Deutsche Forschungsgemeinschaft
Priority Programme 1962

Optimality conditions, approximate stationarity, and applications — a story beyond Lipschitzness

Alexander Y. Kruger, Patrick Mehlitz

SPP 1962
Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization

Preprint Number SPP1962-179

received on October 14, 2021

Edited by
SPP1962 at Weierstrass Institute for Applied Analysis and Stochastics (WIAS)
Leibniz Institute in the Forschungsverbund Berlin e.V.
Mohrenstraße 39, 10117 Berlin, Germany
E-Mail: spp1962@wias-berlin.de

World Wide Web: http://spp1962.wias-berlin.de/

Optimality conditions, approximate stationarity, and applications – a story beyond Lipschitzness

Alexander Y. Kruger* Patrick Mehlitz†

October 14, 2021

Approximate necessary optimality conditions in terms of Fréchet subgradients and normals for a rather general optimization problem with a potentially non-Lipschitzian objective function are established with the aid of Ekeland's variational principle, the fuzzy Fréchet subdifferential sum rule, and a novel notion of lower semicontinuity relative to a set-valued mapping or set. Feasible points satisfying these optimality conditions are referred to as approximately stationary. As applications, we derive a new general version of the extremal principle. Furthermore, we study approximate stationarity conditions for an optimization problem with a composite objective function and geometric constraints, a qualification condition guaranteeing that approximately stationary points of such a problem are M-stationary, and a multiplier-penalty-method which naturally computes approximately stationary points of the underlying problem. Finally, necessary optimality conditions for an optimal control problem with a non-Lipschitzian sparsity-promoting term in the objective function are established.

Keywords: Approximate stationarity, Generalized separation, Non-Lipschitzian programming, Optimality conditions, Sparse control
MSC: 49J52 49J53, 49K27, 90C30, 90C48

*Federation University Australia, Centre for Informatics and Applied Optimization, School of Engineering, Information Technology and Physical Sciences, Ballarat VIC 3353, Australia, a.kruger@federation.edu.au, https://asterius.federation.edu.au/akruger/, ORCID: 0000-0002-7861-7380

†Brandenburgische Technische Universität Cottbus–Senftenberg, Institute of Mathematics, 03046 Cottbus, Germany, mehlitz@b-tu.de, https://www.b-tu.de/fg-optimale-steuerung/team/dr-patrick-mehlitz, ORCID: 0000-0002-9355-850X

1 Introduction

Approximate stationarity conditions, claiming that, along a convergent sequence, a classical stationarity condition (like a multiplier rule) holds up to a tolerance which tends to zero, have proved to be a powerful tool in mathematical optimization throughout the last decades. The particular interest in such conditions is based on two prominent features. First, they often serve as necessary optimality conditions even in the absence of constraint qualifications. Second, different classes of solution algorithms for the computational treatment of optimization problems naturally produce sequences whose accumulation points are approximately stationary. Approximate stationarity conditions can be traced back to the early 1980s, see Kruger and Mordukhovich [1980], Kruger [1985], where they popped up as a consequence of the famous *extremal principle*. The latter geometric result, when formulated in infinite dimensions in terms of Fréchet normals, can itself be interpreted as a kind of approximate stationarity, see Kruger and Mordukhovich [1980], Kruger [2003], Mordukhovich [2006]. In Andreani et al. [2010, 2011], this fundamental concept, which is referred to as *Approximate Karush–Kuhn–Tucker (AKKT) stationarity* in these papers, has been rediscovered due to its significant relevance in the context of numerical standard nonlinear programming. A notable feature of AKKT-stationary points is the potential unboundedness of the associated sequence of Lagrange-multipliers. The latter already depicts that AKKT-stationary points do not need to satisfy the classical KKT conditions. This observation gave rise to the investigation of conditions ensuring that AKKT-stationary points actually are KKT points, see e.g. Andreani et al. [2016]. The resulting constraint qualifications for the underlying nonlinear optimization problem turned out to be comparatively weak. During the last decade, reasonable notions of approximate stationarity have been introduced for more challenging classes of optimization problems like programs with complementarity, see Andreani et al. [2019b], Ramos [2021], cardinality, see Kanzow et al. [2021], conic, see Andreani et al. [2020], nonsmooth, see Helou et al. [2020], Mehlitz [2020, 2021], and geometric constraints, see Jia et al. [2021], in the finite-dimensional situation. A generalization to optimization problems in abstract Banach spaces can be found in Börgens et al. [2020]. In all these papers, the underlying optimization problem's objective function is assumed to be locally Lipschitzian. Note that the (local) Lipschitz property of the (all but one) functions involved is a key assumption in most conventional subdifferential calculus results in infinite dimensions in convex and nonconvex settings, see e.g. the sum rules in Lemma 2.2. However, as several prominent applications like sparse portfolio selection, compressed sensing, edge-preserving image restoration, low-rank matrix completion, or signal processing, where the objective function is often only lower semicontinuous, demonstrate, Lipschitz continuity might be a restrictive property of the data. The purpose of this paper is to provide a reasonable extension of approximate stationarity to a rather general class of optimization problems in Banach spaces with a lower semicontinuous objective function and generalized equation constraints generated by a set-valued mapping in order to open the topic up to the aforementioned challenging applications.

Our general approach to a notion of approximate stationarity, which serves as a necessary optimality condition, is based on two major classical tools: *Ekeland's variational* *principle*, see Ekeland [1974], and the *fuzzy calculus* of Fréchet normals, see Ioffe [2017], Kruger [2003]. Another convenient ingredient of the theory is a new notion of lower semicontinuity of extended-real-valued functions relative to a given set-valued mapping which holds for free in finite dimensions. We illustrate our findings in the context of generalized set separation and derive a novel extremal principle which differs from the traditional one which dates back to Kruger and Mordukhovich [1980]. On the one hand, its prerequisites regarding the position of the involved sets relative to each other is slightly more restrictive than in Kruger and Mordukhovich [1980] when the classical notion of extremality, meaning that the sets of interest can be "pushed apart from each other", is used. On the other hand, our new extremal principle covers settings where extremality is based on functions which are just lower semicontinuous, and, thus, applies in more general situations. The final part of the paper is dedicated to the study of optimization problems with so-called geometric constraints, where the feasible set equals the preimage of a closed set under a smooth transformation, whose objective function is the sum of a smooth part and a merely lower semicontinuous part. First, we apply our concept of approximate stationarity to this problem class in order to obtain necessary optimality conditions. Furthermore, we introduce an associated qualification condition which guarantees M-stationarity of approximately stationary points. As we will show, this generalizes related considerations from Chen et al. [2017], Guo and Ye [2018] which were done in a completely finite-dimensional setting. Second, we suggest an augmented Lagrangian method for the numerical solution of geometrically constrained programs and show that it computes approximately stationary points in our new sense. Finally, we use our theory in order to state necessary optimality conditions for optimal control problems with a non-Lipschitzian so-called *sparsity-promoting* term in the objective function, see Ito and Kunisch [2014], Wachsmuth [2019], which enforces optimal controls to be zero on large parts of the domain.

The remaining parts of the paper are organized as follows. In Section 2, we comment on the notation which is used in this manuscript and recall some fundamentals from variational analysis. Section 3 is dedicated to the study of a new notion of lower semicontinuity of an extended-real-valued function relative to a given set-valued mapping or set. We derive necessary optimality conditions of approximate stationarity type for rather general optimization problems in Section 4. This is used in Section 5 in order to derive a novel extremal principle in generalized set separation. Furthermore, we apply our findings from Section 4 in Section 6 in order to state necessary optimality conditions of approximate stationarity type for optimization problems in Banach spaces with geometric constraints and a composite objective function. Based on that, we derive a new qualification condition ensuring M-stationarity of local minimizers, see Section 6.1, an augmented Lagrangian method which naturally computes approximately stationary points, see Section 6.2, and necessary optimality conditions for optimal control problems with a sparsity-promoting term in the objective function, see Section 6.3. Some concluding remarks close the paper in Section 7.

2 Notation and preliminaries

2.1 Basic notation

Our basic notation is standard, see e.g. Ioffe [2017], Mordukhovich [2006], Rockafellar and Wets [1998]. The symbols $\mathbb{R}$ and $\mathbb{N}$ denote the sets of all real numbers and all positive integers, respectively. Throughout the paper, X and Y are either metric or Banach spaces (although many facts, particularly, most of the definitions in Section 2.2, are valid in arbitrary normed vector spaces, i.e., do not require the spaces to be complete). For brevity, we use the same notations $d(\cdot,\cdot)$ and $\|\cdot\|$ for distances and norms in all spaces. Banach spaces are often treated as metric spaces with the distance determined by the norm in the usual way. The distance from a point $x \in X$ to a set $\Omega \subset X$ in a metric space X is defined by $\operatorname{dist}_\Omega(x) := \inf_{u\in\Omega} d(x,u)$, and we use the convention $\operatorname{dist}_\varnothing(x) := +\infty$. Throughout the paper, $\overline{\Omega}$ and $\operatorname{int}\Omega$ denote the closure and the interior of Ω, respectively. Whenever X is a Banach space, $\{x_k\}_{k\in\mathbb{N}} \subset X$ is a sequence, and $\bar{x} \in X$ is some point, we exploit $x_k \to \bar{x}$ ($x_k \rightharpoonup \bar{x}$) in order to denote the strong (weak) convergence of $\{x_k\}_{k\in\mathbb{N}}$ to $\bar{x}$. Similarly, we use $x_k^* \overset{*}{\rightharpoonup} x^*$ in order to express that a sequence $\{x_k^*\}_{k\in\mathbb{N}} \subset X^*$ converges weakly* to $x^* \in X^*$. Finally, $x_k \to_\Omega \bar{x}$ means that $\{x_k\}_{k\in\mathbb{N}} \subset \Omega$ converges strongly to $\bar{x}$. In case where X is a Hilbert space and $K \subset X$ is a closed, convex set, we denote by $P_K\colon X \to X$ the projection map associated with K.

If X is a Banach space, its topological dual is denoted by X^*, while $\langle\cdot,\cdot\rangle\colon X^* \times X \to \mathbb{R}$ denotes the bilinear form defining the pairing between the two spaces. If not explicitly stated otherwise, products of (primal) metric or Banach spaces are equipped with the maximum distances or norms, e.g., $\|(x,y)\| := \max(\|x\|,\|y\|)$ for all $(x,y) \in X \times Y$. Note that the corresponding dual norm is the sum norm given by $\|(x^*,y^*)\| := \|x^*\| + \|y^*\|$ for all $(x^*,y^*) \in X^* \times Y^*$. The open unit balls in the primal and dual spaces are denoted by $\mathbb{B}$ and $\mathbb{B}^*$, respectively, while the corresponding closed unit balls are denoted by $\overline{\mathbb{B}}$ and $\overline{\mathbb{B}}^*$, respectively. The notations $B_\delta(x)$ and $\overline{B}_\delta(x)$ stand, respectively, for the open and closed balls with center x and radius $\delta > 0$ in X.

For an extended-real-valued function $\varphi\colon X \to \mathbb{R}_\infty := \mathbb{R} \cup \{+\infty\}$, its domain and epigraph are defined by $\operatorname{dom}\varphi := \{x \in X \,|\, \varphi(x) < +\infty\}$ and $\operatorname{epi}\varphi := \{(x,\mu) \in X \times \mathbb{R} \,|\, \varphi(x) \le \mu\}$, respectively. For each set $\Omega \subset X$, we set $\varphi_\Omega := \varphi + i_\Omega$ where $i_\Omega\colon X \to \mathbb{R}_\infty$ is the so-called indicator function of Ω which equals zero on Ω and is set to $+\infty$ on $X \setminus \Omega$.

A set-valued mapping $\Upsilon\colon X \rightrightarrows Y$ between metric spaces X and Y is a mapping, which assigns to every $x \in X$ a (possibly empty) set $\Upsilon(x) \subset Y$. We use the notations $\operatorname{gph}\Upsilon := \{(x,y) \in X \times Y \,|\, y \in \Upsilon(x)\}$, $\operatorname{Im}\Upsilon := \bigcup_{x\in X}\Upsilon(x)$, and $\operatorname{dom}\Upsilon := \{x \in X \,|\, \Upsilon(x) \ne \varnothing\}$ for the graph, the image, and the domain of Υ, respectively. Furthermore, $\Upsilon^{-1}\colon Y \rightrightarrows X$ given by $\Upsilon^{-1}(y) := \{x \in X \,|\, y \in \Upsilon(x)\}$ for all $y \in Y$ is referred to as the inverse of Υ. Assuming that $\bar{x} \in \operatorname{dom}\Upsilon$ is fixed,

$$\limsup_{x\to\bar{x}} \Upsilon(x) := \{y \in Y \,|\, \exists \{(x_k,y_k)\}_{k\in\mathbb{N}} \subset \operatorname{gph}\Upsilon\colon\; x_k \to \bar{x},\, y_k \to y\}$$

is referred to as the (strong) outer limit of Υ at $\bar{x}$. Finally, if X is a Banach space, for a set-valued mapping $\Xi\colon X \rightrightarrows X^*$ and $\bar{x} \in \operatorname{dom}\Xi$, we use

$$\mathrm{w}^* - \limsup_{x\to\bar{x}} \Xi(x) := \left\{ x^* \in X^* \,\middle|\, \exists\{(x_k, x_k^*)\}_{k\in\mathbb{N}} \subset \operatorname{gph}\Xi\colon\; x_k \to \bar{x},\, x_k^* \overset{*}{\rightharpoonup} x^* \right\}$$

in order to denote the outer limit of Ξ at $\bar{x}$ when equipping X^* with the weak* topology. Let us note that both outer limits from above are limits in the sense of Painlevé–Kuratowski.

Recall that a Banach space is a so-called Asplund space if every continuous, convex function on an open convex set is Fréchet differentiable on a dense subset, or equivalently, if the dual of each separable subspace is separable as well. We refer the reader to Phelps [1993], Mordukhovich [2006] for discussions about and characterizations of Asplund spaces. We would like to note that all reflexive, particularly, all finite-dimensional Banach spaces possess the Asplund property.

2.2 Variational analysis

The subsequently introduced notions of variational analysis and generalized differentiation are standard, see e.g. Kruger [2003], Mordukhovich [2006].

Given a subset Ω of a Banach space X, a point $\bar{x} \in \Omega$, and a number $\varepsilon \geq 0$, the nonempty, closed, convex set

$$N_{\Omega,\varepsilon}(\bar{x}) := \left\{ x^* \in X^* \,\middle|\, \limsup_{x\to_\Omega \bar{x},\, x\neq\bar{x}} \frac{\langle x^*, x-\bar{x}\rangle}{\|x-\bar{x}\|} \leq \varepsilon \right\} \tag{2.1}$$

is the set of ε-normals to Ω at $\bar{x}$. In case $\varepsilon = 0$, it is a closed, convex cone called Fréchet normal cone to Ω at $\bar{x}$. In this case, we drop the subscript ε in the above notation and simply write

$$N_\Omega(\bar{x}) := \left\{ x^* \in X^* \,\middle|\, \limsup_{x\to_\Omega \bar{x},\, x\neq\bar{x}} \frac{\langle x^*, x-\bar{x}\rangle}{\|x-\bar{x}\|} \leq 0 \right\}.$$

Based on (2.1), one can define the more robust limiting normal cone to Ω at $\bar{x}$ by means of a limiting procedure:

$$\overline{N}_\Omega(\bar{x}) := \mathrm{w}^* - \limsup_{x\to_\Omega \bar{x},\, \varepsilon\downarrow 0} N_{\Omega,\varepsilon}(x).$$

Whenever X is an Asplund space, the above definition admits the following simplification:

$$\overline{N}_\Omega(\bar{x}) = \mathrm{w}^* - \limsup_{x\to_\Omega \bar{x}} N_\Omega(x).$$

If Ω is a convex set, the Fréchet and limiting normal cones reduce to the normal cone in the sense of convex analysis, i.e.,

$$N_\Omega(\bar{x}) = \overline{N}_\Omega(\bar{x}) = \{x^* \in X^* \,|\, \langle x^*, x-\bar{x}\rangle \leq 0 \,\forall x \in \Omega\}.$$

For a lower semicontinuous function $\varphi\colon X\to\mathbb{R}_\infty$, defined on a Banach space X, its Fréchet subdifferential at $\bar x\in\operatorname{dom}\varphi$ is defined as

$$\begin{aligned}\partial\varphi(\bar x) : &= \left\{x^*\in X^*\,\middle|\,\liminf_{x\to\bar x,\,x\neq\bar x}\frac{\varphi(x)-\varphi(\bar x)-\langle x^*,x-\bar x\rangle}{\|x-\bar x\|}\geq 0\right\}\\ &= \{x^*\in X^*\,|\,(x^*,-1)\in N_{\operatorname{epi}\varphi}(\bar x,\varphi(\bar x))\}\,.\end{aligned}$$

The limiting and singular limiting subdifferential of φ at $\bar x$ are defined, respectively, by means of

$$\begin{aligned}\overline{\partial}\varphi(\bar x) &:= \left\{x^*\in X^*\,\middle|\,(x^*,-1)\in\overline{N}_{\operatorname{epi}\varphi}(\bar x,\varphi(\bar x))\right\},\\ \overline{\partial}^\infty\varphi(\bar x) &:= \left\{x^*\in X^*\,\middle|\,(x^*,0)\in\overline{N}_{\operatorname{epi}\varphi}(\bar x,\varphi(\bar x))\right\}.\end{aligned}$$

Note that in case where X is an Asplund space, we have

$$\begin{aligned}\overline{\partial}\varphi(\bar x) &= \mathrm{w}^*-\limsup_{x\to\bar x,\,\varphi(x)\to\varphi(\bar x)}\partial\varphi(x),\\ \overline{\partial}^\infty\varphi(\bar x) &= \mathrm{w}^*-\limsup_{x\to\bar x,\,\varphi(x)\to\varphi(\bar x),\,t\downarrow 0} t\,\partial\varphi(x),\end{aligned}$$

see [Mordukhovich, 2006, Theorems 2.34 and 2.38]. If φ is convex, the Fréchet and limiting subdifferential reduce to the subdifferential in the sense of convex analysis, i.e.,

$$\partial\varphi(\bar x)=\overline{\partial}\varphi(\bar x)=\{x^*\in X^*\,|\,\varphi(x)-\varphi(\bar x)-\langle x^*,x-\bar x\rangle\geq 0\,\forall x\in X\}\,.$$

By convention, we set $N_\Omega(x)=\overline{N}_\Omega(x):=\varnothing$ if $x\notin\Omega$ and $\partial\varphi(x)=\overline{\partial}\varphi(x)=\overline{\partial}^\infty\varphi(x):=\varnothing$ if $x\notin\operatorname{dom}\varphi$. It is easy to check that $N_\Omega(\bar x)=\partial i_\Omega(\bar x)$ and $\overline{N}_\Omega(\bar x)=\overline{\partial}i_\Omega(\bar x)$.

For a set-valued mapping $\Upsilon\colon X\rightrightarrows Y$ between Banach spaces, its Fréchet coderivative at $(\bar x,\bar y)\in\operatorname{gph}\Upsilon$ is defined as

$$\forall y^*\in Y^*:\quad D^*\Upsilon(\bar x,\bar y)(y^*):=\{x^*\in X^*\,|\,(x^*,-y^*)\in N_{\operatorname{gph}\Upsilon}(\bar x,\bar y)\}\,.$$

The proof of our main result Theorem 4.1 relies on certain fundamental results of variational analysis: Ekeland's variational principle, see e.g. [Aubin and Frankowska, 2009, Section 3.3] or Ekeland [1974], and two types of subdifferential sum rules which address the subdifferential in the sense of convex analysis, see e.g. [Phelps, 1993, Theorem 3.16], and the Fréchet subdifferential, see e.g. [Fabian, 1989, Theorem 3]. Below, we provide these results for completeness.

Lemma 2.1. *Let X be a complete metric space, $\varphi\colon X\to\mathbb{R}_\infty$ be lower semicontinuous and bounded from below, $\bar x\in\operatorname{dom}\varphi$, and $\varepsilon>0$. Then there exists a point $\hat x\in X$ which satisfies the following conditions:*

(a) $\varphi(\hat x)\leq\varphi(\bar x)$;

(b) $\forall x\in X:\quad \varphi(x)+\varepsilon d(x,\hat x)\geq\varphi(\hat x)$.

Lemma 2.2. *Let X be a Banach space, $\varphi_1, \varphi_2 \colon X \to \mathbb{R}_\infty$, and $\bar{x} \in \operatorname{dom} \varphi_1 \cap \operatorname{dom} \varphi_2$. Then the following assertions hold.*

(a) ***Convex sum rule****. Let φ_1 and φ_2 be convex, and φ_1 be continuous at a point in $\operatorname{dom} \varphi_2$. Then $\partial(\varphi_1 + \varphi_2)(\bar{x}) = \partial\varphi_1(\bar{x}) + \partial\varphi_2(\bar{x})$.*

(b) ***Fuzzy sum rule****. Let X be Asplund, φ_1 be Lipschitz continuous around $\bar{x}$, and φ_2 be lower semicontinuous in a neighborhood of $\bar{x}$. Then, for each $x^* \in \partial(\varphi_1+\varphi_2)(\bar{x})$ and $\varepsilon > 0$, there exist $x_1, x_2 \in X$ with $\|x_i - \bar{x}\| < \varepsilon$ and $|\varphi_i(x_i) - \varphi_i(\bar{x})| < \varepsilon$, $i = 1, 2$, such that $x^* \in \partial\varphi_1(x_1) + \partial\varphi_2(x_2) + \varepsilon\mathbb{B}^*$.*

We will need representations of the subdifferentials of the distance function collected in the next lemma. These results are taken from [Kruger, 2003, Proposition 1.30], [Ioffe, 2017, Theorem 4.40], and [Penot, 2013, Section 3.5.2, Exercise 6].

Lemma 2.3. *Let X be a Banach space, $\Omega \subset X$ be nonempty and closed, and $\bar{x} \in X$. Then the following assertions hold.*

(a) If $\bar{x} \in \Omega$, then $\partial \operatorname{dist}_\Omega(\bar{x}) = N_\Omega(\bar{x}) \cap \overline{\mathbb{B}}^*$.

(b) If $\bar{x} \notin \Omega$ and either X is Asplund or Ω is convex, then, for each $x^* \in \partial \operatorname{dist}_\Omega(\bar{x})$ and each $\varepsilon > 0$, there exist $x \in \Omega$ and $u^* \in N_\Omega(x)$ such that $\|x - \bar{x}\| < \operatorname{dist}_\Omega(\bar{x}) + \varepsilon$ and $\|x^* - u^*\| < \varepsilon$.

Let us briefly mention that assertion (b) of Lemma 2.3 can obviously be improved when the set of projections of $\bar{x}$ onto Ω is nonempty, see [Mordukhovich, 2006, Proposition 1.102]. This is always the case if Ω is a nonempty, closed, convex subset of a reflexive Banach space, since in this case Ω is weakly sequentially compact while the norm is weakly sequentially lower semicontinuous.

The conditions in the final definition of this subsection are standard, see e.g. Klatte and Kummer [2002], Kruger [2009].

Definition 2.4. *Let X be a metric space, $\varphi \colon X \to \mathbb{R}_\infty$, and $\bar{x} \in \operatorname{dom} \varphi$.*

(a) We call $\bar{x}$ a stationary point of φ *if $\liminf_{x\to\bar{x},\, x\neq\bar{x}} \frac{\varphi(x)-\varphi(\bar{x})}{d(x,\bar{x})} \geq 0$.*

(b) Let $\varepsilon > 0$ and $U \subset X$ with $\bar{x} \in U$. We call $\bar{x}$ an ε-minimal point of φ on U *if $\inf_{x\in U} \varphi(x) > \varphi(\bar{x}) - \varepsilon$. If $U = X$, $\bar{x}$ is called a* globally ε-minimal point of φ.

In the subsequent remark, we interrelate the concepts from Definition 2.4.

Remark 2.5. *For a metric space X, $\varphi \colon X \to \mathbb{R}_\infty$, and $\bar{x} \in \operatorname{dom} \varphi$, the following assertions hold.*

(a) If $\bar{x}$ is a local minimizer of φ, then it is a stationary point of φ.

(b) If $\bar{x}$ is a stationary point of φ, then, for each $\varepsilon > 0$ and each sufficiently small $\delta > 0$, $\bar{x}$ is an $\varepsilon\delta$-minimal point of φ on $B_\delta(\bar{x})$.

(c) If X is a normed space, then $\bar{x}$ is a stationary point of φ if and only if $0 \in \partial\varphi(\bar{x})$.

3 Novel notions of semicontinuity

In this paper, we exploit new notions of lower semicontinuity of extended-real-valued functions relative to a given set-valued mapping or set. Here, we first introduce the concepts of interest before studying their properties and presenting sufficient conditions for their validity.

3.1 Lower semicontinuity of a function relative to a set-valued mapping or set

Let us start with the definition of the property of our interest.

Definition 3.1. *Fix metric spaces X and Y, $\Phi\colon X \rightrightarrows Y$, $\varphi\colon X \to \mathbb{R}_\infty$, and $\bar{y} \in Y$.*

(a) *Let a subset $U \subset X$ be such that $U \cap \Phi^{-1}(\bar{y}) \cap \operatorname{dom} \varphi \neq \varnothing$. The function φ is* lower semicontinuous on U relative to Φ at $\bar{y}$ *if*

$$\inf_{u\in\Phi^{-1}(\bar{y})\cap U} \varphi(u) \leq \inf_{\substack{U'+\rho\mathbb{B}\subset U,\\ \rho>0}} \liminf_{\substack{x\in U',\, y\to\bar{y},\\ \operatorname{dist}_{\operatorname{gph}\Phi}(x,y)\to 0}} \varphi(x). \tag{3.1}$$

(b) *Let $\bar{x} \in \Phi^{-1}(\bar{y}) \cap \operatorname{dom} \varphi$. The function φ is* lower semicontinuous near $\bar{x}$ relative to Φ at $\bar{y}$ *if there is a $\delta_0 > 0$ such that, for each $\delta \in (0, \delta_0)$, φ is lower semicontinuous on $\overline{B}_\delta(\bar{x})$ relative to Φ at $\bar{y}$.*

Inequality (3.1) can be strict, see Example 3.4 below. Note that whenever (3.1) holds with a subset $U \subset X$, it also holds with $\overline{U}$ in place of U. The converse implication is not true in general, see Example 3.5 below. Particularly, a function which is lower semicontinuous on a set U relative to Φ at $\bar{y}$ may fail to have this property on a smaller set. This shortcoming explains the idea behind Definition 3.1 (b). Furthermore, we have the following result.

Lemma 3.2. *Fix metric spaces X and Y, $\Phi\colon X \rightrightarrows Y$, $\varphi\colon X \to \mathbb{R}_\infty$, $(\bar{x}, \bar{y}) \in \operatorname{gph} \Phi$, and a subset $U \subset X$ with $\bar{x} \in U \cap \operatorname{dom} \varphi$. Assume that $\bar{x}$ is a minimizer of φ on U. If φ is lower semicontinuous on U relative to Φ at $\bar{y}$, then it is lower semicontinuous on $\widehat{U}$ relative to Φ at $\bar{y}$ for each subset $\widehat{U}$ satisfying $\bar{x} \in \widehat{U} \subset U$.*

Proof. For each subset $\widehat{U}$ satisfying $\bar{x} \in \widehat{U} \subset U$, we find

$$\begin{aligned}\inf_{u\in\Phi^{-1}(\bar{y})\cap\widehat{U}} \varphi(u) = \varphi(\bar{x}) &= \inf_{u\in\Phi^{-1}(\bar{y})\cap U} \varphi(u)\\ &\leq \inf_{\substack{U'+\rho\mathbb{B}\subset U,\\ \rho>0}} \liminf_{\substack{x\in U',\, y\to\bar{y},\\ \operatorname{dist}_{\operatorname{gph}\Phi}(x,y)\to 0}} \varphi(x) \leq \inf_{\substack{U'+\rho\mathbb{B}\subset \widehat{U},\\ \rho>0}} \liminf_{\substack{x\in U',\, y\to\bar{y},\\ \operatorname{dist}_{\operatorname{gph}\Phi}(x,y)\to 0}} \varphi(x),\end{aligned}$$

which shows the claim. $\square$

The properties in the next definition are particular cases of the ones in Definition 3.1, corresponding to the set-valued mapping $\Phi\colon X \rightrightarrows Y$ whose graph is given by $\operatorname{gph} \Phi := \Omega \times Y$, where $\Omega \subset X$ is a fixed set and Y can be an arbitrary metric space, e.g., one can take $Y := \mathbb{R}$. Observe that in this case, $\Phi^{-1}(y) = \Omega$ is valid for all $y \in Y$.

Definition 3.3. *Fix a metric space X, $\varphi\colon X \to \mathbb{R}_\infty$, and $\Omega \subset X$.*

(a) Let a subset $U \subset X$ be such that $U \cap \Omega \cap \operatorname{dom}\varphi \neq \varnothing$. The function φ is lower semicontinuous on U relative to Ω *if*

$$\inf_{u\in\Omega\cap U}\varphi(u) \leq \inf_{\substack{U'+\rho\mathbb{B}\subset U,\\ \rho>0}} \liminf_{\substack{x\in U',\\ \operatorname{dist}_\Omega(x)\to 0}} \varphi(x). \tag{3.2}$$

(b) Let $\bar{x} \in \Omega \cap \operatorname{dom}\varphi$. The function φ is lower semicontinuous near $\bar{x}$ relative to Ω *if there is a $\delta_0 > 0$ such that, for each $\delta \in (0, \delta_0)$, φ is lower semicontinuous on $\overline{B}_\delta(\bar{x})$ relative to Ω.*

The subsequent example shows that (3.2) can be strict.

Example 3.4. *Consider the lower semicontinuous function $\varphi\colon \mathbb{R} \to \mathbb{R}$ given by $\varphi(x) := 0$ if $x \leq 0$ and $\varphi(x) := 1$ if $x > 0$, and the sets $\Omega = U := [0, 1] \subset \mathbb{R}$. Then $\inf_{u\in\Omega\cap U}\varphi(u) = 0$, while if a subset U' satisfies $U' + \rho\mathbb{B} \subset U$ for some $\rho > 0$, then $U' \subset (0, 1)$, and consequently $\varphi(x) = 1$ for all $x \in U'$. Hence, the right-hand side of (3.2) equals 1.*

A function which is lower semicontinuous on a set U relative to Ω may fail to have this property on a smaller set.

Example 3.5. *Consider the function $\varphi\colon \mathbb{R} \to \mathbb{R}$ given by $\varphi(x) := 0$ if $x \leq 0$, and $\varphi(x) := -1$ if $x > 0$, the set $\Omega := \{0, 1\} \subset \mathbb{R}$, and the point $\bar{x} := 0$. Consider the closed interval $U_1 := [-1, 1]$. We find $\inf_{u\in\Omega\cap U_1}\varphi(u) = -1$ which is the global minimal value of φ on $\mathbb{R}$. Hence, φ is lower semicontinuous on U_1 relative to Ω by Definition 3.3. For $U_2 := (-1, 1)$, we find $\inf_{u\in\Omega\cap U_2}\varphi(u) = 0$. Moreover, choosing $U' := (-1/2, 1/2)$ and $x_k := 1/(k + 2)$ for each $k \in \mathbb{N}$, we find $U' + \frac{1}{2}\mathbb{B} \subset U_2$, $\{x_k\}_{k\in\mathbb{N}} \subset U'$, $d(x_k, \bar{x}) \to 0$, and $\varphi(x_k) \to -1$, i.e., φ is not lower semicontinuous on U_2 relative to Ω by definition. Note that $\bar{x}$ is a local minimizer of φ on Ω but not on U_1 or U_2.*

In the next two statements, we present sequential characterizations of the properties from Definition 3.1 (a) and Definition 3.3 (a).

Proposition 3.6. *Fix metric spaces X and Y, $\Phi\colon X \rightrightarrows Y$, $\varphi\colon X \to \mathbb{R}_\infty$, $\bar{y} \in Y$, and a subset $U \subset X$ with $U \cap \Phi^{-1}(\bar{y}) \cap \operatorname{dom}\varphi \neq \varnothing$. Then φ is lower semicontinuous on U relative to Φ at $\bar{y}$ if and only if*

$$\inf_{u\in\Phi^{-1}(\bar{y})\cap U}\varphi(u) \leq \liminf_{k\to+\infty}\varphi(x_k)$$

for all sequences $\{(x_k, y_k)\}_{k\in\mathbb{N}} \subset X \times Y$ satisfying $y_k \to \bar{y}$, $\operatorname{dist}_{\operatorname{gph}\Phi}(x_k, y_k) \to 0$, and $\{x_k\}_{k\in\mathbb{N}} + \rho\mathbb{B} \subset U$ for some $\rho > 0$.

Proof. We need to show that the right-hand side of (3.1) equals the infimum over all numbers $\liminf_{k\to+\infty}\varphi(x_k)$ where the sequence $\{(x_k, y_k)\}_{k\in\mathbb{N}} \subset X \times Y$ needs to satisfy

$y_k \to \bar{y}$, $\operatorname{dist}_{\operatorname{gph}\Phi}(x_k, y_k) \to 0$, and $\{x_k\}_{k\in\mathbb{N}} + \rho\mathbb{B} \subset U$ for some $\rho > 0$. Let $\{(x_k, y_k)\}_{k\in\mathbb{N}}$ be such a sequence. Then

$$\inf_{\substack{U'+\rho\mathbb{B}\subset U,\\ \rho>0}} \liminf_{\substack{x\in U',\, y\to\bar{y},\\ \operatorname{dist}_{\operatorname{gph}\Phi}(x,y)\to 0}} \varphi(x) \le \liminf_{\substack{x\in\{x_k\}_{k\in\mathbb{N}},\, y\to\bar{y},\\ \operatorname{dist}_{\operatorname{gph}\Phi}(x,y)\to 0}} \varphi(x) \le \liminf_{k\to+\infty} \varphi(x_k).$$

Conversely, let the right-hand side of (3.1) be finite, and choose $\varepsilon > 0$ arbitrarily. Then there exist a subset $\widehat{U} \subset U$ and a number $\hat{\rho} > 0$ such that $\widehat{U} + \hat{\rho}\mathbb{B} \subset U$ and

$$\liminf_{k\to+\infty} \inf_{\substack{x\in\widehat{U},\, d(y,\bar{y})<\frac{1}{k},\\ \operatorname{dist}_{\operatorname{gph}\Phi}(x,y)<\frac{1}{k}}} \varphi(x) = \liminf_{\substack{x\in\widehat{U},\, y\to\bar{y},\\ \operatorname{dist}_{\operatorname{gph}\Phi}(x,y)\to 0}} \varphi(x) < \inf_{\substack{U'+\rho\mathbb{B}\subset U,\\ \rho>0}} \liminf_{\substack{x\in U',\, y\to\bar{y},\\ \operatorname{dist}_{\operatorname{gph}\Phi}(x,y)\to 0}} \varphi(x) + \varepsilon.$$

For each $k \in \mathbb{N}$ such that $\inf_{x\in\widehat{U},\, d(y,\bar{y})<\frac{1}{k},\, \operatorname{dist}_{\operatorname{gph}\Phi}(x,y)<\frac{1}{k}} \varphi(x)$ is finite, there is a tuple $(x_k, y_k) \in X \times Y$ such that $x_k \in \widehat{U}$, $d(y_k, \bar{y}) < 1/k$, $\operatorname{dist}_{\operatorname{gph}\Phi}(x_k, y_k) < 1/k$, and

$$\varphi(x_k) < \inf_{\substack{x\in\widehat{U},\, d(y,\bar{y})<\frac{1}{k},\\ \operatorname{dist}_{\operatorname{gph}\Phi}(x,y)<\frac{1}{k}}} \varphi(x) + \frac{1}{k}.$$

Considering the tail of the sequences, if necessary, we have $\{x_k\}_{k\in\mathbb{N}} + \hat{\rho}\mathbb{B} \subset U$, $y_k \to \bar{y}$, $\operatorname{dist}_{\operatorname{gph}\Phi}(x_k, y_k) \to 0$, and

$$\liminf_{k\to+\infty} \varphi(x_k) < \inf_{\substack{U'+\rho\mathbb{B}\subset U,\\ \rho>0}} \liminf_{\substack{x\in U',\, y\to\bar{y},\\ \operatorname{dist}_{\operatorname{gph}\Phi}(x,y)\to 0}} \varphi(x) + \varepsilon.$$

As the number ε has been chosen arbitrarily, this proves the converse part in the present setting. If the right-hand side of (3.1) equals $-\infty$, then for each $M > 0$, we find a subset $\widehat{U} \subset U$ and a number $\hat{\rho} > 0$ such that $\widehat{U} + \hat{\rho}\mathbb{B} \subset U$ and

$$\liminf_{\substack{x\in\widehat{U},\, y\to\bar{y},\\ \operatorname{dist}_{\operatorname{gph}\Phi}(x,y)\to 0}} \varphi(x) < -M.$$

Hence, there is a sequence $\{(x_k, y_k)\}_{k\in\mathbb{N}} \subset X \times Y$ such that $\{x_k\}_{k\in\mathbb{N}} + \hat{\rho}\mathbb{B} \subset U$, $y_k \to \bar{y}$, and $\operatorname{dist}_{\operatorname{gph}\Phi}(x_k, y_k) \to 0$ as $k \to +\infty$ while $\liminf_{k\to+\infty} \varphi(x_k) < -M$. Taking the infimum over all $M > 0$ now completes the proof of the assertion. $\square$

Corollary 3.7. *Let X be a metric space, $\varphi\colon X \to \mathbb{R}_\infty$, and $\Omega, U \subset X$ be sets with $\Omega \cap U \cap \operatorname{dom}\varphi \neq \varnothing$. Then φ is lower semicontinuous on U relative to Ω if and only if*

$$\inf_{u\in\Omega\cap U} \varphi(u) \le \liminf_{k\to+\infty} \varphi(x_k) \tag{3.3}$$

for all sequences $\{x_k\}_{k\in\mathbb{N}} \subset X$ satisfying $\operatorname{dist}_\Omega(x_k) \to 0$, and $\{x_k\}_{k\in\mathbb{N}} + \rho\mathbb{B} \subset U$ for some $\rho > 0$.

3.2 Sufficient conditions for lower semicontinuity of a function relative to a set-valued mapping

As we will demonstrate below, the property from Definition 3.1 (a) is valid whenever the involved function φ and the set-valued mapping Φ enjoy certain semicontinuity properties, i.e., it can be decomposed into two independent properties regarding the two main data objects. This will be beneficial in order to identify scenarios where the new concept applies.

The upper semicontinuity properties of a set-valued mapping that we state in the following two definitions seem to fit well for this purpose (in combination with the corresponding lower semicontinuity properties of a function).

Definition 3.8. *Fix metric spaces X and Y, $S\colon Y \rightrightarrows X$, and $\bar{y} \in \operatorname{dom} S$. The mapping S is* upper semicontinuous at $\bar{y}$ *if*

$$\lim_{x \in S(y),\, y \to \bar{y}} \operatorname{dist}_{S(\bar{y})}(x) = 0.$$

Definition 3.9. *Fix a Banach space X, a metric space Y, $S\colon Y \rightrightarrows X$, and $\bar{y} \in \operatorname{dom} S$. The mapping S is* partially weakly sequentially upper semicontinuous at $\bar{y}$ *if $x \in S(\bar{y})$ holds for each sequence $\{(y_k, x_k)\}_{k\in\mathbb{N}} \subset \operatorname{gph} S$ which satisfies $y_k \to \bar{y}$ and $x_k \rightharpoonup x$.*

For a discussion of the property in Definition 3.8, we refer the reader to [Klatte and Kummer, 2002, p. 10]. The property in Definition 3.9 can be interpreted as the usual sequential upper semicontinuity if X is equipped with the weak topology. In case where Y is a Banach space, this property is inherent whenever the graph of the underlying set-valued mapping is weakly sequentially closed which is naturally given whenever the latter is convex and closed. Obviously, each closed-graph set-valued mapping with a finite-dimensional image space is partially weakly sequentially upper semicontinuous.

Proposition 3.10. *Fix metric spaces X and Y, $\Phi\colon X \rightrightarrows Y$, and $\varphi\colon X \to \mathbb{R}_\infty$. Let $\bar{y} \in Y$ and a subset $U \subset X$ with $U \cap \Phi^{-1}(\bar{y}) \cap \operatorname{dom} \varphi \neq \varnothing$ be arbitrarily chosen. Define $S\colon Y \rightrightarrows X$ by $S(y) := \Phi^{-1}(y) \cap U$ for all $y \in Y$. If one of the following criteria holds, then φ is lower semicontinuous on U relative to Φ at $\bar{y}$:*

(a) φ is lower semicontinuous on U relative to $\Phi^{-1}(\bar{y})$ and S is upper semicontinuous at $\bar{y}$;

(b) X is a reflexive Banach space, U is closed and convex, φ is weakly sequentially lower semicontinuous on U, and S is partially weakly sequentially upper semicontinuous at $\bar{y}$.

Proof. Let a sequence $\{(x_k, y_k)\}_{k\in\mathbb{N}} \subset X \times Y$ satisfying $y_k \to \bar{y}$, $\operatorname{dist}_{\operatorname{gph}\Phi}(x_k, y_k) \to 0$, and $\{x_k\}_{k\in\mathbb{N}} + \rho\mathbb{B} \subset U$ for some $\rho > 0$ be arbitrarily chosen. There exists a sequence $\{(x'_k, y'_k)\}_{k\in\mathbb{N}} \subset \operatorname{gph}\Phi$ such that $d((x'_k, y'_k), (x_k, y_k)) \to 0$. Hence, $y'_k \to \bar{y}$ and, for all sufficiently large $k \in \mathbb{N}$, we have $d(x'_k, x_k) < \rho$, and, consequently, $x'_k \in U$.

(a) By Definition 3.8, $\operatorname{dist}_{\Phi^{-1}(\bar{y})}(x'_k) \to 0$. Then $\operatorname{dist}_{\Phi^{-1}(\bar{y})}(x_k) \to 0$ and, by Corollary 3.7, inequality (3.3) holds, where $\Omega := \Phi^{-1}(\bar{y})$.

(b) Passing to a subsequence (without relabeling), we can assume $x_k \rightharpoonup \hat{x}$ for some $\hat{x} \in \overline{\operatorname{conv}}\{x_k\}_{k\in\mathbb{N}} \subset U$ since $\{x_k\}_{k\in\mathbb{N}}$ is a bounded sequence of a reflexive Banach space and U is convex as well as closed. Hence, we find $\varphi(\hat{x}) \leq \liminf_{k\to+\infty} \varphi(x_k)$ by weak sequential lower semicontinuity of φ. Obviously, we have $x'_k \rightharpoonup \hat{x}$. By Definition 3.9, $\hat{x} \in \Phi^{-1}(\bar{y})$ holds true. Thus, $\inf_{u\in\Phi^{-1}(\bar{y})\cap U} \varphi(u) \leq \varphi(\hat{x}) \leq \liminf_{k\to+\infty} \varphi(x_k)$.

As the sequence $\{(x_k, y_k)\}_{k\in\mathbb{N}}$ has been chosen arbitrarily, the conclusion follows from Proposition 3.6. □

The next assertion is an immediate consequence of Proposition 3.10 with the conditions from (b).

Corollary 3.11. *Fix a reflexive Banach space X, a closed and convex set $U \subset X$, $\varphi\colon X \to \mathbb{R}_\infty$ which is weakly sequentially lower semicontinuous on U, $\Phi\colon X \rightrightarrows Y$ where Y is another Banach space, and some $\bar{y} \in Y$ such that $U \cap \Phi^{-1}(\bar{y}) \cap \operatorname{dom}\varphi \neq \varnothing$. Then φ is lower semicontinuous on U relative to Φ at $\bar{y}$ provided that one of the following conditions is satisfied:*

(a) $\operatorname{gph}\Phi \cap (U \times Y)$ *is weakly sequentially closed;*

(b) X *is finite-dimensional and* $\operatorname{gph}\Phi \cap (U \times Y)$ *is closed.*

Particularly, whenever $\bar{x} \in \Phi^{-1}(\bar{y}) \cap \operatorname{dom}\varphi$ is fixed, φ is weakly sequentially lower semicontinuous, and either $\operatorname{gph}\Phi$ is weakly sequentially closed or $\operatorname{gph}\Phi$ is closed while X is finite-dimensional, then φ is lower semicontinuous near $\bar{x}$ relative to Φ at $\bar{y}$.

In the upcoming subsections, we discuss sufficient conditions for the semicontinuity properties of a set-valued mapping and an extended-real-valued function appearing in the conditions (a) of Proposition 3.10.

3.3 Sufficient conditions for lower semicontinuity of a function relative to a set

In the statement below, we present some simple situations where a function is lower semicontinuous relative to a set in the sense of Definition 3.3 (a).

Proposition 3.12. *Let X be a metric space, $\varphi\colon X \to \mathbb{R}_\infty$, and $\Omega, U \subset X$ be sets with $\Omega \cap U \cap \operatorname{dom}\varphi \neq \varnothing$. Then φ is lower semicontinuous on U relative to Ω provided that one of the following conditions is satisfied:*

(a) $U \subset \Omega$;

(b) $\Omega \cap U = \{\bar{x}\}$, *and φ is lower semicontinuous at $\bar{x}$;*

(c) $\bar{x} \in \Omega \cap U$ *is a minimizer of φ on U;*

(d) φ *is uniformly continuous on U.*

Proof. Under each of the condition (a), (b), and (c), the conclusion is straightforward since inequality (3.2) is an immediate consequence of the following simple relations, respectively, holding with any $U' \subset U$:

(a) $\inf\limits_{u\in\Omega\cap U} \varphi(u) = \inf\limits_{u\in U} \varphi(u), \quad \liminf\limits_{x\in U',\, \mathrm{dist}_\Omega(x)\to 0} \varphi(x) = \inf\limits_{x\in U'} \varphi(x) \geq \inf\limits_{x\in U} \varphi(x);$

(b) $\inf\limits_{u\in\Omega\cap U} \varphi(u) = \varphi(\bar{x}), \quad \liminf\limits_{x\in U',\, \mathrm{dist}_\Omega(x)\to 0} \varphi(x) = \liminf\limits_{x\to_{U'}\bar{x}} \varphi(x) \geq \liminf\limits_{x\to\bar{x}} \varphi(x) \geq \varphi(\bar{x});$

(c) $\inf\limits_{u\in\Omega\cap U} \varphi(u) = \varphi(\bar{x}), \quad \liminf\limits_{x\in U',\, \mathrm{dist}_\Omega(x)\to 0} \varphi(x) \geq \varphi(\bar{x}).$

It remains to prove the claim under condition (d). Let a number $\varepsilon > 0$ be arbitrarily chosen. Let a subset $U' \subset X$ and a number $\rho > 0$ be such that $U' + \rho\mathbb{B} \subset U$. By (d), there is a $\delta > 0$ such that

$$\forall x, x' \in U\colon \quad d(x, x') < \delta \quad \Longrightarrow \quad |\varphi(x) - \varphi(x')| < \varepsilon.$$

Let a point $x \in U'$ satisfy $\mathrm{dist}_\Omega(x) < \delta' := \min(\rho, \delta)$. Then there is a point $x' \in \Omega$ satisfying $d(x, x') < \delta'$. Hence, $x, x' \in U$, $d(x, x') < \delta$, and, consequently, $|\varphi(x) - \varphi(x')| < \varepsilon$. Thus, we have $\inf_{u\in\Omega\cap U} \varphi(u) \leq \varphi(x') < \varphi(x) + \varepsilon$, and, consequently,

$$\inf_{u\in\Omega\cap U} \varphi(u) \leq \liminf_{x\in U',\, \mathrm{dist}_\Omega(x)\to 0} \varphi(x) + \varepsilon.$$

Taking the infimum on the right-hand side of the last inequality over ε and U', we arrive at (3.2). $\square$

As a corollary, we obtain sufficient conditions for the lower semicontinuity property from Definition 3.3 (b).

Corollary 3.13. *Let X be a metric space, $\varphi\colon X \to \mathbb{R}_\infty$, $\Omega \subset X$, and $\bar{x} \in \Omega \cap \mathrm{dom}\,\varphi$. Then φ is lower semicontinuous near $\bar{x}$ relative to Ω provided that one of the following conditions is satisfied:*

(a) $\bar{x} \in \mathrm{int}\,\Omega$;

(b) $\bar{x}$ is an isolated point of Ω, and φ is lower semicontinuous at $\bar{x}$;

(c) $\bar{x}$ is an (unconditional) local minimizer of φ;

(d) φ is uniformly continuous near $\bar{x}$.

It follows from Corollary 3.13 (d) that each locally Lipschitz function is lower semicontinuous near a reference point relative to any set containing this point.

The subsequent result can be directly distilled from Corollary 3.11.

Proposition 3.14. *Fix a reflexive Banach space X, a closed and convex set $U \subset X$, and $\varphi\colon X \to \mathbb{R}_\infty$ which is weakly sequentially lower semicontinuous on U. Let $\Omega \subset X$ be chosen such that $\Omega \cap U \cap \mathrm{dom}\,\varphi \neq \varnothing$ while $\Omega \cap U$ is weakly sequentially closed. Then φ is lower semicontionuous on U relative to Ω.*

As a corollary, we obtain the subsequent result.

Corollary 3.15. *Fix a reflexive Banach space X, $\varphi\colon X \to \mathbb{R}_\infty$ which is weakly sequentially lower semicontinuous, and a weakly sequentially closed set $\Omega \subset X$. Then, for each $\bar{x} \in \Omega \cap \operatorname{dom}\varphi$, φ is lower semicontinuous near $\bar{x}$ relative to Ω.*

Note that whenever X is finite-dimensional, $\varphi\colon X \to \mathbb{R}_\infty$ is lower semicontinuous, and $\Omega \subset X$ is closed, then the assumptions of Corollary 3.15 hold trivially.

The following statement shows that lower semicontinuity relative to a set is preserved under decoupled summation.

Proposition 3.16. *Fix $n \in \mathbb{N}$ with $n \geq 2$. For each $i \in \{1, \ldots, n\}$, let X_i be a metric space, $\varphi_i\colon X_i \to \mathbb{R}_\infty$, $\Omega_i, U_i \subset X_i$, and $\Omega_i \cap U_i \cap \operatorname{dom}\varphi_i \neq \varnothing$. Suppose that φ_i is lower semicontinuous on U_i relative to Ω_i. Then $\varphi\colon X_1 \times \ldots \times X_n \to \mathbb{R}_\infty$ given by*

$$\forall (x_1, \ldots, x_n) \in X_1 \times \ldots \times X_n\colon \quad \varphi(x_1, \ldots, x_n) := \varphi_1(x_1) + \ldots + \varphi_n(x_n)$$

is lower semicontinuous on $U := U_1 \times \ldots \times U_n$ relative to $\Omega := \Omega_1 \times \ldots \times \Omega_n$.

Proof. The assertion is a direct consequence of Definition 3.3 (a). More precisely, we find

$$\begin{aligned}\inf_{u \in \Omega \cap U} \varphi(u) = \sum_{i=1}^{n} \inf_{u_i \in \Omega_i \cap U_i} \varphi_i(u_i) &\leq \sum_{i=1}^{n} \inf_{\substack{U_i' + \rho_i \mathbb{B} \subset U_i,\\ \rho_i > 0}} \liminf_{\substack{x_i \in U_i',\\ \operatorname{dist}_{\Omega_i}(x_i) \to 0}} \varphi_i(x_i) \\ &= \inf_{\substack{U' + \rho \mathbb{B} \subset U,\\ \rho > 0}} \liminf_{\substack{x \in U',\\ \operatorname{dist}_{\Omega}(x) \to 0}} \varphi(x),\end{aligned}$$

and this proves the claim. $\square$

3.4 A sufficient condition for upper semicontinuity of the inverse of a set-valued mapping

The next statement presents a condition ensuring validity of the upper semicontinuity assumption which appears in Proposition 3.10 (a).

Proposition 3.17. *Let X and Y be metric spaces, $\Phi\colon X \rightrightarrows Y$, and $(\bar{x}, \bar{y}) \in \operatorname{gph}\Phi$. Assume that Φ is metrically subregular at $(\bar{x}, \bar{y})$, i.e., that there exist a neighborhood U of $\bar{x}$ and a constant $L > 0$ such that*

$$\forall x \in U\colon \quad \operatorname{dist}_{\Phi^{-1}(\bar{y})}(x) \leq L \operatorname{dist}_{\Phi(x)}(\bar{y}). \tag{3.4}$$

Then, for each set $U' \subset U$ satisfying $\bar{x} \in U'$, the mapping $S_{U'}\colon Y \rightrightarrows X$, given by $S_{U'}(y) := \Phi^{-1}(y) \cap U'$ for each $y \in Y$, is upper semicontinuous at $\bar{y}$.

Proof. Let a number $\varepsilon > 0$ as well as $U' \subset U$ with $\bar{x} \in U'$ be given. Choose a number $\delta \in (0, \varepsilon/L)$. Then, for each $y \in B_\delta(\bar{y})$ and each $x \in S_{U'}(y)$, condition (3.4) yields $\operatorname{dist}_{S_{U'}(\bar{y})}(x) = \operatorname{dist}_{\Phi^{-1}(\bar{y})}(x) \leq L d(y, \bar{y}) < L\delta < \varepsilon$. By Definition 3.8, $S_{U'}$ is upper semicontinuous at $\bar{y}$. $\square$

We note that the metric subregularity condition (3.4) from Proposition 3.17 already amounts to a qualification condition addressing sets of type $\{x \in X \,|\, \bar{y} \in \Phi(x)\}$, see [Gfrerer, 2013, Section 5]. Sufficient conditions for metric subregularity can be found e.g. in Bai et al. [2019], Dontchev and Rockafellar [2014], Dontchev et al. [2020], Ioffe [2017], Kruger [2015], Maréchal [2018], Zheng and Ng [2010].

We would like to point the reader's attention to the fact that metric subregularity of Φ is a quantitative continuity property coming along with a *modulus* of subregularity $L > 0$ while upper semicontinuity of the mappings $S_{U'}$ in Proposition 3.17 is just a qualitative continuity property. In this regard, there exist weaker sufficient conditions ensuring validity of the upper semicontinuity requirements from Proposition 3.10 (a). However, it is not clear if such conditions can be easily checked in terms of initial problem data while this is clearly possible for metric subregularity as the aforementioned list of references underlines. Finally, we would like to mention that in case where one wants to avoid fixing the component $\bar{x} \in X$ in the preimage space in Proposition 3.17, it is possible to demand that Φ^{-1} is Lipschitz upper semicontinuous at $\bar{y}$ in the sense of [Klatte and Kummer, 2002, p. 10]. Again, this is a quantitative continuity property.

Example 3.18. *Let $G\colon X \to Y$ be a single-valued mapping between Banach spaces. Furthermore, let $C \subset X$ and $K \subset Y$ be nonempty, closed sets. We investigate the feasibility mapping $\Phi\colon X \rightrightarrows Y \times X$ given by $\Phi(x) := (G(x) - K, x - C)$ for all $x \in X$ as well as some point $\bar{x} \in X$ such that $(\bar{x}, (0,0)) \in \operatorname{gph} \Phi$ and some neighborhood U of $\bar{x}$. Let us define $S\colon Y \times X \rightrightarrows X$ by means of $S(y,z) := \Phi^{-1}(y,z) \cap U$ for each pair $(y,z) \in Y \times X$. One can check that S is upper semicontinuous at $(0,0)$ if and only if*

$$\operatorname{dist}_{K\times C}((G(x_k), x_k)) \to 0 \quad \Longrightarrow \quad \lim_{k\to+\infty} \operatorname{dist}_{G^{-1}(K)\cap C}(x_k) = 0$$

for each sequence $\{x_k\}_{k\in\mathbb{N}} \subset U$, and this is trivially satisfied if G is continuous and X is finite-dimensional. For the purpose of completeness, let us also mention that S is partially weakly sequentially upper semicontinuous at $(0,0)$ if and only if

$$x_k \rightharpoonup x, \quad \operatorname{dist}_{K\times C}((G(x_k), x_k)) \to 0 \quad \Longrightarrow \quad x \in G^{-1}(K) \cap C \tag{3.5}$$

is valid for each sequence $\{x_k\}_{k\in\mathbb{N}} \subset U$ and each point $x \in U$. Again, this is inherent if G is continuous while X is finite-dimensional and U is closed.

In infinite-dimensional situations, whenever G is continuously Fréchet differentiable and C as well as K are convex, Robinson's constraint qualification, given by

$$G'(\bar{x}) \left[\bigcup\nolimits_{\alpha\in[0,+\infty)} \alpha(C - \bar{x}) \right] - \bigcup\nolimits_{\alpha\in[0,+\infty)} \alpha(K - G(\bar{x})) = Y,$$

is equivalent to so-called metric regularity of Φ at $(\bar{x}, (0,0))$, see [Bonnans and Shapiro, 2000, Proposition 2.89], and the latter is sufficient for metric subregularity of Φ at $(\bar{x}, (0,0))$.

The final corollary of this section now follows from Propositions 3.10 and 3.17 and Corollary 3.13.

Corollary 3.19. *Fix metric spaces X and Y, $\Phi\colon X \rightrightarrows Y$, $\varphi\colon X \to \mathbb{R}_\infty$, $\bar y \in Y$, and $\bar x \in \Phi^{-1}(\bar y) \cap \operatorname{dom}\varphi$. Assume that Φ is metrically subregular at $(\bar x, \bar y)$ and that φ satisfies one of the conditions (a)-(d) of Corollary 3.13. Then φ is lower semicontinuous near $\bar x$ relative to Φ at $\bar y$.*

4 Optimality conditions and approximate stationarity

We consider the optimization problem

$$\min\{\varphi(x) \,|\, \bar y \in \Phi(x)\}, \tag{P}$$

where $\varphi\colon X \to \mathbb{R}_\infty$ is an arbitrary function, $\Phi\colon X \rightrightarrows Y$ is a set-valued mapping between Banach spaces, and $\bar y \in \operatorname{Im}\Phi$. Let us mention that the model (P) is quite general and covers numerous important classes of optimization problems, see e.g. Gfrerer [2013], Mehlitz [2020] for a discussion. The constrained problem (P) is obviously equivalent to the unconditional minimization of the restriction $\varphi_{\Phi^{-1}(\bar y)}$ of φ to $\Phi^{-1}(\bar y)$. We say that $\bar x$ is an ε-minimal point of problem (P) on U if it is an ε-minimal point of $\varphi_{\Phi^{-1}(\bar y)}$ on U. Analogously, stationary points of (P) are defined.

The next theorem presents dual (i.e., subdifferential/coderivative based) necessary conditions for ε-minimal points of problem (P).

Theorem 4.1. *Let X and Y be Banach spaces, $\varphi\colon X \to \mathbb{R}_\infty$ be lower semicontinuous, $\Phi\colon X \rightrightarrows Y$ have closed graph, and fix $\bar y \in Y$, $\bar x \in \operatorname{dom}\varphi \cap \Phi^{-1}(\bar y)$, $U \subset X$, $\varepsilon > 0$, as well as $\delta > 0$. Assume that $B_\delta(\bar x) \subset U$, and*

(a) on U, φ is bounded from below and lower semicontinuous relative to Φ at $\bar y$;

(b) either X and Y are Asplund, or φ and $\operatorname{gph}\Phi$ are convex.

Suppose that $\bar x$ is an ε-minimal point of problem (P) on U. Then, for each $\eta > 0$, there exist points $x_1, x_2 \in B_\delta(\bar x)$ and $y_2 \in \Phi(x_2) \cap B_\eta(\bar y)$ such that $\|x_2 - x_1\| < \eta$, $\operatorname{dist}_{\operatorname{gph}\Phi}(x_1, \bar y) < \eta$, $\varphi(x_1) < \varphi(\bar x) + \eta$, and

$$0 \in \partial\varphi(x_1) + \operatorname{Im} D^*\Phi(x_2, y_2) + \frac{2\varepsilon}{\delta}\mathbb{B}^*.$$

Moreover, if φ and $\operatorname{gph}\Phi$ are convex, then $\varphi(x_1) \le \varphi(\bar x)$.

Proof. Since φ is bounded from below on U, and $\bar x$ is an ε-minimal of problem (P) on U, there exist numbers $c > 0$ and $\varepsilon' \in (0, \varepsilon)$ such that

$$\forall x \in U\colon \qquad \varphi(x) > \varphi(\bar x) - c, \tag{4.1a}$$
$$\forall x \in \Phi^{-1}(\bar y) \cap U\colon \quad \varphi(x) > \varphi(\bar x) - \varepsilon'. \tag{4.1b}$$

For $\gamma > 0$ and $\gamma_1 > 0$, let the functions $\phi_\gamma, \hat\phi_{\gamma,\gamma_1}\colon X \times Y \to \mathbb{R}_\infty$ be given by

$$\forall (x, y) \in X \times Y\colon \qquad \phi_\gamma(x, y) := \varphi(x) + \gamma\big(\|y - \bar y\| + \operatorname{dist}_{\operatorname{gph}\Phi}(x, y)\big), \tag{4.2a}$$

$$\hat{\phi}_{\gamma,\gamma_1}(x,y) := \phi_\gamma(x,y) + \gamma_1 \|x-\bar{x}\|^2 . \tag{4.2b}$$

Set $\delta_0 := \delta\varepsilon'/\varepsilon$, and choose numbers $\delta' \in (\delta_0,\delta)$ and $\xi \in (0,\delta-\delta')$ such that $\xi(\delta'+2) < 2(\varepsilon\delta'/\delta - \varepsilon')$. Fix an arbitrary $\eta > 0$ and a positive number $\eta' < \min(\eta, 2(\delta-\delta'))$. Set $\gamma_1 := (\varepsilon'+\xi)/(\delta')^2$. Observe that $\hat{\phi}_{\gamma,\gamma_1}(\bar{x},\bar{y}) = \varphi(\bar{x})$, and $\hat{\phi}_{\gamma,\gamma_1}$ is bounded from below on $\overline{B}_{\delta'}(\bar{x}) \times Y$ due to (4.1a). By Ekeland's variational principle, see Lemma 2.1, for each $k \in \mathbb{N}$, there exists a point $(x_k,y_k) \in \overline{B}_{\delta'}(\bar{x}) \times Y$ such that

$$\hat{\phi}_{k,\gamma_1}(x_k,y_k) \le \varphi(\bar{x}), \tag{4.3a}$$

$$\forall (x,y) \in \overline{B}_{\delta'}(\bar{x}) \times Y\colon \quad \hat{\phi}_{k,\gamma_1}(x,y) + \xi\, \|(x,y)-(x_k,y_k)\| \ge \hat{\phi}_{k,\gamma_1}(x_k,y_k). \tag{4.3b}$$

It follows from (4.1a), (4.2), and (4.3a) that

$$k\big(\|y_k - \bar{y}\| + \operatorname{dist}_{\operatorname{gph}\Phi}(x_k,y_k)\big) + \gamma_1 \|x_k - \bar{x}\|^2 \le \varphi(\bar{x}) - \varphi(x_k) < c,$$

and, consequently,

$$\|y_k - \bar{y}\| + \operatorname{dist}_{\operatorname{gph}\Phi}(x_k,y_k) < c/k, \tag{4.4a}$$

$$\gamma_1 \|x_k - \bar{x}\|^2 \le \varphi(\bar{x}) - \varphi(x_k) \tag{4.4b}$$

are valid for all $k \in \mathbb{N}$. By (4.4a), $y_k \to \bar{y}$ and $\operatorname{dist}_{\operatorname{gph}\Phi}(x_k,y_k) \to 0$ as $k \to +\infty$, and $y_k \in B_{\eta'/4}(\bar{y})$ as well as $\operatorname{dist}_{\operatorname{gph}\Phi}(x_k,y_k) < \eta'/4$ follow for all $k > 4c/\eta'$. Recall that $\{x_k\}_{k\in\mathbb{N}} + \rho\mathbb{B} \subset B_\delta(\bar{x}) \subset U$ for any positive $\rho < \delta - \delta'$. By Proposition 3.6, there exist an integer $\bar{k} > 4c/\eta'$ and a point $x' \in \Phi^{-1}(\bar{y}) \cap U$ such that $\varphi(x') < \varphi(x_{\bar{k}}) + \xi$. By (4.1b), we have $\varphi(\bar{x}) - \varphi(x') < \varepsilon'$. Set $\gamma := \bar{k}$, $\hat{x} := x_{\bar{k}}$, and $\hat{y} := y_{\bar{k}}$. Thus, $\hat{y} \in B_{\eta'/4}(\bar{y})$ and $\operatorname{dist}_{\operatorname{gph}\Phi}(\hat{x},\hat{y}) < \eta'/4$. By (4.4b),

$$\gamma_1 \|\hat{x} - \bar{x}\|^2 \le (\varphi(\bar{x}) - \varphi(x')) + (\varphi(x') - \varphi(\hat{x})) < \varepsilon' + \xi = \gamma_1(\delta')^2.$$

Hence, we find $\|\hat{x} - \bar{x}\| < \delta'$. In view of (4.2b), condition (4.3a) is equivalent to

$$\phi_\gamma(\hat{x},\hat{y}) + \gamma_1 \|\hat{x} - \bar{x}\|^2 \le \varphi(\bar{x}). \tag{4.5}$$

For each $(x,y) \in B_{\delta'}(\bar{x}) \times Y$ different from $(\hat{x},\hat{y})$, it follows from (4.2b) that

$$\begin{aligned}
\frac{\phi_\gamma(\hat{x},\hat{y}) - \phi_\gamma(x,y)}{\|(x,y)-(\hat{x},\hat{y})\|} &= \frac{\hat{\phi}_{\gamma,\gamma_1}(\hat{x},\hat{y}) - \hat{\phi}_{\gamma,\gamma_1}(x,y) + \gamma_1\big(\|x-\bar{x}\|^2 - \|\hat{x}-\bar{x}\|^2\big)}{\|(x,y)-(\hat{x},\hat{y})\|} \\
&\le \frac{\hat{\phi}_{\gamma,\gamma_1}(\hat{x},\hat{y}) - \hat{\phi}_{\gamma,\gamma_1}(x,y) + \gamma_1 \|x-\hat{x}\| \left(\|x-\bar{x}\| + \|\hat{x}-\bar{x}\|\right)}{\|(x,y)-(\hat{x},\hat{y})\|} \\
&\le \frac{\hat{\phi}_{\gamma,\gamma_1}(\hat{x},\hat{y}) - \hat{\phi}_{\gamma,\gamma_1}(x,y)}{\|(x,y)-(\hat{x},\hat{y})\|} + \gamma_1\big(\|x-\bar{x}\| + \|\hat{x}-\bar{x}\|\big),
\end{aligned}$$

and, consequently, in view of (4.3b),

$$\sup_{(x,y)\in(B_{\delta'}(\bar{x})\times Y)\setminus\{(\hat{x},\hat{y})\}} \frac{\phi_\gamma(\hat{x},\hat{y}) - \phi_\gamma(x,y)}{\|(x,y)-(\hat{x},\hat{y})\|} < \xi + 2\gamma_1\delta' = \frac{2\varepsilon' + \xi(\delta'+2)'}{\delta} < \frac{2\varepsilon}{\delta}.$$

Since $\hat{x}$ is an interior point of $\overline{B}_{\delta'}(\bar{x})$, it follows that

$$\limsup_{(x,y)\to(\hat{x},\hat{y})} \frac{\phi_\gamma(\hat{x},\hat{y}) - \phi_\gamma(x,y)}{\|(x,y)-(\hat{x},\hat{y})\|} < \frac{2\varepsilon}{\delta}. \tag{4.6}$$

By (4.2a) and (4.5), we find $\varphi(\hat{x}) \le \varphi(\bar{x})$, and due to (4.6), there is a number $\hat{\varepsilon} \in (0, \frac{2\varepsilon}{\delta})$ such that

$$\liminf_{(x,y)\to(\hat{x},\hat{y})} \frac{\phi_\gamma(x,y) + \hat{\varepsilon}\,\|(x,y)-(\hat{x},\hat{y})\| - \phi_\gamma(\hat{x},\hat{y})}{\|(x,y)-(\hat{x},\hat{y})\|} \ge 0.$$

Set $\xi' := 2\varepsilon/\delta - \hat{\varepsilon} > 0$. By definition of the Fréchet subdifferential, the above inequality yields

$$(0,0) \in \partial\,(\phi_\gamma + \hat{\varepsilon}\,\|(\cdot,\cdot)\|)\,(\hat{x},\hat{y}). \tag{4.7}$$

Condition (4.7) can be rewritten as $(0,0) \in \partial\,(\varphi + \gamma g + h)\,(\hat{x},\hat{y})$, where the functions $g, h \colon X \times Y \to \mathbb{R}$ are given by

$$\forall (x,y) \in X \times Y\colon \quad \begin{aligned} g(x,y) &:= \operatorname{dist}_{\operatorname{gph}\Phi}(x,y),\\ h(x,y) &:= \gamma\|y-\bar{y}\| + \hat{\varepsilon}\|(x,y)-(\hat{x},\hat{y})\|. \end{aligned}$$

Note that g and h are Lipschitz continuous, and h is convex. We distinguish two cases.

Case 1: Let X and Y be Asplund spaces. Let us recall the estimates $\|\hat{x}-\bar{x}\| < \delta' < \delta$, $\|\hat{y}-\bar{y}\| < \eta'/4 < \eta/4$, $\operatorname{dist}_{\operatorname{gph}\Phi}(\hat{x},\hat{y}) < \eta'/4 < \eta/4$, and $\varphi(\hat{x}) \le \varphi(\bar{x})$. By the fuzzy sum rule, see Lemma 2.2 (b), there exist points $(x_1,y_1), (u_2,v_2) \in X \times Y$ arbitrarily close to $(\hat{x},\hat{y})$ with $\varphi(x_1)$ arbitrarily close to $\varphi(\hat{x})$, so that

$$\|x_1-\bar{x}\| < \delta, \quad \|u_2-\bar{x}\| < \delta', \quad \varphi(x_1) < \varphi(\bar{x}) + \eta, \quad \|y_1-\bar{y}\| < \frac{\eta}{2},$$
$$\|v_2-\bar{y}\| < \frac{\eta'}{2}, \quad \|u_2-x_1\| < \frac{\eta'}{2}, \quad \operatorname{dist}_{\operatorname{gph}\Phi}(x_1,y_1) < \frac{\eta}{2}, \quad \operatorname{dist}_{\operatorname{gph}\Phi}(u_2,v_2) < \frac{\eta'}{4},$$

and subgradients $x_1^* \in \partial\varphi(x_1)$ and $(u_2^*, v_2^*) \in \partial g(u_2,v_2)$ satisfying

$$\|x_1^* + \gamma u_2^*\| < \hat{\varepsilon} + \frac{\xi'}{2}.$$

Thus, $x_1 \in B_\delta(\bar{x})$ and $\operatorname{dist}_{\operatorname{gph}\Phi}(x_1,\bar{y}) < \operatorname{dist}_{\operatorname{gph}\Phi}(x_1,y_1) + \|y_1-\bar{y}\| < \eta$. In view of Lemma 2.3 (b), there exist $(x_2,y_2) \in \operatorname{gph}\Phi$ and $(u_2^{*\prime}, v_2^{*\prime}) \in N_{\operatorname{gph}\Phi}(x_2,y_2)$ such that

$$\|(x_2,y_2)-(u_2,v_2)\| < \operatorname{dist}_{\operatorname{gph}\Phi}(u_2,v_2) + \frac{\eta'}{4} < \frac{\eta'}{2}, \quad \|u_2^{*\prime} - u_2^*\| < \frac{\xi'}{2\gamma}.$$

Set $x_2^* := \gamma u_2^{*\prime}$ and $y^* := -\gamma v_2^{*\prime}$. Thus, $x_2^* \in D^*\Phi(x_2,y_2)(y^*)$, and we have

$$\begin{aligned} \|y_2-\bar{y}\| &\le \|v_2-\bar{y}\| + \|y_2-v_2\| < \eta' < \eta,\\ \|x_2-\bar{x}\| &\le \|u_2-\bar{x}\| + \|x_2-u_2\| < \delta' + \frac{\eta'}{2} < \delta, \end{aligned}$$

$$
\begin{aligned}
&\|x_2 - x_1\| \le \|x_2 - u_2\| + \|u_2 - x_1\| < \eta' < \eta,\\
&\|x_1^* + x_2^*\| \le \|x_1^* + \gamma\, u_2^*\| + \gamma\, \|u_2^{*\prime} - u_2^*\| < \hat\varepsilon + \xi' = \frac{2\varepsilon}{\delta}.
\end{aligned}
$$

Case 2: Let φ and $\operatorname{gph}\Phi$ be convex. We have $\hat x \in B_{\delta'}(\bar x) \subset B_\delta(\bar x)$, $\varphi(\hat x) \le \varphi(\bar x)$, $\|\hat y - \bar y\| < \eta'/4$, and $\operatorname{dist}_{\operatorname{gph}\Phi}(\hat x, \hat y) < \eta'/4 < \eta$. By the convex sum rule, see Lemma 2.2 (a), there exist subgradients $x_1^* \in \partial\varphi(\hat x)$ and $(u_2^*, v_2^*) \in \partial g(\hat x, \hat y)$ satisfying

$$
\|x_1^* + \gamma u_2^*\| \le \hat\varepsilon.
$$

In view of Lemma 2.3 (b), there exist $(x_2, y_2) \in \operatorname{gph}\Phi$ and $(u_2^{*\prime}, v_2^{*\prime}) \in N_{\operatorname{gph}\Phi}(x_2, y_2)$ such that

$$
\|(x_2, y_2) - (\hat x, \hat y)\| < \operatorname{dist}_{\operatorname{gph}\Phi}(\hat x, \hat y) + \frac{\eta'}{4}, \quad \|u_2^{*\prime} - u_2^*\| < \frac{\xi'}{\gamma}.
$$

Set $x_1 := \hat x$, $x_2^* := \gamma u_2^{*\prime}$, and $y^* := -\gamma v_2^{*\prime}$. Thus, $x_1 \in B_\delta(\bar x)$, $\operatorname{dist}_{\operatorname{gph}\Phi}(x_1, \bar y) < \eta'/2 < \eta$, $\varphi(x_1) \le \varphi(\bar x)$, and $x_2^* \in D^*\Phi(x_2, y_2)(y^*)$. Replacing (u_2, v_2) with $(\hat x, \hat y)$ in the corresponding estimates established in Case 1, we obtain

$$
\begin{aligned}
&\|y_2 - \bar y\| < \eta, \quad \|x_2 - \bar x\| < \delta, \quad \|x_2 - x_1\| < \eta,\\
&\|x_1^* + x_2^*\| \le \|x_1^* + \gamma\, u_2^*\| + \gamma\, \|u_2^{*\prime} - u_2^*\| < \hat\varepsilon + \xi' = \frac{2\varepsilon}{\delta}.
\end{aligned}
$$

This completes the proof. □

Clearly, Theorem 4.1 provides dual necessary conditions for ε-minimality of a feasible point of problem (P) under some additional structural assumptions on the data which are almost for free in the finite-dimensional setting, see Corollary 3.11, and of reasonable strength in the infinite-dimensional one. In the subsequent remark, we comment on additional primal and dual conditions for ε-minimality which can be distilled from the proof of Theorem 4.1.

Remark 4.2. *(a) In the proof of Theorem 4.1, more sets of necessary conditions for local ε-minimality of a feasible point of problem (P) have been established along the way. Moreover, the first part of the proof does not use the linear structure of the spaces, i.e., the arguments work in the setting of general complete metric spaces X and Y. The conditions can be of independent interest and are listed below. We assume that X and Y are complete metric spaces and all the other assumptions of Theorem 4.1 are satisfied, except condition (b).*

Necessary conditions for local ε-minimality. *There is a $\delta_0 \in (0, \delta)$ such that, for each $\delta' \in (\delta_0, \delta)$ and $\eta > 0$, there exist points $\hat x \in B_{\delta'}(\bar x)$ and $\hat y \in B_\eta(\bar y)$ satisfying $\operatorname{dist}_{\operatorname{gph}\Phi}(\hat x, \hat y) < \eta$, and numbers $\gamma > 0$ and $\gamma_1 > 0$ such that, with the function $\phi_\gamma \colon X \times Y \to \mathbb{R}_\infty$ given by*

$$
\forall (x, y) \in X \times Y\colon \quad \phi_\gamma(x, y) := \varphi(x) + \gamma\big(d(y, \bar y) + \operatorname{dist}_{\operatorname{gph}\Phi}(x, y)\big),
$$

the following conditions hold:

- $\phi_\gamma(\hat{x},\hat{y})+\gamma_1 d(\hat{x},\bar{x})^2 \le \varphi(\bar{x})$, *and*
- *primal nonlocal condition* **(PNLC)**: $\displaystyle\sup_{\substack{(x,y)\neq(\hat{x},\hat{y})\\ x\in B_{\delta'}(\bar{x})}} \frac{\phi_\gamma(\hat{x},\hat{y})-\phi_\gamma(x,y)}{d((x,y),(\hat{x},\hat{y}))} < \frac{2\varepsilon}{\delta}$,
- *primal local condition* **(PLC)**: $\displaystyle\limsup_{(x,y)\to(\hat{x},\hat{y})} \frac{\phi_\gamma(\hat{x},\hat{y})-\phi_\gamma(x,y)}{d((x,y),(\hat{x},\hat{y}))} < \frac{2\varepsilon}{\delta}$,
- *dual condition* **(DC)** *(X and Y are Banach spaces): condition (4.7) is satisfied with some $\hat{\varepsilon}\in(0,\frac{2\varepsilon}{\delta})$.*

The relationship between the conditions is as follows: **(PNLC)** $\Rightarrow$ **(PLC)** $\Rightarrow$ **(DC)**. *The dual conditions in Theorem 4.1 are consequences of the above conditions.*

Let us note that the left-hand side in **(PNLC)** *is the* nonlocal slope, *see Fabian et al. [2010], of the function $\phi_\gamma+i_{B_{\delta'}(\bar{x})}$ at $(\hat{x},\hat{y})$, while the left-hand side in* **(PLC)** *is the conventional slope of ϕ_γ at $(\hat{x},\hat{y})$.*

(b) *Since the function φ in Theorem 4.1 is assumed to be lower semicontinuous, it is automatically bounded from below on* some *neighborhood of $\bar{x}$. We emphasize that Theorem 4.1 requires all the conditions to hold on the same set U containing a neighborhood of $\bar{x}$.*

As a consequence of Theorem 4.1, we obtain necessary conditions characterizing local minimizers of (P).

Corollary 4.3. *Let X and Y be Banach spaces, $\varphi\colon X\to\mathbb{R}_\infty$ lower semicontinuous, $\Phi\colon X\rightrightarrows Y$ have closed graph, $\bar{y}\in Y$, and $\bar{x}\in\operatorname{dom}\varphi\cap\Phi^{-1}(\bar{y})$. Assume that*

(a) *the function φ is lower semicontinuous near $\bar{x}$ relative to Φ at $\bar{y}$;*

(b) *either X and Y are Asplund, or φ and $\operatorname{gph}\Phi$ are convex.*

Suppose that $\bar{x}$ is a local minimizer of (P). Then, for each $\varepsilon>0$, there exist points $x_1,x_2\in B_\varepsilon(\bar{x})$ and $y_2\in\Phi(x_2)\cap B_\varepsilon(\bar{y})$ such that $|\varphi(x_1)-\varphi(\bar{x})|<\varepsilon$ and

$$0\in\partial\varphi(x_1)+\operatorname{Im}D^*\Phi(x_2,y_2)+\varepsilon\mathbb{B}^*.$$

Moreover, if φ and $\operatorname{gph}\Phi$ are convex, then $\varphi(x_1)\le\varphi(\bar{x})$.

Proof. Let a number $\varepsilon>0$ be arbitrarily chosen. Set $\varepsilon':=\varepsilon/2$. By the assumptions and Remark 2.5, there exists a $\delta\in(0,\varepsilon)$ such that on $U:=\overline{B}_\delta(\bar{x})$ the function φ is bounded from below and lower semicontinuous relative to Φ at $\bar{y}$, and $\bar{x}$ is an $\varepsilon'\delta$-minimal point of $\varphi_{\Phi^{-1}(\bar{y})}$ on U. Thus, all the assumptions of Theorem 4.1 are satisfied. Moreover, $2(\varepsilon'\delta)/\delta=\varepsilon$ and, since φ is lower semicontinuous, one can ensure that $\varphi(x_1)>\varphi(\bar{x})+\varepsilon$. Hence, taking any $\eta\in(0,\varepsilon)$, the assertion follows from Theorem 4.1. $\square$

In the subsequent remark, we comment on the findings in Corollary 4.3.

Remark 4.4. *(a) The analogues of the necessary conditions in Remark 4.2 (a) are valid in the setting of Corollary 4.3, too. More precisely, it suffices to replace $\frac{2\varepsilon}{\delta}$ with just ε in the involved conditions.*

(b) The necessary conditions in Corollary 4.3 hold for each stationary point (not necessarily a local minimizer) of problem (P).

We now consider an important particular case of problem (P), namely

$$\min\{\varphi(x)\,|\,x\in\Omega\}, \tag{$\widetilde{\mathrm{P}}$}$$

where $\Omega\subset X$ is a nonempty subset of a Banach space. To obtain this setting from the one in (P), it suffices to consider the set-valued mapping $\Phi\colon X\rightrightarrows Y$ whose graph is given by $\operatorname{gph}\Phi:=\Omega\times Y$. Here, Y can be an arbitrary Asplund space, e.g., one can take $Y:=\mathbb{R}$. Observe that $\Phi^{-1}(y)=\Omega$ holds for all $y\in Y$. Hence, by Definition 3.8, for all $y\in Y$, the mapping Φ^{-1} is upper semicontinuous at y. Moreover, $N_{\operatorname{gph}\Phi}(x,y)=N_\Omega(x)\times\{0\}$. Thus, the next statement is a consequence of Proposition 3.10 and Theorem 4.1.

Theorem 4.5. *Let X be a Banach space, $\varphi\colon X\to\mathbb{R}_\infty$ lower semicontinuous, $\Omega\subset X$ a closed set, and fix $\bar{x}\in\operatorname{dom}\varphi\cap\Omega$, $U\subset X$, $\varepsilon>0$, and $\delta>0$. Assume that $B_\delta(\bar{x})\subset U$, and*

(a) on U, the function φ is bounded from below and lower semicontinuous relative to Ω;

(b) either X is Asplund, or φ and Ω are convex.

Suppose that $\bar{x}$ is an ε-minimal point of problem ($\widetilde{\mathrm{P}}$) on U. Then, for each $\eta>0$, there exist points $x_1\in B_\delta(\bar{x})$ and $x_2\in\Omega\cap B_\delta(\bar{x})$ such that $\|x_2-x_1\|<\eta$, $\varphi(x_1)<\varphi(\bar{x})+\eta$, and

$$0\in\partial\varphi(x_1)+N_\Omega(x_2)+\frac{2\varepsilon}{\delta}\mathbb{B}^*.$$

Moreover, if φ and Ω are convex, then $\varphi(x_1)\le\varphi(\bar{x})$.

The next corollary follows immediately.

Corollary 4.6. *Let X be a Banach space, $\varphi\colon X\to\mathbb{R}_\infty$ lower semicontinuous, $\Omega\subset X$ a closed set, and $\bar{x}\in\operatorname{dom}\varphi\cap\Omega$. Assume that*

(a) the function φ is lower semicontinuous near $\bar{x}$ relative to Ω;

(b) either X is Asplund, or φ and Ω are convex.

Suppose that $\bar{x}$ is a local minimizer of ($\widetilde{\mathrm{P}}$). Then, for each $\varepsilon>0$, there exist $x_1\in B_\varepsilon(\bar{x})$ and $x_2\in\Omega\cap B_\varepsilon(\bar{x})$ such that $|\varphi(x_1)-\varphi(\bar{x})|<\varepsilon$ and

$$0\in\partial\varphi(x_1)+N_\Omega(x_2)+\varepsilon\mathbb{B}^*.$$

Moreover, if φ and Ω are convex, then $\varphi(x_1)\le\varphi(\bar{x})$.

Whenever φ is Lipschitz continuous around $\bar{x}$, the assertion of Corollary 4.6 is an immediate consequence of Fermat's rule and the sum rules stated in Lemma 2.2. We note that Corollary 4.6 is applicable in more general situations, exemplary, if φ is only uniformly continuous in a neighborhood of the investigated local minimizer, see Corollary 3.13, or if X is finite-dimensional, see Corollary 3.15.

Note that the dual necessary optimality conditions in Corollaries 4.3 and 4.6 do not hold at the reference point but at some other points arbitrarily close to it. Such conditions describe certain properties of admissible points which can be interpreted as a kind of dual approximate stationarity. The precise meaning of approximate stationarity will be discussed in Section 6.1 in the setting of geometrically-constrained optimization problems.

5 Generalized separation and extremal principle

Below, we discuss certain generalized extremality and separation properties of a collection of closed subsets $\Omega_1, \ldots, \Omega_n$ of a Banach space X, having a common point $\bar{x} \in \bigcap_{i=1}^n \Omega_i$. Here, n is an integer satisfying $n \geq 2$. We write $\{\Omega_1, \ldots, \Omega_n\}$ to denote the collection of sets as a single object.

We begin with deriving necessary conditions for so-called $\mathcal{F}$-extremality of a collection of sets. The property in the definition below is determined by a nonempty family $\mathcal{F}$ of nonnegative lower semicontinuous functions $f\colon X^n \to \mathbb{R}_\infty$ and mimics the corresponding conventional one, see e.g. Kruger and Mordukhovich [1980].

Definition 5.1. *Let a family $\mathcal{F}$ described above be given. Suppose that, for each $f \in \mathcal{F}$, the function $\hat{f}\colon X^n \to \mathbb{R}_\infty$ is defined by*

$$\forall z := (x_1, \ldots, x_n) \in X^n\colon \quad \hat{f}(z) := f(x_1 - x_n, \ldots, x_{n-1} - x_n, x_n). \tag{5.1}$$

*The collection $\{\Omega_1, \ldots, \Omega_n\}$ is $\mathcal{F}$-*extremal at $\bar{x}$ *if, for each $\varepsilon > 0$, there exist a function $f \in \mathcal{F}$ and a number $\rho > 0$ such that $f(0, \ldots, 0, \bar{x}) < \varepsilon$ and*

$$\forall x_i \in \Omega_i + \rho\mathbb{B}\ (i = 1, \ldots, n)\colon \quad \hat{f}(x_1, \ldots, x_n) > 0. \tag{5.2}$$

The following theorem, which is based on Theorem 4.5, provides a general necessary condition for $\mathcal{F}$-extremality.

Theorem 5.2. *Assume that*

(a) there is a neighborhood U of $\bar{x}$ such that, for each $f \in \mathcal{F}$, the function $\hat{f}\colon X^n \to \mathbb{R}_\infty$ defined by (5.1) is lower semicontinuous on U^n relative to $\Omega := \Omega_1 \times \ldots \times \Omega_n$;

(b) either X is Asplund, or $\Omega_1, \ldots, \Omega_n$ and all $f \in \mathcal{F}$ are convex.

Suppose that the collection $\{\Omega_1, \ldots, \Omega_n\}$ is $\mathcal{F}$-extremal at $\bar{x}$. Then, for each $\varepsilon > 0$ and $\eta > 0$, there exist a function $f \in \mathcal{F}$ with $f(0, \ldots, 0, \bar{x}) < \varepsilon$ and points $x_i \in \Omega_i \cap B_\varepsilon(\bar{x})$, $x_i' \in B_\eta(x_i)$, and $x_i^* \in X^*$ $(i = 1, \ldots, n)$ such that

$$\sum_{i=1}^n \operatorname{dist}_{N_{\Omega_i}(x_i)}(x_i^*) < \varepsilon, \tag{5.3a}$$

$$0 < f(w) < f(0,\ldots,0,\bar{x}) + \eta, \tag{5.3b}$$

$$-\left(x_1^*,\ldots,x_{n-1}^*,\sum_{i=1}^n x_i^*\right) \in \partial f(w), \tag{5.3c}$$

where $w := (x_1'-x_n',\ldots,x_{n-1}'-x_n',x_n') \in X^n$. Moreover, if f and $\Omega_1,\ldots,\Omega_n$ are convex, then $f(w) \le f(0,\ldots,0,\bar{x})$.

Proof. Let arbitrary numbers $\varepsilon > 0$ and $\eta > 0$ be fixed. Choose a number $\delta \in (0,\varepsilon)$ so that $B_\delta(\bar{x}) \subset U$, and set $\varepsilon' := \varepsilon \min(\delta/2, 1)$. By Definition 5.1, there exist a function $f \in \mathcal{F}$ and a number $\rho > 0$ such that $f(0,\ldots,0,\bar{x}) < \varepsilon' \le \varepsilon$, and condition (5.2) holds, where the function $\hat{f}\colon X^n \to \mathbb{R}_\infty$ is defined by (5.1). Observe that Ω is a closed subset of the Banach space X^n, $\bar{z} := (\bar{x},\ldots,\bar{x}) \in \Omega$, and $\hat{f}(\bar{z}) = f(0,\ldots,0,\bar{x}) < \varepsilon'$. Since the function f is nonnegative, so is $\hat{f}$, and, consequently, $\bar{z}$ is an ε'-minimal point of $\hat{f}_\Omega$ (as well as $\hat{f}$) on X^n. Set $\eta' := \min(\eta,\rho)$. By Theorem 4.5, there exist points $z := (x_1,\ldots,x_n) \in \Omega \cap B_\delta(\bar{z})$, $z' := (x_1',\ldots,x_n') \in B_{\eta'}(z)$, and $x^* := (x_1^*,\ldots,x_n^*) \in (X^*)^n$ such that $f(w) = \hat{f}(z') < \hat{f}(\bar{z}) + \eta = f(0,\ldots,0,\bar{x}) + \eta$, and

$$-x^* \in \partial\hat{f}(z'), \qquad \operatorname{dist}_{N_\Omega(z)}(x^*) < \frac{2\varepsilon'}{\delta} \le \varepsilon. \tag{5.4}$$

Moreover, if f and $\Omega_1,\ldots,\Omega_n$ are convex, then $f(w) \le f(0,\ldots,0,\bar{x})$. Observe that $x_i' \in \Omega_i + \rho\mathbb{B}$ $(i = 1,\ldots,n)$, and it follows from (5.2) that $f(w) = \hat{f}(z') > 0$ which shows (5.3b).

The function $\hat{f}$ given by (5.1) is a composition of f and the continuous linear mapping $A\colon X^n \to X^n$ given as follows:

$$\forall (u_1,\ldots,u_n) \in X^n\colon \quad A(u_1,\ldots,u_n) := (u_1 - u_n,\ldots,u_{n-1} - u_n, u_n).$$

The mapping A is obviously a bijection. It is easy to check that the adjoint mapping $A^*\colon (X^*)^n \to (X^*)^n$ is of the form

$$\forall (u_1^*,\ldots,u_n^*) \in (X^*)^n\colon \quad A^*(u_1^*,\ldots,u_n^*) := \left(u_1^*,\ldots,u_{n-1}^*, u_n^* - \sum_{i=1}^{n-1} u_i^*\right). \tag{5.5}$$

By the Fréchet subdifferential chain rule, which can be distilled from [Mordukhovich, 2006, Theorem 1.66, Proposition 1.84]), we obtain $\partial\hat{f}(z') = A^*\partial f(w)$, where $w = Az' = (x_1'-x_n',\ldots,x_{n-1}'-x_n',x_n')$. In view of (5.5), the inclusion in (5.4) is equivalent to (5.3c). It now suffices to observe that $N_\Omega(z) = N_{\Omega_1}(x_1) \times \ldots \times N_{\Omega_n}(x_n)$, and, consequently, the inequality in (5.4) yields (5.3a). $\square$

For the conclusions of Theorem 5.2 to be non-trivial, one must ensure that the family $\mathcal{F}$ satisfies the following conditions:

(a) $\inf_{f\in\mathcal{F}} f(0,\ldots,0,\bar{x}) = 0$;

(b) $\liminf\limits_{w\to(0,\ldots,0,\bar{x}),\, f\in\mathcal{F},\, f(w)\downarrow 0,\, w^*\in\partial f(w)} \|w^*\| > 0.$

A typical example of such a family is given by the collection $\mathcal{F}_A$ of functions of type

$$\forall z := (x_1, \ldots, x_n) \in X^n: \quad f_a(z) := \max_{1 \le i \le n-1} \|x_i - a_i\|, \tag{5.6}$$

where $a := (a_1, \ldots, a_{n-1}) \in X^{n-1}$. The proofs of the conventional extremal principle and its extensions usually employ such functions. Note that functions from $\mathcal{F}_A$ are constant in the last variable.

It is easy to see that, for each $f_a \in \mathcal{F}_A$ and $z := (x_1, \ldots, x_n) \in X^n$, the value $f_a(z)$ is the maximum norm of $(x_1 - a_1, \ldots, x_{n-1} - a_{n-1})$ in X^{n-1}. Thus, $f_a(z) > 0$ if and only if $(x_1, \ldots, x_{n-1}) \neq a$, and

$$f_a(0, \ldots, 0, \bar{x}) = \max_{1 \le i \le n-1} \|a_i\| \to 0 \quad \text{as} \quad a \to 0$$

showing (a). Moreover, $\partial f_a(z) \neq \varnothing$ for all $z \in X^n$ and, if $f_a(z) > 0$, then $\|w^*\| = 1$ for all $w^* \in \partial f_a(w)$, i.e., the limit in (b) equals 1. Observe also that, since each function $f_a \in \mathcal{F}_A$ is convex and Lipschitz continuous, the same holds true for the corresponding function $\hat{f}_a$ defined by (5.1). Hence, $\hat{f}_a$ is automatically lower semicontinuous near each point of X^n relative to each set containing this point, see Corollary 3.13.

When $f_a \in \mathcal{F}_A$ is given by (5.6), condition (5.2) takes the following form:

$$\bigcap_{i=1}^{n-1} (\Omega_i + \rho\mathbb{B} - a_i) \cap (\Omega_n + \rho\mathbb{B}) = \varnothing. \tag{5.7}$$

With this in mind, the extremality property in Definition 5.1 admits a geometric interpretation.

Proposition 5.3. *The collection $\{\Omega_1, \ldots, \Omega_n\}$ is $\mathcal{F}_A$-extremal at $\bar{x}$ if and only if, for each $\varepsilon > 0$, there exist vectors $a_1, \ldots, a_{n-1} \in X$ and a number $\rho > 0$ such that $\max_{1 \le i \le n-1} \|a_i\| < \varepsilon$, and condition (5.7) holds.*

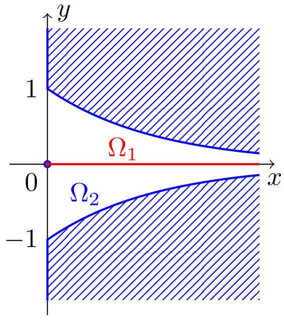
The characterization in Proposition 5.3 means that sets with nonempty intersection can be “pushed apart” by arbitrarily small translations in such a way that even small enlargements of the sets become nonintersecting. Note that condition (5.7) is stronger than the conventional extremality property originating from Kruger and Mordukhovich [1980], which corresponds to setting $\rho = 0$ in (5.7). The converse statement is not true as the next example shows.

Example 5.4. *Consider the closed sets $\Omega_1, \Omega_2 \subset \mathbb{R}^2$ given by*

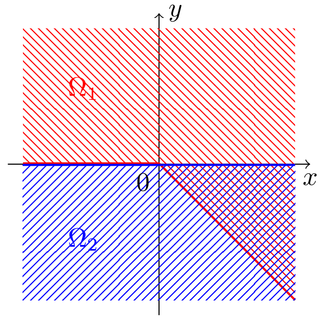
$$\Omega_1 := \{(x, y) \mid x \ge 0,\ y = 0\}, \quad \Omega_2 := \left\{(x, y) \mid x \ge 0,\ |y| \ge e^{-x}\right\} \cup \{(0, 0)\},$$

see Figure 5.1(a). We have $\Omega_1 \cap \Omega_2 = \{(0, 0)\}$ and $(\Omega_1 - (t, 0)) \cap \Omega_2 = \varnothing$ for each $t < 0$. At the same time, $(\Omega_1 + \rho\mathbb{B} - a) \cap (\Omega_2 + \rho\mathbb{B}) \neq \varnothing$ for all $a \in \mathbb{R}^2$ and $\rho > 0$. By Proposition 5.3, $\{\Omega_1, \Omega_2\}$ is not $\mathcal{F}_A$-extremal at $(0, 0)$.

Theorem 5.2 produces the following necessary condition for $\mathcal{F}_A$-extremality.



(a) Sets from Example 5.4.



(b) Sets from Example 5.6.

Figure 5.1: Visualization of the sets Ω_1 and Ω_2 from Examples 5.4 and 5.6.

Corollary 5.5. *Assume that either X is Asplund, or $\Omega_1, \ldots, \Omega_n$ are convex. Suppose that the collection $\{\Omega_1, \ldots, \Omega_n\}$ is $\mathcal{F}_A$-extremal at $\bar{x}$. Then, for each $\varepsilon > 0$, there exist points $x_i \in \Omega_i \cap B_\varepsilon(\bar{x})$ and $x_i^* \in X^*$ $(i = 1, \ldots, n)$ satisfying (5.3a) and*

$$\sum_{i=1}^{n} x_i^* = 0, \tag{5.8a}$$

$$\sum_{i=1}^{n-1} \|x_i^*\| = 1. \tag{5.8b}$$

Moreover, for each $\tau \in (0, 1)$, the points x_i and x_i^* $(i = 1, \ldots, n)$ can be chosen so that

$$\sum_{i=1}^{n-1} \langle x_i^*, x_n - x_i + a_i \rangle > \tau \max_{1 \le i \le n-1} \|x_n - x_i + a_i\|, \tag{5.9}$$

where $a_1, \ldots, a_{n-1}$ are vectors satisfying the characterization in Proposition 5.3.

Proof. Fix $\varepsilon > 0$ arbitrarily. Recall that, for each $f_a \in \mathcal{F}_A$, the function $\hat{f}_a \colon X^n \to \mathbb{R}_\infty$ defined according to (5.1) is lower semicontinuous near $\bar{z} := (\bar{x}, \ldots, \bar{x})$ relative to $\Omega := \Omega_1 \times \ldots \times \Omega_n$. By definition of $\mathcal{F}_A$, Proposition 5.3, and Theorem 5.2, for each $\eta > 0$, there exist vectors $a_1, \ldots, a_{n-1} \in X$, points $x_i \in \Omega_i \cap B_\varepsilon(\bar{x})$, $x_i' \in B_\eta(x_i)$, and $x_i^* \in X^*$ $(i = 1, \ldots, n)$, and a number $\rho > 0$ such that $\max_{1 \le i \le n-1} \|a_i\| < \varepsilon$, and conditions (5.3) and (5.7) hold, where $w := (x_1' - x_n', \ldots, x_{n-1}' - x_n', x_n')$ and the function f is replaced by f_a defined by (5.6). Clearly, we find

$$\partial f_a(w) = \partial \|\cdot\|_{X^{n-1}}(x_1' - x_n' - a_1, \ldots, x_{n-1}' - x_n' - a_{n-1}) \times \{0\},$$

where $\|\cdot\|_{X^{n-1}}$ is the maximum norm in X^{n-1}. Condition (5.8a) follows immediately from (5.3c). Moreover, since $f_a(w) > 0$, we can apply [Zălinescu, 2002, Corollary 2.4.16] to find that condition (5.8b) is satisfied, and

$$\sum_{i=1}^{n-1} \langle x_i^*, x_n' - x_i' + a_i \rangle = f_a(w). \tag{5.10}$$

Let an arbitrary number $\tau \in (0,1)$ be fixed, and let $\eta := \rho(1-\tau)/4$. In view of (5.7), we have

$$\max_{1\le i\le n-1} \|x_n - x_i + a_i\| \ge \rho. \tag{5.11}$$

Using (5.6), (5.8b), (5.10), and (5.11), we can prove the remaining estimate (5.9):

$$\begin{aligned}
\sum_{i=1}^{n-1} \langle x_i^*, x_n - x_i + a_i\rangle &\ge \sum_{i=1}^{n-1} \left(\langle x_i^*, x_n' - x_i' + a_i \rangle - 2\,\|x_i^*\| \max_{1\le j\le n} \|x_j - x_j'\| \right) \\
&> \sum_{i=1}^{n-1} \langle x_i^*, x_n' - x_i' + a_i \rangle - 2\eta \\
&= \max_{1\le i\le n-1} \|x_n' - x_i' + a_i\| - 2\eta \\
&> \max_{1\le i\le n-1} \|x_n - x_i + a_i\| - 4\eta \\
&= \max_{1\le i\le n-1} \|x_n - x_i + a_i\| - \rho(1-\tau) \\
&\ge \tau \max_{1\le i\le n-1} \|x_n - x_i + a_i\|.
\end{aligned}$$

This completes the proof. □

The next example illustrates application of Theorem 5.2 in the case where $\mathcal{F}$ consists of discontinuous functions.

Example 5.6. *Consider the closed sets $\Omega_1, \Omega_2 \subset \mathbb{R}^2$ given by*

$$\Omega_1 := \{(x,y) \mid \max(y, x+y) \ge 0\}, \quad \Omega_2 := \{(x,y) \mid y \le 0\}.$$

Let us equip $\mathbb{R}^2$ with the Euclidean norm. We have $(0,0) \in \Omega_1 \cap \Omega_2$ and $\operatorname{int}(\Omega_1 \cap \Omega_2) = \{(x,y) \,|\, y > 0,\ x+y > 0\}$. Hence, these sets cannot be "pushed apart", and $\{\Omega_1, \Omega_2\}$ is not extremal at $(0,0)$ in the conventional sense, see Figure 5.1(b) for an illustration. Let the family $\mathcal{F}$ consist of all nonnegative lower semicontinuous functions $f_t\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}_\infty$ of the type

$$\forall (x,y),(u,v) \in \mathbb{R}^2 \times \mathbb{R}^2\colon \quad f_t((x,y),(u,v)) := \|(x, y+t)\| + i_{(-\infty,0]}(u), \tag{5.12}$$

corresponding to all $t \ge 0$.

We now show that $\{\Omega_1, \Omega_2\}$ is $\mathcal{F}$-extremal at $(0,0)$. Indeed, for each $\varepsilon > 0$ and $t \in (0,\varepsilon)$, we have $f_t((0,0),(0,0)) = t < \varepsilon$. The function from (5.1) takes the form

$$\forall (x,y),(u,v) \in \mathbb{R}^2 \times \mathbb{R}^2\colon \quad \hat{f}_t((x,y),(u,v)) := \|(x-u, y-v+t)\| + i_{(-\infty,0]}(u).$$

Let $\rho \in (0, t/3)$, $(x,y) \in \Omega_1 + \rho\mathbb{B}$, and $(u,v) \in \Omega_2 + \rho\mathbb{B}$. If $u > 0$ or $x \ne u$, then $\hat{f}_t((x,y),(u,v)) > 0$. Let $x = u \le 0$. Then $y > -2\rho$, $v < \rho$, and, consequently, $\hat{f}_t((x,y),(u,v)) = |y - v + t| > -3\rho + t > 0$. Hence, condition (5.2) holds, i.e., $\{\Omega_1, \Omega_2\}$ is $\mathcal{F}$-extremal at $(0,0)$.

For each $t \geq 0$, $\hat{f}_t$ is Lipschitz continuous on $\operatorname{dom}\hat{f}_t = \mathbb{R}^2 \times ((-\infty, 0] \times \mathbb{R})$ *and, for every point* $((x, y), (u, v)) \in \operatorname{dom}\hat{f}_t$*, the distance* $\operatorname{dist}_{\Omega_1\times\Omega_2}((x, y), (u, v))$ *is attained at some point* $((x', y'), (u', v'))$ *with* $u' = u$*, i.e.,* $((x', y'), (u', v')) \in \operatorname{dom}\hat{f}_t$*. Using this, it is easy to see from Definition 3.3 (b) that* $\hat{f}_t$ *is lower semicontinuous near* $((0, 0), (0, 0))$ *relative to* $\Omega_1 \times \Omega_2$*.*

By Theorem 5.2, for each $\varepsilon > 0$*, there exist a number* $t \in (0, \varepsilon)$ *and points* $(x, y) \in \Omega_1 \cap B_\varepsilon(0, 0)$*,* $(u, v) \in \Omega_2 \cap B_\varepsilon(0, 0)$*,* $(x^*, y^*), (u^*, v^*) \in \mathbb{R}^2$*, and* $w \in X^2 \times X^2$ *such that* $0 < f_t(w) < \infty$ *and*

$$\operatorname{dist}_{N_{\Omega_1}(x,y)}((x^*, y^*)) + \operatorname{dist}_{N_{\Omega_2}(u,v)}((u^*, v^*)) < \varepsilon, \tag{5.13a}$$

$$-((x^*, y^*), (x^*, y^*) + (u^*, v^*)) \in \partial f_t(w). \tag{5.13b}$$

In view of (5.12), it follows from (5.13b) that $\|(x^*, y^*)\| = 1$*,* $x^* + u^* \leq 0$*, and* $y^* + v^* = 0$*. When* ε *is sufficiently small, condition (5.13a) implies one of the following situations:*

- $x < 0$*,* $y = v = 0$*, and* (x^*, y^*) *as well as* (u^*, v^*) *can be made arbitrarily close to* $(0, -1)$ *and* $(0, 1)$*, respectively,*
- $x > 0$*,* $y = -x$*,* $v = 0$*, and* (x^*, y^*) *as well as* (u^*, v^*) *can be made arbitrarily close to* $(-\sqrt{2}/2, -\sqrt{2}/2)$ *and* $(0, \sqrt{2}/2)$*, respectively.*

This can be interpreted as a kind of generalized separation.

6 Geometrically-constrained optimization problems with composite objective function

In this section, we are going to apply the theory of Section 4 to the optimization problem

$$\min\{f(x) + q(x) \,|\, G(x) \in K,\, x \in C\} \tag{Q}$$

where $f\colon X \to \mathbb{R}$ is continuously Fréchet differentiable, $q\colon X \to \mathbb{R}_\infty$ is lower semicontinuous, $G\colon X \to Y$ is continuously Fréchet differentiable, and $C \subset X$ as well as $K \subset Y$ are nonempty and closed. Here, X and Y are assumed to be Banach spaces. Throughout the section, the feasible set of (Q) will be denoted by $\mathcal{S}$, and we implicitly assume $\mathcal{S} \cap \operatorname{dom} q \neq \varnothing$ in order to avoid trivial situations.

Observe that the objective function $\varphi := f + q$ can be decomposed into a regular part f and some challenging irregular part q while the constraints in (Q) are stated in so-called geometric form. In this regard, the model (Q) still covers numerous applications ranging from data science and image processing (in case where q is a sparsity-promoting functional) over conic programs (in which case K is a convex cone) to disjunctive programs which comprise, exemplary, complementarity- and cardinality-constrained problems (in this situation, K is a nonconvex set of combinatorial structure).

In the subsequently stated remark, we embed program (Q) into the rather general framework which has been discussed in Section 4.

Remark 6.1. *Observing that f is differentiable, we find*

$$\forall x \in X: \quad \partial\varphi(x) = \partial(f+q)(x) = f'(x) + \partial q(x)$$

from the sum rule stated in [Kruger, 2003, Corollary 1.12.2]. The feasibility mapping $\Phi\colon X \rightrightarrows Y \times X$ associated with (Q) is given by means of $\Phi(x) := (G(x) - K, x - C)$ for all $x \in X$, see Example 3.18. We find

$$\operatorname{gph}\Phi = \{(x,(y,x')) \in X \times Y \times X \,|\, (G(x) - y, x - x') \in K \times C\}. \tag{6.1}$$

Observing that the continuously differentiable mapping $(x, y, x') \mapsto (G(x) - y, x - x')$ possesses a surjective derivative, we can apply the change-of-coordinates formula from [Mordukhovich, 2006, Corollary 1.15] in order to obtain

$$N_{\operatorname{gph}\Phi}(x,(y,x')) = \left\{ (G'(x)^*\lambda + \eta, -\lambda, -\eta) \in X^* \times Y^* \times X^* \,\middle|\, \begin{array}{l} \lambda \in N_K(G(x) - y), \\ \eta \in N_C(x - x') \end{array} \right\}$$

for each triplet $(x,(y,x')) \in \operatorname{gph}\Phi$, and this yields

$$D^*\Phi(x,(y,x'))(\lambda,\eta) = \begin{cases} G'(x)^*\lambda + \eta & \text{if } \lambda \in N_K(G(x) - y),\ \eta \in N_C(x - x'), \\ \varnothing & \text{otherwise} \end{cases}$$

for arbitrary $\lambda \in Y^*$ and $\eta \in X^*$.

6.1 Approximate stationarity and uniform qualification condition

The subsequent theorem is a simple consequence of Corollary 4.3 and Remark 6.1, and provides a necessary optimality condition for (Q).

Theorem 6.2. *Fix $\bar{x} \in \mathcal{S} \cap \operatorname{dom} q$ and assume that*

(a) the function $f + q$ is lower semicontinuous near $\bar{x}$ relative to Φ from Remark 6.1 at $(0, 0)$;

(b) either X and Y are Asplund, or f, q, and $\operatorname{gph}\Phi$ from (6.1) are convex.

Suppose that $\bar{x}$ is a local minimizer of (Q). Then, for each $\varepsilon > 0$, there exist points $x, x', x'' \in B_\varepsilon(\bar{x})$ and $y \in \varepsilon\mathbb{B}$ such that $|q(x) - q(\bar{x})| < \varepsilon$ and

$$0 \in f'(x) + \partial q(x) + G'(x')^* N_K(G(x') - y) + N_C(x'') + \varepsilon\mathbb{B}^*. \tag{6.2}$$

In the subsequent remark, we comment on some special situations where the assumptions of Theorem 6.2 are naturally valid and which can be checked in terms of initial data.

Remark 6.3. *Let $\bar{x} \in \mathcal{S} \cap \operatorname{dom} q$. Due to Proposition 3.10, Corollaries 3.11 and 3.19, and Example 3.18, each of the following conditions implies condition (a) of Theorem 6.2:*

(a) *the function* $f+q$ *satisfies one of the conditions (a)-(d) in Corollary 3.13 and the mapping* Φ *from Remark 6.1 is metrically subregular at* $(\bar{x},(0,0))$*, see Example 3.18;*

(b) X *is reflexive, the functions* f *and* q *are weakly sequentially lower semicontinuous, and condition (3.5) holds for all sequences* $\{x_k\}_{k\in\mathbb{N}} \subset X$ *and all points* $x \in X$*.*

Furthermore, condition (b) of Theorem 6.2 is valid whenever X *and* Y *are Asplund, or if* f*,* q*, and* C *are convex,* K *is a convex cone, and* G *is* K*-convex in the following sense:*

$$\forall x, x' \in X\, \forall s \in [0,1]\colon \quad G(sx+(1-s)x') - s\,G(x) - (1-s)G(x') \in K.$$

We note that (Q) already satisfies condition (b) of Remark 6.3 as soon as X and Y are finite-dimensional. In the presence of condition (b) from Remark 6.3, Theorem 6.2 is closely related to [Börgens et al., 2020, Proposition 3.3] as soon as q is absent.

Due to Theorem 6.2, the following definition is reasonable.

Definition 6.4. *A point* $\bar{x} \in \mathcal{S} \cap \operatorname{dom} q$ *is an* approximately stationary point *of (Q) if, for each* $\varepsilon > 0$*, there exist points* $x, x', x'' \in B_\varepsilon(\bar{x})$ *and* $y \in \varepsilon\mathbb{B}$ *such that* $|q(x) - q(\bar{x})| < \varepsilon$ *and (6.2) are valid.*

Approximate necessary optimality conditions in terms of Fréchet subgradients and normals can be traced back to the 1980s, see e.g. Kruger and Mordukhovich [1980], Kruger [1985] and the references therein.

In order to compare the notion of stationarity from Definition 6.4 to others from the literature, let us mention an equivalent characterization of asymptotic stationarity in terms of sequences.

Remark 6.5. *A point* $\bar{x} \in \mathcal{S} \cap \operatorname{dom} q$ *is approximately stationary if and only if there are sequences* $\{x_k\}_{k\in\mathbb{N}}, \{x'_k\}_{k\in\mathbb{N}}, \{x''_k\}_{k\in\mathbb{N}} \subset X$*,* $\{y_k\}_{k\in\mathbb{N}} \subset Y$*, and* $\{\eta_k\}_{k\in\mathbb{N}} \subset X^*$ *such that* $x_k \to \bar{x}$*,* $x'_k \to \bar{x}$*,* $x''_k \to \bar{x}$*,* $y_k \to 0$*,* $\eta_k \to 0$*,* $q(x_k) \to q(\bar{x})$*, and*

$$\forall k \in \mathbb{N}\colon \quad \eta_k \in f'(x_k) + \partial q(x_k) + G'(x'_k)^* N_K(G(x'_k) - y_k) + N_C(x''_k).$$

In case where X and Y are finite-dimensional while q is locally Lipschitzian, a similar approximate stationarity condition in terms of sequences has been investigated in [Mehlitz, 2020, Sections 4, 5.1]. In Börgens et al. [2020], the authors considered the model (Q) with convex sets K and C in the absence of q. Generally, using approximate notions of stationarity in nonlinear programming can be traced back to Andreani et al. [2010, 2011]. Let us mention that in all these papers, the authors speak of *asymptotic* or *sequential* stationarity conditions. A sequential Lagrange multiplier rule for convex programs in Banach spaces can be found already in Thibault [1997].

During the last decade, the concept of approximate stationarity has been extended to several classes of optimization problems comprising, exemplary, complementarity- and cardinality-constrained programs, see Andreani et al. [2019b], Kanzow et al. [2021], Ramos [2021], conic optimization problems, see Andreani et al. [2020], smooth geometrically-constrained optimization problems in Banach spaces, see Börgens et al. [2020], and nonsmooth Lipschitzian optimization problems in finite-dimensional spaces, see Mehlitz

[2020, 2021]. In each of the aforementioned situations, it has been demonstrated that approximate stationarity, on the one hand, provides a necessary optimality condition in the absence of constraint qualifications, and Theorem 6.2 demonstrates that this is the case for our concept from Definition 6.4 as well under reasonable assumptions. On the other hand, the results from the literature underline that approximate stationarity is naturally satisfied for accumulation points of sequences generated by some solution algorithms. In Section 6.2, we extend these observations to the present setting.

Assume that $\bar{x} \in \mathcal{S} \cap \operatorname{dom} q$ is an approximately stationary point of (Q). Due to Remark 6.5, we find sequences $\{x_k\}_{k\in\mathbb{N}}, \{x'_k\}_{k\in\mathbb{N}}, \{x''_k\}_{k\in\mathbb{N}} \subset X$, $\{y_k\}_{k\in\mathbb{N}} \subset Y$, and $\{\eta_k\}_{k\in\mathbb{N}} \subset X^*$ satisfying $x_k \to \bar{x}$, $x'_k \to \bar{x}$, $x''_k \to \bar{x}$, $y_k \to 0$, $\eta_k \to 0$, $q(x_k) \to q(\bar{x})$, and $\eta_k \in f'(x_k) + \partial q(x_k) + G'(x'_k)^* N_K(G(x'_k) - y_k) + N_C(x''_k)$ for each $k \in \mathbb{N}$. Particularly, we find sequences $\{\lambda_k\}_{k\in\mathbb{N}} \subset Y^*$ and $\{\mu_k\}_{k\in\mathbb{N}} \subset X^*$ of multipliers and a sequences $\{\xi_k\}_{k\in\mathbb{N}} \subset X^*$ of subgradients such that $\eta_k = f'(x_k) + \xi_k + G'(x'_k)^*\lambda_k + \mu_k$, $\lambda_k \in N_K(G(x'_k) - y_k)$, $\mu_k \in N_C(x''_k)$, and $\xi_k \in \partial q(x_k)$ for each $k \in \mathbb{N}$. Assuming for a moment $\lambda_k \overset{*}{\rightharpoonup} \lambda$, $\mu_k \overset{*}{\rightharpoonup} \mu$, and $\xi_k \overset{*}{\rightharpoonup} \xi$ for some $\lambda \in Y^*$ and $\mu, \xi \in X^*$, we find $\lambda \in \overline{N}_K(G(\bar{x}))$, $\mu \in \overline{N}_C(\bar{x})$, and $\xi \in \overline{\partial} q(\bar{x})$ by definition of the limiting normal cone and subdifferential, respectively, as well as $0 = f'(\bar{x}) + \xi + G'(\bar{x})^*\lambda + \mu$, i.e., a multiplier rule is valid at $\bar{x}$ which is referred to as M-stationarity in the literature.

Definition 6.6. *A feasible point $\bar{x} \in \mathcal{S} \cap \operatorname{dom} q$ is an* M-stationary *point of* (Q) *if*

$$0 \in f'(\bar{x}) + \overline{\partial} q(\bar{x}) + G'(\bar{x})^* \overline{N}_K(G(\bar{x})) + \overline{N}_C(\bar{x}).$$

Let us note that in the case of standard nonlinear programming, where q vanishes while $C := X$, $Y := \mathbb{R}^{m_1+m_2}$, and $K := (-\infty, 0]^{m_1} \times \{0\}^{m_2}$ for $m_1, m_2 \in \mathbb{N}$, the system of M-stationarity coincides with the classical Karush–Kuhn–Tucker system.

One can easily check by means of simple examples that approximately stationary points of (Q) do not need to be M-stationary even in finite dimensions. Roughly speaking, this phenomenon is caused by the fact that the multiplier and subgradient sequences $\{\lambda_k\}_{k\in\mathbb{N}}$, $\{\mu_k\}_{k\in\mathbb{N}}$, and $\{\xi_k\}_{k\in\mathbb{N}}$ in the considerations which prefixed Definition 6.6 do not need to be bounded, see [Mehlitz, 2020, Section 3.1] for related observations. The following example is inspired by [Mehlitz, 2020, Example 3.3].

Example 6.7. *We consider $X = Y = C := \mathbb{R}$, set $f(x) := x$, $q(x) := 0$, as well as $G(x) := x^2$ for all $x \in \mathbb{R}$, and fix $K := (-\infty, 0]$. Let us investigate $\bar{x} := 0$. Note that this is the only feasible point of the associated optimization problem* (Q) *and, thus, its uniquely determined global minimizer. Due to $f'(\bar{x}) = 1$ and $G'(\bar{x}) = 0$, $\bar{x}$ cannot be an M-stationary point of* (Q)*. On the other hand, setting*

$$x_k := 0, \quad x'_k := -\frac{1}{2k}, \quad y_k := \frac{1}{4k^2}, \quad \eta_k := 0, \quad \lambda_k := k$$

for each $k \in \mathbb{N}$, we have $x_k \to \bar{x}$, $x'_k \to \bar{x}$, $y_k \to 0$, $\eta_k \to 0$, as well as $\eta_k = f'(x_k) + G'(x'_k)^*\lambda_k$ and $\lambda_k \in N_K(G(x'_k) - y_k)$ for each $k \in \mathbb{N}$, i.e., $\bar{x}$ is approximately stationary for (Q)*. Observe that $\{\lambda_k\}_{k\in\mathbb{N}}$ is unbounded.*

Let us underline that the above example demonstrates that local minimizers of (Q) do not need to be M-stationary in general while approximate stationarity serves as a necessary optimality condition under some assumptions on the data which are inherent in finite dimensions, see Theorem 6.2 and Remark 6.3. Nevertheless, M-stationarity turned out to be a celebrated stationarity condition in finite-dimensional optimization. On the one hand, it is restrictive enough to exclude non-reasonable feasible points of (Q) when used as a necessary optimality condition. On the other hand, it is weak enough to hold at the local minimizers of (Q) under very mild qualification conditions. Exemplary, we would like to refer the reader to Flegel et al. [2007] where this is visualized by so-called disjunctive programs where K is the union of finitely many polyhedral sets. Another interest in M-stationarity arises from the fact that this system can often be solved directly in order to identify reasonable feasible points of (Q), see e.g. Guo et al. [2015], Harder et al. [2021]. In infinite-dimensional optimization, particularly, in optimal control, M-stationarity has turned out to be of limited practical use since the limiting normal cone to nonconvex sets in function spaces is uncomfortably large due to convexification effects arising when taking weak limits, see e.g. Harder and Wachsmuth [2018], Mehlitz and Wachsmuth [2018].

Due to this interest in M-stationarity, at least from the finite-dimensional point of view, we aim to find conditions guaranteeing that a given approximately stationary point of (Q) is already M-stationary.

Definition 6.8. *We say that the* uniform qualification condition *holds at* $\bar{x} \in \mathcal{S} \cap \operatorname{dom} q$ *whenever*

$$\begin{aligned}\limsup_{\substack{x\to\bar{x},\, x'\to\bar{x},\, x''\to\bar{x},\\ y\to 0,\, q(x)\to q(\bar{x})}} &\big(\partial q(x) + G'(x')^* N_K(G(x') - y) + N_C(x'')\big)\\ &\subset \overline{\partial} q(\bar{x}) + G'(\bar{x})^* \overline{N}_K(G(\bar{x})) + \overline{N}_C(\bar{x}).\end{aligned}$$

By construction, the uniform qualification condition guarantees that a given approximately stationary point of (Q) is already M-stationary as desired.

Proposition 6.9. *Let* $\bar{x} \in \mathcal{S} \cap \operatorname{dom} q$ *satisfy the uniform qualification condition. If* $\bar{x}$ *is an approximately stationary point of* (Q)*, then it is M-stationary.*

Proof. By definition of approximate stationarity, for each $k \in \mathbb{N}$, we find $x_k, x'_k, x''_k \in B_{1/k}(\bar{x})$, $y_k \in \frac{1}{k}\mathbb{B}$, and $\eta_k \in \frac{1}{k}\mathbb{B}^*$ such that $|q(x_k) - q(\bar{x})| < \frac{1}{k}$ and $\eta_k - f'(x_k) \in \partial q(x_k) + G'(x'_k)^* N_K(G(x'_k) - y_k) + N_C(x''_k)$. Since f is assumed to be continuously differentiable, we find $\eta_k - f'(x_k) \to -f'(\bar{x})$. Thus, by validity of the uniform qualification condition, it holds

$$\begin{aligned}-f'(\bar{x}) &\in \limsup_{k\to+\infty} \big(\partial q(x_k) + G'(x'_k)^* N_K(G(x'_k) - y_k) + N_C(x''_k)\big)\\ &\subset \overline{\partial} q(\bar{x}) + G'(\bar{x})^* \overline{N}_K(G(\bar{x})) + \overline{N}_C(\bar{x}),\end{aligned}$$

i.e., $\bar{x}$ is an M-stationary point of (Q). ☐

Combining this with Theorem 6.2 yields the following result.

Corollary 6.10. *Let $\bar{x} \in \mathcal{S} \cap \operatorname{dom} q$ be a local minimizer of (Q) which satisfies the assumptions of Theorem 6.2 as well as the uniform qualification condition. Then $\bar{x}$ is M-stationary.*

Observe that we do not need any so-called *sequential normal compactness condition*, see [Mordukhovich, 2006, Section 1.1.4], for the above statement to hold which pretty much contrasts the results obtained in [Mordukhovich, 2006, Section 5]. Indeed, sequential normal compactness is likely to fail in the function space context related to optimal control, see Mehlitz [2019].

Let us point the reader's attention to the fact that the uniform qualification condition is not a constraint qualification in the narrower sense for (Q) since it also depends on (parts of) the objective function. Nevertheless, Corollary 6.10 shows that it may serve as a qualification condition for M-stationarity of local minimizers under mild assumptions on the data. In the absence of q, the uniform qualification condition is related to other prominent so-called *sequential* or *asymptotic* constraint qualifications from the literature which address several different kinds of optimization problems, see e.g. Andreani et al. [2019a,b, 2016], Börgens et al. [2020], Mehlitz [2020, 2021], Ramos [2021]. In Section 6.3, we demonstrate by means of a prominent setting from optimal control that the uniform qualification condition may hold in certain situations where q is present, see Lemma 6.17.

Remark 6.11. *Note that in the particular setting $q \equiv 0$, the uniform qualification condition from Definition 6.8 at some point $\bar{x} \in \mathcal{S}$ simplifies to*

$$\limsup_{x' \to \bar{x},\, x'' \to \bar{x},\, y \to 0} \left(G'(x')^* N_K(G(x') - y) + N_C(x'') \right) \subset G'(\bar{x})^* \overline{N}_K(G(\bar{x})) + \overline{N}_C(\bar{x}). \quad (6.3)$$

In the light of Proposition 6.9 and Corollary 6.10, (6.3) serves as a constraint qualification guaranteeing M-stationarity of $\bar{x}$ under mild assumptions as soon as this point is a local minimizer of the associated problem (Q). One may, thus, refer to (6.3) as the uniform constraint qualification.

Observations related to the ones from Remark 6.11 have been made in Börgens et al. [2020], [Jia et al., 2021, Section 2.2], and [Mehlitz, 2020, Section 5.1] and underline that (6.3) is a comparatively weak constraint qualification whenever $q \equiv 0$. Exemplary, let us mention that whenever X and Y are finite-dimensional the *generalized Mangasarian–Fromovitz constraint qualification*

$$-G'(\bar{x})^* \lambda \in \overline{N}_C(\bar{x}),\ \lambda \in \overline{N}_K(G(\bar{x})) \quad \Longrightarrow \quad \lambda = 0 \quad (6.4)$$

is sufficient for (6.3) to hold, but the uniform constraint qualification is often much weaker than (6.4) which corresponds to metric regularity of Φ from Remark 6.1 at $(\bar{x}, (0,0))$, see [Mehlitz, 2020, Section 3.2] for related discussions. Let us also mention that (6.4) is sufficient for metric subregularity of Φ at $(\bar{x}, (0,0))$ exploited in Corollary 3.19.

The following proposition provides a sufficient condition for validity of the uniform qualification condition in case where X is finite-dimensional.

Proposition 6.12. *Let X be finite-dimensional and $\bar{x} \in \mathcal{S} \cap \operatorname{dom} q$. Suppose that the uniform constraint qualification (6.3) is valid at $\bar{x}$, and*

$$\left(G'(\bar{x})^* \overline{N}_K(G(\bar{x})) + \overline{N}_C(\bar{x})\right) \cap (-\overline{\partial}^\infty q(\bar{x})) = \{0\}. \tag{6.5}$$

Then the uniform qualification condition holds at $\bar{x}$.

Proof. Let us fix

$$x^* \in \limsup_{\substack{x \to \bar{x},\, x' \to \bar{x},\, x'' \to \bar{x}, \\ y \to 0,\ q(x) \to q(\bar{x})}} \left(\partial q(x) + G'(x')^* N_K(G(x') - y) + N_C(x'')\right).$$

Then we find sequences $\{x_k\}_{k\in\mathbb{N}}, \{x'_k\}_{k\in\mathbb{N}}, \{x''_k\}_{k\in\mathbb{N}} \subset X$, $\{y_k\}_{k\in\mathbb{N}} \subset Y$, and $\{x^*_k\}_{k\in\mathbb{N}} \subset X^*$ such that $x_k \to \bar{x}$, $x'_k \to \bar{x}$, $x''_k \to \bar{x}$, $y_k \to \bar{y}$, $q(x_k) \to q(\bar{x})$, and $x^*_k \to x^*$ as well as $x^*_k \in \partial q(x_k) + G'(x'_k)^* N_K(G(x'_k) - y_k) + N_C(x''_k)$ for all $k \in \mathbb{N}$. Thus, there are sequences $\{u^*_k\}_{k\in\mathbb{N}}, \{v^*_k\}_{k\in\mathbb{N}} \subset X^*$ satisfying $x^*_k = u^*_k + v^*_k$, $u^*_k \in \partial q(x_k)$, and $v^*_k \in G'(x'_k)^* N_K(G(x'_k) - y_k) + N_C(x''_k)$ for all $k \in \mathbb{N}$.

Let us assume that $\{u^*_k\}_{k\in\mathbb{N}}$ is unbounded. Then, due to $x^*_k \to x^*$, $\{v^*_k\}_{k\in\mathbb{N}}$ is unbounded, too. For each $k \in \mathbb{N}$, we define $\tilde{u}^*_k := u^*_k/(\|u^*_k\| + \|v^*_k\|)$ and $\tilde{v}^*_k := v^*_k/(\|u^*_k\| + \|v^*_k\|)$, i.e., the sequence $\{(\tilde{u}^*_k, \tilde{v}^*_k)\}_{k\in\mathbb{N}}$ belongs to the unit sphere of $X^* \times X^*$. Without loss of generality, we may assume $\tilde{u}^*_k \to \tilde{u}^*$ and $\tilde{v}^*_k \to \tilde{v}^*$ for some $\tilde{u}^*, \tilde{v}^* \in X^*$ since X is finite-dimensional. We note that $\tilde{u}^*$ and $\tilde{v}^*$ cannot vanish at the same time. Taking the limit in $x^*_k/(\|u^*_k\| + \|v^*_k\|) = \tilde{u}^*_k + \tilde{v}^*_k$, we obtain $0 = \tilde{u}^* + \tilde{v}^*$. By definition of the singular limiting subdifferential, we have $\tilde{u}^* \in \overline{\partial}^\infty q(\bar{x})$ while

$$\tilde{v}^* \in \limsup_{k\to+\infty} \left(G'(x'_k)^* N_K(G(x'_k) - y_k) + N_C(x''_k)\right) \subset G'(\bar{x})^* \overline{N}_K(G(\bar{x})) + \overline{N}_C(\bar{x})$$

follows by the uniform constraint qualification (6.3). Thus, we find $\tilde{u}^* = \tilde{v}^* = 0$ from condition (6.5). The latter, however, contradicts $(\tilde{u}^*, \tilde{v}^*) \neq (0, 0)$.

From above, we now know that $\{u^*_k\}_{k\in\mathbb{N}}$ and $\{v^*_k\}_{k\in\mathbb{N}}$ are bounded. Without loss of generality, we may assume $u^*_k \to u^*$ and $v^*_k \to v^*$ for some $u^*, v^* \in X^*$. By definition of the limiting subdifferential we have $u^* \in \overline{\partial} q(\bar{x})$, and $v^* \in G'(\bar{x})^* \overline{N}_K(G(\bar{x})) + \overline{N}_C(\bar{x})$ is guaranteed by the uniform constraint qualification (6.3). Thus, we end up with $x^* \in \overline{\partial} q(\bar{x}) + G'(\bar{x})^* \overline{N}_K(G(\bar{x})) + \overline{N}_C(\bar{x})$ which completes the proof. $\square$

Proposition 6.12 shows that in case where X is finite-dimensional, validity of the uniform qualification condition can be guaranteed in the presence of two conditions. The first one, represented by condition (6.3), is a sequential constraint qualification which guarantees regularity of the constraints at the reference point. The second one, given by condition (6.5), ensures in some sense that the challenging part of the objective function and the constraints of (Q) are somewhat compatible at the reference point. A similar *decomposition* of qualification conditions has been used in Chen et al. [2017], Guo and Ye [2018] in order to ensure M-stationarity of standard nonlinear problems in finite dimensions with a composite objective function. In the latter papers, the authors referred to a condition of type (6.5) as *basic qualification*, and this terminology can be traced back to the works of Mordukhovich, see e.g. Mordukhovich [2006].

Note that in order to transfer Proposition 6.12 to the infinite-dimensional setting, one would be in need to postulate sequential compactness properties on q or the constraint data which are likely to fail in several interesting function spaces, see Mehlitz [2019] again.

6.2 Augmented Lagrangian methods for optimization problems with non-Lipschitzian objective functions

We consider the optimization problem (Q) such that X is an Asplund space, Y is a Hilbert space with $Y \cong Y^*$, and K is convex. Let us note that the assumption on Y can be relaxed by assuming the existence of a Hilbert space H with $H \cong H^*$ such that (Y, H, Y^*) is a Gelfand triplet, see [Börgens et al., 2020, Section 7] or Börgens et al. [2019], Kanzow et al. [2018] for a discussion. Furthermore, we will exploit the following assumption which is standing throughout this section.

Assumption 6.13. *At least one of the following assumptions is valid.*

(a) The space X is finite-dimensional.

(b) The function q is uniformly continuous.

(c) The functions f, q, and $x \mapsto \operatorname{dist}_K^2(G(x))$ are weakly sequentially lower semicontinuous and C is weakly sequentially closed. Furthermore, X is reflexive.

Throughout this subsection, we assume that C is a comparatively simple set, e.g., a box if X is equipped with a (partial) order relation, while the constraints $G(x) \in K$ are difficult and will be treated with the aid of a multiplier-penalty approach. In this regard, for some penalty parameter $\theta > 0$, we investigate the (partial) augmented Lagrangian function $\mathcal{L}_\theta \colon X \times Y \to \mathbb{R}_\infty$ given by

$$\forall (x, \lambda) \in X \times Y \colon \quad \mathcal{L}_\theta(x, \lambda) := f(x) + \frac{\theta}{2} \operatorname{dist}_K^2 \left(G(x) + \frac{\lambda}{\theta} \right) + q(x).$$

We would like to point the reader's attention to the fact that the second summand in the definition of $\mathcal{L}_\theta$ is continuously differentiable since the squared distance to a convex set possesses this property. For the control of the penalty parameter, we make use of the function $V_\theta \colon X \times Y \to \mathbb{R}$ given by

$$\forall (x, y) \in X \times Y \colon \quad V_\theta(x, \lambda) := \| G(x) - P_K(G(x) + \lambda/\theta) \| .$$

The method of interest is now given as stated in Algorithm 1.

We would like to point the reader's attention to the fact that Algorithm 1 is a so-called *safeguarded* augmented Lagrangian method since the multiplier estimates u_k are chosen from the bounded set B. In practice, one typically chooses B as a (very large) box, and defines u_k as the projection of λ_k onto B in (S.2). Note that without safeguarding, one obtains the classical augmented Lagrangian method. However, it is well known that the safeguarded version possesses superior global convergence properties, see Kanzow and

Algorithm 1 Safeguarded augmented Lagrangian method for (Q).

(S.0) Choose $(x_0, \lambda_0) \in (\operatorname{dom} q) \times Y$, $\theta_0 > 0$, $\gamma > 1$, $\tau \in (0,1)$, and a nonempty, bounded set $B \subset Y$ arbitrarily. Set $k := 0$.

(S.1) If (x_k, λ_k) satisfies a suitable termination criterion, then stop.

(S.2) Choose $u_k \in B$ and find an approximate solution $x_{k+1} \in C \cap \operatorname{dom} q$ of

$$\min\{\mathcal{L}_{\theta_k}(x, u_k) \,|\, x \in C\}. \tag{6.6}$$

(S.3) Set

$$\lambda_{k+1} := \theta_k \left[G(x_{k+1}) + u_k/\theta_k - P_K\left(G(x_{k+1}) + u_k/\theta_k\right)\right].$$

(S.4) If $k = 0$ or $V_{\theta_k}(x_{k+1}, u_k) \le \tau\, V_{\theta_{k-1}}(x_k, u_{k-1})$, then set $\theta_{k+1} := \theta_k$. Otherwise, set $\theta_{k+1} := \gamma\, \theta_k$.

(S.5) Go to (S.1).

Steck [2017]. An overview of augmented Lagrangian methods in constrained optimization can be found in Birgin and Martínez [2014].

Let us comment on potential termination criteria for Algorithm 1. On the one hand, Algorithm 1 is designed for the computation of M-stationary points of (Q) which, at the latest, will become clear in Corollary 6.16. Thus, one may check approximate validity of these stationarity conditions in (S.1). However, if q or C is variationally challenging, this might be a nontrivial task. On the other hand, at its core, Algorithm 1 is a penalty method, so it is also reasonable to check approximate feasibility with respect to the constraints $G(x) \in K$ in (S.1).

In Chen et al. [2017], the authors suggest to solve (Q), where all involved spaces are instances of $\mathbb{R}^n$ while the constraints $G(x) \in K$ are replaced by smooth inequality and equality constraints, with the classical augmented Lagrangian method. In case where q is not present and X as well as Y are Euclidean spaces, Algorithm 1 recovers the partial augmented Lagrangian scheme studied in Jia et al. [2021] where the authors focus on situations where C is nonconvex and of challenging variational structure. We note that, technically, Algorithm 1 is also capable of handling this situation. However, it might be difficult to solve the appearing subproblems (6.6) if both q and C are variationally complex. Note that we did not specify in (S.2) how precisely the subproblems have to be solved. Exemplary, one could aim to find stationary or globally ε-minimal points of the function $\mathcal{L}_{\theta_k}(\cdot, u_k)_C$ here. We comment on both situations below.

Our theory from Section 4 can be used to show that Algorithm 1 computes approximately stationary points of (Q) when the subproblems (6.6) are solved up to stationarity of $\mathcal{L}_{\theta_k}(\cdot, u_k)_C$.

Theorem 6.14. *Let $\{x_k\}_{k\in\mathbb{N}}$ be a sequence generated by Algorithm 1 such that x_{k+1}*

is a stationary point of $\mathcal{L}_{\theta_k}(\cdot, u_k)_C$ for each $k \in \mathbb{N}$. Assume that, along a subsequence (without relabeling), we have $x_k \to \bar{x}$ and $q(x_k) \to q(\bar{x})$ for some $\bar{x} \in X$ which is feasible to (Q). Then $\bar{x}$ is an approximately stationary point of (Q).

Proof. Observe that Assumption 6.13 guarantees that $\mathcal{L}_{\theta_k}(\cdot, u_k)$ is lower semicontinuous relative to C near each point from $C \cap \operatorname{dom} q$, see Corollaries 3.13 and 3.15. Since x_{k+1} is a stationary point of $\mathcal{L}_{\theta_k}(\cdot, u_k)_C$, we can apply Remark 2.5 and Theorem 4.5 in order to find $x'_{k+1} \in B_{1/k}(x_{k+1})$ and $x''_{k+1} \in C \cap B_{1/k}(x_{k+1})$ such that $|q(x'_{k+1}) - q(x_{k+1})| < \frac{1}{k}$ and

$$0 \in \partial \mathcal{L}_{\theta_k}(x'_{k+1}, u_k) + N_C(x''_{k+1}) + \tfrac{1}{k}\,\mathbb{B}^*$$

for each $k \in \mathbb{N}$. From $x_k \to \bar{x}$ and $q(x_k) \to q(\bar{x})$ we have $x'_k \to \bar{x}$, $x''_k \to \bar{x}$, and $q(x'_k) \to q(\bar{x})$. Noting that f, G, and, by convexity of K, the squared distance function dist_K^2 are continuously differentiable, we find

$$\begin{aligned} 0 \in f'(x'_{k+1}) + \theta_k\, G'(x'_{k+1})^* \left[G(x'_{k+1}) + u_k/\theta_k - P_K\left(G(x'_{k+1}) + u_k/\theta_k\right)\right] \\ + \partial q(x'_{k+1}) + N_C(x''_{k+1}) + \tfrac{1}{k}\,\mathbb{B}^* \end{aligned} \tag{6.7}$$

for each $k \in \mathbb{N}$ where we used the subdifferential sum rule from [Kruger, 2003, Corollary 1.12.2]. Let us set $y_{k+1} := G(x'_{k+1}) - P_K(G(x'_{k+1}) + u_k/\theta_k)$ for each $k \in \mathbb{N}$. By definition of the projection and convexity of K, we find

$$\theta_k(y_k + u_k/\theta_k) \in N_K(P_K(G(x'_{k+1}) + u_k/\theta_k)) = N_K(G(x'_{k+1}) - y_{k+1}),$$

so we can rewrite (6.7) by means of

$$0 \in f'(x'_{k+1}) + \partial q(x'_{k+1}) + G'(x'_{k+1})^* N_K(G(x'_{k+1}) - y_{k+1}) + N_C(x''_{k+1}) + \tfrac{1}{k}\,\mathbb{B}^* \tag{6.8}$$

for each $k \in \mathbb{N}$.

It remains to show $y_{k+1} \to 0$. We distinguish two cases.

First, assume that $\{\theta_k\}_{k\in\mathbb{N}}$ remains bounded. By construction of Algorithm 1, this yields $V_{\theta_k}(x_{k+1}, u_k) \to 0$ as $k \to +\infty$. Recalling that the projection P_K is Lipschitz continuous with modulus 1 by convexity of K, we have

$$\begin{aligned} \|y_{k+1}\| &\le V_{\theta_k}(x_{k+1}, u_k) + \left\|G(x'_{k+1}) - G(x_{k+1})\right\| \\ &\quad + \left\|P_K(G(x'_{k+1}) + u_k/\theta_k) - P_K(G(x_{k+1}) + u_k/\theta_k)\right\| \\ &\le V_{\theta_k}(x_{k+1}, u_k) + 2\left\|G(x'_{k+1}) - G(x_{k+1})\right\| \end{aligned}$$

for each $k \in \mathbb{N}$. Due to $x_k \to \bar{x}$ and $x'_k \to \bar{x}$ as well as continuity of G, this yields $y_{k+1} \to 0$.

Finally, suppose that $\{\theta_k\}_{k\in\mathbb{N}}$ is unbounded. Since this sequence is monotonically increasing, we have $\theta_k \to +\infty$. By boundedness of $\{u_k\}_{k\in\mathbb{N}}$, continuity of G as well as the projection P_K, $x'_k \to \bar{x}$, and feasibility of $\bar{x}$ for (Q), it holds

$$y_{k+1} = G(x'_{k+1}) - P_K(G(x'_{k+1}) + u_k/\theta_k) \to G(\bar{x}) - P_K(G(\bar{x})) = 0,$$

and this completes the proof. $\square$

Let us mention that the assumption $q(x_k) \to q(\bar{x})$ is trivially satisfied as soon as q is continuous on its domain. For other types of discontinuity, however, this does not follow by construction of the method and has to be presumed. Let us note that this convergence is also implicitly used in the proof of the related result [Chen et al., 2017, Theorem 3.1] but does not follow from the postulated assumptions, i.e., this assumption is missing there.

Note that demanding feasibility of accumulation points is a natural assumption when considering augmented Lagrangian methods. This property naturally holds whenever the sequence $\{\theta_k\}_{k\in\mathbb{N}}$ remains bounded or if q is bounded from below while the sequence $\{\mathcal{L}_{\theta_k}(x_{k+1}, u_k)\}_{k\in\mathbb{N}}$ remains bounded. The latter assumption is typically satisfied whenever globally ε_k-minimal points of $\mathcal{L}_{\theta_k}(\cdot, u_k)_C$ can be computed in order to approximately solve the subproblems (6.6) in (S.2), where $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset [0, +\infty)$ is a bounded sequence. Indeed, we have

$$\forall x \in \mathcal{S}\colon \quad \mathcal{L}_{\theta_k}(x_{k+1}, u_k) \le \mathcal{L}_{\theta_k}(x, u_k) + \varepsilon_k \le f(x) + \|u_k\|^2/(2\theta_k) + q(x) + \varepsilon_k \tag{6.9}$$

in this situation, and this yields the claim by boundedness of $\{u_k\}_{k\in\mathbb{N}}$ and monotonicity of $\{\theta_k\}_{k\in\mathbb{N}}$. If $\{\varepsilon_k\}_{k\in\mathbb{N}}$ is a null sequence, we obtain an even stronger result.

Theorem 6.15. *Let $\{x_k\}_{k\in\mathbb{N}} \subset X$ be a sequence generated by Algorithm 1 and let $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset [0, +\infty)$ be a null sequence such that x_{k+1} is a globally ε_k-minimal point of $\mathcal{L}_{\theta_k}(\cdot, u_k)_C$ for each $k \in \mathbb{N}$. Then each accumulation point $\bar{x} \in X$ of $\{x_k\}_{k\in\mathbb{N}}$ is a global minimizer of (Q) and, along the associated subsequence, we find $q(x_k) \to q(\bar{x})$.*

Proof. Without loss of generality, we assume $x_k \to \bar{x}$. By closedness of C, we have $\bar{x} \in C$. The estimate (6.9) yields

$$f(x_{k+1}) + q(x_{k+1}) + \frac{\theta_k}{2} \operatorname{dist}_K^2\left(G(x_{k+1}) + \frac{u_k}{\theta_k}\right) - \frac{\|u_k\|^2}{2\theta_k} \le f(x) + q(x) + \varepsilon_k \tag{6.10}$$

for each $x \in \mathcal{S}$. We show the statement of the theorem by distinguishing two cases.

In case where $\{\theta_k\}_{k\in\mathbb{N}}$ remains bounded, we find $\operatorname{dist}_K(G(x_{k+1})) \le V_{\theta_k}(x_{k+1}, u_k) \to 0$ from (S.4), so the continuity of the distance function dist_K and G yields $G(\bar{x}) \in K$, i.e., $\bar{x}$ is feasible to (Q). Using the triangle inequality, we also obtain

$$\operatorname{dist}_K(G(x_{k+1}) + u_k/\theta_k) \le \operatorname{dist}_K(G(x_{k+1})) + \|u_k\|/\theta_k \le V_{\theta_k}(x_{k+1}, u_k) + \|u_k\|/\theta_k$$

for each $k \in \mathbb{N}$. Squaring on both sides, exploiting the boundedness of $\{u_k\}_{k\in\mathbb{N}}$ and $V_{\theta_k}(x_{k+1}, u_k) \to 0$ yields

$$\limsup_{k\to+\infty} \left(\operatorname{dist}_K^2\left(G(x_{k+1}) + u_k/\theta_k\right) - (\|u_k\|/\theta_k)^2\right) \le 0.$$

The boundedness of $\{\theta_k\}_{k\in\mathbb{N}}$ and (6.10) thus show $\limsup_{k\to+\infty}(f(x_{k+1}) + q(x_{k+1})) \le f(x) + q(x)$ for each $x \in \mathcal{S}$. Exploiting the lower semicontinuity of q, this leads to $f(\bar{x}) + q(\bar{x}) \le f(x) + q(x)$, i.e., $\bar{x}$ is a global minimizer of (Q). On the other hand, we have

$$f(\bar{x}) + q(\bar{x}) \le \liminf_{k\to+\infty} (f(x_{k+1}) + q(x_{k+1})) \le \limsup_{k\to+\infty} (f(x_{k+1}) + q(x_{k+1})) \le f(\bar{x}) + q(\bar{x})$$

from the particular choice $x := \bar{x}$, so the continuity of f yields $q(x_k) \to q(\bar{x})$ as claimed.

Now, let us assume that $\{\theta_k\}_{k\in\mathbb{N}}$ is not bounded. Then we have $\theta_k \to +\infty$ from (S.4). By choice of x_{k+1}, we have $\mathcal{L}_{\theta_k}(x_{k+1}, u_k) \le \mathcal{L}_{\theta_k}(x, u_k) + \varepsilon_k$ for all $x \in C$ and each $k \in \mathbb{N}$, so the definition of the augmented Lagrangian function yields

$$f(x_{k+1})+q(x_{k+1})+\frac{\theta_k}{2}\operatorname{dist}_K^2\left(G(x_{k+1})+\frac{u_k}{\theta_k}\right) \le f(x)+q(x)+\frac{\theta_k}{2}\operatorname{dist}_K^2\left(G(x)+\frac{u_k}{\theta_k}\right)+\varepsilon_k$$

for each $x \in C$. By continuity of f and lower semicontinuity of q, $\{f(x_{k+1})+q(x_{k+1})\}_{k\in\mathbb{N}}$ is bounded from below. Thus, dividing the above estimate by θ_k and taking the limit inferior, we find

$$\begin{aligned}\operatorname{dist}_K^2(G(\bar{x})) &= \liminf_{k\to+\infty}\operatorname{dist}_K^2\left(G(x_{k+1})+u_k/\theta_k\right)\\ &\le \liminf_{k\to+\infty}\operatorname{dist}_K^2\left(G(x)+u_k/\theta_k\right) = \operatorname{dist}_K^2(G(x))\end{aligned}$$

for each $x \in C$ from $\theta_k \to +\infty$ and continuity of dist_K and G. Hence, $\bar{x}$ is a global minimizer of $\operatorname{dist}_K^2 \circ G$ over C. Since $\mathcal{S}$ is assumed to be nonempty, we infer $\operatorname{dist}_K^2(G(\bar{x})) = 0$, i.e., $\bar{x}$ is feasible to (Q). Exploiting boundedness of $\{u_k\}_{k\in\mathbb{N}}$, nonnegativity of the distance function, and $\theta_k \to +\infty$, we now obtain $\limsup_{k\to+\infty}(f(x_{k+1}) + q(x_{k+1})) \le f(x) + q(x)$ for each $x \in \mathcal{S}$ from (6.10). Proceeding as in the first case now yields the claim. ☐

It remains to clarify how the subproblems (6.6) can be solved in practice. If the non-Lipschitzness of q is, in some sense, structured while C is of simple form, it should be reasonable to solve (6.6) with the aid of a nonmonotone proximal gradient method, see [Chen et al., 2017, Section 3.1]. On the other hand, in situations where q is not present while C possesses a variational structure which allows for the efficient computation of projections, a nonmonotone spectral gradient method might be used to solve (6.6), see [Jia et al., 2021, Section 3]. Finally, it might be even possible to solve (6.6) up to global optimality in analytic way in some practically relevant applications where q is a standard sparsity-promoting term and the remaining data is simple enough.

Coming back to the assertion of Theorem 6.14, the following is now clear from Corollary 6.10.

Corollary 6.16. *Let $\{x_k\}_{k\in\mathbb{N}}$ be a sequence generated by Algorithm 1 such that x_{k+1} is a stationary point of $\mathcal{L}_{\theta_k}(\cdot, u_k)_C$ for each $k \in \mathbb{N}$. Assume that, along a subsequence (without relabeling), we have $x_k \to \bar{x}$ and $q(x_k) \to q(\bar{x})$ for some $\bar{x} \in X$ which is feasible to (Q) and satisfies the uniform qualification condition. Then $\bar{x}$ is M-stationary.*

Note that in the light of Proposition 6.12, Corollary 6.16 drastically generalizes and improves [Chen et al., 2017, Theorem 3.1] which shows global convergence of a related augmented Lagrangian method to certain stationary points under validity of a basic qualification, see condition (6.5), and the *relaxed constant positive linear dependence constraint qualification* which is more restrictive than condition (6.3) in the investigated setting, see [Jia et al., 2021, Lemma 2.7] as well. Let us mention that such a result has been foreshadowed in [Jia et al., 2021, Section 5.4]. We would like to point the reader's attention to the fact that working with strong accumulation points in the context of Theorems 6.14 and 6.15 and Corollary 6.16 is indispensable as long as q or the sets K and C are not convex since the limiting variational tools rely on strong convergence in the primal space. In the absence of q and if K and C are convex, some convergence results based on weak accumulation points are available, see e.g. [Börgens et al., 2020, Section 7] and Börgens et al. [2019], Kanzow et al. [2018]. Clearly, in finite dimensions, both types of convergence are equivalent and the consideration of strong accumulation points is not restrictive at all.

6.3 Sparsity-promotion in optimal control

In this section, we apply the theory derived earlier to an optimal control problem with a sparsity-promoting term in the objective function. As it is common to denote control functions by u in the context of optimal control, we will use the same notation here for the decision variable for notational convenience.

For some bounded domain $D \subset \mathbb{R}^d$ and some $p \in (0,1)$, we define a function $q\colon L^2(D) \to \mathbb{R}$ by means of

$$\forall u \in L^2(D)\colon \quad q(u) := \int_D |u(\omega)|^p \,\mathrm{d}\omega. \tag{6.11}$$

Above, $L^2(D)$ denotes the standard Lebesgue space of (equivalence classes of) measurable functions whose square is integrable and is equipped with the usual norm. In optimal control, the function q is used as an additive term in the objective function in order to promote sparsity of underlying control functions, see Ito and Kunisch [2014], Natemeyer and Wachsmuth [2021], Wachsmuth [2019]. A reason for this behavior is that the integrand $t \mapsto |t|^p$ possesses a unique global minimizer and infinite growth at the origin. In Mehlitz and Wachsmuth [2021], the authors explore the variational properties of the functional q. It has been shown to be uniformly continuous in [Mehlitz and Wachsmuth, 2021, Lemma 2.3]. Furthermore, in [Mehlitz and Wachsmuth, 2021, Theorem 4.6], the following formula has been proven for each $\bar u \in L^2(D)$:

$$\overline{\partial} q(\bar u) = \partial q(\bar u) = \bigl\{\eta \in L^2(D) \,|\, \eta = p\,|\bar u|^{p-2}\,\bar u \text{ a.e. on } \{\bar u \neq 0\}\bigr\}. \tag{6.12}$$

Let us emphasize that this means that the Fréchet and limiting subdifferential actually coincide and can be empty if the reference point is a function which tends to zero too fast somewhere on its domain. This underlines the sparsity-promoting properties of q.

Now, for a continuously differentiable function $f\colon L^2(D) \to \mathbb{R}$ and functions $u_a, u_b \in L^2(D)$ satisfying $u_a < 0 < u_b$ almost everywhere on D, we consider the optimization problem

$$\min_u \{f(u) + q(u) \,|\, u \in C\} \tag{OC}$$

where $C \subset L^2(D)$ is given by the box

$$C := \{u \in L^2(D) \,|\, u_a \le u \le u_b \text{ a.e. on } D\}.$$

For later use, let us mention that, for each $u \in C$, the (Fréchet) normal cone to C at u is given by the pointwise representation

$$N_C(u) = \left\{ \eta \in L^2(D) \;\middle|\; \begin{array}{ll} \eta \leq 0 & \text{a.e. on } \{u < u_b\} \\ \eta \geq 0 & \text{a.e. on } \{u_a < u\} \end{array} \right\}. \tag{6.13}$$

Typically, in optimal control, f is a function of type

$$\forall u \in L^2(D)\colon \quad f(u) := \tfrac{1}{2} \|S(u) - y_\mathrm{d}\|^2 + \tfrac{\sigma}{2} \|u\|^2 \tag{6.14}$$

where $S\colon L^2(D) \to H$ is the continuously differentiable control-to-observation operator associated with a given system of differential equations, H is a Hilbert space, $y_\mathrm{d} \in H$ is the desired state, and $\sigma \geq 0$ is a regularization parameter. Clearly, by means of the chain rule, f is continuously differentiable with derivative given by

$$\forall u \in L^2(D)\colon \quad f'(u) = S'(u)^*[S(u) - y_\mathrm{d}] + \sigma u.$$

The presence of q in the objective functional of (OC) enforces sparsity of its solutions, i.e., the support of optimal controls is likely to be small. It already has been mentioned in Ito and Kunisch [2014], Natemeyer and Wachsmuth [2021] that one generally cannot show existence of solutions to optimization problems of type (OC). Nevertheless, the practical need for sparse controls makes it attractive to consider the model and to derive necessary optimality conditions in order to identify reasonable stationary points.

In the subsequent lemma, we show that the feasible points of (OC) satisfy the uniform qualification condition stated in Definition 6.8.

Lemma 6.17. *Let $\bar u \in L^2(D)$ be a feasible point of (OC). Then the uniform qualification condition holds at $\bar u$.*

Proof. Recalling that q is continuous while C is convex, the uniform qualification condition takes the simplified form

$$\limsup_{u \to \bar u,\, u' \to \bar u} \left(\partial q(u) + N_C(u') \right) \subset \overline{\partial} q(\bar u) + N_C(\bar u).$$

Let us fix some point $\eta \in \limsup_{u\to\bar u,\, u'\to\bar u} \left(\partial q(u) + N_C(u')\right)$. Then we find sequences $\{u_k\}_{k\in\mathbb{N}}, \{u'_k\}_{k\in\mathbb{N}}, \{\eta_k\}_{k\in\mathbb{N}} \subset L^2(D)$ such that $u_k \to \bar u$, $u'_k \to \bar u$, $\eta_k \to \eta$, as well as $\eta_k \in \partial q(u_k) + N_C(u'_k)$ for all $k \in \mathbb{N}$. Particularly, there are sequences $\{\xi_k\}_{k\in\mathbb{N}}, \{\mu_k\}_{k\in\mathbb{N}} \subset L^2(D)$ such that $\xi_k \in \partial q(u_k)$, $\mu_k \in N_C(u'_k)$, and $\eta_k = \xi_k + \mu_k$ for all $k \in \mathbb{N}$. From (6.12) we find $\xi_k = p\,|u_k|^{p-2}\,u_k$ almost everywhere on $\{u_k \neq 0\}$ for each $k \in \mathbb{N}$. Furthermore, we have $\mu_k \leq 0$ almost everywhere on $\{u'_k = u_a\}$, $\mu_k \geq 0$ almost everywhere on $\{u'_k = u_b\}$, and $\mu_k = 0$ almost everywhere on $\{u_a < u'_k < u_b\}$ for each $k \in \mathbb{N}$ from (6.13). Along a subsequence (without relabeling) we can ensure the convergences $u_k(\omega) \to \bar u(\omega)$, $u'_k(\omega) \to \bar u(\omega)$, and $\eta_k(\omega) \to \eta(\omega)$ for almost every $\omega \in D$. Thus, for almost every $\omega \in \{\bar u = u_a\}$, we can guarantee $u_k(\omega) < 0$ and $u'_k(\omega) \in [u_a(\omega), 0)$, i.e., $\eta_k(\omega) = \xi_k(\omega) + \mu_k(\omega) \leq p|u_k(\omega)|^{p-2}u_k(\omega)$ for all large enough $k \in \mathbb{N}$, so, taking the limit yields $\eta(\omega) \leq p\,|\bar u(\omega)|^{p-2}\,\bar u(\omega)$. Similarly, we find $\eta(\omega) \geq p\,|\bar u(\omega)|^{p-2}\,\bar u(\omega)$ for almost every

$\omega \in \{\bar{u} = u_b\}$. Finally, for almost every $\omega \in \{\bar{u} \neq 0\} \cap \{u_a < \bar{u} < u_b\}$, we have $u_k(\omega) \neq 0$ and $u_a(\omega) < u_k'(\omega) < u_b(\omega)$, i.e., $\eta_k(\omega) = p\,|u_k(\omega)|^{p-2}\, u_k(\omega)$ for large enough $k \in \mathbb{N}$, so taking the limit, we have $\eta(\omega) = p\,|\bar{u}(\omega)|^{p-2}\, \bar{u}(\omega)$. Again, from (6.12) and (6.13), we have $\eta \in \overline{\partial} q(\bar{u}) + N_C(\bar{u})$, and this yields the claim. $\square$

Recalling that q is uniformly continuous, the subsequent result now directly follows from Corollary 6.10, the above lemma, and formulas (6.12) as well as (6.13).

Theorem 6.18. *Let $\bar{u} \in L^2(D)$ be a local minimizer of (OC). Then there exists a function $\eta \in L^2(D)$ such that*

$$f'(\bar{u}) + \eta = 0, \tag{6.15a}$$

$$\eta = p|\bar{u}|^{p-2}\bar{u} \quad \text{a.e. on } \{\bar{u} \neq 0\} \cap \{u_a < \bar{u} < u_b\}, \tag{6.15b}$$

$$\eta \leq p\,|u_a|^{p-2}\, u_a \quad \text{a.e. on } \{\bar{u} = u_a\}, \tag{6.15c}$$

$$\eta \geq p\,|u_b|^{p-2}\, u_b \quad \text{a.e. on } \{\bar{u} = u_b\}. \tag{6.15d}$$

We note that our approach to obtain necessary optimality conditions for (OC) is much different from the one used in Ito and Kunisch [2014], Natemeyer and Wachsmuth [2021] where *Pontryagin's maximum principle* has been used to derive pointwise conditions characterizing local minimizers under more restrictive assumptions than we needed to proceed. On the one hand, this led to optimaility conditions which also provide information on the subset of D where the locally optimal control is zero, and one can easily see that this is not the case in Theorem 6.18. On the other hand, a detailed inspection of (6.15) makes clear that our necessary optimality conditions provide helpful information regarding the structure of the optimal control as the multiplier η possesses L^2-regularity while (6.15b) causes η to possess singularities as the optimal control tends to zero somewhere on the domain. Thus, this condition clearly promotes sparse controls which either are zero, tend to zero (if at all) slowly enough, or are bounded away from it. Note that this differs from the conditions derived in Ito and Kunisch [2014], Natemeyer and Wachsmuth [2021] which are multiplier-free.

7 Concluding remarks

In this paper, we established a theory on approximate stationarity conditions for optimization problems with potentially non-Lipschitzian objective functions in a very general setting. In contrast to the finite-dimensional situation, where approximate stationarity has been shown to serve as a necessary optimality condition for local optimality without any additional assumptions, some additional semicontinuity properties need to be present in the infinite-dimensional context. We exploited our findings in order to re-address the classical topic of set extremality and were in position to derive a novel version of the popular extremal principle. This may serve as a starting point for further research which compares the classical as well as the new version of the extremal principle in a more detailed way. Moreover, we used our results in order to derive an approximate notion of stationarity as well as an associated qualification condition related to M-stationarity for optimization problems with a composite objective function and geometric constraints in the Banach space setting. This theory then has been applied to study the convergence properties of an associated augmented Lagrangian method for the numerical solution of such problems. Furthermore, we demonstrated how these findings can be used to derive necessary optimality conditions for optimal control problems with control constraints and a sparsity-promoting term in the objective function. Some future research may clarify whether our approximate stationarity conditions can be used to find necessary optimality conditions for optimization problems in function spaces where nonconvexity or nonsmoothness pop up in a different context. Exemplary, it would be interesting to study situations where the solution operator S appearing in (6.14) is nonsmooth, see e.g. Christof et al. [2018], Hintermüller et al. [2014], Rauls and Wachsmuth [2020], where the set of feasible controls is nonconvex, see e.g. Clason et al. [2017, 2020], Mehlitz and Wachsmuth [2018], or where the function q is a term promoting sharp edges in continuous image denoising or deconvolution, see e.g. [Bredies and Lorenz, 2018, Section 6].

Acknowledgments

The authors are grateful to Hoa Bui who suggested Example 5.4.

This work is supported by the Australian Research Council, project DP160100854, and the DFG Grant *Bilevel Optimal Control: Theory, Algorithms, and Applications* (Grant No. WA 3636/4-2) within the Priority Program SPP 1962 (Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization). The first author benefited from the support of the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska–Curie Grant Agreement No. 823731 CONMECH, and Conicyt REDES program 180032.

References

R. Andreani, J. M. Martínez, and B. F. Svaiter. A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal on Optimization*, 20(6):3533–3554, 2010. doi: 10.1137/090777189.

R. Andreani, G. Haeser, and J. M. Martínez. On sequential optimality conditions for smooth constrained optimization. *Optimization*, 60(5):627–641, 2011. doi: 10.1080/02331930903578700.

R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva. A cone-continuity constraint qualification and algorithmic consequences. *SIAM Journal on Optimization*, 26(1): 96–110, 2016. doi: 10.1137/15M1008488.

R. Andreani, N. S. Fazzio, M. L. Schuverdt, and L. D. Secchin. A sequential optimality condition related to the quasi-normality constraint qualification and its algorithmic consequences. *SIAM Journal on Optimization*, 29(1):743–766, 2019a. doi: 10.1137/17M1147330.

R. Andreani, G. Haeser, L. D. Secchin, and P. J. S. Silva. New sequential optimality conditions for mathematical programs with complementarity constraints and algorithmic consequences. *SIAM Journal on Optimization*, 29(4):3201–3230, 2019b. doi: 10.1137/18M121040X.

R. Andreani, G. Haeser, and D. S. Viana. Optimality conditions and global convergence for nonlinear semidefinite programming. *Mathematical Programming*, 180(1):203–235, 2020. doi: 10.1007/s10107-018-1354-5.

J.-P. Aubin and H. Frankowska. *Set-Valued Analysis*. Birkhäuser, Boston, 2009.

K. Bai, J. J. Ye, and J. Zhang. Directional quasi-/pseudo-normality as sufficient conditions for metric subregularity. *SIAM Journal on Optimization*, 29(4):2625–2649, 2019. doi: 10.1137/18M1232498.

E. G. Birgin and J. M. Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization*. SIAM, Philadelphia, 2014.

J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer, New York, 2000.

E. Börgens, C. Kanzow, and D. Steck. Local and global analysis of multiplier methods for constrained optimization in Banach spaces. *SIAM Journal on Control and Optimization*, 57(6):3694–3722, 2019. doi: 10.1137/19M1240186.

E. Börgens, C. Kanzow, P. Mehlitz, and G. Wachsmuth. New constraint qualifications for optimization problems in Banach spaces based on asymptotic KKT conditions. *SIAM Journal on Optimization*, 30(4):2956–2982, 2020. doi: 10.1137/19M1306804.

K. Bredies and D. Lorenz. *Mathematical Image Processing*. Birkhäuser, Cham, 2018.

X. Chen, L. Guo, Z. Lu, and J. J. Ye. An augmented Lagrangian method for non-Lipschitz nonconvex programming. *SIAM Journal on Numerical Analysis*, 55(1):168–193, 2017. doi: 10.1137/15M1052834.

C. Christof, C. Meyer, S. Walther, and C. Clason. Optimal control of a non-smooth semilinear elliptic equation. *Mathematical Control & Related Fields*, 8(1):247–276, 2018. doi: 10.3934/mcrf.2018011.

C. Clason, A. Rund, and K. Kunisch. Nonconvex penalization of switching control of partial differential equations. *Systems & Control Letters*, 106:1–8, 2017. doi: 10.1016/j.sysconle.2017.05.006.

C. Clason, Y. Deng, P. Mehlitz, and U. Prüfert. Optimal control problems with control complementarity constraints: existence results, optimality conditions, and a penalty method. *Optimization Methods and Software*, 35(1):142–170, 2020. doi: 10.1080/10556788.2019.1604705.

A. L. Dontchev and R. T. Rockafellar. *Implicit Functions and Solution Mappings*. Springer, New York, 2014.

A. L. Dontchev, H. Gfrerer, A. Y. Kruger, and J. V. Outrata. The radius of metric subregularity. *Set-Valued and Variational Analysis*, 28(3):451–472, 2020. doi: 10.1007/s11228-019-00523-2.

I. Ekeland. On the variational principle. *Journal of Mathematical Analysis and Applications*, 47:324–353, 1974. doi: 10.1016/0022-247X(74)90025-0.

M. J. Fabian. Subdifferentiability and trustworthiness in the light of a new variational principle of Borwein and Preiss. *Acta Universitatis Carolinae*, 30(2):51–56, 1989. URL http://dml.cz/dmlcz/701793.

M. J. Fabian, R. Henrion, A. Y. Kruger, and J. V. Outrata. Error bounds: necessary and sufficient conditions. *Set-Valued and Variational Analysis*, 18(2):121–149, 2010. doi: 10.1007/s11228-010-0133-0.

M. L. Flegel, C. Kanzow, and J. V. Outrata. Optimality conditions for disjunctive programs with application to mathematical programs with equilibrium constraints. *Set-Valued Analysis*, 15(2):139–162, 2007. doi: 10.1007/s11228-006-0033-5.

H. Gfrerer. On directional metric regularity, subregularity and optimality conditions for nonsmooth mathematical programs. *Set-Valued and Variational Analysis*, 21(2): 151–176, 2013. doi: 10.1007/s11228-012-0220-5.

L. Guo and J. J. Ye. Necessary optimality conditions and exact penalization for non-Lipschitz nonlinear programs. *Mathematical Programming*, 168:571–598, 2018. doi: 10.1007/s10107-017-1112-0.

L. Guo, G.-H. Lin, and J. J. Ye. Solving mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 166(1):234–256, 2015. doi: 10.1007/s10957-014-0699-z.

F. Harder and G. Wachsmuth. Comparison of optimality systems for the optimal control of the obstacle problem. *GAMM-Mitteilungen*, 40(4):312–338, 2018. doi: 10.1002/gamm.201740004.

F. Harder, P. Mehlitz, and G. Wachsmuth. Reformulation of the M-stationarity conditions as a system of discontinuous equations and its solution by a semismooth Newton method. *SIAM Journal on Optimization*, 31(2):1459–1488, 2021. doi: 10.1137/20M1321413.

E. S. Helou, S. A. Santos, and L. E. A. Simões. A new sequential optimality condition for constrained nonsmooth optimization. *SIAM Journal on Optimization*, 30(2):1610–1637, 2020. doi: 10.1137/18M1228608.

M. Hintermüller, B. S. Mordukhovich, and T. M. Surowiec. Several approaches for the derivation of stationarity conditions for elliptic MPECs with upper-level control constraints. *Mathematical Programming*, 146:555–582, 2014. doi: 10.1007/s10107-013-0704-6.

A. D. Ioffe. *Variational Analysis of Regular Mappings. Theory and Applications.* Springer, Cham, 2017.

K. Ito and K. Kunisch. Optimal control with $L^p(\Omega)$, $p \in [0, 1)$, control cost. *SIAM Journal on Control and Optimization*, 52(2):1251–1275, 2014. doi: 10.1137/120896529.

X. Jia, C. Kanzow, P. Mehlitz, and G. Wachsmuth. An augmented Lagrangian method for optimization problems with structured geometric constraints. Technical report, preprint arXiv, 2021. URL `https://arxiv.org/abs/2105.08317`.

C. Kanzow and D. Steck. An example comparing the standard and safeguarded augmented Lagrangian methods. *Operations Research Letters*, 45(6):598–603, 2017. doi: 10.1016/j.orl.2017.09.005.

C. Kanzow, D. Steck, and D. Wachsmuth. An augmented Lagrangian method for optimization problems in Banach spaces. *SIAM Journal on Optimization*, 56(1):272–291, 2018. doi: 10.1137/16M1107103.

C. Kanzow, A. B. Raharja, and A. Schwartz. Sequential optimality conditions for cardinality-constrained optimization problems with applications. *Computational Optimization and Applications*, 80:185–211, 2021. doi: 10.1007/s10589-021-00298-z.

D. Klatte and B. Kummer. *Nonsmooth Equations in Optimization. Regularity, Calculus, Methods and Applications.* Kluwer Academic, Dordrecht, 2002.

A. Y. Kruger. Generalized differentials of nonsmooth functions and necessary conditions for an extremum. *Sibirian Mathematical Journal*, 26:370–379, 1985.

A. Y. Kruger. On Fréchet subdifferentials. *Journal of Mathematical Sciences (New York)*, 116(3):3325–3358, 2003. doi: 10.1023/A:1023673105317.

A. Y. Kruger. About stationarity and regularity in variational analysis. *Taiwanese Journal of Mathematics*, 13(6A):1737–1785, 2009. doi: 10.11650/twjm/1500405612.

A. Y. Kruger. Error bounds and metric subregularity. *Optimization*, 64(1):49–79, 2015. doi: 10.1080/02331934.2014.938074.

A. Y. Kruger and B. S. Mordukhovich. Extremal points and the Euler equation in nonsmooth optimization problems. *Doklady Akademii Nauk BSSR*, 24(8):684–687, 1980.

M. Maréchal. Metric subregularity in generalized equations. *Journal of Optimization Theory and Applications*, 176:541–558, 2018. doi: 10.1007/s10957-018-1246-0.

P. Mehlitz. On the sequential normal compactness condition and its restrictiveness in selected function spaces. *Set-Valued and Variational Analysis*, 27(3):763–782, 2019. doi: 10.1007/s11228-018-0475-6.

P. Mehlitz. Asymptotic stationarity and regularity for nonsmooth optimization problems. *Journal of Nonsmooth Analysis and Optimization*, 1:6575, 2020. doi: 10.46298/jnsao-2020-6575.

P. Mehlitz. Asymptotic regularity for Lipschitzian nonlinear optimization problems with applications to complementarity-constrained and bilevel programming. Technical report, 2021. URL `https://arxiv.org/abs/2105.01985`. preprint arXiv.

P. Mehlitz and G. Wachsmuth. The limiting normal cone to pointwise defined sets in Lebesgue spaces. *Set-Valued and Variational Analysis*, 26(3):449–467, 2018. doi: 10.1007/s11228-016-0393-4.

P. Mehlitz and G. Wachsmuth. Subdifferentiation of nonconvex sparsity-promoting functionals on Lebesgue spaces. Technical report, 2021. URL `https://arxiv.org/abs/2107.09340`. preprint arXiv.

B. S. Mordukhovich. *Variational Analysis and Generalized Differentiation, Part I: Basic Theory, Part II: Applications*. Springer, Berlin, 2006.

C. Natemeyer and D. Wachsmuth. A proximal gradient method for control problems with nonsmooth and nonconvex control cost. *Computational Optimization and Applications*, 80:639–677, 2021. doi: 10.1007/s10589-021-00308-0.

J.-P. Penot. *Calculus Without Derivatives*. Springer, New York, 2013.

R. R. Phelps. *Convex Functions, Monotone Operators and Differentiability*. Springer, Berlin, 1993.

A. Ramos. Mathematical programs with equilibrium constraints: a sequential optimality condition, new constraint qualifications and algorithmic consequences. *Optimization Methods and Software*, 36(1):45–81, 2021. doi: 10.1080/10556788.2019.1702661.

A.-T. Rauls and G. Wachsmuth. Generalized derivatives for the solution operator of the obstacle problem. *Set-Valued and Variational Analysis*, 28:259–285, 2020. doi: 10.1007/s11228-019-0506-y.

R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer, Berlin, 1998.

L. Thibault. Sequential convex subdifferential calculus and sequential Lagrange multipliers. *SIAM Journal on Control and Optimization*, 35(4):1434–1444, 1997. doi: 10.1137/S0363012995287714.

D. Wachsmuth. Iterative hard-thresholding applied to optimal control problems with $L^0(\Omega)$ control cost. *SIAM Journal on Control and Optimization*, 57(2):854–879, 2019. doi: 10.1137/18M1194602.

C. Zălinescu. *Convex Analysis in General Vector Spaces*. World Scientific, River Edge, 2002.

X. Y. Zheng and K. F. Ng. Metric subregularity and calmness for nonconvex generalized equations in Banach spaces. *SIAM Journal on Optimization*, 20(5):2119–2136, 2010. doi: 10.1137/090772174.