2} = \frac{(m_1 - m_2)z_k}{2} = \frac{|y_k - x_k|}{4} = \frac{z_k}{2}.$$

The latter shows that inequality (37) *cannot* be satisfied for  $x = x_k, y = y_k$ , and  $t = 1/2$  if  $\gamma > 0$  is chosen to be sufficiently small (say  $\gamma < 1/2$ ) however small  $\eta$  is. Thus function (38) is *not* approximately convex at  $\bar{x}_k = 1/k$  for any large  $k \in \mathbb{N}$ .

An important fact established in [23, Theorem 3.6] shows that for every l.s.c. function  $\psi: X \rightarrow \overline{\mathbb{R}}$  on an arbitrary Banach space  $X$  the *major subdifferentials* of variational analysis (Clarke-Rockafellar, Fréchet, Ioffe, Mordukhovich) *coincide* at a point  $\bar{x} \in \text{dom } \psi$  where  $\psi$  is approximately convex and they agree with the *convex-type subdifferential*

$$\partial\psi(\bar{x}) := \{x^* \in X^* \mid \langle x^*, v \rangle \leq \psi'(\bar{x}; v) \text{ for all } v \in X\} \tag{39}$$

defined via the classical *directional derivative*

$$\psi'(\bar{x}; v) := \lim_{t \downarrow 0} \frac{\psi(\bar{x} + tv) - \psi(\bar{x})}{t} \tag{40}$$

of  $\psi$  at  $\bar{x}$  in the direction  $v$ , which exists and is sublinear on  $X$ . If  $\psi$  is convex, the subdifferential (39) reduces to the classical subdifferential of convex analysis. Thus we keep the notation  $\partial\psi(\bar{x})$  for the *subdifferential of the approximately convex function*  $\psi$  at  $\bar{x}$  that encompasses all the aforementioned subdifferentials.

The next proposition contains some useful properties of approximately convex functions  $\psi: X \rightarrow \overline{\mathbb{R}}$  around the reference point employed, in particular, in the proof of necessary optimality conditions of Section 5. Observe that we assume the “around” approximate convexity of  $\psi$  to make sure that  $\partial\psi(\cdot)$  in (39) is the subdifferential of the function  $\psi$  not only at  $\bar{x}$  but also at all the points  $x \in \text{dom } \psi$  sufficiently close to  $\bar{x}$ . In fact, certain modifications of the proofs below allow us to justify the necessary optimality conditions obtained in Section 5 in the more general case when the constraint set  $\Theta$  in (1) is approximately convex only *at* the optimal point; see Remark 5.7.

**Proposition 4.3 (Properties of approximately convex functions).** *Let  $\psi: X \rightarrow \overline{\mathbb{R}}$  be approximately convex around  $\bar{x}$  on a Banach space  $X$ . Then there is an upper semicontinuous function  $\theta: (0, \infty) \rightarrow [0, \infty)$  such that  $\theta(\tau) \downarrow 0$  as  $\tau \downarrow 0$  and the following hold:*

(i) *For all  $x, y \in X$  sufficiently close to  $\bar{x}$  and all  $t \in (0, 1)$  we have*

$$\begin{cases} \psi(x_t) \leq (1-t)\psi(x) + t\psi(y) + \theta(r_{[x,y]}(\bar{x}))t(1-t)\|x-y\|, \\ \frac{\psi(x_t) - \psi(x)}{\|x_t - x\|} \leq \frac{\psi(y) - \psi(x)}{\|y - x\|} + \theta(r_{[x,y]}(\bar{x}))(1-t), \end{cases} \tag{41}$$

where  $r_{[x,y]}(\bar{x}) := \max\{\|x - \bar{x}\|, \|y - \bar{x}\|\}$  and where  $x_t := x + t(y - x)$ .



(ii) Let  $x^* \in \partial\psi(x)$ , where  $x \in X$  is sufficiently close to  $\bar{x}$ . Then for all  $y \in X$  close to  $x$  we have the estimate

$$\langle x^*, y - x \rangle \leq \psi(y) - \psi(x) + \theta(r_{[x,y]}(\bar{x}))\|y - x\|. \tag{42}$$

(iii) If (42) holds for some  $x \in X$  close to  $\bar{x}$  and all  $y \in X$  close to  $x$ , then

$$\langle x^*, v \rangle \leq \psi'(x; v) + \theta(\|x - \bar{x}\|)\|v\| \quad \text{whenever } v \in X. \tag{43}$$

**Proof.** Define the function  $\theta: (0, \infty) \rightarrow [0, \infty)$  by

$$\theta(\tau) := \limsup_{\eta \rightarrow \tau} \omega(\eta), \quad \tau \in (0, \infty), \tag{44}$$

where  $\omega(\eta) := \inf\{\gamma > 0 \mid (37) \text{ holds for all } (x, y) \in B(\bar{x}; \eta)\}$ . It is easy to check that function (44) satisfies all the requirements asserted in the theorem. Let us justify the three properties (i)–(iii) with this function  $\theta$ .

To proceed with (i), observe that the first inequality in (41) follows directly from (37) by construction of the function  $\theta$  in (44). Subtracting  $\psi(x)$  from both sides of the first inequality in (41) and dividing then each term by  $\|x_t - x\| = t\|y - x\|$ , we arrive at the second inequality in (41) and thus justify property (i) of the proposition.

To prove (ii), fix  $x \in X$  sufficiently close to  $\bar{x}$  and take any  $y \in X$  close to  $x$ . Then the second inequality in (41) implies that

$$\frac{\psi(x + tv) - \psi(x)}{t} \leq \psi(x + v) - \psi(x) + \theta(r_{[x,y]}(\bar{x}))(1 - t)\|v\| \tag{45}$$

with  $v := y - x$  for all  $t > 0$  sufficiently small. By passing to the limit in (45) as  $t \downarrow 0$  and taking into account the existence of the directional derivative in (40), we conclude that

$$\psi'(x; v) \leq \psi(x + v) - \psi(x) + \theta(r_{[x,x+v]}(\bar{x}))\|v\| \quad \text{with } v = y - x. \tag{46}$$

Since  $\langle x^*, y - x \rangle \leq \psi'(x; y - x)$  for any  $x^* \in \partial\psi(x)$  by (39), it follows from (46) that estimate (42) is satisfied, which justifies property (ii).

Finally, let  $x^*$  satisfy (42) for some fixed  $x$  close to  $\bar{x}$  and any  $y$  close to  $x$ . Taking an arbitrary direction  $v \in X$  and setting  $y := x + tv$  for small  $t > 0$ , we get from (42) that

$$\langle x^*, v \rangle \leq \frac{\psi(x + tv) - \psi(x)}{t} + \theta(r_{[x,x+tv]}(\bar{x}))\|v\|,$$

which gives (43) by passing to the limit as  $t \downarrow 0$  by (40) due to the upper semicontinuity of  $\theta(\cdot)$ . This justifies (iii) and completes the proof of the proposition.  $\square$

### 5. Case Study for Approximately Convex Constraints

The concluding section of the paper is devoted to the implementation and specification of the general necessary optimality conditions for problem (1) defined on complete metric spaces in the case of *approximately convex* constraint sets  $\Theta$  that belong to a broad class of Banach spaces admitting uniformly Gâteaux differentiable renorms.

Recall that a norm  $\|\cdot\|$  on a Banach space  $X$  is *uniformly Gâteaux differentiable* if for every  $h \in X$  with  $\|h\| = 1$  the limit

$$\lim_{t \rightarrow 0} \frac{\|x + th\| - \|x\|}{t}$$

exists and the convergence is uniform in  $x \in X$  with  $\|x\| = 1$ . We say that a Banach space  $X$  is *uniformly Gâteaux smooth* if it admits a uniformly Gâteaux differentiable renorming, i.e., an equivalent uniformly Gâteaux differentiable norm at nonzero points. The class of Gâteaux smooth Banach spaces is sufficiently broad containing, in particular, all *weakly compactly generated* Banach spaces and thus every separable and every reflexive space. We refer the reader to [7, 9] and the bibliographies therein for a variety of results on Gâteaux smooth spaces including equivalent descriptions, sufficient conditions, examples, and more discussions. Let us mention, in particular, a nice characterization established recently in [9]:  $X$  is uniformly Gâteaux smooth if and only if the dual unit ball  $B^* \subset X^*$  is a uniform Eberlein compact in the weak\* topology of  $X^*$ .

In our proof below we need the following equivalent descriptions from [7, Proposition 6.2(ii) and Theorem 6.7] of the uniformly Gâteaux differentiable norm  $\|\cdot\|$  on  $X$  via the dual norm on  $X^*$ ; for simplicity we keep the same notation  $\|\cdot\|$  for the corresponding dual norm on  $X^*$ .

**Proposition 5.1 (Equivalent dual descriptions of uniformly Gâteaux differentiable norms).** *The norm  $\|\cdot\|$  on  $X$  is uniformly Gâteaux differentiable if and only if the dual norm on  $X^*$  is weak\*-uniformly rotund in the sense that for any sequences of dual elements  $x_k^* \in X^*$  and  $y_k^* \in X^*$  as  $k \in \mathbb{N}$  satisfying the relations*

$$\|x_k^*\| = \|y_k^*\| = 1 \quad \text{for all } k \in \mathbb{N} \quad \text{and} \quad \|x_k^* + y_k^*\| \rightarrow 2 \quad \text{as } k \rightarrow \infty \quad (47)$$

*we have the weak\* convergence  $(x_k^* - y_k^*) \xrightarrow{w^*} 0$  as  $k \rightarrow \infty$  in  $X^*$ . Furthermore, the conditions  $\|x_k^*\| = \|y_k^*\| = 1$  as  $k \in \mathbb{N}$  in (47) can be equivalently replaced by those of  $\|x_k^*\| \rightarrow 1$  and  $\|y_k^*\| \rightarrow 1$  as  $k \rightarrow \infty$  in the above characterization of uniform Gâteaux differentiable norms.*

For the main result of this section we need also the following property for the constraint set  $\Theta$  at the reference optimal solution  $\bar{x} = f(\bar{w})$  to (1), which ensures the *nontriviality* of multipliers in the corresponding necessary optimality conditions.

**Definition 5.2 (Tangential relative interior points).** We say that a subset  $\Theta$  of a Banach space  $X$  has a *tangential relative interior point* at  $\bar{x} \in \Theta$  if there exist  $x_0 \in X$ , numbers  $\eta > 0$ ,  $\gamma > 0$  and a compact set  $C \subset X$  such that

$$B(x_0; \eta) \subset [t^{-1}(\Theta - \bar{x})] \cap \mathcal{B} + C \quad \text{for all } t \in (0, \gamma). \quad (48)$$

Note that condition (48) automatically holds with  $x_0 = 0$  for every closed and convex set  $\Theta \subset X$  such that the linear subspace spanned by  $\Theta$  is *closed* and *finite-codimensional* in  $X$  and its *relative interior*,  $\text{ri } \Theta$ , is *nonempty*. Indeed, it follows from [4, Theorem 2.5] that in this case there is a convex compact set  $C \subset X$  such that  $0 \in \text{int}[(\Theta - \bar{x}) \cap \mathcal{B} + C]$ , i.e.,

$$B(0; \eta) \subset (\Theta - \bar{x}) \cap \mathcal{B} + C \quad \text{for some } \eta > 0. \quad (49)$$

Since  $\Theta$  convex and  $0 \in (\Theta - \bar{x})$ , we have  $\Theta - \bar{x} \subset t^{-1}(\Theta - \bar{x})$ , and hence (49) implies (48).

In what follows we pay the main attention to *approximately convex* sets admitting *tangential relative interior points* in *uniformly Gâteaux smooth* Banach spaces. The next theorem shows that the *subdifferential* (39) of the distance functions  $d_\Theta$  for such sets, which encompasses the major subdifferentials of variational analysis, is an *outer regular subdifferential* in the sense of Definition 2.5 satisfying furthermore the *nontriviality condition* (8) of Theorem 3.1. Note that, since the dual unit ball  $B^* \subset X^*$  is *sequentially weak\* compact* for any (uniformly) Gâteaux smooth space  $X$  by the discussion in Section 2, there is *no difference* between topological and sequential outer regular subdifferentials in the setting under consideration in the next theorem.

**Theorem 5.3 (Outer regular subdifferentials for the distance functions of approximately convex sets).** *Let  $X$  be a uniformly Gâteaux smooth Banach space, and let  $\Omega \subset X$  be a nonempty subset locally closed around  $\bar{x} \in \Omega$ . The following assertions hold:*

- (i) *If  $\Theta$  is approximately convex around  $\bar{x}$ , then the subdifferential  $\partial d_\Theta$  in (39) of the distance function  $d_\Theta$  encompassing the major subdifferentials of variational analysis is an outer regular subdifferential around  $\bar{x}$ .*
- (ii) *If in addition  $\Theta$  has a tangential relative interior point at  $\bar{x}$ , then the nontriviality condition  $0 \notin \partial' d_\Theta(\bar{x})$  holds for the subdifferential  $\partial d_\Theta$ .*

**Proof.** To justify (i), observe first that, as discussed in Section 4, the approximate convexity around  $\bar{x}$  ensures the existence of a neighborhood of  $\bar{x}$  on which the subdifferential  $\partial d_\Theta(x)$  in (39) of the distance function  $d_\Theta$  encompasses the major subdifferentials of variational analysis. Property (P1) in Definition 2.5 follows for the subdifferential (39) of  $d_\Theta$  directly from its definition. The outer robustness property (P2) and EMVI property (P4) with  $\omega \equiv 0$  in (7) hold for  $\partial d_\Theta$  due to, e.g., their validity for Clarke’s generalized gradient of Lipschitz continuous functions; see [6, Proposition 2.1.5 and Theorem 2.3.7].

To complete the proof of (i), it remains to justify the outer regularity property (P3) of  $\partial d_\Theta$  from Definition 2.5. The case of  $\bar{x} \in \text{int } \Theta$  is trivial, since in this case there is no sequence of  $x_k \in \Theta'$  converging to  $\bar{x}$ . Thus we consider the boundary case  $\bar{x} \in \text{bd } \Theta$ , fix an arbitrary sequence  $x_k \xrightarrow{\Theta'} \bar{x}$  as  $k \rightarrow \infty$ , and let  $\|\cdot\|$  be a *uniformly Gâteaux differentiable norm* on  $X$ . Take now any sequence of subgradients  $x_k^* \in \partial d_\Theta(x_k)$  from (39) and establish first the norm convergence

$$\|x_k^*\| \rightarrow 1 \text{ as } k \rightarrow \infty. \tag{50}$$

To proceed, let  $\varepsilon_k := 1/k$  for all  $k \in \mathbb{N}$  and choose  $y_k \in \Theta$  such that

$$d_\Theta(x_k) \geq (1 - \varepsilon_k)\|x_k - y_k\|, \quad k \in \mathbb{N}.$$

Apply now property (42) of the approximately convex function  $\psi(x) = d_\Theta(x)$  with  $x^* = x_k^*$  and  $(x, y) = (x_k, y_k)$  therein to get the estimate

$$\langle x_k^*, y_k - x_k \rangle \leq -(1 - \varepsilon_k)\|x_k - y_k\| + \theta(r_{[x_k, y_k]}(\bar{x}))\|y_k - x_k\|, \quad k \in \mathbb{N},$$

since  $d_\Theta(y_k) = 0$ . Dividing then each term of the above inequality by  $\|x_k - y_k\| \neq 0$  for all  $k \in \mathbb{N}$ , we conclude that

$$\left\langle x_k^*, \frac{x_k - y_k}{\|x_k - y_k\|} \right\rangle \geq 1 - \varepsilon_k - \theta(r_{[x_k, y_k]}(\bar{x})),$$

which gives  $1 \geq \|x_k^*\| \geq 1 - \varepsilon_k - \theta(r_{[x_k, y_k]}(\bar{x}))$ . Passing to the limit in the latter estimates as  $k \rightarrow \infty$  and taking into account that  $x_k, y_k \rightarrow \bar{x}$  and  $r_{[x_k, y_k]}(\bar{x}) \rightarrow 0$ , we arrive at (50).

Since  $x_k^* \in \mathbb{B}^*$  for all  $k \in \mathbb{N}$  and the dual ball  $\mathbb{B}^* \subset X^*$  is *sequentially* weak\* compact in  $X^*$  (by the Gâteaux smoothness of  $X$ ), the sequence  $\{x_k^*\}$  contains a subsequence that weak\* converges to some  $x^* \in X^*$ . Without loss of generality, assume that the sequence  $\{x_k^*\}$  itself converges to  $x^*$  as  $k \rightarrow \infty$ . To justify the outer regularity property (6), we thus need to show that any weak\* convergent sequence of  $y_k^* \in \partial d_\Theta(x_k)$ ,  $k \in \mathbb{N}$ , has the same weak\* limit  $x^*$ , i.e.,

$$x_k^* - y_k^* \xrightarrow{w^*} 0 \text{ as } k \rightarrow \infty \text{ whenever } y_k^* \in \partial d_\Theta(x_k), \quad k \in \mathbb{N}, \tag{51}$$

and the sequence  $\{y_k^*\}$  weak\* converges in  $X^*$ . Indeed, by the obvious convexity of the set  $\partial d_\Theta(x_k)$  in (39), we have the inclusion  $(x_k^* + y_k^*)/2 \in \partial d_\Theta(x_k)$  for all  $k \in \mathbb{N}$ . Therefore, the above relation (50) implies the norm convergence

$$\|y_k^*\| \rightarrow 1 \quad \text{and} \quad \|x_k^* + y_k^*\| \rightarrow 2 \quad \text{as } k \rightarrow \infty. \tag{52}$$

It easily follows from (52) and from the equivalent dual description of the uniformly Gâteaux differentiable norm  $\|\cdot\|$  from Proposition 5.1 that  $x_k^* - y_k^* \xrightarrow{w^*} 0$  as  $k \rightarrow \infty$ . This justifies (51) and thus completes the proof of the outer regularity assertion (i) of the theorem.

Next we justify assertion (ii) of the theorem ensuring the validity of the *nontriviality condition*  $0 \notin \partial' d_\Theta(\bar{x})$  for the outer limit (5) of the subdifferential (39) for the distance function  $d_\Theta$  under the *tangential relative interiority property* (48) of the approximately convex set  $\Theta$  under consideration. Take any  $x^* \in \partial' d_\Theta(\bar{x})$  and by the (sequential) construction in (5) find sequences  $x_k \xrightarrow{\Theta'} \bar{x}$  and  $x_k^* \in \partial d_\Theta(x_k)$  such that  $x_k^* \xrightarrow{w^*} x^*$  as  $k \rightarrow \infty$ . We need to show that  $x^* \neq 0$ . To proceed, employ the tangentially relative interiority property of  $\Theta$  at  $\bar{x}$  from Definition 5.2 assuming without loss of generality that  $x_0 = 0$  therein. In this way, using the function  $\theta(\cdot)$  from Proposition 4.3 and the constants from Definition 5.2, select  $t \in (0, \gamma)$  so small that  $\theta(t) \leq \eta/4$  and suppose in what follows that  $k \in \mathbb{N}$  is so large that  $\|x_k - \bar{x}\| \leq t$ . Applying inequality (42) from Proposition 4.3 to  $d_\Theta$  with  $x = x_k$  and taking into account that  $d_\Theta(x_k) \geq 0$  and  $d_\Theta(y) = 0$ , we get

$$\langle x_k^*, y - x_k \rangle \leq \theta(r_{[x_k, y]}(\bar{x})) \|y - x_k\| \quad \text{for all } y \in \Theta \cap B(\bar{x}; t). \tag{53}$$

Since  $r_{[x_k, y]}(\bar{x}) \leq \max\{\|x_k - \bar{x}\|, \|y - \bar{x}\|\} \leq t$ ,  $\|y - x_k\| \leq 2t$ , and  $\theta(t) \leq \eta/4$  in (53), this estimate yields that

$$\langle x_k^*, y - x_k \rangle \leq \eta t/2 \quad \text{for large } k \in \mathbb{N}. \tag{54}$$

Take further any point  $u \in B(0; \eta)$  and represent it by the tangential relative interiority condition (48) in Definition 5.2 as

$$u = x/t + z \text{ for some } x \in (\Theta - \bar{x}) \cap B(0; t) \text{ and } z \in C.$$

Letting  $y := x + \bar{x} = t(u - z) + \bar{x} \in \Theta \cap B(\bar{x}; t)$ , we get from (54) that

$$\langle x_k^*, t(u - z) + \bar{x} - x_k \rangle \leq \eta t/2,$$

which immediately implies the estimate

$$\langle x_k^*, u \rangle \leq \frac{\langle x_k^*, x_k - \bar{x} \rangle}{t} + \langle x_k^*, z \rangle + \frac{\eta}{2} \leq \frac{|\langle x_k^*, \bar{x} - x_k \rangle|}{t} + \max_{z \in C} |\langle x_k^*, z \rangle| + \frac{\eta}{2}.$$

Since the latter also holds with  $u$  replaced by  $-u \in B(0; \eta)$ , we arrive at

$$\eta \|x_k^*\| = \sup_{u \in B(0; \eta)} |\langle x_k^*, u \rangle| \leq \max_{z \in C} |\langle x_k^*, z \rangle| + \frac{|\langle x_k^*, \bar{x} - x_k \rangle|}{t} + \frac{\eta}{2} \tag{55}$$

for all large  $k \in \mathbb{N}$ . Let us finally show that estimate (55) ensures that  $x^* \neq 0$  for the weak\* limit of  $x_k^* \xrightarrow{w^*} x^*$  as  $k \rightarrow \infty$ .

Assuming the contrary and taking into account the compactness of  $C$  in  $X$ , we get

$$\max_{z \in C} |\langle x^*, z \rangle| \rightarrow 0 \text{ as } k \rightarrow \infty$$

from the weak\* convergence  $x_k^* \xrightarrow{w^*} 0$ . Furthermore, it follows from the norm convergence  $x_k \rightarrow \bar{x}$  and from the boundedness of  $\{x_k^*\}$  in  $X^*$  by the *uniform boundedness principle* that

$$|\langle x_k^*, x - \bar{x} \rangle| \rightarrow 0 \text{ as } k \rightarrow \infty$$

The latter two relations allow us to conclude from (55) that  $\|x_k^*\| \leq 2/3$  for all large  $k \in \mathbb{N}$  that clearly contradicts the norm convergence (50) derived above. Thus  $x^* \neq 0$ , which completes the proof of assertion (ii) and of the whole theorem.  $\square$

Now we are ready to establish the main result of this section providing verifiable necessary optimality conditions for the original problem (1) on *metric spaces* with operator constraints given by general *nonsmooth mappings* and *approximately convex* sets in uniformly *Gâteaux smooth Banach spaces*. This result is an efficient specification in the setting under consideration of the general necessary optimality conditions of Section 3 obtained via abstract outer regular subdifferentials. To formulate the new result, we recall the following well-known constructions of variational analysis; see, e.g., [20, Chapter 1].

Given a nonempty set  $\Theta \subset X$  in a Banach space  $X$  and a point  $\bar{x} \in \Theta$ , the *Fréchet normal cone* to  $\Theta$  at  $\bar{x}$  is defined by

$$N(\bar{x}; \Theta) := \left\{ x^* \in X^* \mid \limsup_{x \xrightarrow{\Theta} \bar{x}} \frac{\langle x^*, x - \bar{x} \rangle}{\|x - \bar{x}\|} \leq 0 \right\} \tag{56}$$

via the standard upper limit of scalar functions. The *weak contingent cone* to  $\Theta$  at  $\bar{x}$  is defined via the *weak convergence* “ $\xrightarrow{w}$ ” on  $X$  by

$$T_w(\bar{x}; \Theta) := \left\{ v \in X \mid \begin{array}{l} \exists \text{ sequences } x_k \xrightarrow{\Theta} \bar{x} \text{ and } \alpha_k \geq 0 \\ \text{such that } \alpha_k(x_k - \bar{x}) \xrightarrow{w} v \text{ as } k \rightarrow \infty \end{array} \right\}. \tag{57}$$

If the weak convergence in (57) is replaced by the *norm* convergence on  $X$ , construction (57) reduces to the classical *Bouligand-Severi contingent cone*  $T(\bar{x}; \Theta)$ ; see [20, Subsection 1.1.2] for more details, discussions, and references. We obviously have the inclusion

$$T(\bar{x}; \Theta) \subset T_w(\bar{x}; \Theta),$$

where the equality holds if  $X$  is finite-dimensional. Furthermore, the *polarity inclusion*

$$N(\bar{x}; \Theta) \subset \{x^* \in X^* \mid \langle x^*, v \rangle \leq 0 \text{ for all } v \in T_w(\bar{x}; \Theta)\} \tag{58}$$

is satisfied in arbitrary Banach spaces, where the equality holds in (58) if  $X$  is reflexive; see [20, Theorem 1.10]. Observe that the Fréchet normal cone (56) is always convex while neither  $T_w(\bar{x}; \Theta)$  nor  $T(\bar{x}; \Theta)$  must be convex even in finite dimensions.

**Theorem 5.4 (Necessary optimality conditions for operator-constrained problems on metric spaces with approximately convex constraint sets).** *Let  $\bar{w}$  be a local minimizer for problem (1) in the framework of Theorem 3.1. Assume in addition that  $X$  is a Gâteaux smooth Banach space, that the constraint set  $\Theta \subset X$  is approximately convex around  $\bar{x} := f(\bar{w})$ , and that  $\Theta$  admits a tangential relative interior point at  $\bar{x}$ . Then there are multipliers  $(\lambda, x^*) \in \mathbb{R} \times X^*$  such that*

$$(\lambda, x^*) \neq (0, 0), \quad \lambda \geq 0, \quad x^* \in N(\bar{x}; \Theta), \tag{59}$$

and the strict derivate relation

$$\lambda \vartheta + \langle x^*, v \rangle \geq 0 \text{ for all } (\vartheta, v) \in D_s(\varphi, f)(\bar{w}) \tag{60}$$

is satisfied. Furthermore, the normal cone inclusion  $x^* \in N(\bar{x}; \Theta)$  in (59) implies that

$$\langle x^*, v \rangle \leq 0 \text{ for all } v \in T_w(\bar{x}; \Theta) \tag{61}$$

via the weak contingent cone (57), where the equivalence between  $x^* \in N(\bar{x}; \Theta)$  and (61) holds if the Banach space  $X$  is reflexive.

**Proof.** Theorem 5.3 tells us that the subdifferential  $\partial d_\Theta$  in (39) of the approximately convex distance function  $d_\Theta$ , which encompasses the major subdifferentials of variational analysis, is an *outer regular subdifferential* of  $d_\Theta$  around  $\bar{x}$  under the assumptions made. Thus we can apply the sequential version of Theorem 3.1 (equivalent to the topological one) to the case under consideration in the uniformly Gâteaux smooth space  $X$ . By assertion (ii) of Theorem 5.3 the *nontriviality condition* (8) with  $\mathcal{D}'d_\Theta(\bar{x}) = \partial'd_\Theta(\bar{x})$  holds, and thus Theorem 3.1 ensures the existence of multipliers  $(\lambda, x^*) \in \mathbb{R} \times X^*$  such that

$$(\lambda, x^*) \neq (0, 0), \quad \lambda \geq 0, \quad x^* \in \text{cone } \partial'd_\Theta(\bar{x}), \tag{62}$$

and the *strict derivate relation* (10)=(60) is satisfied. To complete the proof of the theorem, it remains to show that the inclusion  $x^* \in \text{cone } \partial' d_\Theta(\bar{x})$  in (62) implies that  $x^* \in N(\bar{x}; \Omega)$ , which in turn yields (61).

Indeed, it follows directly from the *outer robustness* property (5) of the subdifferential (39) at  $x = \bar{x}$  that  $\partial' d_\Theta(\bar{x}) \subset \partial d_\Theta(\bar{x})$ . Since the subdifferential (39) for the approximate convex function  $\psi = d_\Theta$  reduces to the Fréchet subdifferential of  $d_\Theta$  at  $\bar{x}$ , we get from [20, Corollary 1.96] that  $x^* \in N(\bar{x}; \Theta)$  for the Fréchet normal cone defined in (56). Furthermore, inequality (61) in arbitrary Banach spaces  $X$  and its equivalence to  $x^* \in N(\bar{x}; \Theta)$  in (59) if  $X$  is reflexive follow from the *polarity inclusion* (58) and from the case of equality therein mentioned above. This completes the proof of the theorem.  $\square$

We conclude this section with several remarks discussing some specifications and extensions of the major results obtained in the paper.

**Remark 5.5 (Multiplier rule in the case of convex constraint sets).** If the constraint set  $\Theta$  in (1) is *convex*, then condition  $x^* \in N(\bar{x}; \Omega)$  in (59) reduces to

$$\langle x^*, x - \bar{x} \rangle \leq 0 \text{ for all } x \in \Theta.$$

This follows from the fact that in the convex case the normal cone (56) agrees with the classical normal cone of convex analysis. This version of Theorem 5.4 significantly extends the main result of [17] obtained in the case when the cost function  $\varphi$  is continuous and the space  $X$  has a strictly convex/norm-rotund dual (instead of the *weak\** rotundity as in Theorem 5.4) and the set  $\Theta$  is convex and finite-codimensional with closed span and nonempty relative interior. The latter assumptions imply the tangential relative interior condition (48) as discussed after Definition 5.2. Note that the proof of the nontriviality condition  $(\lambda, x^*) \neq (0, 0)$  in [17] is based on Lemma 3.6 from Chapter 4 in [16], which cannot be applied in the setting of Theorem 5.3.

**Remark 5.6 (Nontriviality condition under sequential normal compactness).** The nontriviality condition  $(\lambda, x^*) \neq (0, 0)$  in Theorem 5.4 based on assertion (ii) of Theorem 5.3 holds in fact under the replacement of the tangential relative interiority assumption (48) by generally less restrictive *sequential normal compactness* (SNC) property of  $\Theta$  at  $\bar{x} \in \Theta$  formulated via the normal cone (56) as follows:

$$[x_k^* \in N(x_k; \Theta) \text{ with } x_k \rightarrow \bar{x}, x_k^* \xrightarrow{w^*} 0] \implies \|x_k^*\| \rightarrow 0 \text{ as } k \rightarrow \infty. \tag{63}$$

This property is automatic in finite dimensions while playing a crucial role in variational analysis and its applications in infinite-dimensional spaces; see [20, 21] for a comprehensive theory and numerous applications. It has been well recognized that the SNC property (63) is implied in arbitrary Banach spaces by certain Lipschitzian requirements imposed on the set in question, in particular, by the *compactly epi-Lipschitzian* (CEL) property of  $\Theta$  around  $\bar{x}$  in the sense of Borwein and Strójas that follows from (48); see [20, Subsection 1.1.4] and [11] for more details and references.

**Remark 5.7 (Case of approximately convex constraint sets *at* versus *around* the reference point).** Some modifications of the proofs given in Proposition 4.3, Theorem 5.3, and Theorem 5.4 allow us to justify the necessary optimality conditions of Theorem 5.4 under the assumption that the constraint set  $\Theta$  is approximately convex

only *at* (versus *around*) the reference point  $\bar{x}$ . The main idea behind these changes is to keep the subdifferential construction (39) via the classical directional derivative  $\psi'(\bar{x}; v)$  at  $\bar{x}$  for a locally Lipschitzian function  $\psi$  while replacing  $\psi'(x; v)$  by the *robust* Clarke's generalized directional derivative  $\psi^\circ(x; v)$  of  $\psi$  at points nearby. This robust approximation allows us to conduct the limiting procedure in the proof of Theorem 5.3 and consequently in Theorem 5.4.

**Remark 5.8 (Extensions to other classes of regular functions and sets).** Approximate convexity is not the only type of nice/regular behavior of functions and sets. Other classes of functions and sets exhibiting locally nice convex-like properties have been extensively studied and applied in variational analysis and optimization; see, e.g., [2, 5, 21, 24, 26] and the references therein. Recently many of such notions have been unified in [24] under the name of  $\varphi$ -regularity. The latter notion postulates a property of type (42) from Proposition 4.3(ii) with respect to Fréchet subgradients. The class of  $\varphi$ -regular functions contains, in particular, all *prox-regular* functions that are highly important in many aspects of variational analysis and its applications; see the discussion and references presented after Definition 4.1. As the reader can observe from the proofs presented above, the methods developed in this paper would allow us to modify and extend the major results obtained to the case of  $\varphi$ -regularity.

**Acknowledgements.** The authors are gratefully indebted to the anonymous referee and also to Marián Fabian, Alex Kruger, Jiří Outrata, and Lionel Thibault for their valuable suggestions and remarks that allowed us to improve the original presentation.

## References

- [1] D. Aussel, J.-N. Corvellec, M. Lassonde: Mean value property and subdifferential criteria for lower semicontinuous functions, *Trans. Amer. Math. Soc.* 347 (1995) 4147–4161.
- [2] D. Aussel, A. Daniilidis, L. Thibault: Subsmooth sets: functional characterizations and related concepts, *Trans. Amer. Math. Soc.* 357 (2004) 1275–1302.
- [3] F. Bernard, L. Thibault, N. Zlateva: Characterizations of prox-regular sets in uniformly convex Banach spaces, *J. Convex Analysis* 12 (2006) 525–559.
- [4] J. M. Borwein, Y. Lucet, B. S. Mordukhovich: Compactly epi-Lipschitzian convex sets and functions in normed spaces, *J. Convex Analysis* 7 (2000) 375–393.
- [5] J. M. Borwein, Q. J. Zhu: *Techniques of Variational Analysis*, CMS Books in Mathematics 20, Springer, New York (2005).
- [6] F. H. Clarke: *Optimization and Nonsmooth Analysis*, CMS Books in Mathematics 1, Wiley, New York (1983).
- [7] R. Deville, G. Godefroy, V. Zizler: *Smoothness and Renorming in Banach Spaces*, Wiley, New York (1993).
- [8] J. Diestel: *Geometry of Banach Spaces. Selected Topics*, Lecture Notes in Mathematics 485, Springer, New York (1975).
- [9] M. Fabian, G. Godefroy, V. Montesinos, V. Zizler: Inner characterizations of weakly compactly generated Banach spaces and their relatives, *J. Math. Anal. Appl.* 297 (2004) 419–455.



- [10] M. Fabian, P. D. Loewen, B. S. Mordukhovich: Subdifferential calculus in Asplund generated spaces, *J. Math. Anal. Appl.* 322 (2006) 787–795.
- [11] M. Fabian, B. S. Mordukhovich: Sequential normal compactness versus topological normal compactness in variational analysis, *Nonlinear Anal., Theory Methods Appl.* 54A (2003) 1057–1067.
- [12] A. D. Ioffe: Coderivative compactness, metric regularity and subdifferential calculus, in: *Constructive, Experimental, and Nonlinear Analysis*, M. Théra (ed.), CMS Conf. Proc. 27, AMS, Providence (2000) 123–164.
- [13] A. D. Ioffe, J. V. Outrata: On metric and calmness qualification conditions in subdifferential calculus, *Set-Valued Anal.* 16 (2008) 199–227.
- [14] M. Lassonde: First-order rules for nonsmooth constrained optimization, *Nonlinear Anal., Theory Methods Appl.* 44A (2001) 1031–1056.
- [15] Yu. S. Ledyayev, Q. J. Zhu: Nonsmooth analysis on smooth manifolds, *Trans. Amer. Math. Soc.* 359 (2007) 3687–3732.
- [16] X. Li, J. M. Yong: *Optimal Control Theory for Infinite Dimensional Systems*, Birkhäuser, Boston (1995).
- [17] M. McAsey, L. Mou: A multiplier rule on a metric space, *J. Math. Anal. Appl.* 337 (2008) 1064–1071.
- [18] M. McAsey, L. Mou: A proof of a general maximum principle for optimal controls via a multiplier rule on metric space, *J. Math. Anal. Appl.* 337 (2008) 1072–1088.
- [19] B. S. Mordukhovich: Abstract extremal principle with applications to welfare economics, *J. Math. Anal. Appl.* 251 (2000) 187–216.
- [20] B. S. Mordukhovich: *Variational Analysis and Generalized Differentiation. I: Basic Theory*, Grundlehren der Mathematischen Wissenschaften 330, Springer, Berlin (2006).
- [21] B. S. Mordukhovich: *Variational Analysis and Generalized Differentiation. II: Applications*, Grundlehren der Mathematischen Wissenschaften 331, Springer, Berlin (2006).
- [22] B. S. Mordukhovich, N. M. Nam: Subgradient of distance functions with applications to Lipschitzian stability, *Math. Program., Ser. B* 104 (2005) 635–668.
- [23] H. V. Ngai, D. T. Luc, M. Théra: Approximate convex functions, *J. Nonlinear Convex Anal.* 1 (2000) 155–176.
- [24] H. V. Ngai, M. Théra:  $\varphi$ -regular functions in Asplund spaces, *Control Cybernet.* 35 (2007) 755–774.
- [25] R. A. Poliquin, R. T. Rockafellar, L. Thibault: Local differentiability of distance functions, *Trans. Amer. Math. Soc.* 332 (2000) 5231–5249.
- [26] R. T. Rockafellar, R. J-B. Wets: *Variational Analysis*, Grundlehren der Mathematischen Wissenschaften 317, Springer, Berlin (1998).
- [27] W. Schirotzek: *Nonsmooth Analysis*, Series Universitext, Springer, Berlin (2007).
- [28] L. Thibault, D. Zagrodny: Integration of subdifferentials of lower semicontinuous functions, *J. Math. Anal. Appl.* 189 (1995) 22–58.
- [29] Z. Wu, J. J. Ye: Equivalence between various derivatives and subdifferentials of the distance function, *J. Math. Anal. Appl.* 282 (2003) 629–647.
- [30] L. Zajíček: Differentiability of approximately convex, semiconvex and strongly paraconvex functions, *J. Convex Analysis* 15 (2008) 1–15.