1-1-2011

New variational principles with applications to optimization theory and algorithms

Hung Minh Phan
Wayne State University,

NEW VARIATIONAL PRINCIPLES WITH APPLICATIONS TO OPTIMIZATION THEORY AND ALGORITHMS

by

HUNG MINH PHAN

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2011

MAJOR: MATHEMATICS

Approved by:

Advisor Date

DEDICATION

To my parents

ACKNOWLEDGEMENTS

It is a pleasure to express my regards and blessings to all of those who supported me in any respect during the completion of this dissertation.

I could not find enough words to express my deepest gratitude to my advisor, Prof. Boris Mordukhovich, who introduces me to the beautiful world of variational analysis, proposes me exciting problems, teaches me lots of attractive topics in mathematics, and constantly helps as well as encourages me for all 5 years I study at Wayne Sate University. This dissertation would not have been completed without his guidance and endless support.

I am taking this opportunity to thank Prof. George Yin, Prof. Alexander Korostelev, Prof. Tze Chien Sun and Prof. Alper Murat for serving in my committee.

I owe my thanks to Prof. Marco López for his many significant suggestions in Section 3.7; to Prof. Christian Kanzow and Dr. Tim Hoheisel for their very important contribution to Chapter 5. I am grateful to Prof. Phan Quoc Khanh, Prof. Nguyen Dinh and Dr. Lam Quoc Anh, who recommended me to Wayne State University and helped me a lot when I worked in Vietnam; to Dr. Nguyen Mau Nam and Dr. Truong Quang Bao for their special help and support during these years.

I would like to share this moment with my family. I am indebted to my parents, my sister and grandmothers for their unconditional and unlimited love and support since I was born. My special gratitude goes to my wife for her love and encouragement.

Lastly, I would like to express my appreciation to Ms. Mary Klamo, Ms. Patricia Bonesteel, Prof. Daniel Frohardt, Prof. Choon-Jai Rhee, Prof. John Breckenridge, Ms. Tiana Fluker; and to the entire Department of Mathematics, who have made available their support in a number of ways.

TABLE OF CONTENTS

Chapter 1
Introduction

1.1 Overview

Optimization has been a major motivation and a driving force for developing differential and integral calculus. Indeed, solving an optimization problem leads to introducing the notion of *derivative* and the invention of the *Fermat stationary principle*, which is credited to the famous mathematician Pierre de Fermat. Briefly, the stationary principle states that:

"if a function $f(x)$ attains its local maximum/minimum at a point $\bar{x}$ then $f'(\bar{x}) = 0$".

This fundamental principle led to the rising of *differential calculus*, which was developed systematically by Gottfried Leibnitz and Isaac Newton. It has been well recognized that differential calculus is a very powerful mathematical tool, which is efficiently used in various applications. However, the limitation of differential calculus amounts to the requirement of differentiability of the data, while *nonsmooth* structures arise naturally and frequently in many mathematical models. Nonsmooth analysis refers to the study of generalized differential properties of sets, functions, and set-valued mappings without smoothness (differentiability) assumptions on their initial data. The term "*variational analysis*" is usually used to indicate a broader area based on variational principles, which includes nonsmooth and set-valued analysis, optimization, equilibrium, control, stability and sensitivity with respect to data perturbations, etc.

In searching for solutions to optimization problems, it would be reasonable, though not trivial, to think of some general "extremality" in the behavior of the objective functions and the constraints. This encourages the use of convex separation as well as the development of extremal principles. Let us mention that although convex separation is a well known principle in mathematics, it has not played a pivotal role in optimization theory until being employed by Jean-Jacques Moreau and R. Tyrrell Rockafellar in their study of generalized differential properties of convex sets and convex functions in the 1960s. The main ingredient used is the derivative-like object for convex functions at reference points. This object is called *subdifferential*, which is defined to be a set of subgradients in contrast to classical derivatives/gradients which are singletons. This area, now known as convex analysis, plays a fundamental role in mathematics and applied science. The next breakthrough development was the definition of the subdifferential for locally Lipschitz functions, given by Francis Clarke in his Ph.D dissertation in 1973 under the supervision of Rockafellar. Clarke's subdifferential is always a convex set, and convexity still plays a crucial role in this area. However, the beauty of convexity inevitably faces challenges in many applications. In the mid 1970s, Boris Mordukhovich made a further significant contribution to Variational Analysis by developing a *dual* approach, where the convexity limitations were avoided. Basically, the "extremality" idea mentioned above was successfully involved in nonconvex structures. This eventually led to the invention of *extremal principles*, which furnished the calculus of the Mordukhovich/limiting/basic subdifferential. Since the latter subdifferential is generally nonconvex and always smaller than Clarke's subdifferential, it provides a sharper and more efficient tool for the analysis and applications. It is important to note that practical problems in, e.g., engineering and economics, always demand more and more effective tools to reduce cost and increase productivity. Besides the theoretical beauty and various applications of generalized differentiation to optimization and control theory, it is also important to develop numerical algorithms of nonsmooth optimization, e.g., subgradient and Newton-type methods, which are among the most efficient tools in applications. It has been widely recognized that Newton's method and its extensions play a vital role in solving systems of nonlinear equations and computing solutions to a variety of practical problems.

Great progress has been made in the developments and applications of variational analysis. Basics of variational analysis in finite dimensional spaces can be found in the book "*Variational Analysis*" by Rockafellar and Wets [61]. Further developments and applications including infinite dimensional problems are thoroughly studied in the book "*Techniques of Variational Analysis*" by Borwein and Zhu [7], and in the two volume monograph "*Variational Analysis and Generalized Differentiations*" by Mordukhovich [47, 48]. The development of many aspects of variational analysis and its applications to optimal control, variational inequalities, etc, can be found in the excellent books by Aubin and Frankowska [3], Clarke [12], Clarke et al. [13], Facchinei and Pang [21], Schirotzek [62], Vinter [68], and the references therein.

1.2 About this dissertation

Besides Chapter 2 which is reserved for preliminary concepts, the main topics of this dissertation has two parts. The first part concerns some extensions of the aforementioned extremal principles. The second part develops some numerical algorithms of Newton-type.

In the first part, the major motivation for our work is to develop and apply extremal principles of variational analysis the first version of which was formulated in [35] for finitely many sets via ε-normals, which is defined by (2.3) in the next chapter; see [47, Chapter 2] for more details and discussions. Recall [47, Definition 2.5] that a set system $\{\Omega_1, \ldots, \Omega_m\}$, $m \geq 2$, satisfies the *approximate extremal principle* at $\bar{x} \in \cap_{i=1}^m \Omega_i$ if for every $\varepsilon > 0$ there are $x_i \in \Omega_i \cap (\bar{x} + \varepsilon I\!B)$ and $x_i^* \in \widehat{N}(x_i; \Omega_i) + \varepsilon I\!B^*$, $i = 1, \ldots, m$, such that

$$x_1^* + \ldots + x_m^* = 0 \quad \text{and} \quad \|x_1^*\|^2 + \ldots + \|x_m^*\|^2 = 1. \tag{1.1}$$

If the dual vectors x_i^* can be taken from the limiting normal cone $N(\bar{x}; \Omega_i)$ (2.5), then we say that the system $\{\Omega_1, \ldots, \Omega_m\}$ satisfies the *exact extremal principle* at $\bar{x}$.

Note that the known extremal principles do not involve any tangential approximations of sets in primal spaces and do not employ convex separation. This dual-space approach exhibits a number of significant advantages in comparison with convex separation techniques and opens new perspectives in variational analysis, generalized differentiation, and their numerous applications. On the other hand, we are not familiar with any versions of extremal principles in the scope of [47, 48] for infinite systems of sets; it is not even clear how to formulate them appropriately in the lines of the developed methodology. Among primary motivations for considering infinite systems of sets we mention problems of semi-infinite programming, especially those concerning the most difficult case of countably many constraints vs. conventional ones with compact indexes; cf. [24].

Efficient conditions ensuring the fulfillment of both approximate and exact versions of the extremal principle can be found in [47, Chapter 2] and the references therein. Roughly speaking, the approximate extremal principle in terms of Fréchet normals holds for locally extremal points of any closed subsets in Asplund spaces ([47, Theorem 2.20]) while the exact extremal principle requires additional sequential normal compactness assumptions that are automatic in finite dimensions; see [47, Theorem 2.22].

Recall [35, 47] that a point $\bar{x} \in \cap_{i=1}^m \Omega_i$ is *locally extremal* for the system $\{\Omega_1, \ldots, \Omega_m\}$ if there are sequences $\{a_{ik}\} \subset X$, $i = 1, \ldots, m$, and a neighborhood U of $\bar{x}$ such that $a_{ik} \to 0$ as $k \to \infty$ and

$$\bigcap_{i=1}^{m} \Big(\Omega_i - a_{ik}\Big) \cap U = \emptyset \;\text{ for all large }\; k \in \mathbb{N}. \tag{1.2}$$

As shown in [47], this extremality notion for sets encompasses standard notions of local optimality for various optimization-related and equilibrium problems as well as for set systems arising in proving calculus rules and other frameworks of variational analysis.

In Chapter 3, we propose and justify extremal principles of a new type, which can be applied to infinite set systems while also provide independent results for finitely many nonconvex sets. To achieve this goal, we develop a novel approach that incorporates and unifies some ideas from both tangential approximations of sets in primal spaces and nonconvex normal cone approximations in dual spaces. The essence of this approach is as follows. Employing a variational technique, we first derive a new *conic extremal principle*, which concerns countable systems of general *nonconvex* cones in finite dimensions and describes their extremality at the origin via an appropriate countable version of the generalized Euler equation formulated in terms of the nonconvex limiting normal cone in [45]. Then we introduce a notion of *tangential extremal points* for infinite (in particular, finite) systems of closed sets involving their tangential approximations. The corresponding *tangential extremal principles* are induced in this way by applying the conic extremal principle to the collection of selected tangential approximations. The major attention is paid to the case of tangential approximations generated by the (nonconvex) Bouligand-Severi *contingent cone*, which exhibits remarkable properties that are most appropriate for implementing the proposed scheme and subsequent applications. The contingent cone is replaced by its *weak* counterpart when the space in question is infinite-dimensional. Selected applications of the developed theory to problems of semi-infinite programming and multiobjective optimization are also given.

At the same time, the above tangential extremal principles concern the so-called tangential extremality (and only in finite dimensions) and do not reduce to the conventional extremal principles of [47] for finite systems of sets even in simple frameworks. In Chapter 4, we develop new *rated extremal principles* for both finite and infinite systems of closed sets in finite-dimensional and infinite-dimensional spaces. Besides being applied to conventional local extremal points of finite set systems and reducing to the known results for them, the rated extremal principles provide enhanced information in the case of finitely many sets while open new lines of development for countable set systems. The results obtained in this way allow us, in particular, to derive intersection rules for generalized normals of infinite intersections of closed sets, which imply in turn new necessary optimality conditions for mathematical programs with countable constraints in finite and infinite dimensions.

The second part studies some generalized algorithms of Newton-type. It is well-recognized that Newton's method is one of the most powerful and useful methods in optimization and in the related area of solving systems of nonlinear equations

$$H(x) = 0 \tag{1.3}$$

defined by continuous vector-valued mappings $H : \mathbb{R}^n \to \mathbb{R}^n$. In the classical setting when H is a continuously differentiable (smooth, C^1) mapping, Newton's method builds the following iteration procedure

$$x^{k+1} := x^k + d^k \quad \text{for all} \quad k = 0, 1, 2, \ldots, \tag{1.4}$$

where $x^0 \in \mathbb{R}^n$ is a given starting point, and where $d^k \in \mathbb{R}^n$ is a solution to the linear system of equations (often called "Newton equation")

$$H'(x^k)d = -H(x^k). \tag{1.5}$$

A detailed analysis and numerous applications of the classical Newton's method (1.4), (1.5) and its modifications can be found, e.g., in the books [15, 32, 51] and the references therein.

However, in the vast majority of applications, including those to optimization, variational inequalities, complementarity and equilibrium problems, etc, the underlying mapping H in (1.3) is nonsmooth. Indeed, the aforementioned optimization-related models can be written via Robinson's formalism of "generalized equations," which in turn can be reduced to standard equations of the form above (using, e.g., the projection operator) while with *intrinsically nonsmooth* mappings H; see [21, 43, 59, 54] for more details, discussions, and references.

Robinson originally proposed (see [58] and also [60] based on his earlier preprint) a point-based approximation approach to solve nonsmooth equations (1.3), which then was developed by his student Josephy [29] to extend Newton's method for solving variational inequalities and complementarity problems. Other approaches replace the classical derivative $H'(x_k)$ in the Newton equation (1.5) by some generalized derivatives. In particular, the B-differentiable Newton's method developed by Pang [52, 53] uses the iteration scheme (1.4) with d^k being a solution to the subproblem

$$H'(x^k; d) = -H(x^k), \tag{1.6}$$

where $H'(x^k; d)$ denotes the classical directional derivative of H at x^k in the direction d. Besides the existence of the classical directional derivative in (1.6), a number of strong assumptions are imposed in [52, 53] to establish appropriate convergence results.

In another approach developed by Kummer [36] and Qi and Sun [57], the direction d^k in (1.4) is taken as a solution to the linear system of equations

$$A_k d = -H(x^k), \tag{1.7}$$

where A_k is an element of Clarke's generalized Jacobian $\partial_C H(x_k)$ of a Lipschitz continuous mapping H. In [56], Qi suggested to replace $A_k \in \partial_C H(x^k)$ in (1.7) by the choice of A_k from the so-called B-subdifferential $\partial_B H(x^k)$ of H at x^k, which is a proper subset of $\partial_C H(x^k)$; see Section 4 for more details. We also refer the reader to [21, 33, 60] and bibliographies therein for wide overviews, historical remarks, and other developments on Newton's method for nonsmooth Lipschitz equations as in (1.3) and to [31] for some recent applications.

It is proved in [57] and [56] that the Newton type method based on implementing the generalized Jacobian and B-subdifferential in (1.7), respectively, superlinearly converges to a solution of (1.3) for a class of *semismooth* mappings H; see Section 4 for the definition and discussions. This subclass of Lipschitz continuous and directionally differentiable mappings is rather broad and useful in applications to optimization-related problems. However, not every mapping arising in applications is either directionally differentiable or Lipschitz continuous. The reader can find valuable classes of functions and mappings of this type in [47, 61] and overwhelmingly in spectral function analysis, eigenvalue optimization, studying of roots of polynomials, stability of control systems, etc.; see, e.g., [8] and the references therein.

In Chapter 5, we propose a new Newton-type algorithm to solve nonsmooth equations (1.3) described by general continuous mappings H that is based on *graphical derivatives*. It reduces to the classical Newton's method (1.5) when H is smooth, being different from previously known versions of Newton's method in the case of Lipschitz continuous mappings H. Based on advanced tools of variational analysis involving *metric regularity* and *coderivatives*, we justify well-posedness of the new algorithm and its *superlinear local* and *global* (of the Kantorovich type) convergence under verifiable assumptions that hold for semismooth mappings but are *not* restricted to them. Detailed comparisons of our algorithm and results with the semismooth and B-differentiable Newton's methods are given and certain improvements are justified.

In Chapter 6, we follow the stream of ideas in Pang [52] to consider the so-called *merit function*

$$q(x) = \frac{1}{2}\|H(x)\|^2 = \frac{1}{2}H(x)^T H(x), \tag{1.8}$$

where $x^T y$ is the usual scalar product for $x, y \in \mathbb{R}^n$. One can easily observe that solving (1.3) is equivalent to solving the equation

$$q(x) = 0.$$

Since $q(x)$ is nonnegative, one idea to solve this equation is to find some iterations $\{x^k\}$ such that certain level of decrease is obtained, i.e., $q(x^k) > q(x^{k+1})$, $k \in \mathbb{N}$. To furnish, we replace the Newton iteration (1.4) by a *damped* iteration

$$x^{k+1} := x^k + \alpha_k d^k \ \text{ for all } \ k = 0, 1, 2, \ldots, \tag{1.9}$$

where d^k is a solution of Newton equation (1.5), and $\alpha_k \in (0, 1]$ is a suitable chosen scalar. This method also has an advantage that we can guarantee the *global* convergence of the iterations. Similar to Chapter 5, we present a damped Newton's method based on graphical derivatives. We also employ the advanced tools of variational analysis, the metric regularity and its coderivative criterion to justify the well-posedness and the local/global convergence of the proposed algorithm.

Note metric regularity and related concepts of variational analysis has been employed in the analysis and justification of numerical algorithms starting with Robinson's seminal contribution; see, e.g., [1, 38, 49] and their references for the recent account. However, we are not familiar with any usage of graphical derivatives and coderivatives for these purposes.

Chapter 2

Preliminary

In this chapter we briefly overview some basic tools of variational analysis and generalized differentiation that are widely used in what follows. Our notation is basically standard in variational analysis; see, e.g., [47, 61] for more details and references. Unless otherwise stated, the space X in question is Banach with the norm $\|\cdot\|$ and the canonical pairing $\langle\cdot,\cdot\rangle$ between X and its topological dual X^*. Recall that $B(\bar{x}, r)$ stands for a closed ball centered at $\bar{x}$ with radius $r > 0$, that $I\!B$ and $I\!B^*$ are the closed unit ball of the space in question and its dual, respectively, and that $\mathbb{N} := \{1, 2, \ldots\}$. The symbols $\stackrel{w}{\to}$ and $\stackrel{w^*}{\to}$ indicate the weak convergence in X and the weak* convergence in X^*, respectively. The notation $x \stackrel{\Omega}{\to} \bar{x}$ means that $x \to \bar{x}$ with $x \in \Omega$. Finally, $\mathbb{N} := \{1, 2, \ldots\}$ signifies the collection of all natural numbers.

Given a set-valued mapping $F\colon X \rightrightarrows Y$ between Banach spaces X and Y, we denote by

$$\mathop{\mathrm{Lim\,sup}}_{x\to\bar{x}} F(x) := \Big\{ y \in Y \Big| \begin{array}{l} \exists \text{ sequences } \ x_k \to \bar{x} \ \text{ and } \ y_k \to y \ \text{ as } \ k \to \infty \\ \text{such that } \ y_k \in F(x_k) \text{ for all } \ k \in \mathbb{N} \end{array} \Big\} \tag{2.1}$$

the *sequential Painlevé-Kuratowski outer limit* of F at $\bar{x}$. its When Y is the topological dual X^*, we use the weak* limit $y_k \stackrel{w^*}{\to} y$ by convention. Given $\emptyset \neq \Omega \subset X$, denote by

$$\mathrm{cone}\,\Omega := \bigcup_{\lambda\geq 0} \lambda\Omega = \bigcup_{\lambda \geq 0} \Big\{ \lambda v \Big| \ v \in \Omega \Big\}$$

the *conic hull* of Ω and by

$$\mathrm{co}\,\Omega := \Big\{ \sum_{i\in I} \lambda_i u_i \Big| \ I \ \text{ finite }, \ \lambda_i \geq 0, \ \sum_{i\in I} \lambda_i = 1, \ u_i \in \Omega \Big\}$$

the *convex hull* of this set.

2.1 Tangents and Normal to Nonconvex Sets

We recall some basic notions of tangent and normal cones to nonempty sets closed around the reference points. Given $\Omega \subset X$ and $\bar{x} \in \Omega$, the closed (while often nonconvex) cone

$$T(\bar{x};\Omega) := \operatorname*{Lim\,sup}_{t \downarrow 0} \frac{\Omega - \bar{x}}{t}, \tag{2.2}$$

which is also known as the *Bouligand-Severi tangent/contingent cone* to Ω at $\bar{x}$. When the "Lim sup" is taken with respect to the weak topology, we have the *weak contingent cone* to Ω at $\bar{x}$ denoted by $T_w(\bar{x};\Omega)$. For any $\varepsilon \geq 0$, the collection

$$\widehat{N}_\varepsilon(\bar{x};\Omega) := \left\{ x^* \in X^* \middle| \limsup_{x \xrightarrow{\Omega} \bar{x}} \frac{\langle x^*, x - \bar{x} \rangle}{\|x - \bar{x}\|} \leq \varepsilon \right\} \tag{2.3}$$

is called the set of *ε-normals* to Ω at $\bar{x}$. In the case of $\varepsilon = 0$ the set $\widehat{N}(\bar{x};\Omega) := \widehat{N}_0(\bar{x};\Omega)$ is a cone known as the *Fréchet/regular normal cone* (or the prenormal cone) to Ω at this point. Note that the Fréchet normal cone is always convex while it may be trivial (i.e., reduced to $\{0\}$) at boundary points of simple nonconvex sets in finite dimensions as for $\Omega = \{(x_1, x_2) \in \mathbb{R}^2 | \; x_2 \geq -|x_1|\}$ at $\bar{x} = (0,0)$. If the space X is reflexive, then

$$\widehat{N}(\bar{x};\Omega) = T_w^*(\bar{x};\Omega) := \big\{ x^* \in X^* \big| \; \langle x^*, v \rangle \leq 0, \; \forall v \in T_w(\bar{x};\Omega) \big\}. \tag{2.4}$$

The *Mordukhovich/basic/limiting normal cone* to Ω at a point $\bar{x} \in \Omega$ is defined by

$$N(\bar{x};\Omega) := \operatorname*{Lim\,sup}_{\substack{x \to \bar{x} \\ \varepsilon \downarrow 0}} \widehat{N}_\varepsilon(x;\Omega) \tag{2.5}$$

via the sequential outer limit Painlevé-Kuratowski outer limit (2.1) of ε-normals (2.3) as $x \to \bar{x}$ and $\varepsilon \downarrow 0$. If the space X is Asplund (i.e., each of its separable subspace has a separable dual that holds, in particular, when is reflexive) and the set Ω is locally closed around $\bar{x}$, we can equivalently put $\varepsilon_k = 0$ in (2.5); see [47] for more details. If $X = \mathbb{R}^n$, the basic normal cone (2.5) can be equivalently described as

$$N(\bar{x};\Omega) = \operatorname*{Lim\,sup}_{x\to\bar{x}} \Big\{\text{cone}\,\big[x - \Pi(x;\Omega)\big]\Big\} \tag{2.6}$$

via the Euclidian projector $\Pi(x;\Omega) := \{w \in \Omega|\ \|x - w\| = \text{dist}\,(x;\Omega)\}$ of $x \in \mathbb{R}^n$ onto Ω, which was the original definition in [45].

It is worth mentioning that the limiting normal cone (2.5) is often nonconvex as, e.g., for the set $\Omega \subset \mathbb{R}^2$ considered above, where $N(0;\Omega) = \{(u_1,u_2) \in \mathbb{R}^2|\ u_2 = -|u_1|\}$. It does not happen when Ω is *normally regular* at $\bar{x}$ in the sense that $N(\bar{x};\Omega) = \widehat{N}(\bar{x};\Omega)$. The latter class includes convex sets when both cones (2.3) as $\varepsilon = 0$ and (2.5) reduce to the classical normal cone of convex analysis and also some other collections of "nice" sets of a certain locally convex type. At the same time it excludes a number of important settings that frequently appear in applications; see, e.g., the books [47, 48, 61] for precise results and discussions. Being nonconvex, the normal cone $N(\bar{x};\Omega)$ in (2.5) cannot be tangentially generated by duality of type (2.4), since the duality/polarity operation automatically implies convexity. Nevertheless, in contrast to Fréchet normals, this limiting normal cone enjoys *full calculus* in general Asplund spaces, which is mainly based on extremal principles of variational analysis and related variational techniques; see [47] for a comprehensive calculus account and further references.

The next simple observation is useful in what follows.

Proposition 2.1 (generalized normals to cones). *Let $\Lambda \subset X$ be a cone, and let $w \in \Lambda$.*

Then we have the inclusion

$$\widehat{N}(w;\Lambda) \subset N(0;\Lambda).$$

Proof. Pick any $x^* \in \widehat{N}(w;\Lambda)$ and get by definition (2.3) of the Fréchet normal cone that

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

Fix $x \in \Lambda$, $t > 0$ and let $u := x/t$. Then $(x/t) \in \Lambda$, $tw \in \Lambda$, and

$$\limsup_{x \xrightarrow{\Lambda} tw} \frac{\langle x^*, x-tw\rangle}{\|x-tw\|} = \limsup_{x \xrightarrow{\Lambda} w} \frac{t\langle x^*, (x/t)-w\rangle}{t\|(x/t)-w\|} = \limsup_{u \xrightarrow{\Lambda} w} \frac{\langle x^*, u-w\rangle}{\|u-w\|} \le 0,$$

which gives $x^* \in \widehat{N}(tw;\Lambda)$ by (2.3). Letting finally $t \to 0$, we get $x^* \in N(0;\Lambda)$ and thus complete the proof of the proposition. $\square$

2.2 Subdifferentials and Coderivatives

Given a set-valued mapping $F : X \rightrightarrows Y$ between Banach spaces with the graph

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

we define the (normal) *coderivative* of F at $(\bar{x},\bar{y}) \in \operatorname{gph} F$ via the normal cone (2.6) by

$$D^*_N F(\bar{x},\bar{y})(y^*) := \big\{x^* \in X^* \big|\ (x^*,-y^*) \in N\big((\bar{x},\bar{y}); \operatorname{gph} F\big)\big\}, \quad y^* \in Y^*, \tag{2.7}$$

where $\bar{y} = f(\bar{x})$ is omitted if $F = f : X \to Y$ is single-valued. The notion of normal coderivative and its counter part $D^*_M F$, the so-called *mixed coderivative*, were originally introduced in [47]. Both notions coincide when Y is finite dimensional and will be denoted by D^*F.

Observe that the coderivative (2.7) is a positively homogeneous mapping $D_N^* F(\bar{x},\bar{y}) : Y^* \rightrightarrows X^*$, which reduces to the single-valued adjoint derivative operator

$$D_N^* f(\bar{x})(y^*) = \big\{\nabla f(\bar{x})^* y^*\big\} \;\text{ for all }\; y^* \in Y^* \tag{2.8}$$

if f is *strictly differentiable* at $\bar{x}$ in the sense that

$$\lim_{\substack{x\to\bar{x}\\ u\to\bar{x}}} \frac{f(x) - f(u) - \langle \nabla f(\bar{x}), x - u\rangle}{\|x - u\|} = 0;$$

the latter is automatic if f when C^1 around $\bar{x}$.

Given an extended-real-valued function $\varphi\colon X \to \overline{\mathbb{R}} := (-\infty,\infty]$, recall that the *Fréchet/regular subdifferential* of φ at $\bar{x}$ with $\varphi(\bar{x}) < \infty$ is defined by

$$\widehat{\partial}\varphi(\bar{x}) := \Big\{x^* \in X^* \Big|\; \liminf_{x\to\bar{x}} \frac{\varphi(x) - \varphi(\bar{x}) - \langle x^*, x - \bar{x}\rangle}{\|x - \bar{x}\|} \ge 0\Big\}. \tag{2.9}$$

It is easy to see that $\widehat{N}(\bar{x};\Omega) = \widehat{\partial}\delta(\bar{x};\Omega)$ for the indicator function $\delta(\cdot;\Omega)$ of Ω defined by $\delta(x;\Omega) := 0$ when $x \in \Omega$ and $\delta(x;\Omega) = \infty$ otherwise. We define its *limiting/basic subdifferential* at $\bar{x}$ by

$$\partial\varphi(\bar{x}) := \big\{x^* \in X^* \big|\; (x^*, -1) \in N\big((\bar{x},\varphi(\bar{x})); \operatorname{epi}\varphi\big)\big\} \tag{2.10}$$

via the normal cone (2.5) to the epigraph $\operatorname{epi}\varphi := \{(x,\mu) \in X \times \mathbb{R} |\; \mu \ge \varphi(x)\}$. When X is Asplund, the subdifferential (2.10) can be equivalently represented as

$$\partial\varphi(\bar{x}) = \operatorname*{Lim\,sup}_{x \xrightarrow{\varphi} \bar{x}} \widehat{\partial}\varphi(x). \tag{2.11}$$

2.3 Graphical Derivatives and Related Notions

Given a mapping $F:\mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and $(\bar{x},\bar{y})\in \operatorname{gph} F$, the *graphical/contingent derivative* of F at $(\bar{x},\bar{y})$ is introduced in [2] as a mapping $DF(\bar{x},\bar{y})\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ with the values

$$DF(\bar{x},\bar{y})(z) := \big\{w\in\mathbb{R}^m \big|\ (z,w)\in T\big((\bar{x},\bar{y});\operatorname{gph} F\big)\big\},\quad z\in\mathbb{R}^n, \tag{2.12}$$

defined via the contingent cone (2.2) to the graph of F at the point $(\bar{x},\bar{y})$; see [3, 61] for various properties, equivalent representation, and applications. We also drop $\bar{y}$ in the graphical derivative notation when the mapping in question is single-valued at $\bar{x}$. Note that the graphical derivative (2.12) and coderivative constructions (2.7) are *not dual* to each other, since the basic normal cone (2.5) is *nonconvex* and hence cannot be tangentially generated.

The following modified derivative construction for mappings, which seems to be new in generality although constructions of this (radial, Dini-like) type have been widely used for extended-real-valued functions.

Definition 2.2 (restrictive graphical derivative of mappings). *Let $F\colon\mathbb{R}^n \rightrightarrows \mathbb{R}^m$, and let $(\bar{x},\bar{y})\in\operatorname{gph} F$. Then a set-valued mapping $\widetilde{D}F(\bar{x},\bar{y})\colon\mathbb{R}^n \rightrightarrows \mathbb{R}^m$ given by*

$$\widetilde{D}F(\bar{x},\bar{y})(z) := \operatorname*{Lim\,sup}_{t\downarrow 0}\frac{F(\bar{x}+tz)-\bar{y}}{t},\quad z\in\mathbb{R}^n, \tag{2.13}$$

is called the RESTRICTIVE GRAPHICAL DERIVATIVE *of F at $(\bar{x},\bar{y})$.*

The next proposition collects some properties of the graphical derivative (2.12) and its restrictive counterpart (2.13) needed in what follows.

Proposition 2.3 (properties of graphical derivatives). *Let $F:\mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and $(\bar{x},\bar{y})\in \operatorname{gph} F$. Then the following assertions hold:*

(*a*) *We have* $\widetilde{D}F(\bar{x},\bar{y})(z) \subset DF(\bar{x},\bar{y})(z)$ *for all* $z \in \mathbb{R}^n$.

(*b*) *There are inverse derivative relationships*

$$DF(\bar{x},\bar{y})^{-1} = DF^{-1}(\bar{y},\bar{x}) \;\; \text{and} \;\; \widetilde{D}F(\bar{x},\bar{y})^{-1} = \widetilde{D}F^{-1}(\bar{y},\bar{x}).$$

(*c*) *If* F *is single-valued and locally Lipschitzian around* $\bar{x}$, *then*

$$\widetilde{D}F(\bar{x})(z) = DF(\bar{x})(z) \;\; \text{for all} \;\; z \in \mathbb{R}^n.$$

(*d*) *If* F *is single-valued and directionally differentiable at* $\bar{x}$, *then*

$$\widetilde{D}F(\bar{x})(z) = \big\{F'(\bar{x};z)\big\} \;\; \text{for all} \;\; z \in \mathbb{R}^n.$$

(*e*) *If* F *is single-valued and Gâteaux differentiable at* $\bar{x}$ *with the Gâteaux derivative* $F'_G(\bar{x})$, *then we have*

$$\widetilde{D}F(\bar{x})(z) = \big\{F'_G(\bar{x})z\big\} \;\; \text{for all} \;\; z \in \mathbb{R}^n.$$

(*f*) *If* F *is single-valued and* (*Fréchet*) *differentiable at* $\bar{x}$ *with the derivative* $F'(\bar{x})$, *then*

$$DF(\bar{x})(z) = \big\{F'(\bar{x})z\big\} \;\; \text{for all} \;\; z \in \mathbb{R}^n.$$

Proof. It is shown in [61] that the graphical derivative (2.12) admits the representation

$$DF(\bar{x},\bar{y})(z) = \operatorname*{Lim\,sup}_{t\downarrow 0,\, h\to z} \frac{F(\bar{x}+th)-\bar{y}}{t}, \quad z \in \mathbb{R}^n. \tag{2.14}$$

The inclusion in (a) is an immediate consequence of Definition 2.2 and representation (2.14).

The first equality in (b), observed from the very beginning [2], easily follows from definition (2.12). We can similarly check the second one in (b).

To justify the equality in (c), it remains to verify by (a) the opposite inclusion '$\supset$' when F is single-valued and locally Lipschitzian around $\bar{x}$. In this case fix $z \in \mathbb{R}^n$, pick any $w \in DF(\bar{x})(z)$, and find by representation (2.14) sequences $h_k \to z$ and $t_k \downarrow 0$ such that

$$\frac{F(\bar{x}+t_k h_k)-F(\bar{x})}{t_k} \to w \;\text{ as }\; k \to \infty.$$

The local Lipschitz continuity of F around $\bar{x}$ with constant $L \geq 0$ implies that

$$\begin{aligned}\Big\|\frac{F(\bar{x}+t_k h_k)-F(\bar{x})}{t_k}-\frac{F(\bar{x}+t_k z)-F(\bar{x})}{t_k}\Big\| &= \Big\|\frac{F(\bar{x}+t_k h_k)-F(\bar{x}+t_k z)}{t_k}\Big\| \\ &\leq L\|h_k - z\|\end{aligned}$$

for all $k \in \mathbb{N}$ sufficiently large, and hence we have the convergence

$$\frac{F(\bar{x}+t_k z)-F(\bar{x})}{t_k} \to w \;\text{ as }\; k \to \infty.$$

Thus $w \in \widetilde{D}F(\bar{x})(z)$, which justifies (c). Assertions (d) and (e) follow directly from the definitions. Finally, assertion (f) is implied by (e) in the local Lipschitzian case (c) while it can be easily derived from the (Fréchet) differentiability of F at $\bar{x}$ with no Lipschitz assumption; see, e.g., [61, Exercise 9.25(b)]. $\square$

Proposition 2.3 reveals important differences between the graphical derivative (2.12) and the coderivative (2.7). Indeed, assertions (c) and (d) of this proposition show that the graphical derivative of locally Lipschitzian and *directionally differentiable* mappings $F\colon \mathbb{R}^n \to \mathbb{R}^m$ is always *single-valued.* At the same time, the coderivative single-valuedness for locally Lipschitzian mappings is equivalent to the *strict/strong* Fréchet *differentiability* of F at the point in question; see [47, Theorem 3.66]. It follows from the well-known formula

$$\mathrm{co} D^*F(\bar{x})(z) = \big\{A^T z\big|\ A \in \partial_C F(\bar{x})\big\} \tag{2.15}$$

where $\partial_C F(\bar{x})$ denotes the Clarke's genelaized Jacobian of F at $\bar{x}$ and that the strict differentiability of F characterizes also the single-valuedness of $\partial_C F$. In the case of $F = (f_1, \ldots, f_m)\colon \mathbb{R}^n \to \mathbb{R}^m$ being locally Lipschitzian around $\bar{x}$ the coderivative (2.7) admits the subdifferential description

$$D^*F(\bar{x})(z) = \partial\Big(\sum_{i=1}^m z_i f_i\Big)(\bar{x}) \ \text{ for any } \ z = (z_1, \ldots, z_m) \in \mathbb{R}^m, \tag{2.16}$$

Finally, we recall the notion of metric regularity and its coderivative characterization that play a significant role in algorithm designs. A mapping $F \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *metrically regular* around $(\bar{x}, \bar{y}) \in \mathrm{gph}\, F$ if there are neighborhoods U of $\bar{x}$ and V of $\bar{y}$ as well as a number $\mu > 0$ such that

$$\mathrm{dist}\big(x; F^{-1}(y)\big) \le \mu\, \mathrm{dist}\big(y; F(x)\big) \ \text{ for all } \ x \in U \ \text{ and } \ y \in V. \tag{2.17}$$

Observe that it is sufficient to require the fulfillment of (2.17) just for those $y \in V$ satisfying the estimate $\mathrm{dist}(y; F(x)) \le \gamma$ for some $\gamma > 0$; see [47, Proposition 1.48].

We will see later that metric regularity is crucial for justifying the well-posedness of our generalized Newton algorithm and establishing its local and global convergence. It is also worth mentioning that, in the opposite direction, a Newton-type method (known as the Lyusternik-Graves iterative process) leads to verifiable conditions for metric regularity of smooth mappings;

see, e.g., the proof of [47, Theorem 1.57] and the commentaries therein. The latter procedure is replaced by variational/extremal principles of variational analysis in the case of nonsmooth and set-valued mappings under consideration; cf. [27, 47, 61].

We will broadly use the following *coderivative characterization* of the metric regularity property for an arbitrary set-valued mapping F with closed graph, known also as the Mordukhovich criterion (see [46, Theorem 3.6], [61, Theorem 9.45], and the references therein): F is metrically regular around $(\bar{x}, \bar{y})$ *if and only if* the inclusion

$$0 \in D^*F(\bar{x}, \bar{y})(z) \ \text{ implies that } \ z = 0, \tag{2.18}$$

which amounts the kernel condition $\ker D^*F(\bar{x}, \bar{y}) = \{0\}$.

Part A: Variational Extremal Principles

Chapter 3

Tangential Extremal Principles

3.1 Tangential Extremal Systems and Extremality Conditions

We start the chapter with this section introducing the notions of conic and tangential extremal systems for finite and countable collections of sets and discuss extremality conditions, which are at the heart of the conic and tangential extremal principles justified in the subsequent sections. These new extremality concepts are compared with conventional notions of local extremality for set systems.

We start with the new definitions of extremal points and extremal systems of a countable or finite number of cones and general sets in normed spaces.

Definition 3.1 (conic and tangential extremal systems). *Let X be an arbitrary normed space. Then we say that:*

(a) *A countable system of cones $\{\Lambda_i\}_{i\in\mathbb{N}} \subset X$ with $0 \in \cap_{i=1}^{\infty}\Lambda_i$ is* EXTREMAL AT THE ORIGIN, *or simply is an* EXTREMAL SYSTEM OF CONES, *if there is a bounded sequence $\{a_i\}_{i\in\mathbb{N}} \subset X$ with*

$$\bigcap_{i=1}^{\infty}\big(\Lambda_i - a_i\big) = \emptyset. \tag{3.1}$$

(b) *Let $\{\Omega_i\}_{i\in\mathbb{N}} \subset X$ be an countable system of sets with $\bar{x} \in \cap_{i=1}^{\infty}\Omega_i$, and let $\Lambda := \{\Lambda_i(\bar{x})\}_{i\in\mathbb{N}}$ with $0 \in \cap_{i=0}^{\infty}\Lambda_i(\bar{x}) \subset X$ be an approximating system of cones. Then $\bar{x}$ is a* Λ-TANGENTIAL LOCAL EXTREMAL POINT *of $\{\Omega_i\}_{i\in\mathbb{N}}$ if the system of cones $\{\Lambda_i(\bar{x})\}_{i\in\mathbb{N}}$ is extremal at the origin. In this case the collection $\{\Omega_i, \bar{x}\}_{i\in\mathbb{N}}$ is called a* Λ-TANGENTIAL EXTREMAL SYSTEM.

(c) *Suppose that* $\Lambda_i(\bar{x}) = T(\bar{x};\Omega_i)$ *are the contingent cones to* Ω_i *at* $\bar{x}$ *in* (b). Then $\{\Omega_i,\bar{x}\}_{i\in\mathbb{N}}$ is called a CONTINGENT EXTREMAL SYSTEM with the CONTINGENT LOCAL EXTREMAL POINT $\bar{x}$. We use the terminology of WEAK CONTINGENT EXTREMAL SYSTEM and WEAK CONTINGENT LOCAL EXTREMAL POINT if $\Lambda_i(\bar{x}) = T_w(\bar{x};\Omega_i)$ are the weak contingent cones to Ω_i at $\bar{x}$.

Note that all the notions in Definition 3.1 obviously apply to the case of systems containing *finitely* many sets; indeed, in such a case the other sets reduce to the whole space X. Observe also that both parts in part (c) of this definition are equivalent in finite dimensions. Furthermore, they both reduce to (a) in the general case if all the sets Ω_i are cones and $\bar{x} = 0$.

Let us now compare the new notions of Definition 3.1 with the conventional notion of locally extremal points for finitely many sets in (1.2).

We first observe that for finite systems of cones the local extremality of the origin in the sense of (1.2) is equivalent to the validity of condition (3.1) of Definition 3.1.

Proposition 3.2 (equivalent description of cone extremality). *The finite system of cones* $\{\Lambda_1,\ldots,\Lambda_m\}$ *is extremal at the origin in the sense of Definition* 3.1(a) *if and only if* $\bar{x} = 0$ *is a local extremal point of* $\{\Lambda_1,\ldots,\Lambda_m\}$ *in the sense of* (1.2).

Proof. The "only if" part is obvious. To justify the "if" part, assume that there are elements $a_1,\ldots,a_m \in X$ such that

$$\bigcap_{i=1}^{m} \big(\Lambda_i - a_i\big) = \emptyset. \tag{3.2}$$

Now for any $\eta > 0$ we have by (3.2) and the conic structure of Λ_i that

$$\emptyset = \bigcap_{i=1}^{m} \eta\big(\Lambda_i - a_i\big) = \bigcap_{i=1}^{m} \big(\eta\Lambda_i - \eta a_i\big) = \bigcap_{i=1}^{m} \big(\Lambda_i - \eta a_i\big).$$

Letting $\eta \downarrow 0$ implies that the extremality condition (1.2) holds, i.e., the origin is a local extremal point of the cone system $\{\Lambda_1, \ldots, \Lambda_m\}$. $\square$

Next we show that the local extremality (1.2) and the contingent extremality from Definition 3.1(c) are independent notions even in the case of two sets in $\mathbb{R}^2$.

Example 3.3 (contingent extremality versus local extremality).

(i) Consider two closed subsets in $\mathbb{R}^2$ defined by

$$\Omega_1 := \operatorname{epi}\varphi \;\text{ with }\; \varphi(x) := x\sin(1/x) \;\text{ as }\; x \neq 0,\; \varphi(0) = 0 \;\text{ and }\; \Omega_2 := (\mathbb{R} \times \mathbb{R}_-) \setminus \operatorname{int}\Omega_1.$$

Take the point $\bar{x} = (0,0) \in \Omega_1 \cap \Omega_2$ and observe that the contingent cones to Ω_1 and Ω_2 at $\bar{x}$ are computed, respectively, by

$$T(\bar{x};\Omega_1) = \operatorname{epi}(-|\cdot|) \;\text{ and }\; T(\bar{x};\Omega_2) = \mathbb{R} \times \mathbb{R}_-.$$

It is easy to see that $\bar{x}$ is a local extremal point of $\{\Omega_1, \Omega_2\}$ but not a contingent local extremal point of this set system.

(ii) Define two closed subsets of $\mathbb{R}^2$ by

$$\Omega_1 := \big\{(x_1,x_2) \in \mathbb{R}^2 \big|\; x_2 \geq -x_1^2\big\} \;\text{ and }\; \Omega_2 := \mathbb{R} \times \mathbb{R}_-.$$

The contingent cones to Ω_1 and Ω_2 at $\bar{x} = (0,0)$ are computed by

$$T(\bar{x};\Omega_1) = \mathbb{R} \times \mathbb{R}_+ \;\text{ and }\; T(\bar{x};\Omega_2) = \mathbb{R} \times \mathbb{R}_-.$$

We can see that $\{\Omega_1, \Omega_2, \bar{x}\}$ is a contingent extremal system but not an extremal system of sets.

Our further intention is to derive verifiable *extremality conditions* for tangentially extremal points of set systems in certain *countable* forms of the *generalized Euler equation* expressed via the limiting normal cone (2.5) at the points in question. Let us first formulate and discuss the desired conditions, which reflect the essence of the tangential extremal principles of this chapter.

Definition 3.4 (extremality conditions for countable systems). *We say that:*

(a) *The system of cones* $\{\Lambda_i\}_{i\in\mathbb{N}}$ *in* X *satisfies the* CONIC EXTREMALITY CONDITIONS *at the origin if there are normals* $x_i^*\in N(0;\Lambda_i)$ *for* $i=1,2,\ldots$ *such that*

$$\sum_{i=1}^{\infty}\frac{1}{2^i}x_i^*=0 \quad \textit{and} \quad \sum_{i=1}^{\infty}\frac{1}{2^i}\|x_i^*\|^2=1. \tag{3.3}$$

(b) *Let* $\{\Omega_i\}_{i\in\mathbb{N}}$ *with* $\bar{x}\in\cap_{i=1}^{\infty}\Omega_i$ *and* $\Lambda:=\{\Lambda_i\}_{i\in\mathbb{N}}$ *with* $0\in\cap_{i=1}^{\infty}\Lambda_i$ *be, respectively, systems of arbitrary sets and approximating cones in* X*. Then the system* $\{\Omega_i\}_{i\in\mathbb{N}}$ *satisfies the* Λ-TANGENTIAL EXTREMALITY CONDITIONS *at* $\bar{x}$ *if the systems of cones* $\{\Lambda_i\}_{i\in\mathbb{N}}$ *satisfies the conic extremality conditions at the origin. We specify the* CONTINGENT EXTREMALITY CONDITIONS *and the* WEAK CONTINGENT EXTREMALITY CONDITIONS *for* $\{\Omega_i\}_{i\in\mathbb{N}}$ *at* $\bar{x}$ *if* $\Lambda=\{T(\bar{x};\Omega_i)\}_{i\in\mathbb{N}}$ *and* $\Lambda=\{T_w(\bar{x};\Omega_i)\}_{i\in\mathbb{N}}$*, respectively.*

(c) *The system of sets* $\{\Omega_i\}_{i\in\mathbb{N}}$ *in* X *satisfies the* LIMITING EXTREMALITY CONDITIONS *at* $\bar{x}\in\cap_{i=1}^{\infty}\Omega_i$ *if there are limiting normals* $x_i^*\in N(\bar{x};\Omega_i)$*,* $i=1,2,\ldots$*, satisfying* (3.3).

Let us briefly discuss the introduced extremality conditions.

Remark 3.5 (discussions on extremality conditions).

(i) All the conditions of Definition 3.4 can be obviously specified to the case of *finite systems* of sets by considering all the other sets as the whole space therein. Then the series in (3.3) become finite sums and the coefficients 2^{-i} can be dropped by rescaling.

(ii) It easily follows from the constructions involved that the contingent, weak contingent, and limiting extremality conditions are are *equivalent* to each other if all the sets Ω_i are either *cones* with $\bar{x} = 0$ or *convex* near $\bar{x}$.

(iii) As we show below, the weak contingent extremality conditions *imply* the limiting extremality conditions in any reflexive space X and also in Asplund spaces under a certain additional assumption, which is automatic under reflexivity. Thus the contingent extremality conditions imply the limiting ones in finite dimensions. The opposite implication does *not hold* even for two sets in $\mathbb{R}^2$. To illustrate it, consider the two sets from Example 3.3(i) for which $\bar{x} = (0,0)$ is a local extremal point in the usual sense, and hence the limiting extremality conditions hold due to [47, Theorem 2.8]. However, it is easy to see that the contingent extremality conditions are violated for this system.

Observe that for the case of *finitely many* sets $\{\Omega_1, \ldots, \Omega_m\}$ the limiting extremality conditions of Definition 3.4(c) correspond to the generalized Euler equation in the *exact extremal principle* of [47, Definition 2.5(iii)] applied to local extremal points of sets. A natural version of the "fuzzy" Euler equation in the *approximate extremal principle* of [47, Definition 2.5(ii)] for the case of a *countable* set system $\{\Omega_i\}_{i\in\mathbb{N}}$ at $\bar{x} \in \cap_{i=1}^{\infty}\Omega_i$ can be formulated as follows: for any $\varepsilon > 0$ there are

$$x_i \in \Omega_i \cap (\bar{x} + \varepsilon I\!B) \ \text{ and } \ x_i^* \in \widehat{N}(x_i;\Omega_i) + \frac{1}{2^i}\varepsilon I\!B^*, \quad i \in \mathbb{N}, \tag{3.4}$$

such that the relationships in (3.3) is satisfied. It turns out that such a countable version of the approximate extremal principle always *holds trivially*, at least in Asplund spaces, for *any* system of closed sets $\{\Omega_i\}_{i\in\mathbb{N}}$ at *every* boundary point $\bar{x}$ of infinitely many sets Ω_i.

Proposition 3.6 (triviality of the approximate extremality conditions for countable

set systems). *Let* $\{\Omega_i\}_{i\in\mathbb{N}}$ *be a countable system of sets closed around some point* $\bar{x}\in\cap_{i=1}^{\infty}\Omega_i$*, and let* $\varepsilon>0$*. Assume that for infinitely many* $i\in\mathbb{N}$ *there exist* $x_i\in\Omega_i\cap(\bar{x}+\varepsilon I\!B)$ *such that* $\widehat{N}(x_i;\Omega_i)\neq\{0\}$*; this is the case when* X *is Asplund and* $\bar{x}$ *belongs to the boundary of infinitely many sets* Ω_i*. Then we always have* $\{x_i^*\}_{i\in\mathbb{N}}$ *satisfying conditions* (3.3) *and* (3.4)*.*

Proof. Observe first that the fulfillment of the assumption made in the proposition for the case of Asplund spaces follows from the density of Fréchet normals on boundaries of closed sets in such spaces; see, e.g., [47, Corollary 2.21]. To proceed further, fix $\varepsilon>0$ and find $j\in\mathbb{N}$ so large that

$$\frac{\sqrt{2^j}}{2^{j-1}}\le\frac{1}{2}\varepsilon\quad\text{and}\quad\widehat{N}(x_j;\Omega_j)\neq\{0\}\quad\text{with}\quad x_j\in\Omega_j\cap(\bar{x}+\varepsilon I\!B).$$

This allows us to get $0\neq x_j^*\in\widehat{N}(x_j;\Omega_j)$ such that $\|x_j^*\|=\sqrt{2^j}$ and then choose

$$\begin{aligned}&x_1^*:=-\frac{1}{2^{j-1}}x_j^*\in 0+\frac{1}{2}\varepsilon I\!B^*\subset\widehat{N}(x_1;\Omega_2)+\frac{1}{2}\varepsilon I\!B^*,\quad x_j^*\in\widehat{N}(x_j;\Omega_j)+\frac{1}{2^j}\varepsilon I\!B^*,\\&\text{and}\quad x_i^*:=0\in\widehat{N}(x_i;\Omega_i)+\frac{1}{2^i}\varepsilon I\!B^*\quad\text{for all}\quad i\neq 1,j.\end{aligned}$$

Thus we have the sequence $\{x_i^*\}_{i\in\mathbb{N}}$ satisfying (3.4) and the relationships

$$\sum_{i=1}^{\infty}\frac{1}{2^i}x_i^*=\frac{1}{2}\Big(-\frac{1}{2^{j-1}}x_j^*\Big)+0+\ldots+\frac{1}{2^j}x_j^*+\ldots=0,\quad\sum_{i=1}^{\infty}\frac{1}{2^i}\|x_i^*\|^2>1,$$

which give (3.3) and complete the proof of the proposition. □

3.2 Conic Extremal Principle for Countable Systems of Sets

This section addresses the *conic extremal principle* for countable systems of cones in finite-dimensional spaces. This is the first extremal principle for infinite systems of sets, which ensures the fulfillment of the conic extremality conditions of Definition 3.4(a) for a conic extremal system at the origin under a natural nonoverlapping assumption. We present a number of examples illustrating the results obtained and the assumptions made.

To derive the main result of this section, we extend the *method of metric approximations* initiated in [45] to the case of countable systems of cones; cf. an essentially different realization of this method in the proof of the extremal principle for local extremal points of finitely many sets in $\mathbb{R}^n$ given in [47, Theorem 2.8]. First observe an elementary fact needed in what follows.

Lemma 3.7 (series differentiability). *Let $\|\cdot\|$ be the usual Euclidian norm in $\mathbb{R}^n$, and let $\{z_i\}_{i\in\mathbb{N}}\subset\mathbb{R}^n$ be a bounded sequence. Then a function $\varphi\colon\mathbb{R}^n\to\mathbb{R}$ defined by*

$$\varphi(x):=\sum_{i=1}^{\infty}\frac{1}{2^i}\big\|x-z_i\big\|^2,\quad x\in\mathbb{R}^n,$$

is continuously differentiable on $\mathbb{R}^n$ with the derivative $\nabla\varphi(x)=\displaystyle\sum_{i=1}^{\infty}\frac{1}{2^{i-1}}\big(x-z_i\big)$, $x\in\mathbb{R}^n$.

Proof. It is easy to see that both series above converge for every $x\in\mathbb{R}^n$. Taking further any $u,\xi\in\mathbb{R}^n$ with the norm $\|\xi\|$ sufficiently small, we have

$$\|u+\xi\|^2-\|u\|^2-2\langle u,\xi\rangle=\|u\|^2+2\langle u,\xi\rangle+\|\xi\|^2-\|u\|^2-2\langle u,\xi\rangle=\|\xi\|^2=o(\|\xi\|).$$

Thus it follows for any $x\in\mathbb{R}^n$ and y close to x that

$$\begin{aligned}\varphi(y)-\varphi(x)-\Big\langle\nabla\varphi(x),y-x\Big\rangle&=\sum_{i=1}^{\infty}\frac{1}{2^i}\Big[\|y-z_i\|^2-\|x-z_i\|^2-2\big\langle x-z_i,y-x\big\rangle\Big]\\&=\sum_{i=1}^{\infty}\frac{1}{2^i}\,\|y-x\|^2=o(\|y-x\|),\end{aligned}$$

which justifies that $\nabla\varphi(x)$ is the derivative of φ at x, which is obviously continuous on $\mathbb{R}^n$. $\Box$

Here is the extremal principle for a countable systems of cones, which plays a crucial role in the subsequent applications in this chapter.

Theorem 3.8 (conic extremal principle in finite dimensions). *Let $\{\Lambda_i\}_{i\in\mathbb{N}}$ be an extremal system of closed cones in $X=\mathbb{R}^n$ satisfying the* NONOVERLAPPING CONDITION

$$\bigcap_{i=1}^{\infty}\Lambda_i=\{0\}. \tag{3.5}$$

Then the conic extremal principle holds, i.e., there are $x_i^*\in N(0;\Lambda_i)$ for $i=1,2,\ldots$ such that

$$\sum_{i=1}^{\infty}\frac{1}{2^i}x_i^*=0 \quad \textit{and} \quad \sum_{i=1}^{\infty}\frac{1}{2^i}\|x_i^*\|^2=1.$$

Moreover, one can find $w_i\in\Lambda_i$ for which $x_i^*\in\widehat{N}(w_i;\Lambda_i)$, $i=1,2,\ldots$.

Proof. Pick a bounded sequence $\{a_i\}_{i\in\mathbb{N}}\subset\mathbb{R}^n$ from Definition 3.1(a) satisfying

$$\bigcap_{i=1}^{\infty}\big(\Lambda_i-a_i\big)=\emptyset$$

and consider the unconstrained optimization problem:

$$\text{minimize }\varphi(x):=\left[\sum_{i=1}^{\infty}\frac{1}{2^i}\text{dist}^{\,2}(x+a_i;\Lambda_i)\right]^{\frac{1}{2}},\quad x\in\mathbb{R}^n. \tag{3.6}$$

Let us prove that problem (3.6) has an optimal solution. Since the function φ in (3.6) is continuous on $\mathbb{R}^n$ due the continuity of the distance function and the uniform convergence of the series therein, it suffices to show that there is $\alpha>0$ for which the nonempty level set $\{x\in\mathbb{R}^n|\ \varphi(x)\le\inf_x\varphi+\alpha\}$ is bounded and then to apply the classical Weierstrass theorem. Suppose by the contrary that the level sets are unbounded whenever $\alpha>0$, for any $k\in\mathbb{N}$ find $x_k\in\mathbb{R}^n$ satisfying

$$\|x_k\|>k\ \text{ and }\ \varphi(x_k)\le\inf_x\varphi+\frac{1}{k}.$$

Setting $u_k := x_k/\|x_k\|$ with $\|u_k\| = 1$ and taking into account that all Λ_i are cones, we get

$$\frac{1}{\|x_k\|}\varphi(x_k) = \left[\sum_{i=1}^{\infty}\frac{1}{2^i}\text{dist}^2\Big(u_k + \frac{a_i}{\|x_k\|};\Lambda_i\Big)\right]^{\frac{1}{2}} \le \frac{1}{\|x_k\|}\Big(\inf_x \varphi + \frac{1}{k}\Big) \to 0 \;\text{ as }\; k \to \infty. \tag{3.7}$$

Furthermore, there is $M > 0$ such that for large $k \in \mathbb{N}$ we have

$$\text{dist}\Big(u_k + \frac{a_i}{\|x_k\|};\Lambda_i\Big) \le \Big\|u_k + \frac{a_i}{\|x_k\|}\Big\| \le M.$$

Without relabeling, assume $u_k \to u$ as $k \to \infty$ with some $u \in \mathbb{R}^n$. Passing now to the limit as $k \to \infty$ in (3.7) and employing the uniform convergence of the series therein and the fact that $a_i/\|x_k\| \to 0$ uniformly in $i \in \mathbb{N}$ due the boundedness of $\{a_i\}_{i\in\mathbb{N}}$, we have

$$\left[\sum_{i=1}^{\infty}\frac{1}{2^i}\text{dist}^2(u;\Lambda_i)\right]^{\frac{1}{2}} = 0.$$

This implies by the closedness of the cones Λ_i and the *nonoverlapping condition* (3.5) of the theorem that $u \in \bigcap_{i=1}^{\infty}\Lambda_i = \{0\}$. The latter is impossible due to $\|u\| = 1$, which contradicts our intermediate assumption on the unboundedness of the level sets for φ and thus justifies the existence of an optimal solution $\widetilde{x}$ to problem (3.6).

Since the system of closed cones $\{\Lambda_i\}_{i\in\mathbb{N}}$ is *extremal at the origin*, it follows from the construction of φ in (3.6) that $\varphi(\widetilde{x}) > 0$. Taking into account the nonemptiness of the projection $\Pi(x;\Lambda)$ of $x \in \mathbb{R}^n$ onto an arbitrary closed set $\Lambda \subset \mathbb{R}^n$, pick any $w_i \in \Pi(\widetilde{x} + a_i;\Lambda_i)$ as $i \in \mathbb{N}$ and observe from Proposition 2.1 above and the proof of [47, Theorem 1.6] that

$$\widetilde{x} + a_i - w_i \in \Pi^{-1}(w_i;\Lambda_i) - w_i \subset \widehat{N}(w_i;\Lambda_i) \subset N(0;\Lambda_i). \tag{3.8}$$

Furthermore, the sequence $\{a_i - w_i\}_{i\in\mathbb{N}}$ is bounded in $\mathbb{R}^n$ due to

$$\|x + a_i - w_i\| = \text{dist}\,(x + a_i; \Lambda_i) \le \|x + a_i\|.$$

Next we consider another unconstrained optimization problem:

$$\text{minimize } \psi(x) := \left[\sum_{i=1}^{\infty} \frac{1}{2^i}\|x + a_i - w_i\|^2\right]^{\frac{1}{2}}, \quad x \in \mathbb{R}^n. \tag{3.9}$$

It follows from $\psi(x) \ge \varphi(x) \ge \varphi(\widetilde{x}) = \psi(\widetilde{x})$ for all $x \in \mathbb{R}^n$ that problem (3.9) has the same optimal solution $\widetilde{x}$ as (3.6). The main difference between these two problems is that the cost function ψ in (3.9) is *smooth* around $\widetilde{x}$ by Lemma 3.7, the smoothness of the function $\sqrt{t}$ around nonzero points, and the fact that $\psi(\widetilde{x}) \ne 0$ due to the cone extremality. Applying now the *classical Fermat rule* to the smooth unconstrained minimization problem (3.9) and using the derivative calculation in Lemma 3.7, we arrive at the relationships

$$\nabla\psi(\widetilde{x}) = \sum_{i=1}^{\infty} \frac{1}{2^i} x_i^* = 0 \ \text{ with } \ x_i^* := \frac{1}{\psi(\widetilde{x})}\Big(\widetilde{x} + a_i - w_i\Big), \quad i \in \mathbb{N}. \tag{3.10}$$

The latter implies by (3.8) that $x_i^* \in \widehat{N}(w_i; \Lambda_i) \subset N(0; \Lambda_i)$ for all $i \in \mathbb{N}$. Furthermore, it follows from the constructions of x_i^* in (3.10) and of ψ in (3.9) that $\displaystyle\sum_{i=1}^{\infty} \frac{1}{2^i}\|x_i^*\|^2 = 1$, which thus completes the proof of the theorem. $\square$

In the remaining part of this section, we present three examples showing that all the assumptions made in Theorem 3.8 (nonoverlapping, finite dimension, and conic structure) are *essential* for the validity of this result.

Example 3.9 (nonoverlapping condition is essential). Let us show that the conic extremal principle may fail for countable systems of *convex* cones in $\mathbb{R}^2$ if the nonoverlapping condition (3.5) is violated. Define the convex cones $\Lambda_i \subset \mathbb{R}^2$ as $i \in \mathbb{N}$ by

$$\Lambda_1 := \mathbb{R} \times \mathbb{R}_+ \ \text{ and } \ \Lambda_i := \left\{(x,y) \in \mathbb{R}^2 \middle|\ y \le \frac{x}{i}\right\} \ \text{ for } \ i = 2,3,\ldots.$$

Observe that for any $\nu > 0$ we have

$$\Big(\Lambda_1 + (0,\nu)\Big) \bigcap_{k=2}^{\infty} \Lambda_k = \emptyset,$$

which means that the cone system $\{\Lambda_i\}_{i\in\mathbb{N}}$ is extremal at the origin. On the other hand,

$$\bigcap_{i=1}^{\infty} \Lambda_i = \mathbb{R}_+ \times \{0\},$$

i.e., the nonoverlapping condition (3.5) is violated. Furthermore, we can easily compute the corresponding normal cones by

$$N(0;\Lambda_1) = \big\{\lambda(0,-1)\big|\ \lambda \ge 0\big\} \ \text{ and } \ N(0;\Lambda_i) = \big\{\lambda(-1,i)\big|\ \lambda \ge 0\big\}, \ \ i = 2,3,\ldots.$$

Taking now any $x_i^* \in N(0;\Lambda_i)$ as $i \in \mathbb{N}$, observe the equivalence

$$\Big[\sum_{i=1}^{\infty} \frac{1}{2^i} x_i^* = 0\Big] \Longleftrightarrow \Big[\frac{\lambda_1}{2}(0,-1) + \sum_{i=2}^{\infty} \frac{\lambda_i}{2^i}\big(-1,i\big) = 0 \ \text{ with } \ \lambda_i \ge 0 \ \text{ as } \ i \in \mathbb{N}\Big].$$

The latter implies that $\lambda_i = 0$ and hence $x_i^* = 0$ for all $i \in \mathbb{N}$. Thus the nontriviality condition in (3.3) is not satisfied, which shows that the conic extremal principle fails for this system.

Example 3.10 (conic structure is essential). If all the sets Ω_i for $i \in \mathbb{N}$ are *convex* but some of them are *not cones*, then the equivalent extremality conditions of Definition 3.4(b,c) are natural extensions of the conic extremality conditions in Theorem 3.8. We show nevertheless that the corresponding extension of the conic extremal principle under the nonoverlapping requirement

$$\bigcap_{i=1}^{\infty} \Omega_i = \{0\} \tag{3.11}$$

fails without imposing a conic structure on all the sets involved. Indeed, consider a countable system of closed and convex sets in $\mathbb{R}^2$ defined by

$$\Omega_1 := \big\{(x,y) \in \mathbb{R}^2 \big|\ y \geq x^2\big\} \ \text{ and } \ \Omega_i := \big\{(x,y) \in \mathbb{R}^2 \big|\ y \leq \frac{x}{i}\big\} \ \text{ for } \ i = 2, 3, \ldots.$$

We can see that only the set Ω_1 is not a cone and that the nonoverlapping requirement (3.11) is satisfied. Furthermore, the system $\{\Omega_i\}_{i\in\mathbb{N}}$ is extremal at the origin in the sense that (3.1) holds. However, the arguments similar to Example 3.9 show that the extremality conditions (3.3) with $x_i^* \in N(0;\Omega_i)$ as $i \in \mathbb{N}$ fail to fulfill. Note that, as shown in Section 3.4, both contingent and limiting extremal principles hold for countable systems of general nonconvex sets if nonoverlapping condition (3.11) is replaced by another one reflecting the *contingent extremality*.

Example 3.11 (failure of the conic extremal principle in infinite dimensions). This last example demonstrates that the conic extremal principle of Theorem 3.8 with the nonoverlapping condition (3.5) may fail for countable systems of *convex cones* (in fact, half-spaces) in an arbitrary infinite-dimensional *Hilbert space*. To proceed, consider a Hilbert space X with the orthonormal basis $\{e_i |\ i \in \mathbb{N}\}$ and define a countable system of closed half-spaces by $\Lambda_1 := \big\{x \in X \big|\ \langle x, e_1\rangle \leq 0\big\}$ and $\Lambda_i := \big\{x \in X \big|\ \langle x, e_i - e_{i-1}\rangle \leq 0\big\}$ for $i = 2, 3, \ldots$

It is easy to compute the corresponding normal cones to the above sets:

$$N(0;\Lambda_1)=\big\{\lambda e_1\big|\ \lambda\ge 0\big\}\ \text{ and }\ N(0;\Lambda_i)=\big\{\lambda(e_i-e_{i-1})\big|\ \lambda\ge 0\big\}\ \text{ for }\ i=2,3,\ldots.$$

Now let us check that the nonoverlapping condition (3.5) is satisfied. Indeed, picking any point

$$x=\sum_{i=1}^{\infty}\alpha_i e_i\in\bigcap_{i=1}^{\infty}\Lambda_i,$$

we have $\alpha_1=\langle x,e_1\rangle\le 0$ and $\alpha_i=\langle x,e_i\rangle\le\langle x,e_{i-1}\rangle=\alpha_{i-1}$ for $i=2,3,\ldots$. This clearly leads to $\alpha_i=0$ for all $i\in\mathbb{N}$, which yields $x=0$ and thus justifies (3.5). The same arguments show that

$$(\Lambda_1-e_1)\cap\bigcap_{i=2}^{\infty}\Lambda_i=\emptyset,$$

i.e., $\{\Lambda_i\}_{i\in\mathbb{N}}$ is a *conic extremal system*. However, the conic extremality conditions of Definition 3.4(a) *fail* for this system. To check this, suppose that there exist $x_i^*\in N(0;\Lambda_i)$ as $i\in\mathbb{N}$ satisfying the relationships

$$\sum_{i=1}^{\infty}x_i^*=0\ \text{ and }\ \sum_{i=1}^{\infty}\|x_i^*\|>0. \tag{3.12}$$

By the above structure of $N(0;\Lambda_i)$ we have $x_1^*=\lambda_1 e_1$ and $x_i^*=\lambda_i(e_i-e_{i-1})$ as $i=2,3,\ldots$ for some $\lambda_i\ge 0$ as $i\in\mathbb{N}$. Thus the first condition in (3.12) reduces to

$$\lambda_1 e_1+\sum_{i=2}^{\infty}\lambda_i\big(e_i-e_{i-1}\big)=0.$$

The latter is possible if either (a): $\lambda_i=1$ for all $i\in\mathbb{N}$ or (b): $\lambda_i=0$ for all $i\in\mathbb{N}$. Case (a) surely contradicts the convergence of the series in the second condition of (3.12) while in case (b) the latter series converges to zero. Hence the conic extremal principle of Theorem 3.8 does not hold in this infinite-dimensional setting.

3.3 Tangential Normal Enclosedness and Approximate Normality

In this section we introduce and study two important properties of tangents cones that are of their own interest while allow us make a bridge between the extremal principles for cones and the limiting extremality conditions for arbitrary closed sets at their tangential extremal points. The main attention is paid to the contingent and weak contingent cones, which are proved to enjoy these properties under natural assumptions.

Let us start with introducing a new property of sets that is formulated in terms of the limiting normal cone (2.5) and plays a crucial role of what follows.

Definition 3.12 (tangential normal enclosedness). *Given a nonempty subset $\Omega \subset X$ and a subcone $\Lambda \subset X$ of a Banach space X, we say that Λ is* TANGENTIALLY NORMALLY ENCLOSED (TNE) *into Ω at a point $\bar{x} \in \Omega$ if*

$$N(0; \Lambda) \subset N(\bar{x}; \Omega). \tag{3.13}$$

The word "tangential" in Definition 3.12 reflects the fact that this normal enclosedness property is applied to tangential approximations of sets at reference points. Observe that if the set Ω is convex near $\bar{x}$, then its classical tangent cone at $\bar{x}$ enjoys the TNE property; indeed, in this case inclusion (3.13) holds as equality. We establish below a remarkable fact on the validity of the TNE property for the weak contingent cone to any closed subset of a reflexive Banach space.

To study this and related properties, fix $\Omega \subset X$ with $\bar{x} \in \Omega$ and denote by $\Lambda_w := T_w(\bar{x}; \Omega)$

the weak contingent cone to Ω at $\bar{x}$ without indicating Ω and $\bar{x}$ for brevity. Given a direction $d \in \Lambda_w$, let $\mathcal{T}_d^w$ be the collection of all sequences $\{x_k\} \subset \Omega$ such that

$$\frac{x_k - \bar{x}}{t_k} \xrightarrow{w} d \text{ for some } t_k \downarrow 0.$$

It follows from definition of $\Lambda_w = T_w(\bar{x}; \Omega)$ that $\mathcal{T}_d^w \neq \emptyset$ whenever $d \in \Lambda_w$.

Definition 3.13 (tangential approximate normality). *We say that $\Omega \subset X$ has the* TANGENTIAL APPROXIMATE NORMALITY (TAN) *property at $\bar{x} \in \Omega$ if whenever $d \in \Lambda_w$ and $x^* \in \widehat{N}(d; \Lambda_w)$ are chosen there is a sequence $\{x_k\} \in \mathcal{T}_d^w$ along which the following holds: for any $\varepsilon > 0$ there exists $\delta \in (0, \varepsilon)$ such that*

$$\limsup_{k \to \infty} \Big[\sup \Big\{ \frac{\langle x^*, z - x_k \rangle}{t_k} \Big| \, z \in \Omega \cap (x_k + t_k \delta I\!B) \Big\} \Big] \leq 2\varepsilon\delta, \tag{3.14}$$

where $t_k \downarrow 0$ is taken from the construction of $\mathcal{T}_d^w$.

The meaning of this property that gives the name is as follows: any $x^* \in \widehat{N}(d; \Lambda_w)$ for the tangential approximation of Ω at $\bar{x}$ behaves approximately like a true normal at appropriate points x_k near $\bar{x}$. It occurs that the TAN property holds for any closed subset of a reflexive Banach space. The next proposition provides even a stronger result.

Proposition 3.14 (approximate tangential normality in reflexive spaces). *Let Ω be a subset of a reflexive space X, and let $\bar{x} \in \Omega$. Then given any $d \in \Lambda_w = T_w(\bar{x}; \Omega)$ and $x^* \in \widehat{N}(d; \Lambda_w)$, we have (3.14) whenever sequences $\{x_k\} \in \mathcal{T}_d^w$ and $t_k \downarrow 0$ are taken from the construction of $\mathcal{T}_d^w$. In particular, the set Ω enjoys the TAN property at $\bar{x}$.*

Proof. Assume that $\bar{x} = 0$ for simplicity. Pick any $\varepsilon > 0$ and by the definition of Fréchet normals find $\delta \in (0, \varepsilon)$ such that

$$\langle x^*, v - d\rangle \leq \frac{\varepsilon}{2}\|v - d\| \text{ for all } v \in \Lambda_w \cap (d + \delta I\!B). \tag{3.15}$$

Fix any sequences $\{x_k\} \in \mathcal{T}_d^w$ and $t_k \downarrow 0$ from the formulation of the proposition and show that property (3.14) holds with the numbers ε and δ chosen above. Supposing the contrary, find $\{x_k\} \in \mathcal{T}_d^w$ and the corresponding sequence $t_k \downarrow 0$ such that

$$\lim_{k\to\infty} \left\{ \sup \frac{\langle x^*, z - x_k\rangle}{t_k} \middle| \, z \in \Omega \cap (B(x_k + t_k \delta I\!B) \right\} > 2\varepsilon\delta$$

along some subsequence of $k \in \mathbb{N}$, with no relabeling here and in what follows. Hence there is a sequence of $z_k \in \cap(x_k + t_k\delta I\!B)$ along which

$$\frac{\langle x^*, z_k - x_k\rangle}{t_k} > \varepsilon\delta \text{ for } k \in \mathbb{N}.$$

Taking into account the relationships

$$\left\|\frac{z_k}{t_k} - \frac{x_k}{t_k}\right\| \leq \delta \text{ and } \frac{x_k}{t_k} \xrightarrow{w} d \text{ as } k \to \infty,$$

we get that the sequence $\left\{\dfrac{x_k}{t_k}\right\}$ is bounded in X, and so is $\left\{\dfrac{z_k}{t_k}\right\}$. Since any bounded sequence in a reflexive Banach space contains a weakly convergent subsequence, we may assume with no loss of generality that the sequence $\left\{\dfrac{z_k}{t_k}\right\}$ weakly converges to some $v \in X$ as $k \to \infty$. It follows from the weak convergence of this sequence that

$$\|v - d\| \leq \liminf_{k\to\infty} \left\|\frac{z_k}{t_k} - \frac{x_k}{t_k}\right\| \leq \delta.$$

This allows us to conclude that

$$\langle x^*, v - d\rangle \geq \varepsilon\delta > \frac{\varepsilon}{2}\delta \geq \frac{\varepsilon}{2}\|v - d\|,$$

which contradicts (3.15) and thus completes the proof of the proposition. □

The next theorem is the main result of this section showing that the TAN property of a closed set in an Asplund space implies the TNE property of the weak contingent cone to this set at the reference point. This unconditionally justifies the latter property in reflexive spaces.

Theorem 3.15 (TNE property in Asplund spaces). *Let Ω be a closed subset of an Asplund space X, and let $\bar{x} \in \Omega$. Assume that Ω has the tangential approximate normality property at $\bar{x}$. Then the weak contingent cone $\Lambda_w = T(\bar{x}; \Omega)$ is tangentially normally enclosed into Ω at this point. Furthermore, the latter TNE property holds for any closed subset of a reflexive space.*

Proof. We are going show that the following holds in the Asplund space setting under the TAN property of Ω at $\bar{x}$:

$$\widehat{N}(d; \Lambda_w) \subset N(\bar{x}; \Omega) \text{ for all } d \in \Lambda, \ \|d\| = 1, \tag{3.16}$$

which is obviously equivalent to $N(0; \Lambda_w) \subset N(\bar{x}; \Omega)$, the TNE property of the weak contingent cone Λ_w. Then the second conclusion of the theorem in reflexive spaces immediately follows from Proposition 3.14. Assume without loss of generality that $\bar{x} = 0$. To justify (3.16), fix $d \in \Lambda_w$ and $x^* \in \widehat{N}(d; \Lambda_w)$ with $\|d\| = 1$ and $\|x^*\| = 1$. Taking $\{x_k\} \in \mathcal{T}_d^w$ from Definition 3.13, it follows that for any ε there is $\delta < \varepsilon$ such that (3.14) holds with $\bar{x} = 0$. Hence

$$\langle x^*, z - x_k\rangle \leq 3t_k\varepsilon\delta \text{ whenever } z \in Q := \Omega \cap (x_k + t_k\delta I\!B), \quad k \in \mathbb{N}. \tag{3.17}$$

Consider further the function $\varphi(z) := -\langle x^*, z - x_k\rangle$, $z \in Q$, for which we have by (3.17) that

$$\varphi(x_k) = 0 \le \inf_{z\in Q} \varphi(z) + 3t_k\varepsilon\delta.$$

Setting $\lambda := \frac{t_k\delta}{3}$ and $\widetilde{\varepsilon} := 3t_k\varepsilon\delta$, we apply the Ekeland variational principle (see, e.g., [47, Theorem 2.26]) with λ and $\widetilde{\varepsilon}$ to the function φ on Q. In this way we find $\widetilde{x} \in Q$ such that $\|\widetilde{x} - x_k\| \le \lambda$ and $\widetilde{x}$ minimizes the perturbed function

$$\psi(z) := -\langle x^*, z - x_k\rangle + \frac{\widetilde{\varepsilon}}{\lambda}\|z - \widetilde{x}\| = -\langle x^*, z - x_k\rangle + 9\varepsilon\|z - \widetilde{x}\|, \quad z \in Q.$$

Applying now the generalized Fermat rule to ψ at $\widetilde{x}_k$ and then the fuzzy sum rule in the Asplund space setting (see, e.g., [47, Lemma 2.32]) gives us

$$0 \in -x^* + (9\varepsilon + \lambda)I\!B^* + \widehat{N}(\widetilde{x}_k; Q) \tag{3.18}$$

with some $\widetilde{x}_k \in \Omega \cap (\widetilde{x} + \lambda I\!B)$. The latter means that

$$\|\widetilde{x}_k - x_k\| \le \|\widetilde{x}_k - \widetilde{x}\| + \|\widetilde{x} - x_k\| \le 2\lambda < t_k\delta.$$

Hence $\widetilde{x}_k$ belongs to the interior of the ball centered at $\widetilde{x}$ with radius $t_k\delta$, which implies that $\widehat{N}(\widetilde{x}_k; Q) = \widehat{N}(\widetilde{x}_k; \Omega)$. Thus we get from (3.18) that

$$x^* \in \widehat{N}(\widetilde{x}_k; \Omega) + (9\varepsilon + \lambda)I\!B^*, \quad k \in \mathbb{N}.$$

Letting there $k \to \infty$ and then $\varepsilon \downarrow 0$ gives us $\widetilde{x}_k \to \bar{x}$ and $x^* \in N(\bar{x}; \Omega)$. This justifies (3.16) and completes the proof of the theorem. □

Corollary 3.16 (TNE property of the contingent cone in finite dimensions). *Let a set $\Omega \subset \mathbb{R}^n$ be closed around $\bar{x} \in \Omega$. Then the contingent cone $T(\bar{x};\Omega)$ to Ω at $\bar{x}$ is tangentially normally enclosed into Ω at this point, i.e., we have*

$$N(0;\Lambda) \subset N(\bar{x};\Omega) \quad \text{with} \quad \Lambda := T(\bar{x};\Omega). \tag{3.19}$$

Proof. It follows from Theorem 3.15 due to $T(\bar{x};\Omega) = T_w(\bar{x};\Omega)$ in $\mathbb{R}^n$. □

Note that another proof of inclusion (3.19) in $\mathbb{R}^n$ can be found in [61, Theorem 6.27].

3.4 Contingent and Weak Contingent Extremal Principles for Countable and Finite Systems of Closed Sets

By *tangential extremal principles* we understand results justifying the validity of extremality conditions defined in Section 3.1 for countable and/or finite systems of closed sets at the corresponding *tangential extremal points.* Note that, given a system of $\Lambda = \{\Lambda_i\}$-approximating cones to a set system $\{\Omega_i\}$ at $\bar{x}$, the results ensuring the fulfillment of the Λ-tangential extremality conditions at Λ-tangential local extremal points are directly induced by an appropriate conic extremal principle applied to the cone system $\{\Lambda_i\}$ at the origin. It is remarkable, however, that for *tangentially normally enclosed* cones $\{\Lambda_i\}$ we simultaneously ensure the fulfillment of the *limiting extremality conditions* of Definition 3.4(c) at the corresponding tangential extremal points. As shown in Section 3.3, this is the case of the contingent cone in finite dimensions and of the weak contingent cone in reflexive (and also in Asplund) spaces.

In this section we pay the main attention to deriving the *contingent* and *weak contingent extremal principle* involving the aforementioned extremality conditions for countable and finite systems of sets and finite-dimensional and infinite-dimensional spaces. Observe that in the case of countable collections of sets the results obtained are the first in the literature, while in the case of finite systems of sets they are independent of the those known before being applied to different notions of tangential extremal points; see the discussions in Section 3.1.

We begin with the contingent extremal principle for countable systems of arbitrary closed sets in finite-dimensional spaces.

Theorem 3.17 (contingent extremal principle for countable sets systems in finite dimensions). *Let* $\bar{x} \in \bigcap_{i=1}^{\infty} \Omega_i$ *be a contingent local extremal point of a countable system of closed sets* $\{\Omega_i\}_{i\in\mathbb{N}}$ *in* $\mathbb{R}^n$*. Assume that the contingent cones* $T(\bar{x};\Omega_i)$ *to* Ω_i *at* $\bar{x}$ *are nonoverlapping*

$$\bigcap_{i=1}^{\infty} \Big\{T(\bar{x};\Omega_i)\Big\} = \{0\}.$$

Then there are normal vectors

$$x_i^* \in N(0;\Lambda_i) \subset N(\bar{x};\Omega_i) \ \ \textit{for} \ \ \Lambda_i := T(\bar{x};\Omega_i) \ \ \textit{as} \ \ i \in \mathbb{N}$$

satisfying the extremality conditions in (3.3).

Proof. This result follows from combining Theorem 3.8 and Corollary 3.16. □

Consider further systems of finitely many sets $\{\Omega_1,\ldots,\Omega_m\}$ in Asplund spaces and derive for them the weak contingent extremal principle. Recall that a set $\Omega \subset X$ is *sequentially normally compact* (SNC) at $\bar{x} \in \Omega$ if for any sequence $\{(x_k,x_k^*)\}_{k\in\mathbb{N}} \subset \Omega \times X^*$ we have the implication

$$\big[x_k \to \bar{x},\ x_k^* \stackrel{w^*}{\to} 0 \ \text{ with } \ x_k^* \in \widehat{N}(x_k;\Omega),\ k \in \mathbb{N}\big] \Longrightarrow \|x_k^*\| \to 0 \ \text{ as } \ k \to \infty.$$

In [47, Subsection 1.1.4], the reader can find a number of efficient conditions ensuring the SNC property, which holds in rather broad infinite-dimensional settings. The next proposition shows that the SNC property of TAN sets is inherent by their weak contingent cones.

Proposition 3.18 (SNC property of weak contingent cones). *Let Ω be a closed subset of an Asplund space X satisfying the tangential approximate normality property at $\bar{x} \in \Omega$. Then the weak contingent cone $T_w(\bar{x};\Omega)$ is SNC at the origin provided that Ω is SNC at $\bar{x}$. In particular, in reflexive spaces the SNC property of a closed subset Ω at $\bar{x}$ unconditionally implies the SNC property of its weak contingent cone $T_w(\bar{x};\Omega)$ at the origin.*

Proof. To justify the SNC property of $\Lambda_w := T_w(\bar{x};\Omega)$ at the origin, take sequences $d_k \to 0$ and $x_k^* \in \widehat{N}(d_k;\Lambda_w)$ satisfying $x_k^* \stackrel{w^*}{\to} 0$ as $k \to \infty$. Using the TAN property of Ω at $\bar{x}$ and following the proof of Theorem 3.15, we find sequences $\varepsilon_k \downarrow 0$ and $\widetilde{x}_k \stackrel{\Omega}{\to} \bar{x}$ such that

$$x_k^* \in \widehat{N}(\widetilde{x}_k;\Omega) + \varepsilon_k I\!B^* \quad \text{for all} \quad k \in \mathbb{N}.$$

Hence there are $\widetilde{x}_k^* \in \widehat{N}(\widetilde{x}_k;\Omega)$ with $\|\widetilde{x}_k^* - x_k^*\| \le \varepsilon_k$, which implies that $\widetilde{x}_k^* \stackrel{w^*}{\to} 0$ as $k \to \infty$. By the SNC property of Ω at $\bar{x}$ we get that $\|\widetilde{x}_k^*\| \to 0$, which yields in turn that $\|x_k^*\| \to 0$ as $k \in \infty$. This justifies the SNC property of Λ_w at the origin. The second assertion of this proposition immediately follows from Proposition 3.14. □

Now we are ready to establish the weak contingent extremal principle for systems of finitely many closed subsets of Asplund spaces in both approximate and exact forms.

Theorem 3.19 (weak contingent extremal principle for finite systems of sets in Asplund spaces). *Let $\bar{x} \in \bigcap_{i=1}^m \Omega_i$ be a weak contingent local extremal point of the system $\{\Omega_1,\ldots,\Omega_m\}$ of closed sets in an Asplund space X. Assume that all the sets Ω_i, $i = 1,\ldots,m$,* *have the TAN property at $\bar{x}$, which is automatic in reflexive spaces. Then the following versions of the weak contingent extremal principle hold:*

(i) Approximate version*: for any $\varepsilon > 0$ there are $x_i^* \in N(\bar{x};\Omega_i)$ as $i = 1,\ldots,m$ satisfying*

$$\|x_1^* + \ldots + x_m^*\| \le \varepsilon \quad \textit{and} \quad \|x_1^*\| + \ldots + \|x_m^*\| = 1. \tag{3.20}$$

(ii) Exact version*: if in addition all but one of the sets Ω_i as $i = 1,\ldots,m$ are SNC at $\bar{x}$, then there exist $x_i^* \in N(\bar{x};\Omega_i)$ as $i = 1,\ldots,m$ satisfying*

$$x_1^* + \ldots + x_m^* = 0 \quad \textit{and} \quad \|x_1^*\| + \ldots + \|x_m^*\| = 1. \tag{3.21}$$

Proof. It follows from Proposition 3.2 that the cone system $\{\Lambda_w^i = T_w(\bar{x};\Omega_i)\}$ as $i = 1,\ldots,m$ is extremal at the origin in the conventional sense (1.2). Applying to it the approximate extremal principle from [47, Theorem 2.20], for any $\varepsilon > 0$ we find $x_i \in \Lambda_w^i$ and $x_i^* \in \widehat{N}(x_i;\Lambda_w^i)$ as $i = 1,...,m$ such that all the relationships in (3.20) hold. Then

$$x_i^* \in \widehat{N}(x_i;\Lambda_w^i) \subset N(0;\Lambda_w^i) \subset N(\bar{x};\Omega_i), \quad i = 1,\ldots,m,$$

by Proposition 2.1 and Theorem 3.15, which justifies assertion (i).

Now to justify (ii), observe that all but one of the cones Λ_w^i are SNC at the origin by Proposition 3.18. Thus (ii) follows from [47, Theorem 2.22] and Theorem 3.15. □

3.5 Fréchet Normals to Countable Intersections of Cones

In this section we present applications of the conic extremal principle established in Theorem 3.8 to deriving several representations, under appropriate assumptions, of Fréchet normals to *countable intersections* of cones in finite-dimensional spaces. These calculus results are certainly of their independent interest while their are largely employed to problems of semi-infinite programming and multiobjective optimization.

To begin with, we introduce the following qualification condition for countable systems of cones formulated in terms of limiting normals (2.5), which plays a significant role in deriving the results of this section as well as in the subsequent applications.

Definition 3.20 (normal qualification condition for countable systems of cones). *Let* $\{\Lambda_i\}_{i\in\mathbb{N}}$ *be a countable system of closed cones in* X. *We say that it satisfies the* NORMAL QUALIFICATION CONDITION *at the origin if*

$$\Big[\sum_{i=1}^{\infty} x_i^* = 0,\ \ x_i^* \in N(0;\Lambda_i)\Big] \Longrightarrow \big[x_i^* = 0,\ \ i \in \mathbb{N}\big]. \tag{3.22}$$

This definition corresponds to the normal qualification condition of [47] for finite systems of sets; see the discussions and various applications of the latter condition therein. In this section we use the normal qualification condition of Definition 3.20 to represent Fréchet normals to countable intersections of cones in terms of limiting normals to each of the sets involved. Let us start with the following "fuzzy" intersection rule at the origin.

Theorem 3.21 (fuzzy intersection rule for Fréchet normals to countable intersections of cones). *Let* $\{\Lambda_i\}_{i\in\mathbb{N}}$ *be a countable system of arbitrary closed cones in* $\mathbb{R}^n$ *satisfying the normal qualification condition* (3.22). *Then given a Fréchet normal* $x^* \in \widehat{N}\big(0;\bigcap_{i=1}^{\infty}\Lambda_i\big)$ *and a number* $\varepsilon > 0$, *there are limiting normals* $x_i^* \in N(0;\Lambda_i)$ *as* $i \in \mathbb{N}$ *such that*

$$x^* \in \sum_{i=1}^{\infty} \frac{1}{2^i} x_i^* + \varepsilon I\!B^*. \tag{3.23}$$

Proof. Fix $x^* \in \widehat{N}\Big(0; \bigcap_{i=1}^{\infty} \Lambda_i\Big)$ and $\varepsilon > 0$. By definition (2.3) of Fréchet normals we have

$$\langle x^*, x\rangle - \varepsilon\|x\| < 0 \ \text{ whenever } \ x \in \bigcap_{i=1}^{\infty} \Lambda_i \setminus \{0\}. \tag{3.24}$$

Define a countable system of closed cones in $\mathbb{R}^{n+1}$ by

$$O_1 := \big\{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R}\big|\ x \in \Lambda_1,\ \alpha \le \langle x^*, x\rangle - \varepsilon\|x\|\big\} \ \text{ and } \ O_i := \Lambda_i \times \mathbb{R}_+ \ \text{ for } \ i = 2, 3, \ldots. \tag{3.25}$$

Let us check that all the assumptions for the validity of the conic extremal principle in Theorem 3.8 are satisfied for the system $\{O_i\}_{i\in\mathbb{N}}$. Picking any $(x, \alpha) \in \bigcap_{i=1}^{\infty} O_i$, we have $x \in \bigcap_{i=1}^{\infty} \Lambda_i$ and $\alpha \ge 0$ from the construction of Ω_i as $i \ge 2$. This implies in fact that $(x, \alpha) = (0, 0)$. Indeed, supposing $x \ne 0$ gives us by (3.24) that

$$0 \le \alpha \le \langle x^*, x\rangle - \varepsilon\|x\| < 0,$$

which is a contradiction. On the other hand, the inclusion $(0, \alpha) \in O_1$ yields that $\alpha \le 0$ by the construction of O_1, i.e., $\alpha = 0$. Thus the *nonoverlapping condition*

$$\bigcap_{i=1}^{\infty} O_i = \{(0, 0)\}$$

holds for $\{O_i\}_{i\in\mathbb{N}}$. Similarly we check that

$$\Big(O_1 - (0, \gamma)\Big) \cap \bigcap_{i=2}^{\infty} O_i = \emptyset \ \text{ for any fixed } \ \gamma > 0, \tag{3.26}$$

i.e., $\{O_i\}_{i\in\mathbb{N}}$ is a *conic extremal system* at the origin. Indeed, violating (3.26) means the existence of $(x,\alpha)\in\mathbb{R}^n\times\mathbb{R}$ such that

$$(x,\alpha)\in\Big[O_1-(0,\gamma)\Big]\cap\bigcap_{i=2}^{\infty}O_i,$$

which implies that $x\in\bigcap_{i=1}^{\infty}O_i$ and $\alpha\geq 0$. Then by the construction of O_1 in (3.25) we get

$$\gamma+\alpha\leq\langle x^*,x\rangle-\varepsilon\|x\|\leq 0,$$

a contradiction due the positivity of γ in (3.26).

Applying now the second conclusion of Theorem 3.8 to the system $\{O_i\}_{i\in\mathbb{N}}$ gives us the pairs $(w_i,\alpha_i)\in O_i$ and $(x_i^*,\lambda_i)\in\widehat{N}\big((w_i,\alpha_i);O_i\big)$ as $i\in\mathbb{N}$ satisfying the relationships

$$\sum_{i=1}^{\infty}\frac{1}{2^i}\big(x_i^*,\lambda_i\big)=0\quad\text{and}\quad\sum_{i=1}^{\infty}\frac{1}{2^i}\big\|(x_i^*,\lambda_i)\big\|^2=1. \tag{3.27}$$

It immediately follows from the constructions of O_i as $i\geq 2$ in (3.25) that $\lambda_i\leq 0$ and $x_i^*\in\widehat{N}(w_i;\Lambda_i)$; thus $x_i^*\in N(0;\Lambda_i)$ for $i=2,3,\ldots$ by Proposition 2.1. Furthermore, we get

$$\limsup_{(x,\alpha)\overset{O_1}{\to}(w_1,\alpha_1)}\frac{\langle x_1^*,x-w_1\rangle+\lambda_1(\alpha-\alpha_1)}{\|x-w_1\|+|\alpha-\alpha_1|}\leq 0 \tag{3.28}$$

by the definition of Fréchet normals to O_1 at $(w_1,\alpha_1)\in O_1$ with $\lambda_1\geq 0$ and

$$\alpha_1\leq\langle x^*,w_1\rangle-\varepsilon\|w_1\| \tag{3.29}$$

by the construction of O_1. Examine next the two possible cases in (3.27): $\lambda_1=0$ and $\lambda_1>0$.

Case 1: $\lambda_1=0$. If inequality (3.29) is strict in this case, find a neighborhood U of w_1 such that $\alpha_1 < \langle x^*, x\rangle - \varepsilon\|x\|$ for all $x \in U$, which ensures that $(x, \alpha_1) \in O_1$ for all $x \in \Lambda_1 \cap U$. Substituting (x, α_1) into (3.28) gives us

$$\limsup_{x \xrightarrow{\Lambda_1} w_1} \frac{\langle x_1^*, x - w_1\rangle}{\|x - w_1\|} \leq 0,$$

which means that $x_1^* \in \widehat{N}(w_1; \Lambda_1)$. If (3.29) holds as equality, we put $\alpha := \langle x^*, x\rangle - \varepsilon\|x\|$ and get

$$|\alpha - \alpha_1| = \big|\langle x^*, x - w_1\rangle + \varepsilon(\|w_1\| - \|x\|)\big| \leq \big(\|x^*\| + \varepsilon\big)\|x - w_1\|.$$

Furthermore, it follows from (3.28) that

$$\limsup_{(x,\alpha) \xrightarrow{O_1} (w_1, \alpha_1)} \frac{\langle x_1^*, x - w_1\rangle}{\|x - w_1\| + |\alpha - \alpha_1|} \leq 0.$$

Thus for any $\nu > 0$ sufficiently small and α chosen above, we have

$$\langle x_1^*, x - w_1\rangle \leq \nu\big(\|x - w_1\| + |\alpha - \alpha_1|\big) \leq \nu\big(1 + \|x^*\| + \varepsilon\big)\|x - w_1\|$$

whenever $x \in \Lambda_1$ is sufficiently closed to w_1. The latter yields that

$$\limsup_{x \xrightarrow{\Lambda_1} w_1} \frac{\langle x_1^*, x - w_1\rangle}{\|x - w_1\|} \leq 0, \quad \text{i.e.,} \quad x_1^* \in \widehat{N}(w_1; \Lambda_1).$$

Thus in both cases of the strict inequality and equality in (3.29), we justify that $x_1^* \in \widehat{N}(w_1; \Lambda_1)$ and thus $x_1^* \in N(0; \Lambda_1)$ by Proposition 2.1. Summarizing the above discussions gives us

$$x_i^* \in N(0; \Lambda_i) \ \text{ and } \ \lambda_i = 0 \ \text{ for all } \ i \in \mathbb{N}$$

in Case 1 under consideration. Hence it follows from (3.27) that there are $\widetilde{x}_i^* := (1/2^i)x_i^* \in N(0;\Lambda_i)$ as $i \in \mathbb{N}$, not equal to zero simultaneously, satisfying

$$\sum_{i=1}^{\infty} \widetilde{x}_i^* = 0.$$

This contradicts the normal qualification condition (3.22) and thus shows that the case of $\lambda_1 = 0$ is actually *not possible* in (3.29).

Case 2: $\lambda_1 > 0$. If inequality (3.29) is strict, put $x = w_1$ in (3.28) and get

$$\limsup_{\alpha \to \alpha_1} \frac{\lambda_1(\alpha - \alpha_1)}{|\alpha - \alpha_1|} \le 0.$$

That yields $\lambda_1 = 0$, a contradiction. Hence it remains to consider the case when (3.29) holds as equality. To proceed, take $(x, \alpha) \in O_1$ satisfying

$$x \in \Lambda_1 \setminus \{w_1\} \ \text{ and } \ \alpha = \langle x^*, x\rangle - \varepsilon\|x\|.$$

By the equality in (3.29) we have

$$\alpha - \alpha_1 = \langle x^*, x - w_1\rangle + \varepsilon(\|w_1\| - \|x\|) \ \text{ and thus } \ |\alpha - \alpha_1| \le (\|x^*\| + \varepsilon)\|x - w_1\|.$$

On the other hand, it follows from (3.28) that for any $\gamma > 0$ sufficiently small there exists a neighborhood V of w_1 such that

$$\langle x_1^*, x - w_1\rangle + \lambda_1(\alpha - \alpha_1) \le \lambda_1\gamma\varepsilon\big(\|x - w_1\| + |\alpha - \alpha_1|\big) \tag{3.30}$$

whenever $x \in \Lambda_1 \cap V$. Substituting (x, α) with $x \in \Lambda_1 \cap V$ into (3.30) gives us

$$\begin{aligned} \langle x_1^*, x - w_1 \rangle + \lambda_1(\alpha - \alpha_1) &= \langle x_1^* + \lambda_1 x^*, x - w_1 \rangle + \lambda_1 \varepsilon(\|w_1\| - \|x\|) \\ &\le \lambda_1 \gamma \varepsilon(\|x - w_1\| + |\alpha - \alpha_1|) \\ &\le \lambda_1 \gamma \varepsilon\big[\|x - w_1\| + (\|x^*\| + \varepsilon)\|x - w_1\|\big] \\ &= \lambda_1 \gamma \varepsilon\big(1 + \|x^*\| + \varepsilon\big)\|x - w_1\|. \end{aligned}$$

It follows from the above that for small $\gamma > 0$ we have

$$\langle x_1^* + \lambda_1 x^*, x - w_1 \rangle + \lambda_1 \varepsilon(\|w_1\| - \|x\|) \le \lambda_1 \varepsilon \|x - w_1\|$$

and thus arrive at the estimates

$$\langle x_1^* + \lambda_1 x^*, x - w_1 \rangle \le \lambda_1 \varepsilon \|x - w_1\| + \lambda_1 \varepsilon(\|x\| - \|w_1\|) \le 2\lambda_1 \varepsilon \|x - w_1\|$$

for all $x \in \Lambda_1 \cap V$. The latter implies by definition (2.3) of ε-normals that

$$x_1^* + \lambda_1 x^* \in \widehat{N}_{2\lambda_1\varepsilon}(w_1; \Lambda_1). \tag{3.31}$$

Furthermore, it is easy to observe from the above choice of λ_1 and the structure of O_1 in (3.25) that $\lambda_1 \le 2 + 2\varepsilon$. Employing now the representation of ε-normals in (3.31) from [47, formula (2.51)] held in finite dimensions, we find $v \in \Lambda_1 \cap (w_1 + 2\lambda_1 \varepsilon I\!B)$ such that

$$x_1^* + \lambda_1 x^* \in \widehat{N}(v; \Lambda_1) + 2\lambda_1 \varepsilon I\!B^* \subset N(0; \Lambda_1) + 2\lambda_1 \varepsilon I\!B^*. \tag{3.32}$$

Since $\lambda_1 > 0$ in the case under consideration and by $-x_1^* = 2\sum_{i=2}^{\infty} \frac{1}{2^i} x_i^*$ due to the first equality in (3.27), it follows from (3.32) that

$$x^* \in N(0;\Lambda_1) + \frac{2}{\lambda_1}\sum_{i=2}^{\infty}\frac{1}{2^i}x_i^* + 2\varepsilon I\!B^*,$$

and hence there exists $\widetilde{x}_1^* \in N(0;\Lambda_1)$ such that

$$x^* \in \sum_{i=1}^{\infty}\frac{1}{2^i}\widetilde{x}_i^* + 2\varepsilon I\!B^* \ \text{ with } \ \widetilde{x}_i^* := \frac{2x_i^*}{\lambda_1} \in N(0;\Lambda_i) \ \text{ for } \ i = 2,3,\ldots.$$

This justifies (3.23) and completes the proof of the theorem. □

Our next result shows that we can put $\varepsilon = 0$ in representation (3.23) under an additional assumption on Fréchet normals to cone intersections.

Theorem 3.22 (refined representation of Fréchet normals to countable intersections of cones). *Let $\{\Lambda_i\}_{i\in\mathbb{N}}$ be a countable system of arbitrary closed cones in $\mathbb{R}^n$ satisfying the normal qualification condition* (3.22)*. Then for any Fréchet normal $x^* \in \widehat{N}\left(0;\bigcap_{i=1}^{\infty}\Lambda_i\right)$ satisfying*

$$\langle x^*, x\rangle < 0 \ \ \textit{whenever} \ \ x \in \bigcap_{i=1}^{\infty}\Lambda_i \setminus \{0\} \tag{3.33}$$

there are limiting normals $x_i^* \in N(0;\Lambda_i)$, $i = 1,2,\ldots$, such that

$$x^* = \sum_{i=1}^{\infty}\frac{1}{2^i}x_i^*. \tag{3.34}$$

Proof. Fix a Fréchet normal $x^* \in \widehat{N}\left(0;\bigcap_{i=1}^{\infty}\Lambda_i\right)$ satisfying condition (3.33) and construct a countable system of closed cones in $\mathbb{R}^n \times \mathbb{R}$ by

$$O_1 := \big\{(x,\alpha) \in \mathbb{R}^n \times \mathbb{R} \big|\ x \in \Lambda_1,\ \alpha \le \langle x^*, x\rangle\big\} \text{ and } O_i := \Lambda_i \times \mathbb{R}_+ \text{ for } i = 2,3,\ldots. \tag{3.35}$$

Similarly to the proof Theorem 3.21 with taking (3.33) into account, we can verify that all the assumptions of Theorem 3.8 hold. Applying the conic extremal principle from this theorem gives us pairs $(w_i, \alpha_i) \in O_i$ and $(x_i^*, \lambda_i) \in \widehat{N}\big((w_i,\alpha_i); O_i\big)$ such that the extremality conditions in (3.27) are satisfied. We obviously get $\lambda_i \le 0$ and $x_i^* \in \widehat{N}(w_i;\Lambda_i)$ for $i = 1,2,\ldots$, which ensures that $x_i^* \in N(0;\Lambda_i)$ as $i \ge 2$ by Proposition 2.1. It follows furthermore that for $i = 1$ the limiting inequality (3.28) holds. The latter implies by the structure of the set O_1 in (3.35) that

$$\lambda_1 \ge 0 \text{ and } \alpha_1 \le \langle x^*, w_1\rangle. \tag{3.36}$$

Similarly to the proof of Theorem 3.21 we consider the two possible cases $\lambda_1 = 0$ and $\lambda_1 > 0$ in (3.36) and show that the first case contradicts the normal qualification condition (3.22). In the second case we arrive at representation (3.34) based on the extremality conditions in (3.27) and the structures of the sets O_i in (3.35). □

The next theorem in this section provides constructive upper estimates of the Fréchet normal cone to countable intersections of closed cones in finite dimensions and of its interior via limiting normals to the sets involved at the origin.

Theorem 3.23 (Fréchet normal cone to countable intersections). *Let $\{\Lambda_i\}_{i\in\mathbb{N}}$ be a countable system of arbitrary closed cones in $\mathbb{R}^n$ satisfying the normal qualification condition*

(3.22)*, and let* $\Lambda := \bigcap_{i=1}^{\infty} \Lambda_i$*. Then we have the inclusions*

$$\text{int}\, \widehat{N}(0;\Lambda) \subset \Big\{ \sum_{i=1}^{\infty} x_i^* \Big|\ x_i^* \in N(0;\Lambda_i) \Big\}, \tag{3.37}$$

$$\widehat{N}(0;\Lambda) \subset \text{cl} \Big\{ \sum_{i \in I} x_i^* \Big|\ x_i^* \in N(0;\Lambda_i),\ I \in \mathcal{L} \Big\}, \tag{3.38}$$

where $\mathcal{L}$ *stands for the collection of all finite subsets of the natural series* $\mathbb{N}$*.*

Proof. First we justify inclusion (3.37) assuming without loss of generality that $\text{int}\, N(0;\Lambda) \neq \emptyset$. Pick any $x^* \in \text{int}\, \widehat{N}(0;\Lambda)$ and also $\gamma > 0$ such that $x^* + 3\gamma I\!B^* \subset \widehat{N}(0;\Lambda)$. Then for any $x \in \Lambda \setminus \{0\}$ find $z^* \in \mathbb{R}^n$ satisfying the relationships

$$\|z^*\| = 2\gamma \ \text{ and } \ \langle z^*, x \rangle < -\gamma \|x\|.$$

Since $x^* - z^* \in x^* + 3\gamma I\!B^* \subset \widehat{N}(0;\Lambda)$, we have $\langle x^* - z^*, x \rangle \leq 0$ and hence

$$\langle x^*, x \rangle = \langle x^* - z^*, x \rangle + \langle z^*, x \rangle < -\gamma \|x\| < 0.$$

This allows us to employ Theorem 3.22 and thus justify the first inclusion (3.37).

To prove the remaining inclusion (3.38), pick pick $x^* \in \widehat{N}(0;\Lambda)$ and for any fixed $\varepsilon > 0$ apply Theorem 3.21. In this way we find $x_i^* \in N(0;\Lambda_i)$, $i \in \mathbb{N}$, such that $x^* \in \displaystyle\sum_{i=1}^{\infty} \frac{1}{2^i} x_i^* + \varepsilon I\!B^*$.

Since $\varepsilon > 0$ was chosen arbitrarily, it follows that

$$x^* \in A := \text{cl} \left\{ \sum_{i=1}^{\infty} \frac{1}{2^i} x_i^* \Big|\ x_i^* \in N(0;\Lambda_i) \right\}.$$

Let us finally justify the inclusion

$$A \subset \operatorname{cl} C \quad \text{with} \quad C := \Big\{ \sum_{i \in I} x_i^* \Big| \; x_i^* \in N(0; \Lambda_i), \; I \in \mathcal{L} \Big\}.$$

To proceed, pick $z^* \in A$ and for any fixed $\varepsilon > 0$ find $x_i^* \in N(0; \Lambda_i)$ satisfying

$$\Big\| z^* - \sum_{i=1}^{\infty} \frac{1}{2^i} x_i^* \Big\| \le \frac{\varepsilon}{2}.$$

Then choose a number $k \in \mathbb{N}$ so large that

$$\Big\| z^* - \sum_{i=1}^{k} \frac{1}{2^i} x_i^* \Big\| \le \varepsilon.$$

Since $\displaystyle\sum_{i=1}^{k} \frac{1}{2^i} x_i^* \in C$, we get $(z^* + \varepsilon I\!B^*) \cap C \neq \emptyset$, which means that $z^* \in \operatorname{cl} C$. This justifies (3.38) and completes the proof of the theorem. □

3.6 Tangents and Normals to Infinite Intersections of Sets

The main purpose of this section is to derive calculus rules for representing generalized normals to countable intersections of arbitrary closed sets under appropriate qualification conditions. Besides employing the tangential extremal principle, one of the major ingredients in our approach is relating calculus rules for generalized normals to countable set intersections with the so-called "conical hull intersection property" defined in terms of tangents to sets, which was intensively studied and applied in the literature for the case of finite intersections of convex sets; see, e.g., [6, 11, 16, 19, 41] and the references therein. In what follows, we keep the terminology of convex analysis (that goes back probably to [11]) replacing the tangent and normal cones therein by the nonconvex extension (2.2) and (2.6).

Definition 3.24 (CHIP for countable intersections). *A set system* $\{\Omega_i\}_{i\in\mathbb{N}}$ *in* $\mathbb{R}^n$ *is said to have the* CONICAL HULL INTERSECTION PROPERTY (*CHIP*) *at* $\bar{x} \in \bigcap_{i=1}^{\infty} \Omega_i$ *if*

$$T\Big(\bar{x}; \bigcap_{i=1}^{\infty} \Omega_i\Big) = \bigcap_{i=1}^{\infty} T(\bar{x}; \Omega_i). \tag{3.39}$$

In convex analysis and its applications the CHIP is often related to the so-called "strong CHIP" for finite set intersections expressed via the normal cone to the convex sets in question; see also [40] for infinite intersections of convex sets. Following this terminology in the case of infinite intersections of nonconvex sets, we say that a countable system of sets $\{\Omega_i\}_{i\in\mathbb{N}}$ has the *strong conical hull intersection property* (or the *strong CHIP*) at $\bar{x} \in \bigcap_{i=1}^{\infty} \Omega_i$ if

$$N\Big(\bar{x}; \bigcap_{i=1}^{\infty} \Omega_i\Big) = \Big\{ \sum_{i\in I} x_i^* \Big| \; x_i^* \in N(\bar{x}; \Omega_i), \; I \in \mathcal{L} \Big\}. \tag{3.40}$$

When all the sets Ω_i as $i \in \mathbb{N}$ are *convex* in (3.40), the strong CHIP of the system $\{\Omega_i\}_{i\in\mathbb{N}}$ can be equivalently written in the form

$$N\Big(\bar{x}; \bigcap_{i=1}^{\infty} \Omega_i\Big) = \operatorname{co} \bigcup_{i=1}^{\infty} N(\bar{x}; \Omega_i). \tag{3.41}$$

We say that a countable set system $\{\Omega_i\}_{i\in\mathbb{N}}$ has the *asymptotic strong CHIP* at $\bar{x} \in \bigcap_{i=1}^{\infty} \Omega_i$ if the latter representation is replaced by

$$N\Big(\bar{x}; \bigcap_{i=1}^{\infty} \Omega_i\Big) = \operatorname{cl} \operatorname{co} \bigcup_{i=1}^{\infty} N(\bar{x}; \Omega_i). \tag{3.42}$$

The next result shows the *equivalence* between the CHIP and the asymptotic strong CHIP for intersections of convex sets in finite dimensions. It follows from the proof that this equivalence holds for *arbitrary* intersections of convex sets, not only for countable ones studied in this chapter.

Theorem 3.25 (characterization of CHIP for intersections of convex sets). *Let $\{\Omega_i\}_{i\in\mathbb{N}}$ be a system of convex sets in $\mathbb{R}^n$, and let $\bar{x}\in\bigcap_{i=1}^{\infty}\Omega_i$. The following are equivalent:*

(a) *The system $\{\Omega_i\}_{i\in\mathbb{N}}$ has the CHIP at $\bar{x}$.*

(b) *The system $\{\Omega_i\}_{i\in\mathbb{N}}$ has the asymptotic strong CHIP at $\bar{x}$.*

In particular, the strong CHIP implies the CHIP but not vice versa.

Proof. Observe first that for convex sets in finite dimensions, in addition to the duality property (2.4) with $\widehat{N}(\bar{x};\Omega)$ replaced by $N(\bar{x};\Omega)$, we have the reverse duality representation

$$T(\bar{x};\Omega)=N^*(\bar{x};\Omega):=\big\{v\in\mathbb{R}^n\big|\ \langle x^*,v\rangle\le 0\ \text{ for all }\ x^*\in N(\bar{x};\Omega)\big\}. \tag{3.43}$$

Let us now justify the equality

$$\Big(\bigcap_{i=1}^{\infty}T(\bar{x};\Omega_i)\Big)^*=\operatorname{cl}\operatorname{co}\bigcup_{i=1}^{\infty}N(\bar{x};\Omega_i). \tag{3.44}$$

The inclusion "$\supset$" follows from (2.4) by the observation $N(\bar{x};\Omega_i)=T^*(\bar{x};\Omega_i)\subset\Big(\bigcap_{i=1}^{\infty}T(\bar{x};\Omega_i)\Big)^*$ due the closedness and convexity of the polar set on the right-hand side of the latter inclusion.

To prove the opposite inclusion "$\subset$" in (3.44), pick some $x^*\notin\operatorname{cl}\operatorname{co}\bigcup_{i=1}^{\infty}N(\bar{x};\Omega_i)$. Then the classical separation theorem for convex sets ensures the existence of a vector $v\in\mathbb{R}^n$ such that

$$\langle x^*,v\rangle>0\ \text{ and }\ \langle u^*,v\rangle\le 0\ \text{ for all }\ u^*\in\operatorname{cl}\operatorname{co}\bigcup_{i=1}^{\infty}N(\bar{x};\Omega_i). \tag{3.45}$$

Hence for each $i\in\mathbb{N}$ we get $\langle u^*,v\rangle\le 0$ whenever $u^*\in N(\bar{x};\Omega_i)$, which implies that $v\in$

$N^*(\bar{x};\Omega_i)$ and therefore $v \in T(\bar{x};\Omega_i)$ by (3.43). This gives us $v \in \bigcap_{i=1}^{\infty} T(\bar{x};\Omega_i)$, and so $x^* \notin \Big(\bigcap_{i=1}^{\infty} T(\bar{x};\Omega_i)\Big)^*$ due to $\langle x^*, v\rangle > 0$ in (3.45). It justifies the inclusion "$\subset$" in (3.44), which holds as equality. Taking into account that the set $\cap_{i=1}^{\infty} T(\bar{x};\Omega_i)$ is a closed convex cone and agrees hence with its second dual, we get

$$\bigcap_{i=1}^{\infty} T(\bar{x};\Omega_i) = \Big(\mathrm{cl}\,\mathrm{co} \bigcup_{i=1}^{\infty} N(\bar{x};\Omega_i)\Big)^*. \tag{3.46}$$

Assuming that the CHIP in (a) holds and employing (2.4) and (3.44) for the set intersection $\Omega := \bigcap_{i=1}^{\infty} \Omega_i$ allow us to arrive at the equalities

$$N(\bar{x};\Omega) = T^*(\bar{x};\Omega) = \Big(\bigcap_{i=1}^{\infty} T(\bar{x};\Omega_i)\Big)^* = \mathrm{cl}\,\mathrm{co} \bigcup_{i=1}^{\infty} N(\bar{x};\Omega_i),$$

which give the asymptotic strong CHIP in (b). Conversely, assume that (b) holds. Then employing (3.43) and (3.46) implies the relationships

$$T(\bar{x};\Omega) = N^*(\bar{x};\Omega) = \Big(\mathrm{cl}\,\mathrm{co} \bigcup_{i=1}^{\infty} N(\bar{x};\Omega_i)\Big)^* = \bigcap_{i=1}^{\infty} T(\bar{x};\Omega_i),$$

which ensure the fulfillment of the CHIP in (a) and thus establish the equivalence the properties in (a) and (b). Since the strong CHIP implies the asymptotic strong CHIP due to the closedness of $N(\bar{x};\Omega)$, it also implies the CHIP. The converse implication does not hold even for finitely many sets; counterexamples are presented, in particular, in [6, 19]. $\square$

The following simple consequence of Theorem 3.25 computes the normal cone to set of feasible solutions in linear semi-infinite programming with countable inequality constraints; cf. [9].

Corollary 3.26 (normal cone to sets of feasible solutions of linear semi-infinite programs with countable constraints). *Consider the set*

$$\Omega := \big\{x \in \mathbb{R}^n \big|\ \langle a_i, x\rangle \le 0,\ i \in \mathbb{N}\big\}, \tag{3.47}$$

where the vectors $a_i \in \mathbb{R}^n$ are fixed. Then the normal cone to Ω at the origin is computed by

$$N(0;\Omega) = \mathrm{cl}\,\mathrm{co}\Big[\bigcup_{i=1}^{\infty}\big\{\lambda a_i \big|\ \lambda \ge 0\big\}\Big]. \tag{3.48}$$

Proof. It is easy to see that the set (3.47) is represented as a countable intersection of sets having the CHIP. Furthermore, the asymptotic strong CHIP for this system is obviously (3.48). Thus the result follows immediately from Theorem 3.25. □

There are also interesting connections of Corollary 3.26 with the results of [24, Theorem 5.3(i)] and with the so-called "local Farkas-Minkowski qualification condition" for infinite systems of linear inequalities [55], which happens to be equivalent to the strong CHIP in this setting. The reader can find more discussions on related conditions for infinite convex inequality systems in Section 3.7.

Now let us show that the CHIP may be violated in rather simple situations involving finite and infinite intersections of convex sets defined by inequalities with convex functions.

Example 3.27 (failure of CHIP for finite and infinite intersections of convex sets).

(i) First consider the two convex sets

$$\Omega_1 := \big\{(x_1,x_2) \in \mathbb{R}^2 \big|\ x_2 \ge x_1^2\big\} \ \text{ and } \ \Omega_2 := \big\{(x_1,x_2) \in \mathbb{R}^2 \big|\ x_2 \le -x_1^2\big\}$$

and their intersection at $\bar{x} = (0,0)$. We have

$$\Omega_1 \cap \Omega_2 = \{\bar{x}\}, \;\; T(\bar{x};\Omega_1) = \mathbb{R} \times \mathbb{R}_+, \;\text{ and }\; T(\bar{x};\Omega_2) = \mathbb{R} \times \mathbb{R}_-.$$

Thus the CHIP does not hold in this case, since

$$T(\bar{x};\Omega_1 \cap \Omega_2) = \{(0,0)\} \neq T(\bar{x};\Omega_1) \cap T(\bar{x};\Omega_2) = \mathbb{R} \times \{0\}.$$

(ii) In the next case we have the CHIP violation for the countable intersection of convex sets, with the intersection set having nonempty interior. For each $i \in \mathbb{N}$, define $\varphi_i(x) := ix^2$ if $x < 0$ and $\varphi_i(x) := 0$ if $x \geq 0$. Let $\Omega_i := \operatorname{epi} \varphi_i$ and $\bar{x} = (0,0)$. It is easy to see that

$$\bigcap_{i=1}^{\infty} \Omega_i = \mathbb{R}_+ \times \mathbb{R}_+ \;\text{ and }\; T(\bar{x},\Omega_i) = \mathbb{R} \times \mathbb{R}_+ \;\text{ for }\; i \in \mathbb{N}.$$

It gives therefore the relationships

$$T\Big(\bar{x}, \bigcap_{i=1}^{\infty} \Omega_i\Big) = \mathbb{R}_+ \times \mathbb{R}_+ \neq \bigcap_{i=1}^{\infty} T(\bar{x};\Omega_i) = \mathbb{R} \times \mathbb{R}_+, \quad i \in \mathbb{N},$$

which show that the CHIP fails for this system of sets at the chosen point $\bar{x} = (0,0)$.

Of course, we cannot expect to extend the equivalence of Theorem 3.25 to intersections of nonconvex sets. In what follows we are mainly interested in obtaining calculus rules for generalized normals as in the strong CHIP using the nonconvex CHIP from Definition 3.24 (i.e., a calculus rule for tangents) as an appropriate assumption together with additional qualification conditions. Observe that the implication CHIP $\Longrightarrow$ strong CHIP does not hold even for finite intersections of convex sets; see Theorem 3.25.

To implement this strategy, we first intend to obtain some sufficient conditions for the CHIP of countable intersections of nonconvex sets. Note that a number of sufficient conditions for the CHIP has been proposed for finite intersections of convex sets, where convex interpolation techniques play a particularly important role; see [6, 11, 16, 41] and the references therein. However, such techniques do not seem to be useful in nonconvex settings. To proceed in deriving sufficient conditions for the CHIP of countable nonconvex intersections, we explore some other possibilities.

Let us start with extending the concept and techniques of linear regularity in the direction of [6, 41, 65] to the case of infinite nonconvex systems; cf. various results and discussions therein on particular cases of linear regularity and its applications. Given a countable system of closed sets $\{\Omega_i\}_{i\in\mathbb{N}}$, we say that it is *linearly regular* at $\bar{x}\in\Omega:=\bigcap_{i=1}^{\infty}\Omega_i$ if there exist a neighborhood U of $\bar{x}$ and a number $C>0$ such that

$$\mathrm{dist}\,(x;\Omega)\le C\sup_{i\in\mathbb{N}}\big\{\mathrm{dist}\,(x;\Omega_i)\big\}\ \text{ for all }\ x\in U. \tag{3.49}$$

In the next proposition we denote for convenience the distance function $\mathrm{dist}(x;\Omega)$ by $d_\Omega(x)$ and employ the standard notion of equi-convergence for families of functions.

Proposition 3.28 (sufficient conditions for CHIP in terms of linear regularity). *Let $\{\Omega_i\}_{i\in\mathbb{N}}$ be a countable system of closed sets in $\mathbb{R}^n$ with the intersection $\Omega:=\bigcap_{i=1}^{\infty}\Omega_i$, and let $\bar{x}\in\Omega$. Assume that the system of sets $\{\Omega_i\}_{i\in\mathbb{N}}$ is linearly regular at $\bar{x}$ with some $C>0$ in* (3.49) *and that the family of functions $\{d_{\Omega_i}(\cdot)\}_{i\in\mathbb{N}}$ is equi-directionally differentiable at $\bar{x}$ in the sense that for any $h\in\mathbb{R}^n$ the functions*

$$\left\{\frac{d_{\Omega_i}(\bar{x}+th)}{t},\ i\in\mathbb{N}\right\}$$

of $t > 0$ converge as $t \downarrow 0$ to the corresponding directional derivatives $d'_{\Omega_i}(\bar{x}; h)$ uniformly in $i \in \mathbb{N}$. Then for all $h \in \mathbb{R}^n$ and the positive constant C from (3.49) we have the estimate

$$\operatorname{dist}(h; \Lambda) \le C \sup_{i \in \mathbb{N}} \big\{\operatorname{dist}(h; \Lambda_i)\big\} \ \ \textit{with} \ \ \Lambda := T(\bar{x}; \Omega) \ \ \textit{and} \ \ \Lambda_i := T(\bar{x}; \Omega_i) \ \ \textit{as} \ \ i \in \mathbb{N}. \tag{3.50}$$

In particular, the set system $\{\Omega_i\}_{i \in \mathbb{N}}$ satisfies the CHIP at $\bar{x}$.

Proof. Fixing $h \in \mathbb{R}^n$ and using definition (2.2) and [61, Exercise 4.8], we get

$$\operatorname{dist}(h; \Lambda) = \liminf_{t \downarrow 0} \operatorname{dist}\Big(h; \frac{\Omega - \bar{x}}{t}\Big) = \liminf_{t \downarrow 0} \frac{\operatorname{dist}(\bar{x} + th; \Omega)}{t}.$$

When t is small, the assumed linear regularity yields that

$$\frac{\operatorname{dist}(\bar{x} + th; \Omega)}{t} \le C \sup_{i \in \mathbb{N}} \frac{\operatorname{dist}(\bar{x} + th; \Omega_i)}{t}.$$

Applying further the equi-directional differentiability gives us

$$\frac{\operatorname{dist}(\bar{x} + th; \Omega_i)}{t} \to d'_{\Omega_i}(\bar{x}; h) = \operatorname{dist}(h; \Lambda_i) \ \ \text{uniformly in} \ \ i \ \ \text{as} \ \ t \downarrow 0,$$

i.e., for any $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $t \in (0, \delta)$ we have

$$\Big|\frac{\operatorname{dist}(\bar{x} + th; \Omega_i)}{t} - \operatorname{dist}(h; \Lambda_i)\Big| \le \varepsilon \ \ \text{for all} \ \ i \in \mathbb{N}.$$

Hence it holds for any $t \in (0, \delta)$ that

$$\sup_{i \in \mathbb{N}} \frac{\operatorname{dist}(\bar{x} + th; \Omega_i)}{t} \le \sup_{i \in \mathbb{N}} \big\{\operatorname{dist}(h; \Lambda_i)\big\} + \varepsilon.$$

Combining all the above, we get the estimates

$$\operatorname{dist}(h;\Lambda) \le C \liminf_{t\downarrow 0} \sup_{i\in\mathbb{N}} \frac{\operatorname{dist}(\bar{x}+th;\Omega_i)}{t} \le C \sup_{i\in\mathbb{N}} \big\{\operatorname{dist}(h;\Lambda_i)\big\} + C\varepsilon,$$

which imply (3.50), since ε was chosen arbitrarily. Finally, the CHIP of the system $\{\Omega_i\}_{i\in\mathbb{N}}$ at $\bar{x}$ follows directly from (3.50) and the definitions. □

Now we present a consequence of Proposition 3.28 that simplifies the verification of linear regularity for countable set systems.

Corollary 3.29 (CHIP via simplified linear regularity). *Let $\{\Omega_i\}_{i\in\mathbb{N}}$ be a countable system of closed subsets in $\mathbb{R}^n$, and let $\bar{x} \in \Omega = \bigcap_{i=1}^{\infty} \Omega_i$. Assume that the family $\{d(\cdot;\Omega_i)\}_{i\in\mathbb{N}}$ is equi-directionally differentiable at $\bar{x}$ and that there are numbers $C > 0$, $j \in \mathbb{N}$, and a neighborhood U of $\bar{x}$ such that*

$$\operatorname{dist}(x;\Omega) \le C \sup_{i\ne j} \big\{\operatorname{dist}(x;\Omega_i)\big\} \quad \text{for all } x \in \Omega_j \cap U.$$

Then the set system $\{\Omega_i\}_{i\in\mathbb{N}}$ satisfies the CHIP at $\bar{x}$.

Proof. Employing Proposition 3.28, it suffices to show that the set system $\{\Omega_i\}_{i\in\mathbb{N}}$ is linearly regular at $\bar{x}$. To proceed, take $r > 0$ so small that $\operatorname{dist}(x;\Omega) \le C \sup_{i\ne j} \big\{\operatorname{dist}(x;\Omega_i)\big\}$ for all $x \in \Omega_j \cap (\bar{x}+3r\mathbb{B})$. Since the distance function is nonexpansive, for every $y \in \Omega_j \cap (\bar{x}+3r\mathbb{B})$ and $x \in \mathbb{R}^n$ we have

$$\begin{aligned} 0 &\le C \sup_{i\ne j} \big\{\operatorname{dist}(y;\Omega_i)\big\} - \operatorname{dist}(y;\Omega) \le C \sup_{i\ne j} \Big(\big\{\operatorname{dist}(x;\Omega_i)\big\} + \|x-y\|\Big) - \operatorname{dist}(x;\Omega) + \|x-y\| \\ &\le C \sup_{i\ne j} \big\{\operatorname{dist}(x;\Omega_i)\big\} - \operatorname{dist}(x;\Omega) + (C+1)\|x-y\|. \end{aligned}$$

Then it follows for all $x \in \mathbb{R}^n$ that

$$\operatorname{dist}(x;\Omega) \le (2C+1)\max\big[\sup_{i\ne j}\big\{\operatorname{dist}(x;\Omega_i)\big\},\ \operatorname{dist}\big(x;\Omega_j\cap(\bar{x}+3r\mathbb{B})\big)\big].$$

Thus the linear regularity of $\{\Omega_i\}_{i\in\mathbb{N}}$ at $\bar{x}$ in the form of

$$\operatorname{dist}(x;\Omega) \le (2C+1)\sup_{i\in\mathbb{N}}\big\{\operatorname{dist}(x;\Omega_i)\big\}$$

would follow now from the relationship

$$\operatorname{dist}\big(x;\Omega_j\cap(\bar{x}+3r\mathbb{B})\big) = \operatorname{dist}(x;\Omega_j) \ \text{ for all } \ x\in\bar{x}+r\mathbb{B}. \tag{3.51}$$

To show (3.51), fix a vector $x\in\bar{x}+r\mathbb{B}$ above and pick any $y\in\Omega_j\setminus(\bar{x}+3r\mathbb{B})$. This readily gives us $\|x-y\|\ge\|y-\bar{x}\|-\|\bar{x}-x\|\ge 3r-r=2r$ and implies that

$$\operatorname{dist}\big(x;\Omega_j\setminus(\bar{x}+3r\mathbb{B})\big)\ge 2r \ \text{ while } \ \operatorname{dist}\big(x;\Omega_j\cap(\bar{x}+3r\mathbb{B})\big)\le\|x-\bar{x}\|\le r.$$

Hence we get the equalities

$$\begin{aligned}\operatorname{dist}(x;\Omega_j) &= \min\big\{\operatorname{dist}\big(x;\Omega_j\setminus(\bar{x}+3r\mathbb{B})\big),\ \operatorname{dist}\big(x;\Omega_j\cap(\bar{x}+3r\mathbb{B})\big)\big\}\\ &= \operatorname{dist}\big(x;\Omega_j\cap(\bar{x}+3r\mathbb{B})\big),\end{aligned}$$

which justify (3.51) and thus complete the proof of the corollary. □

The next proposition, which holds in fact for arbitrary (not only countable) intersections of sets, establishes a new sufficient condition for the CHIP of $\{\Omega_i\}_{i\in\mathbb{N}}$. To formulate it, we introduce a notion of the *tangential rank* of the intersection $\Omega := \bigcap_{i=1}^{\infty} \Omega_i$ at $\bar{x} \in \Omega$ by

$$\rho_\Omega(\bar{x}) := \inf_{i \in \mathbb{N}} \left\{ \limsup_{\substack{x \to \bar{x} \\ x \in \Omega_i \setminus \{\bar{x}\}}} \frac{\text{dist}\,(x; \Omega)}{\|x - \bar{x}\|} \right\}, \tag{3.52}$$

where we put $\rho_\Omega(\bar{x}) := 0$ if $\Omega_i = \{\bar{x}\}$ for at least one $i \in \mathbb{N}$.

Proposition 3.30 (sufficient condition for CHIP via tangential rank of intersection). *Given a countable system of closed sets $\{\Omega_i\}_{i \in \mathbb{N}}$ in $\mathbb{R}^n$, suppose that $\rho_\Omega(\bar{x}) = 0$ for the tangential rank of their intersection $\Omega := \bigcap_{i=1}^{\infty} \Omega_i$ at $\bar{x} \in \Omega$. Then this system exhibits the CHIP at $\bar{x}$.*

Proof. The result holds trivially if $\Omega_i = \{\bar{x}\}$ for some $i \in \mathbb{N}$. Assume that $\Omega_i \setminus \{\bar{x}\} \neq \emptyset$ for all $i \in \mathbb{N}$ and observe that $T(\bar{x}; \Omega) \subset T(\bar{x}; \Omega_i)$ whenever $i \in \mathbb{N}$. Thus we always have

$$T(\bar{x}; \Omega) \subset \bigcap_{i \in \mathbb{N}} T(\bar{x}; \Omega_i).$$

To prove the reverse inclusion, fix an arbitrary vector $0 \neq v \in \bigcap_{i=1}^{\infty} T(\bar{x}; \Omega_i)$. By $\rho_\Omega(\bar{x}) = 0$ and definition (3.52), for any $k \in \mathbb{N}$ we find a set Ω_k from the system under consideration such that

$$\limsup_{\substack{x \to \bar{x} \\ x \in \Omega_k \setminus \{\bar{x}\}}} \frac{\text{dist}\,(x; \Omega)}{\|x - \bar{x}\|} < \frac{1}{k}.$$

Since $v \in T(\bar{x}; \Omega_k)$, there are sequences $\{x_j\}_{j \in \mathbb{N}} \subset \Omega_k$ and $t_j \downarrow 0$ satisfying

$$x_j \to \bar{x} \quad \text{and} \quad \frac{x_j - \bar{x}}{t_j} \to v \quad \text{as } j \to \infty,$$

which in turn implies the limiting estimate

$$\limsup_{j \to \infty} \frac{\text{dist}\,(x_j; \Omega)}{\|x_j - \bar{x}\|} < \frac{1}{k}.$$

The latter allows us to find a vector $x_k \in \{x_j\}_{j\in\mathbb{N}}$ with $\|x_k - \bar{x}\| \le 1/k$ and the corresponding number $t_k \le 1/k$ such that

$$\left\|\frac{x_k - \bar{x}}{t_k} - v\right\| \le \frac{1}{k} \quad \text{and} \quad \frac{\operatorname{dist}(x_k;\Omega)}{\|x_k - \bar{x}\|} < \frac{1}{k}.$$

Then it follows that there exists $z_k \in \Omega$ satisfying the relationships

$$\|z_k - x_k\| < \frac{1}{k}\|x_k - \bar{x}\| \le \frac{1}{k^2}.$$

Combining all the above together gives us the estimates

$$\left\|\frac{z_k - \bar{x}}{t_k} - v\right\| \le \left\|\frac{z_k - x_k}{t_k}\right\| + \left\|\frac{x_k - \bar{x}}{t_k} - v\right\| \le \frac{1}{k}\left\|\frac{x_k - \bar{x}}{t_k}\right\| + \frac{1}{k} \le \frac{1}{k}\Big(\|v\| + \frac{1}{k}\Big) + \frac{1}{k}, \quad k \in \mathbb{N}.$$

Now letting $k \to \infty$, we get $z_k \xrightarrow{\Omega} \bar{x}$, $t_k \downarrow 0$, and $\left\|\frac{z_k - \bar{x}}{t_k} - v\right\| \longrightarrow 0$. The latter verifies that $v \in T(\bar{x};\Omega)$ and thus completes the proof of the proposition. □

To conclude our discussions on the CHIP, we give yet another verifiable condition ensuring the fulfillment of this property for countable intersections of closed sets. We say that a set A is of *invex type* if it can be represented as the complement to a union with respect to $t \in T$ of some open convex sets A_t, i.e.,

$$A = \mathbb{R}^n \setminus \bigcup_{t\in T} A_t, \tag{3.53}$$

The following lemma needed for the next proposition is also used in Section 5.

Lemma 3.31 (sets of invex type). *Let $A \subset \mathbb{R}^n$ be a set of invex type* (3.53)*, and let $\bar{x} \in \bigcap_{t\in T} \operatorname{bd} A_t \cap \operatorname{bd} A$ be taken from the boundary intersections. Then we have the inclusion involving the tangent cone $T(\bar{x};A)$: $\bar{x} + T(\bar{x};A) \subset A$.*

Proof. To justify desired inclusion, suppose on the contrary that there is $v \in T(\bar{x}; A)$ such that $\bar{x} + v \notin A$. For this v we find by definition (2.2) sequences $s_k \downarrow 0$ and $x_k \in A$ such that $\frac{x_k - \bar{x}}{s_k} \to v$ as $k \to \infty$. Since $\bar{x} + v \notin A$, by invexity (3.53) there is an index $t_0 \in T$ for which $\bar{x} + v \in A_{t_0}$. Thus we get the inclusion

$$\bar{x} + \frac{x_k - \bar{x}}{s_k} \in A_{t_0} \ \text{ for all } \ k \in \mathbb{N} \ \text{ sufficiently large.}$$

Then employing the convexity of A_{t_0} gives us that

$$x_k = (1 - s_k)\bar{x} + s_k\Big(\bar{x} + \frac{x_k - \bar{x}}{s_k}\Big) \in A_{t_0}$$

for the fixed index $t_0 \in T$ and all large numbers $k \in \mathbb{N}$. This contradicts the fact that of $x_k \in A$ and thus justifies the claimed inclusion. $\square$.

Now we are ready to derive the aforementioned sufficient condition for the CHIP.

Proposition 3.32 (CHIP for countable intersections of invex-type sets). *Given a countable system $\{\Omega_i\}_{i \in \mathbb{N}}$ in $\mathbb{R}^n$, assume that there is a (possibly infinite) index subset $J \subset \mathbb{N}$ such that each Ω_i for $i \in J$ is the complement to an open and convex set in $\mathbb{R}^n$ and that*

$$\bar{x} \in \Big(\bigcap_{i \in J} \operatorname{bd} \Omega_i\Big) \cap \operatorname{int} \bigcap_{i \notin J} \Omega_i \tag{3.54}$$

for some $\bar{x}$. Then the system $\{\Omega_i\}_{i \in \mathbb{N}}$ enjoys the CHIP at $\bar{x}$.

Proof. Take any Ω_i with $i \in J$ and consider the convex and open set $A \subset \mathbb{R}^n$ such that $\Omega = R^n \setminus A$. Then $\bar{x} \in \operatorname{bd} A \cap \operatorname{bd} \Omega_i$ by (3.54). Then Lemma 3.31 ensures that $\bar{x} + T(\bar{x}; \Omega_i) \subset \Omega_i$

for this index $i \in J$. By the choice of $\bar{x}$ in (3.54) we have furthermore that

$$\bigcap_{i=1}^{\infty} T(\bar{x};\Omega_i) = \bigcap_{i\in J} T(\bar{x};\Omega_i) \subset \bigcap_{i\in J}(\Omega_i - \bar{x}).$$

Since the set on the left-hand side of the latter inclusion is a cone, it follows that

$$\bigcap_{i=1}^{\infty} T(\bar{x};\Omega_i) \subset T\Big(0;\bigcap_{i\in J}(\Omega_i - \bar{x})\Big) = T\Big(\bar{x};\bigcap_{i\in J}\Omega_i\Big) = T\Big(\bar{x};\bigcap_{i=1}^{\infty}\Omega_i\Big). \tag{3.55}$$

As the opposite inclusion in (3.55) is obvious, we conclude that the CHIP is satisfied for the countable set system $\{\Omega_i\}_{i\in\mathbb{N}}$ at $\bar{x}$. □

For countable systems of linear inequalities we have a useful consequence of Proposition 3.32.

Corollary 3.33 (CHIP for countable linear systems). *Consider the set system $\{\Omega_i\}_{i\in\mathbb{N}}$ defined by linear inequalities*

$$\Omega_i := \big\{x \in \mathbb{R}^n \big|\ \langle a_i, x\rangle \le b_i\big\},$$

where $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$ are fixed as $i \in \mathbb{N}$. Given a point $\bar{x} \in \Omega$ and the associated set $J(\bar{x})$ of active indices, suppose that

$$\bar{x} \in \operatorname{int}\big\{x \in \mathbb{R}^n \big|\ \langle a_i, x\rangle \le b_i,\ i \in \mathbb{N}\setminus J(\bar{x})\big\}.$$

Then the countable linear system $\{\Omega_i\}_{i\in\mathbb{N}}$ enjoys the CHIP at $\bar{x}$.

Proof. It obviously follows from Proposition 3.32. Note that one of the referees suggested an alternative proof of this result based on Farkas' lemma with no usage of Lemma 3.31. □

In the last part of this section we show that the CHIP for countable intersections of nonconvex sets, combined with some other classification conditions, allows us to derive principal calculus rule for representing *generalized normals to infinite set intersections.* Thus the verifiable sufficient conditions for the CHIP established above largely contribute to the implementation of these calculus rules. Note that the results obtained in this direction provide new information even for convex set intersections, since in this case they furnish the required implication CHIP $\Longrightarrow$ strong CHIP, which does not hold in general nonconvex settings; see Theorem 3.25 for more discussions.

First we formulate and discuss appropriate qualification conditions for countable systems of sets in terms of the basic normal cone (2.6).

Definition 3.34 (normal closedness and qualification conditions for countable set systems). *Let $\{\Omega_i\}_{i\in\mathbb{N}}$ be a countable system of sets, and let $\bar{x}\in\bigcap_{i=1}^{\infty}\Omega_i$. We say that:*

(a) *The set system $\{\Omega_i\}_{i\in\mathbb{N}}$ satisfies the* NORMAL CLOSEDNESS CONDITION (NCC) *at $\bar{x}$ if the combination of basic normals*

$$\Big\{\sum_{i\in I}x_i^*\Big|\ x_i^*\in N(\bar{x};\Omega_i),\ I\in\mathcal{L}\Big\}\ \text{ is closed in }\ \mathbb{R}^n, \tag{3.56}$$

where $\mathcal{L}$ stands for the collection of all the finite subsets of $\mathbb{N}$.

(b) *The system $\{\Omega_i\}_{i\in\mathbb{N}}$ satisfies the* NORMAL QUALIFICATION CONDITION (NQC) *at $\bar{x}$ if the following implication holds:*

$$\Bigg[\sum_{i=1}^{\infty}x_i^*=0,\ \ x_i^*\in N(\bar{x};\Omega_i)\Bigg]\Longrightarrow\Big[x_i^*=0\ \text{ for all }\ i\in\mathbb{N}\Big]. \tag{3.57}$$

The NCC in Definition 3.34(a) relates to various versions of the so-called *Farkas-Minkowski qualification condition* and its extensions for finite and infinite systems of sets. We refer the reader to, e.g., [17, 18] and the bibliographies therein, as well as to subsequent discussions in

Section 4, for a number of results in this direction concerning convex infinite inequality systems and to [10] for more details on linear inequality systems with arbitrary index sets in general Banach spaces.

The NQC in Definition 3.34(b) is a direct extension of the corresponding condition (3.20)) for system of cones. The counterpart of (3.57) for finite systems of sets is studied and applied in [47, 48] under the same name. The following proposition presents a simple sufficient condition for the validity of the NQC in the case of countable systems of convex sets.

Proposition 3.35 (NQC for countable systems of convex sets). *Let* $\{\Omega_i\}_{i\in\mathbb{N}}$ *be a system of convex sets for which there is an index* $i_0 \in \mathbb{N}$ *such that*

$$\Omega_{i_0} \cap \bigcap_{i\neq i_0} \operatorname{int}\Omega_i \neq \emptyset. \tag{3.58}$$

Then the NQC in (3.57) *is satisfied for the system* $\{\Omega_i\}_{i\in\mathbb{N}}$ *at any* $\bar{x} \in \bigcap_{i=1}^{\infty}\Omega_i$.

Proof. Suppose without loss of generality that $i_0 = 1$ and fix some $w \in \Omega_1 \cap \bigcap_{i=2}^{\infty} \operatorname{int}\Omega_i$. Taking any normals $x_i^* \in N(\bar{x};\Omega_i)$ with $i \in \mathbb{N}$ satisfying

$$\sum_{i=1}^{\infty} x_i^* = 0,$$

we get by the convexity of the sets Ω_i that $\langle x_i^*, w-\bar{x}\rangle \leq 0$ for all $i \in \mathbb{N}$. Then it follows that

$$\langle x_i^*, w-\bar{x}\rangle = -\sum_{j\neq i}\langle x_j^*, w-\bar{x}\rangle \geq 0, \quad i \in \mathbb{N},$$

which yields $\langle x_i^*, w-\bar{x}\rangle = 0$ whenever $i \in \mathbb{N}$. Picking $u \in \mathbb{R}^n$ with $\|u\| = 1$ and taking into account that $w \in \cap_{i=2}^{\infty} \text{int}\, \Omega_i$, we get

$$\lambda \langle x_i^*, u \rangle = \langle x_i^*, w + \lambda u - \bar{x} \rangle \le 0, \quad i = 2, 3, \ldots,$$

whenever $\lambda > 0$ is sufficiently small. Since u is any vector satisfying $\|u\| = 1$, it follows that $x_i^* = 0$ for $i = 2, 3, \ldots$ and therefore $x_i^* = 0$ for all $i \in \mathbb{N}$. $\square$

Finally, we obtain the main result of this section, which expresses Fréchet normal to infinite set intersections via basic normals to the sets involved under the above CHIP and qualification conditions. This major calculus rule for arbitrary closed sets employs the corresponding intersection rule for cones from Theorem 3.23, which is based on the tangential extremal principle.

Theorem 3.36 (generalized normals to countable set intersections). *Let $\{\Omega_i\}_{i \in \mathbb{N}}$ be a countable system of closed sets in $\mathbb{R}^n$, and let $\bar{x} \in \Omega := \bigcap_{i=1}^{\infty} \Omega_i$. Assume that the CHIP in* (3.39) *and NQC in* (3.57) *are satisfied for $\{\Omega_i\}_{i \in \mathbb{N}}$ at $\bar{x}$. Then we have the inclusion*

$$\widehat{N}(\bar{x}; \Omega) \subset \text{cl} \Big\{ \sum_{i \in I} x_i^* \Big|\, x_i^* \in N(\bar{x}; \Omega_i),\ I \in \mathcal{L} \Big\}, \tag{3.59}$$

where $\mathcal{L}$ stands for the collection of all the finite subsets of $\mathbb{N}$. If in addition the NCC in (3.56) *holds for $\{\Omega_i\}_{i \in \mathbb{N}}$ at $\bar{x}$, then the closure operation can be omitted on the right-hand side of* (3.59).

Proof. Using the assumed CHIP for $\{\Omega_i\}_{i \in \mathbb{N}}$ at $\bar{x}$, constructions (2.2) and (2.3), and the duality correspondence (2.4) gives us

$$\widehat{N}(\bar{x}; \Omega) = \widehat{N}\big(0; T(\bar{x}; \Omega)\big) = \widehat{N}\Big(0; \bigcap_{i=1}^{\infty} T(\bar{x}; \Omega_i)\Big). \tag{3.60}$$

It follows from (3.19) that $N\big(0;T(\bar{x};\Omega_i)\big)\subset N(\bar{x};\Omega_i)$ for all $i\in\mathbb{N}$, and thus the assumed NQC in (3.57) implies the conic one in (3.20). Applying Theorem 3.23, we have

$$\widehat{N}\Big(0;\bigcap_{i=1}^{\infty}T(\bar{x};\Omega_i)\Big)\subset \operatorname{cl}\Big\{\sum_{i\in I}x_i^*\Big|\ x_i^*\in N\big(0;T(\bar{x};\Omega_i)\big),\ I\in\mathcal{L}\Big\}.$$

Now the intersection rule (3.59) follows from (3.19) and (3.60). Finally, the closure operation in (3.59) can be obviously dropped if the system $\{\Omega_i\}_{i\in\mathbb{N}}$ satisfies the NCC at $\bar{x}$. □

3.7 Applications to Semi-Infinite Programming

This section is devoted to deriving necessary optimality conditions for various problems of semi-infinite programming (SIP) with *countable* constraints. Problems with countable constraints are among the most difficult in SIP, in comparison with conventional ones involving constraints indexed by compact sets. In fact, SIP problems with countable constraints are not different from seemingly more general problems with arbitrary index sets. Problems of the latter class have drawn particular attention in a number of recent publications, where some special structures of this type (mostly with linear and convex inequality constraints) have been considered; see, e.g., [10, 17, 18] and the references therein. In this section we derive, based on the tangential extremal principle and its calculus consequences, new optimality conditions for SIP with various types of countable constraints and compare them with those known in the literature.

Let us start with SIP involving countable constraints of the *geometric type*:

$$\text{minimize }\ \varphi(x)\ \text{ subject to }\ x\in\Omega_i\ \text{ as }\ i\in\mathbb{N},\tag{3.61}$$

where $\varphi\colon\mathbb{R}^n\to\overline{\mathbb{R}}$ is an extended-real-valued function, and where $\{\Omega_i\}_{i\in\mathbb{N}}\subset\mathbb{R}^n$ is a countable system of constraint sets. Considering in general problems with nonsmooth and nonconvex cost functions and following the classification of [48, Chapter 5], we derive necessary optimality conditions of two kinds for (3.61) and other SIP *minimization* problems: *lower subdifferential* and *upper subdifferential* ones. Conditions of the "lower" kind are more conventional for minimization dealing with usual (lower) subdifferential constructions. On the other hand, conditions of the "upper" kind employ upper subdifferential (or superdifferential) constructions, which seem to be more appropriate for maximization problems while bringing significantly stronger information for special classes of minimizing cost functions in comparison with lower subdifferential ones; see [48] for more discussions, examples, and references.

We begin with upper subdifferential optimality conditions for (3.61). Given $\varphi\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ finite at $\bar{x}$, the *upper subdifferential* of φ at $\bar{x}$ used in this paper is of the Fréchet type defined by

$$\widehat{\partial}^+\varphi(\bar{x}) := -\widehat{\partial}(-\varphi)(\bar{x}) = \Big\{x^* \in \mathbb{R}^n \Big| \limsup_{x\to\bar{x}} \frac{\varphi(x) - \varphi(\bar{x}) - \langle x^*, x - \bar{x}\rangle}{\|x - \bar{x}\|} \le 0\Big\} \tag{3.62}$$

via (2.9). Note that $\widehat{\partial}^+\varphi(\bar{x})$ reduces to the upper subdifferential (or superdifferential) of convex analysis if φ is concave. Furthermore, the subdifferential sets $\widehat{\partial}\varphi(\bar{x})$ and $\widehat{\partial}^+\varphi(\bar{x})$ are nonempty simultaneously if and only if φ is Fréchet differentiable at $\bar{x}$.

As before, in the next theorem and in what follows the symbol $\mathcal{L}$ stands for the collection of all the finite subsets of the natural series $\mathbb{N}$.

Theorem 3.37 (upper subdifferential conditions for SIP with countable geometric constraints). *Let $\bar{x}$ be a local optimal solution to problem* (3.61), *where $\varphi\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is an arbitrary extended-real-valued function finite at $\bar{x}$, and where the sets $\Omega_i \subset \mathbb{R}^n$ for $i \in \mathbb{N}$ are locally closed around $\bar{x}$. Assume that the system $\{\Omega_i\}_{i\in\mathbb{N}}$ has the CHIP at $\bar{x}$ and satisfies the*

NQC of Definition 3.34(b) *at this point. Then we have the set inclusion*

$$-\widehat{\partial}^+\varphi(\bar{x}) \subset \mathrm{cl}\Big\{\sum_{i\in I} x_i^*\Big|\; x_i^* \in N(\bar{x};\Omega_i),\; I \in \mathcal{L}\Big\}, \tag{3.63}$$

which reduces to that of

$$0 \in \nabla\varphi(\bar{x}) + \mathrm{cl}\Big\{\sum_{i\in I} x_i^*\Big|\; x_i^* \in N(\bar{x};\Omega_i),\; I \in \mathcal{L}\Big\}. \tag{3.64}$$

if φ is Fréchet differentiable at $\bar{x}$. If in addition the NCC of Definition 3.34(a) *holds for $\{\Omega_i\}_{i\in\mathbb{N}}$ at $\bar{x}$, then the closure operations can be omitted in* (3.63) *and* (3.64).

Proof. It follows from [48, Proposition 5.2] that

$$-\widehat{\partial}^+\varphi(\bar{x}) \subset \widehat{N}\Big(\bar{x};\bigcap_{i=1}^{\infty}\Omega_i\Big). \tag{3.65}$$

Applying now to (3.65) the representation of Fréchet normals to countable set intersections from Theorem 3.36 under the assumed CHIP and NQC, we arrive at (3.63), where the closure operation can be omitted when the NCC holds at $\bar{x}$. If φ is Fréchet differentiable at $\bar{x}$, it follows that $\widehat{\partial}^+\varphi(\bar{x}) = \{\nabla\varphi(\bar{x})\}$, and thus (3.63) reduces to (3.64). □

Note that the set inclusion (3.63) is trivial if $\widehat{\partial}^+\varphi(\bar{x}) = \emptyset$, which is the case of, e.g., nonsmooth convex functions. On the other hand, the upper subdifferential necessary optimality condition (3.63) may be much more selective than its lower subdifferential counterparts when $\widehat{\partial}^+\varphi(\bar{x}) \neq \emptyset$, which happens, in particular, for some remarkable classes of functions including concave, upper regular, semiconcave, upper-C^1, and other ones important in various applications. The reader can find more information and comparison in [48, Subsection 5.1.1] and the commentaries therein concerning problems with finitely many geometric constraints.

Next let us present a lower subdifferential condition for the SIP problem (3.61) involving the basic subdifferential (2.10), which is nonempty for majority of nonsmooth functions; in particular, for any local Lipschitzian one. To formulate this condition, recall the notion of the *singular subdifferential* of φ at $\bar{x}$ defined by

$$\partial^{\infty}\varphi(\bar{x}) := \big\{x^* \in \mathbb{R}^n \big|\ (x^*, 0) \in N\big((\bar{x}; \varphi(\bar{x})); \operatorname{epi}\varphi\big)\big\}. \tag{3.66}$$

Note that $\partial^{\infty}\varphi(\bar{x}) = \{0\}$ if φ is locally Lipschitzian around $\bar{x}$. Recall also that a set Ω is *normally regular* at $\bar{x}$ if $N(\bar{x}; \Omega) = \widehat{N}(\bar{x}; \Omega)$. This is the case, in particular, of locally convex and other "nice" sets; see, e.g., [47, 61] and the references therein.

Theorem 3.38 (lower subdifferential conditions for SIP with countable geometric constraints.) *Let $\bar{x}$ be a local optimal solution to problem* (3.61) *with a lower semicontinuous cost function $\varphi\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ finite at $\bar{x}$ and a countable system $\{\Omega_i\}_{i\in\mathbb{N}}$ of sets locally closed around $\bar{x}$. Assume that the feasible solution set $\Omega := \bigcap_{i=1}^{\infty}\Omega_i$ is normally regular at $\bar{x}$, that the system $\{\Omega_i\}_{i\in\mathbb{N}}$ satisfies the CHIP* (3.39) *and the NQC* (3.57) *at $\bar{x}$, and that*

$$\operatorname{cl}\Big\{\sum_{i\in I} x_i^* \Big|\ x_i^* \in N(\bar{x}; \Omega_i),\ I \in \mathcal{L}\Big\} \bigcap \big(-\partial^{\infty}\varphi(\bar{x})\big) = \{0\}, \tag{3.67}$$

which holds, in particular, when φ is locally Lipschitzian around $\bar{x}$. Then we have

$$0 \in \partial\varphi(\bar{x}) + \operatorname{cl}\Big\{\sum_{i\in I} x_i^* \Big|\ x_i^* \in N(\bar{x}; \Omega_i),\ I \in \mathcal{L}\Big\}. \tag{3.68}$$

The closure operations can be omitted in (3.67) *and* (3.68) *if the NCC* (3.56) *is satisfied at $\bar{x}$.*

Proof. It follows from [48, Proposition 5.3] that

$$0 \in \partial\varphi(\bar{x}) + N(\bar{x};\Omega) \ \text{ provided that } \ \partial^\infty\varphi(\bar{x}) \cap \big(- N(\bar{x};\Omega)\big) = \{0\} \tag{3.69}$$

for the optimal solution $\bar{x}$ to the problem under consideration with the feasible solution set $\Omega := \bigcap_{i=1}^\infty \Omega_i$. Since the set Ω is normally regular at $\bar{x}$, we can replace $N(\bar{x};\Omega)$ by $\widehat{N}(\bar{x};\Omega)$ in (3.69). Applying now Theorem 3.36 to the countable set intersection Ω in (3.69) under the assumptions made, we arrive at all the conclusions of this theorem. □

Next we consider a SIP problem with *countable operator constraints* defined by:

$$\text{minimize } \ \varphi(x) \ \text{ subject to } \ f(x) \in \Theta_i \ \text{ as } \ i \in \mathbb{N}, \tag{3.70}$$

where $\varphi\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, $\Theta_i \subset \mathbb{R}^m$ for $i \in \mathbb{N}$, and $f\colon \mathbb{R}^n \to \mathbb{R}^m$. The following statements are consequences of Theorems 3.37 and 3.38, respectively.

Corollary 3.39 (upper and lower subdifferential conditions for SIP with operator constraints). *Let $\bar{x}$ be a local optimal solution to* (3.70), *where the cost function φ is finite at $\bar{x}$, where the mapping $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is strictly differentiable at $\bar{x}$ with the surjective* (*full rank*) *derivative, and where the sets $\Theta_i \subset \mathbb{R}^m$ as $i \in \mathbb{N}$ are locally closed around $f(\bar{x})$ while satisfying the CHIP* (3.39) *and NQC* (3.57) *conditions at this point. The following assertions holds:*

(i) *We have the upper subdifferential optimality condition:*

$$-\,\widehat{\partial}^+\varphi(\bar{x}) \subset \text{cl}\Big\{ \sum_{i\in I} \nabla f(\bar{x})^* y_i^* \Big|\ y_i^* \in N\big(f(\bar{x});\Theta_i\big),\ I \in \mathcal{L}\Big\}, \tag{3.71}$$

(ii) *If φ is lower semicontinuous around $\bar{x}$ and*

$$\mathrm{cl}\Big\{\sum_{i\in I}\nabla f(\bar{x})^*y_i^*\Big|\ y_i^*\in N\big(f(\bar{x});\Theta_i\big),\ I\in\mathcal{L}\Big\}\bigcap\big(-\partial^\infty\varphi(\bar{x})\big)=\{0\}, \tag{3.72}$$

then we have the inclusion

$$0\in\partial\varphi(\bar{x})+\mathrm{cl}\Big\{\sum_{i\in I}\nabla f(\bar{x})^*y_i^*\Big|\ y_i^*\in N\big(f(\bar{x});\Theta_i\big),\ I\in\mathcal{L}\Big\}. \tag{3.73}$$

Furthermore, the closure operations can be omitted in (3.71)–(3.73) *if the set system* $\{\Theta_i\}_{i\in\mathbb{N}}$ *satisfies the NCC* (3.56) *at* $f(\bar{x})$.

Proof. Observe that problem (3.70) can be equivalently rewritten in the geometric form (3.61) with $\Omega_i:=f^{-1}(\Theta_i)$, $i\in\mathbb{N}$. Then employing the well-known results on representing the tangent and normal cones in (2.2) and (2.6) to inverse images of sets under strict differentiable mappings with surjective derivatives (see, e.g., [47, Theorem 1.17] and [61, Exercise 6.7]), we have

$$T\big(\bar{x};f^{-1}(\Theta)\big)=\nabla f(\bar{x})^{-1}T\big(f(\bar{x});\Theta\big)\ \text{ and }\ N\big(\bar{x};f^{-1}(\Theta)\big)=\nabla f(\bar{x})^*N\big(f(\bar{x});\Theta\big). \tag{3.74}$$

It follows from the surjectivity of $\nabla f(\bar{x})$ that the CHIP and NQC for $\{\Theta_i\}_{i\in\mathbb{N}}$ at $f(\bar{x})$ are equivalent, respectively, to the CHIP and NQC of $\{\Omega_i\}_{i\in\mathbb{N}}$ at $\bar{x}$; see [47, Lemma 1.18]. This implies the equivalence between the qualification and optimality conditions (3.71)–(3.73) for problem (3.70) under the assumptions made and the corresponding conditions (3.63), (3.67), and (3.68) for problem (3.61) established in Theorems 3.37 and 3.38. To complete the proof of the corollary, it suffices to observe similarly to (3.74) that the assumed NCC for $\{\Theta_i\}_{i\in\mathbb{N}}$ at $f(\bar{x})$

is equivalent under the surjectivity of $\nabla f(\bar{x})$ to the NCC (3.56) for the inverse images $\{\Omega_i\}_{i\in\mathbb{N}}$ at $\bar{x}$. Thus the possibility to omit the closure operations in the framework of the corollary follows directly from the corresponding statements of Theorems 3.37 and 3.38. □

The rest of this section concerns SIP problems with *countable inequality constraints*:

$$\text{minimize } \varphi(x) \text{ subject to } \varphi_i(x) \le 0 \text{ as } i \in \mathbb{N}, \tag{3.75}$$

where the cost function φ is as in problems (3.61) and (3.70) while the constraint functions $\varphi_i\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, $i \in \mathbb{N}$, are lower semicontinuous around the reference optimal solution. Note that problems with infinite inequality constraints are considered in the vast majority of publications on semi-infinite programming, where the main attention is paid to the case of convex or linear infinite inequalities; see below some comparison with known results for SIP of the latter types.

Although our methods are applied to problems (3.75) of the general inequality type, for simplicity and brevity we focus here on the case when the constraint functions φ_i, $i \in \mathbb{N}$, are locally Lipschitzian around the optimal solution. In the general case we need to involve the singular subdifferential (3.66) of these functions; see the proofs below. Let us first introduce subdifferential counterparts of the normal qualification and closedness conditions from Definition 3.34.

Definition 3.40 (subdifferential closedness and qualification conditions for countable inequality constraints). *Consider a countable constraint system $\{\Omega_i\}_{i\in\mathbb{N}} \subset \mathbb{R}^n$ with*

$$\Omega_i := \big\{x \in \mathbb{R}^n \big|\ \varphi_i(x) \le 0\big\}, \quad i \in \mathbb{N}, \tag{3.76}$$

where the functions φ_i are locally Lipschitzian around $\bar{x} \in \bigcap_{i=1}^{\infty} \Omega_i$. We say that:

(a) *The system* $\{\Omega_i\}_{i\in\mathbb{N}}$ *in* (3.76) *satisfies the* SUBDIFFERENTIAL CLOSEDNESS CONDITION (*SCC*) *at* $\bar{x}$ *if the set*

$$\Big\{\sum_{i\in I}\lambda_i\partial\varphi_i(\bar{x})\Big|\ \lambda_i\ge 0,\ \lambda_i\varphi_i(\bar{x})=0,\ I\in\mathcal{L}\Big\}\ \textit{ is closed in }\ \mathbb{R}^n. \tag{3.77}$$

(b) *The system* $\{\Omega_i\}_{i\in\mathbb{N}}$ *in* (3.76) *satisfies the* SUBDIFFERENTIAL QUALIFICATION CONDITION (*SQC*) *at* $\bar{x}$ *if the following implication holds:*

$$\Big[\sum_{i=1}^{\infty}\lambda_i x_i^*=0,\ x_i^*\in\partial\varphi_i(\bar{x}),\ \lambda_i\ge 0,\ \lambda_i\varphi_i(\bar{x})=0\Big]\Longrightarrow\big[\lambda_i=0\ \textit{ for all }\ i\in\mathbb{N}\big]. \tag{3.78}$$

The next theorem provides necessary optimality conditions of both upper and lower subdifferential types for SIP problems (3.75) without any smoothness and/or convexity assumptions.

Theorem 3.41 (upper and lower subdifferential conditions for general SIP with inequality constraints). *Let* $\bar{x}$ *be a local optimal solution to problem* (3.75)*, where the constraint functions* $\varphi_i\colon\mathbb{R}^n\to\overline{\mathbb{R}}$ *are locally Lipschitzian around* $\bar{x}$ *for all* $i\in\mathbb{N}$*. Assume that the level set system* $\{\Omega_i\}_{i\in\mathbb{N}}$ *in* (3.76) *has the CHIP at* $\bar{x}$ *and that the SQC* (3.78) *is satisfied at this point. Then the following assertions hold:*

(i) *We have the upper subdifferential optimality condition:*

$$-\widehat{\partial}^+\varphi(\bar{x})\subset\operatorname{cl}\Big\{\sum_{i\in I}\lambda_i\partial\varphi_i(\bar{x})\Big|\ \lambda_i\ge 0,\ \lambda_i\varphi_i(\bar{x})=0,\ I\in\mathcal{L}\Big\}, \tag{3.79}$$

where the closure operation can be omitted if the SCC (3.77) *is satisfied at* $\bar{x}$*.*

(ii) *Assume in addition that φ is lower semicontinuous around $\bar{x}$ and that*

$$\mathrm{cl}\Big\{\sum_{i\in I}\lambda_i\partial\varphi_i(\bar{x})\Big|\ \lambda_i\ge 0,\ \lambda_i\varphi_i(\bar{x})=0,\ I\in\mathcal{L}\Big\}\bigcap\big(-\partial^\infty\varphi(\bar{x})\big)=\{0\}, \tag{3.80}$$

which is automatic if φ is locally Lipschitzian around $\bar{x}$. Then

$$0\in\partial\varphi(\bar{x})+\mathrm{cl}\Big\{\sum_{i\in I}\lambda_i\partial\varphi_i(\bar{x})\Big|\ \lambda_i\ge 0,\ \lambda_i\varphi_i(\bar{x})=0,\ I\in\mathcal{L}\Big\} \tag{3.81}$$

with removing the closure operation in (3.80) *and* (3.81) *when the SCC* (3.77) *holds at $\bar{x}$.*

Proof. It is well known from the basic subdifferential calculus (see, e.g., [47, Theorem 3.86]) that

$$N(\bar{x};\Omega)\subset\mathbb{R}_+\partial\vartheta(\bar{x}):=\big\{\lambda x^*\in\mathbb{R}^n\big|\ x^*\in\partial\vartheta(\bar{x}),\ \lambda\ge 0\big\}\ \text{ for }\ \Omega:=\big\{x\in\mathbb{R}^n\big|\ \vartheta(x)\le 0\big\} \tag{3.82}$$

provided that $\vartheta\colon\mathbb{R}^n\to\overline{\mathbb{R}}$ is locally Lipschitzian around $\bar{x}$ and that $0\notin\partial\vartheta(\bar{x})$, which is ensured by the assumed SQC. Now we apply inclusion (3.82) to each set Ω_i in (3.76) and substitute this into the NQC (3.57) as well as into the qualification condition (3.67) and the optimality conditions (3.63) and (3.68) for problem (3.61) with the constraint sets (3.76). It follows in this way that the SQC (3.78) and all the relationships (3.79)–(3.81) imply the aforementioned conditions of Theorems 3.37 and (3.38) in the setting (3.75) under consideration. It shows furthermore that the SCC (3.77) yields the NCC (3.56) for the sets Ω_i in (3.76), which thus completes the proof of the theorem. $\square$

Now we consider in more detail the case of *convex* constraint functions φ_i in (3.75). Note that the validity of the SQC (3.78) is ensured in the case by the interior-type condition (3.58) of

Proposition 3.35. The next theorem justifies necessary optimality conditions for problems with countable convex inequalities, which does not require either interiority-type or SQC constraint qualifications while containing a qualification condition that implies both the CHIP and SCC in (3.77). Let us first recall (see [17, 18] and the references therein) that the SIP problem (3.75) with the constraints given by convex functions φ_i, $i \in \mathbb{N}$, satisfies the *Farkas-Minkowski constraint qualification* (FMCQ) if the set

$$\mathrm{co}\left[\mathrm{cone}\bigcup_{i=1}^{\infty}\mathrm{epi}\,\varphi_i^*\right] \text{ is closed in } \mathbb{R}^n \times \mathbb{R}, \tag{3.83}$$

where $\vartheta^*(x^*) := \sup\{\langle x^*, x\rangle - \vartheta(x)|\ x \in \mathbb{R}^n\}$ stands for the Fenchel conjugate function to $\vartheta\colon \mathbb{R}^n \to \overline{\mathbb{R}}$. In fact, this condition can be considered as a consequence of the Farkas-Minkowski property for linear inequality systems [24] via a linearization of convex inequalities by using the Fenchel conjugates. Following [24, Section 7.5] and [23, Definition 5.12], we say that system (3.76) defined by convex inequalities satisfies the *local Farkas-Minkowski* (LFM) property at $\bar{x} \in \Omega := \cap_{i=1}^{\infty}\Omega_i$ if

$$N(\bar{x};\Omega) = \mathrm{co}\left[\mathrm{cone}\bigcup_{i\in J(\bar{x})}\partial\varphi_i(\bar{x})\right] =: B(\bar{x}), \tag{3.84}$$

where $J(\bar{x}) := \{i \in \mathbb{N}|\ \varphi_i(\bar{x}) = 0\}$ is the the collection of active indices at $\bar{x}$. Note that the LFM property (3.84) is called the "basic constraint qualification" in [39, 42].

It has been observed for the convex systems under consideration that FMCQ$\Longrightarrow$LFM. We refer the reader to [22] for a comprehensive study of relationships between various qualification conditions for systems of convex inequalities.

Having this in hand, we get the following results for infinite convex inequality systems, where we assumed for simplicity that the cost function in (3.75) is locally Lipschitzian. The final formulation and the proof of the theorem below is suggested to us by Marco López.

Theorem 3.42 (upper and lower subdifferential conditions for SIP with convex inequality constraints). *Let all the general assumptions but SQC* (3.78) *of Theorem* 3.41 *be fulfilled at the local optimal solution* $\bar{x}$ *to* (3.75)*. Suppose in addition that the cost function* φ *is locally Lipschitzian around* $\bar{x}$*, that the constraint functions* φ_i*,* $i \in \mathbb{N}$*, are convex, and that the LFM property* (3.84) *holds at* $\bar{x}$*. Then the SCC* (3.77) *and CHIP* (3.39) *also hold, and both necessary optimality conditions* (3.79) *and* (3.81) *are satisfied with the closure operations omitted therein.*

Proof. Observe that the SCC in (3.77) is nothing else but the closedness of the set $B(\bar{x})$, and hence we have the implication LFM$\Longrightarrow$SCC by the closedness of the normal cone $N(\bar{x};\Omega)$. Furthermore, we always have the inclusions

$$B(\bar{x}) \subset \text{co} \bigcup_{i \in J(\bar{x})} N(\bar{x};\Omega_i) \subset N(\bar{x};\Omega). \tag{3.85}$$

Hence the LFM property combined with (3.85) implies the strong CHIP. By Theorem 3.25 we have the CHIP as well since $N(\bar{x};\Omega_i) = \{0\}$ whenever $i \notin J(\bar{x})$. Taking all this into account, we get under the assumptions made the inclusions

$$-\widehat{\partial}^+\varphi(\bar{x}) \subset N(\bar{x};\Omega) \;\text{ and }\; 0 \in \partial\varphi(\bar{x}) + N(\bar{x};\Omega),$$

which imply in turn the validity of

$$-\widehat{\partial}^+\varphi(\bar{x}) \subset B(\bar{x}) \;\text{ and }\; 0 \in \partial\varphi(\bar{x}) + B(\bar{x}),$$

and thus complete the proof of the theorem. □

Next we present specifications of both upper and lower subdifferential optimality conditions derived above for SIP (3.75) with linear inequality constraints. In the finite-dimensional countable case under consideration the results obtained in this way reduce to those from [10, Theorems 3.1 and 4.1] while it is *not* assumed here the strong Slater condition and the coefficient boundedness imposed in [10]. For simplicity we consider the case of homogeneous constraints and suppose that $\bar{x} = 0$ is a local optimal solution.

Proposition 3.43 (upper and lower subdifferential conditions for SIP with linear inequality constraints). *Let $\bar{x} = 0$ be a local optimal optimal solution to the SIP problem*

$$\text{minimize } \varphi(x) \text{ subject to } \langle a_i, x\rangle \le 0 \text{ for all } i \in \mathbb{N}, \tag{3.86}$$

where $\varphi\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is finite at the origin. Then we have the inclusions

$$-\widehat{\partial}^+\varphi(0) \subset \operatorname{cl}\operatorname{co}\Big[\bigcup_{i=1}^{\infty}\big\{\lambda a_i \big|\ \lambda \ge 0\big\}\Big]. \tag{3.87}$$

$$0 \in \partial\varphi(0) + \operatorname{cl}\operatorname{co}\Big[\bigcup_{i=1}^{\infty}\big\{\lambda a_i \big|\ \lambda \ge 0\big\}\Big], \tag{3.88}$$

where (3.88) *holds provided that φ is lower semicontinuous around the origin and*

$$\Big(\operatorname{cl}\operatorname{co}\Big[\bigcup_{i=1}^{\infty}\big\{\lambda a_i \big|\ \lambda \ge 0\big\}\Big]\Big) \cap \big(-\partial^{\infty}\varphi(0)\big) = \{0\}. \tag{3.89}$$

Furthermore, the LFM property implies that the closure operations can be omitted in (3.87)*–*(3.89)*.*

Proof. It follows the lines in the proof of Theorem 3.42 with the usage of the normal cone representation (3.48) from Corollary 3.26. □

Finally in this section, we present several examples illustrating the qualification conditions imposed in Theorem 3.42 and their comparison with known results in the literature.

Example 3.44 (comparison of qualification conditions). All the examples below concern lower subdifferential conditions for SIP problems (3.75) with convex cost and constraint functions.

(i) The CHIP (3.39) **and the SCC** (3.77) **are independent**. Consider a linear constraint system in (3.43) at $\bar{x} = (0,0) \in \mathbb{R}^2$ for $\varphi_i(x) = \langle a_i, x\rangle$ with $a_i = (1,i)$ as $i \in \mathbb{N}$, which has the CHIP. At the same time the set

$$\mathrm{co} \bigcup_{i=0}^{\infty} \mathbb{R}_+ \partial\varphi_i(\bar{x}) = \mathrm{co}\left\{\lambda(1,i) \in \mathbb{R}^2 \big|\ \lambda \geq 0,\ i \in \mathbb{N}\right\} = \mathbb{R}^2_+ \setminus \left\{(0,\lambda) \big|\ \lambda > 0\right\}$$

is not closed, and hence the SCC (3.77) does not hold. On the other hand, for the quadratic functions $\varphi_i(x) = ix_1^2 - x_2$ as $i \in \mathbb{N}$ as $x = (x_1,x_2) \in \mathbb{R}^2$, we get $\partial\varphi_i(0) = \nabla\varphi_i(0) = (0,-1)$, and hence the SCC (3.77) holds at the origin while the CHIP is violated at this point similarly to Example 3.27(ii).

(ii) (CHIP and SCC versus FMCQ and CQC). Besides the FMCQ (3.83), another qualification condition is employed in [17, 18] to obtain necessary optimality conditions of *Karush-Kuhn-Tucker* (KKT) type (no closure operation in (3.81)) for *fully convex* SIP problems (3.75) involving all the convex functions φ and φ_i. This condition, named the *closedness qualification condition* (CQC) is formulated as follows via the convex conjugate functions: the set

$$\mathrm{epi}\,\varphi^* + \mathrm{co}\left[\mathrm{cone} \bigcup_{i=1}^{\infty} \mathrm{epi}\,\varphi_i^*\right] \text{ is closed in } \mathbb{R}^n \times \mathbb{R}. \tag{3.90}$$

The next example presents a fully convex SIP problem satisfying both CHIP and SCC but neither the CQC nor the FMCQ. This shows that Theorem 3.42 holds in this case to produce the KKT optimality condition while the corresponding result of [17] is not applicable.

Consider the SIP (3.42) with $x = (x_1, x_2) \in \mathbb{R}^2$, $\bar{x} = (0,0)$, $\varphi(x) = -x_2$, and

$$\varphi_i(x_1, x_2) = \begin{cases} ix_1^2 - x_2 & \text{if } \ x_1 < 0, \\ -x_2 & \text{if } \ x_1 \geq 0, \end{cases} \qquad i \in \mathbb{N}.$$

We have $\partial\varphi_i(\bar{x}) = \nabla\varphi_i(\bar{x}) = (0, -1)$ for all $i \in \mathbb{N}$, and hence the SCC (3.77) holds. It is easy to check that the CHIP holds at $\bar{x}$, since

$$T\Big(\bar{x}; \bigcap_{i=1}^{\infty} \Omega_i\Big) = T(\bar{x}; \Omega_i) = \mathbb{R} \times \mathbb{R}_+ \ \text{ for } \ \Omega_i := \big\{x \in \mathbb{R}^2 \big|\ \varphi_i(x) \leq 0\big\}, \quad i \in \mathbb{N}.$$

On the other hand, for $x^* = (\lambda_1, \lambda_2) \in \mathbb{R}^n$ we compute the conjugate functions by

$$\varphi^*(x^*) = \begin{cases} 0 & \text{if } \ (\lambda_1, \lambda_2) = (0, -1), \\ \infty & \text{otherwise} \end{cases} \qquad \text{and } \ \varphi_i^*(x^*) = \begin{cases} \dfrac{\lambda_1^2}{4i} & \text{if } \ \lambda_1 \leq 0, \ \lambda_2 = -1, \\ \infty & \text{otherwise.} \end{cases}$$

This shows that the convex sets

$$\text{co}\Big[\text{cone} \bigcup_{i=0}^{\infty} \text{epi}\, \varphi_i^*\Big] \quad \text{and} \quad \text{epi}\, \varphi^* + \text{co}\Big[\text{cone} \bigcup_{i=0}^{\infty} \text{epi}\, \varphi_i^*\Big]$$

are not closed in $\mathbb{R}^2 \times \mathbb{R}$, and hence the FMCQ (3.83) and the CQC (3.90) are not satisfied.

(iii) (SQC does not imply CHIP for countable systems). As noted by one of the referees, the SQC (3.78) implies the CHIP for finitely many sets described by smooth inequalities. However, it is not the case for countably many inequalities described by smooth convex functions. Indeed, consider the functions $\varphi_i\colon \mathbb{R}^2 \to \mathbb{R}$ of this type given by

$$\varphi_i(x_1, x_2) := ix_1^2 - x_2, \quad i \in \mathbb{N},$$

for which the SQC holds at $(0,0)$. On the other hand, the sets $\Omega_i := \{(x_1, x_2) \in \mathbb{R}^2 |\ \varphi_i(x_1, x_2) \leq 0\}$ reduce to those in Example 3.27(ii) for which the CHIP is violated at the origin.

3.8 Applications to Multiobjective Optimization

The last section of this chapter concerns problems of multiobjective optimization with set-valued objectives and countable constraints. Although optimization problems with single-valued/vector and (to a lesser extent) set-valued objectives have been widely considered in optimization and equilibrium theories as well as in their numerous applications (see, e.g., the books [25, 28, 48] and the references therein), we are not familiar with the study of such problems involving countable constraints. Our interest is devoted to deriving necessary optimality conditions for problems of this type based on the dual-space approach to the general multiobjective optimization theory developed in [4, 5, 48] and the new tangential extremal principle established in Section 3.1.

The main problem of our consideration is as follows:

$$\text{minimize } \ F(x) \ \text{ subject to } \ x \in \Omega := \bigcap_{i=1}^{\infty} \Omega_i \subset \mathbb{R}^n, \tag{3.91}$$

where Ω_i, $i \in \mathbb{N}$, are closed subsets of $\mathbb{R}^n$, where $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a set-valued mapping of closed graph, and where "minimization" is understood with respect to some partial ordering "$\leq$" on $\mathbb{R}^m$. We pay the main attention to the multiobjective problems with the *Pareto-type* ordering:

$$y_1 \leq y_2 \ \text{ if and only if } \ y_2 - y_1 \in \Theta,$$

where $\emptyset \neq \Theta \subset \mathbb{R}^m$ is a closed, convex, and pointed ordering cone. In the aforementioned references the reader can find more discussions on this and other ordering relations.

Recall that a point $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ with $\bar{x} \in \Omega$ is a *local minimizer* of problem (3.91) if there exists a neighborhood U of $\bar{x}$ such that there is no $y \in F(\Omega \cap U)$ preferred to $\bar{y}$, i.e.,

$$F(\Omega \cap U) \cap (\bar{y} - \Theta) = \{\bar{y}\}. \tag{3.92}$$

Note that notion (3.92) does not take into account the image localization of minimizers around $\bar{y} \in F(\bar{x})$, which is essential for certain applications of set-valued minimization, e.g., to economic modeling; see [5]. A more appropriate notion for such problems is defined in [5] under the name of *fully localized minimizers* as follows: there are neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that

$$F(\Omega \cap U) \cap (\bar{y} - \Theta) \cap V = \{\bar{y}\}. \tag{3.93}$$

The next result establishes necessary optimality conditions of the coderivative type for fully localized minimizers of problem (3.91) with countable constraints based on the approach of [48] to problems of multiobjective optimizations, whose implementations in [4, 5] focus specifically on problems with set-valued criteria, and the tangential extremal principle for countable sets in Section 3.1. We address here fully localized minimizers for multiobjective problems (3.91) with normally regular feasible sets, i.e., when $N(\bar{x}; \Omega) = \widehat{N}(\bar{x}; \Omega)$, which particularly includes the case of convex set Ω_i, $i \in \mathbb{N}$.

Theorem 3.45 (optimality conditions for fully localized minimizers of multiobjective problems with countable constraints and normally regular feasible sets). *Let the pair* $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ *be a fully localized minimizer for* (3.91) *with the CHIP system of countable* *constraints* $\{\Omega_i\}_{i\in\mathbb{N}}$. *Assume that the feasible set* $\Omega=\bigcap_{i=1}^\infty\Omega_i$ *is normally regular at* $\bar{x}\in\Omega$ *and that the NQC* (3.57) *and the coderivative qualification condition*

$$D^*F(\bar{x},\bar{y})(0)\bigcap\Big[-\operatorname{cl}\Big\{\sum_{i\in I}x_i^*\Big|\ x_i^*\in N(\bar{x};\Omega_i),\ I\in\mathcal{L}\Big\}\Big]=\{0\} \tag{3.94}$$

are satisfied. Then there is $0\neq y^*\in-N(0;\Theta)$ *such that*

$$0\in D^*F(\bar{x},\bar{z})(y^*)+\operatorname{cl}\Big\{\sum_{i\in I}x_i^*\Big|\ x_i^*\in N(\bar{x};\Omega_i),\ I\in\mathcal{L}\Big\}. \tag{3.95}$$

Proof. Applying [5, Theorem 3.4] for fully localized minimizers of set-valued optimization problems with abstract geometric constraints $x\in\Omega$ (cf. also [4, Theorem 5.3] for the case of local minimizers (3.92) and [48, Theorem 5.59] for vector single-objective counterparts), we find

$$0\neq -y^*\in N(0;\Theta)\ \text{ and }\ x^*\in D^*F(\bar{x},\bar{y})(y^*)\cap\big(-N(\bar{x};\Omega)\big) \tag{3.96}$$

provided the fulfillment of the qualification condition

$$D^*F(\bar{x},\bar{y})(0)\cap\big(-N(\bar{x};\Omega)\big)=\{0\}. \tag{3.97}$$

To complete the proof of the theorem, it suffices to employ in (3.96) and (3.97) the sum rule for countable set intersections from Theorem 3.36 by taking into account the assumed normal regularity of the intersection set Ω at $\bar{x}$. □

Note that the qualification condition (3.94) holds automatically if the objective mapping F is *Lipschitz-like* (or has the Aubin property) around $(\bar{x},\bar{y})\in\operatorname{gph}F$, i.e., there are neighborhoods

U of $\bar{x}$ and V of $\bar{y}$ such that

$$F(x) \cap V \subset F(u) + \ell \|x - u\| I\!B \quad \text{for all} \quad x, u \in U$$

with some number $\ell \geq 0$. Indeed, it follows from the Mordukhovich criterion in [61, Theorem 9.40] (see also [47, Theorem 4.10] and the references therein) that $D^*F(\bar{x}, \bar{y})(0) = \{0\}$ in this case.

Next we introduce two kinds of "graphical" minimizers for multiobjective problems for which, in particular, we can avoid the normal regularity assumption in optimality conditions of type (3.95) in Theorem 3.45. The definition below concerns multiobjective optimization problems with general geometric constraints that may not be represented as countable set intersections.

Definition 3.46 (graphical and tangential graphical minimizers). *Let* $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ *with* $\bar{x} \in \Omega$. *We say that:*

(i) $(\bar{x}, \bar{y})$ *is a* LOCAL GRAPHICAL MINIMIZER *to problem* (3.91) *if there are neighborhoods* U *of* $\bar{x}$ *and* V *of* $\bar{y}$ *such that*

$$\operatorname{gph} F \cap \Big[\Omega \times (\bar{y} - \Theta)\Big] \cap \big(U \times V\big) = \big\{(\bar{x}, \bar{y})\big\}. \tag{3.98}$$

(ii) $(\bar{x}, \bar{y})$ *is a* LOCAL TANGENTIAL GRAPHICAL MINIMIZER *to problem* (3.91) *if*

$$T\big((\bar{x}, \bar{y}); \operatorname{gph} F\big) \cap \big[T(\bar{x}; \Omega) \times (-\Theta)\big] = \{0\}. \tag{3.99}$$

Similarly to the discussions and examples on relationships between local extremal and tangentially extremal points of set systems given in Section 3.1, we observe that the optimality notions in Definition 3.46 are *independent* of each other. Let us now compare the the graphical optimality of Definition 3.46(i) with fully localized minimizers of (3.93).

Proposition 3.47 (relationships between fully localized and graphical minimizers). *Let* $(\bar{x},\bar{y}) \in \operatorname{gph} F$ *be a feasible solution to problem* (3.91) *with general geometric constraints. Then the following assertions are satisfied:*

(i) $(\bar{x},\bar{y})$ *is a local graphical minimizer if it is a fully localized minimizer for this problem.*

(ii) *The opposite implication holds if there is a neighborhood* U *of* $\bar{x}$ *such that* $\bar{y} \notin F(x)$ *for every* $x \in \Omega \cap U$, $x \neq \bar{x}$.

Proof. To justify (i), assume that $(\bar{x},\bar{y})$ is a local graphical minimizer, take its neighborhood $U \times V$ from Definition 3.46(i), and pick any

$$y \in F(\Omega \cap U) \cap (\bar{y} - \Theta) \cap V.$$

Then there is $x \in \Omega \cap U$ such that $y \in F(x)$, and so

$$(x,y) \in \operatorname{gph} F \cap \big[\Omega \times (\bar{y} - \Theta)\big] \cap \big(U \times V\big) = \big\{(\bar{x},\bar{y})\big\}.$$

Thus $F(\Omega \cap U) \cap (\bar{y} - \Theta) \cap V = \{\bar{y}\}$, i.e., $(\bar{x},\bar{y})$ is a fully localized minimizer for (3.91).

Next we prove (ii). Suppose that $(\bar{x},\bar{y})$ is a fully localized minimizer with a neighborhood $U \times V$, shrink U so that the assumption in (ii) holds, and take

$$(x,y) \in \operatorname{gph} F \cap \big[\Omega \times (\bar{y} - \Theta)\big] \cap \big(U \times V\big).$$

Since $y \in F(x)$, it follows that $y \in F(\Omega \cap U) \cap (\bar{y} - \Theta) \cap V = \{\bar{y}\}$. If $x \neq \bar{x}$, the latter contradicts the assumption in (ii). Thus $x = \bar{x}$, which completes the proof of the proposition. $\square$

The next theorem uses the full strength of the tangential extremal principle in Section 3.1 justifying the necessary optimality conditions of Theorem 3.45 for *tangential* graphical minimizers of the multiobjective problem (3.91) with countable constraints without imposing the normal regularity requirement of the feasible set.

Theorem 3.48 (optimality conditions for tangential graphical minimizers). *Let* $(\bar{x}, \bar{y})$ *be a local tangential graphical minimizer for problem* (3.91) *under the fulfillment all the assumptions of Theorem* 3.45 *but the normal regularity of* Ω *at* $\bar{x}$. *Suppose in addition that* $\text{int}\Theta \neq \emptyset$. *Then there is* $0 \neq y^* \in -N(0;\Theta)$ *such that the necessary optimality condition* (3.95) *is satisfied.*

Proof. We have by Definition 3.46(ii) that $T((\bar{x},\bar{y});\text{gph}\, F) \cap \big[\Lambda \times (-\Theta)\big] = \{0\}$ with $\Lambda := T(\bar{x};\Omega)$. Since the system $\{\Omega_i\}_{i\in\mathbb{N}}$ has the CHIP at $\bar{x}$, it follows that

$$\Lambda = \bigcap_{i=1}^{\infty} \Lambda_i \ \text{ with } \ \Lambda_i := T(\bar{x};\Omega_i).$$

Further, define the closed cones $\Gamma_0 := T((\bar{x},\bar{y});\text{gph}\, F)$ and $\Gamma_i := \Lambda_i \times (-\Theta)$ as $i \in \mathbb{N}$ with $\bigcap_{i=0}^{\infty} \Gamma_i = \{0\}$ and show that for any $\xi \in \Theta \setminus \{0\}$ we get

$$\bigcap_{i=1}^{\infty} \Gamma_i \bigcap \Big[\Gamma_0 + (0,\xi)\Big] = \emptyset. \tag{3.100}$$

Indeed, supposing the contrary gives us a vector $(x,y) \in \mathbb{R}^n \times \mathbb{R}^m$ with $(x, y-\xi) \in \Gamma_0$ and $(x,y) \in \Lambda_i \times (-\Theta)$ for all $i \in \mathbb{N}$. Since Θ is a convex cone, we also have the inclusion $(x, y-\xi) \in \Lambda_i \times (-\Theta) = \Gamma_i$ as $i \in \mathbb{N}$, and hence $(x, y-\xi) \in \bigcap_{i=0}^{\infty} \Gamma_i = \{0\}$.

It follows therefore that $y = \xi \in -\Theta$, which implies by the pointedness of the cone Θ that $\xi \in (-\Theta) \cap \Theta = \{0\}$, a contradiction justifying (3.100).

The latter means that $\{\Gamma_i\}, i = 0,1,\ldots$, is a countable system of cones extremal at the origin with the nonoverlapping condition $\bigcap_{i=0}^{\infty}\Gamma_i = \{0\}$. Now applying the tangential extremal principle of Theorem 3.17 to this system of cones, we get elements (x_i^*, y_i^*) as $i = 0,1,\ldots$ satisfying the relationships

$$(x_0^*, y_0^*) \in N(0;\Gamma_0) \subset N\big((\bar{x},\bar{y});\operatorname{gph} F\big), \tag{3.101}$$

$$(x_i^*, y_i^*) \in N(0;\Gamma_i) \subset N(\bar{x};\Omega_i) \times \big[-N(0;\Theta)\big], \quad i \in \mathbb{N}, \tag{3.102}$$

$$\sum_{i=0}^{\infty}\frac{1}{2^i}\Big(x_i^*, y_i^*\Big) = 0, \quad \text{and} \quad \sum_{i=0}^{\infty}\frac{1}{2^i}\Big(\|x_i^*\|^2 + \|y_i^*\|^2\Big) = 1. \tag{3.103}$$

It follows from (3.101)–(3.103) that

$$x_0^* \in D^*F(\bar{x},\bar{y})(-y_0^*) \;\text{ and }\; -y_0^* = \sum_{i=1}^{\infty}\frac{1}{2^i}y_i^* \in -N(0;\Theta), \tag{3.104}$$

where the latter inclusion holds by the convexity and closedness of the cone $N(0;\Theta)$.

There are the two possible cases in (3.104): $y_0^* \neq 0$ and $y_0^* = 0$. In the first case we get

$$0 \in D^*F(\bar{x},\bar{y})(-y_0^*) + \sum_{i=1}^{\infty}\frac{1}{2^i}x_i^*,$$

which readily implies the optimality condition (3.95) with $0 \neq y^* := -y_0^* \in -N(0;\Theta)$; cf. the proof of the second part of Theorem 3.23.

To complete the proof of this theorem, it remains to show that the case of $y_0^* = 0$ in (3.104) cannot be realized under the imposed qualification conditions (3.57) and (3.94). Indeed, for

$y_0^* = 0$ we have from (3.102) and (3.104) that

$$-\frac{1}{2}y_1^* = \sum_{i=2}^{\infty} \frac{1}{2^i} y_i^* \in \big[-N(0;\Theta)\big] \cap N(0;\Theta). \tag{3.105}$$

Since the cone Θ is convex, it follows from (3.105) that

$$\langle y_1^*, y\rangle \le 0 \ \text{ and } \ \langle y_1^*, y\rangle \ge 0 \ \text{ for any } \ y \in \Theta,$$

i.e., $\langle y_1^*, y\rangle = 0$ on Θ. The latter implies that $y_1^* = 0$ by $\mathrm{int}\Theta \neq \emptyset$.

Proceeding in this way by induction gives us that $y_i^* = 0$ for all $i \in \mathbb{N}$. Now it follows from (3.102) and the first inclusion in (3.104) that $x_0^* = 0$ by the assumed coderivative qualification condition (3.94). Hence we get from (3.103) the relationships

$$\sum_{i=0}^{\infty} \frac{1}{2^i} x_i^* = 0 \ \text{ and } \ \sum_{i=0}^{\infty} \frac{1}{2^i} \|x_i^*\|^2 = 1,$$

which contradict the assumed NQC (3.57) and thus complete the proof of the theorem. □

Note in conclusion that, similarly to Section 3.7, we can develop necessary optimality conditions for multiobjective problems with countable constraints of operation and inequality types.

Chapter 4

Rated Extremal Principles

4.1 Rated Extremality of Finite Systems of Sets

In this first section of this chapter, we introduce a new notion of *rated extremality* for finite systems of sets, which essentially broader the previous notion (1.2) of local extremality. We show nevertheless that both exact and approximate versions of the extremal principle hold for this rated extremality under the same assumptions as in [47] for locally extremal points. Let us start with the definition of rated extremal points. For simplicity we drop the word "local" for rated extremal points in what follows.

Definition 4.1 (Rated extremal points of finite set systems). *Let $\Omega_1,\ldots,\Omega_m$ as $m\ge 2$ be nonempty subsets of X, and let $\bar{x}$ be a common point of these sets. We say that $\bar{x}$ is a* (*local*) RATED EXTREMAL POINT *of rank α, $0\le\alpha<1$, of the set system $\{\Omega_1,\ldots,\Omega_m\}$ if there are $\gamma>0$ and sequences $\{a_{ik}\}\subset X$, $i=1,\ldots,m$, such that $r_k:=\max_i\|a_{ik}\|\to 0$ as $k\to\infty$ and*

$$\bigcap_{i=1}^{m}\big(\Omega_i-a_{ik}\big)\cap B(\bar{x},\gamma r_k^{\alpha})=\emptyset\quad \textit{for all large } k\in\mathbb{N}. \tag{4.1}$$

In this case we say that $\{\Omega_1,\ldots,\Omega_m\}$ is a RATED EXTREMAL SYSTEM *at $\bar{x}$.*

The case of local extremality (1.2) obviously corresponds to (4.1) with rate $\alpha=0$. The next example shows that there are rated extremal points for systems of two simple sets in $\mathbb{R}^2$, which are not locally extremal in the conventional sense of (1.2).

Example 4.2 (Rated extremality versus local extremality). Consider the sets $\Omega_1:=\big\{(x_1,x_2)\in\mathbb{R}^2\big|\ x_2-x_1^2\le 0\big\}$ and $\Omega_2:=\big\{(x_1,x_2)\in\mathbb{R}^2\big|\ -x_2-x_1^2\le 0\big\}$. Then it is easy to check that $(\bar{x}_1,\bar{x}_2)=(0,0)\in\Omega_1\cap\Omega_2$ is a rated extremal point of rank $\alpha=\frac{1}{2}$ for the system

$\{\Omega_1, \Omega_2\}$ but not a local extremal point of this system.

Prior to proceeding with the results in this section, we briefly discuss relationships between the rated extremality and the *tangential extremality* of set systems introduced in Chapter 3. The next proposition result and the subsequent example reveal relationships between the rated extremality and tangential extremality of set systems.

Proposition 4.3 (relationships between rated and tangential extremality of finite systems of sets). *Let $\{\Omega_1, \ldots, \Omega_m\}$ as $m \geq 2$ be a Λ-tangential extremal system of sets at $\bar{x}$. Assume that there are real numbers $C > 0$, $p \in (0,1)$ and a neighborhood U of $\bar{x}$ such that*

$$\text{dist}\,(x - \bar{x}; \Lambda_i) \leq C\|x - \bar{x}\|^{1+p} \quad \text{for all} \;\; x \in \Omega_i \cap U \quad \text{and} \;\; i = 1, \ldots, m. \tag{4.2}$$

Then $\{\Omega_1, \ldots, \Omega_m\}$ is a rated extremal system at $\bar{x}$.

Proof. Since the general case of $m \geq 2$ can be derived by induction, it suffices to justify the result in the case of $m = 2$. Let $\{\Lambda_1, \Lambda_2\}$ be an extremal system of approximation cones and find by definition elements $a_1, a_2 \in X$ such that

$$(\Lambda_1 - a_1) \cap (\Lambda_2 - a_2) = \emptyset.$$

Without loss of generality, assume that $a_1 = -a_2 =: a$. Take $\alpha \in (0,1)$ with $\beta := \alpha(1 + p) > 1$ and show that for all small $t > 0$ we have

$$(\Omega_1 - ta) \cap (\Omega_2 + ta) \cap B(\bar{x}, \|ta\|^\alpha) = \emptyset. \tag{4.3}$$

Suppose by contradiction that there exists

$$x \in (\Omega_1 - ta) \cap (\Omega_2 + ta) \cap B(\bar{x}, \|ta\|^{\alpha}). \tag{4.4}$$

That implies by using condition (4.2) that

$$\text{dist}\,(x - \bar{x}; \Lambda_1 - ta) = \text{dist}\,(x + ta - \bar{x}; \Lambda_1) \le C\|x + ta - \bar{x}\|^{1+p},$$

$$\text{dist}\,(x - \bar{x}; \Lambda_2 + ta) = \text{dist}\,(x - ta - \bar{x}; \Lambda_2) \le C\|x - ta - \bar{x}\|^{1+p}.$$

Thus we have for some constant $\widetilde{C}$ that

$$\|x + ta - \bar{x}\|^{1+p} \le \widetilde{C} \max\left\{\|x - \bar{x}\|, \|ta\|\right\}^{1+p} \le \widetilde{C} \max\left\{\|ta\|^{\beta}, \|ta\|^{1+p}\right\} = o(\|ta\|) \ \text{ as } \ t \downarrow 0$$

and similarly $\|x - ta - \bar{x}\|^{1+p} = o(\|ta\|)$. Put then $d := \text{dist}\,(\Lambda_1 - a, \Lambda_2 + a) > 0$ and observe due the conic structures of Λ_1 and Λ_2 that

$$td = \text{dist}\,(\Lambda_1 - ta; \Lambda_2 + ta) > 0$$

for all $t > 0$ sufficiently small. Combining all the above gives us

$$td = \text{dist}\,(\Lambda_1 - ta; \Lambda_2 + ta) \le \text{dist}\,(x - \bar{x}; \Lambda_1 - ta) + \text{dist}\,(x - \bar{x}; \Lambda_2 + ta) = o(\|ta\|),$$

which is a contradiction. Thus $\{\Omega_1, \Omega_2, \bar{x}\}$ is a rated extremal system at $\bar{x}$ with rank α chosen above. This completes the proof of the proposition. □

One of the most important special cases of tangential extremality is the so-called *contingent extremality* when the approximating cones to Ω_i are given by the Bouligand-Severi contingent cones to this sets. The following example (of two parts) shows that the notions of rated extremality and contingent extremality are independent from each other in a simple setting of two sets in $\mathbb{R}^2$.

Example 4.4 (independence of rated and contingent extremality). Let $X = \mathbb{R}^2$, and let $\bar{x} = (0,0)$.

(i) Consider two closed sets in $\mathbb{R}^2$ given by

$$\Omega_1 := \operatorname{epi} f \quad \text{and} \quad \Omega_2 := \mathbb{R} \times \mathbb{R}_- \setminus \operatorname{int} \Omega_1,$$

where $f(x) := x \sin \frac{1}{x}$ for $x \in \mathbb{R}$ with $f(0) := 0$. It is easy to see that the contingent cones to Ω_1 and Ω_2 at $\bar{x}$ are computed by

$$\Lambda_1 = \operatorname{epi}(-|\cdot|) \quad \text{and} \quad \Lambda_2 = \mathbb{R} \times \mathbb{R}_-.$$

We can check that the set system $\{\Omega_1, \Omega_2\}$ is locally extremal at $\bar{x}$, and hence $\bar{x}$ is a rated extremal point of this system of sets with rank $\alpha = 0$. On the other hand, the contingent extremality is obviously violated for $\{\Omega_1, \Omega_2\}$ at $\bar{x}$ as follows from the above computations of Λ_1 and Λ_2.

(ii) Now we define two closed sets in $\mathbb{R}^2$ by

$$\Omega_1 := \mathbb{R} \times \mathbb{R}_- \quad \text{and} \quad \Omega_2 := \operatorname{epi} f \quad \text{with} \quad f(x) := -x^{1+\frac{1}{\ln^2 |x|}} \quad \text{for} \quad x \neq 0 \quad \text{and} \quad f(0) := 0.$$

The contingent cones to Ω_1 and Ω_2 at $\bar{x}$ are easily computed by $\Lambda_1 = \mathbb{R} \times \mathbb{R}_-$ and $\Lambda_2 = \mathbb{R} \times \mathbb{R}_+$. We can check that $\bar{x}$ is not a rated extremal point of $\{\Omega_1, \Omega_2\}$ whenever $\alpha \in [0,1)$, while the contingent extremality obviously holds for this system at $\bar{x}$.

The next theorem justifies the fulfillment of the exact extremal principle for any rated extremal point of a finite system of closed sets in $\mathbb{R}^n$. It extends the extremal principle of [47, Theorem 2.8] obtained for local extremal points, i.e., when $\alpha = 0$ in Definition 4.1.

Theorem 4.5 (Exact extremal principle for rated extremal systems of sets in finite dimensions). *Let $\bar{x}$ be a rated extremal point of rank $\alpha \in [0,1)$ for the system of sets $\{\Omega_1, \ldots, \Omega_m\}$ as $m \geq 2$ in $\mathbb{R}^n$. Assume that all the sets Ω_i are locally closed around $\bar{x}$. Then the exact extremal principle holds for $\{\Omega_1, \ldots, \Omega_m\}$ at $\bar{x}$, i.e, there are $x_i^* \in N(\bar{x}; \Omega_i)$ for $i = 1, \ldots, m$ satisfying the relationships in* (1.1).

Proof. Given a rated extremal point $\bar{x}$ of the system $\{\Omega_1, \ldots, \Omega_m\}$, take numbers $\alpha \in [0,1)$ and $\gamma > 0$ as well as sequences $\{a_{ik}\}$ and $\{r_k\}$ from Definition 4.1. Consider the following unconstrained minimization problem for any fixed $k \in \mathbb{N}$:

$$\text{minimize } \; d_k(x) := \left[\sum_{i=1}^m \text{dist}^{\,2}\big(x + a_{ik}; \Omega_i\big)\right]^{\frac{1}{2}} + \frac{\sqrt{m}}{\gamma^{\frac{1}{\alpha}}} \|x - \bar{x}\|^{\frac{1}{\alpha}} \;, \; x \in \mathbb{R}^n. \tag{4.5}$$

Since the function d_k is continuous and its level sets are bounded, there exists an optimal solution x_k to (4.5) by the classical Weierstrass theorem. We obviously have the relationships

$$d_k(x_k) \leq d_k(\bar{x}) = \left[\sum_{i=1}^m \text{dist}^{\,2}\big(\bar{x} + a_{ik}; \Omega_i\big)\right]^{\frac{1}{2}} \leq \left[\sum_{i=1}^m \|a_{ik}\|^2\right]^{\frac{1}{2}} \leq r_k\sqrt{m},$$

which readily imply the estimate

$$\frac{\sqrt{m}}{\gamma^{\frac{1}{\alpha}}} \|x_k - \bar{x}\|^{\frac{1}{\alpha}} \leq r_k\sqrt{m}, \;\; \text{i.e.,} \;\; \|x_k - \bar{x}\| \leq \gamma r_k^{\alpha}.$$

Taking the latter into account, we get

$$\nu_k := \left[\sum_{i=1}^{m} \text{dist}^2\big(x_k + a_{ik}; \Omega_i\big) \right]^{\frac{1}{2}} > 0,$$

since the opposite statement $\nu_k = 0$ contradicts the rated extremality of $\bar{x}$. Furthermore, the optimality of x_k in (4.5) and choice of $\{a_{ik}\}$ give us the relationships

$$d_k(x_k) = \nu_k + \frac{\sqrt{m}}{\gamma^{\frac{1}{\alpha}}} \|x_k - \bar{x}\|^{\frac{1}{\alpha}} \leq \left[\sum_{i=1}^{m} \|a_{ik}\|^2 \right]^{\frac{1}{2}} \downarrow 0 \text{ as } k \to \infty,$$

which ensure in turn that $x_k \to \bar{x}$ and $\nu_k \downarrow 0$ as $k \to \infty$.

We now arbitrarily pick $w_{ik} \in \Pi(x_k + a_{ik}; \Omega_i)$ for $i = 1, \ldots, m$ in the closed set Ω_i and for each $k \in \mathbb{N}$ consider the problem:

$$\text{minimize } \rho_k(x) := \left[\sum_{i=1}^{m} \|x + a_{ik} - w_{ik}\|^2 \right]^{\frac{1}{2}} + \frac{\sqrt{m}}{\gamma^{\frac{1}{\alpha}}} \|x - \bar{x}\|^{\frac{1}{\alpha}}, \quad x \in \mathbb{R}^n, \tag{4.6}$$

which obviously has the same optimal solution x_k as for (4.5). Since $\nu_k > 0$ and the norm $\|\cdot\|$ is Euclidian, the function $\rho_k(\cdot)$ in (4.6) is continuously differentiable around x_k. Thus applying the classical Fermat rule to the *smooth* unconstrained minimization problem (4.6), we get

$$\nabla \rho_k(x_k) = \sum_{i=1}^{m} x_{ik}^* + C\|x_k - \bar{x}\|^{\frac{1-2\alpha}{\alpha}}(x_k - \bar{x}) = 0 \quad \text{for some constant } C,$$

where $x_{ik}^* := (x_k + a_{ik} - w_{ik})/\nu_k$ for $i = 1, \ldots, m$ with $\|x_{1k}^*\|^2 + \ldots + \|x_{mk}^*\|^2 = 1$.

Observe that $\|x_k - \bar{x}\|^{\frac{1-2\alpha}{\alpha}}(x_k - \bar{x}) = \|x_k - \bar{x}\|^{\frac{1-\alpha}{\alpha}} \dfrac{x_k - \bar{x}}{\|x_k - \bar{x}\|} \to 0$ as $x_k \to \bar{x}$. Due to the compactness of the unit sphere in $\mathbb{R}^n$, we find $x_i^* \in \mathbb{R}^n$ as $i = 1, \ldots, m$ such that $x_{ik}^* \to x_i^*$ as $k \to \infty$ without relabeling. It follows from the equivalent description (2.6) of the limiting normal cone that $x_i^* \in N(\bar{x};\Omega_i)$ for all $i = 1,\ldots,m$. Moreover, we get from the constructions above that

$$\|x_1^*\|^2 + \ldots + \|x_m^*\|^2 = 1 \;\text{ and }\; x_1^* + \ldots + x_m^* = 0.$$

This gives all the conclusions of the exact extremal principle and completes the proof of the theorem. □

The next example shows that the exact extremal principle is violated if we take $\alpha = 1$ in Definition 4.1.

Example 4.6 (Violating the exact extremal principle for rated extremal points of rank $\alpha = 1$). Define two closed sets in $\mathbb{R}^2$ by

$$\Omega_1 := \operatorname{epi}(-\|\cdot\|) \;\text{ and }\; \Omega_2 := \mathbb{R} \times \mathbb{R}_-.$$

Taking any $a_k \downarrow 0$, we see that

$$\big(\Omega_1 + (0,a_k)\big) \cap \big(\Omega_1 - (0,a_k)\big) \cap B(\bar{x}, a_k/2) = \emptyset,$$

i.e., $\bar{x} = (0,0)$ is a rated extremal point of $\{\Omega_1,\Omega_2\}$ of rank $\alpha = 1$. However, it is easy to check that the relationships of the exact extremal principle do not hold for this system at $\bar{x}$.

Observe that Example 4.6 shows that the relationships of the approximate extremal principle are also violated when $\alpha = 1$. However, for rated extremal systems of rank $\alpha \in [0,1)$ the approximate extremal principle holds in general infinite-dimensional settings. Let us proceed with justifying this statement extending the corresponding results of [47] obtained for the rank

$\alpha = 0$ in Definition 4.1.

Theorem 4.7 (Approximate extremal principle for rated extremal systems in Fréchet smooth spaces). *Let X be a Banach space admitting an equivalent norm Fréchet differentiable off the origin, and let $\bar{x}$ be a rated extremal point of rank $\alpha \in [0,1)$ for a system of sets $\Omega_1, \ldots, \Omega_m$ locally closed around $\bar{x}$. Then the approximate extremal principle holds for $\{\Omega_1, \ldots, \Omega_m\}$ at $\bar{x}$.*

Proof. Choose an equivalent norm $\|\cdot\|$ on X differentiable off the origin and consider first the case of $m = 2$ in the theorem. Let $\bar{x} \in \Omega_1 \cap \Omega_2$ be a rated extremal point of rank $\alpha \in [0,1)$ with $\gamma > 0$ taken from Definition 4.1. Denote $r := \max\{\|a_1\|, \|a_2\|\}$ and for any $\varepsilon > 0$ find a_1, a_2 such that

$$r^{1-\alpha} \le \min\left\{\frac{\gamma}{2}, \frac{\varepsilon}{(2\gamma)^{(1-\alpha)/\alpha}}\right\} \text{ and } \big(\Omega_1 - a_1\big) \cap \big(\Omega_2 - a_2\big) \cap B\big(\bar{x}, \gamma r^{\alpha}\big) = \emptyset.$$

We also select a constant $C > 0$ with $(\frac{2}{C})^{\alpha} = \frac{\gamma}{2}$ and denote $\beta := \frac{1}{\alpha} > 1$. Define the function

$$\varphi(z) := \|(x_1 - a_1) - (x_2 - a_2)\| \quad \text{for} \quad z = (x_1, x_2) \in X \times X \tag{4.7}$$

with the product norm $\|z\| := (\|x_1\|^2 + \|x_2\|^2)^{1/2}$ on $X \times X$, which is Fréchet differentiable off the origin under this property of the norm on X. Next fix $z_0 = (\bar{x}, \bar{x})$ and define the set

$$W(z_0) := \Big\{z \in \Omega_1 \times \Omega_2 \big| \varphi(z) + C\|z - z_0\|^{\beta} \le \varphi(z_0)\Big\}, \tag{4.8}$$

which is obviously nonempty and closed. For each $z = (x_1, x_2) \in W(z_0)$ we have $i = 1, 2$: $C\|x_i - \bar{x}\|^{\beta} \le C\|z - \bar{z}\|^{\beta} \le \varphi(z_0) = \|-a_1 + a_2\| \le 2r$.

That implies $\|x_i - \bar{x}\| \le \left(\frac{2}{C}\right)^{\frac{1}{\beta}} r^{\frac{1}{\beta}} = \left(\frac{2}{C}\right)^{\alpha} r^{\alpha} = \frac{\gamma}{2} r^{\alpha}$ and thus

$$W(z_0) \subset B(\bar{x}, \gamma r^{\alpha}) \times B(\bar{x}, \gamma r^{\alpha}) \subset B\big(\bar{x}, \tfrac{1}{2}\varepsilon^{\frac{\alpha}{1-\alpha}}\big) \times B\big(\bar{x}, \tfrac{1}{2}\varepsilon^{\frac{\alpha}{1-\alpha}}\big).$$

It follows from Definition 4.1 and constructions (4.7) and (4.8) that $\varphi(z) > 0$ for all $z \in W(x_0)$. Indeed, assuming on the contrary that $\varphi(z) = 0$ for some $z = (x_1, x_2) \in W(x_0)$ gives us

$$\|x_1 - a_1 - \bar{x}\| \le \|x_1 - \bar{x}\| + \|a_1\| \le \tfrac{\gamma}{2} r^{\alpha} + r = \left(\tfrac{\gamma}{2} + r^{1-\alpha}\right) r^{\alpha} \le \gamma r^{\alpha}$$

and thus $x_1 - a_1 = x_2 - a_2 \in \big(\Omega_1 - a_1\big) \cap \big(\Omega_2 - a_2\big) \cap B(\bar{x}, \gamma r^{\alpha}) \neq \emptyset$, a contradiction.

Hence φ is Fréchet differentiable at any point $z \in W(z_0)$. Pick any $z_1 \in \Omega_1 \times \Omega_2$ satisfying

$$\varphi(z_1) + C\|z_1 - z_0\|^{\beta} \le \inf_{W(z_0)} \Big\{\varphi(z) + C\|z - z_0\|^{\beta}\Big\} + \frac{r}{2}$$

and define further the nonempty and closed set

$$W(z_1) := \left\{ z \in \Omega_1 \times \Omega_2 \,\Big|\, \varphi(z) + C\|z - z_0\|^{\beta} + C\frac{\|z - z_1\|^{\beta}}{2} \le \varphi(z_1) + C\|z_1 - z_0\|^{\beta} \right\}.$$

Arguing inductively, suppose we have z_k and $W(z_k)$, then pick $z_{k+1} \in W(z_k)$ such that

$$\varphi(z_{k+1}) + C\sum_{i=0}^{k} \frac{\|z_{k+1} - z_i\|^{\beta}}{2^i} \le \inf_{W(z_k)} \Big\{\varphi(z) + C\sum_{i=0}^{k} \frac{\|z - z_i\|^{\beta}}{2^i}\Big\} + \frac{r}{2^{2k+1}}$$

and construct the subsequent nonempty and closed set

$$W(z_{k+1}) := \left\{ z \in \Omega_1 \times \Omega_2 \,\Big|\, \varphi(z) + C\sum_{i=0}^{k+1} \frac{\|z - z_i\|^{\beta}}{2^i} \le \varphi(z_{k+1}) + C\sum_{i=0}^{k} \frac{\|z_{k+1} - z_i\|^{\beta}}{2^i} \right\}.$$

It is easy to see that the sequence $\{W(z_k)\} \subset \Omega_1 \times \Omega_2$ is nested. Let us check that

$$\operatorname{diam} W(z_{k+1}) := \sup\big\{\|z-w\| \,\big|\, z,w \in W(z_{k+1})\big\} \to 0 \ \text{ as } \ k \to \infty. \tag{4.9}$$

Indeed, for each $z \in W(z_{k+1})$ and $k \in \mathbb{N}$ we have

$$\begin{aligned} C\frac{\|z-z_{k+1}\|^\beta}{2^{k+1}} &\le \varphi(z_{k+1}) + C\sum_{i=0}^{k}\frac{\|z_{k+1}-z_i\|^\beta}{2^i} - \left(\varphi(z) + C\sum_{i=0}^{k}\frac{\|z-z_i\|^\beta}{2^i}\right) \\ &\le \varphi(z_{k+1}) + C\sum_{i=0}^{k}\frac{\|z_{k+1}-z_i\|^\beta}{2^i} - \inf_{W(z_k)}\left\{\varphi(z) + C\sum_{i=0}^{k}\frac{\|z-z_i\|^\beta}{2^i}\right\} \le \frac{r}{2^{2k+1}}, \end{aligned}$$

which implies that $\operatorname{diam} W(z_{k+1}) \le 2\left(\frac{r}{C2^k}\right)^{\frac{1}{\beta}}$ and thus justifies (4.9). Due to the completeness of X the classical Cantor theorem ensures the existence of $\bar z = (\bar x_1, \bar x_2) \in W(z_0)$ such that $\bigcap_{k=0}^{\infty} W(z_k) = \{\bar z\}$ with $z_k \to \bar z$ as $k \to \infty$. Now we show that $\bar z$ is a minimum point of the function

$$\phi(z) := \varphi(z) + C\sum_{i=0}^{\infty}\frac{\|z-z_i\|^\beta}{2^i} \tag{4.10}$$

over the set $\Omega_1 \times \Omega_2$. To proceed, take any $\bar z \ne z \in \Omega_1 \times \Omega_2$ and observe that $z \notin W(z_k)$ for all $k \in \mathbb{N}$ sufficiently large while $\bar z \in W(z_k)$. This yields the estimates

$$\phi(z) \ge \varphi(z) + C\sum_{i=0}^{k}\frac{\|z-z_i\|^\beta}{2^i} \ge \varphi(z_k) + C\sum_{i=0}^{k-1}\frac{\|z_k-z_i\|^\beta}{2^i} \ge \varphi(\bar z) + C\sum_{i=0}^{k}\frac{\|\bar z-z_i\|^\beta}{2^i}$$

and hence justifies the claimed inequality $\phi(z) \ge \phi(\bar z)$ by letting $k \to \infty$.

We get therefore that the function $\phi(z) + \delta(z;\Omega_1 \times \Omega_2)$ attains at $\bar z$ its minimum on the whole space $X \times X$. The generalized Fermat rule gives us the inclusion $0 \in \widehat\partial\big(\phi(z) + \delta(z;\Omega_1 \times \Omega_2)\big)$. Since $\varphi(\bar z) > 0$ and the norm $\|\cdot\|^\beta$ is smooth, the function ϕ in (4.10) is Fréchet differentiable at $\bar{z}$. Applying the sum rule from [47, Proposition 1.107], the Fréchet subdifferential formula for the indicator function, and the product formula for Fréchet normal cone (2.3) from [47, Proposition 1.2], we get

$$-\nabla\phi(\bar{z}) = -(u_1^*, u_2^*) \in \widehat{N}(\bar{z}; \Omega_1 \times \Omega_2) = \widehat{N}(\bar{x}_1; \Omega_1) \times \widehat{N}(\bar{x}_2; \Omega_2),$$

where the dual elements u_i^*, $i = 1, 2$, are computed by

$$u_1^* = x^* + \sum_{j=0}^{\infty} w_{1j}^* \frac{\|\bar{x}_1 - x_{1j}\|^{\beta-1}}{2^j} \quad \text{and} \quad u_2^* = -x^* + \sum_{j=0}^{\infty} w_{2j}^* \frac{\|\bar{x}_2 - x_{2j}\|^{\beta-1}}{2^j}$$

with $z_j = (x_{1j}, x_{2j})$, $x^* = \nabla\big(\|\cdot\|\big)\big((\bar{x}_1 - a_1) - (\bar{x}_2 - a_2)\big)$, and

$$w_{ij}^* = \begin{cases} \nabla(\|\cdot\|)(\bar{x}_i - x_{ij}) & \text{if } \bar{x}_i - x_{ij} \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

for $i = 1, 2$ and $j = 0, 1, \ldots$ due to the construction of the function ϕ in (4.10). Observing further that $\|x^*\| = 1$ and that $\bar{z}, z_i \in W(z_0)$ gives us

$$\|\bar{x}_i - x_{ij}\| \le \varepsilon^{\frac{1-\alpha}{\alpha}} = \varepsilon^{\frac{1}{\beta-1}},$$

which implies the estimates $\|\bar{x}_i - x_{ij}\|^{\beta-1} \le \varepsilon$ and

$$\sum_{j=0}^{\infty} \|w_{ij}^*\| \frac{\|\bar{x}_i - x_{ij}\|^{\beta-1}}{2^j} \le 2\varepsilon, \quad i = 1, 2.$$

Setting finally $x_1^* := -x^*/2$, $x_2^* := x^*/2$, and $x_i := \bar{x}_i$ for $i = 1, 2$, we arrive at the relationships

$$\|x_1^*\| + \|x_2^*\| = 1, \quad \text{and} \quad x_1^* + x_2^* = 0,$$

with $x_i^* \in \widehat{N}(x_i; \Omega_i) + \varepsilon B^*$, $x_i \in B(\bar{x}, \varepsilon)$ for $i = 1, 2$, which show that the approximate extremal principle holds for rated extremal points of two sets.

Consider now the general case of $m > 2$ sets. Observe that if $\bar{x}$ as a rated extremal point of the system $\{\Omega_1, \ldots, \Omega_m\}$ with some rank $\alpha \in [0, 1)$, then the point $\bar{z} := (\bar{x}, \ldots, \bar{x}) \in X^{n-1}$ is a local rated extremal point of the same rank for the system of two sets

$$\Theta_1 := \Omega_1 \times \ldots \times \Omega_{n-1} \quad \text{and} \quad \Theta_2 := \big\{(x, \ldots, x) \in X^{n-1} \big| x \in \Omega_m\big\}. \tag{4.11}$$

To justify this, take numbers $\alpha \in [0, 1)$ and $\gamma > 0$ and the sequences $(a_{1k}, \ldots, a_{mk})$ from Definition 4.1 for m sets and check that

$$\Big(\Theta_1 - (a_{1k}, \ldots, a_{n-1,k})\Big) \cap \Big(\Theta_2 - (a_{nk}, \ldots, a_{nk})\Big) \cap B\big((\bar{x}, \ldots, \bar{x}); \gamma r_k^\alpha\big) = \emptyset \tag{4.12}$$

with $r_k := \max\{\|a_{1k}\|, \ldots, \|a_{nk}\|\}$. Indeed, the violation of (4.12) means that there are $x_m \in \Omega_m$ and $(x_1, ..., x_{n-1}) \in \Omega_1 \times ... \times \Omega_{n-1}$ satisfying

$$x_1 - a_{1k} = \ldots = x_{m-1} - a_{m-1,k} = x_m - a_{mk} \in B(\bar{x}, \gamma r_k^\alpha),$$

which clearly contradicts the rated extremality of $\bar{x}$ with rank α for the system $\{\Omega_1, \ldots, \Omega_m\}$. Applying finally the relationships of the approximate extremal principle to the system of two sets in (4.11) and taking into account the structures of these sets as well as the aforementioned product formula for Fréchet normals, we complete the proof of the theorem. $\square$

The next theorem elevates the fulfillment of the approximate extremal principle for rated extremal points from Fréchet smooth to Asplund spaces by using the method of *separable reduction*; see [20, 47].

Theorem 4.8 (Approximate extremal principle for rated extremal systems in Asplund spaces). *Let X be an Asplund space, and let $\bar{x}$ be a rated extremal point of rank $\alpha \in [0,1)$ for a system of sets $\Omega_1,\ldots,\Omega_m$ locally closed around $\bar{x}$. Then the approximate extremal principle holds for $\{\Omega_1,\ldots,\Omega_m\}$ at $\bar{x}$.*

Proof. Taking a rated extremal point $\bar{x}$ for the system $\{\Omega_1,\ldots,\Omega_m\}$ of rank $\alpha \in [0,1)$, find a number $\gamma > 0$ and sequences $\{a_{ik}\}$, $i = 1,\ldots,m$, from Definition 4.1. Consider a separable subspace Y_0 of the Asplund space X defined by

$$Y_0 := \operatorname{span}\big\{\bar{x}, a_{ik}\big|\; i = 1,\ldots,m,\; k \in \mathbb{N}\big\}.$$

Pick now a closed and separable subspace $Y \subset X$ with $Y \supset Y_0$ and observe that $\bar{x}$ is a rated extremal point of rank α for the system $\{\Omega_1 \cap Y,\ldots,\Omega_m \cap Y\}$. Indeed, we have

$$\begin{aligned}&\Big((\Omega_1 \cap Y) - a_{1k}\Big) \cap \ldots \cap \Big((\Omega_m \cap Y) - a_{mk}\Big) \cap B_Y(\bar{x};\gamma r_k^{\alpha})\\ &\subset \Big(\Omega_1 - a_{1k}\Big) \cap \ldots \cap \Big(\Omega_m - a_{mk}\Big) \cap B_X(\bar{x};\gamma r_k^{\alpha}) = \emptyset,\end{aligned}$$

where $r_k := \max\{\|a_{1k}\|,\ldots,\|a_{mk}\|\}$, and where B_X and B_Y are the closed unit balls in the space X and Y, respectively. The rest of the proof follows the one in [47, Theorem 2.20] by taking into account that Y admits an equivalent Fréchet differentiable norm off the origin. $\square$

We conclude this section with deriving the exact extremal principle for rated extremal systems of rank $\alpha \in [0,1)$ in Asplund spaces extending the corresponding result of [47, Theorem 2.22] obtained for $\alpha = 0$.

Theorem 4.9 (Exact extremal principle for rated extremal systems in Asplund spaces). *Let X be an Asplund space, and let $\bar{x}$ be a rated extremal point of rank $\alpha \in [0,1)$ for a system of sets $\Omega_1, \ldots, \Omega_m$ locally closed around $\bar{x}$. Assume that all but one of the sets Ω_i, $i = 1, \ldots, m$, are SNC at $\bar{x}$. Then the exact extremal principle holds for $\{\Omega_1, \ldots, \Omega_m\}$ at $\bar{x}$.*

Proof. Follows the lines in the proof of [47, Theorem 2.22] by passing to the limit in the relationships of the rated approximate extremal principle obtained in Theorem 4.8. □

4.2 Rated Extremal Principles for Infinite Set Systems

This section concerns new notions of rated extremality and deriving rated extremal principles for infinite systems of closed sets. The main results are obtained in the framework of Asplund spaces.

Let us start with introducing a notion of rated extremality for arbitrary (may be infinite and not even countable) systems of sets in general Banach spaces. We say that $R(\cdot)\colon \mathbb{R}_+ \to \mathbb{R}_+$ is a *rate function* if there is a real number M such that

$$rR(r) \le M \quad \text{and} \quad \lim_{r \downarrow 0} R(r) = \infty. \tag{4.13}$$

In what follow we denote by $|I|$ the cardinality (number of elements) of a finite set I.

Definition 4.10 (Rated extremality for infinite systems of sets). *Let $\{\Omega_i\}_{i \in T}$ be a system of closed subsets of X indexed by an arbitrary set T, and let $\bar{x} \in \bigcap_{t \in T} \Omega_i$. Given a rate function $R(\cdot)$, we say that $\bar{x}$ is an R-RATED EXTREMAL POINT of the system $\{\Omega_i\}_{i \in T}$ if there exist sequences $\{a_{ik}\} \subset X$, $i \in T$ and $k \in \mathbb{N}$, with $r_k := \sup_{i \in T} \|a_{ik}\| \to 0$ as $k \to \infty$ such* *that whenever* $k \in \mathbb{N}$ *there is a finite index subset* $I_k \subset T$ *of cardinality* $|I_k|^{3/2} = o(R_k)$ *with* $R_k := R(r_k)$ *satisfying*

$$\bigcap_{i \in I_k} \big(\Omega_i - a_{ik}\big) \cap B\big(\bar{x}; r_k R_k\big) = \emptyset \ \text{ for all large } \ k. \tag{4.14}$$

In this case we say that $\{\Omega_i\}_{i \in T}$ *is an* R-RATED EXTREMAL SYSTEM *at* $\bar{x}$.

It is easy to see that a finite rated extremal system of sets from Definition 4.1 is a particular case of Definition 4.10. Indeed, suppose that $\bar{x}$ is a rated extremal point of rank $\alpha \in [0, 1)$ for a finite set system $\{\Omega_1, \ldots, \Omega_m\}$, i.e., condition (4.1) is satisfied. Defining $R(r) := \frac{\gamma}{r^{1-\alpha}}$, we have that $rR(r) \to 0$ and $R(r) \to \infty$ as $r \to 0$; thus $R(\cdot)$ is a rate function while condition (4.14) is satisfied.

Let us discuss some specific features of the rated extremality in Definition 4.10 for the case of infinite systems. For simplicity we denote $R = R(r)$ in what follows if no confusion arises.

Remark 4.11 (Growth condition in rated extremality). Observe that, although $\{\Omega_i\}_{i \in T}$ is an infinite system in Definition 4.10, the rated extremality therein involves only *finitely many* sets for each given accuracy $\varepsilon > 0$. The imposed requirement $|I|^{3/2} = o(R)$ guarantees that $|I|^{3/2}$ *grows slower* than R, which is very crucial in our proof of the extremal principle below. In other words, the number of sets involved must not be *too large*; otherwise the result is trivial. We prove in Theorem 4.15 that the rate $|I|^{3/2} = o(R)$ ensures the validity of the rated extremal principle, where the number r measures *how far* the sets are shifted.

Define next extremality conditions for infinite systems of sets, which will be justified as an appropriate extremal principle in what follows. These conditions are of the approximate extremal principle type expressed in terms of Fréchet normals at nearby points.

Definition 4.12 (Rated extremality conditions for infinite systems). *Let* $\{\Omega_i\}_{i\in T}$ *be a system of nonempty subsets of* X *indexed by an arbitrary set* T*, and let* $\bar{x} \in \bigcap_{t\in T} \Omega_i$*. We say that the set system* $\{\Omega_i\}_{i\in T}$ *satisfies the* RATED EXTREMAL PRINCIPLE *at* $\bar{x}$ *if for any* $\varepsilon > 0$ *there exist a number* $r \in (0, \varepsilon)$*, a finite index subset* $I \subset T$ *with cardinality* $|I|r < \varepsilon$*, points* $x_i \in \Omega_i \cap B(\bar{x}, \varepsilon)$*, and dual elements* $x_i^* \in \widehat{N}(x_i; \Omega_i) + r\mathbb{B}^*$ *for* $i \in I$ *such that*

$$\sum_{i\in I} x_i^* = 0 \quad \textit{and} \quad \sum_{i\in I} \|x_i^*\|^2 = 1. \tag{4.15}$$

Observe that when a system consists of finitely many sets $\{\Omega_1, \ldots, \Omega_m\}$ with $|I| = m$, we put the other sets equal to the whole space X and reduce Definition 4.10 in this case to the conventional conditions of the approximate extremal principle for finite systems of sets; see Section 2.

Now we address the nontriviality issue for the introduced version of the extremal principle for infinite set systems. It is appropriate to say (roughly speaking) that a version of the extremal principle is *trivial* if all the information is obtained from only one set of the system while the other sets contribute nothing; i.e., if $y_i^* = 0 \in \widehat{N}(x_i; \Omega_i)$ for all but one index i. It has been shown in Chapter 3 that a “natural” extension of the approximate extremal principle for countable systems is trivial.

The next proposition justifies the nontriviality of the rated extremal principle for infinite set systems proposed in Definition 4.12.

Proposition 4.13 (Nontriviality of rated extremality conditions for infinite systems). *Let* $\{\Omega_i\}_{i\in T}$ *be a system of set satisfying the extremality conditions of Definition* 4.12 *at some point* $\bar{x} \in \bigcap_{t\in T} \Omega_i$*. Then the rated extremal principle defined by these conditions is nontrivial.*

Proof. Suppose on the contrary that the rated extremal principle of Definition 4.12 is trivial, i.e., there is $i_0 \in T$ (say $i_0 = 1$) and $y_i^* \in X^*$ as $i \in T$ such that

$$x_i^* \in y_i^* + r\mathbb{B}^* \subset \widehat{N}(x_i; \Omega_i) + r\mathbb{B}^* \ \text{ for all } \ i \in I,$$

$$\sum_{i \in I} x_i^* = 0, \ \sum_{i \in I} \|x_i^*\|^2 = 1, \ \text{ and } \ y_i^* = 0 \ \text{ whenever } \ i \in I \setminus \{1\}$$

in the notation of Definition 4.10. It follows that $\|x_i^*\| \le r$ for all $i \in I \setminus \{1\}$ implying that

$$\Big\| y_1^* + \sum_{i \neq 1} x_i^* \Big\| \le r \ \text{ and } \ \|y_1^*\| \le |I|r.$$

Thus we arrive at the relationships

$$\sum_{i \in I} \|x_i^*\|^2 < (\|y_1^*\| + r)^2 + \sum_{i \neq 1} r^2 \le |I|^2 r^2 + 2|I|r^2 + r^2 + (|I| - 1)r^2 < C\varepsilon^2 \downarrow 0$$

as $\varepsilon \downarrow 0$, a contradiction. This justifies the nontriviality of the rated extremal principle. □

Observe further that the extremal principle of Definition 4.12 may be trivial is the rate condition $|I|r < \varepsilon$ is not imposed. The following example describes a general setting when this happens.

Example 4.14 (The rate condition is essential for nontriviality). Assume that the condition $|I|r < \varepsilon$ is violated in the framework of Definition 4.12. Fix $\nu > 0$, suppose that $I = \{1, \ldots, N\}$ with $Nr > \nu$, pick some $u^* \in \widehat{N}(x_1; \Omega_1)$ with the norm $\|u^*\| = \nu$, and define the dual elements $x_1^* := u^* - \frac{u^*}{N} \in \widehat{N}(x_1; \Omega_1) + r\mathbb{B}^*$ and $x_i^* := 0 - \frac{u^*}{N} \in \widehat{N}(x_i; \Omega_i) + r\mathbb{B}^*$ for all $i = 2, \ldots, N$.

Then we have the relationships

$$x_1^*+\ldots+x_N^*=0 \ \text{ and } \ \|x_1^*\|^2+\ldots+\|x_N^*\|^2>\frac{\nu^2}{4},$$

which imply the triviality of the rated extremal principle by rescaling.

Now we are ready to derive the main result of this section, which justifies the validity of the rated extremal principle for rated extremal points of infinite systems of closed sets in Asplund spaces.

Theorem 4.15 (Rated extremal principle for infinite systems). *Let $\{\Omega_i\}_{i\in T}$ be a system of closed sets in an Asplund space X, and let $\bar{x}$ be a rated extremal point of this system. Then the rated extremality conditions of Definition* 4.12 *are satisfied for $\{\Omega_i\}_{i\in T}$ at $\bar{x}$.*

Proof. Given $\varepsilon>0$, take $r=\sup_i\|a_i\|$ sufficiently small and pick the corresponding index subset $I=\{1,\ldots,N\}$ with $N^{3/2}=o(R)$ from Definition 4.10. Consider the product space X^N with the norm of $z=(x_1,\ldots,x_N)\in X^N$ given by $\|z\|:=(\|x_1\|^2+\ldots+\|x_N\|^2)^{\frac{1}{2}}$ and define a function $\varphi\colon X^N\to\mathbb{R}$ by

$$\varphi(z):=\left(\sum_{i=2}^N\|(x_1-a_1)-(x_i-a_i)\|^2\right)^{\frac{1}{2}}. \tag{4.16}$$

To proceed, denote $\bar{z}:=(\bar{x},\bar{x},\ldots,\bar{x})\in\Omega_1\times\ldots\times\Omega_N$ and form the set

$$W:=\Big(\Omega_1\times\ldots\times\Omega_N\Big)\cap\Big(B\big(\bar{x},(R-1)r\big)\times\ldots\times B\big(\bar{x},(R-1)r\big)\Big), \tag{4.17}$$

which is nonempty and closed. We conclude that $\varphi(z)>0$ for all $z\in W$. Indeed, suppose on the contrary that $\varphi(z)=0$ for some $z=(x_1,\ldots,x_N)\in W$ and get by the estimates

$\|x_1 - a_1 - \bar{x}\| \le \|x_1 - \bar{x}\| + \|a_1\| \le (R-1)r + r = Rr$ the relationships

$$x_1 - a_1 = \ldots = x_N - a_N \in \bigcap_{i=1}^{N} (\Omega_i - a_i) \cap B(\bar{x}, Rr) \neq \emptyset,$$

which contradict the extremality condition (4.14). Observe further that

$$\varphi(\bar{z}) = \Big(\sum_{i=2}^{N} \|a_1 - a_i\|^2 \Big)^{\frac{1}{2}} < 2r\sqrt{N} \le \inf_{z \in W} \varphi(z) + 2rN^{\frac{1}{2}}.$$

Now we apply Ekeland's variational principle (see, e.g., [47, Theorem 2.26]) with the parameters $\varepsilon := 2rN^{\frac{1}{2}}$ and $\lambda := rR^{\frac{1}{2}}N^{\frac{3}{4}}$ to the lower semicontinuous and bounded from below function $\varphi(z)+\delta(z;W)$ on X^N and find in this way $z_0 \in W$ such that $\|z_0 - \bar{z}\| \le \lambda$ and that z_0 minimizes the perturbed function

$$\varphi(z) + \beta\|z - z_0\| + \delta(z;W) \;\text{ on }\; z \in X^N \;\text{ with }\; \beta := \frac{\varepsilon}{\lambda} = \frac{2}{R^{\frac{1}{2}}N^{\frac{1}{4}}}. \tag{4.18}$$

By the imposed growth condition $N^{\frac{3}{2}} = o(R)$ as $r \downarrow 0$ we have

$$\varepsilon = 2rN^{\frac{1}{2}} = r \cdot o(R^{\frac{1}{3}}) \le r \cdot o\Big(\frac{1}{r}\Big)^{\frac{1}{3}} \le r \cdot o\Big(\frac{1}{r}\Big) \to 0,$$

and similarly,

$$\begin{aligned} \frac{\lambda}{Rr} &= \frac{rR^{\frac{1}{2}}N^{\frac{3}{4}}}{Rr} = \frac{N^{\frac{3}{4}}}{R^{\frac{1}{2}}} \to 0, \\ N\beta &= \frac{2N}{R^{\frac{1}{2}}N^{\frac{1}{4}}} = \frac{2N^{\frac{3}{4}}}{R^{\frac{1}{2}}} = 2\Big(\frac{N^{\frac{3}{2}}}{R}\Big)^{\frac{1}{2}} \to 0 \;\text{ as }\; r \downarrow 0. \end{aligned}$$

Thus $\lambda = o(Rr)$ and $\beta \downarrow 0$ as $r \downarrow 0$ for the quantity β defined in (4.18). Taking into account that the function $\varphi(\cdot) + \beta\|\cdot - z_0\|$ is obviously Lipschitz continuous around $\bar{z}$, we apply to this sum the subdifferential fuzzy sum rule from [47, Lemma 2.32]. This allows us to find, for any given number $\eta > 0$, elements $z_1 = (y_1, \ldots, y_N) \in z_0 + \eta I\!B$ and $z_2 = (x_1, \ldots, x_N) \in z_0 + \eta I\!B$ such that

$$\big|\varphi(z_1) + \beta\|z_1 - z_0\| - \varphi(z_0)\big| \le \eta,\ z_2 \in W, \text{ and} \tag{4.19}$$

$$0 \in \widehat{\partial}\Big(\varphi(\cdot) + \beta\|\cdot - z_0\|\Big)(z_1) + \widehat{N}(z_2; W) + \eta I\!B^*. \tag{4.20}$$

Our next step is to explore formula (4.20). Since $\varphi(z_0) > 0$, we choose

$$\eta \le \min\Big\{\beta, \lambda, \frac{\varphi(z_0)}{2(1+\beta)}\Big\}.$$

Then it follows from (4.19) that

$$|\varphi(z_1) - \varphi(z_0)| \le (1+\beta)\eta \le (1+\beta)\frac{\varphi(z_0)}{2(1+\beta)} = \frac{\varphi(z_0)}{2},$$

which implies that $\varphi(z_1) =: \alpha > 0$. It is easy to see that the function $\varphi(\cdot)$ in (4.16) is convex. Applying the Moreau-Rockafellar theorem of convex analysis gives us

$$\widehat{\partial}\Big(\varphi(\cdot) + \beta\|\cdot - z_0\|\Big)(z_1) = \widehat{\partial}\varphi(z_1) + \beta I\!B^*, \tag{4.21}$$

where the Fréchet subdifferentials on both sides of (4.21) reduce to the classical subdifferential of convex functions. By the structure of φ in (4.16) and that of z_1 we have

$$\varphi(z_1) = \Big(\sum_{i=2}^{N} \|(y_1 - a_1) - (y_i - a_i)\|^2\Big)^{\frac{1}{2}}.$$

Denote further $\xi_i := y_1 - a_1 - y_i + a_i$ for $i = 2, \ldots, N$ and observe that $\alpha = \varphi(z_1) = \Big(\sum_{i=2}^{N} \|\xi_i\|^2\Big)^{\frac{1}{2}}$. Since the square root function is smooth at nonzero point, we apply the chain rule of convex analysis to derive that any element $(y_1^*,\ldots,y_N^*)\in\widehat{\partial}\varphi(z_1)$ has the representation

$$y_i^*=\begin{cases}-\dfrac{u_i^*}{\alpha}\cdot\|\xi_i\| & \text{if } \xi_i\neq 0,\\ 0 & \text{if } \xi_i=0,\end{cases}\qquad i=2,\ldots,N,$$

and $y_1^*=-y_2^*-y_3^*-\ldots-y_N^*$, where $u_i^*\in\widehat{\partial}\|\cdot\|(\xi_i)$ is a subgradient of the norm function calculated at the nonzero point ξ_i; hence $\|u_i^*\|=1$. This yields that

$$\|y_2^*\|^2+\ldots+\|y_N^*\|^2=1\ \text{ and }\ \|y_1^*\|^2+\ldots+\|y_N^*\|^2\geq 1.$$

On the other hand, we have the estimates

$$\|z_2-\bar z\|\leq|z_2-z_0\|+\|z_0-\bar z\|\leq\eta+\lambda\leq 2\lambda=o(Rr)$$

for $z_2=(x_1,\ldots,x_N)$ and hence $\|x_i-\bar x\|<\|z_2-\bar z\|=o(Rr)$ for $i=1,\ldots,N$. The latter ensures that each component x_i lies in the interior of the ball $B(\bar x,(R-1)r)$. Furthermore, it follows from the structure of W in (4.17) and the product formula for Fréchet normals that

$$\widehat{N}(z_2;W)=\widehat{N}\big(z_2;\Omega_1\times\ldots\times\Omega_N\big)=\widehat{N}(x_1;\Omega_1)\times\ldots\times\widehat{N}(x_N;\Omega_N),$$

which implies by combining with (4.20) and (4.21) the existence of $(y_1^*,\ldots,y_N^*)\in\widehat{\partial}\varphi(z_1)$ satisfying

$$y_1^*+\ldots+y_N^*=0,\ \text{ and }\ \|y_1^*\|^2+\ldots+\|y_N^*\|^2>1,$$

with $0\in y_i^*+\widehat{N}(x_i;\Omega_i)+2\beta I\!B^*$, $\|x_i-\bar x\|<2\lambda\to 0$ as $r\downarrow 0$.

Finally, replace y_i^* by $-y_i^*$ and get from the above that

$$y_i^* \in \widehat{N}(x_i;\Omega_i) + 2\beta I\!B^*, \quad \|x_i - \bar{x}\| < 2\lambda \to 0,$$

$$\text{for} \;\; i = 1,\ldots,N, \;\; N\beta \to 0 \;\; \text{as} \;\; r \downarrow 0,$$

$$y_1^* + \ldots + y_N^* = 0, \quad \text{and} \quad \|y_1^*\|^2 + \ldots + \|y_N^*\|^2 \geq 1,$$

which gives all the relationships of the rated extremal principle and completes the proof of the theorem. □

From the proof above we can distill some quantitative estimates for the elements involved in the relationships of the rated extremal principle.

Remark 4.16 (Quantitative estimates in the rated extremal principle). The proof of Theorem 4.15 essentially uses the growth assumptions $N^{3/2} = o(R)$ and $R \leq \frac{M}{r}$ on rated extremal points. Observe in fact that the given proof allows us to make the following *quantitative conclusions*: For any $\varepsilon > 0$ there exist a number $r \in (0,\varepsilon)$, an index subset $I = \{j_1,\ldots,j_N\}$ with $N^{3/2} = o(R(r))$, and elements

$$y_i^* \in \widehat{N}(x_i;\Omega_i) \;\; \text{with} \;\; \|x_i - \bar{x}\| \leq 2rR^{\frac{1}{2}}N^{\frac{3}{4}} \;\; \text{for all} \;\; i \in I$$

satisfying the relationships

$$\|y_{j_1}^* + \ldots + y_{j_N}^*\| \leq 2N\beta = \frac{4N^{\frac{3}{4}}}{R^{\frac{1}{2}}} \;\; \text{and} \;\; \|y_{j_1}^*\|^2 + \ldots + \|y_{j_N}^*\|^2 \geq 1.$$

Similar but somewhat different quantitative statement can be also made: For any rated extremal point $\bar{x}$ of the system $\{\Omega_i\}_{i\in T}$ with a rate function $R(r) = O(r)$ there is a constant $C > 0$ such that whenever $\varepsilon > 0$ there exist a number $r \in (0,\varepsilon)$, an index subset $I = \{j_1,\ldots,j_N\}$ with

$N^{3/2} = o(\frac{1}{r})$, and elements

$$y_i^* \in \widehat{N}(x_i; \Omega_i) \ \text{ with } \ \|x_i - \bar{x}\| \le C\sqrt{rN^{\frac{3}{2}}} \ \text{ for all } \ i \in I$$

satisfying the estimates

$$\|y_{j_1}^* + \ldots + y_{j_N}^*\| \le C\sqrt{rN^{\frac{3}{2}}} \ \text{ and } \ \|y_{j_1}^*\|^2 + \ldots + \|y_{j_N}^*\|^2 \ge 1.$$

In the last part of this section we introduce and study a certain notion of *perturbed extremality* for arbitrary (finite or infinite) set systems and compare it, in particular, with the notion of linear subextremality known for systems of two sets. Given two sets $\Omega_1, \Omega_2 \subset X$, the number

$$\vartheta(\Omega_1, \Omega_2) := \sup\left\{\nu \ge 0 \middle|\, \nu I\!B \subset \Omega_1 - \Omega_2\right\}$$

is known as the *measure of overlapping* for these sets [34]. We say that the system $\{\Omega_1, \Omega_2\}$ is *linear subextremal* [48, Subsection 5.4.1] around $\bar{x}$ if

$$\vartheta_{lin}(\Omega_1, \Omega_2, \bar{x}) := \liminf_{\substack{x_1 \xrightarrow{\Omega_1} \bar{x}, x_2 \xrightarrow{\Omega_2} \bar{x} \\ r \downarrow 0}} \frac{\vartheta\Big([\Omega_1 - x_1] \cap rI\!B, [\Omega_2 - x_2] \cap rI\!B\Big)}{r} = 0, \tag{4.22}$$

which is called "weak stationarity" in [34]; see [34, 48] for more discussions and references. It is proved in [34] and [48, Theorem 5.88] that the linear subextremality of a closed set system $\{\Omega_1, \Omega_2\}$ around $\bar{x}$ is equivalent, in the Asplund space setting, to the validity of the approximate extremal principle for $\{\Omega_1, \Omega_2\}$ at $\bar{x}$.

Our goal in what follows is to define a perturbed version of rated extremality, which is applied to infinite set systems while extends linear subextremality for systems of two sets as well. Given an R-rated extremal system of sets $\{\Omega_i\}_{i\in T}$ from Definition 4.10, we get that for any $\varepsilon>0$ there are $r=\sup\|a_i\|$, $R=R(r)$, and $I\subset T$ satisfying

$$\bigcap_{i\in I}\big(\Omega_i-\bar{x}-a_i\big)\cap(rR)I\!B=\emptyset. \tag{4.23}$$

Let us now perturb (4.23) by replacing $\bar{x}$ with some $x_i\in\Omega_i\cap B_\varepsilon(\bar{x})$ and arrive at the following construction.

Definition 4.17 (Perturbed extremal systems). *Let $\{\Omega_i\}_{i\in T}$ be a system of nonempty sets in X, and let $\bar{x}\in\bigcap_{i\in T}\Omega_i$. We say that $\bar{x}$ is R-PERTURBED EXTREMAL POINT of $\{\Omega_i,i\in T\}$ if for any $\varepsilon>0$ there exist $r=\sup_{i\in I}\|a_i\|<\varepsilon$, $I\subset T$ with $|I|^{3/2}=o(R)$, and $x_i\in\Omega_i\cap B_\varepsilon(\bar{x})$ as $i\in I$ such that*

$$\bigcap_{i\in I}\big(\Omega_i-x_i-a_i\big)\cap(rR)I\!B=\emptyset. \tag{4.24}$$

In this case we say that $\{\Omega_i\}_{i\in T}$ is an R-PERTURBED EXTREMAL SYSTEM at $\bar{x}$.

The next proposition establishes a connection between linear subextremality and perturbed extremality for systems of two sets $\{\Omega_1,\Omega_2\}$.

Proposition 4.18 (Perturbed extremality from linear subextremality). *Let a set system $\{\Omega_1,\Omega_2,\bar{x}\}$ be linearly subextremal around $\bar{x}$. Then it is an R-perturbed extremal system at this point.*

Proof. Employing the definition of linear subextremality, for any $\varepsilon>0$ sufficiently small we find $x_i\in\Omega_i\cap B_\varepsilon(\bar{x})$ and $r'<\varepsilon$ such that

$$\vartheta\big([\Omega_1-x_1]\cap r'I\!B,[\Omega_2-x_2]\cap r'I\!B\big)<r'\varepsilon.$$

This implies the existence of a vector $a \in X$ satisfying $\|a\| \leq r'\varepsilon$ and

$$a \notin \Big([\Omega_1 - x_1] \cap r'I\!B\Big) - \Big([\Omega_2 - x_2] \cap r'I\!B\Big),$$

which ensures in turn that

$$\Big([\Omega_1 - x_1] \cap r'I\!B - \frac{a}{2}\Big) \cap \Big([\Omega_2 - x_2] \cap r'I\!B + \frac{a}{2}\Big) = \emptyset. \tag{4.25}$$

Let us show that the latter implies the fulfillment of

$$\Big[\Omega_1 - x_1 - \frac{a}{2}\Big] \cap \Big[\Omega_2 - x_2 + \frac{a}{2}\Big] \cap \frac{r'}{2}I\!B = \emptyset. \tag{4.26}$$

Indeed, suppose that (4.26) does not hold and pick $\xi \in X$ from the left-hand side set in (4.26). Since $\xi + \frac{a}{2} \in \Omega_1 - x_1$ and $\|\xi\| \leq \frac{r'}{2}$, we have

$$\Big\|\xi_{\frac{a}{2}}\Big\| \leq \frac{r'}{2} + \frac{r'\varepsilon}{2} \leq \frac{r'}{2} + \frac{r'}{2} = r'$$

and consequently $\xi \in [\Omega_1 - x_1] \cap r'I\!B - \dfrac{a}{2}$. Similarly we get $\xi \in [\Omega_2 - x_2] \cap r'I\!B - \dfrac{a}{2}$. This clearly contradicts (4.25) and thus justifies the claimed relationship (4.26).

By setting $r := \dfrac{\|a\|}{2}$, out remaining task is to construct a continuous function $: \mathbb{R}_+ \to \mathbb{R}_+$ such that $R(r) \to \infty$ as $r \downarrow 0$ and that for each $\varepsilon > 0$ there is $r < \varepsilon$ satisfying

$$\Big[\Omega_1 - x_1 - \frac{a}{2}\Big] \cap \Big[\Omega_2 - x_2 + \frac{a}{2}\Big] \cap (rR)I\!B = \emptyset.$$

We first construct such a function along a sequence $r_k \downarrow 0$ as $k \to \infty$. Picking $\varepsilon_k \downarrow 0$, find

$r'_k < \varepsilon_k$ and select $a_k \in X$ with $\|a_k\| \le r'_k \varepsilon_k$ such that the sequence of $\|a_k\|$ is decreasing. Then define $r_k := \dfrac{\|a_k\|}{2}$ and $R(\varepsilon_k) := \dfrac{1}{\varepsilon_k}$. It follows from the constructions above that

$$r_k R(r_k) \le r'_k \varepsilon_k \frac{1}{\varepsilon_k} = r'_k, \quad k \in \mathbb{N}.$$

We clearly see that the sequence $\{R(r_k)\}$ is increasing as $r_k \downarrow 0$. Extending $R(\cdot)$ piecewise linearly to $\mathbb{R}_+$ brings us to the framework of Definition 4.17 and thus completes the proof of the proposition. □

Finally in this section, we show the rated extremality conditions of Definition 4.12 holds for R-perturbed extremal points of infinite set systems from Definition 4.17.

Theorem 4.19 (Rated Extremal Principle for Perturbed Systems). *Let $\bar{x}$ be an R-perturbed extremal point of a closed set system $\{\Omega_i\}_{i\in T}$ in an Asplund space X. Then the rated extremal principle holds for this system at $\bar{x}$.*

Proof. Fix $\varepsilon > 0$ and find I, $\{x_i\}_{i\in I}$, and $\{a_i\}_{i\in I}$ from Definition 4.17 such that

$$\bigcap_{i\in I} \big(\Omega_i - x_i - a_i\big) \cap (rR)I\!B = \emptyset.$$

For convenience denote $I := \{1, \ldots, N\}$ and define

$$\Omega := \Big\{(u_1, \ldots, u_N) \in X^N \;\Big|\; u_i \in \Omega_i \cap (x_i + rRI\!B),\ i \in I\Big\}.$$

For any $z = (u_1, \ldots, u_N) \in \Omega$ consider the function

$$\varphi(z) := \Big(\sum_{i=2}^{N} \|(u_1 - x_1 - a_1) - (u_i - x_i - a_i)\|^2\Big)^{\frac{1}{2}} > 0.$$

Furthermore, for $\bar{z} = (x_1, \ldots, x_N)$ we have the estimates

$$\varphi(\bar{z}) = \Big(\sum_{i=2}^{N} \|a_1 - a_i\|^2\Big)^{\frac{1}{2}} < 2r\sqrt{N} \le \inf_{z\in\Omega} \varphi(z) + 2rN^{\frac{1}{2}}.$$

The rest of the proof follows the arguments in the proof of Theorem 4.15. □

4.3 Calculus Rules for Rated Normals to Infinite Intersections

In the last section of this chapter we apply the rated extremal principle of Section 4.2 to deriving some calculus rules for general normals to infinite set intersections, which are closely related to necessary optimality conditions in problems of semi-infinite and infinite programming. Unless otherwise stated, the spaces below are Asplund and the sets under consideration are closed around reference points. As in Section 4.2, we often drop the subscript "r" for simplicity in the notation of rate functions $R_r = R(r)$ if no confusion arises. In addition, we always assume that rate functions are continuous.

We start with the following definition of *rated normals* to set intersections.

Definition 4.20 (Rated normals to set intersection). *Let $\Omega := \bigcap_{i\in T} \Omega_i$, and let $\bar{x} \in \Omega$. We say that a dual element $x^* \in X^*$ is an R-NORMAL to the set intersection Ω if for any $r \downarrow 0$ there is $I = I(r) \subset T$ of cardinality $|I|^{3/2} = o(R_r)$ such that*

$$\langle x^*, x - \bar{x}\rangle - r\|x - \bar{x}\| < r \quad \text{for all} \quad x \in \bigcap_{i\in I} \Omega_i \cap B(\bar{x}, rR_r). \tag{4.27}$$

The next proposition reveals relationships between Fréchet and R-normals to set intersections.

Proposition 4.21 (Rated normals versus Fréchet normals to set intersections). *Let $\bar{x} \in \Omega = \bigcap_{i\in I} \Omega_i$. Then any R-normal to Ω at $\bar{x}$ is a Fréchet normal to Ω at $\bar{x}$. The converse* *holds if I is finite.*

Proof. Assume x^* is an R-normal to Ω at $\bar{x}$ with some rate function $R(r)$ while x^* is not a Fréchet normal to Ω at this point. Hence there are $\delta > 0$ and a sequence $x_k \xrightarrow{\Omega} \bar{x}$ such that $\delta\|x_k - \bar{x}\| < \langle x^*, x_k - \bar{x}\rangle$ for all $k \in \mathbb{N}$. Hence $x_k \neq \bar{x}$ and

$$\delta\|x_k - \bar{x}\| < \langle x^*, x_k - \bar{x}\rangle < r\|x_k - \bar{x}\| + r$$

whenever $\|x_k - \bar{x}\| \leq rR$. Now suppose that $rR = M > 0$ for some M and then fix a number $k \in \mathbb{N}$ such that $\|x_k - \bar{x}\| \leq rR$. Letting $r \downarrow 0$, we arrive at the contradiction $\delta\|x_k - \bar{x}\| \leq 0$.

Consider next the remaining case when $rR \to 0$ as $r \downarrow 0$ and find $r_k > 0$ sufficiently small so that $\|x_k - \bar{x}\| = r_k R(r_k)$ due to the continuity of R and the convergence $rR \xrightarrow{r\downarrow 0} 0$. It follows that

$$\delta r_k R(r_k) < r_k^2 R(r_k) + r_k \ \text{ and hence } \ \delta < r_k + \frac{1}{R(r_k)}, \quad k \in \mathbb{N},$$

which gives a contradiction as $k \to \infty$. Thus x^* is a Fréchet normal to Ω at $\bar{x}$.

Conversely, assume that the index set I is finite, i.e., $I = \{1, \ldots, N\}$, and that x^* is a Fréchet normal. Then for any $r > 0$ we have by (2.3) that

$$\langle x^*, x - \bar{x}\rangle - r\|x - \bar{x}\| \leq 0 \ \text{ for all } \ x \in \bigcap_{i=1}^{N} \Omega_i \cap U,$$

where U is a neighborhood of $\bar{x}$. This clearly implies (4.27) with any rate function R, which ensures that x^* is an R-normal to Ω at $\bar{x}$ and thus completes the proof of the proposition. $\square$

The next example concerns infinite systems of convex sets in $\mathbb{R}^2$. It illustrates the way of computing R-normals to infinite intersections and shows that R-normals in this case reduce to usual ones.

Example 4.22 (Rated normals for infinite systems). Let $m \geq 4$ be a fixed integer. Consider an infinite system of convex sets $\{\Omega_k\}_{k\in\mathbb{N}}$ in $\mathbb{R}^2$ defined as the epigraphs of the convex and smooth functions

$$g_k(x) := \begin{cases} k^m x^2 & \text{for } x \geq 0, \\ 0 & \text{for } x < 0, \end{cases} \qquad k = 1, 2, \ldots .$$

Let $\bar{x} := (0,0)$, $\Omega := \bigcap_{k=1}^{\infty} \Omega_k$, and let $R = R(r) = r^{\alpha-1}$ for some $\alpha \in (0, \frac{2}{11})$. We obviously get $\Omega = \mathbb{R}_- \times \mathbb{R}_+$ and $N(\bar{x};\Omega) = \mathbb{R}_+ \times \mathbb{R}_-$. Let us verify that $x^* = (1,0)$ is an R-normal to Ω at $\bar{x}$, which implies the whole normal cone $N(\bar{x};\Omega)$ consists of R-normals.

To proceed, fix any $r > 0$ sufficiently small and denote by k_0 the *smallest* integer such that

$$\max\left\{\frac{1}{4r^2}, \frac{1}{4r^{2+\alpha}}\right\} = \frac{1}{4r^{2+\alpha}} \leq k_0^m.$$

Now consider $I := \{1, \ldots, k_0\}$ and check that

$$k_0 \leq \left(\frac{1}{4r^{2+\alpha}}\right)^{1/m} + 1 < \frac{1}{r^{\frac{2+\alpha}{m}}}.$$

Since $1 - \frac{3}{2m}(2+\alpha) - \alpha \geq 1 - \frac{3}{8}(2+\alpha) - \alpha \geq \frac{1}{4} - \frac{11}{8}\alpha > 0$, it follows that

$$\frac{|I|^{3/2}}{R} < \frac{r^{1-\alpha}}{r^{\frac{3(2+\alpha)}{2m}}} = r^{1-\frac{3}{2m}(2+\alpha)-\alpha} \to 0 \quad \text{when} \quad r \downarrow 0.$$

Defining further $\Omega_0 := \bigcap_{k=1}^{k_0} \Omega_k$, it remains to show that

$$\langle x^*, x\rangle - r\|x\| < r \quad \text{for all} \quad x \in \Omega_0 \cap B(0; rR). \tag{4.28}$$

To verify (4.28), take $x := (t, s)$ and consider only the case when $t > 0$, since the other case of $t \le 0$ is obvious. For $t > 0$ we have $s \ge k_0^m t^2$ and

$$\langle x^*, x\rangle - r\|x\| = t - r\sqrt{t^2+s^2} \le t\Big(1 - r\sqrt{1+k_0^{2m}t^2}\Big) < t\big(1 - rk_0^m t\big) = -rk_0^m t^2 + t =: f(t). \quad (4.29)$$

It follows from $\|x\| \le rR = r^\alpha$ that

$$r^\alpha \ge \sqrt{t^2+s^2} \ge t\sqrt{1+k_0^{2m}t^2} > k_0^m t^2$$

and hence $t < \big(\frac{r^\alpha}{K^m}\big)^{1/2}$. The latter implies that for all $x = (t, s) \in \Omega_0 \cap B(0; rR)$ with $t > 0$ we have

$$\langle x^*, x\rangle - r\|x\| < f(t) \le \sup_{[0,a]} f(t) \quad \text{with} \quad a := \Big(\frac{r^\alpha}{k_0^m}\Big)^{1/2} \ge \frac{1}{2rk_0^m}.$$

Observe finally that the function $f(t)$ in (4.29) attains its maximum on [0,a] at the point $t = \frac{1}{2rk_0^m}$ and that

$$\sup_{[0,a]} f(t) = -rk_0 \frac{1}{4r^2k_0^{2m}} + \frac{1}{2rk_0^m} = \frac{1}{4rk_0^m} \le r.$$

Combining all the above, we arrive at (4.28) and thus achieve our goals in this example.

The next example related to the previous one involves the notion of equicontinuity for systems of mappings. Given $f_i\colon X \to Y$, $i \in T$, we say that the system $\{f_i\}_{i\in T}$ is *equicontinuous* at $\bar{x}$ if for any $\varepsilon > 0$ there is $\delta > 0$ such that $\|f_i(x) - f_i(\bar{x})\| < \varepsilon$ for all $x \in B(\bar{x}, \delta)$ and $i \in T$. This notion has been recently exploited in [63] in the framework of variational analysis; see Remark 4.33.

Example 4.23 (Non-equicontinuity of gradient and normal systems). Given an integer

$m \geq 4$, define an infinite systems of functions $\varphi_k\colon \mathbb{R}^2 \to \mathbb{R}$ for $k \in \mathbb{N}$ by

$$\varphi_k(x_1,x_2) := \begin{cases} k^m x_1^2 - x_2 & \text{for } \ x_1 > 0, \\ \\ -x_2 & \text{for } \ x_1 \leq 0. \end{cases} \tag{4.30}$$

It is easy to check that the system of gradients $\{\nabla\varphi_k\}_{k\in\mathbb{N}}$ is not equicontinuous at $\bar{x} = (0,0)$.

Furthermore, observe that the sets Ω_k in Example 4.22 can be defined by

$$\Omega_k := \big\{x \in \mathbb{R}^2 \big|\ \varphi_k(x) \leq 0\big\}, \quad k \in \mathbb{N}. \tag{4.31}$$

Given any boundary point (x_1,x_2) of the set Ω_k, we compute the unit normal vector to Ω_k at (x_1,x_2) by

$$\xi_k(x_1,x_2) = \begin{cases} \dfrac{1}{\sqrt{4k^{2m}x_1^2+1}}(2k^m x_1, -1) & \text{for } \ x_1 > 0, \\ \\ (0,-1) & \text{for } \ x_1 \leq 0. \end{cases}$$

and then check the relationships for $x_1 > 0$:

$$\|\xi_k(x_1,x_2) - \xi_k(0,0)\|^2 = \frac{8k^{2m}x_1^2 - 2\sqrt{4k^{2m}x_1^2+1}}{4k^{2m}x_1^2+1} \to 2 \ \text{ as } \ k \to \infty.$$

The latter means that the system of $\{\xi_k\}_{k\in\mathbb{N}}$ is not equicontinuous at $\bar{x} = (0,0)$.

The next major result of this chapter establishes a certain "fuzzy" intersection rule for rated normals to infinite set intersections. Its proof is based on the rated extremal principle for infinite set systems obtained above in Theorem 4.15. Parts of this proof are similar to deriving a fuzzy sum rule for Fréchet normals to intersections of two sets in Asplund spaces given in [50] and in [47, Lemma 3.1] on the base of the approximate extremal principle for such set systems.

Theorem 4.24 (Fuzzy intersection rule for R-normals). *Let $\bar{x} \in \Omega := \bigcap_{i \in T} \Omega_i$, and let $x^* \in X^*$ be an R-normal to Ω at $\bar{x}$. Then for any $\varepsilon > 0$ there exist an index subset I, Fréchet normals $x_i^* \in \widehat{N}(x_i; \Omega_i)$ with $\|x_i - \bar{x}\| < \varepsilon$ for $i \in I$, and a number $\lambda \ge 0$ such that*

$$\lambda x^* \in \sum_{i \in I} x_i^* + \varepsilon I\!B^* \quad \text{and} \quad \lambda^2 + \lambda^2 \|x^*\|^2 + \sum_{i \in I} \|x_i^*\|^2 = 1. \tag{4.32}$$

Proof. Without loss of generality, assume that $\bar{x} = 0$. Pick any $x^* \in \widehat{N}(0; \Omega)$ and by Definition 4.20 for any $r > 0$ sufficiently small find an index subset $|I|^{3/2} = o(R)$ such that

$$\langle x^*, x \rangle - r\|x\| < r \quad \text{whenever} \quad x \in \bigcap_{i \in I} \Omega_i \cap (rR) I\!B. \tag{4.33}$$

Then we form the following closed subsets of the Asplund space $X \times \mathbb{R}$:

$$\begin{aligned} &O_1 := \Big\{ (x, \alpha) \in X \times \mathbb{R} \Big| \, x \in \Omega_1, \ \alpha \le \langle x^*, x \rangle - r\|x\| \Big\}, \\ &O_i := \Omega_i \times \mathbb{R}_+ \ \text{ for } \ i \in I \setminus \{1\}, \end{aligned} \tag{4.34}$$

where $I = \{1, \ldots, N\}$ with "1" denoting the first element of I for simplicity. This leads us to

$$\Big(O_1 - (0, r) \Big) \cap \bigcap_{i \in I \setminus \{1\}} O_i \cap (rR_r) I\!B = \emptyset. \tag{4.35}$$

Indeed, if on the contrary (4.35) does not hold, we get (x, α) from the above intersection satisfying $\alpha \ge 0$, $x \in \bigcap_{i \in I} \Omega_i \cap (\varepsilon R_\varepsilon) I\!B$, and

$$r \le \alpha + r \le \langle x^*, x \rangle - r\|x\|,$$

where the latter is due to $(x, \alpha + r) \in O_1$. This clearly contradicts (4.33) and so justifies (4.35).

Thus we have that $(0,0) \in X \times \mathbb{R}$ is a rated extremal point of the set system $\{O_1, O_2\}$ from (4.34) in the sense of Definition 4.10. Applying to this system the rated extremal principle from Theorem 4.15 with taking into account Remark 4.16 to find elements (w_i, α_i) and $(x_i^* \lambda_i)$ for $i = 1, \ldots, N$ satisfying the relationships

$$\begin{cases} (x_i^*, \lambda_i) \in \widehat{N}\big((w_i, \alpha_i); O_i\big), \ \|(w_i, \alpha_i)\| \le 2rR^{\frac{1}{2}}N^{\frac{3}{4}}, \quad i \in I, \\ \Big\|(x_1^*, \lambda_1) + \ldots + (x_N^*, \lambda_N)\Big\| \le \dfrac{4N^{\frac{3}{4}}}{R^{\frac{1}{2}}} =: \eta \downarrow 0 \ \text{ as } \ r \downarrow 0, \\ \|(x_1^*, \lambda_1)\|^2 + \ldots + \|(x_N^*, \lambda_N)\|^2 = 1. \end{cases} \tag{4.36}$$

By the structure of O_i as $i = 1, \ldots, N$ we have from the first line of (4.36) that $x_i^* \in \widehat{N}(w_i; \Omega_i)$, that $\lambda_i \le 0$ for $i = 2, \ldots, N$, and that

$$\limsup_{(x,\alpha) \overset{O_1}{\to} (w_1, \alpha_1)} \frac{\langle x_1^*, x - w_1 \rangle + \lambda_1(\alpha - \alpha_1)}{\|x - w_1\| + |\alpha - \alpha_1|} \le 0 \tag{4.37}$$

by the definition of Fréchet normals. It also follows from the structure of O_1 that $\lambda_1 \ge 0$ and

$$\alpha_1 \le \langle x^*, w_1 \rangle - r\|w_1\|. \tag{4.38}$$

This allows us to split the situation into the follows two cases.

Case 1: $\lambda_1 = 0$. If inequality (4.38) is strict in this case, there is a neighborhood W of w_1 such that $\alpha_1 \le \langle x^*, x \rangle - r\|x\|$ for all $x \in \Omega_1 \cap W$.

This implies that $(x, \alpha_1) \in O_1$ for $x \in \Omega_1 \cap W$. Substituting (x, α_1) into (4.37) gives us

$$\limsup_{x \overset{\Omega_1}{\to} w_1} \frac{\langle x_1^*, x - w_1 \rangle}{\|x - w_1\|} \le 0, \ \text{ i.e., } \ x_1^* \in \widehat{N}(w_1; \Omega_1).$$

If (4.38) holds as equality, we denote $\alpha := \langle x^*, x\rangle - r\|x\|$ and get

$$|\alpha - \alpha_1| = \Big|\langle x^*, x - w_1\rangle + r(\|w_1\| - \|x\|)\Big| \le \Big(\|x^*\| + r\Big)\|x - w_1\|,$$

which implies by (4.37) that

$$\limsup_{(x,\alpha)\overset{O_1}{\to}(w_1,\alpha_1)} \frac{\langle x_1^*, x - w_1\rangle}{\|x - w_1\| + |\alpha - \alpha_1|} \le 0.$$

Thus it follows for any $\varepsilon' > 0$ sufficiently small and the number α chosen above that

$$\langle x_1^*, x - w_1\rangle \le \varepsilon'\Big(\|x - w_1\| + |\alpha - \alpha_1|\Big) \le \varepsilon'\Big(1 + \|x^*\| + r\Big)\|x - w_1\|$$

for all $x \in \Omega_1$ sufficiently closed to w_1. This ensures that

$$\limsup_{x\overset{\Lambda_1}{\to}w_1} \frac{\langle x_1^*, x - w_1\rangle}{\|x - w_1\|} \le 0, \quad \text{i.e.,} \quad x_1^* \in \widehat{N}(w_1; \Omega_1)$$

when (4.38) holds as equality as well as the strict inequality. Since $\lambda_1 = 0$ in Case 1 under consideration and since $\lambda_i \le 0$ for all $i \ge 2$, it follows that

$$\lambda_2^2 + \ldots + \lambda_N^2 \le (\lambda_2 + \ldots + \lambda_N)^2 \le \eta^2.$$

This leads us to the estimates

$$\|x_1^*\|^2 + \ldots + \|x_N^*\|^2 \ge 1 - (\lambda_2^2 + \ldots + \lambda_N^2) \ge \frac{1}{2},$$

and thus we get from (4.36) all the conclusion of the theorem with $\lambda = 0$ in (4.32) in this case.

Case 2: $\lambda_1 > 0$. If inequality (4.38) is strict in this case, put $x := w_1$ and get from (4.37) that

$$\limsup_{\alpha \to \alpha_1} \frac{\lambda_1(\alpha - \alpha_1)}{|\alpha - \alpha_1|} \leq 0,$$

which yields $\lambda_1 = 0$, a contradiction. It remains therefore to consider the case when (4.38) holds as equality. Take then a pair $(x, \alpha) \in O_1$ with $x \in \Omega_1 \setminus \{w_1\}$ and $\alpha = \langle x^*, x\rangle - r\|x\|$, and hence get from (4.38) that: $\alpha - \alpha_1 = \langle x^*, x - w_1\rangle + r(\|w_1\| - \|x\|)$.

This implies the relationships

$$\langle x_1^*, x - w_1\rangle + \lambda_1(\alpha - \alpha_1) = \langle x_1^* + \lambda_1 x^*, x - w_1\rangle + \lambda_1 r(\|w_1\| - \|x\|),$$

$$|\alpha - \alpha_1| \leq (\|x^*\| + r)\|x - w_1\|.$$

On the other hand, it follows from (4.37) that for any $\varepsilon' > 0$ sufficiently small there exists a neighborhood V of w_1 such that

$$\langle x_1^*, x - w_1\rangle + \lambda_1(\alpha - \alpha_1) \leq \lambda_1 \varepsilon' r\Big(\|x - w_1\| + |\alpha - \alpha_1|\Big),$$

whenever $x \in \Omega_1 \cap V$ and that

$$\begin{aligned} &\langle x_1^* + \lambda_1 x^*, x - w_1\rangle + \lambda_1 r(\|w_1\| - \|x\|) \leq \lambda_1 \varepsilon' r(\|x - w_1\| + |\alpha - \alpha_1|) \\ &\leq \lambda_1 \varepsilon' r\Big[\|x - w_1\| + (\|x^*\| + r)\|x - w_1\|\Big] = \lambda_1 \varepsilon' r\big(1 + \|x^*\| + r\big)\|x - w_1\|. \end{aligned}$$

Let us now choose $\varepsilon' > 0$ sufficiently small so that

$$\langle x_1^* + \lambda_1 x^*, x - w_1\rangle + \lambda_1 r(\|w_1\| - \|x\|) \leq \lambda_1 r\|x - w_1\|.$$

and for all $x \in \Omega_1 \cap V$ get the estimate

$$\langle x_1^* + \lambda_1 x^*, x - w_1 \rangle \le \lambda_1 r \|x - w_1\| + \lambda_1 r(\|x\| - \|w_1\|) \le 2\lambda_1 r \|x - w_1\|.$$

It follows from definition (2.3) of ε-normals that $x_1^* + \lambda_1 x^* \in \widehat{N}_{2\lambda_1 r}(w_1; \Omega_1)$, where $\lambda_1 \le 1$ by the third line of (4.36). Using the representation of ε-normals in Asplund spaces from [47, (2.51)], we find $v \in \Omega_1 \cap (w_1 + 2\lambda_1 r)I\!B)$ such that

$$x_1^* + \lambda_1 x^* \in \widehat{N}(v; \Omega_1) + 2\lambda_1 r I\!B^*.$$

Hence $\|v\| \le \|v - w_1\| + \|w_1\| \le 2\lambda_1 r + 2rR^{\frac{1}{2}}N^{\frac{3}{4}} \le 3rR^{\frac{1}{2}}N^{\frac{3}{4}}$ and there is $\widetilde{x}_1^* \in \widehat{N}(v; \Omega_1)$ with

$$\lambda_1 x^* \in \widetilde{x}_1^* - x_1^* + 2\lambda_1 r I\!B^*.$$

Taking into account that $x_1^* + \ldots + x_N^* \in \eta I\!B^*$, we get

$$\lambda_1 x^* \in \widetilde{x}_1^* + x_2^* + \ldots + x_N^* + (2\lambda_1 r + \eta) I\!B^*.$$

On the other hand, it follows from $-x_1^* = \lambda_1 x^* - \widetilde{x}_1^* - u^*$ with some $\|u^*\| \le 2\lambda_1 r \le 2r$ that

$$\|x_1^*\|^2 \le \big(\lambda_1 \|x^*\| + \|\widetilde{x}_1^*\| + 2r\big)^2 \le 2\lambda_1^2 \|x^*\|^2 + 2\|\widetilde{x}_1^*\|^2 + \frac{1}{4}.$$

Moreover, since $|\lambda_1 + \lambda_2 + \ldots + \lambda_N| \le \eta \downarrow 0$ as $r \downarrow 0$ by the second line of (4.36) and since

$\lambda_1 \geq 0$ while $\lambda_i \leq 0$ for $i = 2, \ldots, N$, we have

$$\eta^2 > \lambda_1^2 + (\lambda_2 + \ldots + \lambda_N)^2 + 2\lambda_1(\lambda_2 + \ldots + \lambda_N) > \lambda_1^2 + (\lambda_2 + \ldots + \lambda_N)^2 + 2\lambda_1(-\lambda_1 - \eta)$$

It also follows from (4.36) and $0 < \lambda_1 < 1$ that

$$\lambda_1^2 \geq (\lambda_2 + \ldots \lambda_N)^2 - \eta^2 - 2\eta\lambda_1 \geq \lambda_2^2 + \ldots + \lambda_N^2 - \frac{1}{4},$$

which leads us to the subsequent estimates: $\lambda_1^2 + \ldots + \lambda_N^2 \leq 2\lambda_1^2 + \dfrac{1}{4}$ and

$$\begin{aligned} 1 &\leq \Big(\lambda_1^2 + \ldots + \lambda_N^2\Big) + \Big(\|x_1^*\|^2 + \ldots + \|x_N^*\|^2\Big) \\ &\leq 2\lambda_1^2 + 2\lambda_1^2\|x^*\|^2 + 2\|\tilde{x}_1^*\|^2 + \Big(\|x_2^*\|^2 + \ldots + \|x_N^*\|^2\Big) + \frac{1}{2}. \end{aligned}$$

This finally ensures that

$$\frac{1}{4} \leq \lambda_1^2 + \lambda_1^2\|x^*\|^2 + \|\widetilde{x}_1^*\|^2 + \|x_2^*\|^2 + \ldots + \|x_N^*\|^2$$

and brings us to all the conclusions of the theorem with $\lambda := \lambda_1$ in (4.32). □

Remark 4.25 (Quantitative estimates in the intersection rule). It can be observed directly from the proof of Theorem 4.24 that we get in fact the following quantitative estimates in intersection rule obtained for infinite set systems when $r > 0$ is sufficiently small: $|I|^{3/2} = o(R)$,

$$\|x_i - \bar{x}\| < 3rR^{\frac{1}{2}}|I|^{\frac{3}{4}}, \text{ and } \lambda x^* \in \sum_{i\in I} x_i^* + \Big(2r + 4\frac{|I|^{\frac{3}{4}}}{R^{\frac{1}{2}}}\Big)I\!B^*.$$

In particular, for $R = O\big(\frac{1}{r}\big)$, there is $C > 0$ such that all the conclusions hold with $|I|^{3/2} =$

$N^{3/2} = o\big(\frac{1}{r}\big)$,

$$\|x_i - \bar{x}\| < C\sqrt{rN^{\frac{3}{2}}}, \text{ and } \lambda x^* \in \sum_{i\in I} x_i^* + C\sqrt{rN^{\frac{3}{2}}}I\!B^*.$$

Remark 4.26 (Perturbed rated normals to infinite intersections). Inspired by our consideration of perturbed extremal systems, we define a perturbed version of R-normals to infinite set intersections as follows: $x^* \in X^*$ is a *perturbed R-normal* to the intersection $\Omega := \bigcap_{i\in T}\Omega_i$ at $\bar{x} \in \Omega$ if for any $\varepsilon > 0$ there exist a number $r > 0$, an index subset I with cardinality $|I|^{3/2} = o(R_r)$, and points $x_i \in \Omega_i \cap B(\bar{x}, \varepsilon)$ as $i \in I$ such that $r|I| < \varepsilon$ and

$$\langle x^*, x\rangle - r\|x\| < r \text{ whenever } x \in \bigcap_{i\in I}\big(\Omega_i - x_i\big) \cap (rR_r)I\!B.$$

Then the corresponding version of the intersection rule from Theorem 4.24 can be derived for perturbed rated normals to infinite intersections by a similar way with replacing in the proof the rated extremal principle from Theorem 4.15 by its perturbed version from Theorem 4.19.

We proceed with deriving calculus rules for the so-called *limiting R-normals* (defined below) to infinite intersections of sets. First we propose a new qualification conditions for infinite systems.

Definition 4.27 (Approximate qualification condition) *We say that a system of sets* $\{\Omega_i\}_{i\in T} \subset X$ *satisfies the* APPROXIMATE QUALIFICATION CONDITION (AQC) *at* $\bar{x} \in \bigcap_{i\in T}\Omega_i$ *if for any* $\varepsilon \downarrow 0$*, any finite index subset* $I_\varepsilon \subset T$*, and any Fréchet normals* $x_{i\varepsilon}^* \in \widehat{N}(x_{i\varepsilon};\Omega_i) \cap I\!B^*$ *with* $\|x_{i\varepsilon} - \bar{x}\| \le \varepsilon$ *as* $i \in I_\varepsilon$ *the following implication holds:*

$$\Big\|\sum_{i\in I_\varepsilon} x_{i\varepsilon}^*\Big\| \xrightarrow{\varepsilon\downarrow 0} 0 \Longrightarrow \sum_{i\in I_\varepsilon}\|x_{i\varepsilon}^*\|^2 \xrightarrow{\varepsilon\downarrow 0} 0. \tag{4.39}$$

The next proposition presents verifiable conditions ensuring the validity of AQC for finite systems of sets under the SNC property; see [47] for more details.

Proposition 4.28 (AQC for finite set systems under SNC assumptions) *Let* $\{\Omega_1,\ldots,\Omega_m\}$ *be a finite set system satisfying the limiting qualification condition at* $\bar{x}\in\bigcap_{i=1}^m\Omega_i$*: for any sequences* $x_{ik}\overset{\Omega_i}{\to}\bar{x}$ *and* $x_{ik}^*\overset{w^*}{\to}x_i^*$ *with* $x_{ik}^*\in\widehat{N}(x_{ik};\Omega_i)$ *as* $k\to\infty$ *and* $i=1,\ldots,m$ *we have*

$$\|x_{1k}^*+\ldots+x_{mk}^*\|\to 0\Longrightarrow x_1^*=\ldots=x_m^*=0,$$

which is automatic under the normal qualification condition via the basic normal cone (2.5)*:*

$$\big[x_1^*+\ldots+x_m^*=0\;\text{ and }\;x_i^*\in N(\bar{x};\Omega_i),\;i=1,\ldots,m\big]\Longrightarrow x_i^*=0\;\text{ for all }\;i=1,\ldots,m.$$

Assume in addition that all but one of Ω_i *are SNC at* $\bar{x}$*. Then the AQC is satisfied for* $\{\Omega_1,\ldots,\Omega_m\}$ *at* $\bar{x}$*.*

Proof. Pick $\varepsilon_k\downarrow 0$, $x_{ik}^*\in\widehat{N}(x_{ik};\Omega_i)\cap I\!B^*$, $\|x_{ik}-\bar{x}\|\le\varepsilon_k$ as $i=1,\ldots,m$ and assume that

$$\|x_{1k}^*+\ldots+x_{mk}^*\|\to 0\;\text{ as }\;k\to\infty. \tag{4.40}$$

Taking into account that the sequences $\{x_{ik}^*\}\subset X^*$ are bounded when X is Asplund, we extract from them weak* convergent subsequences and suppose with no relabeling that $x_{ik}^*\overset{w^*}{\to}x_i^*$ as $k\to\infty$ for all $i=1,\ldots,m$. It follows from the imposed limiting qualification condition for $\{\Omega_1,\ldots,\Omega_m\}$ at $\bar{x}$ that $x_1^*=\ldots=x_m^*=0$. Since all but one (say for $i=1$) of the sets Ω_i are SNC at $\bar{x}$, we have that $\|x_{ik}^*\|\to 0$ as $k\to\infty$ for $i=2,\ldots,m$. Then (4.40) implies that $\|x_{1k}^*\|\to 0$ as well, which verifies implication (4.39) and thus completes the proof of the proposition. □

The following example illustrates the validity of the AQC for infinite systems of sets.

Example 4.29 (AQC for infinite systems) We verify that the AQC holds in the framework of Example 4.23 at the origin $\bar{x} = (0,0) \in \mathbb{R}^2$. Recall that for each $k \in \mathbb{N}$ the normal cone to a convex set Ω_k from (4.31) at a boundary point $x = (x_1, x_2)$ is computed by

$$N(x;\Omega_k) = \mathbb{R}_+\xi_k(x) \ \text{ with } \ \xi_k(x) = \xi_k(x_1,x_2) = \begin{cases} (2k^m x_1, -1) & \text{for } x_1 > 0, \\ (0,-1) & \text{for } x_1 \le 0. \end{cases}$$

If according to the left-hand side of (4.39) we have

$$\Big\| \sum_{k \in I_\varepsilon} \lambda_{\varepsilon k} \xi_k(x_{\varepsilon k}) \Big\| \to 0 \ \text{ as } \ \varepsilon \downarrow 0,$$

then it follows from the above representation of ξ_k that its component goes to zero as $k \to \infty$. Thus

$$\sum_{k \in I_\varepsilon} \|\lambda_{\varepsilon k} \xi_k(x_{\varepsilon k})\|^2 \to 0 \ \text{ as } \ \varepsilon \downarrow 0,$$

which verifies the AQC property of the system $\{\Omega_k\}_{k \in \mathbb{N}}$ at $\bar{x}$.

Now we are ready to define limiting R-normals and derive infinite intersection rules for them. In the definition below R_k stands for a rate function for each x_k^*; these functions may be different from each other.

Definition 4.30 (Limiting R-normals to infinite set intersections) *Consider an arbitrary set system $\{\Omega_i\}_{i \in T} \subset X$, and let $\Omega := \bigcap_{i \in T} \Omega_i$ with $\bar{x} \in \Omega$. We say that a dual element x^* is a* LIMITING R-NORMAL *to Ω at $\bar{x}$ if there exist sequences $\{(x_k, x_k^*)\}_{k \in \mathbb{N}} \subset X \times X^*$ such that*

$x_k \xrightarrow{\Omega} \bar{x}$, $x_k^* \xrightarrow{w^*} x^*$ *as* $k \to \infty$ *and that each element* x_k^* *is an* R_k*-normal to* Ω *at* x_k*,*

It is clear from the definition and Proposition 4.21 that any limiting R-normal is a basic/limiting normal to Ω at $\bar{x}$. Conversely, if T is a finite index set and X is an Asplund space, then we the reverse implication holds, i.e., any limiting/basic normal is a limiting R-normal.

The next theorem provides a representation of limiting R-normals to infinite set intersections via Fréchet normals to each set under consideration. In particular, it implies a useful calculus rule for the basic normal cone (2.5) to infinite intersections.

Theorem 4.31 (Representation of limiting R-normals to infinite intersections) *Let* $\Omega := \bigcap_{i\in T}\Omega_i$ *with* $\bar{x} \in \Omega$ *for the system* $\{\Omega_i\}_{i\in T} \subset X$ *satisfying the AQC property from Definition* 4.27 *at* $\bar{x}$*. Then for any given limiting R-normal to* Ω *at* $\bar{x}$ *and any* $\varepsilon > 0$ *we have the inclusion*

$$x^* \in \operatorname{cl}^*\Big\{\sum_{i\in I} x_i^* + \varepsilon I\!B^* \;\Big|\; x_i^* \in \widehat{N}(x_i;\Omega_i),\ \|x_i - \bar{x}\| < \varepsilon,\ I \subset T\Big\},$$

where $I \subset T$ *is a finite index subset. In particular, if all the limiting/basic normals to* Ω *at* $\bar{x}$ *are limiting R-normals in this setting, then*

$$N(\bar{x};\Omega) \subset \bigcap_{\varepsilon>0}\operatorname{cl}^*\Big\{\sum_{i\in I} x_i^* + \varepsilon I\!B^* \;\Big|\; x_i^* \in \widehat{N}(x_i;\Omega_i),\ \|x_i - \bar{x}\| < \varepsilon,\ I \subset T\Big\}. \tag{4.41}$$

Proof. Take a sequence $\{x_k^*\}$ of R-normals to Ω at x_k with $x_k \to \bar{x}$ and $x_k^* \xrightarrow{w^*} x^*$ as $k \to \infty$. The latter convergence ensures by the Uniform Boundedness Principle that the set $\{\|x_k^*\|\}_{k\in\mathbb{N}}$ is bounded in X^*. Picking $\varepsilon > 0$ sufficiently small, we find $x_k \in \Omega$ with $\|x_k - \bar{x}\| < \varepsilon$. Applying Theorem 4.24 to x_k^* for each $k \in \mathbb{N}$ gives us sequences $x_{ik}^* \in \widehat{N}(x_{ik};\Omega_i)$ with $\|x_{ik} - x_k\| < \varepsilon$ for

$i \in I_k \subset T$ and $\lambda_k \geq 0$ satisfying

$$\lambda_k x_k^* \in \sum_{i \in I_k} x_{ik}^* + \varepsilon I\!B^* \ \text{ and } \ \lambda_k^2 + \lambda_k^2 \|x_k^*\|^2 + \sum_{i \in I_k} \|x_{ik}^*\|^2 = 1, \quad k \in \mathbb{N}. \tag{4.42}$$

Let us show that the sequence $\{\lambda_k\}$ is bounded away from 0. Assuming on the contrary $\lambda_k \downarrow 0$ as $k \to \infty$, we have

$$\Big\| \sum_{i \in I_k} x_{ik}^* \Big\| \longrightarrow 0 \text{ as } \ k \to \infty$$

from the inclusion in (4.42). Then the imposed AQC leads us to

$$\sum_{i \in I_k} \|x_{ik}^*\|^2 \to 0 \ \text{ as } \ k \to \infty,$$

which contradicts the equality in (4.42) and thus shows that there is constant $C > 0$ with $\lambda_k > C$ for all $k \in \mathbb{N}$ sufficiently large. Rescaling finally the inclusion in (4.42), we get

$$x_k^* \in \sum_{i \in I} \frac{x_{ik}^*}{\lambda_k} + \frac{\varepsilon}{C} I\!B^*, \quad k \in \mathbb{N},$$

which ensures that $x_k^* \xrightarrow{w^*} x^*$ as $k \to \infty$ and thus justifies the first conclusion of the theorem. The second ones on basic normals follows immediately. $\square$

The next corollary provides more explicit results for the case of infinite systems of cones, with the replacement of Fréchet normals in Theorem 4.31 by basic normals at the origin.

Corollary 4.32 (Limiting R-normals to intersection of cones). *Let $\{\Lambda_i\}_{i \in T}$ be a system of cones in X, and let $\Lambda := \bigcap_{i \in T} \Lambda_i$. Suppose that $x^* \in X^*$ is a limiting R-normal to Λ at the origin and that the AQC property from Definition* 4.27 *holds at $\bar{x} = 0$. Then for any $\varepsilon > 0$ we* *have the representation*

$$x^* \in \mathrm{cl}^*\Big\{\sum_{i\in I} x_i^* + \varepsilon I\!B^* \;\Big|\; x_i^* \in N(0;\Lambda_i),\ I \subset T\Big\}$$

via finite index subsets $I \subset T$. If furthermore all the limiting/basic normals to Λ at the original are limiting R-normals in this setting, then

$$N(0;\Lambda) \subset \bigcap_{\varepsilon>0} \mathrm{cl}^*\Big\{\sum_{i\in I} x_i^* + \varepsilon I\!B^* \;\Big|\; x_i^* \in N(0;\Lambda_i),\ I \subset T\Big\}.$$

Proof. It follows from Proposition 2.1 that $\widehat{N}(w_i;\Lambda_i) \subset N(0;\Lambda_i)$ for any cone Λ_i and any $w_i \in \Lambda_i$. Then we have both conclusions of the corollary from Theorem 4.31. □

Remark 4.33 (Comparison with known results). For the case of finite set systems the intersection rules of Theorems 4.24 and 4.31 go back to the well-known results of [47]. In fact, not much has been known for representations of generalized normals to infinite intersections. Chapter 3 presents our first results in this direction obtained on the base of the tangential extremal principle in finite dimensions, have a different nature and do not generally reduce to those in [47] for finite set systems.

An interesting representation of the basic normal cone (2.5) has been recently established in [63, Theorem 3.1] for infinite intersections of sets given by inequality constraints with smooth functions. This result essentially exploits specific features of the sets and functions under consideration and imposes certain assumptions, which are not required by our Theorem 4.31. In particular, [63, Theorem 3.1] requires the equicontinuity of the constraint functions involved, which is not the case of our Theorem 4.31 as shown in Examples 4.22 and 4.23. Note to this end that all the limiting normals are limiting R-normals in the framework of Example 4.22 and that the AQC assumption is satisfied therein; see Example 4.29.

We finish the paper with deriving necessary optimality conditions for problems of semi-infinite and infinite programming with geometric constraints given by

$$\text{minimize} \quad \varphi(x) \ \text{ subject to } x \in \Omega_i, \quad t \in T, \tag{4.43}$$

with a general cost function $\varphi\colon X \to \overline{\mathbb{R}}$ and constraints sets $\Omega_t \subset X$ indexed by an arbitrary (possibly infinite) set T. We refer the reader to [10, 24] and the bibliographies therein for various results, discussions, and examples concerning optimization problems of type (4.43) and their specifications. The limiting normal cone representation (4.41) for infinite set intersections in Theorem 4.31, combined with some basic principles in constrained optimization, leads us to necessary optimality conditions for local optimal solutions to (4.43) expressed via its initial data.

The next theorem contains results of this kind in both *lower subdifferential* and *upper subdifferential* forms; see Section 3.7.

Theorem 4.34 (Necessary optimality condition for semi-infinite and infinite programs with general geometric constraints). *Let $\bar{x}$ be a local optimal solution to problem* (4.43)*. Assume that any basic normal to $\Omega := \bigcap_{i\in T}\Omega_i$ at $\bar{x}$ is a limiting R-normal in this setting, and that the AQC requirements is satisfied for $\{\Omega_i\}_{i\in T}$ at $\bar{x}$. Then the following conditions, involving finite index subsets $I \subset T$, hold:*

(i) *For general cost functions φ finite at $\bar{x}$ we have*

$$-\widehat{\partial}\varphi(\bar{x}) \subset \bigcap_{\varepsilon>0} \mathrm{cl}^*\Big\{\sum_{i\in I} x_i^* + \varepsilon I\!B^* \ \Big|\ x_i^* \in \widehat{N}(x_i;\Omega_i),\ \|x_i-\bar{x}\| < \varepsilon,\ I \subset T\Big\}. \tag{4.44}$$

(ii) *If in addition φ is locally Lipschitzian around $\bar{x}$, then*

$$0 \in \partial\varphi(\bar{x}) + \bigcap_{\varepsilon>0} \operatorname{cl}^*\Big\{ \sum_{i\in I} x_i^* + \varepsilon I\!\!B^* \;\Big|\; x_i^* \in \widehat{N}(x_i;\Omega_i),\ \|x_i - \bar{x}\| < \varepsilon,\ I \subset T \Big\}. \tag{4.45}$$

Proof. It follows from [48, Proposition 5.2] that

$$-\widehat{\partial}\varphi(\bar{x}) \subset \widehat{N}(\bar{x};\Omega) \subset N(\bar{x};\Omega) \tag{4.46}$$

for the general constrained optimization problem

$$\text{minimize } \ \varphi(x) \ \text{ subject to } \ x \in \Omega. \tag{4.47}$$

Employing now in (4.46) the intersection formula (4.41) for basic normals to $\Omega = \bigcap_{i\in T} \Omega_i$, we arrive at the upper subdifferential necessary optimality condition (4.44) for problem (4.43).

To justify (4.45), we get from [48, Propostion 5.3] the lower subdifferential necessary optimality condition

$$0 \in \partial\varphi(\bar{x}) + N(\bar{x};\Omega) \tag{4.48}$$

for problem (4.47) provided that φ is locally Lipschitzian around $\bar{x}$. Using the intersection formula (4.41) in (4.48) completes the proof of the theorem. $\square$

Part B: Generalized Newton's methods

Chapter 5

Pure Newton's method

5.1 The Pure Newton's Algorithm

This section presents a new generalized Newton's method for nonsmooth equations, which is based on graphical derivatives. We first precisely describe the algorithm and justify its well-posedness/solvability. Then a local superlinear convergence result under appropriate assumptions will be presented. Finally, we establish a global convergence result of the Kantorovich type for our generalized Newton algorithm.

5.1.1 Description and Justification of the Algorithm

Keeping in mind the classical scheme of the smooth Newton's method in (1.4), (1.5) and taking into account the graphical derivative representation of Proposition 2.3(f), we propose an extension of the Newton equation (1.5) to nonsmooth mappings given by:

$$-H(x^k) \in DH(x^k)(d^k), \quad k = 0, 1, 2, \ldots. \tag{5.1}$$

This leads us to the following generalized Newton algorithm to solve (1.3):

Algorithm 5.1 (generalized Newton's method).

Step 0: *Choose a starting point* $x^0 \in \mathbb{R}^n$.

Step 1: *Check a suitable termination criterion.*

Step 2: *Compute* $d^k \in \mathbb{R}^n$ *such that* (5.1) *holds.*

Step 3: *Set* $x^{k+1} := x^k + d^k$*,* $k \leftarrow k+1$*, and go to Step* 1*.*

The proposed Algorithm 5.1 does not require a priori any assumptions on the underlying mapping $H\colon \mathbb{R}^n \to \mathbb{R}^n$ in (1.3) besides its continuity, which is the standing assumption in this paper. Other assumptions are imposed below to justify the well-posedness and (local and global) convergence of the algorithm. Observe that Proposition 2.3(c,d) ensures that Algorithm 5.1 reduces to scheme (1.6) in the B-differentiable Newton method provided that H is directionally differentiable and locally Lipschitzian around the solution point in question. In Section 5 we consider in detail relationships with known results for the B-differentiable Newton's method, while Section 4 compares Algorithm 5.1 and the assumptions made with the corresponding semismooth versions in the framework of (1.7).

To proceed further, we need to make sure that the generalized Newton equation (5.1) is *solvable*, which is a major part of the well-posedness of Algorithm 5.1. The next proposition shows that an appropriate assumption to ensure the solvability of (5.1) is *metric regularity.*

Proposition 5.2 (solvability of the generalized Newton equation). *Assume that $H\colon \mathbb{R}^n \to \mathbb{R}^n$ is metrically regular around $\bar{x}$ with $\bar{y} = H(\bar{x})$, i.e., we have $\ker D^*H(\bar{x}) = \{0\}$. Then there is a constant $\varepsilon > 0$ such that for all $x \in B_\varepsilon(\bar{x})$ the equation*

$$-H(x) \in DH(x)(d) \tag{5.2}$$

admits a solution $d \in \mathbb{R}^n$. Furthermore, the set $S(x)$ of solutions to (5.2) is computed by

$$S(x) = \operatorname*{Lim\,sup}_{t\downarrow 0,\, h\to -H(x)} \frac{H^{-1}\big(H(x)+th\big)-x}{t} \neq \emptyset. \tag{5.3}$$

Proof. By the assumed metric regularity (2.17) of H we find a number $\mu > 0$ and neighborhoods U of $\bar{x}$ and V of $H(\bar{x})$ such that

$$\text{dist}\big(x; H^{-1}(y)\big) \le \mu\,\text{dist}\big(y; H(x)\big) \quad \text{for all} \quad x \in U \quad \text{and} \quad y \in V.$$

Pick now an arbitrary vector $x \in U$ and select sequences $h_k \to -H(x)$ and $t_k \downarrow 0$ as $k \to \infty$. Suppose with no loss of generality that $H(x) + t_k h_k \in V$ for all $k \in \mathbb{N}$. Then we have

$$\text{dist}\big(x; H^{-1}(H(x) + t_k h_k)\big) \le \mu t_k \|h_k\|, \quad k \in \mathbb{N},$$

and hence there is a vector $u_k \in H^{-1}(H(x) + t_k h_k)$ such that $\|u_k - x\| \le \mu t_k \|h_k\|$ for all $k \in \mathbb{N}$. This shows that the sequence $\{\|u_k - x\|/t_k\}$ is bounded, and thus it contains a subsequence that converges to some element $d \in \mathbb{R}^n$. Passing to the limit as $k \to \infty$ and recalling the definitions of the outer limit (2.1) and of the tangent cone (2.2), we arrive at

$$\big(d, -H(x)\big) \in \operatorname*{Lim\,sup}_{t \downarrow 0} \frac{\text{gph}\, H - \big(x, H(x)\big)}{t} = T\big((x, H(x)); \text{gph}\, H\big),$$

which justifies the desired inclusion (5.2). The solution representation (5.3) follows from (2.12) and Proposition 2.3(b) in the case of single-valued mappings, since

$$S(x) = DH(x)^{-1}\big(-H(x)\big)$$

due to (5.2). This completes the proof of the proposition. □

5.1.2 Local Convergence

In this subsection we first formulate major assumptions of our generalized Newton's method and then show that they ensure the superlinear local convergence of Algorithm 5.1.

(H1) There exist a constant $C > 0$, a neighborhood U of $\bar{x}$, and a neighborhood V of the origin in $\mathbb{R}^n$ such that the following holds:

For all $x \in U$, $z \in V$, and for any $d \in \mathbb{R}^n$ with $-H(x) \in DH(x)(d)$ there is a vector $w \in \tilde{D}H(x)(z)$ such that

$$C\|d - z\| \leq \|w + H(x)\| + o(\|x - \bar{x}\|).$$

(H2) There exists a neighborhood U of $\bar{x}$ such that for all $u \in U$ and for all $v \in \tilde{D}H(x)(\bar{x} - x)$ we have

$$\|H(x) - H(\bar{x}) + v\| = o(\|x - \bar{x}\|).$$

A detailed discussion of these two assumptions and sufficient conditions for their fulfillment are given in the next section. Note that assumption (H2) means, in the terminology of [21, Definition 7.2.2] focused on locally Lipschitzian mappings H, that the family $\{\widetilde{D}H(x)\}$ provides a *Newton approximation scheme* for H at $\bar{x}$.

Now we establish our principal local convergence result that makes use of the major assumptions (H1) and (H2) together with metric regularity.

Theorem 5.3 (superlinear local convergence of the generalized Newton's method). *Let $\bar{x} \in \mathbb{R}^n$ be a solution to* (1.3) *for which the underlying mapping $H\colon \mathbb{R}^n \to \mathbb{R}^n$ is metrically regular around $\bar{x}$ and assumptions* (H1) *and* (H2) *are satisfied. Then there is a number $\varepsilon > 0$ such that for all $x^0 \in B_\varepsilon(\bar{x})$ the following assertions hold:*

(i) *Algorithm* 5.1 *is well defined and generates a sequence $\{x^k\}$ converging to $\bar{x}$.*

(ii) *The rate of convergence $x^k \to \bar{x}$ is at least superlinear.*

Proof. To justify (i), pick $\varepsilon > 0$ such that assumptions (H1) and (H2) hold with $U := B_\varepsilon(\bar{x})$ and $V := B_\varepsilon(0)$ and such that Proposition 5.2 can be applied. Then we choose a starting point $x^0 \in B_\varepsilon(\bar{x})$ and conclude by Proposition 5.2 that the subproblem

$$-H(x^0) \in DH(x^0)(d)$$

has a solution d^0. Thus the next iterate $x^1 := x^0 + d^0$ is well defined. Let further $z^0 := \bar{x} - x^0$ and get $\|z^0\| \leq \varepsilon$ by the choice of the starting point x^0. By assumption (H1), find a vector $w^0 \in \widetilde{D}H(x^0)(z^0)$ such that

$$C\|x^1 - \bar{x}\| = C\|(x^1 - x^0) - (\bar{x} - x^0)\| = C\|d^0 - z^0\| \leq \|w^0 + H(x^0)\| + o(\|x^0 - \bar{x}\|).$$

Taking this into account and employing assumption (H2), we get the relationships

$$\begin{aligned} C\|x^1 - \bar{x}\| &\leq \|w^0 + H(x^0)\| + o(\|x^0 - \bar{x}\|) = \|H(x^0) - H(\bar{x}) + w^0\| + o(\|x^0 - \bar{x}\|) \\ &= o(\|x^0 - \bar{x}\|) \leq \tfrac{C}{2}\|x^0 - \bar{x}\|, \end{aligned}$$

which imply that $\|x^1 - \bar{x}\| \leq \frac{1}{2}\|x^0 - \bar{x}\|$. The latter yields, in particular, that $x^1 \in B_\varepsilon(\bar{x})$. Now standard induction arguments allow us to conclude that the iterative sequence $\{x^k\}$ generated by Algorithm 5.1 is well defined and converges to the solution $\bar{x}$ of (1.3) with at least a linear rate. This justifies assertion (i) of the theorem.

Next we prove assertion (ii) showing that the convergence $x^k \to \bar{x}$ is in fact *superlinear* under the validity of assumption (H2). To proceed, we basically follow the proof of assertion (i) and construct by induction sequences $\{d^k\}$ satisfying $-H(x^k) \in DH(x^k)(d^k)$ for all $k \in \mathbb{N}$,

$\{z^k\}$ with $z^k := \bar{x} - x^k$, and $\{w^k\}$ with $w^k \in \widetilde{D}H(x^k)(z^k)$ such that

$$C\|x^{k+1} - \bar{x}\| \le \|w^k + H(x^k)\| + o(\|x^k - \bar{x}\|), \quad k \in \mathbb{N}.$$

Applying then assumption (H2) gives us the relationships

$$C\|x^{k+1} - \bar{x}\| \le \|H(x^k) - H(\bar{x}) + w^k\| + o(\|x^k - \bar{x}\|) = o(\|x^k - \bar{x}\|),$$

which ensure the superlinear convergence of the iterative sequence $\{x^k\}$ to the solution $\bar{x}$ of (1.3) and thus complete the proof of the theorem. □

5.1.3 Global Convergence

Besides the local convergence in the Newton's method based on suitable assumptions imposed at the (unknown) solution, there are global (or semi-local) convergence results of the Kantorovich type [30] which show that, under certain conditions at the starting point x^0 and a number of assumptions to hold in a suitable region around x^0, Newton's iterates are well defined and converge to a solution belonging to this region; see [15, 30] for more details and references. In the case of nonsmooth equations (1.3) results of the Kantorovich type were obtained in [57, 60] for the corresponding versions of Newton's method. Global convergence results of different types can be found in, e.g., [14, 21, 26, 53] and their references.

Here is a global convergence result for our generalized Newton's method to solve (1.3).

Theorem 5.4 (global convergence of the generalized Newton's method). *Let x^0 be a starting point of Algorithm* 5.1*, and let*

$$\Omega := \left\{x \in \mathbb{R}^n \middle| \ \|x - x^0\| \le r\right\} \tag{5.4}$$

with some $r > 0$. *Impose the following assumptions:*

(a) *The mapping* $H\colon \mathbb{R}^n \to \mathbb{R}^n$ *in* (1.3) *is metrically regular on* Ω *with modulus* $\mu > 0$, *i.e., it is metrically regular around every point* $x \in \Omega$ *with the same modulus* μ.

(b) *The set-valued map* $DH(x)(z)$ *uniformly on* Ω *converges to* $\{0\}$ *as* $z \to 0$ *in the sense that: for all* $\varepsilon > 0$ *there is* $\delta > 0$ *such that*

$$\|w\| \le \varepsilon \ \text{ whenever } \ w \in DH(x)(z),\ \|z\| \le \delta, \ \text{ and } \ x \in \Omega.$$

(c) *There is* $\alpha \in (0, 1/\mu)$ *such that*

$$\mu\|H(x^0)\| \le r(1 - \alpha\mu) \tag{5.5}$$

and for all $x, y \in \Omega$ *we have the estimate*

$$\|H(x) - H(y) - v\| \le \alpha\|x - y\| \ \text{ whenever } \ v \in DH(x)(y - x). \tag{5.6}$$

Then Algorithm 5.1 *is well defined, the sequence of iterates* $\{x^k\}$ *remains in* Ω *and converges to a solution* $\bar{x} \in \Omega$ *of* (1.3). *Moreover, we have the error estimate*

$$\|x^k - \bar{x}\| \le \frac{\alpha\mu}{1 - \alpha\mu}\|x^k - x^{k-1}\| \ \text{ for all } \ k \in \mathbb{N}. \tag{5.7}$$

Proof. The metric regularity assumption (a) allows us to employ Proposition 5.2 and, for any $x \in \Omega$ and $d \in \mathbb{R}^n$ satisfying the inclusion $-H(x) \in DH(x)(d)$, to find sequences of

$h_k \to -H(x)$ and $t_k \downarrow 0$ as $k \to \infty$ such that

$$\|d\| = \lim_{k\to\infty} \left\| \frac{H^{-1}\big(H(x) + t_k h_k\big) - x}{t_k} \right\| \le \lim_{k\to\infty} \mu\|h_k\| = \mu\|H(x)\|.$$

In view of assumption (5.5) in (c) and the iteration procedure of the algorithm, this implies

$$\|x^1 - x^0\| = \|d^0\| \le \mu\|H(x^0)\| \le r(1 - \alpha\mu),$$

which ensures that $x^1 \in \Omega$ due to the form of Ω in (5.4) and the choice of α. Proceeding further by induction, suppose that $x^1, \ldots, x^k \in \Omega$ and get the relationships

$$\begin{aligned}
\|x^{k+1} - x^k\| = \|d^k\| \le \mu\|H(x^k)\| &\le \mu\|H(x^k) - H(x^{k-1}) + H(x^{k-1})\| \\
&\le \alpha\mu\|x^k - x^{k-1}\| \quad \Big(\text{using (5.6) and } -H(x^{k-1}) \in DH(x^{k-1})(x^k - x^{k-1})\Big) \\
&\le (\alpha\mu)^k\|x^1 - x^0\| \le r(\alpha\mu)^k(1 - \alpha\mu),
\end{aligned}$$

which imply the estimates

$$\|x^{k+1} - x^0\| \le \sum_{j=0}^{k} \|x^{j+1} - x^j\| \le \sum_{j=0}^{k} r(\alpha\mu)^j(1 - \alpha\mu) \le r$$

and hence justify that $x^{k+1} \in \Omega$. Thus all the iterates generated by Algorithm 5.1 remain in Ω. Furthermore, for any natural numbers k and m, we have

$$\|x^{k+m+1} - x^k\| \le \sum_{j=k}^{k+m} \|x^{j+1} - x^j\| \le \sum_{j=k}^{k+m} r(\alpha\mu)^j(1 - \alpha\mu) \le r(\alpha\mu)^k,$$

which shows that the generated sequence $\{x^k\}$ is a Cauchy sequence. Hence it converges to some point $\bar{x}$ that obviously belongs to the underlying closed set (5.4).

To show next that $\bar{x}$ is a solution to the original equation (1.3), we pass to the limit as $k \to \infty$ in the iterative inclusion $-H(x^k) \in DH(x^k)(x^{k+1} - x^k)$, $k \in \mathbb{N}$.

It follows from assumption (b) that $\lim_{k\to\infty} H(x^k) = 0$. The continuity of H then implies that $H(\bar{x}) = 0$, i.e., $\bar{x}$ is a solution to (1.3).

It remains to justify the error estimate (5.7). To this end, first observe by (5.5) that

$$\|x^{k+m+1} - x^k\| \leq \sum_{j=k}^{k+m} \|x^{j+1} - x^j\| \leq \sum_{j=0}^{m} (\alpha\mu)^{j+1} \|x^k - x^{k-1}\| \leq \frac{\alpha\mu}{1 - \alpha\mu} \|x^k - x^{k-1}\|$$

for all $k, m \in \mathbb{N}$. Passing now to the limit as $m \to \infty$, we arrive at (5.7) thus completes the proof of the theorem. $\square$

5.2 Discussion of Major Assumptions and Comparison with Semismooth Newton's methods

In this section we pursue a twofold goal: to discuss the major assumptions made in Section 3 and to compare our generalized Newton's method based on graphical derivatives with the semismooth versions of the generalized Newton's method developed in [56, 57]. As we will see from the discussions below, these two aims are largely interrelated. Let us begin with sufficient conditions for metric regularity in terms of the constructions used in the semismooth versions of the generalized Newton's method. Given a locally Lipschitz continuous vector-valued mapping $H\colon \mathbb{R}^n \to \mathbb{R}^m$, we have by the classical Rademacher theorem that the set of points

$$S_H := \{x \in \mathbb{R}^n \big| \; H \text{ is differentiable at } x\} \tag{5.8}$$

is of full Lebesgue measure in $\mathbb{R}^n$.

Thus for any mapping $H\colon \mathbb{R}^n \to \mathbb{R}^m$ locally Lipschitzian around $\bar{x}$ the set

$$\partial_B H(\bar{x}) := \Big\{ \lim_{k\to\infty} H'(x^k) \Big| \, \exists \{x^k\} \subset S_H \ \text{ with } \ x^k \to \bar{x} \Big\} \tag{5.9}$$

is nonempty and obviously compact in $\mathbb{R}^m$. It was introduced in [66] for $m = 1$ as the set of "almost-gradients" and then was called in [56] the *B-subdifferential* of H at $\bar{x}$. Clarke's *generalized Jacobian* [12] of H at $\bar{x}$ is defined by the convex hull

$$\partial_C H(\bar{x}) := \text{co} \big\{ \partial_B H(\bar{x}) \big\}. \tag{5.10}$$

We also make use of the *Thibault derivative/limit set* [67] (called sometimes the "strict graphical derivative" [61]) of H at $\bar{x}$ defined by

$$D_T H(\bar{x})(z) := \mathop{\text{Lim}\,\text{sup}}_{\substack{x\to\bar{x}\\ t\downarrow 0}} \frac{H(x+tz) - H(x)}{t}, \quad z \in \mathbb{R}^n. \tag{5.11}$$

Observe the known relationships [33, 67] between the above derivative sets

$$\partial_B H(\bar{x}) z \subset D_T H(\bar{x})(z) \subset \partial_C H(\bar{x}) z, \quad z \in \mathbb{R}^n. \tag{5.12}$$

The next result gives a sufficient condition for metric regularity of Lipschitzian mappings in terms of the Thibault derivative (5.11).

Proposition 5.5 (sufficient condition for metric regularity in terms of Thibault's derivative). *Let $H\colon \mathbb{R}^n \to \mathbb{R}^n$ be locally Lipschitzian around $\bar{x}$, and let*

$$0 \notin D_T H(\bar{x})(z) \ \ \textit{whenever} \ \ z \neq 0. \tag{5.13}$$

Then the mapping H is metrically regular around $\bar{x}$.

Proof. Kummer's inverse function theorem [37, Theorem 1.1] ensures that condition (5.13) implies (actually is equivalent to) the fact that there are neighborhoods U of $\bar{x}$ and V of $H(\bar{x})$ such that the mapping $H\colon U \to V$ is one-to-one with a locally Lipschitzian inverse $H^{-1}\colon V \to U$. Let $\mu > 0$ be a Lipschitz constant of H^{-1} on V. Then for all $x \in U$ and $y \in V$ we have the relationships

$$\begin{aligned}\operatorname{dist}\big(x; H^{-1}(y)\big) &= \|x - H^{-1}(y)\| = \|H^{-1}\big(H(x)\big) - H^{-1}(y)\| \\ &\le \mu \|H(x) - y\| = \mu \operatorname{dist}\big(y; H(x)\big),\end{aligned}$$

which thus justify the metric regularity of H around $\bar{x}$. □

To proceed further with sufficient conditions for the validity of our assumption (H1), we first introduce the notion of directional boundedness.

Definition 5.6 (directional boundedness). *A mapping $H\colon \mathbb{R}^n \to \mathbb{R}^m$ is said to be* DIRECTIONALLY BOUNDED *around $\bar{x}$ if*

$$\limsup_{t \downarrow 0} \left\| \frac{H(x + tz) - H(x)}{t} \right\| < \infty \tag{5.14}$$

for all x near $\bar{x}$ and for all $z \in \mathbb{R}^n$.

It is easy to see that if H is either directionally differentiable around $\bar{x}$ or locally Lipschitzian around this point, then it is directionally bounded around $\bar{x}$. The following example shows that the converse does not hold in general.

Example 5.7 (directionally bounded mappings may be neither directionally differ-

entiable nor locally Lipschitzian). Define a real function $H\colon \mathbb{R} \to \mathbb{R}$ by

$$H(x) := \begin{cases} x \sin\left(\frac{1}{x}\right) & \text{if } \ x \neq 0, \\ 0 & \text{if } \ x = 0. \end{cases}$$

It is easy to see that this function is not either directionally differentiable at $\bar{x} = 0$ or locally Lipschitzian around this point. However, it is directionally bounded around $\bar{x}$. Indeed, for any $x \neq 0$ near $\bar{x}$ condition (5.14) holds because H is simply differentiable at $x \neq 0$. For $x = 0$ we have

$$\limsup_{t \downarrow 0} \left| \frac{H(tz) - H(0)}{t} \right| = \limsup_{t \downarrow 0} \frac{|H(tz)|}{t} = \limsup_{t \downarrow 0} \left| z \sin\left(\frac{1}{tz}\right) \right| = |z| < \infty.$$

The next proposition and its corollary present verifiable sufficient conditions for the fulfillment of assumption (H1).

Proposition 5.8 (assumption (H1) from metric regularity). *Let $H\colon \mathbb{R}^n \to \mathbb{R}^n$, and let $\bar{x}$ be a solution to* (1.3). *Suppose that H is metrically regular around $\bar{x}$ (i.e.,* $\ker D^*H(\bar{x}) = 0$*), that it is directionally bounded and one-to-one around this point. Then assumption* (H1) *is satisfied.*

Proof. Recall that the metric regularity of H around $\bar{x}$ is equivalent to the condition $\ker D^*H(\bar{x}) = \{0\}$ by the coderivative criterion (2.18). Let $U \subset \mathbb{R}^n$ be a neighborhood of $\bar{x}$ such that H is metrically regular and one-to-one on U. Choose further a neighborhood $V \subset \mathbb{R}^n$ of $H(\bar{x}) = 0$ from the definition of metric regularity of H around $\bar{x}$. Then pick $x \in U$, $z \in V$ and an arbitrary direction $d \in \mathbb{R}^n$ satisfying $-H(x) \in DH(x)(d)$. Employing now Proposition 5.2, we get

$$d \in \operatorname*{Lim\,sup}_{h \to -H(x),\, t \downarrow 0} \frac{H^{-1}\big(H(x) + th\big) - x}{t}.$$

By the local single-valuedness of H^{-1} and the metric regularity of H around $\bar{x}$ there exists a number $\mu > 0$ such that

$$\left\| \frac{H^{-1}(H(x)+th)-x}{t} - z \right\| \le \mu \left\| \frac{H(x)+th-H(x+tz)}{t} \right\| = \mu \left\| \frac{H(x+tz)-H(x)}{t} - h \right\|$$

for all $t > 0$ sufficiently small. It follows that

$$\|d-z\| \le \limsup_{\substack{t\downarrow 0 \\ h\to -H(x)}} \left\| \frac{H^{-1}\big(H(x)+th\big)-x}{t} - z \right\| \le \mu \limsup_{\substack{t\downarrow 0 \\ h\to -H(x)}} \left\| \frac{H(x+tz)-H(x)}{t} - h \right\| < \infty$$

by the directional boundedness of H around $\bar{x}$. The boundedness of the family

$$\Big\{ v(t) := \frac{H(x+tz)-H(x)}{t} \Big\}, \quad t \downarrow 0,$$

allows us to select a sequence $t_k \downarrow 0$ such that $v(t_k) \to w$ for some $w \in \mathbb{R}^n$. By passing to the limit above as $k \to \infty$ and employing Definition 2.2 we get that

$$w \in \widetilde{D}H(x)(z) \quad \text{and} \quad \frac{1}{\mu}\|d-z\| \le \|w + H(x)\|,$$

which completes the proof of the proposition. $\square$

Corollary 5.9 (sufficient conditions for (H1) via Thibault's derivative). *Let $\bar{x}$ be a solution to* (1.3), *where $H\colon \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitzian around $\bar{x}$ and such that condition* (5.13) *holds, which is automatic when* $\det A \neq 0$ *for all $A \in \partial_C H(\bar{x})$. Then* (H1) *is satisfied with H being both metrically regular and one-to-one around $\bar{x}$.*

Proof. Indeed, both metric regularity and bijectivity of H around $\bar{x}$ assumed in Proposition 5.8 follow from Proposition 5.5 and its proof. Nonsingularity of all $A \in \partial_C H(\bar{x})$ clearly implies (5.13) by the second inclusion in (5.12). □

Note that other conditions ensuring the fulfillment of assumption (H1) for Lipschitzian and non-Lipschitzian mappings $H\colon \mathbb{R}^n \to \mathbb{R}^n$ can be formulated in terms of Warga's *derivate containers* by [69, Theorems 1 and 2] on "fat homeomorphisms" that also imply the metric regularity and one-to-one properties of H.

Next we proceed with the discussion of assumption (H2) and present, in particular, sufficient conditions for their fulfillment via semismoothness. First observe the following.

Proposition 5.10 (relationship between graphical derivative and generalized Jacobian). *Let $H\colon \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitzian around $\bar{x}$. Then we have*

$$DH(\bar{x})(z) \subset \partial_C H(\bar{x})z \ \ \textit{for all} \ \ z \in \mathbb{R}^n. \tag{5.15}$$

Proof. Pick $w \in DH(\bar{x})(z)$ and get by Proposition 2.3(c) and Definition 2.2 a sequence of $t_k \downarrow 0$ as $k \to \infty$ such that

$$w = \lim_{k\to\infty} \frac{H(\bar{x}+t_k z) - H(\bar{x})}{t_k}. \tag{5.16}$$

It is easy to see from (5.16) and the definition of the Thibault derivative (5.11) that we have $w \in D_T H(\bar{x})(z)$. Then the desired result $w \in \partial_C H(\bar{x})$ follows from (5.12). □

Inclusion (5.15), which may be strict as illustrated by Example 5.11 below, shows that our generalized Newton algorithm 5.1 based on the graphical derivative provides in the case of Lipschitz equations (1.3) a more accurate choice of the iterative direction d^k via (5.1) in comparison with the iterative relationship

$$-H(x^k) \in \partial_C H(x^k)d^k, \quad k = 0, 1, 2, \ldots, \tag{5.17}$$

used in the semismooth Newton's method [57] and related developments [33, 36] based on the generalized Jacobian. If in addition to the assumptions of Proposition 5.10 the mapping H is directionally differentiable at $\bar{x}$, then $DH(\bar{x})(z) = \{H'(\bar{x}; z)\}$ by Proposition 2.3(c,d). Thus in this case we have from Proposition 5.10 that for any $z \in \mathbb{R}^n$ there is $A \in \partial_C H(\bar{x})$ such that $H'(\bar{x}; z) = Az$, which recovers a well-known result from [57, Lemma 2.2].

The following example shows that the converse inclusion in Proposition 5.10 is not satisfied in general even with the replacement of the set $DH(\bar{x})(z)$ in (5.15) by its convex hull $\text{co}DH(\bar{x})(z)$ in the case of real functions. Furthermore, the same holds if we replace the generalized Jacobian in (5.15) by the smaller B-subdifferential $\partial_B H(\bar{x})$ from (5.9).

Example 5.11 (graphical derivative is strictly smaller than B-subdifferential and generalized Jacobian). Consider the simplest nonsmooth convex function $H(x) = |x|$ on $\mathbb{R}$. In this case $\partial_B H(0) = \{-1, 1\}$ and $\partial_C H(0) = [-1, 1]$. Thus

$$\partial_B H(0)z = \{-1, 1\} \ \text{ and } \ \partial_C H(0)z = [-1, 1] \ \text{ for } \ z = 1.$$

Since $H(x) = |x|$ is locally Lipschitzian and directionally differentiable, we have

$$DH(0)(z) = \big\{H'(0; z)\big\} = |z| = \{1\} \ \text{ for } \ z = 1.$$

Hence it gives the relationships

$$DH(0)(z) = \text{co}\big\{DH(0)(z)\big\} \subset \partial_B H(0)z \subset \partial_C H(0)z,$$

where both inclusions are strict. Observe also the difference between the convexification of the graphical derivative and of the coderivative; in the latter case we have equality (2.15).

As mentioned in Section 1, there is an improvement [56] of the iterative procedure (5.17) with the replacement the generalized Jacobian therein by the B-subdifferential

$$-H(x^k) \in \partial_B H(x^k) d^k, \quad k = 0, 1, 2, \ldots. \tag{5.18}$$

Note that, along with obvious advantages of version (5.18) over the one in (5.17), in some settings it is easier to deal with the generalized Jacobian than with its B-subdifferential counterpart due to much better calculus and convenient representations for $\partial_C H(\bar{x})$ in comparison with the case of $\partial_B H(\bar{x})$, which does not even reduce to the classical subdifferential of convex analysis for simple convex functions as, e.g., $H(x) = |x|$. A remarkable common feature for both versions in (5.17) and (5.18) is the efficient semismoothness assumption imposed on the underlying mapping H to ensure its local superlinear convergence. This assumption, which unifies and labels versions (5.17) and (5.18) as the "semismooth Newton's method", is replaced in our generalized Newton's method by assumption (H2). Let us now recall the notion of semismoothness and compare it with (H2).

A mapping $H\colon \mathbb{R}^n \to \mathbb{R}^m$, locally Lipschitzian and directionally differentiable around $\bar{x}$, is *semismooth* at this point if the limit

$$\lim_{\substack{h \to z,\ t \downarrow 0 \\ A \in \partial_C H(\bar{x}+th)}} \{Ah\} \tag{5.19}$$

exists for all $z \in \mathbb{R}^n$; see [21, Definition 7.4.2]. This notion was introduced in [44] for real-valued functions and then extended in [57] to the vector mappings for the purpose of applications to a nonsmooth Newton's method. It is not hard to check [57, Proposition 2.1] that the existence of the limit in (5.19) implies the directional differentiability of H at $\bar{x}$ (but may not around this point) with

$$H'(\bar{x};z) = \lim_{\substack{h\to z,\ t\downarrow 0\\ A\in\partial_C H(\bar{x}+th)}} \big\{Ah\big\} \ \text{ for all } \ z \in \mathbb{R}^n.$$

One of the most useful properties of semismooth mappings is the following representation for them obtained in [54, Proposition 1]:

$$\|H(\bar{x}+z) - H(\bar{x}) - Az\| = o(\|z\|) \ \text{ for all } \ z \to 0 \ \text{ and } \ A \in \partial_C H(\bar{x}+z), \tag{5.20}$$

which we exploit now to relate semismoothness to our assumption (H2).

Proposition 5.12 (semismoothness implies assumption (H2)). *Let $H\colon \mathbb{R}^n \to \mathbb{R}^m$ be semismooth at $\bar{x}$. Then assumption* (H2) *is satisfied.*

Proof. Since any semismooth mapping is Lipschitz continuous on a neighborhood U of $\bar{x}$, we have by Proposition 2.3(c) that

$$\widetilde{D}H(x)(\bar{x}-x) = DH(x)(\bar{x}-x) \ \text{ for all } \ x \in U.$$

Proposition 5.10 yields therefore that

$$\widetilde{D}H(x)(\bar{x}-x) \subset \partial_C H(x)(\bar{x}-x) \ \text{ whenever } \ x \in U.$$

Given any $v \in \widetilde{D}H(x)(\bar{x}-x)$ and using the latter inclusion, find a matrix $A \in \partial_C H(x)$ such that $v = A(\bar{x}-x)$. Applying finally property (5.20) of semismooth mappings, we get

$$\|H(x) - H(\bar{x}) + v\| = \|H(x) - H(\bar{x}) - A(x-\bar{x})\| = o(\|x-\bar{x}\|) \ \text{ for all } \ x \in U,$$

which thus verifies (H2) and completes the proof of the proposition. □

Note that the previous proposition actually shows that condition (5.20) implies (H2). The next result states that the converse is also true, i.e., we have that assumption (H2) is completely equivalent to (5.20) for locally Lipschitzian mappings.

Proposition 5.13 (equivalent description of (H2)). *Let $H\colon \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitzian around $\bar{x}$, and let assumption* (H2) *hold with some neighborhood U therein. Then*

$$\|H(x) - H(\bar{x}) - A(x - \bar{x})\| = o(\|\bar{x} - x\|) \ \text{ for all } \ x \in U \ \text{ and } \ A \in \partial_B H(x). \tag{5.21}$$

Therefore assumption (H2) *is equivalent to* (5.20)*.*

Proof. Arguing by contradiction, suppose that (5.21) is violated and find sequences $x_k \to \bar{x}$, $A_k \in \partial_B H(x_k)$ and a constant $\gamma > 0$ such that

$$\|H(x_k) - H(\bar{x}) - A_k(x_k - \bar{x})\| \geq \gamma \|\bar{x} - x_k\|, \quad k \in \mathbb{N}.$$

By the Lipschitz property of H and by construction (5.9) of the B-subdifferential there are points of differentiability $u_k \in S_H$ close to x_k with $H'(u_k)$ sufficiently close to A_k satisfying

$$\|H(u_k) - H(\bar{x}) - H'(u_k)(u_k - \bar{x})\| \geq \tfrac{\gamma}{2} \|\bar{x} - u_k\|, \quad k \in \mathbb{N}.$$

Then Proposition 2.3(c,f) gives us the representations

$$\widetilde{D}H(u_k)(\bar{x} - u_k) = DH(u_k)(\bar{x} - u_k) = -H'(u_k)(u_k - \bar{x})$$

for all $k \in \mathbb{N}$, which imply therefore that

$$\|H(u_k) - H(\bar{x}) + v\| \geq \frac{\gamma}{2}\|\bar{x} - u_k\| \text{ whenever } v \in \widetilde{D}H(u_k)(\bar{x} - u_k), \quad k \in \mathbb{N}.$$

This clearly contradicts assumption (H2) for k sufficiently large and thus ensures property (5.21). The equivalence between (H2) and (5.20) follows now from the implication (H2)$\Longrightarrow$(5.21) and the proof of Proposition 5.12. □

It is well known that, for the class of locally Lipschitzian and directionally differentiable mappings, condition (5.20) is equivalent to the original definition of semismoothness; see, e.g., [21, Theorem 7.4.3]. Proposition 5.13 above establishes the equivalence of (5.20) to our major assumption (H2) provided that H is locally Lipschitzian around the reference point while it may *not* be directionally differentiable therein. In fact, it follows from Example 5.15 that assumption (H2) may hold for locally Lipschitzian functions, which are not directionally differentiable and hence not semismooth. Let us now illustrate that (H2) may also be satisfied for non-Lipschitzian mappings, in which case it is *not* equivalent to property (5.20).

Example 5.14 (assumption (H2) holds for non-Lipschitzian one-to-one mappings). Consider the mapping $H\colon \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$H(x_1, x_2) := \Big(x_2\sqrt{|x_1| + |x_2|^3}, x_1\Big) \text{ for } x_1, x_2 \in \mathbb{R}. \tag{5.22}$$

It is easy to check that this mapping is one-to-one around $(0, 0)$. Focusing for definiteness on the nonnegative branch of the mapping H, observe that at any point $(x_1, x_2) \in \mathbb{R}^2$ with either

$x_1, x_2 > 0$, the classical Jacobian $JH(x_1,x_2)$ is computed by

$$JH(x_1,x_2) = \begin{bmatrix} \dfrac{x_2}{2\sqrt{x_1+x_2^3}} & \sqrt{x_1+x_2^3} + \dfrac{3x_2^3}{2\sqrt{x_1+x_2^3}} \\ 1 & 0 \end{bmatrix}.$$

Setting $x_1 = x_2^3$, we see that the first component

$$\frac{x_2}{2\sqrt{x_1+x_2^3}} = \frac{x_2}{2\sqrt{x_2^3+x_2^3}}$$

is unbounded when $x_1, x_2 \downarrow 0$. This implies that the Jacobian $JH(x_1,x_2)$ is unbounded around $(\bar{x}_1,\bar{x}_2) = (0,0)$, and hence H is not locally Lipschitzian around the origin.

Let us finally verify that the underlying assumption (H2) is satisfied for the mapping H in (5.22). First assume that $x_1, x_2 > 0$. Then we need to check that

$$\begin{aligned} &\|H(x_1,x_2) - H(\bar{x}_1,\bar{x}_2) + JH(x_1,x_2)(-x_1,-x_2)\| \\ &= \left| x_2\sqrt{x_1+x_2^3} - \frac{x_1x_2}{2\sqrt{x_1+x_2^3}} - x_2\sqrt{x_1+x_2^3} - \frac{3x_2^4}{2\sqrt{x_1+x_2^3}} \right| \\ &= \left| \frac{x_1x_2}{2\sqrt{x_1+x_2^3}} + \frac{3x_2^4}{2\sqrt{x_1+x_2^3}} \right| = o\big(\sqrt{x_1^2+x_2^2}\big). \end{aligned}$$

The latter surely holds as $(x_1,x_2) \to (0,0)$ due to the estimates

$$\frac{x_1x_2}{2\sqrt{x_1+x_2^3}\sqrt{x_1^2+x_2^2}} \le \frac{x_1}{\sqrt{x_1+x_2^3}} \le \sqrt{x_1},$$

$$\frac{3x_2^4}{2\sqrt{x_1+x_2^3}\sqrt{x_1^2+x_2^2}} \le \frac{3x_2^3}{2\sqrt{x_1+x_2^3}} \le 3x_2,$$

which thus justify the fulfillment of assumption (H2) in this case. The other cases where

$x_1 > 0, x_2 \le 0$ or $x_1 < 0, x_2 > 0$ or $x_1 < 0, x_2 \le 0$ or, finally, $x_1 = 0, x_2$ arbitrary (here H is not differentiable) can be treated in a similar way.

To complete our discussion on the major assumptions in this section, let us present an example of a locally Lipschitzian function, which satisfies assumptions (H1) and (H2) being locally one-to-one and metrically regular around the point in question while not being directionally differentiable and hence not semismooth at this point. Other examples of this type involving Lipschitzian while not directionally differentiable functions useful for different versions of the generalized Newton's method can be found in [21, 33, 36].

Example 5.15 (non-semismooth but metrically regular, Lipschitzian, and one-to-one functions satisfying (H1) and (H2)). We construct a function $H\colon [-1,1] \to \mathbb{R}$ in the following way. First set $H(\bar{x}) := 0$ at $\bar{x} = 0$. Then define H on the interval $(1/2, 1]$ staying between two lines

$$\Big(1 - \frac{1}{2}\Big)x + \frac{1}{4} \le H(x) \le x$$

in the following way: start from $(1,1)$ and let H be continuous piecewise linear when x goes from 1 to 1/2 with the slope 1+1/4 and then with the slope $1/2 - 1/4$ alternatively until x reaches 1/2. Consider further each interval $(2^{-k}, 2^{-(k-1)}]$ for $k = 2, 3, \ldots$ and, starting from the point $\big(2^{-(k-1)}, 2^{-(k-1)}\big)$, define H to be continuous piecewise linear with the corresponding slopes of either $1 + 2^{-2k}$ or $1 - 2^{-k} - 2^{-2k}$ staying between the two lines

$$\Big(1 - \frac{1}{2^k}\Big)x + \frac{1}{2^{2k}} \le H(x) \le x. \tag{5.23}$$

Thus we have constructed H on the whole interval $[0, 1]$. On the interval $[-1, 0]$, define the function H symmetrically with respect to the origin. Then it is easy to see that H in continuous on $[-1,1]$ and satisfies the following properties:

- H is clearly Lipschitz continuous around $\bar{x}=0$.

- Since H is continuous and monotone with a positive uniform slope, it is one-to-one and metrically regular around $\bar{x}$, which directly follows, e.g., from the coderivative criterion (2.18). This ensures the fulfillment of assumption (H1) by Proposition 5.8.

- To verify assumption (H2), fix $k\in\mathbb{N}$ and $x\in(2^{-k},2^{-(k-1)}]$ and then pick any

$$v\in DH(x)(\bar{x}-x)\subset\Big[1-\frac{1}{2^k}-\frac{1}{2^{2k}},1+\frac{1}{2^{2k}}\Big](\bar{x}-x).$$

  Since $\bar{x}=0$, the latter implies that

$$-\Big(1+\frac{1}{2^{2k}}\Big)x\le v\le\Big(1-\frac{1}{2^k}-\frac{1}{2^{2k}}\Big)x$$

  Thus we have by (5.23) and simple computations that

$$|H(x)-H(\bar{x})+v|\le\frac{1}{2^k}|x|+\frac{1}{2^{2k}}+\frac{1}{2^{2k}}=o\Big(\frac{1}{2^k}\Big)=o(|x-\bar{x}|),$$

  which shows that assumption (H2) is satisfied. In fact, it follows from above that the latter value is $O(2^{-2k})=O(\|x-\bar{x}\|^2)$.

- Let us finally check that H is not directionally differentiable at $x_k=2^{-k}$ for any $k\in\mathbb{N}$; therefore it is not directionally differentiable around the reference point $\bar{x}=0$ and hence not semismooth at $\bar{x}$. Indeed, this follows directly from computing the graphical derivative by

$$DH(x_k)(1)=\Big[1-\frac{1}{2^k},1\Big],\quad k\in\mathbb{N},$$

which is not single-valued at x_k, and thus H is not directionally differentiable at x_k due to Proposition 2.3(c,d).

5.3 Application to the B-differentiable Newton Method

In this section we present applications of the graphical derivative-based generalized Newton's method developed above to the B-differentiable Newton's method for nonsmooth equations (1.3) originated by Pang [52].

Throughout this section, suppose that $H\colon \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitzian and directionally differentiable around the reference solution $\bar{x}$ to (1.3). Proposition 2.3(c,d) yields in this setting that the generalized Newton equation (5.1) in our Algorithm 5.1 reduces to

$$-H(x^k) = H'(x^k; d^k) \tag{5.24}$$

with respect to the new search direction d^k and that the new iterate x^{k+1} is computed by

$$x^{k+1} := x^k + d^k, \quad k = 0, 1, 2, \ldots. \tag{5.25}$$

Note that Pang's B-differentiable Newton's method and its further developments (see, e.g., [21, 26, 53, 56, 57]) are based on Robinson's notion of the B(ouligand)-derivative [58] for nonsmooth mappings; hence the name. As was then shown in [64], the B-derivative of a locally Lipschitzian mapping agrees with the classical directional derivative. Thus the iteration scheme in Pang's B-differentiable method reduces to (5.24) and (5.25) in the Lipschitzian and directionally differentiable case, and so we keep the original name of [52].

The next theorem shows what we get from applying our local convergence result from Theorem 5.3 and the subsequent analysis developed in Sections 3 and 4 to the B-differentiable

Newton's method. This theorem employs an equivalent description of assumption (H2) held in the setting under consideration and the coderivative criterion (2.18) for metric regularity of the underlying Lipschitzian mapping H ensuring the validity of assumption (H1).

Theorem 5.16 (solvability and local convergence of the B-differentiable Newton's method via metric regularity). *Let $H\colon \mathbb{R}^n \to \mathbb{R}^n$ be semismooth, one-to-one, and metrically regular around a reference solution $\bar{x}$ to* (1.3), *i.e.,*

$$0 \in \partial\langle z, H\rangle(\bar{x}) \Longrightarrow z = 0. \tag{5.26}$$

Then the B-differentiable Newton's method (5.24), (5.25) *is well defined* (*meaning that equation* (5.24) *is solvable for d^k as $k \in \mathbb{N}$*) *and converges at least superlinearly to the solution $\bar{x}$.*

Proof. Since H is locally Lipschitzian around $\bar{x}$, the coderivative criterion (2.18) is equivalently written in form (5.26) via the limiting subdifferential (2.10) due to the scalarization formula (2.16). Applying Theorem 5.3 to the B-differentiable Newton's method, we need to check that assumptions (H1) and (H2) are satisfied in the setting under consideration. Indeed, it follows from Proposition 5.13 and the discussion right after it that (H2) is equivalent to the semismoothness for locally Lipschitzian and directionally differentiable mappings. The fulfillment of assumption (H1) is guaranteed by Proposition 5.8. □

More specific sufficient conditions for the well-posedness and superlinear convergence of the B-differentiable Newton's method are formulated via of the Thibault derivative (5.11).

Corollary 5.17 (B-differentiable Newton method via Thibault's derivative). *Let $H\colon \mathbb{R}^n \to \mathbb{R}^n$ be semismooth at the reference solution point $\bar{x}$ of equation* (1.3), *and let condition* (5.13) *be satisfied. Then the B-subdifferential Newton's method* (5.24), (5.25) *is well defined and* *converges superlinearly to the solution* $\bar{x}$*.*

Proof. Follows from Theorem 5.16 and Proposition 5.9. □

Observe by the second inclusion in (5.12) that the assumptions of Corollary 5.17 are satisfied when all the matrices from the generalized Jacobian $\partial_C H(\bar{x})$ are nonsingular. In the latter case the solvability of subproblem (5.24) and the superlinear convergence of the B-differentiable Newton's method follow from the results of [57] that in turn improve the original ones in [52], where H is assumed to be strongly Fréchet differentiable at the solution point.

Further, it is shown in [56] that the B-differentiable method for semismooth equations (1.3) converges superlinearly to the solution $\bar{x}$ if just matrices $A \in \partial_B H(\bar{x})$ are nonsingular while *assuming* in addition that subproblem (5.24) is *solvable*. As illustrated by the example presented on pp. 243–244 of [56], without the latter assumption the B-differentiable Newton method may not be well defined for semismooth mappings H on the plane with all the nonsingular matrices from $\partial_B H(\bar{x})$. We want to emphasize that the solvability assumption for (5.24) is not imposed in Theorem 5.16—it is *ensured* by *metric regularity*.

Let us now discuss interconnections between the metric regularity property of locally Lipschitzian mappings $H\colon \mathbb{R}^n \to \mathbb{R}^n$ via its coderivative characterization (5.26) and the nonsingularity of the generalized Jacobian and B-subdifferential of H at the reference point. To this end, observe the following relationships between the corresponding constructions.

Proposition 5.18 (relationships between the B-subdifferential, generalized Jacobian, and coderivative of Lipschitzian mappings). *Let* $H\colon \mathbb{R}^n \to \mathbb{R}^m$ *be locally Lipschitzian around* $\bar{x}$*. Then we have*

$$\partial_B H(\bar{x})^T z \subset \partial\langle z, H\rangle(\bar{x}) \subset \partial_C H(\bar{x})^T z \ \text{ for all } \ z \in \mathbb{R}^m, \tag{5.27}$$

where both inclusions in (5.27) *are generally strict.*

Proof. Recall that the middle term in (5.27) expressed via the limiting subdifferential (2.10) is exactly the coderivative $D^*H(\bar{x})(z)$ due to the scalarization formula (2.16) for locally Lipschitzian mappings. Thus the second inclusion in (5.27) follows immediately from the well-known equality (2.15) involving convexification, and it is strict as a rule due to the usual nonconvexity of the limiting subdifferential; see [47, 61].

To justify the first inclusion in (5.27), observe that the limiting subdifferential $\partial f(\bar{x})$ of every function $f\colon \mathbb{R}^n \to \mathbb{R}$ continuous around $\bar{x}$ admits the representation

$$\partial f(\bar{x}) = \operatorname*{Lim\,sup}_{x \to \bar{x}} \widehat{\partial} f(x) \tag{5.28}$$

via the outer limit (2.1) of the Fréchet/regular subdifferentials

$$\widehat{\partial} f(x) := \Big\{ p \in \mathbb{R}^n \Big| \liminf_{u \to x} \frac{f(u) - f(x) - \langle p, u - x \rangle}{\|u - x\|} \geq 0 \Big\} \tag{5.29}$$

of f at x; see, e.g., [47, Theorem 1.89]. We obviously have from (5.29) that $\widehat{\partial} f(\bar{x}) = \{f'(\bar{x})\}$ if f is (Fréchet) differentiable at $\bar{x}$ with its derivative/gradient $f'(\bar{x})$.

Having the mapping $H = (h_1, \ldots, h_m)\colon \mathbb{R}^n \to \mathbb{R}^m$ in the proposition and fixing an arbitrary vector $\bar{z} = (\bar{z}_1, \ldots, \bar{z}_m) \in \mathbb{R}^m$, form now a scalar function $f_{\bar{z}}\colon \mathbb{R}^n \to \mathbb{R}$ by

$$f_{\bar{z}}(x) := \sum_{i=1}^{m} \bar{z}_i h_i(x), \quad x \in \mathbb{R}^n. \tag{5.30}$$

Then the first inclusion in (5.27) amounts to say that

$$\partial_B H(\bar{x})^T \bar{z} \subset \partial f_{\bar{z}}(\bar{x}). \tag{5.31}$$

To proceed with proving (5.31), pick any matrix $A \in \partial_B H(\bar{x})^T \bar{z}$ and denote by $a_i \in \mathbb{R}^n$, $i = 1, \ldots, n$, its vector rows. By definition (5.9) of the B-subdifferential $\partial_B H(\bar{x})$ there is a sequence $\{x^k\} \subset S_H$ from the set of differentiability (5.8) such that $x^k \to \bar{x}$ and $H'(x^k) \to A$ as $k \to \infty$. It is clear from (5.30) that the function $f_{\bar{z}}$ is differentiable at each x^k with

$$f'_{\bar{z}}(x^k) = \sum_{i=1}^{m} \bar{z}_i h'_i(x^k) \to \sum_{i=1}^{m} \bar{z}_i a_i = A^T \bar{z} \quad \text{as} \quad k \to \infty.$$

Since $\widehat{\partial} f_{\bar{z}}(x^k) = \{f'_{\bar{z}}(x^k)\}$ at all the points of differentiability, we arrive at (5.31) by representation (5.28) of the limiting subdifferential and thus justify the first inclusion in (5.27).

To illustrate that the latter inclusion may be strict, consider the function $H(x) := |x|$ on $\mathbb{R}$. Then $\partial_B H(0)z = \{-z, z\}$ for all $z \in \mathbb{R}$, while

$$\partial(zH)(0) = D^*H(0)(z) = \begin{cases} [-z, z] & \text{for } \; z \geq 0, \\ \{-z, z\} & \text{for } \; z < 0. \end{cases}$$

This completes the proof of the proposition. □

It follows from Proposition 5.18 in the case of Lipschitzian transformations $H \colon \mathbb{R}^n \to \mathbb{R}^n$ that the nonsingularity of all the matrices $A \in \partial_C H(\bar{x})$ is a sufficient condition for the metric regularity of H around $\bar{x}$ due to the coderivative criterion (5.26) while the nonsingularity of all $A \in \partial_B H(\bar{x})$ is a necessary condition for this property. Note however, as it has been discussed above, that the nonsingularity condition for $\partial_B H(\bar{x})$ alone does not ensure the solvability of subproblem (5.24) in the B-differentiable Newton's method, and thus it cannot be used alone for the justification of algorithm (5.24), (5.25) in the B-differentiable semismooth case. Furthermore, we are not familiar with any verifiable condition to support the nonsingularity of $\partial_B H(\bar{x})$ in the full justification of the B-differentiable Newton's method.

In contrast to this, the metric regularity itself, via its verifiable pointwise characterization (5.26), ensures the solvability of (5.24) and fully justifies the B-differentiable Newton's method with its superlinear convergence provided that the mapping H is semismooth and locally invertible around the reference solution point. Note that the nonsingularity of the generalized Jacobian $\partial_C H(\bar{x})$ implies not only the metric regularity but simultaneously the semismoothness and local invertibility of a Lipschitzian transformation $H\colon \mathbb{R}^n \to \mathbb{R}^n$. However, the latter condition fails to spot a number of important situations when all the assumptions of Theorem 5.16 are satisfied; see, in particular, Corollary 5.17 and the corresponding conditions in terms of Warga's derivate containers discussed right after Corollary 5.9. We refer the reader to the specific mappings $H\colon \mathbb{R}^2 \to \mathbb{R}^2$ from [37, Example 2.2] and [69, Example 3.3] that can be used to illustrate the above statement.

5.4 Concluding Remarks

In this chapter we develop a new generalized Newton's method for solving systems of nonsmooth equations $H(x) = 0$ with $H\colon \mathbb{R}^n \to \mathbb{R}^n$. Local superlinear convergence and global (of the Kantorovich type) convergence results are derived under relatively mild conditions. In particular, the local Lipschitz continuity and directional differentiability of H are not necessarily required. We show that the new method and its specifications have some advantages in comparison with previously known results on the semismooth and B-differentiable versions of the generalized Newton's method for nonsmooth Lipschitz equations.

Our approach is heavily based on advanced tools of variational analysis and generalized differentiation. The algorithm itself is built by using the graphical/contingent derivative of H, while other graphical derivatives and coderivatives are employed in formulating appropriate assumptions and proving solvability and convergence results. The fundamental property of metric regularity and its pointwise coderivative characterization play a crucial role in the justification of the algorithm and its satisfactory performance.

In the other lines of developments, it seems appealing to develop an alternative Newton-type algorithm, which is constructed by using the basic coderivative instead of the graphical derivative. This requires certain symmetry assumptions for the given problem, since the coderivative is an extension of the adjoint derivative operator. Major advantages of a coderivative-based Newton's method would be comprehensive calculus rules held for the coderivative in contrast to the contingent derivative, complete coderivative characterizations of Lipschitzian stability, and explicit calculations of the coderivative in a number of settings important for applications. The details of these ideas are part of our future research.

Chapter 6

Damped Newton's method

6.1 The Damped Newton's Algorithm

In this section, we present the damped Newton's algorithm together with the standing assumptions. These conditions are needed to realize the algorithm and its the convergence results. We first let $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous nondecreasing function satisfying $\gamma(t) \downarrow 0$ as $t \downarrow 0$. Let $\eta : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ be a continuous function satisfying

$$\bar{\eta} := \limsup_{\substack{\|H(x)\| \to 0 \\ \|u\| \to 0}} \eta(x, u) < \frac{1}{M} \quad \text{for some given} \quad M > 0.$$

To begin our analysis, we assume the following assumptions

(G1) $DH(\cdot)$ satisfies the γ-range

$$\operatorname{diam} DH(x)(d) \leq \gamma(\|H(x)\|) \quad \text{for all} \quad \|d\| = 1.$$

(G2) $DH(\cdot)$ satisfies the η-approximation

$$H(x + u) - H(x) \in DH(x)(u) + \eta(x, u)\|u\|I\!B,$$

for all small $u \in \mathbb{R}^n$.

(G3) H is *canonically uniformly continuous* in the following sense: For all $\varepsilon, r > 0$, there exists $\delta = \delta(r, \varepsilon) > 0$ such that

$$\|H(x + u) - H(x)\| \leq \varepsilon \quad , \quad \text{for all } x \in \Omega, \|H(x)\| \geq r \ , \ \|u\| \leq \delta.$$

Briefly, assumption (G1) requires the "size" of the graphical derivative of H at each direction should not be too large as $\|H(x)\| \to 0$. In this scheme, γ serves as a *gauge* function depending on $\|H(x)\|$ only. Assumption (G2) indicates that the graphical derivative could well-approximate the function difference $H(x+u) - H(x)$ up to some *error* proportional to $\|u\|$ (or probably even superlinear to $\|u\|$ when we impose some assumptions on η). Finally, assumption (G3) suggests some *uniform continuity* away from the zero points of H, which is obviously fulfilled in the case H is Lipschitzian. Moreover, this assumption makes sense in the case H is *not* Lipschitzian at zero points.

We now present the damped Newton's method, also known as Newton's method with *line-search*. The algorithm was first introduced in [52], and later studied in [54, 56]. To start, we define the *region of convergence* Ω by

$$\Omega := \left\{ x \in \mathbb{R}^n \;\middle|\; \|H(x)\| < \gamma^{-1}\left(\tfrac{1-M\bar{\eta}}{M}\right) \right\},$$

where we mean $\gamma^{-1}(a) = \sup\{z|\gamma(z) = a\}$ as convention. The fact that Ω might be open, however, does not interfere our analysis in the sequel. In our argument, we will always assume all the *level set* is *bounded* and hence Ω is bounded. We also choose some "*slope*"-parameter $\sigma \in (0, 1)$. The parameter σ will play some role in the convergence of the algorithm. Indeed, for each starting point $x^0 \in \Omega$, we will choose a suitable σ such that the algorithm is executable.

Algorithm 6.1 (Generalized Damped Newton's method). *Let $\beta \in (0, 1)$ be a given scalar.*

(S.0) Choose a starting point $x^0 \in \Omega$.

(S.1) Check a suitable termination criterion.

(S.2) Compute $d^k \in \mathbb{R}^n$ *such that* (5.1) *holds, i.e.,*

$$-H(x^k) \in DH(x^k)(d^k),$$

and $\|d^k\| \le M\|H(x^k)\|$.

(S.3) Let $\alpha_k = \beta^{m_k}$ *where* m_k *is the first nonnegative integer* m *for which*

$$\frac{q(x^k + \beta^m d^k) - q(x^k)}{\beta^m} \le -\sigma q(x^k) \tag{6.1}$$

(S.4) Set $x^{k+1} = x^k + \alpha_k d^k$, $k \leftarrow k+1$, *and go to (S.1).*

The crucial matter in Newton's method is that solving the Newton equation (5.1) must be easier than solving the original equation (1.3), otherwise the Newton's method would be useless. Therefore, the solvability of the Newton equation should be taken into account. Let us mention that Proposition 5.2 provides a result of solvability based on *metric regularity*. In addition, the proof of Proposition 5.2 shows that the solution d of (5.2) also satisfies $\|d\| \le \mu\|H(x)\|$. Thus, it implies that Step (S.2) is always accomplished if we set M to be any constant larger than the metric regularity modulus of H.

Our next task is to verify that direction d in Step (S.2) produces a *descent direction*, which consequently implies the realization of Step (S.3).

Lemma 6.2 (descent direction). *Let (G1) hold and assume that* $d \neq 0$ *is taken from Step (S.2), i.e.,* d *is a solution to*

$$-H(x) \in DH(x)(d)$$

with $\|d\| \le M\|H(x)\|$. *The following hold* (*i*) *If* $x \in \Omega$ *and* $H(x) \neq 0$ *then* d *is a descent direction of* q *at* x.

(*ii*) *If the parameter* $\sigma < 1 - M\gamma(\|H(x)\|)$, *then step (S.3) is realized.*

Proof. By assumption we have $-\frac{H(x)}{\|d\|} \in DH(x)\big(\frac{d}{\|d\|}\big)$. Using (G1) we have

$$DH(x)(\tfrac{d}{\|d\|}) \subset -\tfrac{H(x)}{\|d\|} + \gamma I\!B.$$

Since $\|d\| \leq M\|H(x)\|$, it follows that

$$DH(x)(d) \subset -H(x) + M\|H(x)\|\gamma I\!B.$$

Since $H(x)$ is finite, one has $DH(x)(d)$ is bounded. It follows that

$$\operatorname*{Lim\,sup}_{\lambda \to 0} \left\| \frac{H(x+\lambda d) - H(x)}{\lambda} \right\| < +\infty. \tag{6.2}$$

We define the distance

$$r_\lambda := r(\lambda) = \operatorname{dist}\left(\frac{H(x+\lambda d) - H(x)}{\lambda}; DH(x)(d)\right),$$

and that,

$$\frac{H(x+\lambda d) - H(x)}{\lambda} \in DH(x)(d) + r_\lambda I\!B.$$

We claim that $r_\lambda \downarrow 0$ as $\lambda \downarrow 0$. Indeed, suppose it is not the case and there is $\delta > 0$, $\lambda_k \downarrow 0$ such that $r_{\lambda_k} > \delta > 0$, from (6.2) we may assume that

$$\frac{H(x+\lambda_k d) - H(x)}{\lambda_k} \to v \in DH(x)(d).$$

This is a contradiction to $r_{\lambda_k} > \delta$ and verifies our claim. Now one has

$$\begin{aligned}\frac{q(x+\lambda d)-q(x)}{\lambda} &= \frac{1}{2}\Big[H(x+\lambda d)+H(x)\Big]^T\Big[\frac{H(x+\lambda d)-H(x)}{\lambda}\Big]\\ &\subset \frac{1}{2}\Big[H(x+\lambda d)+H(x)\Big]^T\Big[DH(x)(d)+r_\lambda I\!B\Big]\\ &\subset \frac{1}{2}\Big[H(x+\lambda d)+H(x)\Big]^T\Big[-H(x)+M\|H(x)\|\gamma I\!B+r_\lambda I\!B\Big]\\ &\subset -\frac{1}{2}\Big[H(x+\lambda d)+H(x)\Big]^T H(x)+\\ &\quad+\frac{1}{2}\Big[H(x+\lambda d)+H(x)\Big]^T M\|H(x)\|\gamma I\!B+\frac{r_\lambda}{2}\Big[H(x+\lambda d)+H(x)\Big]^T I\!B\end{aligned}$$

That implies

$$\begin{aligned}\frac{q(x+\lambda d)-q(x)}{\lambda} &\le -\frac{1}{2}\Big[H(x+\lambda d)+H(x)\Big]^T H(x)+\\ &\quad+\frac{1}{2}\|H(x+\lambda d)+H(x)\|.M\|H(x)\|\gamma+\frac{r_\lambda}{2}\|H(x+\lambda d)+H(x)\|\end{aligned}$$

Letting $\lambda \downarrow 0$, we arrive

$$\begin{aligned}\limsup_{\lambda\downarrow 0}\frac{q(x+\lambda d)-q(x)}{\lambda} &\le -\|H(x)\|^2+\|H(x)\|^2 M\gamma\\ &= \Big(-1+M\gamma\Big)\|H(x)\|^2 \le 2\Big(-1+M\gamma\Big)q(x).\end{aligned}$$

The last inequality follows from $\gamma = \gamma(\|H(x)\|) \le \frac{1-M\bar{\eta}}{M}$ due to the fact that $x \in \Omega$. This guarantees d is a descent direction of q at x. Moreover, since $\sigma \le 1-\gamma M$, one has for m large

$$\frac{q(x+\beta^m d)-q(x)}{\beta^m} < -\sigma q(x).$$

This verifies Step (S.3) in Algorithm 6.1 and completes the proof. □

Note that when the graphical derivative is *singleton*, i.e., H is directionally differentiable, one has $\gamma \equiv 0$ and Lemma 6.2 goes back to Pang [52, Lemma 1].

Theorem 6.3 (Algorithm executability). *Assume that (G1) holds and H is metrically regular with modulus $\mu \leq M$. For a starting point $x^0 \in \Omega$, we choose the parameter*

$$0 < \sigma < 1 - M\gamma(\|H(x^0)\|).$$

Then Algorithm 6.1 generates a sequence $\{x^k\}$ satisfying $\|H(x^{k+1})\| < \|H(x^k)\|$ for all k, hence $\{x_k\}$ remains in Ω.

Proof. Starting at x^0, by Lemma 6.2 we have d^0 is a descent direction and find α_0 by step (S.3) such that

$$\frac{q(x^0 + \alpha_0 d^0) - q(x^0)}{\alpha_0} \leq -\sigma q(x^0).$$

Hence $q(x^1) < (1 - \sigma)q(x^0)$ with $x^1 = x^0 + \alpha_0 d^0$, i.e., $\|H(x^1)\| < \|H(x^0)\|$. That implies $x^1 \in \Omega$ and that

$$\sigma < 1 - M\gamma(\|H(x^0)\|) < 1 - M\gamma(\|H(x^1)\|),$$

which in turn implies the procedure is executable at x^1 by Lemma 6.2. The proof is then complete by using induction. □

6.2 Convergence Analysis

In this section, we justify the convergence of Algorithm 6.1 under the assumptions made. The following results present our convergence analysis. Let us mention that Theorem 6.4 is modified from Pang [52].

Theorem 6.4 (Preliminary convergence analysis). *Assume (G1) holds and Algorithm 6.1 is executable, in particular, H is metrically regular on Ω with some modulus $\mu \leq M$. Let $\{x^k\}$ with the corresponding α_k be the sequence generated by Algorithm 6.1. Assume further that*

$H(x^k) \neq 0$ for all k and that $\limsup_k \alpha_k > 0$. Then $\{x_k\}$ converges to a zero of H.

Proof. The sequence $\{q(x^k)\}$ is nonnegative and decreasing, thus it converges and

$$\lim_{k\to\infty} \big(q(x^k) - q(x^{k+1})\big) = 0.$$

Due to (6.1) one has

$$\lim_{k\to\infty} \alpha_k q(x^k) = 0.$$

By extracting a subsequence, we assume that $\lim\limits_{l\to\infty} \alpha_{k_l} = C > 0$ then $\lim\limits_{l\to\infty} q(x^{k_l}) = 0$. Since $\{q(x^k)\}$ is decreasing, it implies the whole sequence $\{q(x^k)\}$ decreases to zero as $k \to \infty$.

From (6.1), we have

$$\sqrt{q(x^k)} - \sqrt{q(x^{k+1})} \geq \frac{\sigma \alpha_k q(x^k)}{\sqrt{q(x^k)} + \sqrt{q(x^{k+1})}} \geq \frac{\sigma \alpha_k}{2} \sqrt{q(x^k)},$$

which implies

$$\|H(x^k)\| - \|H(x^{k+1})\| \geq \frac{\sigma \alpha_k}{2} \|H(x^k)\|.$$

Now one has

$$\|x^{k+1} - x^k\| = \alpha_k \|d^k\| \leq M\alpha_k \|H(x^k)\| \leq \frac{2M}{\sigma}\Big(\|H(x^k)\| - \|H(x^{k+1})\|\Big).$$

Inductively, one has for all p that

$$\|x^{k+p} - x^k\| \leq \frac{2M}{\sigma}\Big(\|H(x^k)\| - \|H(x^{k+p})\|\Big) \leq \frac{2M}{\sigma}\|H(x^k)\| \to 0 \quad \text{as } k \to \infty.$$

This verifies $\{x^k\}$ is a Cauchy sequence, thus converges to some $\bar{x}$ which is obviously a zero of

H. The proof is now complete. □

Theorem 6.5 (Convergence analysis). *Assume (G1)-(G3) hold and Algorithm 6.1 is executable, in particular, H is metrically regular in Ω with some modulus $\mu \leq M$. Then for any starting point $x^0 \in \Omega$, there exists $\sigma > 0$ for Algorithm 6.1 such that the generated sequence $\{x^k\}$ converges to a zero of H.*

Proof. Since $x^0 \in \Omega$, one has $\|H(x^0)\| < \gamma^{-1}\left(\frac{1-M\bar{\eta}}{M}\right)$. Therefore, we can choose σ such that

$$\|H(x^0)\| < \gamma^{-1}\left(\tfrac{1-\sigma-M\bar{\eta}}{M}\right),$$

which means

$$\sigma < 1 - M\bar{\eta} - M\gamma(\|H(x^0)\|).$$

The choice of σ guarantees Algorithm 6.1 is executable by Theorem 6.3. We denote $\{x^k\}$ the sequence generated by Algorithm 6.1 with our choice of σ. Then it is clear from Lemma 6.2 and Theorem 6.4 that $\|H(x^k)\|$ is decreasing.

Suppose $\limsup_{k\to\infty} \alpha_k > 0$ then the conclusion follows from Theorem 6.4 and we finish the proof. So our remaining task is to prove the remaining case when $\lim_{k\to\infty} \alpha_k = 0$.

To furnish, we first claim that $\|H(x^k)\| \to 0$. Suppose by contrary that there is $r > 0$ such that $\|H(x^k)\| > r$ for all k. By the choice of α_k, the scalar $\frac{\alpha_k}{\beta}$ does not satisfy (6.1), i.e.,

$$\Delta_k := \frac{q(x^k + \frac{\alpha_k}{\beta}d^k) - q(x^k)}{\frac{\alpha_k}{\beta}} > -\sigma q(x^k) = -\frac{\sigma}{2}\|H(x^k)\|^2. \tag{6.3}$$

Employing (G2) with $u = \lambda d$, we find $\delta > 0$ such that

$$\frac{H(x+\lambda d) - H(x)}{\lambda} \in DH(x)(d) + \bar{\eta}\|d\|I\!B$$

whenever $\|\lambda d\| \le \delta$ and $x \in \Omega$. Now employing (G3), take $\varepsilon > 0$ small we shrink δ such that

$$\|H(x + \lambda d) - H(x)\| \le \varepsilon \quad \text{whenever } \|\lambda d\| \le \delta.$$

Now since $C := \|H(x^0)\| \ge \|H(x^k)\| \ge r$, $\alpha_k \downarrow 0$ and $\|d^k\| \le M\|H(x^k)\|$, we have

$$\|\tfrac{\alpha_k}{\beta} d^k\| \le \delta \quad \text{for all large } k.$$

Therefore, (G2) implies for large k that

$$\frac{H(x^k + \frac{\alpha_k}{\beta} d^k) - H(x^k)}{\frac{\alpha_k}{\beta}} \in DH(x^k)(d^k) + \bar{\eta}\|d^k\| I\!B.$$

Similar to Lemma 6.2, we have

$$DH(x^k)(d^k) + \eta\|d^k\| I\!B \subset -H(x^k) + \|H(x^k)\|.M\gamma_k I\!B + M\bar{\eta}\|H(x^k)\| I\!B,$$

where we denote $\gamma_k = \gamma(\|H(x^k)\|)$ for brevity in notation. Define $2u := H(x^k + \frac{\alpha_k}{\beta} d^k) - H(x^k)$, then due to (G3) we have $\|2u\| \le \varepsilon$, and thus

$$\begin{aligned} \Delta_k = \frac{q(x^k + \frac{\alpha_k}{\beta} d^k) - q(x^k)}{\frac{\alpha_k}{\beta}} &= \frac{1}{2}\Big[H(x^k + \tfrac{\alpha_k}{\beta} d^k) + H(x^k)\Big]^T \Big[\frac{H(x^k + \frac{\alpha_k}{\beta} d^k) - H(x^k)}{\frac{\alpha_k}{\beta}}\Big] \\ &\in \Big[H(x^k) + u\Big]^T \Big[- H(x^k) + \|H(x^k)\|.M\gamma_k I\!B + M\bar{\eta}\|H(x^k)\| I\!B\Big] \end{aligned}$$

So we have the estimate

$$\Delta_k \le -\|H(x^k)\|^2 + \|H(x^k)\|^2(M\gamma_k + M\bar{\eta}) + \frac{\varepsilon}{2}\|H(x^k)\|\Big(1 + M\gamma_k + M\bar{\eta}\Big).$$

Choose ε small such that the second term on the right hand side is *less* than $\frac{\sigma}{4}r^2$, then

$$\Delta_k \leq \|H(x^k)\|^2\Big(-1 + M\gamma_k + M\bar{\eta}\Big) + \frac{\sigma}{4}r^2 \leq \|H(x^k)\|^2\Big(-1 + M\gamma_k + M\bar{\eta} + \frac{\sigma}{4}\Big).$$

Combining this with the estimate (6.3) we have

$$-\frac{\sigma}{2} \leq -1 + M\gamma_k + M\bar{\eta} + \frac{\sigma}{4} \tag{6.4}$$

Due to $\|H(x^k)\| < \|H(x^0)\| < \gamma^{-1}\left(\frac{1-\sigma-\bar{\eta}M}{M}\right)$, we have $\gamma_k = \gamma(\|H(x^k\|) \leq \frac{1-\sigma-\bar{\eta}M}{M}$. Thus from (6.4),

$$-\frac{\sigma}{2} \leq -1 + (1 - \sigma - \bar{\eta}M) + M\bar{\eta} + \frac{\sigma}{4} = -\frac{3\sigma}{4},$$

which is a contradiction. This verifies our claim that $\|H(x^k)\| \to 0$.

Finally, using the same argument in Theorem 6.4, we conclude that $\{x^k\}$ converges to some $\bar{x}$ which is a zero of H. The proof is now complete. □

Remark 6.6 Comparing to [52], Theorem 6.4 and 6.5 show one special feature that the sequence of iterates $\{x^k\}$ converges *solely* to a single zero of H. Of course, the original equation might have more than one solution.

In the rest of this section, we provide some result on the convergence rate of the algorithm. Suppose we can choose $\alpha_k = 1$ for all k large in Step (S.3) of Algorithm 6.1. In this case, the damped Newton's method becomes the Newton's method in classical sense, which is also known as *pure* Newton's method. Generally, the pure Newton's method is very appealing due to its *superlinear* convergence under mild assumptions.

Theorem 6.7 (pure Newton's method). *Assume (G1), (G2), (G3) hold and Algorithm 6.1* *is executable, in particular, H is metrically regular in Ω with some modulus $\mu \le M$. Assume further that*

$$\bar{\eta} = \limsup_{\substack{\|H(x)\|\to 0 \\ \|u\|\to 0}} \eta(x,u) < \tfrac{\sqrt{2}-1}{M}.$$

Then:

(i) *For any starting point $x^0 \in \Omega$, there exists $\sigma > 0$ for Algorithm 6.1 such that the generated sequence $\{x^k\}$ converges to a zero of H.*

(ii) *Algorithm 6.1 eventually becomes the pure Newton's method, i.e., $\alpha_k = 1$ for k large, or equivalently, for all k large the iterations are given by*

$$x^{k+1} = x^k + d^k \quad \text{with } d^k \text{ solves } \; -H(x^k) \in DH(x^k)(d^k).$$

(iii) *There exists a constant $C > 0$ such that one has the error estimate*

$$\|x^k - \bar{x}\| \le C(1-\sigma)^{k/2} \quad \text{for all } k \text{ large.} \tag{6.5}$$

where $\bar{x}$ is a zero of H.

Proof. Let $\tilde{\eta} := \limsup_{\|H(x)\|,\|u\|\downarrow 0} \eta(x,u)$ and choose the parameter σ such that

$$0 < \sigma < \min\left\{4 - 2(1+\tilde{\eta}M)^2, 1 - M\bar{\eta} - M\gamma(\|H(x^0)\|)\right\}.$$

Let $\{x^k\}$ be the iterations generated by Algorithm 6.1. It is clear that all conclusions in Theorem 6.5 hold, which verifies (i). To prove (ii), we show that eventually $\alpha^k = 1$ for all k large.

Since $\|H(x^k)\| \to 0$, one has $\|d^k\| \le M\|H(x^k)\| \to 0$ and $\gamma_k = \gamma(\|H(x^k)\|) \to 0$ as well. For

k large one has by (G2) that

$$H(x^k + d^k) - H(x^k) \in DH(x^k)(d^k) + \eta_k \|d^k\| I\!\!B$$
$$\subset -H(x^k) + (M\gamma_k + M\eta_k)\|H(x^k)\| I\!\!B,$$

where $\eta_k = \eta(r_k)$ for $r_k = \sup_{\|v\| \le \|d^k\|} \|H(x^k + v)\|$. This implies

$$H(x^k + d^k) + H(x^k) \in H(x^k) + (M\gamma_k + M\eta_k)\|H(x^k)\| I\!\!B.$$

One has

$$2\Big[q(x^k + d^k) - q(x^k)\Big] = \Big[H(x^k + d^k) + H(x^k)\Big]^T \Big[H(x^k + d^k) - H(x^k)\Big]$$
$$\subset \Big[H(x^k) + (M\gamma_k + M\eta_k)\|H(x^k)\| I\!\!B\Big]^T \Big[-H(x^k) + (M\gamma_k + M\eta_k)\|H(x^k)\| I\!\!B\Big].$$

So

$$2\Big[q(x^k + d^k) - q(x^k)\Big] \le -\|H(x^k)\|^2 + 2\|H(x^k)\|^2(M\gamma_k + M\eta_k) + \|H(x^k)\|^2(M\gamma_k + M\eta_k)^2$$
$$= \|H(x^k)\|^2\Big(-1 + 2(M\gamma_k + M\eta_k) + (M\gamma_k + M\eta_k)^2\Big)$$
$$= q(x^k)\Big(-4 + 2(1 + M\gamma_k + M\eta_k)^2\Big).$$

Take $k \to \infty$, then $x^k \to \bar{x}$ and $\|d^k\|, \|H(x^k)\| \to 0$, so $r_k \to 0$ by the continuity of H. This implies $\gamma_k \to 0$ and $\eta_k \to \tilde{\eta}$. It follows that for k sufficiently large

$$-4 + 2(1 + M\gamma_k + M\eta_k)^2 < -\sigma.$$

Hence one has

$$q(x^k + d^k) - q(x^k) \le -\sigma q(x^k). \tag{6.6}$$

This implies (6.1) in step (S.3) of Algorithm 6.1 is satisfied with $\alpha_k = 1$, for all k large.

It remains to prove (iii). From the argument in Theorem 6.4, one has

$$\|x^k - x^{k+p}\| \leq \tfrac{2M}{\sigma}\|H(x^k)\|,$$

Let $p \to \infty$, one has

$$\|x^k - \bar{x}\| \leq \tfrac{2M}{\sigma}\|H(x^k)\| = C\sqrt{q(x^k)}.$$

From (6.6) one has

$$q(x^k) \leq (1-\sigma)q(x^{k-1}) \quad \text{for } k \text{ large.}$$

Combining the last two relations, we verify (iii). The proof is now complete. □

Theorem 6.8 (pure Newton's method with superlinear convergence). *Assume (G1), (G2) and (G3) hold and Algorithm 6.1 is executable, in particular, H is metrically regular in Ω with some modulus $\mu \leq M$. Assume further that*

$$\lim_{\substack{\|H(x)\|\to 0\\ \|u\|\to 0}} \eta(x,u) = 0.$$

Then all conclusions of Theorem 6.7 hold. Moreover, the rate of convergence is superlinear.

Proof. It is obvious that all conclusions in Theorem 6.7 hold. It remains to prove that the convergence rate is superlinear. Let $\{x^k\}$ be the iterations generated by Algorithm 6.1 which converges to a zero $\bar{x}$ of H. It follows from the proof of Theorem 6.7 that

$$\|x^{k+1} - \bar{x}\| \leq \tfrac{2M}{\sigma}\|H(x^{k+1})\| \quad \text{for all } k.$$

Using $-H(x^k) \in DH(x^k)(x^{k+1} - x^k)$ for k large, we continue the estimate as follows

$$\begin{aligned}
\|H(x^{k+1})\| &= \|H(x^{k+1}) - H(x^k) + H(x^k)\| \\
&\leq \operatorname{dist}\Big(H(x^{k+1}) - H(x^k); DH(x^k)(x^{k+1} - x^k)\Big) + \operatorname{diam} DH(x^k)(x^{k+1} - x^k) \\
&\leq \eta_k \|x^{k+1} - x^k\| + \gamma_k \|x^{k+1} - x^k\| \\
&= (\eta_k + \gamma_k)\|x^{k+1} - x^k\|,
\end{aligned}$$

where $\gamma_k = \gamma(\|H(x^k)\|)$ and $\eta_k = \eta(x^k, d^k)$. Since $\|H(x^k)\|, \|d^k\| \downarrow 0$, we have that both $\gamma_k, \eta_k \to 0$.

Now for any small $\varepsilon > 0$, with k sufficiently large one has

$$\|x^{k+1} - \bar{x}\| \leq \varepsilon \|x^{k+1} - x^k\| \leq \varepsilon \|x^{k+1} - \bar{x}\| + \varepsilon \|x^k - \bar{x}\|.$$

It follows that

$$\|x^{k+1} - \bar{x}\| \leq \frac{\varepsilon}{1 - \varepsilon}\|x^k - \bar{x}\| \leq 2\varepsilon \|x^k - \bar{x}\|,$$

which implies $\|x^{k+1} - \bar{x}\| = o(\|x^k - \bar{x}\|)$, i.e., the convergence rate is superlinear. □

Remark 6.9 Under the assumptions in Theorem 6.8, one can consider the damped Newton's method as a "hybrid" algorithm. That means it has two phases: the first phase is applied for *global* convergence as we need to be sufficiently closed to the solution, while the second phase is the pure Newton's method which will converge *superlinearly.*

Theorem 6.10 (pure Newton's method with locally superlinear convergence). *Assume (G1), (G2) and (G3) hold on some region Ω' and Algorithm 6.1 generates a sequence $\{x^k\}$*

converges to an isolated zero $\bar{x} \in \Omega'$ *of* H*. Assume further that*

$$\lim_{\substack{x \to \bar{x} \\ \|u\| \downarrow 0}} \eta(x, u) = 0.$$

Then $\{x^k\}$ *becomes the pure Newton's iterations in some neighborhood* U *of* $\bar{x}$ *and the rate of convergence is superlinear.*

Proof. Pick some neighborhood U of $\bar{x}$ such that

$$\eta(x, u) < \tfrac{\sqrt{2}-1}{M} \quad \text{for all} \quad x, x + u \in U.$$

Since $\{x^k\}$ converges to $\bar{x}$, we may assume $x^0 \in U$. Then the rest of the proof is similar to Theorem 6.7 and Theorem 6.8. $\square$

6.3 Discussion on Major Assumptions

In this section, we discuss some sufficient conditions for our major assumptions (G1), (G2), and (G3). Notice that assumption (G1) is automatic when H is *directionally differentiable.* We will also use the notions in Section 5.2 and their relationships.

In the next two lemmas, we provide sufficient condition for assumption (G2) and (G3).

Proposition 6.11 (assumptions (G2) and (G3) for Lipschitz functions). *Let* $H : \mathbb{R}^n \to \mathbb{R}^n$ *be Lipschitz on some closed bounded set* $\Omega \subset \mathbb{R}^n$*, hence (G3) holds on* Ω *automatically. The following hold*

(*i*) *Assume for some* $\eta > 0$ *one has* $\eta M < 1$ *and*

$$\operatorname{diam}\left[\partial_C H(x)(d)\right] < \eta \quad \text{for all } x \in \Omega, \|d\| \leq 1. \tag{6.7}$$

Then (G2) holds on Ω *with* $\eta(\cdot) \equiv \eta$.

(*ii*) *Assume for some* $\eta > 0$ *one has* $\eta M < \sqrt{2} - 1$ *satisfying* (6.7). *Then (G2) holds on* Ω *and the damped Newton's method eventually becomes pure Newton's method.*

Proof. By contrary assume there exist $x_k \in \Omega$, $\|\lambda_k d_k\| \downarrow 0$ such that (G2) does not hold, i.e.,

$$w_k := \frac{H(x_k + \lambda_k d_k) - H(x_k)}{\lambda_k} \notin DH(x_k)(d_k) + \eta\|d_k\|\mathbb{B},$$

Hence $\|d_k\| > 0$, we divide by $\|d_k\|$ and arrive

$$\frac{H(x_k + \lambda'_k d'_k) - H(x_k)}{\lambda'_k} \notin DH(x_k)(d'_k) + \eta\mathbb{B},$$

where $\lambda'_k := \lambda_k\|d_k\| \downarrow 0$ and $d'_k := \frac{d_k}{\|d_k\|}$. So we may assume $\lambda_k \downarrow 0$ and $\|d_k\| = 1$. To proceed, we pick a sequence $v_k \in DH(x_k)(d_k) \subset \partial_C H(x_k)(d_k)$ such that

$$\|w_k - v_k\| > \eta.$$

Due to Lipschitz property and Ω is bounded, by taking subsequences we can assume that $x_k \to \bar{x} \in \Omega$, $d_k \to d$ for some $\|d\| = 1$ and

$$w_k \to \xi \in D_T H(\bar{x})(d) \subset \partial_C H(\bar{x})(d),$$

$$v_k \to \tilde{v} \in \partial_C H(\bar{x})(d).$$

Combining all we have

$$\eta \le \|\xi - \tilde{v}\| \le \operatorname{diam} \partial_C H(\bar{x})(d) < \eta,$$

which is a contradiction. Hence the proof is complete. □

Corollary 6.12 (Damped Newton's method for Lipschitz functions). *Let $H : \mathbb{R}^n \to \mathbb{R}^n$ be a Lipschitz function satisfying (G1) and assume that H is metrically regular on Ω with some modulus $\mu \le M$. Assume further that there is $\eta > 0$ with $\eta M < 1$ such that*

$$\operatorname{diam}\left[\partial_C H(x)(d)\right] < \eta \quad \textit{for all } \|d\| \le 1, x \in \mathbb{R}^n.$$

Then for any starting point $\|H(x^0)\| < \gamma^{-1}\left(\frac{1-\eta M}{M}\right)$, there is $\sigma > 0$ such that Algorithm 6.1 generates a sequence $\{x^k\}$ converges to a zero $\bar{x}$ of H. Moreover if η satisfies $\eta M < \sqrt{2} - 1$ then there exist an algorithm parameter σ and $C > 0$ such that the error estimate (6.5) *holds.*

Proof. Using Lemma 6.11, then assumption (G2) and (G3) are also satisfied. Thus the conclusions follow from Theorem 6.4 and Theorem 6.5, and Theorem 6.7. □

Corollary 6.13 (Damped Newton's method for directional differentiability functions). *Let $H : \mathbb{R}^n \to \mathbb{R}^n$ be a directionally differentiable Lipschitz function and assume that H is metrically regular on Ω with some modulus $\mu \le M$. Assume further that there is $\eta > 0$ with $\eta M < 1$ such that*

$$\operatorname{diam}\left[\partial_C H(x)(d)\right] < \eta \quad \textit{for all } \|d\| \le 1, x \in \mathbb{R}^n.$$

Then for any starting point x^0, there is $\sigma > 0$ such that Algorithm 6.1 generates a sequence $\{x^k\}$ converges to a zero $\bar{x}$ of H. Moreover if η satisfies $\eta M < \sqrt{2} - 1$ then there exist an algorithm parameter σ and $C > 0$ such that the error estimate (6.5) *holds.*

Proof. We first check all assumptions (G1),(G2) and (G3). Obviously, assumption (G3) holds due to Lipschitz property. Since H is directionally differentiable, the graphical derivative is a singleton and coincides with the directionally derivative:

$$DH(x)(d) = H'(x,d) \quad \text{for all } x, d \in \mathbb{R}^n.$$

So (G1) is satisfied with $\gamma \equiv 0$. We now choose σ such that

$$1 - \sigma - \eta M > 0.$$

For any starting point x^0, define the level set

$$\Omega := \left\{ x \,\middle|\, \|H(x)\| \leq \|H(x^0)\| \right\}$$

which is bounded by our standing assumption. Finally assumption (G2) holds on Ω due to by Proposition 6.11. Hence, we can now proceed similar to Theorem 6.5 and Theorem 6.7 and derive the convergence result. □

6.4 Future Development

In this chapter we develop a new generalized damped Newton's method for solving systems of nonsmooth equations $H(x) = 0$. Several global convergence results are derived under relatively mild conditions. Similar to Chapter 5, variational analysis and generalized differentiation play a fundamental role in our study. The algorithm is also built based on the graphical/contingent derivatives. The metric regularity and its pointwise coderivative characterization play a crucial role in the justification of the algorithm and its convergence. Besides, we also give some results on its convergence rate, which occur under some special situations.

The damped Newton's method has its own advantage since it generally provides global convergence results, which are extremely important in applications. Our next development will concentrate on its applications to problems with complicated structures, e.g., variational inequalities, nonlinear complementarity problems, etc. On the other hand, we also continue to study its performance comparing with other well-known methods. The details of these ideas are part of our future research.

REFERENCES

[1] F. J. Aragón Artacho and M. H. Geoffroy, *Uniformity and inexact version of a proximal method for metrically regular mappings*, J. Math. Anal. Appl., 335 (2007), pp. 168–183.

[2] J.-P. Aubin, *Contingent derivatives of set-valued maps and existence of solutions to non-linear inclusions and differential inclusions*, in Mathematical analysis and applications, Academic Press, New York, 1981, pp. 159–229.

[3] J.-P. Aubin and H. Frankowska, *Set-valued analysis*, Birkhäuser, Boston, 1990.

[4] T. Q. Bao and B. S. Mordukhovich, *Relative Pareto minimizers for multiobjective problems: existence and optimality conditions*, Math. Program., 122 (2010), pp. 301–347.

[5] T. Q. Bao and B. S. Mordukhovich, *Set-valued optimization in welfare economics*, Adv. Math. Econ., 13 (2010), pp. 113–153.

[6] H. H. Bauschke, J. M. Borwein, and W. Li, *Strong conical hull intersection property, bounded linear regularity, Jameson's property (G), and error bounds in convex optimization*, Math. Program., 86 (1999), pp. 135–160.

[7] J. M. Borwein and Q. J. Zhu, *Techniques of variational analysis*, Springer-Verlag, New York, 2005.

[8] J. V. Burke, A. S. Lewis, and M. L. Overton, *Variational analysis of functions of the roots of polynomials*, Math. Program., 104 (2005), pp. 263–292.

[9] M. J. Cánovas, M. A. López, B. S. Mordukhovich, and J. Parra, *Variational analysis in semi-infinite and infinite programming. I. Stability of linear inequality systems of feasible solutions*, SIAM J. Optim., 20 (2009), pp. 1504–1526.

[10] M. J. Cánovas, M. A. López, B. S. Mordukhovich, and J. Parra, *Variational analysis in semi-infinite and infinite programming, II: necessary optimality conditions*, SIAM J. Optim., 20 (2010), pp. 2788–2806.

[11] C. K. Chui, F. Deutsch, and J. D. Ward, *Constrained best approximation in Hilbert space*, Constr. Approx., 6 (1990), pp. 35–64.

[12] F. H. Clarke, *Optimization and nonsmooth analysis*, John Wiley & Sons Inc., New York, 1983. A Wiley-Interscience Publication.

[13] F. H. Clarke, Y. S. Ledyaev, R. J. Stern, and P. R. Wolenski, *Nonsmooth analysis and control theory*, Springer-Verlag, New York, 1998.

[14] T. De Luca, F. Facchinei, and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Math. Programming, 75 (1996), pp. 407–439.

[15] J. E. Dennis, Jr. and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, Prentice Hall Inc., Englewood Cliffs, NJ, 1983.

[16] F. Deutsch, W. Li, and J. D. Ward, *A dual approach to constrained interpolation from a convex subset of Hilbert space*, J. Approx. Theory, 90 (1997), pp. 385–414.

[17] N. Dinh, B. S. Mordukhovich, and T. T. A. Nghia, *Qualification and optimality conditions for DC programs with infinite constraints*, Acta Math. Vietnam., 34 (2009), pp. 125–155.

[18] N. Dinh, B. S. Mordukhovich, and T. T. A. Nghia, *Subdifferentials of value functions and optimality conditions for DC and bilevel infinite and semi-infinite programs*, Math. Program., 123 (2010), pp. 101–138.

[19] E. Ernst and M. Théra, *Boundary half-strips and the strong chip*, SIAM J. Optim., 18 (2007), pp. 834–852 (electronic).

[20] M. Fabian and B. S. Mordukhovich, *Separable reduction and extremal principles in variational analysis*, Nonlinear Anal., 49 (2002), pp. 265–292.

[21] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems. Vol. I and II*, Springer-Verlag, New York, 2003.

[22] D. H. Fang, C. Li, and K. F. Ng, *Constraint qualifications for optimality conditions and total Lagrange dualities in convex infinite programming*, Nonlinear Anal., 73 (2010), pp. 1143–1159.

[23] M. A. Goberna, V. Jeyakumar, and M. A. López, *Necessary and sufficient constraint qualifications for solvability of systems of infinite convex inequalities*, Nonlinear Anal., 68 (2008), pp. 1184–1194.

[24] M. A. Goberna and M. A. López, *Linear semi-infinite optimization*, John Wiley & Sons Ltd., Chichester, 1998.

[25] A. Göpfert, H. Riahi, C. Tammer, and C. Zălinescu, *Variational methods in partially ordered spaces*, Springer-Verlag, New York, 2003.

[26] S.-P. Han, J.-S. Pang, and N. Rangaraj, *Globally convergent Newton methods for nonsmooth equations*, Math. Oper. Res., 17 (1992), pp. 586–607.

[27] A. D. Ioffe, *Metric regularity and subdifferential calculus*, Uspekhi Mat. Nauk, 55 (2000), pp. 103–162.

[28] J. Jahn, *Vector optimization*, Springer-Verlag, Berlin, 2004. Theory, applications, and extensions.

[29] N. H. Josephy, *Newton's method for generalized equations and the PIES energy model*, Ph.D Dissertation, Department of Industrial Engineering, University of Wisconsin-Madison, (1979).

[30] L. V. Kantorovich and G. P. Akilov, *Functional analysis in normed spaces*, The Macmillan Co., New York, 1964.

[31] C. Kanzow, I. Ferenczi, and M. Fukushima, *On the local convergence of semismooth Newton methods for linear and nonlinear second-order cone programs without strict complementarity*, SIAM J. Optim., 20 (2009), pp. 297–320.

[32] C. T. Kelley, *Solving nonlinear equations with Newton's method*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2003.

[33] D. Klatte and B. Kummer, *Nonsmooth equations in optimization*, Kluwer Academic Publishers, Dordrecht, 2002. Regularity, calculus, methods and applications.

[34] A. Y. Kruger, *About regularity of collections of sets*, Set-Valued Anal., 14 (2006), pp. 187–206.

[35] A. Y. Kruger and B. S. Mordukhovich, *Extremal points and the Euler equation in nonsmooth optimization problems*, Dokl. Akad. Nauk BSSR, 24 (1980), pp. 684–687.

[36] B. Kummer, *Newton's method for nondifferentiable functions*, in Advances in mathematical optimization, Akademie-Verlag, Berlin, 1988, pp. 114–125.

[37] B. Kummer, *Lipschitzian inverse functions, directional derivatives, and applications in $C^{1,1}$ optimization*, J. Optim. Theory Appl., 70 (1991), pp. 561–581.

[38] A. S. Lewis, D. R. Luke, and J. Malick, *Local linear convergence for alternating and averaged nonconvex projections*, Found. Comput. Math., 9 (2009), pp. 485–513.

[39] C. Li and K. F. Ng, *Constraint qualification, the strong CHIP, and best approximation with convex constraints in Banach spaces*, SIAM J. Optim., 14 (2003), pp. 584–607.

[40] C. Li and K. F. Ng, *Strong CHIP for infinite system of closed convex sets in normed linear spaces*, SIAM J. Optim., 16 (2005), pp. 311–340.

[41] C. Li, K. F. Ng, and T. K. Pong, *The SECQ, linear regularity, and the strong CHIP for an infinite system of closed convex sets in normed linear spaces*, SIAM J. Optim., 18 (2007), pp. 643–665.

[42] W. Li, C. Nahak, and I. Singer, *Constraint qualifications for semi-infinite systems of convex inequalities*, SIAM J. Optim., 11 (2000), pp. 31–52 (electronic).

[43] Z.-Q. Luo, J.-S. Pang, and D. Ralph, *Mathematical programs with equilibrium constraints*, Cambridge University Press, Cambridge, 1996.

[44] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optimization, 15 (1977), pp. 959–972.

[45] B. S. Mordukhovich, *Maximum principle in the problem of time optimal response with nonsmooth constraints*, Prikl. Mat. Meh., 40 (1976), pp. 1014–1023.

[46] B. S. Mordukhovich, *Complete characterization of openness, metric regularity, and Lipschitzian properties of multifunctions*, Trans. Amer. Math. Soc., 340 (1993), pp. 1–35.

[47] B. S. Mordukhovich, *Variational analysis and generalized differentiation. I: Basic theory*, Springer-Verlag, Berlin, 2006.

[48] B. S. Mordukhovich, *Variational analysis and generalized differentiation. II: Applications*, Springer-Verlag, Berlin, 2006.

[49] B. S. Mordukhovich, J. Peña, and V. Roshchina, *Applying metric regularity to compute a condition measure of smoothing algorithm for matrix games*, SIAM J. Optim., to appear.

[50] B. S. Mordukhovich and B. Wang, *Extensions of generalized differential calculus in Asplund spaces*, J. Math. Anal. Appl., 272 (2002), pp. 164–186.

[51] J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, Academic Press, New York, 1970.

[52] J.-S. Pang, *Newton's method for B-differentiable equations*, Math. Oper. Res., 15 (1990), pp. 311–341.

[53] J.-S. Pang, *A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems*, Math. Programming, 51 (1991), pp. 101–131.

[54] J.-S. Pang and L. Q. Qi, *Nonsmooth equations: motivation and algorithms*, SIAM J. Optim., 3 (1993), pp. 443–465.

[55] R. Puente and V. N. Vera de Serio, *Locally Farkas-Minkowski linear inequality systems*, Top, 7 (1999), pp. 103–121.

[56] L. Q. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res., 18 (1993), pp. 227–244.

[57] L. Q. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Programming, 58 (1993), pp. 353–367.

[58] S. M. Robinson, *Local structure of feasible sets in nonlinear programming. III. Stability and sensitivity*, Math. Programming Stud., (1987), pp. 45–66, Nonlinear analysis and optimization (Louvain-la-Neuve, 1983).

[59] S. M. Robinson, *An implicit-function theorem for a class of nonsmooth functions*, Math. Oper. Res., 16 (1991), pp. 292–309.

[60] S. M. Robinson, *Newton's method for a class of nonsmooth functions*, Set-Valued Anal., 2 (1994), pp. 291–305, Set convergence in nonlinear analysis and optimization.

[61] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, Springer-Verlag, Berlin, 1998.

[62] W. Schirotzek, *Nonsmooth analysis*, Springer, Berlin, 2007.

[63] T. I. Seidman, *Normal cones to infinite intersections*, Nonlinear Anal., 72 (2010), pp. 3911–3917.

[64] A. Shapiro, *On concepts of directional differentiability*, J. Optim. Theory Appl., 66 (1990), pp. 477–487.

[65] W. Song and R. Zang, *Bounded linear regularity of convex sets in Banach spaces and its applications*, Math. Program., 106 (2006), pp. 59–79.

[66] N. Z. Šor, *The class of almost differentiable functions and a certain method of minimizing functions of that class*, Kibernetika (Kiev), (1972), pp. 65–70.

[67] L. Thibault, *Subdifferentials of compactly Lipschitzian vector-valued functions*, Ann. Mat. Pura Appl. (4), 125 (1980), pp. 157–192.

[68] R. Vinter, *Optimal control*, Birkhäuser, Boston, 2000.

[69] J. Warga, *Fat homeomorphisms and unbounded derivate containers*, J. Math. Anal. Appl., 81 (1981), pp. 545–560.

ABSTRACT

NEW VARIATIONAL PRINCIPLES WITH APPLICATIONS TO OPTIMIZATION THEORY AND ALGORITHMS

by

HUNG MINH PHAN

August 2011

Advisor: Dr. Boris S. Mordukhovich

Major: Mathematics (Applied)

Degree: Doctor of Philosophy

In this dissertation we investigate some applications of variational analysis in optimization theory and algorithms. In the first part we develop new extremal principles in variational analysis that deal with finite and infinite systems of convex and nonconvex sets. The results obtained, under the name of tangential extremal principles and rated extremal principles, combine primal and dual approaches to the study of variational systems being in fact first extremal principles applied to infinite systems of sets. These developments are in the core geometric theory of variational analysis. Our study includes the basic theory and applications to problems of semi-infinite programming and multiobjective optimization. The second part of this dissertation concerns developing numerical methods of the Newton-type to solve systems of nonlinear equations. We propose and justify a new generalized Newton algorithm based on graphical derivatives. Based on advanced tools of variational analysis and generalized differentiation, we establish the well-posedness and convergence results of the algorithm. Besides, we present a new generalized damped Newton algorithm, which is also known as Newton's method with line-search. Some global convergence results are also justified.

AUTOBIOGRAPHICAL STATEMENT

HUNG MINH PHAN

Education

- *Ph.D. in Mathematics*, Wayne State University (August 2011, expected)
- *B.S. in Mathematics*, Cantho University, Vietnam (2004)

Awards

- *Research grants DMS-0603846 and DMS-1007132 supported by the US National Science Foundation, Research grant DP-12092508 supported by the Australian Research Council (Prof. Boris S. Mordukhovich is the Principle Investigator)*
- *Outstanding Graduate Research Award* (2010, 2011), Department of Mathematics, Wayne State University
- *Best Poster Prize* (2008), International Workshop on Variational Analysis, Santiago, Chile
- *Andre G. Laurent Award for Outstanding Achievement in the Ph.D Program* (2011), *Maurice J. Zelonka Endowed Scholarship* (2011), *M. F. Janowitz Endowed Mathematics Scholarship* (2009), *Karl W. and Helen L. Folley Endowed Mathematics Scholarship* (2008), Department of Mathematics, Wayne State University
- *Thomas C. Rumble Fellowship* (2006-2007), Wayne State University
- *Graduate Student Professional Travel Award* (2007, 2008, 2009, 2010, 2011), College of Liberal Arts and Sciences, Wayne State University

Publications

1. B. S. Mordukhovich and H. M. Phan, *Tangential extremal principle for finite and infinite systems, I: Basic theory*, Math. Program., to appear.
2. B. S. Mordukhovich and H. M. Phan, *Tangential extremal principle for finite and infinite systems, II: Applications to semi-infinite and multiobjective optimization*, Math. Program., to appear.
3. B. S. Mordukhovich and H. M. Phan, *Rated extremal principle for finite and infinite systems with applications to optimization*, Optimization, to appear.
4. T. Hoheisel, C. Kanzow, B. S. Mordukhovich and H. M. Phan, *Generalized Newton's method based on graphical derivatives*, Nonlinear Anal., to appear.
5. B. S. Mordukhovich, N. M. Nam and H. M. Phan, *Variational analysis of marginal function with applications to bilevel programming problems*, J. Optim. Theory Appl., submitted.
6. B. S. Mordukhovich and H. M. Phan, *Generalized damped Newton's method for nonsmooth equations*, in preparation.