Boris S. Mordukhovich

Variational Analysis and Generalized Differentiation II

Applications

Springer

Boris S. Mordukhovich
Department of Mathematics
Wayne State University
College of Science
Detroit, MI 48202-9861, U.S.A.
E-mail: *boris@math.wayne.edu*

Library of Congress Control Number: 2005932550

Mathematics Subject Classification (2000): 49J40, 49J50, 49J52, 49K24, 49K27, 49K40, 49N40, 58C06, 54C06, 58C20, 58C25, 65K05, 65L12, 90C29, 90C31, 90C48, 39B35

ISSN 0072-7830
ISBN-10 3-540-25438-2 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-25438-6 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springeronline.com

Printed in The Netherlands

Typesetting: by the author and TechBooks using a Springer LaTeX macro package
Cover design: *design & production* GmbH, Heidelberg

Printed on acid-free paper SPIN: 11398325 41/TechBooks 5 4 3 2 1 0

To Margaret, as always

Preface

Namely, because the shape of the whole universe is most perfect and, in fact, designed by the wisest creator, nothing in all of the world will occur in which no maximum or minimum rule is somehow shining forth.

Leonhard Euler (1744)

We can treat this firm stand by Euler [411] ("...nihil omnino in mundo contingint, in quo non maximi minimive ratio quapiam eluceat") as the most fundamental principle of *Variational Analysis*. This principle justifies a variety of striking implementations of *optimization/variational* approaches to solving numerous problems in mathematics and applied sciences that may not be of a variational nature. Remember that optimization has been a major motivation and driving force for developing differential and integral calculus. Indeed, the *very concept of derivative* introduced by Fermat via the tangent slope to the graph of a function was motivated by solving an optimization problem; it led to what is now called the *Fermat stationary principle*. Besides applications to optimization, the latter principle plays a crucial role in proving the most important calculus results including the mean value theorem, the implicit and inverse function theorems, etc. The same line of development can be seen in the infinite-dimensional setting, where the Brachistochrone was the first problem not only of the calculus of variations but of all functional analysis inspiring, in particular, a variety of concepts and techniques in infinite-dimensional differentiation and related areas.

Modern variational analysis can be viewed as an outgrowth of the calculus of variations and mathematical programming, where the focus is on optimization of functions relative to various constraints and on sensitivity/stability of optimization-related problems with respect to perturbations. Classical notions of variations such as moving away from a given point or curve no longer play a critical role, while concepts of problem *approximations* and/or *perturbations* become crucial.

One of the most characteristic features of modern variational analysis is the intrinsic presence of *nonsmoothness*, i.e., the necessity to deal with nondifferentiable functions, sets with nonsmooth boundaries, and set-valued mappings. Nonsmoothness naturally enters not only through initial data of optimization-related problems (particularly those with inequality and geometric constraints) but largely via *variational principles* and other optimization, approximation, and perturbation techniques applied to problems with even smooth data. In fact, many fundamental objects frequently appearing in the framework of variational analysis (e.g., the distance function, value functions in optimization and control problems, maximum and minimum functions, solution maps to perturbed constraint and variational systems, etc.) are inevitably of nonsmooth and/or set-valued structures requiring the development of new forms of analysis that involve *generalized differentiation*.

It is important to emphasize that even the simplest and historically earliest problems of *optimal control* are *intrinsically nonsmooth*, in contrast to the classical calculus of variations. This is mainly due to pointwise constraints on control functions that often take only discrete values as in typical problems of automatic control, a primary motivation for developing optimal control theory. Optimal control has always been a major source of inspiration as well as a fruitful territory for applications of advanced methods of variational analysis and generalized differentiation.

Key issues of variational analysis in finite-dimensional spaces have been addressed in the book "Variational Analysis" by Rockafellar and Wets [1165]. The development and applications of variational analysis in infinite dimensions require certain concepts and tools that cannot be found in the finite-dimensional theory. The *primary goals* of this book are to present basic concepts and principles of variational analysis unified in finite-dimensional and infinite-dimensional space settings, to develop a comprehensive generalized differential theory at the same level of perfection in both finite and infinite dimensions, and to provide valuable applications of variational theory to broad classes of problems in constrained optimization and equilibrium, sensitivity and stability analysis, control theory for ordinary, functional-differential and partial differential equations, and also to selected problems in mechanics and economic modeling.

Generalized differentiation lies at the heart of variational analysis and its applications. We systematically develop a *geometric dual-space approach* to generalized differentiation theory revolving around the *extremal principle*, which can be viewed as a local *variational* counterpart of the classical convex separation in nonconvex settings. This principle allows us to deal with *nonconvex* derivative-like constructions for sets (normal cones), set-valued mappings (coderivatives), and extended-real-valued functions (subdifferentials). These constructions are defined directly in dual spaces and, being nonconvex-valued, cannot be generated by any derivative-like constructions in primal spaces (like tangent cones and directional derivatives). Nevertheless, our basic nonconvex constructions enjoy comprehensive calculi, which happen to be significantly better than those available for their primal and/or convex-valued counterparts. Thus passing to *dual spaces*, we are able to achieve more beauty and harmony in comparison with primal world objects. In some sense, the dual viewpoint does indeed allow us to meet the perfection requirement in the fundamental statement by Euler quoted above.

Observe to this end that dual objects (multipliers, adjoint arcs, shadow prices, etc.) have always been at the center of variational theory and applications used, in particular, for formulating principal optimality conditions in the calculus of variations, mathematical programming, optimal control, and economic modeling. The usage of variations of optimal solutions in primal spaces can be considered just as a convenient tool for deriving necessary optimality conditions. There are no essential restrictions in such a "primal" approach in smooth and convex frameworks, since primal and dual derivative-like constructions are equivalent for these classical settings. It is not the case any more in the framework of modern variational analysis, where even *nonconvex primal space* local approximations (e.g., tangent cones) inevitably yield, *under duality, convex sets* of normals and subgradients. This convexity of dual objects leads to significant restrictions for the theory and applications. Moreover, there are many situations particularly identified in this book, where primal space approximations simply cannot be used for variational analysis, while the employment of dual space constructions provides comprehensive results. Nevertheless, tangentially generated/primal space constructions play an important role in some other aspects of variational analysis, especially in finite-dimensional spaces, where they recover in duality the nonconvex sets of our basic normals and subgradients at the point in question by *passing to the limit* from points nearby; see, for instance, the afore-mentioned book by Rockafellar and Wets [1165]

Among the abundant bibliography of this book, we refer the reader to the monographs by Aubin and Frankowska [54], Bardi and Capuzzo Dolcetta [85], Beer [92], Bonnans and Shapiro [133], Clarke [255], Clarke, Ledyaev, Stern and Wolenski [265], Facchinei and Pang [424], Klatte and Kummer [686], Vinter [1289], and to the comments given after each chapter for significant aspects of variational analysis and impressive applications of this rapidly growing area that are not considered in the book. We especially emphasize the concurrent and complementing monograph "Techniques of Variational Analysis" by Borwein and Zhu [164], which provides a nice introduction to some fundamental techniques of modern variational analysis covering important theoretical aspects and applications not included in this book.

The book presented to the reader's attention is self-contained and mostly collects results that have not been published in the monographical literature. It is split into two volumes and consists of eight chapters divided into sections and subsections. Extensive comments (that play a special role in this book discussing basic ideas, history, motivations, various interrelations, choice of terminology and notation, open problems, etc.) are given for each chapter. We present and discuss numerous references to the vast literature on many aspects of variational analysis (considered and not considered in the book) including early contributions and very recent developments. Although there are no formal exercises, the extensive remarks and examples provide grist for further thought and development. Proofs of the major results are complete, while there is plenty of room for furnishing details, considering special cases, and deriving generalizations for which guidelines are often given.

Volume I "Basic Theory" consists of four chapters mostly devoted to basic constructions of generalized differentiation, fundamental extremal and variational principles, comprehensive generalized differential calculus, and complete dual characterizations of fundamental properties in nonlinear study related to Lipschitzian stability and metric regularity with their applications to sensitivity analysis of constraint and variational systems.

Chapter 1 concerns the generalized differential theory in arbitrary *Banach spaces*. Our basic normals, subgradients, and coderivatives are directly defined in dual spaces via *sequential* weak* limits involving more primitive ε-normals and ε-subgradients of the Fréchet type. We show that these constructions have a variety of nice properties in the general Banach spaces setting, where the usage of ε-enlargements is crucial. Most such properties (including first-order and second-order calculus rules, efficient representations, variational descriptions, subgradient calculations for distance functions, necessary coderivative conditions for Lipschitzian stability and metric regularity, etc.) are collected in this chapter. Here we also define and start studying the so-called *sequential normal compactness* (SNC) properties of sets, set-valued mappings, and extended-real-valued functions that automatically hold in finite dimensions while being one of the most essential ingredients of variational analysis and its applications in infinite-dimensional spaces.

Chapter 2 contains a detailed study of the *extremal principle* in variational analysis, which is the main single tool of this book. First we give a direct variational proof of the extremal principle in finite-dimensional spaces based on a smoothing penalization procedure via the method of *metric approximations*. Then we proceed by infinite-dimensional variational techniques in Banach spaces with a Fréchet smooth norm and finally, by separable reduction, in the larger class of *Asplund spaces*. The latter class is well-investigated in the geometric theory of Banach spaces and contains, in particular, every reflexive space and every space with a separable dual. Asplund spaces play a prominent role in the theory and applications of variational analysis developed in this book. In Chap. 2 we also establish relationships between the (geometric) extremal principle and (analytic) variational principles in both conventional and enhanced forms. The results obtained are applied to the derivation of novel variational characterizations of Asplund spaces and useful representations of the basic generalized differential constructions in the Asplund space setting similar to those in finite dimensions. Finally, in this chapter we discuss abstract versions of the extremal principle formulated in terms of axiomatically defined normal and subdifferential structures on appropriate Banach spaces and also overview in more detail some specific constructions.

Chapter 3 is a cornerstone of the generalized differential theory developed in this book. It contains comprehensive *calculus rules* for basic normals, subgradients, and coderivatives in the framework of Asplund spaces. We pay most of our attention to *pointbased* rules via the limiting constructions *at* the points in question, for both assumptions and conclusions, having in mind that pointbased results indeed happen to be of crucial importance for applications. A number of the results presented in this chapter seem to be new even in the finite-dimensional setting, while overall we achieve the same level of perfection and generality in Asplund spaces as in finite dimensions. The main issue that distinguishes the finite-dimensional and infinite-dimensional settings is the necessity to invoke *sufficient amounts of compactness* in infinite dimensions that are not needed at all in finite-dimensional spaces. The required compactness is provided by the afore-mentioned SNC properties, which are included in the assumptions of calculus rules and call for their own calculus ensuring the preservation of SNC properties under various operations on sets and mappings. The absence of such a *SNC calculus* was a crucial obstacle for many successful applications of generalized differentiation in infinite-dimensional spaces to a range of infinite-dimensions problems including those in optimization, stability, and optimal control given in this book. Chapter 3 contains a broad spectrum of the SNC calculus results that are decisive for subsequent applications.

Chapter 4 is devoted to a thorough study of Lipschitzian, metric regularity, and linear openness/covering properties of set-valued mappings, and to their applications to sensitivity analysis of parametric constraint and variational systems. First we show, based on variational principles and the generalized differentiation theory developed above, that the necessary coderivative conditions for these fundamental properties derived in Chap. 1 in arbitrary Banach spaces happen to be *complete characterizations* of these properties in the Asplund space setting. Moreover, the employed variational approach allows us to obtain verifiable formulas for computing the *exact bounds* of the corresponding moduli. Then we present detailed applications of these results, supported by generalized differential and SNC calculi, to sensitivity and stability analysis of parametric constraint and variational systems governed by perturbed sets of feasible and optimal solutions in problems of optimization and equilibria, implicit multifunctions, complementarity conditions, variational and hemivariational inequalities as well as to some mechanical systems.

Volume II "Applications" also consists of four chapters mostly devoted to applications of basic principles in variational analysis and the developed generalized differential calculus to various topics in constrained optimization and equilibria, optimal control of ordinary and distributed-parameter systems, and models of welfare economics.

Chapter 5 concerns constrained optimization and equilibrium problems with possibly nonsmooth data. Advanced methods of variational analysis based on extremal/variational principles and generalized differentiation happen to be very useful for the study of constrained problems even with smooth initial data, since nonsmoothness naturally appears while applying penalization, approximation, and perturbation techniques. Our primary goal is to derive necessary optimality and suboptimality conditions for various constrained problems in both finite-dimensional and infinite-dimensional settings. Note that conditions of the latter – *suboptimality* – type, somehow underestimated in optimization theory, don't assume the existence of optimal solutions (which is especially significant in infinite dimensions) ensuring that "almost" optimal solutions "almost" satisfy necessary conditions for optimality. Besides considering problems with constraints of conventional types, we pay serious attention to rather new classes of problems, labeled as *mathematical problems with equilibrium constraints* (MPECs) and *equilibrium problems with equilibrium constraints* (EPECs), which are intrinsically nonsmooth while admitting a thorough analysis by using generalized differentiation. Finally, certain concepts of *linear subextremality* and *linear suboptimality* are formulated in such a way that the necessary optimality conditions derived above for conventional notions are seen to be *necessary and sufficient* in the new setting.

In *Chapter 6* we start studying problems of *dynamic optimization* and *optimal control* that, as mentioned, have been among the primary motivations for developing new forms of variational analysis. This chapter deals mostly with optimal control problems governed by *ordinary* dynamic systems whose state space may be infinite-dimensional. The main attention in the first part of the chapter is paid to the Bolza-type problem for evolution systems governed by constrained *differential inclusions*. Such models cover more conventional control systems governed by parameterized evolution equations with control regions generally dependent on state variables. The latter don't allow us to use control variations for deriving necessary optimality conditions. We develop the *method of discrete approximations*, which is certainly of numerical interest, while it is mainly used in this book as a direct vehicle to derive optimality conditions for continuous-time systems by passing to the limit from their discrete-time counterparts. In this way we obtain, strongly based on the generalized differential and SNC calculi, necessary optimality conditions in the extended Euler-Lagrange form for nonconvex differential inclusions in infinite dimensions expressed via our basic generalized differential constructions.

The second part of Chap. 6 deals with constrained optimal control systems governed by ordinary evolution equations of *smooth dynamics* in arbitrary Banach spaces. Such problems have essential specific features in comparison with the differential inclusion model considered above, and the results obtained (as well as the methods employed) in the two parts of this chapter are generally independent. Another major theme explored here concerns *stability* of the maximum principle under discrete approximations of nonconvex control systems. We establish rather surprising results on the *approximate maximum principle* for discrete approximations that shed new light upon both qualitative and quantitative relationships between continuous-time and discrete-time systems of optimal control.

In *Chapter 7* we continue the study of optimal control problems by applications of advanced methods of variational analysis, now considering systems with *distributed parameters.* First we examine a general class of *hereditary systems* whose dynamic constraints are described by both delay-differential inclusions and linear algebraic equations. On one hand, this is an interesting and not well-investigated class of control systems, which can be treated as a special type of variational problems for *neutral functional-differential inclusions* containing time delays not only in state but also in velocity variables. On the other hand, this class is related to differential-algebraic systems with a linear link between "slow" and "fast" variables. Employing the method of discrete approximations and the basic tools of generalized differentiation, we establish a strong variational convergence/stability of discrete approximations and derive extended optimality conditions for continuous-time systems in both Euler-Lagrange and Hamiltonian forms.

The rest of Chap. 7 is devoted to optimal control problems governed by *partial differential equations* with *pointwise* control and state constraints. We pay our primary attention to evolution systems described by *parabolic* and *hyperbolic* equations with controls functions acting in the Dirichlet and Neumann boundary conditions. It happens that such *boundary control* problems are the most challenging and the least investigated in PDE optimal control theory, especially in the presence of pointwise state constraints. Employing approximation and perturbation methods of modern variational analysis, we justify variational convergence and derive necessary optimality conditions for various control problems for such PDE systems including *minimax* control under *uncertain disturbances.*

The concluding *Chapter 8* is on applications of variational analysis to *economic modeling.* The major topic here is *welfare economics*, in the general nonconvex setting with infinite-dimensional commodity spaces. This important class of competitive equilibrium models has drawn much attention of economists and mathematicians, especially in recent years when nonconvexity has become a crucial issue for practical applications. We show that the methods of variational analysis developed in this book, particularly the extremal principle, provide adequate tools to study Pareto optimal allocations and associated price equilibria in such models. The tools of variational analysis and generalized differentiation allow us to obtain extended nonconvex versions of the so-called "second fundamental theorem of welfare economics" describing marginal equilibrium prices in terms of minimal collections of generalized normals to nonconvex sets. In particular, our approach and variational descriptions of generalized normals offer new economic interpretations of market equilibria via "nonlinear marginal prices" whose role in nonconvex models is similar to the one played by conventional linear prices in convex models of the Arrow-Debreu type.

The book includes a Glossary of Notation, common for both volumes, and an extensive Subject Index compiled separately for each volume. Using the Subject Index, the reader can easily find not only the page, where some notion and/or notation is introduced, but also various places providing more discussions and significant applications for the object in question.

Furthermore, it seems to be reasonable to title all the statements of the book (definitions, theorems, lemmas, propositions, corollaries, examples, and remarks) that are numbered in sequence within a chapter; thus, in Chap. 5 for instance, Example 5.3.3 precedes Theorem 5.3.4, which is followed by Corollary 5.3.5. For the reader's convenience, all these statements and numerated comments are indicated in the List of Statements presented at the end of each volume. It is worth mentioning that the list of acronyms is included (in alphabetic order) in the Subject Index and that the common principle adopted for the book notation is to use lower case Greek characters for numbers and (extended) real-valued functions, to use lower case Latin characters for vectors and single-valued mappings, and to use Greek and Latin upper case characters for sets and set-valued mappings.

Our notation and terminology are generally consistent with those in Rockafellar and Wets [1165]. Note that we try to distinguish everywhere the notions defined *at* the point and *around* the point in question. The latter indicates *robustness/stability* with respect to perturbations, which is critical for most of the major results developed in the book.

The book is accompanied by the abundant bibliography (with English sources if available), common for both volumes, which reflects a variety of topics and contributions of many researchers. The references included in the bibliography are discussed, at various degrees, mostly in the extensive commentaries to each chapter. The reader can find further information in the given references, directed by the author's comments.

We address this book mainly to researchers and graduate students in mathematical sciences; first of all to those interested in nonlinear analysis, optimization, equilibria, control theory, functional analysis, ordinary and partial differential equations, functional-differential equations, continuum mechanics, and mathematical economics. We also envision that the book will be useful to a broad range of researchers, practitioners, and graduate students involved in the study and applications of variational methods in operations research, statistics, mechanics, engineering, economics, and other applied sciences.

Parts of the book have been used by the author in teaching graduate classes on variational analysis, optimization, and optimal control at Wayne State University. Basic material has also been incorporated into many lectures and tutorials given by the author at various schools and scientific meetings during the recent years.

Acknowledgments

My first gratitude go to Terry Rockafellar who has encouraged me over the years to write such a book and who has advised and supported me at all the stages of this project.

Special thanks are addressed to Rafail Gabasov, my doctoral thesis adviser, from whom I learned optimal control and much more; to Alec Ioffe, Boris Polyak, and Vladimir Tikhomirov who recognized and strongly supported my first efforts in nonsmooth analysis and optimization; to Sasha Kruger, my first graduate student and collaborator in the beginning of our exciting journey to generalized differentiation; to Jon Borwein and Marián Fabian from whom I learned deep functional analysis and the beauty of Asplund spaces; to Ali Khan whose stimulating work and enthusiasm have encouraged my study of economic modeling; to Jiři Outrata who has motivated and influenced my growing interest in equilibrium problems and mechanics and who has intensely promoted the implementation of the basic generalized differential constructions of this book in various areas of optimization theory and applications; and to Jean-Pierre Raymond from whom I have greatly benefited on modern theory of partial differential equations.

During the work on this book, I have had the pleasure of discussing its various aspects and results with many colleagues and friends. Besides the individuals mentioned above, I'm particularly indebted to Zvi Artstein, Jim Burke, Tzanko Donchev, Asen Dontchev, Joydeep Dutta, Andrew Eberhard, Ivar Ekeland, Hector Fattorini, René Henrion, Jean-Baptiste Hiriart-Urruty, Alejandro Jofré, Abderrahim Jourani, Michal Kočvara, Irena Lasiecka, Claude Lemaréchal, Adam Levy, Adrian Lewis, Kazik Malanowski, Michael Overton, Jong-Shi Pang, Teemu Pennanen, Steve Robinson, Alex Rubinov, Andrzej Świech, Michel Théra, Lionel Thibault, Jay Treiman, Hector Sussmann, Roberto Triggiani, Richard Vinter, Nguyen Dong Yen, George Yin, Jack Warga, Roger Wets, and Jim Zhu for valuable suggestions and fruitful conversations throughout the years of the fulfillment of this project.

The continuous support of my research by the National Science Foundation is gratefully acknowledged.

As mentioned above, the material of this book has been used over the years for teaching advanced classes on variational analysis and optimization attended mostly by my doctoral students and collaborators. I highly appreciate their contributions, which particularly allowed me to improve my lecture notes and book manuscript. Especially valuable help was provided by Glenn Malcolm, Nguyen Mau Nam, Yongheng Shao, Ilya Shvartsman, and Bingwu Wang. Useful feedback and text corrections came also from Truong Bao, Wondi Geremew, Pankaj Gupta, Aychi Habte, Kahina Sid Idris, Dong Wang, Lianwen Wang, and Kaixia Zhang.

I'm very grateful to the nice people in Springer for their strong support during the preparation and publishing this book. My special thanks go to Catriona Byrne, Executive Editor in Mathematics, to Achi Dosajh, Senior Editor in Applied Mathematics, to Stefanie Zoeller, Assistant Editor in Mathematics, and to Frank Holzwarth from the Computer Science Editorial Department.

I thank my younger daughter Irina for her interest in my book and for her endless patience and tolerance in answering my numerous question on English. I would also like to thank my poodle Wuffy for his sharing with me the long days of work on this book. Above all, I don't have enough words to thank my wife Margaret for her sharing with me everything, starting with our high school years in Minsk.

Ann Arbor, Michigan
August 2005

Boris Mordukhovich

Contents

Volume I Basic Theory

Volume II Applications

Volume II

Applications

5

Constrained Optimization and Equilibria

This chapter is devoted to applications of the basic tools of variational analysis and generalized differential calculus developed above to the study of constrained optimization and equilibrium problems with possibly nonsmooth data. Actually it is a *two-sided process*, since *optimization ideas lie at the very heart of variational analysis* as clearly follows from the previous material. Let us particularly mention variational descriptions of the normals and subgradients under consideration in both finite and infinite dimensions; see Theorem 1.6, Subsect. 1.1.4, and Theorem 1.88 for more details. Moreover, the main instrument of our analysis—the *extremal principle*—itself gives necessary conditions for set extremality, which are at the core of the basic results on generalized differential calculus and related characterizations of Lipschitzian stability and metric regularity developed in Chaps. 2–4.

The primary objective of this chapter is to derive *necessary optimality* and *suboptimality* conditions for various problems of constrained optimization and equilibria in infinite-dimensional spaces. Note that results of the latter (suboptimality) type ensure that "almost" optimal solutions "almost" satisfy necessary conditions for optimality *without* imposing assumptions on the *existence* of exact optimizers, which is essential in infinite dimensions. Starting with problems of mathematical programming under functional and geometric constraints, we consider then various problems of multiobjective optimization, minimax problems and equilibrium constraints, some concepts of extended extremality, etc. The key tools of our analysis are based on the extremal principle and its modifications together with generalized differential calculus. A major role is played by the SNC calculus that is crucial for applications to constrained optimization and equilibrium problems in infinite dimensions.

5.1 Necessary Conditions in Mathematical Programming

This section concerns first-order necessary optimality and suboptimality conditions for general problems of mathematical programming with operator, functional, and geometric constraints. We derive such conditions in various forms depending on the type of assumptions imposed on the initial data. Let us first examine optimization problems with only geometric constraints given by arbitrary nonempty subsets of Banach and Asplund spaces.

5.1.1 Minimization Problems with Geometric Constraints

Given a function $\varphi\colon X \to \overline{I\!R}$ finite at a reference point and a nonempty subset Ω of a Banach space X, we consider the following minimization problem with geometric constraints:

$$\text{minimize } \varphi(x) \text{ subject to } x \in \Omega \subset X\,. \tag{5.1}$$

The constrained problem (5.1) is obviously equivalent to the problem of *unconstrained* minimization:

$$\text{minimize } \varphi(x) + \delta(x;\Omega), \quad x \in X,$$

where the indicator function $\delta(\cdot;\Omega)$ imposes an "infinite penalty" on the constraint violation. Thus, given a local optimal solution $\bar{x}$ to (5.1), we get

$$0 \in \widehat{\partial}\big(\varphi + \delta(\cdot;\Omega)\big)(\bar{x}) \subset \partial\big(\varphi + \delta(\cdot;\Omega)\big)(\bar{x}) \tag{5.2}$$

by the generalized Fermat rule from Proposition 1.114. To pass from (5.2) to efficient necessary optimality conditions in terms of the initial data (φ,Ω), one needs to employ subdifferential *sum rules* for $\varphi + \delta(\cdot;\Omega)$. The simplest result in this direction follows from the sum rule in Proposition 1.107(i) provided that φ is Fréchet differentiable at $\bar{x}$.

Proposition 5.1 (necessary conditions for constrained problems with Fréchet differentiable costs). *Let $\bar{x}$ be a local optimal solution to problem* (5.1) *in a Banach space X. Assume that φ is Fréchet differentiable at $\bar{x}$. Then*

$$-\nabla\varphi(\bar{x}) \in \widehat{N}(\bar{x};\Omega), \qquad -\nabla\varphi(\bar{x}) \in N(\bar{x};\Omega)\,.$$

Proof. Applying Proposition 1.107(i) to the first inclusion in (5.2) and using the relationship $\widehat{\partial}\delta(\bar{x};\Omega) = \widehat{N}(\bar{x};\Omega)$, we arrive at $-\nabla\varphi(\bar{x}) \in \widehat{N}(\bar{x};\Omega)$. This immediately implies the second necessary condition in the proposition due to the inclusion $\widehat{N}(\bar{x};\Omega) \subset N(\bar{x};\Omega)$. △

If φ is not Fréchet differentiable at $\bar{x}$, one cannot proceed in the above way using Fréchet-like subgradient constructions, which don't possess a satisfactory calculus even in finite dimensions. The picture is completely different for our basic constructions $\partial\varphi$ and $N(\cdot,\Omega)$, which enjoy a full calculus in general nonsmooth settings of Asplund spaces. Before going in this direction, let us present a rather surprising result providing *upper subdifferential* necessary conditions in the *minimization* problem (5.1) that happen to be very efficient for a special class of functions φ. These necessary optimality conditions generalize those in Proposition 5.1 and actually reduce to them in the proof due to the *variational description* of Fréchet subgradients in Theorem 1.88(i) applied to the Fréchet upper subdifferential $\widehat{\partial}^+\varphi(\bar{x})$ defined in (1.52).

Proposition 5.2 (upper subdifferential conditions for local minima under geometric constraints). *Let $\bar{x}$ be a local optimal solution to the minimization problem* (5.1) *in a Banach space X, where $\varphi: X \to \overline{I\!R}$ is finite at $\bar{x}$. Then one has the inclusions*

$$-\widehat{\partial}^+\varphi(\bar{x}) \subset \widehat{N}(\bar{x};\Omega), \qquad -\widehat{\partial}^+\varphi(\bar{x}) \subset N(\bar{x};\Omega)\,. \tag{5.3}$$

Proof. We only need to prove the first inclusion in (5.3), which is trivial when $\widehat{\partial}^+\varphi(\bar{x}) = \emptyset$. Assume that $\widehat{\partial}^+\varphi(\bar{x}) \neq \emptyset$ and take $x^* \in \widehat{\partial}^+\varphi(\bar{x}) = -\widehat{\partial}(-\varphi)(\bar{x})$. Applying Theorem 1.88(i) to the Fréchet subgradient $-x^*$ from $\widehat{\partial}(-\varphi)(\bar{x})$, we find a function $s: X \to I\!R$ with $s(\bar{x}) = \varphi(\bar{x})$ and $s(x) \geq \varphi(x)$ whenever $x \in X$ such that $s(\cdot)$ is Fréchet differentiable at $\bar{x}$ with $\nabla s(\bar{x}) = x^*$. It gives

$$s(\bar{x}) = \varphi(\bar{x}) \leq \varphi(x) \leq s(x)$$

for all $x \in \Omega$ around $\bar{x}$. Thus $\bar{x}$ is a local optimal solution to the constrained minimization problem:

$$\text{minimize } \; s(x) \; \text{ subject to } \; x \in \Omega$$

with a Fréchet differentiable objective. Applying Proposition 5.1 to the latter problem, we conclude that $-x^* \in \widehat{N}(\bar{x};\Omega)$, which gives (5.3) and complete the proof of the proposition. △

When φ is Fréchet differentiable at $\bar{x}$, the result of Proposition 5.2 reduces to the inclusion $-\nabla\varphi(\bar{x}) \in \widehat{N}(\bar{x};\Omega)$ in Proposition 5.1 due to $\widehat{\partial}^+\varphi(\bar{x}) = \{\nabla\varphi(\bar{x})\}$ in this case. An interesting class of optimization problems satisfying the assumptions of Proposition 5.2 contains problems of *concave minimization* when φ is concave and continuous around $\bar{x}$, and hence $\widehat{\partial}^+\varphi(\bar{x})$ agrees with the (nonempty) upper subdifferential of convex analysis. If X is Asplund, $\widehat{\partial}^+\varphi(\bar{x}) \neq \emptyset$ when φ is Lipschitz continuous around $\bar{x}$ and *upper regular* at $\bar{x}$, i.e., $\widehat{\partial}^+\varphi(\bar{x}) = \partial^+\varphi(\bar{x})$. Indeed, in this case one has $\partial^+\varphi(\bar{x}) = -\partial(-\varphi)(\bar{x}) \neq \emptyset$ by Corollary 2.25. Observe that the latter class contains, besides strictly differentiable functions and concave continuous functions, the so-called *semiconcave* functions that are very important for many applications; see more discussions in Subsect. 5.5.4 containing comments to this chapter.

Note that the condition $\widehat{\partial}^+\varphi(\bar{x}) = \emptyset$, when the inclusions in (5.3) are trivial, *itself is an easy checkable necessary optimality condition* for (5.1) whenever the constraints are not into account and φ is not Fréchet differentiable at $\bar{x}$. Indeed, since $0 \in \widehat{\partial}\varphi(\bar{x}) \neq \emptyset$ at a point of local minimum, then $\widehat{\partial}^+\varphi(\bar{x})$ must be empty by Proposition 1.87. However, it is a trivial necessary condition that doesn't carry much information for constrained minimization problems. The following *lower subdifferential* conditions, expressed in terms of basic and singular lower subgradients of the cost function φ, are more conventional for constrained minimization.

Proposition 5.3 (lower subdifferential conditions for local minima under geometric constraints). *Let $\bar{x}$ be a local optimal solution to the minimization problem* (5.1)*, where Ω is locally closed and φ is l.s.c. around $\bar{x}$ while X is Asplund. Assume that*

$$\partial^\infty\varphi(\bar{x}) \cap \big(-N(\bar{x};\Omega)\big) = \{0\} \tag{5.4}$$

and that either Ω is SNC at $\bar{x}$ or φ is SNEC at $\bar{x}$; all these assumptions hold if φ is locally Lipschitzian around $\bar{x}$. Then one has

$$\partial\varphi(\bar{x}) \cap \big(-N(\bar{x};\Omega)\big) \neq \emptyset, \quad i.e., \quad 0 \in \partial\varphi(\bar{x}) + N(\bar{x};\Omega)\ . \tag{5.5}$$

Proof. It follows from the subdifferential sum rule in Theorem 3.36 applied to the basic subdifferential of the sum in (5.2). △

Remark 5.4 (upper subdifferential versus lower subdifferential conditions for local minima). Observe that, despite the broader applicability of Proposition 5.3, the upper subdifferential conditions of Proposition 5.2 may give an essentially stronger result for special classes of nonsmooth problems, even in the case of Lipschitzian functions φ in finite dimensions. In particular, for *concave* continuous functions φ one has, by Theorem 1.93, that

$$\partial\varphi(\bar{x}) \subset \partial^+\varphi(\bar{x}) = \widehat{\partial}^+\varphi(\bar{x}) \neq \emptyset\ .$$

Then comparing the second inclusion in (5.3) (which is even weaker than the first inclusion therein) with the one in (5.5), we see that the necessary condition of Proposition 5.2 requires that *every* element x^* of the set $\widehat{\partial}^+\varphi(\bar{x})$ must belong to $-N(\bar{x};\Omega)$, instead of that *some* element x^* from the *smaller* set $\partial\varphi(\bar{x})$ belongs to $-N(\bar{x};\Omega)$ by Proposition 5.3. This shows that the upper subdifferential necessary conditions for local minima may have sizeable advantages over the lower subdifferential conditions above when the former apply. Let us illustrate it by a simple *example*:

$$\text{minimize} \;\; \varphi(x) := -|x| \;\; \text{subject to} \;\; x \in \Omega := [-1,0] \subset I\!R\ .$$

Obviously $\bar{x} = 0$ is not an optimal solution to this problem. However, it cannot be taken away by the lower subdifferential necessary condition (5.5), which gives in this case the relations

$$\partial\varphi(0) = \{-1,1\}, \quad N(0;\Omega) = [0,\infty), \quad \text{and} \quad -1 \in -N(0;\Omega)\ .$$

On the other hand, the upper subdifferential necessary conditions in (5.3), which are the same in this case, don't hold for $\bar{x} = 0$, since

$$\widehat{\partial}^+\varphi(0) = [-1,1] \quad \text{and} \quad [-1,1] \not\subset N(0;\Omega) .$$

This confirms non-optimality of $\bar{x} = 0$ in the example problem by Proposition 5.2 in contrast to Proposition 5.3.

Observe that the class of minimization problems for the *difference of two convex functions* (i.e., for the so-called *DC-functions* important in various applications) can be equivalently reduced to minimizing *concave* functions subject to convex constraints; see, e.g., Horst, Pardalos and Thoai [583] for more developments and discussions.

Note also that, when φ is *upper regular* at $\bar{x}$ and Lipschitz continuous around this point, one has the relationship

$$\partial_C\varphi(\bar{x}) = \text{cl}^*\widehat{\partial}^+\varphi(\bar{x})$$

between Clarke's generalized gradient and the Fréchet *upper* subdifferential of φ at $\bar{x}$ provided that X is Asplund. Indeed, it follows from the symmetry (2.71) of the generalized gradient and its representation via the basic subdifferential in Theorem 3.57(ii). Moreover, the weak* closure operation above is *redundant* if X is WCG; see Theorem 3.59(i). Thus, if one replaces in this case the basic subdifferential $\partial\varphi(\bar{x})$ in Proposition 5.3 by its Clarke counterpart, the obtained lower subdifferential result is *substantially weaker* than the upper subdifferential condition of Proposition 5.2 with $\widehat{\partial}^+\varphi(\bar{x}) = \partial_C\varphi(\bar{x})$.

In many areas of the variational theory and applications (in particular, to optimal control) geometric constraints are usually given as *intersections of sets*; see, e.g., the next section and Chap. 6. Based on the above results for problem (5.1) and calculus rules for basic normals to set intersections, one can derive necessary optimality conditions for optimization problems with many geometric constraints. To furnish this in the case of upper subdifferential conditions, we employ the second inclusion in (5.3), since the first one doesn't lead to valuable *pointbased* results for set intersections due to the lack of calculus for Fréchet normals.

Let us present general results in both lower and upper subdifferential forms considering for simplicity the case of two set intersections in geometric constraints given in products of Asplund spaces. In the next theorem we use the qualification and PSNC conditions introduced in Subsect. 3.1.1; see also discussions therein.

Theorem 5.5 (local minima under geometric constraints with set intersections). *Let $\bar{x}$ be a local optimal solution to problem* (5.1) *with $\Omega = \Omega_1 \cap \Omega_2$, where the sets $\Omega_1, \Omega_2 \subset \prod_{j=1}^m X_j$ are locally closed around $\bar{x}$ and the spaces X_j are Asplund. Then the following assertions hold:*

(i) *Assume that the system $\{\Omega_1, \Omega_2\}$ satisfies the limiting qualification condition at $\bar{x}$. Given $J_1, J_2 \subset \{1,\ldots,m\}$ with $J_1 \cup J_2 = \{1,\ldots,m\}$, we also assume that Ω_1 is PSNC at $\bar{x}$ with respect to J_1 and that Ω_2 is strongly PSNC at $\bar{x}$ with respect to J_2. Then one has*

$$-\widehat{\partial}^+\varphi(\bar{x}) \subset N(\bar{x};\Omega_1) + N(\bar{x};\Omega_2)\,.$$

(ii) *In addition to the assumptions in* (i), *suppose that* φ *is l.s.c. around* $\bar{x}$ *and SNEC at this point and that*

$$\big(-\partial^\infty\varphi(\bar{x})\big) \cap \big[N(\bar{x};\Omega_1) + N(\bar{x};\Omega_2)\big] = \{0\} \tag{5.6}$$

(*all the additional assumptions are satisfied if* φ *is Lipschitz continuous around* $\bar{x}$). *Then one has*

$$0 \in \partial\varphi(\bar{x}) + N(\bar{x};\Omega_1) + N(\bar{x};\Omega_2)\,. \tag{5.7}$$

(iii) *Assume that* φ *is l.s.c. around* $\bar{x}$, *that both* Ω_1 *and* Ω_2 *are SNC at this point, and that the qualification condition*

$$\begin{aligned}&\Big[x^* \in \partial^\infty\varphi(\bar{x}), \quad x_1^* \in N(\bar{x};\Omega_1), \quad x_2^* \in N(\bar{x};\Omega_2)\,,\\ &x^* + x_1^* + x_2^* = 0\Big] \Longrightarrow x^* = x_1^* = x_2^* = 0\end{aligned} \tag{5.8}$$

holds. Then one has (5.7).

Proof. To prove (i), we use Proposition 5.2 and then apply the intersection rule from Theorem 3.4 to the basic normal cone $N(\bar{x};\Omega)$ in (5.3). This gives

$$N(\bar{x};\Omega) = N(\bar{x};\Omega_1 \cap \Omega_2) \subset N(\bar{x};\Omega_1) + N(\bar{x};\Omega_2)\,, \tag{5.9}$$

and thus we arrive at the upper subdifferential inclusion in (i).

Assertion (ii) follows from Proposition 5.3 under the SNEC assumption on φ and from the intersection rule of Theorem 3.4 by substituting (5.9) into (5.4) and (5.5). Recall finally that every function φ locally Lipschitzian around $\bar{x}$ is SNEC at $\bar{x}$ due to Corollary 1.69 (see the discussion after Definition 1.116) with $\partial^\infty\varphi(\bar{x}) = \{0\}$ by Corollary 1.81.

It remains to prove (iii). Using Proposition 5.3 in the case of SNC sets Ω, we need to express the SNC assumption on Ω and the other conditions of that proposition in terms of Ω_1, Ω_2, and φ. By Corollary 3.81 the set intersection $\Omega = \Omega_1 \cap \Omega_2$ is SNC at $\bar{x}$ if both Ω_i are SNC at this point and satisfy the qualification condition

$$N(\bar{x};\Omega_1) \cap \big(-N(\bar{x};\Omega_2)\big) = \{0\}\,, \tag{5.10}$$

which also ensures the intersection formula (5.9); see Corollary 3.5. It is easy to check that (5.8) implies both qualification conditions (5.4) and (5.10). Indeed, (5.10) follows right from (5.8) with $x^* = 0$. To get (5.4), we take $x^* \in N(\bar{x};\Omega_1 \cap \Omega_2)$ with $-x^* \in \partial^\infty\varphi(\bar{x})$ and find $x_i^* \in N(\bar{x};\Omega_i)$, $i = 1, 2$, such that $x_1^* + x_2^* = x^*$ by Corollary 3.5. Thus $x^* + x_1^* + x_2^* = 0$, which gives $x^* = 0$ by (5.8) and ends the proof of the theorem. △

Let us present a corollary of Theorem 5.5 that unifies and simplifies its assumptions for the case of finitely many geometric constraints.

Corollary 5.6 (local minima under many geometric constraints). *Let $\bar{x}$ be a local optimal solution to problem* (5.1) *with $\Omega = \Omega_1 \cap \ldots \cap \Omega_n$, where each Ω_i is locally closed around $\bar{x}$ in the Asplund space X. Assume that all but one of Ω_i are SNC at $\bar{x}$ and that*

$$\Big[x_1^* + \ldots + x_n^* = 0,\ \ x_i^* \in N(\bar{x};\Omega_i)\Big] \Longrightarrow x_i^* = 0, \quad i = 1,\ldots,n\ . \tag{5.11}$$

Then the upper subdifferential necessary condition

$$-\widehat{\partial}^+\varphi(\bar{x}) \subset N(\bar{x};\Omega_1) + \ldots + N(\bar{x};\Omega_n)$$

holds. If in addition φ is l.s.c. around $\bar{x}$ and SNEC at this point and if (5.11) *is replaced by the stronger qualification condition*

$$\Big[x^* \in \partial^\infty\varphi(\bar{x}), \quad x_i^* \in N(\bar{x};\Omega_i),\ i = 1,\ldots,n\ ,$$

$$x^* + \sum_{i=1}^n x_i^* = 0\Big] \Longrightarrow x^* = x_1^* = \ldots = x_n^* = 0\ ,$$

then one has the lower subdifferential inclusion

$$0 \in \partial\varphi(\bar{x}) + N(\bar{x};\Omega_1) + \ldots + N(\bar{x};\Omega_n)$$

Furthermore, the latter necessary optimality condition still holds if the SNEC property of φ at $\bar{x}$ is replaced by the SNC property of all $\Omega_1,\ldots,\Omega_n$ at this point in the assumptions above.

Proof. It is clear that the qualification condition (5.11) together with the SNC property of all but one Ω_i imply the assumptions of Theorem 5.5(i) for two and then for n sets, by induction, and hence ensure the intersection rule

$$N(\bar{x};\Omega_1 \cap \ldots \cap \Omega_n) \subset N(\bar{x};\Omega_1) + \ldots + N(\bar{x};\Omega_n)\ ;$$

cf. Corollary 3.37. This justifies the upper subdifferential necessary condition of the corollary. The lower subdifferential condition is derived by induction from assertion (ii) of Theorem 5.5 under the SNEC assumption on φ and from assertion (iii) of this theorem under the SNC assumption on all Ω_i. △

5.1.2 Necessary Conditions under Operator Constraints

In this subsection we derive necessary optimality conditions in extended problems of mathematical programming that contain, along with geometric constraints, also *operator constraints* given by set-valued and single-valued mappings with values in infinite-dimensional spaces. Our analysis is mainly based on the reduction to minimization problems containing only geometric constraints given by intersections of two sets one of which is an inverse image of some set under a set-valued or single-valued mapping. Then we apply results on the generalized differential calculus developed in Chaps. 1 and 3 including efficient rules that ensure the fulfillment and preservation of SNC properties. In this way we derive general necessary optimality conditions of both lower and upper subdifferential types under certain constraint qualifications ensuring the so-called normal/qualified form of optimality conditions as well as necessary conditions without such qualifications.

Let us consider the following constrained optimization problem:

$$\text{minimize } \varphi_0(x) \text{ subject to } x \in F^{-1}(\Theta) \cap \Omega , \tag{5.12}$$

where $\varphi_0: X \to \overline{I\!R}$, $F: X \rightrightarrows Y$, $\Omega \subset X$, and $\Theta \subset Y$, and where

$$F^{-1}(\Theta) := \{x \in X |\ F(x) \cap \Theta \neq \emptyset\}$$

is the inverse image of the set Θ under the set-valued mapping F. Model (5.12) covers many special classes of optimization problems, in particular, classical problems of nonlinear programming with equality and inequality constraints; see the next subsection.

Observe that (5.12) reduces to the problem of constrained minimization admitting only geometric constraints given by the intersection of two sets: $\Omega_1 = F^{-1}(\Theta)$ and $\Omega_2 = \Omega$. Thus one can apply the results of the preceding subsection and then *calculus rules* for the normal cones to inverse images and intersections as well as those preserving the SNC property, which are developed in Chaps. 1 and 3. In this way we arrive at necessary optimality conditions for the general problem (5.12) obtained in the *normal form*, i.e., with a nonzero multiplier corresponding to the cost function φ_0. Let us first derive *upper subdifferential* necessary conditions for optimality in the minimization problem (5.12).

Theorem 5.7 (upper subdifferential conditions for local minima under operator constraints). *Given a local optimal solution $\bar{x}$ to problem* (5.12) *with Banach spaces X and Y, we have the assertions:*

(i) *Assume that $\Omega = X$, that $F = f: X \to Y$ is Fréchet differentiable at $\bar{x}$ with the surjective derivative $\nabla f(\bar{x})$, and that either f is strictly differentiable at $\bar{x}$ or it is continuous around this point with* $\dim Y < \infty$. *Then*

$$-\widehat{\partial}^+\varphi_0(\bar{x}) \subset \nabla f(\bar{x})^* \widehat{N}(f(\bar{x}); \Theta) .$$

(ii) *Assume that X is Asplund, that Ω is locally closed around $\bar{x}$, that $F = f: X \to Y$ is strictly differentiable at $\bar{x}$ with the surjective derivative, and that the qualification condition*

$$\nabla f(\bar{x})^* N(f(\bar{x}); \Theta) \cap \big(- N(\bar{x}; \Omega)\big) = \{0\}$$

holds. Then one has

$$-\widehat{\partial}^+\varphi_0(\bar{x}) \subset \nabla f(\bar{x})^* N(f(\bar{x}); \Theta) + N(\bar{x}; \Omega)$$

provided that either Ω *or* Θ *is SNC at* $\bar{x}$ *and* $f(\bar{x})$*, respectively.*

(iii) *Assume that both spaces* X *and* Y *are Asplund, that the sets* Ω, Θ, *and* $\operatorname{gph} F$ *are closed, and that the set-valued mapping* $S(\cdot) := F(\cdot) \cap \Theta$ *is inner semicompact around* $\bar{x}$*. Then*

$$-\widehat{\partial}^+\varphi_0(\bar{x}) \subset \bigcup \Big[D_N^* F(\bar{x},\bar{y})(y^*) \Big|\ \bar{y} \in S(\bar{x}),\ \ y^* \in N(\bar{y};\Theta) \Big] + N(\bar{x};\Omega) \tag{5.13}$$

under one of the following requirements on (F,Θ,Ω)*:*

(a) Ω *is SNC at* $\bar{x}$*, the qualification conditions*

$$\bigcup \Big[D_N^* F(\bar{x},\bar{y})(y^*) \Big|\ \bar{y} \in S(\bar{x}),\ \ y^* \in N(\bar{y};\Theta) \Big] \bigcap \big(- N(\bar{x};\Omega) \big) = \{0\}, \tag{5.14}$$

$$N(\bar{y};\Theta) \cap \ker \widetilde{D}_M^* F(\bar{x},\bar{y}) = \{0\} \ \ \textit{for all} \ \ \bar{y} \in S(\bar{x}) \tag{5.15}$$

are satisfied, and either the inverse mapping F^{-1} *is PSNC at* $(\bar{y},\bar{x})$ *or* Θ *is SNC at* $\bar{y}$ *for all* $\bar{y} \in S(\bar{x})$*.*

(b) *The qualification conditions* (5.14) *and*

$$N(\bar{y};\Theta) \cap \ker D_N^* F(\bar{x},\bar{y}) = \{0\} \ \ \textit{for all} \ \ \bar{y} \in S(\bar{x}) \tag{5.16}$$

are satisfied, and either F *is PSNC at* $(\bar{x},\bar{y})$ *and* Θ *is SNC at* $\bar{y}$*, or* F *is SNC at* $(\bar{x},\bar{y})$ *for all* $\bar{y} \in S(\bar{x})$*.*

Proof. To justify (i) in the Banach space setting, we are based on the first upper subdifferential condition in Proposition 5.2 and then employ the equality for computing $\widehat{N}(\bar{x}; f^{-1}(\Theta))$ from Corollary 1.15.

To prove (ii) when X is Asplund (while Y may be arbitrarily Banach) and f is strictly differentiable at $\bar{x}$ with the surjective derivative, we apply assertion (i) of Theorem 5.5 with $\Omega_1 = f^{-1}(\Theta)$ and $\Omega_2 = \Omega$ assuming that either Ω or $f^{-1}(\Theta)$ is SNC at $\bar{x}$ and $f(\bar{x})$, respectively, and that

$$N(\bar{x}; f^{-1}(\Theta)) \cap \big(- N(\bar{x};\Omega) \big) = \{0\} \, .$$

When Ω is SNC at $\bar{x}$, the result of (ii) follows from Theorem 1.17 providing the normal cone representation

$$N(\bar{x}; f^{-1}(\Theta)) = \nabla f(\bar{x})^* N(f(\bar{x});\Theta) \, .$$

When Ω is not assumed to be SNC at $\bar{x}$, we need to involve the SNC property of $f^{-1}(\Theta)$ at $\bar{x}$, which is equivalent to the SNC property of Θ at $f(\bar{x})$ by Theorem 1.22. This justifies (ii).

To prove assertion (iii), we apply Theorem 5.5(i) for $\Omega_1 = F^{-1}(\Theta)$ and $\Omega_2 = \Omega$. Then we use the upper estimate of $N(\bar{x}; F^{-1}(\Theta))$ for general multifunctions from Theorem 3.8, which requires that both spaces X and Y are Asplund. To employ this theorem, we first observe that the set $F^{-1}(\Theta)$ is locally closed around $\bar{x}$ under the assumptions of (iii); see the proof of Theorem 3.8 noting that $S(\cdot)$ is assumed to be lower semicompact *around* $\bar{x}$. By Theorem 3.8 we get

$$N(\bar{x}; F^{-1}(\Theta)) \subset \bigcup \Big[D_N^* F(\bar{x}, \bar{y})(y^*) \Big| \; \bar{y} \in S(\bar{x}), \; y^* \in N(\bar{y}; \Theta) \Big]$$

under the assumptions on F and Θ made in (a). If now Ω is supposed to be SNC at $\bar{x}$, then we arrive at (5.13) by using the upper subdifferential inclusion

$$-\widehat{\partial}^+ \varphi_0(\bar{x}) \subset N(\bar{x}; F^{-1}(\Theta)) + N(\bar{x}; \Omega)$$

of Theorem 5.5 under the qualification condition

$$N(\bar{x}; F^{-1}(\Theta)) \cap \big(-N(\bar{x}; \Omega) \big) = \{0\} \, .$$

If Ω is not supposed to be SNC at $\bar{x}$, we need to use the SNC property of $F^{-1}(\Theta)$ at $\bar{x}$ that is ensured by Theorem 3.84 under the assumptions made in (b). This completes the proof of the theorem. △

Note that, by Proposition 1.68, the PSNC property of F holds in (b) if F is *Lipschitz-like* around $(\bar{x}, \bar{y})$. Observe also that the result of assertion (ii)in Theorem 5.7 reduces to the one in assertion (iii) of this theorem if X is additionally assumed to be Asplund while Θ is locally closed around $f(\bar{x})$.

Next we derive *lower subdifferential* optimality conditions in the *normal form* for problem (5.12) based on assertions (ii) and (iii) of Theorem 5.5 and employing the calculus results used in the proof of Theorem 5.7.

Theorem 5.8 (lower subdifferential conditions for local minima under operator constraints). *Given a local optimal solution $\bar{x}$ to problem* (5.12)*, suppose that X is Asplund, that Ω is locally closed around $\bar{x}$, and that φ_0 is l.s.c. around this point. Then we have the assertions:*

(i) *Let Y be Banach, and let $F = f\colon X \to Y$ be strictly differentiable at $\bar{x}$ with the surjective derivative $\nabla f(\bar{x})$. Then*

$$0 \in \partial \varphi_0(\bar{x}) + \nabla f(\bar{x})^* N(f(\bar{x}); \Theta) + N(\bar{x}; \Omega)$$

provided that

$$\Big[x^* \in \partial^\infty \varphi_0(\bar{x}), \quad x_1^* \in \nabla f(\bar{x})^* N(f(\bar{x}); \Theta), \quad x_2^* \in N(\bar{x}; \Omega) \, ,$$

$$x^* + x_1^* + x_2^* = 0 \Big] \Longrightarrow x^* = x_1^* = x_2^* = 0$$

and that one of the following requirements holds:

(a) φ_0 *is SNEC at* $\bar{x}$*, and either* Ω *is SNC at* $\bar{x}$ *or* Θ *is SNC at* $f(\bar{x})$*;*

(b) *both* Ω *and* Θ *have the SNC property at* $\bar{x}$ *and* $f(\bar{x})$*, respectively.*

(ii) *Let* Y *be Asplund, let the sets* Θ *and* $\operatorname{gph} F$ *be closed, and let the set-valued mapping* $S(\cdot) = F(\cdot) \cap \Theta$ *be inner semicompact around* $\bar{x}$*. Then*

$$0 \in \partial\varphi_0(\bar{x}) + \bigcup \Big[D_N^* F(\bar{x}, \bar{y})(y^*) \Big| \; \bar{y} \in S(\bar{x}), y^* \in N(\bar{y}; \Theta) \Big] + N(\bar{x}; \Omega) \tag{5.17}$$

under one of the following requirements on $(\varphi_0, F, \Theta, \Omega)$*:*

(c) φ_0 *is SNEC at* $\bar{x}$ *and*

$$\Big[x^* \in \partial^\infty \varphi_0(\bar{x}), \; x_1^* \in \bigcup \Big[D_N^* F(\bar{x}, \bar{y})(y^*) \Big| \; \bar{y} \in S(\bar{x}), \; y^* \in N(\bar{y}; \Theta) \Big], \; x_2^* \in N(\bar{x}; \Omega), \quad x^* + x_1^* + x_2^* = 0 \Big] \Longrightarrow x^* = x_1^* = x_2^* = 0 \tag{5.18}$$

in addition to the assumptions in either (a) *or* (b) *of Theorem* 5.7(iii)*, where* (5.14) *is superseded by* (5.18)*.*

(d) Ω *is SNC at* $\bar{x}$*, the qualification conditions* (5.16) *and* (5.18) *are satisfied, and either* F *is PSNC at* $(\bar{x}, \bar{y})$ *and* Θ *is SNC at* $\bar{y}$*, or* F *is SNC at* $(\bar{x}, \bar{y})$ *for all* $\bar{y} \in S(\bar{x})$*.*

Proof. To prove assertion (i), we base on Theorem 5.5(ii) with $\Omega_1 = f^{-1}(\Theta)$ and $\Omega_2 = \Omega$. Then the desired result in case (a) follows from the representation of the normal cone $N(\bar{x}; f^{-1}(\Theta))$ in the proof of Theorem 5.7(ii). When φ_0 is not assumed to be SNEC at $\bar{x}$, we need to use conditions ensuring the SNC property of the intersection $f^{-1}(\Theta) \cap \Omega$ at $\bar{x}$. Since both sets $f^{-1}(\Theta)$ and Ω are SNC at this point under the assumptions made in (b) and since $\nabla f(\bar{x})$ is surjective, the SNC property of the intersection follows from Corollary 3.81.

The proof of assertion (ii) is similar based on Theorem 5.5(ii) with $\Omega_1 = F^{-1}(\Theta)$ and $\Omega_2 = \Omega$ and the upper estimate of $N(\bar{x}; F^{-1}(\Theta))$ from the proof of Theorem 5.7(iii). This gives the subdifferential inclusion (5.17) in case (c). To justify (5.17) in case (d), we observe that both sets in the intersection $F^{-1}(\Theta)$ and Ω are SNC at $\bar{x}$ under the assumption made, and the qualification condition (5.18) ensures the SNC property of this intersection by Corollary 3.81. This completes the proof of the theorem. △

Note that the result in assertion (i) of Theorem 5.8 follows from the one in assertion (ii) provided that the space Y is Asplund and the set Θ is closed. However, these assumptions are not imposed in (i). Observe also that the qualification conditions (5.15) and (5.16) coincide when X is finite-dimensional while (5.15) is weaker in general. The main advantage of (5.15) is that it always holds together with the PSNC property of F^{-1} at $(\bar{y}, \bar{x})$ if F is *metrically* *regular* around this point. Thus we arrive at following efficient corollary of Theorems 5.7 and 5.8, where the cost function φ_0 is supposed to be locally Lipschitzian to simplify the assumptions in the latter theorem.

Corollary 5.9 (upper and lower subdifferential conditions under metrically regular constraints). *Let $\bar{x}$ be a local optimal solution to problem* (5.12) *under the common assumptions in Theorem* 5.7(iii) *together with* (5.14) *and the SNC requirement on Ω at $\bar{x}$. Suppose also that F is metrically regular around $(\bar{x}, \bar{y})$ for all $\bar{y} \in S(\bar{x})$. Then the upper subdifferential condition* (5.13) *holds. If in addition φ_0 is locally Lipschitzian around $\bar{x}$, then the lower subdifferential condition* (5.17) *holds as well.*

Proof. The upper subdifferential condition (5.15) follows from Theorem 5.7(iii) in case (a) due to the *coderivative characterization* of metric regularity in Theorem 4.18(c). To derive the lower subdifferential condition (5.17) from case (c) of Theorem 5.8(ii), we observe that φ_0 is automatically SNEC at $\bar{x}$ when it is locally Lipschitzian and that (5.18) reduces to (5.14) under this assumption. △

Both upper subdifferential (5.13) and lower subdifferential (5.17) necessary optimality conditions for problem (5.12) admit essential simplifications if F is assumed to be single-valued and *strictly Lipschitzian* at a minimum point $\bar{x}$. This is due to the *scalarization* formula for the normal coderivative established in Theorem 3.28 under the assumption that f is w^*-strictly Lipschitzian. Observe that for mappings between Asplund spaces the notions of strictly Lipschitzian and w^*-strictly Lipschitzian mappings from Definition 3.25 are equivalent by Proposition 3.26.

Corollary 5.10 (upper and lower subdifferential conditions under strictly Lipschitzian constraints). *Let $\bar{x}$ be a local solution to problem* (5.12) *in Asplund spaces X and Y, where $F = f\colon X \to Y$ is single-valued and strictly Lipschitzian at $\bar{x}$. Then one has*

$$-\widehat{\partial}^+\varphi_0(\bar{x}) \subset \bigcup \Big[\partial\langle y^*, f\rangle(\bar{x})\Big|\ y^* \in N(f(\bar{x});\Theta)\Big] + N(\bar{x};\Omega)\,, \tag{5.19}$$

$$0 \in \partial\varphi_0(\bar{x}) + \bigcup \Big[\partial\langle y^*, f\rangle(\bar{x})\Big|\ y^* \in N(f(\bar{x});\Theta)\Big] + N(\bar{x};\Omega) \tag{5.20}$$

under the corresponding assumptions of Theorems 5.7(iii) *and* 5.8(ii), *where $S(\bar{x}) = \{f(\bar{x})\}$ and f is PSNC at $\bar{x}$ automatically.*

Proof. By Theorem 3.28 we have

$$D^*_N f(\bar{x})(y^*) = \partial\langle y^*, f\rangle(\bar{x}) \text{ for all } y^* \in Y^*$$

if $f\colon X \to Y$ is a mapping between Asplund spaces that is strictly Lipschitzian at $\bar{x}$. Thus the upper subdifferential condition (5.13) and lower subdifferential condition (5.17) reduce to (5.19) and (5.20), respectively. △

As we have mentioned, the necessary optimality conditions obtained above for problem (5.12) are given in the *normal/qualified* form under certain *constraint qualifications* that ensure such a normality. What happens if such constraint qualifications are not fulfilled? Then we expect to get necessary conditions in a generalized *non-qualified* form (sometimes called the *Fritz John form*) with a nonnegative (may be zero) multiplier corresponding to the cost function. Let us formulate upper and lower subdifferential conditions in this form that actually follow from Theorems 5.7 and 5.8.

Theorem 5.11 (necessary optimality conditions without constraint qualifications). *Given a local optimal solution $\bar{x}$ to problem* (5.12)*, we have the assertions:*

(i) *Assume that X and Y are Banach, that $\Omega = X$ and $\Theta = \{0\}$, and that $F = f\colon X \to Y$ is Fréchet differentiable at $\bar{x}$. Then there exists $\lambda_0 \geq 0$ such that for every $x^* \in \widehat{\partial}^+\varphi_0(\bar{x})$ there is $y^* \in Y^*$ for which*

$$0 = \lambda_0 x^* + \nabla f(\bar{x})^* y^*, \quad (\lambda_0, y^*) \neq 0\,, \tag{5.21}$$

provided that either f is strictly differentiable at $\bar{x}$ and the image space $\nabla f(\bar{x})X$ is closed in Y, or f is continuous around $\bar{x}$ and $\dim Y < \infty$.

(ii) *Assume that X is Asplund while Y is Banach, that $f\colon X \to Y$ is strictly differentiable at $\bar{x}$ with the surjective derivative $\nabla f(\bar{x})$, and that Ω is locally closed around $\bar{x}$. Then there exists $\lambda_0 \geq 0$ such that for every $x^* \in \widehat{\partial}^+\varphi_0(\bar{x})$ there is $y^* \in N(f(\bar{x});\Theta)$ for which*

$$-\lambda_0 x^* - \nabla f(\bar{x})^* y^* \in N(\bar{x};\Omega), \quad (\lambda_0, y^*) \neq 0\,,$$

provided that either Ω is SNC at $\bar{x}$ or Θ is SNC at $f(\bar{x})$.

(iii) *Assume that both X and Y are Asplund, that Ω and Θ are closed, and that $S(\cdot) = F(\cdot) \cap \Theta$ is inner semicompact around $\bar{x}$. Then there exists $\lambda_0 \geq 0$ such that for every $x^* \in \widehat{\partial}^+\varphi_0(\bar{x})$ there are $\bar{y} \in S(\bar{x})$ and dual elements $y^* \in N(\bar{y};\Theta)$, $x_1^* \in D_N^*F(\bar{x},\bar{y})(y^*)$, and $x_2^* \in N(\bar{x};\Omega)$ satisfying*

$$0 = \lambda_0 x^* + x_1^* + x_2^*, \quad (\lambda_0, y^*, x_1^*) \neq 0\,, \tag{5.22}$$

provided that one of the following properties holds for every $\bar{y} \in S(\bar{x})$:

(a) *Ω is SNC at $\bar{x}$ and F^{-1} is PSNC at $(\bar{y},\bar{x})$;*

(b) *Ω is SNC at $\bar{x}$ and Θ is SNC at $\bar{y}$;*

(c) *F is PSNC at $(\bar{x},\bar{y})$ and Θ is SNC at $\bar{y}$;*

(d) *F is SNC at $(\bar{x},\bar{y})$.*

(iv) *Let φ_0 be locally Lipschitzian around $\bar{x}$ in addition to the assumptions in* (iii)*. Then there are $\lambda_0 \geq 0$, $x^* \in \partial\varphi_0(\bar{x})$, $\bar{y} \in S(\bar{x})$, $y^* \in N(\bar{y};\Theta)$, $x_1^* \in D_N^*F(\bar{x},\bar{y})(y^*)$, and $x_2^* \in N(\bar{x};\Omega)$ such that* (5.22) *holds provided that one of the properties* (a)–(d) *in* (iii) *is fulfilled for every $\bar{y} \in S(\bar{x})$.*

Proof. To prove (i), observe that it follows from Theorem 5.7(i) with the "normal" multiplier $\lambda_0 = 1$ if $\nabla f(\bar{x})\colon X \to Y$ is surjective under the assumptions made. If $\nabla f(\bar{x})$ is not surjective and the space $\nabla f(\bar{x})X$ is closed in Y, then it is easy to show (by the separation theorem; cf. the proof of Theorem 1.57) that $\ker \nabla f(\bar{x})^* \neq \{0\}$, i.e., there is $0 \neq y^* \in Y^*$ such that $\nabla f(\bar{x})^* y^* = 0$. Thus we get (5.21) with $\lambda_0 = 0$ and $y^* \neq 0$.

Let us derive the upper subdifferential conditions in (iii) from the ones in Theorem 5.7(iii) noting that the proof of (ii) is entirely similar (it is actually contained in the proof below) based on assertion (ii) of Theorem 5.7. Observe that Theorem 5.7(iii) implies the desired result of (iii) with $\lambda_0 = 1$ if the qualification conditions (5.14) and (5.16) are satisfied. Assuming the opposite, we need to show that the relations in (iii) hold with $\lambda_0 = 0$ and $(y^*, x_1^*) \neq 0$. Indeed, if (5.14) is not satisfied, then there are $\bar{y} \in S(\bar{x})$ and dual elements $y^* \in N(\bar{y}; \Theta)$ and $0 \neq x^* \in D_N^* F(\bar{x}, \bar{y})(y^*)$ such that $-x^* \in N(\bar{x}; \Omega)$. This gives (5.22) with $\lambda_0 = 0$, $x_1^* = x^*$, and $x_2^* = -x^*$. If (5.16) is not satisfied, then there are $\bar{y} \in S(\bar{x})$ and $0 \neq y^* \in N(\bar{y}; \Theta)$ such that $0 \in D_N^* F(\bar{x}, \bar{y})(y^*)$. This gives (5.22) with $\lambda_0 = 0$, $y^* \neq 0$, and $x_1^* = x_2^* = 0$.

It remains to prove the lower subdifferential necessary conditions in assertion (iv) provided that the cost function φ_0 is Lipschitz continuous around $\bar{x}$. We have mentioned above that under the latter assumption φ_0 is automatically SNEC at $\bar{x}$ and the qualification condition (5.18) reduces to (5.14). Hence we conclude from Theorem 5.8 that (5.22) holds with $\lambda_0 = 1$ and some $x^* \in \partial\varphi_0(\bar{x})$ under the constraint qualifications (5.14) and (5.16). If either (5.14) or (5.16) is not satisfied, we justify(5.22) with $\lambda_0 = 0$ similarly to the proof of the upper subdifferential conditions in assertion (iii). △

Note that assertion (i) of Theorem 5.11 gives a *upper subdifferential extension* of the classical Lyusternik version of the *Lagrange multiplier rule* for problems with equality operator constraints in *Banach spaces* that reduces to our result when f is strictly differentiable at $\bar{x}$. When $\dim Y < \infty$ and f is merely Fréchet differentiable *at* $\bar{x}$, this result also follows from Theorems 6.37 and 6.38 in the case of equality constraints; cf. the proof in Subsect. 6.3.4. It is easy to check that assertions (ii)–(iv) of Theorem 5.11 are actually *equivalent* to the corresponding assertions of Theorems 5.7 and 5.8 if the qualification condition (5.16) is assumed instead of (5.15) in Theorem 5.7 and if φ_0 is assumed to be locally Lipschitzian in Theorem 5.8(ii). In general Theorems 5.7 and 5.8 contain more subtle requirements ensuring the upper subdifferential optimality conditions in the normal form.

It is interesting to observe that the version of the Lagrange multiplier rule in assertion (i) of Theorem 5.11 is *not valid* even in the case of finite-dimensional spaces X, Y and a linear cost function φ_0 if f is assumed to be *merely Fréchet differentiable at* $\bar{x}$ with *no continuity* requirement on it *around* this point. This is demonstrated by the following example.

Example 5.12 (violation of the multiplier rule for problems with Fréchet differentiable but discontinuous equality constraints). *Necessary optimality conditions with Lagrange multipliers don't hold for a two-dimensional problem of minimizing a linear cost function subject to the equality constraint given by a function that is Fréchet differentiable at a point of the global minimum but not continuous around this point.*

Proof. Consider the problem of minimizing $\varphi_0(x_1, x_2) := x_1$ subject to

$$0 = f(x_1, x_2) := \begin{cases} x_2 + x_1^2 & \text{if } x_2 \ge 0 \,, \\ \\ x_2 - x_1^2 & \text{otherwise} \,. \end{cases}$$

It is easy to check that $\bar{x} = (0, 0)$ is a global minimizer for this problem, where f is Fréchet differentiable *at* $\bar{x}$ but not continuous *around* this point. Since $\nabla\varphi_0(0, 0) = (1, 0)$ and $\nabla f(0, 0) = (0, 1)$, the only pair $(\lambda_0, \lambda_1) = (0, 0)$ satisfies the optimality condition (5.21) given by

$$0 = \lambda_0 \nabla\varphi_0(\bar{x}) + \lambda_1 \nabla f(\bar{x}) \,,$$

a contradiction. Note that f is *not strictly* differentiable at $\bar{x}$. △

Let us formulate efficient consequences of Theorem 5.11 in the case of strictly Lipschitzian mappings $F = f \colon X \to Y$ between Asplund spaces.

Corollary 5.13 (strictly Lipschitzian constraints with no qualification). *Let $\bar{x}$ be a local optimal solution to problem* (5.12)*, where X and Y are Asplund, Ω and Θ are closed, and $F = f$ is single-valued and strictly Lipschitzian at $\bar{x}$. Then there exists $\lambda_0 \ge 0$ such that for every $x^* \in \widehat{\partial}^+\varphi_0(\bar{x})$ there is $y^* \in N(f(\bar{x}); \Theta)$ satisfying*

$$-\lambda_0 x^* \in \partial\langle y^*, f\rangle(\bar{x}) + N(\bar{x}; \Omega), \quad (\lambda_0, y^*) \ne 0 \,,$$

provided that one of the following properties is fulfilled:

(a) *Ω is SNC at $\bar{x}$ and f^{-1} is PSNC at $(f(\bar{x}), \bar{x})$;*
(b) *Θ is SNC at $f(\bar{x})$.*

If in addition φ_0 is Lipschitz continuous around $\bar{x}$, then there are $\lambda_0 \ge 0$ and $y^* \in N(f(\bar{x}); \Theta)$ satisfying

$$0 \in \lambda_0 \partial\varphi_0(\bar{x}) + \partial\langle y^*, f\rangle(\bar{x}) + N(\bar{x}; \Omega), \quad (\lambda_0, y^*) \ne 0 \,,$$

provided that either (a) *or* (b) *holds.*

Proof. Both upper and lower subdifferential conditions of the corollary follow directly from Theorem 5.11 and the coderivative scalarization formula, which ensures that $x_1^* = 0$ if $y^* = 0$ in the conditions above. In this case the requirements in (b) and (c) of Theorem 5.11 reduce to the SNC property of Θ at $f(\bar{x})$, since f is automatically PSNC at $\bar{x}$ due to its locally

Lipschitz continuity. Let us show that the SNC property of f in (d) of Theorem 5.11 is redundant in the case of strictly Lipschitzian mappings. Indeed, by Corollary 3.30 such mappings $f\colon X \to Y$ are SNC if and only if Y is finite-dimensional, which is included to the SNC requirement on Θ. Thus properties (a)–(d) of Theorem 5.11 reduce to (a) and (b) in the corollary. △

Remark 5.14 (lower subdifferential conditions via the extremal principle). Note that the *lower subdifferential* (but not upper subdifferential) necessary optimality conditions obtained above can be derived by the direct application of the *extremal principle* with the subsequent use of calculus rules and SNC properties for basic normals to inverse images. Indeed, it is easy to observe that, given a local optimal solution $\bar{x}$ to the constrained problem (5.12), the point $(\bar{x}, \varphi_0(\bar{x}))$ is *locally extremal* for the system of three sets in the space $X \times I\!R$:

$$\Omega_0 := \operatorname{epi}\varphi_0, \quad \Omega_1 := F^{-1}(\Theta) \times I\!R, \quad \Omega_2 := \Omega \times I\!R\,.$$

Applying the exact extremal principle from Theorem 2.22 to this system and then using the calculus results as above, we arrive at necessary conditions for $\bar{x}$ of the subdifferential type expressed in terms of basic normals and subgradients. Note that this way leads us not only to *exact/pointbased* optimality conditions of the above type but also to necessary conditions in an *approximate/fuzzy* form expressed via Fréchet normals and subgradients at points nearby the local minimizer *without any SNC* assumptions. To derive necessary conditions of the latter type, one needs to employ the approximate version of the extremal principle from Theorem 2.20 and then the corresponding rules of *fuzzy calculus*; see Theorem 1.14, Lemma 3.1, and Remark 3.21. We are going to present more results of this direction in the subsequent parts of this chapter for special classes of constrained optimization problems (5.12) and their multiobjective counterparts.

This subsection is concluded by considering a special class of optimization problems with *operator constraints of the equality type* $f(x) = 0$ given by single-valued mappings with *infinite-dimensional* range spaces. Note that the specific feature of the latter constraints in comparison with the general ones in problem (5.12) is that the set $\Theta = \{0\}$ is *never SNC* unless the range space for f is finite-dimensional.

We explore a fruitful approach to necessary optimality conditions for such problems, under additional finitely many inequality constraints as well as that of the geometric type, based on reducing the constrained problems to *unconstrained* minimization by some *exact penalization* technique. This reduction becomes possible under the following *weakened metric regularity* property of operator constraint mappings relative to geometric constraints *at* the reference point versus to *around* it as in Definition 1.47.

Definition 5.15 (weakened metric regularity). *A single-valued mapping $f\colon X \to Y$ between Banach spaces is* METRICALLY REGULAR AT *a point $\bar{x} \in \Omega$*

RELATIVE *to a set* $\Omega \subset X$ *if there are a constant* $\mu > 0$ *and a neighborhood* U *of* $\bar{x}$ *such that*

$$\operatorname{dist}(x; S) \le \mu \|f(x) - f(\bar{x})\| \quad \text{for all} \quad x \in U \cap \Omega ,$$

where $S := \{x \in \Omega |\ f(x) = f(\bar{x})\}$.

It is easy to see that the above regularity holds if the Ω-*restrictive* mapping $f_\Omega(x) := f(x) + \Delta(x; \Omega)$ defined on the whole space X is locally metrically regular *around* $\bar{x}$ in the sense of Definition 1.47(ii). Thus the *sufficient* conditions for the latter metric regularity established in Chap. 4 ensure the fulfillment of the Ω-relative metric regularity of f at the reference point $\bar{x}$. It is not hard to observe that they are definitely *necessary* for the weakened metric regularity of nonsmooth mappings. This is largely related to the fact that the metric regularity concept from Definition 5.15 is *not robust* with respect to perturbations of the initial point, in contrast to the case of Definition 1.47.

The next result establishes the desired reduction of constrained optimization problems of the mentioned type to unconstrained problems via a certain exact penalization, which is convenient for the subsequent applications to necessary conditions of the *lower subdifferential* type in constrained minimization.

Theorem 5.16 (exact penalization under equality constraints). *Let* $\bar{x}$ *be a local optimal solution to the constrained problem* **(CP)***:*

$$\text{minimize } \varphi_0(x) \text{ subject to } \varphi_i(x) \le 0,\ i = 1, \ldots, m, \quad f(x) = 0, \quad x \in \Omega ,$$

where $f: X \to Y$ *is a mapping between Banach spaces, and where* φ_i *are real-valued functions. Assume that* f *is locally Lipschitzian around* $\bar{x}$ *and metrically regular at this point relative to* Ω*. Denoting*

$$I(\bar{x}) := \big\{i \in \{1, \ldots, m\} \big|\ \varphi_i(\bar{x}) = 0\big\} ,$$

we suppose also that the functions φ_i *are locally Lipschitzian around* $\bar{x}$ *for* $i \in I(\bar{x}) \cup \{0\}$ *and upper semicontinuous at* $\bar{x}$ *for* $i \in \{1, \ldots, m\} \setminus I(\bar{x})$*. Then* $\bar{x}$ *is a local optimal solution to the unconstrained problem* **(UP)** *of minimizing the objective:*

$$\max\Big\{\varphi_0(x) - \varphi_0(\bar{x}),\ \max_{i \in I(\bar{x})} \varphi_i(x)\Big\} + \mu\big(\|f(x)\| + \operatorname{dist}(x; \Omega)\big)$$

for all $\mu > 0$ *sufficiently large.*

Proof. It is easy to see that $\bar{x}$ is a local solution to the problem of minimizing

$$\varphi(x) := \max\Big\{\varphi_0(x) - \varphi_0(\bar{x}), \max_{i \in I(\bar{x})} \varphi_i(x)\Big\} \text{ subject to } f(x) = 0, \quad x \in \Omega$$

under the assumptions imposed on φ_i. Since f is continuous and metrically regular at $\bar{x}$ relative to Ω, there exist a number $\mu_1 > 0$ and a neighborhood U of $\bar{x}$ such that for any $x \in U \cap \Omega$ there is $u \in \Omega$ satisfying

$$\varphi(u) \geq \varphi(\bar{x}), \quad f(u) = 0, \quad \|x - u\| \leq \mu_1 \|f(x)\| .$$

Let ℓ be a common Lipschitz constant for φ and f on U, and let $\mu_2 \geq \ell\mu_1$. Then for any $x \in U \cap \Omega$ and the above $u \in \Omega$ corresponding to x one has

$$\varphi(x) \geq \varphi(x) - \varphi(u) + \varphi(\bar{x}) \geq -\ell\|x - u\| + \varphi(\bar{x})$$

$$\geq -\ell\mu_1\|f(x)\| + \varphi(\bar{x}) \geq -\mu_2\|f(x)\| + \varphi(\bar{x}) ,$$

i.e., $\bar{x}$ is a local solution to the problem

$$\text{minimize } \varphi(x) + \mu_2\|f(x)\| \text{ subject to } x \in \Omega .$$

Observe now that $x \in \Omega$ is equivalent to $\operatorname{dist}(x;\Omega) = 0$, that the latter function is obviously metrically regular at $\bar{x}$ relative to Ω, and that $\varphi(x) + \mu_2\|f(x)\|$ is Lipschitz. Using the above arguments, we find $\mu_3 > 0$ such that $\bar{x}$ is a local solution to the problem:

$$\text{minimize } \varphi(x) + \mu_2\|f(x)\| + \mu_3\operatorname{dist}(x;\Omega) .$$

To complete the proof of the theorem, it remains to take $\mu := \max\{\mu_2, \mu_3\}$. $\triangle$

Based on the above exact penalization result and employing subdifferential and SNC calculus results of Chap. 3 together with pointbased coderivative criteria of metric regularity from Chap. 4, we derive efficient conditions for optimal solutions to constrained problems of the (CP) type treated in Theorem 5.16.

Theorem 5.17 (necessary conditions for problems with operator constraints of equality type). *Let $\bar{x}$ be a local optimal solution to problem (CP), where both spaces X and Y are Asplund, where the functions φ_i satisfy the assumptions of Theorem* 5.16*, and where the set Ω is locally closed around $\bar{x}$. Assume also that the mapping f is strictly Lipschitzian at $\bar{x}$ and such that f_Ω^{-1} is PSNC at $(f(\bar{x}), \bar{x})$. Then there are numbers $\lambda_i \geq 0$ for $i \in I(\bar{x}) \cup \{0\}$ and a linear functional $y^* \in Y^*$ not equal to zero simultaneously and satisfying*

$$0 \in \partial\Big(\sum_{i \in I(\bar{x}) \cup \{0\}} \lambda_i\varphi_i\Big)(\bar{x}) + \partial\langle y^*, f\rangle(\bar{x}) + N(\bar{x};\Omega) .$$

Proof. Assume first that f is metrically regular at $\bar{x}$ relative to Ω. Then there is $\mu > 0$ such that $\bar{x}$ is a local optimal solution to the unconstrained problem (UP) in Theorem 5.16. Hence

$$0 \in \partial\Big(\max\big\{\varphi_0(\cdot) - \varphi_0(\bar{x}),\ \max_{i \in I(\bar{x})} \varphi_i(\cdot)\big\} + \mu\big(\|f(\cdot)\| + \operatorname{dist}(\cdot;\Omega)\big)\Big)(\bar{x}) .$$

Applying now the subdifferential sum rule from Theorem 3.36 to the latter function and then using the maximum rule from Theorem 3.46(ii), the chain rule from Corollary 3.43 for the composition $\|f(x)\| = (\psi \circ f)(x)$ with $\psi(y) := \|y\|$, and the subdifferential formula for the distance function $\mathrm{dist}(x; \Omega)$ from Theorem 1.97, we arrive at the necessary optimality conditions of the theorem with $\big(\lambda_i \big|\, i \in I(\bar{x}) \cup \{0\}\big) \neq 0$.

If f is not supposed to be metrically regular at $\bar{x}$ relative to Ω, then mapping $f_\Omega(x) := f(x) + \Delta(x; \Omega)$ is *not* metrically regular around $\bar{x}$ in the sense of Definition 1.47(ii). By Theorem 4.18(c) this happens when either $\ker \widetilde{D}^*_M f_\Omega(\bar{x}) \neq \{0\}$ or f_Ω^{-1} is not PSNC at $(f(\bar{x}), \bar{x})$. The latter is impossible due to the assumption of this theorem. Thus there is $y^* \neq 0$ such that

$$0 \in \widetilde{D}^*_M f_\Omega(\bar{x})(y^*) \subset D^*_N f_\Omega(\bar{x})(y^*) = D^*_N\big(f + \Delta(\cdot; \Omega)\big)(\bar{x})(y^*)\ .$$

Using the coderivative sum rule from Proposition 3.12 whose qualification assumption holds due to Lipschitz continuity of f and then employing the scalarization formula of Theorem 3.28, since f is strictly Lipschitzian, we arrive at the inclusion

$$0 \in D^*_N f(\bar{x})(y^*) + N(\bar{x}; \Omega) = \partial\langle y^*, f\rangle(\bar{x}) + N(\bar{x}; \Omega)\ .$$

This ensures the conclusion of the theorem with $y^* \neq 0$. △

Note that if f is assumed to be *merely Lipschitz* continuous around $\bar{x}$ (but not strictly Lipschitzian at this point), then the conclusion of Theorem 5.17 holds in the form of

$$0 \in \partial\Big(\sum_{i \in I(\bar{x}) \cup \{0\}} \lambda_i \varphi_i\Big)(\bar{x}) + D^*_N f(\bar{x})(y^*) + N(\bar{x}; \Omega)$$

with $(\lambda_i, y^*) \neq 0$. This directly follows from the proof of the theorem.

The next corollary describes a broad class of operator constraints involving *generalized Fredholm* mappings that satisfy the assumptions of the above theorem. This result is especially important for applications to problems of optimal control; see Chap. 6.

Corollary 5.18 (necessary conditions for optimization problems with generalized Fredholm operator constraints). *Let $\bar{x}$ be a local optimal solution to the above problem (CP) with operator constraints. Assume that f is generalized Fredholm at $\bar{x}$, that Ω is SNC at $\bar{x}$, and that all the other data in (CP) satisfy the assumptions of Theorem* 5.17*. Then the necessary optimality conditions of the theorem hold.*

Proof. As proved in Theorem 3.35, f_Ω^{-1} is PSNC at $(f(\bar{x}), \bar{x})$ under the assumptions imposed on f and Ω. Since every compactly strictly Lipschitzian mapping is automatically strictly Lipschitzian and the addition of a linear bounded operator doesn't violate this property, we conclude that f is strictly Lipschitzian at $\bar{x}$ and thus complete the proof of the corollary. △

5.1.3 Necessary Conditions under Functional Constraints

In this subsection we study in more detail a special class of the constrained problems (5.12) having finitely many *functional constraints* of *equality and inequality* types defined by real-valued functions on infinite-dimensional spaces. Namely, given $\varphi_i\colon X \to I\!R$ for $i = 0, 1, \ldots, m + r$ and $\Omega \subset X$, we consider the following problem of *nondifferentiable programming*:

$$\begin{cases} \text{minimize } \varphi_0(x) \quad \text{subject to} \\ \varphi_i(x) \le 0, \quad i = 1, \ldots, m\ , \\ \varphi_i(x) = 0, \quad i = m + 1, \ldots, m + r\ , \\ x \in \Omega\ . \end{cases} \tag{5.23}$$

Note that the functions φ_i may be extended-real-valued while the assumption about their real-valuedness doesn't restrict the generality due to the additional geometric constraints in (5.23). It is clear that (5.23) is a particular case of (5.12) with $F = (\varphi_1, \ldots, \varphi_{m+r})\colon X \to I\!R^{m+r}$ and

$$\Theta = \Big\{(\alpha_1, \ldots, \alpha_{m+r}) \in I\!R^{m+r} \Big|\ \alpha_i \le 0 \ \text{ for } \ i = 1, \ldots, m \quad \text{and} \\ \alpha_i = 0 \ \text{ for } \ i = m + 1, \ldots, m + r\Big\}\ . \tag{5.24}$$

Thus the results of the preceding subsection directly imply necessary optimality conditions for problem (5.23) by taking into account the form of the set Θ in (5.24). However, the specific structure of (5.23) allows us to derive also more subtle necessary conditions for local minima than those induced by the general scheme (5.12).

Let us first obtain *upper subdifferential* conditions for local minima in (5.23). The next theorem contains new results that are specific for problems with *inequality constraints* together with necessary optimality conditions for (5.23) that follow from the results of Subsect. 5.1.2. As always, we use the common coderivative symbol D^* for the basic coderivatives of mappings with values in finite-dimensional spaces. For brevity we present only necessary optimality conditions without constraint qualifications; the normal counterparts of these conditions either follow from the corresponding results of the preceding subsection or can be derived in a similar way.

Theorem 5.19 (upper subdifferential conditions in nondifferentiable programming). *Let $\bar{x}$ be a local optimal solution to problem* (5.23)*, where the set Ω is locally closed around $\bar{x}$ and the functions φ_i are continuous around this point for $i = m + 1, \ldots, m + r$. The following assertions hold:*

(i) *Assume that X admits a Lipschitzian $\mathcal{C}^1$ bump function* (*this is automatic when X admits a Fréchet differentiable renorm, in particular, when X*

is reflexive), and that either Ω or $f := (\varphi_{m+1}, \ldots, \varphi_{m+r})$ is SNC at $\bar{x}$. Then for any Fréchet upper subgradients $x_i^* \in \widehat{\partial}^+\varphi_i(\bar{x})$, $i = 0, \ldots, m$, there are $(\lambda_0, \ldots, \lambda_{m+r}) \in I\!R^{m+r+1}$, $x^* \in D^*f(\bar{x})(\lambda_{m+1}, \ldots, \lambda_{m+r})$, and $\widetilde{x}^* \in N(\bar{x}; \Omega)$ satisfying the relations

$$\lambda_i \geq 0 \;\; \text{for} \;\; i = 0, \ldots, m, \quad \lambda_i\varphi_i(\bar{x}) = 0 \;\; \text{for} \;\; i = 1, \ldots, m\;, \tag{5.25}$$

$$0 = \sum_{i=0}^{m} \lambda_i x_i^* + x^* + \widetilde{x}^*, \quad (\lambda_0, \ldots, \lambda_{m+r}, x^*) \neq 0\;. \tag{5.26}$$

If φ_i are Lipschitz continuous around $\bar{x}$ for $i = m+1, \ldots, m+r$, then in addition to (5.25) *one has*

$$-\sum_{i=0}^{m} \lambda_i x_i^* \in \partial\Big(\sum_{i=m+1}^{m+r} \lambda_i\varphi_i\Big)(\bar{x}) + N(\bar{x}; \Omega), \quad (\lambda_0, \ldots, \lambda_{m+r}) \neq 0\;, \tag{5.27}$$

with no other assumptions on (φ_i, Ω) besides the local closedness of Ω.

(ii) *Assume that X is Asplund, that $f := (\varphi_1, \ldots, \varphi_{m+r})$ is continuous around $\bar{x}$, and that either Ω or f is SNC at $\bar{x}$. Then there exists $\lambda_0 \geq 0$ such that for every Fréchet upper subgradient $x_0^* \in \widehat{\partial}^+\varphi_0(\bar{x})$ there are $(\lambda_1, \ldots, \lambda_{m+r}) \in I\!R^{m+r}$, $x^* \in D^*f(\bar{x})(\lambda_1, \ldots, \lambda_{m+r})$, and $\widetilde{x}^* \in N(\bar{x}; \Omega)$ satisfying* (5.25) *and*

$$0 = \lambda_0 x_0^* + x^* + \widetilde{x}^*, \quad (\lambda_0, \ldots, \lambda_{m+r}, x^*) \neq 0\;. \tag{5.28}$$

If φ_i are Lipschitz continuous around $\bar{x}$ for $i = 1, \ldots, m+r$, then in addition to (5.25) *one has*

$$-\lambda_0 x_0^* \in \partial\Big(\sum_{i=1}^{m+r} \lambda_i\varphi_i\Big)(\bar{x}) + N(\bar{x}; \Omega), \quad (\lambda_0, \ldots, \lambda_{m+r}) \neq 0\;, \tag{5.29}$$

with no other assumptions on (φ_i, Ω) besides the local closedness of Ω.

Proof. To prove (i) under the general assumptions made, we take arbitrary elements $x_i^* \in \widehat{\partial}^+\varphi_i(\bar{x})$ for $i = 0, \ldots, m$ and apply the variational description from Theorem 1.88(ii) with $\mathcal{S} = \mathcal{LC}^1$ to the subgradients $-x_i^* \in \widehat{\partial}(-\varphi_i)(\bar{x})$. In this way we find functions $s_i\colon X \to I\!R$ for $i = 0, \ldots, m$ satisfying

$$s_i(\bar{x}) = \varphi_i(\bar{x}) \;\; \text{and} \;\; s_i(x) \geq \varphi_i(x) \;\; \text{around} \;\; \bar{x}$$

such that each $s_i(x)$ is continuously differentiable around $\bar{x}$ with $\nabla s_i(\bar{x}) = x_i^*$. It is easy to check that $\bar{x}$ is a local solution to the following optimization problem of type (5.23) but with the cost and inequality constraint functions continuously differentiable around $\bar{x}$:

$$\begin{cases} \text{minimize} \;\; s_0(x) \quad \text{subject to} \\ s_i(x) \le 0, \quad i = 1, \ldots, m \;, \\ \varphi_i(x) = 0, \quad i = m+1, \ldots, m+r \;, \\ x \in \Omega \;. \end{cases} \tag{5.30}$$

Apply now the necessary conditions of Theorem 5.11(iii) to problem (5.30), which corresponds to (5.12) with the single-valued mapping

$$F := (s_1, \ldots, s_m, \varphi_{m+1}, \ldots, \varphi_{m+r})$$

and the set Θ defined in (5.24). Observe that

$$N\big((\varphi_1(\bar{x}), \ldots, \varphi_{m+r}(\bar{x})); \Theta\big) = \Big\{ (\lambda_1, \ldots, \lambda_{m+r}) \in I\!R^{m+r} \Big| \; \lambda_i \ge 0 \;,$$

$$\lambda_i \varphi_i(\bar{x}) = 0 \;\text{ for }\; i = 1, \ldots, m \Big\}$$

with $s_i(\bar{x}) = \varphi_i(\bar{x})$, $i = 1, \ldots, m$, and that

$$F(x) = \big(s(x), 0\big) + \big(0, \varphi_{m+1}(x), \ldots, \varphi_{m+r}(x)\big) \tag{5.31}$$

for the above F, where $s := (s_1, \ldots, s_m)\colon X \to I\!R^m$ is continuously differentiable around $\bar{x}$. Thus the condition $y^* \in N(\bar{y}; \Theta)$ in Theorem 5.11(iii) with $y^* = (\lambda_1, \ldots, \lambda_{m+r})$ reduces to the sign and complementary slackness conditions in (5.25) as $i = 1, \ldots, m$.

Since $Y = I\!R^{m+r}$ in Theorem 5.11(iii), the SNC and PSNC properties of F in (5.31) are equivalent to the SNC property of $f = (\varphi_{m+1}, \ldots, \varphi_{m+r})$ by Theorem 1.70. It is easy also to see that one of the requirements (a)–(d) in Theorem 5.11(iii) holds if and only if either Ω or f is SNC at $\bar{x}$. The coderivative sum rule from Theorem 1.62(ii) applied to the sum in (5.31) ensures that relation (5.22) with $x_1^* \in D^*F(\bar{x}, \bar{y})(y^*)$ and $x_2^* \in N(\bar{x}; \Omega)$ therein is equivalent to the conditions

$$0 = \sum_{i=0}^{m} \lambda_i \nabla s_i(\bar{x}) + x^* + \widetilde{x}^*, \quad (\lambda_0, \ldots, \lambda_{m+r}, \widetilde{x}^*) \ne 0 \;,$$

with $x^* \in D^* f(\bar{x})(\lambda_{m+1}, \ldots, \lambda_{m+r})$, $\widetilde{x}^* \in N(\bar{x}; \Omega)$, and $\lambda_0 \ge 0$. Recalling that $\nabla s_i(\bar{x}) = x_i^*$ for $i = 0, \ldots, m$, we arrive at (5.26). To derive (5.27) from (5.26) when φ_i are locally Lipschitzian for $i = m+1, \ldots, m+r$, it is sufficient to observe that f is automatically SNC at $\bar{x}$ in this case and then to apply the scalarization formula to the coderivative $D^* f(\bar{x})$, which gives

$$D^* f(\bar{x})(\lambda_{m+1}, \ldots, \lambda_{m+r}) = \partial\Big(\sum_{i=m+1}^{m+r} \lambda_i \varphi_i \Big)(\bar{x}) \;.$$

It remains to prove (ii). To proceed, we use directly Theorem 5.11(iii) with $F = f := (\varphi_1, \ldots, \varphi_{m+r})$ and Θ defined in (5.24). In this way one has (5.25) and (5.28) under the general assumptions made in (ii) with some $x^* \in D^* f(\bar{x})(\lambda_1, \ldots, \lambda_{m+r})$. When all $\varphi_1, \ldots, \varphi_{m+r}$ are Lipschitz continuous around $\bar{x}$, the latter implies (5.29) by the coderivative scalarization. △

Note that the necessary conditions of Theorem 5.19 are given in terms of either coderivatives of the "condensed" mappings $(\varphi_{m+1}, \ldots, \varphi_{m+r})\colon X \to I\!R^r$ and $(\varphi_1, \ldots, \varphi_{m+r})\colon X \to I\!R^{m+r}$ or via subgradients of the sums in (5.27) and (5.29). Based on coderivative and subdifferential calculus rules, they may be expressed in a *separated* form involving coderivatives and subgradients of *single* functions φ_i by some weakening of the results. In particular, for the coderivative result of Theorem 5.19 it can be done by applying the coderivative sum rule of Theorem 3.10 to

$$f(x) = \big(\varphi_{m+1}(x), 0, \ldots, 0\big) + \ldots + \big(0, \ldots, 0, \varphi_{m+r}(x)\big)$$

and then by using Theorems 1.80 and 2.40 to express coderivatives of φ_i via basic and singular subgradients of both φ_i and $-\varphi_i$. For brevity we present the results of this type just for Lipschitzian functions φ_i when the corresponding conditions simply follow from the subdifferential calculus rule of Theorem 3.36. In this case it is convenient to use the two-sided *symmetric subdifferential*

$$\partial^0 \varphi(\bar{x}) := \partial \varphi(\bar{x}) \cup \partial^+ \varphi(\bar{x})$$

for each function φ_i, $i = m+1, \ldots, m+r$, describing the equality constraints in the optimization problem (5.23) under consideration.

Corollary 5.20 (upper subdifferential conditions with symmetric subdifferentials for equality constraints). *Let $\bar{x}$ be a local optimal solution to problem* (5.23), *where the set Ω is locally closed around $\bar{x}$ and the functions φ_i are Lipschitz continuous around this point for $i = m+1, \ldots, m+r$. Then the following assertions hold:*

(i) *Assume that X admits a Lipschitzian $\mathcal{C}^1$ bump function. Then for any $x_i^* \in \widehat{\partial}^+ \varphi_i(\bar{x})$, $i = 0, \ldots, m$, there are multipliers $(\lambda_0, \ldots, \lambda_{m+r}) \neq 0$ satisfying* (5.25) *and such that*

$$-\sum_{i=0}^{m} \lambda_i x_i^* \in \sum_{i=m+1}^{m+r} \lambda_i \partial^0 \varphi_i(\bar{x}) + N(\bar{x}; \Omega) .$$

(ii) *Assume that X is Asplund and that φ_i are Lipschitz continuous around $\bar{x}$ for $i = 1, \ldots, m$ as well. Then there is $\lambda_0 \geq 0$ such that for every Fréchet upper subgradient $x_0^* \in \widehat{\partial}^+ \varphi_0(\bar{x})$ there are multipliers $(\lambda_1, \ldots, \lambda_{m+r}) \in I\!R^{m+r}$ satisfying* (5.25) *and*

$$-\lambda_0 x_0^* \in \sum_{i=1}^{m} \lambda_i \partial \varphi_i(\bar{x}) + \sum_{i=m+1}^{m+r} \lambda_i \partial^0 \varphi_i(\bar{x}) + N(\bar{x}; \Omega), \quad (\lambda_0, \ldots, \lambda_{m+r}) \neq 0 .$$

Proof. The inclusion in (i) follows from (5.27) due to the subdifferential sum rule in Theorem 3.36 and the relationships

$$\partial(\lambda\varphi)(\bar{x}) = \lambda\partial\varphi(\bar{x}) \ \text{ for } \ \lambda \ge 0 \quad \text{and} \quad \partial(\lambda\varphi)(\bar{x}) \subset \lambda\partial^0\varphi(\bar{x}) \ \text{ for } \ \lambda \in I\!R\,.$$

Similarly we derive the inclusion in (ii) from (5.29) in Theorem 5.19(ii). △

Another way (actually more precise than in Corollary 5.20) to describe necessary optimality conditions in terms of single functions for problems with equality constraints, is to use the *even subdifferential set* for φ at $\bar{x}$ given by

$$\partial\varphi(\bar{x}) \cup \partial(-\varphi)(\bar{x})$$

with only *nonnegative* multipliers. This is due to

$$\partial(\lambda\varphi)(\bar{x}) \subset |\lambda|\big[\partial\varphi(\bar{x}) \cup \partial(-\varphi)(\bar{x})\big] \ \text{ for all } \ \lambda \in I\!R\,. \tag{5.32}$$

We are going to use this description in what follows. Note that the above "even" set is the same for the functions φ and $-\varphi$; this is where the name comes from, although the set $\partial\varphi(\bar{x}) \cup \partial(-\varphi)(\bar{x})$ doesn't reduce to the classical gradient when φ is smooth.

Next let us derive necessary optimality conditions of the *lower subdifferential* type for problem (5.23) with inequality, equality, and geometric constraints. By results of this type we mean, similarly to Subsect. 5.1.2, such necessary optimality conditions that involve, instead of upper subgradients of the cost and inequality constraint functions, their lower subgradients or normal vectors to their epigraphs. We obtain several results in this direction depending on the assumptions made on the initial data by using different techniques. As in the case of upper subdifferential results, we focus on general optimality conditions without constraint qualifications related to the normal (qualified) form in the same way as in preceding subsection.

The first theorem of this type provides necessary optimality conditions in problem (5.23) given via normals and subgradients for *each constraint* separately. It is based on the direct application of the extremal principle even without using any calculus rule. We present necessary conditions in the *approximate* and *exact* forms depending on the corresponding version of the extremal principle used in the proof. The latter conditions are also specified for problems with Lipschitzian data.

Theorem 5.21 (necessary conditions via normals and subgradients of separate constraints). *Let $\bar{x}$ be a local optimal solution to problem* (5.23)*, where the space X is Asplund and the set Ω is locally closed around $\bar{x}$. The following assertions hold:*

(i) *Assume that the functions φ_i are l.s.c. around $\bar{x}$ for $i = 0, \ldots, m$ and continuous around this point for $i = m+1, \ldots, m+r$. Then for any $\varepsilon > 0$ there are points*

$$(x_0,\alpha_0)\in\operatorname{epi}\varphi_0\cap\big[(\bar{x},\varphi_0(\bar{x}))+\varepsilon I\!B\big],\quad \widehat{x}\in\Omega\cap\big(\bar{x}+\varepsilon I\!B\big)\ ,$$

$$(x_i,\alpha_i)\in\operatorname{epi}\varphi_i\cap\big[(\bar{x},0)+\varepsilon I\!B\big],\quad i=1,\ldots,m\ ,$$

$$(x_i,\alpha_i)\in\operatorname{gph}\varphi_i\cap\big[(\bar{x},0)+\varepsilon I\!B\big],\quad i=m+1,\ldots,m+r\ ,$$

and dual elements

$$(x_i^*,-\lambda_i)\in\widehat{N}((x_i,\alpha_i);\operatorname{epi}\varphi_i)+\varepsilon I\!B^*,\quad i=0,\ldots,m\ ,$$

$$(x_i^*,-\lambda_i)\in\widehat{N}((x_i,\alpha_i);\operatorname{gph}\varphi_i)+\varepsilon I\!B^*,\quad i=m+1,\ldots,m+r\ ,$$

$$\widehat{x}^*\in\widehat{N}(\widehat{x};\Omega)+\varepsilon I\!B^*$$

satisfying the relations

$$x_0^*+\ldots+x_{m+r}^*+\widehat{x}^*=0\ , \tag{5.33}$$

$$\|(x_0^*,\lambda_0)\|+\ldots+\|(x_{m+r}^*,\lambda_{m+r})\|+\|\widehat{x}^*\|=1\ . \tag{5.34}$$

(ii) *Assume that all but one of the sets* $\operatorname{epi}\varphi_i$ $(i=0,\ldots,m)$*,* $\operatorname{gph}\varphi_i$ $(i=m+1,\ldots,m+r)$*, and* Ω *are SNC at the points* $(\bar{x},\varphi_0(\bar{x}))$*,* $(\bar{x},0)$*, and* $\bar{x}$*, respectively. Then there are*

$$(x_0^*,-\lambda_0)\in N((\bar{x},\varphi_0(\bar{x}));\operatorname{epi}\varphi_0),\quad \widehat{x}^*\in N(\widehat{x};\Omega)\ ,$$

$$(x_i^*,-\lambda_i)\in N((\bar{x},0);\operatorname{epi}\varphi_i)\ \text{ for }\ i=1,\ldots,m\ ,$$

$$(x_i^*,-\lambda_i)\in N((\bar{x},0);\operatorname{gph}\varphi_i)\ \text{ for }\ i=m+1,\ldots,m+r$$

satisfying relations (5.33) *and* (5.34) *with* $\lambda_i\ge 0$ *for* $i=0,\ldots,m$*. If in addition* φ_i *is assumed to be upper semicontinuous at* $\bar{x}$ *for those* $i=1,\ldots,m$ *where* $\varphi_i(\bar{x})<0$*, then*

$$\lambda_i\varphi_i(\bar{x})=0\ \text{ for }\ i=1,\ldots,m\ .$$

(iii) *Assume that the functions* φ_i *are Lipschitz continuous around* $\bar{x}$ *for all* $i=0,\ldots,m+r$*. Then there are multipliers* $(\lambda_0,\ldots,\lambda_{m+r})\ne 0$ *such that*

$$0\in\sum_{i=0}^{m}\lambda_i\partial\varphi_i(\bar{x})+\sum_{i=m+1}^{m+r}\lambda_i\Big[\partial\varphi_i(\bar{x})\cup\partial(-\varphi_i)(\bar{x})\Big]+N(\bar{x};\Omega)\ ,$$

$$\lambda_i\ge 0\ \text{ for all }\ i=0,\ldots,m+r,\quad \text{and}\ \ \lambda_i\varphi_i(\bar{x})=0\ \text{ for }\ i=1,\ldots,m\ .$$

Proof. To prove (i), we assume without loss of generality that $\varphi_0(\bar{x}) = 0$. Then it is easy to observe that $(\bar{x}, 0)$ is a local extremal point of the following system of closed sets in the Asplund space $X \times I\!R^{m+r+1}$:

$$\Omega_i := \{(x, \alpha_0, \ldots, \alpha_{m+r}) \mid \alpha_i \geq \varphi_i(x)\}, \quad i = 0, \ldots, m \ ,$$

$$\Omega_i := \{(x, \alpha_0, \ldots, \alpha_{m+r}) \mid \alpha_i = \varphi_i(x)\}, \quad i = m+1, \ldots, m+r \ ,$$

$$\Omega_{m+r+1} := \Omega \times \{0\} \ .$$

Now approximate optimality conditions in (i) follow directly from the approximate version of the extremal principle in Theorem 2.20. Similarly applying the exact version of the extremal principle under the SNC assumptions in Theorem 2.22, we find elements (x^*, λ_i) and $\widehat{x}^*$ satisfying (5.33), (5.34), and the normal cone inclusions in (ii). It follows from Proposition 1.76 on basic normals to epigraphs that $\lambda_i \geq 0$ for $i = 0, \ldots, m$. To establish (ii), it remains to show that the complementary slackness conditions hold under the additional assumption on φ_i. Indeed, if $\varphi_i(\bar{x}) < 0$ for some $i \in \{1, \ldots, m\}$, then $\varphi_i(x) < 0$ for all x around $\bar{x}$ provided that φ_i is upper semicontinuous at $\bar{x}$. The latter implies that $(\bar{x}, 0)$ is an interior point of the epigraph of φ_i. Thus $N((\bar{x}, 0); \operatorname{epi} \varphi_i) = \{0\}$ and $x_i^* = \lambda_i = 0$ for this i, which completes the proof of assertion (ii).

To prove (iii), we observe by Proposition 1.76 and Corollary 1.81 that

$$(x^*, -\lambda) \in N((\bar{x}, \varphi(\bar{x})); \operatorname{epi} \varphi) \Longleftrightarrow x^* \in \lambda \partial \varphi(\bar{x}), \ \lambda \geq 0$$

if φ is Lipschitz continuous around $\bar{x}$. On the other hand,

$$(x^*, -\lambda) \in N((\bar{x}, \varphi(\bar{x})); \operatorname{gph} \varphi) \Longleftrightarrow x^* \in D^* \varphi(\bar{x})(\lambda) = \partial \langle \lambda, \varphi \rangle (\bar{x})$$

by the coderivative scalarization for locally Lipschitzian functions. Invoking finally (5.32) and taking (5.33) and (5.34) into account, we complete the proof of (iii) and the whole theorem. △

Remark 5.22 (comparison between different forms of necessary optimality conditions). Similarly to Corollary 5.20 we can write down necessary optimality conditions in more conventional form replacing, for the case of *equality constraints*, the *even* subdifferential set $\partial \varphi_i(\bar{x}) \cup \partial(-\varphi_i)(\bar{x})$ with a *nonnegative* multiplier λ_i by the *two-sided* symmetric subdifferential $\partial^0 \varphi_i(\bar{x})$ with an *arbitrary* multiplier λ_i. It follows from the definitions that the symmetric form of necessary conditions is more precise than the latter one. Let us illustrate this by thefollowing example in $I\!R^2$:

$$\text{minimize } \ x_1 \ \text{ subject to } \ \varphi(x_1, x_2) := |\,|x_1| + x_2| + x_1 = 0 \ .$$

Based on the computation of subgradients for the function φ in Example 2.49, we conclude that $\bar{x} = (0, 0)$ is not an optimal solution to the above problem due to Theorem 5.21(iii), while the usage of $\lambda\partial^0\varphi(0)$ with $\lambda \in I\!R$ doesn't allow us to make such a conclusion. Of course, such a conclusion cannot be made by using Clarke's generalized gradient $\partial_C\varphi(\bar{x})$ and Warga's minimal derivate container $\Lambda^0\varphi(\bar{x})$ for φ that are two-sided subdifferential constructions *always containing* $\partial^0\varphi(\bar{x})$, which is illustrated by the computation in Example 2.49.

Since both $\partial_C\varphi(\bar{x})$ and $\Lambda^0\varphi(\bar{x})$ may be essentially bigger (never smaller) than $\partial\varphi(\bar{x})$, the usage of the basic subdifferential in the results above leads us to more precise necessary optimality conditions for local minima in problems with *nonsmooth cost functions* and *inequality constraints*. The simplest illustrative example is given by the *unconditional* one-dimensional problem

$$\text{minimize } \varphi(x) := -|x|, \quad x \in I\!R ,$$

where $\bar{x} = 0$ is not a minimum (but maximum) point, while $0 \in \partial_C\varphi(0) = [-1, 1]$. On the other hand, $0 \notin \partial\varphi(0) = \{-1, 1\}$.

For the two-dimensional problem

$$\text{minimize } x_1 \text{ subject to } \varphi(x_1, x_2) := |x_1| - |x_2| \le 0$$

we have $\partial\varphi(0,0) = \{(v_1, v_2)|\ -1 \le v_1 \le 1,\ v_2 = 1 \text{ or } v_2 = -1\}$, and hence the point $\bar{x} = (0,0)$ is ruled out from being optimal by Theorem 5.21(iii), while the use of $\partial_C\varphi(0,0) = \{(v_1, v_2)|\ -1 \le v_1 \le 1,\ -1 \le v_2 \le 1\}$ doesn't allow us to do it. Another example of a two-dimensional problem with a nonsmooth inequality constraint is given by

$$\text{minimize } x_2 \text{ subject to } \varphi(x_1, x_2) := |\,|x_1| + x_2| + x_2 \le 0 ,$$

where $\partial\varphi(0,0) = \{(v_1, v_2)|\ |v_1| + 1 \le v_2 \le 2\} \cup \{(v_1, v_2)|\ 0 \le v_2 \le -|v_1| + 1\}$; see Example 2.49 for details. Thus the result of Theorem 5.21(iii) allows us to rule out the non-optimal point $\bar{x} = (0,0)$, while it cannot be done with the help of either $\partial_C\varphi(0,0)$ or $\Lambda^0\varphi(0,0)$.

The next optimization result we are going to obtain in the form of the *Lagrange principle*, which says that necessary optimality conditions in constrained problems can be given as necessary conditions for unconstrained local minima of some Lagrange functions (Lagrangian) built upon the original constraints with suitable multipliers. For the minimization problem (5.23) we consider the standard *Lagrangian*

$$L(x, \lambda_0, \ldots, \lambda_{m+r}) := \lambda_0\varphi_0(x) + \ldots + \lambda_{m+r}\varphi_{m+r}(x) \tag{5.35}$$

involving the cost function and the functional (but not geometric) constraints, and also the *essential Lagrangian*

$$L_\Omega(x; \lambda_0, \ldots, \lambda_{m+r}) := \lambda_0\varphi_0(x) + \ldots + \lambda_{m+r}\varphi_{m+r}(x) + \delta(x; \Omega) \tag{5.36}$$

involving the geometric constraints as well.

To derive general results of the Lagrange principle type, let us first establish a calculus lemma that is certainly of independent interest and will be also used in the sequel. Given a single-valued mapping $f\colon X \to Z$ between Banach spaces and subsets $\Omega \subset X$ and $\Theta \subset Z$, we consider the set

$$\mathcal{E}(f,\Theta,\Omega) := \big\{(x,z) \in X \times Z \big|\, f(x) - z \in \Theta,\ x \in \Omega\big\}\,, \tag{5.37}$$

which can be viewed as a *generalized epigraph* of the function f on Ω with respect to Θ. If, in particular, $f = (\varphi_0,\ldots,\varphi_m)\colon X \to I\!R^{m+1}$ and if $\Theta = I\!R^{m+1}_-$ is the nonnegative orthant of $Z = I\!R^{m+1}$, then the set (5.37) is the epigraph of the vector function f with respect to the standard order on $I\!R^{m+1}$. For $\Theta = \{0\}$ the set (5.37) is just the graph of f. If, more generally, $\Theta \subset Z$ is a convex cone inducing an order on Z, then (5.37) is the epigraph of the restriction $f_\Omega := f|_\Omega$ of the mapping $f\colon X \to Z$ on the set Ω with respect to this order on Z. Note that we can always write

$$f_\Omega(x) = f(x) + \Delta(x;\Omega)\ \text{ for all }\ x \in X$$

via the indicator mapping $\Delta(x;\Omega) := 0 \in Z$ if $x \in \Omega$ and $\Delta(x) := \emptyset$ otherwise.

In the next lemma we use the property of strong coderivative normality given in Definition 4.8; some sufficient conditions for this property are listed in Proposition 4.9.

Lemma 5.23 (basic normals to generalized epigraphs). *Let $f\colon X \to Z$ be a mapping between Banach spaces, and let $\Omega \subset X$ and $\Theta \subset Z$ be such sets that $\bar{x} \in \Omega$ and $f(\bar{x}) - \bar{z} \in \Theta$. The following assertions hold:*

(i) *Assume that f is locally Lipschitzian around $\bar{x}$ relative to Ω. Then*

$$D^*_M f_\Omega(\bar{x})(z^*) = \partial\langle z^*, f_\Omega\rangle(\bar{x})\ \text{ for all }\ z^* \in Z^*\,.$$

(ii) *One always has*

$$(x^*,z^*) \in N\big((\bar{x},\bar{z});\mathcal{E}(f,\Omega,\Theta)\big) \Longrightarrow -z^* \in N(f(\bar{x}) - \bar{z};\Theta)\,.$$

Assume further that both X and Z are Asplund, that f is continuous around $\bar{x}$ relative to Ω, and that both Ω and Θ are locally closed around $\bar{x}$ and $f(\bar{x})-\bar{z}$, respectively. Then

$$N\big((\bar{x},\bar{z});\mathcal{E}(f,\Omega,\Theta)\big) \subset \Big\{(x^*,z^*) \in X^* \times Z^* \Big|\, x^* \in D^*_N f_\Omega(\bar{x})(z^*)\,,\ -z^* \in N(f(\bar{x}) - \bar{z};\Theta)\Big\}\,. \tag{5.38}$$

(iii) *Assume in addition to* (ii) *that f is locally Lipschitzian around $\bar{x}$ relative to Ω and that f_Ω is strongly coderivatively normal at $\bar{x}$. Then*

$$N\big((\bar{x},\bar{z});\mathcal{E}(f,\Omega,\Theta)\big)\subset\Big\{(x^*,z^*)\in X^*\times Z^*\Big|\ x^*\in\partial\langle z^*,f_\Omega\rangle(\bar{x})\ ,\ -z^*\in N(f(\bar{x})-\bar{z};\Theta)\Big\}\ . \tag{5.39}$$

(iv) *Assume that* f *is locally Lipschitzian around* $\bar{x}$ *relative to* Ω*. Then the opposite inclusion holds in* (5.39) *in the case of arbitrary Banach spaces* X *and* Z*. If in addition* f_Ω *is strongly coderivatively normal at* $\bar{x}$*, then the opposite inclusion holds in* (5.38) *as well.*

Proof. Assertion (i) is an extension of the mixed scalarization formula in Theorem 1.90 and can be proved in the exactly same way. In fact, one can observe that the linear structure of $\Omega = X$ is never used in the proof of Theorem 1.90 in contrast to the proof of the normal scalarization formula in Lemma 3.27 and Theorem 3.28. Note that assertion (i) provides a bridge between assertions (ii) and (iii) of this theorem.

The first inclusion in (ii) follows directly from the definition of basic normals via the limit of ε-normals due to the structure of the set $\mathcal{E}(f,\Omega,\Theta)$ in (5.37). To prove the second inclusion in (ii), we observe that the latter set is represented as the inverse image

$$\mathcal{E}(f,\Omega,\Theta)=g^{-1}(\Theta)\ \text{ with }\ g(x,z):=f_\Omega(x)-z;$$

so we can apply Theorem 3.8 on basic normals to inverse images in Asplund spaces with $F=g$. It is easy to see that $g(\cdot)\cap\Theta$ is inner semicompact at $\bar{x}$ with $g(\bar{x},\bar{z})\cap\Theta=f(\bar{x})-\bar{z}$ under the continuity and closedness assumptions made. Let us show that for the mapping g of the above special structure one automatically has

$$\ker\widetilde{D}^*_M g(\bar{x},\bar{z})=\{0\}\ \text{ and }\ g^{-1}\ \text{ is PSNC at }\ (f(\bar{x})-\bar{z},\bar{x},\bar{z})\ .$$

First check the kernel condition. Picking any $z^*\in Z^*$ with $0\in\widetilde{D}^*_M g(\bar{x},\bar{z})(z^*)$, we find $(x_k,z_k)\to(\bar{x},\bar{z})$ and $(u_k^*,v_k^*)\in\widehat{D}^*g(x_k,z_k)(z_k^*)$ such that

$$x_k\in\Omega,\quad \|(u_k^*,v_k^*)\|\to 0,\quad\text{and}\quad z_k^*\xrightarrow{w^*}z^*\ \text{ as }\ k\to\infty\ .$$

Since $g(x,z)=f_\Omega(x)-z$, one has

$$\widehat{D}^*g(x_k,z_k)(z_k^*)=\big(\widehat{D}^*f_\Omega(x_k)(z_k^*),0\big)+(0,-z_k^*)$$

by Theorem 1.62(i). Hence

$$u_k^*\in\widehat{D}^*f_\Omega(x_k)(z_k^*)\quad\text{and}\quad v_k^*=-z_k^*,\quad k\in I\!N\ ,$$

which gives $\|z_k^*\|\to 0=z^*$, i.e., $\ker\widetilde{D}^*_M g(\bar{x},\bar{z})=\{0\}$. To check the PSNC property of g^{-1} at $(f(\bar{x})-\bar{z},\bar{x},\bar{z})$, we proceed in a similar way taking

$$(u_k^*,v_k^*,z_k^*)\in\widehat{N}((x_k,z_k);\operatorname{gph}g)\ \text{ with }\ \|(u_k^*,v_k^*)\|\to 0\ \text{ and }\ z_k^*\xrightarrow{w^*}0\ .$$

Then, by the above arguments, one has $\|z_k^*\| \to 0$, which justifies the PSNC property of g^{-1} at $(f(\bar{x}) - \bar{z}, \bar{x}, \bar{z})$. Thus all the assumptions of Theorem 3.8 are verified, and we arrive at (5.38) due to

$$(u^*, v^*) \in D_N^* g(\bar{x}, \bar{z})(z^*) \Longleftrightarrow u^* \in D_N^* f_\Omega(\bar{x})(z^*),\ v^* = -z^*\ ,$$

as follows from the sum rule of Theorem 1.62(ii) applied to $g(x, z) = f_\Omega(x) - z$.

Let us prove inclusion (5.39) in (iii) under the additional assumptions made therein. Taking into account assertion (i) of the theorem, we have

$$D_N^* f_\Omega(\bar{x})(z^*) = D_M^* f_\Omega(\bar{x})(z^*) = \partial\langle z^*, f_\Omega\rangle(\bar{x}), \quad z^* \in Z^*\ , \qquad (5.40)$$

which shows that (5.39) follows from (5.38) in this case.

It remains to prove (iv). First let us justify the opposite inclusion in (5.39). Picking any $z^* \in -N(f(\bar{x}) - \bar{z}; \Theta)$ and $x^* \in \partial\langle z^*, f_\Omega\rangle(\bar{x})$, we are going to show that $(x^*, z^*) \in N\big((\bar{x}, \bar{z}); \mathcal{E}(f, \Omega, \Theta)\big)$. By definitions of basic normals and subgradients one has sequences $\varepsilon_{1k} \downarrow 0$, $\varepsilon_{2k} \downarrow 0$, $x_k \xrightarrow{\Omega} \bar{x}$, $z_k \to \bar{z}$, $x_k^* \xrightarrow{w^*} x^*$, and $z_k^* \xrightarrow{w^*} z^*$ as $k \to \infty$ such that

$$x_k^* \in \widehat{\partial}_{\varepsilon_{1k}}\langle z_k^*, f_\Omega\rangle(x_k) \ \text{ and } \ -z_k^* \in \widehat{N}_{\varepsilon_{2k}}(f(x_k) - z_k; \Theta) \ \text{ for all } \ k \in I\!N\ .$$

It is easy to deduce from the definitions of ε-normals and ε-subgradients with the use of the local Lipschitz continuity of f_Ω around x_k for k sufficiently large that the above inclusions yield

$$(x_k^*, z_k^*) \in \widehat{N}_{\varepsilon_k}\big((x_k, z_k); \mathcal{E}(f, \Omega, \Theta)\big) \ \text{ for large } \ k \in I\!N\ ,$$

where $\varepsilon_k := \varepsilon_{1k} + (\ell + 1)\varepsilon_{2k} \downarrow 0$ with the Lipschitz constant ℓ of f_Ω around $\bar{x}$. The latter implies the opposite inclusion in (5.39) as $k \to \infty$. The opposite inclusion in (5.38) follows from the one in (5.39) due to the normal coderivative representation (5.40) under the coderivative normality assumption. △

Now we come back to the main optimization problem (5.23) under consideration in this subsection and define the set

$$\mathcal{E}(\varphi_0, \ldots, \varphi_{m+r}, \Omega) := \Big\{(x, \alpha_0, \ldots, \alpha_{m+r}) \in X \times I\!R^{m+r+1}\Big|\ x \in \Omega,\ \varphi_i(x) \le \alpha_i\ ,$$
$$i = 0, \ldots, m;\ \varphi_i(x) = \alpha_i,\ i = m+1, \ldots, m+r\Big\}\ ,$$

which corresponds to (5.37) with $f = (\varphi_0, \ldots, \varphi_{m+r})\colon X \to I\!R^{m+r+1}$ and $\Theta = I\!R_-^{m+1} \times \{0\} \subset I\!R^{m+r+1}$. The next result, based on the extremal principle, provides necessary optimality conditions for (5.23) via basic normals to the generalized epigraph $\mathcal{E}(\varphi_0, \ldots, \varphi_{m+r}, \Omega)$ in a very broad framework and can be equivalently expressed in an extended form of the *Lagrange principle* under Lipschitzian assumptions on φ_i, $i = 0, \ldots, m + r$. For convenience we assume in what follows that $\varphi_0(\bar{x}) = 0$ at the optimal solution under consideration, which doesn't restrict the generality.

Theorem 5.24 (extended Lagrange principle). *Let $\bar{x}$ be a local optimal solution to problem* (5.23), *where the space X is Asplund. Assume that the set Ω is locally closed around $\bar{x}$ and that the functions φ_i are l.s.c. around $\bar{x}$ relative to Ω for $i = 0, \ldots, m$ and continuous around this point relative to Ω for $i = m+1, \ldots, m+r$. Then there are Lagrange multipliers $(\lambda_0, \ldots, \lambda_{m+r}) \in I\!R^{m+r+1}$, not equal to zero simultaneously, such that*

$$(0, -\lambda_0, \ldots, -\lambda_{m+r}) \in N\big((\bar{x}, 0); \mathcal{E}(\varphi_0, \ldots, \varphi_{m+r}, \Omega)\big) , \tag{5.41}$$

which automatically implies the sign and complementary slackness conditions in (5.25). *If in addition the functions φ_i, $i = m+1, \ldots, m+r$, are continuous around $\bar{x}$ relative to Ω, then* (5.41) *implies also that*

$$0 \in D_N^*\big((\varphi_0, \ldots, \varphi_{m+r}) + \Delta(\cdot; \Omega)\big)(\bar{x})(\lambda_0, \ldots, \lambda_{m+r}) . \tag{5.42}$$

Moreover, if all the functions φ_i, $i = 0, \ldots, m + r$, are Lipschitz continuous around $\bar{x}$ relative to the set Ω, then the coderivative inclusion (5.42) *is equivalent to the subdifferential one*

$$0 \in \partial L_\Omega(\cdot, \lambda_0, \ldots, \lambda_{m+r})(\bar{x}) \tag{5.43}$$

in terms of the essential Lagrangian (5.36). *In this case the necessary condition* (5.41) *is equivalent to the simultaneous fulfillment of* (5.25) *and* (5.43).

Proof. Since $\bar{x}$ is a local optimal solution to (5.23), there is a neighborhood U of $\bar{x}$ such that $\bar{x}$ provides the minimum to φ_0 over $x \in U$ subject to the constraints in (5.23). Consider the sets

$$\Omega_1 := \mathcal{E}(\varphi_0, \ldots, \varphi_{m+r}, \Omega) \ \text{ and } \ \Omega_2 := \operatorname{cl} U \times \{0\}$$

in the Asplund space $X \times I\!R^{m+r+1}$ and observe that $(\bar{x}, 0)$ is an extremal point of the system $\{\Omega_1, \Omega_2\}$. Indeed, one obviously has $(\bar{x}, 0) \in \Omega_1 \cap \Omega_2$ and $\big(\Omega_1 - (0, \nu_k, 0, \ldots, 0)\big) \cap \Omega_2 = \emptyset$, $k \in I\!N$, for any sequence of negative numbers $\nu_k \uparrow 0$ by the local optimality of $\bar{x}$ in (5.23). Taking into account that both sets Ω_1 and Ω_2 are locally closed around $(\bar{x}, 0)$ and that Ω_2 is SNC at this point due to $\bar{x} \in \operatorname{int} U$ and $0 \in I\!R^{m+r+1}$, we apply the exact version of the extremal principle from Theorem 2.22 and arrive at (5.41) with $(\lambda_0, \ldots, \lambda_{m+r}) \neq 0$. By Lemma 5.23(ii) with $\Theta = I\!R_-^{m+1} \times \{0\}$ one immediately has from (5.41) the complementary slackness and sign conditions (5.25) and, under the continuity assumption on φ_i for $i = m+1, \ldots, m+r$, the coderivative inclusion (5.42). If φ_i are locally Lipschitzian around $\bar{x}$ for all $i = 0, \ldots, m+r$, the equivalence statements in the theorem follow from assertions (iii) and (iv) of Lemma 5.23, since the coderivative normality assumption holds for any mapping with a finite-dimensional image space. △

Using further calculus rules for basic normals, coderivatives, and subgradients, we can derive various consequences of inclusions (5.41)–(5.43). Let us present some results expressed in terms of subgradients of the standard Lagrangian (5.35) involving the cost function and functional (but not geometric) constraints in the optimization problem (5.23).

Corollary 5.25 (Lagrangian conditions and abstract maximum principle). *Let $\bar{x}$ be a local optimal solution to* (5.23). *Assume that the space X is Asplund, that the functions φ_i are Lipschitz continuous around $\bar{x}$ for all $i = 0, \ldots, m + r$ and that the set Ω is locally closed around this point. Then there are Lagrange multipliers $\lambda_0, \ldots, \lambda_{m+r}$, not all zero, such that the conditions* (5.25) *and*

$$0 \in \partial L(\cdot, \lambda_0, \ldots, \lambda_{m+r})(\bar{x}) + N(\bar{x}; \Omega) \tag{5.44}$$

hold. If in addition the set Ω is convex, then

$$\langle x^*, \bar{x} \rangle = \max \big[\langle x^*, x \rangle \big| \; x \in \Omega \big] \tag{5.45}$$

for some $x^* \in -\partial L(\cdot, \lambda_0, \ldots, \lambda_{m+r})(\bar{x})$.

Proof. Inclusion (5.44) follows from (5.43) due to the subdifferential sum rule from Theorem 2.33(c). It implies the maximum condition (5.45) in the case of convex geometric constraints by the representation of basic normals to convex sets from Proposition 1.5. △

Note that the second assertion in Corollary 5.25 gives an *abstract maximum principle*, which is directly induced by the *convex structure* via the normal cone representation for convex geometric constraints. Note also that the results obtained imply those in terms of separate constraints similarly to Corollary 5.20 and Theorem 5.21(iii).

Passing to the next topic, we observe that lower subdifferential conditions for the minimization problem (5.23) obtained in Theorem 5.21(iii) employ the basic subgradient sets $\partial\varphi_i(\bar{x})$ for $i = 0, \ldots, m$ and $\partial\varphi_i(\bar{x}) \cup \partial(-\varphi_i)(\bar{x})$ for $i = m + 1, \ldots, m + r$, which are nonsmooth extensions of the classical *strict* derivative. While the results of this type seem to be *unimprovable* for general equality constraints in infinite dimensions, we may derive more subtle in certain situations (generally independent) results employing extensions of the *usual*—not strict—Fréchet derivative for nonsmooth cost and inequality constraint functions. Some results in this direction are given in Theorem 5.19 via Fréchet *upper subdifferentials*. Now we derive *lower subdifferential* conditions in a *mixed* form that involve subgradient extensions of the strict derivative for equality constraint functions and those of the usual Fréchet derivative for functions describing the objective and inequality constraints.

To proceed, recall some notions of nonsmooth analysis related to *convex directional approximations* of functions and sets. Given $\varphi\colon X \to \overline{I\!R}$ finite at $\bar{x}$, the extended-real-valued function

$$d^+\varphi(\bar{x};h) := \limsup_{\substack{z\to h \\ t\downarrow 0}} \frac{\varphi(\bar{x}+tz)-\varphi(\bar{x})}{t} \tag{5.46}$$

is the *upper Dini-Hadamard directional derivative* of φ at $\bar{x}$ in the direction h. One can put $z = h$ in (5.46) if φ is Lipschitzian around $\bar{x}$. A function $p(\bar{x};\cdot)\colon X \to \overline{I\!R}$ is an *upper convex approximation* of φ at $\bar{x}$ if it is convex, l.s.c., and positively homogeneous with $p(\bar{x};h) \ge d^+\varphi(\bar{x};h)$ for all $h \in X$. Then the subdifferential of $p(\bar{x};\cdot)$ at $h = 0$ in the sense of convex analysis is called the *p-subdifferential* of φ at $\bar{x}$ and is denoted by

$$\partial_p\varphi(\bar{x}) := \partial p(\bar{x};0) = \big\{x^* \in X^* \big|\ \langle x^*,h\rangle \le p(\bar{x};h)\ \text{ for all }\ h \in X\big\}\ . \tag{5.47}$$

Observe that the subdifferential (5.47) depends on an upper convex approximation $p(\bar{x};\cdot)$, i.e., is not uniquely defined. For example, the function $\varphi(x) = -|x|$ on $I\!R$ admits a family of upper convex approximations at $\bar{x} = 0$ given by $p(0;h) = \gamma h$ for any $\gamma \in [-1,1]$. It follows from Subsect. 2.5.2A that Clarke's generalized directional derivative $\varphi^\circ(\bar{x};h)$ automatically provides an upper convex approximation for any locally Lipschitzian function φ. However, this approximation may not be the best one, as we see from the above example of $\varphi(x) = -|x|$. Note also that $p(\bar{x};h) = \langle\nabla\varphi(\bar{x}),h\rangle$ is an upper convex approximation of φ whenever φ is Gâteaux differentiable at $\bar{x}$, i.e., p-subdifferentials are nonsmooth extensions of the usual (not strict) derivative of a function at a reference point. There are various efficient realizations of the idea to build a *convex-valued* subdifferential in the scheme (5.47) corresponding to specific classes of functions admitting upper convex approximations, which are along the initial line of developing nonsmooth analysis; see the comments and references in Subsect. 1.4.1.

Recall also the construction of the *contingent cone* to a set $\Omega \subset X$ at $\bar{x} \in \Omega$ defined in Subsect. 1.1.2 by

$$T(\bar{x};\Omega) := \mathop{\mathrm{Lim\,sup}}_{t\downarrow 0} \frac{\Omega - \bar{x}}{t}\ . \tag{5.48}$$

This is a nonempty and closed cone that reduces to the classical tangent cone for convex sets Ω, while (5.48) is nonconvex in general. Note that

$$\widehat{N}_\varepsilon(\bar{x};\Omega) \subset \big\{x^* \in X^* \big|\ \langle x^*,v\rangle \le \varepsilon\|x\|\ \text{ for all }\ v \in T(\bar{x};\Omega)\big\} \tag{5.49}$$

whenever $\varepsilon \ge 0$. Moreover, (5.49) holds as equality if X is finite-dimensional. Thus in the latter case the Fréchet normal cone $\widehat{N}(\bar{x};\Omega)$ is polar/dual to the contingent cone (5.48) due to the equality relationship in (5.49).

Theorem 5.26 (mixed subdifferential conditions for local minima). *Let $\bar{x}$ be a local optimal solution to problem* (5.23)*, where the space X is Asplund, where the set Ω is locally closed around $\bar{x}$, and where all the functions φ_i are locally Lipschitzian around this point. Assume also that there exists a* *convex closed subcone* M *of* $T(\bar{x};\Omega)$ *with* $M^* \subset N(\bar{x};\Omega)$ *and that the functions* φ_i, $i \in I(\bar{x}) \cup \{0\}$, *admit upper convex approximations at* $\bar{x}$, *which are continuous at some point of* M. *Denote* $\vartheta(x) := \|(\varphi_{m+1}(x),\ldots,\varphi_{m+r}(x))\|$ *and suppose that this function admits an upper convex approximation at* $\bar{x}$ *whose subdifferential* (5.47) *is contained in* $\partial\vartheta(\bar{x})$. *Then there are Lagrange multipliers* $\lambda_i \geq 0$ *for* $i \in I(\bar{x}) \cup \{0\}$ *and* $(\lambda_{m+1},\ldots,\lambda_{m+r}) \in I\!R^r$, *not equal to zero simultaneously, such that*

$$0 \in \sum_{i\in I(\bar{x})\cup\{0\}} \lambda_i \partial_p \varphi_i(\bar{x}) + \partial\Big(\sum_{i=m+1}^{m+r} \lambda_i \varphi_i\Big)(\bar{x}) + N(\bar{x};\Omega)\,, \tag{5.50}$$

where $\partial_p\varphi_i(\bar{x})$ *stand for the subdifferentials* (5.47) *corresponding to the upper convex approximations* $p_i(\bar{x};\cdot)$ *of* φ_i *for* $i \in I(\bar{x}) \cup \{0\}$.

Proof. First we consider the case when $f := (\varphi_{m+1},\ldots,\varphi_{m+r})\colon X \to I\!R^r$ is *metrically regular* at $\bar{x}$ relative to Ω. Then Theorem 5.16 ensures that, for some $\mu > 0$, $\bar{x}$ is a local optimal solution to the unconstrained minimization problem (UP) defined therein. Invoking the form of the cost function in (UP) and employing the definition of upper convex approximations as well as the convexity assumption on $M \subset T(\bar{x};\Omega)$, one can derive by standard separation arguments of convex programming with the usage of standard subdifferential formulas of convex analysis that there are numbers $\lambda_i \geq 0$, $i \in I(\bar{x})$, sum of which is 1, such that

$$0 \in \sum_{i\in I(\bar{x})} \lambda_i \partial_p \varphi_i(\bar{x}) + \mu\partial_p\vartheta(\bar{x}) + M^*\,.$$

Taking into account that $M^* \subset N(\bar{x};\Omega)$ and $\partial_p\vartheta(\bar{x}) \subset \partial\vartheta(\bar{x})$ and then applying the chain rule from Corollary 3.43 to the latter subdifferential of the composition $\vartheta = (\psi \circ f)$ with $\psi(y) := \|y\|$, we arrive at (5.50) under the metric regularity assumption.

If f is *not* metrically regular at $\bar{x}$ relative to Ω, then $f_\Omega = f + \Delta(\cdot;\Omega)$ is not metrically regular around $\bar{x}$ in the sense of Definition 1.47(ii). By Theorem 4.18(c) this happens when either $\ker \widetilde{D}^*_M f_\Omega(\bar{x}) \neq \{0\}$ or f_Ω^{-1} is not PSNC at $(f(\bar{x}),\bar{x})$. Since the image space for f_Ω is finite-dimensional, the latter PSNC condition automatically holds, and hence the absence of the metric regularity means that there is nonzero $y^* \in I\!R^r$ such that $0 \in \widetilde{D}^*_M f_\Omega(\bar{x})(y^*)$. The rest of the proof follows the one in Theorem 5.17. △

Note that Theorem 5.26 and the previous results in terms of basic subgradients give *generally independent* conditions even in the case of problems with only inequality constraints. In particular, one can check that the function $\varphi(x_1,x_2) = |\,|x_1| + x_2| + x_2$ from the last example considered in Remark 5.22 doesn't admit upper convex approximations at $\bar{x} = 0$ whose subdifferentials are proper subsets of the basic subdifferential $\partial\varphi(0)$. On the other hand, Theorem 5.26 allows us to establish non-optimality of the point $\bar{x} = 0$ in the one-dimensional optimization problem:

$$\begin{cases} \text{minimize } \ \varphi_0(x) := x \ \text{ subject to} \\ \\ \varphi_1(x) := x^2 \sin(1/x) \le 0 \ \text{ as } \ x \ne 0 \ \text{ with } \ \varphi_1(0) = 0 \ , \end{cases}$$

while the above necessary conditions in terms of basic subgradients don't work, since $\partial\varphi_1(0) = [-1, 1]$.

The final result of this subsection concerns *lower subdifferential* necessary optimality conditions for problems (5.23) with *non-Lipschitzian* data. The previous results obtained for such problems are expressed in terms of normals to graphical and epigraphical sets and cannot be reduced to subgradients of the corresponding functions in the absence of Lipschitzian assumptions. Now we are going to derive new necessary conditions for non-Lipschitzian problems in a *fuzzy subdifferential form* that involve Fréchet subgradients of the cost and constraint functions in (5.23). To proceed, we need the following lemma, which is a *weak* non-Lipschitzian counterpart of the (strong) fuzzy sum rule given in Theorem 2.33(b) under the semi-Lipschitzian assumption. Note that this result involves a weak* neighborhood of the origin in X^* instead of a small dual ball as in Theorem 2.33(b). This lemma is derived from the density result of Corollary 2.29 by using properties of infimal convolutions; see Fabian [414, 415] for a complete proof and more discussions.

Lemma 5.27 (weak fuzzy sum rule). *Let X be an Asplund space, and let $\varphi_1, \ldots, \varphi_n$ be extended-real-valued l.s.c. functions on X. Then for any $\bar{x} \in X$, $\varepsilon > 0$, $x^* \in \widehat{\partial}(\varphi_1 + \ldots + \varphi_n)(\bar{x})$, and any weak* neighborhood V^* of the origin in X^* there are $x_i \in \bar{x} + \varepsilon I\!B$ and $x_i^* \in \widehat{\partial}\varphi_i(x_i)$ such that $|\varphi_i(x_i) - \varphi_i(\bar{x})| \le \varepsilon$ for all $i = 1, \ldots, n$ and*

$$x^* \in \sum_{i=1}^{n} x_i^* + V^* \ .$$

Now we are ready to establish a *weak approximate version* of the Lagrange multiplier rule for local optimal solutions to problem (5.23) with non-Lipschitzian functional constraints.

Theorem 5.28 (weak subdifferential optimality conditions for non-Lipschitzian problems). *Let $\bar{x}$ be a local optimal solution to problem* (5.23) *in an Asplund space X. Assume that the functions φ_i are l.s.c. around $\bar{x}$ for $i = 0, \ldots, m$ and continuous around this point for $i = m+1, \ldots, m+r$, and that the set Ω is locally closed around $\bar{x}$. Then for any $\varepsilon > 0$ and any weak* neighborhood V^* of the origin in X^* there are*

$$x_i \in \bar{x} + \varepsilon I\!B \ \ \textit{with} \ \ |\varphi_i(x_i) - \varphi_i(\bar{x})| \le \varepsilon \ \ \textit{for} \ \ i = 0, \ldots, m + r \ ,$$

$$x_i^* \in \widehat{\partial}\varphi_i(x_i) \ \ \textit{for} \ \ i = 0, \ldots, m \ ,$$

$$x_i^* \in \widehat{\partial}\varphi_i(x_i) \cup \widehat{\partial}(-\varphi_i)(x_i) \ \ \textit{for} \ \ i = m+1, \ldots, m + r \ ,$$

$$\widehat{x}^* \in \widehat{N}(\widehat{x}; \Omega) \ \ \textit{with} \ \ \widehat{x} \in \Omega \cap (\bar{x} + \varepsilon I\!B), \quad \textit{and}$$

$$\lambda_i \ge 0 \ \ \textit{for} \ \ i = 0, \ldots, m + r \ \ \textit{with} \ \ \sum_{i=0}^{m+r} \lambda_i = 1$$

satisfying the relation

$$0 \in \sum_{i=0}^{m+r} \lambda_i x_i^* + \widehat{x}^* + V^* \ .$$

Proof. Consider the constraint sets

$$\Omega_i = \begin{cases} \{x \in X \mid \varphi_i(x) \le 0\} & \text{for } \ i = 1, \ldots, m \ , \\ \{x \in X \mid \varphi_i(x) = 0\} & \text{for } \ i = m+1, \ldots, m + r \end{cases}$$

and observe that the original constraint problem (5.23) is obviously equivalent to the unconstrained problem with "infinite penalties":

$$\text{minimize} \ \ \varphi_0(x) + \delta(x; \Omega_1 \cap \ldots \cap \Omega_{m+r} \cap \Omega), \quad x \in X \ .$$

By the generalized Fermat principle and the cost function structure in the latter problem we have

$$0 \in \widehat{\partial}\Big(\varphi_0 + \sum_{i=1}^{m+r} \delta(\cdot; \Omega_i) + \delta(\cdot; \Omega)\Big)(\bar{x}) \ .$$

Picking any $\varepsilon > 0$ and a weak* neighborhood V^* of the origin in X^* and then applying Lemma 5.27 to the above sum, we find

$$x_0^* \in \widehat{\partial}\varphi_0(x_0) \ \ \text{with} \ \ \|(x_0, \varphi_0(x_0)) - (\bar{x}, \varphi_0(\bar{x}))\| \le \varepsilon \ ,$$

$$\widetilde{x}^* \in \widehat{N}(\widehat{x}; \Omega) \ \ \text{with} \ \ \widehat{x} \in \Omega \cap (\bar{x} + \varepsilon I\!B) \ ,$$

$$\widetilde{x}_i^* \in \widehat{N}(\widetilde{x}_i; \Omega_i) \ \ \text{with} \ \ \widetilde{x}_i \in \Omega_i \cap (\bar{x} + (\varepsilon/2) I\!B) \ \ \text{for} \ \ i = 1, \ldots, m + r$$

satisfying the relation

$$0 \in x_0^* + \sum_{i=1}^{m+r} \widetilde{x}_i^* + \widetilde{x}^* + \Big(\frac{1}{m + r + 1}\Big) V^* \ .$$

Taking into account the structures of the set Ω_i, we now consider the following two cases.

Case (a). There are either $i \in \{1,\ldots,m\}$ and $\lambda \neq 0$ satisfying $(0,\lambda) \in N((\widetilde{x}_i,0);\operatorname{epi}\varphi_i)$ or $i \in \{m+1,\ldots,m+r\}$ satisfying $0 \in \partial\varphi_i(\widetilde{x}_i) \cup \partial(-\varphi_i)(\widetilde{x}_i)$. Let this happen for some $i \in \{1,\ldots,m\}$. Then by the basic cone representation from Theorem 2.35 in Asplund spaces we find $(x_i,\alpha_i) \in \operatorname{epi}\varphi_i$ and $(x_i^*,-\lambda_i) \in \widehat{N}((x_i,\alpha_i);\operatorname{epi}\varphi_i)$ such that

$$\|x_i - \widetilde{x}_i\| \le \varepsilon/2, \quad \lambda_i > 0, \quad \text{and} \quad x_i^* \in \lambda_i V^* \ .$$

It is easy to see that $\widehat{N}((x_i,\alpha_i);\operatorname{epi}\varphi_i) \subset \widehat{N}((x_i,\varphi_i(x_i));\operatorname{epi}\varphi_i)$ due to $\alpha_i \ge \varphi_i(x_i)$, and so $(x_i^*,-\lambda_i) \in N((x_i,\varphi_i(x_i));\operatorname{epi}\varphi_i)$. Thus we have in this situation the inclusions

$$x_i^*/\lambda_i \in \widehat{\partial}\varphi_i(x_i) \quad \text{and} \quad x_i^*/\lambda_i \in V^* \ ,$$

which ensure that all the required relations in the theorem hold with $\lambda_i = 1$ for the reference index i (the other λ_i are zero) and $\widehat{x}^* = 0$.

Consider further case (a) with some $i \in \{m+1,\ldots,m+r\}$. Using the basic subdifferential representation from Theorem 2.34(b) for continuous functions on Asplund spaces, we find $(x_i,x_i^*) \in X \times X^*$ such that

$$x_i \in \widetilde{x}_i + (\varepsilon/2)I\!B \quad \text{and} \quad x_i^* \in \big[\widehat{\partial}\varphi_i(x_i) \cup \widehat{\partial}(-\varphi_i)(x_i)\big] \cap V^* \ .$$

This also implies the conclusions of the theorem with $\lambda_i = 1$ for the reference index i, $\widehat{x}^* = 0$, and all other λ_i equal to zero.

Case (b). Otherwise to the assumptions in case (a). First we consider an index $i \in \{m+1,\ldots,m+r\}$ corresponding to the equality constraints, i.e., when $\Omega_i = \{x \in X|\ \varphi_i(x) = 0\}$. Observe that $\Omega_i \times \{0\} = \Lambda_1 \cap \Lambda_2$ for

$$\Lambda_1 := \operatorname{gph}\varphi_i \quad \text{and} \quad \Lambda_2 := \big\{(x,\alpha) \in X \times I\!R\big|\ \alpha = 0\big\} \ ,$$

where the second set is SNC at $(\widetilde{x}_i,0)$ and the qualification condition of Corollary 3.5 reduces to $0 \notin \partial\varphi_i(\widetilde{x}_i) \cup \partial(-\varphi_i)(\widetilde{x}_i)$. Applying now the intersection formula from this corollary and then using Theorems 1.80 and 2.40(ii) that give the subdifferential representations of coderivatives for continuous functions, we arrive at the inclusion

$$N(\widetilde{x}_i;\Omega_i) \subset \partial^\infty\varphi_i(\widetilde{x}_i) \cup \partial^\infty(-\varphi_i)(\widetilde{x}_i) \cup I\!R_+\partial\varphi_i(\widetilde{x}_i) \cup I\!R_+\partial(-\varphi_i)(\widetilde{x}_i) \ ,$$

where $I\!R_+S := \{\nu s|\ \nu \ge 0,\ s \in S\}$. This imply, invoking the normal and subdifferential representations form Theorems 2.34(b), 2.35(b), and 2.38, that for all $i = m+1,\ldots,m+r$ there are $x_i \in \widetilde{x}_i + (\varepsilon/2)I\!B$, $\nu_i \ge 0$, and

$$x_i^* \in \widehat{\partial}\varphi_i(x_i) \cup \widehat{\partial}(-\varphi_i)(x_i) \ \text{ with } \ \nu_i x_i^* \in \widetilde{x}_i^* + \tfrac{1}{m+r+1}V^* \ .$$

Next let us consider an index $i \in \{1, \ldots, m\}$ corresponding to the inequality constraints, i.e., when $\Omega_i = \{x \in X|\ \varphi_i(x) \le 0\}$. Representing Ω_i via the intersection form as $\Omega_i \times \{0\} = \Lambda_1 \cap \Lambda_2$ with

$$\Lambda_1 := \operatorname{epi}\varphi_i \quad \text{and} \quad \Lambda_2 := \big\{(x,\alpha) \in X \times I\!R \big|\ \alpha = 0\big\} ,$$

we observe that the assumptions of Corollary 3.5 hold at $(\widetilde{x}_i, 0)$, since Λ_2 is SNC at this point and the qualification condition of the corollary reduces to

$$(0,\lambda) \in N((\widetilde{x}_i,0); \operatorname{epi}\varphi_i) \Longrightarrow \lambda = 0 .$$

Hence, taking the above $(\widetilde{x}_i, \widetilde{x}_i^*)$ with

$$\widetilde{x}_i^* \in \widehat{N}(\widetilde{x}_i; \Omega_i) \subset N(\widetilde{x}_i; \Omega_i), \quad \widetilde{x}_i \in \Omega_i \cap (\bar{x} + (\varepsilon/2)I\!B) ,$$

we find $\widetilde{\nu}_i \ge 0$ such that $(\widetilde{x}_i^*, -\widetilde{\nu}_i) \in N((\widetilde{x}_i, 0); \operatorname{epi}\varphi_i)$. Then using the limiting representation of basic normals from Theorem 2.35(b), we approximate $(\widetilde{x}_i^*, \widetilde{\nu}_i)$ in the weak* topology of $X^* \times I\!R$ by elements $(\widehat{x}_i^*, -\widehat{\nu}_i) \in \widehat{N}((\widehat{x}_i, \alpha_i); \operatorname{epi}\varphi_i)$ with $(\widehat{x}_i, \alpha_i)$ sufficiently close to $(\widetilde{x}_i, 0)$. Without loss of generality we may assume that $\alpha_i = \varphi_i(\widehat{x}_i)$; cf. case (a). If $\widehat{\nu}_i \ne 0$, we put $x_i := \widehat{x}_i$, $\nu_i := \widehat{\nu}_i$, and $x_i^* := \widehat{x}_i^*/\nu_i \in \widehat{\partial}\varphi_i(x_i)$ to get

$$x_i^* \in \widehat{\partial}\varphi_i(x_i) \ \text{ with } \ \nu_i x_i^* \in \widetilde{x}_i^* + \tfrac{1}{m+r+1} V^* .$$

If $\widehat{\nu}_i = 0$, we use Lemma 2.37 to find a *strong* approximation $(\nu_i x_i^*, x_i)$ of $(\widehat{x}_i^*, \widehat{x}_i)$ in the norm topology of $X \times X^*$ such that $\nu_i \ge 0$ and $x_i^* \in \widehat{\partial}\varphi_i(x_i)$.

Combining the above relationships, one has

$$\|\widehat{x} - \bar{x}\| \le \varepsilon, \quad \|x_i - \bar{x}\| \le \varepsilon \ \text{ for } \ i = 0, \ldots, m+r, \quad |\varphi_0(x_0) - \varphi_0(\bar{x})| \le \varepsilon ,$$

$$\widetilde{x}^* \in \widehat{N}(\widehat{x}; \Omega), \quad x_i^* \in \widehat{\partial}\varphi_i(x_i) \ \text{ for } \ i = 0, \ldots, m ,$$

$$x_i^* \in \widehat{\partial}\varphi_i(x_i) \cup \widehat{\partial}(-\varphi_i)(x_i) \ \text{ for } \ i = m+1, \ldots, m+r, \ \text{and}$$

$$0 \in x_0^* + \sum_{i=1}^{m+r} \nu_i x_i^* + \widehat{x}^* + V^* \ \text{ with } \ \nu_i \ge 0 \ \text{ for all } \ i = 1, \ldots, m+r .$$

Letting now $\lambda := 1/(1 + \sum_{i=1}^{m+r} \nu_i)$, $\widehat{x}^* := \lambda \widetilde{x}^*$, $\lambda_0 := \lambda$, and $\lambda_i := \lambda \nu_i$ for $i = 1, \ldots, m+r$, we arrive at all the conclusions of the theorem but

$$|\varphi_i(x_i) - \varphi_i(\bar{x})| \le \varepsilon \ \text{ for } \ i = 1, \ldots, m \tag{5.51}$$

noticing that for $i = m+1, \ldots, m+r$ the latter estimates are automatic due to the continuity of φ_i for these i. It is not the case in (5.51) when φ_i are supposed to be merely l.s.c. around $\bar{x}$ for $i = 1, \ldots, m$. Observe that if $\varphi_i(\bar{x}) = 0$, then (5.51) directly follows from the lower semicontinuity of φ_i for this $i \in \{1, \ldots, m\}$. If otherwise $\varphi_i(\bar{x}) < 0$ for some $i \in \{1, \ldots, m\}$, we replace this constraint by $\phi_i(x) := \varphi_i(x) - \varphi_i(\bar{x}) \le 0$ and observe that $\bar{x}$ is an optimal solution to the new problem with $\phi_i(\bar{x}) = 0$ and $\widehat{\partial}\phi_i(\bar{x}) = \widehat{\partial}\varphi_i(\bar{x})$. This fully justifies (5.51) and completes the proof of the theorem. △

5.1.4 Suboptimality Conditions for Constrained Problems

This subsection is devoted to *suboptimality* conditions for problems of mathematical programming in infinite-dimensional spaces. This means that we *don't assume the existence of optimal solutions* and obtained conditions held for suboptimal (ε-optimal) solutions, which always exist. The latter is particularly important for infinite-dimensional optimization problems, where the existence of optimal solutions requires quite restrictive assumptions. As pointed out by L. C. Young, any theory of necessary optimality conditions is "naive" until the existence of optimal solutions is clarified. This was the primary motivation for developing theories of generalized curves/relaxed controls in problems of the calculus of variations and optimal control to automatically ensure the existence of optimal solutions; see Chap. 6 for more details and discussions. However, the approaches developed in the mentioned areas of infinite-dimensional optimization are substantially based on specific features of continuous-time dynamic constraints governed by differential and related equations. This doesn't apply to general optimization problems in infinite dimensions. A natural approach to avoiding troubles with the existence of optimal solutions in general optimization problems is to show that "almost" optimal (i.e., suboptimal) solutions "almost" satisfy necessary conditions for optimality. From the practical viewpoint this has about the same effect and applications as necessary optimality conditions.

In what follows we are going to derive necessary optimality conditions of the subdifferential type for problems of nondifferentiable programming (5.23) with equality and inequality constraints under both Lipschitzian and non-Lipschitzian assumptions on the initial data. Similar results can be obtained for more general problems with operator constraints of type (5.12) that are not considered in this subsection for brevity.

Let us start with suboptimality conditions for problems (5.23) with *non-Lipschitzian data*. The following result is similar to Theorem 5.28. The only essential difference is that the obtained *weak suboptimality* conditions don't include conclusion (5.51) for the inequality constraints given by l.s.c. functions. The proof of the next theorem is also similar to the proof of Theorem 5.28, but it is somewhat more involved with the usage of the *lower subdifferential variational principle* from Theorem 2.28 instead the Fermat stationary one for the corresponding unconstrained problem.

Recall that *feasible solutions* to the optimization problem (5.23) are those x satisfying all the constraints, and that by $\inf\varphi_0$ we mean the infimum of thecost function with respect to all feasible solutions to (5.23). We always assume that $\inf\varphi_0 > -\infty$. It is natural to say that a point x is an *ε-optimal* solution to (5.23) if it is feasible to this problem with

$$\varphi_0(x) \le \inf\varphi_0 + \varepsilon \ .$$

Theorem 5.29 (weak suboptimality conditions for non-Lipschitzian problems). *Let X be an Asplund space, and let V^* be an arbitrary weak**

neighborhood of the origin in X^*. *Assume that* Ω *is closed, that* $\varphi_0, \ldots, \varphi_m$ *are l.s.c., and that* $\varphi_{m+1}, \ldots, \varphi_{m+r}$ *are continuous on the set of* ε*-optimal solutions to* (5.23) *for all* $\varepsilon > 0$ *sufficiently small. Then there exists* $\bar{\varepsilon} > 0$ *such that for every* $0 < \varepsilon < \bar{\varepsilon}$ *and every* ε^2*-optimal solution* $\bar{x}$ *to* (5.23) *there are* (x_i, x_i^*, λ_i) *satisfying the conditions:*

$$x_i \in \bar{x} + \varepsilon I\!B \quad \text{for } i = 0, \ldots, m + r \text{ with } |\varphi_0(x_0) - \varphi_0(\bar{x})| \le \varepsilon ,$$

$$x_i^* \in \widehat{\partial}\varphi_i(x_i) \quad \text{for } i = 0, \ldots, m ,$$

$$x_i^* \in \widehat{\partial}\varphi_i(x_i) \cup \widehat{\partial}(-\varphi_i)(x_i) \quad \text{for } i = m + 1, \ldots, m + r ,$$

$$\widehat{x}^* \in \widehat{N}(\widehat{x}; \Omega) \quad \text{with } \widehat{x} \in \Omega \cap (\bar{x} + \varepsilon I\!B) ,$$

$$\lambda_i \ge 0 \quad \text{for } i = 0, \ldots, m + r \quad \text{with } \sum_{i=0}^{m+r} \lambda_i = 1, \quad \text{and}$$

$$0 \in \sum_{i=0}^{m+r} \lambda_i x_i^* + \widehat{x}^* + V^* .$$

Proof. For any $v \in X$ and $\gamma > 0$ we consider a family of weak* neighborhoods of the origin in X^* given by

$$V^*(v; \gamma) := \left\{ x^* \in X^* \middle| \, |\langle x^*, v \rangle| < \gamma \right\}$$

that form a basis of the weak* topology. Taking an arbitrary weak* neighborhood V^* in the theorem, we find $\bar{\gamma} > 0$, $p \in I\!N$, and $v_j \in X$ with $\|v_j\| = 1$, $1 \le j \le p$, such that

$$\bigcap_{j=1}^{p} V^*(v_j; 2\bar{\gamma}) \subset V^* .$$

Let us show that the conclusions of the theorem hold for every ε satisfying

$$0 < \varepsilon < \bar{\varepsilon} := \min \left\{ \bar{\gamma}, 1 \right\} .$$

Indeed, take any feasible $\bar{x}$ with $\varphi_0(\bar{x}) < \inf \varphi_0 + \varepsilon^2$ and find $\eta \in (0, \varepsilon)$ such that $\varphi_0(\bar{x}) < \inf \varphi_0 + (\varepsilon - \eta)^2$. Considering the constraint sets Ω_i as defined in the proof of Theorem 5.28, observe that for the function

$$\varphi(x) := \varphi_0(x) + \delta(x; \Omega_1) + \ldots + \delta(x; \Omega_{m+r}) + \delta(x; \Omega), \quad x \in X ,$$

one has $\varphi(\bar{x}) < \inf_X \varphi + (\varepsilon - \eta)^2$. Then applying the *lower subdifferential variational principle* from Theorem 2.28(b), we find a feasible solution $u \in X$ to (5.23) and $u^* \in \widehat{\partial}\varphi(u)$ satisfying $\|u - \bar{x}\| < \varepsilon - \eta$ and

$$\|u^*\| < \varepsilon - \eta < \bar{\gamma}, \quad \varphi_0(u) < \inf \varphi_0 + (\varepsilon - \eta)^2 < \inf \varphi_0 + \varepsilon - \eta ,$$

which implies that $|\varphi_0(u) - \varphi_0(\bar{x})| < \varepsilon - \eta$.

Now we take $\gamma := \bar{\gamma}/(m + r + 1)$ and consider the weak* neighborhood

$$\widehat{V}^* := \bigcap_{j=1}^{p} V^*(v_j; \gamma)$$

of $0 \in X^*$. Employing the weak fuzzy sum rule from Lemma 5.27 to $u^* \in \widehat{\partial}\varphi(u)$ with the neighborhood $\widehat{V}^*$ and the number η and then following the proof of Theorem 5.28, we arrive at all the conclusions of this theorem. △

Our next result provides *strong* suboptimality conditions in a *qualified/normal* form for problems with partly Lipschitzian data under appropriate constraint qualifications. In what follows we use the notation

$$I(x) := \big\{ i \in \{1, \ldots, m + r\} \big| \ \varphi_i(x) = 0 \big\} \ \text{ and } \ \Lambda(x) := \big\{ \lambda_i \geq 0 \big| \ i \in I(x) \big\}$$

for any feasible solution x of problem (5.23).

Theorem 5.30 (strong suboptimality conditions under constraint qualifications). *Let X be Asplund, and let $\varepsilon > 0$. Assume that φ_0 is l.s.c., that Ω is closed, and that either φ_0 is SNEC or Ω is SNC on the set of ε-optimal solutions to* (5.23)*. Suppose also that on this set the functions $\varphi_1, \ldots, \varphi_{m+r}$ are locally Lipschitzian around x and the following qualification condition holds:*

if $x^*_\infty \in \partial^\infty \varphi_0(x)$, $x^* \in N(x; \Omega)$, $x^*_i \in \partial \varphi_i(x)$, $i \in \{1, \ldots, m\} \cap I(x)$,

$x^*_i \in \partial\varphi_i(x) \cup \partial(-\varphi_i)(x)$, $i \in \{m + 1, \ldots, m + r\}$, $\lambda_i \in \Lambda(x)$, and

$$x^*_\infty + \sum_{i \in I(x)} \lambda_i x^*_i + x^* = 0 ,$$

then $x^*_\infty = x^* = 0$ and $\lambda_i = 0$ for all $i \in I(x)$.

Under these assumptions one has the suboptimality conditions as follows: for every ε-optimal solution $\bar{x}$ to (5.23) *and every $\nu > 0$ there is an ε-optimal solution $\widehat{x}$ to this problem such that $\|\widehat{x} - \bar{x}\| \leq \nu$ and the estimate*

$$\Big\| \widehat{x}^*_0 + \sum_{i \in I(\widehat{x})} \widehat{\lambda}_i \widehat{x}^*_i + \widehat{x}^* \Big\| \leq \frac{\varepsilon}{\nu}$$

is satisfied with some $\widehat{x}^*_0 \in \partial\varphi_0(\widehat{x})$, $\widehat{x}^* \in N(\widehat{x}; \Omega)$, $\widehat{\lambda}_i \in \Lambda(\widehat{x})$,

$$\widehat{x}^*_i \in \partial\varphi_i(\widehat{x}) \ \ \textit{for} \ \ i \in \big\{1, \ldots, m\big\} \cap I(\widehat{x}), \quad \textit{and}$$

$$\widehat{x}^*_i \in \partial\varphi_i(\widehat{x}) \cup \partial(-\varphi_i)(\widehat{x}) \ \ \textit{for} \ \ i = m + 1, \ldots, m + r .$$

Conversely, if the above suboptimality conditions hold for any problem of minimizing a l.s.c. function $\varphi_0\colon X \to \overline{I\!R}$ *concave on its domain in a Banach space* X*, then* X *must be Asplund.*

Proof. As in the proof of Theorem 5.29, we consider the penalized function

$$\varphi(x) := \varphi_0(x) + \delta(x;\Omega_1) + \ldots + \delta(x;\Omega_{m+r}) + \delta(x;\Omega), \quad x \in X\ ,$$

and observe that $\bar{x}$ is an ε-optimal solution to the unconstrained problem of minimizing φ. Applying the *lower subdifferential variational principle* to this function with the given $\nu > 0$, we find an ε-optimal solution to the original problem (5.23) and $\widehat{x}^* \in \partial\varphi(\widehat{x})$ satisfying the estimates $\|\widehat{x} - \bar{x}\| \le \nu$ and $\|\widehat{x}^*\| \le \varepsilon/\nu$. Having the subdifferential equality

$$\partial\varphi(\widehat{x}) = \partial\Big(\varphi_0 + \sum_{i \in I(\widehat{x})} \delta(\cdot;\Omega_i) + \delta(\cdot;\Omega)\Big)(\widehat{x})\ ,$$

we apply to the latter sum the basic subdifferential sum rule from Theorem 3.36 under the assumptions made with the efficient subdifferential conditions for the SNC property of the constraint sets Ω_i obtained in Corollary 3.85. Then taking into account the representations of basic normals to the sets Ω_i from the proof of case (b) in Theorem 5.28 when all φ_i are Lipschitz continuous, we arrive at the desired suboptimality conditions.

It remains to prove the converse statement of the theorem. Let X be a Banach space, and let $\varphi\colon X \to I\!R$ be an arbitrary concave continuous function. By the continuity of φ, for any $\bar{x} \in X$ and $\varepsilon > 0$ there is $0 < \varepsilon_1 < \varepsilon$ such that $\varphi(\bar{x}) < \varphi(x) + 2\varepsilon$ whenever $x \in \bar{x} + \varepsilon_1 I\!B$. Consider the unconstrained optimization problem:

$$\text{minimize } \varphi_0(x) := \varphi(x) + \delta(x;\bar{x} + \varepsilon_1 I\!B), \quad x \in X\ .$$

Applying to the this problem the suboptimality conditions of the theorem, we find $\widehat{x} \in \bar{x} + (\varepsilon/2)I\!B$ such that $\partial\varphi_0(\widehat{x}) = \partial\varphi(\widehat{x}) \neq \emptyset$. Due to the basic subdifferential representation

$$\partial\varphi(\widehat{x}) = \mathop{\mathrm{Lim\,sup}}_{\substack{x \to \widehat{x} \\ \epsilon \downarrow 0}} \widehat{\partial}_\epsilon \varphi(x)$$

for arbitrary continuous functions on Banach spaces (see Theorem 1.89), for every $\epsilon > 0$ there is $x_\epsilon \in \bar{x} + \varepsilon I\!B$ with $\widehat{\partial}_\epsilon\varphi(x_\epsilon) \neq \emptyset$. This implies that, for any concave continuous function $\varphi\colon X \to I\!R$ and any $\epsilon > 0$, the set of points $\{x \in X|\ \widehat{\partial}_\epsilon\varphi(x) \neq \emptyset\}$ is dense in X. Then by Corollary 2.29 (see also the discussion after it) the space X must be Asplund. △

If φ_0 is Lipschitzian continuous on the set of ε-optimal solutions to (5.23), then φ_0 is automatically SNC and $\partial^\infty\varphi_0(x) = \{0\}$. In this case the qualification condition of Theorem 5.30 is a *constraint qualification.* Moreover, it reduces to the classical *Mangasarian-Fromovitz constraint qualification* when the functions $\varphi_1, \ldots, \varphi_{m+r}$ are strictly differentiable at such x and $\Omega = X$. Thus we arrive at the following consequence of the above theorem.

Corollary 5.31 (suboptimality under Mangasarian-Fromovitz constraint qualification). *Let X be Asplund, and let φ_0 be locally Lipschitzian on the set of ε-optimal solutions to* (5.23) *with $\Omega = X$ for some $\varepsilon > 0$. Assume that $\varphi_1, \ldots, \varphi_{m+r}$ are strictly differentiable and satisfy the Mangasarian-Fromovitz constraint qualification on the latter set. Then for every ε-optimal solution $\bar{x}$ to* (5.23) *and every $\nu > 0$ there are an ε-optimal solution $\widehat{x}$ to this problem, a subgradient $x_0^* \in \partial\varphi_0(\widehat{x})$, and multipliers $(\lambda_1, \ldots, \lambda_{m+r}) \in I\!\!R^{m+r}$ satisfying $\|\widehat{x} - \bar{x}\| \le \nu$, $\lambda_i \ge 0$ and $\lambda_i \varphi_i(\widehat{x}) = 0$ for $i = 1, \ldots, m$, and*

$$\Big\| x_0^* + \sum_{i=1}^{m+r} \lambda_i \nabla\varphi_i(\widehat{x}) \Big\| \le \frac{\varepsilon}{\nu} .$$

Proof. Follows directly from Theorem 5.30 due to $\partial\varphi(x) = \{\nabla\varphi(x)\}$ for strictly differentiable functions. △

Our final result in this section provides *strong suboptimality* conditions for Lipschitzian problems (5.23) with *no constraint qualifications.*

Corollary 5.32 (strong suboptimality conditions without constraint qualifications). *Let X be Asplund, and let $\varepsilon > 0$. Assume that Ω is closed and that all $\varphi_0, \ldots, \varphi_{m+r}$ are locally Lipschitzian on the set of ε-optimal solutions to* (5.23)*. Then for every $\nu > 0$ and every ε-optimal solution $\bar{x}$ to* (5.23) *there is an ε-optimal solution $\widehat{x}$ to this problem such that $\|\widehat{x} - \bar{x}\| \le \nu$ and*

$$\Big\| \sum_{i \in I(\widehat{x}) \cup \{0\}} \lambda_i \widehat{x}_i^* + \widehat{x}^* \Big\| \le \frac{\varepsilon}{\nu}, \qquad \sum_{i \in I(\widehat{x}) \cup \{0\}} \lambda_i = 1$$

with some $\lambda_i \ge 0$ for $i \in I(\widehat{x}) \cup \{0\}$, $\widehat{x}^* \in N(\widehat{x}; \Omega)$, $\widehat{x}_0^* \in \partial\varphi_0(\widehat{x})$,

$$\widehat{x}_i^* \in \partial\varphi_i(\widehat{x}) \ \text{ for } \ i \in \big\{1, \ldots, m\big\} \cap I(\widehat{x}), \quad \text{ and }$$

$$\widehat{x}_i^* \in \partial\varphi_i(\widehat{x}) \cup \partial(-\varphi_i)(\widehat{x}) \ \text{ for } \ i = m+1, \ldots, m+r .$$

Proof. Suppose first that the qualification condition of Theorem 5.30 is fulfilled. Then we have the suboptimality conditions of this theorem with some $\widehat{\lambda}_i \in \Lambda(\widehat{x})$. Now letting

$$\lambda := 1 + \sum_{i \in I(\bar{x})} \widehat{\lambda}_i, \quad \lambda_0 := \frac{1}{\lambda}, \ \text{ and } \ \lambda_i := \frac{\widehat{\lambda}_i}{\lambda} \ \text{ for } \ i \in I(\widehat{x}) ,$$

we arrive at the relations of the corollary with the multipliers $(\lambda_0, \ldots, \lambda_{m+r})$.

Assuming finally that the qualification condition of Theorem 5.30 is not fulfilled and taking into account that $x^*_\infty = 0$ by the Lipschitz continuity of φ_0, we find an ε-optimal solution $\widehat{x}$ to problem (5.23), multipliers $\widehat{\lambda}_i \geq 0$ for $i \in I(\widehat{x})$, not all zero, as well as dual elements $x^* \in N(\widehat{x};\Omega)$, $\widehat{x}^*_i \in \partial\varphi_i(\widehat{x})$ for $i \in \{1,\ldots,m\} \cap I(\widehat{x})$, and $\widehat{x}^*_i \in \partial\varphi_i(\widehat{x}) \cup \partial(-\varphi_i)(\widehat{x})$ for $i \in \{m+1,\ldots,m+r\}$ satisfying the equality

$$\sum_{i\in I(\widehat{x})} \widehat{\lambda}_i \widehat{x}^*_i + x^* = 0 .$$

Dividing the latter by $\lambda := \sum_{i\in I(\widehat{x})} \widehat{\lambda}_i > 0$, one has at the suboptimality conditions of the corollary with $\lambda_0 := 0$, $\lambda_i := \widehat{\lambda}_i/\lambda$ for $i \in I(\widehat{x})$, $\widehat{x}^* := x^*/\lambda$, and the same $\widehat{x}^*_i$ as above. △

Observe that if problem (5.23), as well as that considered in Subsect. 5.1.2, has *many geometric constraints* $x \in \Omega_i$, $i = 1,\ldots,n$, they can be obviously reduced to the one $x \in \Omega := \cap_{i=1}^n \Omega_i$ given by the set intersection. Then we may handle these constraints by using intersection rules for basic normals as in Subsect. 5.1.1 and thus extend the corresponding necessary optimality and suboptimality conditions of Subsects. 5.1.2–5.1.4 to optimization problems with many geometric constraints. To extend necessary optimality and suboptimality conditions expressed in terms of Fréchet normals to geometric constraints, as those in Theorems 5.21(i) and 5.29, one may use fuzzy intersection rules for Fréchet normals discussed in Subsect. 3.1.1.

Note also, besides deriving lower suboptimality conditions by applying the lower subdifferential variational principle, we can obtain their *upper* counterparts from the upper subdifferential variational principle of Theorem 2.30; see the paper [938] by Mordukhovich, Nam, and Yen for various results and discussions in this direction.

5.2 Mathematical Programs with Equilibrium Constraints

In this section we consider a special class of optimization problems known as *mathematical programs with equilibrium constraints* (MPECs). A characteristic feature of these problems is the presence, among other constraints, "equilibrium constraints" of the type $y \in S(x)$, where $S(x)$ often represents the solution map to a "lower-level" problem of parametric optimization. MPECs naturally appear in various aspects of hierarchical optimization and equilibrium theory as well as in many practical applications, especially those related to mechanical and economic modeling. We refer the reader to the books by Luo, Pang and Ralph [820], Outrata, Kočvara and Zowe [1031], and Facchinei and Pang [424] for systematic expositions, examples, and applications of such problems in finite-dimensional spaces.

Typically the equilibrium constraints $y \in S(x)$ in MPECs are solution maps to parametric variational inequalities and complementarity problems of different types. An important class of MPECs, which was actually a starting point of this active area of research and applications, contains problems of *bilevel programming* (that go back to Stackelberg games), where $S(x)$ is the solution map to a parametric problem of linear or nonlinear programming. Note that most MPECs, even in relatively simple cases of mathematical programs with *complementarity constraints*, are *essentially different* from standard problems of *nonlinear programming* with equality and inequality constraints; possible reductions lead to various *irregularities*, e.g., to the violation of the Mangasarian-Fromovitz constraint qualification and the like.

A general class of MPECs considered first in Subsect. 5.2.1 is given in the abstract form:

$$\text{minimize } \varphi(x,y) \text{ subject to } y \in S(x), \quad x \in \Omega\ , \tag{5.52}$$

where $S\colon X \rightrightarrows Y$ be a set-valued mapping between Banach spaces, $\varphi\colon X \times Y \to \overline{I\!R}$, and $\Omega \subset X$. Our main attention is paid to the case when the equilibrium map S is given in the form

$$S(x) := \big\{ y \in Y \big|\ 0 \in f(x,y) + Q(x,y) \big\} \tag{5.53}$$

with $f\colon X \times Y \to Z$ and $Q\colon X \times Y \rightrightarrows Z$, i.e., S describes solution maps to the *parametric variational systems*

$$0 \in f(x,y) + Q(x,y)$$

considered in Chap. 4 with their various specifications. As we know, model (5.53) covers solution maps to the classical variational inequalities and complementarity problems as well as to their extensions and modifications studied in Sect. 4.4. In what follows we are going to derive first-order necessary optimality conditions for general MPECs given in (5.52), (5.53) and for their important special cases. Our approach is mainly based on reducing MPECs to the optimization problems with *geometric* constraints considered in Sect. 5.1, with taking into account their special structures, and then on employing the sensitivity theory for parametric variational systems (coderivative estimates and efficient conditions for Lipschitzian stability) developed in Sect. 4.4. The results obtained involve second-order subdifferentials of extended-real-valued potentials defining variational and hemivariational inequalities in composite forms, which are the most interesting for applications.

5.2.1 Necessary Conditions for Abstract MPECs

In this subsection we consider abstract MPECs of type (5.52) and present necessary optimality conditions in lower and upper subdifferential forms. Such conditions are derived by reducing (5.52) to the standard form (5.1) with *two* geometric constraints and then employing the results of Theorem 5.5. In this way we take an advantage of the *product structure* on $X \times Y$, which allows us to impose mild qualification and SNC assumptions on the initial data of (5.52). Let us start with *upper subdifferential* necessary optimality conditions for general MPECs. Unless otherwise stated, we suppose that $\varphi\colon X \times Y \to \overline{I\!R}$ is an extended-real-valued function finite at reference points.

Theorem 5.33 (upper subdifferential optimality conditions for abstract MPECs). *Let $(\bar{x}, \bar{y})$ be a local optimal solution to* (5.52). *Assume that the spaces X and Y are Asplund and that the sets* $\operatorname{gph} S$ *and Ω are locally closed around $(\bar{x}, \bar{y})$ and $\bar{x}$, respectively. Assume also that either S is PSNC at $(\bar{x}, \bar{y})$ or Ω is SNC at $\bar{x}$, and that the mixed qualification condition*

$$D^*_M S(\bar{x}, \bar{y})(0) \cap \big(-N(\bar{x}; \Omega)\big) = \{0\}$$

is fulfilled. Then one has

$$-x^* \in D^*_N S(\bar{x}, \bar{y})(y^*) + N(\bar{x}; \Omega)$$

for every $(x^*, y^*) \in \widehat{\partial}^+ \varphi(\bar{x}, \bar{y})$.

Proof. Observe that $(\bar{x}, \bar{y})$ provides a local minimum to the function φ subject to the constraints $(x, y) \in \Omega_1 := \operatorname{gph} S$ and $(x, y) \in \Omega_2 := \Omega \times Y$ in the Asplund space $X \times Y$. Applying the upper subdifferential conditions of Theorem 5.5(i) to the latter problem, one can easily see that the PSNC property of Ω_1 at $(\bar{x}, \bar{y})$ with respect to X reduces to the PSNC property of the mapping S at this point, and that Ω_2 is always strongly PSNC at $(\bar{x}, \bar{y})$ with respect to Y being also SNC at this point if and only if Ω is SNC at $\bar{x}$. Moreover, the mixed qualification condition of the theorem clearly implies that the set system $\{\Omega_1, \Omega_2\}$ satisfies the limiting qualification condition at $(\bar{x}, \bar{y})$ in the sense of Definition 3.2. Thus we have, by Theorem 5.5(i), that

$$-\widehat{\partial}^+ \varphi(\bar{x}, \bar{y}) \subset N\big((\bar{x}, \bar{y}); \operatorname{gph} S\big) + N(\bar{x}; \Omega) \times \{0\}\ ,$$

which surely implies the upper subdifferential condition of this theorem. $\triangle$

The upper subdifferential conditions of Theorem 5.33 carry nontrivial information for MPECs when $\widehat{\partial}^+ \varphi(\bar{x}, \bar{y}) \neq \emptyset$. We have discussed in Subsect. 5.1.1 some important classes of functions φ satisfying this requirement. Recall that one automatically has $\widehat{\partial}^+ \varphi(\bar{x}, \bar{y}) \neq \emptyset$ when φ is either Fréchet differentiable at $(\bar{x}, \bar{y})$ or concave and continuous around $(\bar{x}, \bar{y})$. If both X and Y are Asplund, the latter case can be extended to the class of functions Lipschitz continuous around $(\bar{x}, \bar{y})$ and *upper regular* at this point, in particular, to semiconcave functions.

The next more conventional *lower subdifferential* conditions of for local minima in MPEC problems (5.52) have a different nature in comparison with the above upper subdifferential conditions being generally independent of them; cf. the discussions in Remark 5.4.

Theorem 5.34 (lower subdifferential optimality conditions for abstract MPECs). *Let* $(\bar{x}, \bar{y})$ *be a local optimal solution to* (5.52)*, where* X *and* Y *are Asplund, where* φ *is l.s.c. around* $(\bar{x}, \bar{y})$*, and where* Ω *and* $\operatorname{gph} S$ *are locally closed around* $(\bar{x}, \bar{y})$ *and* $\bar{x}$*, respectively. The following hold:*

(i) *In addition to the assumptions of Theorem* 5.33*, suppose that* φ *is SNEC at* $(\bar{x}, \bar{y})$*, and that the conditions*

$$(x^*_\infty, y^*_\infty) \in \partial^\infty \varphi(\bar{x}, \bar{y}), \quad 0 \in x^*_\infty + D^*_N S(\bar{x}, \bar{y})(y^*_\infty) + N(\bar{x}; \Omega)$$

are satisfied only for $x^*_\infty = y^*_\infty = 0$*; these additional assumptions are automatically fulfilled if* φ *is locally Lipschitzian around* $(\bar{x}, \bar{y})$*. Then there is* $(x^*, y^*) \in \partial\varphi(\bar{x}, \bar{y})$ *such that*

$$0 \in x^* + D^*_N S(\bar{x}, \bar{y})(y^*) + N(\bar{x}; \Omega) \ . \tag{5.54}$$

(ii) *Assume that both* S *and* Ω *are SNC at* $(\bar{x}, \bar{y})$ *and* $\bar{x}$*, respectively, and that the qualification condition*

$$\Big[(x^*_\infty, y^*_\infty) \in \partial^\infty \varphi(\bar{x}, \bar{y}),\ x^*_1 \in D^*_N S(\bar{x}, \bar{y})(y^*_\infty),\ x^*_2 \in N(\bar{x}; \Omega)\ ,$$

$$x^*_\infty + x^*_1 + x^*_2 = 0\Big] \Longrightarrow x^*_\infty = y^*_\infty = x^*_1 = x^*_2 = 0$$

is fulfilled. Then there is $(x^*, y^*) \in \partial\varphi(\bar{x}, \bar{y})$ *such that the optimality condition* (5.54) *holds.*

Proof. As in the proof of Theorem 5.33, we reduce (5.52) to minimizing $\varphi(x, y)$ subject to the geometric constraints: $(x, y) \in \Omega_1 = \operatorname{gph} S$ and $(x, y) \in \Omega_2 = \Omega \times Y$. Applying Theorem 5.5 to the latter problem, it is easy to check that the qualification condition (5.7) reduces to the one assumed in (i) of this theorem, and the lower subdifferential optimality condition (5.7) gives (5.54). Similarly we see that the qualification condition (5.8) reduces to the one assumed in (ii) of this theorem, which completes the proof. △

Based on the coderivative characterization of the Lipschitz-like property, we arrive at the following effective corollary of Theorems 5.33 and 5.34 that ensures the fulfillment of both upper and lower subdifferential optimality conditions above via *intrinsic requirements* on the initial data of the MPEC problem (5.52) under consideration.

Corollary 5.35 (upper and lower subdifferential conditions under Lipschitz-like equilibrium constraints). *Let* $(\bar{x}, \bar{y})$ *be a local optimal solution to the MPEC problem* (5.52) *with Asplund spaces* X *and* Y *and with locally closed sets* $\operatorname{gph} S$ *and* Ω*. Assume that* S *is Lipschitz-like around* $(\bar{x}, \bar{y})$*. Then one has*

$$-x^* \in D^*_N S(\bar{x}, \bar{y})(y^*) + N(\bar{x}; \Omega) \ \text{ for every } \ (x^*, y^*) \in \widehat{\partial}^+ \varphi(\bar{x}, \bar{y}) \ .$$

If in addition φ is locally Lipschitzian around $(\bar{x}, \bar{y})$, then

$$-x^* \in D_N^* S(\bar{x}, \bar{y})(y^*) + N(\bar{x}; \Omega) \ \text{ for some } \ (x^*, y^*) \in \partial\varphi(\bar{x}, \bar{y}) \, .$$

Proof. We know from Theorem 4.10 that S is Lipschitz-like around $(\bar{x}, \bar{y})$ if and only if $D_M^* S(\bar{x}, \bar{y})(0) = \{0\}$ and S is PSNC at $(\bar{x}, \bar{y})$ for closed-graph mappings between Asplund spaces. Thus the Lipschitz-like property of S implies the fulfillment of all the assumptions in Theorem 5.33 ensuring the upper subdifferential optimality condition. If in addition φ is Lipschitz continuous around $(\bar{x}, \bar{y})$, then $\partial^\infty \varphi(\bar{x}, \bar{y}) = \{0\}$, and hence we have the stated lower subdifferential optimality condition by Theorem 5.34(i). △

As follows from Corollary 5.35, the *Lipschitz-like property* of equilibrium constraints is a *constraint qualification* ensuring the *normal form* of both upper and lower subdifferential conditions for general MPECs. If now S is given in a parametric form of constraint and/or variational systems considered in Sects. 4.3 and 4.4, one can derive necessary optimality conditions of the normal (Karush-Kuhn-Tucker) type for problems (5.52) with the corresponding equilibrium constraints $y \in S(x)$ using the results of these sections, which provide upper estimates for $D_N^* S(\bar{x}, \bar{y})$ and efficient conditions for the Lipschitz-like property of such mappings.

We are *not* going to utilize here the results of Sect. 4.3 on parametric constraint systems in the form

$$S(x) := \big\{ y \in Y \big| \; g(x, y) \in \Theta, \ \ (x, y) \in \Omega \big\} \, ,$$

since constraints in this form are not specific for MPECs, and necessary optimality conditions for (5.52) obtained in this way don't actually bring new information in comparison with those, which have been derived in Subsects. 5.1.2 and 5.1.3 for problems with operator and functional constraints. Our main attention will be paid to necessary optimality conditions for MPECs in form (5.52) with equilibrium constraints governed by the *parametric variational systems* (5.53) considered in Sect. 4.4.

Before establishing such conditions in what follows, we conclude this subsection by deriving general necessary optimality conditions for abstract MPECs of type (5.52) in the *non-qualified* (Fritz John) form without imposing any constraint qualification.

Theorem 5.36 (upper and lower subdifferential conditions for non-qualified MPECs). *Let $(\bar{x}, \bar{y})$ be a local optimal solution to problem* (5.52). *Assume that the spaces X and Y are Asplund and that the sets* $\operatorname{gph} S$ *and Ω are locally closed around $(\bar{x}, \bar{y})$ and $\bar{x}$, respectively. Then there is $\lambda \in \{0, 1\}$ such that for every $(x^*, y^*) \in \widehat{\partial}^+ \varphi(\bar{x}, \bar{y})$ there exist $x_1^* \in D_N^* S(\bar{x}, \bar{y})(\lambda y^*)$ and $x_2^* \in N(\bar{x}; \Omega)$ satisfying*

$$\lambda x^* + x_1^* + x_2^* = 0, \quad (\lambda, x_1^*) \neq 0 \tag{5.55}$$

provided that either S is PSNC at $(\bar{x}, \bar{y})$ or Ω is SNC at $\bar{x}$.

If in addition φ is locally Lipschitzian around $(\bar{x}, \bar{y})$, then there are $\lambda \geq 0$, $(x^*, y^*) \in \partial\varphi(\bar{x}, \bar{y})$, $x_1^* \in D_N^* S(\bar{x}, \bar{y})(\lambda y^*)$, and $x_2^* \in N(\bar{x}; \Omega)$ satisfying (5.55).

Proof. This result follows from Theorems 5.33 and 5.34. Let us first justify the upper subdifferential conditions. If the mixed qualification condition of Theorem 5.33 is satisfied, we have (5.55) with $\lambda = 1$ by the assertion of that theorem under the assumptions made. If the latter qualification condition doesn't hold, there is $x^* \neq 0$ satisfying

$$x^* \in D_M^* S(\bar{x}, \bar{y})(0) \subset D_N^* S(\bar{x}, \bar{y})(0) \text{ and } -x^* \in N(\bar{x}; \Omega)\ .$$

Thus one gets (5.55) with $\lambda = 0$, $x_1^* = x^* \neq 0$, and $x_2^* = -x^*$.

Assume in addition that φ is Lipschitz continuous around $(\bar{x}, \bar{y})$ and apply assertion (i) of Theorem 5.34. Then we have relations (5.55) with $\lambda = 1$ and some $(x^*, y^*) \in \partial\varphi(\bar{x}, \bar{y})$ due to this assertion, provided that the mixed qualification condition

$$D_M^* S(\bar{x}, \bar{y})(0) \cap (-N(\bar{x}; \Omega)) = \{0\}$$

and the other assumptions of Theorem 5.34(i) are satisfied. If the latter qualification condition doesn't hold, we arrive at (5.55) with $\lambda = 0$ and $x_1^* \neq 0$ as in the above proof of the upper subdifferential condition. △

5.2.2 Variational Systems as Equilibrium Constraints

In this subsection we consider MPECs with equilibrium constraints defined by parameter-dependent generalized equations:

$$\text{minimize } \varphi(x, y) \text{ subject to } 0 \in f(x, y) + Q(x, y), \ \ x \in \Omega\ , \tag{5.56}$$

where $f\colon X \times Y \to Z$ and $Q\colon X \times Y \rightrightarrows Z$ are, respectively, single-valued and set-valued mappings between Banach (mostly Asplund) spaces. In other words, model (5.56) describes MPECs of type (5.52) governed by parametric variational systems $S(x)$, which are the solution maps (5.53) to perturbed generalized equations. Our goal is to derive necessary optimality conditions for local solutions to problem (5.56) in terms of its initial data (φ, f, Q, Ω). We are going to derive both *upper and lower subdifferential* optimality conditions for (5.56) based on the results of Theorems 5.33 and 5.34, i.e., to obtain necessary conditions in the *normal/qualified form*. Similarly to Theorem 5.36, one can deduce from the corresponding non-qualified necessary optimality conditions with possibly zero multipliers associated with the cost function.

To derive the desired necessary conditions from the above theorems, we need to express the assumptions and conclusions of these theorems involving the equilibrium constraints

$$y \in S(x) \Longleftrightarrow 0 \in f(x,y) + Q(x,y)$$

in (5.56) via the initial data f and Q. This can be done by employing the results of Sect. 4.4, which provide upper estimates for the coderivatives of such mappings (variational systems) S as well as sufficient conditions for their PSNC and Lipschitz-like properties. What one actually may derive from the results of Sect. 4.4 concerning applications to necessary optimality conditions in MPECs are upper estimates for $D_N^*S(\bar{x},\bar{y})$ and sufficient conditions for the SNC property of S via f and Q. In this way we get efficient conditions for the fulfillment of the constraint qualification

$$D_N^*S(\bar{x},\bar{y})(0) \cap \big(-N(\bar{x};\Omega)\big) = \{0\} \tag{5.57}$$

and the other assumptions and conclusions of Theorem 5.33 and 5.34 in terms of the initial data of the MPEC problem (5.56) and its specification.

Let us start with upper subdifferential necessary optimality conditions for the MPEC problem (5.56). The first theorem provides necessary conditions of this type for the case of equilibrium constraints governed by general parametric variational systems in (5.56).

Theorem 5.37 (upper subdifferential conditions for MPECs with general variational constraints). *Let* $(\bar{x},\bar{y})$ *be a local optimal solution to the MPEC problem* (5.56), *where* $f\colon X \times Y \to Z$ *and* $Q\colon X \times Y \rightrightarrows Z$ *are mappings between Asplund spaces. Assume that* f *is continuous around* $(\bar{x},\bar{y})$, *that* Ω *is locally closed around* $\bar{x}$, *and that the graph of* Q *is locally closed around* $(\bar{x},\bar{y},\bar{z})$ *with* $\bar{z} := -f(\bar{x},\bar{y})$. *Suppose also that one of the following assumptions* (a)–(c) *holds:*

(a) Ω *and* Q *are SNC at* $\bar{x}$ *and* $(\bar{x},\bar{y},\bar{z})$, *respectively, and the two qualification conditions are satisfied:*

$$\Big[(x^*,0) \in D_N^*f(\bar{x},\bar{y})(z^*) + D_N^*Q(\bar{x},\bar{y},\bar{z})(z^*),\;\; -x^* \in N(\bar{x};\Omega)\Big] \Longrightarrow x^* = 0\,,$$

$$\Big[(x^*,y^*) \in D_N^*f(\bar{x},\bar{y})(z^*) \cap \big(-D_N^*Q(\bar{x},\bar{y},\bar{z})(z^*)\big)\Big] \Longrightarrow x^* = y^* = z^* = 0\,;$$

the latter is equivalent to

$$\Big[0 \in \partial\langle z^*,f\rangle(\bar{x},\bar{y}) + D_N^*Q(\bar{x},\bar{y},\bar{z})(z^*)\Big] \Longrightarrow z^* = 0 \tag{5.58}$$

when f *is strictly Lipschitzian at* $(\bar{x},\bar{y})$.

(b) Ω *is SNC at* $\bar{x}$, $\dim Z < \infty$, f *is Lipschitz continuous around* $(\bar{x},\bar{y})$, *and the qualification conditions*

$$\Big[(x^*,0) \in \partial\langle z^*,f\rangle(\bar{x},\bar{y}) + D_N^*Q(\bar{x},\bar{y},\bar{z})(z^*),\;\; -x^* \in N(\bar{x};\Omega)\Big] \Longrightarrow x^* = 0$$

and (5.58) *are satisfied.*

(c) *Q is SNC at $(\bar{x},\bar{y},\bar{z})$, f is PSNC at $(\bar{x},\bar{y})$ (which is automatic when it is Lipschitz continuous around this point), and the qualification conditions from part* (a) *hold.*

Then for every $(x^*,y^*)\in\widehat{\partial}^+\varphi(\bar{x},\bar{y})$ there are $\widetilde{x}^*\in N(\bar{x};\Omega)$ and $z^*\in Z^*$ supporting the necessary optimality condition

$$(-x^*-\widetilde{x}^*,-y^*)\in D_N^*f(\bar{x},\bar{y})(z^*)+D_N^*Q(\bar{x},\bar{y},\bar{z})(z^*)\ .$$

Proof. Let us apply the upper subdifferential optimality conditions from Theorem 5.33 to problem (5.56), i.e., in the case when the equilibrium constraints $y\in S(x)$ are given in the variational/generalized equation form (5.53). It is easy to see that the continuity and closedness assumptions made on f and Q ensure the local closedness of S. To proceed further, we first assume that Ω is SNC at $\bar{x}$ and use the coderivative upper estimate for such mappings $S(\cdot)$ obtained in Theorem 4.46. Then one has

$$D_N^*S(\bar{x},\bar{y})(y^*)\subset\Big\{x^*\in X^*\Big|\ \exists z^*\in Z^*\ \text{ with }$$
$$(x^*,-y^*)\in D_N^*f(\bar{x},\bar{y})(z^*)+D_N^*Q(\bar{x},\bar{y},\bar{z})(z^*)\Big\}$$

and, substituting the latter into the qualification condition (5.57) and the upper subdifferential necessary condition of Theorem 5.33, we arrive at the conclusions of this theorem under the assumptions in (a) and (b).

Now we consider the remaining case when S is PSNC in Theorem 5.33 and provide efficient conditions in terms of f and Q ensuring the latter (even SNC) property for S. Actually it was done in the proof of Theorem 4.59 as a part of checking the coderivative criterion for the Lipschitz-like property of S based on the application of the SNC calculus from Theorem 3.84. Using these results, we arrive at the upper subdifferential optimality condition of the theorem under the assumptions in (c). △

Next we derive lower subdifferential optimality conditions for the MPEC (5.56) based on the application of Theorem 5.34 with the treatment of the equilibrium constraint S in (5.56) by the results of Theorem 4.46 and 3.84.

Theorem 5.38 (lower subdifferential conditions for MPECs with general variational constraints). *Let $(\bar{x},\bar{y})$ be a local optimal solution to* (5.56)*, where $f\colon X\times Y\to Z$ and $Q\colon X\times Y\rightrightarrows Z$ are mappings between Asplund spaces. Assume that φ is l.s.c. around $(\bar{x},\bar{y})$, that f is continuous around $(\bar{x},\bar{y})$, that Ω is locally closed around $\bar{x}$, and that the graph of Q is locally closed around $(\bar{x},\bar{y},\bar{z})$ with $\bar{z}=-f(\bar{x},\bar{y})$. The following assertions hold:*

(i) *Suppose that in addition to the assumptions of Theorem* 5.37 *the function φ is SNEC at $(\bar{x},\bar{y})$ and that the conditions*

$$(x^*_\infty - \widehat{x}^*, -y^*_\infty) \in D^*_N f(\bar{x},\bar{y})(z^*) + D^*_N Q(\bar{x},\bar{y},\bar{z})(z^*)\;,$$

$$(x^*_\infty, y^*_\infty) \in \partial^\infty\varphi(\bar{x},\bar{y}) \;\; \text{with some} \;\; \widehat{x}^* \in N(\bar{x};\Omega),\;\; z^* \in Z^*$$

are satisfied only when $x^*_\infty = y^*_\infty = 0$*; both of the latter assumptions are automatically fulfilled if* φ *is Lipschitz continuous around* $(\bar{x},\bar{y})$*. Then there are* $(x^*,y^*) \in \partial\varphi(\bar{x},\bar{y})$*,* $\widetilde{x}^* \in N(\bar{x};\Omega)$*, and* $z^* \in Z^*$ *such that*

$$(-x^* - \widetilde{x}^*, -y^*) \in D^*_N f(\bar{x},\bar{y})(z^*) + D^*_N Q(\bar{x},\bar{y},\bar{z})(z^*)\;. \tag{5.59}$$

(ii) *Suppose that both* Ω *and* Q *are SNC at* $\bar{x}$ *and* $(\bar{x},\bar{y},\bar{x})$*, respectively, that* f *is PSNC at* $(\bar{x},\bar{y})$*, and that the qualification conditions*

$$\Big[(x^*,y^*) \in D^*_N f(\bar{x},\bar{y})(z^*) \cap \big(-D^*_N Q(\bar{x},\bar{y},\bar{z})(z^*)\big)\Big] \Longrightarrow x^* = y^* = z^* = 0\;,$$

$$\Big[(x^*_\infty, y^*_\infty) \in \partial^\infty\varphi(\bar{x},\bar{y}),\; (x^*_1, -y^*_\infty) \in D^*_N f(\bar{x},\bar{y})(z^*) + D^*_N Q(\bar{x},\bar{y},\bar{z})(z^*)\;,$$

$$x^* \in N(\bar{x};\Omega),\; z^* \in Z^*,\; x^*_\infty + x^*_1 + x^*_2 = 0\Big] \Longrightarrow x^*_\infty = y^*_\infty = x^*_1 = x^*_2 = 0$$

are fulfilled. Then there are $(x^*,y^*) \in \partial\varphi(\bar{x},\bar{y})$*,* $\widetilde{x}^* \in N(\bar{x};\Omega)$*, and* $z^* \in Z^*$ *satisfying the optimality condition* (5.59).

Proof. To justify (i), we use Theorem 5.34(i) and then proceed similarly to the proof of Theorem 5.37 employing the upper coderivative estimate and the efficient conditions for the SNC property of the equilibrium map S obtained in Theorems 4.46 and 4.59, respectively. Assertion (ii) can be proved by the same based on Theorem 5.34(ii). △

Let us present efficient consequences of Theorems 5.37 and 5.38(i) ensuring the validity of the Lipschitz-like property of the equilibrium map S in (5.56) and hence the fulfillment of the qualification and PSNC conditions in Theorems 5.33 and 5.34(i).

Corollary 5.39 (upper and lower subdifferential conditions under Lipschitz-like variational constraints). *Let* $(\bar{x},\bar{y})$ *be a local optimal solution to* (5.56)*, where* $f\colon X \times Y \to Z$ *and* $Q\colon X \times Y \rightrightarrows Z$ *are mappings between Asplund spaces. Assume that* f *is continuous around* $(\bar{x},\bar{y})$*, that* Ω *is locally closed around* $\bar{x}$*, that the graph of* Q *is locally closed around* $(\bar{x},\bar{y},\bar{z})$ *with* $\bar{z} = -f(\bar{x},\bar{y})$*, and that the qualification conditions*

$$\Big[(x^*,0) \in D^*_N f(\bar{x},\bar{y})(z^*) + D^*_N Q(\bar{x},\bar{y},\bar{z})(z^*)\Big] \Longrightarrow x^* = 0\;,$$

$$\Big[(x^*,y^*) \in D^*_N f(\bar{x},\bar{y})(z^*) \cap \big(-D^*_N Q(\bar{x},\bar{y},\bar{z})(z^*)\big)\Big] \Longrightarrow x^* = y^* = z^* = 0$$

are satisfied. Then for every $(x^*,y^*) \in \widehat{\partial}^+\varphi(\bar{x},\bar{y})$ *there are* $\widetilde{x}^* \in N(\bar{x},\Omega)$ *and* $z^* \in Z^*$ *such that the optimality condition* (5.59) *holds. If in addition*

φ is Lipschitz continuous around $(\bar{x}, \bar{y})$, then there are $(x^*, y^*) \in \partial\varphi(\bar{x}, \bar{y})$, $\widetilde{x}^* \in N(\bar{x}; \Omega)$, and $z^* \in Z^*$ satisfying (5.59).

Proof. These assertions follow directly from Theorems 5.37 and 5.38(i), respectively. They are also consequences of Corollary 5.35 and Theorem 4.59 ensuring the Lipschitz-like property of the equilibrium constraint S in the MPEC problem (5.56). △

One can easily derive concretizations and simplifications of the results obtained in some special cases using coderivative representations for f and/or Q; compare, in particular, Sect. 4.4 for the cases of strictly differentiable mappings f and convex-graph multifunctions Q, as well as for parameter-independent fields $Q = Q(y)$.

In what follows we are going to discuss in more details the most interesting cases of variational constraints in (5.56) when the equilibrium map S is given in a *subdifferential form*, which covers the classical variational inequalities and complementarity problems as well as hemivariational inequalities and further generalizations. Let us pay the main attention to the two classes of *generalized variational inequalities* (GVIs) with a composite subdifferential structure considered in Sect. 4.4, where the equilibrium mapping S is given in forms (4.66) and (4.67). The first class of GVIs induces MPECs of the type:

$$\text{minimize } \varphi(x, y) \text{ subject to } 0 \in f(x, y) + \partial(\psi \circ g)(x, y),\ x \in \Omega \quad (5.60)$$

governed by single-valued mappings $f: X \times Y \to X^* \times Y^*$ and $g: X \times Y \to W$ between Banach spaces and by an extended-real-valued function $\psi: W \to \overline{I\!R}$. Let us derive both upper and lower subdifferential necessary optimality conditions in (5.60) for simplicity considering locally Lipschitzian cost functions φ in the case of lower subdifferential conditions. We start with the case of smooth and parameter-independent mappings $g: Y \to W$ in (5.60) with surjective derivatives allowing the space generality in necessary optimality conditions for (5.60) expressed in terms of the (normal) second-order subdifferential of $\varphi: W \to \overline{I\!R}$. Following the terminology of Sect. 4.4, we label such problems as MPECs governed by parametric *hemivariational inequalities* (HVIs) with *composite potentials*.

Theorem 5.40 (upper, lower subdifferential conditions for MPECs governed by HVIs with composite potentials). *Let $(\bar{x}, \bar{y})$ be a local optimal solution to problem* (5.60) *with $f: X \times Y \to Y^*$, $g: Y \to W$, and $\psi: W \to \overline{I\!R}$. Suppose that W is Banach, X is Asplund, Y is finite-dimensional and that the following assumptions hold:*

(a) *f is strictly differentiable at $(\bar{x}, \bar{y})$ with the surjective partial derivative $\nabla_x f(\bar{x}, \bar{y}): X \to Y^*$.*

(b) *g is $\mathcal{C}^1$ around $\bar{y}$ with the surjective derivative $\nabla g(\bar{y}): Y \to W$, and the mapping $\nabla g: Y \to \mathcal{L}(Y, W)$ is strictly differentiable at $\bar{y}$.*

(c) *Ω is locally closed around $\bar{x}$ and the graph of $\partial\psi$ is locally closed around $(\bar{w},\bar{v})$, where $\bar{w} := g(\bar{y})$ and where $\bar{v} \in W^*$ is a unique functional satisfying the relations*

$$-f(\bar{x},\bar{y}) = \nabla g(\bar{y})^*\bar{v}, \quad \bar{v} \in \partial\psi(\bar{w}) .$$

Then for every $(x^*,y^*) \in \widehat{\partial}^+\varphi(\bar{x},\bar{y})$ there is $u \in Y$ such that

$$\begin{aligned} -x^* &\in \nabla_x f(\bar{x},\bar{y})^*u + N(\bar{x};\Omega) , \\ -y^* &\in \nabla_y f(\bar{x},\bar{y})^*u + \nabla^2\langle\bar{v},g\rangle(\bar{y})^*u + \nabla g(\bar{y})^*\partial_N^2\psi(\bar{w},\bar{v})\big(\nabla g(\bar{y})u\big) \end{aligned} \tag{5.61}$$

provided that $u = 0$ is the only vector satisfying the system of inclusions

$$\begin{cases} 0 \in \nabla_x f(\bar{x},\bar{y})^*u + N(\bar{x};\Omega) , \\ 0 \in \nabla_y f(\bar{x},\bar{y})^*u + \nabla^2\langle\bar{v},g\rangle(\bar{y})^*u + \nabla g(\bar{y})^*\partial_N^2\psi(\bar{w},\bar{v})\big(\nabla g(\bar{y})u\big) . \end{cases}$$

In in addition φ is locally Lipschitzian around $(\bar{x},\bar{y})$, then there are $u \in Y$ and $(x^*,y^*) \in \partial\varphi(\bar{x},\bar{y})$ satisfying (5.61).

Proof. To establish the upper subdifferential conditions of the theorem, we employ the results of Theorem 5.37 under the assumptions in (c) for $Q(y) := \partial(\psi\circ g)(y)$. Taking into account the strict differentiability of f at $(\bar{x},\bar{y})$ with the surjectivity of $\nabla_x f(\bar{x},\bar{y})$ and the parameter-independence of Q, one has condition (5.58) automatically fulfilled while the first qualification condition in Theorem 5.37(a) reduces to

$$\Big[0 \in \nabla_x f(\bar{x},\bar{y})^*u + N(\bar{x};\Omega),\ 0 \in \nabla_y f(\bar{x},\bar{y})^*u + \partial^2(\psi\circ g)(\bar{y},\bar{z})(u)\Big] \Longrightarrow u=0$$

with $\bar{z} := -f(\bar{x},\bar{y})$ provided that the mapping $\partial(\psi\circ g)(\cdot)$ is locally closed-graph around $(\bar{y},\bar{z})$. Observe the SNC property of Q and PSNC property of f at the reference points follow immediately from the finite dimensionality of Y and the strict differentiability of f. Then, by the upper subdifferential optimality conditions of Theorem 5.37 applied to (5.60), for every upper subgradients $(x^*,y^*) \in \widehat{\partial}^+\varphi(\bar{x},\bar{y})$ there is $u \in Y$ such that

$$-x^* \in \nabla_x f(\bar{x},\bar{y})^*u + N(\bar{x};\Omega), \quad -y^* \in \nabla_y f(\bar{x},\bar{y})^*u + \partial^2(\psi\circ g)(\bar{y},\bar{z})(u) .$$

Using now the first-order subdifferential chain rule of Proposition 1.112(i), we have the equality

$$\partial(\psi\circ g)(y) = \nabla g(y)^*\partial\psi(w)$$

for all y close to $\bar{y}$ and $w = g(y)$, which implies that the graph of $\partial(\psi\circ g)(\cdot)$ is locally closed around $(\bar{y},\bar{z})$ if and only if the subdifferential mapping $\partial\psi(\cdot)$ is closed-graph around $(\bar{w},\bar{v})$. Applying further the second-order subdifferential chain rule of Theorem 1.127 to $\partial^2(\psi\circ g)(\bar{y},\bar{z})$ in the above relationships and taking into account that $\nabla g(\bar{y})^{**}=\nabla g(\bar{y})$ under the assumptions made, we arrive at the upper subdifferential conditions stated in the theorem.

If φ is locally Lipschitzian around $(\bar{x},\bar{y})$, the lower subdifferential conditions of the theorem are deduced by a similar way from Theorem 5.38(i). △

Recall that the assumption in (c) of the above theorem on the closed graph of $\partial\psi$ around $(\bar{w},\bar{v})$ automatically holds if ψ is either *continuous* around this point or *amenable* at $(\bar{w},\bar{v})$; see the definition in Subsect. 3.2.5.

The next result provides necessary optimality conditions for the MPEC problem (5.60) governed by *parameter-dependent* GVIs with composite potentials in finite-dimensional spaces under essentially less restrictive assumptions on f and g (but more restrictive on ψ) than those imposed in Theorem 5.40.

Theorem 5.41 (upper, lower subdifferential conditions for MPECs governed by GVIs with composite potentials). *Let* $(\bar{x},\bar{y})$ *be a local optimal solution to* (5.60), *where* $f\colon X\times Y\to X^*\times Y^*$ *and* $g\colon X\times Y\to W$ *are mappings between finite-dimensional spaces. Suppose that* f *is continuous around* $(\bar{x},\bar{y})$, *that* g *is twice continuously differentiable around this point, that* ψ *is l.s.c. around* $\bar{w}:=g(\bar{x},\bar{y})$, *and that* Ω *is locally closed around* $\bar{x}$. *Denote* $\bar{z}:=-f(\bar{x},\bar{y})\in\partial(\psi\circ g)(\bar{x},\bar{y})$ *and*

$$M(\bar{x},\bar{y}):=\big\{\bar{v}\in W^*\big|\ \bar{v}\in\partial\psi(\bar{w}),\quad \nabla g(\bar{x},\bar{y})^*\bar{v}=\bar{z}\big\}$$

and assume that:

(a) *The function* ψ *is lower regular around* $\bar{w}$ *and the graphs of* $\partial\psi$ *and* $\partial^\infty\psi$ *are closed when* w *is near* $\bar{w}$.

(b) *The following first-order and second-order qualification conditions for the composition* $\psi\circ g$ *hold:*

$$\partial^\infty\psi(\bar{w})\cap\ker\nabla g(\bar{x},\bar{y})^*=\{0\}\ ,$$

$$\partial^2\psi(\bar{w},\bar{v})(0)\cap\ker\nabla g(\bar{x},\bar{y})^*=\{0\}\ \ \textit{for all}\ \ \bar{v}\in M(\bar{x},\bar{y})\ .$$

(c) *One has the two relationships:*

$$\Big[(x^*,y^*)\in\bigcup_{\bar{v}\in M(\bar{x},\bar{y})}\Big(\nabla^2\langle\bar{v},g\rangle(\bar{x},\bar{y})u+\nabla g(\bar{x},\bar{y})^*\partial^2\psi(\bar{w},\bar{v})\big(\nabla g(\bar{x},\bar{y})u\big)\Big)$$

$$\bigcap\Big(-D^*f(\bar{x},\bar{y})(u)\Big)\Big]\Longrightarrow(x^*,y^*,u)=(0,0,0)\ ,$$

$$\Big[(x^*,0)\in D^*f(\bar{x},\bar{y})(u)+\bigcup_{\bar{v}\in M(\bar{x},\bar{y})}\Big(\nabla^2\langle\bar{v},g\rangle(\bar{x},\bar{y})(u)$$

$$+\nabla g(\bar{x},\bar{y})^*\partial^2\psi(\bar{w},\bar{v})\big(\nabla g(\bar{x},\bar{y})u\big)\Big),\ \ -x^*\in N(\bar{x};\Omega)\Big]\Longrightarrow x^*=0\ .$$

Then for every $(x^*, y^*) \in \widehat{\partial}^+\varphi(\bar{x}, \bar{y})$ *there is* $u \in X \times Y$ *such that*

$$(-x^*, -y^*) \in D^* f(\bar{x}, \bar{y})(u) + \bigcup_{\bar{v} \in M(\bar{x}, \bar{y})} \Big[\nabla^2\langle \bar{v}, g\rangle(\bar{x}, \bar{y})(u) + \nabla g(\bar{x}, \bar{y})^* \partial^2 \psi(\bar{w}, \bar{v})\big(\nabla g(\bar{x}, \bar{y})u\big)\Big] + N(\bar{x}; \Omega)\ . \tag{5.62}$$

If in addition φ *is Lipschitz continuous around* $(\bar{x}, \bar{y})$*, then there are elements* $(x^*, y^*) \in \partial\varphi(\bar{x}, \bar{y})$ *and* $u \in X \times Y$ *satisfying* (5.62).

Proof. Apply the upper and lower subdifferential optimality conditions of Theorems 5.33 and 5.34(i) in the case of

$$S(x) := \big\{ y \in Y \big|\ 0 \in f(x, y) + \partial(\psi \circ g)(x, y)\big\}\ .$$

Then use the upper estimate of $D^*_N S(\bar{x}, \bar{y})$ obtained in Theorem 4.50 based on the second-order subdifferential sum rule. In this way we arrive at the conclusions of the theorem under the assumptions made. △

Observe that the first relationship in (c) of the above theorem reduces to

$$0 \in \partial\langle u, f\rangle(\bar{x}, \bar{y}) + \bigcup_{\bar{v} \in M(\bar{x}, \bar{y})} \Big[\nabla^2\langle \bar{v}, g\rangle(\bar{x}, \bar{y})^* u + \nabla g(\bar{x}, \bar{y})^* \partial^2 \psi(\bar{w}, \bar{v})(\nabla g(\bar{x}, \bar{y})u)\Big] \Longrightarrow u = 0$$

when f is locally Lipschitzian around $(\bar{x}, \bar{y})$. The latter holds automatically if $g = g(y)$ and f is strictly differentiable at $(\bar{x}, \bar{y})$ with the surjective partial derivative $\nabla_x f(\bar{x}, \bar{y})$.

It happens that the first-order assumptions of Theorem 5.41 are automatically satisfied if the potential $\phi := \psi \circ g$ of the equilibrium constraints in (5.60) is *strongly amenable*; see the definition in Subsect. 3.2.5.

Corollary 5.42 (optimality conditions for MPECs with amenable potentials). *Let* $(\bar{x}, \bar{y})$ *be a local optimal solution to the MPEC problem* (5.60) *in finite dimensions with* Ω *closed around* $\bar{x}$*,* f *continuous around* $(\bar{x}, \bar{y})$*, and with the potential* $\phi = \psi \circ g$ *strongly amenable at this point. Suppose that the assumptions in* (c) *and the second-order qualification condition in* (b) *of Theorem* 5.41 *are satisfied. Then one has the upper subdifferential optimality condition of this theorem with no other assumptions. The above lower subdifferential condition holds as well if in addition* φ *is Lipschitz continuous around* $(\bar{x}, \bar{y})$.

Proof. It follows from Theorem 5.41 due to the properties of strongly amenable functions discussed in Subsect. 3.2.5 and the second-order subdifferential chain rule given in Corollary 3.76. △

Now we consider MPECs governed by the generalized variational inequalities with *composite fields*:

$$\text{minimize } \varphi(x,y) \text{ subject to } 0 \in f(x,y) + (\partial\psi \circ g)(x,y),\ x \in \Omega\ , \quad (5.63)$$

where $g\colon X \times Y \to W$, $\psi\colon W \to \overline{I\!R}$, and $f\colon X \times Y \to W^*$. The next theorem provides general necessary optimality conditions of the upper and lower subdifferential types for such MPECs.

Theorem 5.43 (upper, lower subdifferential conditions for MPECs governed by GVIs with composite fields). *Let $(\bar{x},\bar{y})$ be a local optimal solution to problem* (5.63) *with Ω closed around $\bar{x}$, $\bar{w} := g(\bar{x},\bar{y})$, and $\bar{z} := -f(\bar{x},\bar{y})$. The following assertions hold:*

(i) *Assume that X,Y are Asplund while W is Banach, that $g = g(y)$ is strictly differentiable at $\bar{y}$ with the surjective derivative $\nabla g(\bar{y})$, that f is strictly differentiable at $(\bar{x},\bar{y})$ with the surjective partial derivative $\nabla_x f(\bar{x},\bar{y})$, and that $u = 0 \in W^{**}$ is the only element satisfying*

$$0 \in \nabla_x f(\bar{x},\bar{y})^* u + N(\bar{x};\Omega), \quad 0 \in \nabla_y f(\bar{x},\bar{y})^* u + \nabla g(\bar{y})^* \partial_N^2 \psi(\bar{w},\bar{z})(u)\ .$$

Then for every $(x^*,y^*) \in \widehat{\partial}^+\varphi(\bar{x},\bar{y})$ there is $u \in W^{**}$ such that

$$\begin{aligned} -x^* &\in \nabla_x f(\bar{x},\bar{y})^* u + N(\bar{x};\Omega)\ , \\ -y^* &\in \nabla_y f(\bar{x},\bar{y})^* u + \nabla g(\bar{y})^* \partial_N^2 \psi(\bar{w},\bar{z})(u) \end{aligned} \quad (5.64)$$

provided that either Ω is SNC at $\bar{x}$ or $\partial\psi$ is SNC at $(\bar{w},\bar{z})$.

(ii) *Assume that X,Y,W,W^* are Asplund, that f and g are continuous around $(\bar{x},\bar{y})$, that the graph of $\partial\psi$ is norm-closed around $(\bar{w},\bar{z})$, that*

$$\partial_N^2 \psi(\bar{w},\bar{z})(0) \cap \ker D_N^* g(\bar{x},\bar{y}) = \{0\}\ ,$$

that $x^* = 0$ is the only element satisfying

$$(x^*,0) \in D_N^* f(\bar{x},\bar{y})(u) + D_N^* g(\bar{x},\bar{y}) \circ \partial_N^2 \psi(\bar{w},\bar{z})(u), \quad -x^* \in N(\bar{x};\Omega)$$

for some $u \in W^{**}$, and that $(x^*,y^*,u) = (0,0,0)$ is the only one satisfying

$$(x^*,y^*) \in D_N^* f(\bar{x},\bar{y})(u) \cap \big(-D_N^* g(\bar{x},\bar{y}) \circ \partial_N^2 \psi(\bar{w},\bar{z})(u)\big)\ .$$

Then for every upper subgradient $(x^*,y^*) \in \widehat{\partial}^+\varphi(\bar{x},\bar{y})$ there are $\widetilde{x}^* \in N(\bar{x};\Omega)$ and $u \in W^{**}$ such that

$$(-x^* - \widetilde{x}^*, -y^*) \in D_N^* f(\bar{x},\bar{y})(u) + D_N^* g(\bar{x},\bar{y}) \circ \partial_N^2 \psi(\bar{w},\bar{z})(u) \quad (5.65)$$

provided that either f is Lipschitz continuous around $(\bar{x},\bar{y})$ and $\dim W < \infty$, or g is PSNC at $(\bar{x},\bar{y})$ and $\partial\psi$ is SNC at $(\bar{w},\bar{z})$, or g is SNC at $(\bar{x},\bar{y})$ and $\partial\psi^{-1}$ is PSNC at $(\bar{z},\bar{w})$.

(iii) *Assume that φ is Lipschitz continuous around $(\bar{x},\bar{y})$ in addition to the assumptions in either* (i) *or* (ii). *Then there are, respectively, $(x^*,y^*) \in \partial\varphi(\bar{x},\bar{y})$ and $u \in W^{**}$ satisfying* (5.64) *and $(x^*,y^*) \in \partial\varphi(\bar{x},\bar{y})$, $\widetilde{x}^* \in N(\bar{x};\Omega)$, $u \in W^{**}$ satisfying* (5.65).

Proof. To establish (i), we employ the upper subdifferential optimality conditions of Theorem 5.33 and the coderivative formula of Proposition 4.53 for the equilibrium map

$$S(x) := \big\{ y \in Y \big|\, 0 \in f(x,y) + (\partial\psi \circ g)(x,y) \big\}$$

in (5.63). As follows from Theorem 1.22, the SNC property of S at $(\bar{x},\bar{y})$ is equivalent to the one of $\partial\psi$ at $(\bar{w},\bar{z})$ under the surjectivity assumption on $\nabla g(\bar{y})$. Then combining the assumptions and conclusions of Theorem 5.33 and Proposition 4.53, we justify (i). The proof of (ii) is similar based on the optimality conditions of Theorem 5.33 and the upper coderivative estimate of Theorem 4.54. The sufficient conditions of the SNC property of the composition $\partial\psi \circ g$ are derived from Theorem 3.98 as in the proof of Theorem 4.54.

The lower subdifferential optimality conditions in (iii) follow from Theorem 5.34(i) by employing the above arguments. △

Let us present some consequences of the upper and lower subdifferential assertions (ii) and (iii) of Theorem 5.43 in the case of strictly differentiable mappings f and g with finite-dimensional image spaces and possibly *non-surjective* derivatives when the relationships of the theorem admit essential simplifications.

Corollary 5.44 (optimality conditions for special MPECs governed by GVIs with composite fields). *Let $(\bar{x},\bar{y})$ be a local optimal solution to problem* (5.63) *with $f\colon X \times Y \to I\!R^m$ and $g\colon X \times Y \to I\!R^m$ strictly differentiable at $(\bar{x},\bar{y})$ and with $\Omega \subset X$ closed around $\bar{x}$. Assume that X and Y are Asplund, that* $\operatorname{gph}\partial\psi$ *is closed around $(\bar{w},\bar{z})$ (which is automatic for continuous and for amenable functions), that*

$$\partial^2\psi(\bar{w},\bar{z})(0) \cap \ker \nabla g(\bar{x},\bar{y})^* = \{0\}\,.$$

and that the system of inclusions

$$\begin{cases} x^* \in \nabla_x f(\bar{x},\bar{y})^* u + \nabla_x g(\bar{x},\bar{y})^* \partial^2\psi(\bar{w},\bar{z})(u), \quad -x^* \in N(\bar{x};\Omega)\,, \\[2ex] 0 \in \nabla_y f(\bar{x},\bar{y})^* u + \nabla_y g(\bar{x},\bar{y})^* \partial^2\psi(\bar{w},\bar{z})(u) \end{cases}$$

has only the trivial solution $x^* = u = 0$. Then for every $(x^*,y^*) \in \widehat{\partial}^+\varphi(\bar{x},\bar{y})$ there is $u \in I\!R^m$ such that

$$\begin{aligned} &-x^* \in \nabla_x f(\bar{x},\bar{y})^* u + \nabla_x g(\bar{x},\bar{y})^* \partial^2\psi(\bar{w},\bar{z})(u) + N(\bar{x};\Omega)\,, \\ &-y^* \in \nabla_y f(\bar{x},\bar{y})^* u + \nabla_y g(\bar{y})^* \partial^2\psi(\bar{w},\bar{z})(u)\,. \end{aligned} \tag{5.66}$$

If in addition the cost function φ is Lipschitz continuous around $(\bar{x},\bar{y})$, then there are $(x^*,y^*) \in \partial\varphi(\bar{x},\bar{y})$ and $u \in I\!R^m$ satisfying (5.66).

Proof. This easily follows from Theorem 5.43(ii,iii) due to the coderivative representation for strictly differentiable functions. △

Remark 5.45 (optimality conditions for MPECs under canonical perturbations). Consider the class of MPECs

$$\text{minimize } \varphi(x,z,y) \text{ subject to } z \in f(x,y) + Q(x,y), \quad (x,z) \in \Omega\ , \tag{5.67}$$

with equilibrium constraints given by solution maps to *canonically perturbed* generalized equations

$$\Sigma(x,z) := \big\{ y \in Y \big| \, z \in f(x,y) + Q(x,y) \big\} \ .$$

One can treat (5.67) as a particular case of the MPECs (5.56) with respect to the parameter pair $p := (x,z)$. Hence the above necessary optimality conditions obtained for (5.56) readily imply the corresponding results for (5.67). One the other hand, the canonical structure of parameter-dependent equilibrium constraints in (5.67) allows us to derive special results for this class of MPECs. This can be done on the base of the upper and lower subdifferential optimality conditions of Theorem 5.33 and 5.34 (see also Corollary 5.35) under the *Lipschitz-like* property of the equilibrium map $\Sigma(\cdot)$ efficient conditions for which are obtained in Subsect. 4.4.3. The latter results automatically induce necessary optimality conditions for MPECs (5.67) and also for (5.56) that are generally independent of those obtained in Corollary 5.39 and their specifications. We refer the reader to the corresponding results and discussions in Subsect. 4.4.3.

5.2.3 Refined Lower Subdifferential Conditions for MPECs via Exact Penalization

Here we develop another approach to necessary optimality conditions for MPECs governed by parametric variational systems of type (5.56). In contrast to the preceding subsection, this approach is not directly based on applying calculus rules to the general optimality conditions of Subsect. 5.2.1 but involves a preliminary penalization procedure, which leads to more subtle *lower subdifferential* results in some settings. On the other hand, the penalization approach doesn't allow us to derive necessary optimality conditions of the upper subdifferential type given in Subsect. 5.2.2.

To begin with, we define a Lipschitzian property of set-valued mappings *at* reference points of their graphs.

Definition 5.46 (calmness of set-valued mappings). *Let $F\colon X \rightrightarrows Y$ be a set-valued mapping between Banach spaces, and let $(\bar{x},\bar{y}) \in \operatorname{gph} F$. Then F is* CALM *at $(\bar{x},\bar{y})$ with modulus $\ell \geq 0$ if there are neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that*

$$F(x) \cap V \subset F(\bar{x}) + \ell \|x - \bar{x}\| I\!B \ \text{ for all } \ x \in U\ . \tag{5.68}$$

If one may choose $V = Y$ in (5.68) with $\bar{x} \in \operatorname{dom} F$, the mapping F is CALM *at the point $\bar{x}$.*

The latter calmness property of set-valued mappings at points of their domains is also known as *upper Lipschitzian* property of F at $\bar{x} \in \operatorname{dom} F$ (in the sense of Robinson). Following the terminology of this book, the graph-localized calmness property (5.68) may be alternatively called the *upper-Lipschitz-like* property of F at $(\bar{x}, \bar{y}) \in \operatorname{gph} F$.

One can see that the above calmness/upper Lipschitzian properties of set-valued mappings are less restrictive than their "full" counterparts from Definition 1.40, where $\bar{x}$ is replaced by $u \in U$ that varies around $\bar{x}$ together with x. On the other hand, the calmness properties, in contrast to the full Lipschitzian ones, are *not robust* with respect to perturbations of the reference point $\bar{x}$. Moreover, the above calmness properties don't imply that $(\bar{x}, \bar{y}) \in \operatorname{int}(\operatorname{gph} F)$ and $\bar{x} \in \operatorname{int}(\operatorname{dom} F)$, respectively. Note also that for single-valued mappings $F = f\colon X \to Y$ the calmness property of f *doesn't reduce* to the standard local Lipschitzian property of single-valued mappings. A classical setting, due to Robinson, when a mapping $F\colon I\!R^n \rightrightarrows I\!R^m$ is calm/upper Lipschitzian at every point $\bar{x} \in \operatorname{dom} F$ but may not be locally Lipschitzian around $\bar{x}$ is the one when F is *piecewise polyhedral*, i.e., its graph is expressible as the union of finitely many (convex) polyhedral sets. Such mappings are important for applications in mathematical programming with finitely many linear constraints of equality and inequality types.

In this subsection we use the calmness property (5.68) for the study of MPECs governed by parametric variational systems. First let us consider the following optimization problem containing constraints in the form of *nonparametric generalized equations*:

$$\text{minimize } \varphi(t) \text{ subject to } 0 \in F(t), \quad t \in \Omega , \tag{5.69}$$

where $F\colon T \rightrightarrows Z$ is a set-valued mapping between Banach spaces, $\varphi\colon T \to \overline{I\!R}$, and $\Omega \subset T$. Since the constraints in (5.69) can be written as $t \in F^{-1}(0) \cap \Omega$, this problem is a special case of the optimization problem (5.12) considered in Subsect. 5.1.2. Applying the necessary optimality conditions obtained there for the latter problem unavoidably requires the Lipschitz-like property of F^{-1} (or the metric regularity property of F) around a minimum point due to the qualification condition (5.15) with $\Theta = \{0\}$. However, this property may be relaxed by using preliminary an *exact penalization* procedure.

Indeed, problem (5.69) can be equivalently written as:

$$\text{minimize } \varphi(t) \text{ subject to } z \in F(t), \quad z = 0, \quad t \in \Omega .$$

The next auxiliary result, which is strongly related to Theorem 5.16, provides a reduction of (5.69) to general MPECs considered in Subsect. 5.2.1.

Lemma 5.47 (exact penalization under generalized equation constraints). *Let $\bar{t}$ be a local optimal solution to problem* (5.69) *in the framework of Banach spaces. Assume that φ is Lipschitz continuous around $\bar{t}$ with modulus ℓ_φ and that the mapping $(F^{-1} \cap \Omega)(z) := F^{-1}(z) \cap \Omega$ is calm at $(0, \bar{t})$*

with modulus ℓ. Then there are neighborhoods V of $\bar{t}$ and U of $0 \in Z$ such that $(\bar{t}, 0) \in T \times Z$ solves the penalized problem

$$\text{minimize } \psi(t,z) := \varphi(t) + \mu\|z\| \text{ subject to } z \in F(t) \cap U, \quad t \in \Omega \cap V$$

provided that $\mu \geq \ell_\varphi \cdot \ell$.

Proof. Since $F^{-1} \cap \Omega$ is calm at $(0, \bar{t})$ with modulus $\ell \geq 0$, there are neighborhoods V of $\bar{t}$ and U of $0 \in Z$ such that for some $\widehat{t} \in F^{-1}(0) \cap \Omega$ one has the estimate

$$\|t - \widehat{t}\| \leq \ell\|z\| \text{ whenever } t \in F^{-1}(z) \cap \Omega \cap V, \quad z \in U\ .$$

Using this and the Lipschitz continuity of φ with modulus ℓ_φ, we get

$$\begin{aligned}\varphi(\bar{t}) &\leq \varphi(\widehat{t}) = \varphi(t) + \big(\varphi(\widehat{t}) - \varphi(t)\big)\\ &\leq \varphi(t) + \ell_\varphi\|\widehat{t} - t\| \leq \varphi(t) + \ell_\varphi \cdot \ell\|z\|\\ &\leq \varphi(t) + \mu\|z\|\end{aligned}$$

whenever $t \in F^{-1}(z) \cap \Omega \cap V$, $z \in U$, and $\mu \geq \ell_\varphi \cdot \ell$. △

Theorem 5.48 (necessary optimality conditions under generalized equation constraints). *Let $\bar{t}$ be a local optimal solution to problem* (5.69)*, where T and Z are Asplund and where Ω and* $\operatorname{gph} F$ *are locally closed around $\bar{t}$ and $(\bar{t}, 0)$, respectively. Assume that φ is locally Lipschitzian around $\bar{t}$ with modulus ℓ_φ, that $F^{-1} \cap \Omega$ is calm at $(0, \bar{t})$ with modulus ℓ, and that the mixed qualification condition*

$$D^*_M F(\bar{t}, 0)(0) \cap \big(-N(\bar{t}; \Omega)\big) = \{0\}$$

is fulfilled. Suppose also that either F is PSNC at $(\bar{t}, 0)$ or Ω is SNC at $\bar{t}$. Then for any $\mu \geq \ell_\varphi \cdot \ell$ there is $z^* \in Z^*$ with $\|z^*\| \leq \mu$ such that

$$0 \in \partial\varphi(\bar{t}) + D^*_N F(\bar{t}, 0)(z^*) + N(\bar{t}; \Omega)\ .$$

If in particular F is given in the form

$$F(t) := g(t) + \Theta \quad \textit{with} \quad g\colon T \to Z \quad \textit{and} \quad \Theta \subset Z\ ,$$

then there is $z^* \in -N(-g(\bar{t}); \Theta)$ with $\|z^*\| \leq \mu$ such that

$$0 \in \partial\varphi(\bar{t}) + D^*_N g(\bar{t})(z^*) + N(\bar{t}; \Omega) \tag{5.70}$$

provided that g is continuous around $\bar{t}$, that Θ is locally closed around $-g(\bar{t})$, that the qualification condition

$$D^*_M g(\bar{t})(0) \cap \big(-N(\bar{t}; \Omega)\big) = \{0\} \tag{5.71}$$

holds, and that either g is PSNC at $\bar{t}$ or Ω is SNC at this point.

Proof. From the viewpoint of necessary optimality conditions the penalized optimization problem in Lemma 5.47 can be equivalently written as :

$$\text{minimize } \varphi(t)+\mu\|z\| \text{ subject to } z\in F(t),\ \ t\in\Omega\ ,$$

which is a special form of the general MPECs (5.52). Now applying to this problem the result of Theorem 5.34(i) in the case of Lipschitzian cost functions and then using the subdifferential sum rule of Theorem 2.33(c) for $\varphi(t)+\mu\|z\|$, we justify the first part of the theorem.

Now let $F(t):=g(t)+\Theta$ and apply the general statement of the theorem to this particular mapping, which is the sum of g and $\Theta(t):=\Theta$ for all $t\in T$. It is easy to see that the latter mapping is PSNC at any $(\bar{t},\bar{z})\in T\times\Theta$ and that its both coderivatives $D^*=D^*_N,D^*_M$ are computed by

$$D^*\Theta(\bar{t},\bar{z})(z^*)=\begin{cases}0 & \text{if } -z^*\in N(\bar{z};\Theta)\ ,\\ \emptyset & \text{otherwise.}\end{cases}$$

Then we have by the coderivative sum rules of Theorem 3.10 applied to both coderivatives $D^*=D^*_N,D^*_M$ of the sum $f+\Theta$ that

$$D^*F(\bar{t},0)(z^*)\subset\begin{cases}D^*g(\bar{t})(z^*) & \text{if } -z^*\in N(-g(\bar{t});\Theta)\ ,\\ \emptyset & \text{otherwise.}\end{cases}$$

Substituting this into the general qualification and necessary optimality conditions of the theorem, we arrive at relations (5.71) and (5.70), respectively. It remains to observe that the PSNC property of g at $\bar{t}$ implies the one for $F=g+\Theta$ at $(\bar{t},0)$ due to Theorem 3.88. △

Note that the qualification condition (5.71) holds and g is PSNC at $\bar{t}$ if it is Lipschitz continuous around this point. Observe also that the above approach based on the exact penalization *doesn't* allow us to deduce *upper subdifferential* optimality conditions for (5.69) from the ones for (5.52), since the required sum rule is not generally valid for the Fréchet upper subdifferential of the sum $\varphi(\cdot)+\mu\|\cdot\|$ unless φ is Fréchet differentiable at a minimum point.

Next we derive necessary optimality conditions for the MPEC problem with equilibrium constraints governed by parametric variational systems:

$$\text{minimize } \varphi(x,y) \text{ subject to } 0\in f(x,y)+Q(x,y),\ \ (x,y)\in\Omega\ , \tag{5.72}$$

where $f\colon X\times Y\to Z$, $Q\colon X\times Y\rightrightarrows Z$, and $\Omega\subset X\times Y$. Observe that problem (5.72) is more general than (5.56), where the geometric constraints don't depend on y. The results obtained below are based on reducing the MPEC problem (5.72) to the one in (5.69) governed by nonparametric generalized equations and then on employing Theorem 5.48 and calculus rules. Note that these results are generally different from those obtained in Subsect. 5.2.2 even in the case of y-independent geometric constraints.

There are at least *two ways* of reducing (5.72) to (5.69). The first one is directly by considering

$$F(t) = F(x, y) := f(x, y) + Q(x, y)$$

and then using the general optimality conditions of Theorem 5.48. The second way consists of reducing (5.72) to a special form of Theorem 5.48 with

$$\begin{aligned} &F(x, y) := g(x, y) + \Theta, \quad \Theta := \operatorname{gph} Q, \quad \text{and} \\ &g(x, y) := (-x, -y, f(x, y)) . \end{aligned} \tag{5.73}$$

Let us explore just the latter way for brevity. It leads to the following necessary optimality conditions for the MPEC problem (5.72).

Theorem 5.49 (optimality conditions for MPECs via penalization). *Let* $(\bar{x}, \bar{y})$ *be a local optimal solution to problem* (5.72)*, where* $f\colon X \times Y \to Z$ *and* $Q\colon X \times Y \rightrightarrows Z$ *are mappings between Asplund spaces. Assume that* φ *is Lipschitz continuous around* $(\bar{x}, \bar{y})$ *with modulus* ℓ_φ*, that* f *is continuous around this point, and that the sets* Ω *and* $\operatorname{gph} Q$ *are locally closed around* $(\bar{x}, \bar{y})$ *and* $(\bar{x}, \bar{y}, \bar{z})$ *with* $\bar{z} := -f(\bar{x}, \bar{y})$*, respectively. Suppose also that the mapping* $G\colon X \times Y \times Z \rightrightarrows X \times Y$ *given by*

$$G(u, v, w) := \big\{(x, y) \in \Omega \big|\ \big(u + x, v + y, w - f(x, y)\big) \in \operatorname{gph} Q\big\}$$

is calm at $(0, 0, 0, \bar{x}, \bar{y})$ *with modulus* ℓ*, that the qualification condition*

$$D^*_M f(\bar{x}, \bar{y})(0) \cap \big(-N((\bar{x}, \bar{y}); \Omega)\big) = \{0\}$$

is fulfilled, and that either f *is PSNC at* $(\bar{x}, \bar{y})$ *or* Ω *is SNC at this point. Then there are* $(x^*, y^*, z^*) \in X^* \times Y^* \times Z^*$ *with* $\|(x^*, y^*, z^*)\| \le \ell_\varphi \cdot \ell$ *and* $(x^*, y^*) \in D^*_N Q(\bar{x}, \bar{y}, \bar{z})(z^*)$ *satisfying*

$$(-x^*, -y^*) \in \partial\varphi(\bar{x}, \bar{y}) + D^*_N f(\bar{x}, \bar{y})(z^*) + N((\bar{x}, \bar{y}); \Omega) ,$$

which implies that

$$0 \in \partial\varphi(\bar{x}, \bar{y}) + D^*_N f(\bar{x}, \bar{y})(z^*) + D^*_N Q(\bar{x}, \bar{y}, \bar{z})(z^*) + N((\bar{x}, \bar{y}); \Omega) .$$

Proof. Apply the special case of Theorem 5.48 with the data of (5.73). Since

$$g(x, y) = (-x, -y, 0) + (0, 0, f(x, y)) ,$$

it is easy to observe from Theorem 1.70 that g is PSNC at $(\bar{x}, \bar{y})$ if and only if f is PSNC at this point. Then using the sum rules from Theorem 1.62(ii) for both coderivatives $D^* = D^*_N, D^*_M$, we have

$$D^*g(\bar{x}, \bar{y})(x^*, y^*, z^*) = (-x^*, -y^*) + D^*f(\bar{x}, \bar{y})(z^*) .$$

Thus we get the qualification condition and the necessary optimality condition of the theorem directly from (5.71) and (5.70) of Theorem 5.48. △

For further applications of Theorem 5.49 one needs to provide efficient conditions ensuring the calmness property of the mapping G in this theorem. As we know, G is calm at the reference point if it is Lipschitz-like around it. Since G is given in the form of *constraint systems*, sufficient conditions for the latter property follow from the results of Subsect. 4.3.2. Let us implement these results considering for simplicity the case when the *base* f in the equilibrium constraint of (5.72) is *strictly Lipschitzian* at $(\bar{x}, \bar{y})$. In this case f is automatically PSNC at $(\bar{x}, \bar{y})$ and the qualification condition of Theorem 5.49 is satisfied; hence the Lipschitz-like property of G implies the necessary optimality conditions for the MPEC problem in the latter theorem.

Corollary 5.50 (equilibrium constraints with strictly Lipschitzian bases). *In the general framework of Theorem* 5.49*, suppose that* f *is strictly Lipschitzian at* $(\bar{x}, \bar{y})$*, that* Q *is SNC at* $(\bar{x}, \bar{y}, \bar{z})$*, and that the relation*

$$(x^*, y^*) \in \big[\partial\langle z^*, f\rangle(\bar{x}, \bar{y}) + N((\bar{x}, \bar{y}); \Omega)\big] \cap \big(- D_N^* Q(\bar{x}, \bar{y}, \bar{z})(z^*)\big) \quad (5.74)$$

holds only for $x^* = y^* = z^* = 0$*. Then there is* $z^* \in Z^*$ *such that the necessary optimality condition*

$$0 \in \partial\varphi(\bar{x}, \bar{y}) + \partial\langle z^*, f\rangle(\bar{x}, \bar{y})(z^*) + D_N^* Q(\bar{x}, \bar{y}, \bar{z})(z^*) + N((\bar{x}, \bar{y}); \Omega)$$

is satisfied.

Proof. Let $h: X \times Y \times Z \times X \times Y \to X \times Y \times Z$ be defined by

$$h(u, v, w, x, y) := \big(u + x, v + y, w - f(x, y)\big) .$$

Then the mapping G in Theorem 5.49 is represented as the constraint system

$$G(u, v, w) = \Big\{(x, y) \in X \times Y \Big| \; h(u, v, w, x, y) \in \operatorname{gph} Q ,$$
$$(u, v, w, x, y) \in X \times Y \times Z \times \Omega\Big\} .$$

To ensure the Lipschitz-like property of G around $(0, 0, 0, \bar{x}, \bar{y})$, we apply the result of Corollary 4.41. It is easy to see from the structure of G that h is strictly Lipschitzian at $(0, 0, 0, \bar{x}, \bar{y})$ *if and only if* f is strictly Lipschitzian at $(\bar{x}, \bar{y})$ and that the set $\{0\} \times N((\bar{x}, \bar{y}); \Omega)$ is PSNC at $(0, 0, 0, \bar{x}, \bar{y})$ with respect to the first three components in the product space $X \times Y \times Z \times X \times Y$. Then the qualification condition (4.44) of Corollary 4.41 applied to the above mapping G reads that $(u^*, v^*, w^*, x^*, y^*, z^*) = (0, 0, 0, 0, 0, 0)$ is the only solution to the inclusion system

$$\begin{cases} (u^*, v^*, w^*, 0, 0) \in \partial\big\langle (x^*, y^*, z^*), h\big\rangle(0, 0, 0, \bar{x}, \bar{y}) + \{0\} \times N((\bar{x}, \bar{y}); \Omega) , \\ (x^*, y^*, z^*) \in N((\bar{x}, \bar{y}, \bar{z}); \operatorname{gph} Q), \end{cases}$$

which is equivalent to require that the above relations yields $x^* = y^* = z^* = 0$. By the elementary subdifferential sum rule we have

$$\partial\big\langle (x^*, y^*, z^*), h\big\rangle(0, 0, 0, \bar{x}, \bar{y}) = \big(x^*, y^*, z^*, (x^*, y^*) + \partial\big\langle -z^*, f\big\rangle(\bar{x}, \bar{y})\big) ,$$

and therefore the above qualification condition is equivalent to say that system (5.74) has only the trivial solution $(x^*, y^*, z^*) = (0, 0, 0)$. This completes the proof of the corollary. △

Similarly to the preceding subsection one can derive, based on calculus rules, further consequences of Theorem 5.49 and Corollary 5.50 for the cases of equilibrium constraints in (5.72) governed by generalized variational inequalities with *composite potentials* and *composite fields*, i.e., when

$$Q(x, y) = \partial(\psi \circ g)(x, y) \quad \text{and} \quad Q(x, y) = (\partial\psi \circ g)(x, y) .$$

The results obtained in this way are expressed in terms of the *second-order subdifferentials* of extended-real-valued functions ψ. Leaving this to the reader, we present next another corollary of Theorem 5.49 for a special class of MPECs in finite dimensions, where the mapping G in Theorem 5.49 may *not be Lipschitz-like* but still satisfies the weaker *calmness* property.

Consider the following MPEC problem with both variational and nonvariational constraints of the *polyhedral* type:

$$\begin{aligned} &\text{minimize } \ \varphi(x, y) \ \text{ subject to} \\ &0 \in A_1x + B_1y + c_1 + Q(A_2x + B_2y + c_2), \quad Lx + My + e \le 0 , \end{aligned} \tag{5.75}$$

where $x \in I\!\!R^n$, $y \in I\!\!R^m$, $Q\colon I\!\!R^s \rightrightarrows I\!\!R^m$, $C \in R^{p\times n}$, $D \in I\!\!R^{p\times m}$, and A_i, B_i, c_i $(i = 1, 2)$ are matrices and vectors of appropriate dimensions.

Corollary 5.51 (optimality conditions for MPECs with polyhedral constraints). *Let* $(\bar{x}, \bar{y})$ *be a local optimal solution to* (5.75). *Assume that* φ *is Lipschitz continuous around* $(\bar{x}, \bar{y})$ *and that* Q *is piecewise polyhedral, i.e., its graph is a union of finitely many polyhedral sets. Then there are vectors* $(x^*, y^*) \in \partial\varphi(\bar{x}, \bar{y})$, $(u^*, v^*) \in I\!\!R^s \times I\!\!R^m$, *and* $z^* := (\lambda_1, \ldots, \lambda_p) \in I\!\!R^p$ *satisfying the relations*

$$0 = x^* + A_2^*u^* + A_1^*v^* + L^*z^*, \quad 0 = y^* + B_2^*u^* + B_1^*v^* + M^*z^* ,$$

$$u^* \in D^*Q\big(A_2\bar{x} + B_2\bar{y} + c_2, -A_1\bar{x} - B_1\bar{y} - c_1\big)(v^*), \quad \textit{and}$$

$$\lambda_j \ge 0, \quad \lambda_j\big((L\bar{x})_j + (M\bar{y})_j + e_j\big) = 0 \ \textit{ for } \ j = 1, \ldots, p .$$

Proof. As mentioned above, a *piecewise polyhedral set is calm* at every point of its domain. Since both sets gph Q and $\Omega := \{(x, y) \in I\!R^{n+m} |\ Lx+My+e \le 0\}$ are piecewise polyhedral, the mapping

$$G(u, v) := \Big\{(x, y) \in \Omega \Big|\ (u + A_2x + B_2y + c_2, v - A_1x - B_1y - c_1) \in \text{gph}\, Q\Big\}$$

is calm at $(0, 0)$, and hence all the assumptions of Theorem 5.49 are fulfilled. Then taking into account the particular structure of the initial data in (5.75) as a special form of (5.72), we deduce the necessary optimality conditions of the corollary directly from the ones in Theorem 5.49. △

To illustrate the obtained necessary optimality conditions for MPECs, we consider the following *example*, where the equilibrium constraint is governed by a one-dimensional *variational inequality* of the so-called *second kind*, i.e., defined by the subdifferential of a convex continuous function:

$$\text{minimize } \tfrac{1}{2}x - y \ \text{ subject to } \ 0 \in y - x + \partial|y|, \quad x \in [-2, 0]\ .$$

It is simple to observe that $(\bar{x}, \bar{y}) = (-1, 0)$ is the unique global solution to this problem, which is a special case of (5.75) with $Q(y) = \partial|y|$. Since this mapping Q is obviously piecewise polyhedral, all the assumptions of Corollary 5.51 are fulfilled. To check the necessary optimality conditions of this corollary, we need to compute the coderivative $D^*Q(0, 1)$, i.e., the basic normal cone to the graph of $\partial|y|$ at the reference point. It can be easily done geometrically applying the representation of Theorem 1.6, which gives

$$N((0, -1); \text{gph}\, \partial|\cdot|) = \Big\{(u, v) \in I\!R^2 \Big|\ \text{ either } \ uv = 0, \ \text{ or } \ u > 0 \ \text{ and } \ v < 0\Big\}\ .$$

Then the necessary optimality conditions of Corollary 5.51 reduce to

$$(\tfrac{1}{2}, -\tfrac{1}{2}) \in N\big((0, -1); \text{gph}\, \partial|\cdot|\big)\ ,$$

which is definitely satisfied.

Remark 5.52 (implementation of optimality conditions for MPECs). The most challenging task in applications of the above optimality conditions to specific MPECs is to compute (or to obtain efficient upper estimates) of the coderivatives for the field multifunctions Q. In the cases when Q is given in the subdifferential form $Q(\cdot) = \partial\psi(\cdot)$, as well as in the composite subdifferential forms considered in Subsect. 5.2.2, this reduces to computing or estimating the second-order subdifferentials of the corresponding potentials. Some examples and discussions on such calculations were presented in Subsects. 1.3.5 and 4.4.2; see, in particular, Example 4.67 related to mechanical applications. As mentioned above, the second-order subdifferentials for general classes of nonsmooth functions important in optimization and various applications were computed in the papers of Dontchev and Rockafellar [364] and Mordukhovich and Outrata [939]. Many specific calculations and applications in this direction can be found in the papers by Kočvara, Kružik and

Outrata [689], Kočvara and Outrata [691, 690], Lucet and Ye [816], Mordukhovich, Outrata and Červinka [940], Outrata [1024, 1025, 1027, 1030], Ye [1338, 1339], Ye and Ye [1343], Zhang [1360], and the references therein.

In particular, complete calculations have been done by Outrata [1027] for MPECs with *implicit complementarity constraints* given by

$$f(x,y) \geq 0, \quad y - g(x,y) \geq 0, \quad \langle f(x,y), y - g(x,y) \rangle = 0 \; ,$$

where f and g are smooth single-valued mappings from $I\!R^n \times I\!R^m$ into $I\!R^m$. Such problems are important for various engineering, economic, and mechanical applications. They correspond to the standard nonlinear complementarity problems when $g = 0$. It is easy to see that the implicit complementarity constraints can be equivalently written as the equilibrium constraints

$$0 \in f(x,y) + (\partial\varphi \circ h)(x,y)$$

with $\varphi(\cdot) := \delta(\cdot; I\!R^m_+)$ and $h(x,y) := y - g(x,y)$. The main part of calculations for such MPECs consists of computing the basic normal cone to the graph of $N(\cdot; I\!R^m_+)$, which is done by Outrata in [1024]. Based on the nonsmooth calculus developed above, we can extend these results to *nonsmooth* complementarity problems with nondifferentiable mappings f and g.

5.3 Multiobjective Optimization

This section is devoted to *multiobjective* constrained optimization problems, where objective/cost functions may not be real-valued, i.e., optimization is conducted with respect to more general *preference relations*. Such problems, which probably first arose in economic modeling (see, e.g., Chap. 8), are certainly important for applications. They are also interesting mathematically having often significant differences in comparison with single-objective minimization/maximimization problems and requiring special considerations.

In what follows we study general classes of multiobjective/vector optimization problems with various constraints in infinite-dimensional spaces. The involved concepts of optimality (efficiency, equilibrium) are given by preference relations that cover the standard ones well-recognized in the theory and applications while extending and generalizing them in several directions.

First we consider multiobjective problems, where the notion of optimality for a cost mapping $f\colon X \to Z$ between Banach spaces is described by means of a *generalized order relation* defined by a given subset $M \subset Z$, which may be generally nonconic and nonconvex having an empty interior. Such a notion of (f,M)-*optimality* is actually induced by the concept of *local extremal points* of set systems (see Sect. 2.1) and extends the classical concepts of Pareto/weak Pareto optimality as well as their generalizations. To derive necessary optimality conditions for multiobjective problems of this type with various constraints, we employ the *extremal principle* of Sect. 2.2, together with the developed generalized differential and SNC calculi, that lead us to comprehensive results for such multiobjective as well as related minimax problems in terms of our basic normals and subgradients. Note that our approach *doesn't* rely on any *scalarization* techniques and results that are conventionally used in the study of multiobjective optimization problems.

Along with the multiobjective problems of the above type, we consider some classes of constrained problems, where the optimality concept is generally described by an *abstract nonreflexive preference* relation satisfying certain transitivity and local satiation requirements. Such preference relations may go far beyond generalized Pareto/weak Pareto concepts of optimality being useful for some important applications. To handle multiobjective problems of the latter type, we develop an *extended extremal principle* that applies not just to system of sets but to systems of *set-valued mappings*. Roughly speaking, the main difference between the conventional and extended extremal principle is that the latter allows us to take into account a local *deformation* of sets, rather than their (linear) *translation*, in extremal systems. In this way we derive new necessary optimality conditions for constrained multiobjective problems with general nonreflexive preference relations under reasonable assumptions. We discuss some specifications of the results obtained and their relationships with previous developments.

5.3.1 Optimal Solutions to Multiobjective Problems

Let us start with an abstract concept of optimality that covers conventional notions of optimal solutions to multiobjective problems and is induced by the concept of set extremality from Definition 2.1.

Definition 5.53 (generalized order optimality). *Given a single-valued mapping $f\colon X \to Z$ between Banach spaces and a set $0 \in \Theta \subset Z$, we say that a point $\bar{x} \in X$ is* LOCALLY (f, Θ)-OPTIMAL *if there are a neighborhood U of $\bar{x}$ and a sequence $\{z_k\} \subset Z$ with $\|z_k\| \to 0$ as $k \to \infty$ such that*

$$f(x) - f(\bar{x}) \notin \Theta - z_k \quad \text{for all} \quad x \in U \quad \text{and} \quad k \in I\!N\,. \tag{5.76}$$

The set Θ in Definition 5.53 may be viewed as a generator of an extended *order/preference relation* between $z_1, z_2 \in Z$ defined via $z_1 - z_2 \in \Theta$. In the scalar case of $Z = I\!R$ and $\Theta = I\!R_-$, the above optimality notion is clearly reduced to the standard local optimality.

Note that we don't generally assume that Θ is either convex or its interior is nonempty. If Θ is a convex subcone of Z with $\operatorname{ri}\Theta \neq \emptyset$, then the above optimality concept covers the conventional concept of optimality (called sometimes *Slater optimality*) requiring that there is no $x \in U$ with $f(x) - f(\bar{x}) \in \operatorname{ri}\Theta$. This extends the notion of *weak Pareto optimality/efficiency* corresponding to $f(x) - f(\bar{x}) \in \operatorname{int}\Theta$ in the above relations. To reduce it to the notion in Definition 5.53, we take $z_k := -z_0/k$ for $k \in I\!N$ in (5.76) with some $z_0 \in \operatorname{ri}\Theta$. The standard notion of *Pareto optimality* can be formulated in these terms as the absence of $x \in U$ for which $f(x) - f(\bar{x}) \in \Theta$ and $f(\bar{x}) - f(x) \notin \Theta$. Of course, the Pareto-type notions can be written in the classical terms of *utility functions* when $\Theta = I\!R^m_-$.

On the other hand, it is convenient for the further study to formulate the following *minimax problem* over a compact set as a problems of of multiobjective optimization.

Example 5.54 (minimax via multiobjective optimization). *Let $\bar{x}$ be a local optimal solution to the minimax problem:*

$$\text{minimize } \varphi(x) := \max\big\{\langle z^*, f(x)\rangle \big| \, z^* \in \Lambda\big\}, \quad x \in X \; ,$$

where $f\colon X \to Z$ and where $\Lambda \subset Z^*$ is weak* sequentially compact subset of Z^* such that there is $z_0 \in Z$ with $\langle z^*, z_0\rangle > 0$ for all $z^* \in \Lambda$. Suppose for simplicity that $\varphi(\bar{x}) = 0$. Then $\bar{x}$ is locally (f, Θ)-optimal in the sense of Definition 5.53 *with*

$$\Theta := \big\{z \in Z \big| \, \langle z^*, z\rangle \le 0 \;\; \text{whenever} \;\; z^* \in \Lambda\big\} \; .$$

Proof. Taking z_0 given above, one can easily check that (5.76) holds with the sequence $z_k := z_0/k$, $k \in I\!N$. △

We'll show in the next subsection that the (f, Θ)-optimality under general constraints can be comprehensively handled on the base of the *extremal principle* of Sect. 2.2 and the efficient representations of basic normals to generalized epigraphs obtained in Lemma 5.23 together with the SNC calculus in infinite dimensions.

However, there are multiobjective problems arising, e.g., in control applications and game-theoretical frameworks, where appropriate concepts of optimality require *nonlinear transformations* of sets in extremal systems instead of their *linear translations* as in Definition 5.53. This can be formalized by considering general preference relations on Z satisfying certain requirements that allow us to use suitable techniques of variational analysis.

Given a subset $Q \subset Z \times Z$, we say that z_1 is *preferred to* z_2 and write $z_1 \prec z_2$ if $(z_1, z_2) \in Q$. A preference $\prec$ is *nonreflexive* if the corresponding set Q doesn't contain the diagonal (z, z). In the sequel we consider nonreflexive preference relations satisfying the following requirements.

Definition 5.55 (closed preference relations). *Let*

$$\mathcal{L}(z) := \big\{u \in Z \big| \, u \prec z\big\}$$

be a LEVEL SET *at $z \in Z$ with respect to the given preference $\prec$. We say that $\prec$ is* LOCALLY SATIATED *around $\bar{z}$ if $z \in \operatorname{cl} \mathcal{L}(z)$ for all z in some neighborhood of $\bar{z}$. Furthermore, $\prec$ is* ALMOST TRANSITIVE *on Z provided that for all $u \prec z$ and $v \in \operatorname{cl} \mathcal{L}(u)$ one has $v \prec z$. The preference relation $\prec$ is called* CLOSED *around $\bar{z}$ if it is locally satiated and almost transitive simultaneously.*

Note that, while the local satiation property definitely holds for any reasonable preference, the almost transitivity requirement may be violated for some natural preferences important in applications, in particular, for those related to the (f, Θ)-optimality in Definition 5.53. Indeed, consider the case of the so-called "generalized Pareto" preference induced by a closed cone $\Theta \subset Z$ such that $z_1 \prec z_2$ if and only if $z_1 - z_2 \in \Theta$ and $z_1 \neq z_2$. This is, of course, a particular case of Definition 5.53. The next proposition completely describes the requirements on Θ under which this preference is almost transitive. Recall that a cone Θ is *pointed* if $\Theta \cap (-\Theta) = \{0\}$.

Proposition 5.56 (almost transitive generalized Pareto). *The generalized Pareto preference $\prec$ defined above is almost transitive if and only if the cone $\Theta \subset Z$ is convex and pointed.*

Proof. Let us first show that the cone Θ is convex if the above preference $\prec$ is almost transitive. Taking arbitrary elements $z_1, z_2 \in \Theta \setminus \{0\}$, $\lambda \in (0,1)$, and $a \in Z$, we define $u := a + \lambda z_1$ and $v := a - (1-\lambda)z_2$. Since $\lambda z_1 \neq 0$, one has $u \prec a$ and $a \prec v$. By the almost transitivity property we have $u \prec v$, which means that $\lambda z_1 + (1-\lambda)z_2 = u - v \in \Theta$, i.e., Θ is convex.

To prove that Θ is pointed under the transitivity of $\prec$, we take $z \in \Theta \cap (-\Theta)$ and put $u := a + z$ and $v := a - (-z)$. If $z \neq 0$, then the almost transitivity property implies that $u \prec v$, which gives $0 = u - v \in \Theta \setminus \{0\}$. This is a contradiction, and so $z = 0$.

To prove the converse statement of the proposition, we assume that Θ is convex and pointed, and take $v \in \operatorname{cl} \mathcal{L}(u)$ with $v \prec z$. Then there are $z_1, z_2 \in \Theta$ such that $v = u + z_1$, $z = u - z_2$, and $z_2 \neq 0$. By the convexity of Θ one has $(v - z)/2 = z_1/2 + z_2/2 \in \Theta$, and so $v \in \operatorname{cl} \mathcal{L}(z)$. The assumption on $v = z$ yields $z_1 = -z_2 \neq 0$, which contradicts the pointedness of Θ. Thus we have $v \prec z$ and complete the proof of the proposition. △

Invoking the characterization of Proposition 5.56, we observe that the almost transitivity condition of Definition 5.55 may fail to fulfill for important special cases of generalized Pareto preferences (and hence in the setting of Definition 5.53). It happens, in particular, for the preference described by the following *lexicographical ordering* on $\mathbb{R}^m$.

Example 5.57 (lexicographical order). *Let $\prec$ be a preference on $\mathbb{R}^m$, $m \geq 3$, defined by the lexicographical order, i.e., $u \prec v$ if there is an integer $j \in \{0, \ldots, m-1\}$ such that $u_i = v_i$ for $i = 1, \ldots, j$ and $u_{j+1} < v_{j+1}$ for the corresponding components of the vectors $u, v \in \mathbb{R}^m$. Then this preference is locally satiated but not almost transitive on $\mathbb{R}^m$.*

Proof. It is easy to check that the lexicographical preference $\prec$ is locally satiated on $\mathbb{R}^m$. On the other hand, this preference is generated by the convex cone $\Theta := \{(z_1, \ldots, z_m) \in \mathbb{R}^m |\ z_1 \leq 0\}$, which is *not pointed*, and thus the almost transitivity property is violated by Proposition 5.56. To illustrate this, let us consider the vectors

$$z := (0, 0, 1, \ldots, 0), \quad u := (0, \ldots, 0), \quad v := (0, 1, 1, 0, \ldots, 0)$$

in $I\!R^m$ and the sequence $v_k := (-1/k, 1, 1, 0, \ldots, 0) \to v$ as $k \to \infty$. Then $u \prec z$, $v_k \prec u$, but $v \not\prec z$ while $v \in \operatorname{cl} \mathcal{L}(u)$. △

In the rest of this section we derive necessary optimality conditions in constrained multiobjective problems, where concepts of local optimality for a (vector) mapping $f\colon X \to Z$ at $\bar{x}$ are given by a generalized order Θ on Z in the sense of Definition 5.53 as well as by closed preferences on Z in the sense of Definition 5.55. The results obtained in both cases are based on somewhat different techniques and are generally independent.

5.3.2 Generalized Order Optimality

This subsection concerns necessary optimality conditions for constrained multiobjective problems with local optimal solutions understood in the sense of Definition 5.53. This definition suggests the possibility of using the extremal principle for set systems to derive necessary conditions for such a generalized order optimality being actually inspired by the concept of local extremal points for system of sets. Our main goal is to obtain necessary conditions in the *pointbased/exact* form involving generalized differential constructions at the reference optimal solution. We mostly focus on *qualified* necessary optimality conditions taking into account that they directly imply, in our *dual-space* approach, the corresponding non-qualified optimality conditions similarly to the derivation in Sects. 5.1 and 5.2.

To get general results on necessary condition for generalized order optimality under minimal assumptions, we need an extended version of the *exact extremal principle* from Theorem 2.22 for the case of two sets in *products* of Asplund spaces. This result involves the PSNC and strong PSNC properties of sets in the product space $X_1 \times X_2$ with respect to an index set $J \subset \{1, 2\}$ that may be empty; see Definition 3.3. Note that both PSNC and strong PSNC properties are automatic if $J = \emptyset$, and both reduce to the SNC property of sets when $J = \{1, 2\}$. Our primary interest in the following lemma is an *intermediate* case, which takes into account the product structure that is essential for the main result of this subsection.

Lemma 5.58 (exact extremal principle in products of Asplund spaces). *Let $\bar{x} \in \Omega_1 \cap \Omega_2$ be a local extremal point of the sets $\Omega_1, \Omega_2 \subset X_1 \times X_2$ that are supposed to be locally closed around $\bar{x}$, and let $J_1, J_2 \subset \{1, 2\}$ with $J_1 \cup J_2 = \{1, 2\}$. Assume that both spaces X_1 and X_2 are Asplund, and that Ω_1 is PSNC at $\bar{x}$ with respect to J_1 while Ω_2 is strongly PSNC at $\bar{x}$ with respect to J_2. Then there exists $x^* \neq 0$ satisfying*

$$x^* \in N(\bar{x}; \Omega_1) \cap \big(-N(\bar{x}; \Omega_2) \big) .$$

Proof. Applying the approximate extremal principle of Theorem 2.20 to the extremal system $\{\Omega_1, \Omega_2, \bar{x}\}$ and taking a sequence $\varepsilon_k \downarrow 0$ as $k \to \infty$, we find $u_k \in \Omega_1$, $v_k \in \Omega_2$, $u_k^* \in \widehat{N}(u_k; \Omega_1)$, and $v_k^* \in \widehat{N}(v_k; \Omega_2)$ such that

$$\|u_k - \bar{x}\| < \varepsilon_k, \quad \|v_k - \bar{x}\| < \varepsilon_k, \quad \|u_k^* + v_k^*\| < \varepsilon_k \ ,$$

$$\tfrac{1}{2} - \varepsilon_k < \|u_k^*\| < \tfrac{1}{2} + \varepsilon_k, \quad \tfrac{1}{2} - \varepsilon_k < \|v_k^*\| < \tfrac{1}{2} + \varepsilon_k$$

for all $k \in I\!N$. Since the sequences $\{u_k^*\}$ and $\{v_k^*\}$ are bounded in the duals to Asplund spaces, they weak* converge to some u^* and v^* along subsequences. Thus $u^* \in N(\bar{x}; \Omega_1)$, $v^* \in N(\bar{x}; \Omega_2)$, and $u^* + v^* = 0$.

It remains to show that $x^* := u^* \neq 0$. Assuming the contrary and using the strong PSNC property of Ω_2 at $\bar{x}$ with respect to J_2, we get $\|v_{jk}^*\| \to 0$ as $k \to \infty$ for each $j \in J_2$. This implies by the relations of the approximate extremal principle that $\|u_{jk}^*\| \to 0$ for each $j \in J_2$ as well. Since Ω_1 is assumed to be PSNC at $\bar{x}$ with respect to J_1, this gives that $\|u_{jk}^*\| \to 0$ for all $j \in J_1$. Due to $J_1 \cup J_2 = \{1, 2\}$, we conclude that $\|u_k^*\| \to 0$ as $k \to \infty$. This is a contradiction, which completes the proof of the lemma. △

Based on the above lemma and generalized differential calculus, we next derive necessary optimality conditions for constrained multiobjective problems, where the optimality is understood in the sense of Definition 5.53. Given a mapping $f: X \to Z$ and sets $\Omega \subset X$ and $\Theta \subset Z$, we consider the generalized epigraph $\mathcal{E}(f, \Omega, \Theta)$ defined in (5.37) and the restriction $f_\Omega := f|_\Omega$ of f on Ω. Recall that the notion of strong coderivative normality needed for the last statement of the theorem is introduced in Definition 4.8 and some sufficient conditions for this property are presented in Proposition 4.9.

Theorem 5.59 (necessary conditions for generalized order optimality). *Let $f: X \to Z$ be a mapping between Asplund spaces, and let $\Omega \subset X$ and $\Theta \subset Z$ be such sets that $\bar{x} \in \Omega$ and $0 \in \Theta$. Suppose that the point $\bar{x}$ is locally (f, Θ)-optimal relative to Ω (i.e., subject to the constraint $x \in \Omega$). The following assertions hold:*

(i) *Assume that the set*

$$\mathcal{E}(f, \Omega, \Theta) := \{(x, z) \in X \times Z \,|\, f(x) - z \in \Theta, \ x \in \Omega\}$$

is locally closed around $(\bar{x}, \bar{z})$ with $\bar{z} := f(\bar{x})$ and that $\dim Z < \infty$. Then there is $z^* \in Z^*$ satisfying

$$(0, -z^*) \in N\big((\bar{x}, \bar{z}); \mathcal{E}(f, \Omega, \Theta)\big), \quad z^* \neq 0\ , \tag{5.77}$$

which always implies that $z^* \in N(0; \Theta)$, and it also implies that $0 \in D_N^* f_\Omega(\bar{x})(z^*)$ provided that f is continuous around $\bar{x}$ relative to Ω and that Ω and Θ are locally closed around $\bar{x}$ and 0, respectively. If in addition f is Lipschitz continuous around $\bar{x}$ relative to Ω, then (5.77) is equivalent to

$$0 \in \partial\langle z^*, f_\Omega\rangle(\bar{x}), \quad z^* \in N(0;\Theta) \setminus \{0\}\ . \tag{5.78}$$

(ii) *Let f be continuous around $\bar{x}$ relative to Ω, let Ω and Θ be locally closed around $\bar{x}$ and* 0, *respectively, and let*

(a) *either Θ be SNC at* 0,

(b) *or f_Ω^{-1} be PSNC at $(\bar{z}, \bar{x})$.*

Then there is $z^* \in Z^*$ satisfying

$$0 \neq z^* \in N(0;\Theta) \cap \ker D_N^* f_\Omega(\bar{x})\ , \tag{5.79}$$

which is equivalent to (5.78) *and to* (5.77) *provided that f is Lipschitz continuous around $\bar{x}$ relative to Ω and that the restriction f_Ω is strongly coderivatively normal at this point.*

Proof. Assume for simplicity that $\bar{z} = f(\bar{x}) = 0$. Then, according to Definition 5.53, the point $(\bar{x}, 0) \in X \times Z$ is a *local extremal point* of the set system $\{\Omega_1, \Omega_2\}$, where

$$\Omega_1 := \mathcal{E}(f, \Omega, \Theta), \quad \Omega_2 := \mathrm{cl}\, U \times \{0\}\ ,$$

and where U is a neighborhood of the local optimality from (5.76) with $x \in \Omega$.

Consider first the framework of assertion (i), where the set Ω_1 is locally closed around $(\bar{x}, 0)$ even if f may not be continuous around $\bar{x}$; cf. Theorem 5.24. Since U is a neighborhood of $\bar{x}$ and Z is finite-dimensional, the set Ω_2 is SNC at $(\bar{x}, 0)$. Thus in this case we can use the conventional version of the exact extremal principle from Theorem 2.22, which immediately gives (5.77) and hence $z^* \in N(0;\Theta)$. The other conclusions in (i) follow from Lemma 5.23 under the assumptions made.

Next we consider the general Asplund space setting of assertion (ii), where the continuity assumption on f and the closedness assumptions on Ω and Θ directly imply the local closedness of $\mathcal{E}(f, \Omega, \Theta)$ around $(\bar{x}, 0)$. Thus we may employ the product space version of the exact extremal principle from Lemma 5.58 provided that there are index sets $J_1, J_2 \subset \{1, 2\}$ with $J_1 \cup J_2 = \{1, 2\}$ such that Ω_1 is PSNC at $(\bar{x}, 0)$ with respect to J_1 while Ω_2 is strongly PSNC at $(\bar{x}, 0)$ with respect to J_2.

Let us take $J_1 = \{2\}$ and $J_2 = \{1\}$, i.e., $X_1 = Z$ and $X_2 = X$ in the framework of Lemma 5.58. It is easy to see that Ω_2 is *strongly* PSNC at $(\bar{x}, 0)$ with respect to X, since U is a neighborhood of $\bar{x}$; note that Ω_2 is never SNC at $(\bar{x}, 0)$ unless Z is finite-dimensional. It remains to show that the set $\Omega_1 = \mathcal{E}(f, \Omega, \Theta)$ is PSNC at $(\bar{x}, 0)$ with respect to Z.

Since the latter set is represented as the inverse image

$$\mathcal{E}(f, \Omega, \Theta) = g^{-1}(\Theta) \ \text{ with } \ g(x, z) := f_\Omega(x) - z\ ,$$

one may apply Theorem 3.84 to ensure the stronger SNC property of this set. However, in this way we arrive at excessive conditions for the required

PSNC property. Let us establish more subtle sufficient conditions for the latter property by taking into account the specific structure of the mapping g. We need to show that, given arbitrary sequences $(x_k, z_k) \to (\bar{x}, 0)$ with $x_k \in \Omega$ and $g(x_k, z_k) \in \Theta$ and also $(x_k^*, z_k^*) \in \widehat{N}\big((x_k, z_k); \mathcal{E}(f, \Omega, \Theta)\big)$, one has

$$\Big[\|x_k^*\| \to 0, \quad z_k^* \xrightarrow{w^*} 0\Big] \Longrightarrow \|z_k^*\| \to 0 \,.$$

Consider locally closed sets $\Lambda_1, \Lambda_2 \subset X \times Z \times Z$ defined by $\Lambda_1 := \operatorname{gph} g$, $\Lambda_2 := X \times Z \times \Theta$ and observe that

$$(x_k^*, z_k^*, 0) \in \widehat{N}((x_k, z_k, v_k); \Lambda_1 \cap \Lambda_2) \;\text{ for all }\; k \in I\!N \,, \tag{5.80}$$

where $v_k := g(x_k, z_k)$. We are going to justify, using the full strength of Theorem 3.79 on the PSNC property of set intersections in the product of three spaces $X \times Z \times Z = X_1 \times X_2 \times X_3$, that the set $\Lambda_1 \cap \Lambda_2$ is PSNC at $(\bar{x}, 0, 0)$ with respect to $Z = X_2$.

First let us consider case (a) in (ii) when Θ is assumed to be SNC at 0. In this case we take $J_1 = \{2\}$ and $J_2 = \{1, 2, 3\}$ in the notation of Theorem 3.79 and observe that Λ_2 is SNC at $(\bar{x}, 0, 0)$, i.e., its strong PSNC property at $(\bar{x}, 0, 0)$ with respect to $J_2 \setminus J_1 = \{1, 3\}$ is automatic. Let us check that Λ_1 is PSNC at $(\bar{x}, 0, 0)$ with respect to $J_1 = \{2\}$. The latter means that for any sequences $(x_k, z_k, v_k) \to (\bar{x}, 0, 0)$ and $(x_k^*, z_k^*, v_k^*) \in \widehat{N}((x_k, z_k, v_k); \operatorname{gph} g)$ as $k \to \infty$ one has

$$\Big[\|(x_k^*, v_k^*)\| \to 0, \quad z_k^* \xrightarrow{w^*} 0\Big] \Longrightarrow \|z_k^*\| \to 0 \,.$$

Indeed, since the relation $(x_k^*, z_k^*, v_k^*) \in \widehat{N}((x_k, z_k, v_k); \operatorname{gph} g)$ can be rewritten as $(x_k^*, z_k^*) \in \widehat{D}^* g(x_k, z_k)(-v_k^*)$ and hence it gives

$$x_k^* \in \widehat{D}^* f_\Omega(x_k)(-v_k^*) \quad \text{and} \quad z_k^* = v_k^* \tag{5.81}$$

by the structure of g and Theorem 1.62(i), we derive from (5.81) that $\|z_k^*\| = \|v_k^*\| \to 0$ as $k \to \infty$, which justifies the PSNC property of Λ_1 at $(\bar{x}, 0, 0)$ with respect to $J_1 = \{2\}$.

To employ Theorem 3.79 in case (a), it remains to verify the *mixed* qualification condition of Definition 3.78 for the set system $\{\Lambda_1, \Lambda_2\} \subset X_1 \times X_2 \times X_3$ at $(\bar{x}, 0, 0)$ with respect to $(J_1 \setminus J_2) \cup (J_2 \setminus J_1) = \{1, 3\}$. Taking into account the structures of Λ_1 and Λ_2, the latter condition reduces to the following: for any sequences $(x_k^*, z_k^*, v_k^*) \in \widehat{N}((x_k, z_k, v_k); \operatorname{gph} g)$ and $u_k^* \in \widehat{N}(u_k; \Theta)$ satisfying

$$(x_k, z_k, v_k, u_k) \to (\bar{x}, 0, 0, 0), \quad \|x_k^*\| \to 0, \quad z_k^* \xrightarrow{w^*} 0, \quad (u_k^*, v_k^*) \xrightarrow{w^*} (u^*, -u^*)$$

as $k \to \infty$ with some $u^* \in N(0; \Theta)$ one has $u^* = 0$. It follows from the previous discussion that relations (5.81) hold for the above sequences whatever $k \in I\!N$. Thus the mentioned mixed qualification condition is equivalent to

$$N(0;\Theta)\cap\ker\widetilde{D}^*_M f_\Omega(\bar{x})=\{0\} \tag{5.82}$$

for the sets Λ_1 and Λ_2 under consideration. If (5.82) doesn't hold, then we immediately arrive at the optimality condition (5.79), since the mixed coderivative is never larger than the normal one. Thus we may assume that (5.82) is fulfilled, and then Theorem 3.79 ensures the PSNC property of the intersection $\Lambda_1\cap\Lambda_2$ at $(\bar{x},0,0)$ with respect to Z. By the latter PSNC property one has $\|z_k^*\|\to 0$ from (5.80), and therefore the set $\mathcal{E}(f,\Omega,\Theta)$ is PSNC at $(\bar{x},0)$ with respect to Z. This allows us to apply Lemma 5.58 to the above set system $\{\Omega_1,\Omega_2\}$ and arrive at the optimality condition (5.77). In turn, (5.77) implies (5.79) by Lemma 5.23, which also ensures the other conclusions of the theorem in case (a).

To complete the proof of the theorem, it remains to consider case (b) in (ii). The only difference between the above proof in case (a) is that now we need to justify the PSNC property of the intersection $\Lambda_1\cap\Lambda_2$ at $(\bar{x},0,0)$ with respect to $Z=X_2$ in the product space $X\times Z\times Z=X_1\times X_2\times X_3$ under the PSNC assumption on f_Ω^{-1} at $(0,\bar{x})$. To proceed, we again use Theorem 3.79 with another arrangement of the index sets therein in the case under consideration. Namely, let us now take $J_1=\{2,3\}$ and $J_2=\{1,2\}$ in the notation of Theorem 3.79. Then $J_2\setminus J_1=\{1\}$, and the set Λ_2 is obviously *strongly* PSNC at $(\bar{x},0,0)$ with respect to J_2. Let us check that Λ_1 is PSNC at $(\bar{x},0,0)$ with respect to J_1 under the assumption in (b). Indeed, the required PSNC property means that for any sequences $(x_k,z_k,v_k)\to(\bar{x},0,0)$ and $(x_k^*,z_k^*,v_k^*)\in\widehat{N}((x_k,z_k,v_k);\operatorname{gph} g)$ one has

$$\Big[\|x_k^*\|\to 0,\quad (z_k^*,v_k^*)\xrightarrow{w^*}(0,0)\Big]\Longrightarrow\|(z_k^*,v_k^*)\|\to 0\ \text{ as }\ k\to\infty\,.$$

By the above arguments in case (a) the latter is equivalent to say that for any sequences (x_k,x_k^*,z_k^*) satisfying

$$x_k^*\in\widehat{D}^*f_\Omega(x_k)(z_k^*)\ \text{ and }\ x_k\to\bar{x},\ \|x_k^*\|\to 0,\ z_k^*\xrightarrow{w^*}0$$

one has $\|z_k^*\|\to 0$ as $k\to\infty$, which is obviously equivalent to the PSNC property of f_Ω^{-1} at $(0,\bar{x})$ assumed in (b).

Finally, the application of Theorem 3.79 in case (b) requires the fulfillment of the mixed qualification condition from Definition 3.78 for $\{\Lambda_1,\Lambda_2\}$ at $(\bar{x},0,0)$ with respect to $(J_1\setminus J_2)\cup(J_2\setminus J_1)=\{1,3\}$, which happens to be the same as in case (a). This completes the proof of the theorem. △

Taking into account that $f_\Omega(x)=f(x)+\Delta(x;\Omega)$ with the indicator mapping $\Delta(\cdot;\Omega)$ of the set Ω and employing the coderivative/subdifferential and PSNC sum rules developed in Chap. 3, one may easily derive from Theorem 5.59 the corresponding (generally more restrictive) conditions expressed in terms of f and Ω separately. It can be also done in the framework of (5.77) by using intersection rules for the set

$$\mathcal{E}(f,\Omega,\Theta)=\big\{(x,z)\big|\ f(x)-z\in\Theta\big\}\cap\big(\Omega\times Z\big)\ .$$

This allows us (again based on comprehensive calculus rules) to deal efficiently with constrained multiobjective problems related to generalized order optimality, where constraints are given in various forms similar to those studied in Sect. 5.1 for single-objective minimization. In this way we obtain necessary optimality conditions for multiobjective problems with geometric and operator constraints described as $x\in G^{-1}(\Lambda)\cap\Omega$, which particularly include functional constraints of equality and inequality types. Let us present a corollary of Theorem 5.59 in case (a) therein considering for simplicity only multiobjective problems with operator (no geometric) constraints given by inverse images of sets under set-valued mappings between Asplund spaces.

Corollary 5.60 (multiobjective problems with operator constraints). *Let $f\colon X\to Z$ and $G\colon X\rightrightarrows Y$ be mappings between Asplund spaces, and let $\Theta\subset Z$ and $\Lambda\subset Y$ be nonempty subsets. Suppose that $\bar{x}$ is (f,Θ)-optimal subject to $x\in G^{-1}(\Lambda)$, where f is continuous, G is closed-graph, and Θ and Λ are closed around the corresponding points. Suppose also that Θ is SNC at* 0 *and that the set-valued mapping $S(\cdot):=G(\cdot)\cap\Lambda$ is inner semicompact around $\bar{x}$. Then there is $z^*\in N(0;\Theta)\setminus\{0\}$ such that*

$$\bigcup\Big[D_N^*G(\bar{x},\bar{y})(y^*)\Big|\ \bar{y}\in S(\bar{x}),\ \ y^*\in N(\bar{y};\Lambda)\Big]\bigcap\Big(-D_N^*f(\bar{x})(z^*)\Big)\neq\emptyset$$

under one of the following requirements on (f,G,Λ):

(a) *f is PSNC at $\bar{x}$, the qualification conditions*

$$N(\bar{y};\Lambda)\cap\ker\widetilde{D}_M^*G(\bar{x},\bar{y})=\{0\}\ \ \textit{for all}\ \ \bar{y}\in S(\bar{x})\ ,$$

$$\bigcup\Big[D_N^*G(\bar{x},\bar{y})(y^*)\Big|\ \bar{y}\in S(\bar{x}),\ \ y^*\in N(\bar{y};\Lambda)\Big]\bigcap\Big(-D_M^*f(\bar{x})(0)\Big)=\{0\}$$

hold, and either G^{-1} is PSNC at $(\bar{y},\bar{x})$ or Λ is SNC at $\bar{y}$ for all $\bar{y}\in S(\bar{x})$.

(b) *The second qualification condition in* (a) *holds together with*

$$N(\bar{y};\Lambda)\cap\ker D_N^*G(\bar{x},\bar{y})=\{0\}\ \ \textit{for all}\ \ \bar{y}\in S(\bar{x})\ ,$$

and either G is PSNC at $(\bar{x},\bar{y})$ and Λ is SNC at $\bar{y}$, or G is SNC at $(\bar{x},\bar{y})$ for all $\bar{y}\in S(\bar{x})$.

Proof. Using assertion (ii) in Theorem 5.59 in case (a) therein, we find $z^*\in N(0;\Theta)$ satisfying

$$0\in D_N^*\big[f+\Delta(\cdot;\Omega)\big](\bar{x})(z^*)$$

with $\Omega:=G^{-1}(\Lambda)$. The latter implies, by the coderivative sum rule of Theorem 3.10, that

$$N(\bar{x};G^{-1}(\Lambda))\cap\big(-D_N^*f(\bar{x})(z^*)\big)\neq\emptyset$$

provided that

$$N(\bar{x}; G^{-1}(\Lambda)) \cap \big(- D^*_M f(\bar{x})(z^*) \big) = \{0\} \ ,$$

and that either f is PSNC at $\bar{x}$ or $G^{-1}(\Lambda)$ is SNC at this point. If f is supposed to be PSNC at $\bar{x}$, then the corollary follows from Theorem 3.8 giving the representation of basic normals to the inverse image $G^{-1}(\Lambda)$ under the assumptions made in (a). Otherwise, one needs to employ Theorem 3.84 ensuring the SNC property of $G^{-1}(\Lambda)$ at $\bar{x}$, which gives the conclusions of the corollary under the assumptions made in (b). △

Note that the PSNC conditions on f and G^{-1} and both qualification conditions in (a) of Corollary 5.60 *automatically hold if f is Lipschitz continuous around $\bar{x}$ and G is metrically regular around* $(\bar{x}, \bar{y})$ *for all* $\bar{y} \in S(\bar{x})$. Note also the results of Corollary 5.60 provide *qualified* necessary optimality conditions for multiobjective problems under the most suitable constraint qualifications. They easily imply necessary conditions in a non-qualified form admitting the violations of constraint qualifications; cf. Subsect. 5.1.2 for problems of minimizing extended-real-valued functions (i.e., with a single objective).

Similarly to Subsect. 5.1.3 one may derive from Theorem 5.59 and Corollary 5.60 the corresponding necessary optimality conditions for *multiobjective* problems with *functional constraints* given by equalities and inequalities. However, some results of Subsect. 5.1.3 (and the preceding material of Sect. 5.1) essentially exploit specific features of single-objective minimization problems. It particularly concerns upper subdifferential conditions for minimizing extended-real-valued functions. Nevertheless, for multiobjective problems we can obtain necessary optimality conditions of the *upper subdifferential* type involving *inequality* constraints (but not objective mappings) under additional assumptions on the Asplund space X in question. For simplicity we present such necessary optimality conditions with no constraint qualifications for multiobjective problems having only inequality constraints, since only these constraints are of special interest from the viewpoint of upper subdifferential optimality conditions.

Theorem 5.61 (upper subdifferential optimality conditions for multiobjective problems). *Given $f: X \to Z$ and $\Theta \subset Z$ closed around* 0, *suppose that $\bar{x}$ is (f, Θ)-optimal subject to the inequality constraints*

$$\varphi_i(x) \le 0, \quad i = 1, \ldots, m \ ,$$

with $\varphi_i: X \to \overline{I\!R}$ finite at $\bar{x}$ for all $i = 1, \ldots, m$. Assume that Z is Asplund while X admits a Lipschitzian $\mathcal{C}^1$-smooth bump function (which is automatic when X admits a Fréchet differentiable renorm), that f is Lipschitz continuous around $\bar{x}$, and that Θ is SNC at the origin. Then for any Fréchet upper subgradients $x^*_i \in \widehat{\partial}^+ \varphi_i(\bar{x})$, $i = 1, \ldots, m$, there are $z^* \in N(0; \Theta)$ and $(\lambda_1, \ldots, \lambda_m) \in I\!R^m$ satisfying

$$\lambda_i \geq 0, \quad \lambda_i \varphi(\bar{x}) = 0 \;\; \textit{for all} \;\; i = 1, \ldots, m$$

such that $(z^*, \lambda_1, \ldots, \lambda_m) \neq 0$ *and one has*

$$0 \in D_N^* f(\bar{x})(z^*) + \sum_{i=1}^{m} \lambda_i x_i^* \, .$$

If in addition f *is strictly Lipschitzian at* $\bar{x}$, *then*

$$0 \in \partial\langle z^*, f\rangle(\bar{x}) + \sum_{i=1}^{m} \lambda_i x_i^*, \quad (z^*, \lambda_1, \ldots, \lambda_m) \neq 0 \, .$$

Proof. Given $x_i^* \in \widehat{\partial}\varphi_i(\bar{x})$ for $i = 1, \ldots, m$ and using the variational description of Fréchet subgradients $-x_i^* \in \widehat{\partial}(-\varphi_i)(\bar{x})$ from Theorem 1.88(ii) with $\mathcal{S} = \mathcal{LC}^1$ (i.e., in spaces admitting Lipschitzian $\mathcal{C}^1$ bump functions), we find $s_i \colon X \to I\!R$ for $i = 1, \ldots, m$ and a neighborhood U of $\bar{x}$ such that

$$s_i(\bar{x}) = \varphi_i(\bar{x}), \quad s_i(x) \geq \varphi_i(x) \;\; \text{for all} \;\; x \in U, \quad i = 1, \ldots, m \, ,$$

and each s_i is continuously differentiable on U with $\nabla s_i(\bar{x}) = x_i^*$, $i = 1, \ldots, m$. It follows from the construction of s_i that $\bar{x}$ is an (f, Θ)-optimal solution not only subject to the original constraints $\varphi_i(x) \leq 0$ but also subject to the *smooth* inequality constraints

$$s_i(x) \leq 0, \quad i = 1, \ldots, m \, .$$

Let us apply to the latter problem the necessary optimality conditions from Corollary 5.60 with $G(x) := \big(s_1(x), \ldots, s_m(x)\big)$ and $\Lambda := I\!R^m_-$. In this case

$$D^*G(\bar{x})(\lambda_1, \ldots, \lambda_m) = \sum_{i=1}^{m} \lambda_i \nabla s_i(x) = \sum_{i=1}^{m} \lambda_i x_i^* \quad \text{and}$$

$$N(\bar{y}; \Lambda) = \big\{(\lambda_1, \ldots, \lambda_m) \in I\!R^m \big|\; \lambda_i \geq 0, \;\; \lambda_i \varphi_i(\bar{x}) = 0, \;\; i = 1, \ldots, m\big\} \, ,$$

where $\bar{y} := \big(\varphi_1(\bar{x}), \ldots, \varphi_m(\bar{x})\big) = \big(s_1(\bar{x}), \ldots, s_m(\bar{x})\big)$. It is easy to see that all the assumptions of Corollary 5.60(b) are fulfilled for the problem under consideration, except the qualification condition involving the kernel of $D^*G(\bar{x})$. If the latter is satisfied, we get the conclusion of the theorem with $z^* \neq 0$ by Corollary 5.60. Otherwise, we obviously have the conclusion of the theorem with $(\lambda_1, \ldots, \lambda_m) \neq 0$, which completes the proof. △

By specifying the ordering set Θ one may derive from the general results obtained above necessary optimality conditions for particular multiobjective problems. Observe that if Θ is a *convex cone*, which is the case in many typical applications, then the optimality condition $z^* \in N(0; \Theta)$ in Theorems 5.59, 5.61 and Corollary 5.60 reads as

$$\langle z^*, z\rangle \le 0 \ \text{ for all } \ z \in \Theta \ .$$

Let us present an application of the above results to the constrained *minimax* problem:

$$\text{minimize } \ \varphi(x) := \max\big\{\langle z^*, f(x)\rangle \big|\, z^* \in \Lambda\big\} \ \text{ subject to } \ x \in \Omega \ , \tag{5.83}$$

where $f\colon X \to Z$, $\Omega \subset X$, and $\Lambda \subset Z^*$. As shown in Example 5.54, the minimax objective in (5.83) can be reduced to the *generalized order optimality* considered in Theorem 5.59. The next result gives a concretization and refinement of the latter theorem applied to (5.83) with a new *complementary slackness condition* specific for the minimax problems under consideration. For simplicity we formulate a minimax counterpart of Theorem 5.59 only in the case of the coderivative condition (5.79) therein.

Theorem 5.62 (optimality conditions for minimax problems). *Let $\bar x$ be a local optimal solution to the constrained minimax problem* (5.83) *with $\bar z := f(\bar x)$, where $f\colon X \to Z$ is a mapping between Asplund spaces that is continuous around $\bar x$ relative to Ω. Suppose that Ω is locally closed around $\bar x$, that $\Lambda \subset Z^*$ is weak* sequentially compact, and that there is $z_0 \in Z$ for which $\langle z^*, z_0\rangle = 1$ whenever $z^* \in \Lambda$. Then there is $\bar z^* \in Z^*$ satisfying the inclusions*

$$0 \in D_N^* f_\Omega(\bar x)(\bar z^*) \ \ \textit{with} \ \ \bar z^* \ne 0 \,, \tag{5.84}$$

$$\bar z^* \in \mathrm{cl}^*\mathrm{co}\,\big[\mathrm{cone}\,\Lambda\big] = \mathrm{cl}^*\Big\{\sum_{i=1}^m \alpha_i z_i^* \Big|\, \alpha_i \ge 0,\ z_i^* \in \Lambda,\ m \in I\!N\Big\} \tag{5.85}$$

and the complementary slackness condition

$$\big\langle \bar z^*, \bar z - \varphi(\bar x) z_0\big\rangle = 0 \tag{5.86}$$

provided that:

(a) *either the set*

$$\Theta := \big\{z \in Z \big|\, \langle z^*, z\rangle \le 0 \ \ \textit{for all} \ \ z^* \in \Lambda\big\}$$

is SNC at $f(\bar x) - \varphi(\bar x) z_0$,

(b) *or the inverse mapping f_Ω^{-1} is PSNC at $(\bar z, \bar x)$.*

Proof. First observe that the maximum is attached in (5.83) under the assumptions imposed on Λ and f. Build the set

$$\overline{\Theta} := \Theta + \big(\varphi(\bar x) z_0 - \bar z\big)$$

upon the given data in the theorem and show that $\bar x$ is locally $(f, \overline{\Theta})$-optimal relative to Ω. Since $\bar z - \varphi(\bar x) z_0 \in \Theta$, one has $0 \in \overline{\Theta}$. We need to check that condition (5.76) holds with some $z_k \to 0$. Assuming the contrary, take

$z_k := z_0/k$ for $k \in I\!N$ and find $x \in U$ from a neighborhood of $\bar{x}$ such that $x \in \Omega$ and one has

$$\langle z^*, f(x)\rangle - \varphi(\bar{x}) = \langle z^*, f(x)\rangle - \varphi(\bar{x})\langle z^*, z_0\rangle$$
$$\le -\langle z^*, z_0\rangle/k < 0 \ \text{ for all } \ z^* \in \Lambda$$

as $k \to \infty$, which contradicts the local optimality of $\bar{x}$ in the minimax problem (5.84). Applying Theorem 5.59(ii) in this setting and taking into account the convexity of $\overline{\Theta}$, we find $\bar{z}^*$ satisfying (5.84) and

$$\big\langle \bar{z}^*, z - \big(\bar{z} - \varphi(\bar{x})z_0\big)\big\rangle \le 0 \ \text{ for all } \ z \in \Theta\ . \tag{5.87}$$

It remains to show that (5.87) implies (it is actually equivalent to) both conditions (5.85) and (5.86). Indeed, we have from (5.87) and the conic structure of Θ the inequality

$$\big\langle \bar{z}^*, \alpha z - \big(\bar{z} - \varphi(\bar{x})z_0\big)\big\rangle \le 0 \ \text{ for all } \ \alpha > 0 \ \text{ and } \ z \in \Theta\ .$$

It gives, by passing to the limit as $\alpha \to \infty$, that $\langle \bar{z}^*, z\rangle \le 0$ whenever $z \in \Theta$, and hence (5.85) holds. Moreover, one has the inequality

$$\big\langle \bar{z}^*, \bar{z} - \varphi(\bar{x})z_0\big\rangle \le 0\ ,$$

since $\bar{z} - \varphi(\bar{x})z_0 \in \Theta$. The opposite inequality follows from (5.87) with $z = 0$. Thus we get (5.86) and complete the proof of the theorem. △

Observe that all the assumptions of Theorem 5.62 involving the set Λ automatically hold if this set consists of finitely many linearly independent elements. In the latter case the general minimax problem (5.83) actually reduces to minimizing the maximum of a finite number of real-valued functions:

$$\text{minimize } \ \varphi(x) = \max\big\{\varphi_i(x)\big|\ i = 1,\ldots,n\big\} \ \text{ subject to } \ x \in \Omega\ . \tag{5.88}$$

Let us present an easy consequence of Theorem 5.62 for problem (5.88) assuming for simplicity that all φ_i are locally Lipschitzian.

Corollary 5.63 (minimax over finite number of functions). *Let $\bar{x}$ be a local optimal solution to problem* (5.88) *in an Asplund space X, where all $\varphi_i\colon X \to I\!R$ are Lipschitz continuous around $\bar{x}$ and where Ω is locally closed around this point. Then there are numbers $\lambda_i \ge 0$, $i = 1,\ldots,n$, such that*

$$\lambda_i\big(\varphi_i(\bar{x}) - \varphi(\bar{x})\big) = 0 \ \ \textit{for} \ \ i = 1,\ldots,n \quad \textit{and}$$

$$0 \in \partial\Big(\sum_{i=1}^n \lambda_i\varphi_i + \delta(\cdot;\Omega)\Big)(\bar{x}) \subset \partial\Big(\sum_{i=1}^n \lambda_i\varphi_i\Big)(\bar{x}) + N(\bar{x};\Omega)\ .$$

Proof. It follows directly from Theorem 5.62 with $Z = I\!R^n$ and Λ consisting of the basic unit vectors. Note that these necessary optimality conditions with the last inclusion in the theorem can be also derived from the ones in scalar optimization employing the calculus rule of Theorem 3.46(ii) for subdifferentiation of maximum functions. △

Similarly to the above multiobjective results one can get, based on Theorem 5.62 and Corollary 5.63, necessary optimality conditions for minimax problems with operator and other constraints as well as upper subdifferential conditions in the case of constraints given by inequalities.

5.3.3 Extremal Principle for Set-Valued Mappings

Our next goal is to derive necessary optimality conditions for constrained multiobjective problems, where the notion of optimality is described by a closed preference relation in the sense of Definition 5.55. As observed in Subsect. 5.3.1, this notion may be different from the generalized order optimality studied in the preceding subsection. From the viewpoint of variational geometry, general closed preferences lead to considering extremal systems of *set-valued mappings/multifunctions* (but not just systems of sets) related to nonlinear *deformations* vs. (linear) translations. In this subsection we study such extremal systems of multifunctions and derive appropriate versions of the *extremal principle* for them in both approximate and exact forms. The latter form of the extremal principle for set-valued mappings requires certain extensions of limiting normals and the SNC property in the case of *moving sets* that take into account the dependence of sets on deformation parameters. We begin with the definition of local extremality for systems of multifunctions.

Definition 5.64 (extremal systems of multifunctions). *Let $S_i\colon M_i \rightrightarrows X$, $i = 1,\ldots,n$, be set-valued mappings from metric spaces (M_i, d_i) into a Banach space X. We say that $\bar{x}$ is a* LOCAL EXTREMAL POINT *of the system $\{S_1,\ldots,S_n\}$ at $(\bar{s}_1,\ldots,\bar{s}_n)$ provided that $\bar{x} \in S_1(\bar{s}_1) \cap \ldots \cap S_n(\bar{s}_n)$ and there exists a neighborhood U of $\bar{x}$ such that for every $\varepsilon > 0$ there are $s_i \in \operatorname{dom} S_i$ satisfying the conditions*

$$d(s_i, \bar{s}_i) \le \varepsilon, \quad \operatorname{dist}\big(\bar{x}; S_i(s_i)\big) \le \varepsilon \;\; \textit{as} \;\; i = 1,\ldots,n, \quad \textit{and}$$

$$S_1(s_1) \cap \ldots \cap S_n(s_n) \cap U = \emptyset \,. \tag{5.89}$$

In this case $\{S_1,\ldots,S_n,\bar{x}\}$ is called the EXTREMAL SYSTEM *at $(\bar{s}_1,\ldots,\bar{s}_n)$.*

It is easy to see that the above definition extends to set-valued mappings the notion of set extremality from Definition 2.1. Indeed, considering for simplicity an extremal system of two sets $\{\Omega_1, \Omega_2, \bar{x}\}$, we reduce it to the above notion for set-valued mappings by letting

$$M_1 := X, \quad M_2 := \{0\}, \quad S_1(s_1) := \Omega_1 + s_1, \quad S_2(0) := \Omega_2 \ ,$$

which corresponds to the *linearity* of set-valued mappings (or to the *translation* of sets) in Definition 5.64. The next example shows that the extremal systems involving deformations of sets *cannot* be reduced to those obtained by their translations. Indeed, consider the moving sets (i.e., set-valued mappings) defined by

$$\begin{aligned} S_1(s_1) &:= \big\{(x,y) \in I\!R^2 \big|\ |x| - 2|y| \ge s_1\big\} \ , \\ S_2(s_2) &:= \big\{(x,y) \in I\!R^2 \big|\ |y| - 2|x| \ge s_2\big\} \ , \end{aligned} \tag{5.90}$$

which can be viewed as deformations of the initial sets $\Omega_1 := S_1(0)$ and $\Omega_2 := S_2(0)$. One can check that $(0,0)$ is a local extremal point of $\{S_1, S_2\}$ in the sense of Definition 5.64, while it is not the case with respect to Definition 2.1.

Our major example of extremal systems involving set deformations relates to problems of multiobjective optimization with respect to closed preference relations described in Definition 5.55.

Example 5.65 (extremal points in multiobjective optimization with closed preferences). *Let $f\colon X \to Z$ be a mapping between Banach spaces, let $\prec$ be a closed preference relation on Z with the level set $\mathcal{L}(z)$, and let $\bar{x}$ be an optimal solution to the constrained multiobjective problem:*

$$\textit{minimize} \quad f(x) \quad \textit{subject to} \quad x \in \Omega \ ,$$

where "minimization" is understood with respect to the preference $\prec$. Then $(\bar{x}, f(\bar{x}))$ is a local extremal point at $(f(\bar{x}), 0)$ for the system of multifunctions $S_i\colon M_i \rightrightarrows X \times Z$, $i = 1, 2$, defined by

$$S_1(s_1) := \Omega \times \operatorname{cl} \mathcal{L}(s_1) \quad \textit{with} \quad M_1 := \mathcal{L}(f(\bar{x})) \cup \{f(\bar{x})\} \ ,$$

$$S_2(s_2) = S_2 := \big\{(x, f(x)) \big|\ x \in X\big\} \quad \textit{with} \quad M_2 := \{0\} \ .$$

Proof. First we observe that $(\bar{x}, f(\bar{x})) \in S_1(f(\bar{x})) \cap S_2$ due to the local satiation property of $\prec$. To establish (5.89), we assume the contrary and find, given any neighborhood U of $(\bar{x}, f(\bar{x}))$, a point $s_1 \in \mathcal{L}(f(\bar{x}))$ close to $f(\bar{x})$ but not equal to the latter by the preference nonreflexivity, for which

$$S_1(s_1) \cap S_2 \cap U \neq \emptyset \ .$$

This yields the existence of x near $\bar{x}$ with $(x, f(x)) \in S_1(s_1) = \Omega \times \operatorname{cl} \mathcal{L}(s_1)$. Hence $x \in \Omega$ and $f(x) \prec f(\bar{x})$ by the almost transitivity property of $\prec$. This contradicts the local optimality of $\bar{x}$ in the constrained multiobjective problem under consideration. △

Before establishing the extremal principle for set-valued mappings and its applications to multiobjective optimization, let us present two other examples of extremal systems that are certainly of independent interest.

Example 5.66 (extremal points in two-player games). *Let $(\bar{x},\bar{y}) \in \Omega\times\Theta$ be a* SADDLE POINT *of a payoff function $\varphi\colon X\times Y\to I\!R$ over subsets $\Omega\subset X$ and $\Theta\subset Y$ of Banach spaces, i.e.,*

$$\varphi(x,\bar{y}) \le \varphi(\bar{x},\bar{y}) \le \varphi(\bar{x},y) \quad \textit{whenever} \quad (x,y)\in\Omega\times\Theta\ .$$

Define a set-valued mapping $S_1\colon[\varphi(\bar{x},\bar{y}),\infty)\times(-\infty,\varphi(\bar{x},\bar{y})]\rightrightarrows\Omega\times I\!R\times\Theta\times I\!R$ and a set $S_2\subset\Omega\times I\!R\times\Theta\times I\!R$ by

$$S_1(\alpha,\beta):=\Omega\times[\alpha,\infty)\times\Theta\times(-\infty,\beta],\quad S_2:=\text{hypo}\,\varphi(\cdot,\bar{y})\times\text{epi}\,\varphi(\bar{x},\cdot)\ .$$

Then the point $\big(\bar{x},\varphi(\bar{x},\bar{y}),\bar{y},\varphi(\bar{x},\bar{y})\big)$ is locally extremal for the system $\{S_1,S_2\}$ at $\big(\varphi(\bar{x},\bar{y}),\varphi(\bar{x},\bar{y})\big)$.

Proof. One obviously has

$$\big(\bar{x},\varphi(\bar{x},\bar{y}),\bar{y},\varphi(\bar{x},\bar{y})\big)\in S_1\big(\varphi(\bar{x},\bar{y}),\varphi(\bar{x},\bar{y})\big)\cap S_2\ .$$

Furthermore, it follows from the definition of the saddle point $(\bar{x},\bar{y})$ that

$$S_1(\alpha,\beta)\cap S_2=\emptyset \quad \text{whenever} \quad (\alpha,\beta)\in[\varphi(\bar{x},\bar{y}),\infty)\times(-\infty,\varphi(\bar{x},\bar{y})]$$

and $(\alpha,\beta)\ne\big(\varphi(\bar{x},\bar{y}),\varphi(\bar{x},\bar{y})\big)$. Thus $\big(\bar{x},\varphi(\bar{x},\bar{y}),\bar{y},\varphi(\bar{x},\bar{y})\big)$ is a local extremal point of $\{S_1,S_2\}$ in the above sense. △

Example 5.67 (extremal points in time optimal control). *Let $\bar{\tau}$ be an optimal solution to the following optimal control problem: minimize the transient time $\tau>0$ subject to the endpoint constraint $x(\tau)=0$ over absolutely continuous trajectories $x\colon[0,\tau]\to I\!R^n$ of the ordinary differential equation*

$$\dot{x}(t)=f\big(x(t),u(t)\big),\quad x(0)=x_0,\quad u(t)\in U \quad \textit{a.e.}\ \ t\in[0,\tau] \tag{5.91}$$

corresponding to measurable controls $u(\cdot)$. Consider the REACHABLE SET *multifunction $S_1\colon(0,\infty)\rightrightarrows I\!R^n$ defined by*

$$S_1(s_1):=\Big\{x(s_1)\in I\!R^n\Big|\ x(\cdot)\ \ \textit{is feasible in (5.91) on } [0,s_1]\Big\}\ ,$$

and let $S_2=\{0\}\subset I\!R^n$. Then $0\in I\!R^n$ is a local extremal point of the system $\{S_1,S_2\}$ at $(\bar{\tau},0)$ in the sense of Definition 5.64 with $M_1=(0,\infty)$ and $M_2=\{0\}\subset I\!R$.

Proof. Follows directly from the definitions. △

Next we derive the *extremal principle for systems of multifunctions* in an *approximate form* similar to the one in Theorem 2.20. This result is actually *equivalent* to the approximate extremal principle for systems of sets in Theorem 2.20 and happens to be yet another *characterization* of Asplund spaces.

Theorem 5.68 (approximate extremal principle for multifunctions). *Let $S_i\colon M_i \rightrightarrows X$ be set-valued mappings from metric spaces (M_i, d_i) into a Banach space X for $i = 1, \ldots, n$. Then the following are equivalent:*

(a) *X is Asplund.*

(b) *For any extremal system $\{S_1, \ldots, S_n, \bar{x}\}$ at $(\bar{s}_1, \ldots, \bar{s}_n)$ the* APPROXIMATE EXTREMAL PRINCIPLE *holds provided that each S_i is closed-valued around $\bar{s}_i$. This means that for every $\varepsilon > 0$ there are $s_i \in \operatorname{dom} S_i$, $x_i \in S_i(s_i)$, and $x_i^* \in X^*$, $i = 1, \ldots, n$, satisfying*

$$d(s_i, \bar{s}_i) \le \varepsilon, \quad \|x_i - \bar{x}\| \le \varepsilon, \quad x_i^* \in \widehat{N}\big(x_i; S_i(s_i)\big) + \varepsilon I\!B^* , \tag{5.92}$$

$$x_1^* + \ldots + x_n^* = 0, \qquad \|x_1^*\| + \ldots + \|x_n^*\| = 1 . \tag{5.93}$$

(c) *For any extremal system $\{S_1, \ldots, S_n, \bar{x}\}$ at $(\bar{s}_1, \ldots, \bar{s}_n)$ the ε-normal counterpart of the approximate extremal principle holds with*

$$\widehat{N}\big(x_i; S_i(s_i)\big) + \varepsilon I\!B^* \quad \text{replaced by} \quad \widehat{N}_\varepsilon\big(x_i; S_i(s_i)\big)$$

in (5.92), $i = 1, \ldots, n$, provided that each S_i is closed-valued around $\bar{s}_i$.

Proof. First note that (b)⇒(c), since one always has

$$\widehat{N}(\bar{x}; \Omega) + \varepsilon I\!B^* \subset \widehat{N}_\varepsilon(\bar{x}; \Omega) .$$

Observe further that the ε-extremal principle for multifunctions in (c) implies the one for systems of sets from Definition 2.5(i). Thus implication (c)⇒(a) in the theorem follows from (c)⇒(a) in Theorem 2.20. It remains to prove that (a)⇒(b), i.e., that the approximate extremal principle holds for any extremal system of multifunctions in Asplund spaces. It can be done similarly to the procedure in Sect. 2.2 based on the direct variational arguments in Fréchet smooth spaces and the method of separable reduction. In what follows we give another proof that employs the Ekeland variational principle in Theorem 2.26(i) and the fuzzy subgradient condition for minimum points of semi-Lipschitzian sum in Lemma 2.32, which is equivalent to the approximate extremal principle for systems of sets.

Let $\bar{x}$ be a local extremal point of the system $\{S_1, \ldots, S_n\}$ at $(\bar{s}_1, \ldots, \bar{s}_n)$, where X is Asplund in Definition 5.64. Take $U := \bar{x} + rI\!B$ and, given $\varepsilon > 0$, choose $\varepsilon' > 0$ satisfying

$$\varepsilon' < \min\Big\{\varepsilon^2/(5\varepsilon + 12n^2 + \varepsilon^2), r^2/4\Big\} .$$

Then we take $s_1, \ldots, s_n$ from Definition 5.64 corresponding to ε'. Denote $\Omega := S_1(s_1) \times \ldots \times S_n(s_n)$ and form the function

$$\varphi(y_1, \ldots, y_n) := \sum_{i,j=1}^{n} \|y_i - y_j\| + \delta\big((y_1, \ldots, y_n); \Omega\big)$$

as $(y_1,\ldots,y_n)\in U^n$, which is l.s.c. and positive on the complete metric space U^n. Whenever $y_i'\in S_i(s_i)$ are chosen by

$$\|y_i'-y_j'\|\le \operatorname{dist}\big(\bar{x};S_i(s_i)\big)+\operatorname{dist}\big(\bar{x};S_j(s_j)\big)+\varepsilon'\le 3\varepsilon'$$

one has $\varphi(y_1',\ldots,y_n')\le 3n^2\varepsilon'<\varepsilon^2/4$. By the Ekeland variational principle from Theorem 2.26(i) applied to the above function φ we find $x_i'\in y_i'+(\varepsilon/2)I\!B\subset \bar{x}+\varepsilon I\!B$ for $i=1,\ldots,n$ such that the perturbed function

$$\sum_{i,j=1}^{n}\|y_i-y_j\|+\frac{\varepsilon}{2}\sum_{i=1}^{n}\|y_i-x_i'\|+\delta\big((y_1,\ldots,y_n);\Omega\big) \tag{5.94}$$

attains its global minimum at $(x_1',\ldots,x_n')$ on U^n. Assume that $U^n=X^n$ without loss of generality and denote

$$\psi(y_1,\ldots,y_n):=\sum_{i,j=1}^{n}\|y_i-y_j\|,\quad (y_1,\ldots,y_n)\in X^n\ ,$$

for which $\psi(x_1',\ldots,x_n')>0$ by the construction. Now applying Theorem 2.20 and Lemma 2.32(i) to (5.94) and taking into account that

$$\widehat{\partial}\delta\big((y_1,\ldots,y_n);\Omega\big)=\widehat{N}\big(y_1;S_1(y_1)\big)\times\ldots\times\widehat{N}\big(y_n;S_n(s_n)\big)\ \text{ for any }\ y_i\in S(s_i)\ ,$$

we find $x_i\in S_i(s_i)\cap(x_i'+\varepsilon' I\!B)\subset(\bar{x}+\varepsilon I\!B)$, $z_i\in x_i'+\varepsilon' I\!B$ for $i=1,\ldots,n$, and $(-x_1^*,\ldots,-x_n^*)\in\widehat{\partial}\psi(z_1,\ldots,z_n)$ such that

$$0\in(-x_1^*,\ldots,-x_n^*)+\widehat{N}\big(x_1;S_1(s_1)\big)\times\ldots\times\widehat{N}\big(x_n;S_n(s_n)\big)+\varepsilon'(n+1)(I\!B^*)^n\ .$$

The latter relation clearly implies that

$$x_i^*\in\widehat{N}\big(x_i;S_i(s_i)\big)+\varepsilon I\!B^*\ \text{ whenever }\ i=1,\ldots,n$$

for the chosen number ε, which gives (5.92).

Let us show that $x_1^*,\ldots,x_n^*$ satisfy (5.93) as well. Shrinking ε' further if necessary, we can make $\psi(z_1,\ldots,z_n)>0$. Observe that the inclusion $(-x_1^*,\ldots,-x_n^*)\in\widehat{\partial}\psi(z_1,\ldots,z_n)$ yields

$$\begin{aligned}&\langle -x_1^*-\ldots-x_n^*,h\rangle\\ &\le\liminf_{t\to 0}\frac{\psi(z_1+th,\ldots,z_n+th)-\psi(z_1,\ldots,z_n)}{t}\\ &=\liminf_{t\to 0}\frac{\sum_{i,j=1}^{n}\|(z_i+th)-(z_j+th)\|-\sum_{i,j=1}^{n}\|z_i-z_j\|}{t}=0\end{aligned}$$

for any unit vector $h\in X$. This gives the first relation (Euler equation) in (5.93). It remains to show that

$$\|x_1^*\| + \ldots + \|x_n^*\| \geq 1 ,$$

which implies the second relations in (5.93) by normalization. To proceed, we observe that the function ψ is positively homogeneous, and hence the inclusion $(-x_1^*, \ldots, -x_n^*) \in \widehat{\partial}\psi(z_1, \ldots, z_n)$ implies

$$\sum_{i=1}^n \langle -x_i^*, -z_i \rangle \leq \liminf_{t\to 0} \frac{\psi(z_1 - tz_1, \ldots, z_n - tz_n) - \psi(z_1, \ldots, z_n)}{t}$$

$$= -\psi(z_1, \ldots, z_n) .$$

Using $-x_1^* = x_2^* + \ldots + x_n^*$ from the Euler equation, one has

$$\psi(z_1, \ldots, z_n) \leq \sum_{i=1}^n \langle -x_i^*, z_i \rangle = \sum_{i=2}^n \langle x_i^*, z_1 - z_i \rangle$$

$$\leq \max\Big\{\|x_i^*\| \ \Big|\ i = 2, \ldots, n\Big\} \sum_{i=2}^n \|z_1 - z_i\|$$

$$\leq \max\Big\{\|x_i^*\| \ \Big|\ i = 1, \ldots, n\Big\} \psi(z_1, \ldots, z_n) .$$

Since $\psi(z_1, \ldots, z_n) > 0$, the latter gives the estimate

$$\max\Big\{\|x_i^*\| \ \Big|\ i = 1, \ldots, n\Big\} \geq 1 ,$$

which completes the proof of the theorem. △

Our next intention is to obtain the extremal principle for multifunctions in the *exact/limiting form* similar to the one in Theorem 2.22 for systems of sets. It is natural to derive such a result by passing to the limit as $\varepsilon \downarrow 0$ in relations (5.92) and (5.93) of the approximate extremal principle. However, the situation here is somewhat different from the case of the extremal principle for sets, since now the sets $S_i(s_i)$ in (5.92) are *moving*, i.e., they depend on certain points that converge to $\bar{s}_i$ as $\varepsilon \downarrow 0$. To perform the limiting procedure and to obtain the extremal principle in a suitable limiting form, we need to describe limiting normals to moving sets and also to impose appropriate normal compactness requirements that allow us to pass to the limit in infinite-dimensional settings. Let us first define the cone of limiting normals to moving sets that is useful in both finite and infinite dimensions.

Definition 5.69 (limiting normals to moving sets). *Let $S\colon Z \rightrightarrows X$ be a set-valued mapping from a metric space Z into a Banach space X, and let $(\bar{z}, \bar{x}) \in \operatorname{gph} S$. Then*

$$N_+\big(\bar{x}; S(\bar{z})\big) := \mathop{\mathrm{Lim\,sup}}_{\substack{(z,x)\overset{\operatorname{gph} S}{\to}(\bar{z},\bar{x})\\ \varepsilon\downarrow 0}} \widehat{N}_\varepsilon\big(x; S(z)\big) \tag{5.95}$$

is the EXTENDED NORMAL CONE *to* $S(\bar{z})$ *at* $\bar{x}$. *The mapping* S *is* NORMALLY SEMICONTINUOUS *at* $(\bar{z},\bar{x})$ *if*

$$N_+\big(\bar{x};S(\bar{z})\big)=N\big(\bar{x};S(\bar{z})\big)\ . \tag{5.96}$$

Observe that one can equivalently put $\varepsilon=0$ in (5.95) if X is *Asplund* and S is *closed-valued* around $\bar{x}$. This follows directly from formula (2.51) giving a representation of ε-normals in Asplund spaces. Note also that the normality notion in (5.95) has nothing to do with a (generalized) differentiability of the set-valued mapping $S(\cdot)$: the variable z there is just a *parameter* of moving sets, which is involved in the limiting process.

One always has the inclusion "$\supset$" in (5.96), i.e., more limiting normals may obviously appear during the process in (5.95) involving the moving sets $S(\cdot)$ than during the one in (1.2) that takes only the set $S(\bar{x})$ into account. However, $N_+\big(\bar{x};S(\bar{z})\big)$ agrees with the basic normal cone $N\big(\bar{x};S(\bar{z})\big)$ when the sets $S(z)$ behave reasonably well as $z\to\bar{z}$, not merely when they are parameter-independent. Let us present simple sufficient conditions for property (5.96); see also Commentary to this chapter for more results in this direction.

Proposition 5.70 (normal semicontinuity of moving sets). *Let* $S\colon Z\rightrightarrows X$ *be a multifunction from a metric space* Z *into a Banach space* X. *Then* S *is normally semicontinuous at* $(\bar{z},\bar{x})\in\operatorname{gph}S$ *in the following two cases:*

(i) $S(z)=g(z)+\Omega$ *around* $\bar{z}$, *where* $\Omega\subset X$ *is an arbitrary nonempty set and* $g\colon Z\to X$ *is continuous at* $\bar{z}$.

(ii) S *is convex-valued near* $\bar{z}$ *and inner semicontinuous at this point, i.e.,*

$$S(\bar{z})\subset\operatorname*{Lim\,inf}_{z\to\bar{z}}S(z)\ .$$

Proof. In case (i) the normal semicontinuity property follows directly from definitions (1.2) and (5.95) and from the continuity of $g(\cdot)$. Note that this case is sufficient for applications to the exact extremal principle involving the translation of fixed sets.

Let us consider case (ii). Taking $x^*\in N_+\big(\bar{x};S(\bar{z})\big)$, we find sequences $\varepsilon_k\downarrow 0$, $x_k\to\bar{x},\ z_k\to\bar{z}$, and $x_k^*\xrightarrow{w^*}x^*$ such that

$$x_k\in S(z_k)\quad\text{and}\quad x_k^*\in\widehat{N}_{\varepsilon_k}\big(x_k;S(z_k)\big)\quad\text{for all}\quad k\in I\!N\ .$$

Employing Proposition 1.3 on the representation of ε-normals to convex sets, one has the explicit description

$$\langle x_k^*,u-x_k\rangle\le\varepsilon_k\|u-x_k\|\ \text{ for all }\ u\in S(z_k)\ .$$

Let us show that the inner semicontinuity assumption in (ii) implies that

$$\langle x^*,u-\bar{x}\rangle\le 0\ \text{ for all }\ u\in S(\bar{z})\ ,$$

which means that $x^* \in N\big(\bar{x}; S(\bar{z})\big)$, since the basic normal cone agrees with the normal cone of convex analysis for convex sets. Indeed, assume on the contrary that the latter is violated at some $\bar{u} \in S(\bar{z})$, i.e., $\langle x^*, \bar{u} - \bar{x}\rangle > 0$. Using the inner semicontinuity of S at $\bar{z}$, for the given $\bar{u}$ and the sequence $z_k \to \bar{z}$ we find a sequence $u_k \to \bar{u}$ such that $u_k \in S(z_k)$ for all $k \in I\!N$. We have the representation

$$\langle x_k^*, u_k - x_k\rangle = \langle x^*, \bar{u} - \bar{x}\rangle + \Big[\langle x_k^* - x^*, \bar{u} - \bar{x}\rangle + \langle x_k^*, u_k - \bar{u}\rangle - \langle x_k^*, x_k - \bar{x}\rangle\Big] .$$

One can see that all the terms in the square brackets tend to zero as $k \to \infty$ due to the corresponding convergence of x_k, u_k, x_k^* and the boundedness of $\{x_k^*\}$. This allows us to conclude that

$$\langle x_k^*, u_k - x_k\rangle > \varepsilon_k \|u_k - x_k\| \ \text{ for large } k \in I\!N ,$$

which contradicts the above representation of ε-normals and completes the proof of the proposition. △

To proceed towards the exact extremal principle for multifunctions in the case of infinite-dimensional image spaces, we need the following normal compactness property of set-valued mappings, which involves their *images* but not graphs as in the basic SNC definition.

Definition 5.71 (SNC property of moving sets). *We say that a set-valued mapping $S\colon Z \rightrightarrows X$ between a metric space Z and a Banach space X is* IMAGELY SNC *(or just ISNC) at $(\bar{z}, \bar{x}) \in \operatorname{gph} S$ if for any sequences $(\varepsilon_k, z_k, x_k, x_k^*)$ satisfying*

$$x_k^* \in \widehat{N}_{\varepsilon_k}\big(x_k; S(z_k)\big), \quad \varepsilon_k \downarrow 0, \quad (z_k, x_k) \overset{\operatorname{gph} S}{\to} (\bar{z}, \bar{x}), \quad x_k^* \overset{w^*}{\to} 0$$

one has $\|x_k^*\| \to 0$ as $k \to \infty$.

This property is automatic, besides the finite-dimensional setting for X, when S admits the representation

$$S(z) = g(z) + \Omega \quad \text{around} \quad \bar{z}$$

provided that $g\colon Z \to X$ is continuous at $\bar{z}$ and that $\Omega \subset X$ is SNC at $\bar{x} - g(\bar{z})$. One may equivalently put $\varepsilon_k = 0$ in Definition 5.71 if X is Asplund and S is closed-valued around $\bar{z}$. Similarly to the case of fixed sets, there are strong relationships between the above ISNC property and the corresponding counterparts of the CEL property for moving sets. In particular, a mapping $S\colon Z \rightrightarrows X$ between Banach spaces is ISNC at $(\bar{z}, \bar{x})$ if there are numbers $\alpha, \eta > 0$ and a compact set $C \subset X$ such that

$$\widehat{N}_\varepsilon\big(x; S(z)\big) \subset \Big\{x^* \in X^* \Big|\ \eta\|x^*\| \le \varepsilon\alpha + \max_{c \in C} |\langle x^*, c\rangle|\Big\}$$

whenever $(z,x) \in \operatorname{gph} S \cap \big((\bar{z},\bar{x}) + \eta I\!B_{Z\times X}\big)$. The latter surely holds if S is *uniformly CEL* around $(\bar{x},\bar{z})$ in the sense that there are a compact set $C \subset Z$, neighborhoods $V \times U$ of $(\bar{x},\bar{z})$ and O of the origin in Z, and a number $\gamma > 0$ such that one has

$$S(x) \cap U + tO \subset S(x) + \gamma C \text{ for all } x \in U \text{ and } t \in (0,\gamma)\,;$$

cf. the proof of Theorem 1.26. In accordance with Definition 1.24, S is said to be *uniformly epi-Lipschitzian* around $(\bar{x},\bar{z})$ if C can be selected as a singleton. The latter is always fulfilled for any $\bar{x} \in S(\bar{z})$ if there is a neighborhood V of $\bar{z}$ such that $S(z)$ is convex for $z \in V$ and $\operatorname{int}\big(\cap_{z\in V} S(z)\big) \neq \emptyset$; cf. the proof of Proposition 1.25. Similarly to Subsect. 1.2.5 we can define the *partial* ISNC property of set-valued mappings and ensure the fulfillment of this property for *uniformly* Lipschitz-like as well as partially CEL multifunctions.

It is worth mentioning that the extended normal cone (5.95) and the ISNC property from Definition 5.71, as well their mapping/function counterparts and partial analogs, enjoy *full calculi* similar to those for the basic constructions and SNC properties developed in this book. We are not going to present and applied such results in what follows; their formulations and proofs are parallel to those for "non-moving" objects.

Now we are ready to establish the exact/limiting extremal principle for systems of multifunctions, which extends (is actually equivalent to) the exact extremal principle for systems of sets obtained in Theorem 2.22.

Theorem 5.72 (exact extremal principle for multifunctions).

(i) *Let $S_i\colon M_i \rightrightarrows X$, $i = 1,\ldots,n$, be multifunctions from metric spaces (M_i,d_i) into an Asplund space X. Assume that $\bar{x}$ is a local extremal point of the system $\{S_1,\ldots,S_n\}$ at $(\bar{s}_1,\ldots,\bar{s}_n)$, where each S_i is closed-valued around $\bar{s}_i$ and all but one of them are ISNC at the corresponding points $(\bar{s}_i,\bar{x})$ of their graphs. Then the following* EXACT EXTREMAL PRINCIPLE *holds: there are*

$$x_i^* \in N_+\big(\bar{x}; S_i(\bar{s}_i)\big) \quad for \quad i = 1,\ldots,n$$

satisfying the generalized Euler equation

$$x_1^* + \ldots + x_n^* = 0 \quad with \quad (x_1^*,\ldots,x_n^*) \neq 0\,.$$

(ii) *Conversely, let the exact extremal principle hold for every extremal system of two multifunctions $\{S_1,S_2,\bar{x}\}$ with the image space X, where both mappings S_i are closed-valued around the corresponding points $\bar{s}_i$ and one of them is ISNC at $(\bar{s}_i,\bar{x})$. Then X is Asplund.*

Proof. Part (ii) follows directly from Theorem 2.22(ii), since the exact extremal principle for systems of multifunctions implies the one for systems of sets, while the ISNC property for moving sets reduces to the standard SNC property when sets are fixed. It remains to justify part (i) of the theorem.

To proceed, we apply the approximate extremal principle given in Theorem 5.68(b) when X is Asplund. It ensures that for each $k \in I\!N$ there are

s_{ik} with $d(s_{ik},\bar{s}_i)\le\frac{1}{k}$, $x_{ik}\in\bar{x}+\frac{1}{k}I\!B$, and $x^*_{ik}\in\widehat{N}\big(x_{ik};S_i(s_{ik})\big)$, $i=1,\ldots,n$, satisfying the relations

$$\|x^*_{1k}\|+\ldots+\|x^*_{nk}\|\ge 1-1/k\quad\text{and}\quad\|x^*_{1k}+\ldots+x^*_{nk}\|\le 1/k\ . \tag{5.97}$$

By normalization if necessary one can always select bounded sequences $\{x^*_{ik}\}$, $i=1,\ldots,n$, satisfying (5.97). Since the dual ball $I\!B^*\subset X^*$ is sequentially weak* compact by the Asplund property of X, we find $x^*_i\in X^*$ such that $x^*_{ik}\xrightarrow{w^*}x^*_i$ along a subsequence of $k\to\infty$ for all $i=1,\ldots,n$. Now passing to the limit as $k\to\infty$ and using definition (5.95), we arrive at the desired relationships in the theorem except the nontriviality of $(x^*_1,\ldots,x^*_n)$.

To establish the latter, suppose that all x^*_i are zero and assume for definiteness that the first $n-1$ mappings S_i are ISNC at $(\bar{s}_i,\bar{x})$, $i=1,\ldots,n-1$. Then $\|x^*_{ik}\|\to 0$ as $k\to\infty$ for $i=1,\ldots,n-1$ by the construction of x^*_i. Passing to the limit at the second relation in (5.97), we conclude that $\|x^*_{nk}\|\to 0$ as well. This clearly contradicts the first relation in (5.97) for large $k\in I\!N$ and completes the proof of the theorem. △

Note that the extended normal cone (5.95) *cannot* be generally replaced in Theorem 5.72 by the basic one (1.2) unless the corresponding mapping S_i is assumed to be normally semicontinuous. Indeed, consider the extremal system of multifunctions $\{S_1,S_2\}$ defined in (5.90) with the local extremal point $\bar{x}=0\in I\!R^2$ at $(\bar{s}_1,\bar{s}_2)=(0,0)$. It is easy to check that neither S_1 nor S_2 is normally semicontinuous at $(0,0,0)$, and that

$$N\big(0,S_1(0)\big)\cap\big[-N\big(0,S_2(0)\big)\big]=\{0\}\ .$$

Hence an analog of Theorem 5.72 with N_+ replaced by N doesn't hold for this extremal system of multifunctions.

5.3.4 Optimality Conditions with Respect to Closed Preferences

In this subsection we present some applications of the extended extremal principle to general problems of constrained multiobjective optimization, where objective mappings are "minimized" with respect to closed preference relations. Let us first consider the following multiobjective problem with only geometric constraints:

$$\text{minimize}\ \ f(x)\ \ \text{with respect to}\ \ \prec\ \ \text{subject to}\ \ x\in\Omega\ , \tag{5.98}$$

where $f\colon X\to Z$ is a mapping between Banach spaces, where $\Omega\subset X$, and where $\prec$ is a nonreflexive preference relation on Z with the moving level set $\mathcal{L}(\cdot)$ from Definition 5.55. The next theorem gives necessary optimality conditions for (5.98) in both approximate/fuzzy and exact/limiting forms.

Theorem 5.73 (optimality conditions for problems with closed preferences and geometric constraints). *Let $\bar{x}$ be a local optimal solution to problem* (5.98) *with $\bar{z} := f(\bar{x})$, where the preference $\prec$ is closed and where both spaces X and Z are Asplund. Assume that f is continuous around $\bar{x}$ and that Ω is locally closed around this point. The following assertions hold:*

(i) *For every $\varepsilon > 0$ there are $(x_0, x_1, z_0, z_1, x^*, z^*) \in X^2 \times Z^2 \times X^* \times Z^*$ satisfying $x_0, x_1 \in \bar{x} + \varepsilon I\!B_X$, $z_0, z_1 \in \bar{z} + \varepsilon I\!B_Z$,*

$$x^* \in \widehat{N}(x_1; \Omega), \quad z^* \in \widehat{N}\big(z_1; \mathrm{cl}\, \mathcal{L}(z_0)\big)$$

with $\|(x^*, z^*)\| = 1$, and

$$0 \in x^* + \widehat{D}^* f(x_0)(z^*) + \varepsilon I\!B_{X^*} \; .$$

Moreover, one has

$$0 \in \widehat{\partial}\langle z^*, f\rangle(x_0) + \widehat{N}(x_1; \Omega) + \varepsilon I\!B_{X^*} \quad \textit{with} \quad \|z^*\| = 1$$

if f is Lipschitz continuous around $\bar{x}$.

(ii) *Assume that either f is SNC at $\bar{x}$, or Ω is SNC at $\bar{x}$ and the mapping $\mathrm{cl}\, \mathcal{L}\colon Z \rightrightarrows Z$ generated by the level sets of $\prec$ is ISNC at $(\bar{z}, \bar{z})$. Then there are x^* and z^*, not both zero, satisfying*

$$x^* \in D^*_N f(\bar{x})(z^*) \cap \big(- N(\bar{x}; \Omega)\big) \;\; \textit{and} \;\; z^* \in N_+\big(\bar{z}; \mathrm{cl}\, \mathcal{L}(\bar{z})\big) \; .$$

Furthermore, one has

$$0 \in \partial\langle z^*, f\rangle(\bar{x}) + N(\bar{x}; \Omega) \;\; \textit{with} \;\; z^* \in N_+\big(\bar{z}; \mathrm{cl}\, \mathcal{L}(\bar{z})\big) \setminus \{0\}$$

provided that f is strictly Lipschitzian at $\bar{x}$ and either $\dim Z < \infty$, or Ω is SNC at $\bar{x}$ and $\mathrm{cl}\, \mathcal{L}$ is ISNC at $(\bar{z}, \bar{z})$.

Proof. First we prove (i) based on the approximate extremal principle from Theorem 5.68. It is shown in Example 5.65 that $(\bar{x}, \bar{z})$ is a local extremal point of the system $\{S_1, S_2\}$ at $(\bar{z}, 0)$, where $S_i\colon M_i \rightrightarrows X \times Z$, $i = 1, 2$, are defined therein with

$$S_1(z) = \Omega \times \mathrm{cl}\, \mathcal{L}(z) \quad \text{and} \quad S_2 \equiv \mathrm{gph}\, f \; .$$

Since the space $X \times Z$ is Asplund and both S_i are locally closed-valued under the general assumptions made, we apply the assertion of Theorem 5.68(b), which gives $z_0 \in \bar{z} + \varepsilon I\!B_Z$ and $(x_i, z_i) \in (\bar{x}, \bar{z}) + \varepsilon I\!B_{X \times Z}$ for $i = 1, 2$ satisfying

$$(x_1^*, z_1^*) \in \widehat{N}\big((x_1, z_1); S_1(z_0)\big), \quad (x_2^*, z_2^*) \in \widehat{N}\big((x_2, z_2); S_2\big) \; ,$$

$$\|(x_1^*, z_1^*) + (x_2^*, z_2^*)\| \le \varepsilon, \quad \|(x_1^*, z_1^*)\| + \|(x_2^*, z_2^*)\| \ge 1 - \varepsilon \; ,$$

where $x_1 \in \Omega$, $z_1 \in \mathrm{cl}\,\mathcal{L}(z_0)$, and $z_2 = f(x_2)$. Taking into account the structures of S_1, S_2 and the product formula for $\widehat{N}$ from Proposition 1.2, we have from the first line above that

$$x_1^* \in \widehat{N}(x_1;\Omega), \quad z_1^* \in \widehat{N}\big(z_1;\mathrm{cl}\,\mathcal{L}(z_0)\big), \quad x_2^* \in \widehat{D}^* f(x_2)(-z_2^*)\ .$$

Put $x_0 := x_2$ $x^* := x_1^*$, $z^* := z_1^*$ and employ normalization to ensure $\|(x^*,z^*)\| = 1$. Then using the second line above and shrinking ε if necessary, one easily gets that the pair (x^*,z^*) satisfies all the conclusions in (i) when f is supposed to be merely continuous around $\bar{x}$. If f is assumed to be Lipschitz continuous around this point, then we know that

$$\widehat{D}^* f(x_0)(z^*) = \widehat{\partial}\langle z^*, f\rangle(x_0)\ ,$$

which therefore completes the proof of assertion (i).

To prove (ii), we apply the exact extremal principle from Theorem 5.72(i) to the extremal system $\{S_1, S_2, (\bar{x},\bar{z})\}$ under consideration. The structures of S_i and the product formulas in Proposition 1.2 ensure that the ISNC/SNC assumptions of the theorem imply the required ISNC properties in Theorem 5.72, and also that

$$N_+\big((\bar{x},\bar{z});S_1(\bar{z})\big) = N(\bar{x};\Omega) \times N_+\big(\bar{z};\mathrm{cl}\,\mathcal{L}(\bar{z})\big)\ ,$$

$$N_+\big((\bar{x},\bar{z});S_2\big) = \big\{(x^*,z^*)\big|\ x^* \in D_N^* f(\bar{x})(-z^*)\big\}\ .$$

Thus all the conclusions in the first part of (ii) follow directly from the exact extremal principle of Theorem 5.72.

To justify the necessary optimality conditions in the second part of (ii), it suffices to observe that, by Theorem 3.28,

$$D_N^* f(\bar{x})(z^*) = \partial\langle z^*, f\rangle(\bar{x}) \ \text{ when } \ f \ \text{ is strictly Lipschitzian at } \bar{x}\ ,$$

and that f is SNC at $\bar{x}$ if it Lipschitz continuous around $\bar{x}$ while $\dim Z < \infty$; see Corollary 1.69(i). This completes the proof of the theorem. △

It is worth mentioning that when $f\colon X \to Z$ is strictly Lipschitzian at $\bar{x}$ and X is Asplund, the SNC property of f at $\bar{x}$ is *equivalent* to the finite dimensionality of Z due to Corollary 3.30. Observe also that the only difference between the ISNC property of the mapping $\mathrm{cl}\,\mathcal{L}$ at $(\bar{z},\bar{z})$ in Theorem 5.73(ii) and the one for the level set mapping $\mathcal{L}$ is that $\bar{z} \in \mathrm{cl}\,\mathcal{L}(\bar{z})$ while $\bar{z} \notin \mathcal{L}(\bar{z})$, since the preference $\prec$ is locally satiated and nonreflexive.

Remark 5.74 (comparison between optimality conditions for multiobjective problems). We obtained above the two basic results on necessary optimality conditions in problems of multiobjective optimization with geometric constraints: Theorem 5.59 and Theorem 5.73. Although both concepts of multiobjective optimality considered in these theorems extend most of the conventional notions, they are generally different; see the results and discussions in Subsect. 5.3.1. Nevertheless, necessary optimality conditions obtained in Theorems 5.59 and 5.73 have a lot in common. Compare, in particular, the coderivative conditions in assertions (ii) of these theorems. Employing the coderivative sum rule from Proposition 3.12 to $f_\Omega(x) = f(x) + \Delta(x;\Omega)$ with the qualification condition

$$D_N^* f(\bar{x})(0) \cap \big(-N(\bar{x};\Omega)\big) = \{0\}\ ,$$

we derive from (5.79) and the normal compactness conditions imposed in Theorem 5.59(ii) that the (f,Θ)-optimality of $\bar{x}$ relative to Ω implies the existence of $(x^*,z^*) \neq 0$ satisfying

$$0 \in x^* + D_N^* f(\bar{x})(z^*), \quad x^* \in N(\bar{x};\Omega), \quad z^* \in N(0;\Theta)$$

provided that either f is SNC at $\bar{x}$, or Ω is SNC at $\bar{x}$ and Θ is SNC at 0. In the general setting (even in finite dimensions) Theorem 5.59(ii) gives more delicate necessary conditions for generalized order optimality. On the other hand, Theorem 5.73(ii) applies to multiobjective optimization problems with respect to closed preference relations that cannot be handled by conventional translations of fixed sets in extremal systems but involve nonlinear deformations of moving sets.

Similarly to the case of generalized order optimality in Subsect. 5.3.2, as well as to previous developments in this chapter, one can derive various consequences of Theorem 5.73 in multiobjective problems with closed preference relations under operator and functional constraints. All these consequences are based on applications of the comprehensive generalized differential and SNC calculi developed in Chap. 3. As an example of such results, let us present the following corollary of the coderivative optimality conditions in Theorem 5.73(ii) to multiobjective problems with *operator constraints.*

Corollary 5.75 (optimality conditions for problems with closed preferences and operator constraints). *Let $\prec$ be a closed preference on Z with the level set $\mathcal{L}(\cdot)$, and let $\bar{x}$ be a local optimal solution to the multiobjective optimization problem:*

$$\text{minimize } \ f(x) \ \text{ with respect to } \ \prec \ \text{ subject to } \ x \in G^{-1}(\Lambda)\ ,$$

where $f\colon X \to Z$ and $G\colon X \rightrightarrows Y$ are mappings between Asplund spaces with $\bar{z} := f(\bar{x})$, and where $\Lambda \subset Y$. Suppose that f is continuous and $S(\cdot) := G(\cdot)\cap\Lambda$ is inner semicompact around $\bar{x}$, and that the sets $\operatorname{gph} G$ *and Λ are locally closed around the corresponding points. Then there are x^* and z^*, not both zero, such that*

$$-x^* \in D_N^* f(\bar{x})(z^*), \quad z^* \in N_+\big(\bar{z};\operatorname{cl}\mathcal{L}(\bar{z})\big), \quad \text{and}$$

$$x^* \in \bigcup \Big[D_N^* G(\bar{x}, \bar{y})(y^*) \Big| \; y^* \in N(\bar{y}; \Lambda), \;\; \bar{y} \in S(\bar{x}) \Big]$$

under one of the following requirements on (f, G, Λ)*:*

(a) f *is SNC at* $\bar{x}$*, the qualification condition*

$$N(\bar{y}; \Lambda) \cap \ker \widetilde{D}_M^* G(\bar{x}, \bar{y}) = \{0\} \;\; \text{for all} \;\; \bar{y} \in S(\bar{x})$$

is satisfied, and either G^{-1} *is PSNC at* $(\bar{y}, \bar{x})$ *or* Λ *is SNC at* $\bar{y}$ *for all* $\bar{y} \in S(\bar{x})$*.*

(b) $\mathrm{cl}\, \mathcal{L}$ *is ISNC at* $(\bar{z}, \bar{z})$*, the qualification condition*

$$N(\bar{y}; \Lambda) \cap \ker D_N^* G(\bar{x}, \bar{y}) = \{0\} \;\; \text{for all} \;\; \bar{y} \in S(\bar{x})$$

is satisfied, and either G *is PSNC at* $(\bar{x}, \bar{y})$ *and* Λ *is SNC at* $\bar{y}$*, or* G *is SNC at* $(\bar{x}, \bar{y})$ *for all* $\bar{y} \in S(\bar{x})$*.*

Proof. To derive this corollary from the coderivative optimality conditions of Theorem 5.73(ii), it suffices to apply Theorem 3.8 that gives the representation of basic normals to $G^{-1}(\Lambda)$ under the assumptions in (a), and Theorem 3.84 that ensures the SNC property of $G^{-1}(\Lambda)$ under the assumptions in (b). △

Let us next consider multiobjective problems with respect to closed preferences under *functional constraints* of equality and inequality types. Similarly to Subsect. 5.3.2, we may derive necessary optimality conditions for such problems of the two types: involving basic *lower subgradients* of constraint functions and also those using Fréchet *upper subgradients* of functions describing inequality constraints. For simplicity we present results only for problems with inequality constraints, since only these constraints distinguish between lower and upper subdifferential conditions.

Theorem 5.76 (lower and upper subdifferential conditions for multiobjective problems with inequality constraints). *Let* $\prec$ *be a closed preference on* Z *with the level set* $\mathcal{L}(\cdot)$*, and let* $\bar{x}$ *be a local optimal solution to the multiobjective problem:*

$$\text{minimize} \;\; f(x) \;\; \text{with respect to} \;\; \prec \;\; \text{subject to} \;\; \varphi_i(x) \le 0, \quad i = 1, \ldots, m \, ,$$

where $f\colon X \to Z$ *is continuous around* $\bar{x}$ *with* $\bar{z} := f(\bar{x})$*, while* $\varphi_i\colon X \to \overline{I\!R}$ *are merely finite at* $\bar{x}$ *for all* $i = 1, \ldots, m$*. Suppose that either* f *is SNC at* $\bar{x}$ *or* $\mathrm{cl}\, \mathcal{L}$ *is ISNC at* $(\bar{z}, \bar{z})$*. The following assertions hold:*

(i) *Assume that both spaces* X *and* Z *are Asplund, and that each* φ_i *is Lipschitz continuous around* $\bar{x}$*. Then there are* $z^* \in Z^*$ *and multipliers* $(\lambda_1, \ldots, \lambda_m) \in I\!R^m$ *satisfying*

$$z^* \in N_+\big(\bar{x}; \mathrm{cl}\, \mathcal{L}(\bar{z})\big), \quad \lambda_i \ge 0, \quad \lambda_i \varphi_i(\bar{x}) = 0 \;\; \text{as} \;\; i = 1, \ldots, m \qquad (5.99)$$

such that $(z^*, \lambda_1, \ldots, \lambda_m) \ne 0$ *and one has*

$$0 \in D_N^* f(\bar{x})(z^*) + \partial\Big(\sum_{i=1}^m \lambda_i \varphi_i\Big)(\bar{x}) \subset D_N^* f(\bar{x})(z^*) + \sum_{i=1}^m \lambda_i \partial\varphi_i(\bar{x}) \ .$$

(ii) *Assume that Z is Asplund while X admits a Lipschitzian $\mathcal{C}^1$ bump function (which is automatic when X admits a Fréchet smooth renorm). Then for any $x_i^* \in \widehat{\partial}^+\varphi_i(\bar{x})$, $i = 1, \ldots, m$, there are $0 \neq (z^*, \lambda_1, \ldots, \lambda_m) \in Z^* \times I\!R^m$ satisfying* (5.99) *and*

$$0 \in D_N^* f(\bar{x})(z^*) + \sum_{i=1}^m \lambda_i x_i^* \ .$$

Proof. The lower subdifferential optimality conditions in assertions (i) of the theorem follow directly from assertions (i) and (ii) of Corollary 5.73 with $Y = I\!R^m$, $G(x) = \big(\varphi_1(x), \ldots, \varphi_m(x)\big)$, and $\Lambda = R^m_-$. Indeed, it suffices to observe that in this case one has

$$N(\bar{y}; \Lambda) = \big\{(\lambda_1, \ldots, \lambda_m) \in I\!R^m \big|\ \lambda_i \geq 0, \ \ \lambda_i \varphi_i(\bar{x}) = 0, \ \ i = 1, \ldots, m\big\} \ ,$$

$$D^* G(\bar{x})(\lambda_1, \ldots, \lambda_m) = \partial\Big(\sum_{i=1}^m \lambda_i \varphi_i\Big)(\bar{x}) \subset \sum_{i=1}^m \lambda_i \partial\varphi_i(\bar{x}) \ .$$

To justify the upper subdifferential condition in (ii), we take arbitrary elements $x_i^* \in \widehat{\partial}^+\varphi_i(\bar{x})$ for $i = 1, \ldots, m$ and find, by the variational descriptions of Fréchet subgradients from Theorem 1.88(ii), functions $s_i\colon X \to \overline{I\!R}$ continuously differentiable in some neighborhood U of $\bar{x}$ and such that

$$s_i(\bar{x}) = \varphi_i(\bar{x}), \quad \nabla s_i(\bar{x}) = x_i^*, \quad s_i(x) \geq \varphi_i(x) \ \text{ for all } \ i = 1, \ldots, m \ .$$

It is easy to see that $\bar{x}$ is a local optimal solution to the multiobjective problem

$$\text{minimize } \ f(x) \ \text{ with respect to } \ \prec \ \text{ subject to } \ s_i(x) \leq 0, \quad i = 1, \ldots, m \ .$$

Applying now the optimality condition from assertion (i) of this theorem to the latter problem, we complete the proof of (ii). △

In the conclusion of this subsection let us briefly discuss some applications of the (extended) extremal principle to a class of *multiobjective games with many players*. Such problems can be roughly described as games with n players, where each player wants to choose a strategy $\bar{x}_i$ from a space X_i such that they $\prec_i$ optimize (with respect to the preference $\prec_i$ on Y) an objective mapping $f : X_1 \times \ldots \times X_n \to Z$ given all other players choices $\bar{x}_j$, $j \neq i$.

This is a general game setting that covers, in particular, the case when each of the players can have a different objective mapping $f_i\colon X_1 \times \ldots \times X_n \to Z_i$. In the latter case one has $f := (f_1, \ldots, f_n)\colon X_1 \times \ldots \times X_n \to Z := Z_1 \times \ldots \times Z_n$ with the ordering $\prec_i$ on Z defined by

$$z \prec_i v \;\text{ for }\; z,v \in Z \quad \text{provided that} \quad z_i \prec_i v_i \;\text{ for }\; z_i, v_i \in Z_i \ .$$

It is well known that an essential concept in all game theory is that of a *saddle point.* Let us give a generalized version of this concept for the above multiobjective setting, where $\prec$ stands for $(\prec_1, \ldots, \prec_n)$.

Definition 5.77 (saddle points for multiobjective games). *A point $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)$ is a* LOCAL $\prec$-SADDLE POINT *of $f\colon X_1 \times \ldots \times X_n \to Z$ if for each $i = 1, \ldots, n$ there is a neighborhood U_i of $\bar{x}_i$ such that*

$$f(\bar{x}) \prec_i f(\bar{x}_1, \ldots, \bar{x}_{i-1}, x_i, \bar{x}_{i+1}, \ldots, \bar{x}_n) \quad \textit{for all} \quad x_i \in U_i \ .$$

Observe that this notion of saddle points may be different from the usual concept considered in Example 5.66 with preferences not depending on players and spaces. Indeed, let the payoff mapping $f : I\!R^4 \to I\!R^2$ be given by

$$f(x,y,u,v) := (x^2 + u, -y^2 - e^v) \ ,$$

and let us group the variables so that x and y are for the first player and u and v are for the second one. This means that $X_1 = X_2 = Z = I\!R^2$. The order $\prec_1$ on $Z = I\!R^2$ for the first player is that $(w,s) \prec_1 (\tilde{w}, \tilde{s})$ if $w < \tilde{w}$ and $s \geq \tilde{s}$ or $w \leq \tilde{w}$ and $s > \tilde{s}$. The order $\prec_2$ on $Z = I\!R^2$ for the second player is that $(w,s) \prec_2 (\tilde{w}, \tilde{s})$ if $w < \tilde{w}$ and $s < \tilde{s}$. This is a *mixture of Pareto and weak Pareto optimality.* One can check that any point of the form $(0,0,u,v)$ is a local $\prec$*-saddle point* for these orderings.

Now we present necessary optimality conditions for multiobjective games with additional constraints on player strategies. For simplicity we formulate results only for the case of geometric constraints.

Given $f\colon X_1 \times \ldots \times X_n \to Z$ and $\prec_i$ as in Definition 5.77 and constraint sets $\Omega_i \subset X_i$ for $i = 1, \ldots, n$, we consider the following multiobjective constrained game $\mathcal{G}$: find local $\prec$-saddle points of f subject to the constraints $x_i \in \Omega_i \subset X_i$ for each $i = 1, \ldots, n$. Let $\bar{x}$ be a local optimal solution to game $\mathcal{G}$. Then one has, by Definition 5.77 of $\prec$-saddle points, that the i-th component $\bar{x}_i$ of $\bar{x}$ is a local solution to the following multiobjective constrained optimization problem for each player i:

$$\text{minimize } \; f(\bar{x}_1, \ldots, \bar{x}_{i-1}, x_i, \bar{x}_{i+1}, \ldots, \bar{x}_n) \; \text{ subject to } \; x_i \in \Omega_i \ ,$$

where “minimization” is understood with respect to the preference $\prec_i$ on Z.

Denote $f_i(x_i) := f(\bar{x}_1, \ldots, \bar{x}_{i-1}, x_i, \bar{x}_{i+1}, \ldots, \bar{x}_n)$, $\bar{z}_i := f_i(\bar{x}_i)$ for $i = 1, \ldots, n$ and consider the level sets $\mathcal{L}_i(z)$ induced by the preferences $\prec_i$ on Z. Employing the above results for problems of multiobjective optimization, based on the approximate and exact versions of the extremal principle for multifunctions, we arrive at necessary optimality conditions in multiobjective games. For brevity these results are formulated only for the case of Lipschitzian objective mappings.

Theorem 5.78 (optimality conditions for multiobjective games). *Let $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)$ be a local optimal solution to the above game $\mathcal{G}$, where the preferences $\prec_i$ are closed on Z and where the spaces $X_1, \ldots, X_n, Z$ are Asplund. Suppose that the mapping $f\colon X_1 \times \ldots \times X_n \to Z$ is Lipschitz continuous around $\bar{x}$ and that the sets $\Omega_i \subset X_i$ are locally closed around $\bar{x}_i$ for all $i = 1, \ldots, n$. The following assertions hold:*

(i) *For every $\varepsilon > 0$ there are $(x_i, u_i, z_i, v_i, z_i^*) \in X_i \times X_i \times Z \times Z \times Z^*$ satisfying $x_i, u_i \in \bar{x} + \varepsilon I\!B_{X_i}$, $z_i, v_i \in \bar{z}_i + \varepsilon I\!B_Z$, $\|z_i^*\| = 1$, and*

$$0 \in \widehat{\partial}\langle z_i^*, f_i\rangle(x_i) + \widehat{N}(u_i; \Omega_i) + \varepsilon I\!B_{X_i^*}, \quad z_i^* \in \widehat{N}\big(v_i; \mathrm{cl}\, \mathcal{L}_i(z_i)\big), \quad i = 1, \ldots, n\ .$$

(ii) *Assume that f is strictly Lipschitzian at $\bar{x}$ and either $\dim Z < \infty$, or Ω_i is SNC at $\bar{x}_i$ and $\mathrm{cl}\, \mathcal{L}_i$ is ISNC at $(\bar{z}_i, \bar{z}_i)$ for each $i = 1, \ldots, n$. Then there are $z_1^*, \ldots, z_n^* \in Z^*$ such that $\|z_i^*\| = 1$ and*

$$0 \in \partial\langle z_i^*, f_i\rangle(\bar{x}_i) + N(\bar{x}_i; \Omega_i), \quad z_i^* \in N_+\big(\bar{z}_i; \mathrm{cl}\, \mathcal{L}_i(\bar{z}_i)\big) \quad \textit{as} \quad i = 1, \ldots, n\ .$$

Proof. Since for each player $i = 1, \ldots, n$ the i-th component $\bar{x}_i$ of $\bar{x}$ is a local optimal solution to the multiobjective optimization problem formulated above, we apply both assertions of Theorem 5.73 to these problems and get the necessary optimality conditions in (i) and (ii). △

5.3.5 Multiobjective Optimization with Equilibrium Constraints

The last subsection of this section is devoted to problems of *multiobjective optimization* that involve *equilibrium constraints* of the type

$$0 \in q(x, y) + Q(x, y)$$

governed by parametric variational systems. We have considered such constraints in Sect. 5.3 in the framework of MPECs with single (real-valued) objective functions. Now we are going to study multiobjective optimization problems with equilibrium constraints, where optimal solutions are understood either in the sense of generalized order optimality from Definition 5.53 or in the sense of closed preference relations from Definition 5.55. As discussed in Subsect. 5.3.1, both of these multiobjective notions cover, in particular, standard equilibrium concepts related to Pareto-type optimality/efficiency and the like. Thus the multiobjective optimization problems studied in what follows include the so-called *equilibrium problems with equilibrium constraints* (EPECs) that are important for many applications. Note that equilibrium concepts on the upper level of multiobjective problems can be described by *vector variational inequalities*; see, in particular, Giannessi [504] and the references therein. For convenience we adopt the abbreviation *EPECs* (or *EPEC problems*, slightly abusing the language) for all the multiobjective problems with equilibrium-type constraints considered in this subsection.

Although EPECs may have constraints of other types (geometric, operator, functional) along with equilibrium ones, they are not included for brevity; it can be done similarly to Sect. 5.2. We pay the main attention to *point-based/exact* necessary optimality conditions for EPECs formulated at the reference optimal solution.

First let us study EPECs, where optimal solutions are understood in the sense of generalized order optimality from Definition 5.53. The following result gives necessary optimality conditions for an *abstract version* of such problems with equilibrium constraints described by a general parameter-dependent multifunction. In its formulation we use the *strong PSNC* property of $F\colon X \rightrightarrows Y$ at $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ that, in accordance with Definitions 1.67 and 3.3, means that for any sequences $(\varepsilon_k, x_k, y_k, x_k^*, y_k^*) \in [0, \infty) \times (\operatorname{gph} F) \times X^* \times Y^*$ satisfying

$$\varepsilon_k \downarrow 0,\ (x_k, y_k) \to (\bar{x}, \bar{y}),\ x_k^* \in \widehat{D}^*_{\varepsilon_k} F(x_k, y_k)(y^*),\ \text{and}\ (x_k^*, y_k^*) \stackrel{w^*}{\to} (0,0)$$

one has $\|x_k^*\| \to 0$ as $k \to \infty$. It holds, in particular, for mappings $F\colon X \rightrightarrows Y$ that are *partially CEL* around $(\bar{x}, \bar{y})$; see Theorem 1.75. Note that one can equivalently put $\varepsilon_k = 0$ in the relations above for closed-graph mappings between Asplund spaces.

Theorem 5.79 (generalized order optimality for abstract EPECs). *Let $f\colon X \times Y \to Z$, $\Theta \subset Z$ with $0 \in \Theta$, and $S\colon X \rightrightarrows Y$ with $(\bar{x}, \bar{y}) \in \operatorname{gph} S$. Suppose that the point $(\bar{x}, \bar{y})$ is locally (f, Θ)-optimal subject to $y \in S(x)$. The following assertions hold:*

(i) *Assume that the set*

$$\mathcal{E}(f, S, \Theta) := \big\{(x, y, z) \in X \times Y \times Z \big|\ f(x, y) - z \in \Theta,\ \ y \in S(x)\big\}$$

is locally closed around $(\bar{x}, \bar{y}, \bar{z})$ with $\bar{z} := f(\bar{x}, \bar{y})$ and that $\dim Z < \infty$. Then there is $z^* \in Z^*$ satisfying

$$(0, -z^*) \in N\big((\bar{x}, \bar{y}, \bar{z}); \mathcal{E}(f, S, \Theta)\big), \quad z^* \in N(0; \Theta) \setminus \{0\}\ .$$

(ii) *Assume that Z is Asplund, that f is continuous around $(\bar{x}, \bar{y})$, and that $\operatorname{gph} S$ and Θ are locally closed around $(\bar{x}, \bar{y})$ and 0, respectively. Then there is $z^* \in N(0; \Theta) \setminus \{0\}$ satisfying*

$$0 \in D^*_N f(\bar{x}, \bar{y})(z^*) + N\big((\bar{x}, \bar{y}); \operatorname{gph} S\big) \tag{5.100}$$

in each of the following cases:

(a) *Θ is SNC at 0,*

$$\Big[(x^*, y^*) \in D^*_M f(\bar{x}, \bar{y})(0), \quad -x^* \in D^*_N S(\bar{x}, \bar{y})(y^*)\Big] \Longrightarrow x^* = y^* = 0\ ,$$

and either S is SNC at $(\bar{x}, \bar{y})$ or f is PSNC at this point; the latter property and the above qualification condition are automatic when f is Lipschitz continuous around $(\bar{x}, \bar{y})$.

(b) *f is Lipschitz continuous around $(\bar{x},\bar{y})$, f^{-1} is strongly PSNC at $(\bar{z},\bar{x},\bar{y})$, and*

$$\Big[(x^*,y^*)\in D_N^*f(\bar{x},\bar{y})(0),\quad -x^*\in D_N^*S(\bar{x},\bar{y})(y^*)\Big]\Longrightarrow x^*=y^*=0\,.$$

Moreover, (5.100) *is equivalent to*

$$0\in\partial\langle z^*,f\rangle(\bar{x},\bar{y})+N\big((\bar{x},\bar{y});\operatorname{gph}S\big)$$

if f is strictly Lipschitzian at $(\bar{x},\bar{y})$.

Proof. Observe the EPEC problem under consideration is equivalent to the multiobjective optimization problem studied in Theorem 5.59 for the mapping f of two variables under the geometric constraints $(x,y)\in\Omega:=\operatorname{gph}S$. Thus assertion (i) of this theorem follows directly from assertion (i) of Theorem 5.59. To prove (ii), we use Theorem 5.59(ii) that ensures the existence of $z^*\in N(0;\Theta)\setminus\{0\}$ satisfying

$$0\in D_N^*\big(f+\Delta(\cdot;\operatorname{gph}S)\big)(\bar{x},\bar{y})$$

when either Θ is SNC at 0 or $\big(f+\Delta(\cdot;\operatorname{gph}S)\big)^{-1}$ is PSNC at $(\bar{z},\bar{x},\bar{y})$. To proceed, we apply the coderivative sum rule from Proposition 3.12 to the special sum $f+\Delta(\cdot;\operatorname{gph}S)$. This gives

$$0\in D_N^*f(\bar{x},\bar{y})(z^*)+N\big((\bar{x},\bar{y});\operatorname{gph}S\big)$$

under the limiting qualification condition (3.25) of that proposition, which is automatically fulfilled if

$$D_M^*f(\bar{x},\bar{y})(0)\cap\big(-N((\bar{x},\bar{y});\operatorname{gph}S)\big)=\{0\}$$

and if either f is PSNC at $(\bar{x},\bar{y})$ or S is SNC at this point; it certainly holds when f is Lipschitz continuous around $(\bar{x},\bar{y})$. Thus we get (5.100) under the assumptions in (a).

To justify (5.100) in case (b), one needs to check that the assumptions in (b) yield that $\big(f+\Delta(\cdot;\operatorname{gph})\big)^{-1}$ is PSNC at $(\bar{z},\bar{x},\bar{y})$. Indeed, the latter property means that for any sequences $(x_k,y_k,x_k^*,y_k^*,z_k^*)$ with $(x_k,y_k)\to(\bar{x},\bar{y})$ satisfying

$$(x_k^*,y_k^*)\in\widehat{D}^*\big(f+\Delta(\cdot;\operatorname{gph}S)\big)(x_k,y_k)(z_k^*),\quad\|(x_k^*,y_k^*)\|\to 0,\ \text{ and }\ z_k^*\xrightarrow{w^*}0$$

one has $\|z_k^*\|\to 0$ as $k\to\infty$. It follows from the proof of Theorem 3.10 that the qualification condition in (b) implies the fuzzy sum rule for the Fréchet coderivative $\widehat{D}^*\big(f+\Delta(\cdot;\operatorname{gph}S)\big)(x_k,y_k)$ considered above, which ensures the existence of $\varepsilon_k\downarrow 0$, $(x_{ik},y_{ik})\to(\bar{x},\bar{y})$ for $i=1,2$, and $(\tilde{x}_k^*,\tilde{y}_k^*,\tilde{z}_k^*)$ such that

$$(\tilde{x}_k^*, \tilde{y}_k^*) \in \widehat{D}^* f(x_{1k}, y_{1k})(\tilde{z}_k^*) + \widehat{N}\big((x_{2k}, y_{2k}); \operatorname{gph} S\big)$$

and $\|(\tilde{x}_k^*, \tilde{y}_k^*, \tilde{z}_k^*) - (x_k^*, y_k^*, z_k^*)\| \le \varepsilon_k$ for all $k \in I\!N$. Thus $(\tilde{x}_k^*, \tilde{y}_k^*) = (x_{1k}^*, y_{1k}^*) + (x_{2k}^*, y_{2k}^*)$ for some

$$(x_{1k}^*, y_{1k}^*) \in \widehat{D}^* f(x_{1k}, y_{1k})(\tilde{z}_k^*) \quad \text{and} \quad (x_{2k}^*, y_{2k}^*) \in \widehat{N}\big((x_{2k}, y_{2k}); \operatorname{gph} S\big)\ .$$

Since f is locally Lipschitzian around $(\bar{x}, \bar{y})$, the sequence (x_{1k}^*, y_{1k}^*) in bounded in $X^* \times Y^*$; hence, by the Asplund property of $X \times Y$, it contains a subsequence weak* converging to some $(x^*, y^*) \in D_N^* f(\bar{x}, \bar{y})(0)$. By $\|(\tilde{x}_k^*, \tilde{y}_k^*)\| \to 0$ and $(x_2^*, y_2^*) = (\tilde{x}_k^*, \tilde{y}_k^*) - (x_{1k}^*, y_{1k}^*)$ one has that $(x_{2k}^*, y_{2k}^*) \stackrel{w^*}{\to} (-x^*, -y^*) \in N\big((\bar{x}, \bar{y}); \operatorname{gph} S\big)$ along a subsequence of $k \to \infty$. By the qualification condition in (b) we get $x^* = y^* = 0$. The latter implies that $(x_{1k}^*, y_{1k}^*, z_k^*) \stackrel{w^*}{\to} 0$ with $(x_{1k}^*, y_{1k}^*) \in \widehat{D}^* f(x_{1k}, y_{1k})(z_k^*)$. Employing now the strong PSNC property of f^{-1} at $(\bar{z}, \bar{x}, \bar{y})$, we conclude that $\|z_k^*\| \to 0$. The last statement in the theorem follows from the scalarization formula of Theorem 3.28. △

Necessary optimality conditions for abstract EPECs obtained in Theorem 5.79 are given in the *normal form* under general constraint qualification. Let us present a corollary of these results providing necessary optimality conditions in the *non-qualified* (Fritz John) form with no qualification conditions imposed on the initial data.

Corollary 5.80 (non-qualified conditions for abstract EPECs). *Let $(\bar{x}, \bar{y})$ be locally (f, Θ)-optimal subject to $y \in S(x)$, where $f: X \times Y \to Z$, $\Theta \subset Z$, and $S: X \rightrightarrows Y$ satisfy the common assumptions of Theorem* 5.79(ii)*. Then there are $0 \ne (x^*, y^*, z^*) \in X^* \times Y^* \times Z^*$ such that the necessary optimality conditions*

$$(x^*, y^*) \in D_N^* f(\bar{x}, \bar{y})(z^*), \quad -x^* \in D_N^* S(\bar{x}, \bar{y})(y^*), \quad z^* \in N(0; \Theta)$$

hold in each of the following cases:

(a) *f is PSNC at $(\bar{x}, \bar{y})$ and Θ is SNC at* 0*;*

(b) *S and Θ are SNC at $(\bar{x}, \bar{y})$ and* 0*, respectively;*

(c) *f is Lipschitz continuous around $(\bar{x}, \bar{y})$ and f^{-1} is strongly PSNC at $(\bar{z}, \bar{x}, \bar{y})$ with $\bar{z} = f(\bar{x}, \bar{y})$.*

Proof. If the qualification conditions in either case (a) or (b) of Theorem 5.79(ii) are fulfilled, then one has the optimality conditions in the corollary with $z^* \ne 0$. The violation of these constraint qualifications directly implies that the desired optimality conditions are satisfied with $(x^*, y^*) \ne 0$. △

Our next step is to derive necessary optimality conditions for multiobjective problems with equilibrium constraints governed by parameter-dependent generalized equations/variational systems. They correspond to the above abstract framework with

$$S(x) := \left\{ y \in Y \middle| \, 0 \in q(x,y) + Q(x,y) \right\} . \tag{5.101}$$

To derive optimality conditions for EPECs with equilibrium/variational constraints of type (5.101), one needs to apply the results of Theorem 5.79 and Corollary 5.80 to the mapping $S(\cdot)$ given in (5.101). For simplicity we present below only those optimality conditions for such problems that don't require constraint qualifications, i.e., correspond to Corollary 5.80. Optimality conditions of normal form can be derived via Theorem 5.79 similarly to lower subdifferential conditions for MPECs in Subsect. 5.2.2.

Theorem 5.81 (generalized order optimality for EPECs governed by variational systems). *Let $f\colon X \times Y \to Z$ be a mapping between Asplund spaces with $\bar{z} := f(\bar{x},\bar{y})$, let $\Theta \subset Z$ with $0 \in Z$, and let $(\bar{x},\bar{y})$ be locally (f,Θ)-optimal subject to the constraints*

$$0 \in q(x,y) + Q(x,y) ,$$

where $q\colon X \times Y \to P$ and $Q\colon X \times Y \rightrightarrows P$ are mappings into an Asplund space P with $\bar{p} := -q(\bar{x},\bar{y})$. Assume that f and q are continuous around $(\bar{x},\bar{y})$, that Θ is closed around 0, and that Q is closed-graph around $(\bar{x},\bar{y},\bar{p})$. Then there are $(x^*,y^*,z^*,p^*) \in X^* \times Y^* \times Z^* \times P^*$ satisfying the relations $(x^*,y^*,z^*) \neq 0$, $z^* \in N(0;\Theta)$, and

$$(x^*,y^*) \in D_N^* f(\bar{x},\bar{y})(z^*) \bigcap \Big(-D_N^* q(\bar{x},\bar{y})(p^*) - D_N^* Q(\bar{x},\bar{y},\bar{p})(p^*) \Big)$$

in each of the following cases:

(a) *f is PSNC at $(\bar{x},\bar{y})$, Θ and Q are SNC at 0 and $(\bar{x},\bar{y},\bar{p})$, respectively, and one has the qualification condition*

$$\Big[(x^*,y^*) \in D_N^* q(\bar{x},\bar{y})(p^*) \cap \big(-D_N^* Q(\bar{x},\bar{y},\bar{p})(p^*) \big) \Big] \Longrightarrow x^* = y^* = p^* = 0 ,$$

which is equivalent to

$$\Big[0 \in \partial \langle p^*, q \rangle (\bar{x},\bar{y}) + D_N^* Q(\bar{x},\bar{y},\bar{p})(p^*) \Big] \Longrightarrow p^* = 0 \tag{5.102}$$

when q is strictly Lipschitzian at $(\bar{x},\bar{y})$.

(b) *f is PSNC at $(\bar{x},\bar{y})$, Θ is SNC at 0, $\dim P < \infty$, q is Lipschitz continuous around $(\bar{x},\bar{y})$, and (5.102) holds.*

(c) *Θ is SNC at 0, q is PSNC at $(\bar{x},\bar{y})$, and the qualification condition in* (a) *is satisfied.*

(d) *f is Lipschitz continuous around $(\bar{x},\bar{y})$, f^{-1} is strongly PSNC at $(\bar{z},\bar{x},\bar{y})$, Q is SNC at $(\bar{x},\bar{y},\bar{p})$, and the qualification condition in* (a) *holds.*

(e) *f and q are Lipschitz continuous around $(\bar{x},\bar{y})$, $\dim P < \infty$, f^{-1} is strongly PSNC at $(\bar{z},\bar{x},\bar{y})$, and (5.102) is satisfied.*

Furthermore, if f is strictly Lipschitzian at $(\bar{x},\bar{y})$, then the above optimality conditions can be equivalently written as follows: there are $z^* \in N(0;\Theta) \setminus \{0\}$ and $p^* \in P^*$ such that

$$0 \in \partial \langle z^*, f \rangle (\bar{x},\bar{y}) + D_N^* q(\bar{x},\bar{y})(p^*) + D_N^* Q(\bar{x},\bar{y},\bar{p})(p^*) .$$

Proof. Based on the optimality conditions from Corollary 5.80, where S is defined in (5.101), we need to give efficient conditions under which the coderivative $D^*_N S(\bar{x}, \bar{y})$ of S and the SNC property of this mappings can be efficiently expressed in terms of the initial data (q, Q) of (5.101). To proceed, we first use Theorem 4.46 giving the upper coderivative estimate (4.63) for S via the coderivatives of q and Q under the assumptions made therein. Combining these assumptions with those in (a) and (c) of Corollary 5.80, we arrive at the conclusion of the theorem in cases (a), (b), (d), and (e). It remains to consider case (c) in Corollary 5.80, which requires efficient conditions for the SNC property of the mapping S from (5.101). In this case we employ the proof of Theorem 4.59, where it is shown (on the base of Theorem 3.84) that mapping (5.101) is SNC at $(\bar{x}, \bar{y})$ if q is PSNC at this point in addition to the qualification condition in (a) and the SNC property of Q at $(\bar{x}, \bar{y}, \bar{p})$. Combining these assumptions with those (b) of Corollary 5.80, we justify the result of this theorem in case (c). The last statement of the theorem follows as usual from the scalarization formula of Theorem 3.28. △

We can derive many consequences of Theorem 5.81 similarly to our considerations in Sections 4.4 and 5.2. Let us present just some of them related to equilibrium constraints given in *composite subdifferential forms* that are the most interesting for applications. The first result concerns EPECs with equilibrium constraints governed by the so-called *hemivariational inequalities with composite potentials.*

Corollary 5.82 (optimality conditions for EPECs governed by HVIs with composite potentials). *Let $f: X \times Y \to Z$ be a continuous mapping with $\bar{z} := f(\bar{x}, \bar{y})$, let $\Theta \subset Z$ be a closed set with $0 \in \Theta$, and let $(\bar{x}, \bar{y})$ be locally (f, Θ)-optimal subject to the equilibrium constraints*

$$0 \in q(x, y) + \partial(\psi \circ g)(y) \,,$$

where $q: X \times Y \to Y^*$, $g: Y \to W$, and $\psi: W \to \overline{I\!R}$. Suppose that W is Banach, that X and Z are Asplund, that $\dim Y < \infty$, and that the following assumptions hold:

(a) *Either f is PSNC at $(\bar{x}, \bar{y})$ and Θ is SNC at 0, or f is Lipschitz continuous around $(\bar{x}, \bar{y})$ and f^{-1} is strongly PSNC at $(\bar{z}, \bar{x}, \bar{y})$.*

(b) *q is strictly differentiable at $(\bar{x}, \bar{y})$ with the surjective partial derivative $\nabla_x q(\bar{x}, \bar{y})$.*

(c) *g is continuously differentiable around $\bar{y}$ with the surjective derivative $\nabla g(\bar{y})$, and the mapping $\nabla g(\cdot)$ from Y to the space of linear bounded operators from Y to W is strictly differentiable at $\bar{y}$.*

(d) *The graph of $\partial\psi$ is locally closed around $(\bar{w}, \bar{v})$, where $\bar{w} := g(\bar{y})$ and where $\bar{v} \in W^*$ is a unique functional satisfying the relations*

$$-q(\bar{x}, \bar{y}) = \nabla g(\bar{y})^* \bar{v}, \quad \bar{v} \in \partial\psi(\bar{w}) \,.$$

Then there are $(y^*, z^*, u) \in Y^* \times Z^* \times Y$ *such that* $(y^*, z^*) \neq 0$, $z^* \in N(0; \Theta)$, $\big(-\nabla_x q(\bar{x}, \bar{y})^* u, y^*\big) \in D^*_N f(\bar{x}, \bar{y})(z^*)$, *and*

$$-y^* \in \nabla_y q(\bar{x}, \bar{y})^* u + \nabla^2 \langle \bar{v}, g\rangle(\bar{y})^* u + \nabla g(\bar{y})^* \partial^2_N \psi(\bar{w}, \bar{v})\big(\nabla g(\bar{y}) u\big) \; .$$

Furthermore, if f *is strictly Lipschitzian at* $(\bar{x}, \bar{y})$, *then the above optimality conditions can be written as follows: there are* $z^* \in N(0; \Theta) \setminus \{0\}$ *and* $u \in Y$ *such that*

$$0 \in \partial\langle z^*, f\rangle(\bar{x}, \bar{y}) + \nabla q(\bar{x}, \bar{y})^* u$$

$$+\Big(0, \nabla^2 \langle \bar{v}, g\rangle(\bar{y})^* u + \nabla g(\bar{y})^* \partial^2_N \psi(\bar{w}, \bar{v})\big(\nabla g(\bar{y}) u\big)\Big) \; .$$

Proof. This follows from Theorem 5.81 with $Q(y) = \partial(\psi \circ g)(y)$ by computing

$$D^* Q(\bar{y}, \bar{p})(u) = \partial^2(\psi \circ g)(\bar{y}, \bar{p})(u) \;\text{ with }\; \bar{p} := -q(\bar{x}, \bar{y})$$

using the second-order subdifferential chain rule from Theorem 1.127. Observe that Q is SNC by $\dim Y < \infty$ and that the qualification condition in Theorem 5.81 holds automatically, since $\nabla_x q(\bar{x}, \bar{y})$ is surjective and Q doesn't depend on the parameter x. △

The next corollary provides necessary optimality conditions for EPECs, where equilibrium constraints are given by *parameter-dependent* variational systems (labeled as *generalized variational inequalities*–GVIs) with composite potentials. For brevity and simplicity we consider only the case of *amenable* potentials in finite dimensions. Note that no surjectivity assumptions on derivatives are imposed.

Corollary 5.83 (generalized order optimality for EPECs governed by GVIs with amenable potentials). *Let* $f\colon X \times Y \to Z$ *be a continuous mapping, let* $\Theta \subset Z$ *be a closed set with* $0 \in \Theta$, *and let* $(\bar{x}, \bar{y})$ *be locally* (f, Θ)*-optimal subject to the parameter-dependent equilibrium constraints*

$$0 \in q(x, y) + \partial(\psi \circ g)(x, y) \; ,$$

where $q\colon X \times Y \to X^* \times Y^*$, $g\colon X \times Y \to W$, *and* $\psi\colon W \to \overline{I\!R}$, $\dim(X \times Y \times W) < \infty$, *and* Z *is Asplund. Assume that* q *is Lipschitz continuous around* $(\bar{x}, \bar{y})$ *and the potential* $\varphi := \psi \circ g$ *is strongly amenable at this point. Denote* $\bar{p} := -q(\bar{x}, \bar{y}) \in \partial(\psi \circ g)$, $\bar{w} := g(\bar{x}, \bar{y})$,

$$M(\bar{x}, \bar{y}) := \Big\{\bar{v} \in W^* \Big|\; \bar{v} \in \partial\psi(\bar{w}), \quad \nabla g(\bar{x}, \bar{y})^* \bar{v} = \bar{p}\Big\}$$

and impose the second-order qualification conditions:

$$\partial^2 \psi(\bar{w}, \bar{v})(0) \cap \ker \nabla g(\bar{x}, \bar{y})^* = \{0\} \;\text{ for all }\; \bar{v} \in M(\bar{x}, \bar{y}) \;\text{ and}$$

$$\Big[0 \in \partial\langle u, q\rangle(\bar{x}, \bar{y}) + \bigcup_{\bar{v} \in M(\bar{x}, \bar{y})} \Big[\nabla^2\langle\bar{v}, g\rangle(\bar{x}, \bar{y})(u) + \nabla g(\bar{x}, \bar{y})^* \partial^2 \psi(\bar{w}, \bar{v})\big(\nabla g(\bar{x}, \bar{y})u\big)\Big] \Longrightarrow u = 0 \, .$$

Then there are $0 \neq (x^*, y^*, z^*)$ *with* $z^* \in N(0; \Theta)$ *satisfying the relations* $(-x^*, -y^*) \in D^*_N f(\bar{x}, \bar{y})(z^*)$ *and*

$$(x^*, y^*) \in \partial\langle u, q\rangle(\bar{x}, \bar{y}) + \bigcup_{\bar{v} \in M(\bar{x}, \bar{y})} \Big[\nabla^2\langle\bar{v}, g\rangle(\bar{x}, \bar{y})(u) + \nabla g(\bar{x}, \bar{y})^* \partial^2 \psi(\bar{w}, \bar{v})\big(\nabla g(\bar{x}, \bar{y})u\big)\Big]$$

with some $u \in X \times Y$ *in each of the following cases:*

(a) f *is PSNC at* $(\bar{x}, \bar{y})$ *and* Θ *is SNC at* 0*;*

(b) f *is Lipschitz continuous around* $(\bar{x}, \bar{y})$ *and* f^{-1} *is strongly PSNC at* $(\bar{z}, \bar{x}, \bar{y})$*, where* $\bar{z} := f(\bar{x}, \bar{y})$.

Furthermore, these optimality conditions are equivalent to the existence of $z^* \in N(0; \Theta) \setminus \{0\}$ *and* $u \in X \times Y$ *satisfying*

$$0 \in \partial\langle z^*, f\rangle(\bar{x}, \bar{y}) + \partial\langle u, q\rangle(\bar{x}, \bar{y}) + \bigcup_{\bar{v} \in M(\bar{x}, \bar{y})} \Big[\nabla^2\langle\bar{v}, g\rangle(\bar{x}, \bar{y})(u) + \nabla g(\bar{x}, \bar{y})^* \partial^2 \psi(\bar{w}, \bar{v})\big(\nabla g(\bar{x}, \bar{y})u\big)\Big]$$

when f *is strictly Lipschitzian at* $(\bar{x}, \bar{y})$.

Proof. This follows from Theorem 5.81 with $Q(x, y) = \partial(\psi \circ g)(x, y)$ by applying the second-order subdifferential chain rule for amenable functions derived in Corollary 3.76. △

The last corollary of Theorem 5.81 concerns EPECs involving equilibrium/variational constraints governed by parametric generalized equations with *composite fields.* Constraints of this type may be considered in full generality similarly to MPECs in Sect. 2.2. For simplicity we present necessary optimalityconditions only for a special class of such EPECs under some smoothness assumptions.

Corollary 5.84 (optimality conditions for EPECs with composite fields). *Let* $(\bar{x}, \bar{y})$ *be locally* (f, Θ)*-optimal subject to*

$$0 \in q(x, y) + (\partial\psi \circ g)(x, y) \, ,$$

where $f\colon X \times Y \to Z$ *and* $\Theta \subset Z$ *are the same as in the previous corollary while* $g\colon X \times Y \to W$, $\psi\colon W \to \overline{I\!R}$, *and* $q\colon X \times Y \to W^*$. *Suppose that* X *and*

Y are Asplund while $\dim W < \infty$*, that both q and g are strictly differentiable at $(\bar{x}, \bar{y})$, and that* $\operatorname{gph} \partial \psi$ *is locally closed around $(\bar{w}, \bar{p})$ with $\bar{w} = g(\bar{x}, \bar{y})$ and $\bar{p} = -q(\bar{x}, \bar{y})$; the latter is automatic for continuous and for amenable functions. Assume also the qualification conditions*

$$\partial^2 \psi(\bar{w}, \bar{p})(0) \cap \ker \nabla g(\bar{x}, \bar{y})^* = \{0\} \quad \textit{and}$$

$$\Big[0 \in \nabla q(\bar{x}, \bar{y})^* u + \nabla g(\bar{x}, \bar{y})^* \partial^2 \psi(\bar{w}, \bar{p})(u)\Big] \Longrightarrow u = 0 \, .$$

Then there are $0 \neq (x^*, y^*, z^*)$ with $z^* \in N(0; \Theta)$ satisfying

$$(x^*, y^*) \in D_N^* f(\bar{x}, \bar{y})(z^*) \bigcap \Big[- \nabla q(\bar{x}, \bar{y})^* u + \nabla g(\bar{x}, \bar{y})^* \partial^2 \psi(\bar{w}, \bar{p})(u)\Big]$$

for some $u \in X \times Y$ in each of the cases (a) *and* (b) *of the previous corollary.*

Furthermore, these optimality conditions are equivalent to

$$0 \in \partial \langle z^*, f \rangle(\bar{x}, \bar{y}) + \nabla q(\bar{x}, \bar{y})^* u + \nabla g(\bar{x}, \bar{y})^* \partial^2 \psi(\bar{w}, \bar{p})(u)$$

with $z^* \in N(0; \Theta) \setminus \{0\}$ when f is strictly Lipschitzian at $(\bar{x}, \bar{y})$.

Proof. This follows from Theorem 5.81 with $Q(x, y) = (\partial \psi \circ g)(x, y)$ and the upper estimate for the coderivative $D^* Q(\bar{x}, \bar{y}, \bar{p})$ derived in Theorem 4.54 under the assumptions made. △

The results obtained directly imply necessary optimality conditions for EPECs with specific types of equilibria, as well as for *minimax problems with equilibrium constraints*, as discussed in Subsects. 5.3.1 and 5.3.2.

Next we derive some results for EPECs with respect to *closed preferences* that are similar to but generally independent of those given above. As before, we present only pointbased/exact optimality conditions for the problems under consideration. Let us start with optimality conditions for EPECs with *abstract* equilibrium constraints governed by general set-valued mappings.

Proposition 5.85 (optimality conditions for abstract EPECs with closed preferences). *Let $(\bar{x}, \bar{y})$ be a local optimal solution to the multiobjective optimization problem:*

$$\textit{minimize} \;\; f(x, y) \;\; \textit{with respect to} \;\; \prec \;\; \textit{subject to} \;\; y \in S(x) \, ,$$

where $f: X \times Y \to Z$ is a mapping between Asplund spaces that is continuous around $(\bar{x}, \bar{y})$ with $\bar{z} := f(\bar{x}, \bar{y})$, where the preference $\prec$ is closed on Z with the level set $\mathcal{L}(\cdot)$, and where $S: X \rightrightarrows Y$ is closed-graph around $(\bar{x}, \bar{y})$. Assume that either f is SNC at $(\bar{x}, \bar{y})$, or S is SNC at this point and $\operatorname{cl} \mathcal{L}: Z \rightrightarrows Z$ *is ISNC at $(\bar{z}, \bar{z})$. Then there are $0 \neq (x^*, y^*, z^*) \in X^* \times Y^* \times Z^*$ satisfying*

$$(x^*, y^*) \in D_N^* f(\bar{x}, \bar{y})(z^*), \quad -x^* \in D_N^* S(\bar{x}, \bar{y})(y^*), \quad \textit{and} \;\; z^* \in N_+\big(\bar{z}; \operatorname{cl} \mathcal{L}(\bar{z})\big) .$$

Furthermore, one has

$$0 \in \partial\langle z^*, f\rangle(\bar{x}, \bar{y}) + N\big((\bar{x}, \bar{y}); \operatorname{gph} S\big) \;\; \textit{with} \;\; z^* \in N_+\big(\bar{z}; \operatorname{cl} \mathcal{L}(\bar{z})\big) \setminus \{0\}$$

provided that f *is strictly Lipschitzian at* $(\bar{x}, \bar{y})$ *and either* $\dim Z < \infty$*, or* S *is SNC at* $(\bar{x}, \bar{y})$ *and* $\operatorname{cl} \mathcal{L}$ *is ISNC at* $(\bar{z}, \bar{z})$*.*

Proof. This follows directly from Theorem 5.73(ii) with the constraint set $\Omega := \operatorname{gph} S$ in the Asplund space $X \times Y$. △

Now we are ready to derive necessary optimality conditions for EPECs involving closed preference relations and equilibrium constraints governed by parametric variational systems/generalized equations (5.101).

Theorem 5.86 (optimality conditions for EPECs with closed preferences and variational constraints). *Let* $(\bar{x}, \bar{y})$ *be a local optimal solution to the multiobjective optimization problem:*

$$\textit{minimize} \;\; f(x, y) \;\; \textit{with respect to} \;\; \prec \;\; \textit{subject to} \;\; 0 \in q(x, y) + Q(x, y) ,$$

where $f\colon X \times Y \to Z$*,* $q\colon X \times Y \to P$ *and* $Q\colon X \times Y \rightrightarrows P$ *are mappings between Asplund spaces, and where* $\prec$ *is a closed preference relation on* Z*. Suppose that* f *and* q *are continuous around* $(\bar{x}, \bar{y})$*, and that* Q *is closed-graph around* $(\bar{x}, \bar{y}, \bar{p})$ *with* $\bar{p} := -q(\bar{x}, \bar{y}) \in Q(\bar{x}, \bar{y})$*. Then there are* $(x^*, y^*, z^*, p^*) \in X^* \times Y^* \times Z^* \times P^*$ *satisfying the relations*

$$(x^*, y^*, z^*) \neq 0, \quad z^* \in N_+\big(\bar{z}; \operatorname{cl} \mathcal{L}(\bar{z})\big) \;\; \textit{with} \;\; \bar{z} := f(\bar{x}, \bar{y}), \quad \textit{and}$$

$$(x^*, y^*) \in D_N^* f(\bar{x}, \bar{y})(z^*) \bigcap \Big(- D_N^* q(\bar{x}, \bar{y})(p^*) - D_N^* Q(\bar{x}, \bar{y}, \bar{p})(p^*)\Big)$$

in each of the following cases:

(a) f *is SNC at* $(\bar{x}, \bar{y})$*,* Q *is SNC at* $(\bar{x}, \bar{y}, \bar{p})$*, and the qualification condition*

$$\Big[(x^*, y^*) \in D_N^* q(\bar{x}, \bar{y})(p^*) \cap \big(- D_N^* Q(\bar{x}, \bar{y}, \bar{p})(p^*)\big)\Big] \Longrightarrow x^* = y^* = p^* = 0$$

holds, which is equivalent to (5.102) *when* q *is strictly Lipschitzian at* $(\bar{x}, \bar{y})$*.*

(b) f *is SNC at* $(\bar{x}, \bar{y})$*,* $\dim P < \infty$*,* q *is Lipschitz continuous around* $(\bar{x}, \bar{y})$*, and* (5.102) *is satisfied.*

(c) $\operatorname{cl} \mathcal{L}$ *is ISNC at* $(\bar{z}, \bar{z})$*,* g *is PSNC at* $(\bar{x}, \bar{y})$*, and the qualification condition in* (a) *holds.*

Furthermore, for f *strictly Lipschitzian at* $(\bar{x}, \bar{y})$ *the above optimality conditions can be equivalently written as follows: there is a nonzero element* $z^* \in N_+\big(\bar{z}; \operatorname{cl} \mathcal{L}(\bar{z})\big)$ *satisfying*

$$0 \in \partial\langle z^*, f\rangle(\bar{x}, \bar{y}) + D_N^* q(\bar{x}, \bar{y})(p^*) + D_N^* Q(\bar{x}, \bar{y}, \bar{p})(p^*)$$

with some $p^* \in P^*$. *In this case the SNC assumption on* f *in* (a) *and* (b) *implies that* $\dim Z < \infty$.

Proof. Apply Proposition 5.85 with S given in (5.101). To proceed, we need to use efficient conditions ensuring an upper estimate of the coderivative $D_N^* S(\bar{x}, \bar{y})$ and the SNC property of S at $(\bar{x}, \bar{y})$ in terms of the initial data (q, Q) in (5.101). It can be done similarly to the proof of Theorem 5.81 based on the corresponding results of Sect. 4.4. △

Similarly to the above setting of generalized order optimality we can derive from Theorem 5.86 the corresponding counterparts of Corollaries 5.82, 5.83, and 5.84 that give necessary optimality conditions for EPECs with closed preference relations and equilibrium constraints governed by the composite variational systems considered above.

5.4 Subextremality and Suboptimality at Linear Rate

This section is devoted to the study of *less restrictive* concepts of set extremality and of (sub)optimal solutions to standard minimization problems as well as multiobjective optimization problems than the ones considered before. It happens that the necessary extremality and optimality conditions obtained above for the conventional notions are *necessary and sufficient* for the new notions studied in the section.

The main difference between the conventional notions and those introduced and studied below is that the latter relate to extremality/optimality not at the point in question but in a *neighborhood* of it, and that they involve a *linear rate* in the sense precisely defined in what follows. To some extent, this is similar to the *linear rate of openness* that distinguishes the covering properties described in Definition 1.51 from general openness properties in the framework of the classical open mapping theorems. We also mention the relationship between general continuity and *Lipschitz* continuity properties; the latter actually mean "continuity at a linear rate." It happens that, as in the case of covering and Lipschitzian properties admitting *complete dual characterizations*, similar characterizing results hold for properly defined extremality and optimality notions with a linear rate. The main goal of this section is to realize these proper definitions, to clarify their specific features, and to justify the corresponding necessary and sufficient extremality/optimality conditions.

We start with set extremality first defining the notion of *linear subextremality* (or subextremality at a linear rate) for systems of sets and showing that such systems are *fully characterized* by the generalized Euler equations of the *extremal principle*, in both approximate and exact forms. Then we consider *linear suboptimality* for constrained multiobjective optimization problems and obtain *necessary and sufficient conditions* for this concept via coderivatives. The final part of this section is devoted to *characterizing linear subminimality* of lower semicontinuous functions in terms of their subdifferentials and to the subsequent derivation of necessary and sufficient conditions for linear subminimality in constrained problems. We illustrate by striking examples essential differences between the standard minimality and linear subminimality notions for real-valued functions. Note that for strictly differentiable functions the linear subminimality reduces to the classical *stationarity* in the sense of vanishing the strict derivative at the reference point.

5.4.1 Linear Subextremality of Set Systems

Given two subsets Ω_1 and Ω_2 of a Banach space X, we consider the constant

$$\vartheta(\Omega_1, \Omega_2) := \sup\left\{\nu \geq 0 \middle|\, \nu I\!B \subset \Omega_1 - \Omega_2\right\} \tag{5.103}$$

describing the *measure of overlapping* for these sets. Note that one has $\vartheta(\Omega_1, \Omega_2) = -\infty$ in (5.103) if $\Omega_1 \cap \Omega_2 = \emptyset$. It is easy to observe that a point $\bar{x} \in \Omega_1 \cap \Omega_2$ is *locally extremal* for the set system $\{\Omega_1, \Omega_2\}$ in the sense of Definition 2.1 *if and only if*

$$\vartheta\big(\Omega_1 \cap B_r(\bar{x}), \Omega_2 \cap B_r(\bar{x})\big) = 0 \quad \text{for some} \quad r > 0\,, \tag{5.104}$$

where $B_r(\bar{x}) := \bar{x} + rI\!B$ as usual. Modifying the constant $\vartheta(\cdot,\cdot)$ in (5.104), we come up to the following notion of *linear subextremality* for systems of two sets in Banach spaces.

Definition 5.87 (linear subextremality for two sets). *Given $\Omega_1, \Omega_2 \subset X$ and $\bar{x} \in \Omega_1 \cap \Omega_2$, we say that the set system $\{\Omega_1, \Omega_2\}$ is* LINEARLY SUBEXTREMAL *around the point $\bar{x}$ if $\vartheta_{\rm lin}(\Omega_1, \Omega_2, \bar{x}) = 0$, where*

$$\vartheta_{\rm lin}(\Omega_1, \Omega_2, \bar{x}) := \liminf_{\substack{x_i \xrightarrow{\Omega_i} \bar{x} \\ r \downarrow 0}} \frac{\vartheta\big([\Omega_1 - x_1] \cap rI\!B, [\Omega_2 - x_2] \cap rI\!B\big)}{r} \tag{5.105}$$

with $i = 1, 2$, and where the measure of overlapping $\vartheta(\cdot,\cdot)$ is defined in (5.103).

It is clear that the set extremality in the sense of (5.104) implies the linear subextremality in the sense of (5.105), but not vice versa. Let us discuss some specific features of linear subextremality for set systems that distinguish this notion from the concept of (5.104):

(a) The constant $\vartheta_{\rm lin}(\Omega_1, \Omega_2, \bar{x})$ defined in (5.105), in contrast to the one $\vartheta\big(\Omega_1 \cap B_r(\bar{x}), \Omega_2 \cap B_r(\bar{x})\big)$ from (5.103), involves a *linear rate* of set perturbations as $r \downarrow 0$. Therefore condition (5.105) describes a *local nonoverlapping at* *linear rate* for the sets Ω_1 and Ω_2, while the condition in (5.104) corresponds to a local nonoverlapping of these sets with an *arbitrary rate* as $r \downarrow 0$,

(b) Condition (5.105) requires not the *precise* local nonoverlapping of the given sets but up to their *infinitesimally small deformations*. It follows from the representation

$$\vartheta_{\rm lin}(\Omega_1, \Omega_2, \bar{x}) = \liminf_{x_i \xrightarrow{\Omega_i} \bar{x}} \vartheta_{\rm lin}(\Omega_1 - x_1, \Omega_2 - x_2), \quad \text{where} \quad i = 1, 2 \quad \text{and}$$

$$\vartheta_{\rm lin}(\Omega_1, \Omega_2) := \liminf_{r \downarrow 0} \frac{\vartheta(\Omega_1 \cap r I\!\!B, \Omega_2 \cap r I\!\!B)}{r}$$

with $\vartheta(\cdot, \cdot)$ defined in (5.103).

(c) Condition (5.105) doesn't require that the sets Ω_1 and Ω_2 nonoverlap *exactly at* the point $\bar{x}$. One can see from the relations in (b) that (5.105) holds if, given any neighborhood U of $\bar{x}$, there are points $x_1 \in \Omega_1 \cap U$ and $x_2 \in \Omega_2 \cap U$ ensuring an *approximate nonoverlapping* of the translated sets $\Omega_1 - x_1$ and $\Omega_2 - x_2$ with a linear rate.

We have proved in Theorem 2.20 that, for arbitrary Asplund spaces, the relations of the extremal principle in the approximate form of Definition 2.5 provide necessary conditions for the local set extremality in the sense of Definition 2.1 equivalently described in (5.104). It happens in fact that these relations are *necessary and sufficient* for the linear set subextremality defined above. The exact statements are given in the next theorem.

Theorem 5.88 (characterization of linear subextremality via the approximate extremal principle). *Let Ω_1 and Ω_2 be subsets of a Banach space X, and let $\bar{x} \in \Omega_1 \cap \Omega_2$. The following assertions hold:*

(i) *Assume that for every positive ε there are $\widehat{x}_i \in \Omega_i \cap (\bar{x} + \varepsilon I\!\!B)$ and $x_i^* \in \widehat{N}_\varepsilon(\widehat{x}_i; \Omega_i)$ for $i = 1, 2$ such that*

$$\|x_1^* + x_2^*\| \le \varepsilon \quad \text{and} \quad \|x_1^*\| + \|x_2^*\| = 1\,. \tag{5.106}$$

Then $\{\Omega_1, \Omega_2\}$ is linearly subextremal around $\bar{x}$.

(ii) *Conversely, assume that both sets Ω_i are locally closed and that the system $\{\Omega_1, \Omega_2\}$ is linearly subextremal around $\bar{x}$. Then for every $\varepsilon > 0$ there are $\widehat{x}_i \in \Omega_i \cap (\bar{x} + \varepsilon I\!\!B)$ and $x_i^* \in \widehat{N}(\widehat{x}_i; \Omega_i)$, $i = 1, 2$, satisfying* (5.106) *provided that X is Asplund. Moreover, if the latter property holds for any linearly subextremal system $\{\Omega_1, \Omega_2\} \subset X$ around some point $\bar{x} \in \Omega_1 \cap \Omega_2$, than the space X must be Asplund.*

Proof. To prove (i), we suppose that $\{\Omega_1, \Omega_2\}$ is *not* linearly subextremal around $\bar{x}$, i.e., one has

$$\vartheta_{\rm lin}(\Omega_1, \Omega_2, \bar{x}) =: \alpha > 0$$

for the constant $\vartheta_{\rm lin}$ in (5.105). The latter means that there is $\bar{r} > 0$ such that

$$\vartheta\big([\Omega_1 - x_1] \cap r I\!B, [\Omega_2 - x_2] \cap r I\!B\big) > (\alpha r)/2 \tag{5.107}$$

for any positive $r \le \bar{r}$ and every $x_i \in \Omega_i \cap r I\!B$, $i = 1, 2$, where $\vartheta(\cdot, \cdot)$ is defined in (5.103). On the other hand, it follows from the conditions assumed in (i) with $\varepsilon := \min\{\alpha/16, 1/4\}$ and from the *very definition* (1.1) of ε-normals, which actually fits well the subextremality at a linear rate, that there is a positive number $r < \bar{r}$ such that

$$\langle x_i^*, x\rangle \le \tfrac{\alpha}{32}\|x\| \quad \text{whenever} \quad x \in \big[\Omega_i - x_i\big] \cap r I\!B, \quad i = 1, 2 \; .$$

Since $x_2^* = -x_1^* + (x_1^* + x_2^*)$, one has

$$-\langle x_1^*, x\rangle \le \big(\tfrac{\alpha}{32} + \varepsilon\big)\|x\| \le \tfrac{3\alpha}{32}\|x\| \quad \text{for all} \quad x \in \big[\Omega_2 - x_2\big] \cap r I\!B \quad \textit{and}$$

$$\langle x_1^*, x\rangle \le (\alpha r)/8 \quad \text{for all} \quad x \in \big[(\Omega_1 - x_1) \cap r I\!B\big] - \big[(\Omega_2 - x_2) \cap r I\!B\big] \; .$$

Now it follows from (5.107) and the first relations in (5.106) that

$$\|x_1^*\| \le 1/4 \quad \text{and} \quad \|x_2^*\| \le \|x_1^*\| + \varepsilon \le 1/2 \; ,$$

which contradicts the second relations in (5.106) and justifies assertion (i).

Next let us justify assertion (ii) of the theorem following the procedure in the proofs of Lemma 2.32(ii) and Theorem 2.51(i) related to establishing necessary conditions for set extremality. It happens that the same ideas work for the more general notion of set subextremality at a linear rate.

Let $\{\Omega_1, \Omega_2\}$ be linearly subextremal around $\bar{x}$, i.e., (5.105) holds. Given $\varepsilon \in (0, 1)$, we find $x_i \in \Omega_i \cap (\varepsilon/2) I\!B$ for $i = 1, 2$ and $0 < r < \varepsilon$ such that

$$\vartheta\big([\Omega_1 - x_1] \cap r I\!B, [\Omega_2 - x_2] \cap r I\!B\big) < (r\varepsilon)/8 \; .$$

This implies, by definition (5.103) of the overlapping constant $\vartheta(\cdot, \cdot)$, the existence of $a \in (r\varepsilon/8) I\!B$ satisfying

$$a \notin \big([\Omega_1 - x_1] \cap r I\!B\big) - \big([\Omega_2 - x_2] \cap r I\!B\big) \; .$$

Therefore one has

$$\|u - x_1 - v + x_2 - a\| > 0 \quad \text{if} \quad u \in \Omega_1 \cap (x_1 + r I\!B), \quad v \in \Omega_2 \cap (x_2 + r I\!B) \; .$$

Since X is assumed to be Asplund, the product space $X \times X$ is Asplund as well; for convenience we equipped it with the maximum norm $\|(u, v)\| := \max\{\|u\|, \|v\|\}$. Define a real-valued function on $X \times X$ by

$$\varphi(u, v) := \|u - x_1 - v + x_2\|, \quad (u, v) \in X \times X \; ,$$

and observe from the above that

$$\varphi(u,v) > 0 \text{ for } (u,v) \in \Omega := \big[\Omega_1 \cap (x_1 + r I\!B)\big] \times \big[\Omega_2 \cap (x_2 + r I\!B)\big]$$

$$\text{with} \quad \varphi(x_1,x_2) = \|a\| \le (r\varepsilon)/8 \ .$$

It follows from Ekeland's variational principle (Theorem 2.26) that there are $\bar{u} \in \Omega_1 \cap (x_1 + (r/4)I\!B)$ and $\bar{v} \in \Omega_2 \cap (x_2 + (r/4)I\!B)$ such that $(\bar{u},\bar{v})$ is the minimum point to the extended-real-valued function

$$\varphi(u,v) + \tfrac{\varepsilon}{2}\|(u,v) - (\bar{u},\bar{v})\| + \delta((u,v);\Omega), \quad (u,v) \in X \times X \ .$$

Applying now the subgradient description of the approximate extremal principle given in Lemma 2.32(i) in any Asplund space and taking into account that the first two terms in the above sum are convex, we find

$$(y_1,y_2) \in (\bar{u},\bar{v}) + \tfrac{r}{4}I\!B \subset (x_1,x_2) + \tfrac{r}{2}I\!B, \quad (z_1,z_2) \in \Omega \cap \big[(\bar{u},\bar{v}) + \tfrac{r}{4}I\!B\big] \ ,$$

and (x^*_{1j}, x^*_{2j}), $j = 1,2,3$, satisfying

$$(x^*_{11}, x^*_{21}) \in \widehat{\partial}\varphi(y_1,y_2), \quad \|(x^*_{12}, x^*_{22})\| \le \varepsilon/2 \ ,$$

$$x^*_{13} \in \widehat{N}(z_1;\Omega_1), \quad x^*_{23} \in \widehat{N}(z_2;\Omega_2), \quad \text{and}$$

$$\|(x^*_{11}, x^*_{21}) + (x^*_{12}, x^*_{22}) + (x^*_{13}, x^*_{23})\| \le \varepsilon/2 \ .$$

Moreover, $x^*_{11} = -x^*_{21} =: x^*$, where x^* is a subgradient of the norm calculated at the nonzero point $y_1 - x_1 - y_2 + x_2 - a$. Thus $\|x^*\| = 1$ and

$$\|x^*_{13} + x^*_{23}\| \le \|x^*_{13} + x^*\| + \|x^*_{23} - x^*\| = \|(x^*_{11}, x^*_{21}) + (x^*_{13}, x^*_{23})\| \le \varepsilon \ ,$$

$$2 - \varepsilon \le \|x^*_{13}\| + \|x^*_{23}\| \le 2 + \varepsilon \ .$$

Denote $\widehat{x}_1 := x_{13}$, $\widehat{x}_2 := x_{23}$,

$$x^*_1 := x^*_{13}\big/(\|x^*_{13}\| + \|x^*_{23}\|), \quad \text{and} \quad x^*_2 := x^*_{23}\big/(\|x^*_{13}\| + \|x^*_{23}\|) \ .$$

Then one has $x^*_i \in \widehat{N}(\widehat{x}_i;\Omega_i)$, $\widehat{x}_i \in \Omega_i \cap (\bar{x} + \varepsilon I\!B)$ for $i = 1,2$, and

$$\|x^*_1\| + \|x^*_2\| = 1, \quad \|x^*_1 + x^*_2\| \le \varepsilon\big/(2 - \varepsilon) \le \varepsilon \ ,$$

which gives all the relations of the approximate extremal principle for linearly subextremal systems in Asplund spaces. The last statement of the theorem follows from implication (b)⇒(a) in Theorem 2.20. △

The next result, which is a consequence of Theorem 5.88, characterizes the linear suboptimality of set systems via the relations of the *exact extremal principle* under additional assumptions.

Theorem 5.89 (characterization of linear subextremality via the exact extremal principle). *Let the system $\{\Omega_1, \Omega_2\} \subset X$ be linearly subextremal around $\bar{x} \in \Omega_1 \cap \Omega_2$. Assume that X is Asplund, that the sets Ω_1 and Ω_2 are locally closed around $\bar{x}$, and that one of them is SNC at this point. Then there is $x^* \in X^*$ satisfying*

$$0 \neq x^* \in N(\bar{x}; \Omega_1) \cap \big(- N(\bar{x}; \Omega_2) \big) \; . \tag{5.108}$$

Furthermore, condition (5.108) *is necessary and sufficient for the linear subextremality of $\{\Omega_1, \Omega_2\}$ around $\bar{x}$ if* $\dim X < \infty$.

Proof. Let us justify the first statement of the theorem based on assertion (ii) of Theorem 5.88. Picking $\varepsilon_k \downarrow 0$ as $k \to \infty$ and using the latter result, we find sequences $x_{ik} \to \bar{x}$ and $x^*_{ik} \in \widehat{N}(x_{1k}; \Omega_i)$ for $i = 1, 2$ such that

$$\|x^*_{1k} + x^*_{2k}\| \leq \varepsilon_k \quad \text{and} \quad \|x^*_{1k}\| + \|x^*_{2k}\| = 1 \;\; \text{whenever} \;\; k \in I\!\!N \; . \tag{5.109}$$

Since X is Asplund and the sequences $\{x^*_{1k}\}$ and $\{x^*_{2k}\}$ are bounded in X^*, there are subsequences of them that weak* converge to x^*_1 and x^*_2, respectively. It follows from the first relations in (5.109) and the lower semicontinuity of the norm function in the weak* topology of X^* that $x^*_1 = -x^*_2 =: x^*$. Furthermore, $x^* \in N(\bar{x}; \Omega_1) \cap \big(- N(\bar{x}; \Omega_2) \big)$ by the definition of the basic normal cone. It remains to show that $x^* \neq 0$ if one of the sets (say Ω_1) is SNC at $\bar{x}$.

On the contrary, assume that $x^* = 0$. Then $x^*_{1k} \xrightarrow{w^*} 0$, and hence $\|x^*_{1k}\| \to 0$ by the SNC property of Ω_1 at $\bar{x}$. It follows from the first relation in (5.109) that $\|x^*_{2k}\| \to 0$ as well. This obviously contradicts the second relation in (5.109) and finishes the proof of (5.108) for linearly subextremal systems of closed sets in Asplund spaces.

Assume now that (5.108) holds for $\{\Omega_1, \Omega_2, \bar{x}\}$ with $\|x^*\| = 1$ while X is finite-dimensional. Using representation (1.8) of the basic normal cone in finite dimensions, we find sequences $x_{ik} \xrightarrow{\Omega_i} \bar{x}$, $x^*_{1k} \to x^*$, and $x^*_{2k} \to -x^*$ such that $x^*_{ik} \in \widehat{N}(x_{ik}; \Omega_i)$ for $i = 1, 2$ and all $k \in I\!\!N$. Since $x^*_{1k} + x^*_{2k} \to 0$ and $\|x^*_{1k}\| + \|x^*_{2k}\| \to 2\|x^*\| = 2$ as $k \to \infty$, one concludes by the standard normalization that for every $\varepsilon > 0$ there are $x_i \in \Omega_i \cap (\bar{x} + \varepsilon I\!\!B)$ and $x^*_i \in \widehat{N}(x_i; \Omega_i)$, $i = 1, 2$, satisfying (5.106). Thus $\{\Omega_1, \Omega_2\}$ is linearly subextremal around $\bar{x}$ by assertion (i) of Theorem 5.88. This completes the proof of the theorem. $\triangle$

Note that the above proof of the second part of Theorem 5.89 essentially employs the finite dimensionality of the space X ensuring the agreement between the norm and weak* topology on X^*; cf. the fundamental Josefson-Nissenzweig theorem discussed, e.g., in Subsect. 1.1.3. On the other hand, the latter assumption can be relaxed for sets Ω_i of special functional structures; see the next two subsections for more details.

Remark 5.90 (linear subextremality for many sets). The above definition of linear set subextremality concerns the case of two sets. Given a system of *finitely many* sets $\{\Omega_1, \ldots, \Omega_n\}$, $n \geq 2$, in a Banach space X, we define its linear subextremality in the following way: $\{\Omega_1, \ldots, \Omega_2\}$ is *linearly subextremal* around $\bar{x} \in \Omega_1 \cap \ldots \cap \Omega_n$ if the system of two sets

$$\widetilde{\Omega}_1 := \prod_{i=1}^{n} \Omega_i \quad \text{and} \quad \widetilde{\Omega}_2 := \big\{(x, \ldots, x) \in X^n \big|\, x \in X\big\}$$

is linearly subextremal around $(\bar{x}, \ldots, \bar{x}) \in X^n$ in the sense of Definition 5.87. This is equivalent to say that, given any $j \in \{1, \ldots, n\}$, the system of two sets

$$\overline{\Omega}_1 := \prod_{i \in \{1,\ldots,n\} \setminus j} \Omega_i \quad \text{and} \quad \overline{\Omega}_2 := \big\{(x, \ldots, x) \in X^{n-1} \big|\, x \in \Omega_j\big\}$$

is linearly subextremal around $(\bar{x}, \ldots, \bar{x}) \in X^{n-1}$.

Based on the above results for the case of two sets and elementary calculations, one can obtain the corresponding counterparts of Theorems 5.88 and 5.89 for systems of finitely many sets. In particular, a system of locally closed sets $\{\Omega_1, \ldots, \Omega_n\}$, $n \geq 2$, in an Asplund space X is linearly subextremal around $\bar{x}$ *if and only if* the following relations of the *approximate extremal principle* holds: for every $\varepsilon > 0$ there are $x_i \in \Omega_i \cap (\bar{x} + \varepsilon I\!B)$ and $x_i^* \in \widehat{N}(x_i; \Omega_i)$ for $i = 1, \ldots, n$ satisfying

$$\|x_1^* + \ldots + x_n^*\| \leq \varepsilon, \quad \|x_1^*\| + \ldots + \|x_n^*\| = 1 \, .$$

If in addition all but one Ω_i are SNC at x_i, then for any system $\{\Omega_1, \ldots, \Omega_n\}$ linearly subextremal around $\bar{x}$ one has the relations of the *exact extremal principle*: there are $x_i^* \in N(\bar{x}; \Omega_i)$, $i = 1, \ldots, n$, satisfying

$$x_1^* + \ldots + x_n^* = 0, \quad \|x_1^*\| + \ldots + \|x_n^*\| = 1 \, .$$

Furthermore, the latter relations are *necessary and sufficient* for the linear subextremality of $\{\Omega_1, \ldots, \Omega_n\}$ around $\bar{x}$ when X is finite-dimensional.

5.4.2 Linear Suboptimality in Multiobjective Optimization

In this subsection we consider some problems of *constrained multiobjective optimization* and study a new notion of *linearly suboptimal solutions* to such problems. This notion closely relates to (is actually induced by) the linear subextremality of set systems studied in the preceding subsection (similarly to the relationship between the generalized order optimality and set extremality in Subsect. 5.3.1), while we formulate it independently via the initial data. Our primary intention is to obtain *necessary and sufficient* conditions (as well as merely necessary conditions) for linearly suboptimal solutions in both approximate/fuzzy and exact/pointbased forms. Although the former conditions will be derived under more general assumptions, the latter ones have some advantages due to the possibility of using well-developed *calculus* for our basic normal/coderivative/subdifferential constructions. This is crucial to cover various constraints in multiobjective problems.

Given a mapping $f\colon X \to Z$ between Banach spaces, subsets $\Omega \subset X$ and $\Theta \subset Z$, and a point $\bar{x} \in \Omega$ with $f(\bar{x}) \in \Theta$, we introduce the constant

$$\vartheta_{\mathrm{lin}}(f, \Omega, \Theta, \bar{x}) := \liminf_{\substack{x \xrightarrow{\Omega} \bar{x},\ z \xrightarrow{\Theta} f(x) \\ r \downarrow 0}} \frac{\vartheta\big(f(B_r(x) \cap \Omega) - f(x), \Theta - z\big)}{r}, \quad (5.110)$$

where $\vartheta(\cdot, \cdot)$ is defined in (5.103).

Definition 5.91 (linearly suboptimal solutions to multiobjective problems). *Given* $(f, \Omega, \Theta, \bar{x})$ *as above, we say that* $\bar{x}$ *is* LINEARLY SUBOPTIMAL *with respect to* (f, Ω, Θ) *if one has*

$$\vartheta_{\mathrm{lin}}(f, \Omega, \Theta, \bar{x}) = 0$$

for the constant $\vartheta_{\mathrm{lin}}(f, \Omega, \Theta, \bar{x})$ *defined in* (5.110).

It is easy to check that every $\bar{x}$ locally (f, Θ)-optimal in the sense of Definition 5.53 (with $f(\bar{x}) = 0$ for simplicity) subject to the constraint $x \in \Omega$ happens to be also linearly suboptimal with respect to (f, Ω, Θ). Thus the above notion of linearly suboptimal solutions is an extension of the (exact) generalized order optimality for constrained multiobjective problems studied in Subsect. 5.3.5. Besides suboptimality versus optimality, another crucial difference between the solution notions in Definitions 5.91 and 5.53 is the *linear rate*; cf. the discussion on the relationships between the set extremality and linear subextremality after Definition 5.87. This allows us to obtain *necessary and sufficient* conditions for linearly suboptimal solutions in general settings. First we derive a "fuzzy" result in this direction, which is closely related (being actually equivalent) to the characterization of the linear subextremality via the approximate extremal principle from Theorem 5.88. To formulate this result, we define a set-valued mapping $F\colon X \rightrightarrows Z$ built upon (f, Ω, Θ) by

$$F(x) := \begin{cases} f(x) - \Theta & \text{if } x \in \Omega, \\ \emptyset & \text{otherwise}. \end{cases} \quad (5.111)$$

Note that the graph of this mapping F agrees with the generalized epigraph set $\mathcal{E}(f, \Omega, \Theta)$ considered in Subsect. 5.3.2.

Theorem 5.92 (fuzzy characterization of linear suboptimality in multiobjective optimization). *Let* X *and* Y *be Banach, and let* $\bar{x} \in \Omega$ *with* $f(\bar{x}) \in \Theta$. *The following assertions hold:*

(i) *Assume that for every* $\varepsilon > 0$ *there are* $(x, z) \in (\bar{x}, 0) + \varepsilon I\!B_{X \times Z}$ *with* $z \in F(x)$ *and* $z^* \in Z^*$ *with* $1 - \varepsilon \le \|z^*\| \le 1 + \varepsilon$ *satisfying the inclusion*

$$0 \in \widehat{D}^*_\varepsilon F(x,z)(z^*) \,. \tag{5.112}$$

Then $\bar{x}$ *is linearly suboptimal with respect to* (f,Ω,Θ).

(ii) *Conversely, assume that* $\bar{x}$ *is linearly suboptimal with respect to* (f,Ω,Θ). *Then for every* $\varepsilon > 0$ *there are* $(x,z) \in (\bar{x},0) + \varepsilon I\!B_{X\times Z}$ *with* $z \in F(x)$ *and* $z^* \in Z^*$ *with* $1-\varepsilon \le \|z^*\| \le 1+\varepsilon$ *satisfying the inclusion*

$$0 \in \widehat{D}^* F(x,z)(z^*)$$

provided that gph F *is locally closed around* $(\bar{x},0)$ *and that both spaces* X *and* Z *are Asplund.*

Proof. It is easy to see that $\bar{x}$ is linearly suboptimal with respect to (f,Ω,Θ) *if and only if* the system of two sets

$$\Omega_1 := \operatorname{gph} F \quad \text{and} \quad \Omega_2 := X \times \{0\} \subset X \times Z$$

is linearly subextremal around $(\bar{x},0) \in X \times Z$. Then applying the characterization of the linear subextremality from Theorem 5.88 to this set system $\{\Omega_1,\Omega_2\}$ and taking into account that $\widehat{N}_\varepsilon((x,0);\Omega_2) = (\varepsilon I\!B^*) \times Z^*$ and that

$$(0,-z^*) \in \widehat{N}_\varepsilon((x,z);\Omega_1) \Longleftrightarrow 0 \in \widehat{D}^*_\varepsilon F(x,z)(z^*)$$

for all $\varepsilon \ge 0$, we arrive at all the conclusions of the theorem. △

Corollary 5.93 (consequences of fuzzy characterization of linear suboptimality). *Condition* (5.112) *always implies that* $z^* \in \widehat{N}_\varepsilon(f(x);\Theta)$ *for all* $\varepsilon \ge 0$. *Moreover, for any* $x \in X$ *close to* $\bar{x}$ *with* $z = f(x) \in \Theta$ *one has*

$$0 \in \widehat{D}^* F(x,z)(z^*) \Longleftrightarrow 0 \in \widehat{\partial}\langle z^*, f_\Omega\rangle(x), \quad z^* \in \widehat{N}(f(x);\Theta)$$

with $f_\Omega = f + \delta(\cdot;\Omega)$ *provided that* f *is Lipschitz continuous around* $\bar{x}$ *relative to the constraint set* Ω.

Proof. Follows directly from the definitions and the (easy) scalarization formula for the Fréchet coderivative of locally Lipschitzian functions. △

Our next theorem provides *necessary* conditions and *sufficient* conditions (as well as *pointbased characterizations*) for linearly suboptimal solutions to multiobjective optimization problems given in the *condensed form*, i.e., via the mapping F built in (5.111) upon the initial data (f,Ω,Θ). These results are expressed in terms of the *mixed coderivative* (1.25) and the *reversed mixed coderivative* (1.40) of the mapping F calculated *exactly* at the reference solution. Note that the PSNC property of the mapping F^{-1} imposed in assertion (ii) of next theorem agrees with the PSNC property of the set $\mathcal{E}(f,\Omega,\Theta)$ in Theorem 5.59. Hence either one of the assumptions (a) and (b) of Theorem 5.59(ii) with $\bar{z}=0$ ensures the required PSNC property of F^{-1} at $(0,\bar{x})$; see the proof of Theorem 5.59. Recall also that sufficient conditions for the strong coderivative normality of F in assertion (iii) of the next theorem are listed in Proposition 4.9.

Theorem 5.94 (condensed pointbased conditions for linear suboptimality in multiobjective problems). *Let F be a mapping between Banach spaces built in* (5.111) *upon* (f, Ω, Θ). *The following hold:*

(i) *Assume that* $\dim X < \infty$ *and that there is* $0 \neq z^* \in Z^*$ *satisfying*

$$0 \in D^*_M F(\bar{x}, 0)(z^*) \; .$$

Then $\bar{x}$ is linearly suboptimal with respect to (f, Ω, Θ).

(ii) *Conversely, assume that $\bar{x}$ is linearly suboptimal with respect to* (f, Ω, Θ). *Then there is* $0 \neq z^* \in Z^*$ *satisfying*

$$0 \in \widetilde{D}^*_M F(\bar{x}, 0)(z^*)$$

provided that both X and Z are Asplund, that $\operatorname{gph} F$ *is locally closed around* $(\bar{x}, 0)$, *and that F^{-1} is PSNC at* $(0, \bar{x})$; *the latter is automatic when either* $\dim Z < \infty$ *or F is metrically regular around* $(\bar{x}, 0)$.

(iii) *Let* $\dim X < \infty$, *let Z be Asplund, and let F be closed-graph around* $(\bar{x}, 0)$. *Assume also that F is SNC and strongly coderivatively normal at* $(\bar{x}, 0)$ *with* $D^*F(\bar{x}, 0) := D^*_M F(\bar{x}, 0) = D^*_N F(\bar{x}, 0)$. *Then $\bar{x}$ is linearly suboptimal with respect to* (f, Ω, Θ) *if and only if there is* $0 \neq z^* \in Z^*$ *satisfying*

$$0 \in D^* F(\bar{x}, 0)(z^*) \; .$$

Proof. Let us first justify (i). Using $0 \in D^*_M F(\bar{x}, 0)(z^*)$ and the definition of the mixed coderivative with $\dim X < \infty$, we find $\varepsilon_k \downarrow 0$, $x_k \to \bar{x}$, $z_k \to 0$, $x^*_k \to 0$, and $z^*_k \to z^*$ such that

$$z_k \in F(x_k) \quad \text{and} \quad x^*_k \in \widehat{D}^*_{\varepsilon_k} F(x_k, z_k)(z^*_k) \quad \text{whenever} \quad k \in I\!N \; .$$

Note that the first inclusion above implies, due to the construction of F in (5.111), that $x_k \in \Omega$ and $z_k = f(x_k) \in \Theta$. Furthermore, since $\|z^*_k - z^*\| \to 0$ and $\|z^*\| = 1$, we may assume without loss of generality that $\|z^*_k\| = 1$ for all $k \in I\!N$. From $x^*_k \in \widehat{D}^*_{\varepsilon_k} F(x_k, z_k)(z^*_k)$ one has

$$\langle x^*_k, x - x_k \rangle - \langle z^*_k, z - z_k \rangle \le \varepsilon_k \big(\|x - x_k\| + \|z - z_k\| \big)$$

whenever the pair (x, z) is sufficiently close to $(\bar{x}, 0)$. This implies the estimate

$$-\langle z^*_k, z - z_k \rangle \le \big(\varepsilon_k + \|x^*_k\| \big) \big(\|x - x_k\| + \|z - z_k\| \big) \; ,$$

which means that

$$0 \in \widehat{D}^*_{\gamma_k} F(x_k, z_k)(z^*_k) \quad \text{with} \quad \gamma_k := \varepsilon_k + \|x^*_k\| \downarrow 0 \quad \text{as} \quad k \to \infty \; .$$

Applying now assertion (i) of Theorem 5.92, we conclude that $\bar{x}$ is linearly suboptimal with respect to (f, Ω, Θ).

To prove (ii), let us take a point $\bar{x}$ linearly suboptimal with respect to (f, Ω, Θ) and pick an arbitrary sequence $\varepsilon_k \downarrow 0$ as $k \to \infty$. Using assertion

(ii) of Theorem 5.92, we find sequences $(x_k, z_k) \to (\bar{x}, 0)$ with $z_k \in F(x_k)$ and $z_k^* \in Z^*$ with $\|z_k^*\| = 1$ satisfying $0 \in \widehat{D}^* F(x_k, z_k)(z_k^*)$ for all $k \in I\!N$. Since Z is Asplund, there is $z^* \in Z^*$ such that $z_k^* \stackrel{w^*}{\to} z^*$ as $k \to \infty$ along a subsequence, and one clearly has $0 \in \widetilde{D}_M^* F(\bar{x}, 0)(z^*)$ by passing to the limit. Furthermore, $z^* \neq 0$ by the PSNC assumption made. The latter assumption obviously holds if Z is finite-dimensional. It is also fulfilled when F is metrically regular around $(\bar{x}, 0)$ by Proposition 1.68 and the equivalence between the Lipschitz-like property of F and the metric regularity of F^{-1}. Thus we arrive at all the conclusions in (ii).

The final assertion (iii) is a direct combination of (i) and (ii). Note that $\widetilde{D}_M^* F(\bar{x}, 0) = D_N^* F(\bar{x}, 0)$ and the PSNC property of F^{-1} is equivalent to the SNC property of F in this case, since $\dim X$ is finite-dimensional. △

Using full calculus, we deduce from the condensed results of Theorem 5.94(ii) comprehensive *necessary* conditions for linear suboptimality in multiobjective problems and their specifications subject to various (in particular, equilibrium) constraints expressed *separately* via the initial data (f, Ω, Θ), i.e., in terms of generalized differential constructions for each of f, Ω, and Θ; cf. the results of Subsects. 5.3.2 and 5.3.5 for generalized order optimality. The situation for *sufficient* conditions and also for the *characterization* of linear suboptimality is more delicate: we have to employ calculus rules with *equalities*, which are essentially more restrictive than those we need for necessity. Let us present some results in this direction providing the characterization of linear suboptimality in terms of the initial data (f, Ω, Θ) based on the condensed conditions of Theorem 5.94(iii).

Theorem 5.95 (separated pointbased criteria for linear suboptimality in multiobjective problems). *Let $f: X \to Z$ be Lipschitz continuous around $\bar{x}$ with* $\dim X < \infty$*, and let $\Omega \subset X$ and $\Theta \subset Z$ be locally closed around $\bar{x}$ and $\bar{z} := f(\bar{x}) \in \Theta$, respectively. Impose one of the following assumptions* (a)–(c) *on the initial data:*

(a) $\dim Z < \infty$ *and either $\Omega = X$, or f strictly differentiable at $\bar{x}$.*

(b) *Z is Asplund, $\Omega = X$, Θ is normally regular and SNC at $\bar{z}$, and f is strictly Lipschitzian at $\bar{x}$.*

(c) *Z is Asplund, Ω is normally regular at $\bar{x}$, Θ is normally regular and SNC at $\bar{z}$, and f is N-regular at $\bar{x}$.*

Then $\bar{x}$ is linearly suboptimal with respect to (f, Ω, Θ) if and only if there is $0 \neq z^* \in Z^*$ satisfying

$$0 \in \partial\langle z^*, f\rangle(\bar{x}) + N(\bar{x}; \Omega), \quad z^* \in N(\bar{z}; \Theta)\ .$$

Proof. Since $\operatorname{gph} F = \mathcal{E}(f, \Omega, \Theta)$ with the latter set defined in (5.37), we have

$$D_N^* F(\bar{x}, 0)(z^*) = \Big\{x^* \in X^* \Big|\ (x^*, -z^*) \in N\big((\bar{x}, 0); \mathcal{E}(f, \Omega, \Theta)\big)\Big\}\ .$$

Then assertions (iii) and (iv) of Lemma 5.23 ensure the representation

$$D_N^* F(\bar{x},0)(z^*) = \begin{cases} \partial\langle z^*, f_\Omega\rangle(\bar{x}) & \text{if } z^* \in N(\bar{z};\Theta)\,, \\ \emptyset & \text{otherwise} \end{cases} \tag{5.113}$$

provided that Z is Asplund and that the restriction f_Ω of f on Ω is locally Lipschitzian around $\bar{x}$ and strongly coderivatively normal at this point.

Consider first the case of $\Omega = X$. Then we observe from (5.113) that F is strongly coderivatively normal at $(\bar{x},0)$ if either $\dim Z < \infty$, or f is strictly Lipschitzian at $\bar{x}$ and Θ is normally regular at $\bar{z}$; see Proposition 4.9. To meet all the assumptions of Theorem 5.94(iii), one needs also to check (in the case of $\dim Z = \infty$) that F^{-1} is PSNC at $(0,\bar{x})$. Invoking the proof of Theorem 5.59(ii), we ensure this property if either Θ is SNC at $\bar{z}$ or f^{-1} is PSNC at $(\bar{z},\bar{x})$. Since X is finite-dimensional, the latter is equivalent to the SNC property of f at $(\bar{x},\bar{z})$ and, by Corollary 3.30, reduces to $\dim Z < \infty$ for strictly Lipschitzian mappings. Thus we complete the proof of the theorem in the case of $\Omega = X$.

To proceed in the constraint case of $\Omega \neq X$ under the assumptions made, it remains to ensure the equality

$$\partial\langle z^*, f_\Omega\rangle(\bar{x}) = \partial\langle z^*, f\rangle(\bar{x}) + N(\bar{x};\Omega)$$

in (5.113). By the sum rule of Proposition 1.107(ii) we have this equality when f is strictly differentiable at $\bar{x}$. Moreover, this equality holds and f_Ω is also N-regular (and hence strongly coderivatively normal) at $\bar{x}$ if f is N-regular and Ω is normally regular at this point; see Propositions 3.12 and 4.9. Combining these facts with the assumptions on Θ in (c) needed in the case of $\dim Z = \infty$ similarly to the above proof for $\Omega = X$, we arrive at all the requirements of Theorem 5.94(iii) and complete the proof of the theorem. △

Let us present a corollary of the last theorem giving a characterization of linearly suboptimal solutions to multiobjective problems with *operator constraints*. Note that the corresponding *necessary* optimality conditions obtained in Corollary 5.60 hold true *without any change* for linearly suboptimal solutions under general operator constraints given by set-valued and nonsmooth mappings. However, the *necessary and sufficient* conditions presented below require essentially more restrictive assumptions on the initial data ensuring the *equality* in the calculus rule for inverse images and in addition the *normal regularity* of inverse images in infinite dimensions. This inevitably confines our consideration to strictly differentiable mappings describing operator and functional constraints in multiobjective problems.

Corollary 5.96 (pointbased criteria for linear suboptimality under operator constraints). *Let* $f\colon X \to Z$, $g\colon X \to Y$, $\Theta \subset Z$, *and* $\Lambda \subset Y$. *Assume that* $\dim X < \infty$, *that* Θ *and* Λ *are locally closed around* $\bar{z}$ *and* $\bar{y} :=$

$g(\bar{x})$, respectively, and that f is strictly differentiable at $\bar{z}$ while g has this property at $\bar{y}$. Suppose also that one of the following assumptions holds:

(a) *Y is Banach,* $\dim Z < \infty$, *and $\nabla g(\bar{y})$ is surjective.*

(b) $\dim Y < \infty$, *Z is Asplund, Λ is normally regular at $\bar{y}$, Θ is normally regular and SNC at $\bar{z}$, and*

$$N(\bar{y};\Lambda)\cap\ker\nabla g(\bar{x})^* = \{0\}\ .$$

Then $\bar{x}$ is linearly suboptimal with respect to $(f, g^{-1}(\Lambda), \Theta)$ if and only if there is $0 \neq z^* \in Z^*$ *satisfying*

$$0 \in \nabla f(\bar{x})^* z^* + \nabla g(\bar{x})^* N(\bar{y};\Lambda), \quad z^* \in N(\bar{z};\Theta)\ .$$

Proof. We use Theorem 5.95 with $\Omega := g^{-1}(\Lambda)$. First apply Theorem 1.17 to ensure the calculus formula

$$N(\bar{x};\Omega) = \nabla g(\bar{x})^* N(\bar{y};\Lambda)$$

under the surjectivity assumption on $\nabla g(\bar{y})$ made in (a) when Y is Banach. Then we arrive at the conclusion of this corollary due to Theorem 5.95(a).

To ensure the normal regularity of $\Omega = g^{-1}(\Lambda)$, needed in Theorem 5.95(c) in addition to the above calculus formula, we employ Theorem 3.13(iii) with $F(y) = \delta(y;\Lambda)$ therein, which justifies the conclusion of the corollary under the assumptions made in (b). Note that we cannot get anything but strict differentiability from the N-regularity condition on g in the latter theorem, since the graphical regularity of g is equivalent to its strict differentiability at the reference point due to Corollary 3.69 with $\dim X < \infty$. △

The result obtained has a striking consequence for the case of multiobjective problems with *functional constraints* in the classical form of equalities and inequalities given by strictly differentiable functions. In this case an appropriate multiobjective version of the *Lagrange multiplier rule* in the *normal form* provides *necessary and sufficient* conditions for linear suboptimality under the *Mangasarian-Fromovitz constraint qualification.*

Corollary 5.97 (linear suboptimality in multiobjective problems with functional constraints). *Let $f\colon X \to Z$ be strictly differentiable at $\bar{x}$ with* $\dim X < \infty$ *and Z Asplund, let Θ be normally regular and SNC at $\bar{z}$, and let*

$$\Omega := \big\{x \in X\big|\ \varphi_i(x) \le 0,\ i = 1,\ldots,m;\ \varphi_i(x) = 0,\ i = m+1,\ldots,m+r\big\}\ ,$$

where each φ_i is strictly differentiable at $\bar{x}$. Assume the Mangasarian-Fromovitz constraint qualification:

(a) $\nabla\varphi_{m+1}(\bar{x}),\ldots,\nabla\varphi_{m+r}(\bar{x})$ *are linearly independent, and*
(b) *there is* $u\in X$ *satisfying*

$$\langle\nabla\varphi_i(\bar{x}),u\rangle<0,\quad i\in\{1,\ldots,m\}\cap I(\bar{x})\;,$$

$$\langle\nabla\varphi_i(\bar{x}),u\rangle=0,\quad i=m+1,\ldots,m+r\;,$$

where $I(\bar{x}):=\big\{i=1,\ldots,m+r\big|\;\varphi_i(\bar{x})=0\big\}$.

Then $\bar{x}$ *is linearly suboptimal with respect to* (f,Ω,Θ) *if and only if there is* $z^*\in N(\bar{z};\Theta)\setminus\{0\}$ *and* $(\lambda_1,\ldots,\lambda_{m+r})\in I\!R^{m+r}$ *such that*

$$\nabla f(\bar{x})^*z^*+\sum_{i=1}^{m+r}\lambda_i\nabla\varphi_i(\bar{x})=0\;,$$

$$\lambda_i\ge 0\;\;\textit{and}\;\;\lambda_i\varphi_i(\bar{x})=0\;\;\textit{for all}\;\;i=1,\ldots,m\;.$$

Proof. Follows from Corollary 5.96(b) with

$$\Lambda:=\Big\{(\alpha_1,\ldots,\alpha_{m+r})\in I\!R^{m+r}\Big|\;\alpha_i\le 0\;\text{ for }\;i=1,\ldots,m\qquad\text{and}$$
$$\alpha_i=0\;\text{ for }\;i=m+1,\;\ldots,m+r\Big\}$$

and $g:=(\varphi_1,\;\ldots,\varphi_{m+r})\colon X\to I\!R^{m+r}$. △

Let us next derive *necessary and sufficient* conditions for linear suboptimality in *multiobjective problems with equilibrium constraints*, i.e., in *EPECs* in the terminology of Subsect. 5.3.5. Taking into account the above discussions, the general framework for such problems is formulated as follows. Given $f\colon X\times Y\to Z$, $S\colon X\rightrightarrows Y$, and $\Theta\subset Z$, we say that $(\bar{x},\bar{y})$ is *linearly suboptimal* with respect to (f,S,Θ) if it is linearly suboptimal with respect to $(f,\operatorname{gph}S,\Theta)$ in the sense of Definition 5.91. We are mostly interested in equilibrium constraints described by solution maps to *parametric variational systems* of the type

$$0\in q(x,y)+Q(x,y)\;.$$

First observe, based on Theorem 5.94(ii) and calculus rules of the inclusion type, that all the *necessary* conditions obtained in Subsect. 5.3.5 for generalized order optimality hold true for linearly suboptimal solutions to the EPECs under consideration. To derive *criteria* for linear suboptimality, we need to employ more restrictive calculus rules of the *equality* type that provide exact formulas for computing coderivatives of solution maps given by equilibrium constraints and also ensure *graphical regularity* of these maps in appropriate settings. To proceed, we rely on the results of Theorem 5.95 with $\Omega=\operatorname{gph}S\subset X\times Y$ and on the corresponding coderivative formulas and regularity assertions established in Subsect. 4.4.1 for parametric variational systems. In the next theorem we impose for simplicity the strict differentiability assumption on f (instead of N-regularity in (c) of Theorem 5.95), which is unavoidable when $\dim Z < \infty$ while it may be relaxed in infinite dimensions; see Theorem 3.68.

Theorem 5.98 (characterization of linear suboptimality for general EPECs). *Let $f\colon X \times Y \to Z$ and $q\colon X \times Y \to P$ be strictly differentiable at $(\bar{x}, \bar{y})$ with $\bar{z} := f(\bar{x}, \bar{y}) \in \Theta$ and $\bar{p} := -q(\bar{x}, \bar{y})$; let $\Theta \subset Z$ and the graph of $Q\colon X \times Y \rightrightarrows P$ be locally closed around $\bar{z}$ and $(\bar{x}, \bar{y}, \bar{p}) \in \operatorname{gph} Q$, respectively; let both X and Y be finite-dimensional; and let*

$$S(x) := \big\{ y \in Y \big|\ 0 \in q(x, y) + Q(x, y) \big\} .$$

Assume in addition that one of the following requirements holds:

(a) $\dim Z < \infty$, *P is Banach, $\nabla_x q(\bar{x}, \bar{y})$ is surjective, and $Q = Q(y)$.*

(b) *Z and P are Asplund, Θ is SNC and normally regular at $\bar{z}$, $Q = Q(x, y)$ is SNC and N-regular at $(\bar{x}, \bar{y}, \bar{p})$, and the adjoint generalized equation*

$$0 \in \nabla q(\bar{x}, \bar{y})^* p^* + D^*_N Q(\bar{x}, \bar{y}, \bar{p})(p^*)$$

has only the trivial solution $p^* = 0$.

Then $(\bar{x}, \bar{y})$ is linearly suboptimal with respect to (f, S, Θ) if and only if there are $z^* \in N(\bar{z}; \Theta) \setminus \{0\}$ and $p^* \in P^*$ satisfying

$$0 \in \nabla f(\bar{x}, \bar{y})^* z^* + \nabla q(\bar{x}, \bar{y})^* p^* + D^*_N Q(\bar{x}, \bar{y}, \bar{p})(p^*) .$$

Proof. Employing Theorem 5.95 with $\Omega = \operatorname{gph} S \subset X \times Y$, we conclude that $(\bar{x}, \bar{y})$ is linearly suboptimal with respect to (f, S, Θ) *if and only if* there is $z^* \in N(\bar{z}; \Theta) \setminus \{0\}$ satisfying

$$0 \in \nabla f(\bar{x}, \bar{y})^* z^* + N\big((\bar{x}, \bar{y}); \operatorname{gph} S\big)$$

provided that both X and Y are finite-dimensional, that f is strictly differentiable at $(\bar{x}, \bar{y})$, and that either $\dim Z < \infty$ or Z is Asplund, Θ is SNC and normally regular at $\bar{z}$, and S is N-regular at $(\bar{x}, \bar{y})$.

To obtain results in terms of the initial data for the solution map S, we thus need to represent $N\big((\bar{x}, \bar{y}); \operatorname{gph} S\big)$ via (q, Q) and also to invoke additional conditions ensuring the N-regularity of S at $(\bar{x}, \bar{y})$ when $\dim Z = \infty$. First consider the case of $\dim Z < \infty$, when we don't need to ensure the regularity of S. In this case one has by Theorem 4.44(i) that

$$N\big((\bar{x}, \bar{y}); \operatorname{gph} S\big) = \Big\{ (x^*, y^*) \in X^* \times Y^* \Big|\ x^* = \nabla_x q(\bar{x}, \bar{y})^* p^* ,$$
$$y^* \in \nabla_y q(\bar{x}, \bar{y})^* p^* + D^*_N Q(\bar{y}, \bar{p})(p^*) \ \text{ for some } \ p^* \in P^* \Big\}$$

when P is Banach, $Q = Q(\bar{y})$, and $\nabla_x q(\bar{x}, \bar{y})$ is surjective. This gives the conclusion of the theorem in case (a).

If $Q = Q(x, y)$ and Z is Asplund, we employ assertion (ii) of Theorem 4.44, which gives the representation formula for $N\big((\bar{x}, \bar{y}); \operatorname{gph} S\big)$ and simultaneously ensures the N-regularity of S at $(\bar{x}, \bar{y})$ under the regularity assumption on Q but with no surjectivity of $\nabla_x q(\bar{x}, \bar{y})$. Combining this with the assumptions in Theorem 5.95(c), we complete the proof of the theorem. △

The most restrictive assumption in (b) of Theorem 5.98 is the N-regularity of the field Q. It holds, in particular, when Q is *convex-graph*. The reader can easily get a specification of Theorem 5.98 in this case from the results of Corollary 4.45 expressed *explicitly* in terms of Q but not its coderivative.

Let us present a specification of Theorem 5.98 in the case of $Q = \partial(\psi \circ g)$, i.e., when the field of the generalized equation under consideration is given in the *subdifferential form* with a *composite potential*. As discussed in Subsect. 4.4.1, such a model covers classical variational inequalities and their extensions. To obtain characterizations of linear suboptimality for EPECs of this type, we involve *second-order* subdifferential chain rules giving a representation of $D^*Q = \partial^2(\psi \circ g)$ via the initial data (ψ, g). Again, we may apply only those calculus results that ensure chain rules as *equalities*. Since graphical regularity is not a realistic property for subdifferential mappings with nonsmooth potentials, we restrict ourselves to case (a) of Theorem 5.98 combined with the coderivative calculation in Theorem 4.49 for solution maps to parametric *hemivariational inequalities* (HVIs).

Corollary 5.99 (linear suboptimality for EPECs governed by HVIs with composite potentials). *Let $Q(y) = \partial(\psi \circ g)(y)$ under the assumptions in case* (a) *of Theorem* 5.98, *where*

$$S(x) := \big\{ y \in Y \big|\ 0 \in q(x, y) + \partial(\psi \circ g)(y) \big\} ,$$

$q\colon X \times Y \to Y^*$, $g\colon Y \to W$, $\psi\colon W \to \overline{I\!R}$, and W is Banach. Suppose in addition that $g \in \mathcal{C}^1$ with the surjective derivative $\nabla g(\bar{y})$, that $\nabla g(\cdot)$ is strictly differentiable at $\bar{y}$, and that the graph of $\partial\psi$ is locally closed around $(\bar{w}, \bar{v})$, where $\bar{w} := g(\bar{y})$ and where $\bar{v} \in W^*$ is a unique functional satisfying

$$-q(\bar{x}, \bar{y}) = \nabla g(\bar{y})^* \bar{v} .$$

Then $(\bar{x}, \bar{y})$ is linearly suboptimal with respect to (f, S, Θ) if and only if there are $z^* \in N(\bar{z}; \Theta) \setminus \{0\}$ and (*uniquely defined*) *$u \in Y$ such that*

$$0 = \nabla_x f(\bar{x}, \bar{y})^* z^* + \nabla_x q(\bar{x}, \bar{y})^* u \quad \textit{and}$$

$$0 \in \nabla_y f(\bar{x}, \bar{y})^* z^* + \nabla_y q(\bar{x}, \bar{y})^* u + \nabla g(\bar{y})^* \partial_N^2 \psi(\bar{w}, \bar{v})\big(\nabla g(\bar{y}) u\big) .$$

Proof. Follows from Theorem 5.98(a) due to the calculation of $D_N^* S(\bar{x}, \bar{y})$ for the above mapping S given in Theorem 4.49, which is based on the second-order subdifferential formula from Theorem 1.127. △

Finally in this subsection, we present a criterion for linear suboptimality for EPECs governed by parametric generalized equations with *composite fields.*

Corollary 5.100 (linear suboptimality for EPECs governed by HVIs with composite fields). *Let* $Q(y) = (\partial\psi \circ g)(y)$ *under the assumptions in case* (a) *of Theorem* 5.98*, where* $P = W^*$ *for some Banach space* W*, where*

$$S(x) := \big\{ y \in Y \big|\, 0 \in q(x,y) + (\partial\psi \circ g)(y) \big\}$$

with $g\colon Y \to W$ *and* $\psi\colon W \to \overline{I\!R}$*, and where* g *is strictly differentiable at* $\bar{y}$ *with the surjective derivative* $\nabla g(\bar{y})$*. Denoting* $\bar{w} := g(\bar{y})$ *and* $\bar{p} := -q(\bar{x},\bar{y})$*, we assume that the graph of* $\partial\psi$ *is locally closed around* $(\bar{w},\bar{p})$*, which is automatic when* ψ *is either continuous or amenable. Then* $(\bar{x},\bar{y})$ *is linearly suboptimal with respect to* (f,S,Θ) *if and only if there are* $z^* \in N(\bar{z};\Theta)\setminus\{0\}$ *and* (*uniquely defined*) $u \in W^{**}$ *satisfying*

$$0 = \nabla_x f(\bar{x},\bar{y})^* z^* + \nabla_x q(\bar{x},\bar{y})^* u \quad \text{and}$$

$$0 \in \nabla_y f(\bar{x},\bar{y})^* z^* + \nabla_y q(\bar{x},\bar{y})^* u + \nabla g(\bar{y})^* \partial_N^2 \psi(\bar{w},\bar{p})(u)\ .$$

Proof. Follows from Theorem 5.98(a) due to the calculation of $D_N^* S(\bar{x},\bar{y})$ for the above mapping S given in Proposition 4.53 based on the coderivative chain rule from Theorem 1.66. △

5.4.3 Linear Suboptimality for Minimization Problems

In the concluding subsection of Sect. 5.4 (and of the whole chapter) we study the above notion of linear suboptimality for usual *minimization* problems; thus we refer to this notion as to *linear subminimality.* Minimization problems form, of course, a special subclass of the multiobjective optimization problems considered in the preceding subsection with a single (real-valued) objective f and with $\Theta = I\!R_-$. On the other hand, such problems and their linearly suboptimal solutions have some specific features in comparison with general multiobjective problems. We present characterizing results for linear subminimality in both approximate and pointbased forms for unconstrained and constrained problems. Some striking illustrative examples will be given as well. Besides necessary and sufficient conditions for linear subminimality involving lower subgradients, we obtain also refined necessary conditions via upper subgradients, which are specific for minimization problems.

Definition 5.101 (linear subminimality). *Let* $\Omega \subset X$*, and let* $\varphi\colon X \to \overline{I\!R}$ *be finite at* $\bar{x} \in \Omega$*. We say that* $\bar{x}$ *is* LINEARLY SUBMINIMAL *with respect to* (φ,Ω) *if one has*

$$\limsup_{\substack{x \xrightarrow{\Omega} \bar{x} \\ \varphi(x)\to\varphi(\bar{x}) \\ r\downarrow 0}} \inf_{u\in B_r(x)\cap\Omega} \frac{\varphi(u)-\varphi(x)}{r} = 0 .$$

The point $\bar{x}$ is said to be linearly subminimal for φ if $\Omega = X$ in the above.

Observe that the linear subminimality of $\bar{x}$ with respect to (φ, Ω) corresponds to the linear suboptimality of $\bar{x}$ with respect to (f, Ω, Θ) from Definition 5.91 when $f(x) = \varphi(x) - \varphi(\bar{x})$ and $\Theta = I\!R_-$.

It is easy to see that any local minimizer for the function φ subject to $x \in \Omega$ is linearly subminimal with respect to (φ, Ω), but not vice versa. The next example illustrates some striking differences that occur even for unconstrained problems involving one-dimensional functions.

Example 5.102 (specific features of linear subminimality). One can check directly from the definition that $\bar{x} = 0 \in I\!R$ is *linearly subminimal* for each of the following functions: $\varphi(x) := x^2$, $\varphi(x) := -x^2$, and $\varphi(x) := x^3$. These functions are different from the viewpoint of minimization having $\bar{x} = 0$ as a *minimizer*, a *maximizer*, and just a *stationary point*, respectively.

The point $\bar{x} = 0$ is also linearly subminimal for the piecewise constant and l.s.c. function

$$\varphi(x) := \begin{cases} -\dfrac{1}{n}, & -\dfrac{1}{n} < x \le -\dfrac{1}{n+1}, \quad n \in I\!N , \\[2ex] 0, & x = 0 , \\[2ex] \dfrac{1}{n}, & \dfrac{1}{n+1} < x \le \dfrac{1}{n}, \quad n \in I\!N . \end{cases}$$

Although this point is *not* a local minimizer for φ, *every neighborhood* of $\bar{x} = 0$ contains a point of local minimum to this function. However, this is not the case for the function

$$\varphi(x) := \begin{cases} -\dfrac{1}{n} + \dfrac{1}{n^2}\Big(x + \dfrac{1}{n}\Big), & -\dfrac{1}{n} < x \le -\dfrac{1}{n+1}, \quad n \in I\!N , \\[2ex] 0, & x = 0 , \\[2ex] \dfrac{1}{n} + \dfrac{1}{n^2}\Big(x - \dfrac{1}{n+1}\Big), & \dfrac{1}{n+1} < x \le \dfrac{1}{n}, \quad n \in I\!N , \end{cases}$$

which is l.s.c., piecewise linear, and *doesn't have local minimizers at all*, while $\bar{x} = 0$ is *linearly subminimal* for it.

Let us present some *equivalent descriptions* of linear subminimality in primal spaces that clarify its relation to *perturbed minimization* as well as to *generalized stationary* points defined via *limiting slopes* instead of classical derivatives. For simplicity we put $\Omega = X$ in Definition 5.101 taking into account that φ is extended-real-valued.

Theorem 5.103 (equivalent descriptions of linear subminimality). *Let $\varphi\colon X \to \overline{I\!R}$ be finite at $\bar{x}$ and l.s.c. around this point, and let X be Banach. The following properties are equivalent:*

(a) *The point $\bar{x}$ is linearly subminimal for φ.*

(b) *For any $\varepsilon_k \downarrow 0$ there exists $x_k \xrightarrow{\varphi} \bar{x}$ as $k \to \infty$ such that*

$$\varphi(x_k) \le \varphi(x) + \varepsilon_k \|x - x_k\| \ \text{ for all } \ x \ \text{ around } \ x_k \text{ and } k \in I\!N\ .$$

(c) *One has*

$$\liminf_{x \xrightarrow{\varphi} \bar{x}} |\nabla\varphi|(x) = 0, \quad \text{where} \quad |\nabla\varphi|(x) := \limsup_{u \xrightarrow{\varphi} x} \frac{\max\{\varphi(x) - \varphi(u), 0\}}{\|u - x\|}$$

is called the (strong) SLOPE *of φ at x.*

Proof. It is easy to observe that the property in (b) can be equivalently described as

$$\limsup_{\substack{x \xrightarrow{\varphi} \bar{x} \\ r \downarrow 0}} \tau_\varphi(x, r) = 0, \quad \text{where} \quad \tau_\varphi(x, r) := \inf_{\|u - x\| < r} \min\left\{\frac{\varphi(u) - \varphi(x)}{\|u - x\|}, 0\right\} .$$

Thus (b)$\Rightarrow$(a) due to

$$\tau_\varphi(x, r) \le \inf_{\|u - x\| \le r} \frac{\varphi(u) - \varphi(x)}{r} \le 0 \ \text{ for all } \ r > 0\ .$$

To prove (a)$\Rightarrow$(b), assume that $\bar{x}$ is linearly subminimal for φ and find by Definition 5.101 sequences $u_k \xrightarrow{\varphi} \bar{x}$ and $\varepsilon_k \downarrow 0$ such that

$$\varphi(x) - \varphi(u_k) \ge -\varepsilon_k^2 \ \text{ for all } \ x \in u_k + 2\varepsilon_k I\!B\ .$$

By the Ekeland variational principle of Theorem 2.26, employed for each number $k \in I\!N$, there is $x_k \in u_k + \varepsilon_k I\!B$ satisfying

$$\varphi(x_k) \le \varphi(u_k) \quad \text{and} \quad \varphi(x) - \varphi(x_k) \ge -\varepsilon_k \|x - x_k\| \ \text{ whenever } \ x \ \text{ near } \ x_k\ .$$

One obviously has $x_k \xrightarrow{\varphi} \bar{x}$ and $\tau_\varphi(x_k, r) \ge -\varepsilon_k$ for small $r > 0$, which implies property (b) due to its description above.

Observing finally that

$$\limsup_{\substack{x \xrightarrow{\varphi} \bar{x} \\ r \downarrow 0}} \tau_\varphi(x, r) = -\liminf_{x \xrightarrow{\varphi} \bar{x}} |\nabla\varphi|(x)\ ,$$

we get the equivalence (b)$\Longleftrightarrow$(c) and complete the proof of the theorem. △

One of the most principal differences between standard local minimizers and linearly subminimal points is that the former are *not stable* with respect to small perturbations while the latter *are*. Indeed, consider the simplest quadratic function $\varphi(x) := x^2$ for which $\bar{x} = 0$ gives the global minimum. Perturbing φ by $\psi(x) := -|x|^{3/2}$ around $\bar{x}$, we see that $\bar{x}$ is no longer a local minimizer for the function $\varphi(x) + \psi(x)$; actually the latter achieves its global maximum at $\bar{x} = 0$. On the other hand, the notion of linear subminimality is *stable* relative to any smooth perturbations with *vanishing derivatives*.

Proposition 5.104 (stability of linear subminimality). *Let $\Omega \subset X$, and let $\varphi: X \to \overline{I\!R}$ and $\psi: X \to I\!R$ be functions on a Banach space X such that ψ is strictly differentiable at $\bar{x}$ with $\nabla\psi(\bar{x}) = 0$. Then $\bar{x}$ is linearly subminimal with respect to (φ, Ω) if and only if it is linearly subminimal with respect to $(\varphi + \psi, \Omega)$.*

Proof. It follows directly from Definition 5.101 and from $\nabla\psi(\bar{x}) = 0$ for the strict derivative of ψ that the linear subminimality of $\bar{x}$ with respect to (φ, Ω) yields the one of $\bar{x}$ with respect to $(\varphi + \psi, \Omega)$. Applying this to the functions $\varphi + \psi$ and $-\psi$, we have the opposite implication. △

An immediate consequence of this observation is that, in any Banach space, *linearly subminimal* points of any *smooth* function reduce to its *stationary* points in the classical sense.

Corollary 5.105 (linearly subminimal and stationary points of strictly differentiable functions). *Let $\varphi: X \to I\!R$ be strictly differentiable at $\bar{x}$. Then $\bar{x}$ is linearly subminimal for φ if and only if $\bar{x}$ is a φ-stationary point, i.e., one has $\nabla\varphi(\bar{x}) = 0$.*

Proof. It follows from Proposition 5.104 with $\varphi = 0$ and $\Omega = X$. △

One can see from Corollary 5.105 that, for strictly differentiable real-valued functions on Banach spaces, the notion of *linear subminimality* and the symmetric one of *linear submaximality* are *equivalent*. This is *not* however the case for *nonsmooth* functions. Thus both notions of linear subminimality and linear submaximality can be treated as *one-sided* extensions of the classical stationary concepts to nonsmooth functions.

Let us now derive, based on Theorems 5.92 and 5.94, *necessary and sufficient* conditions for linear subminimality in both fuzzy/approximate and pointbased/exact forms. For brevity we formulate only criteria for this property but not necessary conditions and sufficient conditions separately. The next theorem contains *condensed* results in this direction via Fréchet and basic subgradients of the restriction $\varphi_\Omega(x) := \varphi(x) + \delta(x; \Omega)$ of the cost function φ on the constraint set Ω in the general setting.

Theorem 5.106 (condensed subdifferential criteria for linear subminimality). *Let $\varphi_\Omega: X \to \overline{I\!R}$ be l.s.c. around $\bar{x} \in \Omega$ with $|\varphi(\bar{x})| < \infty$, and let X be Asplund. The following assertions hold:*

(i) *The point* $\bar{x}$ *is linearly subminimal with respect to* (φ, Ω) *if and only if for every* $\varepsilon > 0$ *there are* $x \in \Omega \cap (\bar{x} + \varepsilon I\!B)$ *with* $|\varphi(x) - \varphi(\bar{x})| \le \varepsilon$ *and* $x^* \in \partial\varphi_\Omega(x)$ *with* $\|x^*\| \le \varepsilon$.

(ii) *Assume that* $\dim X < \infty$. *Then* $\bar{x}$ *is linearly subminimal with respect to* (φ, Ω) *if and only if* $0 \in \partial\varphi_\Omega(\bar{x})$.

Proof. Assertion (i) of the theorem follows from the fuzzy characterization of linear suboptimality in Theorem 5.92 with $\Theta = I\!R_-$ and $f(x) = \varphi(x) - \varphi(\bar{x})$.

To prove (ii), we use the pointbased characterization of linear suboptimality in (iii) of Theorem 5.94 with the same f as in (i) and F defined in (5.111). Note that this F is automatically SNC and strongly coderivatively normal at $(\bar{x}, 0)$ due to $Z = I\!R$, and one obviously has

$$0 \in D^*F(\bar{x}, 0)(1) \Longleftrightarrow 0 \in \partial\varphi_\Omega(\bar{x}) \; .$$

This completes the proof of the theorem. △

Observe that the ε-subdifferential condition in Theorem 5.106(i) *cannot* be replaced with $0 \in \partial\varphi_\Omega(x)$; a counterexample is provided by the second function from Example 5.102.

The second assertion of Theorem 5.106 and subdifferential sum rules of the *equality* type imply the next result providing a *pointbased characterization* of linear subminimality in terms of basic subgradients of φ and basic normals to Ω calculated at the reference solution $\bar{x}$.

Corollary 5.107 (separated pointbased characterization of linear subminimality). *Let* $\dim X < \infty$, *and let* $\bar{x} \in \Omega$ *with* $|\varphi(\bar{x})| < \infty$. *Suppose also that one of the following assumptions* (a)–(c) *holds:*

(a) φ *is l.s.c. around* $\bar{x}$ *and* $\Omega = X$.

(b) φ *is strictly differentiable at* $\bar{x}$ *and* Ω *is closed around this point.*

(c) φ *is l.s.c. around* $\bar{x}$ *and lower regular at this point,* Ω *is locally closed and normally regular at* $\bar{x}$, *and one has the qualification condition*

$$\partial^\infty\varphi(\bar{x}) \cap \big(-N(\bar{x}; \Omega)\big) = \{0\} \; .$$

Then $\bar{x}$ *is linearly subminimal with respect to* (φ, Ω) *if and only if*

$$0 \in \partial\varphi(\bar{x}) + N(\bar{x}; \Omega) \; . \tag{5.114}$$

Proof. Condition (5.114) coincides with the one in Theorem 5.106(ii) when $\Omega = X$. When φ is strictly differentiable, condition $0 \in \partial\varphi_\Omega(\bar{x})$ is equivalent to (5.114) by the equality

$$\partial\varphi_\Omega(\bar{x}) = \nabla\varphi(\bar{x}) + N(\bar{x}; \Omega)$$

due to Proposition 1.107(ii). Under the assumptions in (c) we have the equality

$$\partial\varphi_\Omega(\bar{x}) = \partial\varphi(\bar{x}) + N(\bar{x}; \Omega)$$

due to the equality sum rule in Theorem 3.36. △

Note that in case (b) the characterization (5.114) of linear subminimality follows directly from Theorem 5.95(a) on multiobjective optimization, while in case (a) it follows Theorem 5.95(b) when φ is locally Lipschitzian. However, in case (c) the assumptions ensuring (5.114) by Corollary 5.107 are essentially weaker than those induced by Theorem 5.95(c). Indeed, the N-regularity assumption on $f(x) = \varphi(x) - \varphi(\bar{x})$ with $Z = I\!R$ in Theorem 5.95(c), which is the graphical regularity of φ at $\bar{x}$, is *equivalent* to the strict differentiability of φ at this point due to Proposition 1.94. On the other hand, the lower regularity of φ assumed in Corollary 5.107(c) holds for important classes of nonsmooth functions encountered in minimization problems. In particular, this includes convex functions and a broader class of amenable functions discussed above. Such a difference between the results of Theorem 5.95 in the case of minimization problems and the ones of Corollary 5.107 is due to the *one-sided* specific character of minimizing extended-real-valued functions, which is missed by separated conditions in the vector framework.

Based on the results of Corollary 5.107 in the constraint case $\Omega \neq X$, one may derive their consequences providing *necessary and sufficient* conditions for *linear subminimality* in problems with *specific types of constraints.* For problems with operator, functional, and/or equilibrium constraints (i.e., MPECs) it can be done as in Corollaries 5.96, 5.97, Theorem 5.98, and its two corollaries. Moreover, in addition to the above results requiring the strict differentiability of the objective mapping, we get also characterizations of linear subminimality in those problems with regular constraints and *lower regular* cost functions. We leave details to the reader.

Finally in this subsection, we obtain *necessary* conditions for *linear subminimality* in nonsmooth constrained problems, where *upper subgradients* are used for functions describing a single objective and inequality constraints.

Let us consider a cost function $\varphi_0 \colon X \to \overline{I\!R}$ finite at $\bar{x}$ and a constraint set $\Delta \subset X$ given by

$$\Delta := \big\{ x \in \Omega \subset X \ \text{ with } \ \varphi_i(x) \le 0 \ \text{ for } \ i = 1, \ldots, m \big\} ,$$

where $\varphi_i \colon X \to I\!R$ for all i. The next theorem gives upper subdifferential necessary conditions for linearly subminimal solutions with respect to (φ_0, Δ).

Theorem 5.108 (upper subdifferential necessary conditions for linearly subminimal solutions). *Let $\bar{x} \in \Delta$ be linearly subminimal with respect to (φ_0, Δ), where Ω is locally closed around $\bar{x}$. Assume that either X admits a Lipschitzian $\mathcal{C}^1$ bump function, or X is Asplund and $\varphi_i(\bar{x}) < 0$ for all $i = 1, \ldots, m$. Then for any Fréchet upper subgradients $x_i^* \in \widehat{\partial}^+ \varphi_i(\bar{x})$, $i = 0, \ldots, m$, there are $0 \neq (\lambda_0, \ldots, \lambda_m) \in I\!R^{m+1}$ such that*

$$\lambda_i \ge 0 \ \text{ for } \ i = 0, \ldots, m, \quad \lambda_i \varphi_i(\bar{x}) = 0 \ \text{ for } \ i = 1, \ldots, m, \quad \text{and}$$

$$-\sum_{i=0}^{m} \lambda_i x_i^* \in N(\bar{x}; \Omega) .$$

Proof. Suppose that $\widehat{\partial}^+\varphi_i(\bar{x}) \neq \emptyset$ for all $i = 0, \ldots, m$ (otherwise the conclusion of theorem holds trivially) and pick arbitrary $x^* \in \widehat{\partial}^+\varphi_i(\bar{x})$ for each i. Now applying the variational description of Fréchet subgradients from Theorem 1.88(i) to $-x_i^* \in \widehat{\partial}(-\varphi_i)(\bar{x})$, we find functions $s_i\colon X \to \overline{I\!R}$ for $i = 0, \ldots, m$ that are Fréchet differentiable at $\bar{x}$ satisfying

$$s_i(\bar{x}) = \varphi_i(\bar{x}), \quad \nabla s_i(\bar{x}) = x_i^*, \quad \text{and} \quad s_i(x) \geq \varphi_i(x) \quad \text{around} \quad \bar{x}\ .$$

Consider another constraint set

$$\widetilde{\Delta} := \big\{x \in \Omega \ \text{ with } \ s_i(x) \leq 0 \ \text{ for all } \ i = 1, \ldots, m\big\}$$

and observe that $\bar{x} \in \widetilde{\Delta}$ and that $\bar{x}$ is linearly subminimal with respect to $(\varphi_0, \widetilde{\Delta})$. Moreover, the definitions of linear subminimality and of Fréchet upper subgradients imply by the construction of s_0 that $\bar{x}$ is linearly subminimal with respect to $(s_0, \widetilde{\Delta})$. If $s_i(\bar{x}) = \varphi_i(\bar{x}) < 0$ for all $i = 1, \ldots, m$, we have by Corollary 5.97 with $f(x) = \varphi(x) - \varphi(\bar{x})$ and $\Theta = I\!R_-$, the necessary part of which clearly holds in any Asplund space (see Theorem 5.94(ii) and the subsequent arguments based on calculus rules in Asplund spaces), that

$$-\nabla s_0(\bar{x}) = -x_0^* \in N(\bar{x}; \widetilde{\Delta}) = N(\bar{x}; \Omega)\ .$$

It remains to consider the alternative case in the theorem when at least one of the inequality constraints is active at $\bar{x}$. In this case all the functions s_i may be chosen to be *continuously differentiable* around $\bar{x}$ by Theorem 1.88(ii) with $\mathcal{S} = \mathcal{LC}^1$. Then using again the necessary conditions for linear subminimality from Corollary 5.97 held in Asplund spaces, we get the inclusion

$$-\sum_{i=0}^{m} \lambda_i \nabla s_i(\bar{x}) \in N(\bar{x}; \Omega)$$

with some $(\lambda_0, \ldots, \lambda_m) \neq 0$ satisfying the above sign and complementary slackness conditions. The last relation in the theorem is now follows from $\nabla s_i(\bar{x}) = x_i^*$ for $i = 0, \ldots, m$. △

Specifying the constraint set Ω is the form of equality, operator, equilibrium, and/or other types of constraints and using the fully developed calculus, one may derive from Theorem 5.108 necessary conditions for linear suboptimality involving Fréchet upper subgradients of cost functions similarly to the upper subdifferential necessary conditions for minimization problems established in Sects. 5.1 and 5.2 of this chapter.

5.5 Commentary to Chap. 5

5.5.1. Two-Sided Relationships between Analysis and Optimization. This chapter is on applications of the basic tools of variational analysis developed in Volume I (Chaps. 1–4) to optimization and equilibrium problems. More specifically, we consider in this chapter a variety of problems in non-dynamic *constrained optimization* (including those with *equilibrium constraints*) and problems of *multiobjective optimization*, which cover classical and generalized concepts of *equilibrium*. Our main attention is devoted to deriving *necessary optimality* and *suboptimality conditions* of various types for the problems under consideration using the basic extremal/variational principles and the tools of generalized differentiation (with their comprehensive calculi) developed above.

It has been well recognized that *optimization/variational ideas* and techniques play a *crucial role* in all the areas of mathematical analysis, including those which seem to be far removed from optimization. Among the striking examples mentioned in Preface, recall the very first (Fermat) *derivative* concept the introduction of which was motivated by solving an optimization problem; the classical (Lagrange) *mean value* theorem, which is probably the most fundamental result of differential and integral calculi whose proof is based on the reduction to optimization and the usage of Fermat's stationary principle; and Bernoulli's *brachistochrone* problem, which actually inspired the development of all (infinite-dimensional) *functional analysis*.

Yet another powerful illustration of the mightiness of optimization is the *generalized differential calculus* developed in Volume I, which is strongly based on variational ideas, mainly on the *extremal principle*. Remember that the extremal principle provides *necessary conditions* for *set extremality*, which can be viewed as a geometric concept of optimality extending classical and generalized notions of optimal solutions to various optimization-related and equilibrium problems. Thus the application and specification of the extremal principle in concrete situations of constrained optimization and equilibria directly provide necessary optimality conditions in such settings. However, much more developed and diverse results can be derived while involving the power of generalized differential calculus together with the associated SNC calculus in infinite dimensions. This is the main contents of Chap. 5.

It is worth mentioning that the approach to necessary optimality conditions based on the extremal principle, as well as the extremal principle itself and its proof, essentially distinguish from the conventional approach to deriving necessary optimality conditions in constrained optimization, which was suggested and formalized by Dubovitskii and Milyutin [369, 370] and then was developed in many subsequent publications; see some reference and discussions in Subsect. 1.4.1. The Dubovitskii-Milyutin formalism contains the following *three* major components:

(a) to treat local minima via the *empty intersection* of certain sets in the *primal* space built upon the initial cost and constraint data;

(b) to approximate the above sets by *convex cones* with no intersection;

(c) to arrive at *dual* necessary optimality conditions in the form of an *abstract Euler equation* by employing *convex separation.*

The fundamental difference of our *extremal principle* approach from the formalism by Dubovitskii and Milyutin is the absence of *any convex approximation* in the primal space, while a generalized Euler equation is obtained via *nonconvex* constructions directly in the *dual* space by reduction to an approximating sequence of *smooth unconstrained* optimization problems; see Chap. 2.

5.5.2. Lower and Upper Subgradients in Nonsmooth Analysis and Optimization. Considering *minimization* problems for extended-real-valued functions, we distinguish in the results presented in this book between *lower subgradient* and *upper subgradient* optimality conditions. Conditions of these two kinds are significantly different for the case of *nonsmooth* cost functions and agree, of course, for smooth objectives as in Proposition 5.1. Note that the first result of the latter type for an arbitrary set $\Omega \subset X$ that admits a *convex cone* approximation K was obtained by Kantorovich [664] as early as in 1940 in the form

$$-\nabla\varphi(\bar{x}) \in K^*$$

via the dual/conjugate cone K^* to K in the general topological spaces X. Kantorovich's paper, published in Russian, was probably the first result of the general theory of extremal problems. Unfortunately, it didn't draw any attention either in the USSR or in the West being definitely ahead of its time. We refer the reader to the brilliant analysis by Polyak [1099] of this and other earlier developments on optimization, involving the related social environment, in the former Soviet Union.

In nonsmooth optimization, the concept of *subgradient* (or of *subdifferential* as a collection of subgradients) has been traditionally related to "lower" properties of nonsmooth functions and thus to minimization vs. maximization problems. On the other hand, subgradients/subdifferential of *concave* functions were defined by Rockafellar [1142] in the way different from (while symmetric to) that for convex functions. It corresponded in fact to what we now call *upper subgradients/subdifferential*; the latter terminology was explicitly introduced in Rockafellar and Wets [1165], although upper subgradient constructions were not actually employed in that book.

Another terminology, which has been fully accepted in the theory of viscosity solutions to nonlinear partial differential equations as well as in a number of publications on nonsmooth analysis, is that of "subdifferential" and "superdifferential." It is interesting to observe that (lower) subgradients are used to define viscosity "supersolutions," while "subsolutions" are defined via "supergradients." In this book we choose, after discussion with Rockafellar and Wets, to employ the *lower* and *upper* subgradient terminology as more natural and appropriate for optimization, taking "lower" for granted to describe subdifferential constructions extending the one for convex functions and using "upper" instead of "super" for symmetric constructions generalizing that for concave functions in the framework of convex analysis.

It is worth recalling to this end that Clarke's *generalized gradient* (good name!) on the class of locally Lipschitzian functions, being a *lower* subdifferential construction (i.e., extending the subdifferential for convex but not for concave functions), *coincides* at the same time with its *upper* subdifferential counterparts, due to its plus-minus symmetry $\partial_C(-\varphi)(\bar{x}) = -\partial_C\varphi(\bar{x})$ like for the classical gradient. This implies, in particular, that any conditions formulated via Clarke's generalized gradient, *don't* distinguished between minimization and maximization of nonsmooth (even convex) functions, between inequality constraints of the opposite signs, etc. However, as stated by Rockafellar [1142], "the theory of the maximum of a convex function relative to a convex set has an entirely different character from the theory of the minimum." In contrast, the lower and upper Fréchet-like and basic/limiting subdifferential constructions of this book are essentially *one-sided* and different from each other. We efficiently exploit these differences while deriving lower and upper subdifferential optimality conditions for constrained minimization of nonsmooth functions presented in Chap. 5.

5.5.3. Maximization Problems for Convex Functions and Their Differences. To the best of our knowledge, the first necessary optimality condition, which indeed distinguishes maximization and minimization, was obtained by Rockafellar [1142, Section 32] for the problem of *maximizing* a *convex* function φ over a convex set Ω in finite dimensions. This condition, for a *local* maximizer $\bar{x} \in \Omega$, was given in the *set-inclusion* form

$$\partial\varphi(\bar{x}) \subset N(\bar{x}; \Omega) \tag{5.115}$$

that obviously reduces to both inclusions (5.3) in Proposition 5.2 for the problem of *minimizing* the *concave* function $-\varphi$ over Ω. As mentioned in Subsect. 5.1.1, there is a very important class of *DC-functions*, represented as the difference of two convex functions $\varphi_1 - \varphi_2$, which can be reduced to minimizing concave function over convex sets. An analog of the necessary condition (5.115) for *DC*-functions reads as

$$\partial\varphi_1(\bar{x}) \subset \partial\varphi_2(\bar{x}); \tag{5.116}$$

see Hiriart-Urruty [573]. Then some modified versions of (5.115) and (5.116) were used to derive *necessary and sufficient* conditions for *global maximization* of convex functions, *DC*-functions, and closely related to them functions over convex sets; see particularly Strekalovsky [1226, 1227, 1228], Hiriart-Urruty [573], Hiriart-Urruty and Ledyaev [574], Flores-Bazán [461], Flores-Bazán and Oettli [462], and Tsevendorj [1272]. The reader can find more details and discussions on major achievements in this direction in the survey paper by Dür, Horst and Locatelli [373] and in the recent research by Ernst and Théra [410], where some other striking differences between maximizing and minimizing convex functions have been discovered. We also refer the reader to the recent study by Dutta [375] who derived *characterizations* of *global* maximizers for some classes of "pseudoconvex" and "quasiconvex" functions on convex sets in finite dimensions via Clarke's generalized gradient. Furthermore, he obtained *sufficient* conditions for global maximization of general Lipschitzian functions over such sets via our basic subdifferential constructions.

5.5.4. Upper Subdifferential Conditions for Constrained Minimization. A systematic study of *upper subdifferential* conditions for *constrained minimization* problems involving general (may be non-Lipschitzian) cost functions was conducted by Mordukhovich [925] in infinite-dimensional spaces. Most results of this type presented in Chap. 5 are taken from that paper. The results obtained seem to be new even in finite dimensions. They apply to *local* minimizers, same as more conventional *lower subdifferential* conditions, which are given in Chap. 5 in a parallel way. As discussed, these two kinds of necessary optimality conditions are generally *independent*, while upper subdifferential ones may be *essentially stronger* for some classes of minimization problems involving nonsmooth cost functions φ. Although the relation $\widehat{\partial}^+\varphi(\bar{x}) = \emptyset$ is itself an easy verifiable necessary condition for a local minimizer $\bar{x}$, the most efficient applications of upper subdifferential optimality conditions require the *nontriviality* $\widehat{\partial}^+\varphi(\bar{x}) \neq \emptyset$ of the Fréchet upper subdifferential. This is automatic, in particular, for those locally Lipschitzian functions on Asplund spaces that happen to be *upper regular* at the minimum point in question; see Remark 5.4 for more details. The latter class contains, besides smooth and concave continuous functions, a large class of *semiconcave* functions important in various applications, especially to *optimal control* and *viscosity solutions* to nonlinear PDEs.

Recall that a function $\varphi\colon \Omega \to I\!R$ defined on a convex set Ω is *semiconcave* if there is a nondecreasing upper semicontinuous function $\omega\colon I\!R_+ \to I\!R_+$ with $\omega(\rho) \to 0$ as $\rho \downarrow 0$ such that

$$\begin{aligned}&\lambda\varphi(x_1) + (1-\lambda)\varphi(x_2) - \varphi\big(\lambda x_1 + (1-\lambda)x_2\big)\\ &\le \lambda(1-\lambda)\,\|x_1 - x_2\|\,\omega\big(\|x_1 - x_2\|\big)\end{aligned} \tag{5.117}$$

whenever $x_1, x_2 \in \Omega$ and $\lambda \in [0,1]$; see the recent book by Cannarsa and Sinestrari [217] and the references therein. The most important case for both the theory and applications of semiconcavity corresponds to a *linear* modulus $\omega(\cdot)$ in (5.117). The latter class of functions (in an equivalent form and with the opposite sign) was probably introduced and employed for the first time in optimization by Janin [629] under the name "convexity up to a square" (or "presque convexes du deuxiéme ordre", PC2, in French). However, the origin of this construction goes back to partial differential equations, where the class of semiconcave (with a linear modulus) functions was exactly the one used by Kruzhkov [720] and Douglis [368] to establish the first global existence and uniqueness results for solutions to Hamilton-Jacobi equations. Furthermore, semiconcave functions have played a remarkable role in powerful uniqueness theories for generalized (viscosity, minimax, etc.) solutions to Hamilton-Jacobi and the like equations and their applications to optimal control and differential games; see particularly [85, 86, 216, 217, 295, 296, 297, 458, 471, 472, 789, 793, 1230] with the comprehensive bibliographies therein.

It is interesting to observe close relationships (in fact the equivalence) between semiconcave functions and the major subclasses of *subsmooth*, or *upper-*$\mathcal{C}^k$, functions introduced by Rockafellar [1151] in the form

$$\varphi(x) := \min_{t \in T} \phi(x,t) ,$$

where T is a compact space and where $\phi(x,t)$ is k times continuously differentiable in $x \in I\!R^n$ on an open set uniformly in $t \in T$; see also Penot [1069] for some infinite-dimensional extensions. As proved by Rockafellar [1151, 1165], the class of upper-$\mathcal{C}^2$ functions fully agrees with the class of functions semiconcave with a linear modulus, i.e., concave up to a square. The equivalence between the general class of semiconcave functions (5.117) and the class of upper-$\mathcal{C}^1$ functions was established by Cannarsa and Sinestrari [217]. Furthermore, upper-$\mathcal{C}^2$ functions happen to be equivalent to "weakly concave" functions in the sense of Vial [1286], while upper-$\mathcal{C}^1$ functions agree (in finite dimensions) with "approximately concave" ones considered by Ngai, Luc and Théra [1006]. We refer the reader to the recent paper by Aussel, Daniilidis and Thibault [63] for more discussions on these and related classes of nonsmooth functions and for the comprehensive study of associate geometric concepts.

Observe also that semiconcave functions with linear and more general moduli are closely related to functions called *paraconcave* in the theory of generalized convexity; see [534, 697, 1040, 1072] and the references therein. This name was suggested by Rolewicz [1169, 1170] who independently introduced and studied paraconvexity/paraconcavity in the framework of set-valued mappings. A strong interest to such functions has been motivated by approximation and regularization procedures via the *infimal convolution* and related operations, which have been proved to be locally $\mathcal{C}^{1,1}$ in many important cases due to the following characterization first established by Hiriart-Urruty and Plazanet [576]: *a function is* $\mathcal{C}^{1,1}$ *around* $\bar{x}$ *if it simultaneously paraconvex and paraconcave around this point.* We particularly refer the reader to the papers by Eberhard et al. [381, 386, 387] for various applications of this result to *second-order generalized differentiation.*

5.5.5. Lower Subdifferential Optimality and Qualification Conditions for Constrained Minimization. In contrast to upper subdifferential conditions for nonsmooth minimization, their *lower subdifferential* counterparts are more conventional, with a variety of modifications, and have a much longer history. Of course, for optimization problems with *smooth* data both lower and upper subdifferential necessary optimality conditions reduce to classical results of constrained optimization that go back to the standard versions of Lagrange multipliers in the *qualified* (sometimes called normal or Karush-Kuhn-Tucker) and *non-qualified* (sometimes called Fritz John) forms. Results of the first type contain *qualification conditions*, which ensure the nontriviality ($\lambda_0 \neq 0$) of a multiplier corresponding to the objective/cost function. We refer the reader to the fundamental contributions by Lagrange [737], Karush [665] (published in the survey paper by Kuhn [723]), John [638], Kuhn and Tucker [724], and Mangasarian and Fromovitz [841] for the origin of such optimality and qualification conditions and the main motivations behind them. Further developments with more detailed historical accounts and various applications can be particularly found in [7, 9, 111, 112, 89, 158, 163, 164, 249, 255, 370, 376, 432, 499, 504, 512, 544, 571, 588, 595, 602, 618, 707, 718, 801, 824, 860, 840, 892, 902, 962, 1009, 1097, 1119, 1152, 1155, 1160, 1165, 1216, 1256, 1264, 1265, 1267, 1268, 1289, 1315, 1319, 1340, 1341, 1373, 1378] and the numerous references therein.

Note that the qualification conditions for *optimization* given in Chap. 5 have the same nature as the qualification conditions obtained in Volume I from the viewpoint of *generalized differential calculus*; they are very much interrelated. Furthermore, both optimality and qualification conditions of this book are derived in *dual spaces* being generally *less restrictive* than their primal space counterparts. Thus the *common dual space structure* of these optimality and qualification conditions allows us to make a natural *bridge* between the optimization results of the qualified and non-qualified types developed in Chap. 5.

In this book we concentrate on *first-order* necessary optimality (as well as suboptimality) conditions for various classes of optimization problems. However, we use not only first-order but also *second-order* subdifferential constructions for problems with *equilibrium constraints*, which is due to the first-order *variational* nature of such constraints; see Sects. 5.2 and 5.3 and the corresponding comments to them given below. The reader can find more information on second-order optimality conditions in [37, 64, 65, 102, 111, 132, 133, 153, 176, 234, 236, 282, 283, 372, 384, 387, 502, 575, 486, 516, 601, 613, 624, 628, 704, 756, 764, 771, 857, 858, 877, 1037, 1038, 1039, 1067, 1092, 1156, 1165, 1307, 1308, 1310, 1337, 1358] and their bibliographies.

5.5.6. Optimization Problems with Operator Constraints. The material of Subsect. 5.1.2 is devoted to necessary optimality conditions of both *lower* and *upper* subdifferential types for minimization problems with the so-called *operator constraints* defined in the general form $x \in F^{-1}(\Theta) \cap \Omega$ via inverse images/preimages of sets under set-valued mappings. Traditionally operator constraints are defined in the equality form $f(x) = 0$, where $f\colon X \to Y$ is a single-valued mapping with an *infinite-dimensional* range space Y. This name appeared (probably first in the Russian literature) from the observation that dynamic constraints in typical problems of the calculus of variations and optimal control can be written in such a form, where f is a certain differential or integral operator into an infinite-dimensional space; see, e.g., Dubovitskii and Milyutin [370].

It seems that the first general result for such problems in an infinite-dimensional form of Lagrange multipliers was obtained by Lyusternik in his seminal work [824], where f is a $\mathcal{C}^1$ operator between Banach spaces. To establish this result, Lyusternik developed his now classical iterative process and arrived at the "distance estimate," which is nowadays called *metric regularity*. Lyusternik's version of the Lagrange principle (in the *qualified* form) was obtained under the *Lyusternik regularity condition* $\ker\nabla f(\bar{x})^* = \{0\}$, which signifies the *surjectivity* of the derivative operator $\nabla f(\bar{x})\colon X \to Y$. It is not difficult to derive from Lyusternik's qualified necessary optimality condition the non-qualified version

$$\lambda\nabla\varphi(\bar{x}) + \nabla f(\bar{x})^* y^* = 0, \quad \lambda \ge 0, \;\; (\lambda, y^*) \ne 0\;, \tag{5.118}$$

of the Lagrange multiplier rule for a local minimizer $\bar{x}$ of a smooth function φ subject to the operator constraint $f(x) = 0$ *provided* that the derivative image $\nabla f(\bar{x})X$ is *closed* in Y; see, e.g., Ioffe and Tikhomirov [618]. As well known, the multiplier rule (5.118) *doesn't generally hold*, even in the simplest infinite-dimensional case of $Y = \ell^2$ for smooth problems, without the latter closedness assumption.

First necessary optimality conditions for problems of minimizing a cost function $\varphi_0(x)$ subject to *nonsmooth* operator constraints $f(x) = 0$ given by a Lipschitzian mapping $f\colon X \to Y$ between Banach spaces, together with more standard constraints

$$\varphi_i(x) \le 0, \;\; i = 1, \ldots, m, \;\; \text{and} \;\; x \in \Omega\;,$$

were obtained by Ioffe [595], via Clarke's generalized gradient and normal cone, in the generalized Lagrange form

$$0 \in \partial_C\Big(\sum_{i=0}^{m}\lambda_i\varphi_i + \langle y^*, f\rangle\Big)(\bar{x}) + N_C(\bar{x};\Omega), \quad (\lambda_0, \ldots, \lambda_m, y^*) \ne 0\;, \tag{5.119}$$

accompanied by the usual sign and complementary slackness conditions. Besides the conventional local Lipschitzian property of φ_i, $i = 0, \ldots, m$, it was assumed in [595] that: Y has an equivalent norm whose dual is strictly convex; Ω has a certain "tangential lower semicontinuous property" at $\bar{x}$ formulated in terms of Clarke's tangent cone and directional derivative; and f has a "strict prederivative" with norm compact values satisfying a version of the "finite codimension property" relative to $T_C(\bar{x};\Omega)$. This result was improved by Ioffe [598] and by Ginsburg and Ioffe [506] who established significantly stronger counterparts of (5.119), with the usage of the "approximate" subdifferential and normal cone instead of the convex-valued constructions by Clarke, under much more subtle versions of the finite codimension property formulated via the above "approximate" normal and subgradient constructions. Note that the latter *advanced* formulations of the *finite codimension* property happened to be closely related to a *topological* counterpart of the *partial* sequential normal compactness (PSNC) property for mappings as well as to the partial CEL property by Jourani and Thibault [655]; see Subsects. 1.2.5, 4.5.4 and the corresponding discussions in Ioffe [607].

5.5.7. Operator Constraints via Basic Calculus. Theorem 5.11 giving *non-qualified* necessary conditions in both *upper* and *lower* subdifferential forms for general problems with operator constraints was obtained in Mordukhovich [925], while its *qualified* counterparts from Theorems 5.7 and 5.8 are new. Observe that the qualified optimality conditions *imply* in fact the corresponding non-qualified ones, but *not vice versa*. This is due to the structure of the qualification conditions in Theorems 5.7 and 5.8 (as well as in the subsequent necessary optimality conditions presented in the book), which usually contain *more subtle* dual-space information than is needed for the associated non-qualified optimality conditions. Note also that the developed SNC calculus allows us to derive a variety of normal compactness-like requirements, generally less restrictive than the afore-mentioned finite codimension property, ensuring the fulfillment of *pointbased* necessary optimality conditions for problems with operator constraints.

It is worth mentioning that in our approach to necessary optimality conditions we treat operator constraints as *geometric constraints* and then employ generalized differential and SNC calculi to derive results via the initial data. The presence of *both* these calculi based on the *extremal principle*, being characteristic for the basic constructions used in the book, undoubtedly happens to be *the most crucial factor* for successful implementing our approach. Note that this approach doesn't have any restriction to deal with *many* geometric constraints, which is significant for various classes of optimization problems, in particular, for optimal control; see Chaps. 6 and 7. As well known, the presence of *only one* geometric constraint with possibly empty interior (or that of the operator/equality type) has been a substantial obstacle in the Dubovitskii-Milyutin formalism and its subsequent developments.

To conclude the discussion around Theorems 5.7, 5.8, and 5.11, let us comment on those parts of assertions (i) of Theorems 5.7 and 5.11 that *don't* impose the *strict* (or continuous) differentiability assumptions on the *equality type* constraints with *values* in *finite-dimensional* spaces. These results are essentially due to calculating the Fréchet normal cone to inverse images given in Corollary 1.15, which is based on the *Brouwer fixed-point theorem*; cf. also Halkin [543] and Ioffe [595]. Results of this type were developed by Ioffe [595, 602] and Ye [1340, 1341] to derive necessary optimality conditions for *Lipschitzian* problems with finitely many equality and inequality constraints via small *convex-valued* subdifferentials (of Michel-Penot's [870, 871] and Treiman's [1264, 1265] types) that don't possess *any robustness* property. Note that the corresponding results of Theorems 5.7(i) and 5.11(i) don't require the local Lipschitz continuity of constraint functions, while still imposing the *continuity* requirement on equality constraint functions *around* the point in question. The latter requirement is *essential* for the validity of Lagrange-type necessary optimality conditions as demonstrated by Example 5.12, which is due to Uderzo [1274].

5.5.8. Exact Penalization and Weakened Metric Regularity. The remainder of Subsect. 5.1.2 concerns another method to deal with minimization problems involving operator constraints of the classical equality type $f(x) = 0$ given by Lipschitzian mappings. This method known as *exact penalization* goes back to Eremin [406] and Zangwill [1354] in the context of convex programming and has been well developed in connection with numerical optimization; see, e.g., Bertsekas [111], Burke [188, 189], Polyak [1097], and the references therein. Regarding applications to necessary optimality conditions in nonsmooth optimization, this method was first suggested probably by Ioffe [588] who established Theorem 5.16; cf. also a somewhat different result by Clarke [249, 255] on exact penalization that didn't specifically address operator constraints. We refer the reader to the recent book by Demyanov [318] and its bibliography for various applications of exact penalization techniques to necessary optimality conditions in problems of constrained optimization, the calculus of variations, and optimal control.

The main concept implemented in Theorem 5.16 is *regularity at a point* (called *weakened metric regularity* in Definition 5.15) introduced by Ioffe in [587]. This notion, which is closely related to *subregularity* in the terminology by Dontchev and Rockafellar [366], is generally different from the basic concept of metric regularity *around* the point used throughout the book. The weakened metric regularity is *not robust* and doesn't allow adequate characterizations as well as calculus/preservation properties similar to the basic metric regularity. At the same time, this weakened metric regularity and the associated (inverse) notion of *calmness* happen to be convenient for various applications; see more comments below in Subsect. 5.5.16.

Theorem 5.17 giving *lower* subdifferential optimality conditions for Lipschitzian problems with *equality operator constraints* and its Corollary 5.18 providing an efficient specification for operator constraints of the *generalized Fredholm* type are new. In comparison with the afore-mentioned results by Ioffe [598] and by Ginsburg and Ioffe [506] discussed in Subsect. 5.5.6, the new results impose *milder* sequential normal compactness assumptions than the finite codimension property and employ *smaller* sets of subgradients and normals. On the other hand, our results require the Asplund (generally nonseparable) space structure of both spaces X and Y, while those in [598, 506] apply to arbitrary Banach spaces X and to (close to separable) spaces Y admitting an equivalent norm whose dual is strictly convex.

Note also that the *strict Lipschitzian* assumption on the operator constraint mapping f in Theorem 5.17, which is milder than the strict prederivative assumption on f imposed in [506, 595, 598], can be relaxed to the merely local Lipschitz continuity of f, but in this case the basic subdifferential of the scalarization $\partial\langle y^*, f\rangle(\bar{x})$ in the multiplier rule of Theorem 5.17 should be replaced with the (larger) *normal coderivative* $D_N^* f(\bar{x})(y^*)$. Observe that such a *coderivative form* doesn't have any counterparts in terms of Clarke's constructions even in finite dimensions.

5.5.9. Necessary Optimality Conditions in the Presence of Finitely Many Functional Constraints. Subsection 5.1.3 concerns the more conventional form (5.23) of mathematical programs with *finitely many* equality, inequality, and geometric constraints. Such constrained optimization problems are specifications of those with operator constraints considered in Subsect. 5.1.2, while the specific form (5.23) allows us to develop a *greater variety* of methods and results on necessary optimality conditions.

The *upper subdifferential* conditions of Theorem 5.19 are partly new and partly taken from Mordukhovich [925]. Note that the optimality conditions of Theorem 5.19(i) employ Fréchet upper subgradients not only for the cost function φ_0 as in Subsect. 5.2.1 but also for the functions φ_i, $i = 1, \ldots, m$, describing the *inequality constraints* in (5.23). This however requires a special "smooth bump" structure of the space X in question for applying the needed *smooth variational descriptions* of Fréchet subgradients that happen to be crucial in the proof.

The subsequent results of Subsect. 5.1.3 deal with *lower subdifferential* conditions for the nondifferentiable programming problem (5.23), which include not only those via lower subgradients of the cost and inequality constraints functions but also necessary optimality conditions expressed in terms of *generalized normals* to the corresponding epigraphs.

We start with such *geometric* conditions in assertions (i) and (ii) of Theorem 5.21, which give both *approximate/fuzzy* and *exact/pointbased* forms of necessary optimality conditions for (5.23) in the general *Asplund space* framework derived by a *direct application* of the corresponding form of the *extremal principle* with *no Lipschitzian* assumptions on the functions involved. These results in full generality were first presented in Mordukhovich [922], but in fact the results as well as the methods employed go back (sometimes as *transversality conditions* in optimal control) to the original publications by Mordukhovich [887, 889, 892] and by Kruger and Mordukhovich [717, 718, 719], where necessary optimality conditions of this type were established for various specifications of (5.23) in finite-dimensional and Fréchet smooth spaces.

The *subdifferential form* of the pointbased conditions as given in Theorem 5.21(iii), with the replacement of basic normals to epigraphs by basic subgradients of the corresponding functions under their local *Lipschitz continuity*, can be also found in the afore-mentioned papers in the finite-dimensional, Fréchet smooth, and Asplund space frameworks. Note that in [706, 707] Kruger obtained an extension of these results to problems with *infinitely many inequality* constraints given in the inclusion form $f(x) \in \Theta$, where f is a single-valued Lipschitzian mapping while $\Theta \subset Y$ is an *epi-Lipschitzian* subset of a Fréchet smooth space. The latter requirement reduces to int $\Theta \neq \emptyset$ when Θ is convex; that is where the name "infinite many inequalities" comes from. Such "inequality type" results can be derived from Theorem 5.8(iv) under the *much milder* SNC assumption on Θ in Asplund spaces.

Let us next discuss the treatment of the *equality constraints*

$$\varphi_i(x) = 0, \quad i = m+1, \ldots, m+r \;,$$

in problem (5.23), which is the same for the upper and lower subdifferential conditions of Subsect. 5.1.3 being *significantly different* from that for the inequality constraints as well as for the cost function under consideration. When φ_i, $i = m+1, \ldots, m+r$, are locally *Lipschitzian*, the equality constraints can be reflected in the necessary optimality conditions by the "condensed term"

$$\partial\Big(\sum_{i=m+1}^{m+r} \lambda_i \varphi_i\Big)(\bar{x}), \quad (\lambda_{m+1}, \ldots, \lambda_{m+r}) \in I\!R^r \;, \tag{5.120}$$

via the basic subdifferential of the sum $\lambda_{m+1}\varphi_{m+1} + \ldots + \lambda_{m+r}\varphi_{m+r}$ with *arbitrary* (no sign) Lagrange multipliers; see, in particular, condition (5.27) in Theorem 5.19. Since λ_i are *not* nonnegative in (5.120) and since the basic subdifferential ∂ is a *one-sided* construction while satisfying a subdifferential sum rule, we can replace (5.120) by the *larger sum* of sets

$$\sum_{i=m+1}^{m+r} \lambda_i \partial^0 \varphi_i(\bar{x}), \quad (\lambda_{m+1}, \ldots, \lambda_{m+r}) \in I\!R^r \;,$$

formed via the *symmetric subdifferentials* $\partial^0\varphi_i(\bar{x}) = \partial\varphi_i(\bar{x}) \cup \partial^+\varphi_i(\bar{x})$ of the separate equality constraint functions φ_i, but not just via the basic subdifferentials $\partial\varphi_i(\bar{x})$; cf. Corollary 5.20. We prefer however to use the *more exact* while less conventional subdifferential expression with *all* the *nonnegative* multipliers

$$\sum_{i=m+1}^{m+r} \lambda_i \Big[\partial\varphi_i(\bar{x}) \cup \partial(-\varphi_i)(\bar{x})\Big] \;\text{ with }\; \lambda_i \geq 0, \quad i = m+1, \ldots, m+r \;,$$

reflecting the equality constraints in necessary optimality conditions for (5.23) and related problems; see Theorem 5.21(iii) and its proof that contains, by taking into account inclusion (5.32), the derivation of the latter expression from the geometric conditions $(\bar{x}, -\lambda_i) \in N((\bar{x}, 0); \operatorname{gph}\varphi_i)$ in Theorem 5.21(ii) when the constraint functions φ_i are locally *Lipschitzian* around $\bar{x}$ as $i = m+1, \ldots, m+r$. This *significantly distinguishes* the lower subdifferential optimality conditions of Theorem 5.21(iii) from other versions of Lagrange multiplier rules in nondifferentiable programming, particularly from those established by Clarke [249] and Warga [1319] in terms of their *two-sided* subdifferential constructions that equally treat inequality and equality constraints; see Remark 5.22 for more discussions and illustrative examples.

5.5.10. The Lagrange Principle. The next topic of Subsect. 5.1.3 relating to lower subdifferential conditions for constrained optimization problems of type (5.23) concerns *nonsmooth* extensions of the so-called *Lagrange principle*. This name was suggested by Tikhomirov (see, in particular, his books with Ioffe [618], with Alekseev and Fomin [7], and with Brinkhuis [178]) who observed that necessary optimality conditions in many extremal problems arising in various areas of mathematics and applied sciences (nonlinear programming, calculus of variations, optimal control, approximation theory, inequalities, classical mechanics, astronomy, optics, etc.) could be obtained in the following scheme: define the *Lagrangian* $L(x, \lambda_0, \ldots, \lambda_{m+r})$ by formula (5.35) with multipliers $(\lambda_0, \ldots, \lambda_{m+r})$ corresponding to the cost function and to all the *functional* (equality and inequality type) constraints and then consider the problem of minimizing the Lagrangian subject to the remaining *geometric constraints*. The Lagrange principle says, in accordance with the primary idea of Lagrange [737], that necessary optimality conditions for the original constrained problem can be derived as necessary optimality conditions for minimizing the Lagrangian subject *only* to the *geometric constraints* (i.e., fully *unconstrained* if there are no geometric constraints in the original problem) with some nontrivial set of Lagrange multipliers.

Of course, the validity of the Lagrange principle should be justified for each class of optimization problems under consideration. Ioffe and Tikhomirov did this in their book [618] (originally published in Russian in 1974) for extremal problems with the so-called "smooth-convex" structure, which cover problems of optimal control involving smooth dynamics, state constraints of the inequality type, and general geometric constraints on control functions.

The first *nonsmooth* version of the Lagrange principle was obtained by Hiriart-Urruty [571] for Lipschitzian problems of type (5.23) via Clarke's generalized gradient and normal cone. Further results on the nonsmooth Lagrange principle were developed by Ioffe [595] for problems with operator constraints via Clarke's constructions (see Subsect. 5.5.6) and then by Kruger [707], Mordukhovich [897, 901], and by Ginsburg and Ioffe [506] in terms of nonconvex subdifferential constructions.

The results of Lemma 5.23 and Theorem 5.24 are new; some special cases and consequences can be found in [707, 708, 897, 901]. Corollary 5.25 on the "abstract maximum principle" reveals the fact well understood in variational analysis that *maximum-type* optimality conditions relate to the *convexity* of geometric constraints by an *extremal* structure of the normal cone to convex sets. Note to this end that the maximum principle in optimal control of continuous-time systems *doesn't* generally require any explicit convexity assumptions due to a certain "hidden convexity" inherent in such systems; see Chap. 6 for more details and discussions.

5.5.11. Mixed Multiplier Rules. It has been well recognized in optimization theory that *equality* and *inequality constraints* are of a fundamentally different nature, and hence they should be treated differently. As seen above, equality and inequality constraints in nonsmooth optimization problems can be distinguished by using *different subgradient sets* in the corresponding necessary optimality conditions. Note that the cost function is usually treated in necessary conditions as those describing inequality constraints.

Theorem 5.26 presents *lower subdifferential* optimality conditions of yet another type for problem (5.23) that significantly distinguish between the equality and inequality constraints in this problem. The essence of this theorem, first established by Mordukhovich [897, 901] in finite dimensions, is that it provides a *mixed* subdifferential generalization of the Lagrange multiplier rule. Indeed, while our basic *robust* subdifferential (an extension of the *strict* derivative) is used for *equality* constraints in Lagrangian necessary optimality conditions, a *non-robust* extension of the classical derivative is employed for *inequality* constraints and the *objective* function.

The notion of the "upper convex approximation" used in Theorem 5.26 and the generated "p-subdifferential" construction (5.47) are due to Pshenichnyi [1108, 1109]. Observe that these objects are defined *non-uniquely* and generally non-constructively. One of the possible upper convex approximations is Clarke's directional derivative, which is usually *not the best* one as demonstrated in the afore-mentioned work by Pshenichnyi. On the other hand, it is easy to show that any *Gâteaux* differentiable function admits the best upper convex approximation via its Gâteaux derivative (see the discussion after Theorem 5.26), which thus provides a version of the Lagrange multiplier rule generally *independent* on the previous necessary optimality conditions of Subsect. 5.3.1.

5.5.12. Necessary Conditions for Problems with Non-Lipschitzian Data. As seen from the results and discussions given above, all the *lower subdifferential* versions of necessary optimality conditions in a generalized form of Lagrange multipliers for the problem of nondifferentiable programming (5.23) were derived under the local *Lipschitzian* assumption on the functions φ_i, $i = 0, \ldots, m + r$, describing the objective and functional constraints. There are also results on Lagrange multipliers in the classical differential form assuming only *differentiability* but not *strict/continuous* differentiability (i.e., generally not the local Lipschitz continuity) of functional data partly discussed above; see Theorems 5.7(i) and 5.11(i) as well as the papers by Halkin [543], Ioffe [595], and Ye [1340, 1341]. At the same time Theorem 5.21(ii) gives necessary conditions for problem (5.23) at the reference minimizer $\bar{x}$ with *no Lipschitzian* assumptions but *not* in a conventional *subdifferential* form: it involves *basic normals* to *graphs* and *epigraphs*, i.e., eventually not only basic but also *singular subgradients* of the cost and constraint functions.

Alternative lower subdifferential optimality conditions for *non-Lipschitzian* problems in the *approximate/fuzzy* form of Theorem 5.28 were first obtained by Borwein, Treiman and Zhu [158] in *reflexive* spaces, where the reflexivity of the space in question was essentially used in the proof; see also Borwein and Zhu [163, 164]. The *Asplund space* version of such weak fuzzy optimality conditions were independently derived with different proofs by Mordukhovich and B. Wang [962] and by Ngai and Théra [1009]. The proof given in the book is taken from [962]. Results of this type were also obtained by Zhu [1373] for nondifferentiable programming problems in Banach spaces admitting *smooth renorms* with respect to some bornology.

5.5.13. Suboptimality Conditions. The last subsection of Sect. 5.1 is devoted to *suboptimality* conditions for constrained optimization problems. By such results we understand those, which justify the fulfillment of *almost necessary optimality conditions for almost optimal solutions*, where “almost” means “up to an arbitrary $\varepsilon > 0$.”

It seems to be clear that from the viewpoint of practical applications, as well as from that of justifying numerical algorithms based on necessary conditions, there are *no much difference* between necessary *optimality* and *suboptimality* conditions. The main advantage of suboptimality vs. necessary optimality conditions is that dealing with suboptimality allows us to avoid principal difficulties with the *existence* of *optimal solutions* that may either *not exist* (especially in infinite dimensions), or it may be hard to verify their existence.

The importance of suboptimality conditions has been well recognized in the classical calculus of variations after the seminal publications by Young [1349, 1350] and McShane [861, 862]. Recall that the principal purpose of those fundamental developments was not only to construct variational problems admitting optimal solutions in the class of “generalized curves” that can be approximated by *suboptimal* solutions to the original problem, but also to establish necessary optimality conditions for generalized curves that happened to provide suboptimality conditions for *minimizing sequences* of “ordinary curves.”

This line of development was continued in optimal control theory by Gamkrelidge [495, 496, 497] and Warga [1313, 1314, 1315] who independently constructed a proper *relaxation* (the term coined by Warga) of the original control problem using somewhat different but equivalent *convexification* procedures and eventually obtaining suboptimality conditions for *minimizing sequences* of original controls via necessary optimality conditions and approximations of their generalized/relaxed counterparts; see also Ioffe and Tikhomirov [617], McShane [863], and Young [1351] for discussing relationships of these approaches and results with the classical calculus of variations. Suboptimality conditions for dynamic optimization and control problems of various kinds were later derived, *without* employing any relaxation procedures, by Gabasov, Kirillova and Mordukhovich [488], Gavrilov and Sumin [500], Medhin [867], Mordukhovich [901], Moussaoui and Seeger [987], Plotnikov and

Sumin [1084], Seeger [1199], Sumin [1233, 1234], and Zhou [1367, 1368, 1369] among other researchers.

For general (not necessarily dynamic) optimization problems in Banach spaces the first suboptimality conditions were established by Ekeland [396, 397, 399] via his powerful *variational principle.* As mentioned in [397], such *suboptimality* issues were among the *primary motivations* for developing Ekeland's variational principle. Concerning problems of mathematical programming with equality and inequality constraints as in (5.23), Ekeland derived in [397] qualified suboptimality conditions of the ε-multiplier type under *smoothness* assumptions on the initial data imposing *linearly independence* condition on the gradients of all the constraint functions; this is a stronger qualification condition than that of Mangasarian and Fromovitz.

Based on the Ekeland variational principle and necessary optimality conditions in appropriately *perturbed* problems, *lower* subdifferential suboptimality conditions were developed in both qualified and non-qualified forms for various classes of nonsmooth optimization problems by using mostly the generalized differential constructions of Clarke. The reader can find a number of results and applications in this direction in the research by Attouch and Wets [47], Auslender and Teboulle [60], Bustos [207], Dutta [374], Gupta, Shiraishi and Yokoyama [526], Hamel [546], Kusraev and Kutateladze [733], Loridan [811], Loridan and Morgan [812], Moussaoui and Seeger [986], and their references.

The results presented in Subsect. 5.5.13 are taken from the paper by Mordukhovich and B. Wang [962] based on the application of the *lower subdifferential variational principle* from Theorem 2.28 and appropriate techniques of the generalized differential calculus. We distinguish between two major types of suboptimality conditions: the *weak* conditions from Theorem 5.29 and the *strong* ones from Theorem 5.30 and its corollaries given in both qualified and non-qualified forms.

The *weak suboptimality conditions* of Theorem 5.29 don't practically impose any assumptions on the initial data in the *Asplund space* setting (besides the minimal local requirements on lower semicontinuity of the cost and inequality constraint functions, continuity of those describing the equality constraints, and closedness of the geometric constraint), but the results obtained involve a *weak*[*] *neighborhood* $V^* \subset X^*$ of the origin providing a *weak fuzzy* counterpart of the Lagrange multiplier rule expressed via Fréchet normals and subgradients near the reference minimizer.

In contrast, the *strong suboptimality conditions* in both the *qualified* form of Theorem 5.30 and the *non-qualified* form of Corollary 5.32 establish a more appropriate *strong* version of the *approximate Lagrange multiplier rule*, with a *small dual ball* $\varepsilon I\!B^*$ replacing the weak* neighborhood V^* from Theorem 5.29, expressed via basic normals and subgradients under additional Lipschitzian and SNC assumptions. We particularly note the result of Corollary 5.31, which provides strong suboptimality conditions for *smooth* problems of nonlinear programming under the classical *Mangasarian-Fromovitz constraint qualification.*

5.5.14. Mathematical Programs with Equilibrium Constraints. The general class of constrained optimization problems studied in Sect. 5.2 is now known as *mathematical programs with equilibrium constraints* (MPECs). This name appeared in the book by Luo, Pang and Ralph [820] containing a variety of qualitative and numerical results as well as practical applications for this remarkable class of mathematical programs in finite dimensions. Another (nonsmooth) approach to the study of MPECs and related optimization problems was developed in the book by Outrata, Kočvara and Zowe [1031].

Historically MPECs have their origin in the economic literature of the 1930s concerning problems of *hierarchical optimization* known now as *Stackelberg games*; see the book by von Stackelberg [1222] for the initial motivations and applications and the paper by Leitmann [758] for a modern revisiting. This class of hierarchical problems is closely related to *bilevel programming*, where the focus is on two-level mathematical programs interrelated in such a way that the set of optimal solutions to the lower-level parametric problems is the set of feasible solutions to the upper-level one. The reader can find more results, references, and discussions on bilevel programming in the book by Dempe [316], his comprehensive (till 2003) annotated bibliography [317], and the recent paper by Dutta and Dempe [377].

It is often appropriate to consider in hierarchical optimization not just optimal solutions to the lower-level problem but a larger set of the so-called *KKT points*, which contains the collection of optimal (or stationary) solutions together with the corresponding Lagrange multipliers from the first-order optimality conditions. In such a way the description of the feasible solution set to the upper-level problem involves the classical *complementary slackness conditions* for mathematical programs with inequality constraints. Conditions of this type are of great importance for their own sake; they have been studied for years in *complementarity theory* well developed in mathematical programming; see, e.g., the book by Cottle, Pang and Stone [294] and the recent two-volume monograph by Facchinei and Pang [424] for comprehensive studies of various classes of complementarity problems in finite-dimensional spaces. Considering complementarity conditions in the (lower-level) framework of hierarchical optimization gives rise to the study of *mathematical problems with complementarity constraints* (MPCC), which is one of the most significant parts of both MPEC theory and applications.

On the other hand, there are important classes of MPECs for which feasible solutions are given by more general conditions than complementarity; in particular, those defined by parametric *variational inequalities*; see, e.g., the afore-mentioned book [424]. It has been well recognized that the most natural and convenient setup for describing feasible solutions to MPECs, which covers the previously mentioned settings as well as other remarkable classes of non-conventional mathematical programs, is Robinson's framework of parametric *generalized equations* (5.53). This way is proved to be appropriate for defining not only sets of optimal solutions/KKT points to lower-level optimization, complementarity, and variational inequality problems but also for various type of *equilibria* arising in economics, mechanics, and other applied sciences. Thus it fully justifies the name "equilibrium constraints" widely spread in the optimization-related literature.

A characteristic feature of both MPCCs and MPECs is the presence of *intrinsic nonsmoothness*, even for problems with smooth initial data. Such a nonsmoothness is sometimes hidden while still playing a *crucial role* in the theory and numerical algorithms. It is not thus surprising that the methods of nonsmooth analysis and generalized differentiation happen to provide fundamental tools in developing various theoretical and computational aspects of MPCCs and MPECs, particularly those related to necessary optimality conditions, sensitivity and stability analysis, convergence rate and error estimates.

The usage of appropriate concepts of generalized differentiation and associated calculi gives rise to the corresponding notions of *stationarity* important for the MPEC theory and applications: particularly ***B***(ouligand)-stationarity, ***C***(larke)-stationarity, and ***M***(ordukhovich)-stationarity. The latter notion has drawn a major attention in recent years due to some practical applications (especially to mechanical equilibria) and requiring the *weakest constraint qualifications* as a first-order necessary optimality condition for MPECs. The reader can find various qualitative and numerical results dealing with MPEC stationarity in Anitescu [20], Facchinei and Pang [424], Flegel [454], Flegel and Kanzov [455, 456], Flegel, Kanzov and Outrata [457], Fukushima and Pang [480], Hu and Ralph [584], Kočvara, Kružik and Outrata [689], Kočvara and Outrata [690, 691], Outrata [1024, 1025, 1026, 1027, 1028, 1029, 1030], Ralph [1116], Ralph and Wright [1117], Scheel and Scholtes [1191], Scholtes [1192], Scholtes and Stöhr [1194], Treiman [1268], Ye [1338, 1339, 1342], Ye and Ye [1343], Zhang [1361], etc.

5.5.15. Necessary Optimality Conditions for MPECs via Basic Calculus. The approach to necessary optimality conditions for general MPECs and their specifications developed in Subsects. 5.2.1 and 5.2.2 is based on considering first *abstract* MPECs of type (5.52) with equilibrium constraints given by general set-valued mappings $y \in S(x)$, then reducing them to mathematical programs with only *geometric* constraints studied in Subsect. 5.1.1 while defined in *product* spaces, and finally using generalized differential and SNC calculi involving our basic normal, coderivative, and subdifferential constructions. Let us emphasize that this approach to derive necessary optimality conditions for general (as well as for more specified) MPECs allows us to *avoid* well-recognized *obstacles* in the study of MPECs, which occur while employing various conventional methods of reducing MPECs to usual mathematical programs when, however, standard constraint conditions are *not satisfied* even in the case of simple MPCCs with smooth data; see, e.g., Ye [1338, 1339] for more references and discussions.

Most of the *lower* and *upper* subdifferential optimality conditions for general MPECs and their specifications presented in Subsects. 5.2.1 and 5.2.2 are taken from the recent paper by Mordukhovich [911]; some of them are new.

Note that necessary optimality and qualification conditions of the lower subdifferential type were previously developed by Outrata, Treiman, Ye, Zhang, and their collaborators for various classes of MPECs and MPCCs via basic normals, coderivatives, and subgradients in finite-dimensional spaces; see [457, 689, 690, 691, 816, 1024, 1025, 1026, 1026, 1028, 1030, 1268, 1338, 1339, 1342, 1343, 1360, 1361], where the reader can find many efficient conditions expressed in terms of the initial problem data as well as numerous examples and applications.

Regarding necessary optimality conditions for general/abstract MPECs obtained in Subsect. 5.2.1, observe a crucial role of the *mixed coderivative* and the *partial* SNC property of the equilibrium map $S(\cdot)$ in the constraint qualification and normal compactness assumptions of Theorems 5.33 and 5.34. In such a way we strongly explore (in infinite dimensions) a *product structure* of the underlying decision-parameter space inherent in MPECs, which considerably distinguishes this remarkable class of constrained optimization problems from general mathematical programs with geometric constraints. Since these assumptions are *automatic* for *Lipschitz-like* mappings, in both finite and infinite dimensions, the results obtained single out a significant and rather general class of MPECs for which the first-order *qualified* necessary optimality conditions are always satisfied; see Corollary 5.35.

The subsequent necessary optimality conditions obtained for *structured* MPECs in Subsect. 5.2.2 can be viewed as consequences of the "abstract" MPEC results from Subsect. 5.2.1 and well-developed generalized differential and SNC *calculi*. Moreover, we broadly use the calculation and upper estimates for coderivatives of parametric variational systems derived in Sect. 4.4 for the purpose of sensitivity analysis. Now it is fully employed for necessary optimality conditions in MPECs, which reveals *close relationships* between these seemingly different issues. Observe also the usage of the *second-order subdifferentials* in the first-order optimality conditions for the most important classes of MPECs governed by *generalized variational inequalities* (GVIs) of types (5.60) and (5.63) and their specifications. It is not however surprising, since MPECs constraints themselves accumulate a first-order information about lower-level parametric problems; see the discussions above.

5.5.16. Exact Penalization and Calmness in Optimality Conditions for MPECs. The results of Subsect. 5.2.3 are based on another approach to deriving necessary optimality conditions for MPECs with equilibrium constraints governed by parametric variational systems of type (5.56): it involves, besides employing generalized differential and SNC calculi, a preliminary *exact penalization procedure*; cf. the corresponding development in Subsect. 5.1.2 for optimization problems with operator constraints of the equality type. In this way we obtain refinements of some *lower* (but *not upper*) subdifferential conditions for MPECs governed by parametric variational systems that were established in Subsect. 5.2.2.

Lemma 5.47 on the exact penalization for optimization problems under the generalized equation constraints (5.69) was established by Ye and Ye [1343]; see also Zhang [1360] for a preceding upper Lipschitzian version. Observe its similarity to the exact penalization result of Theorem 5.16 for optimization problems under equality constraints, which is due to Ioffe [588]. Furthermore, we can view the *calmness* condition from Definition 5.46 used in Lemma 5.47 as an (inverse) set-valued counterpart of the *weakened metric regularity* from Definition 5.15.

The calmness terminology for set-valued mappings in the framework of Definition 5.46 was suggested by Rockafellar and Wets [1165]. As mentioned in Subsect. 5.2.3 after this definition, the calmness property of F at $\bar{x} \in \operatorname{dom} F$, with $V = Y$ in (5.68), was introduced by Robinson [1130] as the "upper Lipschitzian" property of set-valued mappings. In [1132], Robinson established a major fact about the fulfillment of his upper Lipschitzian property for *piecewise polyhedral* multifunctions between finite-dimensional spaces. The graph-localized version of the calmness (upper Lipschitzian) property at $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ was later developed, under different names, by Klatte [684] and then independently by Ye and Ye [1343] in the context of Lemma 5.47 with subsequent MPEC applications.

On the other hand, the calmness property of *optimal value functions*, in the sense consonant with the usage of this word in the context of Definition 5.46, was developed by Clarke [249, 255] (while suggested by Rockafellar; see [249, p. 172]) as a kind of *constraint qualification* for necessary optimality conditions. The latter property is closely related to the notion of "Φ_1-subdifferential" introduced by Dolecki and Rolewicz [341] in the framework of exact penalization. We also refer the reader to Burke [188, 189], Facchinei and Pang [424], Henrion and Jourani [559], Henrion, Jourani and Outrata [560], Henrion and Outrata [561, 562], Klatte and Kummer [686], Outrata [1027, 1030], Ye [1338, 1339], Zhang [1361, 1362], Zhang and Treiman [1363], and the bibliographies therein for numerous applications of calmness and related properties to various aspects of optimization and variational analysis.

The necessary optimality conditions of Theorems 5.48, 5.49 and Corollary 5.50 are new in full generality; their finite-dimensional versions and concretizations were obtained by Outrata [1024, 1027], Ye [1338, 1339], Ye and Ye [1343], and Zhang [1360] with a variety of applications to some special classes of MPECs, particularly to bilevel programming. Corollary 5.51 and the subsequent example for polyhedral problems are taken from Outrata [1027].

5.5.17. Constrained Problems of Multiobjective Optimization and Equilibria. Section 5.3 is devoted to the study of constrained problems of *multiobjective optimization*, where the objectives are given by general preference relations that particularly cover a number of diverse *equilibria*. There is a vast literature dealing with various aspects of multiobjective/vector optimization and equilibrium models including the existence of optimal and equilibrium solutions, optimality conditions, numerical algorithms, and applications.

We refer the reader to [90, 93, 178, 230, 255, 265, 293, 306, 378, 395, 402, 424, 446, 480, 504, 516, 532, 534, 550, 627, 628, 636, 697, 707, 813, 820, 897, 901, 926, 928, 958, 995, 1000, 1001, 1002, 1029, 1031, 1040, 1046, 1119, 1134, 1181, 1195, 1214, 1312] and the bibliographies therein for a variety of approaches, results, and discussions. Note that most of the references above don't particularly deal with economic modeling and the corresponding concepts of competitive equilibria, which are considered in Chap. 8.

The material presented in Sect. 5.3 mainly concerns *general concepts* of optimal solutions to *multiobjective optimization* and *equilibrium* problems with their specification. The primary goal of the obtained results is the derivation of *necessary optimality conditions* for certain remarkable classes of multiobjective optimization problems with various constraints including a new class of the so-called *equilibrium problems with equilibrium constraints* (EPECs) important in many practical applications. It is demonstrated by the results presented in this section that the developed methods of variational analysis and generalized differentiation happen to be very useful to handle such problems and lead to powerful optimality conditions most of them are either new and have been just recently published. We don't consider here existence and numerical issues in multiobjective optimization and equilibria, which are far removed from the methods developed in this book.

Our main attention is paid to the study of *two different approaches* to multiobjective optimization and equilibria, which significantly distinct from each other even from the viewpoint of solution concepts. At the same time, necessary optimality conditions for constrained problems obtained via these approaches are based on generally different versions of the *extremal principle.*

5.5.18. Solution Concepts in Multiobjective Optimization. The notion of *generalized order optimality* from Definition 5.53 goes back to the early work by Kruger and Mordukhovich (see [707, 719, 897, 901]); it is directly induced by the concept of *local extremal points* for systems of sets. Observe that this abstract optimality notion *doesn't* impose any assumptions on *convexity* and/or *nonempty interiority* on the ordering set Θ; compare, e.g., Gamkrelidge [496], Gorokhovik [516], Neustadt [1001, 1002], Rubinov [1181], Singer [1214], Warga [1319], and other publications on abstract optimality. The particular notions of optimality discussed after Definition 5.53 are essentially classical; they largely go back to the seminal work by Pareto [1053]. Observe that it is much easier to investigate *weak Pareto* optimal solutions in comparison with (proper) *Pareto* ones; the major results in vector optimization have been actually obtained for *weak* Pareto solutions.

Definition 5.55 of *closed preferences* is due to Mordukhovich, Treiman and Zhu [958], while various abstract preference relations have been long considered and applied in vector optimization and especially in economic modeling; see, e.g., Debreu [310], Mas-Colell [854, 855], Pallaschke and Rolewicz [1040], Zhu [1372], and the references therein. The results of Proposition 5.56 characterizing the almost transitivity property of the generalized Pareto preference via the special properties of the ordering cone and then of Example 5.57 showing that it may *fail* for the lexicographical order on finite-dimensional spaces are taken from the dissertation by Eisenhart [395] conducted under supervision of Zhu.

5.5.19. Necessary Conditions for Generalized Order Optimality. Subsection 5.3.2 presents necessary optimality conditions for general constrained multiobjective optimization problems and their specifications, where the notion of *generalized order optimality* is understood in the sense of Definition 5.53. The results obtained are based on the version of the *exact extremal principle* given in Lemma 5.58. Its main difference from the version established in Subsect. 2.2.3 is that it takes into account the *product structure* of the underlying space *inherent* in multiobjective optimization. In this way more subtle conditions for the exact extremal principle (involving PSNC but not SNC properties) are derived in *infinite dimensions*; see Mordukhovich and B. Wang [963] for further results in this direction.

Theorem 5.59 and its Corollary 5.60 are new; some particular results under significantly more restrictive assumptions were given in Kruger [707] and Mordukhovich [897, 901]. The *upper* subdifferential conditions from Theorem 5.61 are new as well.

Minimax optimization problems have drawn a strong attention of mathematicians, applied scientists, and practitioners for many years due to their particular importance for the theory and applications. Such problems, which are intrinsically *nonsmooth*, were among the first classes of nonstandard optimization problems studied by (mostly specific) methods of nonsmooth analysis; see, e.g., Clarke [255], Danskin [307], Demyanov and Malosemov [319], Dubovitskii and Milyutin [370], Ioffe and Tikhomirov [618], Krasovskii and Subbotin [702], Neustadt [1002], Pshenichnyi [1106], Rockafellar and Wets [1165], and the references therein.

One (traditional) approach to deriving necessary optimality conditions for minimax problems is to employ those for general nonsmooth problems of *scalar* optimization and then to use formulas for computing/estimating the corresponding subdifferentials of *maximum* functions. Employing in this way the calculus rules of Subsect. 3.2.1 for basic subgradients of the maximum function over a finite set, we easily arrive at the necessary optimality conditions of Corollary 5.63. This result was first established in Mordukhovich [892] by a direct application of the method of metric approximations in finite-dimensional spaces.

The approach we employ to prove Theorem 5.62, based on the *reduction to generalized order optimality*, seems to be more appropriate and convenient to handle the minimax problem (5.83) involving maximization over a *weak** *compact* subset of a dual space. The results obtained in Theorem 5.62 are new in full generality, while some special cases for the compact set under maximization were previously considered by Kruger [706] and Mordukhovich [901].

5.5.20. Extended Versions of the Extremal Principle for Set-Valued Mappings. Subsection 5.3.3 contains *extended* versions of the *extremal principle* particularly needed for applications to necessary optimality conditions for problems of multiobjective optimization described via closed preference. Such extensions should be able to deal not with just (linear) translations of sets but with *nonlinear deformations* of set-valued mappings. An appropriate result in the form of the *approximate extremal principle* for set-valued mappings is given in Theorem 5.68 that is taken from the paper by Mordukhovich, Treiman and Zhu [958], where the reader can find the presented and additional examples illustrating Definition 5.64 of *extended extremal systems*.

To establish an *exact* version of the extremal principle for set-valued mappings, the notion of limiting normals to *moving* (i.e., parameterized) sets is required. An appropriate definition is given in [958], where we put $\varepsilon = 0$ in construction (5.95), which doesn't restrict the generality in the Asplund space setting under consideration; cf. also Bellaassali and Jourani [93] in finite dimensions. The notion of *normal semicontinuity* for moving sets from (5.96) was introduced earlier by Mordukhovich [894] motivated by applications to the covering property of set-valued mappings.

The *sufficient* conditions for the *normal semicontinuity* from Proposition 5.70 are taken from Mordukhovich [894, 901], while in Bellaassali and Jourani [93] the reader can find an interesting example of violating this property for a set-valued mapping $S(z) = \mathrm{cl}\,\mathcal{L}(z)$ generated by level sets of the preference determined by a Lipschitz continuous *utility function* on $I\!R^2$. Other sufficient conditions ensuring the normal semicontinuity of *uniformly prox-regular* mappings have been recently obtained by Bounkhel and Jofré [171] in finite dimensions and by Bounkhel and Thibault [173] in Hilbert spaces motivated by applications to nonconvex economies and to nonconvex sweeping processes, respectively.

As in the case of fixed sets, we need some amount of *normal compactness* to derive results of the exact/pointbased type in *infinite dimensions*. An appropriate extension of the SNC property for moving sets/set-valued mappings is formulated in Definition 5.71 under the name of "*imagely SNC*". This property, together with its *partial* versions as well as with the construction of the limiting normal cone from Definition 5.69 and the corresponding subdifferential and coderivative notions, were investigated in detail by Mordukhovich and B. Wang [966, 969]; see some discussions after Definition 5.71. It occurs that the extended limiting constructions for moving sets and mappings satisfy *calculus rules* similar to their basic counterparts, while the relationships between the basic and extended SNC properties are more complicated depending on a properly defined *uniformity*.

The *exact extremal principle* for set-valued mappings formulated in Theorem 5.72 was proved in Mordukhovich, Treiman and Zhu [958]. The *converse* implication in Theorem 5.72 follows directly from the corresponding result for extremal systems of closed sets established in Theorem 2.22(ii).

5.5.21. Necessary Conditions for Multiobjective Problems with Closed Preference Relations. Subsection 5.3.4 contains necessary optimality conditions for multiobjective optimization problems under various constraints (of geometric, operator, and functional types), where "multiobjective minimization" is defined by *closed preference relations.* The results obtained in this subsection are based on applying the extended versions of the extremal principle from Subsect. 5.3.3 and are given in both *approximate/fuzzy* and *exact/limiting* forms.

The *fuzzy* optimality condition from Theorem 5.73(i) for problems with geometric constraints are taken from Mordukhovich, Treiman and Zhu [958], where the reader can also find necessary conditions in "strong" and "weak" fuzzy forms for multiobjective problems with finitely many functional constraints of equality and inequality types. The *limiting* optimality conditions obtained in Theorem 5.73(ii), Corollary 5.75, and Theorem 5.76 are new; partial results for the mentioned problem with equality and inequality constraints were obtained in [958] under the finite dimensionality assumption on the range space Z for the objective mapping $f\colon X \to Z$. We refer the reader to Remark 5.74 for the discussion on comparison between the corresponding optimality conditions obtained for multiobjective problems with "generalized order" and "closed preference" concepts of vector optimality.

The material of Subsect. 5.3.4 on *multiobjective games* is taken from Mordukhovich, Treiman and Zhu [958].

5.5.22. Equilibrium Problems with Equilibrium Constraints. Subsection 5.3.5 is devoted to *multiobjective optimization problems with equilibrium constraints.* We treat this class of vector optimization problems as a multiobjective counterpart/extension of MPECs considered in Sect. 5.2. Indeed, they involve the same type of (equilibrium) constraints as MPECs, while the optimization is conducted with respect to the general multiobjective criteria discussed in Subsect. 5.3.1. As shown therein, the concepts of multiobjective optimization under consideration include various notions of *equilibria*, and thus such problems can be viewed as *equilibrium problems with equilibrium constraints* (EPECs).

The EPEC terminology has appeared quite recently; it was coined by Scholtes in his talk [1193] at the 2002 International Conference on Complementarity Problems. Practical applications were among the primary motivations to study this class of multiobjective optimization problems; see the concurrent work by Hu, Ralph, Ralph, Bardsley and Ferris [585] on the competition equilibrium model in deregulated electricity markets. The main attention in [585, 1193] was paid to EPECs, where the behavior of both Leaders (upper level) and Followers (lower level) were modeled via the *noncooperative* Nash (or Cournot-Nash) equilibrium; cf. [995, 1031]. We also refer the reader to Fukushima and Pang [480] and Outrata [1029] for related developments. The latter paper contains, in particular, a deep insight into the nature of various EPECs and presents necessary optimality conditions for a class of noncooperative (regarding both levels) EPECs by reducing them to coupled MPECs and employing our basic generalized differential constructions.

EPECs of the *other kind* were examined by Mordukhovich [926, 928] from the viewpoint of *multiobjective optimization* at the *upper* hierarchical level and *equilibrium constraints* governed by parametric variational systems at the *lower* level of hierarchy. Such EPECs can be naturally treated as those where all the Leaders *cooperate* with each other seeking a generalized *Pareto-type* equilibrium; they cannot be just reduced to systems of MPECs and require special considerations. Necessary optimality conditions for EPECs obtained in the afore-mentioned papers [926, 928] were derived, in finite dimensions, from general results of multiobjective optimization (see the preceding material of this section) by using generalized differential calculus for our basic constructions. More special results of this type were obtained by Ye and Zhu [1345] for finite-dimensional multiobjective problems with variational inequality constraints, where the upper-level optimality was defined by some "regular" preference relations.

The recent work by Mordukhovich, Outrata and Červinka [940] contained the development and specification of the approach from [924, 928] to an important class of finite-dimensional EPECs governed by *complementarity constraints* at the lower level with the classical *weak Pareto* optimality at the upper level. Taking into account specific features of the complementarity constraints in finite dimensions, the necessary optimality conditions in [940] were expressed constructively via the initial problem data and were used for building an efficient *numerical algorithm* based on the *implicit programming* approach developed in the book by Outrata, Kočvara and Zowe [1031] in the context of MPECs. Furthermore, in [940] the reader can find applications of the results obtained to the modeling of hierarchical *oligopolistic markets* involving many Leaders and Followers.

The results presented in Subsect. 5.3.5 are mostly new developing the previous optimality conditions obtained by Mordukhovich [926, 928] in finite dimensions. Note that the infinite-dimensional setting happens to be significantly more involved, since it requires employing, besides calculus rules of generalized differentiation, appropriate results of *SNC calculus* to express necessary optimality and qualification conditions via the EPEC initial data. Observe a crucial role of Theorem 5.59 on generalized order optimality as well as of the chain rules in second-order subdifferentiation to derive the necessary optimality conditions for EPECs in Subsect. 5.3.5.

5.5.23. Subextremality and Suboptimality at Linear Rate. The issues brought up in Sect. 5.4 are non-conventional in optimization theory, where *necessary* conditions are usually (except for convex problems and the like) *not sufficient* for standard notions of optimality. Observe that all the major necessary optimality conditions in all the branches of the classical and modern optimization theory (Lagrange multipliers and Karush-Kuhn-Tucker conditions in nonlinear programming, Euler-Lagrange equation in the calculus of variations, Pontryagin maximum principle in optimal control, etc.) are expressed in *dual* forms involving *adjoint* variables. This is the case of all the generalized extremality and optimality conditions developed in this book. At the same time, the very *notions of optimality*, both scalar and vector, are formulated of course in *primal* terms.

A challenging question is to find certain modified notions of extremality/optimality so that known necessary conditions for the previously recognized notions become *necessary and sufficient* in the new framework. Such a study was initiated by Kruger [710, 711], and then it has been continued in his subsequent publications [713, 714, 715]. The new notions of set extremality and the associated optimality for vector and scalar optimization problems were first called "extended extremality/optimality" [710, 711, 712, 713], while recently [714, 715, 716] Kruger has started to use the name of "weak stationarity" for them. We suggest to use the term *linear subextremality/suboptimality* for these notions by the reasons explained below; cf. also the introductory part of Sect. 5.4.

Indeed, the new notions, being weaker than the conventional ones, actually concern an extremal/optimal behavior of sets and mappings at points *nearby* those in question; thus it makes sense to use the prefix "sub" to identify such a behavior.

The other crucial feature of the new notions is that, in contrast to the conventional ones, they involve a *linear rate* of extremality and optimality, similarly to the linear openness/covering, metric regularity, and Lipschitz-like properties comprehensively studied in this book. As seen, the *linear rate nature* of these fundamental properties, which has been fully recognized just in the framework of modern variational analysis (even regarding the classical settings), is the *key issue* allowing us to establish their *complete characterizations* in terms of generalized differentiation.

Precisely the same *linear rate essence* of the (sub)extremality and (sub)optimality concepts studied in Sect. 5.4 is the *driving force* ensuring the possibility to justify the validity of the known extremality and optimality conditions for the conventional notions as *necessary and sufficient* conditions for the new notions under consideration. Moreover, there are direct connections between covering/metric regularity/Lipschitz-like properties and the linear subextremality/subotimality notions that reveal via both proofs (see, e.g., the proof of Theorem 5.88) as well as the corresponding constant relationships from the recent papers by Kruger [715, 716].

5.5.24. Linear Set Subextremality and Linear Suboptimality for Multiobjective Problems. Definition 5.87 of *linear set subextremality* is due to Kruger [711] called originally "extended extremality" and then [715] "weak extremality" of set systems. *Necessary and sufficient* conditions for linear subextremality in the form equivalent to (5.106) was first announced by Kruger [711] in Fréchet smooth spaces and then proved in [712, 713] in the Asplund space setting. Note that the proof of assertion (ii) is similar to those of

Lemma 2.32(ii) and Theorem 2.51(i) taken, respectively, from Mordukhovich and Shao [948] and Mordukhovich [920]. Theorem 5.89 on *characterizing* this notion via the *exact extremal principle* is new.

The notion of *linear suboptimality* for *multiobjective* problems from Definition 5.91 was introduced by Kruger in [710] under the name of "extended (f, Ω, Θ)-extremality." A *fuzzy* characterization of this notion in the form equivalent to (5.112) of Theorem 5.92 was first announced in [710] for Fréchet smooth spaces and then was proved in [712] in Asplund spaces. All the other results of this subsection regarding *exact/pointbased* characterizations of linear suboptimality for multiobjective problems are new.

To derive these pointbased characterizations, we involve our basic normals, coderivatives (both mixed and normal), as well as first-order and second-order subgradients at the points in question. This allows us to employ the well-developed generalized differential calculus for these constructions, together with the associated SNC calculus in infinite dimensions. It is important to emphasize that to obtain in this way *necessary and sufficient* conditions for linear suboptimality of structured multiobjective problems (including those for EPECs), we have to use calculus results of not just the "right" inclusion type as in the vast majority of applications of generalized differentiation, but of the *equality* type—which are more restrictive but still well developed in the book—at all the calculus levels. Likewise, we need to employ SNC calculus results ensuring the *equivalence* between the corresponding SNC properties under various operations in infinite dimensions.

5.5.25. Linear Subminimality in Constrained Optimization. The last subsection of Chap. 5 concerns the notion of *linear subminimality* for optimization problems involving *scalar* (extended-real-valued) functions. This notion was introduced by Kruger [712] under the name of "almost minimality," and then it was studied in [713] as "extended minimality" and in [714] as "weak inf-stationarity." Although one can always treat the linear subminimality from Definition 5.101 as a particular case of the linear suboptimality concept formulated in Definition 5.91 for mappings to generalized ordering spaces, there are certain specific features of the scalar case that should be taken into account in the study and applications. As illustrated by the simple functions from Example 5.102, which is due to Kruger [712], the behavior of linearly subminimal points may be dramatically different from that of points of local minimum. On the other hand, it has been observed in [712] that linearly subminimal points are *stable* with respect to perturbations by smooth functions with *vanishing derivatives*, in contrast to local minimizers. This implies that for *smooth* functions the notion of linear subminimality is *equivalent* to the classical *stationarity*, which is *not* however the case in *nonsmooth* settings.

Another Kruger's observation made later in [713] is that, in the general case of l.s.c. functions on Banach spaces, the linear subminimality from Definition 5.101 is *equivalent* to the notion introduced by Kummer [728] under the name of "stationarity points with respect to minimization," which is formulated in part (b) of Theorem 5.103. The latter theorem also establishes an efficient description of linear suboptimality via the powerful construction of *strong slope* introduced by De Giorgi, Marina and Tosques [312] in the theory of evolution equations and then efficiently employed by Azé, Corvellec and Lucchetti [70] and by Ioffe [608] to variational stability and metric regularity; see discussions in Subsect. 4.5.2.

The *fuzzy* subdifferential criterion for linear subminimality from assertion (i) of Theorem 5.106 is due to Kruger [712] following directly from Theorem 5.92. Assertion (ii) of Theorem 5.106 is new. It provides a "condensed" *pointbased characterization* of linear subminimality via basic subgradients of the restricted function $\varphi_\Omega = \varphi + \delta(\cdot;\Omega)$ and then allows us to get the convenient "separated" criterion (5.114) under each of the assumptions (a)–(c) of Corollary 5.107, which ensure the subdifferential sum rule as *equality*; see also the discussion after this corollary. The latter result implies more specific criteria for linear subminimality of *structured* constrained minimization problems (in particular, for MPECs) similarly to the derivation of Subsect. 5.4.2 based on the *equality* type first-order and second-order calculus rules of our basic generalized differentiation.

Finally, Theorem 5.108 gives new *necessary* conditions for linear subminimality in problems with *inequality constraints.* In contrast to all the previous results on linear suboptimality and subminimality, it establishes conditions of the *upper subdifferential* type, which again can be developed for other structured problems of constrained optimization similarly to the necessary conditions for conventional optimality studied in detail in Sects. 5.1 and 5.2.

6

Optimal Control of Evolution Systems in Banach Spaces

The next two chapters are on *optimal control*, which is among the most important motivations and fruitful applications of modern methods of variational analysis and generalized differentiation. It is not accidental that the very concepts of basic normals, subgradients, and coderivatives used in this book were introduced and applied by the author in connection with problems of optimal control. In fact, already the simplest and historically first problems of optimal control are *intrinsically nonsmooth*, even in the case of smooth functional data describing dynamics and constraints on feasible arcs. The crux of the matter is that a characteristic feature of optimal control problems, in contrast to the classical calculus of variations, is the presence of *pointwise* constraints on control functions, which may be (and often are) defined by *highly irregular* sets consisting, e.g., of finitely many points. In particular, this is the case of typical problems in automatic control that provided the primary motivation for developing optimal control theory.

The principal goal of the following developments is to derive necessary optimality conditions in a range of optimal control problems for evolution systems by using methods of variational analysis and generalized differentiation. This chapter concerns dynamical systems governed by *ordinary* differential equations and inclusions in Banach spaces; control problems for systems with *distributed parameters* governed by functional-differential and partial differential relations will be mostly considered in Chap. 7.

The main attention is paid in this chapter to optimal control/dynamic optimization problems of the Bolza and Mayer types governed by infinite-dimensional evolution inclusions and control systems with both *discrete-time* and *continuous-time* dynamics in the presence of endpoint constraints. Along with the variational principles in infinite dimensions and tools of generalized differentiation developed above, we employ special techniques of dynamic optimization and optimal control. The basic approach developed below is the *method of discrete approximations*, which allows us to approximate continuous-time control problems by those involving discrete dynamics. The relationship between continuous-time and discrete-time control systems is one of the *central topics* of this chapter. The results obtained in this direction shed new light upon both *qualitative* and *numerical* aspects of optimal control from the viewpoint of the theory and applications.

6.1 Optimal Control of Discrete-Time and Continuous-time Evolution Inclusions

This section concerns optimal control problems for dynamic/evolution systems governed by *differential inclusions* and their *finite-difference approximations* in appropriate (quite general) *Banach spaces*. The models under consideration capture more conventional problems of optimal control described by parameterized differential equations. Our primary method to study continuous-time control systems is to construct *well-posed discrete approximations* and to establish their *variational stability* with respect to the *value convergence* as well as a suitable *strong convergence* of their optimal solutions. Then we derive necessary optimality conditions for discrete-time optimal control problems governed by *finite-difference inclusions*. The latter problems can be reduced to non-dynamic optimization problems considered in the previous chapter in the presence of many geometric constraints. On the other hand, they have specific structural features exploited in what follows. In this way, applying generalized differential and SNC calculi from Chap. 3, we obtain necessary optimality conditions for discrete approximations in both fuzzy and exact forms under fairly general assumptions on the initial data. Passing to the limit with the use of coderivative characterizations of Lipschitzian stability from Chap. 4 allows us to derive necessary optimality conditions for *intermediate local minimizers* (that provide a local minimum lying between the classical weak and strong ones) in the *extended Euler-Lagrange* form for continuous-time systems under certain relaxation/convexification with respect to velocity variables. To avoid such a relaxation under appropriate assumptions, we develop an additional approximation procedure in the next section.

6.1.1 Differential Inclusions and Their Discrete Approximations

Let X be a Banach space (called the *state space* in what follows), and let $T := [a, b]$ be a *time interval* of the real line. Consider a set-valued mapping $F\colon X \times T \rightrightarrows X$ and define the *differential/evolution inclusion*

$$\dot{x}(t) \in F(x(t), t) \quad \text{a.e. } t \in [a, b] \tag{6.1}$$

generated by F, where $\dot{x}(t)$ stands for the time derivative of $x(t)$, and where a.e. (almost everywhere) means as usual that the relation holds up to the Lebesgue measure zero on $I\!R$. Let us give the precise definition of solutions to the differential inclusion (6.1), which is used in this chapter.

Definition 6.1 (solutions to differential inclusions). *By a* SOLUTION *to inclusion* (6.1) *we understand a mapping* $x: T \to X$, *which is Fréchet differentiable for a.e.* $t \in T$ *and satisfies* (6.1) *and the* NEWTON-LEIBNIZ FORMULA

$$x(t) = x(a) + \int_a^t \dot{x}(s)\, ds \quad \text{for all} \quad t \in T \ ,$$

where the integral in taken in the BOCHNER SENSE.

It is well known that for $X = I\!R^n$, $x(t)$ is a.e. differentiable on T and satisfies the Newton-Leibniz formula if and only if it is *absolutely continuous* on T in the standard sense, i.e., for any $\varepsilon > 0$ there is δ such that

$$\sum_{j=1}^{l} \|x(t_{j+1}) - x(t_j)\| \le \varepsilon \quad \text{whenever} \quad \sum_{j=1}^{l} |t_{j+1} - t_j| \le \delta$$

for the disjoint intervals $(t_j, t_{j+1}] \subset T$. However, for infinite-dimensional spaces X even the Lipschitz continuity may not imply the a.e. differentiability. On the other hand, there is a *complete characterization* of Banach spaces X, where the absolute continuity of every $x: T \to X$ is *equivalent* to its a.e. differentiability and the fulfillment of the Newton-Leibniz formula. This is the class of spaces with the so-called *Radon-Nikodým property* (RNP).

Definition 6.2 (Radon-Nikodým property). *A Banach space* X *has the* RADON-NIKODÝM PROPERTY *if for every finite measure space* $(\varXi, \varSigma, \mu)$ *and for each* μ*-continuous vector measure* $m: \varSigma \to X$ *of bounded variation there is* $g \in L^1(\mu; \varXi)$ *such that*

$$m(E) = \int_E g\, d\mu \quad \text{for} \quad E \in \varSigma \ .$$

This fundamental property is well investigated in the general vector measure theorem and the geometric theory of Banach spaces; we refer the reader to the classical texts by Diestel and Uhl [334] and Bourgin [169] for the comprehensive study of the RNP and its applications. In particular, in [334, pp. 217–219] one can find the summary of equivalent formulations/charactetizations of the RNP and the list of specific Banach spaces for which the RNP automatically holds. It is important to observe that the latter list contains every *reflexive* space and every *weakly compactly generated dual* space, hence all *separable duals*. On the other hand, the classical spaces c_0, c, l^∞, $L^1[0,1]$, and $L^\infty[0,1]$ *don't* have the RNP. Let us mention a nice relationship between the RNP and Asplund spaces used in what follows: *given a Banach space* X, *the dual space* X^* *has the RNP if and only if* X *is Asplund*.

Thus for Banach spaces with the RNP (and only for such spaces) the solution concept of Definition 6.1 agrees with the standard definition of

Carathéodory solutions dealing with absolutely continuous mappings. In general, Definition 6.1 postulates what we actually need for our purposes without appealing to Carathéodory solutions and the RNP. However, the RNP along with the Asplund property of X are essentially used for deriving major results in this chapter (but not all of them) from somewhat different prospectives not directly related to the adopted concept of solutions to differential inclusions.

It has been well recognized that differential inclusions, which are certainly of their own interest, provide a useful generalization of *control systems* governed by differential/evolution *equations* with control parameters:

$$\dot{x} = f(x,u,t), \quad u \in U(t) \;, \tag{6.2}$$

where the control sets $U(\cdot)$ may also depend on the state variable x via $F(x,t) = f(x,U(x,t),t)$. In some cases, especially when the sets $F(x,t)$ are convex, the differential inclusions (6.1) admit parametric representations of type (6.2), but in general they cannot be reduced to parametric control systems and should be studied for their own sake. Note also that the *ODE form* (6.2) in Banach spaces is strongly related to various control problems for evolution *partial differential equations* of parabolic and hyperbolic types, where solutions may be understood in some other appropriate senses; see, e.g., the books by Fattorini [432] and by Li and Yong [789] as well as the results and discussions presented in Remark 6.26 and Chap. 7 below.

Our principal method to study differential inclusions involves *finite-difference* replacements of the derivative

$$\dot{x}(t) \approx \frac{x(t+h)-x(t)}{h}, \quad h \to 0 \;,$$

where the *uniform Euler scheme* is considered for simplicity. To formalize this process, we take any natural number $N \in I\!N$ and consider the *discrete grid/mesh* on T defined by

$$T_N := \big\{a, a+h_N, \ldots, b-h_N, b\big\}, \quad h_N := (b-a)/N \;,$$

with the *stepsize of discretization* h_N and the *mesh points* $t_j := a + jh_N$ as $j = 0,\ldots,N$, where $t_0 = a$ and $t_N = b$. Then the differential inclusion (6.1) is replaced by a sequence of its *finite-difference/discrete approximations*

$$x_N(t_{j+1}) \in x_N(t_j) + h_N F(x_N(t_j),t_j), \quad j = 0,\ldots,N-1 \;. \tag{6.3}$$

Given a discrete trajectory $x_N(t_j)$ satisfying (6.3), we consider its *piecewise linear extension* $x_N(t)$ to the continuous-time interval T, i.e., the *Euler broken lines.* We also define the *piecewise constant extension* to T of the corresponding *discrete velocity* by

$$v_N(t) := \frac{x_N(t_{j+1})-x_N(t_j)}{h_N}, \quad t \in [t_j,t_{j+1}), \quad j = 0,\ldots,N-1 \;.$$

It follows from the very definition of the Bochner integral that

$$x_N(t) = x_N(a) + \int_a^t v_N(s)\,ds \quad \text{for} \quad t \in T \ .$$

Our first goal is to show that *every* solution to the differential inclusion (6.1) can be *strongly approximated*, under reasonable assumptions, by extended trajectories to the discrete inclusions (6.3). By strong approximation we understand the convergence in the norm topology of the classical Sobolev space $W^{1,2}\big([a,b];X\big)$ with the norm

$$\|x(\cdot)\|_{W^{1,2}} := \max_{t\in[a,b]} \|x(t)\| + \Big(\int_a^b \|\dot{x}(t)\|^2\,dt\Big)^{1/2} ,$$

where the norm on the right-hand side is taken in the space X. Note that the convergence in $W^{1,2}\big([a,b];X\big)$ implies the (uniform) convergence of the trajectories on $[a,b]$ and the *pointwise* (a.e. $t \in [a,b]$) convergence of (some subsequence of) their derivatives. The latter is crucial for our purposes, especially in the case of *nonconvex* values $F(x,t)$.

Let us formulate the *basic assumptions* for our study that apply not only to the next theorem but also to the subsequent results on differential inclusions via discrete approximations. Nevertheless, these assumptions can be relaxed in some settings; see the remarks and discussions below. Roughly speaking, we assume that the set-valued mapping $F\colon X \times [a,b] \rightrightarrows X$ is compact-valued, locally Lipschitzian in x, and Hausdorff continuous in t a.e. on $[a,b]$. More precisely, the following hypotheses are imposed along a given trajectory $\bar{x}(\cdot)$ to (6.1), which is arbitrary in the next theorem but then will be a reference optimal solution to the variational problem under consideration.

(H1) There are an open set $U \subset X$ and positive numbers m_F and ℓ_F such that $\bar{x}(t) \in U$ for all $t \in [a,b]$, the sets $F(x,t)$ are nonempty and compact for all $(x,t) \in U \times [a,b]$, and one has the inclusions

$$F(x,t) \subset m_F I\!\!B \quad \text{for all} \quad (x,t) \in U \times [a,b] \ , \tag{6.4}$$

$$F(x_1,t) \subset F(x_2,t) + \ell_F \|x_1 - x_2\| I\!\!B \quad \text{for all} \quad x_1, x_2 \in U, \quad t \in [a,b] \ . \tag{6.5}$$

(H2) $F(x,\cdot)$ is Hausdorff continuous for a.e. $t \in [a,b]$ uniformly in $x \in U$.

Note that inclusion (6.5) is equivalent to the uniform Lipschitz continuity

$$\operatorname{haus}\big(F(x,t), F(u,t)\big) \le \ell_F \|x - u\|, \quad x, u \in U \ ,$$

of $F(\cdot,t)$ with respect to the *Pompieu-Hausdorff metric* $\operatorname{haus}(\cdot,\cdot)$ on the space of nonempty and compact subsets of X; see Subsect. 1.2.2.

To handle efficiently the Hausdorff continuity of $F(x,\cdot)$ for a.e. $t \in [a,b]$, define the *averaged modulus of continuity* for F in $t \in [a,b]$ while $x \in U$ by

$$\tau(F;h) := \int_a^b \sigma(F;t,h)\,dt, \tag{6.6}$$

where $\sigma(F;t,h) := \sup\big\{\omega(F;x,t,h)\big|\ x \in U\big\}$ with

$$\omega(F;x,t,h) := \sup\Big\{\mathrm{haus}\big(F(x,t_1),F(x,t_2)\big)\Big|\ t_1,t_2 \in \big[t-\tfrac{h}{2},t+\tfrac{h}{2}\big]\cap[a,b]\Big\}\ .$$

The following observation is easily implied by the definitions.

Proposition 6.3 (averaged modulus of continuity). *Property* (H2) *holds if and only if* $\tau(F;h) \to 0$ *as* $h \to 0$.

Note that for single-valued mapping $f\colon[a,b] \to X$ the property $\tau(f;h) \to 0$ as $h \to 0$ is *equivalent to the Riemann integrability* of f on $[a,b]$; see Sendov and Popov [1201]. The latter holds, as well known, if and only if f is continuous at almost all $t \in [a,b]$.

The following *strong approximation* theorem plays a crucial role in further results based on discrete approximations.

Theorem 6.4 (strong approximation by discrete trajectories). *Let* $\bar{x}(\cdot)$ *be a solution to the differential inclusion* (6.1) *under assumptions* (H1) *and* (H2), *where* X *is an arbitrary Banach space. Then there is a sequence of solutions* $\widehat{x}_N(t_j)$ *to the discrete inclusions* (6.3) *such that*

$$\widehat{x}_N(a) = \bar{x}(a)\ \ \text{for all}\ \ N \in I\!N$$

and the extensions $\widehat{x}_N(t)$, $a \le t \le b$, *converge to* $\bar{x}(t)$ *strongly in the space* $W^{1,2}\big([a,b];X\big)$ *as* $N \to \infty$.

Proof. By Definition 6.1 involving the Bochner integral, the derivative mapping $\dot{\bar{x}}(\cdot)$ is *strongly measurable* on $[a,b]$, and hence we can find (rearranging the mesh points t_j if necessary) a sequence of *simple/step mappings* $w_N(\cdot)$ on T such that $w_N(t)$ are constant on $[t_j,t_{j+1})$ for every $j = 0,\ldots,N-1$ and $w_N(\cdot)$ converge to $\dot{\bar{x}}(\cdot)$ in the norm topology of $L^1\big([a,b];X\big)$ as $N \to \infty$. Combining this convergence with (6.1) and (6.4), we get

$$\int_a^b \|w_N(t)\|\,dt = \sum_{j=0}^{N-1}\|w_N(t_j)\|\,(t_{j+1}-t_j) \le (m_F+1)(b-a) \tag{6.7}$$

for all large N. In the estimates below we use the numerical sequence

$$\xi_N := \int_a^b \|\dot{\bar{x}}(t) - w_N(t)\|\,dt \to 0\ \ \text{as}\ \ N \to \infty\ .$$

Let us define the discrete functions $u_N(t_j)$ by

$$u_N(t_{j+1}) = u_N(t_j) + h_N w_N(t_j), \quad j = 0, \ldots, N-1, \quad u_N(t_0) := \bar{x}(a)$$

and observe that the functions

$$u_N(t) := \bar{x}(a) + \int_a^t w_N(s)\,ds, \quad a \le t \le b\ ,$$

are piecewise linear extensions of $u_N(t_j)$ to the interval $[a,b]$ and that

$$\|u_N(t) - \bar{x}(t)\| \le \int_a^t \|w_N(s) - \dot{\bar{x}}(s)\|\,ds \le \xi_N \ \text{ for } \ t \in [a,b]\ . \tag{6.8}$$

Therefore $u_N(t) \in U$ for all $t \in [a,b]$ whenever N is sufficiently large.

Taking the *distance function* $\operatorname{dist}(\cdot;\Omega)$ to a set in X, one can check that the Lipschitz condition (6.5) is equivalent to

$$\operatorname{dist}\big(w; F(x_1,t)\big) \le \operatorname{dist}\big(w; F(x_2,t)\big) + \ell_F\|x_1 - x_2\|$$

whenever $w \in X$, $x_1, x_2 \in U$, and $t \in [a,b]$; cf. the proof of Theorem 1.41. By the construction of $\tau(F;h)$ in (6.6) and the obvious relation

$$\operatorname{dist}\big(w; F(x,t_1)\big) \le \operatorname{dist}\big(w; F(x,t_2)\big) + \operatorname{haus}\big(F(x,t_1), F(x,t_2)\big)$$

one has the estimate

$$\begin{aligned}\zeta_N :&= \sum_{j=0}^{N-1} h_N \operatorname{dist}\big(w_N(t_j); F(u_N(t_j),t_j)\big)\\ &= \textstyle\sum_{j=0}^{N-1}\int_{t_j}^{t_{j+1}} \operatorname{dist}\big(w_N(t_j); F(u_N(t_j),t)\big)\,dt\\ &\le \sum_{j=0}^{N-1}\int_{t_j}^{t_{j+1}} \operatorname{dist}\big(w_N(t_j); F(u_N(t),t)\big)\,dt + \tau(F;h_N)\ .\end{aligned}$$

The Lipschitz property of F and the construction of $w_N(\cdot)$ imply

$$\operatorname{dist}\big(w_N(t_j); F(u_N(t_j),t)\big) \le \operatorname{dist}\big(w_N(t); F(u_N(t_j),t)\big) + \ell_F w_N(t_j)(t - t_j)$$

whenever $t \in [t_j, t_{j+1})$, and then

$$\begin{aligned}\operatorname{dist}\big(w_N(t); F(u_N(t),t)\big) &\le \operatorname{dist}\big(w_N(t); F(\bar{x}(t),t)\big) + \ell_F\|u_N(t) - \bar{x}(t)\|\\ &\le \|w_N(t) - \dot{\bar{x}}(t)\| + \ell_F\xi_N \ \text{ a.e. } \ t \in [a,b]\ .\end{aligned}$$

Employing further (6.7) and (6.8), we arrive at the estimate

$$\zeta_N \le \gamma_N := \big(1 + \ell_F(b-a)\big)\xi_N + \ell_F(b-a)(m_F+1)/2 + \tau(F;h_N)\ . \tag{6.9}$$

Observe that the functions $u_N(t_j)$ built above are *not* trajectories for the discrete inclusions (6.3), since one doesn't have $w_N(t_j) \in F(u_N(t_j),t_j)$. Now we use $w_N(t_j)$ to construct *actual trajectories* $\widehat{x}_N(t_j)$ for (6.3) that are close to $u_N(t_j)$ and enjoy the convergence property stated in the theorem.

Let us define $\widehat{x}_N(t_j)$ recurrently by the following *proximal algorithm*, which is realized due to the compactness assumption on the values of F:

$$\begin{cases} \widehat{x}_N(t_0)=\bar{x}(a), \quad \widehat{x}_N(t_{j+1})=\widehat{x}_N(t_j)+h_N v_N(t_j), \quad j=0,\ldots,N-1\,, \\ \text{where } v_N(t_j)\in F(\widehat{x}_N(t_j),t_j) \text{ with} \\ \|v_N(t_j)-w_N(t_j)\|=\operatorname{dist}\big(w_N(t_j);F(\widehat{x}_N(t_j),t_j)\big)\,. \end{cases} \tag{6.10}$$

First we prove that algorithm (6.10) keeps $\widehat{x}_N(t_j)$ inside the neighborhood U from (H1) whenever N is sufficiently large. Indeed, let us consider any number $N\in I\!N$ satisfying $\bar{x}(t)+\eta_N I\!B\subset U$ for all $t\in[a,b]$, where

$$\eta_N:=\gamma_N\exp\big(\ell_F(b-a)\big)+\xi_N$$

with ξ_N and γ_N defined above. We have $\eta_N\to 0$ as $N\to\infty$, since $\xi_N\to 0$ by the construction of ξ_N and since $\gamma_N\to 0$ due to assumption (H2) is equivalent to $\tau(F;h_N)\to 0$ by Proposition 6.3. Arguing by induction, we suppose that $\widehat{x}_N(t_i)\in U$ for all $i=0,\ldots,j$ and show that this also holds for $i=j+1$. Using (6.5), (6.9), and (6.10), one gets

$$\begin{aligned} \|\widehat{x}_N(t_{j+1})-u_N(t_{j+1})\| &\le \|\widehat{x}_N(t_j)-u_N(t_j)\|+h_N\|v_N(t_j)-w_N(t_j)\| \\ &\le \|\widehat{x}_N(t_j)-u_N(t_j)\|+h_N\operatorname{dist}\big(w_N(t_j);F(u_N(t_j),t_j)\big) \\ &+\ell_F\|\widehat{x}_N(t_j)-u_N(t_j)\|\le\ldots \\ &\le h_N\sum_{i=0}^{j}(1+\ell_F h_N)^{j-i}\operatorname{dist}\big(w_N(t_i);F(u_N(t_i),t_i)\big) \\ &\le h_N\exp\big[\ell_F(b-a)\big]\sum_{i=0}^{j}\operatorname{dist}\big(w_N(t_i);F(u_N(t_i),t_i)\big) \\ &\le \gamma_N\exp\big(\ell_F(b-a)\big)\,. \end{aligned}$$

Due to (6.8) the latter implies that

$$\|\widehat{x}_N(t_{j+1})-\bar{x}_N(t_{j+1})\|\le\gamma_N\exp\big(\ell_F(b-a)\big)+\xi_N=:\eta_N, \tag{6.11}$$

which proves that $\widehat{x}_N(t_j)\in U$ for all $j=0,\ldots,N$. Taking this into account, we have by the previous arguments that

$$\sum_{j=0}^{N}\|\widehat{x}_N(t_j)-u_N(t_j)\|\le(b-a)\exp\big(\ell_F(b-a)\big)\sum_{j=0}^{N-1}\operatorname{dist}\big(w_N(t_j);F(u_N(t_j),t_j)\big)\,.$$

Now let us estimate the quantity

$$\vartheta_N := \int_a^b \|\dot{\widehat{x}}_N(t) - w_N(t)\|\, dt \quad \text{as} \quad N \to \infty \; .$$

Using the last estimate above together with (6.9) and (6.11), we have

$$\vartheta_N = \sum_{j=0}^{N-1} h_N \|\dot{\widehat{x}}_N(t_j) - w_N(t_j)\| = \sum_{j=0}^{N-1} h_N \mathrm{dist}\big(w_N(t_j); F(\widehat{x}_N(t_j), t_j)\big)$$

$$\le \sum_{j=0}^{N-1} h_N \mathrm{dist}\big(w_N(t_j); F(u_N(t_j), t_j)\big) + \ell_F \sum_{j=0}^{N-1} h_N \|\widehat{x}_N(t_j) - u_N(t_j)\|$$

$$\le \gamma_N \big(1 + \ell_F(b-a) \exp\big(\ell_F(b-a)\big)\big) \; .$$

Thus one finally gets

$$\begin{aligned}\int_a^b \|\dot{\widehat{x}}_N(t) - \dot{\bar{x}}(t)\|\, dt &\le \int_a^b \|\dot{\widehat{x}}_N(t) - \dot{\bar{x}}(t)\|\, dt + \int_a^b \|w_N(t) - \dot{\bar{x}}(t)\|\, dt \\ &\le \gamma_N \big(1 + \ell_F(b-a) \exp\big(\ell_F(b-a)\big)\big) + \xi_N := \alpha_N \; .\end{aligned} \tag{6.12}$$

Since $\alpha_N \to 0$ as $N \to \infty$, this obviously implies the desired convergence $\widehat{x}_N(\cdot) \to \bar{x}(\cdot)$ in the norm of $W^{1,2}\big([a,b]; X\big)$ due to the Newton-Leibniz formula for $\widehat{x}_N(t)$ and $\bar{x}(t)$ and due to the boundedness assumption (6.4). △

Remark 6.5 (numerical efficiency of discrete approximations). It follows from (6.12) by the Newton-Leibniz formula that

$$\|\widehat{x}_N(t) - \bar{x}(t)\| \le \int_a^b \|\dot{\widehat{x}}_N(s) - \dot{\bar{x}}(s)\|\, ds \le \alpha_N \quad \text{for all} \quad t \in [a,b] \; .$$

Thus the error estimate and numerical efficiency of the discrete approximation in Theorem 6.4 depend on the evaluation of the averaged modulus of continuity $\tau(F;h)$ from (6.6) and the approximating quantity ξ_N defined in the proof of Theorem 6.4. Denoting

$$v(F) := \sup_k \Big\{ \sum_{i=1}^{k-1} \sup_x \big[\mathrm{haus}\big(F(x,t_{i+1}), F(x,t_i)\big), x \in U\big], a \le t_1 \le \ldots \le t_k \le b \Big\} \; ,$$

it is not hard to check that

$$\tau(F;h) \le v(F)h = O(h)$$

whenever $F(x,\cdot)$ has a *bounded variation* $v(F) < \infty$ uniformly in $x \in U$; see Dontchev and Farkhi [354]. Furthermore, one has the estimate

$$\xi_N \le 2\tau(\dot{\bar{x}}; h_N)$$

by taking $w_N(t) = \dot{\widehat{x}}_N(t) = \dot{\bar{x}}(t_j)$ for $t \in [t_j, t_j + h_N)$ if $\dot{\bar{x}}(\cdot)$ is *Riemann integrable* on $[a,b]$.

Remark 6.6 (discrete approximations of one-sided Lipschitzian differential inclusions). The Lipschitz continuity and compact-valuedness assumptions on F in Theorem 6.4 can be relaxed under additional requirements on the state space X in question. In particular, some counterparts of the $\mathcal{C}\big([a,b];X\big)$-approximation and $W^{1,2}\big([a,b];X\big)$-approximation results in the above theorem are obtained by Donchev and Mordukhovich [346] for the Hilbert pace setting with replacing the classical Lipschitz continuity in (H1) by the following *one-sided Lipschitzian property* of F in x: there is a constant $\ell \in I\!R$ (not necessarily positive) such that

$$\sigma\big(x_1 - x_2; F(x_1,t)\big) \le \ell\|x_1 - x_2\|^2 \text{ whenever } x_1, x_2 \in U, \text{ a.e. } t \in [a,b] ,$$

where $\sigma(x;Q) := \sup_{q\in Q}\langle x, q\rangle$ stands for the *support function* of $Q \subset X$. Moreover, the compact-valuedness assumption on the mapping $F(\cdot,t)$ may be replaced by imposing its *boundedness on bounded sets*: see the mentioned paper for more details and discussions.

6.1.2 Bolza Problem for Differential Inclusions and Relaxation Stability

In this subsection we start considering the following problem of dynamic optimization over solutions (in the sense of Definition 6.1) to differential inclusions in Banach spaces: minimize the *Bolza functional*

$$J[x] := \varphi\big(x(a), x(b)\big) + \int_a^b \vartheta\big(x(t), \dot{x}(t), t\big)\,dt \tag{6.13}$$

over trajectories $x\colon [a,b] \to X$ for the differential inclusion (6.1) such that $\vartheta\big(x(t), \dot{x}(t), t\big)$ is Bochner integrable on the fixed time interval $T := [a,b]$ subject to the *endpoint constraints*

$$\big(x(a), x(b)\big) \in \Omega \subset X^2 . \tag{6.14}$$

This problem is labeled by (P) and called the (generalized) *Bolza problems for differential inclusions*. We use the term *arc* for any solution $x = x(\cdot)$ to (6.1) with $J[x] < \infty$ and the term *feasible arc* for arcs $x(\cdot)$ satisfying the endpoint constraints (6.14). Since the dynamic (6.1) and endpoint (6.14) constraints are given explicitly, we may assume that both functions φ and ϑ in the cost functional (6.13) take finite values.

The formulated problem (P) covers a broad range of various problems of dynamic optimization in finite-dimensional and infinite-dimensional spaces. In particular, it contains both standard and nonstandard models in optimal control for parameterized control systems (6.2) with possibly closed-loop control sets $U(x,t)$. Note also that problems with free time (non-fixed time intervals), with integral constraints on $(x,\dot{x})$, and with some other types of state constraints can be reduced to the form of (P).

Aiming to derive necessary conditions for optimal solutions to (P) that would apply not only to *global* but also to *local* minimizers, we first introduce appropriate concepts of local minima. Our basic notion is as follows.

Definition 6.7 (intermediate local minima). *A feasible arc $\bar{x}$ is an* INTERMEDIATE LOCAL MINIMIZER *(i.l.m.) of rank $p \in [1,\infty)$ for (P) if there are numbers $\varepsilon > 0$ and $\alpha \geq 0$ such that $J[\bar{x}] \leq J[x]$ for any feasible arcs to (P) satisfying*

$$\|x(t) - \bar{x}(t)\| < \varepsilon \ \ \textit{for all} \ \ t \in [a,b] \quad \textit{and} \tag{6.15}$$

$$\alpha \int_a^b \|\dot{x}(t) - \dot{\bar{x}}(t)\|^p \, dt < \varepsilon \ . \tag{6.16}$$

Relationships (6.15) and (6.16) actually mean that we consider a neighborhood of $\bar{x}$ in the Sobolev space $W^{1,p}\big([a,b];X\big)$. If there is only requirement (6.15) in Definition 6.7, i.e., $\alpha = 0$ in (6.16), that one gets the classical *strong* local minimum corresponding to a neighborhood of $\bar{x}$ in the norm topology of $\mathcal{C}\big([a,b];X\big)$. If instead of (6.16) one puts the more restrictive requirement

$$\|\dot{x}(t) - \dot{\bar{x}}(t)\| < \varepsilon \ \ \text{a.e.} \ \ t \in [a,b] \ ,$$

then we have the classical *weak* local minimum in the framework of Definition 6.7. This corresponds to considering a neighborhood of $\bar{x}$ in the topology of $W^{1,\infty}\big([a,b];X\big)$. Thus the introduced notion of i.l.m. takes, for any $p \in [1,\infty)$, an *intermediate* position between the classical concepts of strong $(\alpha = 0)$ and weak $(p = \infty)$ local minima. Clearly all the necessary conditions for i.l.m. automatically hold for strong (and hence for global) minimizers. Let us consider some examples that illustrate relationships between weak, intermediate, and strong local minimizers in variational problems.

The first example is standard showing that the notions of weak and strong minimizers are distinct in the simplest problems of the classical calculus of variations with endpoint constraints.

Example 6.8 (weak but not strong minimizers). *There is a problem of the classical calculus of variations for which a weak local minimizer is not a strong local minimizer.*

Proof. Consider the variational problem:

$$\text{minimize} \ \ J[x] := \int_0^\pi x^2(t)[1 - \dot{x}^2(t)] \, dt$$

over absolutely continuous functions $x\colon[0,\pi] \to I\!R$ satisfying the endpoint constraints $x(0) = x(\pi) = 0$. Let us first show that $\bar{x}(\cdot) \equiv 0$ is a *weak local minimizer*. Indeed, taking any $\varepsilon \in (0,1)$ and any feasible arc $x \neq \bar{x}$ satisfying

$$|x(t) - \bar{x}(t)| \le \varepsilon,\ t \in [0,\pi], \quad \text{and} \quad |\dot{x}(t) - \dot{\bar{x}}(t)| \le \varepsilon \ \text{ a.e. } \ t \in [0,\pi]\,,$$

one has $0 < 1-\varepsilon^2 \le 1-\dot{x}^2(t)$ for almost all $t \in [0,\pi]$. Thus $x^2(t)[1-\dot{x}^2(t)] > 0$ a.e. $t \in [0,\pi]$ with $J[x] > 0 = J[\bar{x}]$, i.e., $\bar{x}$ is a weak local minimizer. On the other hand, $\bar{x}$ is *not a strong local minimizer*, which can be justified as follows. Take feasible arcs $x_k(t) := (1/\sqrt{k})\sin(kt)$ for any $k \in I\!N$ and observe that

$$J[x_k] = \frac{\pi}{2}\Big(\frac{1}{k} - \frac{1}{4}\Big) < 0 \ \text{ for } \ k \ge 5$$

while $|x_k(t) - \bar{x}(t)| \le 1/\sqrt{k}$ for all $t \in [0,\pi]$ and $k \in I\!N$. Thus, given any $\varepsilon > 0$, we can always find a feasible arc x_k that belongs to the ε-neighborhood of $\bar{x}$ in $\mathcal{C}([0,\pi]; I\!R)$ with $J[x_k] < J[\bar{x}]$. △

Next let us consider a less standard situation when a weak local minimizer may not be an intermediate local minimizer in the sense of Definition 6.7 for any rank $p \in [1,\infty)$. Again it happens in the one-dimensional framework of the classical calculus of variations.

Example 6.9 (weak but not intermediate minimizers). *There is a one-dimensional problem of the calculus of variations for which a weak local minimizer is not an intermediate local minimizer of any rank* $p \ge 1$.

Proof. Consider the variational problem:

$$\text{minimize} \ \ J[x] := \int_0^1 \big[\dot{x}^3(t) + 3\dot{x}^2(t)\big]\,dt$$

over absolutely continuous function $x\colon[0,1] \to I\!R$ satisfying the endpoint constraints $x(0) = x(1) = 0$. To show that $\bar{x}(\cdot) \equiv 0$ is a *weak local minimizer*, we observe that the integrand is non-negative whenever $\dot{x}(t) \ge -3$, and hence $J[x] > 0$ for every feasible arc x with

$$0 < |\dot{x}(t) - \dot{\bar{x}}(t)| \le \varepsilon < 3 \ \text{ a.e. } \ t \in [0,1]\,.$$

Given any $p \ge 1$, let us now prove that $\bar{x}$ is *not an intermediate local minimizer* of rank p. To proceed, we consider the family of feasible arcs

$$x_k(t) := \begin{cases} -k^{\frac{1}{2p}}t & \text{if } 0 \le t \le \frac{1}{k}\,, \\[2ex] \dfrac{-k^{\frac{1}{2p}}(1-t)}{k-1} & \text{if } \frac{1}{k} < t \le 1 \end{cases}$$

for natural numbers $k \geq 3^{4p}$. One can check that

$$J[x_k] = -\frac{k^{\frac{1}{p}}}{(k-1)^2}\Big[(k^{\frac{1}{2p}} - 3)(k-2) - 3\Big] < 0 \quad \text{and}$$

$$\int_0^1 |\dot{x}_k(t) - \dot{\bar{x}}(t)|^p = \frac{1}{\sqrt{k}^p}\Big(1 + \frac{1}{(k-1)^{p-1}}\Big)^p \leq \Big(\frac{2}{\sqrt{k}}\Big)^p .$$

Thus for any $\varepsilon > 0$ and any $p \geq 1$ we have

$$\int_0^1 |\dot{x}_k(t) - \dot{\bar{x}}(t)|^p \leq \varepsilon^p \;\text{ with }\; J[x_k] < 0 \;\text{ whenever }\; k \geq \max\big\{\varepsilon^{-2p}, 3^{4p}\big\} ,$$

which shows that $\bar{x}$ cannot be an intermediate minimizer of rank p.

Considering the *simplified version*

$$\text{minimize } \; J[x] := \int_0^1 \dot{x}^3(t)\,dt \;\text{ subject to }\; x(0) = 0, \;\; x(1) = 1$$

of the above problem, observe that the arc $\bar{x}(t) = t$ is a weak local minimizer while not an intermediate local minimizer of any rank $p \geq 2$ (but not of $p \geq 1$). To show the latter, we take the functions $x_k(t) = \bar{x}(t) + y_k(t)$ with $y_k(0) = y_k(1) = 0$ and

$$\dot{y}_k(t) = \begin{cases} -\sqrt{k} & \text{if } 0 \leq t \leq \frac{1}{k} , \\ \sqrt{k}(k-1)^{-1} & \text{if } \frac{1}{k} < t \leq 1 \end{cases}$$

and check directly that

$$J[x_k] = -\sqrt{k} + O(1) \to -\infty \;\text{ while }\; \int_0^1 |\dot{x}_k(t) - \dot{\bar{x}}(t)|^p\,dt \to 0 \;\text{ as }\; k \to \infty$$

for each $p \in [2, \infty)$, which completes the discussion. △

The previous examples concerned problems of the calculus of variations with no differential inclusion/dynamic constraints. The next example deals with *autonomous, convex-valued, Lipschitzian* differential inclusions and demonstrates that the concepts of strong and intermediate local minimizers may be different in this case.

Example 6.10 (intermediate but not strong minimizers for bounded, convex-valued, and Lipschitzian differential inclusions). *There is an optimal control problem of minimizing a linear cost function over trajectories of an autonomous, uniformly bounded, and Lipschitzian differential inclusion with compact and convex values for which an intermediate local minimizer of any rank* $p \in [1, \infty)$ *is not a strong local minimizer.*

Proof. Let $x = (x_1, x_2, x_3, x_4) \in I\!R^4$, and let

$$\psi(x_1, x_2) := \begin{cases} x_2^2 \cos\left(\frac{\pi x_1}{x_2}\right) & \text{for } x_2 \neq 0\,, \\ 0 & \text{for } x_2 = 0\,. \end{cases}$$

It is easy to check that ψ is continuously differentiable on $I\!R^4$. Consider the following problem:

$$\text{minimize} \;\; J[x] := -x_2(1)$$

over absolutely continuous trajectories for the differential inclusion

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{pmatrix} \in \left\{ \begin{bmatrix} 1 \\ 0 \\ \upsilon \\ |\psi(x_1, x_2) - x_2 x_3| \end{bmatrix} \;\middle|\; \upsilon \in [-4, 4] \right\} \quad \text{a.e. } \; t \in [0, 1]$$

with the endpoint constraints

$$x_1(0) = x_4(0) = x_4(1) = 0, \qquad x_1(1) = 1\,.$$

Take a feasible arc $\bar{x}(t) = (t, 0, 0, 0)$ and show first that it is *not a strong local minimizer*. Indeed, for any $\varepsilon \in (0, 2\sqrt{2})$ the function

$$x(t) = \left(t,\; \frac{\varepsilon}{\sqrt{2}},\; \frac{\varepsilon}{\sqrt{2}} \cos\left(\frac{\sqrt{2}\pi t}{\varepsilon}\right),\; 0\right)$$

is a feasible arc from the ε-neighborhood of $\bar{x}$ in the space $\mathcal{C}\big([0, 1]; I\!R^4\big)$ with the cost $J[x] = -\varepsilon/\sqrt{2} < 0 = J[\bar{x}]$.

Next let us show that $\bar{x}$ is an *intermediate local minimizer* of rank $p = 1$, and hence of any rank $p \in [1, \infty)$, for the problem under consideration. Choose any $\varepsilon \in (0, 1/2)$ and assume on the contrary that there is a feasible arc $x(\cdot)$ satisfying the relations (6.15) and (6.16) in Definition 6.7 and such that $J[x] < J[\bar{x}]$. Then

$$x_1(t) = t, \quad x_2(t) \equiv c, \quad \text{and} \quad |\psi(t, c) - c x_3(t)| \equiv 0$$

on $[0, 1]$ for some $c \in (0, 1/2)$. This gives

$$x_3(t) = \psi(t, c) = c \cos\left(\frac{\pi t}{c}\right), \quad \text{and hence} \quad \dot{x}_3 = \pi \sin\left(\frac{\pi t}{c}\right)\,.$$

Therefore one has

$$\int_0^1 \|\dot{x}(t) - \dot{\bar{x}}(t)\|\, dt = \pi \int_0^1 \left|\sin\left(\frac{\pi t}{c}\right)\right| dt = \pi c \int_0^{c^{-1}} \big|\sin(\pi s)\big|\, ds$$

$$\geq \pi c \int_0^{[c^{-1}]} \big|\sin(\pi s)\big|\, ds = 2c\left[\frac{1}{c}\right] \geq \frac{2}{3}$$

due to $c \in (0, 1/2)$, where $[a]$ stands as usual for the greatest integer less than or equal to $a \in I\!R$. The latter clearly contradicts the choice of $\varepsilon < 1/2$, which proves that $\bar{x}$ is an intermediate local minimizer of rank $p = 1$. △

In what follows, along with the original problem (P), we consider its *relaxed* counterpart that, roughly speaking, is obtained from (P) by the *convexification* procedure with respect to the velocity variable. Taking the integrand $\vartheta(x, v, t)$ in (6.13), we consider its restriction

$$\vartheta_F(x, v, t) := \vartheta(x, v, t) + \delta\big(v; F(x, t)\big)$$

to the sets $F(x, t)$ in (6.1) and denote by $\widehat{\vartheta}_F(x, v, t)$ the *biconjugate* (bypolar) function to $\vartheta_F(x, \cdot, t)$, i.e.,

$$\widehat{\vartheta}_F(x, v, t) = \big(\vartheta_F\big)_v^{**}(x, v, t) \;\text{ for all }\; (x, v, t) \in X \times X \times [a, b]\,.$$

It is well known that $\widehat{\vartheta}_F(x, v, t)$ is the *greatest proper, convex, l.s.c.* function with respect to v, which is *majorized* by ϑ_F. Moreover, $\vartheta_F = \widehat{\vartheta}_F$ *if and only if* ϑ_F is proper, convex, and l.s.c. with respect to v.

Given the original variational problem (P), we define the *relaxed problem* (R), or the *relaxation* of (P), as follows:

$$\text{minimize }\; \widehat{J}[x] := \varphi\big(x(a), x(b)\big) + \int_a^b \widehat{\vartheta}_F\big(x(t), \dot{x}(t), t\big)\, dt \tag{6.17}$$

over a.e. differentiable arcs $x\colon [a, b] \to X$ that are Bochner integrable on $[a, b]$ together with $\vartheta_F\big(x(t), \dot{(x)}(t), t\big)$, satisfy the Newton-Leibniz formula on $[a, b]$ and the endpoint constraints (6.14). Note that, in contrast to (6.13), the integrand in (6.17) is extended-real-valued. Furthermore, the natural requirement $\widehat{J}[x] < \infty$ yields that $x(\cdot)$ is a solution (in the sense of Definition 6.1) to the *convexified differential inclusion*

$$\dot{x}(t) \in \operatorname{clco} F\big(x(t), \dot{x}(t), t\big) \;\text{ a.e. }\; t \in [a, b]\,. \tag{6.18}$$

Thus the relaxed problem (R) can be considered under explicit dynamic constrained given by the convexified differential inclusion (6.18). Any trajectory for (6.18) is called a *relaxed trajectory* for (6.1), in contrast to *original trajectories*/arcs for the latter inclusion.

There are deep relationships between relaxed and original trajectories for differential inclusion, which reflect *hidden convexity* inherent in continuous-time (nonatomic measure) dynamic systems defined by differential operators. We'll see various realizations of this phenomenon in the subsequent material of this chapter. In particular, *any relaxed trajectory* of compact-valued and Lipschitz in x differential inclusion in Banach spaces may be *uniformly approximated* (in the space $\mathcal{C}\big([a, b]; X\big)$ by original trajectories starting with the same initial state $x(a) = x_0$; see, e.g., Theorem 2.2.1 in Tolstonogov [1258] with the references therein. We need a version of this approximation/density property involving not only differential inclusions but also minimizing functionals. The following result, which holds when the underlying Banach space is *separable*, is proved by De Blasi, Pianigiani and Tolstonogov [308]. Results of this type go back to the classical theorems of Bogolyubov [121] and Young [1350] in the calculus of variations.

Theorem 6.11 (approximation property for relaxed trajectories). *Let $x(\cdot)$ be a relaxed trajectory for the differential inclusion* (6.1)*, where X is separable, and where $F\colon X \times [a,b] \rightrightarrows X$ is compact-valued and uniformly bounded by a summable function, locally Lipschitzian in x, and measurable in t. Assume also that the integrand ϑ in* (6.13) *is continuous in (x,v), measurable in t, and uniformly bounded by a summable function near $x(\cdot)$. Then there is sequence of the original trajectories $x_k(\cdot)$ for* (6.1) *satisfying the relations*

$$x_k(a) = x(a), \quad x_k(\cdot) \to x(\cdot) \ \text{ in } \ \mathcal{C}\big([a,b];X\big), \quad \text{and}$$

$$\liminf_{k\to\infty} \int_a^b \vartheta\big(x_k(t),\dot{x}_k(t),t\big)\,dt \le \int_a^b \widehat{\vartheta}_F\big(x(t),\dot{x}(t),t\big)\,dt\ .$$

Note that Theorem 6.11 *doesn't* assert that the approximating trajectories $x_k(\cdot)$ satisfy the endpoint constraints (6.14). Indeed, there are examples showing that the latter may not be possible. If they do, then problem (P) has the property of *relaxation stability*:

$$\inf(P) = \inf(R)\ , \tag{6.19}$$

where the infima of the cost functionals (6.13) and (6.17) are taken over all the feasible arcs in (P) and (R), respectively.

An obvious sufficient condition for the relaxation stability is the *convexity* of the sets $F(x,t)$ and of the integrand ϑ in v. However, the relaxation stability goes far beyond the standard convexity due to the hidden convexity property of continuous-time differential systems. In particular, Theorem 6.11 ensures the relaxation stability of nonconvex problems (P) with no constraints on $x(b)$. There are other efficient conditions for the relaxation stability of nonconvex problems discussed, e.g., in Ioffe and Tikhomirov [617], Mordukhovich [888, 915], and Tolstonogov [1258]. Let us mention the classical Bogolyubov theorem ensuring the relaxation stability in variational problems of minimizing (6.13) with endpoint constraint (6.14) but with *no dynamic constraints* (6.1); relaxation stability of control systems *linear in state variables* via the fundamental Lyapunov theorem on the range convexity of nonatomic vector measures that largely justifies the hidden convexity; the *calmness* condition by Clarke [246, 255] for differential inclusion problems with endpoint constraints of the inequality type; the *normality* condition by Warga [1315, 1321] involving parameterized control systems (6.2), etc.

An essential part of our study relates to *local minima* that are *stable with respect to relaxation*. The corresponding counterpart of Definition 6.7 is formulated as follows.

Definition 6.12 (relaxed intermediate local minima). *The arc $\bar{x}$ is a* RELAXED INTERMEDIATE LOCAL MINIMIZER *(r.i.l.m.) of rank $p \in [1, \infty)$ for the original problem (P) if $\bar{x}$ is a feasible solution to (P) and provides an intermediate local minimum of this rank to the relaxed problem (R) with the same cost $J[\bar{x}] = \widehat{J}[\bar{x}]$.*

The notions of *relaxed weak* and *relaxed strong local minima* are defined similarly, with the same relationships between them as discussed above. Of course, there is no difference between the corresponding relaxed and usual (non-relaxed) notions of local minima for problems (P) with convex sets $F(x,t)$ and integrands ϑ convex with respect to velocity. It is also clear that any relaxed intermediate (weak, strong) minimizer for (P) provides the corresponding non-relaxed local minimum to the original problem. The opposite requires a kind of *local* relaxation stability. Note that any necessary condition for r.i.l.m. holds for relaxed strong local minima, and hence for optimal solutions to (P) (global or absolute minimizers) under the relaxation stability (6.19) of this problem.

Our primary goal is to derive general necessary optimality conditions for r.i.l.m. in the Bolza problem (P) under consideration; some results will be later obtained without any relaxation as well. To proceed, we employ the *method of discrete approximations*, which relates variational/optimal control problems for continuous-time systems to their finite-difference counterparts. The first step in this direction is to build *well-posed* discrete approximations of a *given* r.i.l.m. $\bar{x}(\cdot)$ in problem (P) such that optimal solutions to discrete-time problems *strongly converge* to $\bar{x}(\cdot)$ in the space $W^{1,2}([a,b];X)$. This will be accomplished in the next subsection.

6.1.3 Well-Posed Discrete Approximations of the Bolza Problem

Considering differential inclusions and their finite-difference counterparts in Subsect. 6.1.1, we established there that *every* trajectory for a differential inclusion in a general Banach space can be *strongly approximated* by extended trajectories for finite-difference inclusions under the natural assumptions made. This result doesn't directly relate to optimization problems involving differential inclusions, but we are going to employ it now in the optimization framework. The primary objective of this subsection is as follows.

Given a trajectory $\bar{x}(\cdot)$ for the differential inclusion (6.1), which provides a *relaxed intermediate local minimum* (r.i.l.m.) to the optimization problem (P) defined above, construct a *well-posed* family of approximating optimization problems (P_N) for finite-difference inclusions (6.3) such that (extended) *optimal solutions* $\bar{x}_N(\cdot)$ to (P_N) *strongly converge* to $\bar{x}(\cdot)$ in the norm topology of $W^{1,2}([a,b];X)$.

Imposing the standing hypotheses (H1) and (H2) formulated in Subsect. 6.1.1, we observe that the boundedness assumption (6.4) implies that the notion of r.i.l.m. from Definition 6.12 *doesn't depend on rank* p from the interval $[1,\infty)$. This means that $\bar{x}(\cdot)$ is an r.i.l.m. of some rank $p \in [1,\infty)$, then it is also an r.i.l.m. of any other rank $p \ge 1$. In what follows we take $p = 2$ and $\alpha = 1$ in (6.16) for simplicity.

To proceed, one needs to impose proper assumptions on the other data ϑ, φ, and Ω of problem (P) in addition to those imposed on F. Dealing with the Bochner integral, we always identify measurability of mappings $f\colon [a,b] \to X$ with *strong measurability*. Recall that f is strongly measurable if it can be a.e. approximated by a sequence of step X-valued functions on measurable subsets of $[a,b]$. Given a neighborhood U of $\bar{x}(\cdot)$ and a constant m_F from (H1), we further assume that:

(H3) $\vartheta(\cdot,\cdot,t)$ is continuous on $U \times (m_F I\!B)$ uniformly in $t \in [a,b]$, while $\vartheta(x,v,\cdot)$ is measurable on $[a,b]$ and its norm is majorized by a summable function uniformly in $(x,v) \in U \times (m_F I\!B)$.

(H4) φ is continuous on $U \times U$; $\Omega \subset X \times X$ is locally closed around $(\bar{x}(a),\bar{x}(b))$ and such that the set $\mathrm{proj}_1\Omega \cap (\bar{x}(a) + \varepsilon I\!B)$ is compact for some $\varepsilon > 0$, where $\mathrm{proj}_1\Omega$ stands for the projection of Ω on the first space X in the product space $X \times X$.

Note that symmetrically we may assume the local compactness of the second projection of $\Omega \subset X \times X$; the first one is selected in (H4) just for definiteness.

Now taking the r.i.l.m. $\bar{x}(\cdot)$ under consideration, let us apply to this feasible arc Theorem 6.4 on the strong approximation by discrete trajectories. Thus we find a sequence of the extended discrete trajectories $\widehat{x}_N(\cdot)$ approximating $\bar{x}(\cdot)$ and compute the numbers η_N in (6.11). Having $\varepsilon > 0$ from relations (6.15) and (6.16) of the intermediate minimizer $\bar{x}(\cdot)$ with $p = 1$ and $\alpha = 1$, we always suppose that $\bar{x}(t) + \varepsilon/2 \in U$ for all $t \in [a,b]$. Let us construct the sequence of discrete approximation problems (P_N), $N \in I\!N$, as follows: minimize the discrete-time Bolza functional

$$
\begin{aligned}
J_N[x_N] :&= \varphi\big(x_N(t_0), x_N(t_N)\big) + \|x_N(t_0) - \bar{x}(a)\|^2 \\
&+ \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \vartheta\Big(x_N(t_j), \frac{x_N(t_{j+1}) - x_N(t_j)}{h_N}, t\Big)\,dt \\
&+ \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \Big\|\frac{x_N(t_{j+1}) - x_N(t_j)}{h_N} - \dot{\bar{x}}(t)\Big\|^2\,dt
\end{aligned}
\tag{6.20}
$$

over discrete trajectories $x_N = x_N(\cdot) = \big(x_N(t_0),\ldots,x_N(t_N)\big)$ for the difference inclusions (6.3) subject to the constraints

$$\big(x(t_0), x_N(t_N)\big) \in \Omega + \eta_N I\!B, \tag{6.21}$$

$$\|x_N(t_j) - \bar{x}(t_j)\| \le \frac{\varepsilon}{2} \quad \text{for} \quad j = 1,\ldots,N, \quad \text{and} \tag{6.22}$$

$$\sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \Big\| \frac{x_N(t_{j+1}) - x_N(t_j)}{h_N} - \dot{\bar{x}}(t) \Big\|^2 dt \le \frac{\varepsilon}{2}\,. \tag{6.23}$$

As in Subsect. 6.1.1, we consider (without mentioning any more) piecewise linear extensions of $x_N(\cdot)$ to the whole interval $[a,b]$ with piecewise constant derivatives for which one has

$$\begin{cases} x_N(t) = x_N(a) + \displaystyle\int_a^t \dot{x}_N(s)\,ds \quad \text{for all } t \in [a,b] \quad \text{and} \\ \dot{x}_N(t) = \dot{x}_N(t_j) \in F(x_N(t_j),t_j), \quad t \in [t_j,t_{j+1}), \quad j = 0,\ldots,N-1\,. \end{cases} \tag{6.24}$$

The next theorem establishes that the given local minimizer $\bar{x}(\cdot)$ to (P) can be approximated by *optimal solutions* to (P_N) *strongly* in $W^{1,2}\big([a,b];X\big)$, which implies the a.e. *pointwise* convergence of the derivatives essential in what follows. To justify such an approximation, we need to impose both the Asplund structure and the Radon-Nikodým property (RNP) on the space X in question, which ensure the validity of the classical Dunford theorem on the weak compactness in $L^1\big([a,b];X\big)$. It is worth noting that every *reflexive* space is Asplund and has the RNP simultaneously. Furthermore, the second dual space X^{**} enjoys the RNP (and hence so does $X \subset X^{**}$) if X^* is Asplund. Thus the class of Banach spaces X for which both X and X^* are Asplund satisfies the properties needed in the next theorem. As discussed in the beginning of Subsect. 3.2.5, there are *nonreflexive* (even separable) spaces that fall into this category.

Theorem 6.13 (strong convergence of discrete optimal solutions). *Let $\bar{x}(\cdot)$ be an r.i.l.m. for the Bolza problem (P) under assumptions* (H1)–(H4)*, and let (P_N), $N \in I\!N$, be a sequence of discrete approximation problems built above. The following hold:*

(i) *Each (P_N) admits an optimal solution.*

(ii) *If in addition X is Asplund and has the RNP, then any sequence $\{\bar{x}_N(\cdot)\}$ of optimal solutions to (P_N) converges to $\bar{x}(\cdot)$ strongly in $W^{1,2}([a,b];X)$.*

Proof. To justify (i), we observe that the set of feasible trajectories to each problem (P_N) is nonempty for all large N, since the extended functions $\widehat{x}_N(\cdot)$

from Theorem 6.4 satisfy (6.3) and the constraints (6.21)–(6.23) by construction. This follows immediately from (6.11) in the case of (6.21) and (6.22). In the case of (6.23) we get from (6.4) and (6.12) that

$$\sum_{j=0}^{N-1}\int_{t_j}^{t_{j+1}}\left\|\frac{\widehat{x}_N(t_{j+1})-\widehat{x}_N(t_j)}{h_N}-\dot{\bar{x}}(t)\right\|^2 dt=\int_a^b\|\dot{\widehat{x}}_N(t)-\dot{\bar{x}}(t)\|^2\,dt \le 2m_F\alpha_N\le\frac{\varepsilon}{2}$$

for large N by the formula for α_N in (6.12). The existence of optimal solutions to (P_N) follows now from the classical Weierstrass theorem due to the compactness and continuity assumptions made in (H1), (H3), and (H4).

It remains to prove the convergence assertion (ii). Check first that

$$J_N[\widehat{x}_N]\to J[\bar{x}]\quad\text{as}\quad N\to\infty \tag{6.25}$$

along some sequence of $N\in\mathbb{N}$. Considering the expression (6.20) for $J_N[\widehat{x}_N]$, we deduce from Theorem 6.4 that the second terms therein vanishes, the forth term tends to zero due to (6.4) and (6.12), and the first term tends to $\varphi\big(\bar{x}(a),\bar{x}(b)\big)$ due to the continuity assumption on φ in (H4). It is thus sufficient to show that

$$\sigma_N:=\sum_{j=0}^{N-1}\int_{t_j}^{t_{j+1}}\vartheta\Big(\widehat{x}_N(t_j),\frac{\widehat{x}_N(t_{j+1})-\widehat{x}_N(t_j)}{h_N},t\Big)\,dt\to\int_a^b\vartheta(\bar{x}(t),\dot{\bar{x}}(t),t)\,dt$$

as $N\to\infty$. Using the sign "$\sim$" for expressions that are equivalent as $N\to\infty$, we get the relationships

$$\sigma_N=\sum_{j=0}^{N-1}\int_{t_j}^{t_{j+1}}\vartheta\big(\widehat{x}_N(t_j),\dot{\widehat{x}}_N(t),t\big)\,dt\sim\int_a^b\vartheta\big(\widehat{x}_N(t),\dot{\widehat{x}}_N(t),t\big)\,dt$$

$$\sim\int_a^b\vartheta\big(\bar{x}(t),\dot{\widehat{x}}_N(t),t\big)\,dt\sim\int_a^b\vartheta\big(\bar{x}(t),\dot{\bar{x}}(t),t\big)\,dt$$

by Theorem 6.4 ensuring the a.e. convergence $\dot{\widehat{x}}_N(t)\to\dot{\bar{x}}(t)$ along a subsequence of $N\to\infty$ and by the Lebesgue dominated convergence theorem for the Bochner integral that is valid under (H3).

Note that we have justified (6.25) for any intermediate (not relaxed) local minimizer $\bar{x}(\cdot)$ for the original problem (P) in an arbitrary Banach space X. Next let us show that (6.25) implies that

$$\lim_{N\to\infty}\Big[\beta_N:=\|\bar{x}_N(a)-\bar{x}(a)\|^2+\int_a^b\|\dot{\bar{x}}_N(t)-\dot{\bar{x}}(t)\|^2\,dt\Big]=0 \tag{6.26}$$

for every sequence of optimal solutions $\bar{x}_N(\cdot)$ to (P_N) provided that $\bar{x}(\cdot)$ is a *relaxed* intermediate local minimizer for the original problem, where the state space X is assumed to be Asplund and to satisfy the RNP.

Suppose that (6.26) is not true. Take a limiting point $\beta > 0$ of the sequence $\{\beta_N\}$ in (6.26) and let for simplicity that $\beta_N \to \beta$ for all $N \to \infty$. We are going to apply the Dunford theorem on the relative *weak compactness* in the space $L^1([a,b];X)$ (see, e.g., Diestel and Uhl [334, Theorem IV.1]) to the sequence $\{\dot{\bar{x}}_N(\cdot)\}$, $N \in I\!N$. Due to (6.24) and (H1) this sequence satisfies the assumptions of the Dunford theorem. Furthermore, both spaces X and X^* have the RNP, since the latter property for X^* is equivalent to the Asplund structure on X, as mentioned above. Hence we suppose without loss of generality that there is $v \in L^1([a,b];X)$ such that

$$\dot{\bar{x}}_N(\cdot) \to v(\cdot) \text{ weakly in } L^1\big([a,b];X\big) \text{ as } N \to \infty .$$

Since the Bochner integral is a linear continuous operator from $L^1([a,b];X)$ to X, it remains continuous if the spaces $L^1([a,b];X)$ and X are endowed with the weak topologies. Due to (6.21) and the assumptions on Ω in (H4), the set $\{\bar{x}_N(a)|\ N \in I\!N\}$ is relatively compact in X. Using (6.24) and the *compactness* property of solution sets for differential inclusions under the assumptions made in (H1) and (H2) (see, e.g., Tolstonogov [1258, Theorem 3.4.2]), we conclude that the sequence $\{\bar{x}_N(\cdot)\}$ contains a subsequence that converges to some $\widetilde{x}(\cdot)$ in the norm topology of the space $\mathcal{C}([a,b];X)$. Now passing to the limit in the Newton-Leibniz formula for $\bar{x}_N(\cdot)$ in (6.24) and taking into account the above convergences, one has

$$\widetilde{x}(t) = \widetilde{x}(a) + \int_a^t v(s)\,ds \text{ for all } t \in [a,b] ,$$

which implies the absolute continuity and a.e. differentiability of $\widetilde{x}(\cdot)$ on $[a,b]$ with $v(t) = \dot{\widetilde{x}}(t)$ for a.e. $t \in [a,b]$. Observe that $\widetilde{x}(\cdot)$ is a solution to the convexified differential inclusion (6.18). Indeed, since a subsequence of $\{\dot{\bar{x}}_N(\cdot)\}$ converges to $\dot{\widetilde{x}}(\cdot)$ weakly in $L^1([a,b];X)$, some *convex combinations* of $\dot{\bar{x}}_N(\cdot)$ converge to $\dot{\widetilde{x}}(\cdot)$ in the norm topology of $L^1([a,b];X)$, and hence *pointwisely* for a.e. $t \in [a,b]$. Passing to the limit in the differential inclusions for $\bar{x}_N(\cdot)$ in (6.24), we conclude that $\widetilde{x}(\cdot)$ satisfies (6.18). By passing to the limit in (6.21) and (6.22), we also conclude that $\widetilde{x}(\cdot)$ satisfies the endpoint constraints in (6.14) as well as

$$\|\widetilde{x}(t) - \bar{x}(t)\| \le \varepsilon/2 \text{ for all } t \in [a,b] .$$

Furthermore, the integral functional

$$I[v] := \int_a^b \|v(t) - \dot{\bar{x}}(t)\|^2\,dt$$

is lower semicontinuous in the weak topology of $L^2([a,b];X)$ due to the convexity of the integrand in v. Since the weak convergence of $\dot{\bar{x}}_N(\cdot) \to \dot{\widetilde{x}}(\cdot)$ in $L^1([a,b];X)$ implies the one in $L^2([a,b];X)$ by the boundedness assumption (6.4), and since

$$\int_a^b \|\dot{\bar{x}}_N(t) - \dot{\bar{x}}(t)\|^2\,dt = \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \Big\| \frac{\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)}{h_N} - \dot{\bar{x}}(t) \Big\|^2\,dt\ ,$$

the above lower semicontinuity and relation (6.23) imply that

$$\int_a^b \|\dot{\widetilde{x}}(t) - \dot{\bar{x}}(t)\|^2\,dt \le \liminf_{N\to\infty} \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \Big\| \frac{\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)}{h_N} - \dot{\bar{x}}(t) \Big\|^2\,dt \le \frac{\varepsilon}{2}\ .$$

Thus the arc $\widetilde{x}(\cdot)$ belongs to the ε-neighborhood of $\bar{x}(\cdot)$ in the space $W^{1,2}\big([a,b];X\big)$.

Let us finally show that the arc $\widetilde{x}(\cdot)$ gives a smaller value to cost functional (6.17) than $\bar{x}(\cdot)$. One always has

$$J_N[\bar{x}_N] \le J_N[\widehat{x}_N] \quad \text{for all large} \quad N \in I\!\!N\ ,$$

since each $\widehat{x}_N(\cdot)$ is feasible to (P_N). Now passing to the limit as $N \to \infty$ and taking into account the previous discussions as well as the construction of the convexified integrand $\widehat{\vartheta}_F$ in (6.17), we get from (6.25) that

$$\varphi\big(\widetilde{x}(a), \widetilde{x}(b)\big) + \int_a^b \widehat{\vartheta}_F(\widetilde{x}(t), \dot{\widetilde{x}}(t), t)\,dt + \beta \le J[\bar{x}]\ ,$$

which yields by $\beta > 0$ that $\widehat{J}[\widetilde{x}] < J[\bar{x}] = \widehat{J}[\bar{x}]$. The latter is impossible, since $\bar{x}(\cdot)$ is an r.i.l.m. for (P). Thus (6.26) holds, which obviously implies the desired convergence $\bar{x}_N(\cdot) \to \bar{x}(\cdot)$ in the norm topology of the space $W^{1,2}\big([a,b];X\big)$ and completes the proof of the theorem. △

The arguments developed in the proof of Theorem 6.13 allow us to establish efficient conditions for the *value convergence* of discrete approximations, which means that the optimal/infimal values of the cost functionals in the discrete approximation problems converge to the one in the original problem (P). Moreover, using the approximation property for relaxed trajectories from Theorem 6.11, we obtain in fact a *necessary and sufficient* condition for the value convergence in terms of an intrinsic property of the original problems.

Observe that the cost functional (6.20) as well as the constraints (6.22) and (6.23) in the discrete approximation problems (P_N) explicitly contain the given local minimizer $\bar{x}(\cdot)$ to (P). Considering below the value convergence of discrete approximations, we are *not* going to involve *any local minimizer* in the construction of discrete approximations and/or even to assume the *existence of optimal solutions* to the original problem. To furnish this, we consider a sequence of new discrete approximation problems $(\widetilde{P}_N)$ built as follows: minimize

$$\widetilde{J}_N[x_N] := \varphi\big(x_N(t_0), x_N(t_N)\big) + \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \vartheta\Big(x_N(t_j), \frac{x_N(t_{j+1}) - x_N(t_j)}{h_N}, t\Big)\,dt$$

subject to the discrete inclusions (6.3) and the *perturbed* endpoint constraints (6.21), where the sequence η_N is not yet specified. Note that problems $(\widetilde{P}_N)$ are constructively built upon the initial data of the original continuous-time problem. In the next theorem the notation $\widetilde{J}_N^0 := \inf(\widetilde{P}_N)$, $\inf(P)$, and $\inf(R)$ stands for the optimal value of the cost functional in problems $(\widetilde{P}_N)$, (P), and (R), respectively. Observe that optimal solutions to the discrete-time problems $(\widetilde{P}_N)$ always *exist* due to the assumptions (H1)–(H4) made in Theorem 6.13 under proper perturbations η_N of the endpoint constraints; see its proof.

Theorem 6.14 (value convergence of discrete approximations). *Let $U \subset X$ be an open subset of a Banach space X such that $x_k(t) \in U$ as $t \in [a,b]$ and $k \in I\!N$ for a minimizing sequence of feasible solutions to (P). Assume that hypotheses* (H1)–(H4) *are fulfilled with this set U, where $\bar{x}(a)+\varepsilon I\!B$ is replaced by* $\operatorname{cl} U$ *in* (H4). *The following assertions hold:*

(i) *There is a sequence of the endpoint constraint perturbations $\eta_N \downarrow 0$ in* (6.21) *such that*

$$\inf(R) \le \liminf_{N\to\infty} \widetilde{J}_N^0 \le \limsup_{N\to\infty} \widetilde{J}_N^0 \le \inf(P)\ , \tag{6.27}$$

where the left-hand side inequality requires that X is Asplund and has the RNP. Therefore the relaxation stability (6.19) *of (P) is sufficient for the value convergence of discrete approximations*

$$\inf(\widetilde{P}_N) \to \inf(P) \quad \textit{as} \quad N \to \infty$$

provided that X is Asplund and has the RNP.

(ii) *Conversely, the relaxation stability of (P) is also a necessary condition for the value convergence* $\inf(\widetilde{P}_N) \to \inf(P)$ *of the discrete approximations with arbitrary perturbations $\eta_N \downarrow 0$ of the endpoint constraints provided that X is reflexive and separable.*

Proof. Let us first prove that the right-hand side inequality in (6.27) holds in any Banach space X. Taking the minimizing sequence of feasible arcs $x_k(\cdot)$ to (P) specified in the theorem, we apply to each $x_k(\cdot)$ Theorem 6.4 on the strong approximation by discrete trajectories. Involving the diagonal process, we build the extended discrete trajectories $\widehat{x}_N(\cdot)$ for (6.3) such that

$$\eta_N := \|\big(\widehat{x}_N(a), \widehat{x}_N(b)\big) - \big(x_{k_N}(a), x_{k_N}(b)\big)\| \to 0 \quad \text{as} \quad N \to \infty$$

and consider the sequence of discrete approximation problems $(\widetilde{P}_N)$ with these constraint perturbations η_N in (6.21). Similarly to the proof of the first part of Theorem 6.13, we show that each $(\widetilde{P}_N)$ admits an optimal solution and, arguing by contradiction, one has the right-hand side inequality in (6.27). To justify the left-hand side inequality in (6.27), we follow the proof of the second part of Theorem 6.13 assuming that X is Asplund and enjoys the RNP. This automatically implies the value convergence of $\inf(\widetilde{P}_N) \to \inf(P)$ under the relaxation stability of (P).

To prove the converse assertion (ii) in the theorem, we first observe that the relaxed problem (R) admits an optimal solution under the assumptions made; see Tolstonogov [1258, Theorem A.1.3]. It follows from the arguments in the second part of Theorem 6.13 that actually justify, under the assumptions made, the compactness of feasible solutions to the relaxed problem and the lower semicontinuity of the minimizing functional (6.17) in the topology on the set of feasible solutions $x(\cdot)$ induced by the weak convergence of the derivatives $\dot{x}(\cdot) \in L^1\big([a,b];X\big)$ provided that X is Asplund and has the RNP. Assume now that X is reflexive and separable and, employing Theorem 6.11, approximate a certain relaxed optimal trajectory $\bar{x}(\cdot)$ by a sequence of the original trajectories $x_k(\cdot)$ converging to $\bar{x}(\cdot)$ as established in that theorem. In turn, each $x_k(\cdot)$ can be strongly approximated in $W^{1,2}\big([a,b];X\big)$ by discrete trajectories $\widehat{x}_{k_N}(\cdot)$ due to Theorem 6.4. Using the diagonal process, we get a sequence of the discrete trajectories $\widehat{x}_N(\cdot)$ approximating $\bar{x}(\cdot)$ and put

$$\eta_N := \big\|\big(\widehat{x}_N(a), \widehat{x}_N(b)\big) - \big(\bar{x}(a), \bar{x}(b)\big)\big\| \to \infty \quad \text{as} \quad N \to \infty\,.$$

Now assume that problem (P) is not stable with respect to relaxation, i.e., $\inf(R) < \inf(P)$, and show that

$$\liminf_{N\to\infty} \widetilde{J}_N^0 < \inf(P)$$

for a sequence of discrete approximation problems $(\widetilde{P}_N)$ with some perturbations η_N of the endpoint constraints (6.21). Indeed, having

$$\inf(R) = \varphi\big(\bar{x}(a), \bar{x}(b)\big) + \int_a^b \widehat{\vartheta}_F(\bar{x}(t), \dot{\bar{x}}(t), t)\,dt < \inf(P)$$

for the relaxed optimal trajectory $\bar{x}(\cdot)$, we build η_N as above and consider problems $(\widetilde{P}_N)$ with these perturbations of the endpoint constraints. Taking into account the approximation of $\bar{x}(\cdot)$ by $x_k(\cdot)$ due to Theorem 6.11, the strong approximation of $x_k(\cdot)$ by the discrete trajectories $\widehat{x}_N(\cdot)$ in Theorem 6.4, and the relations

$$\widetilde{J}_N^0 \le \varphi\big(\widehat{x}_N(t_0), \widehat{x}_N(t_N)\big) + \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \vartheta\left(\widehat{x}_N(t_j), \frac{\widehat{x}_N(t_{j+1}) - \widehat{x}_N(t_j)}{h_N}, t\right) dt$$

$$= \varphi\big(\widehat{x}_N(a), \widehat{x}_N(b)\big) + \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \vartheta\big(\widehat{x}_N(t_j), \dot{\widehat{x}}_N(t), t\big)\,dt,$$

we get by the absence of the relaxation stability that

$$\liminf_{N\to\infty} \widehat{J}_N^0 \le \liminf_{N\to\infty} \Big[\varphi\big(\widehat{x}_N(a), \widehat{x}_N(b)\big) + \int_a^b \vartheta\big(\widehat{x}_N(t), \dot{\widehat{x}}_N(t), t\big)\,dt\Big]$$

$$\le \varphi\big(\bar{x}(a), \bar{x}(b)\big) + \int_a^b \widehat{\vartheta}_F\big(\bar{x}(t), \dot{\bar{x}}(t), t\big)\,dt < \inf(P)\,.$$

Therefore we don't have the value convergence of discrete approximations for problems $(\widetilde{P}_N)$ corresponding to the above perturbations of the endpoint constraints. This justifies (ii) and completes the proof of the theorem. △

Thus the relaxation stability of (P), which is an intrinsic and natural property of continuous-time dynamic optimization problems, is actually a *criterion* for the value convergence of discrete approximations under appropriate perturbations of the endpoint constraints in (6.21). It follows from the proof of Theorem 6.14 that one can express the corresponding perturbations η_N via the averaged modulus of continuity (6.6) by

$$\eta_N = \tau(\dot{\bar{x}}; h_N) \to \infty \quad \text{as} \quad N \to \infty$$

provided that (P) admits an optimal solution $\bar{x}(\cdot)$ with the Riemann integrable derivative $\dot{\bar{x}}(\cdot)$ on $[a, b]$. Moreover, $\eta_N = O(h_N)$ if $\dot{\bar{x}}(t)$ is of bounded variation on this interval; see Subsect. 6.1.1.

Remark 6.15 (simplified form of discrete approximations). Observe that if $\vartheta(x, v, \cdot)$ is *a.e. continuous* on $[a, b]$ uniformly in (x, v) in some neighborhood of the optimal solution $\bar{x}(\cdot)$, then the cost functional in (6.20) in problem (P_N) can be replaced in Theorem 6.13 by

$$\begin{aligned} J_N[x_N] :&= \varphi\big(x_N(t_0), x_N(t_N)\big) + \|x_N(t_0) - \bar{x}(a)\|^2 \\ &+ h_N \sum_{j=0}^{N-1} \vartheta\Big(x_N(t_j), \frac{x_N(t_{j+1}) - x_N(t_j)}{h_N}, t_j\Big) \\ &+ \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \Big\|\frac{x_N(t_{j+1}) - x_N(t_j)}{h_N} - \dot{\bar{x}}(t)\Big\|^2 dt; \end{aligned} \tag{6.28}$$

and similarly for the cost functional in problem $(\widetilde{P}_N)$ used in Theorem 6.14. Indeed, this is an easy consequence of the fact that $\tau(\vartheta; h_N) \to 0$ as $N \to \infty$ for the averaged modulus of continuity (6.6) when $\vartheta(x, v, \cdot)$ is a.e. continuous. Denote by $(\overline{P}_N)$ the discrete approximation problem that differs from (P_N) of that the cost functional (6.20) is replaced by the simplified one (6.28). In what follows we consider both problems (P_N) and $(\overline{P}_N)$ using them to derive necessary optimality conditions for the original problem. The results obtained in these ways are distinguished by the assumptions on the initial data that allow us to justify the desired necessary optimality conditions. Namely, while the use of the simplified problems $(\overline{P}_N)$ as $N \to \infty$ requires the a.e. continuity assumption on ϑ with respect of t (versus the measurability), it relaxes the requirements on the state space X needed in the case of (P_N); see below.

6.1.4 Necessary Optimality Conditions for Discrete-Time Inclusions

Theorem 6.13 on the strong convergence of discrete approximations makes a *bridge* between optimal solutions to the discrete-time problems (P_N), as well as their simplified versions $(\overline{P}_N)$ from Remark 6.15, and the given relaxed intermediate local minimizer $\bar{x}(\cdot)$ for the original continuous-time problem (P). Our further strategy is as follows: first to establish necessary optimality conditions in the sequences of discrete approximation problems (P_N) and $(\overline{P}_N)$ and then to obtain, by passing to the limit as $N \to \infty$, necessary conditions for the given local minimizer to the original optimal control problem (P) governed by differential inclusions.

This subsection is devoted to the derivation of necessary optimality conditions in general discrete-time Bolza problems and their special counterparts for the discrete approximations problems (P_N) and $(\overline{P}_N)$. We explore *two approaches* to these issues. The first one involves reducing general dynamic optimization problems for discrete-time inclusions to non-dynamic problems of *mathematical programming with operator constraints* and then employing necessary optimality conditions for such problems obtained in Subsect. 5.1.2. The second approach is based on the specific features of the discrete approximation problems (P_N) and $(\overline{P}_N)$ and the use of *fuzzy calculus* results from Chaps. 2–4. The results derived by using these two approaches are not reduced to each other, and they require different assumptions. It happens, however, that the *approximate* necessary optimality conditions obtained via the second approach are more suitable for deriving the corresponding results for the continuous-time problem (P) in the next subsection, while those obtained via the first one are definitely of independent interest.

Let us start with the *first approach* and consider the following (non-dynamic) problem of *mathematical programming* (MP) with operator, inequality, and geometric constraints to which we can reduce our discrete-time problems of dynamic optimization:

$$\begin{cases} \text{minimize } \ \varphi_0(z) \ \text{ subject to} \\ \varphi_j(z) \le 0, \quad j = 1, \ldots, s\ , \\ f(z) = 0\ , \\ z \in \Xi_j \subset Z, \quad j = 1, \ldots, l\ , \end{cases} \tag{6.29}$$

where φ_j are real-valued functions on Z, where $f\colon Z \to E$ is a mapping between Banach spaces, and where $\Xi_j \subset Z$. This is a problem with operator constraints of the type considered in the end of Subsect. 5.1.2 with the only difference that now we have *many geometric constraints* given by the sets Ξ_j. As we see below, the geometric constraints in (6.29) arise from the discretized differential inclusions (6.3), and the number l of them is increasing as $N \to \infty$. Note that problem (MP) is *intrinsically nonsmooth*, even in the case of the smooth data f and φ_j in (6.29) and in the generating dynamic problems. Indeed, the nonsmoothness comes from the geometric constraints in (6.29), which reflect the *dynamics* governed by differential and finite-difference inclusions in the original problem (P) and its discrete approximations.

To derive necessary optimality conditions in problem (MP), one may apply Corollary 5.18 that concerns the problem like (6.29) but with many geometric constraints. Denote

$$\Xi := \Xi_1 \cap \ldots \cap \Xi_l$$

and replace the geometric constraints in (6.29) by $z \in \Xi$. Employing now the result of Corollary 5.18, we need to present necessary optimality conditions for problem (MP) via its initial data. This can be done by using calculus rules for generalized normals and the SNC property of set intersections developed in Chap. 3.

Proposition 6.16 (necessary conditions for mathematical programming with many geometric constraints). *Let $\bar z$ be a local optimal solution to problem* (6.29)*, where the spaces Z and E are Asplund and where the sets Ξ_j are locally closed around $\bar z$. Assume also that all φ_i are Lipschitz continuous around $\bar z$, that f is generalized Fredholm at $\bar z$, and that each Ξ_j is SNC at this point. Then there are real numbers $\{\mu_j \in I\!R|\ j = 0, \ldots, s\}$ as well as linear functionals $e^* \in E^*$ and $\{z_j^* \in Z^*|\ j = 1, \ldots, l\}$, not all zero, such that $\mu_j \ge 0$ for $j = 0, \ldots, s$ and*

$$\mu_j \varphi_j(\bar z) = 0 \quad \text{for} \quad j = 1, \ldots, s\ , \tag{6.30}$$

$$z_j^* \in N(\bar z; \Xi_j) \quad \text{for} \quad j = 1, \ldots, l\ , \tag{6.31}$$

$$-\sum_{j=1}^{l} z_j^* \in \partial\Big(\sum_{j=0}^{s} \mu_j \varphi_j\Big)(\bar z) + D_N^* f(\bar z)(e^*)\ . \tag{6.32}$$

Proof. Apply Corollary 5.18 to problem (6.29) with the condensed geometric constraint $z \in \Xi$ given by the intersection of the sets Ξ_j. Then we find $\{\mu_j \ge 0|\ j = 0, \ldots, s\}$ and $e^* \in E^*$, not all zero, such that μ_j satisfy the complementary slackness conditions in (6.30) and

$$0 \in \partial\Big(\sum_{j=0}^{s} \mu_j \varphi_j\Big)(\bar{z}) + D_N^* f(\bar{z})(e^*) + N(\bar{z}; \Xi) \tag{6.33}$$

provided that the intersection set Ξ is SNC at $\bar{z}$. The latter holds, by Corollary 3.81, if each Ξ_j is SNC at this point and the qualification condition

$$\Big[z_1^* + \ldots + z_l^* = 0, \quad z_j^* \in N(\bar{z}; \Xi_j)\Big] \Longrightarrow \Big[z_j^* = 0, \quad j = 1, \ldots, s\Big]$$

is fulfilled. Furthermore, the same qualification condition ensures, by Corollary 3.37, the intersection formula

$$N(\bar{z}; \Xi) \subset N(\bar{z}; \Xi_1) + \ldots + N(\bar{z}; \Xi_l)$$

when all but one of Ξ_j are SNC at $\bar{z}$. Substituting this into (6.33), we conclude that the fulfillment of the above qualification condition implies (6.32) with $(\mu_j, e^*) \neq 0$. At the same time, the violation of the qualification condition means that (6.32) holds with $(z_1^*, \ldots, z_l^*) \neq 0$ and all zero μ_j and e^*. This completes the proof of the proposition. △

Now let us consider the application of Proposition 6.16 to the following constrained *Bolza problem for discrete-time inclusions* labeled as (DP):

$$\begin{cases} \text{minimize} \;\; \varphi(x_0, x_N) + h \displaystyle\sum_{j=0}^{N-1} \vartheta_j\Big(x_j, \frac{x_{j+1} - x_j}{h}\Big) \;\; \text{subject to} \\ x_{j+1} \in x_j + hF_j(x_j) \;\; \text{for} \;\; j = 0, \ldots, N-1\,, \\ (x_0, x_N) \in \Xi \subset X^2\,, \end{cases}$$

where $F_j\colon X \rightrightarrows X$, where φ and ϑ_j are real-valued functions on X^2, and where $h > 0$ and $N \in I\!N$ are fixed. Observe that problem (DP) incorporates the basic structure of discrete approximation problems from the preceding subsection, for any fixed N, without taking into account the terms therein related to approximating the given intermediate local minimizer $\bar{x}(\cdot)$ for the original continuous-time problem (P).

Theorem 6.17 (necessary optimality conditions for discrete-time inclusions). *Let $\{\bar{x}_j |\; j = 0, \ldots, N\}$ be a local optimal solution to problem (DP). Assume that X is Asplund, that the sets Ξ and F_j are locally closed and SNC at $(\bar{x}_0, \bar{x}_N)$ and $\big(\bar{x}_j, (\bar{x}_{j+1} - \bar{x}_j)/h\big)$, respectively, and that the functions φ and ϑ_j are locally Lipschitzian around the corresponding points $\bar{x}_j$ for all $j = 0, \ldots, N$. Then there are $\lambda \geq 0$ and $\{p_j \in X^* |\; j = 0, \ldots, N\}$, not simultaneously zero, such that one has the extended Euler-Lagrange inclusion*

$$\Big(\frac{p_{j+1} - p_j}{h}, p_{j+1}\Big) \in \lambda \partial \vartheta_j\Big(\bar{x}_j, \frac{\bar{x}_{j+1} - \bar{x}_j}{h}\Big) + N\Big(\Big(\bar{x}_j, \frac{\bar{x}_{j+1} - \bar{x}_j}{h}\Big); \operatorname{gph} F_j\Big)$$

for all $j = 0, \ldots, N-1$ with the transversality inclusion

$$(p_0, -p_N) \in \lambda \partial \varphi(\bar{x}_0, \bar{x}_N) + N\big((\bar{x}_0, \bar{x}_N)); \Xi\big)\,.$$

Proof. It is easy to see that the discrete-time *dynamic optimization* problem (DP) can be equivalently written in the *non-dynamic* form of mathematical programming given by (6.29) with

$$z := (x_0, \ldots, x_N, v_0, \ldots, v_{N-1}) \in Z := X^{2N+1}, \; E := X^N, \; l := N \; ,$$

$$\varphi_0(z) := \varphi(x_0, x_N) + h \sum_{j=0}^{N-1} \vartheta_j(x_j, v_j), \quad \varphi_j(z) := 0 \;\text{ as }\; j \geq 1 \; ,$$

$$f(z) = \big(f_0(z), \ldots, f_{N-1}(z)\big) \;\text{ with}$$

$$f_j(z) := x_{j+1} - x_j - hv_j, \quad j = 0, \ldots, N-1 \; ,$$

$$\Xi_j := \big\{z \in X^{2N+1} \big| \; v_j \in F_j(x_j)\big\} \;\text{ for }\; j = 0, \ldots, N-1 \; ,$$

$$\Xi_N := \big\{z \in X^{2N+1} \big| \; (x_0, x_N) \in \Xi\big\}$$

Thus $\bar{z} := \big(\bar{x}_0, \ldots, \bar{x}_N, (\bar{x}_1 - \bar{x}_0)/h, \ldots, (\bar{x}_N - \bar{x}_{N-1})/h\big)$ is a local optimal solution to the (MP) problem (6.29) with the data defined above. The operator constraint mapping f is surely *generalized Fredholm* at $\bar{z}$; moreover, the sets Ξ_j, $j = 0, \ldots, N$, are obviously SNC at $\bar{z}$ under the assumptions imposed on F_j and Ξ. Since the cost function φ_0 is locally Lipschitzian around $\bar{z}$ and the product spaces Z and E are Asplund, we apply the necessary optimality conditions from Proposition 6.16 to the (MP) problem under consideration, which give us a number $\mu_0 \geq 0$ as well as linear functionals $z_j^* = (x_{0j}^*, \ldots, x_{Nj}^*, v_{0j}^*, \ldots, v_{(N-1)j}^*) \in (X^*)^{2N+1}$ for $j = 0, \ldots, N$ and $e^* = (e_0^*, \ldots, e_{N-1}^*) \in (X^*)^N$, not all zero, such that conditions (6.30)–(6.32) hold with the data defined above. It follows from the structure of Ξ_j in (6.37) that (6.31) is equivalent to

$$\begin{cases} (x_{ij}^*, v_{ij}^*) \in N\Big(\Big(\bar{x}_j, \dfrac{\bar{x}_{j+1} - \bar{x}_j}{h}\Big); \operatorname{gph} F_j\Big) \quad \text{and} \\[2ex] x_{ij}^* = v_{ij}^* = 0 \;\text{ if }\; i \neq j \;\text{ for all }\; j = 0, \ldots, N-1 \; ; \\[2ex] (x_{0N}^*, x_{NN}^*) \in N\big((\bar{x}_0, \bar{x}_N); \Xi\big) \;\text{ and }\; x_{iN}^* = v_{iN}^* = 0 \;\text{ otherwise} \; . \end{cases}$$

Denoting $\lambda := \mu_0$ and employing the sum rule for basic subgradients of locally Lipschitzian functions in Theorem 3.36, we get from (6.32) and the structures of φ_0 and f that there are

$$(x_0^*, x_N^*) \in \partial\varphi(\bar{x}_0, \bar{x}_N) \;\text{ and }\; (u_j^*, w_j^*) \in \partial\vartheta_j\Big(\bar{x}_j, \frac{\bar{x}_{j+1} - \bar{x}_j}{h}\Big)$$

for $j = 0, \ldots, N-1$ satisfying the relations

$$\begin{cases} -x_{00}^* - x_{0N}^* = \lambda\big(x_0^* + hu_0^*\big) - e_0^* \;, \\ -x_{jj}^* = \lambda hu_j^* + e_{j-1}^* - e_j^*, \quad j = 0,\ldots,N-1 \;, \\ -x_{NN}^* = \lambda x_N^* + e_{N-1}^* \;, \\ -v_{jj}^* = h\big(\lambda w_j^* - e_j^*\big), \quad j = 0,\ldots,N-1 \;. \end{cases}$$

Denoting finally

$$p_0 := -x_{0N}^* - \lambda x_0^* + e_0^* \;\text{ and }\; p_j := e_{j-1}^*, \quad j = 1,\ldots,N \;,$$

we arrive at the desired Euler-Lagrange and transversality inclusions with $\lambda \geq 0$ and $\{p_j \in X^* |\; j = 0,\ldots,N\}$ not equal to zero simultaneously. This completes the proof of the theorem. △

Let us return to our *discrete approximation* problems (P_N) and $(\overline{P}_N)$. Fixed any $N \in I\!N$, observe that problem $(\overline{P}_N)$ defined in (6.3), (6.21)–(6.23), and (6.28) reduces to the form of mathematical programming (6.29) that is just slightly different from the one for (DP). Indeed, letting

$$z := (x_0,\ldots,x_N,v_0,\ldots,v_{N-1}) \in Z := X^{2N+1},\; E := X^N,\; s := N+2,\; l := N \;,$$

we rewrite $(\overline{P}_N)$ as (6.29) with the following data:

$$\begin{aligned} \varphi_0(z) :&= \varphi(x_0,x_N) + \|x_0 - \bar{x}(a)\|^2 + h_N \sum_{j=0}^{N-1} \vartheta_j(x_j,v_j) \\ &+ \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \|v_j - \dot{\bar{x}}(t)\|^2\, dt \;, \end{aligned} \tag{6.34}$$

$$\varphi_j(z) := \begin{cases} \|x_{j-1} - \bar{x}(t_{j-1})\| - \varepsilon/2 \;\text{ for }\; j = 1,\ldots,N+1 \;, \\ \displaystyle\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|v_i - \dot{\bar{x}}(t)\|^2\, dt - \varepsilon/2 \;\text{ for }\; j = N+2 \;, \end{cases} \tag{6.35}$$

$$\begin{aligned} & f(z) = \big(f_0(z),\ldots,f_{N-1}(z)\big) \;\text{ with} \\ & f_j(z) := x_{j+1} - x_j - h_N v_j, \quad j = 0,\ldots,N-1 \;, \end{aligned} \tag{6.36}$$

$$\begin{aligned} & \Xi_j := \big\{z \in X^{2N+1} \big|\; v_j \in F_j(x_j)\big\} \;\text{ for }\; j = 0,\ldots,N-1 \;, \\ & \Xi_N := \big\{z \in X^{2N+1} \big|\; (x_0,x_N) \in \Omega_N\big\} \;, \end{aligned} \tag{6.37}$$

where $\vartheta_j(x,v) := \vartheta(x,v,t_j)$, $F_j(x) := F(x,t_j)$, and $\Omega_N := \Omega + \eta_N I\!B$. Notice that the only difference between the (MP) forms for (DP) and $(\overline{P}_N)$ is reflected by the terms in the cost functions and inequality constraints involving the given intermediate local minimizer $\bar{x}(\cdot)$ for the original continuous-time problem (P). These terms can be easily treated in deriving necessary optimality conditions similarly to the proof of Theorem 6.17. Moreover, the impact of these terms to necessary optimality conditions *disappears* in the limiting procedure as $N \to \infty$, i.e., they can be actually *ignored* from the viewpoint of necessary optimality conditions in the original problem (P); see below.

Similarly we observe that problem (P_N) defined in (6.3), (6.20)–(6.23) equivalently reduces to the (MP) form (6.29) with the cost function

$$\begin{aligned}\varphi_0(z) :&= \varphi(x_0, x_N) + \|x_0 - \bar{x}(a)\|^2 \\ &+ \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \Big[\vartheta(x_j, v_j, t) + \|v_j - \dot{\bar{x}}(t)\|^2\Big]\, dt\end{aligned} \tag{6.38}$$

and the same constraints (6.35)–(6.37). The difference between (6.34) and (6.38) consists of replacing

$$h_N \sum_{j=0}^{N-1} \vartheta_j(x_j, v_j) \quad \text{by} \quad \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \vartheta(x_j, v_j, t)\, dt\ ,$$

where the latter allows us to deal with *summable* (in Bochner's sense) integrands $\vartheta(x,v,\cdot)$. In order to derive necessary optimality conditions for problems involving measurable/summable integrands, we need an auxiliary result (certainly important for its own sake) ensuring the *subdifferentiation under the integral sign*, which can be viewed as an "infinite sum" (continuous measure) extension of the subdifferential sum rule for finite sums of Lipschitzian functions obtained in Subsect. 3.2.1. However, the validity of the integral result requires more restrictions on the space in question: we assume its reflexivity and separability versus the Asplund structure in the finite sum rule used in Theorem 6.17. Although the following subdifferential formula holds in rather general measure spaces, we present it only for the case of real intervals, say $T = [0,1]$, needed in subsequent applications. Recall that the *integral of a set-valued mapping* is always understood as the collection of integrals of its summable selections.

Lemma 6.18 (basic subgradients of integral functional). *Let X be a reflexive and separable Banach space. Given $\bar{x} \in X$, assume that $\varphi\colon X \times [0,1] \to I\!R$ is measurable in t for each x near $\bar{x}$ and locally Lipschitzian around $\bar{x}$ with a summable modulus on $[0,1]$. Then one has*

$$\partial\Big(\int_0^1 \varphi(\cdot,t)\,dt\Big)(\bar{x}) \subset \mathrm{cl} \int_0^1 \partial\varphi(\bar{x},t)\,dt\ , \tag{6.39}$$

where the subdifferential on the right-hand side is taken with respect to x*, and where the closure "*cl*" is taken with respect to the norm topology in* X^**.*

Proof. First we observe that the mapping $\partial\varphi(\bar{x},\cdot)\colon [0,1] \rightrightarrows X^*$ is closed-valued and *measurable* in the standard sense for set-valued mappings $F\colon T \rightrightarrows Y$, i.e., that the inverse image $F^{-1}(\Theta)$ is measurable for any open subset $\Theta \subset Y$; for closed-valued mappings such a measurability admits many other equivalent descriptions; see, e.g., Theorems 14.3 and 14.56 in Rockafellar and Wets [1165] that hold in infinite dimensions. Note also that, in the case of separable image spaces, this measurability is equivalent to *strong measurability* (i.e., the possibility of the a.e. pointwise approximation by a sequence of step mappings) that is specific for the Bochner integral under consideration. By the well-known theorems on *measurable selections* (see, e.g., the afore-mentioned book [1165] as well as the early book by Castaing and Valadier [229]) there are measurable singe-valued mappings $\xi\colon [0,1] \to X^*$ such that $\xi(t) \in \partial\varphi(\bar{x},t)$ for a.e. $t \in [0,1]$. Moreover, since X^* is separable and $\partial\varphi(\bar{x};\cdot)$ is *integrably bounded* by the summable Lipschitz modulus of $\varphi(\cdot,t)$ as easily follows from the assumptions made (see Corollary 1.81), every measurable selector ξ of $\partial\varphi(\bar{x};\cdot)$ is Bochner integrable on $[0,1]$. Hence the multivalued integral on the right-hand side of (6.39) is well-defined and nonempty.

It follows from Clarke [255, Theorem 2.7.2] that a counterpart of (6.39) holds with the replacement of the basic subdifferential by the Clarke generalized gradient of Lipschitz functions on both sides. Using now Theorem 3.57 and the reflexivity of X, we have

$$\partial\Big(\int_0^1 \varphi(\cdot,t)\,dt\Big)(\bar{x}) \subset \int_0^1 \operatorname{clco}\partial\varphi(\bar{x},t)\,dt\ ,$$

since the weak closure agrees with the norm closure for convex sets in reflexive spaces by the Mazur theorem. On the other hand, it is known as an infinite-dimensional extension of the celebrated Lyapunov-Aumann theorem (see, e.g., Sect. 1.1 in Tolstonogov [1258]) that

$$\int_0^1 \operatorname{clco} F(t)\,dt = \operatorname{cl}\int_0^1 F(t)\,dt$$

for every compact-valued, strongly measurable, and integrable bounded mapping. This gives (6.39) and ends the proof of the lemma. △

Based on Theorem 6.17 and the subsequent discussions, we can similarly formulate and justify the extended Euler-Lagrange and transversality inclusions for optimal solutions to both discrete approximation problems (P_N) and $(\overline{P}_N)$. The differences between the above ones for problem (DP) in Theorem 6.17 and those for problem $(\overline{P}_N)$ are just in terms converging to zero as $N \to \infty$. The Euler-Lagrange inclusion for problem (P_N) is parallel to the one in $(\overline{P}_N)$ with replacing

$$\lambda_N \partial \vartheta \Big(\bar{x}_N(t_j), \frac{\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)}{h_N}, t_j \Big)$$

by the norm-closure of

$$\frac{\lambda_N}{h_N} \int_{t_j}^{t_{j+1}} \partial \vartheta \Big(\bar{x}_N(t_j), \frac{\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)}{h_N}, t \Big) \, dt$$

on the right-hand side, which comes from the integration formula of Lemma 6.18. The latter terms converges to $\lambda \partial \vartheta (\bar{x}(t), \dot{\bar{x}}(t), t)$ as $N \to \infty$ for a.e. $t \in [a, b]$; see the proof of Theorem 6.21 in the next subsection.

The results obtained by this approach employing the exact/limiting optimality conditions in the general mathematical programming problems from Theorem 6.16 require the *SNC assumptions* on the sets gph F_j and Ω_N in problems (P_N) and $(\overline{P}_N)$. These assumptions may be restrictive for the limiting procedure to derive necessary optimality conditions in the original continuous-time problem (P); so we'll try to avoid or essentially relax them in what follows. This can be done by starting with *approximate/fuzzy* necessary optimality conditions for problems of mathematical programming that strongly take into account specific features of the discrete-time problems (P_N) and $(\overline{P}_N)$. It happens that to realize this approach, we need to impose the *Lipschitz-like* property of the set-valued mappings F_j generated the graphical geometric constraints in problem (DP), and hence in (P_N) and $(\overline{P}_N)$, which is not assumed in Theorem 6.17. On the other hand, the Lipschitz continuity of the original mapping $F(\cdot, t)$ in (6.1) is among our standing assumptions (see (H1) in Subsect. 6.1.1), and thus we don't have any reservations to employ it in the context of necessary optimality conditions for discrete approximations.

The next two theorems give *approximate* necessary optimality conditions for local minimizers in *sequences* of discrete-time problem $(\overline{P}_N)$ and (P_N). Their proofs involve the use of some *fuzzy/neighborhood* calculus results from the prior chapters. In particular, we employ the semi-Lipschitzian sum rule for Fréchet subgradients from Theorem 2.33 and the fuzzy intersection rule for Fréchet normals from Lemma 3.1. These results provide representations of Fréchet subgradients and normals of sums and intersections at the reference points via those at points that are arbitrarily close to the reference ones. *Just for notational simplicity* we suppose in the formulation and proof of the following theorem that these *arbitrarily close points reduce to the reference points in question.* This agreement doesn't actually restrict the generality from the viewpoint of our main goal in this section to derive necessary optimality conditions in the continuous-time problem (P), which is finalized in the next subsection. Indeed, the possible difference between the mentioned points obviously disappears in the limiting procedure. The interested reader may readily proceed with all the details.

Let us start with approximate necessary optimality conditions for the *simplified* discrete approximation problems $(\overline{P}_N)$ as $N \to \infty$ described in Remark 6.15, which are efficient under the a.e. continuity assumption on the integrand $\vartheta(x, v, \cdot)$ in the original problem (P). In what follows $I\!B^*$ stands as usual for the dual closed unit ball *regardless* of the space in question, and subdifferential of ϑ is taken with respect to the first two variables.

Theorem 6.19 (approximate Euler-Lagrange conditions for simplified discrete-time problems). *Let $\bar{x}_N(\cdot) = \{\bar{x}_N(t_j)|\ j = 0, \ldots, N\}$ be local optimal solutions to problems $(\overline{P}_N)$ as $N \to \infty$. Assume that X is Asplund, that Ω_N is locally closed around $\big(\bar{x}_N(t_0), \bar{x}_N(t_N)\big)$, that F_j is closed-graph and Lipschitz-like around $\big(\bar{x}_N(t_j), [\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)]/h_N\big)$, and that the functions φ and $\vartheta(\cdot, \cdot, t_j)$ are locally Lipschitzian around $\bar{x}_N(\cdot)$ for every $j = 0, \ldots, N-1$. Consider the quantities*

$$\theta_{Nj} := 2 \int_{t_j}^{t_{j+1}} \left\| \frac{\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)}{h_N} - \dot{\bar{x}}(t) \right\| dt, \quad j = 0, \ldots, N-1 \,.$$

Then there exists a number $\gamma > 0$ independent of N and such that for some sequences of natural numbers $N \to \infty$ and positive numbers $\varepsilon_N \downarrow 0$ there are multipliers $\lambda_N \geq 0$ and adjoint trajectories $p_N(\cdot) = \{p_N(t_j) \in X^*|\ j = 0, \ldots, N\}$ satisfying the nontriviality condition

$$\lambda_N + \|p_N(t_N)\| \geq \gamma \quad \text{as} \quad N \to \infty \,, \tag{6.40}$$

the approximate Euler-Lagrange inclusion

$$\begin{aligned}
&\Big(\frac{p_N(t_{j+1}) - p_N(t_j)}{h_N}, p_N(t_{j+1}) - \lambda_N \frac{\theta_{Nj}}{h_N} b^*_{N_j}\Big) \\
&\in \lambda_N \widehat{\partial}\vartheta\Big(\bar{x}_N(t_j), \frac{\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)}{h_N}, t_j\Big) \\
&+ \widehat{N}\Big(\Big(\bar{x}_N(t_j), \frac{\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)}{h_N}\Big); \text{gph}\, F_j\Big) + \varepsilon_N I\!B^*
\end{aligned} \tag{6.41}$$

for $j = 0, \ldots, N-1$, and the approximate transversality inclusion

$$\begin{aligned}
&\Big(p_N(t_0) - 2\lambda_N b^*_N \|\bar{x}(a) - \bar{x}_N(t_0)\|, -p_N(t_N)\Big) \\
&\in \lambda_N \widehat{\partial}\varphi\big(\bar{x}_N(t_0), \bar{x}_N(t_N)\big) + \widehat{N}\big((\bar{x}_N(t_0), \bar{x}_N(t_N)); \Omega_N\big) + \varepsilon_N I\!B^*
\end{aligned} \tag{6.42}$$

with some $b^*_N, b^*_{Nj} \in I\!B^*$.

Proof. Fixed $N \in I\!N$, consider problem $(\overline{P}_N)$ in the equivalent (MP) form (6.29) with the data defined in (6.34)–(6.37). Denote

$$\bar{z} := \big(\bar{x}_N(t_0), \ldots, \bar{x}_N(t_N), \bar{v}_N(t_0), \ldots, \bar{v}_N(t_{N-1})\big)$$

and take N so large that constraints (6.22) and (6.23) for $\bar{x}_N(\cdot)$ hold with the strict inequality. The latter can be clearly done by the strong convergence result of Theorem 6.13.

Suppose first that f in (6.36) is *metrically regular* at $\bar{z}$ relative to the intersection $\Xi := \Xi_0 \cap \ldots \cap \Xi_N$, where the sets Ξ_j are constructed in (6.37). Since φ_0 in (6.34) is locally Lipschitzian around $\bar{z}$ and by the choice of N, we employ Theorem 5.16 and find $\mu > 0$ such that $\bar{z}$ is a local optimal solution to the unconstrained problem:

$$\text{minimize} \;\; \varphi_0(z) + \mu \left(\|f(z)\| + \operatorname{dist}(z; \Xi) \right) .$$

Therefore, by the generalized Fermat rule, one has

$$0 \in \widehat{\partial}\Big(\varphi_0(\cdot) + \mu \, \|f(\cdot)\| + \mu \operatorname{dist}(\cdot; \Xi)\Big)(\bar{z}) .$$

Now using the fuzzy sum rule from Theorem 2.33 and remembering our notational agreement, we fix any $\varepsilon > 0$ and get

$$0 \in \widehat{\partial}\varphi_0(\bar{z}) + \mu\widehat{\partial}\|f(\cdot)\|(\bar{z}) + \mu \, \widehat{\partial}\operatorname{dist}(\bar{z}; \Xi) + (\varepsilon/3)\mathbb{B}^* .$$

By Proposition 1.95 on Fréchet subgradients of the distance function and by the elementary chain rule for the composition $\|f(z)\| = (\psi \circ f)(z)$ with $\psi(y) := \|y\|$ and the smooth mapping f from (6.36) one has

$$0 \in \widehat{\partial}\varphi_0(\bar{z}) + \sum_{j=0}^{N-1} \nabla f_j(\bar{z})^* e_j^* + \widehat{N}(\bar{z}; \Xi) + (\varepsilon/3)\mathbb{B}^*$$

with some $e_j^* \in X^*$. Observe that

$$\sum_{j=0}^{N-1} \nabla f_j(\bar{z})^* e_j^* = \big(- e_0^*, e_0^* - e_1^*, \ldots, e_{N-2}^* - e_{N-1}^*, e_{N-1}^*, -h_N e_0^*, \ldots, -h_N e_N^* \big)$$

by the structure of $f(z)$ in (6.36). Further, it follows from the fuzzy intersection rule in Lemma 3.1 and the discussion right after it that, taking into account the notational agreement, we get

$$\widehat{N}(\bar{z}; \Xi) \subset \widehat{N}(\bar{z}; \Xi_0) + \ldots + \widehat{N}(\bar{z}; \Xi_N) + (\varepsilon/3)\mathbb{B}^* .$$

To justify it, one needs to check the *fuzzy qualification condition* (3.9) for the sets involved. It obviously holds for the set intersections of Ξ_j, with $j = 0, \ldots, N-1$ by the structure of these sets in (6.37). To verify this condition at the last step, let us show that there is $\gamma > 0$ for which

$$\Big(\widehat{N}\Big(z; \bigcap_{j=0}^{N-1} \Xi_j\Big) + \gamma \, \mathbb{B}^*\Big) \bigcap \Big(- \widehat{N}\big(z_N; \Xi_N\big) + \gamma \, \mathbb{B}^*\Big) \bigcap \mathbb{B}^* \subset \frac{1}{2}\mathbb{B}^*$$

whenever $z \in \Xi_j \cap (\bar{z} + \gamma \, \mathbb{B})$, $j = 0, \ldots, N-1$, and $z_N \in \Xi_N \cap (\bar{z} + \gamma \, \mathbb{B})$. It follows directly from the set structures in (6.37) that for any $z_j^* \in \widehat{N}(z_j; \Xi_j)$ with $z_j^* =$

$(x^*_{0j},\ldots,x^*_{Nj},v^*_{0j},\ldots,v^*_{N-1j})$ and $z_j=(x_{0j},\ldots,x_{Nj},v_{0j},\ldots,v_{N-1j})$ close to $\bar z$ one has the relations

$$x^*_{ij}\in\widehat D^*F_j(x_{jj},v_{jj})(-v^*_{jj}),\quad x^*_{ij}=v^*_{ij}=0\ \text{ if }\ i\neq j,\quad j=0,\ldots,N-1;$$

$$(x^*_{0N},x^*_{NN})\in\widehat N\big((x_{0N},x_{NN});\Omega_N\big)\ \text{ with }\ x^*_{iN}=v^*_{iN}=0\ \text{ otherwise}\,.$$

Therefore, by Theorem 1.43 on Fréchet coderivatives of Lipschitzian mappings, we get the estimates

$$\|x^*_{jj}\|\le\ell\|v^*_{jj}\|\ \text{ for all }\ j=0,\ldots,N-1$$

provided that F_j are Lipschitz-like around (x_{jj},v_{jj}) with modulus ℓ. This easily implies the above fuzzy qualification condition at the last step by taking into account that it holds at all the previous steps with $\varepsilon_N:=\varepsilon/N$.

Next we proceed with estimating Fréchet subgradients of the cost function φ_0 in (6.34). It is well known from convex analysis that

$$\partial\|\cdot\|^2(x)\subset2\|x\|I\!B^*\ \text{ for any }\ x\in X$$

in arbitrary Banach spaces. Using this and applying the fuzzy sum rule from Theorem 2.33 to the specific form of φ_0 in (6.34), we have

$$\widehat\partial\varphi_0(\bar z)\subset\widehat\partial\varphi\big(\bar x_N(t_0),\bar x_N(t_N)\big)+2\|\bar x_N(t_0)-\bar x(a)\|I\!B^*$$
$$+h_N\sum_{j=0}^{N-1}\Big[\widehat\partial\vartheta_j\big(\bar x_N(t_j),\bar v_N(t_j)\big)+\big(0,2\theta_{Nj}I\!B^*\big)\Big]+(\varepsilon/3)I\!B^*$$

with taking into account our notational agreement and the construction of θ_{Nj}. Now combining the above relationships and estimates in generalized Fermat rule, one gets

$$\begin{cases}-x^*_{00}-x^*_{0N}-x^*_0-2b^*_N\big\|\bar x_N(t_0)-\bar x(a)\big\|-u^*_0+e^*_0\in\varepsilon I\!B^*\,,\\[1ex] -x^*_{jj}-h_Nu^*_j-e^*_{j-1}+e^*_j\in\varepsilon I\!B^*,\quad j=0,\ldots,N-1\,,\\[1ex] -x^*_{NN}-x^*_N-e^*_{N-1}\in\varepsilon I\!B^*\,,\\[1ex] -v^*_{jj}-h_Nw^*_j-\theta_{Nj}b^*_{Nj}+h_Ne^*_j\in\varepsilon I\!B^*,\quad j=0,\ldots,N-1\end{cases}$$

with some $b^*_{Nj},b^*\in I\!B^*$,

$$(x^*_{ij},v^*_{ij})\in\widehat N\Big(\Big(\bar x_N(t_j),\frac{\bar x_N(t_{j+1})-\bar x_N(t_j)}{h_N}\Big);\text{gph}\,F_j\Big),\quad\text{and}$$

$$(x^*_0,x^*_N)\in\widehat\partial\varphi\big(\bar x_N(t_0),\bar x_N(t_N)\big),\quad(u^*_j,w^*_j)\in\widehat\partial\vartheta_j\Big(\bar x_N(t_j),\frac{\bar x_N(t_{j+1})-\bar x_N(t_j)}{h_N}\Big)$$

for $j = 0, \ldots, N-1$. Denoting

$$p_N(t_0) := -x_{0N}^* - \lambda_N x_0^* + e_0^* \quad \text{and} \quad p_N(t_j) := e_{j-1}^*, \quad j = 1, \ldots, N \; ,$$

we arrive at the approximate Euler-Lagrange and transversality inclusions (6.41) and (6.42) with $\lambda_N = 1$ for *any* $N \in I\!N$ sufficiently large and *any* $\varepsilon = \varepsilon_N$. Note that the nontriviality condition (6.40) is obviously fulfilled with $\gamma_N = 1$ in the metric regularity case under consideration.

It remains to consider the case when the mapping f from (6.36) is *not metrically regular* at $\bar{z}$ relative to the set intersection $\Xi := \Xi_0 \cap \ldots \cap \Xi_N$. In this case the extended mapping $f_\Xi(z) := -f(z) + \Delta(z; \Xi)$ is not metrically regular around $\bar{z}$ in the sense of Definition 1.47(ii). We now apply the *neighborhood characterization* of metric regularity in Asplund spaces obtained in Theorem 4.5. It is not hard to observe that this criterion can be equivalently written as follows: a closed-graph mapping $F\colon X \rightrightarrows Y$ between Asplund spaces is metrically regular around $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ *if and only if* there is a positive number ν such that

$$\ker \widehat{D}^* F(x, y) \subset I\!B^* \quad \text{whenever} \quad x \in \bar{x} + \nu I\!B, \quad y \in F(x) \cap (\bar{y} + \nu I\!B) \; .$$

Applying this result to the mapping $-f(z) + \Delta(z; \Xi)$ that is *not* metrically regular around $\bar{z}$, we have the following assertion as N is fixed: for any $\eta > 0$ there are $z \in \bar{z} + \eta I\!B$ and $e^* \in \ker \widehat{D}^* f_\Xi(z)$ with $e^* = (e_0^*, \ldots, e_{N-1}^*) \in (X^*)^N$ satisfying $\|e^*\| > 1$. Thus

$$0 \in \widehat{D}^* f_\Xi(z)(e^*) \quad \text{for some} \quad \|e^*\| > 1 \quad \text{and} \quad z \in \bar{z} + \nu I\!B \; .$$

Fixed $\varepsilon > 0$, we employ the coderivative sum rule from Theorem 1.62(i) and then the above intersection rule for Fréchet normals that give

$$0 \in \sum_{j=0}^{N-1} \nabla f_j(z)^* e_j^* + \sum_{j=0}^{N} \widehat{N}(z_j; \Xi_j) + \varepsilon I\!B^*$$

with some $z_j \in \Xi_j \cap (z + \varepsilon I\!B)$. According to our notation agreement we may put $z_j = z = \bar{z}$ for simplicity. Thus there are $z_j^* \in \widehat{N}(\bar{z}; \Xi_j)$ satisfying

$$-\sum_{j=0}^{N} z_j \in \sum_{j=0}^{N-1} \nabla f_j(z)^* e_j^* + \varepsilon I\!B^* \; .$$

Taking into account the structures of the mapping f in (6.36) and the sets Ξ_j in (6.37), we find as above dual elements

$$(x_{ij}^*, v_{ij}^*) \in \widehat{N}\Big(\Big(\bar{x}_N(t_j), \frac{\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)}{h_N}\Big); \operatorname{gph} F_j\Big)$$

for $j = 0, \ldots, N-1$ and

$$(x_{0N}^*, x_{NN}^*) \in \widehat{N}\big((\bar{x}_N(t_0), \bar{x}_N(t_N)); \Omega_N\big)$$

satisfying the relations

$$\begin{cases} -x_{00}^* - x_{0N}^* + e_0^* \in \varepsilon I\!B^*\ , \\ -x_{jj}^* - e_{j-1}^* + e_j^* \in \varepsilon I\!B^*, \quad j = 0, \ldots, N-1\ , \\ -x_{NN}^* - x_N^* - e_{N-1}^* \in \varepsilon I\!B^*\ , \\ -v_{jj}^* + h_N e_j^* \in \varepsilon I\!B^*, \quad j = 0, \ldots, N-1\ . \end{cases}$$

Define the adjoint discrete trajectory $p_N(t_j)$, $j = 0, \ldots, N$, by

$$p_N(t_0) := -x_{0N}^* + e_0^* \ \text{ and } \ p_N(t_j) := e_{j-1}^*, \quad j = 1, \ldots, N\ .$$

It follows from the above constructions that the pair $\big(\bar{x}_N(\cdot), p_N(\cdot)\big)$ satisfies the Euler-Lagrange inclusion (6.41) and the transversality inclusion (6.42) with $\lambda_N = 0$ and arbitrary $\varepsilon_N = \varepsilon > 0$. Moreover, the adjoint trajectory $p_N(\cdot)$ obeys the following nontriviality condition:

$$\|p_N(t_1)\| + \ldots + \|p_N(t_N)\| \ge 1 \ \text{ for all large } \ N \in I\!N\ .$$

Let us finally prove that, by the *Lipschitz-like* assumption on F_j, the nontriviality condition in this case can be equivalently written as $\|p_N(t_N)\| \ge 1$, which agrees with (6.40) as $\lambda_N = 0$. The approximate Euler-Lagrange inclusion (6.41) can be now rewritten in the form

$$\frac{p_N(t_{j+1}) - p_N(t_j)}{h_N} \in \widehat{D}^* F_j\Big(\bar{x}_N(t_j), \frac{\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)}{h_N}\Big)\big(-p_N(t_{j+1}) + \varepsilon I\!B^*\big)$$

$$+\varepsilon I\!B^* \quad \text{for} \quad j = 0, \ldots, N-1\ .$$

Then the Lipschitz-like property of F_j assumed in the theorem with modulus $\ell = \ell_F$ yields by Theorem 1.43 that

$$\|x_j^*\| \le \ell \|v_j^*\| \ \text{ whenever } \ x_j^* \in \widehat{D}^* F_j(x_j, v_j)(v_j^*)$$

and (x_j, v_j) around $\big(\bar{x}_N(t_j), [\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)]/h_N\big)$. Thus

$$\|p_N(t_{N-1})\| \le \|p_N(t_N)\|\big(1 + h_N \ell\big) + h_N \varepsilon(\ell + 1)\ .$$

Continuing this process, one has

$$\|p_N(t_j)\| \le \exp\big(\ell(b-a)\big)\|p_N(t_N)\| + \varepsilon(b-a)(1+\ell) \ \text{ for all } \ j = 0, \ldots, N\ .$$

Suppose that the nontriviality condition (6.40) doesn't hold along with (6.41) and (6.42) in the case of $\lambda_N = 0$ under consideration. Take a sequence $\gamma_k \downarrow 0$ as $k \to \infty$ and choose numbers N_k and ε_k such that

$$N_k := \big[1/\gamma_k\big], \;\; \varepsilon_k \le \gamma_k^2, \text{ and } \|p_N(t_N)\| \le \gamma_k^2, \quad k \in I\!N \,,$$

where $[\cdot]$ stands for the greatest integer less than or equal to the given real number. By the adjoint trajectory estimate we have

$$\sum_{j=1}^{N_k} \|p_{N_k}(t_j)\| \le N_k\gamma_k \exp\big(\ell(b-a)\big) + \varepsilon_k N_k(b-a)(1+\ell)$$
$$\le \gamma_k \exp\big(\ell(b-a)\big) + \gamma_k(b-a)(1+\ell) \downarrow 0 \;\text{ as }\; k \in I\!N \,,$$

which contradicts the fact established above. This therefore completes the proof of the theorem. △

Finally in this subsection, we obtain *approximate* necessary optimality conditions for the sequence of discrete-time problems (P_N) defined in (6.3), (6.20)–(6.23). The difference between these problems and the simplified problems $(\overline{P}_N)$ is that (P_N) deal with approximating *summable* integrands $\vartheta(x, \upsilon, \cdot)$ in the original problem (P), which is reflected by the integral term involving ϑ in the cost function (6.20). The latter term makes the analysis of problems (P_N) to be more complicated in comparison with the one for $(\overline{P}_N)$. To proceed, we need to use Lemma 6.18 on the subdifferentiation under the (Bochner) integral sign, which requires additional assumptions on the space X. The next theorem incorporates these developments in the framework of the extended Euler-Lagrange inclusion for (P_N). We keep our notational agreement discussed before the formulation of Theorem 6.19.

Theorem 6.20 (approximate Euler-Lagrange conditions for discrete problems involving summable integrands). *Let $\bar{x}_N(\cdot) = \{\bar{x}_N(t_j)|\; j = 0, \ldots, N\}$ be local optimal solutions to problems (P_N) as $N \to \infty$. Assume that X is reflexive and separable, that φ, F_j, Ω_N, and θ_{N_j} are the same as in Theorem* 6.19*, and that ϑ satisfies assumption* (H3) *of Subsect.* 6.1.3 *with the replacement of continuity by Lipschitz continuity. Then there exists a number $\gamma > 0$ independent of N and such that for some sequences of natural numbers $N \to \infty$ and positive numbers $\varepsilon_N \downarrow 0$ there are multipliers $\lambda_N \ge 0$ and adjoint trajectories $p_N(\cdot) = \{p_N(t_j) \in X^*|\; j = 0, \ldots, N\}$ satisfying the nontriviality condition* (6.40)*, the approximate transversality inclusion* (6.42)*, and the Euler-Lagrange inclusion in the modified form*

$$\Big(\frac{p_N(t_{j+1}) - p_N(t_j)}{h_N}, p_N(t_{j+1}) - \lambda_N \frac{\theta_{Nj}}{h_N} b_{Nj}^*\Big)$$
$$\in \frac{\lambda_N}{h_N} \,\mathrm{cl} \int_{t_j}^{t_{j+1}} \partial\vartheta\Big(\bar{x}_N(t_j), \frac{\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)}{h_N}, t\Big)\, dt \tag{6.43}$$
$$+\widehat{N}\Big(\Big(\bar{x}_N(t_j), \frac{\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)}{h_N}\Big); \mathrm{gph}\, F_j\Big) + \varepsilon_N I\!B^*$$

for all $j = 0, \ldots, N-1$ with some $b_{Nj}^* \in I\!B^*$.

Proof. Each problem (P_N) can be equivalently written in the (MP) form (6.29) with the data defined in (6.35)–(6.38). Now we proceed similarly to the proof of Theorem 6.19 using additionally Lemma 6.18 to calculate subgradients of integral function. This becomes possible under the additional assumptions on X made in the theorem and gives the modified form (6.43) of the approximate Euler-Lagrange inclusion. $\triangle$

Taking into account the value convergence results of Theorem 6.14, we can treat the necessary optimality conditions obtained in this subsection for the discrete approximation problems under consideration as *suboptimality conditions* for the original problem (P). Moreover, the strong convergence results presented in Theorem 6.13 and Remark 6.15 allow us to view the above necessary optimality conditions for the discrete-time problems as suboptimality conditions concerning a *given* relaxed intermediate local minimizer for the original problem. Note that the assumptions made in Theorems 6.13 and 6.14 ensure the *existence* of optimal solutions to the discrete approximations, while it is *not* the case for the original continuous-time problem (P) in either finite-dimensional or infinite-dimensional setting. Necessary *optimality* conditions for relaxed local minimizers to problem (P) are considered next.

6.1.5 Euler-Lagrange Conditions for Relaxed Minimizers

The aim of this subsection is to derive necessary conditions for the underlying r.i.l.m. to the original Bolza problem (P) involving constrained differential inclusions by *passing to the limit* from the ones for discrete approximations obtained in the preceding subsection. This is based on the strong convergence result for discrete approximations given in Theorem 6.13, on the approximate necessary optimality conditions for the discrete problems (P_N) and $(\overline{P}_N)$ from Theorems 6.19 and 6.20, and on stability properties of the generalized differential constructions. The major ingredient involved in this limiting procedure is the possibility to establish an appropriate *convergence of adjoint trajectories*, which allows us to pass to the limit in the approximate Euler-Lagrange inclusions. This is done below by employing the *coderivative characterization of Lipschitzian stability* used also in the preceding subsection.

Let us first clarify the assumptions needed for the main results of this subsection. They involve of course those ensuring the strong convergence of discrete approximations and the fulfillment of the (approximate) necessary optimality conditions in discrete-time problems (P_N) and $(\overline{P}_N)$ used below. In fact, not too much has to be added for furnishing the limiting process to derive pointwise necessary optimality conditions in the original Bolza problem (P) via discrete approximations.

In what follows we keep assumptions (H1) and (H2) from Subsect. 6.1.1 on the mapping F in (6.1) and consider the Lipschitzian modification of assumptions (H3) and (H4) from Subsect. 6.1.3:

(H3′) $\vartheta(\cdot,\cdot,t)$ is Lipschitz continuous on $U \times (m_F I\!B)$ uniformly in $t \in [a,b]$, while $\vartheta(x,v,\cdot)$ is measurable on $[a,b]$ and its norm is majorized by a summable function uniformly in $(x,v) \in U \times (m_F I\!B)$.

(H4′) φ is Lipschitz continuous on $U \times U$; $\Omega \subset X \times X$ is locally closed around $(\bar{x}(a),\bar{x}(b))$ and such that the set $\text{proj}_1 \Omega \cap \big(\bar{x}(a) + \varepsilon I\!B\big)$ is compact for some $\varepsilon > 0$.

Note that (H3′) contains the *measurability* assumption on $\vartheta(x,v,\cdot)$, which corresponds to Theorem 6.20. The latter imposes more restrictive requirement on the state space X in comparison with Theorem 6.19, which however relates to the *a.e. continuity* of $\vartheta(x,v,\cdot)$ in the convergence result for problem $(\overline{P}_N)$; see Remark 6.15. Taking this into account, we consider also another modification of (H3) that is an alternative to the above assumption (H3′):

(H3″) $\vartheta(x,v,\cdot)$ is a.e. continuous on $[a,b]$ and bounded on this interval uniformly in $(x,v) \in U \times (m_F I\!B)$, while $\vartheta(\cdot,\cdot,t)$ is Lipschitz continuous on

$$\Theta_\nu(t) := \Big\{(x,v) \in U \times (m_F + \nu)I\!B \Big|\, \exists \tau \in (t-\nu,t] \;\text{ with }\; v \in F(x,\tau)\Big\}$$

uniformly in $t \in [a,b]$ for some $\nu > 0$.

Dealing with the a.e. continuous mappings $F(x,\cdot)$ and $\vartheta(x,v,\cdot)$ in the limiting procedures involving t, we use *extended* normal cone N_+ from Definition 5.69 to the *moving* sets $\text{gph}\, F(\cdot)$ and the corresponding subdifferential of $\vartheta(x,v,t)$. Although these constructions may be different from the basic normal cone and subdifferential in the case of *non-autonomous* objects, they agree with the latter in general settings ensuring *normal semicontinuity*; see the results and discussions after Definition 5.69. Note that we *don't need* to replace the basic subdifferential of the integrand ϑ by the extended one assuming the *measurability* of ϑ in t as in (H3′). We also don't need to replace the basic normal cone to $\text{gph}\, F$ in the next Subsect. 6.1.6 dealing with measurable set-valued mappings in differential inclusions.

Recall that, given $(\bar{x},\bar{v},\bar{t})$ with $\bar{v} \in F(\bar{x},\bar{t})$, the *extended normal cone* to the moving set $\text{gph}\, F(t)$ at $(\bar{x},\bar{v}) \in \text{gph}\, F(\bar{t})$ is, in the case of closed subsets in Asplund spaces,

$$N_+\big((\bar{x},\bar{v});\text{gph}\, F(\bar{t})\big) := \mathop{\text{Lim sup}}_{(x,v,t)\to(\bar{x},\bar{v},\bar{t})} \widehat{N}\big((x,v);\text{gph}\, F(t)\big)\ .$$

Correspondingly, the *extended subdifferential* of $\vartheta(\cdot,\cdot,\bar{t})$ at $(\bar{x},\bar{v})$ is

$$\partial_+\vartheta(\bar{x},\bar{v},\bar{t}) := \mathop{\text{Lim sup}}_{(x,v,t)\to(\bar{x},\bar{v},\bar{t})} \widehat{\partial}\vartheta(x,v,t)\ ,$$

where $\widehat{\partial}\vartheta(\cdot,\cdot,t)$ is taken with respect to (x,v) under fixed t. Note that $\partial_+\vartheta(\bar{x},\bar{v},\bar{t})$ can be equivalently described via the extended normal cone N_+ to the moving epigraphical set $\operatorname{epi}\vartheta(t)$. One can see that these extended objects reduce to the basic ones $N(\cdot;\operatorname{gph}F)$ and $\partial\vartheta$ when F and ϑ are independent of t, as well as in the more general settings discussed above.

Now we are ready to formulate and prove the *extended Euler-Lagrange conditions* for relaxed intermediate minimizers in the original Bolza problem (P). We consider separately the two cases: when the integrand ϑ is a.e. continuous in t, and when it is summable. Although the second case imposes less requirements on the integrand and gives a better form of the Euler-Lagrange inclusion, in the first case we are able to obtain necessary optimality conditions in more general Banach spaces. Let us start with the first one. The *strong PSNC* property used below is defined and discussed in Subsect. 3.1.1.

Theorem 6.21 (extended Euler-Lagrange conditions for relaxed local minimizers in Bolza problems with a.e. continuous integrands). *Let $\bar{x}(\cdot)$ be a relaxed intermediate local minimizer for the Bolza problem (P) under assumptions* (H1), (H2), (H4′), *and* (H3″). *Suppose also that both spaces X and X^* are Asplund and that the set Ω is strongly PSNC at $(\bar{x}(a),\bar{x}(b))$ with respect to the second component. Then there are $\lambda\ge 0$ and an absolutely continuous mapping $p\colon[a,b]\to X^*$, not both zero, satisfying the extended Euler-Lagrange inclusion*

$$\dot{p}(t)\in\operatorname{clco}\Big\{u\in X^*\Big|\ \big(u,p(t)\big)\in\lambda\partial_+\vartheta\big(\bar{x}(t),\dot{\bar{x}}(t),t\big)+N_+\big((\bar{x}(t),\dot{\bar{x}}(t));\operatorname{gph}F(t)\big)\Big\}\tag{6.44}$$

for a.e. $t\in[a,b]$ and the transversality inclusion

$$\big(p(a),-p(b)\big)\in\lambda\partial\varphi\big(\bar{x}(a),\bar{x}(b)\big)+N\big((\bar{x}(a),\bar{x}(b));\Omega\big)\,.\tag{6.45}$$

Proof. We derive these conditions by passing to the limit in the necessary optimality conditions for discrete-time problems $(\overline{P}_N)$ from Theorem 6.19 with taking into account the strong convergence of the simplified discrete approximations; see Theorem 6.13 and Remark 6.15. Recall that the Asplund property of X is *equivalent* to the Radon-Nicodým property of X^*; see Subsect. 6.1.1. Since X is a closed subspace of X^{**} and X^* is assumed to be Asplund, this yields that X has the Radon-Nicodým property. Thus all the assumptions of Theorem 6.13 are fulfilled, which allows us to employ the strong convergence of discrete approximations.

Note that the assumptions made clearly ensure the fulfillment of the ones in Theorem 6.19. Employing the necessary optimality conditions for $(\overline{P}_N)$ obtained therein, we find (sub)sequences of numbers $\lambda_N\ge 0$ and discrete adjoint trajectories $p_N(\cdot)=\{p_N(t_j)|\ j=0,\ldots,N\}$ satisfying inclusions (6.40)–(6.42) with some $\varepsilon_N\downarrow 0$ as $N\to\infty$. Observe that without loss of generality the nontriviality condition (6.40) can be equivalently written as

$$\lambda_N + \|p_N(t_N)\| = 1 \text{ for all } N \in I\!N ,$$

because the number $\gamma > 0$ is independent of N. Also one can always suppose that $\lambda_N \to \lambda \geq 0$ as $N \to \infty$.

In what follows we use the notation $\bar{x}_N(t)$ and $p_N(t)$ for piecewise linear extensions of the corresponding discrete trajectories to $[a, b]$ with their piecewise constant derivatives $\dot{\bar{x}}_N(t)$ and $\dot{p}_N(t)$. Having θ_{Nj} defined in Theorem 6.19, we consider a sequence of functions $\theta_N\colon [a, b] \to I\!R$ given by

$$\theta_N(t) := \frac{\theta_{Nj}}{h_N} b^*_{Nj} \text{ for } t \in [t_j, t_{j+1}), \quad j = 0, \ldots, N-1 .$$

Invoking Theorem 6.13, we get

$$\begin{aligned}\int_a^b \|\theta_N(t)\|\, dt &\leq \sum_{j=0}^{N-1} \theta_{Nj} \leq 2 \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \left\| \frac{\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)}{h_N} - \dot{\bar{x}}_N(t) \right\| dt \\ &= 2 \int_a^b \|\dot{\bar{x}}_N(t) - \dot{\bar{x}}(t)\|\, dt =: \nu_N \to 0 \text{ as } N \to \infty .\end{aligned}$$

This allows us to suppose without loss of generality that

$$\dot{\bar{x}}_N(t) \to \dot{\bar{x}}(t) \text{ and } \theta_N(t) \to 0 \text{ a.e. } t \in [a, b] \text{ as } N \to \infty .$$

Consider the approximate discrete Euler-Lagrange inclusions (6.41) along the designated sequence of $N \to \infty$, which is identified with the whole set of natural numbers $I\!N$. By (6.41) we find

$$(x^*_{Nj}, v^*_{Nj}) \in \widehat{\partial}\vartheta_j\Big(\bar{x}_N(t_j), \frac{\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)}{h_N}\Big), \quad j = 0, \ldots, N-1 ,$$

and $e^*_{N_j}, \widetilde{e}^*_{Nj} \in I\!B^*$ such that the inclusions

$$\begin{aligned}&\Big(\frac{p_N(t_{j+1}) - p_N(t_j)}{h_N} - \lambda_N x^*_{Nj}\Big) + \varepsilon_N e^*_{N_j} \\ &\in \widehat{D}^* F_j\Big(\bar{x}_N(t_j), \frac{\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)}{h_N}\Big)\Big(\lambda_N v^*_{Nj} + \lambda_N \frac{\theta_{Nj}}{h_N} b^*_{Nj} - p_N(t_{j+1}) + \varepsilon_N \widetilde{e}^*_{N_j}\Big)\end{aligned}$$

hold for all $j = 0, \ldots, N-1$ and all $N \in I\!N$. It follows from the local Lipschitz continuity of ϑ assumed in (H3$'$) and from Proposition 1.85 that

$$\|(x^*_{Nj}, y^*_{Nj})\| \leq \ell_\vartheta \text{ for all } j = 0, \ldots, N-1 \text{ and } N \in I\!N ,$$

where ℓ_ϑ is a uniform Lipschitz modulus of $\vartheta(\cdot, \cdot, t)$ independent of $t \in [a, b]$. By the Lipschitz continuity of F in (H1) and the coderivative condition of Theorem 1.43 we get the estimates

$$\left\| \frac{p_N(t_{j+1}) - p_N(t_j)}{h_N} - \lambda_N x^*_{Nj} + \varepsilon_N e^*_{N_j} \right\|$$

$$\le \ell_F \left\| \lambda_N v^*_{Nj} + \lambda_N \frac{\theta_{Nj}}{h_N} b^*_{Nj} - p_N(t_{j+1}) + \varepsilon_N \widetilde{e}^*_{N_j} \right\|$$

for $j = 0, \ldots, N-1$. Similarly to the proof of Theorem 6.19 with taking $\|p_N(t_N)\| \le 1$ into account, we derive from these estimates that $p_N(t)$ is uniformly bounded on $[a,b]$ and that

$$\|\dot{p}_N(t)\| \le \alpha + \beta \|\theta_N(t)\| \quad \text{a.e. } t \in [a,b]$$

with some positive numbers α and β independent of N. Since both spaces X and X^* have the RNP, it follows from the Dunford theorem on the weak compactness in $L^1\big([a,b];X^*\big)$ that a subsequence of $\{\dot{p}_N(\cdot)\}$ converges to some $v(\cdot) \in L^1\big([a,b];X^*\big)$ weakly in this space. Employing the weak continuity of the Bochner integral as a linear operator from $L^1\big([a,b];X^*\big)$ to X^* and the estimate $\|p_N(b)\| \le 1$, we conclude that there is an absolutely continuous mapping $p\colon [a,b] \to X^*$ satisfying

$$p(t) := p(b) + \int_t^b v(s)\,ds, \quad a \le t \le b\,,$$

where $p(b)$ is a limiting point of $\{p_N(b)\}$ in the weak* topology of X^*, and such that the values $p_N(t)$ converge to $p(t)$ weakly in X^* (and hence weak* in this space) for all $t \in [a,b]$. Furthermore, $\dot{p}_N(\cdot) \to \dot{p}(\cdot) = v(t)$ in the weak topology of $L^1\big([a,b];X^*\big)$. Then the classical Mazur theorem ensures that some sequence of *convex combinations* of $\{\dot{p}_N(\cdot)\}$ converges to $\dot{p}(\cdot)$ strongly in $L^1\big([a,b];X^*\big)$ as $N \to \infty$, and hence (passing to a subsequence with no relabeling) it converges almost everywhere on $[a,b]$.

Given any $N \in I\!N$, the approximate Euler-Lagrange inclusion (6.41) can be rewritten as

$$\dot{p}_N(t) \in \Big\{ u \in X^* \Big| \; \big(u, p_N(t_{j+1}) - \lambda_N \theta_N(t)\big) \in \lambda_N \widehat{\partial} \vartheta(\bar{x}_N(t_j), \dot{\bar{x}}_N(t), t_j)$$

$$+ \widehat{N}\big((\bar{x}_N(t_j), \dot{\bar{x}}_N(t)); \operatorname{gph} F(t_j)\big) + \varepsilon_N I\!B^* \Big\}$$

for $t \in [t_j, t_{j+1})$ with $j = 0, \ldots, N-1$. Now passing to the limit as $N \to \infty$ and using the *pointwise* convergence results established below, we arrive at the extended Euler-Lagrange inclusion (6.44).

To derive the transversality inclusion (6.45), we take the limit in the discrete ones (6.42) as $N \to \infty$. The only thing to clarify is the possibility to pass from Fréchet normals to $\Omega_N = \Omega + \eta_N I\!B$ to the basic normals to Ω. The latter can be easily done by using the sum rule from Theorem 3.7(i) and the fact that $\eta_N \downarrow 0$ as $N \to \infty$.

It remains to justify the nontriviality condition $\big(\lambda, p(\cdot)\big) \ne 0$. Assuming that $\lambda = 0$, one may put $\lambda_N = 0$ for all $N \in I\!N$ without loss of generality.

We need to show that $p(\cdot)$ is not identically equal to zero on $[a,b]$. Suppose the contrary, i.e., $p(t) = 0$ whenever $t \in [a,b]$. Then it follows from the above proof that $p_N(t) \xrightarrow{w^*} 0$ for all $t \in [a,b]$; in particular, $p_N(t_0) \xrightarrow{w^*} 0$ and $p_N(t_N) \xrightarrow{w^*} 0$ as $N \to \infty$. The discrete transversality inclusion (6.42) is written in this case as

$$\big(p_N(t_0), -p_N(t_N)\big) \in \widehat{N}\big((\bar{x}_N(t_0), \bar{x}_N(t_N)); \Omega + \eta_N I\!B\big) + \varepsilon_N I\!B^* \,. \tag{6.46}$$

Using again Theorem 3.7(i) for the Fréchet normals cone to the sum in (6.46) and then employing the strong PSNC property of Ω at $\big(\bar{x}(a), \bar{x}(b)\big)$ with respect to the second component, we get $\|p_N(t_N)\| \to 0$ as $N \to \infty$, which contradicts the nontriviality condition (6.42) in Theorem 6.19 and completes the proof of this theorem. △

The next theorem gives necessary optimality conditions in the extended Euler-Lagrange form for the original Bolza problem (P) derived by passing to the limit from the approximate necessary optimality in the discrete-time problems (P_N). In contrast to Theorem 6.21, this theorem applies to the *summable* integrands $\vartheta(x, v, \cdot)$ and gives a better form of the Euler-Lagrange inclusion. On the other hand, it imposes more restrictive assumptions on the state space X in question. In the formulations and proof of this theorem we keep the same notational agreement as for Theorem 6.21 discussed above.

Theorem 6.22 (extended Euler-Lagrange conditions for relaxed local minimizers in Bolza problems with summable integrands). *Let $\bar{x}(\cdot)$ be a relaxed intermediate local minimizer for the Bolza problem (P) under assumptions* (H1), (H2), (H3′), *and* (H4′). *Suppose also that the space X is reflexive and separable and that the set Ω is strongly PSNC at $\big(\bar{x}(a), \bar{x}(b)\big)$ with respect to the second component. Then there are a number $\lambda \geq 0$ and an absolutely continuous mapping $p\colon [a,b] \to X^*$, not both zero, satisfying the extended Euler-Lagrange inclusion*

$$\dot{p}(t) \in \operatorname{co}\Big\{u \in X^* \Big|\ \big(u, p(t)\big) \in \lambda\partial\vartheta(\bar{x}(t), \dot{\bar{x}}(t), t) + N_+\big((\bar{x}(t), \dot{\bar{x}}(t)); \operatorname{gph} F(t)\big)\Big\} \tag{6.47}$$

for a.e. $t \in [a,b]$ and the transversality inclusion (6.45).

Proof. We follow the lines in the proof of Theorem 6.21 using the sequence of discrete approximation problems (P_N) instead of $(\overline{P}_N)$. The only difference is in the justification of the extended Euler-Lagrange inclusion (6.47) in comparison with (6.44) that are based on generally different discrete-time counterparts (6.43) and (6.41) under somewhat different assumptions.

To proceed, we suppose for notation convenience that the discrete Euler-Lagrange inclusions (6.43) hold as $N \to \infty$ without taking the closure of the set-valued integral therein; this doesn't restrict the generality as follows from the proof below. Then, by (6.43) and the definition of the Fréchet coderivative, there are dual elements

$$(x^*_{Nj}, v^*_{Nj}) \in \int_{t_j}^{t_{j+1}} \partial\vartheta\Big(\bar{x}_N(t_j), \frac{\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)}{h_N}, t\Big)\, dt, \quad j = 0, \ldots, N-1\ ,$$

as well as $e^*_{Nj}, \widetilde{e}^*_{Nj} \in I\!B^*$ satisfying the inclusions

$$\Big(\frac{p_N(t_{j+1}) - p_N(t_j)}{h_N} - \lambda_N x^*_{Nj}\Big) + \varepsilon_N e^*_{N_j}$$

$$\in \widehat{D}^* F_j\Big(\bar{x}_N(t_j), \frac{\bar{x}_N(t_{j+1}) - \bar{x}_N(t_j)}{h_N}\Big)\Big(\lambda_N v^*_{Nj} + \lambda_N \frac{\theta_{Nj}}{h_N} b^*_{Nj} - p_N(t_{j+1}) + \varepsilon_N \widetilde{e}^*_{N_j}\Big)$$

that are fulfilled for all $j = 0, \ldots, N-1$ along a sequence of $N \to \infty$; put below $N \in I\!N$ for simplicity. Following the proof of Theorem 6.21, we find an absolutely continuous mapping $p\colon [a,b] \to X^*$ such that $p_N(t) \to p(t)$ weakly in X^* for all $t \in [a,b]$ and a sequence of convex combinations of $\dot{p}_N(t)$ converges to $\dot{p}(t)$ almost everywhere on $[a,b]$ as $N \to \infty$. Then rewrite the above discrete-time inclusions in the form

$$\dot{p}_N(t) \in \Big\{u \in X^* \Big|\ \big(u, p_N(t_{j+1}) - \lambda_N \theta_N(t)\big) \in \frac{\lambda_N}{h_N}(x^*_{Nj}, v^*_{Nj})$$

$$+ \widehat{N}\big((\bar{x}_N(t_j), \dot{\bar{x}}_N(t)); \operatorname{gph} F(t_j)\big) + \varepsilon_N I\!B^*\Big\}$$

for $t \in [t_j, t_{j+1})$ with $j = 0, \ldots, N-1$. By the construction of (x^*_{Nj}, v^*_{Nj}) there are summable mappings $u^*_{Nj}\colon [t_j, t_{j+1}] \to X^*$ and $w^*_{Nj}\colon [t_j, t_{j+1}] \to X^*$ satisfying the relations

$$\big(u^*_{Nj}(t), w^*_{Nj}(t)\big) \in \partial\vartheta\Big(\bar{x}_N(t_j), \frac{\bar{x}_N(t_j) - \bar{x}_N(t_{j+1})}{h_N}, t\Big) \ \text{ a.e. } \ t \in [t_j, t_{j+1}]\ ,$$

$$\frac{(x^*_{Nj}, v^*_{Nj})}{h_N} = \frac{1}{h_N}\int_{t_j}^{t_{j+1}} \big(u^*_{Nj}(t), w^*_{Nj}(t)\big)\, dt \ \text{ for } \ j = 0, \ldots, N-1\ .$$

Define the sequences of mappings $u^*_N\colon [a,b] \to X^*$ and $w^*_N\colon [a,b] \to X^*$ on the whole interval $[a,b]$ by

$$\big(u^*_N(t), w^*_N(t)\big) := \big(u^*_{Nj}(t), w^*_{Nj}(t)\big) \ \text{ for } \ t \in [t_j, t_{j+1}), \quad j = 0, \ldots, N-1\ .$$

Since $u^*_N(\cdot)$ and $w^*_N(\cdot)$ are integrable bounded on $[a,b]$, there are subsequences of them that converge, by the Dunford theorem, to some $u^*(\cdot)$ and $w^*(\cdot)$ in the weak topology of $L^1\big([a,b]; X^*\big)$. Invoking again the Mazur weak closure theorem and using the strong convergence of $\bar{x}_N(\cdot) \to \bar{x}(\cdot)$ from Theorem 6.13, one has the relations

$$\big(u^*(t), w^*(t)\big) \in \operatorname{clco} \partial\vartheta\big(\bar{x}(t), \dot{\bar{x}}(t), t\big) = \operatorname{co} \partial\vartheta(\bar{x}(t), \dot{\bar{x}}(t), t) \text{ a.e. } t \in [a, b] ,$$

where the *closure operation can be omitted* due to the reflexivity of X and the compactness of $\operatorname{co} \partial\vartheta(\bar{x}(t), \dot{\bar{x}}(t), t)$ in the weak topology of X^*, and hence its closedness in the strong topology of this space. Employing now the infinite-dimensional counterpart of the Lyapunov-Aumann theorem mentioned in the proof of Lemma 6.18, the well-known property

$$\lim_{h\to 0} \frac{1}{h} \int_t^{t+h} f(s)\,ds = f(t) \text{ a.e. } t \in [a, b]$$

of the Bochner integral, and also the weak closedness of the basic subdifferential for locally Lipschitzian functions on reflexive spaces (cf. Theorem 3.59), we conclude that there are subgradients $\big(x^*(t), v^*(t)\big)$ of $\vartheta(\cdot, \cdot, t)$ such that

$$\frac{\lambda_N}{h_N}(x^*_{Nj}, v^*_{Nj}) \xrightarrow{w^*} \big(x^*(t), v^*(t)\big) \in \partial\vartheta(\bar{x}(t), \dot{\bar{x}}(t), t) \text{ a.e. } t \in [a, b] .$$

Passing finally to the limit in the above inclusions for $\dot{p}_N(\cdot)$ as $N \to \infty$, we arrive at the desired extended Euler-Lagrange inclusion (6.47), where the closure operation can be dropped in the reflexive case under consideration due to the uniform boundedness of $p_N(\cdot)$ and $\dot{p}_N(\cdot)$; see the discussion above. Note that it is sufficient to use the basic subdifferential in the integrand $\vartheta(\cdot, \cdot, t)$ in (6.47), but not the extended one as in (6.44), in the case under consideration. Thus we complete the proof of the theorem. △

The nontriviality condition in both Theorems 6.21 and 6.22 ensures that the pair $\big(\lambda, p(\cdot)\big)$ satisfying the Euler-Lagrange and transversality inclusions is not zero. The next result presents additional assumptions under which we have the *enhanced* nontriviality conditions: $\big(\lambda, p(b)\big) \neq 0$.

Corollary 6.23 (extended Euler-Lagrange conditions with enhanced nontriviality). *Let $\bar{x}(\cdot)$ be an r.i.l.m. for the Bolza problem (P). In addition to the assumptions in Theorems* 6.21 *and* 6.22*, respectively, suppose that*

(a) *either $\Omega = \Omega_a \times \Omega_b$, where Ω_b is SNC at $\bar{x}(b)$;*

(b) *or Ω is strongly PSNC at $\big(\bar{x}(a), \bar{x}(b)\big)$ relative to the second component, $F(\cdot, t)$ is strongly coderivatively normal at $\big(\bar{x}(t), \dot{\bar{x}}(t)\big)$, and* $\operatorname{gph} F(t)$ *is normally semicontinuous at this point for a.e. $t \in [a, b]$.*

Then one has the extended Euler-Lagrange and transversality inclusions (6.44) *and* (6.45) $\big($*respectively,* (6.47) *and* (6.45)$\big)$ *with the replacement of*

$$N_+\big((\bar{x}(t), \dot{\bar{x}}(t)); \operatorname{gph} F(t)\big) \quad \text{by} \quad N\big((\bar{x}(t), \dot{\bar{x}}(t)); \operatorname{gph} F(t)\big)$$

in case (b) *and with the enhanced nontriviality condition $\lambda + \|p(b)\| = 1$.*

Proof. Following the (same) proof of the nontriviality condition in Theorems 6.21 and 6.22, one has the transversality inclusion (6.46) for the adjoint trajectories $p_N(\cdot)$ in the discrete approximations with $\lambda_N = 0$. Assuming (a), we arrive at

$$-p_N(t_N) \in \widehat{N}\big(\bar{x}_N(t_N); \Omega_b + \eta I\!\!B\big) + \varepsilon_N I\!\!B^* \quad \text{as} \quad N \to \infty\ ,$$

which implies, by Theorem 3.7(i) and the SNC property of Ω_b at $\bar{x}(b)$, that $\|p_N(t_N)\| \to 0$ whenever $p_N(t_N) \stackrel{w^*}{\to} 0$ as $N \to \infty$. This clearly contradicts the nontriviality condition for the discrete-time problems $(\overline{P}_N)$ and (P_N) from Theorems 6.19 and 6.20, respectively.

It remains to justify the nontriviality condition $\lambda + \|p(b)\| \neq 0$ in case (b). It follows from the fact that, under the assumptions made in (b), $p(t) = 0$ for all $t \in [a,b]$ whenever $p(\cdot)$ satisfies the extended Euler-Lagrange inclusion (6.44) with $\lambda = 0$ and $p(b) = 0$. Indeed, invoking the normal semicontinuity of $\operatorname{gph} F(t)$ in this case, we write (6.44) as

$$\dot{p}(t) \in \operatorname{clco}\Big\{u \in X^*\Big|\ \big(u, p(t)\big) \in N\big((\bar{x}(t), \dot{\bar{x}}(t)); \operatorname{gph} F(t)\big)\Big\} \quad \text{a.e.} \quad t \in [a,b]$$

that is equivalent, by the strong coderivative normality assumption in (b), to

$$\dot{p}(t) \in \operatorname{clco} D^*_M F\big(\bar{x}(t), \dot{x}(t)\big)\big(-p(t)\big) \quad \text{a.e.} \quad t \in [a,b]\ .$$

The latter clearly implies, due to the mixed coderivative condition for the Lipschitz continuity from Theorem 1.44, that

$$p(t) \equiv 0 \quad \text{on} \quad [a,b] \quad \text{when} \quad p(b) = 0\ ,$$

which completes the proof of the corollary. △

If X is *finite-dimensional*, any set is SNC and any mapping $F\colon X \rightrightarrows X$ is strongly coderivatively normal at every point. Thus we automatically have the extended Euler-Lagrange conditions in Theorem 6.22 and Corollary 6.23. Another setting that *doesn't require any SNC/PSNC assumptions* on the constraint set Ω is the case of endpoint constraints given by a *finite number of equalities and inequalities* with locally Lipschitzian functions considered next.

Corollary 6.24 (extended Euler-Lagrange conditions for problems with functional endpoint constraints). *Let the endpoint constraint set Ω in problem (P) be given by*

$$\Omega := \Big\{(x_a, x_b) \in X^2\Big|\ \varphi_i(x_a, x_b) \le 0, \quad i = 1, \ldots, m\ ,$$
$$\varphi_i(x_a, x_b) = 0, \quad i = m+1, \ldots, m+r\Big\}\ ,$$

where each φ_i is locally Lipschitzian around $\big(\bar{x}(a), \bar{x}(b)\big)$ together with the cost function $\varphi_0 := \varphi$. Suppose that all the assumptions of Corollary 6.23 *hold except those related to the SNC/PSNC properties of Ω. Then there are nonnegative multipliers $(\lambda_0,\ldots,\lambda_{m+r})\neq 0$ with*

$$\lambda_i\varphi_i\big(\bar{x}(a),\bar{x}(b)\big)=0,\quad i=1,\ldots,m\ ,$$

and an absolutely continuous adjoint arc $p\colon[a,b]\to X^*$ satisfying the extended Euler-Lagrange inclusions mentioned therein as well as the following transversality condition:

$$\big(p(a),-p(b)\big)\in\sum_{i=0}^{m}\lambda_i\partial\varphi_i\big(\bar{x}(a),\bar{x}(b)\big)$$
$$+\sum_{i=m+1}^{m+r}\lambda_i\Big[\partial\varphi_i\big(\bar{x}(a),\bar{x}(b)\big)\bigcup\partial\big(-\varphi_i\big)\big(\bar{x}(a),\bar{x}(b)\big)\Big]\ .$$

If, in particular, all φ_i are strictly differentiable at $\big(\bar{x}(a),\bar{x}(b)\big)$, then there are $(\lambda_0,\ldots,\lambda_{m+r})\neq 0$ satisfying the above complementary slackness condition and the standard sign condition

$$\lambda_i\geq 0\ \ \textit{for}\ \ i=0,\ldots,m$$

and such that the transversality condition

$$\big(p(a),-p(b)\big)=\sum_{i=0}^{m+r}\lambda_i\nabla\varphi_i\big(\bar{x}(a),\bar{x}(b)\big)$$

supplements the corresponding Euler-Lagrange inclusion of Corollary 6.23.

Proof. Suppose first that the locally Lipschitzian functions $\varphi_1,\ldots,\varphi_{m+r}$ satisfy the nonsmooth counterpart of the Mangasarian-Fromovitz constraint qualification formulated in Theorem 3.86. Then the constraint set Ω defined in this corollary is SNC at $\big(\bar{x}(a),\bar{x}(b)\big)$. Furthermore, it follows from the calculus rule of Theorem 3.8 specified for $F:=(\varphi_1,\ldots,\varphi_{m+r})$ and

$$\Theta:=\Big\{(\alpha_1,\ldots,\alpha_{m+r})\in I\!R^{m+r}\Big|\ \alpha_i\leq 0,\ \ i=1,\ldots,m\ ,$$
$$\alpha_i=0,\ \ i=m+1,\ldots,m+r\Big\}$$

therein that the same constraint qualification ensures the inclusion

$$N(\bar{z};\Omega)\subset\Big\{\sum_{i=1}^{m}\lambda_i\partial\varphi_i(\bar{z})+\sum_{i=m+1}^{m+r}\lambda_i\Big[\partial\varphi_i(\bar{z})\bigcup\partial\big(-\varphi_i\big)(\bar{z})\Big]\Big|$$
$$\lambda_i\geq 0,\ \ i=1,\ldots,m+r;\ \ \lambda_i\varphi_i(\bar{z})=0,\ \ i=1,\ldots,m\Big\}$$

for basic normals to the constraint set Ω at the point $\bar{z} := \big(\bar{x}(a), \bar{x}(b)\big)$. Then the transversality inclusion formulated at this corollary follows from (6.45) with $\lambda_0 = \lambda$, where the nontriviality condition $\big(\lambda, p(b)\big) \neq 0$ is equivalent to $(\lambda_0, \ldots, \lambda_{m+r}) \neq 0$. Assuming finally that the qualification conditions of Theorem 3.86 don't hold, we immediately arrive at the desired transversality inclusion with $(\lambda_1, \ldots, \lambda_{m+r}) \neq 0$ and complete the proof. △

Note that the enhanced nontriviality condition $\big(\lambda_0, p(b)\big) \neq 0$, inspired by the one in Corollary 6.23, may *not* hold in the framework of Corollary 6.24 if the constraint set Ω is not SNC (or strongly PSNC); in particular, when the Mangasarian-Fromovitz type constraint qualification of Theorem 3.86 is not fulfilled. It may happen, for instance, for a *two-point boundary problem* with $x(a) = x_0$ and $x(b) = x_1$ involving smooth parabolic systems of optimal control; see the well-known examples in Fattorini [432] and Li and Yong [789]. On the other hand, the SNC requirement is met in case (a) of Corollary 6.23 when $x(a) = x_0$ and $x(b) \in x_1 + r I\!B$ with $r > 0$, since the latter ball is always SNC (it is actually epi-Lipschitzian by Proposition 1.25).

Observe also that, using the *smooth variational description* of Fréchet subgradients similarly to the proof of Theorem 5.19 for nondifferentiable programming and employing the results of Corollary 6.24 in the case of smooth endpoint functions, we can derive counterparts of Theorems 6.21 and 6.22 with *upper subdifferential* transversality conditions; see Remark 6.30 for the exact formulation and more details.

To conclude this section, let us discuss some particular issues mostly related to the above Euler-Lagrange conditions for differential inclusions with infinite-dimensional state spaces.

Remark 6.25 (discussion on the Euler-Lagrange conditions).

(i) It follows from the proof of Theorems 6.21 and 6.22 that the strong PSNC assumption imposed on Ω to ensure the nontriviality condition may be replaced by the following *alternative* assumption on F written as: there is $t \in [a, b]$ such that for any sequences $t_k \to t$, $x_k \to \bar{x}(t)$, $v_k \in F(x_k, t_k)$, and $(x_k^*, v_k^*) \in \widehat{N}\big((x_k, v_k); \operatorname{gph} F(t_k)\big)$ one has

$$(x_k^*, v_k^*) \xrightarrow{w^*} (0, 0) \Longrightarrow \|v_k^*\| \to 0 \quad \text{as} \quad k \to \infty\,.$$

This property is closely related to the *strong PSNC* property of F at $(\bar{x}(t), t)$ with respect to the *image component*; cf. also its SNC analog for moving sets in Definition 5.71.

(ii) Recall that the SNC property of *convex sets* with nonempty relative interiors is *equivalent* by Theorem 1.21 to the *finite codimension* property of their closed affine hulls. The *strong PSNC* property may be essentially weaker than the SNC one; see, e.g., Theorem 1.75.

(iii) If the velocity sets $F(x, t)$ and the integrand $\vartheta(x, \cdot, t)$ are *convex* around the given local minimizer, then the Euler-Lagrange inclusion of Theorem 6.21 easily implies the *Weierstrass-Pontryagin maximum condition*

$$\langle p(t), \dot{\bar{x}}(t)\rangle - \lambda\vartheta\big(\bar{x}(t), \dot{\bar{x}}(t), t\big) = \max_{v\in F(\bar{x}(t),t)} \Big\{\langle p(t), v\rangle - \lambda\vartheta\big(\bar{x}(t), v, t\big)\Big\}$$

for a.e. $t \in [a,b]$. It can be directly derived from the extremal property of the coderivative of convex-valued mappings in Theorem 1.34. The latter is the underlying condition of the results unified under the label "(Pontryagin) maximum principle" in optimal control. It will be shown in the next subsection that the maximum condition supplements, at least in the case of reflexive and separable state spaces under some additional assumptions, the extended Euler-Lagrange inclusion with *no convexity* requirements. To this end we note that the SNC (actually strong PSNC) properties required in Theorems 6.21 and 6.22 may be viewed as *nonconvex counterparts* of *finite codimension* requirements in the theory of necessary optimality conditions for controlled *evolution equations* of type (6.2) and their *PDE specifications* known in the case of *smooth* velocity mappings f and *convex* constraint/target sets Ω; cf. the afore-mentioned books by Fattorini [432] and Li and Yong [789] with the references and discussions therein.

Remark 6.26 (optimal control of semilinear unbounded differential inclusions). Many important models involving *semilinear partial differential equations* can be appropriately described by $\mathcal{C}_0$ *semigroups*; we again refer to the books by Fattorini [432] and Li and Yong [789] as well as to the subsequent material of Sects. 7.2–7.4 in this book. In this way an analog of the optimal control problem (P) from this section can be considered with the replacement of the differential inclusion (6.1) by the evolution model

$$\dot{x}(t) \in Ax(t) + F\big(x(t), t\big)\ ,$$

where A is an *unbounded* infinitesimal generator of a *compact* $\mathcal{C}_0$ semigroup on X, and where *continuous* solutions $x(\cdot)$ to this inclusion are understood in the *mild* sense. The latter means that there is a Bochner integrable mapping $v(\cdot) \in L^1\big([a,b]; X\big)$ such that

$$v(t) \in F\big(x(t), t\big) \quad \text{a.e.}\ \ t \in [a,b] \quad \text{and}$$

$$x(t) = e^{A(t-a)}x(a) + \int_a^t e^{A(t-s)}v(s)\,ds, \qquad t \in [a,b]\ .$$

Developing the above approach in the case of the *Mayer cost functional*

$$\text{minimize}\ \ \varphi\big(x(a), x(b)\big)\ \ \text{with}\ \ \big(x(a), x(b)\big) \in \Omega \subset X^2\ ,$$

we derive necessary optimality conditions under the additional *convexity* assumption of the velocity sets $F(x,t)$ around the optimal solution. Then the *extended Euler-Lagrange inclusion* in the case of reflexive and separable state spaces X and autonomous systems (for simplicity) is formulated as follows:

$$\begin{cases} p(t) \in e^{A^*(b-t)}p(b) \\ \qquad + \displaystyle\int_b^t \Big\{ e^{A^*(s-t)} D_N^* F\big(\bar{x}(s), v\big)\big(-p(s)\big) \Big| \, v \in M\big(\bar{x}(s), p(s)\big)\Big\}\, ds \end{cases}$$

for all $t \in [a,b]$, where $p\colon [a,b] \to X^*$ is a *continuous* mapping satisfying the *transversality* and *nontriviality* conditions

$$\big(p(a), -p(b)\big) \in \lambda\partial\varphi\big(\bar{x}(a), \bar{x}(b)\big) + N\big((\bar{x}(a), \bar{x}(b)); \Omega\big), \quad \lambda + \|p(b)\| \neq 0$$

with $\lambda \geq 0$, where the *argmaximum sets* $M(x,p)$ are defined by

$$M(x,p) := \big\{ v \in F(x) \big| \, \langle p, v\rangle = \mathcal{H}(x,p)\big\}$$

with

$$\mathcal{H}(x,p) := \max\big\{ \langle p, v\rangle \big| \, v \in F(x)\big\}\ .$$

Moreover, the extended Euler-Lagrange inclusion implies in this case the *Weierstrass-Pontryagin maximum condition*

$$\langle p(t), \bar{v}(t)\rangle = \mathcal{H}\big(\bar{x}(t), p(t)\big) \ \text{ a.e. } \ t \in [a,b]$$

with a measurable mapping $\bar{v}(t) \in F\big(\bar{x}(t)\big)$ satisfying

$$p(t) \in e^{A^*(b-t)}p(b) + \int_b^t \Big\{ e^{A^*(s-t)} D_N^* F\big(\bar{x}(s), \bar{v}(s)\big)\big(-p(s)\big)\Big\}\, ds, \quad t \in [a,b];$$

see Mordukhovich and D. Wang [970, 971] for proofs and more discussions on these and related results.

6.2 Necessary Optimality Conditions for Differential Inclusions without Relaxation

This section is mainly devoted to deriving necessary optimality conditions for nonconvex differential inclusions *without any relaxation* based on approximating the original constrained problem by a family of nonsmooth Bolza problems with no differential inclusions and no endpoint constraints. The extended Euler-Lagrange conditions for the latter class of *unconstrained* Bolza problems and the assumptions made allow essential specifications in comparison with the general results established in the preceding section. By passing to the limit, we obtain necessary optimality conditions of the Euler-Lagrange type for arbitrary (i.e., *non-relaxed*) intermediate minimizers for the original control problems with *reflexive and separable* state spaces. Moreover, they are supplemented by the *Weierstrass-Pontryagin maximum condition* valid in the general nonconvex setting. If the state space X is *finite-dimensional* and the velocity sets $F(x,t)$ are *convex*, the above Euler-Lagrange and maximum conditions are *equivalent* to the *extended Hamiltonian inclusion* expressed via a *partial convexification* of the basic subdifferential of the Hamiltonian function associated with $F(x,t)$. We also discuss various generalizations of the results obtained and present some illustrative examples.

6.2.1 Euler-Lagrange and Maximum Conditions for Intermediate Local Minimizers

The realization of the approach mentioned above requires some additional assumptions on the initial data in comparison with Theorem 6.22, while the a.e. continuity assumption on the velocity mapping $F(x,\cdot)$ can be replaced by its measurability; see below. Furthermore, it is more convenient in this section to consider the following *Mayer form* (P_M) of problem (P) studied in the preceding section, with a fixed left endpoint of feasible arcs:

$$\text{minimize } \varphi\big(x(b)\big) \text{ subject to } x(b)\in\Omega\subset X$$

over absolutely continuous trajectories of the differential inclusion

$$\dot{x}(t)\in F(x(t),t) \text{ a.e. } t\in[a,b], \quad x(a)=x_0 . \tag{6.48}$$

The general case of nonzero integrands f in the Bolza problem can be reduced to the Mayer one by standard state augmentation techniques. Note also that, since the state space X is assumed to be reflexive and separable in what follows, this notion of absolutely continuous solutions to (6.48) agrees with the one given in Definition 6.1.

We first formulate the assumptions on the set-valued mapping F in (6.48) that are *weaker* than those imposed in Theorem 6.22. Keeping assumption (H1) from Subsect. 6.1.1 on the compactness and Lipschitz continuity of F in x with possibly *summable* functions $m_F(\cdot)$ and $\ell_F(\cdot)$ on $[a,b]$ (although it may also be loosen in some directions by various standard reductions as, e.g., in [255, 261, 598, 1289]), we replace the a.e. continuity assumption (H2) by the *measurability* assumption on F in the time variable $t\in[a,b]$. Note that all the reasonable notions of measurability are *equivalent* for set-valued mappings with closed values in *separable* spaces (cf. the discussion in the proof of Lemma 6.18), which is the case in this section.

(H2′) $F(x,\cdot)$ is measurable on the interval $[a,b]$ uniformly in x on the open set $U\subset X$ taken from (H1).

We also weaken the continuity and Lipschitz continuity assumptions on the cost function $\varphi=\varphi(x)$ from (H4) and (H4′) observing that this leads to the modified (more general) transversality condition for the Mayer problem under consideration. Namely, we replace the latter assumptions by the following one:

(H4″) φ is l.s.c. around $\bar{x}(b)$ relative to Ω, which is suppose to be locally closed around this point.

On the other hand, the following theorem imposes the *additional* coderivative normality and SNC assumptions on F in comparison with Theorem 6.22 and Corollary 6.23. Observe that the coderivative form of the extended Euler-Lagrange inclusion given below is *equivalent* to the one from Corollary 6.23 for $\vartheta = 0$ *without* imposing the normal semicontinuity assumptions on $\operatorname{gph} F(t)$. In the rest of this subsection we study intermediate local minimizers of *rank one* from Definition 6.7. Recall that $\varphi_\Omega(\cdot) = \varphi(\cdot) + \delta(\cdot;\Omega)$ as usual.

Theorem 6.27 (Euler-Lagrange and Weierstrass-Pontryagin conditions for nonconvex differential inclusions). *Let $\bar{x}(\cdot)$ be an intermediate local minimizer for the Mayer problem (P_M) under assumptions* (H1), H2′), *and* (H4″). *Suppose in addition that:*

(a) *the Banach space X is reflexive, separable, and admits an equivalent Kadec norm;*

(b) *the function φ_Ω is SNEC at $\bar{x}(b)$, and its epigraph is weakly closed;*

(c) *the mapping $F(\cdot,t)\colon X \rightrightarrows X$ is SNC at $(\bar{x}(t),\dot{\bar{x}}(t))$, strongly coderivatively normal around this point, and its graph is weakly closed for a.e. $t \in [a,b]$.*

Then there exist a number $\lambda \ge 0$ and an absolutely continuous adjoint arc $p\colon [a,b] \to X^*$, not both zero, satisfying the Euler-Lagrange inclusion

$$\dot{p}(t) \in \operatorname{co} D_x^* F\big(\bar{x}(t),\dot{\bar{x}}(t),t\big)\big(-p(t)\big) \quad \text{a.e. } t \in [a,b]\,, \tag{6.49}$$

the Weierstrass-Pontryagin maximum condition

$$\big\langle p(t),\dot{\bar{x}}(t)\big\rangle = \max_{v \in F(\bar{x}(t),t)} \big\langle p(t),v\big\rangle \quad \text{a.e. } t \in [a,b]\,, \tag{6.50}$$

and the transversality inclusion

$$\big(-p(b),-\lambda\big) \in N\big((\bar{x}(b),\bar{\beta});\operatorname{epi}\varphi_\Omega\big)\,. \tag{6.51}$$

Moreover, (6.51) *always implies*

$$-p(b) \in \partial\big[\lambda\varphi + \delta(\cdot;\Omega)\big]\big(\bar{x}(b)\big) \tag{6.52}$$

being equivalent to the latter condition if φ is Lipschitz continuous around $\bar{x}(b)$ relative to Ω.

Proof. Consider the parametric functional

$$\theta_\beta(x) := \operatorname{dist}\big((x(b),\beta);\operatorname{epi}\varphi_\Omega\big) \quad \text{as} \quad \beta \in I\!R$$

over feasible arcs/trajectories to the original differential inclusion (6.1) with no other constraints. In what follows we fix the open set $U \subset X$ from assumption (H1) regarding $\bar{x}(\cdot)$. For every $\varepsilon > 0$ one obviously has

$$\theta_\beta(\bar{x}) \le |\beta - \bar{\beta}|$$

whenever β is sufficiently close to $\bar{\beta} = \varphi(\bar{x}(b))$. Since $\bar{x}(\cdot)$ is an intermediate local minimizer for (P_M) and by the structure of $\theta_\beta(x)$, we get

$$\theta_\beta(x) > 0 \;\text{ for any }\; \beta < \bar{\beta}$$

whenever a trajectory $x(t)$ for (6.48) belongs to some $W^{1,1}$-neighborhood of the local minimizer under consideration and such that

$$x(t) \in U \;\text{ for all }\; t \in (a, b] \; .$$

Form now the space $\mathcal{X}$ of all the trajectories $x(\cdot)$ for (6.48) satisfying the only constraint $x(t) \in \operatorname{cl} U$ as $t \in (a, b]$ with the metric

$$d(x, y) := \int_a^b \|\dot{x}(t) - \dot{y}(t)\| \, dt \; .$$

It is easy to see, from Definition 6.1 of solutions to the original differential inclusion and standard properties of the Bochner integral, that the metric space $\mathcal{X}$ is complete and that the function $\theta_\beta(\cdot)$ is (Lipschitz) continuous on $\mathcal{X}$ for any $\beta \in I\!R$. It follows from the above constructions that for every $\varepsilon > 0$ there is $\beta_\varepsilon < \bar{\beta}$ such that $\beta_\varepsilon \to \bar{\beta}$ as $\varepsilon \downarrow 0$ and

$$0 \le \theta_\varepsilon(\bar{x}) < \varepsilon \le \inf_{x \in \mathcal{X}} \theta_\varepsilon(x) + \varepsilon \quad \text{with } \; \theta_\varepsilon := \theta_{\beta_\varepsilon} \; .$$

Applying the *Ekeland variational principle* from Theorem 2.26(i), we find an arc $x_\varepsilon(\cdot) \in \mathcal{X}$ satisfying

$$d(x_\varepsilon, \bar{x}) \le \sqrt{\varepsilon} \;\text{ and }\; \theta_\varepsilon(x) + \sqrt{\varepsilon} d(x, x_\varepsilon) \ge \theta_\varepsilon(x_\varepsilon)$$

for all $x \in \mathcal{X}$. Note that the distance estimate above yields that $x_\varepsilon(t) \in U$ as $t \in (a, b]$ and that $x_\varepsilon(\cdot)$ belongs to the fixed $W^{1,1}$-neighborhood of the intermediate local minimizer $\bar{x}(\cdot)$ for small $\varepsilon > 0$. Hence $\theta_\varepsilon(x_\varepsilon) > 0$.

Next, given any $\alpha, \varepsilon > 0$ and the summable Lipschitz constant $\ell_F(\cdot)$ from (6.5), we define the Bolza-type functional

$$J_\varepsilon^\alpha[x] := \theta_\varepsilon(x) + \sqrt{\varepsilon} d(x, x_\varepsilon) + \alpha \int_a^b \sqrt{1 + \ell_F^2(t)} \operatorname{dist}\big((x(t), \dot{x}(t)); \operatorname{gph} F(t)\big) \, dt$$

on the sets of all absolutely continuous mappings $x \colon [a, b] \to X$, not necessarily trajectories for (6.48), satisfying $x(t) \in U$ as $t \in (a, b]$. To proceed, we need the following auxiliary result.

Claim. *There is a number* $\alpha \geq 1$ *such that for every* $\varepsilon \in (0, 1/\alpha)$ *the absolutely continuous mapping* $x_\varepsilon\colon [a,b] \to X$ *built above provides an intermediate local minimum for the Bolza functional* J_ε^α *subject to*

$$x(a) = x_0 \quad and \quad x(t) \in U \quad for \quad t \in (a,b] \,.$$

To prove this claim, we first observe that there are positive numbers ν, γ such that for every arc $y(\cdot)$ satisfying $y(a) = x_0$, $y(t) \in U$ as $t \in (a,b]$, and

$$\int_a^b \operatorname{dist}\big(\dot{y}(t); F(y(t),t)\big)\, dt < \nu$$

there exists a trajectory $x(\cdot)$ for (6.28) with

$$d(x,y) \leq \gamma \int_a^b \sqrt{1 + \ell_F^2(t)}\, \operatorname{dist}\big((y(t), \dot{y}(t)); \operatorname{gph} F(t)\big)\, dt \,. \tag{6.53}$$

Indeed, this follows directly from Filippov's theorem on quasitrajectories of differential inclusions (see, e.g., Theorem 1 on p. 120 in Aubin and Cellina [50] whose proof holds true for infinite-dimensional inclusions under the assumptions made in (H1) and (H2$'$)) and from the estimate

$$\operatorname{dist}\big(v, F(u,t)\big) \leq \sqrt{1 + \ell_F^2(t)}\, \operatorname{dist}\big((u,v); \operatorname{gph} F(t)\big)$$

that is obviously valid under (H1). Suppose now that the above claim doesn't hold. Then for each $k \in I\!N$ there are $\varepsilon_k \in (0, 1/k)$ and an arc $y_i(\cdot) \in \mathcal{X}$ satisfying $y_k(t) \in U$ as $t \in (a,b]$,

$$\max_{t \in [a,b]} \|y_k(t) - x_{\varepsilon_k}(t)\| + \int_a^b \|\dot{y}_k(t) - \dot{x}_{\varepsilon_k}(t)\|\, dt < \frac{1}{k} \,,$$

and $J_{\varepsilon_k}^k[x_{\varepsilon_k}] > J_{\varepsilon_k}^k[y_k]$. Hence $y_k(\cdot) \to \bar{x}(\cdot)$ in the norm topology of $W^{1,1}\big([a,b]; X\big)$ and, moreover,

$$J_{\varepsilon_k}^k[x_{\varepsilon_k}] = \theta_{\varepsilon_k}(x_{\varepsilon_k}) \downarrow 0 \quad \text{as} \quad k \to \infty \,.$$

Therefore, given any $\nu > 0$, we get

$$\int_a^b \operatorname{dist}\big(\dot{y}_k(t); F(y_k(t),t)\big)\, dt < J_{\varepsilon_k}^k[x_{\varepsilon_k}] < \nu$$

for large k. This implies, by (6.53), that there are a number $\gamma > 0$ independent of k and trajectories $x_k(\cdot)$ for (6.28) as $k \to \infty$ such that

$$d(x_k, y_k) \leq \gamma \int_a^b \sqrt{1 + \ell_F^2(t)}\, \operatorname{dist}\big((y_k(t), \dot{y}_k(t)); \operatorname{gph} F(t)\big)\, dt \,. \tag{6.54}$$

Since the right-hand side of (6.54) converges to zero and since $y_k(\cdot) \to \bar{x}(\cdot)$ strongly in $W^{1,1}([a,b];X)$, we get the strong $W^{1,1}$-convergence $x_k(\cdot) \to \bar{x}(\cdot)$ as $k \to \infty$, which ensures that all the trajectories $x_k(\cdot) \in \mathcal{X}$ belong to the fixed $W^{1,1}([a,b];X)$-neighborhood of the intermediate local minimizer $\bar{x}(\cdot)$ for large $k \in I\!N$. This gives

$$J_{\varepsilon_k}^k[x_k] \ge J_{\varepsilon_k}^k[x_{\varepsilon_k}] > J_{\varepsilon_k}^k[y_k] = \theta_{\varepsilon_k}(y_k) + \sqrt{\varepsilon_k} d(y_k, x_{\varepsilon_k})$$

$$+k \int_a^b \operatorname{dist}\big(\dot{y}_k(t); F(y_k(t),t)\big)\, dt =: k\xi_k \ .$$

Now taking into account (6.54) and the construction of θ_ε, we arrive at

$$k\xi_k < \sqrt{\varepsilon}_k\big(d(x_k, x_{\varepsilon_k}) - d(y_k, x_{\varepsilon_k})\big) + \theta_{\varepsilon_k}(x_k) - \theta_{\varepsilon_k}(y_k) \le 3\gamma\, \xi_k$$

for large k. This is a contradiction, which ends the proof of the claim.

Note that, since U is open in X, the constraint $x(t) \in U$ as $t \in (a,b]$ can be *ignored* from the viewpoint of necessary optimality conditions. Thus we may treat $x_\varepsilon(\cdot)$ is an intermediate local minimizer for the *unconstrained Bolza problem* with finite-valued and *Lipschitzian data*:

$$\text{minimize} \ \ \varphi_\varepsilon\big(x(b)\big) + \int_a^b \vartheta_\varepsilon(x(t), \dot{x}(t), t)\, dt \tag{6.55}$$

over absolutely continuous arcs $x\colon [a,b] \to X$ satisfying $x(a) = x_0$ and lying in a $W^{1,1}$-neighborhood of $\bar{x}(\cdot)$, where the endpoint cost function is given by

$$\varphi_\varepsilon(x) := \operatorname{dist}\big((x, \beta_\varepsilon); \operatorname{epi} \varphi_\Omega\big) \ , \tag{6.56}$$

and where the integrand is

$$\vartheta_\varepsilon(x,v,t) := \alpha\sqrt{1+\ell_F^2(t)}\, \operatorname{dist}\big((x,v); \operatorname{gph} F(t)\big) + \sqrt{\varepsilon}\|v - \dot{x}_\varepsilon(t)\| \ . \tag{6.57}$$

Note that *any intermediate local minimizer* for the unconstrained problem (6.55) provides a *relaxed* intermediate local minimum to this problem. It can be observed from the relaxation result in Theorem 6.11 and its "intermediate" modification given by Ioffe and Rockafellar in Theorem 4 of [616], which is valid in infinite dimensions under the assumptions made. Note also that assumptions (H1), (H2′), and (H3″) ensure that problem (6.55) with the data defined in (6.56) and (6.57) satisfies all the assumptions of Theorem 6.22 except for the compactness of the velocity sets in (P), which in fact is *not needed* in the unconstrained and $W^{1,1}$-bounded framework of (6.55); cf. the proof of Theorem 6.22 and the preceding results it is based on.

We now apply the necessary optimality conditions from Theorem 6.22 to problem (6.55) for any fixed $\varepsilon > 0$. Using the extended Euler-Lagrange inclusion (6.47) with the integrand ϑ_ε in (6.57) and then employing the sum rule from Theorem 2.33(c), find an absolutely continuous adjoint arc $p_\varepsilon\colon[a,b]\to X^*$ satisfying

$$\dot p_\varepsilon(t)\in\operatorname{co}\Big\{u\in X^*\Big|\ \big(u,p_\varepsilon(t)\big)\in\mu(t)\partial\operatorname{dist}\big((x_\varepsilon(t),\dot x_\varepsilon(t));\operatorname{gph}F(t)\big)$$
$$+\sqrt{\varepsilon}\big(0,I\!B^*\big)\Big\}$$

for a.e. $t\in[a,b]$ with $\mu(t):=\alpha\sqrt{1+\ell_F^2(t)}$. Fixed $t\in[a,b]$, consider the two cases regarding $\big(x_\varepsilon(t),\dot x_\varepsilon(t)\big)$:

$$\textbf{(i)}\ \ \dot x_\varepsilon(t)\in F(x_\varepsilon(t),t)\quad\text{and}\quad\textbf{(ii)}\ \ \dot x_\varepsilon(t)\notin F(x_\varepsilon(t),t)\ .$$

In case (i) we use Theorem 1.97 on basic subgradients of the distance function at set points, which gives the *approximate adjoint inclusion*

$$\dot p_\varepsilon(t)\in\operatorname{co}\Big\{u\in X^*\Big|\ \big(u,p_\varepsilon(t)\big)\in N\big((x_\varepsilon(t),\dot x_\varepsilon(t));\operatorname{gph}F(t)\big)+\sqrt{\varepsilon}\big(0,I\!B^*\big)\Big\}\ .$$

Considering case (ii) and employing the first projection formula from Theorem 1.105 for basic subgradients of the distance function at out-of-set points under the Kadec norm structure of X assumed in (a) (see Corollary 1.106 of that theorem), we have the inclusion

$$\partial\operatorname{dist}\big((x_\varepsilon(t),\dot x_\varepsilon(t));\operatorname{gph}F(t)\big)\subset\bigcup_{(x,v)\in\Pi\left((x_\varepsilon(t),\dot x_\varepsilon(t));\operatorname{gph}F(t)\right)}N\big((x,v);\operatorname{gph}F(t)\big)\ .$$

Taking now into account the pointwise convergence $\big(x_\varepsilon(t),\dot x_\varepsilon(t)\big)\to\big(\bar x(t),\dot{\bar x}(t)\big)$ as $\varepsilon\downarrow 0$, one has

$$\partial\operatorname{dist}\big((x_\varepsilon(t),\dot x_\varepsilon(t));\operatorname{gph}F(t)\big)\subset N\big((\widetilde x_\varepsilon,\widetilde v_\varepsilon);\operatorname{gph}F(t)\big)$$

for some $(\widetilde x_\varepsilon,\widetilde v_\varepsilon)\in\operatorname{gph}F(t)$ converging to $\big(\bar x(t),\dot{\bar x}(t)\big)$ as $\varepsilon\downarrow 0$. Thus in case (ii) we get the *approximate adjoint inclusion*

$$\dot p_\varepsilon(t)\in\operatorname{co}\Big\{u\in X^*\Big|\ \big(u,p_\varepsilon(t)\big)\in N\big((\widetilde x_\varepsilon,\widetilde v_\varepsilon);\operatorname{gph}F(t)\big)+\sqrt{\varepsilon}\big(0,I\!B^*\big)\Big\}\ .$$

To derive the extended Euler-Lagrange inclusion (6.49) in problem (P_M), one needs to pass to the limit as $\varepsilon\downarrow 0$ in the approximate adjoint inclusions for $p_\varepsilon(\cdot)$ in both cases (i) and (ii). Since the two approximate adjoint inclusions are similar, we may consider only the first one for definiteness. Observe that

$$\operatorname*{Lim\,sup}_{\varepsilon\downarrow 0}N\big((x_\varepsilon(t),\dot x_\varepsilon(t));\operatorname{gph}F(t)\big)=N\big((\bar x(t),\dot{\bar x}(t));\operatorname{gph}F(t)\big)$$

by the pointwise convergence of $\big(x_\varepsilon(t),\dot x_\varepsilon(t)\big)\to\big(\bar x(t),\dot{\bar x}(t)\big)$ and the robustness property of the basic normal cone from Theorem 3.60 held due to the SNC assumption on F. Note also that the approximate adjoint inclusion for $p_\varepsilon(\cdot)$ can be equivalently rewritten via the *normal* coderivative of F and hence, by the strong coderivative normality assumption of the theorem, in terms of the *mixed* coderivative $D^*_M F$. Proceeding similarly to the proof of Theorem 6.21 with the use of the mixed coderivative condition for the Lipschitzian continuity from Theorem 1.44 as well as the classical Dunford and Mazur theorems as above, we surely arrive at (6.49).

Consider next the transversality inclusion for $p_\varepsilon(b)$ in problem (6.55) with the cost function φ_ε in (6.56). Employing the transversality condition (6.45) from Theorem 6.22 in this setting, we have just the first terms in (6.45), where $\lambda = 1$ and $\varphi(x_a, x_b) = \varphi_\varepsilon(x_b)$. The crucial condition

$$\operatorname{dist}\big((x_\varepsilon(b), \beta_\varepsilon); \operatorname{epi}\varphi_\Omega\big) > 0$$

ensures that $(x_\varepsilon(b), \beta_\varepsilon) \notin \operatorname{epi}\varphi_\Omega$ for all $\varepsilon > 0$ sufficiently small. Employing again Theorem 1.105/Corollary 1.106, one has

$$\big(-p_\varepsilon(b), -\lambda_\varepsilon\big) \in \bigcup_{(x,b)\in\Pi((x_\varepsilon,\beta_\varepsilon);\,\operatorname{epi}\varphi_\Omega)} N\big((x,\beta); \operatorname{epi}\varphi_\Omega\big)$$

with some $\lambda_\varepsilon \geq 0$. Moreover, we can put $\lambda_\varepsilon + \|p_\varepsilon(b)\| = 1$ due to the SNEC property of φ_Ω at $\bar{x}(b)$ and hence *around* this point; see Remark 1.27(ii). Passing to the limit as $\varepsilon \downarrow 0$ and taking into account the robustness result of Theorem 3.60, we arrive at the desired transversality inclusion (6.51) with $\lambda \geq 0$ by putting $\varepsilon \downarrow 0$. The nontriviality condition $\lambda + \|p(b)\| = 1$ follows from the one for $\big(\lambda_\varepsilon, p_\varepsilon(b)\big)$ due to the SNEC property of φ_Ω that surely holds if Ω is SNC at $\bar{x}(b)$ and φ is Lipschitz continuous around this point. The latter is an easy consequence of Theorem 3.90, which ensures even the stronger SNC property of φ at $\bar{x}(b)$. The equivalence between the transversality inclusions (6.51) and (6.52) whenever φ is locally Lipschitzian around $\bar{x}(b)$ relative to Ω follows from Lemma 5.23. Note that inclusion (6.52) further implies

$$-p(b) \in \lambda\partial\varphi\big(\bar{x}(b)\big) + N(\bar{x}(b); \Omega)$$

for Lipschitz continuous cost functions.

The above proof justifies the extended Euler-Lagrange and transversality conditions in the theorem for arbitrary intermediate local minimizers to problem (P_M) with *no relaxation*. In this general nonconvex setting the extended Euler-Lagrange inclusion (6.49) doesn't automatically imply the maximum condition (6.50). To establish the latter condition supplementing (6.49) and (6.51), we follow the proof of Theorem 7.4.1 in Vinter [1289] given for a Mayer problem of the type (P_M) involving nonconvex differential inclusions in finite-dimensional spaces. The proof of the latter theorem is based on reducing the constrained Mayer problem for nonconvex differential inclusions to an unconstrained Bolza (finite Lagrangian) problem, which in turn is reduced to a problem of optimal control with *smooth dynamics* admitting a direct way to derive the maximum principle; cf. also Sect. 6.3. One can check that the tools of infinite-dimensional variational analysis developed above and the assumptions made allow us to extend the given proof to the case of reflexive and separable spaces under consideration. In this way we establish the maximum condition (6.50) in addition to the other necessary optimality conditions of the theorem and complete the proof. △

Remark 6.28 (necessary conditions for nonconvex differential inclusions under weakened assumptions). Some assumptions of Theorem 6.27, particularly those on the Kadec norm and on the weakly closed graph and epigraph in (a)–(c), can be relaxed under a certain modification of the proof. This concerns the application of necessary optimality conditions from Theorem 6.22 to the unconstrained Bolza problem (6.55). The latter conditions are expressed in terms of the basic/limiting constructions and then require the usage of the projection result from Corollary 1.106 to efficiently estimate basic subgradients of the distance function at out-of-set points under the mentioned assumptions. To avoid these extra requirements, one may apply first a *fuzzy* discrete approximation version of Theorem 6.27 to the unconstrained problem (6.55), involving Fréchet normals and subgradients as in the proof of Theorem 6.21, and then pass to the limit as $N \to \infty$ and $\varepsilon \downarrow 0$. In this way, the realization of which is more involved, we replace the usage of the distance function result of Corollary 1.106 via basic subgradients by its Fréchet subgradient counterpart from Theorem 1.103 that holds under milder assumptions.

Observe that the SNC and strong coderivative normality properties of F are automatic when X is *finite-dimensional*, which also implies the SNEC property of the extended endpoint function φ_Ω assumed in Theorem 6.27. Furthermore, the latter property is *not needed* (actually it holds *automatically* under qualification conditions of the Mangasarian-Fromovitz type) in the general infinite-dimensional case of the theorem if the cost function is locally Lipschitzian and the endpoint constraint set given via a *finite number* of equalities and inequalities defined by locally Lipschitzian functions.

Corollary 6.29 (transversality conditions for differential inclusions with equality and inequality constraints). *Let $\bar{x}(\cdot)$ be an intermediate local minimizer for the Mayer problem (P_M) with the endpoint constraint set*

$$\Omega := \left\{ x \in X \,\middle|\, \varphi_i(x) \le 0,\ i = 1, \ldots, m;\ \ \varphi_i(x) = 0,\ i = m+1, \ldots, m+r \right\},$$

where each φ_i is locally Lipschitzian around $\bar{x}(b)$ together with the cost function $\varphi_0 := \varphi$. Suppose that all the assumptions of Theorem 6.27 hold except the SNEC property of the extended endpoint function φ_Ω. Then there are nonnegative multipliers $(\lambda_0, \ldots, \lambda_{m+r}) \ne 0$ and an absolutely continuous adjoint arc $p\colon [a,b] \to X^*$ satisfying the Euler-Lagrange and maximum conditions (6.49) *and* (6.50) *together with the complementary slackness condition*

$$\lambda_i \varphi_i\big(\bar{x}(b)\big) = 0 \ \ \textit{for} \ \ i = 1, \ldots, m$$

and the transversality inclusion

$$-p(b) \in \sum_{i=0}^{m} \lambda_i \partial\varphi_i\big(\bar{x}(b)\big) + \sum_{i=m+1}^{m+r} \lambda_i \Big[\partial\varphi_i\big(\bar{x}(b)\big) \bigcup \partial\big(-\varphi_i\big)\big(\bar{x}(b)\big)\Big] .$$

If furthermore all φ_i, $i = 0, \ldots, m+r$, are strictly differentiable at $\bar{x}(b)$, then there are multipliers $(\lambda_0, \ldots, \lambda_{m+r}) \neq 0$ with $\lambda_i \geq 0$ as $i = 0, \ldots, m$ and an adjoint arc $p\colon [a,b] \to X^*$ satisfying

$$-p(b) = \sum_{i=0}^{m+r} \lambda_i \nabla\varphi_i\big(\bar{x}(b)\big)$$

together with the above Euler-Lagrange, Weierstrass-Pontryagin, and complementary slackness conditions.

Proof. It follows from (6.52) with $\lambda := \lambda_0$ that

$$-p(b) \in \lambda_0 \partial\varphi_0\big(\bar{x}(b)\big) + N\big(\bar{x}(b); \Omega\big) .$$

Moreover, φ_Ω is SNEC at $\bar{x}(b)$ provided that Ω is SNC at this point; see Corollary 3.89. Then we proceed similarly to the proof of Corollary 6.24 and complete the proof of this corollary. △

6.2.2 Discussion and Examples

In this subsection we consider certain generalizations and variants of the above results, discuss some interrelations and examples. First note that the comprehensive generalized differential and SNC calculi developed in Chap. 3 allow us to derive various consequences and extensions of Theorem 6.27 in the case of *operator* endpoint constraints given by

$$x(b) \in F^{-1}(\Theta) \cap \Omega$$

with $F\colon X \rightrightarrows Y$ and $\Theta \subset Y$; cf. Sect. 5.1 for problems of mathematical programming. Let us discuss in more details some other important issues related to obtained necessary optimality conditions for differential inclusions.

Remark 6.30 (upper subdifferential transversality conditions). Suppose in addition to the assumptions of Theorem 6.21 that the space X admits a $\mathcal{C}^1$ Lipschitzian bump function; this is automatic under the reflexivity assumption on X in Theorems 6.22 and 6.27. Then employing the results of Sects. 6.1 and 6.2 together with the *smooth variational description* of Fréchet subgradients in Theorem 1.88(ii), we derive necessary optimality conditions for problems (P) and (P_M), as well as for their discrete-time counterparts, with transversality relations expressed via *upper subgradients* of functions that describe the objective and inequality constraints. This can be done by reducing them to the case of *smooth* functions describing the objective and inequality constraints; cf. the proof of Theorem 5.19 for nondifferentiable programming. Considering, in particular, the Mayer problem of minimizing $\varphi_0(x(b))$ over absolutely continuous trajectories $x\colon [a,b] \to X$ for the differential inclusion (6.48) subject to the endpoint constraints

$$\varphi_i\big(x(b)\big) \le 0, \quad i = 1,\ldots,m\ ,$$

under the assumptions made on F and X in Theorem 6.27 and *no* assumptions on φ_i, we have the following necessary optimality conditions for an *intermediate* local minimizer $\bar{x}(\cdot)$: given *every* set of Fréchet *upper subgradients* $x_i^* \in \widehat{\partial}^+\varphi_i\big(\bar{x}(b)\big)$, $i = 0,\ldots,m$, there are multipliers

$$(\lambda_0,\ldots,\lambda_m) \ne 0 \;\text{ with }\; \lambda_i \ge 0 \;\text{ for all }\; i = 0,\ldots,m$$

and an absolutely continuous mapping $p\colon [a,b] \to X^*$ satisfying the Euler-Lagrange and maximum conditions (6.49) and (6.50) together with

$$\lambda_i\varphi_i\big(\bar{x}(b)\big) = 0 \;\text{ for }\; i = 1,\ldots,m \;\text{ and}$$
$$p(b) + \sum_{i=0}^{m} \lambda_i x_i^* = 0\ .$$

To justify these conditions via the above arguments, it remains to check the SNEC property of the extended endpoint function φ_Ω in Theorem 6.27 with

$$\Omega := \big\{x \in X\big|\ \varphi_i(x) \le 0,\ \ i = 1,\ldots,m\big\}$$

and the *smooth* data φ, φ_i. It follows from Corollary 3.87 ensuring the SNC property of the classical constraint set in nonlinear programming; cf. the proof of Corollaries 6.24 and 6.29.

Remark 6.31 (necessary optimality conditions for multiobjective control problems). The methods and results developed above can be extended to *multiobjective optimization* problems governed by differential inclusions. Given a mapping $f\colon X \to Z$ and a subset $\Theta \subset Z$ of a Banach space with $0 \in \Theta$, consider a multiobjective counterpart of the above Mayer problem (P_M), where the *generalized order* (f,Θ)*-optimality* of a trajectory $\bar{x}(\cdot)$ for (6.48) subject to $x(b) \in \Omega$ is understood in the sense that there is a sequence $\{z_k\} \subset Z$ with $z_k \to 0$ as $k \to \infty$ such that

$$f\big(x(b)\big) - f\big(\bar{x}(b)\big) \notin \Theta - z_k, \quad k \in I\!N\ ,$$

for any feasible trajectory $x(\cdot)$ from a $W^{1,1}\big([a,b];X\big)$-neighborhood of $\bar{x}(\cdot)$; cf. Definition 5.53 and the related discussions in Subsect. 5.3.1. Let

$$\mathcal{E}(f,\Omega,\Theta) = \big\{(x,z) \in X \times Z\big|\ f(x) - z \in \Theta,\ \ x \in \Omega\big\}$$

be the "generalized epigraph" of the restrictive mapping $f_\Omega = f + \Delta(\cdot;\Omega)$ with respect to the ordering set Θ. Taking a sequence $z_k \to 0$ from the above definition of the (f,Θ)-optimality for $\bar{x}(\cdot)$, we define the functions

$$\theta_k(x) := \operatorname{dist}\big((x, f(\bar{x}) - z_k); \mathcal{E}(f,\Omega,\Theta)\big), \quad k \in I\!N .$$

and proceed similarly to the proof of Theorem 6.27 with the replacement of $\theta_\beta(x)$ therein by the sequence of $\theta_k(x)$. In this way we arrive at necessary optimality conditions in the multiobjective control problem under consideration that are different from the ones in Theorem 6.27 only in transversality relations. Namely, suppose in addition to the assumptions on X and F in Theorem 6.27 that the space Z is WCG and Asplund and that the generalized epigraphical set $\mathcal{E}(f,\Omega,\Theta)$ is locally closed around $(\bar{x},\bar{z})$ and SNC at this point with $\bar{z} := f(\bar{x})$. Then there are an adjoint arc $p\colon [a,b] \to X^*$ and an adjoint vector $z^* \in N(0;\Theta)$, not both zero, satisfying the extended Euler-Lagrange inclusion (6.49), the Weierstrass-Pontryagin maximum condition (6.50), and the *transversality inclusion*

$$\big(-p(b), -z^*\big) \in N\big((\bar{x}(b),\bar{z}); \mathcal{E}(f,\Omega,\Theta)\big) .$$

The latter inclusion is equivalent, by Lemma 5.23, to

$$-p(b) \in \partial\langle z^*, f_\Omega\rangle(\bar{x}), \quad z^* \in N(0;\Theta)$$

if the mapping f is Lipschitz continuous around $\bar{x}$ relative to Ω and strongly coderivatively normal at this point, and if the sets Ω and Θ are locally closed around the points $\bar{x}$ and 0, respectively. Note that multiobjective optimal control problems of the above type but with respect to *closed preference relations* can be treated similarly; cf. Subsect. 5.3.4. In this way we can also derive necessary optimality conditions for multiobjective (as well as of the Mayer and Bolza types) optimal control problems governed by differential inclusions with *equilibrium constraints*, which are dynamic counterparts of MPEC and EPEC problems studied in Sect. 5.2 and Subsect. 5.3.5.

Remark 6.32 (Hamiltonian inclusions). When $X = I\!R^n$, an additional optimality condition can be obtained for *relaxed* intermediate local minimizers to problem (P_M) (as well as to (P) and the counterparts of these problems discussed in the preceding remarks), which is expressed via basic subgradients to the *Hamiltonian* function defined by

$$\mathcal{H}(x,p,t) := \sup\big\{\langle p, v\rangle \,\big|\, v \in F(x,t)\big\} .$$

It follows from Rockafellar's dualization theorem ([1162, Theorem 3.3]) that

$$\operatorname{co}\Big\{u \in I\!R^n \,\Big|\, (u,p) \in N\big((\bar{x},\bar{v}); \operatorname{gph} F\big)\Big\} = \operatorname{co}\Big\{u \in I\!R^n \,\Big|\, (-u,\bar{v}) \in \partial\mathcal{H}(\bar{x},p)\Big\}$$

if F is *convex-valued* and satisfies some requirements around $(\bar{x},\bar{v})$ that are automatic under the assumptions made on F in (H1); dependence on t is not important and is thus suppressed. The proof of the latter *dualization relationship* is essentially finite-dimensional; cf. also the proofs in Ioffe [604, Theorem 4] and in Vinter [1289, Theorem 7.6.5]. Since the Hamiltonian of the convexified inclusion (6.18) is obviously agrees with the original one $\mathcal{H}(x,p,t)$, we deduce from the above duality relation that the Euler-Lagrange inclusion (6.49) in Theorem 6.27 implies the *extended Hamiltonian inclusion*

$$\dot p(t)\in\operatorname{co}\Big\{u\in I\!R^n\Big|\ \big(-u,\dot{\bar x}(t)\big)\in\partial\mathcal{H}\big(\bar x(t),p(t),t\big)\Big\}\ \text{ a.e }\ t\in[a,b] \tag{6.58}$$

as a *necessary optimality condition* for *relaxed minimizers* in the case of finite-dimensional state spaces. Moreover, the Euler-Lagrange inclusion (6.49) and the Hamiltonian inclusion (6.58) are *equivalent* for problems (P_M) with the *convex* velocity sets $F(x,t)$. Note that (6.58) is a refined Hamiltonian inclusion involving a *partial convexification* of the basic subdifferential $\partial\mathcal{H}(\bar x(t),p(t),t)$, which clearly supersedes the *fully convexified* one

$$\big(-\dot p(t),\dot{\bar x}(t)\big)\in\operatorname{co}\partial\mathcal{H}(\bar x(t),p(t),t)\ \text{ a.e. }\ t\in[a,b] \tag{6.59}$$

involving Clarke's generalized gradient $\partial_C\mathcal{H}(\bar x(t),p(t),t)=\operatorname{co}\partial\mathcal{H}(\bar x(t),p(t),t)$ of the Hamiltonian with respect to (x,p). It is worth observing that both Hamiltonian inclusions (6.58) and (6.59) are *invariant* with respect to the convexification of $F(x,t)$, which is *not* the case for the extended Euler-Lagrange inclusion (6.49).

Remark 6.33 (local controllability). The approach developed in the preceding subsection for necessary optimality conditions allows us to study also related issues concerning the so-called *local controllability* of nonconvex differential inclusions in the case of *finite-dimensional* spaces. Given $x_0\in X$, we denote by $\mathcal{R}(x_0)$ the *reachable set* for the differential inclusion (6.48), which is the set of all $z\in X$ such that $x(b)=z$ for some arc $x\colon[a,b]\to X$ admissible to (6.48). The meaning of local controllability is to derive efficient conditions for *boundary trajectories* of the differential inclusion (6.48), in a certain generalized sense. To be more precise, we consider a mapping $g\colon X\to X$ locally Lipschitzian mapping around $\bar x(b)$ and a trajectory $\bar x\colon[a,b]\to X$ for (6.48) such that $g\big(\bar x(b)\big)\in\operatorname{bd}\mathcal{R}(x_0)$. Then assuming that $X=I\!R^n$ in addition to (H1) and (H2′), we find a vector $x^*\in I\!R^n$ with $\|x^*\|=1$ and an adjoint arc $p(\cdot)$ satisfying the extended Euler-Lagrange inclusion (6.49) with the *boundary/transversality condition*

$$-p(b)\in\partial\langle x^*,g\rangle\big(\bar x(b)\big) \tag{6.60}$$

and the Weierstrass-Pontryagin maximum condition (6.50). Moreover, if the reachable set $\mathcal{R}(x_0)$ is *locally closed* around $\bar x(b)$, then the extended Hamiltonian inclusion (6.58) is also satisfied.

To justify the Euler-Lagrange and maximum conditions (6.49) and (6.50) with the new transversality condition (6.60), we follow the proof of Theorem 6.27 and, given any $\varepsilon>0$, find a vector $c_\varepsilon\in I\!R^n$ and a trajectory $x_\varepsilon(\cdot)$ for (6.48) such that $\|g\big(x_\varepsilon(b)\big)-c_\varepsilon\|>0$,

$$c_\varepsilon \to g(\bar{x}(b)),\ \ x_\varepsilon(\cdot) \to \bar{x}(\cdot)\ \text{ strongly in }\ W^{1,1}\big([a,b];I\!R^n\big)\ \text{ as }\ \varepsilon \downarrow 0\,,$$

and $x_\varepsilon(\cdot)$ is an unconditional *strong* local minimizer for problem (6.55) with the same integrand (6.57) and the endpoint function

$$\varphi_\varepsilon(z) := \|g(z) - c_\varepsilon\|\,.$$

Then we proceed as in the proof of Theorem 6.27 with the only difference that now we need to compute the basic subdifferential of the new function $\varphi_\varepsilon(\cdot)$ at the point $\bar{x}_\varepsilon(b)$ with $\|g(x_\varepsilon(b)) - c_\varepsilon\| > 0$. Using the subdifferential chain rule of Corollary 3.43 and then passing to the limit as $\varepsilon \downarrow 0$ while taking into account the *compactness of the unit sphere* in $I\!R^n$, we arrive at the transversality condition (6.60) that supplements (6.49) and (6.50). To justify the extended Hamiltonian inclusion (6.58), we observe that the assumptions made ensure the closedness of the reachable set $\widetilde{\mathcal{R}}(x_0)$ generated by the *convexified* differential inclusion

$$\dot{x}(t) \in \text{co}\, F(x(t),t)\ \text{ a.e. }\ t \in [a,b], \quad x(a) = x_0$$

and the density of $\mathcal{R}(x_0)$ in $\widetilde{\mathcal{R}}(x_0)$; cf. Theorem 6.11. Thus the local closedness assumption on $\mathcal{R}(x_0)$ yields that $\bar{x}(b)$ is a boundary point of $\widetilde{\mathcal{R}}(x_0)$, and so (6.58) follows from the discussion in Remark 6.32.

Note that the finite dimensionality of the state space X is needed in the above proof for local controllability to guarantee the compactness of the dual unit sphere in the weak* topology of X^*, which never holds in infinite dimensions due to the fundamental Josefson-Nissenzweig theorem. Such a difference with the infinite-dimensional setting of Theorem 6.27 is due to the fact that in the proof of the latter theorem we actually applied the *exact extremal principle* to the local extremal system of sets $\mathcal{R}(x_0) \times \{\varphi(\bar{x}(b))\}$ and $\text{epi}\,\varphi_\Omega$ (in the notation of Theorem 6.27) with the SNC assumption imposed on the second set in the extremal system. In the setting of local controllability we deal with the local extremal system of sets $\mathcal{R}(x_0)$ and $\{\bar{x}(b)\}$, where the second singleton set is *never SNC* in infinite dimensions. Observe however that we didn't explore in the proof of Theorem 6.27, as well as in the framework of local controllability, the possibility of imposing a *SNC requirement on the reachable set* $\mathcal{R}(x_0)$, which may lead to *alternative* assumptions ensuring the fulfillment of necessary optimality and local controllability conditions in infinite dimensions; cf. the result and discussion in Remark 6.25(i).

To conclude this section, we present some examples illustrating the results obtained and the relationships between them. First let us show that the *partial convexification* can *not* be avoided in both extended Euler-Lagrange and Hamiltonian inclusion (6.49) and (6.58).

Example 6.34 (partial convexification is essential in Euler-Lagrange and Hamiltonian optimality conditions). *There is a two-dimensional*

Mayer problem of minimizing a linear function over absolutely continuous trajectories of a convex-valued differential inclusion with no endpoint constraints such that analogs of the Euler-Lagrange inclusion (6.49) *and the Hamiltonian inclusion* (6.58) *with no* (*partial*) *convexification* "co" *therein don't hold as necessary optimality conditions.*

Proof. Consider the following Mayer problem for a convex-valued differential inclusion with $x = (x_1, x_2) \in I\!R^2$:

$$\begin{cases} \text{minimize} \;\; J[x] := x_2(1) \;\; \text{subject to} \\ \dot{x}_1 \in [-\nu, \nu], \quad x_1(0) = 0 \;, \\ \dot{x}_2 = |x_1|, \quad x_2(0) = 0 \;, \\ \;\text{for a.e. } \; t \in [0,1] \;\; \text{with some} \;\; \nu > 0 \;. \end{cases}$$

It is easy to see that $\bar{x}(t) \equiv 0$ is the only optimal solution to this problem, and that an analog of the Euler-Lagrange inclusion (6.49) for the adjoint arc $(p(t), -1) \in I\!R^2$ without "co" therein gives, along this $\bar{x}(\cdot)$, the relation

$$\dot{p}(t) \in \big\{ -1, 1 \big\} \;\; \text{a.e. } t \in [0,1]$$

with the transversality condition $p(1) = 0$. Furthermore, the maximum condition, implied by the Euler-Lagrange inclusion in this case due to Theorem 1.34, takes the form

$$\langle p(t), \dot{\bar{x}}(t) \rangle = \max_{v \in [-\nu, \nu]} \langle p(t), v \rangle \;\; \text{a.e.} \;\; t \in [0,1] \;,$$

which yields that $p(t) \equiv 0$; a contradiction. Since $\mathcal{H}(p, x) = \nu \,\text{sign}\, p - |x_1|$, the Hamiltonian inclusion

$$\big(-\dot{p}(t), \dot{\bar{x}}(t) \big) \in \partial \mathcal{H}\big(\bar{x}(t), p(t) \big) \;\; \text{a.e.} \;\; t \in [0,1] \;,$$

which is (6.58) with no "co" therein, leads to the same relations as above and hence doesn't hold as a necessary optimality condition. △

The next two examples illustrate relationships between the extended Euler-Lagrange inclusion (6.49) and the extended Hamiltonian inclusion (6.58) with the (fully) convexified Hamiltonian inclusion (6.59).

Example 6.35 (extended Euler-Lagrange inclusion is strictly better than convexified Hamiltonian inclusion). *There is a compact-valued and convex-valued multifunction $F\colon I\!R^2 \rightrightarrows I\!R^2$, which is Lipschitz continuous on $I\!R^2$ and such that*

$$(-w, v) \in \text{co}\, \partial \mathcal{H}(x, p) \;\; \textit{but} \;\; w \notin \text{co} \big\{ u \in I\!R^2 \big| \; u \in D^* F(x, v)(-p) \big\}$$

for some points x, v, w, p in the plane.

Proof. Define $F\colon I\!R^2 \rightrightarrows I\!R^2$ by

$$F(x_1,x_2) := \big\{(\tau,\tau\,|x_1|+\nu)\in I\!R^2\big|\ \tau\in[-1,1],\ \nu\in[0,\mu]\big\}\ \text{ with some }\ \mu>0\,,$$

where the sets $F(x)$ are parallelograms in the plane for all $x=(x_1,x_2)\in I\!R^2$. The corresponding Hamiltonian is

$$\mathcal{H}(x_1,x_2,p_1,p_2)=\big|p_1+p_2|x_1|\big|+\max\big\{p_2,0\big\}\,.$$

Considering the points $x=(0,0)$, $v=(0,0)$, and $p=(0,-1)$, we see that the corresponding set $F(x)$ is the rectangle $[-1,1]\times[0,\mu]$, and that p is an outward normal vector to this set at the boundary point v. The crucial feature of this example is that the hyperplane $x_2=0$ supporting the set $F(x)$ at v intersects this set in *more than one point*. In other words, the maximum of $\langle p,v\rangle$ over $v\in F(x)$ is attained at *infinitely many points*. The basic subdifferential of $\mathcal{H}$ at the point $(0,0,0,-1)$ and its convexification (Clarke's generalized gradient) are actually calculated in Example 2.49; thus

$$\operatorname{co}\partial\mathcal{H}(0,0,0,-1)=[-1,1]\times\{0\}\times[-1,1]\times\{0\}\subset I\!R^4\,.$$

Taking $w=(-1,0)$, one has $(-w,v)\in\operatorname{co}\partial\mathcal{H}(0,0,0,-1)$. Let us show that

$$(w,p)=(-1,0,0,-1)\notin\operatorname{clco}N\big((x,v);\operatorname{gph}F\big)\,,$$

which definitely justifies the claim of this example.

To proceed, we note that, up to a permutation of the coordinates, the graph of F can be represented as

$$\operatorname{gph}F=E\times I\!R\ \text{ with }\ E:=\big\{(x_1,\tau,|x_1|\tau+\nu)\in I\!R^3\big|\ \tau\in[-1,1],\ \nu\in[0,\mu]\big\}\,,$$

where the set E obviously coincides around the point $(0,0,0)$ with the *epigraph* of the Lipschitzian function $\varphi\colon I\!R^2\to I\!R$ defined by $\varphi(y,\tau):=\tau|y|$. It is easy to see that

$$\operatorname{co}\partial\varphi(0,0)=\partial\varphi(0,0)=\big\{(0,0)\big\}\,.$$

One therefore calculates

$$N\big((0,0,\varphi(0,0));\operatorname{epi}\varphi\big)=\bigcup_{\lambda\ge 0}\lambda\big[\partial\varphi(0,0)\times\{-1\}\big]=\big\{(0,0)\big\}\times(-\infty,0]\,,$$

and hence we deduce that

$$\operatorname{clco}N\big((0,0,0,0);\operatorname{gph}F\big)=\big\{(0,0,0)\big\}\times(-\infty,0]\,.$$

In particular, the latter cone doesn't contain the point $(w,p)=(-1,0,0,-1)$, even though $(-w,v)\in\operatorname{co}\partial\mathcal{H}(x,p)$. △

The last example shows that the extended/refined Hamiltonian condition (6.58) *strictly supersedes* the fully convexified one (6.59) in both settings of convex-valued and nonconvex-valued differential inclusions.

Example 6.36 (partially convexified Hamiltonian condition strictly improves its fully convexified counterpart). *There is a set-valued mapping* $F\colon I\!R^n \rightrightarrows I\!R^n$ *in the form* $F(x) = g(x)S$, *where* $S \subset I\!R^n$ *is a compact set and where* $g(x)$, *for each* x, *is a linear isomorphism of* $I\!R^n$ *depending continuously on* x, *such that for some* $(\bar{x}, \bar{v}, \bar{p})$ *one has*

$$\mathrm{co}\,\big\{u \in I\!R^n \big|\ (u, \bar{v}) \in \partial\mathcal{H}(\bar{x}, \bar{p})\big\} \neq \big\{u \in I\!R^n \big|\ (u, \bar{v}) \in \mathrm{co}\,\partial\mathcal{H}(\bar{x}, \bar{p})\big\}\ .$$

Proof. If F is given in the above form, then its Hamiltonian is calculated by

$$\mathcal{H}(x, p) = \sup\big\{\langle p, v\rangle \big|\ v \in g(x)S\big\} = \sup\{\langle p, g(x)s\rangle \big|\ s \in S\big\} =: \delta^*\big(g^*(x)p; S\big),$$

where $\delta^*(\cdot; S)$ stands for the standard support function of the set S. Since S is bounded, its support function is continuous. Denote

$$\psi_s(x, p) := \langle s, g^*(x)p\rangle = \langle g(x)s, p\rangle$$

and suppose that $g(\cdot)$ is Lipschitz continuous. Employing the scalarization formula and taking into account the structure of ψ, we have

$$\partial\mathcal{H}(\bar{x}, \bar{p}) = \bigcup_{s \in \partial\delta^*(0;S)} \partial\psi_s(\bar{x}, \bar{p})$$

at any given point $(\bar{x}, \bar{p})$. The linearity of ψ in p yields that

$$\partial\psi_s(\bar{x}, \bar{p}) = \big(\partial_x\psi_s(\bar{x}, \bar{p}), g(\bar{x})s\big)\ .$$

Therefore the inclusion $(u, 0) \in \partial\psi_s(\bar{x}, \bar{p})$ implies that $s = 0$ and thus $u = 0$.

Based on the above discussion, we need to find a set S, a Lipschitz continuous family of linear isomorphisms $g(x)$ of $I\!R^n$, and a point $(\bar{x}, \bar{p}) \in I\!R^n \times I\!R^n$ such that $0 \in S$ and $\mathrm{co}\,\partial\mathcal{H}(\bar{x}, \bar{p})$ contains a pair $(u, 0)$ with $u \neq 0$. In particular, it can be done as follows for $n = 2$. Let

$$S := \big\{(y_1, y_2) \in I\!R^2 \big|\ |y_1| \le 1,\ y_2 = 0\big\}, \quad g^*(x) := \begin{pmatrix} 1 & |x_1| \\ 1 & 1 \end{pmatrix}\ ,$$

$\bar{x} := (0, 0)$, and $\bar{p} := (0, 1)$. Then

$$\delta^*\big((w_1, w_2); S\big) = w_1 \ \text{ and } \ \mathcal{H}(x, p) = \big|p_1 + p_2|x_1|\big|\ .$$

One can directly calculate (cf. Example 2.49) that the set $\mathrm{co}\,\partial\mathcal{H}(\bar{x}, \bar{p})$ is the convex hull of the following four points: $(1, 0, 1, 0)$, $(-1, 0 - 1, 0)$, $(1, 0, -1, 0)$, and $(-1, 0, 1, 0)$. Thus

$$\big\{u \in I\!R \big|\ (u, 0) \in \mathrm{co}\,\partial\mathcal{H}(\bar{x}, \bar{p})\big\} = [-1, 1]\ ,$$

which justifies the claim of this example. △

6.3 Maximum Principle for Continuous-Time Systems with Smooth Dynamics

In this section we study optimal control problems governed by ordinary differential equations in infinite-dimensional spaces that explicitly involve constrained control inputs $u(\cdot)$ as follows:

$$\dot{x} = f(x,u,t), \quad u(t) \in U \quad \text{a.e. } t \in [a,b] , \tag{6.61}$$

where $f\colon X \times U \times [a,b] \to X$ with a Banach state space X and a metric control space U. Although control systems of this type can be reduced to differential inclusions $\dot{x} \in F(x,t)$ with $F(x,t) := f(x,U,t)$, the explicit control input in (6.61) with the control region U independent of x (it may depend on t) allows us to develop efficient methods of studying such dynamic systems that take into account their specific features.

Throughout the section we assume that system (6.61) is of *smooth dynamics*, which means that the velocity mapping f is *continuously differentiable* ($\mathcal{C}^1$) with respect to the state variable x around an optimal solution to be considered. Despite this assumption, the control system (6.61) and optimization problems over its feasible controls and trajectories intrinsically involve *nonsmoothness* due to the control geometric constraints $u(t) \in U$ a.e. $t \in [a,b]$ defined by control sets U of a general nature. For instance, it is the case of the simplest/classical optimal control problems with $U = \{0,1\}$.

In this section the main attention is paid to the Mayer-type control problem for systems (6.61) of smooth dynamics subject to *finitely many* endpoint constraints given by equalities and inequalities with functions merely *Fréchet differentiable* (possibly not strictly) *at* points of minima. Our goal is to derive necessary optimality conditions in the form of the *Pontryagin maximum principle* (PMP) for such problems in *general Banach spaces*, with *no* additional assumptions on the reflexivity and separability of X as well as on the sequential normal compactness and strong coderivative normality of $F(x,t) = f(x,U,t)$ imposed in Theorem 6.27 of the preceding section. The technique used for this purpose is different from those employed in Sects. 6.1 and 6.2; it goes back to the classical approach in optimal control theory involving *needle variations* of optimal controls. We also derive enhanced results of the maximum principle type with *upper subdifferential* transversality conditions in the case of nondifferentiable cost and inequality constraint functions. Such conditions are obtained without imposing *any smoothness* assumptions on the state space in question needed for the corresponding necessary optimality conditions derived above in both mathematical programming and dynamic optimization settings; cf. Theorem 5.19 and Remark 6.30. Thus the results of this section, which essentially exploit the specific structure of smooth control systems (6.61) and the imposed endpoint constraints, are generally independent of those obtained in Sects. 6.1 and 6.2.

This section is organized as follows. Subsect. 6.3.1 contains the formulation of the main assumptions and results as well as the derivation of the maximum principle with upper subdifferential transversality conditions from the one with Fréchet differentiable endpoint functions. We also discuss possible extensions of the maximum principle to control problems with intermediate state constraints as well as to some classes of time-delay systems. Subsection 6.3.2 is devoted to the proof of the PMP for free-endpoint control problems in Banach spaces, which is substantially simpler than that for problems with endpoint constraints. Subsection 6.3.3 deals with optimal control problems involving endpoint constraints of the inequality type. Finally, in Subsect. 6.3.4 we derive, with the use of the Brouwer fixed-point theorem, transversality conditions in the case of equality constraints given by continuous functions that are just differentiable at optimal endpoints.

6.3.1 Formulation and Discussion of Main Results

It is more simple and convenient (and in fact does't much restrict the generality) to formulate and then to prove the main results of this section for the case of control systems (6.61) with a *fixed left endpoint* $x(a) = x_0$; we discuss various extensions of the main results in the end of this subsection.

Denote by $\mathcal{A}$ the collection of *admissible* control-trajectory pairs $\{u(\cdot), x(\cdot)\}$ generated by *measurable* controls $u(\cdot)$ satisfying the *pointwise* constraints $u(t) \in U$ for a.e. $t \in [a, b]$ and the corresponding solutions $x(\cdot)$ to (6.61) with $x(a) = x_0$ defined by

$$x(t) = x_0 + \int_a^t f(x(s), u(s), s)\, ds \quad \text{for all} \quad t \in [a, b] , \tag{6.62}$$

where the integral is understood in the Bochner sense; cf. Definition 6.1. As is well known, any solution to (6.62) is absolutely continuous on $[a, b]$. Moreover, it is a.e. differentiable on $[a, b]$ and satisfies the differential equation (6.61) for a.e. $t \in [a, b]$ provided that X has the Radon-Nikodým property (see Subsect. 6.1.1), which is *not* assumed here. What we need in this section is the integral representation (6.62), which is taken as the definition of admissible solutions/arcs to the differential equation (6.61) in Banach spaces.

Given real-valued functions φ_i, $i = 0, \ldots, m + r$, on the state space X, we now formulate the optimal control problem studied below:

$$\text{minimize} \quad J[u, x] = \varphi_0\big(x(b)\big) \quad \text{over} \quad (u, x) \in \mathcal{A} \tag{6.63}$$

subject to the endpoint constraints

$$\varphi_i\big(x(b)\big) \le 0 \quad \text{for} \quad i = 1, \ldots, m , \tag{6.64}$$

$$\varphi_i\big(x(b)\big) = 0 \quad \text{for} \quad i = m + 1, \ldots, m + r . \tag{6.65}$$

Admissible solutions $(u, x) \in \mathcal{A}$ satisfying the endpoint constraints (6.64) and (6.65) are called *feasible solutions* to problem (6.63)–(6.65). So we don't distinguish between admissible and feasible solutions for problems with *free endpoints*, i.e., with no endpoint constraints (6.64) and (6.65). We always assume that the set of feasible solutions to (6.63)–(6.65) is *not empty.*

A feasible solution $\{\bar{u}(\cdot), \bar{x}(\cdot)\}$ is *optimal* to (6.63)–(6.65) if

$$J[\bar{u}, \bar{x}] \le J[u, x] \text{ for all } (u, x) \in \mathcal{A}$$

satisfying the endpoint constraints (6.64) and (6.65). Our goal is to derive necessary conditions of the PMP type for a given optimal solution $\{\bar{u}(\cdot), \bar{x}(\cdot)\}$ to the problem under consideration. Although we present necessary conditions for (global) optimal solutions, one can observe from the proofs provided below that the results obtained hold true for *local minimizers* $\{\bar{u}(\cdot), \bar{x}(\cdot)\}$ in the sense that $J[\bar{u}, \bar{x}] \le J[x, u]$ whenever (u, x) is feasible to (6.63)–(6.65) and $\|x(t) - \bar{x}(t)\| < \varepsilon$ for all $t \in [a, b]$ with some $\varepsilon > 0$. This corresponds to *strong* local minimizers in Subsect. 6.1.2 for $F(x, t) = f(x, U, t)$.

Given an optimal solution $\{\bar{u}(\cdot), \bar{x}(\cdot)\}$ to (6.63)–(6.65), we impose the following *standing assumptions* throughout the whole section:

—the state space X is Banach;

—the control set U is a *Souslin subset* (i.e., a continuous image of a Borel subset) in a complete and separable metric space;

—there is an open set $O \subset X$ containing $\bar{x}(t)$ such that f is Fréchet differentiable in x with both $f(x, u, t)$ and $\nabla_x f(x, u, t)$ continuous in (x, u), measurable in t, and norm-bounded by a summable function for all $x \in O$, $u \in U$, and a.e. $t \in [a, b]$;

—the functions φ_i are continuous around $\bar{x}(b)$ and Fréchet differentiable at this point for $i = m + 1, \ldots, m + r$.

Note that the control set U may depend on t in a general *measurable* way, which allows one to use standard *measurable selection* results; see, e.g., the books [54, 229, 1165] with the references therein.

Appropriate assumptions on the functions φ_i, $i = 0, \ldots, m$, describing the objective and inequality constraints will be presented in the main theorems stated below. Note that the basic assumptions on them require their Fréchet *differentiability at* $\bar{x}(b)$ (*not* even their *continuity around* this point), while upper subdifferential conditions hold for a broader class of nondifferentiable functions on *arbitrary* Banach spaces.

To formulate the relations of the maximum principle, let us define the *Hamilton-Pontryagin* function for system (6.61) by

$$H(x, p, u, t) := \langle p, f(x, u, t) \rangle \text{ with } p \in X^* .$$

Observe that the *Hamiltonian* defined in Sect. 6.2 for $F(x, t) = f(x, U, t)$ corresponds to the *maximization* of the function $H(x, p, u, t)$ with respect to u over the whole the control region:

$$\mathcal{H}(x,p,t) = \max\left\{H(x,p,u,t)\,\middle|\, u \in U\right\} .$$

Note also that H is *smooth* with respect to the state and adjoint variables (x,p), which of course is not the case for $\mathcal{H}$.

Theorem 6.37 (maximum principle for smooth control systems). *Let $\{\bar{u}(\cdot),\bar{x}(\cdot)\}$ be an optimal solution to problem* (6.63)–(6.65) *under the standing assumptions made. Suppose also that the functions φ_i, $i=0,\ldots,m$, are Fréchet differentiable at the optimal endpoint $\bar{x}(b)$. Then there are multipliers $(\lambda_0,\ldots,\lambda_{m+r}) \neq 0$ satisfying*

$$\lambda_i \ge 0 \quad \text{for} \quad i = 0,\ldots,m ,$$

$$\lambda_i\varphi_i\big(\bar{x}(b)\big) = 0 \quad \text{for} \quad i = 1,\ldots,m ,$$

and such that the following maximum condition holds:

$$H\big(\bar{x}(t),p(t),\bar{u}(t),t\big) = \max_{u\in U} H\big(\bar{x}(t),p(t),u,t\big) \quad \text{a.e.} \quad t \in [a,b] , \tag{6.66}$$

where an absolutely continuous mapping $p\colon [a,b] \to X^*$ is a trajectory for the adjoint system

$$\dot{p} = -\nabla_x H(\bar{x},p,\bar{u},t) \quad \text{a.e.} \quad t \in [a,b] \tag{6.67}$$

with the transversality condition

$$p(b) = -\sum_{i=0}^{m+r}\lambda_i\nabla\varphi_i\big(\bar{x}(b)\big) . \tag{6.68}$$

Note that a solution (adjoint arc) to system (6.67) is understood in the *integral/mild* sense similarly to (6.61), i.e.,

$$p(t) = p(b) + \int_t^b \nabla_x H\big(\bar{x}(s),p(s),\bar{u}(t),s\big)\,ds, \quad t \in [a,b] ,$$

with $\nabla_x H(\bar{x},p,\bar{u}.t) = \langle p, \nabla_x f(\bar{x},\bar{u},t)\rangle$. Observe also that the transversality condition (6.66) agrees with the one in Corollary 6.29. However, now the endpoint functions is *not* assumed to be *strictly* differentiable at $\bar{x}(b)$.

The proof of Theorem 6.37 will be given in Subsects. 6.3.2–6.3.4. Meantime let us formulate and prove an *upper subdifferential* counterpart of this theorem, which gives on one hand an extension of the transversality condition (6.68) to the case of nondifferentiable functions φ_i, $i=0,\ldots,m$, while on the other hand follows from Theorem 6.37 and the *smooth variational description* of Fréchet subgradients.

Theorem 6.38 (maximum principle with transversality conditions via Fréchet upper subgradients). *Let* $\{\bar{u}(\cdot), \bar{x}(\cdot)\}$ *be an optimal solution to the control problem* (6.63)–(6.65) *under the standing assumptions made. Then for every collection of Fréchet upper subgradients* $x_i^* \in \widehat{\partial}^+\varphi_i\big(\bar{x}(b)\big)$, $i = 0, \ldots, m$, *there are multipliers* $(\lambda_0, \ldots, \lambda_{m+r}) \neq 0$ *satisfying the sign and complementary slackness conditions of Theorem* 6.37 *and such that the maximum condition* (6.66) *holds with the corresponding trajectory* $p(\cdot)$ *of the adjoint system* (6.67) *satisfying the transversality condition*

$$p(b) + \sum_{i=0}^{m+r} \lambda_i x_i^* = 0\ . \tag{6.69}$$

Proof. Take an arbitrary set of Fréchet upper subgradients $x_i^* \in \widehat{\partial}^+\varphi_i\big(\bar{x}(b)\big)$, $i = 0, \ldots, m$, and employ the smooth variational description of $-x_i^*$ from assertion (i) of Theorem 1.88 held in any Banach space. In this way we find functions $s_i\colon X \to I\!R$ for $i = 0, \ldots, m$ satisfying the relations

$$s_i\big(\bar{x}(b)\big) = \varphi_i\big(\bar{x}(b)\big), \quad s_i(x) \geq \varphi_i(x) \ \text{ around } \bar{x}(b)\ ,$$

and such that each $s_i(\cdot)$ is Fréchet differentiable at $\bar{x}(b)$ with $\nabla s_i\big(\bar{x}(b)\big) = x_i^*$, $i = 0, \ldots, m$. From the construction of these functions we easily deduce that the process $\{\bar{u}(\cdot), \bar{x}(\cdot)\}$ is an optimal solution to the following control problem:

$$\text{minimize } \ \widetilde{J}[u, x] = s_0\big(x(b)\big) \ \text{ over } \ (u, x) \in \mathcal{A}$$

subject to the inequality and equality endpoint constraints

$$s_i\big(x(b)\big) \leq 0 \ \text{ for } \ i = 1, \ldots, m$$

and (6.65), where $\mathcal{A}$ is the collection of admissible control-trajectory pairs defined in the beginning of this subsection. The initial data of the latter optimal control problem satisfy all the assumptions of Theorem 6.37. Thus applying the above maximum principle to this problem and taking into account that $\nabla s_i\big(\bar{x}(b)\big) = x_i^*$ for $i = 0, \ldots, m$, we complete the proof of the theorem. △

One can observe the difference between the formulations and proofs of Theorem 6.38, in the part related to upper subdifferential transversality conditions, and of Theorem 5.19 on upper subdifferential optimality conditions in mathematical programming. Both results reduce to their *smooth* (in difference senses) counterparts based on smooth variational descriptions of Fréchet subgradients. In the case of Theorem 5.19 we need to require the continuous differentiability (more precisely, *strict differentiability*) of the cost and constraint functions to be able to apply the corresponding necessary conditions in smooth nonlinear programming. In this way an additional assumption on the geometry of Banach spaces comes into play to ensure the $\mathcal{C}^1$ description of

Fréchet subgradients by Theorem 1.88(ii). On the other hand, Theorem 6.38 relies, by a milder smooth variational description from Theorem 1.88(i), on the preceding Theorem 6.37 that requires only the *Fréchet differentiability* of the endpoint functions *at* the optimal point. Note that Theorems 6.37 and 6.38 concerning optimal control problems obviously imply, by putting $f = 0$ in (6.61), the corresponding *improvements* of the results in Subsect. 5.1.3 for *mathematical programming* problems with equality and inequality constraints.

Remark 6.39 (control problems with constraints at both endpoints and at intermediate points of trajectories). One can see from the proof of Theorem 6.37 given in Subsects. 6.3.2–6.3.4 that a minor modification of this proof allows us to derive similar necessary optimality conditions (including those of the upper subdifferential type) for optimal control problems with endpoint constraints of form (6.64) and (6.65) at both $t = a$ and $t = b$ and with the cost function φ_0 depending on both $x(a)$ and $x(b)$ under the same assumptions on the initial data. In this case the transversality condition (6.68) on the absolutely continuous adjoint arc $p\colon [a,b] \to X^*$ is replaced by

$$\big(p(a), -p(b)\big) = \sum_{i=0}^{m+r} \lambda_i \nabla\varphi_i\big(\bar{x}(a), \bar{x}(b)\big) \ .$$

Furthermore, we may similarly derive necessary optimality conditions for control problems involving *intermediate state constraints*, i.e., with constraints on trajectories given at intermediate points $\tau_i \in [a,b]$ of the time interval. For example, consider the modified problem (6.63)–(6.65) with

$$\varphi_i = \varphi_i\big(x(a), x(\tau), x(b)\big), \quad i = 0, \ldots, m+r \ ,$$

where $\tau \in (a,b)$ is an intermediate moment of the time interval. Then the difference between the necessary optimality conditions of Theorem 6.37 and the ones for the modified state-constrained problem is that we now have a *discontinuous* adjoint arc $p(\cdot)$ with the *jump condition* at the intermediate point $t = \tau$ incorporated into the transversality conditions as follows:

$$\big(p(a), p(\tau+0) - p(\tau-0), -p(b)\big) = \sum_{i=0}^{m+r} \lambda_i \nabla\varphi_i\big(\bar{x}(a), \bar{x}(\tau), \bar{x}(b)\big) \ .$$

We can similarly modify the upper subdifferential conditions of Theorem 6.38 in the case of control problems with intermediate state constraints.

Remark 6.40 (maximum principle in time-delay control systems). The results of Theorems 6.37 and 6.38 can be extended to various systems with time delays in state and control variables. For example, let us consider the standard system with a constant time delay $\theta > 0$ in the state variable:

$$\begin{cases} \dot{x}(t) = f\big(x(t), x(t-\theta), u(t), t\big) & \text{a.e. } t \in [a,b]\,, \\ x(t) = c(t), \quad t \in [a-\theta, a]\,, \\ u(t) \in U & \text{a.e. } t \in [a,b] \end{cases}$$

over measurable controls and absolutely continuous trajectories with a Banach state space X and the initial "tail" mapping $c\colon [a-\theta, a] \to X$ that is necessary to start the time-delay process. Denote by $\mathcal{A}$ the collection of admissible pairs $\{u(\cdot), x(\cdot)\}$ satisfying the above delay system and define the corresponding Hamilton-Pontryagin function

$$H(x, y, p, u, t) := \big\langle p, f(x, y, u, t)\big\rangle, \quad p \in X^*\,,$$

where y stands for the delay variable $x(t-\theta)$. Considering now problem (6.63)–(6.65) with $\mathcal{A}$ signifying the collection of admissible pairs for the delay system, we get counterparts of Theorems 6.37 and Theorem 6.38 with the *adjoint system* given by

$$-\dot{p}(t) = \begin{cases} \nabla_x H\big(x(t), x(t-\theta), p(t), u(t), t\big) \\ +\nabla_y H\big(x(t+\theta), x(t), p(t+\theta), u(t+\theta), t\big) & \text{a.e. } t \in [a, b-\theta]\,; \\ \nabla_x H\big(x(t), x(t-\theta), p(t), u(t), t\big) & \text{a.e. } t \in [b-\theta, b]\,. \end{cases}$$

These results can be actually proved by reducing the time-delay control system in X to the one with *no* delay in the state space X^N, for some natural number N sufficiently large. Furthermore, the methods developed in the proofs of Theorems 6.37 and 6.38 allow us to derive similar results for control problems with more *general delays* depending on both time and state variables, as well as with time-distributed delays.

Remark 6.41 (functional-differential control systems of neutral type). The dynamics of such control systems is described by differential equations with time delays not only in state variables but in *velocity* variables as well. A typical model is given by

$$\dot{x}(t) = f\big(x(t), x(t-\theta), \dot{x}(t-\theta), u(t), t\big), \quad u(t) \in U, \quad \text{a.e. } t \in [a,b]$$

with proper initial conditions on $[a-\theta, a]$. Systems of this type are *fundamentally different* from the standard ODE control systems and time-delay systems considered in the preceding remark. They are substantially more difficult for variational analysis and exhibit a number of phenomena that are not inherent in the control systems considered above; the reader may find more discussions in Commentary to Chap. 7, where we consider such systems and their extensions in more details. Now observe that, although necessary optimality conditions in the form of Theorems 6.37 and 6.38 can be derived by similar methods in the case of *convex velocity sets* $f(x,y,z,U,t)$ with a Banach state space, a proper analog of the Pontryagin maximum principle *doesn't generally hold* for neutral control systems even with no endpoint constraints in finite dimensions. It happens, in particular, for the optimal control

$$\bar{u}(t)=0 \ \text{ as } \ t\in[0,1) \ \text{ and } \ \bar{u}(t)=1 \ \text{ as } \ t\in[1,2]$$

to the following two-dimensional control problem:

$$\begin{cases} \text{minimize } \ J[u,x]=x_2(2) \ \text{ subject to} \\ \dot{x}_1(t)=u(t), \quad \dot{x}_2(t)=\dot{x}_1^2(t-1)-u^2(t), \ \ t\in[0,2]\ , \\ x_1(t)=x_2(t)=0, \ \ t\in[-1,0]; \quad |u(t)|\le 1, \ \ t\in[0,2]\ . \end{cases}$$

The reader can find complete calculations for this example in the book by Gabasov and Kirillova [485, Sect. 3.6]; see also Example 6.70 in Subsect. 6.4.6 below for similar calculations in a finite-difference analog of this control problem.

6.3.2 Maximum Principle for Free-Endpoint Problems

In this subsection we study problem (6.63), where $\mathcal{A}$ is the collection of admissible pairs $\{u(\cdot),x(\cdot)\}$ for the control system (6.61) with the fixed left endpoint $x(a)=x_0$; see the beginning of the preceding subsection for the exact formulation. This problem is labeled as a *free-endpoint problem of optimal control* despite the left endpoint is always fixed; we have in mind the absence of the constraints (6.64) and (6.65) on the right endpoint of admissible trajectories. As follows from the proofs below, the *free-endpoint* problem (6.63) is *significantly different* from the constrained problem (6.63)–(6.65); moreover, the problems with *inequality* and *equality* endpoint constraints are essentially different from each other as well. The principal difference between the unconstrained and constrained problems is that in case of (6.63) all admissible trajectories are feasible, and one doesn't need to care about satisfying the endpoint constraints while varying admissible controls $u(\cdot)\in U$. Note that the control constraints of the above (arbitrary) geometric type are *always present* in the problems under consideration, they distinguish optimal control problems from the classical calculus of variations and signify *intrinsic nonsmoothness* inherent in optimal control.

This subsection is devoted to the proof of the maximum principle from Theorem 6.37 for problem (6.63) under the assumptions made in the theorem on the given data (U,X,f,φ_0). Note that the transversality condition (6.68) reduces in this case to

$$p(b)=-\nabla\varphi_0\big(\bar{x}(b)\big)\ , \tag{6.70}$$

i.e., with $\lambda_0 = 1$ and $\lambda_i = 0$, $i = 1, \ldots, m + r$, in (6.68). Indeed, if $\lambda_0 = 0$ and $p(b) = 0$ in (6.68), then $p(t) \equiv 0$ for all $t \in [a, b]$ due to the linearity of the adjoint system (6.67) with respect to p, which would contradict the nontriviality condition $\big(p(\cdot), \lambda_0\big) \neq 0$ in Theorem 6.37.

The proof of Theorem 6.37 for the free-endpoint problem (6.63) is *purely analytic*, in the sense that it doesn't invoke any geometric facts and arguments in the vein of the convex separation theorem and the like. This is significantly different from the proofs of Theorem 6.37 in the case of inequality and equality endpoint constraints given in Subsect. 5.3.3 and 5.3.4. The basic ingredients in the proof of Theorem 6.37 for problem (6.63) are the *increment formula* for the cost functional in (6.63) and the use of the so-called *needle* variations (sometimes called "McShane variations") of the optimal control.

Let us start with the increment formula. Given two admissible controls $\bar{u}(t), u(t) \in U$ (observe that $\bar{u}(\cdot)$ may not be optimal before resuming it in the sequel) and the corresponding solutions $\bar{x}(\cdot), x(\cdot)$ in (6.62), we denote

$$\Delta\bar{u}(t) := u(t) - \bar{u}(t), \quad \Delta\bar{x}(t) := x(t) - \bar{x}(t), \quad \Delta J[\bar{u}] := \varphi_0\big(x(b)\big) - \varphi_0\big(\bar{x}(b)\big) .$$

Our intention is to obtain a convenient representation of the cost functional increment $\Delta J[\bar{u}]$ in terms of the Hamilton-Pontryagin function evaluated along the admissible pair $\{\bar{u}(\cdot), \bar{x}(\cdot)\}$ and the corresponding trajectory $p(\cdot)$ of the adjoint system (6.67) with the boundary condition (6.70). Recall that we use the same standard symbol $o(\cdot)$ for *all* expressions of this category.

Lemma 6.42 (increment formula for the cost functional). *Let*

$$\Delta_u H\big(\bar{x}(t), p(t), \bar{u}(t), t\big) := H\big(\bar{x}(t), p(t), u(t), t\big) - H\big(\bar{x}(t), p(t), \bar{u}(t), t\big)$$

in the notation above. Then one has

$$\Delta J[\bar{u}] = -\int_a^b \Delta_u H\big(\bar{x}(t), p(t), \bar{u}(t), t\big)\, dt + \eta ,$$

where the remainder η is given by $\eta = \eta_1 + \eta_2 + \eta_3$ with

$$\eta_1 := o\big(\|\Delta\bar{x}(b)\|\big), \quad \eta_2 := -\int_a^b o\big(\|\Delta\bar{x}(t)\|\big)\, dt, \quad \text{and}$$

$$\eta_3 := -\int_a^b \Big\langle \frac{\partial \Delta_u H\big(\bar{x}(t), p(t), \bar{u}(t), t\big)}{\partial x}, \Delta\bar{x}(t) \Big\rangle\, dt .$$

Proof. Since φ_0 is assumed to be Fréchet differentiable at $\bar{x}(b)$, we have the representation

$$\Delta J[\bar{u}] = \varphi_0\big(x(b)\big) - \varphi_0\big(\bar{x}(b)\big) = \big\langle \nabla\varphi_0\big(\bar{x}(b)\big), \Delta\bar{x}(b)\big\rangle + o\big(\|\Delta\bar{x}(b)\|\big) .$$

Taking into account that solutions to the state and adjoint equations satisfy (by definition) the Newton-Leibniz formula and using *integration by parts* held for the Bochner integral, one gets the identity

$$\langle p(b), \Delta\bar{x}(b)\rangle = \int_a^b \langle \dot{p}(t), \Delta\bar{x}(t)\rangle\,dt + \int_a^b \langle p(t), \Delta\dot{\bar{x}}(t)\rangle\,dt\,,$$

where $p\colon [a,b] \to X^*$ is an arbitrary absolutely continuous mapping from the solution class. Imposing the boundary condition (6.70) on $p(b)$, we arrive at

$$\Delta J[\bar{u}] = -\int_a^b \langle \dot{p}(t), \Delta\bar{x}(t)\rangle\,dt - \int_a^b \langle p(t), \Delta\dot{\bar{x}}(t)\rangle\,dt + o\big(\|\Delta\bar{x}(b)\|\big)\,.$$

Let us transform the second integral above. Using the equation

$$\Delta\dot{\bar{x}}(t) = f\big(\bar{x}(t) + \Delta\bar{x}(t), \bar{u}(t) + \Delta\bar{u}(t), t\big) - f\big(\bar{x}(t), \bar{u}(t), t\big)\,,$$

the definition of the Hamilton-Pontryagin function $H(x,p,u,t)$, and the smoothness of f in x, we have

$$\begin{aligned}
&\int_a^b \langle p(t), \Delta\dot{\bar{x}}(t)\rangle\,dt \\
&= \int_a^b \Big[H\big(\bar{x}(t) + \Delta\bar{x}(t), p(t), \bar{u}(t) + \Delta\bar{u}(t), t\big) - H\big(\bar{x}(t), p(t), \bar{u}(t), t\big)\Big]\,dt \\
&= \int_a^b \Big[H\big(\bar{x}(t), p(t), \bar{u}(t) + \Delta\bar{u}(t), t\big) - H\big(\bar{x}(t), p(t), \bar{u}(t), t\big)\Big]\,dt \\
&+ \int_a^b \Big\langle \frac{\partial H\big(\bar{x}(t), p(t), \bar{u}(t), t\big)}{\partial x}, \Delta\bar{x}(t)\Big\rangle\,dt + \int_a^b o\big(\|\Delta\bar{x}(t)\|\big)\,dt\,.
\end{aligned}$$

Remembering finally that $p(\cdot)$ is a solution to the adjoint system (6.67) generated by $\{\bar{u}(\cdot), \bar{x}(\cdot)\}$, we complete the proof of the lemma. △

In the above increment formula both controls $\bar{u}(\cdot)$ and $u(\cdot)$ are arbitrary measurable mappings satisfying the pointwise control constraints. Now we build $u(\cdot)$ as a special perturbation of the reference control $\bar{u}(\cdot)$ that is called a *needle variation*, or sometimes a *single needle variation*, of this control. Namely, fix arbitrary numbers $\tau \in [a,b)$ and $\varepsilon > 0$ with $\tau + \varepsilon < b$, take an arbitrary point $\upsilon \in U$, and construct an admissible control $u(t)$, $t \in [a,b]$, in the following form

$$u(t) := \begin{cases} \upsilon, & t \in [\tau, \tau+\varepsilon)\,, \\ \bar{u}(t), & t \notin [\tau, \tau+\varepsilon)\,. \end{cases} \tag{6.71}$$

The obtained perturbed control differs from the reference one only on the small time interval $[\tau, \tau+\varepsilon)$, where its value is *arbitrary* in the control set U; the name "needle variation" comes from this. For the corresponding trajectory increment $\Delta\bar{x}(t)$, depending on the parameters $(\tau, \varepsilon, \upsilon)$, one clearly has

$$\Delta\bar{x}(t) = 0 \ \text{ for all } \ t \in [a, \tau] \ .$$

Let us estimate $\Delta\bar{x}(t)$ for $t \in (\tau, b]$, which is given in the next lemma. In what follows we denote by ℓ the uniform Lipschitz constant for $f(\cdot, \upsilon, t)$ whose existence is guaranteed by the standing assumptions. For simplicity we suppose that ℓ is independent of t although the assumptions made allow it to be summable on $[a, b]$ with no change of the result.

Lemma 6.43 (increment of trajectories under needle variations). *Let $\Delta\bar{x}(\cdot)$ be the increment of $\bar{x}(\cdot)$ corresponding to the needle variation* (6.71) *of $\bar{u}(\cdot)$ with parameters $(\tau, \varepsilon, \upsilon)$. Then there is a constant $K > 0$ independent of (τ, ε) (it may depend on υ) such that*

$$\|\Delta\bar{x}(t)\| \le K\varepsilon \ \ \textit{for all} \ \ t \in [a, b] \ .$$

Proof. Since $\Delta\bar{x}(\tau) = 0$, one has by (6.62) that

$$\Delta\bar{x}(t) = \int_\tau^t \Big[f\big(\bar{x}(s) + \Delta\bar{x}(s), \upsilon, s\big) - f\big(\bar{x}(s), \bar{u}(s), s\big)\Big]\, ds, \quad \tau \le t \le \tau + \varepsilon \ .$$

Taking into account the uniform Lipschitz continuity of f in x with the constant ℓ and denoting $\Delta_\upsilon f\big(\bar{x}(s), \bar{u}(s), s\big) := f\big(\bar{x}(s), \upsilon, s\big) - f\big(\bar{x}, \bar{u}(s), s\big)$, we have

$$\begin{aligned} \|\Delta\bar{x}(t)\| &= \int_\tau^t \big\|f\big(\bar{x}(s) + \Delta\bar{x}(s), \upsilon, s\big) - f\big(\bar{x}(s), \bar{u}(s), s\big)\big\|\, ds \\ &\le \int_\tau^t \big\|\Delta_\upsilon f\big(\bar{x}(s), \bar{u}(s), s\big)\big\| \ ds + \ell \int_\tau^t \|\Delta\bar{x}(s)\|\, ds \ . \end{aligned}$$

Using the notation

$$\alpha(t) := \int_\tau^t \big\|\Delta_\upsilon f\big(\bar{x}(s), \bar{u}(s), s\big)\big\|\, ds \ \text{ and } \ \beta(t) := \|\Delta\bar{x}(t)\| \ ,$$

the above estimate can be rewritten as

$$\beta(t) \le \alpha(t) + \ell \int_\tau^t \beta(s)\, ds, \quad \tau \le t \le \tau + \varepsilon \ ,$$

which yields by the classical Gronwall lemma that

$$\|\Delta\bar{x}(t)\| \le \Big(\int_\tau^t \big\|\Delta_\upsilon f\big(\bar{x}(s), \bar{u}(s), s\big)\big\|\, ds\Big) \exp\big(\ell(t - \tau)\big) \le K\varepsilon$$

for $t \in [\tau, \tau + \varepsilon]$, where $K = K(\upsilon)$ is independent of ε and τ.

It remains to estimate $\Delta\bar{x}(t)$ on the last interval $[\tau + \varepsilon, b]$, where it satisfies the equation

$$\Delta\dot{\bar{x}}(t) = f\big(\bar{x}(t) + \Delta\bar{x}(t), \bar{u}(t), t\big) - f\big(\bar{x}(t), \bar{u}(t), t\big) \text{ with } \|\Delta\bar{x}(\tau+\varepsilon)\| \le K\varepsilon$$

the solution of which is understood in the integral sense (6.62). Since

$$\|\Delta\bar{x}(t)\| \le \|\Delta\bar{x}(\tau+\varepsilon)\| + \int_{\tau+\varepsilon}^{t} \big\|f\big(\bar{x}(s) + \Delta\bar{x}(s), \bar{u}(s), s\big) - f\big(\bar{x}(s), \bar{u}(s), s\big)\big\|\, ds$$

$$\le K\varepsilon + \ell \int_{\tau+\varepsilon}^{t} \|\Delta\bar{x}(s)\|\, ds, \quad \tau + \varepsilon \le t \le b\,,$$

we again apply the Gronwall lemma and arrive, by increasing K if necessary, at the desired estimate of $\|\Delta\bar{x}(t)\|$ on the whole interval $[a, b]$. △

Now we are ready to justify the maximum principle of Theorem 6.37 for the free-endpoint control problem under consideration.

Proof of Theorem 6.37 for the free-endpoint problem. Let $\{\bar{u}(\cdot), \bar{x}(\cdot)\}$ be an optimal solution to problem (6.63), and let $p(\cdot)$ be the corresponding solution to the adjoint system (6.67) with the boundary/transversality condition (6.70). We are going to show that the maximum condition (6.66) holds for a.e. $t \in [a, b]$. Assume on the contrary that there is a set $T \subset [a, b]$ of a positive measure such that

$$H\big(\bar{x}(t), p(t), \bar{u}(t), t\big) < \sup_{u \in U} H\big(\bar{x}(t), p(t), u, t\big) \text{ for } t \in T\,.$$

Then using standard results on *measurable selections* under the assumptions made, we find a measurable mapping $v: T \to U$ satisfying

$$\Delta_v H(t) := H\big(\bar{x}(t), p(t), v(t), t\big) - H\big(\bar{x}(t), p(t), \bar{u}(t), t\big) > 0, \quad t \in T\,.$$

Let $T_0 \subset [a, b]$ be a set of *Lebesgue regular* points (or points of approximate continuity) for the function $H(t)$ on the interval $[a, b]$, which is of *full measure* on $[a, b]$ due to the classical Denjoy theorem. Given $\tau \in T_0$ and $\varepsilon > 0$, consider a *needle variation* of the optimal control built by

$$u(t) := \begin{cases} v(t), & t \in T_\varepsilon := [\tau, \tau+\varepsilon) \cap T_0\,, \\ \bar{u}(t), & t \in [a, b] \setminus T_\varepsilon\,, \end{cases}$$

and apply to $\bar{u}(\cdot)$ and $u(\cdot)$ the *increment formula* for the cost functional from Lemma 6.42. By this formula we have the relation

$$\Delta J[\bar{u}] = -\int_{\tau}^{\tau+\varepsilon} \Delta_v H(t)\, dt + \eta_1 + \eta_2 + \eta_3$$

with the above positive increment of the Hamilton-Pontryagin function $\Delta_v H(t)$ and the remainders η_i, $i = 1, 2, 3$, defined in Lemma 6.42 along the trajectory increment $\Delta\bar{x}(\cdot)$ corresponding to the needle variation $u(\cdot)$ under consideration. It follows from the proof of Lemma 6.43, with an easy modification to take into account the variable perturbation $v(\cdot)$ on T_ε instead of the constant one in (6.71), that $\|\Delta\bar{x}(t)\| = O(\varepsilon)$ for $t \in [a,b]$. Hence

$$\eta_1 = o\big(\|\Delta\bar{x}(b)\|\big) = o(\varepsilon), \quad \eta_2 = \int_a^b o\big(\|\Delta\bar{x}(t)\|\big)\,dt = o(\varepsilon), \quad \text{and}$$

$$\begin{aligned}\eta_3 &\le \int_\tau^{\tau+\varepsilon} \Big|\Big\langle \frac{\partial \Delta H_v\big(\bar{x}(t), p(t), \bar{u}(t), t\big)}{\partial x}, \Delta\bar{x}(t)\Big\rangle\Big|\,dt \\ &\le K\varepsilon \int_\tau^{\tau+\varepsilon} \Big\|\frac{\partial \Delta H_v\big(\bar{x}(t), p(t), \bar{u}(t), t\big)}{\partial x}\Big\|\,dt = o(\varepsilon)\ .\end{aligned}$$

The choice of $\tau \in T_0$ as a Lebesgue regular point of the function $\Delta_v H(t)$ and the construction of the Bochner integral yield

$$\int_\tau^{\tau+\varepsilon} \Delta_v H(t)\,dt = \varepsilon\Big[H\big(\bar{x}(\tau), p(\tau), v(\tau), \tau\big) - H\big(\bar{x}(\tau), p(\tau), \bar{u}(\tau), \tau\big)\Big] + o(\varepsilon)\ .$$

Thus we get the representation

$$\Delta J[\bar{u}] = -\varepsilon\Big[H\big(\bar{x}(\tau), p(\tau), v(\tau), \tau\big) - H\big(\bar{x}(\tau), p(\tau), \bar{u}(\tau), \tau\big)\Big] + o(\varepsilon)\ ,$$

which implies that $\Delta J[\bar{u}] < 0$ along the above needle variation of the optimal control $\bar{u}(\cdot)$ for all $\varepsilon > 0$ sufficiently small. This clearly contradicts the optimality of $\bar{u}(\cdot)$ in problem (6.63) and completes the proof of Theorem 6.37 for the free-endpoint optimal control problem. △

6.3.3 Transversality Conditions for Problems with Inequality Constraints

One can see from the preceding subsection that the analytic proof of the maximum principle given there for the free-endpoint optimal control problem doesn't hold in the case of endpoint constraints of types (6.64) and/or (6.65). Indeed, in that proof we didn't care about the *feasibility* with respect to these constraints of trajectories corresponding to needle control variations. Dealing with endpoint constraint problems requires a more sophisticated technique that involves the *geometry* of the reachable set for system (6.61) and its interaction with the cost functional and endpoint constraints. The crux of the matter is to show that there is a *convex* set generated by feasible endpoint variations of the given optimal trajectory that doesn't intersect some convex set "forbidden" by optimality, which allows us to employ the *convex separation.* This can be achieved by using *multineedle* variations of the optimal control in question. The latter is realized by the *continuity of time* in $[a,b]$ and actually reflects the *hidden convexity* of continuous-time control problems.

In this subsection we consider optimal control problems that involve only endpoint constraints of the *inequality type* (6.64). Control problems with the equality constraints (6.65) are somewhat different (more complicated); they will be studied in the next subsection. Our main goal is to derive the transversality condition (6.68) in the relations of the maximum principle from Theorem 6.37 in the case of inequality constraints given by differentiable functions. As discussed in Subsect. 6.3.1, transversality conditions in more general control problems and under less restrictive assumptions can be either reduced to the one in (6.68) or derived similarly.

Let us emphasize that, although we study optimal control problems with a *Banach* state space X, they involve only *finitely many* endpoint constraints on system trajectories. The method we develop allows us to take an advantage of this setting (which is somehow related to the *finite codimension* property of the constraint set; cf. Corollaries 6.29, 6.24 and Remark 6.25) and to deal with *finite-dimensional images* of endpoint variations under the derivative operators for the cost and constraint functions, employing thus the convex separation theorem in finite dimensions.

In the rest of this subsection we consider the optimal control problem (6.63) with the inequality endpoint constraints (6.64) and fix an optimal solution $\{\bar{u}(\cdot),\bar{x}(\cdot)\}$ to this problem. Assume without loss of generality that $\varphi_i\big(\bar{x}(b)\big)=0$ for all $i=1,\ldots,m$. It is easy to see from the proof (as usually with inequality constraints) that $\lambda_i=0$ if $\varphi_i\big(\bar{x}(b)\big)<0$ for some $i\in\{1,\ldots,m\}$, i.e., the corresponding function φ_i can be excluded from consideration. In this setting the complementary slackness conditions of Theorem 6.37 hold automatically, and we need to establish relations (6.66)–(6.68) with $r=0$ and $0\neq(\lambda_0,\ldots,\lambda_m)\in I\!\!R^m_+$.

Along with (single) needle variations introduced in the preceding subsection we now invoke "multineedle variations" built as follows. Fix a natural number $M\geq 1$ and M points $\tau_j\in[a,b]$ of the original time interval with $a\leq\tau_1<\tau_2\leq\ldots<\tau_M<b$. Consider also arbitrary numbers $N_j\in I\!\!N$ for $j=1,\ldots,M$ and $\alpha_{ij}\in[0,1]$ for $i=1,\ldots,N_j$ satisfying the relations

$$\tau_j+\varepsilon_0\sum_{i=1}^{N_j}\alpha_{ij}<\tau_{j+1},\quad j=1,\ldots,M-1,\quad\text{and}\quad\tau_M+\varepsilon_0\sum_{i=1}^{N_M}\alpha_{iM}<b$$

with some $\varepsilon_0>0$. We are going to construct a perturbation $u(\cdot)$ of the reference control $\bar{u}(\cdot)$ that is different from $\bar{u}(\cdot)$ on $N_1+\ldots+N_M$ time intervals of a *small total length*, while the difference between $u(\cdot)$ and $\bar{u}(\cdot)$ on these intervals is up to *any element* from the feasible control region U. To proceed, let us take arbitrary $v_{ij}\in U$ and $\varepsilon\in(0,\varepsilon_0]$ and define a *multineedle variation* $u(\cdot)$ of the reference control $\bar{u}(\cdot)$ by

$$u(t) := \begin{cases} v_{ij}, & t \in \Big[\tau_j + \sum_{\nu=0}^{i-1} \alpha_{\nu j}\varepsilon, \tau_j + \sum_{\nu=1}^{i} \alpha_{\nu j}\varepsilon\Big), \quad \alpha_{0j} := 0,\ i = 1, \ldots, N_j\ , \\ \bar{u}(t), & t \notin \Big[\tau_j, \tau_j + \sum_{i=1}^{N_j} \alpha_{ij}\varepsilon\Big), \quad j = 1, \ldots, M\ . \end{cases} \tag{6.72}$$

Note that, although there are M basic points τ_j, the multineedle variation (6.72) involves $N_1 + \ldots + N_M$ points of needle-type perturbations; this is different from a single needle variation (6.71) even in the case of $M = 1$. Actually the multineedle variation (6.72) is a *collection* of $N_1 + \ldots + N_M$ *single* needle variations of type (6.71) with the given parameters $(\tau_j, v_{ij}, \alpha_{ij}, \varepsilon)$.

Let $\Delta\bar{x}_{\tau_j,v_{ij},\alpha_{ij},\varepsilon}(b)$ be the *endpoint increment* of the trajectory $\bar{x}(\cdot)$ corresponding to the *single needle variation* of type (6.71) with the parameters $(\tau_j, v_{ij}, \alpha_{ij}, \varepsilon)$. Dealing with the differential equation (6.61) of smooth dynamics and its *linearization* in x along the process $\{\bar{u}(\cdot), \bar{x}(\cdot)\}$ as in the proof of Lemma 6.42, we can check the relationship

$$\Delta\bar{x}_{\tau_j,v_{ij},\alpha_{ij},\varepsilon}(b) = \big[\alpha_{ij}\,\Lambda\bar{x}_{\tau_j,v_{ij},1}(b)\big]\varepsilon + o(\varepsilon) \tag{6.73}$$

between $\Delta\bar{x}_{\tau_j,v_{ij},\alpha_{ij},\varepsilon}(b)$ and the corresponding *linearized endpoint increment* $\Lambda\bar{x}_{\tau_j,v_{ij},\alpha_{ij}}(b)$ computed by

$$\Lambda\bar{x}_{\tau_j,v_{ij},\alpha_{ij}}(b) = \alpha_{ij} R(b, \tau_j) \Delta_{v_{ij}} f\big(\bar{x}(\tau_j), \bar{u}(\tau_j), \tau_j\big) =: \alpha_{ij}\,\Lambda\bar{x}_{\tau_j,v_{ij},1}$$

via the *resolvent* (Green function) $R(t, \tau)$ of the linearized homogeneous equation for (6.61) with respect to x along $\{\bar{u}(\cdot), \bar{x}(\cdot)\}$ given as

$$\dot{x} = \nabla_x f\big(\bar{x}(t), \bar{u}(t), t\big)x\ .$$

Furthermore, the *endpoint increment* $\Delta\bar{x}(b)$ generated by the *multineedle variation* (6.72) is represented by

$$\Delta\bar{x}(b) = \Big[\sum_{j=1}^{M}\sum_{i=1}^{N_j} \alpha_{ij}\,\Lambda_{\tau_j,v_{ij},1}\bar{x}(b)\Big]\varepsilon + o(\varepsilon)\ .$$

Now we form the following finite-dimensional *linearized image set* generated by inner products involving derivatives of the cost and constraint functions and the linearized endpoint increments corresponding to *all the multineedle variations* (6.72) of the reference optimal control $\bar{u}(\cdot)$:

$$S := \Big\{(y_0, \ldots, y_m) \in I\!R^{m+1}\Big|\ y_0 = \sum_{j=1}^{M}\sum_{i=1}^{N_j} \Big\langle \nabla\varphi_0\big(\bar{x}(b)\big), \Lambda_{\tau_j,v_{ij},\alpha_{ij}}\bar{x}(b)\Big\rangle, \ldots\ , \\ y_m = \sum_{j=1}^{M}\sum_{i=1}^{N_j} \Big\langle \nabla\varphi_m\big(\bar{x}(b)\big), \Lambda_{\tau_j,v_{ij},\alpha_{ij}}\bar{x}(b)\Big\rangle\Big\} \tag{6.74}$$

with arbitrary $\tau_j \in [a,b)$, $v_{ij} \in U$, $\alpha_{ij} \in [0,1]$, $i = 1,\ldots,N_j$, $N_j \in I\!N$, $j = 1,\ldots,M$, and $M \in I\!N$.

There are *two crucial facts* regarding the set S in (6.74). First of all, it happens to be *convex*, which is mainly due to the possibility of using arbitrary $\alpha_{ij} \in [0,1]$ in multineedle variations (6.72). The latter is based on the *time continuity* of $[a,b]$ and, as mentioned above, reflects the *hidden convexity* of continuous-time control systems. The second fact is due to the *optimality* of $\bar{u}(\cdot)$ in the constrained control problem (6.63), (6.64): it ensures that the linearized image set (6.74) *doesn't intersect* the convex set of *forbidden points* (from the viewpoint of optimality and inequality constraints in the problem under consideration) given by

$$I\!R^{m+1}_{<} := \left\{ (y_0,\ldots,y_m) \in I\!R^{m+1} \,\middle|\, y_i < 0 \ \text{ for all } \ i = 0,\ldots,m \right\} .$$

Both of these facts are proved in the following lemma.

Lemma 6.44 (hidden convexity and primal optimality condition in control problems with inequality constraints). *Let* $\{\bar{u}(\cdot),\bar{x}(\cdot)\}$ *be an optimal solution to the inequality constrained problem* (6.63) *and* (6.64)*, where all the functions* φ_i *are supposed to be Fréchet differentiable at* $\bar{x}(b)$ *in addition to the standing assumptions of Subsect.* 6.3.1*. Then the linearized image set* S *in* (6.74) *is convex and doesn't intersect the set of forbidden points* $I\!R^{m+1}_{<}$.

Proof. Let us fix a collection of parameters (τ_i, v_{ij}, N_j, M) and show that the set (6.74), still denoted by S, is *convex* while the numbers α_{ij} are arbitrarily taken from $[0,1]$. This clearly implies the convexity of the "full" set S. Indeed, taking two different collections of (τ_i, v_{ij}, N_j, M), we may always unify them, which again gives an admissible multineedle variation (6.72). It is therefore sufficient to justify the convexity of S only in the case when parameters α_{ij} take values on the interval $[0,1]$.

To proceed, we fix (τ_i, v_{ij}, N_j, M) and take two collections $\{\alpha_{ij}^{(1)}\}$ and $\{\alpha_{ij}^{(2)}\}$ such that the corresponding points $y^{(1)}$ and $y^{(2)}$ in (6.74) belong to the linearized image set S. Then considering the point $\lambda y^{(1)} + (1-\lambda)y^{(2)}$ for any $\lambda \in [0,1]$ and taking into account the linear dependence of $\Lambda\bar{x}_{\tau_j,v_{ij},\alpha_{ij}}(b)$ on α_{ij}, we conclude that $\lambda y^{(1)} + (1-\lambda)y^{(2)}$ is an element of S corresponding to $\{\lambda\alpha_{ij}^{(1)} + (1-\lambda)\alpha_{ij}^{(2)}\}$, which justifies the convexity of S.

It remains to show that $S \cap I\!R^{m_1}_{<} = \emptyset$, where S stands for the "full" image set in (6.74) corresponding to all the admissible multineedle variations (6.72). Assuming the contrary, we find a multineedle variation (6.72) with some admissible parameters $(\tau_i, v_{ij}, \alpha_{ij}, N_j, M)$ such that

$$\sum_{j=1}^{M}\sum_{i=1}^{N_j}\Big\langle\nabla\varphi_0\big(\bar{x}(b)\big),\,\varDelta_{\tau_j,v_{ij},\alpha_{ij}}\bar{x}(b)\Big\rangle<0,\dots\,,$$

$$\sum_{j=1}^{M}\sum_{i=1}^{N_j}\Big\langle\nabla\varphi_m\big(\bar{x}(b)\big),\,\varDelta_{\tau_j,v_{ij},\alpha_{ij}}\bar{x}(b)\Big\rangle<0\,.$$

Then using the Fréchet differentiability of the functions $\varphi_0,\dots,\varphi_m$ at $\bar{x}(b)$ and the above relationship between the endpoint increment $\varDelta\bar{x}(b)$ generated by (6.72) and the linearized ones $\varDelta_{\tau_j,v_{ij},\alpha_{ij}}$ corresponding to each collection $(\tau_j,v_{ij},\alpha_{ij},N_j,M)$, we get

$$\begin{aligned}\varphi_k\big(x(b)\big)-\varphi_k\big(\bar{x}(b)\big)&=\big\langle\nabla\varphi_k\big(\bar{x}(b)\big),\,\varDelta\bar{x}(b)\big\rangle+o(\varepsilon)\\&=\Big[\sum_{j=1}^{M}\sum_{i=1}^{N_j}\Big\langle\nabla\varphi_k\big(\bar{x}(b)\big),\,\varDelta_{\tau_j,v_{ij},\alpha_{ij}}\bar{x}(b)\Big\rangle\Big]\varepsilon+o(\varepsilon)<0\end{aligned}$$

for all $k=0,\dots,m$ and all $\varepsilon>0$ sufficiently small. The latter means that there is a multineedle control variation (6.72) such that the corresponding trajectory $x(\cdot)$ satisfies all the inequality constraints (6.64), being therefore *feasible* for the problem under consideration, and gives a smaller value to the cost functional in (6.63) in comparison with $\bar{x}(\cdot)$. This contradicts the optimality of the process $\{\bar{u}(\cdot),\bar{x}(\cdot)\}$ in problem (6.63), (6.64) and thus completes the proof of the lemma. △

The obtained relation $S\cap I\!R^{m+1}_{<}=\emptyset$ can be viewed as a *primal* necessary optimality condition, which is of course not efficient, since it depends on control variations and is not expressed in terms of the initial data of the problem under consideration. To proceed further, we pass to its *dual* form employing the *convex separation* theorem and then invoking the Hamilton-Pontryagin function by the constructions of the increment method in Lemma 6.42; see the arguments below.

Proof of Theorem 6.37 for problems with inequality constraints. Applying the classical *separation theorem* to the convex sets S and $I\!R^{m+1}_{<}$ from Lemma 6.44, we find a *nonzero* vector $(\lambda_0,\dots,\lambda_m)\in I\!R^{m+1}$ such that

$$\sum_{i=0}^{m}\lambda_i y_i\ge\sum_{i=0}^{m}\lambda_i z_i\quad\text{for all}\quad(y_0,\dots,y_m)\in S\quad\text{and}\quad(z_0,\dots,z_m)\in I\!R^{m+1}_{<}\,.$$

This easily implies that $\lambda_i\ge 0$ for all $i=0,\dots,m$ and that

$$\sum_{i=0}^{m}\lambda_i y_i\ge 0\quad\text{whenever}\quad(y_0,\dots,y_m)\in S\,.\tag{6.75}$$

Note that the vector $(\lambda_0,\dots,\lambda_m)$ *doesn't depend* on a specific multineedle variation (6.72); it separates the set of all such variations from $0\in I\!R^{m+1}$. In particular, employing (6.75) just for vectors $(y_0, \ldots, y_m)$ generated by *single* needle variations (6.71) with parameters (τ, v, ε) and taking into account the relationship (6.73) between the full and linearized increments of the optimal trajectory along (single) needle variations, one has

$$\sum_{i=0}^{m} \lambda_i \Big\langle \nabla\varphi_i\big(\bar{x}(b)\big), \varDelta_{\tau,v,\varepsilon}\bar{x}(b)\Big\rangle + o(\varepsilon) \geq 0$$

for all $\tau \in [a, b)$, $v \in U$, and $\varepsilon > 0$ sufficiently small. Putting now

$$p(b) := -\sum_{i=0}^{m} \lambda_i \nabla\varphi_i\big(\bar{x}(b)\big)$$

and proceeding as in the proof of Lemma 6.42 and Theorem 6.37 for the free-endpoint control problem in Subsect. 6.3.2 with the replacement of the boundary condition (6.70) by the latter one, we end the proof of Theorem 6.37 for problems with inequality endpoint constraints. △

6.3.4 Transversality Conditions for Problems with Equality Constraints

To complete the proof of Theorem 6.37, it remains to justify it for the case of *equality* endpoint constraints in the problem under consideration. Without loss of generality we focus here on the optimal control problem given by (6.63) and (6.65), i.e., with no inequality constraints considered in the preceding subsection. For convenience, suppose that the equality constraints are given by the first m functions φ_i as

$$\varphi_i\big(x(b)\big) = 0, \quad i = 1, \ldots, m \; . \tag{6.76}$$

Having this in mind, form again the *linearized image set* S in (6.74) generated now by the images of multineedle variations under the gradient mappings for the cost and *equality* constraint functions. The set of *forbidden points* in the equality constrained problem is given by

$$S^< := \Big\{(y_0, \ldots, y_m) \in I\!R^{m+1} \Big|\; y_0 < 0, y_1 = 0, \ldots, y_m = 0\Big\} \; .$$

Our goal is to investigate all the possible relationships between the image set S and the above set of forbidden points that are allowed by the optimality of $\{\bar{u}(\cdot), \bar{x}(\cdot)\}$. The most difficult case is considered in the next lemma, which establishes that the origin cannot be an *interior point* of the intersection $S \cap S^<$. The proof given below involves the *Brouwer fixed-point theorem.* Note that, although this fundamental topological result is heavily finite-dimensional, it allows us to deal with the optimal control problems described by evolution equations in *infinite dimensions.* The crux of the matter is, as mentioned, that the control problem has *finitely many* endpoint constraints, which ensures the *finite codimension* property of the constraint set.

Lemma 6.45 (endpoint variations under equality constraints). *Let* $\{\bar{u}(\cdot), \bar{x}(\cdot)\}$ *be an optimal solution to the control problem* (6.63), (6.76) *under the standing assumptions on* X, U, *and* f. *Assume also that the functions* $\varphi_0, \ldots, \varphi_m$ *are Fréchet differentiable at* $\bar{x}(b)$ *and that* $\varphi_1, \ldots, \varphi_m$ *are in addition continuous around this point. Then one has*

$$0 \notin \operatorname{int}\big(\operatorname{proj}_{I\!R^m} S\big) ,$$

where the linearized image set S *is generated in* (6.74) *by the endpoint equality constraints* (6.76).

Proof. Assume the contrary and denote by B_η a closed ball in $I\!R^m$ of radius $\eta > 0$ centered at the origin. Let $\mathcal{T}$ be a regular "tetrahedron" with the vertices $q^{(s)}$, $s = 1, \ldots, m+1$, inscribed into $\mathcal{T}$. If η is sufficiently small, then for each $s = 1, \ldots, m+1$ there are numbers $\{\alpha_{ij}^{(s)}\}$ in the multineedle variation (6.72) and $\nu < 0$ such that

$$\begin{cases} \displaystyle\sum_{j=1}^{M}\sum_{i=1}^{N_j}\Big\langle \nabla\varphi_0\big(\bar{x}(b)\big), \varLambda_{\tau_j, v_{ij}, \alpha_{ij}^{(s)}}\bar{x}(b)\Big\rangle < \nu < 0 \ \text{ and} \\[2ex] \displaystyle q_k^{(s)} = \sum_{j=1}^{M}\sum_{i=1}^{N_j}\Big\langle \nabla\varphi_k\big(\bar{x}(b)\big), \varLambda_{\tau_j, v_{ij}, \alpha_{ij}^{(s)}}\bar{x}(b)\Big\rangle \end{cases}$$

for all $k = 1, \ldots, m$, where $q_k^{(s)}$ stands for the kth component of the vertex $q^{(s)}$. Each point $q = q(\beta) \in \mathcal{T}$ can be represented as a *convex combination* of the tetrahedron vertices by

$$q(\gamma) = \sum_{s=1}^{m+1} \gamma_s q^{(s)} \ \text{ with } \ \gamma = (\gamma_1, \ldots, \gamma_{m+1}) \in P ,$$

where P connotes the m-dimensional simplex. Let $u_{\gamma,\varepsilon}(\cdot)$ be a multineedle variation (6.72) with the parameters $(\tau_j, v_{ij}, \alpha_{ij}(\gamma), \varepsilon)$, where

$$\alpha_{ij}(\gamma) := \sum_{s=1}^{m+1} \gamma_s \alpha_{ij}^{(s)}, \quad \gamma = (\gamma_1, \ldots, \gamma_m) \in P .$$

Consider now an ε-parametric family of mappings $g(\cdot, \varepsilon)\colon P \to I\!R^m$ defined by

$$g(\gamma, \varepsilon) := \Big(\frac{\varphi_1\big(x_{\gamma,\varepsilon}(b)\big) - \varphi_1\big(\bar{x}(b)\big)}{\varepsilon}, \ldots, \frac{\varphi_m\big(x_{\gamma,\varepsilon}(b)\big) - \varphi_m\big(\bar{x}(b)\big)}{\varepsilon}\Big) ,$$

where $x_{\gamma,\varepsilon}(\cdot)$ signifies a trajectory for (6.61) corresponding to the multineedle control variation $u_{\gamma,\varepsilon}(\cdot)$. Putting also

$$g(\gamma,0):=\Big(\sum_{j=1}^{M}\sum_{i=1}^{N_j}\Big\langle\nabla\varphi_1\big(\bar x(b)\big),\alpha_{ij}(\gamma)\Lambda_{\tau_j,v_{ij},1}\bar x(b)\Big\rangle,\ldots\ ,$$

$$\sum_{j=1}^{M}\sum_{i=1}^{N_j}\Big\langle\nabla\varphi_m\big(\bar x(b)\big),\alpha_{ij}(\gamma)\Lambda_{\tau_j,v_{ij},1}\bar x(b)\Big\rangle\Big)\ ,$$

we conclude that the mapping $g(\cdot,\cdot)$ is *continuous* on $P\times[0,\varepsilon_0]$ with ε_0 sufficiently small. This is due to the standing assumptions on the Fréchet differentiability of $\varphi_1,\ldots,\varphi_m$ *at* $\bar x(b)$ and the continuity of these functions *around* this point. It follows from the above constructions that

$$g(\gamma,0)=\sum_{s=1}^{m+1}\gamma_s q^{(s)}\ \text{ and }\ G(P,0)=\mathcal{T}\ ;$$

thus the set $g(P,0)$ contains the origin as an interior point. Let us show that there is $\widehat\varepsilon>0$ such that

$$0\in\operatorname{int}g(P,\varepsilon)\ \text{ for all }\ \varepsilon<\widehat\varepsilon\ .$$

To proceed, we observe that the mapping $g(\cdot,0)$ is one-to-one and continuous from P into $\mathcal{T}$. Hence its *inverse mapping* is single-valued and continuous; let us denote it by $p(y)$ and put

$$h(y,\varepsilon):=g\big(p(y),\varepsilon\big)\ \text{ for all }\ y\in\mathcal{T}\ \text{ and }\ \varepsilon\in[0,\varepsilon_0]\ .$$

Take $\eta>0$ so small that the ball B_η of radius η centered at the origin belongs to the tetrahedron $\mathcal{T}$. Then the continuity of the mapping $h(\cdot,\cdot)$ yields the existence of $\widehat\varepsilon>0$ such that

$$\|h(y,0)-h(y,\varepsilon)\|<\eta\ \text{ whenever }\ \varepsilon<\widehat\varepsilon\ .$$

Thus, given any $\varepsilon\in(0,\widehat\varepsilon)$, the continuous mapping $h(y,0)-h(y,\varepsilon)$ *transforms the ball* B_η *into itself*. Employing the *Brouwer fixed-point theorem*, we find a point $y^\varepsilon\in B_\eta$ satisfying

$$h(y^\varepsilon,0)-h(y^\varepsilon,\varepsilon)=y^\varepsilon\ \text{ for all }\ \varepsilon\in(0,\widehat\varepsilon)\ .$$

This implies by $h(y,0)\equiv y$ that

$$h(y^\varepsilon,\varepsilon)=g\big(p(y^\varepsilon),\varepsilon\big)=g(\gamma^\varepsilon,\varepsilon)\ \text{ for some }\ \gamma^\varepsilon\in P\ \text{ with }\ g(\gamma^\varepsilon,0)=y^\varepsilon\ .$$

Taking into account the construction of $g(\cdot,\cdot)$, we conclude that the trajectories $x_{\gamma^\varepsilon,\varepsilon}(\cdot)$ generated by the multineedle variations $u_{\gamma^\varepsilon,\varepsilon}(\cdot)$ under consideration *satisfy the equality constraints* (6.76) for all $\varepsilon\in(0,\widehat\varepsilon)$. Moreover, for the variations along the cost functional one has

$$\sum_{j=1}^{M}\sum_{i=1}^{N_j}\Big\langle\nabla\varphi_0\big(\bar{x}(b)\big),\,\Lambda_{\tau_j,v_{ij},\alpha_{ij}(\gamma^\varepsilon)}\bar{x}(b)\Big\rangle$$

$$=\sum_{s=1}^{m+1}\gamma_s^\varepsilon\Big(\sum_{j=1}^{M}\sum_{i=1}^{N_j}\Big\langle\nabla\varphi_0\big(\bar{x}(b)\big),\,\Lambda_{\tau_j,v_{ij},\alpha_{ij}^{(s)}}\bar{x}(b)\Big\rangle\Big)$$

$$<\sum_{s=1}^{m+1}\gamma_s^\varepsilon\nu<\nu\quad\text{whenever}\quad\varepsilon\in(0,\widehat{\varepsilon})\ .$$

The latter implies, similarly to the case of inequality constraints, that

$$\varphi_0\big(x_{\gamma^\varepsilon,\varepsilon}(b)\big)<\varphi_0\big(\bar{x}(b)\big)$$

along some *feasible solutions* to the equality constrained problem (6.63), (6.65). This contradicts the *optimality* of the process $\{\bar{u}(\cdot),\bar{x}(\cdot)\}$ in this problem and completes the proof of the lemma. △

Based on Lemma 6.45 and the arguments developed in Subsects. 6.3.2 and 6.3.3, we finally justify Theorem 6.37 in the remaining case of equality constraints and thus complete the whole proof of this theorem.

Proof of Theorem 6.37 for problems with equality constraints. Taking into account Lemma 6.45, there are the following two possible relationships between the linearized image set S in (6.72) corresponding the equality constraints (6.76) and the set of forbidden points $S^<$:

(a) $S\cap S^<=\emptyset$;

(b) $S\cap S^<\neq\emptyset$ and $0\in\text{bd}\big(\text{proj}_{I\!R^m}S\big)$.

Consider first case (a). Since both sets S and $S^<$ are convex, we employ the classical *separation theorem* for convex sets and find a *nonzero* vector $(\lambda_0,\ldots,\lambda_m)\in I\!R^{m+1}$ such that

$$\sum_{i=0}^{m}\lambda_i y_i\ge\sum_{i=0}^{m}\lambda_i z_i\quad\text{for all}\quad(y_0,\ldots,y_m)\in S\quad\text{and}\quad(z_0,\ldots,z_m)\in S^<\ .$$

It easily implies, by the structure of the forbidden set $S^<$, that $\lambda_0\ge 0$ and that the relation (6.75) holds. To complete the proof of the theorem in this case, we now proceed exactly as in the case of inequality constraints at the very end of Subsect. 6.3.3.

It remains to examine case (b). Denote $\Omega:=\text{proj}_{I\!R^m}S$ and observe that this set is closed and convex in $I\!R^m$. Since $0\in\text{bd}\,\Omega$, we apply the supporting hyperplane theorem for convex sets and find a nonzero m-vector $(\lambda_1,\ldots,\lambda_m)$ supporting Ω at the origin. Then we again arrive at the basic relation (6.75) with the nontrivial $(m+1)$-vector $(0, \lambda_1, \ldots, \lambda_m)$ and complete the proof of the theorem similarly to the case of inequality constraints. △

Note that the *continuity* assumption on the *equality constraint* functions φ_i *around* $\bar{x}(b)$, an addition to their Fréchet differentiability at this point, is essential for the validity of Theorem 6.37 even in the case of finite-dimensional state space X with the trivial dynamics $f = 0$; see Example 5.12.

6.4 Approximate Maximum Principle in Optimal Control

This section is devoted to optimal control problems for a *parametric family* of dynamical systems governed by *discrete approximations* of control systems with continuous time. Discrete/finite-difference approximations play a prominent role in both qualitative and numerical aspects of optimal control. While considered as a *process* with a decreasing step of discretization, they occupy an *intermediate position* between continuous-time control systems and discrete-time control systems with fixed steps. Recall that discrete approximations of general control problems for differential inclusions have been studied in Sect. 6.1, but the attitude there was different from that in this section. Our previous direction was *from discrete to continuous*: to establish necessary optimality conditions for discrete-time systems with *fixed* discretization steps and then to use well-posed discrete approximations as a *vehicle* in deriving optimality conditions for continuous-time control systems. The results obtained in this way in Sect. 6.1 provide necessary conditions of a *maximum principle* type only under some *convexity/relaxation* assumptions imposed *a priori* on the system dynamics.

Now we are going to explore the other direction in the relationship between discrete-time and continuous-time control systems: *from continuous to discrete*. Having in mind that the *Pontryagin maximum principle* (PMP) and its extensions to nonsmooth problems and differential inclusions hold *without any convexity/relaxation* assumptions on the continuous-time dynamics, it is challenging to clarify the possibility to establish necessary optimality conditions of the *maximum principle type* for discrete approximations. The results obtained in this direction are *rather surprising*; see below.

6.4.1 Exact and Approximate Maximum Principles for Discrete-Time Control Systems

As seen in Sects. 6.2 and 6.3, the relations of the maximum principle involving the Weierstrass-Pontryagin maximum condition hold for continuous-time control systems with *no* a priori *convexity* assumptions. This happens due to specific features of the continuous-time dynamics that generates some *hidden convexity* property inherent in such control systems. Probably the most striking and deep manifestation of the hidden convexity for continuous-time systems is given by the fundamental Lyapunov theorem on the range convexity of *nonatomic/continuous* vector measures, which is equivalent to the Aumann convexity theorem for set-valued integrals; see, e.g., the discussion in the proof of Lemma 6.18 and the references therein. In the proof of the maximum principle for control systems with smooth dynamics given in Sect. 6.3 we didn't invoke these results while exploiting *directly the time continuity* in the construction of needle (and multineedle) variations generating the *automatic convexity* of the linearized image set as in Lemma 6.44. One cannot expect such properties for *discrete-time* systems described by the general discrete inclusions of the type

$$x(t+1) \in F(x(t),t), \quad t = 0,\ldots,K-1\ ,$$

or by their parameterized control representations

$$x(t+1) = f\big(x(t),u(t),t\big), \quad u(t) \in U, \quad t = 0,\ldots,K-1\ ,$$

where $K \in I\!N$ signifies the number of steps (final discrete time) for the discrete dynamic process. However, the *discrete maximum principle* holds if the sets of "discrete velocities" $F(x,t)$, or their counterparts $f(x,U,t)$ for the parameterized control systems, are *assumed to be convex*. In this case the maximum condition is actually a *direct consequence* of the *Euler-Lagrange inclusion* as discussed above. Indeed, it follows from the extremal property of the coderivative to convex-valued mappings from Theorem 1.34 due to a special representation of the normal cone to *convex* sets.

As well known, the discrete maximum principle *may not hold*, even for simple control systems with smooth dynamics, if the above velocity sets are *not convex*. We now present an example of the failure of the discrete maximum principle (as a natural analog of the Pontryagin maximum principle for discrete-time control systems) for a *family* of simple free-endpoint problems with smooth dynamics. In this example the Hamilton-Pontryagin function achieves its *global minimum* (instead of maximum) along *any* feasible control. As always in this chapter, a "free-endpoint" problem means that there are no constraints on the right endpoint of the system trajectories, while the left endpoint may be fixed.

Example 6.46 (failure of the discrete maximum principle). *There is a family of optimal control problems of minimizing a linear function over two-dimensional discrete-time control systems with smooth dynamics and no endpoint constraints such that any feasible control for these problems doesn't satisfy the discrete maximum principle.*

Proof. Consider the following family of optimal control problems with a two-dimensional state vector $x = (x_1, x_2) \in I\!R^2$:

$$\begin{cases} \text{minimize } \; J[u,x] = \varphi\big(x(K)\big) := x_2(3) \;\text{ subject to} \\[2ex] x_1(t+1) = \vartheta(u(t),t), \quad x_1(0) = 0\,, \\[2ex] x_2(t+1) = \gamma\,\big(x_1(t)\big)^2 + \eta x_2(t) - (\gamma/\eta)\big(\vartheta(u(t),t)\big)^2, \quad x_2(0) = 0\,, \\[2ex] u(t) \in U, \quad t = 0,1,2\,, \end{cases}$$

where the scalar function $\vartheta(\cdot,\cdot)$, the numbers γ, η, and the control set U are arbitrary. Then (a natural discrete counterpart of) the Hamilton-Pontryagin function for this system is

$$\begin{aligned} H\big(x(t), p(t+1), u, t\big) :&= \big\langle p(t+1), f\big(x(t), u, t\big)\big\rangle \\ &= p_1(t+1)\vartheta(u,t) + \gamma\, p_2(t+1)\big(x_1(t)\big)^2 \\ &\quad + \eta p_2(t+1) x_2(t) - (\gamma/\eta)\big(\vartheta(u,t)\big)^2\,, \end{aligned}$$

where the adjoint trajectory $p(\cdot)$ satisfies the corresponding discrete analog of the system (6.67) given by

$$p(t) = \nabla_x H\big(x(t), p(t+1), u(t), t\big), \quad t \in \{0,\ldots,K-1\} = \{0,1,2\}\,,$$

with the boundary/transversality condition

$$p(K) = -\nabla\varphi\big(x(K)\big) = (0,-1) \;\text{ at }\; K = 3\,.$$

For the problem under consideration one has

$$\begin{aligned} &p_2(3) = -1, \quad p_2(2) = -\eta, \quad p_2(1) = -\eta^2\,, \\ &p_1(3) = 0, \quad p_1(2) = -\gamma\, x_1(2) = -2\gamma\,\vartheta\big(u(1),1\big)\,, \\ &p_1(1) = -2\gamma\,\eta x_1(1) = -2\gamma\,\eta\vartheta\big(u(0),0\big)\,. \end{aligned}$$

Then considering only the terms depending on u in the Hamilton-Pontryagin function, we get

$$\begin{aligned} &H(u,0) = -\gamma\,\eta\big[2\vartheta\big(u(0),0\big)\vartheta(u,0) - \big(\vartheta(u,0)\big)^2\big]\,, \\ &H(u,1) = -\gamma\,\big[2\vartheta\big(u(1),1\big)\vartheta(u,1) - \big(\vartheta(u,1)\big)^2\big]\,. \end{aligned}$$

This shows that, given an arbitrary $\vartheta(\cdot,\cdot)$ and U, the functions $H(u,0)$ and $H(u,1)$ attain their *global minimum* at any $u(0)$ and $u(1)$ whenever $\gamma > 0$ and $\gamma\,\eta > 0$, respectively. Thus the above relationships of the discrete maximum principle are *not necessary for optimality* in the family of optimal control problems under consideration. △

It is worth mentioning that the Hamilton-Pontryagin function in the above example *does attain its global maximum* over $u \in U$ for optimal controls when

$t = K - 1 = 2$. This can be shown by using the increment formula applied to *concave* cost functionals along needle variations of optimal controls; cf. the arguments below in Subsect. 6.4.2. Moreover, the discrete maximum principle *holds true* in the family of problems from Example 6.46 for *all* t, i.e., it provides necessary optimality conditions along optimal controls at every time moment, *if and only if*

$$\gamma \leq 0 \text{ and } \eta \geq 0 .$$

This follows from the above consideration and the results of Sect. 17 in Mordukhovich's book [901], where some *individual conditions* for the validity of the discrete maximum principle are given. Thus the simultaneous fulfillment of the conditions $\gamma \leq 0$ and $\eta \geq 0$ *fully describes* the relationships between the initial data of the problems from Example 6.46, which ensure the fulfillment of the discrete maximum principle. Note that overall the results in this direction obtained in the afore-mentioned book [901] strongly take into account *interconnections* between the *initial* data of *nonconvex* discrete-time control systems; see more discussions and examples therein.

The main attention in this section is paid not to optimal control problems governed by dynamical systems with *fixed* discrete time but to *finite-difference/discrete approximations* of continuous-time problems studied in the preceding section. This means that instead of the continuous-time control system (6.61) we consider a *sequence* of its finite-difference analogs given by

$$\begin{cases} x_N(t+h_N) = x_N(t) + h_N f\big(x_N(t), u_N(t), t\big), \quad x_N(a) = x_0 \in X , \\ u(t) \in U, \quad t \in T_N := \big\{a, a+h_N, \ldots, b-h_N\big\} , \end{cases} \tag{6.77}$$

with $N \in I\!N$ and $h_N := (b-a)/N$. Recall that discrete approximations of differential/evolution inclusions have been studied in Sect. 6.1 being used there as a vehicle to derive necessary optimality conditions for continuous-time control problems. Now our goal is quite opposite: to look at optimal control problems for discrete approximations from the viewpoint of their continuous-time counterparts. **The key question is**:

Would it be possible to obtain a certain natural analog of the Pontryagin maximum principle for optimal control problems governed by nonconvex finite-difference systems of type (6.77) *as* $N \to \infty$?

If the answer is *no*, then such a potential *instability* of the PMP may pose *serious challenges* to its implementation in any numerical algorithm involving finite-difference approximations of time derivatives.

To begin with, for each $N \in I\!N$ we consider the problem of minimizing a smooth endpoint function $\varphi_0\big(x(b)\big)$ over discrete-time process $\{u_N(\cdot), x_N(\cdot)\}$ satisfying (6.77). The *exact* PMP analog for each of these problems, the *discrete maximum principle*, is written as follows: given an optimal process $\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\}$, there is an adjoint arc $p_N(\cdot)$, $t \in T_N \cup \{b\}$, satisfying

$$p_N(t) = p_N(t+h_N) + h_N \nabla_x H\big(\bar{x}_N(t), p_N(t+h_N), \bar{u}_N(t), t\big) \qquad (6.78)$$

as $t \in T_N$ with the transversality condition

$$p_N(b) = -\nabla\varphi_0\big(\bar{x}_N(b)\big) \qquad (6.79)$$

and such that the *exact maximum condition*

$$H\big(\bar{x}_N(t), p_N(t+h_N), \bar{u}_N(t), t\big) = \max_{u \in U} H\big(\bar{x}_N(t), p_N(t+h_N), u, t\big), \quad t \in T_N \ .$$

is valid whenever $N \in I\!N$, with the usual Hamilton-Pontryagin function

$$H(x, p, u, t) := \big\langle p, f(x, u, t)\big\rangle \ .$$

It follows from Example 6.46 (via standard rescaling) and the discussion above that this (*exact*) *discrete maximum principle* may be generally *violated* even for simple classes of optimal control problems governed by discrete approximation systems of type (6.77) whenever $N \in I\!N$. This may signify a possible instability of the PMP under discrete approximations. Note, however, that to require the fulfillment of such an *exact* counterpart of the PMP for discrete approximation systems is *too much* to ensure the PMP stability under discretization of continuous-time control systems.

What we really need for this purpose is the validity, along *every sequence* of optimal solutions $\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\}$ to the discrete approximation problems while $N \in I\!N$ becomes sufficiently large, of the *approximate maximum condition*

$$H\big(\bar{x}_N(t), p_N(t+h_N), \bar{u}_N(t), t\big) = \max_{u \in U} H\big(\bar{x}_N(t), p_N(t+h_N), u, t\big) + \varepsilon(t, h_N)$$

for all $t \in T_N$ with some $\varepsilon_N(t, h_N) \to 0$ as $N \to \infty$ uniformly in $t \in T_N$, where $p_N(\cdot)$ are the corresponding adjoint trajectories satisfying (6.78) and (6.79). In this case we say that the *approximate maximum principle* (AMP) holds for the discrete approximation problems under consideration. Such an approximate analog of the PMP ensures the *discretization stability* of the latter and thus justifies the possibility to employ the PMP in computer calculations and simulations of nonconvex continuous-time control systems. Furthermore, giving necessary optimality conditions for sequences of discrete approximation problems, the AMP plays *essentially the same role* as the (exact) discrete maximum principle in solving discrete-time control problems with sufficiently *small steps*; see particularly Example 6.68. However, in the case of large stepsizes h the approximate maximum condition, still being necessary for optimality, may be *far removed* from the exact maximum.

It is proved in Subsect. 6.4.3 that the *AMP holds*, with $\varepsilon(h_N, t) = O(h_N)$ in arbitrary Banach state spaces X, for *smooth free-endpoint* problems of optimal control, i.e., for problems of minimizing smooth (continuously differentiable) cost functions over discrete approximation systems (6.77) with smooth dynamics and no endpoint constraints. The proof is purely analytic based on using (single) needle control variations and a discrete counterpart of the increment formula from Subsect. 6.3.2.

The *crucial difference* between the PMP for continuous-time systems and the AMP for discrete approximations is that the latter result *doesn't have* an expected (lower) *subdifferential analog* for optimal control problems involving the simplest *nonsmooth* (even convex) cost functions! The corresponding counterexample is presented in Subsect. 6.4.3, together with those showing the violation of the AMP for optimal control problems with Fréchet *differentiable* (but *not continuously* differentiable) cost functions as well as for control problems with *nonsmooth dynamics*.

Thus the AMP happens to be very *sensitive to nonsmoothness.* On the other hand, in Subsect. 6.4.3 we derive an *upper subdifferential* version of the AMP, parallel to that in Subsect. 6.3.1 for continuous-time systems, which holds however for a more restrictive class of cost functions in comparison with the one for continuous-time systems. This class of *uniformly upper subdifferentiable* functions is introduced and studied in Subsect. 6.4.2.

The case of optimal control problems for discrete approximation systems (6.77) *with endpoint constraints* is much more involved. Considering control systems with *smooth inequality constraints* of the type

$$\varphi_i\big(x_N(b)\big) \le 0, \quad i = 1,\ldots,m\ ,$$

we formulate in Subsect. 6.4.4 the AMP with *perturbed complementary slackness conditions* under some *properness* assumption on the sequence of optimal controls, which can be treated as a discrete counterpart of piecewise continuity. The latter assumption happens to be essential for the validity of the AMP for nonconvex constrained systems as demonstrated by an example. The proof of the AMP given in Subsect. 6.4.5 reveals an *approximate counterpart* of the *hidden convexity* property for finite-difference control problems under consideration; see below for more details and discussions. We also derive the *upper subdifferential* form of the AMP for inequality constrained problems with uniformly upper subdifferentiable endpoint functions φ_i, $i = 0,\ldots,m$.

A proper setup for discrete approximations of continuous-time control problems with endpoint constraints of the *equality type*

$$\varphi_i\big(x(b)\big) = 0, \quad i = m+1,\ldots,m+r\ ,$$

involves the *constraint perturbations*

$$\big|\varphi_i\big(x_N(b)\big)\big| \le \xi_{iN}, \quad i = m+1,\ldots,m+r\ ,$$

with $\xi_{iN} \downarrow 0$ as $N \to \infty$. It is proved in Subsect. 6.4.5 that the *AMP holds* for discrete approximation problems with perturbed equality constraints described by smooth functions provided that the following *consistency condition*

$$\limsup_{N\to\infty} \frac{h_N}{\xi_{iN}} = 0 \ \text{ for all } \ i = m+1,\ldots,m+r \ . \tag{6.80}$$

is imposed. This means that the equality constraint perturbations ξ_{iN} should tend to zero *slower* than the discretization stepsize h_N, which particularly requires that $\xi_{iN} \neq 0$. We give an example showing the consistency condition (6.80) is *essential* for the fulfillment of the AMP, which may be violated even when $\xi_{iN} = O(h_N)$.

The results obtained admit an extension to discrete approximations of systems with *time delays* in state variables, which relates to the case of *incommensurability* between the length $b-a$ of the time interval and the approximation stepsize h_N; see Subsect. 6.4.6. On the other hand, we present an example showing the *AMP doesn't hold* for discrete approximations of *neutral systems*, even in the case of smooth free-endpoint control problems.

Before deriving the mentioned results on the AMP, let us describe and study the class of *uniformly upper subdifferentiable* functions on Banach spaces for which the *upper subdifferential form* of the AMP will be developed. This class particularly includes every continuously differentiable function as well as every concave continuous function that are of special interest for applications.

6.4.2 Uniformly Upper Subdifferentiable Functions

The main object of this subsection is the class of functions defined as follows.

Definition 6.47 (uniform upper subdifferentiability). *A real-valued function defined on a Banach space X is* UNIFORMLY UPPER SUBDIFFERENTIABLE *around a point $\bar{x}$ if for every x from some neighborhood V of $\bar{x}$ there exists a nonempty set $\mathcal{D}^+\varphi(x) \subset X^*$ described by: for any given $\varepsilon > 0$ there is $\nu > 0$ such that $x^* \in \mathcal{D}^+\varphi(x)$ if and only if*

$$\varphi(v) - \varphi(x) - \langle x^*, v - x\rangle \le \varepsilon \|v - x\| \tag{6.81}$$

whenever $v \in V$ with $\|v - x\| \le \nu$ and $x^* \in \mathcal{D}^+\varphi(x)$.

It is easy to see that this class contains every *smooth* (i.e., $\mathcal{C}^1$ around $\bar{x}$)) function with $\mathcal{D}^+\varphi(x) = \{\nabla\varphi(x)\}$ and also every *concave* continuous function with $\mathcal{D}^+\varphi(x) = \partial^+\varphi(x)$ as x is around $\bar{x}$ in *any Banach space*. Furthermore, one can derive from the definition that the above class is closed with respect to taking the *minimum* over compact sets. Note that even if φ is Lipschitz continuous around $\bar{x}$ and Fréchet differentiable *at* $\bar{x}$, it may *not* be uniformly upper subdifferentiable around this point. A simple example is provided by the standard function $\varphi\colon I\!R \to I\!R$ defined by $\varphi(x) := x^2 \sin(1/x)$ for $x \neq 0$ and $\varphi(0) := 0$ with $\bar{x} = 0$.

Before formulating the main result of this subsection, we consider an arbitrary function $\varphi\colon X \to \overline{I\!R}$ finite at $\bar{x}$ and describe relationships between the *Fréchet upper subdifferential* of φ at $\bar{x}$ defined in (1.52) by

$$\widehat{\partial}^+\varphi(\bar{x}) := \left\{x^* \in X^* \middle| \ \limsup_{x\to\bar{x}} \frac{\varphi(x)-\varphi(\bar{x})-\langle x^*, x-\bar{x}\rangle}{\|x-\bar{x}\|} \le 0\right\}$$

and the two modifications of the so-called *Dini* (or Dini-Hadamard) *upper directional derivative* of φ at $\bar{x}$ defined by

$$d^+(\bar{x}; z) := \limsup_{\substack{y\to z\\ t\downarrow 0}} \frac{\varphi(\bar{x}+ty)-\varphi(\bar{x})}{t}$$

for the standard (strong) version and by

$$d^+_w(\bar{x}; z) := \limsup_{\substack{y\xrightarrow{w} z\\ t\downarrow 0}} \frac{\varphi(\bar{x}+ty)-\varphi(\bar{x})}{t}$$

for its *weak* counterpart, where $y \xrightarrow{w} z$ signifies the weak convergence in X. The next proposition used below is definitely interesting for its own sake; it reveals the *duality* between the subgradient and directional derivative constructions under consideration that generally holds in *reflexive* spaces for the *weak* directional derivative and in *finite dimensions* for the *strong* one. We formulate it for the case of upper constructions needed in this section; it readily implies the lower counterpart.

Proposition 6.48 (relationships between Fréchet subgradients and Dini directional derivatives). *One always has*

$$\widehat{\partial}^+\varphi(\bar{x}) \subset \left\{x^* \in X^* \middle| \ \langle x^*, z\rangle \ge d^+_w\varphi(\bar{x}; z) \ \text{ for all } \ z \in X\right\}$$

$$\subset \left\{x^* \in X^* \middle| \ \langle x^*, z\rangle \ge d^+\varphi(\bar{x}; z) \ \text{ for all } \ z \in X\right\},$$

where the equality holds in the first inclusion when X is reflexive, while it holds in the second one when $\dim X < \infty$*. Moreover,*

$$d^+\varphi(\bar{x}; z) = \limsup_{t\downarrow 0} \frac{\varphi(\bar{x}+tz)-\varphi(\bar{x})}{t} \tag{6.82}$$

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

Proof. To prove the final inclusion in the proposition, it is sufficient to observe that for every $x^* \in \widehat{\partial}^+\varphi(\bar{x})$ and $z \in X$ one has

$$d^+\varphi(\bar{x}; z) - \langle x^*, z\rangle = \|z\| \cdot \limsup_{\substack{y\to z\\ t\downarrow 0}} \frac{\varphi(\bar{x}+ty)-\varphi(\bar{x})-t\langle x^*, y\rangle}{t} \le 0\,;$$

the other is similar. Let us prove that the first inclusion holds as equality if X is reflexive. To proceed, we pick $x^* \notin \widehat{\partial}^+\varphi(\bar{x})$ and take any $\gamma > 0$. Then there is a sequence $x_k \to \bar{x}$ such that

$$\varphi(x_k) - \varphi(\bar{x}) - \langle x^*, x_k - \bar{x}\rangle - \gamma \|x_k - \bar{x}\| > 0 \ \text{ for all } \ k \in I\!N \ .$$

Since X is reflexive, we suppose without loss of generality that the sequence $(x_k - \bar{x})/\|x_k - \bar{x}\|$ weakly converges to some $z \in X$. Then

$$d^+\varphi(\bar{x}; z) \geq \limsup_{k\to\infty} \frac{\varphi(x_k) - \varphi(\bar{x})}{\|x_k - \bar{x}\|} \geq \langle x^*, z\rangle + \gamma \ ,$$

which ensures the required equality, since γ was chosen arbitrarily.

It remains to justify representation (6.82) if φ is locally Lipschitzian around $\bar{x}$ with some modulus $\ell > 0$. Then we get

$$|\varphi(\bar{x} + ty) - \varphi(\bar{x} + tz)| \leq t\ell\|y - z\| \ \text{ for any } \ y, z \in X$$

when $t > 0$ is sufficiently small. Thus one has

$$d^+\varphi(\bar{x}; z) = \limsup_{\substack{y\to z\\ t\downarrow 0}} \Big[\frac{\varphi(\bar{x} + tz) - \varphi(\bar{x})}{t} + \frac{\varphi(\bar{x} + ty) - \varphi(\bar{x} + tz)}{t}\Big]$$

$$= \limsup_{t\downarrow 0} \frac{\varphi(\bar{x} + tz) - \varphi(\bar{x})}{t} \ \text{ whenever } \ z \in X \ ,$$

which justifies (6.82) and completes the proof of the proposition. △

Now we are ready to establish important properties of uniformly upper subdifferentiable functions that are employed in what follows being certainly of independent interest. It shows, in particular, that such functions enjoy the *upper regularity* property formulated right after Definition 1.91.

Theorem 6.49 (properties of uniformly upper subdifferentiable functions). *Let X be reflexive, and let φ be continuous at $\bar{x}$ and uniformly upper subdifferentiable around this point with the subgradient sets $\mathcal{D}^+\varphi(x)$ from Definition* 6.47*. Then there is a neighborhood of $\bar{x}$ in which φ is Lipschitz continuous and one can choose*

$$\mathcal{D}^+\varphi(x) = \widehat{\partial}^+\varphi(x) = \partial^+\varphi(x) \ .$$

Proof. The subgradient sets $\mathcal{D}^+\varphi(x)$ are obviously convex. Moreover, it is easy to check that each of these sets is norm-closed in X^* and hence also weakly closed due to its convexity and the assumed reflexivity of X. Let us show that $\mathcal{D}^+\varphi(x)$ is *uniformly bounded* in X^* around $\bar{x}$. Assume the contrary and select some sequences $x_k \to \bar{x}$ and $x_k^* \in \mathcal{D}^+\varphi(x_k)$ with $\|x_k^*\| \to \infty$ as $k \to \infty$. Then employing the Hahn-Banach theorem and taking into account the reflexivity of X, we find $u_k \in X$ satisfying the relations

$$\langle x_k^*, u_k\rangle = \|x_k^*\|^{1/2} \ \text{ and } \ \|u_k\| = \|x_k^*\|^{-1/2} \ \text{ for all } \ k \in I\!N \ .$$

Setting now $v_k := x_k + u_k$, one has from (6.81) that

$$\varphi(v_k) - \varphi(x_k) \le -\langle x_k^*, u_k \rangle + \varepsilon \|u_k\|$$

with $\|u_k\| \to 0$ and $\langle x_k^*, u_k \rangle \to \infty$ by the construction above. This yields that $\varphi(v_k) - \varphi(x_k) \to -\infty$ while $x_k, v_k \to \bar{x}$ as $k \to \infty$, which contradicts the required continuity of φ at $\bar{x}$ and thus justifies the uniform boundedness of $\mathcal{D}^+\varphi(x)$ around this point.

Next we show that φ is *locally Lipschitzian* around $\bar{x}$. It can be done similarly to the proof of Theorem 3.52 based on the mean value inequality from Theorem 3.49 that holds for $\mathcal{D}^+\varphi(\cdot)$. However, we may easier proceed directly invoking the uniform boundedness of the sets $\mathcal{D}^+\varphi(x)$ around $\bar{x}$ and property (6.81). Indeed, assume the contrary and find sequences $x_k \to \bar{x}$ and $v_k \to \bar{x}$ satisfying

$$|\varphi(v_k) - \varphi(x_k)| > k\|v_k - x_k\| \quad \text{as} \quad k \to \infty\ .$$

Suppose for definiteness that $\varphi(v_k) - \varphi(x_k) > k\|v_k - x_k\|$; the other case is symmetric. Now using the uniform upper subdifferentiability of φ, we find a sequence of $x_k^* \in \mathcal{D}^+\varphi(x_k)$ satisfying

$$\begin{aligned} k\|v_k - x_k\| &< \varphi(v_k) - \varphi(x_k) \le \langle x_k^*, v_k - x_k \rangle + \varepsilon \|v_k - x_k\| \\ &\le \big(\|x_k^*\| + \varepsilon\big)\|v_k - x_k\| \end{aligned}$$

for any given $\varepsilon > 0$ when k is sufficiently large. This yields that $\|x_k^*\| \to \infty$ as $k \to \infty$, which contradicts the uniform boundedness of the sets $\mathcal{D}^+\varphi(x)$ around $\bar{x}$ and thus justifies the local Lipschitzian property of φ.

It follows from the definition of Fréchet upper subgradients in (1.52) and the construction of $\mathcal{D}^+\varphi(x)$ in (6.81) that one always has $\mathcal{D}^+\varphi(x) \subset \widehat{\partial}^+\varphi(x)$. Let us show in fact that $\mathcal{D}^+\varphi(x) = \widehat{\partial}^+\varphi(x)$ around $\bar{x}$. First observe that the set-valued mapping $\mathcal{D}^+\varphi\colon V \rightrightarrows X^*$ is *closed-graph* in the norm$\times$weak topology of $X \times X^*$ on any closed subset of V. Using this fact and the local Lipschitz continuity of φ around $\bar{x}$, we derive from (6.81) that φ is *directionally differentiable* in the classical sense

$$\varphi'(x; z) := \lim_{t \downarrow 0} \frac{\varphi(\bar{x} + tz) - \varphi(\bar{x})}{t}, \quad z \in X\ ,$$

whenever x is sufficiently close to $\bar{x}$; moreover, we have the representation

$$\varphi'(x; z) = \min\big\{\langle x^*, z \rangle \big|\ x^* \in \mathcal{D}^+\varphi(x)\big\}\ , \tag{6.83}$$

where the minimum is attained due to the weak closedness of $\mathcal{D}^+\varphi(x)$ in X^*. Since $\mathcal{D}^+\varphi(x)$ is also convex, one gets from (6.83) and the results of Proposition 6.48 that $\widehat{\partial}^+\varphi(x) \subset \mathcal{D}^+\varphi(x)$. Indeed, assuming the opposite and then separating $x^* \notin \mathcal{D}^+\varphi(x)$ from the convex and norm-closed set $\mathcal{D}^+\varphi(x) \subset X^*$, we arrive at a contradiction with (6.82) and (6.83). Finally, the equality $\mathcal{D}^+\varphi(x) = \partial^+\varphi(x)$ and the upper regularity of φ around $\bar{x}$ follows from the mention closed-graph property of $\mathcal{D}^+\varphi(\cdot)$ by the upper subdifferential version of Theorem 2.34 on the limiting representation of basic subgradients. This completes the proof of the theorem. △

As mentioned above, properties of uniformly upper subdifferentiable functions allow us to derive the AMP in optimal control problems for discrete approximations with *upper subdifferential* transversality conditions; see the following subsections. This requires more from the functions and spaces under consideration in comparison with the assumptions needed to justify upper subdifferential transversality conditions in the PMP for continuous-time systems as well as upper subdifferential optimality conditions in problems of mathematical programming; cf. Sects. 5.1, 5.2, and 6.3. These significantly more restrictive requirements needed for the AMP are due to the *parametric* nature of finite-difference systems treated as a *process* as $N \to \infty$. We'll see in the next subsection that, even in the case of *differentiable* cost functions in free-endpoint control problems with finite-dimensional state spaces, the *continuity of the derivatives* is essential for the validity of the AMP in sequences of discrete approximations.

6.4.3 Approximate Maximum Principle for Free-Endpoint Control Systems

This subsection is devoted to optimal control problems for sequences of finite-difference systems (6.77) with *no endpoint constraints* on the right-hand end of trajectories. As in the case of continuous-time systems, free-endpoint problems for discrete approximations are essentially different from their constrained counterparts. The main *positive result* of this subsection is the *approximate maximum principle* for free-endpoint problems in Banach spaces with *upper subdifferential* transversality conditions valid for uniformly upper subdifferentiable cost functions. In particular, this justifies the AMP for control problems with continuously differentiable cost functions, where the boundary/transversality condition for the adjoint system (6.78) is written in the classical form (6.79). On the other hand, we present an example showing that the AMP *doesn't hold* when the cost function is *differentiable at* the point of interest but *not* $\mathcal{C}^1$ *around* it. Other examples show that the AMP is very *sensitive to nonsmoothness*: it doesn't hold for control problems with nonsmooth dynamics and—which is even more striking—for nice systems with *convex* nonsmooth cost functions.

Consider the *sequence* of optimal control problems (P_N^0) for discrete-time systems studied in this subsection:

$$\text{minimize} \;\; J_N[u_N, x_N] := \varphi_0\big(x_N(b)\big) \tag{6.84}$$

over control-trajectory pairs $\{u_N(\cdot), x_N(\cdot)\}$ satisfying the control system (6.77) as $N \to \infty$. Given a sequence of optimal solutions $\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\}$ to problems (P_N^0), we impose the following *standing assumptions*:

—the control space U is metric, the state space X is Banach;

—there is an open set O containing $\bar{x}_N(t)$ for all $t \in T_N \cup \{b\}$ such that f is Fréchet differentiable in x with both $f(x,u,t)$ and its state derivative $\nabla_x f(x,u,t)$ continuous in (x,u,t) and uniformly norm-bounded whenever $x \in O$, $u \in U$, and $t \in T_N \cup \{b\}$ as $N \to \infty$;

—the sequence $\{\bar{x}_N(b)\}$ belongs to a compact subset of X.

The latter assumption is not restrictive at all in finite dimensions: it follows from standard conditions ensuring the uniform boundedness of admissible trajectories for continuous-time control systems. In infinite dimensions it can be derived from the conditions imposed in (H1) of Subsect. 6.1.1; cf. the proof of Theorem 6.13 and the references therein.

Here is the main *positive* result of this subsection.

Theorem 6.50 (AMP for free-endpoint control problems with upper subdifferential transversality conditions). *Let the pairs $\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\}$ be optimal to problems (P_N^0) under the standing assumptions made. Suppose in addition that the cost function φ_0 is uniformly upper subdifferentiable around the limiting point(s) of the sequence $\{\bar{x}_N(b)\}$ with the corresponding subgradient sets $\mathcal{D}^+(x)$. Then for every sequence of upper subgradients $x_N^* \in \mathcal{D}^+\varphi_0\big(\bar{x}(b)\big)$ there is $\varepsilon(t,h_N) \to 0$ as $N \to \infty$ uniformly in $t \in T_N$ such that one has the approximate maximum condition*

$$H\big(\bar{x}_N(t), p_N(t+h_N), \bar{u}_N(t), t\big) = \max_{u \in U} H\big(\bar{x}_N(t), p_N(t+h_N), u, t\big) \\ +\varepsilon(t,h_N), \quad t \in T_N\ , \tag{6.85}$$

where each $p_N(\cdot)$ is the corresponding trajectory for the adjoint system (6.78) *with the boundary/transversality condition*

$$p_N(b) = -x_N^* \ \text{ for all } \ N \in I\!N\ . \tag{6.86}$$

Furthermore, this result holds with any $x_N^* \in \widehat{\partial}^+\varphi\big(\bar{x}_N(b)\big)$ in (6.86) *if in addition X is reflexive and φ_0 is continuous at the optimal points.*

Proof. Considering a sequence of optimal solutions $\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\}$ to (P_N^0), we suppose that the trajectories $\bar{x}_N(t)$ belong to the uniform neighborhoods fixed in the assumptions made for all $N \in I\!N$. It follows from Definition 6.47 of the uniform upper subdifferentiability for φ_0 that $\mathcal{D}^+\varphi_0\big(\bar{x}_N(b)\big) \neq \emptyset$ and that inequality (6.81) holds for any $x_N^* \in \mathcal{D}^+\varphi_0\big(\bar{x}_N(b)\big)$ as $N \to \infty$. Now taking an arbitrary sequence of $x_N^* \in \mathcal{D}^+\varphi_0\big(\bar{x}_N(b)\big)$, we get

$$\varphi_0(x) - \varphi_0\big(\bar{x}_N(b)\big) \le \langle x_N^*, x - \bar{x}_N(b)\rangle + o\big(\|x - \bar{x}_N(b)\|\big) \tag{6.87}$$

$$\text{with } \frac{o\big(\|x-\bar{x}_N(b)\|\big)}{\|x-\bar{x}_N(b)\|} \to 0 \text{ as } x \to x_N(b) \text{ } \textit{uniformly} \text{ in } N .$$

Letting $p_N(b) := -x_N^*$ as in (6.86), we derive from (6.87) that

$$J[u_N, x_N] - J[\bar{u}_N, \bar{x}_N] \le -\big\langle p_N(b),\, \varDelta x_N(b)\big\rangle + o\big(\|\varDelta x_N(b)\|\big) ,$$

with $\varDelta x_N(t) := x_N(t) - \bar{x}_N(t)$, for all admissible processes in (P_N^0) whenever $x_N(b)$ is sufficiently close to $\bar{x}_N(b)$. Taking into account the identity

$$\begin{aligned}\big\langle p_N(b),\, \varDelta x_N(b)\big\rangle &= \sum_{t\in T_N} \Big\langle p_N(t+h_N) - p_N(t),\, \varDelta x_N(t)\Big\rangle \\ &\quad + \sum_{t\in T_N} \Big\langle p_N(t+h_N),\, \varDelta x_N(t+h_N) - \varDelta x_N(t)\Big\rangle\end{aligned}$$

and using the smoothness of f in x, we get from the above inequality that

$$\begin{aligned}0 \le J[u_N, x_N) - J[\bar{u}_N, \bar{x}_N] &\le -\sum_{t\in T_N} \Big\langle p_N(t+h_N) - p_N(t),\, \varDelta x_N(t)\Big\rangle \\ -h_N \sum_{t\in T_N} &\Big\langle p_N(t+h_N), \nabla_x f\big(\bar{x}_N(t), \bar{u}_N(t), t\big)\, \varDelta x_N(t)\Big\rangle \\ -h_N \sum_{t\in T_N} &\varDelta_u H\big(\bar{x}_N(t), p_N(t+h_N), \bar{u}_N(t), t\big) \\ -h_N \sum_{t\in T_N} &\eta_N(t) + o\big(\|\varDelta x_N(b)\|\big) ,\end{aligned} \tag{6.88}$$

where the remainder $\eta_N(t)$ is computed by

$$\begin{aligned}\eta_N(t) = \Big\langle &\nabla_x H\big(\bar{x}_N(t), p_N(t+h_N), u_N(t), t\big) \\ &-\nabla_x H\big(\bar{x}_N(t), p_N(t+h_N), \bar{u}_N(t), t\big),\, \varDelta x_N(t)\Big\rangle + o\big(\|\varDelta x_N(t)\|\big)\end{aligned}$$

with the quantity $o(\|\varDelta x_N(t)\|)$ being uniform in N due to the assumptions on $\nabla_x f$, and where the increment $\varDelta_u H$ is defined similarly to the one in Subsect. 6.3.2 for continuous-time systems.

Now we consider (single) *needle variations* of the optimal controls $\bar{u}_N(\cdot)$ in the following form:

$$u_N(t) = \begin{cases} v & \text{if } t = \tau \ , \\ \bar{u}_N(t) & \text{if } t \in T_N \setminus \{\tau\} \ , \end{cases}$$

where $v \in U$ and $\tau = \tau(N) \in T_N$ as $N \in I\!N$. All these controls are obviously feasible for the discrete approximation problems under consideration, which are not subject to endpoint constraints. The trajectory increments corresponding to the needle variations satisfy the relations

$$\Delta x_N(t) = 0 \;\text{ for }\; t = a, \ldots, \tau; \quad \|\Delta x_N(t)\| = O(h_N) \;\text{ for }\; t = \tau + h_N, \ldots, b\,.$$

Taking this into account and substituting the needle variations $u_N(\cdot)$ into the increment inequality (6.88), one gets

$$0 \le J[u_N, x_N] - J[\bar{u}_N, \bar{x}_N] \le -h_N \Delta_u H\big(\bar{x}_N(\tau), p_N(\tau + h_N), \bar{u}_N(\tau), \tau\big) + o(h_N)\,.$$

Arguing by contradiction, we directly derive from the latter inequality the approximative maximum condition (6.85).

To complete the proof of the theorem, it remains to apply Theorem 6.49 on uniform upper subdifferentiability to the cost function φ_0. This ensures that x_N^* may be taken from the whole Fréchet upper subgradient sets $\widehat{\partial}^+\varphi_0\big(\big(\bar{x}(b)\big)$ in the transversality conditions (6.86) as $N \to \infty$ provided that X is reflexive and that φ_0 is assumed to be continuous a priori. △

Remark 6.51 (discrete approximations versus continuous-time systems.) Observe that the proof of Theorem 6.50 is similar to the one for continuous-time systems with free endpoints; cf. the proofs of Theorem 6.37 in Subsect. 6.3.2 and of its upper subdifferential version (Theorem 6.38) in Subsect. 6.3.1. The given proofs in both continuous-time and discrete-time settings are based on using the *increment formulas* for cost functionals and (*single*) *needle variations* of optimal controls. In a sense, the proof for discrete approximations problems is a simplified version of that given for systems with continuous time (which is definitely not the case when endpoint constraints are involved; see the next subsection). On the other hand, there are two *significant differences* between the results and proofs for continuous-time systems and those for discrete approximations.

Firstly, in the case of continuous-time systems there is a possibility of using a *small parameter* ε as the length of needle variations, which ensures the smallness of trajectory increments $\Delta x(t) = O(\varepsilon)$ and happens to be *crucial* for establishing the *exact* maximum principle in continuous-time optimal control. In systems of discrete approximations the smallness of trajectory increments is provided by the *decreasing stepsize* h_N, which is a parameter of the problem but not of variations. This leads to the *approximate* maximum condition with the error as small as the step of discretization. Of course, such a device is not possible when $h_N \not\to 0$.

The *second* difference concerns the *parametric nature* of discrete approximation problems in contrast to problems with continuous time. This requires the more restrictive *uniformity* assumptions imposed on cost functions in comparison with the case of continuous-time systems.

The following two consequences of Theorem 6.50 and its proof deal with important classes of cost functions that are automatically uniformly upper subdifferentiable and admit *more precise* versions of the AMP. Note that these results don't require the reflexivity assumption on the state space X as in the second part of Theorem 6.50; they are valid in *arbitrary Banach spaces.*

Corollary 6.52 (AMP for free-endpoint control problems with smooth cost functions). *Let the pairs $\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\}$ be optimal to problems (P_N^0) under the standing assumptions made. Suppose in addition that the cost function φ_0 is continuously differentiable around the limiting point(s) of $\{\bar{x}_N(b)\}$. Then the approximate maximum principle of Theorem* 6.50 *holds with the transversality condition* (6.79) *for the corresponding adjoint trajectory $p_N(\cdot)$ whenever $N \in I\!N$. Moreover, we can take $\varepsilon(t, h_N) = O(h_N)$ in the maximum condition* (6.85) *if both $\nabla_x f(\cdot, u, t)$ and $\nabla \varphi_0(\cdot)$ are locally Lipschitzian around $\bar{x}_N(\cdot)$ uniformly in $u \in U$, $t \in T_N$, and $N \to \infty$.*

Proof. As mentioned above, in any Banach space X we have $\mathcal{D}^+\varphi(x) = \{\nabla\varphi(x)\}$ in a neighborhood of $\bar{x}$ if φ is $\mathcal{C}^1$ around this point. It can be easily shown that (6.87) holds as equality for smooth functions φ_0; moreover, one has $|o(\eta)| \le \ell\eta^2$ therein if $\nabla\varphi_0$ is locally Lipschitzian. Note further that the Lipschitzian assumption imposed on $\nabla_x f(\cdot, u, t)$ in the corollary implies that

$$o\big(\|\varDelta x_N(t)\|\big) = O\big(\|\varDelta x_N(t)\|^2\big)$$

uniformly in N for the "o" term in the remainder $\eta_N(\cdot)$ in the proof of the theorem. This yields that $\varepsilon(t, h_N) = O(h_N)$ in the approximate maximum condition (6.85) under the assumptions made. △

Corollary 6.53 (AMP for free-endpoint control problems with concave cost functions). *Let the pairs $\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\}$ be optimal to problems (P_N^0) under the standing assumptions made. Suppose in addition that the cost function φ_0 is concave on some open set containing all $\bar{x}_N(b)$. Then the approximate maximum principle of Theorem* 6.50 *holds along every sequence of subgradients $x_N^* \in \partial^+\varphi_0\big(\bar{x}_N(b)\big)$. Moreover, one can take $\varepsilon(t, h_N) = O(h_N)$ in* (6.85) *if $\nabla_x f(\cdot, u, t)$ is locally Lipschitzian around $\bar{x}_N(\cdot)$ uniformly in $u \in U$, $t \in T_N$, and $N \to \infty$.*

Proof. Recall that $\mathcal{D}^+\varphi(x) = \partial^+\varphi(x)$ for concave continuous functions in arbitrary Banach spaces. Furthermore, $o\big(\|x - \bar{x}_N(b)\|\big) \equiv 0$ in the inequality (6.87) under the concavity assumption of the corollary. Combining this with the estimate of $\eta_N(\cdot)$ in the proof of Corollary 6.52, we conclude that $\varepsilon(t, h_N) = O(h_N)$ in (6.85) under the assumptions made. △

Now we proceed with *counterexamples*, i.e., examples showing that the AMP may be violated if some of the assumptions in Theorem 6.50 are not satisfied. All the examples below are given for finite-dimensional control systems with nonconvex velocity sets. Our first example demonstrates that the

AMP doesn't hold in the *expected lower subdifferential form* (as the maximum principle for continuous-time control systems) even in the simplest nonsmooth case of minimizing convex functions over systems with linear dynamics.

Example 6.54 (AMP may not hold for linear control systems with nonsmooth and convex minimizing functions). *There is a one-dimensional control problem of minimizing a nonsmooth and convex cost function over a linear system with no endpoint constraints for which the AMP is violated.*

Proof. Consider the following sequence of one-dimensional optimal control problem (P_N^0), $N \in I\!N$, for discrete-time systems:

$$\begin{cases} \text{minimize } \ \varphi\big(x_N(1)\big) := |x_N(1) - \vartheta| \\ \\ \text{subject to} \\ \\ x_N(t + h_N) = x_N(t) + h_N u_N(t), \quad x_N(0) = 0\ , \\ \\ u_N(t) \in U := \big\{0, 1\big\}, \quad t \in T_N := \big\{0, h_N, \ldots, 1 - h_N\big\}\ , \end{cases} \tag{6.89}$$

where ϑ is a positive *irrational* number less than 1 whose choice will be specified below. The dynamics in (6.89) is a discretization of the simplest ODE control system $\dot{x} = u$. Observe that, since ϑ is irrational and h_N is rational, we have $\bar{x}_N(1) \neq \vartheta$ for the endpoint of an optimal trajectory to (6.89) as $N \in I\!N$, while obviously $\bar{x}(1) = \vartheta$ for optimal solutions to the continuous-time counterpart. It is also clear that for all sufficiently small stepsizes h_N an optimal control to (6.89) is neither $u_N(t) \equiv 0$ nor $u_N(t) \equiv 1$, but it has at least one point of *control switch*.

Suppose that for some subsequence $N_k \to \infty$ one has $\bar{x}_{N_k}(1) > \vartheta$; put $\{N_k\} = I\!N$ without loss of generality. Let us show that in this case the approximate maximum condition *doesn't hold* at points $t \in T_N$ for which $\bar{u}_N(t) = 1$. Indeed, we have

$$H\big(\bar{x}_N(t), p_N(t + h_N), u\big) = p_N(t + h_N)u \quad \text{and} \quad p_N(t) \equiv -1$$

for the Hamilton-Pontryagin function and the adjoint trajectory for this problems, since $\bar{x}_N(1) > \vartheta$ along the optimal solution to (6.89). Thus

$$\max_{u \in U} H\big(\bar{x}_N(t), p_N(t + h_N), u\big) = 0, \quad t \in T_N\ ,$$

$$\text{while } \ H\big(\bar{x}_N(s), p_N(s + h_N), \bar{u}_N(s)\big) = -1$$

at the points $s \in T_N$ of control switch, where $\bar{u}_N(s) = 1$ regardless of h_N.

Let us specify the choice of ϑ in (6.89) ensuring that $\bar{x}_N(1) > \vartheta$ along some subsequence of natural numbers. We claim that $\bar{x}_N(1) > \vartheta$ if $\vartheta \in (0, 1)$ is an irrational number whose decimal representation contains infinitely many digits from the set $\{5,6,7,8,9\}$; e.g., $\vartheta = 0.676676667\ldots$. Indeed, put $h_N := 10^{-N}$, which is a subsequence of $h_N = N^{-1}$ as required in (6.89). It is easy to see that in this case the set of all reachable points at $t = 1$ is the set of rational numbers between 0 and 1 with exactly N digits in the fractional part of their decimal representations. In particular, for $N = 3$ this set is $\{0, 0.001, 0.002, \ldots, 0.999, 1\}$. Therefore, by the construction of ϑ, the closest point to ϑ from the reachable set is greater than ϑ, and this point must be the endpoint of the optimal trajectory $\bar{x}_N(1)$. △

The next example, complemented to Example 6.54, shows that the AMP fails even for problems with *differentiable* but *not continuously differentiable* cost functions.

Example 6.55 (AMP may not hold for linear systems with differentiable but not $\mathcal{C}^1$ cost functions). *There is a one-dimensional control problem of minimizing a Fréchet differentiable but not continuously differentiable cost function over a linear system with no endpoint constraints for which the AMP is violated.*

Proof. Consider the same control system as in (6.89) and construct a minimizing function $\varphi(x)$ that satisfies the requirements listed above. Let $\psi(x)$ be a $\mathcal{C}^1$ function with the properties:

$$\psi(x) \ge 0, \quad \psi(x) = \psi(-x), \quad \psi(x) \equiv 0 \ \text{ if } \ |x| > 2 ,$$

$$|\nabla\psi(x)| \le 1 \ \text{ for all } \ x, \quad \text{and} \ \ \nabla\psi(-1) = \vartheta > 0 .$$

Define the cost function $\varphi(x)$ by

$$\varphi(x) := \Big(x - \frac{1}{9}\Big)^2 + \sum_{n=1}^{\infty} 10^{-2n-3}\psi\Big(10^{2n+3}\big(x - \sum_{k=1}^{n} 10^{-k}\big) - 1\Big) ,$$

which is continuously differentiable around every point but $x = \frac{1}{9}$, where it is differentiable and attains its absolute minimum. As in Example 6.54, we put $h_N := 10^{-N}$, and then the point $x = \frac{1}{9}$ *cannot be reached* by discretization. It is not hard to check that the endpoint of the optimal trajectory $\bar{x}_N(\cdot)$ for each N is computed by

$$\bar{x}_N(1) = \sum_{k=1}^{N} 10^{-k} \quad \text{with} \quad \nabla\varphi\big(\bar{x}_N(1)\big) = \vartheta + \varepsilon_N ,$$

where $\varepsilon_N \downarrow 0$ as $N \to \infty$. Proceeding as in Example 6.54 with the same sequence of optimal controls, we have

$$H\big(\bar{x}_N(t), p_N(t + h_N), u\big) \equiv -\vartheta u + O(\varepsilon_N) ,$$

and the approximate maximum condition (6.85) doesn't hold at the points of control switch, where $\bar{u}_N(t) = 1$. △

The last example in this subsection concerns systems with *nonsmooth dynamics.* We actually consider a finite-difference analog of minimizing an integral functional subject to a one-dimensional control system, which is equivalent to a two-dimensional optimal control problem of the Mayer type. The discrete "integrand" in this problem is nonsmooth with respect to the state variable x; it happens to be continuously differentiable with respect to x *along* the optimal process $\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\}$ under consideration but *not uniformly* in N. Thus the example below demonstrates that the *uniform smoothness* assumption on f over an open set containing all the optimal trajectories $\bar{x}_N(\cdot)$ is essential for the validity of the AMP.

Example 6.56 (violation of AMP for control problems with nonsmooth dynamics). *The AMP doesn't hold in discrete approximations of a minimization problem for an integral functional over a one-dimensional linear control system with no endpoint constraints such that the integrand is linear with respect to the control variable while convex and nonsmooth with respect to the state one. Moreover, the integrand in this problem happens to be $\mathcal{C}^1$ with respect to the state variable along the sequence of optimal solutions to the discrete approximations (P_N^0) for all $N \in \mathbb{N}$ but not uniformly in N.*

Proof. First we consider the following continuous-time optimal control problem of the Bolza type:

$$\begin{cases} \text{minimize } \; J[u,x] := \displaystyle\int_0^b \big(u(t) + |x(t) - t^2/2|\big)\,dt \\[2ex] \text{subject to} \\[2ex] \dot{x} = tu, \quad x(0) = 0\,, \\[2ex] u(t) \in U := \big\{1, c\big\}, \quad 0 \le t \le b\,, \end{cases}$$

where the terminal time b and the number $c > 1$ will be specified below. It is obvious that the optimal control to this problem is $\bar{u}(t) \equiv 1$ and the corresponding optimal trajectory is $\bar{x}(t) = t^2/2$.

By discretization we get the sequence of finite-difference control problems:

$$\begin{cases} \text{minimize } \; J[u_N,x_N] := h_N \displaystyle\sum_{t \in T_N} \big(u_N(t) + |x_N(t) - t^2/2|\big) \\[2ex] \text{subject to} \\[2ex] x_N(t+h_N) = x_N(t) + h_N t u_N(t), \quad x_N(0) = 0\,, \\[2ex] u_N(t) \in U = \big\{1, c\big\}, \quad t \in T_N := \big\{0, \ldots, (N-1)h_N\big\}\,. \end{cases} \tag{6.90}$$

We first show that $\bar{u}_N(t) \equiv 1$ remains to be the (unique) optimal control to (6.90) if the stepsize h_N is sufficiently small and the numbers (b, c) are chosen appropriately. It is easy to check that the corresponding trajectory $\bar{x}(\cdot)$ is computed by

$$\bar{x}_N(t) = \frac{t^2}{2} - \frac{th_N}{2} \quad \text{for all} \quad N \in I\!N \ .$$

Then the value $\bar{J}_N$ of the cost functional at $\bar{u}_N(\cdot)$ equals

$$\bar{J}_N = b + h_N^2 \sum_{t \in T_N} \frac{t}{2} = b + \frac{b^2 h_N}{4} + o(h_N) \ .$$

If we replace $u_N(t) = 1$ by $u_N(t) = c$ at some point $t \in T_N$, then the increment of the summation $h_N \sum_{t \in T_N} u_N(t)$ equals $(c-1)h_N$. Hence the corresponding value of the cost functional is

$$J[u_N, x_N] = h_N \sum_{t \in T_N} u_N(t) + h_N \sum_{t \in T_N} |x_N(t) - t^2/2|$$

$$> h_N \sum_{t \in T_N} u_N(t) \geq b + (c-1)h_N$$

for any feasible control $u_N(t)$ to (6.90) different from $\bar{u}_N(t) \equiv 1$. Comparing the latter with $\bar{J}_N$, we conclude that the control $\bar{u}_N(t) \equiv 1$ is indeed *optimal* to (6.90) if $b^2/4 < c - 1$ and N is sufficiently large.

We finally show that for $b > 2$ and $c > b^2/4 + 1$ (e.g., for $b = 3$ and $c = 4$) the sequence of optimal controls $\bar{u}_N(t) \equiv 1$ *doesn't* satisfy the approximate maximum condition (6.85) at all points $t \in T_N$ sufficiently close to $t = b/2$. Compute the Hamilton-Pontryagin function as a function of $t \in T_N$ and of $u \in U$ at the optimal trajectory $\bar{x}_N(t)$ corresponding to the optimal control under consideration with the adjoint trajectory $p_N(t)$ for (6.78). Reducing (6.90) to the standard Mayer form and taking into account that $\bar{x}_N(t) < t^2/2$ for all $t \in T_N$ due to above formula for $\bar{x}_N(t)$, we get

$$H\big(\bar{x}_N(t), p_N(t+h_N), u, t\big) = tp_N(t+h_N)u - u - |\bar{x}_N(t) - t^2/2|$$

$$= \big(tp_N(t+h_N) - 1\big)u + \big(\bar{x}_N(t) - t^2/2\big) \ ,$$

where $p_N(t)$ satisfies the equation

$$p_N(t) = p_N(t+h_N) + h_N, \quad p_N(b) = 0 \ ,$$

whose solution is $p_N(t) = b - t$. Therefore one has

$$H\big(\bar{x}_N(t), p_N(t+h_N), u, t\big) = \big(t(b - t + h_N) - 1\big)u + O(h_N)$$

$$= \big(-t^2 + bt - 1\big)u + O(h_N) \ .$$

The multiplier $-t^2 + bt - 1$ is positive in the neighborhood of $t = b/2$ if its discriminant $b^2 - 4$ is positive. Thus $u = c$, but not $u = 1$, provides the maximum to the Hamilton-Pontryagin function around $t = b/2$ if h_N is sufficiently small, which justifies the claim of this example. △

Finally in this subsection, we give a modification of Theorem 6.50 in the general case of possible *incommensurability* of the time interval $b - a$ and the stepsize h_N; note that $b - a = Nh_N$ as $N \in I\!N$ in Theorem 6.50. This is particularly important for the extension of the AMP to finite-difference approximations of time-delay systems in Subsect. 6.4.5. For simplicity we use the notation

$$f(x_N, u_N, t) := f\big(x_N(t), u_N(t), t\big) \ .$$

Given the time interval $[a, b]$, define the grid T_N on $[a, b]$ by

$$T_N := \big\{a, a + h_N, \ldots, b - \widetilde{h}_N - h_N\big\}$$

$$\text{with } \ h_N := \frac{b-a}{N} \quad \text{and} \quad \widetilde{h}_N := b - a - h_N \left[\frac{b-a}{h_N}\right] ,$$

where $[z]$ stands for the greatest integer less than or equal to the real number z. The modified discrete approximation problems $(\widetilde{P}_N^0)$ are written as

$$\begin{cases} \text{minimize } J[u_N, x_N] := \varphi_0\big(x_N(b)\big) \ \text{ subject to} \\ x_N(t + h_N) = x_N(t) + h_N f(x_N, u_N, t), \quad t \in T_N, \quad x_N(a) = x_0 \in X \ , \\ x_N(b) = x_N(b - \widetilde{h}_N) + \widetilde{h}_N f(x_N, u_N, b - \widetilde{h}_N) \ , \\ u_N(t) \in U, \quad t \in T_N \ . \end{cases}$$

Theorem 6.57 (AMP for problems with incommensurability). *Let the pairs $\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\}$ be optimal to problems $(\widetilde{P}_N^0)$. In addition to the standing assumptions, suppose that φ_0 is uniformly upper subdifferentiable around the limiting point(s) of the sequence $\{\bar{x}_N(b)\}$, $N \in I\!N$. Then for every sequence of upper subgradients $x_N^* \in \widehat{\mathcal{D}}^+\varphi_0\big(\bar{x}_N(b)\big)$ there is $\varepsilon(t, h_N) \to 0$ as $N \to \infty$ uniformly in $t \in T_N$ such that the approximate maximum condition*

$$H(\bar{x}_N, p_N, \bar{u}_N, t) = \max_{u \in U} H(\bar{x}_N, p_N, u, t) + \varepsilon(t, h_N)$$

holds for all $t \in \widetilde{T}_N := T_N \cup \{b - \widetilde{h}_N\}$, where the Hamilton-Pontryagin function is defined by

$$H(\bar{x}_N, p_N, u, t) := \begin{cases} \big\langle p_N(t + h_N), f(\bar{x}_N, u, t)\big\rangle & \text{if } t \in T_N \ , \\ \big\langle p_N(t), f(\bar{x}_N, u, t - \widetilde{h}_N)\big\rangle & \text{if } t = b - \widetilde{h}_N \ , \end{cases}$$

and where each $p_N(\cdot)$ *satisfies the adjoint system*

$$\begin{cases} p_N(t) = p_N(t+h_N) + h_N \nabla_x f(\bar{x}_N, \bar{u}_N, t)^* \, p_N(t+h_N), \quad t \in T_N \ , \\ p_N(b-\widetilde{h}_N) = p_N(b) + \widetilde{h}_N \nabla_x f(b-\widetilde{h}_N, \bar{x}_N, \bar{u}_N, t)^* \, p_N(b) \end{cases}$$

with the transversality condition $p_N(b) = -x_N^*$. *Furthermore, specifications similar to the second part of Theorem* 6.50 *as well as Corollaries* 6.52 *and* 6.53 *are also fulfilled.*

Proof. It is similar to the proof of Theorem 6.50 and its corollaries with the following modification of the increment formula for the minimizing functional:

$$\begin{aligned} 0 &\le J[u_N, x_N] - J[\bar{u}_N, \bar{x}_N] \le -\big\langle p_N(b), \varDelta x_N(b)\big\rangle + o\big(\|\varDelta x_N(b)\|\big) \\ &= -\sum_{t \in T_N} \big\langle p_N(t+h_N) - p_N(t), \varDelta x_N(t)\big\rangle \\ &-\big\langle p_N(b) - p_N(b-\widetilde{h}_N), \varDelta x_N(b-\widetilde{h}_N)\big\rangle \\ &-h_N \sum_{t \in T_N} \big\langle p_N(t+h_N), \nabla f_x(\bar{x}_N, \bar{u}_N, t)\, \varDelta x_N(t)\big\rangle \\ &-\widetilde{h}_N \big\langle p_N(b), \nabla_x f(\bar{x}_N, \bar{u}_N, b-\widetilde{h}_N)\, \varDelta x_N(b-\widetilde{h}_N)\big\rangle \\ &-h_N \sum_{t \in \widetilde{T}_N} \varDelta_u H(\bar{x}_N, p_N, \bar{u}_N) + h_N \sum_{t \in \widetilde{T}_N} \eta_N(t) + o\big(\|\varDelta x_N(b)\|\big) \ , \end{aligned}$$

where $\varDelta_u H$ and $\eta_N(t)$ are defined similarly to the non-delay problems. Substituting the adjoint trajectory into this formula and using needle variations of the optimal control, we arrive at the conclusions of the theorem. △

6.4.4 Approximate Maximum Principle under Endpoint Constraints: Positive and Negative Statements

This subsection concerns discrete approximations of optimal control problems with endpoint constraints. Our primary goal here is to formulate the approximate maximum principle for discrete approximation problems under appropriate assumptions and to clarify whether these assumptions are essential for its validity; the proof of the AMP is given in the next subsection.

Constructing discrete approximations, it is natural to *perturb* endpoint constraints and to consider the following *sequence* of optimal control problems (P_N) for discrete-time systems:

$$\begin{cases} \text{minimize } \ J[u_N, x_N] := \varphi_0\big(x_N(b)\big) \ \text{ subject to} \\ x_N(t+h_N) = x_N(t) + h_N f\big(x_N(t), u_N(t), t\big), \quad x_N(a) = x_0 \in X \ , \\ u_N(t) \in U, \quad t \in T_N := \big\{a, a+h_N, \ldots, b-h_N\big\} \ , \\ \varphi_i\big(x_N(t_1)\big) \le \gamma_{iN}, \quad i = 1, \ldots, m \ , \\ |\varphi_i\big(x_N(t_1)\big)| \le \xi_{iN}, \quad i = m+1, \ldots, m+r \ , \\ h_N := \dfrac{b-a}{N}, \quad N = 1, 2, \ldots \ , \end{cases}$$

where $\gamma_{iN} \to 0$ and $\xi_{iN} \downarrow 0$ as $N \to \infty$ for all i. The main result of this subsection shows that, under standard smoothness assumptions on the initial data, the AMP holds for *proper* sequences of optimal controls to problems (P_N) with *arbitrary* perturbations of *inequality* constraints (in particular, one can put $\gamma_{iN} = 0$) while with *consistent* perturbations of *equality* constraints matched the step of discretization. Then we demonstrate that the mentioned properness and consistency requirements are *essential* for the validity of the AMP, and we also derive an appropriate *upper subdifferential* analog of the AMP for problems with nonsmooth cost and inequality constraint functions.

Throughout this subsection we keep the *standing assumptions* on the initial data listed in Subsect. 6.4.3 supposing in addition that the state space X is *finite-dimensional*, which is needed in the proofs below. Along with the conventional notation for the matrix product, we use the agreement

$$\prod_{j}^{k=i} A_k := \begin{cases} A_i A_{i-1} \cdots A_j & \text{if } i \ge j \ , \\ I & \text{if } i = j-1 \ , \\ 0 & \text{if } i < j-1 \ , \end{cases}$$

where i, j are any integers and where I stands as usual for the identity matrix.

As in the case of continuous-time systems, the proof of the AMP for problems (P_N) with endpoint constraints is essentially different and more involved in comparison with free-endpoint problems. Recalling the proof of Theorem 6.37 for continuous-time systems with inequality endpoint constraints in Subsect. 6.3.3, we observe that a crucial part of this proof is Lemma 6.44, which verifies that the linearized image set S in (6.74) is convex and doesn't intersect the set of forbidden points. These facts are definitely due to the time continuity reflecting the *hidden convexity* of continuous-time control systems. Note that the mentioned image set S in (6.74) is generated by *multineedle* variations of the optimal control the very construction of which in (6.82) is essentially based on the time continuity.

In what follows we establish a certain *finite-difference analog* of the hidden convexity property for control systems in (P_N) involving *convex hulls* of some linearized image sets S_N generated by *single needle* variations of optimal controls. We show that *small shifts* (up to $o(h_N)$) of these convex hulls don't intersect the set of forbidden points as $N \to \infty$. This basically leads, via the *convex separation* theorem, to the approximate maximum principle for problems (P_N) under endpoint constraints of the inequality type, with appropriately *perturbed complementary slackness* conditions.

Such a device (as well as any finite-difference counterparts of the construction in Subsect. 6.3.4) doesn't apply to problems (P_N) with *arbitrarily* perturbed *equality* constraints (in particular, when $\xi_N = 0$) for which the AMP is *generally violated*. Nevertheless, the complementary slackness conditions mentioned above allow us to derive a natural version of the AMP for problems (P_N) with *appropriately perturbed* equality constraints by reducing them to the case of inequalities.

Before formulating the main result of this subsection, we introduce an important notion specific for *sequences* of finite-difference control problems.

Definition 6.58 (control properness in discrete approximations). *Let $d(\cdot,\cdot)$ stand for the distance in the control space U is problems (P_N). We say that the sequence of discrete-time controls $\{u_N(\cdot)\}$ in (P_N) is* PROPER *if for every increasing subsequence $\{N\}$ of natural numbers and every sequence of mesh points $\tau_{\theta(N)} \in T_N$ satisfying*

$$\tau_{\theta(N)} = a + \theta(N)h_N \quad \text{as} \quad \theta(N) = 0,\ldots,N-1 \quad \text{and} \quad \tau_{\theta(N)} \to t \in [t_0,t_1]$$

one of the following properties holds:

$$\text{either} \quad d\big(u_N(\tau_{\theta(N)}),u_N(\tau_{\theta(N)+q})\big) \to 0 \quad \text{or} \quad d\big(u_N(\tau_{\theta(N)}),u_N(\tau_{\theta(N)-q})\big) \to 0$$

as $N \to \infty$ with any natural constant q.

The notion of properness for sequences of feasible controls in discrete approximation problems is a *finite-difference counterpart* of the piecewise continuity for continuous-time systems. It turns out that the situation when sequences of optimal controls are not proper in discrete approximations of constrained systems with nonconvex velocities is not unusual, and this leads to the violation of the AMP for standard problems with inequality constraints. Note that the properness assumption is *not needed* for the validity of the AMP in free-endpoint problems; see Theorem 6.50.

Now we are ready to formulate the AMP for constrained control problems (P_N) with endpoint constraints described by smooth functions.

Theorem 6.59 (AMP for control problems with smooth endpoint constraints). *Let the pairs $\{\bar{u}_N(\cdot),\bar{x}_N(\cdot)\}$ be optimal to (P_N) for all $N \in \mathbb{N}$ under the standing assumptions made. Suppose in addition that all the functions φ_i, $i = 0,\ldots,m+r$, are continuously differentiable around the limiting point(s) of $\{\bar{x}_N(b)\}$ and that:*

(a) *the sequence of optimal controls* $\{\bar{u}_N(\cdot)\}$ *is proper;*

(b) *the consistency condition* (6.80) *holds for the perturbations* ξ_{iN} *of all the equality constraints.*

Then there are numbers $\{\lambda_{iN} \big| \ i = 0, \ldots, m + r\}$ *satisfying*

$$\lambda_{iN}\big(\varphi_i(\bar{x}_N(b)) - \gamma_{iN}\big) = O(h_N), \quad i = 1, \ldots, m \ , \tag{6.91}$$

$$\lambda_{iN} \geq 0, \ \ i = 0, \ldots, m, \qquad \sum_{i=0}^{m+r} \lambda_{iN}^2 = 1 \ , \tag{6.92}$$

and such that the approximate maximum condition (6.85) *is fulfilled with* $\varepsilon_N(t, h_N) \to 0$ *uniformly in* $t \in T_N$ *as* $N \to \infty$*, where each* $p_N(t)$, $t \in T_N \cup \{b\}$*, is the corresponding trajectory of the adjoint system* (6.78) *with the endpoint transversality condition*

$$p_N(b) = -\sum_{i=0}^{m+r} \lambda_{iN} \nabla\varphi_i\big(\bar{x}_N(b)\big) \ . \tag{6.93}$$

We postpone the proof of this major theorem till the next subsection and now present two *counterexamples* showing the *properness* and *consistency* conditions are *essential* for the validity of the AMP under the other assumptions held. Our first example concerns the properness condition from Definition 6.58.

Example 6.60 (AMP may not hold in smooth control problems with no properness condition). *There is a two-dimensional linear control problem with an inequality constraint such that optimal controls in the sequence of its discrete approximations are not proper and don't satisfy the approximate maximum principle.*

Proof. Consider a linear continuous-time optimal control problem (P) with a two-dimensional state $x = (x_1, x_2) \in I\!R^2$ in the following form:

$$\begin{cases} \text{minimize } \varphi\big(x(1)\big) := -x_1(1) \ \text{ subject to} \\ \dot{x}_1 = u, \quad \dot{x}_2 = x_1 - ct, \quad x_1(0) = x_2(0) = 0 \ , \\ u(t) \in U := \{0, 1\}, \quad 0 \leq t \leq 1 \ , \\ x_2(1) \leq -\dfrac{c-1}{2} \ , \end{cases}$$

where $c > 1$ is a given constant. Observe that the only "unpleasant" feature of this problem is that the control set $U = \{0, 1\}$ is *nonconvex*, and hence the feasible velocity sets $f(x, U, t)$ are nonconvex as well. It is clear that $\bar{u}(t) \equiv 1$

is the unique optimal solution to problem (P) and that the corresponding optimal trajectory is $\bar{x}_1(t) = t$, $\bar{x}_2(t) = -\dfrac{c-1}{2}t^2$. Moreover, the inequality constraint is *active*, since $\bar{x}_2(1) = -\dfrac{c-1}{2}$.

Let us now discretize this problem with the stepsize $h_N := \frac{1}{2N}$, $N \in I\!N$. For the notation convenience we omit the index N in what follows. Thus the discrete approximation problems (P_N) corresponding to the above problem (P) are written as:

$$\begin{cases} \text{minimize } \varphi\big(x(1)\big) = -x_1(1) \ \text{ subject to} \\ x_1(t+h) = x_1(t) + hu(t), \quad x_1(0) = 0\,, \\ x_2(t+h) = x_2(t) + h\big(x_1(t) - ct\big), \quad x_2(0) = 0\,, \\ u(t) \in \big\{0,1\big\}, \quad t \in \big\{0, h, \ldots, 1-h\big\}\,, \\ x_2(1) \le -\dfrac{c-1}{2} + h^2\,, \end{cases}$$

i.e., we put $\gamma_N := h_N^2$ in the constraint perturbation for (P_N).

To proceed, we compute the trajectories in (P_N) corresponding to $u(t) \equiv 1$. It is easy to see that $x_1(t) = t$ for this $u(\cdot)$. To compute $x_2(t)$, observe that

$$\big[x(t+h) = y(t) + ht,\ x(0) = 0\big] \Longrightarrow x(t) = \frac{t^2}{2} - \frac{th}{2}\,.$$

Indeed, one has by the direct calculation that

$$x(t) = h\sum_{\tau=0}^{t-h} = \big[\text{put } \tau = kh\big] = h^2 \sum_{k=0}^{\frac{t}{h}-1} k = h^2 \frac{\frac{t}{h}\big(\frac{t}{h}-1\big)}{2} = \frac{t^2}{2} - \frac{th}{2}\,.$$

Therefore for $x_2(t)$ corresponding to $u(t) \equiv 1$ in (P_N) we have

$$x_2(t) = h\sum_{\tau=0}^{t-h}\big(\tau - c\tau\big) = -\frac{c-1}{2}t^2 + \frac{c-1}{2}ht\,.$$

By this calculation we see that, for h sufficiently small, $x_2(t_1)$ no longer satisfies the endpoint constraint, and thus $u(t) \equiv 1$ is not a feasible control to problem (P_N) for all h close to zero. This implies that an optimal control to (P_N) for small h, which obviously exists, must have at least one *switching point* s such that $u(s) = 0$, and hence the maximum value of the corresponding endpoint $x_1(1)$ will be less than or equal to $1 - h$. Put

$$u(t) := \begin{cases} 1 & \text{if } t \ne s\,, \\ 0 & \text{if } t = s \end{cases}$$

and justify the formula

$$x_2(t) = \begin{cases} -\dfrac{c-1}{2}t^2 + \dfrac{c-1}{2}ht, & t \le s\ , \\ \\ -\dfrac{c-1}{2}t^2 + \dfrac{c-1}{2}ht - h(t-s) + h^2, & t \ge s+h\ , \end{cases}$$

for the corresponding trajectories in (P_N) depending on h and s. We only need to justify the second part of this formula. To compute $x_2(t)$ for $t \ge s+h$, substitute $x_1(t) = t-h$ into the discrete system in (P_N). It is easy to see that the increment $\Delta x_2(t)$ compared to the case when $u(t) \equiv 1$ is

$$h \sum_{\tau=s+h}^{t-h} (-h) = -h(t-h-s) = -h(t-s) + h^2\ ,$$

which justifies the above formula for $x_2(t)$.

Now we specify the parameters of the above control putting $c = 2$ and $s = 0.5$ for all N, i.e., considering the discrete-time function

$$\bar{u}(t) := \begin{cases} 1 & \text{if } \ t \ne 0.5\ , \\ \\ 0 & \text{if } \ t = 0.5\ . \end{cases}$$

Note that the point $t = 0.5$ belongs to the grid T_N for all N due to $h_N := \frac{1}{2N}$. Observe further that the sequence of these controls *doesn't satisfy the properness property* in Definition 6.58. It follows from the above formula for $x_2(t)$ that the corresponding trajectories obey the endpoint constraint in (P_N) whenever $N \in I\!N$, since $\bar{x}_2(1) = -\frac{1}{2}t^2 + h^2$. Moreover, it is clear from the given calculations that the control $\bar{u}(t)$ is optimal to problem (P_N) for any N.

Let us show that this sequence of optimal controls $\bar{u}(\cdot)$ doesn't satisfy the approximate maximum condition (6.85) at the point of switch. Indeed, the adjoint system (6.78) for the problems (P_N) under consideration is

$$p(t) = p(t+h) + h\nabla_x f(\bar{x}_1, \bar{x}_2, \bar{u}, t)^* \, p(t+h)\ ,$$

where the Jacobian matrix $\nabla_x f$ and its adjoint/transposed one are equal to

$$\nabla_x f = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad \nabla_x f^* = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\ .$$

Thus we have the adjoint trajectories

$$p_1(t) = p_1(t+h) + hp_2(t+h) \quad \text{and} \quad p_2(t) \equiv \text{const},$$

where the pair (p_1, p_2) satisfies the transversality condition (6.93) with the corresponding sign and nontriviality conditions (6.92) written as

$$p_1(1) = \lambda_0,\ p_2(1) = -\lambda_1; \quad \lambda_0 \ge 0,\ \lambda_1 \ge 0,\ \lambda_0^2 + \lambda_1^2 = 1 .$$

This implies that $p_1(t)$ is a linear nondecreasing function. The corresponding Hamilton-Pontryagin function is equal to

$$H\big(x(t), p(t+h), u(t)\big) = p_1(t+h)u(t) + \text{terms not depending on } u .$$

Examining the latter expression and taking into account that the optimal controls are equal to $\bar{u}(t) = 1$ for all t but $t = 0.5$, we conclude that the approximate maximum condition (6.85) holds only if $p_1(t)$ is either nonnegative or tends to zero everywhere except $t = 0.5$. Observe that $p_1(t) \equiv 0$ yields $\lambda_1 = \lambda_2 = 0$, which contradicts the nontriviality condition. Hence $p_1(t)$ must be positive away from $t = 0$. Therefore a sequence of controls having a point of switch not tending to zero as $h \downarrow 0$ *cannot* satisfy the approximate maximum condition at this point. This shows that the AMP *doesn't hold* for the sequence of optimal controls to the problems (P_N) built above. △

Many examples of this type can be constructed based on the above idea, which essentially means the following. Take a continuous-time problem with active inequality constraints and *nonconvex* admissible velocity sets $f(x, U, t)$. It often happens that after the discretization the "former" optimal control becomes not feasible in discrete approximations, and the "new" optimal control in the sequence of discrete-time problems has a *singular point of switch* (thus making the sequence of optimal controls not proper), where the approximate maximum condition is not satisfied.

The next example shows that the AMP may be violated for *proper* sequences of optimal controls to discrete approximation problems for continuous-time systems with *equality* endpoint constraints if such constraints are *not perturbed consistently* with the step of discretization.

Example 6.61 (AMP may not hold with no consistent perturbations of equality constraints). *There is a two-dimensional linear control problem with a linear endpoint constraint of the equality type such that a proper sequence of optimal controls to its discrete approximations doesn't satisfy the AMP without consistent constraint perturbations.*

Proof. Consider first the following optimal control problem for a two-dimensional system with an endpoint constraint of the equality type:

$$\begin{cases} \text{minimize } \ \varphi_0\big(x(1)\big) := x_2(1) \ \text{ subject to} \\[2ex] \dot{x} = u, \quad t \in T := [0,1], \ \ x(0) = 0 , \\[2ex] u(t) \in U := \Big\{(0,0),\ (0,-1),\ (1,-\sqrt{2}),\ (-\sqrt{2},-3)\Big\} , \\[2ex] \varphi_1\big(x(1)\big) := x_1(1) = 0 , \end{cases}$$

where $x = (x_1, x_2) \in I\!R^2$ and $u = (u_2, u_2) \in I\!R^2$. One can see that this linear problem is as standard and simple as possible with the only exception regarding the *nonconvexity* of the control region U.

Construct a sequence of discrete approximation problems (P_N) in the standard way of Theorem 6.59 by taking *zero perturbation* of the endpoint constraint, i.e., with $\xi_N = 0$. Thus we have:

$$\begin{cases} \text{minimize } \ \varphi_0\big(x_N(1)\big) = x_{2N}(1) \ \text{ subject to} \\ x_N(t + h_N) = x_N(t) + h_N u_N(t), \ \ x_N(0) = 0 \in I\!R^2 , \\ u(t) \in U, \ \ t \in T_N := \big\{0, h_N, \ldots, 1 - h_N\big\} , \\ \varphi_1\big(x_N(1)\big) = x_{1N}(1) = 0 \ \text{ with } \ h_N = N^{-1}, \ \ N \in I\!N . \end{cases}$$

It is easy to check that the only optimal solutions to problems (P_N) are

$$\bar{u}_N(t) = (0, -1), \quad \bar{x}_N(t) = (0, -t) \ \text{ for all } \ t \in T_N, \ \ N \in I\!N ,$$

which give the minimal value of the cost functional $\bar{J}_N = -1$. Note that the sequence $\bar{u}_N(\cdot)\}$ is obviously *proper* in the sense of Definition 6.58. The corresponding trajectories $p_N(\cdot)$ of the adjoint system (6.78) satisfying the transversality condition (6.93) are

$$p_N(t) = (-\lambda_{1N}, -\lambda_{0N}) \ \text{ for all } \ t \in T_N \cup \{1\} ,$$

where the sign and nontriviality conditions (6.92) for the multipliers $(\lambda_{0N}, \lambda_{1N})$ are written as

$$\lambda_{0N} \geq 0, \quad \lambda_{0N}^2 + \lambda_{1N}^2 = 1 \ \text{ whenever } \ N \in I\!N .$$

Furthermore, for each $N \in I\!N$ the Hamilton-Pontryagin function in the discrete-time system computed along $\bar{x}_N(\cdot)$ and the corresponding adjoint trajectory $p_N(\cdot)$ reduces to

$$H_N(u, t) = -\lambda_{1N} u_1 - \lambda_{0N} u_2, \quad t \in T_N ,$$

that gives $H_N(\bar{u}_N) = \lambda_{0N}$ for the optimal control.

Let us justify the estimate

$$\delta_N := \max\big\{H_N(u)\big|\ u \in U\big\} - H(\bar{u}_N) \geq 1 \ \text{ for all } \ N \in I\!N ,$$

which shows that the approximate maximum condition (6.85) is violated in the above sequence of problems (P_N). To proceed, consider the two possible cases for the multipliers $(\lambda_{0N}, \lambda_{1N})$:

(a) $\lambda_{0N} \geq 0, \ \ \lambda_{1N} \geq 0, \ \ \lambda_{0N}^2 + \lambda_{1N}^2 = 1$;

(b) $\lambda_{0N} \geq 0, \ \ \lambda_{1N} < 0, \ \ \lambda_{0N}^2 + \lambda_{1N}^2 = 1$.

In case (a) we have that

$$\delta_N = \lambda_{1N}\sqrt{2} + 3\lambda_{0N} - \lambda_{0N} \ge \sqrt{2}\big(\lambda_{1N} + \lambda_{0N}\big) \ge \sqrt{2}\ ,$$

while case (b) allows the estimate

$$\delta_N \ge |\lambda_{1N}| + 2\lambda_{0N} - \lambda_{0N} \ge 1\ .$$

Thus the AMP doesn't hold in the sequence of discrete approximation problems under consideration. △

We can observe from the above discussion that the failure of the AMP in Example 6.61 is due to the fact that the equality constraint is not perturbed (or *not sufficiently perturbed*) in the process of discrete approximation, while the optimal value of the cost functional is *not stable* with respect to such perturbations. Indeed, any control $u_N(t)$ equal to either $(1,-2)$ or $(-\sqrt{2},-3)$ at some $t \in T_N$ and giving the value $J_N[u_N] < -1$ to the cost functional is *not feasible* for the constraint $x_{1N}(1) = 0$, being however feasible for appropriate perturbations of this constraint. On the other hand, these very points of U provide the *maximum* to the Hamilton-Pontryagin function. Such a situation occurs in the discrete-time systems of Example 6.61 due to the *incommensurability* of irrational numbers in the control set U and just the rational mesh T_N for all $N \in N$. Of course, this is *not possible* in *continuous-time* systems by the completeness of real numbers.

6.4.5 Approximate Maximum Principle under Endpoint Constraints: Proofs and Applications

After all the discussions above, let us start proving Theorem 6.59. We split the proof into *three major steps* including two lemmas of independent interest, which contribute to our understanding of an appropriate counterpart of the *hidden convexity for discrete approximations*. Then we derive an *upper subdifferential* extension of the AMP to constrained problems with inequality constraint described by uniformly upper subdifferential functions. Finally, we present some typical applications of the AMP to discrete-time (with small stepsize) and continuous-time systems.

Let $u_N(t) \in U$ for all $t \in T_N$ as $N \in I\!N$. Given an integer number r with $1 \le r \le N-1$, we define needle-type variations of the control $u_N(\cdot)$ as follows. Consider a set of parameters $\{\theta_j(N), v_j(N)\}_{j=1}^r$, where $v_j(N) \in U$ and where $\theta_j(N)$ are integers satisfying

$$0 \le \theta_j(N) \le N-1 \ \text{ with } \ \theta_j(N) \ne \theta_i(N) \ \text{ if } \ j \ne i\ .$$

Denoting $\tau_{\theta_j(N)} := a + \theta_j(N)h_N$, we call

$$\widetilde{u}_N(t) := \begin{cases} v_j(N), & t = \tau_{\theta_j(N)}\ , \\ u_N(t), & t \in T_N,\ \ t \ne \tau_{\theta_j(N)},\ \ j = 1,\ldots,r\ , \end{cases} \tag{6.94}$$

the *r-needle variation* of the control $u_N(\cdot)$ with the parameters $\{\theta_j(N), v_j(N)\}$. When $r = 1$, control (6.94) is a (single) *needle* variation of $u_N(\cdot)$, while it is a *multineedle* variation of $u_N(\cdot)$ for $r > 1$. The variations introduced are discrete-time counterparts of the corresponding needle-type variations (6.71) and (6.72) of continuous-time controls, being however essentially different from the latter especially in the multineedle case.

Let $\widetilde{x}_N(\cdot)$ be the trajectory of the finite-difference system

$$x_N(t + h_N) = x_N(t) + h_N f\big(x_N(t), u_N(t), t\big), \quad x_N(a) = x_0\ , \tag{6.95}$$

corresponding to the control variation $\widetilde{u}_N(\cdot)$ with the parameters $\{\theta_j(N), v_j(N)\}$; in what follows we usually skip indicating their dependence on N. Then the difference $\widetilde{x}_N(\cdot) - x_N(\cdot)$ is denoted by $\varDelta^r_{\{\theta_j, v_j\}} x_N(\cdot)$ for $r > 1$ and by $\varDelta_{\theta,v} x_N(\cdot)$ for $r = 1$; it is called for convenience the *multineedle* (or *r-needle*) and the (single) *needle trajectory increment*, respectively. We speak about the corresponding *endpoint increments* when $t = b$.

Our first intention is to establish relationships between *integer combinations* of endpoint trajectory increments generated by *single needle* variations of the reference controls $u_N(\cdot)$ as $N \to \infty$ and some *multineedle* endpoint trajectory increments. The result derived below can be essentially viewed as an *approximate finite-difference* analog of the *hidden convexity* property crucial for continuous-time systems.

Let $\{u_N(t)\}$, $t \in T_N$, be the reference control sequence, and let $(\theta_j(N), v_j(N))$ be parameters of *single* needle variations of $u_N(\cdot)$ for each $j = 1, \ldots, p$, where p is a natural number independent of N. Given nonnegative integers m_j as $j = 1, \ldots, p$ also independent of N, consider the corresponding needle trajectory increments $\varDelta_{\theta_j, v_j} x_N(b)$ and denote them by $\varDelta_{\theta,v,j} x(b)$ for simplicity. Form the *integer combination*

$$\varDelta_N(p, m_j) := \sum_{j=1}^{p} m_j\, \varDelta_{\theta,v,j} x_N(b)$$

of the (single) needle trajectory increments for each $N = p, p+1, \ldots$ and show that it can be represented, up to a *small quantity* of order $o(h_N)$, as a *multineedle* variation of the reference control.

Lemma 6.62 (integer combinations of needle trajectory increments). *Let $\{u_N(\cdot)\}$, $N \in I\!N$, be a proper sequence of reference controls, let $p \in I\!N$ and $m_j \in I\!N \cup \{0\}$ for $j = 1, \ldots, p$ be independent of N, and let $\big(\theta_j(N), v_j(N)\big)$, $j = 1, \ldots, p$, be parameters of* (*single*) *needle variations. Then there are $r \in I\!N$ independent of N and parameters $\{\widetilde{\theta}_j(N), \widetilde{v}_j(N)\}_{j=1}^r$, of r-needle variations of type* (6.94) *such that*

$$\varDelta_N(p, m_j) = \varDelta^r_{\{\widetilde{\theta}_j, \widetilde{v}_j\}} x_N(b) + o(h_N) \quad \text{as} \quad N \to \infty\ .$$

for the corresponding endpoint trajectory increments.

Proof. First we obtain convenient representation of endpoint trajectory increments generated by needle and multineedle variations of the reference controls, which are *not* required to form a proper sequence in this setting. Recall the above notation for matrix products and denote by $K>0$ a common uniform norm bound of f and $\nabla_x f$ along $\{u_N(\cdot),x_N(\cdot)\}$, which exists due to the standing assumptions formulated in Subsect. 6.4.3. Note that, for applications to the main theorems, below but not in this lemma, we actually need the uniform boundedness along the reference sequence of *optimal* solutions to (P_N).

We start with *single* needle variations generated by parameters $\big(\theta(N),v(N)\big)$. It immediately follows from (6.95) and the smoothness of f in x that

$$\varDelta_{\theta,v}x_N(\tau_i)=0,\quad i=0,\ldots,\theta\ ,$$

$$\varDelta_{\theta,v}x_N(\tau_{\theta+1})=h_N\big[f\big(x_N(\tau_\theta),v,\tau_\theta\big)-f\big(x_N(\tau_\theta),u_N(\tau_\theta),\tau_\theta\big)\big]=:h_N y\ ,$$

$$\begin{aligned}\varDelta_{\theta,v}x_N(\tau_{\theta+2})=h_N\big[I+h_N\nabla_x f\big(x_N(\tau_{\theta+1}),u_N(\tau_{\theta+1}),\tau_{th+1}\big)\big]y\\ +h_N\,o\big(\|\varDelta_{\theta,v}x_N(\tau_{\theta+1})\|\big)\ .\end{aligned}$$

Then we easily have by induction that

$$\begin{aligned}&\varDelta_{\theta,v}x_N(b)=h_N\Big\{\prod_{\theta+1}^{i=N-1}\Big[I+h_N\nabla_x f\big(x_N(\tau_i),u_N(\tau_i),\tau_i\big)\Big]\Big\}y\\ &+h_N\sum_{k=\theta+2}^{N-1}\Big\{\prod_{k}^{i=N-1}\Big[I+h_N\nabla_x f\big(x_N(\tau_i),u_N(\tau_i),\tau_i\big)\Big]\Big\}o\big(\|\varDelta_{\theta,v}x_N(\tau_{k-1}\|\big)\\ &+h_N\,o\big(\|\varDelta_{\theta,v}x_N(\tau_{N-1})\|\big)\ .\end{aligned}$$

Observe from (6.95) and the assumptions made that $\varDelta_{\theta,v}x_N(t)=O(h_N)$ for all $t\in T_N$ uniformly in N. Thus given any $\varepsilon>0$, there is $N_\varepsilon\in I\!N$ such that

$$\big\|o\big(\|\varDelta_{\theta,v}x_N(\tau_k)\|\big)\big\|\le\varepsilon h_N,\quad k=\theta+2,\ldots,N-1,\ \ N\ge N_\varepsilon\ ,$$

which implies the estimate

$$\begin{aligned}&\Big\|\sum_{k=\theta+2}^{N-1}\Big\{\prod_{k}^{i=N-1}\Big(I+h_N\nabla_x f\big(x_N(\tau_i),u_N(\tau_i),\tau_i\big)\Big)\Big\}o\big(\|\varDelta_{\theta,v}x_N(\tau_{k-1})\|\big)\Big\|\\ &\le\varepsilon h_N\sum_{k=\theta+2}^{N-1}\prod_{k}^{i=N-1}(1+h_N K)\le\frac{\varepsilon}{K}\exp\big(K(b-a)\big)\ .\end{aligned}$$

Combining this with the above formula for $\varDelta_{\theta,v}x_N(b)$, we arrive at the efficient representation

$$\varDelta_{\theta,v}x_N(b) = h_N\Big\{\prod_{\theta+1}^{i=N-1}\Big[I + h_N\nabla_x f\big(x_N(\tau_i), u_N(\tau_i), \tau_i\big)\Big]\Big\}y \tag{6.96}$$
$$+o(h_N) \quad \text{as} \quad N \to \infty$$

for the endpoint trajectory increments generated by *single needle* variations of the reference controls, where $o(h_N)/h_N \to 0$ independently of the needle parameters $\theta = \theta(N)$ and $v = v(N)$ as $N \to \infty$.

Consider now endpoint trajectory increments generated by *multineedle* variations (6.74) with parameters $\{\theta_j(N), v_j(N)\}_{j=1}^r$. Similarly to (6.96) we derive the following representation:

$$\varDelta^r_{\{\theta_j,v_j\}}x_N(b) = h_N\Big\{\sum_{j=1}^{r}\Big[\prod_{\theta_j+1}^{i=N-1}\Big(I + \nabla_x f\big(x_N(\tau_i), u_N(\tau_i), \tau_i\big)\Big)\Big]y_j\Big\} \tag{6.97}$$
$$+o(h_N) \quad \text{as} \quad N \to \infty\ ,$$

where $o(h_N)$ is independent of $\{\theta_j(N), v_j(N)\}$ but depends on the number r of varying points, and where

$$y_j := f\big(x_N(\tau_{\theta_j}), v_j, \tau_{\theta_j}\big) - f\big(x_N(\tau_{\theta_j}), u_N(\tau_{\theta_j}), \tau_{\theta_j}\big) \quad \text{for} \quad j = 1, \ldots, r\ .$$

Next we assume that the control sequence $\{u_N(\cdot)\}$ is *proper* and justify the main relationship formulated in this lemma. Without loss of generality, suppose that the mesh points

$$\tau_{\theta_j(N)} := a + \theta_j(N)h_N, \quad j = 1, \ldots, p\ ,$$

converge to some numbers $\bar{\tau}_j \in [a, b]$, $j = 1, \ldots, p$, as $N \to \infty$. First we examine the case of

$$\bar{\tau}_i \neq \bar{\tau}_j \quad \text{for} \quad i \neq j, \quad \text{and} \quad \bar{\tau}_j \neq b \quad \text{whenever} \quad i, j \in \{1, \ldots, p\}\ . \tag{6.98}$$

Given the parameters of the integer combination $\varDelta_N(p, m_j)$, for each $N \geq p$, we take the number $r := m_1 + \ldots + m_p$ independent of N and consider the endpoint trajectory increment $\varDelta^r_{\{\widetilde{\theta}_{jq},\widetilde{v}_{jq}\}}x_N(b)$ generated by the multineedle control variation

$$\widetilde{u}_N(t) := \begin{cases} v_j(N) & \text{if } t = \tau_{\theta_j+q}(N)\ , \\ u_N(t) & \text{if } t \neq \tau_{\theta_j+q}(N), \quad t \in T_N\ , \end{cases} \tag{6.99}$$

whenever $j = 1, \ldots, p$ and $q = 0, \ldots, m_j - 1$ with

$$\widetilde{\theta}_{jq}(N) := \theta_j(N) + q \quad \text{and} \quad \widetilde{v}_{jq}(N) := v_j(N) \quad \text{for all} \quad j, q \; .$$

By assumptions (6.98) these multineedle control variations are well defined for all large N. Employing representation (6.97) of the corresponding endpoint increments, we have

$$\Delta^r_{\{\widetilde{\theta}_{iq},\widetilde{v}_{jq}\}} x_N(b) = h_N \Big\{ \sum_{j=1}^{p} \sum_{q=1}^{m_j} \Big[\prod_{\theta_j+q}^{i=N-1} \Big(I + h_N \nabla_x f\big(x_N(\tau_i), u_N(\tau_i), \tau_i\big)\Big)\Big] y_{jq-1} \Big\}$$
$$+ o(h_N) \quad \text{as} \quad N \to \infty$$

with a uniform estimate of $o(h_N)$ and with

$$y_{jq} := f\big(x_N(\tau_{\theta_j+q}), v_j, \tau_{\theta_j+q}\big) - f\big(x_N(\tau_{\theta_j+q}), u_N(\tau_{\theta_j+q}), \tau_{\theta_j+q}\big) \; .$$

By the *properness* of $\{u_N(\cdot)\}$ and the continuity of f with respect to all its variables we get $y_{ij} - y_{j0} \to 0$ as $N \to \infty$, which implies the representation

$$\Delta^r_{\{\widetilde{\theta}_{jq},\widetilde{v}_{jq}\}} x_N(b) = \Big\{ \sum_{j=1}^{p} m_j \prod_{\theta_j+1}^{i=N-1} \Big[I + \nabla_x f\big(x_N(\tau_i), u_N(\tau_i), \tau_i\big)\Big] y_j \Big\}$$
$$+ o(h_N) \quad \text{as} \quad N \to \infty \; ,$$

where y_j are defined in (6.97). Comparing the latter representation with formula (6.96) for the endpoint trajectory increment generated by *single* needle variations with the parameters $\big(\theta_j(N), v_j(N)\big)$ as $j = 1, \ldots, p$ and taking into account the expression for $\Delta_N(p, m_j)$, we arrive at the conclusion of the lemma under the above requirements (6.98) on the limiting point $\bar{\tau}_j$.

Suppose now that these requirements are not fulfilled. It is sufficient to examine the following two extreme cases:

(a) $\bar{\tau}_1 = \bar{\tau}_2 = \ldots = \bar{\tau}_p \neq b$,

(b) $\bar{\tau}_1 = \bar{\tau}_2 = \ldots = \bar{\tau}_p = b$,

which being combined with (6.98) cover all the possible locations of the limiting points $\bar{\tau}_j$ in $[a, b]$. Let us present the corresponding modifications of the multineedle variations (6.99) in both cases (a) and (b), which lead to the conclusion of the lemma similarly to the arguments above.

To proceed in case (a), reorder $\big(\theta_j(N), v_j(N)\big)$ as $j = 1, \ldots, p$ in such a way that $\theta_1 < \ldots < \theta_p$ (assuming that all θ_j are different without loss of generality) and identify for convenience the indexes θ_j with the corresponding mesh points τ_{θ_j}. Then construct the variations of $u_N(\cdot)$ at the points θ_1, $\theta_1 + 1, \ldots, \theta_1 + m_1 - 1$ as in (6.99). Assuming that the control variations corresponding to the parameters (θ_i, v_i) as $1 \leq i \leq p - 1$ have been already built, construct them for (θ_{i+1}, v_{i+1}). Denote by θ_0 the greatest point among those of $\{\theta_j\}$ at which we have built the control variations. If $\theta_0 < \theta_{i+1}$, construct variations of $u_N(\cdot)$ at $\theta_{i+1}, \theta_{i+1}+1, \ldots, \theta_{i+1}+m_{i+1}$ as in (6.99). If $\theta_0 \ge \theta_{i+1}$, construct variations of the same type at $\theta_0+1, \ldots, \theta_0+m_{i+1}$. One can check the multineedle variations built in this way ensure the fulfillment of the lemma conclusion in case (a).

In case (b) we proceed by reordering $\big(\theta_j(N), v_j(N)\big)$ as $j = 1, \ldots, p$ so that $\theta_1 > \theta_2 > \ldots > \theta_p$ and then construct the corresponding multineedle variations of $u_N(\cdot)$ symmetrically to case (a), i.e., from the right to the left. In this way we complete the proof of the lemma. △

The next result gives a *sequential* finite-difference analog of Lemma 6.44 and may be treated as a certain *approximate* (not exact/limiting) manifestation of the hidden convexity in discrete approximation problems, with no using the abstraction of time continuity. To proceed, we need to distinguish between *essential* and *inessential* inequality constraints in the process of discrete approximation important in what follows.

Definition 6.63 (essential and inessential inequality constraints for finite-difference systems). *The inequality endpoint constraint*

$$\varphi_i\big(x_N(b)\big) \le \gamma_{iN} \quad \textit{with some} \quad i \in \{1, \ldots, m\}$$

is ESSENTIAL *for a sequence of feasible solutions* $\{u_N(\cdot), x_N(\cdot)\}$ *to problems* (P_N) *along a subsequence of natural numbers* $\mathcal{M} \subset I\!N$ *if*

$$\varphi_i\big(x_N(b)\big) - \gamma_{iN} = O(h_N) \quad \textit{as} \quad h_N \to \infty \ ,$$

i.e., there is a real number $K_i \ge 0$ *such that*

$$-K_i h_N \le \varphi_i\big(x_N(b)\big) - \gamma_{iN} \le 0 \quad \textit{as} \quad N \to \infty, \quad N \in \mathcal{M} \ .$$

This constraint is INESSENTIAL *for the sequence* $\{u_N(\cdot), x_N(\cdot)\}$ *along* $\mathcal{M}$ *if whenever* $K > 0$ *there is* $N_0 \in I\!N$ *such that*

$$\varphi_i\big(x_N(b)\big) - \gamma_{iN} \le -K h_N \quad \textit{for all} \quad N \ge N_0, \quad N \in \mathcal{M} \ .$$

The notion of essential constraints in sequences of discrete approximations corresponds to the notion of *active* constraints in nonparametric optimization problems. Without loss of generality, suppose that for the sequence of optimal solutions $\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\}$ to the parametric problems (P_N) under consideration the *first* $l \in \{1, \ldots, m\}$ inequality constraints are *essential* while the other $m-l$ constraints are inessential along *all* natural numbers, i.e., with $\mathcal{M} = I\!N$.

Given optimal solutions $\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\}$ to problems (P_N) as $N \in I\!N$, we form the *linearized image set*

$$S_N := \big\{(y_0, \ldots, y_l) \in I\!R^{l+1} \big|\ y_i = \big\langle \nabla\varphi_i\big(\bar{x}_N(b)\big),\ \Delta_{\theta,v}\bar{x}_N(b)\big\rangle\big\} \tag{6.100}$$

generated by inner products involving the gradients of the cost and *essential* inequality constraint functions and the endpoint trajectory increments corresponding to *all* the *single needle* variations of the optimal controls. Our goal is to show that the sequence $\{\text{co}\, S_N\}$ of the convex hulls of sets (6.100) can be *shifted* by some quantities of *order* $o(h_N)$ as $h_N \to 0$ so that the resulting sets don't intersect the convex set of *forbidden points* in R^{l+1} given by

$$I\!R^{l+1}_< := \big\{(y_0,\ \ldots, y_l) \in I\!R^{l+1} \big|\ y_i < 0 \ \text{ for all } \ i = 0, \ldots, l\big\} .$$

Lemma 6.64 (hidden convexity and primal optimality conditions in discrete approximation problems with inequality constraints). *Let* $\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\}$ *be a sequence of optimal solutions to problems* (P_N) *with* $\varphi_i = 0$ *as* $i = m+1, \ldots, m+r$ *(no perturbed equality constraints). In addition to the standing assumptions, suppose that the endpoint functions* φ_i *are continuously differentiable around the limiting point(s) of* $\{\bar{x}_N(\cdot)\}$ *for all* $i = 0, \ldots, m$. *Assume also that the control sequence* $\{\bar{u}_N(\cdot)\}$ *is proper and that the first* $l \in \{1, \ldots, m\}$ *inequality constraints are essential for* $\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\}$ *while the other are inessential for these solutions. Then there is a sequence of* $(l+1)$*-dimensional quantities of order* $o(h_N)$ *as* $h_N \to 0$ *such that*

$$\big(\text{co}\, S_N + o(h_N)\big) \cap I\!R^{l+1}_< = \emptyset \ \ \textit{for all large} \ \ N \in I\!N . \tag{6.101}$$

Proof. For each N and fixed $r \in I\!N$ independent of N, consider an endpoint trajectory increment $\Delta^r_{\{\theta_j, v_j\}} \bar{x}_N(b)$ generated by a *multineedle* variation of the optimal control $\bar{u}_N(\cdot)$, where $\{\theta_j(N), v_j(N)\}^r_{j+1}$ are the variation parameters in (6.94). Form a sequence of the vectors

$$y_N = (y_{N0}, \ldots, y_{Nl}) \in I\!R^{l+1} \ \text{ with } \ y_{Ni} := \big\langle \nabla\varphi_i\big(\bar{x}_N(b)\big),\, \Delta^r_{\{\theta_j, v_j\}} \bar{x}_N(b)\big\rangle$$

and show that there is a sequence of $(l+1)$-dimensional quantities of order $o(h_N)$ as $h_N \to 0$ such that

$$y_N + o(h_N) \notin I\!R^{l+1}_< \ \text{ as } N \to \infty . \tag{6.102}$$

Indeed, it follows from representation (6.97) and the assumptions made that

$$\big\|\Delta^r_{\{\theta_j, v_j\}} \bar{x}_N(b)\big\| \le \mu h_N \ \text{ for all } \ t \in T_N \ \text{ and } \ N \in I\!N ,$$

where $\mu > 0$ depends on r but not on $\{\theta_j(N), v_j(N)\}^r_{j=1}$. By optimality of $\bar{x}_N(\cdot)$ in problems (P_N) with no perturbed equality constraints, for each $N \in I\!N$ there is an index $i_0(N) \in \{0, \ldots, m\}$ such that

$$\varphi_{i_0}\big(\bar{x}_N(b) + \Delta^r_{\{\theta_j, v_j\}} \bar{x}_N(b)\big) - \varphi_{i_0}\big(\bar{x}_N(b)\big) \ge 0 .$$

Since only the first l inequality constraints are essential for $\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\}$, the latter inequality holds for some $i_0 \in \{0, \ldots, l\}$ whenever N is sufficiently large. Consider the numbers

$$\delta_N := \max_{0 \le i \le l} \sup\Big\{ \ \big|\varphi_i\big(\bar{x}_N(b) + \Delta x\big) - \varphi_i\big(\bar{x}_N(b)\big)$$

$$-\big\langle \nabla\varphi_i\big(\bar{x}_N(b)\big),\, \Delta x\big\rangle\big| \ \Big| \ \|\Delta x\| \le \mu h_N \Big\}$$

for which $\delta_N/h_N \to 0$ as $N \to \infty$ uniformly with respect to variations due to the smoothness of φ_i assumed. This implies that

$$y_{Ni_0} + \delta_N \geq 0 \text{ as } N \to \infty ,$$

which justifies (6.102) with the quantities $o(h_N) := (0, \ldots, \delta_N, \ldots, 0) \in I\!R^{l+1}$, where δ_N appears at the $i_0(N)$-th position.

Our next goal is to obtain an analog of estimate (6.102) for *convex combinations* of endpoint trajectory increments generated by *single needle* variations of the optimal controls. In the case of such *integer combinations*, the corresponding analog of (6.102) follows directly from this estimate due to the preceding Lemma 6.62. Let us show that the case of convex combinations can be actually reduced to the integer one.

Consider a sequence of parameters $\big(\theta_j(N), v_j(N)\big)$, $j = 1, \ldots, p$, generating single needle variations of the optimal controls $\{\bar{u}_N(\cdot)\}$ with some $p \in I\!N$ and then define the *convex combinations*

$$y_{Ni}(p, \alpha) := \sum_{j=1}^{p} \alpha_j(N) \Big\langle \nabla\varphi_j\big(\bar{x}_N(b)\big), \Delta_{\theta,v,j}\bar{x}_N(b) \Big\rangle , \tag{6.103}$$
$$\text{as } \alpha_j(N) \geq 0, \quad \alpha_1(N) + \ldots + \alpha_p(N) = 1, \quad i = 1, \ldots, l .$$

Fixing (p, α) in the above combinations and taking $y_N(p, \alpha) \in I\!R^{l+1}$ with the components $y_{Ni}(p, \alpha)$, suppose that there is a number $N_0 \in I\!N$ such that

$$y_N(p, \alpha) \in I\!R^{l+1}_{<} \text{ whenever } N \geq N_0 .$$

Let us now show that for each natural number $N \geq N_0$ there is an index $i_0 = i_0(N) \in \{0, \ldots, l\}$ for which

$$0 > y_{Ni_0}(p, \alpha) = o(h_N) \text{ as } h_N \to \infty . \tag{6.104}$$

Assuming the contrary, we find a subsequence $\mathcal{M} \subset I\!N$ such that

$$\lim_{N\to\infty} \frac{y_{Ni}(p, \alpha)}{h_N} := \beta_i < 0 \text{ as } N \in \mathcal{M} \text{ for all } i = 0, \ldots, l .$$

Suppose without loss of generality that $\mathcal{M} = \{p, \ldots, p+1, \ldots\}$, that $\beta_i > -\infty$, and that the sequence $\{\alpha_j(N)\}$ converges to some $\alpha_j^0 \in I\!R$ as $N \to \infty$ for each $j = 1, \ldots, p$. Given $\nu > 0$, define p integers k_j by

$$k_j = k_j(\nu) := \left[\frac{\alpha_j^0}{\nu}\right] \text{ for all } j = 1, \ldots, p$$

and form the *integer combinations* $y_{Ni}(p, k)$ by

$$y_{Ni}(p,k) := \frac{y_{Ni}(p,\alpha^0)}{\nu} + \sum_{j=1}^{p}\Big(k_j - \frac{\alpha_j^0}{\nu}\Big)\Big\langle \nabla\varphi_i\big(\bar{x}_N(b)\big),\, \varDelta_{\theta,v,j}\bar{x}_N(b)\Big\rangle$$

as $i = 0,\ldots,l$, where $k := (k_1,\ldots,k_p)$ and $\alpha^0 := (\alpha_1^0,\ldots\alpha_l^0)$.

Let $\mu > 0$ be the constant selected (with $r = 1$) in the proof of (6.102), and let $\kappa > 0$ be a uniform norm bound for all $\varphi_i\big(\bar{x}_N(b)\big)$ and $\nabla\varphi_i\big(\bar{x}_N(b)\big)$ as $i = 0,\ldots,l$. Choose $i_1 \in \{0,\ldots,l\}$ and define $\nu > 0$ so that

$$|\beta_{i_1}| = \min_{0\le i\le k} |\beta_i| \quad \text{and } \nu := \frac{\beta_{i_1}}{\beta_{i_1} - p\kappa\mu}\ .$$

Then we have the estimates

$$\lim_{N\to\infty} \frac{y_{Ni}(p,k)}{h_N} \le \beta_i - \frac{\beta_i\mu\kappa p}{\beta_{i_1}} + \mu\kappa p \le \beta_i < 0 \ \text{ whenever } \ i = 0,\ldots,l\ ,$$

which clearly contradicts (6.102) by Lemma 6.62 on the representation of integer combinations of endpoint trajectory increments generated by (single) control variations. This proves (6.104).

Finally, we justify the required relationships (6.101). There is nothing to prove when $\text{co}\, S_N \cap I\!R^{l+1}_< = \emptyset$ for all large $N \in I\!N$. Suppose that $\text{co}\, S_N \cap I\!R^{l+1}_< \ne \emptyset$ along a subsequence $\{N\}$, which we put equal to the whole set $I\!N$ of natural numbers without loss of generality. For each $N \in I\!N$ define

$$\sigma_N := -\inf\Big\{ \max_{0\le i\le l} y_i \Big|\ y = (y_0,\ldots,y_l) \in \text{co}\, S_N \cap I\!R^{l+1}_- \Big\}\ ,$$

where the infimum is achieved at some $y_N \in I\!R^{l+1}_<$ under the assumptions made. Invoking the classical Carathéodory theorem, represent y_N in the convex combination form (6.103) with $p = l+2$. Employing now (6.104), we find an index $i_0 = i_0(N)$ such that

$$\sigma_N = -\max\big\{y_{Ni}\big|\ i = 0,\ \ldots, l\big\} \le y_{Ni_0} = o(h_N) \ \text{ as } \ N \to \infty\ ,$$

which implies (6.101) with the $(l+1)$-dimensional shift $o(h_N) := (\sigma_N,\ldots,\sigma_N)$ and thus ends the proof of the lemma. △

Completing the proof of Theorem 6.59. Now we have all the major ingredients to complete the proof of the theorem. Let us start with the case when only the perturbed *inequality constraints* are present in problems (P_N), i.e., $\varphi_i = 0$ for $i = m+1,\ldots,m+r$. Since we suppose without loss of generality that the first $l \le m$ inequality constraints are *essential* for the sequence of optimal solutions $\{\bar{u}_N(\cdot),\bar{x}_N(\cdot)\}$, while the remaining $m-l$ inequality constraints are inessential for this sequence, it gives by Definition 6.63 that

$$\varphi_i\big(\bar{x}_N(b)\big) - \gamma_{iN} = O(h_N) \ \text{ as } \ N\to\infty \ \text{ for } \ i = 1,\ldots,l\ .$$

Employing Lemma 6.64 and the classical *separation theorem* for the convex sets in (6.101), we find a sequence of unit vectors $(\lambda_{0N},\ldots,\lambda_{lN}) \in I\!R^{l+1}$ that separate these sets. Taking into account the structures of the sets in (6.101), one easily has that

$$\lambda_{iN} \geq 0 \text{ for all } i = 0, \ldots, l, \quad \lambda_{0N}^2 + \ldots + \lambda_{lN}^2 = 1, \text{ and}$$

$$\sum_{i=0}^{l} \lambda_{iN} \Big\langle \nabla\varphi_i\big(\bar{x}_N(b)\big), \Delta_{\theta,v}\bar{x}_N(b) \Big\rangle + o(h_N) \geq 0 \text{ as } N \to \infty$$

for any (single) needle variations of the optimal controls with parameters $\big(\theta(N), v(N)\big)$. Putting now

$$\lambda_{iN} := 0 \text{ for } i = l+1, \ldots, m \text{ as } N \to \infty$$

and proceeding similarly to the proof of Theorem 6.50 for free-endpoint problems, we get as N becomes sufficiently large that

$$h_N \Big[H\big(\bar{x}_N(t), p(t+h_N), v, t\big) - H\big(\bar{x}_N(t), p_N(t+h_N), \bar{u}_N(t), t\big) \Big] + o(h_N) \leq 0$$

for all $v \in U$ and $t \in T_N$, where each $p_N(\cdot)$ satisfies the adjoint system (6.86) with the transversality condition (6.93) and where $\lambda_{0N}, \ldots, \lambda_{mN}$ obviously obey conditions (6.91) and (6.92) for the inequality constrained problems (P_N) under consideration. The above Hamiltonian inequality directly implies, arguing by contradiction as in the proof of Theorem 6.50, the approximate maximum condition (6.85). This completes the proof of the theorem in the case of problems (P_N) with inequality constraints.

Consider now the general case of (P_N) when the *perturbed equality constraints* are present as well. Each of the constraints $\big|\varphi_{iN}\big(x_N(b)\big)\big| \leq \xi_{iN}$ can be obviously split into the two inequality constraints

$$\varphi_{iN}^+\big(x_N(b)\big) := \varphi_i\big(x_N(b)\big) - \xi_{iN} \leq 0\,,$$

$$\varphi_{iN}^-\big(x_N(b)\big) := -\varphi_i\big(x_N(b)\big) - \xi_{iN} \leq 0$$

for $i = m+1, \ldots, m+r$. Let us show that if *one* of these constraints is *essential* for $\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\}$ along some subsequence $\mathcal{M} \subset I\!N$, then the *other* is *inessential* along the same subsequence under the *consistency condition* (6.80). Indeed, suppose for definiteness that the constraint $\varphi_{iN}^+\big(\bar{x}_N(b)\big) \leq 0$ is essential for some $i \in \{m+1, \ldots, m+r\}$ along $\mathcal{M}$. Then by (6.80) we have

$$\varphi_{iN}^-\big(\bar{x}_N(b)\big) = -\varphi_i\big(\bar{x}_N(b)\big) + \xi_{iN} - 2\xi_{iN} = -\varphi_{iN}^+\big(\bar{x}_N(b)\big) - 2\xi_{iN} \leq K h_N$$

as $N \in \mathcal{M}$ for any $K > 0$, which means that the constraint $\varphi_{iN}^-\big(\bar{x}_N(t_1)\big) \leq 0$ is inessential. Applying in this way the inequality case of the theorem proved above, we find multipliers λ_{iN}^+ and λ_{iN}^- satisfying

$$\lambda_{iN}^+ \cdot \lambda_{iN}^- = 0 \text{ for } i = m+1, \ldots, m+r \text{ as } N \to \infty\,.$$

Putting finally

$$\lambda_{iN} := \lambda_{iN}^{+} - \lambda_{iN}^{-}, \quad i = m+1, \ldots, m+r \;,$$

we complete the proof of the theorem. △

Remark 6.65 (AMP for control problem with constraints at both endpoints and at intermediate points of trajectories). The approach developed above allows us to derive necessary optimality conditions in the AMP form for more general discrete approximation problems of the type (P_N) with the cost function $\varphi_0\big(x_N(a), x_N(b)\big)$ and the constraints

$$\varphi_i\big(x_N(a), x_N(b)\big) \le \gamma_{iN}, \quad i = 1, \ldots, m \;,$$

$$\big|\varphi_i\big(x_N(a), x_N(b)\big)\big| \le \xi_{iN}, \quad i = m+1, \; \ldots, m+r \;,$$

imposed at both endpoints of feasible trajectories. The AMP holds for such problems, under the same assumptions on the initial data as in Theorems 6.50 and 6.59, with the additional *approximate* transversality condition at the *left endpoints* of optimal trajectories given by

$$\lim_{N\to\infty} \Big[p_N(a) - \sum_{i=0}^{m+r} \lambda_{iN} \nabla_{x_a} \varphi_i\big(\bar{x}_N(a), \bar{x}_N(b)\big)\Big] = 0 \;,$$

where $\nabla_{x_a}\varphi_i$ stands for the partial derivatives of the functions $\varphi_i(x_a, x_b)$ at the optimal endpoints.

Similar results can be derived for analogs of problems (P_N) with the objective $\varphi_0 = \varphi(x_a, x_\tau, x_b)$ and *intermediate state constraints* of the type

$$\varphi_i\big(x_N(a), x_N(\tau), x_N(b)\big) \le \gamma_{iN}, \quad i = 1, \ldots, m \;,$$

$$\big|\varphi_i\big(x_N(a), x_N(\tau_N), x_N(b)\big)\big| \le \xi_{iN}, \quad i = m+1, \ldots, m+r \;,$$

where $\tau_N \in T_N$ is an intermediate point of the mesh. The AMP obtained for such problems involves the additional *exact* condition of the *jump* type:

$$p_N(\tau_N + h_N) - p_N(\tau_N) = \sum_{i=0}^{m+r} \lambda_{iN} \nabla_{x_\tau} \varphi_i\big(\bar{x}_N(a), \bar{x}_N(\tau_N), \bar{x}_N(b)\big)$$

$$-h_N \nabla_x H\big(\bar{x}_N(\tau_N), p_N(\tau_N + h_N), \bar{u}_N(\tau_N), \tau_N\big) \;.$$

Note that in this case the adjoint system (6.86) is required to hold for $p_N(\cdot)$ at points $t \in T_N \setminus \tau_N$.

Next we present an extension of Theorem 6.59 to *nonsmooth* problems (P_N), where the cost and inequality constraint functions φ_i, $i = 0, \ldots, m$, are assumed to be *uniformly upper subdifferentiable*. In this case the transversality conditions are obtained in the *upper subdifferential* form.

Theorem 6.66 (AMP for constrained nonsmooth problems with upper subdifferential transversality conditions). *Let $\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\}$ be optimal solutions to problems (P_N) for $N \in \mathbb{N}$ under all the assumptions of Theorem 6.59 except for the smoothness of φ_i for $i = 0, \ldots, m$. Instead we assume that these functions are uniformly upper subdifferentiable around the limiting point(s) of $\{\bar{x}_N(b)\}$. Then for any sequences of upper subgradients $x^*_{iN} \in \widehat{\partial}^+\varphi_i\big(\bar{x}_N(b)\big)$, $i = 0, \ldots, m$, there are numbers $\big\{\lambda_{iN} \big|\ i = 0, \ldots, m + r\big\}$ such that all the conditions (6.85), (6.86), (6.91), and (6.92) hold with*

$$p_N(b) = -\sum_{i=0}^{m} \lambda_{iN} x^*_{iN} - \sum_{i=m+1}^{m+r} \lambda_{iN} \nabla\varphi_i\big(\bar{x}_N(b)\big) \; .$$

Proof. Given $x^*_{iN} \in \widehat{\partial}^+\varphi_i\big(\bar{x}_N(b)\big)$ for $i = 0, \ldots, m$ and $N \in \mathbb{N}$, construct a nonsmooth counterpart of the set S_N in (6.100) by

$$S_N := \big\{(y_0, \ldots, y_l) \in \mathbb{R}^{l+1} \big|\ y_i = \big\langle x^*_{iN}, \varDelta_{\theta,v}\bar{x}_N(b)\big\rangle\big\} \; .$$

Then we get an analog of Lemma 6.64 with a similar proof. The only difference is that instead of the equalities

$$\varphi_i\big(\bar{x}_N(b) + \varDelta x\big) - \varphi_i\big(\bar{x}_N(b)\big) - \big\langle \nabla\varphi_i\big(\bar{x}_N(b)\big), \varDelta x\big\rangle + o\big(\|\varDelta x\|\big) = 0$$

used in the proof of Lemma 6.64 in the smooth case, we now arrive at the same conclusion based on the inequalities

$$\varphi_i\big(\bar{x}_N(b) + \varDelta x\big) - \varphi_i\big(\bar{x}_N(b)\big) - \big\langle x^*_{iN}, \varDelta x\big\rangle + o\big(\|\varDelta x\|\big) \le 0$$

that are due to the uniform upper subdifferentiability of φ_i for $i = 0, \ldots, l$. The separation theorem applied to the above convex sets gives

$$\sum_{i=0}^{l} \big\langle x^*_{iN}, \varDelta_{\theta,v}\bar{x}_N(b)\big\rangle + o(h_N) \ge 0 \; ,$$

which leads to the approximate maximum principle with the upper subdifferential transversality conditions similarly to the proof of Theorem 6.59. △

Remark 6.67 (suboptimality conditions for continuous-time systems via discrete approximations). The results on the fulfillment of the AMP in discrete approximation problems obtained above allow us to derive *suboptimality* conditions for *continuous-time* systems in the form of a certain *ε-maximum principle*. We have discussed in Subsect. 5.1.4 the importance of suboptimality conditions for the theory and applications of optimization problems, especially in the framework of infinite-dimensional spaces. The results and discussions of Subsect. 5.1.4 mostly concern problems of mathematical programming with functional constraints. In optimal control of continuous-time systems (even with finite-dimensional state spaces) suboptimality conditions are of great demand due to the well-known fact that *optimal solutions* *often fail to exist* in systems with *nonconvex* velocities. In such cases "almost necessary conditions" for "almost optimal" (suboptimal) solutions provide a substantial information about optimization problems that is crucial from both qualitative and quantitative/numerical viewpoints.

It follows from the above results on the *value stability* of discrete approximations (see Theorem 6.14 in Subsect. 6.1.4) that, given any $\varepsilon > 0$, optimal solutions $\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\}$ to the discrete approximation problems (P_N) considered in this subsection allow us to construct ε-*optimal solutions* $\{u_\varepsilon(\cdot), x_\varepsilon(\cdot)\}$ to the corresponding continuous-time counterpart (P) satisfying

$$\varphi_0\big(x_\varepsilon(b)\big) \le \inf J[x,u] + \varepsilon \quad \text{with}$$

$$\varphi_i\big(x_\varepsilon(b)\big) \le \varepsilon, \quad i = 1,\ldots,m, \qquad \big|\varphi_i\big(x_\varepsilon(b)\big)\big| \le \varepsilon, \quad i = m+1,\ldots,m+r\ .$$

Moreover, ε-optimal controls to the continuous-time problem (P) may always be chosen to be *piecewise constant* on $[a,b]$.

Using now the necessary optimality conditions for the discrete approximation problems (P_N) provided by Theorem 6.59 in the AMP form, we arrive at the following ε-*maximum principle* for *suboptimal solutions* to (P): there are multipliers $(\lambda_0,\ldots,\lambda_{m+r}) \in I\!R^{m+r}$ satisfying

$$\lambda_i \ge 0 \;\text{ for }\; i = 0,\ldots,m, \quad \lambda_0^2 + \ldots + \lambda_{m+r}^2 = 1\ ,$$

$$\big|\lambda_i \varphi_i\big(x_\varepsilon(b)\big)\big| \le \varepsilon \;\text{ for }\; i = 1,\ldots,m\ ,$$

and such that, whenever $u \in U$ and $t \in [a,b]$, one has

$$H\big(x_\varepsilon(t), p_\varepsilon(t), u_\varepsilon(t), t\big) \ge H\big(x_\varepsilon(t), p_\varepsilon(t), u, t\big) - \varepsilon\ ,$$

where $p_\varepsilon(\cdot)$ is the corresponding trajectory of the adjoint system

$$\dot{p} = -\nabla H\big(x_\varepsilon(t), p, u_\varepsilon(t), t\big), \quad t \in [a,b]\ ,$$

with the transversality condition

$$p_\varepsilon(b) = -\sum_{i=0}^{m+r} \nabla\varphi_i\big(x_\varepsilon(b)\big)\ .$$

Similar results hold for continuous-time problems with *intermediate state constraints* imposed at some points $\tau_j \in (a,b)$ and also for problems with endpoint constraints at both $t = a$ and $t = b$; cf. Remark 6.65. In the latter case we get an ε-*transversality condition* at $t = a$ given by

$$\Big|p_\varepsilon(a) - \sum_{i=0}^{m+r} \lambda_i \nabla_{x_a}\varphi_i\big(x_\varepsilon(a), x_\varepsilon(b)\big)\Big| \le \varepsilon\ .$$

Note, however, that the *upper subdifferential* form of the AMP in Theorem 6.66 is *not* suitable to induce a similar suboptimality result for continuous-time systems, since the Fréchet upper subdifferential $\widehat{\partial}^+\varphi(\cdot)$ doesn't generally have the required *continuity* property for nonsmooth functions.

To conclude this subsection, we illustrate the application of the AMP to optimizing constrained discrete-time systems with small stepsizes of discretization. First observe from the proof of Theorem 6.50 (and the one for Theorem 6.59) that the *difference in values* of the cost and constraint functions between optimal controls $\bar{u}_N(\cdot)$ to problems (P_N) and controls $u_N(\cdot)$ *maximizing* the Hamilton-Pontryagin function $H\big(\bar{x}_N(t), p_N(t), \cdot, t\big)$ over $u \in U$ is of *order* $o(h_N)$ as $N \to \infty$. This means in fact that the application of the *approximate* maximum principle to optimization of discrete-time systems with small stepsizes h_N leads to practically the *same effects* as in the case of its *exact* counterpart, the discrete maximum principle. Taking this into account, we now use the AMP to solve discrete approximation problems arising in optimization of some chemical processes.

Example 6.68 (application of the AMP to optimization of catalyst replacement). Consider the following optimal control problem (P) for a two-dimensional continuous-time system that appears in the catalyst replacement modeling; see, e.g., Fan and Wang [426]:

$$\begin{cases} \text{minimize} \quad J[u,x] = \varphi_0\big(x(1)\big) := x_1(1) \quad \text{subject to} \\ \dot{x}_1 = -u_1(u_1+u_2), \quad \dot{x}_2 = u_1, \quad x_1(0) = x_2(0) = 0, \quad t \in T := [0,1]\ , \\ u(t) = \big(u_1(t), u_2(t)\big) \in U := \big\{(u_1,u_2) \in I\!R^2 \big|\ 0 \le u_1, u_2 \le 2\big\}\ , \\ \varphi_1\big(x(1)\big) := x_2(1) \le 0\ . \end{cases}$$

To solve this problem numerically, construct a sequence of its discrete approximation problems (P_N):

$$\begin{cases} \text{minimize} \quad J_N[u_N,x_N] := \varphi_0\big(x_N(1)\big) = x_{1N}(1) \quad \text{subject to} \\ x_{1N}(t+h_N) = x_{1N}(t) - h_N u_{1N}(t)\big[u_{1N}(t) + u_{2N}(t)\big], \quad x_{1N}(0) = 0\ , \\ x_{2N}(t+h_N) = x_{2N}(t) + h_N u_{1N}(t), \quad x_{2N}(0) = 0, \quad h_N := N^{-1}\ , \\ 0 \le u_{1N}(t) \le 2, \quad 0 \le u_{2N}(t) \le 2, \quad t \in T_N := \{0, h_N,\ \ldots, 1-h_N\big\}\ , \\ \varphi_2\big(x_N(1)\big) = x_{2N}(1) \le 0 \quad \text{as} \quad N \to \infty\ . \end{cases}$$

Since the sets of "admissible velocities" $f(x,U,t)$ in (P_N) are *not convex*, the (exact) *discrete maximum principle cannot be applied* to find optimal controls for these problems. Let us use for this purpose the *approximate maximum principle* justified in Theorem 6.59.

For each $N \in I\!N$ the corresponding trajectory $p_N(t) = \big(p_{1N}(t), p_{2N}(t)\big)$ of the adjoint system (6.86) with the transversality condition (6.93) is

$$p_{1N}(t) = -\lambda_{0N}, \quad p_{2N}(t) = -\lambda_{1N} \quad \text{whenever} \quad t \in T_N\ ,$$

while the Hamilton-Pontryagin function along this trajectory is given by

$$H_N(u,t) = u_1\big(\lambda_{0N}u_1 + \lambda_{0N}u_2 - \lambda_{1N}\big), \quad t \in T_N \ .$$

Let us determine controls $\widehat{u}_N(t) = (\widehat{u}_{1N}(t), \widehat{u}_{2N}(t))$ that maximize the Hamilton-Pontryagin function over the control region U. One can easily see by the normalization condition in (6.92) that such controls maximize the function

$$H_\lambda(u_1,u_2) := u_1\big(\lambda u_1 + \lambda u_2 - \sqrt{1-\lambda^2}\big) \ \text{ over } \ (u_1,u_2) \in U$$

as $\lambda \in (0,1)$. It is not hard to compute, taking into account the structure of the control set U, that the maximizing controls $\widehat{u}_N(\cdot)$ are as follows depending on the values of the parameter $\lambda \in (0,1)$:

(a) if $\lambda > 1/\sqrt{17}$, then $\widehat{u}_{1N}(t) = 2$, $\widehat{u}_{2N}(t) = 2$ for all $t \in T_N$;

(b) if $\lambda < 1/\sqrt{17}$, then $\widehat{u}_{1N}(t) = 0$, $\widehat{u}_{2N}(t) \in [0,2]$ for all $t \in T_N$;

(c) if $\lambda = 1/\sqrt{17}$, then for each $t \in T_N$ one has either $\widehat{u}_{1N}(t) = \widehat{u}_{2N}(t) = 2$, or $\widehat{u}_{1N}(t) = 0$ and $\widehat{u}_{2N}(t) \in [0,2]$.

We can directly check that the controls $\widehat{u}_N(\cdot)$ in case (a) are not feasible for (P_N), since the corresponding trajectories $\widehat{x}_N(\cdot)$ don't satisfy the endpoint constraint. In case (b) the controls $\widehat{u}_N(\cdot)$ are far from optimality, since $J_N[\widehat{u}_N,\widehat{x}_N] = 0$ while $\inf J[u_N,x_N] \le -1$. In case (c) the controls $\widehat{u}_N(\cdot)$ are feasible for (P_N) provided that the number of points $t \in T_N$ at which $\widehat{u}_{1N}(t) = 2$ and $\widehat{u}_{2N}(t) = 2$ is not greater than $[N/2]$ as $N \in I\!N$. By Theorem 6.59 and the discussion right before this example we conclude that optimal controls $\bar{u}_N(\cdot)$ to (P_N) (which always exist) may be either feasible ones $\widehat{u}_N(\cdot)$ in case (c) satisfying the properness condition, or those for which the values of the cost and constraint functions are different from $\varphi_0\big(\widehat{x}_N(b)\big)$ and $\varphi_1\big(\widehat{x}_N(b)\big)$ by quantities of order $o(h_N)$ as $N \to \infty$.

Thus the AMP allows us to efficiently describe the collection of all feasible controls to (P_N) that are suspicious to optimality. Based on this information, we can finally determine from the structure of problems (P_N) that optimal solutions to the sequence of these problems are given by the controls

$$\begin{cases} \bar{u}_{1N}(t) = \bar{u}_{2N}(t) = 2 & \text{if } t \text{ is the } [N/2]\text{-th point of } T_N \ , \\ \bar{u}_{1N}(t) = 0, \quad \bar{u}_{2N}(t) \in [0,2] & \text{for all other } t \in T_N \ . \end{cases}$$

This completely solves the problems under consideration.

6.4.6 Control Systems with Delays and of Neutral Type

The last subsection of this section is devoted to the extension of the AMP in the *upper subdifferential* form to finite-difference approximations of *time-delay* controls systems with smooth dynamics. For brevity we present results only for free-endpoint problems. The main theorem of this subsection provides a generalization of Theorem 6.50 in the case of delay problems; the corresponding extension of Theorems 6.59 and 6.66 can be derived similarly. On the other hand, we show at the end of this subsection that the AMP may *not hold* for discrete approximations of smooth functional-differential systems of *neutral type* that contain time-delays not only in state variables but in velocity variables as well.

We begin with the following continuous-time problem (D) with a single time delay in the state variable:

$$\begin{cases} \text{minimize } J[u,x] := \varphi\big(x(b)\big) \text{ subject to} \\ \dot{x}(t) = f\big(x(t), x(t-\theta), u(t), t\big) \text{ a.e. } t \in [a,b]\,, \\ x(t) = c(t), \quad t \in [a-\theta, a]\,, \\ u(t) \in U \text{ a.e. } t \in [a,b] \end{cases}$$

over measurable controls $u\colon [a,b] \to U$ and the corresponding absolutely continuous trajectories $x\colon [a,b] \to X$ of the delay system, where $\theta > 0$ is a constant time-delay, and where $c\colon [t_0 - \theta, t_0] \to X$ is a given function defining the initial "tail" condition that is needed to start the delay system; see Remark 6.40, where the results on the maximum principle for such problems have been discussed. Now our goal is to derive an appropriate version of the AMP for discrete approximation of the delay problem (D).

Let us build discrete approximations of (D) based on the Euler finite-difference replacement of the derivative. In the case of time-delay systems we need to ensure that the point $t-\theta$ belongs to the discrete grid whenever t does. It can be achieved by defining the discretization step as $h_N := \dfrac{\theta}{N}$ in contrast to $h_N = \dfrac{b-a}{N}$ for the non-delay problems (P_N^0) considered in Subsect. 6.4.3. In such a scheme the length of the time interval $b-a$ is generally *no longer commensurable* with the discretization step h_N. Define the grid T_N on the main time interval $[a,b]$ by

$$T_N := \big\{a, a+h_N, \ldots, b - \widetilde{h}_N - h_N\big\} \text{ with}$$

$$h_N := \frac{\theta}{N} \text{ and } \widetilde{h}_N := b - a - h_N \left[\frac{b-a}{h_N}\right]$$

and consider the following sequence of finite-difference approximation problems (D_N) with discrete time delays:

$$\begin{cases} \text{minimize } J[u_N, x_N] := \varphi\big(x_N(b)\big) \text{ subject to} \\ x_N(t+h_N) = x_N(t) + h_N f\big(x_N(t), x_N(t-Nh_N), u_N(t), t\big),\ t \in T_N\ , \\ x_N(b) = x_N(b-\widetilde{h}_N) + \widetilde{h}_N f\big(x_N(b-\widetilde{h}_N), u_N(b-\widetilde{h}_N), b-\widetilde{h}_N\big)\ , \\ x_N(t) = c(t), \quad t \in T_{0N} := \big\{a-\theta, a-\theta+h_N, \ldots, a\big\}\ , \\ u_N(t) \in U, \quad t \in T_N\ . \end{cases}$$

To derive the AMP for the sequence of problems (D_N), we reduce these problems to those *without delays* and employ the results of Theorem 6.57, where the *standing assumptions* are similar to the ones formulated in Subsect. 6.4.3 involving now the additional state variable y in $f(x,y,u,t)$ together with x. For convenience we introduce the following notation:

$$z_N(t) := \big(x_N(t), x_N(t-\theta)\big), \quad \bar{z}_N(t) := \big(\bar{x}_N(t), \bar{x}_N(t-\theta)\big)\ ,$$

$$f(z_N, u_N, t) := f\big(x_N(t), x_N(t-\theta), u_N(t), t\big)\ ,$$

$$f(t, \bar{z}_N, u_N) := f\big(\bar{x}_N(t), \bar{x}_N(t-\theta), u_N(t), t\big)$$

in which terms the *adjoint system* to (D_N) is written as

$$p_N(t) = p_N(t+h_N) + h_N \nabla_x f(\bar{z}_N, \bar{u}_N, t)^* p_N(t+h_N)$$

$$+ h_N \nabla_y f(\bar{z}_N, \bar{u}_N, t+\theta)^* p_N(t+\theta+h_N) \quad \text{for } t \in T_N\ ,$$

$$p_N(b-\widetilde{h}_N) = p_N(t_1) + \widetilde{h}_N \nabla_x f(\bar{z}_N, \bar{u}_N, b-\widetilde{h}_N)^* p_N(b)$$

along the optimal processes $\big\{\bar{u}_N(\cdot), \bar{x}_N(\cdot)\big\}$ to the delay problems (D_N) for each $N \in I\!N$. Introducing the corresponding *Hamilton-Pontryagin function*

$$H(x_N, y_N, p_N, u, t) := \begin{cases} \big\langle p_N(t+h_N), f(x_N, y_N, u, t)\big\rangle & \text{if } t \in T_N\ , \\ \big\langle p_N(t), f(x_N, y_N, u, t-\widetilde{h}_N)\big\rangle & \text{if } t = b-\widetilde{h}_N \end{cases}$$

with $y_N(t) := x_N(t-\theta)$, we rewrite the adjoint system as

$$p_N(t) = p_N(t+h_N) + h_N\Big[\nabla_x H(\bar{z}_N, p_N, \bar{u}_N, t) + \nabla_y H(\bar{z}_N, p_N, \bar{u}_N, t+\theta)\Big]$$

when $t \in T_N$ and

$$p_N(b-\widetilde{h}_N) = p_N(b) + \widetilde{h}_N \nabla_x H(\bar{z}_N, p_N, \bar{u}_N, b-\widetilde{h}_N)$$

at the “incommensurable” point. Then we have the following result on the fulfillment of the AMP for time-delay discrete approximations.

Theorem 6.69 (AMP for delay systems). *Let the pairs* $\{\bar u_N(\cdot), \bar x_N(\cdot)\}$ *be optimal to problems* (D_N). *In addition to the standing assumptions, suppose that the cost function* φ *is uniformly upper subdifferentiable around the limiting point(s) of the sequence* $\{\bar x_N(b)\}$, $N \in I\!N$. *Then for every sequence of upper subgradients* $x_N^* \in \mathcal{D}^+\varphi(\bar x_N(b))$ *the approximate maximum condition*

$$H(\bar z_N, p_N, \bar u_N, t) = \max_{u \in U} H(\bar z_N, p_N, u, t) + \varepsilon(t, h_N), \quad t \in \widetilde T_N := T_N \cup \{b - \widetilde h_N\},$$

is fulfilled, where $\varepsilon(t, h_N) \to 0$ *as* $h_N \to 0$ *uniformly in* $t \in \widetilde T_N$, *and where* $p_N(\cdot)$ *satisfies and the transversality relations*

$$p_N(b) = -x_N^*, \qquad p_N(t) = 0 \ \text{ as } \ t > b\ . \tag{6.105}$$

Furthermore, we can take any $x^* \in \widehat\partial^+\varphi(\bar x_N(b))$ *in* (6.105) *if* X *is reflexive and* φ *is continuous around the limiting point(s) of* $\{\bar x_N(b)\}$.

Proof. We reduce the delay discrete approximation problems to those with *no delay* (but with the *incommensurability* between $b - a$ and h_N) by the following *multistep procedure.* Denote

$$y_{1N}(t) := x_N(t - h_N), \quad t \in \{a + 2h_N, \ldots, b - \widetilde h_N\}\ ,$$

$$y_{1N}(t) := c_N(t - h_N), \quad t \in \{a - \theta + h_N, \ldots, a + h_N\}\ ,$$

$$y_{2N}(t) := y_{1N}(t - h_N), \quad t \in \{a - \theta + 2h_N, \ldots, b - \widetilde h_N\}\ ,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$y_{NN}(t) := y_{N-1,N}(t - h_N), \quad t \in \{a, \ldots, b - \widetilde h_N\}\ ,$$

and observe that the values of $y_{1N}(b), \ldots, y_{NN}(b)$ can be defined arbitrarily, since they don't enter either the adjoint system or the cost function. To match the setup of Theorem 6.57, define

$$y_{1N}(b) := x_N(b - \widetilde h_N),\ y_{2N}(b) := y_{1N}(b - \widetilde h_N), \ldots, y_{NN}(b) := y_{N-1,N}(b - \widetilde h_N)\ .$$

After the change of variables we have

$$y_{NN}(t) = \begin{cases} x_N(t - \theta),\ t \in \{a + \theta + h_N, \ldots, b - \widetilde h_N\}\ , \\ c(t - \theta), \quad t \in \{a, \ldots, a + \theta\}\ . \end{cases}$$

The original system in (D_N) is thereby reduced, for each $N \in I\!N$, to the following non-delay system of dimension $I\!R^{(N+1)n}$:

$$\begin{cases} s_N(t + h_N) = s_N(t) + h_N g(s_N, u_N, t), \quad t \in T_N\ , \\ s_N(b) = s_N(b - \widetilde h_N) + \widetilde h_N g(s_N, u_N, b - \widetilde h_N) \end{cases}$$

with the state vector $s_N(t) := \big(x_N(t), y_{1N}(t), \ldots, y_{NN}(t)\big)$ and the "velocity" mapping $g(s_N, u_N, t)$ given by

$$g\big(s_N(t), u_N(t), t\big) := \begin{pmatrix} f\big(x_N(t), y_{NN}(t), u_N(t), t\big) \\ \\ \dfrac{x_N(t) - y_{1N}(t)}{h_N} \\ \\ \ldots\ldots \\ \dfrac{y_{N-1,N}(t) - y_{NN}(t)}{h_N} \end{pmatrix},$$

where h_N should be replaced by $\widetilde{h}_N$ for $t = b - \widetilde{h}_N$ in the last formula.

Let us apply Theorem 6.57 to the problem of minimizing the same functional as in (D_N) over the feasible pairs $\{u_N(\cdot), s_N(\cdot)\}$ of the obtained non-delay system. The adjoint system in this problem, with respect to the new adjoint variable $q \in I\!R^{(N+1)n}$, has the form

$$\begin{cases} q_N(t) = q_N(t + h_N) + h_N \nabla_s g(\bar{z}_N, \bar{u}_N, t)^* \, q(t + h_N), \quad t \in T_N \,, \\ \\ q_N(b - \widetilde{h}_N) = q_N(b) + \widetilde{h}_N \nabla_s g(\bar{s}_N, \bar{u}_N, b - \widetilde{h}_N)^* \, q_N(b) \end{cases}$$

with the transversality condition

$$q_N(b) = -(x_N^*, 0, \ldots, 0) \;\text{ for }\; x_N^* \in \mathcal{D}^+\varphi\big(\bar{x}_N(b)\big)\,,$$

which reduces to $x_N^* \in \widehat{\partial}^+\varphi\big(\bar{x}_N(b)\big)$ when X is reflexive and φ is continuous. Taking into account the above relationship between g and f and performing elementary calculations, we express the operator $\nabla_s g^*$ via $\nabla_x f^*$ and $\nabla_y f^*$ and arrive at the transversality relations (6.105) for the first component $p_N(\cdot)$ of the adjoint trajectory $q_N(\cdot)$. Furthermore, one gets the relationship

$$\begin{aligned} \widetilde{H}(\bar{s}_N, q_N, u, t) &= \big\langle q_N(t + h_N), g(\bar{s}_N, u, t)\big\rangle \\ &= \big\langle p_N(t + h_N), f(\bar{z}_N, u, t)\big\rangle + r(\bar{s}_N, q_N, h_N, t) \\ &= H(\bar{z}_N, p_N, u, t) + r(\bar{s}_N, q_N, h_N, t), \quad t \in T_N \,, \end{aligned}$$

and similarly for $t = b - \widetilde{h}_N$, between the Hamilton-Pontryagin functions of the non-delay and original delay systems considered above, where the remainder $r(\bar{s}_N, q_N, h_N, t)$ doesn't depend on u. Applying now the approximate maximum condition from Theorem 6.57 to the non-delay system, we complete the proof of the theorem. △

To conclude this section, we consider optimal control problems for finite-difference approximations of the so-called *functional-differential systems of neutral type* (cf. also Sect. 7.1) given by

$$\dot{x}(t) = f\big(x(t), x(t-\theta), \dot{x}(t-\theta), u(t), t\big), \quad u(t) \in U, \quad \text{a.e. } t \in [a,b] ,$$

which contain time-delays not only in state but also in *velocity* variables. A finite-difference counterpart of such systems with the stepsize h and with the grid $T := \{a, a+h, \ldots, b-h\}$ is

$$x(t+h) = x(t) + hf\Big(x(t), x(t-\theta), \frac{x(t-\theta+h) - x(t-\theta)}{h}, u(t), t\Big)$$

as $u(t) \in U$ for $t \in T$, and the adjoint system is given by

$$p(t) = p(t+h) + h\nabla_x f(\bar{v}, \bar{u}, t)^* \, p(t+h) + h\nabla_y f(\bar{v}, \bar{u}, t+\theta)^* \, p(t+\theta+h)$$

$$+h\nabla_z f(\bar{v}, \bar{u}, t+\theta-h)^* \, p(t+\theta) - h\nabla_z f(\bar{v}, \bar{u}, t+\theta)^* \, p(t+\theta+h)$$

for $t \in T$, where $\{\bar{u}(\cdot), \bar{x}(\cdot)\}$ is an optimal solution to the neutral analog of problem (D_N), and where

$$\bar{v}(t) := \Big(\bar{x}(t), \bar{x}(t-\theta), \frac{\bar{x}(t-\theta+h) - \bar{x}(t-\theta)}{h}\Big), \quad t \in T .$$

The following example shows that the AMP is *not generally fulfilled* for finite-difference neutral systems, in contrast to ordinary and delay ones, even in the case of *smooth* cost functions.

Example 6.70 (AMP may not hold for neutral systems). *There is a two-dimensional control problem of minimizing a linear function over a smooth neutral system with no endpoint constraints such that some sequence of optimal controls to discrete approximations doesn't satisfy the approximative maximum principle regardless of the stepsize and a mesh point.*

Proof. Consider the following parametric family of discrete optimal control problems for neutral systems with the parameter $h > 0$:

$$\begin{cases} \text{minimize } J[u, x_1, x_2] := x_2(2) \ \text{ subject to} \\[1ex] x_1(t+h) = x_1(t) + hu(t), \quad t \in T := \big\{0, h, \ldots, 2-h\big\} , \\[1ex] x_2(t+h) = x_2(t) + h\Big(\dfrac{x_1(t-1+h) - x_1(t-1)}{h}\Big)^2 - hu^2(t), \quad t \in T , \\[1ex] x_1(t) \equiv x_2(t) \equiv 0, \quad t \in T_0 := \{-1, \ldots, 0\} , \\[1ex] |u(t)| \le 1, \quad t \in T . \end{cases}$$

It is easy to see that

$$x_2(1) = -h\sum_{t=0}^{1-h} u^2(t) \quad \text{and}$$

$$x_2(2) = x_2(1) + h\sum_{t=1}^{2-h}\Big(\frac{x_1(t-1+h)-x_1(t-1)}{h}\Big)^2 - h\sum_{t=1}^{2-h}u^2(t)$$

$$= -h\sum_{t=0}^{1-h}u^2(t) + h\sum_{t=0}^{1-h}u^2(t) - h\sum_{t=1}^{2-h}u^2(t) = -h\sum_{t=1}^{2-h}u^2(t)\,.$$

Thus the control

$$\bar u(t) = \begin{cases} 0,\ t\in\{0,\ldots,1-h\}\,, \\ 1,\ t\in\{1,\ldots,2-h\}\,, \end{cases}$$

is an optimal control to the problems under consideration for any h. The corresponding trajectory is

$$\bar x_1(t) = \begin{cases} 0, & t\in\{0,\ldots,1-h\}\,, \\ t-1, & t\in\{1,\ldots,2-h\}; \end{cases} \qquad \bar x_2(t) = \begin{cases} 0, & t\in\{0,\ldots,1-h\}\,, \\ -t+1, & t\in\{1,\ldots,2-h\}\,. \end{cases}$$

Computing the partial derivatives of the "velocity" mapping f in the above system, we get

$$\nabla_x f = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \nabla_y f = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \text{and}$$

$$\nabla_z f(t+1) = \frac{1}{h}\begin{pmatrix} 0 & 0 \\ 2(x_1(t+h)-x_1(t)) & 0 \end{pmatrix}.$$

Hence the adjoint system reduces to

$$p_1(t) = p_1(t+h) + 2\big(\bar x_1(t) - \bar x_1(t-h)\big)\,p_2(t+1)$$

$$-2\big(\bar x_1(t+h) - \bar x_1(t)\big)\,p_2(t+1+h), \quad t\in\{0,\ldots,2-h\}\,,$$

with $p_2(t)\equiv$ const and with the transversality conditions

$$p_1(2)=0,\ \ p_2(2)=-1; \quad p_1(t)=p_2(t)=0 \ \text{ for } \ t>2\,.$$

The solution of this system is

$$p_1(t)\equiv 0,\ \ p_2(t)\equiv -1 \ \text{ for all } \ t\in\{0,\ldots,2-h\}\,.$$

Thus the Hamilton-Pontryagin function along the optimal solution is

$$H(t,\bar x_1,\bar x_2,p_1,p_2,u) = p_2(t+h)\Big\{\Big(\frac{x_1(t-1+h)-x_1(t-1)}{h}\Big)^2 - u^2\Big\}$$

$$+p_1(t+h)u = u^2, \quad t\in\{0,\ldots,1-h\}\,.$$

This shows that the optimal control $\bar{u}(t) = 0$ *doesn't* provide the approximate maximum to the Hamilton-Pontryagin function regardless of h and mesh points $t \in \{0, \ldots, 1-h\}$. Note at the same time that *another sequence* of optimal controls with $\bar{u}(t) = 1$ for all $t \in \{0, \ldots, 2-h\}$ *does* satisfy the *exact* discrete maximum principle regardless of h. △

6.5 Commentary to Chap. 6

6.5.1. Calculus of Variations and Optimal Control. Chapter 6 is devoted to problems of *dynamic optimization*. This name conventionally reflects the fact that some initial data of a given optimization problem *evolve in time*. The origin of such problems goes back to the classical *calculus of variations*, which was in the beginning of all infinite-dimensional analysis; we refer the reader to the seminal contributions by Euler [411], Lagrange [737], Hamilton [548], Jacobi [625], Mayer [859], Weierstrass [1326], Bolza [130], Tonelli [1260], Carathéodory [222], and Bliss [119] (with his famous Chicago school) among other developments the most influential for the topics considered in this book.

The theory of *optimal control* for *ordinary differential equations* (ODE), which has been well recognized as a modern counterpart of the classical calculus of variations, distinguishes from its predecessor by, first of all, the presence of *hard/pointwise* constraints on control functions generating system trajectories (often called admissible arcs) via the *evolution ODE systems*

$$\dot{x} = f(x, u, t), \quad u(t) \in U, \quad t \in [a, b], \quad x \in I\!R^n . \tag{6.106}$$

Such control constraints given by sets U of a rather *irregular nature*, which appeared already in the very first problems of optimal control arisen from practical applications, have been a permanent source of *intrinsic nonsmoothness* in optimal control theory and have eventually *motivated* the development of many crucial aspects of modern *variational analysis* and *generalized differentiation.*

As mentioned in Subsect. 1.4.1, the fundamental result of optimal control theory widely known as the *Pontryagin maximum principle* (PMP) [1102], which was formulated by Pontryagin and then was proved by Gamkrelidze [494] for linear systems and by Boltyanskii [124] for problems with nonlinear *smooth* dynamics, has played a major role in developing modern variational analysis. It is interesting to observe that the first attempt [129] in formulating the maximum principle—as a *sufficient* condition for *local* optimality—was *wrong*; see the papers by Boltyanskii [128] and Gamkrelidge [498] for (rather different) historical accounts in the discovery of the maximum principle. In these papers and also in the book by Hestenes [565] and in the survey paper by McShane [865], the reader can find various discussions on the relationships between the maximum principle and the preceding results obtained in the

Chicago school on the calculus of variations and in the theory and applications of automatic control; see also the excellent survey by Gabasov and Kirillova [487]. Probably the closest predecessors to optimal control theory were non-standard variational problems and results developed for *optimal systems* of linear *automatic control*, in particular, the so-called "theorem on *n*-intervals" by Feldbaum [440] and the "bang-bang principle" by Bellman, Glicksberg and Gross [95].

Although analogs of many elements in both formulation and proof of the PMP can be found in the calculus of variations (particularly *needle variations* employed by McShane [860], which actually go back to Weierstrass [1326] and his necessary optimality condition for strong minimizers; *tangential convex approximations* and the usage of *convex separation* as in McShane [860]; *canonical* variables and a modified *Hamiltonian* function, etc.), the discovery of the PMP and its proof came as a *surprise* ("sensation" in Pshenichnyi's wording [1106]). It is difficult to overestimate the impact and role of the PMP in the development of modern variational analysis. We refer the reader to [7, 32, 105, 124, 218, 235, 255, 370, 485, 486, 497, 504, 539, 565, 618, 801, 863, 865, 877, 1002, 1106, 1239, 1289, 1315, 1351] for more results and discussions on the relationships between optimal control, the calculus of variations, and mathematical programming.

It seems that among the *most significant new contributions* of the PMP in comparison with the classical calculus of variations was the discovery (by Pontryagin) of the *adjoint system* to (6.106) given by

$$\dot{p} = -\frac{\partial f(\bar{x}, \bar{u}, t)}{\partial x}^{*} p = -\nabla_x H(\bar{x}, p, \bar{u}, t) , \tag{6.107}$$

via the *Hamilton-Pontryagin function*

$$H(x, p, u, t) := \langle p, f(x, u, t) \rangle, \quad p \in I\!R^n , \tag{6.108}$$

computed along the optimal process $(\bar{x}, \bar{u})$, in which terms the crucial pointwise *maximum condition* was written as

$$H\big(\bar{x}(t), p(t), \bar{u}(t), t\big) = \max_{u \in U} H\big(\bar{x}(t), p(t), u, t\big) \quad \text{a.e.} \tag{6.109}$$

It has been recognized, after the discovery of the PMP, that the maximum condition (6.109) is an optimal control counterpart of the *Weierstrass's excess function condition* for *strong minimizers* in the calculus of variations.

6.5.2. Differential Inclusions. A notable disadvantage of the original optimal control model (6.106) is that it doesn't cover problems with *state-dependent* control sets $U = U(x)$ important for both the theory and applications. Problems of this class, as well as of other significant classes in control and dynamic optimization, can be naturally written in the form of *differential inclusions*

$$\dot{x} \in F(x,t), \quad x \in I\!R^n , \tag{6.110}$$

which actually go back to the classes of set-valued differential equations studies (not from the control viewpoint) in the 1930s as "contingent equations" by Marchaud [850] and "paratingent equations" by Zaremba [1355]; see also Nagumo [990] and Wȧzewski [1325] for early developments. Control systems (6.106) equivalently reduce to the differential inclusion form (6.110) by the so-called "Filippov implicit function lemma" [449], which is in fact a result on measurable selections of set-valued mappings; see, e.g., Castaing and Valadier [229] and Rockafellar and Wets [1165] for more references and discussions.

Observe that control systems governed by differential inclusions (6.110) are significantly *more complicated* in comparison with the classical ones (6.106) due to, e.g., the impossibility of employing standard needle variations to derive optimality conditions. Moreover, systems (6.110) *explicitly* reveal the *intrinsic nonsmoothness* inherent even in classical optimal control via, first of all, *hard* control constraints of the type $u(t) \in U$, particularly given by finite sets like $U = \{0, 1\}$ that are typical in automatic control applications. This phenomenon is somehow hidden in the PMP for systems (6.106) of *smooth dynamics* due to using the Hamilton-Pontryagin function (6.108) differentiable in the state-costate variables (x, p). Another manifestation of nonsmoothness in optimal control is provided by the *Hamiltonian* function

$$\mathcal{H}(x,p,t) := \sup\big\{\langle p,v\rangle\big|\ v \in F(x,t)\big\} \tag{6.111}$$

for the differential inclusion (6.110), which corresponds to the "true" Hamiltonian

$$\mathcal{H}(x,p,t) := \sup\big\{H(x,p,u,t)\big|\ u \in U\big\}$$

for the standard/parameterized control systems (6.106). These generalized Hamiltonians can be viewed as control counterparts of the classical Hamiltonian in problems of the calculus of variations and mechanics associated (via the *Legendre transform* if the latter is well-defined) with the Lagrangian, i.e., integrand under minimization.

6.5.3. Optimality Conditions for Smooth or Graph-Convex Differential Inclusions. Nonsmoothness is a *characteristic feature* of the Hamiltonian (6.111) and its above implementation for control systems (6.106); a smooth behavior occurs only under some quite restrictive assumptions. However, the first necessary optimality conditions for control problems governed by differential inclusions were obtained (under the name of "support principle") by Boltyanskii [125] assuming the *smoothness* of (6.111) in the state variable; see also the related papers by Fedorenko [438, 439], Boltyanskii [127], Blagodatskikh [117], Blagodatskikh and Filippov [118] with other (mostly Russian) references therein.

In [1143, 1144, 1145], Rockafellar derived necessary (and sufficient) optimality condition applied to differential inclusions (6.110) under more reasonable assumptions of the *graph-convexity* for $F(\cdot,t)$. In fact, Rockafellar considered a more general framework of the (fully) convex *generalized problem of Bolza*:

$$\text{minimize } \varphi\big(x(a),x(b)\big)+\int_a^b \vartheta\big(x(t),\dot{x}(t),t\big)\,dt\;, \tag{6.112}$$

where, in contrast to the classical Bolza problem [130] and the preceding Mayer problem [859] with $\vartheta=0$, the functions φ and ϑ may be *extended-real-valued*, i.e., (6.112) particularly incorporates the differential inclusion model (6.110) via the indicator function $\vartheta(x,v,t):=\delta\big((x,v);\operatorname{gph}F(\cdot,t)\big)$. The convexity assumption on $\vartheta(x,v,t)$ in both variables (x,v) made in [1143, 1144, 1145] implies that the Hamiltonian (6.111) associated with the differential inclusion (6.110) is convex in p and concave in x, so it is *subdifferentiable* as a *saddle function* with respect to (x,p) in the sense of convex analysis. Using the machinery of convex analysis in infinite-dimensional spaces, Rockafellar obtained necessary and sufficient conditions for optimal solutions $\bar{x}(\cdot)$ to the convex generalized problem of Bolza and thus for convex-graph differential inclusions via the *generalized Hamiltonian equation* [1145] called also the *Hamiltonian condition/inclusion*

$$\big(-\dot{p}(t),\dot{\bar{x}}(t)\big)\in\partial\mathcal{H}\big(\bar{x}(t),p(t),t\big)\quad\text{a.e.}\;, \tag{6.113}$$

where $\partial\mathcal{H}$ stands for the subdifferential of the Hamiltonian function $\mathcal{H}(x,p,t)$ with respect to (x,p). If $\mathcal{H}(x,p,t)$ happens to be differentiable with respect to x and p, inclusion (6.113) reduces to the classical Hamiltonian system

$$\dot{\bar{x}}(t)=\nabla_p\mathcal{H}\big(\bar{x}(t),p(t),t\big)\;\;\text{and}\;\;-\dot{p}(t)=\nabla_x\mathcal{H}\big(\bar{x}(t),p(t),t\big)\;.$$

Somewhat different (while mostly equivalent) results for optimization problems governed by *convex-graph* differential inclusions were later obtained by Halkin [542], Berliocchi and Lasry [107], and Pshenichnyi [1107, 1109].

6.5.4. Clarke's Euler-Lagrange Condition. Observe that although the graph-convexity assumption on $F(\cdot,t)$ is more reasonable in comparison with the smoothness requirement on the Hamiltonian, it is still rather restrictive. In particular, for standard control systems (6.106) this assumption actually reduces to the *linearity* of $f(\cdot,\cdot,t)$ and the convexity of U; see Rockafellar [1143]. A crucial step from *fully convex*, or "biconvex" in Halkin's terminology, problems (i.e., those for which the integrand in (6.112) in convex in *both* (x,v) variables) to problems involving the *convexity* only in the *velocity* variable v, which corresponds to the *convex-valuedness* of $F(x,t)$ in the differential inclusion framework (6.110), was made by Clarke in his pioneering work in the 1970s starting with his dissertation [243].

The initial point for Clarke [243, 245] was the Bolza-type problem (6.112) with *finite* (moreover Lipschitzian) integrand/Lagrangian $\vartheta(\cdot,\cdot,t)$ considered without any smoothness and convexity assumptions on the integrand ϑ as well as on the l.s.c. endpoint function φ, which was allowed to be extended-real-valued. The main necessary optimality condition was obtained in the *Euler-Lagrange form*

$$\big(\dot p(t),p(t)\big)\in\partial_C\vartheta\big(\bar x(t),\dot{\bar x}(t),t\big)\ \text{ a.e.} \tag{6.114}$$

via Clarke's generalized gradient of $\vartheta(\cdot,\cdot,t)$ in (6.114). Inclusion (6.114) gets back the classical Euler-Lagrange equation if $\vartheta(x,v,t)$ is smooth in (x,v); it reduces to the Euler-Lagrange inclusion obtained by Rockafellar [1143] if ϑ is convex in both x and v variables. Furthermore, Clarke's proof of (6.114) in [243, 245] was based on *reducing* the nonconvex Bolza problem under consideration to the *fully convex* problem comprehensively studied by Rockafellar. The *convex-valuedness* of Clarke's generalized gradient and its duality relationship with his generalized directional derivative played a *major role* in the possibility to accomplish the latter reduction and thus in the whole proof of (6.114).

Based on the Euler-Lagrange condition (6.114) for finite Lagrangians, Clarke obtained [247] its counterpart

$$\big(\dot p(t),p(t)\big)\in N_C\big((\bar x(t),\dot{\bar x}(t));\operatorname{gph}F(t)\big)\quad\text{a.e.} \tag{6.115}$$

for *Lipschitzian and bounded* differential inclusions (6.110) via his normal cone to the graph of $F=F(\cdot,t)$. Then he derived [248] the Euler-Lagrange inclusion (6.114) for the generalized Bolza problem (6.112), where ϑ was assumed to be extended-real-valued and epi-Lipschitzian in (x,v). The most notable and restrictive assumption imposed in [247, 248] was the *calmness* condition similar to that discussed in Subsect. 5.5.16 for problems of mathematical programming. This is a kind of constraint qualification/regularity requirement, which ensures the normal form of necessary optimality conditions and holds, in particular, when the endpoint function φ is locally Lipschitzian in either variable; the latter however excludes the corresponding endpoint constraints. Note that the calmness requirement allowed Clarke to avoid formally the convexity assumption on ϑ even in v, while the convexity property was actually present in [247, 248] due to the "admissible relaxation" provided by calmness; see also [246] for a detailed study of these relationships. Moreover, as mentioned in [248, p. 683], "...the [bi]convex case [developed by Rockafellar] lies at the heart of the proof of our result."

The most serious *drawback* of the Euler-Lagrange inclusion in form (6.115), fully recognized only later, is that it involves the Clarke normal cone to the *graph* of $F(\cdot,t)$ from (6.110), which happens to be a *linear subspace* of dimension $d\ge n$ whenever F is *graphically Lipschitzian* near the optimal solution; see Subsect. 1.4.4 for more discussions. Due to this property, the set on the right-hand side of (6.115) may be *too large* to provide an adequate information on adjoint arcs $p(\cdot)$ in many situations important for the theory and applications.

6.5.5. Clarke's Hamiltonian Condition. Besides the Euler-Lagrange condition (6.115), Clarke also established necessary optimality conditions for the generalized Bolza problem and thus for Lipschitzian differential inclusions in the following *Hamiltonian form*:

$$\big(-\dot{p}(t),\dot{\bar{x}}(t)\big)\in\partial_C\mathcal{H}\big(\bar{x}(t),p(t),t\big)\quad\text{a.e.}\tag{6.116}$$

involving his generalized gradient of the Hamiltonian function in both (x,p) variables. The first Hamiltonian results were obtained under the calmness assumption [253, 255] and then without this and other constraint qualifications [256].

Note that, in the absence of regularity/normality assumptions, the validity of the Hamiltonian condition (6.116) was established only for *convex-valued* differential inclusions (which corresponds to the convexity in v of the Lagrangian in the generalized Bolza form); the derivation of (6.116) without convexity originally presented in [251] was incorrect in the proof of Claim on p. 262 therein related to the convexification procedure. Similar approach based on employing the Ekeland variational principle worked nevertheless for proving Clarke's extension [250] of the Pontryagin maximum principle for *nonsmooth* optimal control systems of type (6.106). A long-standing conjecture about the validity of the Hamiltonian necessary optimality condition (6.116) without the above convexity assumption, which resisted the efforts of several authors, has been recently resolved by Clarke [261] for Lipschitzian and bounded differential inclusions by applying *Stegall's variational principle* [1224] instead of Ekeland's one in the framework of his proof. Observe that, in contrast to the classical smooth case and to the fully convex case of Rockafellar, Clarke's Euler-Lagrange condition (6.115) and Hamiltonian condition (6.116) are *not equivalent* even in simple situations. Moreover, they don't follow from each other being truly *independent*; see examples and discussions in Kaśkosz and Lojasiewicz [667] and in Loewen and Rockafellar [805].

It was not even clear till the work by Loewen and Rockafellar [804] whether one could find a *common adjoint arc* $p(\cdot)$ satisfying both Euler-Lagrange condition (6.115) and Hamiltonian condition (6.116) simultaneously. The affirmative answer was given in [804] for *convex-valued* and Lipschitzian differential inclusions with no assumption of calmness or normality. Note that in this case both conditions (6.115) and (6.116) automatically imply the *Weierstrass-Pontryagin maximum condition*

$$\big\langle p(t),\dot{\bar{x}}(t)\big\rangle=\mathcal{H}\big(\bar{x}(t),p(t),t\big)\quad\text{a.e.}\tag{6.117}$$

We refer the reader to [254, 255, 256, 267, 268, 272, 273, 274, 276, 595, 666, 667, 803, 804, 808, 1178, 1291, 1292] and the bibliographies therein for extensions and modifications of necessary optimality conditions of the Euler-Lagrange and Hamiltonian types obtained in terms of Clarke's generalized differential constructions for various problems of dynamic optimization and optimal control.

6.5.6. Transversality Conditions. Necessary optimality conditions in problems of dynamic optimization include, besides dynamic relations of the type discussed above (Euler-Lagrange, Hamiltonian, Weierstrass-Pontryagin), also endpoint relations on adjoint trajectories called *transversality conditions*. They are expressed via appropriate (generalized) differential constructions for cost and constraint functions depending on endpoints of state trajectories. Note that endpoint constraints on $(x(a), x(b))$ can be implicitly included in the endpoint cost function φ if it is assumed to be extended-real-valued as in the generalized problem of Bolza (6.112). However, typically such constraints are given explicitly in the form

$$\big(x(a), x(b)\big) \in \Omega \subset I\!R^n \, , \tag{6.118}$$

where the constraint/target set Ω may be specified in some functional form by, e.g., equalities and inequalities with real-valued (often Lipschitzian) functions.

In the afore-mentioned publications by Clarke and his followers concerning minimization of Lipschitzian cost functions φ as in (6.112) subject to endpoint constraints of type (6.118), the transversality conditions were derived in the form

$$\big(p(a), -p(b)\big) \in \lambda\partial_C\varphi\big(\bar{x}(a), \bar{x}(b)\big) + N_C\big((\bar{x}(a), \bar{x}(b)); \Omega\big) \tag{6.119}$$

with $\lambda \geq 0$ via Clarke's generalized gradient of φ and his normal cone to Ω at the optimal endpoints $(\bar{x}(a), \bar{x}(b))$. When φ and Ω happen to be convex, the transversality inclusion (6.119) reduces to that obtained earlier by Rockafellar [1143]. Note that the normal form $\lambda = 1$ holds under the calmness assumption and that a proper counterpart of (6.119) is expressed in terms of Clarke's normal cone to the epigraph of $\varphi + \delta(\cdot; \Omega)$ if φ is merely l.s.c. around $(\bar{x}(a), \bar{x}(b))$.

Transversality conditions in the significantly more *advanced form*

$$\big(p(a), -p(b)\big) \in \lambda\partial\varphi\big(\bar{x}(a), \bar{x}(b)\big) + N\big((\bar{x}(a), \bar{x}(b)); \Omega\big) \tag{6.120}$$

were first established by Mordukhovich in the mid-1970s via his basic/limiting normal cone and subdifferential: in [887] for time optimal control problems and in [889, 892] for other classes of problems in optimal control and dynamic optimization involving ODE control systems (6.106) and differential inclusions (6.110); see also [717, 897, 900, 901, 902, 904]. These results were obtained by the *method of metric approximations*, which was actually the driving force to introduce the nonconvex-valued normal cone and subdifferential in [887]; more comments and discussions were given in Subsects. 1.4.5 and 2.6.1.

It seems that the transversality conditions in form (6.120) didn't get a proper attention in the Western literature before Mordukhovich's talk at the

Montreal workshop (February 1989) and the publication of Clarke's second book [257], where these conditions were mentioned in footnotes with the reference to Mordukhovich; see Subsect. 1.4.8. However, even after that many papers (see, e.g., those listed in Subsect. 1.4.8) still continued using transversality conditions in form (6.119) instead of the advanced one (6.120).

Nevertheless, it has been eventually recognized the possibility to justify the advanced transversality conditions (6.120) in *any* investigated setting of dynamic optimization. We particularly refer the reader to the publications [33, 40, 93, 113, 258, 260, 261, 264, 265, 275, 443, 444, 506, 605, 611, 616, 801, 805, 806, 807, 845, 847, 878, 880, 914, 915, 916, 921, 932, 955, 959, 970, 971, 973, 974, 976, 1021, 1022, 1074, 1075, 1076, 1077, 1078, 1079, 1080, 1118, 1161, 1162, 1176, 1179, 1211, 1215, 1216, 1233, 1289, 1293, 1294, 1295, 1372], which clearly demonstrated this for various problems of the calculus of variations and optimal control of ordinary differential systems and their distributed-parameter counterparts.

6.5.7. Extended Euler-Lagrange Conditions for Convex-Valued Differential Inclusions. The usage of the nonconvex normal cone from [887] in the framework of *dynamic* optimality conditions for differential inclusions was initiated in the 1980 paper by Mordukhovich [892] for the problem of minimizing the cost function $\varphi(x(a),x(b))$ over absolutely continuous trajectories for the *convex-valued*, bounded, and Lipschitzian (in x) differential inclusion (6.110) subject to the endpoint constraints (6.118). Given an optimal solution $\bar{x}(\cdot)$ to this problem, a dynamic necessary optimality condition was obtained in [892] in the form

$$\begin{aligned}\big(\dot{p}(t),\dot{\bar{x}}(t)\big)\in\operatorname{co}\Big\{(u,v)\in I\!R^{2n}\Big|\big(u,p(t)\big)\in N\big((\bar{x}(t),v);\operatorname{gph}F(t)\big)\ ,\\ v\in M\big(\bar{x}(t),p(t),t\big)\Big\}\ \text{ a.e. }\ t\in[a,b]\end{aligned}\tag{6.121}$$

with the *argmaximum sets* $M(x,p,t)$ defined by

$$M(x,p,t):=\big\{v\in F(x,t)\big|\ \langle p,v\rangle=\mathcal{H}(x,p,t)\big\}$$

and the transversality inclusion (6.120) held when φ is locally Lipschitzian. If the argmaximum set $M(\bar{x}(t),p(t),t)$ is a *singleton* for a.e. $t\in[a,b]$ (it happens, in particular, when the velocity set $F(\bar{x}(t),t)$ is *strictly convex* almost everywhere), condition (6.121) reduces to

$$\big(\dot{p}(t),\dot{\bar{x}}(t)\big)\in\operatorname{co}\Big\{(u,v)\Big|\ \big(u,p(t)\big)\in N\big((\bar{x}(t),\dot{\bar{x}}(t));\operatorname{gph}F(t)\big)\Big\}\ \text{ a.e.}\tag{6.122}$$

It is worth mentioning that these results were derived in [892] with *no* calmness and/or any other qualification conditions by using the method of *discrete approximations*; see Subsect. 6.5.12 for more discussions on this technique.

Observe that in contrast to Clarke's Euler-Lagrange condition (6.115) requiring the *full convexification* of the basic normal cone (since $N_C=\operatorname{clco}N$), both conditions (6.121) and (6.122) involve only a *partial* convexification, which allows us to avoid troubles with the subspace property of the Clarke normal cone to graphical sets.

Condition (6.122) obviously implies the Euler-Lagrange condition in Clarke's form (6.115); it is easy to find examples when (6.122) is strictly better. This is however not the case regarding the comparison between (6.115) and (6.121) when the velocity sets $F(x,t)$ are not strictly convex. Indeed, there are examples in Loewen and Rockafellar [805] showing that these two necessary optimality conditions are generally *independent*. Moreover, it has been subsequently proved by Ioffe [603] and Rockafellar [1162] (as the two complementary implications) that Mordukhovich's initial version of the Euler-Lagrange condition (6.121) for convex-valued differential inclusions happens to be *equivalent* to Clarke's Hamiltonian condition (6.116).

We refer the reader to other publications by Mordukhovich [901, 902, 908] containing the developments of condition (6.121), and thus of (6.122) in the case of strictly convex velocity sets, for various dynamic optimization problems involving convex-valued (or relaxed) differential inclusions; in particular, for problems with free time, intermediate state constraints, Bolza-type functionals, etc. Developing then the discrete approximation techniques of [892, 901, 902, 908], Smirnov [1215] established the validity of the refined Euler-Lagrange condition (6.122) for (*not strictly*) convex-valued, Lipschitzian, bounded, and autonomous differential inclusions by reduction them in fact to the strictly convex case.

Further results in this direction were obtained by Loewen and Rockafellar [805] for convex-valued and *unbounded* differential inclusions of type (6.110), with the replacement of the standard Lipschitzian property of $F(\cdot,t)$ for bounded inclusions by its "integrable sub-Lipschitzian" counterpart in the unbounded case. They derived the Euler-Lagrange condition in the advanced form (6.122) emphasizing that "two simple themes underlie our approach: *truncation* and *strict convexity*." The latter means that they developed an efficient technique allowing them to reduce the general case under consideration to bounded and Lipschitzian differential inclusions, for which condition (6.121) held and agreed with the refined one (6.122). Note that the *convexity* assumption on the sets $F(x,t)$ played a crucial role in the technique developed in [805]. The two subsequent papers by Loewen and Rockafellar [806, 807] contained extensions of these results to the generalized problem of Bolza with state constraints and free time. It is worth mentioning that in [806] the general Bolza case with an extended-real-valued integrand/Lagrangian in (6.112) was reduced under mild "epi-continuity" and growth assumptions to a Mayer problem for an unbounded differential inclusion satisfying the "integrable sub-Lipschitzian" property of [805]; moreover, the *coderivative criterion* for Lipschitz-like behavior established by Mordukhovich [909] (see Theorem 4.10) served as a key technical ingredient in justifying the possibility of such a reduction.

At this point we observe that the Euler-Lagrange inclusion (6.122) can be equivalently written in the *coderivative form*

$$\dot{p}(t) \in \mathrm{co} D_x^* F\big(\bar{x}(t), \dot{\bar{x}}(t), t\big)\big(- p(t)\big) \quad \text{a.e.}\,, \tag{6.123}$$

which was actually the *original motivation* for introducing the coderivative construction in [892] (as the *adjoint mapping* to F) to describe adjoint systems in optimal control problems governed by discrete-time and differential inclusions. Since the coderivative reduces to the *adjoint Jacobian* for smooth single-valued mappings, relation (6.123) can be viewed as an appropriate extension of the *adjoint system* (6.107) to generalized control processes governed by differential inclusions. Note that the Hamiltonian form of necessary optimality conditions as in (6.113) *doesn't offer* such an extension in the nonsmooth setting. Besides an intrinsic esthetic value, form (6.123) carries a powerful *technical component* allowing us to employ comprehensive coderivative calculus and dual characterizations of Lipschitzian and related properties to the study of many issues in control theory for differential inclusions, particularly those concerning limiting processes; see, e.g., the above proofs of the major results presented in Sects. 6.1 and 6.2 of this book.

6.5.8. Extended Euler-Lagrange and Weierstrass-Pontryagin Conditions for Nonconvex-Valued Differential Inclusions. As mentioned, the results discussed in Subsect. 6.5.7 (as well as the previous versions reviewed in Subsect. 6.5.6) were derived under the *convexity* hypothesis imposed on the velocity sets $F(x,t)$ of differential inclusions in the absence of calmness-like assumptions. Necessary optimality conditions for *nonconvex*-valued (while Lipschitzian and bounded) differential inclusions with endpoint constraints involving the *extended Euler-Lagrange* condition (6.123) were first established by Mordukhovich [915] without any constraint qualifications. Observe that the Euler-Lagrange condition in Clarke's *fully convexified* form (6.115) was previously obtained by Kaśkosz and Lojasiewicz [667] for boundary trajectories of nonconvex, bounded, and Lipschitzian differential inclusions. In [915], the reader can find the corresponding version of the extended Euler-Lagrange condition (6.123) for the Bolza problem (6.112) with a finite nonconvex integrand over nonconvex differential inclusions, while another paper by Mordukhovich [916] concerned problems with free time.

The *Weierstrass-Pontryagin maximum condition* (6.117) doesn't play an independent role for convex-valued differential inclusions, since it follows automatically from any version of the Euler-Lagrange conditions discussed above. This is *no* longer true in the nonconvex setting for which the maximum condition was not established in the afore-mentioned papers [667, 915]. Nevertheless, it was asserted in [915, Remark 7.6] that the methods developed therein would allow us to prove (6.117) accompanying the refined Euler-Lagrange condition (6.123) if the classical Weierstrass necessary condition would be established for strong minimizers of the Bolza problem with *finite* Lagrangian and free endpoints without imposing any smoothness and/or convexity assumptions. The latter task was first accomplished by Ioffe and Rockafellar [616] who derived the counterpart

$$\dot{p}(t) \in \operatorname{co}\left\{u \in I\!R^n \,\middle|\, \big(u, p(t)\big) \in \partial\vartheta\big(\bar{x}(t), \dot{\bar{x}}(t), t\big)\right\} \text{ a.e.} \tag{6.124}$$

of the extended Euler-Lagrange condition (6.123) accompanied by the classical Weierstrass condition, valid for all $v \in I\!R^n$ and a.e. t,

$$\vartheta\big(\bar{x}(t), v, t\big) \geq \vartheta\big(\bar{x}(t), \dot{\bar{x}}(t), t\big) + \langle p(t), v - \dot{\bar{x}}(t)\rangle \tag{6.125}$$

for the *nonconvex* Bolza problem (6.112) with the finite (real-valued) integrand ϑ.

Based on Ioffe-Rockafellar's result and the techniques of [915], Mordukhovich derived in [914] the Euler-Lagrange condition (6.123) accompanied by the Weierstrass-Pontryagin maximum condition (6.117) for nonconvex differential inclusions under the *boundedness* and Lipschitzian assumptions on F with respect to x. More general results of this type were then obtained in the concurrent papers by Ioffe [604] and Vinter and Zheng [1294] who derived, by different techniques, the extended Euler-Lagrange (6.123) and Weierstrass-Pontryagin (6.117) necessary optimality conditions for nonconvex and *unbounded* differential inclusions under the *integrable sub-Lipschitzian* assumption by Loewen and Rockafellar [805]. It is interesting to observe that Vinter and Zheng [1294] gave another proof of Ioffe-Rockafellar's results (6.124) and (6.125) for problems with finite Lagrangians based on their reduction to optimal control problems for systems with *smooth dynamics* and *nonsmooth endpoint constraints* employing to them the version of the maximum principle with transversality conditions (6.120) originally obtained in the 1976 paper by Mordukhovich [916]. We also refer the reader to the subsequent papers by Vinter and Zheng [1295, 1296, 1297] for appropriate versions of the extended Euler-Lagrange and Weierstrass-Pontryagin conditions to problems with state constraints and free time, and also to their applications. Furthermore, Rampazzo and Vinter [1118] generalized these results for nonconvex differential inclusions with the so-called *degenerated* state constraints providing nondegenerate necessary optimality conditions for problems in which endpoints may belong to the boundary of state constraints, and so the standard necessary conditions convey no useful information. See also Arutyunov and Aseev [33], Ferreira, Fontes and Vinter [443] with the references therein for previous results concerning degenerate control problems.

Quite recently, Clarke [260, 261] derived necessary optimality conditions in the extended Euler-Lagrange form (6.123) accompanied by the Weierstrass-Pontryagin maximum condition (6.117) for nonconvex and unbounded differential inclusions under *fairly weak* (probably minimal) assumptions on the initial data. In the process of proof, he developed a delicate and powerful technique involving *smooth variational principles* and *decoupling* machinery that allowed him to reduce these conditions under the weak assumptions made to the settings already known and discussed above. The conditions derived in [260, 261] also incorporated a novel *stratified* feature in which both the assumptions and conclusions were formulated relative to a *prescribed* radius function. They also gave rise to new forms of the so-called "hybrid maximum principle" for optimal control problems with cost integrands of a very general nature while with the smooth underlying dynamics.

Note that in certain special situations potentially *stronger* versions of the extended Euler-Lagrange condition can be obtained for minimizing nonconvex and nonsmooth integral functionals of the calculus of variations and related problems. To this end we refer the reader to the papers by Ambrosio, Ascenzi and Buttazzo [17], Marcelli [845, 846], and Marcelli, Outkine and Sytchev [847], where some versions of the Euler-Lagrange conditions via the *subdifferential of convex analysis* were derived for *nonconvex* problems with some special structures. The results of this type are heavily based on *relaxation techniques* particularly involving the Lyapunov convexity theorem [822] and its various extensions and modifications.

6.5.9. Dualization and Extended Hamiltonian Formalism. In Subsects. 6.5.5 and 6.5.7 we have discussed some relationships between the previous versions of the Euler-Lagrange and Hamiltonian optimality conditions for differential inclusions and for the generalized problem of Bolza. Recall that, in contrast to the classical smooth and fully convex cases, Clarke's versions of the Euler-Lagrange (6.115) and Hamiltonian (6.116) conditions are *not equivalent* even in simple settings, while his Hamiltonian condition happens to be *equivalent* to the early Mordukhovich's version of the Euler-Lagrange condition (6.121) for convex-valued differential inclusions. What about an appropriate *Hamiltonian* counterpart of the *extended* Euler-Lagrange condition written as (6.122), or equivalently as (6.123), for differential inclusions and as (6.124) and the problem of Bolza in the *absence of strict convexity*?

This question was first investigated by Rockafellar [1162] in the general framework of the *Legendre-Fenchel transform* (or the *conjugacy correspondence*) of convex analysis defined by the classical formula

$$\vartheta^*(x,p) = \sup_{v\in I\!R^n}\big\{\langle p,v\rangle - \vartheta(x,v)\big\}\,. \tag{6.126}$$

It is well known from convex analysis [1142] that for any proper, *convex*, and l.s.c. function $\vartheta(x,\cdot)\colon I\!R^n \to \overline{I\!R}$ the conjugate function $\vartheta^*(x,\cdot)$ enjoys the same properties on $I\!R^n$ satisfying moreover the symmetric *biconjugacy* relationship

$$\vartheta(x,v) = \sup_{p\in I\!R^n}\big\{\langle p,v\rangle - \vartheta^*(x,p)\big\}\,.$$

The question stated and resolved by Rockafellar [1162] was about relationships between *basic subgradients* of the functions $\vartheta(x,v)$ and $\vartheta^*(x,p)$ with respect to their *both* variables. Under a certain "epi-continuity" assumption, which automatically holds when either ϑ or ϑ^* is locally Lipschitzian around the reference point, it was established in [1162] the following relationship for the *convex hulls*:

$$\mathrm{co}\left\{u \in I\!R^n \,\middle|\, (u, p) \in \partial\vartheta(x, v)\right\} = -\mathrm{co}\left\{u \in I\!R^n \,\middle|\, (u, v) \in \partial\vartheta^*(x, p)\right\}. \quad (6.127)$$

For the case corresponding to differential inclusions, with $\vartheta(x, v) = \delta((x, v); \mathrm{gph}\, F)$, the relationships (6.127) reduces to

$$\mathrm{co}\left\{u \in R^n \,\middle|\, (u, p) \in N\big((x, v); \mathrm{gph}\, F\big)\right\} = \mathrm{co}\left\{u \in I\!R^n \,\middle|\, (-u, v) \in \partial\mathcal{H}(x, p)\right\}$$

by taking into account (6.126) and the Hamiltonian construction (6.111). The proof of the *Rockafellar dualization theorem* (6.127) given in [1162] was rather involved based on advanced tools of convex analysis in finite dimensions including Moreau-Yosida's approximation techniques, Wijsman's epi-continuity theorem, Attouch's theorem on convergence of subgradients, etc.

In view of (6.127), the advanced/*extended Hamiltonian* form *equivalent* to the extended Euler-Lagrange condition (6.123) for *convex*-valued differential inclusions reads as follows:

$$\dot{p}(t) \in \mathrm{co}\left\{u \in I\!R^n \,\middle|\, \big(-u, \dot{\bar{x}}(t)\big) \in \partial\mathcal{H}\big(\bar{x}(t), p(t), t\big)\right\} \ \text{a.e.} \quad (6.128)$$

The same form of the extended Hamiltonian condition holds true for the generalized Bolza problem (6.112), with the Hamiltonian defined accordingly as the conjugate of the Lagrangian integrand $\vartheta(x, p, t)$ in the velocity variable v. The elaboration of the assumptions needed for the fulfillment of the associated Euler-Lagrange condition (6.124) together with the equivalent Hamiltonian form (6.128) in the framework of the generalized problem of Bolza with the integrand $\vartheta(x, v, t)$ *convex* in v was given by Loewen and Rockafellar [806]; see the corresponding discussions on the extended Euler-Lagrange condition in Subsect. 6.5.7, presented right before (6.123), which can now be equally relate to the Hamiltonian condition (6.128) due to Rockafellar's dualization result (6.127).

In [604], Ioffe established the *inclusion* "$\subset$" in (6.127) under significantly weaker assumptions in comparison with those in Rockafellar [1162], while still under the *convexity* of $\vartheta(x, \cdot)$. Employing this result, he justified necessary optimality conditions in both Euler-Lagrange (6.123) and Hamiltonian (6.128) forms for *convex*-valued and *unbounded* differential inclusions with the replacement of the "integrable sub-Lipschitzian" property as in Loewen and Rockafellar [806] by the more general Lipschitz-like (Aubin's "pseudo-Lipschitzian") property of $F(\cdot, t)$. Observe that Ioffe's proof clearly reveals the *pivoting role* of the Euler-Lagrange condition (6.123) in nonsmooth optimal control, which holds with *no* convexity assumptions (see Subsect. 6.5.8) and directly implies the extended Hamiltonian condition (6.128) for convex-valued problems. Note to this end that the validity of the latter Hamiltonian inclusion (6.128) for *nonconvex* problems is still an *open question*, even for bounded and Lipschitzian differential inclusions.

Another proof of the inclusion "⊂" in Rockafellar's dualization theorem (6.127) under about the same hypotheses as in [1162] was later given by Bessis, Ledyaev and Vinter [113] (see also Sect. 7.6 in Vinter's book [1289]). The proof of [113, 1289] employed not Moreau-Yosida's approximations as in [604, 1162] but more direct and conventional (while rather involved) techniques of *proximal analysis.*

6.5.10. Other Techniques and Results in Nonsmooth Optimal Control. It is worth mentioning that, as shown by Ioffe [604], the advanced Euler-Lagrange formalism for *nonconvex* differential inclusions discussed in Subsect. 6.5.8 easily implies a *nonsmooth* version of the *Pontryagin maximum principle* for parameterized control systems of type (6.106) with the adjoint equation

$$-\dot{p}(t) \in \big[J_x f\big(\bar{x}(t), \dot{\bar{x}}(t), t\big)\big]^* p(t) \ \text{ a.e.} \tag{6.129}$$

written via Clarke's generalized Jacobian $J_x f$ of f with respect to x. Recall that the *generalized Jacobian* [252, 255] of a Lipschitzian mapping $f\colon I\!R^n \to I\!R^m$ is defined as the *convex hull* of the classical Jacobian $m \times n$ matrices at points $x_k \to \bar{x}$; the latter set is nonempty and compact by the fundamental Rademacher's theorem [1114]. Such a nonsmooth maximum principle involving the adjoint equation (6.129) was first obtained by Clarke [250, 255] directly for control systems (6.106) based on approximation procedures via *Ekeland's variational principle.* Note also that Ioffe [604] deduced the maximum principle in the somewhat more advanced form suggested by Kaśkosz and Lojasiewicz [666] for parameterized families of *vector fields* from the extended Euler-Lagrange formalism for differential inclusions.

Probably the very *first extension* of the Pontryagin maximum principle to nonsmooth control systems was published by Kugushev [722] who employed a certain constructive technique to approximate the given nonsmooth system by a sequence of *smooth* ones. However, he didn't described efficiently the resulting set of "subgradients" that appeared in this procedure. Other early results on the nonsmooth maximum principle for systems (6.106) were independently obtained by Warga [1316, 1317, 1321] (starting with the end of 1973) using some smooth approximation technique of the *mollifier* type and his *derivate containers* for mappings $f\colon I\!R^n \to I\!R^m$. The latter objects, which are *not uniquely* defined, give more precise results than Clarke's generalized Jacobian in some settings of variational analysis, optimization, and control. However, the *convex hull* of any derivate container provides no more information than the generalized Jacobian (as shown in [1320]), and thus the adjoint system in form (6.129) subsumes that of Warga [1316].

Warga's approach to derive necessary optimality and controllability conditions was extended by Zhu [1370] to nonconvex *differential inclusions* satisfying, besides the standard assumptions of boundedness and Lipschitz continuity, also requirements on the existence of some *local selections*, which were incorporated in the optimality conditions obtained in [1370]. An obvious drawback of such and related conditions (see, e.g., Tuan [1273]) is the absence of *any analytic mechanism* for obtaining required selections, even in the case of convex-valued inclusions. Similar remarks on the possibility to constructively verify assumptions and conclusions explicitly involving certain *auxiliary objects* of approximation and linearization types can be equally addressed to some other necessary optimality conditions for nonsmooth optimal control and variational problems obtained particularly by Frankowska [464, 465, 468] and by Polovinkin and Smirnov [1094, 1095]; cf. also Ahmed and Xiang [6] for problems involving infinite-dimensional differential inclusions.

Note that there is another direction in the theory of necessary optimality conditions for differential inclusions, developed mostly in the Russian school, that aims to derive results for differential inclusions by limiting procedures from the Pontryagin maximum principle for *smooth* optimal control problems involving systems of type (6.106). In this way, using different kinds of *smooth approximations*, some interesting results mainly related to those already known in the theory of *convex-valued* differential inclusions were obtained by Arutyunov, Aseev and Blagodatskikh [34], Aseev [39, 40, 41], and Milyutin [875, 876]; the latter paper was the last work by Alexei Alexeevich Milyutin submitted and published after his death.

On the other way of development, new results for *nonsmooth* control systems (6.106) different from Clarke's version of the nonsmooth maximum principle with the adjoint equation (6.129) were obtained by de Pinho, Vinter, and their collaborators using an appropriate approximation of control systems by differential inclusions with the help of Ekeland's variational principle. These results are described via *joint* subgradients of the Hamilton-Pontryagin function (6.108), called sometimes the *unmaximized indexunmaximized Hamiltonian Hamiltonian*, in the (x, p, u) variables. The first result of this type was derived by de Pinho and Vinter [1078] for standard optimal control problems with endpoint constraints under the name of "Euler-Lagrange inclusion," which didn't seem to be in accordance with the real essence of this condition. Then the name has been appropriately changed, and the results of this type were labeled as necessary optimality conditions for nonsmooth control systems involving the *unmaximized Hamiltonian inclusion* (UHI); see [1076] for more discussions. The subsequent papers of these authors and their collaborators [1074, 1075, 1076, 1077, 1079, 1080] contained various extensions of the UHI type results to optimal control problems with state constraints, with mixed constraints on control and state variables, with algebraic-differential constraints, etc. The results of this type are particularly efficient for *weak minimizers*; cf. also the related paper by Páles and Zeidan [1036]. One of the strongest advantages (as well as the original motivation) of the UHI formalism in comparison with Clarke's version of the nonsmooth maximum principle is that the possibility to get *necessary and sufficient* conditions for optimality in nonsmooth *convex* control problems, which is not the case for Clarke's formalism (6.129).

6.5.11. Dual versus Primal Methods in Optimal Control. Observe that the majority of techniques developed for optimization of *differential inclusions* don't employ the *method of variations* and its modifications that lie at the heart of the classical calculus of variations and optimal control dealing with parameterized control systems of type (6.106). Perhaps the most significant technical reason for this in the context of differential inclusions (6.110) relates to the fact that the method of variations based on the comparison between the given optimal solution and its small (in some sense) local variations doesn't fit well to the very nature of the dynamic constraints $\dot{x} \in F(x)$ and also of control constraints of the type $u \in U(x)$ with the *state-dependent* control region $U(x)$.

Alternative approaches to developing necessary optimality conditions for differential inclusions, as well as for constrained control systems of type (6.106), are based on certain *approximation/pertubation* procedures concerning the *whole problem* under consideration, not only its optimal solution. This may involve various approximations of dynamic optimization problems by those with *no* right-endpoint constraints (which are much easier to handle), exact penalization, decoupling, discrete approximations, etc.; see more details and discussions in Clarke [250, 255], Ioffe [604, 611], Mordukhovich [887, 915], Vinter [1289] with their references.

The techniques and results of the latter type lead to *subgradient-oriented* theories of necessary conditions in nonsmooth optimization and optimal control involving generalized differential constructions in *dual spaces* (normal cones, subdifferential, coderivatives). It seems that the strongest general results of this type are expressed in terms of our basic/limiting dual-space constructions, which *cannot* be generated by derivative-like objects in primal spaces (as tangent cones and directional derivatives) due to their *intrinsic nonconvexity.* This allows us to unify the results obtained in this direction under the name of *dual-space theory.*

On the other line of developments, approaches and results related to the method of variations and its modifications deal with variations and perturbations of optimal solutions in *primal spaces* involving various *tangential* approximations, particularly of reachable sets for control systems; see, e.g., the proof of the Pontryagin maximum principle in [1102] and the subsequent developments by Dubovitskii and Milyutin [370, 877], Halkin [539, 545], Neustadt [1001, 1002], Warga [1315, 1316], and others. We refer to results of this type as to *primal-space theory.* Note that this terminology is *not* in accordance with the one adopted by Vinter [1289, pp. 228–231].

Necessary optimality conditions for nonsmooth optimal control obtained in the dual-space and primal-space theories are generally *independent* from the viewpoints of treated *local minimizers*, employed *analytic machineries*, and imposed *assumptions* on the initial data. In more detail:

—–Types of local minima investigated by primal-space methods depend on the *variations* used, while dual-space methods deal with local minimizers defined *regardless* of variations.

—–Realizations and implementations of primal-space methods heavily depend on using *powerful tools* of nonlinear analysis (including open mapping and implicit function theorems and/or fixed-point results), while dual-space methods are *free* of this machinery employing instead more simple *penalty-type* techniques in finite dimensions as well as modern *variational principles* in infinite-dimensional settings.

—–Assumptions needed for approximation/perturbation techniques in dual-space theory require good behavior *around* points of minima (e.g., Lipschitzian properties and metric regularity), while primal-space techniques may produce results under *at-point* assumptions.

—–Primal-space methods for (smooth and nonsmooth) *constrained* optimization (including constrained optimal control) require finally the usage of *convex separation* for obtaining efficient results in eventually dual terms (Lagrange multipliers, adjoint trajectories, etc.), while dual-space methods *don't appeal* as a rule to convex separation theorems.

In Sect. 6.3, the reader can find some advanced results in the primal-space direction derived in the *conventional PMP form* and its *upper subdifferential* extension. The obtained results concern parameterized control systems of type (6.106) with *smooth dynamics* in *infinite-dimensional* spaces and endpoint equality and inequality constraints described by finitely many real-valued functions. However, these functions may be merely *Fréchet differentiable at* the reference optimal point, *not* even being *continuous* around it (the latter applies only to the functions describing the endpoint objective as well as inequality constraints); see more comments to the material of Sect. 6.3 presented below.

The most general results of the primal type in *nonsmooth* optimal control for *finite-dimensional* systems have been developed by Sussmann during the last decade; see [1235, 1236, 1237, 1238] and the references therein. He started [1235] with the remarkable result called the *Lojasiewicz refinement* of the maximum principle that came out of Lojasiewicz's idea formulated in the unpublished (and probably unfinished) paper [810]. This refinement consists of justifying a version of the PMP by assuming that the velocity mapping $f(x,u,t)$ in (6.106) is not $\mathcal{C}^1$ with respect to x for all $u\in U$ a.e. in t as in the classical PMP and not locally Lipschitzian in x for all $u\in U$ and a.e. t as in Clarke's nonsmooth version of the PMP under "minimal hypotheses" [250] but merely locally Lipschitzian in x *along the given optimal control* $u=\bar{u}(t)$ for a.e. t. A "weak differentiable" version of this result justifies the validity of the

PMP when $f(\cdot, \bar{u}(t), t)$ is *differentiable* (possibly not strictly differentiable) *at one point* $\bar{x}(t)$ along the optimal control $u = \bar{u}(t)$ for a.e. t.

Sussmann proved these results and their far-going generalizations in nonsmooth optimal control developing certain abstract versions of *needle variations* (crucial in the proof of the classical PMP) and *primal-space* constructions of generalized differentials. In the recent paper [38], Arutyunov and Vinter provided a simplified proof of the "weak differentiable" version in the Lojasiewicz refinement of the PMP based on the so-called "inner finite approximations" involving special needle-type variations of the reference optimal control $\bar{u}(\cdot)$ that *don't violate* endpoint constraints on trajectories. The idea of this finite approximation scheme goes back to Tikhomirov being published in [7], where it was applied to the classical PMP in smooth optimal control. Further results in this direction were derived by Shvartsman [1209] for nonsmooth control systems with state constraints.

6.5.12. The Method of Discrete Approximations. Section 6.1 is devoted to a thorough study of dynamic optimization problems in infinite-dimensional spaces by using the method of *discrete approximations*. Although our primary goal is to develop this method as a *vehicle* to derive necessary optimality conditions of the *extended Euler-Lagrange* type (6.123) for dynamic processes governed by nonconvex differential/evolution inclusions, we also present some results of *numerical value* for such processes that concern *well-posedness* and *convergence* issues for discrete approximations of evolution inclusions *with* and *without* optimization involved. It seems that neither necessary optimality conditions for *infinite-dimensional evolution inclusions* nor discrete approximations of such processes have been previously considered in the literature besides the author's recent paper [932], where some of the results obtained in this book were announced. They follow however a series of finite-dimensional developments; see below.

The method of discrete approximations for the study of continuous-time systems goes back to Euler [411] who developed it to establish the famous first-order necessary condition (known now as the Euler or Euler-Lagrange equation) for minimizing integral functionals in the one-dimensional calculus of variations. It is significant to note that Euler regarded the integral under minimization as an *infinite sum* and *didn't employ* limiting processes interpreting instead (via a geometric diagram) the differentials along the minimizing curve as *infinitesimal changes* in comparison with "broken lines," i.e., finite differences. Euler's derivation of the necessary optimality condition in *one equational form* for a "general" (at that time) problem of the calculus of variations signified a major theoretical achievement providing the synthesis of many special cases and examples appeared in the work of earlier researchers. It is worth mentioning that an approximation idea based on replacing a curve by broken lines was partly (and rather vaguely) used by Leibniz [757] in his solution of the brachistochrone problem in the very beginning of the calculus of variations.

Since that time, Euler's finite-difference method and its modifications have been widely employed in various areas of dynamic optimization and numerical analysis of differential systems, with mostly numerical emphasis that has become more significant in the computer era. There is an abundant literature devoted to different aspects of discrete approximations and their numerous applications; we refer the reader to [28, 98, 184, 185, 220, 221, 298, 299, 302, 303, 338, 343, 344, 345, 346, 347, 348, 349, 353, 354, 357, 358, 359, 367, 407, 425, 488, 520, 535, 542, 702, 721, 760, 761, 828, 831, 832, 890, 892, 900, 901, 902, 908, 915, 916, 941, 959, 973, 974, 976, 1012, 1061, 1062, 1086, 1107, 1109, 1215, 1175, 1216, 1280, 1282, 1283, 1284, 1301, 1333, 1379] and the bibliographies therein for representative publications related to dynamic optimization and control systems.

In Sect. 6.1 we extend to the general infinite-dimensional setting of nonconvex evolution/differential inclusions the basic constructions and results of the method of discrete approximations developed previously by Mordukhovich [915] for differential inclusions in finite-dimensional spaces; see also [890, 892, 901, 902, 908, 1107, 1109, 1215, 1216] and the comments below for the preceding work in this direction concerning convex-graph and convex-valued differential inclusions in finite dimensions.

The *underlying idea* and the *basic scheme* of the method of discrete approximations to derive necessary optimality conditions for variational problems involving differential inclusions contain the following *three major components*:

(i) to replace/approximate the original continuous-time variational problem by a *well-posed* sequence of discrete-time optimization problems whose optimal solutions *converge*, in a certain suitable sense, to some (or to the given) optimal solution for the original problem;

(ii) to derive necessary optimality conditions in discrete-time problems of dynamic optimization by reducing them to constrained problems of mathematical programming, which occur to be *intrinsically nonsmooth*, and then by employing *appropriate* tools of *generalized differentiation* with good *calculus*;

(iii) to establish *robust/pointbased* necessary optimality conditions for the original continuous-time dynamic optimization problem by *passing to the limit* from necessary conditions for its discrete approximations and by using the *convergence/stability* results obtained for the discrete approximation procedure together with the corresponding properties of the generalized differential constructions that ensure the required convergence of *adjoint* trajectories.

In Mordukhovich's paper [915], the described discrete approximation scheme was implemented for the general Bolza problem governed by *nonconvex* differential inclusions in *finite-dimensional* spaces; the extended Euler-Lagrange condition of the advanced type (6.123) was first established there in this way for nonconvex problems. The realization of each of the three steps (i)–(iii) listed above for evolution inclusions in *infinite dimensions* requires certain additional developments most of which happen to be significantly different from the finite-dimensional setting.

6.5.13. Discrete Approximations of Evolution Inclusions. The main aspects of the theory of differential inclusions of type (6.1) in infinite-dimensional spaces, called often *evolution inclusions*, are presented in the books by Deimling [314] and by Tolstonogov [1258], while much more is available for differential inclusions in finite dimensions; see, e.g., the books by Aubin and Cellina [50] and by Filippov [450] with the references therein. We follow Deimling [314] in Definition 6.1 of solutions to differential/evolution inclusions in Banach spaces. Note that it differs from Carathéodory solutions in finite dimensions (which go back to [222] in the case of differential equations) by the additional requirement on the validity of the *Newton-Leibniz formula* in terms of the Bochner integral; the latter is not automatic for absolutely continuous mappings with infinite-dimensional values. On the other hand, there is a *precise characterization* of Banach spaces, where the fulfillment of the Newton-Leibniz formula is *equivalent* to the absolute continuity: these are spaces with the *Radon-Nikodým property* (RNP) for which more details are available in the classical monographs by Bourgin [169] and by Diestel and Uhl [334]. The latter property is fundamental in functional analysis; in particular, its validity for the *dual space* X^* is *equivalent* to the *Asplund property* of X. This justifies another line of using the remarkable class of Asplund spaces in the book.

The principal result of Subsect. 6.1.1, Theorem 6.4, justifies a constructive algorithm to *strongly approximate* (in the norm of the Sobolev space $W^{1,2}([a,b];X)$ ensuring particularly the a.e. *pointwise* convergence with respect to *velocities*) of *any* given feasible trajectory for the Lipschitzian differential inclusion (6.1) in arbitrary Banach space X by extended trajectories of its *finite-difference* counterparts (6.3) obtained by using the standard Euler scheme. This result is an infinite-dimensional version of that by Mordukhovich [915, Theorem 3.1] (with just a little change in the proof) extending his previous constructions and results from [901, 902] and those from Smirnov's paper [1215]; see also [1216]. This theorem, besides its independent interest and numerical value to justify an efficient procedure for approximating the set of feasible solutions to a general differential inclusion *regardless of optimization*, provides the foundation for constructing *well-posed* discrete approximations of variational problems for continuous-time evolution systems.

Observe that we don't impose in Theorem 6.4 *any convexity* assumptions on the velocity sets $F(x,t)$ and realize the *proximal algorithm* based on the *projection of velocities* in (6.10). This distinguishes the *velocity approach* from more conventional results on discrete approximations of (convex-graph or convex-valued) differential inclusions involving projections of state vectors and ensuring merely the $\mathcal{C}([a,b];I\!\!R^n)$-norm convergence of trajectories; see, e.g., Pshenichnyi [1107, 1109] and the survey papers by Dontchev and Lempio [359] and by Lempio and Veliov [761]. We emphasize that the latter convergence doesn't allow us to deal with nonconvex inclusions (since the uniform convergence of trajectories corresponds to the *weak* convergence of derivatives and eventually requires the subsequent *convexification* by the Mazur weak closure theorem) and that the achievement of the a.e. *pointwise* convergence of derivatives/velocities plays a *crucial role* in the possibility to establish necessary optimality conditions for *nonconvex problems*.

Let us mention two recent developments on the convergence of discrete approximations in direction (i) listed in Subsect. 6.5.12. In [343], Donchev derived some extensions of the approximation and convergence results from the afore-mentioned paper [915] to finite-dimensional differential inclusions whose right-hand side mappings $F(x,t)$ satisfy the so-called *Kamke* condition with respect to x, where the standard Lipschitz modulus is replaced by a Kamke-type function. The latter property happens to be *generic* (in Baire's sense) in the class of all continuous multifunctions $F(\cdot,t)$. The other work is due to Mordukhovich and Pennanen [941] who established the *epi-convergence* of discrete approximations in the *generalized Bolza* framework under certain *convexity* and Lipschitzian assumptions.

6.5.14. Intermediate Local Minima. In Subsect. 6.2.2 we start studying the Bolza problem for constrained differential/evolution inclusions in Banach spaces following mainly the procedure developed by Mordukhovich [915] in finite dimensions, with some significant infinite-dimensional changes on which we comment below. Note that, in contrast to the generalized Bolza problem in form (6.13) with extended-real-valued functions φ and ϑ implicitly incorporating endpoint and dynamic constraints, we deal with such constraints *explicitly*, since the continuity and Lipschitzian assumptions imposed on φ and ϑ in the results obtained in Sect. 6.1 *exclude* in fact the infinite values of these functions.

The main attention in our study is paid to the notions of *intermediate local minima* of rank $p \in [0,\infty)$ (i.l.m.; see Definition 6.7) and its *relaxed* version (r.i.l.m.; see Definition 6.12). Both notions were introduced by Mordukhovich [915] and were later studied by Ioffe and Rockafellar [616], Ioffe [604], Vinter and Woodford [1293], Woodford [1331], Vinter and Zheng [1294, 1295, 1289], Vinter [1289], and Clarke [260, 261] for various dynamic optimization problems, mostly in the case of $p = 1$, referred to as $W^{1,1}$ local minimizers.

Intermediate local minimizers occupy an *intermediate position* between the classical *weak* and *strong* minimizers for variational problems; that is where this name came from in [915]. Examples 6.8–6.10 show that these three major types of local minimizers may be different even in relatively simple problems of dynamic optimization problems involving particularly convex-valued, bounded, and Lipschitzian differential inclusions. Example 6.8 on the difference between weak and strong minimizers is classical going back to Weierstrass [1326]. The simplified version of Example 6.9 on the difference between weak and intermediate minimizers was presented in [915], while the full version of this example as well as of Example 6.10 are taken from Vinter and Woodford

[1293]. The latter paper and Woodford's dissertation [1331] contain also other examples illustrating the difference between these notions of local minima, particularly the difference between intermediate minimizers of *various ranks* for *convex* and *unbounded* differential inclusions in finite dimensions.

6.5.15. Relaxation Stability and Hidden Convexity. The remainder of Subsect. 6.1.2 presents the construction of the *relaxed* Bolza problem for differential inclusions together with the associated definition and discussions on *relaxation stability*. The idea of proper relaxation (or extension, generalization, regularization) plays a remarkable role in modern variational theory. In general terms, it goes back to Hilbert [567] stating in his famous 20th Problem that "*every problem in the calculus of variations has a solution provided that the word solution is suitably understood.*"

It was fully realized in the 1930s, independently by Bogolyubov [121] and by Young [1349, 1350] for one-dimensional problems of the calculus of variations who showed that adequate extensions of variational problems, which automatically ensure the existence of generalized optimal solutions and their approximations by "ordinary curves," could be achieved by a certain *convexification* with respect to *velocities*. In optimal control, this idea was independently developed by Gamkrelidge [495] and by Warga [1313]; in the latter paper the term "relaxation" was first introduced. Another term broadly used now for similar issues is "Young measures." We refer the reader to [3, 4, 25, 31, 50, 75, 212, 213, 231, 232, 235, 237, 246, 255, 308, 362, 401, 432, 450, 497, 527, 617, 618, 682, 704, 821, 823, 863, 886, 888, 901, 915, 1020, 1049, 1082, 1173, 1174, 1176, 1177, 1258, 1259, 1277, 1315, 1323, 1351] and the bibliographies therein for various relaxation results and their applications to problems of the calculus of variations, optimal control, and related topics.

In this book we follow the constructions developed in [915] for the Bolza problem involving finite-dimensional differential inclusions and employ the relaxation procedure *not* to ensure the existence of generalized solutions but to describe *limiting points* of optimal solutions to discrete approximation problems together with the minimizing functional values. To proceed in this way, the notion of *relaxation stability* formulated in (6.19) plays a crucial role. This property is typically *inherent* in *continuous-time* control systems and differential inclusions relating to their *hidden convexity*; see more discussions and sufficient conditions for relaxation stability presented in Subsect. 6.1.2 and the references therein. We specifically note the approximation property of Theorem 6.11 taken from the recent paper by De Blasi, Pianigiani and Tolstonogov [308], which is a manifestation of the hidden convexity in the framework of the general Bolza problem for infinite-dimensional differential inclusions. Observe also that, in a deep sense, the hidden convexity may be traced to the classical Lyapunov theorem on the range convexity of *nonatomic* vector measures [822] and to its Aumann's version [55] on set-valued integration; see Arkin and Levin [25] and Diestel and Uhl [334] for infinite-dimensional counterparts of such results. We also refer the reader to some other remarkable manifestations of the hidden convexity:

—–Estimates of the "duality gap" in nonconvex programming discovered by Ekeland [398] and then developed by Aubin and Ekeland [51]. These developments are strongly related to the classical Shapley-Folkman theorem in mathematical economics; see the book by Ekeland and Temam [401] for more details and discussions.

—–Convexity of the "nonlinear image of a small ball" recently discovered by Polyak [1098, 1100] who obtained various applications of this phenomenon to optimization, control, and related areas; see also Bobylev, Emel'yanov and Korovin [120] for further developments.

6.5.16. Convergence of Discrete Approximations. While the main attention in Subsect. 6.1.1 was paid to finite-difference approximations of differential/evolution inclusions with *no* optimization involved, the results of Subsect. 6.1.3 concern approximation issues for the *whole* variational problem of Bolza under consideration. This means that we aim to construct well-posed discrete approximations of the original Bolza problem (P) by sequences of discrete-time dynamic optimization problems in such a way that optimal solutions for discrete approximations converge, in a certain prescribed sense, to those for the continuous-time problem. In fact, we present *well-posedness/stability* results that justify the convergence of discrete approximations of the following *two types*:

(I) *Value convergence* ensuring the convergence of *optimal values* of the cost functionals in *constructively* built discrete approximation problems to the optimal value (infimum) of the cost functional in the original problem for which the existence of optimal solutions is not assumed.

(II) *Strong convergence* of optimal solutions for discrete-time problems to the *given* optimal solution for the original problem; the strong convergence is understood in the $W^{1,2}$-norm for piecewise linearly extended discrete trajectories.

Observe that the results of type (II) *explicitly* involve the given optimal solution (actually an *intermediate minimizer*) to the original problem. They are not constructive any more (from the numerical viewpoint) while justifying the way to derive necessary optimality conditions for continuous-time problems by using their discrete approximations (instead of, say, the method of variations, which is not applicable in this framework). The convergence results of type (II) obtained in Subsect. 6.1.3 are of the main interest for deriving necessary optimality conditions in Sect. 6.1 of this book (cf. also Sect. 7.1 for their counterparts concerning functional-differential control systems); they generally impose *milder* assumptions in comparison with those needed to prove the value convergence in (I).

Results of type (I) traditionally relate to *computational* methods in optimal control; they justify "direct" numerical techniques based on approximations of continuous-time control problems by sequences of finite-difference ones, which reduce to problems of *mathematical programming* in finite dimensions provided that state vectors in control systems are finite-dimensional. We are not familiar with any results in this directions for infinite-dimensional differential inclusions, even in the parameterized control form (6.106), besides those presented in Subsect. 6.1.3.

First results on value convergence for standard control systems (6.106) were probably obtained by Budak, Berkovich and Solovieva [184] and Cullum [302] in the late 1960s under rather restrictive assumptions; see also [185, 303, 407] for earlier developments. Then Mordukhovich [890] established the *equivalence* between the *value convergence* of discrete approximations and the *relaxation stability* for general control problems involving parameterized systems (6.106) provided *appropriate perturbations* of state/endpoint constraints *consistent* with the stepsize of discretization. These results were extended in [899, 901, 902] to Lipschitzian differential inclusions; cf. also related results in Dontchev [349] and Dontchev and Zolezzi [367]. Efficient estimates of *convergence rates*, not only with respect to cost functions but also with respect to controls and trajectories, were derived for systems of special structures by Hager [535], Malanowski [831], Dontchev [347], Dontchev and Hager [355], Veliov [1284], and others; see the surveys in [352, 359, 761] for more details and references.

Theorem 6.14 seems to be new even for finite-dimensional differential inclusions developing the corresponding methods and results from Mordukhovich [890, 899, 901]. Observe that the proof of this theorem and the related Theorem 6.13 are more technically involved in comparison with the finite-dimensional case based, besides other things, on the fundamental Dunford theorem ensuring the sequential weak compactness in $L^1([a,b];X)$ provided that both spaces X and X^* satisfy the Radon-Nikodým property, which is the case when *both* X and X^* are *Asplund*. As we remember, the Asplund structure plays a crucial role in the generalized differentiation theory developed in this book from the viewpoint *not related* to the RNP!

Theorem 6.13, which is what we actually need to implement the method of discrete approximations as a *vehicle* for deriving *necessary optimality conditions* for continuous-time systems (i.e., for "*theoretical*" vs. numerical applications) is an infinite-dimensional extension and a modification of Theorem 3.3 from Mordukhovich [915]. The difference between these two results (even in finite dimensions) concerns the way of approximating the original integral functional: we now adopt construction (6.20) instead of the simplified one (6.28) as in [915]. This modification allows us to deal with *measurable* integrands with respect to t that is important for applications in Sect. 6.2, where the integrand *must* be measurable.

Observe the importance of the last term in (6.20) and (6.28) approximating the derivative of the given intermediate minimizer $\bar{x}(\cdot)$. The presence of this term and the usage of the approximation result from Theorem 6.4 allow us to establish the *strong* (in the norm of $W^{1,2}([a,b];X)$) convergence of optimal solutions for the discrete approximation problems to the *given* local minimizer for the original one, which further leads to deriving necessary conditions of type (6.123) for continuous-time problems by passing to the limit from those for their discrete-time counterparts. Besides [915], this approximating term was previously used by Smirnov [1215] (see also his book [1216]) for the Mayer problem involving convex-valued, bounded, and autonomous differential inclusions in finite dimensions. The previous attempts to employ discrete approximations for deriving necessary optimality conditions in the Mayer framework of convex-valued or even convex-graph differential inclusions were able to ensure merely the uniform convergence of extended discrete trajectories to $\bar{x}(\cdot)$ by using an approximating term of the "state type"

$$\sum_{j=0}^{N-1} \|x_N(t_j) - \bar{x}(t_j)\|^2$$

with no derivative $\dot{\bar{x}}(\cdot)$ involved; cf. Halkin [542], Pshenichnyi [1107, 1109], and Mordukhovich [892, 901, 902].

6.5.17. Necessary Optimality Conditions for Discrete Approximations. After establishing the required *strong convergence/stability* of discrete approximations discussed above, the *second step* in realizing the strategy of this method to establish necessary optimality conditions for constrained differential inclusions is to derive *necessary conditions* for *discrete-time problems* formulated in Subsect. 6.1.3. We consider two forms of the discrete approximation problems:

—the "integral" form (P_N) involving the minimization of the cost functional (6.20) subject to the constraints (6.3), (6.21)–(6.23), and

—the "simplified" form $(\overline{P}_N)$ in which the other cost functional (6.28) is minimized under the same constraints.

As discussed, the only distinction between the two functionals (6.20) and (6.28) relates to different ways of approximating the integral functional in the original continuous-time Bolza problem (P): the integral type of (6.20) allows us to consider *measurable* integrands $\vartheta(x,v,\cdot)$ in (6.13), while the summation/simplified type of (6.28) requires the *a.e. continuity* assumption imposing on $\vartheta(x,v,\cdot)$. The reason to consider the latter simplified approximation is that the *summation form* in (6.28) makes it possible to obtain necessary optimality conditions for discrete-time and then for continuous-time problems in more general settings of *Asplund state spaces* X in comparison with the *reflexivity and separability* requirements needed in the case of the integral approximation as in (6.20). This is due to the more developed *subdifferential calculus* for *finite sums* vs. that for *integral functionals*; see below.

In Subsect. 6.1.4 we derived necessary optimality conditions for discrete-time dynamic optimization problems (P_N) and $(\overline{P}_N)$ as well as for their less structured counterpart (DP) called the *Bolza problem for discrete-time inclusions* in infinite dimensions. These problems are certainly of independent interest for discrete systems with fixed steps being important for many applications, particularly to models of economic dynamics; see, e.g., Dyukalov [379] and Dzalilov, Ivanov and Rubinov [380]. Furthermore, necessary optimality conditions for them provide, due to the convergence results of Subsect. 6.1.3, *suboptimality* conditions for the continuous-time Bolza problem under consideration. However, our main interest is to derive such necessary optimality conditions for (P_N) and $(\overline{P}_N)$, which are more convenient for *passing to the limit* in order to establish necessary optimality conditions for the Bolza problem involving infinite-dimensional differential inclusions.

The discrete-time dynamic optimization problems under consideration in Subsect. 6.1.4 can be reduced to the form of *constrained mathematical programming* (MP) given in (6.29). Problems (MP) appeared in this way have *two characteristic features* that distinguish them from other classes of constrained problems in mathematical programming:

(a) They involve *finitely many geometric constraints* the number of which tends to *infinity* when the stepsize of discrete approximations is decreasing to zero. It is worth mentioning that these geometric constraints are of the *graphical* type, which are generated by the discretized inclusions. The presence of such constraints makes the (MP) problem (6.29) *intrinsically nonsmooth* even for smooth functional data in (6.29) and in the generating problems (P_N), $(\overline{P}_N)$, and (P).

(b) If the original state space X is *infinite-dimensional*, the (MP) problem (6.29) unavoidably contains *operator constraints* of the equality type $f(x) = 0$, where the range space for f *cannot* be finite-dimensional. We know that such constraints are among the most difficult in optimization, even for smooth mappings f, which is actually the case for applications to the discrete-time problems under consideration.

The theory of necessary optimality conditions for mathematical programming problems of type (6.29) is available from Chap. 5, where we established necessary conditions in terms of the basic/limiting generalized differential constructions. The main conditions for problems of this type involving extended *Lagrange multipliers* are summarized in Proposition 6.16, where finitely many geometric constraints in (6.29) are incorporated via the *intersection rule* for the basic normal cone and the corresponding *SNC calculus* result in the framework of Asplund spaces. Employing these optimality conditions for (MP) together with *exact/pointwise* calculus rules developed for basic normals and subgradients, we arrive at necessary optimality conditions for the discrete Bolza problem (DP) governed by difference inclusions in the *extended Euler-Lagrange form* of Theorem 6.17. Note that the latter result doesn't impose *any* *convexity* and/or *Lipschitzian* assumptions on the discrete velocity sets $F_j(x)$. The conditions obtained in Theorem 6.17 give an Asplund space version of the finite-dimensional conditions from Mordukhovich [915, Theorem 5.2] under certain SNC requirements needed in infinite dimensions.

The *pointbased* necessary optimality conditions for the discrete Bolza problem (DP) obtained in Theorem 6.17 are important for its own sake and, furthermore, provide a sufficient ground for deriving necessary optimality conditions of the extended Euler-Lagrange type (6.123) for continuous-time problems in finite dimensions; see [915] for more details. However, it is *not* precisely the case in *infinite dimensions*, where the realization of this scheme requires *extra* SNC assumptions ensuring the fulfillment of the pointbased necessary optimality conditions in discrete approximations and then the passage to the limit from them as $N \to \infty$. These extra assumptions can be *avoided* by deriving *approximate/fuzzy* necessary conditions for discrete-time problems, instead of the pointbased ones as in Theorem 6.17. Such approximate optimality conditions are obtained in Theorems 6.19 and 6.20 for the discrete approximation problems $(\overline{P}_N)$ and (P_N), respectively.

The proofs of the afore-mentioned approximate optimality conditions are rather involved requiring, among other things, the usage of *fuzzy* calculus rules as well as *neighborhood* coderivative characterizations of *metric regularity* established by Mordukhovich and Shao [946]. Observe also a significant role of Lemma 6.18 extending to the case of *basic* subgradients the classical *Leibniz rule* on (*sub*)*differentiation under integral sign*. This is an auxiliary result for the proof of Theorem 6.20 allowing us to deal with *summable* integrands in (P) under discrete approximations of type (P_N), while the rule itself is certainly of independent interest. Its proof employs an infinite-dimensional extension of the Lyapunov-Aumann convexity theorem and the corresponding rule for Clarke's subgradients [255, Theorem 2.7.2], which is strongly based in turn on the generalized version of Leibniz's rule established by Ioffe and Levin [612] for subgradients of convex analysis.

6.5.18. Passing to the Limit from Discrete Approximations. In Subsect. 6.1.5 we accomplish the *third step* (labeled as (iii) in Subsect. 6.5.12) in the method of discrete approximations to derive necessary optimality conditions in the original Bolza problem (P) for differential inclusions. The primary goal at this step is to justify the passage to the limit from the obtained necessary conditions in the well-posed discrete approximation problems (P_N) and $(\overline{P}_N)$ and to describe efficiently the resulting necessary optimality conditions for the continuous-time problems that come out of this procedure. As we see, the resulting conditions occur to be those of the *extended Euler-Lagrange* type for *relaxed intermediate local minimizers* in (P) established in Theorems 6.21 and 6.22.

These major results of Subsect. 6.1.5 are somewhat different from each other, in both aspects of the assumptions made and of formulating the extended Euler-Lagrange inclusions in (6.44) and (6.47). The differences came from the corresponding results of Subsect. 6.1.4 for the two types of discrete approximation problems, $(\overline{P}_N)$ and (P_N), as well as from additional requirements needed for passing to the limit in the necessary optimality conditions for these problems.

Theorem 6.21, based on the limiting procedure from the simplified discrete approximations $(\overline{P}_N)$, is an infinite-dimensional generalization of that in Mordukhovich [915, Theorem 6.1] with involving the *extended normal cone* in (6.44). The usage of the basic normal cone in a similar setting of [915] was supported by certain technical hypotheses ensuring the *normal semicontinuity* formulated in Definition 5.69 and discussed after it. Theorem 6.22 is new even in finite dimensions.

One of the main concerns in passing to the limit from the discrete-time necessary optimality conditions in the proofs of both Theorem 6.21 and Theorem 6.22 is to justify appropriate convergences of *adjoint trajectories* and their *derivatives*. To establish the required convergence, we employ a dual *coderivative characterization* of *Lipschitzian behavior* for set-valued mappings used so often in this book; such criteria play a *crucial role* in accomplishing limiting procedures for adjoint systems associated with discrete-time and continuous-time inclusions in dynamic optimization problems described by Lipschitzian mappings.

The principal issue that distinguishes the necessary optimality conditions obtained for *infinite-dimensional* differential inclusions from their finite-dimensional counterparts is the presence of the SNC (actually *strong PSNC*) assumption on the constraint/target set Ω imposed in Theorems 6.21 and 6.22. Assumptions of this type are crucial for optimal control problems for infinite-dimensional evolution systems. In particular, it is well known that *no* analog of the Pontryagin maximum principle holds even for simple optimal control problems governed by the one-dimensional heat equation with a *singleton* target set $\Omega = \{x_1\}$ in Hilbert spaces, which is *never PSNC* in infinite dimensions. The first example of this type was given by Y. Egorov [393]. The reader can also consult with the books by Fattorini [432] and by Li and Yong [789] for more discussions involving the *finite codimension* property equivalent to the SNC one for convex sets; see Remark 6.25. Let us emphasize to this end the result of Corollary 6.24 justifying the extended Euler-Lagrange conditions for the Bolza problem (P) governed by evolution inclusions with *no explicit* (while *hidden*) *SNC/PSNC* assumptions on the constraint set Ω given by *finitely many* equalities and inequalities via Lipschitzian functions.

Lastly, we refer the reader to the recent papers by Mordukhovich and D. Wang [970, 971], where some counterparts of the above results are derived for optimal control problems governed by *semilinear unbounded evolution inclusions* that are particularly convenient for modeling *parabolic PDEs*; see Remark 6.26.

6.5.19. Euler-Lagrange and Maximum Conditions with No Relaxation. As seen, the extended Euler-Lagrange conditions established in

Sect. 6.1 by the method of discrete approximations apply to *relaxed* intermediate local minimizers for the Bolza problem governed by infinite-dimensional differential inclusions. The primary goal of Sect. 6.2 is to derive, based on the conditions obtained in Sect. 6.1 and involving additional variational techniques, refined results of the Euler-Lagrange type accompanied furthermore by the Weierstrass-Pontryagin maximum condition for *nonconvex* differential inclusions *without any relaxation*. The main result, for simplicity formulated in Theorem 6.27 in the case of the Mayer-type problem (P_M) with a fixed left endpoint and arbitrary geometric constraints imposed on right endpoints of trajectories, is new in infinite dimensions; its preceding finite-dimensional versions were discussed in Subsect. 6.5.8.

As in Sect. 6.1, the principal distinction between necessary conditions obtained in finite-dimensional and infinite-dimensional settings relates to the presence of *SNC requirements* unavoidable in infinite dimensions. On the other hand, the technical assumptions made in Theorem 6.27 are *different* from those imposed in Theorems 6.21 and 6.22. Observe also the more general forms (6.51) and (6.52) of the transversality conditions in Theorem 6.27 in comparison with the major results of Sect. 6.1 involving only Lipschitzian cost and constraint functions.

The proof of the pivoting Euler-Lagrange condition (6.49) for intermediate local minimizers to nonconvex problems with *no relaxation* is based, besides applying rather delicate calculus and convergence results of variational analysis, on *two perturbation/approximation* procedures allowing us to reduce the original problem (P_M) to the *unconstrained* (while nonsmooth and nonconvex) Bolza problem (6.55) with finite-valued data that are *Lipschitzian* in the state and velocity variables and *measurable* in t. Since any intermediate local minimizer for the latter problem is automatically a *relaxed* one, it can be treated by the necessary optimality conditions obtained in Theorem 6.22 via discrete approximations.

The first of the afore-mentioned perturbation techniques can be recognized as the *method of metric approximations* originally developed by Mordukhovich [887] to prove the maximum principle for finite-dimensional control problems with smooth dynamics and nonsmooth endpoint constraints by reducing them to free-endpoint problems. The second perturbation technique, involving the *Ekeland variational principle* and *penalization* of dynamic constraints, goes back to Clarke [251] in connections with his results on Hamiltonian and maximum conditions for nonsmooth control systems in finite dimensions. The *claim* in the proof of Theorem 6.27 is an infinite-dimensional extension of the corresponding result by Kaśkosz and Lojasiewicz [667] established there for *strong* minimizers (or *boundary* trajectories). Note the importance of the generalized differential results from Subsect. 1.3.3 for the *distance function* at *in-set* and *out-of-set* points to deal with approximating problems and also a crucial role of the *coderivative criterion* for Lipschitzian behavior that allows us to accomplish the convergence procedure in deriving the extended Euler-Lagrange and transversality inclusions of Theorem 6.27.

The proof of the maximum condition (6.50) supplementing the extended Euler-Lagrange condition (6.49) in the nonconvex case is outlined but not fully presented in Subsect. 6.2.1, since it is technically involved while closely follows the line developed by Vinter and Zheng [1294] (see also Vinter's book [1289, Theorem 7.4.1]) for finite-dimensional differential inclusions; the reader can check all the details. Note that this proof is based on reducing the general Mayer problem for differential inclusions to an optimal control problem with *smooth dynamics* and *nonsmooth endpoint constraints* first treated by Mordukhovich [887] via his nonconvex/limiting normal cone; see Sect. 6.3 for related control problems and techniques in infinite-dimensional settings. It seems that the other available proofs of the maximum condition (6.50) in the Euler-Lagrange framework (6.49) given by Ioffe [598] and by Clarke [261] are restricted to the case of finite-dimensional state spaces.

6.5.20. Related Topics and Results in Optimal Control of Differential Inclusions. The variational methods developed in this book allow us to obtain extensions and counterparts of Theorem 6.27 in various settings partly discussed in Subsect. 6.2.2, which particularly include *upper subdifferential* conditions and *multiobjective* control problems; cf. also Zhu [1372], Bellaassali and Jourani [93], and Eisenhart [395] for related developments in multiobjective dynamic optimization concerning finite-dimensional control systems. It seems however that necessary optimality conditions of the *Hamiltonian* type as well as results on *local controllability* for differential inclusions require the *finite dimensionality* of state spaces; see more details and discussions in Remarks 6.32 and 6.33.

The examples given at the end of Subsect. 6.2.2 illustrate some characteristic features of the results obtained for differential inclusions and the relationships between them. Example 6.34 confirming that the *partial* convexification is *essential* for the validity of both Euler-Lagrange and Hamiltonian optimality conditions of the established extended type is due to Shvartsman (personal communication). Example 6.35 taken from Loewen and Rockafellar [805] shows that the *extended Euler-Lagrange* condition involving only the partial convexification is *strictly better* than the *Hamiltonian condition* in Clarke's fully convexified form even for Lipschitzian control systems with convex velocities. Finally, Example 6.36 given by Ioffe [604] demonstrates that the *partially convexified* Hamiltonian condition, which may *not* be equivalent to its Euler-Lagrange counterpart, also *strictly improves* the *fully convexified* Hamiltonian formalism in rather general settings.

6.5.21. Primal-Space Approach via the Increment Method. Section 6.3 concerns optimal control problems in the more traditional *parameterized* framework (6.61), involving however the *infinite-dimensional dynamics.* Even more, we impose in this section the *continuous differentiability/smoothness* assumption on the velocity function f with respect to the state variable x. Nevertheless, the results presented in Sect. 6.3 are different from those obtained in Sects. 6.1 and 6.2 for dynamic optimization problems governed by nonsmooth evolution inclusions at least in the following major aspects:

—–there are *no* additional geometric assumptions of the state space in question, which is an *arbitrary Banach* space;

—–the objective and (equality and inequality) endpoint constraint functions may *not* be *locally Lipschitzian*, even *not continuous* around the reference point in the case of those functions describing the objective and inequality constraints, while the resulting necessary optimality conditions are obtained in the *conventional PMP form*, whenever the functions are Fréchet differentiable *at* the point in question, and in its *upper subdifferential* extension for special classes of nonsmooth functions.

In contrast to the approximation/perturbation methods employed in Sects. 6.1 and 6.2, we now rely on the more conventional *primal-space* approach that goes back to the classical proof of the Pontryagin maximum principle [124, 1102] with subsequent significant developments in the route paved by Rozonoér [1180] for finite-dimensional control systems. There are *two major ingredients* of the employed primal-space techniques, the traces of which could be found in McShane's paper [860] on the calculus of variations: the usage of *needle variations* and the employment of *convex separation.* Both of these ingredients were crucial in the original proof of the maximum principle [124, 1102], while their clarifications and important modifications came later starting—in different directions—with the papers by Rozonoér [1180] and Dubovitskii and Milyutin [369, 370]; see also other references and discussions in Subsects. 1.4.1 and 6.5.1.

In the proof of the maximum principle formulated in Theorem 6.37 we mainly follow the line initiated in the three-part paper by Rozonoér [1180], who was probably the first to fully recognize a major variational role of the *free-endpoint* "terminal control" (i.e., Mayer) problem in the maximum principle and to develop the so-called *increment method* in proving the PMP for problems of this type employing needle variations. Endpoint constraints were then treated as in finite-dimensional nonlinear programming by using *convex separation* techniques related to the so-called *image space analysis*; cf. Plotnikov [1083], Gabasov and Kirillova [485], and the recent book by Giannessi [504]. A delicate derivation of the transversality conditions for control problems with *equality* endpoint constraints given by merely differentiable functions was developed by Halkin [545] based on the Brouwer fixed-point theorem.

The *upper subdifferential* conditions of the PMP obtained in Theorem 6.38 seems to be new even for finite-dimensional control systems. The closest conditions were derived in the recent book by Cannarsa and Sinestrari [217, Theorem 7.3.1] for free-endpoint control problems in finite dimensions under more restrictive assumptions, while somewhat related results were established by

Mordukhovich and Shvartsman [955, 956] for discrete-time systems and discrete approximations; see Section 6.4. Note that Fréchet upper subgradients (or "supergradients") of the *value function* were used in optimal control for *synthesis* problems via Hamilton-Jacobi equations; see, e.g., Subbotina [1231], Zhou [1366], Cannarsa and Frankowska [216], Cannarsa and Sinestrari [217], Frankowska [472], and their references.

6.5.22. Multineedle Variations and Convex Separation in Image Spaces. In the proof of Theorem 6.37 given in Subsects. 6.3.2–6.3.4 we mainly develop the scheme implemented by Gabasov and Kirillova [485] for finite-dimensional control systems under substantially more restrictive assumptions. As mentioned, the basic idea of the proof for the *free-endpoint* problem in Subsect. 6.3.2 goes back to Rozonoér [1180], while needle variations of *measurable* controls via the increment formula are treated as in Mordukhovich [887, 901]. The reader can find more recent developments on needle variations including their usage for higher-order necessary optimality conditions in the publications by Agrachev and Sachkov [2], Bianchini and Kawski [114], Krener [703], Ledzewicz and Schättler [756], Sussmann [1236, 1238], and in the references therein.

The proof of Theorem 6.37 in the presence of *endpoint constraints* is significantly more involved in comparison with that for the free-endpoint problem. Now it requires taking into account the *geometry* of *reachable sets* for dynamic control systems. The usage of *multineedle variations* occurs to be crucial in the constraint framework. It allows us to construct a *convex tangential* approximation of the reachable set in the *image space*, the dimension of which is equal to the number of endpoint constraints plus one of the cost function. Thus, although the control problem under consideration involves the *infinite-dimensional* dynamics/state space, the proof of the maximum principle relies on the *finite-dimensional convex separation.*

Observe that *no* SNC-type property is involved in Sect. 6.3 to obtain the required *pointbased* results as in the general settings of Sects. 6.1 and 6.2. In fact, the latter is in accordance with the results obtained in the preceding sections, where we observed that the SNC property of the constraint/target set was actually *automatic* in the case of *finitely* many endpoint constraints. This phenomenon relates to the *finite codimension* property of the constraint set, which readily *yields* the sequential normal compactness *unavoidable* in infinite dimensions. Note also that, as one can see from the proofs in Subsects. 6.3.3 and 6.3.4, the *convexity* of the underlying approximation set in the *image space* was reached due to the *continuity* of the time interval; this is yet another manifestation of the *hidden convexity* inherent in continuous-time control systems.

6.5.23. The Discrete Maximum Principle. Section 6.4 again concerns optimal control problems with *discrete time* as well as *discrete approximations* of continuous-time systems. However, now our agenda is completely different from that in Sect. 6.1, where discrete approximations were mostly used as the *driving force* to derive necessary optimality conditions for differential inclusions, although the results obtained therein for *discrete inclusions* are certainly of independent interest. Recall that in Subsect. 6.1.4 we established necessary optimality conditions of the *Euler-Lagrange type* for general (nonconvex and non-Lipschitzian) discrete inclusions by reducing them to nonsmooth mathematical programming with many geometric constraints. When the "discrete velocity" sets $F_j(x)$ are *convex*, the results obtained automatically imply the *maximum-type* conditions by the extremal property of coderivatives for convex-valued mappings from Theorem 1.34, which is actually due to the extremal form of the normal cone to convex sets. It is clear from the general viewpoint of nonsmooth analysis that a certain *convexity* is undoubtedly *needed* for such extremal-type representations. On the other hand, the Pontryagin maximum principle and its nonsmooth extensions hold for *continuous-time* control systems with *no explicit convexity assumptions.* As seen from the results and discussions of Sects. 6.1–6.3, this is due to the *hidden convexity* strongly inherent in the continuous-time dynamics.

Considering optimal control problems for discrete systems with *fixed* stepsizes, we don't have grounds to expect such maximum-type results in the absence of some convexity. Nevertheless, the exact analog of the Pontryagin maximum principle for discrete control problems was first obtained by Rozonoér [1180], under the name of the *discrete maximum principle*, for minimizing a linear function of the right endpoint $x(K)$ without any constraints on $x(K)$ over the discrete-time system

$$\begin{cases} x(t+1) = Ax(t) + b\big(u(t),t\big), \quad x(0) = x_0 \ , \\ \\ u(t) \in U, \quad t = 0, \ldots, K-1 \ , \end{cases} \tag{6.130}$$

with *no* convexity assumptions imposed. The proof of this result was based on the increment formula over needle variations of the reference optimal control at one point $t = \theta$, similarly to the continuous-time case but without involving of course a (nonexistent) interval of "small length." The latter result and its proof given by Rozonoér heavily depended on the specific structure of system (6.130) while probably creating a false impression that the discrete maximum principle might be valid for general nonlinear systems, at least for free-endpoint problems. Note that doubts about such a possibility were clearly expressed in [1180].

A number of papers, mostly in the Western literature, and the book by Fan and Wang [426] were published with incorrect proofs "justifying" that of the discrete maximum principle was necessary for optimality. The first explicit (rather involved) example on violating the discrete maximum principle was given by Butkovsky [208]. Many other examples in this direction, much simpler than the one from [208], can be found in the book by Gabasov and Kirillova [486]; see also the references therein.

Example 6.46 is taken from Mordukhovich [901]. Note that it describes a class of discrete control systems, where the *global minimum* (instead of maximum) condition holds under certain relationships between the initial data. Other examples from [901] show that the discrete maximum principle can be violated even for systems of type (6.130), *linear* in *both* state and control variables, with a nonlinear minimizing function and a nonconvex control set U. In this way we get *counterexamples* to the *conjecture* by Gabasov and Kirillova [486, Commentary to Chap. 5] (repeated later by several authors) on the relationship between the *validity* of the *discrete maximum principle* in discrete-time systems with sufficiently small stepsizes and the *existence of optimal solutions* for continuous-time systems. More striking counterexamples in this direction, showing that the existence of optimal controls in continuous-time systems doesn't imply the fulfillment of even an *approximate* analog of the maximum principle for discrete approximations, are given in Subsects. 6.4.3 and 6.4.4.

The first correct result on the validity of the discrete maximum principle for nonlinear control systems of the type

$$\begin{cases} x(t+1) = f\big(x(t), u(t), t\big), \quad x(0) = x_0 \ , \\ u(t) \in U, \quad t = 0, \ldots, K-1 \ , \end{cases} \tag{6.131}$$

was probably due to Halkin [540] who established it under the *convexity* of the admissible "velocity sets" $f(x, U, t)$; see also the books by Cannon, Cullum and Polak [218], Boltyanskii [127], and Propoi [1105] for further results and discussions in this direction. On the other hand, Gabasov and Kirillova [486] and Mordukhovich [901] singled out special classes of nonlinear free-endpoint control problems for which the discrete maximum principle holds with *no* convexity assumptions. Furthermore, Mordukhovich's book [901] contains the so-called *individual conditions* for the fulfillment of the discrete maximum principle that allow us to describe relationships between the *initial data* of nonconvex systems ensuring either validity or violation of the discrete maximum principle. In particular, these conditions *comprehensively* treat the situation in Example 6.46: the discrete maximum principle holds therein *if and only if* $\gamma \leq 0$ and $\eta \geq 0$.

6.5.24. Necessary Conditions for Free-Endpoint Discrete Parametric Systems. The previous discussions clearly illustrate the *gap* between the Pontryagin maximum principle for continuous-time systems and its discrete-time counterpart in the classical framework of optimal control, even for free-endpoint problems. Besides the striking theoretical value of this phenomenon, it may have a serious *numerical* impact signifying a possible *instability* of the PMP under computing, which inevitably requires the time discretization. Observe however that computer calculations deal not with fixed-step discrete systems of type (6.131) but with parametric *discrete approximation* systems of the type

$$x(t+h) = x(t) + hf\big(x(t), u(t), t\big) \quad \text{as} \quad h \downarrow 0 \, , \tag{6.132}$$

where the stepsize h is a discretization parameter. Thus it is natural to consider necessary optimality conditions for control problems involving parametric systems (6.132) that themselves *depend on the parameter* h.

The first result in this direction was obtained by Gabasov and Kirillova [484, 486] who derived, under the name of "quasimaximum principle," necessary optimality conditions for *free-endpoint parametric* control problems governed by general discrete-time systems of the type

$$x(t+1) = f\big(x(t), u(t), t, h\big), \quad x \in I\!R^n, \quad h \in I\!R^m \, ,$$

imposing rather standard smoothness while *no convexity* assumptions. Their result asserts, for any given $\varepsilon > 0$, the fulfillment of a certain ε-*maximum* condition over a part of the control region that depends on ε and h. Being specified to the discrete approximation systems (6.132), the ε-maximum condition is as close to the one in the Pontryagin maximum principle as smaller ε and h are. Similar results were subsequently derived for discrete approximations of nonconvex free-endpoint control problems in the books by Moiseev [884, 885] and by Ermoliev, Gulenko and Tzarenko [407]; see the aforementioned books and also those by Propoi [1105] and Evtushenko [412] for various discussions and applications of such results to numerical methods in optimal control for continuous-time and discrete-time systems.

The proof of the quasimaximum principle and the related results for free-endpoint problems of discrete approximation given in [484, 486, 884, 885, 407] were similar to each other being, in fact, similar to Rozonoér's proof of the PMP for continuous-time systems with no constraints on trajectories; compare, e.g., the proof of Theorem 6.37 in the unconstrained case of Subsect. 6.3.2 with the one for Theorem 6.50 in the smooth unconstrained case of Subsect. 6.4.3. All these proofs strongly exploited the *unconstrained* nature of the control problems under consideration involving cost increment formulas on *single needle variations* of optimal controls. The only difference between the continuous-time and finite-difference cases concerned the usage of a small discretization stepsize in the parametric family of discrete-time problems *instead* of a small length of needle variations in continuous-time systems. These proofs didn't provide any hint of the possibility to obtain an appropriate counterpart of the PMP for discrete approximations of optimal control problems with endpoint constraints, where some *finite-difference* counterpart of the *hidden convexity* and the geometry of reachable sets must play a crucial role.

6.5.25. The Approximate Maximum Principle for Constrained Discrete Approximations. Necessary optimality conditions in the form of the *approximate maximum principle* (AMP) for optimal control problems of discrete approximation (6.132) with *smooth dynamics* and *smooth endpoint constraints* were first announced by Mordukhovich in [891] and then were developed in the subsequent publications [942, 899, 900, 901, 903]. The final version for smooth control problems presented in Theorem 6.59 was established in [901, 903]; see also [906]. The proof of this major theorem given in Subsect. 6.4.5 goes along the *primal-space* direction, being however significantly different in crucial aspects from its continuous-time counterpart considered in Subsects. 6.3.3 and 6.3.4. There are *three key assumptions* under which we justify the AMP in Theorem 6.59:

——the *consistence* of perturbations of the *equality* constraints;

——the *properness* of the sequence of optimal controls;

——the *smoothness* of the initial data with respect to the state variables.

Each of these assumptions occurs to be *essential* for the validity of the AMP in discrete approximations of *nonconvex constrained* problems as demonstrated by counterexamples of Subsect. 6.4.4.

The crucial role of *consistent perturbations* of endpoint constraints for achieving the *stability* of discrete approximations, from both viewpoints of the *value convergence* and the *validity of the AMP*, has been realized by Mordukhovich since the very beginning of his study; see [890, 891]. Example 6.61 showing that the AMP may be violated if the endpoint equality constraints are not appropriately perturbed (must decrease *slower* than the discretization stepsize) is taken from Mordukhovich and Raketskii [942]; see also [901, 903].

Example 6.60, which is taken from Mordukhovich and Shvartsman [956], demonstrates the significance of the *properness* property along the reference optimal control sequence for the validity of the AMP in constrained nonconvex problems. This property is specific for discrete approximations, although it may be viewed as some analog of the *piecewise continuity*, or generally *Lebesgue points* of measurable controls, that are not of any restriction for continuous-time systems. Note that we *don't need* to impose the properness assumption to ensure the AMP in free-endpoint problems; see Theorem 6.50 and its proof.

6.5.26. Nonsmooth Versions of the Approximate Maximum Principle. One of the most *striking* features of the approximate maximum principle is its *sensitivity to nonsmoothness*. This is probably *the only* result on optimality conditions and related topics of variational analysis we are familiar with that doesn't have any conventional *lower subdifferential* (regarding minimization) extension to nonsmooth (even *convex*) settings. This is demonstrated by examples from the paper of Mordukhovich and Shvartsman [956] presented in Subsect. 6.4.3 for free-endpoint control problems.

On the other hand, the afore-mentioned paper [956] justifies a new form of the approximate maximum principle involving *upper subdifferential transversality conditions* for *free-endpoint* problems with nonsmooth cost functions (Theorem 6.50) and for constrained problems whose *inequality-type* endpoint constraints are described by nonsmooth functions (Theorem 6.66). The results obtained in this direction apply to a special class of nonsmooth functions called *uniformly upper subdifferentiable* in [956]. This class contains, besides smooth and concave functions, also *semiconcave* functions (see Subsect. 5.5.4) being actually closely connected with a localized version of "weakly concave" functions in the sense of Nurminskii [1017] who efficiently used them in numerical optimization. Theorem 6.49 seems to be new in reflexive spaces; some of its conclusions and related properties were established in [956, 1017] with different proofs in finite dimensions.

Theorem 6.50 on the AMP for free-endpoint problems gives an *infinite-dimensional* extension of the upper subdifferential result from Mordukhovich and Shvartsman [956], which smooth version [901] is actually equivalent to the "quasimaximum principle" by Gabasov and Kirillova [484, 486] established under somewhat more restrictive assumptions.

Observe that the *free-endpoint version* of the AMP in Theorem 6.50 doesn't fully follow from the constrained versions of Subsect. 6.4.4 in both smooth and nonsmooth settings. Besides the infinite dimensionality and the *absence* of the properness property for free-endpoint problems, there are *error estimates* of the rate $\varepsilon(t,h_N)=O(h_N)$ for the maximum condition (6.85) in Corollaries 6.52 and 6.53 valid for smooth and concave cost functions in arbitrary Banach spaces.

6.5.27. Applications of the Approximate Maximum Principle. At the end of Subsect. 6.4.5 we present two *typical applications* of the approximate maximum principle. The first one, described in Remark 6.67, follows the route from the paper by Gabasov, Kirillova and Mordukhovich [488] to derive *suboptimality* conditions for continuous-time systems by using the value convergence and necessary optimality conditions for discrete approximations.

Secondly, we consider a more *practical application* of using the approximate maximum principle to solve optimal control problems governed by discrete-time systems with sufficiently *small stepsizes*. Example 6.68 taken from Mordukhovich [901] concerns a (simplified) practical problem of *chemical engineering* described in the book by Fan and Wang [426]. The discrete maximum principle cannot be applied to find optimal solutions to this constrained nonconvex problem, although the authors of [426] mistakenly did it throughout their book and related papers. On the other hand, the application of the approximate maximum principle justified in Theorem 6.59 allows us to find optimal controls.

Other applications of the AMP for constrained discrete approximation problems were developed by Nitka-Styczen [1013, 1014, 1014] who considered the framework of *optimal periodic control* involving *equality* endpoint constraints. Based on the AMP machinery, she designed efficient numerical methods of solving such problems and applied them to practical problems arising in optimization of chemical, biotechnological, and ecological processes. Some of the models considered in [1015] are described by hereditary/delay control systems that require certain modifications of the formulation of the AMP given in [1015] and in Subsect. 6.4.6 of this book.

6.5.28. The Approximate Maximum Principle in Systems with Delays. The results presented in Subsect. 6.4.6 are taken from the paper by Mordukhovich and Shvartsman [956], with their direct extension to delay systems in *infinite-dimensional* spaces. Considering for simplicity only free-endpoint problems, we derive the AMP with *upper subdifferential* transversality conditions for nonlinear systems with *time-delays in state* variables. The proof of this result for delay systems is based on their reduction, following the approach by Warga [1315], to ordinary discrete-time systems with possible *incommensurability* between the length of the underlying time interval $b-a$ and the discretization stepsize h_N.

The final Example 6.70 of Subsect. 6.4.6 draws the reader's attention to a very interesting class of hereditary systems, called functional-differential systems of *neutral type*, that are significantly different from ordinary control systems and their extensions systems with delays only in state variables. Such systems, admitting *time-delays in velocity* variables, are considered in more details in Sect. 7.1; see also Commentary to Chap. 7. Example 6.70, which is a finite-difference adaptation of the continuous-time example from the book by Gabasov and Kirillova [485, Section 3.6], shows that there is *no* natural analog of the AMP held for *smooth free-endpoint* control problems governed by *finite-difference systems of neutral type.*

7

Optimal Control of Distributed Systems

In this chapter we continue our study of optimal control problems from the viewpoint of advanced methods of variational analysis and generalized differentiation. In contrast to the preceding chapter where the main attention was paid to control problems governed by *ordinary* differential equations and inclusions as well as their discrete-time counterparts, we now focus on control systems with *distributed parameters* governed by *functional-differential* and *partial differential* relations. We particularly study optimal control problems for *delayed differential-algebraic inclusions* that cover several important classes of control systems essentially different from ordinary ones, and for *partial differential equations* of *hyperbolic* and *parabolic* types that involve *boundary controls* of both Dirichlet and Neumann types as well as *pointwise state constraints*. All the mentioned problems have not been sufficiently studied in the literature; most of the material presented in this chapter is based on recent results developed by the author and his collaborators.

We start this chapter with studying optimal control problems for the so-called *differential-algebraic systems* with time delays, which describe control processes by interconnected delay-differential inclusions and algebraic equations combining some properties of continuous-time and discrete-time control systems. They include, in particular, functional-differential control systems of *neutral type* briefly discussed in Chap. 6. Then we consider boundary control problems for *hyperbolic systems* with pointwise state constraints. Such problems are essentially more difficult than the ones with distributed controls (due to the lack of regularity) and also different from each other depending on the type of boundary conditions (Neumann or Dirichlet). The final section is devoted to *minimax control* problems for *parabolic systems* in uncertainty conditions with Dirichlet boundary controls and pointwise state constraints. Our main results include necessary optimality and suboptimality conditions and related convergence/stability issues for a number of approximation techniques developed in this chapter in the framework of variational analysis.

7.1 Optimization of Differential-Algebraic Inclusions with Delays

This section deals with dynamic optimization problems for differential-algebraic control systems, which belong to the important while not well-developed area in optimal control. Mathematically differential-algebraic systems provide descriptions of control processes via combinations of interconnected differential and algebraic relations. There are many applications of such dynamic models especially in process systems engineering, robotics, mechanical systems with holonomic and nonholonomic constraints, etc.; see the references in Commentary to this chapter. Despite the significance of dynamic optimization problems governed by differential-algebraic systems, not much has been done for variational analysis of such optimal control problems, in particular, for the derivation of necessary optimality and suboptimality conditions. The most advanced previous results are obtained for control processes described by differential-algebraic *equations* under a rather restrictive "index one" assumption on the dynamics, which doesn't hold in many important applications. An interesting feature of differential-algebraic systems is that optimal processes in such systems *don't satisfy* a natural analog of the Pontryagin maximum principle in the *absence of convexity* assumptions on the velocity sets, even for index one problems.

In this section we study differential-algebraic systems that involve differential *inclusions* vs. equations considered in previous developments. On the other hand, our algebraic equations are assumed to be *linear* with *no* imposing the index one assumption. A principal innovation is introducing a *time delay* in both differential and algebraic relations, which happens to be a *regularization factor* allowing us to *separate* the index one and higher terms in the algebraic equation. The main problem of our study is labeled (DA) and defined as follows: minimize

$$\begin{cases} J[x,z] := \varphi\big(x(a),x(b)\big) + \displaystyle\int_a^b \vartheta\big(x(t),x(t-\theta),z(t),\dot z(t),t\big)\,dt \ \text{ subject to} \\ \dot z(t) \in F\big(x(t),x(t-\theta),z(t),t\big) \ \text{ a.e. } \ t\in[a,b]\ , \\ z(t) = x(t) + Ax(t-\theta), \qquad t\in[a,b]\ , \\ x(t) = c(t), \quad t\in[a-\theta,a)\ , \\ \big(x(a),x(b)\big)\in\Omega\ , \end{cases}$$

where $x:[a-\theta,b]\to X$ is continuous on $[a-\theta,a)$ and $[a,b]$ (with a possible jump at $t=a$) and where $z(\cdot)$ is absolutely continuous on $[a,b]$. For simplicity we suppose in this section that $X=I\!R^n$, i.e., the state space is *finite-dimensional*. Based on the methods developed in Sect. 6.1, one can derive extensions of the results obtained below to the case of infinite-dimensional state spaces X under appropriate assumptions parallel to those required in Sect. 6.1 for ordinary evolution inclusions. Note that, even in the case of $X = I\!R^n$ under consideration, problem (DA) is an object of *infinite-dimensional optimization* for functional-differential control systems, which are significantly different from their ordinary counterparts.

When F doesn't depend on z, the dynamic system in (DA) reduces to the functional-differential system of *neutral type*

$$\frac{d}{dt}\Big[x(t) + Ax(t-\theta)\Big] \in F\big(x(t), x(t-\theta), t\big) \quad \text{a.e.} \quad t \in [a,b]$$

written in the so-called *Hale form.* Thus the Bolza problem (DA) formulated above can be viewed as an extended optimal control problem for neutral systems that corresponds to the case of integrands ϑ independent of $(z, \dot z)$. Let us emphasize that dynamic optimization problems for neutral systems (and, more generally, for the differential-algebraic systems under consideration) are essentially more difficult and exhibit new phenomena in comparison with those for ordinary and delay-differential systems when $A = 0$; see below.

In what follows we always assume that $F\colon I\!R^n \times I\!R^n \times I\!R^n \times [a,b] \rightrightarrows I\!R^n$ is a set-valued mapping of closed graph, that Ω is a closed set, that $\theta > 0$ is a constant delay, and that A is a constant $n \times n$ matrix. Note that the methods used in this section allow us to consider the cases of *multiple delays* $\theta_1 \ge \theta_2 \ge \ldots \ge \theta_m > 0$ as well as *variable delays* $\theta(\cdot)$ with $|\dot\theta(t)| < \alpha \in (0,1)$ for a.e. $t \in [a,b]$.

As in Sect. 6.1 for ordinary differential inclusions, our approach to studying problem (DA) is based on the *method of discrete approximations*, which is of undoubted interest from both qualitative/numerical and quantitative aspects of differential-algebraic inclusions. The realization of this method in the case of problem (DA) is different in several aspects (more involved) from the constructions of Sect. 6.1; it particularly exploits the presence of the *nonzero delay* θ. As before, a crucial issue is to establish *variational stability* of discrete approximations that ensures an appropriate *strong convergence* of optimal solutions.

Subsection 7.1.1 is devoted to the construction of *well-posed* discrete approximations of the differential-algebraic *dynamics* in (DA), with *no* taking into account the cost functional and endpoint constraints. The primary goal is to *strongly approximate* any admissible solution $\{x(\cdot), z(\cdot)\}$ to the differential-algebraic inclusion in (DA) by admissible pairs to its finite-difference counterparts. Such a strong approximation allows us to conduct in Subsect. 7.1.2 the convergence analysis of *optimal solutions* for discrete approximations of (DA), with appropriate *perturbations* of endpoint constraints, to the given optimal solution for the original problem. As in the case of ordinary evolution inclusions, the *relaxation stability* plays an essential role in justifying the required strong variational convergence.

In Subsect. 7.1.3 we derive, employing generalized differential tools of variational analysis, necessary optimality conditions for the *difference-algebraic* systems with discrete-time obtained via the well-posed discrete approximation. The assumed finite dimensionality of the state space essentially simplifies the process of deriving these conditions, although the developed SNC calculus and corresponding "fuzzy" results allow us to eventually extend this device to the case of infinite-dimensional state spaces like in Subsect. 6.1.4 for ordinary evolution systems. Finally, Subsect. 7.1.4 presents the main necessary optimality conditions in extended forms of the Euler-Lagrange and Hamiltonian inclusions for differential-algebraic systems (DA) derived by passing to the limit from discrete approximations.

7.1.1 Discrete Approximations of Differential-Algebraic Inclusions

This subsection deals with discrete approximations of an arbitrary admissible pair for the delayed differential-algebraic system in (DA) *without* taking into account the cost functionals and endpoint constraints. We show that, under fairly general assumptions, any admissible pair to the differential-algebraic system can be *strongly approximated* in the sense indicated below by the corresponding admissible pairs to finite-difference inclusions obtained from it by the classical Euler scheme. This result is *constructive* providing *efficient estimates of the approximation rate*, and hence it is certainly of independent interest for numerical analysis of delayed differential-algebraic inclusions.

Let $\{\bar{x}(\cdot), \bar{z}(\cdot)\}$ be an *admissible pair* for the dynamic system in (DA). This means that $\bar{x}(\cdot)$ is continuous on $[a-\theta, a)$ and $[a, b]$ (with a possible jump at $t = a$), $\bar{z}(\cdot)$ is absolutely continuous on $[a, b]$, and the dynamic relations in (DA) are satisfied. Note that the endpoint constraints in (DA) may not hold for $\{\bar{x}(\cdot), \bar{z}(\cdot)\}$; if they do hold, this pair is *feasible* for (DA). The following *standing assumptions* are imposed throughout the whole section:

(D1) There are two open sets $U \subset I\!R^n$, $V \subset I\!R^n$ and two positive numbers ℓ_F, m_F such that $\bar{x}(t) \in U$ for all $t \in [a-\theta, b]$ and $\bar{z}(t) \in V$ for all $t \in [a, b]$, that the sets $F(x, y, z, t)$ are closed, and that one has

$$F(x, y, z, t) \subset m_F I\!B \quad \text{for all} \quad (x, y, z, t) \in U \times U \times V \times [a, b] ,$$

$$F(x_1, y_1, z_1, t) \subset F(x_2, y_2, z_2, t) + \ell_F\big(\|x_1 - x_2\| + \|y_1 - y_2\| + \|z_1 - z_2\|\big) I\!B$$

whenever $(x_1, y_1, z_1), (x_2, y_2, z_2) \in U \times U \times V$ and $t \in [a, b]$.

(D2) $F(x, y, z, t)$ is Hausdorff continuous for a.e. $t \in [a, b]$ uniformly in $(x, y, z) \in U \times U \times V$.

(D3) The function $c(\cdot)$ is continuous on $[a-\theta, a]$.

Similarly to (6.6) in Subsect. 6.1.1, define the *averaged modulus of continuity* $\tau(F; h)$ for F in $t \in [a, b]$ while $(x, y, z) \in U \times U \times V$ in which terms assumption (D2) is equivalent to $\tau(F; h) \to 0$ as $h \to 0$ by Proposition 6.3.

Construct a sequence of discrete approximations of the delayed differential-algebraic inclusion replacing the derivative by the *Euler finite difference*

$$\dot{z}(t) \approx \frac{z(t+h) - z(t)}{h} .$$

For any $N \in I\!N$ we consider the *step of discretization* $h_N := \theta/N$ and define the *discrete mesh* on $[a,b]$ by

$$t_j := a + jh_N \;\text{ as }\; j = -N, \ldots, k \quad \textit{and} \quad t_{k+1} := b ,$$

where k is a natural number determined from $a + kh_N \le b < a + (k+1)h_N$. Then the corresponding *delayed difference-algebraic inclusions* associated with the dynamics in (DA) are described by

$$\begin{cases} z_N(t_{j+1}) \in z_N(t_j) + h_N F\big(x_N(t_j), x_N(t_j - \theta), z_N(t_j), t_j\big), \quad j = 0, \ldots, k , \\ z_N(t_j) = x_N(t_j) + Ax_N(t_j - \theta), \quad j = 0, \ldots, k+1 , \\ x_N(t_j) = c(t_j) \quad j = -N, \ldots, -1 . \end{cases} \tag{7.1}$$

Given a pair $\{x_N(t_j), z_N(t_j)\}$ satisfying (7.1), consider an *extension* of $x_N(t_j)$ to the continuous-time intervals $[a-\theta, b]$ such that $x_N(t)$ are defined piecewise linearly on $[a,b]$ and piecewise constantly, continuously from the right on $[a-\theta, a)$. We also define piecewise constant *extensions of discrete velocities* on $[a,b]$ by

$$v_N(t) := \frac{z_N(t_{j+1}) - z_N(t_j)}{h_N}, \quad t \in [t_j, t_{j+1}),\; j = 0, \ldots, k .$$

Denoting $z_N(t) := x_N(t) + Ax_N(t - \theta)$, one easily has

$$z_N(t) = z_N(a) + \int_a^t v_N(r)\,dr \;\text{ for }\; t \in [a,b] .$$

The following differential-algebraic counterpart of Theorem 6.4 ensures the *strong approximation* of an *arbitrary* admissible solution to the dynamic system in (DA) by extended pairs $\{x_N(t), z_N(t)\}$ satisfying (7.1). The notation $W^{1,2}[a,b]$ stands for the Sobolev space $W^{1,2}\big([a,b]; I\!R^n\big)$.

Theorem 7.1 (strong approximation for differential-algebraic systems). *Let $\{\bar{x}(\cdot), \bar{z}(\cdot)\}$ be an admissible pair to the dynamic system in (DA) under the assumptions* (D1)–(D3). *Then there is a sequence $\{\widehat{x}_N(t_j), \widehat{z}_N(t_j)\}$ of solutions to difference-algebraic inclusions* (7.1) *with*

$$\widehat{x}_N(t_0) = \bar{x}(a) \;\; \textit{for all} \;\; N \in I\!N$$

such that the extensions $\widehat{x}_N(t)$, $a - \theta \le t \le b$, converge uniformly to $\bar{x}(\cdot)$ on $[a-\theta, b]$ while $\widehat{z}_N(t)$, $a \le t \le b$, converge to $\bar{z}(t)$ in the strong $W^{1,2}$ topology on $[a,b]$ as $N \to \infty$.

Proof. Using the density of step-functions in $L^1[a,b] := L^1\big([a,b];I\!R^n\big)$, first select a sequence $\{w_N(\cdot)\}$, $N \in I\!N$, such that each $w_N(t)$ is constant on the interval $[t_j, t_{j+1})$ for $j = 0, \ldots, k$ and that $w_N(\cdot)$ converge to $\dot{\bar{z}}(\cdot)$ as $N \to \infty$ in the norm topology of $L^1[a,b]$. Similarly to the proof of Theorem 6.4 we have estimate (6.7) for the $L^1[a,b]$-norm of $w_N(\cdot)$, which is sufficient to proceed in the proof of this theorem as in the case of ordinary evolution inclusions in Subsect. 6.1.1. For simplicity of the calculations below, suppose that

$$\|w_N(t)\| \le 1 + m_F \quad \text{whenever} \quad t \in [a,b] \quad \text{and} \quad N \in I\!N \; .$$

Define the numerical sequence

$$\xi_N := \int_a^b \|w_N(t) - \dot{\bar{z}}(t)\|\, dt \to 0 \quad \text{as} \quad N \to \infty \; .$$

Denote $w_{N_j} := w_N(t_j)$ and define $\{u_N(t_j), s_N(t_j)\}$ recurrently by

$$\begin{cases} u_N(t_j) := \bar{x}(t_j) \quad \text{for} \quad j = -N, \ldots, 0 \; , \\ s_N(t_j) := u_N(t_j) + A u_N(t_j - \theta) \quad \text{for} \quad j = 0, \ldots, k+1 \; , \\ s_N(t_{j+1}) := s_N(t_j) + h_N w_{N_j} \quad \text{for} \quad j = 0, \ldots, k \; . \end{cases}$$

Then the extended discrete pairs $\{u_N(t), s_N(t)\}$ satisfy

$$\begin{cases} u_N(t) = \bar{x}(t_j) \quad \text{for} \quad t \in [t_j, t_{j+1}), \; j = -N, \ldots, -1 \; , \\ s_N(t) = u_N(t) + A u_N(t - \theta) \qquad \text{for} \quad t \in [a,b] \; , \\ s_N(t) = \bar{z}(a) + \displaystyle\int_a^t w_N(r)\, dr \quad \text{for} \quad t \in [a,b] \; . \end{cases}$$

Next we want to prove that $u_N(\cdot)$ converge uniformly to $\bar{x}(\cdot)$ on $[a,b]$. Denote $y_N(t) := u_N(t) - \bar{x}(t)$ and $\alpha_N(t) := \|y_N(t) + A y_N(t-\theta)\|$. For any $t \in [a,b]$ one has

$$\alpha_N(t) = \|s_N(t) - \bar{z}(t)\| \le \int_a^t \|w_N(r) - \dot{\bar{z}}(r)\|\, dr \le \; \xi_N \; ,$$

which implies the estimates

$$\begin{aligned} \|y_N(t)\| &\le \alpha_N(t) + \|A\| \cdot \|y_N(t-\theta)\| \\ &\le \alpha_N(t) + \|A\|\alpha_N(t-\theta) + \|A\|^2 \cdot \|y_N(t-2\theta)\| \le \ldots \\ &\le \alpha_N(t) + \|A\|\alpha_N(t-\theta) + \ldots + \|A\|^m \alpha_N(t - m\theta) \\ &+ \|A\|^{m+1} \cdot \|y_N(t-(m+1)\theta)\| \; . \end{aligned}$$

Observe that $c(\cdot)$ is *uniformly continuous* on $[a-\theta, a]$ due to assumption (D3). Picking an arbitrary sequence $\beta_N \downarrow 0$ as $N \to \infty$, we therefore have

$$\|c(t') - c(t'')\| \le \beta_N \quad \text{whenever} \quad t', t'' \in [t_j, t_{j+1}], \quad j = -N, \ldots, -1 .$$

Choose an integer number m such that $a - \theta \le b - (m+1)\theta < a$. Then $t - (m+1)\theta \in [t_j, t_{j+1})$ for some $j \in \{-N, \ldots, -1\}$, which implies that

$$\|y_N(t - (m+1)\theta)\| \le \|c(t_j) - c(t - (m+1)\theta)\| \le \beta_N .$$

Since $m \in I\!N$ doesn't depend on N, this gives

$$\|y_N(t)\| \le \xi_N(1 + \|A\| + \ldots + \|A\|^m) + \|A\|^{m+1}\beta_N := \varrho_N \to 0 \quad \text{as} \quad N \to \infty .$$

Now consider a sequence $\{\zeta_N\}$ defined by

$$\zeta_N := h_N \sum_{j=0}^{k} \operatorname{dist}\big(w_{N_j}; F(u_N(t_j), u_N(t_j - \theta), s_N(t_j), t_j)\big)$$

and show that $\zeta_N \downarrow 0$ as $N \to \infty$. By construction of ζ_N and of the averaged modulus of continuity $\tau(F; h)$ we get the following estimates:

$$\begin{aligned}
\zeta_N &= \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \operatorname{dist}\big(w_{N_j}; F(u_N(t_j), u_N(t_j - \theta) ,\\
&\quad s_N(t_j), t_j)\big)\, dt\\
&= \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \operatorname{dist}\big(w_{N_j}; F(u_N(t_j), u_N(t_j - \theta), s_N(t_j), t)\big)\, dt\\
&\quad + \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \Big[\operatorname{dist}\big(w_{N_j}; F(u_N(t_j), u_N(t_j - \theta), s_N(t_j), t_j)\big)\\
&\quad - \operatorname{dist}\big(w_{N_j}; F(u_N(t_j), u_N(t_j - \theta), s_N(t_j), t)\big)\Big]\, dt\\
&\le \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \operatorname{dist}\big(w_{N_j}; F(u_N(t_j), u_N(t_j - \theta), s_N(t_j), t)\big)\, dt + \tau(F; h_N) .
\end{aligned}$$

Further, by (D1) one has for any $t \in [t_j, t_{j+1})$ with $j = 0, \ldots, k$ that

$$\begin{aligned}
&\operatorname{dist}\big(w_{N_j}; F(u_N(t_j), u_N(t_j - \theta), s_N(t_j), t)\big) - \operatorname{dist}\big(w_{N_j}; F(u_N(t),\\
&u_N(t - \theta), s_N(t), t)\big)\\
&\le \operatorname{dist}\big(F(u_N(t_j), u_N(t_j - \theta), s_N(t_j), t); F(u_N(t), u_N(t - \theta), s_N(t), t)\big)\\
&\le \ell_F\big(\|u_N(t_j) - u_N(t)\| + \|u_N(t_j - \theta) - u_N(t - \theta)\| + \|s_N(t_j) - s_N(t)\|\big) .
\end{aligned}$$

Taking into account that

$$\|s_N(t_j) - s_N(t)\| = \left\| \int_{t_j}^{t} w_N(r)\,dr \right\| \le (1+m_F)(t_{j+1}-t_j)$$
$$= (1+m_F)h_N := \eta_N \downarrow 0 \,,$$

we arrive at the estimates

$$\|u_N(t) - u_N(t_j)\| \le \eta_N + \|A\| \cdot \|u_N(t-\theta) - u_N(t_j-\theta)\|$$
$$\le \eta_N\big(1 + \|A\| + \ldots + \|A\|^m\big) + \|A\|^{m+1} \cdot \|u_N(t-(m+1)\theta)$$
$$-u_N(t_j - (m+1)\theta)\|$$
$$\le \eta_N\big(1 + \|A\| + \ldots + \|A\|^m\big) + \|A\|^{m+1}\beta_N := \delta_N \downarrow 0 \ \text{ as } \ N \to \infty$$

and hence ensure that

$$\text{dist}\big(w_{N_j}; F(u_N(t_j), u_N(t_j-\theta), s_N(t_j), t)\big) - \text{dist}\big(w_{N_j}; F(u_N(t),$$
$$u_N(t-\theta), s_N(t), t)\big)$$
$$\le (\eta_N + 2\delta_N)\ell_F \,.$$

It follows from (D1) and the above estimates that for any $\in [t_j, t_{j+1})$ and $j = 0, \ldots, k$ one has

$$\text{dist}\big(w_{N_j}; F(u_N(t), u_N(t-\theta), s_N(t), t)\big) - \text{dist}\big(w_N(t); F(\bar{x}(t), \bar{x}(t-\theta), \bar{z}(t), t)\big)$$
$$\le \text{dist}\big(F(u_N(t), u_N(t-\theta), s_N(t), t); F(\bar{x}(t), \bar{x}(t-\theta), \bar{z}(t), t)\big)$$
$$\le \ell_F\big(\|u_N(t) - \bar{x}(t)\| + \|u_N(t-\theta) - \bar{x}(t-\theta)\| + \|s_N(t) - \bar{z}(t)\|\big)$$
$$\le (2\varrho_N + \xi_N)\ell_F \,.$$

Denoting $\mu_N := \eta_N + 2\delta_N + 2\varrho_N + \xi_N$, we arrive at

$$\text{dist}\big(w_{N_j}; F(u_N(t_j), u_N(t_j-\theta), s_N(t_j), t)\big)$$
$$\le \ell_F\mu_N + \text{dist}\big(w_{N_j}; F(\bar{x}(t), \bar{x}(t-\theta), \bar{z}(t), t)\big) \le \ell_F\mu_N + \|w_{N_j} - \dot{\bar{z}}(t)\|$$

and finally conclude that

$$\begin{aligned} \zeta_N &\le \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \big(\|w_{N_j} - \dot{\bar{z}}(t)\| + \ell_F\mu_N\big)\,dt + \tau(F; h_N) \\ &= \xi_N + \ell_F\mu_N(b-a) + \tau(F; h_N) := \gamma_N \downarrow 0 \ \text{ as } \ N \to \infty \,. \end{aligned} \tag{7.2}$$

Note that the discrete pair $\{u_N(t_j), s_N(t_j)\}$ may *not* be admissible for (7.1). Using the *proximal algorithm*, we construct

$$\begin{cases} \widehat{x}_N(t_j) = c(t_j), \quad j = -N, \ldots, -1, \quad \widehat{x}_N(t_0) = \bar{x}(a) , \\ \widehat{z}_N(t_{j+1}) = \widehat{z}_N(t_j) + h_N v_{N_j}, \quad j = 0, \ldots, k , \\ \widehat{z}_N(t_j) = \widehat{x}_N(t_j) + A\widehat{x}_N(t_j - \theta), \quad j = 0, \ldots, k+1 , \\ v_{N_j} \in F\big(\widehat{x}_N(t_j), \widehat{x}_N(t_j - \theta), \widehat{z}_N(t_j), t_j\big), \quad j = 0, \ldots, k , \\ \|v_{N_j} - w_{N_j}\| = \operatorname{dist}\big(w_{N_j}; F\big(\widehat{x}_N(t_j), \widehat{x}_N(t_j - \theta), \widehat{z}_N(t_j), t_j\big)\big), \quad j = 0, \ldots, k . \end{cases} \tag{7.3}$$

It follows from the construction in (7.3) that $\{\widehat{x}_N(t_j), \widehat{z}_N(t_j)\}$ is a feasible pair to the discrete inclusion (7.1) for each $N \in I\!N$. Note that

$$\|\widehat{x}_N(t) - \bar{x}(t)\| = \|\widehat{x}_N(t_j) - \bar{x}(t)\| = \|c(t_j) - c(t)\| < \beta_N ,$$

for $t \in [t_j, t_{j+1})$ as $j = -N, \ldots, -1$, which implies that the extensions of $\widehat{x}_N(\cdot)$ converge to $\bar{x}(t)$ uniformly on $[a - \theta, a)$.

Let us analyze the situation on $[a, b]$. First we claim that $\widehat{x}_N(t_j) \in U$ and $\widehat{z}_N(t_j) \in V$ for $j = 0, \ldots, k+1$. Arguing by induction, we obviously have $\widehat{x}_N(t_0) \in U$ and $\widehat{z}_N(t_0) \in V$. Assume that $\widehat{x}_N(t_j) \in U$ and $\widehat{z}_N(t_j) \in U$ for all $j = 1, \ldots, m$ with some fixed number $m \in \{1, \ldots, k\}$. Then

$$\begin{aligned} &\|\widehat{x}_N(t_{m+1}) - u_N(t_{m+1})\| \\ &= \|\widehat{z}_N(t_{m+1}) - A\widehat{x}_N(t_{m+1} - \theta) - s_N(t_{m+1}) + Au_N(t_{m+1} - \theta)\| \\ &\le \|A\| \cdot \|\widehat{x}_N(t_{m+1} - \theta) - u_N(t_{m+1} - \theta)\| + \|\widehat{z}_N(t_{m+1}) - s_N(t_{m+1})\| \\ &\le \|A\| \cdot \|\widehat{x}_N(t_{m+1} - \theta) - u_N(t_{m+1} - \theta)\| + \|A\| \cdot \|\widehat{x}_N(t_m - \theta) - u_N(t_m - \theta)\| \\ &+ \|\widehat{x}_N(t_m) - u_N(t_m)\| + h_N \operatorname{dist}\big(w_{N_m}; F(\widehat{x}_N(t_m), \widehat{x}_N(t_m - \theta), \widehat{z}_N(t_m), t_m)\big) . \end{aligned}$$

Taking into account the estimates

$$\begin{aligned} &\|\widehat{x}_N(t_m) - u_N(t_m)\| \le \|A\| \cdot \|\widehat{x}_N(t_{m-N}) - u_N(t_{m-N})\| \\ &+ \|A\| \cdot \|\widehat{x}_N(t_{m-1-N}) - u_N(t_{m-1-N})\| + \|\widehat{x}_N(t_{m-1}) - u_N(t_{m-1})\| \\ &+ h_N \operatorname{dist}\big(w_{N_{m-1}}; F(\widehat{x}_N(t_{m-1}), \widehat{x}_N(t_{m-1-N}), \widehat{z}_N(t_{m-1}), t_{m-1})\big) , \end{aligned}$$

$$\begin{cases} \text{dist}\big(w_{N_{m-1}}; F(\widehat{x}_N(t_{m-1}), \widehat{x}_N(t_{m-1-N}), \widehat{z}_N(t_{m-1}), t_{m-1})\big) \\ \le \text{dist}\big(w_{N_{m-1}}; F(u_N(t_{m-1}), u_N(t_{m-1-N}), s_N(t_{m-1}), t_{m-1})\big) \\ +\ell_F\big(\|\widehat{x}_N(t_{m-1}) - u_N(t_{m-1})\| + \|\widehat{z}_N(t_{m-1}) - s_N(t_{m-1})\| \\ +\|\widehat{x}_N(t_{m-1-N}) - u_N(t_{m-1-N})\|\big)\ , \end{cases}$$

$$\|\widehat{z}_N(t_m) - s_N(t_m)\| \le \|\widehat{x}_N(t_m) - u_N(t_m)\| + \|A\| \cdot \|\widehat{x}_N(t_{m-N}) - u_N(t_{m-N})\|\ ,$$

and that $\|\widehat{x}_N(t_j) - u_N(t_j)\| = 0$ for $j \le 0$, we get

$$\begin{aligned} &\|\widehat{x}_N(t_{m+1}) - u_N(t_{m+1})\| \\ &\le M_1 h_N \sum_{j=0}^{m} \text{dist}\big(w_{N_j}; F(u_N(t_j), u_N(t_j-\theta), s_N(t_j), t_j)\big) \le M_1\gamma_N \end{aligned} \tag{7.4}$$

with some constant $M_1 > 0$, where the numbers γ_N are defined in (7.2) for each $N \in I\!\!N$. Now invoking the above estimate for $\|y_N(t)\| = \|u_N(t) - \bar{x}(t)\|$ and increasing M_1 if necessary, we arrive at

$$\|\widehat{x}_N(t_{m+1}) - \bar{x}(t_{m+1})\| \le \xi_N + M_1\gamma_N \to 0 \quad \text{as} \quad N \to \infty\ ,$$

which implies that $\widehat{x}_N(t_j) \in U$ for all $j = 0, \ldots, k+1$.

Observing further that

$$\begin{aligned} &\|\widehat{z}_N(t_{m+1}) - s_N(t_{m+1})\| \le \|\widehat{z}_N(t_m) - s_N(t_m)\| + h_N\|v_{N_m} - w_{N_m}\| \\ &\le \|\widehat{z}_N(t_m) - s_N(t_m)\| + h_N \text{dist}\big(w_{N_m}; F(\widehat{x}_N(t_m), \widehat{x}_N(t_m-\theta), \widehat{z}_N(t_m), t_m)\big)\ , \end{aligned}$$

we derive from the above estimate that

$$\begin{aligned} &\|\widehat{z}_N(t_{m+1}) - s_N(t_{m+1})\| \\ &\le M_2 h_N \sum_{j=0}^{m} \text{dist}\big(w_{N_j}; F(u_N(t_j), u_N(t_j-\theta), s_N(t_j), t_j)\big) \le M_2\gamma_N \end{aligned} \tag{7.5}$$

with some constant $M_2 > 0$. Note also that

$$\begin{aligned} &\|\widehat{z}_N(t_{m+1}) - \bar{z}_N(t_{m+1})\| \\ &\le \|\widehat{z}_N(t_{m+1}) - s_N(t_{m+1})\| + \|s_N(t_{m+1}) - \bar{z}_N(t_{m+1})\| \le M_2\gamma_N + \xi_N\ , \end{aligned}$$

which ensures the inclusion $\widehat{z}_N(t_j) \in V$ for all $j = 0, \ldots, k+1$.

It remains to prove that the sequence $\{\widehat{z}_N(\cdot)\}$ converges to $\bar{z}(\cdot)$ in the $W^{1,2}$ norm topology on $[a, b]$, i.e., one has

$$\max_{t\in[a,b]} \|\widehat{z}_N(t) - \bar{z}(t)\| + \int_a^b \|\dot{\widehat{z}}_N(t) - \dot{\bar{z}}(t)\|^2\,dt \to 0 \quad \text{as} \quad N \to \infty\,. \tag{7.6}$$

To furnish this, we use (7.4) and (7.5) to derive

$$\begin{aligned}
\sum_{j=0}^{k+1} \|\widehat{x}_N(t_j) - u_N(t_j)\| &\le \sum_{j=0}^{k+1} M_1 \sum_{m=0}^{j-1} h_N \operatorname{dist}\big(w_{N_m}; F(u_N(t_m), u_N(t_m-\theta),\\
&\quad s_N(t_m), t_m)\big)\\
&\le M_1(b-a)\sum_{j=0}^{k} \operatorname{dist}\big(w_{N_j}; F(u_N(t_j), u_N(t_j-\theta),\\
&\quad s_N(t_j), t_j)\big)\,,\\
\sum_{j=0}^{k+1} \|\widehat{z}_N(t_j) - s_N(t_j)\| &\le \sum_{j=0}^{k+1} M_2 \sum_{m=0}^{j-1} h_N\\
&\quad \operatorname{dist}\big(w_{N_m}; F(u_N(t_m), u_N(t_m-\theta), s_N(t_m), t_m)\big)\\
&\le M_2(b-a)\sum_{j=0}^{k} \operatorname{dist}\big(w_{N_j}; F(u_N(t_j), u_N(t_j-\theta),\\
&\quad s_N(t_j), t_j)\big)\,,
\end{aligned}$$

which imply by (D1) and (7.2)–(7.5) that

$$\begin{aligned}
&\int_a^b \big\|\dot{\widehat{z}}_N(t) - w_N(t)\big\|\,dt = \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \|\dot{\widehat{z}}_N(t) - w_N(t)\|\,dt\\
&= \sum_{j=0}^{k} h_N \operatorname{dist}\big(w_{N_j}; F(u_N(t_j), u_N(t_j-\theta), s_N(t_j), t_j)\big)\\
&+ \sum_{j=0}^{k} h_N \Big[\operatorname{dist}\big(w_{N_j}; F(\widehat{x}_N(t_j), \widehat{x}_N(t_j-\theta), \widehat{z}_N(t_j), t_j)\big)\\
&-\operatorname{dist}\big(w_{N_j}; F(u_N(t_j), u_N(t_j-\theta), s_N(t_j), t_j)\big)\Big]
\end{aligned}$$

$$\le \sum_{j=0}^{k} h_N \text{dist}\big(w_{N_j}; F(u_N(t_j), u_N(t_j-\theta), s_N(t_j), t_j)\big)$$

$$+ \sum_{j=0}^{k} \ell_F h_N \Big[\|\widehat{x}_N(t_j) - u_N(t_j)\| + \|\widehat{x}_N(t_j-\theta) - u_N(t_j-\theta)\| + \|\widehat{z}_N(t_j) - s_N(t_j)\|\Big]$$

$$\le \gamma_N + 2(M_1+M_2)(b-a)\ell_F \sum_{j=0}^{k} h_N \text{dist}\big(w_{N_j}; F(u_N(t_j), u_N(t_j-\theta), s_N(t_j), t_j)\big)$$

$$\le \gamma_N + 2(M_1+M_2)\ell_F(b-a)\gamma_N \ .$$

The latter ensures the estimates

$$\int_a^b \|\dot{\widehat{z}}_N(t) - \dot{\bar{z}}(t)\|\, dt \le \int_a^b \|\dot{\widehat{z}}_N(t) - w_N(t)\|\, dt + \int_a^b \|w_N(t) - \dot{\bar{z}}(t)\|\, dt$$

$$\le \gamma_N\big(1 + 2(M_1+M_2)(b-a)\ell_F\big) + \xi_N \ ,$$

which yield by (D1) and (7.3) that $\|\dot{\widehat{z}}_N(t)\| \le m_F$ and $\|\dot{\bar{z}}(t)\| \le m_F$. Hence

$$\int_a^b \|\dot{\widehat{z}}_N(t) - \dot{\bar{z}}(t)\|^2\, dt = \int_a^b \|\dot{\widehat{z}}_N(t) - \dot{\bar{z}}(t)\| \cdot \big\|\dot{\widehat{z}}_N(t) + \dot{\bar{z}}(t)\big\|\, dt$$

$$\le 2m_F\big[\gamma_N(1 + 2(M_1+M_2)(b-a)\ell_F) + \xi_N\big] \downarrow 0 \ \text{ as } \ N \to \infty \ .$$

Observing finally that

$$\max_{t\in[a,b]} \|\widehat{z}_N(t) - \bar{z}(t)\|^2 \le (b-a)\int_a^b \|\dot{\widehat{z}}_N(t) - \dot{\bar{z}}(t)\|^2\, dt \ ,$$

we arrive at (7.6) and complete the proof of the theorem. △

7.1.2 Strong Convergence of Discrete Approximations

The goal of this subsection is to construct a sequence of *well-posed* discrete approximations of the dynamic optimization problem (DA) such that optimal solutions for discrete approximation problems *strongly converge*, in the sense described below, to a given *optimal solution* for the original optimization problem governed by delayed differential-algebraic inclusions. The following construction, similar to the one in Subsect. 6.1.3 in the case of ordinary evolution inclusions, explicitly involves the optimal solution $\{\bar{x}(\cdot), \bar{z}(\cdot)\}$ to the problem (DA) under consideration for which we aim to derive necessary optimality conditions in the subsequent subsections. As one can see from the proofs, the results obtained hold also for *relaxed intermediate local minimizers* (cf. Subsects. 6.1.2 and 6.1.3), while we restrict ourself to the setting of global solutions/absolute (actually strong) minimizers for simplicity.

For any natural number N, consider the following *discrete-time* dynamic optimization problem (DA_N):

$$\begin{aligned}&\text{minimize}\quad J_N[x_N,z_N]:=\varphi\big(x_N(t_0),x_N(t_{k+1})\big)+\|x_N(t_0)-\bar{x}(a)\|^2\\&+h_N\sum_{j=0}^{k}\vartheta\Big(x_N(t_j),x_N(t_j-\theta),z_N(t_j),\frac{z_N(t_{j+1})-z_N(t_j)}{h_N},t_j\Big)\\&+\sum_{j=0}^{k}\int_{t_j}^{t_{j+1}}\Big\|\frac{z_N(t_{j+1})-z_N(t_j)}{h_N}-\dot{\bar{z}}(t)\Big\|^2\,dt\end{aligned}\tag{7.7}$$

subject to the *dynamic constraints* governed by delayed difference-algebraic inclusions (7.1), the *perturbed endpoint constraints*

$$\big(x_N(t_0),x_N(t_{k+1})\big)\in\Omega_N:=\Omega+\eta_N I\!B,\tag{7.8}$$

where $\eta_N:=\|\widehat{x}_N(t_{k+1})-\bar{x}(b)\|$ with the approximation $\widehat{x}_N(\cdot)$ of $\bar{x}(\cdot)$ from Theorem 7.1, and the *auxiliary constraints*

$$\|x_N(t_j)-\bar{x}(t_j)\|\le\varepsilon,\quad\|z_N(t_j)-\bar{z}(t_j)\|\le\varepsilon,\quad j=1,\ldots,k+1\,,\tag{7.9}$$

with some $\varepsilon>0$. The latter auxiliary constraints are needed to guarantee the *existence* of optimal solutions in (DA_N) and can be *ignored* in the derivation of necessary optimality conditions; see below.

In what follows we select $\varepsilon>0$ in (7.9) such that $\bar{x}(t)+\varepsilon I\!B\subset U$ for all $t\in[a-\theta,b]$ and $\bar{z}(t)+\varepsilon I\!B\subset V$ for all $t\in[a,b]$. Take sufficiently large N ensuring that $\eta_N<\varepsilon$. Note that problems (DA_N) have *feasible* solutions, since the pair $\{\widehat{x}_N(\cdot),\widehat{z}_N(\cdot)\}$ from Theorem 7.1 satisfies all the constraints (7.1), (7.8), and (7.9). Therefore, by the classical Weierstrass theorem, each (DA_N) admits an *optimal* pair $\{\bar{x}_N(\cdot),\bar{z}_N(\cdot)\}$ under the following assumption imposed in addition to (D1)–(D3):

(D4) φ is continuous on $U\times U$, $\vartheta(x,y,z,v,\cdot)$ is continuous for a.e. $t\in[a,b]$ uniformly in $(x,y,z,v)\in U\times U\times V\times m_F I\!B$, $\vartheta(\cdot,\cdot,\cdot,\cdot,t)$ is continuous on $U\times U\times V\times m_F I\!B$ uniformly in $t\in[a,b]$, and Ω is locally closed around $\big(\bar{x}(a),\bar{x}(b)\big)$.

We are going to justify the strong convergence of $\{\bar{x}_N(\cdot),\bar{z}_N(\cdot)\}$ to $\{\bar{x}(\cdot),\bar{z}(\cdot)\}$ in the sense of Theorem 7.1. To proceed, we need to involve an important intrinsic property of the original problem (DA) called *relaxation stability*; cf. Subsect. 6.1.2. Let us consider, along with the original delayed differential-algebraic system in (DA), the *convexified* one

$$\begin{aligned}&\dot{z}(t) \in \operatorname{co} F\big(x(t), x(t-\theta), z(t), t\big) \quad \text{a.e. } t \in [a,b]\,,\\&z(t) = x(t) + Ax(t-\theta), \quad t \in [a,b]\,.\end{aligned} \tag{7.10}$$

Further, given the integrand ϑ in (DA), we take its restriction

$$\vartheta_F(x,y,z,v,t) := \vartheta(x,y,z,v,t) + \delta\big(v; F(x,y,z,t)\big)$$

to the set $F(x,y,z,t)$ for each (x,y,z,t). Denote by $\widehat{\vartheta}_F(x,y,z,v,t)$ the *convexification* of ϑ_F in the v variable and define the *relaxed generalized Bolza problem* $(\overline{DA})$ for delayed differential-algebraic systems as follows: minimize

$$\widehat{J}[x,z] := \varphi\big(x(a), x(b)\big) + \int_a^b \widehat{\vartheta}_F\big(x(t), x(t-\theta), z(t), \dot{z}(t), t\big)\,dt \tag{7.11}$$

over feasible pairs $\{x(\cdot), z(\cdot)\}$ subject to the same tail and endpoint constraints as in (DA). Every feasible pair for $(\overline{DA})$ is called a *relaxed pair* for (DA).

One clearly has $\inf(\overline{DA}) \le \inf(DA)$ for the optimal values of the cost functionals in the relaxed and original problems. We say that the original problem (DA) is *stable with respect to relaxation* if

$$\inf(DA) = \inf(\overline{DA})\,.$$

This property, which obviously holds under the convexity assumptions on the sets $F(x,y,z,t)$ and the integrand ϑ in v, goes far beyond the convexity; cf. the discussion in Subsect. 6.1.2 for ordinary evolution inclusions. There are no difference in fact, from the viewpoint of relaxation stability, between ordinary differential systems and those with time delays only in state variables. However, it is not the case for neutral and differential-algebraic systems. We refer the reader to the book by Kisielewicz [682] for general conditions ensuring the relaxation stability of neutral functional-differential systems with nonconvex velocity sets; similar results hold for differential-algebraic systems under consideration.

Now we are ready to establish the following strong convergence theorem for optimal solutions to discrete approximations, which *makes a bridge* between optimal control problems governed by delayed differential-algebraic and difference-algebraic systems.

Theorem 7.2 (strong convergence of optimal solutions for difference-algebraic approximations). *Let $\{\bar{x}(\cdot), \bar{z}(\cdot)\}$ be an optimal pair for problem (DA), which is assumed to be stable with respect to relaxation. Suppose also that hypotheses* (D1)–(D4) *hold. Then any sequence $\{\bar{x}_N(\cdot), \bar{z}_N(\cdot)\}$, $N \in \mathbb{N}$, of optimal pairs for (DA_N) extended to the continuous interval $[a-\theta, b]$ and $[a,b]$ respectively, strongly converges to $\{\bar{x}(\cdot), \bar{z}(\cdot)\}$ as $N \to \infty$ in the sense that $\bar{x}_N(\cdot)$ converge to $\bar{x}(\cdot)$ uniformly on $[a-\theta, b]$ and $\bar{z}_N(\cdot)$ converge to $\bar{z}(\cdot)$ in the $W^{1,2}$ norm topology on $[a,b]$.*

Proof. We know from the above discussion that (DA_N) has an optimal pair $\{\bar{x}_N(\cdot), \bar{z}_N(\cdot)\}$ for all N sufficiently large; suppose that it happens for all $N \in I\!N$ without loss of generality. Consider the sequence $\{\widehat{x}_N(\cdot), \widehat{z}_N(\cdot)\}$ from the strong approximation result of Theorem 7.1 applied to the given optimal solution $\{\bar{x}(\cdot), \bar{z}(\cdot)\}$. Since each $\{\widehat{x}_N(\cdot), \widehat{z}_N(\cdot)\}$ is feasible for (DA_N), we have

$$J_N[\bar{x}_N, \bar{z}_N] \le J_N[\widehat{x}_N, \widehat{z}_N] \ \text{ whenever } \ N \in I\!N \ .$$

For convenience we represent $J_N[\widehat{x}_N, \widehat{z}_N]$ as the sum of three terms:

$$J_N[\widehat{x}_N, \widehat{z}_N] = I_1 + I_2 + I_2 := \varphi\big(\widehat{x}_N(t_0), \widehat{x}_N(t_{k+1})\big)$$

$$+h_N \sum_{j=0}^{k} \vartheta\left(\widehat{x}_N(t_j), \widehat{x}_N(t_j - \theta), \widehat{z}_N(t_j), \frac{\widehat{z}_N(t_{j+1}) - \widehat{z}_N(t_j)}{h_N}, t_j\right)$$

$$+\sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \left\| \frac{\widehat{z}_N(t_{j+1}) - \widehat{z}_N(t_j)}{h_N} - \dot{\bar{z}}(t) \right\|^2 dt \ .$$

It follows from Theorem 7.1 and the assumption on φ in (D4) that

$$I_1 \to \varphi\big(\bar{x}(a), \bar{x}(b)\big) \ \text{ as } \ N \to \infty \ .$$

Our goal is to show that

$$\limsup_{N\to\infty} J_N[\bar{x}_N, \bar{z}_N] \le J[\bar{x}, \bar{z}], \tag{7.12}$$

which clearly follows from the limiting relation

$$J_N[\widehat{x}_N, \widehat{z}_N] \to J[\bar{x}, \bar{z}] \ \text{ as } \ N \to \infty \ .$$

To justify this, we need to compute the limits of the terms I_2 and I_3 in the above representation for $J_N[\widehat{x}_N, \widehat{z}_N]$. Using the sign "$\sim$" for expressions that are equivalent as $N \to \infty$ and the notation

$$\widehat{v}_N(t) := \frac{\widehat{z}_N(t_{j+1}) - \widehat{z}_N(t_j)}{h_N}, \quad t \in [t_j, t_{j+1}), \ j = 0, \ldots, k \ ,$$

we have the relations:

$$I_2 = h_N \sum_{j=0}^{k} \vartheta\big(\widehat{x}_N(t_j), \widehat{x}_N(t_j-\theta), \widehat{z}_N(t_j), \widehat{v}_N(t_j), t_j\big)$$

$$= \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \vartheta\big(\widehat{x}_N(t_j), \widehat{x}_N(t_j-\theta), \widehat{z}_N(t_j), \widehat{v}_N(t), t\big)\, dt$$

$$+ \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \Big[\vartheta\big(\widehat{x}_N(t_j), \widehat{x}_N(t_j-\theta), \widehat{z}_N(t_j), \widehat{v}_N(t), t_j\big)$$

$$-\vartheta\big(\widehat{x}_N(t_j), \widehat{x}_N(t_j-\theta), \widehat{z}_N(t_j), \widehat{v}_N(t), t\big)\Big]\, dt$$

$$= \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \vartheta\big(\widehat{x}_N(t_j), \widehat{x}_N(t_j-\theta), \widehat{z}_N(t_j), \widehat{v}_N(t), t\big)\, dt + \tau(\vartheta; h_N)$$

$$\sim \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \vartheta\big(\widehat{x}_N(t_j), \widehat{x}_N(t_j-\theta), \widehat{z}_N(t_j), \widehat{v}_N(t), t\big)\, dt$$

$$\to \int_a^b \vartheta\big(\bar{x}(t), \bar{x}(t-\theta), \bar{z}(t), \dot{\bar{z}}(t), t\big)\, dt \quad \text{as} \quad N \to \infty, \qquad \text{and}$$

$$I_3 = \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \big\|\widehat{v}_N(t) - \dot{\bar{z}}(t)\big\|^2\, dt = \int_a^b \big\|\widehat{v}_N(t) - \dot{\bar{z}}(t)\big\|^2\, dt$$

$$= \int_a^b \big\|\dot{\widehat{z}}_N(t) - \dot{\bar{z}}(t)\big\|^2\, dt \to 0 \quad \text{as} \quad N \to \infty\,,$$

which finally imply the required inequality (7.12).

Further, it is easy to observe that the strong convergence asserted in the theorem follows from

$$\beta_N := \|\bar{x}_N(a) - \bar{x}(a)\|^2 + \int_a^b \|\dot{\bar{z}}_N(t) - \dot{\bar{z}}(t)\|^2 dt \to 0 \quad \text{as} \quad N \to \infty\,.$$

On the contrary, suppose that the latter doesn't hold. Then there are $\beta > 0$ and a sequence $\{N_m\} \subset I\!N$ for which $\beta_{N_m} \to \beta$ as $m \to \infty$. Employing the standard compactness arguments based on (7.1) and the boundedness assumption in (D1) in the framework of finite-dimensional state spaces, we find an absolutely continuous mapping $\widetilde{z}\colon [a,b] \to I\!R^n$ and another mapping $\widetilde{x}\colon [a-\theta, b]$ continuous on $[a-\theta, a)$ and $[a,b]$ such that

$$\dot{\bar{z}}_N(t) \to \dot{\widetilde{z}}(t) \quad \text{weakly in } L^2[a,b]\,,$$

that $\bar{x}_N(t) \to \widetilde{x}(t)$ uniformly on $[a-\theta, b]$ as $N \to \infty$ (without loss of generality), and that $\widetilde{z}(t) = \widetilde{x}(t) + A\widetilde{x}(t-\theta)$ for $t \in [a,b]$. By the classical Mazur theorem there is a sequence of *convex combinations* of $\dot{\bar{z}}_N(t)$ that converges to $\dot{\widetilde{z}}(t)$ in the norm topology of $L^2[a,b]$ and hence *pointwisely* for a.e. $t \in [a,b]$ along some subsequence. Therefore

$$\begin{cases} \dot{\widetilde{z}}(t) \in \operatorname{co} F\big(\widetilde{x}(t), \widetilde{x}(t-\theta), \widetilde{z}(t), t\big) \ \text{ a.e. } \ t \in [a,b] , \\ \widetilde{z}(t) = \widetilde{x}(t) + A\widetilde{x}(t-\theta), \quad t \in [a,b] . \end{cases}$$

Since $\widetilde{x}(\cdot)$ obviously satisfies the initial tail condition and the endpoint constraints in (DA), it is feasible for the relaxed problem $(\overline{DA})$. Note that

$$h_N \sum_{j=0}^{k} \vartheta\Big(\bar{x}_N(t_j), \bar{x}_N(t_j-\theta), \bar{z}_N(t_j), \frac{\bar{z}_N(t_{j+1}) - \bar{z}_N(t_j)}{h_N}, t_j\Big)$$

$$= \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \vartheta\big(\bar{x}_N(t_j), \bar{x}_N(t_j-\theta), \bar{z}_N(t_j), \dot{\bar{z}}_N(t), t_j\big)\, dt$$

$$\to \int_a^b \vartheta\big(\widetilde{x}(t), \widetilde{x}(t-\theta), \widetilde{z}(t), \dot{\widetilde{z}}(t), t\big)\, dt \quad \text{as} \quad N \to \infty$$

due to the assumptions made. Observe also that the integral functional

$$I[v] := \int_a^b \big\|v(t) - \dot{\bar{z}}(t)\big\|^2 dt$$

is *lower semicontinuous* in the weak topology of $L^2[a,b]$ by the *convexity* of the integrand in v. Since

$$\sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \left\| \frac{\bar{z}_N(t_{j+1}) - \bar{z}_N(t_j)}{h_N} - \dot{\bar{z}}(t) \right\|^2 dt = \int_a^b \big\|\dot{\bar{z}}_N(t) - \dot{\bar{z}}(t)\big\|^2 dt ,$$

the latter implies that

$$\int_a^b \big\|\dot{\widetilde{z}}(t) - \dot{\bar{z}}(t)\big\|^2 dt \le \liminf_{N\to\infty} \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \left\| \frac{\bar{z}_N(t_{j+1}) - \bar{z}_N(t_j)}{h_N} - \dot{\bar{z}}(t) \right\|^2 dt .$$

Using the above relationships and passing to the limit in the cost functional form (7.7) for $J_N[\bar{x}_N, \bar{z}_N]$ as $N \to \infty$, we arrive at the inequality

$$J[\widetilde{x}, \widetilde{z}] + \beta \le \lim_{N\to\infty} J_N[\bar{x}_N, \bar{z}_N] .$$

By (7.12) one therefore has

$$J[\widetilde{x}, \widetilde{z}] \le J[\bar{x}, \bar{z}] - \beta < J[\bar{x}, \bar{z}] \ \text{ if } \ \beta > 0 \ .$$

This clearly contradicts the optimality of the pair $\{\bar{x}(\cdot), \bar{z}(\cdot)\}$ in the relaxed problem $(\overline{DA})$ due to the assumption on relaxation stability. Thus $\beta = 0$, which completes the proof of the theorem. △

Note that similarly to Subsect. 6.1.3 we can modify Theorem 7.2 in the case of problems with mappings F *measurable* in $t \in [a, b]$ and also to derive an analog of Theorem 6.14 on the *value convergence* of discrete approximations for differential-algebraic systems.

7.1.3 Necessary Optimality Conditions for Difference-Algebraic Systems

In this subsection we derive necessary optimality conditions for the discrete approximation problems (DA_N) by reducing them to nonsmooth mathematical programming problems with *many geometric constraints*. The finite dimensionality of the state space $X = I\!R^n$ allows us to proceed without using the SNC calculus and/or "fuzzy" results as in Subsect. 6.1.4. Denote

$$w := (x_0, \ldots, x_{k+1}, z_0, \ldots, z_{k+1}, v_0, \ldots, v_k) \in I\!R^{n(3k+5)}$$

and define the following mappings and sets built upon the initial data of the approximating problems (DA_N) and eventually of the original problem (DA):

$$\varphi_0(w) : = \varphi(x_0, x_{k+1}) + \|x_0 - \bar{x}(a)\|^2 + h_N \sum_{j=0}^{k} \vartheta\big(x_j, x_{j-N}, z_j, v_j, t_j\big)$$

$$+ \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \big\|v_j - \dot{\bar{z}}(t)\big\|^2 \, dt \ ,$$

$$\varphi_j(w) := \begin{cases} \|x_j - \bar{x}(t_j)\| - \varepsilon, & j = 1, \ldots, k+1 \ , \\ \|z_{j-k-1} - \bar{z}(t_{j-k-1})\| - \varepsilon, & j = k+2, \ldots, 2k+2 \ , \end{cases}$$

$$\Lambda_j := \big\{(x_0, \ldots, v_k) \ \big| \ v_j \in F(x_j, x_{j-N}, z_j, t_j)\big\}, \quad j = 0, \ldots, k \ ,$$

$$\Lambda_{k+1} := \big\{(x_0, \ldots, v_k) \ \big| \ (x_0, x_{k+1}) \in \Omega_N\big\} \ ,$$

$$g_j(w) := z_{j+1} - z_j - h_N v_j, \quad j = 0, \ldots, k \ ,$$

$$s_j(w) := z_j - x_j - A x_{j-N}, \quad j = 0, \ldots, k+1 \ ,$$

where $x_j := c(t_j)$ for $j < 0$. Then each problem (DA_N) *equivalently* reduces to the following problem (MP) of *nonsmooth mathematical programming* in $R^{n(3k+5)}$ with finitely $(k+2)$ many *geometric constraints*:

$$\begin{cases} \text{minimize } \ \varphi_0(w) \ \text{ subject to} \\ \varphi_j(w) \le 0, \quad j = 1, \ldots, r\ , \\ f(w) = 0\ , \\ w \in \Lambda_j, \quad j = 0, \ldots, l\ , \end{cases}$$

where $r = 2k+2$, $l = k+1$, and $f\colon I\!R^{n(3k+5)} \to I\!R^{2k+3}$ is given by

$$f(w) := \big(g_0(w), \ldots, g_k(w), s_0(w), \ldots, s_{k+1}(w)\big), \quad w \in I\!R^{n(3k+5)}\ .$$

For simplicity we skip indicating the dependence of solutions to (DA_N) and the corresponding dual elements on the approximation number N.

Let $\bar{w}$ be an optimal solution to problem (MP) corresponding to those (as $N \in I\!N$) for discrete approximations under consideration. In what follows we assume the local Lipschitz continuity of the functions φ_0 and $\vartheta(\cdot, t)$. Applying now the necessary optimality conditions for (MP) from Proposition 6.16 in the case of finite-dimensional spaces and separating (vector) multipliers for the equality constraint components g_j and s_j of the mapping f, we find $\mu_j \in I\!R$ as $j = 0, \ldots, 2k+2$, $e_j^* \in I\!R^n$ as $j = 0, \ldots, k$, $d_j^* \in I\!R^n$ as $j = 0, \ldots, k+1$, and $w_j^* \in I\!R^{n(3k+5)}$ as $j = 0, \ldots, k+1$ satisfying

$$\begin{cases} \mu_j \ge 0 \quad \text{for} \quad j = 0, \ldots, 2k+2\ , \\ \mu_j \varphi_j(\bar{w}) = 0 \quad \text{for} \quad j = 1, \ldots, 2k+2\ , \\ w_j^* \in N(\bar{w}; \Lambda_j) \quad \text{for} \quad j = 0, \ldots, k+1\ , \\ -\displaystyle\sum_{j=0}^{k+1} w_j^* \in \partial\Big(\sum_{j=0}^{2k+2} \mu_j \varphi_j\Big)(\bar{w}) + \sum_{j=0}^{k} \nabla g_j(\bar{w})^* e_j^* + \sum_{j=0}^{k+1} \nabla s_j(\bar{w})^* d_j^*\ . \end{cases} \tag{7.13}$$

Representing $w_j^* = (x_{0j}^*, \ldots, x_{k+1\,j}^*, z_{0j}^*, \ldots, z_{k+1\,j}^*, v_{0j}^*, \ldots, v_{kj}^*)$, note that all but one components of each w_j^* are zero and the remaining component satisfies

$$(x_{jj}^*, x_{j-N\,j}^*, z_{jj}^*, v_{jj}^*) \in N\big((\bar{x}_j, \bar{x}_{j-N}, \bar{z}_j, \bar{v}_j); \operatorname{gph} F(t_j)\big) \ \text{ for } \ j = 0, \ldots, k\ .$$

Similarly observe that the condition $w_{k+1}^* \in N(\bar{z}^N; \Lambda_{k+1})$ is equivalent to

$$(x_{0\,k+1}^*, x_{k+1\,k+1}^*) \in N\big((\bar{x}_0, \bar{x}_{k+1}); \Omega_N\big)$$

with all the other components of w_{k+1}^* equal to zero. It follows from the construction of φ_j for $j = 1, \ldots, 2k+2$ and the strong convergence of the discrete optimal solutions in Theorem 7.2 that

$$\varphi_j(\bar{w}) < 0 \ \text{ whenever } \ j = 1, \ldots, 2k+2 \ \text{ as } \ N \to \infty \, .$$

Thus $\mu_j = 0$ for all these indexes due to the complementary slackness conditions in (7.13), and we let $\lambda := \mu_0$ for the remaining one. Observe further from the structures of g_j and s_j in problem (MP) that

$$\sum_{j=0}^{k} \nabla g_j(\bar{w})^* e_j^* = \big(0, \ldots, 0, e_0^*, e_0^* - e_1^*, e_{k-1}^* - e_k^*, e_k^*, -h_N e_0^*, \ldots, -h_N e_k^*\big) \ \text{ and}$$

$$\sum_{j=0}^{k+1} \nabla s_j(\bar{w})^* d_j^* = \big(- d_0^* + A^* d_N^*, -d_1^* + A^* d_{N+1}^*, \ldots, -d_{k-N+1}^*$$

$$+ A^* d_{k+1}^*, -d_{k-N+2}^*, \ldots, -d_{k+1}^*, d_0^*, \ldots, d_{k+1}^*, 0, \ldots, 0\big) \, .$$

From the subdifferential sum rule of Theorem 2.33(c) applied to the Lipschitzian sum φ_0 in (MP) one has

$$\partial\varphi_0(\bar{w}) \subset \partial\varphi(\bar{x}_0, \bar{x}_{k+1}) + 2\big(\bar{x}_0 - \bar{x}(a)\big) + h_N \sum_{j=0}^{k} \partial\vartheta(\bar{x}_j, \bar{x}_{j-N}, \bar{z}_j, \bar{v}_j, t_j)$$

$$+ \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} 2\big(\bar{v}_j - \dot{\bar{z}}(t)\big)\, dt$$

with $\partial\vartheta$ standing here and in what follows for the basic subdifferential of ϑ with respect to the first four variables. Thus we get from (7.13) that

$$\begin{cases} -x_{00}^* - x_{0N}^* - x_{0\,k+1}^* = \lambda x_0^* + \lambda h_N u_0^* + \lambda h_N y_0^* \\[1ex] +2\lambda\big(\bar{x}_0 - \bar{x}(a)\big) - d_0^* - A^* d_N^* \, , \\[1ex] -x_{jj}^* - x_{j\,j+N}^* = \lambda h_N u_j^* + \lambda h_N y_j^* - d_j^* - A^* d_{j+N}^*, \quad j = 1, \ldots, k-N+1 \, , \\[1ex] -x_{jj}^* = \lambda h_N u_j^* - d_j^*, \quad j = k-N+2, \ldots, k \, , \\[1ex] -x_{k+1\,k+1}^* = \lambda x_{k+1}^* - d_{k+1}^* \, , \\[1ex] -z_{jj}^* = \lambda h_N z_j^* + d_j^* + e_{j-1}^* - e_j^*, \quad j = 0, \ldots, k \, , \\[1ex] -v_{jj}^* = \lambda h_N v_j^* + \lambda \xi_j - h_N e_j^*, \quad j = 0, \ldots, k \, , \end{cases} \tag{7.14}$$

with the notation

$$(x_0^*, x_{k+1}^*) \in \partial\varphi(\bar{x}_0^N, \bar{x}_{k+1}^N), \quad (u_j^*, y_{j-N}^*, z_j^*, v_j^*) \in \partial\vartheta(\bar{x}_j, \bar{x}_{j-N}, \bar{z}_j, \bar{v}_j^N, t_j) \, ,$$

$$\xi_j := 2\int_{t_j}^{t_{j+1}} \big(\bar{v}_j - \dot{\bar{z}}(t)\big)\, dt \, ,$$

where we don't distinguish between primal and dual vectors in the finite-dimensional spaces under consideration.

Based on the above relationships, we arrive at the following necessary optimality conditions for discrete-time problems (DA_N), where

$$\vartheta_j(\cdot,\cdot,\cdot,\cdot) := \vartheta(\cdot,\cdot,\cdot,\cdot,t_j) \text{ and } F_j(\cdot,\cdot,\cdot) := F(\cdot,\cdot,\cdot,t_j)\,.$$

These conditions hold under milder assumptions on F in comparison with (D1) and (D2), while the continuity requirements on φ and ϑ in (D4) are replaced by their Lipschitz continuity.

Theorem 7.3 (necessary optimality conditions for difference-algebraic inclusions). *Let $\bar{w}$ be an optimal solution to problem (DA_N). Assume that the sets Ω and $\operatorname{gph} F_j$ are locally closed and that the functions φ and ϑ_j are Lipschitz continuous around the points $(\bar{x}_0,\bar{x}_{k+1})$ and $(\bar{x}_j,\bar{x}_{j-N},\bar{z}_j,\bar{v}_j)$, respectively, for all $j=0,\ldots,k$. Then there exist $\lambda\ge 0$, $p_j\in I\!R^n$ as $j=0,\ldots,k+N+1$, $q_j\in I\!R^n$ as $j=-N,\ldots,k+1$, and $r_j\in I\!R^n$ as $j=0,\ldots,k+1$, not all zero, satisfying the conditions*

$$p_j=0,\quad j=k+2,\ldots,k+N+1, \tag{7.15}$$

$$q_j=0,\quad j=k-N+1,\ldots,k+1, \tag{7.16}$$

$$(p_0+q_0,-p_{k+1})\in\lambda\partial\varphi(\bar{x}_0,\bar{x}_{k+1})+N\big((\bar{x}_0,\bar{x}_{k+1});\Omega_N\big), \tag{7.17}$$

and the following difference-algebraic analog of the Euler-Lagrange inclusion:

$$\Big(\frac{\widetilde{p}_{j+1}-\widetilde{p}_j}{h_N},\ \frac{\widetilde{q}_{j-N+1}-\widetilde{q}_{j-N}}{h_N},\ \frac{r_{j+1}-r_j}{h_N},\ -\frac{\lambda\xi_j}{h_N}+p_{j+1}+q_{j+1}+r_{j+1}\Big)$$
$$\in\lambda\partial\vartheta_j(\bar{x}_j,\bar{x}_{j-N},\bar{z}_j,\bar{v}_j)+N\big((\bar{x}_j,\bar{x}_{j-N},\bar{z}_j,\bar{v}_j);\operatorname{gph}F_j\big)$$

for $j=1,\ldots,k$ with the notation

$$\widetilde{p}_j := p_j + A^*p_{j+N}\quad\textit{and}\quad \widetilde{q}_j := q_j + A^*q_{j+N}\ ,$$

Proof. Most of the proof has been actually done above, where we transformed the necessary optimality conditions for (MP) into the ones for (DA_N) written in the form of nonsmooth mathematical programming. What we need to do is to change the notation in the relationships of (7.14). Let us first denote

$$\widetilde{d}^*_j := \begin{cases} d^*_j \text{ for } j=1,\ldots,k+1\,,\\[2mm] 0\quad\text{for } j=k+2,\ldots,k+N\ ,\end{cases}$$

$$\widetilde{y}_j^* := \begin{cases} \lambda y_j^* + \dfrac{x_{j\,j+N}^*}{h_N} & \text{for } j = 1, \dots, k-N+1, \\ 0 & \text{for } j = k-N+2, \dots, k\,, \end{cases}$$

and $\widetilde{r}_j := e_{j-1}^*$ for $j = 1, \dots, k+1$. From (7.14) we have the relationships

$$\begin{cases} \widetilde{d}_j^* + A^* \widetilde{d}_{j+N}^* - \widetilde{y}_j^* = \lambda u_j^* + \dfrac{x_{j\,j}^*}{h_N}\,, \\ \widetilde{y}_{j-N}^* = \lambda y_{j-N}^* + \dfrac{x_{j-N\,j}^*}{h_N}\,, \\ \dfrac{\widetilde{r}_{j+1} - \widetilde{r}_j}{h_N} - \widetilde{d}_j^* = \lambda z_j^* + \dfrac{z_{j\,j}^*}{h_N}\,, \\ -\lambda \dfrac{\xi_j}{h_N} + \widetilde{r}_{j+1} = \lambda v_j^* + \dfrac{v_{j\,j}^*}{h_N} \end{cases} \tag{7.18}$$

for $j = 1, \dots, k$. Define $\widehat{p}_j$ and $\widehat{q}_j$ recurrently by

$$\widehat{p}_j := \widehat{p}_{j+1} - h_N \widetilde{d}_j^* \quad \text{with} \quad \widehat{p}_j = 0 \quad \text{for} \quad j = k+2, \dots, k+N+1\,,$$

$$\widehat{q}_j := \widetilde{q}_{j+1} - h_N \widetilde{y}_j \quad \text{with} \quad \widehat{q}_j = 0 \quad \text{for} \quad j = k-N+1, \dots, k+N+1\,.$$

Putting now $q_j := \widehat{q}_j + A^* \widehat{q}_{j+N}$, we rewrite (7.18) as

$$\begin{cases} \dfrac{(\widehat{p}_{j+1} - q_{j+1}) - (\widetilde{p}_j - q_j)}{h_N} + A^* \dfrac{(\widehat{p}_{j+N+1} - q_{j+N+1}) - (\widehat{p}_{j+N} - q_{j+N})}{h_N} \\ = \lambda u_j^* + \dfrac{x_{j\,j}^*}{h_N}, \quad j = 1, \dots, k\,, \\ \dfrac{(q_{j-N+1} + A^* q_{j+1}) - (q_{j-N} + A^* q_j)}{h_N} = \lambda y_{j-N}^* + \dfrac{x_{j-N\,j}^*}{h_N}, \quad j = 1, \dots, k\,, \\ \dfrac{\widetilde{r}_{j+1} - \widetilde{r}_j}{h_N} - \dfrac{\widehat{p}_{j+1} - \widehat{p}_j}{h_N} = \lambda z_j^* + \dfrac{z_{jj}^*}{h_N}, \quad j = 1, \dots, k\,, \\ -\lambda \dfrac{\xi_j}{h_N} + \widetilde{r}_{j+1} = \lambda v_j^* + \dfrac{v_{j\,j}^*}{h_N}, \quad j = 1, \dots, k\,. \end{cases}$$

Letting finally

$$p_0 := \lambda x_0^* + x_{0\,k+1}^* - q_0\,,$$

$$p_j := \widehat{p}_j - q_j \quad \text{for} \quad j = 1, \dots, k+N+1, \quad \text{and}$$

$$r_j := \widetilde{r}_j - \widehat{p}_j \quad \text{for} \quad j = 1, \dots, k+1\,,$$

we arrive at the necessary optimality conditions of the theorem. △

The following corollary justifies, under additional assumptions, necessary conditions of Theorem 7.3 with some *enhanced nontriviality* used in the next subsection in the proof of optimality conditions for the continuous-time problem (DA) by passing to the limit from discrete approximations.

Corollary 7.4 (necessary conditions for difference-algebraic inclusions with enhanced nontriviality). *In addition to the assumptions of Theorem 7.3, suppose that the mapping F_j is locally bounded and Lipschitz continuous around $(\bar{x}_j, \bar{x}_{j-N}, \bar{z}_j)$ for each $j = 0, \ldots, k$. Then the necessary conditions of the theorem hold with $(\lambda, p_{k+1}, r_{k+1}) \neq 0$, i.e., one can let*

$$\lambda^2 + \|p_{k+1}\|^2 + \|r_{k+1}\|^2 = 1 . \tag{7.19}$$

Proof. If $\lambda = 0$, then the Euler-Lagrange inclusion of the theorem implies, together with conditions (7.15) and (7.16), that

$$\Big(\frac{p_{k+1} - p_k}{h_N}, \frac{-q_{k-N}}{h_N}, \frac{r_{k+1} - r_k}{h_N}\Big) \in D^* F_k(\bar{x}_k, \bar{x}_{k-N}, \bar{z}_k, \bar{v}_k)\big(- p_{k+1} - r_{k+1}\big) .$$

Assuming now that $p_{k+1} = 0$ and $r_{k+1} = 0$, we get

$$\Big(\frac{-p_k}{h_N}, \frac{-q_{k-N}}{h_N}, \frac{-r_k}{h_N}\Big) \in D^* F_k(\bar{x}_k, \bar{x}_{k-N}, \bar{z}_k, \bar{v}_k)(0) ,$$

which yields $p_k = 0$, $q_{k-N} = 0$, and $r_k = 0$ by the coderivative criterion of Corollary 4.11 for the local Lipschitzian property of set-valued mappings in finite dimensions. Repeating this process, we arrive at the contradiction with the nontriviality assertion of Theorem 7.3. △

7.1.4 Euler-Lagrange and Hamiltonian Conditions for Differential-Algebraic Systems

In the final subsection of this section we derive necessary optimality conditions in the extended Euler-Lagrange and Hamiltonian forms for the optimal control problem (DA) governed by differential-algebraic inclusions. Let us start with the Euler-Lagrange conditions, which give the main result of this section under the assumption on *relaxation stability*. The notation N_+ and ∂_+ in the following theorem stand for the extended normal cone and subdifferential of moving object described in Subsect. 6.1.5. Note that, similarly to the case of ordinary evolution inclusions studied in that subsection, we may consider problems (DA) with *summable integrands* and replace the extended subdifferential $\partial_+\vartheta$ in the Euler-Lagrange inclusion by the basic one $\partial\vartheta$.

Theorem 7.5 (Euler-Lagrange conditions for differential-algebraic inclusions). *Let $\{\bar{x}(\cdot), \bar{z}(\cdot)\}$ be an optimal solution to problem (DA) under the standing assumptions* (H1)–(H4)*, where the continuity of the functions φ and $\vartheta(\cdot,\cdot,\cdot,\cdot,t)$ is replaced with the corresponding local Lipschitz continuity. Suppose also that (DA) is stable with respect to relaxation. Then there* *exist a number $\lambda \geq 0$, piecewise continuous arcs $p\colon [a,b+\theta] \to I\!R^n$ and $q\colon [a-\theta,b] \to I\!R^n$ (whose points of discontinuity are confined to multiples of the delay time θ), and an absolutely continuous arc $r\colon [a,b] \to I\!R^n$ such that $p(t)+A^*p(t+\theta)$ and $q(t-\theta)+A^*q(t)$ are absolutely continuous on $[a,b]$ satisfying the relationships*

$$\lambda + \|p(b)\| + \|r(b)\| = 1\,, \tag{7.20}$$

$$p(t) = 0 \ \ for \ \ t \in (b, b+\theta], \quad q(t) = 0 \ \ for \ \ t \in (b-\theta, b], \tag{7.21}$$

$$\big(p(a)+q(a), -p(b)\big) \in \lambda\partial\varphi\big(\bar{x}(a),\bar{x}(b)\big) + N\big((\bar{x}(a),\bar{x}(b));\Omega\big) \tag{7.22}$$

and the extended Euler-Lagrange inclusion

$$\Big(\frac{d}{dt}\big[p(t)+A^*p(t+\theta)\big],\ \frac{d}{dt}\big[q(t-\theta)+A^*q(t)\big], \dot{r}(t)\Big)$$

$$\in \mathrm{co}\Big\{(u,v,w) \mid \big(u,v,w,p(t)+q(t)+r(t)\big) \in \lambda\widetilde{\partial}\vartheta\big(\bar{x}(t),\bar{x}(t-\theta),\bar{z}(t),\dot{\bar{z}}(t),t\big)$$

$$+N_+\big((\bar{x}(t),\bar{x}(t-\theta),\bar{z}(t),\dot{\bar{z}}(t));\mathrm{gph}\,F(t)\big)\Big\} \ \ a.e. \ \ t \in [a,b]\,.$$

Proof. We prove this theorem by using the method of discrete approximations and the previous results ensuring the strong convergence of discrete optimal solutions and necessary optimality conditions in the approximating problems (DA_N). For notational convenience we use in this subsection the *upper* index N to indicate the dependence on this parameter of optimal solutions $(\bar{x}^N, \bar{z}^N)$ to discrete-time problems and the corresponding elements $(\lambda^N, p^N, q^N, r^N)$ in the necessary optimality conditions from Corollary 7.4 used in what follows. Denote by $\bar{x}^N(t)$, $p^N(t)$, $q^N(t-\theta)$, and $r^N(t)$ the piecewise linear extensions of these discrete arcs to the continuous-time interval $[a,b]$ with their corresponding linear combinations $\bar{z}^N(t)$, $\widetilde{p}^N(t)$, and $\widetilde{q}^N(t-\theta)$. It follows from Theorem 7.1 that

$$\int_a^b \|\xi^N(t)\|\,dt = \sum_{j=0}^k \|\xi_j^N\| \leq 2\sum_{j=0}^k \int_{t_j}^{t_{j+1}} \big\|\dot{\bar{z}}(t) - \bar{v}_j^N\big\|\,dt$$

$$= 2\int_a^b \big\|\dot{\bar{z}}(t) - \dot{\bar{z}}^N(t)\big\|\,dt := \nu_N \to 0 \ \text{ as } \ N \to \infty$$

for $\xi^N(t) := \xi_j^N/h_N$ as $t \in [t_j, t_{j+1})$, $j = 0,\ldots,k$, with $\xi_j^N = \xi_j$ from Theorem 7.3. Assume without loss of generality that $\lambda^N \to \lambda \geq 0$,

$$\bar{v}^N(t) := \dot{\bar{z}}^N(t) \to \dot{\bar{z}}(t), \quad \text{and} \quad \xi^N(t) \to 0 \ \text{ a.e. } \ t \in [a,b] \ \text{ as } N \to \infty \, .$$

Let us estimate $\big(p^N(t), q^N(t-\theta), r^N(t)\big)$ for large N. Using (7.15) and (7.16), we derive from the Euler-Lagrange inclusion of Theorem 7.3 that

$$\Big(\frac{p^N_{j+1}-p^N_j}{h_N} - \lambda^N u^*_j, \ \frac{q^N_{j-N+1}-q^N_{j-N}}{h_N} - \lambda^N y^*_{j-N}, \ \frac{r^N_{j+1}-r^N_j}{h_N} - \lambda^N z^*_j \, ,$$

$$-\frac{\lambda^N \xi^N_j}{h_N} + p^N_{j+1} + r^N_{j+1} - \lambda^N v^*_j\Big) \in N\big((\bar{x}^N_j, \bar{x}^N_{j-N}, \bar{z}^N_j, \bar{v}^N_j); \mathrm{gph}\, F_j\big)$$

with some $(u^*_j, y^*_{j-N}, z^*_j, v^*_j) \in \partial\vartheta_j(\bar{x}^N_j, \bar{x}^N_{j-N}, \bar{z}^N_j, \bar{v}^N_j)$ for all $j = k-N+2, \ldots, k+1$. This means, by definition of the coderivative, that

$$\Big(\frac{p^N_{j+1}-p^N_j}{h_N} - \lambda^N u^*_j, \ \frac{q^N_{j-N+1}-q^N_{j-N}}{h_N} - \lambda^N y^*_{j-N}, \frac{r^N_{j+1}-r^N_j}{h_N} - \lambda^N z^*_j\Big)$$

$$\in D^*F_j(\bar{x}^N_j, \bar{x}^N_{j-N}, \bar{z}^N_j, \bar{v}^N_j)\Big(\lambda^N v^*_j + \frac{\lambda^N \xi^N_j}{h_N} - p^N_{j+1} - r^N_{j+1}\Big)$$

for such j. The coderivative criterion of Corollary 4.11 for the local Lipschitzian property of F_j with modulus ℓ_F ensures the estimate

$$\Big\|\Big(\frac{p^N_{j+1}-p^N_j}{h_N} - \lambda^N u^*_j, \ \frac{q^N_{j-N+1}-q^N_{j-N}}{h_N} - \lambda^N y^*_{j-N}, \frac{r^N_{j+1}-r^N_j}{h_N} - \lambda^N z^*_j\Big)\Big\|$$

$$\le \ell_F\Big\|\lambda^N v^*_j + \frac{\lambda^N \xi^N_j}{h_N} - p^N_{j+1} - r^N_{j+1}\Big\| \ \text{ whenever } \ j = k-N+2, \ldots, k+1 \, .$$

Since $\|(u^*_j, y^*_{j-N}, z^*_j, v^*_j)\| \le \ell_\vartheta$ due to the Lipschitz continuity of ϑ with modulus ℓ_ϑ, we derive from the above that

$$\|(p^N_j, q^N_{j-N}, r^N_j)\| \le \ell_F\|\xi^N_j\| + (\ell_F+1)h_N\ell_\vartheta$$

$$+(\ell_F h_N + 1)\|(p^N_{j+1}, q^N_{j-N+1}, r^N_{j+1})\| \le \ell_F\|\xi^N_j\| + (\ell_F h_N+1)\ell_F\|\xi^N_{j+1}\|$$

$$+(\ell_F+1)h_N\ell_\vartheta + (\ell_F h_N+1)(\ell_F+1)h_N\ell_\vartheta$$

$$+(\ell_F h_N+1)^2\|(p^N_{j+2}, q^N_{j-N+2}, r^N_{j+2})\| \le \ldots$$

$$\le \exp\big(\ell_F(b-a)\big)\Big(1 + \frac{\ell_\vartheta}{\ell_F}(\ell_F+1) + \ell_F\nu_N\Big), \quad j = k-N+2, \ldots, k+1 \, ,$$

which implies the uniform boundedness of $\{(p^N_j, q^N_{j-N}, r^N_j)|\ j = k-N+2, \ldots, k+1\}$ and hence that of $\big(p^N(t), q^N(t-\theta), r^N(t)\big)$ on $[b-\theta, b]$.

Next consider indexes $j = k-2N+2, \ldots, k-N+1$ and derive from the discrete Euler-Lagrange inclusion that

$$\Big\|\Big(\frac{p^N_{j+1}-p^N_j}{h_N}-\lambda^N u^*_j,\ \frac{q^N_{j-N+1}-q^N_{j-N}}{h_N}-\lambda^N y^*_{j-N},\ \frac{r^N_{j+1}-r^N_j}{h_N}-\lambda^N z^*_j\Big)\Big\|$$

$$\leq \ell_F\Big\|\lambda^N v^*_j+\frac{\lambda^N\xi^N_j}{h_N}-p^N_{j+1}-q^N_{j+1}-r^N_{j+1}\Big\|$$

$$+\Big\|\Big(\frac{A^*p^N_{j+N+1}-A^*p^N_{j+N}}{h_N},\ \frac{A^*q^N_{j+1}-A^*q^N_j}{h_N},\ 0\Big)\Big\|.$$

This implies, due to the mentioned coderivative criterion and the uniform boundedness of p^N_j and q^N_j from above (by some constant $\alpha>0$), that

$$\Big\|\Big(\frac{p^N_{j+1}-p^N_j}{h_N}-\lambda^N u^*_j,\ \frac{q^N_{j-N+1}-q^N_{j-N}}{h_N}-\lambda^N y^*_{j-N},\ \frac{r^N_{j+1}-r^N_j}{h_N}-\lambda^N z^*_j\Big)\Big\|$$

$$\leq \ell_F\Big\|\lambda^N v^*_j+\frac{\lambda^N\xi^N_j}{h_N}-p^N_{j+1}-q^N_{j+1}-r^N_{j+1}\Big\|+\frac{\alpha}{h_N}$$

for $j=k-2N+2,\ldots,k-N+1$. Therefore we have the estimates

$$\|(p^N_j,q^N_{j-N},r^N_j)\|\leq \ell_F\|\xi^N_j\|+(\ell_F+1)h_N\ell_\vartheta$$

$$+(\ell_F h_N+1)\|(p^N_{j+1},q^N_{j-N+1},r^N_{j+1})\|$$

$$+(\ell_F h_N+1)\alpha\leq \ell_F\|\xi^N_j\|+(\ell_F h_N+1)\ell_F\|\xi^N_{j+1}\|$$

$$+(\ell_F+1)h_N\ell_\vartheta$$

$$+(\ell_F h_N+1)(\ell_F+1)h_N\ell_\vartheta+(\ell_F h_N+1)(\ell_F+1)\alpha+(\ell_F h_N+1)^2$$

$$\|(p^N_{j+2},q^N_{j-N+2},r^N_{j+2})\|$$

$$\leq\ldots\leq \exp\big(\ell_F(b-a)\big)\Big(1+\frac{(\ell_\vartheta+\alpha)(\ell_F+1)}{\ell_F}+\ell_F\nu_N\Big)$$

whenever $j=k-2N+2,\ldots,k-N+1$. This shows that p^N_j, q^N_{j-N}, and r^N_j are uniformly bounded for $j=k-2N+2,\ldots,k-N+1$, and hence the sequence $\{p^N(t),q^N(t-\theta),r^N(t)\}$ is uniformly bounded on $[b-2\theta,b-\theta]$. Repeating the above procedure, we conclude that both sequences $\{p^N(t),q^N(t-\theta),r^N(t)\}$ and $\{\widetilde{p}^N(t),\widetilde{q}^N(t-\theta)\}$ are uniformly bounded on the whole interval $[a,b]$.

Next we estimate $\big(\dot{\widetilde{p}}^N(t),\dot{\widetilde{q}}^N(t-\theta),\dot{r}^N(t)\big)$ on $[a,b]$ using the discrete Euler-Lagrange inclusion and the coderivative characterization of the local Lipschitzian property. This yields, for $t_j\leq t<t_{j+1}$ with $j=0,\ldots,k$, that

$$\|(\dot{\widetilde{p}}^N(t), \dot{\widetilde{q}}^N(t-\theta), \dot{r}^N(t))\| = \Big\|\Big(\frac{\widetilde{p}_{j+1}-\widetilde{p}_j}{h_N}, \frac{\widetilde{q}_{j-N+1}-\widetilde{q}_{j-N}}{h_N}, \frac{r^N_{j+1}-r^N_j}{h_N}\Big)\Big\|$$

$$\le \ell_F\Big\|\lambda^N v^*_j + \frac{\lambda^N \xi^N_j}{h_N} - p^N_{j+1} - q^N_{j+1} - r^N_{j+1}\Big\| + \ell_\vartheta$$

$$\le \ell_F\|\xi^N\| + \ell_F\|p^N_{j+1}\| + \ell_F\|q^N_{j+1}\| + \ell_F\|r^N_{j+1}\| + (\ell_F+1)\ell_\vartheta\ .$$

Thus the sequence $\{\dot{\widetilde{p}}^N(t), \dot{\widetilde{q}}^N(t-\theta), \dot{r}^N(t)\}$ is weakly compact in $L^1[a,b]$. Taking the whole sequence of $N \in I\!N$ without loss of generality, we find three absolutely continuous mappings $\widetilde{p}(\cdot)$, $\widetilde{q}(\cdot-\theta)$, and $r(\cdot)$ on $[a,b]$ such that

$$\dot{\widetilde{p}}^N(t) \to \dot{\widetilde{p}}(t), \quad \dot{\widetilde{q}}^N(t-\theta) \to \dot{\widetilde{q}}(t-\theta), \quad \dot{r}^N(t) \to \dot{r}(t) \ \text{ weakly in } \ L^1[a,b]$$

and $\widetilde{p}^N(t) \to \widetilde{p}(t)$, $\widetilde{q}^N(t-\theta) \to \widetilde{q}(t-\theta)$, $r^N(t) \to r(t)$ uniformly on $[a,b]$ as $N \to \infty$. Since $p^N(t)$ and $q^N(t-\theta)$ are uniformly bounded on $[a,b+\theta]$, they surely converge to some arcs $p(t)$ and $q(t-\theta)$ weakly in $L^1[a,b+\theta]$. Taking into account the above convergence of $\widetilde{p}^N(t)$ and $\widetilde{q}^N(t-\theta)$, we get from (7.16) that $p(\cdot)$ and $q(\cdot)$ satisfy (7.21), that

$$\widetilde{p}(t) = p(t) + A^*p(t+\theta), \quad \widetilde{q}(t-\theta) = q(t-\theta) + A^*q(t), \quad t \in [a,b]\ ,$$

and that $p(t)$ and $q(t)$ are piecewise continuous on $[a,b+\theta]$ and $[a-\theta,b]$, respectively, with possible discontinuity (from the right) at the points $b-i\theta$ at $i=0,1,\ldots$. Conditions (7.20) and (7.22) follow by passing to the limit from (7.19) and (7.17), respectively, by taking into account the robustness of the basic normal cone and subdifferential in finite dimensions.

It remains to justify the extended Euler-Lagrange inclusion in this theorem. To proceed, we rewrite the discrete Euler-Lagrange inclusion of Theorem 7.3 in the form

$$\begin{aligned}
&(\dot{\widetilde{p}}^N(t), \dot{\widetilde{q}}^N(t-\theta), \dot{r}^N(t)) \\
&\in \Big\{(u,v,w)\Big|\ \Big(u,v,w,p^N(t_{j+1}) + q^N(t_{j+1}) + r^N(t_{j+1}) - \frac{\lambda^N\xi^N_j}{h_N}\Big) \\
&\in \lambda^N\partial\vartheta\big(\bar{x}^N(t_j), \bar{x}^N(t_j-\theta), \bar{z}^N(t_j), \bar{v}^N_j\big) \\
&+N\big((\bar{x}^N(t_j), \bar{x}^N(t_j-\theta), \bar{z}^N(t_j), \bar{v}^N_j); \operatorname{gph} F(t_j)\big)\Big\}
\end{aligned} \tag{7.23}$$

for $t \in [t_j, t_{j+1}]$ with $j=0,\ldots,k$. By the classical Mazur theorem there is a sequence of convex combinations of $(\dot{\widetilde{p}}^N(t), \dot{\widetilde{q}}^N(t-\theta), \dot{r}^N(t))$ that converges to $(\dot{\widetilde{p}}(t), \dot{\widetilde{q}}(t-\theta), \dot{r}(t))$ for a.e. $t \in [a,b]$. Passing to the limit in (7.23) and taking into account the pointwise convergence of $\xi^N(t)$ and $\bar{v}^N(t)$ established above as well as the constructions of the extended normal cone and subdifferential and their robustness property with respect to all variables and parameters, we arrive at the required Euler-Lagrange inclusion for problem (DA) and complete the proof of the theorem. △

Observe that for the *Mayer problem* (DA_M), which is (DA) with $\vartheta = 0$, the generalized Euler-Lagrange inclusion of Theorem 7.5 is equivalently expressed in terms of the *extended coderivative* for moving (in $t \in T$) set-valued mapping $S\colon X \times T \rightrightarrows Y$ at $(\bar{x}, \bar{y}, \bar{t})$ with $\bar{y} \in S(\bar{x}, \bar{t})$ defined by

$$D^*_+ S(\bar{x}, \bar{y}, \bar{t})(y^*) := \big\{x^* \in X^* \big|\ (x^*, -y^*) \in N_+\big((\bar{x}, \bar{y}); \operatorname{gph} S(\cdot, \bar{t})\big)\big\},\ \ y^* \in Y^* .$$

Indeed, it can be written in the form

$$\Big(\frac{d}{dt}\big[p(t) + A^* p(t+\theta)\big],\ \frac{d}{dt}\big[q(t-\theta) + A^* q(t)\big], \dot{r}(t)\Big)$$

$$\in \operatorname{co} D^*_+ F\big(\bar{x}(t), \bar{x}(t-\theta), \bar{z}(t), \dot{\bar{z}}(t)\big)\big(-p(t) - q(t) - r(t)\big) \ \text{ a.e. } \ t \in [a, b] .$$

via the extended coderivative of F with respect to the variables (x, y, z), where $t \in [a, b]$ is considered as a moving parameter.

It turns out that the extended Euler-Lagrange inclusion obtained above implies, under the *relaxation stability* of the original problems, two other principal optimality conditions expressed in terms of the Hamiltonian function built upon the velocity mapping F. The first condition called the *extended Hamiltonian inclusion* is given below in terms of a *partial convexification* of the basic subdifferential for the Hamiltonian function. The second one is an analog of the classical *Weierstrass-Pontryagin maximum condition* for the differential-algebraic inclusions under consideration. Recall that an analog of the maximum principle (centered around the maximum condition) *doesn't generally hold* for differential-algebraic systems, even in the case of optimal control problems governed by smooth functional-differential equations of neutral type that are a special case of (DA).

As in the case of ordinary differential inclusions in finite-dimensions (cf. Remark 6.32), the following relationships between the extended Euler-Lagrange and Hamiltonian inclusions are based on Rockafellar's dualization theorem that concerns subgradients of abstract Lagrangian and Hamiltonian associated with set-valued mappings *regardless of the dynamics*. For simplicity we consider the Mayer problem (DA_M) for autonomous differential-algebraic systems. Then the *Hamiltonian* function for the mapping F is

$$\mathcal{H}(x, y, z, p) := \sup\big\{\langle p, v\rangle \big|\ v \in F(x, y, z)\big\} .$$

Corollary 7.6 (extended Hamiltonian inclusion and maximum condition for differential-algebraic inclusions). *Let* $\{\bar{x}(\cdot), \bar{z}(\cdot)\}$ *be an optimal solution to the Mayer problem* (DA_M) *for autonomous delayed differential-algebraic systems under the assumptions of Theorem* 7.5. *Then there exist a number* $\lambda \geq 0$, *piecewise continuous arcs* $p\colon [a, b+\theta] \to I\!R^n$ *and*

$q\colon[a-\theta,b]\to I\!R^n$ (whose points of discontinuity are confined to multiples of the delay time θ), and an absolutely continuous arc $r\colon[a,b]\to I\!R^n$ such that $p(t)+A^*p(t+\theta)$ and $q(t-\theta)+A^*q(t)$ are absolutely continuous on $[a,b]$ and, besides (7.20)–(7.22), *one has the extended Hamiltonian inclusion*

$$\begin{aligned}&\Big(\frac{d}{dt}\big[p(t)+A^*p(t+\theta)\big],\ \frac{d}{dt}\big[q(t-\theta)+A^*q(t)\big],\dot r(t)\Big)\in\operatorname{co}\Big\{(u,v,w)\ \Big|\\&\Big(-u,-v,-w,\dot{\bar z}(t)\Big)\in\partial\mathcal{H}\big(\bar x(t),\bar x(t-\theta),\bar z(t),p(t)+q(t)+r(t)\big)\Big\}\end{aligned}\tag{7.24}$$

and the maximum condition

$$\big\langle p(t)+q(t)+r(t),\dot{\bar z}(t)\big\rangle=\mathcal{H}\big(\bar x(t),\bar x(t-\theta),\bar z(t),p(t)+q(t)+r(t)\big)\tag{7.25}$$

for a.e. $t\in[a,b]$. If moreover F is convex-valued around $\big(\bar x(t),\bar x(t-\theta),\bar z(t)\big)$, then (7.24) *is equivalent to the Euler-Lagrange inclusion*

$$\begin{aligned}&\Big(\frac{d}{dt}\big[p(t)+A^*p(t+\theta)\big],\ \frac{d}{dt}\big[q(t-\theta)+A^*q(t)\big],\dot r(t)\Big)\\&\in\operatorname{co}D^*F\big(\bar x(t),\bar x(t-\theta),\bar z(t),\dot{\bar z}(t)\big)\big(-p(t)-q(t)-r(t)\big)\end{aligned}\tag{7.26}$$

for a.e. $t\in[a,b]$, which automatically implies the maximum condition (7.25) *in the case under consideration.*

Proof. Since (DA_M) is stable with respect to relaxation, the pair $\{\bar x(\cdot),\bar z(\cdot)\}$ is an optimal solution to the relaxed problem $(\overline{DA_M})$ whose only difference from (DA_M) is that the original delayed differential-algebraic inclusion is replaced by its convexification (7.10). By Theorem 7.5 the optimal solution $\{\bar x(\cdot),\bar z(\cdot)\}$ satisfies conditions (7.20)–(7.22) and the relaxed counterpart of the Euler-Lagrange inclusion (7.26) with the replacement of F by its convex hull $\operatorname{co}F$. According to Rockafellar's dualization theorem we have

$$\begin{aligned}&\operatorname{co}\Big\{(u,v,w)\Big|\ (u,v,w,p)\in N\big((x,y,z,q);\operatorname{gph}(\operatorname{co}F)\big)\Big\}\\&=\operatorname{co}\Big\{(u,v,w)\Big|\ (-u,-v,-w,q)\in\partial\overline{\mathcal{H}}(x,y,z,p)\Big\}\,,\end{aligned}$$

where $\overline{\mathcal{H}}$ stands for the Hamiltonian of the relaxed system, i.e., with F replaced by $\operatorname{co}F$. It is easy to check that $\overline{\mathcal{H}}=\mathcal{H}$. Thus the extended Euler-Lagrange inclusion for the relaxed system implies the extended Hamiltonian inclusion (7.24), which surely yields the maximum condition (7.25). When F is *convex-valued*, (7.24) and (7.26) are *equivalent* due to the above dualization equality. Note that, by Theorem 1.34, the Euler-Lagrange inclusion (7.26) implies the maximum condition (7.25) when F is convex-valued. This also happens in the case of relaxation stability with adjoint arcs $\big(p(\cdot),q(\cdot),r(\cdot)\big)$ satisfying the Euler-Lagrange inclusion in the relaxed problem. △

Remark 7.7 (optimal control of delay-differential inclusions). The results obtained can be specified and simplified in the case of optimal control problems governed by *delay-differential inclusions* of the type

$$\dot{x}(t) \in F\big(x((t), x(t-\theta), t\big) \text{ a.e. } t \in [a, b]$$

containing time delays only in state variables. Such systems are actually closer to ordinary differential inclusions than to the differential-algebraic and neutral systems considered in this section. A remarkable specific feature of delay-differential inclusions in comparison with both ordinary and differential-algebraic/neutral ones is that they admit valuable results in the case of *set-valued tail constraints*

$$x(t) \in C(t) \text{ a.e. } t \in [a-\theta, a)$$

given on the initial time interval that provide an additional source for optimization; see Mordukhovich and L. Wang [973] for more details. Furthermore, the approximation procedure and necessary optimality conditions developed in Sect. 6.2 *with no relaxation* assumptions can be extended to the case of delay-differential systems without substantial changes in comparison with ordinary evolution inclusions.

It seems however that similar optimality conditions *cannot* be generally derived for differential-algebraic and neutral inclusions, i.e., when $A \neq 0$ in (DA). The major reason is that the approximation procedure developed in Sect. 6.3 and the results obtained therein are essentially based on the *automatic relaxation stability* of free-endpoint Bolza problems with finite integrands, which is not the case for problems containing delays in velocity variables and/or algebraic relations between state variables.

7.2 Neumann Boundary Control of Semilinear Constrained Hyperbolic Equations

In this section we study optimal control problems for a class of *semilinear hyperbolic equations* with controls acting in *Neumann boundary conditions* in the presence of *pointwise constraints* on control and state functions. It is well known that *state-constrained* control problems are among the most challenging and difficult in dynamic optimization. While such problems have been extensively studied for ordinary and time-delay control systems as well for partial differential equations of the elliptic and parabolic types, it is not the case for hyperbolic equations. In addition, *boundary control* problems happen to be substantially more involved in comparison with those containing control parameters in the body of differential equations, i.e., with problems involving the so-called *distributed controls.*

This section concerns Neumann boundary control problems for hyperbolic systems with state constraints; the corresponding problems with controls in

Dirichlet boundary conditions (which are substantially different from the Neumann ones) are studied in the next section. The main goal here is to establish *necessary optimality conditions* for a state-constrained Neumann boundary control problem governed by the semilinear wave equation that will be established in the *pointwise maximum principle* form under rather mild and natural assumptions. Our approach to derive necessary optimality conditions for this problem is based on *perturbation methods* of variational analysis involving some *penalization* of state constraints and then the passage to the limit from necessary conditions in unconstrained approximating problems; cf. Sect. 6.2 for the case of evolution inclusions. The analysis of approximating control problems for unconstrained hyperbolic equations in this section is however different from the one in Sect. 6.2: it is based on *needle-type variations* as in Sect. 6.3 for ordinary control systems. Details follow.

7.2.1 Problem Formulation and Necessary Optimality Conditions for Neumann Boundary Controls

Given an open bounded set (domain) $\Omega \subset I\!R^n$ with a boundary Γ of class $\mathcal{C}^2$ and given a positive number (time) T, we mainly concern the following optimal control problem governed by the *semilinear wave equation*: minimize

$$J(y,u) = \int_\Omega f\big(x, y(T)\big)dx + \int_Q g(x,t,y)\,dxdt + \int_\Sigma h(s,t,u)\,dsdt$$

over admissible pairs $\{y(\cdot), u(\cdot)\}$ satisfying

$$\begin{cases} y_{tt} - \Delta y + \vartheta(\,\cdot\,, y) = 0 \ \text{ in } Q := \ \Omega \times (0,T)\,, \\ \partial_\nu y = u \qquad\qquad\qquad\quad \text{ in } \Sigma := \ \Gamma \times (0,T)\,, \\ y(0) = y_0, \quad y_t(0) = y_1 \ \text{ in } \Omega \end{cases} \tag{7.27}$$

under the *pointwise constraints* on control and state functions

$$u(\cdot) \in U_{ad} \subset L^2(\Sigma), \quad y(\cdot) \in \Theta \subset \mathcal{C}\big([0,T]; L^2(\Omega)\big)\,,$$

where the operator Δ stands for the classical *Laplacian*, and where ∂_ν stands for the usual *normal derivative* at the boundary. Denote this problem by (NP) and shortly write it as follows:

$$\inf\Big\{ J(y,u) \ \Big| \ \{y(\cdot),u(\cdot)\} \ \text{ satisfies } \ (7.27), \ \ u(\cdot) \in U_{ad}, \ \ y(\cdot) \in \Theta \Big\}\,.$$

Assumptions on the nonlinear function ϑ as well as on the integrands f, g, and h are presented and discussed below. The initial state $(y_0, y_1) \in H^1(\Omega) \times L^2(\Omega)$ is fixed. Note that the main constructions and results of this section can be extended to hyperbolic equations governed by more general *strongly elliptic operators* in (7.27)—not just by the Laplacian Δ—with *time-independent and regular* (in the usual PDE sense) coefficients.

Throughout this and the next sections we use standard notation conventional in the PDE control literature. For the reader's convenience, recall that $\mathcal{M}\big([0,T];L^2(\Omega)\big)$ is the space of measures on $[0,T]$ with values in $L^2(\Omega)$, which is the topological dual of $\mathcal{C}\big([0,T];L^2(\Omega)\big)$. The topological dual of

$$\mathcal{C}_0\big(]0,T];L^2(\Omega)\big) := \big\{y\in\mathcal{C}\big([0,T];L^2(\Omega)\big)\ \big|\ y(0)=0\big\}$$

is denoted by $\mathcal{M}_b\big(]0,T];L^2(\Omega)\big)$ and, similarly, the topological dual of

$$\mathcal{C}_0(]0,T[;L^2(\Omega)) := \big\{y\in\mathcal{C}\big([0,T];L^2(\Omega)\big)\ \big|\ y(0)=0,\ \ y(T)=0\big\}$$

is denoted by $\mathcal{M}_b\big(]0,T[;L^2(\Omega)\big)$. Observe that the spaces $\mathcal{C}_0\big(]0,T];L^2(\Omega)\big)$ and $\mathcal{C}_0\big(]0,T[;L^2(\Omega)\big)$ consist of continuous mappings on the closed interval $[0,T]$ with the prescribed one or both endpoints. In what follows we identify $]0,T]$ and $]0,T[$ with $(0,T]$ and $(0,T)$, respectively.

It is well known that every measure $\mu\in\mathcal{M}_b\big(]0,T];L^2(\Omega)\big)$ can be identified with a measure $\widetilde{\mu}\in\mathcal{M}\big([0,T];L^2(\Omega)\big)$ such that $\widetilde{\mu}(\{0\})=0$ and $\widetilde{\mu}|_{]0,T]}=\mu$, where $\widetilde{\mu}|_{]0,T]}$ denotes the restriction of $\widetilde{\mu}$ to $]0,T]$. Therefore, if $y\in\mathcal{C}\big([0,T];L^2(\Omega)\big)$ and $\mu\in\mathcal{M}_b\big(]0,T];L^2(\Omega)\big)$, we still use the notation

$$\langle y,\mu\rangle_{\mathcal{C}([0,T];L^2(\Omega)),\mathcal{M}_b(]0,T];L^2(\Omega))}\quad\text{for}\quad\langle y,\widetilde{\mu}\rangle_{\mathcal{C}([0,T];L^2(\Omega)),\mathcal{M}([0,T];L^2(\Omega))}\ .$$

Since we have to deal with equations satisfied in the sense of *distributions* in Q, it is also convenient to identify $\mathcal{M}_b\big(]0,T];L^2(\Omega)\big)$ with a subspace of $\mathcal{M}_b\big(\Omega\times]0,T]\big)$; this identification follows from the continuous and dense imbedding $\mathcal{C}_0\big(\Omega\times]0,T]\big)\hookrightarrow\mathcal{C}_0\big(]0,T];L^2(\Omega)\big)$. Thus for $\mu\in\mathcal{M}_b\big(]0,T];L^2(\Omega)\big)$ the notation $\mu|_Q$ —the restriction of μ to Q—is meaningful if μ is considered as a bounded measure on $\Omega\times]0,T]=\Omega\times(0,T]$, and so $\mu|_{\Omega\times\{T\}}$ stands for $\mu(\{T\})$. The same kind of notation is used below in similar settings. For $z\in L^2(Q)$ we denote by z_t (respectively by z_{tt}) the derivative (respectively the second derivative) of z in t in the sense of distributions in Q.

Given a Banach space Z, the duality pairing between Z and Z^* is denoted by $\langle\cdot,\cdot\rangle_{Z,Z^*}$. When there is no ambiguity, we sometimes write $\langle\cdot,\cdot\rangle$ instead of $\langle\cdot,\cdot\rangle_{Z,Z^*}$. To emphasize a specific kind of regularity of solutions to the hyperbolic equations under considerations, we may write, e.g., that $(y,y_t)\in\mathcal{C}\big([0,T];X\big)\times\mathcal{C}\big([0,T];Y\big)$ is a solution to (7.27) instead of just indicating that y is a solution to this system.

If $p(\cdot)$ belongs to $BV\big([0,T];H^1(\Omega)^*\big)$, the space of functions of bounded variation on $[0,T]$ with values in $H^1(\Omega)^*$, one can define $p(t^-)$ and $p(t^+)$ for every $t\in(0,T)$ and also $p(0^+)$ and $p(T^-)$, while the values $p(0)$ and $p(T)$ may be generally different from $p(0^+)$ and $p(T^-)$. There is a unique Radon measure on $[0,T]$ with values in $H^1(\Omega)^*$, denoted by d_tp, such that the restriction of $d_t p$ to $(0,T)$ is the vector-valued distributional derivative of p in $(0,T)$ with $d_t p(\{0\}) = p(0^+) - p(0)$ and $d_t p(\{T\}) = p(T) - p(T^-)$. Moreover, identifying p with its representative right-hand side continuous in $(0,T)$, we have

$$p(0^+) = p(0) + d_t p(\{0\}) \ \text{ and } \ p(t) = p(0) + d_t p\big([0,t]\big) \ \text{ for every } \ t \in]0,T] \ .$$

Recall that if $\{p_k\}$ is a bounded sequence in $BV\big([0,T]; H^1(\Omega)^*\big)$, then there is a subsequence $\{p_{k_m}\}$ and a function $p \in BV\big([0,T]; H^1(\Omega)^*\big)$ such that

$$p_{k_m}(t) \to p(t) \quad \text{weakly in} \quad H^1(\Omega)^* \quad \text{for almost every} \quad t \in [0,T] \ .$$

Note that this convergence may hold for every $t \in [0,T]$ if the above representative right-hand side continuous in $(0,T)$ is not specified; see, e.g., Barbu and Precupanu [84] for more details. In particular,

$$p_{k_m}(T) \to p(T) \ \text{ weakly in } \ H^1(\Omega)^* \ \text{ as } \ m \to \infty \ .$$

Now let us formulate the *standing assumptions* on the initial data of problem (NP) that are needed throughout this paper.

(H1) For every $y \in I\!R$ the function $\vartheta(\cdot,\cdot,y)$ is measurable in Q; for a.e. pairs $(x,t) \in Q$ the function $\vartheta(x,t,\cdot)$ is of class $\mathcal{C}^1$. Moreover, one has

$$\vartheta(\cdot,0) \in L^1\big(0,T; L^2(\Omega)\big), \quad |\vartheta'_y(x,t,y)| \le M \ \text{ in } Q \times I\!R \ \text{ with } \ M > 0 \ ,$$

where ϑ'_y stands for the partial derivative.

(H2) For every $y \in I\!R$ the function $f(\cdot,y)$ is measurable on Ω with $f(\cdot,0)$ belonging to $L^1(\Omega)$. For a.e. $x \in \Omega$ the function $f(x,\cdot)$ is of class $\mathcal{C}^1$. Moreover, there is a constant $C > 0$ such that

$$|f'_y(x,y)| \le C\big(1+|y|\big) \quad \text{whenever} \quad (x,y) \in \Omega \times I\!R \ .$$

(H3) For every $y \in I\!R$ the function $g(\cdot,\cdot,y)$ is measurable on Q with $g(\cdot,0)$ belonging to $L^1(Q)$. For a.e. $(x,t) \in Q$ the function $g(x,t,\cdot)$ is of class $\mathcal{C}^1$. Moreover, there is a constant $C > 0$ such that

$$|g'_y(x,t,y)| \le C\big(1+|y|\big) \quad \text{whenever} \quad (x,t,y) \in Q \times I\!R \ .$$

(H4) For every $u \in I\!R$ the function $h(\cdot,\cdot,u)$ is measurable on Σ with $h(\cdot,0)$ belonging to $L^1(\Sigma)$. For a.e. $(s,t) \in \Sigma$, $h(s,t,\cdot)$ is of class $\mathcal{C}^1$. Moreover, there is a constant $C > 0$ such that

$$|h'_u(s,t,u)| \le C\big(1+|u|\big) \quad \text{whenever} \quad (s,t,u) \in \Sigma \times I\!R \ .$$

(H5) The state constraint set $\Theta \subset \mathcal{C}\big([0,T]; L^2(\Omega)\big)$ is closed and convex with $\text{int}\,\Theta \neq \emptyset$. Furthermore, we suppose that the initial state function $\widehat{y}_0(x,t) := y_0(x)$ belongs to the interior of Θ.

(H6) The control set U_{ad} is given in the form

$$U_{ad} := \left\{ u \in L^2(\Sigma) \;\middle|\; u(s,t) \in K(s,t) \;\text{ a.e. }\; (s,t) \in \Sigma \right\},$$

where $K(\cdot)$ is a measurable multifunction whose values are nonempty and closed subsets of $I\!R$.

Of course, we suppose as usual that the set of *feasible pairs* $\{y(\cdot), u(\cdot)\}$ to (P) is *nonempty*, i.e., there is $u(\cdot) \in U_{ad}$ such that $J(y,u) < \infty$, where $y(\cdot) \in \Theta$ is a *weak solution* of system (7.27) corresponding to u; see the next subsection for the precise definition.

Observe that the above basic assumptions *don't* impose any *convexity* requirements on the integrands in the cost functional with respect to either state or control variables, as well as on the control set U_{ad}. This is different from the Dirichlet boundary control setting considered in Sect. 7.3. The reason is that the Neumann boundary value problem offers *more regularity* in comparison with the Dirichlet one and allows us to employ powerful variational methods to prove necessary optimality conditions that *don't rely on weak convergences*; see more discussion in Sect. 7.3.

To formulate the main result of this section, let us define the (analog of) *Hamilton-Pontryagin* function

$$H(s,t,u,p,\lambda) := pu + \lambda h(s,t,u)$$

for the control problem (NP). The following theorem gives necessary conditions for optimal solutions to (NP), which provide a version of the *Pontryagin maximum principle* in pointwise form for the Neumann boundary control problem under consideration. It is more convenient for us to formulate this result with the *minimum* (but not maximum) condition.

Theorem 7.8 (pointwise necessary optimality conditions for Neumann boundary controls). *Let* $\{\bar y(\cdot), \bar u(\cdot)\}$ *be an optimal solution to problem* (NP) *satisfying assumptions* (H1)–(H6). *Then there exist* $\lambda \ge 0$, $\mu \in \mathcal{M}_b\big(]0,T]; L^2(\Omega)\big)$, *and a measurable subset* $\widetilde\Sigma \subset \Sigma$ *such that* $\mathcal{L}^n(\Sigma \setminus \widetilde\Sigma) = 0$,

$$(\lambda,\mu) \ne 0, \quad \langle \mu, z - \bar y\rangle \le 0 \;\; \textit{for all}\, z \in \Theta, \quad \textit{and} \tag{7.28}$$

$$H\big(s,t,\bar u(s,t), p(s,t), \lambda\big) = \min_{u \in K(s,t)} H\big(s,t,u,p(s,t),\lambda\big) \tag{7.29}$$

for all $(s,t) \in \widetilde\Sigma$, *where* $\mathcal{L}^n$ *denotes the* n*-dimensional Lebesgue measure, and where* $p(\cdot)$ *is the corresponding solution to the adjoint system*

$$\begin{cases} p_{tt} - \Delta p + \vartheta'_y(\cdot,\bar y)p = \lambda g'_y(x,t,\bar y) + \mu|_Q & \textit{in} \quad Q\,, \\ \partial_\nu p = 0 & \textit{in} \quad \Sigma\,, \\ p(T) = y_0, \quad p_t(T) = -\lambda f'_y(x,\bar y(T)\big) - \mu|_{\Omega\times\{T\}} & \textit{in} \quad \Omega\,. \end{cases} \tag{7.30}$$

The proof of Theorem 7.8 is given in Subsect. 7.2.4. The definitions of solutions to the state and adjoint systems in this theorem are formulated and discussed in the next subsection.

7.2.2 Analysis of State and Adjoint Systems in the Neumann Problem

Let us start with the classical nonhomogeneous Neumann boundary value problem for the *linear wave equation*

$$\begin{cases} y_{tt} - \Delta y = \phi & \text{in } Q\,, \\ \partial_\nu y = u & \text{in } \Sigma\,, \\ y(0) = y_0, \quad y_t(0) = y_1 & \text{in } \Omega\,. \end{cases} \tag{7.31}$$

The following fundamental *regularity* result is established by Lasiecka and Triggiani [744, 745]; we refer the reader to the original papers for the (hard) proof, discussions, and PDE applications. Our goal is to incorporate this result in the framework of *variational analysis* of the Neumann boundary *control* problem under consideration. A significant part of our analysis, provided in this subsection, concerns the study of the hyperbolic state system (7.27) with Neumann boundary controls and the corresponding adjoint system.

Lemma 7.9 (basic regularity for the hyperbolic linear Neumann problem). *Assume that* $(\phi, u, y_0, y_1) \in L^1\big(0,T;L^2(\Omega)\big) \times L^2(\Sigma) \times H^1(\Omega) \times L^2(\Omega)$*, and let* $y(\phi,u,y_0,y_1) \in \mathcal{C}\big([0,T];L^2(\Omega)\big) \cap \mathcal{C}^1\big([0,T];H^1(\Omega)^*\big)$ *be the unique weak solution to the linear Neumann boundary value problem* (7.31)*. Then the mapping* $u \mapsto y(0,u,0,0)$ *is bounded from* $L^2(\Sigma)$ *to* $\mathcal{C}\big([0,T];H^{1/2}(\Omega)\big) \cap \mathcal{C}^1\big([0,T];H^{1/2}(\Omega)^*\big)$*, and it is also bounded from* $L^2(\Sigma)$ *to* $H^{3/5-\varepsilon}(Q)$ *for all* $\varepsilon > 0$*. Furthermore, the mapping* $(\phi,y_0,y_1) \mapsto y(\phi,0,y_0,y_1)$ *is bounded from* $L^1\big(0,T;L^2(\Omega)\big) \times H^1(\Omega) \times L^2(\Omega)$ *to* $\mathcal{C}\big([0,T];H^1(\Omega)\big) \cap \mathcal{C}^1\big([0,T];L^2(\Omega)\big)$*.*

Next consider the nonhomogeneous Neumann boundary value problem for the linear wave equation with possibly *nonsmooth data*:

$$\begin{cases} y_{tt} - \Delta y + \theta y = \phi & \text{in } Q\,, \\ \partial_\nu y = u & \text{in } \Sigma\,, \\ y(0) = y_0, \quad y_t(0) = y_1 & \text{in } \Omega\,, \end{cases} \tag{7.32}$$

where the nonsmooth coefficient $\theta(x,t)$ belongs to $L^\infty(Q)$. The following estimate of weak solutions to the *homogeneous* linear Neumann boundary value problem in (7.32) is needed in the sequel.

Lemma 7.10 (solution estimate for the nonsmooth linear Neumann problem in the homogeneous case). *Assume that* $u = 0$ *and that the initial data* (ϕ, y_0, y_1) *belong to* $L^1\big(0,T;L^2(\Omega)\big) \times H^1(\Omega) \times L^2(\Omega)$*. Then the homogeneous Neumann problem in* (7.32) *admits a unique weak solution in* $\mathcal{C}\big([0,T];L^2(\Omega)\big) \cap \mathcal{C}^1\big([0,T];H^1(\Omega)\big)$*. This solution satisfies the estimate*

$$\|y\|_{\mathcal{C}([0,T];H^1(\Omega))} + \|y_t\|_{\mathcal{C}([0,T];L^2(\Omega))} \le C\big(\|\phi\|_{L^1(0,T;L^2(\Omega))}$$
$$+\|y_0\|_{H^1(\Omega)} + \|y_1\|_{L^2(\Omega)}\big)\ ,$$

where the constant $C > 0$ *may depend on* $\|\theta\|_{L^\infty(Q)}$ *and* $\|\phi\|_{L^1(0,T;L^2(\Omega))}$*, but it is invariant with respect to all* $\theta(x,t)$ *having the same* $L^\infty(Q)$*-norm.*

Proof. The proof is standard. It is sufficient to multiply the first equation in (7.32) by y_t, to integrate it over Ω, and then to use the classical Gronwall lemma; see, e.g., Lions' book [791] for more details. △

The next lemma establishes an important *compactness property* of the control–weak solution operator in the nonsmooth and *nonhomogeneous* linear Neumann problem formulated in (7.32).

Lemma 7.11 (compactness of weak solutions to the nonsmooth linear Neumann problem in the nonhomogeneous case). *Assume that* $(\phi, y_0, y_1) = (0,0,0)$ *and that* $u \in L^2(\Sigma)$*. Then the nonhomogeneous Neumann problem in* (7.32) *admits a unique weak solution* $y(u)$ *belonging to* $\mathcal{C}\big([0,T];L^2(\Omega)\big) \cap \mathcal{C}^1\big([0,T];H^1(\Omega)^*\big)$ *and such that the solution mapping* $u \mapsto (y(u), y_t(u))$ *is a bounded operator from* $L^2(\Sigma)$ *into the product space* $\mathcal{C}\big([0,T];H^{1/2}(\Omega)\big) \times \mathcal{C}\big([0,T];H^{1/2}(\Omega)\big)$*. Furthermore, the mapping* $u \mapsto y(u)$ *is a compact operator from* $L^2(\Sigma)$ *into* $\mathcal{C}\big([0,T];L^2(\Omega)\big)$*.*

Proof. The existence and uniqueness of the weak solution to (7.32) can be deduced from the well-known result for the linear system (7.31) by using the standard *fixed-point method* in $L^2\big(0,\bar t;L^2(\Omega)\big)$ as $\bar t$ is sufficiently small and then by iterating the process m times with $m\bar t > T$. In this way we get

$$\|y\|_{\mathcal{C}([0,T];H^{1/2}(\Omega))} + \|y_t\|_{\mathcal{C}([0,T];H^{1/2}(\Omega))} \le C\|u\|_{L^2(\Sigma)}\ ,$$

where the constant $C > 0$ depends on an upper bound of the *norm* $\|\theta\|_{L^\infty(Q)}$ but not on the function $\theta(\cdot)$ itself. Now the compactness statement follows directly from the result by Simon [1212, Corollary 5]. △

Our next goal is to study the Neumann boundary value problem (7.27) for the original *semilinear* wave equation, which is labeled as the *state system* for convenience. First recall the notion of *weak solutions* to the *nonlinear* Neumann problem in (7.27) that is suitable to our study.

Definition 7.12 (weak solutions to the Neumann state system). *A function* $y(\cdot)$ *with* $(y, y_t) \in \mathcal{C}\big([0,T];L^2(\Omega)\big) \times \mathcal{C}\big([0,T];H^1(\Omega)^*\big)$ *is a* WEAK SOLUTION *to the state system* (7.27) *if*

$$\int_Q -\vartheta(\,\cdot\,,y)\,z\,dxdt = \int_Q y\,\varphi\,dxdt - \langle y_t(0), z(0)\rangle_{H^1(\Omega)^*,H^1(\Omega)}$$
$$+\int_\Omega y(0)z_t(0)\,dx + \int_\Sigma z\,u\,dsdt$$

for all $\varphi \in L^1\big(0,T;L^2(\Omega)\big)$, *where* $z(\cdot)$ *solves the homogeneous Neumann boundary value problem*

$$\begin{cases} z_{tt} - \Delta z = \varphi & \text{in } Q \ , \\ \partial_\nu z = 0 & \text{in } \Sigma \ , \\ z(T) = 0, \quad z_t(T) = 0 & \text{in } \Omega \ . \end{cases}$$

The advantage of the above definition is that it allows to establish the existence, uniqueness, and regularity of weak solutions to the original state system under the standing assumptions made in Subsect. 7.2.1.

Theorem 7.13 (existence, uniqueness, and regularity of weak solutions to the Neumann state system). *For every initial triple* $(u, y_0, y_1) \in L^2(\Sigma)\times H^1(\Omega)\times L^2(\Omega)$ *the state system* (7.27) *admits a unique weak solution* $y(\cdot)$ *with* $(y, y_t) \in \mathcal{C}\big([0,T];L^2(\Omega)\big) \times \mathcal{C}\big([0,T];H^1(\Omega)^*\big)$ *such that* (y, y_t) *also belongs to* $\mathcal{C}\big([0,T];H^{1/2}(\Omega)\big) \times \mathcal{C}\big([0,T];H^{1/2}(\Omega)^*\big)$ *and satisfies the estimate*

$$\|y\|_{\mathcal{C}([0,T];H^{1/2}(\Omega))} + \|y_t\|_{\mathcal{C}([0,T];H^{1/2}(\Omega)^*)} \le C\big(\|u\|_{L^2(\Sigma)} + \|y_0\|_{H^1(\Omega)} + \|y_1\|_{L^2(\Omega)} + 1\big)$$

with some constant $C > 0$*. Furthermore, the mapping* $(u, y_0, y_1) \mapsto y$ *is continuous from* $(u, y_0, y_1) \in L^2(\Sigma) \times H^1(\Omega) \times L^2(\Omega)$ *into* $\mathcal{C}\big([0,T];H^{1/2}(\Omega)\big) \cap \mathcal{C}^1\big([0,T];H^{1/2}(\Omega)^*\big)$.

Proof. The existence of weak solutions to the state system (7.27) in the space intersection $\mathcal{C}\big([0,\bar{t}];L^2(\Omega)\big)\cap\mathcal{C}^1\big([0,\bar{t}];H^1(\Omega)^*\big)$ with $\bar{t}$ sufficiently small can be obtained by the standard fixed-point method. Then assumption (H1) and the estimates in Lemmas 7.10 and 7.11 allow us to ensure the existence of solutions in the functional space stated in the theorem. The proof of uniqueness is also standard and is omitted for brevity. The estimate of (y, y_t) in $\mathcal{C}\big([0,T];H^{1/2}(\Omega)\big) \cap \mathcal{C}^1\big([0,T];H^{1/2}(\Omega)^*\big)$ follows from the estimate of y in $\mathcal{C}\big([0,T];L^2(\Omega)\big)$ due to the *basic regularity* of Lemma 7.9. To justify finally the continuity of the mapping $(u, y_0, y_1) \mapsto y$ from $(u, y_0, y_1) \in L^2(\Sigma) \times H^1(\Omega) \times L^2(\Omega)$ into $\mathcal{C}\big([0,T];H^{1/2}(\Omega)\big) \cap \mathcal{C}^1\big([0,T];H^{1/2}(\Omega)^*\big)$, we use again assumption (H1) and the corresponding estimates for the linearized system (7.32) given in Lemmas 7.10 and 7.11. △

Next we consider the (linearized) *adjoint system* to (7.27) given by

$$\begin{cases} p_{tt} - \Delta p + \theta p = \mu|_Q & \text{in } Q \ , \\ \partial_\nu p = 0 & \text{in } \Sigma \ , \\ p(T) = 0, \quad p_t(T) = -\mu|_{\Omega\times\{T\}} & \text{in } \Omega \ , \end{cases} \tag{7.33}$$

where $\mu \in \mathcal{M}_b\big(]0,T];L^2(\Omega)\big)$, where $\mu|_Q$ and $\mu|_{\Omega\times\{T\}}$ denote the restriction of μ to Q and to $\Omega \times \{T\}$, respectively, and where $\theta(x,t) \in L^\infty(Q)$ as in (7.32). In order to introduce and justify an appropriate definition of *weak* solutions to the adjoint system (7.33) with required *well-posedness* properties, we need the following lemma that is certainly of independent interest.

Lemma 7.14 (divergence formula). *The functional space*

$$W := \Big\{\mathbf{V} \in \big(L^2(Q)\big)^{n+1} \;\Big|\; \operatorname{div}(\mathbf{V}) \in \mathcal{M}_b\big(]0,T[;L^2(\Omega)\big)\Big\}$$

endowed with the norm

$$\|\mathbf{V}\|_W := \|\mathbf{V}\|_{(L^2(Q))^{n+1}} + \|\operatorname{div}(\mathbf{V})\|_{\mathcal{M}_b(]0,T[;L^2(\Omega))}$$

is a Banach space. Furthermore, there exists a unique continuous operator γ_{ν_Q} from W into $H^{-1/2}(\partial Q)$ satisfying

$$\gamma_{\nu_Q}(\mathbf{V}) = \gamma_0(\mathbf{V})\cdot \nu_Q \quad \textit{whenever} \quad \mathbf{V} \in \big(\mathcal{C}^1(\overline{Q})\big)^{n+1}$$

and such that the divergence formula

$$\int_Q \mathbf{V}\cdot\nabla\phi + \Big\langle \phi, \operatorname{div}(\mathbf{V})\Big\rangle_{\mathcal{C}([0,T];L^2(\Omega)),\mathcal{M}_b(]0,T[;L^2(\Omega))}$$

$$= \Big\langle \gamma_{\nu_Q}(\mathbf{V}), \gamma_0(\phi)\Big\rangle_{H^{-1/2}(\partial Q),H^{1/2}(\partial Q)}$$

holds for all $\phi \in H^1(Q)$, where ∂Q conventionally denotes the boundary of Q.

Proof. It is easy to check that the space W with the endowed norm is Banach. Let Λ be a continuous extension operator from $H^{1/2}(\partial Q)$ into $H^1(Q)$ that is a bounded linear operator from $H^{1/2}(\partial Q)$ into $H^1(Q)$ satisfying

$$\gamma_0\Lambda\varphi = \varphi \quad \text{for all } \; \varphi \in H^{1/2}(\partial Q)\,.$$

Taking $\mathbf{V} \in \big(\mathcal{C}^1(\overline{Q})\big)^{n+1}$, observe that the functional

$$\varphi \longmapsto \int_Q \mathbf{V}\cdot\nabla\Lambda\varphi + \Big\langle \Lambda\varphi, \operatorname{div}(\mathbf{V})\Big\rangle_{\mathcal{C}([0,T];L^2(\Omega)),\mathcal{M}_b(]0,T[;L^2(\Omega))}$$

is linear and bounded on $H^{1/2}(\partial Q)$. Denoting this functional by $\gamma_{\nu_Q}(\mathbf{V})$, we directly verify that

$$\gamma_{\nu_Q}(\mathbf{V}) = \gamma_0(\mathbf{V})\cdot\nu_Q$$

and that the divergence formula of the theorem is satisfied. This means that $\gamma_{\nu_Q}(\mathbf{V})$ doesn't depend on the extension operator Λ. Furthermore, one has

$$\Big|\int_Q \mathbf{V}\cdot\nabla\Lambda\varphi + \Big\langle \Lambda\varphi, \operatorname{div}(\mathbf{V})\Big\rangle_{\mathcal{C}([0,T];L^2(\Omega)),\mathcal{M}_b(]0,T[;L^2(\Omega))}\Big| \le C\|\varphi\|_{H^{1/2}(\partial Q)}\|\mathbf{V}\|_W\,,$$

which implies the estimate

$$\|\gamma_{\nu_Q}(\mathbf{V})\|_{H^{-1/2}(\partial Q)} \leq C\|\mathbf{V}\|_W \quad \text{for all} \;\; \mathbf{V} \in \big(\mathcal{C}^1(\overline{Q})\big)^{n+1} .$$

Since $\big(\mathcal{C}^1(\overline{Q})\big)^{n+1}$ is dense in W, the proof is complete. △

Next take $(p, p_t) \in L^2\big(0, T; H^1(\Omega)\big) \times L^2\big(0, T; L^2(\Omega)\big)$ and assume that the combination $p_{tt} - \Delta p$, calculated in the sense of distributions on Q, belongs to $\mathcal{M}_b\big(]0, T[; L^2(\Omega)\big)$. Employing Lemma 7.14, we define the *normal trace* on ∂Q of the vectorfield $(-\nabla p, p_t)$ as an element of $H^{-1/2}(\partial Q)$. Then

$$\|\gamma_{\nu_Q}(-\nabla p, p_t)\|_{H^{-1/2}(\partial Q)} \leq C\big(\|p\|_{L^2(0,T;H^1(\Omega))} + \|p_t\|_{L^2(Q)}$$

$$+\|p_{tt} - \Delta p\|_{\mathcal{M}_b(]0,T[;L^2(\Omega))}\big) ,$$

where the constant $C > 0$ is independent of p. Since $\Omega \times \{0\}$ is an open subset of ∂Q, the restriction of the operator $\gamma_{\nu_Q}(-\nabla p, p_t)$ to $\Omega \times \{0\}$ belongs to the space $H^{-1/2}(\Omega)$. Thus we get

$$\gamma_{\nu_Q}(-\nabla p, p_t)|_{\Omega\times\{0\}} = p_t(0) \in H^{-1/2}(\Omega) .$$

Note that this results can be improved. We are going to show in Theorem 7.16 that a properly defined solution $p(\cdot)$ to the adjoint system (3.3) actually has the property of $p_t(0) \in L^2(\Omega)$.

Now we are ready to introduce an appropriate notion of *weak solutions* to the adjoint system (7.33) and justify their basic properties.

Definition 7.15 (weak solutions to the Neumann adjoint system). *A function $p \in L^\infty\big(0, T; L^2(\Omega)\big)$ is a* WEAK SOLUTION *to* (7.33) *if*

$$\Big\langle y(\varphi), \mu\Big\rangle_{\mathcal{C}([0,T];L^2(\Omega))\times\mathcal{M}_b(]0,T];L^2(\Omega))} - \int_Q p\varphi\, dxdt = 0 \tag{7.34}$$

for all $\varphi \in L^1\big(0, T; L^2(\Omega)\big)$, where $y(\varphi)$ is the solution to

$$\begin{cases} y_{tt} - \Delta y + \vartheta y = \varphi & \text{in} \quad Q\ , \\ \partial_\nu y = 0 & \text{in} \quad \Sigma\ , \\ y(0) = 0, \quad y_t(0) = 0 & \text{in} \quad \Omega\ . \end{cases} \tag{7.35}$$

The next theorem establishes the existence, uniqueness, and regularity of weak solutions to the adjoint system (7.33) under the imposed standing assumptions. Note that $\mathcal{C}_w\big([0, T]; H^1(\Omega)\big)$ signifies the space of continuous functions from $[0, T]$ into $H^1(\Omega)$ endowed with the weak topology.

Theorem 7.16 (existence, uniqueness, and regularity of weak solutions to the Neumann adjoint system). *The adjoint system* (7.33) *admits, under the standing assumptions made, a unique weak solution* $p(\cdot)$ *such that* $(p, p_t) \in L^\infty\big(0, T; H^1(\Omega)\big) \times L^\infty\big(0, T; L^2(\Omega)\big)$,

$$p_t \in BV\big([0, T]; H^1(\Omega)^*\big), \quad p \in C_w\big([0, T]; H^1(\Omega)\big), \quad \textit{and}$$

$$p_t(\tau) \in L^2(\Omega) \quad \textit{whenever} \quad \tau \in \big\{t \in [0, T] \;\big|\; \mu(\{t\}) = 0\big\} ,$$

which imply that $p_t(0) \in L^2(\Omega)$. *Furthermore, one has the estimate*

$$\|p\|_{L^\infty(0,T;H^1(\Omega))} + \|p_t\|_{L^\infty(0,T;L^2(\Omega))} \le C\|\mu\|_{\mathcal{M}_b(]0,T];L^2(\Omega))} ,$$

where C *depends on* $\|\vartheta\|_{L^\infty(Q)}$ *but is invariant with respect to functions* $\vartheta(x, t)$ *having the same norm in the space* $L^\infty(Q)$.

Proof. Observe that $p = 0$ when the pair $(p, p_t) \in L^\infty\big(0, T; H^1(\Omega)\big) \times L^\infty\big(0, T; L^2(\Omega)\big)$ satisfies (7.34) with $\mu = 0$. This implies that the adjoint system (7.33) *cannot admit more than one* weak solution. To prove the *existence* of a weak solution, we develop an approximation procedure. First build a sequence $\{\mu_k\} \subset L^1\big(0, T; L^2(\Omega)\big)$ satisfying the properties

$$\begin{cases} \|\mu_k\|_{L^1(0,T;L^2(\Omega))} = \|\mu\|_{\mathcal{M}_b(]0,T[;L^2(\Omega))} \;\; \text{and} \\ \lim\limits_{k\to\infty} \displaystyle\int_Q y\mu_k\, dxdt = \Big\langle y, \mu|_{]0,T[}\Big\rangle_{\mathcal{C}([0,T];L^2(\Omega)),\mathcal{M}_b(]0,T[;L^2(\Omega))} \\ \text{whenever} \;\; y \in \mathcal{C}\big([0, T]; L^2(\Omega)\big) . \end{cases}$$

To define μ_k, we use the following construction. Let $\bar{\mu}$ be the extension of $\mu|_{]0,T[}$ by zero to $I\!R$, let $\{\rho_k\}$ be a sequence of nonnegative symmetric mollifiers on $I\!R$ with their supports in $(-1/k, 1/k)$, and let ψ_0 and ψ_T be the functions on $I\!R$ defined by $\psi_0(t) := -t$ and $\psi_T(t) := 2T - t$. Given $k \ge 2$, we put

$$\bar{\mu}_k(A) := \big(\bar{\mu} * \rho_k\big)(S) + \big(\bar{\mu} * \rho_k\big)\big(\psi_0(S)\big) + \big(\bar{\mu} * \rho_k\big)\big(\psi_T(S)\big)$$

for every Borel subset S in $I\!R$, where the sign $*$ stands for the *convolution product* between $\bar{\mu}$ and the regularizing kernel ρ_k. Since both distributions are with compact supports, the above convolutions are well defined. Then construct the desired measure by

$$\mu_k := \frac{\|\mu\|_{\mathcal{M}_b(]0,T[;L^2(\Omega))}}{\|\bar{\mu}_k|_{]0,T[}\|_{\mathcal{M}_b(]0,T[;L^2(\Omega))}} \; \bar{\mu}_k|_{]0,T[} \, .$$

One can verify that this sequence $\{\mu_k\} \subset L^1\big(0, T; L^2(\Omega)\big)$ satisfies both relations listed above.

Considering now the unique solution p_k to the system

$$\begin{cases} p_{tt} - \Delta p + \vartheta p = \mu_k & \text{in } Q \,, \\ \partial_\nu p = 0 & \text{in } \Sigma \,, \\ p(T) = 0, \quad p_t(T) = -\mu|_{\Omega \times \{T\}} & \text{in } \Omega \end{cases} \tag{7.36}$$

and applying Lemma 7.10, we get the estimate

$$\begin{aligned} &\|p_k\|_{L^\infty(0,T;H^1(\Omega))} + \|p_{kt}\|_{L^\infty(0,T;L^2(\Omega))} + \|p_k(0)\|_{H^1(\Omega)} \\ &+\|p_{kt}(0)\|_{L^2(\Omega)} \le C\|\mu\|_{\mathcal{M}_b(]0,T];L^2(\Omega))} \end{aligned} \tag{7.37}$$

with a constant $C > 0$ independent of k, where p_{kt} stands for the derivative of p_k with respect to $t \in (0, T)$ in the sense of vector-valued distributions. Denoting by p_{ktt} the corresponding derivative of p_{kt} with respect to $t \in (0, T)$ and using (7.36), we arrive at

$$\begin{aligned} p_{ktt} = \pi_k + \mu_k \in L^\infty\big(0, t; H^1(\Omega)^*\big) + \mathcal{M}_b\big(]0, T[; L^2(\Omega)\big) \\ \subset \mathcal{M}_b\big(]0, T[; H^1(\Omega)^*\big) \,, \end{aligned}$$

where the operator π_k is defined by

$$\Big\langle \pi_k, y \Big\rangle_{L^\infty(0,T;H^1(\Omega)^*),L^1(0,T;H^1(\Omega))} := \int_Q \Big(\nabla p_k \cdot \nabla y - \vartheta p_k y\Big)\, dxdt \,.$$

Therefore, in addition to (7.37), the sequences $\{p_{ktt}\}$ and $\{p_{kt}\}$ are bounded in the spaces $\mathcal{M}_b\big(]0, T[; H^1(\Omega)^*\big)$ and $BV\big([0, T]; H^1(\Omega)^*\big)$, respectively. Observing that $\mathcal{M}_b\big(]0, T[; H^1(\Omega)^*\big)$ is the dual of a *separable* Banach space, we select weak* convergent *subsequences* of the above sequences. The same weak* *sequential compactness* property holds for the space $BV\big([0, T]; H^1(\Omega)^*\big)$. Thus we find $p \in L^\infty(0, T; H^1(\Omega))$ with $p_t \in L^\infty\big(0, T; L^2(\Omega)\big) \cap BV\big([0, T]; H^1(\Omega)^*\big)$ and a subsequence $\{p_k\}$ converging to p weak* in $L^\infty\big(0, T; H^1(\Omega)\big)$ and such that $\{p_{kt}\}$ converges weak* in $L^\infty\big(0, T; L^2(\Omega)\big)$ to p_t. Furthermore, since $\gamma_{\nu_Q}(-\nabla p_k, p_{kt})$ is bounded in $L^2(\partial Q)$, we can also deduce that the sequence of $\gamma_{\nu_Q}(-\nabla p_k, p_{kt})$ converges to $\gamma_{\nu_Q}(-\nabla p, p_t)$ in the weak topology of $L^2(\partial Q)$. Taking into account the relations

$$\gamma_{\nu_Q}(-\nabla p_k, p_{kt})|_{\Omega \times \{T\}} = \mu|_{\Omega \times \{T\}} \quad \text{and} \quad \gamma_{\nu_Q}(-\nabla p_k, p_{kt})|_{\Sigma} = 0 \,,$$

one gets that $\gamma_{\nu_Q}(-\nabla p, p_t)|_\Sigma = -\partial_\nu p = 0$ and that

$$\gamma_{\nu_Q}(-\nabla p_k, p_{kt})|_{\Omega \times \{0\}} = p_{kt}(0) \xrightarrow{w} \gamma_{\nu_Q}(-\nabla p, p_t)|_{\Omega \times \{0\}} = p_t(0)$$

in the weak topology of $L^2(\Omega)$. Finally, by passing to the limit in the equality

$$\big\langle y(\varphi), \mu_k \big\rangle_{C([0,T];L^2(\Omega)),\mathcal{M}_b(]0,T];L^2(\Omega))} - \int_Q p_k \varphi \, dxdt = 0 \,,$$

where $y(\varphi)$ is the solution to (7.35), we conclude that $p(\cdot)$ is the desired weak solution to (7.33) and thus complete the proof of the theorem. △

In conclusion of this subsection, let us present a useful Green-type relationship between the corresponding solutions to the (linearized) state and adjoint systems in the Neumann problem under consideration.

Theorem 7.17 (Green formula for the hyperbolic Neumann problem). *Given* $(\phi, y_0, y_1) = (0,0,0)$ *and* $u \in L^2(\Sigma)$*, consider the corresponding weak solution* $y(\cdot)$ *to system* (7.32)*. Given* $\mu \in \mathcal{M}_b(]0,T]; L^2(\Omega))$*, let* p *satisfy the adjoint system* (7.33)*. Then one has*

$$\Big\langle y, \mu \Big\rangle_{\mathcal{C}([0,T];L^2(\Omega)),\mathcal{M}_b(]0,T];L^2(\Omega))} - \int_Q p\,\varphi\, dxdt = \int_\Sigma pu\, dsdt\ .$$

Proof. It follows from the proof of Theorem 7.16 that an *approximate* analog of the Green formula holds for the pairs (y, p_k), where p_k is the corresponding weak solution to the approximating adjoint system (7.36) for each $k \in I\!N$. Passing there to the limit as $k \to \infty$, we obtain the desired Green formula for the Neumann problem as stated in the theorem. △

7.2.3 Needle-Type Variations and Increment Formula

As mentioned above, our approach to deriving necessary optimality conditions in the original state-constrained Neumann problem (NP) includes an *approximation procedure to penalize state constraints*. In this way we arrive at a family of Neumann boundary control problems for hyperbolic equations with *pointwise/hard constraints on controls* but with *no state constraints*. Although the latter approximating problems are significantly easier than the initial state-constrained problem (NP), they still require a delicate variational analysis. As well known in optimal control theory for ordinary differential systems, a key element in deriving maximum-type conditions for problems with hard constraints on control in the absence of state variables is an *increment formula* for minimizing objectives over *needle variations* of optimal controls; cf. Sect. 6.3. In this subsection we obtain some counterparts of such results for the hyperbolic control problems under consideration by using *multidimensional* analogs of needle variations known in the PDE control literature as "diffuse perturbations" and also as "(multi)spike/patch perturbations" of the reference control. We adopt the "diffuse" terminology in what follows.

Given a reference control $\bar{u}(\cdot) \in U_{ad}$, an admissible control $u(\cdot) \in U_{ad}$, and a number $\rho \in (0,1)$, a *diffuse perturbation/variation* of $\bar{u}$ is defined by

$$u_\rho(s,t) := \begin{cases} \bar{u}(s,t) \text{ in } & \Sigma \setminus E_\rho\ , \\ \\ u(s,t) \text{ in } & E_\rho\ , \end{cases} \tag{7.38}$$

where E_ρ is a measurable subset of Σ. The next theorem can be viewed as an increment formula for the cost functional $J(y,u)$ with respect to diffuse perturbations of the reference control. Note that it also contains the corresponding *Taylor expansion* for state trajectory of (7.27), which is an essential ingredient of the increment formula. In what follows we denote the increment of the cost functional J by $\widehat{\Delta}J$ to distinguish it from the Laplacian Δ.

Theorem 7.18 (increment formula in the Neumann problem). *Given arbitrary controls $\bar{u}, u \in U_{ad}$ and a number $\rho \in (0,1)$, consider the diffuse perturbation* (7.38) *and the weak solutions $\bar{y}$ and y_ρ of system* (7.27) *corresponding to $\bar{u}$ and u_ρ, respectively. Then there is a measurable subset $E_\rho \subset \Sigma$ such that the following hold:*

$$\mathcal{L}^n(E_\rho) = \rho\mathcal{L}^n(\Sigma)\ ,$$

$$\int_{E_\rho}\big(h(s,t,\bar{u}) - h(s,t,u)\big)\,dsdt = \rho\int_\Sigma\big(h(s,t,\bar{u}) - h(s,t,u)\big)\,dsdt\ ,$$

$$y_\rho = \bar{y} + \rho z + \rho r_\rho \quad \textit{with} \quad \lim_{\rho\to 0}\|r_\rho\|_{\mathcal{C}([0,T];L^2(\Omega))} = 0, \quad \textit{and} \tag{7.39}$$

$$\begin{aligned} &J(y_\rho,u_\rho) = J(\bar{y},\bar{u}) + \rho\widehat{\Delta}J + o(\rho) \quad \textit{with} \\ &\widehat{\Delta}J := J'_y(\bar{y},\bar{u})z + J(\bar{y},u) - J(\bar{y},\bar{u})\ , \end{aligned} \tag{7.40}$$

where $z(\cdot)$ is the weak solution to the system

$$\begin{cases} z_{tt} - \Delta z + \vartheta'_y(\cdot,\bar{y})z = 0 & \textit{in } Q \\ \partial_\nu z = \bar{u} - u & \textit{in } \Sigma\ , \\ z(0) = 0, \quad z_t(0) = 0 & \textit{in } \Omega\ . \end{cases} \tag{7.41}$$

The proof of this theorem given below relies on the following technical lemma established as Lemma 4.2 in the paper by Raymond and Zidani [1121], where the reader can find all the details. Recall that the notation χ_E stands for the *characteristic function* of the set E equal to 1 on E and to 0 outside.

Lemma 7.19 (properties of diffuse perturbations). *Let $\bar{u}, u \in U_{ad}$. For every $\rho \in (0,1)$ there is a sequence of measurable subsets $E_\rho^k \subset \Sigma$ such that:*

$$\mathcal{L}^n(E_\rho^k) = \rho\mathcal{L}^n(\Sigma)\ ,$$

$$\int_{E_\rho^k}\big(h(s,t,\bar{u}) - h(s,t,u)\big)\,dsdt = \rho\int_\Sigma\big(h(s,t,\bar{u}) - h(s,t,u)\big)\,dsdt, \quad \textit{and}$$

$$\frac{1}{\rho}\chi_{E_\rho^k} \xrightarrow{w^*} 1 \quad \textit{in} \quad L^\infty(\Sigma) \quad \textit{as} \quad k \to \infty\ .$$

Proof of Theorem 7.18. The existence of the subsets E_ρ satisfying the conditions of the theorem is an easy consequence of Lemma 7.19. The main issue is to justify the Taylor expansion (7.39), which clearly implies the increment formula (7.40) due to the construction of diffuse perturbations.

To prove (7.39), we pick a number $\rho \in (0,1)$, take the sets E_ρ^k from Lemma 7.19, and build the diffuse control perturbations by

$$u_\rho^k(s,t) := \begin{cases} \bar{u}(s,t) \text{ in } & \Sigma \setminus E_\rho^k , \\ u(s,t) \text{ in } & E_\rho^k . \end{cases}$$

Let y_ρ^k be the solution of (7.27) corresponding to u_ρ^k, and let z be the (unique) weak solution of (7.41). It is easy to see that for each $\rho \in (0,1)$ and $k \in I\!N$ the function $\xi_\rho^k := \frac{1}{\rho}(y_\rho^k - \bar{y}) - z$ is the unique weak solution to the system

$$\begin{cases} \xi_{tt} - \Delta\xi + \theta_\rho^k \xi = f_\rho^k & \text{in } Q , \\ \partial_\nu \xi = w_\rho^k & \text{in } \Sigma , \\ \xi(0) = 0, \quad \xi_t(0) = 0 \text{ in } \Omega & \end{cases}$$

with the following data: $f_\rho^k := \big(\vartheta_y'(\cdot,\bar{y}) - \theta_\rho^k\big)z$,

$$\theta_\rho^k := \int_0^1 \vartheta_y'\big(\cdot, \bar{y} + \tau(y_\rho^k - \bar{y})\big)\,d\tau, \quad \text{and} \quad w_\rho^k := \Big(1 - \frac{1}{\rho}\chi_{E_\rho^k}\Big)(u - \bar{u}) .$$

Denote by $\xi_\rho^{k,1}$ the solution to

$$\begin{cases} \xi_{tt} - \Delta\xi + \theta_\rho^k \xi = f_\rho^k & \text{in } Q , \\ \partial_\nu \xi = 0 & \text{in } \Sigma , \\ \xi(0) = 0, \quad \xi_t(0) = 0 \text{ in } \Omega , & \end{cases}$$

by $\xi_\rho^{nk,2}$ the solution to

$$\begin{cases} \xi_{tt} - \Delta\xi + \theta_\rho^k \xi = 0 & \text{in } Q , \\ \partial_\nu \xi = w_\rho^k & \text{in } \Sigma , \\ \xi(0) = 0, \quad \xi_t(0) = 0 \text{ in } \Omega , & \end{cases}$$

and by ζ_ρ^k the solution to

$$\begin{cases} \zeta_{tt} - \Delta\zeta + \theta\zeta = 0 & \text{in } Q\ , \\ \partial_\nu \zeta = w_\rho^k & \text{in } \Sigma\ , \\ \zeta(0) = 0, \quad \zeta_t(0) = 0 & \text{in } \Omega\ , \end{cases}$$

where $\theta(x,t) := \vartheta_y'\big(x,t,\bar{y}(x,t)\big)$. One clearly has

$$(\xi_\rho^{k,2} - \zeta_\rho^k)_{tt} - \Delta(\xi_\rho^{k,2} - \zeta_\rho^k) + \theta_\rho^k(\xi_\rho^{k,2} - \zeta_\rho^k) = (\theta - \theta_\rho^k)\zeta_\rho^k \text{ in } Q\ ,$$

$$\partial_\nu(\xi_\rho^{k,2} - \zeta_\rho^k) = 0 \qquad \text{in } \Sigma\ ,$$

$$(\xi_\rho^{nk,2} - \zeta_\rho^k)(0) = 0, \quad (\xi_\rho^{k,2} - \zeta_\rho^k)_t(0) = 0 \qquad \text{in } \Omega\ .$$

By Lemma 7.10 we find a constant $C > 0$ independent of k and ρ such that the estimates

$$\|\xi_\rho^{k,2} - \zeta_\rho^k\|_{\mathcal{C}([0,T];L^2(\Omega))} \le C\|\theta - \theta_\rho^k\|_{L^1(0,T;L^{2n}(\Omega))} \cdot \|\zeta_\rho^k\|_{L^\infty(0,T;L^{2n/(n-1)}(\Omega))}\ ,$$

$$\le C\|\theta - \theta_\rho^k\|_{L^1(0,T;L^{2n}(\Omega))} \cdot \|\zeta_\rho^k\|_{L^\infty(0,T;H^{1/2}(\Omega))} \quad \text{and}$$

$$\|\xi_\rho^{k,1}\|_{\mathcal{C}([0,T];L^2(\Omega))} \le C\|f_\rho^k\|_{L^1(0,T;L^2(\Omega))}$$

hold for all $k \in I\!N$ and $0 < \rho < 1$, where the functions $\|\zeta_\rho^{nk}\|_{L^\infty(0,T;L^{2n/(n-1)}(\Omega))}$ are uniformly bounded due to Lemma 7.9. Employing now the weak* convergence in Lemma 7.19, we conclude that the sequence of w_ρ^k converges to zero in the weak topology of $L^2(\Sigma)$ and, by Lemma 7.11, the sequence of ζ_ρ^k converges to zero strongly in $\mathcal{C}\big([0,T];L^2(\Omega)\big)$ as $k \to \infty$ for all $0 < \rho < 1$. Thus there is an integer $k(\rho)$ such that

$$\|\zeta_\rho^{k(\rho)}\|_{\mathcal{C}([0,T];L^2(\Omega))} \le \rho \qquad \text{whenever } 0 < \rho < 1\ .$$

Observe further that the functions $u_\rho^{k(\rho)}$ converge to $\bar{u}$ strongly in $L^2(\Sigma)$ as $\rho \downarrow 0$. Then it follows from Theorem 7.13 that the functions $y_\rho^{k(\rho)}$ converge to $\bar{y}$ strongly in $\mathcal{C}\big([0,T];L^2(\Omega)\big)$ as $\rho \downarrow 0$. Invoking assumption (H1), one has that the functions $f_\rho^{k(\rho)}$ converge to zero strongly in $L^1\big(0,T;L^2(\Omega)\big)$ and that the functions $(\theta - \theta_\rho^{k(\rho)})$ converge to zero strongly in $L^1\big(0,T;L^{2n}(\Omega)\big)$ as $\rho \downarrow 0$. Taking into account the above estimates, this implies the relations

$$\lim_{\rho\to 0} \|\xi_\rho^{k(\rho)}\|_{\mathcal{C}([0,T];L^2(\Omega))} \le \lim_{\rho\to 0} \big(\|\xi_\rho^{k(\rho),1}\|_{\mathcal{C}([0,T];L^2(\Omega))}$$

$$+\|\xi_\rho^{k(\rho),2} - \zeta_\rho^{k(\rho)}\|_{\mathcal{C}([0,T];L^2(\Omega))} + \|\zeta_\rho^{k(\rho)}\|_{\mathcal{C}([0,T];L^2(\Omega))}\big) = 0\ .$$

Setting finally $E_\rho := E_\rho^{nk(\rho)}$, $u_\rho := u_\rho^{k(\rho)}$, and $\frac{1}{\rho} r_\rho := \xi_\rho^{k(\rho)}$, we complete the proof of the theorem. △

7.2.4 Proof of Necessary Optimality Conditions

This subsection is devoted to the proof of the necessary optimality conditions for the state-constrained Neumann boundary control problem (NP) formulated in Theorem 7.8. The proof involves a *strong approximation* procedure to *penalize the state constraints*, which is based on applying the *Ekeland variational principle* presented in Theorem 2.26. To accomplish this procedure, we first describe a complete metric space and a lower semicontinuous function, which are suitable for the application of Ekeland's principle to our problem.

Given $\bar{u}(\cdot) \in U_{ad}$ and a fixed positive number m, define the set

$$U_{ad}(\bar{u}, m) := \big\{u \in U_{ad} \big|\ |u(s,t) - \bar{u}(s,t)| \le m \ \text{ for a.e. } (s,t) \in \Sigma\big\}$$

and endow this set with the metric $d(\cdot,\cdot)$ defined by

$$d(v,u) := \mathcal{L}^N\big(\{(s,t)\big|\ v(s,t) \ne u(s,t)\}\big)\ ,$$

where $\mathcal{L}^n(\Omega)$ denotes as before the n-dimensional Lebesgue measure of the set $\Omega \subset I\!R^n$. Observe that if $\{u_k\} \subset U_{ad}(\bar{u}, m)$ and $u \in U_{ad}(\bar{u}, m)$ are such that $\lim_{k\to\infty} d(u_k, u) = 0$, then the sequence $\{u_k\}$ *strongly* converges to u in the norm of $L^2(\Sigma)$. The next result provides more information about this space and about the cost functional of (NP) on it, where y_u stands for the weak solution of (7.27) corresponding to u.

Lemma 7.20 (proper setting for Ekeland's principle). *The metric space $\big(U_{ad}(\bar{u}, m), d\big)$ is complete, and the mapping $u \mapsto \big(y_u, J(y_u, u)\big)$ is continuous from $\big(U_{ad}(\bar{u}, m), d\big)$ into $\mathcal{C}\big([0,T]; L^2(\Omega)\big) \times I\!R$.*

Proof. The completeness of the space $\big(U_{ad}(\bar{u}, m), d\big)$ is a well-known fact, which goes back to the original paper by Ekeland [397]. Let us prove the continuity statement of the lemma based on the regularity of weak solutions to the state system (7.27) established in Subsect. 7.2.2.

To proceed, pick $\{u_k\} \subset U_{ad}(\bar{u}, m)$ and $u \in U_{ad}(\bar{u}, m)$ such that the control sequence $\{u_k\}$ converges to u in the above d-metric as $k \to \infty$. Denote by y and by y_k the weak solutions of (7.27) corresponding to u and to u_k, respectively. Since $u_k \to u$ strongly in $L^2(\Sigma)$, the trajectories y_k strongly converge to y in the space $\mathcal{C}\big([0,T]; L^2(\Omega)\big)$ by Theorem 7.13. Furthermore, it follows from the estimates in assumptions (H2)–(H4) that the sequence of values $J(y_k, u_k)$ converges to $J(y, u)$ as $k \to \infty$. This ensures the desired continuity and completes the proof of the lemma. △

Now using the classical results in the geometry of Banach spaces collected, e.g., in the book by Li and Yong [789, Chap. 2], we conclude by the separability of $\mathcal{C}\big([0,T]; L^2(\Omega)\big)$ that there exists an *equivalent norm* $|\cdot|_{\mathcal{C}([0,T];L^2(\Omega))}$ on this space such that is *Gâteaux differentiable* at any nonzero point and its dual norm on $\mathcal{M}\big([0,T]; L^2(\Omega)\big)$—denoted by $|\cdot|_{\mathcal{M}([0,T];L^2(\Omega))}$—is *strictly convex.*

Given the constraint set $\Theta \subset \mathcal{C}\big([0,T];L^2(\Omega)\big)$ in the original problem (NP), we consider the corresponding *distance function*

$$d_\Theta(x) := \inf_{z\in\Theta} |x-z|_{\mathcal{C}\big([0,T];L^2(\Omega)\big)}$$

defined via the new norm $|\cdot|_{\mathcal{C}([0,T];L^2(\Omega))}$ on $\mathcal{C}\big([0,T];L^2(\Omega)\big)$. This function is globally Lipschitzian with modulus $\ell = 1$ and convex on $\mathcal{C}\big([0,T];L^2(\Omega)\big)$ by the convexity of Θ. Furthermore, one has

$$\begin{cases} |\xi|_{\mathcal{M}\big([0,T];L^2(\Omega)\big)} \le 1 & \text{if } x^* \in \partial d_\Theta(x) \text{ and } x \in \Theta, \\ |x^*|_{\mathcal{M}([0,T];L^2(\Omega))} = 1 & \text{if } x^* \in \partial d_\Theta(x) \text{ and } x \notin \Theta; \end{cases}$$

cf. Subsect. 1.3.3. Taking into account that the dual norm $|\cdot|_{\mathcal{M}([0,T];L^2(\Omega))}$ is also strictly convex, we conclude that the subdifferential $\partial d_\Theta(x)$ is a *singleton*, and hence d_Θ is *Gâteaux differentiable* at x for every $x \notin \Theta$.

Let $\{\bar{y}(\cdot),\bar{u}(\cdot)\}$ be an optimal solution to the original problem (P). Using the above distance function d_Θ, we define the *penalized functional* by

$$J_m(y,u) := \left[\left(J(y,u) - J(\bar{y},\bar{u}) + \frac{1}{m^2}\right)^+\right]^2 + d_\Theta^2(y), \quad m \in I\!N\ ,$$

where J is the cost functional in (NP). Since $J_m(\bar{y},\bar{u}) = m^{-4}$, one has that

$$J_m(\bar{y},\bar{u}) < \inf\Big\{ J_m(y,u) \Big|\ u \in U_{ad}(\bar{u},m^{1/3}),\ \ (y,u)\ \text{ satisfies }\ (7.27)\Big\} + \frac{1}{m^2}\ ,$$

for all $m \in I\!N$, i.e., $\{\bar{y}(\cdot),\bar{u}(\cdot)\}$ is an approximate $\frac{1}{m^2}$*-optimal solution* to the penalized problem.

Observe that the functional J_m is *smooth* at points where it *doesn't vanish*, in the sense that it is Gâteaux differentiable at such points; cf. the smoothing procedures in the *metric approximation* proofs of Theorems 2.8 and 2.10 for the extremal principle. This follows from the construction of J_m, assumptions (H2)–(H4), and the above property of d_Θ. Ekeland's principle allows us to *strongly* approximate the reference pair $\{\bar{y}(\cdot),\bar{u}(\cdot)\}$ by a pair $\{y_m(\cdot),u_m(\cdot)\}$ satisfying (7.27) in such a way that $\{y_m(\cdot),u_m(\cdot)\}$ is an *exact solution* to some *perturbed* optimal control problem for system (7.27) with the same control constraints and with *no state constraints.* After all these discussions and preliminary results we are ready to prove the main theorem.

Proof of Theorem 7.8. Divide the proof of this theorem into the following three major steps.

Step 1: Approximating problems via Ekeland's principle. Given an optimal solution $\{\bar{y}(\cdot),\bar{u}(\cdot)\}$ to the original problem (NP), we fix a natural number $m \in I\!N$ and conclude from Lemma 7.20 that the metric space $\big(U_{ad}(\bar{u},m^{1/3}),d\big)$ is complete and that the function $u \longmapsto J_m(y_u,u)$ is lower semicontinuous (even continuous) on this space. By the Ekeland variational principle we find an admissible control u_m satisfying

$$\begin{aligned} &u_m \in U_{ad}(\bar{u}, m^{1/3}), \quad d(u_m, \bar{u}) \le \frac{1}{m}, \quad \text{and} \\ &J_m(y_m, u_m) \le J_m(y_u, u) + \frac{1}{m} d(u_m, u) \end{aligned} \tag{7.42}$$

for all $u \in U_{ad}(\bar{u}, m^{1/3})$, where y_m and y_u are the weak solutions of (7.27) corresponding to u_m and u, respectively. The latter means that, for all natural numbers $m \in I\!N$, the control u_m is an *optimal solution* to the *perturbed problem* (NP_m) defined by:

$$\inf \left\{ J_m(y, u) + \frac{1}{m} \,\middle|\, u \in U_{ad}(\bar{u}, m^{1/3}), \;\; (y, u) \;\text{ satisfies (7.27)} \right\}.$$

Step 2: Necessary conditions in approximating problems. First take an arbitrary control $u_0 \in U_{ad}$ and construct the following modification of the optimal control $\bar{u}$ to (NP) by

$$u_{0m}(s, t) := \begin{cases} u_0(s, t) & \text{if } |u_0(s, t) - \bar{u}(s, t)| \le m^{1/3}, \\ \bar{u}(s, t) & \text{otherwise}. \end{cases}$$

Note that the control u_{0m} is feasible for the approximating problem (NP_m) whenever $m \in I\!N$. Given any $0 < \rho < 1$, define then *diffuse perturbations* of the optimal control u_m to (NP_m) by

$$u_\rho^m(s, t) := \begin{cases} u_m(s, t) & \text{in } \Sigma \setminus E_\rho^m, \\ u_{0m}(s, t) & \text{in } E_\rho^m. \end{cases}$$

Theorem 7.18 ensures the existence of measurable sets $E_\rho^m \subset \Sigma$ for which one has the relations

$$\begin{aligned} &\mathcal{L}^n(E_\rho^m) = \rho \mathcal{L}^n(\Sigma), \quad y_\rho^m = y_m + \rho z_m + \rho r_\rho^m, \\ &\lim_{\rho \to 0} \|r_\rho^m\|_{\mathcal{C}([0,T];L^2(\Omega))} = 0, \quad \text{and} \\ &J(y_\rho^m, u_\rho^m) = J(y_m, u_m) + \rho \widehat{\Delta} J^m + o(\rho), \end{aligned} \tag{7.43}$$

where y_ρ^m is the weak solution of (7.27) corresponding to u_ρ^m, where z_m is the weak solution to

$$\begin{cases} z_{tt} - \Delta z + \vartheta_y'(\cdot, y_m) z = 0 & \text{in } Q, \\ \partial_\nu z = u_m - u_{0m} & \text{in } \Sigma, \\ z(0) = 0, \quad z_t(0) = 0 & \text{in } \Omega, \end{cases}$$

and where $\widehat{\Delta} J^m$ is defined by

$$\widehat{\Delta} J^m := \int_Q g'_y(\cdot, y_m) z_m \, dxdt + \int_\Omega f'_y\big(\cdot, y_m(T)\big) z_m \, dx$$

$$+ \int_\Sigma \big(h(\cdot, u_{0m}) - h(\cdot, u_m)\big) \, dsdt \ .$$

Since each u^m_ρ is feasible for (NP_m), it follows from (7.42) and the construction of the metric $d(\cdot, \cdot)$ therein that

$$\lim_{\rho \to 0} \frac{J_m(y_m, u_m) - J_m(y^m_\rho, u^m_\rho)}{\rho} \le \frac{1}{m} \mathcal{L}^n(\Sigma) \ . \tag{7.44}$$

Observe that $J_m(y_m, u_m) \neq 0$ for all $m \in I\!N$ due the optimality of u_m in (NP_m) and the structure of J_m. Hence J_m is *Gâteaux differentiable* at (y_m, u_m) by the discussion above. Then it easily follows from (7.43) and (7.44) that one has the optimality condition

$$-\lambda_m \widehat{\Delta} J^m - \big\langle \mu_m, z_m \big\rangle \le \frac{1}{m} \mathcal{L}^n(\Sigma) \ , \tag{7.45}$$

where the multipliers λ_m and μ_m are computed by

$$\lambda_m := \frac{\Big(J(y_m, u_m) - J(\bar{y}, \bar{u}) + \dfrac{1}{m^2}\Big)^+}{J_m(y_m, u_m)} \ ,$$

$$\mu_m := \begin{cases} \dfrac{d_\Theta(y_m) \nabla d_\Theta(y_m)}{J_m(y_m, u_m)} & \text{if } y_m \notin \Theta \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

Noting that $\mu_m \in \mathcal{M}\big([0, T]; L^2(\Omega)\big)$, consider the (unique) weak solution p_m to the *adjoint system*

$$\begin{cases} p_{tt} - \Delta p + \vartheta'_y(\cdot, y_m) p = \lambda_m g'_y(\cdot, y_m) + \mu_m|_Q & \text{in } Q \ , \\ \partial_\nu p = 0 & \text{in } \Sigma \ , \\ p(T) = 0, \quad p_t(T) = -\lambda_m f'_y\big(\cdot, y_m(T)\big) - \mu_m|_{\Omega \times \{T\}} & \text{in } \Omega \ , \end{cases}$$

where $\mu_m|_Q$ and $\mu_m|_{\Omega \times \{T\}}$ are the restrictions of μ_m to Q and $\Omega \times \{T\}$, respectively. Employing the *Green formula* from Theorem 7.17, we have

$$\lambda_m \int_Q g'_y(x,t,y_m) z_m \, dxdt + \lambda \int_\Omega f'_y\big(x, y_m(T)\big) z_m(T)\, dx + \langle \mu_m, z_m \rangle$$

$$= \int_Q p_m \Big(z_{ktt} - \Delta z_m + \vartheta'_y(\cdot, y_m) z_m \Big)\, dxdt + \int_\Sigma p_m \, \partial_\nu z_m \, dsdt$$

$$= \int_\Sigma p_m (u_m - u_{0m})\, dsdt \; .$$

The latter implies, by (7.45) and the definition of $\widehat{\Delta} J^m$, that

$$\int_\Sigma \big(\lambda_m h(s,t,u_m) + p_m u_m \big)\, dsdt \le \int_\Sigma \big(\lambda_m h(s,t,u_{0m}) + p_m u_{0m} \big)\, dsdt$$
$$+ \frac{1}{m} \mathcal{L}^n(\Sigma) \;\text{ for all }\; m \in I\!N \; , \tag{7.46}$$

which gives the desired necessary optimality conditions for the solutions u_m to the approximating problems (NP_m).

Step 3: Passing to the limit. To conclude the proof of the theorem, we need to pass to the limit in the above relationships for the optimal solutions u_m to (NP_m) as $m \to \infty$. First observe that

$$\lambda_m^2 + |\mu_m|^2_{\mathcal{M}([0,T];L^2(\Omega))} = 1 \;\text{ for all }\; m \in I\!N \; .$$

Invoking basic functional analysis, we find $(\lambda, \bar\mu) \in I\!R \times \mathcal{M}\big([0,T]; L^2(\Omega)\big)$ with $\lambda \ge 0$ and a subsequence of (λ_m, μ_m), still indexed by m, such that

$$\lambda_m \to \lambda \;\text{ in }\; I\!R \quad\text{and}\quad \mu_m \xrightarrow{w^*} \bar\mu \quad \text{weak}^* \text{ in } \mathcal{M}\big([0,T]; L^2(\Omega)\big) \; .$$

Furthermore, Theorem 7.16 ensures the estimate

$$\|p_m\|_{L^\infty(0,T;H^1(\Omega))} + \|p_{kt}\|_{L^\infty(0,T;L^2(\Omega))}$$

$$\le C\Big(\|\mu\|_{\mathcal{M}([0,T];L^2(\Omega))} + \|g'_y(\cdot, y_m)\|_{L^1(0,T;L^2(\Omega))} + \|f'_y\big(\cdot, y_m(T)\big)\|_{L^2(\Omega)} \Big) \; .$$

Since the sequences $\{\lambda_m\} \subset I\!R$,

$$\{\mu_m\} \subset \mathcal{M}\big([0,T]; L^2(\Omega)\big), \quad \{y_m\} \subset \mathcal{C}\big([0,T]; L^2(\Omega)\big) \; ,$$

and $\{u_m\} \subset L^2(\Sigma)$ are bounded, the sequence $\{(p_m, p_{mt})\}$ is bounded in $L^\infty\big(0,T; H^1(\Omega)\big) \times L^\infty\big(0,T; L^2(\Omega)\big)$. Then there are

$$(p_m, p_{mt}) \xrightarrow{w^*} (p, p_t) \in L^\infty\big(0,T; H^1(\Omega)\big) \times L^\infty\big(0,T; L^2(\Omega)\big) \;\text{ and}$$

$$y_m \xrightarrow{w^*} \bar y \in L^\infty\big(0,T; L^2(\Omega)\big) \;\text{ as }\; m \to \infty$$

in the weak* topologies of the underlying spaces. We know that $u_m \to \bar{u}$ strongly in $L^2(\Sigma)$. Employing the standard arguments as above, it is easy to conclude that $\bar{y}$ is the solution of (7.27) corresponding to $\bar{u}$ and that p is the (unique) weak solution of (7.30) corresponding to $\bar{y}$.

Let us show that the limiting multipliers $(\lambda, \mu) = (\lambda, \bar{\mu}|_{]0,T]})$ are those whose existence is claimed in Theorem 7.8. First justify that $(\lambda, \mu) \neq 0$ due to requirement (H5) on the convexity and nonempty interiority of the set Θ. Suppose the contrary, which yields

$$\lim_{m\to\infty} |\mu_m|^2_{\mathcal{M}([0,T];L^2(\Omega))} = 1 \; . \tag{7.47}$$

By assumption (H5) we have $\widehat{y}_0 \in \operatorname{int}\Theta$. Thus there exists a closed ball $B_\rho(\widehat{y}_0) \subset \mathcal{C}\big([0,T];L^2(\Omega)\big)$ entirely contained in Θ. Employing (7.47) and picking any $m \in I\!N$, we find $z_m \in \rho I\!B$ satisfying

$$\big\langle z_m, \mu_m \big\rangle_{\mathcal{C}([0,T];L^2(\Omega)),\mathcal{M}([0,T];L^2(\Omega))} = \frac{\rho}{2} |\mu_m|_{\mathcal{M}([0,T];L^2(\Omega))} \; .$$

Since $\widehat{y}_0 + z_m \in \Theta$, observe from the definition of μ_m that

$$\big\langle \widehat{y}_0 + z_m - y_m, \mu_m \big\rangle_{\mathcal{C}([0,T];L^2(\Omega)),\mathcal{M}([0,T];L^2(\Omega))} \leq 0, \quad m \in I\!N \; .$$

Passing to the limit as $m \to \infty$, we get

$$\frac{\rho}{2} + \big\langle \widehat{y}_0 - \bar{y}, \bar{\mu} \big\rangle_{\mathcal{C}([0,T];L^2(\Omega)),\mathcal{M}([0,T];L^2(\Omega))} \leq 0 \; .$$

Remember that $\bar{y}(x,0) = \widehat{y}_0(x,0)$ and that $\mu = \bar{\mu}|_{]0,T]}$; therefore

$$\big\langle \widehat{y}_0 - \bar{y}, \bar{\mu} \big\rangle_{\mathcal{C}([0,T];L^2(\Omega)),\mathcal{M}([0,T];L^2(\Omega))} = \big\langle \widehat{y}_0 - \bar{y}, \mu \big\rangle_{\mathcal{C}([0,T];L^2(\Omega)),\mathcal{M}_b(]0,T];L^2(\Omega))} \, ,$$

which clearly implies that

$$\big\langle \widehat{y}_0 - \bar{y}, \mu \big\rangle_{\mathcal{C}([0,T];L^2(\Omega)),\mathcal{M}_b(]0,T];L^2(\Omega))} \leq -\frac{\rho}{2} < 0 \; .$$

The latter contradicts the assumption on $(\lambda, \mu) = 0$ and thus justifies the nontriviality condition in (7.28). The second condition therein easily follows from the above arguments due to the convexity of the constraint set Θ. Note that overall relationships (7.28) are in accordance with general results of constrained optimization, which particularly ensure nontriviality under SNC assumptions on constraint sets; cf. Proposition 1.25 and also Remark 7.30 below.

It remains to verify the minimum condition (7.29). To do this, recall that $u_m \to \bar{u}$ strongly in $L^2(\Sigma)$. Passing to the limit as $m \to \infty$ in (7.46), we get

$$\int_\Sigma \big(\lambda h(s,t,\bar{u}) + p\bar{u}\big)\, dsdt \leq \int_\Sigma \big(\lambda h(s,t,u_0) + pu_0\big)\, dsdt \tag{7.48}$$

for every $u_0 \in U_{ad}$. Finally, taking into account the structure of the admissible control set U_{ad} in (H6) and employing the standard arguments (as those in Sect. 7.4 for parabolic equations with no change required), we derive the pointwise condition (7.29) from the integral one in (7.48). This completes the proof of the theorem. △

Remark 7.21 (existence of optimal solutions to the hyperbolic Neumann problem). For brevity we don't address in this section the *existence* issue for optimal solutions to the Neumann boundary control problem under consideration. However, a general existence theorem for this problem can be derived from the regularity results presented in Subsect. 7.2.2 via the application of the classical Weierstrass theorem on the existence of optimal solutions to abstract problems of minimizing l.s.c. functions over compact sets in suitable topologies; cf. Sects. 7.3 and 7.4 below for similar arguments and results. What we need, however, is to impose additional *convexity* assumptions on the integrand h with respect to the *control variable*, as well as the convexity of the control sets $K(s,t)$ in (H6). Such a convexity, which is not needed for deriving the pointwise necessary optimality conditions, is required for the existence theorem in order to ensure the lower semicontinuity of the cost functional and the closedness of the feasible control set with respect to the corresponding *weak* convergence of controls that implies, by *regularity*, the *strong* convergence of trajectories. Note that we have to impose significantly more restrictive assumptions to handle the *Dirichlet boundary control* problem in the next section, where the *convexity* in both *control and state* variables is needed not only for the existence of optimal solutions but also for deriving necessary optimality conditions. The main reason is that Dirichlet boundary control problems, for hyperbolic as well for parabolic systems, exhibit much *less regularity* in comparison for their Neumann counterparts, and thus they require different methods of variational analysis; see Sects. 7.3 and 7.4.

7.3 Dirichlet Boundary Control of Linear Constrained Hyperbolic Equations

In this section we study a *Dirichlet* counterpart of the Neumann boundary control problem for hyperbolic equations with pointwise state constraints considered in Sect. 7.2. As mentioned, there are significant differences between Neumann and Dirichlet boundary conditions for hyperbolic equations; so the methods and results developed in this section are considerably distinguished from those in the preceding one. Roughly speaking, the requirements imposed on the initial data in the Dirichlet problem are *stronger*, while the results we are able to obtain are *weaker* in comparison with the above case of Neumann boundary controls. This is due to the *lack of regularity* in the Dirichlet case, which forces us to develop a different approach to the variational analysis of the state-constrained Dirichlet boundary control problem in what follows. In particular, necessary optimality conditions are derived in this approach by reducing the Dirichlet control problem of *dynamic optimization* to a problem of mathematical programming in infinite dimensions with *geometric* and *operator constraints* of special types.

7.3.1 Problem Formulation and Main Results for Dirichlet Controls

In what follows we keep the standard notation from Sect. 7.2. The Dirichlet problem under consideration is as follows. Given an open bounded domain $\Omega \subset I\!R^n$ with the boundary Γ of class $\mathcal{C}^2$, consider the problem of minimizing the integral functional

$$J(y,u) := \int_\Omega f\big(x, y(T)\big)dx + \int_Q g(x,t,y)dxdt + \int_\Sigma h(s,t,u)dsdt$$

for a fixed time $T > 0$ over admissible pairs $\{y(\cdot), u(\cdot)\}$ satisfying the multidimensional *linear wave equation* with control functions acting in the *Dirichlet boundary conditions*

$$\begin{cases} y_{tt} - \Delta y = \vartheta & \text{in } Q := \Omega \times (0,T)\,, \\ y = u & \text{in } \Sigma := \Gamma \times (0,T)\,, \\ y(0) = y_0, \quad y_t(0) = y_1 \text{ in } \Omega & \end{cases} \tag{7.49}$$

subject to the *pointwise control and state constraints*

$$u(\cdot) \in U_{ad} \subset L^2(\Sigma), \quad y(\cdot) \in \Theta \subset \mathcal{C}\big([0,T]; L^2(\Omega)\big)\,,$$

where $\vartheta \in L^1\big(0,T; H^{-1}(\Omega)\big)$, $y_0 \in L^2(\Omega)$, and $y_1 \in H^{-1}(\Omega)$ are given functions. Label this problem by (DP) and shortly write it as

$$\inf\Big\{J(y,u)\ \Big|\ \{y(\cdot),u(\cdot)\} \text{ satisfies (7.49)},\ u(\cdot) \in U_{ad},\ y(\cdot) \in \Theta\Big\}\,.$$

Our primary goal in this section is to derive *necessary optimality conditions* for the Dirichlet state-constrained problem (DP) under consideration; the same goal as for the Neumann problem (NP) studied in Sect. 7.2. However, we have to impose significantly more restrictive assumptions on the initial data of (DP), in comparison with those for (NP), to achieve even weaker results; see below. Observe that the hyperbolic dynamics in (DP) is described by the *linear* wave equation with ϑ independent of y, in comparison with the *semilinear* one in (NP). On the other hand, we impose *milder* requirements on the initial state $(y_0, y_1) \in L^2(\Omega) \times H^{-1}(\Omega)$ for the Dirichlet problem in comparison with $(y_0, y_1) \in H^1(\Omega) \times L^2(\Omega)$ for the Neumann case. In fact, the results obtained for (DP) can be extended to more general linear hyperbolic equations with a *strongly elliptic operator* instead of the Laplacian Δ.

Let us now formulate the *standing assumptions* on the initial data in (DP) required for the necessary optimality conditions derived below; only the *first* *four* assumptions, with no $\text{int}\,\Theta \neq \emptyset$ in (H4), are required for the *existence theorem* in what follows.

(H1) For every $y \in I\!R$ the function $f(\cdot, y) \geq 0$ is measurable in Ω with $f(\cdot, 0) \in L^1(\Omega)$. For a.e. $x \in \Omega$ the function $f(x, \cdot)$ is *convex* and continuous on the whole line $I\!R$.

(H2) For every $y \in I\!R$ the function $g(\cdot, \cdot, y) \geq 0$ is measurable in Q with $g(\cdot, \cdot, 0) \in L^1(Q)$. For a.e. $(x, t) \in Q$ the function $g(x, t, \cdot)$ is *convex* and continuous on $I\!R$.

(H3) For every $u \in I\!R$ the function $h(\cdot, u)$ is measurable in Σ with $h(\cdot, 0) \in L^1(\Sigma)$. For a.e. $(s, t) \in \Sigma$ the function $h(s, t, \cdot)$ is *convex* and continuous on $I\!R$. Moreover, h satisfies the following *growth condition*

$$|u|^2 \leq h(s, t, u) \quad \text{whenever} \quad (s, t) \in \Sigma \quad \text{and} \quad u \in I\!R \, .$$

(H4) The state constraint set $\Theta \in \mathcal{C}\big([0, T]; L^2(\Omega)\big)$ is a closed and *convex* with $\text{int}\,\Omega \neq \emptyset$. The control set $U_{ad} \in L^2(\Sigma)$ is also closed and *convex*. Furthermore, $y_0(\cdot) \in \text{int}\,\Theta$ for the initial function $(x, t) \mapsto y_0(x)$, and there is $u \in U_{ad}$ satisfying $y_u \in \Theta$ and $J(y, u) < \infty$ for the corresponding solution $y(\cdot)$ to the Dirichlet system (7.49).

(H5) For a.e. $x \in \Omega$ the function $f(x, \cdot)$ is of class $\mathcal{C}^1$ satisfying

$$|f'_y(x, y)| \leq C\big(1 + |y|\big) \quad \text{with some constant} \quad C > 0 \, .$$

(H6) For a.e. $(x, t) \in Q$ the function $g(x, t, \cdot)$ is of class $\mathcal{C}^1$ satisfying

$$|g'_y(x, t, y)| \leq C\big(1 + |y|\big) \quad \text{with some constant} \quad C > 0 \, .$$

(H7) For a.e. $(s, t) \in \Sigma$ the function $h(s, t, \cdot)$ is of class $\mathcal{C}^1$ satisfying

$$|h'_u(s, t, u)| \leq C\big(1 + |u|\big) \quad \text{with some constant} \quad C > 0 \, .$$

The main difference between the assumptions made for (DP) in comparison with for (NP) is that we now impose the *full convexity* of the integrands f, g, h with respect to the *state and control* variables, together with the convexity of the control set U_{ad}, while *no* convexity is required for the Neumann problem. As mentioned, it is due to the *lack of regularity* for the Dirichlet system (7.49) in comparison with the Neumann one; see Subsect. 7.3.2 for more details and discussions.

Actually the extra convexity assumptions allow us to *compensate*, in a sense, the lack of regularity. Based on the full convexity and the available regularity, we reduce the Dirichlet control problem under consideration to a special problem of *mathematical programming* with geometric and operator constraints in Banach spaces and then deduce necessary optimality conditions for (DP) from an appropriate version of the (abstract) Lagrange multiplier rule for mathematical programming in the line of Subsect. 5.1.2. The necessary optimality conditions for the Dirichlet problem derived in this way are given in the *integral form* of the Pontryagin maximum principle, in contrast to the pointwise form for the Neumann problem in Sect. 7.2. Furthermore, the assumptions made allow us to establish a general *existence theorem* for optimal controls in problem (DP).

Now we are ready to formulate the main results of this section. Note that the appropriate notions of (weak) solutions to the state and adjoint equations needed for these results will be rigorously clarified in Subsect. 7.3.3 and Subsect. 7.3.4, respectively.

Theorem 7.22 (existence of Dirichlet optimal controls). *Suppose that assumptions* (H1)–(H4), *with no* $\operatorname{int}\Theta \neq \emptyset$ *in* (H4), *are satisfied. Then the Dirichlet optimal control problem* (DP) *admits an optimal solution.*

The proof of Theorem 7.22 is given in Subsect. 7.3.3.

Theorem 7.23 (necessary optimality conditions for the hyperbolic Dirichlet problem). *Suppose that assumptions* (H1)–(H7) *are satisfied. Then for every optimal solution* $\{\bar{y}(\cdot), \bar{u}(\cdot)\}$ *to problem* (DP) *the following conditions hold: there are* $\lambda \geq 0$ *and* $\mu \in \mathcal{M}_b\big(]0,T];L^2(\Omega)\big)$ *such that*

$$(\lambda, \mu) \neq 0, \quad \big\langle \mu, y - \bar{y} \big\rangle \leq 0 \ \text{ for all } \ y \in \Theta \quad \text{and} \tag{7.50}$$

$$\int_\Sigma \Big(\frac{\partial p}{\partial \nu} + \lambda h'_u(s,t,\bar{u})\Big)(u - \bar{u})\,dsdt \geq 0 \ \text{ for all } \ u \in U_{ad}, \tag{7.51}$$

where p *is the corresponding solution to the adjoint system*

$$\begin{cases} p_{tt} - \Delta p = \lambda g'_y(x,t,\bar{y}) + \mu|_Q & \text{in} \quad Q\,, \\ p = 0 & \text{in} \quad \Sigma\,, \\ p(T) = y_0, \quad p_t(T) = -\lambda f'_y\big(x,\bar{y}(T)\big) - \mu|_{\Omega\times\{T\}} & \text{in} \quad \Omega\,. \end{cases} \tag{7.52}$$

Moreover, if there exists $\{y(\cdot), u(\cdot)\} \in Y \times (U_{ad} - \bar{u})$ *satisfying*

$$\begin{cases} y_{tt} - \Delta y = 0 \ \text{ in } \ Q, \quad y = u \ \text{ in } \ \Sigma\,, \\ y(0) = 0, \quad y_t(0) = 0 \ \text{ in } \ \Omega, \quad \bar{y} + y \in \operatorname{int}\Theta \end{cases} \tag{7.53}$$

with the state space Y *defined in* (7.54), *then one can take* $\lambda = 1$ *in the above optimality conditions.*

Note that the integral condition (7.51) is formulated as a part of the *minimum* (not maximum) principle, which is more convenient in our framework. The proof of Theorem 7.23 is given in Subsect. 7.3.5 with the preliminary analysis of the adjoint system conducted in Subsect. 7.3.4.

7.3.2 Existence of Dirichlet Optimal Controls

Let us first recall an appropriate notion of solutions to the nonhomogeneous Dirichlet state system (7.49) needed for the purposes of this study. The following notion of *weak solutions* meets our requirements.

Definition 7.24 (weak solutions to the Dirichlet state hyperbolic system). *A function* $y(\cdot)$ *with* $(y, y_t) \in \mathcal{C}\big([0,T]; L^2(\Omega)\big) \times \mathcal{C}\big([0,T]; H^{-1}(\Omega)\big)$ *is a* WEAK SOLUTION *to* (7.49) *if one has*

$$\int_Q f z\, dxdt = \int_Q y\varphi\, dxdt + \big\langle y_t(T), z^0\big\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)}$$

$$-\big\langle y_t(0), z(0)\big\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} - \int_\Omega y(T)z^1\, dx + \int_\Omega y(0)z_t(0)\, dx + \int_\Sigma \frac{\partial z}{\partial \nu_u}\, dsdt$$

for all $(\varphi, z^0, z^1) \in L^1\big(0,T; L^2(\Omega)\big) \times H_0^1(\Omega) \times L^2(\Omega)$, *where* z *solves the homogeneous Dirichlet problem*

$$\begin{cases} z_{tt} - \Delta z = \varphi & \text{in } Q\,, \\ z = 0 & \text{in } \Sigma\,, \\ z(T) = z^0, \quad z_t(T) = z^1 & \text{in } \Omega\,. \end{cases}$$

The importance of the defined notion of weak solutions to the hyperbolic system (7.49) is due to the following fundamental *regularity* result established by Lasiecka, Lions and Triggiani [740], which ensures the *existence, uniqueness*, and *continuous dependence* of weak solutions to (7.49) on the initial and boundary conditions in appropriate Banach spaces. We refer the reader to the afore-mentioned paper for the proof of this result and various applications.

Theorem 7.25 (basic regularity for the Dirichlet hyperbolic problem). *For every* $(\vartheta, u, y_0, y_1) \in L^1\big(0,T; H^{-1}(\Omega)\big) \times L^2(\Sigma) \times L^2(\Omega) \times H^{-1}(\Omega)$ *the Dirichlet system* (7.49) *admits a unique weak solution* $y(\cdot)$ *with* $(y, y_t) \in \mathcal{C}\big([0,T]; L^2(\Omega)\big) \times \mathcal{C}\big([0,T]; H^{-1}(\Omega)\big)$. *Furthermore, the mapping* $(\vartheta, u, y_0, y_1) \mapsto (y, y_t)$ *is linear and continuous from* $L^1\big(0,T; H^{-1}(\Omega)\big) \times L^2(\Sigma) \times L^2(\Omega) \times H^{-1}(\Omega)$ *into* $\mathcal{C}\big([0,T]; L^2(\Omega)\big) \times \mathcal{C}\big([0,T]; H^{-1}(\Omega)\big)$.

Theorem 7.25 plays a crucial role in further considerations. This theorem suggests us to introduce the space of *admissible state functions*, i.e., the space of solutions to system (7.49) when $(\vartheta, u, y_0, y_1) \in L^1\big(0,T; H^{-1}(\Omega)\big) \times L^2(\Sigma) \times L^2(\Omega) \times H^{-1}(\Omega)$, as follows

$$\begin{aligned} Y := \Big\{\, y \in \mathcal{C}\big([0,T]; L^2(\Omega)\big) \,\Big|\, & y_t \in \mathcal{C}\big([0,T]; H^{-1}(\Omega)\big)\,, \\ & y_{tt} - \Delta y \in L^1\big(0,T; H^{-1}(\Omega)\big),\;\; y|_\Sigma \in L^2(\Sigma) \Big\}\,. \end{aligned} \tag{7.54}$$

It is easy to see that the space Y is Banach with the norm $\|\cdot\|$ defined by

$$\|y\|_{\mathcal{C}([0,T];L^2(\Omega))} + \|y_t\|_{\mathcal{C}([0,T];H^{-1}(\Omega))} + \|y_{tt} - \Delta y\|_{L^1(0,T;H^{-1}(\Omega))} + \|y|_\Sigma\|_{L^2(\Sigma)} \cdot$$

Now based on Theorem 7.25 and standard results on the lower semicontinuity of integral functionals in appropriate weak topologies under the assumptions made, we justify the existence of optimal solutions to (DP) by reducing it to the classical Weierstrass theorem in the underlying topological spaces.

Proof of Theorem 7.22. By the existence and uniqueness statements in Theorem 7.25, there is a minimizing sequence $\{(y_k, u_k)\} \subset \mathcal{C}\big([0,T]; L^2(\Omega)\big) \times U_{ad}$ in problem (DP), where y_k is the (unique) solution of (7.49) corresponding to u_k. Due to the growth condition in (H3), the sequence $\{u_k\}$ is bounded in $L^2(\Sigma)$. Thus we suppose without loss of generality that $\{u_k\}$ converges to u in the weak topology of $L^2(\Sigma)$. Since U_{ad} is assumed to be closed and convex in (H4), one has $u(\cdot) \in U_{ad}$. It follows from the continuity statement in Theorem 7.25 that the sequence $\{(y_k, y_{kt})\}$ is bounded in $L^\infty\big(0,T; L^2(\Omega)\big) \times L^\infty\big(0,T; H^{-1}(\Omega)\big)$, where y_{kt} stands for the distributive derivative of y_k. Employing the above continuity, we conclude that $\{(y_k, y_{kt})\}$ converges to (y, y_t) in the weak* topology of $L^\infty\big(0,T; L^2(\Omega)\big) \times L^\infty\big(0,T; H^{-1}(\Omega)\big)$, where y is the solution of (7.49) corresponding to u. Invoking the closedness and convexity of Θ in (H4), one gets that $y(\cdot) \in \Theta$.

It remains to justify the lower semicontinuity of the cost functional, i.e., the limiting relation

$$J(y,u) \le \liminf_{k\to\infty} J(y_k, u_k) .$$

The latter follows directly from the classical results on the lower semicontinuity of integral functionals with respect to the weak topologies under consideration due to the crucial *convexity* assumptions in (H1)–(H3). Thus (y,u) is an optimal solution to the Dirichlet optimal control problem (DP). △

7.3.3 Adjoint System in the Dirichlet Problem

Our primal goal is to prove the necessary optimality conditions formulated in Theorem 7.23. To proceed, we first need to clarify an appropriate notion of solutions to the adjoint system in this theorem and then to establish some properties of adjoint trajectories allowing us to deduce the desired necessary optimality conditions for the hyperbolic control problem from abstract necessary optimality conditions for the auxiliary optimization problem in Banach spaces. Given $\mu \in \mathcal{M}_b\big(]0,T]; L^2(\Omega)\big)$, consider the system

$$\begin{cases} p_{tt} - \Delta p = \mu|_Q & \text{in } \ Q\ , \\ p = 0 & \text{in } \ \Sigma\ , \\ p(T) = 0, \quad p_t(T) = -\mu|_{\Omega\times\{T\}} & \text{in } \ \Omega\ , \end{cases} \tag{7.55}$$

corresponding to the adjoint system (7.52) in Theorem 7.23 with $(\lambda, y_0) = 0$, where $\mu|_Q$ (respectively $\mu|_{\Omega\times\{T\}}$) is the restriction of μ to Q (respectively to $\Omega \times \{T\}$). Observe that these restrictions satisfy $\mu|_Q \in \mathcal{M}_b(]0,T[;L^2(\Omega))$ and $\mu|_{\Omega\times\{T\}} \in L^2(\Omega)$.

To define an appropriate notion of solutions to the adjoint system (7.55), suppose for a moment that $(p, p_t) \in L^2(0,T;H_0^1(\Omega)) \times L^2(0,T;L^2(\Omega))$ and that $p_{tt} - \Delta p \in \mathcal{M}_b(]0,T[;L^2(\Omega))$, where the derivatives are calculated in the sense of distributions in Q. Then, following the corresponding proof for the Neumann problem in Subsect. 7.2.2 based on the *divergence formula* from Lemma 7.14, we define the normal trace on ∂Q for the vectorfield $(-\nabla p, p_t)$ as an element of $H^{-1/2}(\partial Q)$. Moreover, denoting this normal trace by $\gamma_{\nu_Q}(-\nabla p, p_t)$, we have the estimate

$$\|\gamma_{\nu_Q}(-\nabla p, p_t)\|_{H^{-1/2}(\partial Q)} \le C\big(\|p\|_{L^2(0,T;H_0^1(\Omega))} + \|p_t\|_{L^2(Q)} + \|p_{tt} - \Delta p\|_{\mathcal{M}_b(]0,T[;L^2(\Omega))}\big) ,$$

where C is independent of p. This allows us to define $p_t(0)$ as the restriction of this normal trace to $\Omega \times \{0\}$, i.e., as

$$\gamma_{\nu_Q}(-\nabla p, p_t)|_{\Omega\times\{0\}} = p_t(0) \in H^{-1/2}(\Omega) .$$

Thus we arrive at the following definition of *weak solutions* for the adjoint system given in (7.55).

Definition 7.26 (weak solutions to the Dirichlet adjoint system). *A function p with $(p, p_t) \in L^\infty(0,T;H_0^1(\Omega)) \times L^\infty(0,T;L^2(\Omega))$ and $p_{tt} - \Delta p \in \mathcal{M}_b(]0,T[;L^2(\Omega))$ is a* WEAK SOLUTION *to the Dirichlet adjoint system* (7.55) *if one has the equality*

$$-\int_\Omega p(0)y_1\,dx + \big\langle p_t(0), y_0\big\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} + \big\langle y(\vartheta, y_0, y_1), \mu\big\rangle_{\mathcal{C}([0,T];L^2(\Omega))\times\mathcal{M}_b(]0,T];L^2(\Omega))} - \int_Q p\vartheta\,dxdt = 0$$

for all $(\vartheta, y_0, y_1) \in L^2(Q) \times H_0^1(\Omega) \times L^2(\Omega)$, where $y(\vartheta, y_0, y_1)$ denotes the unique solution to the homogeneous Dirichlet problem in (7.49)*, i.e., to*

$$\begin{cases} y_{tt} - \Delta y = \vartheta & in \;\; Q , \\ y = 0 & in \;\; \Sigma , \\ y(0) = y_0, \quad y_t(0) = y_1 & in \;\; \Omega . \end{cases}$$

Let us observe that, since $(p, p_t) \in L^\infty(0,T;H_0^1(\Omega)) \times L^\infty(0,T;L^2(\Omega))$, we have $p \in \mathcal{C}([0,T];L^2(\Omega))$, and thus the term $\int_\Omega p(0)y_1\,dx$ is meaningful. Furthermore, $p_{tt} - \Delta p \in \mathcal{M}_b(]0,T[;L^2(\Omega))$, and hence $p_t(0) =$

$\gamma_{\nu_Q}(-\nabla p, p_t)|_{\Omega\times\{0\}}$ is well defined in $H^{-1/2}(\Omega)$ due to the discussion right before the definition.

The next important result justifies the existence and uniqueness of weak solutions to the adjoint system (7.55) in the sense of Definition 7.26. Moreover, it provides additional regularity properties that are significant for the proof of the main theorem.

Theorem 7.27 (properties of adjoint arcs in the Dirichlet problem). *The adjoint system* (7.55) *admits a unique weak solution* $p(\cdot)$ *such that*

$$(p, p_t) \in L^\infty\big(0,T; H_0^1(\Omega)\big) \times L^\infty\big(0,T; L^2(\Omega)\big).$$

Furthermore, $p_t \in BV\big([0,T]; H^{-1}(\Omega)\big)$,

$$\frac{\partial p}{\partial \nu} = \gamma_{\nu_Q}(\nabla p, -p_t)|_\Sigma \quad \textit{belongs to} \quad L^2(\Sigma)\ ,$$

$p \in \mathcal{C}_w\big([0,T]; H_0^1(\Omega)\big)$, *and*

$$p_t(\tau) \in L^2(\Omega) \quad \textit{for all} \quad \tau \in \Big\{t \in [0,T]\ \Big|\ \mu(\Omega \times \{t\}) = 0\Big\}\ .$$

In particular, we have $p_t(0) \in L^2(\Omega)$.

Proof. First observe that the fulfillment of the equality in the theorem for $(p, p_t) \in L^\infty\big(0,T; H_0^1(\Omega)\big) \times L^\infty\big(0,T; L^2(\Omega)\big)$ with $\mu = 0$ obviously implies that $p = 0$. Thus system (7.55) admits *at most one* solution in the sense of Definition 7.26. We therefore need to justify the *existence* of weak solutions with the additional *regularity properties* listed in the theorem.

Let $\{\mu_k\}$ be a sequence in $L^1\big(0,T; L^2(\Omega)\big)$ satisfying the relations

$$\big\|\mu_k\big\|_{L^1(0,T;L^2(\Omega))} = \big\|\mu|_{]0,T[}\big\|_{\mathcal{M}_b(]0,T[;L^2(\Omega))} \quad \text{and}$$

$$\lim_{k\to\infty}\int_Q y\mu_k\,dxdt = \big\langle y, \mu|_{]0,T[}\big\rangle_{\mathcal{C}([0,T];L^2(\Omega))\times\mathcal{M}_b(]0,T[;L^2(\Omega))}$$

if $y \in \mathcal{C}\big([0,T]; L^2(\Omega)\big)$. Denote by p_k the (unique) solution to

$$\begin{cases} p_{tt} - \Delta p = \mu_k & \text{in } Q\ ,\\ p = 0 & \text{in } \Sigma\ ,\\ p(T) = 0, \quad p_t(T) = -\mu|_{\Omega\times\{T\}} & \text{in } \Omega\ . \end{cases} \tag{7.56}$$

Employing the result of Theorem 2.1 from the afore-mentioned paper by Lasiecka, Lions and Triggiani [740], we have the estimate

$$\|p_k\|_{L^\infty(0,T;H_0^1(\Omega))} + \|p_{kt}\|_{L^\infty(0,T;L^2(\Omega))} + \left\|\frac{\partial p_k}{\partial \nu}\right\|_{L^2(\Sigma)} + \|p_k(0)\|_{H^1(\Omega)}$$

$$+\|p_{kt}(0)\|_{L^2(\Omega)} \le C\|\mu\|_{\mathcal{M}_b(]0,T];L^2(\Omega))}$$

with a constant C independent of k. It follows from (7.56) that the distribution derivative of p_{kt} with respect to t can be represented in the form

$$p_{ktt} = \pi_k + \mu_k \in L^\infty\big(0,T;H^{-1}(\Omega)\big) + \mathcal{M}_b\big(]0,T[;L^2(\Omega)\big)$$

$$\subset \mathcal{M}_b\big(]0,T[;H^{-1}(\Omega)\big) ,$$

where π_k is defined by

$$\langle \pi_k, y\rangle_{L^\infty(0,T;H^{-1}(\Omega))\times L^1(0,T;H_0^1(\Omega))} := -\int_Q \nabla p_k \nabla y \, dxdt .$$

Therefore the sequence $\{p_{ktt}\}$ is bounded in $\mathcal{M}_b\big(]0,T[;H^{-1}(\Omega)\big)$, and hence the corresponding one $\{p_{kt}\}$ is bounded in $BV\big([0,T];H^{-1}(\Omega)\big)$. Then there are $p \in L^\infty\big(0,T;H_0^1(\Omega)\big)$ with $p_t \in BV\big([0,T];H^{-1}(\Omega)\big)$ and a subsequence of $\{p_k\}$ such that $p_k \to p$ in the weak* topology of $L^\infty\big(0,T;H_0^1(\Omega)\big)$ and that $p_{kt} \to p_t$ in the weak* topology of $L^\infty\big(0,T;L^2(\Omega)\big)$ as $k \to \infty$. Since the sequence $\{\gamma_{\nu_Q}(-\nabla p_k, p_{kt})\}$ is bounded in $L^2(\partial Q)$, we may also suppose the weak convergence

$$\gamma_{\nu_Q}(-\nabla p_k, p_{kt}) \to \gamma_{\nu_Q}(-\nabla p, p_t) \text{ weakly in } L^2(\partial Q) .$$

On the other hand, $\gamma_{\nu_Q}(-\nabla p_k, p_{kt})|_{\Omega\times\{T\}} = \mu|_{\Omega\times\{T\}}$ and the sequence of

$$\gamma_{\nu_Q}(\nabla p_k, -p_{kt})|_\Sigma = \frac{\partial p_k}{\partial \nu}$$

is bounded in $L^2(\Sigma)$. Thus

$$\gamma_{\nu_Q}(\nabla p_k, -p_{kt})|_\Sigma \to \gamma_{\nu_Q}(\nabla p, -p_t)|_\Sigma = \frac{\partial p}{\partial \nu} \quad \text{and}$$

$$\gamma_{\nu_Q}(-\nabla p_n, p_{nt})|_{\Omega\times\{0\}} = p_{nt}(0) \to \gamma_{\nu_Q}(-\nabla p, p_t)|_{\Omega\times\{0\}} = p_t(0)$$

in the weak topology of $L^2(\Sigma)$ and $L^2(\Omega)$, respectively. Now passing to the limit as $k \to \infty$ in the equality

$$-\langle p_k(0), y_1\rangle_{L^2(\Omega)} + \langle p_{kt}(0), y_0\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)}$$

$$+\langle y(\vartheta, y_0, y_1), \mu_k\rangle_{\mathcal{C}([0,T];L^2(\Omega))\times\mathcal{M}_b(]0,T];L^2(\Omega))} - \langle p_k, \vartheta\rangle_{L^2(Q)} = 0 ,$$

we conclude that $p(\cdot)$ is the desired weak solution to the adjoint system (7.55) satisfying all but the last displayed relations in the theorem.

To prove the remaining property, suppose that $\mu(\Omega \times \{t\}) = 0$ for some $t \in [0, T]$. Then considering the normal trace of $(-\nabla p, p_t)$ on $\partial(\Omega \times]0, t[)$ as above, we derive the equality

$$\gamma_{\nu_{\Omega \times]0,t[}}(-\nabla p, p_t)|_{\Omega \times \{t\}} = p_t(0) \in L^2(\Omega) ,$$

which completes the proof of the theorem. △

Finally in this section, let us present a useful limiting consequence of Theorem 7.27 that ensures a Green-type relationship between solutions of the adjoint system (7.55) and the original arcs belonging to the space Y of admissible state functions (7.54).

Theorem 7.28 (Green formula for the Dirichlet hyperbolic problem). *Given a measure $\mu \in \mathcal{M}_b(]0, T]; L^2(\Omega))$, consider the unique solution $p(\cdot)$ to the adjoint system* (7.55). *Then for every admissible state function $y \in Y$, the adjoint arc $p(\cdot)$ satisfies the following Green formula*

$$\big\langle y, \mu \big\rangle_{C([0,T];L^2(\Omega)) \times \mathcal{M}_b(]0,T];L^2(\Omega))} - \big\langle p, y_{tt} - \Delta y \big\rangle_{L^\infty(0,T;H_0^1(\Omega)) \times L^1(0,T;H^{-1}(\Omega))}$$

$$= - \int_\Omega y(0) p_t(0)\, dx + \big\langle y_t(0), p(0) \big\rangle_{H^{-1}(\Omega) \times H_0^1(\Omega)} - \int_\Sigma y \frac{\partial p}{\partial \nu}\, dsdt .$$

Proof. As established in Theorem 7.27, the above Green formula holds for the solutions p_k to the approximating adjoint system (7.56). Passing there to the limit as $k \to \infty$, we arrive at the required result. △

7.3.4 Proof of Optimality Conditions

This subsection is devoted to the proof of the main result of Sect. 7.3 formulated in Theorem 7.23. We employ the following strategy:

(a) reduce (DP) to a general optimization/*mathematical programming* problem in Banach spaces in the presence of *geometric* and *operator constraints*, for which necessary optimality conditions of a generalized *Lagrange multiplier* type are known, and then

(b) express the latter optimality conditions and the assumptions under which they hold in terms of the initial data of the original Dirichlet control problem (DP).

Besides the general optimization theory of Chap. 5, the proof is essentially based on the specific results obtained in the preceding subsection for hyperbolic systems under consideration that employ in turn the regularity results of Theorem 7.25. The general optimization problem in Banach spaces, to which we reduce (DP), is called for convenience the *abstract control problem* and is written as follows: minimize

$$\varphi(z,w) \text{ subject to } z \in Z,\quad w \in W_{ad},\quad f_1(z,w)=0,\quad f_2(w) \in \Xi, \tag{7.57}$$

where $\varphi: Z \times W \to I\!R$, $f_1: Z \times W \to Z_1$, $f_2: Z \to Z_2$, where the sets $W_{ad} \subset W$ and $\Xi \subset Z_2$ are closed and convex, and where the spaces Z, Z_1, Z_2, W are Banach with W being separable.

Observe that the "abstract control" problem (7.57) is of a mathematical programming type with geometric and operator constraints studied in Subsect. 5.1.2. Assume in what follows the *Fréchet differentiability* of the cost functional and the *strict differentiability* of the operator constraints with the *surjective* of the strict derivative operator. Taking also into account the *special structure* of problem (7.57) and the *convexity* of the sets describing the geometric constraints, we can derive necessary optimality conditions for (7.57) as an elaboration of Theorem 5.11(ii), where the domain space is arbitrarily *Banach* due to the smoothness and convexity assumptions made. A direct derivation of the next theorem from the viewpoint of convex optimization with smooth operator constraints was given by Alibert and Raymond [9].

Theorem 7.29 (necessary conditions for abstract control problems). *Let $(\bar z,\bar w)$ be an optimal solution to problem* (7.57). *Assume that φ is Fréchet differentiable at $(\bar z,\bar w)$ while f_1 and f_2 are strictly differentiable at $\bar z$ and $(\bar z,\bar w)$, respectively, with the surjective partial derivative $f'_{1z}(\bar z,\bar w): Z \to Z_1$, and that* $\operatorname{int}\Xi \neq \emptyset$. *Then there are adjoint elements $(p,\mu,\lambda) \in Z_1^* \times Z_2^* \times I\!R^+$ such that $(\lambda,\mu) \neq 0$ and the following conditions hold:*

$$\lambda\varphi'_z(\bar z,\bar w)z + \big\langle p, f'_{1z}(\bar z,\bar w)z\big\rangle + \big\langle \mu, f'_2(\bar z)z\big\rangle = 0 \;\; \textit{for every} \;\; z \in Z\,,$$

$$\big\langle \mu, z - f_2(\bar z)\big\rangle \le 0 \;\; \textit{for every} \;\; z \in \Xi, \quad \textit{and}$$

$$\lambda\varphi'_w(\bar z,\bar w)(w-\bar w) + \big\langle p, f'_{1w}(\bar z,\bar w)(w-\bar w)\big\rangle \ge 0 \;\; \textit{for every} \;\; w \in W_{ad}\,.$$

If in addition

$$f'_{1z}(\bar z,\bar w)z_0 + f'_{1w}(\bar y,\bar w)w_0 = 0 \;\; \textit{and} \;\; f_2(\bar z) + f'_2(\bar z)z_0 \in \operatorname{int}\Xi$$

for some $w_0 \in (W_{ad} - \bar w)$ and $z_0 \in Z$, then the above conditions are fulfilled in the normal form, i.e., with $\lambda = 1$.

Now we complete this section by proving the formulated necessary optimality conditions for the original Dirichlet control problem (DP).

Proof of Theorem 7.23. Let $(\bar y,\bar u) \in Y \times U_{ad}$ be the reference optimal solution to (DP). We are going to reduce (DP) to the mathematical programming problem (7.57) considering in Theorem 7.29. To proceed, put:

$$Z := Y,\quad (z,w) := (y,u),\quad W := L^2(\Sigma),\quad W_{ad} := U_{ad},\quad \Xi := \Theta\,,$$

$$Z_1 := L^1\big(0,T;H^{-1}(\Omega)\big)\times L^2(\Sigma)\times L^2(\Omega)\times H^{-1}(\Omega),\ Z_2 := \mathcal{C}\big([0,T];L^2(\Omega)\big)\ ,$$

$$\varphi(y,u) := J(y,u),\quad f_1(y,u) := \Big(y_{tt} - \varDelta y - \vartheta,\, y|_\Sigma - u,\, y(0) - y_0,\, y_t(0) - y_1\Big)\ ,$$

and $f_2(y) := y$. By assumptions (H5)–(H7) the cost functional φ is Fréchet differentiable at $(\bar{y},\bar{u})$, the mapping f_1 is strictly differentiable at $(\bar{y},\bar{u})$, and

$$\varphi'(\bar{y},\bar{u})(y,u) = \int_\Omega f_y'\big(x,\bar{y}(T)\big)y(T)\,dx + \int_Q g_y'(x,t,\bar{y})y\,dxdt + \int_\Sigma h_u'(s,t,\bar{u})u\,dsdt\ ,$$

$$f_1'(\bar{y},\bar{u})(y,u) = f_{1y}'(\bar{y},\bar{u})y + f_{1u}'(\bar{y},\bar{u})u\ ,$$

$$f_{1y}'(\bar{y},\bar{u})y = \Big(y_{tt} - \varDelta y,\, y|_\Sigma,\, y(0),\, y_t(0)\Big)\ ,$$

$$f_{1u}'(\bar{y},\bar{u})u = \big(0,-u,0,0\big)\ \text{ for every }\ (y,u)\in Y\times L^2(\Sigma)\ .$$

Furthermore, it follows from Theorem 7.25 that the linear continuous operator $f_{1y}'(\bar{y},\bar{u})$ is *surjective* from Y to $L^1\big(0,T;H^{-1}(\Omega)\big)\times L^2(\Sigma)\times L^2(\Omega)\times H^{-1}(\Omega)$. Thus all the assumptions of Theorem 7.29 are satisfied.

Applying the latter theorem, find $\lambda \in I\!R^+$, $(\bar{p},\widetilde{p},\widehat{p},\breve{p}) \in L^\infty\big(0,T;H_0^1(\Omega)\big)\times L^2(\Sigma)\times L^2(\Omega)\times H_0^1(\Omega)$, and $\mu\in\mathcal{M}\big([0,T];L^2(\Omega)\big)$ with $(\lambda,\mu)\neq 0$ satisfying the following conditions:

$$\begin{aligned}&\int_\Omega \lambda f_y'\big(x,\bar{y}(T)\big)y(T)\,dx + \int_Q \lambda g_y'(x,t,\bar{y})y\,dxdt + \big\langle \bar{p}, y_{tt} - \varDelta y\big\rangle\\ &+\int_\Sigma \widetilde{p}y\,dsdt + \big\langle \widehat{p}, y(0)\big\rangle + \big\langle \breve{p}, y_t(0)\big\rangle\\ &+\big\langle \mu, y\big\rangle_{\mathcal{M}([0,T];L^2(\Omega))\times\mathcal{C}([0,T];L^2(\Omega))} = 0\end{aligned}\tag{7.58}$$

for every y from the space of admissible state functions Y defined in (7.54),

$$\big\langle \mu, z-\bar{y}\big\rangle_{\mathcal{M}([0,T];L^2(\Omega))\times\mathcal{C}([0,T];L^2(\Omega))} \le 0\ \text{ for every }\ z\in\Theta,\tag{7.59}$$

$$\int_\Sigma \big(\lambda h_u'(x,\bar{y},\bar{u}) + \widetilde{p}\big)(u-\bar{u})dx \ge 0\ \text{ for every }\ u\in U_{ad}\ .\tag{7.60}$$

It follows from (7.59) and (H4) that $\mu|_{\Omega\times\{0\}} = 0$, and thus μ can be identified with a measure belonging to $\mathcal{M}_b\big(]0,T];L^2(\Omega)\big)$. Furthermore, Theorem 7.27

ensures the existence of the unique weak solution $p(\cdot)$ to the adjoint system (7.55) with $(p, p_t) \in L^\infty\big(0,T;H^1(\Omega)\big) \times L^\infty\big(0,T;L^2(\Omega)\big)$. Then the Green formula of Theorem 7.28 and the optimality condition (7.58) yield that

$$\big\langle p+\bar p, y_{tt}-\Delta y\big\rangle + \int_\Sigma \Big(\widetilde p - \frac{\partial p}{\partial \nu}\Big) y\, dsdt$$

$$+\int_\Omega \big(\widehat p - p_t(0)\big)y(0)dx + \int_\Omega \big(\breve p + p(0)\big)y_t(0)dx = 0$$

for every $y \in Y$. Since the mapping $y \longrightarrow \Big(y_{tt}-\Delta y, y|_\Sigma, y(0), y_t(0)\Big)$ is *surjective* from Y to $L^1\big(0,T;H^{-1}(\Omega)\big)\times L^2(\Sigma)\times L^2(\Omega)\times H^{-1}(\Omega)$, the above variational condition gives

$$p = -\bar p \in L^\infty\big(0,T;H_0^1(\Omega)\big), \quad \frac{\partial p}{\partial \nu} = \widetilde p \in L^2(\Sigma)\ ,$$

$$p_t(0) = \widehat p \in L^2(\Omega), \quad \text{and} \quad p(0) = -\breve p \in H_0^1(\Omega)\ .$$

Thus the necessary optimality conditions (7.58)–(7.60) of Theorem 7.29 imply the desired optimality condition (7.50)–(7.52) of Theorem 7.23. Observe finally that the qualification condition (7.53) of Theorem 7.23 reduces to the one in Theorem 7.29, which ensures the normality $\lambda = 1$ and completes the proof of the main theorem. △

Remark 7.30 (SNC state constraints). It follows from the above proof that the assumption on $\operatorname{int}\Theta \neq \emptyset$ may be substantially relaxed by replacing it with the *SNC property* of the convex state constraint set Θ; cf. Theorem 5.11(ii) and also the proof of Theorem 2.51(ii), which shows that merely *sequential* (vs. topological) normal compactness conditions are needed to justify nontriviality via limiting procedures in the general Banach space setting. The same relaxation of the interiority assumption is possible for the Neumann boundary control problems from Sect. 7.2. Note that, in the case of convex subsets of Banach spaces, the SNC property automatically holds for *finite-codimensional* sets with nonempty *relative interiors*; see Theorem 1.21. Moreover, the SNC property is generally *weaker* that the above requirements, which are actually *equivalent* to the CEL (topological) property of convex sets; see Remark 1.27 and Example 3.6. Observe finally that the usage of Theorem 5.11(ii) in the above proof makes it possible to *relax the differentiability assumptions* on the integrands in the Dirichlet boundary control problem (DP) under consideration.

7.4 Minimax Control of Parabolic Systems with Pointwise State Constraints

The last section of this chapter concerns *parabolic control systems* with the *Dirichlet boundary conditions* subject to *pointwise state constraints*. We focus on Dirichlet boundary controls for the following two major reasons. *First*, the Dirichlet case for parabolic systems is much *more challenging* and involved in comparison with the Neumann one; this is substantially *different from hyperbolic systems*, where the Neumann case is considerable more difficult providing nevertheless more regularity; see the preceding two sections with the subsequent comments to them. The *second* reason is that the author's original interest to studying control problems for parabolic systems was primary motivated by practical applications to some *environmental problems* related to automatic regulating the soil water regime; see Mordukhovich [898, 905]. The physical phenomena and engineering constructions in these practical problems lead to mathematical models involving parabolic equations with Dirichlet boundary controls. Furthermore, control processes in the afore-mentioned systems are unavoidably conducted under *uncertain perturbations*, and the most natural optimization criterion is *minimax*. Taking this into account, we consider in this section a *minimax optimal control* problem for linear parabolic systems with controls acting in the Dirichlet boundary conditions in the presence of uncertain distributed perturbations and *hard/poitwise* constraints on the state and control functions. Our primal goal is to establish an *existence theorem* for minimax solutions and to derive *necessary optimality* (as well as *suboptimality*) conditions for open-loop controls under the *worst perturbations*. Finally, we briefly discuss (and refer the reader to the corresponding publications for more details) some issues related to *minimax design* of *closed-loop* parabolic control systems, which involve *feedback controls* in the Dirichlet boundary conditions. Including this material is beyond the scope of the present book.

The minimax control problem under consideration is essentially *nonsmooth* and requires special methods for its variational analysis. To conduct such an analysis, we systematically use *smooth approximation procedures.* Actually we split the original minimax problem into two interrelated optimal control problems for distributed perturbations and boundary controls with *moving* state constraints. Then we approximate state constraints in each of these problems by effective penalizations involving $\mathcal{C}^\infty$-approximations of *maximal monotone* operators. We establish *strong convergence* results for such processes and obtain characterizations of optimal solutions to the approximating problems. Finally imposing proper constraint qualifications, we arrive at necessary optimality conditions for the worst perturbations and optimal controls in the original state-constrained minimax problem.

The most involved part of our variational analysis concerns a state-constrained *Dirichlet boundary control* problem under the worst disturbances. The main complications arise in this case from the presence of *pointwise* state constraints simultaneously with *hard constraints* on L^∞ controls acting in the Dirichlet boundary conditions. It is well known that the latter conditions provide the *lowest regularity properties* of solutions and are related to *unbounded* operators in the framework of variational inequalities. We develop an efficient analysis based an smooth approximation procedures and properties of *mild* solutions to the Dirichlet boundary control problem for parabolic equations.

7.4.1 Problem Formulation and Splitting

Consider the following *parabolic* system

$$\begin{cases} \dfrac{\partial y}{\partial t} + Ay = Bw + \vartheta \ \text{ a.e. in } \ Q := (0,T) \times \Omega \ , \\ y(x,0) = y_0(x), \qquad x \in \Omega \ , \\ y(s,t) = u(s,t), \qquad (s,t) \in \Sigma := (0,T] \times \Gamma \ , \end{cases} \tag{7.61}$$

with *pointwise/hard constraints* on *state* trajectories $y(\cdot)$, uncertain *perturbations/disturbances* $w(\cdot)$, and *Dirichlet boundary controls* $u(\cdot)$ given by:

$$a \le y(x,t) \le b \ \text{ a.e. } \ (x,t) \in Q \ , \tag{7.62}$$

$$c \le w(x,t) \le d \ \text{ a.e. } \ (x,t) \in Q \ , \tag{7.63}$$

$$\mu \le u(s,t) \le \nu \ \text{ a.e. } \ (s,t) \in \Sigma \ , \tag{7.64}$$

where $\Omega \subset \mathbf{R}^n$ is a bounded open set with sufficiently smooth boundary Γ and where each of the intervals $[a,b]$, $[c,d]$, and $[\mu,\nu]$ contains 0.

Let $X := L^2(\Omega; I\!R)$, $U := L^2(\Gamma; I\!R)$, and $W := L^2(\Omega; R)$ be, respectively, spaces of states, controls, and disturbances. In what follows we remove $I\!R$ from the latter and similar space notation for real-valued functions. Denote

$$U_{ad} := \big\{ u \in L^p(0,T;U) \ \big| \ \mu \le u(s,t) \le \nu \ \text{ a.e. } \ (s,t) \in \Sigma \big\}$$

the set of admissible controls, where $L^p(0,T;U)$ is the space of U-valued functions $u(\cdot) = u(s,\cdot)$ on $[0,T]$ with the norm

$$\|u\|_{L^p(0,T;U)} := \Big(\int_0^T \|u(t)\|_U^p dt \Big)^{1/p} = \Big(\int_0^T \Big(\int_\Gamma |u(s,t)|^2 ds)^{p/2} dt \Big)^{1/p} .$$

Similarly we define the set of admissible disturbances

$$W_{ad} := \big\{ w \in L^2(0,T;W) \ \big| \ c \le w(x,t) \le d \ \text{ a.e. } \ (x,t) \in Q \big\} .$$

A pair $(u,w) \in U_{ad} \times W_{ad}$ is called a *feasible solution* to system (7.61) if the corresponding trajectory $y(\cdot)$ satisfies the state constraints (7.62). We always assume that problem (7.61)–(7.64) admits at least one feasible pair (u,w).

Although the constraint sets W_{ad} and U_{ad} are *essentially bounded*, i.e., $W_{ad} \subset L^\infty(Q)$ and $U_{ad} \subset L^\infty(\Sigma)$, we prefer considering them as subsets of the larger spaces $L^2(0,T;W)$ and $L^p(0,T;U)$, respectively, with finite p *sufficiently big*. The reason is that it allows us to take advantages of the *reflexivity* of the latter spaces and of the *differentiability* of their norms away from the origin to efficiently perform our variational analysis.

Throughout this section paper we impose the following *standing assumptions* on the parabolic system under consideration:

(H1) The linear operator

$$A := -\sum_{i,j=1}^{n} \frac{\partial}{\partial x_i}\Big(a_{ij}(x)\frac{\partial}{\partial x_j}\Big) + \sum_{i=1}^{n} a_i(x)\frac{\partial}{\partial x_i} + a_0(x)$$

is *strongly uniformly elliptic* on Ω with real-valued smooth coefficients; i.e., there is $\beta_0 > 0$ such that

$$\sum_{i,j=1}^{n} a_{ij}(x)v_i v_j \geq \beta_0 \sum_{i=1}^{n} v_i^2 \quad \text{for all} \quad x \in \Omega \quad \text{and} \quad (v_1, \ldots, v_n) \in I\!R^n \ .$$

(H2) $\vartheta \in L^\infty(Q)$ and $y_0(x) \in H_0^1(\Omega) \cap H^2(\Omega)$ with

$$a \leq y_0(x) \leq b \ \text{ a.e. } \ x \in \Omega .$$

(H3) $B\colon L^2(0,T;W) \to L^2(0,T;X)$ is a bounded linear operator.

We may always assume that the operator $-A$ generates a *strongly continuous analytic semigroup* $S(\cdot)$ on X satisfying the exponential estimate

$$\|S(t)\| \leq M_1 e^{-\omega t} \tag{7.65}$$

with some constants $\omega > 0$ and $M_1 > 0$, where $\|\cdot\|$ denotes the standard operator norm from X to X. Otherwise, it is a standard procedure to construct a stable translation of the form $\widetilde{A} := A + \widetilde{\omega} I$ that possesses such properties.

Note that since $w \in L^2(0,T;W)$ and $u \in L^p(0,T;U)$, the parabolic system (7.61) may *not* have strong or classical solutions for some $(u,w) \in U_{ad} \times W_{ad}$. In this case, principal difficulties come from *discontinuous* controls in the Dirichlet boundary conditions. Taking advantages of the semigroup approach to parabolic equations, we are going to use for our analysis a concept of *mild solutions* to Dirichlet boundary problems.

Consider the so-called *Dirichlet map* D defined by $y = Du$, where $y(\cdot)$ satisfies the homogeneous elliptic equation

$$\begin{cases} -Ay = 0 & \text{in} \ \ Q \ , \\ y(s,t) = u(s,t), & (s,t) \in \Sigma \ . \end{cases}$$

It is well known (see, e.g., Lions and Magenes [794]) that the Dirichlet operator

$$D\colon L^2(\Gamma) \to \mathcal{D}(A^{1/4-\delta}) = H^{1/2-2\delta}(\Omega), \quad 0 < \delta \leq 1/4 \ , \tag{7.66}$$

is linear and continuous, where $\mathcal{D}$ stands for the domain as usual.

Definition 7.31 (mild solutions to Dirichlet parabolic systems). *A continuous mapping* $y : [0,T] \to X$ *is a* MILD SOLUTION *to system* (7.61) *corresponding to* $(u,w) \in L^p(0,T;U) \times L^2(0,T;W)$ *if for all* $t \in [0,T]$ *one has the representation*

$$y(t) = S(t)y_0 + \int_0^t S(t-\tau)\Big(Bw(\tau) + \vartheta(\tau)\Big)d\tau + \int_0^t AS(t-\tau)Du(\tau)d\tau$$

$$+S(t)y_0 + \int_0^t S(t-\tau)\Big(Bw(\tau) + \vartheta(\tau)\Big)d\tau$$

$$+\int_0^t A^{3/4+\delta}S(t-\tau)A^{1/4-\delta}Du(\tau)d\tau\ ,$$

where D *is the Dirichlet operator defined in* (7.66) *with some* $\delta \in (0,1/4]$.

The reader can find more information about mild solutions to Dirichlet parabolic systems in the paper by Lasiecka and Triggiani [743] and the references therein. Note, in particular, that the assumptions made above ensure the *existence* and *uniqueness* of mild solutions to (7.61) for any $w \in L^2(0,T;W)$ and $u \in L^p(0,T;U)$ provided that $p > 0$ is *sufficiently large*. Observe also that while the X-valued function $y(t)$ from Definition 7.31 is continuous by definition, the real-valued function $y(x,t)$ of two variables is merely *measurable*, since $X = L^2(\Omega)$. This significantly distinguishes mild solutions from other concepts of solutions to parabolic equations. The mild solution approach allows us to deal with *irregular* (measurable) data of parabolic systems involving the Dirichlet boundary conditions considered in this section. On the other hand, the absence of continuity creates substantial difficulties that we are going to overcome in what follows.

Note that δ in Definition 7.31 may be *any* fixed number from the interval $(0,1/4]$. Although the first representation of $y(t)$ in Definition 7.31 doesn't depend on δ at all, this number explicitly appears in some estimates below that are *the better the closer* δ *is to zero.*

Now we introduce the cost functional

$$\begin{aligned} J(u,w) := \int_Q g\big(x,t,y(x,t)\big)\,dxdt + \int_Q f\big(x,t,w(x,t)\big)\,dxdt \\ + \int_\Sigma h\big(s,t,u(s,t)\big)\,dsdt\ , \end{aligned} \tag{7.67}$$

where $y(\cdot)$ is a trajectory (mild solution) to system (7.61) generated by $u(\cdot)$ and $w(\cdot)$. We always suppose that functional (7.67) is well defined and finite for all admissible processes in (7.61)–(7.64). Some additional assumptions on integrands g, f, and h are imposed in Subsects. 7.4.2–7.4.4. The *minimax control problem* under consideration in this section as follow:

(P) find an admissible control $\bar{u} \in U_{ad}$ and a disturbance $\bar{w} \in W_{ad}$ such that $(\bar{u}, \bar{w})$ is a *saddle point* for the functional $J(u, w)$ subject to system (7.61) and state constraints (7.62).

This means, by the definition of saddle points, that

$$J(\bar{u}, w) \le J(\bar{u}, \bar{w}) \le J(u, \bar{w}) \ \text{ for all } \ u \in U_{ad} \ \text{ and } \ w \in W_{ad} \tag{7.68}$$

under conditions (7.61) and (7.62). Such a pair $(\bar{u}, \bar{w})$ is called an *optimal solution* to the minimax problem (P).

For studying optimal solutions to problem (P) we employ the following *splitting procedure*, which significantly exploits the *linearity* of system (7.61). Namely, split the original system (7.61) into two subsystems with separated disturbances and boundary controls. The first system

$$\begin{cases} \dfrac{\partial y_1}{\partial t} + Ay_1 = Bw + \vartheta \ \text{ a.e. in } Q\ , \\ y_1(x,0) = y_0(x), \qquad x \in \Omega\ , \\ y_1(s,t) = 0, \qquad (s,t) \in \Sigma\ , \end{cases} \tag{7.69}$$

has zero (homogeneous) boundary conditions and depends only on disturbances. The second one

$$\begin{cases} \dfrac{\partial y_2}{\partial t} + Ay_2 = 0 \qquad \text{a.e. in } Q\ , \\ y_2(x,0) = 0, \qquad x \in \Omega\ , \\ y_2(s,t) = u(s,t),\ (s,t) \in \Sigma\ , \end{cases} \tag{7.70}$$

is generated by boundary controls and doesn't involve disturbances. It is easy to see that for any $(u, w) \in U_{ad} \times W_{ad}$ one has

$$y(x,t) = y_1(x,t) + y_2(x,t) \ \text{ whenever } \ (x,t) \in Q$$

for the corresponding trajectories of systems (7.61), (7.69), and (7.70).

Let $\bar{y}_1$ and $\bar{y}_2$ be the (unique) trajectories of systems (7.69) and (7.70), respectively, corresponding to $\bar{w}$ and $\bar{u}$. Consider the cost functionals

$$J_1(w, y_1) := \int_Q \Big[g\big(x, t, y_1(x,t) + \bar{y}_2(x,t)\big) + f\big(x, t, w(x,t)\big) \Big]\, dxdt$$

for disturbances $w(\cdot)$ and

$$J_2(u, y_2) := \int_Q g\big(x, t, \bar{y}_1(x,t) + y_2(x,t)\big)\, dxdt + \int_\Sigma h\big(s, t, u(s,t)\big)\, dsdt$$

for boundary controls $u(\cdot)$. Now define two optimization problems corresponding to the cost functionals introduced. The first one is:

$$(P_1) \;\; \text{maximize } J_1(w, y_1) \text{ over } w \in W_{ad} \text{ subject to (7.69) and}$$

$$a - \bar{y}_2(x,t) \le y_1(x,t) \le b - \bar{y}_2(x,t) \;\; \text{a.e. } (x,t) \in Q \; .$$

The second problem is:

$$(P_2) \;\; \text{minimize } J_2(u, y_2) \text{ over } u \in U_{ad} \text{ subject to (7.70) and}$$

$$a - \bar{y}_1(x,t) \le y_2(x,t) \le b - \bar{y}_1(x,t) \;\; \text{a.e. } (x,t) \in Q \; .$$

The next assertion shows that the original minimax problem (P) can be *split* into the two state-constrained dynamic optimization problems (P_1) and (P_2) *separated on disturbances and controls.*

Proposition 7.32 (splitting the minimax problem). *Let* $(\bar{u}, \bar{w})$ *be an optimal solution to problem (P), and let* $\bar{y}_1$ *and* $\bar{y}_2$ *be the corresponding trajectories to systems* (7.69) *and* (7.70)*, respectively. Then* $\bar{w}$ *solves problem* (P_1) *and* $\bar{u}$ *solves problem* (P_2).

Proof. From the above relationship $y(\cdot) = y_1(\cdot) + y_2(\cdot)$ we immediately conclude that $\bar{w}$ is a feasible solution to (P_1), i.e., the corresponding trajectory $\bar{y}_1$ to (7.69) satisfies the state constraints in (P_1). Now the left-hand side of (7.68) implies, due to the structures of the cost functionals J and J_1 in the problems under consideration, that $\bar{w}$ is an optimal solution to (P_1). Arguments for $\bar{u}$ are similar, which completes the proof. △

Thus to obtain necessary conditions for a given optimal solution $(\bar{u}, \bar{w})$ to the minimax problem (P), we consider the two separate problems, (P_1) for $\bar{w}$ and (P_2) for $\bar{u}$, with the connecting *state constraints* in these problems. Note that these constraints depend on (x,t), i.e., they are *moving*. The latter property is essential for studying the original minimax control problem for parabolic systems with uncertain perturbations and irregular boundary controls acting in the Dirichlet boundary conditions.

7.4.2 Properties of Mild Solutions and Minimax Existence Theorem

In this subsection we establish important *regularity* and *convergence* properties of mild solutions to the parabolic system (7.61) needed in what follows, and then we justify the *existence* of minimax optimal solutions.

Let $S(t)$ be an analytic semigroup on X generated by the operator $-A$ and satisfying the exponential estimate (7.65), and let D be the Dirichlet operator with the continuity property (7.66). In what follows we use the estimates

$$\|A^{\delta}D\| \le M_2, \quad \|A^{3/4+\delta}S(t)\| \le \frac{M_3}{t^{3/4+\delta}} \quad \text{for any} \quad \delta \in (0,1/4], \tag{7.71}$$

where $\|\cdot\|$ stands for the corresponding operator norm; see Lasiecka and Triggiani [743] with the references therein. It is clear from Definition 7.31 that the main complications for the study of mild solutions are related to the term involving the Dirichlet map. Consider this term separately via the operator $\mathcal{L}$ from $L^p(0,T;U)$ into $L^r\big(0,T;H^{1/2-\varepsilon}(\Omega)\big)$ defined by

$$\begin{cases} \mathcal{L}u = (\mathcal{L}u)(t) := A\displaystyle\int_0^t S(t-\tau)Du(\tau)\,d\tau \\ \qquad\qquad\qquad = \displaystyle\int_0^t A^{3/4+\delta}S(t-\tau)A^{1/4-\delta}Du(\tau)\,d\tau\ , \end{cases} \tag{7.72}$$

where $p,\ r \in [1,\infty],\ \delta \in (0,1/4]$, and $\varepsilon \in (0,1/2]$. Here the Sobolev space $H^{1/2-\varepsilon}(\Omega) \subset L^2(\Omega) = X$ is equipped with the norm

$$\|y\|_{1/2-\varepsilon} := \|A^{1/4-\varepsilon/2}y\|_X\ .$$

Note that $H^0(\Omega) = L^2(\Omega)$ and that the above norm $\|y\|_{1/2-\varepsilon}$ is stronger than $\|y\|_X$. In the sequel we always take $\delta = \varepsilon/4$ in (7.72) and call $\mathcal{L}$ the *mild solution operator*. Note that this operator is generally *unbounded*. Nevertheless, it enjoys nice *regularity/continuity* properties established next provided that the number p is *sufficiently large*.

Theorem 7.33 (regularity of mild solutions to parabolic Dirichlet systems). *Let $p > 4/\varepsilon$ with some $\varepsilon \in (0,1/2]$. Then $\mathcal{L}u \in \mathcal{C}\big([0,T];H^{1/2-\varepsilon}(\Omega)\big)$ for any $u \in L^p(0,T;U)$. Furthermore, the operator*

$$\mathcal{L}\colon L^p(0,T;U) \to \mathcal{C}([0,T];H^{1/2-\varepsilon}(\Omega))$$

is linear and continuous.

Proof. Obviously $\mathcal{L}$ is linear. To show that $\mathcal{L}$ is continuous, we justify its boundedness; i.e., the estimate

$$\|\mathcal{L}u\|_{\mathcal{C}([0,T];H^{1/2-\varepsilon}(\Omega))} \le K\|u\|_{L^p(0,T;U)} \quad \text{with some} \quad K > 0\ .$$

It follows from (7.71) and (7.72) that, whenever $t \in [0,T]$, one has

$$\begin{aligned} \|(\mathcal{L}u)(t)\|_{1/2-\varepsilon} &= \Big\|\int_0^t A^{1/4-\varepsilon/2}AS(t-\tau)Du(\tau)\,d\tau\Big\|_X \\ &= \Big\|\int_0^t A^{1-\varepsilon/4}S(t-\tau)A^{1/4-\varepsilon/4}Du(\tau)\,d\tau\Big\|_X \\ &\le M_2M_3\int_0^t (t-\tau)^{-(1-\varepsilon/4)}\|u\|_U\,d\tau \\ &\le M_2M_3\Big(\int_0^t (t-\tau)^{-(1-\varepsilon/4)q}\,d\tau\Big)^{1/q}\|u\|_{L^p(0,T;U)}\ , \end{aligned}$$

where $1/p + 1/q = 1$. Since $p > 4/\varepsilon$ yields $q < 4/(4-\varepsilon)$, we get

$$\big\|\big(\mathcal{L}u\big)(t)\big\|_{1/2-\varepsilon} \le M_2 M_3 \Big(\frac{1}{1-(1-\varepsilon/4)q}\Big)^{1/q} t^{(1-(1-\varepsilon/4)q)/q} \|u\|_{L^p(0,T;U)} \,.$$

Prove next that $\mathcal{L}u \in \mathcal{C}\big([0,T]; H^{1/2-\varepsilon}(\Omega)\big)$, i.e., the operator $\big(\mathcal{L}u\big)(t)$ is continuous at any point $t_0 \in [0,T]$ in the norm of $H^{1/2-\varepsilon}(\Omega)$. Indeed, taking $t \ge t_0$ for definiteness, we have

$$\big(\mathcal{L}u\big)(t) - \big(\mathcal{L}u\big)(t_0) = \int_{t_0}^{t} AS(t-\tau)Du(\tau)\,d\tau$$

$$+\big(S(t-t_0) - I\big) \int_0^{t_0} AS(t-\tau)Du(t)\,d\tau \,.$$

The latter implies that

$$\big\|\big(\mathcal{L}u\big)(t) - \big(\mathcal{L}u\big)(t_0)\big\|_{1/2-\varepsilon} \to 0 \;\text{ as }\; t \to t_0$$

due to the above estimate for $\|(\mathcal{L}u)(t)\|_{1/2-\varepsilon}$ and the strong continuity of $S(\cdot)$. Furthermore, from this estimate and the norm definition in $\mathcal{C}\big([0,T]; H^{1/2-\varepsilon}(\Omega)\big)$ we immediately get the required boundedness inequality with

$$K := M_2 M_3 \Big(\frac{1}{1-(1-\varepsilon/4)q}\Big)^{1/q} T^{(1-(1-\varepsilon/4)q)/q} \,.$$

This completes the proof of the theorem. △

Corollary 7.34 (weak continuity of the solution operator). *Let ε and p be chosen as in Theorem 7.33. Then the mild solution operator $\mathcal{L}$ acting from $L^p(0,T;U)$ into $\mathcal{C}\big([0,T]; H^{1/2-\varepsilon}(\Omega)\big)$ is weakly continuous. This implies that for any weak convergent sequence $u_k \to u$ in $L^p(0,T;U)$ one has*

$$\mathcal{L}u_k \to \mathcal{L}u \;\; \textit{weakly in} \;\; \mathcal{C}\big([0,T]; H^{1/2-\varepsilon}(\Omega)\big) \;\; \textit{as} \;\; k \to \infty \,.$$

Proof. It follows from Theorem 7.33 by the standard fact on weak continuity of any linear continuous operator between normed spaces. △

As has been already mentioned, the operator $\mathcal{L}$ from (7.72) plays the key role in the structure of mild solutions from Definition 7.31; for this reason we call it the mild solution operator. It easily follows from the above results that the *strong* (resp. *weak*) convergence of boundary controls in $L^p(0,T;U)$ implies the *strong* (resp. *weak*) convergence of the corresponding trajectories for system (7.61) in the space $\mathcal{C}\big([0,T]; H^{1/2-\varepsilon}(\Omega)\big)$ whenever p is sufficiently large. Observe that if the term with $\mathcal{L}$ disappears in Definition 7.31, i.e., in the case of $u = 0$ in (7.61), mild solutions for (7.61) reduce to standard (strong) solutions in the usual sense. In particular, the *weak* convergence of disturbances $w_k \to w$ in $L^p(0,T;W)$ implies in the latter case the *strong* convergence of the corresponding trajectories $y_k \to y$ in $\mathcal{C}([0,T];X)$ as $k \to \infty$ for any $p \geq 1$.

A specific feature of the original problem (P) and its splitting counterparts is that all the constraints are given in the *hard/pointwise* form via discontinuous real functions of *two* (space and time) variables imposed almost everywhere. At the same time, the semigroup approach applied to the study of these problems operates with continuous time-dependent mappings taking values in functional spaces. To proceed further, we need to establish an appropriate operator convergence that implies the required *a.e. convergence* of state trajectories. The next result, crucial in this direction, gives us what we need for the further variational analysis.

Theorem 7.35 (pointwise convergence of mild solutions). *Let ε and p be chosen as in Theorem* 7.33*. Then the weak convergence of Dirichlet controls $u_k \to u$ in $L^p(0,T;U)$ implies the strong convergence in values of the mild solution operator $\mathcal{L}$ for the original parabolic system, i.e.,*

$$\mathcal{L}u_k \to \mathcal{L}u \ \ \textit{strongly in} \ \ L^2(Q) \ \ \textit{as} \ \ k \to \infty \, .$$

Furthermore, there is a real-valued subsequence of $\{(\mathcal{L}u_k)(x,t)\}$ that converges to $(\mathcal{L}u)(x,t)$ a.e. pointwisely in Q as $k \to \infty$.

Proof. It follows from the weak convergence result of Corollary 7.34 that

$$\big(\mathcal{L}u_k\big)(\cdot,t) \to \big(\mathcal{L}u\big)(\cdot,t) \ \text{ weakly in } \ H^{1/2-\varepsilon}(\Omega) \ \text{ for each } \ t \in [0,T]$$

and also that the sequence $\{\mathcal{L}u_k\}$ is bounded in $\mathcal{C}\big([0,T];H^{1/2-\varepsilon}(\Omega)\big)$. Moreover, the classical embedding result ensures that the embedding of $H^{1/2-\varepsilon}(\Omega)$ into X is *compact*; see, e.g., Lions and Magenes [794, Theorem 16.1]. This yields the *strong convergence*

$$\big(\mathcal{L}u_k\big)(t,\cdot) \to \big(\mathcal{L}u\big)(t,\cdot) \ \text{ in } \ X \ \text{ for each } \ t \in [0,T] \, .$$

Thus we get the following conclusions:

(i) The sequence $\{(\mathcal{L}u_k)(t,\cdot)\}$ is bounded in X, i.e., there is $M \geq 0$ providing the estimate

$$\|\big(\mathcal{L}u_k\big)(t)\|_X \leq M \ \text{ for all } \ t \in [0,T] \ \text{ and } \ k \in I\!N \, .$$

(ii) One has the strong convergence

$$\big\|\big(\mathcal{L}u_k\big)(t) - \big(\mathcal{L}u\big)(t)\big\|_X \to 0 \ \text{ for every } \ t \in [0,T] \ \text{ as } \ k \to \infty \, .$$

Consider now a sequence of real-valued nonnegative functions φ_k on $[0,T]$ defined by the integral

$$\varphi_k(t) := \int_\Omega \Big|\big(\mathcal{L}u_k\big)(x,t) - \big(\mathcal{L}u\big)(x,t)\Big|^2\,dx \quad \text{whenever} \quad t \in [0,T]\,.$$

Then (i) and (ii) imply, respectively, that the functions φ_k are uniformly bounded on $[0,T]$ and $\varphi_k(t) \to 0$ pointwisely in $[0,T]$ as $k \to \infty$. Employing the Lebesgue dominated convergence theorem, we arrive at

$$\int_0^T \varphi_k(t)\,dt \to 0 \quad \text{as} \quad k \to \infty\,,$$

which ensures the strong operator convergence of this theorem. The latter finally implies that $\{(\mathcal{L}u_k)(x,t)\}$ contains a subsequence converging to $(\mathcal{L}u)(x,t)$ for a.e. $(x,t) \in Q$. △

The convergence/continuity results derived above are fundamental for the upcoming variational analysis of the problems (P), (P_1), and (P_2) under consideration, which heavily involves the passage to the limit in various *approximation procedures*. In what follows we always assume (without mentioning this explicitly) that the number p is *sufficiently large* to support the convergence results of Theorem 7.35.

To proceed, we impose next the following assumptions on the integrands in the cost functional (7.67) that ensure the appropriate *lower* and *upper semicontinuity* properties of this integral functional with respect to the u and w variables in the corresponding weak topologies.

(H4a) $g(x,t,y)$ satisfies the *Carathéodory condition*, i.e., it is measurable in $(x,t) \in Q$ for all $y \in I\!R$ and continuous in $y \in I\!R$ for a.e. $(x,t) \in Q$. Moreover, there exist a nonnegative function $\eta(\cdot) \in L^1(Q)$ and a constant $\zeta \ge 0$ such that

$$|g(x,t,y)| \le \eta(x,t) + \zeta|y|^2 \quad \text{a.e.} \quad (x,t) \in Q \quad \text{whenever} \quad y \in I\!R\,.$$

(H5a) $f(x,t,w)$ is measurable in $(x,t) \in Q$, continuous and *concave* in $w \in [c,d]$, and for some function $\kappa(\cdot) \in L^1(Q)$ one has

$$f(x,t,w) \le \kappa(x,t) \quad \text{a.e.} \quad (x,t) \in Q \quad \text{whenever} \quad w \in [c,d]\,.$$

(H6a) $h(s,t,u)$ is measurable in $(s,t) \in \Sigma$, continuous and *convex* in $u \in [\mu,\nu]$, and for some function $\gamma(\cdot) \in L^1(\Sigma)$ one has

$$h(s,t,u) \ge \gamma(s,t) \quad \text{a.e.} \quad (s,t) \in \Sigma \quad \text{whenever} \quad u \in [\mu,\nu]\,.$$

Now we are ready to establish the existence theorem for minimax optimal solutions to the parabolic system under consideration.

Theorem 7.36 (existence of minimax solutions). *Let the assumptions* (H1)–(H3) *and* (H4a)–(H6a) *be fulfilled, and let in addition the integrand g be linear in y. Then the cost functional $J(u,w)$ in* (7.67) *has a saddle point*

$(\bar{u}, \bar{w})$ on $U_{ad} \times W_{ad}$ subject to the parabolic system (7.61). *Moreover, if the corresponding trajectory to* (7.61) *satisfies the state constraints* (7.62), *then $(\bar{u}, \bar{w})$ is an optimal solution to the original minimax problem (P).*

Proof. Consider the functional $J(u, w)$ defined on the set $U_{ad} \times W_{ad} \subset L^p(0, T; U) \times L^2(0, T; W)$ with p sufficiently large. Observe that both sets U_{ad} and W_{ad} are convex and *sequentially* weakly compact in the reflexive spaces $L^p(0, T; U)$ and $L^2(0, T; W)$. Furthermore, it is easy to see that J is *convex-concave* on $U_{ad} \times W_{ad}$ by the assumptions made, where the *linearity* of g in y plays a crucial role.

Let us check the appropriate *semicontinuity* needed for applying the classical von Neumann saddle-point theorem in the infinite-dimensional spaces under consideration (see, e.g., Simons [1213]), where the sequential and topological weak convergences are *equivalent*. Show first that J is sequentially weakly lower semicontinuous with respect to u in the space $L^p(0, T; U)$ for any fixed $w \in L^2(0, T; W)$. To proceed, let a sequence $\{u_k\}$ weakly converge in $L^p(0, T; U)$ to some $\widetilde{u}$ as $k \to \infty$. By Mazur's theorem, find a sequence of convex combinations of u_k converging to $\widetilde{u}$ *strongly* in $L^p(0, T; U)$. Since $U = L^2(\Sigma)$, the latter sequence also converges to $\widetilde{u}$ strongly in $L^2(\Sigma)$. By standard arguments based on the convexity of h with respect to u and the other assumptions in (H6a), we conclude that

$$\int_\Sigma h\big(s, t, \widetilde{u}(s, t)\big)\, dsdt \le \liminf_{k\to\infty} \int_\Sigma h\big(s, t, u_k(s, t)\big)\, dsdt \ .$$

Consider further the trajectories (mild solutions) y_k and $\widetilde{y}$ to system (7.61) generated, respectively, by u_k and $\widetilde{u}$ for any fixed w. Then, by Theorem 7.35, $y_k \to \widetilde{y}$ *strongly* in $L^2(Q)$ as $k \to \infty$. To get the convergence

$$\int_Q g\big(x, t, \widetilde{y}(x, t)\big)\, dxdt = \lim_{k\to\infty} \int_Q g\big(x, t, y_k(x, t)\big)\, dxdt \ ,$$

we apply Polyak's result from [1096, Theorem 2], which ensures that the growth condition in (H4a) is *necessary and sufficient* for the *strong continuity* of the integral functional

$$I(y) := \int_Q g(x, t, y)\, dxdt$$

in $L^2(Q)$ provided that g satisfies the Carathéodory condition formulated in (H4a). Hence the cost functional $J(\cdot, w)$ in (7.67) is sequentially weakly *lower semicontinuous* on U_{ad} for any fixed w under the assumptions made.

To prove the sequential weak *upper semicontinuity* of $J(u, \cdot)$ on W_{ad} for any fixed u, we use the same (symmetric) arguments taking into account that the weak convergence of $w_k \to \widetilde{w}$ in $L^2(0, T; W)$ directly implies the *strong* convergence in $\mathcal{C}\big([0, T]; X\big)$ of the corresponding trajectories $y_k \to \widetilde{y}$; see the discussion right after Corollary 7.34. Thus the cost functional $J(u, w)$ in (7.67) is convex and weakly lower semicontinuous in u on the convex and weakly compact set $U_{ad} \subset L^p(0,T;U)$, and it is concave and weakly upper semicontinuous in w on the convex and weakly compact set $W_{ad} \subset L^2(0,T;W)$. Now the existence of a saddle point $(\bar{u},\bar{w})$ for J on $U_{ad} \times W_{ad}$ subject to system (7.61) follows from the classical minimax theorem in infinite dimensions. It is obvious furthermore that $(\bar{u},\bar{w})$ is an optimal solution to the original minimax problem (P) if the corresponding trajectory $\bar{y}$ satisfies the state constraints (7.62). This completes the proof of the theorem. △

Remark 7.37 (relaxation of linearity). Assumptions (H4a)–(H6a) on the integrands in (7.67) are required throughout this section and play a substantial role in the subsequent main results on the stability of approximations and their variational analysis. On the contrary, a restrictive *linearity* requirement on g in y is made just in Theorem 7.36 to ensure the existence of a saddle point; it is needed in fact only to conclude that the cost functional $J(u,w)$ is convex-concave. This assumption can be *removed* by considering saddle points in the framework of *mixed strategies*, which is similar to the *relaxation* procedures developed for optimal control problems in this and preceding chapters. Observe also that, due to the regularity results obtained in this subsection, the linearity of g in y is *not required* to ensure the *existence* of solutions in *optimal control* (not minimax) problems corresponding to either Dirichlet boundary controls, which provide the most difficult case, or distributed controls as well as controls in the Neumann boundary conditions, which are easier to handle for parabolic systems; see Mordukhovich and Zhang [979] for more details.

7.4.3 Suboptimality Conditions for Worst Perturbations

This subsection concerns the first subproblem (P_1) formulated in Subsect. 7.4.1. We treat (P_1) as an *optimal control problem* with *distributed controls* located on the right-hand side of the parabolic equation. Thus the *worst perturbations* $\bar{w}(\cdot)$ for the minimax problem (P) happen to be optimal solutions (in the sense of *maximizing* the cost functional J_1) to the distributive optimal control problem (P_1) under consideration in the presence of the *moving* state constraints therein. Note that these moving state constraints involve the *irregular* (measurable) function $\bar{y}_2(x,t)$, a mild solutions to the Dirichlet boundary control problem (7.70), that creates substantial complications. First we develop an *approximation method* for removing the latter constraints with justifying the appropriate *strong convergence* of these approximations. Then we provide a detailed variational analysis of the approximating problems to derive necessary *suboptimality* conditions for the worst perturbations. The limiting procedure allowing us to establish necessary *optimality* conditions for the worst perturbations will be developed in Subsect. 7.4.5.

To proceed, consider a set-valued *maximal monotone operator* $\alpha\colon I\!R \rightrightarrows I\!R$ given in the form

$$\alpha(r) := \begin{cases} [0,\infty) & \text{if } \ r = b\ , \\ (-\infty, 0] & \text{if } \ r = a\ , \\ 0 & \text{if } \ a < r < b\ , \\ \emptyset & \text{if either } \ r < a \text{ or } r > b\ . \end{cases}$$

We may construct a parametric family of *smooth* single-valued approximations $\alpha_\varepsilon\colon I\!R \to I\!R$ of the set-valued operator $\alpha(\cdot)$ using first the classical *Moreau-Yosida approximation* and then a $\mathcal{C}_0^\infty$*-mollifier* procedure in $I\!R$. The following realization is convenient for our purposes: construct $\alpha_\varepsilon\colon I\!R \to I\!R$ as $\varepsilon > 0$ by

$$\alpha_\varepsilon(r) := \begin{cases} \varepsilon^{-1}(r-b) - 1/2 & \text{if } \ r \ge b+\varepsilon\ , \\ (2\varepsilon^2)^{-1}(r-b)^2 & \text{if } \ b \le r < b+\varepsilon\ , \\ \varepsilon^{-1}(r-a) + 1/2 & \text{if } \ r \le a-\varepsilon\ , \\ -(2\varepsilon^2)^{-1}(r-a)^2 & \text{if } \ a-\varepsilon < r \le a\ , \\ 0 & \text{if } \ a < r < b. \end{cases} \tag{7.73}$$

It is easy to check computing the derivative of $\alpha_\varepsilon(\cdot)$ that

$$\varepsilon\alpha_\varepsilon'(r) = \begin{cases} 1 & \text{if } \ r \ge b+\varepsilon\ , \\ \varepsilon^{-1}(r-b) & \text{if } \ b \le r < b+\varepsilon\ , \\ 1 & \text{if } \ r \le a-\varepsilon\ , \\ -\varepsilon^{-1}(r-a) & \text{if } \ a-\varepsilon < r \le a, \\ 0 & \text{if } \ a < r < b \end{cases}$$

with $|\varepsilon\alpha_\varepsilon'(r)| \le 1$ for all $r \in I\!R$.

Let $(\bar{u}, \bar{w})$ be the given optimal solution to the minimax problem (P), and let $\bar{y}_1$ and $\bar{y}_2$ be the corresponding trajectories of systems (7.69) and (7.70), respectively. Consider the following ε*-parametric family* of control problems with *no state constraints* that approximate the first subproblem (P_1) in Subsect. 7.4.1 and depends on the given trajectory $\bar{y}_2$ of the Dirichlet system boundary control (7.70):

$(P_{1\varepsilon})$ maximize the penalized functional

$$J_{1\varepsilon}(w, y_1) := \int_Q \Big[g\big(x,t,y_1(x,t) + \bar{y}_2(x,t)\big) + f\big(x,t,w(x,t)\big)\Big]\,dxdt$$

$$-\big\|w-\bar{w}\big\|^2_{L^2(0,T;W)}-\varepsilon\big\|\alpha_\varepsilon(y_1+\bar{y}_2)\big\|^2_{L^2(0,T;X)}$$

subject to $w \in W_{ad}$ and system (7.69) .

Since $w \in W_{ad}$ and $\vartheta \in L^\infty(Q)$, the classical results ensure that the parabolic system (7.69) with the homogeneous Dirichlet boundary conditions admits a *unique strong solution* $y_1 \in W^{1,2}\big([0,T];X\big)$ satisfying the estimate

$$\Big\|\frac{\partial y_1}{\partial t}\Big\|_{L^2(0,T;X)}+\|Ay_1\|_{L^2(0,T;X)}$$

$$\leq C\big(\|y_0\|_{H^1_0(\Omega)\cap H^2(\Omega)}+\|Bw+\vartheta\|_{L^2(0,T;X)}\big)\ .$$

Taking $\{w_k\} \subset W_{ad}$ and the corresponding sequence $\{y_{1k}\}$ of strong solutions to system (7.69) and employing standard arguments in this setting (cf. Subsect. 7.4.2), we conclude that if $w_k \to w \in W_{ad}$ *weakly* in $L^2(0,T;W)$, then $y_{1k} \to y_1$ *strongly* in $\mathcal{C}\big([0,T];X\big)$ as $k \to \infty$ and that y_1 is also a strong solution of (7.69) corresponding to w.

We further proceed with the study of the approximating family $(P_{1\varepsilon})$. Our first goal is to justify the *existence* of optimal solutions to $(P_{1\varepsilon})$ for each $\varepsilon > 0$. This can be done by reducing the existence issue to the classical Weierstrass theorem ensuring the existence of global maximizers for upper semicontinuous cost functions over compact sets in appropriate topologies. The main complications in our case come from the perturbation term in the cost functional that depends on the *irregular* mild solution $\bar{y}_2$ to the Dirichlet system (7.70). Here is the result and its technically involved proof.

Theorem 7.38 (existence of optimal solutions to approximating problems for distributed perturbations). *Let the initial state y_0 in* (7.69) *satisfy assumption* (H2), *and let $\varepsilon > 0$. Then the approximating problem $(P_{1\varepsilon})$ admits at least one optimal solution with $(w_\varepsilon, y_{1\varepsilon}) \in W_{ad} \times W^{1,2}\big([0,T];X\big)$.*

Proof. Observe that the set of feasible solutions to problem (P_1) is nonempty, since the pair $(\bar{w},\bar{y}_1)$ is definitely a feasible solution to $(P_{1\varepsilon})$ for any $\varepsilon > 0$. We intent to show that the cost functional $J_{1\varepsilon}$ in $(P_{1\varepsilon})$ is *proper* and *uniformly upper bounded*, i.e.,

$$j_{1\varepsilon} := \sup_{w\in W_{ad}} J_{1\varepsilon}(w,y_1) < \infty\ , \tag{7.74}$$

where $y_1 \in W^{1,2}\big([0,T];X\big)$ is the corresponding strong solution to system (7.69). Indeed, assumptions (H4a) and (H5a) immediately imply the uniform upper boundedness of

$$\int_Q g\big(x,t,y_1(x,t)+\bar{y}_2(x,t)\big)\,dxdt+\int_Q f\big(x,t,w(x,t)\big)\,dxdt$$

over $w \in W_{ad}$. Furthermore, there obviously exists $\gamma > 0$ such that

$$\|w-\bar{w}\|_{L^2(0,T;W)} < \gamma \quad \text{whenever} \quad w \in W_{ad} \; .$$

It remains to analyze the last term of $J_{1\varepsilon}$ depending on $\bar{y}_2$. Due to estimates (7.71) and Definition 7.31 of mild solutions we have

$$\|\bar{y}_2(t)\|_X \le \frac{4M_2M_3 \max\big\{|\mu|, \nu\big\}\sqrt{\text{mes}(\Gamma)}}{1-4\delta} t^{\frac{1-4\delta}{4}} \quad \text{as} \quad \delta \in (0, 1/4) \; ,$$

where *mes* stands for the standard Lebesgue measure of a set. To estimate the term $\|\alpha_\varepsilon(y_1+\bar{y}_2)\|_{L^2(0,T;X)}$, consider the sets

$$\Omega_{1a}^t := \big\{x \in \Omega \;\big|\; a-\varepsilon < y_1(x,t)+\bar{y}_2(x,t) \le a\big\} \; ,$$

$$\Omega_{2a}^t := \big\{x \in \Omega \;\big|\; y_1(x,t)+\bar{y}_2(x,t) \le a-\varepsilon\big\} \; ,$$

$$\Omega_{1b}^t := \big\{x \in \Omega \;|\; b \le y_1(x,t)+\bar{y}_2(x,t) < b+\varepsilon\big\} \; ,$$

$$\Omega_{2b}^t := \big\{x \in \Omega \;\big|\; y_1(x,t)+\bar{y}_2(x,t) \ge b+\varepsilon\big\} \; ,$$

which are Lebesgue measurable subsets of Ω for a.e. $t \in [0,T]$. Taking into account the approximating structure $\alpha_\varepsilon(\cdot)$ in (7.73) and the trivial inequality $2(r^2+s^2) \ge (r+s)^2$ whenever $r, s \in I\!R$, we obtain the following estimates:

$$\begin{aligned}
&\big\|\alpha_\varepsilon(y_1+\bar{y}_2)\big\|_{L^2(0,T;X)} = \Big[\int_0^T \int_\Omega \alpha_\varepsilon^2\big(y_1(x,t)+\bar{y}_2(x,t)\big)\,dxdt\Big]^{1/2} \\
&= \Big[\int_0^T \Big(\int_{\Omega_{1a}^t} (2\varepsilon^2)^{-2}\big(y_1(x,t)+\bar{y}_2(x,t)-a\big)^4 dx \\
&+ \int_{\Omega_{2a}^t} \Big(\varepsilon^{-1}\big(y_1(x,t)+\bar{y}_2(x,t)-a\big)+\frac{1}{2}\Big)^2 dx \\
&+ \int_{\Omega_{1b}^t} (2\varepsilon^2)^{-2}\big(y_1(x,t)+\bar{y}_2(x,t)-b\big)^4 dx \\
&+ \int_{\Omega_{2b}^t} \Big(\varepsilon^{-1}(y_1(x,t)+\bar{y}_2(x,t)-b)-\frac{1}{2}\Big)^2 dx\Big)dt\Big]^{1/2} \\
&\le \Big[\int_0^T \Big(\frac{1}{4}\text{mes}(\Omega_{1a}^t)+\frac{1}{4}\text{mes}(\Omega_{1b}^t)\Big)dt\Big]^{1/2} \\
&+\Big[\int_0^T \int_{\Omega_{2a}^t} \Big(\varepsilon^{-1}\big(y_1(x,t)+\bar{y}_2(x,t)\big)-\varepsilon^{-1}a+\frac{1}{2}\Big)^2 dxdt\Big]^{1/2} \\
&+\Big[\int_0^T \int_{\Omega_{2b}^t} \Big(\varepsilon^{-1}\big(y_1(x,t)+\bar{y}_2(x,t)\big)-\varepsilon^{-1}b-\frac{1}{2}\Big)^2 dxdt\Big]^{1/2} \\
&\le \frac{1}{2}\sqrt{\text{mes}(Q)} + \Big[\int_0^T \int_{\Omega_{2a}^t} \Big(2\varepsilon^{-2}\big(y_1(x,t)+\bar{y}_2(x,t)\big)^2+2\Big(\varepsilon^{-1}|a|+\frac{1}{2}\Big)^2\Big)\,dxdt\Big]^{1/2} \\
&+\Big[\int_0^T \int_{\Omega_{2b}^t} \Big(2\varepsilon^{-2}\big(y_1(x,t)+\bar{y}_2(x,t)\big)^2+2\Big(\varepsilon^{-1}b+\frac{1}{2}\Big)^2\Big)\,dxdt\Big]^{1/2}
\end{aligned}$$

$$\le \frac{1}{2}\sqrt{\mathrm{mes}(Q)} + \Big[2\varepsilon^{-2}\int_0^T\int_{\Omega^t_{2a}}\big(y_1(x,t)+\bar{y}_2(x,t)\big)^2\,dxdt\Big]^{1/2}$$
$$+\Big[2\Big(\frac{1}{2}+\varepsilon^{-1}|a|\Big)^2\int_0^T\int_{\Omega^t_{2a}}dxdt\Big]^{1/2}$$
$$+\Big[2\varepsilon^{-2}\int_0^T\int_{\Omega^t_{2b}}\big(y_1(x,t)+\bar{y}_2(x,t)\big)^2\,dxdt\Big]^{1/2}+\Big[2\Big(\frac{1}{2}+\varepsilon^{-1}b\Big)^2\int_0^T\int_{\Omega^t_{2b}}dxdt\Big]^{1/2}$$
$$\le \frac{1}{2}\sqrt{\mathrm{mes}(Q)}+\sqrt{2}\varepsilon^{-1}\big\|y_1+\bar{y}_2\big\|_{L^2(0,T;X)}+\sqrt{2}\Big(\frac{1}{2}+\varepsilon^{-1}|a|\Big)\sqrt{\mathrm{mes}(Q)}$$
$$+\sqrt{2}\varepsilon^{-1}\|y_1+\bar{y}_2\|_{L^2(0,T;X)}+\sqrt{2}\Big(\frac{1}{2}+\varepsilon^{-1}b\Big)\sqrt{\mathrm{mes}(Q)}$$
$$=\Big(\frac{1}{2}+\sqrt{2}+\varepsilon^{-1}b\sqrt{2}+\varepsilon^{-1}|a|\sqrt{2}\Big)\sqrt{\mathrm{mes}(Q)}+2\sqrt{2}\varepsilon^{-1}\big\|y_1+\bar{y}_2\big\|_{L^2(0,T;X)}\,.$$

Combining this with the above estimate of $\|\bar{y}_2(t)\|_X$ and that of the strong solution $y_1 \in W^{1,2}\big([0,T];X\big)$ to (7.69), we justify the uniform boundedness of $\|\alpha_\varepsilon(y_1+\bar{y}_2)\|_{L^2(0,T;X)}$ as $\varepsilon > 0$ and thus arrive at (7.74).

Further, for every fixed $\varepsilon > 0$ that may be omitted for simplicity, we have a *maximizing sequence* of feasible solutions $\{w_k, y_{1k}\}$ to $(P_{1\varepsilon})$ such that

$$j_{1\varepsilon}-\frac{1}{k}\le J_{1\varepsilon}(w_k,y_{1k})\le j_{1\varepsilon}\ \text{ whenever }\ k\in I\!N\,. \tag{7.75}$$

Since W_{ad} is bounded, closed, and convex in $L^2(0,T;W)$, we extract a subsequence of $\{w_k\}$ (no relabeling) that converges weakly in $L^2(0,T;W)$ to some function $\widetilde{w} \in W_{ad}$. Taking the corresponding (strong) solution $\widetilde{y}_1$ to system (7.69) and recalling the discussion above, we have

$$y_{1k}\to\widetilde{y}_1\ \text{ strongly in }\ \mathcal{C}\big([0,T];X\big)\ \text{ as }\ k\to\infty\,.$$

It follows from assumptions (H4a) and (H5a) as well as from the concavity and continuity of the function $-\|\cdot\|^2_{L^2(0,T;W)}$ that

$$\limsup_{k\to\infty}\Big(\int_Q\Big[g\big(x,t,y_{1k}(x,t)+\bar{y}_2(x,t)\big)+f\big(x,t,w_k(x,t)\big)\Big]\,dxdt$$
$$-\big\|w_k-\bar{w}\big\|^2_{L^2(0,T;W)}\Big)\le\int_Q\Big[g\big(x,t,\widetilde{y}_1(x,t)+\bar{y}_2(x,t)\big)+f\big(x,t,\widetilde{w}(x,t)\big)\Big]\,dxdt$$
$$-\big\|\widetilde{w}-\bar{w}\big\|^2_{L^2(0,T;W)};$$

cf. the proof of Theorem 7.36. Hence

$$\lim_{k\to\infty}\alpha^2_\varepsilon(y_{1k}+\bar{y}_2)=\alpha^2_\varepsilon(\widetilde{y}_1+\bar{y}_2)\ \text{ a.e. in }\ Q\,,$$

which yields $j_{1\varepsilon}=J_{1\varepsilon}(\widetilde{w},\widetilde{y}_1)$ by passing to the limit in (7.75) as $k\to\infty$ and thus completes the proof of the theorem. △

The next technical lemma is important to justify the preservation of *state constraints* in the approximating procedures developed in this section.

Lemma 7.39 (preservation of state constraints). *Let $y_k(x,t)$, $k \in I\!N$, and $y(x,t)$ be nonnegative functions belonging to the space $L^1(Q)$. Given $c \ge 0$, consider the sets*

$$Q_k := \Big\{(x,t) \in Q \;\Big|\; y_k(x,t) > c + \frac{1}{k}\Big\}, \quad k \in I\!N .$$

Assume that $y_k(x,t) \to y(x,t)$ a.e. in Q and that

$$\int_{Q_k} y_k(x,t)\,dxdt \to 0 \;\; as \;\; k \to \infty .$$

Then we have the state constraint inequalities

$$0 \le y(x,t) \le c \;\; a.e. \; in \;\; Q.$$

Proof. Proving by contradiction, suppose that the conclusion of the lemma doesn't hold. Then for every $\rho > 0$ sufficiently small there is a measurable set $Q_\rho \subset Q$ such that $\operatorname{mes}(Q_\rho) > 0$ and

$$y(x,t) > c + \rho \;\; \text{whenever} \;\; (x,t) \in Q_\rho . \tag{7.76}$$

Taking into account the convergence $y_k(x,t) \to y(x,t)$ a.e. in Q and using the classical Egorov theorem from the theory of real functions, we conclude that for each $\varepsilon > 0$ and $\rho > 0$ there exist a measurable set $Q_\varepsilon \subset Q$ and a number $k_\varepsilon \in I\!N$, both independent of (x,t), such that

$$\rho - \frac{1}{k} > \frac{\rho}{2} > 0, \quad \operatorname{mes}(Q \setminus Q_\varepsilon) < \varepsilon, \quad \text{and}$$

$$|y_k(x,t) - y(x,t)| < \frac{\rho}{2} < \rho - \frac{1}{k} \;\; \text{whenever} \;\; k \ge k_\varepsilon \;\; \text{and} \;\; (x,t) \in Q_\varepsilon .$$

Choose $\varepsilon > 0$ so that $\operatorname{mes}(Q_\rho \cap Q_\varepsilon) \ne 0$. It follows from (7.76) that

$$y_k(x,t) > y(x,t) - \rho + \frac{1}{k} > c + \rho - \rho + \frac{1}{k} = c + \frac{1}{k} \;\; \text{whenever} \;\; k > k_\varepsilon$$

for any $(x,t) \in Q_\rho \cap Q_\varepsilon$, which gives $(Q_\rho \cap Q_\varepsilon) \subset Q_k$ for all $k > k_\varepsilon$. Then the convergence assumption of the lemma implies that

$$\int_{Q_\rho \cap Q_\varepsilon} y_k(x,t)\,dxdt \to 0 \;\; \text{as} \;\; k \to \infty .$$

The latter easily yields the condition

$$\int_{Q_\rho \cap Q_\varepsilon} y(x,t)\,dxdt = 0$$

due to the uniform convergence $y_k(x,t) \to y(x,t)$ in $Q_\rho \cap Q_\varepsilon$ as $k \to \infty$. Since $y(x,t) \ge 0$, we arrive at the conclusion

$$y(x,t) = 0 \ \text{ a.e. in } \ Q_\rho \cap Q_\varepsilon \ ,$$

which contradicts (7.76) and completes the proof of the lemma. △

The next theorem establishes a *strong convergence* of the approximation procedure developed in this subsection and thus shows that optimal solutions to the approximating problem $(P_{1\varepsilon})$, which do exist by Theorem 7.38, happen to be *suboptimal* solutions to the state-constrained problem (P_1) corresponding to the worst perturbations in the original minimax problem.

Theorem 7.40 (strong convergence of approximating problems for worst perturbations). *Let $(\bar{w}, \bar{y}_1)$ be the given optimal solution to problem (P_1), and let $\{(w_\varepsilon, y_{1\varepsilon})\}$ be a sequence of optimal solutions to problems $(P_{1\varepsilon})$. Then there is a subsequence of positive numbers ε such that*

$$w_\varepsilon \to \bar{w} \ \textit{ strongly in } \ L^2(0,T;W) \ ,$$

$$y_{1\varepsilon} \to \bar{y}_1 \ \textit{ strongly in } \ \mathcal{C}\big([0,T];X\big) \ ,$$

$$J_{1\varepsilon}(w_\varepsilon, y_{1\varepsilon}) \to J_1(\bar{w}, \bar{y}_1) \ \textit{ as } \ \varepsilon \downarrow 0 \ .$$

Proof. Using the same weak compactness arguments as in the proof of Theorem 7.38, we find a function $\widetilde{w} \in W_{ad}$ and a subsequence of $\{w_\varepsilon\}$ with

$$w_\varepsilon \to \widetilde{w} \ \text{ weakly in } \ L^2(0,T;W) \ \text{ as } \ \varepsilon \downarrow 0 \ .$$

As shown, there is the (unique) strong solution $\widetilde{y}_1 \in W^{1,2}\big([0,T];X\big)$ to system (7.69) generated by $\widetilde{w}$ such that

$$y_{1\varepsilon} \to \widetilde{y}_1 \ \text{ strongly in } \ \mathcal{C}\big([0,T];X\big) \ \text{ as } \ \varepsilon \downarrow 0 \ .$$

We need to prove that the pair $(\widetilde{w}, \widetilde{y}_1)$ is a feasible solution to problem (P_1). It actually remains to justify that $\widetilde{y}_1$ satisfies the state constraints (7.62), i.e.,

$$a \le \widetilde{y}_1(x,t) + \bar{y}_2(x,t) \le b \ \text{ a.e. in } \ Q \ .$$

To proceed, first note that the pair $(\bar{w}, \bar{y}_1)$, optimal to (P_1), is feasible to $(P_{1\varepsilon})$ with $\alpha_\varepsilon(\bar{y}_1 + \bar{y}_2) = 0$ a.e. in Q for all $\varepsilon > 0$. Due to the optimality of $(w_\varepsilon, y_{1\varepsilon})$ in the latter problem we have

$$J_1(\bar{w}, \bar{y}_1) = J_{1\varepsilon}(\bar{w}, \bar{y}_1) \le J_{1\varepsilon}(w_\varepsilon, y_{1\varepsilon}) \ \text{ for all } \ \varepsilon > 0 \ . \tag{7.77}$$

Using (7.77) and taking into account the structure of the cost functional in $(P_{1\varepsilon})$ as well as assumptions (H4a) and (H5a), conclude that the sequence $\{\varepsilon^{1/2}\|\alpha_\varepsilon(y_{1\varepsilon} + \bar{y}_2)\|_{L^2(0,T;X)}\}$ is bounded. The latter yields

$$\varepsilon\big\|\alpha_\varepsilon(y_{1\varepsilon}+\bar{y}_2)\big\|_{L^2(0,T;X)}\to 0\ \text{ as } \varepsilon\downarrow 0\ ,$$

which gives, by the above construction of $\alpha_\varepsilon(\cdot)$ and the partition of Ω, that

$$\int_0^T\int_{\Omega^t_{1a}}(2\varepsilon)^{-2}\Big(y_{1\varepsilon}(x,t)+\bar{y}_2(x,t)-a\Big)^4\,dxdt$$

$$+\int_0^T\int_{\Omega^t_{2a}}\Big(\big(y_{1\varepsilon}(x,t)+\bar{y}_2(x,t)-a\big)+\frac{\varepsilon}{2}\Big)^2\,dxdt$$

$$+\int_0^T\int_{\Omega^t_{1b}}(2\varepsilon)^{-2}\Big(y_{1\varepsilon}(x,t)+\bar{y}_2(x,t)-b\Big)^4\,dxdt$$

$$+\int_0^T\int_{\Omega^t_{2b}}\Big(\big(y_{1\varepsilon}(x,t)+\bar{y}_2(x,t)-b\big)-\frac{\varepsilon}{2}\Big)^2\,dxdt\to 0$$

as $\varepsilon\downarrow 0$. Note that for a.e. $t\in[0,T]$ we have

$$\Big(y_{1\varepsilon}(x,t)+\bar{y}_2(x,t)-a\Big)^4\le\varepsilon^4\ \text{ a.e. in }\ \Omega^t_{1a}\ ,$$

$$\Big(y_{1\varepsilon}(x,t)+\bar{y}_2(x,t)-b\Big)^4\le\varepsilon^4\ \text{ a.e. in }\ \Omega^t_{1b}\ .$$

Taking this into account together with Lemma 7.39, we get that the limiting pair $(\widetilde{w},\widetilde{y}_1)$ satisfies the state constraints (7.62), and hence it is feasible to (P_1). Thus $J_1(\widetilde{w},\widetilde{y}_1)\le J_1(\bar{w},\bar{y}_1)$.

Using this fact, let us now justify the desired strong convergence results of the theorem. First rewrite (7.77) in the form

$$J_1(\bar{w},\bar{y}_1)+\varepsilon\big\|\alpha_\varepsilon(y_{1\varepsilon}+\bar{y}_2)\big\|^2_{L^2(0,T;X)}$$

$$+\big\|w_\varepsilon-\bar{w}\big\|^2_{L^2(0,T;W)}\le J_1(w_\varepsilon,y_{1\varepsilon})$$

and take the upper limit in the both side of this inequality. Remember that under the assumptions made the functional $J_1(w,y)$ is upper semicontinuous in the weak topology of $L^2(0,T;W)$ and in the norm topology of $\mathcal{C}\big([0,T];X\big)$; cf. the proof of Theorem 7.38. Employing this observation together with the weak convergence of $w_\varepsilon\to\widetilde{w}$ and the strong convergence of $y_{1\varepsilon}\to\widetilde{y}_1$ established above, we derive that

$$J_1(\bar{w},\bar{y}_1)+\limsup_{\varepsilon\downarrow 0}\Big(\varepsilon\big\|\alpha_\varepsilon(y_{1\varepsilon}+\bar{y}_2)\big\|^2_{L^2(0,T;X)}+\big\|w_\varepsilon-\bar{w}\big\|^2_{L^2(0,T;W)}\Big)$$

$$\le\limsup_{\varepsilon\tau\downarrow 0}J_1(w_\varepsilon,y_{1\varepsilon})\le J_1(\widetilde{w},\widetilde{y}_1)\le J_1(\bar{w},\bar{y}_1)\ .$$

The latter yields

$$\lim_{\varepsilon\downarrow 0}\varepsilon\big\|\alpha_\varepsilon(y_{1\varepsilon}+\bar{y}_2)\big\|^2_{L^2(0,T;X)}=0\quad\text{and}\quad\lim_{\varepsilon\downarrow 0}\big\|w_\varepsilon-\bar{w}\big\|^2_{L^2(0,T;W)}=0\ ,$$

i.e., $w_\varepsilon \to \bar{w}$ *strongly* in $L^2(0,T;W)$ and therefore $y_{1\varepsilon}\to\bar{y}_1$ *strongly* in $\mathcal{C}\big([0,T];X\big)$ as $\varepsilon\downarrow 0$. Finally, the value convergence in the theorem follows from the continuity of $J_1(\cdot)$ in the strong topology of $L^2(Q)$ discussed in the proof of Theorem 7.36. Thus we complete the proof of this theorem. △

We conclude this subsection with deriving *necessary optimality conditions* in the approximation problems $(P_{1\varepsilon})$ for any $\varepsilon>0$. Due to Theorem 7.40 and the splitting procedure, the results obtained in this way can be treated as *suboptimality conditions* for the worst perturbations in the original minimax problem. Necessary optimality conditions for problem (P_1) will be established in the final Subsect. 7.4.5 by passing to the limit from those in $(P_{1\varepsilon})$ as $\varepsilon\downarrow 0$ with the help of Theorem 7.40.

Taking into account the convexity of the admissible perturbation set W_{ad} (which is now the control set in the problems $(P_{1\varepsilon})$ under consideration) and the absence of state constraints in $(P_{1\varepsilon})$, we conduct a variational analysis for each of these problems by using classical control variations and the regularity results of Subsect. 7.4.2. To simplify the issue, impose certain *smoothness* assumptions on the integrands with respect to both control/perturbation and state variables. Involving needle variations, as in Sects. 6.4 and 7.2, allows us to *relax* the smoothness and convexity assumptions made, but we are not going to pursue this goal here. Assume the following:

(H4b) $g(x,t,y)$ is continuously differentiable in y for a.e. $(x,t)\in Q$ and $(\partial g/\partial y)(x,t,y)$ is measurable in (x,t) for any $y\in I\!R$. Furthermore, there exist a nonnegative function $\eta_1\in L^2(Q)$ and a constant $\zeta_1\ge 0$ such that

$$\Big|\frac{\partial g}{\partial y}(x,t,y)\Big|\le\eta_1(x,t)+\zeta_1|y|\ \text{ a.e. }\ (x,t)\in Q,\quad\text{whenever}\ \ y\in I\!R\ .$$

(H5b) $f(x,t,w)$ is continuously differentiable in w for a.e. $(x,t)\in Q$ with $(\partial f/\partial w)(x,t,w)$ measurable in (x,t) for all $w\in[c,d]$. Furthermore, there is a nonnegative function $\kappa_1\in L^1(Q)$ such that

$$\Big|\frac{\partial f}{\partial w}(x,t,w)\Big|\le\kappa_1(x,t)\ \text{ a.e. }\ (x,t)\in Q\ \text{ whenever }\ w\in[c,d]\ .$$

Consider the *adjoint* parabolic system with the homogeneous Dirichlet boundary conditions:

$$\begin{cases} \dfrac{\partial p_1}{\partial t} - A^* p_1 = -\dfrac{\partial g}{\partial y}(x,t,y_{1\varepsilon}+\bar{y}_2) \\ +2\varepsilon\alpha'_\varepsilon(y_{1\varepsilon}+\bar{y}_2)\,\alpha_\varepsilon(y_{1\varepsilon}+\bar{y}_2) & \text{a.e. in } Q\ , \\ p_1(T,x)=0, & x\in \operatorname{cl}\Omega\ , \\ p_1(s,t)=0, & (s,t)\in\Sigma\ , \end{cases} \tag{7.78}$$

where $\operatorname{cl}\Omega = \Omega\cup\Gamma$. It follows from (H4b) that

$$\frac{\partial g}{\partial y}\big(x,t,y_1(x,t)+\bar{y}_2(x,t)\big)\in L^2(Q) \quad \text{whenever} \quad y_1\in\mathcal{C}\big([0,T];X\big)\ .$$

As well known from the classical parabolic theory, system (7.78) admit a unique *strong* solution $p_{1\varepsilon}\in W^{1,2}\big([0,T];X\big)$ satisfying

$$p_{1\varepsilon}\in\mathcal{C}\big([0,T];X\big)\cap L^2\big(0,T;H_0^1(\Omega)\cap H^2(\Omega)\big)\ .$$

The next theorem gives necessary optimality conditions for the approximating problems $(P_{1\varepsilon})$ in the *integral form* of the (linearized) *maximum principle*. It easily implies the corresponding *pointwise* result in the *bang-bang* form due to the constraint structure of W_{ad}; see the corollary below. The approximating parameter $\varepsilon>0$ is fixed in what follows.

Theorem 7.41 (suboptimality conditions for worst perturbation in integral form). *Let $(w_\varepsilon, y_{1\varepsilon})$ be an optimal solution to problem $(P_{1\varepsilon})$, and let $p_{1\varepsilon}$ be the corresponding strong solution to the adjoint system* (7.78). *Then for any $w\in L^2(0,T;W)$ such that $w_\varepsilon+\theta w\in W_{ad}$ whenever $\theta\in[0,\theta_0]$ with some $\theta_0>0$ we have*

$$\int_Q\Big(B^*p_{1\varepsilon}+\frac{\partial f}{\partial w}(x,t,w_\varepsilon)-2(w_\varepsilon-\bar{w})\Big)w\,dxdt\le 0\ .$$

Proof. Let $y_{1\varepsilon w}$ be the strong solution of (7.69) corresponding to $w_\varepsilon+\theta w$. It is easy to check that $y_{1\varepsilon w}\to y_{1\varepsilon}$ in the norm topology of $\mathcal{C}\big([0,T];X\big)$ as $\theta\downarrow 0$ with the representation

$$\frac{y_{1\varepsilon w}(x,t)-y_{1\varepsilon}(x,t)}{\theta}=z_{1\varepsilon}(x,t)\ \text{a.e.}\ \ (x,t)\in Q\ \ \text{as}\ \ \theta>0\ , \tag{7.79}$$

where $z_{1\varepsilon}$ is a strong solution to the system

$$\begin{cases} \dfrac{\partial z_1}{\partial t}+Az_1=Bw \ \ \text{a.e. in } Q\ , \\ z_1(x,0)=0, & x\in\Omega\ , \\ z_1(s,t)=0, & (s,t)\in\Sigma\ . \end{cases}$$

Defining the limits

$$\delta_1 := \limsup_{\theta\downarrow 0} \int_Q \frac{g\big(x,t,y_{1\varepsilon w}(x,t)+\bar{y}_2(x,t)\big) - g\big(x,t,y_{1\varepsilon}(x,t)+\bar{y}_2(x,t)\big)}{\theta}\, dxdt\ ,$$

$$\delta_2 := \limsup_{\theta\downarrow 0} \int_Q \frac{\varepsilon\alpha_\varepsilon^2\big(y_{1\varepsilon w}(x,t)+\bar{y}_2(x,t)\big) - \varepsilon\alpha_\varepsilon^2\big(y_{1\varepsilon}(x,t)+\bar{y}_2(x,t)\big)}{\theta}\, dxdt$$

and applying the mean value theorem to the integrands therein, we get

$$\delta_1 = \limsup_{\theta\downarrow 0} \int_Q \frac{\partial g}{\partial y}\Big(x,t,y_{1\varepsilon}+\bar{y}_2+\theta_1(y_{1\varepsilon w}-y_{1\varepsilon})\Big)\frac{y_{1\varepsilon w}(x,t)-y_{1\varepsilon}(x,t)}{\theta}\, dxdt\ ,$$

$$\delta_2 = \varepsilon \limsup_{\theta\downarrow 0} \int_Q \big(\alpha_\varepsilon\big(y_{1\varepsilon w}+\bar{y}_2\big)+\alpha_\varepsilon(y_{1\varepsilon}+\bar{y}_2)\big)\,\alpha_\varepsilon'\big(y_{1\varepsilon}+\bar{y}_2+\theta_2(y_{1\varepsilon w}-y_{1\varepsilon})\big)$$

$$\times \frac{y_{1\varepsilon w}(x,t)-y_{1\varepsilon}(x,t)}{\theta}\, dxdt\ ,$$

where $\theta_1 = \theta_1(x,t),\ \theta_2 = \theta_2(x,t) \in [0,1]$ a.e. in Q. Then using (7.79), (H4b), and the Lebesgue dominated convergence theorem, one has

$$\Big|\int_Q \Big(\frac{\partial g}{\partial y}\big(x,t,y_{1\varepsilon}+\bar{y}_2+\theta_1(y_{1\varepsilon w}-y_{1\varepsilon})\big)\frac{y_{1\varepsilon w}-y_{1\varepsilon}}{\theta}$$

$$-\frac{\partial g}{\partial y}\big(x,t,y_{1\varepsilon}+\bar{y}_2\big)z_{1\varepsilon}\Big)\, dxdt\Big| \le \int_Q \Big|\frac{\partial g}{\partial y}\big(x,t,y_{1\varepsilon}+\bar{y}_2+\theta_1(y_{1\varepsilon w}-y_{1\varepsilon})\big)$$

$$-\frac{\partial g}{\partial y}\big(x,t,y_{1\varepsilon}+\bar{y}_2\big)\Big|\cdot|z_{1\varepsilon}|\, dxdt \to 0 \ \text{ as } \ \theta\downarrow 0\ ,$$

which implies the expression

$$\delta_1 = \int_Q \frac{\partial g}{\partial y}\Big(x,t,y_{1\varepsilon}(x,t)+\bar{y}_2(x,t)\Big)z_{1\varepsilon}(x,t)\, dxdt\ .$$

Noting further that $\alpha_\varepsilon'(\cdot)$ is continuous with $|\varepsilon\alpha_\varepsilon'(\cdot)| \le 1$ and that

$$\alpha_\varepsilon(y_{1\varepsilon w}+\bar{y}_2)+\alpha_\varepsilon(y_{1\varepsilon}+\bar{y}_2) \in L^2(0,T;X)\ ,$$

we deduce from (7.79) and the calculation above that

$$\delta_2 = 2\varepsilon \int_Q \alpha_\varepsilon'\big(y_{1\varepsilon}(x,t)+\bar{y}_2(x,t)\big)\,\alpha_\varepsilon\big(y_{1\varepsilon}(x,t)+\bar{y}_2(x,t)\big)\, z_{1\varepsilon}(x,t)\, dxdt\ .$$

Since $w_\varepsilon + \theta w \to w_\varepsilon$ strongly in $L^2(Q)$ as $\theta \downarrow 0$ for all w satisfying the conditions of the theorem, we deduce from the assumptions in (H5b) and the mean value theorem that

$$\left|\frac{f(x,t,w_\varepsilon+\theta w)-f(x,t,w_\varepsilon)}{\theta}\right| = \left|\frac{\partial f}{\partial w}\big(x,t,w_\varepsilon+\theta_3\theta w\big)\,w\right|$$

$$\le \kappa_1(x,t)\,|w(x,t)|\,,$$

$$\frac{\partial f}{\partial w}\big(x,t,w_\varepsilon+\theta_3\theta w\big)\,w \to \frac{\partial f}{\partial w}\big(x,t,w_\varepsilon\big)\,w \quad \text{a.e. in } Q \text{ as } \theta\downarrow 0\,,$$

where $\theta_3 = \theta_3(x,t) \in [0,1]$ a.e. in Q. Thus the Lebesgue dominated convergence theorem yields

$$\int_Q \frac{\partial f}{\partial w}\big(x,t,w_\varepsilon+\theta_3\theta w\big)w\,dxdt \to \int_Q \frac{\partial f}{\partial w}\big(x,t,w_\varepsilon\big)w\,dxdt \text{ as } \theta\downarrow 0\,.$$

Now employing the *optimality* of $(w_\varepsilon, y_{1\varepsilon})$ in problem $(P_{1\varepsilon})$, we get

$$0 \ge \limsup_{\theta\downarrow 0} \frac{J_{1\varepsilon}(w_\varepsilon+\theta w, y_{1\varepsilon w}) - J_{1\varepsilon}(w_\varepsilon, y_{1\varepsilon})}{\theta}$$

$$\ge \limsup_{\theta\downarrow 0}\int_Q \Big[\frac{g(x,t,y_{1\varepsilon w}+\bar y_2)-g(x,t,y_{1\varepsilon}+\bar y_2)}{\theta}$$

$$+\frac{f(x,t,w_\varepsilon+\theta w)-f(x,t,w_\varepsilon)}{\theta}\Big]\,dxdt$$

$$-\limsup_{\theta\downarrow 0}\int_Q \frac{(w_\varepsilon+\theta w-\bar w)^2-(w_\varepsilon-\bar w)^2}{\theta}\,dxdt$$

$$-\varepsilon\,\limsup_{\theta\downarrow 0}\int_Q \frac{\alpha_\varepsilon^2(y_{1\varepsilon w}+\bar y_2)-\alpha_\varepsilon^2(y_{1\varepsilon}+\bar y_2)}{\theta}\,dxdt\,.$$

Combining all the above, we arrive at the inequality

$$0 \ge \int_Q \Big(\frac{\partial g}{\partial y}\big(x,t,y_{1\varepsilon}+\bar y_2\big) - 2\varepsilon\alpha_\varepsilon'(y_{1\varepsilon}+\bar y_2)\,\alpha_\varepsilon(y_{1\varepsilon}+\bar y_2)\Big)\,z_{1\varepsilon}\,dxdt$$

$$+\int_Q \Big(\frac{\partial f}{\partial w}\big(x,t,w_\varepsilon\big) - 2(w_\varepsilon-\bar w)\Big)\,w\,dxdt\,.$$

Substituting finally the solution $p_{1\varepsilon}$ to (7.70) into the latter inequality and integrating it by parts, we complete the proof of the theorem. △

Corollary 7.42 (suboptimality conditions for worst perturbations in pointwise form). *For each $\varepsilon > 0$ the maximal perturbation w_ε in problem $(P_{1\varepsilon})$ satisfies the following bang-bang relations:*

$$w_\varepsilon(x,t) = c \quad a.e. \quad \begin{cases} (x,t) \in Q \;\Big|\; \big(B^* p_{1\varepsilon}\big)(x,t) + \dfrac{\partial f}{\partial w}\big(x,t,w_\varepsilon\big) \\ \qquad\qquad -2\big(w_\varepsilon(x,t) - \bar{w}(x,t)\big) < 0\,, \end{cases}$$

$$w_\varepsilon(x,t) = d \quad a.e. \quad \begin{cases} (x,t) \in Q \;\Big|\; \big(B^* p_{1\varepsilon}\big)(x,t) + \dfrac{\partial f}{\partial w}\big(x,t,w_\varepsilon\big) \\ \qquad\qquad -2\big(w_\varepsilon(x,t) - \bar{w}(x,t)\big) > 0\,, \end{cases}$$

where $p_{1\varepsilon}$ *is the corresponding solution to the adjoint system* (7.79).

Proof. Taking $\widetilde{w} := w - w_\varepsilon$ for any $w \in W_{ad}$, we have $w_\varepsilon + \theta\widetilde{w} = (1-\theta)w_\varepsilon + \theta w \in W_{ad}$ whenever $\theta \in [0,1]$. Replacing w with $\widetilde{w}$ in the optimality conditions of Theorem 7.41 gives us

$$\int_Q \Big(B^* p_{1\varepsilon} + \frac{\partial f}{\partial w}\big(x,t,w_\varepsilon\big) - 2(w_\varepsilon - \bar{w})\Big)\,(w - w_\varepsilon)\,dxdt \le 0$$

for all $w \in W_{ad}$, which implies the bang-bang relations. △

7.4.4 Suboptimal Controls under Worst Perturbations

In this subsection we study the boundary optimal control problem (P_2) formulated in Subsect. 7.4.1. According to the splitting procedure, optimal solutions to (P_2) allow us to find optimal boundary controls to the original minimax problem (P) under the worst perturbations.

The problem (P_2) under consideration is a *boundary optimal control* problem for *parabolic systems* with *pointwise/hard* control constraints acting in the *Dirichlet boundary conditions* and with the *moving state constraints* generated by the splitting procedure. To remove/approximate the latter constraints, we develop a *penalization technique* that provides useful *suboptimality* information for the original minimax problem.

Let $\alpha(\cdot)$ be the maximal monotone operator defined in the preceding subsection, and let $\alpha_\varepsilon(\cdot)$ be a *smooth approximation* of $\alpha(\cdot)$ in the form (7.73). For each $\varepsilon > 0$ consider a parametric family of *approximating problems* for (P_2) formulated as follows:

$$(P_{2\varepsilon}) \quad \text{minimize } J_{2\varepsilon}(u,y_2) := \int_Q g\big(x,t,\bar{y}_1(x,t) + y_2(x,t)\big)\,dxdt$$

$$+ \int_\Sigma h\big(s,t,u(s,t)\big)\,dsdt + \big\|u - \bar{u}\big\|^p_{L^p(0,T;U)} + \varepsilon\big\|\alpha_\varepsilon(\bar{y}_1 + y_2)\big\|^2_{L^2(0,T;X)}$$

$$\text{subject to } u \in U_{ad} \ \text{ and system } \ (7.70)\,.$$

Remember that solutions to system (7.70) are understood in the *mild* sense, i.e., as $y_2 \in \mathcal{C}\big([0,T];X\big)$ satisfying

$$y_2(t) = \mathcal{L}u := A\int_0^t S(t-\tau)Du(\tau)d\tau \quad \text{for all} \quad t \in [0,T] \ .$$

The next result justifies the *existence of optimal solutions* to the approximating minimization problem $(P_{2\varepsilon})$ for every $\varepsilon > 0$.

Theorem 7.43 (existence of optimal solutions to approximating Dirichlet problems). *For each* $\varepsilon > 0$ *the approximating problem* $(P_{2\varepsilon})$ *admits at least one optimal solution pair* $(u_\varepsilon, y_{2\varepsilon}) \in U_{ad} \times \mathcal{C}\big([0,T];X\big)$.

Proof. First note that each problem $(P_{2\varepsilon})$ has a feasible pair $(\bar{u}, \bar{y}_2)$ generated by the given optimal solution $(\bar{u}, \bar{w})$ to the original minimax problem (P). Let (u, y_2) be an arbitrary feasible pair to $(P_{2\varepsilon})$. It follows from assumptions (H4a) and (H6a) that the integral sum

$$\int_Q g\big(x,t,\bar{y}_1(x,t) + y_2(x,t)\big)\,dxdt + \int_\Sigma h\big(s,t,u(s,t)\big)\,dsdt$$

is uniformly bounded from below over $u \in U_{ad}$ and the corresponding trajectories y_2 of the Dirichlet system (7.70). To estimate the given trajectory $\bar{y}_1$ of the distributed system (7.69), use the exponential semigroup estimate (7.65) in the mild solution representation of Definition 7.31, which gives

$$\big\|\bar{y}_1(t)\big\|_X \le M_1\Big(e^{-\omega t}\big\|y_0\big\|_X + \frac{\big(\|B\|\max\{|c|,d\} + \|\vartheta\|_\infty\big)\sqrt{\text{mes}(\Omega)}}{\omega}\big(1-e^{-\omega t}\big)\Big) \ .$$

Then employing arguments similar to the proof of Theorem 7.38, we deduce the uniform boundedness of the penalization term $\|\alpha_\varepsilon(\bar{y}_1 + y_2)\|_{L^2(0,T;X)}$. This yields the *uniform lower boundedness* of the cost functional $J_{2\varepsilon}(u, y_2)$ and thus the *finiteness* of the infimum $\inf J_{2\varepsilon}(u, y_2)$ in problem $(P_{2\varepsilon})$ for each $\varepsilon > 0$.

Further, taking into account the existence and *uniqueness* of mild solutions y_2 to system (7.70) corresponding to any given admissible control $u \in U_{ad}$, we consider the cost functional $J_{2\varepsilon}$ in $(P_{2\varepsilon})$ as a function of $u \in U_{ad} \subset L^p(0,T;U)$, where the latter space is equipped with the *weak* topology. Now employing the *regularity/convergence* results for mild solutions given in Corollary 7.34 and Theorem 7.35, as well as the *convexity* of the integrand h in u, we conclude that the cost functional in $(P_{2\varepsilon})$ is *weakly semicontinuous* in $L^p(0,T;U)$ on the *weakly compact* set U_{ad}; cf. the proof of Theorem 7.38. Thus the existence of optimal solutions to $(P_{2\varepsilon})$ follows from the classical Weierstrass existence theorem in the topological setting under consideration. △

Next we establish the following *strong convergence* of optimal solutions for the approximating problems $(P_{2\varepsilon})$ to the given optimal solution $(\bar{u}, \bar{y}_2)$ for the state-constrained problem (P_2).

Theorem 7.44 (strong convergence of approximating Dirichlet boundary control problems). *Let $(\bar{u}, \bar{y}_2)$ be the given optimal solution to the state-constrained problem (P_2), and let $\{(u_\varepsilon, y_{2\varepsilon})\}$ be a sequence of optimal solutions to the approximating problems $(P_{2\varepsilon})$. Then there is a subsequence of positive numbers ε such that*

$$u_\varepsilon \to \bar{u} \;\; \textit{strongly in} \;\; L^p(0,T;U) \; ,$$

$$y_{2\varepsilon} \to \bar{y}_2 \;\; \textit{strongly in} \;\; \mathcal{C}\big([0,T];X\big) \; ,$$

$$J_{2\varepsilon}(u_\varepsilon, y_{2\varepsilon}) \to J_2(\bar{u}, \bar{y}_2) \;\; \textit{as} \;\; \varepsilon \downarrow 0 \; .$$

Proof. From the optimality of $(u_\varepsilon, y_{2\varepsilon})$ in $(P_{2\varepsilon})$ and the feasibility of $(\bar{u}, \bar{y}_2)$ in this problem we have

$$J_{2\varepsilon}(u_\varepsilon, y_{2\varepsilon}) \le J_{2\varepsilon}(\bar{u}, \bar{y}_2) = J_2(\bar{u}, \bar{y}_2) \;\; \text{for all} \;\; \varepsilon > 0 \; . \tag{7.80}$$

This implies, in particular, that there is $M > 0$ independent of ε such that

$$\varepsilon\big\|\alpha_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\big\|^2_{L^2(0,T;X)} \le M \;\; \text{and} \;\; \varepsilon\big\|\alpha_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\big\|_{L^2(0,T;X)} \to 0 \;\; \text{as} \;\; \varepsilon \downarrow 0 \; .$$

Due to the weak compactness of U_{ad} in $L^p(0,T;U)$, find a function $\widetilde{u} \in U_{ad}$ and a subsequence of $\{u_\varepsilon\}$ along which

$$u_\varepsilon \to \widetilde{u} \;\; \text{weakly in} \;\; L^p(0,T;U) \;\; \text{as} \;\; \varepsilon \downarrow 0 \; .$$

Denote by $\widetilde{y}_2$ the (unique) mild solution of (7.70) corresponding to $\widetilde{u}$. Using Theorem 7.35, select a subsequence of $\varepsilon \downarrow 0$ such that

$$y_{2\varepsilon}(x,t) \to \widetilde{y}_2(x,t) \;\; \text{a.e. in} \;\; Q$$

provided that p is sufficiently large. Then following the proof of Theorem 7.40 with the use of Lemma 7.39, we justify the validity of the state constraints

$$a \le \bar{y}_1(x,t) + \widetilde{y}_2(x,t) \le b \;\; \text{a.e. in} \;\; Q \; .$$

This ensures that $(\widetilde{u}, \widetilde{y}_2)$ is a *feasible solution* to the state-constrained problem (P_2), and thus

$$J_2(\widetilde{u}, \widetilde{y}_2) \ge J_2(\bar{u}, \bar{y}_2) \; .$$

Now pass to the limit in (7.80) as $\varepsilon \downarrow 0$ with taking into account the *lower semicontinuity* of the cost functional J_2 on the control set U_{ad} in the weak topology of $L^p(0,T;U)$; cf. the proof of Theorem 7.36. This yields

$$\lim_{\varepsilon\downarrow 0}\big\|u_\varepsilon-\bar u\big\|^p_{L^p(0,T;U)}=0 \ \text{ and } \ \lim_{\varepsilon\downarrow 0}\varepsilon\big\|\alpha_\varepsilon(\bar y_1+y_{2\varepsilon})\big\|^2_{L^2(0,T;X)}=0\,. \quad (7.81)$$

The first equality in (7.81) means that $u_\varepsilon\to\bar u$ *strongly* in $L^p(0,T;U)$ as $\varepsilon\downarrow 0$. We know from Theorem 7.33 that, for p sufficiently large, the latter implies the *strong* convergence $y_{2\varepsilon}\to\bar y_2$ in $\mathcal{C}\big([0,T];X\big)$. Finally, the value convergence in this theorem follows from the second equality in (7.81) due to the penalization structure of $J_{2\varepsilon}$. This completes the proof. △

Next we derive necessary optimality conditions for the approximating problems under the following assumptions parallel to those in Subsect. 7.4.3; cf. also the discussions therein.

(H6b) $h(s,t,u)$ is continuously differentiable in u with the derivative measurable in (s,t). Furthermore, there is a nonnegative function $\gamma_1\in L^q(\Sigma)$, with $1/p+1/q=1$, providing the estimate

$$\Big|\frac{\partial h}{\partial u}(s,t,u)\Big|\le\gamma_1(s,t) \ \text{ a.e. } \ (s,t)\in\Sigma \ \text{ whenever } \ \in u\in[\mu,\nu]\,.$$

Let $(u_\varepsilon,y_{2\varepsilon})$ be an optimal solution to the approximating $(P_{2\varepsilon})$ with an arbitrary fixed $\varepsilon>0$. Consider *feasible variations* of the control u_ε in the form $u_\varepsilon+\theta u\in U_{ad}$ with $u\in L^p(0,T;U)$, where $\theta\in[0,\theta_0]$ for some $\theta_0>0$. Denote by $y_{2\varepsilon u}$ the mild solution of (7.70) corresponding to $u_\varepsilon+\theta u$ and consider a function $\varphi\colon[0,\theta_0]\to I\!R$ defined by

$$\varphi(\theta):=J_{2\varepsilon}(u_\varepsilon+\theta u,y_{2\varepsilon u})\,.$$

This function obviously attains its minimum at $\theta=0$. Moreover, it follows from the definition of mild solutions that

$$y_{2\varepsilon u}\to y_{2\varepsilon} \ \text{ strongly in } \ \mathcal{C}\big([0,T];H^{1/2-\varepsilon}(\Omega)\big) \ \text{ as } \ \theta\downarrow 0, \quad \text{and}$$

$$\frac{y_{2\varepsilon u}(x,t)-y_{2\varepsilon}(x,t)}{\theta}=\mathcal{L}u \ \text{ a.e. } \ (x,t)\in Q \ \text{ whenever } \ \theta>0$$

provided that p is sufficiently large.

Now we are ready to derive necessary optimality conditions for the approximating Dirichlet boundary control problems $(P_{2\varepsilon})$. First we establish the *integral form* of the result and then deduce its consequence in the (pointwise) form of the *bang-bang principle.* As we know from the strong convergence of Theorem 7.44, these results provide *suboptimality* conditions for the state-constrained problem (P_2).

Theorem 7.45 (suboptimality conditions for Dirichlet boundary controls under worst perturbations). *Let $(u_\varepsilon,y_{2\varepsilon})$ be an optimal solution to the approximating problem $(P_{2\varepsilon})$ with any fixed $\varepsilon>0$, and let*

$$\mathcal{L}^*\colon\mathcal{C}([0,T];X)^*\to L^q(0,T;U)$$

be the adjoint operator to the mild solution mapping $\mathcal{L}$ *defined in* (7.72)*. Then*

$$0 \le \int_\Sigma \Big[\mathcal{L}^*\Big(\frac{\partial g}{\partial y}(x,t,\bar y_1 + y_{2\varepsilon}) + 2\varepsilon\, \alpha'_\varepsilon(\bar y_1 + y_{2\varepsilon})\, \alpha_\varepsilon(\bar y_1 + y_{2\varepsilon})\Big)$$

$$+\frac{\partial h}{\partial u}(s,t,u_\varepsilon)\Big] u\, dsdt + 2p\int_0^T \|u_\varepsilon - \bar u\|_U^{p-2}\Big(\int_\Gamma \big(u_\varepsilon - \bar u\big)\, u\, ds\Big) dt$$

for all $u \in L^p(0,T;U)$ *satisfying*

$$u_\varepsilon + \theta u \in U_{ad} \quad \text{whenever} \quad \theta \in [0,\theta_0] \quad \text{with some} \quad \theta_0 > 0\ .$$

Proof. Taking into account that the above function φ obviously attains its minimum at $\theta = 0$ and then using the classical mean value theorem, we get the following relationships:

$$0 \le \liminf_{\theta\downarrow 0} \frac{\varphi(\theta) - \varphi(0)}{\theta} + \liminf_{\theta\downarrow 0} \frac{1}{\theta}\Big[\int_Q \big(g(x,t,\bar y_1 + y_{2\varepsilon u}) - g(x,t,\bar y_1 + y_{2\varepsilon})\big)\, dxdt$$

$$+\int_\Sigma \big(h(s,t,u_\varepsilon + \theta u) - h(s,t,u_\varepsilon)\big)\, dsdt + \big(\|u_\varepsilon + \theta u - \bar u\|^p_{L^p(0,T;U)}$$

$$-\|u_\varepsilon - \bar u\|^p_{L^p(0,T;U)}\big) + \varepsilon\big(\|\alpha_\varepsilon(\bar y_1 + y_{2\varepsilon u})\|^2_{L^2(0,T;X)} - \|\alpha_\varepsilon(\bar y_1 + y_{2\varepsilon})\|^2_{L^2(0,T;X)}\big)\Big]$$

$$= \liminf_{\theta\downarrow 0} \frac{1}{\theta}\Big[\int_Q \frac{\partial g}{\partial y}\big(x,t,\bar y_1 + y_{2\varepsilon} + \theta_1(y_{2\varepsilon u} - y_{2\varepsilon})\big)\,\big(y_{2\varepsilon u} - y_{2\varepsilon}\big)\, dxdt$$

$$+\int_\Sigma \frac{\partial h}{\partial u}\big(s,t,u_\varepsilon + \theta_2\theta\, u\big)\theta u\, dsdt$$

$$+\int_0^T \big(\|u_\varepsilon + \theta u - \bar u\|_U^{p-2} + \ldots + \|u_\varepsilon - \bar u\|_U^{p-2}\big)\Big(\int_\Gamma \theta u\big(2u_\varepsilon - 2\bar u + \theta u\big)\, ds\Big)\, dt$$

$$+\varepsilon\int_Q \big(\alpha_\varepsilon(\bar y_1 + y_{2\varepsilon u}) + \alpha_\varepsilon(\bar y_1 + y_{2\varepsilon})\big)\, \alpha'_\varepsilon\big(\bar y_1 + y_{2\varepsilon}$$

$$+\theta_3(y_{2\varepsilon u} - y_{2\varepsilon})\big)\,(y_{2\varepsilon u} - y_{2\varepsilon})\, dxdt\Big]\ ,$$

where $\theta_i = \theta_i(x,t) \in [0,1]$ a.e. in Q for $i = 1,2,3$. Observe that $\theta_i(y_{2\varepsilon u} - y_{2\varepsilon}) \to 0$ strongly in $L^2(Q)$ as $\theta \downarrow 0$ for $i = 1,2,3$ and that

$$\alpha_\varepsilon(\bar y_1 + y_{2\varepsilon u}) + \alpha_\varepsilon(\bar y_1 + y_{2\varepsilon}) \in L^2(0,T;X)\ .$$

Then similarly to Subsect. 7.4.3, by using assumptions (H4b), (H6b) and the Lebesgue dominated convergence theorem, we arrive at the inequality

$$0 \le \int_Q \Big(\frac{\partial g}{\partial y}(x,t,\bar{y}_1 + y_{2\varepsilon}) + 2\varepsilon \alpha'_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\,\alpha_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\Big) \mathcal{L}u\,dxdt$$

$$+ \int_\Sigma \frac{\partial h}{\partial u}(s,t,u_\varepsilon) u\,dsdt + 2p \int_0^T \|u_\varepsilon - \bar{u}\|_U^{p-2} \Big(\int_\Gamma (u_\varepsilon - \bar{u}) u\,ds \Big) dt\ ,$$

which implies the necessary optimality condition stated in the theorem. △

The necessary optimality condition of Theorem 7.45 easily yields the *bang-bang* form of optimal controls to the perturbed problems $(P_{2\varepsilon})$. Note that the bang-bang relations presented below don't carry any information on optimal controls at the *singular set* of points (s,t) from the boundary Σ that vanish the corresponding integrand in the integral inequality of Theorem 7.45.

Corollary 7.46 (bang-bang suboptimality conditions for Dirichlet boundary controls). *For each $\varepsilon > 0$ the optimal control u_ε to $(P_{2\varepsilon})$ satisfies the following bang-bang relations:*

$$u_\varepsilon(s,t) = \mu \quad a.e. \begin{cases} (s,t) \in \Sigma \;\Big|\; \mathcal{L}^* \Big(\dfrac{\partial g}{\partial y}(x,t,\bar{y}_1 + y_{2\varepsilon}) \\ +2\varepsilon \alpha'_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\,\alpha_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\Big) + \dfrac{\partial h}{\partial u}(s,t,u_\varepsilon) \\ +2p\|u_\varepsilon - \bar{u}\|_U^{p-2}(u_\varepsilon - \bar{u}) < 0\ , \end{cases}$$

$$u_\varepsilon(s,t) = \nu \quad a.e. \begin{cases} (s,t) \in \Sigma \;\Big|\; \mathcal{L}^* \Big(\dfrac{\partial g}{\partial y}(x,t,\bar{y}_1 + y_{2\varepsilon}) \\ +2\varepsilon \alpha'_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\,\alpha_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\Big) + \dfrac{\partial h}{\partial u}(s,t,u_\varepsilon) \\ +2p\|u_\varepsilon - \bar{u}\|_U^{p-2}(u_\varepsilon - \bar{u}) > 0\ . \end{cases}$$

Proof. Put $u := \upsilon - \bar{u}$ in the result of Theorem 7.45 for any $\upsilon \in U_{ad}$. Taking $\theta_0 = 1$, we have $u_\varepsilon + \theta u = (1-\theta)u_\varepsilon + \theta\upsilon \in U_{ad}$ for all $\theta \in [0,1]$. Then the formulated bang-bang relations follow directly from the integral inequality condition obtained in the theorem. △

7.4.5 Necessary Optimality Conditions under State Constraints

In the concluding part of this section (and of the whole Chap. 7) we develop the limiting procedures to derive necessary optimality conditions for the original *minimax control* problem (P). They are based on passing to the limit in the necessary optimality conditions for the approximating problems $(P_{1\varepsilon})$ and $(P_{2\varepsilon})$ by taking into account the splitting procedure and the strong convergence results proved in the previous subsections. First we summarize the approximation and suboptimality results obtained for the given optimal solution $(\bar{u}, \bar{w})$ to the original problem (P).

Theorem 7.47 (suboptimality conditions for minimax solutions). *Let $(\bar{u}, \bar{w})$ be an optimal solution to the minimax control problem (P), and let $\bar{y}$ be the corresponding mild trajectory to system* (7.61) *under all the assumptions* (H1)–(H6) *with $p > 0$ sufficiently large. Then for each $\varepsilon > 0$ there are optimal solutions $(w_\varepsilon, y_{1\varepsilon})$ and $(u_\varepsilon, y_{2\varepsilon})$ to problems $(P_{1\varepsilon})$ and $(P_{2\varepsilon})$, respectively, which strongly approximate $(\bar{u}, \bar{w})$ in the sense of*

$$(u_\varepsilon, w_\varepsilon, y_{1\varepsilon} + y_{2\varepsilon}) \to (\bar{u}, \bar{w}, \bar{y}) \quad \text{in} \quad L^p(0,T;U) \times L^2(0,T;W) \times \mathcal{C}\big([0,T];X\big)$$

as $\varepsilon \downarrow 0$ and which satisfy the corresponding necessary optimality conditions of Theorems 7.41 *and* 7.45.

Analyzing the above necessary conditions of Theorems 7.41 and 7.45, observe that the possibility of passing to the limit therein as $\varepsilon \downarrow 0$ requires the *uniform boundedness* of the approximating term $\varepsilon \alpha'_\varepsilon(\cdot)\,\alpha_\varepsilon(\cdot)$. This doesn't hold without additional assumptions. Let us impose certain *qualification conditions* on the state constraints in the original minimax problem that seem to be the most appropriate for developing the limiting procedures. It what follows $\|\cdot\|_\infty$ signifies the norm in $L^\infty(Q)$.

(CQ1) There are $\widetilde{w} \in W_{ad}$ and $\eta_1 > 0$ such that for all $\zeta \in L^\infty(Q)$ with $\|\zeta\|_\infty \le 1$ one has

$$a \le \widetilde{y}_1(x,t) + \bar{y}_2(x,t) + \eta_1 \zeta(x,t) \le b \quad \text{a.e. in} \quad Q\ ,$$

where $\widetilde{y}_1$ is the (unique) strong solution to system (7.69) generated by $\widetilde{w}$.

(CQ2) There are $\widetilde{u} \in U_{ad}$ and $\eta_2 > 0$ such that for all $\zeta \in L^\infty(Q)$ with $\|\zeta\|_\infty \le 1$ one has

$$a \le \bar{y}_1(x,t) + \widetilde{y}_2(x,t) + \eta_2 \zeta(x,t) \le b \quad \text{a.e. in} \quad Q\ ,$$

where $\widetilde{y}_2$ is the (unique) mild solution to system (7.70) generated by $\widetilde{y}$.

Note that the qualification conditions imposed above are *different* from infinite-dimensional counterparts of the classical *Slater constraint qualification* for convex programs in the underlying spaces of feasible solutions. In particular, they *don't* imply that the sets of feasible trajectories y_1 and y_2 to problems (P_1) and (P_2) have *nonempty interiors* in the spaces $W^{1,2}\big([0,T];X\big)$ and $\mathcal{C}\big([0,T];X\big)$, respectively.

The next lemma provides the desired *uniform estimates* of the approximation terms that are *crucial* for developing the limiting procedures.

Lemma 7.48 (uniform estimates under constraint qualifications). *Let $(\bar{u}, \bar{w}, \bar{y})$, $(w_\varepsilon, y_{1\varepsilon})$, and $(u_\varepsilon, y_{2\varepsilon})$ satisfy the conditions in Theorem* 7.47.

Assume in addition that the constraint qualifications (CQ1) *and* (CQ2) *are fulfilled. Then there is a constant* $C > 0$ *independent of* ε *such that for any* $\varepsilon > 0$ *we have the estimates*

$$\left\|\varepsilon\alpha'_\varepsilon(y_{1\varepsilon}+\bar{y}_2)\,\alpha_\varepsilon(y_{1\varepsilon}+\bar{y}_2)\right\|_1 \le C, \quad \left\|\varepsilon\alpha'_\varepsilon(\bar{y}_1+y_{2\varepsilon})\,\alpha_\varepsilon(\bar{y}_1+y_{2\varepsilon})\right\|_1 \le C\,,$$

where $\|\cdot\|_1$ *stands for the norm in* $L^1(Q)$.

Proof. Let $\widetilde{w}$ satisfy the constraint qualification condition (CQ1). Consider the perturbation

$$w := \widetilde{w} - w_\varepsilon \quad \text{whenever} \quad \varepsilon > 0$$

and substitute it into the last inequality given in the proof of Theorem 7.41. Using the monotonicity of $\alpha_\varepsilon(\cdot)$, we have the estimates

$$\begin{aligned}
0 &\ge \int_Q \frac{\partial g}{\partial y}(x,t,y_{1\varepsilon}+\bar{y}_2)(\widetilde{y}_1 - y_{1\varepsilon})\,dxdt \\
&+ \int_Q \Big(\frac{\partial f}{\partial w}(x,t,w_\varepsilon) - 2(w_\varepsilon - \bar{w})\Big)(\widetilde{w} - w_\varepsilon)\,dxdt \\
&-2\int_Q \varepsilon\alpha'_\varepsilon(y_{1\varepsilon}+\bar{y}_2)\,\alpha_\varepsilon(y_{1\varepsilon}+\bar{y}_2)\,(\widetilde{y}_1 - y_{1\varepsilon})\,dxdt \\
&= \int_Q \frac{\partial g}{\partial y}(x,t,y_{1\varepsilon}+\bar{y}_2)\,(\widetilde{y}_1 - y_{1\varepsilon})\,dxdt \\
&+ \int_Q \Big(\frac{\partial f}{\partial w}(x,t,w_\varepsilon) - 2(w_\varepsilon - \bar{w})\Big)(\widetilde{w} - w_\varepsilon)\,dxdt \\
&+ \int_Q \varepsilon\alpha'_\varepsilon(y_{1\varepsilon}+\bar{y}_2)\,\big(\alpha_\varepsilon(y_{1\varepsilon}+\bar{y}_2) - \alpha_\varepsilon(\tilde{y}_1+\bar{y}_2+\eta_1\zeta)\big) \\
&\times\big(y_{1\varepsilon}+\bar{y}_2-\tilde{y}_1-\bar{y}_2-\eta_1\zeta\big)\,dxdt + 2\int_Q \varepsilon\alpha'_\varepsilon(y_{1\varepsilon}+\bar{y}_2)\,\alpha_\varepsilon(y_{1\varepsilon}+\bar{y}_2)\,\eta_1\zeta\,dxdt \\
&\ge \int_Q \frac{\partial g}{\partial y}(x,t,y_{1\varepsilon}+\bar{y}_2)(\widetilde{y}_1 - y_{1\varepsilon})\,dxdt \\
&+ \int_Q \Big(\frac{\partial f}{\partial w}(x,t,w_\varepsilon) - 2(w_\varepsilon - \bar{w})\Big)(\widetilde{w} - w_\varepsilon)\,dxdt \\
&+2\int_Q \varepsilon\alpha'_\varepsilon(y_{1\varepsilon}+\bar{y}_2)\,\alpha_\varepsilon(y_{1\varepsilon}+\bar{y}_2)\,\eta_1\zeta\,dxdt
\end{aligned}$$

whenever $\zeta \in L^\infty(Q)$ with $\|\zeta\|_\infty \le 1$. Now it easily follows from assumptions (H4b), (H5b) and Theorem 7.40 that there is a constant $C > 0$ independent of ε ensuring the desired uniform boundedness

$$\int_Q \varepsilon\alpha'_\varepsilon(y_{1\varepsilon} + \bar{y}_2)\,\alpha_\varepsilon(y_{1\varepsilon} + \bar{y}_2)\,\zeta\,dxdt \le C$$

for all $\varepsilon > 0$ and $\zeta \in L^\infty(Q)$ with $\|\zeta\|_\infty \le 1$. The latter obviously implies the first estimate in the lemma.

To prove the second estimate claimed in the lemma, we take $\widetilde{u}$ satisfying the constraint qualification condition (CQ2) and substitute the control

$$u := \widetilde{u} - u_\varepsilon \quad \text{whenever} \quad \varepsilon > 0$$

into the last inequality given in the proof of Theorem 7.45. Again using the monotonicity of $\alpha_\varepsilon(\cdot)$, we have

$$\begin{aligned}
0 \le & \int_Q \frac{\partial g}{\partial y}\big(x,t,\bar{y}_1 + y_{2\varepsilon}\big)\,\mathcal{L}(\widetilde{u} - u_\varepsilon)\,dxdt + \int_\Sigma \frac{\partial h}{\partial u}\big(s,t,u_\varepsilon\big)(\widetilde{u} - u_\varepsilon)\,dsdt \\
& +2p\int_0^T \|u_\varepsilon(t) - \bar{u}(t)\|_U^{p-2}\Big(\int_\Gamma (u_\varepsilon - \bar{u})(\widetilde{u} - u_\varepsilon)\,ds\Big)dt \\
& +2\int_Q \varepsilon\alpha'_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\,\alpha_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\,(\mathcal{L}\widetilde{u} - \mathcal{L}u_\varepsilon)\,dxdt \\
\le & \int_Q \frac{\partial g}{\partial y}\big(x,t,\bar{y}_1 + y_{2\varepsilon}\big)(\widetilde{y}_2 - y_{2\varepsilon})\,dxdt + \int_\Sigma \frac{\partial h}{\partial u}\big(s,t,u_\varepsilon\big)(\widetilde{u} - u_\varepsilon)\,dsdt \\
& +2p\int_0^T \|u_\varepsilon(t) - \bar{u}(t)\|_U^{p-2}\Big(\int_\Gamma (u_\varepsilon - \bar{u})(\widetilde{u} - u_\varepsilon)\,ds\Big)dt \\
& -2\int_Q \varepsilon\alpha'_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\,\big(\alpha_\varepsilon(\bar{y}_1 + y_{2\varepsilon}) - \alpha_\varepsilon(\bar{y}_1 + \widetilde{y}_2 + \eta_2\zeta)\big) \\
& \times\big(\bar{y}_1 + y_{2\varepsilon} - \bar{y}_1 - \widetilde{y}_2 - \eta_2\,\zeta\big)\,dxdt - 2\eta_2\int_Q \varepsilon\alpha'_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\,\alpha_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\,\zeta\,dxdt \\
\le & \int_Q \frac{\partial g}{\partial y}\big(x,t,\bar{y}_1 + y_{2\varepsilon}\big)(\widetilde{y}_2 - y_{2\varepsilon})\,dxdt + \int_\Sigma \frac{\partial h}{\partial u}\big(s,t,u_\varepsilon\big)(\widetilde{u} - u_\varepsilon)\,dsdt \\
& +2p\int_0^T \|u_\varepsilon(t) - \bar{u}(t)\|_U^{p-2}\Big(\int_\Gamma (u_\varepsilon - \bar{u})(\widetilde{u} - u_\varepsilon)\,ds\Big)dt \\
& -2\eta_2\int_Q \varepsilon\alpha'_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\,\alpha_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\,\zeta\,dxdt
\end{aligned}$$

for all $\zeta \in L^\infty(Q)$ with $\|\zeta\|_\infty \le 1$. Then we conclude, using assumptions (H4b), (H6b) and Theorem 7.45, that there is a constant $C > 0$ independent of $\varepsilon > 0$ such that

$$\int_Q \varepsilon\alpha'_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\,\alpha_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\,\zeta\,dxdt \le C$$

whenever $\varepsilon > 0$ and $\zeta \in L^\infty(Q)$ with $\|\zeta\|_\infty \le 1$. The latter implies the second estimate in the lemma and completes the proof. △

Given the optimal trajectory to the minimax problem (P), define the set

$$Q_{ab} := \Big\{(x,t) \in Q \;\Big|\; \bar{y}(x,t) = a \;\text{ or }\; \bar{y}(x,t) = b\Big\}\,,$$

where the state constraints (7.62) are *active*. This set plays a significant role in characterizing topological limits of the functions estimated in Lemma 7.48 that are considered as elements of the dual space $L^\infty(Q)^*$. It is well known that the space $L^\infty(Q)^*$ can be identified with the space $ba(Q)$ of those bounded additive functions, sometimes called *generalized measures*, on subsets of Q that vanish on sets of Lebesgue measure zero. This means that for any $\Lambda \in L^\infty(Q)^*$ there is a unique measure $\lambda \in ba(Q)$ such that

$$\Lambda(\beta) = \int_Q \beta\lambda(dxdt) \;\text{ for all }\; \beta \in L^\infty(Q)\,.$$

It what follows we don't distinguish between the spaces $L^\infty(Q)^*$ and $ba(Q)$, i.e., we identify Λ and λ in the above relation. For any $\lambda \in L^\infty(Q)^*$ consider its *support set*, $\operatorname{supp}\lambda$, on which λ is not zero. Recall that support sets are defined up to subsets of Lebesgue measure zero on Q. The convergence in $L^\infty(Q)^*$ is always understood in the *topological* sense as the convergence of *nets*, which is substantially different from the sequential weak* convergence due to a *highly nonsequential* nature of this space. Considering families $\{\Lambda_\varepsilon\}_{\varepsilon>0}$ from $L^\infty(Q)^*$ and extracting weak* convergent subnets of them, we speak for convenience about convergent subnets of $\{\varepsilon\}$.

Lemma 7.49 (net convergence of penalization terms). *Let all the assumptions of Lemma* 7.48 *be fulfilled. Then there are measures* $\lambda_i \in L^\infty(Q)^*$ *with* $\operatorname{supp}\lambda_i \subset Q_{ab}$, $i = 1,2$, *and a subnet of* $\varepsilon \downarrow 0$ *along which*

$$2\varepsilon\alpha'_\varepsilon(y_{1\varepsilon} + \bar{y}_2)\,\alpha_\varepsilon(y_{1\varepsilon} + \bar{y}_2) \to \lambda_1 \;\; \textit{weak}^* \;\textit{in}\;\; L^\infty(Q)^*\,,$$

$$2\varepsilon\alpha'_\varepsilon(\bar{y}_1 + y_{2\varepsilon})\,\alpha_\varepsilon(\bar{y}_1 + y_{2\varepsilon}) \to \lambda_2 \;\; \textit{weak}^* \;\textit{in}\;\; L^\infty(Q)^*\,.$$

Proof. We justify only the first convergence relationship of the lemma; the proof of the second one is similar. For any $\varepsilon > 0$ define a linear functional on the space $L^\infty(Q)$ by

$$\Lambda_{1\varepsilon}(\beta) := 2\int_Q \varepsilon\alpha'_\varepsilon(y_{1\varepsilon} + \bar{y}_2)\,\alpha_\varepsilon(y_{1\varepsilon} + \bar{y}_2)\beta\,dxdt, \quad \beta \in L^\infty(Q)\,.$$

By Lemma 7.48 we have

$$\left|\Lambda_{1\varepsilon}(\beta)\right| \le C\left\|\beta\right\|_\infty \text{ for all } \beta \in L^\infty(Q)\,,$$

which ensures the continuity of each $\Lambda_{1\varepsilon}$ on $L^\infty(Q)$ and, moreover, the *unform boundedness* of the set $\left\{\Lambda_{1\varepsilon}\middle|\ \varepsilon > 0\right\}$ in the space $L^\infty(Q)^*$. Employing the classical result on the weak* (topological) compactness of the unit ball in dual Banach spaces, we find $\Lambda_1 \in L^\infty(Q)^*$ and a subnet of $\{\varepsilon\}$ along which

$$\lim_{\varepsilon\downarrow 0}\Lambda_{1\varepsilon}(\beta) = \lim_{\varepsilon\downarrow 0} 2\int_Q \varepsilon\alpha'_\varepsilon(y_{1\varepsilon}+\bar{y}_2)\,\alpha_\varepsilon(y_{1\varepsilon}+\bar{y}_2)\beta\,dxdt = \Lambda_1(\beta)$$

whenever $\beta \in L^\infty(Q)$. The latter actually justifies the desired convergence for the limiting measure $\lambda_1 \in ba(Q)$, which we identify with Λ_1.

It remains to show that $\operatorname{supp}\lambda_1 \subset Q_{ab}$. Note that due to the state constraints (7.62) the set

$$\left\{(x,t) \in Q\middle|\ \bar{y}(x,t) < a \text{ or b } \bar{y}(x,t) > b\right\}$$

has measure zero. Thus arguing by contradiction and assuming that the support of λ_1 is not contained in Q_{ab}, we find a set $\widetilde{Q}$ such that

$$\begin{cases} \text{mes}\,\widetilde{Q} > 0, \quad \lambda_1(\widetilde{Q}) \ne 0, \quad \text{and} \\ \widetilde{Q} \subset \left\{(x,t) \in Q\middle|\ a < \bar{y}_1(x,t) + \bar{y}_2(x,t) < b\right\}. \end{cases}$$

The latter clearly implies that

$$\widetilde{Q} \subset \bigcup_{r>0} Q_r \text{ with } Q_r := \left\{(x,t) \in Q\middle|\ a + r \le \bar{y}_1(x,t) + \bar{y}_2(x,t) \le b - r\right\}.$$

Noting that $Q_{r_1} \subset Q_{r_2}$ if $r_1 > r_2$, we get

$$\text{mes}\left(\widetilde{Q} \cap Q_r\right) \ne 0 \text{ for all small } r > 0\,.$$

Furthermore, given $\delta > 0$, there is $\tilde{r} > 0$ such that

$$\text{mes}\left(\widetilde{Q} \setminus Q_{\tilde{r}}\right) \le \text{mes}\left(\bigcup_{r>0} Q_r \setminus Q_{\tilde{r}}\right) < \delta\,.$$

Employing the strong convergence $y_{1\varepsilon} \to \bar{y}$ from Theorem 7.40 and then using the classical Egorov theorem, we find $Q_\rho \subset Q_{\tilde{r}} \cap \tilde{Q}$ with $\text{mes}\left((Q_{\tilde{r}} \cap \tilde{Q}) \setminus Q_\rho\right) < \rho$ and a subsequence of $\{y_{1\varepsilon}(x,t)\}$ that converges to $\bar{y}_1(x,t)$ uniformly in Q_ρ. When $\rho > 0$ is sufficiently small, one has $\text{mes}\,(Q_\rho) \ne 0$ and

$$a < y_{1\varepsilon}(x,t) + \bar{y}_2(x,t) < b \text{ in } Q_\rho \text{ for all small } \varepsilon\,.$$

By the structure of $\alpha_\varepsilon(\cdot)$ in (7.73) the latter yields

$$\varepsilon\alpha'_\varepsilon\big(y_{1\varepsilon}(x,t) + \bar{y}_2(x,t)\big)\,\alpha_\varepsilon\big(y_{1\varepsilon}(x,t) + \bar{y}_2(x,t)\big) = 0 \text{ in } Q_\rho$$

whenever $\varepsilon > 0$ is sufficiently small. Observe in addition that

$$\widetilde{Q} = (\widetilde{Q} \cap Q_{\tilde{r}}) \cup (\widetilde{Q} \setminus Q_{\tilde{r}}) = Q_\rho \cup \big((\widetilde{Q} \cap Q_{\tilde{r}}) \setminus Q_\rho\big) \cup (\widetilde{Q} \setminus Q_{\tilde{r}}) \ .$$

Considering now any $\beta \in L^\infty(Q)$ with $\operatorname{supp}\beta \subset \widetilde{Q}$ and denoting

$$\gamma_\varepsilon(x,t) := 2\varepsilon\alpha'_\varepsilon\big(y_{1\varepsilon}(x,t) + \bar{y}_2(x,t)\big)\, \alpha_\varepsilon\big(y_{1\varepsilon}(x,t) + \bar{y}_2(x,t)\big)\, \beta(x,t) \ ,$$

we get the representation

$$\Lambda_{1\varepsilon}(\beta) = \int_{Q_\rho} \gamma_\varepsilon(x,t)\, dxdt + \int_{(\widetilde{Q}\cap Q_{\tilde{r}})\setminus Q_\rho} \gamma_\varepsilon(x,t)\, dxdt$$
$$+ \int_{\widetilde{Q}\setminus Q_{\tilde{r}}} \gamma_\varepsilon(x,t)\, dxdt \ .$$

Since $\gamma_\varepsilon \in L^1(Q)$ and since δ was chosen to be sufficiently small, this implies

$$\Big| \int_{\widetilde{Q}\setminus Q_{\tilde{r}}} \gamma_\varepsilon(x,t)\, dxdt \Big| < \varepsilon \quad \text{whenever} \quad \varepsilon > 0 \ .$$

Taking the above relations into account and combining them with the first uniform estimate of Lemma 7.48, we find $c(\rho) \downarrow 0$ as $\rho \to 0$ such that

$$\big|\Lambda_1(\beta)\big| \le c(\rho) \quad \text{whenever} \quad \beta \in L^\infty(Q) \quad \text{with} \quad \operatorname{supp}\beta \subset \widetilde{Q}$$

for all ρ sufficiently small. Thus $\Lambda_1(\beta) = 0$ for such β, which contradicts our assumption and completes the proof of the lemma. △

Now we are ready to prove necessary optimality conditions for the original minimax problem (P) with state constraints. First we obtain results that characterize the *worst perturbations* in (P). Given elements $\bar{y} \in \mathcal{C}\big([0,T];X\big)$ and $\lambda_1 \in L^\infty(Q)^*$, consider the *adjoint system*

$$\begin{cases} \dfrac{\partial p_1}{\partial t} - A^* p_1 = -\dfrac{\partial g}{\partial y}\big(x,t,\bar{y}\big) + \lambda_1 & \text{a.e. in} \quad Q, \\ p_1(T,x) = 0, & x \in \operatorname{cl}\Omega \ , \\ p_1(s,t) = 0, & (s,t) \in \Sigma \ , \end{cases} \tag{7.82}$$

and define its solution $p_1(x,t)$ in the sense of

$$\int_Q p_1(x,t)\Big(\frac{\partial v}{\partial t} + Av\Big)\, dxdt = \int_Q \frac{\partial g}{\partial y}\big(x,t,\bar{y}(x,t)\big) v\, dtdx$$
$$- \int_Q v\lambda_1(dxdt) \quad \text{whenever} \quad v \in W_0^{2,1,\infty}(Q) \ .$$

The next theorem shows that, along optimal processes to (P), there is a solution to the adjoint system (7.82) belonging to the space $BV\big(0,T;H^{-1}(\Omega)\big)$ of $H^{-1}(\Omega)$-valued functions with bounded variation on $[0,T]$ and satisfying some other conditions. To proceed, we need one more assumption.

(H7) The variable coefficients $a_i(x),\ i=0,\ \ldots,n$, of the elliptic operator A satisfy the conditions

$$a_0(x)\ge 0,\quad a_0(x)-\sum_{i=1}^n\frac{\partial a_i(x)}{\partial x_i}\ge 0\ \text{ for all }\ x\in\Omega\ .$$

Theorem 7.50 (necessary conditions for worst perturbations). *Let* $(\bar u,\bar w,\bar y)$ *be an optimal triple for the original minimax problem* (P) *under assumptions* (H1)–(H5) *and* (H7). *Assume also that the qualification condition* (CQ1) *holds. Then there is a measure* $\lambda_1\in L^\infty(Q)^*$ *with* $\operatorname{supp}\lambda_1\subset Q_{ab}$ *and a trajectory* $p\in BV\big(0,T;H^{-1}(\Omega)\big)\cap L^2\big(0,T;H_0^1(\Omega)\big)\cap L^\infty(0,T;X)$ *to the adjoint system* (7.82) *such that*

$$\int_Q\Big[\Big(B^*p_1+\frac{\partial f}{\partial w}(x,t,\bar w)\Big)\big(w-\bar w\big)\Big]\,dxdt\le 0\ \ \textit{for all}\ \ w\in W_{ad}\ . \tag{7.83}$$

Proof. We prove this theorem by passing to the limit in the necessary optimality conditions of Theorem 7.41 for the approximating problems $(P_{1\varepsilon})$. Let $p_{1\varepsilon}$ be the strong solution to the adjoint system (7.78) corresponding to $(u_\varepsilon,y_{1\varepsilon})$ in Theorem 7.41. Multiplying both parts of this system by $v\in W_0^{2,1,\infty}(Q)$ and integrating the latter by parts, we get the *integral identity*

$$\int_Q p_{1\varepsilon}\Big(\frac{\partial v}{\partial t}+Av\Big)\,dxdt=\int_Q\frac{\partial g}{\partial y}\big(x,t,y_{1\varepsilon}+\bar y_2\big)v\,dxdt$$

$$-\int_Q 2\varepsilon\alpha_\varepsilon'(y_{1\varepsilon}+\bar y_2)\,\alpha_\varepsilon(y_{1\varepsilon}+\bar y_2)\,v\,dxdt\ \ \text{whenever}\ \ v\in W_0^{2,1,\infty}(Q)\ .$$

The strong solution $p_{1\varepsilon}$ to (7.78) is represented in the form

$$p_{1\varepsilon}(t)=-\int_t^T S^*(\tau-t)\Big(\frac{\partial g}{\partial y}\big(\tau,x,y_{1\varepsilon}+\bar y_2\big)-2\varepsilon\alpha_\varepsilon'(y_{1\varepsilon}+\bar y_2)\,\alpha_\varepsilon(y_{1\varepsilon}+\bar y_2)\Big)\,d\tau$$

for all $t\in[0,T]$, where $S^*(\cdot)$ is the strongly continuous semigroup generated by the operator $-A^*$. It follows from the result of Brézis and Strauss [177] that assumption (H7) ensures the *contraction property* of $S^*(\cdot)$ in $L^1(\Omega)$. Employing the latter property with the first estimate of Lemma 7.48, we find a constant $M>0$ independent of ε and t such that

$$\begin{aligned}
&\|p_{1\varepsilon}(t)\|_{L^1(\Omega)}\\
&\le \int_t^T \Big\|S^*(\tau-t)\Big(\frac{\partial g}{\partial y}\big(x,\tau,y_{1\varepsilon}+\bar y_2\big)-2\varepsilon\alpha_\varepsilon'(y_{1\varepsilon}+\bar y_2)\,\alpha_\varepsilon(y_{1\varepsilon}+\bar y_2)\Big)\Big\|_{L^1(\Omega)}d\tau\\
&\le \int_t^T \|S^*(\tau-t)\|\cdot\Big\|\frac{\partial g}{\partial y}\big(x,\tau,y_{1\varepsilon}+\bar y_2\big)-2\varepsilon\alpha_\varepsilon'(y_{1\varepsilon}+\bar y_2)\,\alpha_\varepsilon(y_{1\varepsilon}+\bar y_2)\Big\|_{L^1(\Omega)}d\tau\\
&\le \Big\|\frac{\partial g}{\partial y}\Big\|_1+\big\|2\varepsilon\alpha_\varepsilon'(y_{1\varepsilon}+\bar y_2)\,\alpha_\varepsilon(y_{1\varepsilon}+\bar y_2)\big\|_1\le M<\infty
\end{aligned}$$

whenever $t\in[0,T]$ and $\varepsilon>0$. This means that the family $\{p_{1\varepsilon}\}$ as $\varepsilon>0$ is bounded in $\mathcal{C}\big([0,T];L^1(\Omega)\big)$. Moreover, it follows from (7.78) and Lemma 7.48 that the family $\{\partial p_{1\varepsilon}/\partial t-A^*p_{1\varepsilon}\}$ with $\varepsilon>0$ is bounded in $L^1(Q)$. Then the Sobolev *imbedding theorem* ensures that the family $\{\partial p_{1\varepsilon}/\partial t\}$ is bounded in $L^1\big(0,T;H^{-1}(\Omega)\big)$. Furthermore, based on (7.78) and the previous estimates, one gets the boundedness of $\{p_{1\varepsilon}\}$ in $L^2\big(0,T;H_0^1(\Omega)\big)$ and $L^\infty(0,T;X)$. Now involving standard compactness arguments and the fact that $L^\infty(0,T,X)$ is dual to a *separable* Banach space, we find

$$p_1\in BV\big(0,T;H^{-1}(\Omega)\big)\cap L^2\big(0,T;H_0^1(\Omega)\big)\cap L^\infty(0,T;X)$$

and a *subsequence* of $\{p_{1\varepsilon}\}$ (no relabeling) such that

$$\begin{aligned}
&p_{1\varepsilon}(t)\to p_1(t)\ \text{ strongly in }\ H^{-1}(\Omega)\,,\\
&p_{1\varepsilon}\to p_1\ \text{ strongly in }\ L^2\big(0,T;H_0^1(\Omega)\big)\,,\\
&p_{1\varepsilon}\to p_1\ \text{ weak}^*\text{ in }\ L^\infty(0,T;X)
\end{aligned}$$

as $\varepsilon\downarrow 0$. Passing to the limit in the above integral identity as $\varepsilon\to 0$ and taking into account (the sequential version of) Lemma 7.49, we conclude that p_1 satisfies the adjoint system (7.82). Finally, we arrive at the necessary condition (7.83) by passing to the limit as $\varepsilon\downarrow 0$ in that from Theorem 7.41 with the use of the convergence results from Theorem 7.40 as well as the strong convergence $p_{1\varepsilon}\to p_1$ in $L^2(0,T;X)$. △

Corollary 7.51 (bang-bang relations for worst perturbations). *Under the assumptions of Theorem* 7.50 *we have*

$$\begin{aligned}
&\bar w(x,t)=c\quad a.e.\ \Big\{(x,t)\in Q\ \Big|\ \big(B^*p_1\big)(x,t)+\frac{\partial f}{\partial w}\big(x,t,\bar w(x,t)\big)<0\Big\}\,,\\
&\bar w(x,t)=d\quad a.e.\ \Big\{(x,t)\in Q\ \Big|\ \big(B^*p_1\big)(x,t)+\frac{\partial f}{\partial w}\big(x,t,\bar w(x,t)\big)>0\Big\}\,,
\end{aligned}$$

where $p_1(x,t)$ is the corresponding solution to the adjoint system (7.82).

Proof. This easily follows from (7.83). △

Next we derive necessary optimality conditions for *Dirichlet boundary controls* in the original minimax problem (P) by passing to the limit in the necessary optimality conditions for the approximating problems $(P_{2\varepsilon})$. To perform the limiting procedure, it is crucial to justify that the *mild solution operator* $\mathcal{L}$ in (7.72) is *continuous* from $L^\infty(\Sigma)$ into $L^\infty(\Omega)$. The following theorem ensures this property and establishes the desired necessary optimality conditions for Dirichlet boundary controls in the original state-constrained problem.

Theorem 7.52 (necessary optimality conditions for Dirichlet boundary controls). *Let* $(\bar u, \bar w, \bar y)$ *be an optimal triple for the minimax problem* (P) *under assumptions* (H1)–(H4) *and* (H6)*. Assume also that the qualification condition* (CQ2) *hold. Then there is a measure* $\lambda_2 \in L^\infty(Q)^*$ *with support* $\operatorname{supp}\lambda_2 \subset Q_{ab}$ *such that*

$$\begin{cases} 0 \le \displaystyle\int_\Sigma \Big[\mathcal{L}^*\Big(\frac{\partial g}{\partial y}(x,t,\bar y)\Big) + \frac{\partial h}{\partial u}(s,t,\bar u)\Big](u-\bar u)\,dsdt \\ + \displaystyle\int_\Sigma (u-\bar u)\big(\mathcal{L}^*\lambda_2\big)(dsdt) \quad \text{whenever } u \in U_{ad}\,. \end{cases} \tag{7.84}$$

Proof. Let $(u_\varepsilon, y_{2\varepsilon})$ be optimal solutions to problems (P_ε) that *strongly* converge as $\varepsilon \downarrow 0$ to the given optimal solution $(\bar u, \bar y_2)$ by Theorem 7.44 and satisfy necessary optimality conditions of Theorem 7.45 for each $\varepsilon > 0$. It follows directly from Theorem 7.45 that

$$\begin{cases} 0 \le \displaystyle\int_\Sigma \Big[\mathcal{L}^*\Big(\frac{\partial g}{\partial y}(x,t,\bar y_1 + y_{2\varepsilon}) + 2\varepsilon\alpha_\varepsilon'(\bar y_1 + y_{2\varepsilon})\,\alpha_\varepsilon(\bar y_1 + y_{2\varepsilon})\Big) + \\ \displaystyle\frac{\partial h}{\partial u}(s,t,u_\varepsilon)\Big](u-u_\varepsilon)\,dsdt \\ +2p\displaystyle\int_0^T \|u_\varepsilon - \bar u\|_U^{p-2}\Big(\int_\Gamma (u_\varepsilon - \bar u)(u-u_\varepsilon)\,ds\Big)\,dt \end{cases} \tag{7.85}$$

whenever $u \in U_{ad}$. We need to justify the passage to the limit in (7.85) as $\varepsilon \downarrow 0$ along a subnet. The convergence results of Theorem 7.44 and the continuity of the operator $\mathcal{L}^*\colon L^2\big(0,T;L^2(\Omega)\big) \to L^2\big(0,T;L^2(\Gamma)\big)$ (see Lasiecka and Triggiani [743]) implies that

$$\int_\Sigma \Big[\mathcal{L}^*\Big(\frac{\partial g}{\partial y}(x,t,\bar y_1 + y_{2\varepsilon})\Big) + \frac{\partial h}{\partial u}(s,t,u_\varepsilon)\Big]\,(u-u_\varepsilon)\,dsdt$$

$$\to \int_\Sigma \Big[\mathcal{L}^*\Big(\frac{\partial g}{\partial y}(x,t,\bar y)\Big) + \frac{\partial h}{\partial u}(s,t,\bar u)\Big]\,(u-\bar u)\,dsdt$$

for all $u \in U_{ad}$, and that the last term in (7.85) converges to zero as $\varepsilon \downarrow 0$. To derive (7.84) from (7.85), it remains to show that

$$\int_{\Sigma}(u-u_\varepsilon)\mathcal{L}^*\Big(2\varepsilon\alpha'_\varepsilon(\bar{y}_1+y_{2\varepsilon})\,\alpha_\varepsilon(\bar{y}_1+y_{2\varepsilon})\Big)\,dsdt$$

$$\to\int_{\Sigma}(u-\bar{u})\big(\mathcal{L}^*\lambda_2\big)\,(dsdt)\ \text{ whenever }\ u\in U_{ad}$$

as $\varepsilon\downarrow 0$ along a subnet. Taking Lemma 7.49 into account, the latter clearly follows from the weak* continuity of the adjoint operator

$$\mathcal{L}^*\colon L^\infty(Q)^*\to L^\infty(\Sigma)^*\ ,$$

which is a direct consequence of the strong continuity of the mild solution operator $\mathcal{L}$ in (7.72) considered from $L^\infty(\Sigma)$ into $L^\infty(Q)$. To justify the latter continuity, we involve some results from the theory of *generalized solutions* to parabolic equations along with the previous considerations.

Take a function $\upsilon\in L^2(\Sigma)$ in the Dirichlet boundary condition for (7.70). Employing Theorem 9.1 from Lions' book [791], we know that there is a unique $y(\upsilon)\in L^2(Q)$, called a *generalized solution* to (7.70), such that

$$\int_Q y(\upsilon)\Big(-\frac{\partial z}{\partial t}+A^*z\Big)\,dxdt=-\int_{\Sigma}\upsilon\frac{\partial\upsilon}{\partial\nu_A}\,dsdt \tag{7.86}$$

whenever $z\in H^{2,1}(Q)$ with $z(s,t)=0$ as $(s,t)\in\Sigma$ and $z(T,x)=0$. Take a *mild* solution $y=\mathcal{L}\upsilon$ of the system (7.70) generated by some $\upsilon\in L^\infty(\Sigma)$ and show that this y is a generalized solution to (7.70) in the sense of (7.86).

To proceed, consider the given Dirichlet boundary control υ as an element of the space $L^p(0,T;U)$ with p sufficiently large and use the fact that the domain space $\mathcal{D}(\Sigma)$ is *dense* in $L^p(0,T;U)$, i.e., there is a control sequence $\{\upsilon_k\}\subset\mathcal{D}(\Sigma)$ with

$$\upsilon_k\to\upsilon\ \text{ strongly in }\ L^p(0,T;U)\ \text{ as }\ k\to\infty\ .$$

As well known, for each $\upsilon_k\in\mathcal{D}(\Sigma)$ system (7.70) admits a unique *classical* solution y_k that automatically is a mild solution and a generalized solution to this system. Thus $y_k=\mathcal{L}(\upsilon_k)$ and y_k satisfies (7.86) for all $k\in I\!N$. Furthermore, it follows from the regularity result of Theorem 7.33 that

$$\big\|\mathcal{L}\upsilon-y_k\big\|_{C([0,T];X)}=\big\|\mathcal{L}\upsilon-\mathcal{L}\upsilon_k\big\|_{C([0,T];X)}\to 0\ \text{ as }\ k\to\infty\ .$$

Combining all these facts, we get the estimates

$$
\begin{aligned}
&\Big|\int_Q \mathcal{L}v\Big(-\frac{\partial z}{\partial t}+A^*z\Big)\,dxdt+\int_\Sigma v\frac{\partial z}{\partial \nu_A}\,dsdt\Big| \\
&\le \Big|\int_Q \big(\mathcal{L}v-y_k\big)\Big(-\frac{\partial z}{\partial t}+A^*z\Big)\,dxdt\Big|+\Big|\int_\Sigma \big(v-v_k\big)\frac{\partial z}{\partial \nu_A}dsdt\Big| \\
&\le \big\|\mathcal{L}v-y_k\big\|_{\mathcal{C}([0,T];X)}\cdot\Big\|-\frac{\partial z}{\partial t}+A^*z\Big\|_{L^2(0,T;X)}T^{1/2} \\
&+\big\|v-v_k\big\|_{L^p(0,T;U)}\Big\|\frac{\partial z}{\partial \nu_A}\Big\|_{L^2(0,T;U)}T^{1/\widetilde{q}}\to 0\ \text{ as }\ k\to\infty\,,
\end{aligned}
$$

where $\widetilde{q}:=2(p-1)/p-2$. This gives

$$
\int_Q \mathcal{L}v\Big(-\frac{\partial z}{\partial t}+A^*z\Big)\,dxdt=-\int_\Sigma v\frac{\partial z}{\partial \nu_A}\,dsdt
$$

whenever $z\in H^{2,1}(Q)$ with $z(s,t)=0$ as $(s,t)\in\Sigma$ and $z(T,x)=0$. The latter means that the mild solution $y=\mathcal{L}v$ is a generalized solution to (7.70) for any $v\in L^\infty(\Sigma)$. Using finally the uniqueness of generalized solutions and the fact that the generalized solution operator is a *continuous* mapping from $L^\infty(\Sigma)$ into $L^\infty(Q)$ (see the afore-mentioned book by Lions), we conclude that the linear operator $\mathcal{L}$ under consideration is continuous from $L^\infty(\Sigma)$ into $L^\infty(Q)$. This completes the proof of the theorem. △

Summarizing the results obtained, we arrive to the following theorem that provides necessary conditions for both worst disturbances and Dirichlet optimal control to the original minimax problem.

Theorem 7.53 (characterizing minimax optimal solutions). *Let* $(\bar{u},\bar{w})$ *be an optimal solution to the minimax problem* (P)*, and let* $\bar{y}$ *be the corresponding trajectory of the parabolic system* (7.61)*. Assume that all the hypotheses* (H1)–(H7) *and the constraint qualification conditions* (CQ1) *and* (CQ2) *hold. Then there are measures* $\lambda_i\in L^\infty(Q)^*$ *with* $\operatorname{supp}\lambda_i\subset Q_{ab}$ *for* $i=1,2$ *and an adjoint trajectory*

$$
p_1\in BV\big(0,T;H^{-1}(\Omega)\big)\cap L^2\big(0,T;H_0^1(\Omega)\big)\cap L^\infty(0,T;X)
$$

to (7.82) *such that the optimality conditions* (7.83) *and* (7.84) *are satisfied.*

Remark 7.54 (feedback control design). The results derived in this section allow us to determine the structures of worst perturbations and optimal boundary controls for the *open-loop* minimax control problem (P). They are also useful for the *minimax design of closed-loop* parabolic control systems, where the purpose is to construct *feedback controls* depending on *state* variables and ensuring satisfactory (at least *stable*) behavior under arbitrary perturbation from the admissible region W_{ad} with the *best* performance in the case of worst perturbations. Such feedback control problems have been considered in Mordukhovich [905, 918] and Mordukhovich and Shvartsman [957], where the reader can find more discussions and references. A typical problem studied in the afore-mentioned papers is as follows:

$$\text{minimize}\quad J(u) := \max_{w\in W_{ad}} \int_0^T \big|u\big(y(x_0,t)\big)\big|\,dt$$

over $u \in U_{ad}$ subject to the parabolic system

$$\begin{cases} \dfrac{\partial y}{\partial t} + Ay = w(t) & \text{a.e. in } \; Q\ , \\[2ex] y(x,0) = 0, & x \in \Omega\ , \\[2ex] y(s,t) = u(t), & t \in \Sigma\ , \end{cases}$$

the pointwise state constraints

$$\big|y(x_0,t)\big| \le \eta \quad \text{for all} \quad t \in [0,T]\ ,$$

and the *feedback control law*

$$u(t) = u\big(y(x_0,t)\big)\ ,$$

where $x_0 \in \Omega$ is a given point at which all the information about the system output is collected, where the admissible perturbation and control regions are defined in (7.63) and (7.64), respectively, and where A is a self-adjoint and strongly uniformly elliptic operator given by

$$A := -\sum_{i,j=1}^n a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} - c$$

with $c \in I\!R$ and $a_{ij} \in \mathcal{C}^\infty(\mathrm{cl}\,\Omega)$. Besides conducting an efficient *open-loop* variational analysis and approximation procedures, we exploit *monotonicity* properties of the parabolic dynamics and *asymptotic* characteristics of trajectories on the *infinite horizon*, which allow us to develop an efficient minimax design of *feedback suboptimal* controls that ensure the required *stability in the large* of highly nonlinear *closed-loop* control systems; see the afore-mentioned papers for more details, numerical analysis, and open problems.

7.5 Commentary to Chap. 7

7.5.1. Control Systems with Distributed versus Lump Parameters. Chapter 7 is devoted to problems of dynamic optimization and optimal control for some classes of systems with the so-called "distributed parameters." There is a traditional division in control theory between systems governed by ordinary differential equations (with their discrete-time counterparts) and those systems whose dynamics is described by more complicated equations involving, e.g., various *time delays* (or, more generally, by *functional-differential* equations) as well as systems governed by *partial differential* equations of different types (elliptic, parabolic, hyperbolic, etc.). The main issue that determined this separation, starting at least from the early years of optimal control theory, was the *dimension* (finite or infinite) of the underlying *state space*. In particular, the natural state space for the system governed by ordinary differential equations

$$\dot{x}(t) = f\big(x(t), u(t), t\big), \quad x(t) \in I\!R^n \text{ as } t \in [a,b], \tag{7.87}$$

is *finite-dimensional*; the current state $x(t)$ of (7.87) as $t \in [a,b]$ can be fully determined, for each given control $u(t)$, if the initial state vector $x(a) = x_0 \in I\!R^n$ is given. Such finite-dimensional systems are also known as systems with *lump parameters*.

In contrast, the natural state space for the simplest *delay* control system

$$\dot{x}(t) = f\big(x(t), x(t-\theta), u(t), t\big), \quad x(t) \in I\!R^n \text{ as } t \in [a,b], \tag{7.88}$$

with a single time delay $\theta > 0$ in state variables is *infinite-dimensional*, since we need to know the *initial function* $x(t)$ on the whole "initial tail" interval $t \in [a-\theta, a]$ to determine the current state $x(t)$ for $t \in [a,b]$. Similar (often more involved) situations occur in control processes governed by partial differential equations, integral equations, etc., which are therefore unified as systems with *distributed parameters*.

Roughly speaking, systems with lump parameters are associated with control processes described by ordinary differential equations of type (7.87) in finite dimensions (as well as with discrete-time counterparts and more general differential and discrete inclusion models), while distributed-parameter systems are related to various descriptions involving infinite-dimensional state spaces.

It seems that such a broad traditional classification is rather conditional and doesn't reflect many specific features of the state dynamics that cannot be only characterized by the dimension of state spaces. In particular, dynamic optimization and control problems studied in Chap. 6 in general infinite-dimensional state settings definitely belong to the distributed-parameter theory according to this classification. On the other hand, the *ordinary* differential form of the underlying dynamics plays a crucial role in the methods developed and results obtained, although the infinite dimensionality of state spaces indeed requires more involved consideration.

Chapter 7 concerns several classes of evolution control systems with distributed parameters whose dynamical descriptions are significantly different from each other as well as from their ordinary counterparts studied in Chap. 6. Besides general techniques and structures of variational analysis (which are largely in common for the major developments in Chaps. 6 and 7), the methods and results established for the classes of distributed control systems in

Chap. 7 are essentially based on specific features of the systems under consideration.

7.5.2. Systems with Time Delays in State Variables. Section 7.1 is devoted to the study of dynamic optimization problems for systems with *time delays*. Dynamical processes in such systems (known also as time-lag, hereditary, retarded, differential-difference, functional-differential systems as well as systems with aftereffect, with deviating arguments, etc.) depend on the "prehistoty," which makes them infinite-dimensional even in the case of finite-dimensional state vector $x \in I\!R^n$. The qualitative theory of hereditary systems (with no control) began by Volterra (see his book [1298]), while the intensive development has started in 1950s with the publication of the pioneering book by Myshkis [989]; see the subsequent books by Bellman and Cooke [94], Élsgolts and Norkin [404], Hale [538], Kolmanovskii and Myshkis [694], and the references therein.

First results on optimal control of delay systems were obtained by Kharatishvili [678] who derived an analog of the Pontryagin maximum principle for delay-differential equations of type (7.88) involving a single *time delay* in *state variables*. Then these results were extended to systems with variable and distributed time delays in state and control variables, with various constraints including those of essentially infinite-dimensional types, etc. Among early contributions to optimal control theory for delay-differential systems we mention the research by Banks [78], Friedman [476], Gabasov and Churakova [483], Gabasov and Kirillova [486], Halanay [537], Krasovskii [700], Oğuztöreli [1018], and Warga [1315]. More advanced results in this and related directions for controlled hereditary systems with various time delays in state and control variables and more general constraints were subsequently developed in numerous publications; see, e.g., [81, 101, 275, 281, 485, 486, 506, 679, 694, 696, 701, 867, 901, 1015, 1019, 1173, 1174, 1321, 1323] among others.

7.5.3. Hereditary Systems of Neutral Type. A very interesting class of hereditary control systems substantially different from both ODE systems (7.87) and their delay-differential counterparts (7.88) is described by

$$\dot{x}(t) = f\big(x(t), x(t-\theta), \dot{x}(t-\theta), u(t), t\big) \tag{7.89}$$

and is known as the class of *functional-differential systems of neutral type*, or simply as *neutral control systems*. Such systems (with no control) under this name of "differential equations with deviating arguments of neutral type" were first considered by Élsgolts [403] in the qualitative theory of differential equations with aftereffect; see also Bellman and Cooke [94] and Hale [538]. The main difference between systems (7.88) and (7.89) is that the latter contain time *delays* in *velocity* variables, not only in state and/or control ones. This makes control problems for neutral systems (7.89) significantly more complicated in comparison with those for (7.87) and (7.88). In particular, there is *no analog* of the Pontryagin maximum principle held for neutral control systems in general *nonconvex* setting; see more discussion in Remark 6.41 and the example therein, which is taken from Gabasov and Kirillova [485].

Probably first results on necessary optimality conditions for neutral variational and control systems were independently obtained by Hughes [586] and by Kamenskii and Hvilon [663]; see also the early papers by Sabbagh [1185] and Kent [668]. Various subsequent developments can be found in Angell and Kirsch [19], Banks and Kent [79], Banks and Manitius [81], Chukwu [241], Élsgolts and Norkin [404], Gabasov and Kirillova [485], Gorelik and Mordukhovich [513, 514], Gusakova [529], Jacobs and Langenhop [626], Kisielewicz [682], Kharatishvili and Tadumadze [679], Kolmanovskii and Nosov [695], Kolmanovskii and Shaikhet [696], Mansimov [843], Melikov [868], Mordukhovich [895, 896, 901], Mordukhovich and Sasonkin [945], Salamon [1187], Tadumadze and Alkhazishvili [1242], and in the references therein.

Besides the afore-mentioned nonvalidity of the Pontryagin maximum principle for nonconvex neutral control systems, there are other important issues for which the presence of time delays in velocity variables significantly distinguish neutral systems from their ordinary and delay-differential counterparts. Let us particularly recall a variety of the *adjoint systems* in first-order necessary optimality conditions and related topics [81, 485, 513, 529, 663, 668, 679, 868, 901]; the unavoidable presence of *jumps* in optimality conditions; more restrictive conditions ensuring the *relaxation stability* [682]; the essential dependence of various results on the *time behavior* of system components involving delays in velocities [81, 485, 695, 901, 945]; the influence of *discontinuous initial conditions* on the form of major results on controllability, observability, optimality, and duality aspects [895, 896, 901, 1242]; occurring specific *intermediate* (between the first and second order) necessary optimality conditions in neutral systems *nonlinear* with respect to $\dot{x}(t-\theta)$ [513, 901]; new *second-order* conditions of the Legendre-Clebsch type for nonlinear neutral systems with no constraints on control variables [514, 901]; a variety of necessary optimality conditions for *singular controls* [514, 843, 901], etc. Note also that, in contrast to ordinary and delay-differential systems, the *approximative maximum principle* is *not* generally valid for finite-difference approximations of smooth neutral systems with no endpoint constraints; see Example 6.70 from Chap. 6.

Observe that neutral systems exhibit a lot of similarities with *discrete-time* systems of optimal control. In a sense, it is not surprising, since (7.89) may be viewed as a discrete system with respect to *velocities* $\dot{x}(t)$, where the time delay $\theta > 0$ plays a role of the discrete stepsize. On the other hand, certain results obtained for neutral systems have their counterparts in control theory for partial differential equations of the *hyperbolic type*, particularly for the so-called systems of *Goursat-Darboux*; cf. Ashchepkov and Vasiliev [42], Cernea [233], Gavrilov and M. Sumin [500], Mahmudov [827], Mansimov [844], Plotnikov and V. Sumin [1085], Srochko [1221], and Vasiliev [1281]. At the same time, there are some classes of hyperbolic systems (e.g., those governed by *telegraph equations* and the like) that can be equivalently reduced to neutral systems; see Kolmanovskii and Nosov [695] and the references therein.

7.5.4. Delay-Differential Inclusions. A variety of problems in dynamic optimizations for delay systems more involved than (7.88) are described by *delay-differential inclusions* of the type

$$\dot{x}(t) \in F\big(x(t), x(t-\theta), t\big) \ \text{ a.e. } \ t \in [a,b] \tag{7.90}$$

with the initial and endpoint conditions

$$x(t) = c(t), \ \ t \in [a-\theta, a), \quad \big(x(a), x(b)\big) \in \Omega \ ,$$

which were first considered by Clarke and Watkins [274] under the name of "differential-difference inclusions." Assuming the compactness and *convexity* of the sets $F(x,y,t)$ together with the Lipschitz continuity of $F(\cdot,\cdot,t)$, the authors derived necessary optimality conditions for the Mayer problem of minimizing $\varphi(x(b))$ on absolutely continuous trajectories of (7.90). The results obtained in [274] were expressed in Clarke's Hamiltonian form extending that in [255] with the corresponding transversality conditions given in terms of his constructions $\partial_C\varphi$ and $N_C(\cdot;\Omega)$. Besides necessary optimality conditions, the paper [274] contained related results on computing generalized gradients of the value function depending on endpoint perturbations, with their applications to local controllability.

The Hamiltonian conditions of [274] were extended by Clarke and Wolenski [276] to the Bolza problem involving more general hereditary inclusions written as

$$\dot{x}(t) \in F(x_t, t) \ \text{ a.e. } \ t \in [a,b] \ ,$$

with $x_t\colon [-r,0] \to I\!\!R^n$ given by $x_t(s) := x(t+s)$. The necessary optimality conditions of [276] were derived by using *perturbation techniques* of proximal analysis in infinite-dimensional spaces.

Another approach to optimization of delay-differential inclusions in form (7.90) was developed by Minchenko [878] who extended the primal-space constructions by Polovinkin and Smirnov [1094] and Frankowska [465] to the case of delay systems. The necessary optimality conditions obtained in [878] were expressed in terms of *tangential approximations* being generally independent of those in [274]. Further results in this direction can be found in Minchenko and Volosevich [881]. We also refer the reader to the recent paper by Cernea [234], where tangential techniques and directional derivatives were used for deriving *second-order* necessary optimality conditions for some delay-differential inclusions.

The papers by Mordukhovich [921] and by Mordukhovich and Trubnik [959] developed the method of *discrete approximations* to the study of optimization problems governed by delay-differential inclusions (7.90). The results obtained in the vein of [915] justified well-posedness and strong convergence procedures for discrete approximations of (7.90) and established necessary optimality conditions in both *extended Euler-Lagrange* and *Hamiltonian* forms (the latter only for convex-valued inclusions) using a *partial* convexification of basic normals and subgradients, in contrast to the *full* convexification in the Hamiltonian inclusions given in [274, 276]. Further results in this direction were derived by Mordukhovich and L. Wang [973] for the Bolza problem over trajectories of delay-differential inclusions (7.90) satisfying the *multivalued* "initial tail" condition

$$x(t) \in C(t) \ \text{ a.e. } \ t \in [a-\theta, a) \, .$$

The latter is *specific* for time-delay systems and provides an additional source for control and optimization.

7.5.5. Neutral-Differential Inclusions. First necessary conditions for optimization problems governed by *neutral functional-differential inclusions*

$$\frac{d}{dt}\Big[x(t) - Ax(t-\theta)\Big] \in F\big(x(t), x(t-\theta), t\big) \ \text{ a.e. } \ t \in [a,b] \tag{7.91}$$

were established by Mordukhovich and L. Wang: the Mayer problems was considered in [972], and a comprehensive treatment for the Bolza problem with endpoint constraints was given in [974]; see also [977] for a more general and complicated case of nonautonomous systems with the time-dependent matrix $A = A(t)$ in (7.91).

Note that the neutral-type operator on the left-hand side of (7.91) is given in the so-called *Hale form* [538], which is essential for the techniques and results developed in [972, 974, 977]. The approach of these papers followed that by Mordukhovich [915] based on *discrete approximations.* As expected, the case of neutral systems happened to be significantly more involved and, in contrast to ordinary and delay-differential inclusions, did not allow us to obtain any results without *relaxation stability.* The latter property may be rather restrictive for nonconvex neutral systems; see Kisielewicz [682] for some sufficient conditions for its validity. The main necessary optimality conditions of [972, 974, 977] were derived in the *extended Euler-Lagrange form*, which implied the corresponding analog of the Weierstrass-Pontryagin maximum condition and also (by applying Rockafellar's dualization theorem [1162]) the refined Hamiltonian condition under the assumed relaxation stability, particularly for *convex-valued* functional-differential inclusions of neutral type.

The recent paper by Ortiz [1021] concerned necessary conditions for the *generalized Bolza problem* for *neutral systems* with *varying delays* $\theta = \theta(t)$ written as:

$$\text{minimize } \ \varphi\big(x(a), x(b)\big) + \int_a^b \vartheta\big(x(t), x(t-\theta), \dot{x}(t), \dot{x}(t-\theta), t\big)\, dt, \tag{7.92}$$

where both functions φ and ϑ may be extended-real-valued. This paper was an extension of that by Ortiz and Wolenski [1022] devoted to the *delay-in-state* (not neutral) generalized Bolza problem, where the integrand ϑ in (7.92) didn't depend on the delay velocity term $\dot{x}(t-\theta)$ while depending on the delay state variable $x(t-\theta)$. The main assumption made in [1021] was the *joint convexity* of the integrand ϑ with respect to *both velocity variables* corresponding to $\dot{x}(t)$ and $\dot{x}(t-\theta)$. The approach of [1021] was based on the *decoupling technique* proposed by Clarke [258] for the non-delay Bolza problem and then developed by Ortiz and Wolenski [1022] for systems with time delays. Observe that the main results of [1021], as well as of [258] and of [1022], were given in terms of *fully convexified* Euler-Lagrange and Hamiltonian inclusions of Clarke's type, but not in their significantly more refined forms involving *partially convexified* sets of basic normals and subgradients as in Mordukhovich and L. Wang [974], which extended the corresponding forms of Mordukhovich and Rockafellar for non-delay systems; cf. Chap. 6.

7.5.6. Differential-Algebraic Systems. Section 7.1 is devoted to the study of dynamic optimization problems whose dynamic constrains are described by interrelated *delay-differential inclusions* and *linear delay-algebraic equations* of the type

$$\begin{cases} \dot{z}(t) \in F\big(x(t), x(t-\theta), z(t), t\big) \ \text{ a.e. } \ t \in [a,b]\,, \\ z(t) = x(t) + Ax(t-\theta), \quad t \in [a,b]\,. \end{cases} \tag{7.93}$$

This is a new class of optimal control problems with *distributed parameters* that, on one hand, may be treated as variational problems for *extended neutral* inclusions while, on the other hand, it is related to a special class of *delay differential-algebraic* systems governed by general delay-differential inclusions with linear delay-algebraic *links* between "slow" and "fast" variables. Observe that the integrand ϑ in the Bolza functional of problem (DA) considered in Sect. 7.1 depends on both slow and fast variables (denoted by z and x, respectively) as well as on the *time derivative of slow variables*, while fast variables may not be differentiable in time.

Note that system (7.93) can be written in the form

$$\frac{d}{dt}\Big[x(t) + Ax(t-\theta)\Big] \in F\big(x(t), x(t-\theta), x(t) + Ax(t-\theta), t\big) \ \text{ a.e. } \ t \in [a,b]\,,$$

which is an extension of (7.91) and may be reduced to the general neutral inclusion

$$\dot{x}(t) \in G\big(x(t), x(t-\theta), \dot{x}(t-\theta), t\big) \ \text{ a.e. } \ t \in [a,b] \tag{7.94}$$

provided that $x(t)$ is absolutely continuous on $[a,b]$. We never suppose the latter assuming instead that the *combination* $x(t) + Ax(t-\theta)$ is absolutely continuous. Similarly, the cost functional in (DA) transfers under this substitution into the neutral Bolza form (7.92). Thus problem (DA) can be treated as a special case of Bolza-type variational problems for general neutral inclusions. However, in this way we loose the principal feature of (DA), which is *crucial* for the methods applied and the results obtained in Sect. 7.1. This specific feature of (DA) is as follows: both the dynamic constraint (7.94) and the cost functional (7.92) depend in fact *not* on $\dot{x}(t)$ and $\dot{x}(t-\theta)$ but on the derivative of the *same linear combination* $x(t) + Ax(t-\theta)$. That is why we treat this linear combination as a new state variable in (7.93) and consider (DA) in its natural form, which emphasizes both *delay-differential* and *linear algebraic* constraints on the system dynamics.

In the *non-delay* case of $\theta = 0$, system (7.93) is an inclusion extension of the so-called controlled *differential-algebraic equations* (known as DAEs, for short) that arise in many practical applications, particularly to process system engineering, robotics, mechanical systems with holonomic and nonholonomic constraints, etc.; see, e.g., the book by Brennan, Campbell and Pretzold [174] with many examples, discussions, and references. Generally DAE control systems are given by

$$\begin{cases} \dot{z} = f\big(z(t), x(t), u(t), t\big) \quad \text{a.e. } t \in [a,b]\,, \\[2ex] 0 = g\big(z(t), x(t), u(t), t\big) \quad \text{a.e. } t \in [a,b]\,. \end{cases} \tag{7.95}$$

They are closely related to other special classes of control systems known under different names: implicit systems, singular systems, descriptors, etc.; see, e.g., Dai [305], Devdariani and Ledyaev [327], and their references. In the early Russian literature such systems were studied under the long name of "control systems that are not solved with respect to the derivative;" see Vasiliev [1279], Gabasov and Kirillova [485], Gusakova [528], Kurina [730], and Mordukhovich [901] among other publications on optimal control for systems of this type. Note that the DAEs in (7.95) can be viewed as the limiting case of the *singularly perturbed* control systems

$$\begin{cases} \dot{z} = f\big(z(t), x(t), u(t), t\big) \quad \text{a.e. } t \in [a,b]\,, \\[2ex] \varepsilon\dot{x}(t) = g\big(z(t), x(t), u(t), t\big) \quad \text{a.e. } t \in [a,b] \end{cases} \tag{7.96}$$

as $\varepsilon \downarrow 0$. However, it is well known that the realization of convergence procedures for optimal solutions to (7.96) as $\varepsilon \downarrow 0$ requires fairly restrictive assumptions; cf. Artstein and Gaitsgory [30], Bensoussan [100], Dontchev and Zolezzi [367], Kokotović, Khalil and O'Reilly [693], and the references therein.

To the best of our knowledge, the most advanced results on necessary optimality conditions for control systems with the DAE dynamics (7.95) were derived by de Pinho and Vinter [1079] under the so-called *index one* assumption. They demonstrated the violation of the (strong) Pontryagin maximum principle for such systems, justified it under some convexity assumptions, and established necessary conditions in the new *weak* maximum principle form for systems with nonsmooth differential (not algebraic) dynamics and also with nonsmooth cost and endpoint constraint functions. However, the critical index one assumption made in [1079] seems to be quite restrictive and doesn't hold in many differential-algebraic control systems of practical significance; see, e.g., [174, 1048].

7.5.7. Regularization Role of Time Delay. The results presented in Sect. 7.1 are mostly based on the recent papers by Mordukhovich and L. Wang [975, 976]. To obtain necessary optimality conditions for the class of differential-algebraic control systems under consideration, we use a version of the *method of discrete approximations* that takes into account the presence of the *time delay* $\theta > 0$; the latter happens to be a *regularization factor* allowing us to fully *avoid* the *index one* assumption. Following mainly the procedure developed in Sect. 6.1 for ordinary evolution systems, we establish the well-posedness and strong convergence of discrete approximations, derive necessary optimality conditions for approximating difference-algebraic systems, and then justify the passage to the limit from the obtained necessary conditions in discrete approximation. This leads us to new necessary optimality conditions of the extended Euler-Lagrange and Hamiltonian types for the original Bolza problem governed by delay differential-algebraic systems subject to endpoint constraints under relaxation stability.

The realization of the method of discrete approximations for differential-algebraic systems with delays is rather different from the case of ordinary systems and technically much more involved. On the other hand, in Sect. 7.1 we assume for simplicity that the state vectors (x, z) are finite-dimensional, which allows us to avoid complications with the usage of SNC calculus in infinite dimensions. Moreover, at each stage we apply more convenient calculi of basic/limiting normals, subgradients and coderivatives instead of fuzzy calculus rules as in Sect. 6.1. The *open question* remains about the possibility of passing to the limit from the obtained necessary optimality conditions for the delay systems under consideration as $\theta \downarrow 0$ to derive valuable results for differential-algebraic control systems with *no delay.*

7.5.8. PDE Control Systems. The remaining three sections of Chap. 7 concern some optimal control problems for distributed-parameter systems governed by *partial differential equations* (PDEs). The literature on PDE optimal control is enormous; so we mention only (a number of) those publications, which are largely related to the topics discussed in this book. The reader can find more information in the books by Ahmed and Teo [4], Balakrishnan [74], Banks and Kunisch [80], Barbu [82], Bensoussan, Da Prato, Delfour and Mitter [101], Butkovsky [209], Cherkaev [237], Denkowski, Migórski and Papageorgiou [323], A. Egorov [392], Fattorini [432], Friedman [478], Fursikov [481], Lagnese [736], Lasiecka and Triggiani [746], Li and Yong [789], Lions [791, 792], Lurie [821], Lyashko [823], Neittaaanmäki and Tiba [997], Tiba

[1255], Tröltzsch [1271], Vasiliev [1281] and in their numerous references devoted to various aspects of the theory and applications of PDE control; see also the recent survey paper by Burns [206].

Probably the first paper on PDE optimal control was published in 1960 by Butkovsky and Lerner [211]; it was devoted to optimal control of the one-dimensional heat equation. After a while, it was realized that many PDE control problems could be written in the form of abstract *evolution systems* in *infinite-dimensional spaces*, but the original ODE approach developed by Pontryagin et al. [1102] didn't generally apply to establish the maximum principle in PDE optimal control. The major technical limitation for this was that the *convex separation* theorem used in [1102] *couldn't* be employed in *infinite dimensions* without additional assumptions. The first example on the violation of the maximum principle for a *singleton* target set in PDE control systems was constructed by Y. Egorov who also derived an appropriate analog of the Pontryagin maximum principle in infinite-dimensional control problems under rather restrictive *interiority* assumptions; see [393, 394]. On the other hand, A. Egorov [391] proved the maximum principle for parabolic and hyperbolic systems with target/constraint sets described by *finitely many* equalities and inequalities with no interiority assumptions. It hasn't been realized for a long time that such sets enjoy the finite codimension property; cf. Chap. 6. We refer the reader to the survey paper by Butkovsky, A. Egorov and Lurie [210] for other early developments in PDE optimal control, mostly in the Russian literature.

In the West, the pioneering work on infinite-dimensional optimal control was done by Fattorini [427] and Balakrishnan [73] who first applied the theory of strongly continuous *semigroups* to linear control systems. Among other significant early contributions we particularly mention the publications by Conti [284], Friedman [477], Lions [791], Russel [1184], Malanowski [830], and Wang [1302].

The crucial importance of the *finite codimension* property of reachable and/or target sets for the fulfillment of the maximum principle in infinite-dimensional control systems was first observed by Li and Yao [786] and then developed in their paper [787]. Further developments in this direction were accomplished by Fattorini [429] and by Li and Yong [788]; see also the book by the latter authors [789] for more results, discussions, and references concerning optimal control problems for various classes of semilinear and quasilinear partial differential equations and other infinite-dimensional control systems.

7.5.9. Boundary Control of PDE Systems. The previous discussions and the afore-mentioned publications mostly relate to PDE control problems with *distributed controls* acting in *state equations*, similarly to ordinary dynamic systems. A specific feature of PDE problems, important from both viewpoints of the theory and applications, is the possibility to consider *boundary controls*, i.e., the presence of control functions in boundary conditions. It has been well recognized that boundary control problems are significantly more involved in comparison with their distributed control counterparts.

There are two major types of boundary conditions: *Dirichlet* and *Neumann*, which are significantly different from each other. Usually Dirichlet boundary conditions offer *less regularity* of the corresponding solution operators as we have seen, in particular, in Sects. 7.2–7.4 of this book. The *mixed* (or Robin) type of boundary conditions resembles, as a rule, basic properties of the Neumann one.

Historically boundary controls in PDE systems were first considered in the paper by Fattorini [428] of 1968, which was mainly motivated by applications to approximate controllability and employed the *semigroup* operator approach. A significant progress for Dirichlet boundary control problems for linear parabolic equations was achieved by Washburn in his paper [1324] based on the earlier Ph.D. thesis under supervision by Balakrishnan who first considered a motivating example for a rectangle; see [74]. Among further significant achievements in the study of boundary control and related problems for various PDE systems we mention the contributions by Arada and Raymond [23, 24] Barbu [82], Barbu, Lasiecka and Triggiani [83], Bonnans and Casas [131], Bucci [183], Cârjă [223], Casas [225], Casas, Raymond and Zidani [226], Fattorini [433], Fattorini and Murphy [435, 436], Lasiecka [739], Lasiecka and Triggiani [742, 743, 746], Lions [791, 792], Mordukhovich and Raymond [943, 944], Mordukhovich and Zhang [978, 979], Nowakowski and Nowakowska [1016], Osipov, Pandolfi and Maksimov [1023], Raymond [1120], Raymond and Zidani [1121], Tröltzsch [1271], and Zuazua [1379]; see also the bibliographies therein. More references and comments on specific results obtained in some of the afore-mentioned publications can be found above in Sects. 7.2–7.4 and below in the corresponding comments to these sections.

7.5.10. Neumann Boundary Control of Hyperbolic Equations. We start presenting the PDE material of this chapter with the *Neumann boundary control* problem for the *semilinear wave equation* considered in Sect. 7.3. This choice is made *not* because of the problem under consideration is the easiest one among those studied in the book—just the opposite: boundary control problems for hyperbolic equations are among the *most challenging* and not sufficiently investigated in PDE control theory. It seems that the first results for such problems in the presence of *pointwise state constraints* have been obtained only quite recently in the paper by Mordukhovich and Raymond [944] on which the material of Sect. 7.2 is based. Some results on necessary optimality conditions for boundary control problems with *no* state constraints and also for distributed control problems governed by hyperbolic equations can be found, e.g., in Bucci [183], Fattorini [431, 432], Lasiecka and Triggiani [746], Lions [792], Malanowski [830], White [1328] and the references therein. However, hyperbolic optimal control problems have not been sufficiently studied in the literature—definitely *much less* that their elliptic and parabolic counterparts. One of the significant technical reasons for this, from the semigroup viewpoint for evolution equations in infinite dimensions, is that hyperbolic systems are associated with *non-compact* semigroups generated by the underlying unbounded operator of the corresponding semilinear equation.

Our choice for the material arrangement in Chap. 7 is mainly motivated by the possibility to efficiently employ powerful techniques of *modern variational analysis* (variational principles, approximations by unconstrained problems, needle-type variations, etc.) to the Neumann boundary control problem under consideration, rather similarly to the case of ODE systems in Chap. 6. Another reason to place hyperbolic systems right after functional-differential systems of *neutral type* is that there are certain *similarities* between some classes of these distributed-parameter systems (see Subsect. 7.5.3), which however have not been much exploited. Recall that the approach in Sect. 7.1 involves the discrete approximation technique that is significantly different from the other methods of Chap. 7.

Let us emphasize that the study in Sect. 7.2 is based on modern techniques of variational analysis married with the deep PDE *regularity theory* for hyperbolic Neumann boundary value problems developed by Lasiecka and Triggiani in the late 1980s—early 1990s; see [744, 745]. This regularity theory and the related PDE developments of Subsect. 7.2.2 provide the *required basis*, which supports the subsequent applications of variational techniques to establish necessary optimality conditions for Neumann boundary controls to state-constrained hyperbolic systems.

7.5.11. Pointwise State Constraints via Ekeland's Variational Principle. Problems with pointwise state constraints are among the most difficult in optimal control theory. It is worse mentioning that the derivation of satisfactory conditions for the maximum principle in *state-constrained* control problems governed by nonlinear *ordinary* differential equations was probably the *primary motivation* for developing a general theory of extremal problems by Dubovitskii–Milyutin [370]; see also the books by Ioffe and Tikhomirov [618] and by Warga [1315] with their treatments and references related to state-constrained problems for ordinary and functional-differential (in [1315]) control systems.

State-constrained problems for control systems governed by elliptic and parabolic equations were originally treated by their reduction to *infinite-dimensional* problems of mathematical programming; this was called the "method of Lagrange multipliers" in the PDE control literature; see, e.g., Mackenroth [825], Tröltzsch [1271], and the bibliographies therein.

A powerful approach to derive necessary optimality conditions for endpoint-constrained and (pointwise) state-constrained ODE control problems via *approximation procedures* based on *Ekeland's variational principle* was initiated by Ekeland himself [397] (see also his excellent survey [399]) and then was strongly developed by Clarke [250, 251, 255] to nonsmooth systems and differential inclusions.

First applications of the Ekeland principle to optimal control problems governed by partial differential equations were done probably by Li and Yao [786, 787] and by Plotnikov and M. Sumin [1084]. Then this approach to derive necessary optimality conditions of the maximum principle type in various PDE control problems governed by nonlinear (mostly semilinear) *elliptic* and *parabolic* equations was developed by Arada and Raymond [23, 24], Bonnans and Casas [131], Casas [225], Casas, Raymond and Zidani [226], Casas and Yong [227], Fattorini [429, 430, 431, 432, 433], Fattorini and Frankowska [434], Fattorini and Murphy [435, 436], Li and Yong [788, 789], Raymond [1120], Raymond and Zidani [1121], and by other researchers. As mentioned, the first results for *Neumann* boundary control of state-constrained semilinear *hyperbolic* equations were obtained in Mordukhovich and Raymond [944] by using Ekeland's variational principle as one of the basic ingredients of their analysis.

Note that the implementation of approximating procedures based on Ekeland's variational principle is much more involved for control problems governed by partial differential equations in comparison with the case of ODE systems. In particular, there is a significant difference between bounded and *unbounded* controls in deriving necessary optimality conditions for PDE problems via Ekeland's principle. The main reason relates to the fact that it is not easy to create a *complete* metric space and to ensure the lower semicontinuity of the corresponding penalized functional needed for the application of Ekeland's principle in the framework of unbounded controls for PDE systems. First results for unbounded controls were obtained by Fattorini [431] and independently by Raymond and Zidani [1121] for parabolic systems. The approach developed by Mordukhovich and Raymond [944] and reproduced in Sect. 7.2 is an extension of the one from [1121] to the case of state-constrained hyperbolic systems with unbounded controls in the Neumann boundary conditions. The realization of this approach is strongly based on the regularity theory by Lasiecka and Triggiani [744, 745] and for hyperbolic Neumann boundary value problems and the corresponding developments presented in Subsect. 7.2.2.

7.5.12. Needle-Type Diffuse Control Perturbations. A significant part of the perturbation technique developed in Sect. 7.2 for the derivation of necessary optimality conditions in the state-constrained Neumann problem governed by the semilinear wave equation is the variational analysis of the approximating problems with *no* state constraints, which appear in the penalization procedure. Such an analysis is conducted in Subsect. 7.2.3 by developing a technique that can be viewed as a multidimensional (hyperbolic in this case) counterpart of the *increment method* involving *needle-type* control variations used in Subsect. 6.3.2 to prove the Pontryagin maximum principle for *free-endpoint* ODE control systems.

One can see that the PDE case under consideration is much more involved in comparison with its ODE counterpart. The needle-type control variations used in this analysis are known in the PDE control literature as *diffuse perturbations*, and also as "spike/multispike" and "patch" variations; cf. Fattorini [432], Li and Yong [789], and Raymond and Zidani [1121]. Such variations/perturbations of optimal controls were first used by Li and Yao [786, 787] to derive necessary optimality conditions for PDE control problems and then were developed in many publications; see particularly the aforementioned references [432, 789, 1121] and the bibliographies therein. Note that the justification of the required properties of such perturbations and their implementation in the proof of necessary optimality conditions are based on the *Lyapunov-Aumann convexity theorem*—a "hidden convexity" manifestation; see Raymond and Zidani [1121, Lemma 4.2], Mordukhovich and Raymond [944, Lemma 4.2 and Theorem 4.1], and the constructions reproduced in Lemma 7.19 and in the proof of Theorem 7.18 of Subsect. 7.2.3.

Observe that the usage of diffuse control perturbations allows us to derive the Pontryagin maximum principle for the Neumann boundary control problem for hyperbolic equations in the *pointwise form*, for both approximating and state-constrained systems under consideration in Sect. 7.2. This is similar to the case of ODE control systems studied in Chap. 6. At the same, the limiting procedures in the passing from necessary optimality conditions for unconstrained to constrained problems are significantly different in the ODE and hyperbolic PDE cases (cf. Subsects. 6.2.1 and 7.2.4); the latter is strongly based on the regularity theory for *weak solutions* to the Neumann-type developed in Subsect. 7.2.2.

7.5.13. Dirichlet Boundary Control of Hyperbolic Systems. Section 7.3 is devoted to *Dirichlet* boundary control of the *state-constrained* linear *wave equation* in n-dimensional spaces; actually the results obtained can be extended to more general *linear* hyperbolic equations with *strongly elliptic* operators replacing the classical Laplacian as in the "wave" case. We are not familiar with any results for the optimal control problem considered in Sect. 7.3 expect the recent paper by Mordukhovich and Raymond [943] on which Sect. 7.2 is based.

It has been well recognized that the Dirichlet boundary control case exhibits the *lowest regularity* properties in comparison with distributed and Neumann boundary controls for *all* the types of PDE systems; cf. also more discussions in Sect. 7.4 and the comments to it for parabolic equations. However, the *hyperbolic* case probably offers the lowest regularity in comparison with the other PDE types.

To the best of our knowledge, the *sharpest regularity* theory for Dirichlet hyperbolic *boundary value* problems was developed by Lasiecka, Lions and Triggiani [740]; see also the subsequent paper by Lasiecka and Sokolowski [741] and the book by Lasiecka and Triggiani [746] for some additional material and applications. We employ this theory in Sect. 7.3 for the purposes of our variational analysis. However, the regularity properties available in the PDE theory for the hyperbolic Dirichlet setting are *not* sufficient to develop variational methods and results for Dirichlet boundary control in hyperbolic systems similar to those for the Neumann hyperbolic case in Sect. 7.2 as well as for the Dirichlet parabolic case considered in Sect. 7.4. Nevertheless, something can be done by another method, and the results obtained in this direction are presented in Sect. 7.3.

The method developed in the afore-mentioned paper [943] and reproduced in Sect. 7.3 is based on the reduction of the state-constrained Dirichlet boundary control problem to a special problem of infinite-dimensional programming with *operator* and *geometric* constraints. The *lack of regularity* in the hyperbolic Dirichlet boundary control case, which *doesn't* allow us to conduct a perturbation/approximation analysis, is compensated by extra *full convexity* requirements on the integrands of the minimizing cost functional that ensure the applications of an appropriate version of the *Lagrange multiplier rule* in the obtained infinite-dimensional problem of mathematical programming (cf. particularly Alibert and Raymond [9]) due to the properties of *weak solutions* to the *adjoint* Dirichlet system derived in Subsect. 7.3.3. The assumptions made and the available regularity allow us also to establish the *existence* of optimal controls in the state-constrained Dirichlet boundary control problems under consideration.

7.5.14. Minimax Problems in Optimization and Control. Optimization problems with *minimax cost functions* play a significant role in many aspects of optimization and equilibrium theory, in analysis and synthesis of open-loop and closed-loop control systems, in various (static and dynamic) game-theoretical frameworks, as well as in numerous applications. It is worse repeating that, being *intrinsically nonsmooth*, minimax functions and associated minimax problems have always been among primary motivations for developing and implementations of nonsmooth variational analysis and generalized differentiation techniques. Among the enormous literature on various issues in minimax theory and its applications we refer the reader to Başar and Bernhard [87], Chernousko [238], Chikrii [239], Danskin [307], Demyanov and Malozemov [319], Freeman and Kokotović [474], Krasovskii and Subbotin [702], Kryazhimskii and Osipov [721], Kurzhanskii [731], Kurzhanskii and Vályi [732], Moiseev [885], von Neumann and Morgenstern [1000], Rockafellar and Wets [1165], Subbotin [1230], Simons [1213], and the bibliographies therein.

In Subsects. 5.3.2 and 5.5.19 we have discussed certain minimax issues from the viewpoint of multiobjective optimization and generalized differentiation, while the main objective of Sect. 7.4 is to study a *minimax control problem* for *parabolic* systems with the Dirichlet boundary conditions and pointwise state constraints. Note that some minimax problems for control systems governed by partial differential equations were considered by Ahmed and Xiang [5], Arada [21], Arada, Bergounioux and Raymond [22], Lenhart, Protopescu and Stojanović [762], Li and Yong [789], Mordukhovich [905, 918], Mordukhovich and Shvartsman [957], Mordukhovich and Zhang [978] among other publications.

7.5.15. Minimax Control of Constrained Parabolic Systems. The main motivation for considering the minimax control problem studied in Sect. 7.4 came from applications to automatic control of the soil water regime under *uncertainty*; see Mordukhovich [898, 905]. Dynamic processes in such systems are described by linearized parabolic equations of the filtration/diffusion theory with bounded controls acting in the *Dirichlet boundary conditions* and with distributed *uncertain perturbations* modeled on the right-hand side of the parabolic equation. Furthermore, the major technological requirements in these practical problems can be satisfied by imposing *pointwise state constraints* on controlled motions.

Since there is *no* probabilistic information on uncertain perturbations available in the engineering control problems modeled in [898, 905], *minimax* seems to be the most natural criterion for optimization and control design. In this way, some *open-loop* and *feedback* control problems were solved and practically implemented in [898, 905] for the case of *one-dimensional* linear parabolic equations by using certain *specific features* of the dynamic systems and constraints under consideration. More rigorous investigations are required for hard-constrained *multidimensional* parabolic systems, which motivates the study presented in Sect. 7.4 that is mainly based on the paper by Mordukhovich and Zhang [978].

The minimax problem (P) studied in Sect. 7.4 concerns a linear multidimensional parabolic equation described via a strongly uniformly elliptic operator with variable coefficients, which generates an *analytic semigroup* on a Hilbert space. Measurable controls act in the Dirichlet boundary conditions, while uncertain perturbations are additively distributed in the body of the equation. The *solution* notion for linear parabolic equations with the Dirichlet boundary conditions is understood in the *mild* sense; see Subsect. 7.4.1 and more comments below. The notion of *minimax optimality* is taken in the standard sense as a *saddle point* (worst perturbations and best controls) of the given integral functional depending on control, perturbation, and state variables. A significant feature of the minimax problem is the presence of *hard/pointwise constraints* of the magnitude type on controls, perturbations, and trajectories.

The linearity of the dynamic system and the structure of the imposed constraints allow us to *split* the minimax problem into two interrelated problems for *worst perturbations* and for *Dirichlet boundary controls*, which are studied separately by different methods. However, both of these methods involve certain *smooth approximation procedures*, which are strongly based on well-posedness properties of *mild solutions* to parabolic systems with *irregular/measurable* boundary conditions of the Dirichlet type.

7.5.16. Mild Solutions and Their Properties for Parabolic Systems with Dirichlet Boundary Conditions. Mild solutions to parabolic systems with irregular (merely measurable) Dirichlet boundary conditions were particularly studied by Balakrishnan [74], Lasiecka [739], Lasiecka and

Triggiani [742, 743], and Washburn [1324], where the reader can find existence and uniqueness results for such solutions with the basic operator estimates (7.71). The results of Subsect. 7.4.2 are mostly taken from Mordukhovich and Zhang [978], where they were employed to establish the existence of minimax solutions in Theorem 7.36 (on the base of the classical *von Neumann minimax theorem* in *appropriate* topologies used in Subsect. 7.4.2) and then to justify the convergence/well-posedness of the approximation procedures developed for deriving the necessary optimality and suboptimality conditions in Subsects. 7.4.3–7.4.5.

Observe a special situation occurring in the study of the minimax problem (P) and its splitting control/perturbation counterparts, where the state constraints are imposed in the *pointwise form* (7.62) involving generally *discontinuous* real-valued functions $y(x,t)$ of two variables, while the employed semigroup approach deals with *continuous* time-dependent mild solutions from Definition 6.26 that take values in functional spaces. Theorem 7.35 plays a *crucial role* to overcome this discrepancy by establishing the *a.e. pointwise* convergence in *values* of two-variable state solutions implied by the corresponding *weak* convergence of measurable functions in the *Dirichlet* boundary conditions.

7.5.17. Distributed Control of Constrained Parabolic Systems with Irregular/Nonsmooth Data. According to the splitting procedure, the worst perturbations in the original minimax problem happen to be optimal solutions to the parabolic *distributed control* problem with the fixed Dirichlet boundary conditions and pointwise *state constraints*. Necessary optimality conditions for such control problems can be found, e.g., in Casas [225], Bergounioux and Tröltzsch [104], Mackenroth [825], Fattorini [430, 432], Raymond and Zidani [1121], and Tröltzsch [1271]; see also the references therein. Note that for parabolic systems, distributed control problems are fairly close, in methods and results, to their boundary control counterparts with the *Neumann* (but not Dirichlet) boundary conditions.

However, the distributed control problem (P_1) studied in Subsects. 7.4.3 and 7.4.5 is substantially different from its standard versions due to the significant *data irregularity*. The issue is that the cost functional and state constraints in problem (P_1), which appeared via the splitting procedure from the original minimax problem, depend on the given mild solution $\bar{y}_2(x,t)$ to the Dirichlet boundary control problem (P_2) that usually exhibits a high *discontinuity* as a function of two variables; such problems are often called *nonsmooth* in the PDE literature.

To deal with such distributed control problems with irregular/nonsmooth data, we follow the paper by Mordukhovich and Zhang [978], which is mainly based on a certain *smooth approximation technique* developed in the theory of parabolic *variational inequalities*; see, e.g., Barbu [82], Friedman [479], He [554], Neittaaanmäki and Tiba [997], Tiba [1255], and the references therein. This technique employed in Subsect. 7.4.3 is different from that based on

Ekeland's variational principle while also allowing us to efficiently *penalize* the irregular state-constrained problem by a parametric family of *smooth unconstrained* problems, to establish the required *strong convergence* of approximating solutions, and then to derive *necessary optimality conditions* in the well-posed approximation problems.

The next step accomplished in Subsect. 7.4.5 is to justify a *limiting procedure* in the proof of necessary conditions for optimal solutions (worst perturbations) to the *state-constrained* distributed control problem (P_1) by passing to the limit from the necessary optimality conditions in the parametric approximation problems $(P_{1\varepsilon})$ with no state constraints. In our implementation of the limiting procedure we employ a refined *qualification condition* used in somewhat different settings by Bergounioux and Tiba [103] and by Bergounioux and Tröltzsch [104]. This condition (CQ1), which *doesn't* require that the feasible solution set in (P_1) is of *nonempty interior*, happens to be significantly different from standard infinite-dimensional counterparts of the classical Slater constraint qualification in convex programming. Based on (CQ1) and employing a delicate contraction result by Brézis and Strauss [177] as well as the classical Sobolev imbedding, we arrive in this way at necessary optimality conditions for the worst perturbations in the *integral form* of the Pontryagin maximum principle, which easily implies the *pointwise bang-bang* relations.

7.5.18. Dirichlet Boundary Control of Parabolic Systems with Pointwise State Constraints. The second problem (P_2) appearing in the minimax splitting procedure happens to be a state-constrained Dirichlet boundary control problem with *highly irregular* (L^∞) control functions acting in the Dirichlet boundary conditions. It has been well recognized that the *Dirichlet boundary control* case is the *most challenging* in optimal control theory for *parabolic* equations, since such conditions offer the *lowest regularity* properties of the parabolic dynamics. We refer the reader to Arada and Raymond [23, 24], Barbu [82], Fattorini [433], Fattorini and Murphy [435, 436], Lasiecka [739], Lasiecka and Triggiani [742, 743, 746], Mordukhovich and Zhang [978, 979], and Washburn [1324] for various necessary optimality conditions in parabolic systems with Dirichlet boundary controls.

Our study of the Dirichlet parabolic problems in Subsects. 7.4.4 and 7.4.5 mainly follow the developments by Mordukhovich and Zhang [978, 979] based on the *regularity/stability* properties of *mild solutions* given in Subsect. 7.4.2 and on the *approximation methods* from the theory of *parabolic variational inequations* that are largely similar in spirit to those presented in Subsects. 7.4.3 and 7.4.5 for the case of distributed controls/perturbations.

Observe that the previous results obtained in this direction (see, e.g., Barbu's book [82] and the references therein) imposed much stronger *smoothness* requirements on Dirichlet boundary controls assuming, in the best case, that

$$u \in W_p^{2-1/p,1-2/p}(\Sigma) \ \text{ with } \ p \geq 2 \, .$$

This was due to the fact that, instead of the mild solution theory as in Sect. 7.4, the afore-mentioned previous developments in this direction were mostly based on the classical *strong* solution theory for the parabolic Dirichlet boundary value problem dealing particularly with solutions of class $y \in W_p^{2,1}(Q)$ to the Dirichlet parabolic system (7.61); see the book Ladyzhenskaya, Solonnikov and Uralzeva [735] intensively used by Barbu [82] and other researchers.

The usage of mild solutions in our highly *nonsmooth/irregular* case allows us, first at all, to establish the existence of *minimax solutions* as in Theorem 7.36 and also the existence of *optimal controls* in a more general Dirichlet boundary control (not minimax) problem studied in Mordukhovich and Zhang [979]. Note that the latter theorem *removes* the linearity assumption on the integrand $g(x,t,\cdot)$ needed for the minimax existence theorem; see Remark 7.37. Furthermore, the properties of mild solutions given in Subsect. 7.4.2 make it possible to accomplish the *smooth approximation/penalization* procedure in Subsect. 7.4.4 and to establish the *strong* convergence of approximating optimal solutions with *no* restrictive smoothness assumptions as those imposed in Barbu [82].

Finally, Subsect. 7.4.5 justifies the limiting procedure to derive necessary optimality conditions for the state-constrained Dirichlet boundary control problem (P_2); see also Mordukhovich and Zhang [979] for similar results concerning a more general optimal control problem of this type. To proceed, we employ the constraint qualification condition (CQ2), which is in spirit of the corresponding constraint qualifications imposed by Bergounioux and Tiba [103] and by Bergounioux and Tröltzsch [104] in different settings; see Subsect. 7.5.18 for more discussions. The main result obtained in this way in Theorem 7.52 provides an integral-type counterpart of the Pontryagin maximum principle for the Dirichlet boundary control problem under consideration via the *adjoint* operator

$$\mathcal{L}^*\colon L^\infty(Q)^* \to L^\infty(\Sigma)^*$$

to the mild solution operator $\mathcal{L}$ defined in (7.72), where the dual space $L^\infty(Q)^*$ is identified with that of bounded additive functions (*measures*) on the domain Q. Somewhat different methods and results dealing with L^∞-controls acting in the Dirichlet boundary conditions of parabolic systems were developed by Arada and Raymond [23, 24], Fattorini [433], and Fattorini and Murphy [435, 436].

7.5.19. Feedback Synthesis and Minimax Design of Control Systems. As well known, problems of *feedback control*—when control functions depend on state variables—are among the most difficult in control theory and the most important for applications. There are various approaches to feedback control design, which have been mainly developed for control systems governed by *ordinary differential equations*. We are not going to discuss them in more detail just referring the reader to some recent developments employing certain constructions and techniques of modern variational analysis and generalized differentiation; see Bardi and Capuzzo Dolcetta [85], Cannarsa and Sinestrari [217], Clarke, Ledyaev, Sontag and Subbotin [263], Clarke, Ledyaev, Stern and Wolenski [264, 265], Clarke and Stern [269], Fleming and Soner [458], Frankowska [472], Freeman and Kokotović [474], Goebel [511], Rockafellar and Wolenski [1166, 1167], Subbotin [1230], Sontag [1220], Zelikin and Melnikov [1359], and the bibliographies therein. However, most of the results obtained in the afore-mentioned publications are largely theoretical, and their implementation to design feedback controls in more or less practical problems is always a subject of special considerations and additional investigations.

In many situations occurring in practical applications it happens that control systems are functioning in *uncertainty conditions*, where *no* (deterministic or stochastic) information is available for uncertain disturbances/perturbations but *only regions* of their possible deviations. The *minimax approach*, or the principle of *guaranteed result*, paves a natural route for the design/synthesis of feedback control systems in such uncertainty conditions. It has been well recognized that game-theoretic methods provide a general framework for the minimax control design; see Başar and Bernhard [87], Chernousko [238], Chikrii [239], Krasovskii and Subbotin [702], Kryazhimskii and Osipov [721], Kurzhanskii [731], Kurzhanskii and Vályi [732], Subbotin [1230], and their references among a great many publications in this direction. Observe again that the application of game-theoretical methods and results to the minimax design of particular control systems always requires an additional work that takes into account specific features of the problem in question.

As mentioned in the beginning of Sect. 7.4 (and additionally discussed in Subsect. 7.5.14), the original motivation for our study of minimax control came from some *practical* problems of *engineering design* of automatic reclamation systems for regulating the *groundwater* (or soil water) regime; see Mordukhovich [898, 905]. Adequate mathematical models were identified and described in [898, 905] as problems of *minimax synthesis* of state-constrained parabolic systems with distributed uncertain perturbations and with controls acting in the Dirichlet boundary conditions; see Remark 7.54 for a typical problem of this type. Since *no* methods and results have been available in general theory for such problems, we developed *special techniques* for their solution in the case of the dynamics described by the one-dimensional *heat equation*. The developed techniques involved several approximation procedures that particularly exploited, besides necessary and sufficient optimality conditions, some *monotonicity* properties of the one-dimensional parabolic dynamics and its asymptotics on the *infinite horizon*. It has been revealed furthermore that such systems exhibit a certain *turnpike behavior* (as described in the books by Carlson, Haurie and Leizarowitz [224], by Dyukalov [379], and by Zaslavski [1357]), which happened to be crucial for the efficient *feedback control design* conducted in [905].

Some *multidimensional* extensions of the methods and results initiated in [898, 905] were presented in Mordukhovich [918] and in Mordukhovich and Shvartsman [957], although the minimax design problem for the constrained parabolic systems under consideration has largely remained *open*.

8

Applications to Economics

The concluding chapter of this book is devoted to applications of modern techniques of variational analysis and generalized differentiation to competitive equilibrium models of *welfare economics* involving *nonconvex economies* with *infinite-dimensional* commodity spaces. Note that economic modeling has always been a challenging territory for applications of optimization theory, variational methods, and generalized differential constructions. In particular, convex models of welfare economics of the type considered in this chapter were among the most important *motivations* for the development of convex analysis in the beginning of the 1950s. Since that time such models have been an attractive area for applications of advanced variational and generalized differential techniques in convex and nonconvex settings.

Our main single tool in studying nonconvex models of welfare economics is the *extremal principle* of variational analysis, which allows us to establish new versions of the so-called *generalized/extended second welfare theorem* for weak Pareto, Pareto, and strong Pareto optimal allocations in nonconvex economies with *marginal/equilibrium prices* formalized via the *basic normal cone* and its *Fréchet-like approximations* developed in this book.

8.1 Models of Welfare Economics

In this section we describe models of welfare economics, in both classical and advanced frameworks, define the corresponding equilibrium and Pareto-type optimality concepts, and discuss the so-called *net demand qualification conditions* needed for the subsequent study of Pareto and weak Pareto (but not strong Pareto) optimal allocations. Let us start with informal (and then formal) descriptions of such models and mathematical techniques used for their studies and applications.

8.1.1 Basic Concepts and Model Description

The classical Walrasian equilibrium model of welfare economics and its various generalizations have long been recognized as an important part of the economic theory and applications. It has been well understood that the concept of *Pareto efficiency/optimality* and its variants play a crucial role for the study of equilibria and making the best decisions for competitive economies.

A classical approach to the study of Pareto optimality in economic models with smooth data consists of reducing it to conventional problems of mathematical programming and using first-order *necessary optimality conditions* that involve Lagrange multipliers. In this way important results were obtained at the late 1930s and in the 1940s when it has been shown that the *marginal rates of substitution* for consumption and production are *equal* to each other at any Pareto optimal allocation of resources; see the fundamental book by Samuelson [1188] with the discussions and references therein, and also further comments at the end of this chapter.

In the beginning of the 1950s, Arrow [26] and Debreu [309] made the next crucial step in the theory of welfare economics considering economic models with possibly nonsmooth but *convex* data. Based on the classical *separation theorems* for convex sets, they and their followers developed a nice theory that particularly contains *necessary and sufficient* conditions for Pareto optimal allocations and shows that each of such allocations leads to a *decentralized equilibrium* in convex economies. The key result of this theory is known as the classical *second fundamental theorem of welfare economics* stated that any Pareto optimal allocation can be *decentralized at price equilibria*, i.e., it can be sustained by a nonzero price vector at which each *consumer minimizes his/her expenditures* and each *firm maximizes its profit.* The full statement of this result is definitely *due to convexity*, which is crucial in the Arrow-Debreu model and its extensions based on convex analysis. Note also that the Arrow-Debreu general equilibrium theory in welfare economics has played an important motivating role in the development of *convex analysis* as a mathematical discipline with its subsequent numerous applications.

On the other hand, the *relevance of convexity* assumptions is often doubtful for many important applications, which had been recognized even before developing the Arrow-Debreu model; see, e.g., the afore-mentioned book by Samuelson [1188, pp. 231–232] stating that such assumptions are fulfilled "*only by accident...*" It is well known, in particular, that convexity requirements don't hold in the presence of *increasing returns to scale* in the production sector. A common approach to the study of nonconvex models is based on utilizing local *convex tangential approximations* and then employing the classical separation theorems for convex cones. Constructively it has been done by using the Clarke tangent cone, which is *automatically convex.* In this way, marginal prices are formalized in terms of the dual Clarke normal cone that, however, may be *too large* for satisfactory results in nonconvex models and often doesn't impose any restriction on marginal cost pricing; the reader can find many examples, discussions, and references in the paper by Khan [671]. The latter paper contains much more adequate extensions of the second welfare theorem to nonconvex economies with *finite-dimensional* commodity spaces, where marginal prices are formalized via our (nonconvex) basic normal cone. Khan's approach to derive such results are based on reducing, under appropriate constraint qualifications, Pareto optimal allocations to optimal solutions for problems of *nondifferentiable programming* and then applying necessary conditions in nonsmooth optimization established by Mordukhovich [892]. This approach doesn't require the use of convex separation and/or related results of convex analysis.

The primary goal in what follows is to derive comprehensive results on the extended second welfare theorem(s) in nonconvex models of welfare economics, in the general framework of *infinite-dimensional* commodity spaces, based on the *extremal principle*, which is the main single tool of the variational analysis developed in this book. As discussed in Chap. 2, the extremal principle can be viewed as a *variational counterpart* of the (local) separation in nonconvex settings. On the other hand, it provides necessary conditions for extremal points of nonconvex sets that cover, as will be shown below, the case of Pareto-like optimal allocations. Thus using the extremal principle, we actually *unify* both approaches discussed above, which are based on either the reduction of Pareto optimality to mathematical programming or the application of separation theorems for convex sets.

The machinery of the extremal principle developed in Chap. 2 allows us to derive extended versions of the second welfare theorem for nonconvex economies in both *approximate/fuzzy* and *exact/limiting* forms under mild net demand qualification conditions needed in the case of *Pareto* and *weak Pareto* optimal allocations. In this way we obtain efficient conditions ensuring the marginal price *positivity* when commodity spaces are *ordered*. The results obtained bring new information even in the case of *convex economies*, since we *don't impose* either the classical *interiority* condition or the widely implemented *properness* condition by Mas-Colell [855]. Moreover, in contrast to the vast majority of publications on convex economies with ordered commodity spaces, our approach *doesn't require any lattice structure* of commodity spaces in either finite-dimensional or infinite-dimensional settings.

The usage of the extremal principle makes it possible to derive really surprising results on the generalized second welfare theorem in both approximate and exact forms for *strong Pareto* optimal allocations in nonconvex economies with *ordered* commodity spaces. Indeed, in this case we *don't need qualification conditions* of the above type for the validity of our extended versions of the second welfare theorem. This conclusion seems to be new even for classical models involving convex economies with finite-dimensional commodities.

As mentioned, marginal prices in our nonconvex extensions of the second welfare theorem are formalized via the basic normals in the exact version and via their Fréchet-like approximations in the approximate version presented below for all the three types of Pareto optimal allocations. Then the *variational* *descriptions* of approximate normals established in Subsect. 1.1.4 allow us to derive a *convex-type/decentralized equilibrium* interpretation of the extended second welfare theorems in nonconvex economies involving *nonlinear prices*; see Sect. 7.2 for more details and discussions.

Next let us formally describe the basic model of welfare economics under consideration in this chapter. Although the given description and subsequent properties discussed in this section hold in the general framework of *linear topological* spaces equipped with a locally convex Hausdorff topology, the main results involving generalized normals require the *Asplund* space structure; see also Sect. 8.4 for their counterparts in other classes of *Banach* spaces.

Let E be a normed *commodity space* of the economy $\mathcal{E}$ that involves $n \in I\!N$ consumers with *consumption sets* $C_i \subset E,\ i = 1, \ldots, n$, and $m \in I\!N$ firms with *production sets* $S_j \subset E,\ j = 1, \ldots, m$. Each consumer has a *preference set* $P_i(x)$ that consists of elements in C_i preferred to x_i by this consumer at the consumption plan/bundle $x = (x_1, \ldots, x_n) \in C_1 \times \cdots \times C_n$. This is a valuable generalization (with a useful economic interpretation) of ordering relations given by preferences $\prec_i$ as in Sect. 5.3 and, in particular, by utility functions as in classical models of welfare economics. We have by definition that $x_i \notin P_i(x)$ for all $i = 1, \ldots, n$ and always assume that $P_i(x) \neq \emptyset$ for some $i \in \{1, \ldots, n\}$, i.e., at least one consumer is *nonsatiated.* For convenience we put $\mathrm{cl}\, P_i(x) := \{x_i\}$ if $P_i(x) = \emptyset$.

Now we define *feasible allocations* of the economy $\mathcal{E}$ imposing *market constraints* formalized via a given nonempty subset $W \subset E$ of the commodity space; we label W as the *net demand constraint set* in $\mathcal{E}$.

Definition 8.1 (feasible allocations). *Let* $x = (x_i) := (x_1, \ldots, x_n)$, *and let* $y = (y_j) := (y_1, \ldots, y_m)$. *We say that the pair* $(x, y) \in \prod_{i=1}^n C_i \times \prod_{j=1}^m S_j$ *is a* FEASIBLE ALLOCATION *of* $\mathcal{E}$ *if*

$$w := \sum_{i=1}^n x_i - \sum_{j=1}^m y_j \in W \ . \tag{8.1}$$

Introducing the net demand constraint set allows us to unify some conventional situations in economic models and to give a useful economic insight in the general framework. Indeed, in the classical case the set W consists of one element $\{\omega\}$, where ω is an *aggregate endowment* of scarce resources. Then constraint (8.1) reduces to the "markets clear" condition. Another conventional framework appears in (8.1) when the commodity space E is ordered by a closed positive cone E_+ and we put $W := \omega - E_+$, which corresponds to the "implicit free disposal" of commodities. Generally constraint (8.1) describes a natural situation that may particularly happen when the initial aggregate endowment is not exactly known due to, e.g., *incomplete information.* In the latter general case the set W reflects some *uncertainty* in the economic model under consideration.

In what follows we pay the main attention to the three Pareto-type notions of optimality for feasible allocations in the economic model $\mathcal{E}$: *weak Pareto* optimality, *Pareto* optimality, and *strong Pareto* optimality. While the first two notions will be considered in parallel under similar but somewhat different net demand constraint qualifications, the strong Pareto optimality plays a specific role in the case of ordered commodity spaces, where such constraint qualifications are not needed for the validity of the extended versions of the second welfare theorem established below, even for the classical framework of convex economies with finite-dimensional commodities.

Definition 8.2 (Pareto-type optimal allocations). *Let $(\bar{x}, \bar{y})$ be a feasible allocation of the economy $\mathcal{E}$ with the property*

$$\bar{x}_i \in \operatorname{cl} P_i(\bar{x}) \quad \textit{for all} \quad i = 1, \ldots, n \, .$$

We say that:

(i) *$(\bar{x}, \bar{y})$ is a local* WEAK PARETO OPTIMAL ALLOCATION *of $\mathcal{E}$ if there is a neighborhood O of $(\bar{x}, \bar{y})$ such that for every feasible allocation $(x, y) \in O$ one has $x_i \notin P_i(\bar{x})$ for some $i \in \{1, \ldots, n\}$.*

(ii) *$(\bar{x}, \bar{y})$ is a local* PARETO OPTIMAL ALLOCATION *of $\mathcal{E}$ if there is a neighborhood O of $(\bar{x}, \bar{y})$ such that for every feasible allocation $(x, y) \in O$ either $x_i \notin \operatorname{cl} P_i(\bar{x})$ for some $i \in \{1, \ldots, n\}$ or $x_i \notin P_i(\bar{x})$ for all $i = 1, \ldots, n$.*

(iii) *$(\bar{x}, \bar{y})$ is a local* STRONG PARETO OPTIMAL ALLOCATION *of $\mathcal{E}$ is there is a neighborhood O of $(\bar{x}, \bar{y})$ such that for every feasible allocation $(x, y) \in O$ with $(x, y) \neq (\bar{x}, \bar{y})$ one has $x_i \notin \operatorname{cl} P_i(\bar{x})$ for some $i \in \{1, \ldots, n\}$.*

When the preference sets $P_i(x)$ are defined via preference relations $\prec_i$ as in Sect. 5.3 (in particular, by utility functions), the above notions of Pareto and weak Pareto optimal allocations reduce to the corresponding concepts of *multiobjective optimization* under the special type of constraints (8.1). The notion of strong Pareto optimality is non-conventional in multiobjective optimization, even in the classical framework, while playing an important role in economic modeling.

To study Pareto and weak Pareto optimal allocations, we introduce and discuss in the next subsection appropriate *net demand qualification conditions*, which allow us to reduce these types of Pareto optimality to local extremal points of some closed sets. Such qualifications are not needed in the case of strong Pareto optimal allocations, which will be shown in Subsect. 8.3.2.

8.1.2 Net Demand Qualification Conditions for Pareto and Weak Pareto Optimal Allocations

We begin this subsection with two parallel definitions of qualification conditions for the economy $\mathcal{E}$ that play a *crucial role* in the subsequent results on the extended second welfare theorem in the case of *Pareto* and *weak Pareto* optimal allocations, respectively. Obviously the condition in (ii) implies the one in (i), but not vice versa.

Definition 8.3 (net demand qualification conditions). *Let* $(\bar{x}, \bar{y})$ *be a feasible allocation of the economy* $\mathcal{E}$*, and let*

$$\bar{\omega} := \sum_{i=1}^{n} \bar{x}_i - \sum_{j=1}^{m} \bar{y}_j \ . \tag{8.2}$$

Given $\varepsilon > 0$*, we consider the set*

$$\Delta_\varepsilon := \sum_{i=1}^{n} \operatorname{cl} P_i(\bar{x}) \cap (\bar{x}_i + \varepsilon I\!B) - \sum_{j=1}^{m} \operatorname{cl} S_j \cap (\bar{y}_j + \varepsilon I\!B) - \operatorname{cl} W \cap (\bar{\omega} + \varepsilon I\!B)$$

and say that:

(i) *The* NET DEMAND QUALIFICATION (NDQ) CONDITION *holds at* $(\bar{x}, \bar{y})$ *if there are* $\varepsilon > 0$*, a sequence* $\{e_k\} \subset X$ *with* $e_k \to 0$ *as* $k \to \infty$*, and a consumer index* $i_0 \in \{1, \ldots, n\}$ *such that*

$$\Delta_\varepsilon + e_k \subset P_{i_0}(\bar{x}) + \sum_{i \neq i_0} \operatorname{cl} P_i(\bar{x}) - \sum_{j=1}^{m} S_j - W \tag{8.3}$$

for all $k \in I\!N$ *sufficiently large.*

(ii) *The* NET DEMAND WEAK QUALIFICATION (NDWQ) CONDITION *holds at* $(\bar{x}, \bar{y})$ *if there are* $\varepsilon > 0$ *and a sequence* $e_k \to 0$ *as* $k \to \infty$ *such that*

$$\Delta_\varepsilon + e_k \subset \sum_{i=1}^{n} P_i(\bar{x}) - \sum_{j=1}^{m} S_j - W \tag{8.4}$$

for all $k \in I\!N$ *sufficiently large.*

It is easy to observe that both NDQ and NDWQ conditions automatically hold if *either one* among preference, or production, or net demand constraint sets is *epi-Lipschitzian* around the corresponding point in the sense of Definition 1.24(ii). We know from Proposition 1.25 that for epi-Lipschitzian property of a *convex* set $\Omega \subset X$ is equivalent to its nonempty interior $\operatorname{int} \Omega \neq \emptyset$. Thus the above qualification conditions may be viewed as far-going extensions of the classical *nonempty interiority* condition well developed for convex models of welfare economics.

The next proposition contains verifiable conditions that ensure the fulfillment the NDQ and NDWQ properties and significantly extend the epi-Lipschitzian requirements mentioned above. Note to this end that the epi-Lipschitzian property of Ω around $\bar{x}$ implies this property of the *closure* $\operatorname{cl} \Omega$ around this point, but not vice versa. It is also worth mentioning that the *summation* of sets as in (8.6) and (8.7) below (especially for a large number of sets) tends to improve properties related to nonempty interiors, and that the epi-Lipschitzian property of sets falls into this category.

Proposition 8.4 (sufficient conditions for NDQ and NDWQ properties). *Let $(\bar{x}, \bar{y})$ be a feasible allocation of the economy $\mathcal{E}$. The following assertions hold:*

(i) *Assume that the sets S_j, $j = 1, \ldots, m$, and W are closed near the points $\bar{y}_j$ and $\bar{w}$ from* (8.2)*, respectively. Then the NDQ condition is satisfied at $(\bar{x}, \bar{y})$ if there exist a number $\varepsilon > 0$, an index $i \in \{1, \ldots, n\}$, and a desirability sequence $\{e_{ik}\} \subset E$, $e_{ik} \to 0$ as $k \to \infty$, such that*

$$\operatorname{cl} P_i(\bar{x}) \cap (\bar{x}_i + \varepsilon I\!B) + e_{ik} \subset P_i(\bar{x}) \ \ \textit{for all large} \ \ k \in I\!N \ . \tag{8.5}$$

Moreover, the NDWQ condition is satisfied at $(\bar{x}, \bar{y})$ if a desirability sequence $\{v_{ik}\}$ exists for each $i \in \{1, \ldots, n\}$ with some $\varepsilon > 0$ in (8.5)*.*

(ii) *Assume that $\bar{x}_i \in \operatorname{cl} P_i(\bar{x})$ for all $i = 1, \ldots, n$. Then the NDWQ condition is satisfied at $(\bar{x}, \bar{y})$ if the set*

$$\Delta := \sum_{i=1}^{n} P_i(\bar{x}) - \sum_{j=1}^{m} S_j - W \tag{8.6}$$

is epi-Lipschitzian around $0 \in \operatorname{cl} \Delta$. It happens when either one among the sets $P_i(\bar{x})$ for $i = 1, \ldots, n$, S_j for $j = 1, \ldots, m$, and W or some of their partial combinations in (8.6) *is epi-Lipschitzian around the corresponding point.*

(iii) *Assume that $n > 1$. The NDQ condition is satisfied at $(\bar{x}, \bar{y})$ if there is a consumer $i_0 \in \{1, \ldots, n\}$ such that $P_{i_0}(\bar{x}) \neq \emptyset$ and that the set*

$$\Sigma := \sum_{i \neq i_0} \operatorname{cl} P_i(\bar{x}) \tag{8.7}$$

is epi-Lipschitzian around the point $\sum_{i \neq i_0} \bar{x}_i$. It happens when either one among the sets $\operatorname{cl} P_i(\bar{x})$ for $i \in \{1, \ldots, n\} \setminus \{i_0\}$ or some of their partial sums is epi-Lipschitzian around the corresponding point.

Proof. Both statements in (i) easily follow from the definitions and the assumptions made.

Let us prove (ii). Due to the structure of (8.4), it is sufficient to consider the case when the aggregate set Δ in (8.6) is epi-Lipschitzian around the origin. Using Definition 1.24(ii) of the epi-Lipschitzian property, we find $v \in E$ and $\gamma > 0$ satisfying

$$\Delta \cap (\gamma I\!B) + t(v + \gamma I\!B) \subset \Delta \ \text{ for all } \ t \in (0, \gamma) \ . \tag{8.8}$$

Picking an arbitrary sequence $t_k \downarrow 0$ as $k \to \infty$, put

$$e_k := t_k v \ \text{ as } \ k \in I\!N, \quad \varepsilon := \frac{\gamma}{n + m + 2} \tag{8.9}$$

and show that the NDWQ condition (8.6) holds with e_k and ε from (8.9). To proceed, we take any $z_\varepsilon \in \Delta_\varepsilon$ and conclude by the construction of $\bar{w}$ and Δ_ε

in Definition 8.3 that $z_\varepsilon \in (n+m+1)\varepsilon I\!B$. Due to the structure of Δ in (8.6), find a sequence of elements $z_k \in \Delta$ converging to z_ε as $k \to \infty$. Obviously

$$z_k \in \big(n+m+2\big)\varepsilon I\!B = \gamma\, I\!B \;\text{ for large }\; k \in I\!N \tag{8.10}$$

by the choice of ε in (8.9). We can also select z_k so that

$$z_\varepsilon - z_k \in (t_k \gamma) B \;\text{ for large }\; k \in I\!N\;. \tag{8.11}$$

Now combining (8.8)–(8.11), we get

$$z_\varepsilon + e_k = z_k + t_k v + (z_\varepsilon - z_k) \in \Delta \cap (\gamma\, I\!B) + t_k(v + \gamma\, I\!B) \subset \Delta\;,$$

which surely implies (8.4).

It remains to justify (iii) considering the case when the set Σ in (8.7) is epi-Lipschitzian around the reference point. Using this property, we find $v \in E$ and $\gamma > 0$ such that

$$\sum_{i \neq i_0} \operatorname{cl} P_i(\bar{x}) \bigcap \Big(\sum_{i \neq i_0} \bar{x}_i + \gamma\, I\!B\Big) + t\big(v + \gamma\, I\!B\big) \subset \sum_{i \neq i_0} \operatorname{cl} P_i(\bar{x})\;. \tag{8.12}$$

Now select v_k and ε as in (8.9) and proceed similarly to the above proof of (ii). Take $z_\varepsilon \in \Delta_\varepsilon$ with

$$z_\varepsilon = \sum_{i=1}^{n} x_i - \sum_{j=1}^{m} y_j - w, \quad x_i \in \operatorname{cl} P_i(\bar{x}), \quad y_j \in \operatorname{cl} S_j, \quad w \in \operatorname{cl} W$$

and approximate x_{i_0}, y_j, and w by sequences of elements from the corresponding sets $P_{i_0}(\bar{x})$, S_j, and W. In contrast to the proof of (ii), we *don't* approximate x_i for $i \neq i_0$. Proceedings in this way, we deduce the net demand qualification condition (8.3) from the epi-Lipschitzian property (8.8) by arguments similar to those used in justifying assertion (ii). This gives (iii) and completes the proof of the proposition. △

It is important to observe that we *don't need* to impose *any assumption on the preference and production sets* for the fulfillment of both qualification conditions from Definition 8.3 if the *net demand constraint set* W is *epi-Lipschitzian* around $\bar{w}$. This easily follows from Proposition 8.4(ii). It happens, in particular, when E is *ordered* and $W := \omega - E_+$ with $\operatorname{int} E_+ \neq \emptyset$ for the closed positive cone $E_+ \subset E$. The latter covers the conventional case of the so-called "free disposal Pareto optimum" defined by Cornet [288].

8.2 Second Welfare Theorem for Nonconvex Economies

This section contains necessary conditions for Pareto and weak Pareto optimal allocations of the nonconvex economy $\mathcal{E}$ with an *Asplund* commodity space E *without* imposing *any ordering* structure on commodities. Invoking the *extremal principle* from Chap. 2, we derive necessary conditions for these two types of Pareto optimal allocations in the *approximate* and *exact* forms via the prenormal/Fréchet normal cone and the basic normal cone, respectively. The results obtained are appropriate extensions of the *generalized second welfare theorem* to nonconvex economies involving *the same* (common) marginal/equilibrium price for *all* the preference and production sets. We discuss various consequences and interpretations of the main results including rather surprising ones that ensure convex-type *decentralized equilibria* for *nonconvex* models by using *nonlinear prices.*

8.2.1 Approximate Versions of Second Welfare Theorem

This subsection is devoted to *approximate/fuzzy* versions of the extended second welfare theorem, which are formulated and proved in a parallel way for both Pareto and weak Pareto optimal allocations.

Theorem 8.5 (approximate form of the extended second welfare theorem with Asplund commodity spaces). *Let the pair* $(\bar{x}, \bar{y})$ *be a local Pareto* (*resp. weak Pareto*) *optimal allocation of the economy* $\mathcal{E}$ *with an Asplund commodity space* E*. Assume that the net demand qualification condition* (*resp. net demand weak qualification condition*) *is satisfied at* $(\bar{x}, \bar{y})$*. Then for every* $\varepsilon > 0$ *there exist a suboptimal triple*

$$(x, y, w) \in \prod_{i=1}^{n} \operatorname{cl} P_i(\bar{x}) \times \prod_{j=1}^{m} \operatorname{cl} S_j \times \operatorname{cl} W$$

with w *defined in* (8.1) *and a common marginal price* $p^* \in E^* \setminus \{0\}$ *satisfying*

$$-p^* \in \widehat{N}\big(x_i; \operatorname{cl} P_i(\bar{x})\big) + \varepsilon I\!B^* \tag{8.13}$$

with $x_i \in \bar{x}_i + \frac{\varepsilon}{2} I\!B$ *for all* $i = 1, \ldots, n$,

$$p^* \in \widehat{N}(y_j; \operatorname{cl} S_j) + \varepsilon I\!B^* \tag{8.14}$$

with $y_j \in \bar{y}_j + \frac{\varepsilon}{2} I\!B$ *for all* $j = 1, \ldots, m$,

$$p^* \in \widehat{N}(w; \operatorname{cl} W) + \varepsilon I\!B^* \tag{8.15}$$

with $w \in \bar{w} + \frac{\varepsilon}{2} I\!B$, *and*

$$\frac{1-\varepsilon}{2\sqrt{n+m+1}} \le \|p^*\| \le \frac{1+\varepsilon}{2\sqrt{n+m+1}}\ , \tag{8.16}$$

where $\bar{w}$ *is defined in* (8.2).

Proof. Let $(\bar{x}, \bar{y})$ be a feasible allocation of the economy $\mathcal{E}$. We suppose that this allocation is locally optimal in the sense of either Pareto or weak Pareto from Definition 8.2 and proceed in a parallel way for both cases. In fact, the only difference between these cases is in applying the corresponding net demand qualification condition from Definition 8.3, which are actually designed to reduce the Pareto-type optimality under consideration to local extremal points of a special system of sets. Consider the product space $X := E^{n+m+1}$ equipped with the norm

$$\|(v_1, \ldots, v_{n+m+1})\|_X := \left[\|v_1\|^2 + \ldots + \|v_{n+m+1}\|^2\right]^{1/2}.$$

Since E is Asplund, the product space X is Asplund as well. Taking now a number $\varepsilon > 0$ for which the NDQ condition (resp. the NDWQ condition) holds with the corresponding sequence $\{e_k\}$ in (8.3) and (8.4), define the two closed sets in X as follows

$$\begin{aligned} \Omega_1 := \prod_{i=1}^{n} \left[\operatorname{cl} P_i(\bar{x}) \cap (\bar{x}_i + \varepsilon I\!B)\right] \times \prod_{j=1}^{m} \left[\operatorname{cl} S_j \cap (\bar{y}_j + \varepsilon I\!B)\right] \\ \times \left[\operatorname{cl} W \cap (\bar{w} + \varepsilon I\!B)\right], \end{aligned} \tag{8.17}$$

$$\Omega_2 := \left\{(x, y, w) \in X \,\middle|\, \sum_{i=1}^{n} x_i - \sum_{j=1}^{m} y_j - w = 0\right\}. \tag{8.18}$$

Check that $(\bar{x}, \bar{y}, \bar{w})$ is a local extremal point of the set system $\{\Omega_1, \Omega_2\}$ built in (8.17) and (8.18). Indeed, it follows directly from (8.1) and (8.2) that $(\bar{x}, \bar{y}, \bar{w}) \in \Omega_1 \cap \Omega_2$. To justify the local extremality of $(\bar{x}, \bar{y}, \bar{w})$, it is sufficient to find a neighborhood U of this point and a sequence $\{a_k\} \subset X$ such that $a_k \to 0$ as $k \to \infty$ and that

$$(\Omega_1 - a_k) \cap \Omega_2 \cap U = \emptyset \quad \text{for all large} \ \ k \in I\!N \tag{8.19}$$

under the corresponding qualification condition from Definition 8.3. To proceed, we take a neighborhood $O \in E^{n+m}$ of the Pareto (weak Pareto) optimal allocation $(\bar{x}, \bar{y})$ and a sequence $\{e_k\} \subset E$ converging to zero for which one has (8.3) and (8.4), respectively. In both cases we put

$$U := O \times I\!R \subset X \quad \text{and} \quad a_k := (0, \ldots, 0, e_k) \in X$$

and show that (8.19) holds for the same $k \in I\!N$ as in (8.3) and (8.4). Assuming the contrary, we find $z_k \in \Omega_1$ with $z_k - a_k \in \Omega_2$. By the structure of (8.17) and (8.18) and by the construction of a_k and U this implies the existence of (x_k, y_k, w_k) with $(x_k, y_k) \in O$,

$$x_{ik} \in \mathrm{cl}\, P_i(\bar{x}) \cap (\bar{x}_i + \varepsilon I\!B), \quad i = 1, \ldots, n\ ,$$

$$y_{jk} \in \mathrm{cl}\, S_j \cap (\bar{y}_j + \varepsilon I\!B), \quad j = 1, \ldots, m\ ,$$

$$w_k \in \mathrm{cl}\, W \cap (\bar{w} + \varepsilon I\!B), \quad \text{and}$$

$$\sum_{i=1}^{n} x_{ik} - \sum_{j=1}^{m} y_{jk} - w_k + e_k = 0\ .$$

The latter means, by the construction of the set Δ_ε in Definition 8.3, that $0 \in \Delta_\varepsilon + e_k$. Then applying the NDQ condition, we get that the origin belongs to the set on right-hand side of (8.3), while the NDWQ condition ensures that the right-hand side set in (8.4) contains the origin. This definitely contradicts the (local) Pareto optimality of $(\bar{x}, \bar{y})$ in the first case and the weak Pareto optimality of $(\bar{x}, \bar{y})$ in the second one. Thus we arrive at (8.19), which signifies that $(\bar{x}, \bar{y}, \bar{w})$ is a *local extremal point* for the system of closed sets $\{\Omega_1, \Omega_2\}$ under consideration in the Asplund space X.

Now we can apply to this system the approximate version of the *extremal principle* from Theorem 2.20. According to extremal principle in Asplund spaces, for every $\varepsilon > 0$ there are $z := (x_1, \ldots, x_n, y_1, \ldots, y_m, w) \in \Omega_1,\ \tilde{z} \in \Omega_2$, and dual elements (Fréchet normals)

$$z^* \in \widehat{N}(z; \Omega_1), \quad \text{and} \quad \tilde{z}^* \in \widehat{N}(\tilde{z}; \Omega_2) \tag{8.20}$$

satisfying the relations

$$\|x_i - \bar{x}_i\| \le \frac{\varepsilon}{2}, \quad \|y_j - \bar{y}_j\| \le \frac{\varepsilon}{2}, \quad \|w - \bar{w}\| \le \frac{\varepsilon}{2} \tag{8.21}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$ with

$$\frac{1-\varepsilon}{2} \le \|\widetilde{z}^*\| \le \frac{1+\varepsilon}{2} \quad \text{and} \quad \|z^* + \widetilde{z}^*\| \le \frac{\varepsilon}{2}\ . \tag{8.22}$$

Observe that the set Ω_2 in (8.18) is a linear subspace *separated* in all the variables (x_i, y_j, w). Thus $\widehat{N}(\widetilde{z}; \Omega_2)$ is a subspace orthogonal to Ω_2 and

$$\widetilde{z}^* = (p^*, \ldots, p^*, -p^*, \ldots, -p^*)$$

in (8.20), where the minus terms start with the $(n+1)$st position. It follows from (8.22) and the norm definition on X that

$$\frac{1-\varepsilon}{2} \le \sqrt{n+m+1}\, \|p^*\| \le \frac{1+\varepsilon}{2}\ . \tag{8.23}$$

Then we conclude from (8.18) and the last estimate in (8.22) that

$$-\widetilde{z}^* = (-p^*, \ldots, -p^*, p^*, \ldots, p^*) \in \widehat{N}(z; \Omega_1) + \varepsilon I\!B^*\ . \tag{8.24}$$

Now use the Fréchet normal product formula from Proposition 1.2 applied to the set Ω_1 and observe by (8.21) that all the components (x_i, y_j, w) of the point z in (8.24) belong to the *interiors* of the corresponding neighborhoods in (8.17); hence these neighborhoods can be ignored in the calculation of $\widehat{N}(z;\Omega_1)$. Combining finally (8.21), (8.23), and (8.24), we arrive at relations (8.13)–(8.16) and complete the proof of the theorem. △

Observe that, in contrast to the approximate extremal principle of Theorem 2.20 for the general extremal system of closed sets, Theorem 8.5 ensures the existence of a *common* dual element $p^* \in E^* \setminus \{0\}$ for *all* the sets involved in (8.13)–(8.16), instead of generally different elements x_i^* in the extremal principle. This common element, which can be interpreted as an approximate *marginal/equilibrium price* for all the preference and production sets near Pareto and weak Pareto optimal allocations, corresponds to the *very essence* of the classical second welfare theorem ensuring the identity between marginal rates of substitution for consumers and firms. Note that such a specification of the extremal principle in models of welfare economics is due to the specific structure of sets (8.17) and (8.18) in the extremal system, especially due to the *separated* variables in (8.18).

Let us present an *equilibrium* interpretation of the obtained approximate version of the second welfare theorem in the case of *convex economies*; more precisely, for economies with convex preference and production sets. In this case relations (8.13) and (8.14) reduce, respectively, to *global minimization* (*maximization*) of the *perturbed* consumer expenditures (firm profits) over the corresponding preference (production) sets, justifying therefore a *decentralized price equilibrium* in convex models with *no interiority* assumptions on the convex preferences and production sets in question under small perturbations.

Corollary 8.6 (perturbed equilibrium in convex economies). *In addition to the assumptions of Theorem* 8.5*, suppose that all the preferences and production sets* $P_i(\bar{x})$*,* $i = 1, \ldots, n$*, and* S_j*,* $j = 1, \ldots, m$*, are convex. Then for every* $\varepsilon > 0$ *there exist a suboptimal allocation* (x, y) *with the feasible linear combination* w *from* (8.1) *satisfying*

$$\begin{aligned}(x, y, w) \in & \prod_{i=1}^{n} \Big[\mathrm{cl}\, P_i(\bar{x}) \bigcap \Big(\bar{x}_i + \frac{\varepsilon}{2}I\!B\Big)\Big] \\ & \times \prod_{j=1}^{m} \Big[\mathrm{cl}\, S_j \bigcap \Big(\bar{y}_j + \frac{\varepsilon}{2}I\!B\Big)\Big] \times \Big[\mathrm{cl}\, W \bigcap \Big(\bar{w} + \frac{\varepsilon}{2}I\!B\Big)\Big]\end{aligned} \tag{8.25}$$

and an equilibrium price $p^* \in E^* \setminus \{0\}$ *such that one has* (8.15), (8.16)*, and*

$$\langle p^*, u_i - x_i\rangle \ge -\varepsilon\|u_i - x_i\| \ \ \textit{for all} \ \ u_i \in \mathrm{cl}\, P_i(\bar{x}), \quad i = 1, \ldots, n, \tag{8.26}$$

$$\langle p^*, v_j - y_j\rangle \le \varepsilon\|v_j - y_j\| \ \ \textit{for all} \ \ v_j \in \mathrm{cl}\, S_j, \quad j = 1, \ldots, m\,. \tag{8.27}$$

Proof. It follows directly from relations (8.13) and (8.14) of Theorem 8.5 and the representation of ε-normals to convex sets given in Proposition 1.3. $\triangle$

The next theorem establishes rather surprising results about the *equilibrium meaning* of marginal prices from the approximate second welfare theorem for general *nonconvex economies* based on *smooth variational descriptions* of Fréchet-like normals. Indeed, in this way we get a *perturbed decentralized equilibrium* of the convex type as in the preceding corollary, but with *no* convexity assumptions, replacing the linear price p^* in (8.26) and (8.27) by some *nonlinear prices* that are differentiable in certain senses with the derivatives (i.e., rates of change) *arbitrarily close* to p^* at suboptimal allocations.

Theorem 8.7 (decentralized equilibrium in nonconvex economies via nonlinear prices). *Given any* $\varepsilon > 0$, *the following assertions hold:*

(i) *Let all the assumptions of Theorem* 8.5 *be fulfilled. Then there exist a suboptimal triple* (x, y, w) *satisfying* (8.25) *with* w *defined in* (8.1), *a marginal price* $p^* \in E^* \setminus \{0\}$ *satisfying relations* (8.15) *and* (8.16), *as well as real-valued functions* g_i, $i = 1, \ldots, n$, *and* h_j, $j = 1, \ldots, m+1$, *on the commodity space* E *that are Fréchet differentiable at* x_i, y_j, *and* w, *respectively, with*

$$\begin{cases} \|\nabla g_i(x_i) - p^*\| \le \varepsilon, \quad i = 1, \ldots, n, \\[2ex] \|\nabla h_j(y_j) - p^*\| \le \varepsilon, \quad j = 1, \ldots, m, \\[2ex] \|\nabla h_{m+1}(w) - p^*\| \le \varepsilon \end{cases} \tag{8.28}$$

and such that each g_i, $i = 1, \ldots, n$, *achieves its global minimum over* $\operatorname{cl} P_i(\bar{x})$ *at* x_i, *each* h_j, $j = 1, \ldots, m$, *achieves its global maximum over* $\operatorname{cl} S_j$ *at* y_j, *and* h_{j+1} *achieves its global maximum over* $\operatorname{cl} W$ *at* w.

(ii) *In addition to the assumptions of Theorem* 8.5, *suppose that* E *admits an* $\mathcal{S}$*-smooth bump function from the classes* $\mathcal{S}$ *considered in Theorem* 1.30. *Then there exist a suboptimal triple* (x, y, w) *satisfying* (8.25), *a marginal price* $p^* \in E^* \setminus \{0\}$ *satisfying* (8.15) *and* (8.16), *as well as* $\mathcal{S}$*-smooth functions* g_i *and* h_j *on* E *satisfying* (8.28) *such that each* g_i *achieves its global minimum over* $\operatorname{cl} P_i(\bar{x})$ *uniquely at* x_i, *that each* h_j, $j = 1, \ldots, m$, *achieves its global maximum over* $\operatorname{cl} S_j$ *uniquely at* y_j, *and* h_{j+1} *achieves its global maximum over* $\operatorname{cl} W$ *uniquely at* w. *Moreover, we can choose* g_i *and* h_j *to be convex and concave, respectively, if* E *admits a Fréchet smooth renorm.*

Proof. Take p^* satisfying the conclusions of Theorem 8.5 and then, by (8.13)–(8.15), find p_i^* and p_j^* with

$$\begin{cases} -p_i^* \in \widehat{N}\big(x_i; \operatorname{cl} P_i(\bar{x})\big), \;\; \|p_i^* - p^*\| \le \varepsilon, \quad i = 1, \ldots, n, \\[2ex] p_j^* \in \widehat{N}(y_j; \operatorname{cl} S_j), \;\; \|p_j^* - p^*\| \le \varepsilon, \quad j = 1, \ldots, m, \\[2ex] p_{j+1}^* \in \widehat{N}(w; \operatorname{cl} W), \;\; \|p_{j+1}^* - p^*\| \le \varepsilon . \end{cases} \tag{8.29}$$

Applying now the smooth variational descriptions of Fréchet normals from Theorem 1.30, we arrive at all the conclusions of the theorem. △

One can see that the functions g_i and h_j play a role of *nonlinear prices* discussed before the formulation of Theorem 8.7, which ensure the decentralized convex-type equilibrium in nonconvex economies under small perturbations.

8.2.2 Exact Versions of Second Welfare Theorem

Next we derive *pointbased* necessary optimality conditions for *Pareto* and *weak Pareto* optimal allocations of the economy $\mathcal{E}$ expressed via the *basic normal cone* to the preference, production, and net demand constraint sets computed *exactly at* optimal allocations. The results obtained are given in the *exact form* of the extended second welfare theorem, where the *same* marginal price is associated, at the optimal allocation under consideration, with *all* the economy sets listed above, providing thus a *marginal price equilibrium.*

Our proof of the exact second welfare theorem is based on passing to the limit in the relations of the approximate second welfare theorem established in the preceding subsection. To furnish the limiting procedure, we need to impose some *sequential normal compactness* conditions, as always in this book. However, the present economic model is *different* from all the previous settings, in particular, from the one in Theorem 2.22 for the exact extremal principle. The specific feature of the model $\mathcal{E}$ under consideration is that, instead of imposing the SNC condition on *all but one* sets involved, we require this property only for *one* among the preference, production, and net demand constraint sets. Such an essential improvement of the exact extremal principal in the economic framework $\mathcal{E}$ happens to be possible mostly due to the *separated structure* of the set (8.18) involved in the extremal system.

Theorem 8.8 (exact form of the extended second welfare theorem with Asplund commodity spaces). *Let* $(\bar{x}, \bar{y})$ *be a local Pareto* (*resp. weak Pareto*) *optimal allocation of the economy* $\mathcal{E}$ *satisfying the corresponding assumptions of Theorem* 8.5 *with* $\bar{w}$ *defined in* (8.2). *Suppose in addition that one of the sets*

$$\operatorname{cl} P_i(\bar{x}), \quad i = 1, \ldots, n; \quad \operatorname{cl} S_j, \quad j = 1, \ldots, m; \quad \operatorname{cl} W$$

is sequentially normally compact at $\bar{x}_i$, $\bar{y}_j$, *and* $\bar{w}$, *respectively. Then there is a nonzero price* $p^* \in E^*$ *satisfying*

$$-p^* \in N\big(\bar{x}_i; \operatorname{cl} P_i(\bar{x})\big), \quad i = 1, \ldots, n, \tag{8.30}$$

$$p^* \in N(\bar{y}_j; \operatorname{cl} S_j), \quad j = 1, \ldots, m, \tag{8.31}$$

$$p^* \in N(\bar{w}; \operatorname{cl} W)\ . \tag{8.32}$$

Proof. We prove this theorem by passing to the limit in the relations of Theorem 8.5. Pick an arbitrary sequence $\varepsilon_k \downarrow 0$ as $k \to \infty$ and, according to the latter result, find sequences $(x_k, y_k, w_k, p_k^*) \in E \times E \times E \times E^*$ satisfying

$$x_{ik} \in \text{cl}\, P_i(\bar{x}) \cap (\bar{x}_i + \varepsilon_k I\!B), \quad i = 1, \ldots, n \,,$$

$$y_{jk} \in \text{cl}\, S_j \cap (\bar{y}_j + \varepsilon_k I\!B), \quad j = 1, \ldots, m \,,$$

$$w_k = \sum_{i=1}^{n} x_{ik} - \sum_{j=1}^{m} y_{jk} \in \text{cl}\, W \cap (\bar{w} + \varepsilon_k I\!B) \,,$$

and the dual relations (8.13)–(8.16) with $\varepsilon = \varepsilon_k$ for each $k \in I\!N$. Obviously $(x_k, y_k, w_k) \to (\bar{x}, \bar{y}, \bar{w})$ as $k \to \infty$. Since E is Asplund and the prices p_k^* are uniformly bounded by (8.16), there is $p^* \in E^*$ such that the sequence $\{p_k^*\}$ converges to p^* in the weak* topology of E^*. Now passing to the limit in (8.13)–(8.15) as $k \to \infty$ and remembering the construction of the basic normal cone, we arrive at all the relations (8.30)–(8.32).

It remains to prove that $p^* \neq 0$ if *one* of the sets cl $P_i(\bar{x})$, cl S_j, and cl W is SNC at the corresponding point. On the contrary, let $p^* = 0$ and assume for definiteness that the set cl W is SNC at $\bar{w}$. Then by (8.15) there is a sequence of $e_k^* \in E^*$ such that

$$p_k^* - \varepsilon_k e_k^* \in \widehat{N}(w_k; \text{cl}\, W) \;\text{ with }\; \|e_k^*\| = 1 \;\text{ for all }\; k \in I\!N \,. \tag{8.33}$$

Obviously $p_k^* - \varepsilon_k e_k^* \xrightarrow{w^*} 0$ as $k \to \infty$. Then by Definition 1.20 of SNC sets, we conclude from (8.33) that

$$\|p_k^* - \varepsilon_k e_k^*\| \to 0 \;\text{ and hence }\; \|p_k^*\| \to 0 \;\text{ as }\; k \to \infty \,.$$

The latter clearly contradicts the left-hand side inequality in (8.16) for p_k^*. Thus we have $p^* \neq 0$, which completes the proof of the theorem. △

Let us discuss some useful consequences of Theorem 8.8. First we consider a special case of $\mathcal{E}$, where the net demand constraint set W admits the *conic representation*

$$W = \omega + \Gamma \;\text{ with some }\; \omega \in \text{cl}\, W \,. \tag{8.34}$$

If E is an ordered commodity space with the closed positive cone E_+ in it (see the next section), then representation (8.34) with $\Gamma := -E_+$ corresponds to the so-called *implicit free disposal of commodities.* We consider a more general case of Γ being an arbitrary *convex cone* in E, with no ordering structure, and show that (8.32) implies in this case the following *complementary slackness condition*, which economically can be interpreted as *zero value of excess demand* at the marginal price.

Corollary 8.9 (excess demand condition). *In addition to the assumptions of Theorem* 8.8*, suppose that* W *is given as* (8.34)*, where* Γ *is a nonempty convex subcone of the commodity space* E*. Then there is a nonzero price* $p^* \in E^*$ *satisfying* (8.30)*,* (8.31)*, and*

$$\Big\langle p^*, \sum_{i=1}^{n} \bar{x}_i - \sum_{j=1}^{m} \bar{y}_j - \omega \Big\rangle = 0 \,. \tag{8.35}$$

Proof. To justify (8.35), observe that

$$\langle p^*, \bar{w} - \omega \rangle \ge \langle p^*, w - \omega \rangle \ \text{ for all } \ w \in \operatorname{cl} W \tag{8.36}$$

due to (8.32), (8.34), and the normal cone representation for convex sets. Hence $\langle p^*, \bar{w} - \omega \rangle \ge 0$. On the other hand, taking

$$2(\bar{w} - \omega) \in W - \omega = \Gamma$$

due to the conic structure of Γ, we get by (8.36) that $\langle p^*, \bar{w} - \omega \rangle \le 0$, which justifies (8.35) and completes the proof of the corollary. △

In the case of economies with *convex* preference and production sets, relations (8.30) and (8.31) of Theorem 8.8 reduce to the classical *consumer expenditure minimization* and *firm profit maximization* conditions of the second fundamental theorem of welfare economics. We are able, however, *essentially improve* the *nonempty interiority* condition imposed on convex sets in the economy $\mathcal{E}$. Indeed, we know from Theorem 1.21 that the SNC property required in our extension of the second welfare theorem is *equivalent*, for convex sets with nonempty *relative* interiors, to the *finite codimensionality* of such sets. Moreover, convex sets in Asplund spaces may be SNC even having empty relative interiors; see Example 3.6 and also the discussion in Remark 1.27. Thus the following consequence of Theorem 8.8 provides a far-going improvement of the classical second welfare theorem for convex economies with Asplund commodity spaces in both cases of Pareto and weak Pareto optimality.

Corollary 8.10 (improved second welfare theorem for convex economies). *In addition to the assumptions of Theorem* 8.8*, suppose that all the preference and production sets*

$$P_i(\bar{x}), \ \ i = 1, \ldots, n, \quad \textit{and} \quad S_j, \ \ j = 1, \ldots, m, \quad \textit{are convex.}$$

Then there is a nonzero price $p^* \in E^*$ *satisfying* (8.32) *and such that*

$$\bar{x}_i \ \ \textit{minimizes} \ \ \langle p^*, x_i \rangle \ \ \textit{over} \ \ x_i \in \operatorname{cl} P_i(\bar{x}_i) \ \ \textit{whenever} \ \ i = 1, \ldots, n \ ,$$

$$\bar{y}_j \ \ \textit{maximizes} \ \ \langle p^*, y_j \rangle \ \ \textit{over} \ \ y_j \in \operatorname{cl} S_j \ \ \textit{whenever} \ \ j = 1, \ldots, m \ .$$

Proof. This follows directly from (8.30) and (8.31) due to the normal cone representation for convex sets. △

Remark 8.11 (nonconvex equilibria). As shown in the above corollary, the assumptions in Theorem 8.8 justify the *decentralized price equilibrium*, at Pareto and weak Pareto optimal allocations of *convex* models of welfare economics. In contrast to the approximate/suboptimal setting of Theorem 8.7, we may not generally provide a decentralized equilibrium interpretation of the pointbased relations (8.30) and (8.31) via nonlinear prices. Nevertheless, the results obtained allow us to treat the *marginal price equilibrium* given by the first-order necessary optimality conditions of Theorem 8.8 as the *limiting case* of the *convex-type decentralized equilibrium* in *nonconvex* models, which can be achieved by using *nonlinear prices.*

8.3 Nonconvex Economies with Ordered Commodity Spaces

In this section we study a special case of the welfare economic model $\mathcal{E}$ when the commodity space E is an *ordered Banach space.* Our goals are:

(i) Find efficient conditions ensuring the *marginal price positivity* in the framework of the (exact) extension of the second welfare theorem given by Theorem 8.8 for Pareto and weak Pareto optimal allocations of (generally nonconvex) economies $\mathcal{E}$.

(ii) Derive new versions of both approximate and exact second welfare theorems for *strong Pareto* optimal allocations in the case of ordered commodity spaces, without imposing net demand qualification conditions.

We accomplish these goals in the following two subsections. Observe that we *don't* impose a *lattice structure* on the commodity space in question.

8.3.1 Positive Marginal Prices

Let E be an *ordered Banach space* with the *closed positive cone*

$$E_+ := \left\{ e \in E \middle| \, e \geq 0 \right\} ,$$

where the (standard) partial ordering relation is denoted by $\geq$, in accordance with the conventional notation in the economic literature. The corresponding *dual positive cone* E^*_+, which is the closed positive cone of the ordered space E^*, admits the representation

$$E^*_+ := \left\{ e^* \in E^* \middle| \, e^* \geq 0 \right\} = \left\{ e^* \in E^* \middle| \, \langle e^*, e \rangle \geq 0 \text{ whenever } e \in E_+ \right\} ,$$

where the order on E^* is induced by the given one $\geq$ on E.

Our conditions for the marginal price positivity are based on the following lemma of independent interest, which ensures the *positivity* of the basic normal cone to closed subsets of ordered Banach spaces.

Lemma 8.12 (positivity of basic normals in ordered spaces). *Let E be an ordered Banach space, and let Ω be a nonempty closed subset of E satisfying the condition*

$$\Omega - E_+ \subset \Omega \,. \tag{8.37}$$

Then one has the inclusion

$$N(\bar{e};\Omega) \subset E_+^* \quad \textit{whenever} \quad \bar{e} \in \Omega \,. \tag{8.38}$$

Proof. Take $e^* \in N(\bar{e};\Omega)$, where Ω satisfies (8.37). By the definition of basic normals there are sequences

$$\varepsilon_k \downarrow 0, \quad e_k \xrightarrow{\Omega} \bar{e}, \quad \text{and} \quad e_k^* \xrightarrow{w^*} e^* \;\text{ as }\; k \to \infty$$

with $e_k^* \in \widehat{N}_{\varepsilon_k}(e_k;\Omega)$ for all $k \in I\!N$. Due to (8.37) and the *monotonicity* property of ε-normals

$$\widehat{N}_\varepsilon(e;\Omega_1) \subset \widehat{N}_\varepsilon(e;\Omega_2) \;\text{ whenever }\; e \in \Omega_2 \subset \Omega_1 \;\text{ and }\; \varepsilon \ge 0 \,,$$

we have $e_k^* \in \widehat{N}_{\varepsilon_k}(e_k;\Omega - E_+)$ for all $k \in I\!N$. Fix $k \in I\!N$ and take an arbitrary number $\gamma > 0$. Using the definition of ε-normals, find $\eta_k > 0$ such that

$$\langle e_k^*, e - e_k \rangle \le (\varepsilon_k + \gamma)\,\|e - e_k\| \;\text{ if }\; e \in (e_k + \eta_k I\!B) \cap (\Omega - E_+) \,. \tag{8.39}$$

It is easy to see that

$$e_k - \eta_k u \in (e_k + \eta_k I\!B) \cap (\Omega - E_+) \;\text{ for any }\; u \in E_+ \cap I\!B \,.$$

Substituting $e := e_k - \eta_k u$ into (8.39), one has

$$\langle e_k^*, -u \rangle \le (\varepsilon_k + \gamma)\,\|u\| \le \varepsilon_k + \gamma \;\text{ whenever }\; u \in E_+ \cap I\!B \;\text{ and }\; k \in I\!N \,.$$

Passing to the limit in the latter inequality and taking into account that $e_k^* \xrightarrow{w^*} e^*$ as $k \to \infty$, we arrive at

$$\langle e^*, -u \rangle \le \gamma \;\text{ for all }\; u \in E_+ \cap I\!B \,,$$

which implies that $e^* \in E_+^*$, since $\gamma > 0$ was chosen arbitrary. This gives (8.38) and completes the proof of the lemma. △

Note that (8.37) is related to *free disposal* type conditions in economic models. The next theorem contains assumptions in this line imposed on *either* preference, *or* production, *or* net demand constraint sets that ensure the price positivity $p^* \in E_+^* \setminus \{0\}$ in our extended second welfare theorem in the framework of ordered Asplund commodity spaces.

Theorem 8.13 (positive prices for Pareto and weak Pareto optimal allocations). *Let* $(\bar{x}, \bar{y})$ *be a local Pareto (resp. weak Pareto) optimal allocation of the economy* $\mathcal{E}$. *In addition to the corresponding assumptions of Theorem 8.8, suppose that* E *is an ordered space and that one of the following conditions holds:*

(a) *There exists* $i \in \{1, \ldots, n\}$ *such that the* i*-th consumer satisfies the* DESIRABILITY CONDITION *at* $\bar{x}$, *i.e.,*

$$\operatorname{cl} P_i(\bar{x}) + E_+ \subset \operatorname{cl} P_i(\bar{x}) .$$

(b) *There exists* $j \in \{1, \ldots, m\}$ *such that the* j*-th firm satisfies the* FREE DISPOSAL *condition, i.e.,*

$$\operatorname{cl} S_j - E_+ \subset \operatorname{cl} S_j .$$

(c) *The net demand constraint set* W *exhibits the* IMPLICIT FREE DISPOSAL *of commodities, i.e.,*

$$\operatorname{cl} W - E_+ \subset \operatorname{cl} W .$$

Then there is a positive marginal price $p^* \in E_+^* \setminus \{0\}$ *satisfying relations* (8.30)–(8.32) *via the basic normal cone.*

Proof. The marginal price positivity $p^* \in E_+^*$ in cases (b) and (c) follows directly from Lemma 8.12 due to relations (8.31) and (8.32) of Theorem 8.8. Case (a) reduces to the same lemma by (8.30) and the property

$$N(\bar{e}; \Omega) = -N(-\bar{e}; -\Omega) \quad \text{for every} \quad \Omega \subset E \quad \text{and} \quad \bar{e} \in \Omega$$

valid in any Banach space, which can be checked by definition. △

Observe that each of the conditions in (a)–(c) implies the *epi-Lipschitzian* property of the corresponding sets $\operatorname{cl} P_i(\bar{x})$, $\operatorname{cl} S_j$, and $\operatorname{cl} W$ provided that $\operatorname{int} E_+ \neq \emptyset$. Due to the discussions above, the latter *nonempty interior* requirement on the *positive cone* of E ensures also the fulfillment of the qualification and normal compactness conditions of Theorem 8.8 and thus the existence of a *positive marginal price* $p^* \in E_+^* \setminus \{0\}$ in Theorem 8.13.

8.3.2 Enhanced Results for Strong Pareto Optimality

One can see that the net demand qualification conditions (NDQ and NDWQ) play a major role in the proofs of the above extensions of the second welfare theorem for Pareto and weak Pareto optimal allocations. Indeed, they allow us to reduce the corresponding notions of Pareto optimality to extremal points of set systems and then to apply the extremal principle. In the case of *ordered* commodity spaces E these qualification conditions are related to the *nonempty interior* requirement $\operatorname{int} E_+ \neq \emptyset$ on the positive cone of E. Of course, all the above extensions of the second welfare theorem hold true for strong Pareto optimal allocations, which connote a more restrictive notion of Pareto optimality.

In this subsection we show that the net demand qualification conditions are *not needed* at all for *strong Pareto* optimal allocations of convex and nonconvex economies in ordered commodity spaces, where $\operatorname{int} E_+ = \emptyset$ in many settings important for both the theory and applications. It happens that the strong Pareto optimality requirement allows us to reduce the corresponding optimal allocations to *local extremal points* of set systems with *no* qualification conditions imposed. Thus we can employ again the *extremal principle*, which is our main tool of variational analysis.

Recall that the closed positive cone E_+ is *generating* for E if this space can be represented as $E = E_+ - E_+$. The class of Banach spaces ordered by their generation positive cones is sufficiently large including, in particular, all *Banach lattices* (or normed complete *Riesz spaces*) whose generating positive cones typically have *empty interiors*.

The next result establishes several versions of the second welfare theorem for strong Pareto optimal allocations of (generally nonconvex) economies with ordered Asplund commodity spaces. It contains both approximate and exact forms of the second welfare theorem including marginal price positivity under desirability/free disposal type assumptions. Note that the generating requirement is imposed on the positive cone in the first two statements of the theorem, while the third one provides alternative assumptions on the economy ensuring the same conclusions in more general ordered spaces.

Theorem 8.14 (second welfare theorems for strong Pareto optimal allocations). *Let $(\bar{x}, \bar{y})$ be a local strong Pareto optimal allocation of the economy $\mathcal{E}$ with an ordered Asplund commodity space E, and let the sets S_j, W be locally closed near $\bar{y}_j$ and $\bar{w}$, respectively. Then the following hold:*

(i) *Assume that the closed positive cone E_+ is generating and that either the economy exhibits the implicit free disposal of commodities*

$$W - E_+ \subset W \,, \tag{8.40}$$

or the free disposal production condition

$$S_j - E_+ \subset S_j \quad \textit{for some} \quad j \in \{1, \ldots, m\} \tag{8.41}$$

is fulfilled, or $n > 1$ and there is a consumer $i_0 \in \{1, \ \ldots, n\}$ such that $P_{i_0}(\bar{x}) \neq \emptyset$ and one has the desirability condition

$$\operatorname{cl} P_i(\bar{x}) + E_+ \subset \operatorname{cl} P_i(\bar{x}) \quad \textit{for some} \quad i \in \{1, \ldots, n\} \setminus \{i_0\} \,. \tag{8.42}$$

Then for every $\varepsilon > 0$ there exist a suboptimal triple

$$(x, y, w) \in \prod_{i=1}^{n} \Big[\mathrm{cl}\, P_i(\bar{x}) \bigcap \Big(\bar{x}_i + \frac{\varepsilon}{2} I\!B\Big)\Big]$$

$$\times \prod_{j=1}^{m} \Big[S_j \bigcap \Big(\bar{y}_j + \frac{\varepsilon}{2} I\!B\Big)\Big] \times \Big[W \bigcap \Big(\bar{w} + \frac{\varepsilon}{2} I\!B\Big)\Big]$$

with the aggregate commodity w *defined in* (8.1) *and a common marginal price* $p^* \in E^*$ *satisfying relations* (8.13)–(8.16).

(ii) *If in addition to* (i) *one of the sets*

$$\mathrm{cl}\, P_i(\bar{x}), \quad i = 1, \ldots, n, \qquad S_j, \quad j = 1, \ldots, m, \qquad W$$

is SNC at the corresponding point, then there is a positive marginal price $p^* \in E^* \setminus \{0\}$ *satisfying the pointbased relations* (8.30)–(8.32).

(iii) *All the conclusions in* (i) *and* (ii) *hold true if, instead of the assumption that* E_+ *is a generating cone, we suppose that* $E_+ \neq \{0\}$ *and at least two sets among* W, S_j *for* $j = 1, \ldots, m$, *and* $P_i(\bar{x})$ *for* $i = 1, \ldots, n$ *satisfy the corresponding conditions in* (8.40)–(8.42).

Proof. Consider the system of two sets $\{\Omega_1, \Omega_2\}$ defined in (8.17) and (8.18), where the closure operation for S_j and W in (8.17) is omitted, since these sets are locally closed around the points of interest. Taking a *strong* Pareto local optimum $(\bar{x}, \bar{y})$ of $\mathcal{E}$, we show that $(\bar{x}, \bar{y}, \bar{w}) \in \Omega_1 \cap \Omega_2$ is a *local extremal point* of $\{\Omega_1, \Omega_2\}$ if either the assumptions in (i) or those in (iii) hold. Thus these assumptions replace the corresponding net demand qualification conditions in the proof of Theorem 8.5 for Pareto and weak Pareto optimal allocations.

First we consider case (i) when the positive cone E_+ is *generating* and *one* of the sets W, S_j, and $P_i(\bar{x})$ satisfies the corresponding condition in (8.40)–(8.42). For definiteness assume that (8.40) holds; the other two cases are treated similarly.

It is easy to observe that $\bar{w}$ is a *boundary point* of W; otherwise one has a contradiction with Pareto optimality of $(\bar{x}, \bar{y})$ under the standing assumption on $P_i(\bar{x}) \neq \emptyset$ for some $i \in \{1, \ldots, n\}$. Thus we find a sequence $e_k \to 0$ in E satisfying $\bar{w} + e_k \notin W$ for all $k \in I\!N$. Due to the classical *Krein-Šmulian theorem* (see, e.g., the book by Abramovich and Aliprantis [1] for the proof, discussions, and references), in any Banach space E ordered by a closed generating cone there exists a constant $M > 0$ such that for each $e \in E$ there are positive vectors

$$u, v \in E_+ \text{ with } e = u - v \text{ and } \max\big\{\|u\|, \|v\|\big\} \le M\|e\| .$$

This allows us to find sequences $u_k \stackrel{E_+}{\to} 0$ and $v_k \stackrel{E_+}{\to} 0$ satisfying $e_k = u_k - v_k$. Since $v_k \in E_+$ and $W - E_+ \subset W$, we get

$$\bar{w} + u_k \notin W \text{ with } u_k \stackrel{E_+}{\to} 0 \text{ as } k \to \infty . \tag{8.43}$$

Now take a neighborhood $O \subset E^{n+m}$ from the definition of the local strong Pareto optimal allocation $(\bar{x}, \bar{y})$ and show that the extremality condition (8.19) in the proof of Theorem 8.5 holds for all $k \in I\!N$ along the sequence of $a_k := (0, \ldots, 0, u_k) \in E^{n+m+1}$ and the neighborhood $U := O \times E$. This will justify the local extremality of $(\bar{x}, \bar{y}, \bar{w})$ for the system $\{\Omega_1, \Omega_2\}$ under consideration.

Supposing that (8.19) doesn't hold for some $k \in I\!N$, find $(x_k, y_k, w_k) \in \Omega_1$ such that $(x_k, y_k) \in O$ and $(x_k, y_k, w_k - u_k) \in \Omega_2$. By $u_k \in E_+$ and the implicit free disposal assumption (8.40), the latter implies that

$$\sum_{i=1}^{n} x_{ik} - \sum_{j=1}^{m} y_{jk} = w_k - u_k \in W - E_+ \subset W \tag{8.44}$$

for the components of (x_k, y_k). This means that (x_k, y_k) is a *feasible allocation* of the economy $\mathcal{E}$ belonging to the prescribed neighborhood of $(\bar{x}, \bar{y})$. Since $(\bar{x}, \bar{y})$ is a *strong Pareto* optimal allocation of $\mathcal{E}$, we get $(x_k, y_k) = (\bar{x}, \bar{y})$ for all large $k \in I\!N$. Hence one has

$$\begin{aligned}\bar{w} + u_k &= \sum_{i=1}^{n} \bar{x}_i - \sum_{j=1}^{m} \bar{y}_j + u_k = \sum_{i=1}^{n} x_{ik} - \sum_{j=1}^{m} y_{jk} + u_k \\ &= (w_k - u_k) + u_k = w_k \in W ,\end{aligned}$$

which contradicts (8.43) and thus justifies the local extremality of $(\bar{x}, \bar{y}, \bar{w})$ for $\{\Omega_1, \Omega_2\}$ in case (i).

Let us next show that the extremality condition (8.19) also holds in case (iii) of the theorem when the positive cone E_+ may *not* be generating. Suppose for definiteness that the implicit free disposal condition (8.40) is fulfilled and that *one* of the production set (say S_1) satisfies the free disposal condition in (8.41). Choose a sequence $u_k \overset{E_+}{\to} 0$ with $u_k \neq 0$ for all $k \in I\!N$, which is always possible due to $E_+ \neq \{0\}$. Take again $a_k := (0, \ldots, 0, u_k) \in X$ and check (8.19) along this sequence. Assuming the contrary and repeating the arguments as above, we find $(x_k, y_k, w_k) \in \Omega_2 \cap U$ satisfying (8.44). The latter implies that $(x_k, y_k) = (\bar{x}, \bar{y})$ for all large $k \in I\!N$, since $(\bar{x}, \bar{y})$ is a local *strong Pareto* optimal allocation of $\mathcal{E}$. It follows from (8.44) in this case that

$$\sum_{i=1}^{n} x_{ik} - (y_{1k} - u_k) - \sum_{j=2}^{m} y_{jk} = w_k \in W \tag{8.45}$$

for all $k \in I\!N$ sufficiently large. By the free disposal condition (8.41) for $j = 1$ we have $y_{1k} - u_k \in S_1$, and hence (8.45) ensures that $(x_k, y_k - (u_k, 0, \ldots, 0))$ is a feasible allocation of $\mathcal{E}$ belonging to the prescribed neighborhood of the strong Pareto local optimum $(\bar{x}, \bar{y})$. The latter implies that

$$y_{1k} - u_k = \bar{y}_1 - u_k = \bar{y}_1 ,$$

i.e., $u_k = 0$ for all large $k \in N$. This contradiction justifies the local extremality of $(\bar{x}, \bar{y}, \bar{w})$ for the system $\{\Omega_1, \Omega_2\}$ under the assumptions in (iii).

Applying the extremal principle of Theorem 2.20 to this system of sets, we arrive at the conclusions of the approximate second welfare theorem listed in Theorem 8.5, but now in the case of strong Pareto optimal allocations under the assumptions in either (i) or (iii) with no imposing the SNC property. If finally the SNC assumptions from (ii) are additionally imposed, we get the exact relationships (8.30)–(8.32) of the extended second welfare theorem by passing to the limits as in the proof of Theorem 8.8. The price positivity under (8.40)–(8.42) follows from Lemma 8.12 as in the proof of Theorem 8.13. This completes the proof of this theorem. △

Since the consequences of Theorems 8.5 and 8.8 established in Sect. 8.2, including equilibrium interpretations, don't depend on the net demand qualification conditions, they hold true for strong Pareto optimal allocations in the framework of Theorem 8.14.

Remark 8.15 (modified notion and results for strong Pareto optimal allocations). It has been recently observed by Glenn Malcolm (personal communication) that the results on the extended second welfare theorem obtained in Subsect. 8.3.2 for strong Pareto optimal allocations hold true with *no change* in their formulations for a *modified version* of strong Pareto optimality, which is probably more attractive for economic applications. The only difference between the new modified version of (local) strong Pareto optimal allocations and that given in Definition 8.2(iii) is as follows: *instead of the condition* $x_i \notin \mathrm{cl}\, P_i(\bar{x})$ *for some* $i \in \{1, \ldots, n\}$ *along every feasible allocation with* $(x, y) \neq (\bar{x}, \bar{y})$, *we now require (locally) the fulfillment of this condition merely for those* (x, y) *with* $x \neq \bar{x}$.

The latter modification allows us to involve into consideration feasible allocations with different production plans and associated endowments, which seems to be of substantial economic importance.

The reader can check that the above proof of Theorem 8.14 holds, with just small changes needed, to establish both approximate and limiting versions of the extended second welfare theorem for *modified* strong Pareto optimal allocations under *exactly the the same* assumptions as in assertions (i)—(iii) of this theorem, with *no net demand qualification conditions.*

Indeed, consider the set

$$\Sigma := W + \sum_{j=1}^{m} S_j \cap (\bar{y}_j + \nu I\!\!B)\ ,$$

where $\nu > 0$ is sufficiently small. We can easily check that the commodity

$$\bar{w} + \sum_{j=1}^{m} \bar{y}_j$$

gives a *boundary point* of the set Σ. Then we proceed similarly to the proof of assertion (i) in Theorem 8.14 by using the Krein-Šmulian theorem and replacing condition (8.43) by

$$\bar{w} + \sum_{j=1}^{m} \bar{y}_j \notin \Sigma \quad \text{with} \quad u_k \overset{E_+}{\to} 0 \quad \text{as} \quad k \to \infty \; .$$

This allows us to show that the triple $(\bar{x}, \bar{y}, \bar{w})$ is a local *extremal point* of the set system $\{\Omega_1, \Omega_2\}$ as in Theorem 8.14 based on the modified definition of strong Pareto optimal allocations. Similar arguments are applied to justify counterparts of assertions (ii) and (iii) of Theorem 8.14 in the modified case.

8.4 Abstract Versions and Further Extensions

The last section of the chapter contains some results and discussions on economics modeling in more general frameworks in comparison with the basic settings studied above. First we present abstract (pre)normal counterparts of the extended second welfare theorems for the economy $\mathcal{E}$ described in Sect. 8.1 *without* imposing the *Asplund structure* of the commodity space. The final subsection concerns models of welfare economics with *public goods* as well as some further extensions including models with public environment and with direct distribution.

8.4.1 Abstract Versions of Second Welfare Theorem

The model $\mathcal{E}$ of welfare economics considered above makes sense in any linear topological space as mentioned in Subsect. 8.1.1. At the same time the results obtained in Sects. 8.2–8.3 on the extended second welfare theorems are formulated and proved in terms of Fréchet and basic normals in economies with Asplund commodity spaces. Analyzing the proofs of the results given in Sects. 8.2–8.3, we observe the two major points that require the usage of either Fréchet-like constrictions, or the Asplund structure of commodities, or both these properties:

(i) Applying the *extremal principle* of Theorem 2.20 formulated via Fréchet-like normals to general closed sets, we don't have an opportunity to avoid the Asplund property of the space in question due to the characterization of Asplund spaces from this theorem.

(ii) The results of Theorems 8.13 and 8.14 involving *positive prices* are based on Lemma 8.12, which holds in arbitrary Banach spaces but seems to significantly employ some specific features of Fréchet and basic normals. Furthermore, the *decentralized equilibrium descriptions* of the extended second welfare theorem in nonconvex economies via *nonlinear prices* given in Theorem 8.7 and the related discussions presented in Remark 8.11 definitely require the Freéchet-like structure of generalized normals ensuring their *variational* representations. As already mentioned, the special geometric properties of the space in question listed in assertion (ii) of Theorem 8.7 imply the Asplund property of commodity spaces.

It *doesn't* seem possible to get similar variational descriptions for generalized normals of non-Fréchet type, which are *crucial* for the above decentralized interpretations of the marginal price equilibrium. On the other hand, the main results of group (i) based on the extremal principle have their counterparts in non-Asplund spaces via the corresponding *prenormal* and *normal* structures discussed in Sect. 2.5, where some *abstract/axiomatic* versions of the extremal principle have been derived. The goal of this subsection is to clarify what additional assumptions on the prenormal and normal structures are needed for the validity of abstract analogs of the approximate and exact second welfare theorems established in Sects. 8.2 and 8.3.

Let us start with the *approximate* version of the second welfare theorem for Pareto, weak Pareto, and strong Pareto optimal allocations. Note that the net demand qualification conditions of Definition 8.3 and the free disposal/desirability type conditions listed in Theorem 8.14 don't involve generalized normals. We need to recognize properties of generalized normals, in addition to (H) from Definition 2.41 of prenormal structures and those of presubdifferential structures implying (H) by Propositions 2.42 and 2.43, which are sufficient for deriving abstract counterparts of Theorems 8.5 and 8.14(ii,iii) in the corresponding settings of Banach spaces.

In what follows, impose in addition to (H) the following properties of generalized normals that certainly hold for any reasonable prenormal structure $\widehat{\mathcal{N}}(\cdot;\Omega)$ on a Banach space X:

(H1) If $\Omega\subset X$ is a linear subspace of X and if $\bar{x}\in\Omega$, then

$$\widehat{\mathcal{N}}(\bar{x};\Omega)=\Omega^{\perp}:=\big\{x^*\in X^*\big|\ \langle x^*,x\rangle=0\ \text{ whenever }\ x\in\Omega\big\}$$

is a subspace orthogonal to Ω.

(H2) For all closed subsets Ω_1 and Ω_2 of X such that $\Omega_1\times\Omega_2\subset X$ and for every points $\bar{x}_i\in\Omega_i,\ i=1,2$, one has

$$\widehat{\mathcal{N}}\big((\bar{x}_1,\bar{x}_2);\Omega_1\times\Omega_2\big)\subset\widehat{N}(\bar{x}_1;\Omega_1)\times\widehat{N}(\bar{x}_2;\Omega_2)\,.$$

Note that, by Proposition 1.2, the product property (H2) holds as equality for Fréchet normals and that (H2) is always induced by property (S3) of presubdifferentials $\widehat{\mathcal{D}}$ from Subsect. 2.5.1 for subdifferentially generated prenormal structures $\widehat{\mathcal{N}}(\cdot;\Omega)=\widehat{\mathcal{D}}\delta(\cdot;\Omega)$; see the proof of Proposition 2.42.

The next assertion provides an abstract counterpart of the *approximate* second welfare theorem for both *Pareto and weak Pareto* optimal allocations established in Theorem 8.5.

Theorem 8.16 (abstract versions of the approximate second welfare theorem for Pareto and weak Pareto optimal allocations). *Let $\mathcal{E}$ be an economy with a Banach commodity space E, let $X := E^{n+m+1}$, and let $\widehat{\mathcal{N}}$ be a prenormal structure on X with properties* (H1) *and* (H2). *Considering a local Pareto* (*resp. weak Pareto*) *optimal allocation* $(\bar{x}, \bar{y})$ *of $\mathcal{E}$ with $\bar{w}$ defined in* (8.2), *assume that the net demand qualification condition* (*resp. net demand weak qualification condition*) *holds at* $(\bar{x}, \bar{y})$. *Then for every $\varepsilon > 0$ there exist a suboptimal triple*

$$(x, y, w) \in \prod_{i=1}^{n} \operatorname{cl} P_i(\bar{x}) \times \prod_{j=1}^{m} \operatorname{cl} S_j \times \operatorname{cl} W$$

with the aggregate commodity w defined in (8.1) *and a common marginal price $p^* \in E^* \setminus \{0\}$ satisfying relations* (8.13)–(8.16) *with $\widehat{N}$ replaced by $\widehat{\mathcal{N}}$.*

Proof. It follows the procedure in proving Theorem 8.5 with the use of the abstract version of the approximate extremal principle from Theorem 2.51(i), which holds for any prenormal structure, and also using properties (H1) and (H2) needed to accomplish this procedure. △

Since the normal cone of convex analysis always satisfies assumptions (H), (H1), and (H2) of Theorem 8.16, all the conclusions of Corollary 8.6 on the perturbed decentralized equilibrium for *convex* economies holds in arbitrary Banach spaces. It is not however the case for Theorem 8.7 on *nonconvex* economies as discussed above.

An abstract approximate version of the second welfare theorem for *strong Pareto* optimal allocations of economies with ordered commodity spaces *doesn't* require net demand qualification conditions as stated next.

Theorem 8.17 (abstract version of the approximate second welfare theorem for strong Pareto optimal allocations). *Let $(\bar{x}, \bar{y})$ be a local strong Pareto optimal allocation of the economy $\mathcal{E}$ with an ordered Banach commodity space E, let the sets S_j and W be locally closed near $\bar{y}_j$ and $\bar{w}$, respectively, and let $\widehat{\mathcal{N}}$ be a prenormal structure on $X = E^{n+m+1}$ satisfying properties* (H1) *and* (H2). *Suppose that:*

(a) *either E_+ is generating and one of the free disposal/desirability assumptions* (8.40)–(8.42) *is fulfilled,*

(b) *or $\operatorname{int} E_+ \neq \emptyset$ and at least two sets among W, S_j for $j = 1, \ldots, m$, and $P_i(\bar{x})$ for $i = 1, \ldots, n$, satisfy the corresponding free disposal/desirability assumptions in* (8.40)–(8.42).

Then for every $\varepsilon > 0$ there exist a suboptimal triple

$$(x,y,w)\in\prod_{i=1}^{n}\Big[\mathrm{cl}\,P_i(\bar{x})\bigcap\Big(\bar{x}_i+\frac{\varepsilon}{2}I\!B\Big)\Big]$$

$$\times\prod_{j=1}^{m}\Big[S_j\bigcap\Big(\bar{y}_j+\frac{\varepsilon}{2}I\!B\Big)\Big]\times\Big[W\bigcap\Big(\bar{w}+\frac{\varepsilon}{2}I\!B\Big)\Big]$$

with the aggregate commodity w *defined* (8.1) *and a common marginal price* $p^*\in E^*$ *satisfying relations* (8.13)–(8.16), *where* $\widehat{N}$ *is replaced by* $\widehat{\mathcal{N}}$.

Proof. Follows the one for Theorem 8.14(i, iii) with the use of the abstract extremal principle from Theorem 2.51(i). △

Next we derive *exact versions* of the abstract second welfare theorem for Pareto, weak Pareto, and strong Pareto optimal allocations of nonconvex economies with Banach commodity spaces by passing to the limit from the corresponding approximate versions. To proceed, we need to employ the *abstract sequential normal compactness condition* ($\widehat{\mathcal{N}}$-SNC) introduced in Definition 2.50. This condition depends on the given prenormal structure $\widehat{\mathcal{N}}$; it reduces to our basic SNC property for closed subsets of Asplund spaces when $\widehat{\mathcal{N}}=\widehat{N}$, the Fréchet normal cone. Note that the following abstract extensions of the second welfare theorem require the *sequential* normal compactness property although the marginal price relations are generally expressed in terms of *topological* (net) limiting normal structures on Banach spaces. This is a definite advantage of the results obtained.

First we present an abstract extension of the exact second welfare theorem for Pareto and weak Pareto optimal allocations, which imposes the net demand qualification conditions of Definition 8.3 and generalizes the corresponding results of Theorem 8.8. As in the case of Theorem 8.8, observe that the $\widehat{\mathcal{N}}$-SNC condition in the next theorem is imposed only on *one* among the preference, production, and net demand constraint sets, in contrast to the general exact extremal principle of Theorem 2.51(ii), where this condition is required for *all but one* sets in the extremal system.

Theorem 8.18 (abstract versions of the exact second welfare theorem for Pareto and weak Pareto optimal allocations). *Let* $(\bar{x},\bar{y})$ *be a local Pareto* (*resp. weak Pareto*) *optimal allocation of the economy* $\mathcal{E}$ *satisfying the corresponding assumptions of Theorem* 8.16. *Taking a prenormal structure* $\widehat{\mathcal{N}}$ *on* X, *suppose in addition that one of the sets*

$$\mathrm{cl}\,P_i(\bar{x}),\quad i=1,\ldots,n,\qquad \mathrm{cl}\,S_j,\quad j=1,\ldots,m,\qquad \mathrm{cl}\,W$$

is $\widehat{\mathcal{N}}$*-SNC at the corresponding point. Then there is a nonzero price* $p^*\in E^*$ *satisfying the relations*

$$-p^*\in\overline{\mathcal{N}}\big(\bar{x}_i;\mathrm{cl}\,P_i(\bar{x})\big),\quad i=1,\ldots,n, \tag{8.46}$$

$$p^* \in \overline{\mathcal{N}}(\bar{y}_j; \mathrm{cl}\, S_j), \quad j = 1, \ldots, m, \tag{8.47}$$

$$p^* \in \overline{\mathcal{N}}(\bar{w}; \mathrm{cl}\, W), \tag{8.48}$$

where $\overline{\mathcal{N}}$ *stands for the topological normal structure* (2.67) *generated by* $\widehat{\mathcal{N}}$. *Furthermore, the topological structure* $\overline{\mathcal{N}}$ *can be replaced in* (8.46)–(8.48) *by the sequential normal structure* $\mathcal{N}$ *generated by* $\widehat{\mathcal{N}}$ *in* (2.66) *if the closed dual ball* $I\!B^* \subset E^*$ *is weak* sequentially compact.*

Proof. Similarly to the proof of Theorem 8.8, take an arbitrary sequence $\varepsilon_k \downarrow 0$ as $k \to \infty$ and, according to the abstract approximate version of Theorem 8.16, find sequences $(x_k, y_k, w_k, p_k^*) \in E \times E \times E \times E^*$ satisfying

$$x_{ik} \in \mathrm{cl}\, P_i(\bar{x}) \cap (\bar{x}_i + \varepsilon_k I\!B), \quad i = 1, \ldots, n\ ,$$

$$y_{jk} \in \mathrm{cl}\, S_j \cap (\bar{y}_j + \varepsilon_k I\!B), \quad j = 1, \ldots, m\ ,$$

$$w_k = \sum_{i=1}^{n} x_{ik} - \sum_{j=1}^{m} y_{jk} \in \mathrm{cl}\, W \cap (\bar{w} + \varepsilon_k I\!B)$$

and the dual relations (8.13)–(8.16) with $\varepsilon = \varepsilon_k$ and $\widehat{N}$ replaced by $\widehat{\mathcal{N}}$. Obviously $(x_k, y_k, w_k) \to (\bar{x}, \bar{y}, \bar{w})$ as $k \to \infty$. Note that the price sequence $\{p_k^*\}$ is *bounded* in E^*. Invoking basic functional analysis, one gets a *weak* cluster point* (in the sense of convergent *subsets*) $p^* \in \mathrm{cl}^*\{p_k^* |\ k \in I\!N\}$ of this sequence in general Banach commodity spaces. If the closed unit ball $I\!B^*$ of E^* is weak* sequentially compact (as for either Asplund or β-smooth Banach spaces E), then $\{p_k^*\}$ contains a *subsequence* that weak* converges to some $p^* \in E^*$. Passing to the limit in (8.13)–(8.15) with the prenormal structure $\widehat{\mathcal{N}}$ therein, we conclude that the cluster point p^* in both topological and sequential cases satisfies the limiting relations (8.46)–(8.48) in terms of, respectively, the topological and sequential structure generated by $\widehat{\mathcal{N}}$.

It remains to show that we can *choose* $p^* \neq 0$ if *one* of the sets $\mathrm{cl}\, P_i(\bar{x})$, $\mathrm{cl}\, S_j$, and $\mathrm{cl}\, W$ is $\widehat{\mathcal{N}}$-*sequentially* normally compact at the corresponding point. This is straightforward for Banach spaces with weak* sequentially compact dual balls , but requires some arguments in the general (not sequential) case.

Assume for definiteness that the set $\mathrm{cl}\, W$ is $\widehat{\mathcal{N}}$-sequentially normally compact at $\bar{w}$ and that $p^* = 0$ is *the only* weak* cluster point of $\{p_k^*\}$. Then $p_k^* \xrightarrow{w^*} 0$ as $k \to \infty$ for the *whole sequence.* Due to (8.15) via $\widehat{\mathcal{N}}$ we have

$$p_k^* + \varepsilon_k b_k^* \in \widehat{\mathcal{N}}(w_k; \mathrm{cl}\, W) \ \text{ with some } \ b_k^* \in I\!B^* \ \text{ for all } \ k \in I\!N\ ,$$

and hence $p_k^* + \varepsilon_k b_k^* \xrightarrow{w^*} 0$ as $k \to \infty$. This implies, by the $\widehat{\mathcal{N}}$-SNC property of $\mathrm{cl}\, W$, that $\|p_k^* + \varepsilon_k b_k^*\| \to 0$ and thus $\|p_k^*\| \to 0$ as $k \to \infty$. The latter clearly contradicts the nontriviality condition (8.16) for p_k^* in Theorem 8.16 and completes the proof of this theorem. △

It is easy to see that the results of both Corollaries 8.9 and 8.10 of Theorem 8.8 holds true in the abstract framework of Theorem 8.18 provided that the normal structure in this theorem agrees with the normal cone of convex analysis for closed convex subsets of E.

Consider next economies with *ordered* commodity spaces E. First we observe that, under the standard desirability *or* free disposal conditions formulated in (8.40)–(8.42), *all* the qualification and SNC assumptions of Theorem 8.18 are automatic provided that the closed positive cone E_+ is *solid*, i.e., of *nonempty interior*. Thus we arrive at the following abstract version of the second welfare theorem for Pareto and weak Pareto optimal allocations of economies with ordered commodity spaces. For brevity we present this result only for *topological* normal structures in general Banach spaces; its sequential counterpart under the weak* sequential compactness of $I\!B^* \subset E^*$ is formulated similarly, as in the case of Theorem 8.18.

Corollary 8.19 (abstract second welfare theorem for Pareto and weak Pareto optimal allocations in ordered spaces). *Let E be an ordered Banach space with* $\operatorname{int} E \neq \emptyset$*, and let the topological normal structure $\overline{\mathcal{N}}$ in Theorem* 8.18 *be such that $\overline{\mathcal{N}}(\cdot;\Omega)$ is not larger than the Clarke normal cone for closed subsets of E. The following assertions hold:*

(i) *Given a local weak Pareto optimal allocation $(\bar{x},\bar{y})$ of $\mathcal{E}$, assume that either the net demand constraint set W is closed near $\bar{w}$ exhibiting the free disposal of commodities* (8.40)*, or one of the production sets S_j is closed near $\bar{y}_j$ and obeys the free disposal condition* (8.41)*. Then there is a nonzero marginal price $p^* \in E^*$ satisfying relations* (8.46)–(8.48).

(ii) *Given a local Pareto optimal allocation $(\bar{x},\bar{y})$ of $\mathcal{E}$ for $n > 1$, assume that the i-th consumer satisfies the desirability condition* (8.42)*. Then there is a nonzero marginal price $p^* \in E^*$ satisfying relations* (8.46)–(8.48).

Proof. It is easy to observe that, for any subset Ω of a Banach space, the inclusion $\Omega + K \subset \Omega$ with some nonempty open cone K implies the *epi-Lipschitzian* property of Ω around every $\bar{x} \in \operatorname{cl}\Omega$. Thus each of the conditions (8.40)–(8.42) with $\operatorname{int} E_+ \neq \emptyset$ ensures the epi-Lipschitzian property of the corresponding set and hence, by Proposition 8.4, the fulfillment of the net demand (resp. weak) qualification condition imposed in Theorem 8.18. Since the Clarke normal cone is weak* locally compact for epi-Lipschitzian sets in any Banach space, such sets have the sequential normal compactness property with respect to this cone. This yields the latter property for the corresponding sets in (8.40)–(8.42) with respect to any prenormal structure, which is not larger than the Clarke normal cone. Hence all the assumptions of Theorem 8.18 hold, and we arrive at the marginal price relations (8.46)–(8.48) for Pareto and weak Pareto optimal allocations. △

Similarly to the basic case considered in Subsect. 8.3.2, we finally conclude that the abstract version of the exact second welfare theorem for *strong Pareto* optimal allocations doesn't require net demand constraint qualifications and may hold for convex and nonconvex economies with ordered commodity spaces having closed positive cones of *empty interior*.

Theorem 8.20 (abstract versions of the exact second welfare theorem for strong Pareto optimal allocations). *Let $(\bar{x}, \bar{y})$ be a local strong Pareto optimal allocation of the economy $\mathcal{E}$ with an ordered Banach commodity space E, and let the sets S_j and W be locally closed near $\bar{y}_j$ and $\bar{w}$ respectively. Suppose that one of the sets*

$$\mathrm{cl}\, P_i(\bar{x}), \quad i = 1, \ldots, n, \qquad S_j, \quad j = 1, \ldots, m, \qquad W$$

is $\widehat{\mathcal{N}}$-SNC at the corresponding point, where the abstract prenormal structure $\widehat{\mathcal{N}}$ satisfies assumptions (H1) *and* (H2), *and that*

(a) *either E_+ is generating and one of the free disposal/desirability conditions* (8.40)–(8.42) *is fulfilled,*

(b) *or $E_+ \neq \emptyset$ and at least two sets among W, S_j, $j = 1, \ldots, m$, and $P_i(\bar{x})$, $i = 1, \ldots, n$, satisfy the corresponding conditions in* (8.40)–(8.42).

Then there is a dual element $p^* \in E^* \setminus \{0\}$ satisfying the marginal price relations (8.46)–(8.48), *where $\overline{\mathcal{N}}$ stands for the topological normal structure generated by $\widehat{\mathcal{N}}$. Furthermore, the topological structure $\overline{\mathcal{N}}$ can be replaced in* (8.46)–(8.48) *by the sequential normal structure $\mathcal{N}$ generated by $\widehat{\mathcal{N}}$ if the closed unit ball $I\!B^*$ of E^* is weak* sequentially compact.*

Proof. Pass to limit in the approximate relations of Theorem 8.17 for strong Pareto optimal allocations similarly to the proof of Theorem 8.18. △

The abstract results derived in this subsection admit efficient concretizations for the specific prenormal and normal structures on the corresponding classes of Banach spaces discussed in Subsect. 2.5.3.

8.4.2 Public Goods and Restriction on Exchange

In the concluding subsection we briefly discuss extensions of the methods and results developed in this chapter to economies with public goods. We also mention some possible applications of this approach to competitive equilibrium models with public environment and with restriction on exchange.

In contrast to the welfare economic model studied above, economies with *public goods* involve two categories of commodities: private and public. Consumption of the first type is exclusive, i.e., what is taken by any one individual automatically becomes unavailable for all others. On the contrary, goods are public if their consumption is identical across all individuals. Mathematically this means that the commodity space E is represented as the product of two

Banach spaces $E = X \times Z$, where X and Z are the space of private and public commodities, respectively. Thus consumer variables $x_i \in X$, $i = 1, \ldots, n$, stand for private goods, while those of $z_i \in Z$, $i = 1, \ldots, n$, correspond to public goods of commodities; $y_j \in S_j \subset E$ connote production variables as above. Considering for simplicity the "markets clear" setting in (8.1) with the given initial endowment of scare recourses $\omega \in X$ only for *private* goods, we write the market constraints in the economy involving both private and public goods as follows:

$$\sum_{i=1}^{n}(x_i, z_i) - \sum_{j=1}^{m} y_j = (\omega, 0) . \tag{8.49}$$

Note that the market constraint condition (8.49) reflects the fact that there is *no endowment of public goods*, which is the most crucial characteristic feature of public good economies.

Proceeding as in the above case of economies with no public goods, by incorporating the market constraint condition (8.49) into the construction of the set Ω_2 in (8.18), we obtain similar results for economies with public goods applying the *extremal principle*. These results include both approximate and exact forms of the extended second welfare theorem for all the three types of Pareto optimal allocations, as well as abstract versions of these theorems presented in Subsect. 8.4.1. The main changes for public goods economies, in comparison with the basic results of this chapter, are as follows presented only for the case of the exact/limiting conditions from in Theorem 8.8: instead of the existence of a nonzero marginal price $p^* \in E^*$ satisfying (8.30) and (8.31), we have prices $p^* = (p_x^*, p_z^*) \in X^* \times Z^*$ and $p_i^* \in Z^*$ as $i = 1, \ldots, n$ with $(p_x^*, p_i^*) \neq 0$ for at least one $i \in \{1, \ldots, n\}$ and such that

$$-(p_x^*, p_z^*) \in N\big(\bar{x}_i; \mathrm{cl}\, P_i(\bar{x})\big), \quad i = 1, \ldots, n , \tag{8.50}$$

$$(p_x^*, p_z^*) \in N(\bar{y}_j; \mathrm{cl}\, S_j), \quad j = 1, \ldots, m, \quad \text{and} \tag{8.51}$$

$$p_z^* = \sum_{i=1}^{n} p_i^* . \tag{8.52}$$

Observe that, while conditions (8.50) and (8.51) are actually concretizations of those in (8.30) and (8.31) for the product structure of the commodity space $E = X \times Z$, the last one in (8.52) confirms the fundamental conclusion for welfare economics with public goods that goes back to Samuelson [1189]: *the marginal rates of transformation for public goods are equal to the sum of the individual marginal rates of substitution* at Pareto optimal allocations.

As seen above, the main mathematical tool of our approach to the study of Pareto optimality in models of welfare economics is the *extremal principle* of variational analysis applied to the systems of sets Ω_1 and Ω_2 defined in

(8.17) and (8.18) for models with no public goods and modified by (8.49) for public goods economies. Similar considerations work for welfare models with the so-called *public environment* involving combinations of a *private market* section with a *private non-market sector* (e.g., the legal system); see Villar [1287] for more details and examples.

Mathematically such models can be described similarly to the basic welfare model of Subsect. 8.4.1 with consumer, production, and preference sets depending on *parameters*. Appropriate versions of the extended second welfare theorem for such models with Asplund commodity spaces were derived in Habte's dissertation [533] based on the extremal principle.

Observe that all the models considered above don't involve any *restrictions on exchange* between their agents. Such restrictions are taking into account in some other models of competitive economic equilibria based on *rationing schemes*; see, e.g., Makarov, Levin and Rubinov [829] and Rubinov [1182]. Mathematically most of these models may be written in the form similar to those studied above but with more complex relationships between consumer and production variables in the market constraint conditions in comparison with (8.1) and (8.49). This leads to modifications of the *market constraint set* Ω_2 in the corresponding extremal system and can be handled by employing the *extremal principle* of variational analysis. An important *direct distribution model* of this type was studied by Habte [533] who derived for it various versions of the extended second welfare theorem.

8.5 Commentary to Chap. 8

8.5.1. Competitive Equilibria and Pareto Optimality in Welfare Economics. Competitive equilibrium models and the basis ideas of *market price decentralization* between consumers and producers/firms go back to Léon Walras who provided, in his seminal work "Eléments d'Economie Politique Pure" [1300] published in 1874–1877, justified answers to remarkable questions raised by several of his predecessors. In particular, Adam Smith asked in [1217] why a large number of agents motivated by self-interest and making independent decision don't create social chaos in a private ownership economy. Smith himself gained a deep insight into the impersonal coordination between market behavior of consumers and firms making his famous conclusion on *Invisible Hand*. However, only a mathematical model could provide a scientific justification of empirical observations and conclusions. Constructing such a model, Walras laid the foundation of *general equilibrium theory* for competitive economies known also as models of *welfare economics*.

Roughly speaking, a competitive *price equilibrium* is a situation in which all agents of the economy simultaneously achieve their plans at given prices. It has long been recognized, starting with Pareto's work [1053], that there are some relationships between competitive equilibria in welfare economics and appropriate notions of *efficiency* for feasible allocations of resources; the concept of efficiency in economics is generally identified with *Pareto optimality.* According to the classical *Pareto principle*, a feasible allocation is better than another one if it is preferred by *all* the agents, i.e., "better than" means unanimous agreement. There are several versions and modern interpretations of Pareto optimality useful in economic modeling; see particularly Definition 8.2 and the comments below.

Probably the first rigorously justified *necessary condition for Pareto optimality* in models of welfare economics was published by Lange [738] who established the *equality* between the *marginal rates of substitution* in consumption and production sectors at any Pareto optimal allocation; see also Hicks [566], Samuelson [1188], and Khan [671] for comments on previous attempts, related results, and unpublished material. Lange's proof was based on the observation that a Pareto optimal allocations of resources could be interpreted, under certain *qualification conditions*, as an optimal solution to a constrained problem of *nonlinear programming*, which allowed him to use the classical *Lagrange multiplier rule* under, of course, the standard *smoothness* assumptions on the utility and production functions in finite-dimensional commodity spaces. This result is now known as the original version of the *second fundamental theorem of welfare economics*; the name appeared later in the Arrow-Debreu model for *convex* economies. In fact, Lange's result follows, under the differentiability and convexity assumptions, from the necessary condition for Pareto optimality in the Arrow-Debreu model for convex economies labeled as the "second welfare theorem"—see below.

8.5.2. Convex Models of Welfare Economics. The so-called *Arrow-Debreu model of general equilibrium* initiated in the 1951 papers by Arrow [26] and by Debreu [309] has played a fundamental role in mathematical economics and its applications and has also greatly influenced the development of optimization-related areas in mathematics, particularly that of *convex analysis.* There are numerous publications on various aspects of the Arrow-Debreu model for convex economies with finite-dimensional and infinite-dimensional commodity spaces; see, e.g., the books by Aliprantis, Brown and Burkinshaw [10], Debreu [310], Florenzano [459], Mas-Colell, Whinston and Green [856], and the references therein.

In their model, Arrow and Debreu shifted the focus on marginal rates of substitution to the *decentralization* of Pareto optimal allocations as price equilibria. Under the *convexity* and associated assumptions, they established *necessary and sufficient* conditions for Pareto optimality via its *equivalence* to their competitive price *equilibrium* understood in the following sense: there are *prices* at which expenditure *minimization* by consumers and profit *maximization* by producers sustain the given Pareto optimal allocation. Moreover, the *existence* of such a price equilibrium was proved under appropriate assumptions for general convex models with both finite-dimensional and infinite-dimensional commodity spaces. The major mathematical tools employed in the proofs of these results were *separation* theorems for deriving optimality conditions and *fixed-point* theorems (Brouwer and Kakutani) for establishing the existence of equilibria; all of them are based on *convexity*.

The names of the *first* and *second* welfare theorems came from the corresponding parts of the *equivalence* between Pareto optimality and price equilibria. The *first welfare theorem* states that any equilibrium allocation is Pareto optimal (*sufficient* condition for Pareto optimality), while the *second welfare theorem* states the converse: any Pareto optimal allocation provides a price equilibrium (*necessary* condition for Pareto optimality). Observe that the validity of the first welfare theorem heavily depends on convexity; it doesn't have any analogs in nonconvex (even smooth) models. At the same time, the second welfare theorem admits far-going extensions to nonconvex models; this is actually the main topic of Chap. 8. Note also that, being based on convexity, the Arrow-Debreu model doesn't need any differentiability assumptions as in all the previous developments.

Besides the afore-mentioned convexity hypotheses, the Arrow-Debreu model requires *nonempty interiors* of some sets involved in economies with infinitely many commodities. Mathematically this is due to the application of separation theorems in *infinite-dimensional* spaces. In the case of ordered topological spaces, the interiority assumption reduces in fact to the nonempty interior requirement on the positive cone/orthant in the commodity space in question, which is *not* fulfilled in many situations important for economic modeling. To avoid the latter restrictive requirement, Mas-Colell [855] proposed his celebrated *properness* assumption for convex economies whose ordered commodity spaces are topological vector *lattices* with possibly empty interiors. Various extensions and improvements of Mas-Colell's properness condition for convex economies with finite-dimensional and infinite-dimensional commodity spaces can be found in Aliprantis, Tourky and Yannelis [14], Florenzano [459], Mas-Colell, Whinston and Green [856], Tourky [1261], and their references. We finally mention the very recent paper by Naniewicz [993], which develops a new approach to the Arrow-Debreu model with usual convexity but no interiority assumptions in reflexive commodity spaces. The approach developed of [993] is based on reducing the economic model to a system of *variational inequalities* and employing the theory of pseudo-monotone multivalued mappings.

8.5.3. Enter Nonconvexity. As we have mentioned in Subsect. 8.1.1., the *relevance of convexity* assumptions is often doubtful for many important economic applications, as had been realized even before developing the Arrow-Debreu model; see the citation from Samuelson's book [1188] presented above. Indeed, various types of nonconvexity inevitably arise in modeling monopolistic competition, increasing returns to scale, incomplete markets, externalities, etc.; see more examples and discussions in Anderson [18], Cornet [287], Florenzano [459], Khan [671], Quinzii [1113], Villar [1287], and the references therein.

In his pioneering study on price decentralization in nonconvex models of welfare economics, Guesnerie [524] established a generalized version of the second welfare theorem in the form of necessary optimality conditions for Pareto optimal allocations of nonconvex economies. Instead of postulating convexity of the initial production and preference sets, Guesnerie *assumed the convexity* of their (local) *tangential* approximations and then employed the classical *separation* theorem for convex cones. He formalized this procedure by using the "cone of interior displacements" developed by Dubovitskii and Milyutin [370] in general optimization theory.

Guesnerie's approach to the study of Pareto optimality in nonconvex economies was extended in many publications concerning economies with both finite-dimensional and infinite-dimensional commodity spaces; see, e.g., Bonnisseau and Cornet [135], Brown [181], Cornet [286], Khan and Vohra [673, 674], Quinzii [1113], Villar [1287], and their references. Most of these publications employed the *Clarke tangent cone* that has an advantage of being *automatically convex* and hence can be treated by using the classical convex separation. In this way, *marginal prices* (corresponding to marginal rates of substitution/transformation in nonsmooth and nonconvex models) were formalized via the *dual* Clarke normal cone. However, it has been recognized in a while that Clarke's normal cone may often be *too large* for adequate descriptions of marginal pricing; see examples and discussions in Jouini [642] and Khan [671].

In the latter paper [671] (its first version appeared as a preprint of 1987), Khan obtained a significantly *more satisfactory* extension of the second welfare theorem to nonconvex economies with *finite-dimensional* commodity spaces. In his generalized second welfare theorem, marginal prices were formalized via our *basic normal cone*. Note that his approach didn't involve *any convex separation* while employing instead a reduction to necessary optimality conditions (Lagrange multipliers) in *nonsmooth* mathematical programming established by Mordukhovich [892]; cf. Theorem 5.21(iii) from Subsect. 5.1.3 in finite dimensions. In this way, Khan's approach signified the return to the classical "Lagrange multiplier" viewpoint taken by Hicks, Lange, and Samuelson in the foundations of welfare economics versus the separation approach pioneered by Arrow and Debreu. A similar version of the generalized second welfare theorem in terms of our basic normal cone in finite dimensions was derived by Cornet [288] for a somewhat different economic model by using a direct proof of necessary optimality conditions for the corresponding maximization problem.

Further developments on the second welfare theorem in various nonconvex models of welfare economics with finite-dimensional and infinite-dimensional commodity spaces were given by Borwein and Zhu [164], Flåm [452], Flåm and Jourani [453], Florenzano, Gourdel and Jofré [460], Habte [533], Jofré [633], Jofré and Rivera [635], Khan [669, 670], Malcolm and Mordukhovich [836], Mordukhovich [920, 922, 930], Villar [1288], and Zhu [1375] by using the basic normal cone as well as its infinite-dimensional extensions, modifications, and abstract versions. These developments will be discussed below in more details.

8.5.4. Extremal Principle and Nonconvex Separation in Models of Welfare Economics. In Mordukhovich [920, 922], we suggested an approach to studying Pareto optimality and deriving extended versions of the second welfare theorem for nonconvex economies based on the *extremal principle* of variational analysis fully discussed in Chap. 2. Recall that the extremal principle provides necessary optimality conditions for local extremal points of closed set systems covering particularly local solutions to problems in constrained mathematical programming and vector optimization. On the other hand, it gives a *variational* counterpart of the classical *separation* in the case of *nonconvex* sets, playing essentially the same role in nonconvex variational analysis as separation theorems do in the convex framework. Thus the approach to Pareto optimality based on the extremal principle can be viewed as a *unification* of both the previous approaches discussed above for smooth and convex models. Note that all the results presented in Chap. 8 are based on the application of the extremal principle.

A somewhat different but closely related approach to the study of Pareto optimality in models of welfare economics was proposed by Jofré [633]. His approach is based on the application of a subdifferential condition for *boundary points* of *sum of sets* derived by Borwein and Jofré [148]. This property of a set sum, treated in [148] as a *nonconvex separation*, is actually *equivalent* to the local extremality of *another* set system. Furthermore, the subdifferential characterization of boundary points of set sums obtained in [148] happens to be equivalent to the approximate version of the extremal principle; see the recent papers by Kruger [716] and Zhu [1375] for more precise statements and discussions. The results on the extended second welfare theorem for nonconvex economies obtained in the afore-mentioned developments [164, 452, 453, 460, 533, 633, 635, 836, 920, 922, 930, 1375] were derived by either a direct application of the extremal principle, or by using the equivalent boundary/nonconvex separation property from Borwein and Jofré [148].

8.5.5. The Basic Model and Solution Concepts. The general model of welfare economics described in Subsect. 8.1.1 has been widely accepted in the modern microeconomic literature in the case of $W = \{\omega\}$, where ω is the given initial *aggregate endowment* of scare resources; see, e.g., the books by Aliprantis, Brown and Burkinshaw [10] and by Mas-Colell, Whinston and Green [856]. Note that the preference relation in the economy $\mathcal{E}$ is given by set-valued mappings P_i with no use of preodering, utility functions, and other conventional attributes of the classical welfare economics; cf. Debreu [310].

When $W = \{\omega\}$, the crucial relation (8.1) in Definition 8.1 of feasible allocations reduces to the so-called *markets clear* condition. Introducing the *net demand constraint set* W as in Mordukhovich [920, 922] and Malcolm and Mordukhovich [836] happens to be useful for at least the following two reasons:

—it allows us to consider the standard case $W = \{\omega\}$ simultaneously with the so-called *implicit free disposal* of commodities $W = \omega - E_+$ (see, e.g., Cornet [288]) defined via the closed positive cone $E_+ \subset E$ of the ordered commodity space E;

—with an arbitrary set W, the feasibility condition (8.1) reflects an *uncertainty* situation in the market when the initial aggregate endowment is *not exactly known* due to, e.g., *incomplete information*; cf. particularly the discussion on minimax control problems in uncertainty conditions from Subsect. 7.5.19.

The notions of (local) *Pareto* and *weak Pareto* optimal allocations given in Definition 8.2(i,ii) are standard in the economic literature; they are in accordance with the conventional concepts of Pareto and weak Pareto optimality in general vector/multiobjective optimization problems; see, e.g., Sect. 5.3 and the comments to it with the corresponding references.

To the best of our knowledge, the notion of *strong Pareto* optimal allocations in models of welfare economics given in Definition 8.2(iii) was clearly introduced and studied for the first time by Khan [670]. According to the personal communication with Ali Khan, it was in line with Debreu's work [311] and was partially brought out by the previous study on asymptotics from Khan and Rashid [672], which was in turn motivated by Hildenbrand [568] and was further extended by Anderson [18].

8.5.6. Qualification Conditions. As mentioned, certain *constraint qualification conditions* were present in the initial versions of the second-welfare theorem for smooth and convex models of welfare economics as well as in *all* of their further developments concerning Pareto and weak Pareto optimal allocations. The crucial conditions of this type imposed in the Arrow-Debreu convex model (there are several versions and modifications of them) are known as the (nonempty) *interiority conditions*. The net demand qualification conditions formulated in Definition 8.3 can be viewed as far-going extensions of the classical interiority properties to the case of nonconvex models of welfare economics.

Both NDQ and NDWQ conditions imposing the required properties for Pareto and weak Pareto optimal allocations, respectively, first appeared in Mordukhovich [920], while the previous version of the NDQ with $W = \{\omega\}$ and closed production sets S_j was formulated by Jofré in [633] (and in his earlier preprints with Rivera; see [633] and [635] with the references therein) under the name of "asymptotically included condition." Furthermore, if in the latter case the sequence e_k is replaced by αe with a fixed vector $e \in E$ and all $\alpha > 0$ sufficiently small, the NDQ condition (8.3) reduces to the qualification condition that goes back to Cornet [286] and is known as either "radially Lipschitzian condition" (as in Bonnisseau and Cornet [135]) or "Cornet's constraint qualification" (as in Khan [671]); see also more references and discussions in the afore-mentioned papers. We refer the reader to the recent work by Zhu [1375] and Borwein and Zhu [164] for further developments and applications of the NDWQ condition for weak Pareto optimal allocations.

Proposition 8.4 giving, in a parallel way, easily verifiable sufficient conditions for the fulfillment of the NDQ and NDWQ properties was established in full generality by Malcolm and Mordukhovich [836], while some of its previous versions in the case of $W = \{\omega\}$ were given by Bonnisseau and Cornet [135], Cornet [286], Jofré [633], and Khan [670, 671]. Note that the property formulated in assertion (i) of Proposition 8.4 is a direct generalization of the *desirability direction* condition by Mas-Colell [854], which is related to the classical "more is better" assumption for convex economies with commodity spaces ordered by their closed positive cones having nonempty interiors.

As has already been discussed after Definition 8.3, both NDQ and NDWQ conditions are automatic if *at least one* among preference and production is locally *epi-Lipschitzian*, which is equivalent—for *convex* sets—to imposing the classical *nonempty interiority* assumption on the corresponding set. Set properties of the latter type were called by Khan [671] to be "fat in some direction." Note also that the *summation* of sets used in formulating the NDQ and NDWQ conditions tends to improve the epi-Lipschitzian property of the resulting sets, especially in the case of a *larger number* of agents in the market.

One of the advantages of using the net demand constraint set W in our model is that it allows us, by Proposition 8.4, to avoid *any* requirements on the preference and production sets while imposing the epi-Lipschitzian property of W, which never happens when $W = \{\omega\}$. In this way we automatically cover the welfare model involving the so-called "free-disposal Pareto optima" studied by Cornet [288] in the case of finite-dimensional commodity spaces.

Observe that our qualification conditions are *not* related to Mas-Colell's *properness* condition and its modifications for convex models with ordered commodity spaces discussed in Subsect. 8.5.2. Some *nonconvex* versions of Mas-Colell's properness have been recently introduced and studied by Florenzano, Gourdel and Jofré [460] in the *weak* Pareto optimality framework of the generalized second welfare theorem for models whose ordered commodity spaced are endowed with a Banach *lattice structure*. Note that we have never imposed a lattice structure in our study.

8.5.7. Approximate Versions of the Second Welfare Theorem. To avoid the afore-mentioned nonempty interiority and properness assumptions, several *approximate* versions of the second welfare theorem were established and economically interpreted for various microeconomic models dealing with Pareto and weak Pareto optimal allocations under the *convexity* hypotheses. We particularly refer the reader to the papers by Aliprantis and Burkinshaw [11], Hildenbrand [568], Khan and Rashid [672], and Khan and Vohra [675]. Observe that the latter paper employed for these purposes the celebrated

Bishop-Phelps theorem [116] on the boundary density of support points to convex sets in Banach spaces, which was considered by Ekeland [399] as the "grandfather" of modern variational principles.

In Subsect. 8.2.1 we develop, following Mordukhovich [920, 922] and Malcolm and Mordukhovich [836], approximate versions of the second welfare theorem for Pareto and weak Pareto optimal allocations of *nonconvex* economies with marginal prices formalized via the Fréchet normal cone in Asplund spaces as in [836, 922] and also via more general "prenormal" structures in appropriate Banach spaces as in [920]; see also Subsect. 8.4.1. The proofs of these results are based on the corresponding versions of the *approximate extremal principle*, which can be viewed as nonconvex extensions of the Bishop-Phelps theorem; see Proposition 2.6, Corollary 2.21, and the subsequent discussions in Subsect. 2.6.4.

Results of this type were also derived by Jofré [633] as "viscous" versions of the second welfare theorem for Pareto optimal allocations of nonconvex economies in Banach spaces. Considering a welfare model in the "markets clear" setting under the "asymptotically included" qualification condition, Jofré established a subdifferential form of the approximate/viscous second welfare theorem via axiomatically defined subdifferentials satisfying some general requirements. However, not all of these requirements are satisfied for the Fréchet subdifferential in Banach spaces, in contrast to those in Mordukhovich [920]. Thus the results presented in Subsect. 8.2.1 cannot be derived from [633, Theorem 2] based on the application of the nonconvex boundary condition by Borwein and Jofré [148]; see Subsect. 8.5.4.

8.5.8. Exact Versions of the Second Welfare Theorem under Normal Compactness Conditions. By *exact* (or pointbased) versions of the second welfare theorem we understand necessary conditions for Pareto-like optimality with *marginal prices* formalized via normal cone constructions defined *exactly at* optimal allocations. Results of this type are the *most needed* for economic applications; they include all the classical versions of the second welfare theorem for various economies with finite-dimensional and infinite-dimensional commodity spaces under smoothness and convex assumptions. Concerning nonconvex economies, the majority of the results of this "exact" type were obtained as marginal pricing rules formalized via Clarke's normal cone, while their improvements in terms of basic normals were derived by Khan [670] and by Cornet [288] for economies with finite-dimensional commodity spaces; see Subsect. 8.5.3 for more details and discussions.

As has been well recognized, welfare economies with *infinite-dimensional* commodity spaces require some additional *amount of compactness* for the validity of exact versions of the second welfare theorem, which is fully in accordance with general optimization theory in infinite dimensions.; cf. Chap. 5. Observe that for *convex* economies, sufficient amounts of compactness are *implicitly* contained in the classical interiority and properness assumptions.

The situation is different for *nonconvex* economies, where most results of the exact/pointbased type *explicitly* assume the *epi-Lipschitzian* property of certain sets involved in the model; see particularly Bonnisseau and Cornet [135], Khan [670], and Khan and Vohra [673, 674]. As we know, the latter property happens to be an appropriate extension of the classical interiority condition to the nonconvex setting.

The results of the exact type presented in Subsects. 8.2.2 and 8.3.1 can be found in Mordukhovich [922] and Malcolm and Mordukhovich [836], while their (not full) abstract versions given in Subsect. 8.4.1 are taken from Mordukhovich [920]. The extended versions of the second welfare theorem from Subsects. 8.2.2 and 8.3.1 formalize the marginal pricing rule via the *basic normal cone* in *Asplund* commodity spaces, which can be replaced by its abstract either topological or sequential limiting counterparts in the *appropriate Banach* settings of Subsect. 8.4.1.

A remarkable feature of all the exact versions of the extended second welfare theorem from Sects. 8.2–8.4 is the imposes of the basic *sequential normal compactness* (SNC) property and its abstract *sequential* modification on *just one* of either preference, or production, or net demand constraint sets of the welfare models under consideration. This is of a *striking difference* with the other similarly looking situations studied in the book that concerned exact/limiting results for *finitely many* sets in *infinite dimensions*, namely: the exact extremal principle, calculus rules, and necessary conditions in constrained optimization. Indeed, in all the previous settings we required the SNC and/or related properties for *all but one* of the sets involved in the case study. The significant improvement achieved in the economic model under consideration is due to the specific *linearly separated* structure of the constraint set (8.18) from the extremal system to which we apply the extremal principle.

To this end we mention the exact subdifferential versions of the second welfare theorem obtained by Jofré [633] and similarly by Flåm and Jourani [453] via abstract limiting subdifferentials of the distance function in appropriate Banach spaces. These results assumed the *compactly epi-Lipschitzian* property of *one* of the sets involved in the welfare models. As we know from Subsect. 1.1.4, the epi-Lipschitzian property happens to be a *topological* counterpart of the SNC property and *strictly* implies the latter not only in general Banach spaces without any separability-like structure but also in those spaces whose closed dual balls are weak* sequentially compact, particularly in the Asplund space setting (even for *convex* sets as in Example 3.6 from Subsect. 3.1.1). The reader can find further extensions of the exact second welfare theorem under compactly epi-Lipschitzian assumptions in Flåm [452] and in Florenzano, Gourdel and Jofré [460].

8.5.9. Pareto Optimality in Ordered Commodity Spaces. Considering nonconvex economies with *ordered* commodity spaces, it is natural to ask about the *positivity* of the marginal price p^* satisfying the relations of the *exact* extended second welfare theorem, in the sense that $0 \neq p^* \in E^*_+$

for the dual positive cone $E^*_+ \subset E^*$ associated with the closed positive cone E ordering the commodity space E of the economy $\mathcal{E}$. It has been observed by Malcolm and Mordukhovich [836] that the required price positivity follows from *any* of the "exact" normal cone relationships (8.30)–(8.32) for the marginal price p^* in Theorem 8.8 if the corresponding preference, or production, or net demand constraint set satisfies (*one* of) the conventional *desirability*, *free disposal*, or *implicit free disposal* assumptions listed in Theorem 8.13. This is a direct consequence of the positivity property of basic normals to "free disposal" sets from Lemma 8.12 proved in [836].

It should be emphasized again that all the results on the second welfare theorem discussed above concern either (local) *Pareto* or *weak Pareto* optimal allocations of economies under the fulfillment of the corresponding *constraint qualification* condition as NDQ and NDWQ, which extend the classical interiority condition to nonconvex economies. Since any strong Pareto optimal allocation is obviously a Pareto one, the results obtained are also fulfilled for the more restrictive notion of strong Pareto optimality in economic modeling.

As has been already mentioned, the concept of *strong Pareto* optimality in models of welfare economics was introduced by Khan [670] who employed it for deriving an exact version of the second welfare theorem with marginal pricing formalized via Ioffe's "approximate" normal cone (A-normal cone) in locally convex linear topological spaces [597], which is an infinite-dimensional extension of our basic finite-dimensional construction. Khan's main result in [670] justified such a generalized second welfare theorem for strong (locally) Pareto optimal allocations in nonconvex economies whose ordered commodity spaces were assumed to be *lattices* with reflexive preference relations and with "free-disposal" net demand constraint sets of the type $W = \omega - E_+$. Furthermore, it was assumed in [670] the fulfillment of *both* desirability and free disposal conditions from (a) and (b) of Theorem 8.13 for *all* $i = 1, \ldots, n$ and *all* $j = 1, \ldots, m$, and the validity of the *epi-Lipschitzian* property for *every* production and preference sets around the strong Pareto optimal allocation under consideration.

It follows from Theorem 8.13 in the Asplund space setting and from its abstract analog in Corollary 8.19 for nonconvex economies with any ordered Banach commodity spaces that the improved versions of Khan's second welfare theorem hold under *significantly* less restrictive assumptions for arbitrary *weak Pareto* optimal allocations, but not merely for strong ones. Indeed, taking into account the sufficient condition for the NDWQ property established in Proposition 8.4(ii), we conclude that the extended second welfare theorem holds in Khan's framework (with *no lattice* while with *Banach* structure on the commodity space in question) for *any weak Pareto* optimal allocation provided that *only one* among the preference or production sets is *epi-Lipschitzian* around the corresponding point, whereas *neither* the desirability condition (a) *nor* the free disposal condition (b) of Theorem 8.13 is required to be satisfied. Moreover, under these weaker assumptions, our afore-mentioned results improve Khan's marginal pricing rule by using either the basic normal cone in Asplund spaces or Ioffe's G-normal cone in general Banach spaces, which both are *smaller* than Ioffe's A-normal cone employed by Khan.

In his other paper [669], Khan built (based on some constructions from Treiman [1262]) an example of a nonconvex economy with the classical *Asplund* commodity space c_0 (of sequences of real numbers converging to zero and endowed with the supremum norm) such that Ioffe's A-normal cone to the production set is the *entire* dual space ℓ^1 at the Pareto optimal allocation. Another remarkable feature of this example is that not only our basic normal cone but even its weak* *convexification*—Clarke's normal cone—is *strictly smaller* than Ioffe'a A-normal cone and provides therefore a nontrivial marginal price information in the framework of the generalized second welfare theorem. Observe that Khan's results from [670] are not applied in this example, since the production set is not epi-Lipschitzian and doesn't exhibit the free disposal of commodities.

8.5.10. Strong Pareto Optimality with No Qualification Conditions. It surprisingly happens that *strong Pareto* optimal allocations play a distinguished role in welfare economic models (both convex and nonconvex) with ordered commodity spaces: they *don't need* any net demand *qualification conditions* (including the nonempty interiority and properness ones) for the validity of approximate and exact versions of the second welfare theorem. This was observed by Mordukhovich in [920, 922] and then developed in the recent paper [930]; the corresponding material is presented in Subsect. 8.3.2 via our basic constructions in the Asplund space setting and in Subsect. 8.4.1 in the general framework of abstract normal/subdifferential structures in Banach spaces. The results on the *modified* strong Pareto optimal allocations discussed in Remark 8.15 and based on the personal communication by Glenn Malcolm have not been yet published.

Let us emphasize first of all that in our approach to the second welfare theorem developed in Sect. 8.2 for *Pareto* and *weak Pareto* optimal allocations, qualification conditions are needed *only* to show that such allocations can be reduced to *local extremal points* of some system of sets. Then we apply the *extremal principle* and appropriate calculus rules. Analyzing this scheme in the case of *strong Pareto* optimal allocations of economies with *ordered* commodity spaces under *free-disposal-like* conditions, we observe that such constraint qualification (strongly related to the nonempty interiority of positive cones in this case) are *not* required due to the *very nature* of strong Pareto optimality, which directly leads us to extremal points of sets.

More precisely, we need to impose the afore-mentioned free disposal/ desirability assumptions on *at least two* among production, preference, and net demand constraint sets for accomplishing such a reduction. Otherwise, the positive cone of the ordered space is required to be *generating* in the sense of $E = E_+ - E_+$, which doesn't seem to be a restrictive assumption; it holds, e.g., in any *Banach lattice* (or Riesz spaces). Note that in the latter case of generating positive cones, the reduction of strong Pareto optimal allocations to local extremal points is based on the deep *Krein-Šmulian theorem* from the theory of ordered Banach spaces.

8.5.11. Nonlinear Pricing. As well recognized, *shadow prices* conventionally used in economic modeling and involved in various versions of the second welfare theorem are mathematical interpreted as *dual* vectors, or linear continuous functionals, over commodity spaces. From this viewpoint they can be called *linear prices*.

The usage of *nonlinear prices* for the achievement of market equilibria and other desirable value-based characteristics in welfare economics has been explored in the economic literature, especially in models involving price discrimination, progressive tax tariffs, land markets, and portfolio trading. Probably one of the first publications on nonlinear pricing was the paper by Arrow and Hurwicz [27].

Quite recently a new approach to nonlinear pricing has been initiated by Aliprantis, Tourky and Yannelis [14] for *convex* models of welfare economics with *no lattice* structure of commodity spaces. The main motivation came from the fact that Mas-Colell's fundamental theory of welfare economics with no interiority assumptions crucially requires lattice properties of commodity spaces, *even* in finite-dimensional settings. We particularly refer the reader to the paper by Aliprantis, Monteiro and Tourky [13] containing a striking example of the convex economy with two traders and a *three-dimensional* commodity spaces without a lattice structure, where there is *no* Walrasian equilibrium and where the second welfare theorem *fails*.

As shown by Aliprantis, Tourky and Yannelis [14], the usage of new *nonlinear* prices vs. linear ones in the previous developments provides an adequate general equilibrium theory in finite-dimensional and infinite-dimensional *convex* models with *no* lattice structure of commodities. Note that nonlinear prices used in [14] are always *concave* and positively homogeneous while they may be *nonsmooth*. Furthermore, they *reduce* to standard *linear prices* in vector *lattices*; see more details and references in [14] and in the subsequent paper by Aliprantis, Florenzano and Tourky [12] with further developments and applications.

Some results of a completely different type on nonlinear pricing are presented in Sect. 8.2 (see particularly Theorem 8.7 and Remark 8.11); the reader can find more results and discussions in Mordukhovich [930], while probably for the first time such a nonlinear price interpretation of marginal pricing was observed in Malcolm and Mordukhovich [836, Corollary 4.3]. This approach has nothing to do with lattice or even ordering structures of commodity spaces, but signifies the difference between second welfare theorems in *convex* and *nonconvex* models.

Indeed, *linear* prices in *convex* economies support a fully *decentralized* equilibrium, in the sense that each firm *maximizes* its profit and each consumer *minimizes* his/her expenditure at Pareto optimal allocations. Marginal pricing versions of the second welfare theorem provide only *local* descriptions of linear prices at optimal allocations via some normal cones. It happens however that the usage of *smooth nonlinear prices* allows us to support a *decentralized* (convex-type, maximization-minimization) equilibrium *globally* over all the preference and production sets in fully *nonconvex* models. Moreover, the *rates of change* (i.e., derivatives) of these nonlinear prices *at* Pareto optimal allocations are *arbitrarily close* to the linear *marginal price* from the *approximate* second welfare theorem.

Mathematically these nonlinear price conclusions are based on *smooth variational descriptions* of *Fréchet normals* obtained in Theorem 1.30, which contains strong geometric results of variational analysis involving *variational principles* in their proofs. It should be emphasized that the structure of Fréchet normals is *crucial* for such variational descriptions; these results, in contrast to the other approximate and exact versions of the extended second welfare theorem in Sects. 8.2 and 8.3, don't admit *abstract* counterparts presented and discussed in Sect. 8.4.

The afore-mentioned variational descriptions of Fréchet normals allow us to provide useful and economically valuable interpretations of marginal prices from the *exact* versions of the extended second welfare theorem obtained in Sects. 8.2 and 8.3. Since our *basic normals* are approximated by Fréchet normals in *Asplund* spaces, the marginal price equilibrium relations (8.30)–(8.32) established in Theorem 8.8 and used also in the modified exact versions of Sect. 8.3 can be interpreted as a *limiting decentralized equilibrium* in nonconvex models realized via *nonlinear prices.* Observe some similarities between this limiting decentralized equilibrium supported by nonlinear prices and the so-called "virtual equilibrium" introduced recently by Jofré, Rockafellar and Wets [636] in convex Walrasian models of exchange via a limiting procedure from a classical equilibrium supported by linear prices. Their approach was based on reductions to nonmonotone *variational inequalities*; it was further extended in the subsequent paper [637] to a more general convex Walrasian model of consumption and production with market trading.

Observe that it doesn't seem to be possible to derive results of such a limiting decentralized type from the previous formalizations of marginal prices in nonconvex models of welfare economics via the Clarke normal cone and also via Ioffe's extensions of the basic normal cone to the general Banach space setting. On the other hand, in some special settings discussed in Subsect. 2.5.2, basic normals admit limiting representations in terms of other more primitive normals with a variational structure. In particular, this can be done via *proximal normals* in the finite-dimensional and also in Hilbert space settings, which allows us provide an economic interpretation of limiting marginal prices via a certain *perturbed* maximization and minimization of *quadratic* functions; see Jofré [633] and Jofré and Rivera [635] for more details and discussions.

8.5.12. Abstract Versions. In the last section of Chap. 8 (and of the whole two-volume book!) we discuss some further possible counterparts and generalizations of the extended results on the second welfare theorem developed in Sects. 8.1–8.3. Since our approach to economic modeling is mainly based on the *extremal principle* of variational analysis, we briefly consider some settings, where certain versions of the extremal principle can be readily applied.

First we consider the same model as in Sects. 8.1–8.3 analyzing the possibility to apply the *abstract* versions of the extremal principle developed in Subsect. 2.5.3. The reader can see that, while some of the results obtained in Sects. 8.1–8.3 (particularly related to nonlinear prices and positivity) require Fréchet-type normals and/or the Asplund space setting, most of the obtained extensions of the second welfare theorem hold true in other (including arbitrary) Banach space settings with *appropriate* (pre)normal structures satisfying the revealed axiomatic requirements. We have mentioned above various abstract extensions of the second welfare theorem to axiomatically defined normals and subgradients developed by Jofré [633], Flåm and Jourani [453], and Flåm [452]. It seems that the abstract results of Subsect. 8.4.1, based on the paper by Mordukhovich [920], are the most general among other abstract versions of the extended second welfare theorem for the economic model under consideration.

8.5.13. Further Extensions. Welfare economic models with *public goods* were first studied, under *smoothness* assumptions, in the 1954 paper by Samuelson [1189] who established a public goods version of the "foundations" results by Hicks and Lange with the fundamental conclusion that "the marginal rates of transformation for public goods are equal to the sum of the individual marginal rates of substitution." It took thirteen years from Samuelson's work to obtain the corresponding version by Foley [463] of the Arrow-Debreu second welfare theorem for *convex* economies with public goods; see the recent paper by De Simone and Graziano [326] and its references for developing a public goods welfare theory in Mas-Colell's properness framework for convex economies. Various results on the extended second welfare theorem for *nonconvex* models were obtained by Khan and Vohra [673], Khan [670, 671], Flåm and Jourani [453], Villar [1287, 1288], and other researchers.

As follows from the discussions in Subsect. 8.4.2, our methods developed for welfare economies with no public goods can be easily extended to the case of public goods economies, keeping with Samuelson's fundamental conclusion on marginal rates of transformation and substitution; see (8.52).

Other models (with public environment as in Villar [1287] and with direct distribution as in Makarov, Levin and Rubinov [829]) were considered in detail, from the viewpoint of the extended second welfare theorem handled via the extremal principle, in the dissertation by Habte [533].

References

1. Y. A. Abramovich and C. D. Aliprantis (2002), *An Invitation to Operator Theory*, American Mathematical Society, Providence, Rhode Island.
2. A. A. Agrachev and Y. L. Sachkov (2004), *Control Theory from the Geometric Viewpoint*, Springer, Berlin.
3. N. U. Ahmed (2005), Necessary conditions of optimality with discontinuous vector fields, *Nonlinear Funct. Anal. Appl.* **10**, 129–150.
4. N. U. Ahmed and K. L. Teo (1981), *Optimal Control of Distributed Parameter Systems*, Elsevier, New York.
5. N. U. Ahmed and X. Xiang (1997), Nonlinear uncertain systems and necessary conditions of optimality, *SIAM J. Control Optim.* **35**, 1755–1772.
6. N. U. Ahmed and X. Xiang (1997), Necessary conditions of optimality for differential inclusions in Banach spaces, *Nonlinear Anal.* **30**, 5437–5445.
7. V. M. Alekseev, V. M. Tikhomirov and S. V. Fomin (1987), *Optimal Control*, Consultants Bureau, New York.
8. A. D. Alexandrov (1939), Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected with it, *Uchenye Zapiski Leningrad Gos. Univ., Ser. Math.* **6**, 3–35.
9. J. J. Alibert and J.-P. Raymond (1998), A Lagrange multiplier theorem for control problems with state constraints, *Numer. Funct. Anal. Optim.* **19**, 697–704.
10. C. D. Aliprantis, D. J. Brown and O. Burkinshaw (1990), *Existence and Optimality of Competitive Equilibria*, Springer, New York.
11. C. D. Aliprantis and O. Burkinshaw (1988), The fundamental theorems of welfare economics without proper preferences, *J. Math. Econ.* **17**, 41–54.
12. C. D. Aliprantis, M. Florenzano and R. Tourky (2005), Linear and non-linear price decentralization, *J. Econ. Theory*, to appear.
13. C. D. Aliprantis, P. K. Monteiro and R. Tourky (2004), Non-marketed options, non-existence of equilibria, and non-linear prices, *J. Econ. Theory* **114**, 345–357.
14. C. D. Aliprantis, R. Tourky and N. C. Yannelis (2001), A theory of value with non-linear prices: Equilibrium analysis beyond vector lattices, *J. Econ. Theory* **100**, 22–72.
15. K. Allali and L. Thibault (2005), Fréchet subdifferentials of marginal functions, *Control Cybernet.*, to appear.

16. T. Amahroq and N. Gadhi (2001), On the regularity conditions for vector programming problems, *J. Global Optim.* **21**, 435–443.
17. L. Ambrosio, O. Ascenzi and G. Buttazzo (1989), Lipschitz regularity of intergral functionals with highly discontinuous integrands, *J. Math. Anal. Appl.* **142**, 301–316.
18. R. M. Anderson (1988), The second welfare theorem with non-convex preferences, *Econometrica* **56**, 361–282.
19. T. S. Angell and A. Kirsch (1990), On the necessary conditions for optimal control of retarded systems, *Appl. Math. Optim.* **22**, 117–145.
20. M. Anitescu (2005), Global convergence of an elastic mode approach for a class of mathematical programs with complementarity constraints, *SIAM J. Control Optim.*, to appear.
21. N. Arada (2001), Minimax Dirichlet boundary control problem with state constraints, *Nonlinear Anal.* **46**, 653–673.
22. N. Arada, M. Bergounioux and J.-P. Raymond (2000), Minimax controls for uncertain distributed parabolic systems, *SIAM J. Control Optim.* **38**, 1481–1500.
23. N. Arada and J.-P. Raymond (1999), Optimality conditions for state-constrained Dirichlet boundary control problems, *J. Optim. Theory Appl.* **102**, 51–68.
24. N. Arada and J.-P. Raymond (2002), Dirichlet boundary control of semilinear parabolic equations, II: Problems with pointwise state constraints, *Appl. Math. Optim.* **45**, 145–167.
25. V. I. Arkin and V. L. Levin (1973), Convexity of the values of vector integrals, theorems on measurable selection, and variational problems, *Russian Math. Surveys* **27**, 21–85
26. K. J. Arrow (1951), An extension of the basic theorems of classical welfare economics, in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 507–532, University of California, Berkeley, California.
27. K. J. Arrow and L. Hurwicz (1960), Decentralization and computation in resource allocation, in *Essays in Economics and Econometrics in Honor of Harold Hotelling*, edited by R. W. Pfouts, pp. 34–104, University of North Carolina Press, Chapel Hills, North Carolina.
28. Z. Artstein (1994), First order approximations for differential inclusions, *Set-Valued Anal.* **2**, 7–17.
29. Z. Artstein (1995), A calculus for set-valued maps and set-valued evolution equations, *Set-Valued Anal.* **3**, 213–261.
30. Z. Artstein and V. Gaitsgory (2000), The value function of singularly perturbed control systems, *Appl. Math. Optim.* **41**, 425–445.
31. Z. Artstein and C. C. Popa (2004), Orlicz-Young structures of Young measures, *Acta Appl. Math.* **80**, 1–33.
32. A. V. Arutyunov (2000), *Optimality Conditions: Abnormal and Degenerate Problems*, Kluwer, Dordrecht, The Netherlands.
33. A. V. Arutyunov and S. M. Aseev (1997), Investigation of the degeneracy phenomenon of the maximum principle for optimal control problems with state constraints, *SIAM J. Control Optim.* **35**, 930–952.
34. A. V. Arutyunov, S. M. Aseev and V. I. Blagodatskikh (1994), Necessary conditions of the first order in the problem of optimal control of a differential inclusion with phase constraints, *Math. Sbornik* **79**, 117–139.

35. A. V. Arutyunov and A. F. Izmailov (2005), Sensitivity analysis for cone-constrained optimization problems under relaxed constraint qualifications, *Math. Oper. Res.* **30**, 333–353..
36. A. V. Arutyunov and A. F. Izmailov (2005), Directional stability theorem and directional metric regularity, *Math. Oper. Res.*, to appear.
37. A. V. Arutyunov and F. L. Pereira (2005), Second-order necessary conditions of optimality for problems without a priori normality assumptions, *Math. Oper. Res.*, to appear.
38. A. V. Arutyunov and R. B. Vinter (2004), A simple 'finite approximations' proof of the Pontryagin maximum principle under reduced differentiability hypotheses, *Set-Valued Anal.* **12**, 5–24.
39. S. M. Aseev (1991), Smooth approximations of differential inclusions and the time-optimality problem, *Proc. Steklov Inst. Math.* **200**, 27–34.
40. S. M. Aseev (1997), A method of smooth approximations in the theory of necessary optimality conditions for differential inclusions, *Izvestia: Mathematics* **61**, 235–258.
41. S. M. Aseev (2001), Extremal problems for differential inclusions with phase constraints, *Proc. Steklov Inst. Math.* **233**, 1–63.
42. L. T. Ashchepkov and O. V. Vasiliev (1975), On optimality of singular controls in Goursat-Darboux systems, *USSR Comput. Maths. Math. Phys.* **15**, 1157–1167.
43. E. Asplund (1968), Fréchet differentiability of convex functions, *Acta Math.* **121**, 31–47.
44. H. Attouch (1984), *Variational Convergence for Functions and Operators*, Pitman, Boston, Massachusetts.
45. H. Attouch and H. Ruani (1993), Stability results for Ekeland's ε-variational principle and cone extremal solutions, *Math. Oper. Res.* **18**, 173–201.
46. H. Attouch and R. J-B. Wets (1991), Quantitative stability of variational systems, I: The epigraphical distance, *Trans. Amer. Math. Soc.* **338**, 695–729.
47. H. Attouch and R. J-B. Wets (1993), Quantitative stability of variational systems, III: ε-approximate solutions, *Math. Progr.* **61**, 197–214.
48. J.-P. Aubin (1981), Contingent derivatives of set-valued maps and existence of solutions to nonlinear inclusions and differential inclusions, in *Mathematical Analysis and Applications*, edited by L. Nachbin, pp. 159–229, Academic Press, New York.
49. J.-P. Aubin (1984), Lipschitz behavior of solutions to convex minimization problems, *Math. Oper. Res.* **9**, 87–111.
50. J.-P. Aubin and A. Cellina (1984), *Differential Inclusions*, Springer, Berlin.
51. J.-P. Aubin and I. Ekeland (1976), Estimates of the duality gap in nonconvex programming, *Math. Oper. Res.* **1**, 225–245.
52. J.-P. Aubin and I. Ekeland (1984), *Applied Nonlinear Analysis*, Wiley, New York.
53. J.-P. Aubin and H. Frankowska (1987), On inverse function theorems for set-valued maps, *J. Math. Pures Appl.* **66**, 71–89.
54. J.-P. Aubin and H. Frankowska (1990), *Set-Valued Analysis*, Birkhäuser, Boston, Massachusetts.
55. R. J. Aumann (1965), Integrals of set-valued functions, *J. Math. Anal. Appl.* **12**, 1–12.

56. A. AUSLENDER (1978), Differential stability in nonconvex and nondifferentiable programming, *Math. Progr. Study* **10**, 29–41.
57. A. AUSLENDER (1984), Stability in mathematical programming with nondifferentiable data, *SIAM J. Control Optim.* **22**, 239–254.
58. A. AUSLENDER AND R. COMINETTI (1991), A comparable study of multifunction differentiability with applications in mathematical programming, *Math. Oper. Res.* **16**, 240–258.
59. A. AUSLENDER AND J.-P. CROUZEIX (1988), Global regularity theorems, *Math. Oper. Res.* **13**, 243–253.
60. A. AUSLENDER AND M. TEBOULLE (2003), *Asymptotic Cones and Functions in Optimization and Variational Inequalities*, Springer, New York.
61. D. AUSSEL, J.-N. CORVELLEC AND M. LASSONDE (1995), Mean value property and subdifferential criteria for lower semicontinuous functions, *Trans. Amer. Math. Soc.* **347**, 4147–4161.
62. D. AUSSEL, J.-N. CORVELLEC AND M. LASSONDE (1995), Nonsmooth constrained optimization and multidimensional mean value inequalities, *SIAM J. Optim.* **9**, 690–706.
63. D. AUSSEL, A. DANIILIDIS AND L. THIBAULT (2004), Subsmooth sets: Functional characterizations and related concepts, *Trans. Amer. Math. Soc.* **357**, 1275–1302.
64. E. P. AVAKOV (1988), Necessary conditions of minima for nonregular problems in Banach spaces: Maximum principle for abnormal problems of optimal control, *Proc. Steklov Inst. Math.* **185**, 3–29.
65. E. R. AVAKOV, A. A. AGRACHEV AND A. V. ARUTYUNOV (1992), The level set of a smooth mapping in a neighborhood of a singular point, *Math. Sbornik* **73**, 455–466.
66. J. AVELIN (1997), *Differential Calculus cfor Multifunctions and Nonsmooth Functions*, Ph.D. dissertation, Department of Mathematics, Uppsala University, Sweden.
67. J. AVELIN (2000), Differential calculus for complex-valued multifunctions, *J. Appl. Anal.* **6**, 47–76.
68. V. I. AVERBUKH AND O. G. SMOLYANOV (1968), The various definitions of the derivative in linear topological spaces, *Russian Math. Surveys* **23**, 67–113.
69. D. AZÉ AND J.-N. CORVELLEC (2004), Characterizations of error bounds for lower semicontinuous functions on metric spaces, *Control Optim. Calc. Var.*, **10**, 409-425.
70. D. AZÉ, J.-N. CORVELLEC AND R. E. LUCCHETTI (2002), Variational pairs and applications to stability in nonsmooth analysis, *Nonlinear Anal.* **49**, 643–670.
71. D. AZÉ AND J.-B. HIRIART-URRUTY (2002), Optimal Hoffman type estimates in eigenvalue and semidefinite inequality constraints, *J. Global Optim.* **24**, 133–147.
72. R. L. BAIRE (1905), *Lecons sur les Fonctions Discontinues*, Gauthier-Villars, Paris.
73. A. V. BALAKRISHNAN (1965), Optimal control problems in Banach spaces, *SIAM J. Control* **3**, 152–180.
74. A. V. BALAKRISHNAN (1981), *Applied Functional Analysis*, 2nd edition, Springer, New York.
75. E. J. BALDER (2000), Lectures on Young measure theory and its applications to economics, *Rens. Istit. Mat. Univ. Trieste* **31**, 1–69.

76. S. Banach (1932), *Théorie des Opérations Linéaires*, Monografje Matematyczne, Warsaw.
77. B. Bank, J. Guddat, D. Klatte, B. Kummer and K. Tammer (1982), *Non-Linear Parametric Optimization*, Birkhaüser, Basel.
78. H. T. Banks (1968), Necessary conditions for control problems with variable time lags, *SIAM J. Control* **6**, 9–47.
79. H. T. Banks and G. A. Kent (1972), Control of functional differential equations of retarded and neutral type to target sets in function spaces, *SIAM J. Control* **10**, 567–591.
80. H. T. Banks and K. Kunisch (1989), *Estimation Techniques for Distributed Parameter Systems*, Birkhäuser, Boston, Massachusetts.
81. H. T. Banks and A. Manitius (1974), Applications of abstract variational theory to hereditary systems – a survey, *IEEE Trans. Autom. Control* **19**, 524–533.
82. V. Barbu (1993), *Analysis and Control of Nonlinear Infinite Dimensional Systems*, Academic Press, Boston, Massachusetts.
83. V. Barbu, I. Lasiecka and R. Triggiani (2005), Boundary stabilization of Navier-Stokes equations, *Mem. Amer. Math. Soc.*, to appear.
84. V. Barbu and T. Precupanu (1986), *Convexity and Optimization in Banach Spaces*, D. Reidel, Dordrecht, The Netherlands.
85. M. Bardi and I. Capuzzo Dolcetta (1997), *Optimal Control and Viscosity Solutions of Hamilton-Jacobi Equations*, Birkhäuser, Boston, Massachusetts.
86. N. E. Barron and R. Jensen (1991), Optimal control and semicontinuous viscosity solutions, *Proc. Amer. Math. Soc.* **113**, 397–402.
87. T. Başar and P. Bernhard (1995), *H_∞-Optimal Control and Related Minimax Design Problems*, 2nd edition, Birkhäuser, Boston, Massachusetts.
88. H. Bauschke, J. M. Borwein and W. Li (1999), Strong conical hull intersection property, bounded linear regularity, Jameson's property (G), and error bounds in convex optimization, *Math. Progr.* **86**, 135–160.
89. M. S. Bazaraa, J. J. Goode and M. Z. Nashed (1974), An the cone of tangents with applications to mathematical programming, *J. Optim. Theory Appl.* **13**, 389–426.
90. C. R. Bector, S. Chandra and J. Dutta (2004), *Principles of Optimization Theory*, Alpha Science International Publishers, UK.
91. E. M. Bednarczuk (2003), Order-Lipschitzian properties of multifunctions with applications to stability of efficient points, *Control Cybernet.* **32**, 491–502.
92. G. Beer (1993), *Topologies on Closed and Convex Sets*, Kluwer, Dordrecht, The Netherlands.
93. S. Bellaassali and A. Jourani (2004), Necessary optimality conditions in multiobjective dynamic optimization, *SIAM J. Control Optim.* **42**, 2043–2061.
94. R. Bellman and K. L. Cooke (1963), *Differential-Difference Equations*, Academic Press, New York.
95. R. Bellman, I. Glicksberg and O. Gross (1956), On the "bang-bang" control problem, *Quart. Appl. Math.* **14**, 11–18.
96. H. Benabdellah (2000), Existence of solutions to the nonconvex sweeping process, *J. Diff. Eq.* **164**, 286-295.
97. H. Benabdellah, C. Castaing, A. Salvadori and A Syam (1996), Nonconvex sweeping process, *J. Appl. Anal.* **2**, 217–240.

98. M. Benaim, J. Hofbauer and S. Sorin (2005), Stochastic approximations and differential inclusions, *SIAM J. Control Optim.* **44**, 328–348.
99. J. Benoist (1994), Approximation and regularization of arbitrary sets in finite dimensions, *Set-Valued Anal.* **2**, 95–115.
100. A. Bensoussan (1988), *Perturbation Methods in Optimal Control*, Wiley, New York.
101. A. Bensoussan, G. Da Prato, M. Delfour, S. K. Mitter (1992–93), *Representation and control of infinite-dimensional systems*, published in two volumes, Birkhäuser, Boston, Massachusetts.
102. A. Ben-Tal and J. Zowe (1985), Directional derivatives in nonsmooth optimization, *J. Optim. Theory App.* **47**, 483–490.
103. M. Bergounioux and D. Tiba (1996), General optimality conditions for constrained convex control problems, *SIAM J. Control Optim.* **34**, 698–711.
104. M. Bergounioux and F. Tröltzsch (1996), Optimality conditions and generalized bang-bang principle for a state-constrained semilinear parabolic problem, *Numer. Func. Anal. Optim.* **17**, 517–536.
105. L. D. Berkovitz (1974), *Optimal Control Theory*, Springer, New York.
106. L. D. Berkovitz (1976), A penalty function proof of the maximum principle, *Appl. Math. Optim.* **2**, 291–303.
107. H. Berliocchi and J. M. Lasry (1973), Principe de Pontryagin pour des systémes régis par une equation différentielle multivoque, *C. R. Acad. Sci. Paris* **277**, 1103–1105.
108. F. Bernard and L. Thibault (2004), Prox-regularity of functions and sets in Banach spaces, *Set-Valued Anal.* **12**, 25–47.
109. F. Bernard and L. Thibault (2005), Prox-regular functions in Hilbert spaces, *J. Math. Anal. Appl.* **303**, 1–14.
110. F. Bernard and L. Thibault (2005), Uniform prox-regularity of functions and epigraphs in Hilbert spaces, *Nonlinear Anal.* **60**, 187–207.
111. D. P. Bertsekas (1999), *Nonlinear Programming*, Athena Scientific, Boston, Massachusetts.
112. D. P. Bertsekas and A. E. Ozdaglar (2002), Pseudonormality and a Lagrange multiplier theory for constraint optimization, *J. Optim. Theory Appl.* **114**, 287–343.
113. D. Bessis, Y. S. Ledyaev and R. B. Vinter (2001) Dualization of the Euler and Hamiltonian inclusions, *Nonlinear Anal.* **43**, 861–882.
114. R.-M. Bianchini and M. Kawski (2003), Needle variations that cannot be summed, *SIAM J. Control Optim.* **42**, 218–238.
115. J. Birge and L. Qi (1995), Subdifferential convergence in stochastic programming, *SIAM J. Optim.* **5**, 436–453.
116. E. Bishop and R. R. Phelps (1963), The support functionals of convex sets, in *Convexity*, edited by V. Klee, vol. VII of *Proceedings of Symposia in Pure Mathematics*, pp. 27–35, American Mathematical Society, Providence, Rhode Island.
117. V. I. Blagodatskikh (1984), The maximum principle for differential inclusions, *Proc. Steklov Inst. Math.* **166**, 23–43.
118. V. I. Blagodatskikh and A. F. Filippov (1986), Differential inclusions and optimal control, *Proc. Steklov Inst. Math.* **169**, 199–259.
119. G. A. Bliss (1946), *Lectures on the Calculus of Variations*, University of Chicago Press, Chicago, Illinois.

120. N. A. Bobylev, S. V. Emel'yanov and S. K. Korovin (2004), Convexity of images of convex sets under smooth maps, *Comput. Math. Model.* **15**, 213–222.
121. N. N. Bogolyubov (1930), Sur quelques methodes nouvelles dans le calculus des variations, *Ann. Math. Pura Appl.* **7**, 249–271.
122. J. Bolte, A. Daniilidis and A. Lewis (2004), The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems, preprint.
123. J. Bolte, A. Daniilidis and A. Lewis (2005), A nonsmooth Morse-Sard theorem for subanalytic functions, *J. Math. Anal. Appl.*, to appear.
124. V. G. Boltyanskii (1958), The maximum principle in the theory of optimal processes, *Dokl. Akad. Nauk SSSR* **119**, 1070–1073.
125. V. G. Boltyanskii (1968), Local section method in the theory of optimal processes, *Diff. Eq.* **4**, 2166–2183.
126. V. G. Boltyanskii (1973), The maximum principle for problems of optimal steering, *Diff. Eq.* **9**, 1363–1370.
127. V. G. Boltyanskii (1973), *Optimal Control of Discrete Systems*, Nauka, Moscow.
128. V. G. Boltyanskii (1994), The maximum principle – how it came to be?, Report No. 526, Math. Institut, Tech. Univ. München, München, Germany.
129. V. G. Boltyanskii, R. V. Gamkrelidze and L. S. Pontryagin (1956), On the theory of optimal processes, *Dokl. Akad. Nauk SSSR* **110**, 7–10.
130. O. Bolza (1913), Über den anormalen fall beim Lagrangeschen und Mayerschen problem mit gemischten behingugen und variabeln endpunkten, *Mathematische Annalen* **74**, 430–446.
131. J. F. Bonnans and E. Casas (1992), A boundary Pontryagin's principle for the optimal control of state-constrained elliptic systems, *Int. Ser. Numer. Math.* **107**, 241–249.
132. J. F. Bonnans, R. Cominetti and A. Shapiro (1999), Second-order optimality conditions based on parabolic second-order tangent sets, *SIAM J. Optim.* **9**, 466–492.
133. J. F. Bonnans and A. Shapiro (2000), *Perturbation Analysis of Optimization Problems*, Springer, New York.
134. J. F. Bonnans and A. Sulem (1995), Pseudopower expansion of solutions of generalized equations and constraint optimization problems, *Math. Progr.* **70**, 123–148.
135. J.-M. Bonnisseau and B. Cornet (1988), Valuation equilibrium and Pareto optimum in non-convex economies, *J. Math. Econ.* **17**, 293–308.
136. D. Borwein, J. M. Borwein and X. Wang (1996), Approximate subgradients and coderivatives in $I\!R^n$, *Set-Valued Anal.* **4**, 375–398.
137. J. M. Borwein (1986), Stability and regular points of inequality systems, *J. Optim. Theory Appl.* **48**, 9–52.
138. J. M. Borwein (1987), Epi-Lipschitz-like sets in Banach spaces: Theorems and examples, *Nonlinear Anal.* **11**, 1207–1217.
139. J. M. Borwein, M. Fabian, I. Kortezov and P. D. Loewen (2001), The range of the gradient of a continuously differentiable bump, *J. Nonlinear Convex Anal.* **2**, 1–19.
140. J. M. Borwein, M. Fabian and P. D. Loewen (2002), The shape of the range of a C^1 smooth bump in infinite dimensions, *Israel J. Math.* **132**, 239–251.

141. J. M. Borwein and S. P. Fitzpatrick (1995), Weak* sequential compactness and bornological limit derivatives, *J. Convex Anal.* **2**, 59–68.
142. J. M. Borwein and S. P. Fitzpatrick (1995), Characterization of Clarke subgradients among one-dimensional multifunctions, in *Proc. Optimization Miniconference II*, edited by B. M. Glover and V. Jeyakumar, pp. 61–64, University of New South Wales, Sydney, Australia.
143. J. M. Borwein and S. P. Fitzpatrick (2001), Duality inequalities and sandwiched functions, *Nonlinear Anal.* **46**, 365–380.
144. J. M. Borwein, S. P. Fitzpatrick and R. Girgensohn (2003), Subdifferentials whose graphs are not norm×weak* closed, *Canad. Math. Bull.* **46**, 538–545.
145. J. M. Borwein, S. P. Fitzpatrick and J. R. Giles (1987), The differentiability of real functions on normed linear spaces using generalized subgradients, *J. Math. Anal. Appl.* **128**, 512–534.
146. J. M. Borwein and J. R. Giles (1987), The proximal normal formula in Banach spaces, *Trans. Amer. Math. Soc.* **302**, 371–381.
147. J. M. Borwein and A. D. Ioffe (1996), Proximal analysis in smooth spaces, *Set-Valued Anal.* **4**, 1–24.
148. J. M. Borwein and A. Jofré (1998), A nonconvex separation property in Banach spaces, *Math. Meth. Oper. Res.* **48**, 169–179.
149. J. M. Borwein and A. S. Lewis (2000), *Convex Analysis and Nonlinear Optimization: Theory and Examples*, Springer, New York.
150. J. M. Borwein, Y. Lucet and B. S. Mordukhovich (2000), Compactly epi-Lipschitzian convex sets and functions in normed spaces, *J. Convex Anal.* **7**, 375–393.
151. J. M. Borwein, B. S. Mordukhovich and Y. Shao (1999), On the equivalence of some basic principles of variational analysis, *J. Math. Anal. Appl.* **229**, 228–257.
152. J. M. Borwein, W. B. Moors and X. Wang (2001), Generalized subdifferentials: A Baire categorical approach, *Trans. Amer. Math. Soc.* **353**, 3875–3893.
153. J. M. Borwein and D. Noll (1994), Second order differentiability of convex functions in Banach spaces, *Trans. Amer. Math. Soc.* **342**, 43–81.
154. J. M. Borwein and D. Preiss (1987), A smooth variational principle with applications to subdifferentiability and differentiability of convex functions, *Trans. Amer. Math. Soc.* **303**, 517–527.
155. J. M. Borwein and H. M. Strójwas (1985), Tangential approximations, *Nonlinear Anal.* **9**, 1347–1366.
156. J. M. Borwein and H. M. Strójwas (1986), Proximal analysis and boundaries of closed sets in Banach spaces, I: Theory, *Canad. J. Math.* **38**, 431–452.
157. J. M. Borwein and H. M. Strójwas (1987), Proximal analysis and boundaries of closed sets in Banach spaces, II: Applications, *Canad. J. Math.* **39**, 428–472.
158. J. M. Borwein, J. S. Treiman and Q. J. Zhu (1998), Necessary conditions for constrained optimization problems with semicontinuous and continuous data, *Trans. Amer. Math. Soc.* **350**, 2409–2429.
159. J. M. Borwein, J. S. Treiman and Q. J. Zhu (1999), Partially smooth variational principles and applications, *Nonlinear Anal.* **35**, 1031–1059.

160. J. M. Borwein and Q. J. Zhu (1996), Viscosity solutions and viscosity subderivatives in smooth Banach spaces with applications to metric regularity, *SIAM J. Control Optim.* **34**, 1568–1591.
161. J. M. Borwein and Q. J. Zhu (1997), Variational analysis in nonreflexive spaces and applications to control problems with L^1 perturbations, *Nonlinear Anal.* **28**, 889–915.
162. J. M. Borwein and Q. J. Zhu (1998), Limiting convex examples for nonconvex subdifferential calculus, *J. Convex Anal.* **5**, 221–235.
163. J. M. Borwein and Q. J. Zhu (1999), A survey of subdifferential calculus with applications, *Nonlinear Anal.* **38**, 687–773.
164. J. M. Borwein and Q. J. Zhu (2005) *Techniques of Variational Analysis*, Springer, New York.
165. J. M. Borwein and D. M. Zhuang (1988), Verifiable necessary and sufficient conditions for openness and regularity of set-valued and single-valued maps, *J. Math. Anal. Appl.* **134**, 441–459.
166. P. Bosch, A. Jourani and R. Henrion (2004), Sufficient conditions for error bounds and applications, *Appl. Math. Optim.* **50**, 161–181.
167. G. Bouligand (1930), Sur les surfaces dépourvues de points hyperlimits, *Ann. Soc. Polon Math.* **9**, 32–41.
168. G. Bouligand (1932), *Introduction à la Géométrie Infinitésimale Directe*, Gauthier-Villars, Paris.
169. R. D. Bourgin (1983), *Geometric Aspects of Convex Sets with the Radon-Nicodým Property*, Springer, New York, 1993.
170. M. Bounkhel (2004), Scalarization of the normal Fréchet regularity of set-valued mappings, *New Zealand J. Math.* **33**, 1–18.
171. M. Bounkhel and A. Jofré (2004), Subdifferential stability of the distance function and its applications to nonconvex economics and equilibrium, *J. Nonlinear Convex Anal.* **5**, 331–347.
172. M. Bounkhel and L. Thibault (2002), On various notions of regularity of sets in nonsmooth analysis, *Nonlinear Anal.* **48**, 223–246.
173. M. Bounkhel and L. Thibault (2005), Further characterizations of regular sets in Hilbert spaces and their applications to nonconvex sweeping processes, *J. Nonlinear Convex Anal.*, to appear.
174. K. E. Brennan, S. L. Campbell and L. R. Pretzold (1989), *Numerical Solution of Initial Value Problems in Differential-Algebraic Equations*, North-Holland, New York.
175. A. Bressan (1990), Upper and lower semicontinuous differential inclusions: A unified approach, in *Nonlinear Controllability and Optimal Control*, edited by H. J. Sussmann, pp. 21-31, Marcel Dekker, New York.
176. O. A. Brezhneva and A. A. Tret'yakov (2003), Optimality conditions for degenerate extremum problems with equality constraints, *SIAM J. Control Optim.* **42**, 729–745.
177. H. Bréziz and W. Strauss (1973), Semilinear second-order elliptic equations in L^1, *J. Math. Soc. Japan* **25**, 565–590.
178. J. Brinkhuis and V. M. Tikhomirov (2005), *Optimization: Motivations and Applications*, Princeton University Press, Princeton, New Jersey.
179. A. Brøndsted and R. T. Rockafellar (1965), On the subdifferentiability of convex functions, *Proc. Amer. Math. Soc.* **16**, 605–611.
180. L. E. J. Brouwer (1910), Über eineindeutige, stetige transformationen von flächen in sich, *Math. Ann.* **69**, 176–180.

181. D. J. Brown (1991), Equilibrium analysis with nonconvex technologies, in *Handbook of Mathematical Economics*, Vol. 4, edited by W. Hildenbrand and H. Sonnenschein, North-Holland, Amsterdam, The Netherlands.
182. A. Bruckner (1994), *Differentiation of Real Functions*, American Mathematical Society, Providence, Rhode Island.
183. F. Bucci (1992), A Dirichlet boundary control problem for the strongly damped wave equation, *SIAM J. Control Optim.* **30**, 1092–1100.
184. B. M. Budak, E. M. Berkovich and E. N. Solovieva (1969), Difference approximations in optimal control problems, *SIAM J. Control* **7**, 18–31.
185. B. M. Budak and F. P. Vasiliev (1975), *Some Computational Aspects of Optimal Control Problems*, Moscow University Press, Moscow.
186. R. S. Burachik and A. N. Iusem (2006), *Set-Valued Mappings and Enlargements of Monotone Operators*, SIAM Publications, Philadelphia, Pennsylvania.
187. R. S. Burachik, A. N. Iusem and B. F. Svaiter (1997), Enlargement of monotone operators with applications to variational inequalities, *Set-Valued Anal.* **5**, 159–180.
188. J. V. Burke (1991), Calmness and exact penalization, *SIAM J. Control Optim.* **29**, 493–497.
189. J. V. Burke (1991), An exact penalization viewpoint of constrained optimization, *SIAM J. Control Optim.* **29**, 968–998.
190. J. V. Burke and S. Deng (2002), Weak sharp minima revisted, I: Basic theory, *Control Cybernet.* **31**, 439–469.
191. J. V. Burke and S. Deng (2005), Weak sharp minima revisted, II: Application to linear regularity and error bounds, *Math. Progr.*, to appear.
192. J. V. Burke and M. C. Ferris (1993), Weak sharp minima in mathematical programming, *SIAM J. Control Optim.* **31**, 1340–1359.
193. J. V. Burke, M. C. Ferris and M. Quian (1992), *J. Math. Anal. Appl.* **166**, 199–213.
194. J. V. Burke, A. S. Lewis and M. L. Overton (2000), Optimizing matrix stability, *Proc. Amer. Math. Soc.* **129**, 1635–1642.
195. J. V. Burke, A. S. Lewis and M. L. Overton (2001), Optimal stability and eigenvalue multiplicity, *Found. Comput. Math.* **1**, 205–225.
196. J. V. Burke, A. S. Lewis and M. L. Overton (2002), Approximating subdifferentials by random sampling of gradients, *Math. Oper. Res.* **27**, 567–584.
197. J. V. Burke, A. S. Lewis and M. L. Overton (2003), Robust stability and a criss-cross algorithm for pseudospectra, *IMA J. Numerical Anal.* **23**, 1–17.
198. J. V. Burke, A. S. Lewis and M. L. Overton (2004), Variational analysis of the abscissa mapping for polynomials via the Gauss-Lucas theorem, *J. Global Optim.* **28**, 259–268.
199. J. V. Burke, A. S. Lewis and M. L. Overton (2004), A robust gradient sampling algorithm for nonsmooth, nonconvex optimization, *SIAM J. Optim.* **15**, 751–779.
200. J. V. Burke, A. S. Lewis and M. L. Overton (2005), Variational analysis of functions of the roots of polynomials, *Math. Progr.*, to appear.
201. J. V. Burke and D. R. Luke (2003), Variational analysis applied to the problem of optical phase retrieval, *SIAM J. Control Optim.* **42**, 576–595.

202. J. V. Burke and M. L. Overton (1992), in *Nonsmooth Optimization Methods and Applications*, edited by F. Giannessi, pp. 19–29, Gordon and Breach, Philadelphia, Pennsylvania.
203. J. V. Burke and M. L. Overton (2000), Variational analysis of the abscissa mapping for polynomials, *SIAM J. Control Optim.* **39**, 1651–1676.
204. J. V. Burke and M. L. Overton (2001), Variational analysis of non-Lipschitz spectral functions, *Math. Progr.* **90**, 317–351.
205. J. V. Burke and P. Tseng (1996), A unified analysis of Hoffman's bound via Fenchel duality, *SIAM J. Optim.* **6**, 265–282.
206. J. A. Burns (2003), Nonlinear distributed parameter control systems with non-normal linearizations: Applications and approximations, *Front. Appl. Math.* **27**, 17–53.
207. M. Bustos (1994), Epsilon gradients pour functions localements Lipschitziennes et applications, *Numer. Funct. Anal. Optim.* **15**, 435–453.
208. A. G. Butkovsky (1963), Necessary and sufficient optimality conditions for impulse control systems, *Autom. Remote Control* **24**, 1056–1064.
209. A. G. Butkovsky (1969), *Distributed Control Systems*, Elsevier, New York.
210. A. G. Butkovsky, A. I. Egorov and K. A. Lurie (1968), Optimal control of distributed systems (a survey of Soviet publications), *SIAM J. Control* **6**, 437–476.
211. A. G. Butkovsky and A. J. Lerner (1960), Optimal control systems with distributed parameters, *Autom. Remote Control* **21**, 472–477.
212. G. Buttazzo and G. Dal Maso (1982), Γ-convergence and optimal control problems, *J. Optim. Theory Appl.* **38**, 385–407.
213. G. Buttazzo, M. Giaquinta and S. Hildebrandt (1998), *One-Dimensional Variational Problems: An Introduction*, Oxford University Press, New York.
214. F. Camilli and M. Falcone (1996), Approximation of optimal control problems with state constraints: Estimates and applications, in *Nonsmooth Analysis and Geometric Methos in Deterministic Optimal Control*, edited by B. S. Mordukhovich and H. J. Sussmann, pp. 23–58, Springer, New York.
215. P.-M. Cannarsa and G. Da Prato (1991), Second-order Hamilton-Jacobi equations in infinite dimensions, *SIAM J. Control Optim.* **29**, 474–492.
216. P.-M. Cannarsa and H. Frankowska (1991), Some characterizations of optimal trajectories in control theory, *SIAM J. Control Optim.* **29**, 1322–1347.
217. P.-M. Cannarsa and C. Sinestrari (2004), *Semiconvex Functions, Hamilton-Jacobi Equations, and Optimal Control*, Birkhäuser, Boston, Massuchusetts.
218. M. Cannon, C. Cullum and E. Polak (1970), *Theory of Optimal Control and Mathematical Programming*, McGraw-Hill, New York.
219. M. J. Cánovas, A. L. Dontchev, M. A. López and J. Parra (2005), Metric regularity of semi-infinite constraint systems, *Math. Progr.*, to appear.
220. I. Capuzzo Dolcetta (1983), On a discrete approximation of the Hamilton-Jacobi bequation of dynamic programming, *Appl. Math. Optim.* **10**, 367–377.
221. I. Capuzzo Dolcetta and M. Falcone (1989), Discrete dynamic programming and visocity solutions of the Bellman equation, *Ann. Inst. H. Poincaré: Analyse Non Linéaire* **6**, 161–184.
222. C. Carathéodory (1935), *Variationsrechung und Partielle Differentialgleichungen erster Ordnung*, B. G. Teubner, Leipzig and Berlin, 1935.

223. O. Cârjă (1984), The time optimal problem for boundary-distributed control systems, *Bull. Unione Mat. Italiana* **3-B**, 563–581.
224. D. A. Carlson, A. Haurie and A. Leizarowitz (1991), *Infinite Horizon Optimal Control*, Springer, Berlin.
225. E. Casas (1997), Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations, *SIAM J. Control Optim.* **35**, 1297–1327.
226. E. Casas, J.-P. Raymond and H. Zidani (2000), Pontryagin's principle for local solutions of control problems with mixed control-state constraints, *SIAM J. Control Optim.* **39**, 1182–1203.
227. E. Casas and J. Yong (1995), Maximum principle for state-constrained optimal control problems governed by quasilinear elliptic equations, *Diff. Integ. Eq.* **8**, 1–18.
228. C. Castaing, A. Salvadori and L. Thibault (2001), Functional evolution equations governed by nonconvex sweeping process, *J. Nonlinear Convex Anal.* **2**, 217–241.
229. C. Castaing and M. Valadier (1977), *Convex Analysis and Measurable Multifunctions*, Springer, New York.
230. E. Cavazzuti, M. Pappalardo and M. Passacantando (2002), Nash equilibria, variational inequalities, and dynamical systems, *J. Optim. Theory Appl.* **114**, 491–505.
231. A. Cellina (1988), On the set of solutions to Lipschitzian differential inclusions, *Diff. Integ. Eq.* **1**, 459–500.
232. A. Cellina (2003), The classical problem of the calculus of variations in the autonomous case: Relaxation and Lipschitzianity of solutions, *Trans. Amer. Math. Soc.* **356**, 415–426.
233. A. Cernea (2001), Some second-order necessary conditions for non-convex hyperbolic differential inclusions problems, *J. Math. Anal. Appl.* **253**, 616–639.
234. A. Cernea (2005), Second-order necessary conditions for differential-difference inclusion problems, *Nonlinear Anal.* **62**, 963–974.
235. L. Cesari (1983), *Optimization – Theory and Applications*, Springer, New York.
236. R. W. Chaney (1987), Second-order directional derivatives for nonsmooth functions, *J. Math. Anal. Appl.* **128**, 495–511.
237. A. V. Cherkaev (2000), *Variational Methods for Structural Optimization*, Springer, New York.
238. F. L. Chernousko (1994), *State Estimation for Dynamic Systems*, CRC Press, Boca Raton, Florida.
239. A. A. Chikrii (1997), *Conflict-Controlled Processes*, Kluwer, Dordrecht, The Netherlands.
240. I. Chryssochoos and R. B. Vinter (2003), Optimal control on manifolds: A dynamic programming approach, *J. Math. Anal. App.* **287**, 118–140.
241. E. N. Chukwu (1992), *Stability and Time-Optimal Control of Hereditary Systems*, Academic Press, Boston, Massachusetts.
242. M. Ciligot-Travain and S. Traore (2002), On subgradients of spectral functions, *J. Convex Anal.* **9**, 401–414.
243. F. H. Clarke (1973), *Necessary Conditions for Nonsmooth Problems in Optimal Control and the Calculus of Variations*, Ph.D. dissertation, Department of Mathematics, University of Washington, Seattle.

244. F. H. Clarke (1975), Generalized gradients and applications, *Trans. Amer. Math. Soc.* **205**, 247–262.
245. F. H. Clarke (1975), The Euler-Lagrange differential inclusion, *J. Diff. Eq.* **19**, 80–90.
246. F. H. Clarke (1975), Admissible relaxation in variational and control problems, *J. Math. Anal. Appl.* **51**, 557–576.
247. F. H. Clarke (1976), Optimal solutions to differential inclusions, *J. Optim. Theory Appl.* **19**, 469–478.
248. F. H. Clarke (1976), The generalized problem of Bolza, *SIAM J. Control Optim.* **14**, 682–699.
249. F. H. Clarke (1976), A new approach to Lagrange multipliers, *Math. Oper. Res.* **2**, 165–174.
250. F. H. Clarke (1976), The maximum principle under minimal hypotheses, *SIAM J. Control Optim.* **14**, 1078–1091.
251. F. H. Clarke (1976), Necessary conditions for a general control problem, in *Calculus of Variations and Control Theory*, edited by D. L. Russel, pp. 257–278, Academic Press, New York.
252. F. H. Clarke (1976), On the inverse function theorem, *Pacific J. Math.* **64**, 97–102.
253. F. H. Clarke (1977), Extremal arcs and extended Hamiltonian systems, *Trans. Amer. Math. Soc.* **231**, 349–367.
254. F. H. Clarke (1980), The Erdmann condition and Hamiltonian inclusions in optimal control and the calculus of variations, *Can. J. Math.* **32**, 494–509.
255. F. H. Clarke (1983), *Optimization and Nonsmooth Analysis*, Wiley, New York.
256. F. H. Clarke (1987), Hamiltonian analysis of the generalized problem of Bolza, *Trans. Amer. Math. Soc.* **301**, 385–400.
257. F. H. Clarke (1989), *Methods of Dynamic and Nonsmooth Optimization*, SIAM Publications, Philadelphia, Pennsylvania.
258. F. H. Clarke (1993), A decoupling principle in the calculus of variations, *J. Math. Anal. Appl.* **172**, 92–105.
259. F. H. Clarke (1993), An indirect method in the calculus of variations, *Trans. Amer. Math. Soc.* **336**, 655-673.
260. F. H. Clarke (2003), Necessary conditions in optimal control: A new approach, *Math. Progr.* **97**, 71–89.
261. F. H. Clarke (2005), Necessary conditions in dynamic optimization, *Mem. Amer. Math. Soc.* **173**, No. 816.
262. F. H. Clarke and Y. S. Ledyaev (1994), Mean value inequality in Hilbert spaces, *Trans. Amer. Math. Soc.* **344**, 307–324.
263. F. H. Clarke, Y. S. Ledyaev, E. D. Sontag and A. I. Subbotin (1997), Asymptotic controllability implies feedback stabilization, *IEEE Trans. Autom. Control* **42**, 1394–1407.
264. F. H. Clarke, Y. S. Ledyaev, R. J. Stern and P. R. Wolenski (1995), Qualitative properties of trajectories of control systems: A survey, *J. Dynam. Control Systems* **1**, 1–47.
265. F. H. Clarke, Y. S. Ledyaev, R. J. Stern and P. R. Wolenski (1998), *Nonsmooth Analysis and Control Theory*, Springer, New York.
266. F. H. Clarke, Y. S. Ledyaev and P. R. Wolenski (1995), Proximal analysis and minimization principles, *J. Math. Anal. Appl.* **196**, 722–735.

267. F. H. Clarke and P. D. Loewen (1986), The value function in optimal control: Sensitivity, controllability and time-optimality, *SIAM J. Control Optim.* **24**, 243–263.
268. F. H. Clarke and P. D. Loewen (1987), State constraints in optimal control: A case study in proximal normal analysis, *SIAM J. Control Optim.* **25**, 1440–1456.
269. F. H. Clarke and R. J. Stern (2003), State constrained feedback stabilization, *SIAM J. Control Optim.* **42**, 422–441.
270. F. H. Clarke, R. J. Stern and P. R. Wolenski (1993), Subgradient criteria for monotonicity and the Lipschitz condition, *Canad. J. Math.* **45**, 1167–1183.
271. F. H. Clarke, R. J. Stern and P. R. Wolenski (1995), Proximal smoothness and the lower-C^2 property, *J. Convex Anal.* **2**, 117–144.
272. F. H. Clarke and R. B. Vinter (1989), Applications of optimal multiprocesses, *SIAM J. Control Optim.* **25**, 1048–1071.
273. F. H. Clarke and R. B. Vinter (1989), Optimal multiprocesses, *SIAM J. Control Optim.* **25**, 1072–1091.
274. F. H. Clarke and G. G. Watkins (1986), Necessary conditions, controllability and the value function for differential-difference inclusions, *Nonlinear Anal.* **10**, 1155–1179.
275. F. H. Clarke and P. R. Wolenski (1991), The sensitivity of optimal control problems to time delay, *SIAM J. Control Optim.* **29**, 1176–1215.
276. F. H. Clarke and P. R. Wolenski (1996), Necessary conditions for functional differential inclusions, *Appl. Math. Optim.* **34**, 34–51.
277. G. Colombo and V. V. Goncharov (1999), The sweeping processes without convexity, *Set-Valued Anal.* **7**, 357–374.
278. G. Colombo and V. V. Goncharov (2001), Variational inequalities and regularity properties of closed sets in Hilbert spaces, *J. Convex Anal.* **8**, 197–221.
279. G. Colombo and A. Marigonda (2005), Differentiability properties for a class of nonconvex functions, *Calc. Var. Partial Diff. Eq.*, to appear.
280. G. Colombo and P. R. Wolenski (2004), Variational analysis for a class of minimal time functions in Hilbert spaces, *J. Convex Anal.* **11**, 335–361.
281. F. Colonius (1982), The maximum principle for relaxed hereditary differential systems with function space end conditions, *SIAM J. Control Optim.* **20**, 695–712.
282. R. Cominetti (1990), Metric regularity, tangent cones, and second-order optimality conditions, *Appl. Math. Optim.* **21**, 265–287.
283. R. Cominetti and R. Correa (1990), A generalized second order derivative in nonsmooth optimization, *SIAM J. Optim.* **28**, 789–809.
284. R. Conti (1968), Time-optimal solution of a linear evolution equation in Banach spaces, *J. Optim. Theory Appl.* **2**, 277–284.
285. B. Cornet (1981), Regular properties of tangent and normal cones, CEREMADE Publication 8130, Universiteé de Paris IX "Dauphine".
286. B. Cornet (1986), The second welfare theorem in nonconvex economies, CORE discussion paper No. 8630.
287. B. Cornet (1988), General equilibrium theorem and increasing returns: Presentation, *J. Math. Econ.* **17**, 103–118.

288. B. Cornet (1990), Marginal cost pricing and Pareto optimality, in *Essays in Honor of Edmond Malinvaud*, edited by P. Champsaur, Vol. 1, pp. 14–53, MIT Press, Cambridge, Massachusetts.
289. B. Cornet and M.-O. Czarnecki (1999), Smooth normal approximations of epi-Lipschitz subset in $I\!R^n$, *SIAM J. Control Optim.* **37**, 710–730.
290. B. Cornet and M.-O. Czarnecki (2001), Existence of generalized equilibria: Necessary and sufficient conditions, *Nonlinear Anal.* **44**, 555–574.
291. R. Correa, P. Gajardo and L. Thibault (2005), Subdifferential representation formula and subdifferential criteria for the behavior of nonsmooth functions, *Nonlinear Anal.*, to appear.
292. R. Correa, A. Jofré and L. Thibault (1994), Subdifferential monotonicity as characterization of convex functions, *Numer. Funct. Anal. Optim.* **15**, 1167–1183.
293. R. W. Cottle, F. Giannessi and J.-L. Lions, ed. (1980), *Variational Inequalities and Complementarity Problems*, Wiley, New York.
294. R. W. Cottle, J.-S. Pang and R. E. Stone (1992), *The Linear Complementarity Problem*, Academic Press, Boston, Massachusetts.
295. M. G. Crandall, L. C. Evans and P.-L. Lions (1984), Some properties of viscosity solutions, *Trans. Amer. Math. Soc.* **282**, 487–502.
296. M. G. Crandall, H. Ishii and P.-L. Lions (1992), User's guide to viscosity solutions of second-order partial differential equations, *Bull. Amer. Math. Soc.* **27**, 1–67.
297. M. G. Crandall and P.-L. Lions (1983), Viscosity solutions of Hamilton-Jacobi equations, *Trans. Amer. Math. Soc.* **277**, 1–42.
298. B. D. Craven (1994), Convergence of discrete approximations for constrained minimization, *J. Austral. Math. Soc.* **35**, 1–12.
299. B. D. Craven (1995), *Control and Optimization*, Chapman and Hall, London, UK.
300. B. D. Craven and D. V. Luu (1997), A method for establishing optimality conditions for a nonconvex vector-valued minimax problem, *J. Optim. Theory Appl.* **95**, 295–304.
301. G. P. Crespi, D. La Torre and M. Rocca (2003), Second-order mollified derivatives and optimization, *J. Nonlinear Convex Anal.* **3**, 437-454.
302. J. Cullum (1969), Discrete approximations to continuous optimal control problems, *SIAM J. Control* **7**, 32–49.
303. J. Cullum (1972), Finite-dimensional approximations of state constrained continuous optimal control problems, *SIAM J. Control* **10**, 649–670.
304. M.-O. Czarnecki and L. Rifford (2005), Approximation and regularization of Lipschitz functions: Convergence of the gradients, *Trans. Amer. Math. Soc.*, to appear.
305. L. Dai (1989), *Singular Control Systems*, Springer, Berlin.
306. P. Daniele and A. Maugeri (2001), On dynamical equilibrium problems and variational inequalities, in *Equilibrium Problems: Nonsmooth Optimization and Variational Inequality Models*, edited by F. Giannessi, A. Maugeri and P. Pardalos, pp. 59–69, Kluwer, Dordrecht, The Netherlands.
307. J. M. Danskin (1967), *The Theory of Min-Max and Its Application to Weapon Allocations Problems*, Springer, New York.
308. F. S. De Blasi, G. Pianigiani and A. A. Tolstonogov (2004), A Bogolyubov type theorem with a nonconvex constraint in Banach spaces, *SIAM J. Control Optim.* **43**, 466-476.

309. G. Debreu (1951), The coefficient of resource utilization, *Econometrica* **19**, 273–292.
310. G. Debreu (1959), *Theory of Values*, Yale University Press, New Haven, Connecticut.
311. G. Debreu (1970), Economies with a finite set of equilibria, *Econometrica* **38**, 387–392.
312. E. De Giorgi, A. Marino and M. Tosques (1980), Problemi di evoluzione in spazi metric e curve di massima pendenza, *Atti. Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur.* **68**, 180–187.
313. M. Degiovanni, A. Marino and M. Tosques (1985), Evolution equations with lack of convexity, *Nonlinear Anal.* **9**, 1401–1443.
314. K. Deimling (1992), *Multivalued Differential Equations*, De Gruyter, Berlin.
315. J. Demmel (1987), The condition number and the distance to the nearest ill-posed problem, *Numerische Math.* **51**, 251–289.
316. S. Dempe (2002), *Foundations of Bilevel Programming*, Kluwer, Dordrecht, The Netherlands.
317. S. Dempe (2003), Annotated bibliography on bilevel programming and mathematical programs with equilibrium constraints, *Optimization* **52**, 333–359.
318. V. F. Demyanov (2005), *Extremality Conditions and Calculus of Variations*, Vysshaya Shkola, Moscow.
319. V. F. Demyanov and V. N. Malozemov (1974), *Introduction to Minimax*, Wiley, New York.
320. V. F. Demyanov amd A. M. Rubinov (1980), On quasidifferentiable functionals, *Dokl. Akad. Nauk SSSR* **250**, 21–25.
321. V. F. Demyanov and A. M. Rubinov (1995), *Constructive Nonsmooth Analysis*, Peter Lang, Frankfurt, Germany.
322. V. F. Demyanov and A. M. Rubinov, eds. (1995), *Quasidifferentiability and Related Topics* (2000), Kluwer, Dordrecht, The Netherlands.
323. Z. Denkowski, S. Migórski and N. S. Papageorgiou (2004), *An Introduction to Nonlinear Analysis and Its Applications*, published in two volumes, Kluwer, Boston, Massachusetts.
324. D. Dentcheva and W. Römisch (2000), Differential stability of two-stage stochastic programs, *SIAM J. Optim.* **11**, 87–112.
325. D. Dentcheva and A. Ruszczyński (2004), Optimization with stochastic dominance constraints, *SIAM J. Optim.* **14**, 548-566.
326. A. De Simone and M. G. Graziano (2004), The pure theory of public goods: The case of many commodities, *J. Math. Econ.* **40**, 847–868.
327. E. N. Devdariani and Y. S. Ledyaev (1999), Maximum principle for implicit control systems, *Appl. Math. Optim.* **41**, 79–103.
328. R. Deville (1994), Stability of subdifferentials of nonconvex functions in Banach spaces, *Set-Valued Anal.* **2**, 141–157.
329. R. Deville and E. M. El Haddad (1996), The viscosity subdifferential of the sum of two functions in Banach spaces, I: First order case, *J. Convex Anal.* **3**, 295–308.
330. R. Deville, G. Godefroy and V. Zizler (1993), A smooth variational principle with applications to Hamilton-Jacobi equations in infinite dimensions, *J. Funct. Anal.* **111**, 197–212.
331. R. Deville, G. Godefroy and V. Zizler (1993), *Smoothness and Renorming in Banach Spaces*, Wiley, New York.

332. J. Diestel (1975), *Geometry of Banach Spaces–Selected Topics*, Springer, Berlin.
333. J. Diestel (1984) *Sequences and Series in Banach Spaces*, Springer, New York.
334. J. Diestel and J. J. Uhl, Jr. (1977), *Vector Measures*, American Mathematical Society, Providence, Rhode Island.
335. U. Dini (1878), *Fondamenti per la Teoria delle Funzioni di Variabili Reali*, Pisa, Italy.
336. A. V. Dmitruk (2002), An nonlocal Lyusternik estimate and its applications to control systems with sliding modes, in *Nonlinear Control Systems 2001*, edited by A. B. Kurzhanskii and A. L. Fradkov, Vol. 2, pp. 1061–1064, Elsevier, Exeter, UK.
337. A. V. Dmitruk, A. A. Milyutin and N. P. Osmolovskii (1980), Lyusternik's theorem and the theory of extrema, *Russian Math. Surveys* **35**, 11–51.
338. B. Doitchinov and V. Veliov (1993), Parameterizations of integrals of set-valued mappings and applications, *J. Math. Anal. Appl.* **179**, 483–499.
339. S. Dolecki (1980), A general theory of necessary optimality conditions, *J. Math. Anal. Appl.* **78**, 267–308.
340. S. Dolecki (1982), Tangency and differentiation: Some applications of convergence theory, *Ann. Mat. Pura Appl.* **130**, 223–255.
341. S. Dolecki and S. Rolewicz (1979), Exact penalties for local minima, *SIAM J. Control Optim.* **17**, 596–606.
342. T. Donchev (1997), Lower semicontinuous differential inclusions: One-sided Lipschitz approach, *Coll. Math.* **74**, 177–184.
343. T. Donchev (2005), Approximation of solution sets for optimal control problems, in *Large-Scale Scientific Computations*, edited by I. Lirkov, S. Margenov and J. Wasniewski, Lecture Notes Comp. Sci., Springer, Berlin.
344. T. Donchev and E. Farkhi (1998), Stability and Euler approximation of one-sided Lipschitz differential inclusions, *SIAM J. Control Optim.* **36**, 780–796.
345. T. Donchev and E. Farkhi (1999), Euler approximation of discontinuous one-sided Lipschitz convex differential inclusions, in *Calculus of Variations and Differential Equations*, edited by A. Ioffe, S. Reich and I. Shafrir, pp. 101–118, Chapman and Hall/CRC Press, Boca Raton, Florida.
346. T. Donchev and B. S. Mordukhovich (2005), Strong convergence of discrete approximations for one-sided Lipschitzian differential inclusions, preprint.
347. A. L. Dontchev (1981), Error estimates for s discrete approximation to constrained control problems, *SIAM J. Numer. Anal.* **18**, 500–514.
348. A. L. Dontchev (1983), *Perturbations, Approximations and Sensitivity Analysis of Optimal Control Systems*, Springer, Berlin.
349. A. L. Dontchev (1988), Equivalent perturbations and approximations in optimal control, in *International Series of Numerical Mathematics* **84**, pp. 43–54, Birkhäuser, Basel, Switzerland.
350. A. L. Dontchev (1995), Characterization of Lipschitz stability in optimization, in *Recent Developments in Well-Posed Variational Problems and Related Topics*, edited by R. Lucchetti and J. Revalski, pp. 95–116, Kluwer, Dordrecht, The Netherlands.

351. A. L. Dontchev (1995), Implicit function theorems for generalized equations, *Math. Progr.* **70**, 91–106.
352. A. L. Dontchev (1996), The Graves theorem revisited, *J. Convex Anal.* **3**, 45–54.
353. A. L. Dontchev (1996), Discrete approximations in optimal control, in *Nonsmooth Analysis and Geometric Methods in Deterministic Optimal Control*, edited by B. S. Mordukhovich and H. J. Sussmann, pp. 59–81, Springer, New York.
354. A. D. Dontchev and E. M. Farkhi (1989), Error estimates for discretized differential inclusions, *Computing* **41**, 349–358.
355. A. L. Dontchev and W. W. Hager (1993), Lipschitzian stability in nonlinear control and optimization, *SIAM J. Control Optim.* **31**, 569–603.
356. A. L. Dontchev and W. W. Hager (1994), Implicit functions, Lipschitz maps and stability in optimization, *Math. Oper. Res.* **19**, 753–768.
357. A. L. Dontchev and W. W. Hager (1998), A new approach to Lipschitz continuity in state constrained optimal control, *Syst. Control Lett.* **35**, 137–143.
358. A. L. Dontchev and W. W. Hager (2001), The Euler approximation in state constrained optimal control, *Math. Comp.* **70**, 173–203.
359. A. L. Dontchev and F. Lempio (1992), Difference methods for differential inclusions: A survey, *SIAM Rev.* **34**, 263–294.
360. A. L. Dontchev and A. S. Lewis (2005), Perturbations and metric regularity, *Set-Valued Anal.*, to appear.
361. A. L. Dontchev, A. S. Lewis and R. T. Rockafellar (2003), The radius of metric regularity, *Trans. Amer. Math. Soc.* **355**, 493–517.
362. A. L. Dontchev and B. S. Mordukhovich (1983), Relaxation and well-posedness of nonlinear optimal processes, *Syst. Control Lett.* **3**, 177–179.
363. A. L. Dontchev, M. Quincampoix and N. Zlateva (2005), Aubin criterion for metric regularity, *J. Convex Anal.*, to appear.
364. A. L. Dontchev and R. T. Rockafellar (1996), Characterizations of strong regularity for variational inequalities over polyhedral convex sets, *SIAM J. Optim.* **7**, 1087–1105.
365. A. L. Dontchev and R. T. Rockafellar (2001), Ample parameterization of variational inclusions, *SIAM J. Optim.* **12**, 170–187.
366. A. L. Dontchev and R. T. Rockafellar (2004), Regularity and conditioning of solution mappings in variational analysis, *Set-Valued Anal.* **12**, 79–109.
367. A. L. Dontchev and T. Zolezzi (1993), *Well-Posed Optimization Problems*, Springer, New York.
368. A. Douglis (1961), The continuous dependence of generalized solutions of non-linear partial differential equations upon initial data, *Comm. Pure Appl. Math.* **14**, 267–284.
369. A. Y. Dubovitskii and A. A. Milyutin (1963), Extremum problems in the presence of constraints, *Dokl. Akad. Nauk SSSR* **149**, 759–762.
370. A. Y. Dubovitskii and A. A. Milyutin (1965), Extremum problems in the presence of restrictions, *USSR Comput. Maths. Math. Phys.* **5**, 1–80.
371. N. Dunford and J. T. Schwartz (1958), *Linear Operators, I: General Theory*, Interscience, New York.
372. J. C. Dunn (1995), Second order optimality conditions in sets of L^∞ functions with range in a polyhedron, *SIAM J. Control Optim.* **33**, 1603–1635.

373. M. Dür, R. Horst and M. Locatelli (1998), Necessary and sufficient conditions for convex maximization revisited, *J. Math. Anal. Appl.* **217**, 637–649.
374. J. Dutta (2005), Necessary optimality conditions and saddle points for approximate optimization in Banach spaces, *TOP: Spanish J. Oper. Res.*, **13**, 127–143.
375. J. Dutta (2005), Optimality conditions for maximizing a locally Lipschitz function, *Optimization* **54**, 377–389.
376. J. Dutta (2005), Generalized derivatives and nonsmooth optimization: A finite-dimensional tour, *TOP: Spanish J. Oper. Res.*, **13**.
377. J. Dutta and S. Dempe (2005), Bilevel programming with convex lower level problems, in *Optimization with Multivalued Mappings: Theory, Applications and Algorithms*, edited by S. Dempe and V. V. Kalashnikov, Springer, Berlin.
378. J. Dutta and C. Tammer (2005), Lagrangian conditions for vector optimization on Banach spaces, *Math. Meth. Oper. Res.*, to appear.
379. A. N. Dyukalov (1983), *Problems of Applied Mathematical Economics*, Nauka, Moscow.
380. Z. Dzalilov, A. F. Ivanov and A. M. Rubinov (2001), Difference inclusions with delay of economic growth, *Dynam. Syst. Appl.* **10**, 283–293.
381. A. Eberhard (2001), Prox-regularity and subjects, in *Optimization and Related Topics*, edited by A. Rubinov and B. Glover, *Applied Optimization* **47**, pp. 237–313, Kluwer, Dordrecht, The Netherlands.
382. A. Eberhard and M. Nyblom (1998), Jets, generalized convexity, proximal normality and differences of functions, *Nonlinear Anal.* **34**, 319–360.
383. A. Eberhard, M. Nyblom and D. Ralph (1998), Applying generalized convexity notions to jets, in *Generalized Convexity, Generalized Monotonicity: Recent Results*, edited by J.-P. Crouzeix, J. E. Martinez-Legaz and M. Volle, pp. 111–157, Kluwer, Dordrecht, The Netherlands.
384. A. Eberhard and C. E. M. Pearce (2005), A comparison of two approaches to second-order subdifferentiability concepts with application to optimality conditions, in *Optimization and Control with Applications*, edited by L. Qi, K. L. Teo and X. Yang, pp. 36–100, Springer, Berlin.
385. A. Eberhard, C. E. M. Pearce and D. Ralph (1998), A comparative study of jets, graphical derivatives and coderivatives, RMIT Research Report, No. 8, Melbourne, Australia.
386. A. Eberhard, R. Sivakumaran and R. Wenczel (2005), On the variational behavior of the subhessians of the Lasry-Lions envelope, *J. Convex Anal.*, to appear.
387. A. Eberhard and R. Wenczel (2004), Optimality conditions in nonsmooth analysis, RMIT Research Report, No. 3, Melbourne, Australia.
388. C. Eckart and G. Young (1936), The approximation of one matrix by another of lower rank, *Psychometrika* **1**, 211–218.
389. J. E. Edmond and L. Thibault (2002), Inclusions and integration of subdifferentials, *J. Nonlinear Convex Anal.* **3**, 411–434.
390. J. E. Edmond and L. Thibault (2005), Relaxation of an optimal control problem involving a perturbed sweeping process, *Math. Progr.*, to appear.
391. A. I. Egorov (1967), Necessary optimality conditions for distributed parameter systems, *SIAM J. Control* **5**, 352–408.
392. A. I. Egorov (1978), *Optimal Control of Heat and Diffusion Processes*, Nauka, Moscow.

393. Y. V. Egorov (1963), Certain problems in optimal control theory, *USSR Comput. Maths. Math. Phys.* **3**, 1209–1232.
394. Y. V. Egorov (1964), Some necessary conditions for optimality in Banach spaces, *Math. Sbornik* **64**, 79–101.
395. K. J. Eisenhart (2003), *Multiobjective Optimal Control Problems with Endpoint and State Constraints*, Ph.D. dissertation, Departments of Mathematics, Western Michigan University, Kalamazoo, Michigan.
396. I. Ekeland (1972), Sur les problémes variationnels, *C. R. Acad. Sci. Paris* **275**, 1057–1059.
397. I. Ekeland (1974), On the variational principle, *J. Math. Anal. Appl.* **47**, 324–353.
398. I. Ekeland (1974), Une estimation a priori en programmation non convexe, *C. R. Acad. Sci. Paris* **277**, 149–151.
399. I. Ekeland (1979), Nonconvex minimization problems, *Bull. Amer. Math. Soc.* **1**, 432–467.
400. I. Ekeland and G. Lebourg (1976), Generic Fréchet differentiability and perturbed optimization problems in Banach spaces, *Trans. Amer. Math. Soc.* **224**, 193–216.
401. I. Ekeland and R. Temam (1976), *Convex Analysis and Variational Problems*, North-Holland, Amsterdam, The Netherlands.
402. B. El Abdouni and L. Thibault (1992), Lagrange multipliers for Pareto nonsmooth programming problems in Banach spaces, *Optimization* **26**, 277–285.
403. L. E. Élsgolts (1955), *Qualitative Methods in Mathematical Analysis* (1955). GITTL, Moscow.
404. L. E. Élsgolts and S. B. Norkin (1973), *Introduction to the Theory and Applications of Differential Equations with Deviating Arguments*, Academic Press, New York.
405. M. Epelman and R. M. Freund (2002), A new condition measure, preconditioners, and relations between different measures of conditioning for conic linear systems, *SIAM J. Optim.* **13**, 627–655.
406. I. I. Eremin (1966), The penalty method in convex programming, *Soviet Math. Dokl.* **8**, 458–462.
407. Y. M. Ermoliev, V. P. Gulenko and T. I. Tzarenko (1978), *Finite Difference Method in Optimal Control Problems*, Naukova Dumka, Kiev.
408. Y. M. Ermoliev, V. I. Norkin and R. J-B. Wets (1995), The minimization of semicontinuous functions: Mollifier subgradients, *SIAM J. Control Optim.* **33**, 149–167.
409. E. Ernst and M. Théra (2005), A converse to the Eidelgeit theorem in real Hilbert spaces, *Bull. Sci. Math.* **129**, 381–397.
410. E. Ernst and M. Théra (2005), Necessary and sufficient conditions for the existence of a global maximum for convex functions in reflexive Banach spaces, *J. Convex Anal.*, to appear.
411. L. Euler (1774), *Methodus Inveniendi Curvas Lineas Maximi Minimive Proprietate Gaudentes Sive Solution Problematis Viso Isoperimetrici Latissimo Sensu Accepti*, Lausanne; reprinted in *Opera Omnia*, Ser. 1, Vol. 24, 1952.
412. U. G. Evtushenko (1982), *Methods of Solving Extremal Problems and Their Application in Optimization Systems*, Nauka, Moscow.
413. M. Fabian (1986), Subdifferentials, local ε-supports and Asplund spaces, *J. London Math. Soc.* **34**, 568–576.

414. M. Fabian (1988), On classes of subdifferentiability spaces of Ioffe, *Nonlinear Anal.* **12**, 568–576.
415. M. Fabian (1989), Subdifferentiability and trustworthiness in the light of a new variational principle of Borwein and Preiss, *Acta Univ. Carolina, Ser. Math. Phys.* **30**, 51–56.
416. M. Fabian (1997), *Gâteaux Differentiability of Convex Functions and Topology. Weak Asplund Spaces*, Wiley, New York.
417. M. Fabian, P. Hájek and J. Vanderwerff (1996), On smooth variational principles in Banach spaces, *J. Math. Anal. Appl.* **197**, 153–172.
418. M. Fabian, P. D. Loewen and B. S. Mordukhovich (2005), Subdifferential calculus in Asplund generated spaces, *J. Math. Anal. Appl.*, to appear.
419. M. Fabian and B. S. Mordukhovich (1998), Smooth variational principles and characterizations of Asplund spaces, *Set-Valued Anal.* **6**, 381–406.
420. M. Fabian and B. S. Mordukhovich (1999), Separable reduction and supporting properties of Fréchet-like normals in Banach spaces, *Canad. J. Math.* **51**, 26–48.
421. M. Fabian and B. S. Mordukhovich (2002), Separable reduction and extremal principles in variational analysis, *Nonlinear Anal.* **49**, 265–292.
422. M. Fabian and B. S. Mordukhovich (2003), Sequential normal compactness versus topological normal compactness in variational analysis, *Nonlinear Anal.* **54**, 1057–1067.
423. M. Fabian and N. V. Zhivkov (1985). A characterization of Asplund spaces with help of local ε-supports of Ekeland and Lebourg, *C. R. Acad. Bulgare Sci.* **38**, 671–674.
424. F. Facchinei and J.-S. Pang (2003), *Finite-Dimensional Variational Inequalities and Complementary Problems*, published in two volumes, Springer, New York.
425. M. Falcone (1994), Discrete time high-order schemes for viscosity solutions of the Hamilton-Jacobi equation, *Numer. Math.* **67**, 315–344.
426. L.-T. Fan and C.-S. Wang (1964), *The Discrete Maximum Principle: A Study of Multistage Systems Optimization*, Wiley, New York.
427. H. O. Fattorini (1964), Time-optimal control of solutions of operational differential equations, *SIAM J. Control* **2**, 54–59.
428. H. O. Fattorini (1968), Boundary control problems, *SIAM J. Control* **6**, 349–385.
429. H. O. Fattorini (1987), A unified theory of necessary conditions for nonlinear nonconvex control systems, *Appl. Math. Optim.* **15**, 141–185.
430. H. O. Fattorini (1996), Optimal control with state constraints for semilinear distributed parameter systems, *J. Optim. Theory Appl.* **88**, 25–59.
431. H. O. Fattorini (1997), Nonlinear infinite dimensional optimal control problems with state constraints and unbounded control sets, *Rend. Instit. Mat. Univ. Trieste* **28**, 127–146.
432. H. O. Fattorini (1999), *Infinite Dimensional Optimization and Control Theory*, Cambridge University Press, Cambridge, Massachusetts.
433. H. O. Fattorini (2001), The maximum principle for control systems described by linear parabolic equations, *J. Math. Anal. Appl.* **259**, 630–651.
434. H. O. Fattorini and H. Frankowska (1991), Necessary conditions for infinite dimensional control problems, *Math. Contr. Signal Syst.* **4**, 41–67.

435. H. O. Fattorini and T. Murphy (1994), Optimal problems for nonlinear parabolic boundary control problems: Dirichlet boundary conditions, *Diff. Integ. Eq.* **7**, 1367–1388.
436. H. O. Fattorini and T. Murphy (1994), Optimal problems for nonlinear parabolic boundary control systems, *SIAM J. Control Optim.* **32**, 1577-1596.
437. H. Federer (1959), Curvature measures, *Trans. Amer. Math. Soc.* **93**, 418–491.
438. R. P. Fedorenko (1970), Maximum principle for differential inclusions, *USSR Comput. Maths. Math. Phys.* **10**, 57–68.
439. R. P. Fedorenko (1971), Maximum principle for differential inclusions (necessity), *USSR Comput. Maths. Math. Phys.* **11**, 885–893.
440. A. A. Feldbaum (1953), Optimal process in systems of automatic control, *Autom. Telemekh* **14**, No. 6.
441. H. Fenchel (1951), *Convex Cones, Sets and Functions*, Lecture Notes, Princeton University, Princeton, New Jersey.
442. P. de Fermat (1636), Letters to M. Mersenne and J. Roberval, Toulouse, France.
443. M. M. A. Ferreira, F. A. C. C. Fontes and R. B. Vinter (1999), Nondegenerate necessary conditions for nonconvex optimal control problems with state constraints, *J. Math. Anal. Appl.* **233**, 116–129.
444. M. M. A. Ferreira and R. B. Vinter (1994), When is the maximum principle for state-constrained problems degenerate?, *J. Math. Anal. Appl.* **187**, 432–467.
445. M. C. Ferris and O. L. Mangasarian (1993), Error bounds and strong upper semicontinuity for monotone affine variational inequalities, *Ann. Oper. Res.* **47**, 293–305.
446. M. C. Ferris and J.-S. Pang (1997), Engineering and economic applications of complementarity problems, *SIAM Rev.* **39**, 669–713.
447. A. V. Fiacco (1983), *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, Academic Press, New York.
448. A. V. Fiacco and G. P. McCormick (1968), *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York.
449. A. F. Filippov (1962), On certain questions in the theory of optimal control, *SIAM J. Control* **1**, 76–84.
450. A. F. Filippov (1988), *Differential Equations with Discontinuous Right-Hand Sides*, Kluwer, Dordrecht, The Netherlands.
451. S. P. Fitzpatrick (1980), Metric projections and the differentiability of distance functions, *Bull. Austral. Math. Soc.* **22**, 291–312.
452. S. D. Flåm (2005), Upward slopes and inf-convolutions, *Math. Oper. Res.*, to appear.
453. S. D. Flåm and A. Jourani (2000), Prices and Pareto optima, Research Report.
454. M. L. Flegel (2005), *Constraint Qualifications and Stationarity Concepts for Mathematical Programs with Equilibrium Constraints*, Ph.D. Dissertation, Facult. Math., Univ. Würzburg, Germany.
455. M. L. Flegel and C. Kanzow (2003), A Fritz John approach to first order optimality conditions for mathematical programs with equilibrium constraints, *Optimization* **52**, 277–296.

456. M. L. Flegel and C. Kanzow (2005), On M-stationarity points for mathematical programs with equilibrium constraints, *J. Math. Anal. Appl.* **310**, 286–302.
457. M. L. Flegel, C. Kanzow and J. V. Outrata (2005), Optimality conditions for disjunctive programs with equilibrium constraints, preprint.
458. W. H. Fleming and H. M. Soner (1993), *Controlled Markov Processes and Viscosity Solutions*, Springer, New York.
459. M. Florenzano (2003), *General Equilibrium Analysis: Existence and Optimality Properties of Equilibria*, Kluwer, Dordrecht, The Netherlands.
460. M. Florenzano, P. Gourdel and A. Jofré (2003), Supporting weakly Pareto optimal allocations in infinite dimensional nonconvex economies, preprint.
461. F. Flores-Bazán (1997), On minima of the difference functions, *J. Optim. Theory Appl.* **93**, 525–531.
462. F. Flores-Bazán and W. Oettli (2001), Simplified optimality conditions for minimizing the difference of vector-valued convex functions, *J. Optim. Theory Appl.* **108**, 571–586.
463. D. Foley (1967), Resource allocation and the public sector, *Yale Econ. Essays* **7**, 43–98.
464. H. Frankowska (1985), Necessary conditions for the Bolza problem, *Math. Oper. Res.* **10**, 361–366.
465. H. Frankowska (1987), The maximum principle for an optimal control to a differential inclusion with endpoint constraints, *SIAM J. Control Optim.* **25**, 145–157.
466. H. Frankowska (1987), An open mapping principle for set-valued maps, *J. Math. Anal. Appl.* **127**, 172–180.
467. H. Frankowska (1989), Higher order inverse function theorems, *Ann. Inst. H. Poincaré: Analyse Non Linéaire* **6**, 283–304.
468. H. Frankowska (1989), Contingent cones to reachable sets of control systems, *SIAM J. Control Optim.* **27**, 170–198.
469. H. Frankowska (1990), Some inverse mappings theorems, *Ann. Inst. H. Poincaré: Analyse Non Linéaire* **7**, 183–234.
470. H. Frankowska (1990), A priori estimates for operational differential inclusions, *J. Diff. Eq.* **84**, 100–128.
471. H. Frankowska (1993), Lower semicontinuous solutions of Hamilton-Jacobi-Bellman equations, *SIAM J. Control Optim.* **31**, 257–272.
472. H. Frankowska (2005), Optimal synthesis via superdifferentials of value functions, *Control Cybernet.* **34** (2005), No. 3.
473. M. Fréchet (1911), Sur la notion de différentielle, *C. R. Acad. Sci. Paris* **152**, 845–847.
474. R. A. Freeman and P. V. Kokotović (1996), *Robust Nonlinear Control Design: State-Space and Lyapunov Techniques*, Birkhäuser, Boston, Massachusetts.
475. R. M. Freund and J. R. Vera (1999), Some characterizations and properties of the 'distance to ill-posedness' and the condition measure of a conic linear system, *Math. Progr.* **86**, 225–260.
476. A. Friedman (1964), Optimal control for hereditary processes, *J. Math. Anal. Appl.* **15**, 396–414.
477. A. Friedman (1967), Optimal control in Banach spaces, *J. Math. Anal. Appl.* **19**, 35–55.

478. A. Friedman (1982), *Variational Principles and Free-Boundary Problems*, Wiley, New York.
479. A. Friedman (1987), Optimal control for parabolic variational inequalities, *SIAM J. Control Optim.* **25**, 482–497.
480. M. Fukushima and J.-S. Pang (2005), Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games, *Comput. Management Sci.* **1**, 21–56.
481. A. V. Fursikov (2000), *Optimal Control of Distributed Systems. Theory and Applications*, American Mathematical Society, Prividence, Rhode Island.
482. P. Fusek, D. Klatte and B. Kummer (2002), Examples and couterexamples in Lipschitz analysis, *Control Cybernet.* **31**, 471–492.
483. R. Gabasov and S. V. Churakova (1968), Necessary optimality conditions in time-lag systems, *Autom. Remote Control* **29**, 37–54.
484. R. Gabasov and F. M. Kirillova (1966), On the extension of the maximum principle by L. S. Pontryagin to discrete systems, *Autom. Remote Control* **27**, 1878–1882.
485. R. Gabasov and F. M. Kirillova (1974), *Maximum Principle in the Theory of Optimal Processes*, Nauka i Tekhnika, Minsk, Belarus.
486. R. Gabasov and F. M. Kirillova (1976), *Qualitative Theory of Optimal Processes*, Marcel Dekker, New York.
487. R. Gabasov and F. M. Kirillova (1977), Methods of optimal control, *J. Soviet Math.* **7**, 805–849.
488. R. Gabasov, F. M. Kirillova and B. S. Mordukhovich (1983), The ε-maximum principle for suboptimal controls, *Soviet Math. Dokl.* **27**, 95–99.
489. N. Gadhi (2005), Optimality conditions for a d. c. set-valued problem via the extremal principle, in *Optimization with Multivalued Mappings: Theory, Applications and Algorithms*, edited by S. Dempe and V. V. Kalashnikov, Springer, Berlin.
490. G. N. Galbraith (2001), Extended Hamilton-Jacobi characterization of value functions in optimal control, *SIAM J. Control Optim.* **39**, 281–305.
491. G. N. Galbraith (2002), Cosmically Lipschitz set-valued mappings, *Set-Valued Anal.* **10**, 331–360.
492. G. N. Galbraith (2004), Solution regularity in optimal control via subgradient analysis of the value function, *Set-Valued Anal.* **12**, 111–126.
493. G. N. Galbraith and R. B. Vinter (2003), Lipschitz continuity of optimal controls for state constrained problems, *SIAM J. Control Optim.* **42**, 1727–1744.
494. R. V. Gamkrelidze (1957), On the theory of optimal processes in linear systems, *Dokl. Akad. Nauk SSSR* **116**, 9–11.
495. R. V. Gamkrelidze (1962), On sliding optimal regimes, *Soviet Math. Dokl.* **3**, 559–561.
496. R. V. Gamkrelidze (1965), On some extremal problem in the theory of differential equations with applications to the theory of optimal control, *SIAM J. Control* **3**, 106–128.
497. R. V. Gamkrelidze (1978), *Principles of Optimal Control Theory*, Plenum Press, New York.
498. R. V. Gamkrelidze (1999), Discovery of the maximum principle, *J. Dynam. Control Syst.* **5**, 437–451.
499. J. Gauvin and F. Dubeau (1982), Differential properties of the marginal function in mathematical programming, *Math. Progr. Study* **19**, 101–119.

500. V. S. Gavrilov and M. I. Sumin (2004), Parametric optimization of nonlinear Goursat-Darboux systems with state constraints, *Comput. Maths. Math. Phys.* **44**, 949–968.
501. M. Geoffroy and M. Lassonde (2000), On a convergence of lower semicontinuous functions linked with the graph convergence of their subdifferentials, in *Constructive, Experimental and Nonlinear Analysis*, edited by M. Théra, Canad. Math. Soc. Conf. Proc. **27**, 93–109.
502. P. G. Georgiev and N. P. Zlateva (1996), Second-order subdifferential of $C^{1,1}$ functions and optimality conditions, *Set-Valued Anal.* **4**, 101–117.
503. W. Geremew, B. S. Mordukhovich and N. M. Nam (2005), Coderivative analysis of metric regularity for constraint and variational systems, preprint.
504. F. Giannessi (2005), *Constrained Optimization and Image Space Analysis*, published in two volumes Springer, Berlin.
505. J. R. Giles (1982), On the characterization of Asplund spaces, *J. Aust. Math. Soc.* **32**, 134–144.
506. B. Ginsburg and A. D. Ioffe (1996), The maximum principle in optimal control of systems governed by semilinear equations, in *Nonsmooth Analysis and Geometric Methos in Deterministic Optimal Control*, edited by B. S. Mordukhovich and H. J. Sussmann, pp. 81–110, Springer, New York.
507. I. V. Girsanov (1972), *Lectures on Mathematical Theory of Extremum Problems*, Springer, Berlin.
508. B. M. Glover and B. D. Craven (1994), A Fritz John optimality condition using the approximate subdifferential, *J. Optim. Theory Appl.* **82**, 253–265.
509. B. M. Glover, B. D. Craven and S. D. Flåm (1993), A generalized Karush-Kuhn-Tucker optimality condition without constraint qualification using the approximate subdifferential, *Numer. Func. Anal. Optim.* **14**, 333–353.
510. B. M. Glover and D. Ralph (1994), First order approximations to nonsmooth mappings with applications to metric regularity, *Numer. Func. Anal. Optim.* **15**, 599–620.
511. R. Goebel (2004), Regularity of the optimal feedback and the value function in convex problems of optimal control, *Set-Valued Anal.* **12**, 127–145.
512. B. Gollan (1984), On the marginal function in nonlinear programming, *Math. Oper. Res.* **9**, 208–221.
513. S. B. Gorelik and B. S. Mordukhovich (1985), High-order necessary optimality conditions in neutral systems with some applications, *Izv. Akad Nauk BSSR. Ser. Fiz.-Mat. Nauk*, No. 6, 111–112; Depon. VINITI # 1154-85, Moscow.
514. S. B. Gorelik and B. S. Mordukhovich (1986), Legendre-Clebsch and Kelley conditions for nonlinear systems of neutral type with applications to existence of optimal control, *Izv. Akad Nauk BSSR. Ser. Fiz.-Mat. Nauk*, No. 65, 119; Depon. VINITI # 4057-85, Moscow.
515. V. V. Gorokhovik (1986), Epsilon-quasidifferentiability of real-valued functions and optimality conditions in extremal problems, *Math. Progr. Study* **29**, 203–218.
516. V. V. Gorokhovik (1990), *Convex and Nonsmooth Problems of Vector Optimization*, Nauka i Tekhnika, Minsk, Belarus.
517. V. V. Gorokhovik and P. P. Zabreiko (1998), Fréchet differentiability of multimappings, in *Nonlinear Analysis and Applications*, edited by I. V. Gaishun, pp. 34–49, Minsk, Belarus.

518. M. G. Govil and A. Mehra (2005), Epsilon-optimality for nonsmooth programming on a Hilbert space, in *Generalized Convexity, Generalized Monotonicity and Applications*, edited by A. Eberhard et al., Nonconvex Optimization and Its Applications, pp. 287–298, Springer, New York.
519. M. S. Gowda and R. Sznajder (1996), On the Lipshitzian properties of polyhedral multifunctions, *Math. Progr.* **74**, 267–278.
520. G. Grammel (2003), Towards fully discretized differential inclusions, *Set-Valued Anal.* **11**, 1–8.
521. K. A. Grasse (1985), A higher order sufficient condition for surjectivity, *Nonlinear Anal.* **9**, 87–96.
522. L. M. Graves (1950), Some mapping theorems, *Duke Math. J.* **17**, 111–114.
523. J. Guddat, F. Guerra Vasquez and H. T. Jongen (1990), *Parametric Optimization: Singularities, Pathfollowing and Jumps*, B. G. Tebner and John Wiley & Sons, Stuttgart and Chichester.
524. R. Guesnerie (1975), Pareto optimality in non-convex economies, *Econometrica* **43**, 1–29.
525. S. Guillaume (2000), Subdifferential evolution inclusions in nonconvex analysis, *Positivity* **4**, 357–395.
526. P. Gupta, S. Shiraishi and K. Yokoyama (2005), Epsilon-optimality without constraint qualifications for multiobjective fractional programs, *J. Nonlinear Convex Anal.* **6**, 347–357.
527. V. I. Gurman (1997), *The Extension Principle in Control Problems*, 2nd edition, Nauka, Moscow.
528. M. L. Gusakova (1972), Necessary optimality conditions for systems that are not solved with respect to the derivative, *Diff. Eq.* **8**, 1498–1500.
529. M. L. Gusakova (1974), On necessary optimality conditions for systems of neutral type, *Diff. Eq.* **810**, 1894–1897.
530. T. X. D. Ha (2003), The Ekeland variational principle for set-valued maps involving coderivatives, *J. Math. Anal. Appl.* **286**, 509–523.
531. T. X. D. Ha (2005), Some variants of the Ekeland variational principle for a set-valued map, *J. Optim. Theory Appl.* **124**, 187–206.
532. T. X. D. Ha (2005), Lagrange multipliers for set-valued problems associated with coderivatives, *J. Math. Anal. Appl.*, to appear.
533. A. Habte (2005), *Applications of Variational Analysis to Welfare Economics*, Ph.D. dissertation, Department of Mathematics, Wayne State University, Detroit, Michigan.
534. N. Hadjisavvas, S. Komlosi and S. Schaible, eds. (2005), *Handbook of Generalized Convexity and Generalized Monotonicity*, Springer, Berlin.
535. W. W. Hager (1976), Rates of convergence for discrete approximations to unconstrained control problems, *SIAM J. Numer. Anal.* **13**, 1–14.
536. W. W. Hager (1979), Lipschitz continuity for constrained processes, *SIAM J. Control Optim.* **17**, 321–332.
537. A. Halanay (1968), Optimal control for systems with time lag, *SIAM J. Control* **6**, 215–234.
538. J. Hale (1977), *Theory of Functional Differential Equations*, Springer, New York.
539. H. Halkin (1964), On the necessary condition for the optimal control of nonlinear systems, *J. Analyse Math.* **12**, 1–82.

540. H. Halkin (1964), Optimal control for systems by nonlinear difference equations, in *Advances in Control Systems*, Vol. 1, pp. 173–196, Academic Press, New York.
541. H. Halkin (1970), A satisfactory treatment of equality and operator constraints in Dubovitskii-Milyutin optimization formalism, *J. Optim. Theory Appl.* **6**, 138–149.
542. H. Halkin (1972), Extremal properties of biconvex contingent equations, in *Ordinary Differential Equations*, edited by L. Weiss, pp. 109–119, Academic Press, New York.
543. H. Halkin (1974), Implicit functions and optimization problems without continuous differentiability of the data, *SIAM J. Control* **12**, 229–236.
544. H. Halkin (1976), Mathematical programming without differentiability, in *Calculus of Variations and Control Theory*, edited by D. L. Russel, pp. 279–287, Academic Press, New York.
545. H. Halkin (1978), Necessary conditions for optimal control problems with differentiable and nondifferentiable data, in *Mathematical Control Theory*, Lecture Notes Math. **680**, pp. 77–118, Springer, Berlin.
546. A. H. Hamel (2001), An epsilon-Lagrange multiplier rule for a mathematical programming problem on a Banach space, *Optimization* **49**, 137–150.
547. A. H. Hamel (2003), Phelps' lemma, Daneš' drop theorem and Ekeland's variational principle in locally convex spaces, *Proc. Amer. Math. Soc.* **131**, 3025–3038.
548. W. R. Hamilton (1834), On a general method employed in dynamics, by which the study of the motions of all free systems of attracting or repelling points is reduced to the search and differentiation of one central solution or characteristic function, *Philosophical Transactions of the Royal Society of London* **124**, 247–308.
549. W. L. Hare and A. Lewis (2005). Estimating tangent and normal cones, *Math. Oper. Res.*, to appear.
550. P. T. Harker and J.-S. Pang (1990), Finite-dimensional variational inequalities and complementarity problems: a survey of theory, algorithms and applications, *Math. Progr.* **60**, 161–220.
551. J. Haslinger, M. Miettinen and P. D. Panagiotopoulos (1999), *Finite Elements Methods for Hemivariational Inequalities*, Kluwer, Dordrecht, The Netherlands.
552. F. Hausdorff (1927), *Mengenlehre*, Walter de Gruyter, Berlin.
553. R. Haydon (1990), A counterexample in several quastions about scattered compact spaces, *Bull. London Math. Soc.* **22**, 261–268.
554. Z.-X. He (1987), State constrained control problems governed by variational inequalities, *SIAM J. Control Optim.* **25**, 1119–1144.
555. R. Henrion (1995), Topological characterization of the approximate subdifferential in the finite-dimensional cases, *Math. Methods Oper. Res.* **41**, 161–173.
556. R. Henrion (1997), Topological properties of the approximate subdifferential, *J. Math. Anal. Appl.* **207**, 345–360.
557. R. Henrion (1997), Characterization of stability for cone increasing constraints in stochastic programming, *Set-Valued Anal.* **5**, 323–349.
558. R. Henrion (2004), Perturbation analysis of chance-constrained programs under variation of all constraint data, in *Dynamic Stochastic Optimization*, Lecture Notes Econ. Math. Syst. **523**, pp. 257–274, Springer, Berlin.

559. R. Henrion and A. Jourani (2002), Subdifferential conditions for calmness of convex constraints, *SIAM J. Optim.* **13**, 520–534.
560. R. Henrion, A. Jourani and J. V. Outrata (2002), On the calmness of a class of multifunctions, *SIAM J. Optim.* **13**, 603–618.
561. R. Henrion and J. V. Outrata (2001), A subdifferential condition for calmness of multifunctions, *J. Math. Anal. Appl.* **258**, 110–130.
562. R. Henrion and J. V. Outrata (2005), Calmness of constraint systems with applications, *Math. Progr.*, to appear.
563. R. Henrion and W. Römisch (1999), Metric regularity and quantitative stability in stochastic programming with probabilistic constraints, *Math. Progr.* **84**, 55–88.
564. R. Henrion and W. Römisch (2000), Stability of solutions to chance constrained stochastic programs, in *Parametric Optimization and Related Topics V*, edited by J. Guddat et al., pp. 95–114, Peter Lang, Frankfurt, Germany.
565. M. R. Hestenes (1966), *Calculus of Variations and Optimal Control Theory*, Wiley, New York.
566. J. R. Hicks (1939), The foundations of welfare economics, *Econ. J.* **49**, 696–712.
567. D. Hilbert (1902), Mathematical problems, *Bull. Amer. Math. Soc.* **7**, 437–479.
568. W. Hildenbrand (1969), Pareto optimality for a measure space of economic agents, *Intern. Econ. Rev.* **10**, 363–372.
569. J.-B. Hiriart-Urruty (1979), New concepts in nondifferentiable programming, *Bull. Soc. Math. France* **60**, 57–85.
570. J.-B. Hiriart-Urruty (1979), Tangent cones, generalized gradients and mathematical programming in Banach spaces, *Math. Oper. Res.* **4**, 79–97.
571. J.-B. Hiriart-Urruty (1979), Refinements of necessary optimality conditions in nondifferentiable programming I, *Appl. Math. Optim.* **5**, 63–82.
572. J.-B. Hiriart-Urruty (1982), Refinements of necessary optimality conditions in nondifferentiable programming II, *Math. Progr. Study* **19**, 120–139.
573. J.-B. Hiriart-Urruty (1989), From convex optimization to nonconvex optimization: Necessary and sufficient conditions for global optimality, in *Nonconvex Optimization and Related Topics*, pp. 219–239, Plenum Press, New York.
574. J.-B. Hiriart-Urruty and Y. S. Ledyaev (1996), A note on the characterization of the global maxima of a (tangentially) convex function over a convex set, *J. Convex Anal.* **3**, 55–61.
575. J.-B. Hiriart-Urruty and C. Lemaréchal (1993), *Convex Analysis and Minimization Algorithms*, published in two volumes, Springer, Berlin.
576. J.-B. Hiriart-Urruty and P. Plazanet (1989), Moreau's decomposition theorem revisited, *Ann. Inst. H. Poincaré: Analyse Non Linéaire* **6**, 325–338.
577. J.-B. Hiriart-Urruty and A. Seeger (1989), Calculus rules on a new set-valued second-order derivative for convex functions, *Nonlinear Anal.* **13**, 721–738.
578. I. Hlaváček, J. Haslinger, J. Nečas and J. Lovišek (1988), *Solution of Variational Inequalities in Mechanics*, Springer, New York.
579. A. J. Hoffman (1952), On approximate solutions to systems of linear inequalities, *J. Nat. Bureau Stand.* **49**, 263–265.
580. R. B. Holmes (1975), *Geometric Functional Analysis and Its Applications*, Springer, New York.

581. L. Hörmander (1954), Sur la function d'appui des ensdembles convexes dans un espace localement convexe, *Arkiv för Mat.* **3**, 181–186.
582. L. Hörmander (1990), *The Analysis of Linear Partial Differential Operators I*, Springer, Berlin.
583. R. Horst, P. D. Pardalos and N. V. Thoai (2000), *Introduction to Global Optimization*, 2nd edition, Kluwer, Dordrecht, The Netherlands.
584. X. Hu and D. Ralph (2002), A note of sensitivity of value functions of mathematical programs with complementarity constraints, *Math. Progr.* **93**, 265–279.
585. X. Hu, D. Ralph, E. K. Ralph, P. Bardsley and M. C. Ferris (2002), The effect of transmission capacities on competition in deregulated electricity markets, preprint.
586. D. K. Hughes (1968), Variational and optimal control problems with delayed argument, *J. Optim. Theory Appl.* **2**, 1–14.
587. A. D. Ioffe (1979), Regular points of Lipschitz functions, *Trans. Amer. Math. Soc.* **251**, 61–69.
588. A. D. Ioffe (1979), Necessary and sufficient conditions for a local minimum, I: A reduction theorem and first order conditions, *SIAM J. Control Optim.* **17**, 245–250.
589. A. D. Ioffe (1981), Nonsmooth analysis: Differential calculus of non-differentiable mappings, *Trans. Amer. Math. Soc.* **266** (1981), 1–56.
590. A. D. Ioffe (1981), Sous-differentielles approaches de fonctions numériques, *C.R. Acad. Sci. Paris* **292**, 675–678.
591. A. D. Ioffe (1981), Calculus of Dini subdifferentials, CEREMADE Publication 8110, Universiteé de Paris IX "Dauphine".
592. A. D. Ioffe (1981), Approximate subdifferentials of nonconvex functions, CEREMADE Publication 8120, Universiteé de Paris IX "Dauphine".
593. A. D. Ioffe (1983), On subdifferentiability spaces, *Ann. New York Acad. Sci.* **410**, 107–119.
594. A. D. Ioffe (1984), Calculus of Dini subdifferentials and contingent coderivatives of set-valued maps, *Nonlinear Anal.* **8**, 517–539.
595. A. D. Ioffe (1984), Necessary conditions in nonsmooth optimization, *Math. Oper. Res.* **9**, 159–188.
596. A. D. Ioffe (1984), Approximate subdifferentials and applications, I: The finite dimensional theory, *Trans. Amer. Math. Soc.* **281**, 389–415.
597. A. D. Ioffe (1986), Approximate subdifferentials and applications, II: Functions on locally convex spaces, *Mathematika* **33**, 111–128.
598. A. D. Ioffe (1987), On the local surjection property, *Nonlinear Anal.* **11**, 565–592.
599. A. D. Ioffe (1989), Approximate subdifferentials and applications, III: The metric theory, *Mathematika* **36**, 1–38.
600. A. D. Ioffe (1990), Proximal analysis and approximate subdifferentials, *J. London Math. Soc.* **41**, 175–192.
601. A. D. Ioffe (1991), Variational analysis of composite functions: A formula for the lower second order epi-derivative, *J. Math. Anal. Appl.* **160**, 379–405.
602. A. D. Ioffe (1993), A Lagrange multiplier rule with small convex-valued subdifferentials for non-smooth problems of mathematical programming involving equality and non-functional constraints, *Math. Progr.* **72**, 137–145.

603. A. D. Ioffe (1996), Nonsmooth subdifferentials: Their calculus and applications, in *Proceedings of the First World Congress of Nonlinear Analysts*, edited by V. Lakshmikantham, pp. 2299–2310, De Gruyter, Berlin.
604. A. D. Ioffe (1997), Directional compactness and nonsmooth semi-Fredholm mappings, *Nonlinear Anal.* **29**, 201–219.
605. A. D. Ioffe (1997), Euler-Lagrange and Hamiltonian formalisms in dynamic optimization, *Trans. Amer. Amer. Soc.* **349**, 2871–2900.
606. A. D. Ioffe (1998), Fuzzy principles and characterization of trustworthiness, *Set-Valued Anal.* **6**, 265–276.
607. A. D. Ioffe (2000), Codirectional compactness, metric regularity and subdifferential calculus, in *Constructive, Experimental and Nonlinear Analysis*, edited by M. Théra, Canad. Math. Soc. Conf. Proc. **27**, 123–164.
608. A. D. Ioffe (2000), Metric regularity and subdifferential calculus, *Russian Math. Surveys* **55**, 501–558.
609. A. D. Ioffe (2003), On stability estimates for the regularity property of maps, in *Topological Methods, Variational Methods and Their Applications*, edited by H. Breziz et al., pp. 133–142, World Scientific Publishing, River Edge, New Jersey.
610. A. D. Ioffe (2003), On robustness for the regularity property of maps, *Control Cybernet.* **32**, 543–554.
611. A. D. Ioffe (2005), Optimality alternative: A non-variational approach to necessary conditions, in *Variational Analysis and Applications*, edited by F. Giannessi and A. Maugeri, pp. 531–552, Springer, Berlin.
612. A. D. Ioffe and V. L. Levin (1972), Subdifferentials of convex functions, *Trudy Moscow Mat. Ob.* **26**, 3–73.
613. A. D. Ioffe and T. Milosz (2002), On a characterization of $C^{1,1}$ functions, *Cybernetics and Systems Analysis*, No. 3, 3–13.
614. A. D. Ioffe and J.-P. Penot (1996), Subdifferentials of performance functions and calculus of coderivatives of set-valued mappings, *Serdica Math. J.* **22**, 359–384.
615. A. D. Ioffe and J.-P. Penot (1997), Limiting subhessians, limiting subjects and their calculus, *Trans. Amer. Math. Soc.* **349**, 789–807.
616. A. D. Ioffe and R. T. Rockafellar (1996), The Euler and Weierstrass conditions for nonsmooth variational problems, *Calc. Var. Partial Diff. Eq.* **4**, 59–87.
617. A. D. Ioffe and V. M. Tikhomirov (1968), Extensions of variational problems, *Trans. Moscow Math. Soc.* **18**, 207–273.
618. A. D. Ioffe and V. M. Tikhomirov (1979), *Theory of Extremal Problems*, North-Holland, Amsterdam, The Netherlands.
619. A. D. Ioffe and V. M. Tikhomirov (1997), Some remarks on variational principles, *Math. Notes* **61**, 248–253.
620. A. D. Ioffe and A. J. Zaslavski (2000), Variational principles and well-posedness in optimization and calculus of variations, *SIAM J. Control Optim.* **38**, 566–581.
621. A. N. Iusem (1998), On some properties of generalized proximal point methods for variational inequalities, *J. Optim. Theory Appl.* **96**, 337–362.
622. M. Ivanov (2004), Sequential representation formulae for G-subdifferential and Clarke subdifferential in smooth Banach spaces, *J. Convex Anal.* **11**, 179–196.

623. A. F. Izmailov (2006), *Sensitivity in Optimization*, Fizmatlit, Moscow.
624. A. F. Izmailov and M. V. Solodov (2001), Error bounds for 2-regular mappings with Lipschitzian derivatives and its applications, *Math. Progr.* **89**, 413–435.
625. C. G. J. Jacobi (1866), *Volesungen über Dynamik*, edited by A. Clebsch, Verlag Georg Reimer, Berlin.
626. M. Q. Jacobs and C. E. Langenhop (1976), Criteria for function space controllability of linear neutral systems, *SIAM J. Control Optim.* **14**, 1009–1048.
627. J. Jahn (2004), *Vector Optimization: Theory, Applications and Extensions*, Springer, Berlin.
628. J. Jahn, A. A. Khan and P. Zeilinger (2005), Second order optimality conditions in set-valued optimization, *J. Optim. Theory Appl.* **125**, 331–347.
629. R. Janin (1973), Sur une classe de fonctions sous-linéarisables, *C. R. Acad. Sci. Paris* **277**, 265–267.
630. V. Jeyakumar and D. T. Luc (1998), Approximate Jacobian matrices for nonsmooth continuous maps and C^1-optimization, *SIAM J. Control Optim.* **36**, 1815–1832.
631. V. Jeyakumar and N. D. Yen (2004), Solution stability, regularity and implicit functions for nonsmooth continuous systems, *SIAM J. Optim.* **14**, 1106–1127.
632. H. Jiang and D. Ralph (2000), Smooth SQP methods for mathematical programs with nonlinear complementarity constraints, *SIAM J. Optim.* **10**, 779–808.
633. A. Jofré (2000), A second-welfare theorem in nonconvex economics, in *Constructive, Experimental and Nonlinear Analysis*, edited by M. Théra, Canad. Math. Soc. Conf. Proc. **27**, 175–184.
634. A. Jofré, D. T. Luc and M. Théra (1998), Epsilon-sundifferentials and epsilon-monotonicity, *Nonlinear Anal.* **33**, 71–90.
635. A. Jofré and J. Rivera (2005), A nonconvex separation property and some applications, *Math. Progr.*, to appear.
636. A. Jofré, R. T. Rockafellar and R. J-B. Wets (2005), A variational inequality scheme for determining an economic equilibrium of classical or extended type, in *Variational Analysis and Applications*, edited by F. Giannessi and A. Maugeri, pp. 553–578, Springer, Berlin.
637. A. Jofré, R. T. Rockafellar and R. J-B. Wets (2005), Variational inequalities and economic equilibrium, *Math. Oper. Res.*, to appear.
638. F. John (1948), Extremum problems with inequalities as side conditions, in *Studies and Essays: Courant Anniversary Volume*, edited by K. O. Fredrichs, O. E. Neugebauer and J. J. Stoker, pp. 187–204, Wiley, New York.
639. H. T. Jongen, D. Klatte and K. Tammer (1990), Implicit functions and sensitivity of stationary points, *Math. Progr.* **49**, 123–138.
640. H. T. Jongen, T. Möbert, J.-J. Rückmann and K. Tammer (1987), Implicit functions and sensitivity analysis of stationary points, *Linear Algerbra Appl.* **95**, 97–109.
641. H. T. Jongen, J.-J. Rückmann and G.-W. Weber (1994), One-parametric semi-infinite optimization: On the stability of the feasible set, *SIAM J. Optim.* **4**, 637–648.
642. E. Jouini (1988), A remark on Clarke's normal cone and the marginal cost pricing rule, *J. Math. Econ.* **17**, 309–315.

643. A. Jourani (1995), Compactly epi-Lipschitzian sets and A-subdifferentials in WT-spaces, *Optimization* **34**, 1–17.
644. A. Jourani (1995), Intersection formulae and the marginal function in Banach spaces, *J. Math. Anal. Appl.* **192**, 867–891.
645. A. Jourani (1998), Necessary conditions for extremality and separation theorems with applications to multobjective optimization, *Optimization* **44**, 327–350.
646. A. Jourani (1999), Limit superior of subdifferentials of uniformly convergent functions, *Positivity* **3**, 33–47.
647. A. Jourani (2000), Hoffman's error bound, local controllability and sensitivity analysis, *SIAM J. Control Optim.* **38**, 947–970.
648. A. Jourani (2003), On a class of compactly epi-Lipschitzian sets, *Nonlinear Anal.* **54**, 471–483.
649. A. Jourani (2005), Weak regularity of functions and sets in Asplund spaces, *Nonlinear Anal.*, to appear.
650. A. Jourani and M. Théra (1998), On the limiting Fréchet ε-sibdifferentials, in *Generalized Convexity, Generalized Monotonicity: Recent Results*, edited by J. P. Crouzeix et al., Nonconvex Optim. Appl. **27**, pp. 185–198, Kluwer, Dordrecht, The Netherlands.
651. A. Jourani and L. Thibault (1990), Approximate subdifferential and metric regularity: The finite-dimensional case, *Math. Progr.* **47**, 203–218.
652. A. Jourani and L. Thibault (1990), The use of metric graphical regularity in approximate subdifferential calculus rules in finite dimensions, *Optimization* **21**, 1–11.
653. A. Jourani and L. Thibault (1993), The approximate subdifferential of composite functions, *Bull. Austral. Math. Soc.* **47**, 443–445.
654. A. Jourani and L. Thibault (1994), A note on Fréchet and approximate subdifferentials of composite functions, *Bull. Austral. Math. Soc.* **49**, 111–116.
655. A. Jourani and L. Thibault (1995), Verifiable conditions for openness, metric regularity of multivalued mappings, *Trans. Amer. Math. Soc.* **347**, 1225–1268.
656. A. Jourani and L. Thibault (1995), Metric regularity for strongly compactly Lipschitzian mappings, *Nonlinear Anal.* **24**, 229–240.
657. A. Jourani and L. Thibault (1996), Metric regularity and subdifferential calculus in Banach spaces, *Set-Valued Anal.* **3**, 87–100.
658. A. Jourani and L. Thibault (1996), Extensions of subdifferential calculus rules in Banach spaces, *Canad. J. Math.* **48**, 834–848.
659. A. Jourani and L. Thibault (1998), Chain rules for coderivatives of multivalued mappings in Banach spaces, *Proc. Amer. Math. Soc.* **126**, 1479–1485.
660. A. Jourani and L. Thibault (1998), Qualification conditions for calculus rules of coderivatives of multivalued mappings, *J. Math. Anal. Appl.* **218**, 66–81.
661. A. Jourani and L. Thibault (1999), Coderivatives of multivalued mappings, locally compact cones and metric regularity, *Nonlinear Anal.* **35**, 925–945.
662. A. Jourani and J. J. Ye (2005), Error bounds for eigenvalue and semidefinite matrix inequality systems, *Math. Progr.*, to appear.
663. G. A. Kamenskii and E. A. Hvilon (1968), Optimality conditions for systems with deviating argument, in *Trudy Sem. Teor. Diff. Urav. Otklon. Argum.* **6**, 213–222.

664. L. V. Kantorovich (1940), On an efficient method for solving some classes of extremum problems, *Dokl. Akad. Nauk SSSR* **28**, 212–215.
665. W. Karush (1939), *Minima of Functions of Several Variables with Inequalities as Side Conditions*, Master Thesis, Department of Mathematics, University of Chicago.
666. B. Kaśkosz and S. Lojasiewicz, Jr. (1985), A maximum principle for generalized control systems, *Nonlinear Anal.* **9**, 109–130.
667. B. Kaśkosz and S. Lojasiewicz, Jr. (1992), Lagrange-type extremal trajectories in differential inclusions, *Syst. Control Lett.* **19**, 241–247.
668. G. A. Kent (1971), A maximum principle for optimal control problems with neutral functional differential systems, *Bull. Amer. Math. Soc.* **77**, 565–570.
669. M. A. Khan (1988) Ioffe's normal cone and the foundations of welfare economics: An example, *Econ. Lett.* **28**, 5–19.
670. M. A. Khan (1991), Ioffe's normal cone and the foundations of welfare economics: The infinite dimensional theory, *J. Math. Anal. Appl.* **161**, 284–298.
671. M. A. Khan (1999), The Mordukhovich normal cone and the foundations of welfare economics, *J. Public Econ. Theory* **1**, 309–338.
672. M. A. Khan and S. Rashid (1975), Nonconvexity and Pareto optimality in large markets, *Inter. Econ. Rev.* **16**, 222–245.
673. M. A. Khan and R. Vohra (1987), An extension of the second welfare theorem to economies with non-convexities and public goods, *Quartery J. Econ.* **102**, 223–245.
674. M. A. Khan and R. Vohra (1988), Pareto optimal allocations of nonconvex economies in locally convex spaces, *Nonlinear Anal.* **12**, 943–950.
675. M. A. Khan and R. Vohra (1988), On approximate decentralization of Pareto optimal allocations in locally convex spaces, *J. Approx. Theory* **52**, 149–161.
676. P. Q. Khanh (1986), An induction theorem and general open mapping theorem, *J. Math. Anal. Appl.* **118**, 519–536.
677. P. Q. Khanh (1989), On general open mapping theorems, *J. Math. Anal. Appl.* **144**, 305–312.
678. G. L. Kharatishvili (1961), The maximum principle in the theory of optimal processes with a delay, *Soviet Math. Dokl.* **2**, 28–32.
679. G. L. Kharatishvili and T. A. Tadumadze (1998), A nonlinear optimal control problem with variable delays, nonfixed initial moment, and piecewise-continuous prehistory, *Proc. Steklov Inst. Math.* **220**, 233–252.
680. D. Kinderlehrer and G. Stampacchia (1980), *An Introduction to Variational Inequalities and Their Applications*, Academic Press, New York.
681. A. J. King and R. T. Rockafellar (1992), Sensitivity analysis for nonsmooth generalized equations, *Math. Progr.* **55**, 193–212.
682. M. Kisielewicz (1991), *Differential Inclusions and Optimal Control*, Kluwer, Dordrecht, The Netherlands.
683. K. C. Kiwiel (1985), *Methods of Descent for Nondifferentiable Optimization*, Springer, Berlin.
684. D. Klatte (1994), On quantitative stability for non-isolated minima, *Control Cybernet.* **23**, 183–200.
685. D. Klatte and R. Henrion (1998), Regularity and stability in nonlinear semi-infinite optimization, in *Semi-Infinite Programming*, edited by R. Reemtsen and J.-J. Rückmann, pp. 69–102, Kluwer, Dordrecht, The Netherlands.

686. D. Klatte and B. Kummer (2002), *Nonsmooth Equations in Optimization: Regularity, Calculus, Methods, and Applications*, Kluwer, Boston, Massachusetts,
687. D. Klatte and B. Kummer (2002), Constrained minima and Lipschitzian penalties in metric spaces, *SIAM J. Optim.* **13**, 619–633.
688. M. Kocan and A. Świech (1996), Perturbed optimization on product spaces, *Nonlinear Anal.* **26**, 81–90.
689. M. Kočvara, M. Kružik and J. V. Outrata (2005), On the control of an evolutionary equilibrium in micromagnetics, in *Optimization with Multivalued Mappings: Theory, Applications and Algorithms*, edited by S. Dempe and V. V. Kalashnikov, Springer, Berlin.
690. M. Kočvara and J. V. Outrata (2004), Optimization problems with equilibrium constraints and their numerical solutions, *Math. Progr.* **101**, 119–149.
691. M. Kočvara and J. V. Outrata (2005), On the modeling and control of delamination processes, in *Control and Boundary Analysis*, edited by J. Cagnol and J.-P. Zolésio, Lecture Notes Pure Applied Math. **240**, pp. 169–185, Marcel Dekker, New York.
692. M. Kojima (1980), Strongly stable stationary solutions in nonlinear programs, in *Analysis and Computation of Fixed Points*, edited by S. M. Robinson, pp. 93–138, Academic Press, New York.
693. P. V. Kokotović, H. K. Khalil and J. O'Reilly (1986), *Singular Perturbations in Control Analysis and Design*, Academic Press, New York.
694. V. B. Kolmanovskii and A. D. Myshkis (1992), *Applied Theory of Functional Differential Equations*, Kluwer, Dordrecht, The Netherlands.
695. V. B. Kolmanovskii and V. R. Nosov (1984), Neutral-type systems with aftereffects, *Autom. Remote Control* **45**, 1–28.
696. V. B. Kolmanovskii and L. E. Shaikhet (1996), *Control of Systems with Aftereffect*, Academic Press, New York.
697. S. Komlosi, T. Rapcsák and S. Schaible, eds. (1994), *Generalized Convexity*, Lecture Notes Econ. Math. Syst. **405**, Springer, Berlin.
698. E. Kostina and O. Kostyukova (2005), Generalized implicit function theorem and its applications to parametric optimal control, *J. Math. Anal. Appl.*, to appear.
699. S. G. Krantz and H. R. Parks (2002), *The Implicit Function Theorem: History, Theory, and Applications*, Birkhäuser, Boston, Massachusetts.
700. N. N. Krasovskii (1964), Optimal processes in systems with time lag, in *Automatics and Remote Control Theory*, edited by L. Butterths, pp. 327–332, Oldenburg, Munish, Germany.
701. N. N. Krasovskii and N. Y. Lukoyanov (2000), Equations of Hamilton-Jacobi type in hereditary systems: Minimax solutions, *Proc. Steklov Inst. Math.: Control in Dynamic Systems, suppl. 1*, 136–153.
702. N. N. Krasovskii and A. I. Subbotin (1988), *Game-Theoretical Control Problems*, Springer, New York.
703. A. J. Krener (1977), The high order maximum principle and its applications to singular extremals, *SIAM J. Control Optim.* **15**, 256–293.
704. V. F. Krotov (1996), *Global Methods in Optimal Control*, Marcel Dekker, New York.
705. A. Y. Kruger (1981), Epsilon-semidifferentials and epsilon-normal elements, Depon. VINITI #1331-81, Moscow.

706. A. Y. KRUGER (1981), *Generalized Differentials of Nonsmooth Functions and Necessary Conditions for an Extremum*, Ph.D. dissertation, Department of Mathematics, Belarus State University, Minsk, Belarus.
707. A. Y. KRUGER (1985), Generalized differentials of nonsmooth functions and necessary conditions for an extremum, *Siberian Math. J.* **26**, 370–379.
708. A. Y. KRUGER (1985), Properties of generalized differentials, *Siberian Math. J.* **26**, 822–832.
709. A. Y. KRUGER (1988), A covering theorem for set-valued mappings, *Optimization* **19**, 763–780.
710. A. Y. KRUGER (1997), Strict ε-semidifferentials and extremality conditions, *Dokl. Nat. Akad. Belarus* **41**, 21–26.
711. A. Y. KRUGER (1998), On extremality of set systems, *Dokl. Nat. Akad. Belarus* **42**, 24–28.
712. A. Y. KRUGER (2002), Strict (ε, δ)-semidifferentials and extremality conditions, *Optimization* **51**, 539–554.
713. A. Y. KRUGER (2003), On Fréchet subdifferentials, *J. Math. Sci.* **116**, 3325–3358.
714. A. Y. KRUGER (2004), Weak stationarity: Eliminating the gap between necessary and sufficient conditions, *Optimization* **53**, 147–164.
715. A. Y. KRUGER (2005), Stationarity and regularity of set systems, *Pacific J. Optim.* **1**, 101–126.
716. A. Y. KRUGER (2005), About regularity of set systems, *SIAM J. Optim.*, to appear.
717. A. Y. KRUGER AND B. S. MORDUKHOVICH (1978), Minimization of nonsmooth functionals in optimal control problems, *Eng. Cybernetics* **16**, 126–133.
718. A. Y. KRUGER AND B. S. MORDUKHOVICH (1980), Extremal points and the Euler equation in nonsmooth optimization, *Dokl. Akad. Nauk BSSR* **24**, 684–687.
719. A. Y. KRUGER AND B. S. MORDUKHOVICH (1980), Generalized normals and derivatives, and necessary optimality conditions in nondifferential programming, Parts I and II, Depon. VINITI: I# 408-80, II# 494-80, Moscow.
720. S. N. KRUZHKOV (1960), The Cauchy problem in the large for certain nonlinear first order differential equations, *Soviet Math. Dokl.* **1**, 474–477.
721. A. V. KRYAZHIMSKII AND Y. S. OSIPOV (1995), On evolutionary-differential games, *Proc. Steklov Math. Inst.* **211**, 257–287.
722. E. I. KUGUSHEV (1973), Maximum principle in optimal control systems with nonsmooth right-hand side, *Vest. Moskov. Univ. Ser. Mat. Mekh.*, No. 3, 107–113.
723. H. W. KUHN (1976), Nonlinear programming: A historical survey, in *Nonlinear Programming*, edited by R. W. Cottle and C. E. Lemke, pp. 1–26, American Mathematical Society, Providence, Rhode Island.
724. H. W. KUHN AND A. W. TUCKER (1951), Nonlinear programming, in *Proceedings of the Second Berkley Symposium on Mathematical Statistics and Probability*, edited by J. Neyman, pp. 481–492, University of California Press, Berkeley, California.
725. B. KUMMER (1991), Lipschitzian inverse functions, directional derivatives and applications in $C^{1,1}$ optimization, *J. Optim. Theory Appl.* **70**, 561–582.
726. B. KUMMER (1991), An implicit function theorem for $C^{0,1}$-equations and parametric $C^{1,1}$-optimization, *J. Math. Anal. Appl.* **158**, 35–46.

727. B. Kummer (1999), Metric regularity: Characterizations, nonsmooth variations and successive approximation, *Optimization* **46**, 247–281.
728. B. Kummer (2000), Inverse functions of pseudo regular mappings and regularity conditions, *Math. Progr.* **88**, 313–339.
729. L. Kuntz and S. Scholtes (1994), Structural analysis of nonsmooth mappings, inverse functions, and metric projections, *J. Math. Anal. Appl.* **188**, 346–386.
730. G. A. Kurina (1979), Application of the method of tents to an optimal control problem for a differential equation with a singular matrix multiplying the derivative, *Diff. Eq.* **15**, 417–423.
731. A. B. Kurzhanskii (1977), *Control and Observation under Conditions of Uncertainty*, Nauka, Moscow.
732. A. B. Kurzhanskii and I. Vályi (1997), *Ellipsoidal Calculus for Estimation and Control*, Burlhäuser, Boston, Massachusetts.
733. A. G. Kusraev and S. S. Kutateladze (1995), *Subdifferentials: Theory and Applications*, Kluwer, Dordrecht, The Netherlands.
734. J. Kyparisis (1992), Parametric variational inequalities with multivalued solution sets, *Math. Oper. Res.* **17**, 341–364.
735. O. A. Ladyzhenskaya, V. A. Solonnikov and N. N. Uralzeva (1968), *Linear and Quasilinear Equations of Parabolic Type*, American Mathematical Society, Providence, Rhode Island.
736. J. E. Lagnese (1989), *Boundary Stabilization of Thin Plates*, SIAM Publications, Philadelphia, Pennsylvania.
737. J. L. Lagrange (1813), *Thérie des fonctions analytique*, Paris.
738. O. Lange (1942), The foundations of welfare economics, *Econometrica* **10**, 215–228.
739. I. Lasiecka (1980), Unified theory of abstract parabolic boundary problems – a semigroup approach, *Appl. Math. Optim.* **6**, 287–333.
740. I. Lasiecka, J.-L. Lions and R. Triggiani (1986), Nonhomogeneous boundary value problems for second-order hyperbolic operators, *J. Mat. Pures Appl.* **65**, 149–192.
741. I. Lasiecka and J. Sokolowski (1988), Regularity and strong convergence of a variational approximation to a nonhomogeneous Dirichlet hyperbolic boundary problem, *SIAM J. Math. Anal.* **19**, 528–540.
742. I. Lasiecka and R. Triggiani (1983), Dirichlet boundary control problems for parabolic equations with quadratic cost: Analyticity and Riccati feedback synthesis, *SIAM J. Control Optim.* **21**, 41–67.
743. I. Lasiecka and R. Triggiani (1987), The regulator problem for parabolic equations with Dirichlet boundary control, I: Riccati's feedback synthesis and regularity of optimal solutions, *Appl. Math. Optim.* **16**, 147–168.
744. I. Lasiecka and R. Triggiani (1990), Sharp regularity theory for second order hyperbolic equations of Neumann type, I: L_2 nonhomogeneous data, *Ann. Mat. Pura Appl.* **157**, 285–367.
745. I. Lasiecka and R. Triggiani (1991), Regularity theory of hyperbolic equations with non-homogeneous Neumann boundary conditions, II: General boundary data, *J. Diff. Eq.* **94**, 112–164.
746. I. Lasiecka and R. Triggiani (2000), *Control Theory for Partial Differential Equations: Continuous and Approximation Theory*, published in two volumes, Cambridge University Press, Cambridge, UK.

747. M. LASSONDE (2001), First-order rules for nonsmooth constrained optimization, *Nonlinear Anal.* **44**, 1031–1056.
748. E. B. LEACH (1961), A note on inverse function theorem, *Proc. Amer. Math. Soc.* **12**, 694–697.
749. G. LEBOURG (1975), Valeur moyenne pour gradient généraliseé, *C. R. Acad. Sci. Paris* **281**, 795–798.
750. Y. S. LEDYAEV (2004), On generic existence and uniqueness in nonconvex optimal control, *Set-Valued Anal.* **12**, 147–162.
751. Y. S. LEDYAEV AND Q. J. ZHU (1999), Implicit multifunction theorem, *Set-Valued Anal.* **7**, 209–238.
752. Y. S. LEDYAEV AND Q. J. ZHU (2004), Techniques for nonsmooth analysis on smooth manifolds I: Techniques for local problems, in *Optimal Control, Stabilization and Nonsmooth Analysis*, edited by M. de Queiroz, M. Malisoff and P. Wolenski, Lecture Notes Cont. Inf. Sci. **301**, pp. 283–287, Springer, New York.
753. Y. S. LEDYAEV AND Q. J. ZHU (2004), Techniques for nonsmooth analysis on smooth manifolds II: Using deformations and flowes, in *Optimal Control, Stabilization and Nonsmooth Analysis*, edited by M. de Queiroz, M. Malisoff and P. Wolenski, Lecture Notes Cont. Inf. Sci. **301**, pp. 299–311, Springer, New York.
754. Y. S. LEDYAEV AND Q. J. ZHU (2005), Nonsmooth analysis on smooth manifolds, *Trans. Amer. Math. Soc.*, to appear.
755. G. M. LEE, N. N. TAM AND N. D. YEN (2005), Normal coderivative for multifunctions and implicit function theorems, preprint.
756. U. LEDZEWICZ AND H. SCHÄTTLER (1998), High-order approximations and generalized necessary conditions for optimality, *SIAM J. Control Optim.* **37**, 33–56.
757. G. W. LEIBNIZ (1696), Letter to Johann Bernoulli of July 31, 1696, Hanover, Germany.
758. G. LEITMANN (1978), On generalized Stackelberg strategies, *J. Optim. Theory Appl.* **59**, 637–643.
759. C. LEMARÉCHAL, F. OUSTRY AND C. SAGASTIZÁBAL (2000), The $\mathcal{U}$-Lagrangian of a convex function, *Trans. Amer. Math. Soc.* **352**, 711–729.
760. F. LEMPIO (1995), Euler's method revisited, *Proc. Steklov Inst. Math.* **211**, 473–494.
761. F. LEMPIO AND V. M. VELIOV (1998), Discrete approximations to differential inclusions, *Mitteilungen der GAMM* **21**, 101–135.
762. S. LENHART, V. PROTOPESCU AND S. STOJANOVIĆ (1993), A minimax problem for semilinear nonlocal competitive systems, *Appl. Math. Optim.* **28**, 113–132.
763. E. S. LEVITIN (1994), *Perturbation Theory in Mathematical Programming and Its Applications*, Wiley, New York.
764. E. S. LEVITIN, A. A. MILYUTIN AND N. P. OSMOLOVSKII (1978), Higher order conditions for a local minimum in the problems with constraints, *Russian Math. Surveys* **33**, 97–168.
765. A. B. LEVY (1993), Second-order epi-derivatives of integral functionals, *Set-Valued Anal.* **1**, 379–392.
766. A. B. LEVY (1996), Implicit multifunction theorems for the sensitivity analysis of variational conditions, *Math. Progr.* **74**, 333–350.

767. A. B. Levy (2001), Lipschitzian multifunctions and a Lipschitzian inverse function theorem, *Math. Oper. Res.* **26**, 105–118.
768. A. B. Levy (2001), Solution stability from general principles, *SIAM J. Control Optim.* **40**, 209–238.
769. A. B. Levy and B. S. Mordukhovich (2004), Coderivatives in parametric optimization, *Math. Progr.* **99**, 311–327.
770. A. B. Levy and R. A. Poliquin (1997), Characterizing the single-valuedness of multifunctions, *Set-Valued Anal.* **5**, 351–364.
771. A. B. Levy, R. A. Poliquin and R. T. Rockafellar (2000), Stability of locally optimal solutions, *SIAM J. Optim.* **10**, 580–604.
772. A. B. Levy, R. A. Poliquin and L. Thibault (1995), A partial extension of Attouch's theorem and its applications to second-order epi-differentiation, *Trans. Amer. Math. Soc.* **347**, 1269–1294.
773. A. B. Levy and R. T. Rockafellar (1994), Sensitivity analysis of solutions to generalized equations, *Trans. Amer. Math. Soc.* **345**, 661–671.
774. A. B. Levy and R. T. Rockafellar (1996), Variational conditions and the proto-differentiation of partial subgradient mappings, *Nonlinear Anal.* **26**, 1951–1964.
775. A. S. Lewis (1999), Nonsmooth analysis of engenvalues, *Math. Progr.* **84**, 1–24.
776. A. S. Lewis (1999), Ill-conditioned convex processes and conic linear systems, *Math. Oper. Res.* **23**, 829–834.
777. A. S. Lewis (2001), Ill-conditioned inclusions, *Set-Valued Anal.* **9**, 375–381.
778. A. S. Lewis (2003), Active sets, nonsmoothness and sensitivity, *SIAM J. Optim.* **13**, 702–725.
779. A. S. Lewis (2003), The mathematics of eigenvalue optimization, *Math. Progr.* **97**, 155–176.
780. A. S. Lewis (2004), The structured distance to ill-posedness for conic systems, *Math. Oper. Res.* **29**, 776–785.
781. A. S. Lewis and M. L. Overton (1996), Eigenvalue optimization, *Acta Numerica* **5**, 149–190.
782. A. S. Lewis and H. S. Sendov (2005), Nonsmooth analysis of singular values. Part I: Theory, *Set-Valued Anal.* **13**, 213–241.
783. A. S. Lewis and H. S. Sendov (2005), Nonsmooth analysis of singular values. Part II: Applications, *Set-Valued Anal.* **13**, 243–264.
784. W. Li and I. Singer (1998), Global error bounds for convex multifunctions, *Math. Oper. Res.* **23**, 443–462.
785. Y. Li and S. Shi (2000), A generalization of Ekeland's ε-variational principle and its Borwein-Preiss variant, *J. Math. Anal. Appl.* **246**, 308–319.
786. X. Li and Y. Yao (1981), On optimal control for distributed parameter systems, in *Proc. IFAC 8th Triennial World Congress*, pp. 207–212, Kyoto, Japan.
787. X. Li and Y. Yao (1985), Maximum principle of distributed parameter systems with time lags, in *Lecture Notes Cont. Inform. Sci.* **75**, pp. 410–427, Springer, Berlin.
788. X. Li and J. Yong (1991), Necessary conditions of optimal control for distributed parameter systems, *SIAM J. Control Optim.* **29**, 895–908.
789. X. Li and J. Yong (1995), *Optimal Control Theory for Infinite-Dimensional Systems*, Birkhäuser, Boston, Massachusetts.

790. L.-J. Lin and W.-S. Du (2005), Ekeland's variational principles and the existence of nonconvex equilibria and minimax inequalities in complete metric spaces, *J. Math. Anal. Appl.*, to appear.
791. J.-L. Lions (1971), *Optimal Control of Systems Governed by Partial Differential Equations*, Springer, Berlin.
792. J.-L. Lions (1983), *Contrôle des Systémes Distribués Singuliers*, Gauthier-Villars, Paris.
793. P.-L. Lions (1982), *Generalized Solutions of Hamilton-Jacobi Equations*, Pitman, Boston, Massachusetts.
794. J.-L. Lions and E. Magenes (1972), *Nonhomogeneous Boundary Value Problems and Applications*, Springer, New York.
795. J.-L. Lions and G. Stampacchia (1967), Variational inequalities, *Comm. Pure Appl. Math.* , 493–519.
796. R. Lipschitz (1877), *Lehrbuch der Analysis*, Cohen & Sohn, Bonn.
797. J. M. Liu (1995), Strong stability in variational inequalities, *SIAM J. Control Optim.* **33**, 725–749.
798. P. D. Loewen (1987), The proximal normal formula in Hilbert spaces, *Nonlinear Anal.* **11**, 979–995.
799. P. D. Loewen (1988), The proximal subgradient formula in Banach spaces, *Canad. Math. Bull.* **31**, 353–361.
800. P. D. Loewen (1992), Limits of Fréchet normals in nonsmooth analysis, in *Optimization and Nonlinear Analysis*, edited by A. Ioffe, L. Marcus and S. Reich, Pitman Research Notes Math. Ser. **244**, pp. 178–188, Longman, Harlow, Essex, UK.
801. P. D. Loewen, *Optimal Control via Nonsmooth Analysis* (1993), American Mathematical Society, Providence, Rhode Island.
802. P. D. Loewen (1994), A mean value theorem for Fréchet subgradients, *Nonlinear Anal.* **23**, 1365–1381.
803. P. D. Loewen, F. H. Clarke and R. B. Vinter (1988), Differential inclusions with free time, *Ann. Inst. H. Poincaré: Analyse Non Linéaire* **5**, 573–593.
804. P. D. Loewen and R. T. Rockafellar (1991), The adjoint arcs in nonsmooth optimization, *Trans. Amer. Math. Soc.* **325**, 39–72.
805. P. D. Loewen and R. T. Rockafellar (1994), Optimal control of unbounded differential inclusions, *SIAM J. Control Optim.* **32**, 442–470.
806. P. D. Loewen and R. T. Rockafellar (1996), New necessary conditions for the generalized problem of Bolza, *SIAM J. Control Optim.* **34**, 1496–1511.
807. P. D. Loewen and R. T. Rockafellar (1997), Bolza problem with general time constraints, *SIAM J. Control Optim.* **35**, 2050–2069.
808. P. D. Loewen and R. B. Vinter (1987), Pontryagin type necessary conditions for differential inclusions, *Syst. Control Lett.* **9**, 263–265.
809. P. D. Loewen and X. Wang (2001), A generalized variational principle, *Canad. J. Math.* **53**, 1174–1193.
810. S. Lojasiewicz, Jr. (1992), Local controllability of parameterized differential equations, preprint.
811. P. Loridan (1982), Necessary conditions for ε-optimality, *Math. Progr. Study* **19**, 140–152.
812. P. Loridan and J. Morgan (1989), New results on approximate solutions in two level optimization, *Optimization* **20**, 819–836.

813. D. T. Luc (1989), *Theory of Vector Optimization*, Springer, Berlin.
814. D. T. Luc (1996), A strong mean value theorem and applications, *Nonlinear Anal.* **26**, 915–923.
815. D. T. Luc (2005), Chain rules for approximate Jacobians of continuous functions, *Nonlinear Anal.* **61**, 97–114.
816. Y. Lucet and J. J. Ye (2001), Sensitivity analysis of the value function for optimization problems with variational inequality constraints, *SIAM J. Optim.* **40**, 699–723.
817. D. R. Luke, J. V. Burke and R. G. Lyon (2002), Optical wavefront reconstruction: Theory and numerical methods, *SIAM Rev.* **44**, 169–224.
818. Y. Luo and A. Eberhard (2005), Comparizon principles for viscosity solutions of elliptic equations via fuzzy sum rule, *J. Math. Anal. Appl.* **307**, 736–752.
819. Y. Luo and A. Eberhard (2005), An application of $C^{1,1}$ approximation to comparison principles for viscosity solutions of curvature equations, preprint.
820. Z. Q. Luo, J.-S. Pang and D. Ralph (1996), *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, UK.
821. K. A. Lurie (1993), *Applied Optimal Control of Distributed Systems*, Plenum Press, New York.
822. A. A. Lyapunov (1940), Sur les founctions-vecteurs complétement additives, *Izvest. Akad. Nauk SSSR, Ser. Mat.* **3**, 465–478.
823. S. I. Lyashko (2002), *Generalized Optimal Control of Linear Systems with Distributed Parameters*, Kluwer, Dordrecht, The Netherlands.
824. L. A. Lyusternik (1934), On conditional extrema of functionals, *Math. Sbornik* **41**, 390–401.
825. U. Mackenroth (1982), Convex parabolic boundary control problems with pointwise state constraints, *J. Math. Anal. Appl.* **87**, 256–277.
826. G. G. Magaril-Il'yaev (1978), An implicit function theorem for Lipschitzian mappings, *Russian Math. Surveys* **33**, 209–210.
827. E. N. Mahmudov (2005), On duality in problems of optimal control described by convex differential inclusions of Goursat-Darboux type, *J. Math. Anal. Appl.* **307**, 628–640.
828. E. N. Mahmudov (2005), The optimality principle for discrete and first order partial differential inclusions, *J. Math. Anal. Appl.* **308**, 605–619.
829. V. L. Makarov, M. J. Levin and A. M. Rubinov (1995), *Mathematical Economic Theory: Pure and Mixed Types of Economic Mechanisms*, North-Holland, Amsterdam, The Netherlands.
830. K. Malanowski (1969), On optimal control of the vibrating string, *SIAM J. Control* **7**, 260–271.
831. K. Malanowski (1979), On convergence of finite difference approximations to control and state constrained convex optimal control problems, *Arch. Autom. Telemech.* **24**, 319–337.
832. K. Malanowski (1987), *Stability of Solutions to Convex Problems of Optimization*, Springer, Berlin.
833. K. Malanowski (1994), Regularity of solutions in stability analysis of optimization and optimal control problems, *Control Cybernet.* **23**, 61–86.
834. K. Malanowski (2003), Two-norm approach in stability and sensitivity analysis of optimization and optimal control problems, *Advan. Math. Sci. Appl.* **2**, 397–443.

835. K. Malanowski and Y. Sokolowski (1986), Sensitivity of solutions to convex, control constrained optimal control problems for distributed parameter systems, *J. Math. Anal. Appl.* **120**, 240–263.
836. G. G. Malcolm and B. S. Mordukhovich (2001), Pareto optimality in nonconvex economies with infinite-dimensional commodity spaces, *J. Global Optim.* **20**, 323–346.
837. M. Malisoff, L. Rifford and E. Sontag (2003), Remarks on input to state stabilization, *Proc. 42rd IEEE Conf. Dec. Cont.*, pp. 1053–1058, Maui, Hawaii.
838. M. Malisoff, L. Rifford and E. Sontag (2004), Global asymptotic controllability implies input to state stabilization, *SIAM J. Control Optim.* **42**, 1121–1138.
839. O. L. Mangasarian (1985), A condition number for differentiable convex inequalities, *Math. Oper. Res.* **10**, 175–179.
840. O. L. Mangasarian (1994), *Nonlinear Programming*, SIAM Publications, Philadelphia, Pennsylvania.
841. O. L. Mangasarian and S. Fromovitz (1967), The Fritz John condition in the presence of equality and inequality constraints, *J. Math. Anal. Appl.* **17**, 34–47.
842. O. L. Mangasarian and T. H. Shiau (1986), Error bounds for monotone linear complementarity problems, *Math. Progr.* **36**, 81–89.
843. K. B. Mansimov (1998), Singular controls in systems of neutral type, *Autom. Remote Control* **59**, 653–661.
844. K. B. Mansimov (2001), On the theory of necessary conditions for optimality in a problem with distributed parameters, *Comput. Maths. Math. Phys.* **41**, 1429–1443.
845. C. Marcelli (2002), Variational problems with nonconvex, noncoercive, highly discontinuous integrands: characterization and existence of minimizers, *SIAM J. Control Optim.* **40**, 1473–1490.
846. C. Marcelli (2005), Necessary and sufficient conditions for optimality of nonconvex, noncoercive autonomous variational problems with constraints, preprint.
847. C. Marcelli, E. Outkine and M. Sytchev (2002), Remarks on necessary conditions for minimizers of one-dimensional variational problems, *Nonlinear Anal.* **48**, 979–993.
848. S. Marcellin (2004), *Intégration D'epsilon-sous-différentiels et Problémes D'évolution Non Convexes*, Ph.D. dissertation, Department of Mathematics, University of Montpellier II, Montpellier, France.
849. S. Marcellin and L. Thibault (2005), Integration of ε-Fenchel subdifferentials and maximal cyclic monotonicity, *J. Global Optim.* **32**, 83–91.
850. A. Marchaud (1934), Sur les champs de demi-cônes et les équations différentielles du premier ordre, *Bull. Sc. Math.* **62**, 12–38.
851. A. Marino and M. Tosques (1990), Some variational problems with lack of convexity and some partial differential inequalities, in *Methods of Nonconvex Analysis*, edited by A. Cellina, Lecture Notes in Mathematics **1446**, pp. 58–83, Springer, Berlin.
852. J. E. Martinez-Legaz and P. H. Sach (1999), A new subdifferential in quasiconvex analysis, *J. Convex Anal.* **6**, 1–12.
853. J. E. Martinez-Legaz and M. Théra (1996), Epsilon-subdifferentials in terms of subdifferentials, *Set-Valued Anal.* **4**, 327–332.

854. A. Mas-Colell (1985), Pareto optima and equilibria: The inifinite-dimensional case, in *Advances in Economics*, edited by C. D. Aliprantis et al., pp. 25–42, Springer, New York.
855. A. Mas-Colell (1986), Valuation equilibrium and Pareto optimum revisited, in *Contributions to Mathematical Economics*, edited by W. Hildenbrand and A. Mas-Collel, pp. 317–331, North-Holland, Amsterdam, The Netherlands.
856. A. Mas-Colell, M. D. Whinston and J. R. Green (1995), *Microeconomic Theory*, Oxford University Press, Oxford, UK.
857. H. Maurer and N. P. Osmolovskii (2003), Second order conditions, *Control Cybernet.* **32**, 555–584.
858. H. Maurer and J. Zowe (1979), First and second-order necessary and sufficient optimality conditions for infinite-dimensional programming problems, *Math. Progr.* **16**, 98–110.
859. A. Mayer (1884), Zur aufstellung der kriterien des maximums und minimums der einfachen integrale bei variabeln grenwerten, *Leipziger Berichte* **36**, 99–128.
860. E. J. McShane (1939), On multipliers for Lagrange problem, *Amer. J. Math.* **91**, 809–819.
861. E. J. McShane (1940), Generalized curves, *Duke Math. J.* **6**, 513–536.
862. E. J. McShane (1940), Necessary conditions for generalized curve problems in the calculus of variations, *Duke Math. J.* **7**, 1–27.
863. E. J. McShane (1967), Relaxed controls and vbariational problems, *SIAM J. Contr.* **5**, 438–485.
864. E. J. McShane (1973), The Lagrange multiplier rule, *Amer. Math. Monthly* **80**, 922–925.
865. E. J. McShane (1989), The calculus of variations from the beginning through optimal control theory, *SIAM J. Control Optim.* **27**, 916–939.
866. N. G. Medhin (1990), Minimizing sequences for differential inclusion problems in the presence of state constraints, *Comput. Math. Appl.* **19**, 127–134.
867. N. G. Medhin (1995), On optimal control of functional-differential systems, *J. Optim. Theory Appl.* **85**, 363–376.
868. T. K. Melikov (1996), An analogue of Pontryagin's maximum principle in systems with aftereffect of neutral type, *Comput. Maths. Math. Phys.* **26**, 1541–1546.
869. A. A. Melikyan (1998), *Generalized Characteristics of First Order PDEs*, Birkhäuser, Boston, Massachusetts.
870. P. Michel and J.-P. Penot (1984), Calcul sous-différentiel pour des fonctions lipschitziennes et non-lipschitziennes, *C. R. Acad. Sci. Paris* **298**, 684–687.
871. P. Michel and J.-P. Penot (1992), A generalized derivative for calm and stable functions, *Diff. Integ. Eq.* **5**, 189–196.
872. R. Mifflin (1977), Semismooth and semiconvex functions in constrained optimization, *SIAM J. Control Optim.* **15**, 957–972.
873. R. Mifflin and C. Sagastizábal (2003), Primal-dual gradient structured functions: Second-order results; links to epi-derivatives and partly smooth functions, *SIAM J. Optim.* **13**, 1174–1194.
874. R. Mifflin and C. Sagastizábal (2004), $\mathcal{VU}$-smoothness and proximal point results for some nonconvex functions, *Optim. Meth. Soft.* **19**, 463-478.

875. A. A. Milyutin (1999), Convex-valued Lipschitz differential inclusions and Pontryagin's maximum principle, in *Optimal Control*, edited by R. V. Gamkrelidge, Vol. 4, pp. 175–187, VINITI, Moscow.
876. A. A. Milyutin (2003), On strengthening the conditions by Clarke and Smirnov for convex-valued differential inclusions, *Math. Sbornik* **194**, 261–280.
877. A. A. Milyutin and N. P. Osmolovskii (1998), *Calculus of Variations and Optimal Control*, American Mathematical Society, Providence, Rhode Island.
878. L. I. Minchenko (1999), Necessary optimality conditions for differential-difference inclusions, *Nonlinear Anal.* **35**, 307–322.
879. L. I. Minchenko (2003), Multivalued analysis and differential properties of multivalued mappings and marginal functions, *J. Math. Sci.* **116**, 3266–3302.
880. L. I. Minchenko and S. Sirotko (2002), Controllability of nonsmooth discrete systems with delays, *Optimization* **51**, 161–174.
881. L. I. Minchenko and A. A. Volosevich (2003), Value function and necessary optimality conditions in optimal control problems for differential-difference inclusions, *Nonlinear Anal.* **53**, 407–424.
882. H. Minkowski (1911), *Theorie der Konvexen Körper, Insbesondere Begründung inhres Ober Flächenbegriffs*, Gesammelte Abhandlungen, II, B. G. Teubner, Leipzig.
883. V. J. Mizel and T. I. Seidman (1997), An abstract bang-bang principle and time-optimal boundary control of the heat equation, *SIAM J. Control Optim.* **35**, 1204–1216.
884. N. N. Moiseev (1971), *Numerical Methods in Optimal Systems Theory*, Nauka, Moscow.
885. N. N. Moiseev (1975), *Elements of Optimal Systems Theory*, Nauka, Moscow.
886. M. D. P. Monteiro Marques (1993), *Differential Inclusions in Nonsmooth Mechanical Problems*, Birkhäuser, Boston, Massachusetts.
887. B. S. Mordukhovich (1976), Maximum principle in problems of time optimal control with nonsmooth constraints, *J. Appl. Math. Mech.* **40**, 960–969.
888. B. S. Mordukhovich (1977), Existence of optimal controls, *J. Soviet Math.* **7**, 850–886.
889. B. S. Mordukhovich (1977), Approximation and maximum principle for nonsmooth problems of optimal control, *Russian Math. Surveys* **196**, 263–264.
890. B. S. Mordukhovich (1978), On difference approximations of optimal control systems, *J. Appl. Math. Mech.* **42**, 452–461.
891. B. S. Mordukhovich (1978), On the theory of difference approximations of continuous-time control systems, in *Third USSR Conference on Operations Research*, edited by N. N. Moiseev, pp. 330–331, Gorky, USSR.
892. B. S. Mordukhovich (1980), Metric approximations and necessary optimality conditions for general classes of extremal problems, *Soviet Math. Dokl.* **22**, 526–530.
893. B. S. Mordukhovich (1981), Penalty functions and necessary conditions for an extemum in nonsmooth and nonconvex optimization problems, *Russian Math. Surveys* **36**, 242–243.
894. B. S. Mordukhovich (1984), Nonsmooth analysis with nonconvex generalized differentials and adjoint mappings, *Dokl. Akad. Nauk BSSR* **28**, 976–979.

895. B. S. Mordukhovich (1984), Duality theory in systems with aftereffect, *J. Appl. Math. Mech.* **48**, 440–447.
896. B. S. Mordukhovich (1984), Controllability, observability and duality in dynamical systems with aftereffect. *Autom. Remote Control* **45**, 1019–1027.
897. B. S. Mordukhovich (1985), On necessary conditions for an extremum in nonsmooth optimization, *Soviet Math. Dokl.* **32**, 215–220.
898. B. S. Mordukhovich (1986), Optimal control of the groundwater regime on engineering reclamation systems, *Water Resources* **12**, 244–253.
899. B. S. Mordukhovich (1986), On the theory of difference approximations in optimal control, *Dokl. Akad. Nauk BSSR* **30**, 1964–1067.
900. B. S. Mordukhovich (1987), Approximation methods and optimality conditions in nonconvex control problems, *Soviet Math. Dokl.* **36**, 164–168.
901. B. S. Mordukhovich (1988), *Approximation Methods in Problems of Optimization and Control*, Nauka, Moscow.
902. B. S. Mordukhovich (1988), Approximation and optimization of differential inclusions, *Cybernetics* **24**, 781–788.
903. B. S. Mordukhovich (1988), Approximate maximum principle for finite-difference control systems, *USSR Comput. Maths. Math. Phys.* **28**, 106–114.
904. B. S. Mordukhovich (1989), Necessary optimality conditions for nonsmooth control systems with free time, *Diff. Eq.* **25**, 290–299.
905. B. S. Mordukhovich (1990), Minimax design for a class of distributed parameter systems, *Autom. Remote Control* **50**, 262–283.
906. B. S. Mordukhovich (1990), Maximum principle for nonconvex finite difference systems, in *Analysis and Optimization of Systems*, edited by A. Bensoussan and J.-L. Lions, Lecture Notes Cont. Inf. Sci. **144**, pp. 539–548, Springer, Berlin.
907. B. S. Mordukhovich (1992), Sensitivity analysis in nonsmooth optimization, in *Theoretical Aspects of Industrial Design*, edited by D. A. Field and V. Komkov, SIAM Proc. Appl. Math. **58**, pp. 32–46, Philadelphia, Pennsylvania.
908. B. S. Mordukhovich (1992), On variational analysis of differential inclusions, in *Optimization and Nonlinear Analysis*, edited by A. Ioffe, L. Marcus and S. Reich, Pitman Researh Notes Math. Ser. **244**, pp. 199–213, Longman, Harlow, Essex, UK.
909. B. S. Mordukhovich (1993), Complete characterization of openness, metric regularity, and Lipschitzian properties of multifunctions, *Trans. Amer. Math. Soc.* **340**, 1–35.
910. B. S. Mordukhovich (1994), Generalized differential calculus for nonsmooth and set-valued mappings, *J. Math. Anal. Appl.* **183**, 250–288.
911. B. S. Mordukhovich (1994), Lipschitzian stability of constraint systems and generalized equations, *Nonlinear Anal.* **22**, 173–206.
912. B. S. Mordukhovich (1994), Stability theory for parametetic generalized equations and variational inequalities via nonsmooth analysis, *Trans. Amer. Math. Soc.* **343**, 609–658.
913. B. S. Mordukhovich (1994), Sensitivity analysis for constraint and variational systems by means of set-valued differentiation, *Optimization* **31**, 13–46.
914. B. S. Mordukhovich (1994), Necesary optimality and controllability conditions for nonsmooth control systems, in *Proc. 33rd IEEE Conf. Dec. Cont.*, pp. 3992–3997, Orlando, Florida.

915. B. S. Mordukhovich (1995), Discrete approximations and refined Euler-Lagrange conditions for nonconvex differential inclusions, *SIAM J. Control Optim.* **33**, 882–915.
916. B. S. Mordukhovich (1996), Optimization and finite difference approximations of nonconvex differential inclusions with free time, in *Nonsmooth Analysis and Geometric Methods in Deterministic Optimal Control*, edited by B. S. Mordukhovich and H. J. Sussmann, pp. 153–202, Springer, New York.
917. B. S. Mordukhovich (1997), Coderivatives of set-valued mappings: Calculus and applications, *Nonlinear Anal.* **30**, 3059–3070.
918. B. S. Mordukhovich (1999), Minimax design of constrained parabolic systems, in *Control of Distributed Parameter and Stochastic Systems*, edited by S. Chen et al., pp. 111–118, Kluwer, Boston, Massachusetts.
919. B. S. Mordukhovich (1999), On variational analysis in infinite dimensions, in *Systems Modelling and Optimization*, edited by M. P. Polis et al., pp. 189–197, Chapman and Hall/CRC Press, Boca Raton, Florida.
920. B. S. Mordukhovich (2000), Abstract extremal principle with applications to welfare economics, *J. Math. Anal. Appl.* **251**, 187–216.
921. B. S. Mordukhovich (2000), Optimal control of difference, differential, and differential-difference inclusions, *J. Math. Sci.* **100**, 2613–2632.
922. B. S. Mordukhovich (2001), The extremal principle and its applications to optimization and economics, in *Optimization and Related Topics*, edited by A. Rubinov and B. Glover, Applied Optimization **47**, pp. 343–369, Kluwer, Dordrecht, The Netherlands.
923. B. S. Mordukhovich (2002), Calculus of second-order subdifferentials in infinite dimensions, *Control Cybernet.* **31**, 557–573.
924. B. S. Mordukhovich (2004), Coderivative analysis of variational systems, *J. Global Optim.* **28**, 347–362.
925. B. S. Mordukhovich (2004), Necessary conditions in nonsmooth minimization via lower and upper subgradients, *Set-Valued Anal.* **12**, 163–193.
926. B. S. Mordukhovich (2004), Equilibrium problems with equilibrium constraints via multiobjective optimization, *Optim. Meth. Soft.* **19**, 479–492.
927. B. S. Mordukhovich (2004), Lipschitzian stability of parametric constraint systems in infinite dimensions, in *Generalized Convexity, Generalized Monotonicity and Aplications*, edited by A. Eberhard et al., Nonconvex Optimization and Its Applications, pp. 39–59, Springer, New York.
928. B. S. Mordukhovich (2005), Optimization and equilibrium problems with equilibrium constraints, *OMEGA* **33**, 379–384.
929. B. S. Mordukhovich (2005), Sensitivity analysis for variational systems, in *Variational Analysis and Applications*, edited by F. Giannessi and A. Maugeri, pp. 723–743, Springer, Berlin.
930. B. S. Mordukhovich (2005), Nonlinear prices in nonconvex economies with classical Pareto and strong Pareto optimal allocations, *Positivity* **9**.
931. B. S. Mordukhovich (2005), Sensitivity analysis for generalized variational and hemivariational inequalities, in *Advances in Analysis*, edited by H. G. W. Begehr et al., pp. 305–314, World Scientific Publishing, London, UK.
932. B. S. Mordukhovich (2005), Optimal control of evolution inclusions, *Nonlinear Anal.*, to appear.
933. B. S. Mordukhovich (2005), Coderivative calculus and robust Lipschitzian stability of variational systems, *J. Convex Anal.*, to appear.

934. B. S. Mordukhovich and N. M. Nam (2005), Variational stability and marginal functions via generalized differentiation, *Math. Oper. Res.* **30**, 1–18.
935. B. S. Mordukhovich and N. M. Nam (2005), Subgradients of distance functions with some applications, *Math. Progr.*, to appear.
936. B. S. Mordukhovich and N. M. Nam (2005), Subgradients of distance functions at out-of-state points, *Taiwan. J. Math.*, to appear.
937. B. S. Mordukhovich, N. M. Nam and N. D. Yen (2005), Subgradients of marginal functions in parametric mathematical programming, *Math. Progr.*, to appear.
938. B. S. Mordukhovich, N. M. Nam and N. D. Yen (2005), Fréchet subdifferential calculus and optimality conditions in mathematical programming, *Optimization*, to appear.
939. B. S. Mordukhovich and J. V. Outrata (2001), Second-order subdifferentials and their applications, *SIAM J. Optim.* **12**, 139–169.
940. B. S. Mordukhovich, J. V. Outrata and M. Červinka (2005), Equilibrium problems with complementarity constraints: Case study with applications to oligopolistic markets, *Optimization*, to appear.
941. B. S. Mordukhovich and T. Pennanen (2005), Epi-convergent approximations for generalized Bolza problems, preprint.
942. B. S. Mordukhovich and V. M. Raketskii (1980), Necessary optimality conditions of the maximum principle type in discrete approximations of nonconvex control problems with constraints on trajectories, *Vest. Beloruss. Univ.*, Ser. 1, No. 3, 72–73; Depon. VINITI #791-79, Moscow.
943. B. S. Mordukhovich and J.-P. Raymond (2004), Dirichlet boundary control of hyperbolic equations in the presence of state constraints, *Appl. Math. Optim.* **49**, 145–157.
944. B. S. Mordukhovich and J.-P. Raymond (2004), Neumann boundary control of hyperbolic equations with pointwise state constraints, *SIAM J. Control Optim.* **43**, 1354–1372.
945. B. S. Mordukhovich and A. M. Sasonkin (1985), Duality and optimality conditions in control problems for functional-differential systems of neutral type, *Diff. Eq.* **21**, 532–340.
946. B. S. Mordukhovich and Y. Shao (1995), Differential characterizations of covering, metric regularity, and Lipschitzian properties of multifunctions between Banach spaces, *Nonlinear Anal.* **25**, 1401–1424.
947. B. S. Mordukhovich and Y. Shao (1995), On nonconvex subdifferential calculus in Banach spaces, *J. Convex Anal.* **2**, 211–227.
948. B. S. Mordukhovich and Y. Shao (1996), Extremal characterizations of Asplund spaces, *Proc. Amer. Math. Soc.* **124**, 197–205.
949. B. S. Mordukhovich and Y. Shao (1996), Nonsmooth sequential analysis in Asplund spaces, *Trans. Amer. Math. Soc.* **348**, 1235–1280.
950. B. S. Mordukhovich and Y. Shao (1996), Nonconvex coderivative calculus for infinite-dimensional multifunctions, *Set-Valued Anal.* **4**, 205–236.
951. B. S. Mordukhovich and Y. Shao (1997), Stability of multifunctions in infinite dimensions: Point criteria and applications, *SIAM J. Control Optim.* **35**, 285–314.
952. B. S. Mordukhovich and Y. Shao (1997), Fuzzy calculus for coderivatives of multifunctions, *Nonlinear Anal.* **29**, 605–626.
953. B. S. Mordukhovich and Y. Shao (1998), Mixed coderivatives of set-valued mappings in variational analysis, *J. Appl. Anal.* **4**, 269–294.

954. B. S. Mordukhovich, Y. Shao and Q. J. Zhu (2000), Viscosity coderivatives and their limiting behavior in smooth Banach spaces, *Positivity* **4**, 1–39.
955. B. S. Mordukhovich and I. A. Shvartsman (2002), Discrete maximum principle for nonsmooth optimal control problems with delays, *Cybernet. Systems Anal.* **38**, 255–264.
956. B. S. Mordukhovich and I. A. Shvartsman (2004), The approximate maximum principle in constrained optimal control, *SIAM J. Control Optim.* **43**, 1037–1062.
957. B. S. Mordukhovich and I. A. Shvartsman (2004), Optimization and feedback control of constrained parabolic systems under uncertain perturbations, in *Optimal Control, Stabilization and Nonsmooth Analysis*, edited by M. de Queiroz, M. Malisoff and P. Wolenski, Lecture Notes Cont. Inf. Sci. **301**, pp. 121–132, Springer, New York.
958. B. S. Mordukhovich, J. S. Treiman and Q. J. Zhu (2003), An extended extremal principle with applications to multiobjective optimization, *SIAM J. Optim.* **14**, 359–379.
959. B. S. Mordukhovich and R. Trubnik (2001), Stability of discrete approximations and necessary optimality conditions for delay-differential inclusions, *Ann. Oper. Res.* **101**, 149–170.
960. B. S. Mordukhovich and B. Wang (2000), On variational characterizations of Asplund spaces, in *Constructive, Experimental and Nonlinear Analysis*, edited by M. Théra, Canad. Math. Soc. Conf. Proc. **27**, 245–254.
961. B. S. Mordukhovich and B. Wang (2001), Sequential normal compactness in variational analysis, *Nonlinear Anal.* **47**, 717–728.
962. B. S. Mordukhovich and B. Wang (2002), Necessary optimality and suboptimality conditions in nondifferentiable programming via variational principles, *SIAM J. Control Optim.* **41**, 623–640.
963. B. S. Mordukhovich and B. Wang (2002), Extensions of generalized differential calculus in Asplund spaces, *J. Math. Anal. Appl.* **272**, 164–186.
964. B. S. Mordukhovich and B. Wang (2003), Calculus of sequential normal compactness in variational analysis, *J. Math. Anal. Appl.* **282**, 63–84.
965. B. S. Mordukhovich and B. Wang (2003), Differentiability and regularity of Lipschitzian mappings, *Proc. Amer. Math. Soc.* **131**, 389–399.
966. B. S. Mordukhovich and B. Wang (2004), Generalized differentiation for moving objects, in *Optimal Control, Stabilization and Nonsmooth Analysis*, edited by M. de Queiroz, M. Malisoff and P. Wolenski, Lecture Notes Cont. Inf. Sci. **301**, pp. 351-361, Springer, New York.
967. B. S. Mordukhovich and B. Wang (2004), Restrictive metric regularity and generalized differential calculus in Banach spaces, *Int. J. Maths. Math. Sci.* **50**, 2650–2683.
968. B. S. Mordukhovich and B. Wang (2005), Restrictive metric regularity in variational analysis, *Nonlinear Anal.*, to appear.
969. B. S. Mordukhovich and B. Wang (2005), Generalized differential and normal compactness calculi for moving objects, preprint.
970. B. S. Mordukhovich and D. Wang (2005), Optimal control of semilinear unbounded differential inclusions, *Nonlinear Anal.*, to appear.
971. B. S. Mordukhovich and D. Wang (2005), Optimal control of semilinear evolution inclusions via discrete approximations, *Control Cybernet.* **34**, No. 3.

972. B. S. Mordukhovich and L. Wang (2002), Optimal control of hereditary differential inclusions, in *Proc. 41st IEEE Conf. Decis. Control*, pp. 1107–1112, Las Vegas, Nevada.
973. B. S. Mordukhovich and L. Wang (2003), Optimal control of constrained delay-differential inclusions with multivalued initial conditions, *Control Cybernet.* **32**, 585–609.
974. B. S. Mordukhovich and L. Wang (2004), Optimal control of neutral functional-differential inclusions, *SIAM J. Control Optim.* **43**, 111–136.
975. B. S. Mordukhovich and L. Wang (2004), Optimal control of differential-algebraic inclusions, in *Optimal Control, Stabilization and Nonsmooth Analysis*, edited by M. de Queiroz, M. Malisoff and P. Wolenski, Lecture Notes Cont. Inf. Sci. **301**, pp. 73–83, Springer, New York.
976. B. S. Mordukhovich and L. Wang (2005), Optimal control of delay systems with differential and algebraic dynamic constraints, *ESAIM: Control Optim. Calc. Var.* **11**, 285–309.
977. B. S. Mordukhovich and L. Wang (2005), Optimal control of nonautonomous functional-differential inclusions of neutral type, *Nonlinear Anal.*, to appear.
978. B. S. Mordukhovich and K. Zhang (1997), Minimax control of parabolic equations with Dirichlet boundary conditions and state constraints, *Appl. Math. Optim.* **36**, 323–360.
979. B. S. Mordukhovich and K. Zhang (1998), Dirichlet boundary control of parabolic systems with pointwise state constraints, *Int. Ser. Numer. Math.* **126**, 223–236.
980. J.-J. Moreau (1962), Fonctions convexes duales et points proximaux dans un espace hilbertien, *C. R. Acad. Sci. Paris* **255**, 2897–2899.
981. J.-J. Moreau (1963), Fonctionelles sous-différentiables, *C. R. Acad. Sci. Paris* **257**, 4117–4119.
982. J.-J. Moreau (1965), Proximité dualité dans espace hilbertien, *Bull. Soc. Math. France* **93**, 273–299.
983. J.-J. Moreau (1977), Evolution problems associated with moving coinvex set in a Hilbert space, *J. Diff. Eq.* **26**, 347–374.
984. J.-J. Moreau (1988), Bounded variation in time, in *Topics in Nonsmooth Mechanics*, edited by J.-J. Moreau, P. D. Panagiotopoulos and G. Strang, pp. 1–74, Birkhäuser, Boston, Massachusetts.
985. D. Motreanu and N. Pavel (1999), *Tangency, Flow-Invariance for Differential Equations, and Optimization Problems*, Marcel Dekker, New York.
986. M. Moussaoui and A. Seeger (1994), Sensitivity analysis of optimal value functions of convex parametric programs with possibly empty solution sets, *SIAM J. Optim.* **3**, 659–675.
987. M. Moussaoui and A. Seeger (1996), Epsilon-maximum principle of Pontryagin type and perturbation analysis of convex optimal control problems, *SIAM J. Control Optim.* **34**, 407–427.
988. M. Mureşan (2001), *Analiză Nenetedăsi Aplicatii*, Risoprint, Cluj-Napoca, Romania.
989. A. D. Myshkis (1951), *Linear Differential Equations with Retarded Argument*, GITTL, Moscow.
990. N. Nagumo (1942), Ber diw lage der integralkurven gewohlicher differenialgleichungen, *Proc. Phys. Math. Soc. Japan* **24**, 551–559.

991. N. M. Nam and N. D. Yen (2005), Relationships between approximate Jacobians and coderivatives, *J. Nonlinear Convex Anal.*, to appear.
992. I. Namioka and R. R. Phelps (1975), Banach spaces which are Asplund spaces, *Duke Math. J.* **42**, 735–750.
993. Z. Naniewicz (2005), Pseudo-monotone approach to economic equilibrium problem in reflexive Banach space, preprint.
994. Z. Naniewicz and P. D. Panagiotopoulos (1995), *Mathematical Theory of Hemivariational Inequalities and Applications*, Marcel Dekker, New York.
995. J. F. Nash (1951), Non-cooperative games, *Annals Math.* **54**, 286–295.
996. M. H. Nayakkankuppam and M. L. Overton (1999), Conditioning of semilinear programs, *Math. Progr.* **85**, 525–540.
997. P. Neittaaanmäki and D. Tiba (1994), *Optimal Control of Nonlinear Parabolic Systems*, Marcel Dekker, New York.
998. Y. E. Nesterov (2005), Lexicographical differentiation of nonsmooth functions, *Math. Progr.*, to appear.
999. Y. E. Nesterov and A. S. Nemirovsky (1993), *Interior Point Methods in Convex Optimization: Theory and Applications*, SIAM Publications, Philadelphia, Pennsylvania.
1000. J. von Neumann and O. Morgenstern (1944), *Theory of Games and Economic Behaviour*, Princeton University Press, Princeton, New Jersey.
1001. L. W. Neustadt (1966), An abstract variational theory with applications to a broad class of optimization problems, I: General theory, *SIAM J. Control* **4**, 505–527.
1002. L. W. Neustadt (1976), *Optimization: a Theory of Necessary Conditions*, Princeton University Press, Princeton, New Jersey.
1003. K. F. Ng and W. H. Yang (2004), Regularities and their relations to error bounds, *Math. Progr.* **99**, 521–538.
1004. K. F. Ng and X. Y. Zheng (2001), Error bounds for lower semicontinuous functions in normed spaces, SIAM J. Optim. **12**, 1–17.
1005. K. F. Ng and X. Y. Zheng (2003), Global weak sharp minima in Banach spaces, *SIAM J. Control Optim.* **41**, 1858–1885.
1006. H. V. Ngai, D. T. Luc and M. Théra(2000), Approximate convex functions, *J. Nonlinear Convex Anal.* **1**, 155–176.
1007. H. V. Ngai, D. T. Luc and M. Théra (2002), Extensions of Fréchet ε-subdifferential calculus and applications, *J. Math. Anal. Appl.* **268**, 266–290.
1008. N. V. Ngai and M. Théra (2001), Metric regularity, subdifferential calculus and applications, *Set-Valued Anal.* **9**, 187–216.
1009. N. V. Ngai and M. Théra (2002), A fuzzy necessary optimality condition for non-Lipschitz optimization in Asplund spaces, *SIAM J. Optim.* **12**, 656–668.
1010. N. V. Ngai and M. Théra (2004), Error bounds, and implicit multifunction theorem in smooth Banach spaces and applications to optimization, *Set-Valued Anal.* **12**, 195–223.
1011. A. Nijenhuis (1974), Strong derivatives and inverse mappings, *Amer. Math. Monthly* **81**, 969–980.
1012. M. S. Nikolskii (2004), Difference approximation of some optimization problems, *Comput. Maths. Math. Phys.* **44**, 467–475.
1013. K. Nitka-Styczen (1988), Difference approximations of optimal periodic control problems, *Int. J. Control* **47**, 1893–1904.

1014. K. Nitka-Styczen (1989), Approximate discrete maximum principle for the discrete approximation of optimal periodic control problems, *Int. J. Control* **50**, 1863–1871.
1015. K. Nitka-Styczen (1999), *Optimal Periodic Control of Hereditary Processes. Theory, Algorithms and Applications*, Technical University of Wrocław, Wrocław, Poland.
1016. A. Nowakowski and I. Nowakowska (2005), Dirichlet problem for semilinear hyperbolic equation, *Nonlinear Anal.*, to appear.
1017. E. A. Nurminskii (1979), *Methods of Solutions to Deterministic and Stochastic Minimax Problems*, Naukova Dumka, Kiev.
1018. M. N. Oğuztöreli (1966), *Time-Lag Control Systems*, Academic Press, New York.
1019. A. Olbrot (1976), Control of retarded systems with function space constraints: Necessary optimality conditions, *Control Cybernet.* **5**, 5–31.
1020. C. Olech (1976), Existence theory in optimal control problems, in *Control Theory and Topics in Functional Analysis*, Vol. 1, pp. 291–328, International Atomic Energy Agency, Vienna.
1021. N. Ortiz (2005), Necessary conditions for the neutral problem of Bolza with continuously varying time delay, *J. Math. Anal. Appl.* **305**, 513–527.
1022. N. Ortiz and P. R. Wolenski (2004), The decoupling technique for continuously varying time delay systems, *Set-Valued Anal.* **12**, 225–239.
1023. Y. S. Osipov, L. Pandolfi and V. I. Maksimov (2001), Problems of dynamical reconstruction and robust boundary control: The case of the Dirichlet boundary conditions, *J. Inverse Ill-Posed Probl.* **9**, 149–162.
1024. J. V. Outrata (1999), Optimality conditions for a class of mathematical programs with equilibrium constraints, *Math. Oper. Res.* **24**, 627–644.
1025. J. V. Outrata (1999), Optimality conditions for a class of mathematical programs with equilibrium constraints: Strongly regular case, *Kybernetika* **35**, 177–193.
1026. J. V. Outrata (2000), On mathematical programs with complementarity constraints, *Optim. Meth. Soft.* **14**, 117–137.
1027. J. V. Outrata (2000), A generalized mathematical program with equilibrium constraints, *SIAM J. Control Optim.* **38**, 1623–1638.
1028. J. V. Outrata (2001), On constraint qualifications for mathematical programs with mixed complementarity constraints, in *Complementarity, Applications and Extensions*, edited by M. C. Ferris, O. L. Mangasarian and J.-S. Pang, pp. 253–272, Kluwer, Dordrecht, The Netherlands.
1029. J. V. Outrata (2004), A note on a class of equilibrium problems with equilibrium constraints, *Kybernetika* **40**, 585–594.
1030. J. V. Outrata (2005), Mathematical programs with equilibrium constraints: Theory and numerical methods, in *Nonsmooth Mechanics of Solids, CISM Lecture Notes*, edited by J. Haslinger and G. E. Stavroulakis, Springer, New York.
1031. J. V. Outrata, M. Kočvara and J. Zowe (1998), *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*, Kluwer, Dordrecht, The Netherlands.
1032. J. V. Outrata and W. Römisch (2005), On optimality conditions for some nonsmooth optimization problems over L^p spaces, *Optim. Theory Appl.* **126**, 1–28.

1033. M. L. Overton (1988), On minimizing the maximum eigenvalue of a symmetric matrix, *SIAM J. Matrix Anal. Appl.* **9**, 256–268.
1034. M. L. Overton and R. S. Womersley (1993), Optimality conditions and duality theory for minimizing sums of the largest eigenvalue of symmetric matrices, *Math. Progr.* **62**, 321–337.
1035. A. E. Ozdaglar and D. P. Bertsekas (2004), The relation between pseudonormality and quasiregularity in constrained optimization, *Optim. Meth. Soft.* **19**, 493–506.
1036. Z. Páles and V. Zeidan (1994), First and second order necessary conditions for control problems with constraints, *Trans. Amer. Math. Soc.* **346**, 421–453.
1037. Z. Páles and V. Zeidan (1996), Generalized Hessians for $C^{1,1}$ functions in infinite dimensional normed spaces, *Math. Progr.* **74**, 59–78.
1038. Z. Páles and V. Zeidan (2004), Critical and critical tangent cones in optimization problems, *Set-Valued Anal.* **12**, 241–258.
1039. D. Pallaschke, P. Recht and R. Urbański (1987), On extensions of the second-order derivative, *Bull. Polish Acad. Sci., Ser. Math.* **35**, 750–763.
1040. D. Pallaschke and S. Rolewicz (1998), *Foundations of Mathematical Optimization: Convex Analysis without Linearity*, Kluwer, Dordrecht, The Netherlands.
1041. D. Pallaschke and R. Urbański (2002), *Pairs of Compact Convex Sets*, Kluwer, Dordrecht, The Netherlands.
1042. P. D. Panagiotopoulos (1993), *Hemivariational Inequalities: Applications to Mechanics and Engineering*, Springer, Berlin.
1043. J.-S. Pang (1990), Solution differentiability and continuation of Newton's method for variational inequality problems over polyhedral sets, *J. Optim. Theory Appl.* **66**, 121–135.
1044. J.-S. Pang (1993), A degree-theoretical approach to parametric nonsmooth equations with multivalued perturbed solution sets, *Math. Progr.* **62**, 359–383.
1045. J.-S. Pang (1997), Error bounds in mathematical programming, *Math. Progr.* **79**, 299–332.
1046. J.-S. Pang and D. Stewart (2005), Differential variational inequalities, *Math. Progr.*, to appear.
1047. J.-S. Pang and L. Qi (1993), Nonsmooth equations: Motivations and algorithms, *SIAM J. Optim.* **3**, 443–465.
1048. C. Pantelides, D. Gritsis, K. P. Morison and R. W. H. Sargent (1988), The mathematical modelling of transient systems using differential-algebraic equations, *Comput. Chem. Eng.* **12**, 449–454.
1049. N. S. Papageorgiou (1995), On parametric evolution inclusions of the subdifferential type with applications to optimal control problems, *Trans. Amer. Math. Soc.* **347**, 203–231.
1050. M. Papi and S. Sbaraglia (2003), Regularity properties of constrained set-valued mappings, *Nonlinear Anal.* **54**, 1337–1353.
1051. M. Papi and S. Sbaraglia (2005), Lipschitzian estimates in discrete-time constrained stochastic optimal control, *Dynamics Cont. Disc. Impuls. Syst.*, to appear.
1052. M. Papi and S. Sbaraglia (2005), Optimal asset-liability management with constraints: A dynamic programming approach, *Appl. Math. Comput.*, to appear.
1053. V. Pareto (1909), *Manuel d'Economie Politique*, Giard, Paris.

1054. G. Peano (1892), Sur la définition de la dérivée, *Mathesis* **2**, 12–14.
1055. J. Peña (2000), Understanding the geometry of infeasible perturbations of a conic linear system, *SIAM J. Optim.* **10**, 534–550.
1056. J. Peña (2001), Conditioning of convex programs from a primal-dual perspectives, *Math. Oper. Res.* **26**, 206–220.
1057. J. Peña (2003), A characterization of the distance to infeasibility under structured perturbations, *Linear Algebra Appl.* **370**, 193–216.
1058. J. Peña (2004), Conic systems and sublinear mappings: Equivalent approaches, *Operations Research Letters* **32** (2004), 463–467.
1059. J. Peña (2005), On the block-structured distance to non-surjectivity of sublinear mappings, *Math. Progr.* **103**, 561–573.
1060. T. Pennanen (1999), Graph-convex mappings and K-convex functions, *J. Convex Anal.* **6**, 235–266.
1061. T. Pennanen (2005), Epi-convergent discretizations of multistage stochastic programs, *Math. Oper. Res.* **30**, 245–256.
1062. T. Pennanen (2005), Epi-convergent discretizations of multistage stochastic programs via integration quadratures, *Math. Progr.*, to appear.
1063. J.-P. Penot (1974), Sous-différentieles de fonctions numériques nonconvexes, *C. R. Acad. Paris* **278**, 1553–1555.
1064. J.-P. Penot (1978), Calcul sous-différentiel et optimisation, *J. Funct. Anal.* **27**, 248–276.
1065. J.-P. Penot (1981), A characterization of tangential regularity, *Nonlinear Anal.* **5**, 625–643.
1066. J.-P. Penot (1989), Metric regularity, openness and Lipschitzian behavior of multifunctions, *Nonlinear Anal.* **13**, 629–643.
1067. J.-P. Penot (1994), Sub-hessians, super-hessions and conjugation, *Nonlinear Anal.* **23**, 689–702.
1068. J.-P. Penot (1995), Inverse function theorems for mappings and multimappings, *Southeast. Asian Bull. Math.* **19**, 1–16.
1069. J.-P. Penot (1996), Favorable classes of mappings and multimappings in nonlinear analysis and optimization, *J. Convex Anal.* **3**, 97–116.
1070. J.-P. Penot (1997), Metric estimates for the calculus of multimappings, *Set-Valued Anal.* **5**, 291–308.
1071. J.-P. Penot (1998), Compactness properties, openness criteria and coderivatives, *Set-Valued Anal.* **6**, 363–380.
1072. J.-P. Penot (1998), Are generalized derivatives useful for generalized convex functions?, *Generalized Convexity, Generalized Monotonicity: Recent Results*, edited by J.-P. Crouzeix, J. E. Martinez-Legaz and M. Volle, pp. 3–59, Kluwer, Dordrecht, The Netherlands.
1073. R. R. Phelps (1993), *Convex Functions, Monotone Operators and Differentiability*, 2nd edition, Springer, Berlin.
1074. M. d. R. de Pinho (2003), Mixed constrained control problems, *J. Math. Anal. Appl.* **278**, 293–307.
1075. M. d. R. de Pinho, M. M. A. Ferreira and F. A. C. C. Fontes (2002), An Euler-Lagrange inclusion for optimal control problems with state constraints, *J. Dynam. Control Systems* **8**, 23–45.
1076. M. d. R. de Pinho, M. M. A. Ferreira and F. A. C. C. Fontes (2005), Unmaximized inclusion necessary conditions for nonconvex constrained optimal control problems, *ESAIM: Control Optim. Calc. Var.* **11**.

1077. M. d. R. de Pinho and A. Ilchmann (2002), Weak maximum principle for optimal control problems with mixed constraints, *Nonlinear Anal.* **48**, 1179–1196.
1078. M. d. R. de Pinho and R. B. Vinter (1995), An Euler-Lagrange inclusion for optimal control problems, *IEEE Trans. Autom. Control* **40**, 1191–1198.
1079. M. d. R. de Pinho and R. B. Vinter (1997), Necessary conditions for optimal control problems involving nonlinear differential algebraic equations, *J. Math. Anal. Appl.* **212**, 493–516.
1080. M. d. R. de Pinho, R. B. Vinter and H. Zheng (2001), A maximum principle for optimal control problems with mixed constraints, *IMA J. Math. Control Inf.* **18**, 189–205.
1081. S. Plaskasz (1992), On the solution sets for differential inclusions, *Boll. Un. Mat. Ital.* **6**, 387–394.
1082. V. A. Plotnikov, A. V. Plotnikov and A. N. Vityuk (1999), *Differential Equations with a Multivalued Right-Hand Side: Asymptotic Methods*, AstroPrint, Odessa, Ukraine.
1083. V. I. Plotnikov (1972), Necessary and sufficient conditions for optimality and conditions for uniqueness of the optimizing functions for control systems of general form, *Izv. Akad. Nauk SSSR, Ser. Mat.* **36**, 652–679.
1084. V. I. Plotnikov and M. I. Sumin (1982), The construction of minimizing sequences in control problems for systems with distributed paratemers, *USSR Comput. Maths. Math. Phys.* **22**, 292–296.
1085. V. I. Plotnikov and V. I. Sumin (1972), Optimization of objects with distributed parameters that can be described by Coursat-Darboux systems, *USSR Comput. Maths. Math. Phys.* **12**, 49–57.
1086. E. Polak (1999), *Optimization: Algorithms and Consistent Approximations*, Springer, New York.
1087. R. A. Poliquin (1990), Subgradient monotonicity and convex functions, *Nonlinear Anal.* **14**, 305–317.
1088. R. A. Poliquin (1992), An extension of Attouch's theorem and its application to second-order epi-differentiation of convexly composite functions, *Trans. Amer. Math. Soc.* **332**, 861–874.
1089. R. A. Poliquin and R. T. Rockafellar (1992), Amenable functions in optimization, in *Nonsmooth Optimization Methods and Applications*, edited by F. Giannessi, pp. 338–353, Gordon and Breach, Philadelphia, Pennsylvania.
1090. R. A. Poliquin and R. T. Rockafellar (1996), Prox-regular functions in variational analysis, *Trans. Amer. Math. Soc.* **348**, 1805–1838.
1091. R. A. Poliquin and R. T. Rockafellar (1996), Generalized Hessian properties of regularized nonsmooth functions, *SIAM J. Optim.* **6**, 1121–1137.
1092. R. A. Poliquin and R. T. Rockafellar (1998), Tilt stability of a local minimum, *SIAM J. Optim.* **8**, 287–299.
1093. R. A. Poliquin, R. T. Rockafellar and L. Thibault (2000), Local differentiability of distance functions, *Trans. Amer. Math. Soc.* **332**, 5231–5249.
1094. E. S. Polovinkin and G. V. Smirnov (1986), An approach to differentiation of multifunctions and necessary optimality conditions for differential inclusions, *Diff. Eq.* **22**, 660–668.
1095. E. S. Polovinkin and G. V. Smirnov (1986), Time-optimal problem for differential inclusions, *Diff. Eq.* **22**, 940–952.
1096. B. T. Polyak (1969), Semicontinuity of intergral functionals and existence theorems on extremal problems, *Math. Sbornik* **7**, 59–77.

1097. B. T. Polyak (1987), *Introduction to Optimization*, Optimization Software, New York.
1098. B. T. Polyak (1998), Convexity of quadratic transformations and its use in control and optimization, *J. Optim. Theory Appl.* **99**, 553–583.
1099. B. T. Polyak (2000), History of mathematical programming in the USSR: Analyzing the phenomenon, *Math. Progr.* **91**, 401–416.
1100. B. T. Polyak (2001), Convexity of nonlinear image of a small ball with applicatiobs to optimization, *Set-Valued Anal.* **9**, 159–168.
1101. D. Pompeiu (1905), Fonctions de variables complexes, *Ann. Facul. Sci. Univ. Toulouse* **7**, 265–315.
1102. L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze and E. F. Mishchenko (1962), *The Mathematical Theory of Optimal Processes*, Wiley, New York, 1962.
1103. D. Preiss (1984), Gâteaux differentiable functions are somewhere Fréchet differentiable, *Rend. Circ. Mat. Palermo* **33**, 122–133.
1104. D. Preiss (1990), Differentiability of Lipschitz functions on Banach spaces, *J. Funct. Anal.* **91**, 312–345.
1105. A. I. Propoi (1973), *Elements of the Theory of Optimal Discrete Processes*, Nauka, Moscow.
1106. B. N. Pshenichnyi (1971), *Necessary Conditions for an Extremum*, Marcel Dekker, New York.
1107. B. N. Pshenichnyi (1976), Necessary conditions for an extremum for differential inclusions, *Kibernetika* **12**, 60–73.
1108. B. N. Pshenichnyi (1977), On necessary extremality conditions for nonsmooth functions, *Kibernetika* **13**, 92–96.
1109. B. N. Pshenichnyi (1980), *Convex Analysis and Extremal Problems*, Nauka, Moscow.
1110. V. Pták (1974), A quantitative refinement of the closed graph theorem, *Czech. Math. J.* **24**, 503–506.
1111. L. Qi and J. Sun (1993), A nonsmooth version of Newton's method, *Math. Progr.* **58**, 353–368.
1112. Y. P. Qiu and T. L. Magnanti (1992), Sensitivity analysis for variational inequalities, *Math. Oper. Res.* **17**, 61–76.
1113. M. Quinzii (1992), *Increasing Returns and Efficiency*, Oxford University Press, Oxford, UK.
1114. H. Rademacher (1919), Uber partielle und totale differenzierbarkeit von funktionen mehrerer varieabeln und über die transformation der doppelintegrale, *Math. Ann.* **79**, 340–359.
1115. D. Ralph (1994), A chain rule for nonsmooth composite functions via minimization, *Bull. Austral. Math. Soc.* **49**, 129–137.
1116. D. Ralph (2002), A stable homotopy approach to horizontal linear complementarity problems, *Control Cybernet.* **31**, 575–599.
1117. D. Ralph and S. J. Wright (2004), Some properties of regularization and penalization schemes for MPECs, *Optim. Meth. Soft.* **5**, 527–556.
1118. F. Rampazzo and R. B. Vinter (2000), Degenerate optimal control problems with state constraints, *SIAM J. Control Optim.* **39**, 989–1007.
1119. T. Rapcsák (1997), *Smooth Nonlinear Optimization in* $I\!R^n$, Kluwer, Dordrecht, The Netherlands.

1120. J.-P. Raymond (1997), Nonlinear boundary control of semilinear parabolic problems with pointwise state constraints, *Disc. Cont. Dynam. Systems* **3**, 341–370.
1121. J.-P. Raymond and H. Zidani (1998), Pontryagin's principle for state-constrained control problems governed by parabolic equations with unbounded controls, *SIAM J. Control Optim.* **38**, 1853–1879.
1122. J. Renegar (1995), Incorporating condition measures into the complexity theory of linear programming, *SIAM J. Optim.* **5**, 506–524.
1123. J. Renegar (1995), Linear programming, complexity theory and elementary functional analysis, *Math. Progr.* **70**, 279–540.
1124. L. Rifford (2002), Semiconcave control Lyapunov functions and stabilizing feedback, *SIAM J. Control Optim.* **41**, 659–681.
1125. S. M. Robinson (1972), Normed convex processes, *Trans. Amer. Math. Soc.* **174**, 127–140.
1126. S. M. Robinson (1973), Bounds for error in the solution set of a perturbed linear program, *Linear Algebra Appl.* **6**, 69–81.
1127. S. M. Robinson (1975), An application of error bounds for convex programming in a linear space, *SIAM J. Control* **13**, 271–273.
1128. S. M. Robinson (1976), Stability theory for systems of inequalities, II: Differentiable nonlinear systems, *SIAM J. Numer. Anal.* **13**, 130–143.
1129. S. M. Robinson (1976), Regularity and stability for convex multivalued functions, *Math. Oper. Res.* **1**, 130–143.
1130. S. M. Robinson (1979), Generalized equations and their solutions, I: Basic theory, *Math. Progr. Study* **10**, 128–141.
1131. S. M. Robinson (1980), Strongly regular generalized equations, *Math. Oper. Res.* **5**, 43–62.
1132. S. M. Robinson (1981), Some continuity properties of polyhedral multifunctions, *Math. Progr. Study* **14**, 206–214.
1133. S. M. Robinson (1982), Generalized equations and their solutions, II: Applications to nonlinear programming, *Math. Progr. Study* **19**, 200–221.
1134. S. M. Robinson (1983), Generalized equations, in *Mathematical Programming: The State of the Art*, edited by A. Bachem et al., pp. 346–367, Springer, Berlin.
1135. S. M. Robinson (1987), Local epi-continuity and local optimization, *Math. Progr. Study* **37**, 208–222.
1136. S. M. Robinson (1991), An implicit function theorem for a class of nonsmooth functions, *Math. Oper. Res.* **16**, 292–309.
1137. S. M. Robinson (2003), Constraint nondegeneracy in variational analysis, *Math. Oper. Res.* **28**, 201–232.
1138. S. M. Robinson (2003), Variational conditions with smooth constraints: Structure and analysis, *Math. Progr.* **97**, 245–265.
1139. S. M. Robinson (2004), Localized normal maps and the stability of variational conditions, *Set-Valued Anal.* **12**, 259–274.
1140. R. T. Rockafellar (1963), *Convex Functions and Dual Extremum Problems*, Ph.D. dissertation, Department of Mathematics, Harvard University, Cambridge, Massachusetts.
1141. R. T. Rockafellar (1966), Characterization of subdifferentials of convex functions, *Pacific J. Math.* **17**, 497–510.
1142. R. T. Rockafellar (1970), *Convex Analysis*, Princeton University Press, Princeton, New Jersey.

1143. R. T. Rockafellar (1970), Conjugate convex functions in optimal control and the calculus of variations, *J. Math. Anal. Appl.* **32**, 174–222.

1144. R. T. Rockafellar (1970), Generalized Hamiltonian equations for convex problems of Lagrange, *Pacific J. Math.* **33**, 411–427.

1145. R. T. Rockafellar (1971), Existence and duality theorems for convex problems of Bolza, *Trans. Amer. Math. Soc.* **159**, 1–39.

1146. R. T. Rockafellar (1979), Clarke's tangent cone and the boundaries of closed sets in $\mathbb{R}^n$, *Nonlinear Anal.* **3**, 145–154.

1147. R. T. Rockafellar (1979), Directional Lipschitzian functions and subdifferential calculus, *Proc. London Math. Soc.* **39**, 331–355.

1148. R. T. Rockafellar (1980), Generalized directional derivatives and subgradients of nonconvex functions, *Canad. J. Math.* **32**, 157–180.

1149. R. T. Rockafellar (1981), *The Theory of Subgradients and Its Applications to Problems of Optimization: Convex and Nonconvex Functions*, Helderman Verlag, Berlin.

1150. R. T. Rockafellar (1981), Proximal subgradients, marginal values and augmented Lagrangians in nonconvex optimization, *Math. Oper. Res.* **6**, 424–436.

1151. R. T. Rockafellar (1982), Favorable classes of Lipschitz continuous functions in subgradient optimization, in *Progress in Nondifferentiable Optimization*, edited by E. A. Nurminskii, pp. 125–143, IIASA, Laxenburg, Austria.

1152. R. T. Rockafellar (1982), Lagrange multipliers and subderivatives of optimal value functions in nonlinear programming, *Math. Progr. Study* **17**, 28–66.

1153. R. T. Rockafellar (1985), Maximal monotone relations and the second derivatives of nonsmooth functions, *Ann. Inst. H. Poincaré: Analyse Non Linéaire* **2**, 167–184.

1154. R. T. Rockafellar (1985), Lipschitzian properties of multifunctions, *Nonlinear Anal.* **9**, 867–885.

1155. R. T. Rockafellar (1985), Extensions of subgradient calculus with applications to optimization, *Nonlinear Anal.* **9**, 665–698.

1156. R. T. Rockafellar (1988), First- and second-order epi-differentiability in nonlinear programming, *Trans. Amer. Math. Soc.* **307**, 75–108.

1157. R. T. Rockafellar (1989), Letter to B. S. Mordukhovich of March 6, 1989, Seattle, Washington.

1158. R. T. Rockafellar (1989), Derivation of some improved formulas in subdifferential calculus, privately circulated note.

1159. R. T. Rockafellar (1989), Proto-differentiability of set-valued mappings and its applications in optimization, in *Analyse Non Linéaire*, edited by H. Attouch et al., pp. 449–482, Gauthier-Villars, Paris.

1160. R. T. Rockafellar (1993), Lagrange multipliers and optimality, *SIAM Rev.* **35**, 183–238.

1161. R. T. Rockafellar (1993), Dualization of subgradient conditions for optimality, *Nonlinear Anal.* **20**, 627–642.

1162. R. T. Rockafellar (1996), Equivalent subgradient versions of Hamiltonian and Euler-Lagrange equations in variational analysis, *SIAM J. Control Optim.* **34**, 1300–1315.

1163. R. T. Rockafellar (2000), Second-order convex analysis, *J. Nonlinear Convex Anal.* **1**, 1–16.

1164. R. T. Rockafellar (2004), Hamilton–Jacobi theory and parametric analysis in fully convex problems of optimal control, *J. Global Optim.* **28**, 419–431.

1165. R. T. Rockafellar and R. J-B. Wets (1998), *Variational Analysis*, Springer, Berlin.
1166. R. T. Rockafellar and P. R. Wolenski (2000), Convexity in Hamilton-Jacobi theory, I: Dynamics and duality, *SIAM J. Control Optim.* **39**, 1323–1350.
1167. R. T. Rockafellar and P. R. Wolenski (2000), Convexity in Hamilton-Jacobi theory, II: Envelope representations, *SIAM J. Control Optim.* **39**, 1351–1372.
1168. R. T. Rockafelar and D. Zagrodny (1997), A derivative-coderivative inclusion in second-order nonsmooth analysis, *Set-Valued Anal.* **5**, 1–17.
1169. S. Rolewicz (1979), On paraconvex multifunctions, *Oper. Res. Verfahren* **31**, 540–546.
1170. S. Rolewicz (1979), On γ-paraconvex multifunctions, *Math. Japon.* **24**, 293–300.
1171. S. Rolewicz (2005), On differentiability of strongly $\alpha(\cdot)$-paraconvex functions in non-separable Asplund spaces, *Studia Math.*, **167**, 235–244.
1172. S. Rolewicz (2005), Paraconvex analysis on $C_{\mathbf{E}}^{1,u}$ manifolds, *Optimization*, to appear.
1173. J. F. Rosenblueth and R. B. Vinter (1991), Relaxation procedures for time delay systems, *J. Math. Anal. Appl.* **162**, 542–563.
1174. J. F. Rosenblueth, J. Warga and Q. J. Zhu (1997), On the characterization of properly relaxed delayed controls, *J. Math. Anal. Appl.* **209**, 274–290.
1175. I. M. Ross and F. Fahroo (2004), Legendre pseudospectral approximations of optimal control problems, *Lecture Notes Cont. Inf. Sci.* **295**, pp. 327–342, Springer, New York.
1176. R. Rossi and G. Savaré (2005), Gradient flows of nonconvex functionals in Hilbert spaces and applications, *ESAIM: Control Optim. Cal. Var.*, to appear.
1177. T. Roubíček (1997), *Relaxation in Optimization Theory and Variational Calculus*, De Gruyter, Berlin.
1178. J. D. L. Rowland and R. B. Vinter (1992), Pontryagin type conditions for differential inclusions, *J. Math. Anal. Appl.* **165**, 587–597.
1179. J. D. L. Rowland and R. B. Vinter (1993), Dynamic optimization problems with free-time and active state constraints, *SIAM J. Control Optim.* **31**, 677–697.
1180. L. I. Rozonoér (1959), L. S. Pontryagin maximum principle in the theory of optimal systems, I–III, *Autom. Remote Control* **20**, 1288–1302, 1405–1421, 1517–1532.
1181. A. M. Rubinov (2000), *Abstract Convexity and Global Optimization*, Kluwer, Dordrecht, The Netherlands.
1182. A. M. Rubinov (2002), Equilibrium with restriction on exchange, *Cybernetics and Systems Analysis*, No. 2, 55–70.
1183. J.-J. Rückmann (1999), On existence and uniqueness of stationary points in semi-infinite programming, *Math. Progr.* **86**, 387–415.
1184. D. L. Russell (1966), Optimal regulation of linear symmetruc hyperbolic systems with finite dimensional control, *SIAM J. Control* **4**, 276–294.
1185. L. D. Sabbagh (1969), Variational problems with lags, *J. Optim. Theory Appl.* **3**, 34–51.
1186. S. Saks (1937), *Theory of the Integral*, 2nd edition, Hafner Publishing Co., New York.

1187. D. Salamon (1984), *Control and Observation of Neutral Systems*, Pitman, Harlow, Essex, UK.
1188. P. A. Samuelson (1947), *Foundations of Economic Analysis*, Harvard University Press, Cambridge, Massachusetts.
1189. P. A. Samuelson (1954), The pure theory of public expenditures, *Rev. Econ. Stat.* **36**, 387–389.
1190. J. Schauder (1930), Über die umkerhrung linearer, stetiger funktionaloperationen, *Studia Math.* **2**, 1–6.
1191. H. Scheel and S. Scholtes (2000), Mathematical programs with equilibrium constraints: Stationarity, optimality and sensitivity, *Math. Oper. Res.* **25**, 1–22.
1192. S. Scholtes (2001), Convergence properties of a regularization scheme for mathematical programs with complementarity constraints, *SIAM J. Optim.* **11**, 818–936.
1193. S. Scholtes (2002), On the existence and computation of EPEC solutions, talk given at the ICCP Conference in Cambridge, UK.
1194. S. Schotes and M. Stöhr (1999), Exact penalization of mathematical programs with equilibrium constraints, *SIAM J. Control Optim.* **37**, 617–652.
1195. L. Scrimali (2004), Variational inequalities and optimal equilibrium distributions in transportation networks, *Math. Ineq. Appl.* **7**, 439–451.
1196. R. Schultz (2000), Some aspects of stability in stochastic programming, *Ann. Oper. Res.* **100**, 55–84.
1197. L. Schwartz (1966), *Théorie des Distributions*, Hermann, Paris.
1198. A. Seeger (1992), Limiting behavior of the approximate second-order subdifferential of a convex function, *J. Optim. Theory Appl.* **74**, 527–544.
1199. A. Seeger (1994), Approximate Euler-Lagrange inclusion, approximate transversality condition, and sensitivity analysis of convex parametric problems of calculus of variations, *Set-Valued Anal.* **2**, 307–325.
1200. H. S. Sendov (2000), *Variational Spectral Analysis*, Ph.D. dissertation, Department of Combinatorics and Optimization, University of Waterloo, http://etd/uwaterloo.ca/etd/hssendov2000.pdf, Waterloo, Canada.
1201. B. Sendov and V. A. Popov (1988), *The Averaged Moduli of Smoothness*, Wiley, New York.
1202. F. Severi (1930), Su alcune questioni di topologia infinitesimale, *Ann. Soc. Polon. Math.* **9**, 97–108.
1203. A. Shapiro (1988) Sensitivity analysis of nonlinear programs and differentiability property of metric projections, *SIAM J. Control Optim.* **26**, 628–645.
1204. A. Shapiro (1990), On concepts of directional defferentiability, *J. Math. Anal. Appl.* **66**, 477–487.
1205. A. Shapiro (1994), Sensitivity analysis of parameterized programs via generalized equations, *SIAM J. Control Optim.* **32**, 553–571.
1206. A. Shapiro, T. Homem-de-Mello and J. Kim (2002), Conditioning of convex pievewise linear stochastic programs, *Math. Progr.* **94**, 1–19.
1207. N. Z. Shor (1972), On a class of almost-differentiable functions and on a minimization method for functions from this class, *Kibernetika*, No. 4. 65–70.
1208. N. Z. Shor (1985), *Minimization Methods for Non-Differentiable Functions*, Springer, Berlin.
1209. I. A. Shvartsman (2005), New approximation method in the proof of the maximum principle for nonsmooth optimal control problems with state constraints, *J. Math. Anal. Appl.*, to appear.

1210. I. A. Shvartsman and R. B. Vinter (2005), Regularity properties of optimal controls for state constrained problems with time-varying control constraints, *Nonlinear Anal.*, to appear.
1211. G. H. Silva and R. B. Vinter (1998), Necessary conditions for optimal impulsive control problems, *SIAM J. Control Optim.* **35**, 1829–1846.
1212. J. Simon (1987), Compact sets in the space $L^p(0,T;B)$, *Ann Mat. Pura Appl.* **146**, 65–96.
1213. S. Simons (1998), *Minimax and Monotonicity*, Springer, Berlin.
1214. I. Singer (1997), *Abstract Convex Analysis*, Wiley, New York.
1215. G. V. Smirnov (1991), Discrete approximations and optimal solutions to differential inclusions, *Cybernetics* **27**, 101–107.
1216. G. V. Smirnov (2001), *Introduction to the Theory of Differential Inclusions*, American Mathematical Society, Providence, Rhode Island.
1217. A. Smith (1776), *An Inquiry into the Nature and Causes of the Wealth of Nations*, Clarendon Press, Oxford, England.
1218. S. L. Sobolev (1963), *Applications of Functional Analysis in Mathematical Physics*, American Mathematical Society, Providence, Rhode Island.
1219. M. V. Solodov and B. F. Svaiter (2000), Error bounds for proximal point subproblems and associated inexasct proximal point algorithms, *Math. Progr.* **88**, 371–389.
1220. E. Sontag (1999), Stability and stabilization: Discontinuity and the effect of disturbances, in *Nonlinear Analysis, Differential Equations, and Control*, edited by F. H. Clarke and R. J. Stern, pp. 551–598, Kluwer, Dordrecht, The Netherlands.
1221. V. A. Srochko (1984), Optimality conditions of the maximum principle type in Goursat-Darboux systems, *Siberian J. Math.* **25**, 126–132.
1222. H. von Stackelberg (1934), *Marktform und Gleichgewicht*, Springer, Berlin.
1223. G. Stampacchia (1964), Formes bilinéaires coercitives sur les ensemble convexes, *C. R. Acad. Sci. Paris* **258**, 4413–4416.
1224. C. Stegall (1978), Optimization of functions on certain subsets of Banach spaces, *Trans. Amer. Math. Soc.* **236**, 171–176.
1225. O. Stein (2003), *Bilevel Strategies in Semi-Infinite Programming*, Kluwer, Boston, Massachusetts.
1226. A. S. Strekalovsky (1987), On the problem of global extremum, *Soviet Math. Dokl.* **35**, 194–198.
1227. A. S. Strekalovsky (1998), Global optimality conditions for nonconvex optimization, *J. Global Optim.* **12**, 415–434.
1228. A. S. Strekalovsky (2003), *Elements of Nonconvex Optimization*, Nauka, Novosibirsk.
1229. M. Studniarski and D. E. Ward (1999), Weak sharp minima: Characterizations and sufficient conditions, *SIAM J. Control Optim.* **38** (1999), 219–236.
1230. A. I. Subbotin (1995), *Generalized Solutions of First-Order PDEs*, Birkhäuser, Boston, Massuchusetts.
1231. N. N. Subbotina (1989), The maximum principle and the superdifferential of the value function, *Prob. Control Inform. Theory* **18**, 151–160.
1232. F. Sullivan (1981), A characterization of complete metric spaces, *Proc. Amer. Math. Soc.* **83**, 345–346.
1233. M. I. Sumin (2000), Optimal control of semilinear elliptic equations with state constraints: Maximum principle for minimizing sequences, regularity, normality, sensitivity, *Control Cybernet.* **29**, 449–472.

1234. M. I. SUMIN (2001), Suboptimal control of a semilinear elliptic equation with a phase constraint and a boundary control, *Diff. Eq.* **37**, 281–300.
1235. H. J. SUSSMANN (1994), A strong version of the Lojasiewicz maximum principle, in *Optimal Control and Differential Equations*, edited by N. N. Pavel, pp. 293–309, Marcel Dekker, New York.
1236. H. J. SUSSMANN (1998), Geometry and optimal control, in *Mathematical Control Theory*, edited by J. Baillieul and J. C. Willems, pp. 140-198, Springer, New York.
1237. H. J. SUSSMANN (2000), New theories of set-valued differentials and new versions of the maximum principle, in *Nonlinear Control in the Year 2000*, edited by A. Isidori, F. Lamnabhi-Lagarrigue and W. Respondek, pp. 487–526, Springer, Berlin.
1238. H. J. SUSSMANN (2002), Needle variations and almost lower semicontinuous differential inclusions, *Set-Valued Anal.* **10**, 233–285.
1239. H. J. SUSSMANN AND J. V. WILLEMS (1997), 300 years of optimal control: From the brachystochrone to the maximum principle, *IEEE Control Syst. Magaz.* **17**, 32–44.
1240. A. ŚWIECH (1994), Unbounded second order partial differential equations in infinite dimensional Hilbert spaces, *Comm. Part. Diff. Eq.* **19**, 1999-2036.
1241. A. ŚWIECH (1996), Sub- and superoptimality principles of dynamic programming revisited, *Nonlinear Anal.* **26**, 1429–1436.
1242. T. TADUMADZE AND L. ALKHAZISHVILI (2003), Formulas of variations of solutions for nonlinear controlled delay differential equations with discontinuous initial conditions, *Mem. Diff. Eq. Math. Phys.* **29**, 125–150.
1243. W. TAKAHASHI (2000), *Nonlinear Functional Analysis*, Yokohama Publishers, Yokohama, Japan.
1244. L. THIBAULT (1976), Probléme de Bolza dans un espace de Banach séparable, *C. R. Acad. Sci. Paris* **282**, 1303–1306.
1245. L. THIBAULT (1978), Sous-différentiel de fonctions vertorielles compactement lipschitziennes, *C. R. Acad. Sci. Paris* **286**, 995–998.
1246. L. THIBAULT (1980), Subdifferentials of compactly Lipschitzian vector functions, *Ann. Mat. Pura Appl.* **125**, 157–192.
1247. L. THIBAULT (1982), On generalized differentials and subdifferentials of Lipschitz vector-valued functions, *Nonlinear Anal.* **6**, 1037–1053.
1248. L. THIBAULT (1983), Tangent cones and quasi-interiorly tangent cones to multifunctions, *Trans. Amer. Math. Soc.* **277**, 601–621.
1249. L. THIBAULT (1991), On subdifferentials of optimal value functions, *SIAM J. Control Optim.* **29**, 1019–1036.
1250. L. THIBAULT (1992), Lagrange-Kuhn-Tucker multipliers for general mathematical programming problems, in *Optimization and Nonlinear Analysis*, edited by A. Ioffe, L. Marcus and S. Reich, Pitman Researh Notes Math. Ser. **244**, pp. 311–315, Longman, Harlow, Essex, UK.
1251. L. THIBAULT (1995), A note on the Zagrodny mean value theorem, *Optimization* **35**,127–130.
1252. L. THIBAULT (1997), On compactly Lipschitzian mappings, in *Recent Advances in Optimization*, edited by P. Gritzmann et al., Lecture Notes Econ. Math. Syst. **456**, pp. 356–364, Springer, Berlin.
1253. L. THIBAULT (2003), Sweeping process with regular and nonregular sets, *J. Diff. Eq.* **193**, 1–26.

1254. L. Thibault and D. Zagrodny (1995), Integration of subdifferentials of lower semicontinuous functions, *J. Math. Anal. Appl.* **189**, 22–58.
1255. D. Tiba (1990), *Optimal Control of Nonsmooth Distributed Parameter Systems*, Springer, Berlin.
1256. V. M. Tikhomirov (1990), *Stories about Maxima and Minima*, American Mathematical Society, Providence, Rhode Island.
1257. V. M. Tikhomirov (1997), *Elements of the Theory of Extrema*, Tinbergen Institute, TI 97-048/4, Amsterdam, The Netherlands.
1258. A. A. Tolstonogov (2000), *Differential Inclusions in a Banach Spaces*, Kluwer, Dordrecht, The Netherlands.
1259. A. A. Tolstonogov (2004), Properties of the attainable sets of evolution inclusions and control systems of subdifferential type, *Siberian Math. J.* **45**, 763–784.
1260. L. Tonelli (1923), *Fondamenti di Calcolo delle Variazioni*, published in two volumes, Zanichelli, Bologna, Italy.
1261. R. Tourky (1998), A new approach to the limit theorem on the core of an economy in vector lattices, *J. Econ. Theory* **78**, 321–328.
1262. J. S. Treiman (1983), Characterization of Clarke's tangent and normal cones in finite and infinite dimensions, *Nonlinear Anal.* **7**, 771–783.
1263. J. S. Treiman (1986), Clarke's gradients and epsilon-subgradients in Banach spaces, *Trans. Amer. Math. Soc.* **294**, 65–78.
1264. J. S. Treiman (1988), Shrinking generalized gradients, *Nonlinear Anal.* **12**, 1429–1450.
1265. J. S. Treiman (1989), Finite dimensional optimality conditions: B-gradients, *J. Optim. Theory Appl.* **62**, 771–783.
1266. J. S. Treiman (1990), Optimal control with small generalized gradients, *SIAM J. Control Optim.* **28**, 720–732.
1267. J. S. Treiman (1995), The linear nonconvex generalized gradients and Lagrange multipliers, *SIAM J. Optim.* **5**, 670–680.
1268. J. S. Treiman (1999), Lagrange multipliers for nonconvex generalized gradients with equality, inequality, and set constraints, *SIAM J. Control Optim.* **37**, 1313–1329.
1269. J. S. Treiman (2002), The linear generalized gradient in infinite dimensions, *Nonlinear Anal.* **48**, 427–443.
1270. A. A. Tret'yakov and J. E. Marsden (2003), Factor-analysis of nonlinear mappings: p-Regularity theory, *Commun. Pure Appl. Anal.* **2**, 425–445.
1271. F. Tröltzsch (1984), *Optimality Conditions for Parabolic Control Problems and Applications*, Teubner Texte, Leipzig.
1272. I. Tsevendorj (2001), Piecewise-convex maximization problems, *J. Global Optim.* **21**, 1–14.
1273. H. D. Tuan (1995), On controllability and extremality in nonconvex differential inclusions, *J. Optim. Theory Appl.* **85**, 435–472.
1274. A. Uderzo (2002), Notes on metric regularity and open covering, Technical Report MS2.7.8, pp. 1–26, Dipartimento di Sistemi ed Istituzioni per l'Economia, Univ. di L'Aquila.
1275. C. Ursescu (1975), Multifunctions with closed convex graphs, *Czech. Math. J.* **25**, 438–441.
1276. C. Ursescu (1982), Tangent sets' calculus and necessary conditions for extremality, *SIAM J. Control Optim.* **20**, 563–574.

1277. M. Valadier (1990), Young measures, in *Methods of Convex Analysis*, Lecture Notes Math. **1446**, pp. 152–188, Springer, Berlin.
1278. P. P. Varajay (1967), Nonlinear programming in Banach spaces, *SIAM J. Appl. Math.* **15**, 284–293.
1279. F. P. Vasiliev (1969), Optimality conditions for certain classes of systems which are not solved with respect to the derivative, *Soviet Math. Dokl.* **10**, 224–227.
1280. F. P. Vasiliev (1981), *Solution Methods for Extremal Problems*, Nauka, Moscow.
1281. O. V. Vasiliev (1996), *Optimization Methods*, World Federation Publishers, Atlanta, Georgia.
1282. V. Veliov (1994), Differential inclusions with stable subinclusions, *Nonlinear Anal.* **23**, 1027–1038.
1283. V. Veliov (1997), Lipschitz continuity of the value function in optimal control, *J. Optim. Theory Appl.* **94**, 335–363.
1284. V. Veliov (1997), On the time-discretization of control systems, *SIAM J. Control Optim.* **35**, 1470–1486.
1285. I. Y. Verchenko and A. N. Kolmogorov (1934), Continuation of investigation on discontinuity points for functions of two variables, *Dokl. Acad. Nauk SSSR* **4**, 361–364.
1286. J.-P. Vial (1983), Strong and weak convexity of sets and functions, *Math. Oper. Res.* **8**, 231–259.
1287. A. Villar (2000), *Equilibrium and Efficiency in Production Economies*, Springer, Berlin.
1288. A. Villar (2001), On the efficiency of market equilibrium in production economies, *J. Global Optim.* **20**, 375–389.
1289. R. B. Vinter (2000), *Optimal Control*, Birkhäuser, Boston, Massachusetts.
1290. R. B. Vinter and G. Pappas (1982), A maximum principle for non-smooth optimal control problems with state constraints, *J. Math. Anal. Appl.* **89**, 212–232.
1291. R. B. Vinter and F. L. Pereira (1988), A maximum principle for optimal processes with discontinuous trajectories, *SIAM J. Control Optim.* **26**, 205–229.
1292. R. B. Vinter and P. R. Wolenski (1990), Coextremals and the value function for control problems with data measurable in time, *J. Optim. Theory Appl.* **153**, 37–51.
1293. R. B. Vinter and P. D. Woodford (1997), On the occurrence of intermediate local minimizers that are not strong local minimizers, *Systems Cont. Lett.* **31**, 235–342.
1294. R. B. Vinter and H. Zheng (1997), The extended Euler-Lagrange condition for nonconvex variational problems, *SIAM J. Control Optim.* **35**, 56–77.
1295. R. B. Vinter and H. Zheng (1998), The extended Euler-Lagrange condition for nonconvex variational problems with state constraints, *Trans. Amer. Math. Soc.* **350**, 1181–1204.
1296. R. B. Vinter and H. Zheng (2000), Necessary conditions for free end-time, measurable time dependent optimal control problems with state constraints, *Set-Valued Anal.* **8**, 10–29.
1297. R. B. Vinter and H. Zheng (2003), Some finance problems solved with nonsmooth optimization techniques, *J. Optim. Theory Appl.* **119**, 1–18.

1298. V. Volterra (1931), *Théories Mathématoque de la Lutte pour la Vie*, Guathoer-Villar, Paris.
1299. D. W. Walkup and R. J-B. Wets (1969), A Lipschitzian characterization of convex polyhedra, *Proc. Amer. Math. Soc.* **20**, 167–173.
1300. L. Walras (1874-7), *Eléments d'Economie Politique Pure*, L. Corbaz and Company, Lausanne.
1301. L. Wang (2005), Discrete approximations to optimization of neutral functional differential inclusions, *J. Math. Anal. Appl.* **309**, 474–488.
1302. P. K. C. Wang (1964), Control of distributed parameter systems, in *Advances in Control Systems, Theory and Applications*, edited by C. T. Leondes, Vol. 1, pp. 75–172, Academic Press, New York.
1303. X. Wang (2005), Subdifferentiability of real functions, *Real Anal. Exchange* **30**, 137–171.
1304. X. Wang (2005), Extremal characterizations of reflexive spaces, *Pacific J. Optim.*, to appear.
1305. X. Wang and V. Jeyakumar (2000), A sharp Lagrange multiplier rule for nonsmooth mathematical programming problems involving equality constraints, *SIAM J. Optim.* **10**, 1136–1148.
1306. D. E. Ward (1993), Calculus for parabolic second-order derivatives, *Set-Valued Anal.* **1**, 213–246.
1307. D. E. Ward (1994), A chain rule for parabolic second-order epiderivatives, *Optimization* **28**, 223–236.
1308. D. E. Ward (1995), A comparison of second-order epiderivatives: Calculus and optimality conditions, *J. Math. Anal. Appl.* **193**, 465–482.
1309. D. E. Ward (1996), Dini derivatives of the marginal functions of a non-Lipschitzian program, *SIAM J. Optim.* **6**, 198–211.
1310. D. E. Ward (1999), Second-order necessary conditions in nonsmooth programming, in *Systems Modelling and Optimization*, edited by M. P. Polis et al., pp. 216–224, Chapman and Hall/CRC Press, Boca Raton, Florida.
1311. D. E. Ward and J. M. Borwein (1987), Nonconvex calculus in finite dimensions, *SIAM J. Control Optim.* **25**, 1312–1340.
1312. D. E. Ward and G. M. Lee (2001), Generalized properly efficient solutions of vector optimization problems, *Math. Oper. Res.* **53**, 215–232.
1313. J. Warga (1962), Relaxed variational problems, *J. Math. Anal. Appl.* **4**, 111–128.
1314. J. Warga (1962), Necessary conditions for minimum in relaxed variational problems, *J. Math. Anal. Appl.* **4**, 129–145.
1315. J. Warga (1972), *Optimal Control of Differential and Functional Equations*, Academic Press, New York.
1316. J. Warga (1975), Necessary optimality conditions without differentiability assumptions in optimal control, *J. Diff. Eq.* **15**, 41–62.
1317. J. Warga (1976), Necessary conditions withiut differentiability assumptions in unilatateral control problems, *J. Diff. Eq.* **21**, 25–38.
1318. J. Warga (1976), Derivate containers, inverse functions, and controllability, in *Calculus of Variations and Control Theory*, edited by D. L. Russel, pp. 13–46, Academic Press, New York.
1319. J. Warga (1978), Controllability and a multiplier rule for nondifferentiable optimization problems, *SIAM J. Control Optim.* **16**, 803–812.
1320. J. Warga (1981), Fat homeomorphisms and unbounded derivate containers, *J. Math. Anal. Appl.* **81**, 545–560.

1321. J. WARGA (1983), Controllability, extremality and abnormality in nonsmooth optimal control, *J. Optim. Theory Appl.* **41**, 239–260.
1322. J. WARGA (1988), Homeomorphisms and local $\mathcal{C}^1$ approximations, *Nonlinear Anal.* **12**, 593–597.
1323. J. WARGA (1995), A proper relaxation of control with variable shifts, *J. Math. Anal. Appl.* **196**, 783–793.
1324. D. WASHBURN (1979), A bound on the boundary input map for parabolic equations with applications to time optimal control, *SIAM J. Control Optim.* **17**, 652–671.
1325. T. WAŻEWSKI (1961), Systémes de commandes et équations au contingent, *Bull. Acad. Pol. Sci., Ser. Math. Ast. Phys.* **9**, 151–155.
1326. K. WEIERSTRASS (1927), *Vorlesungen über Variationsrechnung*, Akademische Verlagsgesellschaft, Leipzig.
1327. R. J-B. WETS (1996), Challenges in stochastic programming, *Math. Progr.* **75**, 115–135.
1328. L. W. WHITE (1985), Distributed control of a hyperbolic problem with control and stress constraints, *J. Math. Anal. Appl.* **106**, 41–53.
1329. P. R. WOLENSKI (1990), The exponential formula for the reachable set of a Lipschitz differential inclusion, *SIAM J. Control Optim.* **28**, 1148–1161.
1330. P. R. WOLENSKI AND Y. ZHUANG (1998), Proximal analysis and the minimal time function, *SIAM J. Control Optim.* **36**, 1048–1072.
1331. P. D. WOODFORD (1997), *Optimal Control for Nonsmooth Dynamic Systems*, Ph.D. dissertation, Department of Electric. Electron. Eng., Imperial College, London, UK.
1332. M. H. WRIGHT (1999), Ill-conditioning and computational error in interior methods for nonlinear programming, *SIAM J. Optim.* **9**, 84–111.
1333. S. E. WRIGHT (1995), Consistency of primal-dual approximations for convex optimal control, *SIAM J. Control Optim.* **33**, 1489–1509.
1334. Z. WU AND J. J. YE (2002), On error bounds for lower semicontinuous functions, *Math. Progr.* **92**, 421–435.
1335. Z. WU AND J. J. YE (2003), First-order and second-order conditions for error bounds, *SIAM J. Optim.* **14**, 621–645.
1336. Z. WU AND J. J. YE (2003), Equivalence between various derivatives and subdifferentials of the distance function, *J. Math. Anal. Appl.* **282**, 629–647.
1337. X. Q. YANG AND V. JEYAKUMAR (1992), Generalized second-order directional derivatives and optimization with $C^{1,1}$ functions, *Optimization* **26**, 165–185.
1338. J. J. YE (1999), Optimality conditions for optimization problems with complementarity constraints, *SIAM J. Control Optim.* **9**, 374–387.
1339. J. J. YE (2000), Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints, *SIAM J. Optim.* **10**, 943–962.
1340. J. J. YE (2001), Multiplier rules under mixed assumptions on differentiability and Lipschitz continuity, *SIAM J. Control Optim.* **39**, 1441–1460.
1341. J. J. YE (2004), Nondifferentiable multiplier rules for optimization and bilevel optimization problems, *SIAM J. Optim.* **15**, 252–274.
1342. J. J. YE (2005), Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints, *J. Math. Anal. Appl.* **307**, 350–369.

1343. J. J. Ye and X. Y. Ye (1997), Necessary optimality conditions for optimization problems with variational inequality constraints, *Math. Oper. Res.* **22**, 977–997.

1344. J. J. Ye, D. L. Zhu and Q. J. Zhu (1997), Exact penalization and necessary optimality conditions for generalized bilevel programming problems, *SIAM J. Optim.* **7**, 481–507.

1345. J. J. Ye and Q. J. Zhu (2003), Multiobjective optimization problems with variational inequality constraints, *Math. Progr.* **96**, 139–160.

1346. N. D. Yen (1995), Lipschitz continuity of solutions of variational inequalities with a parametric polyhedral constraint, *Math. Oper. Res.* **20**, 695–708.

1347. N. D. Yen (1995), Hölder continuity of solutions to a parametric variational inequality, *Appl. Math. Optim.* **31**, 245–255.

1348. D. Yost (1993), Asplund spaces for beginners, *Acta Univ. Carolinae, Ser. Math. Phys.* **34**, 159–177.

1349. L. C. Young (1933), On approximation by polygons in the calculus of variations, *Proc. Royal Soc.* (A) **141**, 325–341.

1350. L. C. Young (1937), Generalized curves and the existence of an attained absolute minimum in the calculus of variations, *C. R. Soc. Sci. Lett. Varsovie*, Cl. III, **30**, 212–234.

1351. L. C. Young (1969), *Lectures on the Calculus of Variations and Optimal Control Theory*, Saunders, Philadelphia, Pennsylvania.

1352. D. Zagrodny (1988), Approximate mean value theorem for upper subderivatives, *Nonlinear Anal.* **12**, 1413–1428.

1353. C. Zălinescu (2002), *Convex Analysis in General Vector Spaces*, World Scientific, Singapore.

1354. W. I. Zangwill (1967), Nonlinear programming via penalty functions, *Management Sci.* **13**, 344–358.

1355. S. C. Zaremba (1936), Sur les équations au paratingent, *Bull. Sci. Math.* **60**, 139–160.

1356. A. J. Zaslavski (2000), Generic well-posedness of optimal control problems without convexity assumptions, *SIAM J. Control Optim.* **39**, 250–280.

1357. A. J. Zaslavski (2006), *Turnpike Properties in the Calculus of Variations and Optimal Control*, Springer, New York.

1358. V. M. Zeidan (2001), New second-order optimality conditions for variational problems with C^2-Hamiltonians, *SIAM J. Control Optim.* **40**, 577–609.

1359. M. I. Zelikin and N. B. Melnikov (2004), The Bellman function and optimal synthesis in control problems with nonsmooth constraints, *J. Math. Sci.* **121**, 2281–2294.

1360. R. Zhang (1994), Problems of hierarchical optimization in finite dimensions, *SIAM J. Optim.* **4**, 521–536.

1361. R. Zhang (2003), Multistage bilevel programming problems, *Optimization* **52**, 605–616.

1362. R. Zhang (2005), Weakly upper Lipschitzian multifunctions and applications to parametric optimization, *Math. Progr.* **102**, 153–166.

1363. R. Zhang and J. S. Treiman (1995), Upper-Lipschitz multifunctions and inverse subdifferentials, *Nonlinear Anal.* **24**, 273–286.

1364. X. Y. Zheng and K. F. Ng (2005), The Fermat rule for multifunctions in Banach spaces, *Math. Progr.* **104**, 69–90.

1365. X. Y. Zheng and K. F. Ng (2005), Metric regularity and constraint qualifications for generalized equations in Banach spaces, *SIAM J. Optim.*, to appear.
1366. X. Y. Zhou (1990), Maximum principle, dynamic programming and their connections in deterministic control, *J. Optim. Theory Appl.* **65**, 363–373.
1367. X. Y. Zhou (1995), Deterministic near-optimal controls, part I: Necessary and sufficient conditions for near optimality, *J. Optim. Theory Appl.* **85**, 473–488.
1368. X. Y. Zhou (1996), Deterministic near-optimal controls, part II: Dynamic programming and viscosity solution approach, *Math. Oper. Res.* **21**, 655–674.
1369. X. Y. Zhou (1998), Stochastic near-optimal controls: Necessary and sufficient conditions for near-optimality, *SIAM J. Control Optim.* **36**, 929–947.
1370. Q. J. Zhu (1996), Necessary optimality conditions for nonconvex differential inclusions with endpoint constraints, *J. Diff. Eq.* **124**, 186–204.
1371. Q. J. Zhu (1998), The equivalence of several basic theorems for subdifferentials, *Set-Valued Anal.* **6**, 171–185.
1372. Q. J. Zhu (2000), Hamiltonian necessary conditions for a multiobjective optimal control problem, *SIAM J. Control Optim.* **39**, 97–112.
1373. Q. J. Zhu (2002), Necessary conditions for optimization problems in smooth Banach spaces, *SIAM J. Optim.* **12**, 1032–1047.
1374. Q. J. Zhu (2003), Lower semicontinuous Lyapunov functions and stability, *J. Nonlinear Convex Anal.* **4**, 325–332.
1375. Q. J. Zhu (2004), Nonconvex separation theorem for multifunctions, subdifferential calculus and applications, *Set-Valued Anal.* **12**, 275–290.
1376. T. Zolezzi (2002), On the distance theorem in quadratic optimization, *J. Convex Anal.* **9**, 693–700.
1377. T. Zolezzi (2003), Condition number theorems in optimization, *SIAM J. Optim.* **14**, 507–516.
1378. J. Zowe and S. Kurcyusz (1979), Regularity and stability for the mathematical programming problem in Banach spaces, *Appl. Math. Optim.* **5**, 49–62.
1379. E. Zuazua (2005), Propagation, observation, and control of waves approximated by finite difference methods, *SIAM Rev.* **47**, 197–243.

List of Statements

Chapter 1

1.110 Theorem: Subdifferentiation of compositions: equalities
1.111 Corollary: Subdifferentiation of products and quotients
1.112 Proposition: Subdifferentiation of compositions with surjective derivatives of inner mappings
1.113 Proposition: Subdifferentiation of minimum functions
1.114 Proposition: Nonsmooth versions of Fermat's rule
1.115 Proposition: Mean values
1.116 Definition: Sequential normal epi-compactness of functions
1.117 Proposition: SNEC property under compositions with strictly differentiable inner mappings
1.118 Definition: Second-order subdifferentials
1.119 Proposition: Second-order subdifferentials of twice differentiable functions
1.120 Proposition: Mixed second-order subdifferentials of $\mathcal{C}^{1,1}$ functions
1.121 Proposition: Equality sum rule for second-order subdifferentials
1.122 Definition: Weak* extensibility
1.123 Proposition: Sufficient conditions for weak* extensibility
1.124 Example: Violation of weak* extensibility
1.125 Proposition: Stability property for linear operators with weak* extensible ranges
1.126 Lemma: Special chain rules for coderivatives
1.127 Theorem: Second-order chain rules with surjective derivatives of inner mappings
1.128 Theorem: Second-order chain rules with twice differentiable outer mappings
1.4.1 Comment: Motivations and early developments in nonsmooth analysis
1.4.2 Comment: Tangents and directional derivatives
1.4.3 Comment: Constructions by Clarke and related developments
1.4.4 Comment: Motivations to avoid convexity
1.4.5 Comment: Basic normals and subgradients
1.4.6 Comment: Fréchet-like representations
1.4.7 Comment: Approximate subdifferentials
1.4.8 Comment: Further historical remarks
1.4.9 Comment: Some advantages of nonconvexity
1.4.10 Comment: List of major topics and contributors
1.4.11 Comment: Generalized normals in Banach spaces
1.4.12 Comment: Derivatives and coderivatives of set-valued mappings
1.4.13 Comment: Lipschitzian properties
1.4.14 Comment: Metric regularity and linear openness
1.4.15 Comment: Coderivative calculus in Banach spaces
1.4.16 Comment: Subgradients of extended-real-valued functions
1.4.17 Comment: Subgradients of distance functions
1.4.18 Comment: Subdifferential calculus in Banach spaces
1.4.19 Comment: Second-order generalized differentiation
1.4.20 Comment: Second-order subdifferential calculus in Banach spaces

Chapter 2

2.1 Definition: Local extremality of set systems
2.2 Proposition: Interiors of sets in extremal systems

Chapter 3

Chapter 4

Chapter 5

Chapter 6

6.69 Theorem:	AMP for delay systems
6.70 Example:	AMP may not hold for neutral systems
6.5.1 Comment:	Calculus of variations and optimal control
6.5.2 Comment:	Differential inclusions
6.5.3 Comment:	Optimality conditions for smooth or graph-convex differential inclusions
6.5.4 Comment:	Clarke's Euler-Lagrange condition
6.5.5 Comment:	Clarke's Hamiltonian condition
6.5.6 Comment:	Transversality conditions
6.5.7 Comment:	Extended Euler-Lagrange conditions for convex-valued differential inclusions
6.5.8 Comment:	Extended Euler-Lagrange and Weierstrass-Pontryagin conditions for nonconvex problems
6.5.9 Comment:	Dualization and extended Hamiltonian formalism
6.5.10 Comment:	Other techniques and results in nonsmooth optimal control
6.5.11 Comment:	Dual versus primal methods in optimal control
6.5.12 Comment:	The method of discrete approximations
6.5.13 Comment:	Discrete approximations of evolution inclusions
6.5.14 Comment:	Intermediate local minima
6.5.115 Comment:	Relaxation stability and hidden convexity
6.5.16 Comment:	Convergence of discrete approximations
6.5.17 Comment:	Necessary optimality conditions for discrete approximations
6.5.18 Comment:	Passing to the limit from discrete approximations
6.5.19 Comment:	Euler-Lagrange and maximum conditions with no relaxation
6.5.20 Comment:	Related topics and results in optimal control of differential inclusions
6.5.21 Comment:	Primal-space approach via the increment method
6.5.22 Comment:	Multineedle variations and convex separation in image spaces
6.5.23 Comment:	The discrete maximum principle
6.5.24 Comment:	Necessary conditions for free-endpoint discrete parametric systems
6.25 Comment:	The approximate maximum principle for discrete approximations
6.26 Comment:	Nonsmooth versions of the approximate maximum principle
6.27 Comment:	Applications of the approximate maximum principle
6.28 Comment:	The approximate maximum principle in systems with delays

Chapter 7

7.1 Theorem:	Strong approximation for differential-algebraic systems
7.2 Theorem:	Strong convergence of optimal solutions for difference-algebraic approximations
7.3 Theorem:	Necessary optimality conditions for difference-algebraic inclusions
7.4 Corollary:	Necessary conditions for difference-algebraic inclusions with enhanced nontriviality
7.5 Theorem:	Euler-Lagrange conditions for differential-algebraic inclusions
7.6 Corollary:	Hamiltonian inclusion and maximum condition for differential-algebraic inclusions

Chapter 8

Glossary of Notation

Operations and Symbols

$:=$ and $=:$	equal by definition
$\equiv$	identically equal
*	indication of some dual/adjoint/polar operation
$\langle\cdot,\cdot\rangle$	canonical pairing between space X and its topological dual X^*
$x \rightarrow \bar{x}$	x converges to $\bar{x}$ strongly (by norm)
$x \stackrel{w}{\rightarrow} \bar{x}$	x converges to $\bar{x}$ weakly (in weak topology)
$x \stackrel{w^*}{\rightarrow} \bar{x}$	x converges to $\bar{x}$ weak* (in weak* topology)
$x \stackrel{\Omega}{\rightarrow} \bar{x}$	x converges to $\bar{x}$ with $x \in \Omega$
$\liminf$	lower limit for real numbers
$\limsup$	upper limit for real numbers
$\mathop{\rm Lim\,inf}$	lower/inner limit for set-valued mappings
$\mathop{\rm Lim\,sup}$	upper/outer limit for set-valued mappings
$\dim X$ and $\mathrm{codim}\, X$	dimension and codimension of X, respectively
$\prec$	preference relation
$\lVert\cdot\rVert$ or $\lvert\cdot\rvert$ or $\lvert\lvert\lvert\cdot\rvert\rvert\rvert$	norms
$\mathrm{haus}(\Omega_1,\Omega_2)$	Pompieu-Hausdorff distance between sets
$\mathrm{lip}\, F(\bar{x},\bar{y})$	exact Lipschitzian bound of F around $(\bar{x},\bar{y})$
$\mathrm{reg}\, F(\bar{x},\bar{y})$	exact metric regularity bound of F around $(\bar{x},\bar{y})$
$\mathrm{cov}\, F(\bar{x},\bar{y})$	exact covering/linear openness bound of F around $(\bar{x},\bar{y})$
$\mathrm{rad}\, F(\bar{x},\bar{y})$	radius of metric regularity of F around $(\bar{x},\bar{y})$
$\triangle$	end of proof

Spaces

$I\!R := (-\infty,\infty)$	real line
$\overline{I\!R} := [-\infty,\infty]$	extended real line
$I\!R^n$	n-dimensional Euclidean space
$I\!R^n_+$ and $I\!R^n_-$	nonnegative and nonpositive orthant of $I\!R^n$, respectively

$\mathcal{C}([a,b];X)$	space of X-valued continuous mappings with the supremum norm on $[a,b]$
$\mathcal{C}(K)$	space of continuous functions on the compact set K
$\mathcal{C}[0,\omega_1]$	continuous functions on $[0,\omega_1]$, where ω_1 is the first uncountable ordinal
$\mathcal{C}_0$	continuous functions with compact supports
$\mathcal{C}^k$, $1 \le k \le \infty$,	k times differentiable functions with all continuous derivatives
$\mathcal{C}^{1,1}$	continuously differentiable functions with Lipschitzian derivatives
$L^p([a,b];X)$, $1 \le p \le \infty$,	standard Lebesgue spaces of X-valued mappings
$W^{1,p}$ and H^p	standard Sobolev spaces
$\mathcal{M}$ and $\mathcal{M}_b$	measure spaces (dual to spaces of continuous functions)
BV	functions of bounded variation
c	space of real number sequences with the supremum norm
c_0	subspace of c with sequences converging to zero
ℓ^p, $1 \le p \le \infty$,	sequences of real numbers with standard p–norms

Sets

$\emptyset$	empty set
$I\!N$	set of natural numbers
$B_r(x)$	ball centered at x with radius r
$I\!B_X$	closed unit ball of space X
$I\!B$ and $I\!B^*$	closed unit balls of the space and duals space in question
S and S^*	unit spheres of the space and dual space in question
$\operatorname{int}\Omega$ and $\operatorname{ri}\Omega$	interior and relative interior, respectively
$\operatorname{cl}\Omega$ and $\operatorname{cl}^*\Omega$	closure and weak* topological closure, respectively
$\operatorname{bd}\Omega$ or $\partial\Omega$	set boundary
$\operatorname{co}\Omega$ and $\operatorname{clco}\Omega$	convex hull and closed convex hull, respectively
$\operatorname{cone}\Omega$	conic hull
$\operatorname{aff}\Omega$ and $\overline{\operatorname{aff}}\,\Omega$	affine hull and closed affine hull, respectively
$\operatorname{mes}\Omega$ or $\mathcal{L}^n(\Omega)$	Lebesgue (n-dimensional) measure
$\Pi(x;\Omega)$	projection of x to Ω
$T(\bar{x};\Omega)$	contingent cone to Ω at $\bar{x}$
$T_W(\bar{x};\Omega)$	weak contingent cone to Ω at $\bar{x}$
$T_C(\bar{x};\Omega)$	Clarke tangent cone to Ω at $\bar{x}$
$N(\bar{x};\Omega)$	basic/limiting normal cone to Ω at $\bar{x}$
$N_+(\bar{x};\Omega(\bar{y}))$	extended limiting normal cone to $\Omega(\bar{y})$ at $\bar{x}$
$\widehat{N}(\bar{x};\Omega)$	prenormal cone or Fréchet normal cone to Ω at $\bar{x}$
$N_C(\bar{x};\Omega)$	Clarke normal cone to Ω at $\bar{x}$
$N_G(\bar{x};\Omega)$ and $\widetilde{N}_G(\bar{x};\Omega)$	approximate G-normal cone and its nucleus to Ω at $\bar{x}$

$N_P(\bar{x};\Omega)$	proximal normal cone to Ω at $\bar{x}$
$\widehat{N}_\varepsilon(\bar{x};\Omega)$	sets of ε-normals to Ω at $\bar{x}$
$S_\varepsilon(\bar{x};\Omega)$	ε-support to Ω at $\bar{x}$

Functions

$\delta(\cdot;\Omega)$	set indicator function
$\mathrm{dist}(\cdot;\Omega)$ or $d_\Omega(\cdot)$	distance function
$\rho(x,y) := \mathrm{dist}(y;F(x))$	extended distance function
$\mathrm{dom}\,\varphi$	domain of $\varphi\colon X \to \overline{I\!R}$
$\mathrm{epi}\,\varphi$, $\mathrm{hypo}\,\varphi$, and $\mathrm{gph}\,\varphi$	epigraph, hypergraph, and graph of φ, respectively
$x \xrightarrow{\varphi} \bar{x}$	$x \to \bar{x}$ with $\varphi(x) \to \varphi(\bar{x})$
$\mathcal{H}$	Hamiltonian function in optimal control
H	Hamilton-Pontryagin function in optimal control
L	Lagrangian function in optimization
L_Ω	essential Lagrangian relative to Ω
$\tau(F;h)$	averaged modulus of continuity
$\varphi'(\bar{x})$ or $\nabla\varphi(\bar{x})$	Fréchet derivative/gradient of φ at $\bar{x}$
$\varphi'_\beta(\bar{x})$ or $\nabla_\beta\varphi(\bar{x})$	derivative/gradient of φ at $\bar{x}$ with respect to some bornology
$\lvert\nabla\varphi\rvert(\bar{x})$	(strong) slope of φ at $\bar{x}$
$\varphi'(\bar{x};v)$	classical directional derivative of φ at $\bar{x}$ in direction v
$\varphi^\circ(\bar{x};v)$ and $\varphi^\uparrow(\bar{x};v)$	generalized directional derivative and subderivative of φ
$d^-\varphi(\bar{x};v)$ and $d^+\varphi(\bar{x};v)$	Dini-Hadamard lower and upper directional derivative of φ
$\partial\varphi(\bar{x})$	basic/limiting subdifferential of φ at $\bar{x}$
$\partial^+\varphi(\bar{x})$	upper subdifferential of φ at $\bar{x}$
$\partial^0\varphi(\bar{x})$	symmetric subdifferential of φ at $\bar{x}$
$\partial_\geq\varphi(\bar{x})$	right-sided subdifferential of φ at $\bar{x}$
$\partial^\infty\varphi(\bar{x})$	singular subdifferential of φ at $\bar{x}$
$\widehat{\partial}\varphi(\bar{x})$ and $\widehat{\partial}^+\varphi(\bar{x})$	Fréchet subdifferential and upper subdifferential of φ at $\bar{x}$, respectively
$\partial_A\varphi(\bar{x})$ and $\partial_G\varphi(\bar{x})$	approximate A-subdifferential and G-subdifferential of φ at $\bar{x}$
$\partial_C\varphi(\bar{x})$	Clarke subdifferential/generalized gradient of φ at $\bar{x}$
$\partial_\beta\varphi(\bar{x})$	viscosity (bornological) β-subdifferential of φ at $\bar{x}$
$\partial_P\varphi(\bar{x})$	proximal subdifferential of φ at $\bar{x}$
$\widehat{\partial}_\varepsilon\varphi(\bar{x})$, $\widehat{\partial}_{a\varepsilon}\varphi(\bar{x})$, and $\widehat{\partial}_{g\varepsilon}\varphi(\bar{x})$	Fréchet-type ε-subdifferentials of φ at $\bar{x}$
$\partial_\varepsilon^-\varphi(\bar{x})$	Dini ε-subdifferential of φ at $\bar{x}$
$\nabla^2\varphi(\bar{x})$	classical Hessian (matrix of second derivatives if in $I\!R^n$) of φ at $\bar{x}$
$\partial^2\varphi$, $\partial^2_N\varphi$, and $\partial^2_M\varphi$	second-order subdifferentials (generalized Hessians) of φ

Mappings

$f\colon X \to Y$	single-valued mappings from X to Y
$F\colon X \rightrightarrows Y$	set-valued mappings from X to Y
$\operatorname{dom} F$	domain of F
$\operatorname{rge} F$	range of F
$\operatorname{gph} F$	graph of F
$\ker F$	kernel of F
$F^{-1}\colon Y \rightrightarrows X$	inverse mapping to $F\colon X \rightrightarrows Y$
$F(\Omega)$ and $F^{-1}(\Omega)$	image and inverse image/preimage of Ω under F
$F \circ G$	composition of mappings
$F \overset{h}{\circ} G$	h-composition of mappings
$\Delta(\cdot;\Omega)$	set indicator mapping
Ω_ρ	set enlargement mapping
E_φ	epigraphical mapping
$\mathcal{E}(f,\Theta)$	generalized epigraph of $f\colon X \to Y$ with respect to $\Theta \subset Y$
$DF(\bar{x},\bar{y})$	graphical/contingent derivative of F at $(\bar{x},\bar{y}) \in \operatorname{gph} F$
$D^*F(\bar{x},\bar{y})$	(basic) coderivative of F at $(\bar{x},\bar{y}) \in \operatorname{gph} F$
$D^*_N F(\bar{x},\bar{y})$	normal coderivative of F at $(\bar{x},\bar{y}) \in \operatorname{gph} F$
$D^*_M F(\bar{x},\bar{y})$ and $\widetilde{D}^*_M F(\bar{x},\bar{y})$	mixed and reversed mixed coderivative of F at $(\bar{x},\bar{y})$, respectively
$\widehat{D}^* F(\bar{x},\bar{y})$ and $\widehat{D}^*_\varepsilon F(\bar{x},\bar{y})$	Fréchet coderivative and ε-coderivative of F at $(\bar{x},\bar{y})$, respectively
$Jf(\bar{x})$	generalized Jacobian of f at $\bar{x}$
$\Lambda f(\bar{x})$	derivate container of f at $\bar{x}$

Subject Index

Grundlehren der mathematischen Wissenschaften

A Series of Comprehensive Studies in Mathematics

A Selection

246. Naimark/Stern: Theory of Group Representations
247. Suzuki: Group Theory I
248. Suzuki: Group Theory II
249. Chung: Lectures from Markov Processes to Brownian Motion
250. Arnold: Geometrical Methods in the Theory of Ordinary Differential Equations
251. Chow/Hale: Methods of Bifurcation Theory
252. Aubin: Nonlinear Analysis on Manifolds. Monge-Ampère Equations
253. Dwork: Lectures on ρ-adic Differential Equations
254. Freitag: Siegelsche Modulfunktionen
255. Lang: Complex Multiplication
256. Hörmander: The Analysis of Linear Partial Differential Operators I
257. Hörmander: The Analysis of Linear Partial Differential Operators II
258. Smoller: Shock Waves and Reaction-Diffusion Equations
259. Duren: Univalent Functions
260. Freidlin/Wentzell: Random Perturbations of Dynamical Systems
261. Bosch/Güntzer/Remmert: Non Archimedian Analysis – A System Approach to Rigid Analytic Geometry
262. Doob: Classical Potential Theory and Its Probabilistic Counterpart
263. Krasnosel'skiĭ/Zabreĭko: Geometrical Methods of Nonlinear Analysis
264. Aubin/Cellina: Differential Inclusions
265. Grauert/Remmert: Coherent Analytic Sheaves
266. de Rham: Differentiable Manifolds
267. Arbarello/Cornalba/Griffiths/Harris: Geometry of Algebraic Curves, Vol. I
268. Arbarello/Cornalba/Griffiths/Harris: Geometry of Algebraic Curves, Vol. II
269. Schapira: Microdifferential Systems in the Complex Domain
270. Scharlau: Quadratic and Hermitian Forms
271. Ellis: Entropy, Large Deviations, and Statistical Mechanics
272. Elliott: Arithmetic Functions and Integer Products
273. Nikol'skiĭ: Treatise on the shift Operator
274. Hörmander: The Analysis of Linear Partial Differential Operators III
275. Hörmander: The Analysis of Linear Partial Differential Operators IV
276. Liggett: Interacting Particle Systems
277. Fulton/Lang: Riemann-Roch Algebra
278. Barr/Wells: Toposes, Triples and Theories
279. Bishop/Bridges: Constructive Analysis
280. Neukirch: Class Field Theory
281. Chandrasekharan: Elliptic Functions
282. Lelong/Gruman: Entire Functions of Several Complex Variables
283. Kodaira: Complex Manifolds and Deformation of Complex Structures
284. Finn: Equilibrium Capillary Surfaces
285. Burago/Zalgaller: Geometric Inequalities
286. Andrianaov: Quadratic Forms and Hecke Operators

287. Maskit: Kleinian Groups
288. Jacod/Shiryaev: Limit Theorems for Stochastic Processes
289. Manin: Gauge Field Theory and Complex Geometry
290. Conway/Sloane: Sphere Packings, Lattices and Groups
291. Hahn/O'Meara: The Classical Groups and K-Theory
292. Kashiwara/Schapira: Sheaves on Manifolds
293. Revuz/Yor: Continuous Martingales and Brownian Motion
294. Knus: Quadratic and Hermitian Forms over Rings
295. Dierkes/Hildebrandt/Küster/Wohlrab: Minimal Surfaces I
296. Dierkes/Hildebrandt/Küster/Wohlrab: Minimal Surfaces II
297. Pastur/Figotin: Spectra of Random and Almost-Periodic Operators
298. Berline/Getzler/Vergne: Heat Kernels and Dirac Operators
299. Pommerenke: Boundary Behaviour of Conformal Maps
300. Orlik/Terao: Arrangements of Hyperplanes
301. Loday: Cyclic Homology
302. Lange/Birkenhake: Complex Abelian Varieties
303. DeVore/Lorentz: Constructive Approximation
304. Lorentz/v. Golitschek/Makovoz: Construcitve Approximation. Advanced Problems
305. Hiriart-Urruty/Lemaréchal: Convex Analysis and Minimization Algorithms I. Fundamentals
306. Hiriart-Urruty/Lemaréchal: Convex Analysis and Minimization Algorithms II. Advanced Theory and Bundle Methods
307. Schwarz: Quantum Field Theory and Topology
308. Schwarz: Topology for Physicists
309. Adem/Milgram: Cohomology of Finite Groups
310. Giaquinta/Hildebrandt: Calculus of Variations I: The Lagrangian Formalism
311. Giaquinta/Hildebrandt: Calculus of Variations II: The Hamiltonian Formalism
312. Chung/Zhao: From Brownian Motion to Schrödinger's Equation
313. Malliavin: Stochastic Analysis
314. Adams/Hedberg: Function spaces and Potential Theory
315. Bürgisser/Clausen/Shokrollahi: Algebraic Complexity Theory
316. Saff/Totik: Logarithmic Potentials with External Fields
317. Rockafellar/Wets: Variational Analysis
318. Kobayashi: Hyperbolic Complex Spaces
319. Bridson/Haefliger: Metric Spaces of Non-Positive Curvature
320. Kipnis/Landim: Scaling Limits of Interacting Particle Systems
321. Grimmett: Percolation
322. Neukirch: Algebraic Number Theory
323. Neukirch/Schmidt/Wingberg: Cohomology of Number Fields
324. Liggett: Stochastic Interacting Systems: Contact, Voter and Exclusion Processes
325. Dafermos: Hyperbolic Conservation Laws in Continuum Physics
326. Waldschmidt: Diophantine Approximation on Linear Algebraic Groups
327. Martinet: Perfect Lattices in Euclidean Spaces
328. Van der Put/Singer: Galois Theory of Linear Differential Equations
329. Korevaar: Tauberian Theory. A Century of Developments
330. Mordukhovich: Variational Analysis and Generalized Differentiation I: Basic Theory
331. Mordukhovich: Variational Analysis and Generalized Differentiation II: Applications