Springer Series in Operations Research and Financial Engineering

Alexey F. Izmailov
Mikhail V. Solodov

Newton-Type Methods for Optimization and Variational Problems

Springer

Springer Series in Operations Research and Financial Engineering

Series Editors:
Thomas V. Mikosch
Sidney I. Resnick
Stephen M. Robinson

For further volumes:
http://www.springer.com/series/3182

Alexey F. Izmailov • Mikhail V. Solodov

Newton-Type Methods for Optimization and Variational Problems

Springer

Alexey F. Izmailov
Moscow State University
Moscow, Russia

Mikhail V. Solodov
Instituto de Matemática Pura e Aplicada
Rio de Janeiro, Brazil

ISSN 1431-8598 ISSN 2197-1773 (electronic)
ISBN 978-3-319-04246-6 ISBN 978-3-319-04247-3 (eBook)
DOI 10.1007/978-3-319-04247-3
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013958213

Mathematics Subject Classification (2010): 49-02, 90-02, 49M15, 49M37

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book is entirely devoted to mathematical analysis of Newton-type methods for variational and optimization problems in finite-dimensional spaces. Newtonian methods are among the most important tools for solving problems of various kinds, in a variety of areas. Their value is impossible to overestimate. The general idea behind all methods of this type can be informally described as approximating problem data (or perhaps only some part of the data) up to the first or second order, at every step of the iterative process.

This book provides a unified view of classical as well as very recent developments in the field of Newton-type methods and their convergence analyses. It is worth to stress that we take a rather broad view of which methods belong to the Newtonian family. Specifically, we first develop general perturbed Newtonian frameworks for variational and optimization problems and then, for the purposes of local convergence and convergence rate analyses, relate specific algorithms to the general frameworks by means of perturbations (i.e., the difference between the iteration of the method in question and what would be the exact Newton iteration for the given context). This allows us to treat, in a unified manner, various useful modifications of Newton methods, such as methods with inexact data, quasi-Newton approximations to the derivatives, and approximate solutions of iterative subproblems, among others. Moreover, this unified approach allows also to analyze some algorithms that were not traditionally viewed as of Newton type. Some examples are the linearly constrained augmented Lagrangian algorithm and the inexact restoration methods for optimization, where the objective function in the iterative subproblems does not involve any data approximations. It is interesting to note that in addition to providing a useful insight, the unified line of analysis also improves convergence results for a number of classical algorithms, when compared to what could be established using the previous arguments. In this sense, we emphasize that the main focus of this book is theoretical analysis indeed, with the primary goal being to formulate and prove the best convergence results currently available, for every algorithm considered. By this we mean the combination of the weakest assumptions needed and/or the strongest assertions obtained. We hope that the book achieves this goal, giving state-of-the-art local convergence and rate of convergence results for a number of classical algorithms and for some new ones that appear in the monographic literature here for the first time.

The book's structure is as follows. The first two chapters are mostly introductory, containing some theory of variational analysis and optimization, and mostly the usual material on Newton method for equations and unconstrained optimization. That said, already here we lay down our general philosophy of the perturbed Newtonian frameworks, and some related statements are new or at least nonstandard. Chapters 3 and 4 are devoted to local analysis of Newtonian methods for variational problems and for constrained optimization, respectively. Globalization techniques for those methods are discussed in Chaps. 5 and 6, respectively. Chapter 7 is devoted to the behavior of Newtonian methods on degenerate problems and to special techniques designed to deal with degeneracy. We now give some more specific, albeit brief, comments on the book's content.

Chapter 1 collects some facts on constraint qualifications and regularity conditions for variational problems, and on optimality conditions, that are relevant to the analysis of Newtonian methods later on. Elements of nonsmooth analysis necessary for future use are presented as well. Readers who are well familiar with the theory of optimization and variational problems can skip this chapter, returning to it for consultation when specific facts are cited later on. That said, some results on solution stability, sensitivity, and error bounds are actually very recent and are of independent interest in those areas. We then pass to the discussion of Newton methods. For a subject so classical, some introductory material is mostly standard, of course. To much extent, this is the case of Chap. 2 on systems of nonlinear equations and unconstrained optimization. However, some interpretations and facts concerned with perturbed Newton methods reflect our general approach, which will be the basis for the development of related methods for variational and constrained optimization problems later on. This chapter contains also some material on linesearch and trust-region globalization techniques, quasi-Newton methods, and semismooth Newton methods; all without being exhaustive as this material is available elsewhere; e.g., [19, 29, 45, 208], among many other sources.

Chapter 3 is devoted to local convergence analysis of methods for variational problems, starting with the fundamental Josephy–Newton scheme for generalized equations, including its perturbed and semismooth variants. We also consider semismooth Newton and active-set methods for (mixed) complementarity problems, presenting in particular a full set of relations between various regularity conditions relevant in this context. It is worth to mention that a number of items in this comparison of different regularities are not available elsewhere in the monographic literature; we hope it is useful to give the full picture on this subject.

Chapter 4 deals with local analysis of constrained optimization algorithms; it is an important part of this book. Convergence of the fundamental sequential quadratic programming (SQP) method is derived by relating it to the Josephy–Newton framework for generalized equations. This gives state-of-the-art result for SQP, not requiring the linear independence constraint qualification or the strict complementarity condition. This is different from monographs on computational optimization, which require stronger assumptions since the Josephy–Newton framework is out of their scope. We then extend the SQP scheme by allowing perturbations to its iterations. This is one of the main ideas and tools of this chapter, and perhaps of the book as a whole. Perturbed SQP framework allows to treat in a unified fashion certain truncated SQP variants, the augmented Lagrangian modification of the Hessian, second-order corrections, as well as some methods that are not in the SQP class. The latter include the linearly constrained (augmented) Lagrangian methods, inexact restoration, sequential quadratically constrained quadratic programming (SQCQP), and a certain interior feasible directions method.

In Chap. 5, we consider linesearch globalization of some previously discussed local methods for variational problems. Specifically, globalizations of the Newton method for equations and of the semismooth Newton method for complementarity problems. We also discuss the alternative path-search approach for these problems. Moreover, for the important class of monotone problems (equations and general variational inequalities), we present special globalizations based on the inexact proximal point framework with relative-error approximations. This approach gives algorithms with considerably stronger convergence properties than the alternatives based on the use of merit functions (no regularity conditions or even boundedness of the solution set are required, for example). The inexact proximal point framework with relative-error approximations and its application to globalization appears in the monographic literature here for the first time.

In Chap. 6, we discuss linesearch globalization of SQP based on merit functions, including the Maratos effect and two tools for preserving fast local convergence rate (using the nonsmooth augmented Lagrangian as the merit function, and using second-order corrections for the step). The so-called elastic mode for dealing with possible infeasibility of subproblems is illustrated in the related setting of SQCQP, where this modification turns out to be naturally indispensable. The option of trust-region globalization of SQP based on merit functions is discussed only briefly. Instead, as an alternative to the use of merit functions, a general filter framework is described and convergence of one specific filter trust-region SQP algorithm is shown.

Chapter 7 deals with various classes of degenerate problems. It is an important part of the book and consists entirely of the material not available in other monographic literature. We first put in evidence that if constraints are degenerate (Lagrange multipliers associated with a given solution are not unique), then Newtonian methods for constrained optimization have a strong tendency to generate dual sequences that are attracted to certain special Lagrange multipliers, called critical, which violate the second-order sufficient optimality conditions. We further show that this phenomenon is in fact the reason for slow convergence usually observed in the degenerate case, as convergence to noncritical multipliers would have yielded superlinear primal rate despite degeneracy. We then proceed to develop special Newtonian methods to achieve fast convergence regardless of the degeneracy issues: the stabilized Josephy–Newton method for generalized equations, and the associated stabilized Newton method for variational problems and stabilized SQP for optimization. The appealing feature is that for superlinear convergence these methods require certain second-order sufficiency conditions only (or even the weaker noncriticality of the Lagrange multiplier if there are no inequality constraints), and in particular do not need constraint qualifications of any kind. We conclude with discussing mathematical programs with complementarity constraints, an important class of degenerate problems with special structure.

While we strived to be comprehensive in our analysis of Newtonian methods, inevitably some topics are omitted and some issues, related to the methods that are presented, are not discussed. For example, it might seem strange (at first glance) that the important class of interior-point methods is not analyzed. The reason is that while interior-point methods are related to the Newton method in a certain sense, the nature of this relation is completely different from SQP, say. Interior-point methods require different tools and treatment that does not fit the concept of this book. For excellent expositions of interior-point techniques, we cite [29, 207, 208]. As for issues that are not discussed in relation to those methods that are presented, one such example is the details of practical implementations. This was again a conscious decision, not to lose focus in a book devoted to state-of-the-art mathematical analysis (that said, we believe we kept attention on those approaches that do give rise to competitive algorithms). There are excellent modern books discussing practical issues and implementation details, e.g., [29, 45, 208]. On a related note, we should mention that in all our convergence statements, stopping rules that would be used in an actual implementation are ignored, as well as possible finite termination of an iterative process at an exact or approximate solution. Thus, all iterative sequences are always infinite, which is fully consistent with our focus on the theoretical convergence analysis. We finally mention that this book does not attempt to present a comprehensive survey of the literature or the historical accounts; in most cases, we cite only those results that are directly related to our developments.

Some material presented in this book is a product of our research over the past 10 years or so, sometimes joint with current and former students Anna Daryina, Damián Fernández, Alexey Kurennoy, Artur Pogosyan, and Evgeniy Uskov, whom we also thank for reading the draft and pointing out items that required corrections.

Moscow, Russia — Alexey F. Izmailov
Rio de Janeiro, Brazil — Mikhail V. Solodov
February 2013

Contents

1 Elements of Optimization Theory and Variational Analysis 1
1.1 Constraint Systems 1
1.2 Optimization Problems 4
1.2.1 Problem Statements and Some Basic Properties 4
1.2.2 Unconstrained and Simply-Constrained Problems 7
1.2.3 Equality-Constrained Problems 8
1.2.4 Equality and Inequality-Constrained Problems 10
1.3 Variational Problems 15
1.3.1 Problem Settings 15
1.3.2 Stability and Sensitivity 18
1.3.3 Error Bounds 31
1.4 Nonsmooth Analysis 43
1.4.1 Generalized Differentiation 43
1.4.2 Semismoothness 51
1.4.3 Nondifferentiable Optimization Problems and Problems with Lipschitzian Derivatives 55

2 Equations and Unconstrained Optimization 61
2.1 Newton Method 62
2.1.1 Newton Method for Equations 62
2.1.2 Newton Method for Unconstrained Optimization 81
2.2 Linesearch Methods, Quasi-Newton Methods 84
2.2.1 Descent Methods 84
2.2.2 Quasi-Newton Methods 91
2.2.3 Other Linesearch Methods 103
2.3 Trust-Region Methods 107
2.4 Semismooth Newton Method 123
2.4.1 Semismooth Newton Method for Equations 123
2.4.2 Semismooth Newton Method for Unconstrained Optimization 135

3 Variational Problems: Local Methods 139
3.1 Josephy–Newton Method 140
3.2 Semismooth Newton Method for Complementarity Problems 148
3.2.1 Equation Reformulations of Complementarity Conditions 148
3.2.2 Extension to Mixed Complementarity Problems 164
3.3 Semismooth Josephy–Newton Method 182
3.4 Active-Set Methods for Complementarity Problems 194
3.4.1 Identification Based on Error Bounds 195
3.4.2 Extension to Mixed Complementarity Problems 199

4 Constrained Optimization: Local Methods 205
4.1 Equality-Constrained Problems 205
4.1.1 Newton–Lagrange Method 206
4.1.2 Linearly Constrained Lagrangian Methods 215
4.1.3 Active-Set Methods for Inequality Constraints 222
4.2 Sequential Quadratic Programming 225
4.3 Analysis of Algorithms via Perturbed Sequential Quadratic Programming Framework 232
4.3.1 Perturbed Sequential Quadratic Programming Framework 233
4.3.2 Augmented Lagrangian Modification and Truncated Sequential Quadratic Programming 248
4.3.3 More on Linearly Constrained Lagrangian Methods . . . 251
4.3.4 Inexact Restoration Methods 256
4.3.5 Sequential Quadratically Constrained Quadratic Programming 262
4.3.6 Second-Order Corrections 264
4.3.7 Interior Feasible Directions Methods 270
4.4 Semismooth Sequential Quadratic Programming 287

5 Variational Problems: Globalization of Convergence 305
5.1 Linesearch Methods 305
5.1.1 Globalized Newton Method for Equations 306
5.1.2 Globalized Semismooth Newton Methods for Complementarity Problems 310
5.1.3 Extension to Mixed Complementarity Problems 315
5.2 Path-Search Methods 319
5.2.1 Path-Search Framework for Equations 319
5.2.2 Path-Search Methods for Complementarity Problems 325
5.3 Hybrid Global and Local Phase Algorithms 328
5.3.1 Hybrid Iterative Frameworks 328
5.3.2 Preserving Monotonicity of Merit Function Values 333

5.4 Proximal Point-Based Globalization for Monotone Problems 334
5.4.1 Inexact Proximal Point Framework with Relative-Error Approximations 335
5.4.2 Globalized Newton Method for Monotone Equations 347
5.4.3 Globalized Josephy–Newton Method for Monotone Variational Inequalities 354

6 Constrained Optimization: Globalization of Convergence 367
6.1 Merit Functions 367
6.2 Linesearch Methods 377
6.2.1 Globalization of Sequential Quadratic Programming 377
6.2.2 Convergence Rate of Globalized Sequential Quadratic Programming 388
6.2.3 Globalization of Sequential Quadratically Constrained Quadratic Programming 401
6.3 Trust-Region and Filter Methods 415
6.3.1 Trust-Region Approaches to Sequential Quadratic Programming 416
6.3.2 A General Filter Scheme 418
6.3.3 A Trust-Region Filter Sequential Quadratic Programming Method 424

7 Degenerate Problems with Nonisolated Solutions 439
7.1 Attraction to Critical Lagrange Multipliers 440
7.1.1 Equality-Constrained Problems 441
7.1.2 Equality and Inequality-Constrained Problems 463
7.2 Special Methods for Degenerate Problems 470
7.2.1 Stabilized Josephy–Newton Method 470
7.2.2 Stabilized Sequential Quadratic Programming and Stabilized Newton Method for Variational Problems 475
7.2.3 Other Approaches and Further Developments 499
7.3 Mathematical Programs with Complementarity Constraints 511
7.3.1 Theoretical Preliminaries 512
7.3.2 General-Purpose Newton-Type Methods Applied to Mathematical Programs with Complementarity Constraints 519
7.3.3 Piecewise Sequential Quadratic Programming and Active-Set Methods 522

7.3.4 Newton-Type Methods for Lifted Reformulations 528
7.3.5 Mathematical Programs with Vanishing Constraints 538

Appendix A: Miscellaneous Material 543
A.1 Linear Algebra and Linear Inequalities 543
A.2 Analysis 546
A.3 Convexity and Monotonicity 549

References 553

Index 571

Acronyms

BFGS	Broyden–Fletcher–Goldfarb–Shanno
CQ	Constraint qualification
DFP	Davidon–Fletcher–Powell
GE	Generalized equation
KKT	Karush–Kuhn–Tucker
LCL	Linearly constrained Lagrangian
LCP	Linear complementarity problem
LICQ	Linear independence constraint qualification
MCP	Mixed complementarity problem
MFCQ	Mangasarian–Fromovitz constraint qualification
MPCC	Mathematical program with complementarity constraints
MPEC	Mathematical program with equilibrium constraints
MPVC	Mathematical program with vanishing constraints
NCP	Nonlinear complementarity problem
pSQP	Perturbed sequential quadratic programming
QP	Quadratic programming or quadratic programming problem
RNLP	Relaxed nonlinear programming problem
SMFCQ	Strict Mangasarian–Fromovitz constraint qualification
SONC	Second-order necessary optimality condition
SOSC	Second-order sufficient optimality condition
SQCQP	Sequential quadratically constrained quadratic programming
SQP	Sequential quadratic programming
SSOSC	Strong second-order sufficient optimality condition
TNLP	Tightened nonlinear programming problem
VI	Variational inequality
VP	Variational problem

Notation

Spaces

$\mathbf{R}$	The real one-dimensional space
$\mathbf{R}_+$	The set of nonnegative reals
$\mathbf{R}^n$	The real n-dimensional space
$\mathbf{R}^n_+$	The nonnegative orthant in the real n-dimensional space
$\mathbf{R}^{m\times n}$	The space of real $m \times n$-matrices

Scalars and Vectors

e^i	The i-th row of the unit matrix
$x_1, \ldots, x_n$	Components of a vector $x \in \mathbf{R}^n$
x_K	The vector comprised by components x_i, $i \in K$, of a vector $x \in \mathbf{R}^n$ for $K \subset \{1, \ldots, n\}$
$(x_i,\ i \in K)$	The $\lvert K\rvert$-dimensional vector with components x_i, $i \in K$
$\lvert x\rvert$	$= (\lvert x_1\rvert, \ldots, \lvert x_n\rvert)$; componentwise absolute value of a vector $x \in \mathbf{R}^n$
x^s	$= (x_1^s, \ldots, x_n^s)$; componentwise power of a vector $x \in \mathbf{R}^n$
$\sqrt{x}$	$= (\sqrt{x_1}, \ldots, \sqrt{x_n})$; componentwise square root of a vector $x \in \mathbf{R}^n$
$\min\{x, y\}$	$= (\min\{x_1, y_1\}, \ldots, \min\{x_n, y_n\})$; componentwise minimum of vectors $x, y \in \mathbf{R}^n$
$\max\{x, y\}$	$= (\max\{x_1, y_1\}, \ldots, \max\{x_n, y_n\})$; componentwise maximum of vectors $x, y \in \mathbf{R}^n$
$\langle x, y\rangle$	$= \sum_{j=1}^n x_j y_j$; the Euclidean inner product of vectors $x, y \in \mathbf{R}^n$
$\lVert x\rVert$	$= \lVert x\rVert_2 = \sqrt{\langle x, x\rangle}$; the Euclidean norm of a vector $x \in \mathbf{R}^n$
$\lVert x\rVert_1$	$= \sum_{j=1}^n \lvert x_j\rvert$; the l_1-norm of a vector $x \in \mathbf{R}^n$
$\lVert x\rVert_\infty$	$= \max\{\lvert x_1\rvert, \ldots, \lvert x_n\rvert\}$; the l_∞-norm of a vector $x \in \mathbf{R}^n$
$\{x^k\}$	$= \{x^0, x^1, \ldots, x^k, \ldots\}$; a sequence
$\{t_k\}$	$= \{t_0, t_1, \ldots, t_k, \ldots\}$; a sequence of scalars

$\{x^k\} \to x$	A sequence $\{x^k\}$ converges to x
$t_k \to t$ as $k \to \infty$	A sequence $\{t_k\}$ of scalars converges to t

Matrices

a_{ij}	Elements of a matrix $A \in \mathbf{R}^{m\times n}$ $(i = 1, \ldots, m$, $j = 1, \ldots, n)$
A^{T}	Transposed to a matrix A
I	The identity matrix of an appropriate size
$\mathrm{diag}(x)$	The diagonal matrix with diagonal elements equal to the components of a vector $x \in \mathbf{R}^n$
A_i	The i-th row of a matrix $A \in \mathbf{R}^{m\times n}$ $(i = 1, \ldots, m)$
$A_{K_1K_2}$	The submatrix of a matrix $A \in \mathbf{R}^{m\times n}$ with elements a_{ij}, $i \in K_1$, $j \in K_2$ $(K_1 \subset \{1, \ldots, m\}$, $K_2 \subset \{1, \ldots, n\})$
$\mathrm{im}\, A$	$= \{Ax \mid x \in \mathbf{R}^n\}$; image (range space) of a matrix $A \in \mathbf{R}^{m\times n}$
$\ker A$	$= \{x \in \mathbf{R}^n \mid Ax = 0\}$; kernel (null space) of a matrix $A \in \mathbf{R}^{m\times n}$
A^{-1}	The inverse of a nonsingular square matrix A
$\\|A\\|$	$= \sup_{x\in \mathbf{R}^n\setminus\{0\}} \\|Ax\\|/\\|x\\|$; the norm of a matrix $A \in \mathbf{R}^{m\times n}$
$\\|A\\|_F$	The Frobenius norm of a matrix $A \in \mathbf{R}^{m\times n}$ (the Euclidean norm in the space $\mathbf{R}^{m\times n}$ considered as the real nm-dimensional space)

Sets

$\|K\|$	The cardinality of a finite set K
$B(x, \delta)$	$= \{y \in \mathbf{R}^n \mid \\|y - x\\| \leq \delta\}$; the closed ball with center at $x \in \mathbf{R}^n$ and radius δ
$\mathrm{conv}\, S$	The convex hull of a set S (the smallest convex set containing S)
$S^{\perp}$	$= \{\xi \in \mathbf{R}^n \mid \langle \xi, x\rangle = 0\ \forall\, x \in S\}$; the annihilator (orthogonal complement) of a set $S \subset \mathbf{R}^n$
C°	$= \{\xi \in \mathbf{R}^n \mid \langle \xi, x\rangle \leq 0\ \forall\, x \in C\}$; the polar cone to a cone $C \subset \mathbf{R}^n$
$T_S(x)$	The contingent cone to a set S at $x \in S$
$N_S(x)$	$= (T_S(x))^{\circ}$; the normal cone to a set S at $x \in S$ (if $x \notin S$, then $N_S(x) = \emptyset$)
$\mathcal{M}(x)$	The set of Lagrange multipliers associated with a point x for a given optimization problem or a Karush–Kuhn–Tucker system
$C(x)$	The critical cone of a given optimization problem or a Karush–Kuhn–Tucker system at a point x
$A(x)$	$= \{i = 1, \ldots, m \mid g_i(x) = 0\}$; the set of indices of inequality constraints $g_i(x) \leq 0$, $i = 1, \ldots, m$, active at x

$A_+(x,\, \mu)$	$= \{i \in A(x) \mid \mu_i > 0\}$; the set of indices of inequality constraints strongly active at x for a given μ
$A_0(x,\, \mu)$	$= \{i \in A(x) \mid \mu_i = 0\}$; the set of indices of inequality constraints weakly active at x for a given μ
$A_+(x)$	$= A_+(x,\, \mu)$ when μ is uniquely defined
$A_0(x)$	$= A_0(x,\, \mu)$ when μ is uniquely defined
$\mathcal{D}_f(x)$	The cone of descent directions of a function f at a point x
$\mathcal{S}_F$	The set of points where a mapping F is differentiable

Functions and Mappings

$\mathrm{dist}(x,\, S)$	$= \inf_{y \in S} \\|y - x\\|$; the Euclidean distance from a point $x \in \mathbf{R}^n$ to a set $S \subset \mathbf{R}^n$
$\pi_S(x)$	The Euclidean projection of a point $x \in \mathbf{R}^n$ onto a closed convex set $S \subset \mathbf{R}^n$ (the orthogonal projection when S is a linear subspace)
Ax	A linear operator A applied to x
$B[x,\, y]$	A bilinear mapping B applied to x and y
$B[x]$	For a symmetric bilinear mapping B and a given x: the linear operator $\xi \to B[x,\, \xi]$
F^{-1}	The inverse of a one-to-one mapping F
$F'(x)$	The first derivative (Jacobian) of a mapping F at x
$F''(x)$	The second derivative of a mapping F at x (symmetric bilinear mapping)
$\frac{\partial F}{\partial x}(x,\, y)$	The partial derivative of a mapping F with respect to x at $(x,\, y)$
$\frac{\partial^2 F}{\partial x^2}(x,\, y)$	The second partial derivative of a mapping F with respect to x at $(x,\, y)$
$F'(x;\, \xi)$	The directional derivative of a mapping F at x in a direction ξ
$\partial_B F(x)$	The B-differential of a mapping F at x
$\partial F(x)$	$= \mathrm{conv}\, \partial_B F(x)$; Clarke's generalized Jacobian of a mapping (or a subdifferential of a convex function) F at x
$\hat{\partial} F(x)$	The set of Jacobians of active at x smooth pieces of a piecewise smooth mapping F
$(\partial_B)_x F(x,\, y)$	The partial B-differential of a mapping F with respect to x at $(x,\, y)$
$\partial_x F(x,\, y)$	$= \mathrm{conv}(\partial_B)_x F(x,\, y)$; the partial Clarke's generalized Jacobian of a mapping F with respect to x at $(x,\, y)$
$L(x,\, \lambda,\, \mu)$	$= f(x) + \langle \lambda,\, h(x)\rangle + \langle \mu,\, g(x)\rangle$; the value of the Lagrangian of a given optimization problem at $(x,\, \lambda,\, \mu)$
$\mathrm{dom}\, \Psi$	$= \{x \in \mathbf{R}^n \mid \Psi(x) \neq \emptyset\}$; the domain of a set-valued mapping Ψ defined on $\mathbf{R}^n$

Chapter 1
Elements of Optimization Theory and Variational Analysis

In this chapter we state the problem settings that will be investigated in the book and discuss some theoretical issues necessary for the subsequent analysis of numerical methods. We emphasize that the concept of this chapter is rather minimalistic. We provide only the material that would be directly used later on and prove only those facts which cannot be regarded as "standard" (i.e., their proofs cannot be found in well-established sources). For the material that we regard as rather standard, we limit our presentation to the statements, references, and comments. Readers well familiar with the theory of optimization and variational problems can skip this chapter, returning to it for consultation when specific facts would be cited later on. That said, a number of results on solution stability, sensitivity, and error bounds are actually very recent and are of independent interest in those areas.

1.1 Constraint Systems

Candidates for a solution of a given problem must be at least feasible, i.e., belong to a given set. Feasible sets are usually defined by a system of equalities and inequalities, such as

$$h(x) = 0, \quad g(x) \leq 0, \tag{1.1}$$

where $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are mappings called the constraints of the problem in consideration.

Consider a feasible point $\bar{x} \in \mathbf{R}^n$, i.e., a point satisfying (1.1). When developing both theory and numerical methods for variational and optimization problems, it is often crucial to specify some reasonable assumptions guaranteeing that the constraints behave regularly (in some sense) near $\bar{x}$. Conditions of this kind are known under the general name of *constraint qualification* (CQ). The precise meaning of "behaving regularly" can be different in different circumstances; see, e.g., the overview in [250]. Here,

A.F. Izmailov and M.V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-3-319-04247-3_1,

we shall avoid any general discussion of what is a CQ; instead, we give a list of some specific conditions of this type that will be used in our developments.

The *linearity constraint qualification* (or simply linearity of constraints) consists of saying that the mappings h and g are affine (an affine mapping is the sum of a linear mapping and a constant vector). Usually, this CQ is not related to a specific feasible point; although formally it can be, if we assume that h and g are affine in a neighborhood of $\bar{x}$ instead of everywhere.

If the mapping h is affine and each g_i, $i = 1, \ldots, m$, is a convex function on $\mathbf{R}^n$ (for the basic definitions and facts of convex analysis, see Sect. A.3), then the *Slater constraint qualification* holds if there exists $\hat{x} \in \mathbf{R}^n$ such that $h(\hat{x}) = 0$ and $g(\hat{x}) < 0$. This CQ is also not related to a specific feasible point.

A scalar inequality constraint is called *active* at $\bar{x}$ if it holds as equality at this point. Define the set of indices of inequality constraints active at $\bar{x}$ by

$$A(\bar{x}) = \{i = 1, \ldots, m \mid g_i(\bar{x}) = 0\}.$$

This notation of active constraints will now be used throughout the book without further explanations.

Assuming the differentiability of h and g at $\bar{x}$, the *Mangasarian–Fromovitz constraint qualification* (MFCQ, originally introduced in [187]) is said to hold at $\bar{x}$ if

$$\operatorname{rank} h'(\bar{x}) = l \text{ and } \exists\, \bar{\xi} \in \ker h'(\bar{x}) \text{ such that } g'_{A(\bar{x})}(\bar{x})\bar{\xi} < 0.$$

Remark 1.1. The MFCQ at a feasible point $\bar{x}$ of the constraint system (1.1) allows for the following equivalent dual interpretation: the system

$$(h'(\bar{x}))^{\mathrm{T}}\eta + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0, \quad \zeta_{A(\bar{x})} \geq 0, \quad \zeta_{\{1,\ldots,m\}\setminus A(\bar{x})} = 0$$

with respect to $(\eta, \zeta) \in \mathbf{R}^l \times \mathbf{R}^m$ has only the zero solution. The necessity is evident, while the sufficiency can be derived using the Motzkin Theorem of the Alternatives (Lemma A.2).

The stronger *linear independence constraint qualification* (LICQ) is said to hold at $\bar{x}$ if

$$\operatorname{rank} \begin{pmatrix} h'(\bar{x}) \\ g'_{A(\bar{x})}(\bar{x}) \end{pmatrix} = l + |A(\bar{x})|.$$

In the absence of inequality constraints, both MFCQ and LICQ are equivalent to the classical regularity condition for the equality-constrained problems:

$$\operatorname{rank} h'(\bar{x}) = l. \tag{1.2}$$

The distinguished role of the MFCQ in optimization theory is related to the following stability result, which can be obtained from Robinson's stability theorem [233] (it can also be found in [27, Theorem 2.87]). Consider the parametric constraint system

$$h(\sigma,\, x) = 0, \quad g(\sigma,\, x) \leq 0, \tag{1.3}$$

where $\sigma \in \mathbf{R}^s$ is a parameter and $h : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$ are given constraint mappings.

Theorem 1.2. *Let $h : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^m$ be continuous at $(\bar{\sigma},\, \bar{x}) \in \mathbf{R}^s \times \mathbf{R}^n$ and differentiable with respect to x near $(\bar{\sigma},\, \bar{x})$, and let $\frac{\partial h}{\partial x}$ and $\frac{\partial g}{\partial x}$ be continuous at $(\bar{\sigma},\, \bar{x})$. Let $h(\bar{\sigma},\, \bar{x}) = 0$, $g(\bar{\sigma},\, \bar{x}) \leq 0$, and assume that the MFCQ holds at $\bar{x}$ for the constraint system (1.3) with $\sigma = \bar{\sigma}$.*

Then for each $\sigma \in \mathbf{R}^s$ close enough to $\bar{\sigma}$ there exists $x(\sigma) \in \mathbf{R}^n$ satisfying

$$h(\sigma,\, x(\sigma)) = 0, \quad g(\sigma,\, x(\sigma)) \leq 0,$$

and such that

$$x(\sigma) - \bar{x} = O(\|h(\sigma,\, \bar{x})\| + \|\max\{0,\, g(\sigma,\, \bar{x})\}\|)$$

as $\sigma \to \bar{\sigma}$.

Remark 1.3. Both MFCQ and LICQ are stable properties. Specifically, assume that $h : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^m$ are differentiable with respect to x in a neighborhood of $(\bar{\sigma},\, \bar{x}) \in \mathbf{R}^s \times \mathbf{R}^n$, with $\frac{\partial h}{\partial x}$ and $\frac{\partial g}{\partial x}$ continuous at $(\bar{\sigma},\, \bar{x})$, that g is continuous at $(\bar{\sigma},\, \bar{x})$, and that $h(\bar{\sigma},\, \bar{x}) = 0$, $g(\bar{\sigma},\, \bar{x}) \leq 0$. If the MFCQ or the LICQ holds at $\bar{x}$ for the constraints

$$h(\bar{\sigma},\, x) = 0, \quad g(\bar{\sigma},\, x) \leq 0,$$

then the same condition also holds for any $\sigma \in \mathbf{R}^s$ close enough to $\bar{\sigma}$ for the constraints (1.3) at any point $x \in \mathbf{R}^n$ satisfying these constraints and close enough to $\bar{x}$.

An important notion for describing the geometry of a set around a given feasible point is that of the contingent cone. For a given set $S \subset \mathbf{R}^n$, the *contingent cone* to S at a point $x \in S$ is defined as follows:

$$T_S(x) = \{\xi \in \mathbf{R}^n \mid \exists\, \{t_k\} \subset \mathbf{R}_+ :\ \{t_k\} \to 0+,\, \mathrm{dist}(x + t_k\xi,\, S) = o(t_k)\}.$$

One of the roles of CQs is precisely to guarantee that the abstract/geometric definition of the contingent cone takes an explicit/algebraic form.

Corollary 1.4. *Let $h : \mathbf{R}^n \to \mathbf{R}^l$ be differentiable near $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$, let $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable at $\bar{x}$, and let the set D be defined by the constraint system (1.1) above, i.e.,*

$$D = \{x \in \mathbf{R}^n \mid h(x) = 0,\ g(x) \leq 0\}.$$

Assume that $\bar{x} \in D$.

Then the following assertions are valid:

(a) Any $\xi \in T_D(\bar{x})$ satisfies the linearized constraint system

$$h'(\bar{x})\xi = 0, \quad g'_{A(\bar{x})}(\bar{x})\xi \leq 0. \tag{1.4}$$

(b) If the MFCQ holds at $\bar{x}$ for the constraint system (1.1), then $T_D(\bar{x})$ coincides with the set of all those $\xi \in \mathbf{R}^n$ that satisfy (1.4).

Assertion (a) follows immediately from the differentiability of h and g and from the definition of the contingent cone. Assertion (b) can be derived employing Theorem 1.2 and the general fact that the contingent cone to any set at any point is closed. Actually, if we were to assume that g is differentiable near $\bar{x}$ with its derivative continuous at $\bar{x}$, the needed fact would readily follow by applying Theorem 1.2 to the mappings $h : \mathbf{R} \times \mathbf{R}^n \to \mathbf{R}^l$, $h(\sigma, x) = h(\bar{x} + \sigma\xi + x)$ and $g : \mathbf{R} \times \mathbf{R}^n \to \mathbf{R}^m$, $g(\sigma, x) = g(\bar{x} + \sigma\xi + x)$, at the point $(0, 0) \in \mathbf{R} \times \mathbf{R}^n$.

In particular, in the absence of inequality constraints, i.e., when

$$D = \{x \in \mathbf{R}^n \mid h(x) = 0\},$$

for any $\bar{x} \in D$ it holds that $T_D(\bar{x}) \subset \ker h'(\bar{x})$, and if the regularity condition (1.2) is satisfied, then $T_D(\bar{x}) = \ker h'(\bar{x})$.

1.2 Optimization Problems

The material of this section is a summary of basic facts of optimization theory and optimality conditions, except for one stability result which is not standard and is thus supplied with a justification. In particular, we state for further reference, but do not prove, classical first- and second-order optimality conditions. If in doubt about any of those statements, a reader can consult well-known books containing detailed exposition of optimization theory, e.g., [19, 27, 29, 186, 208].

1.2.1 Problem Statements and Some Basic Properties

An important problem setting in what follows is that of optimization. *Optimization problem* consists of determining a point that minimizes a given *objective function* $f : D \to \mathbf{R}$ over a given *feasible set* $D \subset \mathbf{R}^n$. We state this problem as follows:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in D. \end{array} \tag{1.5}$$

Problem (1.5) is called *feasible* if the set D is nonempty.

A (*global*) *solution* of problem (1.5) is any point $\bar{x} \in D$ satisfying the relation

$$f(\bar{x}) \leq f(x) \quad \forall\, x \in D. \tag{1.6}$$

The weaker notion of a *local solution* of problem (1.5) consists of feasibility of $\bar{x}$ and of the existence of a neighborhood U of $\bar{x}$ such that

$$f(\bar{x}) \leq f(x) \quad \forall\, x \in D \cap U. \tag{1.7}$$

If the inequality in (1.6) or (1.7) is strict for $x \neq \bar{x}$, then $\bar{x}$ is referred to as a strict global or strict local solution, respectively.

If the feasible set D is convex, and the objective function f is convex on D, then any local solution of problem (1.5) is global, and the solution set is convex.

As mentioned in Sect. 1.1, the feasible set of problem (1.5) is typically defined by equality and inequality constraints:

$$D = \{x \in \mathbf{R}^n \mid h(x) = 0,\ g(x) \leq 0\}, \tag{1.8}$$

where $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are given mappings (it is convenient and usually appropriate to consider the objective function f as defined on the entire $\mathbf{R}^n$ as well). In this case, problem (1.5) is written as

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0,\ g(x) \leq 0. \end{array} \tag{1.9}$$

As common for the historical reasons, we shall also refer to (1.9) as the *mathematical programming problem*.

If $D = \mathbf{R}^n$, then problem (1.5) is called an *unconstrained optimization problem*, stated as

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathbf{R}^n. \end{array} \tag{1.10}$$

Otherwise (1.5) is a *constrained optimization problem*. Note that (1.10) formally fits the general setting (1.9) taking $l = m = 0$. If $m = 0$, then (1.9) is an *equality-constrained problem*.

Some important special classes of problem (1.9) arise when the mappings h and g are affine, which means that the set D given by (1.8) is polyhedral. In this case, if f is a linear function, then (1.9) is a *linear programming problem*; if f is a quadratic function, then (1.9) is a *quadratic programming problem* (a quadratic function is a sum of a quadratic form and an affine function).

Among the first basic facts of optimization theory are conditions sufficient for the existence of a global solution. The classical Weierstrass Theorem states that problem (1.5) has a global solution provided D is nonempty and compact, and f is continuous on D. However, assuming that the feasible set is bounded often appears too restrictive. This assumption can be alleviated under some additional conditions imposed on the objective function. In particular, a function $f : D \to \mathbf{R}$ is called *coercive* (on D) if

$$\limsup_{k\to\infty} f(x^k) = +\infty$$

holds for any sequence $\{x^k\} \subset D$ such that $\|x^k\| \to \infty$ (or equivalently, if all level sets of this function are bounded). From the Weierstrass Theorem it easily follows that problem (1.5) has a global solution provided its feasible set D is nonempty and closed, and the objective function f is continuous and coercive on D.

An important example of a function which is necessarily continuous and coercive on the entire $\mathbf{R}^n$ is any strongly convex function $f : \mathbf{R}^n \to \mathbf{R}$. Therefore, for any f strongly convex on $\mathbf{R}^n$, and any closed set $D \subset \mathbf{R}^n$, problem (1.5) has a global solution. Moreover, if D is convex, then this solution is unique.

For future reference, our next result summarizes the properties of existence and uniqueness of solution for problem (1.5) with a quadratic objective function. Note that the Hessian of a quadratic function is constant.

Proposition 1.5. *For a quadratic function $f : \mathbf{R}^n \to \mathbf{R}$, the following assertions are valid:*

(a) If the unconstrained optimization problem (1.10) *has a local solution, then the Hessian of f is positive semidefinite (equivalently, f is convex on $\mathbf{R}^n$).*
(b) If the Hessian of f is positive semidefinite and $D \subset \mathbf{R}^n$ is a convex set, then any local solution of problem (1.5) *is necessarily global.*
(c) If the Hessian of f is positive definite (or equivalently, f is strongly convex on $\mathbf{R}^n$), then problem (1.5) *has a global solution for any nonempty closed set $D \subset \mathbf{R}^n$. If, in addition, the set D is convex, then the solution is unique.*

We complete this discussion of the most basic properties of optimization problems with the first-order necessary optimality condition for problem (1.5) with a smooth objective function and an arbitrary feasible set. The following result follows immediately from the definition of differentiability (see Sect. A.2).

Theorem 1.6. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be differentiable at $\bar{x} \in \mathbf{R}^n$, and let $D \subset \mathbf{R}^n$ be an arbitrary set.*

If $\bar{x}$ is a local solution of problem (1.5)*, then*

$$f'(\bar{x}) \in -(T_D(\bar{x}))^\circ. \tag{1.11}$$

Each first-order necessary optimality condition for some optimization problem class gives rise to the associated stationarity concept. Accordingly, the point $\bar{x} \in \mathbf{R}^n$ is a *stationary point* of the general problem (1.5) if $\bar{x} \in D$ and (1.11) holds.

1.2.2 Unconstrained and Simply-Constrained Problems

Consider first the unconstrained optimization problem (1.10), assuming that its objective function f is smooth. The first-order necessary optimality condition for this problem is the classical Fermat principle, and it is an evident particular case of Theorem 1.6.

Theorem 1.7. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be differentiable at $\bar{x} \in \mathbf{R}^n$.*
If $\bar{x}$ is a local solution of problem (1.10)*, then*

$$f'(\bar{x}) = 0. \tag{1.12}$$

The point $\bar{x} \in \mathbf{R}^n$ is a *stationary point* of problem (1.10) (or a *critical point* of function f) if it satisfies (1.12).

If f is convex on $\mathbf{R}^n$, then any stationary point of problem (1.10) is its global solution. Moreover, one can remove the assumption of differentiability of f at $\bar{x}$, replacing (1.12) by the condition employing the subdifferential of f (see Sect. A.3):

$$\partial f(\bar{x}) \ni 0.$$

In the convex case first-order necessary conditions are sufficient for optimality. In the general (nonconvex) case, more subtle necessary conditions are needed. We proceed with characterization of local optimality in terms of second derivatives. The next theorem presents the second-order necessary optimality condition (SONC).

Theorem 1.8. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be twice differentiable at $\bar{x} \in \mathbf{R}^n$.*
If $\bar{x}$ is a local solution of problem (1.10)*, then the Hessian of f at $\bar{x}$ is positive semidefinite:*

$$\langle f''(\bar{x})\xi, \xi\rangle \geq 0 \quad \forall\, \xi \in \mathbf{R}^n. \tag{1.13}$$

This SONC is naturally related to the second-order sufficient optimality condition (SOSC), which is obtained by replacing the nonstrict inequality in (1.13) by the strict one (for $\xi \neq 0$).

Theorem 1.9. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be twice differentiable at $\bar{x} \in \mathbf{R}^n$.*
If $\bar{x}$ satisfies (1.12) *and the Hessian of f at $\bar{x}$ is positive definite, i.e.,*

$$\langle f''(\bar{x})\xi, \xi\rangle > 0 \quad \forall\, \xi \in \mathbf{R}^n \setminus \{0\}, \tag{1.14}$$

then $\bar{x}$ is a strict local solution of problem (1.10)*. Moreover, the quadratic growth condition holds at $\bar{x}$, i.e., there exist a neighborhood U of $\bar{x}$ and $\gamma > 0$ such that*

$$f(x) - f(\bar{x}) \geq \gamma\|x - \bar{x}\|^2 \quad \forall\, x \in U. \tag{1.15}$$

Conversely, the quadratic growth condition at $\bar{x}$ implies the SOSC (1.14).

Abbreviations SONC and SOSC will be used throughout the book without further explanations, and not only for unconstrained optimization but for more general problem classes as well. Various SOSCs will be mostly used as assumptions in convergence and rate of convergence analysis. Theorems 1.8 and 1.9 (and other theorems on SONCs and SOSCs presented below) demonstrate that the assumptions of sufficient conditions are closely related to the necessary ones (the gap between the two is as small as possible, in a sense). Therefore, assuming sufficient optimality conditions for convergence (and especially for rate of convergence) is very natural.

We complete this section with the first-order necessary optimality condition for a general optimization problem (1.5) with a smooth objective function and a convex feasible set. This result is again an evident particular case of Theorem 1.6.

Theorem 1.10. *Let $D \subset \mathbf{R}^n$ be a convex set, and let $f : \mathbf{R}^n \to \mathbf{R}$ be differentiable at $\bar{x} \in \mathbf{R}^n$.*

If $\bar{x}$ is a local solution of problem (1.5), then

$$\langle f'(\bar{x}),\, x - \bar{x}\rangle \geq 0 \quad \forall\, x \in D. \tag{1.16}$$

If in addition D is closed, then (1.16) is equivalent to the following equality employing the projection operator:

$$\pi_D(\bar{x} - f'(\bar{x})) = \bar{x}. \tag{1.17}$$

In the setting of Theorem 1.10, condition (1.16) is equivalent to (1.11). Therefore, if D is convex, then $\bar{x} \in \mathbf{R}^n$ is a stationary point of the problem (1.5) if and only if $\bar{x} \in D$ and (1.16) (or, equivalently (1.17), assuming that D is closed) holds.

The necessary optimality condition of Theorem 1.10 is a useful tool when the feasible set D has simple structure.

1.2.3 Equality-Constrained Problems

Consider now the equality-constrained problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \end{array} \tag{1.18}$$

where we assume that the objective function $f : \mathbf{R}^n \to \mathbf{R}$ and the mapping $h : \mathbf{R}^n \to \mathbf{R}^l$ defining the constraints are smooth.

The *Lagrangian* $L : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}$ of problem (4.1) is given by

$$L(x,\, \lambda) = f(x) + \langle \lambda,\, h(x)\rangle.$$

The following theorem is the celebrated Lagrange principle giving the first-order necessary optimality condition for problem (1.18). It can be derived from Theorem 1.6 combined with Corollary 1.4.

Theorem 1.11. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be differentiable at $\bar{x} \in \mathbf{R}^n$, and let further $h : \mathbf{R}^n \to \mathbf{R}^l$ be differentiable near $\bar{x}$, with its derivative being continuous at $\bar{x}$. Assume that at least one of the following two conditions is satisfied:*

(i) The regularity condition (the LICQ) holds:

$$\operatorname{rank} h'(\bar{x}) = l. \tag{1.19}$$

(ii) The mapping h is affine.

If $\bar{x}$ is a local solution of problem (1.18)*, then there exists $\bar{\lambda} \in \mathbf{R}^l$ such that*

$$\frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda}) = 0. \tag{1.20}$$

The point $\bar{x} \in \mathbf{R}^n$ is a *stationary point* of problem (1.18) if $h(\bar{x}) = 0$ and there exists $\bar{\lambda} \in \mathbf{R}^l$ satisfying (1.20). In this case $\bar{\lambda}$ is referred to as a *Lagrange multiplier* associated with the stationary point $\bar{x}$. Observe that Lagrange multiplier associated with a stationary point $\bar{x}$ of problem (1.18) is unique if and only if the regularity condition (1.19) holds.

Stationary points and associated Lagrange multipliers of problem (1.18) are characterized by the so-called *Lagrange optimality system*

$$\frac{\partial L}{\partial x}(x,\, \lambda) = 0, \quad h(x) = 0.$$

Employing the full gradient of the Lagrangian, this system can be written in the form

$$L'(x,\, \lambda) = 0.$$

Observe that this is a system of $n + l$ equations in the same number of variables.

The next theorem states SONC for problem (1.18).

Theorem 1.12. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice differentiable at $\bar{x} \in \mathbf{R}^n$, and let h be differentiable near $\bar{x}$, with its derivative being continuous at $\bar{x}$. Assume that the regularity condition* (1.19) *holds.*

If $\bar{x}$ is a local solution of problem (1.18)*, then for the unique $\bar{\lambda} \in \mathbf{R}^l$ satisfying* (1.20) *it holds that*

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})\xi,\, \xi \right\rangle \geq 0 \quad \forall\, \xi \in \ker h'(\bar{x}).$$

The following is SOSC for problem (1.18).

Theorem 1.13. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice differentiable at a point $\bar{x} \in \mathbf{R}^n$.*

If $h(\bar{x}) = 0$ and

$$\forall\, \xi \in \ker h'(\bar{x})\setminus\{0\} \quad \exists\, \bar{\lambda} \in \mathbf{R}^l \ \textit{ satisfying } \ (1.20) \ \textit{ and } \ \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})\xi,\, \xi \right\rangle > 0, \tag{1.21}$$

then $\bar{x}$ is a strict local solution of problem (1.18). Moreover, the quadratic growth condition holds at $\bar{x}$, i.e., there exist a neighborhood U of $\bar{x}$ and $\gamma > 0$ such that

$$f(x) - f(\bar{x}) \geq \gamma \|x - \bar{x}\|^2 \quad \forall\, x \in U \ \textit{ such that } \ h(x) = 0.$$

Conversely, if the regularity condition (1.19) holds, then the quadratic growth condition at $\bar{x}$ implies (1.21).

1.2.4 Equality and Inequality-Constrained Problems

Finally, we consider the mathematical programming problem (1.9) with equality and inequality-constraints, assuming that the objective function $f : \mathbf{R}^n \to \mathbf{R}$ and the constraint mappings $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are smooth.

The *Lagrangian* $L : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}$ of problem (1.9) is given by

$$L(x,\, \lambda,\, \mu) = f(x) + \langle \lambda,\, h(x) \rangle + \langle \mu,\, g(x) \rangle.$$

The first-order necessary optimality condition for problem (1.9) is provided by the following Karush–Kuhn–Tucker (KKT) Theorem [166, 174]. This is again a corollary of Theorem 1.6 combined with Corollary 1.4 and Lemma A.1.

Theorem 1.14. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable at $\bar{x} \in \mathbf{R}^n$, and let $h : \mathbf{R}^n \to \mathbf{R}^l$ be differentiable near $\bar{x}$, with its derivative being continuous at $\bar{x}$. Assume that at least one of the following three conditions is satisfied:*

(i) The MFCQ holds at $\bar{x}$.
(ii) The linearity CQ holds.
(iii) The mapping h is affine, the functions g_i, $i = 1, \ldots, m$, are convex on $\mathbf{R}^n$, and the Slater CQ holds.

If $\bar{x}$ is a local solution of problem (1.9), then there exist $\bar{\lambda} \in \mathbf{R}^l$ and $\bar{\mu} \in \mathbf{R}^m_+$ such that

$$\frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) = 0, \quad \langle \bar{\mu},\, g(\bar{x}) \rangle = 0. \tag{1.22}$$

The point $\bar{x} \in \mathbf{R}^n$ is a *stationary point* of problem (1.9) if $h(\bar{x}) = 0$, $g(\bar{x}) \leq 0$, and there exist $\bar{\lambda} \in \mathbf{R}^l$ and $\bar{\mu} \in \mathbf{R}^m_+$ satisfying (1.22). In this case $(\bar{\lambda}, \bar{\mu})$ is referred to as a *Lagrange multiplier* associated with the stationary point $\bar{x}$.

The second equality in (1.22) is known as the *complementarity* (or *complementary slackness*) condition. When $\bar{\mu} \geq 0$ and $g(\bar{x}) \leq 0$ this condition means that

$$\bar{\mu}_{\{1,\, \ldots,\, m\} \setminus A(\bar{x})} = 0,$$

that is, the Lagrange multiplier components corresponding to inequality constraints inactive at $\bar{x}$ must be equal to zero.

The system of equations and inequalities

$$\frac{\partial L}{\partial x}(x,\, \lambda,\, \mu) = 0, \quad h(x) = 0, \quad \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu,\, g(x) \rangle = 0 \tag{1.23}$$

characterizing stationary points and associated Lagrange multipliers of problem (1.9) is commonly known under the name of the *KKT optimality system*.

Let $\mathcal{M}(\bar{x})$ stand for the set of Lagrange multipliers associated with a stationary point $\bar{x}$ of problem (1.9), that is, the set of $(\lambda,\, \mu) \in \mathbf{R}^l \times \mathbf{R}^m$ satisfying (1.23) for $x = \bar{x}$. It can be easily seen from Remark 1.1 that for any stationary point $\bar{x}$ of problem (1.9) the MFCQ at $\bar{x}$ is equivalent to saying that $\mathcal{M}(\bar{x})$ is bounded (the fact pointed out in [98]). Note that under the MFCQ a Lagrange multiplier associated with $\bar{x}$ need not be unique. A sufficient, but not necessary, condition for uniqueness of the multiplier is the LICQ at $\bar{x}$. We shall also use extensively the so-called *strict Mangasarian–Fromovitz constraint qualification* (SMFCQ) at a stationary point $\bar{x}$ of problem (1.9), which consists precisely of saying that the Lagrange multiplier associated with $\bar{x}$ is unique. Since a singleton is a bounded set, SMFCQ implies MFCQ. On the other hand, SMFCQ is implied by LICQ.

For a given stationary point $\bar{x}$ of problem (1.9), define the following partition of the index set $A(\bar{x})$ of inequality constraints active at $\bar{x}$:

$$A_+(\bar{x},\, \bar{\mu}) = \{i \in A(\bar{x}) \mid \bar{\mu} > 0\}, \quad A_0(\bar{x},\, \bar{\mu}) = \{i \in A(\bar{x}) \mid \bar{\mu} = 0\}.$$

Remark 1.15. It can be easily seen that the SMFCQ at a stationary point $\bar{x}$ of problem (1.9) admits the following equivalent interpretation: for a Lagrange multiplier $(\bar{\lambda},\, \bar{\mu})$ associated with $\bar{x}$ the system

$$(h'(\bar{x}))^{\mathrm{T}}\eta + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0, \quad \zeta_{A_0(\bar{x},\, \bar{\mu})} \geq 0, \quad \zeta_{\{1,\, \ldots,\, m\} \setminus A(\bar{x})} = 0$$

with respect to $(\eta,\, \zeta) \in \mathbf{R}^l \times \mathbf{R}^m$ has only the zero solution.

Under the additional convexity assumptions, the KKT necessary optimality condition given by Theorem 1.14 becomes sufficient.

Theorem 1.16. *Let* $f : \mathbf{R}^n \to \mathbf{R}$ *and* $g : \mathbf{R}^n \to \mathbf{R}^m$ *be differentiable at* $\bar{x} \in \mathbf{R}^n$*, let* f *and the components of* g *be convex on* $\mathbf{R}^n$*, and let* $h : \mathbf{R}^n \to \mathbf{R}^l$ *be affine.*

If $\bar{x}$ *is feasible in problem* (1.9)*, and there exist* $\bar{\lambda} \in \mathbf{R}^l$ *and* $\bar{\mu} \in \mathbf{R}^m_+$ *such that* (1.22) *holds, then* $\bar{x}$ *is a solution of problem* (1.9)*.*

However, apart from the convex case, stationarity does not imply even local optimality, in general. In order to state SONC and SOSC, we will need the notion of the *critical cone* of problem (1.9) at a feasible point $\bar{x}$, defined as follows:

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A(\bar{x})}(\bar{x})\xi \leq 0,\ \langle f'(\bar{x}),\, \xi\rangle \leq 0\}.$$

The proof of the following lemma is elementary.

Lemma 1.17. *Let* $f : \mathbf{R}^n \to \mathbf{R}$*,* $h : \mathbf{R}^n \to \mathbf{R}^l$ *and* $g : \mathbf{R}^n \to \mathbf{R}^m$ *be differentiable at* $\bar{x} \in \mathbf{R}^n$*.*

If $\bar{x}$ *is a stationary point of problem* (1.9)*, then*

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A(\bar{x})}(\bar{x})\xi \leq 0,\ \langle f'(\bar{x}),\, \xi\rangle = 0\},$$

and for any Lagrange multiplier $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ *associated with* $\bar{x}$ *it holds that*

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+(\bar{x},\,\bar{\mu})}(\bar{x})\xi = 0,\ g'_{A_0(\bar{x},\,\bar{\mu})}(\bar{x})\xi \leq 0\}.$$

Remark 1.18. Lemma 1.17 immediately implies that when the *strict complementarity condition* $\bar{\mu}_{A(\bar{x})} > 0$ holds at a stationary point $\bar{x}$ for an associated Lagrange multiplier $(\bar{\lambda},\, \bar{\mu})$ of problem (1.9), then

$$C(\bar{x}) = \ker h'(\bar{x}) \cap \ker g'_{A(\bar{x})}(\bar{x}),$$

and in particular, $C(\bar{x})$ is a linear subspace in this case.

Theorem 1.19. *Let* $f : \mathbf{R}^n \to \mathbf{R}$*,* $h : \mathbf{R}^n \to \mathbf{R}^l$ *and* $g : \mathbf{R}^n \to \mathbf{R}^m$ *be twice differentiable at* $\bar{x} \in \mathbf{R}^n$*, and let* h *and* g *be differentiable near* $\bar{x}$*, with their derivatives being continuous at* $\bar{x}$*. Assume that the MFCQ holds at* $\bar{x}$*.*

If $\bar{x}$ *is a local solution of problem* (1.9)*, then*

$$\forall\, \xi \in C(\bar{x}) \quad \exists\, (\bar{\lambda},\, \bar{\mu}) \in \mathcal{M}(\bar{x}) \ \ \textit{such that} \ \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi \right\rangle \geq 0. \tag{1.24}$$

Assuming the SMFCQ at a stationary point $\bar{x}$, condition (1.24) takes the form

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi \right\rangle \geq 0 \quad \forall\, \xi \in C(\bar{x}) \tag{1.25}$$

for the unique Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$ associated with $\bar{x}$. However, in the absence of the SMFCQ, condition (1.24) does not necessarily imply the existence of a "universal" (the same for each $\xi \in C(\bar{x})$) multiplier $(\bar{\lambda}, \bar{\mu})$ satisfying (1.25). For examples illustrating this fact, see, e.g., [13].

Theorem 1.20. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable at $\bar{x} \in \mathbf{R}^n$.*

If $h(\bar{x}) = 0$ and $g(\bar{x}) \leq 0$, and

$$\forall\, \xi \in C(\bar{x}) \setminus \{0\} \quad \exists\, (\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x}) \ \text{ such that } \ \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0, \tag{1.26}$$

then $\bar{x}$ is a strict local solution of problem (1.9). Moreover, the quadratic growth condition holds at $\bar{x}$, i.e., there exist a neighborhood U of $\bar{x}$ and $\gamma > 0$ such that

$$f(x) - f(\bar{x}) \geq \gamma \|x - \bar{x}\|^2 \quad \forall\, x \in U \ \text{ such that } \ h(x) = 0,\ g(x) \leq 0.$$

Conversely, if the MFCQ holds at $\bar{x}$, then the quadratic growth condition at $\bar{x}$ implies (1.26).

In the analysis of algorithms, we shall often use the stronger form of the SOSC (1.26), namely,

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}$$

for some fixed $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$.

We complete this section by considering the parametric mathematical programming problem

$$\begin{array}{ll} \text{minimize} & f(\sigma, x) \\ \text{subject to} & h(\sigma, x) = 0,\ g(\sigma, x) \leq 0, \end{array} \tag{1.27}$$

where the objective function $f : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}$ and the constraint mappings $h : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$ depend on the parameter $\sigma \in \mathbf{R}^s$. The following stability result is a combination of Theorem 1.2 and [14, Theorem 3.1].

Theorem 1.21. *Let the functions $f : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$ have the following properties for a given $\bar{\sigma} \in \mathbf{R}^s$ and a given $\bar{x} \in \mathbf{R}^n$:*

(i) f is continuous at $(\bar{\sigma}, x)$, and $h(\cdot, x)$ and $g(\cdot, x)$ are continuous at $\bar{\sigma}$, for all $x \in \mathbf{R}^n$ close enough to $\bar{x}$.
(ii) $f(\sigma, \cdot)$ is continuous on a neighborhood of $\bar{x}$ for all $\sigma \in \mathbf{R}^s$ close enough to $\bar{\sigma}$.

(iii) *h and g are differentiable with respect to x near $(\bar{\sigma}, \bar{x})$, and $\frac{\partial h}{\partial x}$ and $\frac{\partial g}{\partial x}$ are continuous at $(\bar{\sigma}, \bar{x})$.*

Let $\bar{x}$ be a strict local solution of problem (1.27) with $\sigma = \bar{\sigma}$, and let the MFCQ hold at $\bar{x}$.

Then for any $\sigma \in \mathbf{R}^s$ close enough to $\bar{\sigma}$, problem (1.27) has a local solution $x(\sigma)$ such that $x(\sigma) \to \bar{x}$ as $\sigma \to \bar{\sigma}$.

Proof. Since $\bar{x}$ is a strict local solution of problem (1.27) with $\sigma = \bar{\sigma}$, there exists $\delta > 0$ such that $\bar{x}$ is the unique global solution of the problem

$$\begin{array}{ll} \text{minimize} & f(\bar{\sigma}, x) \\ \text{subject to} & h(\bar{\sigma}, x) = 0, \; g(\bar{\sigma}, x) \leq 0, \; x \in B(\bar{x}, \delta). \end{array} \tag{1.28}$$

From Theorem 1.2 it follows that for all $\sigma \in \mathbf{R}^s$ close enough to $\bar{\sigma}$ the problem

$$\begin{array}{ll} \text{minimize} & f(\sigma, x) \\ \text{subject to} & h(\sigma, x) = 0, \; g(\sigma, x) \leq 0, \; x \in B(\bar{x}, \delta) \end{array} \tag{1.29}$$

has a nonempty feasible set. This feasible set is evidently compact provided δ is small enough. Hence, by the Weierstrass Theorem, problem (1.29) has a global solution $x(\sigma)$.

Consider an arbitrary sequence $\{\sigma^k\} \subset \mathbf{R}^s$ convergent to $\bar{\sigma}$. Since $\{x(\sigma^k)\}$ is contained in the compact set $B(\bar{x}, \delta)$, this sequence has an accumulation point $\tilde{x} \in \mathbf{R}^n$. Then, by continuity, this point satisfies

$$h(\bar{\sigma}, \tilde{x}) = 0, \quad g(\bar{\sigma}, \tilde{x}) \leq 0, \quad \tilde{x} \in B(\bar{x}, \delta). \tag{1.30}$$

Next, according to Theorem 1.2, there exists a sequence $\{x^k\} \subset \mathbf{R}^n$ convergent to $\bar{x}$ such that

$$h(\sigma^k, x^k) = 0, \quad g(\sigma^k, x^k) \leq 0, \quad x^k \in B(\bar{x}, \delta) \quad \forall\, k.$$

Therefore, by the definition of $\{x(\sigma^k)\}$ it holds that

$$f(\sigma^k, x(\sigma^k)) \leq f(\sigma^k, x^k) \quad \forall\, k,$$

and passing onto the limit in this inequality along the appropriate subsequence we derive that

$$f(\bar{\sigma}, \tilde{x}) \leq f(\bar{\sigma}, \bar{x}).$$

According to (1.30), this implies that $\tilde{x}$ is also a global solution of problem (1.28), and taking into account the choice of δ, the latter means that $\tilde{x} = \bar{x}$. Therefore, every accumulation point of $\{x(\sigma^k)\}$ is precisely $\bar{x}$, i.e., the entire sequence converges to $\bar{x}$.

We thus established that $x(\sigma) \to \bar{x}$ as $\sigma \to \bar{\sigma}$. This implies that for all $\sigma \in \mathbf{R}^s$ close enough to $\bar{\sigma}$ the inequality $\|x(\sigma) - \bar{x}\| < \delta$ holds. Then since $x(\sigma)$ is a global solution of problem (1.29), it is a local solution of problem (1.27). □

1.3 Variational Problems

This section introduces various classes of variational problems and discusses their stability and sensitivity properties, regularity conditions, and error bounds. Further details on these subjects can be found in [27, 68, 169, 201, 239].

1.3.1 Problem Settings

The most general problem setting considered in this book is that of a *generalized equation* (GE), originally introduced in [234]. The framework of GEs turned out to be very convenient, covering various more special problem classes. The GE is the commonly used name for the structured inclusion

$$\Phi(u) + N(u) \ni 0, \tag{1.31}$$

where $\Phi : \mathbf{R}^\nu \to \mathbf{R}^\nu$ is a single-valued mapping, sometimes called *base*, and $N(\cdot)$ is a set-valued mapping from $\mathbf{R}^\nu$ to the subsets of $\mathbf{R}^\nu$ (i.e., $N(u) \subset \mathbf{R}^\nu$ for each $u \in \mathbf{R}^\nu$), sometimes called *field* [201]. This is indeed a generalized equation in the sense that (1.31) reduces to the usual system of equations

$$\Phi(u) = 0 \tag{1.32}$$

when at every point the field multifunction is the singleton $\{0\}$. Many modeling paradigms arise from (and often eventually reduce to) systems of equations, and we shall pay special attention to this problem setting in what follows.

Recall that the *normal cone* to a set $Q \subset \mathbf{R}^\nu$ at a point $u \in Q$ is defined by

$$N_Q(u) = (T_Q(u))^\circ$$

with the additional convention that $N_Q(u) = \emptyset$ if $u \in \mathbf{R}^\nu \setminus Q$. The *variational problem* (VP) defined by a mapping Φ and a set Q is the GE (1.31) with the field given by the normal cone multifunction for the set Q:

$$N(u) = N_Q(u) \tag{1.33}$$

for $u \in \mathbf{R}^\nu$. In this case, Q can be regarded as the *feasible set* of the VP. For example, for a given smooth function $\varphi : \mathbf{R}^\nu \to \mathbf{R}$, stationary points of the optimization problem

$$\begin{array}{ll} \text{minimize} & \varphi(u) \\ \text{subject to} & u \in Q \end{array} \tag{1.34}$$

are characterized by the VP given by (1.31), (1.33) with the gradient base mapping:

$$\Phi(u) = \varphi'(u), \quad u \in \mathbf{R}^\nu \tag{1.35}$$

(see Sect. 1.2.1).

The *variational inequality* (VI) is the VP with a convex feasible set. Note that for a convex set Q the equality

$$N_Q(u) = \{w \in \mathbf{R}^\nu \mid \langle w, v - u\rangle \leq 0 \; \forall\, v \in Q\}$$

holds for all $u \in Q$. Therefore, the VI corresponding to the VP (1.31), (1.33), can be written in the form

$$u \in Q, \quad \langle \Phi(u), v - u\rangle \geq 0 \quad \forall\, v \in Q. \tag{1.36}$$

For example, for a given smooth function $\varphi : \mathbf{R}^\nu \to \mathbf{R}$, stationary points of the optimization problem (1.34) with a convex feasible set $Q \subset \mathbf{R}^\nu$ are characterized by the VI (1.36) with Φ defined in (1.35) (see Sect. 1.2.2).

The *mixed complementarity problem* (MCP) is the VI (1.36) with bound constraints, i.e., with $Q = [a, b]$, where

$$[a, b] = \{u \in \mathbf{R}^\nu \mid a_i \leq u_i \leq b_i,\; i = 1, \ldots, \nu\}$$

is a (generalized) box, $a_i \in \mathbf{R} \cup \{-\infty\}$, $b_i \in \mathbf{R} \cup \{+\infty\}$, $a_i < b_i$, $i = 1, \ldots, \nu$. This problem can be equivalently stated as follows:

$$u \in [a, b], \quad \Phi_i(u) \begin{cases} \geq 0 & \text{if } u_i = a_i, \\ = 0 & \text{if } a_i < u_i < b_i, \quad i = 1, \ldots, \nu. \\ \leq 0 & \text{if } u_i = b_i, \end{cases} \tag{1.37}$$

The (*nonlinear*) *complementarity problem* (NCP) is the particular instance of the VI (1.36) with $Q = \mathbf{R}^\nu_+$, or equivalently, of the MCP (1.37) with $a_i = 0$, $b_i = +\infty$, $i = 1, \ldots, \nu$. Therefore, NCP can be written as

$$u \geq 0, \quad \Phi(u) \geq 0, \quad \langle u, \Phi(u)\rangle = 0. \tag{1.38}$$

The *KKT system* is a structured collection of equations and inequalities in the primal-dual variables, stated as follows:

$$\begin{gathered} F(x) + (h'(x))^{\mathrm{T}}\lambda + (g'(x))^{\mathrm{T}}\mu = 0, \quad h(x) = 0, \\ \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu, g(x)\rangle = 0, \end{gathered} \tag{1.39}$$

where $F : \mathbf{R}^n \to \mathbf{R}^n$, $h : \mathbf{R}^n \to \mathbf{R}^l$, and $g : \mathbf{R}^n \to \mathbf{R}^m$ are the given mappings. If $(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ is a solution of the KKT system (1.39), then $(\bar{\lambda}, \bar{\mu})$ is referred to as a *Lagrange multiplier* associated with the primal solution $\bar{x}$. By $\mathcal{M}(\bar{x})$ we shall denote the set of Lagrange multipliers associated with a point $\bar{x}$, that is, the set of $(\lambda, \mu) \in \mathbf{R}^l \times \mathbf{R}^m$ satisfying (1.39) for $x = \bar{x}$. Define the mapping $G : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n$,

$$G(x, \lambda, \mu) = F(x) + (h'(x))^{\mathrm{T}}\lambda + (g'(x))^{\mathrm{T}}\mu. \tag{1.40}$$

Then the KKT system (1.39) is a particular instance of the VI (1.36) with the mapping $\Phi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ given by

$$\Phi(u) = (G(x, \lambda, \mu),\, h(x),\, -g(x)), \quad u = (x, \lambda, \mu), \tag{1.41}$$

and with

$$Q = \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}_+^m. \tag{1.42}$$

Equivalently, the KKT system (1.39) is a particular instance of the MCP (1.37) if we set $\nu = n + l + m$, with Φ defined in (1.41), and with

$$a_i = -\infty,\ i = 1,\, \ldots,\, n + l, \quad a_i = 0,\ i = n + l + 1,\, \ldots,\, n + l + m,$$
$$b_i = +\infty,\ i = 1,\, \ldots,\, n + l + m.$$

On the other hand, consider the VP

$$F(x) + N_D(x) \ni 0 \tag{1.43}$$

with

$$D = \{x \in \mathbf{R}^n \mid h(x) = 0,\ g(x) \leq 0\}.$$

Similarly to Theorem 1.14, one derives that if $\bar{x} \in \mathbf{R}^n$ solves this VP, and if the MFCQ holds at $\bar{x}$, or the linearity CQ holds, then there exist multipliers $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ such that $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ is a solution of the KKT system (1.39).

In particular, for a given smooth objective function $f : \mathbf{R}^n \to \mathbf{R}$, consider the mathematical programming problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0,\ g(x) \leq 0. \end{array} \tag{1.44}$$

Stationary points of problem (1.44) and the associated Lagrange multipliers are characterized by the KKT optimality system (1.39) with

$$F(x) = f'(x), \quad x \in \mathbf{R}^n \tag{1.45}$$

(see Sect. 1.2.4). Observe that in this case

$$G(x,\, \lambda,\, \mu) = \frac{\partial L}{\partial x}(x,\, \lambda,\, \mu), \quad x \in \mathbf{R}^n,\, \lambda \in \mathbf{R}^l,\ \mu \in \mathbf{R}^m, \tag{1.46}$$

where $L : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}$ is the Lagrangian of problem (1.44):

$$L(x,\, \lambda,\, \mu) = f(x) + \langle \lambda,\, h(x) \rangle + \langle \mu,\, g(x) \rangle.$$

Concerning the existence of solutions of VPs, we mention that the VI (1.36) with a nonempty (convex) set $Q \subset \mathbf{R}^\nu$ has a solution in any of the following cases:

- The set Q is compact and Φ is continuous on Q (see [68, Corollary 2.2.5]; this fact extends the Weierstrass Theorem to the variational setting).
- The set Q is closed and Φ is *coercive* on Q, i.e., there exist $\hat{u} \in Q$ and $\gamma > 0$ such that
$$\liminf_{k \to \infty} \langle \Phi(u^k),\, u^k - \hat{u} \rangle \geq \gamma$$

holds for any sequence $\{u^k\} \subset Q$ such that $\|u^k\| \to \infty$ (see [68, Proposition 2.2.7]; this also extends the corresponding fact for optimization problems).

Concerning the more special problem classes, we mention that if Φ is strongly monotone (see Sect. A.3 for the definition), then it is coercive on any closed set Q, and the VI (1.36) necessarily has the unique solution provided Q is nonempty. Furthermore, if Ψ is a maximal monotone (see Sect. A.3) set-valued mapping from $\mathbf{R}^\nu$ to the subsets of $\mathbf{R}^\nu$, then the solution set of the inclusion

$$\Psi(u) \ni 0$$

is convex. If, in addition, Ψ is strongly monotone with constant $\gamma > 0$, then the inclusion

$$\Psi(u) \ni r$$

has the unique solution $u(r)$ for every $r \in \mathbf{R}^\nu$, and the mapping $u(\cdot)$ (which is the inverse of Ψ) is Lipschitz-continuous on $\mathbf{R}^\nu$ with constant $1/\gamma$:

$$\|u(r^1) - u(r^2)\| \leq \frac{1}{\gamma}\|r^1 - r^2\| \quad \forall\, r^1,\, r^2 \in \mathbf{R}^\nu$$

(see [239, Proposition 12.54]). Those facts extend to the variational setting the properties of the existence and uniqueness of minimizers of strongly convex functions.

1.3.2 Stability and Sensitivity

We start with a version of the implicit function theorem for equations, which is a particular case of, e.g., [62, Theorem 2B.5].

Theorem 1.22. *Let $\Phi : \mathbf{R}^s \times \mathbf{R}^\nu \to \mathbf{R}^\nu$ be such that $\Phi(\bar{\sigma},\, \bar{u}) = 0$ for some $(\bar{\sigma},\, \bar{u}) \in \mathbf{R}^s \times \mathbf{R}^\nu$, and suppose that the following assumptions are satisfied:*

(i) There exists a matrix $J \in \mathbf{R}^{\nu \times \nu}$ such that for any $\varepsilon > 0$

$$\|\Phi(\sigma,\, u^1) - \Phi(\sigma,\, u^2) - J(u^1 - u^2)\| \leq \varepsilon\|u^1 - u^2\|$$

holds for all $\sigma \in \mathbf{R}^s$ close enough to $\bar{\sigma}$, and all $u^1,\, u^2 \in \mathbf{R}^n$ close enough to $\bar{u}$.

(ii) The matrix J is nonsingular.

(iii) The mapping $\Phi(\cdot,\, \bar{u})$ is continuous at $\bar{\sigma}$.

Then there exist neighborhoods $\mathcal{U}$ of $\bar{\sigma}$ and U of $\bar{u}$ such that for any $\sigma \in \mathcal{U}$ there exists the unique $u(\sigma) \in U$ such that $\Phi(\sigma,\, u(\sigma)) = 0$, and

$$u(\sigma) - \bar{u} = O(\|\Phi(\sigma,\, \bar{u})\|)$$

as $\sigma \to \bar{\sigma}$.

Note that, by necessity, under assumption (i) the mapping $\Phi(\bar{\sigma}, \cdot)$ is differentiable at $\bar{u}$, and $J = \frac{\partial \Phi}{\partial u}(\bar{\sigma}, \bar{u})$. On the other hand, from the mean-value theorem (Theorem A.10) it easily follows that if the mapping Φ is differentiable with respect to u near $(\bar{\sigma}, \bar{u})$, and $\frac{\partial \Phi}{\partial u}$ is continuous at $(\bar{\sigma}, \bar{u})$, then assumption (i) holds with $J = \frac{\partial \Phi}{\partial u}(\bar{\sigma}, \bar{u})$.

Assumption (ii) means that the Jacobian in question must be nonsingular. The following important and useful regularity concept, introduced in [234], can be regarded as an extension of this assumption to the GE (1.31). Assume that the base mapping Φ is differentiable at the reference point $\bar{u} \in \mathbf{R}^\nu$.

Definition 1.23. Solution $\bar{u}$ of the GE (1.31) is said to be *strongly regular* if for any $r \in \mathbf{R}^\nu$ close enough to 0, the perturbed (partially) linearized GE

$$\Phi(\bar{u}) + \Phi'(\bar{u})(u - \bar{u}) + N(u) \ni r \tag{1.47}$$

has near $\bar{u}$ the unique solution $u(r)$, and the mapping $u(\cdot)$ is locally Lipschitz-continuous at 0 (i.e., Lipschitz-continuous in some neighborhood of 0).

The role of this concept is clarified by the following implicit function-type result (cf. Theorem 1.22), first established in [234] (it can also be found in [27, Theorem 5.13], [62, Theorem 2B.1, Corollary 2B.3]). We emphasize that it is crucial that the field multifunction in the perturbed GE (1.48) below does not depend on the parameter.

Theorem 1.24. *Let $\Phi : \mathbf{R}^s \times \mathbf{R}^\nu \to \mathbf{R}^\nu$ be Lipschitz-continuous and differentiable with respect to u on some neighborhood of $(\bar{\sigma}, \bar{u}) \in \mathbf{R}^s \times \mathbf{R}^\nu$, with $\frac{\partial \Phi}{\partial u}$ being continuous at $(\bar{\sigma}, \bar{u})$. Let N be a multifunction from $\mathbf{R}^\nu$ to the subsets of $\mathbf{R}^\nu$. Assume that $\bar{u}$ is a strongly regular solution of the GE*

$$\Phi(\sigma, u) + N(u) \ni 0 \tag{1.48}$$

for $\sigma = \bar{\sigma}$.

Then there exist neighborhoods $\mathcal{U}$ of $\bar{\sigma}$ and U of $\bar{u}$ such that for any $\sigma \in \mathcal{U}$ there exists the unique $u(\sigma) \in U$ satisfying the GE (1.48), the mapping $u(\cdot)$ is locally Lipschitz-continuous at $\bar{\sigma}$, and

$$u(\sigma) - \bar{u} = O(\|\Phi(\sigma, \bar{u}) - \Phi(\bar{\sigma}, \bar{u})\|)$$

as $\sigma \to \bar{\sigma}$.

Theorem 1.24 can be derived in at least two different ways: from the classical contraction mapping principle, or from the following fact which essentially says that strong regularity is stable subject to small Lipschitzian perturbations. This fact follows, e.g., from the corresponding result for abstract multifunctions in [63, Theorem 4.1].

Theorem 1.25. *For a given $\Phi : \mathbf{R}^\nu \to \mathbf{R}^\nu$ differentiable at $\bar{u} \in \mathbf{R}^\nu$ and a multifunction N from $\mathbf{R}^\nu$ to the subsets of $\mathbf{R}^\nu$, let $\bar{u}$ be a strongly regular solution of the GE (1.31).*

Then for any fixed neighborhood W of $\bar{u}$, any sufficiently small $\ell \geq 0$, any mapping $R : \mathbf{R}^\nu \to \mathbf{R}^\nu$ which is Lipschitz-continuous on W with Lipschitz constant ℓ, and any $r \in \mathbf{R}^\nu$ close enough to $R(\bar{u})$, the GE

$$R(u) + \Phi(\bar{u}) + \Phi'(\bar{u})(u - \bar{u}) + N(u) \ni r$$

has near $\bar{u}$ the unique solution $u(r)$, and the mapping $u(\cdot)$ is locally Lipschitz-continuous at $R(\bar{u})$.

According to the discussion at the end of Sect. 1.3.1, if the multifunction $u \to \Phi'(\bar{u})(u - \bar{u}) + N(u)$ is maximal monotone and strongly monotone, then the solution $\bar{u}$ of the GE (1.31) is strongly regular.

A simple algebraic characterization of strong regularity for the NCP (1.38) was provided in [234], employing the notion of a P-matrix. Recall that $A \in \mathbf{R}^{\nu \times \nu}$ is referred to as a *P-matrix* if all its principal minors are positive. This property is equivalent to the following: for any $v \in \mathbf{R}^\nu \setminus \{0\}$, there exists an index $i \in \{1, \ldots, \nu\}$ such that $v_i(Av)_i > 0$; see [46, Theorem 3.3.4]. An evident sufficient condition for these equivalent properties is positive definiteness of A, that is,

$$\langle Av, v\rangle > 0 \quad \forall\, v \in \mathbf{R}^\nu \setminus \{0\}$$

(in the case of a symmetric A, the latter is also a necessary condition).

For a given solution $\bar{u}$ of the NCP (1.38), define the index sets

$$\begin{aligned} I_0(\bar{u}) &= \{i = 1, \ldots, \nu \mid \bar{u}_i = \Phi_i(\bar{u}) = 0\}, \\ I_1(\bar{u}) &= \{i = 1, \ldots, \nu \mid \bar{u}_i > 0,\ \Phi_i(\bar{u}) = 0\}, \\ I_2(\bar{u}) &= \{i = 1, \ldots, \nu \mid \bar{u}_i = 0,\ \Phi_i(\bar{u}) > 0\}. \end{aligned}$$

Characterization of strong regularity of a solution of an NCP is then the following.

Proposition 1.26. *Let $\Phi : \mathbf{R}^\nu \to \mathbf{R}^\nu$ be differentiable at $\bar{u} \in \mathbf{R}^\nu$, and let $\bar{u}$ be a solution of the NCP (1.38).*

Then $\bar{u}$ is a strongly regular solution if and only if for the index sets $I_0 = I_0(\bar{u})$, $I_1 = I_1(\bar{u})$, $I_2 = I_2(\bar{u})$ it holds that $(\Phi'(\bar{u}))_{I_1 I_1}$ is a nonsingular matrix, and its Schur complement

$$(\Phi'(\bar{u}))_{I_0 I_0} - (\Phi'(\bar{u}))_{I_0 I_1}((\Phi'(\bar{u}))_{I_1 I_1})^{-1}(\Phi'(\bar{u}))_{I_1 I_0}$$

in the matrix

$$\begin{pmatrix} (\Phi'(\bar{u}))_{I_1 I_1} & (\Phi'(\bar{u}))_{I_1 I_0} \\ (\Phi'(\bar{u}))_{I_0 I_1} & (\Phi'(\bar{u}))_{I_0 I_0} \end{pmatrix}$$

is a P-matrix.

For the KKT system (1.39), the following characterization of strong regularity was obtained in [171].

Proposition 1.27. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable at $\bar{x} \in \mathbf{R}^n$, and let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable at $\bar{x}$.*

The solution $(\bar{x}, \bar{\lambda}, \bar{\mu})$ of the KKT system (1.39) is strongly regular if and only if the determinants of matrices

$$\begin{pmatrix} \dfrac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) & (h'(\bar{x}))^{\mathrm{T}} & (g'_{A_+(\bar{x}, \bar{\mu}) \cup K}(\bar{x}))^{\mathrm{T}} \\ h'(x) & 0 & 0 \\ g'_{A_+(\bar{x}, \bar{\mu}) \cup K}(\bar{x}) & 0 & 0 \end{pmatrix},$$

where G is defined by (1.40), are distinct from zero and have the same sign for all index sets $K \subset A_0(\bar{x}, \bar{\mu})$. In particular, the LICQ is necessary for strong regularity of $(\bar{x}, \bar{\lambda}, \bar{\mu})$.

The first assertion of the next result was established in [234], while the second was derived in [28] (see also [27, Proposition 5.38]).

Proposition 1.28. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable at $\bar{x} \in \mathbf{R}^n$. Let $\bar{x}$ be a stationary point of the mathematical programming problem (1.44), and let $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be an associated Lagrange multiplier.*

If $\bar{x}$ and $(\bar{\lambda}, \bar{\mu})$ satisfy the LICQ and the strong second-order sufficient optimality condition (SSOSC)

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall\, \xi \in C_+(\bar{x}, \bar{\mu}) \setminus \{0\}, \tag{1.49}$$

where

$$C_+(\bar{x}, \bar{\mu}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x})\xi = 0\},$$

then $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a strongly regular solution of the KKT system (1.39) with F defined by (1.45).

Moreover, the SSOSC (1.49) is necessary for strong regularity of $(\bar{x}, \bar{\lambda}, \bar{\mu})$ provided $\bar{x}$ is a local solution of problem (1.44).

Another important regularity notion is that of semistability, introduced in [26].

Definition 1.29. Solution $\bar{u}$ of the GE (1.31) is said to be *semistable* if for every $r \in \mathbf{R}^n$ any solution $u(r)$ of the perturbed GE

$$\Phi(u) + N(u) \ni r, \tag{1.50}$$

close enough to $\bar{u}$, satisfies the estimate

$$u(r) - \bar{u} = O(\|r\|)$$

as $r \to 0$.

In other words, semistability of $\bar{u}$ means the Lipschitzian upper estimate of the distance from $\bar{u}$ to the solution set of the GE subject to the so-called right-hand side perturbations. Obviously, semistability implies that $\bar{u}$ is an isolated solution of the GE (1.31).

We start with the following necessary condition for semistability.

Lemma 1.30. *Let $\Phi : \mathbf{R}^\nu \to \mathbf{R}^\nu$ be differentiable at $\bar{u} \in \mathbf{R}^\nu$, and assume that $\bar{u}$ is a semistable solution of the GE (1.31).*

Then $v = 0$ is an isolated solution of the (partially) linearized GE

$$\Phi(\bar{u}) + \Phi'(\bar{u})v + N(\bar{u} + v) \ni 0. \tag{1.51}$$

Proof. We argue by contradiction. Suppose that there exists a sequence $\{v^k\} \subset \mathbf{R}^\nu \setminus \{0\}$ such that it converges to 0, and for each k the point v^k is a solution of (1.51). Then

$$\Phi(\bar{u} + v^k) + N(\bar{u} + v^k) = \Phi(\bar{u}) + \Phi'(\bar{u})v^k + N(\bar{u} + v^k) + r^k \ni r^k,$$

with some $r^k \in \mathbf{R}^\nu$ such that $\|r^k\| = o(\|v^k\|)$. Thus, $u^k = \bar{u} + v^k$ is a solution of the GE (1.50) with $r = r^k$, and semistability of $\bar{u}$ implies the estimates

$$v^k = u^k - \bar{u} = O(\|r^k\|) = o(\|v^k\|),$$

which is only possible if $v^k = 0$ for all k large enough. This gives a contradiction. □

As a sufficient condition for semistability, we mention strong regularity; this can be easily derived directly from Definitions 1.23 and 1.29, or from Theorem 1.24.

Observe also that according to the discussion above, if the multifunction $\Phi(\cdot) + N(\cdot)$ is maximal monotone and strongly monotone, then the unique solution of the GE (1.31) is automatically semistable.

Consider now the VI (1.36) over a closed convex $Q \subset \mathbf{R}^\nu$. The *natural residual mapping* $R : \mathbf{R}^\nu \to \mathbf{R}^\nu$ of this VI is defined as follows:

$$R(u) = u - \pi_Q(u - \Phi(u)).$$

From Lemma A.12 it readily follows that the VI (1.36) is equivalent to the equation

$$R(u) = 0.$$

Moreover, the following characterization of semistability is valid; see [68, Proposition 5.3.7].

Proposition 1.31. *Let $\Phi : \mathbf{R}^\nu \to \mathbf{R}^\nu$ be continuous at $\bar{u} \in \mathbf{R}^\nu$, let $Q \subset \mathbf{R}^\nu$ be closed and convex, and let $\bar{u}$ be a solution of the VI (1.36).*

Then semistability of $\bar{u}$ *is necessary for the error bound*

$$u - \bar{u} = O(\|R(u)\|) \tag{1.52}$$

to hold for $u \in \mathbf{R}^\nu$ *tending to* $\bar{u}$. *Moreover, if* Φ *is locally Lipschitz-continuous at* $\bar{u}$, *then semistability of* $\bar{u}$ *is also sufficient for* (1.52).

For the usual equation (i.e., when $Q = \mathbf{R}^\nu$) with a smooth mapping, Proposition 1.31 follows immediately from the definition of differentiability (see Sect. A.2). We state this fact separately, and in a slightly more general form allowing for the number of equations in the system to (possibly) exceed the number of variables.

Proposition 1.32. *Let* $\Phi : \mathbf{R}^\nu \to \mathbf{R}^\kappa$ *be differentiable at* $\bar{u} \in \mathbf{R}^\nu$, *and let* $\bar{u}$ *be a solution of the equation* (1.32).

Then the condition

$$\ker \Phi'(\bar{u}) = \{0\} \tag{1.53}$$

is necessary and sufficient for the error bound

$$u - \bar{u} = O(\|\Phi(u)\|)$$

to hold as $u \in \mathbf{R}^\nu$ *tends to* $\bar{u}$. *In particular,* (1.53) *implies that* $\bar{u}$ *is an isolated solution of the equation* (1.32).

In the case when the GE (1.31) is a VI over a polyhedral set, a criterion for semistability arises from Lemma 1.30.

Proposition 1.33. *Let* $\Phi : \mathbf{R}^\nu \to \mathbf{R}^\nu$ *be differentiable at* $\bar{u} \in \mathbf{R}^\nu$, *let* $Q \subset \mathbf{R}^\nu$ *be a polyhedral set, and let* $\bar{u}$ *be a solution of the VI* (1.36).

Then $\bar{u}$ *is semistable if and only if* $v = 0$ *is an isolated solution of the linearized VI*

$$\bar{u} + v \in Q, \quad \langle \Phi(\bar{u}) + \Phi'(\bar{u})v,\, u - \bar{u} - v\rangle \geq 0 \quad \forall\, u \in Q. \tag{1.54}$$

Proof. Necessity follows from Lemma 1.30, since (1.54) is equivalent to the GE (1.51) with $N(\cdot) = N_Q(\cdot)$. For the proof of sufficiency, see [26]. $\square$

We proceed with the special cases of the NCP (1.38) and the KKT system (1.39).

Proposition 1.34. *Let* $\Phi : \mathbf{R}^\nu \to \mathbf{R}^\nu$ *be differentiable at* $\bar{u} \in \mathbf{R}^\nu$, *and let* $\bar{u}$ *be a solution of the NCP* (1.38).

Then $\bar{u}$ *is semistable if and only if the system*

$$\begin{gathered} v_i \geq 0, \quad \langle \Phi_i'(\bar{u}),\, v\rangle \geq 0, \quad v_i\langle \Phi_i'(\bar{u}),\, v\rangle = 0, \quad i \in I_0(\bar{u}), \\ \Phi'_{I_1(\bar{u})}(\bar{u})v = 0, \quad v_{I_2(\bar{u})} = 0 \end{gathered} \tag{1.55}$$

has only the zero solution.

Proof. The assertion follows from Proposition 1.33, since near the zero point the VI (1.54) with $Q = \mathbf{R}_+^\nu$ has the same solutions as the system (1.55). $\square$

Proposition 1.35. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable at $\bar{x} \in \mathbf{R}^n$, let the mappings $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable at $\bar{x}$, and let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ with some $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be a solution of the KKT system (1.39).*

Then $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is semistable if and only if the system

$$
\begin{gathered}
\frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi + (h'(\bar{x}))^{\mathrm{T}}\eta + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0, \\
h'(\bar{x})\xi = 0, \quad g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x})\xi = 0, \\
\zeta_{A_0(\bar{x}, \bar{\mu})} \geq 0, \quad g'_{A_0(\bar{x}, \bar{\mu})}(\bar{x})\xi \leq 0, \quad \zeta_i\langle g_i'(\bar{x}), \xi\rangle = 0, \ i \in A_0(\bar{x}, \bar{\mu}), \\
\zeta_{\{1, \ldots, m\} \setminus A(\bar{x})} = 0
\end{gathered}
\tag{1.56}
$$

with G defined by (1.40) has the unique solution $(\xi, \eta, \zeta) = (0, 0, 0)$.

Proof. It is easy to see that near the zero point, for Φ defined in (1.41) and Q defined in (1.42), the VI (1.54) with respect to $v = (\xi, \eta, \zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ has the same solutions as the system (1.56). The needed result now follows from Proposition 1.33. $\square$

We mention, in passing, that both Propositions 1.34 and 1.35 can be easily proved directly, by means of the so-called piecewise analysis [129, 130, 142]. This approach will be used in Sect. 1.3.3 to establish error bounds for KKT systems.

Remark 1.36. One more issue must be mentioned in connection with Proposition 1.35. As discussed above, semistability of $(\bar{x}, \bar{\lambda}, \bar{\mu})$ implies that it is an isolated solution of (1.39), which in turn implies that $(\bar{\lambda}, \bar{\mu})$ is the unique Lagrange multiplier associated with $\bar{x}$ (the latter follows from the fact that the set $\mathcal{M}(\bar{x})$ is polyhedral, hence, convex). Another way to come to the same conclusion is to take $\xi = 0$ in (1.56), thus reducing it to the system

$$
(h'(\bar{x}))^{\mathrm{T}}\eta + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0, \quad \zeta_{A_0(\bar{x}, \bar{\mu})} \geq 0, \quad \zeta_{\{1, \ldots, m\} \setminus A(\bar{x})} = 0. \tag{1.57}
$$

Similarly to Remark 1.15, one can see that the latter system has only the zero solution if and only if $(\bar{\lambda}, \bar{\mu})$ is the unique Lagrange multiplier associated with $\bar{x}$.

Proposition 1.37. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable at $\bar{x} \in \mathbf{R}^n$. Let $\bar{x}$ be a stationary point of the mathematical programming problem (1.44), and let $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be an associated Lagrange multiplier.*

If $\bar{x}$ and $(\bar{\lambda}, \bar{\mu})$ satisfy the SMFCQ and the SOSC

$$
\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{1.58}
$$

where

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A(\bar{x})}(\bar{x})\xi \leq 0,\ \langle f'(\bar{x}),\, \xi\rangle \leq 0\}$$

is the critical cone of problem (1.44) *at* $\bar{x}$*, then* $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ *is a semistable solution of the KKT system* (1.39) *with* F *defined by* (1.45)*.*

Moreover, the SMFCQ is necessary for semistability of $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$*, while the SOSC* (1.58) *is necessary provided* $\bar{x}$ *is a local solution of problem* (1.44)*.*

Proof. According to Proposition 1.35, semistability of $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ is equivalent to the following: the system

$$\begin{gathered}
\frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi + (h'(\bar{x}))^{\mathrm{T}}\eta + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0,\\
h'(\bar{x})\xi = 0, \quad g'_{A_+(\bar{x},\, \bar{\mu})}(\bar{x})\xi = 0,\\
\zeta_{A_0(\bar{x},\, \bar{\mu})} \geq 0, \quad g'_{A_0(\bar{x},\, \bar{\mu})}(\bar{x})\xi \leq 0, \quad \zeta_i\langle g_i'(\bar{x}),\, \xi\rangle = 0,\ i \in A_0(\bar{x},\, \bar{\mu}),\\
\zeta_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})} = 0
\end{gathered} \tag{1.59}$$

has the unique solution $(\xi,\, \eta,\, \zeta) = (0,\, 0,\, 0)$.

Suppose first that the SMFCQ and the SOSC (1.58) hold but the system (1.59) has a solution $(\xi,\, \eta,\, \zeta) \neq (0,\, 0,\, 0)$. Then, according to Lemma 1.17, $\xi \in C(\bar{x})$ and multiplying the first equation in (1.59) by ξ, we derive

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi \right\rangle = -\langle \eta,\, h'(\bar{x})\xi\rangle - \langle \zeta,\, g'(\bar{x})\xi\rangle = 0.$$

According to the SOSC (1.58), the latter is only possible when $\xi = 0$. Therefore, from (1.59) we obtain that the system (1.57) has a solution $(\eta,\, \zeta) \neq (0,\, 0)$, which is in contradiction with the SMFCQ (see Remark 1.15). This proves that the SMFCQ and the SOSC imply semistability.

The necessity of the SMFCQ follows from the discussion preceding this proposition.

It remains to show that the SOSC is also necessary for semistability when $\bar{x}$ is a local solution of problem (1.44). According to Theorem 1.19, local optimality of $\bar{x}$ satisfying the SMFCQ implies the SONC in the form

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi \right\rangle \geq 0 \quad \forall\, \xi \in C(\bar{x}). \tag{1.60}$$

Suppose that the SOSC (1.58) does not hold. According to (1.60), the latter means the existence of $\xi \in C(\bar{x}) \setminus \{0\}$ such that

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi \right\rangle = 0, \tag{1.61}$$

and this ξ is a solution of the problem

$$\begin{array}{ll}
\text{minimize} & \left\langle \dfrac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi \right\rangle\\
\text{subject to} & \xi \in C(\bar{x}).
\end{array} \tag{1.62}$$

Recall that according to Lemma 1.17,

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+(\bar{x},\,\bar{\mu})}(\bar{x})\xi = 0,\ g'_{A_0(\bar{x},\,\bar{\mu})}(\bar{x})\xi \leq 0\}.$$

Therefore, (1.62) is a linearly constrained problem, and by Theorem 1.14 we conclude that ξ satisfies the KKT system of this problem with some associated Lagrange multipliers. But this KKT system is precisely (1.59), which gives a contradiction because $\xi \neq 0$. $\square$

Consider now the general parametric KKT system

$$\begin{gathered} F(\sigma,\, x) + \left(\frac{\partial h}{\partial x}(\sigma,\, x)\right)^{\mathrm{T}} \lambda + \left(\frac{\partial g}{\partial x}(\sigma,\, x)\right)^{\mathrm{T}} \mu = 0, \quad h(\sigma,\, x) = 0, \\ \mu \geq 0, \quad g(\sigma,\, x) \leq 0, \quad \langle \mu,\, g(\sigma,\, x)\rangle = 0, \end{gathered} \tag{1.63}$$

with respect to $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, where $\sigma \in \mathbf{R}^s$ is a parameter, and $F : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^n$, $h : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^m$ are sufficiently smooth mappings. Define the mapping $G : \mathbf{R}^s \times \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n$,

$$G(\sigma,\, x,\, \lambda,\, \mu) = F(\sigma,\, x) + \left(\frac{\partial h}{\partial x}(\sigma,\, x)\right)^{\mathrm{T}} \lambda + \left(\frac{\partial g}{\partial x}(\sigma,\, x)\right)^{\mathrm{T}} \mu. \tag{1.64}$$

Let $\bar{\sigma} \in \mathbf{R}^s$ be some fixed (base) parameter value, and let $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ be a solution of the system (1.63) with $\sigma = \bar{\sigma}$. The following theorem generalizes Proposition 1.35; it gives a sufficient condition for the Lipschitzian upper estimate of the distance from $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ to the solution set of the KKT system subject to general parametric perturbations. This property is known in the literature under various names; see [178] for a discussion of this terminology. With different levels of generality, this result can be found, e.g., in [27, Theorem 5.9], [169, Theorem 8.11 and Corollary 8.13], and in [168, 178, 245]. A simple proof based on the piecewise analysis is given in [130]. An alternative possibility, adopted here, is to use Proposition 1.35 and the observation that under appropriate smoothness assumptions, the needed stability property for general parametric perturbations is in fact equivalent to its counterpart for the right-hand side perturbations of the GE corresponding to the KKT system. This type of perturbations is usually referred to as *canonical perturbations* of the KKT system (see (1.82) below). Note also that the upper Lipschitzian stability subject to canonical perturbations is precisely the semistability property of the solution is question.

Theorem 1.38. *Let $F : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^n$ be differentiable with respect to x at $(\bar{\sigma},\, \bar{x}) \in \mathbf{R}^s \times \mathbf{R}^n$, and let the mappings $h : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^m$ be differentiable with respect to x in a neighborhood of $(\bar{\sigma},\, \bar{x})$ and twice differentiable with respect to x at $(\bar{\sigma},\, \bar{x})$. Let F, g, $\frac{\partial h}{\partial x}$ and $\frac{\partial g}{\partial x}$ be continuous at $(\bar{\sigma},\, \bar{x})$. Let $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ with some $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be a solution of the system (1.63) for $\sigma = \bar{\sigma}$.*

If the system

$$\left(\frac{\partial h}{\partial x}(\bar{\sigma},\, \bar{x})\right)^{\mathrm{T}} \eta + \left(\frac{\partial g}{\partial x}(\bar{\sigma},\, \bar{x})\right)^{\mathrm{T}} \zeta = 0, \\ \zeta_{A_0(\bar{\sigma},\, \bar{x},\, \bar{\mu})} \geq 0, \quad \zeta_{\{1,\, \dots,\, m\} \setminus A(\bar{\sigma},\, \bar{x})} = 0 \tag{1.65}$$

has the unique solution $(\eta,\, \zeta) = (0,\, 0)$*, then for* $\sigma \in \mathbf{R}^s$ *and for any solution* $(x(\sigma),\, \lambda(\sigma),\, \mu(\sigma))$ *of the system* (1.63) *satisfying* $x(\sigma) \to \bar{x}$ *as* $\sigma \to \bar{\sigma}$*, it holds that* $(\lambda(\sigma),\, \mu(\sigma)) \to (\bar{\lambda},\, \bar{\mu})$ *as* $\sigma \to \bar{\sigma}$*.*

Moreover, if the system

$$\begin{gathered} \frac{\partial G}{\partial x}(\bar{\sigma},\, \bar{x},\, \bar{\lambda},\, \bar{\mu})\xi + \left(\frac{\partial h}{\partial x}(\bar{\sigma},\, \bar{x})\right)^{\mathrm{T}} \eta + \left(\frac{\partial g}{\partial x}(\bar{\sigma},\, \bar{x})\right)^{\mathrm{T}} \zeta = 0, \\ \frac{\partial h}{\partial x}(\bar{\sigma},\, \bar{x})\xi = 0, \quad \frac{\partial g_{A_+(\bar{\sigma},\, \bar{x},\, \bar{\mu})}}{\partial x}(\bar{\sigma},\, \bar{x})\xi = 0, \\ \zeta_{A_0(\bar{\sigma},\, \bar{x},\, \bar{\mu})} \geq 0, \quad \frac{\partial g_{A_0(\bar{\sigma},\, \bar{x},\, \bar{\mu})}}{\partial x}(\bar{\sigma},\, \bar{x})\xi \leq 0, \\ \zeta_i \left\langle \frac{\partial g_i}{\partial x}(\bar{\sigma},\, \bar{x}),\, \xi \right\rangle = 0, \quad i \in A_0(\bar{\sigma},\, \bar{x},\, \bar{\mu}), \\ \zeta_{\{1,\, \dots,\, m\} \setminus A(\bar{\sigma},\, \bar{x})} = 0, \end{gathered} \tag{1.66}$$

where G *is defined by* (1.64)*, has the unique solution* $(\xi,\, \eta,\, \zeta) = (0,\, 0,\, 0)$*, and if, in addition,*

$$\|F(\sigma,\, x) - F(\bar{\sigma},\, x)\| = O(\|\sigma - \bar{\sigma}\|) + o(\|x - \bar{x}\|), \tag{1.67}$$

$$\|h(\sigma,\, x) - h(\bar{\sigma},\, x)\| = O(\|\sigma - \bar{\sigma}\|) + o(\|x - \bar{x}\|), \tag{1.68}$$

$$\|g(\sigma,\, x) - g(\bar{\sigma},\, x)\| = O(\|\sigma - \bar{\sigma}\|) + o(\|x - \bar{x}\|), \tag{1.69}$$

$$\left\| \frac{\partial h}{\partial x}(\sigma,\, x) - \frac{\partial h}{\partial x}(\bar{\sigma},\, x) \right\| = O(\|\sigma - \bar{\sigma}\|) + o(\|x - \bar{x}\|), \tag{1.70}$$

$$\left\| \frac{\partial g}{\partial x}(\sigma,\, x) - \frac{\partial g}{\partial x}(\bar{\sigma},\, x) \right\| = O(\|\sigma - \bar{\sigma}\|) + o(\|x - \bar{x}\|) \tag{1.71}$$

as $(\sigma,\, x) \to (\bar{\sigma},\, \bar{x})$*, then for any solution* $(x(\sigma),\, \lambda(\sigma),\, \mu(\sigma))$ *of the system* (1.63) *such that* $x(\sigma)$ *is close enough to* $\bar{x}$*, it holds that*

$$(x(\sigma) - \bar{x},\, \lambda(\sigma) - \bar{\lambda},\, \mu(\sigma) - \bar{\mu}) = O(\|\sigma - \bar{\sigma}\|) \tag{1.72}$$

as $\sigma \to \bar{\sigma}$*.*

(Conditions (1.67)–(1.71) are automatically satisfied if, e.g., F, h, g, $\frac{\partial h}{\partial x}$ and $\frac{\partial g}{\partial x}$ are differentiable at $(\bar{\sigma},\, \bar{x})$.)

Proof. Consider any sequence $\{\sigma^k\} \subset \mathbf{R}^s$ convergent to $\bar{\sigma}$, and any sequence $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ such that $\{x^k\}$ converges to $\bar{x}$ and for each k the triple $(x^k,\, \lambda^k,\, \mu^k)$ is a solution of the system (1.63) with $\sigma = \sigma^k$. We need to show that in this case the sequence $\{(\lambda^k,\, \mu^k)\}$ necessarily converges to $(\bar{\lambda},\, \bar{\mu})$.

If $\{(\lambda^k,\, \mu^k)\}$ were unbounded, then dividing (1.63) by $\|(\lambda^k,\, \mu^k)\|$, where in (1.63) $\sigma = \sigma^k$ and $(x,\, \lambda,\, \mu) = (x^k,\, \lambda^k,\, \mu^k)$, and passing onto the limit along the appropriate subsequence, we obtain the existence of $(\eta,\, \zeta) \in \mathbf{R}^l \times \mathbf{R}^m$, $\|(\eta,\, \zeta)\| = 1$, satisfying (1.65). This contradicts the assumption that (1.65) has only the trivial solution.

Hence, the sequence $\{(\lambda^k,\, \mu^k)\}$ is bounded. Then it has an accumulation point. For any accumulation point $(\tilde{\lambda},\, \tilde{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ of the sequence in question, passing onto the limit along the appropriate subsequence in (1.63) with $\sigma = \sigma^k$ and $(x,\, \lambda,\, \mu) = (x^k,\, \lambda^k,\, \mu^k)$, we obtain that $(\tilde{\lambda},\, \tilde{\mu})$ is a Lagrange multiplier associated with $\bar{x}$ for the system (1.63) with $\sigma = \bar{\sigma}$. As discussed in Remark 1.36, since the system (1.65) (cf. (1.57)) has only the trivial solution, the multiplier in question is unique. Therefore, we have that $(\tilde{\lambda},\, \tilde{\mu}) = (\bar{\lambda},\, \bar{\mu})$. This means that $\{(\lambda^k,\, \mu^k)\}$ converges to $(\bar{\lambda},\, \bar{\mu})$, establishing the first assertion of the theorem.

In order to prove the second assertion, observe first that (1.66) is a counterpart of the system (1.56) for the parametric KKT system (1.63). Therefore, according to Proposition 1.35, under the stated assumptions $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ is a semistable solution of the system (1.63) for $\sigma = \bar{\sigma}$, and in particular, according to Remark 1.36, the system (1.65) has only the trivial solution.

Furthermore, suppose that conditions (1.67)–(1.71) hold. For each $\sigma \in \mathbf{R}^s$ close enough to $\bar{\sigma}$ and each $x(\sigma) \in \mathbf{R}^n$ close enough to $\bar{x}$, define

$$\begin{aligned} a &= a(\sigma) \\ &= \left(F(\bar{\sigma},\, x(\sigma)) + \left(\frac{\partial h}{\partial x}(\bar{\sigma},\, x(\sigma))\right)^{\mathrm{T}} \lambda(\sigma) + \left(\frac{\partial g}{\partial x}(\bar{\sigma},\, x(\sigma))\right)^{\mathrm{T}} \mu(\sigma)\right) \\ &\quad - \left(F(\sigma,\, x(\sigma)) + \left(\frac{\partial h}{\partial x}(\sigma,\, x(\sigma))\right)^{\mathrm{T}} \lambda(\sigma) + \left(\frac{\partial g}{\partial x}(\sigma,\, x(\sigma))\right)^{\mathrm{T}} \mu(\sigma)\right), \end{aligned}$$

$$b = b(\sigma) = h(\bar{\sigma},\, x(\sigma)) - h(\sigma,\, x(\sigma)), \quad c = c(\sigma) = g(\bar{\sigma},\, x(\sigma)) - g(\sigma,\, x(\sigma)).$$

Assuming that $(x(\sigma),\, \lambda(\sigma),\, \mu(\sigma))$ is a solution of the system (1.63), it then follows that this triple is a solution of the canonically perturbed KKT system

$$\begin{gathered} F(\bar{\sigma},\, x) + \left(\frac{\partial h}{\partial x}(\bar{\sigma},\, x)\right)^{\mathrm{T}} \lambda + \left(\frac{\partial g}{\partial x}(\bar{\sigma},\, x)\right)^{\mathrm{T}} \mu = a, \quad h(\bar{\sigma},\, x) = b, \\ \mu \geq 0, \quad g(\bar{\sigma},\, x) \leq c, \quad \langle \mu,\, g(\bar{\sigma},\, x) - c\rangle = 0. \end{gathered}$$

On the other hand, employing (1.67), (1.70), and (1.71), from the definition of a we derive that

$$a(\sigma) = O(\|\sigma - \bar{\sigma}\|) + o(\|x(\sigma) - \bar{x}\|)$$

as $\sigma \to \bar{\sigma}$ and $x(\sigma) \to \bar{x}$. Similarly, employing (1.68) and (1.69), from the definitions of b and c we derive that

$$b(\sigma) = O(\|\sigma - \bar{\sigma}\|) + o(\|x(\sigma) - \bar{x}\|), \quad c(\sigma) = O(\|\sigma - \bar{\sigma}\|) + o(\|x(\sigma) - \bar{x}\|).$$

Therefore, by the semistability of $(\bar{x}, \bar{\lambda}, \bar{\mu})$ we obtain the estimate

$$\|x(\sigma) - \bar{x}\| + \|\lambda(\sigma) - \bar{\lambda}\| + \|\mu(\sigma) - \bar{\mu}\| = O(\|\sigma - \bar{\sigma}\|) + o(\|x(\sigma) - \bar{x}\|),$$

evidently implying that (1.72) holds as $\sigma \to \bar{\sigma}$ provided $x(\sigma)$ is taken close enough to $\bar{x}$. □

The estimate (1.72) certainly subsumes that $x(\sigma) \to \bar{x}$ as $\sigma \to \bar{\sigma}$. An interesting corollary of the first assertion of Theorem 1.38 is that this primal stability property can actually be established without (1.67)–(1.71), but under some additional continuity assumptions on the data mappings and their derivatives.

Corollary 1.39. *Let $F : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^n$ be differentiable with respect to x at $(\bar{\sigma}, \bar{x}) \in \mathbf{R}^s \times \mathbf{R}^n$, and let the mappings $h : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^m$ be differentiable with respect to x in a neighborhood of $(\bar{\sigma}, \bar{x})$ and twice differentiable with respect to x at $(\bar{\sigma}, \bar{x})$. Let F, h, g, $\frac{\partial h}{\partial x}$ and $\frac{\partial g}{\partial x}$ be continuous in a neighborhood of $(\bar{\sigma}, \bar{x})$. Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ with some $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be a solution of the system (1.63) for $\sigma = \bar{\sigma}$.*

If the system (1.66), where G is defined by (1.64), has the unique solution $(\xi, \eta, \zeta) = (0, 0, 0)$, then there exists $\delta > 0$ such that for any solution $(x(\sigma), \lambda(\sigma), \mu(\sigma))$ of the system (1.63) satisfying $\|x(\sigma) - \bar{x}\| \leq \delta$, it holds that

$$(x(\sigma), \lambda(\sigma), \mu(\sigma)) \to (\bar{x}, \bar{\lambda}, \bar{\mu}) \text{ as } \sigma \to \bar{\sigma}. \tag{1.73}$$

Proof. Recall again that by Proposition 1.35, under the stated assumptions $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a semistable (and hence, isolated) solution of the system (1.63) for $\sigma = \bar{\sigma}$. Therefore, there exists $\varepsilon > 0$ such that $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is the unique solution of this system in $B((\bar{x}, \bar{\lambda}, \bar{\mu}), \varepsilon)$. The first assertion of Theorem 1.38 implies the existence of $\delta \in (0, \varepsilon/2]$ such that for each $\sigma \in \mathbf{R}^s$ close enough to $\bar{\sigma}$, and for each solution $(x(\sigma), \lambda(\sigma), \mu(\sigma))$ of the system (1.63) satisfying $\|x(\sigma) - \bar{x}\| \leq \delta$, it holds that

$$\|(\lambda(\sigma) - \bar{\lambda}, \mu(\sigma) - \bar{\mu})\| \leq \frac{\varepsilon}{2}.$$

Consider now any sequence $\{\sigma^k\} \subset \mathbf{R}^s$ convergent to $\bar{\sigma}$, and any sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ such that for each k the triple (x^k, λ^k, μ^k) is a solution of the system (1.63) with $\sigma = \sigma^k$, and $\|x^k - \bar{x}\| \leq \delta$. Then holds true the inclusion $\{(x^k, \lambda^k, \mu^k)\} \subset B((\bar{x}, \bar{\lambda}, \bar{\mu}), \varepsilon)$, and the bounded sequence $\{(x^k, \lambda^k, \mu^k)\}$ has an accumulation point. For any accumulation point $(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \in B((\bar{x}, \bar{\lambda}, \bar{\mu}), \varepsilon)$, passing onto the limit along the appropriate subsequence in (1.63) with $\sigma = \sigma^k$ and $(x, \lambda, \mu) = (x^k, \lambda^k, \mu^k)$, we conclude that $(\tilde{x}, \tilde{\lambda}, \tilde{\mu})$ is a solution of the system (1.63) with $\sigma = \bar{\sigma}$, and by the choice of ε, this is only possible when $(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) = (\bar{x}, \bar{\lambda}, \bar{\mu})$. Therefore, $\{(x^k, \lambda^k, \mu^k)\}$ converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$. □

Consider now the parametric mathematical programming problem

$$\begin{array}{ll} \text{minimize} & f(\sigma,\, x) \\ \text{subject to} & h(\sigma,\, x) = 0,\;\; g(\sigma,\, x) \leq 0, \end{array} \tag{1.74}$$

where $\sigma \in \mathbf{R}^s$ is a parameter, and the objective function $f : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}$ and the constraints mappings $h : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^m$ are given. Let $L : \mathbf{R}^s \times \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}$ be the Lagrangian of problem (1.74):

$$L(\sigma,\, x,\, \lambda,\, \mu) = f(\sigma,\, x) + \langle \lambda,\, h(\sigma,\, x)\rangle + \langle \mu,\, g(\sigma,\, x)\rangle.$$

In the optimization case, primal (and hence dual) stability can be established under the SMFCQ and the SOSC, without any additional continuity or smoothness assumptions.

Theorem 1.40. *Let $f : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}$, and the mappings $h : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^m$ possess the following properties for $\bar{\sigma} \in \mathbf{R}^s$ and $\bar{x} \in \mathbf{R}^n$:*

(i) f is continuous at $(\bar{\sigma},\, x)$, and $h(\cdot,\, x)$ and $g(\cdot,\, x)$ are continuous at $\bar{\sigma}$, for every $x \in \mathbf{R}^n$ close enough to $\bar{x}$.
(ii) f, h, and g are differentiable with respect to x in a neighborhood of $(\bar{\sigma},\, \bar{x})$ and twice differentiable with respect to x at $(\bar{\sigma},\, \bar{x})$.
(iii) g, $\frac{\partial f}{\partial x}$, $\frac{\partial h}{\partial x}$, and $\frac{\partial g}{\partial x}$ are continuous at $(\bar{\sigma},\, \bar{x})$, and there exists a neighborhood of $\bar{x}$ such that $\frac{\partial h}{\partial x}(\sigma,\, \cdot)$ is continuous on this neighborhood for all $\sigma \in \mathbf{R}^s$ close enough to $\bar{\sigma}$.

Let $\bar{x}$ be a local solution of problem (1.74) with $\sigma = \bar{\sigma}$, satisfying the SMFCQ and the SOSC

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{\sigma},\, \bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{\sigma},\, \bar{x}) \setminus \{0\} \tag{1.75}$$

for the associated Lagrange multiplier $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$, where

$$C(\bar{\sigma},\, \bar{x}) = \left\{ \xi \in \mathbf{R}^n \;\middle|\; \frac{\partial h}{\partial x}(\bar{\sigma},\, \bar{x})\xi = 0,\, \frac{\partial g_{A(\bar{x})}}{\partial x}(\bar{\sigma},\, \bar{x})\xi \leq 0,\, \left\langle \frac{\partial f}{\partial x}(\bar{\sigma},\, \bar{x}),\, \xi \right\rangle \leq 0 \right\}$$

is the critical cone of this problem at $\bar{x}$.

Then for every $\sigma \in \mathbf{R}^s$ close enough to $\bar{\sigma}$, problem (1.74) has a local solution $x(\sigma)$ such that it is a stationary point of this problem, and for every associated Lagrange multiplier $(\lambda(\sigma),\, \mu(\sigma))$ the property (1.73) is valid. If, in addition,

$$\left\| \frac{\partial f}{\partial x}(\sigma,\, x) - \frac{\partial f}{\partial x}(\bar{\sigma},\, x) \right\| = O(\|\sigma - \bar{\sigma}\|) + o(\|x - \bar{x}\|), \tag{1.76}$$

and (1.68)–(1.71) hold as $(\sigma,\, x) \to (\bar{\sigma},\, \bar{x})$, then the estimate (1.72) is valid.

(Observe again that conditions (1.68)–(1.71) and (1.76) are automatically satisfied if, e.g., h, g, $\frac{\partial f}{\partial x}$, $\frac{\partial h}{\partial x}$ and $\frac{\partial g}{\partial x}$ are differentiable at $(\bar{\sigma}, \bar{x})$.)

Proof. According to Theorem 1.20, the SOSC (1.75) implies that $\bar{x}$ is a strict local minimizer of problem (1.74) with $\sigma = \bar{\sigma}$, and the SMFCQ implies the MFCQ at $\bar{x}$. From Theorem 1.21 it then follows that for each $\sigma \in \mathbf{R}^s$ close enough to $\bar{\sigma}$, problem (1.74) has a local solution $x(\sigma)$ such that $x(\sigma) \to \bar{x}$ as $\sigma \to \bar{\sigma}$. Since the MFCQ is stable subject to small perturbations (see Remark 1.3), we conclude that for σ close enough to $\bar{\sigma}$, the MFCQ holds at $x(\sigma)$. Hence, by Theorem 1.14, $x(\sigma)$ is a stationary point of problem (1.74) with some Lagrange multiplier $(\lambda(\sigma), \mu(\sigma))$. The needed assertions follow now by Theorem 1.38 and Propositions 1.35 and 1.37. $\square$

1.3.3 Error Bounds

In this section we discuss error bounds for KKT systems, i.e., estimates of the distance from a given point to the solution set of the KKT system (1.39). We consider the case when for a given $\bar{x}$, the associated set $\mathcal{M}(\bar{x})$ of Lagrange multipliers is nonempty but not necessarily a singleton. The following notion is crucial for this subject; it was introduced in [150].

Definition 1.41. A Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ is called *critical* if there exists a triple $(\xi, \eta, \zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, with $\xi \neq 0$, satisfying the system (1.56), and *noncritical* otherwise.

When there are no inequality constraints, it can be seen from (1.56) that a multiplier $\bar{\lambda} \in \mathcal{M}(\bar{x})$ being critical means that there exists $\xi \in \ker h'(\bar{x}) \setminus \{0\}$ such that

$$\frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda})\xi \in \operatorname{im}(h'(\bar{x}))^{\mathrm{T}}, \tag{1.77}$$

where G is defined by (1.40). For purely equality-constrained problems, this concept was first introduced and studied in [128]. In the fully degenerate case (i.e., when $h'(\bar{x}) = 0$), criticality of $\bar{\lambda}$ simply means that the matrix $\frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda})$ is singular. Generally, since $\operatorname{im}(h'(\bar{x}))^{\mathrm{T}} = (\ker h'(\bar{x}))^{\perp}$, condition (1.77) can be equivalently written in the form

$$\pi_{\ker h'(\bar{x})} \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda})\xi = 0.$$

The (square) matrix of the linear operator

$$\xi \to \pi_{\ker h'(\bar{x})} \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda})\xi : \ker h'(\bar{x}) \to \ker h'(\bar{x})$$

can be regarded as the reduced Jacobian of G with respect to x. Then criticality of $\bar{\lambda}$ means that this reduced Jacobian is singular.

From Proposition 1.35, Remark 1.36, and Definition 1.41, it readily follows that semistability of a solution $(\bar{x}, \bar{\lambda}, \bar{\mu})$ of the KKT system (1.39) is equivalent to the combination of two properties: $(\bar{\lambda}, \bar{\mu})$ must be the unique Lagrange multiplier associated with $\bar{x}$, and this multiplier must be noncritical.

Define the *critical cone* of the KKT system (1.39):

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A(\bar{x})}(\bar{x})\xi \leq 0,\ \langle F(\bar{x}),\, \xi\rangle = 0\}.$$

Similarly to Lemma 1.17, it can be easily checked that for any solution $(\bar{x}, \bar{\lambda}, \bar{\mu})$ of (1.39) it holds that

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+(\bar{x},\, \bar{\mu})}(\bar{x})\xi = 0,\ g'_{A_0(\bar{x},\, \bar{\mu})}(\bar{x})\xi \leq 0\}.$$

It can be observed that when $\frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a symmetric matrix, (1.56) is the KKT system for the quadratic programming problem

$$\begin{array}{ll} \text{minimize} & \dfrac{1}{2}\left\langle \dfrac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi,\, \xi \right\rangle \\ \text{subject to} & \xi \in C(\bar{x}). \end{array} \tag{1.78}$$

Therefore, in this case, $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ being noncritical is equivalent to saying that $\xi = 0$ is the unique stationary point of problem (1.78). Furthermore, it then follows that $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ being noncritical is also equivalent to saying that the Jacobian $\frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})$ and the critical cone $C(\bar{x})$ are an R_0-pair in the sense of [68].

Note that multiplying the first equality in (1.56) by ξ and using the other relations in (1.56), it can be easily seen that under the second-order sufficiency condition

$$\left\langle \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{1.79}$$

as well as under its "opposite-sign" counterpart

$$\left\langle \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi,\, \xi \right\rangle < 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{1.80}$$

the multiplier $(\bar{\lambda}, \bar{\mu})$ is necessarily noncritical. However, there may also exist noncritical multipliers that satisfy neither (1.79) nor (1.80).

At the same time, if $\bar{x}$ and $(\bar{\lambda}, \bar{\mu})$ satisfy the SONC (1.60) for mathematical programming problem (1.44), then the multiplier $(\bar{\lambda}, \bar{\mu})$ is noncritical (in the sense of Definition 1.41 applied with F defined by (1.45) and G defined by (1.46)) if and only if the SOSC (1.58) holds. Recall, however, that as discussed in Sect. 1.2.4, in the absence of the SMFCQ, the SONC (1.60) does not necessarily hold at a local solution $\bar{x}$ of problem (1.44) (cf. Propositions 1.28 and 1.37, where the SMFCQ is necessarily valid).

Proposition 1.42. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable at $\bar{x} \in \mathbf{R}^n$. Let $\bar{x}$ be a stationary point of problem (1.44), and let $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be an associated Lagrange multiplier.*

If $\bar{x}$ and $(\bar{\lambda}, \bar{\mu})$ satisfy the SOSC (1.58), then the multiplier $(\bar{\lambda}, \bar{\mu})$ is noncritical.

Moreover, the SOSC (1.58) is necessary for noncriticality of $(\bar{\lambda}, \bar{\mu})$ provided $\bar{x}$ satisfies the SONC (1.60).

Proof. The first assertion is established by the discussion preceding this proposition.

The second assertion is established similarly to the corresponding part of the proof of Proposition 1.37. Suppose that $(\bar{\lambda}, \bar{\mu})$ is noncritical but the SOSC (1.58) does not hold. Under the assumed condition (1.60), this means that there exists $\xi \in C(\bar{x}) \setminus \{0\}$ satisfying (1.61). Then (1.60) implies that ξ is a solution of problem (1.62), and hence, a stationary point of this problem (due to the linearity of its constraints, and to Theorem 1.14). However, according to the discussion above, such $\xi \neq 0$ does not exist in the case of noncritical $(\bar{\lambda}, \bar{\mu})$, which gives a contradiction. □

The result we present next follows from the analysis in [130] and from [85, Theorem 2]: the assumption that the multiplier is noncritical is equivalent to the Lipschitzian error bound estimating the distance to the solution set of the KKT system, or to the upper Lipschitzian behavior of this solution set under the canonical perturbations. A closely related result under the second-order sufficiency condition (1.79) was also established in [113, Lemma 2]. An extension of Proposition 1.43 to Lipschitz-continuous KKT systems is given in [155].

Proposition 1.43. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable at $\bar{x} \in \mathbf{R}^n$, and let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable at $\bar{x}$. Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ with some $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be a solution of the system (1.39).*

Then the following three properties are equivalent:

(a) The Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$ is noncritical.
(b) The error bound

$$\|x - \bar{x}\| + \mathrm{dist}((\lambda, \mu), \mathcal{M}(\bar{x})) = O\left(\left\| \begin{pmatrix} G(x, \lambda, \mu) \\ h(x) \\ \min\{\mu, -g(x)\} \end{pmatrix} \right\|\right) \tag{1.81}$$

holds as $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ tends to $(\bar{x}, \bar{\lambda}, \bar{\mu})$.
(c) For every $\sigma = (a, b, c) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(0, 0, 0)$, any solution $(x(\sigma), \lambda(\sigma), \mu(\sigma))$ of the canonically perturbed KKT system

$$\begin{gathered} F(x) + (h'(x))^{\mathrm{T}}\lambda + (g'(x))^{\mathrm{T}}\mu = a, \quad h(x) = b, \\ \mu \geq 0, \quad g(x) \leq c, \quad \langle \mu, g(x) - c \rangle = 0, \end{gathered} \tag{1.82}$$

close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$*, satisfies the estimate*

$$\|x(\sigma) - \bar{x}\| + \operatorname{dist}((\lambda(\sigma),\, \mu(\sigma)),\, \mathcal{M}(\bar{x})) = O(\|\sigma\|). \tag{1.83}$$

Proof. In order to demonstrate that (a) implies (b), we first prove the primal estimate

$$x - \bar{x} = O\left(\left\| \begin{pmatrix} G(x,\, \lambda,\, \mu) \\ h(x) \\ \min\{\mu,\, -g(x)\} \end{pmatrix} \right\|\right) \tag{1.84}$$

as $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ tends to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$.

We argue by contradiction. Suppose that (1.84) does not hold. Then there exists a sequence $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ convergent to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, with $x^k \neq \bar{x}$ for all k, and such that

$$\|x^k - \bar{x}\| \left\| \begin{pmatrix} G(x^k,\, \lambda^k,\, \mu^k) \\ h(x^k) \\ \min\{\mu^k,\, -g(x^k)\} \end{pmatrix} \right\|^{-1} \to \infty \text{ as } k \to \infty,$$

which is equivalent to saying that

$$G(x^k,\, \lambda^k,\, \mu^k) = o(\|x^k - \bar{x}\|), \tag{1.85}$$

$$h(x^k) = o(\|x^k - \bar{x}\|), \tag{1.86}$$

$$\min\{\mu^k,\, -g(x^k)\} = o(\|x^k - \bar{x}\|) \tag{1.87}$$

as $k \to \infty$.

By (1.86) we have that

$$\begin{aligned} 0 &= h(x^k) + o(\|x^k - \bar{x}\|) \\ &= h(\bar{x}) + h'(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) \\ &= h'(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|). \end{aligned} \tag{1.88}$$

Moreover, since $g_{A_+(\bar{x},\, \bar{\mu})}(\bar{x}) = 0 < \bar{\mu}_{A_+(\bar{x},\, \bar{\mu})}$, from (1.87) we derive that for all k large enough

$$\begin{aligned} 0 &= \min\{\mu^k_{A_+(\bar{x},\, \bar{\mu})},\, -g_{A_+(\bar{x},\, \bar{\mu})}(x^k)\} + o(\|x^k - \bar{x}\|) \\ &= -g_{A_+(\bar{x},\, \bar{\mu})}(x^k) + o(\|x^k - \bar{x}\|) \\ &= -g_{A_+(\bar{x},\, \bar{\mu})}(\bar{x}) - g'_{A_+(\bar{x},\, \bar{\mu})}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) \\ &= -g'_{A_+(\bar{x},\, \bar{\mu})}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \end{aligned} \tag{1.89}$$

and similarly, since $g_{\{1,\, \dots,\, m\} \setminus A(\bar{x})}(\bar{x}) < 0 = \bar{\mu}_{\{1,\, \dots,\, m\} \setminus A(\bar{x})}$,

$$\begin{aligned} 0 &= \min\{\mu^k_{\{1,\, \dots,\, m\} \setminus A(\bar{x})},\, -g_{\{1,\, \dots,\, m\} \setminus A(\bar{x})}(x^k)\} + o(\|x^k - \bar{x}\|) \\ &= \mu^k_{\{1,\, \dots,\, m\} \setminus A(\bar{x})} + o(\|x^k - \bar{x}\|). \end{aligned} \tag{1.90}$$

As the number of different partitions of the set $A_0(\bar{x},\, \bar{\mu})$ is finite, passing onto a subsequence if necessary, we can assume that there exist index sets I_1 and I_2 such that $I_1 \cup I_2 = A_0(\bar{x},\, \bar{\mu})$, $I_1 \cap I_2 = \emptyset$, and for each k it holds that

$$\mu^k_{I_1} \geq -g_{I_1}(x^k), \quad \mu^k_{I_2} < -g_{I_2}(x^k). \tag{1.91}$$

Similar to the above, from (1.87) we then obtain that

$$\begin{aligned} 0 &= \min\{\mu^k_{I_1},\, -g_{I_1}(x^k)\} + o(\|x^k - \bar{x}\|) \\ &= -g_{I_1}(x^k) + o(\|x^k - \bar{x}\|) \\ &= -g_{I_1}(\bar{x}) - g'_{I_1}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) \\ &= -g'_{I_1}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \end{aligned} \tag{1.92}$$

$$0 = \min\{\mu^k_{I_2},\, -g_{I_2}(x^k)\} + o(\|x^k - \bar{x}\|) = \mu^k_{I_2} + o(\|x^k - \bar{x}\|). \tag{1.93}$$

Moreover, by (1.91) and the second equality in (1.92) we obtain that

$$\mu^k_{I_1} \geq -g_{I_1}(x^k) = o(\|x^k - \bar{x}\|). \tag{1.94}$$

Finally, from (1.91) it also follows that

$$\begin{aligned} -\mu^k_{I_2} &> g_{I_2}(x^k) \\ &= g_{I_2}(\bar{x}) + g'_{I_2}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) \\ &= g'_{I_2}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \end{aligned}$$

and hence, by (1.93),

$$g'_{I_2}(\bar{x})(x^k - \bar{x}) \leq o(\|x^k - \bar{x}\|). \tag{1.95}$$

Furthermore, employing the relations (1.85), (1.90), (1.93), and the equalities $\bar{\mu}_{A_0(\bar{x},\, \bar{\mu})} = 0$ and $\bar{\mu}_{\{1,\, \ldots,\, m\} \setminus A(\bar{x})} = 0$, we derive that

$$\begin{aligned} 0 &= G(x^k,\, \lambda^k,\, \mu^k) + o(\|x^k - \bar{x}\|) \\ &= G(x^k,\, \bar{\lambda},\, \bar{\mu}) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu}) \\ &\quad + o(\|x^k - \bar{x}\|) \\ &= G(x^k,\, \bar{\lambda},\, \bar{\mu}) - G(\bar{x},\, \bar{\lambda},\, \bar{\mu}) \\ &\quad + (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu}) + o(\|x^k - \bar{x}\|) \\ &= \frac{\partial G}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})(x^k - \bar{x}) \\ &\quad + (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu}) + o(\|x^k - \bar{x}\|), \\ &= \frac{\partial G}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})(x^k - \bar{x}) \\ &\quad + (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'_{A_+(\bar{x},\, \bar{\mu})}(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu})_{A_+(\bar{x},\, \bar{\mu})} + (g'_{I_1}(\bar{x}))^{\mathrm{T}}\mu^k_{I_1} \\ &\quad + o(\|x^k - \bar{x}\|), \end{aligned} \tag{1.96}$$

where the third equality is by $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$. Taking into account (1.94), we further derive from (1.96) that

$$-\operatorname{im}(h'(\bar{x}))^{\mathrm{T}} - \operatorname{im}(g'_{A_+(\bar{x},\,\bar{\mu})}(\bar{x}))^{\mathrm{T}} - (g'_{I_1}(\bar{x}))^{\mathrm{T}}\mathbf{R}_+^{|I_1|} \ni \frac{\partial G}{\partial x}(\bar{x},\,\bar{\lambda},\,\bar{\mu})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \quad (1.97)$$

where the set in the left-hand side is a closed cone (as a sum of linear subspaces and a polyhedral cone).

Passing onto a subsequence, if necessary, we can assume that the sequence $\{(x^k - \bar{x})/\|x^k - \bar{x}\|\}$ converges to some $\xi \in \mathbf{R}^n$, $\|\xi\| = 1$. Dividing (1.97) and (1.88), (1.89), (1.92), (1.95) by $\|x^k - \bar{x}\|$, passing onto the limit as $k \to \infty$, we obtain the following:

$$\frac{\partial G}{\partial x}(\bar{x},\,\bar{\lambda},\,\bar{\mu})\xi \in -\operatorname{im}(h'(\bar{x}))^{\mathrm{T}} - \operatorname{im}(g'_{A_+(\bar{x},\,\bar{\mu})}(\bar{x}))^{\mathrm{T}} - (g'_{I_1}(\bar{x}))^{\mathrm{T}}\mathbf{R}_+^{|I_1|}, \quad (1.98)$$

$$h'(\bar{x})\xi = 0, \quad g'_{A_+(\bar{x},\,\bar{\mu})}(\bar{x})\xi = 0, \quad (1.99)$$

$$g'_{I_1}(\bar{x})\xi = 0, \quad g'_{I_2}(\bar{x})\xi \le 0. \quad (1.100)$$

Inclusion (1.98) means that there exists $(\eta,\, \zeta) \in \mathbf{R}^l \times \mathbf{R}^m$ satisfying the first equality in (1.56), and such that

$$\zeta_{I_1} \ge 0, \quad \zeta_{I_2 \cup (\{1,\,\ldots,\,m\} \setminus A(\bar{x}))} = 0.$$

Combining this with (1.99), (1.100), we obtain that $(\xi,\, \eta,\, \zeta)$ satisfies (1.56), which contradicts the assumption that $(\bar{\lambda},\, \bar{\mu})$ is noncritical, because $\xi \neq 0$. This completes the proof of estimate (1.84).

To establish the remaining estimate, i.e.,

$$\operatorname{dist}((\lambda,\, \mu),\, \mathcal{M}(\bar{x})) = O\left(\left\| \begin{pmatrix} G(x,\, \lambda,\, \mu) \\ h(x) \\ \min\{\mu,\, -g(x)\} \end{pmatrix} \right\|\right), \quad (1.101)$$

observe again that

$$\mathcal{M}(\bar{x}) = \left\{ (\lambda,\, \mu) \in \mathbf{R}^l \times \mathbf{R}^m \;\middle|\; \begin{array}{l} (h'(\bar{x}))^{\mathrm{T}}\lambda + (g'_{A(\bar{x})}(x))^{\mathrm{T}}\mu_{A(\bar{x})} = -F(\bar{x}), \\ \mu_{A(\bar{x})} \ge 0,\ \mu_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})} = 0 \end{array} \right\}$$

is a polyhedral set. Applying Hoffman's lemma (see Lemma A.4) to this set, and taking into account (1.40), we obtain the existence of $\gamma > 0$ such that the estimate

$$\operatorname{dist}((\lambda,\, \mu), \mathcal{M}(\bar{x})) \le \gamma \left\| \begin{pmatrix} G(\bar{x},\, \lambda,\, \mu) \\ \min\{0,\, \mu_{A(\bar{x})}\} \\ \mu_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})} \end{pmatrix} \right\| \quad (1.102)$$

holds for all $(\lambda,\, \mu) \in \mathbf{R}^l \times \mathbf{R}^m$.

Evidently,

$$G(\bar{x},\, \lambda,\, \mu) = G(x,\, \lambda,\, \mu) + O(\|x - \bar{x}\|) \tag{1.103}$$

as $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ tends to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$. Note further that if for some $i \in A(\bar{x})$ it holds that $\mu_i < 0$, then

$$|\min\{0,\, \mu_i\}| = -\mu_i \leq \max\{-\mu_i,\, g_i(x)\} = -\min\{\mu_i,\, -g_i(x)\},$$

and hence,

$$\|\min\{0,\, \mu_{A(\bar{x})}\}\| \leq \|\min\{\mu_{A(\bar{x})},\, -g_{A(\bar{x})}(x)\}\|. \tag{1.104}$$

On the other hand, since $g_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})}(\bar{x}) < 0$ and $\bar{\mu}_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})} = 0$, it holds that

$$\mu_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})} = \min\{\mu_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})},\, -g_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})}(x)\} \tag{1.105}$$

for all $(x,\, \mu)$ close enough to $(\bar{x},\, \bar{\mu})$.

Now (1.102) and relations (1.103)–(1.105) yield

$$\operatorname{dist}((\lambda,\, \mu),\, \mathcal{M}(\bar{x})) = O\left(\left\|\begin{pmatrix} G(x,\, \lambda,\, \mu) \\ \min\{\mu,\, -g(x)\} \end{pmatrix}\right\|\right) + O(\|x - \bar{x}\|) \tag{1.106}$$

as $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ tends to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, which further implies the needed (1.101), using also the primal estimate (1.84) already established above.

We next show that (b) implies (c). For any $\sigma = (a,\, b,\, c) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and any solution $(x(\sigma),\, \lambda(\sigma),\, \mu(\sigma)) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ of the canonically perturbed KKT system (1.82) close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, from the error bound (1.81) we immediately obtain that

$$\begin{aligned} &\|x(\sigma) - \bar{x}\| + \operatorname{dist}((\lambda(\sigma),\, \mu(\sigma)),\, \mathcal{M}(\bar{x})) \\ &= O\left(\left\|\begin{pmatrix} G(x(\sigma),\, \lambda(\sigma),\, \mu(\sigma)) \\ h(x(\sigma)) \\ \min\{\mu(\sigma),\, -g(x(\sigma))\} \end{pmatrix}\right\|\right) \\ &= O(\|(a,\, b)\| + \|\min\{\mu(\sigma),\, -g(x(\sigma))\}\|). \end{aligned} \tag{1.107}$$

Using the last line in (1.82), it can be directly checked that

$$|\min\{\mu_i(\sigma),\, -g_i(x(\sigma))\}| \leq |c_i| \quad \forall\, i = 1,\, \ldots,\, m,$$

and hence, (1.107) implies the needed estimate (1.83).

Finally, in order to prove that (c) implies (a), suppose that the latter is violated: there exists a triple $(\xi,\, \eta,\, \zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, with $\xi \neq 0$, satisfying the system (1.56). It can then be checked that for all $t \geq 0$ small enough, the triple $(x(t),\, \lambda(t),\, \mu(t)) = (\bar{x} + t\xi,\, \bar{\lambda} + t\eta,\, \bar{\mu} + t\zeta)$ satisfies (1.82) with some

$\sigma(t) = (a(t),\, b(t),\, c(t)) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ such that $\sigma(t) = o(t)$. However, this contradicts (1.83), since $\|x(t) - \bar{x}\| = t\|\xi\|$, while the right-hand side of (1.83) is $o(t)$. $\square$

For the primal rate of convergence analysis in Chap. 4, we shall also need estimates of the distance to the primal solution, obtained in [77]. We start with some considerations that motivate the kind of primal error bounds that will be derived below, sharpening the estimate (1.106) obtained in the proof of Proposition 1.43.

For each $\mu \in \mathbf{R}^m$, define $\tilde{\mu} \in \mathbf{R}^m$ by

$$\tilde{\mu}_i = \begin{cases} \max\{0,\, \mu_i\} & \text{if } i \in A(\bar{x}), \\ 0 & \text{if } i \in \{1,\, \ldots,\, m\} \setminus A(\bar{x}), \end{cases} \tag{1.108}$$

i.e., $\tilde{\mu}$ is the Euclidean projection of μ onto the set

$$\{\mu \in \mathbf{R}^m \mid \mu_{A(\bar{x})} \geq 0,\ \mu_{\{1,\, \ldots,\, m\} \setminus A(\bar{x})} = 0\}.$$

Then, for any $\lambda \in \mathbf{R}^l$ and any $\xi \in C(\bar{x})$ it evidently holds that

$$\langle G(\bar{x},\, \lambda,\, \tilde{\mu}),\, \xi\rangle = \langle F(\bar{x}),\, \xi\rangle + \langle \lambda,\, h'(\bar{x})\xi\rangle + \langle \tilde{\mu}_{A(\bar{x})},\, g'_{A(\bar{x})}(\bar{x})\xi\rangle \leq 0,$$

which means that $G(\bar{x}, \lambda,\, \tilde{\mu}) \in (C(\bar{x}))^\circ$. Therefore,

$$\begin{aligned} G(\bar{x},\, \lambda,\, \mu) &= G(\bar{x},\, \lambda,\, \tilde{\mu}) + O(\|\mu - \tilde{\mu}\|) \\ &= \pi_{(C(\bar{x}))^\circ}(G(\bar{x},\, \lambda,\, \tilde{\mu})) + O(\|\mu - \tilde{\mu}\|) \\ &= \pi_{(C(\bar{x}))^\circ}(G(x,\, \lambda,\, \mu)) + O(\|x - \bar{x}\|) + O(\|\mu - \tilde{\mu}\|) \\ &= \pi_{(C(\bar{x}))^\circ}(G(x,\, \lambda,\, \mu)) + O(\|x - \bar{x}\|) \\ &\quad + O(\|\min\{0,\, \mu_{A(\bar{x})}\}\| + \|\mu_{\{1,\, \ldots,\, m\} \setminus A(\bar{x})}\|) \end{aligned} \tag{1.109}$$

as $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ tends to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ with some $(\bar{\lambda},\, \bar{\mu}) \in \mathcal{M}(\bar{x})$, where the last equality is by (1.108). Furthermore, as in the proof of Proposition 1.43, estimate (1.102) and relations (1.104) and (1.105) are valid. Combining them with (1.109), we derive the estimate

$$\operatorname{dist}((\lambda,\, \mu), \mathcal{M}(\bar{x})) = O\left(\left\| \begin{pmatrix} \pi_{(C(\bar{x}))^\circ}\,(G(x,\, \lambda,\, \mu)) \\ \min\{\mu,\, -g(x)\} \end{pmatrix} \right\|\right) + O(\|x - \bar{x}\|).$$

Comparing this with (1.81) and taking into account that for any $x \in \mathbf{R}^n$ it holds that $x = \pi_{(C(\bar{x}))^\circ}(x) + \pi_{C(\bar{x})}(x)$ (see Lemma A.13), it is natural to conjecture that in the estimate for $\|x - \bar{x}\|$ one might replace $G(x,\, \lambda,\, \mu)$ by the term $\pi_{C(\bar{x})}(G(x,\, \lambda,\, \mu))$ of generally smaller norm. And this is indeed the case, as will be demonstrated in Proposition 1.46 below.

We shall next derive two different estimates for the distance to the primal solution involving projections. The first one assumes the weaker property of noncriticality of the multiplier but makes projection onto the linear subspace

$$C_+(\bar{x},\, \bar{\mu}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+(\bar{x},\, \bar{\mu})}(\bar{x})\xi = 0\},$$

which is generally larger than $C(\bar{x})$, while the second assumes the stronger second-order sufficiency condition (1.79) but makes projection onto the generally smaller cone $C(\bar{x})$.

Proposition 1.44. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable at $\bar{x} \in \mathbf{R}^n$, and let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable at $\bar{x}$. Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ with some $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be a solution of the system (1.39), and assume that $(\bar{\lambda}, \bar{\mu})$ is a noncritical Lagrange multiplier.*

Then the error bound

$$x - \bar{x} = O\left(\left\|\begin{pmatrix} \pi_{C_+(\bar{x}, \bar{\mu})}\left(G(x, \lambda, \mu)\right) \\ h(x) \\ \min\{\mu, -g(x)\} \end{pmatrix}\right\|\right) \tag{1.110}$$

holds as $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ tends to $(\bar{x}, \bar{\lambda}, \bar{\mu})$.

Proof. We again argue by contradiction. Assuming that (1.110) does not hold, similarly to the proof of Proposition 1.43 we obtain that there exists a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ convergent to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ and satisfying

$$\pi_{C_+(\bar{x}, \bar{\mu})}\left(G(x^k, \lambda^k, \mu^k)\right) = o(\|x^k - \bar{x}\|), \tag{1.111}$$

and (1.86), (1.87). Passing onto a subsequence, if necessary, we can then assume that the sequence $\{(x^k - \bar{x})/\|x^k - \bar{x}\|\}$ converges to some element $\xi \in \mathbf{R}^n$, $\|\xi\| = 1$, and that there further exist index sets I_1 and I_2 such that $I_1 \cup I_2 = A_0(\bar{x}, \bar{\mu})$, $I_1 \cap I_2 = \emptyset$, and that relations (1.90), (1.93), (1.94), (1.99), (1.100) are satisfied.

Furthermore, employing Lemma A.13 and (1.111), we derive that

$$\begin{aligned} 0 &= \pi_{C_+(\bar{x}, \bar{\mu})}(G(x^k, \lambda^k, \mu^k) - \pi_{C_+(\bar{x}, \bar{\mu})}(G(x^k, \lambda^k, \mu^k))) \\ &= \pi_{C_+(\bar{x}, \bar{\mu})}(G(x^k, \lambda^k, \mu^k) + o(\|x^k - \bar{x}\|)). \end{aligned}$$

The latter, by Lemma A.13, implies that

$$\begin{aligned} (C_+(\bar{x}, \bar{\mu}))^\perp &= (C_+(\bar{x}, \bar{\mu}))^\circ \\ &\ni G(x^k, \lambda^k, \mu^k) + o(\|x^k - \bar{x}\|) \\ &= G(x^k, \bar{\lambda}, \bar{\mu}) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu}) \\ &\quad + o(\|x^k - \bar{x}\|) \\ &= G(x^k, \bar{\lambda}, \bar{\mu}) - G(\bar{x}, \bar{\lambda}, \bar{\mu}) \\ &\quad + (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu}) + o(\|x^k - \bar{x}\|) \\ &= \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})(x^k - \bar{x}) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) \\ &\quad + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu}) + o(\|x^k - \bar{x}\|), \end{aligned} \tag{1.112}$$

where the first equality is by the fact that $C_+(\bar{x}, \bar{\mu})$ is a linear subspace, and the third is by $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$. Since

$$
\begin{aligned}
(C_+(\bar{x}, \bar{\mu}))^\perp &= (\ker h'(\bar{x}) \cap \ker g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x}))^\perp \\
&= \mathrm{im}(h'(\bar{x}))^{\mathrm{T}} + \mathrm{im}(g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x}))^{\mathrm{T}},
\end{aligned}
$$

taking into account that $\bar{\mu}_{A_0(\bar{x}, \bar{\mu})} = 0$ and $\bar{\mu}_{\{1, \dots, m\} \setminus A(\bar{x})} = 0$, (1.112) implies the inclusion

$$
\begin{aligned}
&- \mathrm{im}(h'(\bar{x}))^{\mathrm{T}} - \mathrm{im}(g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x}))^{\mathrm{T}} \\
&\ni \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})(x^k - \bar{x}) + (g'_{\{1, \dots, m\} \setminus A_+(\bar{x}, \bar{\mu})}(\bar{x}))^{\mathrm{T}} \mu^k_{\{1, \dots, m\} \setminus A_+(\bar{x}, \bar{\mu})} \\
&+ o(\|x^k - \bar{x}\|).
\end{aligned}
\tag{1.113}
$$

Without loss of generality, again passing onto a subsequence if necessary, we can assume that there exist index sets J_1 and J_2 such that $J_1 \cup J_2 = I_1$, $J_1 \cap J_2 = \emptyset$, and for each k it holds that

$$
\mu^k_{J_1} \geq 0, \quad \mu^k_{J_2} < 0. \tag{1.114}
$$

Then from (1.113) we derive the estimate

$$
\begin{aligned}
&- \mathrm{im}(h'(\bar{x}))^{\mathrm{T}} - \mathrm{im}(g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x}))^{\mathrm{T}} - (g'_{J_1}(\bar{x}))^{\mathrm{T}} \mathbf{R}_+^{|J_1|} \\
&\ni \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})(x^k - \bar{x}) \\
&+ (g'_{J_2 \cup I_2 \cup (\{1, \dots, m\} \setminus A(\bar{x}))}(\bar{x}))^{\mathrm{T}} \mu^k_{J_2 \cup I_2 \cup (\{1, \dots, m\} \setminus A(\bar{x}))} \\
&+ o(\|x^k - \bar{x}\|),
\end{aligned}
\tag{1.115}
$$

where the set in the left-hand side is a closed cone. Furthermore, (1.94) and (1.114) imply the estimate

$$
\mu^k_{J_2} = o(\|x^k - \bar{x}\|). \tag{1.116}
$$

Dividing (1.115) by $\|x^k - \bar{x}\|$, passing onto the limit as $k \to \infty$, and employing the relations (1.90), (1.93), and (1.116), we obtain the following:

$$
\frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi \in - \mathrm{im}(h'(\bar{x}))^{\mathrm{T}} - \mathrm{im}(g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x}))^{\mathrm{T}} - (g'_{J_1}(\bar{x}))^{\mathrm{T}} \mathbf{R}_+^{|J_1|}.
$$

This inclusion means that there exists $(\eta, \zeta) \in \mathbf{R}^l \times \mathbf{R}^m$ satisfying the first equality in (1.56), and such that

$$
\zeta_{J_1} \geq 0, \quad \zeta_{J_2 \cup I_2 \cup (\{1, \dots, m\} \setminus A(\bar{x}))} = 0.
$$

Combining this with (1.99), (1.100), we obtain that (ξ, η, ζ) satisfies (1.56), which contradicts the assumption that $(\bar{\lambda}, \bar{\mu})$ is noncritical, because $\xi \neq 0$. $\square$

We now provide an example demonstrating that, under the noncriticality assumption, in the estimate (1.110) one cannot replace $C_+(\bar{x},\, \bar{\mu})$ by the (generally smaller) cone $C(\bar{x})$, even in the case of optimization and even when $\bar{x}$ satisfies the LICQ.

Example 1.45. Let $n = m = 1$, $l = 0$, $f(x) = -x^2/2$, $g(x) = x$, and let F and G be defined by (1.45) and (1.46), respectively. Then $\bar{x} = 0$ is a stationary point of problem (1.44) satisfying the LICQ, and $\bar{\mu} = 0$ is the unique associated multiplier. Note that $C(\bar{x}) = \{\xi \in \mathbf{R} \mid \xi \leq 0\}$ and $C_+(\bar{x},\, \bar{\mu}) = \mathbf{R}$. The multiplier $\bar{\mu} = 0$ is evidently noncritical ((1.80) holds).

Take any $x < 0$ and $\mu = 0$. Then

$$\frac{\partial L}{\partial x}(x,\, \mu) = -x + \mu = -x > 0,$$

and the projection of this element onto $C(\bar{x})$ is zero. Also,

$$\min\{\mu, -x\} = \mu = 0.$$

Hence, if $C_+(\bar{x},\, \bar{\mu})$ were to be replaced by $C(\bar{x})$ in (1.110), the "residual" in the right-hand side would be equal to zero at points in question, failing to provide the needed estimate. At the same time, (1.110) with $C_+(\bar{x},\, \bar{\mu})$ is evidently valid.

We next show that if noncriticality of the multiplier is replaced by the stronger second-order sufficiency condition (1.79), then $C_+(\bar{x},\, \bar{\mu})$ in (1.110) can be replaced by $C(\bar{x})$.

Proposition 1.46. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable at $\bar{x} \in \mathbf{R}^n$, and let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable at $\bar{x}$. Let $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ with some $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be a solution of the system (1.39), and assume that the second-order sufficiency condition (1.79) holds.*

Then the error bound

$$x - \bar{x} = O\left(\left\| \begin{pmatrix} \pi_{C(\bar{x})}\left(G(x,\, \lambda,\, \mu)\right) \\ h(x) \\ \min\{\mu,\, -g(x)\} \end{pmatrix} \right\|\right) \tag{1.117}$$

holds as $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ tends to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$.

Proof. The argument is along the lines of the proof of Proposition 1.44. Assuming that (1.117) does not hold, similarly to the proof of Proposition 1.43 we obtain that there exists a sequence $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ convergent to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ and satisfying

$$\pi_{C(\bar{x})}\left(G(x^k,\, \lambda^k,\, \mu^k)\right) = o(\|x^k - \bar{x}\|) \tag{1.118}$$

and (1.86), (1.87) as $k \to \infty$. Passing onto a subsequence, if necessary, we can then assume that the sequence $\{(x^k - \bar{x})/\|x^k - \bar{x}\|\}$ converges to some element $\xi \in \mathbf{R}^n$, $\|\xi\| = 1$, and that there exist index sets I_1 and I_2 such that $I_1 \cup I_2 = A_0(\bar{x}, \bar{\mu})$, $I_1 \cap I_2 = \emptyset$, and that relations (1.90), (1.93), (1.94), (1.99), (1.100) are satisfied.

Using the same argument as the one leading to (1.112), and employing Lemma A.13, we derive from (1.118) that

$$\begin{aligned}(C(\bar{x}))^\circ \ni\ & G(x^k, \lambda^k, \mu^k) + o(\|x^k - \bar{x}\|)\\ &= \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})(x^k - \bar{x}) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu})\\ &\quad + o(\|x^k - \bar{x}\|). \end{aligned} \tag{1.119}$$

Note that, according to (1.99), (1.100), it holds that $\xi \in C(\bar{x})$. Therefore, taking into account (1.90), (1.93), (1.99), the first relation in (1.100), and the equality $\bar{\mu}_{\{1,\,\ldots,\,m\}\setminus A_+(\bar{x},\,\bar{\mu})} = 0$, from (1.119) we derive that

$$\begin{aligned} 0 &\geq \left\langle \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})(x^k - \bar{x}),\, \xi \right\rangle + \langle \lambda^k - \bar{\lambda},\, h'(\bar{x})\xi\rangle + \langle \mu^k - \bar{\mu},\, g'(\bar{x})\xi\rangle\\ &\quad + o(\|x^k - \bar{x}\|)\\ &= \left\langle \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})(x^k - \bar{x}),\, \xi \right\rangle\\ &\quad + \langle \mu^k_{I_2 \cup (\{1,\,\ldots,\,m\}\setminus A(\bar{x}))},\, g'_{I_2 \cup (\{1,\,\ldots,\,m\}\setminus A(\bar{x}))}(\bar{x})\xi\rangle + o(\|x^k - \bar{x}\|)\\ &= \left\langle \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})(x^k - \bar{x}),\, \xi \right\rangle + o(\|x^k - \bar{x}\|). \end{aligned}$$

Dividing the latter relation by $\|x^k - \bar{x}\|$ and passing onto the limit, we conclude that

$$\left\langle \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi,\, \xi \right\rangle \leq 0,$$

which contradicts the second-order sufficiency condition (1.79), because of $\xi \in C(\bar{x}) \setminus \{0\}$. $\square$

Note that unlike the primal-dual error bound (1.81), estimates (1.110) and (1.117) measure the distance to the primal solution $\bar{x}$ only; the distance to the multiplier set $\mathcal{M}(\bar{x})$ is not estimated. This is the price paid for deriving sharper error bounds for the primal part, with generally smaller right-hand sides. Note also that unlike (1.81), the right-hand sides in the estimates (1.110) and (1.117) are not computable, as they involve $C_+(\bar{x}, \bar{\mu})$ and $C(\bar{x})$ which are defined at the unknown solution. Thus the purpose for deriving the last two estimates is for convergence analysis only, with no intended use within algorithms themselves.

Remark 1.47. Under the "opposite-sign" second-order condition (1.80), we can obtain the following estimate:

$$x - \bar{x} = O\left(\left\| \begin{pmatrix} \pi_{C(\bar{x})}\left(-G(x,\, \lambda,\, \mu)\right) \\ h(x) \\ \min\{\mu,\, -g(x)\} \end{pmatrix} \right\|\right). \tag{1.120}$$

The proof is analogous to that of Proposition 1.46, noting that for this case in (1.119) we have that

$$(C(\bar{x}))^{\circ} \ni -G(x^k,\, \lambda^k,\, \mu^k) + o(\|x^k - \bar{x}\|).$$

1.4 Nonsmooth Analysis

In this section we provide a summary of facts from nonsmooth analysis that will be employed in the sequel. For a detailed exposition of some of those issues, see, e.g., [44, 68, 169, 201, 239]. That said, none of these references seems to cover all the material required for our purposes. Some additional references and explanations will be provided below where needed.

1.4.1 Generalized Differentiation

We first recall some terminology concerned with the notion of Lipschitz continuity, already used in Sect. 1.3. A mapping $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ is refereed to as *Lipschitz-continuous* on a set $X \subset \mathbf{R}^n$ with a constant $L > 0$ if

$$\|\Phi(x^1) - \Phi(x^2)\| \leq L\|x^1 - x^2\| \quad \forall\, x^1,\, x^2 \in X.$$

We say that Φ is *locally Lipschitz-continuous* at $x \in \mathbf{R}^n$ with a constant $L > 0$ if there exists a neighborhood U of x such that Φ is Lipschitz-continuous on U with this constant. Furthermore, Φ is *locally Lipschitz-continuous on an open set* $O \subset \mathbf{R}^n$ if it is locally Lipschitz-continuous at any point of this set. Finally, Φ is *locally Lipschitz-continuous with respect to* x with a constant $L > 0$ if there exist a neighborhood V of 0 such that

$$\|\Phi(x + \xi) - \Phi(x)\| \leq L\|\xi\| \quad \forall\, \xi \in V.$$

We proceed with the celebrated Rademacher Theorem [227], which states that a locally Lipschitz-continuous mapping is differentiable almost everywhere (for a modern proof see [65]). Let $\mathcal{S}_{\Phi}$ stand for the set of all points where Φ is differentiable.

Theorem 1.48. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ be locally Lipschitz-continuous on an open set $O \subset \mathbf{R}^n$.*

Then the Lebesgue measure of the set $O \setminus \mathcal{S}_\Phi$ is zero, and in particular, the set $\mathcal{S}_\Phi$ is everywhere dense in O, i.e., $O \subset \operatorname{cl} \mathcal{S}_\Phi$.

For our purposes, the significance of the Rademacher Theorem is in the fact that it naturally gives rise to the following notion of generalized differentiation. By the *B-differential*[1] of $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ at $x \in \mathbf{R}^n$ we mean the set

$$\partial_B \Phi(x) = \{J \in \mathbf{R}^{m \times n} \mid \exists\, \{x^k\} \subset \mathcal{S}_\Phi : \{x^k\} \to x,\ \{\Phi'(x^k)\} \to J\}.$$

Then Clarke's *generalized Jacobian* of Φ at $x \in \mathbf{R}^n$ is the set

$$\partial \Phi(x) = \operatorname{conv} \partial_B \Phi(x).$$

If Φ is differentiable near x, and its derivative is continuous at x, then evidently $\partial \Phi(x) = \partial_B \Phi(x) = \{\Phi'(x)\}$. Otherwise, $\partial_B \Phi(x)$ (and thus also $\partial \Phi(x)$) is not necessarily a singleton, even if Φ is differentiable at x; in the latter case, it always holds that $\Phi'(x) \in \partial_B \Phi(x)$. The following illustration is taken from [44, Example 2.2.3].

Example 1.49. Consider the function $\varphi : \mathbf{R} \to \mathbf{R}$,

$$\varphi(x) = \begin{cases} x^2 \sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

This function is everywhere differentiable, but its derivative

$$\varphi'(x) = \begin{cases} 2x \sin(1/x) - \cos(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

is discontinuous at $x = 0$. It is evident that $\partial \varphi(0) = \partial_B \varphi(0) = [-1, 1]$.

The following facts can be easily derived from the definition of differentiability.

Lemma 1.50. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable at $x \in \mathbf{R}^n$.*

Then the following statements hold:

(a) Φ is locally Lipschitz-continuous with respect to $x \in \mathbf{R}^n$ with any constant $L > \|\Phi'(x)\|$.
(b) If Φ is locally Lipschitz-continuous with respect to $x \in \mathbf{R}^n$ with a constant $L > 0$, then $\|\Phi'(x)\| \leq L$.

By Theorem 1.48, if Φ is locally Lipschitz-continuous at $x \in \mathbf{R}^n$ with a constant $L > 0$, then there exists a sequence $\{x^k\} \subset \mathcal{S}_\Phi$ convergent to x.

[1] B is in honor of G. Bouligand.

Therefore, for all k large enough, Φ is locally Lipschitz-continuous at x^k with the same constant $L > 0$, and by Lemma 1.50 it holds that $\|\Phi'(x^k)\| \leq L$. This easily implies the following fact (see [44, Proposition 2.6.2]).

Proposition 1.51. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ be locally Lipschitz-continuous at a point $x \in \mathbf{R}^n$ with a constant $L > 0$.*

Then $\partial_B \Phi(x)$ and $\partial \Phi(x)$ are nonempty compact sets, and

$$\|J\| \leq L \quad \forall\, J \in \partial \Phi(x).$$

Moreover, for any $\varepsilon > 0$ there exists a neighborhood U of 0 such that

$$\mathrm{dist}(J,\, \partial_B \Phi(x)) < \varepsilon \quad \forall\, J \in \partial_B \Phi(x + \xi),\ \forall\, \xi \in U,$$

$$\mathrm{dist}(J,\, \partial \Phi(x)) < \varepsilon \quad \forall\, J \in \partial \Phi(x + \xi),\ \forall\, \xi \in U.$$

Example 1.52. Consider $\varphi : \mathbf{R}^n \to \mathbf{R}$, $\varphi(x) = \|x\|$. This function is Lipschitz-continuous on $\mathbf{R}^n$ with the constant $L = 1$. Moreover, it is continuously differentiable on $\mathbf{R}^n \backslash \{0\}$, and $\varphi'(x) = x/\|x\|$ for any x in this set. It is evident that $\partial_B \varphi(0) = \{x \in \mathbf{R}^n \mid \|x\| = 1\}$, and hence, $\partial \varphi(0) = \{x \in \mathbf{R}^n \mid \|x\| \leq 1\}$.

For the function $\varphi(x) = \|x\|_1$, it can be seen that it is Lipschitz-continuous on $\mathbf{R}^n$, and $\partial \varphi(0) = \{x \in \mathbf{R}^n \mid \|x\|_\infty \leq 1\}$.

The following fact was established in [44, Propositions 2.2.6, 2.2.7].

Proposition 1.53. *Let $\varphi : \mathbf{R}^n \to \mathbf{R}$ be convex on $\mathbf{R}^n$.*

Then φ is locally Lipschitz-continuous at any $x \in \mathbf{R}^n$, and $\partial \varphi(x)$ coincides with the subdifferential of φ (see Sect. A.3).

Of course, any notion of generalized differentiability can be practically useful only when the corresponding generalized differential (or at least some of its elements) can actually be computed in cases of interest. Fortunately, for B-differentials and generalized Jacobians this can be done in many cases, using the structure of Φ. Important examples of this kind will be considered in Sect. 3.2.1. Here, we proceed with some calculus rules for these objects.

To begin with, note that

$$\partial_B (t\Phi)(x) = t \partial_B \Phi(x) \quad \forall\, t \in \mathbf{R},$$

and hence,

$$\partial (t\Phi)(x) = t \partial \Phi(x) \quad \forall\, t \in \mathbf{R}. \tag{1.121}$$

The next two propositions easily follow from the definitions of B-differential and of the generalized Jacobian.

Proposition 1.54. *For any $\Phi_1 : \mathbf{R}^n \to \mathbf{R}^{m_1}$ and $\Phi_2 : \mathbf{R}^n \to \times \mathbf{R}^{m_2}$, for the mapping $\Phi(\cdot) = (\Phi_1(\cdot),\, \Phi_2(\cdot))$ it holds that*

$$\partial_B \Phi(x) \subset \partial_B \Phi_1(x) \times \partial_B \Phi_2(x) \quad \forall\, x \in \mathbf{R}^n, \tag{1.122}$$

$$\partial\Phi(x) \subset \partial\Phi_1(x) \times \partial\Phi_2(x) \quad \forall\, x \in \mathbf{R}^n. \tag{1.123}$$

Moreover, if Φ_1 is differentiable near $x \in \mathbf{R}^n$, and its derivative is continuous at x, then both inclusions hold as equalities:

$$\partial_B\Phi(x) = \{\Phi_1'(x)\} \times \partial_B\Phi_2(x),$$

$$\partial\Phi(x) = \{\Phi_1'(x)\} \times \partial\Phi_2(x).$$

Generally, the inclusions in (1.122) and (1.123) can be strict (see, e.g., [68, Example 7.1.15]).

Proposition 1.55. *Let $\varphi_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, m$, be continuous at a point $x \in \mathbf{R}^n$, and let φ_i, $i \in I(x)$, be differentiable near x, with their derivatives being continuous at x, where*

$$I(x) = \{i = 1, \ldots, m \mid \varphi_i(x) = \varphi(x)\}.$$

Then for the function

$$\varphi(\cdot) = \max_{i=1,\ldots,m} \varphi_i(\cdot)$$

it holds that

$$\partial_B\varphi(x) \subset \{\varphi_i'(x) \mid i \in I(x)\},$$

$$\partial\varphi(x) = \mathrm{conv}\{\varphi_i'(x) \mid i \in I(x)\},$$

and the first relation holds as equality provided that for each $i \in I(x)$ there exists a sequence $\{x^{i,k}\} \subset \mathbf{R}^n$ convergent to x and such that $I(x^{i,k}) = \{i\}$ for all k.

The mean-value theorem we present next is established in [119] (cf. Theorem A.10, (a)).

Theorem 1.56. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ be Lipschitz-continuous on an open convex set $O \subset \mathbf{R}^n$.*

Then for any x^1, $x^2 \in O$ there exist $t_i \in (0, 1)$, $J_i \in \partial\Phi(t_ix^1 + (1 - t_i)x^2)$ and $\theta_i \geq 0$, $i = 1, \ldots, m$, such that $\sum_{i=1}^m \theta_i = 1$ and

$$\Phi(x^1) - \Phi(x^2) = \sum_{i=1}^m \theta_i J_i(x^1 - x^2). \tag{1.124}$$

This theorem is a subtler version of the well-known result given in [44, Proposition 2.6.5], which can be proven relying on the improved variant of Carathéodory's theorem [64, Theorem 18, (ii)]. The latter allows to limit by m (rather than the more standard $m+1$) the number of terms in the convex combination in the right-hand side of (1.124), employing the fact that the set $\cup_{t\in(0,1)}\partial\Phi(tx^1 + (1-t)x^2)$ is necessarily connected (the latter can be derived from Proposition 1.51).

Combining Theorem 1.56 with Proposition 1.51, we obtain the following useful property.

Corollary 1.57. *If $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ is locally Lipschitz-continuous at $x \in \mathbf{R}^n$, then the estimate*

$$\min_{J \in \partial\Phi(x)} \|\Phi(x^1) - \Phi(x^2) - J(x^1 - x^2)\| = o(\|x^1 - x^2\|)$$

holds as x^1, $x^2 \in \mathbf{R}^n$ tend to x.

For the following chain rule theorem, we refer to [44, Proposition 2.6.6] and [68, Proposition 7.1.11].

Theorem 1.58. *Let $\Psi : \mathbf{R}^n \to \mathbf{R}^m$ be locally Lipschitz-continuous at a point $x \in \mathbf{R}^n$, and let $\psi : \mathbf{R}^m \to \mathbf{R}$ be locally Lipschitz-continuous at $\Psi(x)$.*

Then the composition function

$$\varphi(\cdot) = \psi(\Psi(\cdot))$$

is locally Lipschitz-continuous at x, and

$$\partial\varphi(x) \subset \operatorname{conv}\{J^{\mathrm{T}} g \mid J \in \partial\Psi(x),\, g \in \partial\psi(\Psi(x))\}.$$

In particular, if ψ is differentiable near $\Psi(x)$, and its derivative is continuous at $\Psi(x)$, then

$$\partial\varphi(x) = \{J^{\mathrm{T}}\psi'(\Psi(x)) \mid J \in \partial\Psi(x)\}.$$

Alternatively, if ψ is convex on $\mathbf{R}^n$, Ψ is differentiable near x, and its derivative is continuous at x, then

$$\partial\varphi(x) = \{(\Psi'(x))^{\mathrm{T}} g \mid g \in \partial\psi(\Psi(x))\}.$$

Hence, the following is valid.

Corollary 1.59. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ be locally Lipschitz-continuous at a point $x \in \mathbf{R}^n$.*

Then

$$\partial\Phi_i(x) = \{J_i \mid J \in \partial\Phi(x)\} \quad \forall\, i = 1,\, \ldots,\, m.$$

The following important fact highlighting the intrinsic nature of Clarke's generalized Jacobian was established in [66]; it says that this concept of generalized differentiation is "blind" to sets of the Lebesgue measure zero.

Proposition 1.60. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ be locally Lipschitz-continuous on an open set $O \subset \mathbf{R}^n$, and let $\mathcal{N} \subset O$ be any set of the Lebesgue measure zero.*

Then for any $x \in O$, for the set defined by

$$\partial^{\mathcal{N}} \Phi(x) = \operatorname{conv} \partial_B^{\mathcal{N}} \Phi(x),$$

where

$$\partial_B^{\mathcal{N}} \Phi(x) = \{J \in \mathbf{R}^{m \times n} \mid \exists \{x^k\} \subset \mathcal{S}_\Phi \setminus \mathcal{N} : \{x^k\} \to x, \{\Phi'(x^k)\} \to J\},$$

it holds that

$$\partial^{\mathcal{N}} \Phi(x) = \partial \Phi(x).$$

Corollary 1.61. *Let $\Phi_1, \Phi_2 : \mathbf{R}^n \to \mathbf{R}^m$ be locally Lipschitz-continuous at $x \in \mathbf{R}^n$.*

Then

$$\partial(\Phi_1 + \Phi_2)(x) \subset \partial \Phi_1(x) + \partial \Phi_2(x),$$

and the inclusion holds as equality if $m = 1$ and the functions Φ_1 and Φ_2 are convex on $\mathbf{R}^n$.

Moreover, if $\partial \Phi_1(x)$ consists of a single matrix J, then the inclusion also holds as equality:

$$\partial(\Phi_1 + \Phi_2)(x) = J + \partial \Phi_2(x).$$

We note, in passing, that according to Corollary 1.57, if $\partial \Phi(x)$ consists of a single matrix J, then

$$\Phi(x^1) - \Phi(x^2) - J(x^1 - x^2) = o(\|x^1 - x^2\|)$$

as $x^1, x^2 \in \mathbf{R}^n$ tend to x. The latter property is often referred to as *strict differentiability* of Φ at x; it implies differentiability at x and the equality $\Phi'(\bar{x}) = J$.

The next example demonstrates that unlike the generalized Jacobian, the B-differential is not "blind" to sets of the Lebesgue measure zero.

Example 1.62. Define the set $\mathcal{R} = \{\pm 1/k \mid k = 1, 2, \ldots\}$ and the function $\varphi : \mathbf{R} \to \mathbf{R}$, $\varphi(x) = \operatorname{dist}(x, \mathcal{R})$ (see Fig. 1.1). This function is Lipschitz-continuous on the entire $\mathbf{R}$ with the constant $L = 1$, and $\mathcal{S}_\varphi$ consists of the zero point, where $\varphi'(0) = 0$, and of $(-\infty, 1)$, $(1, +\infty)$ and the intervals $((-1/k, -(2k+1)/(2k(k+1)))$, $(-(2k+1)/(2k(k+1)), -1/(k+1))$, $(1/(k+1), (2k+1)/(2k(k+1)))$, $((2k+1)/(2k(k+1)), 1/k))$, $k = 1, 2, \ldots$, and for any x in any of these sets it holds that $\varphi'(x) = -1$ or $\varphi'(x) = 1$. Therefore, $\partial_B \varphi(0) = \{-1, 0, 1\}$. However, taking $\mathcal{N} = \{0\}$, we obtain that $\partial_B^{\mathcal{N}} \varphi(0) = \{-1, 1\} \neq \partial_B \varphi(0)$.

Some other concepts of generalized differentiation are based on the usual notion of the *directional derivative* of $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ at $x \in \mathbf{R}^n$ in a direction $\xi \in \mathbf{R}^n$, which is defined by

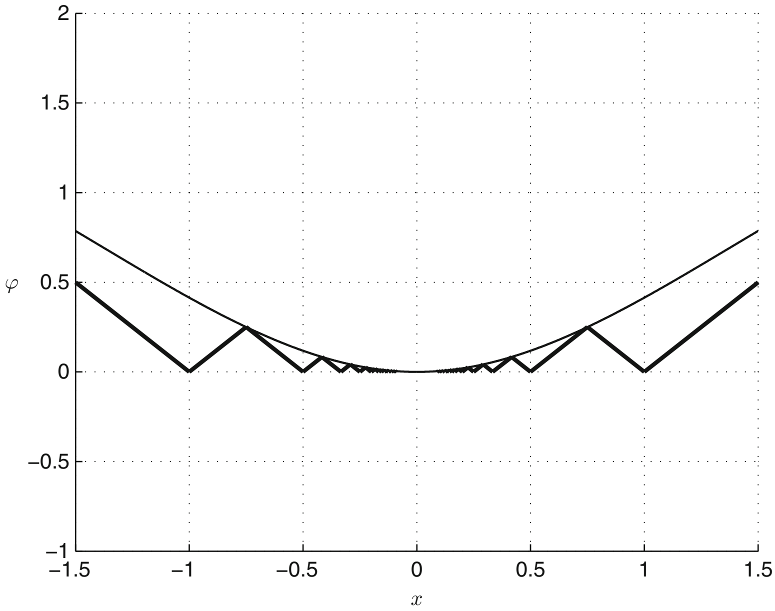

Fig. 1.1 Function from Example 1.62

$$\Phi'(x;\, \xi) = \lim_{t\to 0+} \frac{1}{t}(\Phi(x+t\xi) - \Phi(x)). \tag{1.125}$$

A mapping Φ is called *directionally differentiable* at x in a direction ξ if the limit in (1.125) exists and is finite. Obviously, if Φ is differentiable at x, then it is directionally differentiable at this point in every direction, and $\Phi'(x;\, \xi) = \Phi'(x)\xi$ for all $\xi \in \mathbf{R}^n$.

Proposition 1.63. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ be locally Lipschitz-continuous at $x \in \mathbf{R}^n$ with a constant $L > 0$ and directionally differentiable at x in every direction.*

Then the following statements hold:

(a) The positively homogeneous mapping $\Phi'(x;\, \cdot)$ is Lipschitz-continuous on $\mathbf{R}^n$ with a constant L.
(b) For each $\xi \in \mathbf{R}^n$ there exists a matrix $J \in \partial\Phi(x)$ such that $\Phi'(x;\, \xi) = J\xi$.
(c) As $\xi \in \mathbf{R}^n$ tends to 0, it holds that

$$\Phi(x+\xi) - \Phi(x) - \Phi'(x;\, \xi) = o(\|\xi\|).$$

Assertions (a) and (b) of this proposition were established in [225] (see also [68, Proposition 7.1.17]), while assertion (c) is due to [244] (see also [68, Proposition 3.1.3]). All these properties are easy to check; assertion (b) can be derived from Corollary 1.57.

We proceed with the error bound result generalizing Proposition 1.32 to the nonsmooth case.

Proposition 1.64. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ be directionally differentiable at a point $\bar{x} \in \mathbf{R}^n$ in every direction, and let $\bar{x}$ be a solution of the equation*

$$\Phi(x) = 0. \tag{1.126}$$

Then the condition

$$\{\xi \in \mathbf{R}^n \mid \Phi'(\bar{x};\, \xi) = 0\} = \{0\} \tag{1.127}$$

is necessary for the error bound

$$x - \bar{x} = O(\|\Phi(x)\|) \tag{1.128}$$

to hold as $x \in \mathbf{R}^n$ tends to $\bar{x}$. Moreover, if Φ is locally Lipschitz-continuous at $\bar{x}$, then (1.127) is also sufficient for (1.128), and in particular, (1.127) implies that $\bar{x}$ is an isolated solution of the equation (1.126).

Proof. In order to prove the necessity, consider any $\xi \in \mathbf{R}^n$ such that $\Phi'(\bar{x};\, \xi) = 0$. From (1.128) it follows that

$$\begin{aligned} t\|\xi\| &= \|\bar{x} + t\xi - \bar{x}\| \\ &= O(\|\Phi(\bar{x} + t\xi)\|) \\ &= O(\|\Phi(\bar{x} + t\xi) - \Phi(\bar{x})\|) \\ &= O(t\|\Phi'(\bar{x};\, \xi)\|) + o(t) \\ &= o(t) \end{aligned}$$

as $t \to 0+$, which is possible only when $\xi = 0$.

To prove sufficiency under local Lipschitz-continuity of Φ at $\bar{x}$, suppose that there exists a sequence $\{x^k\} \subset \mathbf{R}^n$ convergent to $\bar{x}$, with $x^k \neq \bar{x}\ \forall\, k$, and such that

$$\frac{\|\Phi(x^k)\|}{\|x^k - \bar{x}\|} \to 0 \text{ as } k \to \infty. \tag{1.129}$$

For each k set $\xi^k = (x^k - \bar{x})/\|x^k - \bar{x}\|$, $t_k = \|x^k - \bar{x}\|$. Passing to a subsequence, if necessary, we can assume that the sequence $\{\xi^k\}$ converges to some $\xi \in \mathbf{R}^n$ with $\|\xi\| = 1$. Then, taking into account (1.129), for all k it holds that

$$\begin{aligned} \frac{\|\Phi(\bar{x} + t_k\xi) - \Phi(\bar{x})\|}{t_k} &\leq \frac{\|\Phi(\bar{x} + t_k\xi^k)\|}{t_k} + \frac{\|\Phi(\bar{x} + t_k\xi) - \Phi(\bar{x} + t_k\xi^k)\|}{t_k} \\ &= \frac{\|\Phi(x^k)\|}{\|x^k - \bar{x}\|} + O(\|\xi^k - \xi\|) \to 0 \text{ as } k \to \infty, \end{aligned}$$

implying the equality $\Phi'(\bar{x};\, \xi) = 0$, which contradicts (1.127). $\square$

Remark 1.65. In Sect. 2.4, the following two regularity conditions will be invoked to establish local convergence of semismooth Newton methods: $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ is referred to as *CD-regular* [224] (*BD-regular* [213]) at $\bar{x} \in \mathbf{R}^n$ if each matrix $J \in \partial\Phi(\bar{x})$ ($J \in \partial_B\Phi(\bar{x})$) is nonsingular. If Φ is locally Lipschitz-continuous at $\bar{x}$ and directionally differentiable at $\bar{x}$ in every direction, then from the second assertion of Proposition 1.63 it readily follows that CD-regularity of Φ at $\bar{x}$ implies (1.127), which further implies the error bound (1.128). It will be demonstrated at the end of the next section that this assertion remains valid with CD-regularity replaced by BD-regularity, provided Lipschitz-continuity and directional differentiability are complemented by an appropriate additional assumption.

Our next result is the inverse function theorem established in [43] (see also [44, Theorem 7.1.1]).

Theorem 1.66. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be locally Lipschitz-continuous at $\bar{x} \in \mathbf{R}^n$, and let $\bar{x}$ be a solution of the equation* (1.126)*. Assume that Φ is CD-regular at $\bar{x}$.*

Then there exist neighborhoods U of $\bar{x}$ and V of 0 such that for all $y \in V$ there exists the unique $x(y) \in U$ such that $\Phi(x(y)) = y$, and the mapping $x(\cdot)$ is locally Lipschitz-continuous at 0.

The proof of the following implicit function theorem associated with Theorem 1.66 can be found in [44, Corollary of Theorem 7.1.1]. Observe also that both Theorems 1.66 and 1.67, as well as Theorem 1.24, are particular cases of the more general results that will be proven in Sect. 3.3.

Theorem 1.67. *Let $\Phi : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^n$ be locally Lipschitz-continuous at $(\bar{\sigma}, \bar{x}) \in \mathbf{R}^s \times \mathbf{R}^n$, and let $\Phi(\bar{\sigma}, \bar{x}) = 0$. Assume that all matrices $J \in \mathbf{R}^{n\times n}$ such that there exists $S \in \mathbf{R}^{n\times s}$ satisfying $(S \;\; J) \in \partial\Phi(\bar{\sigma}, \bar{x})$ are nonsingular.*

Then there exist neighborhoods $\mathcal{U}$ of $\bar{\sigma}$ and U of $\bar{x}$ such that for all $\sigma \in \mathcal{U}$ there exists the unique $x(\sigma) \in U$ such that $\Phi(\sigma, x(y)) = 0$, and the mapping $x(\cdot)$ is locally Lipschitz-continuous at $\bar{\sigma}$.

We complete this section by the following terminology which will be needed below. For a mapping $\Phi : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}^m$, by the *partial B-differential* (*partial generalized Jacobian*) *of Φ with respect to x at* $(x, y) \in \mathbf{R}^n \times \mathbf{R}^l$ we mean the B-differential (generalized Jacobian) of the mapping $\Phi(\cdot, y)$ at x, and denote it by $(\partial_B)_x\Phi(x, y)$ (by $\partial_x\Phi(x, y)$). In particular, it holds that $\partial_x\Phi(x, y) = \operatorname{conv}(\partial_B)_x\Phi(x, y)$.

1.4.2 Semismoothness

The widely used concept unifying Lipschitz-continuity, directional differentiability, and the estimate quantifying the approximation properties of the generalized Jacobian, is provided by the following notion. A mapping $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ is referred to as *semismooth* at $x \in \mathbf{R}^n$ if it is locally Lipschitz-continuous at $x \in \mathbf{R}^n$, directionally differentiable at x in every direction, and the estimate

$$\sup_{J \in \partial\Phi(x+\xi)} \|\Phi(x+\xi) - \Phi(x) - J\xi\| = o(\|\xi\|)$$

holds as $\xi \in \mathbf{R}^n$ tends to 0. If the last estimate is replaced by the stronger one, namely,

$$\sup_{J \in \partial\Phi(x+\xi)} \|\Phi(x+\xi) - \Phi(x) - J\xi\| = O(\|\xi\|^2),$$

then Φ is referred to as *strongly semismooth* at x. It turns out that in many applications, the three conditions constituting the semismoothness property are often satisfied together. This fact makes this concept very convenient when applying the so-called semismooth Newton methods to various important problem classes (see Sect. 3.2).

Example 1.68. Consider again $\varphi : \mathbf{R}^n \to \mathbf{R}$, $\varphi(x) = \|x\|$. As discussed in Example 1.52 above, this function is continuously differentiable on $\mathbf{R}^n \setminus \{0\}$, and hence, $\partial\varphi(x) = \{\varphi'(x)\} = \{x/\|x\|\}$ for any x in this set. This implies that φ is strongly semismooth at any point of this set. Moreover, for any $\xi \in \mathbf{R}^n \setminus \{0\}$ it holds that

$$\begin{aligned}\varphi(\xi) - \varphi(0) - \langle \varphi'(\xi),\, \xi\rangle &= \|\xi\| - \left\langle \frac{\xi}{\|\xi\|},\, \xi \right\rangle \\ &= \|\xi\| - \|\xi\| \\ &= 0,\end{aligned}$$

and therefore, φ is strongly semismooth at 0 as well.

Example 1.69. The function $\varphi : \mathbf{R}^2 \to \mathbf{R}$, $\varphi(x) = \min\{x_1,\, x_2\}$, is strongly semismooth on the entire $\mathbf{R}^2$. This can be seen directly, employing Proposition 1.55. The same is true for the function $\varphi : \mathbf{R}^2 \to \mathbf{R}$ given by $\varphi(x) = \max\{x_1,\, x_2\} = -\min\{-x_1,\, -x_2\}$.

The notion of semismoothness was introduced in [199] for scalar-valued functions; it was extended to the general case in [225]. We next provide some equivalent definitions of semismoothness; the proofs (which we omit) combine the results in [222, 225] with Theorem 1.56 and Proposition 1.63.

Proposition 1.70. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ be locally Lipschitz-continuous at a point $x \in \mathbf{R}^n$.*

Then Φ is semismooth at x if and only if for any $\xi \in \mathbf{R}^n$ there exists finite $\lim_{\tilde{\xi}\to\xi,\, t\to 0+} J\tilde{\xi}$ which does not depend on a choice of $J \in \partial\Phi(x + t\tilde{\xi})$ and, by necessity, this limit is equal to $\Phi'(x;\, \xi)$.

Proposition 1.71. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ be locally Lipschitz-continuous at $x \in \mathbf{R}^n$ and directionally differentiable at x in every direction.*

Then Φ is semismooth at x if and only if any of the following two conditions holds:

(a) The estimate

$$\sup_{J \in \partial\Phi(x+\xi)} \|J\xi - \Phi'(x;\, \xi)\| = o(\|\xi\|)$$

holds as $\xi \in \mathbf{R}^n$ tends to 0.

(b) The estimate

$$\Phi'(x+\xi)\xi - \Phi'(x;\, \xi) = o(\|\xi\|)$$

holds as $\xi \in \mathbf{R}^n$ satisfying $x + \xi \in \mathcal{S}_\Phi$ tends to 0.

Moreover, Φ is strongly semismooth at x if any of the following two conditions holds:

(c) The estimate

$$\sup_{J \in \partial\Phi(x+\xi)} \|J\xi - \Phi'(x;\, \xi)\| = O(\|\xi\|^2)$$

holds as $\xi \in \mathbf{R}^n$ tends to 0.

(d) The estimate

$$\Phi'(x+\xi)\xi - \Phi'(x;\, \xi) = O(\|\xi\|^2)$$

holds as $\xi \in \mathbf{R}^n$ satisfying $x + \xi \in \mathcal{S}_\Phi$ tends to 0.

From this proposition it immediately follows that differentiability of Φ near x and continuity of its derivative at x imply semismoothness of Φ at x. Moreover if, in addition, the derivative of Φ is locally Lipschitz-continuous with respect to x, then Φ is strongly semismooth at x. However, the class of semismooth mappings is actually much wider. The next result is due to [199].

Proposition 1.72. *If $\varphi : \mathbf{R}^n \to \mathbf{R}$ is convex on a convex neighborhood of $x \in \mathbf{R}^n$, then φ is semismooth at x.*

As an easy consequence of Proposition 1.54 and Corollary 1.59, we obtain that a mapping is (strongly) semismooth if and only if its components are (strongly) semismooth.

Proposition 1.73. *For given $\Phi_1 : \mathbf{R}^n \to \mathbf{R}^{m_1}$ and $\Phi_2 : \mathbf{R}^n \to \mathbf{R}^{m_2}$, the mapping $\Phi(\cdot) = (\Phi_1(\cdot),\, \Phi_2(\cdot))$ is (strongly) semismooth at $x \in \mathbf{R}^n$ if and only if Φ_1 and Φ_2 are (strongly) semismooth at x.*

From (1.121) it evidently follows that (strong) semismoothness of the mapping $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ at $x \in \mathbf{R}^n$ implies (strong) semismoothness of $t\Phi$ at x for any $t \in \mathbf{R}$. Moreover, employing Propositions 1.54 and 1.73, and the results in [83, 199], one can see that (strong) semismoothness is preserved under composition.

Proposition 1.74. *Let $\Phi_1 : \mathbf{R}^n \to \mathbf{R}^{m_1}$ be (strongly) semismooth at $x \in \mathbf{R}^n$, and let $\Phi_2 : \mathbf{R}^{m_1} \to \mathbf{R}^m$ be (strongly) semismooth at $\Phi_1(x)$.*

Then the composition mapping $\Phi(\cdot) = \Phi_2(\Phi_1(\cdot))$ is (strongly) semismooth at x.

Combining Propositions 1.73 and 1.74 with the fact from Example 1.69, we derive the following.

Proposition 1.75. *Let $\Phi_1 : \mathbf{R}^n \to \mathbf{R}^m$ and $\Phi_2 : \mathbf{R}^n \to \mathbf{R}^m$ be (strongly) semismooth at $x \in \mathbf{R}^n$.*

Then the sum mapping $\Phi_1 + \Phi_2$, the inner product function $\langle \Phi_1(\cdot), \Phi_2(\cdot)\rangle$, the componentwise maximum mapping $\max\{\Phi_1(\cdot), \Phi_2(\cdot)\}$, and the componentwise minimum mapping $\min\{\Phi_1(\cdot), \Phi_2(\cdot)\}$ are all (strongly) semismooth at the point x.

For semismooth mappings, the second assertion of Proposition 1.63 can be sharpened as follows.

Proposition 1.76. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ be semismooth at $x \in \mathbf{R}^n$.*

Then for each $\xi \in \mathbf{R}^n$ there exists a matrix $J \in \partial_B \Phi(x)$ such that $\Phi'(x;\, \xi) = J\xi$.

Proof. Take any $\xi \in \mathbf{R}^n$ and any sequences $\{t_k\} \subset \mathbf{R}_+$ and $\{J_k\} \subset \mathbf{R}^{m \times n}$ such that $t_k \to 0$ and $J_k \in \partial_B \Phi(x + t_k\xi)$ for all k. According to the first assertion of Proposition 1.71, it holds that

$$t_k J_k \xi - \Phi'(x;\, t_k\xi) = o(t_k)$$

as $k \to \infty$. Then, since $\Phi(x;\, \cdot)$ is positively homogeneous,

$$\{J_k \xi\} \to \Phi'(x;\, \xi). \tag{1.130}$$

Proposition 1.51 implies that the sequence $\{J_k\}$ is bounded (hence, has accumulation points), and every accumulation point J of this sequence belongs to $\partial_B \Phi(x)$. At the same time, (1.130) implies the equality $J\xi = \Phi'(x;\, \xi)$. $\square$

Propositions 1.64 and 1.76 readily imply the following error bound result for semismooth equations.

Proposition 1.77. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ be semismooth at $\bar{x} \in \mathbf{R}^n$, let $\bar{x}$ be a solution of the equation (1.126), and assume that Φ is BD-regular at $\bar{x}$.*

Then condition (1.127) holds. Consequently, the error bound (1.128) holds as well.

Example 1.78. Define the set $\mathcal{R} = \{1/k \mid k = 1, 2, \ldots\}$ and the function $\varphi : \mathbf{R} \to \mathbf{R}$, $\varphi(x) = \mathrm{dist}(x, \mathcal{R})$ (see Fig. 1.2). This function is Lipschitz-continuous on the entire $\mathbf{R}$ with the constant $L = 1$, and $\mathcal{S}_\varphi$ consists of $(-\infty, 0)$, $(1, +\infty)$, and the intervals $(1/(k+1),\, (2k+1)/(2k(k+1)))$, $((2k+

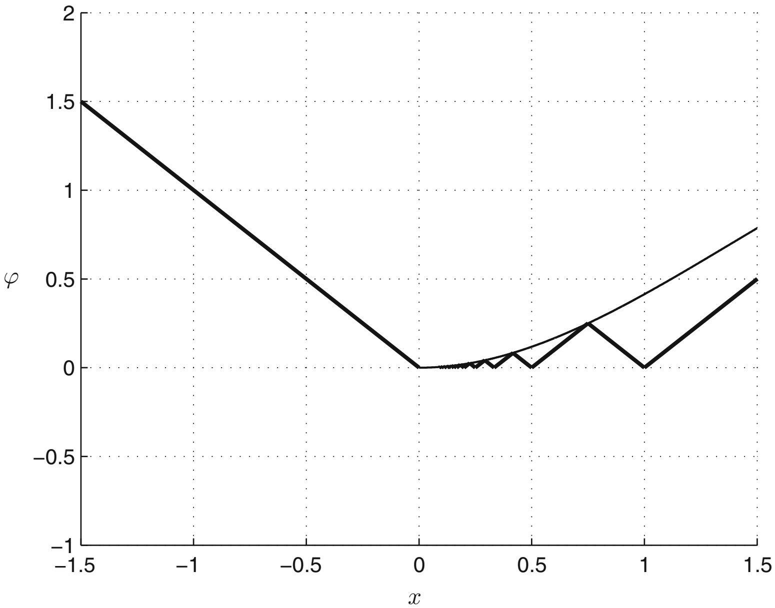

Fig. 1.2 Function from Example 1.78

$1)/(2k(k+1)),\ 1/k)$, $k = 1, 2, \ldots$, and for any x in $\mathcal{S}_\varphi$ it holds that $\varphi'(x) = -1$ or $\varphi'(x) = 1$. It follows that $\partial_B\varphi(0) = \{-1, 1\}$, so that φ is BD-regular at 0. At the same time, $\partial\varphi(0) = [-1, 1]$, and φ is not CD-regular at 0.

It can be easily seen that φ is directionally differentiable at 0 in every direction, $\varphi'(0; 1) = 0$, and hence, (1.127) does not hold. The error bound (1.128) does not hold as well, and moreover, 0 is not an isolated solution of the equation (1.126) as there is the sequence $\{1/k \mid k = 1, 2, \ldots\}$ of solutions convergent to 0. The reason for the observed situation is that φ is not semismooth at 0; that is why Proposition 1.77 is not applicable.

1.4.3 Nondifferentiable Optimization Problems and Problems with Lipschitzian Derivatives

We start with the unconstrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathbf{R}^n, \end{array} \tag{1.131}$$

where the objective function $f : \mathbf{R}^n \to \mathbf{R}$ has restricted smoothness properties. Even though this is a simple unconstrained problem, some of the results presented below seem to be not completely standard; so we provide their simple proofs.

The following theorem generalizes the Fermat principle (Theorem 1.7) to the case when the objective function is not assumed differentiable at the solution.

Theorem 1.79. *Let $\bar{x} \in \mathbf{R}^n$ be a local solution of problem (1.131) for some $f : \mathbf{R}^n \to \mathbf{R}$.*

Then the following assertions hold:

(a) If f is directionally differentiable at $\bar{x}$ in a direction $\xi \in \mathbf{R}^n$, then

$$f'(\bar{x};\, \xi) \geq 0. \tag{1.132}$$

(b) If f is locally Lipschitz-continuous at $\bar{x}$, then

$$0 \in \partial f(\bar{x}). \tag{1.133}$$

Proof. Assertion (a) follows readily from the definition of directional derivative.

In order to prove (b), fix any $\xi \in \mathbf{R}^n$ and any sequence $\{t_k\} \subset \mathbf{R}$ such that $t_k \to 0+$. Then for any $g \in \mathbf{R}^n$ and any k large enough it holds that

$$0 \leq f(\bar{x} + t_k\xi) - f(\bar{x}) \leq t_k\langle g,\, \xi\rangle + |f(\bar{x} + t_k\xi) - f(\bar{x}) - t_k\langle g,\, \xi\rangle|,$$

and hence,

$$t_k\langle g,\, \xi\rangle \geq -|f(\bar{x} + t_k\xi) - f(\bar{x}) - t_k\langle g,\, \xi\rangle|.$$

Therefore,

$$\begin{aligned} t_k \max_{g \in \partial f(\bar{x})} \langle g,\, \xi\rangle &\geq \max_{g \in \partial f(\bar{x})} (-|f(\bar{x} + t_k\xi) - f(\bar{x}) - t_k\langle g,\, \xi\rangle|) \\ &= -\min_{g \in \partial f(\bar{x})} |f(\bar{x} + t_k\xi) - f(\bar{x}) - t_k\langle g,\, \xi\rangle| \\ &= o(t_k) \end{aligned}$$

as $k \to \infty$, where the last estimate is by Corollary 1.57. Dividing both sides of the inequality above by t_k and passing onto the limit as $k \to \infty$ yields

$$\max_{g \in \partial f(\bar{x})} \langle g,\, \xi\rangle \geq 0.$$

We thus established that

$$\forall\, \xi \in \mathbf{R}^n \quad \exists\, g \in \partial f(\bar{x}) \;\text{ such that }\; \langle g,\, \xi\rangle \geq 0. \tag{1.134}$$

Suppose that (1.133) does not hold. Since $\partial f(\bar{x})$ is a closed convex set, applying the appropriate separation theorem from convex analysis (e.g., Theorem A.14), we derive the existence of $\xi \in \mathbf{R}^n$ such that

$$\langle g,\, \xi\rangle < 0 \quad \forall\, g \in \partial f(\bar{x}).$$

This contradicts (1.134), concluding the proof. □

Assertion (b) of this theorem can be found in [44, Proposition 2.3.2].

We proceed with the case when f is differentiable near the point in question, but is not necessarily twice differentiable. The following two theorems generalize Theorems 1.8 (providing SONC) and 1.9 (providing SOSC) to this setting with restricted smoothness properties.

Theorem 1.80. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be differentiable near $\bar{x} \in \mathbf{R}^n$, with its gradient being continuous near $\bar{x}$.*

If $\bar{x}$ is a local solution of problem (1.131)*, then the following assertions hold:*

(a) If f' is directionally differentiable at $\bar{x}$ in a direction $\xi \in \mathbf{R}^n$, then

$$\langle (f')'(\bar{x};\, \xi),\, \xi\rangle \geq 0. \tag{1.135}$$

(b) If f' is locally Lipschitz-continuous at $\bar{x}$, then

$$\forall\, \xi \in \mathbf{R}^n \quad \exists\, H \in \partial f'(\bar{x}) \text{ such that } \langle H\xi,\, \xi\rangle \geq 0.$$

Proof. Observe first that by Theorem 1.7 it holds that

$$f'(\bar{x}) = 0. \tag{1.136}$$

Under the assumptions of (a), employing the integral form of the mean-value theorem (see Theorem A.10, (b)) and (1.136), we derive

$$\begin{aligned}
0 &\leq f(\bar{x} + t\xi) - f(\bar{x}) \\
&= \int_0^1 \langle f'(\bar{x} + \tau t\xi),\, t\xi\rangle\, d\tau \\
&= \int_0^1 \langle f'(\bar{x} + \tau t\xi) - f'(\bar{x}),\, t\xi\rangle\, d\tau \\
&= \int_0^1 \langle \tau t (f')'(\bar{x};\, \xi),\, t\xi\rangle\, d\tau + o(t^2) \\
&= t^2 \langle (f')'(\bar{x};\, \xi),\, \xi\rangle \int_0^1 \tau\, d\tau + o(t^2) \\
&= \frac{1}{2} t^2 \langle (f')'(\bar{x};\, \xi),\, \xi\rangle + o(t^2)
\end{aligned}$$

as $t \to 0+$. Dividing both sides of this relation by t^2 and passing onto the limit as $t \to 0+$, we obtain (1.135).

Assertion (b) easily follows from Corollary 1.57, employing Proposition 1.51 and the equality (1.136). □

Assertion (b) of Theorem 1.80 was suggested in [120]. Assuming that f possesses the locally Lipschitzian gradient, the generalized Jacobian $\partial f'(x)$ used in this assertion is referred to as the *generalized Hessian* of f at x.

Theorem 1.81. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be differentiable near $\bar{x} \in \mathbf{R}^n$, with its gradient being locally Lipschitz-continuous at $\bar{x}$. Assume that $\bar{x}$ satisfies (1.136) and any of the following conditions hold:*

(i) f' is directionally differentiable at $\bar{x}$ in every direction, and

$$\langle (f')'(\bar{x};\, \xi),\, \xi\rangle > 0 \quad \forall\, \xi \in \mathbf{R}^n \setminus \{0\}. \tag{1.137}$$

(ii) It holds that

$$\forall\, H \in \partial f'(\bar{x}) \quad \langle H\xi,\, \xi\rangle > 0 \quad \forall\, \xi \in \mathbf{R}^n \setminus \{0\}. \tag{1.138}$$

Then $\bar{x}$ is a strict local solution of problem (1.131).

Moreover, assuming that f' is directionally differentiable at $\bar{x}$ in every direction, (1.137) is implied by (1.138).

Proof. Assuming (i), we argue by contradiction. Suppose that there exists a sequence $\{x^k\} \subset \mathbf{R}^n$ convergent to $\bar{x}$ and such that $x^k \neq \bar{x}$ for all k, and

$$f(x^k) \leq f(\bar{x}) \quad \forall\, k. \tag{1.139}$$

For each k we set $t_k = \|x^k - \bar{x}\|$ and $\xi^k = (x^k - \bar{x})/\|x^k - \bar{x}\|$; then, $x^k = \bar{x} + t_k\xi^k$, and without loss of generality we can assume that $\{\xi^k\}$ converges to some $\xi \in \mathbf{R}^n$, $\|\xi\| = 1$.

Making use of the integral form of the mean-value theorem (see Theorem A.10, (b)), (1.132) and local Lipschitz-continuity of f' at $\bar{x}$, from (1.139) we derive

$$\begin{aligned}
0 &\geq f(\bar{x} + t_k\xi^k) - f(\bar{x}) \\
&= \int_0^1 \langle f'(\bar{x} + \tau t_k\xi^k),\, t_k\xi^k\rangle\, d\tau \\
&= \int_0^1 \langle f'(\bar{x} + \tau t_k\xi^k) - f'(\bar{x}),\, t_k\xi^k\rangle\, d\tau \\
&\geq \int_0^1 \langle f'(\bar{x} + \tau t_k\xi) - f'(\bar{x}),\, t_k\xi^k\rangle\, d\tau \\
&\quad - \int_0^1 t_k\|f'(\bar{x} + \tau t_k\xi^k) - f'(\bar{x} + \tau t_k\xi)\|\|\xi^k\|\, d\tau \\
&= \int_0^1 \langle \tau t_k(f')'(\bar{x};\, \xi),\, t_k\xi^k\rangle\, d\tau + o(t_k^2) + O(t_k^2\|\xi^k - \xi\|)
\end{aligned}$$

$$
\begin{aligned}
&= t_k^2 \langle (f')'(\bar{x};\, \xi),\, \xi^k \rangle \int_0^1 \tau\, d\tau + o(t_k^2) \\
&= \frac{1}{2} t_k^2 \langle (f')'(\bar{x};\, \xi),\, \xi^k \rangle + o(t_k^2)
\end{aligned}
$$

as $k \to \infty$, where the next-to-last equality is by convergence of $\{\xi^k\}$ to ξ. Dividing both sides of the relation above by t_k^2 and passing onto the limit as $k \to \infty$, we obtain

$$\langle (f')'(\bar{x};\, \xi),\, \xi \rangle \leq 0,$$

which contradicts (1.137).

Assuming (ii), this theorem is a particular case of Theorem 1.82 below.

The last assertion follows from Proposition 1.63. □

Condition (1.138) in Theorem 1.81 can be replaced by the seemingly weaker but actually equivalent condition

$$\forall\, H \in \partial_B f'(\bar{x}) \quad \langle H\xi,\, \xi \rangle > 0 \quad \forall\, \xi \in \mathbf{R}^n \setminus \{0\}. \tag{1.140}$$

The equivalence holds because the set of matrices involved in (1.138) is the convex hull of the set of matrices involved in (1.140).

Finally, consider the mathematical programming problem

$$
\begin{array}{ll}
\text{minimize} & f(x) \\
\text{subject to} & h(x) = 0,\ g(x) \leq 0,
\end{array} \tag{1.141}
$$

where the objective function $f : \mathbf{R}^n \to \mathbf{R}$ and the constraints mappings $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are smooth and have locally Lipschitzian derivatives but may not be twice differentiable. The following SOSC for this problem setting was derived in [170].

Theorem 1.82. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable near $\bar{x} \in \mathbf{R}^n$, with their derivatives being locally Lipschitz-continuous at $\bar{x}$.*

If $h(\bar{x}) = 0$ and $g(\bar{x}) \leq 0$ and there exist $\bar{\lambda} \in \mathbf{R}^l$ and $\bar{\mu} \in \mathbf{R}^m_+$ satisfying

$$\frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) = 0, \quad \langle \bar{\mu},\, g(\bar{x}) \rangle = 0,$$

together with the condition that

$$\forall\, H \in \partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) \quad \langle H\xi,\, \xi \rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{1.142}$$

where

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A(\bar{x})}(\bar{x})\xi \leq 0,\ \langle f'(\bar{x}),\, \xi \rangle \leq 0\}$$

is the critical cone of problem (1.141) *at* $\bar{x}$*, then* $\bar{x}$ *is a strict local solution of problem* (1.141)*.*

Similarly to the unconstrained case, (1.142) can be replaced by the seemingly weaker but in fact equivalent condition

$$\forall\, H \in (\partial_B)_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) \quad \langle H\xi,\, \xi\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}.$$

Chapter 2
Equations and Unconstrained Optimization

In this chapter, we start our discussion of Newton-type methods, which are based on the fundamental principle of linear/quadratic approximation of the problem data (or of some part of the problem data). The underlying idea of Newtonian methods is extremely important, as it serves as a foundation for numerous computationally efficient algorithms for optimization and variational problems.

We start with discussing the basic Newton method for nonlinear equations and unconstrained optimization. High rate of convergence of this scheme is due to using information about the derivatives of the problem data (first derivatives of the operator in the case of nonlinear equations, second derivatives of the objective function in the case of optimization). Thus, each iteration of this basic process should be regarded as relatively expensive. However, one of the main messages of this chapter is that various kinds of inexactness, introduced *intentionally* into the basic Newton scheme, can serve to reduce the cost of the iteration while keeping the convergence rate still high enough. Combined with globalization techniques, such modifications lead to truly practical Newtonian methods for unconstrained optimization problems, the most important of which belong to the quasi-Newton class.

As much of the material covered in this chapter can be considered nowadays quite standard (e.g., linesearch quasi-Newton methods, trust-region methods, etc.), we sometimes mention only the main principles behind certain techniques without going into full details. On the other hand, the general perturbed Newton framework is analyzed very thoroughly, as its natural generalization for optimization and variational problems would be one of the main tools for treating various algorithms throughout the book.

A.F. Izmailov and M.V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-3-319-04247-3_2,

2.1 Newton Method

For historical comments regarding the Newton method, we address the reader to [62].

2.1.1 Newton Method for Equations

The classical Newton method is introduced for the equation

$$\Phi(x) = 0, \tag{2.1}$$

where $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ is a smooth mapping. Let $x^k \in \mathbf{R}^n$ be the current approximation to a solution of (2.1). Then it is natural to approximate the equation (2.1) near the point x^k by its linearization:

$$\Phi(x^k) + \Phi'(x^k)(x - x^k) = 0. \tag{2.2}$$

The linearized equation (2.2) gives the iteration system of the classical *Newton method*. The idea is transparent — the nonlinear equation (2.1) is replaced by the (computationally much simpler) linear equation (2.2). Iterations of the Newton method for the case when $n = 1$ are illustrated in Fig. 2.1.

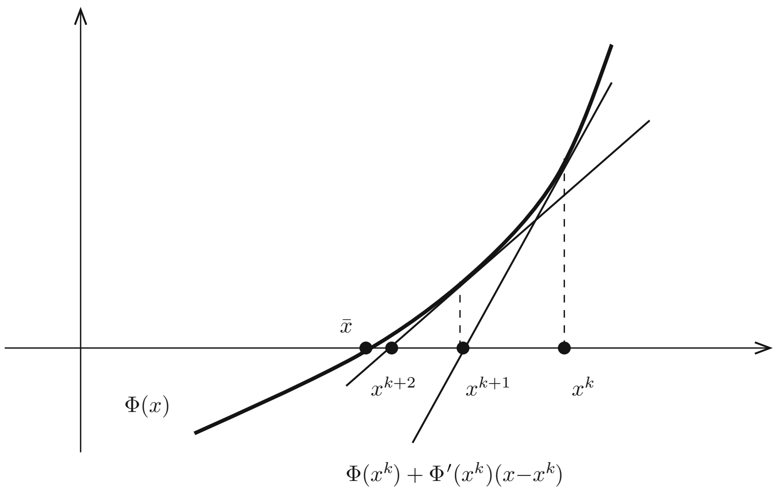

Fig. 2.1 Iterations of the Newton method

Formally, the algorithm is stated as follows.

Algorithm 2.1 Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $\Phi(x^k) = 0$, stop.
2. Compute $x^{k+1} \in \mathbf{R}^n$ as a solution of (2.2).
3. Increase k by 1 and go to step 1.

Assuming that the Jacobian $\Phi'(x^k)$ is nonsingular, the Newton method is often presented in the form of the explicit iterative scheme

$$x^{k+1} = x^k - (\Phi'(x^k))^{-1}\Phi(x^k), \quad k = 0, 1, \ldots, \tag{2.3}$$

with the understanding that an actual implementation of the method need not require computing the complete inverse of the matrix $\Phi'(x^k)$; of interest is only the product $(\Phi'(x^k))^{-1}\Phi(x^k)$.

Under appropriate assumptions, the Newton method is very efficient, which is reflected in the following convergence statements. At the same time, it is clear that in its pure form the method may not converge from points that are not close enough to a solution, even if the latter satisfies all the needed assumptions; see Fig. 2.2 and also Example 2.16 below.

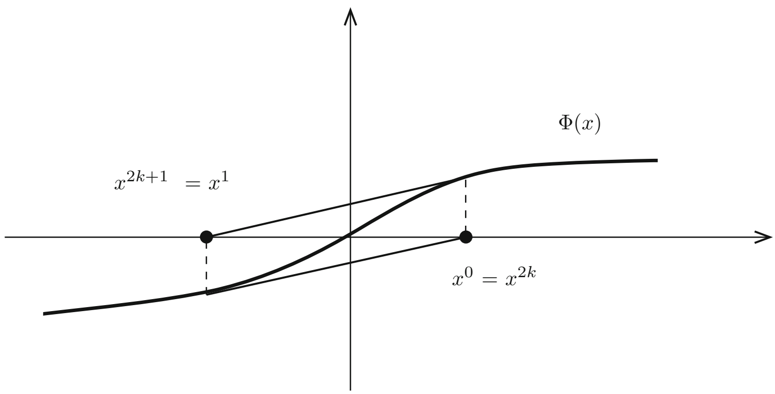

Fig. 2.2 Non-convergence of the Newton method from points far from a solution

The following describes the essential convergence properties of the Newton method.

Theorem 2.2. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighborhood of a point $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$. Let $\bar{x}$ be a solution of the equation (2.1), and assume that $\Phi'(\bar{x})$ is a nonsingular matrix.*

Then the following assertions are valid:

(a) There exists a neighborhood U of $\bar{x}$ and a function $q(\cdot) : U \to \mathbf{R}$ such that $\Phi'(x)$ is nonsingular for all $x \in U$,

$$\|x - (\Phi'(x))^{-1}\Phi(x) - \bar{x}\| \leq q(x)\|x - \bar{x}\| \quad \forall\, x \in U, \tag{2.4}$$

and

$$q(x) \to 0 \text{ as } x \to \bar{x}. \tag{2.5}$$

(b) *Any starting point $x^0 \in \mathbf{R}^n$ close enough to $\bar{x}$ uniquely defines a particular iterative sequence of Algorithm 2.1; this sequence converges to $\bar{x}$, and the rate of convergence is superlinear.*

(c) *If the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$, then $q(\cdot)$ can be chosen in such a way that*

$$q(x) = O(\|x - \bar{x}\|) \tag{2.6}$$

as $x \to \bar{x}$, and the rate of convergence is quadratic.

Assertion (a) means that the Newton step from a point close enough to $\bar{x}$ provides a "superlinear decrease" of the distance to $\bar{x}$, while assertion (c) gives conditions guaranteeing "quadratic decrease" of this distance.

Regarding formal definitions of convergence rates (in particular, superlinear and quadratic), see Sect. A.2.

Proof. According to Lemma A.6, there exist a neighborhood U of $\bar{x}$ and $M > 0$ such that

$$\Phi'(x) \text{ is nonsingular}, \quad \|(\Phi'(x))^{-1}\| \leq M \quad \forall\, x \in U. \tag{2.7}$$

Employing the mean-value theorem (see Theorem A.10, (a)), we can choose U in such a way that the inclusion $x \in U$ implies

$$\begin{aligned} \|x - (\Phi'(x))^{-1}\Phi(x) - \bar{x}\| &\leq \|(\Phi'(x^k))^{-1}\|\|\Phi(x) - \Phi(\bar{x}) - \Phi'(x)(x - \bar{x})\| \\ &\leq q(x)\|x - \bar{x}\|, \end{aligned} \tag{2.8}$$

where

$$q(x) = M \sup\{\|\Phi'(tx + (1-t)\bar{x}) - \Phi'(x)\| \mid t \in [0,\, 1]\}. \tag{2.9}$$

It is clear that this $q(\cdot)$ satisfies (2.5), while (2.8) gives (2.4). This completes the proof of assertion (a).

In particular, for $x^k \in U$, the equation (2.2) has the unique solution x^{k+1} given by (2.3). Moreover, from (2.4) and (2.5) it follows that for any $q \in (0,\, 1)$ there exists $\delta > 0$ such that $B(\bar{x},\, \delta) \subset U$, and the inclusion $x^k \in B(\bar{x},\, \delta)$ implies

$$\|x^{k+1} - \bar{x}\| \leq q\|x^k - \bar{x}\|.$$

In particular, $x^{k+1} \in B(\bar{x},\, \delta)$. It follows that any starting point $x^0 \in B(\bar{x},\, \delta)$ uniquely defines a specific iterative sequence $\{x^k\}$ of Algorithm 2.1; this sequence is contained in $B(\bar{x},\, \delta)$ and converges to $\bar{x}$. Moreover, again employing (2.4), we obtain the estimate

$$\|x^{k+1} - \bar{x}\| \leq q(x^k)\|x^k - \bar{x}\| \quad \forall\, k = 0,\, 1,\, \ldots, \tag{2.10}$$

which, according to (2.5), implies the superlinear rate of convergence. This completes the proof of assertion (b).

Finally, if the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$ with a constant $L > 0$, then, after reducing U if necessary, from (2.9) it follows that the inclusion $x \in U$ implies

$$\begin{aligned} q(x) &\leq M(\sup\{\|\Phi'(\bar{x} + t(x - \bar{x})) - \Phi'(\bar{x})\| \mid t \in [0, 1]\} + \|\Phi'(x) - \Phi'(\bar{x})\|) \\ &\leq 2ML\|x - \bar{x}\|, \end{aligned}$$

which proves (2.6). The quadratic convergence rate now follows from (2.6) and (2.10). This proves (c). □

The main message of the subsequent discussion in this section is that various kinds of inexactness introduced *intentionally* in the basic Newton scheme may lead to more practical Newton-type methods, with lower computational costs per iteration but convergence rate still high enough. To that end, we consider the following general scheme, which we refer to as the *perturbed Newton method*. For a given $x^k \in \mathbf{R}^n$, the next iterate $x^{k+1} \in \mathbf{R}^n$ satisfies the perturbed version of the iteration system (2.2):

$$\Phi(x^k) + \Phi'(x^k)(x - x^k) + \omega^k = 0. \tag{2.11}$$

Here, $\omega^k \in \mathbf{R}^n$ is a perturbation term, which may have various forms and meanings, may play various roles, and may conform to different sets of assumptions depending on the particular algorithms at hand and on the particular purposes of the analysis. At the moment, we are interested in the following general but simple question: under which assumptions regarding ω^k the local convergence and/or the superlinear rate of convergence of the pure Newton method (2.2) is preserved?

We start with some basic (essentially technical) statements, which do not impose any restrictions on the structure of ω^k. Note that this is an a posteriori kind of analysis: the iterative sequence $\{x^k\}$ is given, and the corresponding sequence $\{\omega^k\}$ is then explicitly defined by (2.11). Thus, in this setting the role of $\{\omega^k\}$ is secondary with respect to $\{x^k\}$. Those technical results would be useful later on for analyzing iterative sequences generated by specific Newton-type schemes.

Lemma 2.3. *Under the assumptions of Theorem 2.2, there exist a neighborhood U of $\bar{x}$ and $M > 0$ such that for any $x^k \in U$ and any $x^{k+1} \in \mathbf{R}^n$ and $\omega^k \in \mathbf{R}^n$ satisfying*

$$\omega^k = -\Phi(x^k) - \Phi'(x^k)(x^{k+1} - x^k), \tag{2.12}$$

it holds that

$$\|x^{k+1} - \bar{x}\| \leq M\omega^k + o(\|x^k - \bar{x}\|) \tag{2.13}$$

as $x^k \to \bar{x}$. Moreover, if the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$, then the estimate (2.13) can be sharpened as follows:

$$\|x^{k+1} - \bar{x}\| \leq M\omega^k + O(\|x^k - \bar{x}\|^2). \tag{2.14}$$

Proof. By assertion (a) of Theorem 2.2 and by Lemma A.6, there exist a neighborhood U of $\bar{x}$ and $M > 0$ such that (2.7) holds, and

$$x^k - (\Phi'(x^k))^{-1}\Phi(x^k) - \bar{x} = o(\|x^k - \bar{x}\|) \tag{2.15}$$

as $x^k \in U$ tends to $\bar{x}$. Furthermore, by (2.12),

$$x^{k+1} = x^k - (\Phi'(x^k))^{-1}(\Phi(x^k) + \omega^k).$$

Hence, employing (2.7) and (2.15), we obtain that

$$\begin{aligned}\|x^{k+1} - \bar{x}\| &= \|x^k - (\Phi'(x^k))^{-1}(\Phi(x^k) + \omega^k) - \bar{x}\| \\ &\leq \|(\Phi'(x^k))^{-1}\|\|\omega^k\| + \|x^k - (\Phi'(x^k))^{-1}\Phi(x^k) - \bar{x}\| \\ &\leq M\omega^k + o(\|x^k - \bar{x}\|),\end{aligned}$$

which establishes (2.13).

Finally, if the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$, estimate (2.14) follows by the same argument, but invoking assertion (c) of Theorem 2.2. □

The next result states a necessary and sufficient condition on the perturbation sequence $\{\omega^k\}$ under which superlinear convergence of $\{x^k\}$ is preserved. Note that convergence itself is not established but assumed here.

Proposition 2.4. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$. Let $\bar{x}$ be a solution of the equation (2.1). Let a sequence $\{x^k\} \subset \mathbf{R}^n$ be convergent to $\bar{x}$, and define ω^k according to (2.12) for each $k = 0, 1, \ldots$.*

If the rate of convergence of $\{x^k\}$ is superlinear, then

$$\omega^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \tag{2.16}$$

as $k \to \infty$.

Conversely, if $\Phi'(\bar{x})$ is a nonsingular matrix, and (2.16) holds, then the rate of convergence of $\{x^k\}$ is superlinear. Moreover, the rate of convergence is quadratic, provided the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$ and

$$\omega^k = O(\|x^{k+1} - x^k\|^2 + \|x^k - \bar{x}\|^2) \tag{2.17}$$

as $k \to \infty$.

Proof. By (2.12) and the mean-value theorem (see Theorem A.10), we obtain that for all k large enough

$$\begin{aligned}
\|\omega^k\| &= \|\Phi(x^k) + \Phi'(x^k)(x^{k+1} - x^k)\| \\
&\leq \|\Phi(x^k) - \Phi(\bar{x}) - \Phi'(x^k)(x^k - \bar{x})\| + \|\Phi'(x^k)\|\|x^{k+1} - \bar{x}\| \\
&\leq \sup\{\|\Phi'(tx^k + (1-t)\bar{x}) - \Phi'(x^k)\| \mid t \in [0,\, 1]\}\|x^k - \bar{x}\| \\
&\quad + O(\|x^{k+1} - \bar{x}\|) \\
&= o(\|x^k - \bar{x}\|) + O(\|x^{k+1} - \bar{x}\|)
\end{aligned}$$

as $k \to \infty$. If the sequence $\{x^k\}$ converges to $\bar{x}$ superlinearly, the above implies that $\omega^k = o(\|x^k - \bar{x}\|)$, which in turn implies (2.16).

Suppose now that (2.16) holds. From Lemma 2.3 it then follows that

$$x^{k+1} - \bar{x} = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) = o(\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|),$$

i.e., there exists a sequence $\{t_k\} \subset \mathbf{R}$ such that $t_k \to 0$ and

$$\|x^{k+1} - \bar{x}\| \leq t_k(\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|).$$

for all k large enough. This implies that

$$(1 - t_k)\|x^{k+1} - \bar{x}\| \leq t_k\|x^k - \bar{x}\|.$$

Hence, for all k large enough

$$\|x^{k+1} - \bar{x}\| \leq \frac{t_k}{1 - t_k}\|x^k - \bar{x}\|,$$

i.e.,

$$x^{k+1} - \bar{x} = o(\|x^k - \bar{x}\|)$$

as $k \to \infty$, which gives the superlinear convergence rate.

Finally, if the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$, from Lemma 2.3 it follows that (2.17) implies the estimate

$$x^{k+1} - \bar{x} = O(\|x^{k+1} - x^k\|^2 + \|x^k - \bar{x}\|^2) = O(\|x^{k+1} - \bar{x}\|^2 + \|x^k - \bar{x}\|^2)$$

as $k \to \infty$, which means that there exists $M > 0$ such that

$$\|x^{k+1} - \bar{x}\| \leq M(\|x^{k+1} - \bar{x}\|^2 + \|x^k - \bar{x}\|^2) \tag{2.18}$$

for all k large enough. Since $\{x^k\}$ converges to $\bar{x}$, for any fixed $\varepsilon \in (0,\, 1)$ it holds that $M\|x^{k+1} - \bar{x}\| \leq 1 - \varepsilon$ for all k large enough. Then from (2.18) we derive

$$(1 - M\|x^{k+1} - \bar{x}\|)\|x^{k+1} - \bar{x}\| \leq M\|x^k - \bar{x}\|^2,$$

and hence, for all k large enough

$$\|x^{k+1} - \bar{x}\| \leq \frac{M}{1 - M\|x^{k+1} - \bar{x}\|}\|x^k - \bar{x}\|^2 \leq \frac{M}{\varepsilon}\|x^k - \bar{x}\|^2,$$

which gives the quadratic convergence rate. $\square$

Remark 2.5. If $\{x^k\}$ converges to $\bar{x}$ superlinearly, the estimate (2.16) is, in fact, equivalent to either of the following two (generally stronger) estimates:

$$\omega^k = o(\|x^{k+1} - x^k\|), \tag{2.19}$$

or

$$\omega^k = o(\|x^k - \bar{x}\|). \tag{2.20}$$

Indeed, by (2.16) and the superlinear convergence rate of $\{x^k\}$ to $\bar{x}$, there exist sequences $\{t_k\} \subset \mathbf{R}$ and $\{\tau_k\} \subset \mathbf{R}$ such that $t_k \to 0$, $\tau_k \to 0$, and

$$\|\omega^k\| \leq t_k(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|), \tag{2.21}$$

$$\|x^{k+1} - \bar{x}\| \leq \tau_k \|x^k - \bar{x}\| \tag{2.22}$$

for all k. Then

$$\|x^k - \bar{x}\| \leq \|x^{k+1} - x^k\| + \|x^{k+1} - \bar{x}\| \leq \|x^{k+1} - x^k\| + \tau_k \|x^k - \bar{x}\|,$$

implying that

$$\|x^k - \bar{x}\| \leq \frac{1}{1 - \tau_k} \|x^{k+1} - x^k\|$$

for all k large enough. Combining this with (2.21) we then obtain that

$$\|\omega^k\| \leq t_k \left(1 + \frac{1}{1 - \tau_k}\right) \|x^{k+1} - x^k\| = t_k \frac{2 - \tau_k}{1 - \tau_k} \|x^{k+1} - x^k\|$$

for all k large enough, which gives (2.19). Furthermore, from (2.21) and (2.22) we directly derive that

$$\begin{aligned} \|\omega^k\| &\leq t_k(2\|x^k - \bar{x}\| + \|x^{k+1} - \bar{x}\|) \\ &\leq t_k(2\|x^k - \bar{x}\| + \tau_k \|x^k - \bar{x}\|) \\ &\leq t_k(2 + \tau_k)\|x^k - \bar{x}\| \end{aligned}$$

for all k, which gives (2.20).

The next result provides a sufficient condition on the perturbation sequence $\{\omega^k\}$ for preserving local convergence of $\{x^k\}$.

Proposition 2.6. *Under the assumptions of Theorem 2.2, fix any norm $\|\cdot\|_*$ in $\mathbf{R}^n$, any $q_1, q_2 \geq 0$ such that $2q_1 + q_2 < 1$, and any $\varepsilon \in (0, 1 - 2q_1 - q_2)$.*

Then there exists $\delta > 0$ such that for any sequence $\{x^k\} \subset \mathbf{R}^n$ the following assertions are valid:

(a) If for some $k = 0, 1, \ldots,$ it holds that $x^k \in B(\bar{x}, \delta)$, and ω^k defined according to (2.12) satisfies the condition

$$\|(\Phi'(\bar{x}))^{-1}\omega^k\|_* \leq q_1\|x^{k+1} - x^k\|_* + q_2\|x^k - \bar{x}\|_* \quad \forall\, k = 0, 1, \ldots, \tag{2.23}$$

then

$$\|x^{k+1} - \bar{x}\|_* \leq \frac{q_1 + q_2 + \varepsilon}{1 - q_1} \|x^k - \bar{x}\|_* \tag{2.24}$$

and, in particular, $x^{k+1} \in B(\bar{x}, \delta)$.

(b) If $x^0 \in B(\bar{x}, \delta)$ *and* (2.23) *is satisfied for all* $k = 0, 1, \ldots$, *then* $\{x^k\}$ *converges to* $\bar{x}$, *and the rate of convergence is (at least) linear. More precisely, either* $x^k = \bar{x}$ *for all* k *large enough, or*

$$\limsup_{k \to \infty} \frac{\|x^{k+1} - \bar{x}\|_*}{\|x^k - \bar{x}\|_*} \leq \frac{q_1 + q_2}{1 - q_1}. \tag{2.25}$$

Proof. By (2.12) and (2.23), employing assertion (a) of Theorem 2.2 and the equivalence of norms in $\mathbf{R}^n$, we obtain that for any $\varepsilon \in (0, 1 - 2q_1 - q_2)$ there exists $\delta > 0$ such that for any $x^k \in B(\bar{x}, \delta)$ it holds that

$$\begin{aligned}
\|x^{k+1} - \bar{x}\|_* &= \|x^k - (\Phi'(x^k))^{-1}(\Phi(x^k) + \omega^k) - \bar{x}\|_* \\
&\leq \|(\Phi'(x^k))^{-1}\omega^k\|_* + \|x^k - (\Phi'(x^k))^{-1}\Phi(x^k) - \bar{x}\|_* \\
&\leq q_1\|x^{k+1} - x^k\|_* + q_2\|x^k - \bar{x}\|_* + o(\|x^k - \bar{x}\|_*) \\
&\leq q_1(\|x^{k+1} - \bar{x}\|_* + \|x^k - \bar{x}\|_*) + q_2\|x^k - \bar{x}\|_* + \varepsilon\|x^k - \bar{x}\|_* \\
&\leq q_1\|x^{k+1} - \bar{x}\|_* + (q_1 + q_2 + \varepsilon)\|x^k - \bar{x}\|_*.
\end{aligned}$$

This implies (2.24). Since $(q_1 + q_2 + \varepsilon)/(1 - q_1) \in (0, 1)$, (2.24) implies that $x^{k+1} \in B(\bar{x}, \delta)$. This proves assertion (a).

Furthermore, the inclusion $x^0 \in B(\bar{x}, \delta)$ and assertion (a) imply that the entire sequence $\{x^k\}$ is contained in $B(\bar{x}, \delta)$, and (2.24) shows convergence of this sequence to $\bar{x}$ at a linear rate. Moreover, since ε can be taken arbitrarily small at a price of reducing δ, and since $\{x^k\}$ converges to $\bar{x}$ (hence, the tail of $\{x^k\}$ is contained in $B(\bar{x}, \delta)$ no matter how small δ is), relation (2.24) implies that either $x^k = \bar{x}$ for all k large enough, or (2.25) holds. This proves assertion (b). $\square$

Conditions (2.16) and/or (2.23) are not "practical," because they involve the unknown solution $\bar{x}$ and/or the next iterate x^{k+1}, which is usually computed *after* the perturbation term is settled. Propositions 2.4 and 2.6 are merely technical tools intended for the analysis of some specific algorithms fitting the perturbed Newton method framework.

We start with the class of the so-called *truncated Newton methods*, which were first systematically studied in [55], and which are particular instances of perturbed Newton methods with the perturbation terms satisfying the condition

$$\|\omega^k\| \leq \theta_k\|\Phi(x^k)\|, \quad k = 0, 1, \ldots. \tag{2.26}$$

Here, $\{\theta_k\}$ is a sequence of nonnegative numbers, called *forcing sequence*, which can be either pre-fixed or computed in the course of iterations. The idea of truncated Newton methods consists of solving the iteration system (2.2) not exactly, but up to the accuracy defined by the right-hand side of the inequality in (2.26). Note that (2.26) is totally practical as an approximation criterion for solving the Newton method iteration system (2.2), as it does not involve any unknown objects (such as the solution $\bar{x}$ and/or x^{k+1}, as in the technical conditions (2.16) and (2.23)). Thus, (2.26) can be easily checked in the course of solving (2.2). The most popular strategy is to apply to the linear equation (2.2) some iterative method (e.g., the conjugate gradient method for minimizing its squared residual), and to stop this inner process when (2.26) will be satisfied for ω^k defined in (2.11) with x being the current iterate of the inner process. Once (2.26) is satisfied, x in (2.11) is declared to be the next iterate x^{k+1}. Supplied with a rule for computing the forcing sequence $\{\theta_k\}$ and a choice of an inner iterative scheme, this algorithmic framework results in a specific truncated Newton method.

Employing Propositions 2.4 and 2.6, we obtain the following properties.

Theorem 2.7. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighborhood of a point $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$. Let $\bar{x}$ be a solution the equation (2.1). Let $\{x^k\} \subset \mathbf{R}^n$ be a sequence convergent to $\bar{x}$, and let ω^k be defined according to (2.12) for each $k = 0, 1, \ldots$.*

If the rate of convergence of $\{x^k\}$ is superlinear, then there exists a sequence $\{\theta_k\} \subset \mathbf{R}$ satisfying condition (2.26), and such that $\theta_k \to 0$.

Conversely, if $\Phi'(\bar{x})$ is a nonsingular matrix and there exists a sequence $\{\theta_k\} \subset \mathbf{R}$ satisfying condition (2.26) and such that $\theta_k \to 0$, then the rate of convergence of $\{x^k\}$ is superlinear. Moreover, the rate of convergence is quadratic, provided the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$ and

$$\theta_k = O(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \tag{2.27}$$

as $k \to \infty$.

Proof. To prove the first assertion, observe that by the error bound presented in Proposition 1.32, it holds that

$$x^k - \bar{x} = O(\|\Phi(x^k)\|) \tag{2.28}$$

as $k \to \infty$. By Proposition 2.4 and Remark 2.5, superlinear convergence rate of $\{x^k\}$ implies (2.20). Thus, by (2.28), we have that

$$\omega^k = o(\|x^k - \bar{x}\|) = o(\|\Phi(x^k)\|),$$

which means precisely the existence of a sequence $\{\theta_k\}$ with the needed properties.

The second assertion follows from Proposition 2.4. Indeed,

$$\Phi(x^k) = \Phi(\bar{x}) + \Phi'(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) = O(\|x^k - \bar{x}\|) \tag{2.29}$$

as $k \to \infty$, and therefore, (2.26) with $\theta_k \to 0$ evidently implies (2.20) (and, hence, (2.16)).

Finally, if the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$, and (2.27) holds, quadratic convergence follows by the last assertion of Proposition 2.4, because in this case, taking into account (2.29), we derive that

$$\begin{aligned}\omega^k &= O((\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)\|\Phi(x^k)\|)\\ &= O((\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)\|x^k - \bar{x}\|)\\ &= O(\|x^{k+1} - x^k\|^2 + \|x^k - \bar{x}\|^2)\end{aligned}$$

as $k \to \infty$. □

In the previous result, convergence of $\{x^k\}$ was assumed. But to pass to a constructive result, also establishing convergence, is now easy.

Theorem 2.8. *Suppose that the assumptions of Theorem 2.2 hold, and let $\theta \in (0,\, 1)$ be arbitrary.*

Then for any $x^0 \in \mathbf{R}^n$ close enough to $\bar{x}$ and any sequences $\{x^k\} \subset \mathbf{R}^n$, $\{\omega^k\} \subset \mathbf{R}^n$ and $\{\theta^k\} \subset [0,\, \theta]$ satisfying (2.12) and (2.26) for all $k = 0,\, 1,\, \ldots$, it holds that $\{x^k\}$ converges to $\bar{x}$ and the rate of convergence is (at least) linear. Moreover, the rate of convergence is superlinear provided $\theta_k \to 0$. The rate of convergence is quadratic, provided the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$ and

$$\theta_k = O(\|\Phi(x^k)\|) \tag{2.30}$$

as $k \to \infty$.

Proof. Define the following norm in $\mathbf{R}^n$: $\|x\|_* = \|\Phi'(\bar{x})x\|$, $x \in \mathbf{R}^n$ (as $\Phi'(\bar{x})$ is nonsingular, this is indeed a norm). Then, employing (2.26) and the equivalence of norms in $\mathbf{R}^n$, we obtain that for any $\varepsilon \in (0,\, 1-\theta)$ there exists $\delta > 0$ such that for any $x^k \in B(\bar{x},\, \delta)$ it holds that

$$\begin{aligned}\|(\Phi'(\bar{x}))^{-1}\omega^k\|_* &= \|\omega^k\|\\ &\leq \theta\|\Phi(x^k)\|\\ &= \theta\|\Phi(\bar{x}) + \Phi'(\bar{x})(x^k - \bar{x})\| + o(\|x^k - \bar{x}\|)\\ &= \theta\|x^k - \bar{x}\|_* + o(\|x^k - \bar{x}\|_*)\\ &\leq (\theta + \varepsilon)\|x^k - \bar{x}\|_*,\end{aligned}$$

which is (2.23) with $q = \theta + \varepsilon$. By assertion (a) of Proposition 2.6, this implies the inclusion $x^{k+1} \in B(\bar{x},\, \delta)$, provided δ is chosen small enough. Thus, the inclusion $x^0 \in B(\bar{x},\, \delta)$ implies that the entire sequence $\{x^k\}$ is contained in $B(\bar{x},\, \delta)$, and that (2.23) holds for all $k = 0,\, 1,\, \ldots$. It remains to apply assertion (b) of Proposition 2.6. Superlinear rate of convergence when $\theta_k \to 0$, and quadratic rate when the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$ and (2.30) holds, follow from Theorem 2.7, taking into account (2.29). □

Some specific implementations of truncated Newton methods and related results can be found, e.g., in [208, Chap. 11].

It is interesting to note that the class of *quasi-Newton methods*, which is of great practical importance, can be (theoretically) related to truncated Newton methods, even though the principles behind the two approaches are completely different. Close enough to a solution, a step of any quasi-Newton method is supposed to take the form

$$x^{k+1} = x^k - J_k^{-1}\Phi(x^k), \tag{2.31}$$

where $\{J_k\} \subset \mathbf{R}^{n\times n}$ is a sequence of nonsingular matrices satisfying the so-called *Dennis–Moré condition* (see [57, 58]):

$$(J_k - \Phi'(x^k))(x^{k+1} - x^k) = o(\|x^{k+1} - x^k\|) \tag{2.32}$$

as $k \to \infty$. (As usual, (2.31) does not mean that a matrix is inverted in actual computation.)

Evidently, x^{k+1} is a solution of (2.11) with

$$\omega^k = (J_k - \Phi'(x^k))(x^{k+1} - x^k).$$

Note that (2.32) is merely an asymptotic condition of an a posteriori kind, not relating the properties of two subsequent iterates in any constructive way. Thus, one should certainly not expect any complete convergence results, and even less so, any a priori results (i.e., proving convergence itself) under an assumption so weak. What can be expected, at best, is the superlinear rate of convergence *assuming* convergence of $\{x^k\}$ to a solution $\bar{x}$ of (2.1) with nonsingular $\Phi'(\bar{x})$. And this is indeed valid, according to Proposition 2.4 and Remark 2.5.

Theorem 2.9. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighborhood of a point $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$. Let $\bar{x}$ be a solution of the equation (2.1). Let $\{J_k\} \subset \mathbf{R}^{n\times n}$ be a sequence of nonsingular matrices, and let a sequence $\{x^k\} \subset \mathbf{R}^n$ be convergent to $\bar{x}$, with (2.31) holding for all k large enough.*

If the rate of convergence of $\{x^k\}$ is superlinear, then condition (2.32) holds.

Conversely, if $\Phi'(\bar{x})$ is a nonsingular matrix and condition (2.32) holds, then the rate of convergence of $\{x^k\}$ is superlinear.

For the basic Newton method (2.3), the Dennis–Moré condition (2.32) is, of course, automatic. The idea of practical quasi-Newton methods is to avoid computation of the exact Jacobian $\Phi'(x^k)$ altogether (since this is often too costly and sometimes simply impossible). The task is to approximate $\Phi'(x^k)$ in some sense, employing information about the values of Φ only. It is important to emphasize that this approximation does not subsume that $\|J_k - \Phi'(x^k)\| \to 0$ as $k \to \infty$ and, in fact, this relation indeed does not hold for specific quasi-Newton methods (in general). The needed approximations must be computed according to some recursive formulas, and without using any information about the derivatives of Φ.

For each k, define

$$s^k = x^{k+1} - x^k, \quad r^k = \Phi(x^{k+1}) - \Phi(x^k).$$

Note that these two vectors are already known by the time when J_{k+1} has to be computed. The goal to satisfy (2.32) can be modeled as the equality

$$r^k = J_{k+1} s^k, \tag{2.33}$$

which is usually referred to as the *quasi-Newton* (or *secant*) *equation*. Indeed, from (2.32) it follows that J_{k+1} should be chosen in such a way that the vector $J_{k+1}s^k$ approximates $\Phi'(x^{k+1})s^k$. At the same time,

$$r^k = \int_0^1 \Phi'(x^k + ts^k)s^k \, dt,$$

and implicitly assuming that the matrix $\Phi'(x^k + ts^k)$ in the right-hand side of the last equality approximates $\Phi'(x^{k+1})$ (which is automatic provided the sequence $\{x^k\}$ converges), the idea to impose the equality (2.33) comes naturally.

Therefore, having at hand a nonsingular matrix J_k and vectors s^k and r^k, it is suggested to choose a matrix J_{k+1} satisfying the quasi-Newton equation (2.33). However, such a choice would clearly be not unique. Having in mind stability considerations, it is natural to additionally require the matrix change $J_{k+1} - J_k$ to be "minimal" in some sense: from one iteration to another, the variation of J_k should not be too large. Different understandings of "minimal" lead to different specific quasi-Newton methods. For instance, consider the case when the correction $J_{k+1} - J_k$ is minimal in the Frobenius norm. Taking into account that linearity of constraints is a CQ, by applying the Lagrange principle (Theorem 1.11), we immediately obtain the following.

Proposition 2.10. *For any elements $s^k \in \mathbf{R}^n \setminus \{0\}$ and $r^k \in \mathbf{R}^n$, and for any matrix $J_k \in \mathbf{R}^{n \times n}$, the unique (global) solution of the problem*

$$\begin{array}{ll} \text{minimize} & \|J - J_k\|_F^2 \\ \text{subject to} & Js^k = r^k \end{array}$$

is given by

$$J_{k+1} = J_k + \frac{(r^k - J_k s_k)s_k^{\mathrm{T}}}{\|s^k\|^2}. \tag{2.34}$$

Proposition 2.10 motivates *Broyden's method*, which is one of the popular quasi-Newton methods for systems of equations: J_0 is an arbitrary nonsingular matrix (e.g., $J_0 = I$), and for each k, the matrix J_{k+1} is computed according to (2.34).

If $n = 1$, formula (2.34) reduces to the following:

$$J_{k+1} = \frac{\Phi(x^{k+1}) - \Phi(x^k)}{x^{k+1} - x^k} = J_k \left(1 - \frac{\Phi(x^{k+1})}{\Phi(x^k)}\right), \tag{2.35}$$

which corresponds to the classical *secant method.*

For an excellent survey of practical quasi-Newton methods for nonlinear equations, see [191].

We proceed to an a priori analysis for the cases when the perturbation term has certain structure. The sequence $\{x^k\}$ is not regarded as given anymore, and the role of the perturbation terms $\{\omega^k\}$ is now primary with respect to $\{x^k\}$.

In many practical algorithms based on (2.11), ω^k depends linearly on x, which is only natural: it is highly desirable to preserve linearity of the iteration system of the pure Newton method in its modifications (so that it remains relatively easy to solve). Let $\omega^k = \omega^k(x) = \Omega_k(x - x^k)$, $x \in \mathbf{R}^n$, where $\Omega_k \in \mathbf{R}^{n\times n}$ for each k. Thus, we consider now the process with the iteration system of the form

$$\Phi(x^k) + (\Phi'(x^k) + \Omega_k)(x - x^k) = 0. \tag{2.36}$$

Note that quasi-Newton methods formally fit this instance of perturbed Newton method by setting $\Omega_k = J_k - \Phi'(x^k)$ (It should be remarked, however, that in what follows we assume that the sequence $\{\Omega_k\}$ is at least bounded, a property which is not automatic for quasi-Newton methods).

Theorem 2.11. *Under the assumptions of Theorem 2.2, it holds that for any fixed $\theta \in (0,\, \|(\Phi'(\bar{x}))^{-1}\|^{-1}/2)$ there exists $\delta > 0$ such that for any sequence of matrices $\{\Omega_k\} \subset \mathbf{R}^{n\times n}$ satisfying*

$$\|\Omega_k\| \leq \theta \quad \forall\, k = 0,\, 1,\, \ldots, \tag{2.37}$$

any $x^0 \in B(\bar{x},\, \delta)$ uniquely defines the iterative sequence $\{x^k\} \subset B(\bar{x},\, \delta)$ such that for each $k = 0,\, 1,\, \ldots,$ the point x^{k+1} satisfies the relation (2.11) with $\omega^k = \Omega_k(x^{k+1} - x^k)$; this sequence converges to $\bar{x}$, and the rate of convergence is (at least) linear. Specifically, there exists $q(\theta) \in (0,\, 1)$ such that $q(\theta) = O(\theta)$ as $\theta \to 0$, and either $x^k = \bar{x}$ for all k large enough, or

$$\limsup_{k\to\infty} \frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|} \leq q(\theta). \tag{2.38}$$

Moreover, the rate of convergence is superlinear if $\{\Omega_k\} \to 0$ as $k \to \infty$. The rate of convergence is quadratic, provided the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$ and $\Omega_k = O(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)$ as $k \to \infty$.

Proof. Employing Lemma A.6, by (2.37) and the restriction on θ we obtain that there exists $\delta > 0$ such that for all $x \in B(\bar{x},\, \delta)$ and all $k = 0,\, 1,\, \ldots,$ it holds that

$$\Phi'(x) + \Omega_k \text{ is nonsingular,}$$

$$\|(\Phi'(x) + \Omega_k)^{-1}\| \le \frac{\|(\Phi'(\bar{x}))^{-1}\|}{1 - (\theta + \|\Phi'(x) - \Phi'(\bar{x})\|)\|(\Phi'(\bar{x}))^{-1}\|}.$$

Thus, for any $k = 0, 1, \ldots$, if $x^k \in B(\bar{x}, \delta)$, then the equation (2.36) has the unique solution x^{k+1}, and

$$\begin{aligned}
\|\omega^k\| &= \|\Omega_k(x^{k+1} - x^k)\| \\
&\le \|\Omega_k\|\|(\Phi'(x^k) + \Omega_k)^{-1}\Phi(x^k)\| \\
&\le \|\Omega_k\|\|(\Phi'(x^k) + \Omega_k)^{-1}(\Phi(\bar{x}) + \Phi'(x^k)(x^k - \bar{x})\| + o(\|x^k - \bar{x}\|) \\
&\le \|\Omega_k\|\|x^k - \bar{x} - (\Phi'(x^k) + \Omega_k)^{-1}\Omega_k(x^k - \bar{x})\| + o(\|x^k - \bar{x}\|) \\
&\le \|\Omega_k\| \left(1 + \frac{\theta\|(\Phi'(\bar{x}))^{-1}\|}{1 - \theta\|(\Phi'(\bar{x}))^{-1}\|}\right) \|x^k - \bar{x}\| + o(\|x^k - \bar{x}\|) \\
&\le \frac{\theta}{1 - \theta\|(\Phi'(\bar{x}))^{-1}\|}\|x^k - \bar{x}\| + o(\|x^k - \bar{x}\|) \qquad (2.39)
\end{aligned}$$

as $x^k \to \bar{x}$, where (2.37) was again taken into account. It follows that

$$\|(\Phi'(\bar{x}))^{-1}\omega^k\| \le \|(\Phi'(\bar{x}))^{-1}\|\|\omega^k\| \le q(\theta)\|x^k - \bar{x}\| + o(\|x^k - \bar{x}\|),$$

where $q(\theta) = \theta\|(\Phi'(\bar{x}))^{-1}\|/(1 - \theta\|(\Phi'(\bar{x}))^{-1}\|)$. Note that by the restriction on θ, it holds that $q(\theta) < 1$, and for any $\varepsilon \in (0, 1 - q(\theta))$ the inequality

$$\|(\Phi'(\bar{x}))^{-1}\omega^k\| \le (q(\theta) + \varepsilon)\|x^k - \bar{x}\|$$

is valid provided δ is small enough. By assertion (a) of Proposition 2.6, this implies the inclusion $x^{k+1} \in B(\bar{x}, \delta)$, perhaps for a smaller δ. It follows that any starting point $x^0 \in B(\bar{x}, \delta)$ uniquely defines the iterative sequence $\{x^k\}$ such that for each $k = 0, 1, \ldots$, the point x^{k+1} satisfies (2.11), and this sequence is contained in $B(\bar{x}, \delta)$ and converges to $\bar{x}$. Moreover, by assertion (b) of Proposition 2.6, the rate of convergence is at least linear; more precisely, either $x^k = \bar{x}$ for all k large enough, or

$$\limsup_{k\to\infty} \frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|} \le q(\theta) + \varepsilon.$$

Since ε can be taken arbitrarily small at the price of reducing δ, and since $\{x^k\}$ converges to $\bar{x}$ (and hence, the tail of $\{x^k\}$ is contained in $B(\bar{x}, \delta)$ no matter how small δ is), the latter implies (2.38).

Finally, by the next to last inequality in (2.39), we obtain that if it holds that $\{\Omega_k\} \to 0$, then (2.20) (and, hence, (2.16)) are valid, and the superlinear convergence rate follows from Proposition 2.4.

Similarly, if $\Omega_k = O(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)$ as $k \to \infty$, then (2.17) holds, and Proposition 2.4 gives the quadratic convergence rate. □

The simplest case of a linear perturbation term is when Ω_k is just constant, i.e., $\Omega_k = \Omega$ for all $k = 0, 1, \ldots$, with some fixed $\Omega \in \mathbf{R}^{n\times n}$. Having in mind faster convergence, it is natural to choose Ω_k in such a way that $\Phi'(x^k)+\Omega_k$ is some approximation of $\Phi'(\bar{x})$. One of the possibilities is $\Omega_k = \Phi'(x^0) - \Phi'(x^k)$ for a given starting point $x^0 \in \mathbf{R}^n$. Assuming that $\Phi'(x^0)$ is nonsingular, this iterative scheme can be written in the form

$$x^{k+1} = x^k - (\Phi'(x^0))^{-1}\Phi(x^k), \quad k = 0, 1, \ldots. \tag{2.40}$$

The iteration cost of the basic Newton method is thus reduced, since the derivative of Φ is computed only once (at x^0) and all the iteration linear systems have the same matrix $\Phi'(x^0)$, which has to be factorized also only once (if factorization is used). From Theorem 2.11, it readily follows that the scheme (2.40) possesses local convergence to a solution with a nonsingular Jacobian. The rate of convergence is only linear, though the closer x^0 is to $\bar{x}$ the higher is the rate of convergence, becoming superlinear in the limit. In practice, one can use a modification of this scheme, with $\Phi'(x^k)$ being computed not only for $k = 0$ but on some subsequence of iterations (but not on all iterations). Such compromise between the basic Newton method and method (2.40) is intended for reducing the iteration costs of the former while increasing the rate of convergence of the latter.

It is also sometimes useful to take $\Omega_k = \Omega(x^k)$, $k = 0, 1, \ldots$, where the mapping $\Omega : \mathbf{R}^n \to \mathbf{R}^{n\times n}$ is such that $\Omega(x) \to 0$ as $x \to \bar{x}$. According to Theorem 2.11, any method of this kind possesses local superlinear convergence to a solution $\bar{x}$ whenever $\Phi'(\bar{x})$ is nonsingular.

A particular construction of $\Omega(\cdot)$ in the case when the explicit expression for $\Phi'(\cdot)$ is available can be based on the following observation: if some terms in the expression for $\Phi'(\cdot)$ are known to vanish at a solution, such terms can be dropped (set to zero) in a Newton-type method from the very beginning.

Consider, for example, an over-determined system

$$\Psi(x) = 0,$$

where $\Psi : \mathbf{R}^n \to \mathbf{R}^m$ is twice differentiable near a solution $\bar{x}$, with its second derivative continuous at $\bar{x}$, and with m generally bigger than n. This problem can be reduced to the standard form (2.1), with the number of equations equal to the number of the unknowns, by setting $\Phi(x) = (\Psi'(x))^{\mathrm{T}}\Psi(x)$, $x \in \mathbf{R}^n$. Moreover, if $\bar{x}$ satisfies the condition $\ker \Psi'(\bar{x}) = \{0\}$ (sufficient for $\bar{x}$ to be an isolated solution; see Proposition 1.32), then $\Phi'(\bar{x}) = (\Psi'(\bar{x}))^{\mathrm{T}}\Psi'(\bar{x})$ is nonsingular. At points that are not solutions, the derivative of Ψ depends not only on the first but also on the second derivative of Φ:

$$\Phi'(x)\xi = (\Psi'(x))^{\mathrm{T}}\Psi'(x)\xi + (\Psi''(x)[\xi])^{\mathrm{T}}\Psi(x), \quad x, \xi \in \mathbf{R}^n,$$

which makes the use of the basic Newton method even more costly in this setting. Fortunately, the last term in the expression for $\Phi'(\cdot)$ vanishes at a solution. Dropping this term, we obtain the *Gauss–Newton method*: for a given $x^k \in \mathbf{R}^n$, the next iterate $x^{k+1} \in \mathbf{R}^n$ is computed as a solution of the iteration system

$$(\Psi'(x^k))^{\mathrm{T}}\Psi(x^k) + (\Psi'(x^k))^{\mathrm{T}}\Psi'(x^k)(x - x^k) = 0, \tag{2.41}$$

which corresponds to (2.36) with the linear perturbation term $\Omega_k = \Omega(x^k)$ defined by

$$\Omega(x)\xi = -(\Psi''(x)[\xi])^{\mathrm{T}}\Psi(x), \quad x,\, \xi \in \mathbf{R}^n.$$

Note that if $n = m$, then this iterative process generates the same iterative sequence as the basic Newton method. Note also that the expression in the left-hand side of (2.41) is precisely the gradient of the quadratic objective function of the following linear least-squares problem:

$$\begin{array}{ll} \text{minimize} & \dfrac{1}{2}\|\Psi(x^k) + \Psi'(x^k)(x - x^k)\|^2 \\ \text{subject to} & x \in \mathbf{R}^n. \end{array}$$

The latter can be solved by special algorithms for linear least-squares problems [208, Sect. 10.2], or by conjugate gradient methods [208, Sect. 5.1], without explicitly computing the product $(\Psi'(x^k))^{\mathrm{T}}\Psi'(x^k)$, which could be too expensive.

Local superlinear convergence of the Gauss–Newton method under the assumption $\ker \Psi'(\bar{x}) = \{0\}$ is ensured by Theorem 2.11, according to the discussion above.

Even though keeping the iteration system linear is certainly a reasonable approach, it will be seen below that there exist some practical algorithms (for constrained optimization) fitting the perturbed Newton method framework for which the dependence of the perturbation term on the variables is not necessarily linear. Instead, it satisfies some smoothness-like assumptions, still allowing an a priori analysis via the use of the implicit function theorem. One such example is the linearly constrained augmented Lagrangian method for optimization, discussed in Sect. 4.1.2. This motivates the following results, dealing with nonlinear dependence of perturbations on the problem variables.

Theorem 2.12. *Under the hypotheses of Theorem 2.2, let* $\omega : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$ *satisfy the following assumptions:*

$$\omega(x,\, \xi^1) - \omega(x,\, \xi^2) = o(\|\xi^1 - \xi^2\|) \tag{2.42}$$

as $\xi^1,\, \xi^2 \in \mathbf{R}^n$ *tend to* 0*, uniformly in* $x \in \mathbf{R}^n$ *close enough to* $\bar{x}$*, and there exists* $\theta \in (0,\, \|(\Phi'(\bar{x}))^{-1}\|^{-1})$ *such that the inequality*

$$\|\omega(x,\, 0)\| \le \theta\|x - \bar{x}\| \tag{2.43}$$

holds for all $x \in \mathbf{R}^n$ *close enough to* $\bar{x}$*.*

Then there exists $\delta > 0$ such that any starting point $x^0 \in \mathbf{R}^n$ close enough to $\bar{x}$ uniquely defines the iterative sequence $\{x^k\} \subset \mathbf{R}^n$ such that x^{k+1} satisfies (2.11) with $\omega^k = \omega(x^k,\, x^{k+1} - x^k)$ for each $k = 0,\, 1,\, \ldots,$ and $\|x^{k+1} - x^k\| \le \delta$; this sequence converges to $\bar{x}$, and the rate of convergence is (at least) linear. Specifically, there exists $q(\theta) \in (0,\, 1)$ such that (2.38) holds, and $q(\theta) = O(\theta)$ as $\theta \to 0$.

Moreover, the rate of convergence is superlinear if

$$\omega(x,\, 0) = o(\|x - \bar{x}\|) \tag{2.44}$$

as $x \to \bar{x}$. The rate of convergence is quadratic provided the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$ and

$$\omega(x,\, \xi) = O(\|\xi\|^2 + \|x - \bar{x}\|^2) \tag{2.45}$$

as $x \to \bar{x}$ and $\xi \to 0$.

Proof. Define the mapping $\Psi : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$,

$$\Psi(x,\, \xi) = \Phi(x) + \Phi'(x)\xi + \omega(x,\, \xi).$$

By the assumptions (2.42) and (2.43), the implicit function theorem (Theorem 1.22) is applicable to this mapping at $(x,\, \xi) = (\bar{x},\, 0)$ (here, x is regarded as a parameter). Hence, there exist $\delta > 0$ and $\tilde{\delta} > 0$ such that for each $x \in B(\bar{x},\, \tilde{\delta})$ the equation

$$\Psi(x,\, \xi) = 0$$

has the unique solution $\xi(x) \in B(0,\, \delta)$, and this solution satisfies the estimate

$$\|\xi(x)\| = O(\|\Psi(x,\, 0)\|) = O(\|\Phi(x)\|) + O(\|\omega(x,\, 0)\|) = O(\|x - \bar{x}\|) \tag{2.46}$$

as $x \to \bar{x}$. Then for any $x^k \in B(\bar{x},\, \tilde{\delta})$, the point $x^{k+1} = x^k + \xi(x^k)$ is the only one in $B(x^k,\, \delta)$ satisfying (2.11) with $\omega^k = \omega(x^k,\, x^{k+1} - x^k)$. Furthermore,

$$\begin{aligned}
\|\omega^k\| &= \|\omega(x^k,\, \xi(x^k))\| \\
&\le \|\omega(x^k,\, \xi(x^k)) - \omega(x^k,\, 0)\| + \|\omega(x^k,\, 0)\| \\
&= \|\omega(x^k,\, 0)\| + o(\|\xi(x^k)\|) \\
&= \|\omega(x^k,\, 0)\| + o(\|x^k - \bar{x}\|) \\
&\le \theta\|x^k - \bar{x}\| + o(\|x^k - \bar{x}\|)
\end{aligned} \tag{2.47}$$

as $x^k \to \bar{x}$, where (2.42) and (2.43) were employed again. It follows that

$$\|(\Phi'(\bar{x}))^{-1}\omega^k\| \le \|(\Phi'(\bar{x}))^{-1}\|\|\omega^k\| \le q(\theta)\|x^k - \bar{x}\| + o(\|x^k - \bar{x}\|),$$

where $q(\theta) = \theta\|(\Phi'(\bar{x}))^{-1}\|$. Note that by the restriction on θ, it holds that $q(\theta) < 1$.

The rest of the proof almost literally repeats the corresponding part of the proof of Theorem 2.11. In particular, convergence follows from Proposition 2.6.

The superlinear convergence rate under the assumption (2.44) follows by the third equality in (2.47), and by Proposition 2.4. Moreover, assuming (2.45), the estimate (2.47) can be sharpened as follows:

$$\omega^k = \omega(x^k,\, \xi(x^k)) = O(\|\xi(x^k)\|^2 + \|x^k - \bar{x}\|^2) = O(\|x^k - \bar{x}\|^2)$$

as $x^k \to \bar{x}$, where the last equality is by (2.46). Proposition 2.4 now gives quadratic convergence rate, provided the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$. $\square$

Note that the case discussed above when $\omega^k = \omega^k(x) = \Omega(x^k)(x - x^k)$, $k = 0,\, 1,\, \ldots$, with some mapping $\Omega : \mathbf{R}^n \to \mathbf{R}^{n\times n}$ such that $\Omega(x) \to 0$ as $x \to \bar{x}$ (in particular, the Gauss–Newton method), can be treated both by Theorem 2.11 or 2.12. More interesting examples of the use of Theorem 2.12 will be provided below (see Sects. 4.1, 4.2).

The next result is, in a sense, intermediate between a priori and a posteriori characterizations of perturbed Newton method. We present it here mainly because of the conceptual importance of this kind of analysis for Newton-type methods in the setting of variational problems, where the existence of solutions of subproblems can be guaranteed in general only under rather strong assumptions; see Sect. 3.1. In such cases, it may be useful just to *assume* solvability of subproblems, having in mind that this can be verifiable separately, for more specific algorithms and/or problem classes.

For this analysis, it is natural to replace (2.11) by the generalized equation (GE)

$$\Phi(x^k) + \Phi'(x^k)(x - x^k) + \Omega(x^k,\, x - x^k) \ni 0, \tag{2.48}$$

with a multifunction Ω from $\mathbf{R}^n \times \mathbf{R}^n$ to the subsets of $\mathbf{R}^n$.

Theorem 2.13. *Under the assumptions of Theorem 2.2, let Ω be a multifunction from $\mathbf{R}^n \times \mathbf{R}^n$ to the subsets of $\mathbf{R}^n$, satisfying the following assumptions: for each $x \in \mathbf{R}^n$ close enough to $\bar{x}$, the GE*

$$\Phi(x) + \Phi'(x)\xi + \Omega(x,\, \xi) \ni 0 \tag{2.49}$$

has a solution $\xi(x)$ such that $\xi(x) \to 0$ as $x \to \bar{x}$, and there exist $\theta_1,\, \theta_2 \geq 0$ such that $2\theta_1 + \theta_2 \leq \|(\Phi'(\bar{x}))^{-1}\|^{-1}$ and the inequality

$$\|\Phi(x) + \Phi'(x)\xi\| \leq \theta_1\|\xi\| + \theta_2\|x - \bar{x}\| \tag{2.50}$$

holds for all $x \in \mathbf{R}^n$ close enough to $\bar{x}$ and all $\xi \in \mathbf{R}^n$ close enough to zero, satisfying (2.49).

Then there exists $\delta > 0$ such that for any starting point $x^0 \in \mathbf{R}^n$ close enough to $\bar{x}$, there exists a sequence $\{x^k\} \subset \mathbf{R}^n$ such that x^{k+1} is a solution of the GE (2.48) for each $k = 0,\, 1,\, \ldots$, satisfying

$$\|x^{k+1} - x^k\| \leq \delta; \tag{2.51}$$

any such sequence converges to $\bar{x}$, and the rate of convergence is (at least) linear. Specifically, there exists $q(\theta) \in (0, 1)$, $\theta = \theta_1 + \theta_2$, such that (2.38) holds, and $q(\theta) = O(\theta)$ as $\theta \to 0$.

Moreover, the rate of convergence is superlinear if (2.50) can be replaced by the stronger condition

$$\Phi(x) + \Phi'(x)\xi = o(\|\xi\| + \|x - \bar{x}\|) \tag{2.52}$$

as $x \to \bar{x}$ and $\xi \to 0$. The rate of convergence is quadratic provided the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$, and provided (2.50) can be replaced by the even stronger condition

$$\Phi(x) + \Phi'(x)\xi = O(\|\xi\|^2 + \|x - \bar{x}\|^2). \tag{2.53}$$

Proof. Under the assumptions of the theorem, there exist $\delta > 0$ and $\tilde{\delta} > 0$ such that for any $x^k \in B(\bar{x}, \tilde{\delta})$, there exists $x^{k+1} \in B(x^k, \delta)$ (specifically, $x^{k+1} = x^k + \xi(x^k)$) satisfying (2.48). Assuming that δ and $\tilde{\delta}$ are small enough, for any such x^{k+1}, by setting $\omega^k = -\Phi(x^k) - \Phi'(x^k)(x^{k+1} - x^k)$, we obtain that (2.11) holds with $x = x^{k+1}$, and

$$\|\omega^k\| \leq \theta_1 \|x^{k+1} - x^k\| + \theta_2 \|x^k - \bar{x}\|,$$

where (2.50) was employed. It follows that

$$\begin{aligned} \|(\Phi'(\bar{x}))^{-1}\omega^k\| &\leq \|(\Phi'(\bar{x}))^{-1}\|\|\omega^k\| \\ &\leq q_1(\theta_1)\|x^{k+1} - x^k\| + q_2(\theta_2)\|x^k - \bar{x}\|, \end{aligned}$$

where $q_j(\theta_j) = \theta_j \|(\Phi'(\bar{x}))^{-1}\|$, $j = 1, 2$, satisfy $2q_1(\theta_1) + q_2(\theta_2) < 1$.

The rest of the proof again almost literally repeats the corresponding part of the proof of Theorem 2.11. Convergence follows by Proposition 2.6, and (2.38) holds with $q(\theta) = (q_1(\theta_1) + q_2(\theta_2))/(1 - q_1(\theta_1))$. The superlinear/quadratic convergence rate under the corresponding additional assumptions follows by Proposition 2.4. □

We complete this section with a brief discussion of the case when $\Phi'(\bar{x})$ is not necessarily nonsingular. Such cases will be treated in detail later in this book for (generalized) equations possessing some special (primal-dual) structure, arising from optimization and variational problems. For general equations without any special structure, the behavior of Newton-type methods near solutions with singular Jacobians, as well as various modifications of these methods intended for preserving the efficiency despite singularity, was studied in [151, 152]. Here, we limit the discussion to some comments which may give an initial understanding of the effect of singularity.

Consider the scalar equation

$$x^s = 0,$$

where $s \geq 2$ is an integer parameter. The Newton method iterations for this equation are given by $x^{k+1} = (1-1/s)x^k$, and the sequence $\{x^k\}$ converges to the unique solution $\bar{x} = 0$ from any starting point, but the rate of convergence is only linear. This happens because $\bar{x}$ is a singular solution: the derivative at $\bar{x}$ is zero. At the same time, if we modify the Newton method by introducing the stepsize parameter equal to s, the method hits the exact solution in one step, for any starting point x^0.

More generally, the following fact was established in [242]. Let a function $\Phi : \mathbf{R} \to \mathbf{R}$ be s times differentiable at $\bar{x} \in \mathbf{R}$, $s \geq 2$, where $\bar{x}$ is a root of multiplicity s of the equation (2.1), i.e.,

$$\Phi(\bar{x}) = \Phi'(\bar{x}) = \ldots = \Phi^{(s-1)}(\bar{x}) = 0, \quad \Phi^{(s)}(\bar{x}) \neq 0.$$

Then the Newton method iterates locally converge to $\bar{x}$ at a linear rate, while the method modified by introducing the stepsize parameter equal to s gives the superlinear convergence rate.

2.1.2 Newton Method for Unconstrained Optimization

Consider now the unconstrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathbf{R}^n, \end{array} \tag{2.54}$$

with a twice differentiable objective function $f : \mathbf{R}^n \to \mathbf{R}$. Stationary points of this problem are characterized by the equation (2.1) with $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ being the *gradient mapping* of f:

$$\Phi(x) = f'(x).$$

Thus, one strategy to compute stationary points of the optimization problem (2.54) is to apply some Newton-type method to the equation (2.1) with Φ being the gradient of f.

In the case of the basic Newton method for (2.54), given $x^k \in \mathbf{R}^n$, the next iterate x^{k+1} is computed as a solution of the linear system

$$f'(x^k) + f''(x^k)(x - x^k) = 0. \tag{2.55}$$

Assuming that the Hessian $f''(x^k)$ is nonsingular, the Newton method can be written in the form of the explicit iterative scheme

$$x^{k+1} = x^k - (f''(x^k))^{-1} f'(x^k), \quad k = 0, 1, \ldots .$$

This iteration allows for the following interpretation that puts to the foreground the optimization nature of the original problem. Near the current iterate x^k, the objective function f is naturally approximated by its second-order expansion or, in other words, the original problem (2.54) is approximated by the following subproblem:

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, x - x^k\rangle + \dfrac{1}{2}\langle f''(x^k)(x - x^k),\, x - x^k\rangle \\ \text{subject to} & x \in \mathbf{R}^n. \end{array} \tag{2.56}$$

Since (2.55) is precisely the equation defining stationary points of (2.56), the basic *Newton method* for unconstrained optimization can be presented as follows.

Algorithm 2.14 Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $f'(x^k) = 0$, stop.
2. Compute $x^{k+1} \in \mathbf{R}^n$ as a stationary point of problem (2.56).
3. Increase k by 1 and go to step 1.

Local convergence result for Newton method for unconstrained optimization follows immediately from Theorem 2.2 on local convergence of Newton method for equations.

Theorem 2.15. *Let a function $f : \mathbf{R}^n \to \mathbf{R}$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its Hessian being continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (2.54), and assume that this point satisfies the SOSC*

$$\langle f''(\bar{x})\xi,\, \xi\rangle > 0 \quad \forall\, \xi \in \mathbf{R}^n \setminus \{0\} \tag{2.57}$$

(thus, according to Theorem 1.9, $\bar{x}$ is a strict local solution of problem (2.54)).

Then any starting point $x^0 \in \mathbf{R}^n$ close enough to $\bar{x}$ uniquely defines the iterative sequence of Algorithm 2.14; this sequence converges to $\bar{x}$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the Hessian of f is locally Lipschitz-continuous with respect to $\bar{x}$.

As one specificity of Newton method for optimization, let us mention that a Hessian of a twice differentiable function is a symmetric matrix. Under the assumptions of Theorem 2.15, perhaps the most natural general strategy for solving the iteration system (2.55) appears to be the so-called Cholesky factorization, which provides the LL^{T}-decomposition of a positive definite symmetric matrix (L is a lower triangular matrix with positive diagonal elements) at a price of $n^3/6$ multiplications and the same amount of additions (see, e.g., [100], [261, Lecture 23], [103, Sect. 4.2]). More details on special tools of numerical linear algebra for iteration systems arising in optimization can be found, e.g., in [29, 208].

Note also that the assertion of Theorem 2.15 remains valid if the SOSC (2.57) is replaced by the weaker assumption that $f''(\bar{x})$ is nonsingular. In this respect, Newton method does not distinguish local minimizers from other stationary points of the problem (including the maximizers).

The main advantage of Newton method is its high convergence rate (superlinear, under natural assumptions). However, the basic Newton method has also serious drawbacks, which we discuss next.

First, each step of the Newton method requires computing the Hessian and solving the corresponding linear system, which can be too costly, or simply impossible in some applications. Regarding this issue, we note that perturbed Newton methods for equations discussed in Sect. 2.1.1 can be directly adapted for unconstrained optimization. Indeed, all these methods can be applied to the equation defined by the gradient of f. This may help to reduce the iteration costs significantly. One important example is the class of quasi-Newton methods for unconstrained optimization, discussed in Sect. 2.2.

The second inevitable drawback of pure Newton-type methods is that they possess only local convergence: in all results presented above, a starting point close enough to a solution is required. An iterative sequence of Newton method defined by an inappropriate starting point may not have stationary points of the problem among its accumulations points. In fact, this may happen even in the case of a strongly convex objective function (so that its stationary point is unique, and it is the unique global minimizer).

Example 2.16. Consider the function $f : \mathbf{R} \to \mathbf{R}$,

$$f(x) = \begin{cases} -\dfrac{x^4}{4\sigma^3} + \left(1 + \dfrac{3}{\sigma}\right)\dfrac{x^2}{2} & \text{if } |x| \leq \sigma, \\ \dfrac{x^2}{2} + 2|x| - \dfrac{3\sigma}{4} & \text{if } |x| > \sigma, \end{cases}$$

where $\sigma > 0$ is a parameter. It can be easily checked that for any such σ, the function f is twice continuously differentiable and strongly convex on $\mathbf{R}$, and problem (2.54) with this objective function has the unique stationary point $\bar{x} = 0$. In particular, $f''(\bar{x}) = 1 + 3/\sigma > 0$, and all the assumptions of Theorem 2.15 are satisfied. Take $x^0 = \sigma$. The corresponding iterative sequence $\{x^k\}$ of Algorithm 2.14 is then given by $x^k = 2(-1)^k$, $k = 1, 2, \ldots$, and $\bar{x}$ is not an accumulation point of $\{x^k\}$, no matter how small σ is.

Strategies for globalization of convergence of Newton-type methods for unconstrained optimization is the subject of the rest of this chapter. In particular, linesearch quasi-Newton methods (to be discussed in Sect. 2.2) serve not only for reducing the iteration costs but also for enforcing global convergence of Newton-type methods. (This is the main reason why we present quasi-Newton methods for unconstrained optimization in the context of linesearch methods.)

2.2 Linesearch Methods, Quasi-Newton Methods

In this section, we consider the unconstrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathbf{R}^n, \end{array} \tag{2.58}$$

with a differentiable objective function $f : \mathbf{R}^n \to \mathbf{R}$. One of the most natural approaches to solving (2.58) is the following. For the given iterate, compute a descent direction for f at this point, and make a step of some length along this direction so that the value of f is (sufficiently) reduced. Repeat the procedure for the obtained new iterate, etc. We refer to methods of this kind as descent methods. Evidently, efficiency of any such method depends on two choices: that of the descent direction, and that of the stepsize. Perhaps the most practically important example of good choices for both is the class of linesearch quasi-Newton methods.

2.2.1 Descent Methods

We start with a formal definition of descent directions.

Definition 2.17. A vector $p \in \mathbf{R}^n$ is said to be a *descent direction* for the function $f : \mathbf{R}^n \to \mathbf{R}$ at $x \in \mathbf{R}^n$ if for all $t > 0$ small enough it holds that $f(x + tp) < f(x)$.

The set of all descent directions for f at $x \in \mathbf{R}^n$ is a cone, which will be denoted by $\mathcal{D}_f(x)$. Therefore, $p \in \mathcal{D}_f(x)$ if and only if any sufficiently small displacement of x in the direction p results in a reduction of the function value with respect to $f(x)$. The next statement is elementary.

Lemma 2.18. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be differentiable at $x \in \mathbf{R}^n$.*
Then the following assertions are valid:

(a) For any $p \in \mathcal{D}_f(x)$ it holds that $\langle f'(x), p\rangle \leq 0$.
(b) If for $p \in \mathbf{R}^n$ it holds that $\langle f'(x), p\rangle < 0$, then $p \in \mathcal{D}_f(x)$.

The class of *descent methods* is then given by iterative schemes of the form

$$x^{k+1} = x^k + \alpha_k p^k, \quad p^k \in \mathcal{D}_f(x^k), \quad \alpha_k > 0, \quad k = 0, 1, \ldots, \tag{2.59}$$

where the *stepsize parameters* $\alpha_k > 0$ are chosen in such a way that, at least,

$$f(x^{k+1}) < f(x^k). \tag{2.60}$$

That is, the sequence $\{f(x^k)\}$ must be monotonically decreasing. (If $\mathcal{D}_f(x^k) = \emptyset$ or if an element of $\mathcal{D}_f(x^k)$ cannot be computed by the prescribed tools, the process is terminated.) Note that the inclusion $p^k \in \mathcal{D}_f(x^k)$ implies that the inequality (2.60) holds for all $\alpha_k > 0$ small enough. However, (2.60) is obviously not enough to guarantee convergence: the reduction property must be appropriately quantified.

As mentioned above, a specific descent method is characterized by a specific rule for choosing descent directions, and a specific procedure for computing the appropriate values of the stepsize parameter. Procedures for choosing a stepsize are based on exploring the restriction of the objective function f to the ray spanned by p^k, with its origin at x^k. For this reason, such procedures are usually called *linesearch*. It is interesting to point out the following common feature of optimization algorithms: a choice of search directions p^k is typically based on some approximate model of the objective function f (see below), while linesearch procedures are normally performed for f itself.

By Lemma 2.18, if $f'(x^k) \neq 0$, then one can always take the descent direction $p^k = -f'(x^k)$. The corresponding descent methods (sometimes called *steepest descent methods*) are easy to implement, and their convergence and rate of convergence properties can be fully characterized theoretically. However, such methods are completely impractical: this choice of descent directions usually turns out to be extremely inefficient.

Much more practical descent methods are obtained within the following more general framework. Given $x^k \in \mathbf{R}^n$ take $p^k = -Q_k f'(x^k)$, where $Q_k \in \mathbf{R}^{n\times n}$ is a symmetric positive definite matrix. The matrices, of course, must be chosen in some clever way. Good choices of Q_k will be discussed in Sect. 2.2.2. Right now, we note only that by Lemma 2.18, if $f'(x^k) \neq 0$, then $p^k = -Q_k f'(x^k)$ with a positive definite Q_k is clearly a descent direction for f at x^k, since

$$\langle f'(x^k),\, p^k\rangle = -\langle Q_k f'(x^k),\, f'(x^k)\rangle < 0. \tag{2.61}$$

The "limiting," in some sense, choices for Q_k are $Q_k = (f''(x^k))^{-1}$ corresponding to the (expensive) Newton direction (see Sect. 2.1.2) and $Q_k = I$ corresponding to the (cheap) steepest descent direction. We note that the latter can still be useful sometimes, but only as a "last resort," in those cases when for some reasons more sophisticated options fail.

We next discuss the most important linesearch procedures, assuming that for a given iterate x^k a direction $p^k \in \mathcal{D}_f(x^k)$ is already chosen and fixed. It may seem natural to take the stepsize parameter $\alpha_k > 0$ as a global minimizer of $f(x^k + \alpha p^k)$ over all $\alpha \geq 0$. This *exact linesearch rule* is, formally, ideal: it provides the maximal possible progress in decreasing f along the given direction. If f is a quadratic function with a positive definite Hessian, then such α_k is given by an explicit formula. But beyond the quadratic case, exact linesearch is too expensive and usually impossible anyway. Moreover, even searching for a local minimizer of $f(x^k + \alpha p^k)$ (or, e.g., for the local minimizer closest to zero) is usually not worthwhile—afterall, the eventual goal is to minimize f on the entire space rather than on the given ray. For this reason, much cheaper inexact linesearch rules are used in practice. These rules ensure *sufficient decrease* of the objective function value, instead of searching for (local or global) minimizers of f along the given descent direction.

Armijo rule. Choose the parameters $C > 0$, $\sigma \in (0, 1)$ and $\theta \in (0, 1)$. Set $\alpha = C$.

1. Check the inequality

$$f(x^k + \alpha p^k) \leq f(x^k) + \sigma\alpha\langle f'(x^k), p^k\rangle. \tag{2.62}$$

2. If (2.62) does not hold, replace α by $\theta\alpha$ and go to step 1. Otherwise, set $\alpha_k = \alpha$.

Thus, α_k is the first α of the form $C\theta^j$, $j = 0, 1, \ldots,$ satisfying (2.62) (the needed value is computed by a backtracking procedure starting with the initial trial value C). The quantity $\alpha\langle f'(x^k), p^k\rangle$ in the right-hand side of (2.62) plays the role of "predicted" (by the linear model of f) reduction of the objective function value for the step of length α in the direction p^k. Therefore, inequality (2.62) means that the actual reduction must be no less than a given fraction (defined by the choice of $\sigma \in (0, 1)$) of the "predicted" reduction. Armijo linesearch is illustrated in Fig. 2.3.

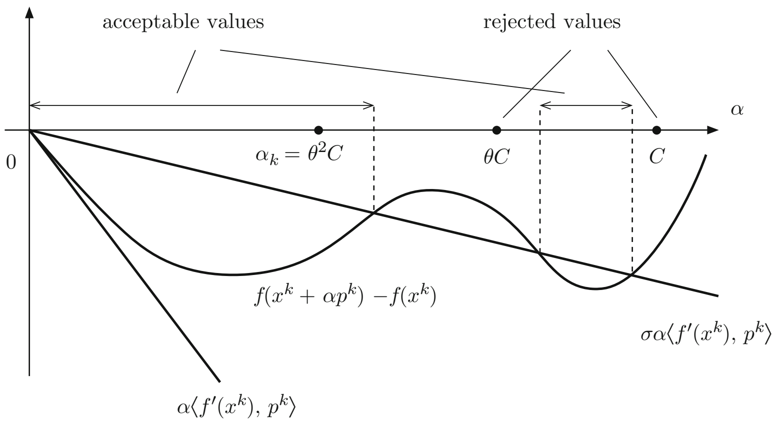

Fig. 2.3 Armijo rule

The next lemma demonstrates that if p^k satisfies the sufficient condition for a descent direction stated in Lemma 2.18, i.e., if

$$\langle f'(x^k), p^k\rangle < 0, \tag{2.63}$$

then the backtracking procedure in the Armijo rule is finite.

Lemma 2.19. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be differentiable at $x^k \in \mathbf{R}^n$.*

Then for any $p^k \in \mathbf{R}^n$ satisfying (2.63), inequality (2.62) holds for all $\alpha > 0$ small enough.

Proof. It holds that

$$\begin{aligned} f(x^k + \alpha p^k) - f(x^k) &= \langle f'(x^k),\, \alpha p^k\rangle + o(\alpha) \\ &= \sigma\alpha\langle f'(x^k),\, p^k\rangle + (1-\sigma)\alpha\langle f'(x^k),\, p^k\rangle + o(\alpha) \\ &= \sigma\alpha\langle f'(x^k),\, p^k\rangle + \alpha\left((1-\sigma)\langle f'(x^k),\, p^k\rangle + \frac{o(\alpha)}{\alpha}\right) \\ &\leq \sigma\alpha\langle f'(x^k),\, p^k\rangle, \end{aligned}$$

because $(1-\sigma)\langle f'(x^k),\, p^k\rangle + o(\alpha)/\alpha < 0$ for any $\alpha > 0$ small enough. □

Evidently, if (2.63) holds, then choosing α_k according to the Armijo rule guarantees the descent property (2.60). Moreover, the inequality (2.62) with $\alpha = \alpha_k$ gives a quantitative estimate of by how much $f(x^{k+1})$ is smaller than $f(x^k)$, and this estimate (unlike (2.60)) is sufficient for establishing convergence under natural assumptions. However, convergence proof is significantly simplified when one can show that the backtracking is finite uniformly with respect to k, i.e., when α_k is separated from zero by some threshold independent of k.

Lemma 2.20. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be differentiable on $\mathbf{R}^n$, and suppose that its gradient is Lipschitz-continuous on $\mathbf{R}^n$ with constant $L > 0$.*

Then for any $x^k \in \mathbf{R}^n$ and $p^k \in \mathbf{R}^n$ satisfying (2.63), the inequality (2.62) holds for all $\alpha \in (0,\, \bar{\alpha}_k]$, where

$$\bar{\alpha}_k = \frac{2(\sigma-1)\langle f'(x^k),\, p^k\rangle}{L\|p^k\|^2} > 0. \tag{2.64}$$

Proof. By Lemma A.11, for all $\alpha > 0$ it holds that

$$f(x^k + \alpha p^k) - f(x^k) - \langle f'(x^k),\, \alpha p^k\rangle \leq \frac{L}{2}\alpha^2\|p^k\|^2.$$

Hence, for all $\alpha \in (0,\, \bar{\alpha}_k]$ we have that

$$\begin{aligned} f(x^k + \alpha p^k) - f(x^k) &\leq \langle f'(x^k),\, \alpha p^k\rangle + \frac{L}{2}\alpha^2\|p^k\|^2 \\ &= \alpha\left(\langle f'(x^k),\, p^k\rangle + \frac{L}{2}\alpha\|p^k\|^2\right) \\ &\leq \sigma\alpha\langle f'(x^k),\, p^k\rangle, \end{aligned}$$

where the last inequality follows from (2.64). □

Lemma 2.21. *Under the assumptions of Lemma 2.20, let* $\{Q_k\} \subset \mathbf{R}^{n\times n}$ *be a sequence of symmetric matrices satisfying*

$$\langle Q_k \xi,\, \xi\rangle \geq \gamma \|\xi\|^2 \quad \forall\, \xi \in \mathbf{R}^n, \quad \|Q_k\| \leq \Gamma \quad \forall\, k, \tag{2.65}$$

with some $\gamma > 0$ *and* $\Gamma > 0$.

Then there exists a constant $c > 0$ *such that for any point* $x^k \in \mathbf{R}^n$ *and for* $p^k = -Q_k f'(x^k)$, *the value* α_k *obtained by the Armijo rule satisfies*

$$\alpha_k \geq c. \tag{2.66}$$

Proof. By (2.65), we obtain that

$$\frac{\langle f'(x^k),\, p^k\rangle}{\|p^k\|^2} = -\frac{\langle f'(x^k),\, Q_k f'(x^k)\rangle}{\|Q_k f'(x^k)\|^2} \leq -\frac{\gamma}{\Gamma^2}.$$

Hence, according to (2.64),

$$\bar{\alpha}_k \geq \frac{2(1-\sigma)\gamma}{L\Gamma^2} > 0.$$

The needed assertion now follows from Lemma 2.20. □

The Armijo rule is simple, clear, and easy to implement. Convergence results presented below refer to this rule. However, more sophisticated linesearch techniques, with better theoretical and practical properties, are often used in practice.

Goldstein rule consists of choosing the stepsize parameter satisfying the inequalities

$$\sigma_1 \leq \frac{f(x^k + \alpha p^k) - f(x^k)}{\alpha \langle f'(x^k),\, p^k\rangle} \leq \sigma_2, \tag{2.67}$$

with fixed $0 < \sigma_1 < \sigma_2 < 1$.

The first inequality in (2.67) is just the Armijo inequality (2.62) with $\sigma = \sigma_1$; it guarantees sufficient decrease of the objective function. Recall that according to Lemma 2.19, this inequality holds for all $\alpha > 0$ small enough. By contrast, the second inequality in (2.67) is evidently violated for all $\alpha > 0$ close to zero. The reason for introducing the second inequality is precisely to avoid stepsize parameters that are too small. The idea is to take larger steps, i.e., prevent the method from slowing down. Goldstein linesearch is illustrated in Fig. 2.4.

Wolfe rule is another realization of the same idea, but instead of (2.67) it employs the inequalities

$$f(x^k + \alpha p^k) \leq f(x^k) + \sigma_1 \alpha \langle f'(x^k),\, p^k\rangle, \tag{2.68}$$

$$\langle f'(x^k + \alpha p^k),\, p^k\rangle \geq \sigma_2 \langle f'(x^k),\, p^k\rangle. \tag{2.69}$$

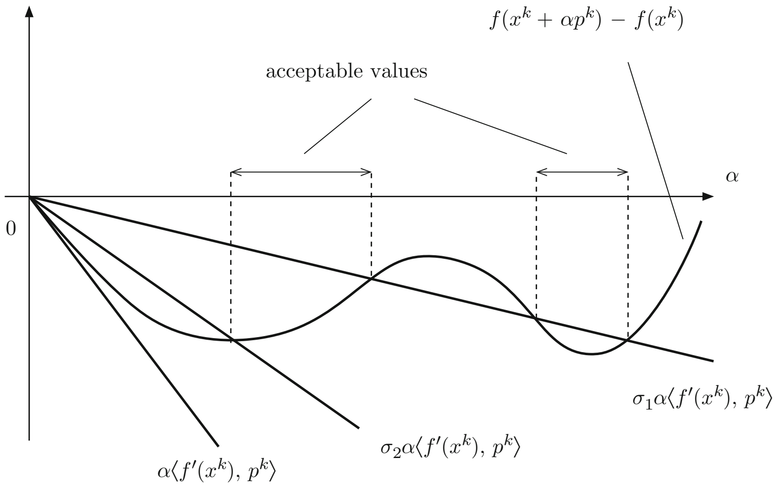

Fig. 2.4 Goldstein rule

Again, (2.68) is the Armijo inequality (2.62) with $\sigma = \sigma_1$. Evidently, analogously to (2.67), the second inequality in (2.69) also does not allow stepsize values that are too small. Note that it involves the gradient of f not only at x^k but also at the trial points $x^k + \alpha p^k$, which entails some additional computational cost. However, when computation of the gradient is not too expensive, the Wolfe rule is often regarded as the most efficient among currently known linesearch options. One important property of this rule is related to quasi-Newton methods; see Sect. 2.2.2. Wolfe linesearch is illustrated in Fig. 2.5.

We next give a simple algorithmic implementation of the Wolfe rule. (The Goldstein rule can be implemented along the same lines.) Let $0 < \sigma_1 < \sigma_2 < 1$ be fixed. Set $c = C = 0$, and choose an initial trial value $\alpha > 0$.

1. Check the inequalities (2.68) and (2.69). If both do hold, go to step 6.
2. If (2.68) does not hold, set $C = \alpha$, and go to step 5.
3. If (2.69) does not hold, set $c = \alpha$.
4. If $C = 0$, choose a new trial value $\alpha > c$ ("extrapolation"), and go to step 1.
5. Choose a new trial value $\alpha \in (c, C)$ ("interpolation"), and go to step 1.
6. Set $\alpha_k = \alpha$.

Violation of (2.68) basically means that the current trial value α is "too large," while violation of (2.69) means that it is "too small." The procedure just described works as follows. Extrapolation steps are performed first, until C becomes positive. Once this happened, interpolation steps are performed. In the course of interpolation C may only decrease, remaining positive, while c may only increase, staying smaller than C.

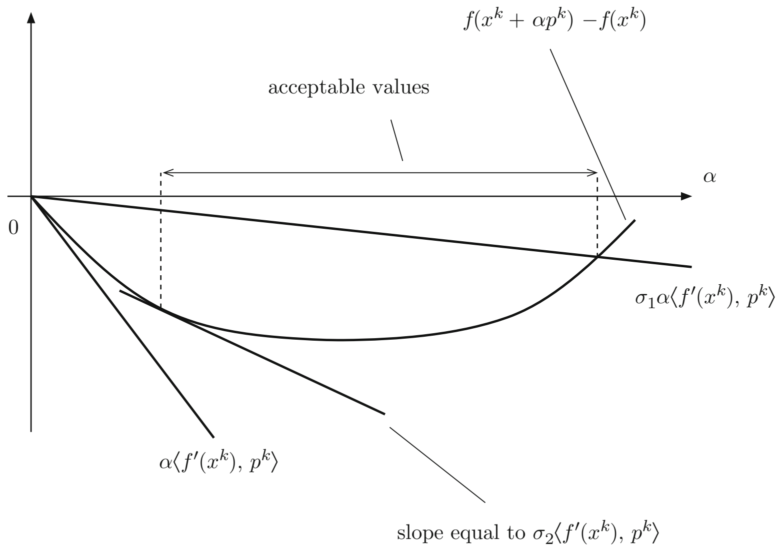

Fig. 2.5 Wolfe rule

Extrapolation and interpolation in the presented procedure can be organized in many ways. For example, one can fix $\theta_1 > 1$, $\theta_2 \in (0,\, 1)$, and replace α by $\theta_1\alpha$ in the case of extrapolation, and set $\alpha = (1-\theta_2)c+\theta_2 C$ in the case of interpolation. More sophisticated options are discussed, e.g., in [29, Chap. 3]. From the theoretical viewpoint, it is important to guarantee the following property: in the case of infinite number of extrapolation steps c must be increasing to infinity, while in the case of infinite number of interpolation steps $(C - c)$ must be tending to zero.

Lemma 2.22. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be continuously differentiable and bounded below on $\mathbf{R}^n$.*

Then for any $x^k \in \mathbf{R}^n$ and $p^k \in \mathbf{R}^n$ satisfying (2.63), the procedure implementing the Wolfe rule such that $c \to +\infty$ in the case of infinite number of extrapolation steps and $(C - c) \to 0$ in the case of infinite number of interpolation steps, is finite.

Proof. Suppose first that there is an infinite number of extrapolation steps. Then the procedure generates an increasing to infinity sequence of values of c, and for each of these values it holds that

$$f(x^k + cp^k) \le f(x^k) + \sigma_1 c\langle f'(x^k),\, p^k\rangle. \tag{2.70}$$

But according to inequality (2.63), the latter contradicts the assumption that f is bounded below. Therefore, the number of extrapolation steps is finite.

Suppose now that the number of interpolation steps is infinite. Then the monotone sequences of values of c and of C converge to a common limit $\bar{\alpha}$. The elements of the first sequence satisfy (2.70) and

$$\langle f'(x^k + cp^k),\, p^k\rangle < \sigma_2 \langle f'(x^k),\, p^k\rangle, \tag{2.71}$$

while the elements of the second sequence satisfy

$$f(x^k + Cp^k) > f(x^k) + \sigma_1 C \langle f'(x^k),\, p^k\rangle. \tag{2.72}$$

By passing onto the limit in (2.70) and (2.72), we obtain the equality

$$f(x^k + \bar{\alpha}p^k) = f(x^k) + \sigma_1 \bar{\alpha} \langle f'(x^k),\, p^k\rangle. \tag{2.73}$$

Taking into account (2.72) and monotone decrease of the values of C, it follows that these values always remain bigger than $\bar{\alpha}$. Employing (2.73), we can rewrite inequality (2.72) in the form

$$\begin{aligned} f(x^k + Cp^k) &> f(x^k) + \sigma_1 \bar{\alpha} \langle f'(x^k),\, p^k\rangle + (C - \bar{\alpha})\langle f'(x^k),\, p^k\rangle \\ &= f(x^k + \bar{\alpha}p^k) + \sigma_1 (C - \bar{\alpha})\langle f'(x^k),\, p^k\rangle. \end{aligned}$$

Taking into account the inequality $C - \bar{\alpha} > 0$, the latter implies

$$\frac{f(x^k + Cp^k) - f(x^k + \bar{\alpha}p^k)}{C - \bar{\alpha}} > \sigma_1 \langle f'(x^k),\, p^k\rangle.$$

Passing onto the limit, and employing the inequalities $\sigma_1 < \sigma_2$ and (2.63), we obtain that

$$\langle f'(x^k + \bar{\alpha}p^k),\, p^k\rangle \geq \sigma_1 \langle f'(x^k),\, p^k\rangle > \sigma_2 \langle f'(x^k),\, p^k\rangle. \tag{2.74}$$

On the other hand, passing onto the limit in (2.71) results in the inequality

$$\langle f'(x^k + \bar{\alpha}p^k),\, p^k\rangle \leq \sigma_2 \langle f'(x^k),\, p^k\rangle,$$

which is in a contradiction with (2.74). □

We conclude this section by mentioning the so-called nonmonotone linesearch methods; see [110]. Allowing an increase of the objective function value on some iterations, these methods tend to produce longer steps. Roughly speaking, the choice of α_k in nonmonotone methods is based on comparison of $f(x^k + \alpha p^k)$ not with $f(x^k)$ but rather with the maximum (or average) value of f along some fixed number of previous iterations. There is computational evidence that such methods may be more efficient in some applications than the usual descent methods.

2.2.2 Quasi-Newton Methods

From now on, we consider descent methods of the specific form

$$x^{k+1} = x^k - \alpha_k Q_k f'(x^k), \quad \alpha_k > 0, \quad k = 0,\, 1,\, \ldots, \tag{2.75}$$

where for each k, $Q_k \in \mathbf{R}^{n \times n}$ is a symmetric positive definite matrix, and the stepsize parameter α_k is chosen by linesearch.

Algorithm 2.23 Choose the parameters $C > 0$, $\sigma \in (0,\, 1)$ and $\theta \in (0,\, 1)$. Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $f'(x^k) = 0$, stop.
2. Choose a symmetric positive definite matrix $Q_k \in \mathbf{R}^{n \times n}$, and compute α_k according to the Armijo rule, employing the direction $p^k = -Q_k f'(x^k)$.
3. Set $x^{k+1} = x^k - \alpha_k Q_k f'(x^k)$.
4. Increase k by 1 and go to step 1.

We first show that the algorithm possesses global convergence (in a certain sense) to stationary points of problem (2.58).

Theorem 2.24. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be differentiable on $\mathbf{R}^n$, and suppose that its gradient is Lipschitz-continuous on $\mathbf{R}^n$. Assume further that there exist $\gamma > 0$ and $\Gamma > 0$ such that the matrices Q_k in Algorithm 2.23 satisfy condition (2.65).*

Then for any starting point $x^0 \in \mathbf{R}^n$, Algorithm 2.23 generates an iterative sequence $\{x^k\}$ such that each of its accumulation points is a stationary point of problem (2.58). Moreover, if an accumulation point exists, or if f is bounded below on $\mathbf{R}^n$, then

$$\{f'(x^k)\} \to 0. \tag{2.76}$$

Proof. The fact that Algorithm 2.23 is well defined follows from Lemma 2.19. Moreover (under the standing assumption that $f'(x^k) \neq 0\ \forall\, k$), the sequence $\{f(x^k)\}$ is monotonically decreasing.

If the sequence $\{x^k\}$ has an accumulation point $\bar{x} \in \mathbf{R}^n$, then $f(\bar{x})$ is an accumulation point of $\{f(x^k)\}$, by the continuity of f. In this case, monotonicity of $\{f(x^k)\}$ implies that the whole sequence $\{f(x^k)\}$ converges to $f(\bar{x})$. If f is bounded below, then the monotone sequence $\{f(x^k)\}$ is bounded below. In this case, $\{f(x^k)\}$ converges even when $\{x^k\}$ does not have any accumulation points.

Since $p^k = -Q_k f'(x^k)$, by the Armijo rule, taking into account Lemma 2.21 and the first inequality in (2.65), we obtain that for all k it holds that

$$f(x^k) - f(x^{k+1}) \geq \sigma \alpha_k \langle Q_k f'(x^k),\, f'(x^k)\rangle \geq \sigma c \gamma \|f'(x^k)\|^2, \tag{2.77}$$

where $c > 0$ is the constant in the right-hand side of (2.66). Since the left-hand side in the relation above tends to zero as $k \to \infty$, we conclude that (2.76) holds. The assertion follows. $\square$

Somewhat more subtle analysis allows to replace Lipschitz-continuity of the gradient of f on the entire $\mathbf{R}^n$ (which is rather restrictive) by simple continuity. The difficulty here is, of course, that under this weaker assumption one cannot guarantee that the values of the stepsize parameter are bounded away from zero.

Theorem 2.25. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be continuously differentiable on $\mathbf{R}^n$. Assume further that there exist $\gamma > 0$ and $\Gamma > 0$ such that the matrices Q_k in Algorithm 2.23 satisfy condition (2.65).*

Then for any starting point $x^0 \in \mathbf{R}^n$ Algorithm 2.23 generates an iterative sequence $\{x^k\}$ such that each of its accumulation points is a stationary point of problem (2.58). Moreover, if the sequence $\{x^k\}$ is bounded, then (2.76) holds.

Proof. The fact that Algorithm 2.23 is well defined follows from Lemma 2.19, as before. Suppose that the sequence $\{x^k\}$ has an accumulation point $\bar{x} \in \mathbf{R}^n$, and let a subsequence $\{x^{k_j}\}$ be convergent to $\bar{x}$ as $j \to \infty$. The case when the corresponding subsequence $\{\alpha_{k_j}\}$ is bounded away from zero is dealt with the same way as in Theorem 2.24 (the only difference is that $\{x^k\}$ in the argument should be replaced by $\{x^{k_j}\}$). Therefore, we consider the case when $\{\alpha_{k_j}\} \to 0$ as $j \to \infty$.

In the latter case, for each j large enough, in the process of backtracking when computing α_{k_j} the initial trial value C was reduced at least once, which means that the value $\alpha = \alpha_{k_j}/\theta$ had been tried and found not to satisfy the Armijo inequality (2.62), i.e.,

$$f\left(x^{k_j} - \frac{\alpha_{k_j}}{\theta} Q_{k_j} f'(x^{k_j})\right) > f(x^{k_j}) - \sigma \frac{\alpha_{k_j}}{\theta} \langle Q_{k_j} f'(x^{k_j}),\, f'(x^{k_j})\rangle.$$

Denoting $\tilde{\alpha}_{k_j} = \alpha_{k_j} \|Q_{k_j} f'(x^{k_j})\|/\theta$ and $\tilde{p}^{k_j} = -Q_{k_j} f'(x^{k_j})/\|Q_{k_j} f'(x^{k_j})\|$, the last inequality can be written in the form

$$f\left(x^{k_j} + \tilde{\alpha}_{k_j} \tilde{p}^{k_j}\right) > f(x^{k_j}) + \sigma \tilde{\alpha}_{k_j} \langle f'(x^{k_j}),\, \tilde{p}^{k_j}\rangle. \tag{2.78}$$

Recalling the second inequality in (2.65), we conclude that $\{\tilde{\alpha}_{k_j}\} \to 0$ as $j \to \infty$. Extracting further subsequences if necessary, we may assume that $\{\tilde{p}^{k_j}\}$ converges to some $\tilde{p} \in \mathbf{R}^n \setminus \{0\}$. With these observations, employing the mean-value theorem (see Theorem A.10, (a)), dividing both sides of (2.78) by $\tilde{\alpha}_{k_j}$ and passing onto the limit as $j \to \infty$, we obtain the inequality

$$\langle f'(\bar{x}),\, \tilde{p}\rangle \geq \sigma \langle f'(\bar{x}),\, \tilde{p}\rangle,$$

which implies that $\langle f'(\bar{x}),\, \tilde{p}\rangle \geq 0$. Then, by (2.65),

$$\begin{aligned}
0 &\geq -\langle f'(\bar{x}),\, \tilde{p}\rangle \\
&= \lim_{j\to\infty} \langle f'(x^{k_j}),\, -\tilde{p}^{k_j}\rangle \\
&= \lim_{j\to\infty} \frac{\langle Q_{k_j} f'(x^{k_j}),\, f'(x^{k_j})\rangle}{\|Q_{k_j} f'(x^{k_j})\|} \\
&\geq \lim_{j\to\infty} \frac{\gamma \|f'(x^{k_j})\|^2}{\Gamma \|f'(x^{k_j})\|} \\
&= \frac{\gamma}{\Gamma} \|f'(\bar{x})\|,
\end{aligned}$$

which is possible only when $f'(\bar{x}) = 0$.

The last assertion of the theorem can be easily derived from the assertion proven above. $\square$

We note that for Algorithm 2.23 with the Armijo linesearch rule replaced by Goldstein or Wolfe rules, global convergence statements analogous to Theorem 2.25 can be obtained.

Observe that neither Theorem 2.24 nor Theorem 2.25 claims the existence of accumulation points for iterative sequences of Algorithm 2.23. However, the latter is evidently guaranteed when f is coercive, since any sequence $\{x^k\} \subset \mathbf{R}^n$ generated by any descent method for problem (2.58) is contained in the level set $\{x \in \mathbf{R}^n \mid f(x) \leq f(x^0)\}$.

The results presented above suggest to try to combine, within a single algorithm, the attractive global convergence properties of descent methods with high convergence rate of Newton-type methods. For that purpose, the Newton-type method should be modified by introducing a stepsize parameter α_k computed by an appropriate linesearch rule. If this rule allows for the full Newton-type step near a qualified solution (i.e., the value $\alpha_k = 1$ is accepted for all k large enough), one can expect that high convergence rate of the Newton-type method would be inherited by the globalized algorithm. At the same time, far from solutions, full Newton-type steps can be too long to ensure monotone decrease of the sequence of the objective function values (and, as a consequence, convergence may not be guaranteed). Far from a solution, the step should therefore be shortened when necessary (i.e., $\alpha_k = 1$ should be reduced). The rest of this section is devoted to formal development of this idea.

Generally, Algorithm 2.23 is referred to as a *quasi-Newton method* for problem (2.58) if, assuming convergence of its iterative sequence to a solution $\bar{x}$, the directions $Q_k f'(x^k)$ approximate Newton directions $(f''(x^k))^{-1} f'(x^k)$ in the sense of the *Dennis–Moré* [57, 58] *condition* (2.80) (or (2.81); cf. (2.32)) stated below. We remark that it is quite natural to discuss quasi-Newton methods for unconstrained optimization in the context of linesearch methods, as it is possible to ensure positive definiteness of Hessian approximations when using some specific quasi-Newton update formulas and the Wolfe rule for computing the stepsize. The resulting algorithms thus fall within the class of descent methods.

The following version of the Dennis–Moré Theorem deals with a linesearch quasi-Newton method, for which the acceptance of full stepsize can be established rather than assumed (see also Theorem 2.29 below).

Theorem 2.26. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its second derivative being continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (2.58). Let $\{x^k\}$ be an iterative sequence of Algorithm 2.23, where $C = 1$ and $\sigma \in (0, 1/2)$, and assume that $\{x^k\}$ converges to $\bar{x}$.*

If the rate of convergence of $\{x^k\}$ is superlinear, then the condition

$$(\alpha_k Q_k - (f''(x^k))^{-1})f'(x^k) = o(\|f'(x^k)\|)$$

holds as $k \to \infty$.

Conversely, if $\bar{x}$ *satisfies the SOSC*

$$\langle f''(\bar{x})\xi,\, \xi\rangle > 0 \quad \forall\, \xi \in \mathbf{R}^n \setminus \{0\}, \tag{2.79}$$

and the condition

$$(Q_k - (f''(x^k))^{-1})f'(x^k) = o(\|f'(x^k)\|) \tag{2.80}$$

holds as $k \to \infty$*, then* $\alpha_k = 1$ *for all* k *large enough, and the rate of convergence of* $\{x^k\}$ *to* $\bar{x}$ *is superlinear.*

Remark 2.27. It can be easily checked that under the assumptions of Theorem 2.26, condition (2.80) is equivalent to

$$(Q_k^{-1} - f''(x^k))(x^{k+1} - x^k) = o(\|x^{k+1} - x^k\|). \tag{2.81}$$

Proof. According to Theorem 2.9 and the fact stated in Remark 2.27, we only need to prove that $\alpha_k = 1$ for all k large enough provided (2.79) and (2.80) hold.

From (2.80) and from the convergence of $\{x^k\}$ to $\bar{x}$, it evidently follows that

$$Q_k f'(x^k) = O(\|f'(x^k)\|) \tag{2.82}$$

as $k \to \infty$.

By the mean-value theorem for scalar-valued functions (see Theorem A.10, (a)), for each k there exists $\tilde{t}_k \in (0,\, 1)$ such that

$$\begin{aligned} f(x^k - Q_k f'(x^k)) = f(x^k) &- \langle f'(x^k),\, Q_k f'(x^k)\rangle \\ &+ \frac{1}{2}\langle f''(\tilde{x}^k)Q_k f'(x^k),\, Q_k f'(x^k)\rangle, \end{aligned}$$

where $\tilde{x}^k = x^k - \tilde{t}_k Q_k f'(x^k)$. It suffices to show that for all k large enough

$$\langle f'(x^k),\, Q_k f'(x^k)\rangle - \frac{1}{2}\langle f''(\tilde{x}^k)Q_k f'(x^k),\, Q_k f'(x^k)\rangle \geq \sigma\langle f'(x^k),\, Q_k f'(x^k)\rangle,$$

that is,

$$(1-\sigma)\langle f'(x^k),\, Q_k f'(x^k)\rangle - \frac{1}{2}\langle f''(\tilde{x}^k)Q_k f'(x^k),\, Q_k f'(x^k)\rangle \geq 0. \tag{2.83}$$

Note that $\{\tilde{x}^k\} \to \bar{x}$, because $\{x^k\} \to \bar{x}$ and $\{Q_k f'(x^k)\} \to 0$ (the latter relation is an immediate consequence of (2.82) and of $\{x^k\} \to \bar{x}$). According to (2.80) and (2.82), we then derive that

$$\begin{aligned}\langle f'(x^k),\, Q_k f'(x^k)\rangle &= \langle f'(x^k),\, (f''(x^k))^{-1} f'(x^k)\rangle + o(\|f'(x^k)\|^2)\\ &= \langle f'(x^k),\, (f''(\bar{x}))^{-1} f'(x^k)\rangle + o(\|f'(x^k)\|^2),\end{aligned}$$

and

$$\begin{aligned}\langle f''(\tilde{x}^k) Q_k f'(x^k),\, Q_k f'(x^k)\rangle &= \langle f''(x^k) Q_k f'(x^k),\, Q_k f'(x^k)\rangle\\ &\quad + o(\|f'(x^k)\|^2)\\ &= \langle f'(x^k),\, Q_k f'(x^k)\rangle + o(\|f'(x^k)\|^2)\\ &= \langle f'(x^k),\, (f''(x^k))^{-1} f'(x^k)\rangle + o(\|f'(x^k)\|^2)\\ &= \langle f'(x^k),\, (f''(\bar{x}))^{-1} f'(x^k)\rangle + o(\|f'(x^k)\|^2),\end{aligned}$$

where the nonsingularity of $f''(\bar{x})$ was taken into account. Hence,

$$\begin{aligned}&(1-\sigma)\langle f'(x^k),\, Q_k f'(x^k)\rangle - \frac{1}{2}\langle f''(x^k) Q_k f'(x^k),\, Q_k f'(x^k)\rangle\\ &= \left(\frac{1}{2} - \sigma\right)\langle (f''(\bar{x}))^{-1} f'(x^k),\, f'(x^k)\rangle + o(\|f'(x^k)\|^2).\end{aligned}$$

The latter implies that (2.83) holds for all k large enough, as $\sigma \in (0,\, 1/2)$ and $(f''(\bar{x}))^{-1}$ is positive definite (by positive definiteness of $f''(\bar{x})$). □

We note that for Algorithm 2.23 with the Armijo linesearch rule replaced by the Goldstein rule (with $0 < \sigma_1 < 1/2 < \sigma_2 < 1$) or the Wolfe rule (with $0 < \sigma_1 < 1/2$, $\sigma_1 < \sigma_2 < 1$) with the initial trial value of the stepsize parameter $\alpha = 1$, results analogous to Theorem 2.26 can be established.

In Theorem 2.26, convergence of the iterates is *assumed*. To obtain a complete result affirming global and locally superlinear convergence, it remains to show that under the assumptions of Theorem 2.24 on global convergence, if the iterates enter a neighborhood of a solution satisfying the SOSC (2.79), then they converge to this solution. Then, if the sequence of matrices $\{Q_k\}$ satisfies the Dennis–Moré condition, Theorem 2.26 guarantees that the rate of convergence is superlinear.

Before stating the needed result, we prove the following local growth property for the norm of the gradient of the objective function, complementing the quadratic growth property in Theorem 1.9.

Lemma 2.28. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be differentiable near $\bar{x} \in \mathbf{R}^n$ and twice differentiable at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (1.10) *satisfying the SOSC* (1.14) *or, equivalently, satisfying*

$$f(x) - f(\bar{x}) \geq \rho\|x - \bar{x}\|^2 \quad \forall\, x \in U \tag{2.84}$$

for some neighborhood U of $\bar{x}$ and some $\rho > 0$.

Then for any $\nu \in (0,\, 4)$, there exists a neighborhood $V \subset U$ of $\bar{x}$ such that

$$\|f'(x)\|^2 \geq \nu\rho(f(x) - f(\bar{x})) \quad \forall\, x \in V. \tag{2.85}$$

Proof. Indeed, for $x \in \mathbf{R}^n$ we have that

$$f'(x) = f'(x) - f'(\bar{x}) = f''(\bar{x})(x - \bar{x}) + o(\|x - \bar{x}\|)$$

as $x \to \bar{x}$, so that

$$\begin{aligned} f(x) - f(\bar{x}) &= \frac{1}{2}\langle f''(\bar{x})(x - \bar{x}),\, x - \bar{x}\rangle + o(\|x - \bar{x}\|^2) \\ &= \frac{1}{2}\langle f'(x),\, x - \bar{x}\rangle + o(\|x - \bar{x}\|^2), \end{aligned}$$

i.e.,

$$\langle f'(x),\, x - \bar{x}\rangle = 2(f(x) - f(\bar{x})) + o(\|x - \bar{x}\|^2).$$

Using (2.84) from Theorem 1.9, we then obtain that for all $x \in U$ close enough to $\bar{x}$ it holds that

$$\begin{aligned} \langle f'(x),\, x - \bar{x}\rangle - \sqrt{\nu}(f(x) - f(\bar{x})) &= (2 - \sqrt{\nu})(f(x) - f(\bar{x})) + o(\|x - \bar{x}\|^2) \\ &\geq (2 - \sqrt{\nu})\rho\|x - \bar{x}\|^2 + o(\|x - \bar{x}\|^2) \\ &\geq 0, \end{aligned}$$

and therefore,

$$\langle f'(x),\, x - \bar{x}\rangle \geq \sqrt{\nu}(f(x) - f(\bar{x})). \tag{2.86}$$

Combining the latter inequality again with (2.84), we obtain that

$$\|f'(x)\|\|x - \bar{x}\| \geq \langle f'(x),\, x - \bar{x}\rangle \geq \sqrt{\nu}(f(x) - f(\bar{x})) \geq \sqrt{\nu}\rho\|x - \bar{x}\|^2,$$

i.e.,

$$\|f'(x)\| \geq \sqrt{\nu}\rho\|x - \bar{x}\|.$$

Hence, using (2.86),

$$\|f'(x)\|^2 \geq \sqrt{\nu}\rho\|x - \bar{x}\|\|f'(x)\| \geq \sqrt{\nu}\rho\langle f'(x),\, x - \bar{x}\rangle \geq \nu\rho(f(x) - f(\bar{x})),$$

which completes the proof. □

Theorem 2.29. *Suppose that the assumptions of Theorem 2.24 are satisfied. Assume, in addition, that $f : \mathbf{R}^n \to \mathbf{R}$ is twice differentiable at $\bar{x} \in \mathbf{R}^n$, which is a stationary point of problem (2.58) satisfying the SOSC (2.79).*

Then if on some iteration k Algorithm 2.23 generates an iterate x^k close enough to $\bar{x}$, it holds that the whole sequence $\{x^k\}$ converges to $\bar{x}$, and the rate of convergence is (at least) geometric.

Moreover, if f is twice differentiable in a neighborhood of $\bar{x}$, with its second derivative being continuous at $\bar{x}$, and if in Algorithm 2.23 we take $C = 1$, $\sigma \in (0,\, 1/2)$, and $\{Q_k\}$ satisfying the Dennis–Moré condition (2.80), then the convergence rate is superlinear.

Proof. By Theorem 1.9 and Lemma 2.28, there exists a neighborhood U of $\bar{x}$ such that the growth conditions

$$f(x) - f(\bar{x}) \geq \rho \|x - \bar{x}\|^2 \quad \forall\, x \in U \tag{2.87}$$

and

$$\|f'(x)\|^2 \geq \nu\rho(f(x) - f(\bar{x})) \quad \forall\, x \in U \tag{2.88}$$

hold with some $\rho > 0$ and $\nu \in (0,\, 4)$. Note also that, by (2.87), it holds that

$$\begin{aligned} \|f'(x)\| = \|f'(x) - f'(\bar{x})\| &\leq L\|x - \bar{x}\| \\ &\leq L\sqrt{\frac{f(x) - f(\bar{x})}{\rho}} \quad \forall\, x \in U, \end{aligned} \tag{2.89}$$

where $L > 0$ is a Lipschitz constant of the gradient of f.

From Lemma 2.21, it follows that (2.77) holds, where $c > 0$ is the constant in the right-hand side of (2.66). Suppose that $x^k \in U$ for some k. Then by (2.77) and (2.88), we have that

$$\begin{aligned} f(x^{k+1}) - f(\bar{x}) &\leq f(x^k) - f(\bar{x}) - \sigma c\gamma \|f'(x^k)\|^2 \\ &\leq (1 - \sigma c\gamma\nu\rho)(f(x^k) - f(\bar{x})) \\ &= q(f(x^k) - f(\bar{x})), \end{aligned} \tag{2.90}$$

where $q = 1 - \sigma c\gamma\nu\rho < 1$.

We next show that if x^k is close enough to $\bar{x}$, then all the subsequent iterates do not leave the neighborhood U of $\bar{x}$. Fix $r > 0$ such that $B(\bar{x},\, r) \subset U$, and define $\delta > 0$ satisfying

$$\delta + \frac{LC\sqrt{(f(x) - f(\bar{x}))/\rho}}{1 - \sqrt{|q|}} \leq r \quad \forall\, x \in B(\bar{x},\, \delta), \tag{2.91}$$

where $C > 0$ is the first trial stepsize value in the Armijo rule. Note that $\alpha_k \leq C$ and $\delta \leq r$. Let $x^k \in B(\bar{x},\, \delta)$. In this case, by (2.89) and (2.91),

$$\begin{aligned} \|x^{k+1} - \bar{x}\| &\leq \|x^k - \bar{x}\| + \|x^{k+1} - x^k\| \\ &\leq \delta + C\|f'(x^k)\| \\ &\leq \delta + LC\sqrt{\frac{f(x^k) - f(\bar{x})}{\rho}} \\ &\leq r, \end{aligned}$$

i.e., $x^{k+1} \in B(\bar{x},\, r)$. From this, using also (2.87), it follows that $q \geq 0$ (as otherwise, (2.90) would not hold).

Suppose that $x^j \in B(\bar{x},\, r)\ \forall\, j = k,\, \ldots,\, s$, for some integer $s \geq k$. Then, by (2.90), we have that

$$\begin{aligned} f(x^j) - f(\bar{x}) &\le q(f(x^{j-1}) - f(\bar{x})) \\ &\vdots \\ &\le q^{j-k}(f(x^k) - f(\bar{x})) \quad \forall\, j = k,\, \ldots,\, s. \end{aligned}$$

Therefore, using also (2.89), we have that

$$\begin{aligned} \|x^{j+1} - x^j\| &\le C\|f'(x^j)\| \\ &\le LC\sqrt{\frac{f(x^j) - f(\bar{x})}{\rho}} \\ &\le (\sqrt{q})^{j-k} LC\sqrt{\frac{f(x^k) - f(\bar{x})}{\rho}} \quad \forall\, j = k,\, \ldots,\, s. \end{aligned}$$

From the latter and (2.91), it follows that

$$\begin{aligned} \|x^{s+1} - \bar{x}\| &\le \|x^s - \bar{x}\| + \|x^{s+1} - x^s\| \\ &\vdots \\ &\le \|x^k - \bar{x}\| + \sum_{l=k}^{s} \|x^{j+1} - x^j\| \\ &\le \delta + LC\sqrt{\frac{f(x^k) - f(\bar{x})}{\rho}} \sum_{l=k}^{s} (\sqrt{q})^{j-k} \\ &\le \delta + \frac{LC\sqrt{(f(x^k) - f(\bar{x}))/\rho}}{1 - \sqrt{q}} \\ &\le r, \end{aligned}$$

i.e., $x^{s+1} \in B(\bar{x},\, r)$.

We have thus established that $x^j \in U$ for all $j = k,\, k+1,\, \ldots$. In particular, (2.90) holds for all k (large enough), which shows that $\{f(x^k)\}$ converges to $f(\bar{x})$ at a linear rate. Then (2.87) implies that $\{x^k\}$ converges to $\bar{x}$ geometrically.

Superlinear convergence rate under the Dennis–Moré condition (2.80) now follows from Theorem 2.26. □

As a direct consequence of Theorem 2.29, we obtain that the usual Newton method with linesearch (for which the Dennis–Moré condition is automatic) is superlinearly convergent whenever its sequence enters a neighborhood of a minimizer satisfying SOSC.

The crucial conclusion from the Dennis–Moré Theorem is that in order to construct a fast optimization method, it is *indispensable* to employ the “second-order information” about the problem either explicitly, or to construct objects describing second-order behavior using first-order information. For the basic Newton method (see Sect. 2.1.2), condition (2.80) is automatic provided the sequence $\{x^k\}$ converges. However, beyond the case of a strongly convex f, there are no reasons to expect $f''(x^k)$ (and hence, $Q_k = (f''(x^k))^{-1}$) to be positive definite for all k. It will indeed be positive definite for x^k close to a solution satisfying the SOSC (2.79) but, in general, not when x^k is far from such a solution. Moreover, even if $f''(x^k)$ is positive definite for all k but the Hessian of f is singular at some (other) points, there is still no guarantee that the iterative sequence would not get stuck near a point of degeneracy, which by no means has to be a stationary point of f. The question concerning the possibility of such behavior was posed in [89] and answered in the affirmative in [196], where an example of nonconvergence of the basic Newton method with the Wolfe linesearch rule is constructed.

The observations above indicate that the basic Newtonian choice of Q_k may be inadequate from the point of view of global convergence, even when exact Hessians are available. Perhaps even more importantly, it turns out that the needed "second-order information" can be constructed without direct computation of Hessians. The main idea of quasi-Newton methods is to completely avoid computing $f''(x^k)$ and solving the corresponding linear system, and instead to approximate the Newton step itself in the sense of the Dennis–Moré condition (2.80). It is important that this approximation does not subsume that $\|Q_k - (f''(x^k))^{-1}\| \to 0$ as $k \to \infty$ and, in fact, this relation indeed does not hold for quasi-Newton methods (in general). The needed approximations must be computed according to some recursive formulas, without using any information about the second derivative of f. Fortunately, such construction can be accomplished, and in many ways.

For each k, define

$$s^k = x^{k+1} - x^k, \quad r^k = f'(x^{k+1}) - f'(x^k). \tag{2.92}$$

Note that these two vectors are already known by the time when Q_{k+1} should be computed, and the goal to achieve (2.81) (which is equivalent to (2.80)) can be modeled as the *quasi-Newton equation*

$$Q_{k+1} r^k = s^k. \tag{2.93}$$

Taking into account that $s^k = -Q_k f'(x^k)$ (by (2.75)), the motivation behind (2.93) is the same as behind the quasi-Newton equation (2.33) for systems of equations; see Sect. 2.1.1.

Therefore, having at hand a symmetric positive definite matrix Q_k and vectors r^k and s^k, it is suggested to choose a symmetric positive definite matrix Q_{k+1} satisfying the quasi-Newton equation (2.93). However, such a choice would be clearly not unique. As in the case of systems of equations, it is natural to additionally require the difference between Q_k and Q_{k+1} to be "minimal" in some sense: from one iteration to another, the variation of Q_k should not be too large. Similarly to quasi-Newton methods for systems of equations, a natural approach is to define Q_{k+1} as a *symmetric* matrix minimizing some matrix norm of $Q_{k+1} - Q_k$ or $Q_{k+1}^{-1} - Q_k^{-1}$. Different norms lead to different specific quasi-Newton methods.

Historically, the first quasi-Newton method is the *Davidon–Fletcher–Powell* (DFP) *method*, in which Q_0 is an arbitrary symmetric positive definite matrix (e.g., $Q_0 = I$), and for each k

$$Q_{k+1} = Q_k + \frac{s^k(s^k)^{\mathrm{T}}}{\langle r^k,\, s^k\rangle} - \frac{(Q_k r^k)(Q_k r^k)^{\mathrm{T}}}{\langle Q_k r^k,\, r^k\rangle}. \tag{2.94}$$

Note that the matrices generated this way remain symmetric and satisfy the quasi-Newton equation (2.93):

$$\begin{aligned} Q_{k+1}r^k &= Q_k r^k + s^k \frac{\langle r^k,\, s^k\rangle}{\langle r^k,\, s^k\rangle} - Q_k r^k \frac{\langle Q_k r^k,\, r^k\rangle}{\langle Q_k r^k,\, r^k\rangle} \\ &= Q_k r^k + s^k - Q_k r^k = s^k. \end{aligned}$$

Moreover, the corresponding Q_{k+1}^{-1} minimizes the weighted Frobenius norm of the correction $Q_{k+1}^{-1} - Q_k^{-1}$ over all the symmetric matrices $Q_{k+1} \in \mathbf{R}^{n\times n}$ satisfying the quasi-Newton equation (2.93); see, e.g., [208, Sect. 11.1] for details. Furthermore, the correction $Q_{k+1} - Q_k$ is a matrix whose rank cannot be greater than 2 (since $\ker(Q_{k+1} - Q_k)$ contains all vectors orthogonal to both r^k and $Q_k s^k$), so the correction is "small" in this sense as well.

Regarding positive definiteness of Q_{k+1}, this depends not only on the quasi-Newton formula used for computing this matrix but also on the choice of the stepsize parameter in (2.75). Specifically, we have the following.

Proposition 2.30. *Let $Q_k \in \mathbf{R}^{n\times n}$ be a symmetric positive definite matrix, and let $s^k,\, r^k \in \mathbf{R}^n$.*

Then formula (2.94) *is well defined and the matrix Q_{k+1} is positive definite if and only if the following inequality holds:*

$$\langle r^k,\, s^k\rangle > 0. \tag{2.95}$$

Proof. The necessity follows immediately from the quasi-Newton equation (2.93), according to which

$$\langle r^k,\, s^k\rangle = \langle Q_{k+1} r^k,\, r^k\rangle.$$

Note also that formula (2.94) is not well defined when $r^k = 0$.

We proceed with sufficiency. From (2.95) it follows that $r^k \neq 0$, and hence, positive definiteness of Q_k implies the inequality $\langle Q_k r^k,\, r^k\rangle > 0$. Combining the latter with (2.95), we obtain that the matrix Q_{k+1} is well defined.

Furthermore, for an arbitrary $\xi \in \mathbf{R}^n$, by (2.94) we derive

$$\begin{aligned}\langle Q_{k+1}\xi,\,\xi\rangle &= \langle Q_k\xi,\,\xi\rangle + \frac{\langle s^k,\,\xi\rangle^2}{\langle r^k,\,s^k\rangle} - \frac{\langle Q_k r^k,\,\xi\rangle^2}{\langle Q_k r^k,\,r^k\rangle}\\ &= \frac{\langle s^k,\,\xi\rangle^2}{\langle r^k,\,s^k\rangle} + \frac{\|Q_k^{1/2}\xi\|^2\|Q_k^{1/2}s^k\|^2 - \langle Q_k^{1/2}\xi,\,Q_k^{1/2}r^k\rangle^2}{\|Q_k^{1/2}r^k\|^2},\end{aligned}$$

where both terms in the right-hand side are nonnegative, according to (2.95) and the Cauchy–Schwarz inequality. Moreover, the equality $\langle Q_{k+1}\xi,\,\xi\rangle = 0$ may hold only when both terms above are equal to zero, i.e., when

$$\langle s^k,\,\xi\rangle = 0, \tag{2.96}$$

and

$$\|Q_k^{1/2}\xi\|\|Q_k^{1/2}r^k\| = |\langle Q_k^{1/2}\xi,\,Q_k^{1/2}r^k\rangle|.$$

The second equality means that $Q_k^{1/2}\xi = tQ_k^{1/2}r^k$ with some $t \in \mathbf{R}$, and since $Q_k^{1/2}$ is nonsingular, this leads to the equality $\xi = tr^k$. Then, by (2.96), $t\langle r^k,\,s^k\rangle = \langle \xi,\,s^k\rangle = 0$, and according to (2.95), the latter is possible only when $t = 0$, i.e., when $\xi = 0$. $\square$

In particular, the inequality (2.95) is always valid if the stepsize parameter in (2.75) is chosen according to the Wolfe rule, while the Armijo rule and the Goldstein rule do not possess this property. This is one of the reasons why the Wolfe rule is recommended for quasi-Newton methods.

Currently, the *Broyden–Fletcher–Goldfarb–Shanno* (BFGS) *method* is regarded as the most efficient general purpose quasi-Newton method. For each k, it defines

$$\begin{aligned}Q_{k+1} = Q_k &+ \frac{(r^k - Q_k s^k)(r^k)^{\mathrm{T}} + r^k(r^k - Q_k s^k)^{\mathrm{T}}}{\langle r^k,\,s^k\rangle}\\ &- \frac{\langle r^k - Q_k s^k,\,s^k\rangle r^k(r^k)^{\mathrm{T}}}{\langle r^k,\,s^k\rangle^2}.\end{aligned} \tag{2.97}$$

It can be immediately verified that (as for the DFP method) the matrices generated according to this formula remain symmetric and satisfy the quasi-Newton equation (2.93), and the rank of corrections $Q_{k+1} - Q_k$ cannot be greater than 2. Moreover, it can be shown that this Q_{k+1} minimizes the weighted Frobenius norm of the correction $Q_{k+1} - Q_k$ over all symmetric matrices $Q_{k+1} \in \mathbf{R}^{n\times n}$ satisfying quasi-Newton equation (2.93); see [208, Sect. 11.1]. For a recent survey of variational origins of the DFP and the BFGS updates, see [111].

Remark 2.31. It can be easily checked that the DFP and BFGS methods can be regarded as "dual" with respect to each other in the following sense. For any symmetric positive definite matrix $Q_k \in \mathbf{R}^{n\times n}$, set $H_k = Q_k^{-1}$. Let Q_{k+1} be generated according to (2.97), let H_{k+1} be generated according to the formula

$$H_{k+1} = H_k + \frac{s^k(s^k)^{\mathrm{T}}}{\langle r^k,\, s^k\rangle} - \frac{(H_k r^k)(H_k r^k)^{\mathrm{T}}}{\langle H_k r^k,\, r^k\rangle}$$

(cf. (2.94)), and suppose that the matrix H_{k+1} is nonsingular. Then the matrix Q_{k+1} is also nonsingular, and $H_{k+1} = Q_{k+1}^{-1}$. From this fact it immediately follows that a counterpart of Proposition 2.30 is valid for the BFGS method as well.

It can be shown that for a quadratic function f with a positive definite Hessian, if α_k is chosen according to the exact linesearch rule, then the DFP and BFGS methods find the unique critical point of f (which is the global solution of problem (2.58), by necessity) from any starting point after no more than $k \le n$ iterations. Moreover, Q_k would coincide with the inverse of the Hessian of f; see, e.g., [18, 19] for details. Recall that for quadratic functions, the exact linesearch rule reduces to an explicit formula. In the non-quadratic case, convergence and rate of convergence results for the DFP and BFGS methods can be found, e.g., in [29, 89, 208]. This analysis is highly nontrivial and is concerned with overcoming serious technical difficulties. In particular, the condition (2.65) is not automatic for the DFP and BFGS methods, and in order to apply Theorems 2.24 or 2.25 one has to verify (2.65), which normally requires some additional assumptions. Here, we present only some general comments.

Known (full) global convergence results for the DFP and BFGS methods are concerned with the case of convex f. The theory of quasi-Newton methods for nonconvex problems is far from being complete, though rich numerical practice puts in evidence that these methods are highly efficient in the nonconvex case as well (especially the BFGS).

Of course, in the non-quadratic case, one cannot expect finite termination of quasi-Newton methods at a solution. However, the rate of convergence usually remains very high. Proving superlinear convergence of a specific quasi-Newton method reduces to the (usually highly nontrivial) verification of the Dennis–Moré condition (2.80), and application of Theorem 2.26.

Quasi-Newton methods are very popular among the users of optimization methods, because they combine high convergence rate with low computational cost per iteration. It is difficult to overestimate the practical value of these methods. For general principles of constructing and analyzing quasi-Newton methods, see [19, 89, 208].

2.2.3 Other Linesearch Methods

Let us briefly mention some other ideas for developing linesearch Newton-type methods, different from the quasi-Newton class. One possibility is to take Q_k as the inverse matrix of a positive definite modification of the Hessian $f''(x^k)$, when the latter is not (sufficiently) positive definite. Specifically, in the process of factorization of $f''(x^k)$, one could replace this matrix by a positive definite matrix if needed, say, of the form $f''(x^k) + \nu_k I$, with the regularization parameter $\nu_k > 0$ large enough. More details on methods of this kind can be found in [208, Sect. 3.4]; for recent contributions, including smart rules for controlling ν_k, see, e.g., [262] and references therein. Note also that instead of modifying the Hessian this way every time when it is not (sufficiently) positive definite, one may choose to do so only when the generated direction p^k is not a descent direction or is not a "good enough" descent direction (positive definiteness of Q_k is sufficient but not necessary for $p^k = -Q_k f'(x^k)$ to be a direction of descent). Other possibilities arise from combining the linesearch strategy with perturbed Newton methods considered in Sect. 2.1.1, and in particular, with the truncated Newton method; see [208, Sect. 7.1] and the corresponding discussion for systems of equations below. See also Sects. 5.1, 6.2, where various ideas of linesearch globalization are applied to constrained optimization and more general variational problems.

We complete this section with a brief discussion of how linesearch can be used to globalize Newton-type methods for a system of equations presented in Sect. 2.1.1. We note that the underlying principles for problems which are more general, or different from unconstrained optimization, are often quite similar nevertheless—the task is to construct an appropriate *merit function* that measures the quality of approximation to a solution in some sense, and for which the direction of a given Newton-type method is of descent, so that linesearch can be applied. We shall get back to those issues more systematically in Sect. 5.1.

Consider the equation

$$\Phi(x) = 0, \tag{2.98}$$

where $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ is a smooth mapping. In this case, it is natural to measure the quality of approximation by a given point $x \in \mathbf{R}^n$ of a solution of (2.98) by the value of the residual $\|\Phi(x)\|$ or, to preserve smoothness, by the value of $\|\Phi(x)\|^2$. Thus, from the optimization point of view, while solving (2.98) we are trying to minimize the merit function $f : \mathbf{R}^n \to \mathbf{R}_+$ given by

$$f(x) = \frac{1}{2}\|\Phi(x)\|^2. \tag{2.99}$$

It is therefore natural to globalize Newton method for (2.98) by introducing linesearch for this merit function f. With this choice,

$$f'(x) = (\Phi'(x))^{\mathrm{T}}\Phi(x), \quad x \in \mathbf{R}^n,$$

and if $\bar{x} \in \mathbf{R}^n$ is a critical point of f, and $\Phi'(\bar{x})$ is a nonsingular matrix, then $\bar{x}$ is a solution of the equation (2.98). On the other hand, if $x \in \mathbf{R}^n$ is not a solution of (2.98), and $\Phi'(x)$ is a nonsingular matrix, then the basic Newtonian direction

$$p = -(\Phi'(x))^{-1}\Phi(x)$$

is a descent direction for the function f at x. This follows from Lemma 2.18, because

$$\langle f'(x),\, p\rangle = -\langle (\Phi'(x))^{\mathrm{T}}\Phi(x),\, (\Phi'(x))^{-1}\Phi(x)\rangle = -\|\Phi(x)\|^2 < 0.$$

Moreover, it holds that

$$p = -(\Phi'(x))^{-1}((\Phi'(x))^{\mathrm{T}})^{-1}(\Phi'(x))^{\mathrm{T}}\Phi(x) = -Q(x)f'(x), \tag{2.100}$$

where the matrix

$$Q(x) = ((\Phi'(x))^{\mathrm{T}}\Phi'(x))^{-1} \tag{2.101}$$

is symmetric and positive definite. Therefore, the corresponding linesearch method can be written in the form (2.75).

Furthermore, the perturbed Newtonian direction has the form

$$p = -(\Phi'(x))^{-1}(\Phi(x) + \omega), \tag{2.102}$$

with some $\omega \in \mathbf{R}^n$ (see (2.11)). Then

$$\begin{aligned} \langle f'(x),\, p\rangle &= -\langle (\Phi'(x))^{\mathrm{T}}\Phi(x),\, (\Phi'(x))^{-1}(\Phi(x) + \omega)\rangle \\ &\leq -\|\Phi(x)\|^2 + \|\Phi(x)\|\|\omega\| \\ &\leq -\|\Phi(x)\|(\|\Phi(x)\| - \|\omega\|), \end{aligned}$$

which is negative provided $\|\omega\| < \|\Phi(x)\|$. Hence, according to Lemma 2.18, in this case p is still a descent direction for f at x. In particular, this will always be the case for a truncated Newton method, characterized by (2.26), if the forcing sequence satisfies $\{\theta_k\} \subset [0,\, \theta]$ with some $\theta \in (0,\, 1)$. More precisely, if $\|\omega\| \leq \theta\|\Phi(x)\|$, then

$$\langle f'(x),\, p\rangle \leq -(1-\theta)\|\Phi(x)\|^2 < 0.$$

As another example of a perturbed Newtonian direction, consider

$$p = -(\Phi'(x) + \Omega)^{-1}\Phi(x), \tag{2.103}$$

with some $\Omega \in \mathbf{R}^{n\times n}$; see (2.36). Such p satisfies (2.102) with

$$\begin{aligned} \omega &= -\Phi(x) + \Phi'(x)(\Phi'(x) + \Omega)^{-1}\Phi(x) \\ &= (\Phi'(x)(\Phi'(x) + \Omega)^{-1} - I)\Phi(x) \\ &= -\Omega(\Phi'(x) + \Omega)^{-1}\Phi(x). \end{aligned}$$

According to the discussion above, it then follows that p is a descent direction for f at x provided $\|\Omega\|\|(\Phi'(x)+\Omega)^{-1}\| < 1$. Employing Lemma A.6, one can easily check that the latter condition is satisfied if $\|\Omega\| < \|(\Phi'(x))^{-1}\|^{-1}/2$.

Note, however, that perturbations resulting from quasi-Newton approximations J of $\Phi'(x)$ may not fit this framework: since such J need not be close to $\Phi'(x)$, there is no reason to expect the corresponding direction

$$p = -J^{-1}\Phi'(x)$$

to be a descent direction for f at $\bar{x}$. This is actually a serious limitation of quasi-Newton methods in the context of nonlinear equations. Observe that this limitation does not exist in the context of optimization, where the objective function serves as a natural merit function.

Note further that writing the Newtonian direction in the form (2.100), (2.101), the associated method can be developed for $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ with m possibly bigger than n. Assuming that $\ker \Phi'(x) = \{0\}$ holds for a given $x \in \mathbf{R}^n$ (which generalizes the assumption that $\Phi'(x)$ is nonsingular to the case of $m \geq n$), this direction is well defined. It is, in fact, the direction produced by the step of the Gauss–Newton method (see (2.41)). According to the discussion above, since $Q(x)$ is positive definite, the resulting linesearch method fits the scheme (2.75).

Thus, one can develop descent methods for systems of equations combining the Newton-type methods from Sect. 2.1.1 with linesearch procedures for the function f defined in (2.99). It should be mentioned, however, that such methods can get stuck near points of singularity of the Jacobian of Φ which need not be solutions of (2.98), in general. In fact, they may not even be critical points of the function f. The possibility of such an undesirable scenario is demonstrated in [35]. In particular, even if the iterative sequence $\{x^k\}$ does not hit any point x where $\ker \Phi'(x) \neq \{0\}$, this sequence can accumulate around such points. Consequently, there is no guarantee that the second inequality in (2.65) will be satisfied with any fixed $\Gamma > 0$ by $Q_k = Q(x^k)$ defined according to (2.101). Therefore, there is no guarantee of global convergence under any reasonable assumptions, unless it is assumed that $\ker \Phi'(x) \neq \{0\}\ \forall\, x \in \mathbf{R}^n$, which is generally too much to expect.

It should be emphasized, however, that appropriate perturbations of the Newtonian directions may not only not harm the descent properties, but in fact can significantly improve them, as well as the overall global behavior of the corresponding linesearch methods. If $n = m$, one can use perturbed directions of the form (2.103), where Ω is selected in such a way that it moves $\Phi'(x) + \Omega$ away from singularity. However, suggesting some general practical rules for defining such Ω could be rather difficult.

The situation changes if we consider the Gauss–Newton direction, which can be naturally perturbed as follows:

$$p = -((\Phi'(x))^{\mathrm{T}}\Phi'(x) + \nu I)^{-1}(\Phi'(x))^{\mathrm{T}}\Phi(x), \tag{2.104}$$

where $\nu \geq 0$ is the regularization parameter. This direction can be regarded as a blend of the pure Gauss–Newton direction, corresponding to $\nu = 0$, and the (anti)gradient direction for the function f defined in (2.99), to which p turns asymptotically as ν tends to $+\infty$.

Algorithms employing directions of the form (2.104) are known under the common name of *Levenberg–Marquardt methods*. Note that these directions can be computed by solving the unconstrained optimization problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|\Phi'(x)p + \Phi(x)\|^2 + \frac{\nu}{2}\|p\|^2 \\ \text{subject to} & p \in \mathbf{R}^n, \end{array}$$

whose quadratic objective function is strongly convex provided $\nu > 0$. Needless to say, a smart control of the parameter ν is essential for the efficiency of this approach. From the global convergence viewpoint, it is desirable to select ν so that the eigenvalues of $(\Phi'(x))^{\mathrm{T}}\Phi'(x)+\nu I$ be separated away from zero by some constant not depending on x. On the other hand, in order to eventually achieve high convergence rate, it is essential that ν must tend to zero as x approaches a solution (so that the method eventually fits the local perturbed Newton framework of Sect. 2.1). An important practical rule for choosing ν arises in the context of trust-region methods; see Sect. 2.3. Observe also that in (2.104), the use of quasi-Newton approximations instead of the exact $\Phi'(x)$ can be considered.

2.3 Trust-Region Methods

Apart from linesearch, another natural strategy to globalize a local method is the following. As discussed above, an iteration of a Newton-type method consists of minimizing a quadratic approximation of the objective function, where the approximation is computed at the current iterate x^k. It is intuitively clear that this approximation can only be "trusted" locally, in some neighborhood of x^k. It is therefore reasonable to minimize the approximation in question in the neighborhood where it is trusted, and not on the whole space. Those considerations give rise to the so-called *trust-region methods*.

Consider again the unconstrained problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathbf{R}^n, \end{array} \tag{2.105}$$

where $f : \mathbf{R}^n \to \mathbf{R}$ is differentiable (twice differentiable in the case of trust-region Newton methods). Let $x^k \in \mathbf{R}^n$ be the current iterate. As discussed earlier, subproblem of the Newton method for problem (2.105) is given by

$$\begin{array}{ll} \text{minimize} & \psi_k(x) \\ \text{subject to} & x \in \mathbf{R}^n, \end{array} \tag{2.106}$$

where

$$\psi_k(x) = f(x^k) + \langle f'(x^k),\, x - x^k\rangle + \frac{1}{2}\langle H_k(x - x^k),\, x - x^k\rangle, \tag{2.107}$$

with $H_k = f''(x^k)$ for the basic choice. Having in mind possible approximations of the Hessian, we shall also consider the case when H_k is a generic matrix, not necessarily $f''(x^k)$. An iteration of the trust-region version of the method consists in solving (possibly approximately) the subproblem

$$\begin{array}{ll} \text{minimize} & \psi_k(x) \\ \text{subject to} & \|x - x^k\| \leq \delta_k, \end{array} \tag{2.108}$$

where $\delta_k > 0$ is the parameter defining a region (in this case, a ball) where the model ψ_k is trusted to be a good approximation of f. Intelligent control of this parameter is clearly the principal issue in this approach.

In what follows, we shall consider only quadratic models ψ_k. Much more general choices are possible, see [45]. In fact, the linesearch approach can also be viewed from the trust-region perspective, making the "extreme" choice of the simple one-dimensional model given by restricting the objective function to the direction of linesearch.

Trust-region methods operate as follows. If a (possibly approximate) solution of the subproblem (2.108) provides a sufficient decrease of the objective function (with respect to what is predicted by the model, see (2.109)), it is accepted as the next iterate. Otherwise, the trust-region parameter is reduced, essentially for similar reasons as reducing the stepsize in linesearch procedures. A very basic version of a trust-region method is the following.

Algorithm 2.32 Choose the parameters $C > 0$ and σ, $\theta \in (0, 1)$, and a starting point $x^0 \in \mathbf{R}^n$. Set $k = 0$.

1. Choose $\delta \geq C$ and a symmetric matrix $H_k \in \mathbf{R}^{n \times n}$.
2. For $\delta_k = \delta$, compute $\tilde{x}^k \in \mathbf{R}^n$, a (possibly approximate) solution of (2.108).
3. If $\tilde{x}^k = x^k$, stop. Otherwise, check the inequality

$$f(\tilde{x}^k) - f(x^k) \leq \sigma(\psi_k(\tilde{x}^k) - f(x^k)). \tag{2.109}$$

   If (2.109) is satisfied, go to step 4; otherwise, set $\delta = \theta\delta_k$ and go to step 2.
4. Set $x^{k+1} = \tilde{x}^k$, increase k by 1 and go to step 1.

We note that it is very important in practice to allow the trust-region radius to increase after successful iterations. While step 1 in Algorithm 2.32 allows this, we shall not go into details of how exactly this should be done. Similarly, when the trust-region is reduced, more sophisticated rules than the simple one in step 3 are often used. We refer the readers to [45] for all the rich details.

Any comparison of efficiency of linesearch and trust-region methods is a complicated issue. In principle, both strategies have some advantages and disadvantages. One could state that trust-region methods became ever more popular over the years. What is clear is that trust-region methods are more robust. Indeed, it is enough to think of the situation when $f'(x^k) = 0$ but there exists a "second-order" descent direction for f from the point x^k, called direction of negative curvature, i.e., some $\xi \in \mathbf{R}^n$ such that $\langle f''(x^k)\xi,\, \xi\rangle < 0$. In such a case, linesearch in the direction $p^k = -Q_k f'(x^k) = 0$ would not make any progress, no matter which matrix Q_k to choose. On the other hand, it is clear that solving the trust-region subproblem (2.108) with $H_k = f''(x^k)$, a point different from x^k would be obtained, and it can be seen that the value of f at this point is smaller than $f(x^k)$; see Proposition 2.37 below. This difference is reflected even in the theoretical properties of the methods: while linesearch descent methods converge to first-order stationary points (see Theorems 2.24, 2.25), the trust-region Newton method converges to second-order stationary points (see Theorem 2.38).

Of course, a trust-region iteration is generally more costly than simple linesearch, especially considering that more than one subproblem of the form (2.108) may need to be solved per iteration (when the trust-region parameter δ_k has to be reduced). However, very efficient special methods had been developed for this purpose [45, 208]. Also, good control rules for δ_k usually avoid the need of solving many subproblems per iteration.

We shall not go into much detail concerning specific methods for solving the trust-region subproblems. But some comments are in order. The following characterization of global solutions of (2.108), due to [202], is the key to the construction.

Proposition 2.33. *For any $g \in \mathbf{R}^n$, any symmetric matrix $H \in \mathbf{R}^{n\times n}$, and any $\delta > 0$, a point $\tilde{\xi}$ is a global solution of the problem*

$$\begin{array}{ll} \text{minimize} & \langle g,\, \xi\rangle + \dfrac{1}{2}\langle H\xi,\, \xi\rangle \\ \text{subject to} & \|\xi\| \le \delta \end{array} \tag{2.110}$$

if and only if there exists $\nu \in \mathbf{R}$ such that

$$g + (H + \nu I)\tilde{\xi} = 0, \quad \nu \ge 0, \quad \|\tilde{\xi}\| \le \delta, \quad \nu(\|\tilde{\xi}\| - \delta) = 0, \tag{2.111}$$

and the matrix $H + \nu I$ is positive semidefinite.

Proof. Denote the objective function of problem (2.110) by

$$q(\xi) = \langle g,\, \xi\rangle + \frac{1}{2}\langle H\xi,\, \xi\rangle, \quad \xi \in \mathbf{R}^n.$$

Suppose first that $\tilde{\xi}$ is a global solution of problem (2.110). If we rewrite the constraint in (2.110) in the equivalent smooth form $\|\xi\|^2/2 \le \delta^2/2$, we observe that it satisfies the LICQ at every feasible point, and the existence of ν satisfying (2.111) follows from the KKT optimality conditions of Theorem 1.14. Moreover, the same holds even for a local solution $\tilde{\xi}$ of problem (2.110), with ν uniquely defined by (2.111).

It remains to show that $H + \nu I$ is positive semidefinite. If $\|\tilde{\xi}\| < \delta$, this follows from the SONC given by Theorem 1.19. (Note, in the passing, that in this case the last condition in (2.111) implies that $\nu = 0$. Hence, H is positive semidefinite. Thus, in this case, (2.110) is a convex problem. In particular, any local solution of this problem is actually global.)

We proceed with the case when $\|\tilde{\xi}\| = \delta$. Since $\tilde{\xi}$ is a global solution of problem (2.110), for any $\xi \in \mathbf{R}^n$ satisfying $\|\xi\| = \delta$ we obtain that

$$\begin{aligned}
\langle g, \tilde{\xi}\rangle + \frac{1}{2}\langle (H + \nu I)\tilde{\xi}, \tilde{\xi}\rangle &= q(\tilde{\xi}) + \frac{\nu}{2}\|\tilde{\xi}\|^2 \\
&= q(\tilde{\xi}) + \frac{\nu}{2}\delta^2 \\
&\leq q(\xi) + \frac{\nu}{2}\delta^2 \\
&= q(\xi) + \frac{\nu}{2}\|\xi\|^2 \\
&= \langle g, \xi\rangle + \frac{1}{2}\langle (H + \nu I)\xi, \xi\rangle.
\end{aligned}$$

Employing the first relation in (2.111), we then derive that

$$-\frac{1}{2}\langle (H + \nu I)\tilde{\xi}, \tilde{\xi}\rangle \leq -\langle (H + \nu I)\tilde{\xi}, \xi\rangle + \frac{1}{2}\langle (H + \nu I)\xi, \xi\rangle,$$

which can be further transformed to

$$\langle (H + \nu I)(\xi - \tilde{\xi}), \xi - \tilde{\xi}\rangle \geq 0.$$

Evidently, the closure of the cone spanned by all $\xi - \tilde{\xi}$ with $\|\xi\| = \delta$ is a half-space defined by a hyperplane containing 0 and orthogonal to $\tilde{\xi}$. Thus, the quadratic form defined by $H + \nu I$ is nonnegative on this half-space, which implies that it is nonnegative on the entire space. Thus, $H + \nu I$ is positive semidefinite.

Suppose now that (2.111) holds with some ν, and the matrix $H + \nu I$ is positive semidefinite. Then $\tilde{\xi}$ is an unconstrained global minimizer of the convex quadratic function $\tilde{q} : \mathbf{R}^n \to \mathbf{R}$ defined by

$$\tilde{q}(\xi) = \langle g, \xi\rangle + \frac{1}{2}\langle (H + \nu I\xi), \xi\rangle = q(\xi) + \frac{\nu}{2}\|\xi\|^2,$$

because $\tilde{q}'(\tilde{\xi}) = g + (H + \nu I)\tilde{\xi} = 0$. Therefore, for any $\xi \in \mathbf{R}^n$ satisfying $\|\xi\| \leq \delta$ it holds that

$$q(\tilde{\xi}) \leq q(\xi) + \frac{\nu}{2}(\|\xi\|^2 - \|\tilde{\xi}\|^2) = q(\xi) + \frac{\nu}{2}(\|\xi\|^2 - \delta^2) \leq q(\xi),$$

where the equality is by the last relation in (2.111). Thus, $\tilde{\xi}$ is a global solution of problem (2.110). □

Let H_k be positive definite, and define

$$p(\nu) = -(H_k + \nu I)^{-1} f'(x^k), \quad \nu \geq 0.$$

According to Proposition 2.33, solution of the subproblem (2.108) is given by $\tilde{x}^k = x^k + p(\nu_k)$ for a certain $\nu_k \geq 0$. If $\|H_k^{-1} f'(x^k)\| \leq \delta_k$, then $\nu_k = 0$; otherwise, $\nu_k \geq 0$ is uniquely defined by the equation

$$\|p(\nu)\| = \delta_k.$$

In particular, computing $\tilde{x}^k$ reduces to the search for the appropriate $\nu_k \geq 0$ (for reasons of conditioning, usually the equivalent equation $1/\|p(\nu)\| = 1/\delta_k$ is solved).

Instead of solving the trust-region subproblems for different values of δ_k, one can try to approximate the solutions curve $p(\cdot)$ directly. Note that changes in δ_k are inverse to the changes in ν, and varying the latter $p(\nu)$ moves in a continuous manner from the Newtonian step $p(0) = -H_k^{-1} f'(x^k)$ to the step in the almost steepest descent direction: $p(\nu) \approx -f'(x^k)/\nu$ for large values $\nu > 0$. The so-called *dog-leg strategy* consists in minimizing the model ψ_k along the piecewise linear trajectory connecting the current iterate x^k, the minimizer x_U^k of ψ_k in the direction of the steepest descent, and the unconstrained minimizer x_N^k of ψ_k (the Newton point), subject to the trust-region constraint. This is illustrated in Fig. 2.6, where the point computed by the dog-leg strategy is denoted by $\tilde{x}^k$ and the exact solution of the trust-region subproblem (2.108) for the given $\delta_k > 0$ is $\bar{x}^k$.

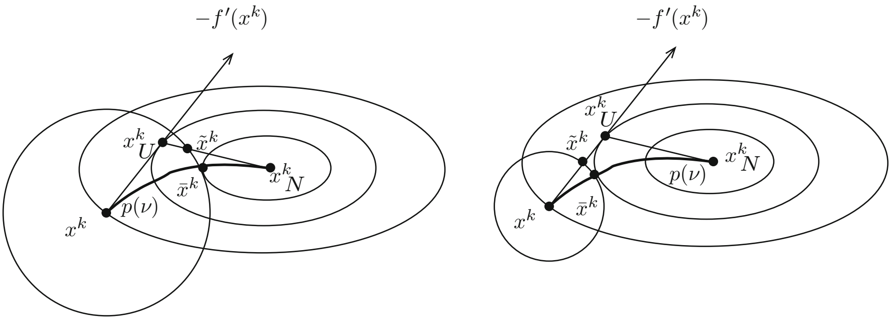

Fig. 2.6 The point computed by the dog-leg strategy

The theory concerning inexact solution of trust-region subproblems is based on the so-called *Cauchy point*, which is the minimizer of the model ψ_k in the direction of steepest descent within the trust-region. Specifically, let $x_C^k = x^k - \alpha_k f'(x^k)$, where $\alpha_k \in \mathbf{R}_+$ is the solution of the problem

$$\begin{array}{ll} \text{minimize} & q_k(\alpha) \\ \text{subject to} & \alpha\|f'(x^k)\| \leq \delta_k,\ \alpha \geq 0, \end{array} \tag{2.112}$$

with the objective function $q_k : \mathbf{R} \to \mathbf{R}$ defined by

$$q_k(\alpha) = \psi_k(x^k - \alpha f'(x^k)).$$

It turns out that any point $\tilde{x}^k$ that decreases the model ψ_k at least as much as the Cauchy point x_C^k can be employed as an approximate solution of the trust-region subproblem (2.108) in Algorithm 2.32. If in the given implementation x_C^k is not computed, the direct verification of the property $\psi_k(\tilde{x}^k) \leq \psi_k(x_C^k)$ is not possible. The following result gives a bound of the progress obtained by the Cauchy point; this bound is always readily available and can serve as a benchmark for judging whether a candidate for approximate solution of the trust-region subproblem (2.108) is acceptable.

Proposition 2.34. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be differentiable at $x^k \in \mathbf{R}^n$.*
Then the Cauchy point x_C^k satisfies

$$f(x^k) - \psi_k(x_C^k) \geq \frac{1}{2}\|f'(x^k)\| \min\left\{\delta_k,\ \frac{\|f'(x^k)\|}{1 + \|H_k\|}\right\}. \tag{2.113}$$

Proof. If $f'(x^k) = 0$, the statement holds trivially. Let $f'(x^k) \neq 0$. Note that

$$q_k(\alpha) = f(x^k) - \alpha\|f'(x^k)\|^2 + \frac{\alpha^2}{2}\langle H_k f'(x^k), f'(x^k)\rangle.$$

If $\langle H_k f'(x^k), f'(x^k)\rangle \leq 0$, then the function q_k is strictly decreasing on $\mathbf{R}_+$. It follows that the solution α_k of (2.112) is on the boundary of the trust-region, i.e., $\alpha_k = \delta_k/\|f'(x^k)\|$. Hence, according to (2.107),

$$\begin{aligned} f(x^k) - \psi_k(x_C^k) &= \alpha_k\|f'(x^k)\|^2 - \frac{\alpha_k^2}{2}\langle H_k f'(x^k), f'(x^k)\rangle \\ &\geq \alpha_k\|f'(x^k)\|^2 \\ &= \delta_k\|f'(x^k)\|, \end{aligned}$$

which verifies that (2.113) holds in this case.

Suppose that $\langle H_k f'(x^k), f'(x^k)\rangle > 0$. In this case the function q_k is strongly convex and has the unique unconstrained global minimizer

$$\bar{\alpha}_k = \frac{\|f'(x^k)\|^2}{\langle H_k f'(x^k), f'(x^k)\rangle}.$$

If $\bar{\alpha}_k \leq \delta_k/\|f'(x^k)\|$, then $\alpha_k = \bar{\alpha}_k$. In that case,

$$\begin{aligned} f(x^k) - \psi_k(x_C^k) &= \bar{\alpha}_k \|f'(x^k)\|^2 - \frac{\bar{\alpha}_k^2}{2} \langle H_k f'(x^k), f'(x^k)\rangle \\ &= \frac{\|f'(x^k)\|^4}{2\langle H_k f'(x^k), f'(x^k)\rangle} \\ &\geq \frac{\|f'(x^k)\|^2}{2\|H_k\|}, \end{aligned}$$

which again verifies (2.113). Finally, if $\bar{\alpha}_k > \delta_k/\|f'(x^k)\|$, then

$$\alpha_k = \frac{\delta_k}{\|f'(x^k)\|} < \frac{\|f'(x^k)\|^2}{\langle H_k f'(x^k), f'(x^k)\rangle},$$

so that

$$\alpha_k^2 \langle H_k f'(x^k), f'(x^k)\rangle < \alpha_k \|f'(x^k)\|^2 = \delta_k \|f'(x^k)\|.$$

Hence,

$$\begin{aligned} f(x^k) - \psi_k(x_C^k) &= \alpha_k \|f'(x^k)\|^2 - \frac{\alpha_k^2}{2} \langle H_k f'(x^k), f'(x^k)\rangle \\ &\geq \delta_k \|f'(x^k)\| - \frac{1}{2}\delta_k \|f'(x^k)\| \\ &= \frac{1}{2}\delta_k \|f'(x^k)\|, \end{aligned}$$

which verifies (2.113) in the last case and completes the proof. □

The next statement shows that if each trust-region subproblem (2.108) is solved to the precision enough to achieve the (computable) bound of reduction guaranteed by the Cauchy point, then the method is well defined. Specifically, a new iterate with a lower objective function value is obtained after reducing the trust-region radius at most a finite number of times.

Let $\gamma \in (0, 1/2]$. For fixed $x^k \in \mathbf{R}^n$ and $H_k \in \mathbf{R}^{n\times n}$, and for each $\delta > 0$, let $\tilde{x}(\delta)$ be some (any) approximate solution of the subproblem (2.108) for $\delta_k = \delta$, satisfying $\|\tilde{x}(\delta) - x^k\| \leq \delta$ and

$$f(x^k) - \psi_k(\tilde{x}(\delta)) \geq \gamma \|f'(x^k)\| \min\left\{\delta,\ \frac{\|f'(x^k)\|}{1 + \|H_k\|}\right\}. \tag{2.114}$$

Points satisfying this property certainly exist, by Proposition 2.34. Moreover, they exist on the piecewise linear dog-leg trajectory, because the Cauchy point itself belongs to this trajectory (see Fig. 2.6, where in the first case x_C^k is the point x_U^k and in the second case it is the dog-leg point $\tilde{x}^k$). Note also that if $\tilde{x}(\delta) = x^k$ so that Algorithm 2.32 stops, by (2.114) we have that $f'(x^k) = 0$.

Proposition 2.35. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be differentiable at $x^k \in \mathbf{R}^n$. Suppose that $f'(x^k) \neq 0$.*

Then if in Algorithm 2.32 approximate solutions $\tilde{x}^k = \tilde{x}(\delta)$ *of trust-region subproblems* (2.108) *satisfy* $\|\tilde{x}(\delta) - x^k\| \leq \delta$ *and* (2.114) *with some* $\gamma > 0$ *and with* $\delta = \delta_k$*, an iterate* x^{k+1} *is generated such that* $f(x^{k+1}) < f(x^k)$*.*

Proof. Since the right-hand side of (2.114) is positive, the assertion would be valid if we show that the acceptance criterion (2.109) is guaranteed to be satisfied after a finite number of reductions of the trust-region radius δ. Because in (2.109) we have $\sigma \in (0, 1)$, it is evidently enough to prove that

$$\lim_{\delta \to 0+} \rho(\delta) = 1, \tag{2.115}$$

where the function $\rho : \mathbf{R}_+ \to \mathbf{R}$ is given by

$$\rho(\delta) = \frac{f(\tilde{x}(\delta)) - f(x^k)}{\psi_k(\tilde{x}(\delta)) - f(x^k)}. \tag{2.116}$$

Note that

$$|\rho(\delta) - 1| = \left| \frac{f(\tilde{x}(\delta)) - \psi_k(\tilde{x}(\delta))}{f(x^k) - \psi_k(\tilde{x}(\delta))} \right|.$$

We have that

$$\begin{aligned} f(\tilde{x}(\delta)) - \psi_k(\tilde{x}(\delta)) &= f(\tilde{x}(\delta)) - f(x^k) - \langle f'(x^k),\, \tilde{x}(\delta) - x^k \rangle \\ &\quad - \frac{1}{2}\langle H_k(\tilde{x}(\delta) - x^k),\, \tilde{x}(\delta) - x^k \rangle \\ &= o(\delta) \end{aligned}$$

as $\delta \to 0$. Combining the latter relation with (2.114), we obtain that

$$|\rho(\delta) - 1| = \frac{o(\delta)}{\delta},$$

which concludes the proof. □

We are now in position to show convergence of the algorithm to stationary points of the problem.

Theorem 2.36. *Let* $f : \mathbf{R}^n \to \mathbf{R}$ *be differentiable on* $\mathbf{R}^n$ *and let its gradient be Lipschitz-continuous on* $\mathbf{R}^n$*. Suppose that there exist* $\Gamma > 0$ *and* $\gamma > 0$ *such that for all* k *in Algorithm 2.32 we have that* $\|H_k\| \leq \Gamma$*, and approximate solutions* $\tilde{x}^k = \tilde{x}(\delta_k)$ *of the trust-region subproblems satisfy* $\|\tilde{x}^k - x^k\| \leq \delta_k$ *and* (2.114) *with* $\delta = \delta_k$*.*

Then each accumulation point of any sequence $\{x^k\}$ *generated by Algorithm 2.32 is a stationary point of problem* (2.105)*.*

Proof. By (2.109), we obtain that

$$\begin{aligned} f(x^k) - f(x^{k+1}) &\geq \sigma(f(x^k) - \psi_k(x^{k+1})) \\ &\geq \gamma\sigma\|f'(x^k)\| \min\left\{\delta_k,\, \frac{\|f'(x^k)\|}{1+\Gamma}\right\}, \end{aligned} \tag{2.117}$$

where the second inequality is by (2.114).

Let $\bar{x}$ be any accumulation point of the sequence $\{x^k\}$. Then $\{f(x^k)\}$ also has an accumulation point, by the continuity of f. Since by (2.117) the sequence $\{f(x^k)\}$ is nonincreasing, it follows that it converges. Hence, $\{f(x^k) - f(x^{k+1})\} \to 0$ as $k \to \infty$.

Suppose that there exists some $\varepsilon > 0$ such that

$$\|f'(x^k)\| \geq \varepsilon \quad \forall\, k. \tag{2.118}$$

Then (2.117) implies that

$$\delta_k \to 0 \text{ as } k \to \infty. \tag{2.119}$$

For any k and any approximate solution $\tilde{x}^k$ of the trust-region subproblem (2.108), let

$$\rho_k = \frac{f(\tilde{x}^k) - f(x^k)}{\psi_k(\tilde{x}^k) - f(x^k)} \tag{2.120}$$

and let $L > 0$ be the Lipschitz constant of the gradient of f. We have that

$$\begin{aligned} |f(\tilde{x}^k) - \psi_k(\tilde{x}^k)| &\leq |f(\tilde{x}^k) - f(x^k) - \langle f'(x^k),\, \tilde{x}^k - x^k\rangle| \\ &\quad + \frac{1}{2}|\langle H_k(\tilde{x}^k - x^k),\, \tilde{x}^k - x^k\rangle| \\ &\leq \frac{(L+\Gamma)\delta_k^2}{2}, \end{aligned}$$

where Lemma A.11 was used. Employing the latter relation and (2.114), we obtain that

$$\begin{aligned} |\rho_k - 1| &= \left|\frac{f(\tilde{x}^k) - \psi_k(\tilde{x}^k)}{f(x^k) - \psi_k(\tilde{x}^k)}\right| \\ &\leq \frac{(L+\Gamma)\delta_k^2}{2\gamma\|f'(x^k)\| \min\{\delta_k,\, \|f'(x^k)\|/(1+\|H_k\|)\}} \\ &\leq \frac{(L+\Gamma)\delta_k^2}{2\gamma\varepsilon \min\{\delta_k,\, \varepsilon/(1+\Gamma)\}} \\ &= \frac{(L+\Gamma)\delta_k}{2\gamma\varepsilon} \\ &\leq 1 - \sigma, \end{aligned}$$

where the second inequality is by (2.118), and the last two relations are valid for all k large enough, once

$$\delta_k \le \bar{\delta} = \min\left\{\frac{\varepsilon}{1+\Gamma},\, \frac{2\gamma\varepsilon(1-\sigma)}{L+\Gamma}\right\},$$

which holds by (2.119) for all k large enough. It follows that the acceptance criterion (2.109), i.e., $\rho_k \ge \sigma$, must hold for the first trust-region radius trial value that is less than $\bar{\delta} > 0$. This contradicts (2.119). We conclude that the hypothesis (2.118) is not valid. Hence,

$$\liminf_{k\to\infty} \|f'(x^k)\| = 0. \tag{2.121}$$

Suppose that

$$\limsup_{k\to\infty} \|f'(x^k)\| > 0,$$

i.e., there exist $\varepsilon > 0$ and a subsequence $\{x^{k_j}\}$ such that

$$\|f'(x^{k_j})\| \ge \varepsilon \quad \forall\, j.$$

For each j, let $k(j)$ be the first index $k > k_j$ such that $\|f'(x^k)\| \le \varepsilon/2$ (this index exists due to (2.121)). Then it holds that

$$\frac{\varepsilon}{2} \le \|f'(x^{k(j)}) - f'(x^{k_j})\| \le L\|x^{k(j)} - x^{k_j}\| \quad \forall\, j.$$

Hence,

$$\frac{\varepsilon}{2L} \le \|x^{k(j)} - x^{k_j}\| \le \sum_{i=k_j}^{k(j)-1} \|x^{i+1} - x^i\| \le \sum_{i=k_j}^{k(j)-1} \delta_i. \tag{2.122}$$

Using the same reasoning as in (2.117), and the fact that $\|f'(x^i)\| \ge \varepsilon/2$ for $i = k_j, \ldots, k(j)-1$ (by the definition of $k(j)$), we then obtain that

$$\begin{aligned} f(x^{k_j}) - \lim_{k\to\infty} f(x^k) &\ge f(x^{k_j}) - f(x^{k(j)}) \\ &= \sum_{i=k_j}^{k(j)-1} (f(x^i) - f(x^{i+1})) \\ &\ge \gamma\sigma \sum_{i=k_j}^{k(j)-1} \|f'(x^i)\| \min\left\{\delta_i,\, \frac{\|f'(x^i)\|}{1+\Gamma}\right\} \\ &\ge \frac{1}{2}\gamma\sigma\varepsilon \sum_{i=k_j}^{k(j)-1} \min\left\{\delta_i,\, \frac{\varepsilon}{2(1+\Gamma)}\right\} \\ &\ge \frac{1}{4}\gamma\sigma\varepsilon^2 \min\left\{\frac{1}{L},\, \frac{1}{1+\Gamma}\right\}, \end{aligned}$$

where the fourth inequality follows from (2.122). Since the left-hand side in the relation above tends to zero as $i \to \infty$, we obtain a contradiction. This completes the proof. $\square$

We next consider the trust-region Newton method, i.e., Algorithm 2.32 with $H_k = f''(x^k)$ and exact solution of subproblems. First note that if for some iteration index k it holds in step 3 that $\tilde{x}^k = x^k$, then this point is stationary for problem (2.105); moreover, it satisfies the second-order necessary condition for optimality. Indeed, since in this case $\tilde{x}^k$ lies in the interior of the feasible set of subproblem (2.108), it is an unconstrained local minimizer of the objective function of this subproblem, and we have that

$$f'(x^k) = f'(x^k) + f''(x^k)(\tilde{x}^k - x^k) = \psi_k'(\tilde{x}^k) = 0,$$

and the matrix

$$f''(x^k) = \psi_k''(\tilde{x}^k)$$

is positive semidefinite (see Theorems 1.7 and 1.8).

Proposition 2.37. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be twice differentiable at $x^k \in \mathbf{R}^n$. Suppose that either $f'(x^k) \neq 0$ or the matrix $f''(x^k)$ is not positive semidefinite.*

Then Algorithm 2.32 with $H_k = f''(x^k)$ and exact solution of subproblems generates an iterate x^{k+1} such that $f(x^{k+1}) < f(x^k)$.

Proof. For each $\delta > 0$, let $\tilde{x}(\delta)$ be some (any) global solution of the subproblem (2.108) with $\delta_k = \delta$, and let $\rho(\delta)$ be given by (2.116). As already seen above, $f'(x^k) \neq 0$ or $f''(x^k)$ not being positive semidefinite imply that $\tilde{x}(\delta) \neq x^k$. Therefore, in particular, it holds that

$$\psi_k(\tilde{x}(\delta)) < \psi_k(x^k) = f(x^k).$$

As in Proposition 2.35, it is sufficient to prove (2.115), as it implies that (2.109) holds for all δ small enough (in which case, an acceptable δ will be computed after a finite number of modifications in step 3 of Algorithm 2.32).

Suppose first that $f'(x^k) \neq 0$. Take any $\xi \in \mathbf{R}^n$ such that $\langle f'(x^k),\, \xi\rangle < 0$, $\|\xi\| = 1$. Since $x^k + \delta\xi$ is a feasible point of subproblem (2.108), we have that

$$\begin{aligned} \psi_k(\tilde{x}(\delta)) &\leq \psi_k(x^k + \delta\xi) \\ &= f(x^k) + \delta\langle f'(x^k),\, \xi\rangle + \frac{\delta^2}{2}\langle f''(x^k)\xi,\, \xi\rangle. \end{aligned}$$

Hence,

$$\begin{aligned} \psi_k(\tilde{x}(\delta)) - f(x^k) &\leq \delta\left(\langle f'(x^k),\, \xi\rangle + \frac{\delta}{2}\|f''(x^k)\|\right) \\ &\leq \frac{\delta}{2}\langle f'(x^k),\, \xi\rangle, \end{aligned} \tag{2.123}$$

where the last inequality holds for all $\delta > 0$ small enough.

On the other hand,

$$\begin{aligned} f(\tilde{x}(\delta)) - \psi_k(\tilde{x}(\delta)) &= f(\tilde{x}(\delta)) - f(x^k) - \langle f'(x^k),\, \tilde{x}(\delta) - x^k\rangle \\ &\quad -\frac{1}{2}\langle f''(x^k)(\tilde{x}(\delta) - x^k),\, \tilde{x}(\delta) - x^k\rangle \\ &= o(\delta^2). \end{aligned} \tag{2.124}$$

Combining (2.124) and (2.123), we obtain that

$$|\rho(\delta) - 1| = \left| \frac{f(\tilde{x}(\delta)) - \psi_k(\tilde{x}(\delta))}{\psi_k(\tilde{x}(\delta)) - f(x^k)} \right| = o(\delta)$$

as $\delta \to 0$, which implies (2.115).

Assume now that $f'(x^k) = 0$, but there exists $\xi \in \mathbf{R}^n$, $\|\xi\| = 1$, such that $\langle f''(x^k)\xi,\, \xi\rangle < 0$. We have that

$$\psi_k(\tilde{x}(\delta)) \leq \psi_k(x^k + \delta\xi) = f(x^k) + \frac{\delta^2}{2}\langle f''(x^k)\xi,\, \xi\rangle.$$

Therefore,

$$\psi_k(\tilde{x}(\delta)) - f(x^k) \leq \frac{\delta^2}{2}\langle f''(x^k)\xi,\, \xi\rangle.$$

Combining this relation with (2.124), we obtain

$$|\rho(\delta) - 1| = \frac{o(\delta^2)}{\delta^2}$$

as $\delta \to 0$, implying (2.115). □

As already mentioned, the trust-region Newton method has the property of convergence to second-order stationary points; this is the result that we prove next.

Theorem 2.38. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be twice continuously differentiable on $\mathbf{R}^n$.*

Then each accumulation point of any sequence $\{x^k\}$ generated by Algorithm 2.32, with $H_k = f''(x^k)$ and exact solution of subproblems, is a stationary point of problem (2.105). Moreover, it satisfies the SONC

$$\langle f''(\bar{x})\xi,\, \xi\rangle \geq 0 \quad \forall\, \xi \in \mathbf{R}^n, \tag{2.125}$$

stated in Theorem 1.8.

Proof. By Proposition 2.37, the sequence $\{f(x^k)\}$ is nondecreasing. Suppose that the sequence $\{x^k\}$ has an accumulation point $\bar{x} \in \mathbf{R}^n$. Then the monotone sequence $\{f(x^k)\}$ has an accumulation point $f(\bar{x})$, which means that it converges (to $f(\bar{x})$). In particular, it holds that $f(x^{k+1}) - f(x^k) \to 0$ as $k \to \infty$. From (2.109) we have that

$$0 \le \sigma(f(x^k) - \psi_k(x^{k+1})) \le f(x^k) - f(x^{k+1}),$$

which implies

$$\psi_k(x^{k+1}) - f(x^k) \to 0 \text{ as } k \to \infty. \tag{2.126}$$

Let $\{x^{k_j}\}$ be any subsequence of $\{x^k\}$ that converges to $\bar{x}$. There are two possibilities:

$$\liminf_{j\to\infty} \delta_{k_j} > 0, \tag{2.127}$$

or

$$\liminf_{j\to\infty} \delta_{k_j} = 0. \tag{2.128}$$

Consider the first case. Define

$$\bar{\delta} = \liminf_{j\to\infty} \delta_{k_j} > 0.$$

Let $\bar{y} \in \mathbf{R}^n$ be a global solution of the problem

$$\begin{aligned} &\text{minimize} \quad \langle f'(\bar{x}),\, x - \bar{x}\rangle + \frac{1}{2}\langle f''(\bar{x})(x-\bar{x}),\, x - \bar{x}\rangle \\ &\text{subject to} \;\; \|x - \bar{x}\| \le \frac{\bar{\delta}}{4}. \end{aligned} \tag{2.129}$$

For all indices j sufficiently large,

$$\|\bar{y} - x^{k_j}\| \le \|\bar{y} - \bar{x}\| + \|x^{k_j} - \bar{x}\| \le \frac{\bar{\delta}}{2} \le \delta_{k_j}.$$

In particular, the point $\bar{y}$ is feasible in the subproblem (2.108) for $k = k_j$. Therefore,

$$\begin{aligned} \psi_{k_j}(x^{k_j+1}) &\le \psi_{k_j}(\bar{y}) \\ &= f(x^{k_j}) + \langle f'(x^{k_j}),\, \bar{y} - x^{k_j}\rangle + \frac{1}{2}\langle f''(x^{k_j})(\bar{y} - x^{k_j}),\, \bar{y} - x^{k_j}\rangle. \end{aligned}$$

Passing onto the limit when $j \to \infty$ and taking into account (2.126), we obtain that

$$0 \le \langle f'(\bar{x}),\, \bar{y} - \bar{x}\rangle + \frac{1}{2}\langle f''(\bar{x})(\bar{y} - \bar{x}),\, \bar{y} - \bar{x}\rangle.$$

Since $\bar{x}$ is obviously feasible in (2.129) and yields zero objective function value, the latter relation implies that the optimal value of (2.129) is exactly zero and $\bar{x}$ is also its global solution. Since $\bar{x}$ lies in the interior of the feasible set of problem (2.129), it is an unconstrained local minimizer of the objective function of this subproblem. Theorems 1.7 and 1.8 imply the assertions of the theorem (in the case of (2.127)).

Suppose now that (2.128) holds. Without loss of generality, we can assume that the whole sequence $\{\delta_{k_j}\}$ converges to zero. Then for all indices j large enough, on iterations indexed by k_j the trust-region parameter had been reduced at least once in step 3 of the algorithm. Hence, for each such j, the value $\tilde{\delta}_{k_j} = \delta_{k_j}/\theta > \delta_{k_j} > 0$ did not satisfy (2.109), i.e., there exists a global solution $\tilde{x}(\tilde{\delta}_{k_j})$ of subproblem (2.108) with $k = k_j$, such that

$$f(\tilde{x}(\tilde{\delta}_{k_j})) > f(x^{k_j}) + \sigma(\psi_{k_j}(\tilde{x}(\tilde{\delta}_{k_j})) - f(x^{k_j})). \tag{2.130}$$

Note that since $\delta_{k_j} \to 0$, we have that $\tilde{\delta}_{k_j} \to 0$ and $\{\tilde{x}(\tilde{\delta}_{k_j}) - x^{k_j}\} \to 0$ as $j \to \infty$.

The rest of the proof follows the lines of that for Proposition 2.37. Suppose the SONC (2.125) does not hold, i.e., there exists $\xi \in \mathbf{R}^n$, $\|\xi\| = 1$, such that

$$\langle f'(\bar{x}),\, \xi\rangle < 0, \tag{2.131}$$

or such that

$$f'(\bar{x}) = 0, \quad \langle f''(\bar{x})\xi,\, \xi\rangle < 0. \tag{2.132}$$

We then obtain that

$$\begin{aligned}\psi_{k_j}(\tilde{x}(\tilde{\delta}_{k_j})) &\leq \psi_{k_j}(x^{k_j} + \tilde{\delta}_{k_j}\xi)\\ &= f(x^{k_j}) + \tilde{\delta}_{k_j}\langle f'(x^{k_j}),\, \xi\rangle + \frac{\tilde{\delta}_{k_j}^2}{2}\langle f''(x^{k_j})\xi,\, \xi\rangle.\end{aligned}$$

In the case of (2.131), from the latter relation we conclude, for all j large enough, that

$$\begin{aligned}\psi_{k_j}(\tilde{x}(\tilde{\delta}_{k_j})) - f(x^{k_j}) &\leq \tilde{\delta}_{k_j}\left(\langle f'(x^{k_j}),\, \xi\rangle + \frac{\tilde{\delta}_{k_j}}{2}\|f''(x^{k_j})\|\right)\\ &\leq \frac{\tilde{\delta}_{k_j}}{2}\langle f'(\bar{x}),\, \xi\rangle.\end{aligned}$$

Defining

$$\rho_{k_j} = \frac{f(\tilde{x}(\tilde{\delta}_{k_j})) - f(x^{k_j})}{\psi_{k_j}(\tilde{x}(\tilde{\delta}_{k_j})) - f(x^{k_j})},$$

we have that

$$|\rho_{k_j} - 1| = \frac{o(\tilde{\delta}_{k_j}^2)}{\tilde{\delta}_{k_j}}$$

as $j \to \infty$. But then

$$\rho_{k_j} \to 1 \text{ as } j \to \infty, \tag{2.133}$$

in contradiction with (2.130).

In the case of (2.132), the relation (2.133) is obtained analogously (see the proof of Proposition 2.37). This again gives a contradiction with (2.130). □

It remains to verify whether, under natural assumptions, the trust-region Algorithm 2.32 locally reduces to the pure Newton method, thus inheriting its fast convergence rate.

Theorem 2.39. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its second derivative being continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (2.105) *that satisfies the SOSC*

$$\langle f''(\bar{x})\xi,\, \xi\rangle > 0 \quad \forall\, \xi \in \mathbf{R}^n \setminus \{0\}.$$

In Algorithm 2.32, let $H_k = f''(x^k)$ and the subproblems be solved exactly.

Then there exists a neighborhood U of $\bar{x}$ such that if $x^k \in U$ for some k, then the next iterate x^{k+1} generated by Algorithm 2.32 coincides with the point $x^k - (f''(x^k))^{-1} f'(x^k)$, the sequence $\{x^k\}$ generated by Algorithm 2.32 converges to $\bar{x}$ and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the Hessian of f is locally Lipschitz-continuous with respect to $\bar{x}$.

Proof. Under the stated assumptions, there exist a neighborhood U of $\bar{x}$, $\mu > 0$ and $M > 0$ such that if $x^k \in U$, then $f''(x^k)$ is positive definite,

$$\langle (f''(x^k))^{-1}\xi,\, \xi\rangle \geq \mu\|\xi\|^2 \quad \forall\, \xi \in \mathbf{R}^n, \quad \|(f''(x^k))^{-1}\| \leq M, \tag{2.134}$$

and in particular, the point $\tilde{x}^k = x^k - (f''(x^k))^{-1} f'(x^k)$ is well defined. Moreover, since this point satisfies $\psi_k'(\tilde{x}^k) = 0$, it is the unique unconstrained global minimizer of the strongly convex quadratic function ψ_k defined in (2.107). Furthermore, since $f'(\bar{x}) = 0$, the neighborhood U can be taken small enough, so that

$$\|\tilde{x}^k - x^k\| = \|(f''(x^k))^{-1} f'(x^k)\| \leq M\|f'(x^k)\| < C,$$

where the second relation in (2.134) was taken into account. In such a case, the global minimizer $\tilde{x}^k$ of ψ_k is feasible in the trust-region subproblem (2.108) for any $\delta_k \geq C$. Recall that at the beginning of each iteration (in step 1 of Algorithm 2.32), we select $\delta \geq C$. Hence, $\tilde{x}^k$ is the unique global solution of (2.108) for the initial choice of trust-region parameter $\delta_k = \delta$, and it remains to show that this point satisfies the sufficient descent condition (2.109), so that the initial choice of δ_k is never reduced.

To this end, observe that by (2.107) and by the definition of $\tilde{x}^k$, the following holds for x^k close enough to $\bar{x}$:

$$
\begin{aligned}
&f(\tilde{x}^k) - f(x^k) - \sigma(\psi_k(\tilde{x}^k) - f(x^k)) \\
&= f(\tilde{x}^k) - f(x^k) - \sigma\left(\langle f'(x^k),\, \tilde{x}^k - x^k\rangle + \frac{1}{2}\langle f''(x^k)(\tilde{x}^k - x^k),\, \tilde{x}^k - x^k\rangle\right) \\
&= (1-\sigma)\Bigg(-\langle f'(x^k),\, (f''(x^k))^{-1} f'(x^k)\rangle \\
&\qquad + \frac{1}{2}\langle f''(x^k)(f''(x^k))^{-1} f'(x^k),\, (f''(x^k))^{-1} f'(x^k)\rangle\Bigg) + o(\|f'(x^k)\|^2) \\
&= -\frac{1-\sigma}{2}\langle (f''(x^k))^{-1} f'(x^k),\, f'(x^k)\rangle + o(\|f'(x^k)\|^2) \\
&\leq -\frac{1-\sigma}{2}\mu\|f'(x^k)\|^2 + o(\|f'(x^k)\|^2) \\
&\leq 0,
\end{aligned}
$$

where the first inequality is by the first relation in (2.134).

We have thus shown that if a neighborhood U of $\bar{x}$ is small enough and $x^k \in U$, then the point $x^k - (f''(x^k))^{-1} f'(x^k)$ is well defined and is accepted by Algorithm 2.32 as x^{k+1}. The assertions follow now from Theorem 2.15. □

As already commented, the basic Newton model ψ_k defined in (2.107) can be replaced in trust-region subproblem (2.108) by other models, and in particular, by those originating from perturbed/modified Newtonian methods. As an example, let us recall the Gauss–Newton method for the equation

$$
\Phi(x) = 0,
$$

where $\Phi : \mathbf{R}^n \to \mathbf{R}^m$ is a smooth mapping, $n \leq m$. According to (2.41), for a given iterate $x^k \in \mathbf{R}^n$, the full Gauss–Newton step is defined by the linear system

$$
(\Phi'(x^k))^{\mathrm{T}} \Phi(x^k) + (\Phi'(x^k))^{\mathrm{T}} \Phi'(x^k)(x - x^k) = 0.
$$

Note that this system characterizes the stationary points of the linear least-squares problem:

$$
\begin{array}{ll}
\text{minimize} & \frac{1}{2}\|\Phi(x^k) + \Phi'(x^k)(x - x^k)\|^2 \\
\text{subject to} & x \in \mathbf{R}^n.
\end{array}
\tag{2.135}
$$

Supplying problem (2.135) with the trust-region constraint, we obtain the subproblem of the Levenberg–Marquardt method:

$$
\begin{array}{ll}
\text{minimize} & \frac{1}{2}\|\Phi(x^k) + \Phi'(x^k)(x - x^k)\|^2 \\
\text{subject to} & \|x - x^k\| \leq \delta_k.
\end{array}
\tag{2.136}
$$

The name fully agrees with terminology introduced in Sect. 2.2.3. Indeed, subproblem (2.136) is convex, and according to Proposition 2.33 (applied with $g = (\Phi'(x^k))^{\mathrm{T}}\Phi(x^k)$ and positive semidefinite $H = (\Phi'(x^k))^{\mathrm{T}}\Phi'(x^k))$, its solution $\tilde{x}^{k+1}$ is characterized by the feasibility condition $\|\tilde{x}^{k+1} - x^k\| \leq \delta_k$ and by the existence of $\nu_k \geq 0$ such that $\nu_k = 0$ provided $\|\tilde{x}^{k+1} - x^k\| < \delta_k$, and

$$(\Phi'(x^k))^{\mathrm{T}}\Phi(x^k) + ((\Phi'(x^k))^{\mathrm{T}}\Phi'(x^k) + \nu_k I)(\tilde{x}^{k+1} - x^k) = 0.$$

The step defined this way corresponds precisely to the direction defined according to (2.104). Observe that if the point $\tilde{x}^{k+1}$ obtained by the full Gauss–Newton step satisfies $\|\tilde{x}^{k+1} - x^k\| < \delta_k$, then this $\tilde{x}^{k+1}$ also solves the trust-region subproblem (2.136).

Convergence analysis of trust-region methods with inexact Newtonian models can be found in [208, Chap. 4].

To conclude, we note that instead of the trust-region subproblem (2.108), where a quadratic function is minimized in a ball, one could consider the subproblem

$$\begin{array}{ll} \text{minimize} & \psi_k(x) \\ \text{subject to} & \|x - x^k\|_\infty \leq \delta_k, \end{array}$$

where the minimization is carried out in a box. In other words, instead of the Euclidean norm $\|\cdot\| = \|\cdot\|_2$, the trust-region is defined using the norm $\|\cdot\|_\infty$ (from computational point of view, other norms can hardly be useful). With this change, essentially the same results can be obtained. As a practical matter, minimization of quadratic functions subject to bound constraints also can be performed efficiently, using specialized software.

Both linesearch and trust-region techniques play important roles for globalizing Newton-type methods in the more general settings of constrained optimization and variational problems, introducing merit functions to measure progress of a given algorithm.

2.4 Semismooth Newton Method

In this section we extend Newton-type methods to nonsmooth equations and to unconstrained optimization problems with nonsmooth derivatives. Applications of the resulting methods to variational problems will be the subject of Sect. 3.2.

2.4.1 Semismooth Newton Method for Equations

Consider the equation

$$\Phi(x) = 0, \tag{2.137}$$

where $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ is no longer assumed to be smooth. Various examples of nonsmooth equations arising from variational and optimization problems are provided in Sect. 3.2 below; see [213] for other examples.

One seemingly appealing idea in this setting is to apply quasi-Newton methods, since the derivatives of Ψ are not involved in their iterations; see Sect. 2.1.1. However, it is known that quasi-Newton methods with standard updates are guaranteed to preserve superlinear convergence only when the equation mapping is actually differentiable at the solution (see [125, 195]), which is an assumption too restrictive.

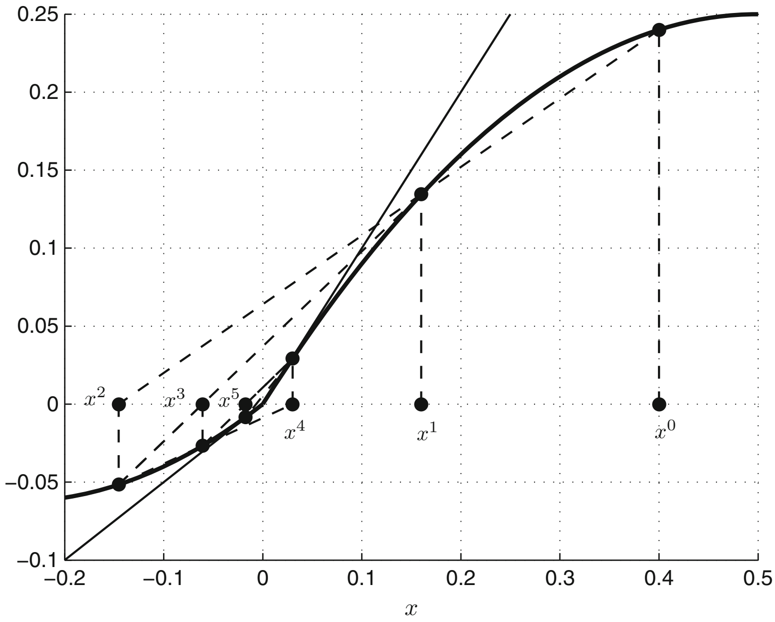

Fig. 2.7 Iterates in Example 2.40

Example 2.40. Let $n = 1$,

$$\Phi(x) = \begin{cases} x/2 + x^2 & \text{if } x \leq 0, \\ x - x^2 & \text{if } x > 0. \end{cases}$$

The solution of interest is $\bar{x} = 0$, and Φ is locally Lipschitz-continuous at 0 and differentiable everywhere except for 0. Evidently, $\partial_B \Phi(0) = \{1/2, 1\}$, $\partial \Phi(0) = [1/2, 1]$, and in particular, Φ is *CD*-regular at 0.

For the equation (2.137) with this Φ, the secant method (which is the one-dimensional instance of Broyden's method; see Sect. 2.1.1) works as follows: being started from $x^0 > 0$ small enough, it generates $x^1 > 0$, then $x^2 < 0$ and $x^3 < 0$, then $x^4 > 0$, then $x^5 < 0$ and $x^6 < 0$, etc. (see Fig. 2.7). After computing two subsequent iterates corresponding to the same smooth piece, namely, $x^{k-2} < 0$ and $x^{k-1} < 0$ for some $k \geq 4$, the method generates good secant approximation of the function, and the next iterate $x^k > 0$ is "superlinearly closer" to the solution than x^{k-1}: $x^k = o(|x^{k-1}|)$. Then, according to (2.35), we derive the estimate

$$\begin{aligned} x^{k+1} &= x^k - \frac{\Phi(x^k)(x^k - x^{k-1})}{\Phi(x^k) - \Phi(x^{k-1})} \\ &= x^k - \frac{(x^k - (x^k)^2)(x^k - x^{k-1})}{x^k - (x^k)^2 - x^{k-1}/2 - (x^{k-1})^2} \\ &= x^k - \frac{-x^k x^{k-1} + o(x^k x^{k-1})}{-x^{k-1}/2 + o(x^{k-1})} \\ &= x^k - 2x^k + o(x^k) \\ &= -x^k + o(x^k), \end{aligned}$$

and superlinear decrease of the distance to the solution is lost.

The idea of the *semismooth Newton method*, originating from [175, 176] and independently developed in [222, 225], is to replace the Jacobian in the Newtonian iteration system (2.2) by an element of Clarke's generalized Jacobian, which may be nonempty even if the true Jacobian does not exist (see Sect. 1.4). Specifically, for the current iterate $x^k \in \mathbf{R}^n$, the next iterate is a solution of the linear equation

$$\Phi(x^k) + J_k(x - x^k) = 0 \tag{2.138}$$

with some $J_k \in \partial\Phi(x^k)$. This is a direct generalization of the Newton method for the differentiable case.

Algorithm 2.41 Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $\Phi(x^k) = 0$, stop.
2. Compute some $J_k \in \partial\Phi(x^k)$. Compute $x^{k+1} \in \mathbf{R}^n$ as a solution of (2.138).
3. Increase k by 1 and go to step 1.

Assuming that J_k is nonsingular, the semismooth Newton method can be written in the form of the explicit iterative scheme

$$x^{k+1} = x^k - J_k^{-1}\Phi(x^k), \quad J_k \in \partial\Phi(x^k), \quad k = 0, 1, \ldots. \tag{2.139}$$

Recall that the nondegeneracy assumption needed for local superlinear convergence of the Newton method to the solution $\bar{x} \in \mathbf{R}^n$ the equation (2.137) in the smooth case, consists of nonsingularity of the Jacobian $\Phi'(\bar{x})$ (see Theorem 2.2). In the context of semismooth Newton methods, this condition is replaced by CD-regularity of Φ at $\bar{x}$, as defined in Sect. 1.4.1.

Theorem 2.42. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be locally Lipschitz-continuous at $\bar{x} \in \mathbf{R}^n$, and assume that the estimate*

$$\sup_{J \in \partial \Phi(x+\xi)} \|\Phi(x+\xi) - \Phi(x) - J\xi\| = o(\|\xi\|) \tag{2.140}$$

holds as $\xi \in \mathbf{R}^n$ tends to 0. Let $\bar{x}$ be a solution of the equation (2.137), and assume that Φ is CD-regular at $\bar{x}$.

Then the following assertions are valid:

(a) There exists a neighborhood U of $\bar{x}$ and a function $q(\cdot) : U \to \mathbf{R}$ such that every $J \in \partial \Phi(x)$ is nonsingular for all $x \in U$,

$$\|x - J^{-1}\Phi(x) - \bar{x}\| \leq q(x)\|x - \bar{x}\| \quad \forall\, x \in U, \tag{2.141}$$

and

$$q(x) \to 0 \ \text{as}\ x \to \bar{x}. \tag{2.142}$$

(b) Any starting point $x^0 \in \mathbf{R}^n$ close enough to $\bar{x}$ defines a particular iterative sequence of Algorithm 2.41, any such sequence converges to $\bar{x}$, and the rate of convergence is superlinear.

(c) If the estimate

$$\sup_{J \in \partial \Phi(x+\xi)} \|\Phi(x+\xi) - \Phi(x) - J\xi\| = O(\|\xi\|^2) \tag{2.143}$$

holds as $\xi \in \mathbf{R}^n$ tends to 0, then

$$q(x) = O(\|x - \bar{x}\|) \tag{2.144}$$

as $x \to \bar{x}$, and the rate of convergence is quadratic.

Proof. From Proposition 1.51 and Lemma A.6, it easily follows that there exist a neighborhood U of $\bar{x}$ and $M > 0$ such that

$$J \text{ is nonsingular}, \quad \|J^{-1}\| \leq M \quad \forall\, J \in \partial \Phi(x),\, \forall\, x \in U. \tag{2.145}$$

Combining this with (2.140), we conclude that the inclusions $x \in U$ and $J \in \partial \Phi(x)$ imply

$$\begin{aligned} \|x - J^{-1}\Phi(x) - \bar{x}\| &\leq \|J^{-1}\|\|\Phi(x) - \Phi(\bar{x}) - J(x - \bar{x})\| \\ &\leq q(x)\|x - \bar{x}\| \end{aligned} \tag{2.146}$$

with some $q(x)$ satisfying (2.142), and (2.146) gives (2.141). This completes the proof of assertion (a).

Assertion (b) follows from (a) by the same argument as in the proof of Theorem 2.2.

Finally, if (2.143) holds, then from (2.145) and the first inequality in (2.146) it follows that the inclusions $x \in U$ and $J \in \partial \Phi(x)$ imply the second inequality with some $q(x)$ satisfying (2.144). The quadratic convergence rate now follows from (2.144) and (2.146). This proves (c). $\square$

Following the lines of previous development of Newton-type methods in Sect. 2.1.1, some useful variations of the semismooth Newton method can be modeled via its perturbed versions. For a given $x^k \in \mathbf{R}^n$, the next iterate $x^{k+1} \in \mathbf{R}^n$ of a perturbed semismooth Newton method must satisfy the system

$$\Phi(x^k) + J_k(x - x^k) + \omega^k = 0 \tag{2.147}$$

with some $J_k \in \partial\Phi(x^k)$ and some perturbation term $\omega^k \in \mathbf{R}^n$.

The next technical lemma follows from Theorem 2.42 the same way as Lemma 2.3 follows from Theorem 2.2.

Lemma 2.43. *Under the assumptions of Theorem 2.42, there exist a neighborhood U of $\bar{x}$ and $M > 0$ such that for any $x^k \in U$, any $J_k \in \partial\Phi(x^k)$, any $\omega^k \in \mathbf{R}^n$, and any $x^{k+1} \in \mathbf{R}^n$ satisfying the equation (2.147), it holds that*

$$\|x^{k+1} - \bar{x}\| \leq M\omega^k + o(\|x^k - \bar{x}\|) \tag{2.148}$$

as $x^k \to \bar{x}$. Moreover, if the estimate (2.143) holds, then the estimate (2.148) can be sharpened as follows:

$$\|x^{k+1} - \bar{x}\| \leq M\omega^k + O(\|x^k - \bar{x}\|^2) \tag{2.149}$$

as $x^k \to \bar{x}$.

The proof of the next a posteriori result mimics the proof of Proposition 2.4, but with (2.140) and (2.143) used instead of smoothness assumptions, and with Lemma 2.43 used instead of Lemma 2.3.

Proposition 2.44. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be locally Lipschitz-continuous at a point $\bar{x} \in \mathbf{R}^n$, and assume that the estimate (2.140) holds for $\xi \in \mathbf{R}^n$. Let $\bar{x}$ be a solution of the equation (2.137). Let a sequence $\{x^k\} \subset \mathbf{R}^n$ be convergent to $\bar{x}$ and such that x^{k+1} satisfies (2.147) with some $J_k \in \partial\Phi(x^k)$ and $\omega^k \in \mathbf{R}^n$ for all k large enough.*

If the rate of convergence of $\{x^k\}$ is superlinear, then

$$\omega^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \tag{2.150}$$

as $k \to \infty$.

Conversely, if Φ is CD-regular at $\bar{x}$, and (2.150) holds, then the rate of convergence of $\{x^k\}$ is superlinear. Moreover, the rate of convergence is quadratic, provided (2.143) holds and

$$\omega^k = O(\|x^{k+1} - x^k\|^2 + \|x^k - \bar{x}\|^2)$$

as $k \to \infty$.

Remark 2.45. By Remark 2.5, we note that in condition (2.150) the right-hand side can be replaced by either $o(\|x^{k+1} - x^k\|)$ or $o(\|x^k - \bar{x}\|)$. The condition modified this way is generally stronger than (2.150). However, if $\{x^k\}$ is assumed to be superlinearly convergent to $\bar{x}$, these conditions become equivalent.

Remark 2.46. The assumptions of the sufficiency part of Proposition 2.44 can be formally relaxed as follows: instead of assuming that (2.150) holds with $\omega^k = -\Phi(x^k) - J_k(x^{k+1} - x^k)$ for the *given* $J_k \in \partial\Phi(x^k)$, one can assume that (2.150) holds for *some* $J_k \in \partial\Phi(x^k)$, which means that

$$\min_{J \in \partial\Phi(x^k)} \|\Phi(x^k) + J(x^{k+1} - x^k)\| = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)$$

as $k \to \infty$. However, due to the necessity part of Proposition 2.44, taking into account Proposition 1.51, this would actually imply that

$$\max_{J \in \partial\Phi(x^k)} \|\Phi(x^k) + J(x^{k+1} - x^k)\| = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)$$

as $k \to \infty$. Therefore, this modification does not make Proposition 2.44 any sharper.

Remark 2.47. Local Lipschitz-continuity of Φ at $\bar{x}$ and the estimate (2.140) are the ingredients of semismoothness of Φ at $\bar{x}$, while estimate (2.143) is the ingredient of strong semismoothness (see Sect. 1.4.2). Therefore, by Theorem 2.42, if Φ is CD-regular at a solution $\bar{x}$ of the equation (2.137), Algorithm 2.41 possesses local superlinear convergence to this solution provided Φ is semismooth at $\bar{x}$. (That is why we use the name "semismooth Newton method.") Moreover, the rate of convergence is quadratic provided Φ is strongly semismooth at $\bar{x}$. The development in [222, 225] relies on semismoothness assumption, while [175, 176] directly employ (2.140), which gives a sharper result. Note also that the condition (2.140) in Theorem 2.42 cannot be dropped; see [68, Example 7.4.1].

Conditions (2.140) and (2.143) can be replaced by semismoothness and strong semismoothness, respectively, in Lemma 2.43 and Proposition 2.44 as well. In the rest of this section, we employ the assumption of semismoothness: this makes the exposition somewhat simpler, even though not all the ingredients of semismoothness are always needed.

As an immediate application of Proposition 2.44, consider the following *semismooth quasi-Newton method*. Let $\{J_k\} \subset \mathbf{R}^{n \times n}$ be a sequence of matrices. For the current iterate $x^k \in \mathbf{R}^n$, let the next iterate x^{k+1} be computed as a solution of the equation (2.138), and assume that $\{J_k\}$ satisfies the Dennis–Moré-type condition (cf. (2.32) and (2.80), (2.81)):

$$\min_{J \in \partial\Phi(x^k)} \|(J_k - J)(x^{k+1} - x^k)\| = o(\|x^{k+1} - x^k\|) \tag{2.151}$$

as $k \to \infty$. The latter implies the existence of $\tilde{J}_k \in \partial\Phi(x^k)$ such that

$$(J_k - \tilde{J}_k)(x^{k+1} - x^k) = o(\|x^{k+1} - x^k\|). \tag{2.152}$$

Therefore, the equation (2.138) can be interpreted as (2.147) with J_k replaced by $\tilde{J}_k$, and with

$$\omega^k = (J_k - \tilde{J}_k)(x^{k+1} - x^k).$$

By (2.152),

$$\omega^k = o(\|x^{k+1} - x^k\|)$$

as $k \to \infty$, and hence, Proposition 2.44 implies the following a posteriori result.

Theorem 2.48. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be semismooth at a point $\bar{x} \in \mathbf{R}^n$. Let $\bar{x}$ be a solution of the equation (2.137), and let Φ be CD-regular at $\bar{x}$. Let $\{J_k\} \subset \mathbf{R}^{n\times n}$ be a sequence of matrices, and let a sequence $\{x^k\} \subset \mathbf{R}^n$ be convergent to $\bar{x}$ and such that x^{k+1} satisfies (2.138) for all k large enough. Assume, finally, that condition (2.151) holds.*

Then the rate of convergence of $\{x^k\}$ is superlinear.

Remark 2.49. Evidently, condition (2.151) in Theorem 2.48 can be replaced by the following formally stronger condition:

$$\max_{J\in\partial\Phi(x^k)} \|(J_k - J)(x^{k+1} - x^k)\| = o(\|x^{k+1} - x^k\|)$$

as $k \to \infty$. Moreover, according to Remark 2.46, in the context of Theorem 2.48 this condition is actually not stronger than (2.151).

It would seem natural to state next a counterpart of Proposition 2.6 for the perturbed semismooth Newton method. However, such a statement would be of less interest in this context, because the special selection of the norm $\|\cdot\|_*$ in such a counterpart of Proposition 2.6 would have now to take care of *all* $J \in \partial\Phi(\bar{x})$ simultaneously. Because of that, the consequences would not be as nice as in the smooth case. In particular, in the counterpart of Theorem 2.8 (see Theorem 2.51 below), it is not enough to assume that forcing parameters are bounded above by some (arbitrary fixed) constant $\theta \in (0, 1)$: we can only establish the *existence* of the needed θ.

Therefore, we proceed directly to *truncated semismooth Newton methods*. These are practically important instances of perturbed semismooth Newton methods, with the perturbation terms in the iteration system (2.147) satisfying the condition

$$\|\omega^k\| \leq \theta_k\|\Phi(x^k)\|, \quad k = 0, 1, \ldots \tag{2.153}$$

for some forcing sequence $\{\theta_k\}$. Note that condition (2.153) is exactly the same as truncation condition (2.26) employed above for smooth equations.

Theorem 2.50. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be semismooth at $\bar{x} \in \mathbf{R}^n$ and let $\bar{x}$ be a solution of the equation (2.137). Let $\{x^k\} \subset \mathbf{R}^n$ be a sequence convergent to $\bar{x}$, and for each $k = 0, 1, \ldots,$ let*

$$\omega^k = -\Phi(x^k) - J_k(x^{k+1} - x^k) \tag{2.154}$$

with some $J_k \in \partial\Phi(x^k)$.

If the rate of convergence of $\{x^k\}$ is superlinear, then there exists a sequence $\{\theta_k\} \subset \mathbf{R}$ satisfying condition (2.153), and such that $\theta_k \to 0$.

Conversely, if Φ is CD-regular at $\bar{x}$, and there exists a sequence $\{\theta_k\} \subset \mathbf{R}$ satisfying condition (2.153), and such that $\theta_k \to 0$, then the rate of convergence of $\{x^k\}$ is superlinear. The rate of convergence is quadratic, provided Φ is strongly semismooth at $\bar{x}$ and

$$\theta_k = O(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \tag{2.155}$$

as $k \to \infty$.

Proof. To prove the first assertion, observe that by Proposition 1.64 and Remark 1.65, for all k large enough it holds that

$$\|x^k - \bar{x}\| = O(\|\Phi(x^k)\|) \tag{2.156}$$

as $k \to \infty$ (recall that semismoothness of Φ at $\bar{x}$ implies directional differentiability of Φ at $\bar{x}$ in every direction, so that Proposition 1.64 is applicable). By Proposition 2.44 and Remark 2.45, the superlinear convergence rate of $\{x^k\}$ implies the estimate

$$\omega^k = o(\|x^k - \bar{x}\|) \tag{2.157}$$

as $k \to \infty$. Combining the latter with (2.156), we have that

$$\omega^k = o(\|\Phi(x^k)\|)$$

as $k \to \infty$, which means precisely the existence of a sequence $\{\theta_k\}$ with the claimed properties.

The second assertion follows from Proposition 2.44, since the relation (2.153) with $\theta_k \to 0$ evidently implies (2.157) (and, hence, (2.150)):

$$\omega^k = o(\|\Phi(x^k) - \Phi(\bar{x})\|) = o(\|x^k - \bar{x}\|)$$

as $k \to \infty$, where local Lipschitz-continuity of Φ at $\bar{x}$ was employed.

Finally, if Φ is strongly semismooth at $\bar{x}$, and (2.155) holds, quadratic convergence follows by the last assertion of Proposition 2.44. □

Theorem 2.51. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be semismooth at $\bar{x} \in \mathbf{R}^n$, let $\bar{x}$ be a solution of the equation (2.137), and assume that Φ is CD-regular at $\bar{x}$.*

Then there exists $\theta > 0$ such that for any starting point $x^0 \in \mathbf{R}^n$ close enough to $\bar{x}$ and any sequences $\{x^k\} \subset \mathbf{R}^n$, $\{\omega^k\} \subset \mathbf{R}^n$ and $\{\theta_k\} \subset [0, \theta]$ satisfying (2.153) and (2.154) with some $J_k \in \partial\Phi(x^k)$ for all $k = 0, 1, \ldots$, it holds that $\{x^k\}$ converges to $\bar{x}$ and the rate of convergence is (at least) linear. Moreover, the rate of convergence is superlinear if $\theta_k \to 0$. The rate of convergence is quadratic, provided Φ is strongly semismooth at $\bar{x}$ and provided $\theta_k = O(\|\Phi(x^k)\|)$ as $k \to \infty$.

Proof. Let $L > 0$ stand for the Lipschitz constant of Φ with respect to $\bar{x}$. Similarly to the proof of Theorem 2.42, we can choose a neighborhood U of $\bar{x}$ and $M > 0$ such that (2.145) holds. Take any $\theta < 1/(LM)$, and any $q \in (\theta LM, 1)$. By (2.153), (2.154) and the inclusion $\theta^k \in [0, \theta]$, employing assertion (a) of Theorem 2.42, we obtain that there exists $\delta > 0$ such that $B(\bar{x}, \delta) \subset U$, and for any $x^k \in B(\bar{x}, \delta)$ it holds that

$$\begin{aligned}
\|x^{k+1} - \bar{x}\| &= \|x^k - J_k^{-1}(\Phi(x^k) + \omega^k) - \bar{x}\| \\
&\leq \|J_k^{-1}\omega^k\| + \|x^k - J_k^{-1}\Phi(x^k) - \bar{x}\| \\
&\leq \theta_k\|J_k^{-1}\|\|\Phi(x^k)\| + o(\|x^k - \bar{x}\|) \\
&\leq \theta M\|\Phi(x^k) - \Phi(\bar{x})\| + o(\|x^k - \bar{x}\|) \\
&\leq \theta LM\|x^k - \bar{x}\| + o(\|x^k - \bar{x}\|) \\
&\leq q\|x^k - \bar{x}\|. \qquad (2.158)
\end{aligned}$$

Since $q \in (0, 1)$, this estimate implies the inclusion $x^{k+1} \in B(\bar{x}, \delta)$. Therefore, the inclusion $x^0 \in B(\bar{x}, \delta)$ implies that the entire sequence $\{x^k\}$ is contained in $B(\bar{x}, \delta)$, and (2.158) shows convergence of this sequence to $\bar{x}$ at a linear rate.

Superlinear rate of convergence when $\theta_k \to 0$, and quadratic rate when Φ is strongly semismooth at $\bar{x}$ and $\theta_k = O(\|\Phi(x^k)\|)$ as $k \to \infty$, follow from Theorem 2.50. □

The fact that in Theorem 2.51 one cannot simply take an arbitrary $\theta \in (0, 1)$ is confirmed by the following counterexample, borrowed from [68, Example 7.5.6].

Example 2.52. Let $n = 2$,

$$\Phi_1(x) = \begin{cases} 2x_1 & \text{if } x_1 \geq 0, \\ x_1 & \text{if } x_1 < 0, \end{cases} \qquad \Phi_2(x) = \begin{cases} x_2 & \text{if } x_2 \geq 0, \\ 2x_2 & \text{if } x_2 < 0. \end{cases}$$

The unique solution of $\Phi(x) = 0$ is $\bar{x} = 0$, and Φ is locally Lipschitz-continuous at 0 and differentiable everywhere except for those points satisfying $x_1 = 0$ or $x_2 = 0$. Evidently, $\partial_B\Phi(0)$ consists of the four matrices

$$\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

In particular, Φ is CD-regular at 0.

Fix any $\delta > 0$ and define the sequence $\{x^k\}$ as follows:

$$x^k = \begin{cases} (\delta,\ \delta/2) & \text{if } k \text{ is even}, \\ (-\delta/2,\ -\delta) & \text{if } k \text{ is odd}. \end{cases}$$

Then Φ is differentiable at each x^k, and it can be directly checked that for all k

$$\|\Phi(x^k) + \Phi'(x^k)(x^{k+1} - x^k)\| = \sqrt{2}\delta,$$

while at the same time

$$\|\Phi(x^k)\| = \frac{\sqrt{17}\delta}{2}.$$

Therefore, if we take any $\theta \geq 2\sqrt{2}/\sqrt{17}$, then

$$\|\Phi(x^k) + \Phi'(x^k)(x^{k+1} - x^k)\| \leq \theta\|\Phi(x^k)\|.$$

Hence, the sequence $\{x^k\}$ can be generated by the truncated semismooth Newton method with $\theta_k = \theta$ for all k. However, $\{x^k\}$ oscillates between the two points distinct from 0, no matter how small δ is (and hence, no matter how close x^0 is to $\bar{x}$).

Various important modifications of semismooth Newton methods arise from the natural intention to develop globally convergent schemes. Far from a CD-regular solution, the generalized Jacobian may contain singular matrices. As a remedy, consider the following *semismooth Levenberg–Marquardt method*: for a given $x^k \in \mathbf{R}^n$, the next iterate $x^{k+1} \in \mathbf{R}^n$ is defined by the equation

$$J_k^{\mathrm{T}}\Phi(x^k) + (J_k^{\mathrm{T}}J_k + \nu_k I)(x - x^k) = 0 \tag{2.159}$$

with some $J_k \in \partial\Phi(x^k)$ and some $\nu_k \geq 0$.

For $\nu_k = 0$, the semismooth Levenberg–Marquardt method reduces to the *semismooth Gauss–Newton method* with the iteration system

$$J_k^{\mathrm{T}}\Phi(x^k) + J_k^{\mathrm{T}}J_k(x - x^k) = 0. \tag{2.160}$$

One potential advantage of the latter compared to the basic iteration system (2.138) is the symmetry of the iteration matrix in (2.160), the property that can be efficiently employed in the methods used to solve this system. In particular, solving (2.160) is equivalent to solving the linear least-squares problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|\Phi(x^k) + J_k(x - x^k)\|^2 \\ \text{subject to} & x \in \mathbf{R}^n. \end{array}$$

See the related discussion for the Gauss–Newton method for smooth equations in Sect. 2.1.1.

Similarly, solving (2.159) is equivalent to solving the unconstrained optimization problem

$$\begin{array}{ll} \text{minimize} & \dfrac{1}{2}\|J_k(x-x^k)+\Phi(x^k)\|^2+\dfrac{\nu_k}{2}\|x-x^k\|^2 \\ \text{subject to} & x\in\mathbf{R}^n, \end{array}$$

whose quadratic objective function is strongly convex when $\nu_k > 0$. As in the smooth case, smart selection of ν_k is essential for efficiency of the semismooth Levenberg–Marquardt method. One possibility arises in the context of trust-region methods; see Sect. 2.3. Speaking about local convergence theory, the following result can be derived by combining considerations similar to those in the proof of Theorem 2.11, with the proof of Theorem 2.51. We omit the details.

Theorem 2.53. *Let $\Phi:\mathbf{R}^n\to\mathbf{R}^n$ be semismooth at $\bar{x}\in\mathbf{R}^n$, let $\bar{x}$ be a solution of the equation (2.137), and assume that Φ is CD-regular at $\bar{x}$.*

Then there exists a constant $\nu > 0$ with the property that for any sequence $\{\nu_k\}\subset[0,\nu]$, any starting point $x^0\in\mathbf{R}^n$ close enough to $\bar{x}$ defines an iterative sequence $\{x^k\}\subset\mathbf{R}^n$ such that for each $k=0,1,\ldots,$ the point x^{k+1} satisfies (2.159) with some $J_k\in\partial\Phi(x^k)$; any such sequence converges to $\bar{x}$, and the rate of convergence is (at least) linear. Moreover, the rate of convergence is superlinear if $\nu_k\to 0$. The rate of convergence is quadratic, provided Φ is strongly semismooth at $\bar{x}$ and $\nu_k=O(\|\Phi(x^k)\|)$ as $k\to\infty$.

The very natural next step (which we again only outline without going into details) is the development of a truncated semismooth Levenberg–Marquardt method, where the iteration system (2.159) is replaced by

$$J_k^{\mathrm{T}}\Phi(x^k)+(J_k^{\mathrm{T}}J_k+\nu_k I)(x-x^k)+\omega^k=0$$

with the perturbation terms $\omega^k\in\mathbf{R}^n$ satisfying the truncation condition (2.153) for some forcing sequence $\{\theta_k\}$. Combining the ideas contained in Theorems 2.51 and 2.53, local superlinear convergence to a solution satisfying CD-regularity can be established if both ν_k and θ_k are kept small enough.

We note that there are various other perturbed/modified versions of semismooth Newton methods in the literature; one interesting example is given in [138].

Remark 2.54. The CD-regularity assumption in Theorem 2.42 can be somewhat weakened if we assume that J_k is chosen in step 2 of Algorithm 2.41 not from $\partial\Phi(x^k)$ but from the generally smaller set $\partial_B\Phi(x^k)$. Specifically, CD-regularity can be replaced by BD-regularity of Φ at $\bar{x}$. The proof of the corresponding counterpart of Theorem 2.42 is exactly the same, but with the reference to the part of Proposition 1.51 concerning B-differentials.

Moreover, sometimes even more special choices of generalized Jacobians can be useful, e.g., employing the specific structure of Φ. For a generic scheme of this kind, suppose that $J_k \in \Delta(x^k)$ for each k, where Δ is a given multifunction from $\mathbf{R}^n$ to the subsets of $\mathrm{R}^{n\times n}$ such that $\Delta(x) \subset \partial\Phi(x)$ for all $x \in \mathbf{R}^n$. Then CD-regularity can be replaced by the assumption of nonsingularity of any matrix in the set

$$\bar{\Delta}(\bar{x}) = \left\{ J \in \mathbf{R}^{n\times n} \;\middle|\; \begin{array}{l} \exists\, \{x^k\} \subset \mathbf{R}^n,\, \{J_k\} \subset \mathbf{R}^{n\times n} :\; J_k \in \Delta(x^k)\ \forall\, k, \\ \{x^k\} \to x,\, \{J_k\} \to J \end{array} \right\}.$$

By Proposition 1.51, it holds that $\bar{\Delta}(\bar{x}) \subset \partial\Phi(\bar{x})$. Hence, this regularity assumption is implied by CD-regularity, and it can be strictly weaker than CD-regularity, depending on the choice of $\Delta(\cdot)$.

Similar modifications apply not only to Theorem 2.42, but to all results of this section.

Remark 2.55. An approach originally suggested in [172] is closely related to the semismooth Newton method, but its domain of application is much narrower. Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be *piecewise smooth* on an open set $O \subset \mathbf{R}^n$, which means that Φ is continuous and there exists a finite collection of continuously differentiable mappings $\Phi_i : O \to \mathbf{R}^n$, $i = 1, \ldots, m$, called smooth pieces of Φ, such that for any $x \in O$ the set

$$I(x) = \{i = 1, \ldots, m \mid \Phi(x) = \Phi_i(x)\}$$

of indices of smooth pieces active at x is nonempty. Then the set

$$\hat{\partial}\Phi(x) = \{\Phi_i'(x) \mid i \in I(x)\}$$

is nonempty as well, and one can define the *piecewise Newton method* for solving the equation (2.137) by the iterative system (2.138) with $J_k \in \hat{\partial}\Phi(x^k)$ (provided $x^k \in O$).

Assuming now that $\bar{x} \in \mathbf{R}^n$ is a solution of the equation (2.137), that Φ is piecewise smooth on some neighborhood O of $\bar{x}$, and that all the matrices in $\hat{\partial}\Phi(\bar{x})$ are nonsingular, the local superlinear convergence of this method to $\bar{x}$ follows easily from the fact that $I(x) \subset I(\bar{x})$ for all $x \in O$ close enough to $\bar{x}$. Indeed, this fact implies that, locally, each iteration of the piecewise Newton method can be interpreted as the usual Newtonian iteration for the smooth equation

$$\Phi_i(x) = 0$$

for some $i \in I(\bar{x})$, and $\bar{x}$ is a solution of this equation as well. The convergence result can now be derived from Theorem 2.2.

It is evident that

$$\partial_B\Phi(x) \subset \hat{\partial}\Phi(x) \quad \forall\, x \in O.$$

Employing this relation, it can be easily seen that piecewise smoothness of Φ on O implies semismoothness of Φ at any point of O, and for equations with piecewise smooth mappings, the semismooth Newton method employing B-differentials (see Remark 2.54) can be regarded as a special realization of the piecewise smooth Newton method. Computing an element of $\hat{\partial}\Phi(x^k)$ can be much simpler than computing an element of $\partial_B\Phi(x^k)$. The price paid for this is the necessity to assume nonsingularity of all the matrices in $\hat{\partial}\Phi(\bar{x})$, which can be a strictly stronger assumption than BD-regularity of Φ at $\bar{x}$. This issue will be illustrated later on by Example 3.15 in the context of complementarity problems.

An important particular instance of a piecewise smooth mapping is a *piecewise linear* (*piecewise affine*) mapping, which has linear (affine) pieces Φ_i, $i = 1, \ldots, m$.

2.4.2 Semismooth Newton Method for Unconstrained Optimization

One possible application of semismooth Newton methods is an unconstrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathbf{R}^n, \end{array} \tag{2.161}$$

with a differentiable objective function $f : \mathbf{R}^n \to \mathbf{R}$ whose gradient is Lipschitz-continuous, but the second derivatives of f may not exist. Similarly to Sect. 2.1.2, stationary points of this problem are characterized by the equation (2.137) with $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ being the gradient mapping of f:

$$\Phi(x) = f'(x).$$

This Φ is Lipschitz-continuous. Thus, one can try to search for stationary points of the optimization problem (2.161) by applying the semismooth Newton-type method to the equation (2.137).

Given the current iterate $x^k \in \mathbf{R}^n$, the next iterate x^{k+1} of the basic semismooth Newton method for (2.161) is computed as a solution of the linear system

$$f'(x^k) + H_k(x - x^k) = 0 \tag{2.162}$$

with H_k being an element of the generalized Hessian of f at x^k, that is, $H_k \in \partial f'(x^k)$. Assuming that H_k is nonsingular, the semismooth Newton method can be written in the form of the explicit iterative scheme

$$x^{k+1} = x^k - H_k^{-1} f'(x^k), \quad H_k \in \partial f'(x^k), \quad k = 0, 1, \ldots.$$

As usual, the optimization nature of the original problem can be put to the foreground by observing that solutions of (2.162) coincide with stationary points of the subproblem

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, x - x^k \rangle + \dfrac{1}{2}\langle H_k(x - x^k),\, x - x^k \rangle \\ \text{subject to} & x \in \mathbf{R}^n. \end{array} \tag{2.163}$$

Therefore, the basic *semismooth Newton method* for unconstrained optimization is stated as follows.

Algorithm 2.56 Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $f'(x^k) = 0$, stop.
2. Compute some $H_k \in \partial f'(x^k)$. Compute $x^{k+1} \in \mathbf{R}^n$ as a stationary point of problem (2.163).
3. Increase k by 1 and go to step 1.

Local convergence properties of the semismooth Newton method for unconstrained optimization follow directly from Theorem 2.42.

Theorem 2.57. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be differentiable in a neighborhood of a point $\bar{x} \in \mathbf{R}^n$, with its gradient being locally Lipschitz-continuous at $\bar{x}$, and assume that the estimate*

$$\sup_{H \in \partial f'(x+\xi)} \|f'(x+\xi) - f'(x) - H\xi\| = o(\|\xi\|)$$

holds as $\xi \in \mathbf{R}^n$ tends to 0. Let $\bar{x}$ be a stationary point of problem (2.161), and assume that this point satisfies the SOSC

$$\forall\, H \in \partial f'(\bar{x}) \quad \langle H\xi,\, \xi \rangle > 0 \quad \forall\, \xi \in \mathbf{R}^n \setminus \{0\} \tag{2.164}$$

(thus, according to Theorem 1.81, $\bar{x}$ is a strict local solution of problem (2.161)).

Then for any starting point $x^0 \in \mathbf{R}^n$ close enough to $\bar{x}$, any iterative sequence generated by Algorithm 2.56 converges to $\bar{x}$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the estimate

$$\sup_{J \in \partial \Phi(x+\xi)} \|\Phi(x+\xi) - \Phi(x) - J\xi\| = O(\|\xi\|^2)$$

holds as $\xi \in \mathbf{R}^n$ tends to 0.

From the symmetry of Hessians of a twice differentiable function, and from the definition of the generalized Hessian, it follows that the latter consists of symmetric matrices as well. This specificity can be used when solving the linear iteration system (2.55); see the related comments in Sect. 2.1.2.

Also, similarly to the smooth case, the assertion of Theorem 2.57 remains valid if the SOSC (2.164) is replaced by the weaker assumption of *CD*-regularity of the gradient mapping f' at $\bar{x}$. As in the smooth case, in this respect the semismooth Newton method does not distinguish local minimizers from other stationary points of the problem.

Needless to say, the perturbed versions of the semismooth Newton method discussed in Sect. 2.4.1 can be directly adapted for unconstrained optimization, with the same motivation as in Sect. 2.4.1.

Semismooth Newton methods for unconstrained optimization can be globalized along the lines of Sects. 2.2 and 2.3, employing f as the natural merit function. In particular, the global convergence analysis of linesearch methods presented in Sects. 2.2.1 and 2.2.2 is fully applicable in this context. An interesting related observation is stated in [67]. Let f be differentiable in a neighborhood of a stationary point $\bar{x}$ of problem (2.58), with its gradient being semismooth at $\bar{x}$, and assume that the SOSC (2.164) holds. Let $\{x^k\}$ be an iterative sequence of the semismooth Newton method equipped with the Armijo or Wolfe linesearch rule with the initial trial stepsize value $\alpha = 1$ (plus some additional requirements regarding the parameters involved in those linesearch rules). Then if $\{x^k\}$ converges to $\bar{x}$, the method eventually accepts the unit stepsize. Hence, the convergence rate is superlinear.

Chapter 3
Variational Problems: Local Methods

In this chapter, we present local analysis of Newton-type algorithms for variational problems, starting with the fundamental Josephy–Newton method for generalized equations. This method is an important extension of classical Newtonian techniques to more general variational problems. For example, as a specific application, the Josephy–Newton method provides a convenient tool for analyzing the sequential quadratic programming (SQP) algorithm for optimization. In fact, the sharpest local convergence results for SQP algorithms available in the literature follow precisely from general properties of the Josephy–Newton method. Having in mind also other optimization algorithms, such as the linearly constrained (augmented) Lagrangian methods, sequential quadratically constrained quadratic programming, and the stabilized version of SQP, we shall develop a general perturbed Josephy–Newton framework that would be a convenient tool in Chap. 4 for treating in a unified manner, in addition to SQP, those other (sometimes seemingly unrelated) optimization techniques.

This chapter also includes the semismooth Newton methods for equations arising from reformulations of (mixed) complementarity conditions. Semismooth Newton methods extend classical Newtonian techniques by relaxing the differentiability assumptions. This relaxation is particularly relevant in applications arising from complementarity, where the associated equation formulations often exhibit natural structural nondifferentiability. The extension of semismooth Newton methods to generalized equations will also be discussed.

Finally, active-set Newton methods are considered. Those methods are based on local identification of some structural properties of a solution, with the aim of making the problem easier to solve by enforcing some inequalities as equalities.

A.F. Izmailov and M.V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-3-319-04247-3_3,

3.1 Josephy–Newton Method

In this section, we discuss Newton-type methods for generalized equations (GE) of the form

$$\Phi(x) + N(x) \ni 0, \tag{3.1}$$

where $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ is a smooth (single-valued) mapping, and $N(\cdot)$ is a set-valued mapping from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$. The so-called *Josephy–Newton method* was originally proposed in [16] (assuming monotonicity of Φ) and in [161, 162] for the GE (3.1) corresponding to a variational inequality (VI), i.e., for the case where $N(\cdot)$ is the normal-cone multifunction associated with a given closed convex set. However, it can be seen that the VI specificity plays no role in the development and analysis of the basic scheme. Therefore, we shall first consider the Josephy–Newton method for an abstract GE of the form (3.1), introducing VI, complementarity and optimization structures later on, as needed.

For the current iterate $x^k \in \mathbf{R}^n$, the next iterate x^{k+1} is computed as a solution of (partial) linearization of the GE (3.1) at the point x^k:

$$\Phi(x^k) + \Phi'(x^k)(x - x^k) + N(x) \ni 0. \tag{3.2}$$

The sharpest currently available theory of local convergence and rate of convergence of the Josephy–Newton method to a given solution of the GE (3.1) is developed in [26], which we essentially follow in the first part of this section. This theory relies on the following two notions. The first one is semistability, as introduced in Definition 1.29. It can be shown (see Proposition 3.4 below) that if the Josephy–Newton method generates an iterative sequence $\{x^k\}$ convergent to a semistable solution $\bar{x}$, the rate of convergence is necessarily superlinear. However, semistability does not guarantee solvability of subproblems (3.2), even for x^k arbitrarily close to $\bar{x}$. To this end, the following notion comes into play.

Definition 3.1. A solution $\bar{x}$ of the GE (3.1) is said to be *hemistable* if for any $x \in \mathbf{R}^n$ close enough to $\bar{x}$, the GE

$$\Phi(x) + \Phi'(x)\xi + N(x + \xi) \ni 0 \tag{3.3}$$

has a solution $\xi(x)$ such that $\xi(x) \to 0$ as $x \to \bar{x}$.

We next state local convergence properties of the Josephy–Newton method. We shall not give a proof, as Theorem 3.2 will be an immediate corollary of a more general result concerning a perturbed version of the Josephy–Newton method, established later on (see Theorem 3.6).

Theorem 3.2. *Let a mapping $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$, and let $N(\cdot)$ be a set-valued mapping from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$. Let $\bar{x}$ be a semistable and hemistable solution of the GE (3.1).*

Then there exists $\delta > 0$ such that for any starting point $x^0 \in \mathbf{R}^n$ close enough to $\bar{x}$, there exists a sequence $\{x^k\} \subset \mathbf{R}^n$ such that x^{k+1} is a solution of the GE (3.2) for each $k = 0, 1, \ldots$, satisfying $\|x^{k+1} - x^k\| \leq \delta$; any such sequence converges to $\bar{x}$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$.

It can be easily seen from Theorem 1.24 that both semistability and hemistability are implied by strong regularity of the solution in question. Concerning semistability, this has already been observed in Sect. 1.3.2. Moreover, strong regularity implies that for any $x \in \mathbf{R}^n$ close enough to $\bar{x}$, the GE (3.3) has the unique solution near $\bar{x}$.

In the case of the usual equation

$$\Phi(x) = 0, \tag{3.4}$$

(i.e., when $N(\cdot) \equiv \{0\}$), both strong regularity and semistability are equivalent to saying that $\Phi'(\bar{x})$ is a nonsingular matrix, and the latter is evidently sufficient for hemistability. Moreover, in this case, the Josephy–Newton method is the usual Newton method for (3.4), and the assertion of Theorem 3.2 can be sharpened as follows: $\delta > 0$ can be taken arbitrarily, and for any $x^0 \in \mathbf{R}^n$ close enough to $\bar{x}$, the corresponding iterative sequence $\{x^k\}$ is unique. With these specifications, Theorem 3.2 reduces to the local convergence and rate of convergence result for the classical Newton method, stated above in Theorem 2.2.

In the case of a VI

$$x \in Q, \quad \langle \Phi(x),\, y - x\rangle \geq 0 \quad \forall\, y \in Q \tag{3.5}$$

with some closed convex set $Q \subset \mathbf{R}^n$ (i.e., when $N(\cdot) = N_Q(\cdot)$), the iteration subproblem (3.2) of the Josephy–Newton method takes the form of the VI

$$x \in Q, \quad \langle \Phi(x^k) + \Phi'(x^k)(x - x^k),\, y - x\rangle \geq 0 \quad \forall\, y \in Q \tag{3.6}$$

with the linearized mapping. In the further special case of the nonlinear complementarity problem (NCP)

$$x \geq 0, \quad \Phi(x) \geq 0, \quad \langle x,\, \Phi(x)\rangle = 0, \tag{3.7}$$

corresponding to (3.5) with $Q = \mathbf{R}^n_+$, the subproblem (3.6) becomes the linear complementarity problem

$$x \geq 0, \quad \Phi(x^k) + \Phi'(x^k)(x - x^k) \geq 0, \quad \langle x,\, \Phi(x^k) + \Phi'(x^k)(x - x^k)\rangle = 0. \tag{3.8}$$

Recall that the characterization of semistability for NCP is given in Proposition 1.34.

The following local interpretation of the Josephy–Newton method for NCP is quite instructive. For a given solution $\bar{x}$ of NCP (3.7), define the index sets

$$\begin{aligned} I_0(\bar{x}) &= \{i = 1, \ldots, n \mid \bar{x}_i = \Phi_i(\bar{x}) = 0\}, \\ I_1(\bar{x}) &= \{i = 1, \ldots, n \mid \bar{x}_i > 0,\ \Phi_i(\bar{x}) = 0\}, \\ I_2(\bar{x}) &= \{i = 1, \ldots, n \mid \bar{x}_i = 0,\ \Phi_i(\bar{x}) > 0\}, \end{aligned}$$

and let $\{x^k\}$ be an iterative sequence of the Josephy–Newton method, convergent to $\bar{x}$. Then from (3.8) it follows that for all k large enough there exists a partition $(I_1^k,\, I_2^k)$ of $I_0(\bar{x})$ (i.e., $I_1^k \cup I_2^k = I_0(\bar{x})$, $I_1^k \cap I_2^k = \emptyset$) such that

$$\Phi_{I_1(\bar{x})\cup I_1^k}(x^k) + \Phi'_{I_1(\bar{x})\cup I_1^k}(x^k)(x^{k+1} - x^k) = 0, \quad x^{k+1}_{I_2(\bar{x})\cup I_2^k} = 0. \tag{3.9}$$

Therefore, the Josephy–Newton iteration can be interpreted as the usual Newton iteration for the system of equations

$$\Phi_{I_1(\bar{x})\cup I_1^k}(x) = 0, \quad x_{I_2(\bar{x})\cup I_2^k} = 0. \tag{3.10}$$

If the *strict complementarity condition* $I_0(\bar{x}) = \emptyset$ holds, then system (3.10) takes the form

$$\Phi_{I_1(\bar{x})}(x) = 0, \quad x_{I_2(\bar{x})} = 0,$$

and in this case, the Josephy–Newton method reduces to the usual Newton method for this system of equations.

However, if the strict complementarity condition does not hold, the partition $(I_1^k,\, I_2^k)$ (and hence, system (3.10)) can change along the iterations. Therefore, the Josephy–Newton method does not reduce to the Newton method for any fixed system of equations in this case.

Generally, neither of the two properties of semistability and hemistability is implied by the other. In fact, semistability is not implied by hemistability even for usual single-valued equations. Indeed, if $N(\cdot) \equiv \{0\}$ and, e.g., $\Phi(\cdot) \equiv 0$, then (3.1) is satisfied by any $\bar{x} \in \mathbf{R}^n$, and any solution is trivially hemistable but, of course, not semistable. The next example demonstrates that beyond the case of usual equations, the converse implication is also not valid, even for the special case of an NCP.

Example 3.3. Consider the NCP (3.7) with $n = 1$ and $\Phi(x) = -x + x^2/2$. This NCP gives the necessary optimality condition for the optimization problem

$$\begin{array}{ll} \text{minimize} & -\dfrac{1}{2}x^2 + \dfrac{1}{6}x^3 \\ \text{subject to} & x \geq 0 \end{array}$$

(see Theorem 1.10). This problem has the unique local solution $\hat{x} = 2$, which is also its global solution. However, in addition to $\hat{x}$, the NCP (3.7) has one more solution, namely $\bar{x} = 0$.

For each $r \in \mathbf{R}$, the perturbed NCP

$$x \geq 0, \quad \Phi(x) \geq r, \quad \langle x,\, \Phi(x) - r\rangle = 0$$

has the following solutions tending to $\bar{x}$ as $r \to 0$: $x^1(r) = 0$ for all $r \geq 0$, and also $x^2(r) = 1 - \sqrt{1-2r}$ for $r \in (0,\, 1/2)$. This implies semistability of $\bar{x}$.

It can be easily checked that the NCP corresponding to the GE (3.3) can be written in the form

$$x+\xi \geq 0, \quad -(1-x)(x+\xi) - \frac{1}{2}x^2 \geq 0, \quad (x+\xi)\left((1-x)(x+\xi) + \frac{1}{2}x^2\right) = 0,$$

and the latter has no solutions for any $x \neq 0$ such that $x \leq 1$. Thus, $\bar{x}$ is not hemistable.

At the same time, it will be shown in Proposition 3.37 that for GEs arising from KKT systems of optimization problems, the notions of semistability and hemistability of a primal-dual solution are equivalent, at least when the primal part is a local solution of the optimization problem.

Local convergence results related to Theorem 3.2 can be found in [62, Sect. 6C], where, however, the basic assumption is the classical property of metric regularity. Generally, metric regularity is neither weaker nor stronger than semistability. At the same time, metric regularity allows to dispense with the explicit assumption of solvability of subproblems, which is of course a good feature. However, the price paid to have solvability automatically is actually quite high, as for variational problems metric regularity is in fact a quite strong requirement: as demonstrated in [61], for the VI (3.5) with a smooth mapping Φ over a polyhedral set Q (thus including systems of equations with the same number of equations and variables, and the cases of an NCP and of a KKT system), metric regularity is in fact equivalent to strong regularity of a solution in question. Therefore, for these important special cases, the results in [62, Corollary 6C.5, Theorem 6C.6] do not go much beyond the original development in [161, 162]. On the other hand, metric regularity allows to consider nonisolated solutions (e.g., underdetermined systems of equations, which generically have nonisolated solutions). In this case, developments under metric regularity are more related to our Sect. 7.2 below. The assertion of [62, Corollary 6C.5, (i)] states that for a given solution, there exists an iterative sequence convergent to this solution, without any possibility to characterize how this "good" iterative sequence can be separated from "bad" iterative sequences. This kind of statement is theoretically interesting and important, but it is in a sense "impractical." By itself, it can hardly be regarded as a final local convergence result. To that end, it is complemented in [62, Corollary 6C.5, (ii), (iii)] by results assuming semistability and strong metric regularity. This gives quite a nice and self-concordant picture but, as mentioned above, the first result corresponds to the analysis in [161, 162] while the second, combined with [62, Corollary 6C.5, (i)], is no sharper than Theorem 3.2.

In what follows, we shall deal with algorithms that can be regarded as a *perturbed Josephy–Newton method*: instead of (3.2), the next iterate x^{k+1} would satisfy the GE

$$\Phi(x^k) + \Phi'(x^k)(x - x^k) + \omega^k + N(x) \ni 0, \tag{3.11}$$

where $\omega^k \in \mathbf{R}^n$ is a perturbation term. This generalized form would be a useful tool for analyzing a number of Newton-related optimization methods that do not fit into the framework of the exact Josephy–Newton iterations. The results presented in the rest of this section were derived in [148].

It is convenient to start with the following a posteriori result concerned with superlinear rate of convergence, assuming convergence itself.

Proposition 3.4. *Let a mapping $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$, and let $N(\cdot)$ be a set-valued mapping from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$. Let $\bar{x}$ be a semistable solution of the GE (3.1). Let a sequence $\{x^k\} \subset \mathbf{R}^n$ be convergent to $\bar{x}$, and assume that x^{k+1} satisfies (3.11) for each $k = 0, 1, \ldots$, with some $\omega^k \in \mathbf{R}^n$ such that*

$$\omega^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \tag{3.12}$$

as $k \to \infty$.

Then the rate of convergence of $\{x^k\}$ is superlinear. Moreover, the rate of convergence is quadratic provided the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$ and

$$\omega^k = O(\|x^{k+1} - x^k\|^2 + \|x^k - \bar{x}\|^2) \tag{3.13}$$

as $k \to \infty$

Proof. For each k, the point x^{k+1} is a solution of the GE

$$\Phi(x) + N(x) \ni r \tag{3.14}$$

with

$$r = r^k = \Phi(x^{k+1}) - \Phi(x^k) - \Phi'(x^k)(x^{k+1} - x^k) - \omega^k, \tag{3.15}$$

and according to the mean-value theorem (see Theorem A.10, (a)) and (3.12), it holds that

$$\begin{aligned}\|r^k\| &\leq \sup\{\|\Phi'(tx^{k+1} + (1-t)x^k) - \Phi'(x^k)\| \mid t \in [0, 1]\}\|x^{k+1} - x^k\| \\ &\quad + \|\omega^k\| \\ &= o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)\end{aligned} \tag{3.16}$$

as $k \to \infty$. Then, by the semistability of $\bar{x}$, it holds that

$$\begin{aligned} x^{k+1} - \bar{x} &= O(\|r^k\|) \\ &= o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \\ &= o(\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|) \end{aligned}$$

as $k \to \infty$. By the same argument as in the proof of Proposition 2.4, the latter implies the superlinear convergence rate.

Furthermore, if the derivative of Φ is locally Lipschitz-continuous at $\bar{x}$, from (3.13) and (3.15) it follows that the estimate (3.16) can be sharpened as follows:

$$r^k = O(\|x^{k+1} - x^k\|^2 + \|x^k - \bar{x}\|^2)$$

as $k \to \infty$. Then, by semistability of $\bar{x}$, we obtain that

$$\begin{aligned} x^{k+1} - \bar{x} &= O(\|r^k\|) \\ &= O(\|x^{k+1} - x^k\|^2 + \|x^k - \bar{x}\|^2) \\ &= O(\|x^{k+1} - \bar{x}\|^2 + \|x^k - \bar{x}\|^2) \end{aligned}$$

as $k \to \infty$. By the same argument as in the proof of Proposition 2.4, the latter implies the quadratic convergence rate. $\square$

As an immediate application of Proposition 3.4, consider the following *quasi-Josephy–Newton method*. Let $\{J_k\} \subset \mathbf{R}^{n \times n}$ be a sequence of matrices. For a current $x^k \in \mathbf{R}^n$, let the next iterate x^{k+1} be computed as a solution of the GE

$$\Phi(x^k) + J_k(x - x^k) + N(x) \ni 0, \tag{3.17}$$

and assume that $\{J_k\}$ satisfies the *Dennis–Moré condition* (cf. (2.32) and (2.80), (2.81)):

$$(J_k - \Phi'(x^k))(x^{k+1} - x^k) = o(\|x^{k+1} - x^k\|) \tag{3.18}$$

as $k \to \infty$. The GE (3.17) can be interpreted as (3.11) with

$$\omega^k = (J_k - \Phi'(x^k))(x^{k+1} - x^k).$$

Then, by (3.18), it follows that

$$\omega^k = o(\|x^{k+1} - x^k\|)$$

as $k \to \infty$, and hence, Proposition 3.4 implies the following a posteriori result.

Theorem 3.5. *Let a mapping $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$, and let $N(\cdot)$ be a set-valued mapping from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$. Let $\bar{x}$ be a semistable solution of the GE (3.1). Let $\{J_k\} \subset \mathbf{R}^{n \times n}$ be a sequence of matrices, and let $\{x^k\} \subset \mathbf{R}^n$ be a sequence convergent to $\bar{x}$ and such that x^{k+1} satisfies (3.17) for all k large enough. Assume, finally, that condition (3.18) holds.*

Then the rate of convergence of $\{x^k\}$ is superlinear.

As discussed in Sect. 2.1.1, in the context of the perturbed Newton method for usual equations a priori analysis is natural and often possible. However, for more complex GEs, solvability of the Josephy–Newton iteration subproblems (whether exact or perturbed) is generally impossible to establish, at least without stronger assumptions which can be avoided otherwise. That is why hemistability had to appear as an assumption in Theorem 3.2, making this result in a sense intermediate between a priori and a posteriori analysis: solvability of subproblems is assumed, having in mind that this can be verifiable separately for more specific algorithms and/or problem classes. Theorem 3.6 below is also of this intermediate nature. Recall that we have seen analysis of this kind for usual equations as well (Theorem 2.13).

Following the pattern of (2.48), we shall consider the perturbed Josephy–Newton method with iteration subproblems of the form

$$\Phi(x^k) + \Phi'(x^k)(x - x^k) + \Omega(x^k,\, x - x^k) + N(x) \ni 0. \tag{3.19}$$

Specifically, the perturbation term in (3.11) now conforms to the inclusion $\omega^k \in \Omega(x^k,\, x^{k+1} - x^k)$.

Theorem 3.6. *Let a mapping $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$, and let $N(\cdot)$ be a set-valued mapping from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$. Let $\bar{x}$ be a semistable solution of the GE (3.1). Let Ω be a multifunction from $\mathbf{R}^n \times \mathbf{R}^n$ to the subsets of $\mathbf{R}^n$, satisfying the following assumptions: for each $x \in \mathbf{R}^n$ close enough to $\bar{x}$, the GE*

$$\Phi(x) + \Phi'(x)\xi + \Omega(x,\, \xi) + N(x + \xi) \ni 0 \tag{3.20}$$

has a solution $\xi(x)$ such that $\xi(x) \to 0$ as $x \to \bar{x}$, and the estimate

$$\omega = o(\|\xi\| + \|x - \bar{x}\|) \tag{3.21}$$

holds as $x \to \bar{x}$ and $\xi \to 0$, uniformly for $\omega \in \Omega(x,\, \xi)$, $x \in \mathbf{R}^n$ and $\xi \in \mathbf{R}^n$ satisfying

$$\Phi(x) + \Phi'(x)\xi + \omega + N(x + \xi) \ni 0. \tag{3.22}$$

Then there exists $\delta > 0$ such that for any starting point $x^0 \in \mathbf{R}^n$ close enough to $\bar{x}$, there exists a sequence $\{x^k\} \subset \mathbf{R}^n$ such that x^{k+1} is a solution of the GE (3.19) for each $k = 0,\, 1,\, \ldots$, satisfying

$$\|x^{k+1} - x^k\| \le \delta; \tag{3.23}$$

any such sequence converges to $\bar{x}$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$, and provided (3.21) can be replaced by the estimate

$$\omega = O(\|\xi\|^2 + \|x - \bar{x}\|^2). \tag{3.24}$$

Proof. Semistability of $\bar{x}$ implies the existence of $\delta_1 > 0$ and $M > 0$ such that for any $r \in \mathbf{R}^n$ and any solution $x(r)$ of the GE (3.14) satisfying the condition $\|x(r) - \bar{x}\| \leq \delta_1$, it holds that

$$\|x(r) - \bar{x}\| \leq M\|r\|. \tag{3.25}$$

Fix any $\delta_2 \in (0,\, \delta_1]$. By the assumptions of the theorem, there exists some $\delta \in (0,\, 3\delta_2/5]$ such that the inequality $\|x^k - \bar{x}\| \leq 2\delta/3$ implies the existence of a solution x^{k+1} of the GE (3.19), such that $\|x^{k+1} - \bar{x}\| \leq \delta_2$. Then x^{k+1} is a solution of the GE (3.14) with $r = r^k$ defined in (3.15), and with some $\omega^k \in \Omega(x^k,\, x^{k+1} - x^k)$, and the inequality in (3.16) and condition (3.21) imply that

$$\|r^k\| \leq \frac{1}{5M}(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|), \tag{3.26}$$

perhaps for a smaller value of δ_2 (and hence, of δ). Since $\delta_2 \leq \delta_1$, (3.25) holds with $r = r^k$ for $x^{k+1} = x(r^k)$, and hence, taking also into account (3.26), we obtain that

$$\|x^{k+1} - \bar{x}\| \leq \frac{1}{5}\|x^{k+1} - x^k\| + \frac{1}{5}\|x^k - \bar{x}\| \leq \frac{1}{5}\|x^{k+1} - \bar{x}\| + \frac{2}{5}\|x^k - \bar{x}\|.$$

This implies the inequality

$$\|x^{k+1} - \bar{x}\| \leq \frac{1}{2}\|x^k - \bar{x}\|, \tag{3.27}$$

which, in particular, implies that

$$\|x^{k+1} - \bar{x}\| \leq \frac{1}{3}\delta. \tag{3.28}$$

Hence,

$$\|x^{k+1} - x^k\| \leq \|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\| \leq \delta.$$

We thus proved that if $\|x^k - \bar{x}\| \leq 2\delta/3$, then the GE (3.19) has a solution x^{k+1} satisfying (3.23).

Suppose now that $\|x^k - \bar{x}\| \leq 2\delta/3$, and x^{k+1} is any solution of the GE (3.19) satisfying (3.23). Then

$$\|x^{k+1} - \bar{x}\| \leq \|x^{k+1} - x^k\| + \|x^k - \bar{x}\| \leq \frac{5}{3}\delta \leq \delta_2 \leq \delta_1.$$

Thus, $x^{k+1} = x(r^k)$ satisfies (3.25) with $r = r^k$, and by the same argument as above, the latter implies (3.27) and (3.28).

Therefore, if $\|x^0 - \bar{x}\| \leq 2\delta/3$, then the next iterate x^1 can be chosen in such a way that (3.23) would hold with $k = 0$, and any such choice would give (3.27) and (3.28) with $k = 0$. The latter implies that $\|x^1 - \bar{x}\| \leq 2\delta/3$ and, hence, the next iterate x^2 can be chosen in such a way that (3.23) would hold with $k = 1$, and any such choice would give (3.27) and (3.28) with $k = 1$.

Continuing this argument, we obtain that there exists a sequence $\{x^k\}$ such that for each k, x^{k+1} is a solution of the GE (3.19) satisfying (3.23), and for any such sequence, (3.27) is valid for all k. But the latter implies that $\{x^k\}$ converges to $\bar{x}$. In order to complete the proof, it remains to refer to Proposition 3.4. $\square$

For the exact Josephy–Newton method (i.e., in the case when $\Omega(\cdot) \equiv \{0\}$), Theorem 3.6 reduces to Theorem 3.2 (in particular, the assumption of solvability of (3.20) reduces to hemistability). On the other hand, in the case of the usual equation (3.4) (i.e., when $N(\cdot) \equiv \{0\}$), Theorem 3.6 essentially reduces to Theorem 2.13.

3.2 Semismooth Newton Method for Complementarity Problems

We now discuss reformulations of complementarity problems, which give rise to equations whose mappings are naturally nondifferentiable but are semismooth. This is one important application where semismooth Newton methods come into play. We also give a comprehensive comparison of regularity conditions which are relevant in this context to ensure applicability of Newton methods.

3.2.1 Equation Reformulations of Complementarity Conditions

Consider first the basic complementarity condition with respect to a pair of scalar variables $(a, b) \in \mathbf{R} \times \mathbf{R}$:

$$a \geq 0, \quad b \geq 0, \quad ab = 0. \tag{3.29}$$

A function $\psi : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ is referred to as a *complementarity function* if the solution set of (3.29) coincides with solutions of the equation

$$\psi(a, b) = 0.$$

Two important and most widely used examples of complementarity functions are the *natural residual function* given by

$$\psi(a, b) = \min\{a, b\}, \tag{3.30}$$

and the *Fischer–Burmeister function* [82] given by

$$\psi(a, b) = a + b - \sqrt{a^2 + b^2}. \tag{3.31}$$

Complementarity functions provide a convenient tool for converting problems that involve complementarity conditions into equations. Consider the NCP

$$x \geq 0, \quad \Phi(x) \geq 0, \quad \langle x,\, \Phi(x)\rangle = 0, \tag{3.32}$$

where $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ is a smooth mapping. Once the complementarity function ψ is chosen, problem (3.32) can be written as the equivalent equation

$$\Psi(x) = 0, \tag{3.33}$$

where $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ is given by

$$\Psi(x) = \psi(x,\, \Phi(x)) \tag{3.34}$$

with the complementarity function ψ applied componentwise. In the case of the natural residual, the mapping Ψ given by (3.34) coincides with the natural residual mapping for the corresponding variational inequality, as defined in Sect. 1.3.1.

The natural residual function is nondifferentiable at any point $(a,\, b)$ satisfying $a = b$, while the Fischer–Burmeister function is nondifferentiable at $(0,\, 0)$. This implies that the mapping Ψ defined in (3.34) with any of these complementarity functions is usually nondifferentiable at a solution $\bar{x} \in \mathbf{R}^n$ violating the strict complementarity condition

$$\bar{x} + \Phi(\bar{x}) > 0, \tag{3.35}$$

no matter how smooth Φ is.

In what follows, we denote by Ψ_{NR} the mapping Ψ defined in (3.34) with ψ from (3.30), and we denote by Ψ_{FB} the mapping Ψ defined in (3.34) with ψ from (3.31).

It should be noted that there also exist smooth complementarity functions; one example is

$$\psi(a,\, b) = 2ab - (\min\{0,\, a + b\})^2, \tag{3.36}$$

which is everywhere differentiable, with the piecewise linear (and hence, Lipschitz-continuous on $\mathbf{R} \times \mathbf{R}$) gradient. However, all smooth complementarity functions have one principal deficiency, expressed by the equality

$$\psi'(0,\, 0) = 0.$$

This equality implies that for every index $i \in \{1,\, \ldots,\, n\}$ for which it holds that $\bar{x}_i = \Phi_i(\bar{x}) = 0$, for the mapping Ψ defined in (3.34) the i-th row of the Jacobian $\Psi'(\bar{x})$ consists of zero entries. Therefore, if the strict complementarity condition (3.35) is violated at the solution $\bar{x}$, then for any smooth complementarity function ψ used in (3.34), the Jacobian $\Psi'(\bar{x})$ is necessarily singular.

This is probably the main reason why nonsmooth complementarity functions are currently much more widely used than smooth ones. As will be demonstrated below, nonsmooth reformulations of complementarity problems are semismooth under natural assumptions, and their solutions can naturally satisfy BD-regularity or CD-regularity conditions even when the strict complementarity does not hold. Therefore, one can search for NCP solutions by applying the semismooth Newton methods discussed in Sect. 2.4 to an appropriate NCP reformulation. In fact, nonsmooth reformulations of complementarity problems provide one of the most important areas of application of semismooth Newton methods. As will be demonstrated below, the special structure of nonsmoothness present in these reformulations allows for easy computation of elements of the generalized Jacobians, which leads to readily implementable algorithms.

In what follows, we deal with nonsmooth complementarity functions only. For the development of Newton-type methods for NCP reformulation employing the smooth complementarity function (3.36) see [138, 209].

The basic *semismooth Newton method* for NCP is the following algorithm, which is a special case of Algorithm 2.41.

Algorithm 3.7 Select a complementarity function $\psi : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$, and define the mapping $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ according to (3.34). Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $\Psi(x^k) = 0$, stop.
2. Compute some $J_k \in \partial_B \Psi(x^k)$. Compute $x^{k+1} \in \mathbf{R}^n$ as a solution of the linear equation
$$\Psi(x^k) + J_k(x - x^k) = 0. \tag{3.37}$$
3. Increase k by 1 and go to step 1.

The first issue to settle is whether the mappings in question are appropriately semismooth.

Proposition 3.8. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be (strongly) semismooth at $x \in \mathbf{R}^n$. Then the mappings Ψ_{NR} and Ψ_{FB} are both (strongly) semismooth at x.*

For Ψ_{NR}, this fact readily follows from Proposition 1.75. For Ψ_{FB}, it follows from Example 1.68 and Propositions 1.73 and 1.74, taking into account the following view on the Fischer–Burmeister function:

$$\psi(a,\, b) = a + b - \|(a,\, b)\|.$$

From Theorem 2.42 and Proposition 3.8, we immediately obtain the following local superlinear convergence result for Algorithm 3.7.

Theorem 3.9. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be semismooth at $\bar{x} \in \mathbf{R}^n$. Let $\bar{x}$ be a solution of the NCP (3.32), and assume that Ψ_{NR} (Ψ_{FB}) is BD-regular at $\bar{x}$.*

Then any starting point $x^0 \in \mathbf{R}^n$ close enough to $\bar{x}$ defines a particular iterative sequence of Algorithm 3.7 with ψ defined according to (3.30) (according to (3.31), respectively), any such sequence converges to $\bar{x}$, and the rate of convergence is superlinear. If Φ is strongly semismooth at $\bar{x}$, then the rate of convergence is quadratic.

We proceed with the outer estimates of B-differentials of Ψ_{NR} and Ψ_{FB}, and with sufficient conditions for the BD-regularity of these mappings at a given solution of the NCP (3.32), assuming that Φ is smooth. The first result readily follows from Propositions 1.54 and 1.55.

Proposition 3.10. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable near $x \in \mathbf{R}^n$, with its derivative being continuous at x.*

Then the rows of any matrix $J \in \partial_B \Psi_{\mathrm{NR}}(x)$ satisfy the equalities

$$J_i = \begin{cases} \Phi_i'(x) \ \text{ or } e^i & \text{if } x_i = \Phi_i(x), \\ \Phi_i'(x) & \text{if } x_i > \Phi_i(x), \\ e^i & \text{if } x_i < \Phi_i(x), \end{cases} \quad i = 1, \ldots, n. \tag{3.38}$$

Observe that if Φ is continuously differentiable on an open set $O \subset \mathbf{R}^n$, then Ψ_{NR} is piecewise smooth on O, and for any $x \in O$ the set of matrices with rows satisfying (3.38) coincides with the set $\hat{\partial}\Psi_{\mathrm{NR}}(x)$ defined in Remark 2.55. Therefore, equation (3.33) with $\Psi = \Psi_{\mathrm{NR}}$ can also be tackled by the piecewise Newton method whose iteration system is (3.37) with J_k being an arbitrary matrix with rows satisfying (3.38) for $x = x^k$ (i.e., any matrix from $\hat{\partial}\Psi_{\mathrm{NR}}(x^k)$, assuming that $x^k \in O$). According to Remark 2.55, local superlinear convergence of this method to a solution $\bar{x}$ can be established assuming continuous differentiability of Φ near $\bar{x}$ and nonsingularity of all such matrices for $x = \bar{x}$ (that is, all matrices in $\hat{\partial}\Psi_{\mathrm{NR}}(\bar{x})$). The semismooth Newton method can be regarded as a special realization of the piecewise Newton method, with a more special choice of matrices from $\hat{\partial}\Psi_{\mathrm{NR}}(x^k)$.

Concerning the B-differential of Ψ_{FB}, probably the easiest way to obtain the next result is to combine Proposition 1.54 with direct considerations based on the definition of the B-differential.

Proposition 3.11. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable near $x \in \mathbf{R}^n$, with its derivative being continuous at x.*

Then the rows of any matrix $J \in \partial_B \Psi_{\mathrm{FB}}(x)$ satisfy the equalities

$$J_i = \begin{cases} \alpha_i \Phi_i'(x) + \beta_i e^i & \text{if } x_i = \Phi_i(x) = 0, \\ \Phi_i'(x) + e^i - \dfrac{\Phi_i(x)\Phi_i'(x) + x_i e^i}{\sqrt{x_i^2 + (\Phi_i(x))^2}} & \text{if } x_i \neq 0 \ \text{ or } \Phi_i(x) \neq 0, \end{cases} \quad i = 1, \ldots, n,$$

with some $(\alpha_i, \beta_i) \in S$ for $i = 1, \ldots, n$ such that $x_i = \Phi_i(x) = 0$, where

$$S = \{(a, b) \in \mathbf{R}^2 \mid (a-1)^2 + (b-1)^2 = 1\}. \tag{3.39}$$

Let now $\bar{x} \in \mathbf{R}^n$ be a solution of the NCP (3.32), and define the index sets

$$I_0(\bar{x}) = \{i = 1, \ldots, n \mid \bar{x}_i = \Phi_i(\bar{x}) = 0\},$$
$$I_1(\bar{x}) = \{i = 1, \ldots, n \mid \bar{x}_i > 0,\ \Phi_i(\bar{x}) = 0\},$$
$$I_2(\bar{x}) = \{i = 1, \ldots, n \mid \bar{x}_i = 0,\ \Phi_i(\bar{x}) > 0\}.$$

Evidently, $(I_0(\bar{x}),\, I_1(\bar{x}),\, I_2(\bar{x}))$ is a partition of the set $\{1, \ldots, n\}$, that is, $I_0(\bar{x}) \cup I_1(\bar{x}) \cup I_2(\bar{x}) = \{1, \ldots, n\}$ and the intersection of any two of the three index sets in question is empty.

Assuming continuity of the derivative of Φ at $\bar{x}$, from Proposition 3.10 it follows that $\partial_B \Psi_{\mathrm{NR}}(\bar{x}) \subset \Delta_{\mathrm{NR}}(\bar{x})$, where the set $\Delta_{\mathrm{NR}}(\bar{x})$ consists of all the matrices $J \in \mathbf{R}^{n \times n}$ with rows satisfying the equalities

$$J_i = \begin{cases} \Phi_i'(\bar{x}) \text{ or } e^i & \text{if } i \in I_0(\bar{x}), \\ \Phi_i'(\bar{x}) & \text{if } i \in I_1(\bar{x}), \\ e^i & \text{if } i \in I_2(\bar{x}). \end{cases}$$

Therefore, BD-regularity of Ψ_{NR} at $\bar{x}$ is implied by nonsingularity of all the matrices $J \in \Delta_{\mathrm{NR}}(\bar{x})$. It can be easily seen that this property is equivalent to saying that the matrix

$$\begin{pmatrix} (\Phi'(\bar{x}))_{I_1(\bar{x}) I_1(\bar{x})} & (\Phi'(\bar{x}))_{I_1(\bar{x}) K} \\ (\Phi'(\bar{x}))_{K I_1(\bar{x})} & (\Phi'(\bar{x}))_{KK} \end{pmatrix}$$

is nonsingular for any index set $K \subset I_0(\bar{x})$. The latter condition is known in the literature under the name of *b-regularity* of the solution $\bar{x}$; see [212]. Observe also that if Φ is continuously differentiable near $\bar{x}$, then $\Delta_{\mathrm{NR}}(\bar{x})$ coincides with $\hat{\partial} \Psi_{\mathrm{NR}}(\bar{x})$, and therefore, b-regularity of $\bar{x}$ is precisely the condition needed for local superlinear convergence of the piecewise Newton method.

Similarly, from Proposition 3.11 it follows that $\partial_B \Psi_{\mathrm{FB}}(\bar{x}) \subset \Delta_{\mathrm{FB}}(\bar{x})$, where the set $\Delta_{\mathrm{FB}}(\bar{x})$ consists of all the matrices $J \in \mathbf{R}^{n \times n}$ with rows satisfying the equalities

$$J_i = \begin{cases} \alpha_i \Phi_i'(\bar{x}) + \beta_i e^i & \text{if } i \in I_0(\bar{x}), \\ \Phi_i'(\bar{x}) & \text{if } i \in I_1(\bar{x}), \\ e^i & \text{if } i \in I_2(\bar{x}), \end{cases} \tag{3.40}$$

with some $(\alpha_i,\, \beta_i) \in S$, $i \in I_0(\bar{x})$. Therefore, BD-regularity of Ψ_{FB} at $\bar{x}$ is implied by nonsingularity of all the matrices $J \in \Delta_{\mathrm{FB}}(\bar{x})$. Moreover, since J_i in (3.40) depends linearly on $(\alpha_i,\, \beta_i)$, by the standard tools of convex analysis (see Sect. A.3) it immediately follows that $\operatorname{conv} \Delta_{\mathrm{FB}}(\bar{x})$ consists of all the matrices $J \in \mathbf{R}^{n \times n}$ with rows satisfying (3.40) with some $(\alpha_i,\, \beta_i) \in \operatorname{conv} S$, $i \in I_0(\bar{x})$, where

$$\operatorname{conv} S = \{(a,\, b) \in \mathbf{R}^2 \mid (a-1)^2 + (b-1)^2 \leq 1\}.$$

Therefore, CD-regularity of Ψ_{FB} at $\bar{x}$ is implied by nonsingularity of all the matrices of this form.

It is evident that $\Delta_{\mathrm{NR}}(\bar{x}) \subset \Delta_{\mathrm{FB}}(\bar{x})$, implying also the inclusion for the convex hulls $\operatorname{conv} \Delta_{\mathrm{NR}}(\bar{x}) \subset \operatorname{conv} \Delta_{\mathrm{FB}}(\bar{x})$. In particular, nonsingularity of all matrices in $\Delta_{\mathrm{FB}}(\bar{x})$ also implies b-regularity of $\bar{x}$. The converse implication is not true; moreover, b-regularity of $\bar{x}$ does not necessarily imply neither CD-regularity of Ψ_{NR} at $\bar{x}$, nor BD-regularity of Ψ_{FB} at $\bar{x}$, as demonstrated by the following example taken from [53, Example 2.1] (see also Example 3.16 below).

Example 3.12. Let $n = 2$, $\Phi(x) = (-x_1 + x_2,\, -x_2)$. Then $\bar{x} = 0$ is the unique solution of the NCP (3.32).

It can be directly checked that

$$\partial_B \Psi_{\mathrm{NR}}(\bar{x}) = \Delta_{\mathrm{NR}}(\bar{x}) = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\, \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix} \right\},$$

and therefore, $\bar{x}$ is b-regular. On the other hand,

$$\frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1/2 \\ 0 & 1 \end{pmatrix}$$

is a singular matrix, and hence, Ψ_{NR} is not CD-regular at $\bar{x}$.

For each $k = 1, 2, \ldots$, set $x^k = (1/k,\, 2/k)$. Then Ψ_{FB} is differentiable at x^k with

$$\Psi'_{\mathrm{FB}}(x^k) = \begin{pmatrix} 0 & 1 - 1/\sqrt{2} \\ 0 & -\sqrt{2} \end{pmatrix},$$

and the sequence $\{x^k\}$ converges to $\bar{x}$. Therefore, the singular matrix in the right-hand side of the last relation belongs to $\partial_B \Psi_{\mathrm{FB}}(\bar{x})$, and hence, Ψ_{FB} is not BD-regular at $\bar{x}$.

One might conjecture that BD-regularity (or at least CD-regularity) of Ψ_{FB} at $\bar{x}$ implies CD-regularity (or at least BD-regularity) of Ψ_{NR} at $\bar{x}$, but this is also not the case, as demonstrated by the following example taken from [50].

Example 3.13. Let $n = 2$, $\Phi(x) = (x_2,\, -x_1 + x_2)$. Then $\bar{x} = 0$ is the unique solution of the NCP (3.32), and it can be seen that $\partial_B \Psi_{\mathrm{NR}}(\bar{x})$ contains the singular matrix

$$\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix},$$

while Ψ_{FB} is CD-regular at $\bar{x}$ (the latter can be shown similarly to Example 3.15 below).

We proceed with the example from [134] showing that even a combination of CD-regularity of Ψ_{NR} at $\bar{x}$ and b-regularity of $\bar{x}$ does not imply BD-regularity (and even less so CD-regularity) of Ψ_{FB} at $\bar{x}$.

Example 3.14. Let $n = 2$, $\Phi(x) = (-x_1 + 3x_2/(2\sqrt{2}),\ 2x_1 + (1 - 3/(2\sqrt{2}))x_2)$. Then $\bar{x} = 0$ is the unique solution of the NCP (3.32).

Consider any sequence $\{x^k\} \subset \mathbf{R}^2$ such that $x_1^k < 0$, $x_2^k = 0$ for all k, and x_1^k tends to 0 as $k \to \infty$. Then for all k it holds that Φ is differentiable at x^k, and

$$\begin{aligned}\Phi'(x^k) &= \begin{pmatrix} -\dfrac{2x_1^k}{\sqrt{2(x_1^k)^2}} & \dfrac{3}{2\sqrt{2}}\left(1 + \dfrac{x_1^k}{\sqrt{2(x_1^k)^2}}\right) \\ 2\left(1 - \dfrac{2x_1^k}{\sqrt{(2x_1^k)^2}}\right) & 1 + \left(1 - \dfrac{3}{2\sqrt{2}}\right)\left(1 - \dfrac{2x_1^k}{\sqrt{(2x_1^k)^2}}\right) \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{2} & \dfrac{3}{2\sqrt{2}}\left(1 - \dfrac{1}{\sqrt{2}}\right) \\ 4 & 1 + 2\left(1 - \dfrac{3}{2\sqrt{2}}\right) \end{pmatrix}.\end{aligned}$$

Therefore, the singular matrix in the right-hand side belongs to $\partial_B \Psi_{\mathrm{FB}}(\bar{x})$.

At the same time, it can be directly checked that

$$\partial_B \Psi_{\mathrm{NR}}(\bar{x}) = \left\{ \begin{pmatrix} 1 & 0 \\ 2 & 1 - \dfrac{3}{2\sqrt{2}} \end{pmatrix},\ \begin{pmatrix} -1 & \dfrac{3}{2\sqrt{2}} \\ 0 & 1 \end{pmatrix} \right\},$$

and hence,

$$\partial \Psi_{\mathrm{NR}}(\bar{x}) = \left\{ \begin{pmatrix} t - (1-t) & (1-t)\dfrac{3}{2\sqrt{2}} \\ 2t & t\left(1 - \dfrac{3}{2\sqrt{2}}\right) + (1-t) \end{pmatrix} \,\middle|\, t \in [0,\, 1] \right\}.$$

By direct computation, for any matrix $J(t)$ in the right-hand side of the last relation we have

$$\det J(t) = \left(2 - \frac{3}{2\sqrt{2}}\right)t - 1 \leq 1 - \frac{3}{2\sqrt{2}} < 0 \quad \forall\, t \in [0,\, 1],$$

implying *CD*-regularity of Ψ_{NR} at $\bar{x}$.

Observe also that

$$\begin{aligned}&\Delta_{\mathrm{NR}}(\bar{x}) \\ &= \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ \begin{pmatrix} 1 & 0 \\ 2 & 1 - \dfrac{3}{2\sqrt{2}} \end{pmatrix},\ \begin{pmatrix} -1 & \dfrac{3}{2\sqrt{2}} \\ 0 & 1 \end{pmatrix},\ \begin{pmatrix} -1 & \dfrac{3}{2\sqrt{2}} \\ 2 & 1 - \dfrac{3}{2\sqrt{2}} \end{pmatrix} \right\},\end{aligned}$$

and it is evident that $\bar{x}$ is a b-regular solution as well.

We have already seen from the previous example that the set $\partial_B \Psi_{\mathrm{NR}}(\bar{x})$ can be smaller than $\Delta_{\mathrm{NR}}(\bar{x})$. The next example, taken from [50], demonstrates that $\partial_B \Psi_{\mathrm{FB}}(\bar{x})$ can be smaller than $\Delta_{\mathrm{FB}}(\bar{x})$. Moreover, Ψ_{NR} and Ψ_{FB} can be both BD-regular, and even CD-regular at $\bar{x}$ when both $\Delta_{\mathrm{NR}}(\bar{x})$ and $\Delta_{\mathrm{FB}}(\bar{x})$ contain singular matrices.

Example 3.15. Let $n = 2$, $\Phi(x) = ((x_1 + x_2)/2,\, (x_1 + x_2)/2)$. Then $\bar{x} = 0$ is the unique solution of the NCP (3.32).

It can be directly checked that

$$\partial_B \Psi_{\mathrm{NR}}(\bar{x}) = \left\{ \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix},\, \begin{pmatrix} 1/2 & 1/2 \\ 0 & 1 \end{pmatrix} \right\},$$

and hence,

$$\partial \Psi_{\mathrm{NR}}(\bar{x}) = \left\{ \begin{pmatrix} t + (1-t)/2 & (1-t)/2 \\ t/2 & (1-t) + t/2 \end{pmatrix} \,\middle|\, t \in [0,\, 1] \right\}.$$

By direct computation, $\det J = 1/2$ for all $J \in \partial \Psi_{\mathrm{NR}}(\bar{x})$, implying CD-regularity (and hence BD-regularity) of Ψ_{NR} at $\bar{x}$. At the same time,

$$\Delta_{\mathrm{NR}}(\bar{x}) = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\, \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix},\, \begin{pmatrix} 1/2 & 1/2 \\ 0 & 1 \end{pmatrix},\, \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \right\},$$

where the matrix

$$J_0 = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \tag{3.41}$$

is singular.

Furthermore, $\Delta_{\mathrm{FB}}(\bar{x})$ consists of matrices of the form

$$J(\alpha,\, \beta) = \begin{pmatrix} \alpha_1/2 + \beta_1 & \alpha_1/2 \\ \alpha_2/2 & \alpha_2/2 + \beta_2 \end{pmatrix} \tag{3.42}$$

for all $\alpha,\, \beta \in \mathbf{R}^2$ such that $(\alpha_i,\, \beta_i) \in S$, $i = 1,\, 2$. By direct computation,

$$\det J(\alpha,\, \beta) = \frac{1}{2}(\alpha_1\beta_2 + \alpha_2\beta_1 + 2\beta_1\beta_2).$$

Since the inclusion $(\alpha_i,\, \beta_i) \in S$ implies that $\alpha_i \geq 0$, $\beta_i \geq 0$ for $i = 1,\, 2$, this determinant equals zero only if

$$\alpha_1\beta_2 = 0, \quad \alpha_2\beta_1 = 0, \quad \beta_1\beta_2 = 0,$$

and hence, $\alpha_1 = \alpha_2 = 1$, $\beta_1 = \beta_2 = 0$, implying that J_0 defined in (3.41) is the only singular matrix in $\Delta_{\mathrm{FB}}(\bar{x})$. However, it can be directly checked that this matrix does not belong to $\partial_B \Psi_{\mathrm{FB}}(\bar{x})$, and therefore, Ψ_{FB} is BD-regular at $\bar{x}$. Moreover, employing the above characterization of the structure of $\operatorname{conv} \Delta_{\mathrm{FB}}(\bar{x})$, it is evident that J_0 is the only singular matrix in this set as well, implying CD-regularity of Ψ_{FB} at $\bar{x}$.

We complete the picture of relations between the properties of BD-regularity and CD-regularity by the following simple example showing that BD-regularity of Ψ_{FB} at $\bar{x}$ does not imply CD-regularity of Ψ_{FB} at $\bar{x}$.

Example 3.16. Let $n = 1$, $\Phi(x) = -x$. Then $\bar{x} = 0$ is the unique solution of the NCP (3.32).

Obviously, $\Psi_{\mathrm{FB}}(x) = -\sqrt{2}|x|$, $\partial_B \Psi_{\mathrm{FB}}(\bar{x}) = \{-\sqrt{2}, \sqrt{2}\}$, implying BD-regularity of Ψ_{FB} at $\bar{x}$. At the same time, $\partial \Psi_{\mathrm{FB}}(\bar{x}) = \Delta_{\mathrm{FB}}(\bar{x}) = [-\sqrt{2}, \sqrt{2}]$ contains 0.

We proceed with the relations of the properties in question with strong regularity. The next proposition combines the results from [53, Theorem 2.5] and [80, Theorem 1].

Proposition 3.17. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable at $\bar{x} \in \mathbf{R}^n$, and let $\bar{x}$ be a strongly regular solution of the NCP* (3.32) *(see Definition 1.23).*

Then all the matrices in $\operatorname{conv} \Delta_{\mathrm{FB}}(\bar{x})$ *(and hence, also in* $\operatorname{conv} \Delta_{\mathrm{NR}}(\bar{x})$*) are nonsingular.*

Proof. We denote

$$I_0 = I_0(\bar{x}), \quad I_1 = I_1(\bar{x}), \quad I_2 = I_2(\bar{x}) \tag{3.43}$$

for brevity. Suppose that $\operatorname{conv} \Delta_{\mathrm{FB}}(\bar{x})$ contains a singular matrix J, i.e., $J\xi = 0$ holds for some $\xi \in \mathbf{R}^n \setminus \{0\}$. Taking into account the structure of the set $\operatorname{conv} \Delta_{\mathrm{FB}}(\bar{x})$, we then obtain the existence of $\alpha = (\alpha_i,\ i \in I_0)$ and $\beta = (\beta_i,\ i \in I_0)$, satisfying

$$\alpha \geq 0, \quad \beta \geq 0, \quad \alpha + \beta > 0 \tag{3.44}$$

and such that (3.40) holds. Hence,

$$\operatorname{diag}(\alpha)\Phi'_{I_0}(\bar{x})\xi + \operatorname{diag}(\beta)\xi_{I_0} = 0, \quad \Phi'_{I_1}(\bar{x})\xi = 0, \quad \xi_{I_2} = 0.$$

Therefore, $\xi_{I_0 \cup I_1} \neq 0$, and

$$\operatorname{diag}(\alpha)(\Phi'(\bar{x}))_{I_0 I_0}\xi_{I_0} + \operatorname{diag}(\alpha)(\Phi'(\bar{x}))_{I_0 I_1}\xi_{I_1} + \operatorname{diag}(\beta)\xi_{I_0} = 0 \tag{3.45}$$

and

$$(\Phi'(\bar{x}))_{I_1 I_0}\xi_{I_0} + (\Phi'(\bar{x}))_{I_1 I_1}\xi_{I_1} = 0. \tag{3.46}$$

Employing Proposition 1.26, we obtain that $\Phi'_{I_1 I_1}(\bar{x})$ is a nonsingular matrix, and from (3.46) it follows that

$$\xi_{I_1} = -((\Phi'(\bar{x}))_{I_1 I_1})^{-1}(\Phi'(\bar{x}))_{I_1 I_0}\xi_{I_0}. \tag{3.47}$$

Substituting this expression into (3.45), we obtain the relation

$$(\mathrm{diag}(\alpha)((\Phi'(\bar{x}))_{I_0I_0} - (\Phi'(\bar{x}))_{I_0I_1}((\Phi'(\bar{x}))_{I_1I_1})^{-1}(\Phi'(\bar{x}))_{I_1I_0})$$
$$+\,\mathrm{diag}(\beta))\xi_{I_0} = 0. \tag{3.48}$$

According to Proposition 1.26, if $\xi_{I_0} \neq 0$, then there exists $i \in I_0$ such that

$$\xi_i(((\Phi'(\bar{x}))_{I_0I_0} - (\Phi'(\bar{x}))_{I_0I_1}((\Phi'(\bar{x}))_{I_1I_1})^{-1}(\Phi'(\bar{x}))_{I_1I_0})\xi_{I_0})_i > 0 \tag{3.49}$$

(and in particular $\xi_i \neq 0$). Multiplying both sides of the i-th equality in (3.48) by ξ_i we derive

$$\alpha_i\xi_i(((\Phi'(\bar{x}))_{I_0I_0} - (\Phi'(\bar{x}))_{I_0I_1}((\Phi'(\bar{x}))_{I_1I_1})^{-1}(\Phi'(\bar{x}))_{I_1I_0})\xi_{I_0})_i + \beta_i\xi_i^2 = 0.$$

However, taking into account (3.44), this contradicts (3.49). Therefore, it holds that $\xi_{I_0} = 0$, and by (3.47) we then further obtain that $\xi_{I_1} = 0$ as well, yielding a contradiction.

The assertion regarding $\Delta_{\mathrm{NR}}(\bar{x})$ follows now immediately from the inclusion $\Delta_{\mathrm{NR}}(\bar{x}) \subset \Delta_{\mathrm{FB}}(\bar{x})$. □

From the inclusions $\partial_B\Psi_{\mathrm{NR}}(\bar{x}) \subset \Delta_{\mathrm{NR}}(\bar{x})$ and $\partial_B\Psi_{\mathrm{FB}}(\bar{x}) \subset \Delta_{\mathrm{FB}}(\bar{x})$, we now conclude that strong regularity of $\bar{x}$ implies CD-regularity (and hence, BD-regularity) of both Ψ_{NR} and Ψ_{FB} at $\bar{x}$.

On the other hand, b-regularity of $\bar{x}$ (and hence, the weaker BD-regularity of Ψ_{NR} at $\bar{x}$) does not imply strong regularity of $\bar{x}$; this is demonstrated by Example 3.12 above. Indeed, by Proposition 3.17, strong regularity would have been implying BD-regularity of Ψ_{FB} which does not hold in that example.

Moreover, in Example 3.16 both b-regularity and BD-regularity of Ψ_{FB} at $\bar{x}$ do hold. In particular, the former holds because

$$\Psi_{\mathrm{NR}}(x) = -|x|, \quad \partial_B\Psi_{\mathrm{NR}}(\bar{x}) = \Delta_{\mathrm{NR}}(\bar{x}) = \{-1,\, 1\}.$$

On the other hand, strong regularity does not hold. The latter follows either from Proposition 1.26 or by directly observing that the perturbed linearized NCP

$$x \geq 0, \quad -x - r \geq 0, \quad x(-x - r) = 0$$

has no solutions for $r > 0$.

Note that both Examples 3.12 and 3.16 violate CD-regularity of both Ψ_{NR} and Ψ_{FB} at $\bar{x}$. However, CD-regularity of Ψ_{NR} and Ψ_{FB} does not imply strong regularity as well. This can be seen from Example 3.15. The absence of strong regularity in that example can be verified either directly or by application of any of Propositions 1.26 or 3.17.

At the same time, nonsingularity of all the matrices in $\Delta_{\mathrm{FB}}(\bar{x})$ implies strong regularity of $\bar{x}$, as we show next.

Lemma 3.18. *If $A \in \mathbf{R}^{m\times m}$ is not a P-matrix, then the following assertions are valid:*

(a) there exist $\alpha \in \mathbf{R}^m$ and $\beta \in \mathbf{R}^m$ such that $(\alpha_i,\, \beta_i) \in S$ for all indices $i = 1,\, \ldots,\, m$, where S is defined in (3.39), and the matrix

$$A(\alpha,\, \beta) = \mathrm{diag}(\alpha)A + \mathrm{diag}(\beta) \tag{3.50}$$

is singular;

(b) there exist $\alpha \in \mathbf{R}^m$ and $\beta \in \mathbf{R}^m$ such that $\alpha_i \geq 0$, $\beta_i \geq 0$, $\alpha_i + \beta_i = 1$ for all $i = 1,\, \ldots,\, m$, and the matrix $A(\alpha,\, \beta)$ defined by (3.50) is singular.

Proof. To prove assertion (a) we argue by induction. If $m = 1$, then the assumption that A is not a P-matrix means that A is a nonpositive scalar. It then follows that the circle S in $(\alpha,\, \beta)$-plane always has a nonempty intersection with the straight line given by the equation $\mathrm{diag}(\alpha)A + \mathrm{diag}(\beta) = 0$. The points in this intersection are the needed pairs $(\alpha,\, \beta)$.

Suppose now that the assertion is valid for any matrix in $\mathbf{R}^{(m-1)\times(m-1)}$, and suppose that the matrix $A \in \mathbf{R}^{m\times m}$ has a nonpositive principal minor. If the only such minor is $\det A$, then set $\alpha_i = 1$, $\beta_i = 0$, $i = 2,\, \ldots,\, m$, and compute

$$\begin{aligned}\det A(\alpha,\, \beta) &= \det \begin{pmatrix} \alpha_1 a_{11} + \beta_1 & \alpha_1 a_{12} & \ldots & \alpha_1 a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mm} \end{pmatrix} \\ &= \alpha_1 \det A + \beta_1 \det A_{\{2,\, \ldots,\, m\}\{2,\, \ldots,\, m\}},\end{aligned}$$

where the last equality can be obtained by expanding the determinant by the first row. Since $\det A \leq 0$ and $\det A_{\{2,\, \ldots,\, m\}\{2,\, \ldots,\, m\}} > 0$, we again obtain that the circle S in $(\alpha_1,\, \beta_1)$-plane always has a nonempty intersection with the straight line given by the equation $\det A(\alpha,\, \beta) = 0$ with respect to $(\alpha_1,\, \beta_1)$, and we again get the needed α and β.

It remains to consider the case of existence of an index set $K \subset \{1,\, \ldots,\, m\}$ such that $\det A_{KK} \leq 0$, and there exists $k \in \{1,\, \ldots,\, m\} \setminus K$. Removing the k-th row and column from A, we then get the matrix $\tilde{A} = \mathbf{R}^{(m-1)\times(m-1)}$ with a nonpositive principal minor. By the hypothesis of the induction, there exists a vector $\tilde{\alpha} = (\alpha_1,\, \ldots,\, \alpha_{k-1},\, \alpha_{k+1},\, \ldots,\, \alpha_m) \in \mathbf{R}^{m-1}$ and there exists a vector $\tilde{\beta} = (\beta_1,\, \ldots,\, \beta_{k-1},\, \beta_{k+1},\, \ldots,\, \beta_m) \in \mathbf{R}^{m-1}$ such that $(\alpha_i,\, \beta_i) \in S$ for all $i = 1,\, \ldots,\, m$, $i \neq k$, and the matrix $\mathrm{diag}(\tilde{\alpha})\tilde{A} + \mathrm{diag}(\tilde{\beta})$ is singular. Setting $\alpha_k = 0$, $\beta_k = 1$, we again obtain the needed α and β.

Assertion (b) follows by dividing $(\alpha_i,\, \beta_i)$ from assertion (a) by $\alpha_i + \beta_i$, for each $i = 1,\, \ldots,\, m$. $\square$

Proposition 3.19. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable at a solution $\bar{x} \in \mathbf{R}^n$ of the NCP (3.32), and assume that at least one of the following assumptions is valid:*

(i) *All the matrices in* $\Delta_{\mathrm{FB}}(\bar{x})$ *are nonsingular.*
(ii) *All the matrices in* $\operatorname{conv}\Delta_{\mathrm{NR}}(\bar{x})$ *are nonsingular.*

Then $\bar{x}$ *is strongly regular.*

Proof. We again use the notation in (3.43). Nonsingularity of all the matrices in $\Delta_{\mathrm{FB}}(\bar{x})$ is equivalent to saying that the matrix

$$D(\alpha,\,\beta)=\begin{pmatrix}(\Phi'(\bar{x}))_{I_1I_1} & (\Phi'(\bar{x}))_{I_1I_0}\\ \operatorname{diag}(\alpha)(\Phi'(\bar{x}))_{I_0I_1} & \operatorname{diag}(\alpha)(\Phi'(\bar{x}))_{I_0I_0}+\operatorname{diag}(\beta)\end{pmatrix}\qquad(3.51)$$

is nonsingular for all $\alpha=(\alpha_i,\ i\in I_0)$ and $\beta=(\beta_i,\ i\in I_0)$ satisfying $(\alpha_i,\,\beta_i)\in S$, $i\in I_0$. Taking $\alpha_i=0$, $\beta_i=1$, $i\in I_0$, we immediately obtain that $(\Phi'(\bar{x}))_{I_1I_1}$ is nonsingular.

Similarly, from the definition of Δ_{NR} and by the standard tools of convex analysis (see Sect. A.3) it can be easily deduced that nonsingularity of all the matrices in $\operatorname{conv}\Delta_{\mathrm{NR}}(\bar{x})$ is equivalent to saying that the matrix defined in (3.51) is nonsingular for all $\alpha=(\alpha_i,\ i\in I_0)$ and $\beta=(\beta_i,\ i\in I_0)$ $\alpha_i\geq 0$, $\beta_i\geq 0$, $\alpha_i+\beta_i=1$, $i\in I_0$. Therefore, by taking $\alpha_i=0$, $\beta_i=1$, $i\in I_0$, we again conclude that $(\Phi'(\bar{x}))_{I_1I_1}$ is nonsingular.

Suppose that $\bar{x}$ is not strongly regular. Then Proposition 1.26 and nonsingularity of $(\Phi'(\bar{x}))_{I_1I_1}$ imply that

$$(\Phi'(\bar{x}))_{I_0I_0}-(\Phi'(\bar{x}))_{I_0I_1}((\Phi'(\bar{x}))_{I_1I_1})^{-1}(\Phi'(\bar{x}))_{I_1I_0}$$

is not a P-matrix. Therefore, by Lemma 3.18 we obtain the existence of $\alpha=(\alpha_i,\ i\in I_0)$ and $\beta=(\beta_i,\ i\in I_0)$ satisfying $(\alpha_i,\,\beta_i)\in S$ for all $i\in I_0$, or $\alpha_i\geq 0$, $\beta_i\geq 0$, $\alpha_i+\beta_i=1$ for all $i\in I_0$, and such that the matrix

$$\operatorname{diag}(\alpha)((\Phi'(\bar{x}))_{I_0I_0}-(\Phi'(\bar{x}))_{I_0I_1}((\Phi'(\bar{x}))_{I_1I_1})^{-1}(\Phi'(\bar{x}))_{I_1I_0})+\operatorname{diag}(\beta)$$

is singular. But this matrix is the Schur complement of the nonsingular matrix $(\Phi'(\bar{x}))_{I_1I_1}$ in $D(\alpha,\,\beta)$, and therefore, by Lemma A.9 we conclude that $D(\alpha,\,\beta)$ is also singular. □

Combining Propositions 3.17 and 3.19, we obtain the following.

Corollary 3.20. *Let* $\Phi:\mathbf{R}^n\to\mathbf{R}^n$ *be differentiable at a solution* $\bar{x}\in\mathbf{R}^n$ *of the NCP* (3.32).

Then the following properties are equivalent:

(a) *All the matrices in* $\Delta_{\mathrm{FB}}(\bar{x})$ *are nonsingular.*
(b) *All the matrices in* $\operatorname{conv}\Delta_{\mathrm{NR}}(\bar{x})$ *are nonsingular.*
(c) *All the matrices in* $\operatorname{conv}\Delta_{\mathrm{FB}}(\bar{x})$ *are nonsingular.*
(d) *Solution* $\bar{x}$ *is strongly regular.*

In particular, any of the equivalent conditions (a)–(d) in Corollary 3.20 implies CD-regularity (and hence, BD-regularity) of both Ψ_{NR} and Ψ_{FB} at a solution $\bar{x}$.

We note that the equivalence of (a), (b), and (c) in Corollary 3.20 can actually be proved directly, regardless of their relation to strong regularity, by the same reasoning as will be used in Proposition 3.31 below in the context of Karush–Kuhn–Tucker systems.

Finally, recall that according to Example 3.14, replacing $\partial_B \Psi_{\mathrm{NR}}(\bar{x})$ by its convexification $\partial \Psi_{\mathrm{NR}}(\bar{x})$, or by its other enlargement $\Delta_{\mathrm{NR}}(\bar{x})$, and assuming nonsingularity of all the matrices in the resulting set does not imply even BD-regularity of Ψ_{FB} at $\bar{x}$. At the same time, applying *both* these enlargements together, that is, assuming nonsingularity of all the matrices in conv $\Delta_{\mathrm{NR}}(\bar{x})$, implies strong regularity of $\bar{x}$, as demonstrated by Proposition 3.19.

In order to implement Algorithm 3.7 with the natural residual and the Fischer–Burmeister complementarity functions, one needs some explicit procedures for computing actual elements of $\partial_B \Psi_{\mathrm{NR}}(x)$ and $\partial_B \Psi_{\mathrm{FB}}(x)$ for a given $x \in \mathbf{R}^n$. (Recall that $\Delta_{\mathrm{NR}}(x)$ and $\Delta_{\mathrm{FB}}(\bar{x})$ give only the outer estimates of the corresponding generalized Jacobians.)

For Ψ_{NR}, the needed procedure was designed in [222]. Differentiability of Φ at x implies that for any index $i \in \{1, \ldots, n\}$, the component $(\Psi_{\mathrm{NR}})_i$ is differentiable at x if and only if either $x_i > \Phi_i(x)$ (in which case we have $(\Psi_{\mathrm{NR}})'_i(x) = \Phi'_i(x)$), or $x_i < \Phi_i(x)$ (in which case we have $(\Psi_{\mathrm{NR}})'_i(x) = e^i$), or $x_i = \Phi_i(x)$ and $\Phi'_i(x) = e^i$ (in which case $(\Psi_{\mathrm{NR}})'_i(x) = \Phi'_i(x) = e^i$).

Assuming continuity of the derivative of Φ at x, set

$$I^0 = \{i = 1, \ldots, n \mid x_i = \Phi_i(x),\ \Phi'_i(x) \neq e^i\},$$

$$J_i = \begin{cases} \Phi'_i(x) & \text{if } x_i > \Phi_i(x), \\ e^i & \text{if } x_i < \Phi_i(x), \\ \Phi'_i(x) = e^i & \text{if } x_i = \Phi_i(x) \text{ and } \Phi'_i(x) = e^i, \end{cases} \qquad i \in \{1, \ldots, n\} \setminus I^0.$$

If the set $I^j \neq \emptyset$ is already generated for some $j \in \{0, 1, \ldots\}$, pick any $\xi^j \in \mathbf{R}^n$ such that at least for some $i \in I^j$ it holds that $\langle \Phi'_i(x), \xi^j \rangle \neq \xi^j_i$. For example, one can take $\xi^j = \Phi'_i(x) - e^i$ for some $i \in I^j$. Set

$$I^{j+1} = \{i \in I^j \mid \langle \Phi'_i(x), \xi^j \rangle = \xi^j_i\},$$

$$J_i = \begin{cases} \Phi'_i(x) & \text{if } \langle \Phi'_i(x), \xi^j \rangle < \xi^j_i \\ e^i & \text{if } \langle \Phi'_i(x), \xi^j \rangle > \xi^j_i, \end{cases} \qquad i \in I^j \setminus I^{j+1}.$$

If $I^{j+1} = \emptyset$, then stop; otherwise, increase j by 1 and loop.

Evidently, this procedure will terminate after a finite number of steps, producing $I^j \neq \emptyset$, $\xi^j \in \mathbf{R}^n$, $j = 0, 1, \ldots, q$, and $I^{q+1} = \emptyset$, with some $q \in \{0, 1, \ldots\}$. It can be easily seen that for any $\theta > 0$ small enough, for $\bar{\xi} = \sum_{j=0}^{q} \theta^j \xi^j$ it holds that $\langle \Phi'_i(x), \bar{\xi} \rangle \neq \xi_i$ for all $i \in I^0$. This implies that for any function $r(\cdot) : \mathbf{R}_+ \to \mathbf{R}^n$ such that $r(t) = o(t)$ as $t \to 0$, for any $i \in \{1, \ldots, n\}$ such that $x_i \neq \Phi_i(x)$, as well as for any $i \in I^0$, for all $t > 0$ small enough it holds that $x_i + t\xi_i + r_i(t) \neq \Phi_i(x + t\xi + r(t))$, the component $(\Psi_{\mathrm{NR}})_i$ is differentiable at $x + t\xi + r(t)$, and $(\Psi_{\mathrm{NR}})'_i(x + t\xi + r(t))$ tends to J_i as $t \to 0+$.

According to the Rademacher Theorem (Theorem 1.48), for any $t > 0$ small enough, in any neighborhood of $x + t\xi$ there exist points of differentiability of Ψ_{NR}, and therefore, we can select $r(t) \in \mathbf{R}^n$ in such a way that $r(t) = o(t)$ and Ψ_{NR} is differentiable at $x + t\xi + r(t)$. As discussed above, for any $i = 1, \ldots, n$, the gradient $(\Psi_{\mathrm{NR}})'_i(x + t\xi + r(t))$ equals either $\Phi'_i(x + t\xi + r(t))$ or e^i, and therefore, for any $i \in \{1, \ldots, n\}$ such that $x_i = \Phi_i(x)$ and $\Phi'_i(x) = e^i$, the gradient $(\Psi_{\mathrm{NR}})'_i(x + t\xi + r(t))$ tends to $J_i = \Phi'_i(x) = e^i$ as $t \to 0+$. Therefore, $\Psi'_{\mathrm{NR}}(x + t\xi + r(t))$ tends to J, implying the needed inclusion $J \in \partial_B \Psi_{\mathrm{NR}}(x)$.

For Ψ_{FB}, the procedure for computing an element of B-differential was developed in [52]. Set

$$K = \{i = 1, \ldots, n \mid x_i = \Phi_i(x) = 0\},$$

and pick any $\xi \in \mathbf{R}^n$ such that $\xi_i \neq 0$ for all $i \in K$. Set

$$J_i = \begin{cases} \Phi'_i(x) + e^i - \dfrac{\Phi_i(x)\Phi'_i(x) + x_i e^i}{\sqrt{x_i^2 + (\Phi_i(x))^2}} & \text{if } i \in \{1, \ldots, n\} \setminus K, \\[2ex] \Phi'_i(x) + e^i - \dfrac{\langle \Phi'_i(x), \xi\rangle \Phi'_i(x) + \xi_i e^i}{\sqrt{x_i^2 + (\Phi_i(x))^2}} & \text{if } i \in K. \end{cases} \tag{3.52}$$

Consider the points of the form $x(t) = x + t\xi$, $t \in \mathbf{R}$. Evidently, for any t close enough to zero and any $i = 1, \ldots, n$, it holds that either $x_i(t) \neq 0$ or $\Phi_i(x_i(t)) \neq 0$, and hence, Ψ_{FB} is differentiable at $x(t)$.

By the continuity of Φ and its derivative at x, for any $i \in \{1, \ldots, n\} \setminus K$ it follows that $(\Psi_{\mathrm{FB}})'_i(x(t))$ tends to J_i defined in (3.52) as $t \to 0$. Furthermore, for $i \in K$ it holds that

$$\Phi_i(x(t)) = t\langle \Phi'_i(x), \xi\rangle + o(t)$$

as $t \to 0$, and hence,

$$\begin{aligned} (\Psi_{\mathrm{FB}})'_i(x(t)) &= \Phi'_i(x(t)) + e^i - \frac{\Phi_i(x(t))\Phi'_i(x(t)) + t\xi_i e^i}{\sqrt{t^2\xi_i^2 + (\Phi_i(x(t)))^2}} \\ &= \Phi'_i(x(t)) + e^i - \frac{t\langle \Phi'_i(x), \xi\rangle \Phi'_i(x(t)) + t\xi_i e^i + o(t)}{\sqrt{t^2\xi_i^2 + t^2\langle \Phi'_i(x), \xi\rangle^2 + o(t^2)}} \\ &= \Phi'_i(x(t)) + e^i - \frac{\langle \Phi'_i(x), \xi\rangle \Phi'_i(x(t)) + \xi_i e^i + o(t)/t}{\sqrt{\xi_i^2 + \langle \Phi'_i(x), \xi\rangle^2 + o(t^2)/t^2}}, \end{aligned}$$

which also tends to the corresponding J_i defined in (3.52) as $t \to 0$. Therefore, $J \in \partial_B \Psi_{\mathrm{FB}}(x)$.

Recall that in the analysis of local convergence of the Josephy–Newton method, the key role is played by the semistability assumption. From Proposition 1.34 we have that $\bar{x}$ is a semistable solution of the NCP (3.32) if and only if the system

$$\begin{gathered} \xi_i \geq 0, \quad \langle \Phi'_i(\bar{x}), \xi\rangle \geq 0, \quad \xi_i\langle \Phi'_i(\bar{x}), \xi\rangle = 0, \quad i \in I_0(\bar{x}), \\ \Phi'_{I_1(\bar{x})}(\bar{x})\xi = 0, \quad \xi_{I_2(\bar{x})} = 0 \end{gathered} \tag{3.53}$$

has the unique solution $\xi = 0$. By the definition of the directional derivative, one can easily conclude that for any $\xi \in \mathbf{R}^n$

$$(\Psi'_{\mathrm{NR}}(\bar{x};\,\xi))_i = \begin{cases} \min\{\xi_i,\, \langle \Phi'_i(\bar{x}),\, \xi\rangle\} & \text{if } i \in I_0(\bar{x}),\\ \langle \Phi'_i(\bar{x}),\, \xi\rangle & \text{if } i \in I_1(\bar{x}),\\ \xi_i & \text{if } i \in I_2(\bar{x}), \end{cases}$$

$$(\Psi'_{\mathrm{FB}}(\bar{x};\,\xi))_i = \begin{cases} \xi_i + \langle \Phi'_i(\bar{x}),\, \xi\rangle - \sqrt{\xi_i^2 + \langle \Phi'_i(\bar{x}),\, \xi\rangle^2} & \text{if } i \in I_0(\bar{x}),\\ \langle \Phi'_i(\bar{x}),\, \xi\rangle & \text{if } i \in I_1(\bar{x}),\\ \xi_i & \text{if } i \in I_2(\bar{x}). \end{cases}$$

These formulas immediately imply the following fact.

Proposition 3.21. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable at $\bar{x} \in \mathbf{R}^n$, and let $\bar{x}$ be a solution of the NCP (3.32).*

Then the set of all $\xi \in \mathbf{R}^n$ satisfying the equality $\Psi'_{\mathrm{NR}}(\bar{x};\,\xi) = 0$, and the set of all $\xi \in \mathbf{R}^n$ satisfying the equality $\Psi'_{\mathrm{FB}}(\bar{x};\,\xi) = 0$, both coincide with the solution set of the system (3.53).

In particular, semistability of a solution $\bar{x}$ of the NCP (3.32) is equivalent to saying that Ψ_{NR} and/or Ψ_{FB} have nonzero directional derivatives at $\bar{x}$ in any nonzero direction, which in its turn is equivalent to any of the equivalent error bounds

$$\|x - \bar{x}\| = O(\|\Psi_{\mathrm{NR}}(x)\|)$$

and

$$\|x - \bar{x}\| = O(\|\Psi_{\mathrm{FB}}(x)\|)$$

to hold as $x \in \mathbf{R}^n$ tends to $\bar{x}$ (the latter follows from Proposition 1.64). Combining these observations with Propositions 1.77 and 3.8, we obtain the following result.

Proposition 3.22. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable near $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$. Let $\bar{x}$ be a solution of the NCP (3.32), and assume that either Ψ_{NR} or Ψ_{FB} is BD-regular at $\bar{x}$.*

Then $\bar{x}$ is semistable.

The absence of converse implications is demonstrated by Examples 3.12 and 3.13; semistability in these examples follows, e.g., from Proposition 1.34, since the corresponding systems (3.53) have only trivial solutions.

The relations between all the regularity conditions discussed in this section are presented as a flowchart in Fig. 3.1, where arrows mean implications, and the absence of arrows means the absence of implications.

In order to compare the semismooth Newton method for NCP with the Josephy–Newton method developed in Sect. 3.1, first recall that local superlinear convergence of the latter was established in Theorem 3.2 under the assumptions of semistability and hemistability of the solution $\bar{x}$ in question. As discussed above, semistability is a weaker property than BD-regularity

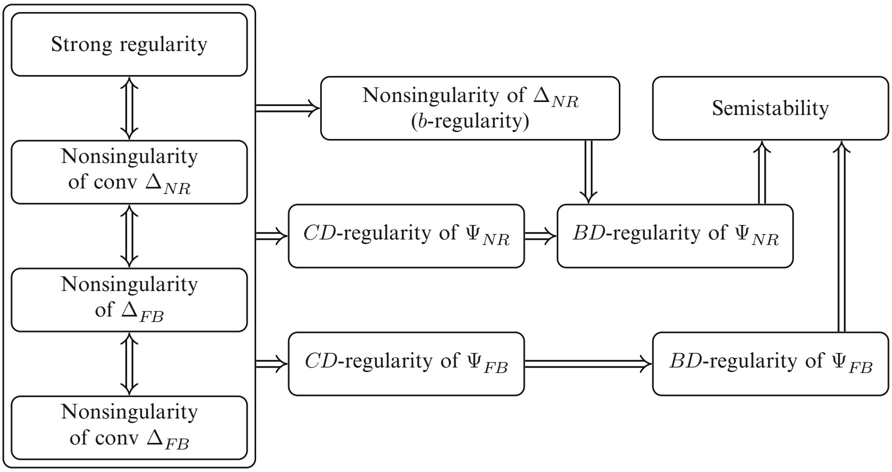

Fig. 3.1 Regularity conditions for complementarity problems

of $\Psi_{\rm NR}$ and/or $\Psi_{\rm FB}$ at $\bar{x}$, the property required in Theorem 3.9 for local superlinear convergence of the semismooth Newton method. On the other hand, Example 3.3 (which is NCP with $\Delta_{\rm NR}(\bar{x}) = \{-1, 1\}$, $\partial_B \Psi_{\rm FB}(\bar{x}) = \{-\sqrt{2}, \sqrt{2}\}$, identically to Example 3.16) demonstrates that neither BD-regularity of $\Psi_{\rm NR}$ nor BD-regularity of $\Psi_{\rm FB}$ implies hemistability. Therefore, neither of the conditions assuring local superlinear convergence of the two methods for NCP (3.32) is implied by the other.

Note also that the iteration subproblems of the semismooth Newton method are linear equations, certainly simpler than the linear complementarity problems solved at each iteration of the Josephy–Newton method for NCP.

It is instructive to consider the following interpretation of an iteration of the semismooth Newton method for NCP reformulated using the natural residual. It is similar to the one given in Sect. 3.1 for the Josephy–Newton method; it is also in the spirit of the piecewise Newton method.

Let $\{x^k\}$ be an iterative sequence of the semismooth Newton method or the piecewise Newton method, convergent to the solution $\bar{x}$ of the NCP (3.32). Then from (3.37) and from Proposition 3.10, it follows that for all k large enough there exists a partition (I_1^k, I_2^k) of $I_0(\bar{x})$ (i.e., a pair of index sets satisfying $I_1^k \cup I_2^k = I_0(\bar{x})$, $I_1^k \cap I_2^k = \emptyset$) such that

$$\{i \in I_0(\bar{x}) \mid x_i^k > \Phi_i(x^k)\} \subset I_1^k, \quad \{i \in I_0(\bar{x}) \mid x_i^k < \Phi_i(x^k)\} \subset I_2^k, \tag{3.54}$$

and

$$\begin{aligned} \min\left\{x^k_{I_1(\bar{x})\cup I_1^k},\, \Phi_{I_1(\bar{x})\cup I_1^k}(x^k)\right\} + \Phi'_{I_1(\bar{x})\cup I_1^k}(x^k)(x^{k+1} - x^k) &= 0, \\ \min\left\{x^k_{I_2(\bar{x})\cup I_2^k},\, \Phi_{I_2(\bar{x})\cup I_2^k}(x^k)\right\} + \left(x^{k+1}_{I_2(\bar{x})\cup I_2^k} - x^k_{I_2(\bar{x})\cup I_2^k}\right) &= 0. \end{aligned}$$

Taking into account (3.54) and the definition of $I_0(\bar{x})$, $I_1(\bar{x})$ and $I_2(\bar{x})$, the last system can be written in the form

$$\Phi_{I_1(\bar{x})\cup I_1^k}(x^k) + \Phi'_{I_1(\bar{x})\cup I_1^k}(x^k)(x^{k+1} - x^k) = 0, \quad x^{k+1}_{I_2(\bar{x})\cup I_2^k} = 0. \tag{3.55}$$

Therefore, the iteration in question can be interpreted as that of the usual Newton method for the system of equations

$$\Phi_{I_1(\bar{x})\cup I_1^k}(x) = 0, \quad x_{I_2(\bar{x})\cup I_2^k} = 0. \tag{3.56}$$

Systems (3.55) and (3.56) coincide with (3.9) and (3.10), respectively. At the same time, the partitions $(I_1^k,\, I_2^k)$ may change differently along the iterations of the semismooth Newton method, of the piecewise Newton method, and of the Josephy–Newton method. Essentially, these are the underlying reasons that lead to different regularity conditions needed for local superlinear convergence of these methods. However, if the strict complementarity condition $I_0(\bar{x}) = \emptyset$ holds, then the system (3.56) takes the form

$$\Phi_{I_1(\bar{x})}(x) = 0, \quad x_{I_2(\bar{x})} = 0.$$

In this case, the semismooth Newton method, the piecewise Newton method, and the Josephy–Newton method all reduce (locally, of course) to the usual Newton method for this system of equations.

Considerations above lead to the following idea. Suppose that one can (locally) identify the index set $I_0(\bar{x})$. Then for any *fixed* partition $(I_1,\, I_2)$ of $I_0(\bar{x})$, the point $\bar{x}$ solves the equation

$$\Phi_{I_1(\bar{x})\cup I_1}(x) = 0, \quad x_{I_2(\bar{x})\cup I_2} = 0.$$

In particular, one can then search for $\bar{x}$ applying the Newton method to this system. The latter can require weaker regularity assumptions than the methods with varying partitions $(I_1^k,\, I_2^k)$. We shall get back to this idea in Sect. 3.4.

As for $\Psi = \Psi_{\mathrm{FB}}$, the iteration of the semismooth Newton method can be interpreted as that of the *perturbed* Newton method for some underlying system of equations, also varying along the iterations in general.

3.2.2 Extension to Mixed Complementarity Problems

Consider now the problem setting more general than NCP, namely, the mixed complementarity problem (MCP), which is the variational inequality (VI) with bound constraints:

$$u \in [a,\, b], \quad \langle \Phi(u),\, v - u\rangle \geq 0 \quad \forall\, v \in [a,\, b], \tag{3.57}$$

where

$$[a,\, b] = \{u \in \mathbf{R}^\nu \mid a_i \le u_i \le b_i,\ i = 1,\, \dots,\, \nu\}$$

is a (generalized) box, $\Phi : \mathbf{R}^\nu \to \mathbf{R}^\nu$ is a smooth mapping, $a_i \in \mathbf{R} \cup \{-\infty\}$, $b_i \in \mathbf{R} \cup \{+\infty\}$, $a_i < b_i$, $i = 1,\, \dots,\, \nu$. Recall that problem (3.57) can be equivalently stated as follows:

$$u \in [a,\, b], \quad \Phi_i(u) \begin{cases} \ge 0 & \text{if } u_i = a_i, \\ = 0 & \text{if } a_i < u_i < b_i, \quad i = 1,\, \dots,\, \nu. \\ \le 0 & \text{if } u_i = b_i, \end{cases} \tag{3.58}$$

Let ψ be a complementarity function satisfying the following additional assumptions:

$$\psi(a,\, b) < 0 \quad \forall\, a > 0,\, \forall\, b < 0, \quad \psi(a,\, b) > 0 \quad \forall\, a > 0,\, \forall\, b > 0.$$

It is easy to check that the natural residual function (3.30), the Fischer–Burmeister function (3.31), and the smooth complementarity function defined in (3.36), all satisfy these additional assumptions.

With a complementarity function at hand, the MCP (3.58) can be equivalently formulated as the equation

$$\Psi(u) = 0,$$

where $\Psi : \mathbf{R}^\nu \to \mathbf{R}^\nu$ is given by

$$\Psi_i(u) = \begin{cases} \Phi_i(u) & \text{if } i \in I_\Phi, \\ \psi(u_i - a_i, \Phi_i(u)) & \text{if } i \in I_a, \\ -\psi(b_i - u_i, -\Phi_i(u)) & \text{if } i \in I_b, \\ \psi(u_i - a_i, -\psi(b_i - u_i,\, -\Phi_i(u))) & \text{if } i \in I_{ab}, \end{cases} \tag{3.59}$$

with

$$\begin{aligned} I_\Phi &= \{i = 1,\, \dots,\, \nu \mid a_i = -\infty,\ b_i = +\infty\}, \\ I_a &= \{i = 1,\, \dots,\, \nu \mid a_i > -\infty,\ b_i = +\infty\}, \\ I_b &= \{i = 1,\, \dots,\, \nu \mid b_i = -\infty,\ b_i < +\infty\}, \\ I_{ab} &= \{i = 1,\, \dots,\, \nu \mid a_i > -\infty,\ b_i < +\infty\}. \end{aligned}$$

If ψ is the natural residual defined in (3.30), it can be seen that (3.59) is equivalent to the single equality

$$\Psi(u) = \min\{u - a, -\min\{b - u,\, -\Phi(u)\}\} \tag{3.60}$$

with the natural convention that $\min\{t,\, +\infty\} = t$ for any $t \in \mathbf{R}$. This partially explains the selection of signs of the components in (3.59). Note also that (3.60) can be equivalently written as follows:

$$\Psi(u) = u - \pi_{[a,\, b]}(u - \Phi(u)),$$

which agrees with the natural residual mapping for VI defined in Sect. 1.3.2.

The developments of the previous section can be extended to MCP and its equation reformulations. The extension itself is rather technical, but is not concerned with any principal difficulties. The semismoothness properties of Ψ are immediate, and it remains to extend the regularity conditions investigated for NCP to this different but largely similar setting. Conceptually, the conclusions obtained above for NCP remain valid for MCP as well. We cite [20, 49, 50, 80, 134] for details.

Consider now another special case of MCP, different from NCP. Namely, we shall concentrate next on the KKT system

$$\begin{gathered} F(x) + (h'(x))^{\mathrm{T}}\lambda + (g'(x))^{\mathrm{T}}\mu = 0, \quad h(x) = 0, \\ \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu,\, g(x)\rangle = 0, \end{gathered} \tag{3.61}$$

where $F : \mathbf{R}^n \to \mathbf{R}^n$, $h : \mathbf{R}^n \to \mathbf{R}^l$, and $g : \mathbf{R}^n \to \mathbf{R}^m$ are smooth mappings. Define the mapping $G : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n$,

$$G(x,\, \lambda,\, \mu) = F(x) + (h'(x))^{\mathrm{T}}\lambda + (g'(x))^{\mathrm{T}}\mu,$$

and recall that the system (3.61) is a particular instance of the MCP (3.57) with $\nu = n + l + m$, with the mapping $\Phi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ given by

$$\Phi(u) = (G(x,\, \lambda,\, \mu),\, h(x),\, -g(x)), \quad u = (x,\, \lambda,\, \mu), \tag{3.62}$$

and with

$$\begin{gathered} a_i = -\infty,\ i = 1,\, \dots,\, n + l, \quad a_i = 0,\ i = n + l + 1,\, \dots,\, n + l + m, \\ b_i = +\infty,\ i = 1,\, \dots,\, n + l + m. \end{gathered} \tag{3.63}$$

Application of semismooth Newton methods to KKT systems was originally studied in [224].

For Φ, a and b defined in (3.62), (3.63), the mapping $\Psi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ defined according to (3.59) takes the form

$$\Psi(u) = (G(x,\, \lambda,\, \mu),\, h(x),\, \psi(\mu,\, -g(x))), \quad u = (x,\, \lambda,\, \mu). \tag{3.64}$$

In the rest of this section, we deal with KKT systems only. Therefore, with no risk of confusion, we shall employ the same notation as in Sect. 3.2.1 for NCP: we denote by Ψ_{NR} the mapping Ψ defined in (3.64) with ψ from (3.30), and we denote by Ψ_{FB} the mapping Ψ defined in (3.64) with the function ψ from (3.31).

The *semismooth Newton method* for the KKT system (3.61) is then stated as follows.

Algorithm 3.23 Choose a complementarity function $\psi : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$, and define the mapping $\Psi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ according to (3.64). Choose $u^0 = (x^0,\, \lambda^0,\, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and set $k = 0$.

1. If $\Psi(u^k) = 0$, stop.
2. Compute some element $J_k \in \partial_B \Psi(u^k)$. Compute the next iterate $u^{k+1} = (x^{k+1}, \lambda^{k+1}, \mu^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ as a solution of the linear equation

$$\Psi(u^k) + J_k(u - u^k) = 0. \tag{3.65}$$

3. Increase k by 1 and go to step 1.

Similarly to Proposition 3.24, we derive the semismoothness of the KKT reformulations.

Proposition 3.24. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be (strongly) semismooth at $x \in \mathbf{R}^n$, and let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable in a neighborhood of x, with their derivatives being (strongly) semismooth at x.*

Then for any $(\lambda, \mu) \in \mathbf{R}^l \times \mathbf{R}^m$, the mappings Ψ_{NR} and Ψ_{FB} are both (strongly) semismooth at (x, λ, μ).

From Theorem 2.42 and Proposition 3.24, we derive the local superlinear convergence result for Algorithm 3.23.

Theorem 3.25. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be semismooth at a point $\bar{x} \in \mathbf{R}^n$, and let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable in a neighborhood of $\bar{x}$, with their derivatives being semismooth at $\bar{x}$. Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ with some $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be a solution of the KKT system (3.61), and assume that Ψ_{NR} (Ψ_{FB}) is BD-regular at $(\bar{x}, \bar{\lambda}, \bar{\mu})$.*

Then any starting point $(x^0, \lambda^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ defines a particular iterative sequence of Algorithm 3.23 with ψ defined according to (3.30) (according to (3.31), respectively), any such sequence converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and the rate of convergence is superlinear. If F and the derivatives of h and g are strongly semismooth at $\bar{x}$, then the rate of convergence is quadratic.

The next task is to interpret the regularity conditions employed in Theorem 3.25. Similarly to Propositions 3.10 and 3.11, we derive the descriptions of B-differentials of the KKT reformulations.

Proposition 3.26. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighborhood of $x \in \mathbf{R}^n$, with its derivative being continuous at x, and let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of x, with their second derivatives being continuous at x.*

Then for any $(\lambda, \mu) \in \mathbf{R}^l \times \mathbf{R}^m$, the set $\partial_B \Psi_{\mathrm{NR}}(u)$ for $u = (x, \lambda, \mu)$ consists of all the matrices

$$J = \begin{pmatrix} \dfrac{\partial G}{\partial x}(x, \lambda, \mu) & (h'(x))^{\mathrm{T}} & (g'(x))^{\mathrm{T}} \\ h'(x) & 0 & 0 \\ -\operatorname{diag}(\delta(x, \mu))g'(x) & 0 & I - \operatorname{diag}(\delta(x, \mu)) \end{pmatrix},$$

where

$$\delta_i(x,\,\mu) = \begin{cases} 1 \;\textit{or}\; 0 & \textit{if}\; \mu_i = -g_i(x), \\ 1 & \textit{if}\; \mu_i > -g_i(x), \\ 0 & \textit{if}\; \mu_i < -g_i(x), \end{cases} \quad i = 1,\,\ldots,\,m.$$

Proposition 3.27. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighborhood of $x \in \mathbf{R}^n$, with its derivative being continuous at x, and let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of x, with their second derivatives being continuous at x.*

Then for any point $(\lambda,\,\mu) \in \mathbf{R}^l \times \mathbf{R}^m$, any matrix $J \in \partial_B \Psi_{\mathrm{FB}}(u)$ for $u = (x,\,\lambda,\,\mu)$ satisfies the equality

$$J = \begin{pmatrix} \dfrac{\partial G}{\partial x}(x,\,\lambda,\,\mu) & (h'(x))^{\mathrm{T}} & (g'(x))^{\mathrm{T}} \\ h'(x) & 0 & 0 \\ -\operatorname{diag}(\delta_\alpha(x,\,\mu))g'(x) & 0 & \operatorname{diag}(\delta_\beta(x,\,\mu)) \end{pmatrix},$$

where

$$(\delta_\alpha)_i(x,\,\mu) = \begin{cases} 1 + \dfrac{g_i(x)}{\sqrt{\mu_i^2 + (g_i(x))^2}} & \textit{if}\; \mu_i \neq 0 \;\textit{or}\; g_i(x) \neq 0, \\ \alpha_i & \textit{if}\; \mu_i = g_i(x) = 0, \end{cases} \quad i = 1,\,\ldots,\,m,$$

$$(\delta_\beta)_i(x,\,\mu) = \begin{cases} 1 - \dfrac{\mu_i}{\sqrt{\mu_i^2 + (g_i(x))^2}} & \textit{if}\; \mu_i \neq 0 \;\textit{or}\; g_i(x) \neq 0, \\ \beta_i & \textit{if}\; \mu_i = g_i(x) = 0, \end{cases} \quad i = 1,\,\ldots,\,m,$$

with some $(\alpha_i,\,\beta_i) \in S$, $i = 1,\,\ldots,\,m$, where S is defined in (3.39).

Moreover, if the gradients $g_i'(x)$, $i \in \{i = 1,\,\ldots,\,m \mid \mu_i = g_i(x) = 0\}$ are linearly independent, then $\partial_B \Psi_{\mathrm{FB}}(x,\,\lambda,\,\mu)$ coincides with the set of all the matrices of the specified form.

Note that unlike in the case of NCP, the statements above for KKT systems provide not only outer estimates of B-differentials but also their exact characterizations (for Ψ_{FB}, this is true under the additional linear independence assumption). This is due to the fact that the primal variable x and the dual variable μ are decoupled in the complementarity part of (3.64).

Let $(\bar{x},\,\bar{\lambda},\,\bar{\mu}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ be a solution of the KKT system (3.61). From Proposition 3.26 it follows that the BD-regularity of Ψ_{NR} at $(\bar{x},\,\bar{\lambda},\,\bar{\mu})$ is equivalent to nonsingularity of the matrices

$$\begin{pmatrix} \dfrac{\partial G}{\partial x}(\bar{x},\,\bar{\lambda},\,\bar{\mu}) & (h'(\bar{x}))^{\mathrm{T}} & (g'_{A_+(\bar{x},\,\bar{\mu})\cup K}(\bar{x}))^{\mathrm{T}} \\ h'(x) & 0 & 0 \\ g'_{A_+(\bar{x},\,\bar{\mu})\cup K}(\bar{x}) & 0 & 0 \end{pmatrix} \tag{3.66}$$

for all index sets $K \subset A_0(\bar{x},\,\bar{\mu})$. The latter condition was introduced in [71] under the name of *quasi-regularity* of the solution $(\bar{x},\,\bar{\lambda},\,\bar{\mu})$. Let us state separately the important fact that quasi-regularity subsumes the LICQ. This follows by considering (3.66) with $K = A_0(\bar{x}, \bar{\mu})$.

Proposition 3.28. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$, and let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x}$, with their second derivatives being continuous at $\bar{x}$. Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ with some $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be a quasi-regular solution of the KKT system (3.61).*

Then $\bar{x}$ satisfies the LICQ.

Furthermore, from Proposition 3.27 it follows that the BD-regularity of Ψ_{FB} at $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is implied by nonsingularity of the matrices

$$\begin{pmatrix} \dfrac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) & (h'(\bar{x}))^{\mathrm{T}} & (g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x}))^{\mathrm{T}} & (g'_{A_0(\bar{x}, \bar{\mu})}(\bar{x}))^{\mathrm{T}} \\ h'(x) & 0 & 0 & 0 \\ -g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x}) & 0 & 0 & 0 \\ -\operatorname{diag}(\alpha) g'_{A_0(\bar{x}, \bar{\mu})}(\bar{x}) & 0 & 0 & \operatorname{diag}(\beta) \end{pmatrix} \tag{3.67}$$

for all $\alpha = (\alpha_i,\ i \in A_0(\bar{x}, \bar{\mu}))$ and $\beta = (\beta_i,\ i \in A_0(\bar{x}, \bar{\mu}))$ with $(\alpha_i, \beta_i) \in S$, $i \in A_0(\bar{x}, \bar{\mu})$. Somewhat surprisingly, the converse implication is also valid, which can be derived from the following counterpart of Proposition 3.28.

Proposition 3.29. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$, and let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x}$, with their second derivatives being continuous at $\bar{x}$. Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ with some $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be a solution of the KKT system (3.61) such that Ψ_{FB} is BD-regular at $(\bar{x}, \bar{\lambda}, \bar{\mu})$.*

Then $\bar{x}$ satisfies the LICQ.

Proof. Fix any sequence $\{\mu^k\} \subset \mathbf{R}^m$ such that $\mu^k_{A(\bar{x})} > 0$, $\mu^k_{\{1, \ldots, m\} \setminus A(\bar{x})} = 0$ for all k, and $\{\mu^k_{A(\bar{x})}\} \to \bar{\mu}_{A(\bar{x})}$. Then the sequence $\{(\bar{x}, \bar{\lambda}, \mu^k)\}$ converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and it can be easily seen that for every k the mapping Ψ_{FB} is differentiable at $(\bar{x}, \bar{\lambda}, \mu^k)$, and the sequence of its Jacobians converges to the matrix

$$J = \begin{pmatrix} \dfrac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) & (h'(\bar{x}))^{\mathrm{T}} & (g'_{A(\bar{x})}(\bar{x}))^{\mathrm{T}} & (g'_{\{1, \ldots, m\} \setminus A(\bar{x})}(\bar{x}))^{\mathrm{T}} \\ h'(x) & 0 & 0 & 0 \\ -g'_{A(\bar{x})}(\bar{x}) & 0 & 0 & 0 \\ 0 & 0 & 0 & I \end{pmatrix}$$

(perhaps after an appropriate re-ordering of rows and columns). The stated result is now evident. $\square$

Combining the last assertion of Proposition 3.27 with Proposition 3.29, we conclude that if Ψ_{FB} is BD-regular at $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$, then $\partial_B \Psi_{\mathrm{FB}}(\bar{u})$

consists of all the matrices of the form (3.67) with $\alpha = (\alpha_i,\ i \in A_0(\bar{x},\, \bar{\mu}))$ and $\beta = (\beta_i,\ i \in A_0(\bar{x},\, \bar{\mu}))$ satisfying $(\alpha_i,\, \beta_i) \in S$, $i \in A_0(\bar{x},\, \bar{\mu})$. Therefore, BD-regularity of Ψ_{FB} at $\bar{u}$ is equivalent to nonsingularity of all such matrices.

One of the evident consequences of the established equivalence is that BD-regularity of Ψ_{FB} at $\bar{u}$ implies quasi-regularity of this solution. The converse implication does not hold. This is demonstrated by Examples 3.3 and 3.16. These examples discuss the primal (NCP) form of optimality conditions for the underlying optimization problems, but it can be easily checked (employing Proposition 3.26, or directly) that for the primal-dual (KKT) optimality systems it holds that

$$\partial_B \Psi_{\mathrm{NR}}(\bar{u}) = \left\{ \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix},\, \begin{pmatrix} -1 & -1 \\ 0 & 1 \end{pmatrix} \right\},$$

and both matrices in the right-hand side are nonsingular. At the same time, according to Proposition 3.27,

$$\partial_B \Psi_{\mathrm{FB}}(\bar{u}) = \left\{ \begin{pmatrix} -1 & -1 \\ \alpha & \beta \end{pmatrix} \,\middle|\, (\alpha,\, \beta) \in S \right\}.$$

Taking $\alpha = \beta = (\sqrt{2} - 1)/\sqrt{2}$ or $\alpha = \beta = (\sqrt{2} + 1)/\sqrt{2}$, we get singular matrices.

Another example demonstrating the absence of the implication in question is taken from [71].

Example 3.30. Let $n = m = 2$, $l = 0$, $F(x) = f'(x)$, $f(x) = x_1^2 + x_2^2 + 4x_1x_2$, $g(x) = (-x_1,\, -x_2)$. Then $\bar{x} = 0$ is the solution and the unique stationary point of the optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \leq 0, \end{array} \tag{3.68}$$

and the unique associated Lagrange multiplier is $\bar{\mu} = 0$. Therefore, $\bar{u} = (\bar{x},\, \bar{\mu})$ is the unique solution of the KKT system (3.61), and it can be directly seen that this solution is quasi-regular.

According to Proposition 3.27,

$$\partial_B \Psi_{\mathrm{FB}}(\bar{u}) = \left\{ \begin{pmatrix} 2 & 4 & -1 & 0 \\ 4 & 2 & 0 & -1 \\ \alpha_1 & 0 & \beta_1 & 0 \\ 0 & \alpha_2 & 0 & \beta_2 \end{pmatrix} \,\middle|\, (\alpha_i,\, \beta_i) \in S,\, i = 1,\, 2 \right\}.$$

Taking $\alpha = (2,\, 2)$, $\beta = (1,\, 1)$, we get a singular matrix.

Furthermore, in [70] it was demonstrated that CD-regularity of Ψ_{FB} at $\bar{u} = (\bar{x},\, \bar{\lambda},\, \bar{\mu})$ is equivalent to strong regularity of this solution. Our next result shows that these properties are further equivalent to CD-regularity of Ψ_{NR} at $\bar{u}$, and to BD-regularity of Ψ_{FB} at $\bar{u}$.

Proposition 3.31. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$, and let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x}$, with their second derivatives being continuous at $\bar{x}$. Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ with some $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be a solution of the KKT system (3.61).*

Then the following properties are equivalent:

(a) Ψ_{NR} is CD-regular at $(\bar{x}, \bar{\lambda}, \bar{\mu})$.
(b) Ψ_{FB} is BD-regular at $(\bar{x}, \bar{\lambda}, \bar{\mu})$.
(c) Ψ_{FB} is CD-regular at $(\bar{x}, \bar{\lambda}, \bar{\mu})$.
(d) Solution $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is strongly regular.

Proof. We prove only the equivalence of (a), (b), and (c); for the equivalence of (c) and (d) we address the reader to [70]. Another way to approach all these equivalences is to develop and then apply the counterpart for the current setting of Corollary 3.20 dealing with an MCP; see [134].

Define the sets

$$S_{(a)} = \{(a,\, b) \in \mathbf{R}^2 \mid a + b = 1,\ a \geq 0,\ b \geq 0\},$$

$$S_{(b)} = S = \{(a,\, b) \in \mathbf{R}^2 \mid (a-1)^2 + (b-1)^2 = 1\},$$

$$S_{(c)} = \operatorname{conv} S = \{(a,\, b) \in \mathbf{R}^2 \mid (a-1)^2 + (b-1)^2 \leq 1\}$$

(see Fig. 3.2). The discussion above (relying on Propositions 3.26, 3.27, and 3.29) allows to conclude that (a) ((b), (c)) is equivalent to saying that the matrix J defined in (3.67) is nonsingular for all $\alpha = (\alpha_i,\ i \in A_0(\bar{x}, \bar{\mu}))$ and $\beta = (\beta_i,\ i \in A_0(\bar{x}, \bar{\mu}))$ such that $(\alpha_i, \beta_i) \in S_{(a)}$ ($S_{(b)}$, $S_{(c)}$, respectively), $i \in A_0(\bar{x}, \bar{\mu})$.

A key observation is the following: for any point $(a, b) \in S_{(c)}$ there exists $t_{(a)} > 0$ such that $(t_{(a)}a,\, t_{(a)}b) \in S_{(a)}$, and there exists $t_{(b)} > 0$ such that $(t_{(b)}a,\, t_{(b)}b) \in S_{(b)}$ (see Fig. 3.2). This implies that any matrix J defined in (3.67) with some $\alpha = (\alpha_i,\ i \in A_0(\bar{x}, \bar{\mu}))$ and $\beta = (\beta_i,\ i \in A_0(\bar{x}, \bar{\mu}))$ such that $(\alpha_i, \beta_i) \in S_{(c)}$, $i \in A_0(\bar{x}, \bar{\mu})$, can be transformed into matrices of the same form but with α and β satisfying the relations $(\alpha_i, \beta_i) \in S_{(a)}$ and $(\alpha_i, \beta_i) \in S_{(b)}$, respectively, $i \in A_0(\bar{x}, \bar{\mu})$, and this transformation can be achieved by multiplication of some rows of J by suitable positive numbers. In particular, such matrices are singular or nonsingular simultaneously, which gives the needed equivalence. □

From Propositions 1.28 and 3.31 we further derive the following sufficient condition for the equivalent properties (a)–(c) in Proposition 3.31 in the case of an optimization-related KKT system. Consider the mathematical programming problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0,\ g(x) \leq 0. \end{array} \tag{3.69}$$

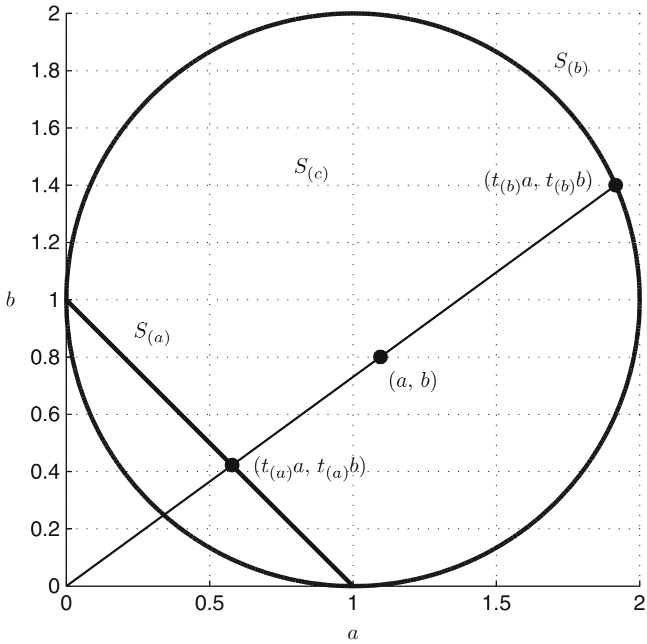

Fig. 3.2 Sets $S_{(a)}$, $S_{(b)}$ and $S_{(c)}$

Recall that stationary points of problem (3.69) and the associated Lagrange multipliers are characterized by the KKT system (3.61) with

$$F(x) = f'(x), \quad x \in \mathbf{R}^n. \tag{3.70}$$

Let $L : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}$ be the Lagrangian of problem (1.44):

$$L(x,\, \lambda,\, \mu) = f(x) + \langle \lambda,\, h(x)\rangle + \langle \mu,\, g(x)\rangle.$$

Proposition 3.32. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (3.69)*, and let $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be an associated Lagrange multiplier.*

If $\bar{x}$ and $(\bar{\lambda},\, \bar{\mu})$ satisfy the LICQ and the SSOSC

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in C_+(\bar{x},\, \bar{\mu}) \setminus \{0\}, \tag{3.71}$$

where

$$C_+(\bar{x},\, \bar{\mu}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+(\bar{x},\, \bar{\mu})}(\bar{x})\xi = 0\}, \tag{3.72}$$

then the equivalent properties (a)–(c) in Proposition 3.31 are valid for F defined in (3.70).

Regarding the CD-regularity of Ψ_{NR} and Ψ_{FB}, this fact can also be easily derived directly, employing only Propositions 3.26 and 3.27.

Following the lines of the corresponding discussion for NCP, we now briefly consider the relation between the BD-regularity of Ψ_{NR} and Ψ_{FB} at a solution $(\bar{x}, \bar{\lambda}, \bar{\mu})$ and the semistability of this solution. Recall that the latter is the key property for local superlinear convergence of the Josephy–Newton method. By Proposition 1.35, $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is semistable if and only if the system

$$\begin{gathered}
\frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi + (h'(\bar{x}))^{\mathrm{T}}\eta + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0,\\
h'(\bar{x})\xi = 0, \quad g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x})\xi = 0,\\
\zeta_{A_0(\bar{x}, \bar{\mu})} \geq 0, \quad g'_{A_0(\bar{x}, \bar{\mu})}(\bar{x})\xi \leq 0, \quad \zeta_i\langle g_i'(\bar{x}), \xi\rangle = 0, \ i \in A_0(\bar{x}, \bar{\mu}),\\
\zeta_{\{1, \ldots, m\}\setminus A(\bar{x})} = 0
\end{gathered} \tag{3.73}$$

has the unique solution $(\xi, \eta, \zeta) = (0, 0, 0)$. By direct computation of directional derivatives of Ψ_{NR} and Ψ_{FB} at $(\bar{x}, \bar{\lambda}, \bar{\mu})$, we obtain the following result.

Proposition 3.33. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable at a point $\bar{x} \in \mathbf{R}^n$, let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable at $\bar{x}$, and let $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ with some $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be a solution of the KKT system (3.61).*

Then the set of all $v = (\xi, \eta, \zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ satisfying the equality $\Psi'_{\mathrm{NR}}(\bar{u}; v) = 0$, and the set of all $v = (\xi, \eta, \zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ satisfying the equality $\Psi'_{\mathrm{FB}}(\bar{u}; v) = 0$, both coincide with the solution set of the system (3.73).

Combining Proposition 3.33 with Propositions 1.77 and 3.24, we immediately obtain the following result.

Proposition 3.34. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable near $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$, and let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable near $\bar{x}$, with their second derivatives being continuous at $\bar{x}$. Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ with some $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be a quasi-regular solution of the KKT system (3.61).*

Then $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is semistable.

Observe that the assertion of Proposition 3.34 is also valid under the stronger equivalent assumptions specified in Proposition 3.31, and in particular, under the BD-regularity of Ψ_{FB} at $(\bar{x}, \bar{\lambda}, \bar{\mu})$.

The absence of the converse implication (from semistability to quasi-regularity) is demonstrated by the following two examples taken from [140]. Recall that according to Proposition 1.37, for the KKT system related to the optimization problem (3.69), semistability of the solution $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is implied by a combination of the SMFCQ and the SOSC

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{3.74}$$

where

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A(\bar{x})}(\bar{x})\xi \le 0,\ \langle f'(\bar{x}),\, \xi\rangle \le 0\}$$

is the critical cone of problem (3.69) at $\bar{x}$. Moreover, recall also that the semistability is in fact equivalent to the combination of the SMFCQ and the SOSC, if $\bar{x}$ is a local solution of problem (3.69).

Example 3.35. Let $n = m = 2$, $l = 0$, $F(x) = f'(x)$, $f(x) = (x_1 + x_2)^2/2$, $g(x) = (-x_1,\, -x_2)$. Then $\bar{x} = 0$ is the solution and the unique stationary point of the optimization problem (3.68), and the unique associated Lagrange multiplier is $\bar{\mu} = 0$. Therefore, $\bar{u} = (\bar{x},\, \bar{\mu})$ is the unique solution of the KKT system (3.61), and this solution is not quasi-regular: the matrix given by (3.66) with $K = \emptyset$ is singular.

At the same time, note that $\bar{x}$ and $\bar{\mu}$ satisfy the LICQ (and hence, the SMFCQ) and the SOSC (but not the SSOSC). Therefore, $\bar{u}$ is semistable by Proposition 1.37.

Example 3.36. Let $n = 2$, $l = 0$, $m = 3$, $f(x) = x_1 + (x_1 + x_2)^2/2$, and $g(x) = (-x_1,\, -x_2,\, -x_1 - x_2)$, $F(x) = f'(x)$. Then $\bar{x} = 0$ is the solution and the unique stationary point of the optimization problem (3.68), and the unique associated Lagrange multiplier is $\bar{\mu} = (1,\, 0,\, 0)$. Therefore, $\bar{u} = (\bar{x},\, \bar{\mu})$ is the unique solution of the KKT system (3.61). Since $\bar{x}$ violates the LICQ, it cannot be quasi-regular.

At the same time $\bar{x}$ and $\bar{\mu}$ satisfy the SMFCQ and the SSOSC (hence, the SOSC), and therefore, $\bar{u}$ is semistable by Proposition 1.37.

Summarizing the relations obtained above for KKT systems, we conclude that each of the four equivalent conditions specified in Proposition 3.31 is strictly stronger than the BD-regularity of Ψ_{NR}, while the latter is strictly stronger than semistability.

Recall now that another important ingredient of local superlinear convergence of the Josephy–Newton method in Theorem 3.2 is hemistability of the solution $\bar{u} = (\bar{x},\, \bar{\lambda},\, \bar{\mu})$ in question. Example 3.3 demonstrates that the BD-regularity of Ψ_{NR} at $\bar{u}$ does not imply hemistability of this solution. As already mentioned above, this example discusses the primal (NCP) form of optimality conditions, but it can be easily checked that the stated conclusions are valid for the primal-dual (KKT) form as well.

However, as we show next, if $\bar{x}$ is a local solution of problem (3.69), then hemistability of $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ for the corresponding KKT system is implied by semistability. Hence, in this case, conditions assuring local superlinear convergence of the Josephy–Newton method are weaker than those required for the semismooth Newton method. The assertion of Proposition 3.37 would also be of relevance for convergence analysis of the SQP method for optimization in Chap. 4.

Proposition 3.37. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives* *being continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (3.69), and let $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be an associated Lagrange multiplier.*

If $\bar{x}$ is a local solution of problem (3.69), then semistability of the solution $(\bar{x}, \bar{\lambda}, \bar{\mu})$ of the KKT system (3.61) with F defined in (3.70) implies hemistability of this solution.

Proof. Assume that $\bar{x}$ is a local solution of problem (3.69), and recall again that according to Proposition 1.37, semistability of $(\bar{x}, \bar{\lambda}, \bar{\mu})$ implies the SMFCQ and the SOSC (3.74).

Consider the parametric optimization problem

$$\begin{array}{ll} \text{minimize} & \langle f'(x), \xi\rangle + \dfrac{1}{2}\left\langle \dfrac{\partial^2 L}{\partial x^2}(x, \lambda, \mu)\xi, \xi\right\rangle \\ \text{subject to} & h(x) + h'(x)\xi = 0, \; g(x) + g'(x)\xi \leq 0, \end{array} \tag{3.75}$$

with $u = (x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ playing the role of a parameter with the base value $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$. One can easily check that $\xi = 0$ is a stationary point of this problem for $u = \bar{u}$, and this point satisfies the SMFCQ and the SOSC for this problem with the associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$. From Theorem 1.40 it then follows that for each $u \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $\bar{u}$, problem (3.75) has a stationary point $\xi(u)$ and an associated Lagrange multiplier $(\tilde{\lambda}(u), \tilde{\mu}(u))$ such that $(\xi(u), \tilde{\lambda}(u), \tilde{\mu}(u)) \to (0, \bar{\lambda}, \bar{\mu})$ as $u \to \bar{u}$. It remains to observe that the KKT system

$$\begin{gathered} f'(x) + \frac{\partial^2 L}{\partial x^2}(x, \lambda, \mu)\xi + (h'(x))^{\mathrm{T}}\tilde{\lambda} + (g'(x))^{\mathrm{T}}\tilde{\mu} = 0, \\ h(x) + h'(x)\xi = 0, \quad \tilde{\mu} \geq 0, \quad g(x) + g'(x)\xi \leq 0, \quad \langle \tilde{\mu}, g(x) + g'(x)\xi\rangle = 0 \end{gathered}$$

of problem (3.75), to which $(\xi(u), \tilde{\lambda}(u), \tilde{\mu}(u))$ is a solution, can be written in the form of the generalized equation

$$\Phi(u) + \Phi'(u)v + N(u + v) \ni 0$$

with $v = (\xi, \tilde{\lambda} - \lambda, \tilde{\mu} - \mu)$, with $\Phi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ defined as

$$\Phi(u) = \left(\frac{\partial L}{\partial x}(x, \lambda, \mu), h(x), -g(x)\right),$$

and with

$$N(\cdot) = N_Q(\cdot), \quad Q = \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m_+.$$

Thus, $v(u) = (\xi(u), \tilde{\lambda}(u) - \lambda, \tilde{\mu}(u) - \mu)$ is a solution of this GE, $v(u) \to 0$ as $u \to \bar{u}$, and hemistability follows now by Definition 3.1. □

The following interpretation of the semismooth Newton method for KKT system is along the lines of the corresponding discussion above for NCP. Let $\Psi = \Psi_{\mathrm{NR}}$, and let $\{u^k\}$ (where $u^k = (x^k, \lambda^k, \mu^k)$) be an iterative sequence of the semismooth Newton method, convergent to the solution $(\bar{x}, \bar{\lambda}, \bar{\mu})$ of the KKT system (3.61). Then from (3.65) and from Proposition 3.26 it follows that for all k large enough there exists a partition (I_1^k, I_2^k) of $A_0(\bar{x}, \bar{\mu})$ (i.e., $I_1^k \cup I_2^k = A_0(\bar{x}, \bar{\mu})$, $I_1^k \cap I_2^k = \emptyset$) such that

$$\begin{aligned}\{i \in A_0(\bar{x},\, \bar{\mu}) \mid \mu_i^k > -g_i(x^k)\} \subset I_1^k,\\ \{i \in A_0(\bar{x},\, \bar{\mu}) \mid \mu_i^k < -g_i(x^k)\} \subset I_2^k,\end{aligned} \tag{3.76}$$

and

$$\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) + \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k)(x^{k+1} - x^k)$$
$$+(h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) = 0,$$
$$h(x^k) + h'(x^k)(x^{k+1} - x^k) = 0,$$
$$\min\left\{\mu^k_{A_+(\bar{x},\,\bar{\mu})\cup I_1^k},\, -g_{A_+(\bar{x},\,\bar{\mu})\cup I_1^k}(x^k)\right\} - g'_{A_+(\bar{x},\,\bar{\mu})\cup I_1^k}(x^k)(x^{k+1} - x^k) = 0,$$
$$\min\left\{\mu^k_{(\{1,\,\dots,\,m\}\setminus A(\bar{x}))\cup I_2^k},\, -g_{(\{1,\,\dots,\,m\}\setminus A(\bar{x}))\cup I_2^k}(x^k)\right\}$$
$$+(\mu^{k+1} - \mu^k)_{(\{1,\,\dots,\,m\}\setminus A(\bar{x}))\cup I_2^k} = 0.$$

Taking into account (3.76) and the definition of $A(\bar{x})$, $A_+(\bar{x},\, \bar{\mu})$, and $A_0(\bar{x},\, \bar{\mu})$, the last system is equivalent to

$$\begin{gathered}\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) + \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k)(x^{k+1} - x^k)\\ +(h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) = 0,\\ h(x^k) + h'(x^k)(x^{k+1} - x^k) = 0,\\ g_{A_+(\bar{x},\,\bar{\mu})\cup I_1^k}(x^k) + g'_{A_+(\bar{x},\,\bar{\mu})\cup I_1^k}(x^k)(x^{k+1} - x^k) = 0,\\ \mu^{k+1}_{(\{1,\,\dots,\,m\}\setminus A(\bar{x}))\cup I_2^k} = 0.\end{gathered} \tag{3.77}$$

Therefore, the iteration in question can be interpreted as that of the usual Newton method for the system of equations

$$\begin{gathered}\frac{\partial L}{\partial x}(x,\, \lambda,\, \mu) = 0, \quad h(x) = 0,\\ g_{A_+(\bar{x},\,\bar{\mu})\cup I_1^k}(x) = 0, \quad \mu_{(\{1,\,\dots,\,m\}\setminus A(\bar{x}))\cup I_2^k} = 0.\end{gathered} \tag{3.78}$$

If the strict complementarity condition $A_0(\bar{x},\, \bar{\mu}) = \emptyset$ holds, then the system (3.78) takes the form

$$\frac{\partial L}{\partial x}(x,\, \lambda,\, \mu) = 0, \quad h(x) = 0, \quad g_{A(\bar{x})}(x) = 0, \quad \mu_{(\{1,\,\dots,\,m\}\setminus A(\bar{x}))} = 0.$$

Therefore, in this case, the semismooth Newton method reduces to the usual Newton method for this system of equations or, in other words, to the Newton–Lagrange method for the equality-constrained problem

$$\begin{array}{ll}\text{minimize} & f(x)\\ \text{subject to} & h(x) = 0,\; g_{A(\bar{x})}(x) = 0.\end{array}$$

But without strict complementarity, the partitions $(I_1^k,\, I_2^k)$ may vary along the iterations, of course; and thus the Newton method is being applied to a changing system of equations.

Motivated by the primal-dual nature of the KKT system, the following discussion is in order. Generally, quadratic rate of convergence of the primal-dual sequence does not imply superlinear (or even linear!) rate of primal convergence, as demonstrated by the following two examples.

Example 3.38. Consider the sequence $\{(x^k,\, \lambda^k)\} \subset \mathbf{R}^2$ generated as follows. Take $x^0 \in \mathbf{R}$ and $\lambda^0 \in (0,\, 1)$, arbitrary. For each k, define $\lambda^{k+1} = (\lambda^k)^2$. For each k odd, define $x^{k+1} = x^k$; and for each k even, define $x^{k+1} = (\lambda^{k+1})^2$.

It can be seen that the sequence $\{(x^k,\, \lambda^k)\}$ converges to $(0,\, 0)$ quadratically. But, obviously, the rate of convergence of $\{x^k\}$ to 0 is not even linear.

In the example above, the sequence $\{\lambda^k\}$ converges superlinearly. In our next example, both sequences $\{x^k\}$ and $\{\lambda^k\}$ fail to possess superlinear convergence rate, despite quadratic convergence of $\{(x^k,\, \lambda^k)\}$. Moreover, the sequences of distances to the corresponding limits are both nonmonotone.

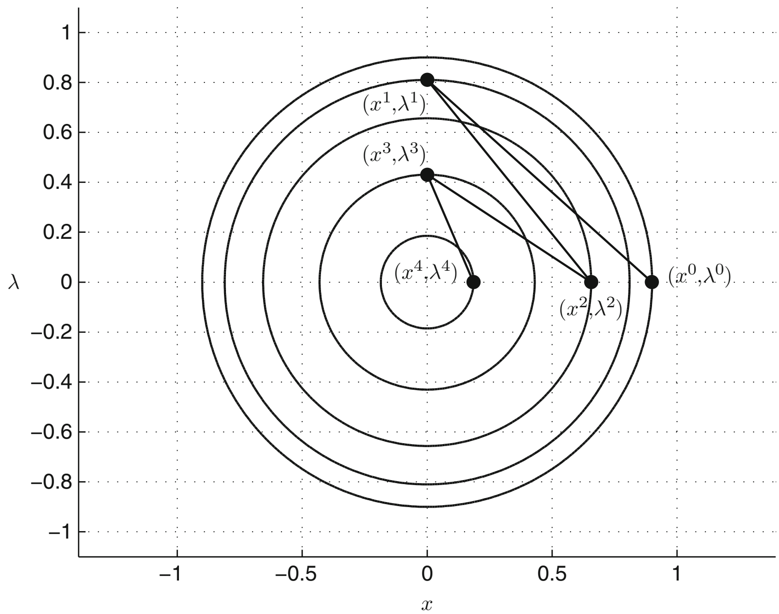

Fig. 3.3 Iterates in Example 3.39

Example 3.39. Fix any $\delta \in (0,\, 1)$ and consider the sequence $\{(x^k,\, \lambda^k)\} \subset \mathbf{R}^2$ generated as follows: $x^k = \delta^{2^k}$, $\lambda^k = 0$ for k even, and $x^k = 0$, $\lambda^k = \delta^{2^k}$ for k odd (see Fig. 3.3). Evidently, the sequence $\{(x^k,\, \lambda^k)\}$ converges to $(0,\, 0)$ quadratically, while x^{k+1} gets farther from 0 than x^k for k odd, and λ^{k+1} gets farther from 0 than λ^k for k even. The latter implies, in particular, that the rates of convergence of $\{x^k\}$ and $\{\lambda^k\}$ are not even linear.

It should be commented that fast primal convergence is often of special importance in practice, and must thus be studied separately (from primal-dual). We shall get back to this issue in Sects. 4.1.1 and 4.3.1 in the context of Newton-type methods for constrained optimization problems. Here, we restrict our presentation to one a posteriori result providing necessary and sufficient conditions for primal superlinear convergence rate of a quasi-Newton version of the semismooth Newton method for the natural residual-based reformulation of the KKT system (3.61).

Therefore, we set $\Psi = \Psi_{\mathrm{NR}}$, and following [224] we consider the method defined by the iteration system (3.65) with $u^k = (x^k,\, \lambda^k,\, \mu^k)$ and with the matrix J_k given by

$$J_k = \begin{pmatrix} H_k & (h'(x^k))^{\mathrm{T}} & (g'_{A_k}(x^k))^{\mathrm{T}} & (g'_{\{1,\,\ldots,\,m\}\setminus A_k}(x^k))^{\mathrm{T}} \\ h'(x^k) & 0 & 0 & 0 \\ -g'_{A_k}(x^k) & 0 & 0 & 0 \\ 0 & 0 & 0 & I \end{pmatrix}, \tag{3.79}$$

where $H_k \in \mathbf{R}^{n\times n}$ is a symmetric matrix, and

$$A_k = A_k(x^k,\, \mu^k) = \{i = 1,\, \ldots,\, m \mid \mu_i^k > -g_i(x^k)\}. \tag{3.80}$$

According to Proposition 3.26, if we take $H_k = \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k)$, then the matrix J_k defined according to (3.79)–(3.80) belongs to $\partial_B \Psi(u^k)$ (after re-ordering of the last m components of Ψ, if needed).

Theorem 3.40. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (3.69), and let $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be an associated Lagrange multiplier. Let $\{H_k\} \subset \mathbf{R}^{n\times n}$ be a sequence of symmetric matrices, let $\{u^k\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ (where $u^k = (x^k,\, \lambda^k,\, \mu^k)$) be convergent to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, and assume that u^{k+1} is a solution of the system (3.65) with $\Psi = \Psi_{\mathrm{NR}}$ and with J_k defined according to (3.79)–(3.80) for each k large enough.*

If the rate of convergence of $\{x^k\}$ is superlinear, then the following condition holds:

$$\pi_{C_0(\bar{x})}\left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k)\right)(x^{k+1} - x^k) = o(\|x^{k+1} - x^k\|) \tag{3.81}$$

as $k \to \infty$, where

$$C_0(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A(\bar{x})}(\bar{x})\xi = 0\}. \tag{3.82}$$

Conversely, if $\bar{x}$ and $(\bar{\lambda},\, \bar{\mu})$ satisfy the SSOSC (3.71), and the condition

$$\pi_{C_+(\bar{x},\, \bar{\mu})}\left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k)\right)(x^{k+1} - x^k) = o(\|x^{k+1} - x^k\|) \tag{3.83}$$

holds as $k \to \infty$, with $C_+(\bar{x},\, \bar{\mu})$ defined in (3.72), then the rate of convergence of $\{x^k\}$ is superlinear.

Proof. By (3.65) and (3.79), the relations

$$\begin{gathered}
\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) + H_k(x^{k+1} - x^k) \\
+(h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) = 0, \\
h(x^k) + h'(x^k)(x^{k+1} - x^k) = 0, \\
\min\{\mu^k_{A_k},\, -g_{A_k}(x^k)\} - g'_{A_k}(x^k)(x^{k+1} - x^k) = 0, \\
\min\left\{\mu^k_{\{1,\, \dots,\, m\}\setminus A_k},\, -g_{\{1,\, \dots,\, m\}\setminus A_k}(x^k)\right\} + (\mu^{k+1} - \mu^k)_{\{1,\, \dots,\, m\}\setminus A_k} = 0
\end{gathered} \tag{3.84}$$

hold for all k large enough. According to (3.80), the last two relations in (3.84) take the form

$$g_{A_k}(x^k) + g'_{A_k}(x^k)(x^{k+1} - x^k) = 0, \quad \mu^{k+1}_{\{1,\, \dots,\, m\}\setminus A_k} = 0, \tag{3.85}$$

and convergence of $\{(x^k,\, \mu^k)\}$ to $(\bar{x},\, \bar{\mu})$ implies the inclusions

$$A_+(\bar{x},\, \bar{\mu}) \subset A_k \subset A(\bar{x}) \tag{3.86}$$

for all k large enough.

Setting

$$\omega^k = \left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k)\right)(x^{k+1} - x^k), \tag{3.87}$$

by the first equality in (3.84) we conclude that

$$\begin{aligned}
\omega^k &= -\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) - \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k)(x^{k+1} - x^k) \\
&\quad -(h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) - (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) \\
&= -\frac{\partial L}{\partial x}(x^{k+1},\, \lambda^k,\, \mu^k) + o(\|x^{k+1} - x^k\|) \\
&\quad -(h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) - (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) \\
&= -\frac{\partial L}{\partial x}(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1}) + o(\|x^{k+1} - x^k\|) \\
&= -\frac{\partial L}{\partial x}(\bar{x},\, \lambda^{k+1},\, \mu^{k+1}) - \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \lambda^{k+1},\, \mu^{k+1})(x^{k+1} - \bar{x}) \\
&\quad +o(\|x^{k+1} - \bar{x}\| + \|x^{k+1} - x^k\|) \\
&= -\frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) - \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda}, \bar{\mu})(x^{k+1} - \bar{x}) \\
&\quad -(h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) - (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}) \\
&\quad +o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \\
&= -\frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda}, \bar{\mu})(x^{k+1} - \bar{x}) - (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) - (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}) \\
&\quad +o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)
\end{aligned} \tag{3.88}$$

as $k \to \infty$, where the convergence of $\{(\lambda^k,\, \mu^k)\}$ to $(\bar{\lambda},\, \bar{\mu})$ was taken into account.

From the second equality in (3.85) and from (3.86) we derive that for all k large enough

$$\mu^{k+1}_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})} = \bar{\mu}_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})} = 0.$$

Hence, for any $\xi \in C_0(\bar{x})$, from (3.82) it follows that

$$\begin{aligned}\langle (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}),\, \xi\rangle &= \langle \lambda^{k+1} - \bar{\lambda},\, h'(\bar{x})\xi\rangle \\ &\quad + \langle \mu^{k+1} - \bar{\mu},\, g'(\bar{x})\xi\rangle \\ &= 0,\end{aligned}$$

which further implies the equality

$$\pi_{C_0(\bar{x})}((h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu})) = 0 \tag{3.89}$$

(here, $C_0(\bar{x})$ is a linear subspace, and $\pi_{C_0(\bar{x})}$ is the orthogonal projector onto this subspace).

Assuming that the rate of convergence of $\{x^k\}$ is superlinear, and combining (3.87) and (3.88) with (3.89), we now obtain the estimate

$$\pi_{C_0(\bar{x})}\left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k)\right)(x^{k+1} - x^k) = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)$$

as $k \to \infty$. Taking into account Remark 2.5, this further implies the needed estimate (3.81).

Suppose now that $\bar{x}$ and $(\bar{\lambda},\, \bar{\mu})$ satisfy the SSOSC (3.71), and condition (3.83) holds.

From the second equality in (3.84) and from the mean-value theorem (see Theorem A.10, (a)) we derive that

$$\begin{aligned}\|h'(\bar{x})(x^{k+1} - \bar{x})\| &= \|h(x^k) + h'(x^k)(x^{k+1} - x^k) - h'(\bar{x})(x^{k+1} - \bar{x})\| \\ &\leq \|h(x^k) - h(\bar{x}) - h'(\bar{x})(x^k - \bar{x})\| \\ &\quad + \|h'(x^k) - h'(\bar{x})\|\|x^{k+1} - x^k\| \\ &= o(\|x^k - \bar{x}\|)\end{aligned} \tag{3.90}$$

as $k \to \infty$, where convergence of $\{x^k\}$ to $\bar{x}$ was again taken into account. Similarly, from the first equality in (3.85) we derive that

$$g'_{A_k}(\bar{x})(x^{k+1} - \bar{x}) = o(\|x^k - \bar{x}\|). \tag{3.91}$$

By Hoffman's error bound for linear systems (see Lemma A.4) and by (3.90)–(3.91), we now conclude that for each k large enough there exists $\xi^k \in \mathbf{R}^n$ such that

$$h'(\bar{x})\xi^k = 0, \quad g'_{A_k}(\bar{x})\xi^k = 0, \tag{3.92}$$

$$x^{k+1} - \bar{x} = \xi^k + o(\|x^k - \bar{x}\|). \tag{3.93}$$

(We emphasize that there is a finite number of possible realizations of A_k, and we simply take the minimum constant in Hoffman's error bound over all these realizations.) Relations (3.72) and (3.92) and the first inclusion in (3.86) imply that $\xi^k \in C_+(\bar{x}, \bar{\mu})$, and in particular

$$\langle x, \xi^k \rangle = \langle \pi_{C_+(\bar{x}, \bar{\mu})}(x), \xi^k \rangle \quad \forall\, x \in \mathbf{R}^n \tag{3.94}$$

(recall that $C_+(\bar{x}, \bar{\mu})$ is a linear subspace, and thus the orthogonal projector $\pi_{C_+(\bar{x}, \bar{\mu})}$ onto this subspace is a symmetric linear operator). Moreover, from (3.92), from the second equality in (3.85), and from the first inclusion in (3.86), we derive the equalities

$$\begin{aligned} \langle (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}), \xi^k \rangle &= \langle \lambda^{k+1} - \bar{\lambda}, h'(\bar{x})\xi^k \rangle \\ &\quad + \langle \mu^{k+1} - \bar{\mu}, g'(\bar{x})\xi^k \rangle \\ &= 0. \end{aligned} \tag{3.95}$$

Furthermore, the SSOSC (3.71) implies the existence of $\gamma > 0$ such that

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle \geq \gamma \|\xi\|^2 \quad \forall\, \xi \in C_+(\bar{x}, \bar{\mu}).$$

Then by (3.92), (3.93), employing also (3.87), (3.88), (3.94), and (3.95), for all k large enough we obtain that

$$\begin{aligned} \gamma \|\xi^k\|^2 &\leq \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi^k, \xi^k \right\rangle \\ &= \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})(x^{k+1} - \bar{x}), \xi^k \right\rangle + o(\|x^k - \bar{x}\| \|\xi^k\|) \\ &= -\langle \omega^k + (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}), \xi^k \rangle \\ &\quad + o((\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)\|\xi^k\|) \\ &= -\langle \pi_{C_+(\bar{x}, \bar{\mu})}(\omega^k), \xi^k \rangle + o((\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)\|\xi^k\|) \\ &= -\left\langle \pi_{C_+(\bar{x}, \bar{\mu})} \left(\left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k) \right) (x^{k+1} - x^k) \right), \xi^k \right\rangle \\ &\quad + o((\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)\|\xi^k\|) \\ &= o((\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)\|\xi^k\|) \end{aligned}$$

as $k \to \infty$, where the last equality is by (3.83). Dividing both sides of the relation above by $\|\xi^k\|$, we obtain the estimate

$$\xi^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|).$$

Combining this with (3.93), we conclude that

$$x^{k+1} - \bar{x} = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|).$$

By the same argument as in the proof of Proposition 2.4, the latter implies superlinear convergence rate. □

Note that Theorem 3.40 does not assume any kind of constraint qualification, and the multiplier $\bar{\lambda}$ need not be unique. This is the first place in the book where we analyze convergence properties of an algorithm in the case of possibly nonunique Lagrange multipliers. Such situations will be thoroughly investigated in Chap. 7, which is devoted entirely to the subject of problems with degenerate constraints (in which case multipliers are necessarily not unique).

3.3 Semismooth Josephy–Newton Method

Consider again the generalized equation (GE)

$$\Phi(x) + N(x) \ni 0, \tag{3.96}$$

where $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ is a single-valued mapping, and $N(\cdot)$ is a set-valued mapping from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$.

Unlike in Sect. 3.1 where Φ is smooth, here we assume that it is locally Lipschitz-continuous but may not be differentiable. Our development follows [158]. One of the motivations to consider this setting is provided by optimality systems for optimization problems whose objective function and constraints are differentiable with locally Lipschitz-continuous first derivatives, but not necessarily twice differentiable. Optimization problems with these smoothness properties arise in stochastic programming and optimal control (the so-called extended linear-quadratic problems [223, 237, 238]), in semi-infinite programming and in primal decomposition procedures (see [170, 221] and references therein). Once but not twice differentiable functions arise also when reformulating complementarity constraints as in [139] or in the lifting approach [156, 258]. Other possible sources are subproblems in penalty or augmented Lagrangian methods with lower-level constraints treated directly and upper-level constraints treated via quadratic penalization or via augmented Lagrangian, which gives certain terms that are not twice differentiable in general; see, e.g., [4].

Our development in this section relies on the following notion.

Definition 3.41. A solution $\bar{x} \in \mathbf{R}^n$ of the GE (3.96) is referred to as *strongly regular with respect to a set* $\Delta \subset \mathbf{R}^{n\times n}$ if for any $J \in \Delta$, the solution $\bar{x}$ of the GE

$$\Phi(\bar{x}) + J(x - \bar{x}) + N(x) \ni 0 \tag{3.97}$$

is strongly regular. If $\Delta = \partial_B \Phi(\bar{x})$ ($\Delta = \partial \Phi(\bar{x})$), then $\bar{x}$ is referred to as a *BD-regular* (*CD-regular*) solution of the GE (3.96).

Therefore, strong regularity of $\bar{x}$ with respect to Δ means that for any $J \in \Delta$ and for any $r \in \mathbf{R}^n$ close enough to 0, the perturbed (partially) linearized GE

$$\Phi(\bar{x}) + J(x - \bar{x}) + N(x) \ni r$$

has near $\bar{x}$ the unique solution $x_J(r)$, and the mapping $x_J(\cdot)$ is locally Lipschitz-continuous at $\bar{x}$.

Evidently, Definition 3.41 generalizes the following two notions that appeared earlier in this book: the notion of strong regularity for the case of a GE with a smooth base mapping (see Definition 1.23), and the notion of BD-regularity (CD-regularity) of a mapping Φ at a solution $\bar{x}$ of the (ordinary) equation

$$\Phi(x) = 0,$$

corresponding to $N(\cdot) \equiv \{0\}$ and $\Delta = \partial_B \Phi(\bar{x})$ ($\Delta = \partial \Phi(\bar{x})$), as defined in Remark 1.65.

The goal of this section is to develop the semismooth Josephy–Newton method for finding strongly regular solutions of the GE (3.96). On the one hand, this method extends the Josephy–Newton method of Sect. 3.1 to the case of a possibly nonsmooth base mapping, while on the other hand, it extends the semismooth Newton method for nonsmooth equations, analyzed in Sect. 2.4, to the case of a GE. We shall provide the related local superlinear convergence theory covering the perturbed versions of the method. In Chap. 4, we shall consider its specific applications to some optimization algorithms.

The following result is an immediate corollary of Theorem 1.25.

Proposition 3.42. *For the given $\Phi : \mathbf{R}^n \to \mathbf{R}^n$, $J \in \mathbf{R}^{n \times n}$ and multifunction N from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$, let $\bar{x}$ be a strongly regular solution of the GE (3.97).*

Then for any fixed neighborhood W of $\bar{x}$ and any sufficiently small $\ell \geq 0$, there exist $\bar{\ell} > 0$ and neighborhoods U of $\bar{x}$ and V of 0 such that for any mapping $R : \mathbf{R}^n \to \mathbf{R}^n$ which is Lipschitz-continuous on W with Lipschitz constant ℓ, and for any $r \in R(\bar{x}) + V$, the GE

$$R(x) + \Phi(\bar{x}) + J(x - \bar{x}) + N(x) \ni r$$

has in U the unique solution $x(r)$, and the mapping $x(\cdot)$ is Lipschitz-continuous on $R(\bar{x}) + V$ with Lipschitz constant $\bar{\ell}$.

We next use Proposition 3.42 to prove solvability of perturbed linearized GEs for all points close enough to a strongly regular solution and all matrices J close enough to the associated set Δ.

Proposition 3.43. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be continuous at $\bar{x} \in \mathbf{R}^n$. For a given multifunction N from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$, let $\bar{x}$ be a solution of the GE (3.96), strongly regular with respect to a compact set $\Delta \subset \mathbf{R}^{n \times n}$.*

Then there exist $\varepsilon > 0$, $\bar{\ell} > 0$ *and neighborhoods* $\tilde{U}$ *and* U *of* $\bar{x}$ *and* V *of* 0 *such that for any* $\tilde{x} \in \tilde{U}$, *any* $J \in \mathbf{R}^{n\times n}$ *satisfying*

$$\mathrm{dist}(J,\, \Delta) < \varepsilon, \tag{3.98}$$

and any $r \in V$, *the GE*

$$\Phi(\tilde{x}) + J(x - \tilde{x}) + N(x) \ni r \tag{3.99}$$

has in U *the unique solution* $x(r)$, *and the mapping* $x(\cdot)$ *is Lipschitz-continuous on* V *with Lipschitz constant* $\bar{\ell}$.

Proof. Fix any $\bar{J} \in \Delta$. For each $\tilde{x} \in \mathbf{R}^n$ and $J \in \mathbf{R}^{n\times n}$ define the mapping $R : \mathbf{R}^n \to \mathbf{R}^n$,

$$R(x) = \Phi(\tilde{x}) - \Phi(\bar{x}) - J\tilde{x} + \bar{J}\bar{x} + (J - \bar{J})x. \tag{3.100}$$

For any pre-fixed $\ell > 0$, the mapping R is Lipschitz-continuous on $\mathbf{R}^n$ with Lipschitz constant ℓ, provided J is close enough to $\bar{J}$. Note also that the element $R(\bar{x}) = \Phi(\tilde{x}) - \Phi(\bar{x}) - J(\tilde{x} - \bar{x})$ tends to 0 as $\tilde{x} \to \bar{x}$. Therefore, by Proposition 3.42 applied with $W = \mathbf{R}^n$, there exist $\varepsilon > 0$, $\bar{\ell} > 0$ and neighborhoods $\tilde{U}$ and U of $\bar{x}$ and V of 0 such that for any $\tilde{x} \in \tilde{U}$ and $J \in \mathbf{R}^{n\times n}$ such that $\|J - \bar{J}\| < \varepsilon$, and for any $r \in V$, the GE

$$R(x) + \Phi(\bar{x}) + \bar{J}(x - \bar{x}) + N(x) \ni r \tag{3.101}$$

has in U the unique solution $x(r)$, and the mapping $x(\cdot)$ is Lipschitz-continuous on V with Lipschitz constant $\bar{\ell}$. Substituting (3.100) into (3.101), observe that the latter coincides with (3.99).

Considering for each $\bar{J} \in \Delta$ the open ball in $\mathbf{R}^{n\times n}$ centered at $\bar{J}$ and of radius ε defined above, we obtain the open cover of the compact set Δ which has a finite subcover. We now re-define $\varepsilon > 0$ in such a way that any $J \in \mathbf{R}^{n\times n}$ satisfying (3.98) belongs to the specified finite subcover. Furthermore, we take the maximum value $\bar{\ell} > 0$ of the corresponding constants and the intersections $\tilde{U}$, U and V of the corresponding neighborhoods defined above over the centers $\bar{J}$ of the balls constituting this subcover. By further shrinking V, if necessary, in order to ensure that for any $\tilde{x} \in \tilde{U}$, any $J \in \mathbf{R}^{n\times n}$ satisfying (3.98), and any $r \in V$, the solution $x(r)$ of (3.99), corresponding to an appropriate element of the subcover, belongs to U, we obtain all the ingredients for the stated assertion. $\square$

Remark 3.44. Assuming local Lipschitz-continuity of Φ at $\bar{x}$, and taking into account Proposition 1.51 (giving compactness of $\partial\Phi(\bar{x})$), we obtain from Proposition 3.43 that the CD-regularity of a solution $\bar{x}$ of the GE (3.96) implies the following: there exist neighborhoods O of $\bar{x}$ and W of 0, and $\ell > 0$, such that for all $J \in \partial\Phi(\bar{x})$ and all $r \in W$ there exists the unique $x_J(r) \in O$ satisfying the GE

$$\Phi(\bar{x}) + J(x - \bar{x}) + N(x) \ni r, \tag{3.102}$$

and the mapping $x_J(\cdot) : W \to O$ is Lipschitz-continuous on W with the constant ℓ. Observe that by necessity, it holds that

$$x_J(0) = \bar{x} \quad \forall\, J \in \partial\Phi(\bar{x}). \tag{3.103}$$

By the *semismooth Josephy–Newton method* for the GE (3.96) we mean the following iterative process: for the current iterate $x^k \in \mathbf{R}^n$, the next iterate is computed as a solution of the GE

$$\Phi(x^k) + J_k(x - x^k) + N(x) \ni 0 \tag{3.104}$$

with some $J_k \in \partial\Phi(x^k)$. Moreover, to enlarge the domain of possible applications, we shall deal the *perturbed semismooth Josephy–Newton method*. Specifically, instead of (3.104), the next iterate x^{k+1} satisfies the GE

$$\omega^k + \Phi(x^k) + J_k(x - x^k) + N(x) \ni 0 \tag{3.105}$$

with some $J_k \in \partial\Phi(x^k)$, where $\omega^k \in \mathbf{R}^n$ is a perturbation term.

We start with the following a posteriori result concerned with superlinear rate of convergence, for now assuming convergence itself.

Proposition 3.45. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be semismooth at $\bar{x} \in \mathbf{R}^n$. Let $\bar{x}$ be a solution of the GE (3.96), strongly regular with respect to some closed set $\bar{\Delta} \subset \partial\Phi(\bar{x})$. Let a sequence $\{x^k\} \subset \mathbf{R}^n$ be convergent to $\bar{x}$, and assume that x^{k+1} satisfies (3.105) for each $k = 0, 1, \ldots$, with some $J_k \in \partial\Phi(x^k)$ and $\omega^k \in \mathbf{R}^n$ such that*

$$\mathrm{dist}(J_k, \bar{\Delta}) \to 0 \tag{3.106}$$

and

$$\omega^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \tag{3.107}$$

as $k \to \infty$.

Then the rate of convergence of $\{x^k\}$ is superlinear. Moreover, the rate of convergence is quadratic provided Φ is strongly semismooth at $\bar{x}$ and

$$\omega^k = O(\|x^{k+1} - x^k\|^2 + \|x^k - \bar{x}\|^2) \tag{3.108}$$

as $k \to \infty$.

Proof. Define $\varepsilon > 0$, $\bar{\ell} > 0$, $\tilde{U}$, U and V according to Proposition 3.43 with $\Delta = \bar{\Delta}$. Then for any $x^k \in \tilde{U}$, any $J_k \in \partial\Phi(x^k)$ satisfying $\mathrm{dist}(J_k, \bar{\Delta}) < \varepsilon$, and any $r \in V$, the GE

$$\Phi(x^k) + J_k(x - x^k) + N(x) \ni r \tag{3.109}$$

has in U the unique solution $x(r)$, and $x(\cdot)$ is Lipschitz-continuous on V with Lipschitz constant $\bar{\ell}$.

For each k set

$$r^k = \Phi(x^k) - \Phi(\bar{x}) - J_k(x^k - \bar{x}). \tag{3.110}$$

Note that by the semismoothness of Φ at $\bar{x}$, it holds that

$$r^k = o(\|x^k - \bar{x}\|) \tag{3.111}$$

as $k \to \infty$. Note also that by (3.110),

$$0 \in \Phi(\bar{x}) + N(\bar{x}) = \Phi(x^k) + J_k(\bar{x} - x^k) + N(\bar{x}) - r^k. \tag{3.112}$$

By the convergence of $\{x^k\}$ to $\bar{x}$, and by (3.106), (3.107), and (3.111), we derive the following: for all k large enough it holds that x^k, $x^{k+1} \in \tilde{U}$, $\mathrm{dist}(J_k, \bar{\Delta}) < \varepsilon$, $-\omega^k \in V$ and $r^k \in V$. Hence, according to Proposition 3.43, x^{k+1} is the unique solution in U of the GE (3.109) with $r = -\omega^k$, i.e., $x^{k+1} = x(-\omega^k)$, while by (3.112), $\bar{x}$ is the unique solution in U of the GE (3.109) with $r = r^k$, i.e., $x^{k+1} = x(r^k)$. Therefore,

$$\begin{aligned}\|x^{k+1} - \bar{x}\| &= \|x(-\omega^k) - x(r^k)\| \\ &\leq \bar{\ell}\|\omega^k + r^k\| \\ &= o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)\end{aligned} \tag{3.113}$$

as $k \to \infty$, where the last estimate is by (3.107) and (3.111). By the same argument as in the proof of Proposition 2.4, the latter implies the superlinear convergence rate.

Finally, if Φ is strongly semismooth at $\bar{x}$, then for r^k defined in (3.110) it holds that

$$r^k = O(\|x^k - \bar{x}\|^2)$$

as $k \to \infty$. Combining this with (3.108), and with the inequality in (3.113), we derive the quadratic convergence rate. □

As an immediate application of Proposition 3.45, consider the following *semismooth quasi-Josephy–Newton method*. Let $\{J_k\} \subset \mathbf{R}^{n \times n}$ be a sequence of matrices. For the current iterate $x^k \in \mathbf{R}^n$, let the next iterate x^{k+1} be computed as a solution of the GE (3.104), and assume that $\{J_k\}$ satisfies the Dennis–Moré-type condition (cf. (2.151)):

$$\min_{J \in \partial\Phi(x^k)} \|(J_k - J)(x^{k+1} - x^k)\| = o(\|x^{k+1} - x^k\|) \tag{3.114}$$

as $k \to \infty$.

In the following a posteriori result, we consider the possibility of somewhat more special choices of matrices J_k.

Theorem 3.46. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be semismooth at $\bar{x} \in \mathbf{R}^n$. Let $\bar{x}$ be a solution of the GE (3.96), strongly regular with respect to some closed set $\bar{\Delta} \subset \partial\Phi(\bar{x})$. Let $\{J_k\} \subset \mathbf{R}^{n \times n}$ be a sequence of matrices, and let a sequence $\{x^k\} \subset \mathbf{R}^n$ be convergent to $\bar{x}$ and such that for all k large enough x^{k+1} satisfies (3.104) and there exists $\tilde{J}_k \in \partial\Phi(x^k)$ satisfying*

$$\mathrm{dist}(\tilde{J}_k, \bar{\Delta}) \to 0 \tag{3.115}$$

and

$$(J_k - \tilde{J}_k)(x^{k+1} - x^k) = o(\|x^{k+1} - x^k\|) \tag{3.116}$$

as $k \to \infty$.

Then the rate of convergence of $\{x^k\}$ *is superlinear.*

Proof. For each k set

$$\omega^k = (J_k - \tilde{J}_k)(x^{k+1} - x^k).$$

Then (3.116) implies (3.107). Employing (3.115), the result now follows immediately from Proposition 3.45. □

Theorem 3.46 generalizes Theorem 2.48. Indeed, the assumption that for each k large enough there exists $\tilde{J}_k \in \partial\Phi(x^k)$ satisfying (3.116) is equivalent to (3.114), which in turn coincides with (2.151). If $\bar{x}$ is a CD-regular solution of the GE (3.96), then Theorem 3.46 is applicable with $\bar{\Delta} = \partial\Phi(\bar{x})$ and, in this case, (3.115) is automatic for any choice of $\tilde{J}_k \in \partial\Phi(x^k)$ according to Proposition 1.51. Note, however, that unlike in Remark 2.49, replacing "min" in (3.114) by "max" would generally lead to a more restrictive assumption, because at this level of generality condition (3.114) cannot be claimed to be necessary for the superlinear convergence rate.

Another appealing possibility is to apply Theorem 3.46 with $\bar{\Delta} = \partial_B\Phi(\bar{x})$, assuming BD-regularity of the solution $\bar{x}$. This corresponds to the development in Sect. 3.2.

We proceed with a priori local convergence analysis.

Theorem 3.47. *Let* $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ *be semismooth at* $\bar{x} \in \mathbf{R}^n$. *Let* $\bar{x}$ *be a solution of the GE* (3.96), *strongly regular with respect to some closed set* $\bar{\Delta} \subset \partial\Phi(\bar{x})$. *Let* Δ *be a multifunction from* $\mathbf{R}^n$ *to the subsets of* $\mathbf{R}^{n\times n}$, *such that*

$$\Delta(x) \subset \partial\Phi(x) \quad \forall\, x \in \mathbf{R}^n, \tag{3.117}$$

and for any $\varepsilon > 0$ *there exists a neighborhood* O *of* $\bar{x}$ *such that*

$$\operatorname{dist}(J,\, \bar{\Delta}) < \varepsilon \quad \forall\, J \in \Delta(x),\ \forall\, x \in O. \tag{3.118}$$

Then there exists $\delta > 0$ *such that for any starting point* $x^0 \in \mathbf{R}^n$ *close enough to* $\bar{x}$, *for each* $k = 0, 1, \ldots$ *and any choice of* $J_k \in \Delta(x^k)$, *there exists the unique solution* x^{k+1} *of the GE* (3.104) *satisfying*

$$\|x^{k+1} - x^k\| \le \delta; \tag{3.119}$$

the sequence $\{x^k\}$ *generated this way converges to* $\bar{x}$, *and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided* Φ *is strongly semismooth at* $\bar{x}$.

Proof. Define $\varepsilon > 0$, $\bar{\ell} > 0$, $\tilde{U}$, U and V according to Proposition 3.43 with $\Delta = \bar{\Delta}$. Moreover, let $\tilde{U}$ be such that (3.118) holds with $O = \tilde{U}$ and with the specified ε.

Then, according to Proposition 3.43, for any $x^k \in \tilde{U}$, any $J_k \in \Delta(x^k)$ and any $r \in V$ the GE

$$\Phi(x^k) + J_k(x - x^k) + N(x) \ni r \tag{3.120}$$

has in U the unique solution $x(r)$, and $x(\cdot)$ is Lipschitz-continuous on V with Lipschitz constant $\bar{\ell}$. In particular, the GE (3.104) has in U the unique solution $x^{k+1} = x(0)$.

Defining r^k according to (3.110) and employing (3.117), by the semismoothness of Φ at $\bar{x}$ we conclude that (3.111) holds, and

$$0 \in \Phi(\bar{x}) + N(\bar{x}) = \Phi(x^k) + J_k(\bar{x} - x^k) + N(\bar{x}) - r^k.$$

Shrinking $\tilde{U}$ if necessary, by (3.111) we conclude that $r^k \in V$ if $x^k \in \tilde{U}$, and hence, $\bar{x}$ is the unique solution of the GE (3.120) with $r = r^k$, i.e., $\bar{x} = x(r^k)$. Therefore,

$$\|x^{k+1} - \bar{x}\| \leq \|x(r^k) - x(0)\| \leq \bar{\ell}\|r^k\| = o(\|x^k - \bar{x}\|) \tag{3.121}$$

as $x^k \to \bar{x}$, where the last estimate is by (3.111).

From (3.121) we derive the following: for any $q \in (0, 1)$, there exists $\delta > 0$ such that $B(\bar{x}, \delta/2) \subset \tilde{U}$, $B(\bar{u}, 3\delta/2) \subset U$, and for any $x^k \in B(\bar{x}, \delta/2)$ it holds that

$$\|x^{k+1} - \bar{x}\| \leq q\|x^k - \bar{x}\|, \tag{3.122}$$

implying that $x^{k+1} \in B(\bar{x}, \delta/2)$. Then

$$\|x^{k+1} - x^k\| \leq \|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\| < \frac{\delta}{2} + \frac{\delta}{2} = \delta,$$

and hence, x^{k+1} is a solution of the GE (3.104) satisfying (3.119). Moreover, for any point $x^{k+1} \in \mathbf{R}^n$ satisfying (3.119), it holds that

$$\|x^{k+1} - \bar{x}\| \leq \|x^{k+1} - x^k\| + \|x^k - \bar{x}\| \leq \delta + \frac{\delta}{2} = \frac{3\delta}{2}.$$

Hence, $x^{k+1} \in U$, implying that x^{k+1} is a solution of the GE (3.104) if and only if it coincides with $x^{k+1} = x(0)$. Thus, the latter is the unique solution of the GE (3.104) satisfying (3.119).

Therefore, the inclusion $x^0 \in B(\bar{x}, \delta/2)$ implies that the entire sequence $\{x^k\}$ is uniquely defined (for any choice of $J_k \in \Delta(x^k)$ for all k) and is contained in $B(\bar{x}, \delta/2)$, and (3.122) shows convergence of this sequence to $\bar{x}$. The convergence rate estimates follow from Proposition 3.45. $\square$

Remark 3.48. Theorem 3.47 generalizes Theorem 2.42. Indeed, the two basic options for $\Delta(\cdot)$ are $\partial_B\Phi(\cdot)$ and $\partial\Phi(\cdot)$. Taking into account Proposition 1.51, for the first option Theorem 3.47 is applicable with $\bar{\Delta} = \partial_B\Phi(\bar{x})$ assuming BD-regularity of $\bar{x}$, while for the second option it is applicable with the set

$\bar{\Delta} = \partial\Phi(\bar{x})$ assuming CD-regularity of $\bar{x}$. However, other choices (e.g., related to the specific problem structure) are also possible.

Recall that for generalized equations with smooth bases, a subtler local convergence result was established in Theorem 3.2, where the assumptions are semistability and hemistability of the solution in question, the combination of which is generally weaker than strong regularity. Note, however, that unlike in Theorem 3.47, the uniqueness of the subproblems' solutions could not be established under these assumptions.

In the case of the variational inequality (VI)

$$x \in Q, \quad \langle \Phi(x),\, y - x\rangle \geq 0 \quad \forall\, y \in Q \tag{3.123}$$

with some closed convex set $Q \subset \mathbf{R}^n$ (i.e., when $N(\cdot) = N_Q(\cdot)$), the iteration subproblem (3.104) of the semismooth Josephy–Newton method takes the form of the VI

$$x \in Q, \quad \langle \Phi(x^k) + J_k(x - x^k),\, y - x\rangle \geq 0 \quad \forall\, y \in Q \tag{3.124}$$

with some $J_k \in \partial\Phi(x^k)$. In particular, for the NCP

$$x \geq 0, \quad \Phi(x) \geq 0, \quad \langle x,\, \Phi(x)\rangle = 0, \tag{3.125}$$

corresponding to the VI (3.123) with $Q = \mathbf{R}^n_+$, the subproblem (3.124) becomes the linear complementarity problem

$$x \geq 0, \quad \Phi(x^k) + J_k(x - x^k) \geq 0, \quad \langle x,\, \Phi(x^k) + J_k(x - x^k)\rangle = 0.$$

According to Definition 3.41, strong regularity of a solution $\bar{x}$ of the NCP (3.125) with respect to a set $\bar{\Delta} \subset \mathbf{R}^{n\times n}$ means that for any $J \in \bar{\Delta}$ the point $\bar{x}$ is a strongly regular solution of the linear complementarity problem

$$x \geq 0, \quad \Phi(\bar{x}) + J(x - \bar{x}) \geq 0, \quad \langle x,\, \Phi(\bar{x}) + J(x - \bar{x})\rangle = 0.$$

The algebraic characterization of the latter property readily follows from Proposition 1.26.

To give some further insight into the roles of the notions introduced in Definition 3.41, we complete this section with some stability results for solutions of the GE (3.96), obtained in [131]. These results unify and extend two classical facts of variational analysis: Theorem 1.24 (which is the implicit function theorem for GEs with smooth base mappings), and Theorems 1.66 and 1.67 (which are the inverse and implicit function theorems, respectively, for usual equations with Lipschitzian mappings).

Theorem 3.49. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be locally Lipschitz-continuous at $\bar{x} \in \mathbf{R}^n$, and let N be a multifunction from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$. Assume that $\bar{x}$ is a CD-regular solution of the GE (3.96).*

Then there exist neighborhoods U of $\bar{x}$ and V of 0 such that for all $y \in V$ there exists the unique $x(y) \in U$ satisfying the perturbed GE

$$\Phi(x) + N(x) \ni y, \tag{3.126}$$

and the mapping $x(\cdot)$ is locally Lipschitz-continuous at 0.

Proof. For any $J \in \partial\Phi(\bar{x})$ and any $y \in \mathbf{R}^n$, the GE (3.126) is equivalent to (3.102) with $r = r_J(x, y)$, where the continuous mapping $r_J : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$ is defined by

$$r_J(x, y) = -(\Phi(x) - \Phi(\bar{x}) - J(x - \bar{x})) + y. \tag{3.127}$$

Let O, W, $\ell > 0$ and the family of mappings $x_J(\cdot) : W \to O$, $J \in \partial\Phi(\bar{x})$, be defined according to Remark 3.44. Since $\partial\Phi(\bar{x})$ is compact, there exist $\delta > 0$ and $\rho > 0$ such that for any $x \in B(\bar{x}, \delta) \subset O$ and any $y \in B(0, \rho)$ it holds that $r_J(x, y) \in W$ for all $J \in \partial\Phi(\bar{x})$. Therefore, for such x, y, and J, the relation (3.126) is further equivalent to the equality

$$x = x_J(r_J(x, y)). \tag{3.128}$$

Fix any $\varepsilon \in (0, 1/(3\ell)]$, and define the function $\omega : \mathbf{R}^n \to \mathbf{R}_+$,

$$\omega(x) = \min_{J \in \partial\Phi(\bar{x})} \|\Phi(x) - \Phi(\bar{x}) - J(x - \bar{x})\|. \tag{3.129}$$

Then from Corollary 1.57 it follows that by further reducing $\delta > 0$, if necessary, we can ensure that

$$\omega(x) \leq \varepsilon\delta \quad \forall\, x \in B(\bar{x}, \delta). \tag{3.130}$$

Now for each $x \in B(\bar{x}, \delta)$ we select the specific $J = J_x \in \partial\Phi(\bar{x})$ as follows. Consider the parametric optimization problem

$$\begin{array}{ll} \text{minimize} & \|\Phi(x) - \Phi(\bar{x}) - J(x - \bar{x})\| + \alpha\|J\|_*^2 \\ \text{subject to} & J \in \partial\Phi(\bar{x}), \end{array} \tag{3.131}$$

where $x \in \mathbf{R}^n$ and $\alpha > 0$ are parameters, and $\|\cdot\|_*$ is any norm defined by an inner product in $\mathbf{R}^{n\times n}$ (e.g., the Frobenius norm). Let $v : \mathbf{R}^n \times \mathbf{R}_+ \to \mathbf{R}_+$ be the optimal-value function of this problem:

$$v(x, \alpha) = \min_{J \in \partial\Phi(\bar{x})} (\|\Phi(x) - \Phi(\bar{x}) - J(x - \bar{x})\| + \alpha\|J\|_*^2). \tag{3.132}$$

Then, according to (3.129), $v(x, 0) = \omega(x)$ for all $x \in \mathbf{R}^n$, and since $\partial\Phi(\bar{x})$ is compact, for any fixed $\bar{\alpha} > 0$ the function v is continuous on the compact set $B(\bar{x}, \delta) \times [0, \bar{\alpha}]$. Since a continuous function on a compact set is uniformly continuous, this further implies the existence of $\alpha > 0$ such that

$$v(x, \alpha) \leq \omega(x) + \varepsilon\delta \quad \forall\, x \in B(\bar{x}, \delta). \tag{3.133}$$

Furthermore, with this positive α fixed, the objective function of problem (3.131) is strongly convex, and therefore, this problem with a convex feasible set has the unique solution J_x for every fixed $x \in \mathbf{R}^n$. This implies that the mapping $x \to J_x : B(\bar{x}, \delta) \to \partial\Phi(\bar{x})$ is continuous. Moreover, according to (3.130), (3.132), and (3.133),

$$\begin{aligned} \|\Phi(x) - \Phi(\bar{x}) - J_x(x - \bar{x})\| &\le v(x, \alpha) \\ &\le \omega(x) + \varepsilon\delta \\ &\le 2\varepsilon\delta \quad \forall\, x \in B(\bar{x}, \delta). \end{aligned} \tag{3.134}$$

For any $y \in B(0, \rho)$, define the mapping $\chi_y : B(\bar{x}, \delta) \to \mathbf{R}^n$,

$$\chi_y(x) = x_{J_x}(r_{J_x}(x, y)). \tag{3.135}$$

By further reducing $\delta > 0$ if necessary, so that $\varepsilon\delta \le \rho$, for any $y \in B(0, \varepsilon\delta)$, from (3.103), (3.127), and (3.134) we derive

$$\begin{aligned} \|\chi_y(x) - \bar{x}\| &= \|x_{J_x}(r_{J_x}(x, y)) - x_{J_x}(0)\| \\ &\le \ell\|r_{J_x}(x, y)\| \\ &\le \ell(\|\Phi(x) - \Phi(\bar{x}) - J_x(x - \bar{x})\| + \|y\|) \\ &\le \ell(2\varepsilon\delta + \varepsilon\delta) \\ &= 3\ell\varepsilon\delta \\ &\le \delta \quad \forall\, x \in B(\bar{x}, \delta), \end{aligned}$$

where the last inequality holds because $\varepsilon \le 1/(3\ell)$. Therefore, χ_y continuously maps $B(\bar{x}, \delta)$ into itself. Hence, by Brouwer's fixed-point theorem (see, e.g., [68, Theorem 2.1.18]), there exists $x(y) \in B(\bar{x}, \delta)$ such that

$$x(y) = \chi_y(x(y)).$$

According to (3.135), this means that for any $y \in B(0, \varepsilon\delta) \subset B(0, \rho)$ the point $x(y) \in B(\bar{x}, \delta)$ satisfies (3.128) with $J = J_{x(y)}$ and, as discussed above, this is equivalent to saying that $x(y)$ solves the GE (3.126).

We thus proved that for any $y \in B(0, \varepsilon\delta)$ the GE (3.126) has a solution $x(y) \in B(\bar{x}, \delta)$. It remains to show that this solution is unique, and the mapping $x(\cdot)$ is Lipschitz-continuous on $B(0, \varepsilon\delta)$, provided $\delta > 0$ is small enough. Then setting $U = B(\bar{x}, \delta)$ and $V = B(0, \varepsilon\delta)$, we shall obtain the needed conclusion.

We first show uniqueness. Suppose that there exist sequences $\{x^{1,k}\} \subset \mathbf{R}^n$, $\{x^{2,k}\} \subset \mathbf{R}^n$ and $\{y^k\} \subset \mathbf{R}^n$ such that both $\{x^{1,k}\}$ and $\{x^{2,k}\}$ converge to $\bar{x}$, $\{y^k\}$ converges to 0, and for every k it holds that $x^{1,k} \ne x^{2,k}$, and the points $x^{1,k}$ and $x^{2,k}$ solve (3.126) with $y = y^k$.

According to Corollary 1.57, for every k we can select $J_k \in \partial\Phi(\bar{x})$ such that

$$\|\Phi(x^{1,k}) - \Phi(x^{2,k}) - J_k(x^{1,k} - x^{2,k})\| = o(\|x^{1,k} - x^{2,k}\|) \tag{3.136}$$

as $k \to \infty$. Since $\partial\Phi(\bar{x})$ is compact, without loss of generality we can assume that $\{J_k\}$ converges to some $J \in \partial\Phi(\bar{x})$, and then (3.136) implies the estimate

$$\|\Phi(x^{1,k}) - \Phi(x^{2,k}) - J(x^{1,k} - x^{2,k})\| = o(\|x^{1,k} - x^{2,k}\|). \tag{3.137}$$

Since $x^{1,k} \in B(\bar{x}, \delta)$, $x^{2,k} \in B(\bar{x}, \delta)$ and $y^k \in B(0, \rho)$ for all k large enough, we have that for such k both points $x^{1,k}$ and $x^{2,k}$ satisfy (3.128) with $y = y^k$. Employing (3.127) and (3.137), we then have that

$$\begin{aligned}\|x^{1,k} - x^{2,k}\| &= \|x_J(r_J(x^{1,k}, y^k)) - x_J(r_J(x^{2,k}, y^k))\| \\ &\le \ell\|r_J(x^{1,k}, y^k) - r_J(x^{2,k}, y^k)\| \\ &\le \ell\|\Phi(x^{1,k}) - \Phi(x^{2,k}) - J(x^{1,k} - x^{2,k})\| \\ &= o(\|x^{1,k} - x^{2,k}\|)\end{aligned}$$

as $k \to \infty$, giving a contradiction.

We proceed with demonstrating Lipschitz-continuity of the solution mapping. Suppose that there exist sequences $\{x^{1,k}\} \subset \mathbf{R}^n$, $\{x^{2,k}\} \subset \mathbf{R}^n$, and $\{y^{1,k}\} \subset \mathbf{R}^n$, $\{y^{2,k}\} \subset \mathbf{R}^n$, such that both $\{x^{1,k}\}$ and $\{x^{2,k}\}$ converge to $\bar{x}$, both $\{y^{1,k}\}$ and $\{y^{2,k}\}$ converge to 0, and for every k it holds that $y^{1,k} \ne y^{2,k}$, the point $x^{i,k}$ solves (3.126) with $y = y^{i,k}$ for $i = 1, 2$, and

$$\frac{\|x^{1,k} - x^{2,k}\|}{\|y^{1,k} - y^{2,k}\|} \to \infty \text{ as } k \to \infty. \tag{3.138}$$

Repeating the argument used to establish uniqueness, we then obtain the estimate

$$\begin{aligned}\|x^{1,k} - x^{2,k}\| &= \|x_J(r_J(x^{1,k}, y^{1,k})) - x_J(r_J(x^{2,k}, y^{2,k}))\| \\ &\le \ell\|r_J(x^{1,k}, y^{1,k}) - r_J(x^{2,k}, y^{2,k})\| \\ &\le \ell(\|\Phi(x^{1,k}) - \Phi(x^{2,k}) - J(x^{1,k} - x^{1,k})\| + \|y^{1,k} - y^{2,k}\|) \\ &= \ell\|y^{1,k} - y^{2,k}\| + o(\|x^{1,k} - x^{2,k}\|)\end{aligned}$$

as $k \to \infty$, giving a contradiction with (3.138). □

Theorem 3.50. *Let $\Phi : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^n$ be locally Lipschitz-continuous at $(\bar{\sigma}, \bar{x}) \in \mathbf{R}^s \times \mathbf{R}^n$, and let N be a multifunction from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$. Assume that $\bar{x}$ is a solution of the GE*

$$\Phi(\sigma, u) + N(u) \ni 0 \tag{3.139}$$

for $\sigma = \bar{\sigma}$, and for all the matrices $J \in \mathbf{R}^{n \times n}$ such that there exists $S \in \mathbf{R}^{n \times s}$ satisfying $(S \;\; J) \in \partial\Phi(\bar{\sigma}, \bar{x})$, the solution $\bar{x}$ of the GE

$$\Phi(\bar{\sigma}, \bar{x}) + J(x - \bar{x}) + N(x) \ni 0$$

is strongly regular.

Then there exist neighborhoods $\mathcal{U}$ of $\bar{\sigma}$ and U of $\bar{x}$ such that for all $\sigma \in \mathcal{U}$ there exists the unique $x(\sigma) \in U$ satisfying the GE

$$\Phi(\sigma,\, x) + N(x) \ni 0, \tag{3.140}$$

and the mapping $x(\cdot)$ is locally Lipschitz-continuous at $\bar{\sigma}$.

Proof. We shall prove the stated assertion using Theorem 3.49, by means of the well-known trick commonly used to derive implicit function theorems from inverse function theorems. In particular, this trick was employed for this purpose in [44].

Define the auxiliary mapping $\Psi : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^s \times \mathbf{R}^n$ by

$$\Psi(u) = (z,\, \Phi(z,\, x)),$$

and the multifunction M from $\mathbf{R}^s \times \mathbf{R}^n$ to the subsets of $\mathbf{R}^s \times \mathbf{R}^n$ by

$$M(u) = \{-\bar{\sigma}\} \times N(x),$$

where $u = (z,\, x)$. Then $\bar{u} = (\bar{\sigma},\, \bar{x})$ is a solution of the GE

$$\Psi(u) + M(u) \ni 0. \tag{3.141}$$

Moreover, the perturbed GE

$$\Psi(u) + M(u) \ni v$$

with $v = (\sigma - \bar{\sigma},\, 0)$, $\sigma \in \mathbf{R}^s$, takes the form of the system

$$z = \sigma, \quad \Phi(z,\, x) + N(x) \ni 0,$$

which is further equivalent to (3.140). Therefore, Theorem 3.50 will readily follow from Theorem 3.49 if we will show that $\bar{u}$ is a CD-regular solution of the GE (3.141).

Evidently, $\partial\Psi(\bar{u})$ consists of matrices of the form

$$\Lambda = \begin{pmatrix} I & 0 \\ S & J \end{pmatrix}$$

with the identity matrix $I \in \mathbf{R}^{s \times s}$, where $(S \;\; J) \in \partial\Phi(\bar{\sigma},\, \bar{x})$. The GE

$$\Psi(\bar{u}) + \Lambda(u - \bar{u}) + M(u) \ni w$$

with such matrix Λ, and with $w = (\zeta,\, \nu) \in \mathbf{R}^s \times \mathbf{R}^n$, takes the form of the system

$$\sigma = \bar{\sigma} + \zeta, \quad \Phi(\bar{\sigma},\, \bar{x}) + S(\sigma - \bar{\sigma}) + J(x - \bar{x}) + N(x) \ni \nu,$$

which is further equivalent to the GE

$$\Phi(\bar{\sigma}, \bar{x}) + J(x - \bar{x}) + N(x) \ni r$$

with $r = \nu - S\zeta$. From the assumptions of the theorem it now evidently follows that the solution $\bar{u}$ of the GE (3.141) is CD-regular. $\square$

3.4 Active-Set Methods for Complementarity Problems

Active-set strategies are tools intended for identifying inequalities that are active at a solution, and enforcing them as equalities on the subsequent iterations. As equalities are generally easier to deal with numerically, such techniques are thus useful for the acceleration of algorithms. In particular, they are often used in conjunction with Newton-type methods in the local phase. Specifically, when a given algorithm approaches a solution, some criteria are applied to identify active inequalities, and once this is done, there is a switch to a Newton-type method for the resulting system of equations in the variational setting, or to an equality-constrained problem in case of optimization.

As we have seen in Sects. 3.1 and 3.2, the Josephy–Newton method and the semismooth Newton methods for complementarity problems have a tendency to eventually identify active inequalities, at least when converging to a solution satisfying the strict complementarity condition. However, this identification is, in a sense, implicit. Also, it is not guaranteed if strict complementarity does not hold. In this section, we are interested in *intentional and explicit* identification of active inequalities. Methods that employ special identification procedures, and make explicit use of the identification predictions, are often referred to as *active-set methods*. Some relatively recent examples of such developments for variational problems can be found in [48–51, 69, 140, 163]. Many successful methods of this kind were developed for constrained optimization problems; some related references will be provided in Sect. 4.1.3.

Identification procedures used in practice are often heuristic by nature, at least in part. Loosely speaking, in many cases one can expect that sufficiently close to a solution, the active set of the current subproblem of the method in question estimates in some sense the active set of the original problem. Some partial theoretical foundations for identification procedures of this kind can be found in [179]. Generally, however, exact identification for approaches of this type cannot be fully proven (unless some strong assumptions, otherwise not needed, are introduced). In this section we focus on identification procedures of a different kind, allowing for strict theoretical justification under very mild assumptions. This approach was suggested in [71] in the context of constrained optimization problems; it relies on computable error bounds estimating the distance from the current iterate to the solution set.

3.4.1 Identification Based on Error Bounds

Consider the nonlinear complementarity problem (NCP)

$$x \geq 0, \quad \Phi(x) \geq 0, \quad \langle x, \Phi(x) \rangle = 0, \tag{3.142}$$

where $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ is a smooth mapping.

Let $\bar{x} \in \mathbf{R}^n$ be a solution of the NCP (3.142), and define the index sets

$$\begin{aligned} I_0(\bar{x}) &= \{i = 1, \ldots, n \mid \bar{x}_i = \Phi_i(\bar{x}) = 0\}, \\ I_1(\bar{x}) &= \{i = 1, \ldots, n \mid \bar{x}_i > 0, \; \Phi_i(\bar{x}) = 0\}, \\ I_2(\bar{x}) &= \{i = 1, \ldots, n \mid \bar{x}_i = 0, \; \Phi_i(\bar{x}) > 0\}. \end{aligned}$$

Recall that $(I_0(\bar{x}), I_1(\bar{x}), I_2(\bar{x}))$ is a partition of the set $\{1, \ldots, n\}$.

If the index sets $I_0 = I_0(\bar{x})$, $I_1 = I_1(\bar{x})$ and $I_2 = I_2(\bar{x})$ can be correctly identified using information available at a point x close enough to the solution $\bar{x}$, then locally the NCP (3.142) can be reduced to the system of nonlinear equations

$$\Phi_{I_0 \cup I_1}(x) = 0, \quad x_{I_0 \cup I_2} = 0.$$

For simplicity of notation, suppose that the components of $x \in \mathbf{R}^n$ are ordered in such a way that $x = (x_{I_1}, x_{I_0 \cup I_2})$. Then solving the NCP (3.142) locally reduces to solving the following system of equations:

$$\Phi_{I_0 \cup I_1}(x_{I_1}, 0) = 0. \tag{3.143}$$

This system can be solved by Newton-type methods discussed in Sect. 2.1.

Observe that in the absence of strict complementarity (when $I_0 \neq \emptyset$), system (3.143) is over-determined: the number of equations is larger than the number of unknowns. For example, one can apply the Gauss–Newton method (see (2.41)): for the current iterate $x^k \in \mathbf{R}^n$, the next iterate x^{k+1} is computed as a solution of the linear system

$$\begin{aligned} \left(\frac{\partial \Phi_{I_0 \cup I_1}}{\partial x_{I_1}}(x^k_{I_1}, 0) \right)^{\mathrm{T}} \Phi_{I_0 \cup I_1}(x^k_{I_1}, 0) \\ + \left(\frac{\partial \Phi_{I_0 \cup I_1}}{\partial x_{I_1}}(x^k_{I_1}, 0) \right)^{\mathrm{T}} \frac{\partial \Phi_{I_0 \cup I_1}}{\partial x_{I_1}}(x^k_{I_1}, 0)(x_{I_1} - x^k_{I_1}) = 0, \\ x_{I_0 \cup I_2} = 0. \end{aligned} \tag{3.144}$$

According to the discussion in Sect. 2.1.1, if Φ is twice differentiable near $\bar{x}$ with its second derivative continuous at $\bar{x}$, local superlinear convergence of this method to $\bar{x}$ holds under the assumption

$$\ker \frac{\partial \Phi_{I_0(\bar{x}) \cup I_1(\bar{x})}}{\partial x_{I_1(\bar{x})}}(\bar{x}) = \{0\}. \tag{3.145}$$

Following [49], we say that a solution $\bar{x}$ of the NCP (3.142) is *weakly regular* if (3.145) holds.

Recall that according to Proposition 1.34, $\bar{x}$ is a semistable solution of the NCP (3.142) if and only if the system

$$\begin{gathered} \xi_i \geq 0, \quad \langle \Phi_i'(\bar{x}),\, \xi\rangle \geq 0, \quad \xi_i\langle \Phi_i'(\bar{x}),\, \xi\rangle = 0, \quad i \in I_0(\bar{x}), \\ \Phi'_{I_1(\bar{x})}(\bar{x})\xi = 0, \quad \xi_{I_2(\bar{x})} = 0 \end{gathered} \tag{3.146}$$

has the unique solution $\xi = 0$.

It turns out that semistability implies weak regularity.

Proposition 3.51. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable at $\bar{x} \in \mathbf{R}^n$, and let $\bar{x}$ be a semistable solution of the NCP (3.142).*

Then $\bar{x}$ is weakly regular.

Proof. Indeed, assuming that there exists $\xi_{I_1(\bar{x})} \in \ker \frac{\partial \Phi_{I_0(\bar{x})\cup I_1(\bar{x})}}{\partial x_{I_1(\bar{x})}}(\bar{x}) \setminus \{0\}$, and setting $\xi_{I_0(\bar{x})\cup I_2(\bar{x})} = 0$, we obtain $\xi \in \mathbf{R}^n$ satisfying (3.146) and such that $\xi \neq 0$, which contradicts semistability. $\square$

The absence of the converse implication is demonstrated by the following example, taken from [49].

Example 3.52. Let $n = 2$, $\Phi(x) = ((x_1 - 1)^2,\, x_1 + x_2 - 1)$. Then $\bar{x} = (1,\, 0)$ is a solution of the NCP (3.142), and we have $I_0(\bar{x}) = \{2\}$, $I_1(\bar{x}) = \{1\}$, $I_2(\bar{x}) = \emptyset$. Weak regularity of $\bar{x}$ certainly holds, as

$$\frac{\partial \Phi_{I_0(\bar{x})\cup I_1(\bar{x})}}{\partial x_{I_1(\bar{x})}}(\bar{x}) \neq 0.$$

At the same time, the system (3.146) takes the form

$$\xi_2 \geq 0, \quad \xi_1 + \xi_2 \geq 0, \quad \xi_2(\xi_1 + \xi_2) = 0,$$

which evidently has nontrivial solutions. Therefore, $\bar{x}$ is not semistable.

We now discuss how the index sets $I_0(\bar{x})$, $I_1(\bar{x})$ and $I_2(\bar{x})$ can be identified. Note first that if $\bar{x}$ satisfies the strict complementarity condition $I_0(\bar{x}) = \emptyset$, then local identification is easy, at least as a matter of theory, and in particular it does not require any additional assumptions. Indeed, by the basic continuity considerations, it holds that

$$I_1(x) = I_1(\bar{x}), \quad I_2(x) = I_2(\bar{x})$$

for all $x \in \mathbf{R}^n$ close enough to $\bar{x}$, where for each $x \in \mathbf{R}^n$ we define

$$I_1(x) = \{i = 1,\, \ldots,\, n \mid x_i \geq \Phi_i(x)\}, \quad I_2(x) = \{i = 1,\, \ldots,\, n \mid x_i < \Phi_i(x)\}.$$

However, without strict complementarity, we only have the inclusions

$$I_1(\bar{x}) \subset I_1(x) \subset I_0(\bar{x}) \cup I_1(\bar{x}), \quad I_2(\bar{x}) \subset I_2(x) \subset I_0(\bar{x}) \cup I_2(\bar{x})$$

for all $x \in \mathbf{R}^n$ close enough to $\bar{x}$.

To guarantee exact identification of the relevant index sets without the restrictive strict complementarity assumption, the construction is as follows. Define the mapping $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ by

$$\Psi(x) = \psi(x, \Phi(x)), \tag{3.147}$$

where $\psi : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ is either the natural residual function

$$\psi(a, b) = \min\{a, b\}, \tag{3.148}$$

or the Fischer–Burmeister function

$$\psi(a, b) = a + b - \sqrt{a^2 + b^2}, \tag{3.149}$$

applied componentwise.

Assume now that $\bar{x}$ is semistable (which is a relatively weak assumption). As discussed in Sect. 3.2, from Propositions 1.34, 1.64, and 3.21 it follows that semistability is equivalent to the error bound

$$\|x - \bar{x}\| = O(\|\Psi(x)\|) \tag{3.150}$$

as $x \in \mathbf{R}^n$ tends to $\bar{x}$.

Define the function $\rho : \mathbf{R}_+ \to \mathbf{R}$ by

$$\rho(t) = \begin{cases} \bar{\rho} & \text{if } t \geq \bar{t}, \\ -1/\log t & \text{if } t \in (0, \bar{t}), \\ 0 & \text{if } t = 0, \end{cases} \tag{3.151}$$

with some fixed $\bar{t} \in (0, 1)$ and $\bar{\rho} > 0$. The choices of $\bar{t}$ and of $\bar{\rho}$ do not affect the theoretical analysis below; [71] suggests to take $\bar{t} = 0.9$ and $\bar{\rho} = -1/\log \bar{t}$. For any $x \in \mathbf{R}^n$, define further the index sets

$$\begin{aligned} I_0(x) &= \{i = 1, \ldots, n \mid \max\{x_i, \Phi_i(x)\} \leq \rho(\|\Psi(x)\|)\}, \\ I_1(x) &= \{i = 1, \ldots, n \mid x_i > \rho(\|\Psi(x)\|), \Phi_i(x) \leq \rho(\|\Psi(x)\|)\}, \\ I_2(x) &= \{i = 1, \ldots, n \mid x_i \leq \rho(\|\Psi(x)\|), \Phi_i(x) > \rho(\|\Psi(x)\|)\}. \end{aligned} \tag{3.152}$$

Then (locally) the prediction provided by those index sets is, in fact, correct.

Proposition 3.53. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable at $\bar{x} \in \mathbf{R}^n$, and let $\bar{x}$ be a semistable solution of the NCP (3.142).*

Then for the index sets defined according to (3.147), (3.152), with the complementarity function $\psi : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ defined either by (3.148) or by (3.149), and with the function $\rho : \mathbf{R}_+ \to \mathbf{R}$ defined by (3.151) with some $\bar{t} \in (0, 1)$ and $\bar{\rho} > 0$, it holds that

$$I_0(x) = I_0(\bar{x}), \quad I_1(x) = I_1(\bar{x}), \quad I_2(x) = I_2(\bar{x}) \tag{3.153}$$

for all $x \in \mathbf{R}^n$ close enough to $\bar{x}$.

Proof. Take any $i \in I_0(\bar{x})$. Then by the error bound (3.150) we have that

$$x_i \leq |x_i - \bar{x}_i| = O(\|\Psi(x)\|) = o(\rho(\|\Psi(x)\|))$$

as $x \in \mathbf{R}^n$ tends to $\bar{x}$, where the last estimate is by (3.151) and by the relation $\lim_{t \to 0+} t \log t = 0$. By similar reasoning, employing the differentiability of Φ at $\bar{x}$, we have that

$$\Phi_i(x) \leq |\Phi_i(x) - \Phi_i(\bar{x})| = O(\|x - \bar{x}\|) = O(\|\Psi(x)\|) = o(\rho(\|\Psi(x)\|)).$$

According to (3.152), the above shows that $i \in I_0(x)$; hence, $I_0(\bar{x}) \subset I_0(x)$ provided x is close enough to $\bar{x}$.

Take now any $i \in \{1, \ldots, n\} \setminus I_0(\bar{x})$. In that case, there exists some $\gamma > 0$ such that for any $x \in \mathbf{R}^n$ close enough to $\bar{x}$ it holds that $\rho(\|\Psi(x)\|) \leq \gamma$, and either $x_i > \gamma$ or $\Phi_i(x) > \gamma$. According to (3.152), it follows that $i \notin I_0(x)$, and hence, $I_0(x) \subset I_0(\bar{x})$.

We have therefore established the first equality in (3.153). The other equalities in (3.153) are verified by very similar considerations. □

Remark 3.54. Of course, (3.151) is not the only possible choice for the function ρ in Proposition 3.53. For example, one can easily check that one alternative appropriate choice is given by

$$\rho(t) = t^{\tau} \tag{3.154}$$

with any fixed $\tau \in (0, 1)$.

We formalize the constructions presented above as the following algorithm, which can be called the *active-set Gauss–Newton method* for the NCP (3.142).

Algorithm 3.55 Choose a complementarity function $\psi : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ defined by (3.148) or by (3.149), and define the mapping $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ according to (3.147). Define the function $\rho : \mathbf{R}_+ \to \mathbf{R}$ by (3.151) with some $\bar{t} \in (0, 1)$ and $\bar{\rho} > 0$, or by (3.154) with some $\tau \in (0, 1)$.

Choose $x^0 \in \mathbf{R}^n$ and define the index sets $I_0 = I_0(x^0)$, $I_1 = I_1(x^0)$ and $I_2 = I_2(x^0)$ according to (3.152). Set $k = 0$.

1. If $\Psi(x^k) = 0$, stop.
2. Compute $x^{k+1} \in \mathbf{R}^n$ as a solution of (3.144).
3. Increase k by 1 and go to step 1.

Combining Propositions 3.51, 3.53 and Remark 3.54 with the discussion above on the Gauss–Newton method, we obtain the following local convergence result.

Theorem 3.56. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its second derivative being continuous at $\bar{x}$. Let $\bar{x}$ be a semistable solution of the NCP (3.142).*

Then any starting point $x^0 \in \mathbf{R}^n$ close enough to $\bar{x}$ uniquely defines a particular iterative sequence of Algorithm 3.55, this sequence converges to $\bar{x}$, and the rate of convergence is superlinear.

3.4.2 Extension to Mixed Complementarity Problems

In this section we consider the mixed complementarity problem (MCP) which is the variational inequality with bound constraints:

$$u \in [a,\, b], \quad \langle \Phi(u),\, v - u\rangle \geq 0 \quad \forall\, v \in [a,\, b], \tag{3.155}$$

where

$$[a,\, b] = \{u \in \mathbf{R}^n \mid a_i \leq u_i \leq b_i,\ i = 1,\, \ldots,\, \nu\}$$

is a (generalized) box, $\Phi : \mathbf{R}^\nu \to \mathbf{R}^\nu$ is a smooth mapping, $a_i \in \mathbf{R} \cup \{-\infty\}$, $b_i \in \mathbf{R} \cup \{+\infty\}$, $a_i < b_i$, $i = 1,\, \ldots,\, \nu$. Recall that problem (3.155) can be equivalently stated as follows:

$$u \in [a,\, b], \quad \Phi_i(u) \begin{cases} \geq 0 & \text{if } u_i = a_i, \\ = 0 & \text{if } a_i < u_i < b_i, \quad i = 1,\, \ldots,\, \nu. \\ \leq 0 & \text{if } u_i = b_i, \end{cases} \tag{3.156}$$

The development of active-set methods for MCP is very similar to that for NCP. Let $\bar{u} \in \mathbf{R}^\nu$ be a solution of the MCP (3.156), and define the index sets

$$\begin{aligned} I_{0a}(\bar{u}) &= \{i = 1,\, \ldots,\, \nu \mid \bar{u}_i = a_i,\ \Phi_i(\bar{u}) = 0\}, \\ I_{0b}(\bar{u}) &= \{i = 1,\, \ldots,\, \nu \mid \bar{u}_i = b_i,\ \Phi_i(\bar{u}) = 0\}, \\ I_1(\bar{u}) &= \{i = 1,\, \ldots,\, \nu \mid a_i < \bar{u}_i < b_i,\ \Phi_i(\bar{u}) = 0\}, \\ I_{2a}(\bar{u}) &= \{i = 1,\, \ldots,\, \nu \mid \bar{u}_i = a_i,\ \Phi_i(\bar{u}) > 0\}, \\ I_{2b}(\bar{u}) &= \{i = 1,\, \ldots,\, \nu \mid \bar{u}_i = b_i,\ \Phi_i(\bar{u}) < 0\}. \end{aligned}$$

Observe that the tuple $(I_{0a}(\bar{u}),\, I_{0b}(\bar{u}),\, I_1(\bar{u}),\, I_{2a}(\bar{u}),\, I_{2b}(\bar{u}))$ is a partition of the set $\{1,\, \ldots,\, \nu\}$.

If the index sets $I_{0a} = I_{0a}(\bar{u})$, $I_{0b} = I_{0b}(\bar{u})$, $I_1 = I_1(\bar{u})$, $I_{2a} = I_{2a}(\bar{u})$ and $I_{2b} = I_{2b}(\bar{u})$ can be correctly identified at a point u close enough to the solution $\bar{u}$, then locally the MCP (3.156) reduces to the system of equations

$$\Phi_{I_0 \cup I_1}(u) = 0, \quad u_{I_a} = a_{I_a},\ u_{I_b} = b_{I_b},$$

where

$$I_0 = I_{0a} \cup I_{0b}, \quad I_a = I_{0a} \cup I_{2a}, \quad I_b = I_{0b} \cup I_{2b}.$$

Suppose that the components of $u \in \mathbf{R}^\nu$ are ordered in such a way that $u = (u_{I_1},\, u_{I_{0a} \cup I_{2a}},\, u_{I_{0b} \cup I_{2b}})$. Then solving the MCP (3.156) locally reduces to solving the system of equations

$$\Phi_{I_0 \cup I_1}(u_{I_1},\, a_{I_a},\, b_{I_b}) = 0. \tag{3.157}$$

In the absence of strict complementarity (when $I_0 \neq \emptyset$), this system is over-determined. Thus, one option is to apply the Gauss–Newton method: for the current iterate $u^k \in \mathbf{R}^\nu$, the next iterate u^{k+1} is computed as a solution of the linear system

$$\begin{gathered}\left(\frac{\partial \Phi_{I_0\cup I_1}}{\partial u_{I_1}}(u^k_{I_1},\, a_{I_a},\, b_{I_b})\right)^{\mathrm{T}} \Phi_{I_0\cup I_1}(u^k_{I_1},\, a_{I_a},\, b_{I_b}) \\ +\left(\frac{\partial \Phi_{I_0\cup I_1}}{\partial u_{I_1}}(u^k_{I_1},\, a_{I_a},\, b_{I_b})\right)^{\mathrm{T}} \frac{\partial \Phi_{I_0\cup I_1}}{\partial u_{I_1}}(u^k_{I_1},\, a_{I_a},\, b_{I_b})(u_{I_1}-u^k_{I_1}) = 0, \\ u_{I_a} = a_{I_a},\ u_{I_b} = b_{I_b}.\end{gathered} \tag{3.158}$$

According to the discussion in Sect. 2.1, if Φ is twice differentiable near $\bar{u}$ with its second derivative continuous at $\bar{x}$, the local superlinear convergence of this method to $\bar{u}$ holds if $\bar{u}$ is a *weakly regular* solution of the MCP (3.156), i.e., if

$$\ker \frac{\partial \Phi_{I_{0a}(\bar{u})\cup I_{0b}(\bar{u})\cup I_1(\bar{u})}}{\partial u_{I_1(\bar{u})}}(\bar{u}) = \{0\}.$$

The index sets $I_{0a}(\bar{u})$, $I_{0b}(\bar{u})$, $I_1(\bar{u})$, $I_{2a}(\bar{u})$, and $I_{2b}(\bar{u})$ can be locally identified similarly to the case of NCP considered in Sect. 3.4.1. In order to do this, the error bounds in terms of the MCP residual mappings defined in Sect. 3.2.2 (see (3.59)) can be employed. If the residual mapping in (3.59) is defined using the natural residual or the Fischer–Burmeister complementarity function, these error bounds hold assuming semistability of the solution $\bar{u}$. See [49, 50] for further details, and in particular, for a characterization of semistability of MCP solutions, which can be readily derived from Proposition 1.33, similarly to Proposition 1.34.

In the rest of this section, we consider in some detail a particular instance of MCP, namely the KKT system

$$\begin{gathered}F(x) + (h'(x))^{\mathrm{T}}\lambda + (g'(x))^{\mathrm{T}}\mu = 0, \quad h(x) = 0, \\ \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu,\, g(x)\rangle = 0,\end{gathered} \tag{3.159}$$

where $F : \mathbf{R}^n \to \mathbf{R}^n$, $h : \mathbf{R}^n \to \mathbf{R}^l$, and $g : \mathbf{R}^n \to \mathbf{R}^m$ are smooth mappings.

It is easy to see that for a given solution $(\bar{x},\, \bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ of (3.159), the system (3.157) in this case takes the form

$$F(x) + (h'(x))^{\mathrm{T}}\lambda + (g'_{A_+}(x))^{\mathrm{T}}\mu_{A_+} = 0, \quad h(x) = 0, \quad g_A(x) = 0, \tag{3.160}$$

with $A = A(\bar{x})$ and $A_+ = A_+(\bar{x},\, \bar{\mu})$.

Define the mapping $G : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n$,

$$G(x,\, \lambda,\, \mu) = F(x) + (h'(x))^{\mathrm{T}}\lambda + (g'(x))^{\mathrm{T}}\mu.$$

Weak regularity of a solution $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ of the KKT system (3.159) means that

$$\ker\begin{pmatrix} \frac{\partial G}{\partial x}(\bar{x},\,\bar{\lambda},\,\bar{\mu}) & (h'(\bar{x}))^{\mathrm{T}} & (g'_{A_+(\bar{x},\,\bar{\mu})}(\bar{x}))^{\mathrm{T}} \\ h'(\bar{x}) & 0 & 0 \\ g'_{A(\bar{x})}(\bar{x}) & 0 & 0 \end{pmatrix} = \{0\}.$$

Assuming that F is twice differentiable near $\bar{x}$ with its second derivative being continuous at $\bar{x}$, and that h and g are thrice differentiable near $\bar{x}$ with their third derivatives being continuous at $\bar{x}$, weak regularity of $(\bar{x},\,\bar{\lambda},\,\bar{\mu})$ implies local superlinear convergence of the Gauss–Newton method applied to the system (3.160). However, it may be more relevant to use different techniques in this setting, in particular preserving the primal-dual structure of this system; see [49, 51, 140].

We proceed to describe the identification procedure for the index sets $A(\bar{x})$ and $A_+(\bar{x},\,\bar{\mu})$. Define the residual mapping $\Psi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ by

$$\Psi(x,\,\lambda,\,\mu) = (G(x,\,\lambda,\,\mu),\,h(x),\,\psi(\mu,\,-g(x))), \tag{3.161}$$

where $\psi : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ is given either by (3.148) or by (3.149). From Propositions 1.35, 1.64, and 3.33, it follows that the semistability of a solution $(\bar{x},\,\bar{\lambda},\,\bar{\mu})$ of the KKT system (3.159) is equivalent to the error bound

$$\|(x-\bar{x},\,\lambda-\bar{\lambda},\,\mu-\bar{\mu})\| = O(\|\Psi(x,\,\lambda,\,\mu)\|) \tag{3.162}$$

as $(x,\,\lambda,\,\mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ tends to $(\bar{x},\,\bar{\lambda},\,\bar{\mu})$.

For the function $\rho : \mathbf{R}_+ \to \mathbf{R}$ defined in (3.151) with some fixed $\bar{t} \in (0,1)$ and $\bar{\rho} > 0$, and for any $(x,\,\lambda,\,\mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, define the index sets

$$\begin{aligned} A(x,\,\lambda,\,\mu) &= \{i = 1,\,\ldots,\,m \mid -g_i(x) \le \rho(\|\Psi(x,\,\lambda,\,\mu)\|)\}, \\ A_+(x,\,\lambda,\,\mu) &= \{i = 1,\,\ldots,\,m \mid \mu_i \ge \rho(\|\Psi(x,\,\lambda,\,\mu)\|)\}. \end{aligned} \tag{3.163}$$

Proposition 3.57. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable at a point $\bar{x} \in \mathbf{R}^n$, let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable at $\bar{x}$, and let $(\bar{x},\,\bar{\lambda},\,\bar{\mu})$ with some $(\bar{\lambda},\,\bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be a semistable solution of the KKT system* (3.159).

Then for the index sets defined according to (3.161), (3.163)*, with the complementarity function $\psi : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ given either by* (3.148) *or by* (3.149)*, and with the function $\rho : \mathbf{R}_+ \to \mathbf{R}$ defined by* (3.151) *with some $\bar{t} \in (0,1)$ and $\bar{\rho} > 0$, it holds that*

$$A(x,\,\lambda,\,\mu) = A(\bar{x}), \quad A_+(x,\,\lambda,\,\mu) = A_+(\bar{x},\,\bar{\mu}) \tag{3.164}$$

for all $(x,\,\lambda,\,\mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x},\,\bar{\lambda},\,\bar{\mu})$.

Proof. Take any $i \in A(\bar{x})$. Then by the error bound (3.162), similarly to the proof of Proposition 3.53, we derive that

$$-g_i(x) \leq |g_i(x)-g_i(\bar{x})| = O(\|x-\bar{x}\|) = O(\|\Psi(x,\, \lambda,\, \mu)\|) = o(\rho(\|\Psi(x,\, \lambda,\, \mu)\|))$$

as $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ tends to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$. According to (3.163), the above shows that $i \in A(x,\, \lambda,\, \mu)$, and hence, $A(\bar{x}) \subset A(x,\, \lambda,\, \mu)$.

Take now any $i \in \{1,\, \ldots,\, m\} \setminus A(\bar{x})$. In that case, there exists some $\gamma > 0$ such that for any $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ it holds that $\rho(\|\Psi(x,\, \lambda,\, \mu)\|) \leq \gamma$ and $-g_i(x) > \gamma$. According to (3.163), it follows that $i \notin A(x,\, \lambda,\, \mu)$, and hence, $A(x,\, \lambda,\, \mu) \subset A(\bar{x})$.

We have therefore established the first equality in (3.164).

Now take any $i \in A_+(\bar{x},\, \bar{\mu})$. Then there exists some $\gamma > 0$ such that for any $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ it holds that $\rho(\|\Psi(x,\, \lambda,\, \mu)\|) \leq \gamma$ and $\mu_i > \gamma$. Then, according to (3.163), it follows that $i \in A_+(x,\, \lambda,\, \mu)$, and hence, $A_+(\bar{x},\, \bar{\mu}) \subset A_+(x,\, \lambda,\, \mu)$.

Take now any $i \in \{1,\, \ldots,\, m\} \setminus A_+(\bar{x},\, \bar{\mu})$. Then $\bar{\mu}_i = 0$ and hence, again employing the error bound (3.162), we derive that

$$\mu_i \leq |\mu_i - \bar{\mu}_i| = O(\|\Psi(x,\, \lambda,\, \mu)\|) = o(\rho(\|\Psi(x,\, \lambda,\, \mu)\|)).$$

Then, according to (3.163), the above shows that $i \notin A_+(x,\, \lambda,\, \mu)$, and therefore, $A_+(x,\, \lambda,\, \mu) \subset A_+(\bar{x},\, \bar{\mu})$.

We have thus established the second equality in (3.164). $\square$

Recall that for KKT systems arising from optimization, semistability of solutions is characterized in Proposition 1.37.

The first equality in (3.164) remains valid if the semistability condition in Proposition 3.57 is replaced by the weaker assumption that a Lagrange multiplier $(\bar{\lambda},\, \bar{\mu})$ associated with $\bar{x}$ is noncritical in the sense of Definition 1.41, thus allowing for the case when this multiplier is not necessarily unique (see also the related discussion in Sect. 1.3.3). Indeed, by Proposition 1.43, the assumption that $(\bar{\lambda},\, \bar{\mu})$ is noncritical implies the error bound

$$\|x - \bar{x}\| = O(\|\Psi(x,\, \lambda,\, \mu)\|)$$

as $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ tends to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$. This error bound (rather than the stronger (3.162)) is precisely what is needed for the proof of the first equality in (3.164).

Proposition 3.58. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable at a point $\bar{x} \in \mathbf{R}^n$, let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable at $\bar{x}$, and let $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ be a solution of the KKT system* (3.159) *with some noncritical Lagrange multiplier $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$.*

Then for the index set $A(x,\, \lambda,\, \mu)$ defined according to (3.147), (3.163), *with the complementarity function $\psi : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ given either by* (3.148) *or by* (3.149), *and with the function $\rho : \mathbf{R}_+ \to \mathbf{R}$ defined by* (3.151) *with some $\bar{t} \in (0, 1)$ and $\bar{\rho} > 0$, it holds that*

$$A(x,\, \lambda,\, \mu) = A(\bar{x})$$

for all $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$.

Remark 3.59. Similarly to Remark 3.54, observe that Propositions 3.57 and 3.58 remain valid if ρ is defined according to (3.154) with any fixed $\tau \in (0, 1)$.

Thus, under the assumptions of Proposition 3.58, we can expect to identify the index set $A = A(\bar{x})$. Then, instead of (3.160), we consider the system

$$F(x) + (h'(x))^{\mathrm{T}}\lambda + (g'_A(x))^{\mathrm{T}}\mu_A = 0, \quad h(x) = 0, \quad g_A(x) = 0. \tag{3.165}$$

Differently from (3.160), in this system the number of equations coincides with the number of variables, and thus the wide range of methods discussed in Sect. 2.1.1 can be applied. Note, however, that nonsingularity of the Jacobian of the system (3.165) at a solution $(\bar{x}, \bar{\lambda}, \bar{\mu})$ implies the LICQ at $\bar{x}$ which, in turn, implies that $(\bar{\lambda}, \bar{\mu})$ is the unique multiplier associated with $\bar{x}$. Therefore, in order to cover the cases of nonunique multipliers, correct identification of active constraints is not enough; some special Newton-type methods for the resulting system of equations need to be developed. We shall get back to active-set methods in Sect. 4.1.3 in the context of optimization problems, and also in Chap. 7 in the context of degenerate problems.

Observe finally that the identification procedures discussed in this section can be used in conjunction with error bounds of a completely different nature, and with different residuals. For example, let F, h and g and the derivatives of h and g be subanalytic near $\bar{x}$ in the sense of [68, Definition 6.6.1], and let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ be a solution of the system (3.159) with some Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$. Then according to [68, Corollary 6.6.4], there exists $\nu > 0$ such that the error bound

$$\|x - \bar{x}\| = O(\|\Psi(x, \lambda, \mu)\|^{\nu})$$

holds as $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ tends to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, where the residual mapping $\Psi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$ is defined as

$$\Psi(x, \lambda, \mu) = (G(x, \lambda, \mu), h(x), \min\{0, \mu\}, \max\{0, g(x)\}, \langle \mu, g(x)\rangle). \tag{3.166}$$

It is important to stress that the value of ν is generally not known in this error bound result for subanalytic objects. Nevertheless, both Propositions 3.57 and 3.58 remain valid with the residual mapping Ψ defined according to (3.166), and with ρ given by (3.151). This is due to the fact that $\lim_{t\to 0+} t^{\nu} \log t = 0$ holds for any fixed ν. At the same time, the use of the function ρ defined according to (3.154) becomes problematic, since one would have to take $\tau \in (0, \nu)$.

Chapter 4
Constrained Optimization: Local Methods

This chapter is devoted to local convergence analysis of Newton-type methods for optimization with constraints. We start with the simpler case of equality constraints, presenting the basic Newton method applied to the Lagrange optimality system, as well as its inexact and quasi-Newton modifications. We also discuss the augmented Lagrangian-based approach, and in particular the linearly constrained Lagrangian (LCL) method, relating it to the perturbed Newton iterations. Then the fundamental sequential quadratic programming (SQP) algorithm for the general case of equality and inequality constraints is presented. Sharp convergence results are derived by relating it to the Josephy–Newton method for generalized equations and applying the results of Sect. 3.1. Then the perturbed SQP framework is introduced, which allows to treat, in addition and in a unified manner, not only various modifications of SQP itself (truncated, augmented Lagrangian based versions, and second-order corrections) but also a number of other algorithms even though their iterative subproblems are different from SQP (sequential quadratically constrained quadratic programming (SQCQP), inexact restoration, and a certain interior feasible directions method). In many cases, this line of analysis allows to derive local convergence results that are sharper than those based on other considerations.

4.1 Equality-Constrained Problems

In this section, we consider the equality-constrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \end{array} \tag{4.1}$$

where the objective function $f : \mathbf{R}^n \to \mathbf{R}$ and the constraints mapping $h : \mathbf{R}^n \to \mathbf{R}^l$ are twice differentiable. The case when there are no inequality

A.F. Izmailov and M.V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-3-319-04247-3_4,

constraints is structurally much simpler, as it allows, for example, direct application of the classical Newton method to the optimality system of problem (4.1), as well as various modifications of Newton method discussed in Sect. 2.1.1. On the other hand, this optimality system is a particular case of the generalized equation or of the Karush–Kuhn–Tucker system, and thus some of the material to be presented can be regarded as a particular case of the developments in Sect. 3.1 or in Sect. 3.2.2. However, we believe that direct analysis free of complications associated with the more general problem settings is more instructive in this case. Therefore, we shall rely only upon results concerning usual equations in Sect. 2.1.1.

4.1.1 Newton–Lagrange Method

According to Sect. 1.2.3, stationary points and associated Lagrange multipliers of problem (4.1) are characterized by the Lagrange optimality system

$$L'(x,\, \lambda) = 0, \tag{4.2}$$

where $L : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}$ is the Lagrangian of problem (4.1):

$$L(x,\, \lambda) = f(x) + \langle \lambda,\, h(x)\rangle.$$

Since (4.2) is a system of $n + l$ equations in the same number of variables, to compute stationary points of problem (4.1) and associated multipliers, we can apply Newton-type methods for usual equations to system (4.2); see Sect. 2.1.1. Assuming that the Hessian $L''(x^k,\, \lambda^k)$ is nonsingular, the iterative scheme of the basic Newton method for (4.2) is then given by

$$(x^{k+1},\, \lambda^{k+1}) = (x^k,\, \lambda^k) - (L''(x^k,\, \lambda^k))^{-1} L'(x^k,\, \lambda^k), \quad k = 0,\, 1,\, \ldots.$$

Computing the derivatives of the Lagrangian, the Newton iteration can be written in the following more detailed form. Let $x^k \in \mathbf{R}^n$ be the current approximation to a stationary point of problem (4.1), and let $\lambda^k \in \mathbf{R}^l$ be an approximation of a Lagrange multiplier associated with this stationary point. Then the next iterate $(x^{k+1},\, \lambda^{k+1})$ is computed as a solution of the linear system

$$\begin{gathered}\frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)(x - x^k) + (h'(x^k))^{\mathrm{T}}(\lambda - \lambda^k) = -\frac{\partial L}{\partial x}(x^k,\, \lambda^k),\\ h'(x^k)(x - x^k) = -h(x^k),\end{gathered} \tag{4.3}$$

with respect to $(x,\, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$.

The resulting iterative procedure is naturally referred to as the *Newton–Lagrange method.*

Algorithm 4.1 Choose $(x^0,\, \lambda^0) \in \mathbf{R}^n \times \mathbf{R}^l$ and set $k = 0$.

1. If $L'(x^k,\, \lambda^k) = 0$, stop.
2. Compute $(x^{k+1},\, \lambda^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^l$ as a solution of the linear system (4.3).
3. Increase k by 1 and go to step 1.

As seen in Sect. 2.1.1, the key issue in applying any Newton method is nonsingularity of the Jacobian of the equation mapping at a solution of interest. In the case of the optimization problem (4.1) and the Newton method applied to its optimality system (4.2), this is guaranteed under the regularity assumption on the constraints and the second-order sufficient optimality condition at the solution. This fact is actually a particular case of any of the Propositions 1.28, 1.37, or 3.32, but we shall give a direct proof.

Lemma 4.2. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice differentiable at $\bar{x} \in \mathbf{R}^n$. Let $\bar{x}$ be a stationary point of problem (4.1), satisfying the regularity condition*

$$\operatorname{rank} h'(\bar{x}) = l. \tag{4.4}$$

Let $\bar{\lambda} \in \mathbf{R}^l$ be the (unique) Lagrange multiplier associated with $\bar{x}$, and assume that the SOSC introduced in Theorem 1.13 holds:

$$\left\langle \frac{\partial^2 L}{\partial x}(\bar{x},\, \bar{\lambda})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in \ker h'(\bar{x}) \setminus \{0\} \tag{4.5}$$

(thus, according to Theorem 1.13, $\bar{x}$ is a strict local solution of problem (4.1)).

Then the full Hessian of the Lagrangian

$$L''(\bar{x},\, \bar{\lambda}) = \begin{pmatrix} \dfrac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda}) & (h'(\bar{x}))^{\mathrm{T}} \\ h'(\bar{x}) & 0 \end{pmatrix}$$

is a nonsingular matrix.

Proof. Take an arbitrary $(\xi,\, \eta) \in \ker L''(\bar{x},\, \bar{\lambda})$. Then

$$\frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})\xi + (h'(\bar{x}))^{\mathrm{T}}\eta = 0, \tag{4.6}$$

$$h'(\bar{x})\xi = 0. \tag{4.7}$$

Multiplying the left- and right-hand sides of (4.6) by ξ and employing (4.7), we derive that

$$0 = \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})\xi,\, \xi \right\rangle + \langle \eta,\, h'(\bar{x})\xi \rangle = \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})\xi,\, \xi \right\rangle.$$

This implies that $\xi = 0$, since otherwise a contradiction with the SOSC (4.5) is obtained. But then (4.6) takes the form

$$(h'(\bar{x}))^{\mathrm{T}}\eta = 0. \tag{4.8}$$

Since the regularity condition (4.4) is equivalent to $\ker(h'(\bar{x}))^{\mathrm{T}} = \{0\}$, it follows that (4.8) can hold only when $\eta = 0$.

It is thus shown that $(\xi, \eta) = 0$, i.e., $\ker L''(\bar{x}, \bar{\lambda}) = \{0\}$, which means that $L''(\bar{x}, \bar{\lambda})$ is nonsingular. □

It can be easily seen that the regularity condition (4.4) is also necessary for the nonsingularity of $L''(\bar{x}, \bar{\lambda})$, while the SOSC (4.5) in Lemma 4.2 can be replaced by the weaker assumption that the multiplier $\bar{\lambda}$ is noncritical. This weaker assumption is also necessary for the nonsingularity of $L''(\bar{x}, \bar{\lambda})$.

The next result on the local superlinear convergence of the Newton–Lagrange method follows readily from Theorem 2.2 and Lemma 4.2.

Theorem 4.3. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (4.1)*, satisfying the regularity condition* (4.4)*. Let $\bar{\lambda} \in \mathbf{R}^l$ be the (unique) Lagrange multiplier associated with $\bar{x}$, and assume that the SOSC* (4.5) *holds.*

Then any starting point $(x^0, \lambda^0) \in \mathbf{R}^n \times \mathbf{R}^l$ close enough to $(\bar{x}, \bar{\lambda})$ uniquely defines the iterative sequence of Algorithm 4.1, this sequence converges to $(\bar{x}, \bar{\lambda})$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the second derivatives of f and h are locally Lipschitz-continuous with respect to $\bar{x}$.

Note that the matrix of the linear system (4.3) is symmetric. However, it is easy to check that this matrix cannot be positive or negative definite (unless $l = 0$, i.e., there are no constraints). System (4.3) can be solved by general-purpose tools of numerical linear algebra, though there also exist special methods taking advantage of the structure therein [29, 208].

Assuming the regularity condition (4.4), if a sufficiently good primal starting point $x^0 \in \mathbf{R}^n$ is available, then a sufficiently good dual starting point λ^0 can be constructed as follows. The regularity condition (4.4) implies that the matrix $h'(\bar{x})(h'(\bar{x}))^{\mathrm{T}}$ is nonsingular. Hence, by the definitions of a stationary point $\bar{x}$ and of an associated Lagrange multiplier $\bar{\lambda}$, it holds that

$$\bar{\lambda} = -(h'(\bar{x})(h'(\bar{x}))^{\mathrm{T}})^{-1}h'(\bar{x})f'(\bar{x}).$$

Therefore, the dual point

$$\lambda^0 = -(h'(x^0)(h'(x^0))^{\mathrm{T}})^{-1}h'(x^0)f'(x^0)$$

is well defined if x^0 is close enough to $\bar{x}$, and λ^0 tends to $\bar{\lambda}$ as x^0 tends to $\bar{x}$.

As the Newton–Lagrange method is just the Newton method applied to the special system of equations, it allows for all the modifications (in particular, those in the perturbed Newton framework) discussed in Sect. 2.1.1. For example, *truncated Newton–Lagrange methods* can be developed along the same lines as in Sect. 2.1.1. In the rest of this section and in Sect. 4.1.2, we concentrate on those instances of the perturbed Newton–Lagrange method which are somehow connected to the specific primal-dual structure of the Lagrange system (4.2).

Consider the following general scheme, which we refer to as the *perturbed Newton–Lagrange method*. For a given $(x^k, \lambda^k) \in \mathbf{R}^n \times \mathbf{R}^l$, the next iterate $(x^{k+1}, \lambda^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^l$ satisfies the perturbed version of the iteration system (4.3):

$$\begin{gathered}\frac{\partial L}{\partial x}(x^k, \lambda^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)(x - x^k) + (h'(x^k))^{\mathrm{T}}(\lambda - \lambda^k) + \omega_1^k = 0,\\ h(x^k) + h'(x^k)(x - x^k) + \omega_2^k = 0,\end{gathered} \tag{4.9}$$

where $\omega_1^k \in \mathbf{R}^n$ and $\omega_2^k \in \mathbf{R}^l$ are the perturbation terms.

Recall that, generally, the quadratic rate of convergence of a primal-dual sequence does not imply superlinear rate for primal convergence; see the discussion in Sect. 3.2.2. At the same time, fast primal convergence is often of particular importance in practice. In the presence of superlinear primal-dual convergence rate, the formal lack of superlinear primal convergence can hardly be regarded as a serious practical deficiency, since the rate of primal convergence is still high: the sequence of the distances to the primal solution is estimated from above by the superlinearly convergent sequence of distances to the primal-dual solution (see Examples 3.38 and 3.39). However, superlinear primal convergence becomes specially important (both theoretically and practically) in the context of perturbed Newton-type methods, when superlinear primal-dual convergence rate is by no means automatic. Then conditions characterizing superlinear primal rate can be regarded as a characterization of "*good perturbations*," i.e., those that preserve the nature and the advantages of Newton-type methods. Another consideration to support this view is that assumptions required for superlinear primal-dual rate can be too strong, unnecessarily restricting the class of potentially promising perturbed Newton-type methods. These strong assumptions may also rule out important types of perturbations useful for globalization of convergence; see Sect. 6.2.1. This viewpoint leads, for example, to a proper understanding of what can be regarded as a quasi-Newton–Lagrange method; see Theorem 4.6 below.

We present next an a posteriori result providing the necessary and sufficient conditions for primal superlinear convergence rate of the perturbed Newton–Lagrange method.

Proposition 4.4. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (4.1), and let $\bar{\lambda} \in \mathbf{R}^l$ be a Lagrange multiplier associated with $\bar{x}$. Let $\{(x^k, \lambda^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l$ be convergent to $(\bar{x}, \bar{\lambda})$, and for each $k = 0, 1, \ldots,$ set*

$$\omega_1^k = -\frac{\partial L}{\partial x}(x^k,\, \lambda^k) - \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)(x^{k+1} - x^k) - (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k), \quad (4.10)$$

$$\omega_2^k = -h(x^k) - h'(x^k)(x^{k+1} - x^k). \quad (4.11)$$

If the rate of convergence of $\{x^k\}$ *is superlinear, then*

$$\pi_{\ker h'(\bar{x})}\omega_1^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|), \quad (4.12)$$

$$\omega_2^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \quad (4.13)$$

as $k \to \infty$*.*

Conversely, if the SOSC (4.5) *holds, and conditions* (4.12)*,* (4.13) *hold as well, then the rate of convergence of* $\{x^k\}$ *is superlinear.*

Proof. For each k, from (4.10) and (4.11) we obtain that

$$\begin{aligned}
\omega_1^k &= -\frac{\partial L}{\partial x}(x^k,\, \lambda^{k+1}) - \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)(x^{k+1} - x^k) \\
&= -\frac{\partial L}{\partial x}(\bar{x},\, \lambda^{k+1}) - \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \lambda^{k+1})(x^k - \bar{x}) - \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)(x^{k+1} - x^k) \\
&\quad + o(\|x^k - \bar{x}\|) \\
&= -\frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda}) - (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) - \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})(x^{k+1} - \bar{x}) \\
&\quad + o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|) \\
&= -(h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) - \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})(x^{k+1} - \bar{x}) \\
&\quad + o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|), \qquad (4.14)
\end{aligned}$$

and

$$\begin{aligned}
\omega_2^k &= -h(\bar{x}) - h'(\bar{x})(x^k - \bar{x}) - h'(\bar{x})(x^{k+1} - x^k) \\
&\quad + o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|) \\
&= -h'(\bar{x})(x^{k+1} - \bar{x}) + o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|) \qquad (4.15)
\end{aligned}$$

as $k \to \infty$, where the convergence of $\{(x^k,\, \lambda^k)\}$ to $(\bar{x},\, \bar{\lambda})$ was taken into account.

Suppose first that the rate of convergence of $\{x^k\}$ is superlinear. Taking into account the equality $\operatorname{im}(h'(\bar{x}))^{\mathrm{T}} = (\ker h'(\bar{x}))^{\perp}$, (4.14) implies the estimate

$$\begin{aligned}
\pi_{\ker h'(\bar{x})}\omega_1^k &= O(\|x^{k+1} - \bar{x}\|) + o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|) \\
&= o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|).
\end{aligned}$$

This gives (4.12). Similarly, (4.15) implies the estimate

$$\begin{aligned}\omega_2^k &= O(\|x^{k+1}-\bar{x}\|) + o(\|x^{k+1}-x^k\|) + o(\|x^k-\bar{x}\|)\\ &= o(\|x^{k+1}-x^k\|) + o(\|x^k-\bar{x}\|),\end{aligned}$$

which gives (4.13).

Suppose now that (4.12), (4.13) hold. Employing (4.15), by Hoffman's error bound for linear systems (see Lemma A.4), for each k large enough there exists $\xi^k \in \mathbf{R}^n$ such that $\xi^k \in \ker h'(\bar{x})$ and

$$x^{k+1}-\bar{x} = \xi^k + o(\|x^{k+1}-x^k\|) + o(\|x^k-\bar{x}\|). \tag{4.16}$$

From (4.14) we now obtain that

$$\begin{aligned}\langle \omega_1^k,\, \xi^k\rangle &= -\langle \lambda^{k+1}-\bar{\lambda},\, h'(\bar{x})\xi^k\rangle - \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})(x^{k+1}-\bar{x}),\, \xi^k\right\rangle\\ &\quad + o(\|x^{k+1}-x^k\|\|\xi^k\|) + o(\|x^k-\bar{x}\|\|\xi^k\|)\\ &= -\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})(x^{k+1}-\bar{x}),\, \xi^k\right\rangle\\ &\quad + o(\|x^{k+1}-x^k\|\|\xi^k\|) + o(\|x^k-\bar{x}\|\|\xi^k\|)\end{aligned} \tag{4.17}$$

as $k \to \infty$.

Furthermore, the SOSC (4.5) implies the existence of $\gamma > 0$ such that

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})\xi,\, \xi\right\rangle \geq \gamma\|\xi\|^2 \quad \forall\, \xi \in \ker h'(\bar{x}).$$

Combining this with (4.17), we then derive

$$\begin{aligned}\gamma\|\xi^k\|^2 &\leq \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})\xi^k,\, \xi^k\right\rangle\\ &= \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})(x^{k+1}-\bar{x}),\, \xi^k\right\rangle\\ &\quad + o(\|x^{k+1}-x^k\|\|\xi^k\|) + o(\|x^k-\bar{x}\|\|\xi^k\|)\\ &= -\langle \omega_1^k,\, \xi^k\rangle + o(\|x^{k+1}-x^k\|\|\xi^k\|) + o(\|x^k-\bar{x}\|\|\xi^k\|)\\ &= -\langle \pi_{\ker h'(\bar{x})}\omega_1^k,\, \xi^k\rangle + o(\|x^{k+1}-x^k\|\|\xi^k\|) + o(\|x^k-\bar{x}\|\|\xi^k\|)\\ &= o((\|x^{k+1}-x^k\| + \|x^k-\bar{x}\|)\|\xi^k\|),\end{aligned}$$

where the third equality is by the inclusion $\xi^k \in \ker h'(\bar{x})$, while the last equality is by (4.12). Thus, $\xi^k = o(\|x^{k+1}-x^k\| + \|x^k-\bar{x}\|)$, and from (4.16) we now derive the estimate

$$\|x^{k+1}-\bar{x}\| = o(\|x^{k+1}-\bar{x}\| + \|x^k-\bar{x}\|)$$

as $k \to \infty$. By the same argument as in the proof of Proposition 2.4, the latter implies superlinear convergence rate. □

Remark 4.5. By Remark 2.5, we note that in each of the conditions (4.12) and (4.13), the right-hand side can be replaced by either $o(\|x^{k+1} - x^k\|)$ or $o(\|x^k - \bar{x}\|)$. The conditions modified this way are generally stronger than (4.12) and (4.13). However, if $\{x^k\}$ is assumed to be superlinearly convergent to $\bar{x}$, all these pairs of conditions become equivalent.

The presented proof of Proposition 4.4 is rather elementary. Employing more sophisticated techniques based on the primal error bounds in Sect. 1.3.3, one can actually replace the SOSC (4.5) in this result by the weaker assumption that $\bar{\lambda}$ is a noncritical multiplier. This will be done in Sect. 4.3.1, for the more general problem involving inequality constraints; see Proposition 4.18.

A number of specific versions of perturbed Newton–Lagrange methods are associated with avoiding the computation of the exact Hessian of the Lagrangian, which in many cases is too expensive and sometimes simply impossible. In *quasi-Newton–Lagrange methods*, the iteration system (4.3) is replaced by

$$\begin{aligned} H_k(x - x^k) + (h'(x^k))^{\mathrm{T}}(\lambda - \lambda^k) &= -\frac{\partial L}{\partial x}(x^k,\, \lambda^k), \\ h'(x^k)(x - x^k) &= -h(x^k), \end{aligned} \tag{4.18}$$

where H_k is some quasi-Newton approximation of $\frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)$. The specific understanding of what can be regarded as a "quasi-Newton approximation" will be given by the Dennis–Moré-type condition (4.22) in Theorem 4.6 below.

For instance, taking into account that Hessian is a symmetric matrix, one can use BFGS approximations: for each $k = 0,\, 1,\, \ldots$, set

$$H_{k+1} = H_k + \frac{s^k(s^k)^{\mathrm{T}}}{\langle r^k,\, s^k\rangle} - \frac{(H_k r^k)(H_k r^k)^{\mathrm{T}}}{\langle H_k r^k,\, r^k\rangle}, \tag{4.19}$$

where

$$s^k = x^{k+1} - x^k, \quad r^k = \frac{\partial L}{\partial x}(x^{k+1},\, \lambda^{k+1}) - \frac{\partial L}{\partial x}(x^k,\, \lambda^{k+1})$$

(see Remark 2.31). Some more sophisticated quasi-Newton update rules for the Newton–Lagrange method can be found, e.g., in [208, Sect. 18.3].

As will be seen in Sect. 5.1, in order to construct efficient linesearch globalizations of quasi-Newton–Lagrange methods it is essential to keep the matrices H_k positive definite. According to Proposition 2.30 and Remark 2.31, if H_k is positive definite, then the matrix H_{k+1} generated according to (4.19) will be positive definite if and only if

$$\langle r^k,\, s^k\rangle > 0. \tag{4.20}$$

However, unlike in the unconstrained case when the Wolfe linesearch is employed, the latter inequality is not automatic in the present context.

The widely used fix for this difficulty is the so-called *Powell's correction* of the BFGS update, suggested in [219]. Let $\theta \in (0, 1)$ be a fixed parameter (the generic value is $\theta = 0.2$). Set

$$t_k = \begin{cases} 1 & \text{if } \langle r^k, s^k \rangle \geq \theta \langle H_k s^k, s^k \rangle, \\ \dfrac{(1-\theta)\langle H_k s^k, s^k \rangle}{\langle H_k s^k, s^k \rangle - \langle r^k, s^k \rangle} & \text{otherwise,} \end{cases} \tag{4.21}$$

$$\tilde{r}^k = t_k r^k + (1 - t_k) H_k s^k,$$

and update H_k by formula (4.19) with r^k replaced by $\tilde{r}^k$:

$$H_{k+1} = H_k + \frac{s^k (s^k)^{\mathrm{T}}}{\langle \tilde{r}^k, s^k \rangle} - \frac{(H_k \tilde{r}^k)(H_k \tilde{r}^k)^{\mathrm{T}}}{\langle H_k \tilde{r}^k, \tilde{r}^k \rangle}.$$

Note that $t_k = 1$ corresponds to the unmodified BFGS update (4.19), while $t_k = 0$ corresponds to keeping the current (positive definite) H_k unchanged: $H_{k+1} = H_k$. The choice of t_k according to (4.21) ensures that H_{k+1} stays close enough to H_k, so that it remains positive definite. Indeed, with this choice, if $s^k \neq 0$, then either $t_k = 1$ and

$$\langle \tilde{r}^k, s^k \rangle = \langle r^k, s^k \rangle \geq \theta \langle H_k s^k, s^k \rangle > 0,$$

or

$$\begin{aligned} \langle \tilde{r}^k, s^k \rangle &= \langle t_k r^k + (1 - t_k) H_k s^k, s^k \rangle \\ &= t_k (\langle r_k, s^k \rangle - \langle H_k s^k, s^k \rangle) + \langle H_k s^k, s^k \rangle \\ &= -(1 - \theta) \langle H_k s^k, s^k \rangle + \langle H_k s^k, s^k \rangle \\ &= \theta \langle H_k s^k, s^k \rangle \\ &> 0. \end{aligned}$$

We proceed with the rate of convergence analysis for quasi-Newton–Lagrange methods. Observe that system (4.18) is a special case of (4.9) with

$$\omega_1^k = \left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k) \right) (x^{k+1} - x^k), \quad \omega_2^k = 0.$$

A posteriori results can be derived along the lines of the analysis of quasi-Newton methods in Sect. 2.1.1. In particular, the necessary and sufficient conditions for superlinear rate of primal-dual convergence of the quasi-Newton–Lagrange method can be readily obtained from Theorem 2.9. At the same time, characterization of primal superlinear convergence of the quasi-Newton–Lagrange method readily follows from Proposition 4.4 and Remark 4.5.

Theorem 4.6. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives continuous at $\bar{x}$.*

Let $\bar{x}$ be a stationary point of problem (4.1), and let $\bar{\lambda} \in \mathbf{R}^l$ be a Lagrange multiplier associated with $\bar{x}$. Let $\{H_k\} \subset \mathbf{R}^{n \times n}$ be a sequence of symmetric matrices, let a sequence $\{(x^k, \lambda^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l$ be convergent to $(\bar{x}, \bar{\lambda})$, and assume that (x^{k+1}, λ^{k+1}) satisfies (4.18) for all k large enough.

If the rate of convergence of $\{x^k\}$ is superlinear, then the following condition holds:

$$\pi_{\ker h'(\bar{x})} \left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k) \right) (x^{k+1} - x^k) = o(\|x^{k+1} - x^k\|) \tag{4.22}$$

as $k \to \infty$.

Conversely, if the SOSC (4.5) holds, and condition (4.22) holds as well, then the rate of convergence of $\{x^k\}$ is superlinear.

Again we note that the SOSC (4.5) in this result can actually be replaced by the weaker assumption that $\bar{\lambda}$ is a noncritical multiplier; see Theorem 4.22.

As another example of a method that can be interpreted as a perturbed Newton–Lagrange method, let us mention the following (see [18, Sect. 4.4.2]). Define the set

$$\mathcal{R} = \{x \in \mathbf{R}^n \mid \operatorname{rank} h'(x) = l\}$$

and the mapping $\lambda_\tau : \mathcal{R} \to \mathbf{R}^l$,

$$\lambda_\tau(x) = (h'(x)(h'(x))^{\mathrm{T}})^{-1}(\tau h(x) - h'(x)f'(x)), \tag{4.23}$$

where $\tau \in \mathbf{R}$ is a parameter. Evidently, if a stationary point $\bar{x}$ of problem (4.1) satisfies the regularity condition (4.4), then $\bar{\lambda} = \lambda_\tau(\bar{x})$ is the unique Lagrange multiplier associated with $\bar{x}$. By explicitly differentiating $\lambda_\tau(\cdot)$, one can construct a primal Newton method for the system of equations

$$\frac{\partial L}{\partial x}(x, \lambda_\tau(x)) = 0, \tag{4.24}$$

with respect to $x \in \mathbf{R}^n$. Moreover, if one takes $\tau = 1$, the following perturbed primal Newton method with a cheaper iteration can be designed: for a current $x^k \in \mathbf{R}^n$, the next iterate x^{k+1} is computed as a solution of the linear system

$$\left(E(x^k) + (I - E(x^k)) \frac{\partial^2 L}{\partial x^2}(x^k, \lambda_1(x^k)) \right) (x - x^k) = -\frac{\partial L}{\partial x}(x^k, \lambda_1(x^k)),$$

where $E : \mathbf{R}^n \to \mathbf{R}^{n \times n}$ is given by

$$E(x) = (h'(x))^{\mathrm{T}}(h'(x)(h'(x))^{\mathrm{T}})^{-1}h'(x).$$

Since this method is completely primal, it can be analyzed directly, as a perturbed Newton method for equation (4.24) with $\tau = 1$. (Note that in this case, $(E(\bar{x}) + (I - E(\bar{x}))\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}))$ coincides with the Jacobian of the mapping in the left-hand side of (4.24) at $\bar{x}$.) However, it appears simpler to interpret this primal method as a perturbed (primal-dual) Newton–Lagrange method. Specifically, one can directly verify that for any $x^k \in \mathcal{R}$, if the point x^{k+1} is computed according to the perturbed primal Newton method, then it coincides with the primal part of the point $(x^{k+1},\, \lambda^{k+1})$ computed as a solution of the perturbed Newton–Lagrange iteration system (4.9) with

$$\omega_1^k = \left(\frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda_1(x^k)) - \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\right)(x^{k+1} - x^k), \quad \omega_2^k = 0,$$

whatever is taken as $\lambda^k \in \mathbf{R}^l$. With this observation at hand, this iterative process can be easily analyzed applying the tools of Sect. 2.1.1. In particular, its local superlinear convergence can be derived from Theorem 2.11.

Speaking about practical versions of the Newton–Lagrange method for problem (4.1), we must mention the so-called *reduced Hessian methods*. The main idea is that for each k, instead of the (full) Hessian $\frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)$, these methods employ the *reduced Hessian* $\Xi_k^{\mathrm{T}} \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\Xi_k$, where the columns of $n \times \dim \ker h'(x^k)$-matrix Ξ_k form a basis of the subspace $\ker h'(x^k)$, and this basis depends smoothly on x^k. For details, see [29, Sect. 14.5], [208, Sect. 18.3].

4.1.2 Linearly Constrained Lagrangian Methods

Observe that for any Lagrange multiplier $\bar{\lambda}$ associated with a stationary point $\bar{x}$ of problem (4.1), it holds that $\bar{x}$ is a stationary point of the unconstrained optimization problem

$$\begin{array}{ll} \text{minimize} & L(x,\, \bar{\lambda}) \\ \text{subject to} & x \in \mathbf{R}^n. \end{array} \tag{4.25}$$

Thus, instead of solving the Lagrange optimality system (4.2), one could (in principle) think of searching for a stationary point of problem (4.25). Of course, the multiplier $\bar{\lambda}$ is not known in advance, but it can be approximated in some way using the primal approximation of $\bar{x}$ (see, e.g., (4.23)). The main drawback of this idea is that under reasonable assumptions (e.g., such as the assumptions of Theorem 4.3), one cannot guarantee positive semidefiniteness (and even less so positive definiteness) of the Hessian $\frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})$. This means, in particular, that $\bar{x}$ need not be a local minimizer in problem (4.25) under natural assumptions; see Example 4.7 below. Moreover, note also that the objective function in (4.25) can easily be unbounded below. For those reasons, one should not expect the usual unconstrained optimization algorithms to efficiently and reliably compute a desired stationary point $\bar{x}$ of (4.25), even if $\bar{\lambda}$ were known.

The preceding discussion is illustrated by the following example.

Example 4.7. Consider the problem

$$\begin{array}{ll} \text{minimize} & x^3 \\ \text{subject to} & x+1=0. \end{array}$$

The solution $\bar{x} = -1$ satisfies the regularity condition, the (unique) associated Lagrange multiplier is $\bar{\lambda} = -3$, and the second-order sufficient condition also holds (trivially, as $\ker h'(\bar{x}) = \{0\}$).

However, as $\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}) = -6 < 0$, the point $\bar{x}$ is not a solution of problem (4.25) (since $\bar{x}$ violates the second-order necessary condition for optimality for problem (4.25)).

A remedy to the difficulty just illustrated is the following modification of the standard Lagrangian, called *augmented Lagrangian*: $L_c : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}$,

$$L_c(x, \lambda) = L(x, \lambda) + \frac{c}{2}\|h(x)\|^2, \tag{4.26}$$

where $c > 0$ is the *penalty parameter*. Note that L_c can be thought of as the standard Lagrangian for the (formally equivalent) penalized version of (4.1), namely

$$\begin{array}{ll} \text{minimize} & f(x) + \dfrac{c}{2}\|h(x)\|^2 \\ \text{subject to} & h(x) = 0. \end{array}$$

As is easily seen, L_c can be used in optimality conditions and algorithms for problem (4.1) instead of L, as for any $c \in \mathbf{R}$ it holds that

$$\frac{\partial L_c}{\partial x}(x, \lambda) = \frac{\partial L}{\partial x}(x, \lambda) \quad \forall\, x \in \mathbf{R}^n \text{ such that } h(x) = 0,\ \forall\, \lambda \in \mathbf{R}^l,$$

$$\frac{\partial L_c}{\partial \lambda}(x, \lambda) = \frac{\partial L}{\partial \lambda}(x, \lambda) = h(x) \quad \forall\, x \in \mathbf{R}^n,\ \forall\, \lambda \in \mathbf{R}^l.$$

Thus, the solution set of the equation

$$L_c'(x, \lambda) = 0$$

coincides with the solution set of the Lagrange optimality system (4.2).

At the same time, going back to Example 4.7, observe that

$$L_c(x, \bar{\lambda}) = x^3 - 3(x+1) + \frac{c}{2}(x+1)^2,$$

and $\bar{x} = -1$ is a strict local minimizer of this function for every $c > 6$, since $\frac{\partial^2 L_c}{\partial x^2}(\bar{x}, \bar{\lambda}) = -6 + c > 0$ for any such c.

The following proposition formalizes the observed advantages of the augmented Lagrangian.

Proposition 4.8. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice differentiable at $\bar{x} \in \mathbf{R}^n$. Let $\bar{x}$ be a stationary point of problem (4.1), and let $\bar{\lambda} \in \mathbf{R}^l$ be* *a Lagrange multiplier associated with $\bar{x}$. Assume finally that the SOSC* (4.5) *holds.*

Then for any $c > 0$ large enough, the Hessian

$$\frac{\partial^2 L_c}{\partial x^2}(\bar{x}, \bar{\lambda}) = \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}) + c(h'(\bar{x}))^{\mathrm{T}} h'(\bar{x})$$

is positive definite. Accordingly, the stationary point $\bar{x}$ of problem

$$\begin{array}{ll} \text{minimize} & L_c(x, \bar{\lambda}) \\ \text{subject to} & x \in \mathbf{R}^n \end{array}$$

satisfies the SOSC for unconstrained local minimizers, as stated in Theorem 1.9:

$$\left\langle \frac{\partial L_c}{\partial x}(\bar{x}, \bar{\lambda})\xi, \, \xi \right\rangle > 0 \quad \forall\, \xi \in \mathbf{R}^n \setminus \{0\}.$$

The proof is by direct application of Lemma A.7, where one should take $H = \dfrac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})$, $A = h'(\bar{x})$.

Augmented Lagrangian algorithms are based on sequentially solving, perhaps approximately, subproblems of the form

$$\begin{array}{ll} \text{minimize} & L_c(x, \lambda) \\ \text{subject to} & x \in \mathbf{R}^n, \end{array} \tag{4.27}$$

with dual estimates λ being updated according to certain rules. These algorithms are treated in detail in, e.g., [18] (see also [5] and [76] for the state-of-the-art results on global and local convergence, respectively). Note that augmented Lagrangian algorithms based on (4.27) are certainly not Newton-type methods, as no part of the problem data is being linearized. Moreover, their superlinear convergence can only be expected when the penalty parameter c tends to $+\infty$.

In what follows, we discuss a different but related class of methods, called *linearly constrained Lagrangian* (LCL) *methods*. Even though those methods are not related to the Newtonian class in any obvious way, we show that they can in fact be interpreted within the perturbed Newton–Lagrange framework introduced above. Moreover, this line of analysis leads to sharp convergence results not available previously. The LCL method was originally proposed in [231], and further developed in [96, 205], among other sources. This method is the basis for solving general nonlinear optimization problems in the widely used MINOS software package [206]. Instead of unconstrained subproblems of the form (4.27), subproblems of this method consist in minimizing the augmented Lagrangian subject to linearized constraints (cf. the second equation in (4.3)).

Algorithm 4.9 Choose $(x^0,\, \lambda^0) \in \mathbf{R}^n \times \mathbf{R}^l$, and set $k = 0$.

1. If $L'(x^k,\, \lambda^k) = 0$, stop.
2. Choose $c_k \geq 0$ and compute $x^{k+1} \in \mathbf{R}^n$ and $\eta^k \in \mathbf{R}^l$ as a stationary point and an associated Lagrange multiplier of the problem

$$\begin{array}{ll} \text{minimize} & L_{c_k}(x,\, \lambda^k) \\ \text{subject to} & h(x^k) + h'(x^k)(x - x^k) = 0. \end{array} \tag{4.28}$$

3. Set $\lambda^{k+1} = \lambda^k + \eta^k$.
4. Increase k by 1 and go to step 1.

In the analysis of local convergence below, we assume that $c_k = c$ for all k, which is quite natural in this (local) context, see the discussion in [96]. Then the Lagrange optimality system of subproblem (4.28) has the form

$$\begin{gathered} f'(x) + (h'(x))^{\mathrm{T}}\lambda^k + c(h'(x))^{\mathrm{T}}h(x) + (h'(x^k))^{\mathrm{T}}\eta = 0, \\ h(x^k) + h'(x^k)(x - x^k) = 0. \end{gathered} \tag{4.29}$$

Observe that the last line of system (4.29), associated with the linearized constraints, is exactly the same as in the basic Newton–Lagrange iteration system (4.3). Structural perturbation that defines the LCL method within the perturbed Newton–Lagrange framework is therefore given by the first line of (4.29). In particular, the LCL method with exact solution of subproblems is a special case of the perturbed Newton–Lagrange method (4.9), with the perturbation terms being

$$\begin{aligned} \omega_1^k &= \frac{\partial L}{\partial x}(x^{k+1},\, \lambda^k) - \frac{\partial L}{\partial x}(x^k,\, \lambda^k) - \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)(x^{k+1} - x^k) \\ &\quad + c(h'(x^{k+1}))^{\mathrm{T}}h(x^{k+1}) \\ &= \frac{\partial L}{\partial x}(x^{k+1},\, \lambda^k) - \frac{\partial L}{\partial x}(x^k,\, \lambda^k) - \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)(x^{k+1} - x^k) \\ &\quad + c(h'(x^{k+1}))^{\mathrm{T}}(h(x^{k+1}) - h(x^k) - h'(x^k)(x^{k+1} - x^k)), \end{aligned} \tag{4.30}$$

where the last equality is by the last line of system (4.29), and

$$\omega_2^k = 0.$$

Note that ω_1^k can be written in the form $\omega_1^k = \omega_1((x^k,\, \lambda^k),\, x^{k+1} - x^k)$, where the mapping $\omega_1 : (\mathbf{R}^n \times \mathbf{R}^l) \times \mathbf{R}^n \to \mathbf{R}^n$ is given by

$$\begin{aligned} \omega_1((x,\, \lambda),\, \xi) = {}& \frac{\partial L}{\partial x}(x + \xi,\, \lambda) - \frac{\partial L}{\partial x}(x,\, \lambda) - \frac{\partial^2 L}{\partial x^2}(x,\, \lambda)\xi \\ & + c(h'(x + \xi))^{\mathrm{T}}(h(x + \xi) - h(x) - h'(x)\xi). \end{aligned} \tag{4.31}$$

The proof of the following a priori local convergence result is by verifying the assumptions of and then applying Theorem 2.12, which can be easily done employing the mean-value theorem (see Theorem A.10) and Lemma 4.2. We omit the details, as a more general result allowing inexact solutions of suproblems is given a bit later, see Theorem 4.11.

Theorem 4.10. *Under the assumptions of Theorem 4.3, for any fixed* $c \geq 0$, *there exists* $\delta > 0$ *such that any starting point* $(x^0, \lambda^0) \in \mathbf{R}^n \times \mathbf{R}^l$ *close enough to* $(\bar{x}, \bar{\lambda})$ *uniquely defines the sequence* $\{(x^k, \lambda^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l$ *such that for each* $k = 0, 1, \ldots,$ *the pair* $(x^{k+1}, \lambda^{k+1} - \lambda^k)$ *satisfies system* (4.29) *and also satisfies*

$$\|(x^{k+1} - x^k, \lambda^{k+1} - \lambda^k)\| \leq \delta; \tag{4.32}$$

any such sequence converges to $(\bar{x}, \bar{\lambda})$, *and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the second derivatives of* f *and of* h *are locally Lipschitz-continuous with respect to* $\bar{x}$.

Taking into account that (4.28) is not a quadratic programming problem, one should not expect the subproblems to be solved exactly in a practical implementation. Motivated by this, apart from interpreting the exact LCL method as a perturbed Newton–Lagrange method, we introduce an extra perturbation associated with inexact solution of the LCL subproblems. Specifically, it is natural to consider the *truncated LCL method*, where system (4.29) is replaced by its relaxed version:

$$\begin{gathered} \|f'(x) + (h'(x))^{\mathrm{T}}\lambda^k + c(h'(x))^{\mathrm{T}}h(x) + (h'(x^k))^{\mathrm{T}}\eta\| \leq \varphi(\|L'(x^k, \lambda^k)\|), \\ \|h(x^k) + h'(x^k)(x - x^k)\| \leq \varphi(\|L'(x^k, \lambda^k)\|), \end{gathered} \tag{4.33}$$

where $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$ is some *forcing function* controlling inexactness, with specific conditions on φ to be imposed later (relating this to the terminology for truncated Newton methods for equations, the forcing sequence $\{\theta_k\}$ in Sect. 2.1.1 would be defined by $\theta_k = \varphi(\|L'(x^k, \lambda^k)\|)/\|L'(x^k, \lambda^k)\|$, for each k).

Note that the first line of (4.33) can be re-written as follows:

$$\begin{aligned} \varphi(\|L'(x^k, \lambda^k)\|) &\geq \left\| \frac{\partial L}{\partial x}(x, \lambda^k) + (h'(x^k))^{\mathrm{T}}\eta + c(h'(x))^{\mathrm{T}}h(x) \right\| \\ &= \left\| \frac{\partial L}{\partial x}(x^k, \lambda^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)(x - x^k) + (h'(x^k))^{\mathrm{T}}\eta \right. \\ &\quad + \left(\frac{\partial L}{\partial x}(x, \lambda^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k) - \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)(x - x^k) \right. \\ &\quad \left. \left. + c(h'(x))^{\mathrm{T}}h(x) \right) \right\|. \end{aligned}$$

We can therefore interpret the subproblem of this perturbed iterative process as the system of inclusions

$$\begin{aligned}\frac{\partial L}{\partial x}(x^k, \lambda^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)(x - x^k) + (h'(x^k))^{\mathrm{T}}\eta + \Omega_1((x^k,\, \lambda^k),\, x - x^k) \ni 0,\\ h(x^k) + h'(x^k)(x - x^k) + \Omega_2(x^k,\, \lambda^k) \ni 0,\end{aligned} \tag{4.34}$$

with the total perturbation multifunctions (combining perturbations with respect to the Newton–Lagrange method with that with respect to the exact LCL method) Ω_1 from $(\mathbf{R}^n \times \mathbf{R}^l) \times \mathbf{R}^n$ to the subsets of $\mathbf{R}^n$, and Ω_2 from $\mathbf{R}^n \times \mathbf{R}^l$ to the subsets of $\mathbf{R}^l$, given by

$$\Omega_1((x,\, \lambda),\, \xi) = \tilde{\omega}_1((x,\, \lambda),\, \xi) + \{\theta_1 \in \mathbf{R}^n \mid \|\theta_1\| \le \varphi(\|L'(x^k,\, \lambda^k)\|)\}, \tag{4.35}$$

$$\Omega_2(x,\, \lambda) = \{\theta_2 \in \mathbf{R}^l \mid \|\theta_2\| \le \varphi(\|L'(x^k,\, \lambda^k)\|)\}, \tag{4.36}$$

where the structural LCL perturbation mapping $\tilde{\omega}_1 : (\mathbf{R}^n \times \mathbf{R}^l) \times \mathbf{R}^n \to \mathbf{R}^n$ is given by

$$\begin{aligned}\tilde{\omega}_1((x,\, \lambda),\, \xi) = \frac{\partial L}{\partial x}(x + \xi,\, \lambda) - \frac{\partial L}{\partial x}(x,\, \lambda) - \frac{\partial^2 L}{\partial x^2}(x,\, \lambda)\xi\\ + c(h'(x + \xi))^{\mathrm{T}} h(x + \xi).\end{aligned} \tag{4.37}$$

Separating the perturbation into the single-valued part and the set-valued part is instructive, because, as explained above, the two parts correspond to inexactness of different kind: structural, defining the method within the general framework, and that associated with inexact solution of subproblems of the method.

The local convergence result for the (perturbed) LCL method now follows from Theorem 2.13.

Theorem 4.11. *Under the assumptions of Theorem 4.3, for any fixed* $c \ge 0$ *and any function* $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$ *such that* $\varphi(t) = o(t)$ *as* $t \to 0$*, there exists* $\delta > 0$ *such that for any starting point* $(x^0,\, \lambda^0) \in \mathbf{R}^n \times \mathbf{R}^l$ *close enough to* $(\bar{x},\, \bar{\lambda})$*, there exists a sequence* $\{(x^k,\, \lambda^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l$ *such that for each* $k = 0,\, 1,\, \ldots,$ *the pair* $(x^{k+1},\, \lambda^{k+1} - \lambda^k)$ *satisfies* (4.33)*, and also satisfies* (4.32)*; any such sequence converges to* $(\bar{x},\, \bar{\lambda})$*, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the second derivatives of* f *and of* h *are locally Lipschitz-continuous with respect to* $\bar{x}$*, and provided* $\varphi(t) = O(t^2)$ *as* $t \to 0$*.*

Proof. In order to apply Theorem 2.13, we need to establish the following properties:

(i) For each $(x,\, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$ close enough to $(\bar{x},\, \bar{\lambda})$, the system of inclusions

$$\begin{aligned}\frac{\partial L}{\partial x}(x,\, \lambda) + \frac{\partial^2 L}{\partial x^2}(x,\, \lambda)\xi + (h'(x))^{\mathrm{T}}\eta + \Omega_1((x,\, \lambda),\, \xi) \ni 0,\\ h(x) + h'(x)\xi + \Omega_2(x,\, \lambda) \ni 0,\end{aligned} \tag{4.38}$$

has a solution $(\xi(x,\, \lambda),\, \eta(x,\, \lambda))$ tending to zero as $(x,\, \lambda) \to (\bar{x},\, \bar{\lambda})$.

(ii) The estimates

$$\left\| \frac{\partial L}{\partial x}(x, \lambda) + \frac{\partial^2 L}{\partial x^2}(x, \lambda)\xi + (h'(x))^{\mathrm{T}}\eta \right\| = o(\|(\xi, \eta)\| + \|(x - \bar{x}, \lambda - \bar{\lambda})\|),$$

$$\|h(x) + h'(x)\xi\| = o(\|(\xi, \eta)\| + \|(x - \bar{x}, \lambda - \bar{\lambda})\|)$$

hold as $(x, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$ tends to $(\bar{x}, \bar{\lambda})$, and $(\xi, \eta) \in \mathbf{R}^n \times \mathbf{R}^l$ tends to $(0, 0)$, for (x, λ) and (ξ, η) satisfying (4.38).

From (4.35) and (4.36), it follows that it is sufficient to verify property (i) with (4.38) replaced by the system of equations

$$\begin{gathered} \frac{\partial L}{\partial x}(x, \lambda) + \frac{\partial^2 L}{\partial x^2}(x, \lambda)\xi + (h'(x))^{\mathrm{T}}\eta + \tilde{\omega}_1((x, \lambda), \xi) = 0, \\ h(x) + h'(x)\xi = 0. \end{gathered} \tag{4.39}$$

According to the relations (4.31), (4.37) and the last equation in (4.39), the term $\tilde{\omega}_1((x, \lambda), \xi)$ in the first equation can be equivalently replaced by $\omega_1((x, \lambda), \xi)$. Once this is done, the needed property can be easily derived from the implicit function theorem (Theorem 1.22) and Lemma 4.2, employing the argument similar to that in the proof of Theorem 2.12. Applicability of Theorem 1.22 can be verified by the mean-value theorem (see Theorem A.10) and Lemma 4.2. Actually, this verification is precisely what is needed in order to establish Theorem 4.10.

Assume now that $(x, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$ and $(\xi, \eta) \in \mathbf{R}^n \times \mathbf{R}^l$ satisfy (4.38). Since $\varphi(t) = o(t)$, we have that

$$\begin{aligned} \varphi(\|L'(x, \lambda)\|) &= o(\|L'(x, \lambda)\|) \\ &= o(\|L'(x, \lambda) - L'(\bar{x}, \bar{\lambda})\|) \\ &= o(\|(x - \bar{x}, \lambda - \bar{\lambda})\|) \end{aligned} \tag{4.40}$$

as $(x, \lambda) \to (\bar{x}, \bar{\lambda})$. According to (4.36) and the second inclusion in (4.38),

$$\|h(x) + h'(x)\xi\| \leq \varphi(\|L'(x, \lambda)\|).$$

Thus, by (4.37), using the mean-value theorem (see Theorem A.10, (a)), we obtain the estimate

$$\begin{aligned} \|\tilde{\omega}_1((x, \lambda), \xi)\| \leq & \left\| \frac{\partial L}{\partial x}(x + \xi, \lambda) - \frac{\partial L}{\partial x}(x, \lambda) - \frac{\partial^2 L}{\partial x^2}(x, \lambda)\xi \right\| \\ & + c\|(h'(x + \xi))^{\mathrm{T}}(h(x + \xi) - h(x) - h'(x)\xi)\| \\ & + O(\varphi(\|L'(x, \lambda)\|)) \\ = & \; o(\|\xi\|) + O(\varphi(\|L'(x, \lambda)\|)) \end{aligned}$$

as $(x, \lambda) \to (\bar{x}, \bar{\lambda})$ and $\xi \to 0$. According to (4.35), (4.36), and (4.40), the last estimate implies that for any elements $\omega_1 \in \Omega_1((x, \lambda), \xi)$ and $\omega_2 \in \Omega_2(x, \lambda)$, it holds that

$$\begin{aligned}\|\omega_1\| &\leq \|\tilde{\omega}_1((x,\, \lambda),\, \xi)\| + \varphi(\|L'(x,\, \lambda)\|)\\ &= o(\|\xi\|) + o(\|(x-\bar{x},\, \lambda-\bar{\lambda})\|),\end{aligned} \tag{4.41}$$

and

$$\|\omega_2\| \leq \varphi(\|L'(x,\, \lambda)\|) = o(\|(x-\bar{x},\, \lambda-\bar{\lambda})\|). \tag{4.42}$$

This proves property (ii).

Moreover, assuming that the second derivatives of f and h are locally Lipschitz-continuous with respect to $\bar{x}$, and that $\varphi(t) = O(t^2)$, the right-hand side of (4.41) can be replaced by $O(\|\xi\|^2 + \|\xi\|\|x-\bar{x}\|) + O(\|(x-\bar{x},\, \lambda-\bar{\lambda})\|^2)$, and the right-hand side of (4.42) by $O(\|(x-\bar{x},\, \lambda-\bar{\lambda})\|^2)$.

The assertions now follow applying Theorem 2.13. □

We complete this section with an a posteriori result regarding primal superlinear convergence of the basic LCL method, which can be obtained as an immediate corollary of Proposition 4.4, without any additional assumptions. Indeed, by the mean-value theorem (see Theorem A.10, (a)) it follows that ω_1^k given by (4.30) satisfies $\omega_1^k = o(\|x^{k+1} - x^k\|)$ as $k \to \infty$, provided $\{x^k\}$ converges to $\bar{x}$, while $\omega_2^k = 0$. Thus, (4.12) and (4.13) are readily satisfied, and Proposition 4.4 implies the following.

Theorem 4.12. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (4.1), and let $\bar{\lambda} \in \mathbf{R}^l$ be a Lagrange multiplier associated with $\bar{x}$, satisfying the SOSC (4.5). Let $c \geq 0$ be fixed, and let an iterative sequence $\{(x^k,\, \lambda^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l$ generated by Algorithm 4.9 with $c_k = c$ for all k large enough be convergent to $(\bar{x},\, \bar{\lambda})$.*

Then the rate of convergence of $\{x^k\}$ is superlinear.

Any reasonable counterpart of this result for the truncated LCL method can hardly be valid, because of the primal-dual nature of the estimate (4.40).

As was the case for Proposition 4.4 and Theorem 4.6, we note that the SOSC (4.5) can actually be relaxed in Theorem 4.12 as well; see Theorem 4.29 below.

4.1.3 Active-Set Methods for Inequality Constraints

As a direct extension of the development above, we complete this section by a brief discussion of active-set methods for problems with additional inequality constraints. A natural idea is to combine the procedures for identification of active inequalities discussed in Sect. 3.4.2, with Newton-type methods for equality-constrained problems analyzed earlier in this section.

As discussed in Sect. 3.4, active-set methods are often based on heuristic identification procedures. Among many others, some examples of such developments for optimization problems can be found in [22, 34, 36, 38, 41, 90, 107, 108, 200]. Our focus here is on fully theoretically justifiable identification procedures relying on error bounds. Such algorithmic developments can also be found in the literature; e.g., [273].

Consider the mathematical programming problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0,\ g(x) \leq 0, \end{array} \tag{4.43}$$

where the objective function $f : \mathbf{R}^n \to \mathbf{R}$ and the constraints mappings $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are twice differentiable.

Let $L : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}$ be the Lagrangian of problem (4.43):

$$L(x,\, \lambda,\, \mu) = f(x) + \langle \lambda,\, h(x)\rangle + \langle \mu,\, g(x)\rangle.$$

Recall that stationary points of problem (4.43) and the associated Lagrange multipliers are characterized by the KKT system

$$\frac{\partial L}{\partial x}(x,\, \lambda,\, \mu) = 0, \quad h(x) = 0, \quad \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu,\, g(x)\rangle = 0. \tag{4.44}$$

Define the residual mapping $\Psi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ of system (4.44) as follows:

$$\Psi(x,\, \lambda,\, \mu) = \left(\frac{\partial L}{\partial x}(x,\, \lambda,\, \mu),\, h(x),\, \psi(\mu,\, -g(x)) \right),$$

where $\psi : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ is either the natural residual function

$$\psi(a,\, b) = \min\{a,\, b\},$$

or the Fischer–Burmeister function

$$\psi(a,\, b) = a + b - \sqrt{a^2 + b^2},$$

applied componentwise.

Define the function $\rho : \mathbf{R}_+ \to \mathbf{R}$ by

$$\rho(t) = \begin{cases} \bar{\rho} & \text{if } t \geq \bar{t}, \\ -1/\log t & \text{if } t \in (0,\, \bar{t}), \\ 0 & \text{if } t = 0, \end{cases}$$

with some fixed $\bar{t} \in (0, 1)$ and $\bar{\rho} > 0$, and for any $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ define the index set

$$A(x,\, \lambda,\, \mu) = \{i = 1,\, \ldots,\, m \mid -g_i(x) \leq \rho(\|\Psi(x,\, \lambda,\, \mu)\|)\}. \tag{4.45}$$

According to Proposition 3.58, if $\bar{x} \in \mathbf{R}^n$ is a stationary point of problem (4.43), and $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ is a noncritical Lagrange multiplier associated with $\bar{x}$, then

$$A(x, \lambda, \mu) = A(\bar{x})$$

for all $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$. Moreover, according to Remark 3.59, this identification result remains valid with the function ρ defined as $\rho(t) = t^\tau$ with any fixed $\tau \in (0, 1)$.

Once the index set $A = A(\bar{x})$ is identified, we can replace problem (4.43) by the equality-constrained problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \; g_A(x) = 0. \end{array} \tag{4.46}$$

This problem can be solved by the Newton-type methods discussed in Sects. 4.1.1 and 4.1.2. In particular, according to Theorem 4.3, the local superlinear convergence of the Newton–Lagrange method for problem (4.46) can be established under the LICQ and the second-order condition of the form

$$\left\langle \frac{\partial^2 L}{\partial x}(\bar{x}, \bar{\lambda})\xi, \xi \right\rangle > 0 \quad \forall\, \xi \in \ker h'(\bar{x}) \cap \ker g'_{A(\bar{x})}(\bar{x}) \setminus \{0\}. \tag{4.47}$$

Moreover, according to Theorem 4.10, a similar result is valid for the LCL method.

Putting all the ingredients together, *active-set Newton–Lagrange method* for problem (4.43) can be formalized as the following iterative procedure. For a given starting point $(x^0, \lambda^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ define the index set $A = A(x^0, \lambda^0, \mu^0)$ according to (4.45), and start the Newton–Lagrange method (Algorithm 4.1) for problem (4.46) from the starting point $(x^0, \lambda^0, \mu^0_A)$, setting $\mu^k_{\{1, \ldots, m\} \setminus A} = 0$ for all k. The *active-set LCL method* for problem (4.43) is the same process but with the Newton–Lagrange method for problem (4.46) replaced by the LCL method (Algorithm 4.9).

Condition (4.47) is evidently implied by the standard SOSC

$$\left\langle \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{4.48}$$

where

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0, \; g'_{A(\bar{x})}(\bar{x})\xi \leq 0, \; \langle f'(\bar{x}), \xi \rangle \leq 0\}$$

is the critical cone of problem (4.43) at $\bar{x}$. Moreover, according to the discussion in Sect. 1.3.3, the SOSC (4.48) implies that $(\bar{\lambda}, \bar{\mu})$ is a noncritical Lagrange multiplier. Therefore, local superlinear convergence of both active-set Newton–Lagrange method and active-set LCL method is justified under the LICQ and the SOSC (4.48).

We shall get back to active-set methods in Chap. 7, with the focus on those cases when the LICQ does not hold and, more generally, the set of Lagrange multipliers associated with the stationary point in question need not be a singleton.

4.2 Sequential Quadratic Programming

In agreement with the name, SQP methods consist of sequentially solving quadratic programming (QP) approximations of the given optimization problem. An appropriately constructed quadratic programming problem turns out to be an adequate local approximation of the original problem. At the same time, QPs are relatively easy to solve: there are well-developed special algorithms for this task (including some with finite termination); see, e.g., [208, Chap. 16]. It is accepted that SQP methods originated from [268]; see [23] for a survey and historical information.

Sequential quadratic programming methods can be considered as an extension of the fundamental idea of the Newton method to constrained optimization. For instance, as will be seen below, in the case of a purely equality-constrained problem the basic SQP algorithm coincides with the Newton–Lagrange method. The idea naturally extends to equality and inequality-constrained problems, in which case, however, SQP can no longer be viewed as a Newton method for some system of equations. Instead, it turns out to be an application of the Josephy-Newton method to the generalized equation (GE) representing the optimality conditions of the problem.

At present, SQP methods (supplied with appropriate globalization techniques; see Chap. 6) certainly belong to the group of the most efficient general-purpose optimization algorithms.

It is instructive to start with the equality-constrained problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \end{array} \tag{4.49}$$

where the objective function $f : \mathbf{R}^n \to \mathbf{R}$ and the constraints mapping $h : \mathbf{R}^n \to \mathbf{R}^l$ are twice differentiable.

Near the current approximation $x^k \in \mathbf{R}^n$ to a stationary point of (4.49), this problem can be approximated by the following quadratic programming problem with linearized constrains:

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, x - x^k\rangle + \dfrac{1}{2}\langle H_k(x - x^k),\, x - x^k\rangle \\ \text{subject to} & h(x^k) + h'(x^k)(x - x^k) = 0, \end{array} \tag{4.50}$$

where $H_k \in \mathbf{R}^{n\times n}$ is some symmetric matrix.

At first sight, it may seem that a natural choice of H_k is the Hessian of the objective function: $H_k = f''(x^k)$, in which case the objective function of (4.50) is the quadratic approximation of the objective function f of problem (4.49). However, as will be demonstrated below, this intuitively reasonable choice is actually inadequate: along with the "second-order information" about the objective function, a good approximation of the original problem must also take into account information about the curvature of the constraints, which cannot be transmitted by their linear approximation. One possibility is to employ the quadratic approximation of the constraints rather than the linear. Indeed, this approach (to be discussed in Sect. 4.3.5) has been attracting some recent attention. However, the corresponding quadratically constrained quadratic subproblems are not QPs anymore, and their efficient numerical solution is more of an issue (apart for some special cases, such as the convex one); see the discussion in Sect. 4.3.5.

Fortunately, the "second-order information" about the constraints can be alternatively passed not to the constraints but to the objective function of subproblem (4.50) by choosing

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k), \tag{4.51}$$

where $\lambda^k \in \mathbf{R}^l$ is some approximation of a Lagrange multiplier associated with a stationary point approximated by x^k, and $L : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}$ is the Lagrangian of problem (4.49):

$$L(x,\, \lambda) = f(x) + \langle \lambda,\, h(x)\rangle.$$

Taking into account the constraints of (4.50), the objective function of this problem is essentially the quadratic approximation of the function $L(\cdot,\, \lambda^k)$ near the current point x^k. More precisely, if we replace the objective function of problem (4.50) (with H_k defined in (4.51)) by the pure quadratic approximation of $L(\cdot,\, \lambda^k)$, stationary points of the subproblem modified in this way will be the same, while the associated multipliers will be shifted by $-\lambda^k$.

The choice of H_k suggested in (4.51) can be formally justified as follows. The stationary points and Lagrange multipliers of problem (4.50) are characterized by the corresponding Lagrange system

$$\begin{gathered} f'(x^k) + H_k(x - x^k) + (h'(x^k))^{\mathrm{T}}\lambda = 0, \\ h(x^k) + h'(x^k)(x - x^k) = 0 \end{gathered} \tag{4.52}$$

with respect to $(x,\, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$. On the other hand, the iteration system of the Newton–Lagrange method for problem (4.49) has the form (see (4.3))

$$\begin{gathered} \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)(x - x^k) + (h'(x^k))^{\mathrm{T}}(\lambda - \lambda^k) = -\frac{\partial L}{\partial x}(x^k,\, \lambda^k), \\ h'(x^k)(x - x^k) = -h(x^k). \end{gathered}$$

If H_k is chosen according to (4.51), the latter system coincides with (4.52), and hence, the iteration defined by subproblem (4.50) is exactly the same as the iteration of the Newton–Lagrange method studied in Sect. 4.1.1. The two methods are therefore "the same." In this sense, perhaps the main role of SQP for equality-constrained problems, as described above, is that it naturally leads to the idea of what SQP might be for equality and inequality-constrained problems, to be introduced below.

Given the equivalence with the Newton–Lagrange method, it follows that the convergence results in Sect. 4.1.1 remain valid here. In particular, SQP for equality-constrained problems has local primal-dual superlinear convergence provided the regularity condition

$$\operatorname{rank} h'(\bar{x}) = l \tag{4.53}$$

and the SOSC

$$\left\langle \frac{\partial^2 L}{\partial x}(\bar{x},\, \bar{\lambda})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in \ker h'(\bar{x}) \setminus \{0\} \tag{4.54}$$

hold at the solution $\bar{x}$ of problem (4.49) with the associated Lagrange multiplier $\bar{\lambda}$.

The interpretation of the Newton–Lagrange method via the optimization subproblems (4.50) is of the same nature as the optimization-based interpretation of the Newton method for unconstrained optimization given in Sect. 2.1.2. Such interpretations are eventually important for globalization of convergence (see Sects. 2.3 and 6.3).

We complete our discussion of the equality-constrained case by observing that it sometimes can make sense to replace the Hessian of the Lagrangian in (4.51) by the approximate Hessian of the augmented Lagrangian function $L_c : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}$,

$$L_c(x,\, \lambda) = L(x,\, \lambda) + \frac{c}{2}\|h(x)\|^2, \tag{4.55}$$

with some value of the penalty parameter $c = c_k > 0$. Specifically, one can take

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k) + c_k (h'(x^k))^{\mathrm{T}} h'(x^k). \tag{4.56}$$

According to Proposition 4.8, if the SOSC (4.54) holds, then the matrix H_k defined in (4.56) is positive definite for all $(x^k,\, \lambda^k)$ close enough to $(\bar{x},\, \bar{\lambda})$ and all $c_k > 0$ large enough, even when the regularity condition (4.53) is not satisfied. This is clearly a desirable property, as positive definiteness of H_k implies that the subproblem (4.50) has the unique (global) solution (which is its unique stationary point) whenever the feasible set of this subproblem is nonempty. Moreover, matrices H_k defined according to (4.56) readily satisfy the Dennis–Moré-type condition (4.22) in Theorem 4.6, at least when $\{c_k\}$ is bounded, and hence, this modification of the Hessian of the Lagrangian would not interfere with high convergence rate.

Note that in the case of this modification, the rule for updating the dual estimates in Algorithm 4.1 must be modified accordingly. Specifically, for all $k = 0, 1, \ldots$, one should take $\lambda^{k+1} = y^{k+1} - c_k h(x^k)$, where y^{k+1} is a Lagrange multiplier associated with the stationary point x^{k+1} of problem (4.50) with H_k defined in (4.56).

We now proceed to the general case of the equality and inequality-constrained problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0,\ g(x) \leq 0, \end{array} \tag{4.57}$$

where the objective function $f : \mathbf{R}^n \to \mathbf{R}$ and the constraints mappings $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are twice differentiable.

Recall the SQP subproblem (4.50) for the equality-constrained case. In the setting under consideration, for the current iterate $x^k \in \mathbf{R}^n$, a natural counterpart of (4.50) is clearly the following:

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, x - x^k \rangle + \dfrac{1}{2}\langle H_k(x - x^k),\, x - x^k \rangle \\ \text{subject to} & h(x^k) + h'(x^k)(x - x^k) = 0,\ g(x^k) + g'(x^k)(x - x^k) \leq 0, \end{array} \tag{4.58}$$

where $H_k \in \mathbf{R}^{n \times n}$ is some symmetric matrix.

Recall that stationary points of problem (4.57) and the associated Lagrange multipliers are characterized by the KKT optimality system

$$\frac{\partial L}{\partial x}(x,\, \lambda,\, \mu) = 0, \quad h(x) = 0, \quad \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu,\, g(x) \rangle = 0, \tag{4.59}$$

where $L : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}$ is the Lagrangian of problem (4.57):

$$L(x,\, \lambda,\, \mu) = f(x) + \langle \lambda,\, h(x) \rangle + \langle \mu,\, g(x) \rangle.$$

The generic *sequential quadratic programming* (SQP) *method* for problem (4.57) is given next.

Algorithm 4.13 Choose $(x^0,\, \lambda^0,\, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and set $k = 0$.

1. If $(x^k,\, \lambda^k,\, \mu^k)$ satisfies the KKT system (4.59), stop.
2. Choose a symmetric matrix $H_k \in \mathbf{R}^{n \times n}$ and compute $x^{k+1} \in \mathbf{R}^n$ as a stationary point of problem (4.58), and $(\lambda^{k+1},\, \mu^{k+1}) \in \mathbf{R}^l \times \mathbf{R}^m$ as an associated Lagrange multiplier.
3. Increase k by 1 and go to step 1.

Regarding possible choices of H_k, the previous discussion of the equality-constrained case suggests that

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k) \tag{4.60}$$

should lead to high convergence rate. Algorithm 4.13 with this choice of H_k will be referred to as the basic SQP method.

Recall again that, under natural assumptions, one cannot expect the matrix H_k defined by (4.60) to be positive definite, even locally. Therefore, the existence of solutions (and stationary points) of the subproblem (4.58) with this matrix is not automatic, even when the feasible set of this subproblem is nonempty. Moreover, even if a stationary point of the subproblem exists, it by no means has to be unique. These issues must be carefully treated within the a priori analysis of Algorithm 4.13, which we present next.

Following [26], we shall derive the sharpest currently known local convergence result for the basic SQP method from Theorem 3.5 on the convergence of the Josephy–Newton method for generalized equations (GE). We assume only the SMFCQ and the standard SOSC at a stationary point $\bar{x}$ of problem (4.57) with the associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$. In particular, the LICQ and strict complementarity are not required, which are common assumptions used in the standard local convergence result for SQP that can be found in most textbooks (see, e.g., [18, Sect. 4.4.3], [208, Theorem 18.1]); the classical result using the LICQ and strict complementarity was originally established in [232].

Recall that the KKT system (4.59) can be written as the GE

$$\Phi(u) + N(u) \ni 0, \tag{4.61}$$

with the mapping $\Phi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ given by

$$\Phi(u) = \left(\frac{\partial L}{\partial x}(x, \lambda, \mu), \, h(x), \, -g(x) \right), \tag{4.62}$$

and with

$$N(\cdot) = N_Q(\cdot), \quad Q = \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m_+, \tag{4.63}$$

where $u = (x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$. Furthermore, the KKT system of the subproblem (4.58) has the form

$$\begin{gathered} f'(x^k) + H_k(x - x^k) + (h'(x^k))^{\mathrm{T}}\lambda + (g'(x^k))^{\mathrm{T}}\mu = 0, \\ h(x^k) + h'(x^k)(x - x^k) = 0, \\ \mu \geq 0, \quad g(x^k) + g'(x^k)(x - x^k) \leq 0, \quad \langle \mu, \, g(x^k) + g'(x^k)(x - x^k) \rangle = 0. \end{gathered} \tag{4.64}$$

With H_k defined according to (4.60), the latter system is evidently equivalent to the GE

$$\Phi(u^k) + \Phi'(u^k)(u - u^k) + N(u) \ni 0,$$

where $u^k = (x^k, \lambda^k, \mu^k)$. But this GE is precisely the iteration system of the Josephy–Newton method (see Sect. 3.1). Therefore, to apply Theorem 3.2 and claim convergence, it remains to figure out conditions for the semistability and hemistability of the solution $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ of the GE (4.61) with Φ and N defined according to (4.62) and (4.63), respectively.

It was already established in Proposition 1.37 that the semistability of $\bar{u}$ is implied by the combination the SMFCQ and the SOSC

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{4.65}$$

where

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A(\bar{x})}(\bar{x})\xi \leq 0,\ \langle f'(\bar{x}),\, \xi\rangle \leq 0\} \tag{4.66}$$

is the critical cone of problem (4.57) at $\bar{x}$. Moreover, the semistability of $\bar{u}$ is in fact equivalent to the combination of the SMFCQ and the SOSC when $\bar{x}$ is a local solution of problem (4.57), and according to Proposition 3.37, in the latter case the hemistability of $\bar{x}$ is implied by semistability.

Combining now Theorem 3.2 with Proposition 3.37, we obtain the local convergence and rate of convergence result for the basic SQP algorithm.

Theorem 4.14. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of a point $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a local solution of problem (4.57), satisfying the SMFCQ and the SOSC (4.65) for the associated Lagrange multiplier $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$.*

Then there exists a constant $\delta > 0$ such that for any starting point $(x^0,\, \lambda^0,\, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, there exists a sequence $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ such that for each $k = 0,\, 1,\, \ldots,$ the point x^{k+1} is a stationary point of problem (4.58) with H_k chosen according to (4.60), and $(\lambda^{k+1},\, \mu^{k+1})$ is an associated Lagrange multiplier satisfying

$$\|(x^{k+1} - x^k,\, \lambda^{k+1} - \lambda^k,\, \mu^{k+1} - \mu^k)\| \leq \delta; \tag{4.67}$$

any such sequence converges to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the second derivatives of f, h, and g are locally Lipschitz-continuous with respect to $\bar{x}$.

Results regarding superlinear primal convergence rate of SQP will be derived in Sect. 4.3.1 in a more general context.

We complete our discussion of the basic SQP method by the following observations. Let the primal-dual iterative sequence $\{(x^k,\, \lambda^k,\, \mu^k)\}$ be generated by Algorithm 4.13, and suppose that the primal sequence $\{x^k\}$ converges to a stationary point $\bar{x}$ of problem (4.57). For each k, define the set of indices of inequality constraints of the SQP subproblems (4.58) active at x^{k+1}:

$$A_k = \{i = 1,\, \ldots,\, m \mid g_i(x^k) + \langle g_i'(x^k),\, x^{k+1} - x^k\rangle = 0\}.$$

If the dual sequence $\{(\lambda^k,\, \mu^k)\}$ converges to a multiplier $(\bar{\lambda},\, \bar{\mu})$ associated with $\bar{x}$ and satisfying the strict complementarity condition, i.e.,

$$\bar{\mu}_{A(\bar{x})} > 0, \tag{4.68}$$

then $A_k = A(\bar{x})$ for all k large enough. Indeed, for large k, taking into account (4.68), we obtain that

$$\mu^{k+1}_{A(\bar{x})} > 0, \quad g_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}(x^k) + g'_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}(x^k)(x^{k+1} - x^k) < 0, \tag{4.69}$$

and since x^{k+1} satisfies (the last line of) (4.64), (4.69) implies that

$$\mu^{k+1}_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})} = 0, \quad g_{A(\bar{x})}(x^k) + g'_{A(\bar{x})}(x^k)(x^{k+1} - x^k) = 0, \tag{4.70}$$

and the equality $A_k = A(\bar{x})$ follows from the last relations in (4.69) and (4.70). Thus, in the case of convergence to a stationary point and multiplier satisfying strict complementarity, active constraints are correctly identified by the SQP algorithm after a finite number of steps.

In general, according to the last line in (4.64) and by the continuity considerations, the inclusions

$$A_+(\bar{x},\, \bar{\mu}) \subset A_k \subset A(\bar{x})$$

always hold for all k large enough. However, the equality $A_k = A(\bar{x})$ is of course not automatic if the strict complementarity condition (4.68) does not hold for the limiting multiplier, or even less so, if the dual sequence $\{(\lambda^k,\, \mu^k)\}$ does not converge. The following is [29, Example 15.3].

Example 4.15. Let $n = m = 1$, $l = 0$, $f(x) = x^2 + x^4$, $g(x) = x$. Then $\bar{x} = 0$ is a solution of problem (4.57), satisfying the LICQ and the SOSC (4.65) with the unique associated multiplier $\bar{\mu} = 0$. The unique constraint is active at $\bar{x}$, and $\bar{\mu}$ violates the strict complementarity condition (4.68).

The SQP subproblem (4.58) with H_k defined according to (4.51) gives the following:

$$\begin{array}{ll} \text{minimize} & 2x^k(1 + 2x^k)(x - x^k) + (1 + 6(x^k)^2)(x - x^k)^2 \\ \text{subject to} & x \le 0. \end{array}$$

If for some k it holds that $x^k < 0$, then this subproblem has the unique solution $x^{k+1} = 4(x^k)^3/(1 + 6(x^k)^2) < 0$, and hence, $A_k = \emptyset$ for all the subsequent values of the iteration index k.

However, if the dual sequence converges to some $(\bar{\lambda},\, \bar{\mu})$, it is still not unnatural to expect that for k sufficiently large, the set A_k will remain unchanged: $A_k = A$, with some fixed set $A \subset \{1,\, \ldots,\, m\}$. This is true in Example 4.15, and generally, this behavior is quite common in practice. In this case, from (4.64) we obtain the equalities

$$
\begin{aligned}
f'(x^k) + H_k(x^{k+1} - x^k) + \left(h'(x^k)\right)^{\mathrm{T}} \lambda^{k+1} + \left(g'_A(x^k)\right)^{\mathrm{T}} \mu_A^{k+1} &= 0,\\
h(x^k) + h'(x^k)(x^{k+1} - x^k) &= 0,\\
g_A(x^k) + g'_A(x^k)(x^{k+1} - x^k) &= 0,\\
\mu^k_{\{1,\,\ldots,\,m\}\setminus A} &= 0
\end{aligned}
\tag{4.71}
$$

for all k large enough. (We suggest to compare this system with (3.77) and to recall the discussion accompanying the latter.) Note that if H_k is defined according to (4.60), these equalities imply that $(x^{k+1},\, \lambda^{k+1},\, \mu_A^{k+1})$ can be regarded as the result of the Newton–Lagrange iteration from the point $(x^k,\, \lambda^k,\, \mu_A^k)$ for the equality-constrained problem

$$
\begin{array}{ll}
\text{minimize} & f(x)\\
\text{subject to} & h(x) = 0,\ g_A(x) = 0.
\end{array}
\tag{4.72}
$$

By passing onto the limit in (4.71), we conclude that $A \subset A(\bar{x})$, $\bar{x}$ is a stationary point of (4.72), and $(\bar{\lambda},\, \bar{\mu}_A)$ is an associated Lagrange multiplier. Thus, in this case, the analysis of the basic SQP method for the original problem (4.57) near $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ reduces to the analysis of the Newton–Lagrange method for the equality-constrained problem (4.72) near its stationary point $\bar{x}$ and the associated multiplier $(\bar{\lambda},\, \bar{\mu}_A)$.

In general, however, it is clear that stabilization of active constraints of SQP subproblems need not happen, and thus the SQP algorithm cannot be regarded as the Newton–Lagrange method for some equality-constrained problem.

4.3 Analysis of Algorithms via Perturbed Sequential Quadratic Programming Framework

In this section, we introduce a perturbed version of the SQP scheme, with motivation to build a general framework suitable for analyzing local convergence properties of a number of useful SQP modifications, as well as some other algorithms that do not belong to the SQP class but whose iterations can be related to it a posteriori.

Consider the equality and inequality-constrained problem

$$
\begin{array}{ll}
\text{minimize} & f(x)\\
\text{subject to} & h(x) = 0,\ g(x) \le 0,
\end{array}
\tag{4.73}
$$

where the objective function $f : \mathbf{R}^n \to \mathbf{R}$ and the constraints mappings $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are twice differentiable. Recall that stationary points of problem (4.73) and the associated Lagrange multipliers are characterized by the KKT optimality system

$$\frac{\partial L}{\partial x}(x,\,\lambda,\,\mu)=0,\quad h(x)=0,\quad \mu\geq 0,\quad g(x)\leq 0,\quad \langle\mu,\,g(x)\rangle=0,\tag{4.74}$$

where $L:\mathbf{R}^n\times\mathbf{R}^l\times\mathbf{R}^m\to\mathbf{R}$ is the Lagrangian of problem (4.73):

$$L(x,\,\lambda,\,\mu)=f(x)+\langle\lambda,\,h(x)\rangle+\langle\mu,\,g(x)\rangle.$$

4.3.1 Perturbed Sequential Quadratic Programming Framework

Recall that for the current iterate $x^k\in\mathbf{R}^n$, the generic SQP subproblem has the form

$$\begin{array}{ll}\text{minimize} & f(x^k)+\langle f'(x^k),\,x-x^k\rangle+\dfrac{1}{2}\langle H_k(x-x^k),\,x-x^k\rangle\\ \text{subject to} & h(x^k)+h'(x^k)(x-x^k)=0,\ g(x^k)+g'(x^k)(x-x^k)\leq 0,\end{array}\tag{4.75}$$

where $H_k\in\mathbf{R}^{n\times n}$ is some symmetric matrix. The basic SQP algorithm corresponds to the choice

$$H_k=\frac{\partial^2 L}{\partial x^2}(x^k,\,\lambda^k,\,\mu^k).\tag{4.76}$$

Recall also that the KKT system (4.74) can be written as the generalized equation (GE)

$$\Phi(u)+N(u)\ni 0,\tag{4.77}$$

with the mapping $\Phi:\mathbf{R}^n\times\mathbf{R}^l\times\mathbf{R}^m\to\mathbf{R}^n\times\mathbf{R}^l\times\mathbf{R}^m$ given by

$$\Phi(u)=\left(\frac{\partial L}{\partial x}(x,\,\lambda,\,\mu),\,h(x),\,-g(x)\right),\tag{4.78}$$

and with

$$N(\cdot)=N_Q(\cdot),\quad Q=\mathbf{R}^n\times\mathbf{R}^l\times\mathbf{R}^m_+,\tag{4.79}$$

where $u=(x,\,\lambda,\,\mu)\in\mathbf{R}^n\times\mathbf{R}^l\times\mathbf{R}^m$. Furthermore, the KKT system of the SQP subproblem (4.75) has the form

$$\begin{gathered}f'(x^k)+H_k(x-x^k)+(h'(x^k))^{\mathrm{T}}\lambda+(g'(x^k))^{\mathrm{T}}\mu=0,\\ h(x^k)+h'(x^k)(x-x^k)=0,\\ \mu\geq 0,\quad g(x^k)+g'(x^k)(x-x^k)\leq 0,\quad \langle\mu,\,g(x^k)+g'(x^k)(x-x^k)\rangle=0.\end{gathered}\tag{4.80}$$

With H_k defined according to (4.76), the latter system is evidently equivalent to the GE

$$\Phi(u^k)+\Phi'(u^k)(u-u^k)+N(u)\ni 0,$$

where $u^k = (x^k,\, \lambda^k,\, \mu^k)$, which is the iteration system of the Josephy–Newton method for the GE (4.77); see Sect. 3.1.

By the *perturbed SQP method*, we mean any iterative procedure fitting the following scheme. For given primal-dual iterate $(x^k,\, \lambda^k,\, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, the next iterate $(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ must satisfy the following perturbed version of the KKT system (4.80) with H_k defined by (4.76):

$$
\begin{gathered}
\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) + \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k)(x - x^k) \\
+(h'(x^k))^{\mathrm{T}}(\lambda - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu - \mu^k) + \omega_1^k = 0, \\
h(x^k) + h'(x^k)(x - x^k) + \omega_2^k = 0, \\
\mu \geq 0, \quad g(x^k) + g'(x^k)(x - x^k) + \omega_3^k \leq 0, \\
\langle \mu,\, g(x^k) + g'(x^k)(x - x^k) + \omega_3^k \rangle = 0,
\end{gathered}
\tag{4.81}
$$

where $\omega_1^k \in \mathbf{R}^n$, $\omega_2^k \in \mathbf{R}^l$, and $\omega_3^k \in \mathbf{R}^m$ are the perturbation terms.

It is important to emphasize that for some forms of perturbations, the perturbed SQP framework includes algorithms which may not be modifications of SQP per se, in the sense that subproblems of those algorithms are not even quadratic. The point is that iterates of all methods in the considered class can be related to a perturbation of SQP given by (4.81) a posteriori. Specific relations would be made clear in the applications described in Sects. 4.3.3, 4.3.5 below. Our framework is to some extent related to inexact SQP in [270]. However, the latter was intended to deal with certain inexactness specifically within the SQP class, while our approach is designed to include also other algorithms. Also, the assumptions to be imposed on the perturbations terms are rather different in our developments.

Evidently, any perturbed SQP method can be regarded as a perturbed Josephy–Newton method for the GE (4.77) with $\Phi(\cdot)$ and $N(\cdot)$ defined according to (4.78) and (4.79), respectively, as considered in Sect. 3.1 (see (3.11)). In particular, Proposition 3.4 and Theorem 3.5 can be applied in this context to obtain a posteriori results on sufficient conditions for primal-dual superlinear convergence rate of perturbed SQP methods. In addition, similarly to the development in Sect. 4.1.1, we next obtain a posteriori results providing necessary and sufficient conditions for the primal superlinear convergence rate. These results were derived in [77]. We first state separately the necessary conditions; these are a direct generalization of the necessity part of Proposition 4.4.

Proposition 4.16. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (4.73), and let $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be a Lagrange multiplier associated with $\bar{x}$. Let $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ be convergent to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, and assume that for each k large enough the triple $(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1})$ satisfies the system (4.81) with some $\omega_1^k \in \mathbf{R}^n$, $\omega_2^k \in \mathbf{R}^l$ and $\omega_3^k \in \mathbf{R}^m$.*

If the rate of convergence of $\{x^k\}$ *is superlinear, then*

$$\omega_2^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|), \tag{4.82}$$

$$(\omega_3^k)_{A_+(\bar{x},\,\bar{\mu})} = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \tag{4.83}$$

as $k \to \infty$*. If in addition*

$$\{(\omega_3^k)_{\{1,\,\dots,\,m\}\setminus A(\bar{x})}\} \to 0, \tag{4.84}$$

then

$$\pi_{C(\bar{x})}(-\omega_1^k) = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \tag{4.85}$$

as $k \to \infty$*, where*

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A(\bar{x})}(\bar{x})\xi \leq 0,\ \langle f'(\bar{x}),\, \xi\rangle \leq 0\} \tag{4.86}$$

is the critical cone of problem (4.73) *at* $\bar{x}$*.*

Proof. The same way as in (3.88), from (4.81) and from the superlinear convergence of $\{x^k\}$ to $\bar{x}$ we derive the estimate

$$\begin{aligned}\omega_1^k = &-(h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) - (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu})\\ &+o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|)\end{aligned} \tag{4.87}$$

as $k \to \infty$.

Similarly, from (4.81) it also follows that

$$\begin{aligned}\omega_2^k &= -h(x^k) - h'(x^k)(x^{k+1} - x^k)\\ &= -h(x^{k+1}) + o(\|x^{k+1} - x^k\|)\\ &= -h(\bar{x}) - h'(\bar{x})(x^{k+1} - \bar{x}) + o(\|x^{k+1} - \bar{x}\|) + o(\|x^{k+1} - x^k\|)\\ &= o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|)\end{aligned}$$

as $k \to \infty$. The latter relation gives (4.82). Moreover, since $\bar{\mu}_{A_+(\bar{x},\,\bar{\mu})} > 0$, we have that $\mu^k_{A_+(\bar{x},\,\bar{\mu})} > 0$ for all k large enough, and it then follows from the last line in (4.81) that

$$\begin{aligned}(\omega_3^k)_{A_+(\bar{x},\,\bar{\mu})} &= -g_{A_+(\bar{x},\,\bar{\mu})}(x^k) - g'_{A_+(\bar{x},\,\bar{\mu})}(x^k)(x^{k+1} - x^k)\\ &= -g_{A_+(\bar{x},\,\bar{\mu})}(x^{k+1}) + o(\|x^{k+1} - x^k\|)\\ &= -g_{A_+(\bar{x},\,\bar{\mu})}(\bar{x}) - g'_{A_+(\bar{x},\,\bar{\mu})}(\bar{x})(x^{k+1} - \bar{x})\\ &\quad\ +o(\|x^{k+1} - \bar{x}\|) + o(\|x^{k+1} - x^k\|)\\ &= o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|)\end{aligned}$$

as $k \to \infty$, which gives (4.83).

For each k, define

$$\tilde{\omega}_1^k = -(h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) - (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}). \tag{4.88}$$

Then, by (4.87), it holds that

$$\omega_1^k - \tilde{\omega}_1^k = o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|) \tag{4.89}$$

as $k \to \infty$.

Since $\{g_{\{1,\ldots,m\}\setminus A(\bar{x})}(x^k)\} \to g_{\{1,\ldots,m\}\setminus A(\bar{x})}(\bar{x}) < 0$, assuming (4.84) we observe that the last line in (4.81) implies that $\mu^k_{\{1,\ldots,m\}\setminus A(\bar{x})} = 0$ for all k large enough. Taking this into account, we obtain from (4.88) and from Lemma 1.17 that for such k, for any $\xi \in C(\bar{x})$ it holds that

$$\begin{aligned}\langle -\tilde{\omega}_1^k, \xi\rangle &= \langle \lambda^{k+1} - \bar{\lambda}, h'(\bar{x})\xi\rangle + \langle \mu^{k+1} - \bar{\mu}, g'(\bar{x})\xi\rangle \\ &= \langle \mu^{k+1}_{A_0(\bar{x},\bar{\mu})}, g'_{A_0(\bar{x},\bar{\mu})}(\bar{x})\xi\rangle \\ &\le 0,\end{aligned}$$

where the inequality $\mu^{k+1} \ge 0$ was taken into account. This means that $-\tilde{\omega}_1^k \in C^{\circ}$, and hence, by Lemma A.13, $\pi_C(-\tilde{\omega}_1^k) = 0$. Employing now (4.89) and the fact that $\pi_C(\cdot)$ is nonexpansive (see Lemma A.13), we obtain (4.85). □

We proceed with sufficient conditions for the primal superlinear convergence in the perturbed SQP framework. The goal is, of course, to keep these conditions as close as possible to the necessary conditions in Proposition 4.16. We start with the following general fact.

Proposition 4.17. *Suppose that the assumptions of Proposition 4.16 hold. Let $\pi : \mathbf{R}^n \to \mathbf{R}^n$ be any Lipschitz-continuous mapping, and assume that the estimate*

$$x - \bar{x} = O\left(\left\|\begin{pmatrix}\pi\left(\dfrac{\partial L}{\partial x}(x, \lambda, \mu)\right) \\ h(x) \\ \min\{\mu, -g(x)\}\end{pmatrix}\right\|\right) \tag{4.90}$$

holds as $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ tends to $(\bar{x}, \bar{\lambda}, \bar{\mu})$.

If

$$\{\pi(-\omega_1^k)\} \to 0, \ \{\omega_2^k\} \to 0, \ \{\omega_3^k\} \to 0, \tag{4.91}$$

then

$$x^{k+1} - \bar{x} = O(\|(\pi(-\omega_1^k), \omega_2^k, (\omega_3^k)_{A(\bar{x})})\|) + o(\|x^k - \bar{x}\|) \tag{4.92}$$

as $k \to \infty$. In particular, if in addition

$$\pi(-\omega_1^k) = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|), \tag{4.93}$$

and conditions (4.82), (4.83) *and*

$$(\omega_3^k)_{A_0(\bar{x},\,\bar{\mu})} = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \tag{4.94}$$

hold as $k \to \infty$*, then the rate of convergence of* $\{x^k\}$ *is superlinear.*

Proof. For each k, we have that

$$\begin{aligned}\frac{\partial L}{\partial x}(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1}) &= \frac{\partial L}{\partial x}(x^{k+1},\, \lambda^k,\, \mu^k) + (h'(x^{k+1}))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k)\\ &\quad +(g'(x^{k+1}))^{\mathrm{T}}(\mu^{k+1} - \mu^k)\\ &= \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) + \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k)(x^{k+1} - x^k)\\ &\quad +(h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k)\\ &\quad +o(\|x^{k+1} - x^k\|)\\ &= -\omega_1^k + o(\|x^{k+1} - x^k\|)\end{aligned} \tag{4.95}$$

as $k \to \infty$, where the convergence of $\{(x^k,\, \lambda^k,\, \mu^k)\}$ was used in the second equality, and the first relation of (4.81) was used in the last one.

Similarly, using the second relation of (4.81), we have that

$$\begin{aligned}h(x^{k+1}) &= h(x^k) + h'(x^k)(x^{k+1} - x^k) + o(\|x^{k+1} - x^k\|)\\ &= -\omega_2^k + o(\|x^{k+1} - x^k\|)\end{aligned} \tag{4.96}$$

as $k \to \infty$.

Since $\{g_{\{1,\,\dots,\,m\}\setminus A(\bar{x})}(x^k)\} \to g_{\{1,\,\dots,\,m\}\setminus A(\bar{x})}(\bar{x}) < 0$ while at the same time $\{(\omega_3^k)_{\{1,\,\dots,\,m\}\setminus A(\bar{x})}\} \to 0$ (see (4.91)), this evidently implies that for each k large enough it holds that $\mu^{k+1}_{\{1,\,\dots,\,m\}\setminus A(\bar{x})} = 0$, and hence

$$\min\{\mu^{k+1}_{\{1,\,\dots,\,m\}\setminus A(\bar{x})},\, -g_{\{1,\,\dots,\,m\}\setminus A(\bar{x})}(x^{k+1})\} = 0. \tag{4.97}$$

Furthermore, the last line in (4.81) can be written in the form

$$\min\{\mu^{k+1},\, -g(x^k) - g'(x^k)(x^{k+1} - x^k) - \omega_3^k\} = 0.$$

Using this equality and the obvious property

$$|\min\{a,\, b\} - \min\{a,\, c\}| \leq |b - c| \quad \forall\, a,\, b,\, c \in \mathbf{R},$$

we obtain that

$$\begin{aligned}|\min\{\mu^{k+1}_{A(\bar{x})},\, -g_{A(\bar{x})}(x^{k+1})\}| &= |\min\{\mu^{k+1}_{A(\bar{x})},\, -g_{A(\bar{x})}(x^k)\\ &\quad -g'_{A(\bar{x})}(x^k)(x^{k+1} - x^k) + o(\|x^{k+1} - x^k\|)\}\\ &\quad - \min\{\mu^{k+1}_{A(\bar{x})},\, -g_{A(\bar{x})}(x^k)\\ &\quad -g'_{A(\bar{x})}(x^k)(x^{k+1} - x^k) - (\omega_3^k)_{A(\bar{x})}\}|\\ &\leq |(\omega_3^k)_{A(\bar{x})}| + o(\|x^{k+1} - x^k\|)\end{aligned} \tag{4.98}$$

as $k \to \infty$.

From (4.95), (4.96), (4.97), and (4.98), by the error bound (4.90) we derive the estimate

$$\begin{aligned} x^{k+1} - \bar{x} &= O(\|(\pi(-\omega_1^k),\, \omega_2^k,\, (\omega_3^k)_{A(\bar{x})})\|) + o(\|x^{k+1} - x^k\|) \\ &= O(\|(\pi(-\omega_1^k),\, \omega_2^k,\, (\omega_3^k)_{A(\bar{x})})\|) + o(\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|) \end{aligned}$$

as $k \to \infty$, which means the existence of a constant $c > 0$ and of a sequence $\{t_k\} \in \mathbf{R}_+$ such that $\{t_k\} \to 0$, and for all k

$$\|x^{k+1} - \bar{x}\| \leq c\|(\pi(-\omega_1^k),\, \omega_2^k,\, (\omega_3^k)_{A(\bar{x})})\|) + t_k(\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|).$$

The latter implies that for all k large enough

$$\begin{aligned} \frac{1}{2}\|x^{k+1} - \bar{x}\| &\leq (1 - t_k)\|x^{k+1} - \bar{x}\| \\ &\leq c\|(\pi(-\omega_1^k),\, \omega_2^k,\, (\omega_3^k)_{A(\bar{x})})\|) + t_k\|x^k - \bar{x}\|, \end{aligned}$$

which evidently gives the estimate (4.92).

Furthermore, combining (4.92) with conditions (4.82), (4.83), (4.93), and (4.94), we conclude that

$$x^{k+1} - \bar{x} = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) = o(\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|)$$

as $k \to \infty$. By the same argument as in the proof of Proposition 2.4, the latter implies the superlinear convergence rate of $\{x^k\}$. □

We can now combine Proposition 4.17 with the error bounds under the noncriticality of the Lagrange multiplier or under the second-order sufficiency assumptions, established in Sect. 1.3.3. Specifically, combining Proposition 1.44 and Proposition 4.17 gives the following generalization of the sufficiency part of Proposition 4.4.

Proposition 4.18. *Under the assumptions of Proposition 4.16, let* $(\bar{\lambda},\, \bar{\mu})$ *be a noncritical Lagrange multiplier.*

If

$$\pi_{C_+(\bar{x},\, \bar{\mu})}(\omega_1^k) = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \tag{4.99}$$

as $k \to \infty$, *where*

$$C_+(\bar{x},\, \bar{\mu}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+}(\bar{x})\xi = 0\},$$

and conditions (4.82), (4.83), (4.94) *hold, then the rate of convergence of* $\{x^k\}$ *is superlinear.*

Example 1.45 demonstrates that in (4.99) one cannot replace the subspace $C_+(\bar{x},\, \bar{\mu})$ by the (generally smaller) critical cone $C(\bar{x})$, even if $\bar{x}$ satisfies the LICQ. Indeed, in this example, for any $\omega_1^k < 0$ and $\omega_3^k = 0$, the point $(x^{k+1}, \mu^{k+1}) = (\omega_1^k, 0)$ satisfies the system (4.81). If the sequence $\{\omega_1^k\}$ converges to zero, then $\{(x^k, \mu^k)\}$ converges to $(\bar{x}, \bar{\mu})$. However, $\pi_{C(\bar{x})}(-\omega_1^k) = 0$, and condition (4.99) with $C_+(\bar{x}, \bar{\mu})$ replaced by $C(\bar{x})$ does not impose any restrictions on the rate of convergence of $\{\omega_1^k\}$, and hence, on the rate of convergence of $\{x^k\}$.

On the other hand, combining Proposition 1.46 and Proposition 4.17, we conclude that $C_+(\bar{x}, \bar{\mu})$ can be replaced by $C(\bar{x})$ if the noncriticality assumption on the Lagrange multiplier is replaced by the (stronger) SOSC. Note that, unlike $C_+(\bar{x}, \bar{\mu})$, the critical cone $C(\bar{x})$ is not necessarily a linear subspace. Therefore, one cannot remove the minus sign in the left-hand side of (4.85).

Proposition 4.19. *Under the assumptions of Proposition 4.16, let the SOSC*

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in C(\bar{x}) \setminus \{0\}, \tag{4.100}$$

be satisfied.

If conditions (4.82), (4.83), (4.94) hold, then the rate of convergence of $\{x^k\}$ is superlinear.

Remark 4.20. By Remark 2.5, and similarly to Remark 4.5, we note that in each of the conditions (4.82), (4.83), (4.85), (4.93), (4.94), (4.99) of Propositions 4.16–4.19, the right-hand side can be replaced by either $o(\|x^{k+1} - x^k\|)$ or $o(\|x^k - \bar{x}\|)$.

Remark 4.21. Combining the estimate (1.120) from Remark 1.47 with Proposition 4.17, we immediately obtain primal superlinear convergence under the assumptions of Proposition 4.19, but with the SOSC (4.100) replaced by the "symmetric" condition with the opposite sign:

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle < 0 \quad \forall \xi \in C(\bar{x}) \setminus \{0\}, \tag{4.101}$$

and with (4.85) replaced by

$$\pi_{C(\bar{x})}(\omega_1^k) = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|).$$

Observe that in Example 1.45, condition (4.101) is satisfied. If $\omega_1^k < 0$, then $\pi_{C(\bar{x})}(\omega_1^k) = \omega_1^k$, and the assertion stated in Remark 4.21 is evidently valid.

Both Propositions 4.18 and 4.19 rely on the assumptions that the dual sequence converges to a multiplier which satisfies the SOSC or, more generally, is noncritical. Such assumptions are completely natural when the SMFCQ holds at $\bar{x}$. In this case, the unique associated Lagrange multiplier is the unique natural attractor for the dual sequence, and there are no particular reasons for this multiplier to be critical. However, when multipliers are not unique, the asymptotic dual behavior of Newton-type methods is a subtle issue, which will be addressed in detail in Sect. 7.1.

In *quasi-Newton SQP methods*, the matrix H_k in the iteration subproblem (4.75) is chosen not by (4.76) but rather as a quasi-Newton approximation of $\frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)$ (see the discussion of quasi-Newton–Lagrange methods in Sect. 4.1.1). In this case, the KKT system (4.80) of the subproblem (4.75) is a special case of the perturbed SQP (4.81), corresponding to

$$\omega_1^k = \left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)\right)(x^{k+1} - x^k), \quad \omega_2^k = 0, \quad \omega_3^k = 0.$$

Characterization of primal superlinear convergence of both pure and quasi-Newton SQP methods readily follows from Propositions 4.16, 4.18, 4.19, and Remark 4.20; this is a generalization of Theorem 4.6.

Theorem 4.22. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (4.73), and let $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be an associated Lagrange multiplier. Let $\{H_k\} \subset \mathbf{R}^{n\times n}$ be a sequence of symmetric matrices, let $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ be convergent to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and assume that x^{k+1} is a stationary point of problem (4.75), and $(\lambda^{k+1}, \mu^{k+1})$ is an associated Lagrange multiplier for each k large enough.*

If the rate of convergence of $\{x^k\}$ is superlinear, then the Dennis–Moré condition holds:

$$\pi_{C(\bar{x})}\left(\frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k) - H_k\right)(x^{k+1} - x^k) = o(\|x^{k+1} - x^k\|) \qquad (4.102)$$

as $k \to \infty$.

Conversely, if $(\bar{\lambda}, \bar{\mu})$ satisfies the SOSC (4.100) and the Dennis-Moré condition (4.102) holds, then the rate of convergence of $\{x^k\}$ is superlinear.

Alternatively, under the weaker assumption that $(\bar{\lambda}, \bar{\mu})$ is a noncritical Lagrange multiplier, if

$$\pi_{C_+(\bar{x},\, \bar{\mu})}\left(\left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)\right)(x^{k+1} - x^k)\right) = o(\|x^{k+1} - x^k\|) \qquad (4.103)$$

as $k \to \infty$, then the rate of convergence of $\{x^k\}$ is also superlinear.

Observe that for pure SQP with the choice of H_k as in (4.76), the condition (4.103) becomes automatic and the primal convergence rate is superlinear if the relevant multiplier is noncritical. Note also that according to Theorem 4.14, under the assumptions of the SMFCQ and the SOSC, SQP actually does possess local primal-dual convergence, and thus, the assumption of primal-dual convergence is not needed in that case. Theorem 4.22 then implies the primal superlinear rate under the SMFCQ and the SOSC only, whenever the starting point is close enough to the primal-dual solution. Prior to [77] which employed the perturbed SQP tools, stronger assumptions were required for the conclusions above.

Example 1.45 demonstrates that in the absence of the SOSC (4.100), the cone $C_+(\bar{x}, \bar{\mu})$ in (4.103) cannot be replaced by $C(\bar{x})$. Indeed, taking, e.g., $H_k = -2$ for all k, for any $x^k < 0$ we obtain that $x^{k+1} = x^k/2 < 0$ is a stationary point of problem (4.75). Thus, for any $x^0 < 0$, the corresponding primal trajectory converges to $\bar{x} = 0$ but only linearly. At the same time, the left-hand side of (4.102) (but not of (4.103)!) equals to zero for all k.

Employing Remark 4.21, it can be easily seen that the Dennis-Moré condition (4.102) can be replaced in Theorem 4.22 by the "symmetric" condition

$$\pi_{C(\bar{x})}\left(\left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k)\right)(x^{k+1} - x^k)\right) = o(\|x^{k+1} - x^k\|)$$

provided the SOSC (4.100) is replaced by the "symmetric" condition (4.101). In particular, this gives a sufficient condition for the primal superlinear convergence of a quasi-Newton SQP method in Example 1.45: $H_k \to -1$ as $k \to \infty$.

Comparing Theorem 4.22 with Theorem 3.40, we see that primal superlinear convergence of the semismooth Newton method applied to the natural residual-based reformulation of the KKT system requires stronger assumptions than the corresponding property for SQP; specifically, the SOSC (4.100) must be replaced by the stronger SSOSC (3.71).

We complete this section with an a priori result on local superlinear convergence of perturbed SQP methods, originally established in [147]. It is based on Theorems 1.40 and 3.6, where the former will be used to establish the solvability of subproblems. To that end, we first need to consider the perturbed SQP iteration system (4.81) with more structured perturbation terms, corresponding to some optimization problem which can be tackled by Theorem 1.40. We start with the latter: consider an iteration subproblem of the form

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, x - x^k\rangle + \dfrac{1}{2}\left\langle \dfrac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k)(x - x^k),\, x - x^k\right\rangle \\ & +\psi((x^k,\, \lambda^k,\, \mu^k),\, x - x^k) \\ \text{subject to} & h(x^k) + h'(x^k)(x - x^k) + \omega_2((x^k,\, \lambda^k,\, \mu^k),\, x - x^k) = 0, \\ & g(x^k) + g'(x^k)(x - x^k) + \omega_3((x^k,\, \lambda^k,\, \mu^k),\, x - x^k) \leq 0, \end{array} \tag{4.104}$$

with some function $\psi : (\mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m) \times \mathbf{R}^n \to \mathbf{R}$ and some mappings $\omega_2 : (\mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m) \times \mathbf{R}^n \to \mathbf{R}^l$ and $\omega_3 : (\mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m) \times \mathbf{R}^n \to \mathbf{R}^m$, which are smooth with respect to the last variable.

Define the function $\Psi : (\mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m) \times (\mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m) \to \mathbf{R}$,

$$\begin{aligned} \Psi((x,\, \lambda,\, \mu),\, (\xi,\, \eta,\, \zeta)) = \psi((x,\, \lambda,\, \mu),\, \xi) &+ \langle \lambda+\eta,\, \omega_2((x,\, \lambda,\, \mu),\, \xi)\rangle \\ &+ \langle \mu+\zeta,\, \omega_3((x,\, \lambda,\, \mu),\, \xi)\rangle. \end{aligned} \tag{4.105}$$

For $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and $(\xi, \eta, \zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, set

$$\omega_1((x, \lambda, \mu), (\xi, \eta, \zeta)) = \frac{\partial \Psi}{\partial \xi}((x, \lambda, \mu), (\xi, \eta, \zeta)). \tag{4.106}$$

Then we can write the KKT system of problem (4.104) as follows:

$$\begin{gathered} f'(x^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)(x - x^k) \\ +(h'(x^k))^{\mathrm{T}}\lambda + (g'(x^k))^{\mathrm{T}}\mu + \omega_1((x^k, \lambda^k, \mu^k), (x - x^k, \lambda - \lambda^k, \mu - \mu^k)) = 0, \\ h(x^k) + h'(x^k)(x - x^k) + \omega_2((x^k, \lambda^k, \mu^k), x - x^k) = 0, \\ \mu \geq 0, \quad g(x^k) + g'(x^k)(x - x^k) + \omega_3((x^k, \lambda^k, \mu^k), x - x^k) \leq 0, \\ \langle \mu, g(x^k) + g'(x^k)(x - x^k) + \omega_3((x^k, \lambda^k, \mu^k), x - x^k) \rangle = 0, \end{gathered} \tag{4.107}$$

which corresponds to (4.81) with

$$\omega_1^k = \omega_1((x^k, \lambda^k, \mu^k), (x^{k+1} - x^k, \lambda^{k+1} - \lambda^k, \mu^{k+1} - \mu^k)), \tag{4.108}$$

$$\omega_j^k = \omega_j((x^k, \lambda^k, \mu^k), x^{k+1} - x^k), \quad j = 2, 3. \tag{4.109}$$

In what follows, the terms defined by (4.108), (4.109) (employing (4.105), (4.106)) will correspond to structural perturbations characterizing various specific algorithms for problem (4.73) within the general perturbed SQP framework. However, since (4.104) is now not necessarily a QP, one cannot expect that these subproblems will be solved exactly in practical implementations. Therefore, it is natural to consider the perturbed SQP method with the additional inexactness in solving the subproblems. Note also that even if subproblems are QPs, it can still make good sense to solve them only approximately; this is the essence of the truncated SQP discussed in Sect. 4.3.2.

Motivated by the discussion above, we now consider that the next iterate $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ satisfies the following inexact version of system (4.107):

$$\begin{gathered} \left\| f'(x^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)(x - x^k) + (h'(x^k))^{\mathrm{T}}\lambda + (g'(x^k))^{\mathrm{T}}\mu \right. \\ \left. + \, \omega_1((x^k, \lambda^k, \mu^k), (x - x^k, \lambda - \lambda^k, \mu - \mu^k)) \right\| \leq \varphi(\rho(x^k, \lambda^k, \mu^k)), \\ \|h(x^k) + h'(x^k)(x - x^k) + \omega_2((x^k, \lambda^k, \mu^k), x - x^k)\| \leq \varphi(\rho(x^k, \lambda^k, \mu^k)), \\ \mu \geq 0, \quad g(x^k) + g'(x^k)(x - x^k) + \omega_3((x^k, \lambda^k, \mu^k), x - x^k) \leq 0, \\ \langle \mu, g(x^k) + g'(x^k)(x - x^k) + \omega_3((x^k, \lambda^k, \mu^k), x - x^k) \rangle = 0. \end{gathered} \tag{4.110}$$

Here $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$ is some forcing function controlling the additional inexactness in solving the subproblems, and $\rho : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}_+$ is the natural residual of the KKT system (4.74) of the original problem (4.73):

$$\rho(x, \lambda, \mu) = \left\| \left(\frac{\partial L}{\partial x}(x, \lambda, \mu), h(x), \min\{\mu, -g(x)\} \right) \right\|. \tag{4.111}$$

Remark 4.23. We use the particular form given by (4.111) for the right-hand sides of the first two relations in (4.110) merely to be specific, and also to keep this discussion closer to practical truncation rules. However, for the purposes of the analysis here, one can actually use any other residual as ρ, assuming only that it is locally Lipschitz-continuous with respect to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ under natural smoothness assumptions. Moreover, one can even employ any function $\rho : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}_+$ satisfying the property

$$\rho(x, \lambda, \mu) = O(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|)$$

as $(x, \lambda, \mu) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$. It can be seen from the proof of Theorem 4.24 below that its assertions remain valid after these modifications.

Furthermore, it is even possible to replace the entire right-hand sides of the first two relations in (4.110) by $\chi_1(x^k, \lambda^k, \mu^k)$ and $\chi_2(x^k, \lambda^k, \mu^k)$, respectively, where $\chi_1 : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}_+$ and $\chi_2 : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}_+$ are any functions satisfying

$$\chi_j(x, \lambda, \mu) = o(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|), \quad j = 1, 2,$$

for the superlinear convergence of the framework given by (4.110), and

$$\chi_j(x, \lambda, \mu) = O(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|^2), \quad j = 1, 2,$$

for the quadratic convergence.

We note that the extensions commented above are not merely formal, as they are useful for some applications of the perturbed SQP framework; see Sect. 4.3.4 on inexact restoration methods for one specific example.

Note that the conditions in (4.110) corresponding to the inequality constraints of (4.104) do not allow for any additional inexactness: these are precisely the corresponding conditions in (4.107). This is somewhat of a drawback, but at this time it is not known how to treat the additional inexactness in inequality constraints within general local analysis (this is related to some principal difficulties in the areas of stability/sensitivity of perturbed subproblems). See, however, how this can be worked around in the truncated SQP method for the special case of bound constraints [147], the development also summarized in Sect. 4.3.2.

For $u = (x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and $v = (\xi, \eta, \zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, set

$$\omega(u, v) = (\omega_1((x, \lambda, \mu), (\xi, \eta, \zeta)), \omega_2((x, \lambda, \mu), \xi), \omega_3((x, \lambda, \mu), \xi)), \tag{4.112}$$

$$\Theta_1(u) = \{\theta_1 \in \mathbf{R}^n \mid \|\theta_1\| \leq \varphi(\rho(x, \lambda, \mu))\}, \tag{4.113}$$

$$\Theta_2(u) = \{\theta_2 \in \mathbf{R}^l \mid \|\theta_2\| \leq \varphi(\rho(x, \lambda, \mu))\}, \tag{4.114}$$

$$\Theta(u) = \Theta_1(u) \times \Theta_2(u) \times \{0\}, \tag{4.115}$$

where 0 is the zero element in $\mathbf{R}^m$, and

$$\Omega(u,\, v) = \omega(u,\, v) + \Theta(u). \tag{4.116}$$

System (4.110) can then be written as the GE

$$\Phi(u^k) + \Phi'(u^k)(u - u^k) + \Omega(u^k,\, u - u^k) + N(u) \ni 0, \tag{4.117}$$

where $u^k = (x^k,\, \lambda^k,\, \mu^k)$, and Φ and N are defined according to (4.78) and (4.79), respectively.

Separating the perturbation into the single-valued part $\omega(u,\, v)$ and the set-valued part $\Theta(u)$ is instructive, because the two parts correspond to perturbations of different kind: as explained above, $\omega(u,\, v)$ represents structural perturbations with respect to the basic SQP iteration, while $\Theta(u)$ stands for additional inexactness allowed when solving the subproblems of the specific method in consideration.

Note that the GE (4.117) is an iteration of the perturbed Josephy–Newton method considered in Theorem 3.6 (see (3.19)). We then have the following statement, fairly general and technical, but which turns to be a useful tool for treating various specific algorithms in a unified manner.

Theorem 4.24. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a local solution of problem (4.73), satisfying the SMFCQ and the SOSC (4.100) for the associated Lagrange multiplier $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$. Furthermore, let a function $\psi : (\mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m) \times \mathbf{R}^n \to \mathbf{R}$, a mapping $\omega_2 : (\mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m) \times \mathbf{R}^n \to \mathbf{R}^l$ and a mapping $\omega_3 : (\mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m) \times \mathbf{R}^n \to \mathbf{R}^m$ possess the following properties:*

(i) *ψ is continuous at $((\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, \xi)$ and $\omega_2(\cdot,\, \xi)$ and $\omega_3(\cdot,\, \xi)$ are continuous at $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, for every $\xi \in \mathbf{R}^n$ close enough to 0.*

(ii) *ψ, ω_2 and ω_3 are differentiable with respect to ξ in a neighborhood of $((\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, 0)$, and twice differentiable with respect to ξ at $((\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, 0)$.*

(iii) *ω_3, $\frac{\partial \psi}{\partial \xi}$, $\frac{\partial \omega_2}{\partial \xi}$, and $\frac{\partial \omega_3}{\partial \xi}$ are continuous at $((\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, 0)$, and there exists a neighborhood of 0 in $\mathbf{R}^n$ such that $\frac{\partial \omega_2}{\partial \xi}((x,\, \lambda,\, \mu),\, \cdot)$ is continuous on this neighborhood for all $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$.*

(iv) *The equalities*

$$\omega_2((\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, 0) = 0, \quad \omega_3((\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, 0) = 0,$$

$$\frac{\partial \psi}{\partial \xi}((\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, 0) = 0, \quad \frac{\partial \omega_2}{\partial \xi}((\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, 0) = 0, \quad \frac{\partial \omega_3}{\partial \xi}((\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, 0) = 0$$

hold, and for the function Ψ defined by (4.105), it holds that

$$\left\langle \frac{\partial^2 \Psi}{\partial \xi^2}((\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, (0,\, 0,\, 0))\xi,\, \xi \right\rangle \geq 0 \quad \forall\, \xi \in C(\bar{x}). \tag{4.118}$$

Assume further that $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$ *is a function such that* $\varphi(t) = o(t)$ *as* $t \to 0$*, and that the estimates*

$$\omega_j((x,\, \lambda,\, \mu),\, \xi) = o(\|\xi\| + \|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu})\|), \quad j = 2,\, 3, \tag{4.119}$$

$$\frac{\partial \Psi}{\partial \xi}((x,\, \lambda,\, \mu),\, (\xi,\, \eta,\, \zeta)) = o(\|(\xi,\, \eta,\, \zeta)\| + \|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu})\|) \tag{4.120}$$

hold as $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ *tends to* $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$*, and as the element* $(\xi,\, \eta,\, \zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ *tends to zero, for* $(x,\, \lambda,\, \mu)$ *and* $(\xi,\, \eta,\, \zeta)$ *satisfying*

$$\begin{gathered} \left\| f'(x) + \frac{\partial^2 L}{\partial x^2}(x,\, \lambda,\, \mu)\xi + (h'(x))^{\mathrm{T}}(\lambda + \eta) + (g'(x))^{\mathrm{T}}(\mu + \zeta) \right. \\ \left. + \frac{\partial \Psi}{\partial \xi}((x,\, \lambda,\, \mu),\, (\xi,\, \eta,\, \zeta)) \right\| \leq \varphi(\rho(x,\, \lambda,\, \mu)), \\ \|h(x) + h'(x)\xi + \omega_2((x,\, \lambda,\, \mu),\, \xi)\| \leq \varphi(\rho(x,\, \lambda,\, \mu)), \\ \mu + \zeta \geq 0, \quad g(x) + g'(x)\xi + \omega_3((x,\, \lambda,\, \mu),\, \xi) \leq 0, \\ \langle \mu + \zeta,\, g(x) + g'(x)\xi + \omega_3((x,\, \lambda,\, \mu),\, \xi)\rangle = 0. \end{gathered} \tag{4.121}$$

Then there exists a constant $\delta > 0$ *such that for any starting point* $(x^0,\, \lambda^0,\, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ *close enough to* $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$*, there exists a sequence* $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ *such that for each* $k = 0,\, 1,\, \ldots,$ *the triple* $(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1})$ *satisfies the system* (4.110) *with* ω_1 *defined in* (4.106) *and* Ψ *defined in* (4.105)*, and also satisfies the inequality*

$$\|(x^{k+1} - x^k,\, \lambda^{k+1} - \lambda^k,\, \mu^{k+1} - \mu^k)\| \leq \delta; \tag{4.122}$$

any such sequence converges to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$*, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic if the second derivatives of* f*,* h *and* g *are locally Lipschitz-continuous with respect to* $\bar{x}$*, if* (4.119) *and* (4.120) *can be replaced by the estimates*

$$\omega_j((x,\, \lambda,\, \mu),\, \xi) = O(\|\xi\|^2 + \|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu})\|^2), \quad j = 2,\, 3, \tag{4.123}$$

$$\frac{\partial \Psi}{\partial \xi}((x,\, \lambda,\, \mu),\, (\xi,\, \eta,\, \zeta)) = O(\|(\xi,\, \eta,\, \zeta)\|^2 + \|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu})\|^2), \tag{4.124}$$

and if $\varphi(t) = O(t^2)$ *as* $t \to 0$*.*

Proof. According to Proposition 1.37, under the stated assumptions the solution $\bar{u} = (\bar{x},\, \bar{\lambda},\, \bar{\mu})$ of the GE (4.77) is automatically semistable.

In order to apply Theorem 3.6, it remains to establish the following properties:

(a) For each $u = (x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $\bar{u} = (\bar{x},\, \bar{\lambda},\, \bar{\mu})$, the GE

$$\Phi(u) + \Phi'(u)v + \Omega(u,\, v) + N(u + v) \ni 0 \tag{4.125}$$

has a solution $v(u)$ tending to zero as $u \to \bar{u}$.

(b) The estimate
$$\omega = o(\|v\| + \|u - \bar{u}\|) \tag{4.126}$$
holds as $u \to \bar{u}$ and $v \to 0$, uniformly for $\omega \in \Omega(u, v)$, $u \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and $v \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ satisfying the GE
$$\Phi(u) + \Phi'(u)v + \omega + N(u + v) \ni 0. \tag{4.127}$$

For a given $u = (x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, consider the system

$$\begin{gathered}
f'(x) + \frac{\partial^2 L}{\partial x^2}(x, \lambda, \mu)\xi + (h'(x))^{\mathrm{T}}(\lambda + \eta) + (g'(x))^{\mathrm{T}}(\mu + \zeta) \\
+\omega_1((x, \lambda, \mu), (\xi, \eta, \zeta)) = 0, \\
h(x) + h'(x)\xi + \omega_2((x, \lambda, \mu), \xi) = 0, \\
\mu + \zeta \geq 0, \quad g(x) + g'(x)\xi + \omega_3((x, \lambda, \mu), \xi) \leq 0, \\
\langle \mu + \zeta, g(x) + g'(x)\xi + \omega_3((x, \lambda, \mu), \xi)\rangle = 0,
\end{gathered}$$

with respect to $v = (\xi, \eta, \zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$. Since the system (4.128) is equivalent to the GE
$$\Phi(u) + \Phi'(u)v + \omega(u, v) + N(u + v) \ni 0, \tag{4.128}$$
and since any solution of (4.128) evidently satisfies (4.125) (see (4.113)–(4.116)), in order to establish property (a) we only need to show that for each $u \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $\bar{u}$, there exists a solution $v(u) = (\xi(u), \eta(u), \zeta(u))$ of the system (4.128) such that $v(u)$ tends to zero as u tends to $\bar{u}$. The latter can be derived by considering the parametric optimization problem

$$\begin{array}{ll}
\text{minimize} & f(x) + \langle f'(x), \xi\rangle + \dfrac{1}{2}\left\langle \dfrac{\partial^2 L}{\partial x^2}(x, \lambda, \mu)\xi, \xi \right\rangle + \psi((x, \lambda, \mu), \xi) \\
\text{subject to} & h(x) + h'(x)\xi + \omega_2((x, \lambda, \mu), \xi) = 0, \\
& g(x) + g'(x)\xi + \omega_3((x, \lambda, \mu), \xi) \leq 0,
\end{array} \tag{4.129}$$

with $u = (x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ regarded as a parameter with the base value $\bar{u}$. Indeed, one can easily check that $\xi = 0$ is a stationary point of this problem for $u = \bar{u}$, and this point satisfies the SMFCQ and the SOSC for this problem with the associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$. Under the stated assumptions, the needed assertion now readily follows from Theorem 1.40, taking into account that (4.128) with $\omega_1((x, \lambda, \mu), \cdot)$ defined according to (4.105), (4.106) is the KKT system of problem (4.129) (with the dual variables of the form $(\lambda + \eta, \mu + \zeta)$).

Assume now that $\omega \in \Omega(u, v)$, $u = (x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and $v = (\xi, \eta, \zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ satisfy (4.127), or equivalently, satisfy (4.121) (see (4.106), (4.112)–(4.116)). Then

$$\omega - \omega(u,\, v) \in \varTheta_1(u) \times \varTheta_2(u) \times \{0\} \tag{4.130}$$

(see (4.115), (4.116)), and by the assumptions of the theorem, the estimates (4.119) and (4.120) hold, or equivalently,

$$\omega(u,\, v) = o(\|v\| + \|u - \bar{u}\|) \tag{4.131}$$

as $u \to \bar{u}$ and $v \to 0$ (see (4.106)).

Since $\varphi(t) = o(t)$, we have that

$$\begin{aligned}\varphi(\rho(x,\, \lambda,\, \mu)) &= o(\rho(x,\, \lambda,\, \mu))\\ &= o(\rho(x,\, \lambda,\, \mu) - \rho(\bar{x},\, \bar{\lambda},\, \bar{\mu}))\\ &= o(\|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu})\|)\end{aligned} \tag{4.132}$$

as $(x,\, \lambda,\, \mu) \to (\bar{x},\, \bar{\lambda},\, \bar{\mu})$, where the last equality is by the local Lipschitz-continuity of ρ with respect to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ (that the latter holds under the stated assumptions can be easily shown by the mean-value theorem; see Theorem A.10). Then from (4.113), (4.114) we derive that for each $\theta_1 \in \varTheta_1(u)$ and $\theta_2 \in \varTheta_2(u)$ it holds that

$$\theta_1 = o(\|u - \bar{u}\|), \quad \theta_2 = o(\|u - \bar{u}\|) \tag{4.133}$$

as $u \to \bar{u}$. Combining this with (4.131), and employing (4.130), we obtain the needed estimate (4.126), thus establishing property (b).

Moreover, assuming that the second derivatives of f, h, and g are locally Lipschitz-continuous with respect to $\bar{x}$, that the estimates (4.119) and (4.120) can be replaced by (4.123) and (4.124), and that $\varphi(t) = O(t^2)$, the right-hand side of (4.131) can be replaced by $O(\|v\|^2 + \|v\|\|u - \bar{u}\| + \|u - \bar{u}\|^2)$ and the right-hand side of (4.133) by $O(\|u - \bar{u}\|^2)$. Instead of (4.126), we then obtain the sharper estimate

$$\omega = O(\|v\|^2 + \|u - \bar{u}\|^2),$$

and the last assertion of the theorem follows by the last assertion of Theorem 3.6. $\square$

It can be seen that the condition (4.118) in Theorem 4.24 can be relaxed as follows:

$$\left\langle \left(\frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) + \frac{\partial^2 \varPsi}{\partial \xi^2}((\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, (0,\, 0,\, 0)) \right) \xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}.$$

4.3.2 Augmented Lagrangian Modification and Truncated Sequential Quadratic Programming

As the first applications of the perturbed SQP framework developed in the previous section, we present the following two modifications of SQP: using the augmented Lagrangian (instead of the usual Lagrangian) to define the QP matrix H_k, and truncating the solution of subproblems.

Unlike in the equality-constrained case, the development of truncated SQP methods for problems with inequality constraints presents some conceptual difficulties; see the discussion in [106, 109]. If the QP (4.75) is solved by some finite active-set method, the truncation is hardly possible. This is because none of the iterates produced be an active-set method, except for the very last one at termination, can be expected to approximate the eventual solution in any reasonable sense (the algorithm "jumps" to a solution once the working active set is correct, rather than approaches a solution asymptotically). An alternative is to solve QPs approximately by some interior-point method, as suggested, e.g., in [147, 177]. This approach is justifiable for the specific issue of truncation, although there are still some practical challenges to deal with (e.g., warm-starting interior-point methods is not straightforward). Speaking about the theoretical side, for local analysis specifically within the perturbed SQP framework the difficulty is the necessity to avoid any truncation for the part of optimality conditions of the QP that involve the inequality constraints; see (4.110) and the discussion following it. As interior-point methods do not maintain complementarity (this is in fact the essence of this class of algorithms), conditions like (4.110) do not hold along the iterations of solving the QP subproblem. Thus, in the context of our line of analysis interior point methods must be applied with some modifications, as will be discussed below.

The modifications in question can be constructed for the special case of bound constraints. To that end, in this section we consider the problem (4.73) with equality constraints and simple (nonnegative) bounds only, the latter represented by the inequality constraints $g(x) = -x$, $x \in \mathbf{R}^n$. Specifically, the problem is

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \; x \geq 0. \end{array} \tag{4.134}$$

Note that general inequality constraints can be reformulated as equalities and bounds introducing slack variables.

Let $L : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}$ be the Lagrangian of problem (4.134), including only the equality constraints:

$$L(x, \lambda) = f(x) + \langle \lambda, h(x) \rangle.$$

We shall derive the local convergence result for the truncated version of SQP, covering at the same time the augmented Lagrangian choice of the QP matrix H_k, discussed in Sect. 4.2. Specifically, we set

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k, \tilde{\lambda}^k) + c(h'(x^k))^{\mathrm{T}} h'(x^k), \tag{4.135}$$

where $c \geq 0$ is the penalty parameter, and $\tilde{\lambda}^k$ is generated by the rule

$$\tilde{\lambda}^0 = \lambda^0, \quad \tilde{\lambda}^k = \lambda^k - ch(x^{k-1}), \ k = 1, 2, \ldots. \tag{4.136}$$

Note that the value $c = 0$ corresponding to the usual Lagrangian is also allowed. As discussed in Sect. 4.2, the Hessian of the augmented Lagrangian has a much better chance of being positive definite, which is an important feature for efficient solution of subproblems, as well as for globalization of convergence of the method; see Sect. 6.2.1.

For the current iterate $(x^k, \lambda^k, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$, the *truncated SQP method* for problem (4.134) generates the next iterate as follows: $(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ satisfies the system

$$\begin{gathered} \left\| f'(x^k) + H_k(x - x^k) + (h'(x^k))^{\mathrm{T}}\lambda - \mu \right\| \leq \varphi(\rho(x^k, \tilde{\lambda}^k, \mu^k)), \\ \|h(x^k) + h'(x^k)(x - x^k)\| \leq \varphi(\rho(x^k, \tilde{\lambda}^k, \mu^k)), \\ \mu \geq 0, \quad x \geq 0, \quad \langle \mu, x \rangle = 0, \end{gathered} \tag{4.137}$$

where H_k and $\tilde{\lambda}^k$ are given by (4.135), (4.136) with some $c \geq 0$; $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$ is a forcing function, and $\rho : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n \to \mathbf{R}_+$ is the natural residual for optimality conditions of the original problem (see (4.111)).

Some comments about the use of interior-point methods within truncated SQP are in order. Primal-dual interior-point methods applied to the QP subproblem generate a sequence of points $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ satisfying $x > 0$ and $\mu > 0$, and therefore, the last line of relations in (4.137) can never hold except in the limit. The suggestion is to employ the following simple *purification procedure*. For (x, λ, μ) produced by an iteration of the given interior-point method, define $\hat{x} \in \mathbf{R}^n$ and $\hat{\mu} \in \mathbf{R}^n$ (automatically satisfying the last line in (4.137)) by

$$\hat{x}_j = \begin{cases} x_j & \text{if } x_j \geq \mu_j, \\ 0 & \text{if } x_j < \mu_j, \end{cases} \quad \hat{\mu}_j = \begin{cases} 0 & \text{if } x_j \geq \mu_j, \\ \mu_j & \text{if } x_j < \mu_j, \end{cases} \quad j = 1, \ldots, n,$$

and verify (4.137) for $(\hat{x}, \lambda, \hat{\mu})$. If (4.137) is satisfied, accept it as the new iterate of the truncated SQP method. Otherwise proceed with the next iteration of the interior-point method for the QP subproblem to obtain a new (x, λ, μ), perform purification again, verify (4.137) for the purified point, etc. Under very reasonable assumptions, the interior-point method drives both (x, λ, μ) and its purification $(\hat{x}, \lambda, \hat{\mu})$ to the solution of the QP subproblem, and therefore, the next iterate satisfying (4.137) is obtained after a finite number of interior-point iterations if $\varphi(\rho(x^k, \tilde{\lambda}^k, \mu^k)) > 0$. See [147] for details.

Observe that (4.137) is precisely the system (4.110) with $g : \mathbf{R}^n \to \mathbf{R}^n$, $g(x) = -x$, with λ^k replaced by $\tilde{\lambda}^k$, with the variable λ shifted by $-ch(x^k)$, and with

$$\omega_1((x,\, \lambda,\, \mu),\, (\xi,\, \eta,\, \zeta)) = \omega_1(x,\, \xi) = c(h'(x))^{\mathrm{T}}(h(x) + h'(x)\xi),$$

$$\omega_2((x,\, \lambda,\, \mu),\, \xi) = 0, \quad \omega_3((x,\, \lambda,\, \mu),\, \xi) = 0$$

for $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ and $\xi \in \mathbf{R}^n$. Moreover, (4.137) is the "truncated" optimality system for the perturbed SQP subproblem (4.104), where one takes

$$\psi((x,\, \lambda,\, \mu),\, \xi) = \psi(x,\, \xi) = \frac{c}{2}\|h(x) + h'(x)\xi\|^2.$$

With these definitions of ψ, ω_1, ω_2, and ω_3, it holds that Ψ coincides with ψ (see (4.105)), and the equality (4.106) is valid.

Applying Theorem 4.24 (with λ^k substituted by $\tilde{\lambda}^k$!), we obtain a local convergence result for the truncated SQP method with (or without) the augmented Lagrangian modification of the QP matrix.

Theorem 4.25. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives continuous at $\bar{x}$. Let $\bar{x}$ be a local solution of problem (4.134), satisfying the SMFCQ and the SOSC for the associated Lagrange multiplier $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^n$.*

Then for any fixed $c \geq 0$, there exists $\delta > 0$ such that for any function $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$ such that $\varphi(t) = o(t)$ as $t \to 0$, and for any starting point $(x^0,\, \lambda^0,\, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, there exists a sequence $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ such that for each $k = 0,\, 1,\, \ldots,$ the triple $(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1})$ satisfies (4.122) and (4.137) with H_k and $\tilde{\lambda}^k$ defined by (4.135) and (4.136); any such sequence converges to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, the sequence $\{\tilde{\lambda}^k\}$ converges to $\bar{\lambda}$, and the rate of convergence of $\{(x^k,\, \tilde{\lambda}^k,\, \mu^k)\}$ is superlinear. Moreover, the rate of convergence of $\{(x^k,\, \tilde{\lambda}^k,\, \mu^k)\}$ is quadratic, if the second derivatives of f and h are locally Lipschitz-continuous with respect to $\bar{x}$, and if $\varphi(t) = O(t^2)$ as $t \to 0$.

Proof. The proof is by direct verification of the assumptions of Theorem 4.24. First, note that

$$\frac{\partial \psi}{\partial \xi}(\bar{x},\, 0) = \omega_1(\bar{x},\, 0) = 0.$$

Also, according to (4.86), for all $\xi \in C(\bar{x})$ the equality $h'(\bar{x})\xi = 0$ holds, and hence,

$$\frac{\partial^2 \psi}{\partial \xi^2}(\bar{x},\, 0)\xi = c(h'(\bar{x}))^{\mathrm{T}} h'(\bar{x})\xi = 0,$$

implying (4.118). Furthermore, if the second relation in (4.121) holds for some $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ and $\xi \in \mathbf{R}^n$ with ω_2 identically equal to zero, and with φ satisfying $\varphi(t) = o(t)$, then by (4.132) it follows that

$$\begin{aligned} \|h(x) + h'(x)\xi\| &\leq \varphi(\rho(x,\, \lambda,\, \mu)) \\ &= o(\|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu})\|) \end{aligned} \tag{4.138}$$

as $(x,\, \lambda,\, \mu) \to (\bar{x},\, \bar{\lambda},\, \bar{\mu})$. This implies the estimate

$$\begin{aligned}\frac{\partial \psi}{\partial \xi}(x,\, \xi) = \omega_1(x,\, \xi) &= c(h'(x))^{\mathrm{T}}(h(x) + h'(x)\xi) \\ &= o(\|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu})\|), \end{aligned} \tag{4.139}$$

which gives (4.120). Moreover, provided that $\varphi(t) = O(t^2)$, the estimate (4.138) can be sharpened as follows:

$$h(x) + h'(x)\xi = O(\|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu})\|^2). \tag{4.140}$$

We then obtain the sharper version of the estimate (4.139):

$$\frac{\partial \psi}{\partial \xi}(x,\, \xi) = O(\|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu})\|^2),$$

which gives (4.124).

All the assertions now follow from Theorem 4.24. □

4.3.3 More on Linearly Constrained Lagrangian Methods

Linearly constrained (augmented) Lagrangian methods discussed above in Sect. 4.1.2 for equality-constrained problems extend also to the case when inequality constraints are present. These methods are traditionally stated for problems with equality and bound constraints, which means that general inequality constraints are reformulated as equalities introducing slack variables; see [96, 205, 231]. This transformation is also applied in the MINOS software package [206]. We therefore follow the tradition and consider the optimization problem in the format of (4.134).

Iteration subproblems of the *linearly constrained Lagrangian* (LCL) *method* for (4.134) consist in minimizing the (augmented) Lagrangian subject to bounds and linearized equality constraints.

Algorithm 4.26 Choose $(x^0,\, \lambda^0,\, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ and set $k = 0$.

1. If $(x^k,\, \lambda^k,\, \mu^k)$ satisfies the KKT system (4.74), stop.
2. Choose $c_k \geq 0$ and compute $x^{k+1} \in \mathbf{R}^n$ and $(\eta^k,\, \mu^{k+1}) \in \mathbf{R}^l \times \mathbf{R}^n$ as a stationary point and an associated Lagrange multiplier of the problem

$$\begin{array}{ll} \text{minimize} & L(x,\, \lambda^k) + \dfrac{c_k}{2}\|h(x)\|^2 \\ \text{subject to} & h(x^k) + h'(x^k)(x - x^k) = 0,\ x \geq 0. \end{array} \tag{4.141}$$

3. Set $\lambda^{k+1} = \lambda^k + \eta^k$.
4. Increase k by 1 and go to step 1.

In the original LCL method proposed in [231], the subproblems involve the usual Lagrangian rather than the augmented Lagrangian, i.e., $c_k = 0$ for all k. However, in practice, it is often important to use $c_k > 0$ [206].

The KKT system defining stationary points and multipliers of subproblem (4.141) has the form

$$\begin{gathered}\frac{\partial L}{\partial x}(x,\, \lambda^k) + c_k(h'(x))^{\mathrm{T}}h(x) + (h'(x^k))^{\mathrm{T}}\eta - \mu = 0,\\ h(x^k) + h'(x^k)(x - x^k) = 0,\\ \mu \geq 0, \quad x \geq 0, \quad \langle \mu,\, x\rangle = 0,\end{gathered} \tag{4.142}$$

with the dual variables $\eta \in \mathbf{R}^l$ and $\mu \in \mathbf{R}^n$. It turns out that the subproblem (4.141) actually does not directly fit the assumptions of Theorem 4.24. Nevertheless, there is a way to make Theorem 4.24 applicable. To this end, consider the following modification of the subproblem (4.141), which is equivalent to it:

$$\begin{array}{ll}\text{minimize} & L(x,\, \lambda^k) - \langle \lambda^k,\, h'(x^k)(x - x^k)\rangle + \dfrac{c_k}{2}\|h(x)\|^2\\ \text{subject to} & h(x^k) + h'(x^k)(x - x^k) = 0,\ x \geq 0.\end{array} \tag{4.143}$$

The corresponding KKT system is then given by

$$\begin{gathered}\frac{\partial L}{\partial x}(x,\, \lambda^k) + c_k(h'(x))^{\mathrm{T}}h(x) + (h'(x^k))^{\mathrm{T}}(\lambda - \lambda^k) - \mu = 0,\\ h(x^k) + h'(x^k)(x - x^k) = 0,\\ \mu \geq 0, \quad x \geq 0, \quad \langle \mu,\, x\rangle = 0,\end{gathered} \tag{4.144}$$

with the dual variables $\lambda \in \mathbf{R}^l$ and $\mu \in \mathbf{R}^n$. Comparing (4.142) and (4.144), we observe that stationary points x^{k+1} of problems (4.141) and (4.143) coincide, and the associated multipliers are of the form $(\eta^k,\, \mu^{k+1})$ and $(\lambda^{k+1},\, \mu^{k+1})$, with $\lambda^{k+1} = \lambda^k + \eta^k$. Thus, for the purposes of convergence analysis, we can deal with the modified subproblems (4.143). It turns out that this does allow to apply Theorem 4.24.

For asymptotic analysis, we may consider that c_k is fixed at some value $c \geq 0$ for all k sufficiently large; this happens for typical update rules under natural assumptions (see, e.g., the discussion in [96]).

Observe first that the constraints of the subproblem (4.143) are exactly the same as in the SQP subproblem (4.75) for the optimization problem (4.73) in consideration. Structural perturbation that defines the LCL method within the perturbed SQP framework presented in Sect. 4.3.1 is therefore given by the objective function of (4.143). Specifically, the subproblem (4.143) can be seen as a particular case of the perturbed SQP subproblem (4.104), where for $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ and $\xi \in \mathbf{R}^n$ one takes

$$\begin{aligned}\psi((x,\,\lambda,\,\mu),\,\xi) &= \psi((x,\,\lambda),\,\xi)\\ &= L(x+\xi,\,\lambda) - \langle \lambda,\, h'(x)\xi\rangle + \frac{c}{2}\|h(x+\xi)\|^2\\ &\quad -f(x) - \langle f'(x),\,\xi\rangle - \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(x,\,\lambda)\xi,\,\xi\right\rangle\\ &= L(x+\xi,\,\lambda) - \left\langle \frac{\partial L}{\partial x}(x,\,\lambda),\,\xi\right\rangle - \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(x,\,\lambda)\xi,\,\xi\right\rangle\\ &\quad +\frac{c}{2}\|h(x+\xi)\|^2 - f(x), \end{aligned} \tag{4.145}$$

$$\omega_2((x,\,\lambda,\,\mu),\,\xi) = 0, \quad \omega_3((x,\,\lambda,\,\mu),\,\xi) = 0. \tag{4.146}$$

Employing the mean-value theorem (see Theorem A.10, (a)), one can directly verify that under the appropriate smoothness assumptions on f and h, the function ψ and the mappings ω_2 and ω_3 defined by (4.145), (4.146) possess all the properties required in Theorem 4.24 with $\varphi(\cdot) \equiv 0$. We omit the technical details, and only mention that

$$\frac{\partial^2 \psi}{\partial \xi^2}((\bar{x},\,\bar{\lambda}),\,0) = c(h'(\bar{x}))^{\mathrm{T}} h'(\bar{x}), \tag{4.147}$$

and if $\xi \in C(\bar{x})$, then according to (4.86) we have that $h'(\bar{x})\xi = 0$. Hence, (4.147) implies the equality $\frac{\partial^2 \psi}{\partial \xi^2}((\bar{x},\,\bar{\lambda}),\,0)\xi = 0$. Observe also that for $(x,\,\lambda) \in \mathbf{R}^n \times \mathbf{R}^l$ and $\xi \in \mathbf{R}^n$ it holds that

$$\begin{aligned}\frac{\partial \psi}{\partial \xi}((x,\,\lambda),\,\xi) &= \frac{\partial L}{\partial x}(x+\xi,\,\lambda) - \frac{\partial L}{\partial x}(x,\,\lambda) - \frac{\partial^2 L}{\partial x^2}(x,\,\lambda)\xi\\ &\quad +c(h'(x+\xi))^{\mathrm{T}} h(x+\xi)\\ &= c(h'(x+\xi))^{\mathrm{T}} h(x+\xi) + o(\|\xi\|) \end{aligned} \tag{4.148}$$

as $x \to \bar{x}$ and $\xi \to 0$. If the second relation in (4.121) holds with ω_2 and φ identically equal to zero, then (4.147) implies the estimate

$$\begin{aligned}\frac{\partial \psi}{\partial \xi}((x,\,\lambda),\,\xi) &= c(h'(x+\xi))^{\mathrm{T}}(h(x+\xi) - h(x) - h'(x)\xi) + o(\|\xi\|)\\ &= o(\|\xi\|) \end{aligned} \tag{4.149}$$

as $x \to \bar{x}$ and $\xi \to 0$, which gives (4.120). Moreover, under stronger smoothness assumptions, the last estimate can be sharpened as follows:

$$\frac{\partial \psi}{\partial \xi}((x,\,\lambda),\,\xi) = O(\|\xi\|^2 + \|\xi\|\|x-\bar{x}\|),$$

which gives (4.124). Applying Theorem 4.24, we obtain the following result on local convergence of the LCL method.

Theorem 4.27. *Under the assumptions of Theorem 4.25, for any fixed value $c \geq 0$, there exists a constant $\delta > 0$ such that for any starting point $(x^0, \lambda^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, there exists a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ such that for each $k = 0, 1, \ldots,$ the point x^{k+1} is a stationary point of problem (4.141) and $(\lambda^{k+1} - \lambda^k, \mu^{k+1})$ is an associated Lagrange multiplier satisfying (4.122); any such sequence converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the second derivatives of f and h are locally Lipschitz-continuous with respect to $\bar{x}$.*

Taking into account that (4.141) is not a quadratic programming problem, it makes good practical sense to solve the LCL subproblems approximately. It is then natural to introduce an extra perturbation associated with inexact solution of subproblems. Consider the *truncated LCL method*, where the system (4.142) is replaced by a version where the parts corresponding to general nonlinearities are relaxed:

$$\begin{gathered}\left\| \frac{\partial L}{\partial x}(x, \lambda^k) + c_k (h'(x))^{\mathrm{T}} h(x) + (h'(x^k))^{\mathrm{T}} \eta - \mu \right\| \leq \varphi(\rho(x^k, \lambda^k, \mu^k)), \\ \|h(x^k) + h'(x^k)(x - x^k)\| \leq \varphi(\rho(x^k, \lambda^k, \mu^k)), \\ \mu \geq 0, \quad x \geq 0, \quad \langle \mu, x \rangle = 0, \end{gathered} \tag{4.150}$$

with some forcing function $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$.

Similarly to the proof of Theorem 4.25, we conclude that if the second relation in (4.121) holds for some $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ and $\xi \in \mathbf{R}^n$ with ω_2 identically equal to zero, and with φ satisfying $\varphi(t) = o(t)$, then the estimate (4.138) is valid. Thus, instead of (4.149), we derive the estimate

$$\begin{aligned} \frac{\partial \psi}{\partial \xi}((x, \lambda), \xi) &= c(h'(x + \xi))^{\mathrm{T}}(h(x + \xi) - h(x) - h'(x)\xi) \\ &\quad + o(\|\xi\|) + o(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|) \\ &= o(\|\xi\|) + o(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|) \end{aligned} \tag{4.151}$$

as $(x, \lambda, \mu) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$ and $\xi \to 0$, which gives (4.120). Moreover, provided $\varphi(t) = O(t^2)$, the estimate (4.138) can be replaced by the sharper (4.140). Under stronger smoothness assumptions, we then obtain the sharper version of the estimate (4.151):

$$\frac{\partial \psi}{\partial \xi}((x, \lambda), \xi) = O(\|\xi\|^2 + \|\xi\|\|x - \bar{x}\|) + O(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|^2),$$

which gives (4.124). Local convergence of the truncated LCL method now follows from Theorem 4.24.

Theorem 4.28. *Under the assumptions of Theorem 4.25, for any fixed value $c \geq 0$, there exists $\delta > 0$ such that for any function $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$ such that*

$\varphi(t) = o(t)$ as $t \to 0$, and for any starting point $(x^0, \lambda^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, there exists a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ such that for each $k = 0, 1, \ldots,$ the triple $(x^{k+1}, \lambda^{k+1} - \lambda^k, \mu^{k+1})$ satisfies (4.150) and (4.122); any such sequence converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic if the second derivatives of f and h are locally Lipschitz-continuous with respect to $\bar{x}$, and if $\varphi(t) = O(t^2)$ as $t \to 0$.

Theorem 4.28 can be regarded as a generalization of Theorem 4.27.

Finally, an a posteriori result regarding primal superlinear convergence of the basic LCL method follows readily from Proposition 4.18, similarly to Theorem 4.12 for equality-constrained problems.

Theorem 4.29. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (4.134), and let $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^n$ be an associated noncritical Lagrange multiplier. Let $c \geq 0$ be fixed, and let an iterative sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ generated by Algorithm 4.26 with $c_k = c$ for all k large enough be convergent to $(\bar{x}, \bar{\lambda}, \bar{\mu})$.*

Then the rate of convergence of $\{x^k\}$ is superlinear.

Proof. The KKT system defining stationary points and multipliers of the subproblem (4.141) has the form

$$\begin{gathered}\frac{\partial L}{\partial x}(x, \lambda^k) + c(h'(x))^{\mathrm{T}} h(x) + (h'(x^k))^{\mathrm{T}} \eta - \mu = 0,\\ h(x^k) + h'(x^k)(x - x^k) = 0,\\ \mu \geq 0, \quad x \geq 0, \quad \langle \mu, x \rangle = 0,\end{gathered}$$

with the dual variables $\eta \in \mathbf{R}^l$ and $\mu \in \mathbf{R}^n$. Within the perturbed SQP framework (4.81), this corresponds to setting

$$\begin{aligned}\omega_1^k &= \frac{\partial L}{\partial x}(x^{k+1}, \lambda^k) + c(h'(x^{k+1}))^{\mathrm{T}} h(x^{k+1}) + (h'(x^k))^{\mathrm{T}} \eta^k - \mu^{k+1}\\ &\quad - \frac{\partial L}{\partial x}(x^k, \lambda^k) + \mu^k - \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)(x^{k+1} - x^k)\\ &\quad - (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (\mu^{k+1} - \mu^k)\\ &= \frac{\partial L}{\partial x}(x^{k+1}, \lambda^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k) - \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)(x^{k+1} - x^k)\\ &\quad - (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k - \eta^k) + c(h'(x^{k+1}))^{\mathrm{T}} h(x^{k+1})\\ &= c(h'(\bar{x}))^{\mathrm{T}} h(x^{k+1}) + o(\|x^{k+1} - x^k\|) + o(\|x^k - \bar{x}\|)\end{aligned} \tag{4.152}$$

as $k \to \infty$, and

$$\omega_2^k = 0, \quad \omega_3^k = 0.$$

Since $\operatorname{im}(h'(\bar{x}))^{\mathrm{T}} = (\ker h'(\bar{x}))^{\perp}$, it follows that the first term in the right-hand side of (4.152) is orthogonal to the subspace $\ker h'(\bar{x})$ containing $C_+(\bar{x}, \bar{\mu})$. This implies (4.99), and the stated assertion follows from Proposition 4.18. $\square$

Note that Theorem 4.29 implies, in particular, that under the same assumptions needed for the primal-dual superlinear convergence of LCL methods (the SMFCQ and the SOSC), we have the primal superlinear convergence as well. Again, under these assumptions the LCL method converges (locally), and so the corresponding assumption in Theorem 4.29 is not needed in this case, if the starting point is close enough to the primal-dual solution in question.

4.3.4 Inexact Restoration Methods

Similarly to the LCL algorithm discussed above, inexact restoration methods are traditionally stated for problems with equality and bound constraints. Therefore, our problem setting in this section remains that of (4.134). We consider inexact restoration methods along the lines of the local framework in [21]. For some other references, see [78, 86, 192, 193], and [194] for a survey.

Yet again, our line of analysis is by interpreting those methods within the perturbed SQP framework of Sect. 4.3.1. This approach is from [159].

It is convenient and instructive to start with the *exact restoration method*, which is not a practical algorithm but rather a motivation for inexact restoration.

Algorithm 4.30 Choose $(x^0,\, \lambda^0,\, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ and set $k = 0$.

1. If $(x^k,\, \lambda^k,\, \mu^k)$ satisfies the KKT system (4.74), stop.
2. (Feasibility phase.) Compute π^k as a projection of x^k onto the feasible set of the problem (4.134), i.e., a global solution of the subproblem

$$\begin{array}{ll} \text{minimize} & \|\pi - x^k\| \\ \text{subject to} & h(\pi) = 0,\ \pi \geq 0. \end{array} \tag{4.153}$$

3. If $(\pi^k,\, \lambda^k,\, \mu^k)$ satisfies the KKT system (4.74), stop.
4. (Optimality phase.) Compute x^{k+1} and $(\eta^k,\, \mu^{k+1})$ as a stationary point and an associated Lagrange multiplier of the subproblem

$$\begin{array}{ll} \text{minimize} & L(x,\, \lambda^k) \\ \text{subject to} & h'(\pi^k)(x - \pi^k) = 0,\ x \geq 0. \end{array} \tag{4.154}$$

5. Set $\lambda^{k+1} = \lambda^k + \eta^k$.
6. Increase k by 1 and go to step 1.

Similarly to the LCL method in Sect. 4.3.3, (in)exact restoration does not directly fit the perturbed SQP framework in terms of the assumptions imposed on it in Theorem 4.24. But there is an equivalent transformation that does the job. To that end, we replace the optimality phase subproblem (4.154) by the following:

$$\begin{array}{ll} \text{minimize} & L(x,\,\lambda^k) - \langle \lambda^k,\, h'(\pi^k)(x - x^k)\rangle \\ \text{subject to} & h'(\pi^k)(x - \pi^k) = 0,\ x \geq 0. \end{array} \tag{4.155}$$

It is easily seen that stationary points x^{k+1} of the problems (4.154) and (4.155) coincide, and the associated multipliers are of the form $(\eta^k,\, \mu^{k+1})$ and $(\lambda^{k+1},\, \mu^{k+1})$, respectively, with $\lambda^{k+1} = \lambda^k + \eta^k$. Thus, for the purposes of convergence analysis, we can deal with the modified subproblems (4.155).

For a given $x \in \mathbf{R}^n$, let $\bar{\pi}(x)$ be a projection of x onto the feasible set of the problem (4.134), computed as at the feasibility phase of Algorithm 4.30 for $x = x^k$. In order to formally apply Theorem 4.24, we need to assume that $\bar{\pi}(\cdot)$ is a fixed single-valued function. As the feasible set is not convex, the projection onto it need not be unique, in general. However, an algorithm used to solve (4.153) follows its internal patterns and computes one specific projection. It is further reasonable to assume that, if at some future iteration projection of the same point needs to be computed, the algorithm would return the same result. With this in mind, considering that $\bar{\pi}(\cdot)$ is a single-valued function is justified for all practical purposes. In Theorem 4.31 below this assumption is stated more formally.

Then the subproblem (4.155) can be seen as a particular case of the perturbed SQP subproblem (4.104), where for $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ and $\xi \in \mathbf{R}^n$ one takes

$$\begin{aligned} \psi((x,\,\lambda,\,\mu),\,\xi) &= \psi((x,\,\lambda),\,\xi) \\ &= L(x+\xi,\,\lambda) - \langle \lambda,\, h'(\bar{\pi}(x))\xi\rangle \\ &\quad - f(x) - \langle f'(x),\,\xi\rangle - \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(x,\,\lambda)\xi,\,\xi\right\rangle \\ &= L(x+\xi,\,\lambda) - \left\langle \frac{\partial L}{\partial x}(x,\,\lambda),\,\xi\right\rangle - \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(x,\,\lambda)\xi,\,\xi\right\rangle \\ &\quad - \langle \lambda,\, (h'(\bar{\pi}(x)) - h'(x))\xi\rangle - f(x), \end{aligned} \tag{4.156}$$

$$\begin{aligned} \omega_2((x,\,\lambda,\,\mu),\,\xi) &= \omega_2(x,\,\xi) \\ &= h'(\bar{\pi}(x))(x + \xi - \bar{\pi}(x)) - h(x) - h'(x)\xi, \end{aligned} \tag{4.157}$$

$$\omega_3((x,\,\lambda,\,\mu),\,\xi) = 0. \tag{4.158}$$

The function ψ and the mappings ω_2, ω_3 defined by (4.156)–(4.158) satisfy all the requirements in Theorem 4.24 with $\varphi(\cdot) \equiv 0$, provided f and h are smooth enough. We omit technical verification of this (which is direct), and only mention the following. For $\Psi((x,\,\lambda,\,\mu),\,(\xi,\,\eta,\,\zeta)) = \Psi((x,\,\lambda),\,(\xi,\,\eta))$ defined according to (4.105), it holds that

$$\frac{\partial^2 \Psi}{\partial \xi^2}((\bar{x},\,\bar{\lambda}),\,(0,\,0)) = 0.$$

Furthermore, it is clear that $\|\bar{\pi}(x) - x\| \leq \|x - \bar{x}\|$ for any $x \in \mathbf{R}^n$, and in particular, $\bar{\pi}(x) \to \bar{x} = \bar{\pi}(\bar{x})$ as $x \to \bar{x}$. Then employing the mean-value theorem (see Theorem A.10, (a)) it can be seen that

$$\begin{aligned}\frac{\partial \Psi}{\partial \xi}((x,\, \lambda),\, (\xi,\, \eta)) &= \frac{\partial L}{\partial x}(x + \xi,\, \lambda) - \frac{\partial L}{\partial x}(x,\, \lambda) - \frac{\partial^2 L}{\partial x^2}(x,\, \lambda)\xi \\ &\quad + (h'(\bar{\pi}(x)) - h'(x))^{\mathrm{T}}\eta \\ &= o(\|\xi\|) + O(\|\eta\|\|x - \bar{x}\|), \end{aligned} \tag{4.159}$$

$$\begin{aligned}\omega_2(x,\, \xi) &= (h'(\bar{\pi}(x)) - h'(x))\xi + h(\bar{\pi}(x)) - h(x) - h'(\bar{\pi}(x))(\bar{\pi}(x) - x) \\ &= O(\|\xi\|\|x - \bar{x}\| + \|x - \bar{x}\|^2) \end{aligned} \tag{4.160}$$

as $(x,\, \lambda) \to (\bar{x},\, \bar{\lambda})$ and $(\xi,\, \eta) \to (0,\, 0)$. Moreover, under stronger smoothness assumptions, the first estimate can be sharpened as follows:

$$\frac{\partial \Psi}{\partial \xi}((x,\, \lambda),\, (\xi,\, \eta)) = O(\|\xi\|^2 + (\|\xi\| + \|\eta\|)\|x - \bar{x}\|). \tag{4.161}$$

Applying now Theorem 4.24, we obtain local superlinear convergence of the exact restoration scheme.

Theorem 4.31. *Under the hypotheses of Theorem 4.25, assume that whenever $x^k = x^j$ for any two iteration indices k and j, then step 2 of Algorithm 4.30 computes $\pi^k = \pi^j$.*

Then there exists a constant $\delta > 0$ such that for any starting point $(x^0,\, \lambda^0,\, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, there exists a sequence $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ of Algorithm 4.30, such that (4.122) holds for each $k = 0,\, 1,\, \ldots$; any such sequence converges to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the second derivatives of f and h are locally Lipschitz-continuous with respect to $\bar{x}$.

An a posteriori result on primal superlinear convergence of Algorithm 4.30 follows readily from Proposition 4.18.

Theorem 4.32. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (4.134), and let $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^n$ be an associated noncritical Lagrange multiplier. Let an iterative sequence $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ generated by Algorithm 4.30 be convergent to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$.*

Then the rate of convergence of $\{x^k\}$ is superlinear.

Proof. From the discussion above, and in particular, from (4.159) and (4.160), it follows that (4.81) holds for each k with

$$\begin{aligned}\omega_1^k &= \frac{\partial \Psi}{\partial \xi}((x^k,\, \lambda^k),\, (x^{k+1} - x^k,\, \lambda^{k+1} - \lambda^k))\\ &= o(\|x^{k+1} - x^k\|) + O(\|\lambda^{k+1} - \lambda^k)\|\|x^k - \bar{x}\|)\\ &= o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|),\end{aligned}$$

$$\begin{aligned}\omega_2^k &= \omega_2(x^k,\, x^{k+1} - x^k)\\ &= O(\|x^{k+1} - x^k\|\|x^k - \bar{x}\| + \|x^k - \bar{x}\|^2)\\ &= o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)\end{aligned}$$

as $k \to \infty$, and with $\omega_3^k = 0$. The assertion now follows from Proposition 4.18.
□

Clearly, the exact restoration scheme of Algorithm 4.30 cannot be anything more than a conceptual motivation, as solving the subproblems (4.153) and (4.154) exactly is impractical (in most cases, simply impossible). Therefore, the question is what kind of inexactness can be allowed when solving these subproblems, while still maintaining the local convergence and rate of convergence properties. To that end, consider the following framework, which we refer to as the *inexact restoration method*.

Algorithm 4.33 Choose functions $\varphi_0,\, \varphi_1,\, \varphi_2 : \mathbf{R}_+ \to \mathbf{R}_+$. Choose a starting point $(x^0,\, \lambda^0,\, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ and set $k = 0$.

1. If $(x^k,\, \lambda^k,\, \mu^k)$ satisfies the KKT system (4.74), stop.
2. (Feasibility phase.) Compute $\pi^k \in \mathbf{R}^n$ satisfying

$$\|h(\pi)\| \le \varphi_0(\|h(x^k)\|), \quad \pi \ge 0. \tag{4.162}$$

3. If $(\pi^k,\, \lambda^k,\, \mu^k)$ satisfies the KKT system (4.74), stop.
4. (Optimality phase.) Compute x^{k+1} and $(\eta^k,\, \mu^{k+1})$ satisfying

$$\left\|\frac{\partial L}{\partial x}(x,\, \lambda^k) + (h'(\pi^k))^{\mathrm{T}}\eta - \mu\right\| \le \varphi_1\left(\left\|\frac{\partial L}{\partial x}(\pi^k,\, \lambda^k) - \mu^k\right\|\right), \tag{4.163}$$

$$\|h'(\pi^k)(x - \pi^k)\| \le \varphi_2\left(\left\|\frac{\partial L}{\partial x}(\pi^k,\, \lambda^k) - \mu^k\right\|\right), \tag{4.164}$$

$$\mu \ge 0, \quad x \ge 0, \quad \langle \mu,\, x\rangle = 0. \tag{4.165}$$

1. Set $\lambda^{k+1} = \lambda^k + \eta^k$.
2. Increase k by 1 and go to step 1.

In the analysis below, it is further assumed that π^k computed at the feasibility phase of Algorithm 4.33 is within a controllable distance from x^k. Specifically, in addition to (4.162), it must satisfy the localization condition

$$\|\pi - x^k\| \le K\|(x^k - \bar{x},\, \lambda^k - \bar{\lambda},\, \mu^k - \bar{\mu})\| \tag{4.166}$$

for some $K > 0$ independent of $(x^k,\, \lambda^k,\, \mu^k)$. In practice, this can be achieved, e.g., by approximately solving the subproblem (4.153), or by other feasibility restoration strategies. In [21], (4.166) is replaced by a "more practical" condition

$$\|\pi - x^k\| \leq \tilde{K}\|h(x^k)\| \tag{4.167}$$

with some $\tilde{K} > 0$, which certainly implies (4.166) with $K = \ell\tilde{K}$ if h is Lipschitz-continuous with respect to $\bar{x}$ with constant $\ell > 0$. As a practical implementation of (4.167), in [21] it is suggested to fix $\tilde{K} > 0$ as a parameter of the algorithm, and to compute π^k as an approximate solution of the subproblem

$$\begin{array}{ll} \text{minimize} & \|h(\pi)\|^2 \\ \text{subject to} & \|\pi - x^k\| \leq \tilde{K}\|h(x^k)\|,\ \pi \geq 0. \end{array}$$

An acceptable approximate solution of this subproblem must satisfy its constraints and the condition (4.162). However, the difficulty with this approach is that for a given $\tilde{K}$, approximate solutions of this kind may not exist, in which case the algorithm in [21] simply declares failure at the feasibility phase. For this reason, we do not employ (4.167) as an actual constraint of the feasibility subproblem of the algorithm, and instead use the condition (4.166) as an ingredient of the analysis.

Employing (4.162)–(4.165), by the previous discussion of the exact restoration scheme, an iteration of Algorithm 4.33 can be interpreted in the perturbed SQP framework as (4.110) with ω_1 defined by (4.105), (4.106), where ψ is given by (4.156), with ω_2 and ω_3 defined by (4.157), (4.158), and with an appropriate φ. Moreover, according to Remark 4.23, the right-hand sides of the first two relations in (4.110) can be replaced by $\chi_1(x^k, \lambda^k, \mu^k)$ and $\chi_2(x^k,\, \lambda^k,\, \mu^k)$, respectively, where the functions $\chi_1 : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n \to \mathbf{R}_+$ and $\chi_2 : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n \to \mathbf{R}_+$ are defined by

$$\chi_1(x,\, \lambda,\, \mu) = \varphi_1\left(\left\|\frac{\partial L}{\partial x}(\pi(x,\, \lambda,\, \mu),\, \lambda) - \mu\right\|\right), \tag{4.168}$$

$$\chi_2(x,\, \lambda,\, \mu) = \varphi_2\left(\left\|\frac{\partial L}{\partial x}(\pi(x,\, \lambda,\, \mu),\, \lambda) - \mu\right\|\right). \tag{4.169}$$

In the above, $\pi(x,\, \lambda,\, \mu)$ is a point selected as at the feasibility phase of Algorithm 4.33 for $(x,\, \lambda,\, \mu) = (x^k,\, \lambda^k,\, \mu^k)$. Observe that for every point $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$, the point $\pi = \bar{x}$ satisfies both of the following two conditions:

$$\|h(\pi)\| \leq \varphi_0(\|h(x)\|), \quad \pi \geq 0,$$

and

$$\|\pi - x\| \leq K\|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu})\| \tag{4.170}$$

(cf. (4.162) and (4.166)) with any $K \geq 1$, and hence, can be selected as $\pi(x, \lambda, \mu)$. Therefore, $\pi(x, \lambda, \mu)$ with the required properties always exists.

Similarly to the case of exact restoration, we can reasonably assume that if step 2 of Algorithm 4.33 is applied at equal primal-dual points on different iterations, then the same result is produced. In particular, $\pi(\cdot)$ is a fixed single-valued function.

Assuming the differentiability of h at $\bar{x}$, it holds that

$$h(x) = h(x) - h(\bar{x}) = O(\|x - \bar{x}\|) \tag{4.171}$$

as $x \to \bar{x}$. Moreover, assuming the twice differentiability of f and h at $\bar{x}$, and taking into account (4.170), we have that

$$\begin{aligned}\left\|\frac{\partial L}{\partial x}(\pi(x, \lambda, \mu), \lambda) - \mu\right\| &\leq \left\|\frac{\partial L}{\partial x}(x, \lambda) - \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda})\right\| + \|\mu - \bar{\mu}\| \\ &\quad + \left\|\frac{\partial L}{\partial x}(\pi(x, \lambda, \mu), \lambda) - \frac{\partial L}{\partial x}(x, \lambda)\right\| \\ &= O(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|) \\ &\quad + O(\|\pi(x, \lambda, \mu) - x\|) \\ &= O(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|) \end{aligned} \tag{4.172}$$

as $(x, \lambda, \mu) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$.

We then derive, from our general principles, the following.

Theorem 4.34. *Under the hypotheses of Theorem 4.25, assume that whenever $(x^k, \lambda^k, \mu^k) = (x^j, \lambda^j, \mu^j)$ for any two iteration indices k and j, then step 2 of Algorithm 4.33 computes $\pi^k = \pi^j$.*

Then for any functions $\varphi_0, \varphi_1, \varphi_2 : \mathbf{R}_+ \to \mathbf{R}_+$ such that $\varphi_0(t) = o(t)$, $\varphi_1(t) = o(t)$ and $\varphi_2(t) = o(t)$ as $t \to 0$, and any $K \geq 1$, there exists $\delta > 0$ such that for any starting point $(x^0, \lambda^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, there exists a sequence $\{(x^k, \pi^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ of Algorithm 4.33 such that

$$\|\pi^k - x^k\| \leq K\|(x^k - \bar{x}, \lambda^k - \bar{\lambda},\ \mu^k - \bar{\mu})\|$$

and (4.122) holds for each $k = 0, 1, \ldots$; for any such sequence the part $\{(x^k, \lambda^k, \mu^k)\}$ converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic if the second derivatives of f and h are locally Lipschitz-continuous with respect to $\bar{x}$, and if $\varphi_0(t) = O(t^2)$, $\varphi_1(t) = O(t^2)$ and $\varphi_2(t) = O(t^2)$ as $t \to 0$.

Proof. Combining (4.168), (4.169) with (4.171), (4.172), and taking into account the assumptions regarding φ_0, φ_1, and φ_2, we obtain that

$$\chi_j(x, \lambda, \mu) = o(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|), \quad j = 1, 2,$$

as $(x, \lambda, \mu) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$. Moreover, if $\varphi_0(t) = O(t^2)$, $\varphi_1(t) = O(t^2)$ and $\varphi_2(t) = O(t^2)$, then

$$\chi_j(x,\, \lambda,\, \mu) = O(\|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu})\|^2), \quad j = 1,\, 2.$$

Observe now that all the previous considerations concerning the exact restoration scheme remain valid if we replace $\bar{\pi}(x)$ by $\pi(x,\, \lambda,\, \mu)$ (with the evident modifications of estimates (4.159)–(4.161), where one should replace $\|x - \bar{x}\|$ by $\|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu})\|$). This claim follows from (4.162), (4.166), and (4.95), implying, in particular, that $\pi(x,\, \lambda,\, \mu) \to \bar{x} = \pi(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ and

$$h(\pi(x,\, \lambda,\, \mu)) = o(\|h(x)\|) = o(\|x - \bar{x}\|)$$

as $(x,\, \lambda,\, \mu) \to (\bar{x},\, \bar{\lambda},\, \bar{\mu})$.

All the assertions now follow by applying Theorem 4.24. □

4.3.5 Sequential Quadratically Constrained Quadratic Programming

The *sequential quadratically constrained quadratic programming* (SQCQP) *method* passes second-order information about the constraints directly to the constraints of the subproblem, rather than to its objective function as in the SQP approach. Specifically, for the general equality and inequality-constrained problem (4.73), SQCQP is the following algorithm.

Algorithm 4.35 Choose $(x^0,\, \lambda^0,\, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and set $k = 0$.

1. If $(x^k,\, \lambda^k,\, \mu^k)$ satisfies the KKT system (4.74), stop.
2. Compute $x^{k+1} \in \mathbf{R}^n$ and $(\lambda^{k+1},\, \mu^{k+1}) \in \mathbf{R}^l \times \mathbf{R}^m$ as a stationary point and an associated Lagrange multiplier of the problem

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, x - x^k\rangle + \dfrac{1}{2} f''(x^k)[x - x^k,\, x - x^k] \\ \text{subject to} & h(x^k) + h'(x^k)(x - x^k) + \dfrac{1}{2} h''(x^k)[x - x^k,\, x - x^k] = 0, \\ & g(x^k) + g'(x^k)(x - x^k) + \dfrac{1}{2} g''(x^k)[x - x^k,\, x - x^k] \le 0. \end{array} \tag{4.173}$$

3. Increase k by 1 and go to step 1.

We note, in the passing, that SQCQP is in principle a primal algorithm, since dual variables are not used to formulate the subproblems: the primal sequence $\{x^k\}$ can be generated independently of the dual sequence $\{(\lambda^k,\, \mu^k)\}$. Nevertheless, dual behavior is certainly important, as most standard stopping criteria for the problem (4.73) are based on some primal-dual measure of optimality, such as the natural residual (4.111) of the KKT system (4.74).

The subproblem (4.173) of the SQCQP method has a quadratic objective function and *quadratic constraints*, and is generally more difficult to solve than, say, the SQP subproblem. This is probably the reason why SQCQP methods were not attracting serious attention until relatively recently [9, 74, 97, 173, 215, 248, 266]. In the convex case, subproblem (4.173) can be cast as a second-order cone program [182, 207], which can be solved efficiently by interior-point algorithms. Another possibility for the convex case is [124]. In [9], nonconvex subproblems were also handled quite efficiently by using other computational tools. In the nonconvex case, for solving the subproblems one may also use the approaches from [3, 15].

The subproblem (4.173) can be seen as a particular case of the perturbed SQP subproblem (4.104), where for $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and $\xi \in \mathbf{R}^n$ one takes

$$\psi((x,\, \lambda,\, \mu),\, \xi) = -\frac{1}{2}\langle \lambda,\, h''(x)[\xi,\, \xi]\rangle - \frac{1}{2}\langle \mu,\, g''(x)[\xi,\, \xi]\rangle, \tag{4.174}$$

$$\omega_2((x,\, \lambda,\, \mu),\, \xi) = \omega_2(x,\, \xi) = \frac{1}{2}h''(x)[\xi,\, \xi], \tag{4.175}$$

$$\omega_3((x,\, \lambda,\, \mu),\, \xi) = \omega_3(x,\, \xi) = \frac{1}{2}g''(x)[\xi,\, \xi]. \tag{4.176}$$

It can be verified directly that under appropriate smoothness assumptions on f, h and g, the function ψ and the mappings ω_2 and ω_3 defined by (4.174)–(4.176) possess all the properties required in Theorem 4.24, whatever is taken as φ. In particular, for the function Ψ defined by (4.105), it holds that $\frac{\partial^2 \Psi}{\partial \xi^2}((\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, (0,\, 0,\, 0)) = 0$, and

$$\omega_2(x,\, \xi) = O(\|\xi\|^2), \quad \omega_3(x,\, \xi) = O(\|\xi\|^2), \tag{4.177}$$

$$\begin{aligned}\frac{\partial \Psi}{\partial \xi}((x,\, \lambda,\, \mu),\, (\xi,\, \eta,\, \zeta)) &= (h''(x)[\xi])^{\mathrm{T}}\eta + (g''(x)[\xi])^{\mathrm{T}}\zeta \\ &= O(\|\xi\|(\|\eta\| + \|\zeta\|))\end{aligned} \tag{4.178}$$

as $(x,\, \lambda,\, \mu)$ tends to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, and as $(\xi,\, \eta,\, \zeta)$ tends to zero, which gives (4.123), (4.124). Applying Theorem 4.24 with $\varphi(\cdot) \equiv 0$, we obtain the following result on local convergence of the SQCQP method.

Theorem 4.36. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$, and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a local solution of problem (4.73), satisfying the SMFCQ and the SOSC (4.100) for the associated Lagrange multiplier $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$.*

Then there exists a constant $\delta > 0$ such that for any starting point $(x^0,\, \lambda^0,\, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, there exists a sequence $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ such that for each $k = 0,\, 1,\, \ldots$, the point x^{k+1} is a stationary point of problem (4.173) and $(\lambda^{k+1},\, \mu^{k+1})$ is an *associated Lagrange multiplier satisfying* (4.122)*; any such sequence converges to* $(\bar{x}, \bar{\lambda}, \bar{\mu})$*, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the second derivatives of* f*,* h*, and* g *are locally Lipschitz-continuous with respect to* $\bar{x}$*.*

We complete this section with an a posteriori result regarding the primal superlinear convergence of the SQCQP method, which follows immediately from Proposition 4.18, similarly to the proof of Theorem 4.29 but employing the estimates (4.177), (4.178).

Theorem 4.37. *Let* $f : \mathbf{R}^n \to \mathbf{R}$*,* $h : \mathbf{R}^n \to \mathbf{R}^l$ *and* $g : \mathbf{R}^n \to \mathbf{R}^m$ *be twice differentiable in a neighborhood of a point* $\bar{x} \in \mathbf{R}^n$*, with their second derivatives being continuous at* $\bar{x}$*. Let* $\bar{x}$ *be a stationary point of problem* (4.73)*, and let* $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ *be an associated noncritical Lagrange multiplier. Let an iterative sequence* $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ *generated by Algorithm 4.35 be convergent to* $(\bar{x}, \bar{\lambda}, \bar{\mu})$*.*

Then the rate of convergence of $\{x^k\}$ *is superlinear.*

Proof. It can be readily seen that the KKT conditions of the subproblem (4.173) correspond to the perturbed SQP framework (4.81) with

$$\begin{aligned}\omega_1^k &= (h''(x^k)[x^{k+1} - x^k])^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g''(x^k)[x^{k+1} - x^k])^{\mathrm{T}}(\mu^{k+1} - \mu^k)\\ &= o(\|x^{k+1} - x^k\|),\end{aligned}$$

$$\omega_2^k = \frac{1}{2}h''(x^k)[x^{k+1} - x^k,\, x^{k+1} - x^k] = O(\|x^{k+1} - x^k\|^2),$$

$$\omega_3^k = \frac{1}{2}g''(x^k)[x^{k+1} - x^k,\, x^{k+1} - x^k] = O(\|x^{k+1} - x^k\|^2)$$

as $k \to \infty$. The needed result follows immediately from Proposition 4.18. □

4.3.6 Second-Order Corrections

Second-order corrections, originally proposed in [88] and some other concurrent works (see [29, Sect. 17, p. 310], [208, Sect. 18, commentary] for related references), are special perturbations of the Newton-type steps for problem (4.73), aimed at “improving feasibility” (which usually actually means reducing infeasibility) of the next primal iterate. Such corrections must be small relative to the main step to avoid interference with high convergence rate provided by the latter. As will be demonstrated in Sect. 6.2.2, second-order corrections are an important ingredient of practical Newton-type algorithms, as they allow to guarantee global convergence without sacrificing the local superlinear convergence rate. On the other hand, in Sect. 4.3.7, second-order corrections will help to preserve feasibility of the primal iterates, which is crucial for the method discussed there.

We shall describe the second-order corrections technique for the equality and inequality-constrained problem (4.73), assuming that we have at our disposal some tools to eventually identify inequality constraints active at the solution. Specifically, for a given stationary point $\bar{x} \in \mathbf{R}^n$ of problem (4.73) and a given associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$, we assume that for the current iterate $(x^k, \lambda^k, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ we can define the index set $A_k = A_k(x^k, \lambda^k, \mu^k) \subset \{1, \ldots, m\}$ in such a way that $A_k = A(\bar{x})$ provided (x^k, λ^k, μ^k) is close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$. For specific examples of such identification tools, see Sects. 3.4.2, 4.1.3, 4.3.7.

Let $p^k \in \mathbf{R}^n$ be the primal direction produced at (x^k, λ^k, μ^k) by some Newton-type method, i.e., the Newtonian step would consist of taking the next iterate as $x^{k+1} = x^k + p^k$. But, instead, we shall compute another direction $\bar{p}^k \in \mathbf{R}^n$ as the minimum-norm solution to the linear system

$$h(x^k + p^k) + h'(x^k)p = 0, \quad g_{A_k}(x^k + p^k) + g'_{A_k}(x^k)p = -w_k e, \tag{4.179}$$

where e is the vector of ones of an appropriate dimension, and $w_k \geq 0$ is a scalar whose role (when it is positive) is to make $g'_{A_k}(x^k)\bar{p}^k$ "more negative," therefore pointing more to the interior of the domain defined by the inequality-constraints. Then we set $x^{k+1} = x^k + p^k + \bar{p}^k$.

The direction $\bar{p}^k$ can be computed as the minimum-norm solution of (4.179) with respect to different norms. Consider, e.g., the Euclidean norm, in which case $\bar{p}^k$ is defined by the quadratic programming problem

$$\begin{array}{ll} \text{minimize} & \|p\|^2 \\ \text{subject to} & h(x^k + p^k) + h'(x^k)p = 0,\ g_{A_k}(x^k + p^k) + g'_{A_k}(x^k)p = -w_k e. \end{array} \tag{4.180}$$

Lemma 4.38. *Let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their derivatives being continuous at $\bar{x}$. Assume that $h(\bar{x}) = 0$, $g(\bar{x}) \leq 0$, and that $\bar{x}$ satisfies the LICQ.*

For arbitrary $p^k \in \mathbf{R}^n$ and $w_k \in \mathbf{R}$, if $x^k \in \mathbf{R}^n$ is close enough to $\bar{x}$, then problem (4.180) with $A_k = A(\bar{x})$ has the unique solution $\bar{p}^k \in \mathbf{R}^n$, and

$$\bar{p}^k = O(\|h(x^k + p^k)\| + \|g_{A(\bar{x})}(x^k + p^k)\| + w_k) \tag{4.181}$$

as $x^k \to \bar{x}$, uniformly in $p^k \in \mathbf{R}^n$ and $w_k \in \mathbf{R}$. In particular, if the derivatives of h and g are locally Lipschitz-continuous at $\bar{x}$, and p^k satisfies the equalities

$$h(x^k) + h'(x^k)p^k = 0, \quad g_{A(\bar{x})}(x^k) + g'_{A(\bar{x})}(x^k)p^k = 0, \tag{4.182}$$

then

$$h(x^k + p^k) = O(\|p^k\|^2), \quad g_{A(\bar{x})}(x^k + p^k) = O(\|p^k\|^2),$$
$$\bar{p}^k = O(\|p^k\|^2 + w_k) \tag{4.183}$$

as $x^k \to \bar{x}$ and $p^k \to 0$, uniformly in $w_k \in \mathbf{R}$.

(The last estimate of this lemma justifies the name "second-order corrections.")

Proof. The LICQ at $\bar{x}$ implies that for all $x^k \in \mathbf{R}^n$ close enough to $\bar{x}$

$$\operatorname{rank} J_k = l + |A(\bar{x})|, \tag{4.184}$$

where

$$J_k = \begin{pmatrix} h'(x^k) \\ g'_{A(\bar{x})}(x^k) \end{pmatrix}.$$

Then the constraints in (4.180) are consistent (there exist feasible points). Hence, by Proposition 1.5, this problem with linear constraints and a strongly convex quadratic objective function has the unique solution $\bar{p}^k \in \mathbf{R}^n$.

Condition (4.184) evidently implies that the matrix $J_k J_k^{\mathrm{T}}$ is nonsingular. Applying Theorem 1.11 to problem (4.180) at $\bar{p}^k$, it can be easily seen that

$$\bar{p}^k = J_k^{\mathrm{T}}(J_k J_k^{\mathrm{T}})^{-1}(h(x^k + p^k),\, g_{A(\bar{x})}(x^k + p^k) + w_k e). \tag{4.185}$$

Moreover, by Lemma A.6, the matrices $(J_k J_k^{\mathrm{T}})^{-1}$ are uniformly bounded provided x^k is close enough to $\bar{x}$. Hence, (4.185) implies the estimate (4.181).

The last assertion follows from (4.181) and Lemma A.11. $\square$

The origins of the second-order corrections technique become more transparent if we consider the following interpretation. Most of the Newton-type methods for problem (4.73) (like SQP and LCL) employ some linearization-based approximation of the constraints. Sometimes, this approximation appears not good enough, and it would be desirable to replace it by the more accurate quadratic approximation (like in SQCQP):

$$\begin{aligned} &h(x^k) + h'(x^k)(x - x^k) + \frac{1}{2}h''(x^k)[x - x^k,\, x - x^k] = 0, \\ &g(x^k) + g'(x^k)(x - x^k) + \frac{1}{2}g''(x^k)[x - x^k,\, x - x^k] \leq 0. \end{aligned} \tag{4.186}$$

However, as also discussed in Sect. 4.3.5, quadratic constraints certainly yield more difficult subproblems. The idea is then to replace the quadratic terms by something more tractable, employing the Newtonian direction p^k which has been already computed.

Specifically, with the expectation that the hypothetical new iterate defined employing the constraints in (4.186) would not be too far from $x^k + p^k$, we replace the quadratic terms by their values at $x = x^k + p^k$. Furthermore, since locally p^k is expected to be small, the approximations

$$h(x^k + p^k) \approx h(x^k) + h'(x^k)p^k + \frac{1}{2}h''(x^k)[p^k,\, p^k]$$

and

$$g(x^k + p^k) \approx g(x^k) + g'(x^k)p^k + \frac{1}{2}g''(x^k)[p^k,\, p^k]$$

should be quite accurate. Hence, we can completely avoid the need to compute the quadratic terms if we substitute them in the two relations of (4.186) by $h(x^k+p^k) - h(x^k) - h'(x^k)p^k$ and $g(x^k+p^k) - g(x^k) - g'(x^k)p^k$, respectively. By doing so, we obtain the linear constraints

$$\begin{aligned} h(x^k + p^k) + h'(x^k)(x - x^k - p^k) &= 0, \\ g(x^k + p^k) + g'(x^k)(x - x^k - p^k) &\le 0. \end{aligned} \tag{4.187}$$

If for those indices which are expected to belong to the active set $A(\bar{x})$ we replace the corresponding inequality constraints in (4.187) by equalities and drop the other inequality constraints in (4.187), we obtain precisely the linear system (4.179) with $w_k = 0$ with respect to $p = x - x^k - p^k$. The subproblem (4.180) means taking the closest to zero solution of this system, which is only natural, since we want to keep the second-order correction term as small as possible.

Second-order corrections with subproblems of the form (4.180) most naturally combine with Newton-type methods whose iteration subproblems are linear systems of equations, rather than quadratic programming problems. One immediate example is given by the Newton-Lagrange method for equality-constrained problems, considered in Sect. 4.1. In Sect. 6.2.2, second-order corrections will be used to speed up the globalized Newton-Lagrange method. Local analysis is based on the estimate (4.183) from Lemma 4.38, which holds when p^k is computed in some specific way such that (4.182) holds. Then the results on local convergence and superlinear rate of convergence can be readily derived by applying Propositions 2.4, 2.6, 4.18 and 4.19. Another example of this use of second-order corrections will be provided in Sect. 4.3.7.

Note that the constraints in (4.187) can also be used directly, substituting the linearized constraints of the Newtonian subproblem for computing the next iterate x^{k+1}, without any attempts to identify the active constraints of the original problem. Our goal in the rest of this section is to show that in conjunction with the basic SQP method, second-order corrections of this kind fit the perturbed SQP framework. Consequently, they do not destroy the local convergence and rate of convergence properties of the SQP method, discussed in Sect. 4.2.

We shall consider the following algorithm.

Algorithm 4.39 Choose $(x^0,\, \lambda^0,\, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and set $k = 0$.

1. If $(x^k,\, \lambda^k,\, \mu^k)$ satisfies the KKT system (4.59), stop.
2. Choose a symmetric matrix $H_k \in \mathbf{R}^{n\times n}$, compute $\tilde{x}^{k+1} \in \mathbf{R}^n$ as a stationary point of problem (4.75), and set $p^k = \tilde{x}^{k+1} - x^k$.
3. Compute $x^{k+1} \in \mathbf{R}^n$ as a stationary point of the problem

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, x - x^k\rangle + \dfrac{1}{2}\langle H_k(x - x^k),\, x - x^k\rangle \\ \text{subject to} & h(x^k + p^k) + h'(x^k)(x - x^k - p^k) = 0, \\ & g(x^k + p^k) + g'(x^k)(x - x^k - p^k) \leq 0, \end{array} \tag{4.188}$$

and $(\lambda^{k+1},\, \mu^{k+1}) \in \mathbf{R}^l \times \mathbf{R}^m$ as an associated Lagrange multiplier.

4. Increase k by 1 and go to step 1.

The local superlinear convergence of Algorithm 4.39 is established in the next theorem.

Theorem 4.40. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a local solution of problem (4.73), satisfying the SMFCQ and the SOSC (4.100) for the associated Lagrange multiplier $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$.*

Then there exists a constant $\delta > 0$ such that for any starting point $(x^0,\, \lambda^0,\, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, there exist sequences $\{(\tilde{x}^k,\, \tilde{\lambda}^k,\, \tilde{\mu}^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ such that for each $k = 0,\, 1,\, \ldots,$ the point $\tilde{x}^{k+1}$ is a stationary point of problem (4.75) with H_k chosen according to (4.60), $(\tilde{\lambda}^k,\, \tilde{\mu}^k)$ is an associated with $\tilde{x}^{k+1}$ Lagrange multiplier, x^{k+1} is a stationary point of problem (4.188) with the same H_k and with $p^k = \tilde{x}^{k+1} - x^k$, and $(\lambda^{k+1},\, \mu^{k+1})$ is an associated with x^{k+1} Lagrange multiplier satisfying

$$\|(\tilde{x}^{k+1} - x^k,\, \tilde{\lambda}^{k+1} - \lambda^k,\, \tilde{\mu}^{k+1} - \mu^k)\| \leq \delta \tag{4.189}$$

and (4.67); any such sequence $\{(x^k,\, \lambda^k,\, \mu^k)\}$ converges to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the second derivatives of f, h, and g are locally Lipschitz-continuous with respect to $\bar{x}$.

Proof. As the algorithm in question has a two-stage iteration, the proof also goes in two steps.

The first step consists of applying Theorem 4.14, which gives the existence of $\delta > 0$ such that for any $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, there exists a stationary point $p \in \mathbf{R}^n$ of the problem

$$\begin{array}{ll} \text{minimize} & \langle f'(x),\, \xi\rangle + \dfrac{1}{2}\left\langle \dfrac{\partial^2 L}{\partial x^2}(x,\, \lambda,\, \mu)\xi,\, \xi\right\rangle \\ \text{subject to} & h(x) + h'(x)\xi = 0,\; g(x) + g'(x)\xi \leq 0, \end{array}$$

and an associated Lagrange multiplier $(\tilde{\lambda},\, \tilde{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ such that

$$\|(p,\, \tilde{\lambda} - \lambda,\, \tilde{\mu} - \mu)\| \leq \delta,$$

and any $(\tilde{x},\, \tilde{\lambda},\, \tilde{\mu})$ with these properties satisfies

$$\|(x + p - \bar{x},\, \tilde{\lambda} - \bar{\lambda},\, \tilde{\mu} - \bar{\mu})\| = o(\|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu})\|) \tag{4.190}$$

as $(x,\, \lambda,\, \mu) \to (\bar{x},\, \bar{\lambda},\, \bar{\mu})$. For any fixed choice of such $p = p(x,\, \lambda,\, \mu)$, the relation (4.190) implies that

$$\begin{aligned}\|p(x,\, \lambda,\, \mu)\| &\leq \|x - \bar{x}\| + \|x + p(x,\, \lambda,\, \mu) - \bar{x}\| \\ &= O(\|x - \bar{x}\|) + o(\|(\lambda - \bar{\lambda},\, \mu - \bar{\mu})\|)\end{aligned} \tag{4.191}$$

as $(x,\, \lambda,\, \mu) \to (\bar{x},\, \bar{\lambda},\, \bar{\mu})$.

The second stage consists of applying Theorem 4.24 with

$$\psi((x,\, \lambda,\, \mu),\, \xi) = 0,$$

$$\omega_2((x,\, \lambda,\, \mu),\, \xi) = h(x + p(x,\, \lambda,\, \mu)) - h(x) - h'(x)p(x,\, \lambda,\, \mu),$$

$$\omega_3((x,\, \lambda,\, \mu),\, \xi) = g(x + p(x,\, \lambda,\, \mu)) - g(x) - g'(x)p(x,\, \lambda,\, \mu)$$

for $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and $\xi \in \mathbf{R}^n$. In particular, by (4.191),

$$\omega_2((x,\, \lambda,\, \mu),\, \xi) = O(\|p(x,\, \lambda,\, \mu)\|^2) = O(\|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu})\|^2),$$

and similarly,

$$\omega_3((x,\, \lambda,\, \mu),\, \xi) = O(\|(x - \bar{x},\, \lambda - \bar{\lambda},\, \mu - \bar{\mu})\|^2)$$

as $(x,\, \lambda,\, \mu) \to (\bar{x},\, \bar{\lambda},\, \bar{\mu})$ (observe that in fact, ω_2 and ω_3 do not depend on ξ), which gives (4.123). Thus, Theorem 4.24 is indeed applicable and gives the stated assertion, perhaps with a smaller $\delta > 0$. $\square$

We complete this section with an a posteriori result on the superlinear primal convergence of Algorithm 4.39.

Theorem 4.41. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (4.73), and let $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be an associated Lagrange multiplier. Let iterative sequences $\{\tilde{x}^k\} \subset \mathbf{R}^n$ and $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ generated by Algorithm 4.39 be convergent to $\bar{x}$ and $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$.*

If $(\bar{\lambda},\, \bar{\mu})$ satisfies the SOSC (4.100) and the Dennis–Moré condition (4.102) holds, then the rate of convergence of $\{x^k\}$ is superlinear.

Alternatively, under the weaker assumption that $(\bar{\lambda},\, \bar{\mu})$ is a noncritical multiplier, if (4.103) holds, then the rate of convergence of $\{x^k\}$ is also superlinear.

Proof. For each k, set

$$\omega_1^k = H_k - \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k), \tag{4.192}$$

and follow the proof of Proposition 4.17 with the iterate x^{k+1} substituted by $\tilde{x}^{k+1}$, and with $(\lambda^{k+1}, \mu^{k+1})$ substituted by a Lagrange multiplier $(\lambda^{k+1}, \mu^{k+1}) \in \mathbf{R}^l \times \mathbf{R}^m$ of problem (4.75) associated with $\tilde{x}^{k+1}$, and with any Lipschitz-continuous mapping $\pi : \mathbf{R}^n \to \mathbf{R}^n$ satisfying (4.90) for $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close to $(\bar{x}, \bar{\lambda}, \bar{\mu})$. This gives the estimate

$$\tilde{x}^{k+1} - \bar{x} = O(\|\pi(-\omega_1^k)\|) + o(\|x^k - \bar{x}\|)$$

as $k \to \infty$. Assuming the SOSC (4.100), taking $\pi = \pi_{C(\bar{x})}$, and combining the fact above with Proposition 1.46 and condition (4.102), we derive the estimate

$$\tilde{x}^{k+1} - \bar{x} = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \tag{4.193}$$

as $k \to \infty$. Alternatively, under the assumption that $(\bar{\lambda}, \bar{\mu})$ is a noncritical multiplier, taking $\pi = \pi_{C_+(\bar{x},\, \bar{\mu})}$, estimate (4.193) follows by Proposition 1.44 if we replace (4.102) by the stronger condition (4.103).

For $p^k = \tilde{x}^{k+1} - x^k$, we obtain from (4.193) that

$$p^k = O(\|x^k - \bar{x}\|) + o(\|x^{k+1} - x^k\|)$$

as $k \to \infty$. Taking this into account, the KKT conditions of the subproblem (4.188) correspond to the perturbed SQP framework (4.81) with ω_1^k defined in (4.192), and with

$$\begin{aligned}\omega_2^k = h(x^k + p^k) - h(x^k) - h'(x^k)p^k &= O(\|p^k\|^2)\\ &= o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|),\end{aligned}$$

$$\begin{aligned}\omega_3^k = g(x^k + p^k) - g(x^k) - g'(x^k)p^k &= O(\|p^k\|^2)\\ &= o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)\end{aligned}$$

as $k \to \infty$. It remains to apply Propositions 4.18 (assuming (4.103)) and 4.19 (assuming (4.102)). $\square$

4.3.7 Interior Feasible Directions Methods

A family of Newton-type methods for problem (4.73) relies on the following idea [116–118, 214, 260]. We apply the Newton method to the equalities in the KKT system (4.74), as if forgetting about the inequalities. Of course, this approach must be applied with some modifications, so that the inequalities are also enforced, at least in the limit. The modification consists in introducing a special perturbation into the Newton iteration for the equalities, which ensures that the inequalities actually hold strictly along the iterations (under certain assumptions, of course).

We shall interpret this approach in the spirit of our perturbed Newtonian frameworks, where the basic (unperturbed) scheme is given by the usual Newton method applied to the equations in the KKT system (4.74), writing the complementarity condition componentwise:

$$\frac{\partial L}{\partial x}(x,\,\lambda,\,\mu)=0,\quad h(x)=0,\quad \mu_i g_i(x)=0,\ i=1,\,\ldots,\,m. \tag{4.194}$$

This is a system of $n+l+m$ equations in the variables $(x,\,\lambda,\,\mu)\in\mathbf{R}^n\times\mathbf{R}^l\times\mathbf{R}^m$, and thus the Newton method for systems of equations is applicable (see Sect. 2.1.1). We shall start precisely with the latter, introducing perturbations that account for inequalities in the KKT system later on.

The basic Newton method for the system (4.194) takes the following form. Let $(x^k,\,\lambda^k,\,\mu^k)\in\mathbf{R}^n\times\mathbf{R}^l\times\mathbf{R}^m$ be the current primal-dual iterate. Then the next iterate $(x^{k+1},\,\lambda^{k+1},\,\mu^{k+1})$ is computed as a solution of the linear system

$$\begin{gathered}\frac{\partial^2 L}{\partial x^2}(x^k,\,\lambda^k,\,\mu^k)(x-x^k)\\ +(h'(x^k))^{\mathrm{T}}(\lambda-\lambda^k)+(g'(x^k))^{\mathrm{T}}(\mu-\mu^k)=-\frac{\partial L}{\partial x}(x^k,\,\lambda^k,\,\mu^k),\\ h'(x^k)(x-x^k)=-h(x^k),\\ \mu_i^k\langle g_i'(x^k),\,x-x^k\rangle+(\mu_i-\mu_i^k)g_i(x^k)=-\mu_i^k g_i(x^k),\quad i=1,\,\ldots,\,m,\end{gathered} \tag{4.195}$$

with respect to $(x,\,\lambda,\,\mu)\in\mathbf{R}^n\times\mathbf{R}^l\times\mathbf{R}^m$.

Algorithm 4.42 Choose $(x^0,\,\lambda^0,\,\mu^0)\in\mathbf{R}^n\times\mathbf{R}^l\times\mathbf{R}^m$ and set $k=0$.

1. If $(x^k,\,\lambda^k,\,\mu^k)$ satisfies (4.194), stop.
2. Compute $(x^{k+1},\,\lambda^{k+1},\,\mu^{k+1})\in\mathbf{R}^n\times\mathbf{R}^l\times\mathbf{R}^m$ as a solution of the linear system (4.195).
3. Increase k by 1 and go to step 1.

The following lemma concerning the nonsingularity of the Jacobian of the system (4.194) at its solution, which in turn makes the Newton method applicable, is an easy generalization of Lemma 4.2.

Lemma 4.43. *Let $f:\mathbf{R}^n\to\mathbf{R}$, $h:\mathbf{R}^n\to\mathbf{R}$ and $g:\mathbf{R}^n\to\mathbf{R}^m$ be twice differentiable at $\bar{x}\in\mathbf{R}^n$. Let $\bar{x}$ be a stationary point of problem (4.73), satisfying the LICQ. Let $(\bar{\lambda},\,\bar{\mu})\in\mathbf{R}^l\times\mathbf{R}^m$ be the (unique) Lagrange multiplier associated with $\bar{x}$, and assume that the strict complementarity condition $\bar{\mu}_{A(\bar{x})}>0$ and the SOSC (4.100) hold.*

Then the Jacobian

$$J=\begin{pmatrix}\dfrac{\partial^2 L}{\partial x^2}(\bar{x},\,\bar{\lambda},\,\bar{\mu}) & (h'(\bar{x}))^{\mathrm{T}} & (g'(\bar{x}))^{\mathrm{T}}\\ h'(\bar{x}) & 0 & 0\\ \mathrm{diag}(\bar{\mu})g'(\bar{x}) & 0 & \mathrm{diag}(g(\bar{x}))\end{pmatrix} \tag{4.196}$$

of the system (4.194) at $(\bar{x},\,\bar{\lambda},\,\bar{\mu})$ is a nonsingular matrix.

Proof. Take an arbitrary $(\xi,\,\eta,\,\zeta) \in \ker J$. Then, taking into account that $\bar{\mu}_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})} = 0$, and using the strict complementarity condition, we obtain from (4.196) that

$$\frac{\partial^2 L}{\partial x^2}(\bar{x},\,\bar{\lambda},\,\bar{\mu})\xi + (h'(\bar{x}))^{\mathrm{T}}\eta + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0, \tag{4.197}$$

$$h'(\bar{x})\xi = 0, \tag{4.198}$$

$$g'_{A(\bar{x})}(\bar{x})\xi = 0, \quad \zeta_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})} = 0. \tag{4.199}$$

Relation (4.198), the first equality in (4.199), and strict complementarity imply that $\xi \in C(\bar{x})$ (see Lemma 1.17). Multiplying the left- and right-hand sides of (4.197) by ξ and employing (4.198) and (4.199), we derive that

$$\begin{aligned} 0 &= \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\,\bar{\lambda},\,\bar{\mu})\xi,\,\xi \right\rangle + \langle \eta,\, h'(\bar{x})\xi\rangle + \langle \zeta,\, g'(\bar{x})\xi\rangle \\ &= \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\,\bar{\lambda},\,\bar{\mu})\xi,\,\xi \right\rangle. \end{aligned}$$

This implies that $\xi = 0$, since otherwise a contradiction with the SOSC (4.100) is obtained. But then (4.197) takes the form

$$(h'(\bar{x}))^{\mathrm{T}}\eta + (g'_{A(\bar{x})}(\bar{x}))^{\mathrm{T}}\zeta_{A(\bar{x})} = 0,$$

where the second equality in (4.199) was taken into account. According to the LICQ, the latter may hold only when $(\eta,\,\zeta_{A(\bar{x})}) = 0$. Combining this with the second equality in (4.199), we conclude that $(\eta,\,\zeta) = 0$.

It is thus shown that $(\xi,\,\eta,\,\zeta) = 0$, i.e., $\ker J = \{0\}$, which means that J is nonsingular. □

It can be easily seen that the LICQ and the strict complementarity condition are also necessary for the nonsingularity of the Jacobian of system (4.194) at $(\bar{x},\,\bar{\lambda},\,\bar{\mu})$, while the SOSC (4.100) in Lemma 4.43 can be replaced by the weaker assumption that the multiplier $(\bar{\lambda},\,\bar{\mu})$ is noncritical. Moreover, the latter is also necessary for the nonsingularity of the Jacobian in question.

Theorem 2.2 and Lemma 4.2 immediately imply the following result on the local superlinear convergence of Algorithm 4.42.

Theorem 4.44. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a local solution of problem (4.57), satisfying the LICQ. Let $(\bar{\lambda},\,\bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be the (unique) Lagrange multiplier associated with $\bar{x}$, and assume that the strict complementarity condition $\bar{\mu}_{A(\bar{x})} > 0$ and the SOSC (4.100) hold.*

Then any starting point $(x^0,\,\lambda^0,\,\mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x},\,\bar{\lambda},\,\bar{\mu})$ uniquely defines the iterative sequence of Algorithm 4.42, this sequence converges to $(\bar{x},\,\bar{\lambda},\,\bar{\mu})$, and the rate of convergence is superlinear.

Moreover, the rate of convergence is quadratic provided the second derivatives of f, h, and g are locally Lipschitz-continuous with respect to $\bar{x}$.

It is clear from the analysis and the discussion above that the LICQ and the strict complementarity are inevitable ingredients of local convergence analysis of Algorithm 4.42; they cannot be dropped or relaxed. Comparing this with the local convergence results for various methods presented in the previous sections, we emphasize that the assumptions required by the method in (4.195) are rather strong. On the other hand, an iteration of this method consists of solving just one system of linear equations (4.194) rather than a quadratic programming subproblem (like in SQP) or even a more difficult subproblem (like in LCL or SQCQP). This gives significant savings in the computational cost of each iteration. Moreover, this simple and seemingly naive approach is a basis for a class of quite successful feasible descent algorithms originating from [116, 117], and further developed in [118, 214, 260]. Of course, another key issue is globalization of convergence of such algorithms. Any detailed discussion of this is beyond the scope of this book, as these algorithms are more in the spirit of interior-point methods: primal iterates remain strictly feasible with respect to inequality constraints, and dual iterates corresponding to these constraints remain strictly positive. Consequently, these methods certainly do not directly fit the perturbed SQP framework (4.81), where the multiplier estimates corresponding to inactive linearized inequality constraints must be equal to zero. However, later on in this section, we discuss some modifications of Algorithm 4.42 that bring it closer to the practical algorithms mentioned above.

Before proceeding to those modifications, note again that Algorithm 4.42 does not fit the perturbed SQP framework (4.81), because the components of μ^{k+1} corresponding to inequality constraints inactive at the solution can be negative along iterations (and approach the eventual zero limit "from the left"). Moreover, the associated complementarity conditions in (4.81) may not hold with any tractable perturbation term ω_3^k, even locally. For this reason, a priori convergence results for Algorithm 4.42 cannot be derived from the perturbed SQP framework. However, a posteriori analysis is still possible.

Theorem 4.45. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of a point $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (4.73), and let $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be an associated noncritical Lagrange multiplier. Let an iterative sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ generated by Algorithm 4.42 be convergent to $(\bar{x}, \bar{\lambda}, \bar{\mu})$.*

Then the rate of convergence of $\{x^k\}$ is superlinear.

Proof. Define the auxiliary sequence $\{\breve{\mu}^k\} \subset \mathbf{R}^m$ by setting

$$\breve{\mu}^k_{A(\bar{x})} = \mu^k_{A(\bar{x})}, \quad \breve{\mu}^k_{\{1,\,\dots,\,m\}\setminus A(\bar{x})} = 0 \tag{4.200}$$

for each k. Furthermore, for each k define $\omega_1^k \in \mathbf{R}^n$ and $\omega_3^k \in \mathbf{R}^m$ by setting

$$\omega_1^k = (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \breve{\mu}^{k+1}) + (g''(x^k)[x^{k+1} - x^k])^{\mathrm{T}}(\mu^k - \breve{\mu}^k), \quad (4.201)$$

$$(\omega_3^k)_i = \left(\frac{\mu_i^{k+1}}{\mu_i^k} - 1\right) g_i(x^k), \ i \in A(\bar{x}), \quad (\omega_3^k)_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})} = 0. \quad (4.202)$$

Then the first equality in (4.195) and (4.201) imply that

$$\begin{aligned} &\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \breve{\mu}^k) + \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \breve{\mu}^k)(x^{k+1} - x^k) \\ &\quad + (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\breve{\mu}^{k+1} - \breve{\mu}^k) + \omega_1^k = 0. \end{aligned} \quad (4.203)$$

Furthermore, combining the last block of equalities in (4.195) with (4.200) and (4.202), we conclude that for all k large enough

$$\breve{\mu}^k_{A(\bar{x})} > 0, \quad g_{A(\bar{x})}(x^k) + g'_{A(\bar{x})}(x^k)(x^{k+1} - x^k) + (\omega_3^k)_{A(\bar{x})} = 0. \quad (4.204)$$

Finally, according to (4.200), for all k large enough it holds that

$$\begin{gathered} \breve{\mu}^{k+1}_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})} = 0, \\ g_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})}(x^k) + g'_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})}(x^k)(x^{k+1} - x^k) < 0. \end{gathered} \quad (4.205)$$

Combining (4.203)–(4.205) with the second equality in (4.195), we obtain that $(x^{k+1},\, \lambda^{k+1},\, \breve{\mu}^{k+1})$ satisfies (4.81) with μ^k replaced by $\breve{\mu}^k$, and with ω_1^k defined in (4.201), $\omega_2^k = 0$, and ω_3^k defined in (4.202).

We proceed with estimating the perturbation terms. According to (4.200), for all k it holds that

$$\mu^k_{A(\bar{x})} - \breve{\mu}^k_{A(\bar{x})} = 0, \quad \mu^k_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})} - \breve{\mu}^k_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})} = \mu^k_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})},$$

and hence, by (4.201),

$$\begin{aligned} \omega_1^k &= O(\|\mu^{k+1}_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})}\| + \|\mu^k_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})}\|\|x^{k+1} - x^k\|) \\ &= O(\|\mu^{k+1}_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})}\|) + o(\|x^{k+1} - x^k\|) \end{aligned} \quad (4.206)$$

as $k \to \infty$, where the last estimate is by the convergence of $\{\mu^k_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})}\}$ to $\bar{\mu}_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})} = 0$. For all k large enough it holds that $g_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})}(x^k) < 0$, and the last block of relations in (4.195) implies that

$$\begin{aligned} \mu_i^{k+1} &= -\frac{\mu_i^k}{g_i(x^k)}\langle g_i'(x^k),\, x^{k+1} - x^k\rangle \\ &= o(\|x^{k+1} - x^k\|) \quad \forall\, i \in \{1,\, \ldots,\, m\} \setminus A(\bar{x}) \end{aligned} \quad (4.207)$$

as $k \to \infty$, where the last estimate is again by the convergence of the sequence $\{\mu^k_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})}\}$ to 0, and by the convergence of $\{g_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})}(x^k)\}$ to $g_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})}(\bar{x}) < 0$. Putting together (4.206) and (4.207), we finally get the estimate

$$\omega_1^k = o(\|x^{k+1} - x^k\|) \quad (4.208)$$

as $k \to \infty$.

Furthermore, according to (4.202),

$$(\omega_3^k)_i = \left(\frac{\mu_i^{k+1} - \mu_i^k}{\mu_i^k}\right) \langle g_i'(\bar{x}),\, x^k - \bar{x}\rangle + o(\|x^k - \bar{x}\|)$$
$$= o(\|x^k - \bar{x}\|) \quad \forall\, i \in A(\bar{x}) \tag{4.209}$$

as $k \to \infty$, where the last estimate is by the convergence of $\{\mu^k_{A(\bar{x})}\}$ to $\bar{\mu}_{A(\bar{x})} > 0$.

From (4.200) and from the equality $\bar{\mu}_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})} = 0$, it evidently follows that $\{\breve{\mu}^k\}$ converges to the same limit $\bar{\mu}$ of $\{\mu^k\}$. The stated assertion follows now from the estimates (4.208) and (4.209), and from Proposition 4.18. □

We next discuss a local version of the algorithms developed in [118, 214, 260]. Following the pattern of these works, we first consider the problem with inequality constraints only:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \le 0. \end{array} \tag{4.210}$$

Later on, we get back to the general problem (4.73) and demonstrate how equality constraints can be incorporated.

The system collecting equalities in the KKT conditions for problem (4.210) (recall also (4.194)) is given by

$$\frac{\partial L}{\partial x}(x,\, \mu) = 0, \quad \mu_i g_i(x) = 0,\ i = 1,\, \ldots,\, m. \tag{4.211}$$

For the current iterate $(x^k,\, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^m$ such that x^k is strictly feasible for problem (4.210) (i.e., $g(x^k) < 0$) and $\mu^k > 0$, consider the iteration system

$$\begin{gathered} H_k(x - x^k) + (g'(x^k))^{\mathrm{T}}(\mu - \mu^k) = -\frac{\partial L}{\partial x}(x^k,\, \mu^k), \\ \mu_i^k \langle g_i'(x^k),\, x - x^k\rangle + (\mu_i - \mu_i^k) g_i(x^k) = -\mu_i^k(g_i(x^k) + v_i^k), \quad i = 1,\, \ldots,\, m, \end{gathered} \tag{4.212}$$

with some symmetric matrix $H_k \in \mathbf{R}^{n\times n}$, and some element $v^k \in \mathbf{R}^m$. If $H_k = \frac{\partial^2 L}{\partial x^2}(x^k,\, \mu^k)$ and $v^k = 0$, this is the counterpart of the system (4.195) for problem (4.210). If $v^k \neq 0$, it is a linear system with the same matrix but with a perturbed right-hand side.

Let $(\tilde{x}^{k+1},\, \tilde{\mu}^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^m$ be a solution of the system (4.212) with $v^k = 0$, and set $\tilde{p}^k = \tilde{x}^{k+1} - x^k$. If $\tilde{p}^k = 0$, then necessarily $\tilde{\mu}^{k+1} = 0$, x^k is a stationary point of problem (4.210) with this multiplier, and no further progress can be expected from the method. Otherwise,

$$f'(x^k) = -H_k \tilde{p}^k - (g'(x^k))^{\mathrm{T}} \tilde{\mu}^{k+1},$$

$$\tilde{\mu}_i^{k+1} = -\frac{\mu_i^k}{g_i(x^k)} \langle g_i'(x^k),\, \tilde{p}^k \rangle, \quad i = 1,\, \ldots,\, m,$$

and combining these two relations, we derive that

$$\langle f'(x^k),\, \tilde{p}^k \rangle = -\langle H_k \tilde{p}^k,\, \tilde{p}^k \rangle + \sum_{i=1}^{m} \frac{\mu_i^k}{g_i(x^k)} \langle g_i'(x^k),\, \tilde{p}^k \rangle^2 < 0$$

if, e.g., H_k is positive definite. Hence, by Lemma 2.18, in this case $\tilde{p}^k$ is a descent direction for f at x^k. Moreover, since $g(x^k) < 0$, this descent direction is also feasible in a sense that any sufficiently short step from x^k in this direction would preserve feasibility. However, this step can be too short when x^k is close to the boundary of the feasible set, i.e., some constraints are almost active at x^k. Indeed,

$$\langle g_i'(x^k),\, \tilde{p}^k \rangle = -\frac{\tilde{\mu}_i^{k+1}}{\mu_i^k} g_i(x^k), \quad i = 1,\, \ldots,\, m,$$

and if for some $i = 1,\, \ldots,\, m$, it appears that $g_i(x^k)$ is close to zero, then $\tilde{p}^k$ can be almost orthogonal to $g_i'(x^k)$.

A remedy is to use the perturbation term v^k with positive components. Let $(x^{k+1},\, \mu^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^m$ be a solution of the system (4.212) with $v^k > 0$, and set $p^k = x^{k+1} - x^k$. Then

$$\langle g_i'(x^k),\, p^k \rangle = -v^k - \frac{\mu_i^{k+1}}{\mu_i^k} g_i(x^k), \quad i = 1,\, \ldots,\, m,$$

and for any $i = 1,\, \ldots,\, m$ such that $g_i(x^k)$ is close to zero, the larger are the components of v^k the better are the descent properties of the direction p^k for g_i at x^k.

However, the components of v^k should not be taken too large, as this can destroy the descent properties of p^k for f. In addition, this perturbation can interfere with the high convergence rate of the original Algorithm 4.42. For a compromise, fix some $\sigma \in (0,\, 1)$ and $\nu \geq 2$ (in [214, 260] it is suggested to take $\nu > 2$, whereas [118] suggests $\nu = 2$), and for each k, select the biggest $t_k \in (0,\, 1]$ such that for p^k computed with $v_i^k = t_k \|\tilde{p}^k\|^\nu$, $i = 1,\, \ldots,\, m$, it holds that

$$\langle f'(x^k),\, p^k \rangle \leq \sigma \langle f'(x^k),\, \tilde{p}^k \rangle.$$

If $\sigma \langle f'(x^k),\, \tilde{p}^k \rangle < 0$ (which is guaranteed in the case of a positive-definite H_k), the needed t_k always exists. This is because $p^k = \tilde{p}^k$ when $t_k = 0$, and p^k depends continuously on t_k, provided the matrix of the linear system (4.212) is nonsingular. Moreover, this dependence is actually affine, and thus, the needed t_k, p^k and μ^{k+1} can be easily computed as follows. Solve (4.212) with, say, $v_i^k = \|\tilde{p}^k\|^\nu$, $i = 1,\, \ldots,\, m$ (corresponding to $t_k = 1$) to obtain $\hat{p}^k$ and $\hat{\mu}^{k+1}$; compute

$$t_k = \begin{cases} 1 & \text{if } \langle f'(x^k),\, \hat{p}^k\rangle \leq \sigma\langle f'(x^k),\, \tilde{p}^k\rangle, \\ (1-\sigma)\dfrac{\langle f'(x^k),\, \tilde{p}^k\rangle}{\langle f'(x^k),\, \tilde{p}^k - \hat{p}^k\rangle} & \text{otherwise;} \end{cases} \tag{4.213}$$

and set

$$p^k = (1-t_k)\tilde{p}^k + t_k\hat{p}^k, \quad \mu^{k+1} = (1-t_k)\tilde{\mu}^{k+1} + t_k\hat{\mu}^{k+1}.$$

Recall that only the right-hand side of the system (4.212) is perturbed, and so no extra factorization of the matrix is needed if direct methods are applied.

In Algorithm 4.46 below, the formula for the next dual iterate will be slightly modified, reflecting the intention to keep the dual iterates positive (but not "too positive," for inactive constraints).

The algorithms developed in [214, 260] are also equipped with second-order corrections of some kind. Moreover, in this context, second-order corrections serve not only the goal of achieving superlinear convergence of the globalized algorithms (which is their general purpose; see Sects. 4.3.6 and 6.2.2), but also ensure (strict) feasibility of the new iterate x^{k+1}. In this sense, second-order corrections are here an intrinsic ingredient of the local algorithm as well (unlike in the case of SQP).

Define the index set

$$A_k = \{i = 1,\, \ldots,\, m \mid \mu_i^{k+1} \geq -g_i(x^k)\}, \tag{4.214}$$

estimating active constraints $A(\bar{x})$, and compute another direction $\bar{p}^k \in \mathbf{R}^n$ as a solution of the subproblem

$$\begin{array}{ll} \text{minimize} & \|p\|^2 \\ \text{subject to} & g_{A_k}(x^k + p^k) + g'_{A_k}(x^k)p = -w_k e. \end{array} \tag{4.215}$$

Here the scalar $w_k > 0$ is chosen so that it places $x^{k+1} = x^k + p^k + \bar{p}^k$ into the interior of the feasible set. At the same time, the value of w_k must be coordinated with $\|p^k\|$, because the correction $\bar{p}^k$ must remain relatively small with respect to p^k. Note that the norm in (4.215) can be different from the Euclidean. Also, [214, 260] suggest to compute $\bar{p}^k$ as a solution of a different subproblem, namely

$$\begin{array}{ll} \text{minimize} & \langle H_k p,\, p\rangle \\ \text{subject to} & g_{A_k}(x^k + p^k) + g'_{A_k}(x^k)p = -w_k e. \end{array}$$

However, for the local algorithm presented below, the simplest subproblem (4.215) with the Euclidean norm does the job. We therefore omit details for the other options.

Regarding the specific choice of w_k, in the algorithm below we use the rule (4.216) borrowed from [214, 260].

Algorithm 4.46 Choose the parameters $\nu \geq 2$, $\sigma \in (0, 1)$, $\tau_1 \in [2, 3)$, $\tau_2 \in [0, 1)$, $\tau > 0$. Choose $(x^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^m$ and set $k = 0$.

1. If (x^k, μ^k) satisfies (4.211), stop.
2. Choose symmetric matrix $H_k \in \mathbf{R}^{n\times n}$; compute $(\tilde{x}^{k+1}, \tilde{\mu}^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^m$ as a solution of the linear system (4.212) with $v^k = 0$. Set $\tilde{p}^k = \tilde{x}^{k+1} - x^k$.
3. Compute $(\hat{x}^{k+1}, \hat{\mu}^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^m$ as a solution of the system (4.212) with $v_i^k = \|\tilde{p}^k\|^\nu$, $i = 1, \ldots, m$, and set $\hat{p}^k = \hat{x}^{k+1} - x^k$.
4. Compute t_k according to (4.213) and set
$$p^k = (1 - t_k)\tilde{p}^k + t_k\hat{p}^k, \quad \bar{\mu}^{k+1} = (1 - t_k)\tilde{\mu}^{k+1} + t_k\hat{\mu}^{k+1}.$$
5. Define the index set A_k according to (4.214). Set
$$w_k = \max\left\{\|p^k\|^{\tau_1}, \max_{i\in A_k}\left|\frac{\mu_i^k}{\bar{\mu}_i^{k+1}} - 1\right|^{\tau_2}\|p^k\|^2\right\}, \tag{4.216}$$
and compute $\bar{p}^k \in \mathbf{R}^n$ as a solution of problem (4.215).
6. Set
$$x^{k+1} = x^k + p^k + \bar{p}^k,$$
$$\mu_i^{k+1} = \max\{\bar{\mu}_i^{k+1}, \|p^k\|^\tau\}, \quad i = 1, \ldots, m. \tag{4.217}$$
7. Increase k by 1 and go to step 1.

Regarding the choice of parameter τ, a priori local superlinear primal-dual convergence result (Theorem 4.49 below) requires $\tau > 1$, while a posteriori superlinear primal convergence result (Theorem 4.51 below) allows any $\tau > 0$.

Lemma 4.47. *Let $g : \mathbf{R}^n \to \mathbf{R}^m$ be continuous at $\bar{x} \in \mathbf{R}^n$. Assume that $g(\bar{x}) \leq 0$, and let $\bar{\mu} \in \mathbf{R}^m$ be such that $\bar{\mu}_{A(\bar{x})} > 0$ and $\bar{\mu}_{\{1,\ldots,m\}\setminus A(\bar{x})} = 0$.*

If $(x^k, \bar{\mu}^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^m$ is close enough to $(\bar{x}, \bar{\mu})$, then for the index set A_k defined according to (4.214) it holds that
$$A_k = A(\bar{x}), \tag{4.218}$$
and
$$\bar{\mu}^{k+1}_{A(\bar{x})} > 0. \tag{4.219}$$

Moreover, if in addition $f : \mathbf{R}^n \to \mathbf{R}$ is differentiable in a neighborhood of $\bar{x}$, and g is twice differentiable in a neighborhood of $\bar{x}$ with its second derivative being continuous at $\bar{x}$, if $\bar{x}$ satisfies the LICQ, and if $p^k \in \mathbf{R}^n$ is defined by Algorithm 4.46 for some $\mu^k \in \mathbf{R}^m$, then Algorithm 4.46 uniquely defines $\bar{p}^k \in \mathbf{R}^n$, and
$$\bar{p}^k = O\left(\|\tilde{p}^k\|^\nu + \|p^k\|^2 + \max_{i\in A(\bar{x})}\left|\frac{\mu_i^k}{\bar{\mu}_i^{k+1}} - 1\right|\|p^k\|\right) \tag{4.220}$$
as $(x^k, \mu^k) \to (\bar{x}, \bar{\mu})$, $p^k \to 0$, and $\bar{\mu}^{k+1} \to \bar{\mu}$, uniformly in $\tilde{p}^k \in \mathbf{R}^n$.

Proof. Relations (4.218) and (4.219) evidently follow from the strict complementarity condition. Therefore, problem (4.215) is equivalent to the quadratic programming problem

$$\begin{array}{ll} \text{minimize} & \|p\|^2 \\ \text{subject to} & g_{A(\bar{x})}(x^k + p^k) + g'_{A(\bar{x})}(x^k)p = -w_k e. \end{array}$$

According to Lemma 4.38, the LICQ implies that for all x^k close enough to $\bar{x}$, this problem has the unique solution $\bar{p}^k$, and

$$\bar{p}^k = O(\|g_{A(\bar{x})}(x^k + p^k)\| + w_k) \tag{4.221}$$

as $x^k \to \bar{x}$, uniformly in $p^k \in \mathbf{R}^n$ and $w_k \in \mathbf{R}$.

Furthermore, since $(x^k + p^k,\, \bar{\mu}^{k+1})$ satisfies (4.212) with $v_i^k = t_k\|\tilde{p}^k\|^\nu$, $i = 1, \ldots, m$, it holds that

$$\begin{aligned} g_i(x^k + p^k) &= g_i(x^k) + \langle g_i'(x^k),\, p^k\rangle + O(\|p^k\|^2) \\ &= -\left(\frac{\mu_i^k}{\bar{\mu}_i^{k+1}} - 1\right)\langle g_i'(x^k),\, p^k\rangle - \frac{\mu_i^k}{\bar{\mu}_i^{k+1}} t_k\|\tilde{p}^k\|^\nu + O(\|p^k\|^2) \\ &= O\left(\|\tilde{p}^k\|^\nu + \|p^k\|^2 + \left|\frac{\mu_i^k}{\bar{\mu}_i^{k+1}} - 1\right|\|p^k\|\right) \quad \forall\, i \in A(\bar{x}) \end{aligned}$$

as $(x^k,\, \mu^k) \to (\bar{x},\, \bar{\mu})$, $p^k \to 0$, and $\bar{\mu}^{k+1} \to \bar{\mu}$, uniformly in $\tilde{p}^k \in \mathbf{R}^n$.

Combining the latter with (4.221), and taking into account relations (4.216), (4.218) and the restrictions $\tau_1 \geq 2$ and $\tau_2 \geq 0$, we conclude that

$$\begin{aligned} \bar{p}^k &= O\Bigg(\|\tilde{p}^k\|^\nu + \|p^k\|^2 + \max_{i \in A(\bar{x})}\left|\frac{\mu_i^k}{\bar{\mu}_i^{k+1}} - 1\right|\|p^k\| \\ &\qquad + \|p^k\|^{\tau_1} + \max_{i \in A(\bar{x})}\left|\frac{\mu_i^k}{\bar{\mu}_i^{k+1}} - 1\right|^{\tau_2}\|p^k\|^2\Bigg) \\ &= O\left(\|\tilde{p}^k\|^\nu + \|p^k\|^2 + \max_{i \in A(\bar{x})}\left|\frac{\mu_i^k}{\bar{\mu}_i^{k+1}} - 1\right|\|p^k\|\right), \end{aligned}$$

which establishes (4.220). □

Before proceeding, recall that by Remark 1.18, if a stationary point $\bar{x}$ of problem (4.210) has an associated Lagrange multiplier $\bar{\mu}$ satisfying strict complementarity, then the critical cone of the problem is the subspace of the form $C(\bar{x}) = C_+(\bar{x},\, \bar{\mu}) = \ker g'_{A(\bar{x})}(\bar{x})$.

Lemma 4.48. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a local solution of problem (4.210), satisfying the LICQ. Let*

$\bar{\mu} \in \mathbf{R}^m$ be the (unique) Lagrange multiplier associated with $\bar{x}$, and assume that the strict complementarity condition $\bar{\mu}_{A(\bar{x})} > 0$ and the SOSC

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\mu})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in \ker g'_{A(\bar{x})}(\bar{x}) \setminus \{0\}$$

hold.

Then for any $(x^k,\, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^m$ close enough to $(\bar{x},\, \bar{\mu})$, Algorithm 4.46 with

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k,\, \mu^k) \tag{4.222}$$

uniquely defines $\tilde{p}^k \in \mathbf{R}^n$, $p^k \in \mathbf{R}^n$, $\bar{\mu}^{k+1} \in \mathbf{R}^m$, $\bar{p}^k \in \mathbf{R}^n$, and the next iterate $(x^{k+1},\, \mu^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^m$, and the following relations hold:

$$\tilde{p}^k = O(\|x^k - \bar{x}\|) + o(\|\mu^k - \bar{\mu}\|), \tag{4.223}$$

$$p^k = O(\|x^k - \bar{x}\|) + o(\|\mu^k - \bar{\mu}\|), \tag{4.224}$$

$$\bar{\mu}^{k+1} - \bar{\mu} = o(\|(x^k - \bar{x},\, \mu^k - \bar{\mu})\|), \tag{4.225}$$

$$p^k - \tilde{p}^k = O(\|\tilde{p}^k\|^\nu), \tag{4.226}$$

$$\bar{p}^k = O\left(\|p^k\|^2 + \max_{i \in A(\bar{x})} \left| \frac{\mu_i^k}{\bar{\mu}_i^{k+1}} - 1 \right| \|p^k\| \right) \tag{4.227}$$

as $(x^k,\, \mu^k) \to (\bar{x},\, \bar{\mu})$, and

$$g(x^{k+1}) < 0. \tag{4.228}$$

Proof. First, recall that the matrix

$$J_k = \begin{pmatrix} H_k & (g'(x^k))^{\mathrm{T}} \\ \mathrm{diag}(\mu^k) g'(x^k) & \mathrm{diag}(g(x^k)) \end{pmatrix} \tag{4.229}$$

of the linear system (4.212) with H_k defined in (4.222) coincides with the Jacobian of the system (4.211) at $(x^k,\, \mu^k)$. By Lemmas 4.43 and A.6, there exists $M > 0$ such that for all $(x^k,\, \mu^k)$ close enough to $(\bar{x},\, \bar{\mu})$, this matrix is nonsingular, and $\|J_k^{-1}\| \leq M$. Thus, $\tilde{x}^{k+1}$, $\tilde{p}^k$, $\tilde{\mu}^{k+1}$, p^k and $\bar{\mu}^{k+1}$ are uniquely defined.

Moreover, with $v^k = 0$, the linear system in question is the iteration system of Algorithm 4.42 applied to problem (4.210) or, to put it in other words, of the usual Newton method applied to the system (4.211). Therefore, Theorem 4.44 yields the estimate

$$(\tilde{x}^{k+1} - \bar{x},\, \tilde{\mu}^{k+1} - \bar{\mu}) = o(\|(x^k - \bar{x},\, \mu^k - \bar{\mu})\|)$$

$(x^k,\, \mu^k) \to (\bar{x},\, \bar{\mu})$. In particular, for $\tilde{p}^k = \tilde{x}^{k+1} - x^k$ it holds that

$$\begin{aligned}\|\tilde{p}^k\| &\leq \|x^k - \bar{x}\| + \|\tilde{x}^{k+1} - \bar{x}\| \\ &= \|x^k - \bar{x}\| + o(\|(x^k - \bar{x},\, \mu^k - \bar{\mu})\|) \\ &= O(\|x^k - \bar{x}\|) + o(\|\mu^k - \bar{\mu}\|),\end{aligned}$$

which proves (4.223).

Furthermore, with $v_i^k = t_k\|\tilde{p}^k\|^\nu$, $i = 1, \ldots, m$, the linear system in question is a perturbation of the iteration system of the Newton method applied to the system (4.211), where the perturbation term v^k satisfies the estimate $\|v^k\| \leq \|\tilde{p}^k\|^\nu = o(\|x^k - \bar{x}\| + \|\mu^k - \bar{\mu}\|)$; the latter relations follow from the estimate (4.223) established above and the fact that $\nu > 1$. Applying Lemma 2.3, we derive the estimate

$$(x^k + p^k - \bar{x},\, \bar{\mu}^{k+1} - \bar{\mu}) = o(\|(x^k - \bar{x},\, \mu^k - \bar{\mu})\|)$$

as $(x^k,\, \mu^k) \to (\bar{x},\, \bar{\mu})$, implying (4.224) and (4.225).

Since $(x^k + \tilde{p}^k,\, \tilde{\mu}^{k+1})$ and $(x^k + p^k,\, \bar{\mu}^{k+1})$ solve two linear systems with the same matrix J_k, and with the right-hand sides differing only by the term v^k satisfying $\|v^k\| \leq \|\tilde{p}^k\|^\nu$, we conclude that

$$\|p^k - \tilde{p}^k\| \leq M\|\tilde{p}^k\|^\nu,$$

which proves (4.226). Note that since $\nu \geq 2$, (4.226) further implies the estimate

$$\tilde{p}^k = O(\|p^k\|) \tag{4.230}$$

as $(x^k,\, \mu^k) \to (\bar{x},\, \bar{\mu})$.

The existence and uniqueness of $\bar{p}^k$ was established in Lemma 4.47. Combining (4.220) with (4.230), and employing the restriction $\nu \geq 2$, we derive the estimate (4.227). By the constraints in (4.215) and by (4.227), we further conclude that

$$\begin{aligned} g_i(x^k + p^k + \bar{p}^k) &= g_i(x^k + p^k) + \langle g_i'(x^k + p^k),\, \bar{p}^k\rangle + O(\|\bar{p}^k\|^2) \\ &= g_i(x^k + p^k) + \langle g_i'(x^k),\, \bar{p}^k\rangle + O(\|p^k\|\|\bar{p}^k\| + \|\bar{p}^k\|^2) \\ &= -w_k \\ &\quad + O\left(\|p^k\|^3 + \max_{i\in A(\bar{x})}\left|\frac{\mu_i^k}{\bar{\mu}_i^{k+1}} - 1\right|\|p^k\|^2\right) \quad \forall\, i \in A(\bar{x}) \end{aligned}$$

as $(x^k,\, \mu^k) \to (\bar{x},\, \bar{\mu})$. Moreover, this value is negative for all $(x^k,\, \mu^k)$ close enough to $(\bar{x},\, \bar{\mu})$, because of (4.216) and of the restrictions $\tau_1 < 3$ and $\tau_2 < 1$. This proves (4.228), because, by (4.224) and (4.227), the value of $g_{\{1,\ldots,m\}\setminus A(\bar{x})}(x^k + p^k + \bar{p}^k)$ can be made arbitrary close to the value $g_{\{1,\ldots,m\}\setminus A(\bar{x})}(\bar{x}) < 0$ by taking $(x^k,\, \mu^k)$ close enough to $(\bar{x},\, \bar{\mu})$. $\square$

We are now in position to state the main a priori local convergence result for the class of algorithms in consideration. The analysis is yet again by means of the perturbed Newtonian framework.

Theorem 4.49. *Under the assumptions of Lemma 4.48, let the parameter τ in Algorithm 4.46 satisfy $\tau > 1$.*

Then any starting point $(x^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\mu})$ uniquely defines the iterative sequence of Algorithm 4.46, this sequence converges to $(\bar{x}, \bar{\mu})$, and the rate of convergence is superlinear.

Proof. According to Lemma 4.48, if (x^k, μ^k) is close enough to $(\bar{x}, \bar{\mu})$, Algorithm 4.46 uniquely defines $\tilde{p}^k \in \mathbf{R}^n$, $p^k \in \mathbf{R}^n$, $\bar{\mu}^{k+1} \in \mathbf{R}^m$, $\bar{p}^k \in \mathbf{R}^n$, and the next iterate $(x^{k+1}, \mu^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^m$, satisfying (4.223)–(4.225), (4.227).

By the construction of the algorithm, and by (4.222), the new primal-dual iterate (x^{k+1}, μ^{k+1}) satisfies the following perturbed version of the iteration system of the Newton method for (4.211):

$$\begin{gathered}\frac{\partial L}{\partial x}(x^k, \mu^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \mu^k)(x - x^k) + (g'(x^k))^{\mathrm{T}}(\mu - \mu^k) + \omega_1^k = 0,\\ \mu_i^k \langle g_i'(x^k), x - x^k\rangle + \mu_i g_i(x^k) + (\omega_3^k)_i = 0, \quad i = 1, \ldots, m,\end{gathered}$$

with

$$\omega_1^k = (g'(x^k))^{\mathrm{T}}(\bar{\mu}^{k+1} - \mu^{k+1}) - \frac{\partial^2 L}{\partial x^2}(x^k, \mu^k)\bar{p}^k, \tag{4.231}$$

$$(\omega_3^k)_i = (\bar{\mu}_i^{k+1} - \mu_i^{k+1})g_i(x^k) + \mu_i^k t_k \|\tilde{p}^k\|^\nu - \mu_i^k \langle g_i'(x^k), \bar{p}^k\rangle, \quad i = 1, \ldots, m. \tag{4.232}$$

From (4.217) and (4.225), it follows that if (x^k, μ^k) is close enough to $(\bar{x}, \bar{\mu})$, then

$$(\bar{\mu}^{k+1} - \mu^{k+1})_{A(\bar{x})} = 0, \tag{4.233}$$

$$\begin{aligned}\|(\bar{\mu}^{k+1} - \mu^{k+1})_{\{1,\ldots,m\}\setminus A(\bar{x})}\| &\le \|\bar{\mu}^{k+1}_{\{1,\ldots,m\}\setminus A(\bar{x})}\| + \|p^k\|^\tau\\ &= o(\|(x^k - \bar{x}, \mu^k - \bar{\mu})\|)\end{aligned} \tag{4.234}$$

as $(x^k, \mu^k) \to (\bar{x}, \bar{\mu})$, where the last equality also employs estimate (4.224), the fact that $\bar{\mu}_{\{1,\ldots,m\}\setminus A(\bar{x})} = 0$, and the restriction $\tau > 1$. Furthermore, by (4.227), again employing (4.224), (4.225), we obtain that

$$\bar{p}^k = o(\|(x^k - \bar{x}, \mu^k - \bar{\mu})\|). \tag{4.235}$$

Combining (4.233)–(4.235) with with (4.231), (4.232), and employing (4.223), we finally derive the estimates

$$\begin{gathered}\omega_1^k = O(\|\bar{p}^k\|) + o(\|(x^k - \bar{x}, \mu^k - \bar{\mu})\|) = o(\|(x^k - \bar{x}, \mu^k - \bar{\mu})\|),\\ \omega_3^k = O(\|\tilde{p}^k\|^\nu + \|\bar{p}^k\|) + o(\|(x^k - \bar{x}, \mu^k - \bar{\mu})\|) = o(\|(x^k - \bar{x}, \mu^k - \bar{\mu})\|)\end{gathered}$$

as $(x^k, \mu^k) \to (\bar{x}, \bar{\mu})$, where the restriction $\nu \geq 2$ was again taken into account. The stated assertions now follow from Propositions 2.4 and 2.6. □

Assuming the local Lipschitz-continuity of the second derivatives of f and g with respect to $\bar{x}$, and appropriately modifying the analysis above, one can obtain quadratic convergence rate; we omit the details.

For a posteriori analysis, we need to relate Algorithm 4.46 to the perturbed SQP framework. For problem (4.210), the perturbed SQP iteration system (4.81) takes the form

$$\begin{gathered}\frac{\partial L}{\partial x}(x^k, \mu^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \mu^k)(x - x^k) + (g'(x^k))^{\mathrm{T}}(\mu - \mu^k) + \omega_1^k = 0,\\ \mu \geq 0, \quad g(x^k) + g'(x^k)(x - x^k) + \omega_3^k \leq 0,\\ \langle \mu,\, g(x^k) + g'(x^k)(x - x^k) + \omega_3^k\rangle = 0.\end{gathered} \tag{4.236}$$

The superlinear primal convergence rate for Algorithm 4.46 appears to require rather strong assumptions, relative to various other algorithms considered previously (quasi-Newton SQP, LCL, SQCQP, quasi-Newton SQP with second-order corrections, or even Algorithm 4.42). This is because different ingredients of this complex algorithm come into play with their own requirements, which need to be put together to obtain a complete picture. For instance, the contribution of the second-order correction term $\bar{p}^k$ to the overall size of the step $p^k + \bar{p}^k$ can hardly be efficiently estimated here without the LICQ. On the other hand, p^k depends on $\tilde{p}^k$, and hence, the two terms must be related to each other, which becomes problematic if we do not assure that for the iteration matrices J_k defined by (4.229), the sequence $\{J_k^{-1}\}$ is bounded.

The latter issue is addressed in the following lemma. Recall that when H_k is defined according to (4.222), the needed property is a consequence of Lemmas 4.43 and A.6, which was already employed in the proof of Lemma 4.48. Here, however, we would like to avoid assuming (4.222), or even assuming that H_k is asymptotically close to $\frac{\partial^2 L}{\partial x^2}(x^k, \mu^k)$. The weaker Dennis–Moré condition (4.239) below is much more appropriate in this context.

Lemma 4.50. *Let $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable in a neighborhood of a point $\bar{x} \in \mathbf{R}^n$, with its second derivative being continuous at $\bar{x}$. Assume that $g(\bar{x}) \leq 0$, and let $\bar{\mu} \in \mathbf{R}^m$ be such that $\bar{\mu}_{A(\bar{x})} > 0$ and $\bar{\mu}_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})} = 0$. Let $\{(x^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^m$ be any sequence convergent to $(\bar{x}, \bar{\mu})$, and let $\{H_k\} \subset \mathbf{R}^{(n+m)\times(n+m)}$ be a bounded sequence of symmetric matrices for which there exists $\gamma > 0$ such that*

$$\langle H_k \xi, \xi\rangle \geq \gamma \|\xi\|^2 \quad \forall\, \xi \in \ker g'_{A(\bar{x})}(\bar{x}) \tag{4.237}$$

holds for all k large enough.

Then there exists $M > 0$ such that for all k large enough, the matrix J_k defined by (4.229) is nonsingular, and $\|J_k^{-1}\| \leq M$.

Proof. It suffices to show that there exists $\tilde{\gamma} > 0$ such that

$$\|J_k(\xi, \zeta)\| \geq \tilde{\gamma}\|(\xi, \zeta)\| \quad \forall\, (\xi, \zeta) \in \mathbf{R}^n \times \mathbf{R}^m$$

holds for all k large enough. Then the needed assertion will evidently be valid with $M = 1/\tilde{\gamma}$.

Suppose that there is no such $\tilde{\gamma}$, which means that, passing to a subsequence of $\{J_k\}$ if necessary, we can assume the existence of a sequence $\{(\xi^k, \zeta^k)\} \subset \mathbf{R}^n \times \mathbf{R}^m$ such that $\|(\xi^k, \zeta^k)\| = 1$ for all k, and

$$\{J_k(\xi^k, \zeta^k)\} \to 0. \tag{4.238}$$

Since $\{H_k\}$ is bounded, passing to a further subsequence if necessary, we can assume that it converges to some $H \in \mathbf{R}^{(n+m)\times(n+m)}$, and $\{(\xi^k, \zeta^k)\}$ converges to some $(\xi, \zeta) \in \mathbf{R}^n \times \mathbf{R}^m$ such that $\|(\xi, \zeta)\| = 1$. By (4.229) and (4.238) we then obtain that $J(\xi, \eta) = 0$, where

$$J = \begin{pmatrix} H & (g'(\bar{x}))^{\mathrm{T}} \\ \mathrm{diag}(\bar{\mu})g'(\bar{x}) & \mathrm{diag}(g(\bar{x})) \end{pmatrix},$$

and thus, J is singular. At the same time, (4.237) implies the condition

$$\langle H\xi, \xi\rangle > 0 \quad \forall\, \xi \in \ker g'_{A(\bar{x})}(\bar{x}) \setminus \{0\}.$$

It remains to follow the proof of Lemma 4.43 with $\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})$ replaced by H, and with all the objects concerned with equality constraints dropped. $\square$

Theorem 4.51. *Under the assumptions of Lemma 4.48, let an iterative sequence $\{(x^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^m$ generated by Algorithm 4.46 be convergent to $(\bar{x}, \bar{\mu})$.*

If the sequence $\{H_k\}$ is bounded, and there exists $\gamma > 0$ such that (4.237) holds for all k large enough, and if the Dennis–Moré condition

$$\pi_{\ker g'_{A(\bar{x})}(\bar{x})}\left(\frac{\partial^2 L}{\partial x^2}(x^k, \mu^k) - H_k\right)(x^{k+1} - x^k) = o(\|x^{k+1} - x^k\|) \tag{4.239}$$

holds as $k \to \infty$, then the rate of convergence of $\{x^k\}$ is superlinear.

Proof. Let the sequences $\{\tilde{p}^k\} \subset \mathbf{R}^n$, $\{\tilde{\mu}^k\} \subset \mathbf{R}^m$, $\{p^k\} \subset \mathbf{R}^n$, $\{\bar{\mu}^k\} \subset \mathbf{R}^m$ and $\{\bar{p}^k\} \subset \mathbf{R}^n$ be defined according to Algorithm 4.46. Observe that $(x^k + p^k, \tilde{\mu}^{k+1})$ satisfies the linear system (4.212) with the matrix J_k defined in (4.229), and with the right-hand side tending to zero due to convergence of $\{(x^k, \mu^k)\}$ to $(\bar{x}, \bar{\mu})$. By Lemma 4.50, this implies that $\{\tilde{p}^k\}$ converges to 0, and then, by a similar reasoning, $\{p^k\}$ also converges to 0. Furthermore, according to (4.217), the order of the difference between $\bar{\mu}^{k+1}$ and μ^{k+1} cannot be greater than $O(\|p^k\|^{\tau})$, and since $\tau > 0$, this implies that $\{\bar{\mu}^k\}$ converges to the same limit $\bar{\mu}$ as $\{\mu^k\}$.

We argue similarly to the proof of Theorem 4.51. Define the auxiliary sequence $\{\check{\mu}^k\} \subset \mathbf{R}^m$ by setting

$$\check{\mu}^k_{A(\bar{x})} = \bar{\mu}^k_{A(\bar{x})}, \quad \check{\mu}^k_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})} = 0 \tag{4.240}$$

for all k. Furthermore, for all k define $\omega_1^k \in \mathbf{R}^n$ and $\omega_3^k \in \mathbf{R}^m$ by setting

$$\omega_1^k = \left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k,\, \check{\mu}^k)\right)(x^{k+1} - x^k) + (g'(x^k))^{\mathrm{T}}(\bar{\mu}^{k+1} - \check{\mu}^{k+1}) - H_k \bar{p}^k, \tag{4.241}$$

$$(\omega_3^k)_i = \left(\frac{\bar{\mu}_i^{k+1}}{\mu_i^k} - 1\right) g_i(x^k) + t_k \|\tilde{p}^k\|^{\nu} - \langle g_i'(x^k),\, \bar{p}^k\rangle, \quad i \in A(\bar{x}), \tag{4.242}$$

$$(\omega_3^k)_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})} = 0. \tag{4.243}$$

By the construction of the algorithm, by relation (4.219) in Lemma 4.47, and by (4.240)–(4.243), we obtain that

$$\frac{\partial L}{\partial x}(x^k,\, \check{\mu}^k) + H_k(x^{k+1} - x^k) + (g'(x^k))^{\mathrm{T}}(\check{\mu}^{k+1} - \check{\mu}^k) + \omega_1^k = 0, \tag{4.244}$$

and relations (4.204), (4.205) hold as well for all k large enough.

Combining (4.244) and (4.204), (4.205), we obtain that $(x^{k+1},\, \check{\mu}^{k+1})$ satisfies (4.236) with μ^k replaced by $\check{\mu}^k$, with ω_1^k defined in (4.241), and with ω_3^k defined in (4.242), (4.243).

We proceed with estimating the perturbation terms. First observe that similarly to the corresponding part of the proof of Lemma 4.48, by Lemma 4.50 we derive the estimate (4.230). Then the estimate (4.220) given by Lemma 4.47 transforms into (4.227). Furthermore, since $\{\bar{\mu}^k\}$ and $\{\mu^k\}$ converge to the same limit, the estimate (4.227) yields

$$\bar{p}^k = o(\|p^k\|) \tag{4.245}$$

as $k \to \infty$. Since $p^k + \bar{p}^k = x^{k+1} - x^k$, we obtain $p^k = O(\|x^{k+1} - x^k\|) + o(\|p^k\|)$, which evidently implies the estimate

$$p^k = O(\|x^{k+1} - x^k\|). \tag{4.246}$$

Using the latter, we transform (4.230) and (4.245) into

$$\tilde{p}^k = O(\|x^{k+1} - x^k\|) \tag{4.247}$$

and

$$\bar{p}^k = o(\|x^{k+1} - x^k\|) \tag{4.248}$$

as $k \to \infty$, respectively.

Since $(x^k + p^k,\, \bar{\mu}^{k+1})$ satisfies (4.195) with $v_i^k = t_k\|\tilde{p}^k\|^{\nu}$, $i = 1,\,\ldots,\,m$, by the last block of relations in (4.195) and by (4.246), (4.247) we obtain that

$$\begin{aligned}\bar{\mu}_i^{k+1} &= -\frac{\mu_i^k}{g_i(x^k)}(\langle g_i'(x^k),\, p^k\rangle + t_k\|\tilde{p}^k\|^\nu)\\ &= o(\|x^{k+1} - x^k\|) \quad \forall\, i \in \{1,\, \ldots,\, m\} \setminus A(\bar{x})\end{aligned} \tag{4.249}$$

as $k \to \infty$, because $\{\mu^k_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}\}$ converges to $\bar{\mu}_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})} = 0$, while $\{g_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}(x^k)\}$ converges to $g_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}(\bar{x}) < 0$, and because $\nu \geq 2$.

From (4.240)–(4.242) and (4.247)–(4.249), we now derive the following estimates:

$$\begin{aligned}\omega_1^k &= \left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k,\, \mu^k)\right)(x^{k+1} - x^k)\\ &\quad + (g''(x^k)[x^{k+1} - x^k])^{\mathrm{T}}(\breve{\mu}^k - \mu^k) + (g'(x^k))^{\mathrm{T}}(\bar{\mu}^{k+1} - \breve{\mu}^{k+1}) - H_k\bar{p}^k\\ &= \left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k,\, \mu^k)\right)(x^{k+1} - x^k)\\ &\quad + (g'_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}(x^k))^{\mathrm{T}}\bar{\mu}^{k+1}_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}\\ &\quad + O((\|\bar{\mu}^k_{A(\bar{x})} - \mu^k_{A(\bar{x})}\| + \|\mu^k_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}\|)\|x^{k+1} - x^k\|)\\ &\quad + o(\|x^{k+1} - x^k\|)\\ &= \left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k,\, \mu^k)\right)(x^{k+1} - x^k) + o(\|x^{k+1} - x^k\|),\end{aligned} \tag{4.250}$$

$$\begin{aligned}(\omega_3^k)_i &= \left(\frac{\bar{\mu}_i^{k+1}}{\mu_i^k} - 1\right)\langle g_i'(\bar{x}),\, x^k - \bar{x}\rangle + t_k\|\tilde{p}^k\|^\nu - \langle g_i'(x^k),\, \bar{p}^k\rangle\\ &\quad + O(\|x^k - \bar{x}\|^2)\\ &= O\left(\|x^k - \bar{x}\|^2 + \left|\frac{\bar{\mu}_i^{k+1}}{\mu_i^k} - 1\right|\|x^k - \bar{x}\|\right) + o(\|x^{k+1} - x^k\|)\\ &= o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \quad \forall\, i \in A(\bar{x})\end{aligned} \tag{4.251}$$

as $k \to \infty$, where the second equality in (4.250) employs the assumption that $\{H_k\}$ is bounded, the second equality in (4.251) employs the restriction $\nu > 1$, while the last equalities in (4.250) and (4.251) are due to the fact that $\{\mu^k\}$ and $\{\bar{\mu}^k\}$ converge to the same limit $\bar{\mu}$.

Putting together (4.239), (4.243), (4.250), (4.251), we finally get the estimates

$$\pi_{\ker g'_{A(\bar{x})}(\bar{x})}(\omega_1^k) = o(\|x^{k+1} - x^k\|), \quad \omega_3^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)$$

as $k \to \infty$, and the assertions now follow from Proposition 4.19. $\square$

We now get back to the more general equality and inequality-constrained problem (4.73). The following nice trick for treating equality constraints in algorithms that would generate feasible primal iterates if there were only inequalities was proposed in [197]. Replace (4.73) by the inequality-constrained problem

$$\begin{array}{ll} \text{minimize} & f(x) - c\sum_{j=1}^{l} h_j(x) \\ \text{subject to} & h(x) \leq 0,\ g(x) \leq 0, \end{array} \tag{4.252}$$

where $c > 0$ is a penalty parameter, and apply Algorithm 4.46 to this problem. Note that on the feasible set of problem (4.252), the objective function of this problem coincides with the function $\varphi_c : \mathbf{R}^n \to \mathbf{R}$,

$$\varphi_c(x) = f(x) + c\|h(x)\|_1,$$

where the last term is the l_1-penalty for the equality constraints. As will be shown in Sect. 6.1, under appropriate assumptions the equality constraints in problem (4.73) can be in a certain sense ignored if f is replaced by φ_c with c large enough. However, φ_c is generally a nonsmooth function. But keeping in (4.252) the inequalities corresponding to the equality constraints allows to remove this nonsmoothness along the iterations, since the iterates the algorithm generates are strictly feasible.

Instead of (4.194), in the approach of this section one can also employ the equation reformulations of the KKT system (4.74) considered in Sect. 3.2.2. For one example of such a development, see [226] where some advantages of this modification are discussed.

Finally, we mention that primal-dual interior-point methods for problem (4.210) can be regarded as perturbed versions of the approach of this section.

4.4 Semismooth Sequential Quadratic Programming

Consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0,\ g(x) \leq 0, \end{array} \tag{4.253}$$

where the objective function $f : \mathbf{R}^n \to \mathbf{R}$ and the constraints mappings $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are differentiable but not necessarily twice differentiable, with their derivatives being locally Lipschitz-continuous. Some sources of optimization problems with these smoothness properties have already been discussed in Sect. 3.3.

Recall that stationary points of problem (4.253) and the associated Lagrange multipliers are characterized by the KKT optimality system

$$\frac{\partial L}{\partial x}(x,\, \lambda,\, \mu) = 0, \quad h(x) = 0, \quad \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu,\, g(x)\rangle = 0, \tag{4.254}$$

where $L : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}$ is the Lagrangian of problem (4.253):

$$L(x,\, \lambda,\, \mu) = f(x) + \langle \lambda,\, h(x)\rangle + \langle \mu,\, g(x)\rangle.$$

The generic SQP method for problem (4.253) was presented in Algorithm 4.13 above. We repeat it here for convenience.

Algorithm 4.52 Choose $(x^0,\, \lambda^0,\, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and set $k = 0$.

1. If $(x^k,\, \lambda^k,\, \mu^k)$ satisfies the KKT system (4.254), stop.
2. Choose a symmetric matrix $H_k \in \mathbf{R}^{n\times n}$ and compute $x^{k+1} \in \mathbf{R}^n$ as a stationary point of the problem

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, x - x^k\rangle + \dfrac{1}{2}\langle H_k(x - x^k),\, x - x^k\rangle \\ \text{subject to} & h(x^k) + h'(x^k)(x - x^k) = 0,\ g(x^k) + g'(x^k)(x - x^k) \leq 0, \end{array} \tag{4.255}$$

   and $(\lambda^{k+1},\, \mu^{k+1}) \in \mathbf{R}^l \times \mathbf{R}^m$ as an associated Lagrange multiplier.
3. Increase k by 1 and go to step 1.

By the basic *semismooth SQP method* we mean Algorithm 4.52 with

$$H_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k). \tag{4.256}$$

Methods of this kind were considered, e.g., in [115, 223]. Here we follow the approach in [158].

The semismooth SQP method is naturally related to the semismooth Josephy–Newton method considered in Sect. 3.3, applied to the generalized equation (GE)

$$\Phi(u) + N(u) \ni 0, \tag{4.257}$$

with the mapping $\Phi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ given by

$$\Phi(u) = \left(\frac{\partial L}{\partial x}(x,\, \lambda,\, \mu),\, h(x),\, -g(x)\right), \tag{4.258}$$

and with

$$N(\cdot) = N_Q(\cdot), \quad Q = \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m_+, \tag{4.259}$$

where $u = (x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$.

Indeed, the KKT system of problem (4.255) has the form

$$\begin{gathered} f'(x^k) + H_k(x - x^k) + (h'(x^k))^{\mathrm{T}}\lambda + (g'(x^k))^{\mathrm{T}}\mu = 0, \\ h(x^k) + h'(x^k)(x - x^k) = 0, \\ \mu \geq 0, \quad g(x^k) + g'(x^k)(x - x^k) \leq 0, \quad \langle \mu,\, g(x^k) + g'(x^k)(x - x^k)\rangle = 0. \end{gathered} \tag{4.260}$$

It can be seen that (4.260) is equivalent to the GE

$$\Phi(u^k) + J_k(u - u^k) + N(u) \ni 0, \tag{4.261}$$

where $u^k = (x^k, \lambda^k, \mu^k)$, with Φ and N defined according to (4.258) and (4.259), respectively, and with

$$J_k = \begin{pmatrix} H_k & (h'(x^k))^{\mathrm{T}} & (g'(x^k))^{\mathrm{T}} \\ h'(x^k) & 0 & 0 \\ -g'(x^k) & 0 & 0 \end{pmatrix}. \tag{4.262}$$

Now, the GE (4.261) is precisely an iteration of the Josephy–Newton method for the GE (4.257).

The analysis of the semismooth Josephy–Newton method in Sect. 3.3 relies on the notion of strong regularity in the sense of Definition 3.41. We next give a characterization of strong regularity for optimization problems.

Proposition 4.53. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$, and $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable at $\bar{x} \in \mathbf{R}^n$. Let $\bar{x}$ be a stationary point of problem (4.253), and let $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be an associated Lagrange multiplier. Let $H \in \mathbf{R}^{n\times n}$ be an arbitrary symmetric matrix and let*

$$J = \begin{pmatrix} H & (h'(\bar{x}))^{\mathrm{T}} & (g'(\bar{x}))^{\mathrm{T}} \\ h'(\bar{x}) & 0 & 0 \\ -g'(\bar{x}) & 0 & 0 \end{pmatrix}. \tag{4.263}$$

If $\bar{x}$ and $(\bar{\lambda}, \bar{\mu})$ satisfy the LICQ and the condition

$$\langle H\xi, \xi\rangle > 0 \quad \forall\, \xi \in C_+(\bar{x}, \bar{\mu}) \setminus \{0\}, \tag{4.264}$$

where

$$C_+(\bar{x}, \bar{\mu}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+(\bar{x},\, \bar{\mu})}(\bar{x})\xi = 0\},$$

then $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ is a strongly regular solution of the GE

$$\Phi(\bar{u}) + J(u - \bar{u}) + N(u) \ni 0 \tag{4.265}$$

with Φ and N defined according to (4.258) and (4.259), respectively.

Moreover, the LICQ is necessary for strong regularity of $\bar{u}$, while the condition (4.264) is necessary for strong regularity of $\bar{u}$ if $\bar{x}$ is a local solution of the quadratic programming problem

$$\begin{array}{ll} \text{minimize} & \langle f'(\bar{x}), x - \bar{x}\rangle + \dfrac{1}{2}\langle H(x - \bar{x}), x - \bar{x}\rangle \\ \text{subject to} & h'(\bar{x})(x - \bar{x}) = 0,\ g'_{A(\bar{x})}(\bar{x})(x - \bar{x}) \leq 0. \end{array} \tag{4.266}$$

Proof. Problem (4.266) is locally (near $\bar{x}$) equivalent to the problem

$$\begin{array}{ll} \text{minimize} & \langle f'(\bar{x}), x - \bar{x}\rangle + \dfrac{1}{2}\langle H(x - \bar{x}), x - \bar{x}\rangle \\ \text{subject to} & h(\bar{x}) + h'(\bar{x})(x - \bar{x}) = 0,\ g(\bar{x}) + g'(\bar{x})(x - \bar{x}) \leq 0. \end{array} \tag{4.267}$$

It can be easily seen that the KKT system for problem (4.267) can be stated as the GE (4.265) with Φ and N defined according to (4.258) and (4.259), respectively, and with J defined in (4.263). Moreover, stationarity of $\bar{x}$ in problem (4.253) with an associated Lagrange multiplier $(\bar{\lambda},\, \bar{\mu})$ is equivalent to stationarity of $\bar{x}$ in problem (4.267) with the same Lagrange multiplier $(\bar{\lambda},\, \bar{\mu})$; the sets of active at $\bar{x}$ inequality constraints of the two problems are the same; the LICQ for the two problems at $\bar{x}$ means the same; and finally, condition (4.264) coincides with the SSOSC for problem (4.267) at $\bar{x}$. The stated assertions now follow from Proposition 1.28 applied to problem (4.267). □

We proceed with some technical results concerned with partial generalized Jacobians that will be needed in the sequel.

Lemma 4.54. *Let $B\colon \mathbf{R}^n \to \mathbf{R}^{m\times l}$ be locally Lipschitz-continuous at $\bar{x} \in \mathbf{R}^n$ with Lipschitz constant $\ell_B > 0$, and let $b\colon \mathbf{R}^n \to \mathbf{R}^m$ be an arbitrary mapping. Define the mapping $\Psi\colon \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}^m$,*

$$\Psi(x,\, y) = B(x)y + b(x). \tag{4.268}$$

If Ψ is differentiable with respect to x at points $(\bar{x},\, y^1) \in \mathbf{R}^n \times \mathbf{R}^l$ and $(\bar{x},\, y^2) \in \mathbf{R}^n \times \mathbf{R}^l$ with some $y^1,\, y^2 \in \mathbf{R}^l$, then

$$\left\| \frac{\partial \Psi}{\partial x}(\bar{x},\, y^1) - \frac{\partial \Psi}{\partial x}(\bar{x},\, y^2) \right\| \leq \ell_B \|y^1 - y^2\|. \tag{4.269}$$

Proof. Differentiability of Ψ with respect to x at $(\bar{x},\, y^1)$ and $(\bar{x},\, y^2)$ means that for any $\xi \in \mathbf{R}^n$ and $j = 1,\, 2$, it holds that

$$\begin{aligned}
&B(\bar{x}+\xi)y^j + b(\bar{x}+\xi) - B(\bar{x})y^j - b(\bar{x}) - \frac{\partial \Psi}{\partial x}(\bar{x},\, y^j)\xi \\
&= \Psi(\bar{x}+\xi,\, y^j) - \Psi(\bar{x}+\xi,\, y^j) - \frac{\partial \Psi}{\partial x}(\bar{x},\, y^j)\xi \\
&= o(\|\xi\|)
\end{aligned}$$

as $\xi \to 0$. This implies the relation

$$(B(\bar{x}+\xi) - B(\bar{x}))(y^1 - y^2) - \left(\frac{\partial \Psi}{\partial x}(\bar{x},\, y^1) - \frac{\partial \Psi}{\partial x}(\bar{x},\, y^2) \right) \xi = o(\|\xi\|) \tag{4.270}$$

as $\xi \to 0$.

Fix an arbitrary $\xi \in \mathbf{R}^n$. By (4.270), using that B is locally Lipschitz-continuous at $\bar{x}$ with Lipschitz constant ℓ_B, for all $t > 0$ it holds that

$$\begin{aligned}
t \left\| \left(\frac{\partial \Psi}{\partial x}(\bar{x},\, y^1) - \frac{\partial \Psi}{\partial x}(\bar{x},\, y^2) \right) \xi \right\| &\leq \|B(\bar{x}+t\xi) - B(\bar{x})\| \|y^1 - y^2\| + o(t) \\
&\leq \ell_B t \|y^1 - y^2\| \|\xi\| + o(t)
\end{aligned}$$

as $t \to 0$. Dividing both sides of the latter relation by t and passing onto the limit as $t \to 0$, we obtain that

$$\left\| \left(\frac{\partial \Psi}{\partial x}(\bar{x},\, y^1) - \frac{\partial \Psi}{\partial x}(\bar{x},\, y^2) \right) \xi \right\| \leq \ell_B \|y^1 - y^2\| \|\xi\|.$$

Since ξ is arbitrary, the required estimate (4.269) follows. $\square$

Lemma 4.55. *Let $B\colon \mathbf{R}^n \to \mathbf{R}^{m\times l}$ and $b\colon \mathbf{R}^n \to \mathbf{R}^m$ be locally Lipschitz-continuous at a point $\bar{x} \in \mathbf{R}^n$, and define the mapping $\Psi\colon \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}^m$ according to (4.268). Let the two sequences $\{(x^k,\, y_1^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l$ and $\{(x^k,\, y_2^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l$ be both convergent to $(\bar{x},\, \bar{y})$ with some $\bar{y} \in \mathbf{R}^l$.*

Then for any sequence $\{W_k^1\} \subset \mathbf{R}^{m\times n}$ such that $W_k^1 \in \partial_x \Psi(x^k,\, y_1^k)$ for all k, there exists a sequence $\{W_k^2\} \subset \mathbf{R}^{m\times n}$ such that $W_k^2 \in \partial_x \Psi(x^k,\, y_2^k)$ for all k large enough, and

$$W_k^1 - W_k^2 = O(\|y_1^k - y_2^k\|)$$

as $k \to \infty$.

Proof. Let U be a neighborhood of $\bar{x}$, such that B and b are Lipschitz-continuous on U. Then the mapping $\Psi(\cdot,\, y_2^k)$ is evidently Lipschitz-continuous on U for all k. Therefore, by the Rademacher Theorem (Theorem 1.48), $\Psi(\cdot,\, y_2^k)$ is differentiable everywhere on $U \setminus \Gamma$, where the Lebesgue measure of the set $\Gamma \subset U$ is zero. Let $\mathcal{S}_k \subset U$ stand for the set of points of differentiability of $\Psi(\cdot,\, y_1^k)$. Since Clarke's generalized Jacobian is "blind" to sets of the Lebesgue measure zero (see Proposition 1.60), for all k large enough (so that $x^k \in U$) and for any matrix $W_k^1 \in \partial_x \Psi(x^k,\, y_1^k)$ there exist a positive integer s_k, matrices $W_{k,\,i}^1 \in \mathbf{R}^{m\times l}$ and reals $\alpha_{k,\,i} \geq 0$, $i = 1,\, \ldots,\, s_k$, such that $\sum_{i=1}^{s_k} \alpha_{k,\,i} = 1$, $W_k^1 = \sum_{i=1}^{s_k} \alpha_{k,\,i} W_{k,\,i}^1$, and for each $i = 1,\, \ldots,\, s_k$, there exists a sequence $\{x_j^{k,\,i}\} \subset \mathcal{S}_k \setminus \Gamma$ convergent to x^k and such that $\{\frac{\partial \Psi}{\partial x}(x_j^{k,\,i},\, y_1^k)\} \to W_{k,\,i}^1$ as $j \to \infty$.

Furthermore, by Lemma 4.54, for all k and j large enough it holds that

$$\left\| \frac{\partial \Psi}{\partial x}(x_j^{k,\,i},\, y_1^k) - \frac{\partial \Psi}{\partial x}(x_j^{k,\,i},\, y_2^k) \right\| \leq \ell_B \|y_1^k - y_2^k\| \quad \forall\, i = 1,\, \ldots,\, s_k, \tag{4.271}$$

where ℓ_B is the Lipschitz constant for B on U. For all k large enough and all $i = 1,\, \ldots,\, s_k$, since $\Psi(\cdot,\, y_2^k)$ is locally Lipschitz-continuous at x^k, the sequence $\{\frac{\partial \Psi}{\partial x}(x_j^{k,\,i},\, y_2^k)\}$ is bounded and therefore, passing to a subsequence if necessary, we can assume that it converges to some $W_{k,\,i}^2$ as $j \to \infty$. Then by passing onto the limit in (4.271) we derive the estimate

$$\|W_{k,\,i}^1 - W_{k,\,i}^2\| \leq \ell_B \|y_1^k - y_2^k\| \quad \forall\, i = 1,\, \ldots,\, s_k. \tag{4.272}$$

By the definition of B-differential we have that $W_{k,\,i}^2 \in (\partial_x)_B \Psi(x^k,\, y_2^k)$. Hence, by the definition of the generalized Jacobian, the convex combination $W_k^2 = \sum_{i=1}^{s_k} \alpha_k^i W_{k,\,i}^2$ belongs to $\partial_x \Psi(x^k,\, y_2^k)$, and employing (4.272) we derive the estimate

$$\begin{aligned}\|W_k^1 - W_k^2\| &= \left\| \sum_{i=1}^{s_k} \alpha_{k,\,i} W_{k,\,i}^1 - \sum_{i=1}^{s_k} \alpha_{k,\,i} W_{k,\,i}^2 \right\| \\ &\leq \sum_{i=1}^{s_k} \alpha_{k,\,i} \|W_{k,\,i}^1 - W_{k,\,i}^2\| \\ &\leq \ell_B \|y_1^k - y_2^k\|.\end{aligned}$$

□

Lemma 4.56. *Let $B\colon \mathbf{R}^n \to \mathbf{R}^{m\times l}$ and $b\colon \mathbf{R}^n \to \mathbf{R}^m$ be semismooth at $\bar{x} \in \mathbf{R}^n$, and define the mapping $\Psi\colon \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}^m$ according to (4.268). Let a sequence $\{(x^k,\, y^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l$ be convergent to $(\bar{x},\, \bar{y})$ with some $\bar{y} \in \mathbf{R}^l$.*

Then for any sequence $\{W_k\} \subset \mathbf{R}^{m\times n}$ such that $W_k \in \partial_x \Psi(x^k,\, y^k)$ for all k, it holds that

$$\Psi(x^k,\, y^k) - \Psi(\bar{x},\, y^k) - W_k(x^k - \bar{x}) = o(\|x^k - \bar{x}\|)$$

as $k \to \infty$.

Proof. Applying Lemma 4.55 with $y_1^k = y^k$ and $y_2^k = \bar{y}$ for all k, we conclude that there exists a sequence of matrices $\{\bar{W}_k\} \subset \mathbf{R}^{m\times n}$ for which it holds that $\bar{W}_k \in \partial_x \Psi(x^k,\, \bar{y})$ for all sufficiently large k and $W_k - \bar{W}_k = O(\|y^k - \bar{y}\|)$ as $k \to \infty$. Employing (4.268) and the semismoothness of B and b at $\bar{x}$, we then derive the estimate

$$\begin{aligned}&\|\Psi(x^k,\, y^k) - \Psi(\bar{x},\, y^k) - W_k(x^k - \bar{x})\| \\ &\leq \|(\Psi(x^k,\, y^k) - \Psi(x^k,\, \bar{y})) - (\Psi(\bar{x},\, y^k) - \Psi(\bar{x}, \bar{y}))\| \\ &\quad + \|(W_k - \bar{W}_k)(x^k - \bar{x})\| + \|\Psi(x^k,\, \bar{y}) - \Psi(\bar{x},\, \bar{y}) - \bar{W}_k(x^k - \bar{x})\| \\ &= \|(B(x^k) - B(\bar{x}))(y^k - \bar{y})\| + O(\|x^k - \bar{x}\|\|y^k - \bar{y}\|) + o(\|x^k - \bar{x}\|) \\ &= O(\|x^k - \bar{x}\|\|y^k - \bar{y}\|) + o(\|x^k - \bar{x}\|) \\ &= o(\|x^k - \bar{x}\|)\end{aligned}$$

as $k \to \infty$. □

Remark 4.57. It can be seen that for any $u = (x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$

$$\partial \Phi(u) = \left\{ \left(\begin{array}{ccc} H & (h'(x))^{\mathrm{T}} & (g'(x))^{\mathrm{T}} \\ h'(x) & 0 & 0 \\ -g'(x) & 0 & 0 \end{array} \right) \middle|\, H \in \partial_x \frac{\partial L}{\partial x}(x,\, \lambda,\, \mu) \right\}. \tag{4.273}$$

This readily follows from the results in [133], but we provide a simple direct proof, using the tools developed in Sect. 1.4.

Indeed, if $J \in \partial_B \Phi(u)$, then there exists a sequence $\{u^k\} \subset \mathcal{S}_\Phi$, where $u^k = (x^k,\, \lambda^k,\, \mu^k)$, convergent to u and such that $\{\Phi'(u^k)\} \to J$. Evidently, for all k

$$\Phi'(u^k) = \begin{pmatrix} \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k) & (h'(x^k))^{\mathrm{T}} & (g'(x^k))^{\mathrm{T}} \\ h'(x^k) & 0 & 0 \\ -g'(x^k) & 0 & 0 \end{pmatrix},$$

which implies that (4.263) holds with $H = \lim_{k\to\infty} \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k)$. According to Lemma 4.55, for each k there exists $H_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda,\, \mu)$ such that

$$\frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k) - H_k = O(\|(\lambda^k - \lambda,\, \mu^k - \mu)\|)$$

as $k \to \infty$, and hence, $\{H_k\}$ converges to H. By Proposition 1.51 we then conclude that $H \in \partial_x \frac{\partial L}{\partial x}(x,\, \lambda,\, \mu)$, which proves that $\partial_B \Phi(u)$ is contained in the set in the right-hand side of (4.273). Since this set is convex, by taking the convex hull of $\partial_B \Phi(u)$, we derive that $\partial \Phi(u)$ is also contained in this set.

Conversely, let J be of the form (4.263) with some $H \in (\partial_B)_x \frac{\partial L}{\partial x}(x,\, \lambda,\, \mu)$. Then there exists a sequence $\{x^k\} \subset \mathcal{S}_{\frac{\partial L}{\partial x}(\cdot,\, \lambda,\, \mu)}$ convergent to x and such that $\{\frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda,\, \mu)\} \to H$. Since L is affine in $(\lambda,\, \mu)$, it can be easily seen that for $u^k = (x^k,\, \lambda,\, \mu)$ it holds that $\{u^k\} \subset \mathcal{S}_{\frac{\partial L}{\partial x}} = \mathcal{S}_{\Phi}$, and at the same time, $\{u^k\}$ converges to u and $\{\Phi'(u^k)\}$ converges to J, implying that J belongs to $\partial_B \Phi(u)$. By taking the convex hulls, we derive that the right-hand side of (4.273) is contained in the left-hand side, completing the proof of (4.273).

By the equality (4.273), Proposition 4.53 immediately implies the following: if $\bar{x}$ and $(\bar{\lambda},\, \bar{\mu})$ satisfy the LICQ and the SSOSC

$$\forall\, H \in \partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) \quad \langle H\xi,\, \xi\rangle > 0 \quad \forall\, \xi \in C_+(\bar{x},\, \bar{\mu}) \setminus \{0\}, \tag{4.274}$$

then $\bar{u} = (\bar{x},\, \bar{\lambda},\, \bar{\mu})$ is a *CD*-regular solution of the GE (4.257) with Φ and N defined according to (4.258) and (4.259), respectively. In particular, in the case of twice differentiable data, Proposition 4.53 applied with $H = \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ recovers the characterization of strong regularity in Proposition 1.28.

By Propositions 1.73 and 1.75, semismoothness of the derivatives of f, h, and g at $\bar{x}$ implies semismoothness of Φ at $\bar{u} = (\bar{x},\, \bar{\lambda},\, \bar{\mu})$. Moreover, taking into account Proposition 4.53 and Remark 4.57, Theorem 3.47 is applicable with $\bar{\Delta} = \partial \Phi(\bar{x})$ and $\Delta(\cdot) = \partial \Phi(\cdot)$ provided $\bar{x}$ and $(\bar{\lambda},\, \bar{\mu})$ satisfy the LICQ and the SSOSC (4.274). Therefore, we obtain the following local convergence and rate of convergence result for the basic semismooth SQP algorithm.

Theorem 4.58. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their derivatives being semismooth at $\bar{x}$. Let $\bar{x}$ be a local solution of problem* (4.253)*, satisfying the LICQ, and let SSOSC* (4.274) *hold for the associated Lagrange multiplier $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$.*

Then there exists a constant $\delta > 0$ such that for any starting point $(x^0,\, \lambda^0,\, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, for each $k = 0,\, 1,\, \ldots$

and any choice of H_k satisfying (4.256)*, there exists the unique stationary point x^{k+1} of problem* (4.255) *and the unique associated Lagrange multiplier $(\lambda^{k+1},\, \mu^{k+1})$ satisfying*

$$\|(x^{k+1} - x^k,\, \lambda^{k+1} - \lambda^k,\, \mu^{k+1} - \mu^k)\| \leq \delta; \tag{4.275}$$

the sequence $\{(x^k,\, \lambda^k,\, \mu^k)\}$ converges to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the derivatives of f, h, and g are strongly semismooth at $\bar{x}$.

Theorem 4.58 essentially recovers the local superlinear convergence result in [115], which was obtained by a direct (and rather involved) analysis. Here, this property is an immediate consequence of the general local convergence theory for the semismooth Josephy–Newton method, given by Theorem 3.47. A similar result was derived in [223], but under stronger assumptions including the strict complementarity condition.

For optimization problems with twice differentiable data, Theorem 4.58 can be sharpened. Specifically, it was demonstrated in Theorem 4.14 that the LICQ can be replaced by the generally weaker SMFCQ, while the SSOSC can be replaced by the usual SOSC. However, unlike in Theorem 4.58, these assumptions do not guarantee uniqueness of the iteration sequence satisfying the localization condition (4.275).

We proceed with the primal superlinear convergence analysis for Algorithm 4.52. Having in mind some potentially useful choices of H_k different from the basic choice (4.256), as well as truncation of subproblems solution (see Sect. 4.3.2 and [147]), we consider a perturbed version of the semismooth SQP. For the given primal-dual iterate $(x^k,\, \lambda^k,\, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, the next iterate $(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1})$ satisfies the relations

$$\begin{gathered}
\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) + W_k(x^{k+1} - x^k) + (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) \\
+(g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) + \omega_1^k = 0, \\
h(x^k) + h'(x^k)(x^{k+1} - x^k) + \omega_2^k = 0, \\
\mu^{k+1} \geq 0,\ g(x^k) + g'(x^k)(x^{k+1} - x^k) + \omega_3^k \leq 0, \\
\langle \mu^{k+1}, g(x^k) + g'(x^k)(x^{k+1} - x^k) + \omega_3^k \rangle = 0,
\end{gathered} \tag{4.276}$$

with some $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)$, where $\omega_1^k \in \mathbf{R}^n$, $\omega_2^k \in \mathbf{R}^l$, and $\omega_3^k \in \mathbf{R}^m$ are the perturbation terms.

We first establish necessary conditions for the primal superlinear convergence of the iterates given by (4.276).

Proposition 4.59. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable in a neighborhood of a point $\bar{x} \in \mathbf{R}^n$, with their derivatives being semismooth at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (4.253)*, and let $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be an associated Lagrange multiplier. Let further $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ be convergent to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, and assume that for each k large enough the triple $(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1})$ satisfies the system* (4.276) *with some $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k)$ and some $\omega_1^k \in \mathbf{R}^n$, $\omega_2^k \in \mathbf{R}^l$ and $\omega_3^k \in \mathbf{R}^m$.*

If the rate of convergence of $\{x^k\}$ *is superlinear, then*

$$\frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - W_k(x^{k+1} - x^k) \\ = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|), \tag{4.277}$$

$$\omega_2^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|), \tag{4.278}$$

$$(\omega_3^k)_{A_+(\bar{x},\, \bar{\mu})} = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \tag{4.279}$$

as $k \to \infty$*. If in addition*

$$\{(\omega_3^k)_{\{1,\, \dots,\, m\}\setminus A(\bar{x})}\} \to 0, \tag{4.280}$$

then

$$\pi_{C(\bar{x})}(-\omega_1^k) = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \tag{4.281}$$

as $k \to \infty$*, where*

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A(\bar{x})}(\bar{x})\xi \le 0,\ \langle f'(\bar{x}),\, \xi\rangle \le 0\}$$

is the critical cone of problem (4.253) *at* $\bar{x}$*.*

Proof. Let $\{\tilde{W}_k\}$ be an arbitrary sequence of matrices such that, for each k, $\tilde{W}_k \in \partial_x \frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k)$. Since $\{(x^k,\, \lambda^k,\, \mu^k)\}$ is convergent to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ (and hence, $\{(\lambda^k,\, \mu^k)\}$ is bounded), and the derivatives of f, h, and g are locally Lipschitz-continuous at $\bar{x}$, one can easily see that there exist a neighborhood U of $\bar{x}$ and $\ell > 0$ such that for all k the mapping $\frac{\partial L}{\partial x}(\cdot,\, \lambda^k,\, \mu^k)$ is Lipschitz-continuous on U with constant ℓ, and both x^k and x^{k+1} belong to U for all k large enough. This implies that $\|W_k\| \le \ell$ and $\|\tilde{W}_k\| \le \ell$ for all such k (see Proposition 1.51). In particular, $\{W_k\}$ and $\{\tilde{W}_k\}$ are bounded sequences. Then, employing Lemma 4.56 (with $p = r = n$, $q = l + m$, $B(x) = ((h'(x))^{\mathrm{T}},\, (g'(x))^{\mathrm{T}})$, $b(x) = f'(x)$, $y^k = (\lambda^k,\, \mu^k)$, $\bar{y} = (\bar{\lambda},\, \bar{\mu})$) and taking into account the superlinear convergence of $\{x^k\}$ to $\bar{x}$, we obtain that

$$\begin{aligned}
&\frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - W_k(x^{k+1} - x^k) \\
&= \left(\frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(\bar{x}, \lambda^k, \mu^k) - \tilde{W}_k(x^{k+1} - \bar{x})\right) \\
&\quad - \left(\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(\bar{x}, \lambda^k, \mu^k) - W_k(x^k - \bar{x})\right) \\
&\quad - (W_k - \tilde{W}_k)(x^{k+1} - \bar{x}) \\
&= o(\|x^{k+1} - \bar{x}\|) + o(\|x^k - \bar{x}\|) + O(\|x^{k+1} - \bar{x}\|) \\
&= o(\|x^k - \bar{x}\|)
\end{aligned} \tag{4.282}$$

as $k \to \infty$, which gives (4.277).

Relations (4.278) and (4.279) are derived identically to the proof of Proposition 4.16.

Furthermore, from (4.276), employing the superlinear convergence of $\{x^k\}$ to $\bar{x}$, the boundedness of $\{W_k\}$, Lemma 4.56 and the local Lipschitz-continuity of the derivatives of h and g at $\bar{x}$, we obtain that

$$\begin{aligned}
-\omega_1^k &= \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) + W_k(x^{k+1} - x^k) \\
&\quad +(h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) \\
&= (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu}) \\
&\quad +(h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) \\
&\quad + \left(\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(\bar{x}, \lambda^k, \mu^k) - W_k(x^k - \bar{x})\right) + W_k(x^{k+1} - \bar{x}) \\
&= ((h'(x^k))^{\mathrm{T}} - (h'(\bar{x}))^{\mathrm{T}})(\lambda^{k+1} - \lambda^k) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) \\
&\quad +((g'(x^k))^{\mathrm{T}} - (g'(\bar{x}))^{\mathrm{T}})(\mu^{k+1} - \mu^k) + (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}) \\
&\quad +o(\|x^k - \bar{x}\|) \\
&= (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}) + o(\|x^k - \bar{x}\|)
\end{aligned}$$

as $k \to \infty$. Assuming (4.280), the last needed estimate (4.281) can now be derived the same way as in the proof of Proposition 4.16. $\square$

We now proceed with sufficient conditions for the primal superlinear convergence. Following Sect. 4.3, in this analysis we only assume that the limiting stationary point $\bar{x}$ and the associated limiting multiplier $(\bar{\lambda}, \bar{\mu})$ satisfy the SOSC

$$\forall\, H \in \partial_x \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) \quad \langle H\xi, \xi\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}. \tag{4.283}$$

This condition is indeed sufficient for local optimality of $\bar{x}$; see Theorem 1.82. Our analysis relies on the following primal error bound result that generalizes Proposition 1.46 with respect to its smoothness assumptions.

Proposition 4.60. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$, and $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable in a neighborhood of a point $\bar{x} \in \mathbf{R}^n$, with their derivatives being semismooth at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (4.253), let $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be an associated Lagrange multiplier, and assume that the SOSC (4.283) holds.*

Then the estimate

$$x - \bar{x} = O\left(\left\|\begin{pmatrix} \pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x, \lambda, \mu)\right) \\ h(x) \\ \min\{\mu, -g(x)\} \end{pmatrix}\right\|\right) \tag{4.284}$$

holds as $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ tends to $(\bar{x}, \bar{\lambda}, \bar{\mu})$.

Proof. The argument is along the lines of the proof of Proposition 1.44. Assuming that (4.284) does not hold, we obtain that there exists a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ convergent to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ and satisfying

$$\pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k)\right) = o(\|x^k - \bar{x}\|), \tag{4.285}$$

$$h(x^k) = o(\|x^k - \bar{x}\|), \tag{4.286}$$

$$\min\{\mu^k,\, -g(x^k)\} = o(\|x^k - \bar{x}\|) \tag{4.287}$$

as $k \to \infty$. Passing onto a subsequence, if necessary, we can assume that the sequence $\{(x^k - \bar{x})/\|x^k - \bar{x}\|\}$ converges to some $\xi \in \mathbf{R}^n$, $\|\xi\| = 1$, and that there exist index sets I_1 and I_2 such that $I_1 \cup I_2 = A_0$, $I_1 \cap I_2 = \emptyset$, and that the following relations hold:

$$\mu^k_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})} = o(\|x^k - \bar{x}\|), \tag{4.288}$$

$$\mu^k_{I_2} = o(\|x^k - \bar{x}\|) \tag{4.289}$$

as $k \to \infty$, and

$$h'(\bar{x})\xi = 0, \quad g'_{A_+(\bar{x},\, \bar{\mu})}(\bar{x})\xi = 0, \tag{4.290}$$

$$g'_{I_1}(\bar{x})\xi = 0, \quad g'_{I_2}(\bar{x})\xi \leq 0. \tag{4.291}$$

Furthermore, employing Lemma A.13 and (4.285) we obtain that

$$\begin{aligned} 0 &= \pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) - \pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)\right)\right) \\ &= \pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) + o(\|x^k - \bar{x}\|)\right), \end{aligned}$$

and hence, by Lemmas 4.56 and A.13,

$$\begin{aligned} (C(\bar{x}))^\circ \ni \frac{\partial L}{\partial x}&(x^k,\, \lambda^k,\, \mu^k) + o(\|x^k - \bar{x}\|) \\ &= \left(\frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) - \frac{\partial L}{\partial x}(\bar{x},\, \lambda^k,\, \mu^k)\right) \\ &\quad + \left(\frac{\partial L}{\partial x}(\bar{x},\, \lambda^k,\, \mu^k) - \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\right) + o(\|x^k - \bar{x}\|) \\ &= W_k(x^k - \bar{x}) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu}) \\ &\quad + o(\|x^k - \bar{x}\|) \end{aligned} \tag{4.292}$$

as $k \to \infty$, where $\{W_k\}$ is any sequence of matrices such that the inclusion $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)$ holds for each k.

By Lemma 4.55, there exists sequence $\{\bar{W}_k\}$ such that $\bar{W}_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \bar{\lambda},\, \bar{\mu})$ for k large enough and $W_k - \bar{W}_k = O(\|(\lambda^k - \bar{\lambda},\, \mu^k - \bar{\mu})\|)$ as $k \to \infty$.

Then, since the sequence $\{\bar{W}_k\}$ is bounded (by the Lipschitz-continuity of $\frac{\partial L}{\partial x}(\cdot,\, \bar{\lambda},\, \bar{\mu})$ on some neighborhood of $\bar{x}$, and by Proposition 1.51), and since $\{(\lambda^k,\, \mu^k)\}$ converges to $(\bar{\lambda},\, \bar{\mu})$, passing to a subsequence if necessary, we can assume that both $\{\bar{W}_k\}$ and $\{W_k\}$ converge to some $\bar{W} \in \mathbf{R}^{n\times n}$. By Proposition 1.51 it follows that $\bar{W} \in \partial_x \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})$. Note that, according to (4.290), (4.291), it holds that $\xi \in C(\bar{x})$. Therefore, taking into account (4.288)–(4.290), the first relation in (4.291), and the equality $\bar{\mu}_{\{1,\, \ldots,\, m\}\setminus A_+(\bar{x},\, \bar{\mu})} = 0$, from (4.292) we derive that

$$\begin{aligned}
0 &\geq \langle W_k(x^k - \bar{x}),\, \xi\rangle + \langle \lambda^k - \bar{\lambda},\, h'(\bar{x})\xi\rangle + \langle \mu^k - \bar{\mu},\, g'(\bar{x})\xi\rangle + o(\|x^k - \bar{x}\|)\\
&= \langle W_k(x^k - \bar{x}),\, \xi\rangle + \langle \mu^k_{I_2\cup(\{1,\, \ldots,\, m\}\setminus A(\bar{x}))},\, g'_{I_2\cup(\{1,\, \ldots,\, m\}\setminus A(\bar{x}))}(\bar{x})\xi\rangle\\
&\quad + o(\|x^k - \bar{x}\|)\\
&= \langle W_k(x^k - \bar{x}),\, \xi\rangle + o(\|x^k - \bar{x}\|)
\end{aligned}$$

as $k \to \infty$. Dividing the obtained relation by $\|x^k - \bar{x}\|$ and passing onto the limit, we conclude that

$$\langle \bar{W}\xi, \xi\rangle \leq 0,$$

which contradicts the SOSC (4.283) because $\xi \in C(\bar{x}) \setminus \{0\}$. □

We are now in position to give conditions that are sufficient for the primal superlinear convergence of the perturbed semismooth SQP method.

Proposition 4.61. *Under the assumptions of Proposition 4.59, suppose that the SOSC (4.283) holds.*

If conditions (4.278)–(4.280) hold, and in addition

$$\begin{aligned}
&\pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - W_k(x^{k+1} - x^k) - \omega_1^k\right)\\
&\quad = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)
\end{aligned} \tag{4.293}$$

and

$$(\omega_3^k)_{A_0(\bar{x},\, \bar{\mu})} = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|) \tag{4.294}$$

hold as $k \to \infty$, then the rate of convergence of $\{x^k\}$ is superlinear.

Proof. Employing convergence of $\{(x^k,\, \lambda^k,\, \mu^k)\}$ to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ and the local Lipschitz-continuity of the derivatives of h and g at $\bar{x}$, from the first and the second equalities in (4.276) we derive that

$$\begin{aligned}
\frac{\partial L}{\partial x}(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1}) &= \frac{\partial L}{\partial x}(x^{k+1},\, \lambda^{k},\, \mu^{k}) + (h'(x^{k+1}))^{\mathrm{T}}(\lambda^{k+1} - \lambda^{k}) \\
&\quad +(g'(x^{k+1}))^{\mathrm{T}}(\mu^{k+1} - \mu^{k}) \\
&= \frac{\partial L}{\partial x}(x^{k+1},\, \lambda^{k},\, \mu^{k}) - \frac{\partial L}{\partial x}(x^{k},\, \lambda^{k},\, \mu^{k}) \\
&\quad +\frac{\partial L}{\partial x}(x^{k},\, \lambda^{k},\, \mu^{k}) + (h'(x^{k}))^{\mathrm{T}}(\lambda^{k+1} - \lambda^{k}) \\
&\quad +(g'(x^{k}))^{\mathrm{T}}(\mu^{k+1} - \mu^{k}) + o(\|x^{k+1} - x^{k}\|) \\
&= \frac{\partial L}{\partial x}(x^{k+1},\, \lambda^{k},\, \mu^{k}) - \frac{\partial L}{\partial x}(x^{k},\, \lambda^{k},\, \mu^{k}) \\
&\quad -W_k(x^{k+1} - x^{k}) - \omega_1^k + o(\|x^{k+1} - x^{k}\|),
\end{aligned}$$

and hence, by (4.293),

$$\pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1})\right) = o(\|x^{k+1} - x^{k}\| + \|x^{k} - \bar{x}\|)$$

as $k \to \infty$. Literally repeating the remaining part of the proof of Proposition 4.17 (with $\pi = \pi_{C(\bar{x})}$), and employing Proposition 4.60, the assertion follows. □

Remark 4.62. The condition (4.293) follows from (4.277) and (4.281), and therefore, according to Proposition 4.59, it is in fact also necessary for the primal superlinear convergence rate (assuming (4.280)).

Remark 4.63. In Proposition 4.61 the SOSC (4.283) can be replaced by the following sequential second-order condition:

$$\liminf_{k\to\infty} \max_{W \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)} \langle W\xi,\, \xi\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\} \tag{4.295}$$

(employing Lemma 4.55, one can easily see that (4.295) is implied by (4.283)). This would require the development of a sequential counterpart of the primal error bound established in Proposition 4.60. We omit the details.

Remark 4.64. By Remark 2.5, similarly to Remark 4.20, we note that in each of the conditions (4.277)–(4.279), (4.281), (4.293), (4.294) in Propositions 4.59 and 4.61, the right-hand side can be replaced by either $o(\|x^k - \bar{x}\|)$ or $o(\|x^{k+1} - x^k\|)$.

The analysis of primal superlinear convergence developed above for the general perturbed semismooth SQP framework (4.276) can be applied to some more specific algorithms. In particular, Algorithm 4.52 can be viewed as a special case of this framework with

$$\omega_1^k = (H_k - W_k)(x^{k+1} - x^k), \quad \omega_2^k = 0, \quad \omega_3^k = 0,$$

where $\{W_k\}$ is a sequence of matrices such that $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)$ for each k. From Propositions 4.59, 4.61 and Remarks 4.62, 4.63 it follows that under (4.295) the primal superlinear convergence of this *quasi-Newton semismooth SQP method* is characterized by the condition

$$\pi_{C(\bar{x})}\left(\frac{\partial L}{\partial x}(x^{k+1}, \lambda^k, \mu^k) - \frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k) - H_k(x^{k+1} - x^k)\right)$$
$$= o(\|x^{k+1} - x^k\| \tag{4.296}$$

as $k \to \infty$. This can be regarded as a natural generalization of the Dennis–Moré condition (4.102) for smooth quasi-Newton SQP methods to the case of semismooth first derivatives.

Taking into account the customary Dennis–Moré condition (4.102), one might think of replacing (4.296) by something like

$$\max_{W \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)} \|\pi_{C(\bar{x})}((W - H_k)(x^{k+1} - x^k))\| = o(\|x^{k+1} - x^k\|). \tag{4.297}$$

In particular, this condition corresponds to the one used for similar purposes in [154], where it is shown that (4.297) is necessary and, under (4.295), sufficient for the primal superlinear convergence of Algorithm 4.52 in the case when there are no inequality constraints (thus, extending Theorem 4.6 to the case of semismooth first derivatives). If f, h, and g are twice continuously differentiable near $\bar{x}$, then by Theorem A.10 one can easily see that the conditions (4.296) and (4.297) are equivalent. In the semismooth case the relationship between these two conditions is not so clear. From Proposition 4.59 it easily follows that (4.297) is necessary for the primal superlinear convergence. Therefore, it is implied by (4.296) if (4.295) holds. The converse implication might not be true, but according to the discussion below, it appears difficult to give an example of the lack of this implication. Namely, we shall show that, under a certain reasonable additional assumption, (4.297) is sufficient for the primal superlinear convergence and thus implies (4.296).

Specifically, assume that the set

$$A_k = \{i = 1,\, \ldots,\, m \mid g_i(x^k) + g_i'(x^k)(x^{k+1} - x^k) = 0\} \tag{4.298}$$

of indices of active inequality constraints of the semismooth SQP subproblems (4.255) stabilizes, i.e., it holds that $A_k = A$ for some fixed $A \subset \{1,\, \ldots,\, m\}$ and all k large enough. According to the last line in (4.260), by basic continuity considerations, the inclusions

$$A_+(\bar{x},\, \bar{\mu}) \subset A_k \subset A(\bar{x}) \tag{4.299}$$

always hold for all k large enough. Therefore, as discussed in Sect. 4.2, the stabilization property is automatic with $A = A(\bar{x})$ when $\{(\lambda^k,\, \mu^k)\}$ converges to a multiplier $(\bar{\lambda},\, \bar{\mu})$ satisfying the strict complementarity condition, i.e., such that $\bar{\mu}_{A(\bar{x})} > 0$ (and hence, $A_+(\bar{x}, \bar{\mu}) = A(\bar{x})$). In other cases, the stabilization property may not hold, but as discussed in Sect. 4.2, this still seems to be reasonable numerical behavior, which should be quite typical. Note also that if this stabilization property does not hold, one should hardly expect convergence of the dual trajectory, in general.

The following result extends the sufficiency part of [154, Theorem 2.3] to the case when inequality constraints can be present.

Theorem 4.65. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their derivatives being semismooth at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (4.253), and let $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ be an associated Lagrange multiplier. Let a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ generated by Algorithm 4.52 be convergent to $(\bar{x}, \bar{\lambda}, \bar{\mu})$. Assume that (4.295) and (4.297) hold and that there exists an index set $A \subset \{1, \ldots, m\}$ such that $A_k = A$ for all k large enough, where the index sets A_k are defined according to (4.298).*

Then the rate of convergence of $\{x^k\}$ is superlinear.

Proof. Define the set

$$\tilde{C}(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_A(\bar{x})\xi = 0,\ g'_{A(\bar{x})\setminus A}(\bar{x})\xi \leq 0\}.$$

By Hoffman's error bound for linear systems (see Lemma A.4) we have that

$$\begin{aligned} \operatorname{dist}(x^{k+1} - \bar{x}, \tilde{C}(\bar{x})) = O(&\|h'(\bar{x})(x^{k+1} - \bar{x})\| + \|g'_A(\bar{x})(x^{k+1} - \bar{x})\| \\ &+\| \max\{0, g'_{A(\bar{x})\setminus A}(\bar{x})(x^{k+1} - \bar{x})\}\|) \end{aligned} \tag{4.300}$$

as $k \to \infty$.

From the second line in (4.260), and from the local Lipschitz-continuity of the derivative of h at $\bar{x}$, we obtain that

$$\begin{aligned} h'(\bar{x})(x^{k+1} - \bar{x}) &= h'(\bar{x})(x^{k+1} - x^k) + h'(\bar{x})(x^k - \bar{x}) \\ &\quad -h(x^k) - h'(x^k)(x^{k+1} - x^k) \\ &= -(h'(x^k) - h'(\bar{x}))(x^{k+1} - x^k) \\ &\quad -(h(x^k) - h(\bar{x}) - h'(\bar{x})(x^k - \bar{x})) \\ &= o(\|x^k - \bar{x}\|) \end{aligned} \tag{4.301}$$

as $k \to \infty$. According to the definition of A, for any sufficiently large k it holds that

$$g_A(x^k) + g'_A(x^k)(x^{k+1} - x^k) = 0,$$

and similarly to (4.301) it follows that

$$g'_A(\bar{x})(x^{k+1} - \bar{x}) = o(\|x^k - \bar{x}\|) \tag{4.302}$$

as $k \to \infty$. Finally, if $i \in A(\bar{x}) \setminus A$ and $\langle g'_i(\bar{x}), x^{k+1} - \bar{x}\rangle > 0$, taking into account the last line of (4.260) and the local Lipschitz-continuity of the derivative of g at $\bar{x}$, we obtain that

$$
\begin{aligned}
\max\{0,\, \langle g_i'(\bar{x}),\, x^{k+1} - \bar{x}\rangle\} &= \langle g_i'(\bar{x}),\, x^{k+1} - \bar{x}\rangle \\
&= \langle g_i'(\bar{x}),\, x^{k+1} - x^k\rangle + \langle g_i'(\bar{x}),\, x^k - \bar{x}\rangle \\
&\leq \langle g_i'(\bar{x}),\, x^{k+1} - x^k\rangle + \langle g_i'(\bar{x}),\, x^k - \bar{x}\rangle \\
&\quad -g_i(x^k) - \langle g_i'(x^k),\, x^{k+1} - x^k\rangle \\
&= -\langle g_i'(x^k) - g_i'(\bar{x}),\, x^{k+1} - x^k\rangle \\
&\quad -(g_i(x^k) - g_i(\bar{x}) - \langle g_i'(\bar{x}),\, x^k - \bar{x}\rangle) \\
&= o(\|x^k - \bar{x}\|) \qquad (4.303)
\end{aligned}
$$

as $k \to \infty$. Relations (4.300)–(4.303) imply that

$$
\operatorname{dist}(x^{k+1} - \bar{x},\, \tilde{C}(\bar{x})) = o(\|x^k - \bar{x}\|)
$$

as $k \to \infty$. The latter means that for each k there exists $\xi^k \in \tilde{C}(\bar{x})$, such that

$$
x^{k+1} - \bar{x} = \xi^k + o(\|x^k - \bar{x}\|). \qquad (4.304)
$$

From the first line of (4.260) and from the semismoothness of the derivatives of f, h, and g at the point $\bar{x}$, employing Lemma 4.56 and the convergence of $\{(\lambda^k,\, \mu^k)\}$ to $(\bar{\lambda},\, \bar{\mu})$ we derive that for any choice of matrices $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)$ it holds that

$$
\begin{aligned}
-H_k(x^{k+1} - x^k) &= \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) \\
&\quad +(h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) \\
&= \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) - \frac{\partial L}{\partial x}(\bar{x},\, \lambda^k,\, \mu^k) - W_k(x^k - \bar{x}) \\
&\quad +\frac{\partial L}{\partial x}(\bar{x},\, \lambda^k,\, \mu^k) - \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) + (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) \\
&\quad +(g'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) + W_k(x^k - \bar{x}) \\
&= W_k(x^k - \bar{x}) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^k - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^k - \bar{\mu}) \\
&\quad +(h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \mu^k) \\
&\quad +o(\|x^k - \bar{x}\|) \\
&= W_k(x^k - \bar{x}) + (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}) \\
&\quad +o(\|x^k - \bar{x}\|)
\end{aligned}
$$

as $k \to \infty$. Therefore,

$$
\begin{aligned}
W_k(x^{k+1} - \bar{x}) &= (W_k - H_k)(x^{k+1} - x^k) - (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) \\
&\quad -(g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}) + o(\|x^k - \bar{x}\|). \qquad (4.305)
\end{aligned}
$$

From the definition of A it follows that

$$g_{\{1,\,\ldots,\,m\}\setminus A}(x^k) + g'_{\{1,\,\ldots,\,m\}\setminus A}(x^k)(x^{k+1} - x^k) < 0$$

for all k large enough. Then, by the last line in (4.260),

$$\mu^{k+1}_{\{1,\,\ldots,\,m\}\setminus A} = 0 \tag{4.306}$$

for all such k. Moreover, according to (4.299) it holds that $\tilde{C}(\bar{x}) \subset C(\bar{x})$, and therefore, (4.297) remains true with $C(\bar{x})$ substituted for $\tilde{C}(\bar{x})$. Then, employing (4.305), (4.306) and the fact that $\langle x,\, \xi\rangle \leq \langle \pi_{\tilde{C}(\bar{x})}(x),\, \xi\rangle$ for all $x \in \mathbf{R}^n$ and all $\xi \in \tilde{C}(\bar{x})$ (see Lemma A.13), we further obtain that

$$\begin{aligned}
\langle W_k\xi^k,\, \xi^k\rangle &= \langle W_k(x^{k+1} - \bar{x}),\, \xi^k\rangle + o(\|x^k - \bar{x}\|\|\xi^k\|)\\
&= \langle (W_k - H_k)(x^{k+1} - x^k),\, \xi^k\rangle\\
&\quad - \langle (h'(\bar{x}))^{\mathrm{T}}(\lambda^{k+1} - \bar{\lambda}) + (g'(\bar{x}))^{\mathrm{T}}(\mu^{k+1} - \bar{\mu}),\, \xi^k\rangle\\
&\quad + o(\|x^k - \bar{x}\|\|\xi^k\|)\\
&\leq \langle \pi_{\tilde{C}(\bar{x})}((W_k - H_k)(x^{k+1} - x^k)),\, \xi^k\rangle + o(\|x^k - \bar{x}\|\|\xi^k\|)\\
&= o((\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)\|\xi^k\|)
\end{aligned} \tag{4.307}$$

as $k \to \infty$.

From (4.295) and from the inclusion $\tilde{C}(\bar{x}) \subset C(\bar{x})$, it further follows that there exist $\gamma > 0$ and a sequence $\{W_k\}$ of matrices such that the inclusion $W_k \in \partial_x \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k)$ holds for all k, and

$$\langle W_k\xi^k,\, \xi^k\rangle \geq \gamma\|\xi^k\|^2$$

for all k large enough. Then (4.307) implies

$$\xi^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|),$$

and hence, given (4.304),

$$x^{k+1} - \bar{x} = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)$$

as $k \to \infty$. Repeating the argument completing the proof of Proposition 2.4, the latter implies the superlinear convergence rate of $\{x^k\}$. $\square$

Chapter 5
Variational Problems: Globalization of Convergence

Newtonian methods for variational problems discussed in Chap. 3 have attractive local convergence and rate of convergence properties, but they are local by nature: for guaranteed convergence, they all require a starting point close enough to a solution. Therefore, to arrive to practical algorithms based on these methods, the process of computing and automatically accepting an appropriate "starting point" must be incorporated into the overall iterative scheme. Such extensions of locally convergent methods are referred to as globalization of convergence. This chapter discusses some strategies for globalization of convergence of Newtonian methods for variational problems.

5.1 Linesearch Methods

As seen in Sect. 2.2, descent methods for unconstrained optimization possess natural global convergence properties. Moreover, linesearch quasi-Newton methods combine global convergence of descent methods with high convergence rate of the Newton method. It is then natural to extend this idea to other problem settings. This requires relating the problem in question to unconstrained optimization, in some meaningful way. This task is usually approached by constructing a function measuring the quality of approximations to a solution of the original problem, called a merit function. Then, given a direction produced by a Newtonian method, we can perform linesearch in this direction, evaluating the candidate points using the chosen merit function. In the case of unconstrained optimization, the natural merit function is the objective function of the problem. For constrained optimization and variational problems, the choice of a merit function is not evident and is certainly not unique. Some possibilities will be discussed in this and the next chapters.

Essentially, the idea of linesearch globalization is to reduce the size of the Newtonian step when the full step does not provide a sufficient decrease for

A.F. Izmailov and M.V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-3-319-04247-3_5,

values of the chosen merit function. If the resulting algorithm turns out to be a descent method for the merit function, one can expect global convergence (in some sense). If, in addition, the stepsize is reduced only far from solutions (solutions satisfying certain properties, of course), this would imply that the algorithm asymptotically turns into the full-step Newtonian method possessing high convergence rate. For linesearch quasi-Newton methods for unconstrained optimization, this ideal combination of convergence properties is achieved in Theorems 2.24–2.26. However, as naturally expected and seen below, for more general and different problem settings the situation is more complex.

5.1.1 Globalized Newton Method for Equations

We start with considering linesearch methods for the usual equation

$$\Phi(x) = 0, \tag{5.1}$$

where $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ is a smooth mapping. As discussed in Sect. 2.2.3, the natural choice of a merit function $\varphi : \mathbf{R}^n \to \mathbf{R}$ for (5.1) is the squared Euclidean residual

$$\varphi(x) = \frac{1}{2}\|\Phi(x)\|^2. \tag{5.2}$$

Moreover, for any $x \in \mathbf{R}^n$ which is not a solution of (5.1), if $\Phi'(x)$ is a nonsingular matrix, then the well-defined Newtonian direction

$$p = -(\Phi'(x))^{-1}\Phi(x) \tag{5.3}$$

is a direction of descent for this merit function at x, i.e., $p \in \mathcal{D}_{\varphi}(x)$. In Sect. 2.2.3 also some perturbed counterparts of the Newtonian direction p are discussed, which are relevant in the context of linesearch methods.

Here we only consider the basic choice (5.3). One special feature of this direction is that its quality as a descent direction can be readily estimated via the residual of (5.1):

$$\langle \varphi'(x),\, p\rangle = \langle (\Phi'(x))^{\mathrm{T}}\Phi(x),\, p\rangle = \langle \Phi(x),\, \Phi'(x)p\rangle = -\|\Phi(x)\|^2. \tag{5.4}$$

In particular, $\langle \varphi'(x),\, p\rangle$ may become close to zero (so that p is not a "good" descent direction) only when $\Phi(x)$ is close to zero. One possible way to deal with the latter situation is to employ the Levenberg–Marquardt regularization, blending the Newton direction with the steepest descent direction. However, the drawback of using the Levenberg–Marquardt directions in linesearch methods is the difficulty in choosing the regularization parameters; see Sect. 2.2.3.

Alternatively, the following algorithm implementing the Newton method equipped with the Armijo linesearch rule has the option of resorting directly to the steepest descent step as a safeguard, but only in those cases when the Newtonian direction does not exist or is too large.

Algorithm 5.1 Choose parameters $C > 0$, $\tau > 0$, $\sigma \in (0, 1)$ and $\theta \in (0, 1)$. Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $\Phi(x^k) = 0$, stop.
2. Compute $p^k \in \mathbf{R}^n$ as a solution of the linear equation
$$\Phi(x^k) + \Phi'(x^k)p = 0. \tag{5.5}$$
If such p^k exists and
$$\|p^k\| \leq \max\{C,\, 1/\|\Phi(x^k)\|^\tau\}, \tag{5.6}$$
go to step 4.
3. Set $p^k = -\varphi'(x^k) = -(\Phi'(x^k))^{\mathrm{T}}\Phi(x^k)$, with $\varphi : \mathbf{R}^n \to \mathbf{R}$ defined by (5.2). If $p^k = 0$, stop.
4. Set $\alpha = 1$. If the inequality
$$\varphi(x^k + \alpha p^k) \leq \varphi(x^k) + \sigma\alpha\langle\varphi'(x^k),\, p^k\rangle \tag{5.7}$$
is satisfied, set $\alpha_k = \alpha$. Otherwise, replace α by $\theta\alpha$, check the inequality (5.7) again, etc., until (5.7) becomes valid.
5. Set $x^{k+1} = x^k + \alpha_k p^k$.
6. Increase k by 1 and go to step 1.

Remark 5.2. According to (5.2) and (5.4), the condition (5.7) can be written as follows:

$$\frac{1}{2}\|\Phi(x^k + \alpha p^k)\|^2 \leq \frac{1}{2}\|\Phi(x^k)\|^2 - \sigma\alpha\|\Phi(x^k)\|^2 = \frac{1 - 2\sigma\alpha}{2}\|\Phi(x^k)\|^2,$$

or equivalently,

$$\|\Phi(x^k + \alpha p^k)\| \leq \sqrt{1 - 2\sigma\alpha}\|\Phi(x^k)\| = (1 - \sigma\alpha + o(\alpha))\|\Phi(x^k)\|$$

as $\alpha \to 0$. Therefore, in Algorithm 5.1, the condition (5.7) can be replaced by

$$\|\Phi(x^k + \alpha p^k)\| \leq (1 - \sigma\alpha)\|\Phi(x^k)\|. \tag{5.8}$$

Observe that, by the differentiability of Φ at x^k, and by (5.3),

$$\Phi(x^k + \alpha p^k) = \Phi(x^k) + \alpha\Phi'(x^k)p^k + o(\alpha) = (1 - \alpha)\Phi(x^k) + o(\alpha)$$

as $\alpha \to 0$, and hence, (5.8) is satisfied for all $\alpha > 0$ small enough.

Our global convergence result for Algorithm 5.1 is along the lines of Theorem 2.25.

Theorem 5.3. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be continuously differentiable on $\mathbf{R}^n$.*

Then for any starting point $x^0 \in \mathbf{R}^n$ Algorithm 5.1 generates the iterative sequence $\{x^k\}$ such that each of its accumulation points $\bar{x} \in \mathbf{R}^n$ satisfies

$$(\Phi'(\bar{x}))^{\mathrm{T}}\Phi(\bar{x}) = 0. \tag{5.9}$$

Proof. The fact that Algorithm 5.1 is well defined follows from Lemma 2.19. Indeed, for each k the corresponding direction p^k either satisfies (5.4) or equals $-\varphi'(x^k)$. Therefore, if $\varphi'(x^k) = (\Phi'(x^k))^{\mathrm{T}}\Phi(x^k) \neq 0$, then in either case

$$\langle \varphi'(x^k),\, p^k\rangle < 0. \tag{5.10}$$

Moreover, assuming that $\varphi'(x^k) \neq 0$ for all k, the sequence $\{\varphi(x^k)\}$ is monotonically decreasing. Since this sequence is bounded below (by zero), it converges, and hence, (5.7) implies the equality

$$\lim_{k\to\infty} \alpha_k \langle \varphi'(x^k),\, p^k\rangle = 0. \tag{5.11}$$

Let $\bar{x}$ be an accumulation point of the sequence $\{x^k\}$, and let $\{x^{k_j}\}$ be a subsequence convergent to $\bar{x}$ as $j \to \infty$. Consider the two possible cases:

$$\limsup_{j\to\infty} \alpha_{k_j} > 0 \quad \text{or} \quad \lim_{j\to\infty} \alpha_{k_j} = 0. \tag{5.12}$$

In the first case, passing onto a further subsequence if necessary, we can assume that the entire $\{\alpha_{k_j}\}$ is separated away from zero:

$$\liminf_{j\to\infty} \alpha_{k_j} > 0.$$

Then (5.11) implies that

$$\lim_{j\to\infty} \langle \varphi'(x^{k_j}),\, p^{k_j}\rangle = 0. \tag{5.13}$$

If p^{k_j} is defined by the Newton iteration system (5.5) for infinitely many indices j, by (5.4) we have that

$$\langle \varphi'(x^{k_j}),\, p^{k_j}\rangle = -\|\Phi(x^{k_j})\|^2$$

for these j, and then (5.13) implies that $\Phi(\bar{x}) = 0$, which certainly implies (5.9). On the other hand, if Newton directions are used only for finitely many indices j, then

$$\langle \varphi'(x^{k_j}),\, p^{k_j}\rangle = -\langle \varphi'(x^{k_j}),\, \varphi'(x^{k_j})\rangle = -\|\varphi'(x^{k_j})\|^2$$

for all j large enough. Hence, by (5.13), $(\Phi'(\bar{x}))^{\mathrm{T}}\Phi(\bar{x}) = \varphi'(\bar{x}) = 0$, i.e., (5.9) holds in this case as well.

It remains to consider the second case in (5.12). Suppose first that the sequence $\{p^{k_j}\}$ is unbounded. Note that this can only happen when the Newton directions are used infinitely often, because otherwise $p^{k_j} = -\varphi'(x^{k_j})$ for all j large enough, and hence, $\{p^{k_j}\}$ converges to $-\varphi'(\bar{x})$. But then the condition (5.6) implies that

$$\liminf_{j\to\infty} \|\Phi(x^{k_j})\| = 0,$$

so that $\Phi(\bar{x}) = 0$, and hence, (5.9) is again valid.

Let finally $\{p^{k_j}\}$ be bounded. Taking a further subsequence, if necessary, assume that $\{p^{k_j}\}$ converges to some $\tilde{p}$. Since in the second case in (5.12) for each j large enough the initial stepsize value had been reduced at least once, the value $\alpha_{k_j}/\theta > \alpha_{k_j}$ does not satisfy (5.7), i.e.,

$$\frac{\varphi(x^{k_j} + \alpha_{k_j} p^{k_j}/\theta) - \varphi(x^{k_j})}{\alpha_{k_j}/\theta} > \sigma\langle\varphi'(x^{k_j}),\, p^{k_j}\rangle.$$

Employing the mean-value theorem (Theorem A.10, (a)) and the fact that $\alpha_{k_j} \to 0$ as $j \to \infty$, and passing onto the limit as $j \to \infty$, we obtain that

$$\langle\varphi'(\bar{x}),\, \tilde{p}\rangle \geq \sigma\langle\varphi'(\bar{x}),\, \tilde{p}\rangle,$$

which may only hold when $\langle\varphi'(\bar{x}),\, \tilde{p}\rangle \geq 0$. Combining this with (5.10), we obtain that

$$\langle\varphi'(\bar{x}),\, \tilde{p}\rangle = 0.$$

Considering, as above, the two cases when the number of times the Newton direction had been used is infinite or finite, the latter relation implies that (5.9) holds. $\square$

According to the proof of Theorem 5.3, if along a subsequence convergent to $\bar{x}$ the Newton direction had been used infinitely many times, then $\bar{x}$ is a solution of (5.1). Convergence to a point $\bar{x}$ satisfying (5.9) which is not a solution of (5.1) can only happen when the Newton directions are not used along the corresponding subsequence from some point on at all, and when in addition the Jacobian $\Phi'(\bar{x})$ is singular.

Another important issue is the existence of accumulation points of iterative sequences generated by Algorithm 5.1. This is guaranteed when the residual $\|\Phi(\cdot)\|$ is coercive.

Finally, we show that Algorithm 5.1 preserves fast local convergence of the basic Newton method under natural assumptions.

Theorem 5.4. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be continuously differentiable on $\mathbf{R}^n$. Let a sequence $\{x^k\} \subset \mathbf{R}^n$ be generated by Algorithm 5.1 with $\sigma \in (0,\, 1/2)$, and assume that this sequence has an accumulation point $\bar{x}$ such that $\Phi'(\bar{x})$ is nonsingular.*

Then the entire sequence $\{x^k\}$ converges to $\bar{x}$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$.

Proof. If $x^k \in \mathbf{R}^n$ is close enough to $\bar{x}$, then, according to Theorem 2.2, there exists the unique $p^k \in \mathbf{R}^n$ satisfying (5.5), and

$$x^k + p^k - \bar{x} = o(\|x^k - \bar{x}\|) \tag{5.14}$$

as $x^k \to \bar{x}$. As a consequence, p^k would be accepted by the test (5.6).

Furthermore, Proposition 1.32 implies the estimate

$$x^k - \bar{x} = O(\|\Phi(x^k)\|)$$

as $x^k \to \bar{x}$. Employing this estimate and (5.2), (5.14), and also taking into account the local Lipschitz-continuity of Φ with respect to $\bar{x}$ (following from the differentiability of Φ at $\bar{x}$), we obtain that

$$\begin{aligned}\varphi(x^k + p^k) &= \frac{1}{2}\|\Phi(x^k + p^k) - \Phi(\bar{x})\|^2\\ &= O(\|x^k + p^k - \bar{x}\|^2)\\ &= o(\|x^k - \bar{x}\|^2)\\ &= o(\|\Phi(x^k)\|^2)\end{aligned}$$

as $x^k \to \bar{x}$. The above relation implies that if x^k is close enough to $\bar{x}$, then

$$\begin{aligned}\varphi(x^k + p^k) &\leq \frac{1 - 2\sigma}{2}\|\Phi(x^k)\|^2\\ &= \varphi(x^k) - \sigma\|\Phi(x^k)\|^2\\ &= \varphi(x^k) + \sigma\langle\varphi'(x^k),\, p^k\rangle,\end{aligned}$$

where the last equality is by (5.4) (recall also that $\sigma \in (0,\, 1/2)$). Therefore, $\alpha_k = 1$ is accepted by step 4 of the algorithm: inequality (5.7) holds with $\alpha = 1$. This shows that the iteration of Algorithm 5.1 reduces to that of Algorithm 2.1. The assertions now follow from Theorem 2.2. □

5.1.2 Globalized Semismooth Newton Methods for Complementarity Problems

Consider now the nonlinear complementarity problem (NCP)

$$x \geq 0, \quad \Phi(x) \geq 0, \quad \langle x,\, \Phi(x)\rangle = 0, \tag{5.15}$$

where $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ is a smooth mapping.

Define the mapping $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ by

$$\Psi(x) = \psi(x,\, \Phi(x)), \tag{5.16}$$

where $\psi : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ is the Fischer–Burmeister complementarity function

$$\psi(a,\, b) = a + b - \sqrt{a^2 + b^2}, \tag{5.17}$$

applied componentwise. As discussed in Sect. 3.2, this mapping is semismooth, and problem (5.15) is equivalent to the equation

$$\Psi(x) = 0,$$

which can be locally solved by the semismooth Newton method. For a given $x \in \mathbf{R}^n$, this method generates the direction

$$p = -J^{-1}\Psi(x), \tag{5.18}$$

assuming $J \in \partial\Psi(x)$ is nonsingular.

To perform linesearch along the Newton direction (5.18), define the merit function $\varphi : \mathbf{R}^n \to \mathbf{R}$,

$$\varphi(x) = \frac{1}{2}\|\Psi(x)\|^2. \tag{5.19}$$

It turns out that this merit function φ is continuously differentiable, even though Ψ itself is not. Moreover, the gradient of φ is explicitly computable using any element of the generalized Jacobian of Φ. These two properties are the key for linesearch globalization of this version of the semismooth Newton method for NCP (5.15); they were established in [69].

Proposition 5.5. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable at a point $x \in \mathbf{R}^n$.*

Then the function $\varphi : \mathbf{R}^n \to \mathbf{R}$ defined by (5.16), (5.17), (5.19) is differentiable at x, and it holds that

$$\varphi'(x) = \sum_{i \in I(x)} \Psi_i(x) \left(\Phi_i'(x) + e^i - \frac{\Phi_i(x)\Phi_i'(x) + x_i e^i}{\sqrt{x_i^2 + (\Phi_i(x))^2}} \right), \tag{5.20}$$

where

$$I(x) = \{i = 1,\, \ldots,\, n \mid x_i \neq 0 \text{ or } \Phi_i(x) \neq 0\}.$$

In particular, if Φ is differentiable near x with its derivative being continuous at x, then

$$\varphi'(x) = J^{\mathrm{T}}\Phi(x) \quad \forall\, J \in \partial\Phi(x). \tag{5.21}$$

Moreover, if Φ is continuously differentiable on $\mathbf{R}^n$, then the function φ is continuously differentiable on $\mathbf{R}^n$.

Proof. According to (5.16), (5.17), for each $i \in I(x)$ the function Ψ_i is differentiable at x, and

$$\Psi_i'(x) = \Phi_i'(x) + e^i - \frac{\Phi_i(x)\Phi_i'(x) + x_i e^i}{\sqrt{x_i^2 + (\Phi_i(x))^2}}. \tag{5.22}$$

On the other hand, if $i \in \{1, \ldots, n\} \setminus I(x)$, then $x_i = \Phi_i(x) = 0$, and taking into account the local Lipschitz-continuity of Φ with respect to x (following from the differentiability of Φ at x), we obtain that

$$\begin{aligned}\Psi_i(x+\xi) &= \psi(x_i + \xi_i,\, \Phi_i(x+\xi)) \\ &= \xi_i + \Phi_i(x+\xi) - \Phi_i(x) - \|(\xi_i,\, \Phi_i(x+\xi) - \Phi_i(x))\| \\ &= O(\|\xi\|)\end{aligned}$$

as $\xi \to 0$, and $\Psi_i(x) = 0$. Therefore,

$$(\Psi_i(x+\xi))^2 - (\Psi_i(x))^2 = (\Psi_i(x+\xi))^2 = O(\|\xi\|^2),$$

implying that $(\Psi_i(\cdot))^2$ is differentiable at x, and its gradient at x is equal to zero. Combining these facts with (5.19), we conclude that φ is differentiable at x, and its gradient is given by (5.20), which further implies that

$$\varphi'(x) = J^{\mathrm{T}}\Psi(x),$$

where $J \in \mathbf{R}^{n\times n}$ is any matrix satisfying the equalities $J_i = \Psi_i'(x)$ for all indices $i \in I(x)$ (other rows of J can be arbitrary, since the corresponding components of $\Psi(x)$ are equal to zero). Employing (5.22) and Proposition 3.11, we then obtain (5.21).

Finally, if Φ is continuously differentiable on $\mathbf{R}^n$, then, according to Proposition 3.8, Ψ is semismooth, and hence, locally Lipschitz-continuous on $\mathbf{R}^n$. In this case, continuity of φ' follows from (5.21) and Proposition 1.51. □

For the direction p defined by (5.18) with some $J \in \partial\Psi(x)$, from (5.21) we derive that

$$\langle \varphi'(x),\, p\rangle = \langle J^{\mathrm{T}}\Psi(x),\, p\rangle = \langle \Psi(x),\, Jp\rangle = -\|\Psi(x)\|^2,$$

similarly to (5.4). In particular, if x is not a solution of the NCP (5.15) (i.e., if $\Psi(x) \neq 0$), then p is a direction of descent for φ at x, i.e., $p \in \mathcal{D}_\varphi(x)$.

The following is a version of the semismooth Newton method for the NCP (5.15), equipped with the Armijo linesearch rule. It is designed along the lines of Algorithm 5.1. This specific hybrid construction was first suggested in [156], but the related developments date back to [52].

Algorithm 5.6 Define the mapping $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ according to (5.16), (5.17). Choose the parameters $C > 0$, $\tau > 0$, $\sigma \in (0, 1)$ and $\theta \in (0, 1)$. Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $\Psi(x^k) = 0$, stop.
2. Compute some $J_k \in \partial_B \Psi(x^k)$. Compute $p^k \in \mathbf{R}^n$ as a solution of the linear system
$$\Psi(x^k) + J_k p = 0. \tag{5.23}$$
If such p^k exists and satisfies
$$\|p^k\| \leq \max\{C,\, 1/\|\Psi(x^k)\|^\tau\},$$
go to step 4.
3. Set
$$p^k = -\varphi'(x^k) = -J_k^{\mathrm{T}} \Psi(x^k),$$
with $\varphi : \mathbf{R}^n \to \mathbf{R}$ defined by (5.19). If $p^k = 0$, stop.
4. Set $\alpha = 1$. If the inequality (5.7) is satisfied, set $\alpha_k = \alpha$. Otherwise, replace α by $\theta\alpha$, check the inequality (5.7) again, etc., until (5.7) becomes valid.
5. Set $x^{k+1} = x^k + \alpha_k p^k$.
6. Increase k by 1 and go to step 1.

The global convergence properties of this algorithm are given by the following statement.

Theorem 5.7. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be continuously differentiable on $\mathbf{R}^n$.*

Then for any starting point $x^0 \in \mathbf{R}^n$ Algorithm 5.6 generates an iterative sequence $\{x^k\}$ such that each of its accumulation points $\bar{x} \in \mathbf{R}^n$ satisfies
$$J^{\mathrm{T}} \Psi(\bar{x}) = 0 \quad \forall\, J \in \partial \Psi(\bar{x}), \tag{5.24}$$
where $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ is defined according to (5.16), (5.17).

This result can be established by the same argument as in the proof of Theorem 5.3 with some evident modifications (in particular, Algorithm 5.1 substituted by Algorithm 5.6, and employing Proposition 5.5). In particular, $\bar{x}$ in Theorem 5.7 is a solution of the NCP (5.15) provided semismooth Newton directions are used infinitely many times along the subsequence convergent to $\bar{x}$. Observe also that any point $\bar{x}$ satisfying (5.24) is necessarily a solution of the NCP (5.15) if $\partial\Psi(\bar{x})$ contains at least one nonsingular matrix. Moreover, this nonsingularity condition can be further weakened as follows: for every $d \in \mathbf{R}^n$ there exist $J \in \partial\Psi(\bar{x})$ and $\xi \in \mathbf{R}^n$ such that $J\xi = d$ (i.e., there is no need for a single matrix $J \in \partial\Psi(\bar{x})$ to have this property for all d). The latter is further implied by the assumption that the directional derivative mapping $\Psi'(\bar{x};\, \cdot)$ is surjective (see Proposition 1.63, (b)). Indeed, taking $d = \Psi(\bar{x})$, for the corresponding J and ξ, from (5.24) we derive
$$0 = \langle J^{\mathrm{T}} \Psi(\bar{x}),\, \xi \rangle = \langle \Psi(\bar{x}),\, J\xi \rangle = \|\Psi(\bar{x})\|^2.$$

Some further characterizations in terms of the mapping Φ (that is, in terms of the original problem data) of when (5.24) implies that $\bar{x}$ is a solution can be found in [68, Sect. 9.1.2]. One sufficient condition, first established in [69], is based on the following notion: $A \in \mathbf{R}^{n\times n}$ is referred to as a P_0-*matrix* if all the principal minors of this matrix are nonnegative. This property is equivalent to the following. For every $\xi \in \mathbf{R}^n \setminus \{0\}$, there exists an index $i \in \{1, \ldots, n\}$ such that $\xi_i \neq 0$ and $\xi_i(A\xi)_i \geq 0$; see [46, Theorem 3.4.2]. An evident sufficient condition for these equivalent properties is positive semidefiniteness of A, that is,

$$\langle A\xi, \xi\rangle \geq 0 \quad \forall\, \xi \in \mathbf{R}^n.$$

Proposition 5.8. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable near a point $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$, and assume that (5.24) holds for $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ defined according to (5.16), (5.17).*

If $\Phi'(\bar{x})$ is a P_0-matrix, then $\bar{x}$ is a solution of the NCP (5.15).

Proof. From (5.20), (5.21), and (5.24) we derive that

$$0 = \varphi'(\bar{x}) = (\Phi'(\bar{x}))^{\mathrm{T}}\eta + \xi, \tag{5.25}$$

where $\xi \in \mathbf{R}^n$ and $\eta \in \mathbf{R}^n$ are defined by

$$\xi_i = \Psi_i(\bar{x})\left(1 - \frac{\bar{x}_i}{\sqrt{\bar{x}_i^2 + (\Phi_i(\bar{x}))^2}}\right), \quad \eta_i = \Psi_i(\bar{x})\left(1 - \frac{\Phi_i(\bar{x})}{\sqrt{\bar{x}_i^2 + (\Phi_i(\bar{x}))^2}}\right), \\ i \in I(\bar{x}), \tag{5.26}$$

and

$$\xi_i = 0, \quad \eta_i = 0, \quad i \in \{1, \ldots, n\} \setminus I(\bar{x}). \tag{5.27}$$

Assuming that $\bar{x}$ is not a solution of the NCP (5.15), there exists $i \in I(\bar{x})$ such that $\Psi_i(\bar{x}) \neq 0$, and for any such i there are only three possibilities: either $\bar{x}_i \neq 0$, $\Phi_i(\bar{x}) \neq 0$, or $\bar{x}_i = 0$, $\Phi_i(\bar{x}) < 0$, or $\bar{x}_i < 0$, $\Phi_i(\bar{x}) = 0$. In each of the cases it can be readily seen from (5.26) that $\xi_i\eta_i > 0$, while at the same time (5.25) implies that $\xi_i((\Phi'(\bar{x}))^{\mathrm{T}}\eta)_i < 0$, and hence, $((\Phi'(\bar{x}))^{\mathrm{T}}\eta)_i\eta_i < 0$. On the other hand, $\Phi'(\bar{x})$ is a P_0-matrix if and only if $(\Phi'(\bar{x}))^{\mathrm{T}}$ is a P_0-matrix, and therefore, there must exist $i \in \{1, \ldots, n\}$ such that $\eta_i \neq 0$ (and hence, according to (5.26), (5.27), $i \in I(\bar{x})$ and $\Psi_i(\bar{x}) \neq 0$) and $((\Phi'(\bar{x}))^{\mathrm{T}}\eta)_i\eta_i \geq 0$. This contradiction establishes the assertion. □

Regarding the existence of accumulation points of iterative sequences generated by Algorithm 5.6, we refer to [68, Sect. 9.1.5], where some characterizations of coercivity of the residual $\|\Psi(\cdot)\|$ are given. As one example, we mention the following: if (5.15) is the linear complementarity problem, i.e., $\Phi(x) = Ax + b$ with some $A \in \mathbf{R}^{n\times n}$ and $b \in \mathbf{R}^n$, then $\|\Psi(\cdot)\|$ is coercive if A is nonsingular.

The next result on the superlinear convergence rate of Algorithm 5.6 can be obtained in a way similar to the proof of Theorem 5.4 with some evident modifications (employing Proposition 3.8, and substituting Theorem 2.2 by Theorem 2.42, Proposition 1.32 by Proposition 1.77, and Algorithms 2.1 and 5.1 by Algorithms 3.7 and 5.6, respectively).

Theorem 5.9. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be continuously differentiable on $\mathbf{R}^n$. Let a sequence $\{x^k\} \subset \mathbf{R}^n$ be generated by Algorithm 5.6 with $\sigma \in (0, 1/2)$, and assume that this sequence has an accumulation point $\bar{x}$ such that the mapping $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ defined according to (5.16)–(5.17) is BD-regular at $\bar{x}$.*

Then the entire sequence $\{x^k\}$ converges to $\bar{x}$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the derivative of Φ is locally Lipschitz-continuous with respect to $\bar{x}$.

5.1.3 Extension to Mixed Complementarity Problems

Employing the machinery of Sect. 3.2.2, Algorithm 5.6 and its analysis in the previous section can be more-or-less directly extended to mixed complementarity problems. Below we focus on the special features of this development concerned with the KKT system

$$\begin{gathered} F(x) + (h'(x))^{\mathrm{T}}\lambda + (g'(x))^{\mathrm{T}}\mu = 0, \quad h(x) = 0, \\ \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu, g(x) \rangle = 0, \end{gathered} \tag{5.28}$$

where $F : \mathbf{R}^n \to \mathbf{R}^n$, $h : \mathbf{R}^n \to \mathbf{R}^l$, and $g : \mathbf{R}^n \to \mathbf{R}^m$ are smooth mappings.

Define the mappings $G : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n$,

$$G(x, \lambda, \mu) = F(x) + (h'(x))^{\mathrm{T}}\lambda + (g'(x))^{\mathrm{T}}\mu, \tag{5.29}$$

and $\Psi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$,

$$\Psi(u) = (G(x, \lambda, \mu), h(x), \psi(\mu, -g(x))), \quad u = (x, \lambda, \mu), \tag{5.30}$$

where $\psi : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ is again the Fischer–Burmeister complementarity function given by (5.17). Define the merit function $\varphi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}$,

$$\varphi(u) = \frac{1}{2}\|\Psi(u)\|^2. \tag{5.31}$$

A linesearch version of Algorithm 3.23 and the related theory can be developed by almost literally repeating Proposition 5.5, Algorithm 5.6, and Theorems 5.7 and 5.9, for the newly defined mapping Ψ and function φ. In particular, each accumulation point $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ of any trajectory generated by the algorithm would satisfy the following property:

$$\varphi'(\bar{u}) = J^{\mathrm{T}}\Psi(\bar{u}) = 0 \quad \forall\, J \in \partial\Psi(\bar{u}). \tag{5.32}$$

We now briefly discuss some conditions ensuring that $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ satisfying (5.32) is necessarily a solution of the KKT system (5.28). These conditions, employing the specificity of KKT systems, were derived in [70]. The first result we state without proof, as it is rather technical (see [70]).

Proposition 5.10. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$, and let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x}$, with their second derivatives being continuous at $\bar{x}$. Assume that for some $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ at least one of the following conditions is satisfied:*

(i) It holds that

$$\operatorname{rank} \begin{pmatrix} h'(\bar{x}) \\ g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x}) \end{pmatrix} = l + |A_+(\bar{x}, \bar{\mu})|,$$

and

$$\left\langle \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall\, \xi \in C_+(\bar{x}, \bar{\mu}) \setminus \{0\},$$

where

$$C_+(\bar{x}, \bar{\mu}) = \left\{ \xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x})\xi = 0 \right\}.$$

(ii) It holds that

$$\operatorname{rank} \begin{pmatrix} h'(\bar{x}) \\ g'_{A^0_+(\bar{x}, \bar{\mu})}(\bar{x}) \end{pmatrix} = l + |A^0_+(\bar{x}, \bar{\mu})|,$$

and

$$\left\langle \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall\, \xi \in C^0_+(\bar{x}, \bar{\mu}) \setminus \{0\},$$

where

$$A^0_+(\bar{x}, \bar{\mu}) = \{ i \in A(\bar{x}) \mid \bar{\mu}_i \geq 0 \},$$

$$C^0_+(\bar{x}, \bar{\mu}) = \left\{ \xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A^0_+(\bar{x}, \bar{\mu})}(\bar{x})\xi = 0 \right\}.$$

Then for $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$, and for $\Psi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ defined according to (5.17), (5.29), (5.30), the set $\partial \Psi(\bar{u})$ contains a nonsingular matrix.

Proposition 5.11. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$, and let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x}$, with their second derivatives being continuous at $\bar{x}$. Assume that $\bar{u} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ with some $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ satisfies (5.32) with $\Psi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ defined according to (5.17), (5.29), (5.30). Assume that*

$$\left\langle \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle \geq 0 \quad \forall\, \xi \in \ker h'(\bar{x}), \tag{5.33}$$

$$\left\langle \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall\, \xi \in (\ker h'(\bar{x}) \cap \ker g'(\bar{x})) \setminus \{0\}, \tag{5.34}$$

and at least one of the following conditions is satisfied:

(i) *It holds that* $\operatorname{rank} h'(\bar{x}) = l$.
(ii) *h is affine, and there exists $\hat{x} \in \mathbf{R}^n$ such that $h(\hat{x}) = 0$.*

Then $\bar{u}$ is a solution of the KKT system (5.28).

Proof. By Proposition 3.27 and (5.30), we conclude that (5.32) implies the equalities

$$\begin{gathered}\frac{\partial G}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})G(\bar{x},\, \bar{\lambda},\, \bar{\mu}) + (h'(\bar{x}))^{\mathrm{T}} h(\bar{x}) - (g'(\bar{x}))^{\mathrm{T}} \operatorname{diag}(\delta_\alpha)\psi(\bar{\mu},\, -g(\bar{x})) = 0,\\ h'(\bar{x})G(\bar{x},\, \bar{\lambda},\, \bar{\mu}) = 0,\\ g'(\bar{x})G(\bar{x},\, \bar{\lambda},\, \bar{\mu}) + \operatorname{diag}(\delta_\beta)\psi(\bar{\mu},\, -g(\bar{x})) = 0,\end{gathered} \tag{5.35}$$

where

$$\begin{gathered}(\delta_\alpha)_i = 1 + \frac{g_i(\bar{x})}{\sqrt{\bar{\mu}_i^2 + (g_i(\bar{x}))^2}}, \quad (\delta_\beta)_i = 1 - \frac{\bar{\mu}_i}{\sqrt{\bar{\mu}_i^2 + (g_i(\bar{x}))^2}}\\ \forall\, i = 1,\, \ldots,\, m \ \text{ such that } \ \psi(\bar{\mu}_i,\, -g_i(\bar{x})) \neq 0,\end{gathered}$$

while the other components of $\delta_\alpha \in \mathbf{R}^m$ and $\delta_\beta \in \mathbf{R}^m$ can be taken arbitrarily; we take them equal to 1. Similarly to the proof of Proposition 5.8, observe further that $(\delta_\alpha)_i > 0$ and $(\delta_\beta)_i > 0$ for all $i = 1,\, \ldots,\, m$ such that $\psi(\bar{\mu}_i,\, -g_i(\bar{x})) \neq 0$. Therefore, $\delta_\alpha > 0$ and $\delta_\beta > 0$.

From the last equality in (5.35) we now derive that

$$\psi(\bar{\mu},\, -g(\bar{x})) = -(\operatorname{diag}(\delta_\beta))^{-1} g'(\bar{x})G(\bar{x},\, \bar{\lambda},\, \bar{\mu}). \tag{5.36}$$

Substituting this into the first equality in (5.35) yields

$$\begin{aligned}0 &= \frac{\partial G}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})G(\bar{x},\, \bar{\lambda},\, \bar{\mu}) + (h'(\bar{x}))^{\mathrm{T}} h(\bar{x})\\ &\quad + (g'(\bar{x}))^{\mathrm{T}} \operatorname{diag}(\delta_\alpha)(\operatorname{diag}(\delta_\beta))^{-1} g'(\bar{x})G(\bar{x},\, \bar{\lambda},\, \bar{\mu})\\ &= \left(\frac{\partial G}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) + (g'(\bar{x}))^{\mathrm{T}} \operatorname{diag}(\delta)g'(\bar{x})\right) G(\bar{x},\, \bar{\lambda},\, \bar{\mu}) + (h'(\bar{x}))^{\mathrm{T}} h(\bar{x}),\end{aligned}$$

where $\delta_i = (\delta_\alpha)_i/(\delta_\beta)_i > 0$ for all $i = 1,\, \ldots,\, m$. Multiplying both sides of this equality by $G(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, and employing the second equality in (5.35), we obtain that

$$\begin{aligned}0 &= \left\langle \left(\frac{\partial G}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) + (g'(\bar{x}))^{\mathrm{T}} \operatorname{diag}(\delta)g'(\bar{x})\right) G(\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, G(\bar{x},\, \bar{\lambda},\, \bar{\mu})\right\rangle\\ &\quad + \langle h(\bar{x}),\, h'(\bar{x})G(\bar{x},\, \bar{\lambda},\, \bar{\mu})\rangle\\ &= \left\langle \frac{\partial G}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})G(\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, G(\bar{x},\, \bar{\lambda},\, \bar{\mu})\right\rangle\\ &\quad + \langle \operatorname{diag}(\delta)g'(\bar{x})G(\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, g'(\bar{x})G(\bar{x},\, \bar{\lambda},\, \bar{\mu})\rangle.\end{aligned} \tag{5.37}$$

Since $\frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})$ is positive semidefinite on $\ker h'(\bar{x})$ (i.e., (5.33) holds), and the second equality in (5.35) means that $G(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \ker h'(\bar{x})$, from (5.37) and the inequality $\delta > 0$ it follows that $g'(\bar{x})G(\bar{x}, \bar{\lambda}, \bar{\mu}) = 0$. Therefore,

$$G(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \ker h'(\bar{x}) \cap \ker g'(\bar{x}),$$

and by (5.34) and (5.37) we conclude that $G(\bar{x}, \bar{\lambda}, \bar{\mu}) = 0$. Hence, by (5.36), $\psi(\bar{\mu}, -g(\bar{x})) = 0$.

By the first equality in (5.35) we then have that

$$(h'(\bar{x}))^{\mathrm{T}} h(\bar{x}) = 0. \tag{5.38}$$

Since condition (i) is equivalent to $\ker(h'(\bar{x}))^{\mathrm{T}} = \{0\}$, it follows that (5.38) can hold only when $h(\bar{x}) = 0$. On the other hand, if condition (ii) holds, then $h(x) = Ax + b$ with some $A \in \mathbf{R}^{l \times n}$ and $b \in \mathbf{R}^l$, and $A\hat{x} + b = 0$. Therefore, $b \in \operatorname{im} A$, and hence, $A\bar{x} + b \in \operatorname{im} A = (\ker A^{\mathrm{T}})^{\perp}$. At the same time, (5.38) takes the form $A^{\mathrm{T}}(A\bar{x} + b) = 0$, or equivalently, $A\bar{x} + b \in \ker A^{\mathrm{T}}$, and hence, $h(\bar{x}) = 0$. This completes the proof. $\square$

The most restrictive assumption of Proposition 5.11, even in the optimization context, appears to be the positive semidefiniteness of $\frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu})$ (on $\ker h'(\bar{x})$), especially because $\bar{\mu}$ does not need to be nonnegative. One possibility to tackle this difficulty is to impose the restriction $\mu \geq 0$ directly, and to maintain it along the iterations; see [72]. Introducing this nonnegativity restriction can also be helpful for overcoming another drawback of the merit function φ defined by (5.17), (5.31) in the context of KKT systems: due to "decoupled" primal and dual variables of a KKT system, the level sets of φ are unlikely to be bounded under any reasonable assumptions. We conclude this section by an example demonstrating this phenomenon.

Example 5.12. Let $n = m = 2$, $l = 1$, $F(x) = f'(x)$, $f(x) = (x_1^2 + x_2^2)/2$, $h(x) = x_1 + x_2 - 1$, $g(x) = (-x_1, -x_2)$. Thus, the underlying optimization problem is the quadratic programming problem with a strongly convex objective function, and with the standard simplex as a feasible set; it is difficult to imagine a problem with better regularity properties.

However, e.g., for $u = (0, \lambda, \mu)$ with any $\mu_1 = \mu_2 = \lambda \geq 0$, from (5.17), (5.29)–(5.31) we see that

$$\begin{aligned}\varphi(u) &= \frac{1}{2}\left((\lambda - \mu_1)^2 + (\lambda - \mu_2)^2 + 1 + \left(\mu_1 - \sqrt{\mu_1^2}\right)^2 + \left(\mu_2 - \sqrt{\mu_2^2}\right)^2\right) \\ &= \frac{1}{2},\end{aligned}$$

demonstrating that φ is not coercive.

5.2 Path-Search Methods

Linesearch globalization strategy discussed in Sect. 5.1 is simple and often successful. On the other hand, as discussed below, it is not always applicable. In the context of nonsmooth equations, an alternative path-search approach was developed in [228]. It is the basis for the PATH solver [60], a well-established tool for solving complementarity problems and variational inequalities with linear constraints.

5.2.1 Path-Search Framework for Equations

To motivate the path-search approach [228], recall first the linesearch methods in Sect. 5.1 for the equation

$$\Psi(x) = 0 \tag{5.39}$$

with a given $\Psi : \mathbf{R}^n \to \mathbf{R}^n$.

If a Newtonian direction $p^k \in \mathbf{R}^n$ computed at the current iterate $x^k \in \mathbf{R}^n$ is a direction of descent for the residual $\|\Psi(\cdot)\|$, we consider the "path" defined as the image $p_k([0, 1])$ of the mapping $p_k : [0, 1] \to \mathbf{R}^n$, $p_k(\alpha) = x^k + \alpha p^k$. In this case, the "path" is simply the line segment connecting $p_k(0) = x^k$ and $p_k(1) = x^k + p^k$. The latter point corresponds to the full Newtonian step, and we first try this point by checking if it provides a sufficient decrease for the residual, which can be quantitatively characterized by the Armijo-type condition

$$\|\Psi(p_k(\alpha))\| \leq (1 - \sigma\alpha)\|\Psi(x^k)\| \tag{5.40}$$

with a pre-fixed $\sigma \in (0, 1)$ (see Remark 5.2). If $\alpha = 1$ satisfies (5.40), we accept the full Newtonian step by taking $x^{k+1} = p_k(1)$. Otherwise we perform backtracking along the path $p_k([0, 1])$, i.e., we reduce the stepsize α until some $\alpha_k \in [0, 1]$ satisfying (5.40) is found; then, we define the next iterate as $x^{k+1} = p_k(\alpha_k)$.

As discussed in Sects. 2.2.3 and 5.1.1, if Ψ is differentiable at x^k, then the basic Newton direction

$$p^k = -(\Psi'(x^k))^{-1}\Psi(x^k) \tag{5.41}$$

is a direction of descent for the residual $\|\Psi(\cdot)\|$. Moreover, as mentioned in Remark 5.2, in this case

$$\Psi(p_k(\alpha)) = (1 - \alpha)\Psi(x^k) + o(\alpha) \tag{5.42}$$

as $\alpha \to 0$, implying that (5.40) holds for all $\alpha > 0$ small enough.

Unfortunately, if Ψ is nonsmooth, this nice descent property of Newton-type directions is generally lost. In Sect. 5.1.2 it was demonstrated that this property remains valid for directions generated by the semismooth Newton method applied to equation reformulations of complementarity problems, employing the Fischer–Burmeister complementarity function. However, this is not necessarily the case for directions p^k generated by other Newton-type methods (e.g., by the Josephy–Newton method), or for other equation reformulations of the problem (e.g., employing different complementarity functions). In such cases, backtracking along the line segment connecting x^k and $x^k + p^k$ to produce descent fails, in general. The idea proposed in [228] to overcome this difficulty is to replace the linesearch procedure by backtracking along a path (continuous curve) connecting x^k and $\hat{x}^{k+1} = x^k + p^k$. Of course, the path should have some special properties.

In order to specify these properties, we need to introduce a more concrete understanding of what is meant here by a Newton-type step for (5.39) in the case of a nonsmooth Ψ. A mapping $\mathcal{A} : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$ is referred to as a *first-order approximation* of Ψ at $x^k \in \mathbf{R}^n$ if

$$\Psi(x) - \mathcal{A}(x^k,\, x) = o(\|x - x^k\|) \tag{5.43}$$

as $x^k \to \bar{x}$. (Note that this condition subsumes that $\mathcal{A}(x^k,\, x^k) = \Psi(x^k)$.) Then a Newton-type step consists of computing $\hat{x}^{k+1} \in \mathbf{R}^n$ solving the equation

$$\mathcal{A}(x^k,\, x) = 0. \tag{5.44}$$

If Ψ is differentiable at x^k, then $\mathcal{A}(x^k,\, x) = \Psi(x^k) + \Psi'(x^k)(x - x^k)$ satisfies (5.43), and (5.44) defines the iteration of the basic Newton method for (5.39), producing the direction $p^k = \hat{x}^{k+1} - x^k$ of the form (5.41).

Some examples of first-order approximations of nonsmooth mappings, arising from the Josephy–Newton method for complementarity problems, will be provided in the next section.

Suppose now that there exists $C_k \in (0,\, 1]$ and a continuous mapping $p_k : [0,\, C_k] \to \mathbf{R}^n$ such that

$$\mathcal{A}(x^k,\, p_k(\alpha)) = (1 - \alpha)\Psi(x^k) \quad \forall\, \alpha \in [0,\, C_k], \tag{5.45}$$

and

$$p_k(\alpha) - x^k = O(\alpha) \tag{5.46}$$

as $\alpha \to 0$, subsuming, in particular, that $p_k(0) = x^k$. From (5.43), (5.45) and (5.46) we derive that

$$\begin{aligned} \Psi(p_k(\alpha)) &= \mathcal{A}(x^k,\, p_k(\alpha)) + o(\|p_k(\alpha) - x^k\|) \\ &= (1 - \alpha)\Psi(x^k) + o(\alpha) \end{aligned} \tag{5.47}$$

as $\alpha \to 0$, similar to (5.42). Therefore, the Armijo-type condition (5.40) holds for all $\alpha \in (0,\, C_k]$ small enough, and $\alpha_k > 0$ satisfying this condition can be found by a backtracking procedure along the path $p_k([0,\, C_k])$ (which, of course, does not need to be a line segment anymore). The next iterate is then defined as $x^{k+1} = p_k(\alpha_k)$.

Observe that $\hat{x}^{k+1}$ solving the Newtonian subproblem (5.44) satisfies the equality in (5.45) for $\alpha = 1$, and in principle, it is of course desirable to have $C_k = 1$ and $p_k(1) = \hat{x}^k$, so that the full Newton-type step could be accepted. However, far from solutions of the problem, the subproblem (5.44) may have no solutions at all, while a path with the needed properties can still exist for some $C_k \in (0, 1)$. In practice, such a path is constructed by iteratively increasing C_k until either the path cannot be extended any further or (5.40) is violated for $\alpha = C_k$.

To get an understanding of what does it mean that the path cannot be extended any further, consider the following evident "path lifting" result.

Lemma 5.13. *Let $A : \mathbf{R}^n \to \mathbf{R}^n$ be such that $A(\tilde{x}) \neq 0$ for some $\tilde{x} \in \mathbf{R}^n$, and there exist a neighborhood U of $\tilde{x}$ and $\varepsilon > 0$ such that $B(A(\tilde{x}), \varepsilon) \subset A(U)$, and A is a one-to-one mapping from U to $A(U)$.*

Then it holds that for any $C \in [0, \min\{1, \varepsilon/\|A(\tilde{x})\|\}]$ the unique mapping $p : [0, C] \to U$ such that

$$A(p(\alpha)) = (1 - \alpha)A(\tilde{x}) \quad \forall\, \alpha \in [0, C], \quad p(0) = \tilde{x},$$

is given by

$$p(\alpha) = A^{-1}((1 - \alpha)A(\tilde{x})). \tag{5.48}$$

We shall say that $A : \mathbf{R}^n \to \mathbf{R}^n$ is *continuously* (*Lipschitz-continuously*) *invertible* near $x \in \mathbf{R}^n$ if there exists a neighborhood U of x such that $A(U)$ is a neighborhood of $A(x)$, A is a one-to-one mapping from U to $A(U)$, and the inverse mapping $A^{-1} : A(U) \to U$ is continuous (Lipschitz-continuous) on $A(U)$. Observe that by Brouwer's domain invariance theorem (see, e.g., [68, Theorem 2.1.11, Proposition 2.2.12]), if A is continuously (Lipschitz-continuously) invertible near x, then it is necessarily continuous (Lipschitz-continuous) on some neighborhood of x (see [68, Corollary 2.1.13]).

According to Lemma 5.13, if $C_k \in (0, 1)$ and $\mathcal{A}(x^k, \cdot)$ is continuously invertible near $p_k(C_k)$, then the path can be extended, i.e., C_k can be enlarged, and p_k can be continuously extended to the new segment $[0, C_k]$ while keeping (5.45) valid. To that end, in the path-search algorithm we present next, we consider that the path cannot be extended if either $C_k = 1$ or $\mathcal{A}(x^k, \cdot)$ is not continuously invertible near $p_k(C_k)$.

Algorithm 5.14 Choose the parameters $\sigma \in (0, 1)$ and $\theta \in (0, 1)$. Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $\Psi(x^k) = 0$, stop.
2. Define $C_k \in (0, 1]$ and a continuous mapping $p_k : [0, C_k] \to \mathbf{R}^n$ such that (5.45) holds, and either $C_k = 1$, or $\mathcal{A}(x^k, \cdot)$ is not continuously invertible near $p_k(C_k)$, or (5.40) is violated for $\alpha = C_k$.
3. Set $\alpha = C_k$. If (5.40) is satisfied, set $\alpha_k = \alpha$. Otherwise, replace α by $\theta\alpha$, check the inequality (5.40) again, etc., until (5.40) becomes valid.
4. Set $x^{k+1} = p_k(\alpha_k)$.
5. Increase k by 1 and go to step 1.

Theorem 5.15. *Let $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ be continuous on $\mathbf{R}^n$, and let the mapping $\mathcal{A} : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$ satisfy the following requirements:*

(i) $\mathcal{A}$ is a uniform first-order approximation of the mapping Ψ on the level set $X_0 = \{x \in \mathbf{R}^n \mid \|\Psi(x)\| \le c_0\}$ with some $c_0 > 0$, i.e., there exist $\delta_0 > 0$ and a function $\omega : \mathbf{R}_+ \to \mathbf{R}_+$ such that $\omega(t) = o(t)$ as $t \to 0$, and

$$\|\Psi(x) - \mathcal{A}(\tilde{x},\, x)\| \le \omega(\|x - \tilde{x}\|) \quad \forall\, \tilde{x} \in X_0,\ \forall\, x \in B(\tilde{x},\, \delta_0). \tag{5.49}$$

(ii) $\mathcal{A}(\tilde{x},\, \cdot)$ is uniformly Lipschitz-continuously invertible near each $\tilde{x} \in X_0$, i.e., there exist $\delta > 0$, $\varepsilon > 0$ and $\ell > 0$, and for each $\tilde{x} \in X_0$ there exists a neighborhood $U_{\tilde{x}}$ of $\tilde{x}$ such that $B(\tilde{x},\, \delta) \subset U_{\tilde{x}}$, $B(\Psi(\tilde{x}),\, \varepsilon) \subset \mathcal{A}(\tilde{x},\, U_{\tilde{x}})$, $\mathcal{A}(\tilde{x},\, \cdot)$ is a one-to-one mapping from $U_{\tilde{x}}$ to $\mathcal{A}(\tilde{x},\, U_{\tilde{x}})$, and its inverse is Lipschitz-continuous on $\mathcal{A}(\tilde{x},\, U_{\tilde{x}})$ with Lipschitz constant ℓ.

(iii) For any point $\tilde{x} \in X_0$, any $C \in (0,\, 1]$, and any continuous mapping $p : [0,\, C) \to \mathbf{R}^n$ satisfying

$$\mathcal{A}(\tilde{x},\, p(\alpha)) = (1 - \alpha)\Psi(\tilde{x}) \quad \forall\, \alpha \in [0,\, C), \quad p(0) = \tilde{x}, \tag{5.50}$$

and such that $\mathcal{A}(\tilde{x},\, \cdot)$ is continuously invertible near $p(\alpha)$ for each value $\alpha \in [0,\, C)$, there exists $p_C = \lim_{\alpha \to C-} p(\alpha)$, and $\mathcal{A}(\tilde{x},\, p_C) = (1 - C)\Psi(\tilde{x})$.

Then for any starting point $x^0 \in X_0$, Algorithm 5.14 generates an iterative sequence $\{x^k\}$ convergent to a solution of (5.39), and the rate of convergence is superlinear.

The technical assumption (iii) of this theorem is the so-called *path continuation property*; it is needed in order to establish the existence of paths in Algorithm 5.14. If $\mathcal{A}(\tilde{x},\, \cdot)$ is continuous at p_C, then the fact that the equality $\mathcal{A}(\tilde{x},\, p_C) = (1 - C)\Psi(\tilde{x})$ holds is obvious. So the essence of this condition is the existence of $p_C = \lim_{\alpha \to C-} p(\alpha)$. The latter property is automatic, e.g., when $p(\cdot)$ is piecewise affine on $(0,\, C)$ (see Remark 2.55; this follows from the fact that, as can be easily checked, in this case $p(\cdot)$ is necessarily Lipschitz-continuous on $(0,\, C)$).

Proof. To show that Algorithm 5.14 is well defined, observe first that by assumption (ii) and Lemma 5.13, for any $x^k \in X_0$ such that $\Psi(x^k) \ne 0$, there exist the largest set $S_k \subset [0,\, 1]$ and the unique mapping $p_k : S_k \to \mathbf{R}^n$ such that either $S_k = [0,\, 1]$ or $S_k = [0,\, C_k)$ with some $C_k \in (0,\, 1]$,

$$\mathcal{A}(x^k,\, p_k(\alpha)) = (1 - \alpha)\Psi(x^k) \quad \forall\, \alpha \in S_k, \quad p_k(0) = x^k,$$

and $\mathcal{A}(x^k,\, \cdot)$ is continuously invertible near $p_k(\alpha)$ for all $\alpha \in S_k$.

If $S_k = [0,\, 1]$, then $C_k = 1$ and this p_k will be accepted at step 2 of Algorithm 5.14. On the other hand, if $S_k = [0,\, C_k)$, then according to assumption (iii), p_k can be continuously extended to $[0,\, C_k]$, in particular by setting $p_k(C_k) = \lim_{\alpha \to C_k-} p_k(\alpha)$, and then $\mathcal{A}(x^k,\, p_k(C_k)) = (1 - C_k)\Psi(x^k)$

holds true. In this case, since S_k is maximal, it holds that $\mathcal{A}(x^k, \cdot)$ is not continuously invertible near $p_k(C_k)$, and hence, these C_k and p_k will again be accepted at step 2 of Algorithm 5.14.

According to assumption (ii), the inverse $\mathcal{A}_k^\dagger$ of $\mathcal{A}(x^k, \cdot)$ is Lipschitz-continuous on $\mathcal{A}(x^k, U_{x^k})$ with Lipschitz constant ℓ, and hence, formula (5.48) in Lemma 5.13 implies that

$$\begin{aligned}\|p_k(\alpha) - x^k\| &= \|\mathcal{A}_k^\dagger((1-\alpha)\Psi(x^k)) - \mathcal{A}_k^\dagger(\Psi(x^k))\| \\ &\leq \ell\|(1-\alpha)\Psi(x^k) - \Psi(x^k)\| \\ &= \ell\alpha\|\Psi(x^k)\| \end{aligned} \tag{5.51}$$

for all $\alpha \geq 0$ small enough, giving the estimate (5.46). Therefore, according to the discussion above, assumption (i) implies (5.42). Hence, the value $\alpha_k > 0$ satisfying (5.40) will be found and accepted by step 3 of Algorithm 5.14, and the next iterate x^{k+1} will be defined by step 4 of this algorithm.

Furthermore, (5.40) implies that $\|\Psi(x^{k+1})\| < \|\Psi(x^k)\|$; hence, $x^{k+1} \in X_0$. Therefore, for any $x^0 \in X_0$ Algorithm 5.14 generates the iterative sequence $\{x^k\} \subset X_0$, and assuming that $\Psi(x^k) \neq 0$ for all k, the corresponding sequence $\{\|\Psi(x^k)\|\}$ is monotonically decreasing. Since the latter sequence is bounded below (by zero), it converges.

By the construction of C_k and p_k above, by assumption (ii), and by Lemma 5.13, for all k it holds that

$$C_k \geq \min\left\{1,\ \frac{\varepsilon}{\|\Psi(x^k)\|}\right\}, \tag{5.52}$$

and, similarly to (5.51),

$$\|p_k(\alpha) - x^k\| \leq \ell\alpha\|\Psi(x^k)\| \quad \forall\, \alpha \in \left[0,\ \min\left\{1,\ \frac{\varepsilon}{\|\Psi(x^k)\|}\right\}\right]. \tag{5.53}$$

Furthermore, since according to assumption (i) it holds that $\omega(t) = o(t)$ as $t \to 0$, we can choose $T \in (0,\ \delta_0]$ such that

$$\omega(t) \leq \frac{1-\sigma}{\ell} t \quad \forall\, t \in [0,\ T]. \tag{5.54}$$

Employing (5.53), we further obtain that

$$\|p_k(\alpha) - x^k\| \leq T \quad \forall\, \alpha \in [0,\ \hat{C}_k], \tag{5.55}$$

implying, in particular, that

$$p_k(\alpha) \in B(x^k,\ \delta_0) \quad \forall\, \alpha \in [0,\ \hat{C}_k], \tag{5.56}$$

where

$$\hat{C}_k = \min\left\{1,\, \frac{\varepsilon}{\|\Psi(x^k)\|},\, \frac{T}{\ell\|\Psi(x^k)\|}\right\}. \tag{5.57}$$

Similarly to (5.42), from assumption (i) and from the relations (5.45) and (5.53)–(5.56) we now derive that

$$\begin{aligned}\|\Psi(p_k(\alpha))\| &\leq \|\mathcal{A}(x^k,\, p_k(\alpha))\| + \omega(\|p_k(\alpha) - x^k\|)\\ &\leq (1-\alpha)\|\Psi(x^k)\| + (1-\sigma)\alpha\|\Psi(x^k)\|\\ &= (1-\sigma\alpha)\|\Psi(x^k)\| \quad \forall\, \alpha \in [0,\, \hat{C}_k].\end{aligned} \tag{5.58}$$

From (5.52), (5.57), and from the inclusion $\{x^k\} \subset X_0$, we have that

$$C_k \geq \hat{C}_k \geq \min\left\{1,\, \frac{\varepsilon}{c_0},\, \frac{T}{\ell c_0}\right\},$$

where the positive constant in the right-hand side does not depend on k. Comparing (5.58) with (5.40) we now conclude that there exists $c > 0$ such that

$$\alpha_k \geq c \quad \forall\, k.$$

Combining this with (5.40), we conclude that

$$\|\Psi(x^{k+1})\| \leq (1-\sigma c)\|\Psi(x^k)\| \quad \forall\, k, \tag{5.59}$$

implying that $\{\Psi(x^k)\}$ converges to 0 (at a linear rate).

The latter conclusion combined with (5.52) and (5.57) implies that for all k large enough it holds that $C_k = \hat{C}_k = 1$, and in particular, (5.58) implies that (5.40) is satisfied with $\alpha = 1$. Therefore, for all k large enough it holds that $\alpha_k = 1$, and from (5.53) and (5.59) we derive the estimate

$$\|x^{k+1} - x^k\| = \|p_k(1) - x^k\| \leq \ell\|\Psi(x^k)\| \leq \ell q^k\|\Psi(x^0)\|,$$

where $q = 1 - \sigma c$. Then for any k and j it holds that

$$\begin{aligned}\|x^{k+j} - x^k\| &\leq \sum_{i=k}^{k+j-1} \|x^{i+1} - x^i\|\\ &\leq \sum_{i=k}^{k+j-1} \ell q^i\|\Psi(x^0)\|\\ &\leq \frac{\ell\|\Psi(x^0)\|}{1-q}q^k \to 0 \text{ as } k \to \infty.\end{aligned}$$

Therefore, $\{x^k\}$ is a Cauchy sequence, and hence it converges to some $\bar{x} \in \mathbf{R}^n$. Since $\{\Psi(x^k)\}$ converges to 0, by the continuity of Ψ we obtain that $\bar{x}$ is a solution of (5.39).

Observe, finally, that due to the convergence of $\{x^k\}$ to $\bar{x}$, using also the continuity of Ψ and assumption (i), the inclusions $\bar{x} \in B(x^k, \delta)$ and $\mathcal{A}(x^k, \bar{x}) \in B(\Psi(x^k), \varepsilon)$ hold for all k large enough. Similarly to (5.51), by assumptions (i) and (ii) we then derive that

$$\begin{aligned} \|x^{k+1} - \bar{x}\| &= \|p_k(1) - \bar{x}\| \\ &= \|\mathcal{A}_k^\dagger(0) - \mathcal{A}_k^\dagger(\mathcal{A}(x^k, \bar{x}))\| \\ &\leq \ell\|\Psi(\bar{x}) - \mathcal{A}(x^k, \bar{x})\| \\ &\leq \ell\omega(\|\bar{x} - x^k\|) \\ &= o(\|x^k - \bar{x}\|) \end{aligned}$$

as $k \to \infty$, establishing the superlinear rate of convergence. □

The convergence properties established in this theorem are very strong: we have global convergence of Algorithm 5.14 at a superlinear rate to a solution of (5.39). Properties so strong come at a price, of course: the assumptions (i) and (ii) in Theorem 5.15 are rather restrictive, in general. Apparently, these assumptions are intrinsically related to the nature of the path-search strategy; they appear very difficult to relax.

5.2.2 Path-Search Methods for Complementarity Problems

We now outline how Algorithm 5.14 can be implemented for the nonlinear complementarity problem (NCP)

$$x \geq 0, \quad \Phi(x) \geq 0, \quad \langle x, \Phi(x) \rangle = 0, \tag{5.60}$$

where $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ is a smooth mapping.

We consider the equation reformulation (5.39) of the NCP (5.60), employing the natural residual complementarity function. Namely, define the mapping $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ by

$$\Psi(x) = \min\{x, \Phi(x)\}. \tag{5.61}$$

We remark that [228] uses a different reformulation, employing the so-called normal map as Ψ. This choice can be advantageous [68, Sect. 1.5], in particular because it requires computing the values of Φ within $\mathbf{R}^n_+$ only (and thus Φ is allowed to be undefined outside $\mathbf{R}^n_+$). However, for our discussion here we prefer to use the natural residual, as it is somewhat simpler, and because the normal map is not used anywhere else in this book.

We start with choosing the first-order approximation mapping. Recall from Sect. 3.1 that the iteration subproblem of the Josephy–Newton method for the NCP (5.60) is the linear complementarity problem (LCP)

$$x \geq 0, \quad \Phi(x^k) + \Phi'(x^k)(x - x^k) \geq 0, \quad \langle x, \Phi(x^k) + \Phi'(x^k)(x - x^k) \rangle = 0,$$

where $x^k \in \mathbf{R}^n$ is the current iterate. This subproblem can be equivalently written in the form (5.44), where the mapping $\mathcal{A} : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$ is again defined using the natural residual complementarity function:

$$\mathcal{A}(\tilde{x},\, x) = \min\{x,\, \Phi(\tilde{x}) + \Phi'(\tilde{x})(x - \tilde{x})\}. \tag{5.62}$$

Definitions (5.61) and (5.62), and the obvious property

$$|\min\{a,\, b\} - \min\{a,\, c\}| \leq |b - c| \quad \forall\, a,\, b,\, c \in \mathbf{R},$$

imply that for any $x,\, \tilde{x} \in \mathbf{R}^n$ it holds that

$$\begin{aligned} |\Psi(x) - \mathcal{A}(\tilde{x},\, x)| &= |\min\{x,\, \Phi(x)\} - \min\{x,\, \Phi(\tilde{x}) + \Phi'(\tilde{x})(x - \tilde{x})\}| \\ &\leq |\Phi(x) - \Phi(\tilde{x}) + \Phi'(\tilde{x})(x - \tilde{x})|, \end{aligned}$$

and therefore, by Theorem A.10, (a),

$$\|\Psi(x) - \mathcal{A}(\tilde{x},\, x)\| \leq \sup_{t \in [0,\, 1]} \|\Phi'(tx + (1 - t)\tilde{x}) - \Phi'(\tilde{x})\| \|x - \tilde{x}\|.$$

Hence, if the derivative of Φ is continuous at $\tilde{x}$, then $\mathcal{A}$ is a first-order approximation of Ψ at $\tilde{x}$. Moreover, if, e.g., there exist $\delta_0 > 0$ and $L > 0$ such that for all $\tilde{x} \in X_0$ the derivative of Φ is Lipschitz-continuous on $B(\tilde{x},\, \delta_0)$ with constant L, then this first-order approximation is uniform, i.e., condition (5.49) from assumption (i) of Theorem 5.15 holds with $\omega(t) = Lt^2$. This assumption on the derivative of Φ is quite natural, especially when the set X_0 is bounded.

Furthermore, for any $\tilde{x} \in \mathbf{R}^n$, the mapping $\mathcal{A}(\tilde{x},\, \cdot)$ is piecewise affine. Hence, it is Lipschitz-continuous on $\mathbf{R}^n$, and according to Theorem 1.66, a sufficient condition for the Lipschitz-continuous invertibility of $\mathcal{A}(\tilde{x},\, \cdot)$ near $\tilde{x}$ is the CD-regularity of this mapping at $\tilde{x}$.

To get some understanding of the latter condition, define the index sets

$$\begin{aligned} I_0 &= \{i = 1,\, \ldots,\, n \mid \tilde{x}_i = \Phi_i(\tilde{x})\}, \\ I_1 &= \{i = 1,\, \ldots,\, n \mid \tilde{x}_i > \Phi_i(\tilde{x})\}, \\ I_2 &= \{i = 1,\, \ldots,\, n \mid \tilde{x}_i < \Phi_i(\tilde{x})\}, \end{aligned} \tag{5.63}$$

Observe that $\tilde{\xi} = \tilde{x} - \Psi(\tilde{x})$ is a solution of the auxiliary LCP

$$\xi \geq 0, \quad \tilde{b} + \Phi'(\tilde{x})\xi \geq 0, \quad \langle \xi,\, \tilde{b} + \Phi'(\tilde{x})\xi \rangle = 0, \tag{5.64}$$

where $\tilde{b} = \Phi(\tilde{x}) + \Phi'(\tilde{x})(\Psi(\tilde{x}) - \tilde{x}) - \Psi(\tilde{x})$, and moreover, for the index sets

$$\begin{aligned} I_0(\tilde{\xi}) &= \{i = 1,\, \ldots,\, n \mid \tilde{\xi}_i = (b + \Phi'(\tilde{x})\xi)_i = 0\}, \\ I_1(\tilde{\xi}) &= \{i = 1,\, \ldots,\, n \mid \tilde{\xi}_i > 0,\ (b + \Phi'(\tilde{x})\xi)_i = 0\}, \\ I_2(\tilde{\xi}) &= \{i = 1,\, \ldots,\, n \mid \tilde{\xi}_i = 0,\ (b + \Phi'(\tilde{x})\xi)_i > 0\}, \end{aligned} \tag{5.65}$$

it holds that

$$I_0(\tilde{\xi}) = I_0, \quad I_1(\tilde{\xi}) = I_1, \quad I_2(\tilde{\xi}) = I_2. \tag{5.66}$$

Define the piecewise affine mapping $\tilde{\Psi} : \mathbf{R}^n \to \mathbf{R}^n$,

$$\tilde{\Psi}(\xi) = \min\{\xi,\, \tilde{b} + \Phi'(\tilde{x})\xi\}.$$

From (5.62)–(5.66) and from Proposition 3.10, it follows that the generalized Jacobian of $\mathcal{A}(\tilde{x}, \cdot)$ at $\tilde{x}$ is contained in $\operatorname{conv} \Delta_{\mathrm{NR}}(\tilde{\xi})$, where $\Delta_{\mathrm{NR}}(\tilde{\xi})$ is the outer estimate of $\partial_B \tilde{\Psi}(\tilde{\xi})$ defined as in Sect. 3.2.1: it consists of all the matrices $J \in \mathbf{R}^{n \times n}$ with rows

$$J_i = \begin{cases} \Phi_i'(\tilde{x}) \text{ or } e^i & \text{if } i \in I_0(\tilde{\xi}), \\ \Phi_i'(\tilde{x}) & \text{if } i \in I_1(\tilde{\xi}), \\ e^i & \text{if } i \in I_2(\tilde{\xi}). \end{cases}$$

Therefore, the CD-regularity of $\mathcal{A}(\tilde{x}, \cdot)$ at $\tilde{x}$ is automatic provided all the matrices in $\operatorname{conv} \Delta_{\mathrm{NR}}(\tilde{\xi})$ are nonsingular. Furthermore, according to Corollary 3.20, the latter is equivalent to the strong regularity of the solution $\tilde{\xi}$ of the LCP (5.64). This property is characterized in Proposition 1.26: it holds if and only if $(\Phi'(\tilde{x}))_{I_1 I_1}$ is a nonsingular matrix, and

$$(\Phi'(\tilde{x}))_{I_0 I_0} - (\Phi'(\tilde{x}))_{I_0 I_1}((\Phi'(\tilde{x}))_{I_1 I_1})^{-1}(\Phi'(\tilde{x}))_{I_1 I_0}$$

is a P-matrix, where (5.65), (5.66) were again taken into account.

Moreover, if the latter condition is satisfied at each $\tilde{x} \in X_0$, if Φ is twice differentiable near each $\tilde{x} \in X_0$, with its second derivative being continuous at $\tilde{x}$, and if X_0 is bounded, then it can be derived from Theorem 1.67 and from the compactness of X_0, that $\mathcal{A}(\tilde{x}, \cdot)$ is uniformly Lipschitz-continuously invertible near each $\tilde{x} \in X_0$, i.e., assumption (ii) of Theorem 5.15 holds. In order to see this, one needs to estimate the full generalized Jacobian $\partial \mathcal{A}(\tilde{x}, \tilde{x})$. Similarly to Proposition 3.10 it can be seen that any matrix in this set has the form $(J \;\; 0)$, where $J \in \operatorname{conv} \Delta_{\mathrm{NR}}(\tilde{\xi})$. We omit the technical details.

Finally, since $\mathcal{A}(\tilde{x}, \cdot)$ defined according to (5.62) is piecewise affine, it can be seen that for any $C \in (0, 1]$ and any continuous mapping $p : [0, C) \to \mathbf{R}^n$ satisfying (5.50) and such that $\mathcal{A}(\tilde{x}, \cdot)$ is continuously invertible near $p(\alpha)$ for each $\alpha \in [0, C)$, it holds that $p(\cdot)$ is piecewise affine on $(0, C)$. The latter implies that assumption (iii) of Theorem 5.15 holds as well.

Some explanations about practical computation of points on the path in the case of the NCP (5.60) are in order. For the current iterate $x^k \in \mathbf{R}^n$, we set $p_k(0) = x^k$, and according to (5.45) and (5.62), for any $\alpha \in (0, 1]$ the subproblem for computing $p_k(\alpha)$ has the form

$$\min\{x,\, \Phi(x^k) + \Phi'(x^k)(x - x^k)\} = (1 - \alpha)\Psi(x^k).$$

Then, similarly to the discussion above, $p_k(\alpha) = \xi_k(\alpha) + (1-\alpha)\Psi(x^k)$, where $\xi_k(\alpha) \in \mathbf{R}^n$ solves the LCP

$$\xi \geq 0, \quad b^k + \alpha d^k + \Phi'(x^k)\xi \geq 0, \quad \langle \xi,\, b^k + \alpha d^k + \Phi'(x^k)\xi \rangle = 0, \tag{5.67}$$

with

$$b^k = \Phi(x^k) + \Phi'(x^k)(\Psi(x^k) - x^k) - \Psi(x^k)), \quad d^k = -(\Phi'(x^k)\Psi(x^k) - \Psi(x^k)).$$

Considering α as a variable, and adding to the LCP (5.67) the additional constraint $0 \leq \alpha \leq 1$, the resulting problem can be solved by a variant of Lemke's algorithm [46, Sect. 4.4] to construct the needed $\xi_k(\cdot)$, and hence, $p_k(\cdot)$. See [228] for more details of this procedure.

5.3 Hybrid Global and Local Phase Algorithms

Any algorithm known to provide the superlinear decrease of the distance to a "qualified" solution when close to it can be globalized by a "hybrid" strategy, which means implementing it as a local phase of some globally convergent scheme. This local phase is activated according to some criteria designed to recognize when the iterate appears close enough to a solution. From a different viewpoint, any method with reasonable global convergence properties, which we further refer to as an *outer-phase method*, can be equipped with an option of switching to some potentially fast local method, in order to accelerate convergence. This (rather universal) approach to globalization is the subject of this section.

In principle, in such hybrid constructions the local method and the outer-phase do not need to have anything in common; for instance, when far from a solution search directions generated by the local method do not have to be directions of descent for any merit function used on the outer-phase. However, if this is the case, it is necessary to introduce some safeguards, such as backtracking to restore the normal course of the outer-phase when it is recognized after some steps that switching to the local method turned out to be premature. Thus such safeguards mean disregarding some eventually unsuccessful iterations of the local method. Of course, this can add significantly to the overall computational costs, which is a potential drawback of this universal approach. As will be seen below, in a variety of situations there exist alternative possibilities of controlling the overall process by means of a single merit function.

5.3.1 Hybrid Iterative Frameworks

We first present two hybrid algorithmic frameworks, with a generic local method and a generic outer-phase method. The first algorithm employs backup safeguards, while the second employs the record values of some chosen merit function $\varphi : \mathbf{R}^n \to \mathbf{R}_+$ with nonnegative values. The merit function is the only element in these model algorithms representing the problem to be solved; specifically, solutions of the problem are characterized by the equality

$$\varphi(x) = 0,$$

which also means that they are global minima of φ on $\mathbf{R}^n$. That, in general, only global minima of a merit function are solutions of the underlying problem is a typical situation in the variational setting; see [68].

Algorithm 5.16 Choose the parameter $q \in (0, 1)$. Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $\varphi(x^k) = 0$, stop.
2. If $k = 0$ or the point x^k was computed by the outer-phase method, set $\tilde{k} = k$ and store $x^{\tilde{k}}$. Compute $x^{k+1} \in \mathbf{R}^n$ by the iteration of the local method, if possible. If x^{k+1} is successfully computed by the local method and satisfies the property
$$\varphi(x^{k+1}) \leq q\varphi(x^k), \tag{5.68}$$
go to step 4.
3. If the point x^k was computed by the local method, set $k = \tilde{k}$, $x^k = x^{\tilde{k}}$. Compute $x^{k+1} \in \mathbf{R}^n$ by the iteration of the outer-phase method.
4. Increase k by 1 and go to step 1.

Remark 5.17. Global convergence properties of Algorithm 5.16 are transparent. If in a sequence $\{x^k\}$ generated by this algorithm all iterates from some index on are computed by the local method (i.e., if backup safeguards are used at most a finite number of times), then from (5.68) we conclude that

$$\varphi(x^k) \to 0 \text{ as } k \to \infty, \tag{5.69}$$

and in particular, every accumulation point of $\{x^k\}$ is a solution of the problem in question. The only alternative scenario is that all the iterates are generated by the outer-phase method (by which we mean that iterations of the local method were discarded), and therefore, the algorithm inherits presumably reasonable global convergence properties of the outer-phase.

We now show that Algorithm 5.16 preserves the superlinear convergence rate of the local method if the latter has this property, and if the merit function provides the local error bound for the underlying problem.

Theorem 5.18. *Let $\varphi : \mathbf{R}^n \to \mathbf{R}_+$ be locally Lipschitz-continuous with respect to $\bar{x} \in \mathbf{R}^n$, and assume that $\varphi(\bar{x}) = 0$ and the error bound*

$$x - \bar{x} = O(\varphi(x)) \tag{5.70}$$

holds as $x \in \mathbf{R}^n$ *tends to* $\bar{x}$*. Assume further that for any* $x^k \in \mathbf{R}^n$ *close enough to* $\bar{x}$ *the local method generates a point* $x^{k+1} \in \mathbf{R}^n$ *such that*

$$x^{k+1} - \bar{x} = o(\|x^k - \bar{x}\|) \tag{5.71}$$

as $x^k \to \bar{x}$*.*

If $\bar{x}$ *is an accumulation point of a sequence* $\{x^k\} \subset \mathbf{R}^n$ *generated by Algorithm 5.16, then the entire sequence* $\{x^k\}$ *converges to* $\bar{x}$*, and the rate of convergence is superlinear.*

Proof. If $x^k \in \mathbf{R}^n$ is close enough to $\bar{x}$, then according to (5.70) and (5.71), and also taking into account the local Lipschitz-continuity of φ with respect to $\bar{x}$ and the equality $\varphi(\bar{x}) = 0$, for a point x^{k+1} computed by the step of the local method we obtain that

$$\begin{aligned} \varphi(x^{k+1}) = \varphi(x^{k+1}) - \varphi(\bar{x}) &= O(\|x^{k+1} - \bar{x}\|) \\ &= o(\|x^k - \bar{x}\|) \\ &= o(\varphi(x^k)) \end{aligned}$$

as $x^k \to \bar{x}$. The above relation implies that for any pre-fixed $q \in (0,\, 1)$, if x^k is close enough to $\bar{x}$, then (5.68) holds, and hence, the point x^{k+1} is accepted by the algorithm. This shows that the iteration of Algorithm 5.16 reduces to the iteration of the local method, and the assertion follows from (5.71). □

We proceed with the second algorithmic framework, using records of the best merit function value at previous iterations, instead of backup safeguards.

Algorithm 5.19 Choose the parameter $q \in (0,\, 1)$. Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$. Set $\varphi_{\mathrm{rec}} = \varphi(x^0)$.

1. If $\varphi(x^k) = 0$, stop.
2. Compute $x^{k+1} \in \mathbf{R}^n$ by the iteration of the local method, if possible. If such x^{k+1} was successfully computed and satisfies the property
$$\varphi(x^{k+1}) \leq q\varphi_{\mathrm{rec}}, \tag{5.72}$$
set $\varphi_{\mathrm{rec}} = \varphi(x^{k+1})$ and go to step 4.
3. Compute $x^{k+1} \in \mathbf{R}^n$ by the iteration of the outer-phase method. If
$$\varphi(x^{k+1}) < \varphi_{\mathrm{rec}},$$
set $\varphi_{\mathrm{rec}} = \varphi(x^{k+1})$.
4. Increase k by 1 and go to step 1.

Remark 5.20. Global convergence properties of Algorithm 5.19 are again quite transparent. If infinitely many iterates x^{k_j} in a sequence $\{x^k\}$ generated by this algorithm are computed by the local method, we have that

$$\varphi(x^{k_j}) \to 0 \text{ as } j \to \infty. \tag{5.73}$$

Indeed, by the construction of the algorithm, all these iterates satisfy the acceptance test (5.72) (for $k = k_j - 1$), and the record value φ_{rec} decreases by the factor of q infinitely many times (and never increases). Therefore, there exists a sequence $\{\varphi_k\}$ of nonnegative reals, convergent to zero and such that

$$\varphi(x^{k_j}) \le q\varphi_{k_j}$$

holds for all j. This implies (5.73), and in particular, every accumulation point of $\{x^{k_j}\}$ is a solution of the problem in consideration.

Otherwise, all the iterates from some index on are generated by the outer-phase method, and the algorithm inherits its presumably reasonable global convergence properties.

The superlinear convergence of the local method is again preserved, in the following sense.

Theorem 5.21. *Under the assumptions of Theorem 5.18, if a sequence $\{x^k\} \subset \mathbf{R}^n$ generated by Algorithm 5.19 converges to $\bar{x}$, then the rate of convergence is superlinear.*

Proof. If the record value φ_{rec} changes infinitely many times, we can consider x^k close enough to $\bar{x}$, and such that $\varphi_{\mathrm{rec}} = \varphi(x^k)$. Then we argue identically to the proof of Theorem 5.18, to conclude that x^{k+1} computed by the step of the local method is accepted by Algorithm 5.19, and the algorithm will be working identically to the local method. Then the superlinear rate of convergence of $\{x^k\}$ to $\bar{x}$ follows.

Alternatively, assuming that the record value $\varphi_{\mathrm{rec}} > 0$ is constant from some point on (which of course subsumes that the steps of the local method are accepted at most a finite number of times), we arrive at a contradiction: $\varphi(x^k)$ tends to 0 (since $\{x^k\}$ tends to $\bar{x}$) and therefore, (5.72) will necessarily hold as a strict inequality for a sufficiently large k, whatever is the fixed value φ_{rec}; then, the record would have to be changed. □

Speaking about methods for complementarity problems, a natural choice of a merit function φ is given by residuals based on complementarity functions discussed in Sect. 3.2.

For simplicity, consider the nonlinear complementarity problem (NCP)

$$x \ge 0, \quad \Phi(x) \ge 0, \quad \langle x, \Phi(x)\rangle = 0, \tag{5.74}$$

where $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ is a smooth mapping. Define the mapping $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ by

$$\Psi(x) = \psi(x, \Phi(x)), \tag{5.75}$$

where $\psi : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ is either the natural residual

$$\psi(a, b) = \min\{a, b\}, \tag{5.76}$$

or the Fischer–Burmeister function

$$\psi(a,\, b) = a + b - \sqrt{a^2 + b^2}. \tag{5.77}$$

Define the function $\varphi : \mathbf{R}^n \to \mathbf{R}_+$,

$$\varphi(x) = \|\Psi(x)\|. \tag{5.78}$$

According to Proposition 3.8, if Φ is differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$ and its derivative is continuous at $\bar{x}$, then Ψ is semismooth (and hence, locally Lipschitz-continuous) at $\bar{x}$. The latter evidently implies that φ is locally Lipschitz-continuous at $\bar{x}$. Moreover, as discussed in Sect. 3.2.1, for any solution $\bar{x}$ of NCP (5.74) it holds that $\varphi(\bar{x}) = 0$, and from Propositions 1.34, 1.64, and 3.21 it follows that the semistability of solution $\bar{x}$ is equivalent to the error bound (5.70).

For variational problems, one natural candidate for the role of the local phase in Algorithms 5.16 and 5.19 is the Josephy–Newton method discussed in Sect. 3.1. According to Theorem 3.2, this method possesses local superlinear convergence to a solution which is hemistable and semistable. The Josephy–Newton method is suitable for a natural path-search globalization strategy (see Sect. 5.2). On the other hand, the latter may also be used in conjunction with some outer-phase method, e.g., in order to avoid failures when the first-order approximation provided by the Josephy–Newton scheme fails to be continuously invertible near the current iterate (which can easily happen for iterates far from a solution).

For complementarity problems, other possible candidates for the local phase in Algorithms 5.16 and 5.19 are Newton-type active-set methods discussed in Sect. 3.4. According to Theorem 3.56, if Φ is twice differentiable near a semistable solution $\bar{x} \in \mathbf{R}^n$ of the NCP (5.74), with its second derivative continuous at $\bar{x}$, the active-set Gauss–Newton method given by Algorithm 3.55 possesses local superlinear convergence to $\bar{x}$.

Summarizing, for the NCP (5.74) we now have all the ingredients needed for Algorithms 5.16 and 5.19. Their global convergence properties are stated in Remarks 5.17 and 5.20, and local rates of convergence are given by Theorems 5.18 and 5.21.

As for the outer-phase method, in the case of the NCP (5.74) one can take Algorithm 5.6, for instance. The reason for combining Algorithms 3.55 and 5.6 in such a hybrid way is that the active-set method of the former achieves superlinear convergence under semistability, while the semismooth Newton method of the latter requires the stronger assumption of BD-regularity; see Theorem 5.9. On the other hand, as will be demonstrated in the next section, this choice of an outer-phase method allows to improve and simplify the algorithmic frameworks in consideration.

Extensions to mixed complementarity problems (and in particular, to Karush–Kuhn–Tucker systems) can be obtained employing the constructions from Sects. 3.2.2, 3.4.2, and 5.1.3. Moreover, possible choices of an outer-phase method are of course not limited to the linesearch method of

Algorithm 5.6, or more generally, to any linesearch methods. Just as an example, the hybrid construction close to Algorithm 5.16 was introduced in [22] in the optimization context, with an active-set Newton-type local method, and with the augmented Lagrangian algorithm as an outer-phase method.

5.3.2 Preserving Monotonicity of Merit Function Values

The hybrid Algorithms 5.16 and 5.19 can be simplified if in the outer-phase method the value of the merit function φ cannot increase. In this case, backup safeguards can be dropped in Algorithm 5.16, while in Algorithm 5.19 the record value φ_{rec} is updated at each iteration and therefore there is no need to track it. Then the two algorithms become identical to the following framework.

Algorithm 5.22 Choose the parameter $q \in (0, 1)$. Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $\varphi(x^k) = 0$, stop.
2. Compute $x^{k+1} \in \mathbf{R}^n$ by the iteration of the local method, if possible. If such x^{k+1} is computed successfully and (5.68) holds, go to step 4.
3. Compute $x^{k+1} \in \mathbf{R}^n$ by the iteration of the outer-phase method.
4. Increase k by 1 and go to step 1.

Remark 5.23. If steps of the outer-phase method do not increase the value of the merit function φ, the sequence $\{\varphi(x^k)\}$ corresponding to any sequence $\{x^k\}$ generated by Algorithm 5.22 is nonincreasing. With this fact at hand, if there exists a subsequence $\{x^{k_j}\}$ of iterates computed by the steps of the local method, from (5.68) we obtain that the subsequence $\{\varphi(x^{k_j})\}$ converges to zero; hence, the entire sequence $\{\varphi(x^k)\}$ goes to zero (i.e., (5.69) holds). Therefore, every accumulation point of $\{x^k\}$ is a solution of the problem in question. The only alternative scenario is that all the iterates from some index on are generated by the outer-phase method. In that case, the algorithm again inherits global convergence properties of the latter.

Proof of the superlinear convergence rate in the statement below literally repeats the proof of Theorem 5.18, with Algorithm 5.16 substituted by Algorithm 5.22.

Theorem 5.24. *Under the assumptions of Theorem 5.18, if $\bar{x}$ is an accumulation point of a sequence $\{x^k\} \subset \mathbf{R}^n$ generated by Algorithm 5.22, then the entire sequence $\{x^k\}$ converges to $\bar{x}$, and the rate of convergence is superlinear.*

Getting back to the NCP (5.74), let the merit function φ be defined according to (5.75), (5.78), and (5.77), i.e., using the Fischer–Burmeister complementarity reformulation. If the linesearch method of Algorithm 5.6 is used for the outer-phase method, the needed monotonicity property for the merit function φ is automatic, because linesearch is performed for the square of φ. Global convergence properties of this variant of Algorithm 5.22 are characterized by Remark 5.23, where the properties called inherited are stated in Theorem 5.7.

According to the discussion in the previous section, in the case of complementarity problems some possible candidates for the local method in Algorithm 5.22 are the Josephy–Newton method and the Newton-type active-set methods discussed in Sect. 3.4. The latter result in asymptotic superlinear convergence under weaker assumptions. Employing various active-set methods, hybrid algorithms of this kind were developed, e.g., in [51, 69, 163] for complementarity problems, and in [137, 143] for some special classes of optimization problems.

Another practically relevant choice of a local method for complementarity problems is the semismooth Newton method applied to the equation reformulation using the natural residual (5.76). This variant of the semismooth Newton method is known to be very efficient locally, but it is hardly amenable to practical non-hybrid globalizations. At the same time, it can be used within Algorithm 5.22 in conjunction with Algorithm 5.6 or with other outer-phase methods, and perhaps for other merit functions. See [164] for one example of a construction of this kind, employing the so-called D-gap function [68, Sect. 10.3] as φ.

For mixed complementarity problems, an algorithmic framework closely related to Algorithm 5.22 (with a generic local method and linesearch outer-phase, and with the additional feature of feasibility of generated sequences) is developed in [80].

Observe finally that the linear decrease test (5.68) can also be used in the "less hybrid" algorithms of Sect. 5.1, as an additional option to accept full Newton-type steps before resorting to linesearch [53]. This can be regarded as a hybrid algorithm in which the role of both local and outer-phase methods is played by the semismooth Newton method, equipped with linesearch and the safeguard to switch to the steepest descent iterate in the outer-phase case.

5.4 Proximal Point-Based Globalization for Monotone Problems

In this section we describe an approach to globalizing Newtonian methods for the special but important class of monotone variational problems. Note that most "general-purpose" globalization strategies rely on decreasing the value, from one iteration to the next, of a suitable merit function for the given problem. It is then natural that one can only guarantee convergence to local minimizers (or even merely critical points) of this merit function, while solutions of the original variational problem generally correspond to global minimizers; see Sects. 5.1 and 5.3. Some more detailed discussion of those issues would be provided below in the context of equations in Sect. 5.4.2 and of variational inequalities in Sect. 5.4.3.

The technique presented here is based on the framework of inexact proximal point methods with relative-error approximation criteria. The key to the construction is that, when close to a solution with appropriate properties, the regularized Newton step provides an acceptable approximate solution of the proximal point subproblem, and thus convergence and fast rate of convergence are inherited from the underlying proximal point scheme. Globally, a linesearch may be needed to ensure a certain separation property, with the subsequent projection step to force progress towards solutions. Among the attractive features of this approach are the following desirable properties: the method is globally well defined; the Newtonian directions are never discarded; global convergence of the whole sequence to some solution point holds without any regularity assumptions (the solution set can even be unbounded); and local superlinear convergence rate is achieved under natural conditions.

5.4.1 Inexact Proximal Point Framework with Relative-Error Approximations

Consider the generalized equation (GE)

$$\Psi(x) \ni 0, \tag{5.79}$$

where Ψ is a maximal monotone set-valued mapping from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$ (see Sect. A.3 for the definitions and basic facts related to monotonicity). For now, (5.79) can be an inclusion without any special structure, i.e., we do not split Ψ into a single-valued base and a set-valued field, as in the material covered before. On the other hand, one can consider (5.79) as the GE with zero base mapping. Eventually, we certainly have in mind applications to more structured variational problems discussed in Sect. 1.3. For example, the variational inequality (VI)

$$x \in Q, \quad \langle \Phi(x),\, y - x\rangle \geq 0 \quad \forall\, y \in Q, \tag{5.80}$$

where $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ is continuous and monotone, and $Q \subset \mathbf{R}^n$ is a closed convex set, is equivalent to the GE (5.79) with the maximal monotone multifunction

$$\Psi(x) = \Phi(x) + N_Q(x), \quad x \in \mathbf{R}^n. \tag{5.81}$$

Given the current approximation $x^k \in \mathbf{R}^n$ to the solution of the GE (5.79), the (exact) *proximal point method* defines the next iterate x^{k+1} as the solution of the regularized GE

$$\Psi(x) + \nu_k(x - x^k) \ni 0, \tag{5.82}$$

where $\nu_k > 0$ is a regularization (or proximal) parameter. Since Ψ is maximal monotone, so is the multifunction in the left-hand side of (5.82); in addition, it is strongly monotone and it then holds that the solution to (5.82) exists and is unique.

The proximal point method dates back to [190, 236]. Its attractive feature relates to global convergence: starting from an arbitrary point, the method generates a sequence converging to some solution of the GE (5.79) provided one exists, and an unbounded sequence when there are no solutions. However, it is clear that the regularized subproblem (5.82), even if potentially better conditioned, is in principle as difficult as the original GE (5.79). Thus the proximal point method is a framework rather than a real computational algorithm. To be of any practical (or even relevant theoretical) use, either some specific structure of the multifunction Ψ has to be exploited, or subproblems (5.82) need to be solved only approximately. In fact, usually both of those features have to be appropriately combined. Some specific examples would be presented in what follows.

Note that generating x^{k+1} via (5.82) can be written as

$$v^{k+1} \in \Psi(x^{k+1}), \quad v^{k+1} + \nu_k(x^{k+1} - x^k) = 0.$$

The traditional approach to approximate solutions, dating back to [236], relaxes the second relation above, resulting in

$$v^{k+1} \in \Psi(x^{k+1}), \quad v^{k+1} + \nu_k(x^{k+1} - x^k) + \omega^k = 0,$$

where $\omega^k \in \mathbf{R}^n$ is the error term accounting for inexactness. Then for convergence it is asked that it holds, with some sequence $\{\delta_k\}$ of reals, that

$$\|\omega^k\| \leq \delta_k \quad \forall\, k = 0,\, 1,\, \ldots, \quad \sum_{k=0}^{\infty} \delta_k < \infty,$$

or

$$\|\omega^k\| \leq \delta_k \|x^{k+1} - x^k\| \quad \forall\, k = 0,\, 1,\, \ldots, \quad \sum_{k=0}^{\infty} \delta_k < \infty,$$

or some other related conditions involving exogenous summable sequences not related to the behavior of the algorithm at a given iteration in any constructive way.

A different approach of *relative errors* controlled at each iteration by the quantities generated by the algorithm is proposed in [252, 253, 255] and unified in [256]; its extension to the variable metric regularization is given in [216], applications to globalization of Newtonian methods in various contexts are developed in [251, 254, 257], and applications to decomposition in [183, 247]. Moreover, not only error terms of the type of ω^k above are allowed, but also the values of the multifunction Ψ can be evaluated inexactly. The latter turns out to be crucial for a number of applications, including for globalization of the Josephy–Newton method for variational inequalities in Sect. 5.4.3.

We next describe the inexact proximal point framework with relative errors following [256], simplified to the finite-dimensional setting.

Given $\varepsilon \geq 0$, the ε-*enlargement* of a monotone set-valued mapping Ψ (see [31, 32]) is given by

$$\Psi^{\varepsilon}(x) = \{v \in \mathbf{R}^n \mid \langle v - \tilde{v},\, x - \xi \rangle \geq -\varepsilon \ \ \forall\, \tilde{v} \in \Psi(\xi),\ \forall\, \xi \in \operatorname{dom} \Psi\}, \quad x \in \mathbf{R}^n. \tag{5.83}$$

For Ψ maximal monotone, it holds that $\Psi^0(x) = \Psi(x)$ for all $x \in \mathbf{R}^n$. Furthermore, the relation $\Psi(x) \subset \Psi^{\varepsilon}(x)$ holds for any $\varepsilon \geq 0$ and any $x \in \mathbf{R}^n$. Hence, Ψ^{ε} is a certain outer approximation of Ψ. We refer the reader to [30] for a comprehensive treatment of such enlargements.

Given any $\tilde{x} \in \mathbf{R}^n$ and $\nu > 0$, consider the proximal point problem of solving

$$\Psi(x) + \nu(x - \tilde{x}) \ni 0 \tag{5.84}$$

in the variable $x \in \mathbf{R}^n$, or equivalently, the system of relations

$$v \in \Psi(x), \quad v + \nu(x - \tilde{x}) = 0 \tag{5.85}$$

in the variables $(x,\, v) \in \mathbf{R}^n \times \mathbf{R}^n$. First, the approach is to relax *both* the inclusion and the equation in (5.85). Next, we require the “error terms” $\varepsilon \geq 0$ in the inclusion and $\omega \in \mathbf{R}^n$ in the equation to be small relative to the size of the element v and of the displacement $\nu(x - \tilde{x})$. Specifically, we call a triple $(x,\, v,\, \varepsilon) \in \mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R}_+$ an approximate solution of the proximal system (5.85), with error tolerance $\sigma \in [0,\, 1)$, if it holds that

$$\begin{gathered} v \in \Psi^{\varepsilon}(x), \quad v + \nu(x - \tilde{x}) + \omega = 0, \\ \varepsilon \geq 0, \quad \|\omega\|^2 + 2\nu\varepsilon \leq \sigma^2 \left(\|v\|^2 + \nu^2 \|x - \tilde{x}\|^2\right) \end{gathered} \tag{5.86}$$

with $\omega \in \mathbf{R}^n$ defined by the second relation.

Observe that if $(x,\, v)$ is the exact solution of (5.85), then $(x,\, v,\, 0)$ satisfies the approximation criteria (5.86) for any $\sigma \in [0,\, 1)$. Conversely, if the relaxation parameter σ is set to zero, then (5.86) induces the exact solution to (5.85), which is natural. At the same time, for $\sigma \in (0,\, 1)$ there are typically many approximate solutions to (5.85) in the sense of (5.86). The relative-error nature of the approximation rule (5.86) can be especially easily observed considering its somewhat more restrictive version given by

$$v \in \Psi(x), \quad v + \nu(x - \tilde{x}) + \omega = 0, \quad \|\omega\| \leq \sigma \max\{\|v\|,\, \nu\|x - \tilde{x}\|\},$$

which was the original proposal in [253]. We emphasize that in the framework of (5.86) the value of the relative error parameter σ need not approach zero in the course of iterations; it can be fixed at any positive value less than 1.

Thus the relative error can be kept bounded away from zero, which is precisely the numerically desirable approach and the motivation for this development.

We next establish some basic properties of approximate solutions defined above. The following result gives a further insight into (5.86) by showing that the quantity in the left-hand side in the inequality therein provides an *explicit global bound* for the distance to the exact solution of the proximal system (5.85).

Lemma 5.25. *Let Ψ be a maximal monotone multifunction from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$. Let $\tilde{x} \in \mathbf{R}^n$ and $\nu > 0$ be arbitrary, and let $(\hat{x}, \hat{v}) \in \mathbf{R}^n \times \mathbf{R}^n$ be the solution of the system (5.85).*

Then for any $\varepsilon \geq 0$, any $x \in \mathbf{R}^n$, and any $v \in \Psi^\varepsilon(x)$, where Ψ^ε is defined according to (5.83), it holds that

$$\|v - \hat{v}\|^2 + \nu^2\|x - \hat{x}\|^2 \leq \|v + \nu(x - \tilde{x})\|^2 + 2\nu\varepsilon.$$

Proof. Since $\hat{v} + \nu(\hat{x} - \tilde{x}) = 0$ and $\hat{v} \in \Psi(\hat{x})$, we have that

$$\begin{aligned}
\|v + \nu(x - \tilde{x})\|^2 &= \|v + \nu(x - \tilde{x}) - (\hat{v} + \nu(\hat{x} - \tilde{x}))\|^2 \\
&= \|v - \hat{v}\|^2 + \nu^2\|x - \hat{x}\|^2 + 2\nu\langle v - \hat{v},\, x - \hat{x}\rangle \\
&\geq \|v - \hat{v}\|^2 + \nu^2\|x - \hat{x}\|^2 - 2\nu\varepsilon,
\end{aligned}$$

where the definition (5.83) of Ψ^ε was used for the inequality. $\square$

Next, re-writing the relations in (5.86), we obtain that

$$\begin{aligned}
\sigma^2(\|v\|^2 + \nu^2\|x - \tilde{x}\|^2) &\geq \|v + \nu(x - \tilde{x})\|^2 + 2\nu\varepsilon \\
&= \|v\|^2 + \nu^2\|x - \tilde{x}\|^2 + 2\nu(\langle v,\, x - \tilde{x}\rangle + \varepsilon),
\end{aligned}$$

which rearranging the terms shows that (5.86) is equivalent to

$$v \in \Psi^\varepsilon(x), \quad \varepsilon \geq 0, \quad \langle v,\, \tilde{x} - x\rangle - \varepsilon \geq \frac{1 - \sigma^2}{2\nu}(\|v\|^2 + \nu^2\|x - \tilde{x}\|^2). \tag{5.87}$$

Proposition 5.26. *Let Ψ be a maximal monotone multifunction from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$. Let $\tilde{x} \in \mathbf{R}^n$ and $\nu > 0$ be arbitrary, and let $(x, v, \varepsilon) \in \mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R}_+$ satisfy the conditions in (5.86) for some $\sigma \in [0, 1)$, where Ψ^ε is defined according to (5.83).*

Then it holds that

$$\frac{1 - \sqrt{1 - (1 - \sigma^2)^2}}{1 - \sigma^2}\|v\| \leq \nu\|x - \tilde{x}\| \leq \frac{1 + \sqrt{1 - (1 - \sigma^2)^2}}{1 - \sigma^2}\|v\|. \tag{5.88}$$

Furthermore, the three conditions $\Psi(\tilde{x}) \ni 0$, $v = 0$, and $x = \tilde{x}$, are equivalent and imply $\varepsilon = 0$.

Proof. As (5.86) is equivalent to (5.87), using the Cauchy–Schwarz inequality, we have that

$$\begin{aligned}\nu\|v\|\|x-\tilde{x}\| &\geq \nu(\langle v,\, \tilde{x}-x\rangle - \varepsilon)\\ &\geq \frac{1-\sigma^2}{2}(\|v\|^2+\nu^2\|x-\tilde{x}\|^2).\end{aligned}$$

Denoting $t = \nu\|x-\tilde{x}\|$, and resolving the quadratic inequality in t

$$t^2 - \frac{2\|v\|}{1-\sigma^2}t + \|v\|^2 \leq 0,$$

we obtain (5.88).

Suppose now that $\Psi(\tilde{x}) \ni 0$. Since $v \in \Psi^{\varepsilon}(x)$, we have that

$$-\varepsilon \leq \langle v-0,\, x-\tilde{x}\rangle = \langle v,\, x-\tilde{x}\rangle, \tag{5.89}$$

and it follows that the left-hand side in (5.87) is nonpositive. Therefore, the right-hand side in (5.87) is zero. Hence, $v = 0$ and $x = \tilde{x}$. On the other hand, (5.88) implies that if $v = 0$, then $x = \tilde{x}$, and vice versa. Also, these conditions obviously imply, by (5.86), that $\varepsilon = 0$ and $\Psi(\tilde{x}) \ni 0$. $\square$

The result just established shows, in particular, that if $\tilde{x} \in \mathbf{R}^n$ is not a solution of the GE (5.79), i.e., $\Psi(\tilde{x}) \not\ni 0$, then for any $(x,\, v,\, \varepsilon)$ satisfying (5.86) it holds that $v \neq 0$ and $x \neq \tilde{x}$, so that (5.87) implies that

$$\langle v,\, \tilde{x}-x\rangle - \varepsilon > 0.$$

On the other hand, if $\bar{x} \in \mathbf{R}^n$ is a solution of the GE (5.79), then, similarly to (5.89), we have the inequality

$$\langle v,\, \bar{x}-x\rangle - \varepsilon \leq 0. \tag{5.90}$$

Geometrically these properties mean that the point $\tilde{x}$ is separated from the solution set of the GE (5.79) by the hyperplane $\{y \in \mathbf{R}^n \mid \langle v,\, y-x\rangle = \varepsilon\}$. It follows that projecting onto this hyperplane (whether exactly or with over- or under-relaxation) brings the point $\tilde{x}$ closer to the solution set of the GE (5.79). Such a projection, of course, is a computationally explicit operation. The next result makes these considerations precise.

Lemma 5.27. *Let Ψ be a maximal monotone multifunction from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$. Let $\tilde{x} \in \mathbf{R}^n$ be arbitrary and suppose that*

$$\langle v,\, \tilde{x}-x\rangle - \varepsilon > 0, \tag{5.91}$$

where $x \in \mathbf{R}^n$, $\varepsilon \geq 0$, and $v \in \Psi^{\varepsilon}(x)$, with Ψ^{ε} defined according to (5.83).

Then for any solution $\bar{x}$ of the GE (5.79) and any $\beta > 0$ it holds that

$$\|\hat{x}-\bar{x}\|^2 \leq \|\tilde{x}-\bar{x}\|^2 - t^2\beta(2-\beta)\|v\|^2,$$

where

$$\hat{x} = \tilde{x} - t\beta v, \quad t = \frac{\langle v,\, \tilde{x} - x\rangle - \varepsilon}{\|v\|^2}. \tag{5.92}$$

Proof. Define the closed halfspace

$$H = \{y \in \mathbf{R}^n \mid \langle v,\, y - x\rangle \le \varepsilon\}.$$

Observe that according to (5.90), $\bar{x} \in H$, while by assumption (5.91), $\tilde{x} \notin H$.

It is well known (and easy to check) that the projection of $\tilde{x}$ onto H is given by

$$\pi_H(\tilde{x}) = \tilde{x} - tv,$$

where the quantity t is defined in (5.92). Since $\bar{x} \in H$, by the basic properties of the projection (Lemma A.12), we obtain that

$$\begin{aligned} 0 &\ge \langle \tilde{x} - \pi_H(\tilde{x}),\, \bar{x} - \pi_H(\tilde{x})\rangle \\ &= \langle tv,\, \bar{x} - \pi_H(\tilde{x})\rangle \\ &= \frac{1}{\beta}\langle \tilde{x} - \hat{x},\, \bar{x} - \pi_H(\tilde{x})\rangle, \end{aligned}$$

where (5.92) was also used. Since $\beta > 0$, the last relation implies that

$$\langle \bar{x} - \pi_H(\tilde{x}),\, \hat{x} - \tilde{x}\rangle \ge 0.$$

Using the latter inequality and again (5.92), we derive that

$$\begin{aligned} \|\tilde{x} - \bar{x}\|^2 &= \|\bar{x} - \hat{x}\|^2 + \|\hat{x} - \tilde{x}\|^2 + 2\langle \bar{x} - \hat{x},\, \hat{x} - \tilde{x}\rangle \\ &= \|\bar{x} - \hat{x}\|^2 + \|\hat{x} - \tilde{x}\|^2 \\ &\quad +2\langle \pi_H(\tilde{x}) - \hat{x},\, \hat{x} - \tilde{x}\rangle + 2\langle \bar{x} - \pi_H(\tilde{x}),\, \hat{x} - \tilde{x}\rangle \\ &\ge \|\bar{x} - \hat{x}\|^2 + \|\hat{x} - \tilde{x}\|^2 + 2\langle \pi_H(\tilde{x}) - \hat{x},\, \hat{x} - \tilde{x}\rangle \\ &= \|\bar{x} - \hat{x}\|^2 - \|\hat{x} - \tilde{x}\|^2 - 2t\langle v,\, \hat{x} - \tilde{x}\rangle \\ &= \|\bar{x} - \hat{x}\|^2 + t^2\beta(2 - \beta)\|v\|^2, \end{aligned}$$

which completes the proof. □

Lemma 5.27 shows that if we take $\beta \in (0,\, 2)$, then the point $\hat{x}$ computed according to (5.92) is closer to the solution set of the GE (5.79) than the point $\tilde{x}$. This is the basis for constructing a convergent algorithm based on inexact solutions of proximal point subproblems. The basic value $\beta = 1$ corresponds to the projection onto the separating hyperplane, while the other values in the interval $(0,\, 2)$ give steps with over- or under-relaxation.

Algorithm 5.28 Choose any $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. Choose $\nu_k > 0$ and $\sigma_k \in [0,\, 1)$, and find $(y^k,\, v^k,\, \varepsilon_k) \in \mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R}_+$, an approximate solution of the proximal system (5.85) with $\tilde{x} = x^k$ and $\nu = \nu_k$ in the sense of (5.86), i.e., a triple satisfying

$$v^k \in \Psi^{\varepsilon_k}(y^k), \quad v^k + \nu_k(y^k - x^k) + \omega^k = 0,$$
$$\varepsilon_k \geq 0, \quad \|\omega^k\|^2 + 2\nu_k\varepsilon_k \leq \sigma_k^2(\|v^k\|^2 + \nu_k^2\|y^k - x^k\|^2),$$

where Ψ^ε is defined according to (5.83).
2. If $y^k = x^k$, stop.
3. Choose $\beta_k \in (0, 2)$ and compute

$$x^{k+1} = x^k - t_k\beta_k v^k, \quad t_k = \frac{\langle v^k, x^k - y^k\rangle - \varepsilon_k}{\|v^k\|^2}. \tag{5.93}$$

4. Increase k by 1 and go to step 1.

Note that the algorithm stops when $y^k = x^k$ which, by Proposition 5.26, means that x^k is a solution of the GE (5.79). As usual, in our convergence analysis we assume that this does not occur, and so an infinite sequence of iterates is generated.

In [256], it is shown that if the tighter approximation condition is used in step 1, namely with the last inequality replaced by

$$\|\omega^k\|^2 + 2\nu_k\varepsilon_k \leq (\sigma_k\nu_k\|y^k - x^k\|)^2,$$

then the following simpler iterates update is a special case of (5.93):

$$x^{k+1} = x^k - \frac{1}{\nu_k}v^k.$$

The latter is the version of the method introduced in [252].

Another issue important to mention is that, unlike in the case of summable approximation errors in the solution of proximal subproblems, one cannot simply accept the point y^k obtained with the relative-error condition as the next iterate x^{k+1}, i.e., the "correction" step (5.93) in Algorithm 5.28 is needed in general (but note that this step is of negligible computational cost, as it is given by an explicit formula). To see this, consider the following simple example [253].

Example 5.29. Let $n = 2$ and $\Psi(x) = Ax$, $x \in \mathbf{R}^n$, where

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Then Ψ is a linear mapping with a skew-symmetric matrix; obviously, it is maximal monotone. The GE (5.79) becomes a system of linear equations with the unique solution $\bar{x} = 0$.

Take $x^k = (1, 0)$ and $\nu_k = 1$. It is easy to check that the point $y^k = (0, 1)$ is admissible according to the conditions in step 1 of Algorithm 5.28. Indeed, taking $v^k = \Psi(y^k)$ and $\varepsilon_k = 0$,

$$0 = v^k + \nu_k(y^k - x^k) + \omega^k = (1, 0) + (-1, 1) + \omega^k$$

holds with $\omega^k = (0, -1)$. So we have that $\|\omega^k\| = 1$, $\|y^k - x^k\| = \sqrt{2}$. The approximation condition is satisfied with, say, $\sigma_k = 1/\sqrt{2}$. But since $\|y^k\| = \|x^k\| = 1$, no progress is being made towards $\bar{x} = 0$, the solution of the problem, if we accept y^k as the next iterate x^{k+1}. In fact, taking $\sigma_k = \sigma \in (1/\sqrt{2}, 1)$ fixed for all k, it is even possible to construct a sequence satisfying the approximation criteria at every iteration and diverging if we take $x^{k+1} = y^k$ for each k.

On the other hand, if the projection step 3 of Algorithm 5.28 is added after y^k is computed, it can be seen that we obtain the solution in the example under consideration in just one step, if the basic value $\beta_k = 1$ is used.

We proceed with global convergence analysis of Algorithm 5.28.

Theorem 5.30. *Let Ψ be a maximal monotone multifunction from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$, and let the solution set of the GE (5.79) be nonempty. Let the constants $\hat{\nu} > 0$, $\hat{\sigma} \in [0, 1)$, $\bar{\beta}$, $\hat{\beta} \in (0, 2)$, $\bar{\beta} \le \hat{\beta}$, be such that the choices of the parameters in Algorithm 5.28 satisfy, for all k large enough, the bounds*

$$0 < \nu_k \le \hat{\nu}, \quad 0 \le \sigma_k \le \hat{\sigma}, \quad \bar{\beta} \le \beta_k \le \hat{\beta}.$$

Then for any starting point $x^0 \in \mathbf{R}^n$ Algorithm 5.28 generates an iterative sequence $\{x^k\}$ converging to some solution of the GE (5.79).

Proof. As we assume that the algorithm does not terminate finitely, by Proposition 5.26 we have that $y^k \neq x^k$ and $v^k \neq 0$ for all k, and

$$\langle v^k, x^k - y^k \rangle - \varepsilon_k \ge \frac{1 - \sigma_k^2}{2\nu_k} (\|v^k\|^2 + \nu_k^2 \|y^k - x^k\|^2) > 0. \tag{5.94}$$

Therefore, by the definition of t_k in (5.93),

$$t_k \ge \frac{(1 - \sigma_k^2)(\|v^k\|^2 + \nu_k^2 \|y^k - x^k\|^2)}{2\nu_k \|v^k\|^2}. \tag{5.95}$$

Furthermore, using the fact that $\|v^k\|^2 + \nu_k^2 \|y^k - x^k\|^2 \ge 2\nu_k \|v^k\| \|y^k - x^k\|$, we obtain that

$$t_k \|v^k\| \ge (1 - \sigma_k^2) \|y^k - x^k\|. \tag{5.96}$$

Using (5.94) and Lemma 5.27, we have that for any solution $\bar{x}$ of the GE (5.79), and for all k large enough, it holds that

$$\begin{aligned} \|x^{k+1} - \bar{x}\|^2 &\le \|x^k - \bar{x}\|^2 - t_k^2 \beta_k (2 - \beta_k) \|v^k\|^2 \\ &\le \|x^k - \bar{x}\|^2 - t_k^2 \hat{\beta} (2 - \bar{\beta}) \|v^k\|^2. \end{aligned} \tag{5.97}$$

In particular, the sequence $\{\|x^k - \bar{x}\|\}$ is nonincreasing and thus convergent. It immediately follows that the sequence $\{x^k\}$ is bounded, and also that

$$\sum_{k=0}^{\infty} t_k^2 \|v^k\|^2 < \infty.$$

Therefore, we also have that $\lim_{k\to\infty} t_k\|v^k\| = 0$. Then (5.96) implies the relation $\lim_{k\to\infty} \|y^k - x^k\| = 0$. From (5.95) it also easily follows that

$$t_k\|v^k\| \geq \frac{1-\sigma_k^2}{2\nu_k}\|v^k\| \geq \frac{1-\hat{\sigma}^2}{2\hat{\nu}}\|v^k\|,$$

so that $\lim_{k\to\infty} \|v^k\| = 0$. Since $\varepsilon_k \geq 0$, relation (5.94) now implies that $\lim_{k\to\infty} \varepsilon_k = 0$.

As the sequence $\{x^k\}$ is bounded, it has at least one accumulation point, say $\bar{x}$. Let $\{x^{k_j}\}$ be some subsequence converging to $\bar{x}$.

Since $\|y^k - x^k\| \to 0$ as $k \to \infty$, it follows that the subsequence $\{y^{k_j}\}$ also converges to $\bar{x}$. Moreover, $\{v^{k_j}\} \to 0$ and $\varepsilon_{k_j} \to 0$ as $j \to \infty$, as established above.

Take any $x \in \mathbf{R}^n$ and $v \in \Psi(x)$. Because $v^k \in \Psi^{\varepsilon_k}(y^k)$, for any index j it holds that

$$\langle v - v^{k_j},\, x - y^{k_j}\rangle \geq -\varepsilon_{k_j}.$$

Therefore,

$$\langle v,\, x - y^{k_j}\rangle \geq \langle v^{k_j},\, x - y^{k_j}\rangle - \varepsilon_{k_j}.$$

Since $\{y^{k_j}\}$ converges to $\bar{x}$, $\{v^{k_j}\}$ converges to zero, and $\{\varepsilon_{k_j}\}$ converges to zero, taking the limit as $j \to \infty$ in the above inequality we obtain that

$$\langle v - 0,\, x - \bar{x}\rangle \geq 0.$$

Since $(x,\, v)$ was taken as an arbitrary element in the graph of Ψ, from maximality of Ψ it follows that $0 \in \Psi(\bar{x})$, i.e., $\bar{x}$ is a solution of the GE (5.79).

The argument above shows that every accumulation point of $\{x^k\}$ is a solution. We now show that the sequence converges, i.e., the accumulation point is unique. Let $\bar{x}^1$ and $\bar{x}^2$ be two accumulation points of $\{x^k\}$. As established above, $\bar{x}^1$ and $\bar{x}^2$ are solutions of the GE (5.79), and there exist $\chi_1 = \lim_{k\to\infty} \|x^k - \bar{x}^1\|^2$ and $\chi_2 = \lim_{k\to\infty} \|x^k - \bar{x}^2\|^2$.

Let $\{x^{k_i}\}$ and $\{x^{k_j}\}$ be two subsequences of $\{x^k\}$ convergent to $\bar{x}^1$ and $\bar{x}^2$, respectively. We have that

$$\|x^{k_j} - \bar{x}^1\|^2 = \|x^{k_j} - \bar{x}^2\|^2 + \|\bar{x}^2 - \bar{x}^1\|^2 + 2\langle x^{k_j} - \bar{x}^2,\, \bar{x}^2 - \bar{x}^1\rangle,$$

$$\|x^{k_i} - \bar{x}^2\|^2 = \|x^{k_i} - \bar{x}^1\|^2 + \|\bar{x}^1 - \bar{x}^2\|^2 + 2\langle x^{k_i} - \bar{x}^1,\, \bar{x}^1 - \bar{x}^2\rangle.$$

Taking the limits in the above two relations as $j \to \infty$ and $i \to \infty$, the inner products in the right-hand sides converge to zero because $\bar{x}^1$ and $\bar{x}^2$ are the limits of $\{x^{k_i}\}$ and $\{x^{k_j}\}$, respectively. Therefore, using the definitions of χ_1 and χ_2, we obtain that

$$\chi_1 = \chi_2 + \|\bar{x}^1 - \bar{x}^2\|^2, \quad \chi_2 = \chi_1 + \|\bar{x}^1 - \bar{x}^2\|^2.$$

This implies that $\bar{x}^1 = \bar{x}^2$, i.e., the accumulation point is unique. □

When the GE (5.79) has no solutions, the sequence $\{x^k\}$ generated by Algorithm 5.28 can be shown to be unbounded.

We next prove that Algorithm 5.28 has the linear and superlinear local convergence rates under some natural assumptions. Specifically, as in Sect. 3.1, we shall assume that the GE (5.79) has a solution $\bar{x} \in \mathbf{R}^n$ which is semistable. According to Definition 1.29, the latter means that there exists $\ell > 0$ such that for any $v \in \mathbf{R}^n$ any solution $x(v)$ of the perturbed GE

$$\Psi(x) \ni v, \tag{5.98}$$

which is close enough to $\bar{x}$, satisfies the estimate

$$\|x(v) - \bar{x}\| \leq \ell \|v\|. \tag{5.99}$$

In particular, this implies that $\bar{x}$ is an isolated solution. Hence, under the maximal monotonicity of Ψ this solution is unique (because the solution set of the GE (5.79) is convex in that case). Recall that, as discussed in Sect. 1.3, semistability is implied by strong regularity. Strong regularity, in turn, is automatic when Ψ is maximal monotone and strongly monotone with constant $\gamma > 0$: in that case, (5.99) holds with $\ell = 1/\gamma$ for the unique solution $x(v)$ of (5.98) for any $v \in \mathbf{R}^n$.

It should be noted that the following weaker condition can be used instead of (5.99):

$$\operatorname{dist}(x(v), \bar{X}) \leq \ell \|v\|,$$

where $\bar{X}$ is the solution set of the GE (5.79); see, e.g., [216, 253]. This condition does not presume uniqueness of the solution. Here, we shall employ the stronger (5.99), as regularity conditions needed for fast local convergence of globalizations of Newtonian methods in Sects. 5.4.2, 5.4.3 imply uniqueness anyway.

We shall assume, for the sake of simplicity, that $\beta_k = 1$ for all k, i.e., there is no relaxation in the "projection" step (5.93) of Algorithm 5.28.

Theorem 5.31. *In addition to the assumptions of Theorem 5.30, suppose that the GE (5.79) has the (necessarily unique) semistable solution $\bar{x} \in \mathbf{R}^n$, i.e., there exists $\ell > 0$ such that for any $v \in \mathbf{R}^n$ any solution $x(v)$ of the GE (5.98), which is close enough $\bar{x}$, satisfies the estimate (5.99). Suppose also that in Algorithm 5.28 we take $\beta_k = 1$ for all k.*

Then for any $x^0 \in \mathbf{R}^n$ the sequence $\{x^k\}$ generated by Algorithm 5.28 converges to $\bar{x}$, and for all k sufficiently large it holds that

$$\|x^{k+1} - \bar{x}\| \leq \frac{c_k}{\sqrt{1 + c_k^2}} \|x^k - \bar{x}\|, \tag{5.100}$$

where

$$c_k = \sqrt{\left(\frac{\ell^2\nu_k^2(1+\sqrt{1-(1-\sigma_k^2)^2})^2}{(1-\sigma_k^2)^2}+1\right)\left(\frac{1}{(1-\sigma_k^2)^2}-1\right)} + \frac{\ell\nu_k(1+\sqrt{1-(1-\sigma_k^2)^2})}{(1-\sigma_k^2)^2}.$$

In particular, the rate of convergence is linear, and it is superlinear if $\nu_k \to 0$ and $\sigma_k \to 0$ as $k \to \infty$.

Proof. Theorem 5.30 implies that $\{x^k\}$ is well defined and converges to $\bar{x}$. Using the first inequality in (5.97) with $\beta_k = 1$, we have that

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - t_k^2\|v^k\|^2, \tag{5.101}$$

where $t_k > 0$ is defined in (5.93) of Algorithm 5.28. We therefore need to estimate the last term above. Observe that similarly to the proof of Theorem 5.30, estimate (5.101) implies that

$$\lim_{k\to\infty} t_k\|v^k\| = 0. \tag{5.102}$$

For each k, consider the auxiliary proximal subproblem

$$\Psi(x) + \frac{1}{t_k}(x - x^k) \ni 0.$$

Let $\hat{x}^k \in \mathbf{R}^n$ be its exact solution, i.e., there exists $\hat{v}^k \in \mathbf{R}^n$ such that

$$\hat{v}^k \in \Psi(\hat{x}^k), \quad \hat{v}^k + \frac{1}{t_k}(\hat{x}^k - x^k) = 0. \tag{5.103}$$

Since according to Algorithm 5.28 we have that $v^k \in \Psi^{\varepsilon_k}(y^k)$, by Lemma 5.25 it follows that

$$\begin{aligned}\frac{1}{t_k^2}\|\hat{x}^k - y^k\|^2 + \|\hat{v}^k - v^k\|^2 &\leq \left\|v^k + \frac{1}{t_k}(y^k - x^k)\right\|^2 + \frac{2\varepsilon_k}{t_k}\\ &= \left\|v^k + \frac{1}{t_k}(y^k - x^k)\right\|^2\\ &\quad + \frac{2}{t_k}(\langle v^k,\, x^k - y^k\rangle - t_k\|v^k\|^2)\\ &= \frac{1}{t_k^2}\|y^k - x^k\|^2 - \|v^k\|^2.\end{aligned}$$

Now using (5.96), the latter implies that

$$\frac{1}{t_k^2}\|\hat{x}^k - y^k\|^2 + \|\hat{v}^k - v^k\|^2 \leq \left(\frac{1}{(1-\sigma_k^2)^2} - 1\right)\|v^k\|^2. \tag{5.104}$$

Note that since by (5.93) and (5.103) we have that $x^{k+1} = x^k - t_k v^k$ (recall that $\beta_k = 1$) and $\hat{x}^k = x^k - t_k \hat{v}^k$, it holds that

$$\|\hat{v}^k - v^k\| = \frac{1}{t_k}\|\hat{x}^k - x^{k+1}\|.$$

Using this relation and (5.104), we obtain that

$$\|\hat{x}^k - y^k\|^2 + \|\hat{x}^k - x^{k+1}\|^2 \leq \left(\frac{1}{(1-\sigma_k^2)^2} - 1\right) t_k^2\|v^k\|^2. \tag{5.105}$$

Since $\sigma_k \leq \hat{\sigma} < 1$ for all k, it follows from (5.102) and (5.105) that $\{\hat{x}^k\}$ converges to the same limit $\bar{x}$ as $\{x^k\}$. Therefore, by (5.99) we conclude that for all k large enough

$$\|\hat{x}^k - \bar{x}\| \leq \ell\|\hat{v}^k\| = \frac{\ell}{t_k}\|\hat{x}^k - x^k\|,$$

where we have used (5.103). In the sequel, we assume that k is large enough, so that the above bound holds. We then further obtain

$$\begin{aligned}
\|x^{k+1} - \bar{x}\| &\leq \|\hat{x}^k - \bar{x}\| + \|\hat{x}^k - x^{k+1}\| \\
&\leq \frac{\ell}{t_k}\|\hat{x}^k - x^k\| + \|\hat{x}^k - x^{k+1}\| \\
&\leq \frac{\ell}{t_k}\|\hat{x}^k - y^k\| + \frac{\ell}{t_k}\|y^k - x^k\| + \|\hat{x}^k - x^{k+1}\|.
\end{aligned} \tag{5.106}$$

Using the Cauchy–Schwarz inequality, it follows that

$$\begin{aligned}
\frac{\ell}{t_k}\|\hat{x}^k - y^k\| + \|\hat{x}^k - x^{k+1}\| &\leq \sqrt{\left(\frac{\ell^2}{t_k^2} + 1\right)\left(\|\hat{x}^k - y^k\|^2 + \|\hat{x}^k - x^{k+1}\|^2\right)} \\
&\leq t_k\sqrt{\left(\frac{\ell^2}{t_k^2} + 1\right)\left(\frac{1}{(1-\sigma_k^2)^2} - 1\right)}\|v^k\|,
\end{aligned}$$

where the second inequality is by (5.105). Using the latter relation in (5.106) yields

$$\begin{aligned}
\|x^{k+1} - \bar{x}\| &\leq t_k\sqrt{\left(\frac{\ell^2}{t_k^2} + 1\right)\left(\frac{1}{(1-\sigma_k^2)^2} - 1\right)}\|v^k\| + \frac{\ell}{t_k}\|y^k - x^k\| \\
&\leq \left(\sqrt{\left(\frac{\ell^2}{t_k^2} + 1\right)\left(\frac{1}{(1-\sigma_k^2)^2} - 1\right)} + \frac{\ell}{t_k(1-\sigma_k^2)}\right) t_k\|v^k\|,
\end{aligned} \tag{5.107}$$

where the second inequality is by (5.96).

From (5.95), we obtain that

$$t_k \geq \frac{1-\sigma_k^2}{2\nu_k}\left(1+\frac{\nu_k^2\|y^k-x^k\|^2}{\|v^k\|^2}\right) \geq \frac{1-\sigma_k^2}{\nu_k(1+\sqrt{1-(1-\sigma_k^2)^2})}, \qquad (5.108)$$

where (5.88) and some algebraic transformations have been used for the second inequality.

Using now (5.108) in (5.107), and recalling the definition of c_k in the statement of the theorem, we conclude that

$$\|x^{k+1}-\bar{x}\| \leq c_k t_k \|v^k\|.$$

Then (5.101) yields

$$\|x^{k+1}-\bar{x}\|^2 \leq \|x^k-\bar{x}\|^2 - \frac{1}{c_k^2}\|x^{k+1}-\bar{x}\|^2,$$

and hence, (5.100) holds.

It remains to observe that $\nu_k \leq \hat{\nu}$ and $\sigma_k \leq \hat{\sigma} < 1$ imply that the sequence $\{c_k\}$ is bounded above, so that the estimate (5.100) shows the linear rate of convergence of $\{x^k\}$ to $\bar{x}$. It can also be easily seen that $\nu_k \to 0$ and $\sigma_k \to 0$ imply $c_k \to 0$ as $k \to \infty$, which gives the superlinear convergence rate. □

5.4.2 Globalized Newton Method for Monotone Equations

Consider the system of equations

$$\Phi(x) = 0, \qquad (5.109)$$

where the (single-valued) mapping $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ is monotone. For global convergence of the method to be presented it is enough for Φ to be continuous; for the superlinear convergence rate Φ would be required to be differentiable with locally Lipschitz-continuous derivative.

Note first that if a Newton-type method for (5.109) is globalized using linesearch for the natural merit function $\varphi : \mathbf{R}^n \to \mathbf{R}$, $\varphi(x) = \|\Phi(x)\|^2$, one can expect convergence to local minimizers or even merely stationary points of φ (see Sect. 5.1.1). Under the only assumption of monotonicity, these need not be global minimizers of φ, i.e., solutions of (5.109). In fact, even to ensure existence of accumulation points for the generated sequence additional assumptions are needed, implying boundedness of a suitable level set of φ, which in turn requires boundedness of the solution set of (5.109). Moreover, even if accumulation points exist, one cannot prove convergence of the whole sequence without additional assumptions. In the monotone case under consideration, there is a theoretically better alternative which does not have any of the drawbacks discussed above.

Here we follow the development of [251]. Note first that the mapping Φ is obviously maximal monotone. Hence, according to Theorem 5.30, the inexact proximal point method of Sect. 5.4.1 applied to (5.109) is convergent. The key idea with respect to globalizing the Newton method is that, under appropriate assumptions, just *one* Newton iteration applied to the proximal subproblem is enough to solve it within the relative-error tolerance specified in Sect. 5.4.1 (moreover, even the Newton step itself can be computed approximately). Then convergence and rate of convergence of the Newton method would essentially follow from the properties of the inexact proximal point scheme of Sect. 5.4.1.

We now describe the algorithm. Given the current iterate $x^k \in \mathbf{R}^n$ and a proximal regularization parameter $\nu_k > 0$, the proximal point subproblem for (5.109) is given by

$$\Phi(x) + \nu_k(x - x^k) = 0. \tag{5.110}$$

The basic ingredient of the iteration consists in solving the Newtonian linearization of this subproblem at the point x^k, i.e., the linear system

$$\Phi(x^k) + J_k(x - x^k) + \nu_k(x - x^k) = 0, \tag{5.111}$$

where J_k is a positive semidefinite matrix. Positive semidefiniteness of J_k is natural in the monotone setting, since if Φ is continuously differentiable on $\mathbf{R}^n$, then $J = \Phi'(x)$ is positive semidefinite for all $x \in \mathbf{R}^n$ (see Proposition A.17). We allow (5.111) to be solved approximately, in the spirit of truncated Newton methods in Sect. 2.1.1. Once an approximate solution of (5.111) is computed, we check whether the obtained step gives the separation property for x^k from the solution set of (5.109) (recall Lemma 5.27, which was the key for the inexact proximal point framework in Sect. 5.4.1). If the full step fails to give separation, then an Armijo-type linesearch procedure is activated to find a suitable point in the computed Newtonian direction. Once the separating hyperplane is obtained, the next iterate x^{k+1} is computed by projecting x^k onto it. We now formally state the algorithm.

Algorithm 5.32 Choose the parameters $\sigma \in (0, 1)$ and $\theta \in (0, 1)$. Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $\Phi(x^k) = 0$, stop.
2. Choose a positive semidefinite matrix J_k, $\nu_k > 0$ and $\rho_k \in [0, 1)$. Compute $p^k \in \mathbf{R}^n$ such that for some $w^k \in \mathbf{R}^n$ it holds that

$$\begin{gathered} \Phi(x^k) + (J_k + \nu_k I)p^k + w^k = 0, \\ \|w^k\| \le \rho_k \nu_k \|p^k\|. \end{gathered} \tag{5.112}$$

3. Set $\alpha = 1$. If the inequality

$$-\langle \Phi(x^k + \alpha p^k),\, p^k\rangle \ge \sigma(1 - \rho_k)\nu_k\|p^k\|^2 \tag{5.113}$$

is satisfied, set $\alpha_k = \alpha$. Otherwise, replace α by $\theta\alpha$, check (5.113) again, etc., until (5.113) becomes valid.

4. Set $y^k = x^k + \alpha_k p^k$.
5. Compute

$$x^{k+1} = x^k - \frac{\langle \Phi(y^k),\, x^k - y^k\rangle}{\|\Phi(y^k)\|^2}\Phi(y^k). \tag{5.114}$$

6. Increase k by 1 and go to step 1.

First note that since the matrix $J_k + \nu_k I$ is positive definite, (5.112) always has the exact (unique) solution corresponding to $w^k = 0$. Furthermore, it always holds that $p^k \neq 0$, since if it were the case that $p^k = 0$ it would follow that $w^k = 0$, and then also $\Phi(x^k) = 0$, so that the method would have stopped on step 1.

We next prove that for Algorithm 5.32 the whole sequence of iterates is globally convergent to a solution of (5.109) under no regularity assumptions at all. The solution need not be unique and the solution set can even be unbounded. Moreover, if the mapping Φ is differentiable with locally Lipschitz-continuous derivative, if at the limit $\bar{x}$ of the iterative sequence the Jacobian matrix $\Phi'(\bar{x})$ is nonsingular, and if the parameters of the algorithm are controlled by certain appropriately set rules, then the superlinear rate of convergence is obtained.

Theorem 5.33. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be continuous and monotone on $\mathbf{R}^n$, and let the solution set of (5.109) be nonempty. Let the constants $\hat{\rho} \in [0, 1)$, $\Gamma > 0$, and $c > 0$, $C > 0$, $c \leq C$, be such that the choices of parameters in Algorithm 5.32 satisfy, for all k large enough, the bounds*

$$0 \leq \rho_k \leq \hat{\rho}, \quad \|J_k\| \leq \Gamma, \quad 0 < c\|\Phi(x^k)\| \leq \nu_k \leq C.$$

Then for any starting point $x^0 \in \mathbf{R}^n$ Algorithm 5.32 generates an iterative sequence $\{x^k\}$ convergent to some solution of (5.109).

Proof. As already commented, Algorithm 5.32 always generates $p^k \neq 0$. We first show that the linesearch procedure in step 3 terminates with a positive stepsize α_k.

Suppose that for some iteration index k this is not the case. That is, no matter how small $\alpha > 0$ becomes, (5.113) is not satisfied. In other words, for all $j = 0, 1, \ldots$, we have that

$$-\langle \Phi(x^k + \theta^j p^k),\, p^k\rangle < \sigma(1 - \rho_k)\nu_k\|p^k\|^2. \tag{5.115}$$

Observe that

$$\begin{aligned} -\lim_{j\to\infty}\langle \Phi(x^k + \theta^j p^k),\, p^k\rangle &= -\langle \Phi(x^k),\, p^k\rangle \\ &= \langle (J_k + \nu_k I)p^k + w^k,\, p^k\rangle \\ &\geq \nu_k\|p^k\|^2 - \|w^k\|\|p^k\| \\ &\geq (1 - \rho_k)\nu_k\|p^k\|^2, \end{aligned}$$

where the second equality and the last inequality follow from (5.112), and the first inequality follows from the positive semidefiniteness of J_k and the Cauchy–Schwarz inequality. Now taking the limits as $j \to \infty$ in both sides of (5.115) implies that $\sigma \geq 1$, which contradicts the choice $\sigma \in (0, 1)$. It follows that the linesearch step is well defined. Then (5.113) also shows that $\Phi(y^k) \neq 0$, so that (5.114) is also well defined.

By (5.113) we then have that

$$\begin{aligned}\langle \Phi(y^k),\, x^k - y^k\rangle &= -\alpha_k \langle \Phi(y^k),\, p^k\rangle \\ &\geq \sigma(1-\rho_k)\nu_k\alpha_k\|p^k\|^2 > 0.\end{aligned} \tag{5.116}$$

Let $\bar{x}$ be any solution of (5.109). By (5.114), (5.116) and Lemma 5.92, it follows that

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2. \tag{5.117}$$

Hence, the sequence $\{\|x^k - \bar{x}\|\}$ is nonincreasing and convergent. Therefore, the sequence $\{x^k\}$ is bounded, and also

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0. \tag{5.118}$$

By (5.112), and by the choice of ν_k for k large enough, we have that

$$\begin{aligned}\|\Phi(x^k)\| &\geq \|(J_k + \nu_k I)p^k\| - \|w^k\| \\ &\geq (1-\rho_k)\nu_k\|p^k\| \\ &\geq (1-\rho_k)c\|\Phi(x^k)\|\|p^k\|.\end{aligned}$$

It follows that the sequence $\{p^k\}$ is bounded, and hence so is $\{y^k\}$. Now by the continuity of Φ and the condition $\rho_k \leq \hat{\rho} < 1$, there exists $\gamma > 0$ such that $\sigma(1-\rho_k)/\|\Phi(y^k)\| \geq \gamma$. Using (5.114) and (5.116), we thus obtain that

$$\|x^{k+1} - x^k\| = \frac{\langle \Phi(y^k),\, x^k - y^k\rangle}{\|\Phi(y^k)\|} \geq \gamma\nu_k\alpha_k\|p^k\|^2.$$

From the latter relation and (5.118) it follows that

$$\lim_{k\to\infty} \nu_k\alpha_k\|p^k\|^2 = 0. \tag{5.119}$$

We next consider the two possible cases:

$$\liminf_{k\to\infty} \|\Phi(x^k)\| = 0 \quad \text{or} \quad \liminf_{k\to\infty} \|\Phi(x^k)\| > 0. \tag{5.120}$$

In the first case, the continuity of Φ implies that the sequence $\{x^k\}$ has some accumulation point $\hat{x}$ such that $\Phi(\hat{x}) = 0$ (recall that $\{x^k\}$ is bounded). Since $\bar{x}$ was an *arbitrary* solution of (5.109), we can choose $\bar{x} = \hat{x}$ in (5.117). The sequence $\{\|x^k - \hat{x}\|\}$ converges, and since $\hat{x}$ is an accumulation point of $\{x^k\}$, it must be the case that $\{x^k\}$ converges to $\hat{x}$.

Consider now the second case in (5.120). From the choice of ν_k for k sufficiently large,

$$\liminf_{k\to\infty} \nu_k \geq c \liminf_{k\to\infty} \|\Phi(x^k)\| > 0.$$

Furthermore, by (5.112) it holds that

$$\begin{aligned}\|\Phi(x^k)\| &\leq \|(J_k + \nu_k I)p^k\| + \|w^k\| \\ &\leq (\|J_k\| + \nu_k + \rho_k\nu_k)\,\|p^k\| \\ &\leq (\Gamma + 2C)\|p^k\|.\end{aligned}$$

Hence,

$$\liminf_{k\to\infty} \|p^k\| > 0.$$

Then, by (5.119), it must hold that

$$\lim_{k\to\infty} \alpha_k = 0.$$

This means that the first trial stepsize value $\alpha = 1$ is always reduced at least once for all k large enough. In particular, the value $\alpha_k/\theta > \alpha_k$ did not satisfy (5.113), i.e.,

$$-\langle \Phi(x^k + \alpha_k p^k/\theta),\, p^k\rangle < \sigma(1 - \rho_k)\nu_k\|p^k\|^2.$$

Taking into account the boundedness of $\{x^k\}$, $\{p^k\}$, $\{\nu_k\}$ and $\{\rho_k\}$, and passing onto a subsequence if necessary, as $k \to \infty$ we obtain that

$$-\langle \Phi(\tilde{x}),\, \tilde{p}\rangle \leq \sigma(1 - \tilde{\rho})\tilde{\nu}\|\tilde{p}\|^2,$$

where $\tilde{x}$, $\tilde{p}$, $\tilde{\nu}$, and $\tilde{\rho}$, are the limits of the corresponding subsequences. On the other hand, by (5.112) and the already familiar argument,

$$-\langle \Phi(\tilde{x}),\, \tilde{p}\rangle \geq (1 - \tilde{\rho})\tilde{\nu}\|\tilde{p}\|^2.$$

Taking into account that $\tilde{\nu}\|\tilde{p}\| > 0$ and $\tilde{\rho} \leq \hat{\rho} < 1$, the last two relations are a contradiction because $\sigma \in (0, 1)$. Hence, the second case in (5.120) is not possible. This completes the proof. $\square$

We now turn our attention to the rate of convergence analysis in the differentiable case. We assume that at the limit $\bar{x}$ (convergence to which is already established in Theorem 5.33) the matrix $\Phi'(\bar{x})$ is positive definite, and therefore, $\bar{x}$ is a strongly regular (and hence semistable) solution of (5.109). Moreover, as discussed above, monotonicity of Φ implies that this solution is unique. We shall show that the unit stepsize is always accepted from some iteration on, and that the resulting point is a suitable approximate solution to the proximal point subproblem (5.110) in the sense of the relative-error measure of Sect. 5.4.1. With those facts in hand, the rate of convergence is immediate from Theorem 5.31. The parameters of the method have to be chosen as specified in (5.121) below. For example, the option

$$\nu_k = \max\{c\|\Phi(x^k)\|,\, \|\Phi(x^k)\|^\tau\}, \quad \rho_k = \min\{1/2,\, \|\Phi(x^k)\|\}$$

with some pre-fixed $c > 0$ and $\tau \in (0,\, 1/2)$ satisfies the assumptions of both Theorems 5.33 and 5.34.

Theorem 5.34. *In addition to the assumptions of Theorem 5.33, let Φ be differentiable near $\bar{x} \in \mathbf{R}^n$, with its derivative being locally Lipschitz-continuous at $\bar{x}$, and let $\bar{x}$ be the (necessarily unique) solution of (5.109) with $\Phi'(\bar{x})$ positive definite. Let Algorithm 5.32 use $J_k = \Phi'(x^k)$ for all k large enough, and let the parameters of the algorithm satisfy the following conditions:*

$$\lim_{k\to\infty} \nu_k = \lim_{k\to\infty} \rho_k = \lim_{k\to\infty} \frac{\|\Phi(x^k)\|}{\nu_k^2} = \lim_{k\to\infty} \frac{\rho_k}{\nu_k} = \lim_{k\to\infty} \frac{\rho_k^2}{\|\Phi(x^k)\|} = 0. \tag{5.121}$$

Then for any $x^0 \in \mathbf{R}^n$ the sequence $\{x^k\}$ generated by Algorithm 5.32 converges to $\bar{x}$, and the rate of convergence is superlinear.

Proof. By Theorem 5.33, we already know that the sequence $\{x^k\}$ converges to $\bar{x}$, and $\bar{x}$ is necessarily semistable. Then since $J_k = \Phi(x^k)$ for all k large enough, the positive definiteness of $\Phi'(\bar{x})$ implies that there exists a constant $\gamma > 0$ such that $\|(J_k + \nu_k I)\xi\| \geq \gamma\|\xi\|$ for all $\xi \in \mathbf{R}^n$ and all k sufficiently large. It is assumed from now on that k is sufficiently large so that all the asymptotic properties and assumptions stated above are in place.

By (5.112) we then derive that

$$\|\Phi(x^k)\| \geq \|(J_k + \nu_k I)p^k\| - \|w^k\| \geq \gamma\|p^k\| - \rho_k\nu_k\|p^k\|.$$

Since $\rho_k\nu_k \to 0$ as $k \to \infty$, it follows that

$$\|p^k\| \leq \frac{2}{\gamma}\|\Phi(x^k)\|. \tag{5.122}$$

In particular, the sequence $\{p^k\}$ converges to zero. By the Lipschitz-continuity of the derivative of Φ in the neighborhood of $\bar{x}$, from Lemma A.11 it follows that there exists $L > 0$ such that

$$\Phi(x^k + p^k) = \Phi(x^k) + \Phi'(x^k)p^k + r^k$$

with some $r^k \in \mathbf{R}^n$ satisfying

$$\|r^k\| \leq L\|p^k\|^2. \tag{5.123}$$

By (5.112) and (5.122) we then obtain that

$$\Phi(x^k + p^k) = -\nu_k p^k - w^k + r^k. \tag{5.124}$$

Using the Cauchy–Schwarz inequality and (5.123), it follows that

$$\begin{aligned} -\langle \Phi(x^k + p^k),\, p^k \rangle &= \langle \nu_k p^k + w^k - r^k,\, p^k \rangle \\ &\geq \nu_k \|p^k\|^2 - (\|w^k\| + \|r^k\|)\|p^k\| \\ &\geq (1-\rho_k)\nu_k \|p^k\|^2 - L\|p^k\|^3 \\ &= \left(1 - \frac{L\|p^k\|}{(1-\rho_k)\nu_k}\right)(1-\rho_k)\nu_k\|p^k\|^2. \end{aligned}$$

Note that from (5.121) and (5.122) it follows that

$$\lim_{k\to\infty} \frac{L\|p^k\|}{(1-\rho_k)\nu_k} \leq \frac{2L}{\gamma} \lim_{k\to\infty} \frac{\|\Phi(x^k)\|}{(1-\rho_k)\nu_k} = 0.$$

Therefore (recalling that $\sigma < 1$), for k sufficiently large it holds that

$$-\langle \Phi(x^k + p^k),\, p^k \rangle > \sigma(1-\rho_k)\nu_k\|p^k\|^2,$$

i.e., the condition (5.113) is satisfied for $\alpha = 1$. Hence, in Algorithm 5.32, $\alpha_k = 1$ and $y^k = x^k + p^k$ for all k large enough.

Now, if we show that the point y^k is an appropriate approximate solution of the proximal point subproblem (5.110) in the sense of the framework of Sect. 5.4.1, then the rates of convergence would follow from Theorem 5.31 (since the update (5.114) is clearly a special case of (5.93) in the general framework of Algorithm 5.28).

To prove the claim, in the current setting we have to show that there exists a sequence $\{\beta_k\}$ of reals such that $\beta_k \to 0$ as $k \to \infty$ and

$$\|\Phi(y^k) + \nu_k p^k\|^2 \leq \beta_k^2(\|\Phi(y^k)\|^2 + \nu_k^2\|p^k\|^2). \tag{5.125}$$

By (5.124) we have that

$$\|\Phi(y^k) + \nu_k p^k\| = \|r^k - w^k\| \leq \|p^k\|(\rho_k\nu_k + L\|p^k\|). \tag{5.126}$$

Next, by (5.112) we obtain that

$$\begin{aligned} \|p^k\|^2 &= -\frac{1}{\nu_k}\langle p^k,\, \Phi(x^k) + \Phi'(x^k)p^k + w^k \rangle \\ &\leq -\frac{1}{\nu_k}(\langle p^k,\, \Phi(x^k) + w^k \rangle) \\ &\leq \frac{1}{\nu_k}\|p^k\|(\|\Phi(x^k)\| + \rho_k\nu_k), \end{aligned}$$

where we have used the fact that $\Phi'(x^k)$ is positive semidefinite (by the monotonicity of Φ). Hence,

$$\|p^k\| \leq \frac{1}{\nu_k}(\|\Phi(x^k)\| + \rho_k\nu_k).$$

By combining this relation with (5.126), we conclude that

$$\|\Phi(y^k) + \nu_k p^k\| \leq \frac{1}{\nu_k}(\|\Phi(x^k)\| + \rho_k \nu_k)(\rho_k \nu_k + L\|p^k\|).$$

Using the latter inequality together with

$$\|\Phi(x^k)\| = \|(\Phi'(x^k) + \nu_k I)p^k + w^k\| \leq (\Gamma + \nu_k + \rho_k \nu_k)\|p^k\|,$$

we obtain that

$$\|\Phi(y^k) + \nu_k p^k\| \leq \beta_k \nu_k \|p^k\|,$$

where

$$\beta_k = \frac{L\|\Phi(x^k)\|}{\nu_k^2} + \frac{\rho_k(\Gamma + \nu_k + \rho_k \nu_k)}{\nu_k} + \frac{\rho_k^2(\Gamma + \nu_k + \rho_k \nu_k)}{\|\Phi(x^k)\|} + \frac{L\rho_k}{\nu_k}.$$

This proves (5.125), and it remains to observe that under the conditions (5.121) it holds that $\beta_k \to 0$ as $k \to \infty$. $\square$

5.4.3 Globalized Josephy–Newton Method for Monotone Variational Inequalities

Consider now the VI (5.80), where $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ is monotone and continuously differentiable, and $Q \subset \mathbf{R}^n$ is a closed convex set.

To globalize the Josephy–Newton method of Sect. 3.1 for the VI (5.80), one possibility is to employ a linesearch procedure in the obtained Newton direction (if it exists) aimed at decreasing the value of a suitable merit function, similarly to the developments in Sect. 5.1. Globalizations based on this approach have been proposed in [189, 217, 218, 259].

The method of [189] is based on the use of the so-called gap function as a merit function (see also [68, Sect. 10.2]). For global convergence of this method, Φ has to be monotone, Q compact, and the linesearch has to be the exact minimization of the gap function along the Newton direction. Globalization developed in [259] employs the so-called regularized gap function (see also [68, Sect. 10.2.1]). For convergence, Φ has to be strongly monotone. In both [189] and [259], the strict complementarity condition (for the inequality constraints defining the set Q and the associated multipliers in the Karush–Kuhn–Tucker conditions for the VI (5.80)) is needed for the superlinear convergence rate.

Algorithms proposed in [217, 218] are based on the so-called D-gap function (see also [68, Sect. 10.3]). These globalizations employ the "safeguard" possibility of performing a standard gradient descent step for the merit function whenever the Newton direction does not exist or is not satisfactory, similarly to the algorithms in Sect. 5.1. Thus the methods of [217, 218] are well defined for any Φ. It typically holds that every accumulation point of the generated sequence of iterates is a stationary point of the merit function employed in the algorithm. However, the existence of such accumulation points, and the equivalence of stationary points of merit functions to solutions of the VI (5.80) cannot be guaranteed without further assumptions, usually rather restrictive. For example, in [217] Φ is assumed to be strongly monotone, and in [218] Φ is the so-called uniform P-function [68, Sect. 3.5.2] and Q is a box. Either of those assumptions implies that solution of the VI (5.80) is in fact globally unique.

Here we follow the development in [257] (see also [246] for a somewhat related approach). For the special case of complementarity problems a different method, but also based on the inexact proximal point scheme, is given in [254]. The advantage of those schemes for monotone problems is that the method is globally well defined, the Newtonian directions are never discarded, convergence to a solution is ensured without any further assumptions, and fast local rate is achieved if the solution is regular in a certain sense.

As already commented above, under our assumptions the VI (5.80) is equivalent to the GE (5.79) with the maximal monotone multifunction defined in (5.81). The globalization technique presented here is based on a linesearch in the (regularized) Josephy–Newton direction which finds a trial point and parameters of the proximal point subproblem for which this trial point is an acceptable approximate solution. We emphasize that this requires only checking the corresponding approximation criterion, and in particular, does not entail actually solving any additional Josephy–Newton subproblems in the course of linesearch. Fast rate of convergence is induced by checking, before linesearch, whether the full Josephy–Newton step provides an acceptable approximate solution in the sense of the relative-error tolerance of the inexact proximal point scheme of Sect. 5.4.1.

We start with describing some ingredients of the algorithm. Given any $\tilde{x} \in \mathbf{R}^n$ fixed, consider the family of proximal point subproblems (5.84), where $\nu > 0$ is not necessarily fixed. According to Sect. 5.4.1 (see (5.86)), the following set of conditions guarantees that progress from the point $\tilde{x}$ to the solution set of GE (5.79) can be made:

$$\begin{gathered} v \in \Psi^{\varepsilon}(x), \quad v + \nu(x - \tilde{x}) + \omega = 0, \\ \varepsilon \geq 0, \quad \nu > 0, \quad \|\omega\|^2 + 2\nu\varepsilon \leq \sigma^2 \left(\|v\|^2 + \nu^2\|x - \tilde{x}\|^2\right), \quad \sigma \in [0,\, 1). \end{gathered} \tag{5.127}$$

One important difference from the preceding material is that we now treat the proximal parameter ν and the error tolerance parameter σ as variables. Specifically, *after* the Josephy–Newton step would be computed for the proximal subproblem, those parameters will be adjusted in order to achieve (5.127).

Consider the proximal subproblem

$$\Psi(x) + \rho(x - \tilde{x}) \ni 0,$$

with $\rho > 0$ now fixed. For Ψ given by (5.81) the latter is equivalent to the VI

$$x \in Q, \quad \langle \Phi(x) + \rho(x - \tilde{x}),\, y - x\rangle \geq 0 \quad \forall\, y \in Q. \tag{5.128}$$

Let $\hat{x} \in \mathbf{R}^n$ be the point obtained after one Josephy–Newton iteration (recall Sect. 3.1) applied to (5.128) from the point $\tilde{x}$, i.e., the point satisfying

$$x \in Q, \quad \langle \Phi(\tilde{x}) + (\Phi'(\tilde{x}) + \rho I)(x - \tilde{x}),\, y - x\rangle \geq 0 \quad \forall\, y \in Q. \tag{5.129}$$

Note that, according to the discussion in Sect. 1.3.1, the solution of (5.129) exists and is unique, because due to the monotonicity of Φ the matrix $\Phi'(\tilde{x})$ is positive semidefinite, and so the mapping in (5.129) is strongly monotone (see Proposition A.17).

If $\hat{x}$ solves the Josephy–Newton subproblem (5.129), then

$$-(\Phi(\tilde{x}) + (\Phi'(\tilde{x}) + \rho I)(\hat{x} - \tilde{x})) \in N_Q(\hat{x}).$$

Hence,

$$\Phi(\hat{x}) - \Phi(\tilde{x}) - (\Phi'(\tilde{x}) + \rho I)(\hat{x} - \tilde{x}) \in \Phi(\hat{x}) + N_Q(\hat{x}) = \Psi(\hat{x}).$$

Thus, $\hat{x}$ provides an acceptable approximate solution to the proximal subproblem (5.128) if, for some $\sigma \in [0,\, 1)$,

$$\begin{aligned}\|\Phi(\hat{x}) - \Phi(\tilde{x}) - \Phi'(\tilde{x})(\hat{x} - \tilde{x})\|^2 \leq \sigma^2(&\|\Phi(\hat{x}) - \Phi(\tilde{x}) - (\Phi'(\tilde{x}) + \rho I)(\hat{x} - \tilde{x})\|^2 \\ &+\rho^2\|\hat{x} - \tilde{x}\|^2).\end{aligned}$$

If this is the case, the method would proceed according to the framework of Sect. 5.4.1. However, far from a solution with proper regularity properties, the latter need not be the case. In what follows, we first show that for all $\alpha > 0$ sufficiently small, the point $y(\alpha) = \tilde{x} + \alpha(\hat{x} - \tilde{x})$, the modified regularization parameter $\nu(\alpha) = \rho/\alpha > 0$, and certain associated computable $v(\alpha) \in \mathbf{R}^n$, $\varepsilon(\alpha) \geq 0$ and $\sigma(\alpha) \in [0,\, 1)$ satisfy the approximation criterion (5.127). In other words, performing an Armijo-type linesearch from $\tilde{x}$ in the Josephy–Newton direction $\hat{x} - \tilde{x}$ we can find a proximal point subproblem for which an acceptable approximate solution is readily available. This solution can then be used in the inexact proximal point framework of Sect. 5.4.1, with the purpose being globalization of convergence of the Josephy–Newton algorithm for the VI (5.80).

It is convenient to start with the following auxiliary result.

Proposition 5.35. *Let $Q \subset \mathbf{R}^n$ be convex, and let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be monotone and differentiable at $\tilde{x} \in Q$ which is not a solution of the VI (5.80). Let $\hat{x} \in \mathbf{R}^n$ be the solution of the VI (5.129).*

Then for all $\alpha \in (0,\, 1]$ sufficiently small and all $\sigma \in [0,\, 1)$ close enough to one, points of the form $y(\alpha) = \tilde{x} + \alpha(\hat{x} - \tilde{x})$ belong to the set Q and satisfy the inequality

$$\left\| \Phi(y(\alpha)) + \frac{\rho}{\alpha}(y(\alpha) - \tilde{x}) \right\|^2 \leq \sigma^2 \left(\|\Phi(y(\alpha))\|^2 + \frac{\rho^2}{\alpha^2}\|y(\alpha) - \tilde{x}\|^2 \right). \quad (5.130)$$

If Φ is further Lipschitz-continuous with constant $L > 0$, then the conditions

$$0 \leq \sigma < 1, \quad 0 < \alpha \leq \frac{(1 + \sigma^2)\rho}{4L} \quad (5.131)$$

imply that $\Phi(y(\alpha)) \neq 0$, and (5.130) is satisfied whenever in addition

$$\sigma \geq \sqrt{\max\left\{0,\ 1 - \frac{\rho^2}{2} \frac{\|\hat{x} - \tilde{x}\|^2}{\|\Phi(y(\alpha))\|^2}\right\}}. \quad (5.132)$$

Proof. By the convexity of Q, we have that $y(\alpha) = \alpha\tilde{x} + (1 - \alpha)\hat{x} \in Q$ for all $\alpha \in [0,\ 1]$.

Using (5.129) with $y = \hat{x} \in Q$, we have that

$$\langle \Phi(\tilde{x}), \hat{x} - \tilde{x} \rangle \leq \langle (\Phi'(\tilde{x}) + \rho I)(\hat{x} - \tilde{x}),\ \tilde{x} - \hat{x} \rangle \leq -\rho\|\hat{x} - \tilde{x}\|^2,$$

where the second inequality follows from the positive semidefiniteness of $\Phi'(\tilde{x})$ (since Φ is monotone). By the Cauchy–Schwarz inequality, we further obtain that

$$\begin{aligned} \langle \Phi(y(\alpha)),\ \hat{x} - \tilde{x} \rangle &= \langle \Phi(\tilde{x}),\ \hat{x} - \tilde{x} \rangle + \langle \Phi(y(\alpha)) - \Phi(\tilde{x}),\ \hat{x} - \tilde{x} \rangle \\ &\leq -\rho\|\hat{x} - \tilde{x}\|^2 + \|\Phi(y(\alpha)) - \Phi(\tilde{x})\|\|\hat{x} - \tilde{x}\|. \quad (5.133) \end{aligned}$$

Next, for all $\alpha \in (0,\ 1]$ and $\sigma \in [0,\ 1)$ it holds that

$$\begin{aligned} &\left\| \Phi(y(\alpha)) + \frac{\rho}{\alpha}(y(\alpha) - \tilde{x}) \right\|^2 - \sigma^2 \left(\|\Phi(y(\alpha))\|^2 + \frac{\rho^2}{\alpha^2}\|y(\alpha) - \tilde{x}\|^2 \right) \\ &= (1 - \sigma^2)(\|\Phi(y(\alpha))\|^2 + \rho^2\|\hat{x} - \tilde{x}\|^2) + 2\rho\langle \Phi(y(\alpha)),\ \hat{x} - \tilde{x} \rangle \\ &\leq (1 - \sigma^2)\|\Phi(y(\alpha))\|^2 - (1 + \sigma^2)\rho^2\|\hat{x} - \tilde{x}\|^2 \\ &\quad +2\rho\|\Phi(y(\alpha)) - \Phi(\tilde{x})\|\|\hat{x} - \tilde{x}\|, \end{aligned}$$

where the inequality follows from (5.133). Hence, (5.130) would be satisfied whenever the right-hand side in the relation above is nonpositive, i.e., whenever

$$(1-\sigma^2)\|\Phi(y(\alpha))\|^2 + 2\rho\|\Phi(y(\alpha)) - \Phi(\tilde{x})\|\|\hat{x} - \tilde{x}\| \leq (1+\sigma^2)\rho^2\|\hat{x} - \tilde{x}\|^2. \quad (5.134)$$

As $\alpha \to 0$ and $\sigma \to 1$, the left-hand side in this inequality tends to zero, while the right-hand side tends to $2\rho^2\|\hat{x} - \tilde{x}\|^2 > 0$, provided $\hat{x} \neq \tilde{x}$. Observe that if $\hat{x} = \tilde{x}$, then (5.129) implies that $\tilde{x}$ solves (5.80). This proves the first assertion.

Next, if Φ is Lipschitz-continuous with constant $L > 0$, then (5.134) would be implied by

$$(1 - \sigma^2)\|\Phi(y(\alpha))\|^2 \leq ((1 + \sigma^2)\rho - 2L\alpha)\rho\|\hat{x} - \tilde{x}\|^2. \quad (5.135)$$

In particular, assuming the second condition in (5.131) with some σ and α, if it would hold that $\Phi(y(\alpha)) = 0$, then (5.135) is automatic. Hence, (5.134) holds as well, which in turn implies (5.130). Taking into account the first condition in (5.131), it is evident that (5.130) with $\Phi(y(\alpha)) = 0$ would imply that $\hat{x} = \tilde{x}$, which is impossible by the assumption that $\tilde{x}$ is not a solution of (5.80).

Therefore, $\Phi(y(\alpha)) \neq 0$, and it can then be seen that (5.132) is well defined, and (5.131), (5.132) give a sufficient condition for (5.135). □

Note that in Proposition 5.35 no claim is made that the two conditions (5.131) and (5.132) can be achieved simultaneously. The point here is that if (5.130) is not satisfied, then necessarily at least one of the conditions in (5.131) or (5.132) must be violated.

We next construct all the relevant objects to satisfy the conditions specified in (5.127).

Proposition 5.36. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be monotone, and let $Q \subset \mathbf{R}^n$ be closed and convex.*

Then for any $\tilde{x}$, $x \in Q$, $\nu > 0$ and $\sigma \in [0,\, 1)$, for

$$\begin{gathered} v = \frac{\nu^2}{1-\sigma^2}(\tilde{x} - \pi_Q(u)), \quad u = \tilde{x} - \frac{1-\sigma^2}{\nu^2}\Phi(x), \\ \varepsilon = \langle \Phi(x) - v,\, x - \pi_Q(u)\rangle, \end{gathered} \tag{5.136}$$

it holds that $\varepsilon \geq 0$ and $v \in \Psi^{\varepsilon}(x)$.

Proof. First note that according to (5.136) we have that

$$\varepsilon = \frac{\nu^2}{1-\sigma^2}\langle u - \pi_Q(u),\, \pi_Q(u) - x\rangle.$$

Since $x \in Q$, by the basic properties of the projection it holds that $\varepsilon \geq 0$, and also that

$$\pi_Q(u) = u - w \tag{5.137}$$

for some $w \in N_Q(\pi_Q(u))$; see Lemma A.12. From the first two equalities in (5.136) we then further obtain that

$$v = \frac{\nu^2}{1-\sigma^2}(\tilde{x} - u + w) = \Phi(x) + \eta, \tag{5.138}$$

where

$$\eta = \frac{\nu^2}{1-\sigma^2} w \in N_Q(\pi_Q(u)). \tag{5.139}$$

To verify that $v \in \Psi^{\varepsilon}(x)$ we have to show that

$$\langle v - \tilde{v},\, x - \xi\rangle \geq -\varepsilon \quad \forall\, \tilde{v} \in \Psi(\xi),\ \forall\, \xi \in \mathbf{R}^n.$$

According to (5.81), since $N_Q(\xi) = \emptyset$ for $\xi \notin Q$, the inclusion $\tilde{v} \in \Psi(\xi)$ implies that $\xi \in Q$, and we have $\tilde{v} = \Phi(\xi) + \zeta$ with some $\zeta \in N_Q(\xi)$. Employing (5.138), we then obtain that

$$\begin{aligned}
\langle v - \tilde{v},\, x - \xi\rangle &= \langle \Phi(x) - \Phi(\xi),\, x - \xi\rangle + \langle \eta - \zeta,\, x - \xi\rangle \\
&\geq \langle \eta,\, x - \xi\rangle - \langle \zeta,\, x - \xi\rangle \\
&\geq \langle \eta,\, x - \xi\rangle \\
&= \langle \eta,\, x - \pi_Q(u)\rangle - \langle \eta,\, \xi - \pi_Q(u)\rangle \\
&\geq \langle \eta,\, x - \pi_Q(u)\rangle,
\end{aligned}$$

where the first inequality follows from the monotonicity of Φ, the second follows from $x \in Q$, $\zeta \in N_Q(\xi)$, and the last from $\xi \in Q$, $\eta \in N_Q(\pi_Q(u))$. We conclude that

$$v \in \Psi^{\varepsilon}(x) \quad \text{with} \quad \varepsilon = -\langle \eta,\, x - \pi_Q(u)\rangle \geq 0,$$

where, according to (5.138), $\eta = v - \Phi(x)$. $\square$

Proposition 5.37. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be monotone, and let $Q \subset \mathbf{R}^n$ be closed and convex. Let the condition*

$$\left\|\Phi(x) + \frac{\rho}{\alpha}(x - \tilde{x})\right\|^2 \leq \sigma^2 \left(\|\Phi(x)\|^2 + \frac{\rho^2}{\alpha^2}\|x - \tilde{x}\|^2\right) \tag{5.140}$$

hold for some $\tilde{x}$, $x \in Q$, $\alpha > 0$, $\rho > 0$, and $\sigma \in [0, 1)$.

Then the relations in (5.127) hold for these x and σ, for $\nu = \rho/\alpha$, and for $v \in \mathbf{R}^n$ and $\varepsilon \geq 0$ given by (5.136).

Proof. By Proposition 5.36, we have that $\varepsilon \geq 0$ and $v \in \Psi^{\varepsilon}(y)$. Therefore, it remains to prove the inequality in (5.127). To that end, employing (5.136)–(5.139) we obtain that

$$\begin{aligned}
&\|v + \nu(x - \tilde{x})\|^2 + 2\nu\varepsilon - \sigma^2(\|v\|^2 + \nu^2\|x - \tilde{x}\|^2) \\
&= \|\Phi(x) + \nu(x - \tilde{x})\|^2 + \|\eta\|^2 + 2\langle \eta,\, \Phi(x) + \nu(x - \tilde{x})\rangle - 2\nu\langle \eta,\, x - \pi_Q(u)\rangle \\
&\quad -\sigma^2\|\Phi(x)\|^2 - \sigma^2\|\eta\|^2 - 2\sigma^2\langle \Phi(x),\, \eta\rangle - \sigma^2\nu^2\|x - \tilde{x}\|^2 \\
&\leq (1 - \sigma^2)\|\eta\|^2 + 2\langle \eta,\, (1 - \sigma^2)\Phi(x) + \nu(\pi_Q(u) - \tilde{x})\rangle \\
&= (1 - \sigma^2)\|\eta\|^2 + 2\nu\langle \eta,\, \nu(\tilde{x} - u) + \nu(u - w - \tilde{x})\rangle \\
&= (1 - \sigma^2)\|\eta\|^2 - 2\nu^2\langle \eta,\, w\rangle \\
&= -(1 - \sigma^2)\|\eta\|^2 \\
&\leq 0,
\end{aligned}$$

where the first inequality follows from the hypothesis (5.140). $\square$

We next describe the algorithm. After the Josephy–Newton subproblem (5.129) is solved, we first test whether its solution $\hat{x}$ already provides an approximate solution to the proximal subproblem (5.128), acceptable in the sense of Sect. 5.4.1. If this is so, the iteration is completed as prescribed by the inexact proximal point framework of Sect. 5.4.1. If not, then linesearch based on the guarantees of Proposition 5.37 is performed, i.e., we search for parameters of the proximal point subproblem for which an appropriate approximate solution is available.

Let $R : \mathbf{R}^n \to \mathbf{R}^n$ be the natural residual mapping of the VI (5.80), given by

$$R(x) = x - \pi_Q(x - \Phi(x)).$$

As is discussed in Sect. 1.3.1, $x \in \mathbf{R}^n$ solves the VI (5.80) if and only if $R(x) = 0$ holds.

Algorithm 5.38 Choose the parameters $c > 0$, $C > 0$, τ, $\theta \in (0, 1)$, and $s \in (0, 1)$. Choose $x^0 \in Q$ and set $k = 0$.

1. Compute $R(x^k)$, and stop if $R(x^k) = 0$.
2. Set
$$\rho_k = \min\{C,\, c\|R(x^k)\|^\tau\}. \tag{5.141}$$
Compute $\hat{x}^k$, the solution of the VI
$$x \in Q, \quad \langle \Phi(x^k) + (\Phi'(x^k) + \rho_k I)(x - x^k),\, y - x\rangle \geq 0 \quad \forall\, y \in Q. \tag{5.142}$$
3. Set
$$v^k = \Phi(\hat{x}^k) - \Phi(x^k) - (\Phi'(x^k) + \rho_k I)(\hat{x}^k - x^k).$$
If
$$\|v^k + \rho_k(\hat{x}^k - x^k)\|^2 \leq (1 - s)^2(\|v^k\|^2 + \rho_k^2\|\hat{x}^k - x^k\|^2), \tag{5.143}$$
then set $y^k = \hat{x}^k$, $\varepsilon_k = 0$, $\nu_k = \rho_k$, $\sigma_k = 1 - s$, $\alpha_k = 1$, and go to step 5.
4. Set $\alpha = 1$. If the objects
$$\begin{aligned} y^k &= x^k + \alpha(\hat{x}^k - x^k), \quad \nu_k = \rho_k/\alpha, \quad \sigma_k = 1 - \alpha s, \\ v^k &= \frac{\nu_k^2}{1 - \sigma_k^2}\left(x^k - \pi_Q\left(x^k - \frac{1 - \sigma_k^2}{\nu_k^2}\Phi(y^k)\right)\right), \\ \varepsilon_k &= \left\langle \Phi(y^k) - v^k,\, y^k - \pi_Q\left(x^k - \frac{1 - \sigma_k^2}{\nu_k^2}\Phi(y^k)\right)\right\rangle, \end{aligned} \tag{5.144}$$
satisfy the condition
$$\|v^k + \nu_k(y^k - x^k)\|^2 + 2\nu_k\varepsilon_k \leq \sigma_k^2(\|v^k\|^2 + \nu_k^2\|y^k - x^k\|^2), \tag{5.145}$$
then set $\alpha_k = \alpha$ and go to step 5. Otherwise replace α by $\theta\alpha$, compute the objects in (5.144) and check (5.145) again, etc., until (5.145) becomes valid.

5. Compute
$$x^{k+1} = \pi_Q \left(x^k - \frac{\langle v^k,\, x^k - y^k \rangle - \varepsilon_k}{\|v^k\|^2} v^k \right).$$
6. Increase k by 1, and go to step 1.

We start with establishing global convergence of Algorithm 5.38.

Theorem 5.39. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be monotone and continuously differentiable on $\mathbf{R}^n$, let $Q \subset \mathbf{R}^n$ be closed and convex, and let the solution set of the VI (5.80) be nonempty.*

Then for any starting point $x^0 \in \mathbf{R}^n$ Algorithm 5.38 generates a sequence $\{x^k\}$ convergent to a solution of the VI (5.80).

Proof. Consider any iteration index k. If $R(x^k) = 0$, then x^k is a solution of (5.80) and Algorithm 5.38 terminates finitely. As usual, we assume that this does not happen. Then in (5.142) $\rho_k > 0$. By the monotonicity of Φ, it then follows that (5.142) is a VI with a strongly monotone mapping. Hence, $\hat{x}^k$ exists and is unique. Note also that $x^k \in Q$, by step 5.

By Proposition 5.37, the linesearch procedure in step 4, if activated, is well defined and terminates finitely. Next, the variable update rule is well defined whenever $v^k \neq 0$. Now, if (5.143) were to be satisfied with $v^k = 0$, then from (5.143) it would further follow that $\hat{x}^k = x^k$. But then $\hat{x}^k$ solving (5.142) means that x^k solves the original problem (5.80) and the algorithm would have stopped in step 1. Suppose now that (5.145) were to be satisfied with $v^k = 0$. Then (5.145) would imply that $y^k = x^k$ and $\varepsilon_k = 0$, so that by Proposition 5.36 it holds that $0 \in \Psi(x^k)$. But this again means that x^k is a solution of (5.80) and the algorithm would have stopped in step 1. This concludes the proof of the assertion that Algorithm 5.38 is well defined. We next show that an infinite sequence $\{x^k\}$ generated by this algorithm converges to a solution of the VI (5.80).

Whether the test (5.143) was satisfied or not, with the quantities defined in step 3 or 4 of Algorithm 5.38, we have that (5.145) holds. From that relation, it follows that
$$2\nu_k \langle v^k,\, y^k - x^k \rangle + 2\nu_k \varepsilon_k \leq (\sigma_k^2 - 1)(\|v^k\|^2 + \nu_k^2 \|y^k - x^k\|^2),$$
and hence,
$$\langle v^k,\, x^k - y^k \rangle - \varepsilon_k \geq \frac{1 - \sigma_k^2}{2\nu_k} (\|v^k\|^2 + \nu_k^2 \|y^k - x^k\|^2) > 0. \tag{5.146}$$
On the other hand, for any solution $\bar{x}$ of the VI (5.80), by the relations $0 \in \Psi(\bar{x})$ and $v^k \in \Psi^{\varepsilon_k}(y^k)$ (the latter holds by the construction in (5.144) and Proposition 5.36) we conclude that
$$\langle v^k,\, \bar{x} - y^k \rangle - \varepsilon_k \leq 0.$$

It follows from the above relation and (5.146) that the hyperplane

$$H_k = \{y \in \mathbf{R}^n \mid \langle v^k,\, y - y^k\rangle = \varepsilon_k\}$$

separates x^k from $\bar{x}$, and it is further easily seen that

$$x^k - \frac{\langle v^k,\, x^k - y^k\rangle - \varepsilon_k}{\|v^k\|^2} v^k = \pi_{H_k}(x^k).$$

Since $\bar{x} \in Q$, by the properties of the Euclidean projections (Lemma A.12), it holds that

$$\begin{aligned} \|x^{k+1} - \bar{x}\|^2 &= \|\pi_Q(\pi_{H_k}(x^k)) - \bar{x}\|^2 \\ &\leq \|\pi_{H_k}(x^k) - \bar{x}\|^2 \\ &\leq \|x^k - \bar{x}\|^2 - \|\pi_{H_k}(x^k) - x^k\|^2 \\ &= \|x^k - \bar{x}\|^2 - \frac{(\langle v^k,\, x^k - y^k\rangle - \varepsilon_k)^2}{\|v^k\|^2}. \end{aligned} \tag{5.147}$$

We conclude that $\{\|x^k - \bar{x}\|\}$ converges, $\{x^k\}$ is bounded, and so is $\{R(x^k)\}$ by continuity. Also, (5.147) implies that

$$\lim_{k\to\infty} \frac{\langle v^k,\, x^k - y^k\rangle - \varepsilon_k}{\|v^k\|} = 0. \tag{5.148}$$

Combining the latter relation with (5.146), we conclude that

$$\lim_{k\to\infty} \frac{1-\sigma_k^2}{\nu_k}\|v^k\| = 0. \tag{5.149}$$

If there is a subsequence $\{x^{k_j}\}$ such that $\{R(x^{k_j})\} \to 0$, then every accumulation point of $\{x^{k_j}\}$ is a solution of the VI (5.80). Let $\hat{x}$ be any accumulation point of $\{x^{k_j}\}$. Using $\hat{x}$ in place of $\bar{x}$ in (5.147), we obtain that $\{\|x^k - \hat{x}\|\}$ converges, and since $\hat{x}$ is an accumulation point of $\{x^k\}$, this convergence must be to zero. We conclude that in this case $\{x^k\}$ converges to some solution of the VI (5.80).

We next consider the two possible cases:

$$\liminf_{k\to\infty} \alpha_k > 0 \quad \text{or} \quad \liminf_{k\to\infty} \alpha_k = 0. \tag{5.150}$$

In the first case, $\{\nu_k\}$ would be bounded above (because $\{\rho_k\}$ is; see steps 2 and 4) and $\{\sigma_k\}$ would be bounded away from one. Then from (5.149) we obtain that

$$\lim_{k\to\infty} \|v^k\| = 0. \tag{5.151}$$

Using (5.148), it then follows that

$$\lim_{k\to\infty} (\langle v^k,\, x^k - y^k\rangle - \varepsilon_k) = 0.$$

Using the latter relation, (5.149) and (5.151) when passing onto the limit in (5.146), we obtain that

$$\lim_{k\to\infty} \nu_k \|y^k - x^k\|^2 = 0. \tag{5.152}$$

As the parameter ν_k is nondecreasing within step 4, if $\liminf_{k\to\infty} \nu_k = 0$, then $\liminf_{k\to\infty} \rho_k = 0$, and then (5.141) implies $\liminf_{k\to\infty} \|R(x^k)\| = 0$. Then $\{x^k\}$ has an accumulation point which is a solution of the VI (5.80), and the argument already specified above applies to conclude that $\{x^k\}$ converges to this solution.

If $\liminf_{k\to\infty} \nu_k > 0$, then (5.152) gives

$$\lim_{k\to\infty} \|y^k - x^k\| = 0. \tag{5.153}$$

Observe that (5.146) then also implies

$$\lim_{k\to\infty} \varepsilon_k = 0. \tag{5.154}$$

Now, let $\bar{x}$ be any accumulation point of $\{x^k\}$, and $\{x^{k_j}\}$ be some subsequence converging to $\bar{x}$. By (5.153), $\{y^{k_j}\}$ also converges to $\bar{x}$. Since $v^k \in \Psi^{\varepsilon_k}(y^k)$ for all k, for any $x \in \mathbf{R}^n$ and $v \in \Psi(\xi)$ we have that

$$\langle v - v^{k_j},\, x - y^{k_j}\rangle \geq -\varepsilon_{k_j}.$$

Now passing onto the limit in the above relation, and taking into account (5.151), (5.153), and (5.154), we conclude that

$$\langle v - 0,\, x - \bar{x}\rangle \geq 0.$$

By the maximal monotonicity of Ψ, it now follows that $0 \in \Psi(\bar{x})$, i.e., $\bar{x}$ solves the VI (5.80). The conclusion that the whole sequence $\{x^k\}$ converges to $\bar{x}$ follows as before.

Consider now the second case in (5.150). Let $\{x^{k_j}\}$ be a subsequence of iterates such that

$$\lim_{j\to\infty} \alpha_{k_j} = 0,$$

and note that this means that step 3 was not successful and step 4 was active on the corresponding iterations of the algorithm.

Since Φ is continuously differentiable and $\{x^k\}$ is bounded, Φ is Lipschitz-continuous on some set containing $\{x^k\}$. By Propositions 5.35 and 5.37 (recall relations (5.131)–(5.132)), since the value $\alpha = \alpha_{k_j}/\theta$ was rejected for j large enough, it must have been the case that

$$\frac{\alpha_{k_j}}{\theta} > \frac{\rho_{k_j}\left(1 + \left(1 - \dfrac{\alpha_{k_j}}{\theta} s\right)^2\right)}{4L} \tag{5.155}$$

or/and $\Phi(y^{k_j}) \neq 0$ and

$$1 - \frac{\alpha_{k_j}}{\theta} s < \sqrt{\max\left\{0,\, 1 - \frac{\rho_{k_j}^2}{2} \frac{\|\hat{x}^{k_j} - x^{k_j}\|^2}{\|\Phi(y^{k_j})\|^2}\right\}}. \tag{5.156}$$

At least one of the inequalities above must hold an infinite number of times. If the first one holds an infinite number of times, then passing to the corresponding subsequence of $\{x^{k_j}\}$, and then passing onto the limit in (5.155) yields

$$\liminf_{j\to\infty} \rho_{k_j} = 0. \tag{5.157}$$

Then by step 2 we have that $\liminf_{j\to\infty} \|R(x^{k_j})\| = 0$, and the same arguments as above imply the assertion of the theorem.

If (5.155) holds a finite number of times, then (5.156) must hold for all indices j sufficiently large. Passing onto the limit in (5.156) as $j \to \infty$, we obtain that

$$\lim_{j\to\infty} \frac{\rho_{k_j}^2 \|\hat{x}^{k_j} - x^{k_j}\|^2}{\|\Phi(y^{k_j})\|^2} = 0,$$

which means that

$$\lim_{j\to\infty} \rho_{k_j} \|\hat{x}^{k_j} - x^{k_j}\| = 0 \quad \text{or/and} \quad \lim_{i\to\infty} \|\Phi(y^{k_j})\| = +\infty. \tag{5.158}$$

If (5.157) holds, then the previous argument applies. Suppose (5.157) does not hold: there exists $\bar{\rho} > 0$ such that $\rho_{k_j} \geq \bar{\rho}$ for all j. Since $\hat{x}^{k_j}$ solves (5.142), for each $y \in Q$ it holds that

$$\langle \Phi(x^{k_j}),\, y - \hat{x}^{k_j}\rangle \geq \langle (\Phi'(x^{k_j}) + \rho_{k_j} I)(x^{k_j} - \hat{x}^{k_j}),\, y - \hat{x}^{k_j}\rangle. \tag{5.159}$$

Taking $y = x^{k_j}$, using the Cauchy–Schwarz inequality in the left-hand side of (5.159), and positive semidefiniteness of $\Phi'(x^{k_j})$ in the right-hand side, we obtain that

$$\|\Phi(x^{k_j})\| \geq \bar{\rho}\|\hat{x}^{k_j} - x^{k_j}\|.$$

Since $\{x^k\}$ is bounded, so is $\{\Phi(x^k)\}$. It follows that $\{\hat{x}^{k_j}\}$ is bounded. It then also follows that $\{y^{k_j}\}$ and $\{\Phi(y^{k_j})\}$ are bounded. Hence, in (5.158) the first equality must hold. And since $\rho_{k_j} \geq \bar{\rho} > 0$ for all j, in fact it holds that $\lim_{j\to\infty} \|\hat{x}^{k_j} - x^{k_j}\| = 0$.

Let $\bar{x}$ be a limit of some subsequence of $\{x^{k_j}\}$. Then the corresponding subsequence of $\{\hat{x}^{k_j}\}$ converges to the same limit $\bar{x}$. Fix any $y \in Q$. Passing onto the limit in (5.159) along the subsequences of $\{x^{k_j}\}$ and $\{\hat{x}^{k_j}\}$ converging to $\bar{x}$, and taking into account the boundedness of $\{\rho_{k_j}\}$, we obtain that

$$\langle \Phi(\bar{x}),\, y - \bar{x}\rangle \geq 0.$$

As $\bar{x} \in Q$ and $y \in Q$ is arbitrary, this means that $\bar{x}$ solves the VI (5.80). Convergence of the whole sequence $\{x^k\}$ to $\bar{x}$ follows as before. $\square$

We next show that close to a strongly regular solution of the VI (5.80), step 3 of Algorithm 5.38 is always successful. This means that the method reduces to the inexact proximal point scheme of Sect. 5.4.1, thus inheriting its rate of convergence.

Theorem 5.40. *In addition to the assumptions of Theorem 5.39, let the derivative of Φ be locally Lipschitz-continuous at $\bar{x} \in \mathbf{R}^n$, and let $\bar{x}$ be the (necessarily unique) solution of the VI (5.80) with positive definite $\Phi'(\bar{x})$.*

Then for any $x^0 \in \mathbf{R}^n$ the sequence $\{x^k\}$ generated by Algorithm 5.38 converges to $\bar{x}$, and the rate of convergence is superlinear.

Proof. Because $\Phi'(\bar{x})$ is positive definite, it holds that the multifunction $x \to \Phi'(\bar{x})(x - \bar{x}) + N_Q(x)$ is maximal monotone and strongly monotone (see Proposition A.17), and therefore, as discussed in Sect. 1.3.2, $\bar{x}$ is a strongly regular (hence semistable) solution of the VI (5.80). As discussed in Sect. 5.4.1, this solution is necessarily unique. By Theorem 5.39 we then have that $\{x^k\}$ converges to this $\bar{x}$, and hence, according to (5.141), ρ_k tends to zero as $k \to \infty$. We next show that (5.143) is satisfied for all k large enough.

By the definition of v^k in step 3 of the algorithm, it holds that

$$\begin{aligned}\|v^k + \rho_k(\hat{x}^k - x^k)\| &= \|\Phi(\hat{x}^k) - \Phi(x^k) - \Phi'(x^k)(\hat{x}^k - x^k)\| \\ &= O(\|\hat{x}^k - x^k\|^2)\end{aligned}$$

as $x^k \to \bar{x}$ and $\hat{x}^k \to \bar{x}$, where the last equality is by Lemma A.11. It is now evident that (5.143) is guaranteed to hold for all k sufficiently large if

$$\hat{x}^k - x^k = o(\rho_k) \tag{5.160}$$

as $k \to \infty$.

Consider the VI (5.129) as the parametric GE

$$\Phi(\tilde{x}) + (\Phi'(\tilde{x}) + \rho I)(x - \tilde{x}) + N_Q(x) \ni 0,$$

where $\tilde{x} \in \mathbf{R}^n$ and $\rho \in \mathbf{R}$ are regarded as parameters. For the base values $\tilde{x} = \bar{x}$ and $\rho = 0$ of these parameters, this subproblem has a solution $x = \bar{x}$, and this solution is strongly regular. Applying Theorem 1.24 we then conclude that for all $\tilde{x} \in \mathbf{R}^n$ close to $\bar{x}$, and all $\rho > 0$ close to zero, this subproblem has near $\bar{x}$ the unique solution $\hat{x}(\tilde{x}, \rho)$, depending on $(\tilde{x}, \rho)$ in a Lipschitz-continuous way, and satisfying

$$\begin{aligned}\hat{x}(\tilde{x}, \rho) - \bar{x} &= O(\|\Phi(\tilde{x}) + (\Phi'(\tilde{x}) + \rho I)(\bar{x} - \tilde{x}) - \Phi(\bar{x})\|) \\ &= O(\|\tilde{x} - \bar{x}\|^2 + \rho\|\tilde{x} - \bar{x}\|)\end{aligned} \tag{5.161}$$

as $\tilde{x} \to \bar{x}$ and $\rho \to 0$, where the last estimate is by Lemma A.11. Since under our assumptions the mapping in (5.142) is strongly monotone for each fixed $\tilde{x} \in \mathbf{R}^n$ and $\rho > 0$, we conclude that this $\hat{x}(\tilde{x}, \rho)$ is the unique solution of the VI (5.129) in the entire $\mathbf{R}^n$. Therefore, since (5.129) coincides with (5.142) for $\tilde{x} = x^k$ and $\rho = \rho_k$, for each k it holds that $\hat{x}^k = \hat{x}(x^k, \rho_k)$. This proves that $\{\hat{x}^k\}$ converges to $\bar{x}$, and (5.161) combined with the convergence of $\{\rho_k\}$ to zero yields the estimate

$$\hat{x}^k - \bar{x} = o(\|x^k - \bar{x}\|),$$

and hence,

$$\hat{x}^k - x^k = O(\|x^k - \bar{x}\|) = O(\|R(x^k)\|)$$

as $k \to \infty$, where the last equality is by Proposition 1.31.

Combining the latter relation with (5.141) (where $\tau \in (0,\, 1)$) guarantees (5.160). This completes the proof. $\square$

Chapter 6
Constrained Optimization: Globalization of Convergence

In this chapter we discuss approaches to globalizing the fundamental sequential quadratic programming (SQP) algorithm presented in Sect. 4.2. This can be done in a number of ways. First, candidate points can be produced either using linesearch in the computed SQP direction or solving SQP subproblems with an additional trust-region constraint. (For unconstrained optimization, these two alternatives were considered in Sects. 2.2 and 2.3, respectively.) Second, the generated candidate points can be evaluated either by computing the value of a suitable merit function or by a filter dominance criterion. Different combinations of all the options give rise to a wealth of possibilities. Naturally, we can only cover a few of those. We discuss in detail the linesearch merit function approach, including the issue of preserving the fast local convergence rate of local SQP by the globalized variant (the Maratos effect, second-order corrections, using nonsmooth augmented Lagrangian as a merit function). Dealing with possibly infeasible subproblems via the so-called elastic mode is also discussed, but in the context of the related sequential quadratically constrained quadratic programming (SQCQP) methods, where infeasibility of subproblems is a key issue even in the convex case. Various trust-region merit function based options are discussed only briefly. We then proceed to present an alternative generic filter globalization for constrained optimization, and a specific trust-region SQP method that can be used within this framework.

6.1 Merit Functions

Evidently, merit functions for constrained optimization must combine the objective function of the problem with some measure of violation of its constraints. The goal is to balance the tasks of reducing the objective function and of achieving feasibility, which may be conflicting. In principle, one might also consider using the residuals of the related optimality systems,

A.F. Izmailov and M.V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-3-319-04247-3_6,

following the lines of the development for complementarity problems in Chap. 5. However, the latter is usually not a good option, since the algorithms based on residuals of optimality conditions as merit functions would evaluate candidate points with respect to stationarity rather than minimization. The optimization nature of the problem would be ignored, to some extent. Thus, natural merit functions for constrained optimization are obtained by adding to the objective function a quantity penalizing the constraints violation, perhaps augmented with a primal-dual Lagrangian term.

Consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0,\ g(x) \leq 0, \end{array} \tag{6.1}$$

where $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$, and $g : \mathbf{R}^n \to \mathbf{R}^m$ are given. Let D stand for the feasible set of this problem:

$$D = \{x \in \mathbf{R}^n \mid h(x) = 0,\ g(x) \leq 0\}. \tag{6.2}$$

Choose some function $\psi : \mathbf{R}^n \to \mathbf{R}_+$, called *penalty* for the set D, indicating constraints violation in the sense that $\psi(x) > 0$ if $x \notin D$, and $\psi(x) = 0$ for $x \in D$. Define the related family of *penalty functions* $\varphi_c : \mathbf{R}^n \to \mathbf{R}$ by

$$\varphi_c(x) = f(x) + c\psi(x), \tag{6.3}$$

parameterized by the *penalty parameter* $c \geq 0$, and consider the associated family of unconstrained penalized problems

$$\begin{array}{ll} \text{minimize} & \varphi_c(x) \\ \text{subject to} & x \in \mathbf{R}^n. \end{array} \tag{6.4}$$

In what follows, we shall mostly use the so-called l_1-penalty

$$\psi(x) = \|(h(x),\, \max\{0,\, g(x)\})\|_1 = \|h(x)\|_1 + \sum_{i=1}^{m} \max\{0,\, g_i(x)\}. \tag{6.5}$$

However, all the results in this section hold true (with obvious modifications) also for penalties defined by other norms. One of the other popular choices is the l_∞-penalty

$$\psi(x) = \|(h(x),\, \max\{0,\, g(x)\})\|_\infty = \max\{\|h(x)\|_\infty,\, 0,\, g_1(x),\, \ldots,\, g_m(x)\}, \tag{6.6}$$

where the inclusion of zero in the right-hand side is redundant when the equality constraints are present.

It is well known (and easy to see) that the l_1-norm and the l_∞-norm in $\mathbf{R}^\nu$ are dual to each other in the following sense:

$$\|u\|_\infty = \sup_{v\in\mathbf{R}^\nu:\,\|v\|_1=1} \langle u,\, v\rangle, \quad \|u\|_1 = \sup_{v\in\mathbf{R}^\nu:\,\|v\|_\infty=1} \langle u,\, v\rangle \quad \forall\, u \in \mathbf{R}^\nu. \tag{6.7}$$

This fact will be used below.

Note that the two penalties specified by (6.5) and (6.6) are generally not differentiable, even when h and g are. The advantage of nondifferentiable penalty functions is that, under certain natural assumptions given below, solutions of the original problem (6.1) can be recovered from solutions of the unconstrained problem (6.4) for *finite* values of the penalty parameter $c > 0$. More precisely, under these assumptions, a solution of problem (6.1) must be a solution of problem (6.4) for any $c > 0$ large enough. This desirable property, called *exact penalization*, is not, in general, shared by (primal) penalty functions that employ smooth penalties. Moreover, it so happens that lack of smoothness is not a problem in the current context. Indeed, nonsmooth penalty functions are still directionally differentiable, which makes them suitable for descent methods employing linesearch procedures. The following characterization of directional derivatives of ψ given by (6.5) and (6.6) is obtained by direct computation. The l_∞-penalty is considered in this lemma for the case when there are no equality constraints for simplicity, and because this is precisely what would be needed later on.

Proposition 6.1. *Let $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable at a point $x \in \mathbf{R}^n$.*

Then the following assertions are valid:

(a) The function $\psi : \mathbf{R}^n \to \mathbf{R}_+$ given by (6.5) is directionally differentiable at x in every direction $\xi \in \mathbf{R}^n$, and

$$\begin{aligned}\psi'(x;\, \xi) = &- \sum_{j\in J^-(x)} \langle h_i'(x),\, \xi\rangle + \sum_{j\in J^0(x)} |\langle h_i'(x),\, \xi\rangle| + \sum_{j\in J^+(x)} \langle h_i'(x),\, \xi\rangle \\ &+ \sum_{i\in I^0(x)} \max\{0,\, \langle g_i'(x),\, \xi\rangle\} + \sum_{i\in I^+(x)} \langle g_i'(x),\, \xi\rangle,\end{aligned} \tag{6.8}$$

where

$$\begin{aligned} J^-(x) &= \{j = 1,\, \ldots,\, l \mid h_j(x) < 0\},\\ J^0(x) &= \{j = 1,\, \ldots,\, l \mid h_j(x) = 0\},\\ J^+(x) &= \{j = 1,\, \ldots,\, l \mid h_j(x) > 0\},\\ I^0(x) &= \{i = 1,\, \ldots,\, m \mid g_i(x) = 0\},\\ I^+(x) &= \{i = 1,\, \ldots,\, m \mid g_i(x) > 0\}.\end{aligned}$$

In particular, if $\bar{x} \in \mathbf{R}^n$ is feasible in problem (6.1), it holds that

$$\psi'(\bar{x};\, \xi) = \|h'(\bar{x})\xi\|_1 + \sum_{i\in A(\bar{x})} \max\{0,\, \langle g_i'(\bar{x}),\, \xi\rangle\}. \tag{6.9}$$

(b) *The function $\psi : \mathbf{R}^n \to \mathbf{R}_+$ given by (6.6) is directionally differentiable at x in every direction $\xi \in \mathbf{R}^n$, and if $l = 0$ (i.e., there are no equality constraints), then*

$$\psi'(x;\, \xi) = \begin{cases} \max\limits_{i \in I(x)} \langle g_i'(x),\, \xi\rangle & \text{if } \psi(x) > 0, \\ \max\left\{0,\ \max\limits_{i \in I(x)} \langle g_i'(x),\, \xi\rangle\right\} & \text{if } \psi(x) = 0, \end{cases}$$

where

$$I(x) = \{i = 1,\, \dots,\, m \mid g_i(x) = \psi(x)\}.$$

In particular, if $\bar{x} \in \mathbf{R}^n$ is feasible in problem (6.1), it holds that

$$\psi'(\bar{x};\, \xi) = \max\left\{0,\ \max_{i \in A(\bar{x})} \langle g_i'(\bar{x}),\, \xi\rangle\right\}.$$

We start with conditions that are necessary for exact penalization.

Theorem 6.2. *For any $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$, let $\bar{x} \in \mathbf{R}^n$ be a (strict) local solution of the penalized problem (6.4) for a given $c > 0$, with $\varphi_c : \mathbf{R}^n \to \mathbf{R}$ given by (6.3), (6.5).*

If $\bar{x}$ is feasible in the original problem (6.1), then $\bar{x}$ is a (strict) local solution of this problem.

If, in addition, f, h, and g are differentiable at $\bar{x}$, then $\bar{x}$ is a stationary point of problem (6.1), and there exists a Lagrange multiplier $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ associated with $\bar{x}$, such that

$$\|(\bar{\lambda},\, \bar{\mu})\|_\infty \leq c. \tag{6.10}$$

Proof. As $\bar{x}$ is a local solution of (6.4), there exists a neighborhood U of $\bar{x}$ such that

$$\varphi_c(\bar{x}) \leq \varphi_c(x) \quad \forall\, x \in U. \tag{6.11}$$

Let D be given by (6.2). Since $\varphi_c(x) = f(x)$ for any $x \in D$, if $\bar{x} \in D$, then (6.11) implies

$$f(\bar{x}) \leq f(x) \quad \forall\, x \in D \cap U,$$

i.e., $\bar{x}$ is a local solution of problem (6.1). The case of strict local solutions is also obvious from the same considerations.

Assume now that f, h, and g are differentiable at $\bar{x}$. By assertion (a) of Proposition 6.1, φ_c is differentiable at $\bar{x}$ in every direction $\xi \in \mathbf{R}^n$, and since $\bar{x} \in D$, relation (6.9) combined with (6.3) gives the equality

$$\varphi_c'(\bar{x};\, \xi) = \langle f'(\bar{x}),\, \xi\rangle + c\|h'(\bar{x})\xi\|_1 + c \sum_{i \in A(\bar{x})} \max\{0,\, \langle g_i'(\bar{x}),\, \xi\rangle\}. \tag{6.12}$$

In particular, the function $\varphi_c'(\bar{x};\, \cdot)$ is convex and piecewise linear, and hence, Lipschitz-continuous on $\mathbf{R}^n$.

As $\bar{x}$ is a local solution of (6.4), by Theorem 1.79 we obtain that

$$\varphi_c'(\bar{x};\, \xi) \geq 0 \quad \forall\, \xi \in \mathbf{R}^n,$$

and hence, $\xi = 0$ is a solution of the problem

$$\begin{array}{ll} \text{minimize} & \varphi_c'(\bar{x};\, \xi) \\ \text{subject to} & \xi \in \mathbf{R}^n. \end{array}$$

Therefore, again by Theorem 1.79,

$$0 \in \partial_\xi \varphi_c'(\bar{x};\, 0). \tag{6.13}$$

Using the facts from Proposition 1.55, Theorem 1.58 and its Corollary 1.59, and from Example 1.52, we derive the equality

$$\partial_\xi \varphi_c'(\bar{x};\, 0) = \left\{ f'(\bar{x}) + c(h'(\bar{x}))^{\mathrm{T}}\eta + c \sum_{i \in A(\bar{x})} \zeta_i g_i'(\bar{x}) \;\middle|\; \begin{array}{l} \eta \in \mathbf{R}^l,\ \|\eta\|_\infty \leq 1, \\ \zeta_i \in [0,\, 1],\ i \in A(\bar{x}) \end{array} \right\}.$$

Therefore, the inclusion (6.13) means the existence of $\bar{\eta} \in \mathbf{R}^l$, $\|\bar{\eta}\|_\infty \leq 1$, and $\bar{\zeta}_i \in [0,\, 1]$, $i \in A(\bar{x})$, such that

$$f'(\bar{x}) + c(h'(\bar{x}))^{\mathrm{T}}\bar{\eta} + c \sum_{i \in A(\bar{x})} \bar{\zeta}_i g_i'(\bar{x}) = 0. \tag{6.14}$$

Taking into account that $\bar{x} \in D$ and defining $\bar{\lambda} = c\bar{\eta}$, $\bar{\mu}_i = c\zeta_i$, $i \in A(\bar{x})$, $\bar{\mu}_i = 0$ for all $i \in \{1,\, \ldots,\, m\} \setminus A(\bar{x})$, from (6.14) it is immediate that $\bar{x}$ is a stationary point of problem (6.1), and $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ is an associated Lagrange multiplier satisfying (6.10). This concludes the proof. □

We now continue with conditions that are sufficient for exact penalization. We first show that the value of the Lagrangian $L : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}$,

$$L(x,\, \lambda,\, \mu) = f(x) + \langle \lambda,\, h(x) \rangle + \langle \mu,\, g(x) \rangle,$$

provides a lower bound for the penalty function if the penalty parameter is large enough.

Proposition 6.3. *For any $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$, any $x \in \mathbf{R}^n$, and any $(\lambda,\, \mu) \in \mathbf{R}^l \times \mathbf{R}^m_+$, it holds that*

$$L(x,\, \lambda,\, \mu) \leq f(x) + \|(\lambda,\, \mu)\|_\infty \psi(x), \tag{6.15}$$

where $\psi : \mathbf{R}^n \to \mathbf{R}_+$ is given by (6.5).

In particular, if $(\lambda,\, \mu)$ and c satisfy the inequality $\|(\lambda,\, \mu)\|_\infty \leq c$, then

$$L(x,\, \lambda,\, \mu) \leq \varphi_c(x), \tag{6.16}$$

where $\varphi_c : \mathbf{R}^n \to \mathbf{R}$ is given by (6.3).

Proof. For each $x \in \mathbf{R}^n$ and $(\lambda, \mu) \in \mathbf{R}^l \times \mathbf{R}^m_+$, it holds that

$$\begin{aligned}\langle \lambda, h(x)\rangle + \langle \mu, g(x)\rangle &\leq \|\lambda\|_\infty \|h(x)\|_1 + \sum_{i=1}^m \mu_i \max\{0, g_i(x)\}\\ &\leq \|\lambda\|_\infty \|h(x)\|_1 + \|\mu\|_\infty \sum_{i=1}^m \max\{0, g_i(x)\}\\ &\leq \|(\lambda, \mu)\|_\infty \|(h(x), \max\{0, g(x)\})\|_1\\ &= \|(\lambda, \mu)\|_\infty \psi(x).\end{aligned}$$

This immediately implies (6.15), and hence also (6.16) if $\|(\lambda, \mu)\|_\infty \leq c$. □

It is convenient to consider separately the convex and (possibly) nonconvex cases.

Theorem 6.4. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be convex on $\mathbf{R}^n$, let $h : \mathbf{R}^n \to \mathbf{R}^l$ be affine, and let the components of $g : \mathbf{R}^n \to \mathbf{R}^m$ be convex on $\mathbf{R}^n$. Assume that f and g are differentiable at $\bar{x} \in \mathbf{R}^n$, and that $\bar{x}$ is a stationary point of problem (6.1) with an associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$.*

Then for any c satisfying (6.10) the point $\bar{x}$ is a (global) solution in the penalized problem (6.4), where $\varphi_c : \mathbf{R}^n \to \mathbf{R}$ is given by (6.3), (6.5).

Proof. In the case under consideration, $L(\cdot, \bar{\lambda}, \bar{\mu})$ is a convex function (as a linear combination of convex functions with nonnegative coefficients; see Sect. A.2) and $\frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) = 0$. Thus, as discussed in Sect. 1.2.2, $\bar{x}$ is a global unconstrained minimizer of $L(\cdot, \bar{\lambda}, \bar{\mu})$, i.e.,

$$L(\bar{x}, \bar{\lambda}, \bar{\mu}) \leq L(x, \bar{\lambda}, \bar{\mu}) \quad \forall\, x \in \mathbf{R}^n.$$

Taking also into account that $\psi(\bar{x}) = 0$, $h(\bar{x}) = 0$ and $\langle \bar{\mu}, g(\bar{x})\rangle = 0$, we then obtain (for any $c \geq 0$) that

$$\begin{aligned}\varphi_c(\bar{x}) &= f(\bar{x}) + c\psi(\bar{x})\\ &= f(\bar{x})\\ &= f(\bar{x}) + \langle \bar{\lambda}, h(\bar{x})\rangle + \langle \bar{\mu}, g(\bar{x})\rangle\\ &= L(\bar{x}, \bar{\lambda}, \bar{\mu})\\ &\leq L(x, \bar{\lambda}, \bar{\mu}) \quad \forall\, x \in \mathbf{R}^n.\end{aligned} \tag{6.17}$$

On the other hand, since $\bar{\mu} \geq 0$, for c satisfying (6.10), from (6.16) in Proposition 6.3 we obtain that

$$L(x, \bar{\lambda}, \bar{\mu}) \leq \varphi_c(x) \quad \forall\, x \in \mathbf{R}^n.$$

Combining this with (6.17), we conclude that $\varphi_c(\bar{x}) \leq \varphi_c(x)$ holds for all $x \in \mathbf{R}^n$. □

Consider now the case when f and the components of g are not necessarily convex, assuming that a local solution of problem (6.1) satisfies the SOSC

$$\left\langle \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{6.18}$$

where

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\; g'_{A(\bar{x})}(\bar{x})\xi \leq 0,\; \langle f'(\bar{x}),\, \xi\rangle \leq 0\}. \tag{6.19}$$

Theorem 6.5. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be Lipschitz-continuous in a neighborhood of $\bar{x} \in \mathbf{R}^n$, and twice differentiable at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (6.1)*, satisfying the SOSC* (6.18) *for some associated Lagrange multiplier $(\bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$.*

Then for any c satisfying the inequality

$$\|(\bar{\lambda},\, \bar{\mu})\|_\infty < c, \tag{6.20}$$

the point $\bar{x}$ is a strict local solution of the penalized problem (6.4)*, where $\varphi_c : \mathbf{R}^n \to \mathbf{R}$ is given by* (6.3), (6.5).

Proof. Suppose that $\bar{x}$ is not a strict local solution of problem (6.4), i.e., that there exists a sequence $\{x^k\} \subset \mathbf{R}^n \setminus \{\bar{x}\}$ such that $\{x^k\}$ converges to $\bar{x}$, and

$$\varphi_c(x^k) \leq \varphi_c(\bar{x}) \quad \forall\, k. \tag{6.21}$$

For each k set $\xi^k = (x^k - \bar{x})/\|x^k - \bar{x}\|$ and $t_k = \|x^k - \bar{x}\|$. We can assume, without loss of generality, that the bounded sequence $\{\xi^k\}$ converges to some $\xi \in \mathbf{R}^n$, $\|\xi\| = 1$. Note that

$$x^k = \bar{x} + t_k\xi^k = \bar{x} + t_k\xi + t_k(\xi^k - \xi) = \bar{x} + t_k\xi + o(t_k) \tag{6.22}$$

as $k \to \infty$. In the assumptions of the theorem, $\varphi_c(\cdot)$ is Lipschitz-continuous in a neighborhood of $\bar{x}$. Hence,

$$\varphi_c(x^k) - \varphi_c(\bar{x} + t_k\xi) = O(\|x^k - \bar{x} - t_k\xi\|) = o(t_k)$$

as $k \to \infty$. Since according to assertion (a) of Proposition 6.1 φ_c is differentiable at $\bar{x}$ in any direction, we then obtain that

$$\begin{aligned} \varphi'_c(\bar{x};\, \xi) &= \lim_{k\to\infty} \frac{\varphi_c(\bar{x} + t_k\xi) - \varphi_c(\bar{x})}{t_k} \\ &= \lim_{k\to\infty} \frac{\varphi_c(x^k) - \varphi_c(\bar{x}) + o(t_k)}{t_k} \leq 0, \end{aligned}$$

where the inequality is by (6.21). Since $\bar{x}$ is feasible in problem (6.1), from relation (6.9) in Proposition 6.1 it then follows that

$$\langle f'(\bar{x}),\, \xi\rangle + c\|h'(\bar{x})\xi\|_1 + c \sum_{i\in A(\bar{x})} \max\{0,\, \langle g_i'(\bar{x}),\, \xi\rangle\} \leq 0. \tag{6.23}$$

In particular, as the last two terms in the left-hand side are nonnegative, we conclude that

$$\langle f'(\bar{x}),\, \xi\rangle \leq 0. \tag{6.24}$$

On the other hand, since $(\bar{\lambda},\, \bar{\mu})$ is a Lagrange multiplier associated with $\bar{x}$, it holds that

$$f'(\bar{x}) + (h'(\bar{x}))^{\mathrm{T}}\bar{\lambda} + \sum_{i\in A(\bar{x})} \bar{\mu}_i g_i'(\bar{x}) = 0.$$

Multiplying both sides of this relation by ξ and using (6.23), we obtain that

$$\begin{aligned} 0 &\geq -\langle\bar{\lambda},\, h'(\bar{x})\xi\rangle - \sum_{i\in A(\bar{x})} \bar{\mu}_i\langle g_i'(\bar{x}),\, \xi\rangle \\ &\quad +c\|h'(\bar{x})\xi\|_1 + c\sum_{i\in A(\bar{x})} \max\{0,\, \langle g_i'(\bar{x}),\, \xi\rangle\} \\ &\geq (c - \|(\bar{\lambda},\, \bar{\mu})\|_\infty)\left(\|h'(\bar{x})\xi\|_1 + \sum_{i\in A(\bar{x})} \max\{0,\, \langle g_i'(\bar{x}),\, \xi\rangle\}\right), \end{aligned}$$

where the condition $\bar{\mu} \geq 0$ has been used. Employing (6.20), we then conclude that $h'(\bar{x})\xi = 0$, $g'_{A(\bar{x})}(\bar{x})\xi \leq 0$. Combining this with (6.24), by (6.19) we further derive the inclusion $\xi \in C(\bar{x})$. Therefore, as $\xi \neq 0$, the SOSC (6.18) yields

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi\right\rangle > 0. \tag{6.25}$$

Using the equality $\frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}) = 0$ and (6.22), (6.25), we now obtain that

$$\begin{aligned} L(x^k,\, \bar{\lambda},\, \bar{\mu}) &= L(\bar{x},\, \bar{\lambda},\, \bar{\mu}) + \left\langle \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu}),\, x^k - \bar{x}\right\rangle \\ &\quad +\frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})(x^k - \bar{x}),\, x^k - \bar{x}\right\rangle + o(\|x^k - \bar{x}\|^2) \\ &= L(\bar{x},\, \bar{\lambda},\, \bar{\mu}) + \frac{t_k^2}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi\right\rangle + o(t_k^2) \\ &> L(\bar{x},\, \bar{\lambda},\, \bar{\mu}) \end{aligned} \tag{6.26}$$

for all k large enough.

Then using (6.21) and (6.26), for all k large enough we have that

$$\varphi_c(x^k) \leq \varphi_c(\bar{x}) = f(\bar{x}) = L(\bar{x},\, \bar{\lambda},\, \bar{\mu}) < L(x^k,\, \bar{\lambda},\, \bar{\mu}) \leq \varphi_c(x^k),$$

where the last inequality is by relation (6.16) in Proposition 6.3. This contradiction completes the proof. □

The results presented above indicate that, with proper care with respect to choosing the penalty parameter, exact penalty functions can serve as a basis for constructing descent methods for constrained optimization problems. Taking a somewhat different point of view, exact penalty functions can be used as merit functions for evaluating candidate points generated by algorithms that may not necessarily aim explicitly at minimizing the penalty function itself. In presenting globalizations of Newton-type methods below, we adopt the latter approach, more natural in this context.

We continue our discussion of merit functions for constrained optimization with the nonsmooth augmented Lagrangian, introduced for the equality-constrained problem

$$\begin{array}{ll}\text{minimize} & f(x)\\ \text{subject to} & h(x) = 0.\end{array} \tag{6.27}$$

Let $L : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}$ be the Lagrangian of this problem:

$$L(x,\, \lambda) = f(x) + \langle \lambda,\, h(x)\rangle.$$

Define the family of functions $\varphi_{c,\,\eta} : \mathbf{R}^n \to \mathbf{R}$,

$$\varphi_{c,\,\eta}(x) = L(x,\, \eta) + c\|h(x)\|_1 = f(x) + \langle \eta,\, h(x)\rangle + c\|h(x)\|_1, \tag{6.28}$$

where $c \geq 0$ and $\eta \in \mathbf{R}^l$ are parameters. This function will be used in Sect. 6.2.2 for restoring local superlinear convergence of a globalized SQP method, along the lines of [25]. Its principal properties are the following.

Proposition 6.6. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice differentiable at a point $\bar{x} \in \mathbf{R}^n$. Let $\bar{x}$ be a stationary point of problem (6.27), satisfying the SOSC*

$$\left\langle \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in \ker h'(\bar{x}) \setminus \{0\} \tag{6.29}$$

for some associated Lagrange multiplier $\bar{\lambda} \in \mathbf{R}^l$.

Then there exists $\bar{c} \geq 0$ such that for any $c > \bar{c}$ and any $\eta \in \mathbf{R}^l$ satisfying

$$\|\eta - \bar{\lambda}\|_\infty < c, \tag{6.30}$$

the point $\bar{x}$ is a strict local solution of the problem

$$\begin{array}{ll}\text{minimize} & \varphi_{c,\,\eta}(x)\\ \text{subject to} & x \in \mathbf{R}^n,\end{array}$$

where $\varphi_{c,\,\eta} : \mathbf{R}^n \to \mathbf{R}$ is given by (6.28).

Proof. Recall that according to Proposition 4.8, under the stated assumptions, for $c > 0$ large enough $\bar{x}$ is a strict local unconstrained minimizer of the function $L_c(\cdot,\, \bar{\lambda})$, where $L_c : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}$ is the (smooth) augmented Lagrangian:

$$L_c(x, \lambda) = L(x, \lambda) + \frac{c}{2}\|h(x)\|^2.$$

Hence,

$$\begin{aligned}\varphi_{c,\eta}(\bar{x}) = f(\bar{x}) &= L(\bar{x}, \bar{\lambda}) + \frac{c}{2}\|h(\bar{x})\|^2 \\ &< L(x, \bar{\lambda}) + \frac{c}{2}\|h(x)\|^2 \qquad (6.31)\end{aligned}$$

for all $x \in \mathbf{R}^n \setminus \{\bar{x}\}$ close enough to $\bar{x}$.

On the other hand, since $\|h(x)\|^2 = o(\|h(x)\|_1)$ as $x \to \bar{x}$, employing (6.30), we derive that

$$\begin{aligned}\varphi_{c,\eta}(x) &= L(x, \bar{\lambda}) + \langle \eta - \bar{\lambda}, h(x)\rangle + c\|h(x)\|_1 \\ &\geq L(x, \bar{\lambda}) + (c - \|\eta - \bar{\lambda}\|_\infty)\|h(x)\|_1 \\ &\geq L(x, \bar{\lambda}) + \frac{c}{2}\|h(x)\|^2 \qquad (6.32)\end{aligned}$$

for all $x \in \mathbf{R}^n$ close enough to $\bar{x}$. Combining now (6.31) and (6.32), the assertion follows. □

Another candidate which could be considered for the role of a merit function for problem (6.27) is the usual smooth augmented Lagrangian $L_c : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}$,

$$L_c(x, \lambda) = L(x, \lambda) + \frac{c}{2}\|h(x)\|^2.$$

Indeed, according to Proposition 4.8, any stationary point $\bar{x} \in \mathbf{R}^n$ of problem (6.27) satisfying the SOSC (6.29) for some associated Lagrange multiplier $\bar{\lambda} \in \mathbf{R}^l$ is a strict local solution of the unconstrained optimization problem

$$\begin{array}{ll}\text{minimize} & L_c(x, \bar{\lambda}) \\ \text{subject to} & x \in \mathbf{R}^n\end{array}$$

for all $c > 0$ large enough. Therefore, $L_c(\cdot, \bar{\lambda})$ possesses the exact penalization property. However, the possibility to use the augmented Lagrangian as a merit function is at least questionable because of the need to know the exact Lagrange multiplier $\bar{\lambda}$. For λ arbitrarily close to $\bar{\lambda}$ the function $L_c(\cdot, \lambda)$ is not necessarily an exact penalty for (6.27).

Recall also that, as discussed in Sect. 4.1.2, the pairs $(\bar{x}, \bar{\lambda})$ of stationary points and associated Lagrange multipliers of problem (6.27) are stationary in the problem

$$\begin{array}{ll}\text{minimize} & L_c(x, \lambda) \\ \text{subject to} & (x, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l.\end{array}$$

However, such points need not be local solutions of this problem, which again makes it not obvious how L_c can be used as a (primal-dual) merit function.

Nevertheless, there exist algorithms employing L_c in this role, and perhaps the most well-known and successful development of this kind is the SNOPT solver; see [101, 102].

6.2 Linesearch Methods

We next discuss globalization using linesearch for an appropriately chosen merit function. An important issue in this approach is preserving fast convergence rate of the local algorithm in its globalized modification. It turns out that this is not straightforward (the so-called Maratos effect). We present two different techniques that can be employed to safeguard the superlinear local convergence rate in the globalized method: the second-order corrections, and using the nonsmooth augmented Lagrangian as a merit function. Another difficulty for globalization is the possibility that subproblems of the algorithm may be infeasible when far from a solution with favorable properties. To deal with this issue, the so-called elastic mode is discussed.

6.2.1 Globalization of Sequential Quadratic Programming

Globalization strategy for the SQP algorithm based on linesearch for nonsmooth penalty functions dates back to [114]. The key observation for this construction is that the primal direction given by solving an SQP subproblem is a direction of descent for the penalty function, provided some conditions are satisfied. Specifically, the matrix of the quadratic objective function of the subproblem should be positive definite, and the penalty parameter should be greater than the norm of the multipliers associated with the subproblem solution.

Consider again the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \ g(x) \le 0, \end{array} \tag{6.33}$$

where $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are smooth.

Recall (see Sect. 4.2) that the SQP subproblem is given by

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, p\rangle + \dfrac{1}{2}\langle H_k p,\, p\rangle \\ \text{subject to} & h(x^k) + h'(x^k)p = 0, \ g(x^k) + g'(x^k)p \le 0, \end{array} \tag{6.34}$$

where $H_k \in \mathbf{R}^{n\times n}$ is a symmetric matrix. The solution p^k of this subproblem is then the direction of change for the primal variables. The merit function for linesearch to be used in the sequel is the l_1-penalty function $\varphi_c : \mathbf{R}^n \to \mathbf{R}$ introduced in Sect. 6.1:

$$\varphi_c(x) = f(x) + c\psi(x), \tag{6.35}$$

where $c > 0$ is the penalty parameter and $\psi : \mathbf{R}^n \to \mathbf{R}_+$ is the l_1-penalty for the constraints, given by

$$\psi(x) = \|h(x)\|_1 + \sum_{i=1}^{m} \max\{0,\, g_i(x)\}. \tag{6.36}$$

Before we formally state the algorithm, some comments are in order. The feasibility of subproblems (6.34) had been established in Sect. 4.2 (as part of their solvability) in a neighborhood of a solution with certain properties. Globally, there is no guarantee that (6.34) would always be feasible (the exception is the case when h is affine and the components of g are convex, or when the gradients of all the constraints are globally linearly independent). For now, we shall assume that subproblems are feasible. This issue would be discussed again later on.

Furthermore, even if the subproblem (6.34) is feasible, there is still no guarantee that it has solutions (or stationary points), unless H_k is positive definite. This matrix, however, is a controlled choice within the algorithm, and one way to ensure solvability of subproblems and global convergence is to take it positive definite. That said, such a choice may enter in contradiction with the requirements needed for the local superlinear rate of convergence; we shall not deal with this issue at this stage, but it will be addressed later on. The goal of the current development is to ensure global convergence (in some sense) of a method based on solving subproblems (6.34), to stationary points of the original problem (6.33). The issue of how the introduced modifications interfere with convergence rate, and what can be done to avoid the undesirable phenomena, would be considered in Sect. 6.2.2.

Algorithm 6.7 Choose the parameters $\bar{c} > 0$, $\sigma \in (0, 1)$ and $\theta \in (0, 1)$. Choose $(x^0, \lambda^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ and set $k = 0$.

1. Choose a symmetric matrix $H_k \in \mathbf{R}^{n\times n}$ and compute $p^k \in \mathbf{R}^n$ as a stationary point of problem (6.34), and an associated Lagrange multiplier $(\lambda^{k+1}, \mu^{k+1}) \in \mathbf{R}^l \times \mathbf{R}^m$.
2. If $p^k = 0$, stop.
3. Choose
$$c_k \geq \|(\lambda^{k+1},\, \mu^{k+1})\|_\infty + \bar{c} \tag{6.37}$$
and compute
$$\Delta_k = \langle f'(x^k),\, p^k\rangle - c_k\psi(x^k), \tag{6.38}$$
where $\psi : \mathbf{R}^n \to \mathbf{R}_+$ is defined by (6.36).

4. Set $\alpha = 1$. If the inequality

$$\varphi_{c_k}(x^k + \alpha p^k) \leq \varphi_{c_k}(x^k) + \sigma\alpha\Delta_k \tag{6.39}$$

with $\varphi_{c_k} : \mathbf{R}^n \to \mathbf{R}$ defined by (6.35) is satisfied, set $\alpha_k = \alpha$. Otherwise, replace α by $\theta\alpha$, check the inequality (6.39) again, etc., until (6.39) becomes valid.
5. Set $x^{k+1} = x^k + \alpha_k p^k$.
6. Increase k by 1 and go to step 1.

As already noted in Sect. 6.1, other penalty terms could be used instead of that based on the l_1-norm. However, it is important that the norm appearing in the choice of the penalty parameter in (6.37) should be dual (in the sense of (6.7)) to the norm defining the penalty term.

We first show that solving the SQP subproblem (6.34) gives a direction of descent for the penalty function, whenever H_k is positive definite and $p^k \neq 0$ (as is seen below, $p^k = 0$ means that the current iterate x^k is a stationary point of the original problem (6.33), so the method naturally stops in that case).

Lemma 6.8. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable at a point $x^k \in \mathbf{R}^n$. Let $H_k \in \mathbf{R}^{n\times n}$ be a symmetric matrix, let $p^k \in \mathbf{R}^n$ be a stationary point of problem (6.34), let $(\lambda^{k+1}, \mu^{k+1}) \in \mathbf{R}^l \times \mathbf{R}^m$ be an associated Lagrange multiplier, and let c_k satisfy (6.37). Finally, let $\varphi_{c_k} : \mathbf{R}^n \to \mathbf{R}$ and $\psi : \mathbf{R}^n \to \mathbf{R}_+$ be defined by (6.35) and (6.36), respectively.*

Then it holds that

$$\varphi'_{c_k}(x^k;\, p^k) \leq \Delta_k \leq -\langle H_k p^k,\, p^k\rangle - \bar{c}\psi(x^k), \tag{6.40}$$

where Δ_k is given by (6.38).

Recall that according to Proposition 6.1, under the assumptions of this lemma φ_{c_k} is differentiable at x^k in every direction, and hence, the relation (6.40) is meaningful. This relation shows, in particular, that Δ_k is the upper estimate of $\varphi'_{c_k}(x^k;\, p^k)$. Moreover, the proof below puts in evidence that this estimate is exact in the absence of inequality constraints. Therefore, (6.39) can be regarded as a version of the Armijo inequality.

Proof. The triple $(p^k,\, \lambda^{k+1},\, \mu^{k+1})$ satisfies the KKT system for problem (6.34), i.e., it holds that

$$\begin{gathered} f'(x^k) + H_k p^k + (h'(x^k))^{\mathrm{T}}\lambda^{k+1} + (g'(x^k))^{\mathrm{T}}\mu^{k+1} = 0, \\ h(x^k) + h'(x^k)p^k = 0, \\ \mu^{k+1} \geq 0, \quad g(x^k) + g'(x^k)p^k \leq 0, \quad \langle \mu^{k+1},\, g(x^k) + g'(x^k)p^k\rangle = 0. \end{gathered} \tag{6.41}$$

Define the index sets

$$\begin{aligned}
J_k^- &= \{j = 1, \ldots, l \mid h_j(x^k) < 0\}, \\
J_k^0 &= \{j = 1, \ldots, l \mid h_j(x^k) = 0\}, \\
J_k^+ &= \{j = 1, \ldots, l \mid h_j(x^k) > 0\}, \\
I_k^0 &= \{i = 1, \ldots, m \mid g_i(x^k) = 0\}, \\
I_k^+ &= \{i = 1, \ldots, m \mid g_i(x^k) > 0\}.
\end{aligned}$$

From the second equality in (6.41) we obtain that

$$-|h_j(x^k)| = \begin{cases} -\langle h_j'(x^k),\, p^k\rangle & \text{if } j \in J_k^-, \\ |\langle h_j'(x^k),\, p^k\rangle| & \text{if } j \in J_k^0, \\ \langle h_j'(x^k),\, p^k\rangle & \text{if } j \in J_k^+. \end{cases}$$

Next, from the second inequality in (6.41), we have that

$$-\max\{0,\, g_i(x^k)\} = \begin{cases} 0 = \max\{0,\, \langle g_i'(x^k),\, p^k\rangle\} & \text{if } i \in I_k^0, \\ -g_i(x^k) \geq \langle g_i'(x^k),\, p^k\rangle & \text{if } i \in I_k^+. \end{cases}$$

Using Proposition 6.1, it follows that

$$\begin{aligned}
\varphi_{c_k}'(x^k;\, p^k) &= \langle f'(x^k),\, p^k\rangle + c_k\psi'(x^k;\, p^k) \\
&\leq \langle f'(x^k),\, p^k\rangle - c_k\|h(x^k)\|_1 - c_k \sum_{i \in I_k^0 \cup I_k^+} \max\{0,\, g_i(x^k)\} \\
&= \langle f'(x^k),\, p^k\rangle - c_k\|h(x^k)\|_1 - c_k \sum_{i=1}^{m} \max\{0,\, g_i(x^k)\} \\
&= \langle f'(x^k),\, p^k\rangle - c_k\psi(x^k) \\
&= \Delta_k,
\end{aligned}$$

where the second equality is by the fact that for each $i \in \{1, \ldots, m\} \setminus (I_k^0 \cup I_k^+)$ it holds that $\max\{0,\, g_i(x^k)\} = 0$. This shows the first inequality in (6.40).

Multiplying both sides of the first equality in (6.41) by p^k and making use of the second and the last equalities in (6.41), we derive that

$$\begin{aligned}
\langle f'(x^k),\, p^k\rangle &= -\langle H_k p^k,\, p^k\rangle - \langle \lambda^{k+1},\, h'(x^k)p^k\rangle - \langle \mu^{k+1},\, g'(x^k)p^k\rangle \\
&= -\langle H_k p^k,\, p^k\rangle + \langle \lambda^{k+1},\, h(x^k)\rangle + \langle \mu^{k+1},\, g(x^k)\rangle \\
&\leq -\langle H_k p^k,\, p^k\rangle + \|\lambda^{k+1}\|_\infty \|h(x^k)\|_1 + \sum_{i=1}^{m} \mu_i^{k+1} \max\{0,\, g_i(x^k)\} \\
&\leq -\langle H_k p^k,\, p^k\rangle + \|\lambda^{k+1}\|_\infty \|h(x^k)\|_1 \\
&\quad + \|\mu_i^{k+1}\|_\infty \sum_{i=1}^{m} \max\{0,\, g_i(x^k)\} \\
&\leq -\langle H_k p^k,\, p^k\rangle + \|(\lambda^{k+1},\, \mu^{k+1})\|_\infty\, \psi(x^k).
\end{aligned}$$

Hence, using (6.36)–(6.38), we obtain that

$$\begin{aligned}\Delta_k &= \langle f'(x^k),\, p^k\rangle - c_k\psi(x^k)\\ &\leq \langle f'(x^k),\, p^k\rangle - (\|(\lambda^{k+1},\, \mu^{k+1})\|_\infty + \bar{c})\psi(x^k)\\ &\leq -\langle H_k p^k,\, p^k\rangle - \bar{c}\psi(x^k),\end{aligned}$$

which gives the second inequality in (6.40). □

It is now clear from (6.40) that if H_k is positive definite and Algorithm 6.7 generates $p^k \neq 0$, then

$$\Delta_k < 0, \tag{6.42}$$

and p^k is a descent direction for φ_{c_k} at the point x^k. Hence, the linesearch procedure in step 2 of the algorithm is well defined, in the sense that it generates some stepsize value $\alpha_k > 0$ after a finite number of backtrackings. On the other hand, if $p^k = 0$ is generated for some iteration index k (so that the method stops) conditions (6.41) show that $(x^k,\, \lambda^{k+1},\, \mu^{k+1})$ satisfies the KKT system of the original problem (6.33).

In principle, iterations of Algorithm 6.7 use variable values of the penalty parameter c_k. In the convergence analysis, it is standard to assume that this value becomes fixed, at least for all k sufficiently large: $c_k = c$, where $c > 0$ is a constant. Note that if this happens, then starting from some index k Algorithm 6.7 can be thought of as a descent method for the unconstrained optimization problem

$$\begin{array}{ll}\text{minimize} & \varphi_c(x)\\ \text{subject to} & x \in \mathbf{R}^n.\end{array} \tag{6.43}$$

Recall that according to the results in Sect. 6.1, under certain conditions (including (6.10)) the l_1-penalization is exact, and the fact that (6.37) holds for $c_k = c$ fixed for all k large enough indicates that the value of the penalty parameter being used is indeed adequate.

As is easy to see from (6.37), the possibility to use c_k fixed from some iteration on depends on boundedness of the dual sequence $\{(\lambda^k,\, \mu^k)\}$ generated by the method. From the practical viewpoint, boundedness of the generated sequence of multipliers is quite natural, at least when the multiplier set of the original problem is bounded. In fact, this can indeed be guaranteed under certain assumptions. If $\{(\lambda^k,\, \mu^k)\}$ is bounded, the validity of $c_k = c$ for some c and all k large can be ensured, for example, by the following simple rule to update c_k. On the first iteration, set $c_0 = \|(\lambda^1,\, \mu^1)\|_\infty + \bar{c} + \delta$ with some $\delta > 0$. For each $k = 1,\, 2,\, \ldots$, check (6.37) for $c_k = c_{k-1}$. If (6.37) is satisfied, take this c_k; otherwise, take $c_k = \|(\lambda^{k+1},\, \mu^{k+1})\|_\infty + \bar{c} + \delta$. Of course, more sophisticated rules are possible (and are usually preferred in practical implementations; see, e.g., [29, Sect. 17.1], [208, Sect. 18.3]).

As usual, we define the Lagrangian $L : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}$ of problem (6.33) by

$$L(x,\, \lambda,\, \mu) = f(x) + \langle \lambda,\, h(x)\rangle + \langle \mu,\, g(x)\rangle.$$

The basic global convergence properties of the SQP method are given by Theorem 6.9 below. Note again that the assumption therein that $c_k = c$ for all k large enough implies that the sequence $\{(\lambda^k, \mu^k)\}$ is bounded. The assumptions that the primal and/or the dual sequences are bounded (or that the penalty parameter is fixed from some iteration on) are typical for showing global convergence of SQP methods. Such boundedness assumptions can only be removed for more special problems (e.g., in the case of convex constraints) and after making, in addition, some modifications to the algorithm; e.g., [249].

Theorem 6.9. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable on $\mathbf{R}^n$, with their derivatives being Lipschitz-continuous on $\mathbf{R}^n$. Assume further that there exist $\gamma > 0$ and $\Gamma > 0$ such that the matrices H_k in Algorithm 6.7 satisfy*

$$\langle H_k\xi,\, \xi\rangle \geq \gamma\|\xi\|^2 \quad \forall\, \xi \in \mathbf{R}^n, \quad \|H_k\| \leq \Gamma \quad \forall\, k. \tag{6.44}$$

Let $\{(x^k,\, \lambda^k,\, \mu^k)\}$ be any sequence generated by this algorithm and assume that $c_k = c$ holds with some $c > 0$ for all k sufficiently large.

Then as $k \to \infty$, it holds that either

$$\varphi_{c_k}(x^k) \to -\infty, \tag{6.45}$$

or

$$\begin{gathered}\{p^k\} \to 0, \quad \left\{\frac{\partial L}{\partial x}(x^k,\, \lambda^{k+1},\, \mu^{k+1})\right\} \to 0, \quad \{h(x^k)\} \to 0,\\ \{\max\{0,\, g(x^k)\}\} \to 0, \quad \mu_i^{k+1} g_i(x^k) \to 0,\ i = 1,\, \ldots,\, m.\end{gathered} \tag{6.46}$$

Moreover, for every accumulation point $(\bar{x},\, \bar{\lambda},\, \bar{\mu}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ of the sequence $\{(x^k,\, \lambda^k,\, \mu^k)\}$ it holds that $\bar{x}$ is a stationary point of problem (6.33), and $(\bar{\lambda},\, \bar{\mu})$ is an associated Lagrange multiplier.

Proof. Note first that by (6.37), and by the assumption that $c_k = c$ for all k large enough, the sequences $\{\lambda^k\}$ and $\{\mu^k\}$ are bounded.

We next show that the sequence of stepsizes $\{\alpha_k\}$ is bounded away from zero. Let $\alpha \in (0,\, 1]$ be arbitrary. Let $\ell > 0$ be the Lipschitz constant for the gradient of f and for the gradients of the components of h and g on $\mathbf{R}^n$. By Lemma A.11, for all k and each $i = 1,\, \ldots,\, m$, it holds that

$$\begin{aligned}&\max\{0,\, g_i(x^k + \alpha p^k)\} - \max\{0,\, g_i(x^k) + \alpha\langle g_i'(x^k),\, p^k\rangle\}\\ &\leq \max\{0,\, g_i(x^k + \alpha p^k) - g_i(x^k) - \alpha\langle g_i'(x^k),\, p^k\rangle\}\\ &\leq |g_i(x^k + \alpha p^k) - g_i(x^k) - \alpha\langle g_i'(x^k),\, p^k\rangle|\\ &\leq \frac{\ell\alpha^2}{2}\|p^k\|^2,\end{aligned} \tag{6.47}$$

where we have used the easily verifiable relations

$$\begin{array}{c}\max\{0,\, a\} - \max\{0,\, b\} \le \max\{0,\, a-b\} \quad \forall\, a,\, b \in \mathbf{R}, \\ \max\{0,\, a\} \le |a| \quad \forall\, a,\, b \in \mathbf{R}.\end{array} \tag{6.48}$$

Furthermore,

$$\begin{aligned}&\max\{0,\, g_i(x^k) + \alpha\langle g_i'(x^k),\, p^k\rangle\}\\ &= \max\{0,\, \alpha(g_i(x^k) + \langle g_i'(x^k),\, p^k\rangle) + (1-\alpha)g_i(x^k)\}\\ &\le \alpha \max\{0,\, g_i(x^k) + \langle g_i'(x^k),\, p^k\rangle\} + (1-\alpha)\max\{0,\, g_i(x^k)\}\\ &= (1-\alpha)\max\{0,\, g_i(x^k)\},\end{aligned}$$

where the inequality is by the convexity of the function $\max\{0,\, \cdot\}$, and the last equality follows from the second inequality in (6.41). From (6.47) we then obtain that

$$\begin{aligned}\max\{0,\, g_i(x^k + \alpha p^k)\} &\le \max\{0,\, g_i(x^k) + \alpha\langle g_i'(x^k),\, p^k\rangle\} + \frac{\ell\alpha^2}{2}\|p^k\|^2\\ &\le (1-\alpha)\max\{0,\, g_i(x^k)\} + \frac{\ell\alpha^2}{2}\|p^k\|^2.\end{aligned} \tag{6.49}$$

By Lemma A.11, it also holds that

$$f(x^k + \alpha p^k) \le f(x^k) + \alpha\langle f'(x^k),\, p^k\rangle + \frac{\ell\alpha^2}{2}\|p^k\|^2. \tag{6.50}$$

Similarly, by the second equality in (6.41) and Lemma A.11, it holds that

$$\begin{aligned}|h(x^k + \alpha p^k)| &= |h(x^k + \alpha p^k) - \alpha h(x^k) - \alpha h'(x^k)p^k|\\ &\le (1-\alpha)|h(x^k)| + \frac{\ell\alpha^2}{2}\|p^k\|^2.\end{aligned} \tag{6.51}$$

Using (6.35), (6.36), and combining the relations (6.49)–(6.51), we obtain that

$$\begin{aligned}\varphi_{c_k}(x^k + \alpha p^k) &= f(x^k + \alpha p^k) + c_k\|h(x^k + \alpha p^k)\|_1\\ &\quad + c_k\sum_{i=1}^m \max\{0,\, g_i(x^k + \alpha p^k)\}\\ &\le f(x^k) + \alpha\langle f'(x^k),\, p^k\rangle + c_k(1-\alpha)\|h(x^k)\|_1\\ &\quad + c_k(1-\alpha)\sum_{i=1}^m \max\{0,\, g_i(x^k)\} + C_k\alpha^2\|p^k\|^2\\ &= \varphi_{c_k}(x^k) + \alpha\langle f'(x^k),\, p^k\rangle - \alpha c_k\psi(x^k) + C_k\alpha^2\|p^k\|^2\\ &= \varphi_{c_k}(x^k) + \alpha\Delta_k + C_k\alpha^2\|p^k\|^2,\end{aligned} \tag{6.52}$$

where

$$C_k = \frac{\ell}{2}(1 + (l+m)c_k) > 0. \tag{6.53}$$

Comparing (6.52) with (6.39), we conclude that the inequality (6.39) is satisfied if

$$\Delta_k + C_k\alpha\|p^k\|^2 \leq \sigma\Delta_k,$$

i.e., it is satisfied for all $\alpha \in (0,\, \bar{\alpha}_k]$, where

$$\bar{\alpha}_k = \frac{(\sigma - 1)\Delta_k}{C_k\|p^k\|^2}.$$

By the first inequality in (6.40), and by the first condition in (6.44), it holds that

$$\bar{\alpha}_k \geq \frac{(1-\sigma)\gamma}{C_k},$$

where the sequence $\{C_k\}$ is bounded (by its definition (6.53), because c_k is constant for all k sufficiently large). We then conclude that there exists $\varepsilon > 0$, that does not depend on k, such that

$$\alpha_k \geq \varepsilon. \tag{6.54}$$

From the relations (6.39) and (6.54), we now obtain that for all k large enough it holds that

$$\varphi_{c_{k+1}}(x^{k+1}) \leq \varphi_{c_k}(x^k) + \sigma\varepsilon\Delta_k. \tag{6.55}$$

Recalling that $c_k = c$ for all k large enough, it holds that from some iteration on, the sequence $\{\varphi_c(x^k)\}$ is monotonically decreasing. Thus, either it is unbounded below, which gives (6.45), or it converges.

In the second case, from (6.55) we obtain that

$$\Delta_k \to 0 \text{ as } k \to \infty. \tag{6.56}$$

Using now the second inequality in (6.40) and (6.44), we conclude that as $k \to \infty$,

$$\psi(x^k) \to 0, \quad \{p^k\} \to 0, \quad \{H_k p^k\} \to 0. \tag{6.57}$$

In particular, by (6.36),

$$\{h(x^k)\} \to 0, \quad \{\max\{0,\, g(x^k)\}\} \to 0. \tag{6.58}$$

Also, passing onto the limit in the first equality in (6.41) and using (6.57), we obtain that

$$\left\{\frac{\partial L}{\partial x}(x^k,\, \lambda^{k+1},\, \mu^{k+1})\right\} \to 0.$$

We next verify that the complementarity condition holds in the limit, in the form of the last relation in (6.46). By (6.38), (6.56), and (6.57), we have that

$$\langle f'(x^k),\, p^k\rangle = \Delta_k + c_k\psi(x^k) \to 0$$

as $k \to \infty$, where we have again employed the assumption that c_k is constant from some point on. Multiplying the first equality in (6.41) by p^k and using the last relation together with (6.57), we obtain that

$$\langle \lambda^{k+1},\, h'(x^k)p^k\rangle + \langle \mu^{k+1},\, g'(x^k)p^k\rangle = -\langle f'(x^k),\, p^k\rangle - \langle H_k p^k,\, p^k\rangle \to 0$$

as $k \to \infty$. Since from the second relation in (6.41) and from (6.58) it follows that

$$\{h'(x^k)p^k\} \to 0,$$

recalling that $\{\lambda^k\}$ is bounded we conclude that

$$\langle \mu^{k+1},\, g'(x^k)p^k\rangle \to 0.$$

Then from the last equality in (6.41) we obtain that

$$\langle \mu^{k+1},\, g(x^k)\rangle \to 0 \tag{6.59}$$

as $k \to \infty$. As the sequence $\{\mu^{k+1}\}$ is nonnegative and bounded, from the second relation in (6.58) it is immediate that

$$\limsup_{k\to\infty} \mu_i^{k+1} g_i(x^k) \le 0 \quad \forall\, i = 1,\, \ldots,\, m. \tag{6.60}$$

Therefore, if we show that

$$\liminf_{k\to\infty} \mu_i^{k+1} g_i(x^k) \ge 0 \quad \forall\, i = 1,\, \ldots,\, m, \tag{6.61}$$

the last relation in (6.46) would follow.

To that end, suppose the contrary to (6.61), i.e., that there exist some index $s \in \{1,\, \ldots,\, m\}$, some subsequence $\{x^{k_j}\}$ and a constant $\delta > 0$ such that $\mu_s^{k_j+1} g_s(x^{k_j}) \le -\delta$ for all j large enough. Then, employing (6.59), it follows that

$$\begin{aligned} 0 &= \lim_{k\to\infty} \langle \mu^{k+1},\, g(x^k)\rangle \\ &= \lim_{j\to\infty} \sum_{i=1}^{m} \mu_i^{k_j+1} g_i(x^{k_j}) \\ &\le -\delta + \sum_{\substack{i=1\\ i\neq s}}^{m} \limsup_{j\to\infty} \mu_i^{k_j+1} g_i(x^{k_j}) \\ &< 0, \end{aligned}$$

where the last inequality follows from (6.60). This contradiction establishes (6.61) and, hence, completes the proof of (6.46).

Finally, by the first relation in (6.46), and by the fact that the sequence $\{\alpha_k\}$ is bounded (it is contained in $[0,\, 1]$) it holds that if $\{x^{k_j}\}$ converges to

$\bar{x}$, then $\{x^{k_j-1}\}$ also converges to the same $\bar{x}$ (as $x^{k-1} = x^k - \alpha_{k-1}p^{k-1}$ for all k). Therefore, any accumulation point of $\{(x^k, \lambda^k, \mu^k)\}$ is also an accumulation point of $\{(x^{k-1}, \lambda^k, \mu^k)\}$. At the same time, from the other limiting relations in (6.46) it follows that for each accumulation point $(\bar{x}, \bar{\lambda}, \bar{\mu})$ of $\{(x^{k-1}, \lambda^k, \mu^k)\}$ it holds that $\bar{x}$ is a stationary point of the problem (6.33), and $(\bar{\lambda}, \bar{\mu})$ is an associated Lagrange multiplier. Therefore, this also holds for any accumulation point of $\{(x^k, \lambda^k, \mu^k)\}$, which completes the proof. □

As already commented, the analysis presented above assumes that the SQP subproblems of the form (6.34) are feasible. Except when the constraints are convex or constraints gradients are linearly independent at every point, feasibility of subproblems at every iteration cannot be guaranteed. To deal with the case of infeasibility, one way is to replace the subproblem (6.34) with the following "elastic mode" modification:

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k), p\rangle + c_k t + \dfrac{1}{2}\langle H_k p, p\rangle \\ \text{subject to} & -te \leq h(x^k) + h'(x^k)p \leq te, \; g(x^k) + g'(x^k)p \leq te, \; t \geq 0, \end{array} \tag{6.62}$$

where $c_k > 0$ is the penalty parameter, e is the vector of ones of an appropriate dimension, and the variables are $(p, t) \in \mathbf{R}^n \times \mathbf{R}$.

Define the penalty function $\varphi_{c_k} : \mathbf{R}^n \to \mathbf{R}$ according to (6.35) with the l_∞-penalty $\psi : \mathbf{R}^n \to \mathbf{R}_+$ given by

$$\psi(x) = \max\{\|h(x)\|_\infty, 0, g_1(x), \ldots, g_m(x)\}.$$

It can be seen that if H_k is positive definite and x^k is not a stationary point of the original problem (6.33), then for a stationary point $(p^k, t_k) \in \mathbf{R}^n \times \mathbf{R}$ of the subproblem (6.62) it holds that $p^k \neq 0$ and it is a direction of descent for the function φ_{c_k} at the point x^k. This fact can serve as the basis for constructing a globalized SQP method along the lines of Algorithm 6.7 above, but employing subproblems of the form (6.62) and linesearch for l_∞-penalty function (rather than l_1).

However, introducing elastic mode comes with a price to pay. Specifically, it gives rise to some difficulties with controlling the penalty parameter, at least as a matter of the convergence theory. According to the discussion above, if one were to fix an arbitrary $c_k = c$, the resulting process would still be a well-defined descent method for solving the unconstrained penalized problem (6.43). But this problem is reasonably related to the original problem (6.33) only when the penalty parameter c is large enough, so that the penalty function φ_c is exact; otherwise, minimizers of the latter need not even be feasible in (6.33). According to Sect. 6.1, one can expect to asymptotically achieve the needed property for φ_{c_k} by taking c_k large enough compared to the appropriate norm of the available Lagrange multiplier estimates. The difficulty though is that in the elastic mode (6.62) the multipliers are computed *after* the penalty parameter is chosen. On a related note, one can check that, unlike for SQP without elastic mode, here the equality $p^k = 0$ (which means that no progress can be made from the current iterate) does not imply neither feasibility of the current iterate x^k nor the complementarity condition with the multipliers $(\lambda^{k+1}, \mu^{k+1})$, unless it holds that

$$c_k > \|(\lambda^{k+1}, \mu^{k+1})\|_1. \tag{6.63}$$

The problem with this is clear: the value $(\lambda^{k+1}, \mu^{k+1})$ is given by the Lagrange multiplier of the subproblem (6.62) which is solved *after* the value of c_k is already chosen. Thus, there is no clear way to ensure a condition like (6.63). This results in certain difficulties, both in practical implementations and in theoretical analysis.

We shall not discuss elastic mode for SQP in any more detail here. It will be analyzed in the context of globalization of SQCQP in Sect. 6.2.3, where elastic mode is indispensable even in the case of convex constraints. Many of the ingredients in the convergence analysis in Sect. 6.2.3 are similar to what would have been required for the elastic mode of SQP.

As already mentioned above, another potential issue with Theorem 6.9 that needs to be addressed is the requirement that H_k must be positive definite (the first condition in (6.44)). This means, for example, that one cannot adopt the basic choice

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k), \tag{6.64}$$

even when the true Hessian of the Lagrangian is available. This is because the Hessian need not be positive definite under any reasonable assumptions (other than some strong convexity assumptions), even locally. On the other hand, as discussed in Sect. 4.2, this basic choice is attractive from the point of view of the convergence rate. Thus, when computing the Hessian of the Lagrangian is relatively cheap, it can still be advantageous to try to use this true second-order information, with some safeguards. In [39], it is proposed to try first the basic choice (6.64) at the start of each iteration. If the corresponding SQP subproblem (6.34) is solvable and gives a direction of "sufficient descent" for the merit function (e.g., the directional derivative of the merit function in this directions is "sufficiently negative"), this direction is accepted. Otherwise, H_k is modified by successively adding to it the multiple of the identity matrix, until the direction with the needed descent properties is obtained.

Another possibility to deal with the issue of positive definiteness of H_k is to build these matrices by quasi-Newton approximations of the Hessian with Powell's correction (see Sect. 4.1). There exist indeed very successful implementations of quasi-Newton SQP methods with linesearch; for example, the SNOPT solver; see [102].

For equality-constrained problems of the form

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0 \end{array} \tag{6.65}$$

(perhaps with additional simple bounds), another possibility to tackle the lack of positive definiteness of the Hessian of the Lagrangian is suggested by Proposition 4.8. Specifically, one can employ the augmented Lagrangian choice of H_k, discussed in Sects. 4.2 and 4.3.2, i.e.,

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k) + c(h'(x^k))^{\mathrm{T}} h'(x^k)$$

with a sufficiently large penalty parameter $c \geq 0$, where $L : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}$,

$$L(x,\, \lambda) = f(x) + \langle \lambda,\, h(x) \rangle.$$

is the Lagrangian of problem (6.65).

Finally, as some other useful variants of the SQP algorithm with linesearch we mention its truncated versions (recall Sect. 4.3.2, where the local convergence of the truncated SQP method is analyzed). Globalized algorithms of this kind were developed, e.g., in [37, 39, 47, 160] for purely equality-constrained problems, and in [24, 147, 204] for problems with equality and inequality constraints.

6.2.2 Convergence Rate of Globalized Sequential Quadratic Programming

Whenever the globalized SQP Algorithm 6.7 converges, in principle the superlinear rate could be expected under the assumptions of Theorem 4.14 or Theorem 4.22, which include the SOSC

$$\left\langle \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{6.66}$$

where

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A(\bar{x})}(\bar{x})\xi \leq 0,\ \langle f'(\bar{x}),\, \xi \rangle \leq 0\}$$

is the critical cone of problem (6.33) at $\bar{x}$. But there are two issues here that require attention.

First, as already discussed, Theorem 6.9 asserting global convergence of Algorithm 6.7 requires the matrices H_k to be positive definite. According to Theorem 4.22, to achieve the superlinear convergence rate they must also approximate the Hessian of the Lagrangian, at least on the critical cone of the problem. As under the SOSC, required for superlinear convergence, the Hessian of the Lagrangian is positive definite on the critical cone, approximating it by positive definite matrices does not give any conceptual contradiction. Assuming that this approximation is good enough, Theorem 4.22 would guarantee superlinear convergence of Algorithm 6.7 if we could show that the linesearch procedure of the latter accepts the unit stepsize for all iterations with the index k large enough. It is precisely the latter property of accepting the unit stepsize which is the second issue requiring special consideration in this context.

Unlike in the case of unconstrained optimization (see Theorem 2.26), it turns out that there are situations when the unit stepsize in the SQP direction does not reduce the value of a given nonsmooth penalty function, even arbitrarily close to a solution and under all the assumptions needed for fast convergence of the local SQP method. Moreover, such situations are not uncommon. The phenomenon is usually referred to as the *Maratos effect* [188].

Example 6.10. Let $n = 2$, $l = 1$, $m = 0$, $f(x) = x_1$, $h(x) = x_1^2 + x_2^2 - 1$. The global solution of problem (6.33) with this data is $\bar{x} = (-1, 0)$; it satisfies the regularity condition

$$\operatorname{rank} h'(\bar{x}) = l, \tag{6.67}$$

the unique associated Lagrange multiplier is $\bar{\lambda} = 1/2$, and the SOSC (6.66) is satisfied as well.

For any $x^k \in \mathbf{R}^2$, let $(p^k, \lambda^{k+1}) \in \mathbf{R}^2 \times \mathbf{R}$ be a stationary point and Lagrange multiplier pair of the SQP subproblem, with the "ideal" choice of $H_k = \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}) = I$. Then from the first line in (6.41) we have that

$$1 + p_1^k + 2\lambda^{k+1}x_1^k = 0, \quad p_2^k + 2\lambda^{k+1}x_2^k = 0, \tag{6.68}$$

$$(x_1^k)^2 + (x_2^k)^2 - 1 + 2x_1^k p_1^k + 2x_2^k p_2^k = 0. \tag{6.69}$$

Observe that for any $x^k \neq 0$ relations (6.68), (6.69) define the unique pair (p^k, λ^{k+1}). Multiplying both sides of the first equality in (6.68) by p_1^k, the second equality by p_2^k, and summing up the results, we obtain that

$$p_1^k = \lambda^{k+1}((x_1^k)^2 + (x_2^k)^2 - 1) - (p_1^k)^2 - (p_2^k)^2,$$

where (6.69) was also used. Therefore,

$$f(x^k + p^k) - f(x^k) = p_1^k = \lambda^{k+1}h(x^k) - \|p^k\|^2.$$

By (6.69), we also have that

$$h(x^k + p^k) = (x_1^k + p_1^k)^2 + (x_2^k + p_2^k)^2 - 1 = \|p^k\|^2.$$

Thus, for any $c > 1$,

$$\varphi_c(x^k + p^k) - \varphi_c(x^k) = \lambda^{k+1}h(x^k) - c|h(x^k)| + (c-1)\|p^k\|^2 > 0,$$

if, for example, $h(x^k) = 0$ and x^k is not a stationary point of problem (6.33) (which implies $p^k \neq 0$; see the proof of Theorem 6.9). This means that $\alpha = 1$ does not satisfy the inequality (6.39), because of (6.42).

According to the above, at any point with the exhibited properties (and there are such points arbitrarily close to the solution), the unit stepsize would not be accepted by the linesearch descent test (6.39).

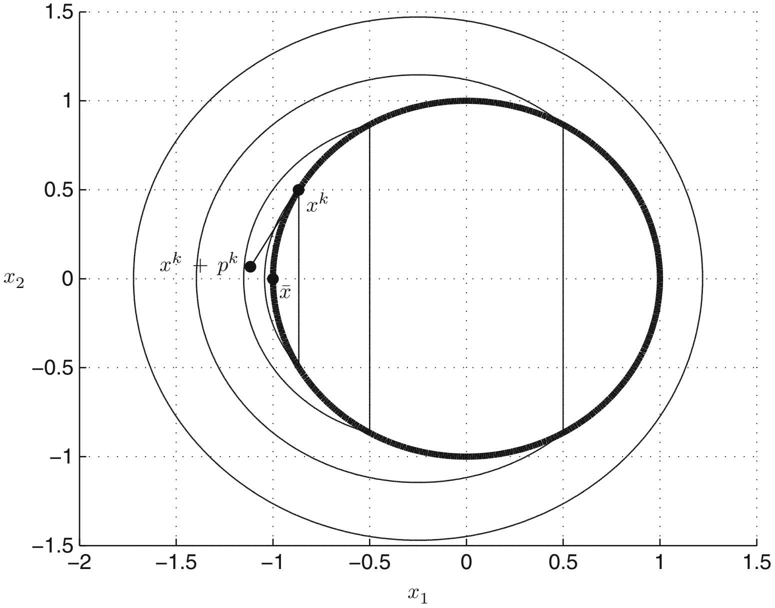

Fig. 6.1 Maratos effect in Example 6.11

One could note that in the previous example, the unit stepsize would be acceptable in linesearch if $c \in (1/2,\, 1)$. But it is not difficult to modify this example in such a way that the unit stepsize is rejected for any value of the penalty parameter.

Example 6.11. In Example 6.10, redefine the objective function as follows: $f(x) = x_1 + x_1^2 + x_2^2$. The point $\bar{x} = (-1,\, 0)$ remains the only global solution of problem (6.33), with the associated unique Lagrange multiplier $\bar{\lambda} = -1/2$. As before, $\bar{x}$ satisfies the LICQ, and the SOSC (6.66) also holds.

Let $x^k \in \mathbf{R}^2$ be any feasible point ($h(x^k) = 0$) which is not stationary for problem (6.33). Let $(p^k,\, \lambda^{k+1}) \in \mathbf{R}^2 \times \mathbf{R}$ be the solution of the Lagrange optimality system of subproblem (6.34) with $H_k = \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda}) = I$. We then have that

$$1 + 2x_1^k + p_1^k + 2\lambda^{k+1}x_1^k = 0, \quad 2x_2^k + p_2^k + 2\lambda^{k+1}x_2^k = 0, \tag{6.70}$$

$$x_1^k p_1^k + x_2^k p_2^k = 0. \tag{6.71}$$

Multiplying the first equality in (6.70) by p_1^k, the second by p_2^k, and summing the results, we obtain

$$p_1^k + (p_1^k)^2 + (p_2^k)^2 = 0,$$

where (6.71) was also used. Hence,

$$f(x^k + p^k) - f(x^k) = p_1^k + (x_1^k + p_1^k)^2 + (x_2^k + p_2^k)^2 - (x_1^k)^2 - (x_2^k)^2 = 0,$$

$$h(x^k + p^k) = (x_1^k + p_1^k)^2 + (x_2^k + p_2^k)^2 - 1 = \|p^k\|^2,$$

where (6.71) was taken into account. We now conclude that

$$\varphi_c(x^k + p^k) - \varphi_c(x^k) = c\|p^k\|^2 > 0 \quad \forall\, c > 0.$$

In Fig. 6.1, thick circle represents the feasible set, and thin lines are the level lines of φ_c for $c = 1$. On can see that the full SQP step from the point x^k increases the value of φ_c.

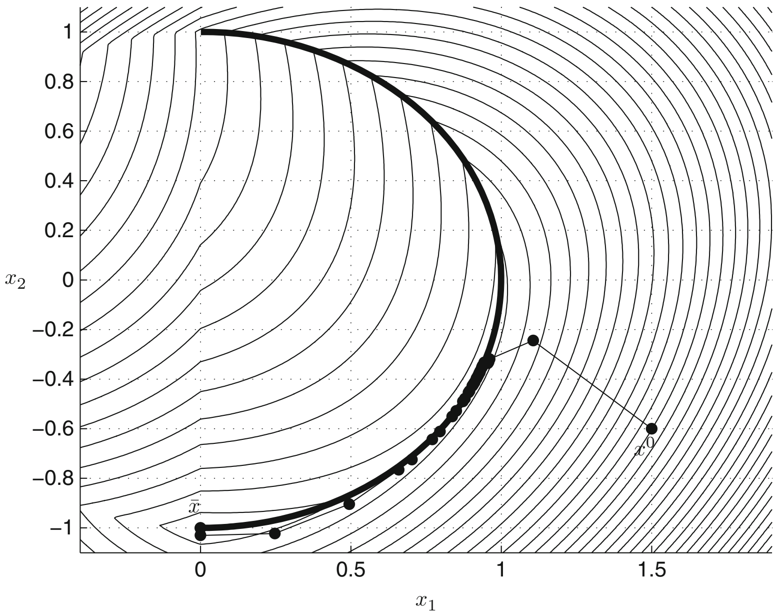

Fig. 6.2 Maratos effect in Example 6.12: primal iterative sequence

Somewhat informally, the Maratos effect can be explained as follows. If the point x^k is close to the feasible region, the step from x^k to $x^k + p^k$ may increase the penalty term ψ measuring constraints violation by a value of the same order as the decrease of the objective function f by the same step. Consequently, for c large enough, the aggregate value of the merit function φ_c can increase, in which case the unit stepsize would be rejected by the linesearch rule.

In practice, the Maratos effect usually shows up as follows: it does not actually destroy the asymptotic superlinear convergence altogether, but it significantly delays the point at which it is observed. The method takes a long sequence of short steps along the boundary of the feasible set, and the eventual superlinear rate does not compensate, for practical purposes, this slowing down at previous iterations. This behavior is demonstrated by the next example.

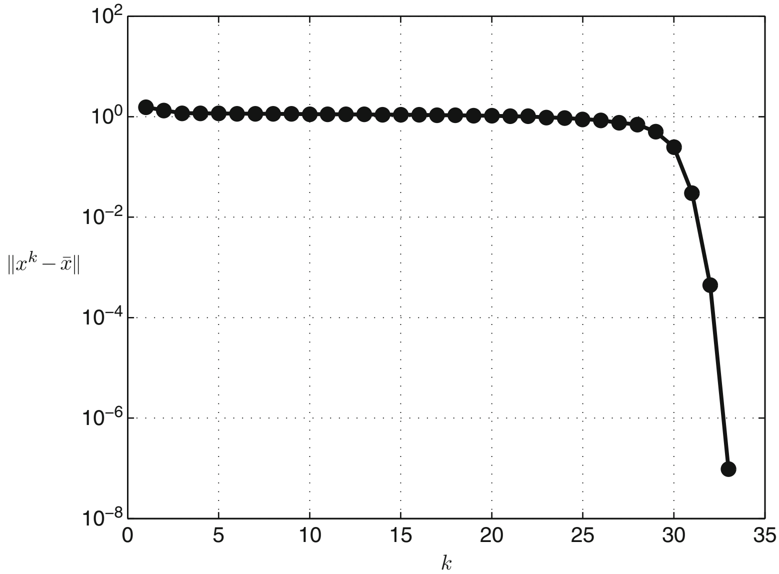

Fig. 6.3 Maratos effect in Example 6.12: distance to solution

Example 6.12. Let $n = 2$, $l = m = 1$, $f(x) = 0.5(x_1 - x_2 - 1)^2 + 0.15(x_1 + x_2)$, $h(x) = x_1^2 + x_2^2 - 1$, $g(x) = -x_1$. The global solution of problem (6.33) with this data is $\bar{x} = (0, -1)$, with the associated Lagrange multiplier being $(\bar{\lambda}, \bar{\mu}) = (0.075, 0.15)$.

Figure 6.2 shows the feasible set (thick line), the sequence generated by Algorithm 6.7 with H_k chosen according to (6.64), and with $\bar{c} = 1$, $\sigma = 0.1$, $\theta = 0.5$, starting from $x^0 = (1.5, -0.6)$ and $(\lambda^0, \mu^0) = (0, 0)$, and the level lines of φ_c (thin lines) for $c = \|(\bar{\lambda}, \bar{\mu})\|_\infty + 1 = 1.15$. Figure 6.3 shows the decrease of the distance to the solution.

To avoid the Maratos effect, Algorithm 6.7 needs to be modified. This can be done in a number of ways. Below, we shall consider two: adding the second-order correction step (recall Sect. 4.3.6), and using the nonsmooth augmented Lagrangian (introduced in Sect. 6.1) as a merit function for linesearch.

Both of those approaches are usually presented for the equality-constrained problem (6.65). The SQP subproblem for (6.65) is given by

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, p\rangle + \dfrac{1}{2}\langle H_k p,\, p\rangle \\ \text{subject to} & h(x^k) + h'(x^k)p = 0, \end{array}$$

with some symmetric matrix $H_k \in \mathbf{R}^{n\times n}$. For an iterate $x^k \in \mathbf{R}^n$, let $p^k \in \mathbf{R}^n$ be a stationary point and $\lambda^{k+1} \in \mathbf{R}^l$ an associated Lagrange multiplier of this subproblem. In other words, $(p^k,\, \lambda^{k+1})$ satisfies the Newton–Lagrange iteration system for problem (6.65):

$$f'(x^k) + H_k p^k + (h'(x^k))^{\mathrm{T}}\lambda^{k+1} = 0, \quad h(x^k) + h'(x^k)p^k = 0. \tag{6.72}$$

As discussed in Sect. 4.3.6, the idea of second-order correction is to bring the usual SQP step closer to the feasible region, so that to decrease the penalty term ψ. If the Euclidean norm is used, the direction of correction $\bar{p}^k \in \mathbf{R}^n$ is computed as the solution of the quadratic programming problem

$$\begin{array}{ll} \text{minimize} & \|p\|^2 \\ \text{subject to} & h(x^k + p^k) + h'(x^k)p = 0. \end{array} \tag{6.73}$$

If the condition

$$\operatorname{rank} h'(x^k) = l \tag{6.74}$$

holds, applying the Lagrange principle (Theorem 1.11) to problem (6.73), one can easily obtain that the unique solution of this problem is given by

$$\bar{p}^k = -(h'(x^k))^{\mathrm{T}}(h'(x^k)(h'(x^k))^{\mathrm{T}})^{-1}h(x^k + p^k). \tag{6.75}$$

According to Lemma 4.38, if h is differentiable in a neighborhood of a feasible point $\bar{x} \in \mathbf{R}^n$ of problem (6.65), with its derivative locally Lipschitz-continuous at $\bar{x}$, and if the regularity condition (6.67) holds, then

$$h(x^k + p^k) = O(\|p^k\|^2), \tag{6.76}$$

$$\bar{p}^k = O(\|p^k\|^2) \tag{6.77}$$

as $x^k \to \bar{x}$ and $p^k \to 0$. As discussed in Sect. 4.3.6, this property combined with the tools developed in Sects. 2.1.1 and 4.1.1 allows to establish local convergence and superlinear rate of convergence of the method producing iterates of the form $x^{k+1} = x^k + p^k + \bar{p}^k$.

On the other hand, using Lemma A.11, from the constraints in (6.73) and from (6.77), we have that

$$\begin{aligned} h(x^k + p^k + \bar{p}^k) &= h(x^k + p^k) + h'(x^k + p^k)\bar{p}^k + O(\|\bar{p}^k\|^2) \\ &= h(x^k + p^k) + h'(x^k)\bar{p}^k + O(\|p^k\|\|\bar{p}^k\|) + O(\|\bar{p}^k\|^2) \\ &= O(\|p^k\|^3) \end{aligned} \tag{6.78}$$

as $x^k \to \bar{x}$ and $p^k \to 0$. Comparing this with (6.76) we see that the step from x^k to $x^k + p^k + \bar{p}^k$ yields better decrease of the constraints infeasibility than the step to $x^k + p^k$. We can thus expect that the step to $x^k + p^k + \bar{p}^k$ would provide the needed decrease of the l_1-penalty function $\varphi_c : \mathbf{R}^n \to \mathbf{R}$ given by

$$\varphi_c(x) = f(x) + c\|h(x)\|_1, \tag{6.79}$$

for the current value $c = c_k \geq 0$ of the penalty parameter.

To embed those considerations into the global scheme, the next iterate x^{k+1} is computed via curvilinear search, i.e., it has the form $x^k + \alpha p^k + \alpha^2 \bar{p}^k$, where $\alpha > 0$ is given by a backtracking procedure starting with the first trial value $\alpha = 1$, until the following descent test is satisfied:

$$\varphi_{c_k}(x^k + \alpha p^k + \alpha^2 \bar{p}^k) \leq \varphi_{c_k}(x^k) + \sigma\alpha\Delta_k, \tag{6.80}$$

where

$$\Delta_k = \varphi'_{c_k}(x^k;\, p^k) = \langle f'(x^k),\, p^k\rangle - c_k\|h(x^k)\|_1 \tag{6.81}$$

(recall (6.38), and Lemma 6.8 and the discussion following its statement).

Another possibility is to check (6.80) for $\alpha = 1$ only: if the corresponding step to the point $x^k + p^k + \bar{p}^k$ is not accepted, one might resort to the usual linesearch in the direction p^k.

As will be shown next, under natural assumptions $\alpha_k = 1$ is locally accepted by (6.80), which results in the superlinear convergence of the corresponding global algorithm. On the other hand, (6.40) implies that under the assumptions of Lemma 6.8, if H_k is positive definite and $p^k \neq 0$, then the inequality (6.80) holds for all $\alpha > 0$ small enough, so that the method is well defined globally and provides a decrease of the merit function from one iterate to the next, just as Algorithm 6.7.

Theorem 6.13. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice differentiable in some neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem* (6.65) *satisfying the regularity condition* (6.67)*. Let $\bar{\lambda} \in \mathbf{R}^l$ be the (unique) Lagrange multiplier associated with $\bar{x}$, and assume that the SOSC holds:*

$$\left\langle \frac{\partial^2 L}{\partial x}(\bar{x},\, \bar{\lambda})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in \ker h'(\bar{x}) \setminus \{0\}. \tag{6.82}$$

Let $c \geq 0$ be such that the matrix $\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}) + c(h'(\bar{x}))^{\mathrm{T}} h'(\bar{x})$ is positive definite. Assume, finally, that the sequence $\{x^k\} \subset \mathbf{R}^n$ converges to $\bar{x}$, the sequence $\{p^k\} \subset \mathbf{R}^n$ converges to 0, the sequence $\{c_k\} \subset \mathbf{R}_+$ is bounded, and for each k a symmetric matrix $H_k \in \mathbf{R}^{n \times n}$ and $\bar{p}^k \in \mathbf{R}^n$ are such that (p^k, λ^{k+1}) satisfies (6.72) with some $\lambda^{k+1} \in \mathbf{R}^l$, $\bar{p}^k \in \mathbf{R}^n$ is the solution of problem (6.73), and it holds that

$$\langle H_k p^k, p^k \rangle \geq \left\langle \left(\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}) + c(h'(\bar{x}))^{\mathrm{T}} h'(\bar{x}) \right) p^k, p^k \right\rangle + o(\|p^k\|^2) \quad (6.83)$$

as $k \to \infty$, and

$$c_k \geq \|\lambda^{k+1}\|_\infty. \quad (6.84)$$

Then for any $\sigma \in (0, 1/2)$ and all k large enough, the inequality (6.80) holds for $\alpha = 1$, where $\varphi_{c_k} : \mathbf{R}^n \to \mathbf{R}$ is given by (6.79).

Concerning the relation (6.83) on the choice of the matrices H_k, recall again that according to Proposition 4.8, under the SOSC (6.82) the matrix $\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}) + c(h'(\bar{x}))^{\mathrm{T}} h'(\bar{x})$ is positive definite for all $c > 0$ large enough. Condition (6.83) means that the quadratic form associated with H_k must estimate from above the quadratic form associated with the latter matrix.

Proof. From the relation (6.83) and the choice of c it follows that there exists $\gamma > 0$ such that

$$\langle H_k p^k, p^k \rangle \geq \gamma \|p^k\|^2 \quad (6.85)$$

for all k large enough. Also, from (6.83) and the fact that $(h'(\bar{x}))^{\mathrm{T}} h'(\bar{x})$ is positive semidefinite, we have that

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}) p^k, p^k \right\rangle \leq \langle H_k p^k, p^k \rangle + o(\|p^k\|^2) \quad (6.86)$$

as $k \to \infty$. Finally, the regularity condition (6.67) implies that (6.74), and hence, (6.75) hold for all k large enough, with the matrices $(h'(x^k)(h'(x^k))^{\mathrm{T}})^{-1}$ uniformly bounded.

Using (6.72), (6.75), (6.76), (6.78), (6.79), (6.81), (6.84)–(6.86), the mean-value theorem (Theorem A.10, (a), applied twice), boundedness of the sequences $\{(h'(x^k)(h'(x^k))^{\mathrm{T}})^{-1}\}$ and $\{c_k\}$, and the convergence of $\{x^k\}$ to $\bar{x}$, we derive the following chain of relations for all k large enough:

$$
\begin{aligned}
&\varphi_{c_k}(x^k + p^k + \bar{p}^k) - \varphi_{c_k}(x^k) - \sigma\Delta_k \\
&= f(x^k + p^k + \bar{p}^k) - f(x^k) + c_k\|h(x^k + p^k + \bar{p}^k)\|_1 - c_k\|h(x^k)\|_1 - \sigma\Delta_k \\
&= \langle f'(x^k),\, p^k + \bar{p}^k\rangle + \frac{1}{2}\langle f''(x^k)p^k,\, p^k\rangle - c_k\|h(x^k)\|_1 - \sigma\Delta_k + o(\|p^k\|^2) \\
&= \langle f'(x^k),\, \bar{p}^k\rangle + \frac{1}{2}\langle f''(x^k)p^k,\, p^k\rangle + (1-\sigma)\Delta_k + o(\|p^k\|^2) \\
&= -\langle f'(x^k),\, (h'(x^k))^{\mathrm{T}}(h'(x^k)(h'(x^k))^{\mathrm{T}})^{-1}h(x^k + p^k)\rangle \\
&\qquad + \frac{1}{2}\langle f''(x^k)p^k,\, p^k\rangle + (1-\sigma)\Delta_k + o(\|p^k\|^2) \\
&= -\langle (h'(x^k)(h'(x^k))^{\mathrm{T}})^{-1}h'(x^k)f'(x^k),\, h(x^k + p^k)\rangle \\
&\qquad + \frac{1}{2}\langle f''(x^k)p^k,\, p^k\rangle + (1-\sigma)\Delta_k + o(\|p^k\|^2) \\
&= \langle (h'(x^k)(h'(x^k))^{\mathrm{T}})^{-1}h'(x^k)(h'(x^k))^{\mathrm{T}}\bar{\lambda},\, h(x^k + p^k)\rangle \\
&\qquad - \left\langle (h'(x^k)(h'(x^k))^{\mathrm{T}})^{-1}h'(x^k)\frac{\partial L}{\partial x}(x^k,\, \bar{\lambda}),\, h(x^k + p^k)\right\rangle \\
&\qquad + \frac{1}{2}\langle f''(x^k)p^k,\, p^k\rangle + (1-\sigma)\Delta_k + o(\|p^k\|^2) \\
&= \langle \bar{\lambda},\, h(x^k + p^k)\rangle + \frac{1}{2}\langle f''(x^k)p^k,\, p^k\rangle + (1-\sigma)\Delta_k + o(\|p^k\|^2) \\
&= \frac{1}{2}\langle f''(x^k)p^k,\, p^k\rangle + \left\langle \bar{\lambda},\, h(x^k) + h'(x^k)p^k + \frac{1}{2}h''(x^k)[p^k,\, p^k]\right\rangle \\
&\qquad + (1-\sigma)\Delta_k + o(\|p^k\|^2) \\
&= \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(x^k,\, \bar{\lambda})p^k,\, p^k\right\rangle + (1-\sigma)\Delta_k + o(\|p^k\|^2) \\
&= (1-\sigma)(\langle f'(x^k),\, p^k\rangle - c_k\|h(x^k)\|_1) + \frac{1}{2}\left\langle \frac{\partial L}{\partial x}(\bar{x},\, \bar{\lambda})p^k,\, p^k\right\rangle + o(\|p^k\|^2) \\
&= (1-\sigma)(-\langle H_k p^k,\, p^k\rangle - \langle \lambda^{k+1},\, h'(x^k)p^k\rangle - c_k\|h(x^k)\|_1) \\
&\qquad + \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})p^k,\, p^k\right\rangle + o(\|p^k\|^2) \\
&= -\left(\frac{1}{2} - \sigma\right)\langle H_k p^k,\, p^k\rangle + (1-\sigma)(\langle \lambda^{k+1},\, h(x^k)\rangle - c_k\|h(x^k)\|_1) \\
&\qquad - \frac{1}{2}\left(\langle H_k p^k,\, p^k\rangle - \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})p^k,\, p^k\right\rangle\right) + o(\|p^k\|^2) \\
&\leq -\left(\frac{1}{2} - \sigma\right)\langle H_k p^k,\, p^k\rangle + (1-\sigma)(\|\lambda^{k+1}\|_\infty - c_k)\|h(x^k)\|_1 + o(\|p^k\|^2) \\
&\leq -\left(\frac{1}{2} - \sigma\right)\langle H_k p^k,\, p^k\rangle + o(\|p^k\|^2) \leq -\left(\frac{1}{2} - \sigma\right)\gamma\|p^k\|^2 + o(\|p^k\|^2) \\
&< 0,
\end{aligned}
$$

which establishes the claim. □

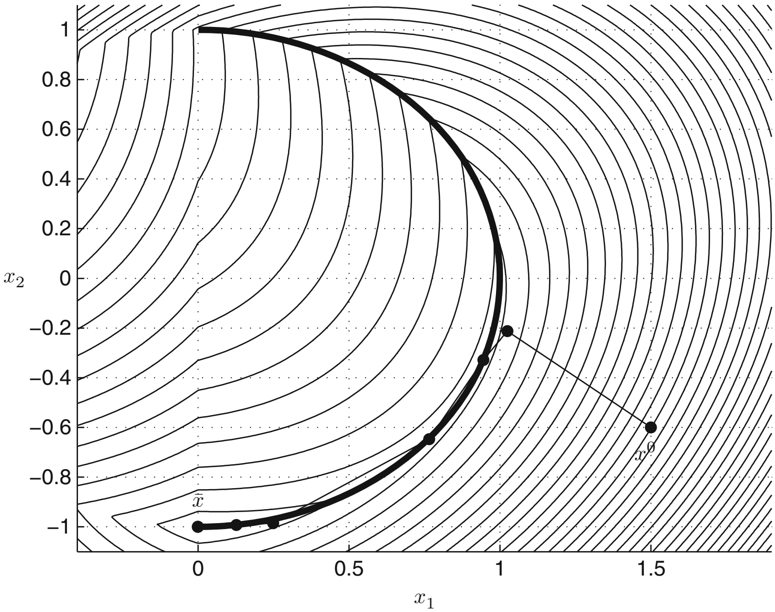

Fig. 6.4 Example 6.12: primal iterative sequence using second-order corrections

Figures 6.4 and 6.5 demonstrate the use of second-order corrections in Example 6.12 (cf. Figs. 6.2 and 6.3). The second-order corrections are computed ignoring the inequality constraint of the problem in question.

As another tool for avoiding the Maratos effect, we next consider the approach proposed in [25]. It consists of performing linesearch in the SQP direction for the nonsmooth augmented Lagrangian function $\varphi_{c,\,\eta} : \mathbf{R}^n \to \mathbf{R}$ defined in Sect. 6.1 as follows:

$$\varphi_{c,\,\eta}(x) = L(x,\,\eta) + c\|h(x)\|_1 = f(x) + \langle \eta,\, h(x)\rangle + c\|h(x)\|_1, \tag{6.87}$$

where $c \geq 0$ and $\eta \in \mathbf{R}^l$ are parameters. Recall that under the SOSC, with the proper choice of parameters this is an exact penalty function for problem (6.65); see Proposition 6.6.

As can be verified using (6.72) and adapting the analysis in Proposition 6.1 and Lemma 6.8, the directional derivative of the nonsmooth augmented Lagrangian in the SQP direction p^k is given by

$$\begin{aligned} \varphi'_{c,\,\eta}(x^k;\,p^k) &= -\langle H_k p^k,\, p^k\rangle + \langle \lambda^{k+1} - \eta,\, h(x^k)\rangle - c\|h(x^k)\|_1 \\ &\leq -\langle H_k p^k,\, p^k\rangle + (\|\lambda^{k+1} - \eta\|_\infty - c)\|h(x^k)\|_1. \end{aligned} \tag{6.88}$$

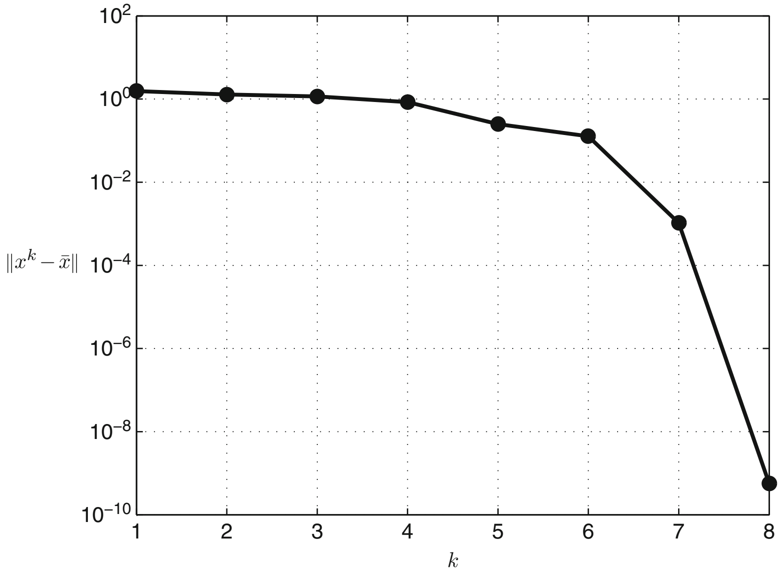

Fig. 6.5 Example 6.12: distance to solution using second-order corrections

Therefore, if $c \geq \|\lambda^{k+1} - \eta\|_\infty$, we have that

$$\varphi'_{c,\,\eta}(x^k;\, p^k) \leq -\langle H_k p^k,\, p^k\rangle, \tag{6.89}$$

and if the matrix H_k is positive definite and $p^k \neq 0$, then p^k is a descent direction for $\varphi_{c,\eta}$ at the point x^k. This allows to construct a method along the lines of Algorithm 6.7, using this modified merit function. The rule for choosing the penalty parameter (6.37) should now be substituted by

$$c_k \geq \|\lambda^{k+1} - \eta^k\|_\infty + \bar{c},$$

and the Armijo inequality for linesearch (6.39) by

$$\varphi_{c_k,\,\eta_k}(x^k + \alpha p^k) \leq \varphi_{c_k,\,\eta_k}(x^k) + \sigma\alpha\varphi'_{c_k,\,\eta_k}(x^k;\, p^k). \tag{6.90}$$

The following result shows that with a proper choice of the matrices H_k (incidentally, the same as in Theorem 6.13 dealing with second-order corrections), the modified global method preserves the superlinear convergence rate of local SQP.

Theorem 6.14. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice differentiable in some neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (6.65), let $\bar{\lambda} \in \mathbf{R}^l$ be a Lagrange multiplier associated with $\bar{x}$, and assume that the SOSC (6.82) holds. Let $c \geq 0$ be such that the matrix $\frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda}) + c(h'(\bar{x}))^{\mathrm{T}} h'(\bar{x})$ is positive definite.*

Assume, finally, that the sequence $\{x^k\} \subset \mathbf{R}^n$ *converges to* $\bar{x}$*, the sequence* $\{p^k\} \subset \mathbf{R}^n$ *converges to* 0*, and for each* k *the symmetric matrix* $H_k \in \mathbf{R}^{n\times n}$ *is such that* (p^k, λ^{k+1}) *satisfies* (6.72) *with some* $\lambda^{k+1} \in \mathbf{R}^l$*,* (6.83) *holds, and*

$$c_k \geq \|\lambda^{k+1} - \eta^k\|_\infty. \tag{6.91}$$

Then for any $\sigma \in (0, 1/2)$ *there exists* $\delta > 0$ *such that for all* k *large enough, if the conditions*

$$c_k \leq \delta, \quad \|\eta^k - \bar{\lambda}\| \leq \delta \tag{6.92}$$

hold, then (6.90) *holds for* $\alpha = 1$*, where* $\varphi_{c_k,\,\eta^k} : \mathbf{R}^n \to \mathbf{R}$ *is given by* (6.87)*.*

Proof. Using (6.72), we have that for each k it holds that

$$\begin{aligned}\langle f'(x^k),\, p^k\rangle &= -\langle H_k p^k,\, p^k\rangle - \langle \lambda^{k+1},\, h'(x^k)p^k\rangle\\ &= -\langle H_k p^k,\, p^k\rangle + \langle \lambda^{k+1},\, h(x^k)\rangle.\end{aligned}$$

By the mean-value theorem (Theorem A.10, (a), applied twice) and the last relation, we obtain that

$$\begin{aligned}f(x^k + p^k) &= f(x^k) + \langle f'(x^k),\, p^k\rangle + \frac{1}{2}\langle f''(\bar{x})p^k,\, p^k\rangle + o(\|p^k\|^2)\\ &= f(x^k) - \langle H_k p^k,\, p^k\rangle + \langle \lambda^{k+1},\, h(x^k)\rangle\\ &\quad + \frac{1}{2}\langle f''(\bar{x})p^k,\, p^k\rangle + o(\|p^k\|^2)\end{aligned} \tag{6.93}$$

as $k \to \infty$, where we also used the convergence of $\{x^k\}$ to $\bar{x}$ and of $\{p^k\}$ to zero. Similarly, using the second equality in (6.72), it holds that

$$\begin{aligned}h(x^k + p^k) &= h(x^k) + h'(x^k)p^k + \frac{1}{2}h''(\bar{x})[p^k,\, p^k] + o(\|p^k\|^2)\\ &= \frac{1}{2}h''(\bar{x})[p^k,\, p^k] + o(\|p^k\|^2).\end{aligned} \tag{6.94}$$

Now, combining (6.93) and (6.94), we obtain that

$$\begin{aligned}
\varphi_{c_k,\,\eta^k}(x^k+p^k) &= f(x^k+p^k)+\langle \eta^k,\, h(x^k+p^k)\rangle + c_k\|h(x^k+p^k)\|_1\\
&= f(x^k)-\langle H_k p^k,\, p^k\rangle + \langle \lambda^{k+1},\, h(x^k)\rangle\\
&\quad +\frac{1}{2}\langle f''(\bar{x})p^k,\, p^k\rangle + \frac{1}{2}\langle \eta^k,\, h''(\bar{x})[p^k,\, p^k]\rangle\\
&\quad +\frac{c_k}{2}\|h''(\bar{x})[p^k,\, p^k]\|_1 + o(\|p^k\|^2)\\
&= f(x^k)-\langle H_k p^k,\, p^k\rangle + \langle \lambda^{k+1},\, h(x^k)\rangle\\
&\quad +\frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \eta^k)p^k,\, p^k\right\rangle + O(c_k\|p^k\|^2)+o(\|p^k\|^2)\\
&= f(x^k)-\langle H_k p^k,\, p^k\rangle + \langle \lambda^{k+1},\, h(x^k)\rangle\\
&\quad +\frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})p^k,\, p^k\right\rangle\\
&\quad +O(c_k\|p^k\|^2)+O(\|\eta^k-\bar{\lambda}\|\|p^k\|^2)+o(\|p^k\|^2)
\end{aligned} \tag{6.95}$$

as $k\to\infty$.

Note that as in the proof of Theorem 6.13, from (6.83) and the choice of c it follows that there exists $\gamma>0$ such that (6.85) holds for all k large enough, and from (6.83) and the fact that $(h'(\bar{x}))^{\mathrm{T}}h'(\bar{x})$ is positive semidefinite it follows that (6.86) holds.

Using the relations (6.86)–(6.88) and (6.95), we conclude that

$$\begin{aligned}
\varphi_{c_k,\,\eta^k}(x^k+p^k)-\varphi_{c_k,\,\eta^k}(x^k) &= -\langle H_k p^k,\, p^k\rangle + \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\bar{\lambda})p^k,\, p^k\right\rangle\\
&\quad +\langle \lambda^{k+1}-\eta^k,\, h(x^k)\rangle - c_k\|h(x^k)\|_1\\
&\quad +O((c_k+\|\eta^k-\bar{\lambda}\|)\|p^k\|^2)+o(\|p^k\|^2)\\
&= \varphi'_{c_k,\,\eta^k}(x^k;\, p^k)+\frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})p^k,\, p^k\right\rangle\\
&\quad +O((c_k+\|\eta^k-\bar{\lambda}\|)\|p^k\|^2)+o(\|p^k\|^2)\\
&\le \varphi'_{c_k,\,\eta^k}(x^k;\, p^k)+\frac{1}{2}\langle H_k p^k,\, p^k\rangle\\
&\quad +O((c_k+\|\eta^k-\bar{\lambda}\|)\|p^k\|^2)+o(\|p^k\|^2)
\end{aligned}$$

as $k\to\infty$. Now, using (6.92), we can write

$$\begin{aligned}
\varphi_{c_k,\,\eta^k}(x^k+p^k)-\varphi_{c_k,\,\eta^k}(x^k) &\le \varphi'_{c_k,\,\eta^k}(x^k;\, p^k)+\frac{1}{2}\langle H_k p^k,\, p^k\rangle + O(\delta\|p^k\|^2)\\
&\le \sigma\varphi'_{c_k,\eta^k}(x^k;p^k)-\frac{1-2\sigma}{2}\langle H_k p^k,\, p^k\rangle\\
&\quad +O(\delta\|p^k\|^2)\\
&\le \sigma\varphi'_{c_k,\,\eta^k}(x^k;\, p^k)-\frac{\gamma(1-2\sigma)}{2}\|p^k\|^2\\
&\quad +O(\delta\|p^k\|^2)\\
&\le \sigma\varphi'_{c_k,\,\eta^k}(x^k;\, p^k),
\end{aligned}$$

where the second inequality follows from (6.89) (which is valid for $c = c_k$ and $\eta = \eta^k$ due to (6.91)), the third from (6.85) and the assumption that $\sigma \in (0, 1/2)$, and the last holds for all k large enough if $\delta > 0$ is sufficiently small. □

Specific rules for choosing the parameters c_k and η^k, that guarantee both global and local superlinear convergence of this modification of SQP, are given in [25].

6.2.3 Globalization of Sequential Quadratically Constrained Quadratic Programming

In this section we consider globalization of the SQCQP method introduced in Sect. 4.3.5, for the problem with inequality constraints

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \leq 0, \end{array} \tag{6.96}$$

where $f : \mathbf{R}^n \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}^m$. Our approach mostly follows [248].

It should be first noted that, unlike in the case of the SQP methods, SQCQP subproblems may be infeasible even if the components of g are convex (which will be the setting in this section). Consider, e.g., $n = m = 1$, $g(x) = x^4 - 1/4$, and the iterate $x^k = 1$. Then the quadratic approximation of the constraint around x^k is given by the inequality $3/4 + 4(x-1) + 6(x-1)^2 \leq 0$, which does not have any solutions.

This observation reveals that any globalization of SQCQP must deal with infeasibility of subproblems, even in the convex case. It is natural to adopt the "elastic mode" idea, already mentioned in Sect. 6.2.1 in connection with SQP (see (6.62)). This leads to the subproblem

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, p\rangle + c_k t + \dfrac{1}{2}\langle H_k p,\, p\rangle \\ \text{subject to} & g_i(x^k) + \langle g_i'(x^k),\, p\rangle + \dfrac{1}{2}\langle G_i^k p,\, p\rangle \leq t,\ i \in I_k,\ t \geq 0, \end{array} \tag{6.97}$$

where $c_k > 0$ is a penalty parameter, $H_k \in \mathbf{R}^{n\times n}$ is a symmetric positive definite matrix, $G_i^k \in \mathbf{R}^{n\times n}$ are symmetric positive semidefinite matrices, $i \in I_k$, and the variables are $(p, t) \in \mathbf{R}^n \times \mathbf{R}$. The index set $I_k \subset \{1, \ldots, m\}$ is defined as follows. Let $\psi : \mathbf{R}^n \to \mathbf{R}_+$ be the l_∞-penalty for the constraints of problem (6.96), given by

$$\psi(x^k) = \max\{0,\, g_1(x^k),\, \ldots,\, g_m(x^k)\}, \tag{6.98}$$

and set

$$I(x^k) = \{i = 1,\, \ldots,\, m \mid g_i(x^k) = \psi(x^k)\}. \tag{6.99}$$

Then I_k can be any index set which satisfies

$$I(x^k) \subset I_k. \tag{6.100}$$

In particular, I_k must contain the constraints most violated at x^k, but need not include all the constraints. The possibility to choose the index set I_k smaller than $\{1, \ldots, m\}$ can be useful (especially at the early stages of the algorithm), because it reduces the number of constraints thus leading to simpler subproblems.

Let $(p^k, t_k) \in \mathbf{R}^n \times \mathbf{R}$ be a solution of (6.97). Then the next iterate would be given by $x^{k+1} = x^k + \alpha_k p^k$, where the stepsize $\alpha_k > 0$ is computed using an Armijo-type linesearch procedure for the l_∞-penalty function $\varphi_{c_k} : \mathbf{R}^n \to \mathbf{R}$,

$$\varphi_{c_k}(x) = f(x) + c_k \psi(x). \tag{6.101}$$

First, note that (6.97) is always feasible and, moreover, it always has the unique solution. Indeed, as is easy to see, for each $p \in \mathbf{R}^n$ fixed, the minimum with respect to t in (6.97) is attained at

$$t_k(p) = \max \left\{ 0, \max_{i \in I_k} \left\{ g_i(x^k) + \langle g_i'(x^k), p \rangle + \frac{1}{2} \langle G_i^k p, p \rangle \right\} \right\}. \tag{6.102}$$

Hence, (6.97) is equivalent to the unconstrained problem

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k), p \rangle + c_k t_k(p) + \frac{1}{2} \langle H_k p, p \rangle \\ \text{subject to} & p \in \mathbf{R}^n. \end{array}$$

Since H_k is positive definite, by the standard tools of convex analysis (see Sects. 1.2.1 and A.3) it can be seen that the objective function of this problem is strongly convex, and hence, this problem has the unique minimizer p^k. It follows that (p^k, t_k) with $t_k = t_k(p^k)$ is the unique solution of (6.97).

Furthermore, since the constraints in (6.97) obviously satisfy the Slater constraint qualification, by Theorems 1.14 and 1.16 we conclude that (p^k, t_k) is the unique stationary point of problem (6.97), and in particular, there exist some $\mu_i^k \in \mathbf{R}$, $i \in I_k$, and $\nu_k \in \mathbf{R}$ such that

$$\begin{gathered} f'(x^k) + H_k p^k + \sum_{i \in I_k} \mu_i^k (g_i'(x^k) + G_i^k p^k) = 0, \quad c_k - \sum_{i \in I_k} \mu_i^k - \nu_k = 0, \\ \mu_i^k \geq 0, \quad g_i(x^k) + \langle g_i'(x^k), p^k \rangle + \frac{1}{2} \langle G_i^k p^k, p^k \rangle \leq t_k, \\ \mu_i^k \left(g_i(x^k) + \langle g_i'(x^k), p^k \rangle + \frac{1}{2} \langle G_i^k p^k, p^k \rangle - t_k \right) = 0, \quad i \in I_k, \\ t_k \geq 0, \quad \nu_k \geq 0, \quad t_k \nu_k = 0. \end{gathered} \tag{6.103}$$

We proceed to state the algorithm (where we formally consider that the result of dividing by zero is $+\infty$).

Algorithm 6.15 Choose the parameters $c_0 > 0$, $\delta_1 > 0$, $\delta_2 > 0$, $\sigma \in (0, 1)$ and $\theta \in (0, 1)$. Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. Choose an index set $I_k \subset \{1, \ldots, m\}$ satisfying (6.100) with $I(x^k)$ defined according to (6.98), (6.99). Choose a symmetric positive definite matrix $H_k \in \mathbf{R}^{n\times n}$ and symmetric positive semidefinite matrices $G_i^k \in \mathbf{R}^{n\times n}$, $i \in I_k$, and compute $(p^k, t_k) \in \mathbf{R}^n \times \mathbf{R}$ as the solution of (6.97), and an associated Lagrange multiplier $((\mu_i^k, i \in I_k), \nu_k) \in \mathbf{R}^{|I_k|} \times \mathbf{R}$. Define $\mu^k \in \mathbf{R}^m$ with components μ_i^k, $i \in I_k$, and $\mu_i^k = 0$, $i \in \{1, \ldots, m\} \setminus I_k$.
2. If $p^k = 0$ and $t_k = 0$, stop.
3. If $p^k = 0$ but $t_k > 0$, set $\alpha_k = 1$ and go to step 4. Otherwise, set $\alpha = 1$. If the inequality (6.39) with $\varphi_{c_k} : \mathbf{R}^n \to \mathbf{R}$ defined by (6.101), and with
$$\Delta_k = \langle f'(x^k), p^k\rangle + \frac{1}{2}\langle H_k p^k, p^k\rangle + c_k(t_k - \psi(x^k)), \tag{6.104}$$
is satisfied, set $\alpha_k = \alpha$. Otherwise, replace α by $\theta\alpha$, check the inequality (6.39) again, etc., until (6.39) becomes valid.
4. Set $x^{k+1} = x^k + \alpha_k p^k$.
5. Compute $r_k = \min\{1/\|p^k\|, \|\mu^k\|_1 + \delta_1\}$. Set
$$c_{k+1} = \begin{cases} c_k & \text{if } c_k \geq r_k, \\ c_k + \delta_2 & \text{if } c_k < r_k. \end{cases} \tag{6.105}$$
6. Increase k by 1 and go to step 1.

Note that the penalty parameter rule above is different from the usual strategies in SQP methods, where typically $r_k = \|\mu^k\|_1 + \delta_1$ is used, or some variation of it. As already mentioned above, when discussing the elastic mode for SQP methods, choosing the penalty parameter in the framework where feasibility is controlled using slacks is problematic. In the specific setting of this section, the modified rule of Algorithm 6.15 resolves the difficulty and guarantees convergence. The idea is to ensure that if $\{c_k\}$ were to be unbounded, then necessarily two things happen: $\|\mu^k\| \to +\infty$ *and* $\{p^k\} \to 0$ as $k \to \infty$. As it will be established below (see Proposition 6.21), this situation cannot occur if the Slater condition holds for the original problem (6.96), thus leading to boundedness of $\{c_k\}$. It is also interesting to note that, unlike in SQP methods, acceptance of the unit stepsize occurs in SQCQP methods naturally, without any special assumptions (see Proposition 6.23 below). In particular, there is no Maratos effect for this algorithm, and thus no special modifications are needed to avoid it.

The following simple lemma will be used several times in the sequel.

Lemma 6.16. *Let the set $D = \{x \in \mathbf{R}^n \mid g_i(x) \leq 0, i = 1, \ldots, m\}$ be nonempty, where $g : \mathbf{R}^n \to \mathbf{R}^m$ is differentiable at $x \in \mathbf{R}^n$, and its components are convex on $\mathbf{R}^n$.*

If the index set
$$J(x) = \{i = 1, \ldots, m \mid g_i(x) > 0\} \tag{6.106}$$

is nonempty, then the relations

$$\sum_{i \in J(x)} \zeta_i g_i'(x) = 0, \quad \zeta_i \geq 0, \; i \in J(x), \tag{6.107}$$

hold only with $\zeta_i = 0$, $i \in J(x)$.

Proof. By the Gordan Theorem of the Alternatives (see Lemma A.3), the assertion follows if we establish that

$$\exists\, \xi \in \mathbf{R}^n \;\text{ such that }\; \langle g_i'(x),\, \xi\rangle < 0 \quad \forall\, i \in J(x).$$

We proceed to exhibit this ξ. Take any $\tilde{x} \in D$, and set $\xi = \tilde{x} - x$. By the convexity of the components of g, employing Proposition A.15 we have that

$$\langle g_i'(x),\, \xi\rangle = \langle g_i'(x),\, \tilde{x} - x\rangle \leq g_i(\tilde{x}) - g_i(x) < 0 \quad \forall\, i \in J(x),$$

thus establishing the claim. □

We start with showing that the method is well defined. This is done in two steps: for the case where $p^k = 0$ but $t_k > 0$, and when $p^k \neq 0$.

Proposition 6.17. *Let the set* $D = \{x \in \mathbf{R}^n \mid g_i(x) \leq 0,\, i = 1,\, \ldots,\, m\}$ *be nonempty, where* $g : \mathbf{R}^n \to \mathbf{R}^m$ *is continuously differentiable on* $\mathbf{R}^n$*, and its components are convex on* $\mathbf{R}^n$*.*

If in Algorithm 6.15 for the current $x^k \in \mathbf{R}^n$ *and* $\mu^k \in \mathbf{R}^m$ *we have* $p^k = 0$ *and* $t_k = 0$*, then* x^k *is a stationary point of problem (6.96), and* μ^k *is an associated Lagrange multiplier.*

If $p^k = 0$ *and* $t_k > 0$*, then there exists a positive integer* j *such that Algorithm 6.15 generates* $p^{k+j} \neq 0$*.*

Proof. Suppose that $p^k = 0$ and $t_k = 0$. Then (6.100) and the second line in (6.103) imply that $g_i(x^k) \leq 0$ for $i \in I(x^k) \subset I_k$. Hence, by (6.98), (6.99), $\psi(x^k) = 0$. Using again (6.100), it is easy to see that $g_{\{1,\, \ldots,\, m\} \setminus I_k}(x^k) < 0$. With this observation, setting $p^k = 0$ and $t_k = 0$ in (6.103), we obtain that $(x^k,\, \mu^k)$ satisfies the KKT conditions for (6.96).

Suppose now that $p^k = 0$, $t_k > 0$, and $p^{k+j} = 0$ for all j. Obviously, we then have $x^{k+j} = x^k$ for all j, by step 4 of Algorithm 6.15. By (6.98) and (6.102), $t_{k+j} = t_{k+j}(0) = \psi(x^{k+j}) = \psi(x^k) = t_k > 0$. Then the last line in (6.103) implies that $\nu_{k+j} = 0$.

Using further the second equality in (6.103), we have that for all j

$$c_{k+j} = \sum_{i \in I_{k+j}} \mu_i^{k+j}. \tag{6.108}$$

In step 5 of Algorithm 6.15, since $1/\|p^{k+j}\| = +\infty$ (by convention), (6.108) implies that $c_{k+j} < r_{k+j}$. Hence, by (6.105), $c_{k+j+1} = c_{k+j} + \delta_2$ for all j. This shows that $c_{k+j} \to +\infty$ as $j \to \infty$, and by (6.108),

$$\|\mu^{k+j}\| \to +\infty \text{ as } j \to \infty.$$

Observe that since $t_{k+j} = \psi(x^{k+j}) > 0$, (6.98), (6.99) and the third line in (6.103) imply that $\mu_i^{k+j} = 0$ for $i \in I_{k+j} \setminus I(x^{k+j})$. Since it holds that $I(x^k) = I(x^{k+j}) \subset I_{k+j}$, the first equality in (6.103) reduces to

$$f'(x^k) + \sum_{i \in I(x^k)} \mu_i^{k+j} g_i'(x^k) = 0.$$

Dividing both sides of the above equality by $\|\mu^{k+j}\|$ and passing onto the limit as $j \to \infty$ along an appropriate subsequence, we obtain the existence of ζ_i, $i \in I(x^k)$, not all equal to zero and such that

$$\sum_{i \in I(x^k)} \zeta_i g_i'(x^k) = 0, \quad \zeta_i \geq 0 \quad \forall\, i \in I(x^k).$$

Since $\psi(x^k) > 0$ implies that $J(x^k) \supset I(x^k) \neq \emptyset$, the latter contradicts Lemma 6.16, completing the proof. $\square$

The following result shows that whenever $p^k \neq 0$, it is a descent direction for φ_{c_k} at x^k. This, in turn, implies that the linesearch step (step 3 of Algorithm 6.15) is well defined. Combining this fact with Proposition 6.17, it follows that the whole Algorithm 6.15 is well defined.

Lemma 6.18. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable at a point $x^k \in \mathbf{R}^n$. Let $\psi : \mathbf{R}^n \to \mathbf{R}_+$ and $\varphi_{c_k} : \mathbf{R}^n \to \mathbf{R}$ be defined by (6.98) and (6.101) for some $c_k > 0$, respectively, and let $I_k \subset \{1, \ldots, m\}$ be an index set satisfying (6.100) with $I(x^k)$ defined according to (6.99). Let $H_k \in \mathbf{R}^{n \times n}$ be a symmetric matrix, let $G_i^k \in \mathbf{R}^{n \times n}$, $i \in I_k$, be positive semidefinite symmetric matrices, let $(p^k, t_k) \in \mathbf{R}^n \times \mathbf{R}$ be a stationary point of (6.97), and let $((\mu_i^k,\ i \in I_k), \nu_k) \in \mathbf{R}^{|I_k|} \times \mathbf{R}$ be an associated Lagrange multiplier.*

Then it holds that

$$\begin{aligned} \varphi_{c_k}'(x^k; p^k) &\leq \Delta_k - \frac{1}{2}\langle H_k p^k, p^k\rangle \\ &\leq -\langle H_k p^k, p^k\rangle - \frac{1}{2}\sum_{i \in I_k} \mu_i^k \langle G_i^k p^k, p^k\rangle - \nu_k \psi(x^k), \end{aligned} \tag{6.109}$$

where Δ_k is defined in (6.104).

Proof. By assertion (b) of Proposition 6.1, it holds that

$$\varphi_{c_k}'(x^k; p^k) = \langle f'(x^k), p^k\rangle + c_k \begin{cases} \max\limits_{i \in I(x^k)} \langle g_i'(x^k), p^k\rangle & \text{if } \psi(x^k) > 0, \\ \max\left\{0, \max\limits_{i \in I(x^k)} \langle g_i'(x^k), p^k\rangle\right\} & \text{if } \psi(x^k) = 0. \end{cases} \tag{6.110}$$

If $I(x^k) = \emptyset$, then it must hold that $\psi(x^k) = 0$ and $g_i(x^k) < 0$ for all $i = 1, \ldots, m$, and in particular $0 \leq t_k = t_k - \psi(x^k)$.

If $I(x^k) \neq \emptyset$, by the second inequality in (6.103), for all $i \in I(x^k) \subset I_k$ (recall (6.100)), it holds that

$$\langle g_i'(x^k),\, p^k\rangle \leq t_k - g_i(x^k) - \frac{1}{2}\langle G_i^k p^k,\, p^k\rangle \leq t_k - \psi(x^k),$$

where the second inequality follows from $g_i(x^k) = \psi(x^k)$ for $i \in I(x^k)$, and from G_i^k being positive semidefinite. If further $\psi(x^k) = 0$, then it holds that $\langle g_i'(x^k),\, p^k\rangle \leq t_k$. Hence,

$$\max\{0,\, \langle g_i'(x^k),\, p^k\rangle\} \leq \max\{0,\, t_k\} = t_k - \psi(x^k).$$

Therefore, in any case (6.110) gives that

$$\varphi_{c_k}'(x^k;\, p^k) \leq \langle f'(x^k),\, p^k\rangle + c_k(t_k - \psi(x^k)), \tag{6.111}$$

which is the first inequality in (6.109) (recall (6.104)).

Multiplying both sides of the first equality in (6.103) by p^k, we further have that

$$\langle f'(x^k),\, p^k\rangle = -\langle H_k p^k,\, p^k\rangle - \sum_{i\in I_k} \mu_i^k(\langle g_i'(x^k),\, p^k\rangle + \langle G_i^k p^k,\, p^k\rangle). \tag{6.112}$$

Employing (6.103), we then obtain that

$$\begin{aligned} -\sum_{i\in I_k} \mu_i^k \langle g_i'(x^k),\, p^k\rangle &= \sum_{i\in I_k} \mu_i^k \left(g_i(x^k) + \frac{1}{2}\langle G_i^k p^k,\, p^k\rangle - t_k\right) \\ &\leq (\psi(x^k) - t_k) \sum_{i\in I_k} \mu_i^k + \frac{1}{2}\sum_{i\in I_k} \mu_i^k \langle G_i^k p^k,\, p^k\rangle \\ &= (\psi(x^k) - t_k)(c_k - \nu_k) + \frac{1}{2}\sum_{i\in I_k} \mu_i^k \langle G_i^k p^k,\, p^k\rangle \\ &= c_k(\psi(x^k) - t_k) - \nu_k\psi(x^k) + \frac{1}{2}\sum_{i\in I_k} \mu_i^k \langle G_i^k p^k,\, p^k\rangle, \end{aligned}$$

where the inequality is by (6.98). Combining the latter relation with (6.104) and (6.112) gives the second inequality in (6.109). □

As a consequence, step 3 of Algorithm 6.15 is well defined and terminates with some $\alpha_k > 0$ after a finite number of backtrackings. Indeed, if $p^k = 0$, this step evidently returns $\alpha_k = 1$. If $p^k \neq 0$ and H_k is positive definite, the second inequality in (6.109) implies that $\Delta_k < 0$. On the other hand, from the first inequality in (6.109) we have that

$$\begin{aligned}\varphi_{c_k}(x^k + \alpha p^k) &= \varphi_{c_k}(x^k) + \alpha\varphi'_{c_k}(x^k;\, p^k) + o(\alpha)\\ &\leq \varphi_{c_k}(x^k) + \alpha\Delta_k + o(\alpha)\\ &\leq \varphi_{c_k}(x^k) + \sigma\alpha\Delta_k\end{aligned}$$

for all $\alpha > 0$ small enough. This means that (6.39) will be satisfied after a finite number of backtrackings.

We next establish that when close to the feasible region of (6.96), the solution of subproblem (6.97) is the same as that of the subproblem without the slack variable:

$$\begin{array}{ll}\text{minimize} & f(x^k) + \langle f'(x^k),\, p\rangle + \dfrac{1}{2}\langle H_k p,\, p\rangle\\ \text{subject to} & g_i(x^k) + \langle g_i'(x^k),\, p\rangle + \dfrac{1}{2}\langle G_i^k p,\, p\rangle \leq 0,\ i \in I_k.\end{array} \tag{6.113}$$

This fact will be used later to establish that the penalty parameters $\{c_k\}$ stay fixed from some point on.

Lemma 6.19. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable at $x^k \in \mathbf{R}^n$, let $I_k \subset \{1, \ldots, m\}$ be some index set, and let $H_k \in \mathbf{R}^{n\times n}$ and $G_i^k \in \mathbf{R}^{n\times n}$, $i \in I_k$, be symmetric matrices.*

If $p^k \in \mathbf{R}^n$ is a stationary point of problem (6.113) with an associated Lagrange multiplier $(\mu_i^k,\, i \in I_k) \in \mathbf{R}^{|I_k|}$, then for any $c_k \geq \sum_{i\in I_k}\mu_i^k$ the point $(p^k,\, 0) \in \mathbf{R}^n \times \mathbf{R}$ is stationary in problem (6.97), with an associated Lagrange multiplier $((\mu_i^k,\, i \in I_k),\, c_k - \sum_{i\in I_k}\mu_i^k)$.

Conversely, if $(p^k,\, 0) \in \mathbf{R}^n \times \mathbf{R}$ is a stationary point of problem (6.97), and $((\mu_i^k,\, i \in I_k),\, \nu_k) \in \mathbf{R}^{|I_k|} \times \mathbf{R}$ is an associated Lagrange multiplier, then p^k is a stationary point of problem (6.113), and $(\mu_i^k,\, i \in I_k)$ is an associated Lagrange multiplier.

Proof. Let $p^k \in \mathbf{R}^n$ be any stationary point of problem (6.113), and let $(\mu_i^k,\, i \in I_k) \in \mathbf{R}^{|I_k|}$ be an associated Lagrange multiplier, which means that

$$\begin{gathered}f'(x^k) + H_k p^k + \sum_{i\in I_k}\mu_i^k(g_i'(x^k) + G_i^k p^k) = 0,\\ \mu_i^k \geq 0, \quad g_i(x^k) + \langle g_i'(x^k),\, p^k\rangle + \frac{1}{2}\langle G_i^k p^k,\, p^k\rangle \leq 0,\\ \mu_i^k\left(g_i(x^k) + \langle g_i'(x^k),\, p^k\rangle + \frac{1}{2}\langle G_i^k p^k,\, p^k\rangle\right) = 0, \quad i \in I_k.\end{gathered} \tag{6.114}$$

If $c_k \geq \sum_{i\in I_k}\mu_i^k$, then (6.114) implies that (6.103) holds with $t_k = 0$ and $\nu_k = c_k - \sum_{i\in I_k}\mu_i^k$. Hence, $(p^k,\, 0)$ is a stationary point of problem (6.97), and $((\mu_i^k,\, i \in I_k),\, c_k - \sum_{i\in I_k}\mu_i^k)$ is an associated Lagrange multiplier.

Conversely, if $(p^k,\, 0) \in \mathbf{R}^n \times \mathbf{R}$ is a stationary point of problem (6.97), and $((\mu_i^k,\, i \in I_k),\, \nu_k) \in \mathbf{R}^{|I_k|} \times \mathbf{R}$ is an associated Lagrange multiplier, then (6.103) implies (6.114). Therefore, p^k is a stationary point of problem (6.113), and $(\mu_i^k,\, i \in I_k)$ is an associated Lagrange multiplier. □

Proposition 6.20. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable on $\mathbf{R}^n$, with their derivatives being continuous at a feasible point $\bar{x} \in \mathbf{R}^n$ of problem (6.96). Let the components of g be convex on $\mathbf{R}^n$, and assume that the Slater CQ holds: there exists $\hat{x} \in \mathbf{R}^n$ such that $g(\hat{x}) < 0$. Let $\{x^k\} \subset \mathbf{R}^n$ be any sequence converging to $\bar{x}$, and for each k, let $I_k \subset \{1, \ldots, m\}$ be some index set, let $H_k \in \mathbf{R}^{n\times n}$ be a symmetric matrix, let $G_i^k \in \mathbf{R}^{n\times n}$, $i \in I_k$, be positive semidefinite symmetric matrices, and let there exist $\gamma > 0$ and $\Gamma > 0$ such that*

$$\langle H_k\xi, \xi\rangle \geq \gamma\|\xi\|^2 \quad \forall\, \xi \in \mathbf{R}^n, \quad \|H_k\| \leq \Gamma, \quad \|G_i^k\| \leq \Gamma \quad \forall\, i \in I_k, \forall\, k. \tag{6.115}$$

Then for all k sufficiently large problem (6.113) is feasible, and hence, has the unique solution p^k, which is the unique stationary point of this problem with some associated Lagrange multiplier $(\mu_i^k, i \in I_k)$. Furthermore, for any choice of the latter, the sequences $\{p^k\}$ and $\{\mu^k\}$, where for each k the vector $\mu^k \in \mathbf{R}^m$ has components μ_i^k, $i \in I_k$, and $\mu_i^k = 0$, $i \in \{1, \ldots, m\} \setminus I_k$, are bounded.

Proof. By the convexity of the components of g and by Proposition A.15, for $\hat{x}$ satisfying the Slater CQ it holds that

$$0 > g_i(\hat{x}) = g_i(\hat{x}) - g_i(\bar{x}) \geq \langle g_i'(\bar{x}), \hat{x} - \bar{x}\rangle \quad \forall\, i \in A(\bar{x}). \tag{6.116}$$

For any $\tau \in (0, 1]$ and any $i = 1, \ldots, m$, employing (6.115), we have that

$$\begin{aligned} & g_i(x^k) + \tau\langle g_i'(x^k), \hat{x} - \bar{x}\rangle + \frac{\tau^2}{2}\langle G_i^k(\hat{x} - \bar{x}), \hat{x} - \bar{x}\rangle \\ & \leq \tau\left(\langle g_i'(\bar{x}), \hat{x} - \bar{x}\rangle + \frac{\tau}{2}\Gamma\|\hat{x} - \bar{x}\|^2\right) + \varepsilon_i^k, \end{aligned} \tag{6.117}$$

where

$$\varepsilon_i^k = g_i(x^k) + \tau\langle g_i'(x^k) - g_i'(\bar{x}), \hat{x} - \bar{x}\rangle.$$

Observe that

$$\varepsilon_i^k \to \begin{cases} g_i(\bar{x}) = 0 & \text{if } i \in A(\bar{x}), \\ g_i(\bar{x}) < 0 & \text{if } i \in \{1, \ldots, m\} \setminus A(\bar{x}) \end{cases} \quad \text{as } k \to \infty. \tag{6.118}$$

For each $i = 1, \ldots, m$ there exists $\tau_i > 0$, and for $i \in A(\bar{x})$ also some $a_i > 0$, such that

$$\tau_i\left(\langle g_i'(\bar{x}), \hat{x} - \bar{x}\rangle + \frac{\tau_i}{2}\Gamma\|\hat{x} - \bar{x}\|^2\right) \leq \begin{cases} -2a_i & \text{if } i \in A(\bar{x}), \\ -g_i(\bar{x})/4 & \text{if } i \in \{1, \ldots, m\} \setminus A(\bar{x}), \end{cases} \tag{6.119}$$

where (6.116) was used for $i \in A(\bar{x})$. Given those choices, we have from (6.118) that

$$\varepsilon_i^k \leq \begin{cases} a_i & \text{if } i \in I(\bar{x}), \\ g_i(\bar{x})/2 & \text{if } i \in \{1, \ldots, m\} \setminus A(\bar{x}) \end{cases} \tag{6.120}$$

holds for all k large enough.

Denoting

$$\tau = \min_{i=1,\,\ldots,\,m} \tau_i, \quad a = \min\left\{\min_{i\in A(\bar{x})} a_i,\, -\frac{1}{4}\max_{i\in\{1,\,\ldots,\,m\}\setminus A(\bar{x})} g_i(\bar{x})\right\},$$

and using (6.117) and (6.119), (6.120), we have that for all k large enough $p = \tau(\hat{x} - \bar{x})$ satisfies

$$g_i(x^k) + \langle g_i'(x^k),\, p\rangle + \frac{1}{2}\langle G_i^k p,\, p\rangle \leq -a < 0 \quad \forall\, i = 1,\, \ldots,\, m, \tag{6.121}$$

and in particular, this p is (strictly) feasible in problem (6.113) for any choice of $I_k \subset \{1,\, \ldots,\, m\}$.

Therefore, for all k sufficiently large, problem (6.113) is a feasible quadratic programming problem whose objective function has a positive definite Hessian (recall the first condition in (6.115)). By Proposition 1.5, this problem has the unique global solution p^k, and there are no other local solutions. Furthermore, by (6.121), the constraints of problem (6.113) satisfy the Slater CQ. Hence, by Theorem 1.14, p^k is a stationary point of this problem, i.e., satisfies (6.114) with some $(\mu_i^k,\, i \in I_k) \in \mathbf{R}^{|I_k|}$, and according to Theorem 1.16, there are no other stationary points.

Next, note that for all k large enough, since $p = \tau(\hat{x} - \bar{x})$ defined above is feasible in (6.113), we have that

$$\begin{aligned} \langle f'(x^k),\, p\rangle + \frac{1}{2}\langle H_k p,\, p\rangle &\geq \langle f'(x^k),\, p^k\rangle + \frac{1}{2}\langle H_k p^k,\, p^k\rangle \\ &\geq \|p^k\|\left(\frac{\gamma}{2}\|p^k\| - \|f'(x^k)\|\right), \end{aligned}$$

where the last inequality employs the first condition in (6.115). Since $\{f'(x^k)\}$ and $\{H_k\}$ are bounded, the above relation implies that $\{p^k\}$ must be bounded.

Using the definition of μ^k, for each k we can rewrite the first equality in (6.114) as follows:

$$f'(x^k) + H_k p^k + \sum_{i=1}^{m} \mu_i^k (g_i'(x^k) + G_i^k p^k) = 0. \tag{6.122}$$

Suppose now that $\{\mu^k\}$ is unbounded. Using the second condition in (6.115), and passing onto a subsequence, if necessary, we can assume that $\|\mu^k\| \to \infty$ as $k \to \infty$, $\{p^k\}$ converges to some $\bar{p} \in \mathbf{R}^n$, and $\{G_i^k\}$ converges to some $G_i \in \mathbf{R}^{n\times n}$, $i = 1,\, \ldots,\, m$. Dividing both sides of the equality (6.122) by $\|\mu^k\|$ and passing onto the limit as $k \to \infty$, we then obtain the existence of $\zeta \in \mathbf{R}^m$, $\zeta \neq 0$, such that

$$\sum_{i=1}^{m} \zeta_i (g_i'(\bar{x}) + G_i\bar{p}) = 0, \quad \zeta \geq 0.$$

By the Gordan Theorem of the Alternatives (see Lemma A.3), the latter is equivalent to saying that there exists no $\xi \in \mathbf{R}^n \setminus \{0\}$ such that

$$\langle g_i'(\bar{x}) + G_i\bar{p},\, \xi\rangle < 0 \quad \forall\, i = 1,\, \ldots,\quad m \text{ such that } \zeta_i > 0. \tag{6.123}$$

For any $i = 1, \ldots, m$ such that $\zeta_i > 0$, there exists a subsequence $\{\mu^{k_j}\}$ such that $i \in I_{k_j}$ and $\mu_i^{k_j} > 0$ for all j. For such i, the last line in (6.114) implies that

$$g_i(x^{k_j}) + \langle g_i'(x^{k_j}),\, p^{k_j}\rangle + \frac{1}{2}\langle G_i^{k_j} p^{k_j},\, p^{k_j}\rangle = 0$$

for all j. Passing onto the limit as $j \to \infty$, we obtain that

$$g_i(\bar{x}) + \langle g_i'(\bar{x}),\, \bar{p}\rangle + \frac{1}{2}\langle G_i\bar{p},\, \bar{p}\rangle = 0. \tag{6.124}$$

Passing onto the limit as $k \to \infty$ in (6.121), we also have that

$$g_i(\bar{x}) + \langle g_i'(\bar{x}),\, p\rangle + \frac{1}{2}\langle G_i p,\, p\rangle < 0.$$

Subtracting (6.124) from the latter inequality, and using the assumption that G_i is positive semidefinite, we have that

$$\begin{aligned} 0 &> \langle g_i'(\bar{x}),\, p - \bar{p}\rangle + \frac{1}{2}\langle G_i p,\, p\rangle - \frac{1}{2}\langle G_i\bar{p},\, \bar{p}\rangle \\ &= \langle g_i'(\bar{x}) + G_i\bar{p},\, p - \bar{p}\rangle + \frac{1}{2}\langle G_i(p - \bar{p}),\, p - \bar{p}\rangle \\ &\geq \langle g_i'(\bar{x}) + G_i\bar{p},\, p - \bar{p}\rangle, \end{aligned}$$

which contradicts the nonexistence of $\xi \in \mathbf{R}^n \setminus \{0\}$ satisfying (6.123). We conclude that $\{\mu^k\}$ is bounded. $\square$

Next, note that by (6.105), either c_k is constant starting from some iteration on or $\{c_k\}$ diverges to infinity. We next show that the latter case cannot occur.

Proposition 6.21. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be continuously differentiable on $\mathbf{R}^n$, let the components of g be convex on $\mathbf{R}^n$, and assume that the Slater CQ holds. Let the sequence $\{x^k\} \subset \mathbf{R}^n$ generated by Algorithm 6.15 be bounded, and assume that there exist $\gamma > 0$ and $\Gamma > 0$ such that* (6.115) *is satisfied.*

Then there exists $c > 0$ such that $c_k = c$ for all k large enough.

Proof. Suppose the opposite, i.e., that $c_k \to +\infty$ as $k \to \infty$. Then (6.105) implies that

$$c_k < r_k = \min\{\|p^k\|^{-1},\, \|\mu^k\|_1 + \delta_1\}$$

happens an infinite number of times for the sequences $\{(p^k, t_k)\}$ and $\{(\mu^k, \nu_k)\}$ generated by the algorithm. It then further follows that there exists a subsequence of iteration indices $\{k_j\}$ such that

$$\{p^{k_j}\} \to 0, \quad \|\mu^{k_j}\| \to \infty \quad \text{as } j \to \infty. \tag{6.125}$$

Taking a further subsequence, if necessary, we can assume that $\{x^{k_j}\}$ converges to some $\tilde{x} \in \mathbf{R}^n$. We next consider the two possible cases: when $\tilde{x}$ is feasible in problem (6.96), and when it is infeasible.

Let $\tilde{x}$ be infeasible, i.e., the set $J(\tilde{x})$ defined according to (6.106) is nonempty. Note that from (6.98) to (6.100) it follows that $J(\tilde{x}) \cap I_{k_j} \neq \emptyset$ for all j large enough. By the second inequality in (6.103),

$$g_i(x^{k_j}) + \langle g_i'(x^{k_j}),\, p^{k_j}\rangle + \frac{1}{2}\langle G_i^{k_j} p^{k_j},\, p^{k_j}\rangle \le t_{k_j} \quad \forall\, i \in I_{k_j}.$$

According to (6.115) and (6.125), for $i \in I_{k_j} \setminus J(\tilde{x})$, as $j \to \infty$ the left-hand side of the inequality above tends to $g_i(\tilde{x}) \le 0$, while according to (6.98)–(6.100) and (6.102), the right-hand side tends to $\psi(\tilde{x}) > 0$. Hence, these constraints of problem (6.97) are inactive for all j large enough and, by the last line in (6.103),

$$\mu_i^{k_j} = 0 \quad \forall\, i \in I_{k_j} \setminus J(\tilde{x}).$$

Therefore, formally defining $G_i^{k_j} \in \mathbf{R}^{n\times n}$ as zero matrices for $i \in J(\tilde{x}) \setminus I_{k_j}$, we can obtain from the first equality in (6.103) that

$$f'(x^{k_j}) + H_{k_j} p^{k_j} + \sum_{i \in J(\tilde{x})} \mu_i^{k_j} (g_i'(x^{k_j}) + G_i^{k_j} p^{k_j}) = 0.$$

Dividing both sides of this equality by $\|\mu^{k_j}\|$ and passing onto the limit as $j \to \infty$, by (6.115) and (6.125), we obtain the existence of ζ_i, $i \in J(\tilde{x})$, not all equal to zero and such that (6.107) holds with $x = \tilde{x}$, which contradicts Lemma 6.16.

Suppose now that $\tilde{x}$ is feasible in problem (6.96). By Proposition 6.20, for all j large enough problem (6.113) with $k = k_j$ has a stationary point $\tilde{p}^{k_j} \in \mathbf{R}^n$ with an associated Lagrange multiplier $(\tilde{\mu}_i^{k_j},\, i \in I_{k_j})$, and the sequence $\{\tilde{\mu}^{k_j}\}$, where for each j the vector $\tilde{\mu}^{k_j} \in \mathbf{R}^m$ has components $\tilde{\mu}_i^{k_j}$, $i \in I_{k_j}$, and $\tilde{\mu}_i^{k_j} = 0$, $i \in \{1, \ldots, m\} \setminus I_{k_j}$, is bounded.

Since $c_k \to +\infty$ as $k \to \infty$, we have that $c_{k_j} \ge \sum_{i \in I_{k_j}} \tilde{\mu}_i^{k_j}$, and therefore, by Lemma 6.19, the point $(\tilde{p}^{k_j},\, 0) \in \mathbf{R}^n \times \mathbf{R}$ is stationary in problem (6.97) with $k = k_j$, and hence, is the unique solution of this problem. Therefore, $(p^{k_j},\, t_{k_j}) = (\tilde{p}^{k_j},\, 0)$. Applying again Lemma 6.19 to this stationary point, and to the associated Lagrange multiplier $((\mu_i^{k_j},\, i \in I_{k_j}),\, \nu_{k_j})$, we now obtain that

$(\mu_i^{k_j},\, i \in I_{k_j})$ is a Lagrange multiplier associated with the stationary point $\tilde{p}^{k_j}$ of problem (6.113) with $k = k_j$, and then the second relation in (6.125) contradicts Proposition 6.20. □

We are now ready to establish global convergence of Algorithm 6.15.

Theorem 6.22. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable on $\mathbf{R}^n$, with their derivatives being Lipschitz-continuous on $\mathbf{R}^n$, let the components of g be convex on $\mathbf{R}^n$, and assume that the Slater CQ holds. Let the sequence $\{x^k\} \subset \mathbf{R}^n$ generated by Algorithm 6.15 be bounded, and assume that for all k large enough $I_k = \{1, \ldots, m\}$, and that there exist $\gamma > 0$ and $\Gamma > 0$ such that (6.115) is satisfied.*

Then the sequence $\{\mu^k\} \subset \mathbf{R}^m$ generated by Algorithm 6.15 is also bounded, and for every accumulation point $(\bar{x}, \bar{\mu}) \in \mathbf{R}^n \times \mathbf{R}^m$ of $\{(x^k, \mu^k)\}$ it holds that $\bar{x}$ is a stationary point of problem (6.96), while $\bar{\mu}$ is an associated Lagrange multiplier.

In particular, if f is convex, then every accumulation point of $\{x^k\}$ is a solution of (6.96).

Proof. We first show that the sequence of stepsizes $\{\alpha_k\}$ is bounded away from zero. Let $\ell > 0$ be the Lipschitz constant for the gradient of f and for the gradients of the components of g on $\mathbf{R}^n$. Similarly to the proof of Theorem 6.9, for all $\alpha \in [0, 1]$ and any $i = 1, \ldots, m$, we have (6.47) and (6.50). For $i \in I_k$, we further obtain that

$$\begin{aligned}
&\max\{0,\, g_i(x^k) + \alpha\langle g_i'(x^k),\, p^k\rangle\} \\
&= \max\{0,\, \alpha(g_i(x^k) + \langle g_i'(x^k),\, p^k\rangle) + (1-\alpha)g_i(x^k)\} \\
&\le \alpha\max\{0,\, g_i(x^k) + \langle g_i'(x^k),\, p^k\rangle\} + (1-\alpha)\max\{0,\, g_i(x^k)\} \\
&\le \alpha\max\left\{0,\, g_i(x^k) + \langle g_i'(x^k),\, p^k\rangle + \frac{1}{2}\langle G_i^k p^k,\, p^k\rangle\right\} \\
&\quad + (1-\alpha)\max\{0,\, g_i(x^k)\} \\
&\le \alpha t_k + (1-\alpha)\max\{0,\, g_i(x^k)\},
\end{aligned} \tag{6.126}$$

where the first inequality is by the convexity of $\max\{0, \cdot\}$, the second is by positive semidefiniteness of G_i^k, and the last is by (6.103). Combining (6.47) and (6.126), we conclude that

$$\max\{0,\, g_i(x^k + \alpha p^k)\} \le \alpha t_k + (1-\alpha)\max\{0,\, g_i(x^k)\} + \frac{\ell\alpha^2}{2}\|p^k\|^2 \quad \forall\, i \in I_k. \tag{6.127}$$

Using (6.50), (6.98), (6.101), (6.104), (6.127), the equality $I_k = \{1, \ldots, m\}$, and positive definiteness of H_k, for all k large enough we obtain that

$$\begin{aligned}
\varphi_{c_k}(x^k+\alpha p^k) &= f(x^k+\alpha p^k)+c_k\max\{0,\, g_1(x^k+\alpha p^k),\, \ldots,\, g_m(x^k+\alpha p^k)\}\\
&\le f(x^k)+\alpha\langle f'(x^k),\, p^k\rangle\\
&\quad +c_k(1-\alpha)\max\{0,\, g_1(x^k),\, \ldots,\, g_m(x^k)\}\\
&\quad +c_k\alpha t_k+C_k\alpha^2\|p^k\|^2\\
&= \varphi_{c_k}(x^k)+\alpha(\langle f'(x^k),\, p^k\rangle-c_k(\psi(x^k)-t_k))+C_k\alpha^2\|p^k\|^2\\
&= \varphi_{c_k}(x^k)+\alpha\left(\Delta_k-\frac{1}{2}\langle H_k p^k,\, p^k\rangle\right)+C_k\alpha^2\|p^k\|^2\\
&\le \varphi_{c_k}(x^k)+\alpha\Delta_k+C_k\alpha^2\|p^k\|^2,
\end{aligned}$$

where

$$C_k=\frac{\ell}{2}(1+c_k)>0. \tag{6.128}$$

By a direct comparison of the latter relation with (6.39), we conclude that (6.39) is satisfied if

$$C_k\alpha\|p^k\|^2\le-(1-\sigma)\Delta_k,$$

i.e., it is satisfied for all $\alpha\in(0,\,\bar{\alpha}_k]$, where

$$\bar{\alpha}_k=\frac{(\sigma-1)\Delta_k}{C_k\|p^k\|^2}. \tag{6.129}$$

By Proposition 6.21, there exists $c>0$ such that $c_k=c$ for all k sufficiently large. Furthermore, by (6.98), (6.103), by the second inequality in (6.109), by positive semidefiniteness of G_i^k, $i\in I_k$, and by the first condition in (6.115), we have that

$$-\Delta_k\ge\frac{1}{2}\langle H_k p^k,\, p^k\rangle\ge\frac{\gamma}{2}\|p^k\|^2. \tag{6.130}$$

Combining this with (6.128) and (6.129) gives

$$\bar{\alpha}_k\ge\frac{(1-\sigma)\gamma}{\ell(1+c)}>0.$$

We then conclude that there exists $\varepsilon>0$, that does not depend on k, such that

$$\alpha_k\ge\varepsilon \tag{6.131}$$

for all k large enough.

Now taking into account (6.39), (6.130), and (6.131), and employing again the equality $c_k=c$, for k large enough we obtain that

$$\varphi_c(x^k)-\varphi_c(x^{k+1})\ge-\sigma\alpha_k\Delta_k\ge\sigma\frac{\varepsilon\gamma}{2}\|p^k\|^2. \tag{6.132}$$

Since $\{x^k\}$ is bounded, $\{\varphi_c(x^k)\}$ is also bounded, by continuity. By (6.132), it is also nonincreasing, and hence, it converges. Then (6.132) further implies that

$$\{p^k\}\to 0,\quad \Delta_k\to 0\quad \text{as } k\to\infty. \tag{6.133}$$

Furthermore, for all k large enough, by step 5 of Algorithm 6.15, the fact that $c_{k+1} = c_k = c$ implies that

$$c \geq r_k = \min\{1/\|p^k\|,\, \|\mu^k\|_1 + \delta_1\}.$$

Taking into account (6.133), we conclude that $\{\mu^k\}$ is bounded and, in particular, $c > \sum_{i\in I_k} \mu_i^k$. Using the second equality in (6.103), the latter implies that $\nu_k > 0$, and hence, by the last line in (6.103), $t_k = 0$ for all k large enough.

Now passing onto the limit as $k \to \infty$ in (6.103) (where $I_k = \{1, \ldots, m\}$), we obtain that every accumulation point of $\{(x^k, \mu^k)\}$ satisfies the KKT optimality system for problem (6.96).

The final assertion of the theorem follows now from Theorem 1.16. □

We next show that if the matrices H_k and G_i^k are (in a certain sense) asymptotically good approximations of $f''(x^k)$ and $g_i''(x^k)$, $i \in I_k$, respectively, then in Algorithm 6.15 we have $\alpha_k = 1$ for all k sufficiently large. Note that this conclusion does not require any conditions on the solution at all, in particular it does not depend on assumptions related to the local convergence rate (a CQ or the SOSC). Therefore, unlike in the case of SQP methods, Maratos effect naturally does not occur in SQCQP methods, and thus no modifications are needed to avoid it. Then, under the additional assumptions that guarantee the local superlinear rate of the SQCQP method (see Sect. 4.3.5), the rate of convergence of Algorithm 6.15 is also superlinear.

Proposition 6.23. *Under the assumptions of Theorem 6.22, let f and g be twice continuously differentiable on $\mathbf{R}^n$, and assume that*

$$\begin{gathered}\langle (f''(x^k) - H_k)p^k,\, p^k\rangle = o(\|p^k\|^2),\\ \max_{i\in I_k} |\langle (g_i''(x^k) - G_i^k)p^k,\, p^k\rangle| = o(\|p^k\|^2)\end{gathered} \tag{6.134}$$

as $k \to \infty$.

Then $\alpha_k = 1$ for all k large enough.

Proof. As established in the proof of Theorem 6.22, $c_k = c > 0$ and $t_k = 0$ for all k large enough, and (6.133) is valid.

For $i \in I_k$ and any sufficiently large k, from the second relation in (6.134) we have that

$$\begin{aligned}&\max\{0,\, g_i(x^k + p^k)\}\\ &= \max\{0,\, g_i(x^k + p^k)\} - \max\left\{0,\, g_i(x^k) + \langle g_i'(x^k),\, p^k\rangle + \frac{1}{2}\langle G_i^k p^k,\, p^k\rangle\right\}\\ &\leq \left| g_i(x^k + p^k) - g_i(x^k) - \langle g_i'(x^k),\, p^k\rangle - \frac{1}{2}\langle g_i''(x^k)p^k,\, p^k\rangle\right|\\ &\quad + \frac{1}{2}|\langle (g_i''(x^k) - G_i^k)p^k,\, p^k\rangle|\\ &= o(\|p^k\|^2)\end{aligned} \tag{6.135}$$

as $k \to \infty$, where the first equality follows from the second inequality in (6.103) with $t_k = 0$, and the inequality follows from (6.48).

Since $I_k = \{1, \ldots, m\}$ for all k large enough, from (6.98), (6.101), (6.104), (6.135), and from the first relation in (6.134) we further obtain that

$$\begin{aligned}
\varphi_c(x^k + p^k) &= f(x^k + d^k) + c \max\{0, g_1(x^k + p^k), \ldots, g_m(x^k + p^k)\} \\
&= f(x^k + p^k) + o(\|p^k\|^2) \\
&= f(x^k) + \langle f'(x^k), p^k \rangle + \frac{1}{2}\langle f''(x^k)p^k, p^k \rangle + o(\|p^k\|^2) \\
&= \varphi_c(x^k) + \langle f'(x^k), p^k \rangle + \frac{1}{2}\langle H_k p^k, p^k \rangle - c\psi(x^k) \\
&\quad + o(\|p^k\|^2) \\
&= \varphi_c(x^k) + \Delta_k + o(\|p^k\|^2) \qquad (6.136)
\end{aligned}$$

as $k \to \infty$, where for the second equality the mean-value theorem (Theorem A.10, (a)) is used, and the last equality employs the fact that $t_k = 0$. By relation (6.130) established in the proof of Theorem 6.22,

$$\|p^k\|^2 = O(-\Delta_k) \qquad (6.137)$$

as $k \to \infty$, and hence, (6.133) and (6.136) imply that (6.39) holds with $\alpha_k = 1$ for all k large enough. □

6.3 Trust-Region and Filter Methods

A natural alternative to using linesearch along the direction obtained by solving a SQP subproblem (like the approach presented in Sect. 6.2) would be to generate candidate points solving SQP subproblems with an additional adjustable trust-region constraint, as described in Sect. 2.3 for unconstrained optimization problems. Moreover, in both the linesearch and the trust-region schemes, instead of a merit function the so-called filter paradigm can be used to evaluate candidate points for acceptance. Trust-region approaches using merit functions are thoroughly treated in [45]; and we shall limit ourselves to a brief survey in Sect. 6.3.1. Instead, we present one specific algorithm based on filter globalization. In Sect. 6.3.2 we describe a generic filter scheme, without specifying how the candidate points are generated. In Sect. 6.3.3 we show that the required assumptions are satisfied when computing the candidate points by a variant of the trust-region SQP method.

6.3.1 Trust-Region Approaches to Sequential Quadratic Programming

We limit our brief discussion of this class of algorithms to the equality-constrained problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \end{array} \tag{6.138}$$

where the objective function $f : \mathbf{R}^n \to \mathbf{R}$ and the constraints mapping $h : \mathbf{R}^n \to \mathbf{R}^l$ are differentiable.

The first attempt might consist of simply adding a trust-region constraint to the SQP subproblem, reducing the trust-region parameter (if needed) to achieve sufficient descent for some merit function for problem (6.138), e.g., one of the merit functions discussed in Sect. 6.1. Given the current iterate $x^k \in \mathbf{R}^n$ and a symmetric matrix $H_k \in \mathbf{R}^{n \times n}$ (possibly depending on the current multiplier estimates and the derivatives information, to approximate the Hessian of the Lagrangian), the iteration subproblem would then be

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k), p \rangle + \dfrac{1}{2}\langle H_k p, p \rangle \\ \text{subject to} & h(x^k) + h'(x^k)p = 0, \ \|p\| \le \delta_k, \end{array} \tag{6.139}$$

for some adjustable trust-region radius $\delta_k > 0$. Then, if for the obtained solution $p^k = p^k(\delta_k)$ the point $x^k + p^k$ is not satisfactory, the parameter δ_k is changed, the subproblem is solved again, the resulting new point is judged for being satisfactory or not, etc. However, it is clear that this simple strategy would not work, in general. Even if the linearized constraint in (6.139) is consistent, the subproblem (6.139) may become infeasible for small values of δ_k; and δ_k may need to be small for the quadratic model "to be trusted" around x^k. Thus more sophisticated techniques are required. There are three main trust-region SQP approaches, although there are numerous variants within each of them.

The *full-space approach* [42, 220, 267] consists in relaxing the linearized constraint in (6.139) enough to ensure that the resulting modification is feasible. Specifically, the subproblem

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k), p \rangle + \dfrac{1}{2}\langle H_k p, p \rangle \\ \text{subject to} & \|h(x^k) + h'(x^k)p\| \le t_k, \ \|p\| \le \delta_k, \end{array}$$

is solved to obtain p^k. The new trial point is then given by $x^k + p^k$. To define a proper value of t_k, one needs to estimate the optimal value of

$$\begin{array}{ll} \text{minimize} & \|h(x^k) + h'(x^k)p\| \\ \text{subject to} & \|p\| \le \delta_k, \end{array}$$

which can be done in a number of ways. For example, based on a fraction of the Cauchy decrease condition for $\|h(x^k) + h'(x^k)p\|^2$, as in [42].

An alternative *composite-step approach* computes the new trial point in two phases. Roughly speaking, the first phase aims at reducing the infeasibility, while the second decreases the quadratic model in the null space of $h'(x^k)$. Methods of this class include, e.g., [33, 59, 93, 210, 230, 263].

Details can be implemented in a number of different ways, especially concerning the feasibility step. So it is convenient to assume that the feasibility step p_n^k (also called *normal step*) is already computed, and first describe the optimality phase. For this phase to have "enough freedom," the previously computed feasibility step should be shorter than the trust-region radius:

$$\|p_n^k\| \leq \beta\delta_k \tag{6.140}$$

for some $\beta \in (0, 1)$. The latter is achieved by simply restricting the size of the step p_n^k in the feasibility phase. Then the optimality step (also called *tangential step*) is given by solving

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, p_t + p_n^k\rangle + \dfrac{1}{2}\langle H_k(p_t + p_n^k),\, p_t + p_n^k\rangle \\ \text{subject to} & h'(x^k)p_t = 0,\ \|p_t\| \leq \sqrt{1-\beta^2}\,\delta_k \end{array} \tag{6.141}$$

to obtain p_t^k. This subproblem may be solved approximately; for example, just enough to give descent as good as the appropriately defined Cauchy point [59]. The new trial point is then given by $x^k + p_n^k + p_t^k$.

We next outline some possibilities for the feasibility phase. Observe first that one can modify the subproblem (6.139) to make it feasible and its solution to satisfy (6.140) by scaling the value of $h(x^k)$ in the linearized constraint [33, 263]. This leads to computing p_n^k by solving

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, p_n\rangle + \dfrac{1}{2}\langle H_k p_n,\, p_n\rangle \\ \text{subject to} & t_k h(x^k) + h'(x^k)p_n = 0,\ \|p_n\| \leq \beta\delta_k, \end{array}$$

for $\beta \in (0, 1)$ and an appropriately chosen $t_k \in (0, 1]$ (for any fixed β there exists $t_k \in (0, 1]$ such that the subproblem above is feasible and thus solvable). However, choosing a suitable value of t_k is clearly nontrivial here. A related approach [210] solves

$$\begin{array}{ll} \text{minimize} & \dfrac{1}{2}\|h(x^k) + h'(x^k)p_n\|^2 \\ \text{subject to} & \|p_n\| \leq \beta\delta_k, \end{array}$$

which has the advantage that no extra parameters need to be estimated. Again, with proper care, this subproblem can also be solved approximately [59].

Another set of assumptions for the feasibility phase is used in [230]. Specifically, the step p_n^k must satisfy the linearized constraint $h(x^k) + h'(x^k)p_n = 0$, the condition (6.140), and be reasonably scaled with respect to the constraints. The latter means that there exist constants $\ell_1 > 0$ and $\ell_2 > 0$ such that if $\|h(x^k)\| \leq \ell_2$, then $\|p_n^k\| \leq \ell_1 \|h(x^k)\|$. This can be achieved under reasonable conditions, assuming that an external feasibility restoration procedure is activated when a point satisfying the linearized constraint and (6.140) cannot be computed. The restoration procedure moves the current iterate closer to the feasible set of the original problem (6.138); for more on feasibility restoration, see Sect. 6.3.3.

Finally, as another possibility, we mention trust-region methods based on unconstrained minimization of some suitable exact penalty function for problem (6.138). If the l_∞-penalty is employed [274], the iteration subproblem can be written in the form

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, p\rangle + c_k t + \dfrac{1}{2}\langle H_k p,\, p\rangle \\ \text{subject to} & \|h(x^k) + h'(x^k)p\|_\infty \leq t,\ \|p\|_\infty \leq \delta_k, \end{array}$$

which is a quadratic programming problem. The key to this method is to properly select/update the penalty parameter $c_k > 0$. This approach is related to the elastic mode SQP (see (6.62)) and to the method of [87], where the l_1-penalty is used (in the latter case the subproblems are not quadratic programs, however).

After a new trial point is computed by any of the above trust-region approaches, it can be evaluated according to the decrease it provides (or not) for a chosen merit function. Since this part is conceptually similar to the use of merit functions in conjunction with linesearch in Sect. 6.2, we shall not go into details, referring the reader to [45]. Instead, we shall present here the alternative strategy of globalizing trust-region SQP by the filter technique.

6.3.2 A General Filter Scheme

Any algorithm for a minimization problem with constraints must deal with two different (and possibly conflicting) criteria related, respectively, to minimality and to feasibility. Minimality is naturally measured by the objective function f; infeasibility is typically measured by penalization of constraints violation. Let the problem be given in the generic format

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in D, \end{array} \tag{6.142}$$

where $f : \mathbf{R}^n \to \mathbf{R}$ is continuous and $D \subset \mathbf{R}^n$ is closed. As in Sect. 6.1, define a penalty $\psi : \mathbf{R}^n \to \mathbf{R}_+$, i.e., a function satisfying

$$\psi(x) > 0 \quad \forall\, x \notin D, \quad \psi(x) = 0 \quad \forall\, x \in D, \tag{6.143}$$

measuring in some reasonable way the feasible set's constraints violation. Both criteria, f and ψ, must be optimized eventually, and the algorithm should follow a certain balance between them, at every step of the iterative process. Traditionally, this task was approached by minimizing a weighted sum of f and ψ, i.e., a penalty function (see Sect. 6.1). More recently, an alternative idea, called the *filter strategy*, has been introduced in [91]. Various filter-based methods and their analyses can be found in [41, 93, 94, 105, 165, 264, 265]. Here, at least conceptually, we follow the approach of [230].

Filter algorithms define a *forbidden region*, where the new iterate cannot belong, by memorizing optimality and infeasibility measures $(f(x^i),\, \psi(x^i))$ at well-chosen past iterations, and then avoiding points dominated by these pairs according to the usual *Pareto domination rule*:

$$\text{“}x \in \mathbf{R}^n \;\text{ dominates }\; y \in \mathbf{R}^n \;\text{ if and only if }\; f(x) \le f(y) \;\text{ and }\; \psi(x) \le \psi(y)\text{”}.$$

Roughly speaking, a candidate point is accepted by the filter as the next iterate whenever it is not dominated by any of the previous iterates. As one would expect, a slightly stronger condition is required for convergence: points that do not make significant enough progress with respect to every pair in the filter must be avoided. In other words, at least one of the measures (f or ψ) should be reduced by a sufficient margin.

Let the two parameters $\sigma_1,\, \sigma_2 \in (0,\, 1)$ be given. These parameters are used to define the forbidden region by shifting the optimality and infeasibility values, as described next. At the k-th iteration, the filter F_k is formed by the union of pairs

$$(\tilde{f}^j,\, \tilde{\psi}^j) = (f(x^j) - \sigma_1\psi(x^j),\, \sigma_2\psi(x^j)),$$

where indices $j \in \{0,\, 1,\, \ldots,\, k-1\}$ correspond to (some of) past iterations, such that no pair in F_k is dominated (in the Pareto sense) by any other pair. Given the previous iterate x^k, at the beginning of the k-th iteration the pair

$$(\tilde{f}^k,\, \tilde{\psi}^k) = (f(x^k) - \sigma_1\psi(x^k),\, \sigma_2\psi(x^k))$$

is *temporarily* introduced into the filter, see Algorithm 6.24 below. Together with pairs in F_k, this pair defines the current forbidden region $\tilde{\mathcal{F}}_k$. Essentially, the algorithm (by generating candidate points) aims at producing a point that is not forbidden; see Fig. 6.6. This point will be the new iterate x^{k+1}.

Obviously, how exactly this is achieved is the core of each specific algorithm. What we describe now is merely a generic framework. After having found the new iterate, the temporary pair enters the new filter F_{k+1} only if this iteration did not produce a decrease in f. Such iterations will be called *ψ-iterations.* Iterations that did reduce the objective function value will be referred to as *f-iterations*; see Algorithm 6.24 below. Whenever the filter has changed, it is cleaned: every pair dominated by the newly added pair is removed, and the algorithm proceeds.

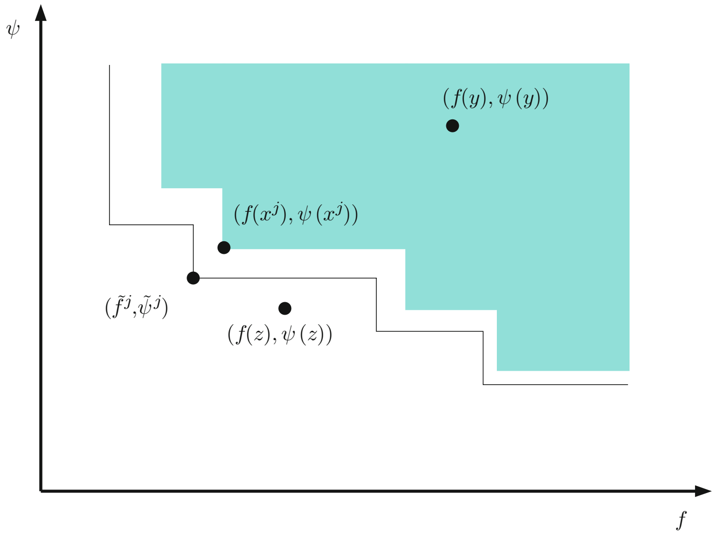

Fig. 6.6 Filter: point z is acceptable; point y is not

Algorithm 6.24 Choose parameters σ_1, $\sigma_2 \in (0, 1)$. Choose $x^0 \in \mathbf{R}^n$ and set $F_0 = \emptyset$, $\mathcal{F}_0 = \emptyset$, $k = 0$.

1. Define the temporary pair, the current filter and the forbidden region, respectively, as follows:

$$(\tilde{f}^k, \tilde{\psi}^k) = (f(x^k) - \sigma_1\psi(x^k), \sigma_2\psi(x^k)),$$

$$\tilde{F}_k = F_k \cup \{(\tilde{f}^k, \tilde{\psi}^k)\},$$

$$\tilde{\mathcal{F}}_k = \mathcal{F}_k \cup \{x \in \mathbf{R}^n \mid f(x) \geq \tilde{f}^k, \psi(x) \geq \tilde{\psi}^k\}.$$

2. Compute (somehow) the next iterate $x^{k+1} \notin \tilde{\mathcal{F}}_k$ (perhaps with some additional requirements).
3. Update the filter as follows. If $f(x^{k+1}) < f(x^k)$, then set

$$F_{k+1} = F_k, \quad \mathcal{F}_{k+1} = \mathcal{F}_k$$

(f-iteration). Else set

$$F_{k+1} = \tilde{F}_k, \quad \mathcal{F}_{k+1} = \tilde{\mathcal{F}}_k$$

(ψ-iteration). In the latter case, remove from the filter every pair dominated by the newly added pair.
4. Adjust k by 1 and go to step 1.

Algorithm 6.24 is, of course, a framework too generic. Many details need to be specified, especially in step 2, for the scheme to be provably convergent and applicable to analyze specific computational methods. We next mention a couple of issues.

First, it may happen in certain undesirable (but not uncommon) situations that the basic iteration of a given method, such as SQP, cannot produce a new point outside of the forbidden region $\tilde{\mathcal{F}}_k$. For example, the SQP subproblem can simply be infeasible at this iteration. In such cases, it is typical that an auxiliary restoration phase is invoked to recover. Basically, it consists in improving the infeasibility measure ψ only, ignoring optimality for the time being; see Sect. 6.3.3.

Second, when close to feasibility (a situation dual to the above), it must be ensured that the method makes enough progress towards optimality, i.e., that the objective function is reduced. To that end, a certain additional assumption concerning step 2 is made; see Proposition 6.27 below. In Sect. 6.3.3, it is shown that this assumption is satisfied by a composite-step trust-region SQP iteration.

We start with some relations which follow directly from the construction of the filter mechanism.

Proposition 6.25. *Let a sequence $\{x^k\}$ be generated by Algorithm 6.24. Then for all k, the following statements are valid:*

(a) At least one of the following two conditions holds:

$$f(x^{k+1}) < f(x^k) - \sigma_1 \psi(x^k),$$

and/or

$$\psi(x^{k+1}) < \sigma_2 \psi(x^k).$$

(b) It holds that $\tilde{\psi}^j > 0$ for all $j \in \{0, 1, \ldots, k\}$ such that $(\tilde{f}^j, \tilde{\psi}^j) \in F_k$.
(c) It holds that $x^{k+j} \notin \mathcal{F}_{k+1}$ for all $j \geq 1$.

Proof. Item (a) is a direct consequence of the fact that $x^{k+1} \notin \tilde{\mathcal{F}}_k$.

We next prove item (b). For any $j \in \{0, 1, \ldots, k\}$, the pair $(\tilde{f}^j, \tilde{\psi}^j)$ can only be in the filter F_k if the corresponding j-th iteration was a ψ-iteration. If it were the case that $\psi(x^j) = 0$, then the second relation in item (a), written for $k = j$, is not possible. Hence, the first relation holds. Therefore, $f(x^{j+1}) < f(x^j)$. But then the j-th iteration was an f-iteration, in which case the corresponding pair would not have entered the filter. It follows that an inclusion into the filter at any k-th iteration can occur only when $\psi(x^k) > 0$, in which case $\tilde{\psi}^k > 0$, implying the assertion.

We finish by establishing item (c). For all $j \geq 1$, by construction, it holds that $x^{k+j} \notin \tilde{\mathcal{F}}_{k+j-1}$. As is easily seen, the forbidden region cannot contract: $\mathcal{F}_{k+1} \subset \mathcal{F}_{k+j}$. Also, the forbidden region is contained in the temporary forbidden region of the previous iteration: $\mathcal{F}_{k+j} \subset \tilde{\mathcal{F}}_{k+j-1}$. It follows that $x^{k+j} \notin \mathcal{F}_{k+1}$. $\square$

We now show that if the method generates a sequence of iterates which has an infeasible accumulation point, then for all the iterates close enough to this accumulation point the resulting step is an f-iteration.

Proposition 6.26. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $\psi : \mathbf{R}^n \to \mathbf{R}_+$ satisfying (6.143) be continuous on $\mathbf{R}^n$. Suppose a subsequence $\{x^{k_i}\}$ of iterates generated by Algorithm 6.24 converges to a point $\bar{x} \notin D$ as $i \to \infty$.*

Then for each index i large enough the iteration indexed by k_i is an f-iteration.

Proof. By the continuity of f and ψ, we have that

$$f(x^{k_i}) \to f(\bar{x}), \quad \psi(x^{k_i}) \to \psi(\bar{x}) > 0$$

as $i \to \infty$.

For contradiction purposes, suppose that there exists a further infinite subsequence $\{x^{k_{i_j}}\}$ of iterates generated by ψ-iterations. Since

$$f(x^{k_i}) - f(x^{k_{i_j}}) + \sigma_1 \psi(x^{k_{i_j}}) \to \sigma_1 \psi(\bar{x}) > 0,$$

and similarly

$$\psi(x^{k_i}) - \sigma_2 \psi(x^{k_{i_j}}) \to (1 - \sigma_2)\psi(\bar{x}) > 0,$$

as i, $j \to \infty$, it follows that for all indices i and j sufficiently large

$$f(x^{k_i}) > f(x^{k_{i_j}}) - \sigma_1 \psi(x^{k_{i_j}})$$

and

$$\psi(x^{k_i}) > \sigma_2 \psi(x^{k_{i_j}}).$$

These two relations mean that $x^{k_i} \in \tilde{\mathcal{F}}_{k_{i_j}}$. Furthermore, since the index k_{i_j} corresponds to a ψ-iteration, the filter update gives $\mathcal{F}_{k_{i_j}+1} = \tilde{\mathcal{F}}_{k_{i_j}}$. As a result, $x^{k_i} \in \mathcal{F}_{k_{i_j}+1}$. Now taking i and j such that $k_i \geq k_{i_j} + 1$, the latter gives a contradiction with Proposition 6.25, (c). □

The final result requires an assumption, already mentioned above, that when close to a feasible point which is not stationary, the algorithm should make sufficient progress towards optimality, i.e., in reducing the value of the objective function. This progress is measured by the following quantity:

$$\gamma_k = \min\{1, \min\{\tilde{\psi}^j \mid (\tilde{f}^j, \tilde{\psi}^j) \in F_k\}\}. \tag{6.144}$$

Note that at the level of generality of Algorithm 6.24, defining explicitly the notion of stationarity is not even necessary here (as it is "hidden" in the assumption in question). In the specific context of SQP, this would be dealt with in Sect. 6.3.3.

The convergence result below also assumes boundedness of the sequence of iterates and affirms that at least one of the accumulation points is stationary. Convergence results of this nature are typical for filter methods.

Theorem 6.27. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $\psi : \mathbf{R}^n \to \mathbf{R}_+$ satisfying (6.143) be continuous on $\mathbf{R}^n$. Suppose Algorithm 6.24 has the following property: for any point $\bar{x}$ which is feasible but not stationary for problem (6.142), there exist a neighborhood U of $\bar{x}$ and $\Delta > 0$ such that whenever $x^k \in U$, the next iterate x^{k+1} satisfies*

$$f(x^k) - f(x^{k+1}) \geq \Delta\gamma_k, \tag{6.145}$$

where γ_k is given by (6.144).

Then if the sequence $\{x^k\}$ is bounded, it has an accumulation point which is stationary for problem (6.142).

Proof. Suppose first that there exists an infinite number of ψ-iterations. Taking a convergent subsequence $\{x^{k_i}\}$ along those indices, we obtain an accumulation point $\bar{x}$. Since all iterations indexed by k_i are ψ-iterations, by Proposition 6.26, $\bar{x}$ cannot be infeasible. Thus it is feasible. If it is also stationary, the needed assertion is obtained. If it is not, by the assumption (6.145), for all indices i large enough

$$f(x^{k_i}) - f(x^{k_i+1}) > 0,$$

where the strict inequality is by Proposition 6.25, (b). But this means that these are f-iterations, contradicting the current assumption. This concludes the proof for the case when there is an infinite number of ψ-iterations.

Now assume that the number of ψ-iterations is finite, i.e., there exists k_0 such that all the iterations indexed by $k \geq k_0$ are f-iterations. Then the sequence $\{f(x^k)\}$ is monotone (for $k \geq k_0$). Since it is bounded (by boundedness of $\{x^k\}$ and continuity of f), it converges. Hence,

$$f(x^k) - f(x^{k+1}) \to 0. \tag{6.146}$$

Also, since the filter is updated on ψ-iterations only, it holds that $F_k = F_{k_0}$ for $k \geq k_0$. Then according to (6.144) and Proposition 6.25, (b), there exists $\bar{\gamma} > 0$ such that

$$\gamma_k = \bar{\gamma} \quad \forall\, k \geq k_0. \tag{6.147}$$

If there exists a subsequence $\{x^{k_i}\}$ for which the first relation in Proposition 6.25, (a), holds, then (6.146) implies that $\psi(x^{k_i}) \to 0$ as $i \to \infty$. If there is no subsequence with the above property, then the second relation in Proposition 6.25, (a), holds for all indices large enough which shows that $\psi(x^k) \to 0$ (at a linear rate). In either case, the conclusion is that $\{x^k\}$ has a feasible accumulation point $\bar{x}$.

If $\bar{x}$ is stationary, the assertion is established. If it is not, then (6.145) holds an infinite number of times, with (6.147), which is in contradiction with (6.146). $\square$

We conclude with some properties of the quantity γ_k defined in (6.144) that will be useful in the treatment of filter-SQP algorithm to follow. To that end, define also

$$\tilde{\gamma}_k = \min\{1, \min\{\tilde{\psi}^j \mid (\tilde{f}^j, \tilde{\psi}^j) \in F_k,\ \tilde{f}^j \leq f(x^k)\}\}. \tag{6.148}$$

Proposition 6.28. *Let a sequence* $\{x^k\}$ *be generated by Algorithm 6.24.*

Then for each k*, the following statements hold for* γ_k *and* $\tilde{\gamma}_k$ *defined by* (6.144) *and* (6.148)*, respectively:*

(a) $0 < \gamma_k \leq \tilde{\gamma}_k \leq 1$ *and* $\gamma_k \leq \sqrt{\tilde{\gamma}_k}$*.*
(b) For a given $\tilde{x}^k \in \mathbf{R}^n$*, if* $f(\tilde{x}^k) < f(x^k) - \sigma_1 \psi(x^k)$ *but* $\tilde{x}^k \in \mathcal{F}_k$*, then* $\psi(\tilde{x}^k) \geq \tilde{\gamma}_k$*.*
(c) $\tilde{\gamma}_k \geq \psi(x^k)$*.*

Proof. Item (a) is immediate by Proposition 6.25, (b), and by the definitions (6.144) and (6.148).

Concerning item (b), in view of the assumption, at least one infeasibility measure involved in the definition of $\tilde{\gamma}_k$ must be increased for having $\tilde{x}^k \in \mathcal{F}_k$.

For item (c), if it were the case that $\tilde{\gamma}_k < \psi(x^k)$, the definition of $\tilde{\gamma}_k$ and of x^k being in the filter would enter into contradiction (x^k would be dominated by some pair in the filter). □

6.3.3 A Trust-Region Filter Sequential Quadratic Programming Method

Consider again the equality-constrained problem (6.138). We next show that one specific SQP composite-step approach mentioned in Sect. 6.3.2 satisfies the hypothesis of Theorem 6.27 on convergence of general filter-globalized algorithms. In particular, the key descent condition (6.145) holds when an iterate is close to feasible nonstationary points. To that end, we now have to be explicit about the stationarity notion to be used.

For problem (6.138), we understand stationarity in the usual way, as it was defined in Sect. 1.2.3: a point is stationary if it satisfies the Lagrange principle stated in Theorem 1.11. But we shall also need the following (scalar) characterization of this stationarity condition. For any $x,\, y \in \mathbf{R}^n$ define the generalized projected gradient (Cauchy) direction

$$d_C(x,\, y) = \pi_{D(x)}(y - f'(x)) - y, \tag{6.149}$$

where

$$D(x) = \{x + \xi \mid \xi \in \mathbf{R}^n,\ h(x) + h'(x)\xi = 0\},$$

and define further the function $S : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$,

$$S(x,\, y) = \begin{cases} -\left\langle f'(x),\, \dfrac{d_C(x,\, y)}{\|d_C(x,\, y)\|} \right\rangle & \text{if } d_C(x,\, y) \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{6.150}$$

Employing Theorem 1.10, it can be easily seen that stationarity of a point $\bar{x} \in \mathbf{R}^n$ which is feasible in (6.138) is equivalent to the condition $d_C(\bar{x},\, \bar{x}) = 0$.

Moreover, if $\bar{x}$ is feasible but nonstationary, then $S(\bar{x},\, \bar{x}) > 0$. If, in addition, the derivatives of f and h are continuous at such $\bar{x}$, and the regularity condition

$$\operatorname{rank} h'(\bar{x}) = l \tag{6.151}$$

holds, then it can be easily seen that S is continuous at $(\bar{x},\, \bar{x})$. In particular, $S(x,\, y) > 0$ for all $x,\, y \in \mathbf{R}^n$ close enough to such $\bar{x}$ (which is feasible, nonstationary, and satisfies (6.151)).

Recall that in composite-step SQP the feasibility phase aims at reducing the infeasibility measure (in the variant considered here—at the same time satisfying the linearized constraints). Then the optimality phase tries to reduce the quadratic model of the problem in the linearized feasible set. Our assumptions on "efficiency" of those two procedures are described next.

Define the infeasibility measure $\psi : \mathbf{R}^n \to \mathbf{R}_+$,

$$\psi(x) = \|h(x)\|,$$

and for a given $x^k \in \mathbf{R}^n$ the linearized feasible set

$$D_k = D(x^k) - x^k = \{p \in \mathbf{R}^n \mid h(x^k) + h'(x^k)p = 0\}.$$

The feasibility step p_n^k must satisfy the linearized constraints, be not too close to the trust-region boundary, and be reasonably scaled with respect to the constraints. Specifically, the requirements are as follows:

$$p_n^k \in D_k, \quad \|p_n^k\| \le \beta\delta_k, \tag{6.152}$$

for some $\beta \in (0,\, 1)$, where $\delta_k > 0$ is the trust-region radius; and close to nonstationary feasible points, the step p_n^k is of the order of constraints violation $\psi(x^k)$ (see (6.163) below). Note that if x^k is feasible ($\psi(x^k) = 0$), then the latter naturally implies that no feasibility action is required, i.e., $p_n^k = 0$. Assuming the linearized constraints are consistent to start with (and also remain consistent with the trust-region constraint added), a step with the needed properties can be computed in a number of ways. For example, by projecting x^k onto D_k or using some other feasibility algorithms.

If it is determined that a step satisfying (6.152) cannot be computed (for example, if $D_k = \emptyset$), then the filter feasibility restoration phase is activated, which produces a not-forbidden point improving the infeasibility, i.e., x^{k+1} for which the inequality $\psi(x^{k+1}) < \psi(x^k)$ holds with a sufficient margin (see Sect. 6.3.2).

If an adequate feasibility step p_n^k had been computed successfully, the method proceeds to the optimality phase of solving

$$\begin{array}{ll} \text{minimize} & \langle f'(x^k) + H_k p_n^k,\, p_t\rangle + \dfrac{1}{2}\langle H_k p_t,\, p_t\rangle \\ \text{subject to} & h'(x^k)p_t = 0,\ \|p_n^k + p_t\| \le \delta_k, \end{array} \tag{6.153}$$

to obtain $\tilde{p}_t^k = \tilde{p}_t^k(\delta_k)$, where $H_k \in \mathbf{R}^{n\times n}$ is a chosen symmetric matrix. In fact, this subproblem can be solved approximately, requiring only that the decrease of the quadratic model given by the optimality step, with respect to the point obtained after the feasibility step, be at least as good as that provided by the generalized Cauchy step.

Specifically, for the quadratic model

$$\Psi_k(p) = f(x^k) + \langle f'(x^k),\, p\rangle + \frac{1}{2}\langle H_k p,\, p\rangle, \tag{6.154}$$

the required condition is

$$\Psi_k(p_n^k) - \Psi_k(p_n^k + \tilde{p}_t^k) \geq \Psi_k(p_n^k) - \Psi_k(p_n^k + p_C^k), \tag{6.155}$$

where $p_C^k = p_C^k(\delta_k)$ is the generalized Cauchy step, defined next. If it holds that $d_C^k(x^k,\, x^k + p_n^k) = 0$ (recall the definition (6.149)) define $p_C^k = 0$; otherwise, define $p_C^k = \alpha_C^k \tilde{d}_C^k$, where

$$\tilde{d}_C^k = \frac{d_C^k(x^k,\, x^k + p_n^k)}{\|d_C^k(x^k,\, x^k + p_n^k)\|}, \tag{6.156}$$

and α_C^k is a solution of

$$\begin{array}{ll} \text{minimize} & \Psi_k(p_n^k + \alpha\tilde{d}_C^k) \\ \text{subject to} & \|p_n^k + \alpha\tilde{d}_C^k\| \leq \delta_k,\ \alpha \geq 0. \end{array} \tag{6.157}$$

The following scheme formalizes step computation by trust-region SQP just outlined above that, when implemented within the filter Algorithm 6.24, guarantees the conclusions of the convergence Theorem 6.27. In particular, Algorithm 6.29 below is a specific (trust-region based SQP) implementation of step 2 of the filter Algorithm 6.24.

Algorithm 6.29 Choose the local parameters $C_1,\, C_2 > 0$ with $C_1 < C_2$, $\delta \in [C_1,\, C_2]$, and $\beta,\, \theta,\, \rho_1,\, \rho_2 \in (0,\, 1)$.

Given (at step 2 of filter Algorithm 6.24) $x^k \in \mathbf{R}^n$, a symmetric matrix $H_k \in \mathbf{R}^{n\times n}$, the current filter $\tilde{F}_k$ and thus the current forbidden region $\tilde{\mathcal{F}}_k$, perform the following steps:

1. Compute a feasibility step $p_n^k = p_n^k(\delta)$ satisfying (6.152) for $\delta_k = \delta$. If this is not possible, activate an external feasibility restoration procedure to compute $x^{k+1} \notin \tilde{\mathcal{F}}_k$, and return to step 3 of Algorithm 6.24.
2. Compute the trial optimality step $\tilde{p}_t^k = \tilde{p}_t^k(\delta)$ by (approximately) solving (6.153) with $\delta_k = \delta$, such that $\tilde{p}_t^k$ is feasible in (6.153), and (6.155) holds. If
$$x^k + p_n^k + \tilde{p}_t^k \notin \tilde{\mathcal{F}}_k \tag{6.158}$$
and either
$$f(x^k) - f(x^k + p_n^k + \tilde{p}_t^k) \geq \rho_1(f(x^k) - \Psi_k(p_n^k + \tilde{p}_t^k)) \tag{6.159}$$

and/or

$$\rho_2(\psi(x^k))^2 > f(x^k) - \Psi_k(p_n^k + \tilde{p}_t^k), \tag{6.160}$$

then set $p_t^k = \tilde{p}_t^k$, $x^{k+1} = x^k + p_n^k + p_t^k$, and return to step 3 of Algorithm 6.24.

3. Replace δ by $\theta\delta_k$. If $\|p_n^k\| \leq \beta\delta$, go to step 2. Otherwise, activate an external feasibility restoration procedure to compute $x^{k+1} \notin \tilde{\mathcal{F}}_k$, and return to step 3 of Algorithm 6.24.

Note that in the above, it is not quite enough to pass the filter tests (i.e., (6.158)) for the candidate point to be accepted. When the predicted by the quadratic model decrease is significant relative to constraints violation ((6.160) does not hold; in a sense, x^k is "more feasible than optimal"), then the decrease of the objective function (6.159) would be required for the point to be accepted (reducing the objective function would be given "priority" with respect to improving further feasibility). This is natural—just consider a feasible x^k so that (6.160) cannot hold, in which case it is clear that f needs to be reduced as in (6.159) to make progress.

We now proceed to formal analysis. The first result concerns estimates for the infeasibility measures, if a new candidate point was produced (restoration was not activated).

Proposition 6.30. *Let $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice continuously differentiable on some convex set $\Omega \subset \mathbf{R}^n$, and let*

$$\|h'(x)\| \leq M_1, \quad \|h''(x)\| \leq M_1 \quad \forall\, x \in \Omega,$$

with some $M_1 > 0$.

If, for a given $x^k \in \Omega$, steps 1 and 2 of Algorithm 6.29 with some $\delta > 0$ compute $p_n^k = p_n^k(\delta)$ and $\tilde{p}_t^k = \tilde{p}_t^k(\delta)$ satisfying $x^k + p_n^k + \tilde{p}_t^k \in \Omega$, then

$$\psi(x^k) \leq M_1\delta, \quad \psi(x^k + p_n^k + \tilde{p}_t^k) \leq M_1\delta^2.$$

Proof. By (6.152),

$$\psi(x^k) = \|h(x^k)\| = \|h'(x^k)p_n^k\| \leq M_1\delta.$$

Similarly, taking into account that $h(x^k) + h'(x^k)p_n^k = 0$ and $h'(x^k)\tilde{p}_t^k = 0$, by the mean-value theorem (see Lemma A.11), it follows that

$$\psi(x^k + p_n^k + \tilde{p}_t^k) = \|h(x^k + p_n^k + \tilde{p}_t^k)\| \leq M_1\delta^2. \quad \square$$

We next estimate variation of the objective function after the feasibility step in terms of constraints violation.

Proposition 6.31. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be continuously differentiable on some convex set $\Omega \subset \mathbf{R}^n$, and let*

$$\|f'(x)\| \leq M_1 \quad \forall\, x \in \Omega. \tag{6.161}$$

Assume that

$$\|H_k\| \leq \Gamma \quad \forall\, k \tag{6.162}$$

with some $\Gamma > 0$.

Given a point $\bar{x} \in \Omega$ *which is feasible for problem* (6.138)*, for any* $\ell > 0$ *there exist a neighborhood* U *of* $\bar{x}$ *and* $M_2 > 0$ *such that if, for a given* $x^k \in U \cap \Omega$*, step 1 of Algorithm 6.29 computes* p_n^k *satisfying* $x^k + p_n^k \in \Omega$ *and*

$$\|p_n^k\| \leq \ell\psi(x^k), \tag{6.163}$$

then

$$|f(x^k) - f(x^k + p_n^k)| \leq M_2\psi(x^k), \tag{6.164}$$

$$|f(x^k) - \Psi_k(p_n^k)| \leq M_2\psi(x^k). \tag{6.165}$$

Proof. As $\psi(\bar{x}) = 0$ and ψ is continuous, there exists a neighborhood U of $\bar{x}$ such that whenever $x^k \in U$, it holds that

$$\frac{\Gamma}{2}(\ell\psi(x^k))^2 \leq M_1\ell\psi(x^k). \tag{6.166}$$

Using the mean-value theorem (Theorem A.10, (b)), property (6.161), and then (6.163), we obtain that

$$|f(x^k) - f(x^k + p_n^k)| \leq M_1\|p_n^k\| \leq M_1\ell\psi(x^k).$$

Similarly,

$$\begin{aligned} |f(x^k) - \Psi_k(p_n^k)| &= \left|\langle f'(x^k),\, p_n^k\rangle + \frac{1}{2}\langle H_k p_n^k,\, p_n^k\rangle\right| \\ &\leq M_1\|p_n^k\| + \frac{\Gamma}{2}\|p_n^k\|^2 \\ &\leq M_1\ell\psi(x^k) + \frac{\Gamma}{2}(\ell\psi(x^k))^2 \\ &\leq 2M_1\ell\psi(x^k), \end{aligned}$$

where the last inequality follows from (6.166).

We thus have the assertion defining $M_2 = 2M_1\ell$. □

The next step is to show that near a feasible nonstationary point, the reductions of the quadratic model and of the objective function are large. This is first done for the optimality step from the point obtained after the feasibility phase, and then for the full step.

Proposition 6.32. *Let* $f : \mathbf{R}^n \to \mathbf{R}$ *be twice continuously differentiable on some open convex set* $\Omega \subset \mathbf{R}^n$*, let* $h : \mathbf{R}^n \to \mathbf{R}^l$ *be continuously differentiable on* Ω*, and let*

$$\|f''(x)\| \leq M_1 \quad \forall\, x \in \Omega. \tag{6.167}$$

Assume that (6.162) holds with some $\Gamma > 0$. Let $\bar{x} \in \Omega$ be a feasible nonstationary point for problem (6.138), satisfying the regularity condition (6.151).

Then for any $\ell > 0$ there exist $\tilde{C} \in (0, C_1)$, $\varepsilon_2 > 0$ and a neighborhood $U \subset \Omega$ of $\bar{x}$ such that if, for a given $x^k \in U$, steps 1 and 2 of Algorithm 6.29 with some $\delta > 0$ compute $p_n^k = p_n^k(\delta)$ and $\tilde{p}_t^k = \tilde{p}_t^k(\delta)$ satisfying $x^k + p_n^k \in \Omega$, $x^k + p_n^k + \tilde{p}_t^k \in \Omega$, and (6.163), then

$$\Psi_k(p_n^k) - \Psi_k(p_n^k + \tilde{p}_t^k) \geq \varepsilon_2 \tilde{\delta} \tag{6.168}$$

holds for all $\tilde{\delta} \in (0, \min\{\delta, \tilde{C}\})$. Moreover, for any $\varepsilon_1 \in (0, 1)$, one can choose those $\tilde{C}$ and U in such a way that

$$f(x^k + p_n^k) - f(x^k + p_n^k + \tilde{p}_t^k) \geq \varepsilon_1(\Psi_k(p_n^k) - \Psi_k(p_n^k + \tilde{p}_t^k)) \tag{6.169}$$

holds provided $\delta \in (0, \tilde{C})$.

Proof. Consider any $\tilde{\delta} \in (0, \delta)$, and set $\alpha = (1 - \beta)\tilde{\delta}$. For $\tilde{d}_C^k$ given by (6.156), as

$$\|p_n^k + \alpha \tilde{d}_C^k\| \leq \|p_n^k\| + \alpha \leq \beta\delta + (1 - \beta)\tilde{\delta} \leq \delta,$$

this α is feasible in (6.157). Hence, by (6.154) and (6.155),

$$\begin{aligned} \Psi_k(p_n^k) - \Psi_k(p_n^k + \tilde{p}_t^k) &\geq \Psi_k(p_n^k) - \Psi_k(p_n^k + p_C^k) \\ &\geq \Psi_k(p_n^k) - \Psi_k(p_n^k + \alpha \tilde{d}_C^k) \\ &= \alpha \left(-\langle f'(x^k), \tilde{d}_C^k \rangle - \langle H_k p_n^k, \tilde{d}_C^k \rangle - \frac{\alpha}{2} \langle H_k \tilde{d}_C^k, \tilde{d}_C^k \rangle \right) \\ &\geq \alpha \left(S(x^k, x^k + p_n^k) - \Gamma \|p_n^k\| - \frac{\alpha \Gamma}{2} \right), \end{aligned} \tag{6.170}$$

where the definition (6.150) was employed for the last inequality.

As the point $\bar{x}$ is feasible nonstationary by the assumption, we have that $\psi(\bar{x}) = 0$ but $S(\bar{x}, \bar{x}) > 0$. Then using also (6.163) and continuity considerations, we conclude that there exist a neighborhood U of $\bar{x}$ and $\hat{\delta} \in (0, C_1)$ such that for any $x^k \in U$ and $\tilde{\delta} \in (0, \hat{\delta})$ it holds that

$$S(x^k, x^k + p_n^k) \geq \frac{1}{2} S(\bar{x}, \bar{x}), \quad \Gamma \|p_n^k\| + \frac{\alpha \Gamma}{2} \leq \frac{1}{4} S(\bar{x}, \bar{x}).$$

Combining this with (6.170) we conclude that

$$\Psi_k(p_n^k) - \Psi_k(p_n^k + \tilde{p}_t^k) \geq \frac{\alpha}{4} S(\bar{x}, \bar{x}) = \frac{(1 - \beta)\tilde{\delta}}{4} S(\bar{x}, \bar{x})$$

holds for $\tilde{\delta} \in (0, \min\{\delta, \hat{\delta}\}]$, which proves the first assertion for any $\tilde{C} \in (0, \hat{\delta}]$ if we define $\varepsilon_2 = (1 - \beta) S(\bar{x}, \bar{x})/4 > 0$.

Concerning the second assertion, note that using (6.154) and the mean-value theorem (Theorem A.10, (a)) we have that for some $\tau \in (0, 1)$ it holds that

$$\begin{aligned}
&|f(x^k + p_n^k) - f(x^k + p_n^k + \tilde{p}_t^k) - (\Psi_k(p_n^k) - \Psi_k(p_n^k + \tilde{p}_t^k))| \\
&= \left| \langle f'(x^k) - f'(x^k + p_n^k + \tau \tilde{p}_t^k),\, \tilde{p}_t^k \rangle + \langle H_k p_n^k,\, \tilde{p}_t^k \rangle + \frac{1}{2} \langle H_k \tilde{p}_t^k,\, \tilde{p}_t^k \rangle \right| \\
&\leq M_1 \|p_n^k + \tau \tilde{p}_t^k\| \|\tilde{p}_t^k\| + \Gamma \|p_n^k\| \|\tilde{p}_t^k\| + \frac{\Gamma}{2} \|\tilde{p}_t^k\|^2 \\
&\leq M_1 \delta \|p_n^k\| + M_1 \delta^2 + \Gamma \delta \|p_n^k\| + \frac{\Gamma}{2} \delta^2 \\
&= (M_1 + \Gamma) \delta \|p_n^k\| + \left(M_1 + \frac{\Gamma}{2} \right) \delta^2.
\end{aligned}$$

Restricting further U if necessary, we can take $\tilde{C} \in (0, \hat{\delta}]$ such that for any $x^k \in U$ and $\delta \in (0, \tilde{C})$, it holds that

$$\frac{(M_1 + \Gamma)\|p_n^k\|}{\varepsilon_2} \leq \frac{1 - \varepsilon_1}{2}, \quad \frac{(M_1 + \Gamma/2)\delta}{\varepsilon_2} \leq \frac{1 - \varepsilon_1}{2}.$$

Then using (6.168) with $\tilde{\delta} = \delta$, we obtain

$$\begin{aligned}
&\left| \frac{f(x^k + p_n^k) - f(x^k + p_n^k + \tilde{p}_t^k)}{\Psi_k(p_n^k) - \Psi_k(p_n^k + \tilde{p}_t^k)} - 1 \right| \\
&= \left| \frac{f(x^k + p_n^k) - f(x^k + p_n^k + \tilde{p}_t^k) - (\Psi_k(p_n^k) - \Psi_k(p_n^k + \tilde{p}_t^k))}{\Psi_k(p_n^k) - \Psi_k(p_n^k + \tilde{p}_t^k)} \right| \\
&\leq \frac{(M_1 + \Gamma)\|p_n^k\|\delta + (M_1 + \Gamma/2)\delta^2}{\varepsilon_2 \delta} \\
&\leq 1 - \varepsilon_1,
\end{aligned}$$

which verifies (6.169). $\square$

Next, the reduction estimates are extended to the full step.

Proposition 6.33. *Under the joint hypotheses of Propositions 6.31 and 6.32, for any $\ell > 0$ there exist $\tilde{C} \in (0, C_1)$, $\varepsilon_2 > 0$ and a neighborhood $U \subset \Omega$ of $\bar{x}$ such that if, for a given $x^k \in U$, steps 1 and 2 of Algorithm 6.29 with some $\delta \in [\theta^2 \tilde{C}, \tilde{C}]$ compute $p_n^k = p_n^k(\delta)$ and $\tilde{p}_t^k = \tilde{p}_t^k(\delta)$ satisfying $x^k + p_n^k \in \Omega$, $x^k + p_n^k + \tilde{p}_t^k \in \Omega$, and (6.163), then it holds that*

$$f(x^k) - \Psi_k(p_n^k + \tilde{p}_t^k) \geq \frac{1}{2} \varepsilon_2 \delta, \tag{6.171}$$

and

$$f(x^k) - f(x^k + p_n^k + \tilde{p}_t^k) \geq \rho_1(f(x^k) - \Psi_k(p_n^k + \tilde{p}_t^k)). \tag{6.172}$$

Proof. Take any $\varepsilon_1 \in (\rho_1, 1)$ and define $\tau = (\varepsilon_1 - \rho_1)/(\varepsilon_1 + \rho_1)$. Let further $\tilde{C} \in (0, C_1)$, $\varepsilon_2 > 0$, $M_2 > 0$, and let a neighborhood U of $\bar{x}$ be defined according to Propositions 6.31 and 6.32.

By further shrinking U if necessary, we can assume that

$$M_2\psi(x) \leq \min\left\{\frac{1}{2}\varepsilon_2\theta^2\tilde{C},\, \tau\varepsilon_1\varepsilon_2\theta^2\tilde{C}\right\} \quad \forall\, x \in U. \tag{6.173}$$

Then, by (6.165),

$$|f(x^k) - \Psi_k(p_n^k)| \leq M_2\psi(x^k) \leq \frac{1}{2}\varepsilon_2\theta^2\tilde{C} \leq \frac{1}{2}\varepsilon_2\delta,$$

where the inclusion $\delta \in [\theta^2\tilde{C},\, \tilde{C}]$ was taken into account. Using this relation and (6.168) with $\tilde{\delta} = \delta$, we obtain

$$\begin{aligned} f(x^k) - \Psi_k(p_n^k + \tilde{p}_t^k) &= f(x^k) - \Psi_k(p_n^k) + \Psi_k(p_n^k) - \Psi_k(p_n^k + \tilde{p}_t^k) \\ &\geq \frac{1}{2}\varepsilon_2\delta, \end{aligned}$$

which is (6.171).

Next, using (6.164), (6.165), (6.168) with $\tilde{\delta} = \delta$, (6.169), and (6.173), we have that

$$\begin{aligned} |f(x^k) - f(x^k + p_n^k)| &\leq M_2\psi(x^k) \\ &\leq \tau\varepsilon_1\varepsilon_2\theta^2\tilde{C} \\ &\leq \tau\varepsilon_1\varepsilon_2\delta \\ &\leq \tau(f(x^k + p_n^k) - f(x^k + p_n^k + \tilde{p}_t^k)), \end{aligned}$$

and

$$\begin{aligned} f(x^k) - \Psi_k(p_n^k) &\leq M_2\psi(x^k) \\ &\leq \tau\varepsilon_2\theta^2\tilde{C} \\ &\leq \tau\varepsilon_2\delta \\ &\leq \tau(\Psi_k(p_n^k) - \Psi_k(p_n^k + \tilde{p}_t^k)). \end{aligned}$$

Hence,

$$\begin{aligned} f(x^k) - f(x^k + p_n^k + \tilde{p}_t^k) &= f(x^k) - f(x^k + p_n^k) \\ &\quad + f(x^k + p_n^k) - f(x^k + p_n^k + \tilde{p}_t^k) \\ &\geq (1 - \tau)(f(x^k + p_n^k) - f(x^k + p_n^k + \tilde{p}_t^k)), \end{aligned}$$

and

$$\begin{aligned} f(x^k) - \Psi_k(p_n^k + \tilde{p}_t^k) &= f(x^k) - \Psi_k(p_n^k) + \Psi_k(p_n^k) - \Psi_k(p_n^k + \tilde{p}_t^k) \\ &\leq (1+\tau)(\Psi_k(p_n^k) - \Psi_k(p_n^k + \tilde{p}_t^k)). \end{aligned}$$

Combining the last two inequalities above and (6.169), we obtain that

$$\begin{aligned} f(x^k) - f(x^k + p_n^k + \tilde{p}_t^k) &\geq (1-\tau)\varepsilon_1(\Psi_k(p_n^k) - \Psi_k(p_n^k + \tilde{p}_t^k)) \\ &\geq \frac{1-\tau}{1+\tau}\varepsilon_1(f(x^k) - \Psi_k(p_n^k + \tilde{p}_t^k)) \\ &= \rho_1(f(x^k) - \Psi_k(p_n^k + \tilde{p}_t^k)), \end{aligned}$$

as claimed. □

We proceed to examine the role of γ_k defined by (6.144), which appears in the right-hand side in the key descent condition (6.145) of the general filter scheme (and of the related $\tilde{\gamma}_k$ defined by (6.148)) in the present SQP context. Recall also Proposition 6.28.

We first exhibit that near a feasible nonstationary point, if a trial point is rejected (and thus the trust-region parameter is reduced), then this is caused by the increase in the infeasibility by this trial point measured by $\tilde{\gamma}_k$.

Proposition 6.34. *Under the joint hypotheses of Propositions 6.31 and 6.32, for any $\ell > 0$ there exist $\tilde{C} \in (0, C_1)$ and a neighborhood $U \subset \Omega$ of $\bar{x}$ such that if, for a given $x^k \in U$, steps 1 and 2 of Algorithm 6.29 with some $\delta \in [\theta^2\tilde{C}, \tilde{C}]$ compute $p_n^k = p_n^k(\delta)$ and $\tilde{p}_t^k = \tilde{p}_t^k(\delta)$ satisfying $x^k + p_n^k \in \Omega$, $x^k + p_n^k + \tilde{p}_t^k \in \Omega$, and (6.163), but the point $x^k + p_n^k + \tilde{p}_t^k$ is rejected by the rules (6.158)–(6.160), then it holds that*

$$\psi(x^k + p_n^k + \tilde{p}_t^k) \geq \tilde{\gamma}_k, \tag{6.174}$$

where $\tilde{\gamma}_k$ is defined in (6.148).

Proof. Define $\tilde{C} \in (0, C_1)$, $\varepsilon_2 > 0$, $M_2 > 0$ and a neighborhood U of $\bar{x}$ according to Propositions 6.31–6.33.

By further shrinking U if necessary, we can assume that

$$M_2\psi(x) \leq \frac{1}{2}\varepsilon_2\theta^2\tilde{C}, \quad \sigma_1\psi(x) < \frac{1}{2}\rho_1\varepsilon_2\theta^2\tilde{C} \quad \forall\, x \in U. \tag{6.175}$$

Then by (6.165) we have that

$$|f(x^k) - \Psi_k(p_n^k)| \leq M_2\psi(x^k) \leq \frac{1}{2}\varepsilon_2\theta^2\tilde{C} \leq \frac{1}{2}\varepsilon_2\delta,$$

where the inclusion $\delta \in [\theta^2\tilde{C}, \tilde{C}]$ was taken into account. Using then (6.168) with $\tilde{\delta} = \delta$,

$$f(x^k) - \Psi_k(p_n^k + \tilde{p}_t^k) \geq \frac{1}{2}\varepsilon_2\delta \geq \frac{1}{2}\varepsilon_2\theta^2\tilde{C}. \tag{6.176}$$

Using further (6.172) and (6.175), we obtain that

$$\begin{aligned} f(x^k) - f(x^k + p_n^k + \tilde{p}_t^k) &\geq \rho_1(f(x^k) - \Psi_k(p_n^k + \tilde{p}_t^k)) \\ &\geq \frac{1}{2}\rho_1\varepsilon_2\theta^2\tilde{C} \\ &> \sigma_1\psi(x). \end{aligned} \tag{6.177}$$

The first inequality of the last chain of relations shows that (6.159) holds. Hence, since the point $x^k + p_n^k + \tilde{p}_t^k$ was rejected, it must be in the forbidden region, i.e., (6.158) is not true. In view of (6.177), Proposition 6.28, (b), implies the claim that (6.174) holds. □

We are now in position to prove the key assumption of Theorem 6.27 on convergence of the general filter scheme, i.e., the property (6.145), when the trial points are computed by composite-step SQP.

Theorem 6.35. *In addition to the joint hypotheses of Propositions 6.30–6.32, assume that $B(\bar{x}, 2\delta) \subset \Omega$, where $\delta > 0$ is the starting value of the trust-region radius in Algorithm 6.29.*

Then there exist $\ell > 0$, a neighborhood U of $\bar{x}$, and $\Delta > 0$, such that

$$U + B(\bar{x}, \delta) \subset \Omega, \tag{6.178}$$

and if $x^k \in U$, then Algorithm 6.29 with the additional requirement (6.163) on a feasibility step p_n^k, successfully generates the next iterate x^{k+1} satisfying (6.145).

(By "successfully generates," we mean here that there exist p_n^k with the needed properties, and once any appropriate p_n^k is computed, Algorithm 6.29 generates some x^{k+1} in a finite number of steps, and this x^{k+1} satisfies (6.145)).

Proof. By the regularity condition (6.151), there exist $\ell > 0$ and a neighborhood V of $\bar{x}$ such that

$$\operatorname{dist}(x, D(x)) \leq \ell\psi(x) \quad \forall\, x \in V. \tag{6.179}$$

This follows from Lemma A.5.

With this ℓ and V at hand, define $\tilde{C} \in (0, C_1)$, $\varepsilon_2 > 0$, $M_2 > 0$ and a neighborhood $U \subset V$ according to Propositions 6.31–6.34. Since $B(\bar{x}, 2\delta) \subset \Omega$, we can assume that (6.178) holds. Furthermore, we can decrease the constant $\tilde{C}$, if necessary, to assume that

$$\tilde{C} < \frac{\theta^2}{M_1} \min\left\{\frac{\beta}{\ell},\, \frac{\varepsilon_2}{2M_2},\, \frac{\varepsilon_2}{2\rho_2},\, \frac{\rho_1\varepsilon_2}{2\sigma_1}\right\}. \tag{6.180}$$

Moreover, restricting further U if necessary, it holds that

$$\ell\psi(x) \leq \beta\theta^2\tilde{C}, \quad \rho_2(\psi(x))^2 \leq \frac{1}{2}\varepsilon_2\theta^2\tilde{C}, \quad \psi(x) \leq 1 \quad \forall\, x \in U. \tag{6.181}$$

From (6.179) we conclude that for any $x^k \in U$ there exists $p_n^k \in D_k$ satisfying (6.163), and the only way the requirements on the feasibility step may not hold is the violation of the trust-region constraint in (6.152). In particular, it is then the only reason for the feasibility restoration to be possibly activated at such an iterate $x^k \in U$.

Algorithm 6.29 starts with some trust-region radius $\delta \geq C_1$ and always terminates with some $\delta_k > 0$ and x^{k+1}. This is because either the restoration procedure is activated in step 1, or otherwise (i.e., if the feasibility step p_n^k is computed) it would be activated after a finite number of trust-region reductions if no trial step is accepted before.

Observe also that according to (6.152), the constraints in (6.153), and inclusion (6.178), it holds that $x^k + p_n^k \in \Omega$, and if the algorithm generates $\tilde{p}_t^k$, then $x^k + p_n^k + \tilde{p}_t^k \in \Omega$.

We shall consider the two possible cases:

$$\delta_k \geq \theta^2\tilde{C} \quad \text{or} \quad \delta_k < \theta^2\tilde{C}. \tag{6.182}$$

In the first case, by (6.163) and the first inequality in (6.181),

$$\|p_n^k\| \leq \ell\psi(x^k) \leq \beta\theta^2\tilde{C} \leq \beta\delta_k,$$

which means the restoration procedure has not been activated and thus the iterate $x^{k+1} = x^k + p_n^k + p_t^k$ has been computed. Then using (6.168) with $\tilde{\delta} = \theta^2\tilde{C}$, we have that

$$\Psi_k(p_n^k) - \Psi_k(p_n^k + p_t^k) \geq \varepsilon_2\theta^2\tilde{C}. \tag{6.183}$$

Now using (6.165), the second inequality in (6.181), and (6.183), we have that

$$f(x^k) - \Psi_k(p_n^k + p_t^k) \geq \frac{1}{2}\varepsilon_2\theta^2\tilde{C} \geq \rho_2(\psi(x^k))^2. \tag{6.184}$$

In particular, (6.160) does not hold. Since the trial point in question has been accepted as x^{k+1}, it follows that the condition (6.159) must have been valid. Using (6.159), the first inequality in (6.184), and also that $\tilde{\gamma}_k \leq 1$, we conclude that

$$\begin{aligned} f(x^k) - f(x^{k+1}) &\geq \rho_1(f(x^k) - \Psi_k(p_n^k + p_t^k)) \\ &\geq \frac{1}{2}\rho_1\varepsilon_2\theta^2\tilde{C} \\ &\geq \frac{1}{2}\rho_1\varepsilon_2\theta^2\tilde{C}\sqrt{\tilde{\gamma}_k}. \end{aligned} \tag{6.185}$$

We now turn our attention to the second case in (6.182). Note that in that case, the trust-region parameter had been reduced at least once, and thus trial optimality steps have been computed by the algorithm. For the second case in (6.182), we further consider the following two possibilities. The first is when $\psi(x^k + p_n^k + \tilde{p}_t^k(\delta)) \geq \tilde{\gamma}_k$ for all $\delta \leq \theta \tilde{C}$ for which the trial step $\tilde{p}_t^k(\delta)$ has been computed. The second is when there exists $\delta \in (0,\, \theta \tilde{C}]$ for which the trial step has been computed such that $\psi(x^k + p_n^k + \tilde{p}_t^k(\delta)) < \tilde{\gamma}_k$.

In the first possibility, since $\delta_k < C_1$ (second case in (6.182)), the trial step $\tilde{p}_t^k = \tilde{p}_t^k(\delta)$ with $\delta = \delta_k/\theta$ has been computed and rejected by step 2 of Algorithm 6.29 (this also subsumes that p_n^k was computed by step 1 of Algorithm 6.29 with this δ). As $\delta_k/\theta < \theta \tilde{C}$, in the case under consideration $\psi(x^k + p_n^k + \tilde{p}_t^k) \geq \tilde{\gamma}_k$. Using Proposition 6.30, we then obtain that

$$\begin{aligned} M_1 \delta_k^2 &= \theta^2 M_1 (\delta_k/\theta)^2 \\ &\geq \theta^2 \psi(x^k + p_n^k + \tilde{p}_t^k) \\ &\geq \theta^2 \tilde{\gamma}_k \\ &\geq \theta^2 \psi(x^k), \end{aligned} \tag{6.186}$$

where the last inequality is by Proposition 6.28, (c). Using (6.163), (6.186), (6.180), and the second condition in (6.182), we have that

$$\begin{aligned} \|p_n^k\| \leq \ell \psi(x^k) &\leq \frac{M_1 \ell}{\theta^2} \delta_k^2 \\ &< \beta \frac{\delta_k^2}{\tilde{C}} < \beta \theta^2 \delta_k < \beta \delta_k \end{aligned}$$

(recall that $\theta \in (0,\, 1)$). This again means that the restoration procedure has not been activated at this iteration. Using (6.168) with $\tilde{\delta} = \delta_k$, we have that

$$\Psi_k(p_n^k) - \Psi_k(p_n^k + p_t^k) \geq \varepsilon_2 \delta_k. \tag{6.187}$$

Since (6.165) is still true, using again (6.186), (6.180), and the second condition in (6.182), we derive

$$\begin{aligned} |f(x^k) - \Psi_k(p_n^k)| \leq M_2 \psi(x^k) &\leq \frac{M_1 M_2}{\theta^2} \delta_k^2 \\ &\leq \frac{1}{2} \varepsilon_2 \frac{\delta_k^2}{\tilde{C}} < \frac{1}{2} \varepsilon_2 \theta^2 \delta_k < \frac{1}{2} \varepsilon_2 \delta_k. \end{aligned} \tag{6.188}$$

Now, combining (6.187) and (6.188), and using (6.180), the second condition in (6.182), the third inequality in (6.181), and (6.186), it follows that

$$
\begin{aligned}
f(x^k) - \Psi_k(p_n^k + p_t^k) &\geq \frac{1}{2}\varepsilon_2\delta_k \\
&\geq \rho_2 \frac{M_1\tilde{C}}{\theta^2}\delta_k \\
&\geq \rho_2 \frac{M_1}{\theta^4}\delta_k^2 \\
&\geq \rho_2 \frac{1}{\theta^2}\psi(x^k) \\
&\geq \rho_2 (\psi(x^k))^2. \qquad (6.189)
\end{aligned}
$$

In particular, (6.160) does not hold and thus for the trial point to have been accepted as x^{k+1}, the condition (6.159) must have been valid. Using (6.159), the next-to-the-last inequality in (6.186), and the first inequality in (6.189), we then conclude that

$$
\begin{aligned}
f(x^k) - f(x^{k+1}) &\geq \rho_1(f(x^k) - \Psi_k(p_n^k + p_t^k)) \\
&\geq \frac{1}{2}\rho_1\varepsilon_2\delta_k \\
&\geq \frac{1}{2}\rho_1\varepsilon_2 \frac{\theta}{\sqrt{M_1}}\sqrt{\tilde{\gamma}_k}. \qquad (6.190)
\end{aligned}
$$

We next consider the second possibility in the second case of (6.182), i.e., for some $\delta \leq \theta\tilde{C}$ a trial step $\tilde{p}_t^k = \tilde{p}_t^k(\delta)$ had been computed for which $\psi(x^k + p_n^k + \tilde{p}_t^k) < \tilde{\gamma}_k$. Let $\bar{\delta}$ be the first value of δ for which this occurred, and let $\bar{p}_t^k = \tilde{p}_t^k(\bar{\delta})$ be the corresponding trial step. We shall show that $\delta_k = \bar{\delta}$, i.e., that step was in fact accepted.

We first establish that for the preceding trial step $\hat{p}_t^k = \tilde{p}_t^k(\bar{\delta}/\theta)$ it holds that

$$\psi(x^k + p_n^k + \hat{p}_t^k) \geq \tilde{\gamma}_k. \qquad (6.191)$$

Indeed, if $\bar{\delta}/\theta \leq \theta\tilde{C}$, then the definition of $\bar{\delta}$ implies (6.191). If $\bar{\delta}/\theta > \theta\tilde{C}$, then $\bar{\delta}/\theta \in [\theta^2\tilde{C},\, \tilde{C}]$, and (6.191) follows from (6.174), since the trial step $\hat{p}_t^k$ was rejected.

Similarly to (6.186), from (6.191) we obtain that

$$M_1\bar{\delta}^2 \geq \theta^2\tilde{\gamma}_k \geq \theta^2\psi(x^k), \qquad (6.192)$$

and then, similarly to the first inequality in (6.189),

$$f(x^k) - \Psi_k(p_n^k + \bar{p}_t^k) \geq \frac{1}{2}\varepsilon_2\bar{\delta}.$$

Then, by (6.172),

$$
\begin{aligned}
f(x^k) - f(x^k + p_n^k + \bar{p}_t^k) &\geq \rho_1(f(x^k) - \Psi_k(p_n^k + \bar{p}_t^k)) \\
&\geq \frac{1}{2}\rho_1\varepsilon_2\bar{\delta}. \qquad (6.193)
\end{aligned}
$$

Using further (6.180) and (6.192),

$$\begin{aligned} f(x^k) - f(x^k + p_n^k + \bar{p}_t^k) &> \sigma_1 \frac{M_1 \tilde{C}}{\theta^2} \bar{\delta} \\ &> \sigma_1 \frac{M_1}{\theta^2} \bar{\delta}^2 \\ &\geq \sigma_1 \psi(x^k), \end{aligned}$$

where the next-to-last inequality is by $\bar{\delta} < \tilde{C}$. By Proposition 6.28, (b), the latter inequality and the definition of $\bar{\delta}$ (which says in particular that $\psi(x^k + p_n^k + \bar{p}_t^k) < \tilde{\gamma}_k$) imply that $x^k + p_n^k + \bar{p}_t^k \notin \tilde{\mathcal{F}}_k$, i.e., the point $x^k + p_n^k + \bar{p}_t^k$ is accepted by the filter. As the first inequality in (6.193) shows that the test (6.159) holds for $\tilde{p}_t^k = \bar{p}_t^k$ as well, it follows that $x^{k+1} = x^k + p_n^k + \bar{p}_t^k$ and $\delta_k = \bar{\delta}$.

Now, the first inequality in (6.192) and the second inequality in (6.193) yield

$$\begin{aligned} f(x^k) - f(x^{k+1}) &\geq \frac{1}{2} \rho_1 \varepsilon_2 \bar{\delta} \\ &\geq \frac{1}{2} \rho_1 \varepsilon_2 \frac{\theta}{2\sqrt{M_1}} \sqrt{\tilde{\gamma}_k}. \end{aligned}$$

Combining this with the outcomes for the other cases, namely, (6.185) and (6.190), and defining

$$\varDelta = \frac{1}{2} \rho_1 \varepsilon_2 \theta \min \left\{ \theta \tilde{C},\, \frac{1}{\sqrt{M_1}} \right\},$$

we obtain

$$f(x^k) - f(x^{k+1}) \geq \varDelta \sqrt{\tilde{\gamma}_k}.$$

The final conclusion, i.e., (6.145), follows from Proposition 6.28, (a). □

Chapter 7
Degenerate Problems with Nonisolated Solutions

The common feature of Newtonian methods discussed in the previous chapters is their local superlinear convergence under certain reasonable assumptions, which, however, always subsumed that the solution in question is isolated (locally unique). On the other hand, sometimes one has to deal with problems (or even problem classes) whose solutions are not isolated, and thus are degenerate in some sense. One important example of degeneracy, to which we shall pay special attention below, is the case of nonunique Lagrange multipliers associated with a stationary point of an optimization problem or with a solution of a variational problem. In that case, the corresponding primal-dual solution of the associated KKT system cannot be isolated, which means that the KKT system is degenerate. When the latter happens, we shall refer to the underlying optimization or variational problem also as degenerate. Nonuniqueness of Lagrange multipliers means that the SMFCQ (and even more so the LICQ) is violated. It is worth to emphasize that there exist important problem classes which necessarily violate not only the SMFCQ but even the usual MFCQ. One example is given by the so-called mathematical programs with complementarity constraints, discussed in Sect. 7.3.

In this chapter we consider Newton-type methods for degenerate problems, i.e., problems for which solutions are known to be nonisolated or where there are good reasons to expect that this is the case. We first discuss the consequences of the nonuniqueness of Lagrange multipliers. In particular, we put in evidence that Newtonian methods for constrained optimization have a strong tendency to generate dual sequences that are attracted to certain special Lagrange multipliers, called critical, which violate the second-order sufficient optimality conditions. We further show that this phenomenon is in fact the reason for slow convergence usually observed in the degenerate case, as convergence to noncritical multipliers would have yielded the superlinear primal rate despite degeneracy. Convergence to critical multipliers appears though to be the typical numerical behavior when such multipliers do exist. We then consider special modifications of Newtonian methods, intended to locally suppress the effect of attraction to critical multipliers, and thus ensure the

A.F. Izmailov and M.V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-3-319-04247-3_7,

superlinear convergence rate despite degeneracy. In particular, we develop the stabilized Josephy–Newton method for generalized equations, and the associated stabilized Newton method for variational problems and stabilized sequential quadratic programming (SQP) for optimization. The appealing feature is that for superlinear convergence these methods require the second-order sufficiency conditions only (or even the weaker noncriticality of the Lagrange multiplier if there are no inequality constraints), and in particular do not need constraint qualifications of any kind.

Finally, we consider the mathematical programs with complementarity constraints. These problems are inherently degenerate, but their degeneracy is structured. This structure can and should be employed for constructing special Newton-type methods. We also consider the related class of mathematical programs with vanishing constraints.

7.1 Attraction to Critical Lagrange Multipliers

In this section, we discuss possible scenarios for behavior of the dual part of sequences generated by primal-dual Newton-type methods when applied to optimization problems with nonunique Lagrange multipliers associated with a solution. It turns out that critical Lagrange multipliers introduced in Definition 1.41 play a crucial role. Specifically, convergence to a critical multiplier appears to be the typical scenario when critical multipliers do exist, and when the dual sequence converges to some point. Moreover, together with the possible absence of dual convergence, attraction to critical multipliers is precisely the reason for slow primal convergence that is usually observed for problems with degenerate constraints. If the dual sequence were to converge to a noncritical limit, the primal rate of convergence would have been superlinear. It is interesting to point out that this negative effect of attraction to critical multipliers was first discovered experimentally; it was first reported in [128], and further studied in [144, 146, 149]. By now there exists compelling numerical evidence demonstrating that this is indeed a persistent phenomenon, the claim also supported by some theoretical considerations discussed below.

Before proceeding, we stress that many theoretical questions concerned with critical multipliers are still open at this time. What is currently available is some characterization of what *would have happened* in the case of convergence of Newton-type methods to a noncritical multiplier, with explanations of why this scenario is unlikely (at least for some specific cases that nevertheless do model the general situations). When these theoretical explanations are put together with the numerical evidence showing that this behavior is indeed unlikely, it can be considered enough for "practical" purposes: there is no doubt that the effect of attraction exists, and thus it should be taken care of when we deal with Newton-type methods for optimization problems with potentially nonunique Lagrange multipliers. However, a completely satisfactory theoretical result should probably be of positive nature, demonstrating that *the set of critical multipliers is indeed an attractor* in some sense. Whether this can be proven or not for more-or-less general problem settings remains a very challenging open question. For the special case of problems with a quadratic objective function and quadratic equality constraints, the result of this kind is given in [153]. However, the analysis in [153] is too involved and technical to be included in this book.

It is convenient to start with the case when there are equality constraints only. Afterwards the discussion is extended to the more involved case of equality and inequality constraints.

7.1.1 Equality-Constrained Problems

Consider the equality-constrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \end{array} \tag{7.1}$$

where the objective function $f : \mathbf{R}^n \to \mathbf{R}$ and the constraints mapping $h : \mathbf{R}^n \to \mathbf{R}^l$ are twice differentiable.

Recall that stationary points and associated Lagrange multipliers of problem (7.1) are characterized by the Lagrange optimality system

$$\frac{\partial L}{\partial x}(x,\, \lambda) = 0, \quad h(x) = 0, \tag{7.2}$$

where $L : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}$ is the Lagrangian of problem (7.1):

$$L(x,\, \lambda) = f(x) + \langle \lambda,\, h(x) \rangle.$$

Let $\bar{x} \in \mathbf{R}^n$ be a stationary point of problem (7.1), and let $\mathcal{M}(\bar{x})$ be the (nonempty) set of Lagrange multipliers associated with $\bar{x}$, that is, the set of $\lambda \in \mathbf{R}^l$ satisfying (7.2) for $x = \bar{x}$. As observed in Sect. 1.2.3, $\mathcal{M}(\bar{x})$ is a singleton if and only if the classical regularity condition holds:

$$\operatorname{rank} h'(\bar{x}) = l. \tag{7.3}$$

The case of interest in this section, however, is when $\mathcal{M}(\bar{x})$ is not a singleton, or equivalently, when (7.3) does not hold.

We start our discussion with the basic Newton–Lagrange method stated in Algorithm 4.1, which is simply the Newton method applied to the system (7.2). For the current primal-dual iterate $(x^k,\, \lambda^k) \in \mathbf{R}^n \times \mathbf{R}^l$, this method generates the next iterate $(x^{k+1},\, \lambda^{k+1})$ as a solution of the linear system

$$\begin{aligned} \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)(x - x^k) + (h'(x^k))^{\mathrm{T}}(\lambda - \lambda^k) &= -\frac{\partial L}{\partial x}(x^k,\, \lambda^k), \\ h'(x^k)(x - x^k) &= -h(x^k), \end{aligned} \tag{7.4}$$

with respect to $(x,\, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$.

According to Theorem 1.13, if there exists a multiplier $\bar{\lambda} \in \mathcal{M}(\bar{x})$ such that the SOSC

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in \ker h'(\bar{x}) \setminus \{0\} \tag{7.5}$$

holds, then $\bar{x}$ is a strict local minimizer of problem (7.1). Then primal convergence to $\bar{x}$ can be expected for (good implementations of) good algorithms, even in degenerate cases. But the convergence is often slow, and it has been observed (e.g., in [271, Sec. 6], and in [146]) that when primal convergence is slow, the reason for this is not so much degeneracy as such, but some undesirable behavior of the dual sequence. Among various scenarios of this behavior, one of the prominent ones appears to be convergence to multipliers violating the SOSC (7.5), or more precisely, to critical Lagrange multipliers. According to Definition 1.41, in the case of the equality-constrained problem (7.1), a multiplier $\bar{\lambda} \in \mathcal{M}(\bar{x})$ is critical if there exists $\xi \in \ker h'(\bar{x}) \setminus \{0\}$ such that

$$\frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})\xi \in \mathrm{im}(h'(\bar{x}))^{\mathrm{T}}.$$

As discussed in Sect. 1.3.3, in the fully degenerate case (i.e., the case of $h'(\bar{x}) = 0$), criticality of $\bar{\lambda}$ means that the Hessian of the Lagrangian $\frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})$ is a singular matrix, while in the general case, criticality amounts to singularity of the reduced Hessian.

We start with the following simple example taken from [128, 146].

Example 7.1. Let $n = l = 1$, $f(x) = x^2$, $h(x) = x^2$. Problem (7.1) with this data has the unique feasible point (hence, the unique solution) $\bar{x} = 0$, with $\mathcal{M}(0) = \mathbf{R}$ (because $h'(0) = 0$). This solution satisfies the SOSC (7.5) for all $\bar{\lambda} > -1$, and the *unique* critical multiplier is $\bar{\lambda} = -1$.

According to (7.4), any primal-dual sequence $\{(x^k,\, \lambda^k)\}$ of the Newton–Lagrange method for this problem satisfies the equalities

$$(1+\lambda^k)(x^{k+1}-x^k)+x^k(\lambda^{k+1}-\lambda^k) = -(1+\lambda^k)x^k, \quad 2x^k(x^{k+1}-x^k) = -(x^k)^2$$

for each k. Suppose that $x^k \neq 0$. Then from the second equality we obtain that

$$x^{k+1} = \frac{1}{2}x^k,$$

and then using the first equality, it follows that

$$\lambda^{k+1} = \frac{1}{2}(\lambda^k - 1).$$

In particular,

$$\lambda^{k+1} + 1 = \frac{1}{2}(\lambda^k + 1).$$

Therefore, if $x^0 \neq 0$, then $x^k \neq 0$ for all k, and the sequence $\{(x^k,\, \lambda^k)\}$ is well defined and converges linearly to $(0,\, -1)$. In particular, $\{\lambda^k\}$ converges to the unique critical multiplier $\bar{\lambda} = -1$. Moreover, both sequences $\{x^k\}$ and $\{\lambda^k\}$ converge linearly.

The behavior observed in this example is fully explained by the following theoretical result, valid for special one-dimensional problems with a single constraint. As discussed previously, at this time results of this nature are not available for general settings. In this sense, it is worth to note that even the proof for the simple special case considered in Proposition 7.2 can hardly be called easy.

Proposition 7.2. *Let $f : \mathbf{R} \to \mathbf{R}$ and $h : \mathbf{R} \to \mathbf{R}$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}$, with their second derivatives being continuous at $\bar{x}$. Assume that $f'(\bar{x}) = h(\bar{x}) = h'(\bar{x}) = 0$, $h''(\bar{x}) \neq 0$.*

Then for any $x^0 \in \mathbf{R} \setminus \{\bar{x}\}$ close enough to $\bar{x}$, and any $\lambda^0 \in \mathbf{R}$, there exists the unique sequence $\{(x^k,\, \lambda^k)\} \subset \mathbf{R} \times \mathbf{R}$ such that $(x^{k+1},\, \lambda^{k+1})$ satisfies (7.4) for all k; this sequence converges to $(\bar{x},\, \bar{\lambda})$, where $\bar{\lambda} = -f''(\bar{x})/h''(\bar{x})$, $x^k \neq \bar{x}$ for all k, and

$$\lim_{k \to \infty} \frac{x^{k+1} - \bar{x}}{x^k - \bar{x}} = \frac{1}{2}. \tag{7.6}$$

Note that the assumption $f'(\bar{x}) = h(\bar{x}) = h'(\bar{x}) = 0$ implies that $\bar{x}$ is a stationary point of problem (7.1), with $\mathcal{M}(\bar{x}) = \mathbf{R}$, while $h''(\bar{x}) \neq 0$ implies that $\bar{\lambda} = -f''(\bar{x})/h''(\bar{x})$ is the unique critical Lagrange multiplier.

Proof. We start with some technical relations. Since $h(\bar{x}) = h'(\bar{x}) = 0$, for $x \in \mathbf{R}$ it holds that

$$h(x) = \frac{1}{2}h''(\bar{x})(x - \bar{x})^2 + o((x - \bar{x})^2), \quad h'(x) = h''(\bar{x})(x - \bar{x}) + o(x - \bar{x})$$

as $x \to \bar{x}$, implying for $x \neq \bar{x}$ the relations

$$\lim_{x \to \bar{x}} \frac{h(x)}{(x - \bar{x})^2} = \frac{1}{2}h''(\bar{x}), \quad \lim_{x \to \bar{x}} \frac{h'(x)}{x - \bar{x}} = h''(\bar{x}).$$

Since $h''(\bar{x}) \neq 0$, these relations yield, in particular, that $h'(x) \neq 0$ for all x close enough to $\bar{x}$, and

$$\lim_{x \to \bar{x}} \frac{h(x)}{h'(x)(x - \bar{x})} = \frac{1}{2}, \quad \lim_{x \to \bar{x}} \frac{h(x)}{(h'(x))^2} h''(\bar{x}) = \frac{1}{2}. \tag{7.7}$$

Moreover, by L'Hôpital's rule, it holds that

$$\lim_{x \to \bar{x}} \frac{f'(x)}{h'(x)} = \lim_{x \to \bar{x}} \frac{f''(x)}{h''(x)} = -\bar{\lambda}. \tag{7.8}$$

Let $(x^k, \lambda^k) \in \mathbf{R} \times \mathbf{R}$ be the current iterate. Then according to (7.4), the next iterate (x^{k+1}, λ^{k+1}) must solve the system

$$\begin{aligned} (f''(x^k) + \lambda^k h''(x^k))(x - x^k) + h'(x^k)(\lambda - \lambda^k) &= -f'(x^k) - \lambda^k h'(x^k), \\ h'(x^k)(x - x^k) &= -h(x^k). \end{aligned} \tag{7.9}$$

If x^k is close enough to $\bar{x}$, the second equation in (7.9) uniquely defines

$$x^{k+1} = x^k - \frac{h(x^k)}{h'(x^k)}, \tag{7.10}$$

and hence,

$$\frac{x^{k+1} - \bar{x}}{x^k - \bar{x}} = 1 - \frac{h(x^k)}{h'(x^k)(x^k - \bar{x})}.$$

According to the first relation in (7.7), the number in the right-hand side can be made arbitrary close to $1/2$ by taking x^k close enough to $\bar{x}$. This implies that for any $x^0 \in \mathbf{R}$ close enough to $\bar{x}$, there exists the unique sequence $\{x^k\}$ such that for each k the point x^{k+1} satisfies the second equation in (7.9), this sequence converges to $\bar{x}$ at a linear rate, and moreover, (7.6) holds.

Furthermore, from the first equation in (7.9) and from (7.10), we see that λ^{k+1} must satisfy the equation

$$f'(x^k) + \lambda h'(x^k) = (f''(x^k) + \lambda^k h''(x^k)) \frac{h(x^k)}{h'(x^k)}$$

uniquely defining

$$\lambda^{k+1} = -\frac{f'(x^k)}{h'(x^k)} + \frac{h(x^k)}{(h'(x^k))^2}(f''(x^k) + \lambda^k h''(x^k)).$$

Therefore,

$$\begin{aligned} \lambda^{k+1} - \bar{\lambda} = &-\left(\frac{f'(x^k)}{h'(x^k)} + \bar{\lambda}\right) + \frac{h(x^k)}{(h'(x^k))^2} h''(x^k) \left(\frac{f''(x^k)}{h''(x^k)} + \bar{\lambda}\right) \\ &+ \frac{h(x^k)}{(h'(x^k))^2} h''(x^k)(\lambda^k - \bar{\lambda}). \end{aligned}$$

Employing the second relation in (7.7) and (7.8), this equality can be written in the form

$$\lambda^{k+1} - \bar{\lambda} = \varepsilon_k + \left(\frac{1}{2} + \delta_k\right)(\lambda^k - \bar{\lambda}),$$

where the sequences $\{\varepsilon_k\} \subset \mathbf{R}$ and $\{\delta_k\} \subset \mathbf{R}$ converge to zero. It can then be seen that the last relation implies convergence of $\{\lambda^k\}$ to $\bar{\lambda}$; we omit the details. $\square$

The situation considered in Proposition 7.2 is very special, of course. In particular, when $n = l = 1$, problem (7.1) is essentially a feasibility problem. However, it turns out that the phenomenon observed in Example 7.1 and justified in Proposition 7.2 is rather "persistent" in general (in higher dimensions, and when there are more constraints). Usually, critical multipliers form a special subclass within the multipliers violating the SOSC (7.5). What is important about them for our discussion here is that they serve as attractors for the dual sequences of Newton-type methods: we claim that convergence to such multipliers is something that should be expected to happen when the dual sequence converges to some point and critical multipliers exist. This is quite remarkable, considering that the set of critical multipliers is normally "thin" in $\mathcal{M}(\bar{x})$ (recall that in Example 7.1 and in Proposition 7.2, the critical multiplier is unique, while the multiplier set $\mathcal{M}(\bar{x})$ is the whole space). Moreover, when the dual sequence converges, the reason for slow primal convergence is precisely dual convergence to critical multipliers; if the dual sequence were to converge to a noncritical multiplier, primal convergence rate would have been superlinear.

We next provide theoretical justifications of these claims for the "most pure" case of degeneracy—full degeneracy, and when the problem data is quadratic (there are no terms of order higher than two). Specifically, let the objective function f and the constraints mapping h be given by

$$f(x) = \frac{1}{2}\langle Ax,\, x\rangle, \quad h(x) = \frac{1}{2}B[x,\, x], \tag{7.11}$$

for $x \in \mathbf{R}^n$, where $A \in \mathbf{R}^{n\times n}$ is a symmetric matrix, and $B : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^l$ is a symmetric bilinear mapping. The stationary point of interest is $\bar{x} = 0$, and the multiplier set is the whole dual space: $\mathcal{M}(0) = \mathbf{R}^l$.

For each $\lambda \in \mathbf{R}^l$ (and each $x \in \mathbf{R}^n$) define

$$H(\lambda) = \frac{\partial^2 L}{\partial x^2}(x,\, \lambda). \tag{7.12}$$

From (7.11) it follows that $H(\lambda)$ is the (symmetric) matrix of the quadratic form $\xi \to \langle A\xi,\, \xi\rangle + \langle \lambda,\, B[\xi,\, \xi]\rangle : \mathbf{R}^n \to \mathbf{R}$, and in particular, this matrix indeed does not depend on x. Since $h'(0) = 0$, a multiplier $\bar{\lambda} \in \mathbf{R}^l$ is critical if and only if $H(\bar{\lambda})$ is a singular matrix.

The following result is established in [149]. Interpretations of the technical assertions of Theorem 7.3, and in particular why they mean that convergence to a noncritical multiplier is an unlikely event, would be provided after its short proof.

Theorem 7.3. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}$ be defined by (7.11), where A is a symmetric $n\times n$-matrix, and $B : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^l$ is a symmetric bilinear mapping. Let $\{(x^k,\, \lambda^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l$ be a sequence such that $(x^{k+1},\, \lambda^{k+1})$ satisfies (7.4) for all k large enough, and assume that this sequence has a subsequence $\{(x^{k_j},\, \lambda^{k_j})\}$ convergent to $(0,\, \bar{\lambda})$ with some $\bar{\lambda} \in \mathbf{R}^l$, and $\{\lambda^{k_j+1}\}$ converges to the same $\bar{\lambda}$.*

Then

$$\left\langle \frac{\partial^2 L}{\partial x^2}(0,\, \bar{\lambda})x^{k_j},\, x^{k_j+1} \right\rangle = o(\|x^{k_j}\|\|x^{k_j+1}\|) \tag{7.13}$$

as $j \to \infty$, and if $\bar{\lambda}$ is a noncritical Lagrange multiplier, then in addition

$$x^{k_j+1} = o(\|x^{k_j}\|), \tag{7.14}$$

$$h''(0)[x^{k_j},\, x^{k_j}] = o(\|x^{k_j}\|^2). \tag{7.15}$$

Proof. Employing (7.12) and (7.4), for all k large enough we derive the equalities

$$H(\lambda^k)x^{k+1} + (B[x^k])^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) = 0, \tag{7.16}$$

$$B[x^k,\, x^{k+1}] = \frac{1}{2}B[x^k,\, x^k]. \tag{7.17}$$

Taking $k = k_j$ and multiplying both sides of (7.16) by x^{k_j}, we obtain the estimate

$$\begin{aligned}
\langle H(\lambda^{k_j})x^{k_j},\, x^{k_j+1}\rangle &= -\langle \lambda^{k_j+1} - \lambda^{k_j},\, B[x^{k_j},\, x^{k_j}]\rangle \\
&= -2\langle \lambda^{k_j+1} - \lambda^{k_j},\, B[x^{k_j},\, x^{k_j+1}]\rangle \\
&= o(\|x^{k_j}\|\|x^{k_j+1}\|)
\end{aligned}$$

as $j \to \infty$, where the second equality is by (7.17), and the third is by the convergence of $\{\lambda^{k_j}\}$ and $\{\lambda^{k_j+1}\}$ to the same limit. This implies (7.13).

Furthermore, if $\bar{\lambda}$ is noncritical, then $H(\bar{\lambda})$ is a nonsingular matrix. Then for all j large enough it holds that $H(\lambda^{k_j})$ is nonsingular and $(H(\lambda^{k_j}))^{-1}$ is well defined and has the norm bounded by some constant independent of j. For such j, from (7.16) we derive that

$$x^{k_j+1} = -(H(\lambda^{k_j}))^{-1}(B[x^{k_j}])^{\mathrm{T}}(\lambda^{k_j+1} - \lambda^{k_j}) = o(\|x^{k_j}\|)$$

as $j \to \infty$, where the last equality is again by the convergence of $\{\lambda^{k_j}\}$ and $\{\lambda^{k_j+1}\}$ to the same limit. This proves (7.14). By substituting the latter into (7.17), we finally obtain (7.15). □

Assuming that the entire sequence $\{(x^k,\, \lambda^k)\}$ converges to $(0,\, \bar{\lambda})$, relation (7.14) in Theorem 7.3 which holds in the case of noncritical dual limit $\bar{\lambda}$, means that the primal sequence $\{x^k\}$ converges to $\bar{x} = 0$ superlinearly. Relation (7.15) additionally says that $\{x^k\}$ converges tangentially to the null set of the quadratic mapping corresponding to $h''(0) = B$. In particular, such behavior is not possible if the null set of B contains only the zero point, which means that dual convergence to a noncritical multiplier is not possible in this case (apart from the cases of finite termination). This is precisely what happens in Example 7.1, and also in Example 7.4 below.

Finally, there is an asymptotic relation (7.13), and we next explain why the behavior characterized by (7.13) and (7.15) is highly unlikely to occur. Suppose there exists a subsequence $\{(x^{k_j},\, \lambda^{k_j})\}$ such that in addition to the assumptions in Theorem 7.3 it holds that $x^{k_j} \neq 0$ and $x^{k_j+1} \neq 0$ for all j, the sequence $\{x^{k_j}/\|x^{k_j}\|\}$ converges to some $\xi \in \mathbf{R}^n$, and the sequence $\{x^{k_j+1}/\|x^{k_j+1}\|\}$ converges either to ξ or to $-\xi$ as $j \to \infty$. (In particular, this holds if the entire $\{x^k\}$ converges to $\bar{x} = 0$ tangentially to a direction $\xi \neq 0$, i.e., $x^k = \|x^k\|\xi + o(\|x^k\|)$ as $k \to \infty$, which is a reasonable numerical behavior, though not automatic, in general.) Then relations (7.13) and (7.15) imply the equalities

$$\left\langle \frac{\partial^2 L}{\partial x^2}(0,\, \bar{\lambda})\xi,\, \xi \right\rangle = 0, \quad h''(0)[\xi,\, \xi] = 0,$$

further implying that

$$\langle f''(0)\xi,\, \xi\rangle = \langle f''(0)\xi,\, \xi\rangle + \langle \bar{\lambda},\, h''(0)[\xi,\, \xi]\rangle = \left\langle \frac{\partial^2 L}{\partial x^2}(0,\, \bar{\lambda})\xi,\, \xi \right\rangle = 0.$$

But then for any $\lambda \in \mathbf{R}^l$ we obtain

$$\left\langle \frac{\partial^2 L}{\partial x^2}(0,\, \lambda)\xi,\, \xi \right\rangle = \langle f''(0)\xi,\, \xi\rangle + \langle \lambda,\, h''(0)[\xi,\, \xi]\rangle = 0,$$

which means, in particular, that the SOSC (7.5) does not hold *for any* multiplier associated with $\bar{x}$. Moreover, the weaker condition

$$\forall\, \xi \in \ker h'(\bar{x}) \setminus \{0\} \quad \exists\, \lambda \in \mathcal{M}(\bar{x}) \text{ such that } \left\langle \frac{\partial^2 L}{\partial x^2}(0,\, \lambda)\xi,\, \xi \right\rangle > 0,$$

which is also sufficient for local optimality of $\bar{x} = 0$ (see Theorem 1.20), cannot hold as well. In addition, the condition

$$\forall\, \xi \in \ker h'(\bar{x}) \setminus \{0\} \quad \exists\, \lambda \in \mathcal{M}(\bar{x}) \text{ such that } \left\langle \frac{\partial^2 L}{\partial x^2}(0,\, \lambda)\xi,\, \xi \right\rangle < 0,$$

which is sufficient for $\bar{x} = 0$ to be a local solution of the maximization counterpart of problem (7.1), again cannot hold. Clearly, all this amounts to a rather special situation, and beyond it convergence to noncritical multipliers is highly unlikely.

Of course, the purely quadratic case considered above is very special. Unfortunately, if we perturb the data by higher-order terms, the argument becomes less clean. More precisely, the asymptotic relations (7.14) and (7.15) in

Theorem 7.3 remain valid (in case of convergence to a noncritical multiplier), but the relation (7.13) takes the cruder form

$$\left\langle \frac{\partial^2 L}{\partial x^2}(0,\, \bar{\lambda})x^{k_j},\, x^{k_j+1} \right\rangle = o(\|x^{k_j}\|\|x^{k_j+1}\|) + o(\|x^{k_j}\|^2),$$

which does not lead to the needed conclusions. The same happens for other kinds of perturbations. Also passing to the general case of degeneracy (when the first derivatives of f and h at $\bar{x}$ may not be zero) and reducing it to the fully degenerate case by means of the Liapunov–Schmidt procedure (see, e.g., [104, Ch. VII]) results in more-or-less the same effect as introducing higher-order terms in the problem data. Actually, beyond the fully degenerate case, convergence to noncritical multipliers appears to be slightly less unlikely and can be encountered in numerical experiments, albeit still rather rarely. At the same time, the theoretical claim of superlinear primal convergence in the case of dual convergence to a noncritical multiplier remains valid in the general case. This will be seen below in the setting of the perturbed Newton–Lagrange method, which covers more algorithms.

For theoretical analysis of attraction to critical Lagrange multipliers beyond the purely quadratic case, we refer to [146]. It should be mentioned though that [146] also does not give a full justification of the phenomenon: it does not show that the set of critical multipliers is indeed an attractor for dual sequences (as in the special case of Proposition 7.2). Also, the analysis in [146] is perhaps too technical for this book; thus we leave it out. Instead, we shall give some more illustrative examples of the effect of attraction to critical multipliers.

The problem in the following example is taken from [13, Example 1.1].

Example 7.4. Let $n = l = 2$,

$$A = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}, \quad B[x,\, x] = 2(x_1^2 - x_2^2,\, x_1 x_2),\ x \in \mathbf{R}^2.$$

Then problem (7.1) with f and h defined according to (7.11) has the unique feasible point (hence, unique solution) $\bar{x} = 0$, and the multiplier set is the whole dual space: $\mathcal{M}(0) = \mathbf{R}^2$. This solution violates the SOSC (7.5) for all $\bar{\lambda} \in \mathbf{R}^2$. Critical multipliers are those $\bar{\lambda} \in \mathbf{R}^2$ that satisfy the equation $4\bar{\lambda}_1^2 + \bar{\lambda}_2^2 = 4$. Observe that the null set of the quadratic mapping associated with B contains only the zero point. Hence, according to Theorem 7.3, primal convergence of the Newton–Lagrange method to $\bar{x} = 0$ with dual convergence to a noncritical multiplier is impossible.

For any k the iteration formulas (7.16) and (7.17) imply the relations

$$x^{k+1} = \frac{1}{2}x^k, \quad \lambda^{k+1} = \frac{1}{2}\lambda^k + \left(-\frac{(x_2^0)^2 - (x_1^0)^2}{2((x_1^0)^2 + (x_2^0)^2)},\, \frac{2x_1^0 x_2^0}{(x_1^0)^2 + (x_2^0)^2} \right), \quad (7.18)$$

and it can be seen that $\{\lambda^k\}$ converges to

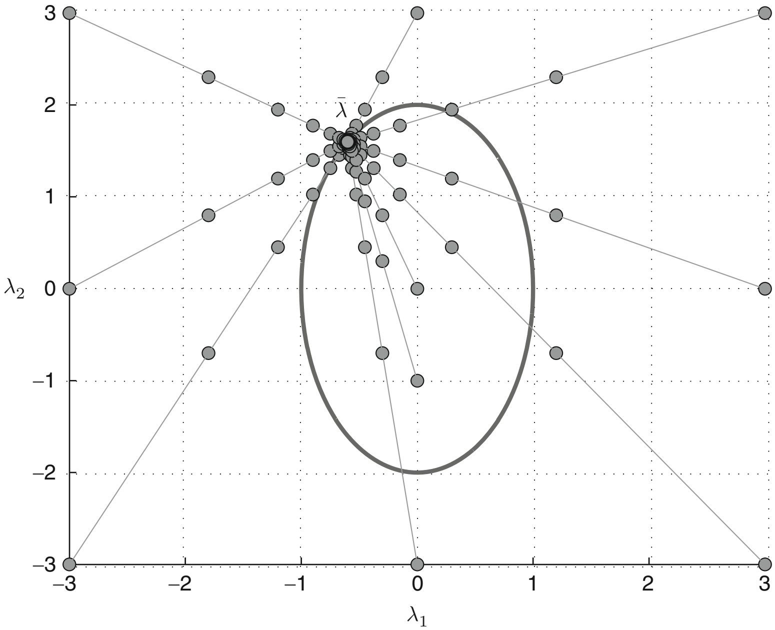

Fig. 7.1 Dual trajectories in Example 7.4 for $x^0 = (1,\, 2)$

$$\bar{\lambda} = \left(-\frac{(x_2^0)^2 - (x_1^0)^2}{((x_1^0)^2 + (x_2^0)^2)},\, \frac{4x_1^0 x_2^0}{(x_1^0)^2 + (x_2^0)^2} \right).$$

Therefore, the dual limit depends exclusively on x^0 and not on λ^0, and by direct substitution it can be checked that this limit is a critical multiplier whatever is taken as x^0. Moreover, from the first relation in (7.18) it is evident that the rate of primal convergence is only linear.

In Figs. 7.1, 7.2, critical multipliers are represented by the thick line (in the form of an oval). Some dual sequences of the Newton–Lagrange method, generated using primal starting points $x^0 = (1,\, 2)$ and $x^0 = (2,\, 1)$, are represented as dots (dual iterates) connected with thin lines.

Example 7.5. Let $n = 3$, $l = 2$,

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 4 \end{pmatrix}, \quad B[x,\, x] = (-x_1^2 + 2x_2^2 - x_3^2,\, 2x_1x_2).$$

Then problem (7.1) with f and h defined according to (7.11) has the unique solution $\bar{x} = 0$, with the multiplier set being $\mathcal{M}(0) = \mathbf{R}^2$, i.e., the whole dual space. This solution satisfies the SOSC (7.5) for $\bar{\lambda} \in \mathbf{R}^2$ such that $\bar{\lambda}_1 \in (0,\, 1)$, $(\bar{\lambda}_1 - 3)^2 - \bar{\lambda}_2^2 > 1$. Critical multipliers are those $\bar{\lambda} \in \mathbf{R}^2$ that satisfy $\bar{\lambda}_1 = 1$ or $(\bar{\lambda}_1 - 3)^2 - \bar{\lambda}_2^2 = 1$.

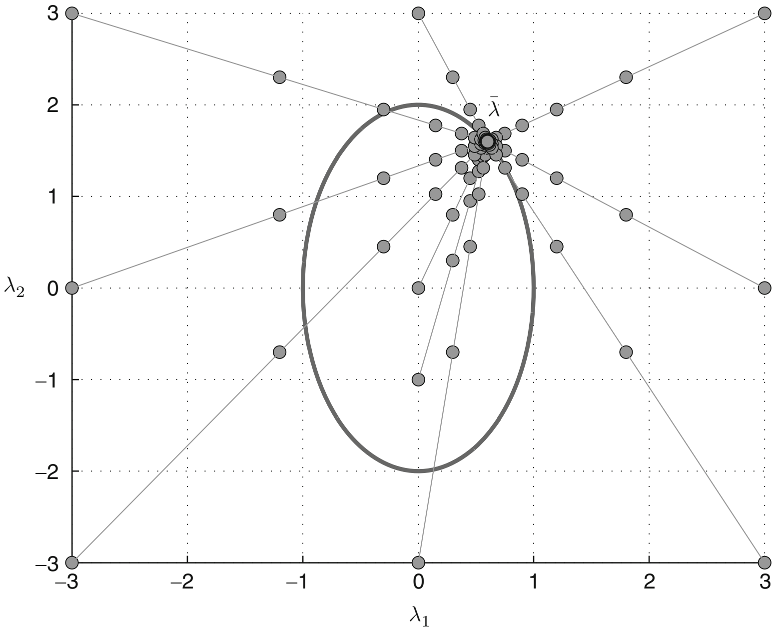

Fig. 7.2 Dual trajectories in Example 7.4 for $x^0 = (2, 1)$

In Figs. 7.3–7.6, critical multipliers are represented by the thick line (they form two branches of a hyperbola and a vertical straight line). Some dual sequences of the Newton–Lagrange method, generated using primal starting points $x^0 = (1, 2, 3)$ and $x^0 = (3, 2, 1)$, are shown in Figs. 7.3, 7.4. Figures 7.5, 7.6 show the distribution of dual iterates at the time of termination of the method according to a natural stopping criterion, for dual trajectories generated starting from the points on the grid in the domain $[-2, 8] \times [-4, 4]$ (step of the grid is 0.25).

Observe that in Fig. 7.6, there is at least one dual iterate λ^* at the time of termination, which is evidently not close to any critical multiplier. This is a result of non-convergence of the dual trajectory at the time the default stopping condition is triggered. Figure 7.7 presents the run that produces this point. It is interesting to note that if the sequence is continued, convergence to a critical multiplier does occur eventually.

Example 7.6. Let $n = 3$, $l = 2$,

$$f(x) = x_1^2 + x_2^2 + x_3^2, \quad h(x) = (x_1 + x_2 + x_3 + x_1^2 + x_2^2 + x_3^2,\ x_1 + x_2 + x_3 + x_1 x_3).$$

Problem (7.1) with this data has the unique solution $\bar{x} = 0$, with the multiplier set being a line in $\mathbf{R}^2$: $\mathcal{M}(0) = \{\lambda \in \mathbf{R}^2 \mid \lambda_1 + \lambda_2 = 0\}$. This solution violates the regularity condition (7.3) (since $\operatorname{rank} h'(0) = 1$), but satisfies the

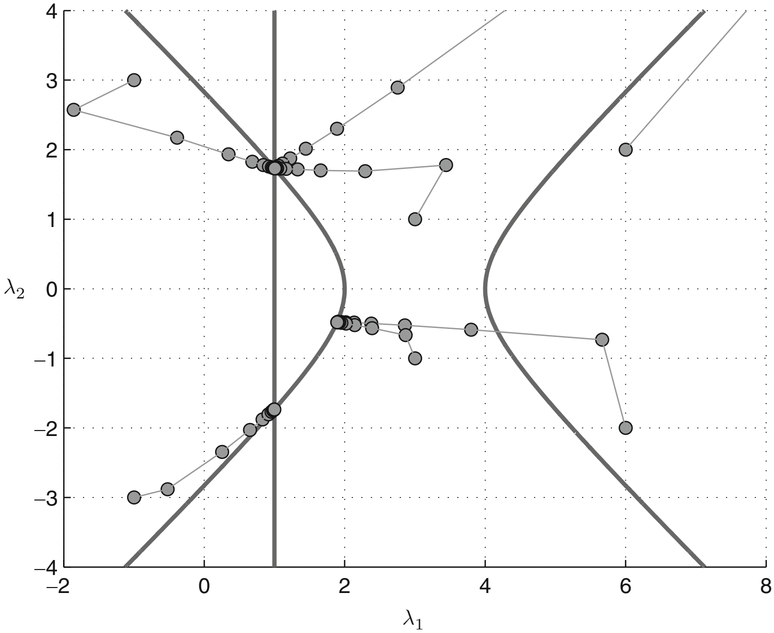

Fig. 7.3 Dual trajectories in Example 7.5 for $x^0 = (1,\, 2,\, 3)$

SOSC (7.5) for $\bar{\lambda} \in \mathcal{M}(0)$ such that $\bar{\lambda}_1 > -2/3$. Critical multipliers are two points on the line of all multipliers: $\bar{\lambda}^1 = (-6/5,\, 6/5)$ and $\bar{\lambda}^2 = (-2/3,\, 2/3)$.

Figure 7.8 shows the set of multipliers (thick gray line) and some dual sequences of the Newton–Lagrange method, generated using the primal starting point $x^0 = (1,\, 2,\, 3)$. Critical multipliers $\bar{\lambda}^1$ and $\bar{\lambda}^2$ are also marked in the figure but they can be barely seen because of dual iterates accumulating around those points.

Once the attraction of the basic Newton–Lagrange iterates to critical multipliers is exhibited, a natural and important question is whether the attraction phenomenon also shows up in relevant modifications of the algorithm. And if so, whether it still causes lack of superlinear convergence. In [149], an affirmative answer to these questions was justified by computational experiments for two other Newtonian algorithms: for the linearly constrained Lagrangian (LCL) method and for a class of quasi-Newton–Lagrange methods (see Sect. 4.1 for their descriptions). In what follows, we present a theoretical result from [149] demonstrating that in the case of dual convergence in the perturbed Newton–Lagrange framework (including LCL and the quasi-Newton modification) to a noncritical multiplier, primal superlinear convergence rate would have been preserved. Since the superlinear convergence rate is usually not observed in practice, this is yet another evidence that dual iterates are attracted to critical multipliers.

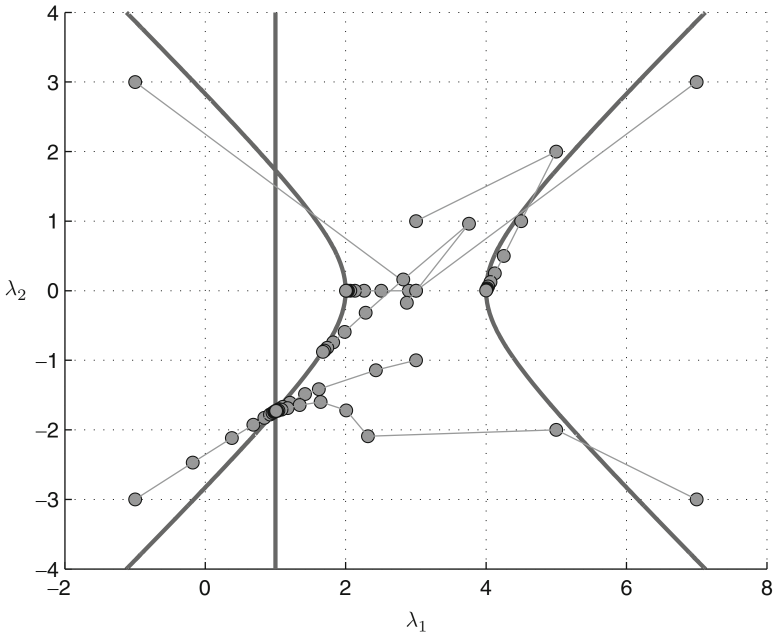

Fig. 7.4 Dual trajectories in Example 7.5 for $x^0 = (3,\, 2,\, 1)$

According to Sect. 4.1.1, in the perturbed Newton–Lagrange framework primal-dual sequences $\{(x^k,\, \lambda^k)\}$ satisfy for all k the equalities

$$\begin{gathered}\frac{\partial L}{\partial x}(x^k,\, \lambda^k) + \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)(x^{k+1} - x^k) + (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + \omega_1^k = 0,\\ h(x^k) + h'(x^k)(x^{k+1} - x^k) + \omega_2^k = 0,\end{gathered} \tag{7.19}$$

where $\omega_1^k \in \mathbf{R}^n$ and $\omega_2^k \in \mathbf{R}^l$ are the perturbation terms.

Primal-dual sequences $\{(x^k,\, \lambda^k)\}$ of quasi-Newton–Lagrange methods satisfy the equalities

$$\begin{aligned} H_k(x^{k+1} - x^k) + (h'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) &= -\frac{\partial L}{\partial x}(x^k,\, \lambda^k),\\ h'(x^k)(x^{k+1} - x^k) &= -h(x^k),\end{aligned} \tag{7.20}$$

where H_k is some quasi-Newton approximation of $\frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)$. We assume that $\{H_k\}$ is a bounded sequence, and that the Dennis-Moré-type condition from Theorem 4.6 holds:

$$\pi_{\ker h'(\bar{x})}\left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\right)(x^{k+1} - x^k) = o(\|x^{k+1} - x^k\|) \tag{7.21}$$

as $k \to \infty$. Obviously, (7.20) is a special case of (7.19) with

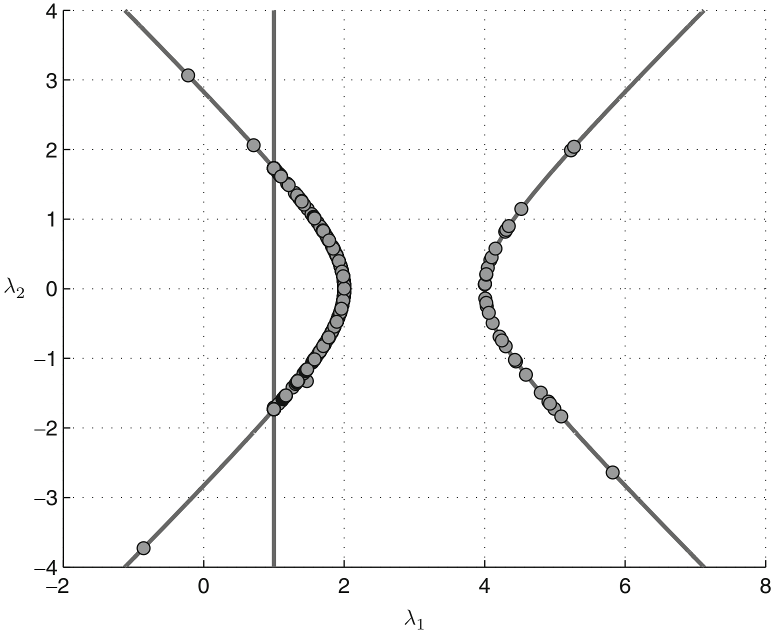

Fig. 7.5 Distribution of dual iterates at the time of termination in Example 7.5 for $x^0 = (1,\ 2,\ 3)$

$$\omega_1^k = \left(H_k - \frac{\partial^2 L}{\partial x^2}(x^k,\ \lambda^k) \right) (x^{k+1} - x^k), \quad \omega_2^k = 0.$$

Observe that ω_1^k depends linearly on $\xi^k = x^{k+1} - x^k$ and we can write $\omega_1^k = \Omega_k \xi^k$, where $\Omega_k \in \mathbf{R}^{n \times n}$. Moreover, under the stated requirements (namely, boundedness of $\{H_k\}$ and condition (7.21)), the operators Ω_k can be assumed to satisfy the following properties:

(QN1) The sequence $\{(I - \pi_{\ker h'(\bar{x})})\Omega_k\}$ is bounded.
(QN2) It holds that $\|\pi_{\ker h'(\bar{x})}\Omega_k\| \to 0$ as $k \to \infty$.

The LCL method was stated above as Algorithm 4.9. Primal-dual sequences $\{(x^k,\ \lambda^k)\}$ of the LCL method are generated as follows: for each k the primal iterate x^{k+1} is a stationary point of the subproblem

$$\begin{aligned} &\text{minimize} \quad f(x) + \langle \lambda^k,\ h(x) \rangle + \frac{c_k}{2}\|h(x)\|^2 \\ &\text{subject to} \ \ h(x^k) + h'(x^k)(x - x^k) = 0, \end{aligned} \tag{7.22}$$

while the corresponding dual iterate has the form $\lambda^{k+1} = \lambda^k + \eta^k$, where η^k is a Lagrange multiplier of the subproblem (7.22) associated with x^{k+1}. Here $c_k \geq 0$ is the penalty parameter, which for the purposes of local analysis we assume to be fixed from some iteration on: $c_k = c$ for all k sufficiently large.

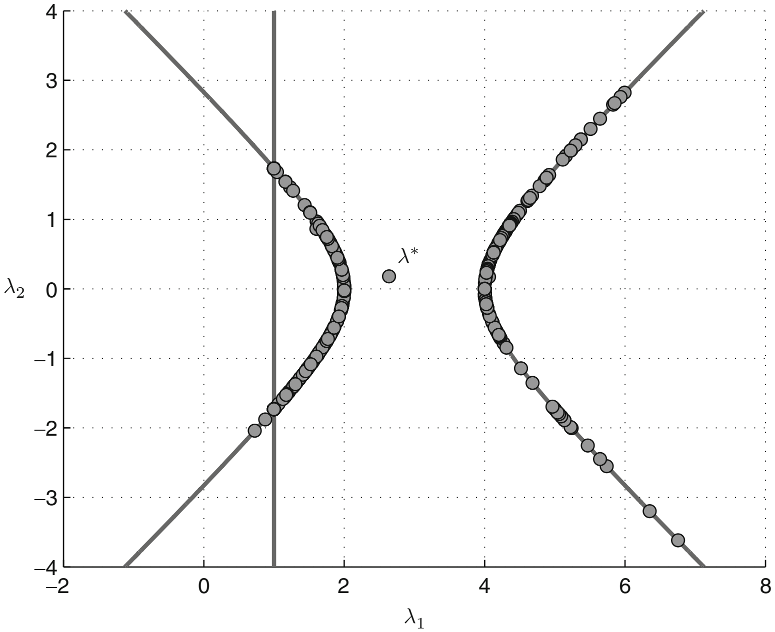

Fig. 7.6 Distribution of dual iterates at the time of termination in Example 7.5 for $x^0 = (3,\, 2,\, 1)$

As demonstrated in Sect. 4.1.2, such primal-dual sequences satisfy (7.19) with

$$\omega_1^k = \frac{\partial L}{\partial x}(x^{k+1},\, \lambda^k) - \frac{\partial L}{\partial x}(x^k,\, \lambda^k) - \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)(x^{k+1} - x^k)$$
$$+c(h'(x^{k+1}))^{\mathrm{T}}(h(x^{k+1}) - h(x^k) - h'(x^k)(x^{k+1} - x^k)$$

and

$$\omega_2^k = 0.$$

We can write $\omega_1^k = \omega_1(x^k,\, \lambda^k,\, x^{k+1} - x^k)$, where the mapping $\omega_1 : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n \to \mathbf{R}^n$ is defined by

$$\omega_1(x,\, \lambda,\, \xi) = \frac{\partial L}{\partial x}(x + \xi,\, \lambda) - \frac{\partial L}{\partial x}(x,\, \lambda) - \frac{\partial^2 L}{\partial x^2}(x,\, \lambda)\xi$$
$$+c(h'(x + \xi))^{\mathrm{T}}(h(x + \xi)) - h(x) - h'(x)\xi.$$

By the mean-value theorem (see Theorem A.10, (a)), one can directly verify that for any fixed $\bar{\lambda} \in \mathbf{R}^l$ the mapping ω_1 satisfies the following assumptions:

(LCL1) The estimate

$$(I - \pi_{\ker h'(\bar{x})})(\omega_1(x,\, \lambda,\, \xi^1) - \omega_1(x,\, \lambda,\, \xi^2)) = O(\|\xi^1 - \xi^2\|)$$

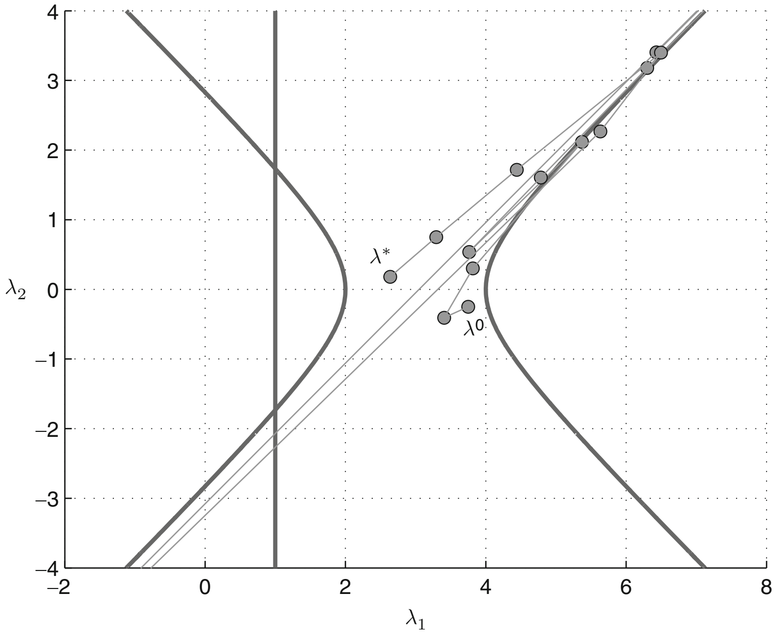

Fig. 7.7 Dual trajectory in Example 7.5 for $x^0 = (3,\, 2,\, 1)$, $\lambda^0 = (3.75,\, -0.25)$

holds as $(x,\, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$ tends to $(\bar{x},\, \bar{\lambda})$, and $\xi^1,\, \xi^2 \in \mathbf{R}^n$ tend to 0, and the value $\|(I - \pi_{\ker h'(\bar{x})})\omega_1(x,\, \lambda,\, 0)\|$ is bounded (by an independent constant) for $(x,\, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$ close to $(\bar{x},\, \bar{\lambda})$.

(LCL2) For each $\varepsilon > 0$ it holds that

$$\|\pi_{\ker h'(\bar{x})}(\omega_1(x,\, \lambda,\, \xi^1) - \omega_1(x,\, \lambda,\, \xi^2))\| \leq \varepsilon\|\xi^1 - \xi^2\|$$

for all $(x,\, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$ close enough to $(\bar{x},\, \bar{\lambda})$, and all $\xi^1,\, \xi^2 \in \mathbf{R}^n$ close enough to 0, and the estimate

$$\pi_{\ker h'(\bar{x})}\omega_1(x,\, \lambda,\, 0) = O(\|x - \bar{x}\|)$$

holds as $(x,\, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$ tends to $(\bar{x},\, \bar{\lambda})$.

Thus, quasi-Newton–Lagrange and LCL methods both fit the perturbed Newton–Lagrange framework, but with perturbations terms satisfying two different sets of assumptions. The next theorem states that in both cases the superlinear primal convergence is achieved if the dual sequence converges to a noncritical multiplier.

Theorem 7.7. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice differentiable in a neighborhood of a point $\bar{x} \in \mathbf{R}^n$, with their second derivatives continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (7.1), and let $\bar{\lambda} \in \mathbf{R}^l$ be a*

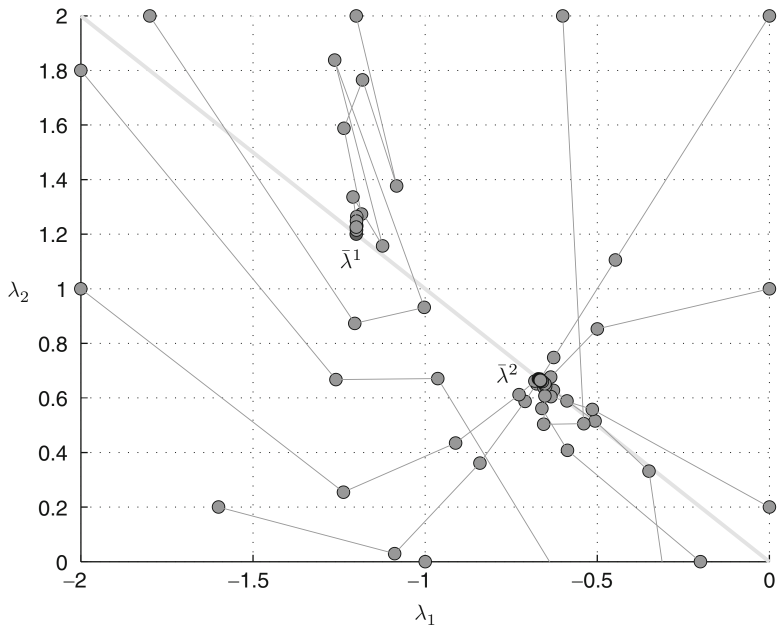

Fig. 7.8 Dual trajectories in Example 7.6 for $x^0 = (1,\, 2,\, 3)$

Lagrange multiplier associated with $\bar{x}$. Let $\{(x^k,\, \lambda^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l$ be a sequence convergent to the point $(\bar{x},\, \bar{\lambda})$ and satisfying (7.19) with $\omega_2^k = 0$ for each k large enough, where either $\omega_1^k = \Omega_k \xi^k$ with $\Omega_k \in \mathbf{R}^{n\times n}$ satisfying the assumptions (QN1), (QN2), or $\omega_1^k = \omega_1(x^k,\, \lambda^k,\, \xi^k)$ with a mapping $\omega_1 : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n \to \mathbf{R}^n$ satisfying the assumptions (LCL1), (LCL2).

If $\bar{\lambda}$ is a noncritical multiplier, then $\{x^k\}$ converges to $\bar{x}$ superlinearly.

Proof. We can assume, without loss of generality, that $\bar{x} = 0$ and $f(0) = 0$. Then for $x \in \mathbf{R}^n$ we can write

$$f(x) = \langle a,\, x\rangle + \frac{1}{2}\langle Ax,\, x\rangle + r(x), \quad h(x) = B_1 x + \frac{1}{2} B_2[x,\, x] + R(x),$$

where $a \in \mathbf{R}^n$; $A \in \mathbf{R}^{n\times n}$ is symmetric; $B_1 \in \mathbf{R}^{l\times n}$; $B_2 : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^l$ is a symmetric bilinear mapping; function $r : \mathbf{R}^n \to \mathbf{R}$ and mapping $R : \mathbf{R}^n \to \mathbf{R}^l$ are twice differentiable near 0, their second derivatives are continuous at 0, and $r(0) = 0$, $r'(0) = 0$, $r''(0) = 0$, $R(0) = 0$, $R'(0) = 0$, $R''(0) = 0$. In the rest of the proof we assume that $r(\cdot) \equiv 0$ and $R(\cdot) \equiv 0$, but strictly in order to simplify the presentation, as it is rather cumbersome even with this simplification. The needed assertions remain valid under the general assumptions regarding r and R that are stated above; we omit those details.

For the specified setting, the Lagrange system (7.2) takes the form

$$a + B_1^{\mathrm{T}}\lambda + H(\lambda)x = 0, \quad B_1 x + \frac{1}{2}B_2[x,\, x] = 0,$$

where for any $\lambda \in \mathbf{R}^l$ the matrix $H(\lambda)$ is defined according to (7.12). In particular, $H(\lambda)$ can be viewed as the (symmetric) matrix of the quadratic form $\xi \to \langle A\xi,\, \xi\rangle + \langle \lambda,\, B_2[\xi,\, \xi]\rangle : \mathbf{R}^n \to \mathbf{R}$. Stationarity of $\bar{x} = 0$ in problem (7.1) then means that $a \in \operatorname{im} B_1^{\mathrm{T}}$, in which case $\mathcal{M}(0)$ is an affine set parallel to $\ker B_1^{\mathrm{T}}$.

Furthermore, for a sequence $\{(x^k,\, \lambda^k)\}$ satisfying (7.19) with $\omega_2^k = 0$ it then holds that $x^{k+1} = x^k + \xi^k$, $\lambda^{k+1} = \lambda^k + \eta^k$, where for each k the primal-dual step $(\xi^k,\, \eta^k) \in \mathbf{R}^n \times \mathbf{R}^l$ satisfies the relations

$$\begin{aligned} H_k\xi^k + B_1^{\mathrm{T}}\eta^k + (B_2[x^k])^{\mathrm{T}}\eta^k &= -a - B_1^{\mathrm{T}}\lambda^k - H_k x^k - \omega_1^k, \\ B_1\xi^k + B_2[x^k,\, \xi^k] &= -B_1 x^k - \frac{1}{2}B_2[x^k,\, x^k], \end{aligned} \tag{7.23}$$

with $H_k = H(\lambda^k)$.

For brevity, set $\mathit{\Pi} = \pi_{\ker B_1}$, $P = \pi_{\ker B_1^{\mathrm{T}}}$. For each $x \in \mathbf{R}^n$ and $\lambda \in \mathbf{R}^l$ define $\hat{H}(\lambda)$ as the (symmetric) matrix of the quadratic form $\xi \to \langle \frac{\partial^2 L}{\partial x^2}(x,\, \lambda)\xi,\, \xi\rangle = \langle A\xi,\, \xi\rangle + \langle \lambda,\, B_2[\xi,\, \xi]\rangle : \ker B_1 \to \mathbf{R}$, that is,

$$\hat{H}(\lambda)\xi = \mathit{\Pi} H(\lambda)\xi, \quad \xi \in \ker B_1. \tag{7.24}$$

This $\hat{H}(\lambda)$ is the reduced Hessian of the Lagrangian, and therefore, $\bar{\lambda} \in \mathcal{M}(0)$ is a noncritical multiplier if and only if the matrix $\hat{H}(\bar{\lambda})$ is nonsingular.

For each k we set $\hat{H}_k = \hat{H}(\lambda^k)$. Evidently, $\{H_k\}$ converges to $H(\bar{\lambda})$, and hence, $\{\hat{H}_k\}$ converges to $\hat{H}(\bar{\lambda})$. Therefore, assuming that $\bar{\lambda}$ is noncritical, for all k large enough it holds that $\hat{H}_k$ is nonsingular and $\hat{H}_k^{-1}$ is well defined and has the norm bounded by some constant independent of k.

In order to estimate $(\xi^k,\, \eta^k)$, we shall make use of the Liapunov–Schmidt procedure, well known in the bifurcation theory (e.g., [104, Ch. VII]). For each index k, we define the decompositions $x^k = x_1^k + x_2^k$, $\xi^k = \xi_1^k + \xi_2^k$, $x_1^k,\, \xi_1^k \in (\ker B_1)^{\perp} = \operatorname{im} B_1^{\mathrm{T}}$, $x_2^k,\, \xi_2^k \in \ker B_1$, and $\lambda^k = \lambda_1^k + \lambda_2^k$, $\eta^k = \eta_1^k + \eta_2^k$, $\lambda_1^k,\, \eta_1^k \in \operatorname{im} B_1$, $\lambda_2^k,\, \eta_2^k \in (\operatorname{im} B_1)^{\perp} = \ker B_1^{\mathrm{T}}$. Applying $(I - \mathit{\Pi})$ and $\mathit{\Pi}$ to both sides of the first equality in (7.23), and taking into account the inclusion $a \in \operatorname{im} B_1^{\mathrm{T}}$, we obtain that

$$\begin{aligned} &B_1^{\mathrm{T}}\eta_1^k + (I - \mathit{\Pi})(H_k\xi^k + (B_2[x^k])^{\mathrm{T}}\eta^k) \\ &= -a - B_1^{\mathrm{T}}\lambda_1^k - (I - \mathit{\Pi})(H_k x^k + \omega_1^k), \end{aligned} \tag{7.25}$$

$$\mathit{\Pi}(H_k\xi^k + (B_2[x^k])^{\mathrm{T}}\eta^k) = -\mathit{\Pi}(H_k x^k + \omega_1^k). \tag{7.26}$$

Furthermore, applying $(I - P)$ and P to both sides of the second equality in (7.23), we obtain that

$$B_1\xi_1^k + (I - P)B_2[x^k, \xi^k] = -B_1x_1^k - \frac{1}{2}(I - P)B_2[x^k, x^k], \tag{7.27}$$

$$PB_2[x^k, \xi^k] = -\frac{1}{2}PB_2[x^k, x^k]. \tag{7.28}$$

Clearly, the linear operators $\mathcal{B} : (\ker B_1)^\perp = \operatorname{im} B_1^{\mathrm{T}} \to \operatorname{im} B_1$, defined by $\mathcal{B}\xi = B_1\xi$, and $\mathcal{B}^* : \operatorname{im} B_1 \to \operatorname{im} B_1^{\mathrm{T}} = (\ker B_1)^\perp$, defined by $\mathcal{B}^*\eta = B_1^{\mathrm{T}}\eta$, are invertible. By Lemma A.6 it follows that for each k large enough, the linear operators

$$\mathcal{B}_k : \operatorname{im} B_1^{\mathrm{T}} \to \operatorname{im} B_1, \quad \mathcal{B}_k\xi_1 = B_1\xi + (I - P)B_2[x^k, \xi],$$

and

$$\mathcal{B}_k^* : \operatorname{im} B_1 \to \operatorname{im} B_1^{\mathrm{T}}, \quad \mathcal{B}_k^*\eta = B_1^{\mathrm{T}}\eta + (I - \Pi)(B_2[x^k])^{\mathrm{T}}\eta,$$

are invertible, and

$$\mathcal{B}_k^{-1} = \mathcal{B}^{-1} + O(\|x^k\|), \quad (\mathcal{B}_k^*)^{-1} = (\mathcal{B}^*)^{-1} + O(\|x^k\|) \tag{7.29}$$

as $x^k \to 0$.

Relation (7.27) evidently implies that

$$\mathcal{B}_k\xi_1^k + (I - P)B_2[x^k, \xi_2^k] = -\mathcal{B}_kx_1^k - \frac{1}{2}(I - P)B_2[x^k, x^k].$$

Applying $\mathcal{B}_k^{-1}$ to both sides of the latter equality, we now obtain that, for a fixed ξ_2^k, there exists the unique ξ_1^k satisfying (7.27), and

$$\xi_1^k = -x_1^k + \tilde{M}_k\xi_2^k + O(\|x^k\|^2) \tag{7.30}$$

as $x^k \to 0$, where we defined the linear operator

$$\tilde{M}_k = \tilde{M}(x^k) = -\mathcal{B}_k^{-1}(I - P)B_2[x^k].$$

Observe that $\tilde{M}(x) = O(\|x\|)$ as $x \to 0$. Note also that $\xi_1^k = \xi_1(x^k, \xi_2^k)$, where $\xi_1 : \mathbf{R}^n \times \ker B_1 \to (\ker B_1)^\perp$ is the affine mapping such that

$$\xi_1(x, \xi^1) - \xi_1(x, \xi^2) = \tilde{M}(x)(\xi^1 - \xi^2) \tag{7.31}$$

holds for all $x \in \mathbf{R}^n$ and $\xi^1, \xi^2 \in \ker B_1$, and

$$\xi_1(x, 0) = O(\|x\|) \tag{7.32}$$

as $x \to 0$.

Furthermore, the relation (7.25) can be written in the form

$$\mathcal{B}_k^*\eta_1^k + (I - \Pi)(H_k\xi^k + (B_2[x^k])^{\mathrm{T}}\eta_2^k) = -a - B_1^{\mathrm{T}}\lambda_1^k - (I - \Pi)(H_kx^k + \omega_1^k).$$

Applying $(\mathcal{B}_k^*)^{-1}$ to both sides of the latter equality, and taking into account the convergence of $\{\lambda^k\}$ to $\bar{\lambda} \in \mathcal{M}(0)$ (implying $\{\eta^k\} \to 0$), the second relation in (7.29), and (7.30), we derive that for fixed ξ_2^k and η_2^k, there exists the unique η_1^k satisfying (7.25) (with uniquely defined $\xi_1^k = \xi_1(x^k,\, \xi_2^k)$), and

$$\begin{aligned}\eta_1^k &= -(\mathcal{B}_k^*)^{-1}(a + B_1^{\mathrm{T}}\lambda_1^k)\\ &\quad -(\mathcal{B}_k^*)^{-1}(I-\Pi)((B_2[x^k])^{\mathrm{T}}\eta_2^k + H_k(x^k+\xi^k)+\omega_1^k)\\ &= -(\mathcal{B}_k^*)^{-1}\mathcal{B}^*((\mathcal{B}^*)^{-1}a + \lambda_1^k)\\ &\quad -(\mathcal{B}_k^*)^{-1}(I-\Pi)((B_2[x^k])^{\mathrm{T}}\eta_2^k + H_k(x^k+\xi^k)+\omega_1^k)\\ &= \hat{\lambda} - \lambda_1^k\\ &\quad -(\mathcal{B}_k^*)^{-1}(I-\Pi)(H_k(x_2^k + \xi_2^k + \tilde{M}_k\xi_2^k) + \omega_1^k) + o(\|x^k\|)\end{aligned} \tag{7.33}$$

as $k \to \infty$, where $\hat{\lambda} = -(\mathcal{B}^*)^{-1}a \in \mathcal{M}(\bar{x}) \cap \operatorname{im} B_1$ is the uniquely defined *normal Lagrange multiplier* (the one with the smallest norm), and that $\{\lambda_1^k\} \to \hat{\lambda}$. We then further obtain that

$$\begin{aligned}\eta_1^k &= -(\lambda_1^k - \hat{\lambda}) - (\mathcal{B}_k^*)^{-1}(I-\Pi)(H_k((I+\tilde{M}_k)(x_2^k+\xi_2^k) - \tilde{M}_k x_2^k) + \omega_1^k)\\ &\quad + o(\|x^k\|)\\ &= -(\lambda_1^k - \hat{\lambda}) + C_k(x_2^k+\xi_2^k) - (\mathcal{B}_k^*)^{-1}(I-\Pi)\omega_1^k + o(\|x^k\|)\end{aligned} \tag{7.34}$$

as $k \to \infty$, where we defined the linear operator

$$C_k = -(\mathcal{B}_k^*)^{-1}(I-\Pi)H_k(I+\tilde{M}_k).$$

Note that the sequence $\{C_k\}$ is bounded.

By substituting (7.30) and (7.34) into (7.26) we obtain that

$$\begin{gathered}\Pi(H_k(\xi_2^k - x_1^k + \tilde{M}_k\xi_2^k)\\ +(B_2[x^k])^{\mathrm{T}}(\eta_2^k - (\lambda_1^k - \hat{\lambda}) + C_k(x_2^k+\xi_2^k) - (\mathcal{B}_k^*)^{-1}(I-\Pi)\omega_1^k))\\ = -\Pi(H_k x^k + \omega_1^k + o(\|x^k\|)),\end{gathered}$$

and hence, taking into account the convergence of $\{\eta^k\}$ to 0,

$$\begin{aligned}&\Pi((H_k(I_n + \tilde{M}_k) + (B_2[x^k])^{\mathrm{T}}C_k))\xi_2^k\\ &= -\Pi(H_k(x^k - x_1^k) - (B_2[x^k])^{\mathrm{T}}(\mathcal{B}_k^*)^{-1}(I-\Pi)\omega_1^k + \omega_1^k + o(\|x^k\|)),\end{aligned}$$

which can be written in the form

$$\begin{aligned}&\Pi(H_k + \hat{M}_k)\xi_2^k\\ &= -\Pi(H_k x_2^k - (B_2[x^k])^{\mathrm{T}}(\mathcal{B}_k^*)^{-1}(I-\Pi)\omega_1^k + \omega_1^k + o(\|x^k\|))\end{aligned} \tag{7.35}$$

as $k \to \infty$, where we defined the linear operator

$$\hat{M}_k = H_k\tilde{M}_k + (B_2[x^k])^{\mathrm{T}}C_k.$$

Note that $\hat{M}_k = O(\|x^k\|)$ as $x^k \to 0$.

Observe further that for each k large enough, the linear operator

$$\mathcal{H}_k : \ker B_1 \to \ker B_1, \quad \mathcal{H}_k \xi = \Pi (H_k + \hat{M}_k)\xi,$$

is invertible, and by Lemma A.6 it holds that

$$\mathcal{H}_k^{-1} = \hat{H}_k^{-1} + O(\|x^k\|),$$

as $x^k \to 0$, where (7.24) was taken into account.

Consider first the case when the perturbation term is $\omega_1^k = \Omega_k \xi^k$, where $\Omega_k \in \mathbf{R}^{n\times n}$ for each k, satisfying (QN1), (QN2).

For all k large enough, applying $\mathcal{H}_k^{-1}$ to both sides of (7.35) and using the convergence of $\{\eta^k\}$ to 0, (7.30) and assumptions (QN1) and (QN2), we obtain that

$$\begin{aligned}
\xi_2^k &= -x_2^k + \hat{H}_k^{-1}\Pi((B_2[x^k])^{\mathrm{T}}((\mathcal{B}_k^*)^{-1}(I-\Pi)\Omega^k\xi^k) - \Omega^k\xi^k) + o(\|x^k\|)\\
&= -x_2^k - \hat{H}_k^{-1}\Pi(I - (B_2[x^k])^{\mathrm{T}}(\mathcal{B}_k^*)^{-1}(I-\Pi))\Omega^k(\xi_2^k - x_1^k + \tilde{M}_k\xi_2^k)\\
&\quad + o(\|x^k\|)\\
&= -x_2^k + S_k\xi_2^k + o(\|x^k\|)
\end{aligned} \tag{7.36}$$

as $k \to \infty$, where we defined the linear operator

$$S_k = \hat{H}_k^{-1}\Pi(I - (B_2[x^k])^{\mathrm{T}}(\mathcal{B}_k^*)^{-1}(I-\Pi)(I+\tilde{M}_k)\Omega^k.$$

Note that $\|S_k\| \to 0$ as $k \to \infty$. Applying now the linear operator $(I - S_k)^{-1}$ to both sides of the equality

$$(I - S_k)\xi_2^k = -x_2^k + o(\|x^k\|)$$

(following from (7.36)), gives

$$\xi_2^k = -x_2^k + o(\|x^k\|), \tag{7.37}$$

and in particular

$$\xi_2^k = O(\|x^k\|) \tag{7.38}$$

as $k \to \infty$.

Before proceeding with the argument, we prove the same relations (7.37) and (7.38) under the second set of assumptions on ω_1^k, i.e., in the case when $\omega_1^k = \omega_1(x^k,\, \lambda^k,\, \xi^k)$, where $\omega_1 : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n \to \mathbf{R}^n$ satisfies the assumptions (LCL1), (LCL2). More precisely, we first show that for all k large enough, the equation (7.26) uniquely defines ξ_2^k, satisfying (7.38).

Taking into account assumption (LCL1) and (7.31), observe that by the first equality in (7.33), one can write $\eta_1^k = \eta_1(x^k,\, \lambda^k,\, \xi_2^k,\, \eta_2^k)$, where the mapping $\eta_1 : \mathbf{R}^n \times \mathbf{R}^l \times \ker B_1 \times (\operatorname{im} B_1)^\perp \to \operatorname{im} B_1$ is affine and such that

$$\eta_1(x,\, \lambda,\, \xi^1,\, \eta_2) - \eta_1(x,\, \lambda,\, \xi^2,\, \eta_2) = O(\|\xi^1 - \xi^2\|)$$

holds as $(x,\, \lambda)$ tends to $(0,\, \bar{\lambda})$, $\xi^1,\, \xi^2 \in \ker B_1$ tend to 0, and $\eta_2 \in (\operatorname{im} B_1)^\perp$ tends to 0, and the value $\|\eta_1(x,\, \lambda,\, 0,\, \eta_2)\|$ is bounded (by an independent constant) for all $(x,\, \lambda)$ close to $(\bar{x},\, \bar{\lambda})$ and all $\eta_2 \in (\operatorname{im} B_1)^\perp$ close to 0.

Consider the mapping $\Phi : \mathbf{R}^n \times \mathbf{R}^l \times (\operatorname{im} B_1)^\perp \times \ker B_1 \to \ker B_1$,

$$\begin{aligned}\Phi(x,\, \lambda,\, \eta_2,\, \xi_2) = \Pi(&H(\lambda)\xi_2 + H(\lambda)(x + \xi_1(x,\, \xi_2))\\ &+(B_2[x])^{\mathrm{T}}(\eta_1(x,\, \lambda,\, \xi_2,\, \eta_2) + \eta_2)\\ &+\omega_1(x,\, \lambda,\, \xi_1(x,\, \xi_2) + \xi_2)).\end{aligned}$$

Then (7.25) is equivalent to

$$\Phi(x,\, \lambda,\, \eta_2,\, \xi_2) = 0 \tag{7.39}$$

with $x = x^k$, $\lambda = \lambda^k$, $\eta_2 = \eta_2^k$, $\xi_2 = \xi_2^k$. Note that $(0,\, \bar{\lambda},\, 0,\, 0)$ is a solution of (7.39). We shall apply the implicit function theorem (Theorem 1.22) to equations (7.39) at this point, treating $(x,\, \lambda,\, \eta_2)$ as a parameter.

Let $J = \hat{H}(\bar{\lambda})$. Employing the assumption (LCL2) and the above-established properties of $\xi_1(\cdot)$ and $\eta_1(\cdot)$, we derive that for each $\varepsilon > 0$

$$\begin{aligned}&\|\Phi(x,\, \lambda,\, \eta_2,\, \xi^1) - \Phi(x,\, \lambda,\, \eta_2,\, \xi^2) - J(\xi^1 - \xi^2)\|\\ &\le \|(\hat{H}(\lambda) - J)(\xi^1 - \xi^2)\| + \|\Pi H(\lambda)(\xi_1(x,\, \xi^1) - \xi_1(x,\, \xi^2))\|\\ &\quad+\|\Pi(B_2[x])^{\mathrm{T}}(\eta_1(x,\, \lambda,\, \xi^1,\, \eta_2) - \eta_1(x,\, \lambda,\, \xi^2,\, \eta_2))\|\\ &\quad+\|\omega_1(x,\, \lambda,\, \xi_1(x,\, \xi^1) + \xi^1) - \omega_1(x,\, \lambda,\, \xi_1(x,\, \xi^2) + \xi^2)\|\\ &\le \varepsilon\|\xi^1 - \xi^2\|\end{aligned}$$

holds for all $(x,\, \lambda)$ close enough to $(0,\, \bar{\lambda})$, all $\eta_2 \in (\operatorname{im} B_1)^\perp$, and all $\xi^1,\, \xi^2 \in \ker B_1$ close enough to 0. Thus, Ψ satisfies near $(0,\, \bar{\lambda},\, 0,\, 0)$ assumption (i) of Theorem 1.22. Moreover, $J = \frac{\partial \Phi}{\partial \xi_2}(0,\, \bar{\lambda},\, 0,\, 0)$ is nonsingular, which is assumption (ii) of Theorem 1.22. Finally, we have that the mapping $(x,\, \lambda,\, \eta_2) \to \Psi(x,\, \lambda,\, \eta_2,\, 0) : \mathbf{R}^n \times \mathbf{R}^l \times (\operatorname{im} B_1)^\perp \to \mathbf{R}^n$ is continuous at $(0,\, \bar{\lambda},\, 0)$, which is assumption (iii) of Theorem 1.22.

Theorem 1.22 now implies that there exists a neighborhood U of 0 in $\ker B_1$ such that for all $(x,\, \lambda,\, \eta_2) \in \mathbf{R}^n \times \mathbf{R}^l \times (\operatorname{im} B_1)^\perp$ close enough to $(0,\, \bar{\lambda},\, 0)$, there exists the unique $\xi_2 = \xi_2(x,\, \lambda,\, \eta_2) \in U$ satisfying (7.39), and

$$\begin{aligned}\xi_2 &= O(\|\Phi(x,\, \lambda,\, \eta_2,\, 0)\|)\\ &= O(\Pi(H(\lambda)(x + \xi_1(x,\, 0)) + (B_2[x])^{\mathrm{T}}(\eta_1(x,\, \lambda,\, 0,\, \eta_2) + \eta_2)\\ &\quad+\omega_1(x,\, \lambda,\, \xi_1(x,\, 0))))\\ &= O(\|x\|)\end{aligned}$$

as $(x,\, \lambda) \to (0,\, \bar{\lambda})$ and $\eta_2 \to 0$, where we have again employed assumption (LCL2) and the above-established properties of $\xi_1(\cdot)$ and $\eta_1(\cdot)$.

Since $\{x^k\}$ converges, it follows that $\{\xi^k\} \to 0$, and in particular, $\xi_2^k \in U$ for all k large enough. Furthermore, for any k, since (ξ^k, η^k) is a solution of (7.23), from the argument above it follows that ξ_2^k solves (7.39) with $x = x^k$, $\lambda = \lambda^k$, $\eta_2 = \eta_2^k$. Taking into account the convergence of $\{(x^k, \lambda^k)\}$ to $(0, \bar{\lambda})$, and of $\{\eta^k\}$ to 0, we conclude that $\xi_2^k = \xi_2(x^k, \lambda^k, \eta_2^k)$ for all k large enough, implying (7.38).

Applying now the linear operator $\mathcal{H}_k^{-1}$ to both sides of (7.35) and using assumptions (LCL1), (LCL2) and (7.30), (7.38), we obtain that

$$\begin{aligned} \xi_2^k &= -x_2^k \\ &\quad +\hat{H}_k^{-1}\Pi((B_2[x^k])^{\mathrm{T}}(\mathcal{B}_k^*)^{-1}(I_n - \Pi)\omega_1(x^k, \lambda^k, \xi^k) - \omega_1(x^k, \lambda^k, \xi^k)) \\ &\quad +o(\|x^k\|) \\ &= -x_2^k + O(\|x^k\|\|\xi^k\|) + o(\|\xi^k\|) + o(\|x^k\|) \\ &= -x_2^k + o(\|x^k\|) \end{aligned} \tag{7.40}$$

as $k \to \infty$, giving (7.37).

To conclude the proof (now, for both sets of assumptions on ω_1^k), combining (7.30), (7.37), and (7.38), we obtain that

$$\begin{aligned} x^{k+1} &= x^k + \xi^k \\ &= (x_1^k + \xi_1^k) + (x_2^k + \xi_2^k) \\ &= \tilde{M}_k\xi_2^k + (x_2^k + \xi_2^k) + O(\|x^k\|^2) \\ &= o(\|x^k\|) \end{aligned}$$

as $k \to \infty$, which means the superlinear rate of convergence of $\{x^k\}$ to $\bar{x}$. $\square$

We can further derive from (7.28), (7.30), and (7.38) that

$$\frac{1}{2}PB_2[x^k, x^k] = PB_2[x^k, x^{k+1}].$$

Combined with the superlinear convergence of $\{x^k\}$ to $\bar{x} = 0$, which holds in the case of dual convergence to a noncritical multiplier, this implies the additional estimate

$$PB_2[x^k, x^k] = o(\|x^k\|^2),$$

or, coming back to the original notation,

$$\pi_{(\mathrm{im}\, h'(0))^\perp} h''(0)[x^k, x^k] = o(\|x^k\|^2).$$

This estimate is the extension of (7.15) to the case beyond the full degeneracy; it means that, generally, the primal trajectory approaches $\bar{x}$ tangentially to the null set of the quadratic mapping associated with $\pi_{(\mathrm{im}\, h'(\bar{x}))^\perp} h''(\bar{x})$.

7.1.2 Equality and Inequality-Constrained Problems

We now discuss how considerations for the equality-constrained problems in the previous section can be extended to the case when there are also inequality constraints. Consider the mathematical programming problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \ g(x) \leq 0, \end{array} \tag{7.41}$$

where the objective function $f : \mathbf{R}^n \to \mathbf{R}$ and the constraints mappings $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are twice differentiable.

Stationary points of problem (7.41) and the associated Lagrange multipliers are characterized by the KKT optimality system

$$\frac{\partial L}{\partial x}(x, \lambda, \mu) = 0, \quad h(x) = 0, \quad \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu, g(x) \rangle = 0, \tag{7.42}$$

where $L : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}$ is the Lagrangian of problem (7.41):

$$L(x, \lambda, \mu) = f(x) + \langle \lambda, h(x) \rangle + \langle \mu, g(x) \rangle.$$

Let $\bar{x} \in \mathbf{R}^n$ be a stationary point of problem (7.41), and let $\mathcal{M}(\bar{x})$ be the (nonempty) set of Lagrange multipliers associated with $\bar{x}$, that is, the set of $(\lambda, \mu) \in \mathbf{R}^l \times \mathbf{R}^m$ satisfying (7.42) for $x = \bar{x}$. According to the terminology introduced in Sect. 1.2.4, the set $\mathcal{M}(\bar{x})$ is not a singleton if and only if it violates the SMFCQ (in which case it violates the stronger LICQ as well).

Possible scenarios of dual behavior of Newton-type algorithms on problems with inequality constraints and nonunique Lagrange multipliers, which we discuss next following [144], are somewhat more diverse than for problems with equality constraints only. However, the effect of attraction to critical multipliers still exists, especially when the set of indices of inequality constraints active in the subproblems stabilizes from some iteration on. And, as before, this does result in slow primal convergence.

From (7.42) it can be immediately derived that if $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ satisfies $\bar{\mu}_{A(\bar{x}) \setminus A} = 0$ for some index set $A \subset A(\bar{x})$, then $\bar{x}$ is a stationary point of the equality-constrained problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \ g_A(x) = 0, \end{array} \tag{7.43}$$

while $(\bar{\lambda}, \bar{\mu}_A)$ is an associated Lagrange multiplier.

It makes sense and is convenient to complement Definition 1.41 by the following notion.

Definition 7.8. A Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ is called *critical with respect to an index set* $A \subset A(\bar{x})$ if $\bar{\mu}_{A(\bar{x}) \setminus A} = 0$ and $(\bar{\lambda}, \bar{\mu}_A)$ is a critical

Lagrange multiplier associated with the stationary point $\bar{x}$ of problem (7.43). Otherwise $(\bar{\lambda}, \bar{\mu})$ is called *noncritical with respect to* A.

From Definition 1.41 one can easily see that if the multiplier $(\bar{\lambda}, \bar{\mu})$ is critical, then there exists $I_1 \subset A_0(\bar{x}, \bar{\mu})$ such that this multiplier is critical with respect to the index set $A = A_+(\bar{x}, \bar{\mu}) \cup I_1$.

Recall that $\bar{x}$ and $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ are said to satisfy the SOSC if

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in C(\bar{x}) \setminus \{0\}, \tag{7.44}$$

where

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A(\bar{x})}(\bar{x})\xi \leq 0,\ \langle f'(\bar{x}), \xi\rangle \leq 0\}$$

is the critical cone of problem (7.41) at $\bar{x}$. Recall further that according to Lemma 1.17, any $\xi \in \mathbf{R}^n$ satisfying the equalities $h'(\bar{x})\xi = 0$, $g'_{A(\bar{x})}(\bar{x})\xi = 0$ belongs to $C(\bar{x})$. Therefore, if the SOSC (7.44) holds, then the stationary point $\bar{x}$ and the associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}_{A(\bar{x})})$ satisfy the SOSC for problem (7.43) with $A = A(\bar{x})$. Therefore, $(\bar{\lambda}, \bar{\mu})$ cannot be a critical multiplier of the original problem (7.41) with respect to $A(\bar{x})$ in that case.

Furthermore, $\bar{x}$ and $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ are said to satisfy the SSOSC if

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in C_+(\bar{x}, \bar{\mu}) \setminus \{0\}, \tag{7.45}$$

where

$$C_+(\bar{x}, \bar{\mu}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x})\xi = 0\}.$$

It is easy to see that in the latter case, for any index set $A \subset A(\bar{x})$ such that $\bar{\mu}_{A(\bar{x})\setminus A} = 0$, the stationary point $\bar{x}$ and the associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}_A)$ satisfy the SOSC for problem (7.43). Therefore, $(\bar{\lambda}, \bar{\mu})$ cannot be a critical multiplier of the original problem (7.41) with respect to A in that case. At the same time, if the SOSC (7.44) holds but the SSOSC (7.45) does not, there may exist index sets $A \subset A(\bar{x})$ $(A \neq A(\bar{x}))$ such that $(\bar{\lambda}, \bar{\mu})$ is critical with respect to this A (see Example 7.33 below and its discussion on p. 521).

Of course, the SOSC (7.44) and the SSOSC (7.45) are equivalent in the case of strict complementarity, i.e., when $\bar{\mu}_{A(\bar{x})} > 0$.

Consider the basic SQP method, that is, Algorithm 4.13 with H_k chosen as the exact Hessian of the Lagrangian with respect to the primal variable at the current primal-dual iterate $(x^k, \lambda^k, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$. The method computes the next primal iterate $x^{k+1} \in \mathbf{R}^n$ as a stationary point of the quadratic programming problem

$$
\begin{aligned}
&\text{minimize} \quad \langle f'(x^k),\, x - x^k\rangle + \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k)(x - x^k),\, x - x^k \right\rangle \\
&\text{subject to} \quad h(x^k) + h'(x^k)(x - x^k) = 0,\ g(x^k) + g'(x^k)(x - x^k) \leq 0,
\end{aligned}
\tag{7.46}
$$

and the next dual iterate $(\lambda^{k+1},\, \mu^{k+1}) \in \mathbf{R}^l \times \mathbf{R}^m$ as an associated Lagrange multiplier.

We next recall the discussion in the end of Sect. 4.2. Let the sequence $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ be generated by this SQP method, and suppose that this sequence converges to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, where $\bar{x}$ is a stationary point $\bar{x}$ of problem (7.41), and $(\bar{\lambda},\, \bar{\mu})$ is an associated Lagrange multiplier. It is not unnatural to assume that the set

$$
A_k = \{i = 1,\, \ldots,\, m \mid g_i(x^k) + \langle g_i'(x^k),\, x^{k+1} - x^k\rangle = 0\}
$$

of indices of inequality constraints of the subproblem (7.46) active at x^k is the same for all k sufficiently large. For one thing, this is actually automatic when $(\bar{\lambda},\, \bar{\mu})$ satisfies the strict complementarity condition. In other cases, the assumption that the set A_k is asymptotically unchanged may not hold, but this still seems to be a reasonable numerical behavior, which is certainly not unusual. Note also that if this stabilization property does not hold, one should hardly expect convergence of the dual trajectory, in general. Assuming that $A_k = A$ for all k large enough, we can interpret the SQP iterations essentially as those of the Newton–Lagrange method for the equality-constrained problem (7.43). Moreover, it necessarily holds that $A \subset A(\bar{x})$, $\bar{x}$ is a stationary point of problem (7.43), and $(\bar{\lambda},\, \bar{\mu}_A)$ is an associated Lagrange multiplier. Therefore, the SQP method inherits properties of the Newton–Lagrange method for problem (7.43), including attraction to multipliers critical with respect to A (when they exist) and its effect on the convergence rate.

We next discuss separately the following three possibilities: nonconvergence of the dual sequence; its convergence to a multiplier noncritical with respect to a stabilized active index set; convergence to a critical multiplier.

Scenario 1: Dual trajectory does not converge. According to numerical experience, and as expected, this usually leads to slow primal convergence. How likely is this scenario? In our experience convergence of the dual sequence is far more typical than nonconvergence. However, for equality-constrained problems, when the regularity condition does not hold and there are no critical multipliers associated with $\bar{x}$, this behavior is at least possible (if not to say typical).

One natural question arising in the presence of inequality constraints is the following: Can the dual trajectory be nonconvergent when the MFCQ holds at the primal limit? The positive answer to this question is actually very obvious if one would allow for completely artificial examples (e.g., with duplicated constraints), and assume further that the QP-solver being used (the algorithm for solving quadratic programming subproblems) can pick up any multiplier of the SQP subproblem (7.46) when there are many. Indeed, consider the problem with two identical inequality constraints. These constraints can satisfy or violate the MFCQ but, in any case, the two constraints of the SQP subproblem (7.46) will always be identical. Hence, the multipliers associated with a stationary point of this subproblem will normally not be unique. Then, by picking up appropriate multipliers, one can enforce any kind of limiting behavior. This is one reason why, in this discussion, we should probably restrict ourselves to the case when the multipliers of subproblems are uniquely defined, so that we could say that the sequence $\{(x^k,\, \lambda^k,\, \mu_A^k)\}$ is uniquely defined by the Newton–Lagrange steps for (7.43). In addition, in practice, QP-solvers pick up the multipliers (when they are not unique) not arbitrarily but according to some internal computational patterns. Thus, when dealing with the case of nonunique multipliers of subproblems one should take these rules into account.

For these reasons, we feel that the answer to the question above should probably result not from consideration of artificial examples but from computational practice, and be restricted to specific implementations of algorithms rather than algorithms themselves.

Scenario 2: Dual trajectory converges to a multiplier which is noncritical with respect to the stabilized index set A. This situation is typical if the constraints of problem (7.43) are regular at $\bar{x}$, i.e.,

$$\operatorname{rank}\begin{pmatrix} h'(\bar{x}) \\ g_A'(\bar{x}) \end{pmatrix} = l + |A|. \tag{7.47}$$

The latter, in turn, may happen when the set A is strictly smaller than $A(\bar{x})$, even when the constraints of the original problem (7.41) violate the SMFCQ (and even more so the LICQ). In this case, the multiplier $(\bar{\lambda},\, \bar{\mu}_A)$ associated with stationary point $\bar{x}$ of problem (7.43) is unique, convergence of the Newton–Lagrange method to this unique multiplier can be expected, and this multiplier has no special reason to be critical. According to the theory presented in Sect. 4.1.1, both primal-dual and primal convergence rates are superlinear in this case. Recall, however, that the set A can be strictly smaller than $A(\bar{x})$ only when the limiting multiplier $(\bar{\lambda},\, \bar{\mu})$ violates the strict complementarity condition.

The problem in the next example is taken from [269].

Example 7.9. Let $n = m = 2$, $l = 0$,

$$f(x) = x_1, \quad g(x) = ((x_1-2)^2 + x_2^2 - 4,\, (x_1-4)^2 + x_2^2 - 16).$$

Problem (7.41) with this data has the unique solution $\bar{x} = 0$, with the associated multiplier set $\mathcal{M}(0) = \{\mu \in \mathbf{R}^2 \mid \mu_1 + 2\mu_2 = 1/4,\ 0 \leq \mu_2 \leq 1/8\}$. This solution satisfies the MFCQ (but not the LICQ), and the SSOSC (7.45) holds with any $\bar{\mu} \in \mathcal{M}(\bar{x})$. Therefore, there are no critical multipliers with respect to any index set $A \subset A(0) = \{1, 2\}$.

It can be seen that for this problem the sequence $\{(x^k, \mu^k)\}$ generated by the SQP method converges to $(0, \bar{\mu})$ with $\bar{\mu} = (0, 1/8)$, and the rate of convergence is superlinear. Moreover, the stabilization condition holds with $A = \{2\}$, the unique constraint of the corresponding equality-constrained problem (7.43) is regular at the stationary point 0 of this problem, and the unique associated Lagrange multiplier is $\bar{\mu}_2 = 1/8$. In particular, the limiting multiplier $\bar{\mu} = (0, 1/8)$ does not satisfy the strict complementary condition, and it is noncritical with respect to the given A.

If the condition (7.47) does not hold but $(\bar{\lambda}, \bar{\mu}_A)$ is a noncritical multiplier for (7.43), then according to Theorem 7.7, the primal convergence rate is still superlinear. However, according to the discussion in Sect. 7.1.1, in the case of violation of (7.47) (in particular, if $A = A(\bar{x})$ and the LICQ does not hold at $\bar{x}$ for the original problem (7.41)), convergence to a multiplier which is noncritical with respect to A is highly unlikely to occur.

Scenario 3: Dual trajectory converges to a critical multiplier. As discussed in Sect. 7.1.1, for equality-constrained problems this scenario appears to be typical, unless critical multipliers do not exist. More generally, if the limiting multiplier $(\bar{\lambda}, \bar{\mu})$ satisfies the strict complementarity condition, then $A = A(\bar{x})$, and therefore, according to the discussion above, if the LICQ does not hold at $\bar{x}$, then it is natural to expect the multiplier $(\bar{\lambda}, \bar{\mu})$ to be critical with respect to A and convergence to be slow.

Consider the following inequality-constrained counterpart of Example 7.1, taken from [144].

Example 7.10. Let $n = m = 1$, $l = 0$, $f(x) = -x^2$, $g(x) = x^2$. Problem (7.41) with this data has the unique feasible point (hence, unique solution) $\bar{x} = 0$, with $\mathcal{M}(0) = \mathbf{R}_+$. This solution violates the MFCQ but satisfies the SOSC (7.44) for any $\bar{\mu} > 1$ (and hence, the SSOSC (7.45) also holds for such multipliers, since they satisfy the strict complementarity condition).

It is easy to check that the SQP step in this case is given by $x^{k+1} = x^k/2$, $\mu^{k+1} = 1 - (1 - \mu^k)/2$, and $\{\mu^k\}$ converges to the strictly complementary multiplier $\bar{\mu} = 1$, which is the unique critical with respect to $A = A(\bar{x}) = \{1\}$ multiplier. Convergence rate is only linear.

The next example taken from [149] (the problem in this examples was also considered in [27, Example 4.23]) demonstrates that this scenario is possible even when the MFCQ holds at the primal limit $\bar{x}$.

Example 7.11. Let $n = m = 2$, $l = 0$,

$$f(x) = -x_1, \quad g(x) = (x_1 - x_2^2,\ x_1 + x_2^2).$$

Problem (7.41) with this data has the unique solution $\bar{x} = 0$, with the multiplier set $\mathcal{M}(0) = \{\mu \in \mathbf{R}^2_+ \mid \mu_1 + \mu_2 = 1\}$. This solution satisfies the MFCQ (but not the LICQ), and the SSOSC (7.45) holds for $\bar{\mu} \in \mathcal{M}(0)$ satisfying the inequality $\bar{\mu}_1 < \bar{\mu}_2$.

The SQP subproblem (7.46) takes the form

$$\begin{array}{ll} \text{minimize} & -(x_1 - x_1^k) - (\mu_1^k - \mu_2^k)(x_2 - x_2^k)^2 \\ \text{subject to} & x_1^k - (x_2^k)^2 + (x_1 - x_1^k) - 2x_2^k(x_2 - x_2^k) \leq 0, \\ & x_1^k + (x_2^k)^2 + (x_1 - x_1^k) + 2x_2^k(x_2 - x_2^k) \leq 0. \end{array}$$

Let, for simplicity, $x_1^k = 0$, $x_2^k \neq 0$, and suppose that μ^k is close enough to $\mathcal{M}(0)$. The primal SQP step is given by $x^{k+1} = x^k/2 = (0,\, x_2^k/2)$. In particular, both constraints of the SQP subproblem (7.46) remain active along the primal trajectory, and hence, $A = A(\bar{x}) = \{1,\, 2\}$.

The multiplier of SQP subproblem is given by

$$\mu^{k+1} = \left(\frac{1}{2} + \frac{1}{4}(\mu_1^k - \mu_2^k),\, \frac{1}{2} - \frac{1}{4}(\mu_1^k - \mu_2^k)\right).$$

It follows that $\{\mu^k\} \to (1/2,\, 1/2)$, and the limiting multiplier satisfies the strict complementarity condition, and it is the unique critical multiplier with respect to $A = \{1,\, 2\}$. Again, convergence rate is only linear.

Our final example demonstrates how this scenario can take place with limiting multiplier violating the strict complementarity condition. This problem comes from [271] and was also discussed in [146].

Example 7.12. Let $n = m = 2$, $l = 0$,

$$f(x) = x_1, \quad g(x) = (-x_1,\, (x_1 - 2)^2 + x_2^2 - 4).$$

Problem (7.41) with this data has the unique solution $\bar{x} = 0$, with the multiplier set $\mathcal{M}(0) = \{\mu \in \mathbf{R}_+^2 \mid \mu_1 = 1 - 4\mu_2,\ 0 \leq \mu_2 \leq 1/4\}$. This solution satisfies the MFCQ (but not the LICQ), and the SOSC (7.44) (and even the SSOSC (7.45)) holds with any $\bar{\lambda} \in \mathcal{M}(\bar{x})$, except for $\bar{\mu} = (1,\, 0)$, which is the unique critical multiplier with respect to $A = \{1,\, 2\}$, violating the strict complementarity condition.

The SQP subproblem (7.46) takes the form

$$\begin{array}{ll} \text{minimize} & (x_1 - x_1^k) + \mu_2^k((x_1 - x_1^k)^2 + (x_2 - x_2^k)^2) \\ \text{subject to} & x_1 \leq 0, \\ & (x_1^k - 2)^2 + (x_2^k)^2 - 4 + 2(x_1^k - 2)(x_1 - x_1^k) + 2x_2^k(x_2 - x_2^k) \leq 0. \end{array}$$

The first primal SQP step gives point with $x_1^1 = 0$. Further steps are given by $x^{k+1} = x^k/2 = (0,\, x_2^k/2)$. In particular, both constraints of the SQP subproblem (7.46) remain active along the primal trajectory; hence, it holds that $A = A(\bar{x}) = \{1,\, 2\}$.

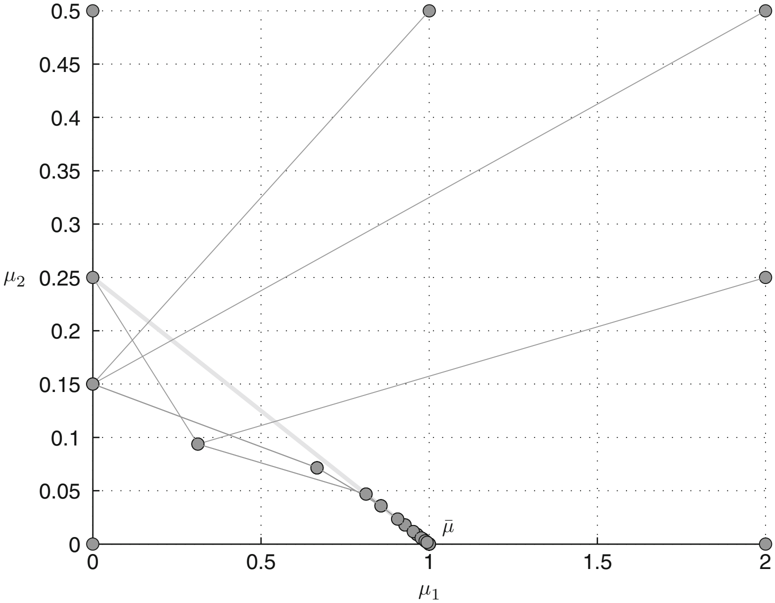

Fig. 7.9 Dual trajectories in Example 7.12 for $x^0 = (1,\, 2)$

The multiplier of the SQP subproblem is given by

$$\mu^{k+1} = (1 - 2\mu_2^k,\ \mu_2^k/2).$$

It follows that $\{\mu^k\} \to \bar{\mu} = (1,\, 0)$.

Figure 7.9 presents the set of multipliers (thick gray line) and some SQP dual trajectories generated starting from $x^0 = (1,\, 2)$. As expected, convergence rate is only linear.

The discussion above readily extends to methods in the perturbed SQP framework (such as quasi-Newton SQP and the linearly constrained Lagrangian methods) when their iterations asymptotically reduce to those of the perturbed Newton–Lagrange method for some underlying equality-constrained problem. See [149] for details. All the considerations also extend to general KKT systems, not necessarily related to optimization problems; in particular, to variational problems with feasible sets given by equality and inequality constraints.

Finally, one might try to analyze along the same lines possible scenarios of dual behavior of semismooth Newton methods applied to equation reformulations of the KKT system (7.42). For instance, as demonstrated in Sect. 3.2.2 (see the discussion around (3.78)), if the reformulation employs the natural residual complementarity function, iteration of the semismooth Newton method for this reformulation can be viewed as the iteration of the Newton–Lagrange method applied to the equality-constrained problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0,\ g_{A_+(\bar{x},\,\bar{\mu})\cup I_1^k}(x) = 0, \end{array}$$

where $(I_1^k,\, I_2^k)$ is some partition of $A_0(\bar{x},\, \bar{\mu})$. If $(I_1^k,\, I_2^k)$ remains unchanged for all k large enough, then the semismooth Newton method reduces to the Newton–Lagrange method for the corresponding fixed equality-constrained problem, and all the previous discussion based on reduction to the equality-constrained case readily applies.

However, according to our experience, stabilization of $(I_1^k,\, I_2^k)$ in the semismooth Newton method appears to be not as natural as stabilization of A_k in the SQP algorithm, unless we assume that the dual sequence converges to a multiplier satisfying the strict complementarity condition. In Example 7.9, dual convergence of the semismooth Newton method fails precisely because the partitions $(I_1^k,\, I_2^k)$ do not stabilize.

For further examples and numerical results demonstrating the effect of attraction to critical Lagrange multipliers, we address the reader to [144, 146, 149].

7.2 Special Methods for Degenerate Problems

The previous section puts in evidence that for KKT systems with nonunique Lagrange multipliers associated with a given primal solution, the superlinear convergence rate of Newton-type methods should not be expected. This often happens because of convergence of the dual sequence to a critical Lagrange multiplier, which is the typical scenario when critical multipliers exist. This state of affairs calls for development of special modifications of Newtonian methods, intended to suppress the attraction to critical multipliers. We describe some approaches to tackle this issue. We start with the stabilized Josephy–Newton framework for generalized equations [85], suitable for dealing with nonisolated solutions. This framework is then used to establish local superlinear convergence of the SQP method for optimization, and for its extension to variational problems. Some alternative approaches for dealing with degeneracy are also discussed.

7.2.1 Stabilized Josephy–Newton Method

In this section we present a very general iterative framework developed in [85]. Consider the generalized equation (GE)

$$\Phi(u) + N(u) \ni 0, \tag{7.48}$$

where $\Phi : \mathbf{R}^\nu \to \mathbf{R}^\nu$ is a continuous (single-valued) mapping, and $N(\cdot)$ is a set-valued mapping from $\mathbf{R}^\nu$ to the subsets of $\mathbf{R}^\nu$. Unlike in the discussion of the Josephy–Newton method in Sect. 3.1, here we are mostly concerned with the case when (7.48) may have nonisolated solutions. Let $\bar{U}$ be the solution set of the GE (7.48). In what follows, we assume that this set is closed.

Consider the class of methods that, given the current iterate $u^k \in \mathbf{R}^\nu$, generate the next iterate u^{k+1} as a solution of the subproblem of the form

$$\mathcal{A}(u^k,\, u) + N(u) \ni 0, \tag{7.49}$$

where for any $\tilde{u} \in \mathbf{R}^\nu$ the mapping $\mathcal{A}(\tilde{u},\, \cdot)$ is some kind of approximation of Φ around $\tilde{u}$. The properties of this approximation are specified in assumption (ii) of Theorem 7.13 below; these properties are related to those of first-order approximation in the path-search approach in Sect. 5.2. In principle, $\mathcal{A}(\tilde{u},\, \cdot)$ can be a set-valued mapping from $\mathbf{R}^\nu$ to the subsets of $\mathbf{R}^\nu$, though in the applications discussed below it will be single-valued.

Note that if Φ is smooth and $\mathcal{A}(\tilde{u},\, u) = \{\Phi(\tilde{u}) + \Phi'(\tilde{u})(u - \tilde{u})\}$, $\tilde{u},\, u \in \mathbf{R}^\nu$, the iteration subproblem (7.49) becomes that of the Josephy–Newton method (see Sect. 3.1). Therefore, the Josephy–Newton method formally fits the framework of (7.49). However, to deal with possibly nonisolated solutions the localization requirement for the iterates in the stabilized method, which we introduce next, is different from the one in the usual Josephy–Newton scheme.

For each $\tilde{u} \in \mathbf{R}^\nu$ define the set

$$U(\tilde{u}) = \{u \in \mathbf{R}^\nu \mid \mathcal{A}(\tilde{u},\, u) + N(u) \ni 0\}, \tag{7.50}$$

so that $U(u^k)$ is the solution set of the iteration subproblem (7.49). As usual, we have to specify which of the solutions of (7.49) are allowed to be the next iterate (solutions "far away" must clearly be discarded from local analysis). In other words, we have to restrict the distance from the current iterate u^k to the next one, i.e., to an element of $U(u^k)$ that can be declared to be u^{k+1}. In the nondegenerate case, it was sufficient to assume that

$$\|u^{k+1} - u^k\| \leq \delta, \tag{7.51}$$

where $\delta > 0$ is fixed and small enough (see Theorem 3.2). In the more subtle/difficult degenerate case, more control is needed; specifically, we shall need to drive δ to zero along the iterations, and at a certain speed. This is the essence of the stabilization mechanism. Specifically, for an arbitrary but fixed $c > 0$ define the subset of the solution set of the subproblem (7.49) by

$$U_c(\tilde{u}) = \{u \in U(\tilde{u}) \mid \|u - \tilde{u}\| \leq c\, \mathrm{dist}(\tilde{u},\, \bar{U})\}, \tag{7.52}$$

and consider the iterative scheme

$$u^{k+1} \in U_c(u^k), \quad k = 0,\, 1,\, \ldots. \tag{7.53}$$

As semistability, which is the key assumption in Theorem 3.2 on the convergence of the Josephy–Newton method, subsumes that the solution in question is isolated, to deal with possibly nonisolated solutions we need to appropriately relax this notion. This is done in assumption (i) of the following theorem. On the other hand, assumption (iii) is a natural counterpart of hemistability, considering that the Josephy–Newton iteration is now replaced by the subproblem (7.49).

Theorem 7.13. *Let a mapping $\Phi : \mathbf{R}^\nu \to \mathbf{R}^\nu$ be continuous in a neighborhood of $\bar{u} \in \mathbf{R}^\nu$, and let $N(\cdot)$ be a set-valued mapping from $\mathbf{R}^\nu$ to the subsets of $\mathbf{R}^\nu$. Let $\bar{U}$ be the solution set of the GE (7.48), let $\bar{u} \in \bar{U}$, and assume that there exists $\bar{\varepsilon} > 0$ such that the set $\bar{U} \cap B(\bar{u}, \bar{\varepsilon})$ is closed. Let $\mathcal{A}$ be a set-valued mapping from $\mathbf{R}^\nu \times \mathbf{R}^\nu$ to the subsets of $\mathbf{R}^\nu$. Assume that the following properties hold with some fixed $c > 0$:*

(i) (Upper Lipschitzian behavior of solutions under canonical perturbations) For any $r \in \mathbf{R}^\nu$ close enough to 0, any solution $u(r)$ of the perturbed GE

$$\Phi(u) + N(u) \ni r \tag{7.54}$$

close enough to $\bar{u}$ satisfies the estimate

$$\operatorname{dist}(u(r), \bar{U}) = O(\|r\|)$$

as $r \to 0$.

(ii) (Precision of approximation) There exists a function $\omega : \mathbf{R}_+ \to \mathbf{R}_+$ such that $\omega(t) = o(t)$ as $t \to 0$ and the estimate

$$\sup\{\|w\| \mid w \in \Phi(u) - \mathcal{A}(\tilde{u}, u),\ \|u - \tilde{u}\| \le c \operatorname{dist}(\tilde{u}, \bar{U})\} \le \omega(\operatorname{dist}(\tilde{u}, \bar{U})) \tag{7.55}$$

holds for all $\tilde{u} \in \mathbf{R}^\nu$ close enough to $\bar{u}$.

(iii) (Solvability of subproblems and localization condition) For any $\tilde{u} \in \mathbf{R}^\nu$ close enough to $\bar{u}$ the set $U_c(\tilde{u})$ defined by (7.50), (7.52) is nonempty.

Then for any starting point $u^0 \in \mathbf{R}^\nu$ close enough to $\bar{u}$, there exists a sequence $\{u^k\} \subset \mathbf{R}^\nu$ satisfying (7.53); any such sequence converges to some $u^* \in \bar{U}$, and it holds that

$$u^{k+1} - u^* = O(\omega(\operatorname{dist}(u^k, \bar{U}))), \tag{7.56}$$

$$\operatorname{dist}(u^{k+1}, \bar{U}) = O(\omega(\operatorname{dist}(u^k, \bar{U}))) \tag{7.57}$$

as $k \to \infty$. In particular, the rates of convergence of $\{u^k\}$ to u^* and of $\{\operatorname{dist}(u^k, \bar{U})\}$ to zero are superlinear. Moreover, those rates of convergence are quadratic if $\omega(t) = O(t^2)$ as $t \to 0$. In addition, for any $\varepsilon > 0$ it holds that $\|u^* - \bar{u}\| < \varepsilon$ provided u^0 is close enough to $\bar{u}$.

Proof. According to (7.50), (7.52), for every $\tilde{u} \in \mathbf{R}^\nu$, any $u \in U_c(\tilde{u})$ satisfies the GE (7.54) with some

$$r \in \Phi(u) - \mathcal{A}(\tilde{u}, u).$$

Moreover, for any $\varepsilon > 0$ it holds that $\|u - \bar{u}\| < \varepsilon$ provided $\tilde{u}$ is close enough to $\bar{u}$. Therefore, for $\tilde{u}$ close enough to $\bar{u}$, assumption (ii) implies the estimate

$$\|r\| \leq \omega(\operatorname{dist}(\tilde{u}, \bar{U})).$$

Hence, assumptions (i) and (iii) imply the existence of $\varepsilon \in (0, \bar{\varepsilon}]$ and $M > 0$ such that

$$U_c(\tilde{u}) \neq \emptyset, \quad \operatorname{dist}(u, \bar{U}) \leq M\omega(\operatorname{dist}(\tilde{u}, \bar{U})) \quad \forall\, u \in U_c(\tilde{u}),\ \forall\, \tilde{u} \in B(\bar{u}, \varepsilon). \tag{7.58}$$

Observe also that since $\varepsilon \leq \bar{\varepsilon}$, the set $\bar{U} \cap B(\bar{u}, \varepsilon)$ is closed.

Employing the assumption $\omega(t) = o(t)$ and further decreasing ε if necessary, we can ensure the property

$$\omega(t) \leq \frac{t}{2M} \quad \forall\, t \in [0, \varepsilon]. \tag{7.59}$$

Set

$$\delta = \frac{\varepsilon}{2c + 1}. \tag{7.60}$$

We now prove by induction that if $u^0 \in B(\bar{u}, \delta)$, then there exists a sequence $\{u^k\} \subset \mathbf{R}^\nu$ satisfying (7.53); moreover, for any such sequence it holds that $u^k \in B(\bar{u}, \varepsilon)$ for all k.

Indeed, since (7.60) evidently implies the inequality $\delta \leq \varepsilon$, we have that $u^0 \in B(\bar{u}, \varepsilon)$. Suppose that there exist $u^j \in \mathbf{R}^n$, $j = 0, 1, \ldots, k$, such that

$$u^j \in U_c(u^{j-1}) \cap B(\bar{u}, \varepsilon) \quad \forall\, j = 1, \ldots, k.$$

Then, $\operatorname{dist}(u^j, \bar{U}) \leq \|u^j - \bar{u}\| \leq \varepsilon$ for all $j = 0, 1, \ldots, k$, and according to (7.58) and (7.59), there exists $u^{k+1} \in U_c(u^k)$, and it holds that

$$\operatorname{dist}(u^{j+1}, \bar{U}) \leq M\omega(\operatorname{dist}(u^j, \bar{U})) \leq \frac{1}{2}\operatorname{dist}(u^j, \bar{U}) \quad \forall\, j = 0, 1, \ldots, k.$$

Therefore, taking into account the inequality in (7.52), we obtain that

$$\begin{aligned}
\|u^{k+1} - u^0\| &\leq \sum_{j=0}^{k} \|u^{j+1} - u^j\| \\
&\leq \sum_{j=0}^{k} c \operatorname{dist}(u^j, \bar{U}) \\
&\leq c \sum_{j=0}^{k} \frac{1}{2^j} \operatorname{dist}(u^0, \bar{U})
\end{aligned}$$

$$
\begin{aligned}
&= c\left(2 - \frac{1}{2^k}\right) \operatorname{dist}(u^0,\, \bar{U}) \\
&\leq 2c \operatorname{dist}(u^0,\, \bar{U}) \\
&\leq 2c\|u^0 - \bar{u}\| \\
&\leq 2c\delta. \qquad (7.61)
\end{aligned}
$$

Hence,

$$\|u^{k+1} - \bar{u}\| \leq \|u^{k+1} - u^0\| + \|u^0 - \bar{u}\| \leq (2c + 1)\delta = \varepsilon,$$

where the last equality is by (7.60). Therefore, $u^{k+1} \in B(\bar{u},\, \varepsilon)$, which completes the induction argument.

Furthermore, according to (7.58) and (7.59), and to the assertion just established, the relations

$$\operatorname{dist}(u^{k+1},\, \bar{U}) \leq M\omega(\operatorname{dist}(u^k,\, \bar{U})) \leq \frac{1}{2} \operatorname{dist}(u^k,\, \bar{U}) \quad \forall\, k \qquad (7.62)$$

are valid for any $u^0 \in B(\bar{u},\, \delta)$ and any sequence $\{u^k\} \subset \mathbf{R}^\nu$ satisfying (7.53). In particular, the sequence $\{\operatorname{dist}(u^k,\, \bar{U})\}$ converges to zero, and the estimate (7.57) holds.

Similarly to (7.61), one can establish the estimate

$$\|u^{k+j} - u^k\| \leq 2c \operatorname{dist}(u^k,\, \bar{U}) \quad \forall\, k,\, \forall\, j, \qquad (7.63)$$

where the right-hand side tends to zero as $k \to \infty$. Therefore, $\{u^k\}$ is a Cauchy sequence. Hence, it converges to some $u^* \in \mathbf{R}^\nu$. Moreover, again employing the convergence of $\{\operatorname{dist}(u^k,\, \bar{U})\}$ to zero, together with the inclusion $\{u^k\} \subset B(\bar{u},\, \varepsilon)$ and the closedness of the set $\bar{U} \cap B(\bar{u},\, \varepsilon)$, we conclude that $u^* \in \bar{U} \cap B(\bar{u},\, \varepsilon)$. Since $\varepsilon > 0$ can be taken arbitrarily small (making δ in (7.60) small enough), the last assertion of the theorem follows.

Regarding the convergence rate of $\{u^k\}$, by (7.62) and (7.63), we derive that

$$\|u^{k+j} - u^{k+1}\| \leq 2c \operatorname{dist}(u^{k+1},\, \bar{U}) \leq 2Mc\omega(\operatorname{dist}(u^k,\, \bar{U})) \quad \forall\, k,\, \forall\, j.$$

Passing onto the limit as $j \to \infty$, it follows that

$$\|u^* - u^{k+1}\| \leq 2Mc\omega(\operatorname{dist}(u^k,\, \bar{U})) \quad \forall\, k.$$

This estimate proves (7.56), and since $\operatorname{dist}(u^k,\, \bar{U}) \leq \|u^k - u^*\|$, again employing the assumption $\omega(t) = o(t)$, this gives the superlinear convergence rate of $\{u^k\}$.

If $\omega(t) = O(t^2)$, the quadratic convergence rate of $\{u^k\}$ and of $\{\operatorname{dist}(u^k,\, \bar{U})\}$ follows from (7.56) and (7.57). $\square$

One can see from the proof of Theorem 7.13 that condition (7.55) in assumption (ii) can be replaced by the following weaker one:

$$\sup\{\|w\| \mid w \in \Phi(u) - \mathcal{A}(\tilde{u}, u),\ u \in U_c(\tilde{u})\} \leq \omega(\mathrm{dist}(\tilde{u}, \bar{U})).$$

Since assumption (i) of Theorem 7.13 is weaker than semistability, this theorem can be used to derive a result related to Theorem 3.6, but with the localization condition of the form (7.51) replaced by the following property:

$$\|u^{k+1} - u^k\| \leq c\,\mathrm{dist}(u^k, \bar{U}).$$

Observe that the latter is not needed in the direct proof of Theorem 3.6 in Sect. 3.1.

In what follows, we shall use the stabilized Josephy–Newton framework to derive convergence of the stabilized SQP method for optimization and of its extension to variational problems. It is interesting to note that modifications of this framework are useful also for the analysis of the augmented Lagrangian method [76], which is not Newtonian.

7.2.2 Stabilized Sequential Quadratic Programming and Stabilized Newton Method for Variational Problems

The stabilized version of the sequential quadratic programming algorithm (stabilized SQP) had been developed in order to achieve fast convergence despite possible degeneracy of constraints of optimization problems, when the Lagrange multipliers associated with a solution are not unique. (Recall that the standard SQP method requires the SMFCQ, i.e., uniqueness of the multiplier, as part of the assumptions; see Sect. 4.2.)

Consider the mathematical programming problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \leq 0, \end{array} \tag{7.64}$$

where the objective function $f : \mathbf{R}^n \to \mathbf{R}$ and the constraints mapping $g : \mathbf{R}^n \to \mathbf{R}^m$ are twice differentiable. As no constraint qualification conditions of any kind are needed in the analysis to be presented, there is no loss of generality in considering inequality constraints only (equality constraints can be simply written as inequalities with the opposite signs to extend the theory). At the same time, when there are equality constraints only (no inequality constraints), the general convergence result can be strengthened. For this reason, the equality-constrained case will be considered separately later on.

Let $L : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ be the Lagrangian of problem (7.64):

$$L(x,\, \mu) = f(x) + \langle \mu,\, g(x) \rangle.$$

For a given stationary point $\bar{x} \in \mathbf{R}^n$ of (7.64), let $\mathcal{M}(\bar{x})$ be the (nonempty) set of the associated Lagrange multipliers, i.e., the set of $\mu \in \mathbf{R}^m$ satisfying the KKT optimality system

$$\frac{\partial L}{\partial x}(x,\, \mu) = 0, \quad \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu,\, g(x) \rangle = 0$$

for $x = \bar{x}$. Recall that the SOSC for problem (7.64) means that

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\mu})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{7.65}$$

where

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid g'_{A(\bar{x})}(\bar{x})\xi \leq 0,\ \langle f'(\bar{x}),\, \xi \rangle \leq 0\} \tag{7.66}$$

is the critical cone of problem (7.64) at $\bar{x}$. Recall also that the SSOSC states that

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\mu})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in C_+(\bar{x},\, \bar{\mu}) \setminus \{0\}, \tag{7.67}$$

where

$$C_+(\bar{x},\, \bar{\mu}) = \{\xi \in \mathbf{R}^n \mid g'_{A_+(\bar{x},\, \bar{\mu})}(\bar{x})\xi = 0\}. \tag{7.68}$$

Stabilized SQP method was introduced in [269] in the form of solving the min-max subproblems

$$\begin{array}{ll} \text{minimize} & \displaystyle\max_{\mu \in \mathbf{R}^m_+} \biggl(\langle f'(x^k),\, x - x^k \rangle + \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(x^k,\, \mu^k)(x - x^k),\, x - x^k \right\rangle \\ & \qquad\qquad + \langle \mu,\, g(x^k) + g'(x^k)(x - x^k) \rangle - \dfrac{\sigma_k}{2}\|\mu - \mu^k\|^2 \biggr) \\ \text{subject to} & x \in \mathbf{R}^n, \end{array}$$

where $(x^k,\, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^m_+$ is the current approximation to a primal-dual solution of (7.64), and $\sigma_k > 0$ is the dual stabilization parameter. It can be seen [180] that this min-max problem is equivalent to the following quadratic programming problem in the primal-dual space:

$$\begin{array}{ll} \text{minimize} & \langle f'(x^k),\, x - x^k \rangle + \dfrac{1}{2} \left\langle \dfrac{\partial^2 L}{\partial x^2}(x^k,\, \mu^k)(x - x^k),\, x - x^k \right\rangle + \dfrac{\sigma_k}{2}\|\mu\|^2 \\ \text{subject to} & g(x^k) + g'(x^k)(x - x^k) - \sigma_k(\mu - \mu^k) \leq 0. \end{array} \tag{7.69}$$

Note that for $\sigma_k = 0$, the subproblem (7.69) becomes a subproblem of the usual SQP method (in particular, stated in the primal space only). One immediate observation is that for $\sigma_k > 0$, the constraints in (7.69) have the "elastic mode" feature and are therefore automatically consistent regardless of any constraint qualification or convexity assumptions (e.g., fixing any

$x \in \mathbf{R}^n$ and taking μ with all the components large enough gives points feasible in (7.69)). This is the first major difference from standard SQP methods.

In [269] superlinear convergence of the stabilized SQP method is established under the MFCQ, the SOSC (7.65) assumed for all $\bar{\mu} \in \mathcal{M}(\bar{x})$, the existence of a multiplier $\bar{\mu} \in \mathcal{M}(\bar{x})$ satisfying the strict complementarity condition $\bar{\mu}_{A(\bar{x})} > 0$, and the assumption that the initial dual iterate is close enough to such a multiplier. Also, [269] gives an analysis in the presence of roundoff errors, which is particularly relevant in this context because the conditioning of the stabilized subproblems deteriorates (in the degenerate case) as $\sigma_k \to 0$. In [271], the assumption of strict complementarity has been removed. Also, [270] suggests a certain inexact SQP framework close in nature to a posteriori perturbed SQP method in Sect. 4.3, which includes stabilized SQP as a particular case. The assumptions, however, still contain the MFCQ. In [112], constraint qualifications are not used, at the expense of employing the SSOSC (7.67) instead of the weaker SOSC (7.65) (and assuming that the dual starting point is close to a multiplier satisfying the SSOSC). In [85], the result of [112] was recovered from more general principles, and somewhat sharpened. The iterative framework of [85], presented in Sect. 7.2.1 as a stabilized Josephy–Newton method for generalized equations, was further used in [75] to prove local superlinear convergence assuming the SOSC (7.65) only. Moreover, the method was extended to variational problems. Quasi-Newton versions of stabilized SQP are analyzed under the SOSC in [73]. In [150] it was shown that the SOSC cannot be relaxed when inequality constraints are present, but for equality-constrained problems the weaker condition of noncriticality of the relevant Lagrange multiplier (see Definition 1.41) is sufficient for convergence. Some numerical results for the stabilized SQP method without attempting globalization are reported in [203], and some suggestions concerning globalization can be found in [78] and [99]. However, building really satisfactory globalization techniques for stabilized SQP is a challenging matter, which can still be considered an open question at this time.

We finally note that stabilized SQP for optimization can also be interpreted within the perturbed SQP framework of Sect. 4.3.1. However, recall that the main convergence result in this framework requires the SMFCQ (Theorem 4.24). If the method is interpreted instead via the stabilized Josephy–Newton framework presented in Sect. 7.2.1, no constraint qualifications are needed to prove local convergence.

We next present a stabilized Newton method for variational problems, which contains stabilized SQP for optimization as a special case. Given a differentiable mapping $F : \mathbf{R}^n \to \mathbf{R}^n$ and a twice differentiable mapping $g : \mathbf{R}^n \to \mathbf{R}^m$, consider the variational problem (VP)

$$F(x) + N_D(x) \ni 0, \tag{7.70}$$

where

$$D = \{x \in \mathbf{R}^n \mid g(x) \le 0\}.$$

Recall that the normal cone $N_D(x)$ to the set D at $x \in \mathbf{R}^n$ is defined as follows: $N_D(x) = (T_D(x))^{\circ}$ for $x \in D$, and $N_D(x) = \emptyset$ for $x \notin D$. As already commented, since no constraint qualifications of any kind would be used in the development that follows, there is no loss of generality in omitting the equality constraints in the definition of the feasible set D.

The KKT system for (7.70) in the variables $(x,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^m$ is given by

$$F(x) + (g'(x))^{\mathrm{T}}\mu = 0, \quad \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu,\, g(x)\rangle = 0 \tag{7.71}$$

(see Sect. 1.3.1). If

$$F(x) = f'(x), \quad x \in \mathbf{R}^n, \tag{7.72}$$

then (7.70) represents the primal, and (7.71) the primal-dual, first-order optimality conditions for the optimization problem (7.64).

A natural extension of the stabilized SQP method for optimization to the current variational setting is the following iterative procedure, which is motivated by the first-order variational formulation of optimality conditions for the subproblems (7.69) in the optimization case.

Define the mapping $G : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^n$ by

$$G(x,\, \mu) = F(x) + (g'(x))^{\mathrm{T}}\mu. \tag{7.73}$$

Let $(x^k,\, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^m$ be the current primal-dual approximation to a solution of (7.71), and let $\sigma_k = \sigma(x^k,\, \mu^k)$ be the dual stabilization parameter (the choice of the function $\sigma : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}_+$ will be specified shortly; it is a computable measure of violation of the KKT conditions (7.71) by the point $(x^k,\, \mu^k)$). Define the affine mapping $\Phi_k : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^m$ by

$$\Phi_k(u) = \left(F(x^k) + \frac{\partial G}{\partial x}(x^k,\, \mu^k)(x - x^k),\, \sigma_k\mu\right), \quad u = (x,\, \mu), \tag{7.74}$$

and consider the affine variational inequality (VI) of the form

$$u \in Q_k, \quad \langle \Phi_k(u), v - u\rangle \geq 0 \quad \forall\, v \in Q_k, \tag{7.75}$$

where

$$Q_k = \left\{u = (x,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^m \mid g(x^k) + g'(x^k)(x - x^k) - \sigma_k(\mu - \mu^k) \leq 0\right\}. \tag{7.76}$$

As can be easily seen, in the optimization case (7.72) the VI (7.75) is precisely the first-order (primal) necessary optimality condition for the stabilized SQP subproblem (7.69). Thus this framework contains stabilized SQP for optimization as a special case. Note that the method makes good sense also in the variational setting, as solving the fully nonlinear VP (7.70) is replaced by solving a sequence of affine VIs (7.75) (the mapping Φ_k is affine and the set Q_k is polyhedral).

By the linearity of its constraints, the subproblem (7.75) is equivalent to solving the KKT system of finding $(x, \mu, \tilde{\mu}) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^m$ such that

$$\begin{gathered}F(x^k) + \frac{\partial G}{\partial x}(x^k, \mu^k)(x - x^k) + (g'(x^k))^{\mathrm{T}}\tilde{\mu} = 0, \quad \sigma_k\mu - \sigma_k\tilde{\mu} = 0,\\ \tilde{\mu} \geq 0, \quad g(x^k) + g'(x^k)(x - x^k) - \sigma_k(\mu - \mu^k) \leq 0,\\ \langle \tilde{\mu}, g(x^k) + g'(x^k)(x - x^k) - \sigma_k(\mu - \mu^k)\rangle = 0.\end{gathered}$$

If $\sigma_k > 0$, noting that $\mu = \tilde{\mu}$ by the second relation in the system above, the subproblem (7.75) is then further equivalent to the affine mixed complementarity problem (MCP) of finding $(x, \mu) \in \mathbf{R}^n \times \mathbf{R}^m$ such that

$$\begin{gathered}G(x^k, \mu^k) + \frac{\partial G}{\partial x}(x^k, \mu^k)(x - x^k) + (g'(x^k))^{\mathrm{T}}(\mu - \mu^k) = 0,\\ \mu \geq 0, \quad g(x^k) + g'(x^k)(x - x^k) - \sigma_k(\mu - \mu^k) \leq 0,\\ \langle \mu, g(x^k) + g'(x^k)(x - x^k) - \sigma_k(\mu - \mu^k)\rangle = 0.\end{gathered} \tag{7.77}$$

Making use of this equivalence, much of the analysis that follows of the convergence of Algorithm 7.14 below would actually refer to the affine MCP (7.77).

Let $\bar{x} \in \mathbf{R}^n$ be a given primal solution of the KKT system (7.71) and let $\bar{\mu} \in \mathcal{M}(\bar{x})$, where $\mathcal{M}(\bar{x})$ is the (nonempty) set of Lagrange multipliers associated with $\bar{x}$ (i.e., the set of $\mu \in \mathbf{R}^m$ satisfying (7.71) for $x = \bar{x}$). Let

$$\begin{aligned}C(\bar{x}) &= \{\xi \in \mathbf{R}^n \mid g'_{A(\bar{x})}(\bar{x})\xi \leq 0,\ \langle F(\bar{x}), \xi\rangle = 0\}\\ &= \{\xi \in \mathbf{R}^n \mid g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x})\xi = 0,\ g'_{A_0(\bar{x}, \bar{\mu})}(\bar{x})\xi \leq 0\}\end{aligned} \tag{7.78}$$

be the critical cone of the KKT system (7.71) at $\bar{x}$ (see Sect. 1.3.1; cf. (7.66)). Suppose that at $(\bar{x}, \bar{\mu})$ the following second-order sufficiency condition holds:

$$\left\langle \frac{\partial G}{\partial x}(\bar{x}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}. \tag{7.79}$$

Note that in the optimization setting (7.72), $C(\bar{x})$ is the usual critical cone and (7.79) is the usual SOSC (7.65). It can be seen that (7.79) implies that the primal part $\bar{x}$ of the solutions of the KKT system (7.71) is locally unique [75].

Let $\sigma : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}_+$ be any function with the following error bound property: there exist a neighborhood V of $(\bar{x}, \bar{\mu})$ and constants $\beta_1 > 0$ and $\beta_2 > 0$ such that

$$\begin{gathered}\beta_1(\|x - \bar{x}\| + \mathrm{dist}(\mu, \mathcal{M}(\bar{x}))) \leq \sigma(x, \mu) \leq \beta_2(\|x - \bar{x}\| + \mathrm{dist}(\mu, \mathcal{M}(\bar{x})))\\ \forall\, (x, \mu) \in V.\end{gathered} \tag{7.80}$$

Note that according to Proposition 1.43, under the second-order sufficiency condition (7.79) the natural residual for the KKT system (7.71), defined by

$$\sigma(x, \mu) = \|(G(x, \mu), \min\{\mu, -g(x)\})\|, \tag{7.81}$$

satisfies the left-most inequality in (7.80). The right-most inequality is immediate by the local Lipschitz-continuity of this σ at $(\bar{x}, \bar{\mu})$, which is implied by our assumptions that F and the derivative of g are locally Lipschitz-continuous at $\bar{x}$.

Of course, (7.81) is not the only possible choice of a function σ with the needed properties. In particular, other complementarity functions can be used to define the residual of the KKT system. Nevertheless, to be specific, in the algorithm stated below we use the function defined in (7.81). The *stabilized Newton method* for solving the VP (7.70) is therefore the following.

Algorithm 7.14 Choose $(x^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^m_+$ and set $k = 0$.

1. Compute $\sigma_k = \sigma(x^k, \mu^k)$, where the function $\sigma : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}_+$ is defined in (7.81). If $\sigma_k = 0$, stop.
2. Compute $(x^{k+1}, \mu^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^m$ as a solution of the affine VI (7.75), where the mapping $\Phi_k : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^m$ and the set $Q_k \subset \mathbf{R}^n \times \mathbf{R}^m$ are given by (7.73), (7.74), and (7.76), respectively.
3. Increase k by 1 and go to step 1.

The proof of local convergence of the stabilized Newton method for the VP (7.70) (or for its KKT system (7.71)), whose iteration is described by the affine VI (7.75) (equivalently, by the affine MCP (7.77)), is via verification of the assumptions of Theorem 7.13 on convergence of the stabilized Josephy–Newton method for GEs. To that end, define $\Phi : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^m$ by

$$\Phi(u) = (G(x, \mu), -g(x)), \quad u = (x, \mu), \tag{7.82}$$

and the set-valued mapping $N(\cdot)$ by

$$N(u) = N(\mu) = \left(0, N_{\mathbf{R}^m_+}(\mu)\right).$$

With these definitions, the problem (7.71) is equivalent to the GE (7.48).

Defining also the mapping $\mathcal{A} : (\mathbf{R}^n \times \mathbf{R}^m) \times (\mathbf{R}^n \times \mathbf{R}^m) \to \mathbf{R}^n \times \mathbf{R}^m$,

$$\mathcal{A}(\tilde{u}, u) = \Phi(\tilde{u}) + (\Phi'(\tilde{u}) + M(\tilde{u}))(u - \tilde{u}), \tag{7.83}$$

where the matrix $M(\tilde{u}) \in \mathbf{R}^{(n+m)\times(n+m)}$ is given by

$$M(\tilde{u}) = \begin{pmatrix} 0 & 0 \\ 0 & \sigma(\tilde{x}, \tilde{\mu})I \end{pmatrix}, \quad \tilde{u} = (\tilde{x}, \tilde{\mu}), \tag{7.84}$$

with the identity matrix $I \in \mathbf{R}^{m\times m}$, the subproblem (7.75) (where we take $\sigma_k = \sigma(x^k, \mu^k)$) of the stabilized Newton method for solving (7.71) is equivalent to (7.49) with $u^k = (x^k, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^m$. (Unlike in Sect. 7.2.1, here we define $\mathcal{A}$ as a single-valued mapping, which, however, should not lead to any ambiguity.)

We need to verify the three assumptions of Theorem 7.13 in the present setting. As will be seen in the proof of Theorem 7.21 below, under the second-order sufficiency condition (7.79), assumption (i) on the upper Lipschitzian behavior of the solutions of the KKT system (7.71) under canonical perturbations follows from some general facts, while assumption (ii) on the quality of approximation of $\Phi(\cdot)$ by $\mathcal{A}(\tilde{u}, \cdot)$ can be easily checked directly. The difficult part is to verify assumption (iii), i.e., to prove that the subproblems (7.75) (equivalently, the affine MCP (7.77)) have solutions, and that there are solutions satisfying the localization property (7.52) (i.e., the iterative step is of the order of the distance from the current iterate to the solution set).

We start with extending the second-order sufficiency condition (7.79) from the positivity property of the matrix in question on the critical cone to the positivity (when $\sigma(x, \mu) > 0$) of the augmented matrix on a certain primal-dual cone, uniform for all (x, μ) close enough to $(\bar{x}, \bar{\mu})$.

For any $(x, \mu) \in \mathbf{R}^n \times \mathbf{R}^m$, define the matrix $B(x, \mu) \in \mathbf{R}^{(n+m)\times(n+m)}$,

$$B(x, \mu) = \begin{pmatrix} \dfrac{\partial G}{\partial x}(x, \mu) & 0 \\ 0 & \sigma(x, \mu)I \end{pmatrix}, \tag{7.85}$$

and the cone

$$K(x, \mu) = \left\{ (\xi, \zeta) \in \mathbf{R}^n \times \mathbf{R}^m \;\middle|\; \begin{array}{c} g'_{A_+(\bar{x}, \bar{\mu})}(x)\xi - \sigma(x, \mu)\zeta_{A_+(\bar{x}, \bar{\mu})} = 0, \\ g'_{A_0(\bar{x}, \bar{\mu})}(x)\xi - \sigma(x, \mu)\zeta_{A_0(\bar{x}, \bar{\mu})} \leq 0, \\ \zeta_{\{1, \dots, m\}\setminus A(\bar{x})} = 0 \end{array} \right\}. \tag{7.86}$$

Proposition 7.15. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$, let $g : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x}$, with its second derivative being continuous at $\bar{x}$, and let $G : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^n$ be defined in (7.73). Let $(\bar{x}, \bar{\mu})$ with some $\bar{\mu} \in \mathbf{R}^m$ be a solution of the KKT system (7.71), satisfying the second-order sufficiency condition (7.79), and let $\sigma : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}_+$ be such that $\sigma(x, \mu) \to 0$ as $(x, \mu) \to (\bar{x}, \bar{\mu})$.*

Then there exists a constant $\gamma > 0$ such that for all $(x, \mu) \in \mathbf{R}^n \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\mu})$ it holds that

$$\langle B(x, \mu)(\xi, \zeta), (\xi, \zeta)\rangle \geq \gamma(\|\xi\|^2 + \sigma(x, \mu)\|\zeta\|^2) \quad \forall\, (\xi, \zeta) \in K(x, \mu), \tag{7.87}$$

where $B(x, \mu)$ and $K(x, \mu)$ are defined by (7.85) and (7.86), respectively.

Proof. Suppose the contrary, i.e., that there exist sequences $\{\varepsilon_k\} \subset \mathbf{R}_+$ converging to 0, $\{(x^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^m$ converging to $(\bar{x}, \bar{\mu})$, and a sequence $\{(\xi^k, \zeta^k)\} \subset \mathbf{R}^n \times \mathbf{R}^m$, such that for all k it holds that $(\xi^k, \zeta^k) \in K(x^k, \mu^k)$ and

$$\left\langle \frac{\partial G}{\partial x}(x^k, \mu^k)\xi^k, \xi^k \right\rangle + \sigma_k\|\zeta^k\|^2 < \varepsilon_k(\|\xi^k\|^2 + \sigma_k\|\zeta^k\|^2), \tag{7.88}$$

where $\sigma_k = \sigma(x^k, \mu^k)$. Note that by the assumption above, $\sigma_k \to 0$ as $k \to \infty$.

Evidently, (7.88) subsumes that $(\xi^k, \sqrt{\sigma_k}\zeta^k) \neq 0$ for each index k. Let $t_k = \|(\xi^k, \sqrt{\sigma_k}\zeta^k)\| > 0$. Passing onto a subsequence, if necessary, we can assume that for some $(\bar{\xi}, \bar{\zeta}) \in \mathbf{R}^n \times \mathbf{R}^m$, $(\bar{\xi}, \bar{\zeta}) \neq 0$, it holds that

$$\left\{ \frac{(\xi^k, \sqrt{\sigma_k}\zeta^k)}{t_k} \right\} \to (\bar{\xi}, \bar{\zeta}). \tag{7.89}$$

Observe that since $\sigma_k \to 0$ while $\{\sqrt{\sigma_k}\zeta^k/t_k\}$ is bounded, it holds that

$$\left\{ \sigma_k \frac{\zeta^k}{t_k} \right\} = \left\{ \sqrt{\sigma_k} \frac{\sqrt{\sigma_k}\zeta^k}{t_k} \right\} \to 0. \tag{7.90}$$

Since $K(x^k, \mu^k)$ is a cone, for all k we have that $(\xi^k/t_k, \zeta^k/t_k) \in K(x^k, \mu^k)$. Dividing now the relations in (7.86) (with $(x, \mu) = (x^k, \mu^k)$ and $(\xi, \zeta) = (\xi^k, \zeta^k)$) by $t_k > 0$ and passing onto the limit as $k \to \infty$, taking into account (7.89) and (7.90), we obtain that

$$g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x})\bar{\xi} = 0, \quad g'_{A_0(\bar{x}, \bar{\mu})}(\bar{x})\bar{\xi} \leq 0,$$

and in particular, $\bar{\xi} \in C(\bar{x})$.

On the other hand, dividing both sides of the relation in (7.88) by $t_k^2 > 0$ and taking the limit as $k \to \infty$, employing (7.89) we have that

$$\left\langle \frac{\partial G}{\partial x}(\bar{x}, \bar{\mu})\bar{\xi}, \bar{\xi} \right\rangle + \|\bar{\zeta}\|^2 \leq 0. \tag{7.91}$$

Now, the second-order sufficiency condition (7.79) and (7.91) imply that $\bar{\xi} = 0$ and $\bar{\zeta} = 0$, in contradiction with $(\bar{\xi}, \bar{\zeta}) \neq 0$. $\square$

Corollary 7.16. *Let the assumptions of Proposition 7.15 hold.*

Then for all $(x, \mu) \in \mathbf{R}^n \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\mu})$ and such that $\sigma(x, \mu) > 0$, the matrix

$$\begin{pmatrix} \dfrac{\partial G}{\partial x}(x, \mu) & (g'_{A(\bar{x})}(x))^{\mathrm{T}} \\ g'_{A(\bar{x})}(x) & -\sigma(x, \mu)I \end{pmatrix} \tag{7.92}$$

is nonsingular.

Proof. By Proposition 7.15, the relation (7.87) holds with some $\gamma > 0$ for all $(x, \mu) \in \mathbf{R}^n \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\mu})$. Fix any such (x, μ), satisfying also $\sigma(x, \mu) > 0$. Take any $(\xi, \zeta) \in \mathbf{R}^n \times \mathbf{R}^m$ such that $\zeta_{\{1, \ldots, m\} \setminus A(\bar{x})} = 0$ and $(\xi, \zeta_{A(\bar{x})})$ belongs to the kernel of the matrix given in (7.92), i.e.,

$$\frac{\partial G}{\partial x}(x, \mu)\xi + (g'_{A(\bar{x})}(x))^{\mathrm{T}}\zeta_{A(\bar{x})} = 0, \quad g'_{A(\bar{x})}(x)\xi - \sigma(x, \mu)\zeta_{A(\bar{x})} = 0. \tag{7.93}$$

Note that the second equality in (7.93) shows that $(\xi, \zeta) \in K(x, \mu)$; see the definition (7.86). Also, multiplying both sides in the second equality in (7.93) by $\zeta_{A(\bar{x})}$ we obtain that

$$\langle \zeta_{A(\bar{x})}, g'_{A(\bar{x})}(x)\xi \rangle = \sigma(x, \mu)\|\zeta_{A(\bar{x})}\|^2 = \sigma(x, \mu)\|\zeta\|^2.$$

Multiplying by ξ both sides in the first equality in (7.93) and using (7.85), (7.87), we then obtain that

$$\begin{aligned} 0 &= \left\langle \frac{\partial G}{\partial x}(x, \mu)\xi, \xi \right\rangle + \langle \zeta_{A(\bar{x})}, g'_{A(\bar{x})}(x)\xi \rangle \\ &= \left\langle \frac{\partial G}{\partial x}(x, \mu)\xi, \xi \right\rangle + \sigma(x, \mu)\|\zeta\|^2 \\ &= \langle B(x, \mu)(\xi, \zeta), (\xi, \zeta) \rangle \\ &\geq \gamma(\|\xi\|^2 + \sigma(x, \mu)\|\zeta\|^2). \end{aligned}$$

Hence, $\xi = 0$ and $\zeta = 0$, implying that the matrix in (7.93) is nonsingular. $\square$

The proof of existence of solutions of stabilized subproblems is done in two steps. We first show that a certain part of relations in the affine MCP (7.77) has a solution. To this end, we make use of the existence result in [68, Theorem 2.5.10] for a VI with an affine mapping on a cone. More specifically, we employ an obvious consequence of the cited result, which we state as follows.

Proposition 7.17. *Let $K \subset \mathbf{R}^\nu$ be a closed convex cone and let $B \in \mathbf{R}^{\nu \times \nu}$ be such that*

$$\langle v, Bv \rangle > 0 \quad \forall\, v \in K \setminus \{0\}. \tag{7.94}$$

Then for all $b \in \mathbf{R}^\nu$, the VI

$$u \in K, \quad \langle Bu + b, v - u \rangle \geq 0 \quad \forall\, v \in K \tag{7.95}$$

has a solution.

Proposition 7.18. *Let the assumptions of Proposition 7.15 hold.*

Then for all $(x, \mu) \in \mathbf{R}^n \times \mathbf{R}^m$ close enough to $(\bar{x}, \bar{\mu})$ and such that $\sigma(x, \mu) > 0$, the affine MCP of finding $(\xi, \zeta) \in \mathbf{R}^n \times \mathbf{R}^m$ such that

$$\begin{gathered} G(x, \mu) + \frac{\partial G}{\partial x}(x, \mu)\xi + (g'(x))^{\mathrm{T}}\zeta = 0, \\ g_{A_+(\bar{x}, \bar{\mu})}(x) + g'_{A_+(\bar{x}, \bar{\mu})}(x)\xi - \sigma(x, \mu)\zeta_{A_+(\bar{x}, \bar{\mu})} = 0, \\ (\mu + \zeta)_{A_0(\bar{x}, \bar{\mu})} \geq 0, \quad g_{A_0(\bar{x}, \bar{\mu})}(x) + g'_{A_0(\bar{x}, \bar{\mu})}(x)\xi - \sigma(x, \mu)\zeta_{A_0(\bar{x}, \bar{\mu})} \leq 0, \\ (\mu_i + \zeta_i)(g_i(x) + \langle g'_i(x), \xi \rangle - \sigma(x, \mu)\zeta_i) = 0, \ i \in A_0(\bar{x}, \bar{\mu}), \\ \zeta_{\{1, \ldots, m\} \setminus A(\bar{x})} = 0, \end{gathered} \tag{7.96}$$

has a solution.

Proof. By Proposition 7.15, for all $(x,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^m$ close enough to $(\bar{x},\, \bar{\mu})$ and such that $\sigma(x,\, \mu) > 0$, and for the matrix $B = B(x,\, \mu)$ defined in (7.85) and the cone $K = K(x,\, \mu)$ defined in (7.86), the property (7.94) is valid (see (7.87)). Hence, by Proposition 7.17, for all such $(x,\, \mu)$ it holds that the affine VI (7.95) has a solution for any $b \in \mathbf{R}^n \times \mathbf{R}^m$. The rest of the proof shows that (7.96) can be reduced to the form of (7.95).

Define $\tilde{b} = (G(x,\, \mu),\, 0) \in \mathbf{R}^n \times \mathbf{R}^m$,

$$Q = \left\{ \begin{array}{c} (\xi,\, \zeta) \\ \in \mathbf{R}^n \times \mathbf{R}^m \end{array} \left| \begin{array}{c} g_{A_+(\bar{x},\, \bar{\mu})}(x) + g'_{A_+(\bar{x},\, \bar{\mu})}(x)\xi - \sigma(x,\, \mu)\zeta_{A_+(\bar{x},\, \bar{\mu})} = 0, \\ g_{A_0(\bar{x},\, \bar{\mu})}(x) + g'_{A_0(\bar{x},\, \bar{\mu})}(x)\xi - \sigma(x,\, \mu)\zeta_{A_0(\bar{x},\, \bar{\mu})} \leq 0, \\ \zeta_{\{1,\, \ldots,\, m\} \setminus A(\bar{x})} = 0 \end{array} \right. \right\}, \tag{7.97}$$

and consider the affine VI

$$u \in Q, \quad \langle Bu + \tilde{b},\, v - u \rangle \geq 0 \quad \forall\, v \in Q. \tag{7.98}$$

Writing $u = (\xi,\, \eta)$, and employing (7.85), the KKT system for (7.98) has the form

$$\begin{gathered} G(x,\, \mu) + \frac{\partial G}{\partial x}(x,\, \mu)\xi + (g'_{A(\bar{x})}(x))^{\mathrm{T}}\tilde{\zeta}_{A(\bar{x})} + \tilde{\zeta}_{\{1,\, \ldots,\, m\} \setminus A(\bar{x})} = 0, \\ \sigma(x,\, \mu)\zeta - \sigma(x,\, \mu)\tilde{\zeta} = 0, \\ g_{A_+(\bar{x},\, \bar{\mu})}(x) + g'_{A_+(\bar{x},\, \bar{\mu})}(x)\xi - \sigma(x,\, \mu)\zeta_{A_+(\bar{x},\, \bar{\mu})} = 0, \\ \tilde{\zeta}_{A_0(\bar{x},\, \bar{\mu})} \geq 0, \quad g_{A_0(\bar{x},\, \bar{\mu})}(x) + g'_{A_0(\bar{x},\, \bar{\mu})}(x)\xi - \sigma(x,\, \mu)\zeta_{A_0(\bar{x},\, \bar{\mu})} \leq 0, \\ \tilde{\zeta}_i(g_i(x) + g'_i(x)\xi - \sigma(x,\, \mu)\zeta_i) = 0,\ i \in A_0(\bar{x},\, \bar{\mu}), \\ \zeta_{\{1,\, \ldots,\, m\} \setminus A(\bar{x})} = 0, \end{gathered}$$

with respect to $(\xi,\, \zeta,\, \tilde{\zeta}) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^m$. Observing that the second relation above implies $\zeta = \tilde{\zeta}$, we conclude that the affine VI (7.98) is equivalent to (7.96).

By Corollary 7.16, the matrix in (7.92) is nonsingular. Let an element $\tilde{u} = (\tilde{\xi},\, \tilde{\zeta}) \in \mathbf{R}^n \times \mathbf{R}^m$ be such that $\tilde{\zeta}_{\{1,\, \ldots,\, m\} \setminus A(\bar{x})} = 0$ and $(\tilde{\xi},\, \tilde{\zeta}_{A(\bar{x})})$ is uniquely defined by the linear system

$$\begin{gathered} \frac{\partial G}{\partial x}(x,\, \mu)\xi + (g'_{A(\bar{x})}(x))^{\mathrm{T}}\tilde{\zeta}_{A(\bar{x})} = -G(x,\, \mu), \\ g'_{A(\bar{x})}(x)\xi - \sigma(x,\, \mu)\tilde{\zeta}_{A(\bar{x})} = -g_{A(\bar{x})}(x). \end{gathered}$$

Comparing (7.86) and (7.97), we then see that $Q = K + \tilde{u}$. Then (7.98) transforms into

$$v \in K, \quad \langle Bv + B\tilde{u} + \tilde{b},\, w - v \rangle \geq 0 \quad \forall\, w \in K, \tag{7.99}$$

which is (7.95) with $b = B\tilde{u} + \tilde{b}$. This completes the proof. $\square$

We next show that the step given by solving the system (7.96), which is part of relations forming the affine MCP (7.77) (equivalently, the subproblem (7.75) of Algorithm 7.14), satisfies the localization property appearing in the iterative framework of Theorem 7.13.

Proposition 7.19. *In addition to the assumptions of Proposition 7.15, let* $\sigma : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}_+$ *be such that there exist a neighborhood* V *of* $(\bar{x}, \bar{\mu})$ *and some* $\beta_1 > 0$ *and* $\beta_2 > 0$ *such that* (7.80) *holds.*

Then there exists a constant $c > 0$ *such that the estimate*

$$\|(\xi, \zeta)\| \leq c\,\sigma(x, \mu) \tag{7.100}$$

holds for all $(x, \mu) \in \mathbf{R}^n \times \mathbf{R}^m_+$ *close enough to* $(\bar{x}, \bar{\mu})$ *and such that* $x \neq \bar{x}$ *or* $\mu \notin \mathcal{M}(\bar{x})$*, and for any solution* (ξ, ζ) *of (7.96).*

Proof. Suppose the contrary: there exist sequences $\{(x^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^m_+$ converging to $(\bar{x}, \bar{\mu})$, and $\{(\xi^k, \zeta^k)\} \subset \mathbf{R}^n \times \mathbf{R}^m$ such that for all k it holds that $x^k \neq \bar{x}$ or $\mu^k \notin \mathcal{M}(\bar{x})$ (implying that $\sigma(x^k, \mu^k) > 0$), and

$$\begin{gathered}
G(x^k, \mu^k) + \frac{\partial G}{\partial x}(x^k, \mu^k)\xi^k + (g'(x^k))^{\mathrm{T}}\zeta^k = 0,\\
g_{A_+(\bar{x}, \bar{\mu})}(x^k) + g'_{A_+(\bar{x}, \bar{\mu})}(x^k)\xi^k - \sigma(x^k, \mu^k)\zeta^k_{A_+(\bar{x}, \bar{\mu})} = 0,\\
(\mu^k + \zeta^k)_{A_0(\bar{x}, \bar{\mu})} \geq 0,\;\; g_{A_0(\bar{x}, \bar{\mu})}(x^k) + g'_{A_0(\bar{x}, \bar{\mu})}(x^k)\xi^k - \sigma(x^k, \mu^k)\zeta^k_{A_0(\bar{x}, \bar{\mu})} \leq 0,\\
(\mu^k_i + \zeta^k_i)(g_i(x^k) + \langle g'_i(x^k), \xi^k\rangle - \sigma(x^k, \mu^k)\zeta^k_i) = 0,\; i \in A_0(\bar{x}, \bar{\mu}),\\
\zeta^k_{\{1, \ldots, m\}\setminus A(\bar{x})} = 0,
\end{gathered} \tag{7.101}$$

while

$$\frac{\|(\xi^k, \zeta^k)\|}{\sigma(x^k, \mu^k)} \to \infty \text{ as } k \to \infty. \tag{7.102}$$

Set $\sigma_k = \sigma(x^k, \mu^k)$, $t_k = \|(\xi^k, \zeta^k)\|$. Then (7.102) implies that $t_k > 0$ for all k large enough, and

$$\frac{\sigma_k}{t_k} \to 0 \text{ as } k \to \infty. \tag{7.103}$$

Note first that under our differentiability assumptions, by the mean-value theorem (see Theorem A.10, (a)) and by (7.80), it holds that

$$F(x^k) - F(\bar{x}) = O(\|x^k - \bar{x}\|) = O(\sigma_k), \tag{7.104}$$

$$\begin{aligned}
g_{A(\bar{x})}(x^k) &= g_{A(\bar{x})}(x^k) - g_{A(\bar{x})}(\bar{x})\\
&= g'_{A(\bar{x})}(\bar{x})(x^k - \bar{x}) + O(\|x^k - \bar{x}\|^2)\\
&= O(\|x^k - \bar{x}\|)\\
&= O(\sigma_k),
\end{aligned} \tag{7.105}$$

$$g'(x^k) - g'(\bar{x}) = O(\|x^k - \bar{x}\|) = O(\sigma_k) \tag{7.106}$$

as $k \to \infty$. For each k, denote $\hat{\mu}^k = \pi_{\mathcal{M}(\bar{x})}(\mu^k)$. Since $\hat{\mu}^k \in \mathcal{M}(\bar{x})$, using the relations (7.73) and (7.104), (7.106), we further obtain that

$$\begin{aligned} G(x^k,\, \mu^k) &= G(x^k,\, \mu^k) - G(\bar{x},\, \hat{\mu}^k) \\ &= F(x^k) + (g'(x^k))^{\mathrm{T}}\mu^k - F(\bar{x}) - (g'(\bar{x}))^{\mathrm{T}}\hat{\mu}^k \\ &= O(\sigma^k) + O(\|\mu^k - \hat{\mu}^k\|) \\ &= O(\sigma_k) \end{aligned} \tag{7.107}$$

as $k \to \infty$, where the last relation again employs (7.80).

Taking a subsequence, if necessary, we can assume that $\{(\xi^k/t_k,\, \zeta^k/t_k)\}$ converges to some $(\xi,\, \zeta) \in \mathbf{R}^n \times \mathbf{R}^m$, $(\xi,\, \zeta) \neq 0$. Dividing both sides of the first equality in (7.101) by t_k and taking the limit as $k \to \infty$, using also (7.103) and (7.107), we obtain that

$$\frac{\partial G}{\partial x}(\bar{x},\, \bar{\mu})\xi + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0. \tag{7.108}$$

Similarly, from the second and third lines in (7.101), taking into account (7.105), we derive that

$$g'_{A_+(\bar{x},\, \bar{\mu})}(\bar{x})\xi = 0, \quad g'_{A_0(\bar{x},\, \bar{\mu})}(\bar{x})\xi \leq 0. \tag{7.109}$$

Furthermore, by the second to fourth lines in (7.101), using also that $\mu^k \geq 0$, we have that

$$\langle (\mu^k + \zeta^k)_{A(\bar{x})},\, g_{A(\bar{x})}(x^k) + g'_{A(\bar{x})}(x^k)\xi^k - \sigma_k \zeta^k_{A(\bar{x})}\rangle = 0,$$

$$\langle \mu^k_{A(\bar{x})},\, g_{A(\bar{x})}(x^k) + g'_{A(\bar{x})}(x^k)\xi^k - \sigma_k \zeta^k_{A(\bar{x})}\rangle \leq 0,$$

which implies that

$$\langle \zeta^k_{A(\bar{x})},\, g_{A(\bar{x})}(x^k) + g'_{A(\bar{x})}(x^k)\xi^k - \sigma_k \zeta^k_{A(\bar{x})}\rangle \geq 0.$$

Dividing both sides of the latter inequality by t_k^2 and taking the limit as $k \to \infty$, using (7.103) and (7.105) we obtain that

$$\langle \zeta_{A(\bar{x})},\, g'_{A(\bar{x})}(\bar{x})\xi\rangle \geq 0. \tag{7.110}$$

Finally, by the last equality in (7.101), we conclude that

$$\zeta_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})} = 0. \tag{7.111}$$

Relations in (7.109) show that $\xi \in C(\bar{x})$. Multiplying by ξ both sides in (7.108) and using (7.110), (7.111), we obtain that

$$\left\langle \frac{\partial G}{\partial x}(\bar{x},\, \bar{\mu})\xi,\, \xi \right\rangle \leq 0,$$

so that (7.79) implies $\xi = 0$. Then the relations (7.108) and (7.111) give $(g'_{A(\bar{x})}(\bar{x}))^{\mathrm{T}}\zeta_{A(\bar{x})} = 0$, i.e.,

$$\zeta_{A(\bar{x})} \in \ker(g'_{A(\bar{x})}(\bar{x}))^{\mathrm{T}}. \tag{7.112}$$

Set $\pi = \pi_{\ker g'_{A(\bar{x})}(\bar{x})}$. By the intermediate relations in (7.105), employing the equality $\ker g'_{A(\bar{x})}(\bar{x}) = (\operatorname{im} g'_{A(\bar{x})}(\bar{x}))^{\perp}$ implying $\pi g'_{A(\bar{x})}(\bar{x}) = 0$, we have that

$$\pi g_{A(\bar{x})}(x^k) = O(\|x^k - \bar{x}\|^2) = O(\sigma_k^2) \tag{7.113}$$

as $k \to \infty$, where the last equality is by (7.80).

Let

$$I_k = \{i \in A(\bar{x}) \mid g_i(x^k) + \langle g_i'(x^k),\, \xi^k\rangle - \sigma_k \zeta_i^k = 0\}.$$

Evidently, there exists an index set $I \subset A(\bar{x})$ such that $I_k = I$ for infinitely many indices k. From now on, without loss of generality we assume that $I_k = I$ for all k. By the second and forth lines in (7.101), we have that $A_+(\bar{x},\, \bar{\mu}) \subset I$ and $(\mu^k + \zeta^k)_{A(\bar{x})\setminus I} = 0$, so that $\zeta^k_{A(\bar{x})\setminus I} = -\mu^k_{A(\bar{x})\setminus I} \leq 0$, and therefore $\zeta_{A(\bar{x})\setminus I} \leq 0$.

Define the cone Ω as the projection of the polyhedral cone defined as $\{z \in \mathbf{R}^m \mid z_I = 0,\ z_{A(\bar{x})\setminus I} \leq 0\}$ onto the $|A(\bar{x})|$-dimensional linear subspace $\{z \in \mathbf{R}^m \mid z_{\{1,\,\dots,\,m\}\setminus A(\bar{x})} = 0\}$. Then Ω is also a polyhedral cone, and since $\zeta_{A(\bar{x})\setminus I} \leq 0$, it holds that $-\zeta_{A(\bar{x})} \in \Omega^{\circ}$.

On the other hand, taking into account the third line in (7.101), we have that for all k

$$\Omega \ni g_{A(\bar{x})}(x^k) + g'_{A(\bar{x})}(x^k)\xi^k - \sigma_k \zeta^k_{A(\bar{x})},$$

and hence, using the equality $\pi g'_{A(\bar{x})}(\bar{x}) = 0$,

$$\pi\Omega \ni \pi g_{A(\bar{x})}(x^k) + \pi(g'_{A(\bar{x})}(x^k) - g'_{A(\bar{x})}(\bar{x}))\xi^k - \sigma_k \pi\zeta^k_{A(\bar{x})}.$$

Since the set in the left-hand side is a closed cone (since it is the projection of a polyhedral cone), dividing the latter relation by $t_k\sigma_k$ and taking the limit as $k \to \infty$, using (7.103), (7.106), (7.113) and the equality $\xi = 0$, we obtain that

$$\pi\Omega \ni -\pi\zeta_{A(\bar{x})} = -\zeta_{A(\bar{x})},$$

where the last equality follows from (7.112). Therefore, there exists $\omega \in \Omega$ such that

$$\zeta_{A(\bar{x})} = -\pi\omega. \tag{7.114}$$

Since $-\zeta_{A(\bar{x})} \in \Omega^{\circ}$, we conclude by (7.114) that

$$0 \geq \langle -\zeta_{A(\bar{x})},\, \omega\rangle = \langle \pi\omega,\, \omega\rangle = \langle \pi\omega,\, \pi\omega\rangle = \|\pi\omega\|^2.$$

Thus $\pi\omega = 0$, so that (7.114) implies that $\zeta = 0$. Combining the latter with (7.111), we obtain a contradiction with $(\xi,\, \zeta) \neq 0$. $\square$

We now extend the solution of (7.96) to the solution of the full affine MCP (7.77) (equivalently of the affine VI subproblem (7.75)), showing also that the needed localization property holds.

Proposition 7.20. *In addition to the assumptions of Proposition 7.15, let* $\sigma : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}_+$ *be such that there exist a neighborhood* V *of* $(\bar{x}, \bar{\mu})$ *and some* $\beta_1 > 0$ *and* $\beta_2 > 0$ *such that* (7.80) *holds.*

Then there exists a constant $c > 0$ *such that for all* $(x, \mu) \in \mathbf{R}^n \times \mathbf{R}^m_+$ *close enough to* $(\bar{x}, \bar{\mu})$ *the affine MCP of finding* $(\xi, \zeta) \in \mathbf{R}^n \times \mathbf{R}^m$ *such that*

$$\begin{gathered} G(x, \mu) + \frac{\partial G}{\partial x}(x, \mu)\xi + (g'(x))^{\mathrm{T}}\zeta = 0, \\ \mu + \zeta \geq 0, \quad g(x) + g'(x)\xi - \sigma(x, \mu)\zeta \leq 0, \\ \langle \mu + \zeta,\, g(x) + g'(x)\xi - \sigma(x, \mu)\zeta \rangle = 0, \end{gathered} \tag{7.115}$$

has a solution satisfying (7.100)*.*

Proof. If $x = \bar{x}$ and $\mu \in \mathcal{M}(\bar{x})$, then $(\xi, \zeta) = (0, 0)$ is a solution of (7.115) trivially satisfying (7.100) for any $c \geq 0$.

Suppose now that $x \neq \bar{x}$ or $\mu \notin \mathcal{M}(\bar{x})$, so that $\sigma(x, \mu) > 0$. By Propositions 7.18, 7.19, there exists a constant $c > 0$ such that if $(x, \mu) \in \mathbf{R}^n \times \mathbf{R}^m_+$ is close enough to $(\bar{x}, \bar{\mu})$, then (7.96) has a solution (ξ, ζ), and any of its solutions satisfies (7.100).

We have that

$$g_{\{1,\ldots,m\}\setminus A(\bar{x})}(x) + g'_{\{1,\ldots,m\}\setminus A(\bar{x})}(x)\xi - \sigma(x, \mu)\zeta_{\{1,\ldots,m\}\setminus A(\bar{x})} < 0 \tag{7.116}$$

provided (x, μ) is sufficiently close to $(\bar{x}, \bar{\mu})$ (so that $\sigma(x, \mu)$ is small enough and, consequently, so are $\|\xi\|$ and $\|\zeta\|$, by (7.100)).

Given (7.96) and (7.116), it remains to observe that

$$(\mu + \zeta)_{A_+(\bar{x}, \bar{\mu})} > 0 \tag{7.117}$$

for all (x, μ) sufficiently close to $(\bar{x}, \bar{\mu})$ (so that $\sigma(x, \mu)$ is small enough and, consequently, so is $\|\zeta\|$, by (7.100)). This concludes the proof. □

Proposition 7.20 establishes that assumption (iii) of Theorem 7.13 is satisfied. In particular, subproblems given by (7.75) (equivalently, by (7.77)) are locally solvable and the distance between consecutive iterates can be bounded above by a measure of violation of the KKT system for the original problem (7.70). Consequently, by (7.80), this distance is of the order of the distance to the solution set.

We are now in position to state the local convergence properties of the stabilized Newton method for VPs (Algorithm 7.14) and, as a direct consequence, of stabilized SQP for optimization.

Theorem 7.21. *Suppose that the assumptions of Proposition 7.15 hold. In addition, let* $\sigma : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}_+$ *be such that there exist a neighborhood* V *of* $(\bar{x}, \bar{\mu})$ *and some* $\beta_1 > 0$ *and* $\beta_2 > 0$ *such that* (7.80) *holds.*

Then for any $c > 0$ *large enough and any staring point* $(x^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^m_+$ *close enough to* $(\bar{x}, \bar{\mu})$*, there exists a sequence* $\{(x^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^m$ *such* *that for each* $k = 0, 1, \ldots,$ *the pair* (x^{k+1}, μ^{k+1}) *solves the affine VI* (7.75) *and satisfies*

$$\|(x^{k+1} - x^k, \mu^{k+1} - \mu^k)\| \leq c \operatorname{dist}((x^k, \mu^k), \{\bar{x}\} \times \mathcal{M}(\bar{x}));$$

any such sequence converges to $(\bar{x}, \mu^*)$ *with some* $\mu^* \in \mathcal{M}(\bar{x})$*, and the rates of convergence of* $\{(x^k, \mu^k)\}$ *to* $(\bar{x}, \mu^*)$ *and of* $\{\operatorname{dist}((x^k, \mu^k), \{\bar{x}\} \times \mathcal{M}(\bar{x}))\}$ *to zero are superlinear. Moreover, the rates of convergence are quadratic provided the derivative of* F *and the second derivative of* g *are locally Lipschitz-continuous at* $\bar{x}$*.*

Proof. The proof is by applying Theorem 7.13, once its assumptions are verified. Observe first that under the stated smoothness assumptions, there exists $\bar{\varepsilon} > 0$ such that the intersection of the solution set of the KKT system (7.71) with $B(\bar{x}, \bar{\varepsilon}) \times \mathbf{R}^m$ is closed.

Assumption (i) is the upper Lipschitzian behavior of the solutions of the generalized equation corresponding to the KKT system (7.71) under canonical perturbations. Since the second-order sufficiency condition (7.79) implies that the multiplier $\bar{\mu}$ is noncritical (Definition 1.41), by Proposition 1.43 we have that (7.79) also implies assumption (i) of Theorem 7.13.

Assumption (ii) of Theorem 7.13 on the quality of approximation of $\Phi(\cdot)$ (defined by (7.82)) by $\mathcal{A}(\tilde{u}, u)$ (defined by (7.83), (7.84)) can be easily checked by the mean-value theorem (Theorem A.10, (a)) as follows:

$$\begin{aligned}\|\Phi(u) - \mathcal{A}(\tilde{u}, u)\| &\leq \|\Phi(u) - \Phi(\tilde{u}) - \Phi'(\tilde{u})(u - \tilde{u})\| + \|M(\tilde{u})\|\|u - \tilde{u}\| \\ &= \sigma(\tilde{x}, \tilde{\mu})\|u - \tilde{u}\| + o(\|u - \tilde{u}\|) \\ &= o(\operatorname{dist}(\tilde{u}, \bar{U}))\end{aligned}$$

as $\operatorname{dist}(\tilde{u}, \bar{U}) \to 0$ for $\tilde{u} = (\tilde{x}, \tilde{\mu}) \in \mathbf{R}^n \times \mathbf{R}^m$ and $\bar{U} = \{\bar{x}\} \times \mathcal{M}(\bar{x})$, provided $\|u - \tilde{u}\| \leq \operatorname{dist}(\tilde{u}, \bar{U})$ (here the right-most inequality in (7.80) is also taken into account). If, in addition, the derivative of F and the second derivative of g are locally Lipschitz-continuous at $\bar{x}$, then by similar considerations we derive the estimate

$$\Phi(u) - \mathcal{A}(\tilde{u}, u) = O((\operatorname{dist}(\tilde{u}, \bar{U}))^2).$$

Finally, assumption (iii) of Theorem 7.13 had been established in Proposition 7.20 above (given the equivalence between the affine VI (7.75) and the affine MCP (7.77)). $\square$

Under a stronger second-order condition, solutions of subproblems (7.77) can be shown to be locally unique. The strong second-order sufficiency condition for the KKT system (7.71) states that

$$\left\langle \frac{\partial G}{\partial x}(\bar{x}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall\, \xi \in C_+(\bar{x}, \bar{\mu}) \setminus \{0\}, \tag{7.118}$$

where the cone $C_+(\bar{x}, \bar{\mu})$ is defined according to (7.68). In the optimization setting (7.72), this corresponds to the SSOSC (7.67). We shall provide the main steps, indicating the changes needed in the preceding analysis.

Regarding the proof of Proposition 7.15, it can be seen that under the strong second-order sufficiency condition (7.118) there exists a constant $\gamma > 0$ such that for all $(x, \mu) \in \mathbf{R}^n \times \mathbf{R}^m$ in a neighborhood of $(\bar{x}, \bar{\mu})$ it holds that

$$\begin{aligned}\langle B(x, \mu)(\xi, \zeta), (\xi, \zeta)\rangle &= \left\langle \frac{\partial G}{\partial x}(x, \mu)\xi, \xi \right\rangle + \sigma(x, \mu)\|\zeta\|^2 \\ &\geq \gamma(\|\xi\|^2 + \sigma(x, \mu)\|\zeta\|^2) \quad \forall\, (\xi, \zeta) \in K_+(x, \mu),\end{aligned} \tag{7.119}$$

where

$$K_+(x, \mu) = \left\{ (\xi, \zeta) \in \mathbf{R}^n \times \mathbf{R}^m \;\middle|\; \begin{array}{c} g'_{A_+(\bar{x}, \bar{\mu})}(x)\xi - \sigma(x, \mu)\zeta_{A_+(\bar{x}, \bar{\mu})} = 0, \\ \zeta_{\{1, \ldots, m\}\setminus A(\bar{x})} = 0 \end{array} \right\}. \tag{7.120}$$

Since $C(\bar{x}) \subset C_+(\bar{x}, \bar{\mu})$, the second-order sufficiency condition (7.79) holds, and so do the assertions of Propositions 7.15 and 7.18, as well as all the ingredients in their proofs. In particular, in the proof of Proposition 7.18, the affine VI (7.99) has a solution. Let $v^1, v^2 \in \mathbf{R}^n \times \mathbf{R}^m$ be solutions of this VI. Then $v^1, v^2 \in K \subset K_+(x, \mu)$ and

$$\begin{aligned}\langle B(v^1 - v^2), v^1 - v^2\rangle &= \langle Bv^1 + B\tilde{u} + \tilde{b}, v^1 - v^2\rangle - \langle Bv^2 + B\tilde{u} + \tilde{b}, v^1 - v^2\rangle \\ &\leq -\langle Bv^1 + B\tilde{u} + \tilde{b}, v^2\rangle - \langle Bv^2 + B\tilde{u} + \tilde{b}, v^1\rangle \\ &\leq 0,\end{aligned} \tag{7.121}$$

where the first inequality follows from (7.99) by taking $w = 0 \in K$. As $v^1, v^2 \in K_+(x, \mu)$, and $K_+(x, \mu)$ is a subspace, $v^1 - v^2 \in K_+(x, \mu)$ holds. Then (7.119) and (7.121) imply that $v^1 - v^2 = 0$. Hence, the MCP (7.96) has the unique solution (ξ, ζ).

Furthermore, by Proposition 7.20, (ξ, ζ) defined this way is a solution of (7.115) satisfying (7.100). Conversely, if $(\xi, \zeta) \in \mathbf{R}^n \times \mathbf{R}^m_+$ is a solution of (7.115) satisfying (7.100), and if (x, μ) is sufficiently close to $(\bar{x}, \bar{\mu})$, we have (7.116) and (7.117). By the complementarity condition in (7.115), these two relations imply

$$g_{A_+(\bar{x}, \bar{\mu})}(x) + g'_{A_+(\bar{x}, \bar{\mu})}(x)\xi - \sigma(x, \mu)\zeta_{A_+(\bar{x}, \bar{\mu})} = 0, \quad \zeta_{\{1, \ldots, m\}\setminus A(\bar{x})} = 0.$$

Hence, (ξ, ζ) is a solution of (7.96), which had been established to be unique.

Following [150], we next show that for equality-constrained problems, local superlinear convergence of the stabilized Newton method can be established with the second-order sufficiency condition replaced by the weaker assumption that the Lagrange multiplier in question is noncritical (Definition 1.41).

Consider the VP (7.70) with its feasible set now defined by equality constraints:

$$D = \{x \in \mathbf{R}^n \mid h(x) = 0\},$$

where $h : \mathbf{R}^n \to \mathbf{R}^l$ is twice differentiable. Conditions characterizing primal-dual solutions of the VP (7.70) are

$$F(x) + (h'(x))^{\mathrm{T}}\lambda = 0, \quad h(x) = 0. \tag{7.122}$$

In the optimization case (7.72), these relations give the Lagrange optimality system for problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0. \end{array}$$

Define $G : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}^n$ and $\sigma : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}_+$ by

$$G(x, \lambda) = F(x) + (g'(x))^{\mathrm{T}}\lambda, \quad \sigma(x, \lambda) = \|(G(x, \lambda), h(x))\|. \tag{7.123}$$

Given $(x^k, \lambda^k) \in \mathbf{R}^n \times \mathbf{R}^l$, the current primal-dual approximation to a solution of (7.122), define the affine mapping $\Phi_k : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}^n \times \mathbf{R}^l$ by

$$\Phi_k(u) = \left(F(x^k) + \frac{\partial G}{\partial x}(x^k, \lambda^k)(x - x^k), \sigma(x^k, \lambda^k)\lambda\right), \quad u = (x, \lambda),$$

and consider the affine VI (7.75), where

$$Q_k = \left\{ \begin{array}{l} u = (x, \lambda) \\ \in \mathbf{R}^n \times \mathbf{R}^l \end{array} \,\middle|\, h(x^k) + h'(x^k)(x - x^k) - \sigma(x^k, \lambda^k)(\lambda - \lambda^k) = 0 \right\}.$$

It can be readily seen that this affine VI is equivalent to the system of linear equations

$$\begin{aligned} \frac{\partial G}{\partial x}(x^k, \lambda^k)(x - x^k) + (h'(x^k))^{\mathrm{T}}(\lambda - \lambda^k) &= -G(x^k, \lambda^k), \\ h'(x^k)(x - x^k) - \sigma(x^k, \lambda^k)(\lambda - \lambda^k) &= -h(x^k). \end{aligned} \tag{7.124}$$

In the optimization setting (7.72), the linear system (7.124) corresponds to the Lagrange optimality system of the stabilized SQP subproblem, which is an equality-constrained quadratic programming problem.

Algorithm 7.22 Choose $(x^0, \lambda^0) \in \mathbf{R}^n \times \mathbf{R}^l$ and set $k = 0$.

1. Compute $\sigma_k = \sigma(x^k, \lambda^k)$, where the function $\sigma : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}_+$ is defined in (7.123). If $\sigma_k = 0$, stop.
2. Compute $(x^{k+1}, \lambda^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^l$ as a solution of the linear system (7.124), where the mapping $G : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}^n$ is defined in (7.123).
3. Increase k by 1 and go to step 1.

Let $\bar{x} \in \mathbf{R}^n$ be a given primal solution of (7.122), and let $\mathcal{M}(\bar{x})$ be the associated (nonempty) set of Lagrange multipliers, i.e., the set of $\lambda \in \mathbf{R}^l$

satisfying (7.122) for $x = \bar{x}$. Recall that in the current setting a multiplier $\bar{\lambda} \in \mathcal{M}(\bar{x})$ is referred to as critical (see Definition 1.41 and (1.77)) if there exists $\xi \in \ker h'(\bar{x}) \setminus \{0\}$ such that

$$\frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda})\xi \in \operatorname{im}(h'(\bar{x}))^{\mathrm{T}}, \tag{7.125}$$

and noncritical otherwise. As discussed in Sect. 1.3.3, noncriticality of a multiplier $\bar{\lambda} \in \mathcal{M}(\bar{x})$ is weaker than the second-order sufficiency condition

$$\left\langle \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda})\xi, \xi \right\rangle > 0 \quad \forall\, \xi \in \ker h'(\bar{x}) \setminus \{0\}.$$

By Proposition 1.43, noncriticality of $\bar{\lambda} \in \mathcal{M}(\bar{x})$ is equivalent to assumption (i) of Theorem 7.13 (upper Lipschitz-continuity of solutions of (7.122) under canonical perturbations), as well as to the following error bound:

$$\|x - \bar{x}\| + \operatorname{dist}(\lambda, \mathcal{M}(\bar{x})) = O(\sigma(x, \lambda)) \tag{7.126}$$

as $(x, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$ tends to $(\bar{x}, \bar{\lambda})$.

Assumption (ii) of Theorem 7.13 (on the quality of approximation) holds by the same simple considerations as for the case of inequality constraints in Theorem 7.21.

It remains to prove that for the equality-constrained problems, if $\bar{\lambda}$ is a noncritical multiplier, then assumption (iii) of Theorem 7.13 (solvability of subproblems) holds. Specifically, we have to prove that there exists $c > 0$ such that for all $(x, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$ close enough to $(\bar{x}, \bar{\lambda})$, the linear system

$$\frac{\partial G}{\partial x}(x, \lambda)\xi + (h'(x))^{\mathrm{T}}\eta = -G(x, \lambda), \quad h'(x)\xi - \sigma(x, \lambda)\eta = -h(x) \tag{7.127}$$

has a solution (ξ, η) satisfying the estimate

$$\|(\xi, \eta)\| \le c(\|x - \bar{x}\| + \operatorname{dist}(\lambda, \mathcal{M}(\bar{x}))). \tag{7.128}$$

We first show that the matrix of the linear system in (7.127) is locally nonsingular.

Proposition 7.23. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with its derivative being continuous at $\bar{x}$, let $h : \mathbf{R}^n \to \mathbf{R}^l$ be twice differentiable in a neighborhood of $\bar{x}$, with its second derivative being continuous at $\bar{x}$, and let $G : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}^n$ and $\sigma : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}_+$ be defined in (7.123). Let $(\bar{x}, \bar{\lambda})$ with some $\bar{\lambda} \in \mathcal{M}(\bar{x})$ be a solution of (7.122), and let $\bar{\lambda}$ be a noncritical Lagrange multiplier.*

Then for all $(x, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$ close enough to $(\bar{x}, \bar{\lambda})$ and such that $x \neq \bar{x}$ or $\lambda \notin \mathcal{M}(\bar{x})$, the matrix

$$\begin{pmatrix} \frac{\partial G}{\partial x}(x,\, \lambda) & (h'(x))^{\mathrm{T}} \\ h'(x) & -\sigma(x,\, \lambda)I \end{pmatrix}$$

is nonsingular.

Proof. As the assumption that $\bar{\lambda}$ is noncritical is equivalent to the error bound (7.126), it holds that $\sigma(x, \lambda) > 0$ if $x \neq \bar{x}$ or $\lambda \notin \mathcal{M}(\bar{x})$.

Suppose that the contrary to the stated assertion holds: there exist a sequence $\{(x^k, \lambda^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l$ converging to $(\bar{x}, \bar{\lambda})$, and a sequence $\{(\xi^k, \eta^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l$, such that for each k it holds that $x^k \neq \bar{x}$ or $\lambda^k \notin \mathcal{M}(\bar{x})$, $(\xi^k, \eta^k) \neq (0, 0)$, and (ξ^k, η^k) belongs to the null space of the matrix in question, that is,

$$\frac{\partial G}{\partial x}(x^k,\, \lambda^k)\xi^k + (h'(x^k))^{\mathrm{T}}\eta^k = 0, \quad h'(x^k)\xi^k - \sigma(x^k,\, \lambda^k)\eta^k = 0.$$

The second equation implies that

$$\eta^k = \frac{1}{\sigma(x^k,\, \lambda^k)} h'(x^k)\xi^k, \tag{7.129}$$

and by substituting this into the first equation, we obtain that

$$\left(\frac{\partial G}{\partial x}(x^k,\, \lambda^k) + \frac{1}{\sigma(x^k,\, \lambda^k)}(h'(x^k))^{\mathrm{T}} h'(x^k) \right) \xi^k = 0. \tag{7.130}$$

Set $H = \frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda})$, $A = h'(\bar{x})$, $H_k = \frac{\partial G}{\partial x}(x^k, \lambda^k)$, $A_k = h'(x^k)$, and $\Omega_k = h'(x^k) - h'(\bar{x})$, $c_k = 1/\sigma(x^k, \lambda^k)$. Observing that, by (7.126), it holds that

$$\Omega_k = O(\|x^k - \bar{x}\|) = O(\sigma(x^k,\, \lambda^k)) = O\left(\frac{1}{c_k}\right),$$

and applying Lemma A.8 with $\tilde{H} = H_k$, $\tilde{A} = A_k$, $\Omega = \Omega_k$ and $c = c_k$, we conclude that for all k large enough the matrix in the left-hand side of (7.130) is nonsingular. Then (7.130) implies that $\xi^k = 0$. By (7.129), we then have that $\eta^k = 0$, contradicting the assumption that $(\xi^k, \eta^k) \neq (0, 0)$. □

It remains to establish the localization estimate (7.128).

Proposition 7.24. *Under the assumptions of Proposition 7.23, there exists* $c > 0$ *such that for all* $(x, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$ *close enough to* $(\bar{x}, \bar{\lambda})$ *the linear system* (7.127) *has a solution satisfying* (7.128).

Proof. If $x = \bar{x}$ and $\lambda \in \mathcal{M}(\bar{x})$, then the right-hand side of (7.127) is equal to zero. Then this system has the trivial solution, satisfying (7.128) for any $c \geq 0$.

By Proposition 7.23, for all $(x, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$ close enough to $(\bar{x}, \bar{\lambda})$ and such that $x \neq \bar{x}$ or $\lambda \notin \mathcal{M}(\bar{x})$, the matrix of the linear system (7.127) is nonsingular. Hence, this system has the unique solution (ξ, η).

Suppose that there exists no $c > 0$ such that the estimate (7.128) is valid. The latter means that there exist sequences $\{(x^k, \lambda^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l$ converging to $(\bar{x}, \bar{\lambda})$, and $\{(\xi^k, \eta^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l$, such that for each k it holds that $x^k \neq \bar{x}$ or $\tilde{\lambda}^k \notin \mathcal{M}(\bar{x})$,

$$\begin{aligned} \frac{\partial G}{\partial x}(x^k, \lambda^k)\xi^k + (h'(x^k))^{\mathrm{T}}\eta^k &= -G(x^k, \lambda^k), \\ h'(x^k)\xi^k - \sigma(x^k, \lambda^k)\eta^k &= -h(x^k), \end{aligned} \tag{7.131}$$

and

$$\frac{\|(\xi^k, \eta^k)\|}{\|x^k - \bar{x}\| + \operatorname{dist}(\lambda^k, \mathcal{M}(\bar{x}))} \to \infty \text{ as } k \to \infty.$$

The last relation above can be equivalently written as

$$\frac{\|x^k - \bar{x}\| + \operatorname{dist}(\lambda^k, \mathcal{M}(\bar{x}))}{\|(\xi^k, \eta^k)\|} \to 0 \text{ as } k \to \infty.$$

Employing this relation and (7.123), as well as the evident estimate

$$\sigma(x, \lambda) = O(\|x - \bar{x}\| + \operatorname{dist}(\lambda, \mathcal{M}(\bar{x})))$$

as $(x, \lambda) \to (\bar{x}, \bar{\lambda})$, we obtain that

$$\frac{\|(G(x^k, \lambda^k), h(x^k))\|}{\|(\xi^k, \eta^k)\|} = \frac{\sigma(x^k, \lambda^k)}{\|(\xi^k, \eta^k)\|} \to 0 \text{ as } k \to \infty. \tag{7.132}$$

Passing onto a convergent subsequence if necessary, we can assume that $\{(\xi^k, \eta^k)/\|(\xi^k, \eta^k)\|\}$ converges to some $(\xi, \eta) \in \mathbf{R}^n \times \mathbf{R}^l$, $(\xi, \eta) \neq 0$. Then dividing the relations in (7.131) by the quantity $\|(\xi^k, \eta^k)\|$, passing onto the limit as $k \to \infty$, and using also (7.132), we obtain that

$$\frac{\partial G}{\partial x}(\bar{x}, \bar{\lambda})\xi + (h'(\bar{x}))^{\mathrm{T}}\eta = 0, \quad h'(\bar{x})\xi = 0. \tag{7.133}$$

If $\xi \neq 0$, then the latter means that $\bar{\lambda}$ is critical (see (7.125)), contradicting the assumption. Thus $\xi = 0$. Then the first equation in (7.133) reduces to

$$(h'(\bar{x}))^{\mathrm{T}}\eta = 0. \tag{7.134}$$

Applying the projection operator $\pi = \pi_{\ker(h'(\bar{x}))^{\mathrm{T}}}$ to both sides of the second equation in (7.131), and taking into account $\ker(h'(\bar{x}))^{\mathrm{T}} = (\operatorname{im} h'(\bar{x}))^{\perp}$ implying that $\pi h'(\bar{x}) = 0$, under the stated differentiability assumptions we obtain that

$$\begin{aligned}
\sigma(x^k,\, \lambda^k)\pi\eta^k &= \pi(h(x^k) + h'(x^k)\xi^k) \\
&= \pi(h(x^k) - h(\bar{x}) - h'(\bar{x})(x^k - \bar{x}) + (h'(x^k) - h'(\bar{x}))\xi^k) \\
&= O(\|x^k - \bar{x}\|^2) + O(\|x^k - \bar{x}\|\|\xi^k\|) \\
&= O((\sigma(x^k,\, \lambda^k))^2) + O(\sigma(x^k,\, \lambda^k)\|\xi^k\|)
\end{aligned}$$

as $k \to \infty$, where we have used (7.126) for the last equality. Thus,

$$\pi\eta^k = O(\sigma(x^k,\, \lambda^k)) + O(\|\xi^k\|).$$

Dividing both sides of the latter relation by $\|(\xi^k,\, \eta^k)\|$ and passing onto the limit as $k \to \infty$, we obtain that $\pi\eta = 0$, where (7.132) and the fact that $\xi = 0$ were used. This shows that $\eta \in \ker \pi = \operatorname{im} h'(\bar{x})$. At the same time, (7.134) means that $\eta \in \ker(h'(\bar{x}))^{\mathrm{T}} = (\operatorname{im} h'(\bar{x}))^{\perp}$. Hence, $\eta = 0$. Since we also have that $\xi = 0$, this gives a contradiction with $(\xi,\, \eta) \neq 0$. $\square$

We have therefore established local convergence of the stabilized Newton method for equality-constrained VPs (Algorithm 7.22), and as a consequence of stabilized SQP for equality-constrained optimization problems, under the only assumption of noncriticality of the relevant Lagrange multiplier.

Theorem 7.25. *Assume the hypothesis of Proposition 7.23.*

Then any starting point $(x^0,\, \lambda^0) \in \mathbf{R}^n \times \mathbf{R}^l$ close enough to $(\bar{x},\, \bar{\lambda})$ uniquely defines the sequence $\{(x^k,\, \lambda^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l$ of Algorithm 7.22, this sequence converges to $(\bar{x},\, \lambda^*)$ with some $\lambda^* \in \mathcal{M}(\bar{x})$, and the rates of convergence of $\{(x^k,\, \lambda^k)\}$ to $(\bar{x},\, \lambda^*)$ and of $\{\operatorname{dist}((x^k,\, \lambda^k),\, \{\bar{x}\} \times \mathcal{M}(\bar{x}))\}$ to zero are superlinear. Moreover, the rates of convergence are quadratic provided the derivative of F and the second derivative of h are locally Lipschitz-continuous with respect to $\bar{x}$.

Getting back to inequality-constrained VPs, recall that by Definition 1.41, a Lagrange multiplier $\bar{\mu} \in \mathbf{R}^m$ associated with a primal solution $\bar{x} \in \mathbf{R}^n$ of the KKT system (7.71) is noncritical if for any solution $(\xi,\, \eta,\, \zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ of the system

$$\begin{gathered}
\frac{\partial G}{\partial x}(\bar{x},\, \bar{\mu})\xi + (g'(\bar{x}))^{\mathrm{T}}\zeta = 0, \quad g'_{A_+(\bar{x},\, \bar{\mu})}(\bar{x})\xi = 0, \\
\zeta_{A_0(\bar{x},\, \bar{\mu})} \geq 0, \quad g'_{A_0(\bar{x},\, \bar{\mu})}(\bar{x})\xi \leq 0, \quad \zeta_i\langle g_i'(\bar{x}),\, \xi\rangle = 0,\ i \in A_0(\bar{x},\, \bar{\mu}), \\
\zeta_{\{1,\, \ldots,\, m\}\setminus A(\bar{x})} = 0
\end{gathered} \tag{7.135}$$

it holds that $\xi = 0$. By Proposition 1.43, noncriticality of $\bar{\mu} \in \mathcal{M}(\bar{x})$ is equivalent to assumption (i) of Theorem 7.13 for the corresponding GE. However, the following example shows that for problems with inequality constraints the weaker assumption that the multiplier is noncritical cannot replace the second-order sufficiency condition (7.79) in Theorem 7.21, even in the case of optimization. The problem in this example is the same as in Example 3.3, where it was used to demonstrate that semistability does not imply hemistability, in general. Similarly, here it demonstrates that for VPs with inequality constraints, assumption (i) of Theorem 7.13 does not imply its assumption (iii).

Example 7.26. Let $n = 1$, $m = 1$, $f(x) = -x^2/2 + x^3/6$, $g(x) = -x$. Problem (7.64) with this data has a global solution $\hat{x} = 2$, and a stationary point $\bar{x} = 0$ which is not even a local solution. The point $\bar{x}$ satisfies the LICQ, and the unique associated Lagrange multiplier is $\bar{\mu} = 0$, giving $A_+(\bar{x}, \bar{\mu}) = \emptyset$, $A(\bar{x}) = A_0(\bar{x}, \bar{\mu})$.

The system (7.135) takes the form

$$-\xi - \zeta = 0, \quad \zeta \geq 0, \quad \xi \geq 0, \quad \zeta\xi = 0.$$

This system has the unique solution $(\xi, \eta) = (0, 0)$. Thus, $\bar{\mu}$ is a noncritical multiplier (and moreover, $(\bar{x}, \bar{\mu})$ is a semistable solution).

We next show that stabilized SQP subproblems may not be solvable arbitrarily close to $(\bar{x}, \bar{\mu})$. For a given $(x^k, \mu^k) \in \mathbf{R} \times \mathbf{R}_+$, consider the stabilized SQP subproblem (7.69):

$$\begin{array}{ll} \text{minimize} & -\left(x^k - \dfrac{1}{2}(x^k)^2\right)(x - x^k) - \dfrac{1}{2}(1 - x^k)(x - x^k)^2 + \dfrac{\sigma_k}{2}(\mu^k)^2 \\ \text{subject to} & x + \sigma_k(\mu - \mu^k) \geq 0, \end{array}$$

where a possible choice of σ_k is

$$\sigma_k = \left\| \left(-x^k + \frac{1}{2}(x^k)^2 - \mu^k, \ \min\{\mu^k, \, x^k\} \right) \right\|.$$

One can directly check that for any $(x^k, \mu^k) \neq (\bar{x}, \bar{\mu})$ close enough to $(\bar{x}, \bar{\mu})$ and satisfying $\mu^k \geq 0$, the latter problem does not have any stationary points.

Indeed, the KKT system of this problem gives the relations

$$\begin{gathered} \frac{1}{2}(x^k)^2 + (1 - x^k)x + \mu = 0, \\ \mu \geq 0, \quad x + \sigma_k(\mu - \mu^k) \geq 0, \quad \mu(x + \sigma_k(\mu - \mu^k)) = 0. \end{gathered} \tag{7.136}$$

Consider first the case of $\mu = 0$. Then the first equation in (7.136) gives

$$x = -\frac{(x^k)^2}{2(1 - x^k)},$$

and hence,

$$x + \sigma_k(\mu - \mu^k) = -\frac{(x^k)^2}{2(1 - x^k)} - \sigma_k \mu^k < 0,$$

which means that the third relation in (7.136) cannot be satisfied.

On the other hand, if $\mu > 0$, then according to the complementary slackness condition (the last relation in (7.136)), the third relation in (7.136) must hold as equality. Combined with the first equation in (7.136), this gives

$$-\frac{(x^k)^2 + 2\mu}{2(1 - x^k)} + \sigma_k(\mu - \mu^k) = 0.$$

The latter further implies for (x^k, μ^k) close to $(\bar{x}, \bar{\mu})$ that

$$\mu = -\frac{\frac{1}{2}(x^k)^2 + \sigma_k(1 - x^k)\mu^k}{1 - \sigma_k + \sigma_k x^k} < 0,$$

which means that the second relation in (7.136) cannot be satisfied.

In the previous example the point $\bar{x}$ is a stationary point of the optimization problem but not a local minimizer. The next example shows that even if $\bar{x}$ is a minimizer, noncriticality still cannot be used instead of the second-order sufficient optimality condition for problems with inequality constraints, in general.

Example 7.27. Let $n = 1$, $m = 2$, $f(x) = -x^2/2$, $g(x) = (-x,\, x^3/6)$. Problem (7.64) with this data has the unique feasible point (hence, the unique solution) $\bar{x} = 0$. This point is stationary, and $\mathcal{M}(\bar{x}) = \{\mu \in \mathbf{R}^2_+ \mid \mu_1 = 0\}$, $C(\bar{x}) = \mathbf{R}_+$. Consider the multiplier $\bar{\mu} = 0$. It can be checked analyzing (7.135) (or by applying condition (1.80)) that $\bar{\mu}$ is a noncritical multiplier (while $(\bar{x}, \bar{\mu})$ is not semistable, since $\mathcal{M}(\bar{x})$ is not a singleton).

For a given $(x^k, \mu^k) \in \mathbf{R} \times \mathbf{R}^2_+$, the stabilized SQP subproblem (7.69) has the form

$$\begin{array}{ll} \text{minimize} & -x^k(x - x^k) - \dfrac{1}{2}(x - x^k)^2 + \dfrac{\sigma_k}{2}(\mu_1^2 + \mu_2^2) \\ \text{subject to} & x + \sigma_k(\mu_1 - \mu_1^k) \geq 0, \\ & -\dfrac{1}{3}(x^k)^3 + \dfrac{1}{2}(x^k)^2 x - \sigma_k(\mu_2 - \mu_2^k) \leq 0, \end{array} \tag{7.137}$$

where

$$\sigma_k = \left\| \left(-x^k - \mu_1^k + \frac{1}{2}(x^k)^2\mu_2^k,\, \min\{\mu_1^k,\, x^k\},\, \min\left\{\mu_2^k,\, -\frac{1}{6}(x^k)^3\right\} \right) \right\|.$$

Take $x^k > 0$, $\mu_1^k > 0$, $\mu_2^k = 0$. Then

$$\sigma_k = \left\| \left(x^k + \mu_1^k,\, \min\{\mu_1^k,\, x^k\},\, -\frac{1}{6}(x^k)^3 \right) \right\|,$$

and there evidently exists a constant $\beta > 0$ (not depending on x^k and μ_1^k) such that the inequality

$$\sigma_k \geq \beta(x^k + \mu_1^k) \tag{7.138}$$

holds for all such x^k and μ^k.

The KKT system of problem (7.137) gives the relations

$$\begin{gathered}
-x-\mu_1+\frac{1}{2}(x^k)^2\mu_2=0,\\
\mu_1\geq 0,\quad x+\sigma_k(\mu_1-\mu_1^k)\geq 0,\quad \mu_1(x+\sigma_k(\mu_1-\mu_1^k))=0,\\
\mu_2\geq 0,\quad -\frac{1}{3}(x^k)^3+\frac{1}{2}(x^k)^2x-\sigma_k\mu_2\leq 0,\\
\mu_2\left(-\frac{1}{3}(x^k)^3+\frac{1}{2}(x^k)^2x-\sigma_k\mu_2\right)=0.
\end{gathered} \tag{7.139}$$

Consider first the case of $\mu_1=\mu_2=0$. Then the first equation in (7.139) gives $x=0$, and hence,

$$x+\sigma_k(\mu_1-\mu_1^k)=-\sigma_k\mu_1^k<0,$$

which means that the second relation in the second line of (7.139) cannot be satisfied.

Furthermore, if $\mu_1=0$, $\mu_2>0$, then according to the complementary slackness condition (the last relation in the last line of (7.139)), the second relation in the last line of (7.139) must hold as equality. Combined with the first equation in (7.139), this gives

$$\left(-\sigma_k+\frac{1}{4}(x^k)^4\right)\mu_2=\frac{1}{3}(x^k)^3. \tag{7.140}$$

By (7.138) we derive

$$-\sigma_k+\frac{1}{4}(x^k)^4\leq -\beta(x^k+\mu_1^k)+\frac{1}{4}(x^k)^4<0$$

provided x^k is small enough, which means that (7.140) may hold only with $\mu_2<0$. Thus, the first relation in the last line of (7.139) cannot be satisfied.

If $\mu_1>0$, $\mu_2=0$, then the first equation in (7.139) gives $x=-\mu_1$, and hence,

$$x+\sigma_k(\mu_1-\mu_1^k)=-(1-\sigma_k)\mu_1-\sigma_k\mu_1^k<0$$

provided x^k and μ_1^k are small enough. Thus, the second relation in the second line of (7.139) cannot be satisfied in this case.

Finally, if $\mu_1>0$ and $\mu_2>0$, then employing the first equation in (7.139) and the complementary slackness conditions (the last relations in the last two lines of (7.139)) we obtain the equations

$$x=-\mu_1+\frac{1}{2}(x^k)^2\mu_2,\quad x=-\sigma_k(\mu_1-\mu_1^k),\quad x=\frac{2}{3}x^k+\frac{2\sigma_k}{(x^k)^2}\mu_2=0.$$

Solving this linear system, we obtain

$$\mu_2 = -\frac{(x^k)^2\left(\frac{2}{3}x^k - \sigma_k\mu_1^k - \sigma_k x^k\right)}{\sigma_k\left(2(1-\sigma_k) + \frac{1}{2}(x^k)^4\right)}.$$

Let, e.g., $\mu_1^k = x^k$. Then

$$\mu_2 = -\frac{(x^k)^2\left(\frac{2}{3}x^k + o(x^k)\right)}{\sigma_k\left(2(1-\sigma_k) + \frac{1}{2}(x^k)^4\right)} < 0$$

provided x^k is small enough, and thus, the first relation in the last line of (7.139) cannot be satisfied.

Summarizing, we have demonstrated that for $x^k = \mu_1^k > 0$ small enough, and for $\mu_2^k = 0$, problem (7.137) does not have any stationary points.

7.2.3 Other Approaches and Further Developments

We proceed with a brief survey of some other approaches to degenerate problems, different from the stabilized SQP method. We also indicate what we believe might be important directions of further developments in this field. We shall concentrate on optimization problems violating some standard constraint qualifications, though the ideas discussed below can also be applicable in the more general variational setting.

Consider the mathematical programming problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0,\ g(x) \leq 0, \end{array} \tag{7.141}$$

where the objective function $f : \mathbf{R}^n \to \mathbf{R}$ and the constraints mappings $h : \mathbf{R}^n \to \mathbf{R}^l$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are twice differentiable.

Let $\bar{x} \in \mathbf{R}^n$ be a stationary point of (7.141) and $\mathcal{M}(\bar{x})$ be the (nonempty) set of associated Lagrange multipliers, that is, the set of $(\lambda, \mu) \in \mathbf{R}^l \times \mathbf{R}^m$ satisfying the KKT system

$$\frac{\partial L}{\partial x}(x, \lambda, \mu) = 0, \quad h(x) = 0, \quad \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu, g(x)\rangle = 0,$$

for $x = \bar{x}$. Here $L : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}$ is the Lagrangian of problem (7.141):

$$L(x, \lambda, \mu) = f(x) + \langle \lambda, h(x)\rangle + \langle \mu, g(x)\rangle.$$

Recall that the SOSC holds at $\bar{x}$ for $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ if

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{7.142}$$

where

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A(\bar{x})}(\bar{x})\xi \leq 0,\ \langle f'(\bar{x}),\, \xi\rangle \leq 0\}$$

is the critical cone of problem (7.141) at $\bar{x}$. The SSOSC holds if

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in C_+(\bar{x},\, \bar{\mu}) \setminus \{0\}, \tag{7.143}$$

where

$$C_+(\bar{x},\, \bar{\mu}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+(\bar{x},\, \bar{\mu})}(\bar{x})\xi = 0\}.$$

An iteration of the method proposed in [84] consists of two steps. First, the primal iterate is obtained by solving the usual SQP subproblem. Then the dual iterate is computed by solving another auxiliary quadratic program whose purpose is dual stabilization. Specifically, for the current primal-dual iterate $(x^k,\, \lambda^k,\, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, the next primal iterate x^{k+1} is given by solving

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, x - x^k\rangle + \frac{1}{2}\langle H_k(x - x^k),\, x - x^k\rangle \\ \text{subject to} & h(x^k) + h'(x^k)(x - x^k) = 0,\ g(x^k) + g'(x^k)(x - x^k) \leq 0, \end{array} \tag{7.144}$$

with

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k).$$

Afterwards, the next dual iterate $(\lambda^{k+1},\, \mu^{k+1})$ is computed as a Lagrange multiplier associated with the solution of the same problem (7.144) but with $H_k = I$.

This approach is evidently different from stabilized SQP. First, unlike in stabilized SQP, subproblems of the form (7.144) can be infeasible. Second, there is no stabilization parameter here tending to zero along the iterations. Nevertheless, computing $(\lambda^{k+1},\, \mu^{k+1})$ becomes an ill-conditioned problem as x^k approaches a degenerate solution. In addition, convergence theory in [84] relies on rather strong assumptions (much stronger than those for the stabilized SQP method). We discuss this next.

For a stationary point $\bar{x}$ of problem (7.141), define the index set

$$A_+(\bar{x}) = \bigcup_{\bar{\mu} \in \mathcal{M}(\bar{x})} A_+(\bar{x},\, \bar{\mu}).$$

Superlinear convergence of a sequence $\{(x^k,\, \lambda^k,\, \mu^k)\}$ generated by the method in question to the set $\{\bar{x}\} \times \mathcal{M}(\bar{x})$ can be established for any starting point $(x^0,\, \lambda^0,\, \mu^0)$ close enough to $\{\bar{x}\} \times \mathcal{M}(\bar{x})$ under the MFCQ and the

SOSC (7.142) for all multipliers $\bar{\mu} \in \mathcal{M}(\bar{x})$, and with the additional assumptions that rank $g'_{A_+(\bar{x})}(x)$ is the same for all $x \in \mathbf{R}^n$ close enough to $\bar{x}$, and

$$\operatorname{im}(g'_{A_+(\bar{x},\,\bar{\mu})}(\bar{x}))^{\mathrm{T}} = \operatorname{im}(g'_{A_+(\bar{x})}(\bar{x}))^{\mathrm{T}} \quad \forall\, \bar{\mu} \in \mathcal{M}(\bar{x}).$$

The last assumption is automatic if there exists $(\bar{\lambda},\, \bar{\mu}) \in \mathcal{M}(\bar{x})$ satisfying the strict complementarity condition $\bar{\mu}_{A(\bar{x})} > 0$.

It is worth to mention that the method of [84] outlined above can also be embedded into the perturbed SQP framework, at least for the purposes of a posteriori analysis; see [270].

Furthermore, it was observed in [270] that for some classes of degenerate problems professional implementations of SQP methods (without any kind of dual stabilization) can actually be quite successful; see, e.g., the discussion of these issues for mathematical programs with complementarity constraints in Sect. 7.3. Some kind of explanation of this was suggested in [270], employing the inexact SQP framework therein. In particular, it was argued that warm-starting the active-set QP solvers (i.e., using the set of active constraints of the SQP subproblem (7.144) identified at the previous iteration to initialize the new one) can be of crucial importance. Another point of significance is that QP solvers used in professional implementations allow for slight violation of inequality constraints that are being enforced as equalities according to the current active set estimation. In [270], these and other subtle points of the available implementations of SQP methods are formalized to produce a new algorithm with local superlinear convergence to $\{\bar{x}\} \times \mathcal{M}(\bar{x})$ under the MFCQ and the SOSC (7.142) for all multipliers $\bar{\mu} \in \mathcal{M}(\bar{x})$, and some additional requirements.

However, it should be kept in mind that the MFCQ cannot hold for degenerate equality-constrained problems or for problems with complementarity constraints. Also, assuming the SOSC for all Lagrange multipliers is too restrictive in any case. Along with stabilized SQP, there exist some other techniques that allow to remove or relax these assumptions. One of them is given in [141], employing the ideas of active-set methods discussed in Sects. 3.4 and 4.1.3. We outline this approach next.

Recall that given a point $(x^0,\, \lambda^0,\, \mu^0)$ close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, where $(\bar{\lambda},\, \bar{\mu}) \in \mathcal{M}(\bar{x})$ is a noncritical multiplier, Proposition 3.58 provides tools to locally identify the active index set $A(\bar{x})$. Recall also that with this identification at hand, the usual active-set approach would apply some Newton-type method to the Lagrange system

$$f'(x) + (h'(x))^{\mathrm{T}}\lambda + (g'_{A(\bar{x})}(x))^{\mathrm{T}}\mu_{A(\bar{x})} = 0,\ h(x) = 0,\ g_{A(\bar{x})}(x) = 0 \quad (7.145)$$

of the equality-constrained problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0,\ g_{A(\bar{x})}(x) = 0. \end{array} \quad (7.146)$$

In the current degenerate setting, the difficulty, of course, is that the modified problem (7.146) is also degenerate. The idea in [141] is to apply a Newton-type method not to the Lagrange system (7.145), but to its regularized version whose construction was originally suggested in [126].

Specifically, set $q = l + |A(\bar{x})|$, and define the mapping $H : \mathbf{R}^n \to \mathbf{R}^q$, $H(x) = (h(x),\, g_{A(\bar{x})}(x))$. Then the Lagrange system of problem (7.145) can be written in the form

$$f'(x) + (H'(x))^{\mathrm{T}}\theta = 0, \quad H(x) = 0, \tag{7.147}$$

where $\theta = (\lambda,\, \mu_{A(\bar{x})}) \in \mathbf{R}^q$. Recall that $\bar{\theta} = (\bar{\lambda},\, \bar{\mu}_{A(\bar{x})})$ is a Lagrange multiplier associated with the stationary point $\bar{x}$ of problem (7.146). Let $P : \mathbf{R}^n \times \mathbf{R}^q \to \mathbf{R}^{q\times q}$ be a mapping continuous at $(\bar{x},\, \bar{\theta})$, such that

$$P(\bar{x},\, \bar{\theta}) = \pi_{(\mathrm{im}\, H'(\bar{x}))^{\perp}}, \tag{7.148}$$

and such that $P(x,\, \theta)$ is a projector, i.e.,

$$(P(x,\, \theta))^2 = P(x,\, \theta) \quad \forall\, (x,\, \theta) \in \mathbf{R}^n \times \mathbf{R}^q. \tag{7.149}$$

Set $P_0 = P(x^0,\, \theta^0)$, where $\theta^0 = (\lambda^0,\, \mu^0_{A(\bar{x})})$. Then the regularized Lagrange system is defined as follows:

$$f'(x) + (H'(x))^{\mathrm{T}}\theta = 0, \quad H(x) + P_0(\theta - \theta^0) = 0. \tag{7.150}$$

The proposal is to apply some Newton-type method to this system, starting from the point $(x^0,\, \theta^0)$. We next give the justification. How to construct the needed P_0 shall be discussed later on.

Proposition 7.28. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $H : \mathbf{R}^n \to \mathbf{R}^q$ be differentiable at $\bar{x} \in \mathbf{R}^n$. Let $\bar{x}$ be a stationary point of problem*

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & H(x) = 0, \end{array} \tag{7.151}$$

and let $\bar{\theta} \in \mathbf{R}^q$ be a Lagrange multiplier associated with $\bar{x}$. Let the mapping $P : \mathbf{R}^n \times \mathbf{R}^q \to \mathbf{R}^{q\times q}$ be continuous at $(\bar{x},\, \bar{\theta})$ and satisfy (7.148), (7.149).

Then for every $(x^0,\, \theta^0) \in \mathbf{R}^n \times \mathbf{R}^q$ close enough to $(\bar{x},\, \bar{\theta})$, there exists the unique element $\theta^* = \theta^*(x^0,\, \theta^0) \in \mathbf{R}^l \times \mathbf{R}^q$ such that $(\bar{x},\, \theta^*)$ is a solution of the system (7.150) with $P_0 = P(x^0,\, \theta^0)$ (which means that θ^* is the unique Lagrange multiplier associated with the stationary point $\bar{x}$ of problem (7.151), satisfying the additional requirement $P_0(\theta^* - \theta^0) = 0$)), and

$$\theta^*(x^0,\, \theta^0) \to \bar{\theta} \text{ as } x^0 \to \bar{x},\ \theta^0 \to \bar{\theta}. \tag{7.152}$$

The additional requirement $P_0(\theta - \theta^0) = 0$ imposed on multipliers is the essence of dual stabilization in this approach.

Proof. Consider the parametric linear system

$$(H'(\bar{x}))^{\mathrm{T}}\theta = -f'(\bar{x}), \quad \pi_{(\operatorname{im} H'(\bar{x}))^{\perp}}\pi\theta = \pi_{(\operatorname{im} H'(\bar{x}))^{\perp}}\pi\theta^0, \tag{7.153}$$

where $\theta \in \mathbf{R}^q$ is the variable, and $\theta^0 \in \mathbf{R}^q$ and projector $\pi \in \mathbf{R}^{q\times q}$ are parameters. One can think of the image space of the operator of this system as the q-dimensional space $\operatorname{im}(H'(\bar{x}))^{\mathrm{T}} \times (\operatorname{im} H'(\bar{x}))^{\perp}$ (recall the fact that $(\operatorname{im} H'(\bar{x}))^{\perp} = \operatorname{im} \pi_{(\operatorname{im} H'(\bar{x}))^{\perp}}$). Note also that the right-hand side of (7.153) certainly belongs to $\operatorname{im}(H'(\bar{x}))^{\mathrm{T}} \times (\operatorname{im} H'(\bar{x}))^{\perp}$, since $-f'(\bar{x}) = (H'(\bar{x}))^{\mathrm{T}}\bar{\theta}$.

With this understanding, we first show that for the choice defined by $\pi = \pi_{(\operatorname{im} H'(\bar{x}))^{\perp}}$, the operator of the system (7.153) mapping $\mathbf{R}^q$ to another q-dimensional space, is nonsingular. Consider an arbitrary $\eta \in \mathbf{R}^q$ satisfying

$$(H'(\bar{x}))^{\mathrm{T}}\eta = 0, \quad (\pi_{(\operatorname{im} H'(\bar{x}))^{\perp}})^2\eta = 0.$$

Taking into account the equality $(\pi_{(\operatorname{im} H'(\bar{x}))^{\perp}})^2 = \pi_{(\operatorname{im} H'(\bar{x}))^{\perp}}$, from the second equation it follows that $\eta \in \ker \pi_{(\operatorname{im} H'(\bar{x}))^{\perp}} = \operatorname{im} H'(\bar{x})$. On the other hand, the first equation can be written as $\eta \in \ker(H'(\bar{x}))^{\mathrm{T}} = (\operatorname{im} H'(\bar{x}))^{\perp}$. We conclude that $\eta = 0$, which establishes the assertion.

Next, note that the system (7.153) with $\pi = \pi_{(\operatorname{im} H'(\bar{x}))^{\perp}}$ and $\theta^0 = \bar{\theta}$ has a solution $\bar{\theta}$, and according to what is already established, this solution is unique. Now, according to Lemma A.6, we conclude that for every element $(\pi, \theta^0) \in \mathbf{R}^{q\times q} \times \mathbf{R}^q$ close enough to $(\pi_{(\operatorname{im} H'(\bar{x}))^{\perp}}, \bar{\theta})$, the system (7.153) has the unique solution θ^* and, moreover, this solution tends to $\bar{\theta}$ as π tends to $\pi_{(\operatorname{im} H'(\bar{x}))^{\perp}}$ and θ^0 tends to $\bar{\theta}$.

Furthermore, taking into account the equality $\pi^2 = \pi$ (which holds since π is a projector), from the second equation in (7.153) it follows that

$$\begin{aligned}
0 &= \pi_{(\operatorname{im} H'(\bar{x}))^{\perp}}\pi(\lambda^* - \lambda^0)\\
&= \pi^2(\lambda^* - \lambda^0) + (\pi_{(\operatorname{im} H'(\bar{x}))^{\perp}} - \pi)\pi(\lambda^* - \lambda^0)\\
&= (I + (\pi_{(\operatorname{im} H'(\bar{x}))^{\perp}} - \pi))\pi(\lambda^* - \lambda^0),
\end{aligned}$$

where, according to Lemma A.6, the matrix $I + (\pi_{(\operatorname{im} H'(\bar{x}))^{\perp}} - \pi)$ is nonsingular provided that π is close enough to $\pi_{(\operatorname{im} H'(\bar{x}))^{\perp}}$. Hence, for such π it must hold that $\pi(\lambda^* - \lambda^0) = 0$. Combined with the first equation in (7.153), this equality means that $(\bar{x}, \lambda^*)$ is a solution of the system

$$f'(x) + (H'(x))^{\mathrm{T}}\theta = 0, \quad H(x) + \pi(\theta - \theta^0) = 0. \tag{7.154}$$

Suppose now that there exists some other $\theta \in \mathbf{R}^q$ such that $(\bar{x}, \theta)$ also solves (7.154). Then θ is a Lagrange multiplier associated with the stationary point $\bar{x}$ of problem (7.151), and $\pi(\theta - \theta^0) = 0$. This evidently implies that $(\bar{x}, \theta)$ satisfies (7.153). But the solution of the latter system is unique, as established above. Hence, $(\bar{x}, \theta^*)$ is the unique solution of (7.154).

To complete the proof, it remains to recall (7.148), (7.149), and the continuity of P at $(\bar{x}, \bar{\theta})$: these properties imply that P_0 is a projector which can be made arbitrarily close to $\pi_{(\operatorname{im} H'(\bar{x}))^\perp}$ by taking $(x^0, \theta^0) \in \mathbf{R}^n \times \mathbf{R}^q$ close enough to $(\bar{x}, \bar{\theta})$. $\square$

Proposition 7.29. *In addition to the assumptions of Proposition 7.28, let the multiplier $\bar{\theta} \in \mathbf{R}^q$ be noncritical.*

Then for every $(x^0, \theta^0) \in \mathbf{R}^n \times \mathbf{R}^q$ close enough to $(\bar{x}, \bar{\theta})$, the Jacobian of the system (7.150) with $P_0 = P(x^0, \theta^0)$ is nonsingular at its solution $(\bar{x}, \theta^*)$, where $\theta^* = \theta^*(x^0, \theta^0)$ is defined in Proposition 7.28.

Proof. Suppose that there exists a sequence $\{(x^k, \theta^k)\} \subset \mathbf{R}^n \times \mathbf{R}^q$ convergent to $(\bar{x}, \bar{\theta})$ and such that for each k the Jacobian of the system (7.150) is singular at $(\bar{x}, \theta^*(x^k, \theta^k))$. This means that for each index k, there exists $(\xi^k, \eta^k) \in \mathbf{R}^n \times \mathbf{R}^q$, $(\xi^k, \eta^k) \neq 0$, such that

$$f''(\bar{x})\xi^k + (H''(\bar{x})[\xi^k])^{\mathrm{T}}\theta^*(x^k, \theta^k) + (H'(\bar{x}))^{\mathrm{T}}\eta^k = 0,$$

$$H'(\bar{x})\xi^k + P(x^k, \theta^k)\eta^k = 0.$$

Normalizing (ξ^k, η^k), and passing onto the limit along the appropriate subsequence, we obtain the existence of $(\xi, \eta) \in \mathbf{R}^n \times \mathbf{R}^q$, $(\xi, \eta) \neq 0$, such that

$$\begin{array}{c} f''(\bar{x})\xi + (H''(\bar{x})[\xi])^{\mathrm{T}}\bar{\theta} + (H'(\bar{x}))^{\mathrm{T}}\eta = 0, \\ H'(\bar{x})\xi + \pi_{(\operatorname{im} H'(\bar{x}))^\perp}\eta = 0, \end{array} \tag{7.155}$$

where (7.148), the continuity of P at $(\bar{x}, \bar{\theta})$, and (7.152) were taken into account. The second equality in (7.155) is evidently equivalent to the pair of equalities

$$H'(\bar{x})\xi = 0, \quad \pi_{(\operatorname{im} H'(\bar{x}))^\perp}\eta = 0. \tag{7.156}$$

If $\xi \neq 0$, then the first equalities in (7.155) and (7.156) contradict the assumption that $\bar{\theta}$ is a noncritical multiplier.

Assuming now that $\xi = 0$, the first equality in (7.155) yields

$$(H'(\bar{x}))^{\mathrm{T}}\eta = 0,$$

which means that $\eta \in \ker(H'(\bar{x}))^{\mathrm{T}} = (\operatorname{im} H'(\bar{x}))^\perp$. At the same time, the second equality in (7.156) means that $\eta \in \ker \pi_{(\operatorname{im} H'(\bar{x}))^\perp} = \operatorname{im} H'(\bar{x})$. It follows that $\eta = 0$, which again gives a contradiction. $\square$

Assuming that the second derivatives of f and of H are continuous at the point $\bar{x}$, Theorem 2.2 and Proposition 7.29 ensure that for every element $(x^0, \theta^0) \in \mathbf{R}^n \times \mathbf{R}^q$ close enough to $(\bar{x}, \bar{\theta})$, the starting point (x^0, θ^0) uniquely defines the sequence $\{(x^k, \theta^k)\} \subset \mathbf{R}^n \times \mathbf{R}^q$ of the Newton method applied to system (7.150) with $P_0 = P(x^0, \theta^0)$, and this sequence converges to $(\bar{x}, \theta^*(x^0, \theta^0))$ superlinearly.

To obtain an implementable algorithm, it remains to discuss how one can go about constructing a mapping P with the required properties in practice.

Let $\Psi : \mathbf{R}^n \times \mathbf{R}^q \to \mathbf{R}^n \times \mathbf{R}^q$ be the residual mapping of the Lagrange system (7.147), that is,

$$\Psi(x,\, \theta) = (f'(x) + (H'(x))^{\mathrm{T}}\theta,\, H(x)).$$

For a given $(x,\, \theta) \in \mathbf{R}^n \times \mathbf{R}^q$, the needed $P(x,\, \theta)$ can be defined using the singular-value decomposition of the Jacobian $H'(x)$, provided the error bound

$$x - \bar{x} = O(\|\Psi(x,\, \theta)\|) \tag{7.157}$$

holds as $(x,\, \theta) \in \mathbf{R}^n \times \mathbf{R}^q$ tends to $(\bar{x},\, \bar{\theta})$. This construction of P actually has much in common with the procedure for identification of active constraints in Proposition 3.58 and Remark 3.59. For instance, fix any $\tau \in (0,\, 1)$, and let $\sigma_j(x)$, $j = 1,\, \ldots,\, s$, $\sigma_1(x) \geq \sigma_2(x) \geq \ldots \geq \sigma_s(x) \geq 0$, be the singular values of $H'(x)$, where $s = \min\{n,\, q\}$. Assuming (7.157), and employing some facts from the perturbation theory for linear operators [167], the following claims can be established. First, for any $(x,\, \theta) \in \mathbf{R}^n \times \mathbf{R}^q$ close enough to $(\bar{x},\, \bar{\theta})$, the integer

$$r(x,\, \theta) = \max\{0,\, \max\{j = 1,\, \ldots,\, s \mid \sigma_j(x) > \|\Psi(x,\, \theta)\|^{\tau}\}\}$$

coincides with $\operatorname{rank} H'(\bar{x})$. Second, if $P(x,\, \theta)$ is defined as the orthogonal projector onto the linear subspace in $\mathbf{R}^q$ spanned by the last $q - r(x,\, \theta)$ left singular vectors of $H'(x)$, then the mapping $P : \mathbf{R}^n \times \mathbf{R}^q \to \mathbf{R}^{q\times q}$ thus defined satisfies all the needed properties. We refer to [141] for details.

Granted, singular-value decomposition is a rather expensive procedure. This is the drawback of this construction, e.g., in comparison with stabilized SQP. However, we emphasize that once computed on the first iteration, P_0 and $(\lambda^0,\, \mu^0_{A(\bar{x})})$ in (7.150) remain fixed over the Newton iterations applied to (7.150). In particular, there is no stabilization parameter that needs to tend to zero in this approach, and thus the conditioning does not deteriorate along the iterations. On the other hand, any globalization strategy when coupled with this local algorithm may require many singular-value decompositions before the local phase is successfully activated. Therefore, the development of cheaper constructions for P would certainly be of interest.

By Proposition 1.43, the error bound (7.157) is implied by the assumption that $\bar{\theta}$ is a noncritical Lagrange multiplier for problem (7.151), which is also assumed in Proposition 7.29. This assumption is further equivalent to saying that $(\bar{\lambda},\, \bar{\mu})$ is a noncritical Lagrange multiplier for problem (7.141) with respect to the index set $A(\bar{x})$ (see Definition 7.8).

Putting all the ingredients together, and assuming that the second derivatives of f, h, and g are continuous at $\bar{x}$, local superlinear convergence of the method from [141] is established under the assumptions that $(\bar{\lambda},\, \bar{\mu})$ is noncritical in the sense of Definition 1.41, and that it is noncritical with respect to $A(\bar{x})$ in the sense of Definition 7.8. Specifically, under these assumptions, for every $(x^0,\, \lambda^0,\, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$, the method uniquely defines the sequence $\{(x^k,\, \lambda^k,\, \mu^k_{A(\bar{x})})\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^{|A(\bar{x})|}$ (where $(\lambda^k,\, \mu^k_{A(\bar{x})}) = \theta^k$), and this sequence converges superlinearly to $(\bar{x},\, \lambda^*,\, \mu^*_{A(\bar{x})})$, where $(\lambda^*,\, \mu^*_{A(\bar{x})}) \in \mathbf{R}^l \times \mathbf{R}^{|A(\bar{x})|}$ satisfies the first equality in (7.145) with $x = \bar{x}$. Also, it holds that $(\lambda^*,\, \mu^*_{A(\bar{x})})$ tends to $(\bar{\lambda},\, \bar{\mu}_{A(\bar{x})})$ as $(x^0,\, \lambda^0,\, \mu^0)$ tends to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$.

The two conditions that $(\bar{\lambda},\, \bar{\mu})$ is noncritical and that it is noncritical with respect to $A(\bar{x})$ become the same if the strict complementarity condition holds. Also, both conditions are implied by the SOSC (7.142) (see the discussion following Definition 7.8). Hence, the local superlinear convergence property of the method in question is valid under the sole assumption of the SOSC (7.142), just as for the stabilized SQP method.

In [180], it is suggested to identify some index set A satisfying the relations $A_+(\bar{x},\, \bar{\mu}) \subset A \subset A(\bar{x})$ for some multiplier $\bar{\mu} \in \mathcal{M}(\bar{x})$, and to apply the stabilized SQP (which can also be called here the stabilized Newton–Lagrange) method to the equality-constrained problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0,\ g_A(x) = 0. \end{array}$$

However, the identification procedure in [180] is fairly primitive, and the assumptions ensuring local superlinear convergence are again rather restrictive. Specifically, it is assumed that

$$\exists\, \bar{\xi} \in \ker h'(\bar{x}) \text{ such that } g'_{A(\bar{x})}(\bar{x})\bar{\xi} < 0,$$

which differs from the MFCQ only by the absence of the regularity condition $\operatorname{rank} h'(\bar{x}) = l$ for the equality constraints, and it is assumed further that the SSOSC (7.143) holds for all the multipliers $(\bar{\lambda},\, \bar{\mu}) \in \mathcal{M}(\bar{x})$.

Finally, [272] employs the same procedure for identification of active constraints as in [141] (the one in Proposition 3.58). But instead of the regularization of the Lagrange system as in (7.145), [272] applies the stabilized Newton–Lagrange method to problem (7.146) (similarly to [180]).

As discussed above, identification is locally correct provided the starting point is close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ with a noncritical multiplier $(\bar{\lambda},\, \bar{\mu}) \in \mathcal{M}(\bar{x})$. According to Theorem 7.25, local superlinear convergence of the stabilized Newton–Lagrange method for problem (7.146) is ensured by assuming that $(\bar{\lambda},\, \bar{\mu})$ is noncritical with respect to $A(\bar{x})$. Therefore, local superlinear convergence of the method in [272] is ensured under the same assumptions as for the method in [141].

Apart from the local convergence theory, [272] suggests a system of tests to verify correctness of identification of the active set $A(\bar{x})$. These tests allow to combine the local algorithm with reasonable "outer strategies" which, in principle, is supposed to lead to globally convergent algorithms with enhanced rate of convergence in the cases of degeneracy. The advantages of this approach over [141] are the following. It appears that coupling of the local phase with some globally convergent outer strategy can be done more naturally, and there is no need for computationally costly procedures (like the singular-value decomposition) to construct the regularized Lagrange system. Both approaches are QP-free: their iteration subproblems are linear systems rather than quadratic programs. Moreover, the linear systems in [272] are closer in structure to those that arise in the Newton–Lagrange method or in primal-dual interior point methods; in particular, this structure preserves sparsity. On the other hand, unlike in [141], the subproblems of the method in [272] become ill-conditioned close to a solution, which is a disadvantage of that approach (at least theoretically; regarding the practical aspect, see [269, 272] for the analysis of the effect of roundoff errors).

The starting point of the local algorithms in [141, 272] is the reduction of problem (7.141) to an equality-constrained problem via identification of the active set, with the subsequent use of some stabilization technique. An alternative way to obtain an equality-constrained problem is to introduce slack variables $z \in \mathbf{R}^m$, and transform (7.141) into

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \quad g(x) + z^2 = 0. \end{array} \tag{7.158}$$

Local solutions of (7.141) and (7.158) are in an obvious correspondence with each other: $\bar{x}$ is a local solution of (7.141) if and only if $(\bar{x}, \bar{z})$ with $\bar{z} = \sqrt{-g(\bar{x})}$ is a well-defined local solution of (7.158). The slacks approach may look attractive because of its universality and simplicity. However, as often happens with such simple transformations, it requires stronger assumptions. For instance, even though any $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ is also a valid Lagrange multiplier associated with the stationary point $(\bar{x}, \bar{z})$ of problem (7.158), it is easy to see that this multiplier can be noncritical in (7.158) only when $(\bar{\lambda}, \bar{\mu})$ satisfies the strict complementarity condition in (7.141).

All the approaches to deal with degeneracy discussed so far are essentially local. Some of them, like the stabilized SQP, possess really impressive local convergence properties which, arguably speaking, would be hard to improve with any new Newtonian method. Given that the methods are local, the key to further developments is globalization of convergence. For stabilized SQP this would entail constructing an appropriate merit/penalty function whose descent can be ensured via linesearch in the stabilized SQP direction, or via adjusting the trust-region constraint. Some globalizations of stabilized SQP are proposed in [78] and [99], but these do not seem to be completely satisfactory. The difficulty is precisely the lack of a merit function that fits naturally with stabilized SQP (like the l_1-penalty function fits the usual SQP, recall Sect. 5.1). Given this state of affairs, an attractive possibility for globalization is the filter technique of Sect. 6.3.2 (or perhaps some other variant of it), avoiding the merit function issues.

For the methods of [141, 272], the key for improvements is effective (in practice) identification of active constraints and computationally cheaper construction of the regularized system for the method in [141]. At this time, these special local methods require a "hybrid" implementation: there is the "local phase" described above, and there must be an "outer phase." The latter is some conventional well-established method with good global convergence properties, playing the role of a safeguard. This can be done, for example, along the lines of hybrid globalization strategies discussed in Sect. 5.3 for active-set Newton methods for complementarity problems. The "local phase" should be activated only when this appears to be justified, i.e., when some (indirect) indications of degeneracy are observed. Otherwise, the "local phase" should not interfere with the conventional algorithm. The switching rules must be designed in a way for the "additions" to the conventional algorithms not to increase their iteration costs significantly. This is, of course, a matter of art of implementation, not an easy one.

Yet another difficulty in the degenerate case is that standard primal-dual Newton-type algorithms (the ones discussed in Sect. 7.1) suffer from attraction to critical Lagrange multipliers when such exist. If the "outer phase" steers the dual trajectory towards a critical multiplier, then the dual iterates may never get close enough to multipliers with the properties needed for local superlinear convergence of the special method employed in the "local phase." This means that the "local phase" may either never get activated or may not work properly. The latter is the global aspect of the negative consequences of attraction to critical multipliers. Therefore, it may be not enough to suppress this effect locally, and some kind of global dual stabilization procedures are also needed.

Another technique useful in the context of degenerate problems is the elastic mode discussed in Sect. 6.2. The constraints of the iteration subproblems are relaxed introducing a slack variable, with the relaxation penalized in the objective function. For the subsequent discussion, we introduce first the following relaxation of the original problem (7.141), given by

$$\begin{array}{ll} \text{minimize} & f(x) + ct \\ \text{subject to} & -te \leq h(x) \leq te,\ g(x) \leq te,\ t \geq 0, \end{array} \tag{7.159}$$

where $c \geq 0$ is the penalty parameter, and e is the vector of ones of appropriate dimension. We can associate with problem (7.159) its quadratic programming approximation

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, x - x^k\rangle + ct + \frac{1}{2}\langle H_k(x - x^k),\, x - x^k\rangle \\ \text{subject to} & -te \leq h(x^k) + h'(x^k)(x - x^k) \leq te,\ g(x^k) + g'(x^k)(x - x^k) \leq te, \\ & t \geq 0, \end{array} \tag{7.160}$$

where $H_k \in \mathbf{R}^{n\times n}$ is a symmetric matrix. This problem is in fact the elastic mode modification of the SQP subproblem (7.144). For computational purposes, other versions of the elastic mode can be used and are often more efficient. For example, instead of the common scalar variable t, each component of the constraints in (7.159) or in (7.160) can be relaxated by a separate variable, also with a separate penalty parameter associated to each.

The main purpose of the elastic mode is to handle possible infeasibility of subproblems approximating (7.141). For example, those that linearize the constraints in (7.141), like in the SQP subproblem (7.144). Possible infeasibility of subproblems is especially relevant when far from a solution, and in the absence of the MFCQ at a solution, even locally. The advantage of subproblems of the form (7.160) is that they are automatically feasible regardless of any assumptions.

Properties of the elastic mode SQP algorithm near a degenerate local solution $\bar{x}$ of problem (7.141) are analyzed in [6, 7]. The method in those works is equipped with a linesearch procedure (see Sect. 6.2) and employs $H_k = I$ for all k. With this choice of H_k this is not a Newton-type method, of course, and one cannot expect a superlinear convergence rate. The assumptions used in [6] are the MFCQ and the quadratic growth condition at a local solution $\bar{x}$ of problem (7.141), i.e., the existence of a neighborhood U of $\bar{x}$ and $\gamma > 0$ such that

$$f(x) - f(\bar{x}) \geq \gamma \|x - \bar{x}\|^2 \quad \forall\, x \in U \ \text{ such that } \ h(x) = 0,\ g(x) \leq 0.$$

Therefore both the algorithm and the assumptions are fully primal. Recall that according to Theorem 1.20, under the MFCQ the quadratic growth condition is equivalent to the SOSC in the form

$$\forall\, \xi \in C(\bar{x}) \setminus \{0\} \quad \exists\, (\bar{\lambda},\, \bar{\mu}) \in \mathcal{M}(\bar{x}) \ \text{ such that } \ \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi \right\rangle > 0. \tag{7.161}$$

As demonstrated in [6], under these assumptions $\bar{x}$ is an isolated stationary point of problem (7.141), and if a sequence generated by the method converges to $\bar{x}$, then the rate of convergence is geometric.

However, the purpose of the elastic mode when it comes to local analysis under the MFCQ is not quite clear. Close to a point satisfying the MFCQ, the linearized constraints are feasible, and it can be seen that for appropriate values of the parameter $c > 0$ the elastic mode becomes inactive, i.e., in (7.160) the solution in the variables t would be zero. Thus the elastic mode subproblem gives the same solutions as the usual SQP subproblem. The results of [6] were further developed in [7], where the MFCQ was removed. More precisely, it was shown that the assumption $\mathcal{M}(\bar{x}) \neq \emptyset$ and the stronger form of the quadratic growth condition (equivalent to the usual one in the case of the MFCQ) imply the following: for any $c \geq 0$ large enough, the point $(x,\, t) = (\bar{x},\, 0)$ is an isolated stationary point of problem (7.159), and the MFCQ and the quadratic growth condition are satisfied at this point. Therefore, the results from [6] are applicable to problem (7.159), which ensures geometric rate of convergence of the method under consideration. Recall that when $\mathcal{M}(\bar{x}) \neq \emptyset$, violation of the MFCQ is equivalent to saying that the multiplier set $\mathcal{M}(\bar{x})$ is unbounded.

Assuming that a local solution $\bar{x}$ of problem (7.141) satisfies the MFCQ and the SOSC (7.142) for all multipliers $\bar{\mu} \in \mathcal{M}(\bar{x})$, a result similar to the one in [6] is given in [10] for a slightly different variant of the SQP algorithm (with a slightly different linesearch procedure, and without the elastic mode, which is only natural in the presence of the MFCQ).

We complete this section by mentioning some literature concerned with the sequential quadratically constrained quadratic programming method (SQCQP; see Sect. 4.3.5), for problems with relaxed constraint qualifications. Recall that for the current iterate $x^k \in \mathbf{R}^n$, the method computes the next iterate x^{k+1} as a stationary point of the problem

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k),\, x - x^k\rangle + \frac{1}{2}\langle f''(x^k)(x - x^k),\, x - x^k\rangle \\ \text{subject to} & h(x^k) + h'(x^k)(x - x^k) + \frac{1}{2}h''(x^k)[x - x^k,\, x - x^k] = 0, \\ & g(x^k) + g'(x^k)(x - x^k) + \frac{1}{2}g''(x^k)[x - x^k,\, x - x^k] \leq 0. \end{array} \quad (7.162)$$

In particular, the method is fully primal.

In [9], the constraints in (7.162) are complemented by the quadratic trust-region constraint

$$\|x - x^k\|^2 \leq \delta, \quad (7.163)$$

where $\delta > 0$ is the trust-region radius. On the one hand, this makes the feasible set of the subproblem bounded. On the other hand, (7.163) formalizes the theoretical requirement for local analysis that the next iterate must be taken close enough to the current one (recall condition (4.122) in Theorem 4.36). The iterations thus defined possess some very attractive properties. Specifically, if the local solution of problem (7.141) satisfies the MFCQ and the quadratic growth condition (equivalent to the SOSC (7.161)), then for any $x^0 \in \mathbf{R}^n$ close enough to $\bar{x}$ there exists an iterative sequence $\{x^k\}$ of this method; any such sequence converges to $\bar{x}$ superlinearly, and the trust-region constraint (7.163) is inactive at x^{k+1} for every k.

Moreover, if SQCQP is applied not to the original problem (7.141) but to its elastic mode modification (7.159), one can dispense with the MFCQ. This possibility is pointed out in [7].

There is still a computational issue though with the cost of solving quadratically constrained quadratic programs. In the general case, these should be considered quite difficult. That said, in [9] the subproblems were solved by applying SQP with a satisfactory outcome. But probably the most promising is the convex case, where the subproblems can be reformulated as second-order cone programs. Significant software advances have been achieved for second-order cone programming over the last decade or so.

A survey of methods discussed in this section can be found in [132].

7.3 Mathematical Programs with Complementarity Constraints

We now discuss one special class of intrinsically degenerate optimization problems, with a variety of important applications. Namely, consider the *mathematical program with complementarity constraints* (MPCC)

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & G(x) \geq 0,\ H(x) \geq 0,\ \langle G(x),\, H(x)\rangle \leq 0, \end{array} \tag{7.164}$$

where the objective function $f : \mathbf{R}^n \to \mathbf{R}$ and the mappings $G : \mathbf{R}^n \to \mathbf{R}^m$ and $H : \mathbf{R}^n \to \mathbf{R}^m$ are assumed smooth enough. We note that "usual" equality and inequality constraints can be added to this problem setting without any substantial theoretical difficulties. Extensions of results derived for the format of (7.164) to the general case are conceptually fairly straightforward, even if they require a certain technical effort. For this reason, we prefer to analyze the simplified setting of (7.164) to keep the exposition less cumbersome, with the understanding that applications do often involve also additional "usual" constraints.

The meaning of complementarity constraints is that for each of the indices $i = 1, \ldots, m$, at least one of the corresponding components $G_i(x)$ or $H_i(x)$ must be zero, while the other must be nonnegative. This immediately reveals a certain combinatorial aspect of the underlying structure, hinting also at the associated difficulties. Note also that the last constraint in (7.164) can be expressed in a number of different ways. For example, as the set of equalities $G_i(x)H_i(x) = 0$, $i = 1, \ldots, m$, or as the set of inequalities $G_i(x)H_i(x) \leq 0$, $i = 1, \ldots, m$, or as the single equality $\langle G(x),\, H(x)\rangle = 0$, all without changing the feasible set. However, there are some reasons for avoiding the equality formulations of this constraint. For instance, using inequality formulations makes the associated set of Lagrange multipliers smaller, which might have both numerical and theoretical advantages. Also, for the inequality formulation, the linearized constraints are more likely to be consistent than if equalities are linearized.

The following simple example demonstrates well the typical structure of the MPCC feasible set, and in particular, the combinatorial aspect mentioned above, and the degeneracy of the constraints.

Example 7.30. Let $n = 2$, $m = 1$, $f(x) = \sigma_1 x_1 + \sigma_2 x_2$, $G(x) = x_1$, $H(x) = x_2$, where $\sigma \in \mathbf{R}^2$ is a parameter.

The feasible set of problem (7.164) with this data is comprised by the two rays forming the border of the nonnegative orthant in $\mathbf{R}^2$. Thus, a feasible point belongs either to one ray, or to the other, or to both (the combinatorial feature is evident). The problem has a solution if and only if $\sigma \geq 0$, in which case $\bar{x} = 0$ always belongs to the solution set. Observe that the contingent cone to the feasible set at $\bar{x} = 0$ is comprised by the same two rays (it coincides with the feasible set). In particular, this cone is not convex. This means, among other things, that the MFCQ does not hold at $\bar{x}$ (see Sect. 1.1). Hence, if there exist Lagrange multipliers associated with the solution $\bar{x}$, then the set of multipliers is necessarily unbounded (see Sect. 1.2). It turns out that this state of affairs is actually intrinsic for the whole problems class, i.e., those properties hold for any MPCC.

Note further that $\bar{x} = 0$ is the unique solution if and only if $\sigma > 0$. If $\sigma_1 = 0$, $\sigma_2 > 0$, then the solutions have the form $\bar{x} = (\bar{x}_1, 0)$ with $\bar{x}_1 \geq 0$. If $\sigma_1 > 0$, $\sigma_2 = 0$, then the solutions have the form $\bar{x} = (0, \bar{x}_2)$ with $\bar{x}_2 \geq 0$. Finally, if $\sigma = 0$, then all the feasible points are solutions.

MPCC is perhaps one of the most important instances of a *mathematical program with equilibrium constraints* (MPEC), where the feasible set is defined by a parametric variational inequality. We address the reader to [68, 184, 211] for numerous applications of MPECs. Many applications arise from the so-called bilevel optimization problems, including Stackelberg games modelling hierarchical market structures (see [40, 56]). In these problems, feasible set is defined as the solution set of a parametric lower-level optimization problem. Replacing the latter by its KKT optimality system leads to the MPCC setting. Among other applications, we mention optimization of elastoplastic mechanical structures of a given topology [79].

We proceed with some elements of the MPCC theory, needed for the subsequent discussions of Newton-type methods for problems of this class.

7.3.1 Theoretical Preliminaries

Let $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R} \to \mathbf{R}$ be the Lagrangian of problem (7.164):

$$L(x, \mu) = f(x) - \langle \mu_G, G(x) \rangle - \langle \mu_H, H(x) \rangle + \mu_0 \langle G(x), H(x) \rangle,$$

where $\mu = (\mu_G, \mu_H, \mu_0) \in \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$. As for any other mathematical programming problem, stationary points of (7.164) and the associated Lagrange multipliers are characterized by the KKT optimality system

$$\begin{gathered} \frac{\partial L}{\partial x}(x, \mu) = 0, \\ \mu_G \geq 0, \ G(x) \geq 0, \ \langle \mu_G, G(x) \rangle = 0, \\ \mu_H \geq 0, \ H(x) \geq 0, \ \langle \mu_H, H(x) \rangle = 0, \\ \mu_0 \geq 0, \ \langle G(x), H(x) \rangle \leq 0. \end{gathered} \tag{7.165}$$

Note that we omit in (7.165) the condition $\mu_0 \langle G(x), H(x) \rangle = 0$, because it is redundant: it is automatic from $\langle G(x), H(x) \rangle = 0$, which is implied by the conditions in (7.165) that state that x is feasible. For $\bar{x} \in \mathbf{R}^n$, let $\mathcal{M}(\bar{x})$

stand for the set of Lagrange multipliers associated with $\bar{x}$, that is, the set of $\mu = (\mu_G,\, \mu_H,\, \mu_0) \in \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$ satisfying (7.165) for $x = \bar{x}$.

As is well known and can be easily checked, the MPCC constraints necessarily violate the MFCQ at every feasible point and, even more so, the stronger constraint qualifications such as the LICQ (recall Example 7.30). Indeed, for any $\bar{x} \in \mathbf{R}^n$ feasible in (7.164), define the index sets

$$\begin{array}{c} I_G = I_G(\bar{x}) = \{i = 1,\, \ldots,\, m \mid G_i(\bar{x}) = 0\}, \\ I_H = I_H(\bar{x}) = \{i = 1,\, \ldots,\, m \mid H_i(\bar{x}) = 0\}, \\ I_0 = I_0(\bar{x}) = I_G \cap I_H. \end{array} \tag{7.166}$$

Observe that $I_G \cup I_H = \{1,\, \ldots,\, m\}$, and any of the tuples $(I_G \setminus I_H,\, I_H)$, $(I_G,\, I_H \setminus I_G)$ and $(I_G \setminus I_H,\, I_H \setminus I_G,\, I_0)$ is a partition of $\{1,\, \ldots,\, m\}$. The MFCQ for (7.164) at $\bar{x}$ consists of the existence of $\bar{\xi} \in \mathbf{R}^n$ satisfying

$$G'_{I_G}(\bar{x})\bar{\xi} > 0, \quad H'_{I_H}(\bar{x})\bar{\xi} > 0, \quad \sum_{i=1}^{m}(H_i(\bar{x})\langle G'_i(\bar{x}), \bar{\xi}\rangle + G_i(\bar{x})\langle H'_i(\bar{x}), \bar{\xi}\rangle) < 0.$$

But if such $\bar{\xi} \in \mathbf{R}^n$ were to exist, then

$$\begin{aligned} 0 &> \sum_{i=1}^{m} H_i(\bar{x})\langle G'_i(\bar{x}), \bar{\xi}\rangle + \sum_{i=1}^{m} G_i(\bar{x})\langle H'_i(\bar{x}), \bar{\xi}\rangle \\ &= \sum_{i \in I_G} H_i(\bar{x})\langle G'_i(\bar{x}), \bar{\xi}\rangle + \sum_{i \in I_H} G_i(\bar{x})\langle H'_i(\bar{x}), \bar{\xi}\rangle \geq 0, \end{aligned}$$

which is a contradiction.

The inevitable lack of the MFCQ at solutions and even feasible points, leads to obvious difficulties for the theoretical analysis of MPCCs, as well as for constructing computational algorithms with guaranteed convergence. Nevertheless, as will be discussed below, $\bar{x}$ being a local solution of (7.164) still implies that it is a stationary point of this problem under reasonable assumptions allowing to cover many cases of interest (but since Theorem 1.14 is never applicable because the MFCQ is violated, special analysis is required to justify this assertion). That said, it is important to note that the lack of the MFCQ implies that if the set of Lagrange multipliers $\mathcal{M}(\bar{x})$ is nonempty it is necessarily unbounded.

In Example 7.30, if $\sigma \geq 0$, any solution $\bar{x}$ is a stationary point of the problem, and the Lagrange multiplier set is

$$\mathcal{M}(\bar{x}) = \left\{ \mu = (\mu_G,\, \mu_H,\, \mu_0) \in \mathbf{R}^3 \;\middle|\; \begin{array}{c} \mu_G = \sigma_1 + \mu_0\bar{x}_2,\; \mu_H = \sigma_2 + \mu_0\bar{x}_1, \\ \mu_0 \geq 0 \end{array} \right\},$$

which is a ray. This type of structure of the multiplier set is rather typical for MPCCs, as will be seen below.

Define further the so-called *MPCC-Lagrangian* $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^m \to \mathbf{R}$ of problem (7.164):

$$\mathcal{L}(x, \lambda) = f(x) - \langle \lambda_G, G(x) \rangle - \langle \lambda_H, H(x) \rangle,$$

where $\lambda = (\lambda_G, \lambda_H) \in \mathbf{R}^m \times \mathbf{R}^m$. A feasible point $\bar{x}$ of (7.164) is said to be a *strongly stationary point* of this problem if there exists an *MPCC-multiplier* $\bar{\lambda} = (\bar{\lambda}_G, \bar{\lambda}_H) \in \mathbf{R}^m \times \mathbf{R}^m$ satisfying

$$\frac{\partial \mathcal{L}}{\partial x}(\bar{x}, \bar{\lambda}) = 0, \ (\bar{\lambda}_G)_{I_H \setminus I_G} = 0, \ (\bar{\lambda}_H)_{I_G \setminus I_H} = 0, \tag{7.167}$$

$$(\bar{\lambda}_G)_{I_0} \geq 0, \ (\bar{\lambda}_H)_{I_0} \geq 0. \tag{7.168}$$

Without (7.168), $\bar{x}$ is called a *weakly stationary point* of (7.164).

Strong stationarity of $\bar{x}$ means that this point is stationary in the usual KKT sense for the so-called *relaxed nonlinear programming problem* (RNLP)

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & G_{I_G \setminus I_H}(x) = 0, \ H_{I_H \setminus I_G}(x) = 0, \\ & G_{I_0}(x) \geq 0, \ H_{I_0}(x) \geq 0, \end{array} \tag{7.169}$$

while weak stationarity means that this point is stationary in the *tightened nonlinear programming problem* (TNLP)

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & G_{I_G}(x) = 0, \ H_{I_H}(x) = 0. \end{array} \tag{7.170}$$

Observe that the point $\bar{x}$ in question is always feasible in the related TNLP (7.170), and hence, in the RNLP (7.169).

We say that *MPCC-linear independence constraint qualification* (MPCC-LICQ) holds at a feasible point $\bar{x}$ of (7.164) if

$$\operatorname{rank} \begin{pmatrix} G'_{I_G}(\bar{x}) \\ H'_{I_H}(\bar{x}) \end{pmatrix} = |I_G| + |I_H|.$$

In other words, MPCC-LICQ holds if the usual regularity condition holds at $\bar{x}$ for the TNLP (7.170), or equivalently, if the usual LICQ holds at $\bar{x}$ for the RNLP (7.169). Furthermore, near $\bar{x}$, the feasible set of the MPCC (7.164) evidently contains the feasible set of the TNLP (7.170) and is contained in the feasible set of the RNLP (7.169). In particular, local optimality of $\bar{x}$ in the MPCC (7.164) implies its local optimality in the TNLP (7.170). If, in addition, the MPCC-LICQ holds at $\bar{x}$, then, according to Theorem 1.11, $\bar{x}$ is a stationary point of the TNLP (7.170), which is equivalent to saying that $\bar{x}$ is a weakly stationary point of the MPCC (7.164).

A much stronger result than the previous assertion is established in [240]. It states that a local solution of the MPCC (7.164) satisfying the

MPCC-LICQ must actually be strongly stationary. Apparently the easiest way to see this is by means of the so-called piecewise analysis that follows.

For each partition $(I_1,\, I_2)$ of I_0, define the *branch* (or *piece*) *problem* at $\bar{x}$ as

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & G_{(I_G\setminus I_H)\cup I_1}(x) = 0,\ H_{(I_H\setminus I_G)\cup I_2}(x) = 0, \\ & G_{I_2}(x) \geq 0,\ H_{I_1}(x) \geq 0. \end{array} \tag{7.171}$$

There is a finite number of such branch problems, $\bar{x}$ is feasible in each of them, and in a neighborhood of $\bar{x}$ the feasible set of the MPCC (7.164) is a union of feasible sets of all the branch problems. Observe now that the MPCC-LICQ at $\bar{x}$ is equivalent to the LICQ holding for each of the branch problems. Therefore, if $\bar{x}$ is a local solution of the MPCC (7.164), then for every partition $(I_1,\, I_2)$ of I_0, this point is a local solution of the corresponding branch problem (7.171). Then according to Theorem 1.14, if in addition the MPCC-LICQ holds at $\bar{x}$, then $\bar{x}$ is a stationary point of the branch problem (7.171). The latter means that for every partition $(I_1,\, I_2)$ of I_0 there exists $\bar{\lambda} = (\bar{\lambda}_G,\, \bar{\lambda}_H) \in \mathbf{R}^m \times \mathbf{R}^m$ satisfying (7.167) and

$$(\bar{\lambda}_G)_{I_2} \geq 0,\ (\bar{\lambda}_H)_{I_1} \geq 0. \tag{7.172}$$

(The subvector $((\bar{\lambda}_G)_{I_G},\, (\bar{\lambda}_H)_{I_H})$ of $\bar{\lambda}$ is the Lagrange multiplier for the branch problem (7.171).) Moreover, since the MPCC-LICQ is the usual regularity condition for the TNLP (7.170) at $\bar{x}$, it implies that there cannot be two distinct multipliers $\bar{\lambda}$ satisfying (7.167). Therefore, $\bar{\lambda}$ defined above for the branch problems is actually the same for all the partitions $(I_1,\, I_2)$ of I_0, and the corresponding $((\bar{\lambda}_G)_{I_G},\, (\bar{\lambda}_H)_{I_H})$ all coincide with the unique multiplier for the TNLP (7.170). Finally, from (7.172) being valid for any partition $(I_1,\, I_2)$ of I_0, it follows that (7.168) holds as well, implying the strong stationarity of $\bar{x}$ with the unique MPCC-multiplier $\bar{\lambda}$. We have therefore established the following.

Theorem 7.31. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be differentiable at a point $\bar{x} \in \mathbf{R}^n$, and let $G : \mathbf{R}^n \to \mathbf{R}^m$ and $H : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable in a neighborhood of $\bar{x}$, with their derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be feasible in the MPCC* (7.164)*, and assume that the MPCC-LICQ holds at $\bar{x}$.*

If $\bar{x}$ is a local solution of the MPCC (7.164)*, then it is a strongly stationary point of this problem, and the associated MPCC-multiplier $\bar{\lambda} = (\bar{\lambda}_G,\, \bar{\lambda}_H)$ is uniquely defined. Moreover, the weak stationarity condition* (7.167) *holds for this $\bar{\lambda}$ only.*

It was shown in [241] that the MPCC-LICQ is a generic property. This observation combined with Theorem 7.31 allows to state that strong stationarity is a reasonable stationarity concept for MPCCs.

In Example 7.30, every feasible point satisfies the MPCC-LICQ, and if $\sigma \geq 0$, then every solution is strongly stationary with the unique associated MPCC-multiplier $\bar{\lambda} = (\bar{\lambda}_G,\, \bar{\lambda}_H) = \sigma$.

It turns out that strong stationarity is actually equivalent to the usual (KKT) stationarity. The following proposition summarizes some results in [95] and [127]. Its proof is by direct verification.

Proposition 7.32. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $G : \mathbf{R}^n \to \mathbf{R}^m$ and $H : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable at $\bar{x} \in \mathbf{R}^n$. Let $\bar{x}$ be feasible in the MPCC (7.164)*

Then $\bar{x}$ is a stationary point of problem (7.164) if and only if it is a strongly stationary point of this problem.

In particular, if $\bar{\mu} = (\bar{\mu}_G, \bar{\mu}_H, \bar{\mu}_0)$ is a Lagrange multiplier associated with $\bar{x}$, then $\bar{\lambda} = (\bar{\lambda}_G, \bar{\lambda}_H)$ defined by

$$(\bar{\lambda}_G)_i = (\bar{\mu}_G)_i - \bar{\mu}_0 H_i(\bar{x}),\ i \in I_G \setminus I_H, \quad (\bar{\lambda}_G)_i = (\bar{\mu}_G)_i,\ i \in I_H, \tag{7.173}$$

$$(\bar{\lambda}_H)_i = (\bar{\mu}_H)_i - \bar{\mu}_0 G_i(\bar{x}),\ i \in I_H \setminus I_G, \quad (\bar{\lambda}_H)_i = (\bar{\mu}_H)_i,\ i \in I_G, \tag{7.174}$$

is an MPCC-multiplier associated with $\bar{x}$.

Conversely, if $\bar{\lambda} = (\bar{\lambda}_G, \bar{\lambda}_H)$ is an MPCC-multiplier associated with $\bar{x}$, then any $\bar{\mu} = (\bar{\mu}_G, \bar{\mu}_H, \bar{\mu}_0)$ satisfying (7.173), (7.174) and

$$\bar{\mu}_0 \geq \bar{\nu}, \tag{7.175}$$

with $\bar{\nu}$ defined by

$$\bar{\nu} = \max\left\{0,\ \max_{i \in I_G \setminus I_H}\left(-\frac{(\bar{\lambda}_G)_i}{H_i(\bar{x})}\right),\ \max_{i \in I_H \setminus I_G}\left(-\frac{(\bar{\lambda}_H)_i}{G_i(\bar{x})}\right)\right\}, \tag{7.176}$$

is a Lagrange multiplier associated with $\bar{x}$.

Furthermore, if f, G, and H are twice differentiable at the point $\bar{x}$, then for every element $\xi \in \mathbf{R}^n$ and every $\bar{\mu} = (\bar{\mu}_G, \bar{\mu}_H, \bar{\mu}_0) \in \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$ and $\bar{\lambda} = (\bar{\lambda}_G, \bar{\lambda}_H) \in \mathbf{R}^m \times \mathbf{R}^m$ satisfying (7.173), (7.174), it holds that

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\mu})\xi, \xi \right\rangle = \left\langle \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\mu})\xi, \xi \right\rangle + 2\bar{\lambda}_0 \sum_{i=1}^{m} \langle G_i'(\bar{x}), \xi\rangle \langle H_i'(\bar{x}), \xi\rangle. \tag{7.177}$$

In particular, if $\bar{\lambda}$ is the unique MPCC-multiplier associated with $\bar{x}$ (e.g., under the MPCC-LICQ), then $\mathcal{M}(\bar{x})$ is the ray defined by (7.173)–(7.175), with its origin corresponding to $\bar{\mu}_0 = \bar{\nu}$.

We proceed with second-order optimality conditions for the MPCC (7.164). It can be easily checked that the standard critical cone of problem (7.164) at a feasible point $\bar{x}$ is given by

$$C(\bar{x}) = \left\{ \xi \in \mathbf{R}^n \,\middle|\, \begin{array}{l} G'_{I_G \setminus I_H}(\bar{x})\xi = 0,\ H'_{I_H \setminus I_G}(\bar{x})\xi = 0, \\ G'_{I_0}(\bar{x})\xi \geq 0,\ H'_{I_0}(\bar{x})\xi \geq 0,\ \langle f'(\bar{x}), \xi\rangle \leq 0 \end{array} \right\}. \tag{7.178}$$

Evidently, this cone coincides with the critical cone of the RNLP (7.169) at the point $\bar{x}$.

We say that the *MPCC-second-order sufficient condition* (MPCC-SOSC) holds at a strongly stationary point $\bar{x}$ of problem (7.164) with the associated MPCC-multiplier $\bar{\lambda}$, if

$$\left\langle \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x},\, \bar{\lambda})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}. \tag{7.179}$$

This condition is equivalent to the usual SOSC for the RNLP (7.169) at $\bar{x}$ for the associated multiplier $((\bar{\lambda}_G)_{I_G},\, (\bar{\lambda}_H)_{I_H})$.

Note that for every $\xi \in C(\bar{x})$, we obtain from (7.178) that (7.177) takes the form

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\mu})\xi,\, \xi \right\rangle = \left\langle \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x},\, \bar{\lambda})\xi,\, \xi \right\rangle + 2\bar{\mu}_0 \sum_{i \in I_0} \langle G_i'(\bar{x}),\, \xi\rangle \langle H_i'(\bar{x}),\, \xi\rangle, \tag{7.180}$$

where the last term in the right-hand side is nonnegative. Thus, according to Proposition 7.32, the MPCC-SOSC implies the usual SOSC

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\mu})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{7.181}$$

for any $\bar{\mu}$ satisfying (7.173)–(7.175). The latter, in turn, implies the local optimality of $\bar{x}$ in the MPCC (7.164); see Theorem 1.20. In particular, under the MPCC-LICQ, the MPCC-SOSC (7.179) (with the unique MPCC-multiplier $\bar{\lambda}$) implies that the SOSC (7.181) holds for any $\bar{\mu}$ in the ray $\mathcal{M}(\bar{x})$, including the origin of this ray.

It is important to point out that the MPCC-SOSC (7.179) is a rather strong condition. In particular, it cannot be linked to any SONC for the MPCC (7.164). By this we mean that a solution of (7.164) that satisfies the MPCC-LICQ (and thus is strongly stationary) need not satisfy the condition obtained from (7.179) by replacing the strict inequality by nonstrict. To that end, we next introduce the weaker and much more natural SOSC for MPCCs, which is related to an appropriate SONC for the MPCC (7.164), as explained below.

It can be directly verified that the union of the critical cones of all the branch problems (7.171) at $\bar{x}$ is given by

$$C_2(\bar{x}) = \left\{ \xi \in \mathbf{R}^n \;\middle|\; \begin{array}{l} G'_{I_G \setminus I_H}(\bar{x})\xi = 0,\ H'_{I_H \setminus I_G}(\bar{x})\xi = 0, \\ G'_{I_0}(\bar{x})\xi \geq 0,\ H'_{I_0}(\bar{x})\xi \geq 0,\ \langle f'(\bar{x}),\, \xi\rangle \leq 0, \\ \langle G_i'(\bar{x}),\, \xi\rangle \langle H_i'(\bar{x}),\, \xi\rangle = 0,\ i \in I_0 \end{array} \right\}, \tag{7.182}$$

where the subscript “2” indicates that, unlike $C(\bar{x})$, this set takes into account the second-order information about the last constraint in (7.164). By direct comparison of (7.178) and (7.182), we see that

$$C_2(\bar{x}) \subset C(\bar{x}). \tag{7.183}$$

We say that the *piecewise SOSC* holds at a strongly stationary point $\bar{x}$ of the MPCC (7.164) with an associated MPCC-multiplier $\bar{\lambda}$, if

$$\left\langle \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\lambda})\xi, \xi \right\rangle > 0 \quad \forall \xi \in C_2(\bar{x}) \setminus \{0\}. \tag{7.184}$$

From (7.167), (7.168) it evidently follows that if $\bar{\lambda} = (\bar{\lambda}_G, \bar{\lambda}_H)$ is an MPCC-multiplier associated with $\bar{x}$, then the pair $((\bar{\mu}_G)_{I_G}, (\bar{\mu}_H)_{I_H})$ is a Lagrange multiplier associated with $\bar{x}$ for the branch problem (7.171) (and hence, also for the TNLP (7.170)). Therefore, the piecewise SOSC (7.184) implies the standard SOSC for each branch problem at $\bar{x}$ (and hence, also for the TNLP (7.170)). By Theorem 1.20 this, in turn, guarantees that $\bar{x}$ is a strict local solution for each branch problem, and hence, for the MPCC (7.164). Thus, the piecewise SOSC is indeed sufficient for optimality, even though it is evidently weaker than the MPCC-SOSC (see (7.183)). In addition, assuming the MPCC-LICQ, we obtain that the LICQ holds for each branch problem, and we derive from Theorem 1.19 that the condition obtained from (7.184) by replacing the strict inequality by nonstrict is necessary for local optimality (the fact first observed in [240]).

The next example is taken from [95].

Example 7.33. Let $n = 2$, $m = 1$, $f(x) = x_1^2 + x_2^2 - 4x_1x_2$, and let the constraints mappings be the same as in Example 7.30: $G(x) = x_1$, $H(x) = x_2$.

The point $\bar{x} = 0$ is the unique solution of the MPCC (7.164) with this data, satisfying the MPCC-LICQ. This point is strongly stationary with the unique associated MPCC-multiplier $\bar{\lambda} = (\bar{\lambda}_G, \bar{\lambda}_H) = (0, 0)$. From (7.178) we have that $C(\bar{x}) = \mathbf{R}_+^2$, and the quadratic form associated with $f''(0)$ is not nonnegative on $C(\bar{x})$. Therefore, the condition (7.179) does not hold, even with the strict inequality replaced by nonstrict.

At the same time, from (7.182) we have that $C_2(\bar{x})$ coincides with the feasible set, and the condition (7.184) is satisfied.

Suppose now that the MPCC-LICQ and the piecewise SOSC (7.184) (with the unique MPCC-multiplier $\bar{\lambda}$) hold at a strongly stationary point $\bar{x}$ of the MPCC (7.164). From (7.180) and (7.182) it can be easily derived that in this case either the SOSC (7.181) holds for all $\bar{\mu}$ in the ray $\mathcal{M}(\bar{x})$, or possibly there exists $\hat{\nu} \geq \bar{\nu}$ such that the SOSC (7.181) does not hold for all $\bar{\mu}$ corresponding to $\bar{\mu}_0 \in [\bar{\nu}, \hat{\nu}]$, and holds for all $\bar{\mu}$ corresponding to $\bar{\mu}_0 > \hat{\nu}$. Conversely, if the SOSC (7.181) holds for some $\bar{\mu} \in \mathcal{M}(\bar{x})$, from (7.180) and (7.182), taking also into account (7.183), it is easy to see that the piecewise SOSC (7.184) holds as well. Thus, under the MPCC-LICQ, the SOSC (with some Lagrange multiplier) is equivalent to the piecewise SOSC.

We complete this section by the following observations. If $\bar{x}$ is a strongly stationary point of the MPCC (7.164), then similarly to Lemma 1.17 it can be easily seen that for any associated MPCC-multiplier $\bar{\lambda} = (\bar{\lambda}_G, \bar{\lambda}_H)$ it holds that

$$C(\bar{x}) = \left\{ \xi \in \mathbf{R}^n \;\middle|\; \begin{array}{l} G'_{I_G \setminus I_H}(\bar{x})\xi = 0,\ H'_{I_H \setminus I_G}(\bar{x})\xi = 0, \\ G'_{I_0}(\bar{x})\xi \geq 0,\ H'_{I_0}(\bar{x})\xi \geq 0, \\ (\bar{\lambda}_G)_i \langle G'_i(\bar{x}),\, \xi\rangle = 0,\ (\bar{\lambda}_H)_i \langle H'_i(\bar{x}),\, \xi\rangle = 0,\ i \in I_0 \end{array} \right\}$$

and

$$C_2(\bar{x}) = \left\{ \xi \in \mathbf{R}^n \;\middle|\; \begin{array}{l} G'_{I_G \setminus I_H}(\bar{x})\xi = 0,\ H'_{I_H \setminus I_G}(\bar{x})\xi = 0, \\ G'_{I_0}(\bar{x})\xi \geq 0,\ H'_{I_0}(\bar{x})\xi \geq 0, \\ (\bar{\lambda}_G)_i \langle G'_i(\bar{x}),\, \xi\rangle = 0,\ (\bar{\lambda}_H)_i \langle H'_i(\bar{x}),\, \xi\rangle = 0, \\ \langle G'_i(\bar{x}),\, \xi\rangle \langle H'_i(\bar{x}),\, \xi\rangle = 0,\ i \in I_0 \end{array} \right\}.$$

It then becomes evident that under the *upper-level strict complementarity condition*

$$(\bar{\lambda}_G)_{I_0} > 0, \quad (\bar{\lambda}_H)_{I_0} > 0, \tag{7.185}$$

it holds that $C(\bar{x}) = C_2(\bar{x}) = K(\bar{x})$, where

$$K(\bar{x}) = \{\xi \in \mathbf{R}^n \mid G'_{I_G}(\bar{x})\xi = 0,\ H'_{I_H}(\bar{x})\xi = 0\}. \tag{7.186}$$

In the case of upper-level strict complementarity (7.185), the MPCC-SOSC (7.179) and the piecewise SOSC (7.184) become equivalent, and they can be stated as

$$\left\langle \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x},\, \bar{\lambda})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in K(\bar{x}) \setminus \{0\}. \tag{7.187}$$

7.3.2 General-Purpose Newton-Type Methods Applied to Mathematical Programs with Complementarity Constraints

Despite the fact that standard constraint qualifications do not hold, there is computational evidence of good performance of some general-purpose optimization codes on MPCCs, i.e., of methods developed for general optimization problems, which do not use any of the specific MPCC structure. One (non-Newtonian) example is the augmented Lagrangian algorithm, also called the method of multipliers. Its global convergence properties when applied to MPCCs, and rather comprehensive computational comparisons with the alternatives on the MacMPEC test problems [185], are given in [157]. Local convergence and rate of convergence analysis without constraint qualifications, thus relevant for degenerate problems and MPCCs, is given in [76]. As augmented Lagrangian methods are not of the Newtonian family, we shall not go into any further details. Instead, we shall discuss another general-purpose algorithm which seems to work quite well on MPCCs, and which is of the Newton type—SQP method. Apart from the computational evidence [8, 11, 92], we note that [95] gives some theoretical justification for local superlinear convergence of the SQP algorithm for MPCCs, at least in some situations. Specifically, in [95] the complementarity constraints are reformulated using slack variables $y_G = G(x)$ and $y_H = H(x)$ (so that $\langle y_G, y_H\rangle = 0$), and local superlinear convergence is fully justified when the complementarity condition for slack variables holds exactly from some iteration on. The set of assumptions also includes the MPCC-LICQ and the MPCC-SOSC, and some requirements on the QP-solver used to solve the subproblems (more on this later), among other things.

More precisely, in [95] it is shown that if the primal sequence hits a point satisfying exact complementarity, that is, $y_G^k \geq 0$, $y_H^k \geq 0$ and $\langle y_G^k,\, y_H^k\rangle = 0$ for some iteration index k, and if this happens close enough to the solution in question, then all the subsequent primal iterates also satisfy exact complementarity. Moreover, the iteration sequence is then the same as that of the SQP method applied to the slack reformulation of the RNLP (7.169). Once this is established, the local superlinear convergence can be derived from (4.14) applied to the RNLP, since the MPCC-LICQ and the MPCC-SOSC imply the LICQ and the SOSC for the RNLP.

However, it is very easy to provide examples satisfying all *natural* in the MPCC context requirements (say, the MPCC-LICQ and the piecewise SOSC), and such that the SQP method does not possess local superlinear convergence. To begin with, due to the lack of the MFCQ, the linearized MPCC constraints can be inconsistent arbitrarily close to a solution, especially in the presence of additional usual constraints (see [95, Sects. 2.2, 6.3] for a discussion of these issues). Moreover, even if SQP subproblems are feasible, superlinear convergence rate may not hold.

Consider, e.g., the problem from Example 7.33. Then for the current primal-dual iterate $(x^k,\, \mu^k) \in \mathbf{R}^n \times (\mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R})$, the SQP subproblem defining the next iterate has the form

$$\begin{array}{ll} \text{minimize} & 2x_1^k(x_1 - x_1^k) + 2x_2^k(x_2 - x_2^k) - 4x_2^k(x_1 - x_1^k) - 4x_1^k(x_2 - x_2^k) \\ & +(x_1 - x_1^k)^2 + (x_2 - x_2^k)^2 - (4 - \mu_0^k)(x_1 - x_1^k)(x_2 - x_2^k) \\ \text{subject to} & x_1 \geq 0,\ x_2 \geq 0,\ x_1^k x_2^k + x_2^k(x_1 - x_1^k) + x_1^k(x_2 - x_2^k) \leq 0. \end{array} \tag{7.188}$$

Suppose that $0 \leq \mu_0^k < 6$, and let, e.g., $x_1^k = x_2^k \neq 0$. It can be easily seen that in this case, the first two constraints in (7.188) cannot be active, and that $x^{k+1} = x^k/2$ is the unique stationary point of this subproblem, with the last constraint being active. The multiplier associated with x^{k+1} is given by

$$\mu^{k+1} = (\mu_G^{k+1},\, \mu_H^{k+1},\, \mu_0^{k+1}) = (0,\, 0,\, 1 + \mu_0^k/2).$$

It follows that $\{x^k\}$ converges linearly to $\bar{x} = 0$, while $\{\mu^k\}$ converges linearly to $\bar{\mu} = (0,\, 0,\, 2)$.

Observe that in this example, for all k, the set of indices of active inequality constraints of the subproblem (7.188) is $A = \{3\}$, that is, only the last constraint in (7.188) is active at x^{k+1}, while the inequality constraints active at the solution are $A(\bar{x}) = \{1, 2, 3\}$. In particular, $\bar{\mu} = (0, 0, 2)$ violates the strict complementarity condition, and it can be verified directly that this $\bar{\mu}$ is a critical Lagrange multiplier with respect to this index set A in the sense of Definition 7.8. Therefore, the observed behavior perfectly agrees with the discussion in Sect. 7.1.2: this is a typical example of attraction of the dual sequence to a critical multiplier.

It is interesting to note that in this example there exists another multiplier critical with respect to the same set A, namely, $\bar{\mu} = (0, 0, 6)$. This multiplier satisfies the SOSC (7.179) (and hence, it is noncritical in the sense of Definition 1.41), but it does not satisfy the SSOSC. However, this multiplier does not attract dual iterates.

Regarding the dual behavior of Newton-type methods when applied to MPCCs, assuming the MPCC-LICQ at the primal limit $\bar{x}$, the following two possibilities should be considered:

- If the upper-level strict complementarity condition (7.185) holds for the unique MPCC-multiplier $\bar{\lambda}$ associated with $\bar{x}$, then there is exactly one Lagrange multiplier violating strict complementarity, namely the origin $\bar{\mu}$ of the ray $\mathcal{M}(\bar{x})$, corresponding to $\bar{\mu}_0 = \bar{\nu}$, and sometimes referred to as the *basic Lagrange multiplier*.
- If the upper-level strict complementarity condition does not hold, then there are no multipliers satisfying strict complementary.

It is tempting to extend the discussion in Sect. 7.1.2 of general degenerate optimization problems to MPCCs. For example, suppose that the upper-level strict complementarity condition holds. Since under the MPCC-LICQ the MPCC-SOSC (7.179) implies that the SOSC (7.181) holds for any Lagrange multiplier, it is natural to conjecture that under the MPCC-SOSC, if the dual trajectory converges, then most likely it converges to the basic multiplier violating strict complementarity. Similarly, under the piecewise SOSC (7.184), there may exist $\hat{\nu} \geq \bar{\nu}$ such that the SOSC (7.181) holds only with those multipliers $\bar{\mu}$ that correspond to $\bar{\mu}_0 > \hat{\nu}$. And then it is natural to conjecture that under the piecewise SOSC, if the dual trajectory converges, then most likely it converges either to the critical multiplier $\bar{\mu}$ corresponding to $\bar{\mu}_0 = \hat{\nu}$ or to the basic multiplier violating strict complementarity. The cases of violation of the upper-level strict complementarity condition can be also analyzed as above.

However, the discussion for general optimization problems should be applied to MPCCs with some care, because of the very special structure of the latter. Specifically, as discussed above, a quite possible (and, in fact, favorable) scenario for the SQP method applied to the MPCC reformulation with slack variables is when the primal sequence hits a point satisfying exact complementarity. In this case, the constraints linearized about this point violate standard constraint qualifications and, hence, the Lagrange multipliers of the SQP subproblem would normally be not unique. Moreover, exact complementarity holds for all the subsequent iterations. Therefore, the specificity of the QP-solver employed to solve the subproblems must be taken into account (otherwise, as already discussed in Sect. 7.1.2, if QP-solver may pick any multiplier, then dual behavior should be considered essentially arbitrary). For example, in [95], this specificity consists of saying that "a QP-solver always chooses a linearly independent basis," and this (together with other things, including the upper-level strict complementarity) results in dual convergence to the basic multiplier. Note, however, that if one assumes the piecewise SOSC (7.184) instead of the stronger MPCC-SOSC (7.179), then dual convergence to the basic multiplier is still an undesirable event, since this multiplier may not satisfy the SOSC (7.181) (even though it is noncritical, in general).

For the case when complementarity is not maintained along the iterations, i.e., if $\langle y_G^k,\, y_H^k\rangle > 0$ for all k (as for the problem from Example 7.33), the discussion for general optimization problems still perfectly applies to MPCCs, and the corresponding conclusions in Sect. 7.1.2 remain valid.

Overall, the existing evidence supporting the use of standard Newtonian algorithms (say, the SQP method) on MPCCs cannot be regarded as completely satisfactory, especially as a matter of theoretical guarantees. Therefore, it still makes sense to consider special methods for degenerate problems discussed in Sect. 7.2, or to develop algorithms which take into account the special structure of MPCCs, and guarantee superlinear convergence under more natural assumptions.

Among the approaches discussed in Sect. 7.2, the stabilized SQP and the techniques from [141, 272] are good candidates for a local method in the MPCC setting. Recall that these methods all guarantee superlinear convergence when started close enough to a primal-dual solution $(\bar{x},\, \bar{\mu})$ satisfying the SOSC (7.181). Assuming the MPCC-LICQ at $\bar{x}$, the SOSC (7.181) with some multiplier (with all multipliers) is implied by the piecewise SOSC (7.184) (by the MPCC-SOSC (7.179)). Therefore, local superlinear convergence of these methods for MPCCs is fully justified under the MPCC-LICQ and the piecewise SOSC (or the MPCC-SOSC allowing dual starting points close enough to an arbitrary $\bar{\mu} \in \mathcal{M}(\bar{x})$).

Finally, one particular tool for tackling degenerate problems which should be recalled in connection with MPCCs is the elastic mode modification of the SQP method, designed to handle potentially infeasible subproblems. We refer to [8, 11, 12] for theoretical and numerical validation of various forms of the elastic mode for MPCCs.

Some other approaches to MPCCs, which make explicit use of its structure, are discussed in the next two sections.

7.3.3 Piecewise Sequential Quadratic Programming and Active-Set Methods

The main idea of the piecewise SQP algorithm for MPCCs, proposed originally in [229] for problems with linear complementarity constraints and then extended in [184] to the nonlinear case, is the following: at each iteration, identify *any* branch problem valid at the solution $\bar{x}$ that is being approximated, and perform a step of the usual SQP method for this branch.

To identify a valid branch problem (7.171), it suffices to (over)estimate the sets $I_G \setminus I_H$ and $I_H \setminus I_G$. Locally, this comes for free, with no significant computational cost, and no regularity assumptions needed. For example, assuming only the continuity of G and H at $\bar{x}$, for any $x^k \in \mathbf{R}^n$ close enough to $\bar{x}$ the index sets

$$\begin{aligned} J_G^k &= J_G(x^k) = \{i = 1, \ldots, m \mid G_i(x^k) < H_i(x^k)\}, \\ J_H^k &= J_H(x^k) = \{i = 1, \ldots, m \mid G_i(x^k) \geq H_i(x^k)\} \end{aligned} \tag{7.189}$$

evidently satisfy the inclusions

$$(I_G \setminus I_H) \subset J_G^k, \quad (I_H \setminus I_G) \subset J_H^k. \tag{7.190}$$

Consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & G_{J_G^k}(x) = 0,\ H_{J_H^k}(x) = 0,\ G_{J_H^k}(x) \geq 0,\ H_{J_G^k}(x) \geq 0. \end{array} \tag{7.191}$$

Since $J_G^k \cup J_H^k = \{1, \ldots, m\}$ and $J_G^k \cap J_H^k = \emptyset$, the index sets $I_1 = J_G^k \cap I_0$ and $I_2 = J_H^k \cap I_0$ form a partition of I_0. Furthermore, from (7.190) it follows that

$$(I_G \setminus I_H) \cup I_1 \subset J_G^k, \quad (I_H \setminus I_G) \cup I_2 \subset J_H^k. \tag{7.192}$$

Conversely, suppose that there exists $i \in J_G^k$ such that $i \notin (I_G \setminus I_H) \cup I_1$, i.e., $i \notin (I_G \setminus I_H)$ and $i \notin I_1$, implying that $i \in I_H \setminus I_0 = I_H \setminus I_G$. Then the second inclusion in (7.190) implies that $i \in J_H^k$, which is in a contradiction with the inclusion $i \in J_G^k$. This shows that the first inclusion in (7.192) holds as equality, and by similar argument it follows that the second inclusion in (7.192) holds as equality as well. Thus, we have that

$$(I_G \setminus I_H) \cup I_1 = J_G^k, \quad (I_H \setminus I_G) \cup I_2 = J_H^k. \tag{7.193}$$

From (7.193) it follows that problem (7.191) differs from the branch problem (7.171) only by the extra inequality constraints

$$G_{J_H^k \setminus I_2}(x) \geq 0,\ H_{J_G^k \setminus I_1}(x) \geq 0. \tag{7.194}$$

Similarly to the argument used above to establish (7.193), it can be shown that

$$(J_H^k \setminus I_2) \cap I_G = \emptyset, \quad (J_G^k \setminus I_1) \cap I_H = \emptyset.$$

Therefore, all the constraints in (7.194) are inactive at $\bar{x}$, and hence, locally (near $\bar{x}$) problems (7.171) and (7.191) are equivalent.

Based on these considerations, the *piecewise SQP* algorithm for problem (7.164) is the following procedure.

Algorithm 7.34 Choose $x^0 \in \mathbf{R}^n$ and $\lambda^0 = (\lambda_G^0,\, \lambda_H^0) \in \mathbf{R}^m \times \mathbf{R}^m$, and set $k = 0$.

1. Define the index sets J_G^k and J_H^k according to (7.189). If $(x^k,\, \lambda^k)$ satisfies the system

$$\begin{gathered}\frac{\partial \mathcal{L}}{\partial x}(x,\, \lambda) = 0,\ G_{J_G^k}(x) = 0,\ H_{J_H^k}(x) = 0,\\ (\lambda_G)_{J_H^k} \geq 0,\ G_{J_H^k}(x) \geq 0,\ \langle (\lambda_G)_{J_H^k},\, G_{J_H^k}(x)\rangle = 0,\\ (\lambda_H)_{J_G^k} \geq 0,\ H_{J_G^k}(x) \geq 0,\ \langle (\lambda_G)_{J_G^k},\, G_{J_G^k}(x)\rangle = 0,\end{gathered}$$

   stop.
2. Compute $x^{k+1} \in \mathbf{R}^n$ as a stationary point of problem

$$\begin{array}{ll}\text{minimize} & \langle f'(x^k),\, x - x^k\rangle + \dfrac{1}{2}\left\langle \dfrac{\partial^2 \mathcal{L}}{\partial x^2}(x^k,\, \lambda^k)(x - x^k),\, x - x^k\right\rangle\\ \text{subject to} & G_{J_G^k}(x^k) + G'_{J_G^k}(x^k)(x - x^k) = 0,\\ & H_{J_H^k}(x^k) + H'_{J_H^k}(x^k)(x - x^k) = 0,\\ & G_{J_H^k}(x^k) + G'_{J_H^k}(x^k)(x - x^k) \geq 0,\\ & H_{J_G^k}(x^k) + H'_{J_G^k}(x^k)(x - x^k) \geq 0,\end{array} \tag{7.195}$$

   and $\lambda^{k+1} = (\lambda_G^{k+1},\, \lambda_H^{k+1}) \in \mathbf{R}^m \times \mathbf{R}^m$ as an associated Lagrange multiplier.
3. Increase k by 1 and go to step 1.

To justify local superlinear convergence of the piecewise SQP method one needs to guarantee local superlinear convergence of the usual SQP iterations applied to each branch problem, and dual convergence to the same multiplier for all the branches. The latter comes from assuming the MPCC-LICQ: according to Theorem 7.31, this assumption guarantees that the unique MPCC-multiplier gives rise to the unique multiplier for each branch problem, the same for all the branches. In addition, the MPCC-LICQ implies the LICQ for each branch, and assuming the piecewise SOSC enforces the SOSC for each branch problem. According to Theorem 4.14, these are all the ingredients needed for local superlinear convergence of the SQP method applied to each branch problem. We summarize these observations in the following theorem, omitting the details of a more formal proof, which are nevertheless clear from the comments provided.

Theorem 7.35. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $G : \mathbf{R}^n \to \mathbf{R}^m$ and $H : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a local solution of the MPCC (7.164), satisfying the MPCC-LICQ and the piecewise SOSC (7.184) for the associated MPCC-multiplier $\bar{\lambda} \in \mathbf{R}^m \times \mathbf{R}^m$.*

Then there exists a constant $\delta > 0$ *such that for any starting point* $(x^0, \lambda^0) \in \mathbf{R}^n \times (\mathbf{R}^m \times \mathbf{R}^m)$ *close enough to* $(\bar{x}, \bar{\lambda})$, *there exists an iterative sequence* $\{(x^k, \lambda^k)\} \subset \mathbf{R}^n \times (\mathbf{R}^m \times \mathbf{R}^m)$ *of Algorithm 7.34, satisfying*

$$\|(x^{k+1} - x^k,\, \lambda^{k+1} - \lambda^k,\, \mu^{k+1} - \mu^k)\| \leq \delta$$

for each $k = 0, 1, \ldots$; *any such sequence converges to* $(\bar{x}, \bar{\lambda})$, *and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the second derivatives of* f, G, *and* H *are locally Lipschitz-continuous with respect to* $\bar{x}$.

Observe that in the local setting of Theorem 7.35, its assertion remains valid if the index sets J_G^k and J_H^k are defined in Algorithm 7.34 not for each k but only for $k = 0$, and then are kept fixed on subsequent iterations. However, identification at each step is practically useful since, intuitively, this makes the algorithm "less local." That said, even with identification at every step, Algorithm 7.34 is difficult to globalize, as explained further below.

The idea of the approach in [143] (to some extent motivated by [181]) is more in the spirit of active-set methods: instead of an arbitrary valid branch problem, identify the TNLP (7.170) and perform the Newton–Lagrange iterations for this equality-constrained problem. (Observe, however, that in the presence of usual inequality constraints in the original MPCC, they will appear in the associated TNLP as well, and the Newton–Lagrange iterations will have to be replaced by the SQP iterations.) The TNLP (7.170) is not a branch problem of the form (7.171), in general, but its feasible set is evidently contained in the feasible set of every branch problem.

According to Theorem 4.3, for local superlinear convergence of the Newton–Lagrange method for the TNLP (7.170), the required assumptions are the regularity condition and the SOSC for this problem. As discussed in Sect. 7.3.1, in terms of the original MPCC (7.164), these conditions are guaranteed by the MPCC-LICQ and the piecewise SOSC (7.184), while the latter is equivalent to the existence of $\bar{\mu} \in \mathcal{M}(\bar{x})$ satisfying the SOSC (7.181).

Identifying the TNLP (7.170) means identifying the index sets I_G and I_H. In one simple case this identification is actually immediate. We say that the *lower-level strict complementarity condition* holds at a feasible point $\bar{x}$ of problem (7.164) if $I_0 = I_0(\bar{x}) = \emptyset$. In this case, by continuity considerations it follows that the feasible set of problem (7.164) near $\bar{x}$ coincides with the feasible set of the TNLP (7.170). Therefore, assuming the lower-level strict complementarity, no relevant information is lost when locally passing from the MPCC (7.164) to the TNLP (7.170); at the same time, the latter might have much better regularity properties, as discussed above.

Moreover, by the continuity of G and H it is evident that under the lower-level strict complementarity assumption, the index sets J_G^k and J_H^k defined according to (7.189) coincide with I_G and I_H, respectively, for all $x^k \in \mathbf{R}^n$ close enough to $\bar{x}$.

However, the lower-level strict complementarity at a solution of the MPCC (7.164) is the kind of assumptions that should be regarded too restrictive. For the parametric MPCC in Example 7.30, the lower-level strict complementarity condition holds at all feasible points with the only exception of $\bar{x} = 0$, but the latter is *always* a solution when any exist. (For other feasible points, the feasible set is a smooth submanifold around them, a simple case). But if $\sigma > 0$, then $\bar{x} = 0$ is the only solution, and if σ has a negative component, then there are no solutions.

To approach the task of identifying I_G and I_H without the lower-level strict complementarity assumption, observe that these sets are nothing else but the sets of indices of all the active inequality constraints in MPCC (7.164), taking out the last constraint which is a priori known to be active at any feasible point (and thus does not need to be "identified"). Therefore, local identification of I_G and I_H can be performed by means of the procedure discussed in Sects. 3.4.2, 4.1.3. This identification technique costs nothing of significance computationally. Regarding the needed assumptions, Proposition 3.58 requires a primal-dual point used for identification to be close enough to $(\bar{x}, \bar{\mu})$ such that $\bar{\mu} \in \mathcal{M}(\bar{x})$ is a noncritical multiplier. The latter is automatic if $\bar{\mu}$ satisfies the SOSC (7.181), and according to our discussion above, the existence of such $\bar{\mu}$ can again be guaranteed under the MPCC-LICQ and the piecewise SOSC (7.184). Hence, for the algorithm outlined, local superlinear convergence is obtained under the same set of assumptions as for the piecewise SQP, namely, the MPCC-LICQ and the piecewise SOSC (7.184) at the solution $\bar{x}$.

We proceed with the formal statement of the *active-set Newton method* for MPCCs. For the KKT system (7.165), we define its residual mapping $\Psi : \mathbf{R}^n \times (\mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}) \to \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$ as follows:

$$\Psi(x, \mu) = \left(\frac{\partial \mathcal{L}}{\partial x}(x, \mu), \psi(\mu_G, G(x)), \psi(\mu_H, H(x)), \psi(\mu_0, -\langle G(x), H(x)\rangle) \right), \tag{7.196}$$

where $\mu = (\mu_G, \mu_H, \mu_0)$, and $\psi : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ is either the natural residual

$$\psi(a, b) = \min\{a, b\}, \tag{7.197}$$

or the Fischer–Burmeister complementarity function

$$\psi(a, b) = a + b - \sqrt{a^2 + b^2}. \tag{7.198}$$

Algorithm 7.36 Choose a complementarity function $\psi : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ defined either by (7.197) or by (7.198), and define further the mapping $\Psi : \mathbf{R}^n \times (\mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}) \to \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$ according to (7.196). Fix some $\tau \in (0, 1)$.

Choose $x^0 \in \mathbf{R}^n$ and $\mu^0 = (\mu_G^0, \mu_H^0, \mu_0^0) \in \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$, and set $k = 0$. Define the index sets

$$I_G = I_G(x^0,\, \mu^0) = \{i = 1,\, \dots,\, m \mid G_i(x^0) \le \|\Psi(x^0,\, \mu^0)\|^\tau\}, \tag{7.199}$$

$$I_H = I_H(x^0,\, \mu^0) = \{i = 1,\, \dots,\, m \mid H_i(x^0) \le \|\Psi(x^0,\, \mu^0)\|^\tau\}. \tag{7.200}$$

Generate an iterative sequence $\{(x^k,\, \lambda^k)\} \subset \mathbf{R}^n \times (\mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R})$, with $\lambda^k = (\lambda_G^k,\, \lambda_H^k) \in \mathbf{R}^m \times \mathbf{R}^m$, as follows:

- Generate the sequence $\{(x^k,\, (\lambda_G^k)_{I_G},\, (\lambda_H^k)_{I_H})\}$ by the Newton–Lagrange method (Algorithm 4.1) applied to problem (7.170), starting from the point $(x^0,\, (\lambda_G^0)_{I_G},\, (\lambda_H^0)_{I_H})$ with $(\lambda_G^0)_{I_G}$ and $(\lambda_H^0)_{I_H}$ defined by

$$\begin{aligned} (\lambda_G^0)_i &= (\mu_G^0)_i - \mu_0^0 H_i(x^0), \quad i \in I_G \setminus I_H, \\ (\lambda_G^0)_i &= (\mu_G^0)_i, \quad i \in I_G \cap I_H, \end{aligned} \tag{7.201}$$

$$\begin{aligned} (\lambda_H^0)_i &= (\mu_H^0)_i - \mu_0^0 G_i(x^0), \quad i \in I_H \setminus I_G, \\ (\lambda_H^0)_i &= (\mu_H^0)_i, \quad i \in I_G \cap I_H. \end{aligned} \tag{7.202}$$

- Set

$$(\lambda_G^k)_{I_H \setminus I_G} = 0, \quad (\lambda_H^k)_{I_G \setminus I_H} = 0 \quad \forall\, k = 0,\, 1,\, \dots. \tag{7.203}$$

Local convergence properties of this active-set Newton method for MPCCs are the following.

Theorem 7.37. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $G : \mathbf{R}^n \to \mathbf{R}^m$ and $H : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a local solution of the MPCC* (7.164)*, satisfying the MPCC-LICQ and the SOSC* (7.181) *for an associated Lagrange multiplier $\bar{\mu} \in \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$. Furthermore, let $\bar{\lambda} \in \mathbf{R}^m \times \mathbf{R}^m$ be the (unique) MPCC-multiplier associated with $\bar{x}$.*

Then any starting point $(x^0,\, \mu^0) \in \mathbf{R}^n \times (\mathbf{R}^m \times \mathbf{R}^m)$ close enough to $(\bar{x},\, \bar{\mu})$ uniquely defines the iterative sequence $\{(x^k,\, \lambda^k)\} \subset \mathbf{R}^n \times (\mathbf{R}^m \times \mathbf{R}^m)$ of Algorithm 7.36, this sequence converges to $(\bar{x},\, \bar{\lambda})$, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the second derivatives of f, G, and H are locally Lipschitz-continuous with respect to $\bar{x}$.

Proof. Proposition 3.58 implies that if $(x^0,\, \mu^0)$ is close enough to $(\bar{x},\, \bar{\mu})$, the index sets I_G and I_H defined in (7.199), (7.200) coincide with their counterparts defined in (7.166).

The definition (7.201), (7.202) of the initial approximation to the MPCC-multiplier $\bar{\lambda}$ is motivated by Proposition 7.32. Specifically, if $(x^0,\, \mu^0)$ is close enough to $(\bar{x},\, \bar{\mu})$, the pair $((\lambda_G^0)_{I_G},\, (\lambda_H^0)_{I_H})$ defined by (7.201), (7.202) will be close enough to $((\bar{\lambda}_G)_{I_G},\, (\bar{\lambda}_H)_{I_H})$ (according to Proposition 7.32, the latter satisfies (7.173), (7.174)). Therefore, according to Theorem 4.3, the Newton–Lagrange method applied to problem (7.170), and using the starting point $(x^0,\, (\lambda_G^0)_{I_G},\, (\lambda_H^0)_{I_H})$, uniquely defines the sequence $\{(x^k,\, (\lambda_G^k)_{I_G},\, (\lambda_H^k)_{I_H})\}$, and this sequence converges to $(\bar{x},\, (\bar{\lambda}_G)_{I_G},\, (\bar{\lambda}_H)_{I_H}))$ superlinearly, or even quadratically under the assumption that the second derivatives of f, G, and H are locally Lipschitz-continuous with respect to $\bar{x}$. At the same time, according to (7.167) and (7.203), it holds that

$$(\lambda_G^k)_{I_H \setminus I_G} = (\bar{\lambda}_G)_{I_H \setminus I_G} = 0, \quad (\lambda_H^k)_{I_G \setminus I_H} = (\bar{\lambda}_H)_{I_G \setminus I_H} = 0 \quad \forall\, k = 0,\, 1,\, \ldots.$$

The assertions follow. □

It is worth to comment that Algorithm 7.36 appears more suitable for globalization than the piecewise SQP (Algorithm 7.34). The key issue in this respect is the following. Algorithm 7.36 uses as a dual starting point an approximation of a Lagrange multiplier rather than an approximation of an MPCC-multiplier. The proximity to points satisfying the KKT system (7.165) (and hence, to Lagrange multipliers) can be controlled by some globally defined merit functions, like the norm of Ψ in (7.196), for example. By contrast, the relations (7.167), (7.168) characterizing MPCC-multipliers involve the index sets I_G and I_H depending on a specific solution $\bar{x}$. Consequently, it is difficult (perhaps impossible) to suggest a reasonable globally defined merit function that involves MPCC-multipliers and can measure proximity to them.

Furthermore, having in mind globalization of convergence, it is useful to consider a modified version of Algorithm 7.36, with identification being performed not only once (at the start of the process) but at the beginning of each iteration. For this, one needs to generate not only the sequence $\{(x^k,\, \lambda^k)\}$ but also an appropriate sequence $\{\mu^k\} \subset \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$ approximating some element of $\mathcal{M}(\bar{x})$, and redefine I_G and I_H accordingly:

$$I_G = I_G(x^k,\, \mu^k) = \{i = 1,\, \ldots,\, m \mid G_i(x^k) \leq \|\Psi(x^k,\, \mu^k)\|^\tau\},$$

$$I_H = I_H(x^k,\, \mu^k) = \{i = 1,\, \ldots,\, m \mid H_i(x^k) \leq \|\Psi(x^k,\, \mu^k)\|^\tau\}$$

for each $k = 0,\, 1,\, \ldots$. This identification is a cheap procedure and, therefore, the modification does not increase computational costs to any extent to be concerned about. However, the required convergence analysis becomes much more involved. For all the conclusions of Theorem 7.37 to remain valid for this modified algorithm, one needs to keep $\{\mu^k\}$ close to $\bar{\mu}$, which can be achieved by keeping it close to μ^0. In particular, one can just take $\mu^k = \mu^0$ for all $k = 1,\, 2,\, \ldots$. A less trivial option, more suitable for globalization purposes (and more in the spirit of SQP methods), is stated and analyzed in [143]. Some related hybrid globalized methods are also developed and tested in this reference.

7.3.4 Newton-Type Methods for Lifted Reformulations

The following idea, called "lifting MPCC," is proposed in [258]. Consider in the space $\mathbf{R}^2$ of variables $(a,\, b)$ the set defined by the basic complementarity condition:

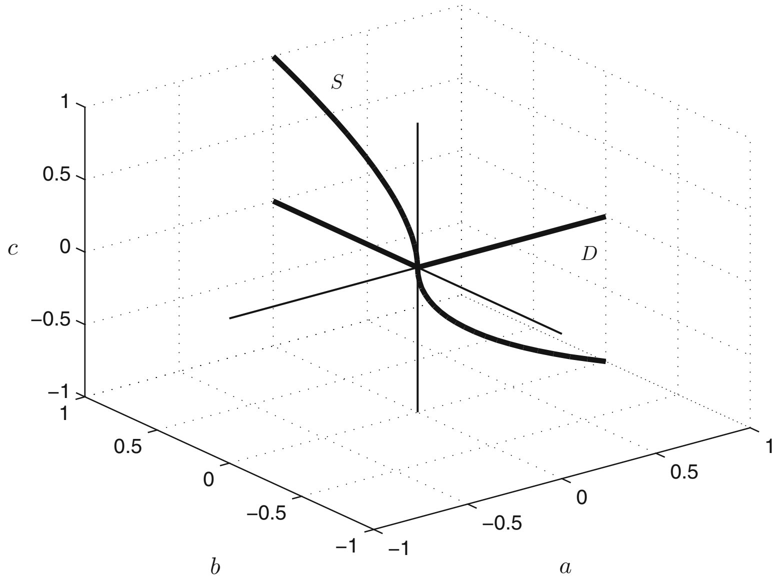

Fig. 7.10 Curves D and S

$$D = \{(a,\, b) \in \mathbf{R}^2 \mid a \geq 0,\ b \geq 0,\ ab = 0\}. \tag{7.204}$$

This set is "nonsmooth," in the sense that it has a kink at the origin. Introducing an artificial variable $c \in \mathbf{R}$, consider a smooth curve S in the space $\mathbf{R}^3$ of variables $(a,\, b,\, c)$ such that the projection of S onto the plane $(a,\, b)$ coincides with the set D. This can be done, for example, as follows:

$$S = \{(a,\, b,\, c) \in \mathbf{R}^3 \mid a = (-\min\{0,\, c\})^s,\ b = (\max\{0,\, c\})^s\} \tag{7.205}$$

with any $s > 1$.

The set D given by (7.204) and the curve S given by (7.205) with $s = 2$, are shown in Fig. 7.10 by bold lines. The curve S is the intersection of two smooth surfaces shown in Fig. 7.11.

The construction above gives rise to the *lifted reformulation* of the original MPCC (7.164), defined by

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & (-\min\{0,\, y\})^s - G(x) = 0,\ (\max\{0,\, y\})^s - H(x) = 0. \end{array} \tag{7.206}$$

As is easy to see, a point $\bar{x} \in \mathbf{R}^n$ is a (local) solution of (7.164) if and only if the point $(\bar{x},\, \bar{y}) \in \mathbf{R}^n \times \mathbf{R}^m$ is a (local) solution of (7.206) with $\bar{y}$ uniquely defined by $\bar{x}$. In [258], it is suggested to use the power $s = 3$. If $s > 2$, the constraints in the reformulation (7.206) are twice differentiable equalities, which appear much simpler than the original complementarity constraints in (7.164). Of course, this does not come for free: a closer look reveals that

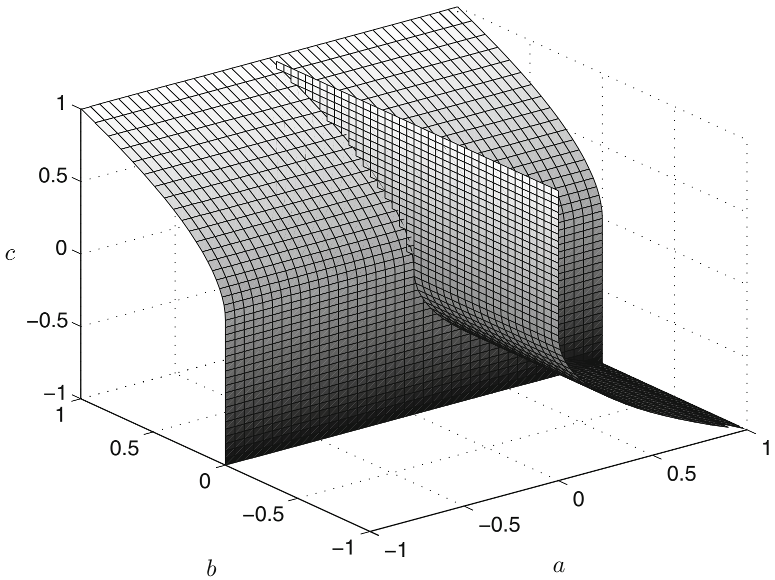

Fig. 7.11 Surfaces $a = (-\min\{0,\, c\})^2$ and $b = (\max\{0,\, c\})^2$

the Jacobian of the Lagrange optimality system of the reformulation (7.206) with $s > 2$ is inevitably degenerate whenever the lower-level strict complementarity does not hold at $\bar{x}$. Geometrically, degeneracy arising in the lifted reformulation has to do with the fact that the tangent of the smooth curve S (regardless of the power $s > 1$ used in its definition) at the point $(0,\, 0,\, 0)$ is always vertical; see Fig. 7.10. Therefore, the objective function is constant along the "vertical" tangent subspace to the feasible set of problem (7.206) at $(\bar{x},\, \bar{y})$, i.e., the subspace consisting of $(\xi,\, \eta) \in \mathbf{R}^n \times \mathbf{R}^m$ satisfying $\xi = 0$, $\eta_{I_H \setminus I_G} = 0$, $\eta_{I_G \setminus I_H} = 0$. Moreover, when $s > 2$, the constraints of problem (7.206) cannot contribute to the Hessian of its Lagrangian along this subspace. These two facts mean that no reasonable second-order sufficient optimality condition can hold for this problem at $(\bar{x},\, \bar{y})$, which leads to degeneracy. Formally, at any feasible point of (7.206) with $s > 2$, the derivatives of the Lagrangian with respect to y_i for indices $i \in I_0$ have zero gradients, and thus the Jacobian matrix has zero rows.

The issue of degeneracy of twice differentiable lifted MPCC is similar in nature to degeneracy of smooth equation reformulations of nonlinear complementarity problems (NCP) in the absence of strict complementarity, as discussed in Sect. 3.2.1. At the same time, nonsmooth equation reformulations of NCP, such as based on the natural residual or the Fischer–Burmeister function, can have appropriate regularity properties without strict complementarity and are thus generally preferred for constructing Newton-type methods for NCP; see Sect. 3.2. Drawing on this experience for NCP, in [154, 156] it is suggested to use, instead of power $s = 3$ (which leads to degeneracy of the Lagrange optimality system of the lifted MPCC reformulation) the power $s = 2$ which leads to its nonsmoothness but can be expected to have better regularity properties. Specifically, consider the reformulation of the MPCC (7.164) given by

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & (\min\{0,\, y\})^2 - G(x) = 0,\ (\max\{0,\, y\})^2 - H(x) = 0. \end{array} \tag{7.207}$$

As we shall show, nonsmoothness of the Lagrange optimality system of the lifted MPCC (7.207) is structured, so that we can apply the semismooth Newton method presented in Sect. 2.4. Moreover, local superlinear convergence of this method is guaranteed under reasonable assumptions. Furthermore, it turns out that the squared residual of the Lagrange optimality system of (7.207) is actually continuously differentiable, even though the system itself is not. This opens the way to a natural globalization of the local semismooth Newton method. The latter is again similar to the nonsmooth Fischer–Burmeister equality reformulation of NCP, for which the squared residual becomes smooth and can be used for globalization, as discussed in Sect. 5.1.

We proceed with establishing some relations between the original MPCC (7.164) and its lifted reformulation (7.207). Note first that the value $\bar{y}$ of the artificial variable y that corresponds to any given feasible point $\bar{x}$ of problem (7.164) is uniquely defined: the point $(\bar{x},\, \bar{y})$ is feasible in (7.207) if and only if

$$\bar{y}_{I_H \setminus I_G} = -(G_{I_H \setminus I_G}(\bar{x}))^{1/2}, \quad \bar{y}_{I_G \setminus I_H} = (H_{I_G \setminus I_H}(\bar{x}))^{1/2}, \quad \bar{y}_{I_0} = 0. \tag{7.208}$$

Furthermore, it is immediate that $\bar{x}$ is a (local) solution of the original problem (7.164) if and only if $(\bar{x},\, \bar{y})$ is a (local) solution of the lifted MPCC reformulation (7.207). In addition, it is also easy to see that the MPCC-LICQ at a point $\bar{x}$ feasible in the MPCC (7.164) is equivalent to the usual regularity condition for the equality-constrained lifted MPCC (7.207) at $(\bar{x},\, \bar{y})$.

Define the usual Lagrangian $L^{\uparrow} : \mathbf{R}^n \times \mathbf{R}^m \times (\mathbf{R}^m \times \mathbf{R}^m) \to \mathbf{R}$ of the lifted problem (7.207):

$$L^{\uparrow}(x,\, y,\, \lambda) = f(x) + \langle \lambda_G,\, (\min\{0,\, y\})^2 - G(x)\rangle + \langle \lambda_H,\, (\max\{0,\, y\})^2 - H(x)\rangle,$$

where $\lambda = (\lambda_G,\, \lambda_H) \in \mathbf{R}^m \times \mathbf{R}^m$. The Lagrange optimality system characterizing stationary points of problem (7.207) and the associated multipliers is given by

$$\begin{array}{cc} \dfrac{\partial L^{\uparrow}}{\partial x}(x,\, y,\, \lambda) = 0, & \dfrac{\partial L^{\uparrow}}{\partial y}(x,\, y,\, \lambda) = 0, \\ (\min\{0,\, y\})^2 - G(x) = 0, & (\max\{0,\, y\})^2 - H(x) = 0, \end{array} \tag{7.209}$$

where

$$\frac{\partial L^{\uparrow}}{\partial x}(x,\, y,\, \lambda) = \frac{\partial \mathcal{L}}{\partial x}(x,\, \lambda), \tag{7.210}$$

$$\frac{\partial L^{\uparrow}}{\partial y_i}(x,\, y,\, \lambda) = 2(\lambda_G)_i \min\{0,\, y_i\} + 2(\lambda_H)_i \max\{0,\, y_i\}, \quad i = 1,\, \ldots,\, m. \tag{7.211}$$

Observe that for any $i \in \{1,\, \ldots,\, m\}$, the right-hand side in (7.211) is not differentiable at points $(x,\, y,\, \lambda)$ such that $y_i = 0$.

The following correspondence between stationary points and multipliers for the original problem (7.164) and its lifted reformulation (7.207) is established by direct verification.

Proposition 7.38. *Let $f : \mathbf{R}^n \to \mathbf{R}$, $G : \mathbf{R}^n \to \mathbf{R}^m$ and $H : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable at $\bar{x} \in \mathbf{R}^n$. Let $\bar{x}$ be feasible in the MPCC (7.164).*

If $\bar{x}$ is a strongly stationary point of (7.164) and $\bar{\lambda} = (\bar{\lambda}_G,\, \bar{\lambda}_H)$ is an associated MPCC-multiplier, then the point $(\bar{x},\, \bar{y})$ with $\bar{y}$ given by (7.208) is a stationary point of the lifted MPCC (7.207), and $\bar{\lambda}$ is an associated Lagrange multiplier.

Conversely, if $(\bar{x},\, \bar{y})$ is a stationary point of the lifted MPCC (7.207), then $\bar{x}$ is a weakly stationary point of the MPCC (7.164). In addition, if there exists a Lagrange multiplier $\bar{\lambda} = (\bar{\lambda}_G,\, \bar{\lambda}_H)$ associated with $(\bar{x},\, \bar{y})$ and such that $(\bar{\lambda}_G)_{I_0} \geq 0$ and $(\bar{\lambda}_H)_{I_0} \geq 0$, then $\bar{x}$ is a strongly stationary point of the MPCC (7.164), and $\bar{\lambda}$ is an associated MPCC-multiplier.

We shall next consider the semismooth Newton method applied to the Lagrange optimality system (7.209) of the lifted MPCC reformulation (7.207). Taking into account the relation (7.210), this system takes the form of the equation

$$\Phi(u) = 0, \tag{7.212}$$

where $\Phi : \mathbf{R}^n \times \mathbf{R}^m \times (\mathbf{R}^m \times \mathbf{R}^m) \to \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}^m$ is given by

$$\Phi(u) = \left(\frac{\partial \mathcal{L}}{\partial x}(x, \lambda),\ \frac{\partial L^{\uparrow}}{\partial y}(x, y, \lambda),\ (\min\{0, y\})^2 - G(x),\ (\max\{0, y\})^2 - H(x)\right), \tag{7.213}$$

with $u = (x,\, y,\, \lambda)$.

Using (7.211) and (7.213), by direct calculations it follows that $\partial_B \Phi(u)$ at an arbitrary point $u = (x,\, y,\, \lambda) \in \mathbf{R}^n \times \mathbf{R}^m \times (\mathbf{R}^m \times \mathbf{R}^m)$ consists of all the matrices of the form

$$J = \begin{pmatrix} \frac{\partial^2 \mathcal{L}}{\partial x^2}(x,\, \lambda) & 0 & -(G'(x))^{\mathrm{T}} & -(H'(x))^{\mathrm{T}} \\ 0 & 2A(y,\, \lambda) & 2B_{\min}(y) & 2B_{\max}(y) \\ -G'(x) & 2B_{\min}(y) & 0 & 0 \\ -H'(x) & 2B_{\max}(y) & 0 & 0 \end{pmatrix}, \tag{7.214}$$

where

$$A(y,\, \lambda) = \operatorname{diag}(a(y,\, \lambda)), \tag{7.215}$$

$$B_{\min}(y) = \operatorname{diag}(\min\{0,\, y\}), \quad B_{\max}(y) = \operatorname{diag}(\max\{0,\, y\}), \tag{7.216}$$

and the vector $a(y,\, \lambda) \in \mathbf{R}^m$ is defined by

$$a_i = \begin{cases} (\lambda_G)_i & \text{if } y_i < 0, \\ (\lambda_G)_i \text{ or } (\lambda_H)_i & \text{if } y_i = 0, \\ (\lambda_H)_i & \text{if } y_i > 0, \end{cases} \quad i = 1,\, \ldots,\, m. \tag{7.217}$$

The *semismooth Newton–Lagrange method for the lifted MPCC reformulation* (7.207) is therefore given by the following.

Algorithm 7.39 Define $\Phi : \mathbf{R}^n \times \mathbf{R}^m \times (\mathbf{R}^m \times \mathbf{R}^m) \to \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}^m$ according to (7.213). Choose $u^0 = (x^0,\, y^0,\, \lambda^0) \in \mathbf{R}^n \times \mathbf{R}^m \times (\mathbf{R}^m \times \mathbf{R}^m)$ and set $k = 0$.

1. If $\Phi(u^k) = 0$, stop.
2. Compute a matrix $J_k = J$ by (7.214)–(7.217) with $(x,\, y,\, \lambda) = (x^k,\, y^k,\, \lambda^k)$. Compute $u^{k+1} = (x^{k+1},\, y^{k+1},\, \lambda^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^m \times (\mathbf{R}^m \times \mathbf{R}^m)$ as a solution of the linear system
$$\Phi(u^k) + J_k(u - u^k) = 0. \tag{7.218}$$
3. Increase k by 1 and go to step 1.

Let $\bar{x}$ be a strongly stationary point of the original MPCC (7.164), and let $\bar{\lambda}$ be an associated MPCC-multiplier. Let $\bar{y}$ be given by (7.208). According to Theorem 2.42 and Remarks 2.47, 2.54, the local superlinear (quadratic) convergence of Algorithm 7.39 to the solution $\bar{u} = (\bar{x},\, \bar{y},\, \bar{\lambda})$ of the equation (7.212) would be established if we show (strong) semismoothness and *BD*-regularity of Φ at $\bar{u}$. The needed semismoothness properties readily follow from Proposition 1.75.

Proposition 7.40. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $G,\, H : \mathbf{R}^n \to \mathbf{R}^m$ be differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, with their derivatives being semismooth at this point.*

Then for any $y \in \mathbf{R}^m$ and $\lambda = (\lambda_G,\, \lambda_H) \in \mathbf{R}^m \times \mathbf{R}^m$, the mapping Φ defined in (7.213) *is semismooth at $(\bar{x},\, y,\, \lambda)$. Moreover, if f, G, and H are twice differentiable around $\bar{x}$, with their derivatives being locally Lipschitz-continuous with respect to $\bar{x}$, then Φ is strongly semismooth at $(\bar{x},\, y,\, \lambda)$.*

Now, according to (7.214)–(7.217), and taking also into account (7.168) and (7.208), we obtain that $\partial_B \Phi(\bar{u})$ consists of all the matrices of the form

$$J = \begin{pmatrix} \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x},\, \bar{\lambda}) & 0 & -(G'(\bar{x}))^{\mathrm{T}} & -(H'(\bar{x}))^{\mathrm{T}} \\ 0 & 2A & 2B_{\min} & 2B_{\max} \\ -G'(\bar{x}) & 2B_{\min} & 0 & 0 \\ -H'(\bar{x}) & 2B_{\max} & 0 & 0 \end{pmatrix}, \tag{7.219}$$

where

$$A = \mathrm{diag}(a), \quad B_{\min} = \mathrm{diag}(b_{\min}), \quad B_{\max} = \mathrm{diag}(b_{\max}), \tag{7.220}$$

with the vector $a \in \mathbf{R}^m$ given by

$$a_i = \begin{cases} 0 & \text{if } i \in \{1, \ldots, m\} \setminus I_0, \\ (\bar{\mu}_G)_i \text{ or } (\bar{\mu}_H)_i & \text{if } i \in I_0, \end{cases} \tag{7.221}$$

and the vectors $b_{\min} \in \mathbf{R}^m$ and $b_{\max} \in \mathbf{R}^m$ given by

$$(b_{\min})_{I_H \setminus I_G} = -(G_{I_H \setminus I_G}(\bar{x}))^{1/2}, \quad (b_{\min})_{I_G} = 0, \tag{7.222}$$

$$(b_{\max})_{I_G \setminus I_H} = (H_{I_G \setminus I_H}(\bar{x}))^{1/2}, \quad (b_{\max})_{I_H} = 0. \tag{7.223}$$

We next show that the mapping Φ is BD-regular at $\bar{u}$ under reasonable assumptions.

Proposition 7.41. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $G, H : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable at $\bar{x} \in \mathbf{R}^n$ which is a strongly stationary point of the MPCC (7.164), satisfying the MPCC-LICQ. Let $\bar{\lambda}$ be the (unique) associated MPCC-multiplier. Assume finally that the upper-level strict complementarity condition (7.185) and the second-order sufficient condition (7.187) with $K(\bar{x})$ defined in (7.186) are satisfied.*

Then Φ defined by (7.213) is BD-regular at $\bar{u} = (\bar{x}, \bar{y}, \bar{\lambda})$, where $\bar{y}$ is given by (7.208).

Proof. Suppose that for some matrix $J \in \partial\Phi_B(\bar{u})$ and some $\xi \in \mathbf{R}^n$, $\eta \in \mathbf{R}^m$, $\zeta_G \in \mathbf{R}^m$ and $\zeta_H \in \mathbf{R}^m$ it holds that $Jv = 0$, where $v = (\xi, \eta, (\zeta_G, \zeta_H))$. According to (7.219)–(7.223), taking into account also (7.185), we then conclude that

$$\frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\lambda})\xi - (G'(\bar{x}))^{\mathrm{T}}\zeta_G - (H'(\bar{x}))^{\mathrm{T}}\zeta_H = 0, \tag{7.224}$$

$$(\zeta_G)_{I_H \setminus I_G} = 0, \quad (\zeta_H)_{I_G \setminus I_H} = 0, \quad \eta_{I_0} = 0, \tag{7.225}$$

$$-\langle G_i'(\bar{x}), \xi\rangle - 2(G_i(\bar{x}))^{1/2}\eta_i = 0, \ i \in I_H \setminus I_G, \quad G'_{I_G}(\bar{x})\xi = 0, \tag{7.226}$$

$$-\langle H_i'(\bar{x}), \xi\rangle + 2(H_i(\bar{x}))^{1/2}\eta_i = 0, \ i \in I_G \setminus I_H, \quad H'_{I_H}(\bar{x})\xi = 0. \tag{7.227}$$

The second relations in (7.226) and (7.227) mean that $\xi \in K(\bar{x})$, see (7.186). Moreover, multiplying both sides of the equality (7.224) by ξ and using the first two relations in (7.225) and the second relations in (7.226) and (7.227), we obtain that

$$0 = \left\langle \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\lambda})\xi, \xi \right\rangle - \langle \zeta_G, G'(\bar{x})\xi\rangle - \langle \zeta_H, H'(\bar{x})\xi\rangle = \left\langle \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\lambda})\xi, \xi \right\rangle.$$

Hence, by (7.187), we conclude that $\xi = 0$.

Substituting now $\xi = 0$ into (7.224) and using the first two relations in (7.225), we have that

$$(G'_{I_G}(\bar{x}))^{\mathrm{T}}(\zeta_G)_{I_G} + (H'_{I_H}(\bar{x}))^{\mathrm{T}}(\zeta_H)_{I_H} = 0.$$

From the latter equality and from the MPCC-LICQ, it follows that

$$(\zeta_G)_{I_G} = 0, \quad (\zeta_H)_{I_H} = 0. \tag{7.228}$$

In addition, using $\xi = 0$ in the first relations of (7.226) and (7.227), we have that

$$\eta_{I_H \setminus I_G} = 0, \quad \eta_{I_G \setminus I_H} = 0. \tag{7.229}$$

Combining (7.225), (7.228), and (7.229) gives that $\eta = 0$, $\zeta_G = 0$, $\zeta_H = 0$, i.e., $v = 0$.

We have thus shown that for any matrix J in question, $\ker J = \{0\}$ holds, i.e., all these matrices are nonsingular. □

Given Theorem 2.42, Remarks 2.47, 2.54, and Propositions 7.40 and 7.41, we immediately obtain local convergence and rate of convergence for Algorithm 7.39.

Theorem 7.42. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $G, H : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbf{R}^n$, and let their second derivatives be continuous at $\bar{x}$. Let $\bar{x}$ be a strongly stationary point of the MPCC (7.164), satisfying the MPCC-LICQ , and let $\bar{\lambda}$ be the (unique) MPCC-multiplier associated with $\bar{x}$. Assume, finally, that the upper-level strict complementarity condition (7.185) and the second-order sufficient condition (7.187) with $K(\bar{x})$ defined in (7.186) are satisfied.*

Then any starting point $(x^0, y^0, \lambda^0) \in \mathbf{R}^n \times \mathbf{R}^m \times (\mathbf{R}^m \times \mathbf{R}^m)$ close enough to $(\bar{x}, \bar{y}, \bar{\lambda})$ defines a particular iterative sequence of Algorithm 7.39, any such sequence converges to $(\bar{x}, \bar{y}, \bar{\lambda})$, and the rate of convergence is superlinear. If the second derivatives of f, G, and H are locally Lipschitz-continuous with respect to $\bar{x}$, then the rate of convergence is quadratic.

We next comment on how the local convergence result in Theorem 7.42 compares to some alternatives.

As already discussed above, using the lifted reformulation (7.206) with $s > 2$ gives the Lagrange optimality system which is smooth but inherently degenerate (at least in the absence of the restrictive lower-level strict complementarity). Nevertheless, this degeneracy is again structured and can be tackled, in principle, by the tools for solving degenerate equations and reformulations of complementarity problems developed in [138]. However, careful consideration of the approach of [138] applied to the lifted reformulation (7.206) shows that it would require the same assumptions as those for Algorithm 7.39 in Theorem 7.42, while the approach itself is quite a bit more complicated. In addition, methods in [138] do not come with natural globalization strategies, while Algorithm 7.39 allows natural globalization by linesearch, as will be discussed below.

On the other hand, local superlinear convergence of the piecewise SQP method and of the active-set Newton method discussed in Sect. 7.3.3 had been shown in Theorems 7.35 and 7.37 under the assumptions weaker than those for Algorithm 7.39 in Theorem 7.42. Specifically, those methods do not require the upper-level strict complementarity (7.185). But, just as in the case of [138], the piecewise SQP and the active-set Newton method for MPCCs come without ready-to-use globalization strategies.

Observe finally that the mapping Φ defined in (7.213) is actually piecewise smooth (under the appropriate smoothness assumptions on the original problem data), and the semismooth Newton method specified in Algorithm 7.39 can be interpreted as the piecewise smooth Newton method discussed in Remark 2.55.

A natural globalization strategy for Algorithm 7.39 can be developed by introducing linesearch for the merit function $\varphi\colon \mathbf{R}^n \times \mathbf{R}^m \times (\mathbf{R}^m \times \mathbf{R}^m) \to \mathbf{R}$,

$$\varphi(u) = \frac{1}{2}\|\Phi(u)\|^2, \tag{7.230}$$

where the mapping Φ is defined in (7.213). This development turns out to be similar in spirit to the linesearch algorithms employing the Fischer–Burmeister function for reformulation of complementarity problems; see Sect. 5.1. The reason is that the merit function φ in (7.230) happens to be continuously differentiable, even though Φ itself is not; moreover, the gradient of φ is explicitly computable using any element of the B-differential of the mapping Φ.

Proposition 7.43. *Let $f : \mathbf{R}^n \to \mathbf{R}$ and $G,\, H : \mathbf{R}^n \to \mathbf{R}^m$ be twice differentiable at $x \in \mathbf{R}^n$.*

Then for any $y \in \mathbf{R}^m$ and $\lambda = (\lambda_G,\, \lambda_H) \in \mathbf{R}^m \times \mathbf{R}^m$, the function φ defined by (7.213) and (7.230) is differentiable at the point $u = (x,\, y,\, \lambda)$, and it holds that

$$\varphi'(u) = J\Phi(u) \quad \forall\, J \in \partial_B \Phi(u). \tag{7.231}$$

Moreover, if f, G, and H are twice continuously differentiable on an open set $O \subset \mathbf{R}^n$, then the function φ is continuously differentiable on the set $O \times \mathbf{R}^m \times (\mathbf{R}^m \times \mathbf{R}^m)$.

Proof. Nonsmoothness of φ can only be induced by the components of Φ that correspond to partial derivatives of $L^{\uparrow}$ with respect to y (see (7.213)); all the other components of Φ are sufficiently smooth under the stated assumptions.

Observe that for any $t \in \mathbf{R}$ it holds that

$$\min\{0,\, t\} \max\{0,\, t\} = 0.$$

Therefore, for each $i = 1,\, \ldots,\, m$, from (7.211) it follows that

$$\left(\frac{\partial L^{\uparrow}}{\partial y_i}(x,\, y,\, \lambda)\right)^2 = 4((\lambda_G)_i^2(\min\{0,\, y_i\})^2 + (\lambda_H)_i^2(\max\{0,\, y_i\})^2), \quad (7.232)$$

where the right-hand side is a differentiable function in the variables $y \in \mathbf{R}^m$ and $\lambda = (\lambda_G,\, \lambda_H) \in \mathbf{R}^m \times \mathbf{R}^m$. This shows that φ has the announced differentiability properties.

Furthermore, from (7.232) it follows that

$$\left(\frac{1}{2}\left\|\frac{\partial L^{\uparrow}}{\partial y}(x,\, y,\, \lambda)\right\|^2\right)' = \begin{pmatrix} 0 \\ 4((\lambda_G)_1^2(\min\{0,\, y_1\}) + (\lambda_H)_1^2(\max\{0,\, y_1\})) \\ \cdots \\ 4((\lambda_G)_1^2(\min\{0,\, y_m\}) + (\lambda_H)_m^2(\max\{0,\, y_m\})) \\ 4(\lambda_G)_1(\min\{0,\, y_1\})^2 \\ \cdots \\ 4(\lambda_G)_m(\min\{0,\, y_m\})^2 \\ 4(\lambda_H)_1(\max\{0,\, y_1\})^2 \\ \cdots \\ 4(\lambda_H)_m(\max\{0,\, y_m\})^2 \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ 2A(y,\, \lambda) \\ 2B_{\min}(y) \\ 2B_{\max}(y) \end{pmatrix} \frac{\partial L^{\uparrow}}{\partial y}(x,\, y,\, \lambda), \quad (7.233)$$

where the last equality is by (7.211) and (7.215)–(7.217). Differentiating the other parts of φ and combining the result with (7.233) and with (7.213)–(7.217) gives the equality (7.231). □

Recall that according to (7.214)–(7.217), all the matrices $J \in \partial_B \Phi(u)$ are symmetric at any $u \in \mathbf{R}^n \times \mathbf{R}^m \times (\mathbf{R}^m \times \mathbf{R}^m)$. Using this fact, as well as (7.231), for any such matrix and for any direction $v \in \mathbf{R}^n \times \mathbf{R}^m \times (\mathbf{R}^m \times \mathbf{R}^m)$ computed as a solution of the linear system

$$Jv = -\Phi(u),$$

it holds that

$$\begin{aligned} \langle \varphi'(u),\, v\rangle &= \langle J\Phi(u),\, v\rangle = \langle \Phi(u),\, Jv\rangle = -\langle \Phi(u),\, \Phi(u)\rangle = -\|\Phi(u)\|^2 \\ &= -2\varphi(u). \end{aligned} \quad (7.234)$$

In particular, if the point u is not a solution of the equation (7.212) or, equivalently, is not a global minimizer of the function φ, then v is a descent direction for φ at the point u. This immediately suggests a natural globalization strategy for the local semismooth Newton method in Algorithm 7.39; see [156] for further details.

Of course, globalization based on the squared residual of the Lagrange system of problem (7.207) is not the only possibility. Alternatively, one can employ the l_1-penalty function $\varphi_c : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$,

$$\varphi_c(x,\, y) = f(x) + c\psi(x,\, y),$$

where $c > 0$ is the penalty parameter, and $\psi : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}_+$ is the l_1-penalty for the constraints in (7.207), i.e.,

$$\psi(x,\, y) = \|((\min\{0,\, y\})^2 - G(x),\, (\max\{0,\, y\})^2 - H(x))\|_1.$$

A strategy following this pattern, more in the spirit of classical SQP algorithms (see Sect. 6.2) is given in [154]; it takes into account special structure of the lifted MPCC (7.207) and can be advantageous compared to the use of the residual of the Lagrange system as a merit function. In particular, it is natural to expect that globalization using the penalty function, being more optimization-oriented, should outperform globalization based on the residual in terms of the "quality" of computed solutions. The former should be less attracted by nonoptimal stationary points than the latter, and this is confirmed by the numerical experiments in [154] on the MacMPEC test collection [185].

It should be also emphasized that it is not appropriate just to apply generic SQP with the l_1-penalty linesearch globalization to the lifted MPCC (7.207). Globalization strategies of this kind require positive definite modifications of the matrices in the generalized Hessian of the Lagrangian of problem (7.207). In principle, this can be done by quasi-Newton updates, such as BFGS with Powell's correction (see Sect. 4.1.1). However, as discussed in Sect. 2.4.1, for systems of nonlinear equations quasi-Newton methods with standard updates are guaranteed to preserve superlinear convergence only when the equation mapping is actually differentiable at the solution, while the Lagrange optimality system (7.209) of (7.207) is not differentiable. For this reason, instead of using standard quasi-Newton updates to make the matrices positive definite, it is more promising to use special modification of the matrices in the generalized Hessian of the Lagrangian, directly linked to the structure of the problem at hand. Employing Theorem 4.65, it can be shown that the specific modification proposed in [154] can be expected to preserve high convergence rate. We refer to [154] for details of this special quasi-Newton scheme, for global convergence analysis of the resulting algorithm, and the description of its performance on the MacMPEC test collection [185].

7.3.5 Mathematical Programs with Vanishing Constraints

In this section we briefly discuss another kind of intrinsically degenerate optimization problems, the so-called *mathematical program with vanishing constraints* (MPVC). MPVCs are closely related to MPCCs, but they are different and there are some good reasons why they should be considered separately. MPVC is the following problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & H_i(x) \geq 0, \quad G_i(x)H_i(x) \leq 0, \; i = 1, \ldots, m, \end{array} \tag{7.235}$$

where $f : \mathbf{R}^n \to \mathbf{R}$ is a smooth function and $G, H : \mathbf{R}^n \to \mathbf{R}^m$ are smooth mappings.

The name of this problem class, introduced in [1], is motivated by the following observation: if for some index $i \in \{1, \ldots, m\}$ and some $x \in \mathbf{R}^n$ it holds that $H_i(x) = 0$, then the corresponding component in the second group of constraints in (7.235) holds automatically, i.e., it "vanishes" from the problem. On the other hand, if $H_i(x) > 0$, then the corresponding component takes the form of the inequality $G_i(x) \leq 0$. MPVC is therefore a prototype of the following problem with switch-on/switch-off constraints:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & G_i(x) \leq 0 \text{ if } x \in D_i, \; i = 1, \ldots, m, \end{array} \tag{7.236}$$

where $D_i \subset \mathbf{R}^n$, $i = 1, \ldots, m$. At points $x \in D_i$ the constraint $G_i(x) \leq 0$ is switched on, while at points $x \notin D_i$ it is switched off. When the sets D_i are represented by the inequalities $H_i(x) > 0$, the general problem (7.236) is precisely the MPVC (7.235). Problem (7.236) is a useful modelling paradigm. In particular, MPVC appears to be a convenient modelling framework for problems of optimal topology design of elastic mechanical structures; see [1, 2] for applications of this kind.

We note that, similarly to MPCCs, we could easily consider the more general problem with the additional "usual" equality and inequality constraints. The usual constraints do not add any significant structural difficulties. We consider the simpler setting (7.235) to keep the notation lighter and the main ideas more transparent.

It turns out that vanishing constraints also usually violate standard constraint qualifications, which again naturally leads to theoretical and computational difficulties. To illustrate this, consider the following model example.

Example 7.44. Let $n = 2$, $m = 1$, $f(x) = x_2 + \sigma x_1^2/2$, $G(x) = x_1$, $H(x) = x_2$, where $\sigma \in \mathbf{R}$ is a parameter.

The feasible set of problem (7.235) with this data is the union of an orthant and a ray; see Fig. 7.12. This problem has a solution if and only if $\sigma \geq 0$, in which case $\bar{x} = 0$ always belongs to the solution set. In particular, $\bar{x} = 0$ is the unique solution if and only if $\sigma > 0$. Evidently, the point $\bar{x}$ does not satisfy the MFCQ.

The situation of Example 7.44 is typical in some sense, as discussed next.

For a feasible point $\bar{x} \in \mathbf{R}^n$ of problem (7.235), define the index sets (associated with $\bar{x}$)

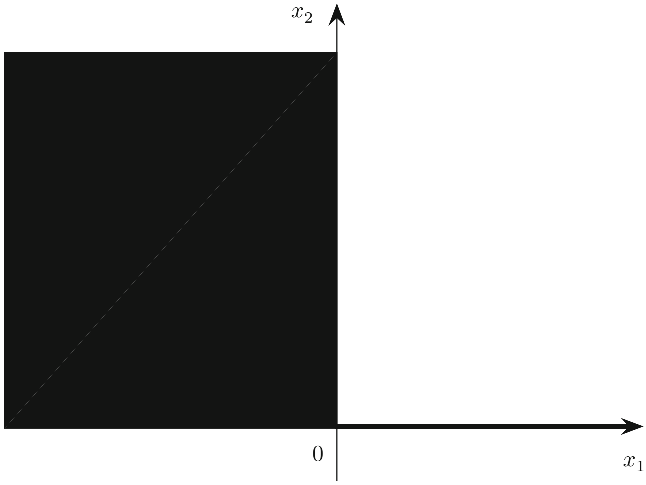

Fig. 7.12 Feasible set of MPVC in Example 7.44

$$
\begin{aligned}
I_+ = I_+(\bar{x}) = \{i = 1, \ldots, m \mid H_i(\bar{x}) > 0\},\\
I_0 = I_0(\bar{x}) = \{i = 1, \ldots, m \mid H_i(\bar{x}) = 0\},
\end{aligned}
$$

and further partitionings of I_+ and I_0:

$$
\begin{aligned}
I_{+0} = I_{+0}(\bar{x}) = \{i \in I_+ \mid G_i(\bar{x}) = 0\},\\
I_{+-} = I_{+-}(\bar{x}) = \{i \in I_+ \mid G_i(\bar{x}) < 0\},
\end{aligned}
$$

and

$$
\begin{aligned}
I_{0+} = I_{0+}(\bar{x}) = \{i \in I_0 \mid G_i(\bar{x}) > 0\},\\
I_{00} = I_{00}(\bar{x}) = \{i \in I_0 \mid G_i(\bar{x}) = 0\},\\
I_{0-} = I_{0-}(\bar{x}) = \{i \in I_0 \mid G_i(\bar{x}) < 0\}.
\end{aligned}
$$

Then locally (near $\bar{x}$) the feasible set of the MPVC (7.235) is defined by the constraints

$$
\begin{gathered}
H_{I_{0+}}(x) = 0, \quad H_{I_{00} \cup I_{0-}}(x) \geq 0, \quad G_{I_{+0}}(x) \leq 0,\\
G_i(x)H_i(x) \leq 0, \; i \in I_{00}.
\end{gathered} \tag{7.237}
$$

The set given by the system (7.237) can be naturally partitioned into branches associated with the partitions of the index set I_{00}. The combinatorial nature of the underlying structure is evident. On the other hand, if $I_{00} = \emptyset$, then the combinatorial aspect disappears, as (7.237) reduces to

$$
H_{I_{0+}}(x) = 0, \quad H_{I_{0-}}(x) \geq 0, \quad G_{I_{+0}}(x) \leq 0.
$$

The property $I_{00} = I_{00}(\bar{x}) = \emptyset$ can be regarded as the *lower-level strict complementarity condition* at a feasible point $\bar{x}$ of the MPVC (7.235). Clearly, this condition simplifies local structure dramatically. However, the lower-level strict complementarity is a rather restrictive assumption; it should not be expected to hold in applications of interest. Observe that in Example 7.44, this condition holds at any feasible point, except for $\bar{x} = 0$. Recall, however, that $\bar{x} = 0$ is always a solution of this problem when $\sigma \geq 0$, and moreover, it is the unique solution when $\sigma > 0$.

If the lower-level strict complementarity condition does not hold, then the constraints of the MPVC (7.235) are inevitably degenerate. Specifically, it is demonstrated in [1] (and can be easily verified) that the following statements are valid:

- If $I_0 \neq \emptyset$, then $\bar{x}$ violates the LICQ.
- If $I_{00} \cup I_{0+} \neq \emptyset$, then $\bar{x}$ violates the MFCQ.

Observe further that MPVC can be reduced, in principle, to MPCC. Specifically, introducing a slack variable $z \in \mathbf{R}^m$, the MPVC (7.235) can be equivalently reformulated as the following MPCC:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & G(x) - z \leq 0, \quad H(x) \geq 0, \quad z \geq 0, \\ & H_i(x) z_i = 0,\ i = 1, \ldots, m. \end{array} \tag{7.238}$$

However, this reformulation has the following serious drawback (apart from the increased dimension, which is undesirable by itself). Given a (local) solution $\bar{x}$ of the MPVC (7.235) such that $I_0 \neq \emptyset$, the corresponding components of the slack variable are not uniquely defined (for any $z_{I_0} \geq G_{I_0}(\bar{x})$, the pair $(\bar{x}, z)$ is a (local) solution of the MPCC (7.238)). And since z does not enter the objective function of (7.238), the corresponding local solutions of this problem will never be strict. This means, in particular, that natural sufficient optimality conditions will never hold for the MPCC (7.238). Consequently, various theoretical results and/or computational methods relying on such conditions will not be applicable to the MPCC (7.238). For this reason, MPVC requires an independent analysis and special methods. Also, MPVC is in a sense “less degenerate” than MPCC: for example, as discussed above, the LICQ may hold at a feasible point of MPVC, while any feasible point of MPCC violates even the weaker MFCQ. Therefore, treating MPVC directly makes more sense than reducing it to MPCC.

That said, much of the theory and Newton-type methods for MPVC can be developed more-or-less along the same lines as those for MPCC in Sects. 7.3.1–7.3.4. In particular, one can employ the decomposition of MPVC into branch problems and piecewise analysis. For example, the relaxed and tightened problems associated with MPVC, and the related stationarity concepts, are defined naturally (and analogously to MPCC). We refer to [1, 121–123, 135, 145] for special optimality conditions and stationarity concepts for MPVC, and to [135–137] for the piecewise SQP method, active-set Newton methods and the lifting approach.

Appendix A
Miscellaneous Material

A.1 Linear Algebra and Linear Inequalities

The following statement can be regarded as a variant of the celebrated Farkas Lemma (e.g., [27, Theorem 2.201]).

Lemma A.1. *For any* $A \in \mathbf{R}^{l\times n}$ *and* $B \in \mathbf{R}^{m\times n}$*, for the cone*

$$C = \{x \in \mathbf{R}^n \mid Ax = 0,\, Bx \leq 0\}$$

it holds that

$$C^\circ = \{x \in \mathbf{R}^n \mid x = A^{\mathrm{T}}y + B^{\mathrm{T}}z,\, y \in \mathbf{R}^l,\, z \in \mathbf{R}^m_+\}.$$

Lemma A.1 can be derived as a corollary of the Motzkin Theorem of the Alternatives [186, p. 28], stated next.

Lemma A.2. *For any* $A \in \mathbf{R}^{l\times n}$*,* $B \in \mathbf{R}^{m\times n}$*,* $B_0 \in \mathbf{R}^{m_0\times n}$*, one and only one of the following statements holds: either there exists* $x \in \mathbf{R}^n$ *such that*

$$Ax = 0, \quad Bx \leq 0, \quad B_0x < 0,$$

or there exists $(y,\, z,\, z^0) \in \mathbf{R}^l \times \mathbf{R}^m \times \mathbf{R}^{m_0}$ *such that*

$$A^{\mathrm{T}}y + B^{\mathrm{T}}z + B_0^{\mathrm{T}}z^0 = 0, \quad z \geq 0, \quad z^0 \geq 0, \quad z^0 \neq 0.$$

The following simplified version of Lemma A.2, convenient in some applications, is known as the Gordan Theorem of the Alternatives.

Lemma A.3. *For any* $B \in \mathbf{R}^{m_0\times n}$*, one and only one of the following two alternatives is valid: either there exists* $x \in \mathbf{R}^n$ *such that*

$$B_0x < 0,$$

or there exists $z^0 \in \mathbf{R}^m$ *such that*

$$B_0^{\mathrm{T}}z^0 = 0, \quad z^0 \geq 0, \quad z^0 \neq 0.$$

A.F. Izmailov and M.V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-3-319-04247-3,

The following is Hoffman's lemma giving a (global) error bound for linear systems (e.g., [27, Theorem 2.200]).

Lemma A.4. *For any $A \in \mathbf{R}^{l\times n}$, $a \in \mathbf{R}^l$, and $B \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$, assume that the set*

$$S = \{x \in \mathbf{R}^n \mid Ax = a,\, Bx \leq b\}$$

is nonempty.

Then there exists $c > 0$ such that

$$\operatorname{dist}(x,\, S) \leq c(\|Ax - a\| + \|\max\{0,\, Bx - b\}\|) \quad \forall\, x \in \mathbf{R}^n.$$

The following is the finite-dimensional version of the classical Banach Open Mapping Theorem.

Lemma A.5. *For any $A \in \mathbf{R}^{l\times n}$, if $\operatorname{rank} A = l$, then there exists $c > 0$ such that for any $B \in \mathbf{R}^{l\times n}$ close enough to A, and any $y \in \mathbf{R}^l$, the equation*

$$Bx = y$$

has a solution $x(y)$ such that

$$\|x(y)\| \leq c\|y\|.$$

This result is complemented by the more exact characterization of invertibility of small perturbations of a nonsingular matrix; see, e.g., [103, Theorem 2.3.4].

Lemma A.6. *Let $A \in \mathbf{R}^{n\times n}$ be a nonsingular matrix.*

Then any matrix $B \in \mathbf{R}^{n\times n}$ satisfying the inequality $\|B - A\| < 1/\|A^{-1}\|$ is nonsingular, and

$$\|B^{-1} - A^{-1}\| \leq \frac{\|A^{-1}\|^2\|B - A\|}{1 - \|A^{-1}\|\|B - A\|}.$$

Lemma A.7 below is well known; it is sometimes called the Finsler Lemma [81], or the Debreu Lemma [54]. The similar in spirit Lemma A.8, on the other hand, is not standard, so we have to give its proof (from [150]). As the proofs of Lemmas A.7 and A.8 are somehow related, it makes sense to provide both.

Lemma A.7. *Let $H \in \mathbf{R}^{n\times n}$ be any symmetric matrix and $A \in \mathbf{R}^{l\times n}$ any matrix such that*

$$\langle H\xi,\, \xi\rangle > 0 \quad \forall\, \xi \in \ker A \setminus \{0\}. \tag{A.1}$$

Then the matrix $H + cA^{\mathrm{T}}A$ is positive definite for all $c \geq 0$ large enough.

Proof. We argue by contradiction. Suppose that there exist $\{c_k\} \subset \mathbf{R}$ and $\{\xi^k\} \subset \mathbf{R}^n$ such that $c_k \to +\infty$ as $k \to \infty$, and for all k it holds that $\|\xi^k\| = 1$ and

$$\langle (H + c_k A^{\mathrm{T}} A)\xi^k,\, \xi^k \rangle \leq 0. \tag{A.2}$$

Without loss of generality, we may assume that $\{\xi^k\} \to \xi$, with some $\xi \in \mathbf{R}^n \setminus \{0\}$. Dividing (A.2) by c_k and passing onto the limit as $k \to \infty$, we obtain that

$$0 \geq \langle A^{\mathrm{T}} A\xi,\, \xi \rangle = \|A\xi\|^2,$$

i.e., $\xi \in \ker A$.

On the other hand, since for each k it holds that

$$\langle A^{\mathrm{T}} A\xi^k,\, \xi^k \rangle = \|A\xi^k\|^2 \geq 0,$$

the inequality (A.2) implies that $\langle H\xi^k,\, \xi^k \rangle \leq 0$. Passing onto the limit as $k \to \infty$, we obtain that $\langle H\xi,\, \xi \rangle \leq 0$, in contradiction with (A.1). $\square$

Another result, as already commented somewhat similar in nature to the Debreu–Finsler Lemma, is stated next.

Lemma A.8. *Let $H \in \mathbf{R}^{n \times n}$ and $A \in \mathbf{R}^{l \times n}$ be such that*

$$H\xi \notin \operatorname{im} A^{\mathrm{T}} \quad \forall\, \xi \in \ker A \setminus \{0\}. \tag{A.3}$$

Then for any $C > 0$, any $\tilde{H} \in \mathbf{R}^{n \times n}$ close enough to H, and any $\tilde{A} \in \mathbf{R}^{l \times n}$ close enough to A, the matrix $\tilde{H} + c(A + \Omega)^{\mathrm{T}} \tilde{A}$ is nonsingular for all $c \in \mathbf{R}$ such that $|c|$ is large enough, and for all $\Omega \in \mathbf{R}^{l \times n}$ satisfying $\|\Omega\| \leq C/|c|$.

Proof. Suppose the contrary, i.e., that there exist sequences $\{H_k\} \subset \mathbf{R}^{n \times n}$, $\{A_k\} \subset \mathbf{R}^{l \times n}$, $\{\Omega_k\} \subset \mathbf{R}^{l \times n}$, $\{c_k\} \subset \mathbf{R}$ and $\{\xi^k\} \subset \mathbf{R}^n \setminus \{0\}$, such that $\{H_k\} \to H$, $\{A_k\} \to A$, $|c_k| \to \infty$ as $k \to \infty$, and for all k it holds that $\|\Omega_k\| \leq C/|c_k|$ and

$$H_k \xi^k + c_k (A + \Omega_k)^{\mathrm{T}} A_k \xi^k = 0. \tag{A.4}$$

We can assume, without loss of generality, that $\|\xi^k\| = 1$ for all k, and $\{\xi^k\} \to \xi$, with some $\xi \in \mathbf{R}^n \setminus \{0\}$. Then since the right-hand side in

$$A^{\mathrm{T}} A_k \xi^k = -\frac{1}{c_k} H_k \xi^k - \Omega_k^{\mathrm{T}} A_k \xi^k,$$

tends to zero as $k \to \infty$, it must hold that $A^{\mathrm{T}} A\xi = 0$.

It is thus established that $A\xi \in \ker A^{\mathrm{T}}$, and since $A\xi \in \operatorname{im} A = (\ker A^{\mathrm{T}})^{\perp}$, this shows that $A\xi = 0$. Thus, $\xi \in \ker A \setminus \{0\}$.

On the other hand, (A.4) implies that the inclusion

$$H_k \xi^k + c_k \Omega_k^{\mathrm{T}} A_k \xi^k = -c_k A^{\mathrm{T}} A_k \xi^k \in \operatorname{im} A^{\mathrm{T}}$$

holds for all k, where the second term in the left-hand side tends to zero as $k \to \infty$ because $\{c_k \Omega_k\}$ is bounded and $\{A_k \xi^k\} \to A\xi = 0$. Hence, $H\xi \in \operatorname{im} A^{\mathrm{T}}$ by the closedness of $\operatorname{im} A^{\mathrm{T}}$. This gives a contradiction with (A.3). $\square$

We complete this section by the following fact concerned with the existence of the inverse of a block matrix; see [243, Proposition 3.9].

Lemma A.9. *If $A \in \mathbf{R}^{n\times n}$ is a nonsingular matrix, $B \in \mathbf{R}^{n\times m}$, $C \in \mathbf{R}^{m\times n}$, $D \in \mathbf{R}^{m\times m}$, then for the matrix*

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

it holds that

$$\det M = \det A \det(D - CA^{-1}B).$$

Under the assumptions of Lemma A.9, the matrix $D - CA^{-1}B$ is referred to as the *Schur complement* of A in M.

A.2 Analysis

Our use of the big-O and little-o notation employs the following conventions. For a mapping $F : \mathbf{R}^n \to \mathbf{R}^m$ and a function $\varphi : \mathbf{R}^n \to \mathbf{R}_+$, and for a given $\bar{x} \in \mathbf{R}^n$, we write $F(x) = O(\varphi(x))$ as $x \to \bar{x}$ if there exists $c > 0$ such that $\|F(x)\| \leq c\varphi(x)$ for all $x \in \mathbf{R}^n$ close enough to $\bar{x}$. We write $F(x) = o(\varphi(x))$ as $x \to \bar{x}$ if for every $\varepsilon > 0$ no matter how small, it holds that $\|F(x)\| \leq \varepsilon\varphi(x)$ for all $x \in \mathbf{R}^n$ close enough to $\bar{x}$. For sequences $\{x^k\} \subset \mathbf{R}^n$ and $\{t_k\} \subset \mathbf{R}_+$, by $x^k = O(t_k)$ as $k \to \infty$ we mean that there exists $c > 0$ such that $\|x^k\| \leq ct_k$ for all k large enough. Accordingly, $x^k = o(t_k)$ as $k \to \infty$ if for every $\varepsilon > 0$ no matter how small, it holds that $\|x^k\| \leq \varepsilon t_k$ for all k large enough. For a sequence $\{\tau_k\} \subset \mathbf{R}$, we write $\tau^k \leq o(t_k)$ as $k \to \infty$ if for any $\varepsilon > 0$ no matter how small it holds that $\tau_k \leq \varepsilon t_k$ for all k large enough.

Concerning convergence rate estimates, the terminology is as follows. Let a sequence $\{x^k\} \subset \mathbf{R}^n$ be convergent to some $\bar{x} \in \mathbf{R}^n$. If there exist $q \in (0, 1)$ and $c > 0$ such that

$$\|x^k - \bar{x}\| \leq cq^k$$

for all k large enough (or, in other words, $\|x^k - \bar{x}\| = O(q^k)$ as $k \to \infty$), then we say that $\{x^k\}$ has *geometric convergence rate*. If there exists $q \in (0, 1)$ such that

$$\|x^{k+1} - \bar{x}\| \leq q\|x^k - \bar{x}\| \tag{A.5}$$

for all k large enough, then we say that $\{x^k\}$ has *linear convergence rate*. Linear rate implies geometric rate, but the converse is not true. If for every $q \in (0, 1)$, no matter how small, the inequality (A.5) holds for all k large enough, then we say that $\{x^k\}$ has *superlinear convergence rate*. To put it in other words, superlinear convergence means that

$$\|x^{k+1} - \bar{x}\| = o(\|x^k - \bar{x}\|)$$

as $k \to \infty$. A particular case of the superlinear rate is *quadratic convergence rate*, meaning that there exists $c > 0$ such that

$$\|x^{k+1} - \bar{x}\| \leq c\|x^k - \bar{x}\|^2$$

for all k large enough or, in other words,

$$\|x^{k+1} - \bar{x}\| = O(\|x^k - \bar{x}\|^2)$$

as $k \to \infty$. Unlike superlinear or quadratic rate, linear convergence rate depends on the norm: linear convergence rate in some norm in $\mathbf{R}^n$ does not necessarily imply linear convergence rate in a different norm.

We next state some facts and notions of differential calculus for mappings (generally vector-valued and with vector variable). It is assumed that the reader is familiar with differential calculus for scalar-valued functions in a scalar variable.

The mapping $F : \mathbf{R}^n \to \mathbf{R}^m$ is said to be *differentiable at* $x \in \mathbf{R}^n$ if there exists a matrix $J \in \mathbf{R}^{m \times n}$ such that for $\xi \in \mathbf{R}^n$ is holds that

$$F(x + \xi) = F(x) + J\xi + o(\|\xi\|)$$

as $\xi \to 0$. The matrix J with this property is necessarily unique; it coincides with the *Jacobian* $F'(x)$ (the matrix of first partial derivatives of the components of F at x with respect to all the variables), and it is also called the *first derivative of* F *at* x. The rows of the Jacobian are the *gradients* $F_1'(x), \ldots, F_m'(x)$ (vectors of first partial derivatives with respect to all variables) of the components of F at x.

The mapping $F : \mathbf{R}^n \to \mathbf{R}^m$ is (*continuously*) *differentiable on a set* $S \subset \mathbf{R}^n$ if it is differentiable at every point of some open set $O \subset \mathbf{R}^n$ such that $S \subset O$ (and the mapping $F'(\cdot)$ defined on O is continuous at every point of S).

The mapping $F : \mathbf{R}^n \to \mathbf{R}^m$ is said to be *twice differentiable at* $x \in \mathbf{R}^n$ if it is differentiable in a neighborhood of x, and the mapping $F'(\cdot)$ defined on this neighborhood is differentiable at x. The derivative $(F')'(x)$ of $F'(\cdot)$ at x can be regarded as a linear operator from $\mathbf{R}^n$ to $\mathbf{R}^{m \times n}$, or alternatively, as a bilinear mapping $F''(x) : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^m$ defined by

$$F''(x)[\xi^1, \xi^2] = ((F')'(x)\xi^1)\xi^2, \quad \xi^1, \xi^2 \in \mathbf{R}^n.$$

This bilinear mapping is necessarily symmetric, that is,

$$F''(x)[\xi^1, \xi^2] = F''(x)[\xi^2, \xi^1] \quad \forall\, \xi^1, \xi^2 \in \mathbf{R}^n.$$

The mapping in question is called the *second derivative of F at x*, and it is comprised by the *Hessians* $F_1''(x), \ldots, F_m''(x)$ (the matrices of second partial derivatives) of the components of F at x:

$$F''(x)[\xi^1, \xi^2] = (\langle F_1''(x)\xi^1, \xi^2\rangle, \ldots, \langle F_m''(x)\xi^1, \xi^2\rangle) \quad \forall\, \xi^1, \xi^2 \in \mathbf{R}^n.$$

Note that the symmetry of the bilinear mapping $F''(x)$ is equivalent to the symmetry of the Hessians of the components of F.

If F is twice differentiable at $\bar{x}$, then for $\xi \in \mathbf{R}^n$ it holds that

$$F(x+\xi) = F(x) + F'(x)\xi + \frac{1}{2}F'(x)[\xi, \xi] + o(\|\xi\|^2)$$

as $\xi \to 0$. This fact can be regarded as a particular case of the Taylor formula.

The mapping F is *twice* (*continuously*) *differentiable on a set* $S \subset \mathbf{R}^n$ if it is twice differentiable at every point of some open set $O \subset \mathbf{R}^n$ such that $S \subset O$ (and the mapping $F''(\cdot)$ defined on the set O is continuous at every point of S).

Furthermore, the mapping $F : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}^m$ is said to be *differentiable at* $(x, y) \in \mathbf{R}^n \times \mathbf{R}^l$ *with respect to* x if the mapping $F(\cdot, y)$ is differentiable at x. The derivative of the latter mapping at x is called the *partial derivative of F with respect to x at* (x, y), and it is denoted by $\frac{\partial F}{\partial x}(x, y)$.

Similarly, the mapping $F : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}^m$ is said to be *twice differentiable at* $(x, y) \in \mathbf{R}^n \times \mathbf{R}^l$ *with respect to* x if the mapping $F(\cdot, y)$ is twice differentiable at x. The second derivative of the latter mapping at x is called the *second partial derivative of F with respect to x at* (x, y), and it is denoted by $\frac{\partial^2 F}{\partial x^2}(x, y)$.

In this book, any of the itemized assertions in the next statement is referred to as a mean-value theorem. The first part of item (a) is a rather subtle and not widely known result; it was established in [198]. The other statements are fairly standard.

Theorem A.10. *For any $F : \mathbf{R}^n \to \mathbf{R}^m$ and any $x^1, x^2 \in \mathbf{R}^n$, the following assertions are valid:*

(a) If F is continuous on $[x^1, x^2] = \{tx^1 + (1-t)x^2 \mid t \in [0, 1]\}$ and differentiable on $(x^1, x^2) = \{tx^1 + (1-t)x^2 \mid t \in (0, 1)\}$, then there exist $t_i \in (0, 1)$ and $\theta_i \geq 0$, $i = 1, \ldots, m$, such that $\sum_{i=1}^m \theta_i = 1$ and

$$F(x^1) - F(x^2) = \sum_{i=1}^{m} \theta_i F'(t_i x^1 + (1-t_i)x^2)(x^1 - x^2),$$

and in particular,

$$\|F(x^1) - F(x^2)\| \leq \sup_{t \in (0, 1)} \|F'(tx^1 + (1-t)x^2)\| \|x^1 - x^2\|.$$

(b) If F is continuously differentiable on the line segment $[x^1,\, x^2]$, then

$$F(x^1) - F(x^2) = \int_0^1 F'(tx^1 + (1-t)x^2)(x^1 - x^2)\, dt.$$

The next fact is an immediate corollary of assertion (b) of Theorem A.10.

Lemma A.11. *For any $F : \mathbf{R}^n \to \mathbf{R}^m$ and any $x^1,\, x^2 \in \mathbf{R}^n$, if F is differentiable on the line segment $[x^1,\, x^2]$, with its derivative being Lipschitz-continuous on this segment with a constant $L > 0$, then*

$$\|F(x^1) - F(x^2) - F'(x^2)(x^1 - x^2)\| \le \frac{L}{2}\|x^1 - x^2\|^2.$$

A.3 Convexity and Monotonicity

For a detailed exposition of finite-dimensional convex analysis, we refer to [235]. In this section we only recall some basic definitions and facts used in this book. For details on (maximal) monotone mappings and related issues, we refer to [17, 30] and [239, Chap. 12].

For a finite number of points $x^1,\, \ldots,\, x^m \in \mathbf{R}^n$, their *convex combinations* are points of the form $\sum_{i=1}^m t_i x^i$ with some $t_i \ge 0$, $i = 1,\, \ldots,\, m$, such that $\sum_{i=1}^m t_i = 1$. In particular, convex combinations of two points $x^1,\, x^2 \in \mathbf{R}^n$ are points of the form $tx^1 + (1-t)x^2$, $t \in [0,\, 1]$, and they form a line segment connecting x^1 and x^2.

A set $S \subset \mathbf{R}^n$ is said to be *convex* if for each pair of points $x^1,\, x^2 \in S$ all convex combinations of these points belong to S (equivalently, for any points $x^1,\, \ldots,\, x^m \in S$, where $m \ge 2$, all convex combinations of these points belong to S).

The *convex hull* of a set $S \subset \mathbf{R}^n$, denoted by $\operatorname{conv} S$, is the smallest convex set in $\mathbf{R}^n$ that contains S (equivalently, the set of all convex combinations of points in S).

By a (Euclidean) *projection* of a point $x \in \mathbf{R}^n$ onto a given set $S \subset \mathbf{R}^n$ we mean a point closest to x among all the points in S, i.e., any global solution of the optimization problem

$$\begin{array}{ll} \text{minimize} & \|y - x\| \\ \text{subject to} & y \in S. \end{array} \tag{A.6}$$

As the objective function in (A.6) is coercive, projection of any point onto any nonempty closed set in $\mathbf{R}^n$ exists. If, in addition, the set is convex, then the following holds.

Lemma A.12. *Let $S \subset \mathbf{R}^n$ be any nonempty closed convex set.*

Then the projection operator onto S, $\pi_S : \mathbf{R}^n \to S$, is well defined and single valued: for any point $x \in \mathbf{R}^n$ its projection $\pi_S(x)$ onto S exists and is unique. Moreover, $\bar{x} = \pi_S(x)$ if and only if

$$\bar{x} \in S, \quad \langle x - \bar{x},\, y - \bar{x}\rangle \leq 0 \quad \forall\, y \in S.$$

In addition, the projection operator is nonexpansive:

$$\|\pi_S(x^1) - \pi_S(x^2)\| \leq \|x^1 - x^2\| \quad \forall\, x^1,\, x^2 \in \mathbf{R}^n.$$

A set $C \subset \mathbf{R}^n$ is called a *cone* if for each $x \in C$ it contains all points of the form tx, $t \geq 0$. The *polar cone* to C is defined by

$$C^\circ = \{\xi \in \mathbf{R}^n \mid \langle \xi,\, x\rangle \leq 0\; \forall\, x \in C\}.$$

Lemma A.13. *For any nonempty closed convex cone $C \subset \mathbf{R}^n$ it holds that*

$$x = \pi_C(x) + \pi_{C^\circ}(x) \quad \forall\, x \in \mathbf{R}^n,$$

and in particular,

$$C^\circ = \{x \in \mathbf{R}^n \mid \pi_C(x) = 0\},$$

$$\pi_C(x - \pi_C(x)) = 0 \quad \forall\, x \in \mathbf{R}^n.$$

An important property concerns separation of (convex) sets by hyperplanes. The following separation theorem can be found in [235, Corollary 11.4.2].

Theorem A.14. *Let $S_1,\, S_2 \subset \mathbf{R}^n$ be nonempty closed convex sets, with at least one of them being also bounded (hence, compact).*

Then $S_1 \cap S_2 = \emptyset$ if and only if there exist $\xi \in \mathbf{R}^n \setminus \{0\}$ and $t \in \mathbf{R}$ such that

$$\langle \xi,\, x^1\rangle < t < \langle \xi,\, x^2\rangle \quad \forall\, x^1 \in S_1,\, x^2 \in S_2.$$

Given a convex set $S \subset \mathbf{R}^n$, a function $f : S \to \mathbf{R}$ is said to be *convex* (on the set S) if

$$f(tx^1 + (1-t)x^2) \leq tf(x^1) + (1-t)f(x^2) \quad \forall\, x^1,\, x^2 \in S,\; \forall\, t \in [0,\, 1].$$

Equivalently, f is convex if its *epigraph* $\{(x,\, t) \in S \times \mathbf{R} \mid f(x) \leq t\}$ is a convex set. It is immediate that a linear combination of convex functions with nonnegative coefficients is a convex function, and the maximum over a finite family of convex functions is a convex function.

Furthermore, f is said to be *strongly convex* (on S) if there exists $\gamma > 0$ such that

$$f(tx^1 + (1-t)x^2) \leq tf(x^1) + (1-t)f(x^2) - \gamma t(1-t)\|x^1 - x^2\|^2$$
$$\forall\, x^1,\, x^2 \in S,\; \forall\, t \in [0,\, 1].$$

The sum of a convex function and a strongly convex function is evidently strongly convex. The following are characterizations of convexity for smooth functions.

Proposition A.15. *Let $O \subset \mathbf{R}^n$ be a nonempty open convex set, and let $f : O \to \mathbf{R}$ be differentiable in O.*

Then the following items are equivalent:

(a) The function f is convex on O.
(b) $f(x^1) \geq f(x^2) + \langle f'(x^2),\, x^1 - x^2\rangle$ for all $x^1, x^2 \in O$.
(c) $\langle f'(x^1) - f'(x^2),\, x^1 - x^2\rangle \geq 0$ for all $x^1, x^2 \in O$.

If f is twice differentiable in O, then the properties above are further equivalent to

(d) The Hessian $f''(x)$ is positive semidefinite for all $x \in O$.

It is clear that a quadratic function is convex if and only if its (constant) Hessian is a positive semidefinite matrix (see item (d) in Proposition A.15). Moreover, a quadratic function is strongly convex if and only if its Hessian is positive definite.

Given a convex function $f : \mathbf{R}^n \to \mathbf{R}$, an element $a \in \mathbf{R}^n$ is called a *subgradient* of f at a point $x \in \mathbf{R}^n$ if

$$f(y) \geq f(x) + \langle a,\, y - x\rangle \quad \forall\, y \in \mathbf{R}^n.$$

The set of all the elements $a \in \mathbf{R}^n$ with this property is called the *subdifferential* of f at x, denoted by $\partial f(x)$.

Proposition A.16. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be convex on $\mathbf{R}^n$.*

Then for each $x \in \mathbf{R}^n$ the subdifferential $\partial f(x)$ is a nonempty compact convex set. Moreover, f is continuous and directionally differentiable at every $x \in \mathbf{R}^n$ in every direction $\xi \in \mathbf{R}^n$, and it holds that

$$f'(x;\, \xi) = \max_{y \in \partial f(x)} \langle y,\, \xi\rangle.$$

For some further calculus rules for subdifferentials see Sect. 1.4.1, where they are presented in a more general (not necessarily convex) setting. Section 1.4.1 provides all the necessary material for the convex calculus in this book.

We complete this section by some definitions and facts concerned with the notion of monotonicity for (multi)functions. For a (generally) set-valued mapping Ψ from $\mathbf{R}^n$ to the subsets of $\mathbf{R}^n$, define its domain

$$\operatorname{dom} \Psi = \{x \in \mathbf{R}^n \mid \Psi(x) \neq \emptyset\}.$$

Then Ψ is said to be *monotone* if

$$\langle y^1 - y^2,\, x^1 - x^2\rangle \geq 0 \quad \forall\, y^1 \in \Psi(x^1),\ \forall\, y^2 \in \Psi(x^2),\ \forall\, x^1,\, x^2 \in \operatorname{dom} \Psi,$$

and *maximal monotone* if, in addition, its *graph* $\{(x, y) \in \mathbf{R}^n \times \mathbf{R}^n \mid y \in \Psi(x)\}$ is not contained in the graph of any other monotone set-valued mapping.

Some examples of maximal monotone mappings are: a continuous monotone function $F : \mathbf{R} \to \mathbf{R}$, the subdifferential multifunction $\partial f(\cdot)$ of a convex function $f : \mathbf{R}^n \to \mathbf{R}$, and the normal cone multifunction $N_S(\cdot)$ for a closed convex set $S \subset \mathbf{R}^n$.

The sum of two monotone mappings is monotone, and the sum of two maximal monotone mappings is maximal monotone if the domain of one intersects the interior of the domain of the other.

Furthermore, Ψ is said to be *strongly monotone* if there exists $\gamma > 0$ such that $\Psi - \gamma I$ is monotone, which is equivalent to the property

$$\langle y^1 - y^2,\, x^1 - x^2 \rangle \geq \gamma \|x^1 - x^2\|^2 \quad \forall\, y^1 \in \Psi(x^1),\, \forall\, y^2 \in \Psi(x^2),\ \forall\, x^1,\, x^2 \in \operatorname{dom} \Psi.$$

In particular, the identity mapping I is strongly monotone, and the sum of a monotone mapping and a strongly monotone mapping is strongly monotone.

The following characterization of monotonicity for smooth mappings can be found, e.g., in [239, Proposition 12.3].

Proposition A.17. *Let $F : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable on $\mathbf{R}^n$.*

Then F is monotone if and only if $F'(x)$ is positive semidefinite for all $x \in \mathbf{R}^n$.

We refer to [17, 30] and [239, Chap. 12] for other details on (maximal) monotone mappings and related issues.

References

1. W. Achtziger, C. Kanzow, Mathematical programs with vanishing constraints: optimality conditions and constraint qualifications. Math. Program. **114**, 69–99 (2007)
2. W. Achtziger, T. Hoheisel, C. Kanzow, A smoothing-regularization approach to mathematical programs with vanishing constraints. Comput. Optim. Appl. **55**, 733–767 (2013)
3. L.T.H. An, An efficient algorithm for globally minimizing a quadratic function under convex quadratic constraints. Math. Program. **87**, 401–426 (2000)
4. R. Andreani, E.G. Birgin, J.M. Martínez, M.L. Schuverdt, On Augmented Lagrangian methods with general lower-level constraints. SIAM J. Optim. **18**, 1286–1309 (2007)
5. R. Andreani, E.G. Birgin, J.M. Martínez, M.L. Schuverdt, Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. Math. Program. **111**, 5–32 (2008)
6. M. Anitescu, Degenerate nonlinear programming with a quadratic growth condition. SIAM J. Optim. **10**, 1116–1135 (2000)
7. M. Anitescu, *Nonlinear Programs with Unbounded Lagrange Multiplier Sets* (Mathematics and Computer Science Division, Argonne National Laboratories, Argonne, 2000). Preprint ANL/MCS-P796-0200
8. M. Anitescu, *On solving mathematical programs with complementarity constraints as nonlinear programs* (Mathematics and Computer Science Division, Argonne National Laboratories, Argonne, 2000). Preprint ANL/MCS-P864-1200
9. M. Anitescu, A superlinearly convergent sequential quadratically constrained quadratic programming algorithm for degenerate nonlinear programming. SIAM J. Optim. **12**, 949–978 (2002)
10. M. Anitescu, On the rate of convergence of sequential quadratic programming with nondifferentiable exact penalty function in the presence of constrain degeneracy. Math. Program. **92**, 359–386 (2002)

A.F. Izmailov and M.V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-3-319-04247-3,

11. M. Anitescu, On using the elastic mode in nonlinear programming approaches to mathematical programs with complementarity constraints. SIAM J. Optim. **15**, 1203–1236 (2005)
12. M. Anitescu, P. Tseng, S.J. Wright, Elastic-mode algorithms for mathematical programs with equilibrium constraints: global convergence and stationarity properties. Math. Program. **110**, 337–371 (2007)
13. A.V. Arutyunov, *Optimality Conditions: Abnormal and Degenerate Problems* (Kluwer Academic, Dordrecht, 2000)
14. A.V. Arutyunov, A.F. Izmailov, Sensitivity analysis for cone-constrained optimization problems under the relaxed constraint qualifications. Math. Oper. Res. **30**, 333–353 (2005)
15. C. Audet, P. Hansen, B. Jaumard, G. Savard, A branch and cut algorithm for nonconvex quadratically constrained quadratic programming. Math. Program. **87**, 131–152 (2000)
16. A.B. Bakushinskii, A regularization algorithm based on the Newton-Kantorovich method for solving variational inequalities. USSR Comput. Math. Math. Phys. **16**, 16–23 (1976)
17. H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics (Springer, New York, 2011)
18. D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods* (Academic Press, New York, 1982)
19. D.P. Bertsekas, *Nonlinear Programming*, 2nd edn. (Athena, Belmont, 1999)
20. S.C. Billups, Algorithms for complementarity problems and generalized equations. Ph.D. thesis. Technical Report 95-14. Computer Sciences Department, University of Wisconsin, Madison, 1995
21. E.G. Birgin, J.M. Martínez, Local convergence of an inexact-restoration method and numerical experiments. J. Optim. Theory Appl. **127**, 229–247 (2005)
22. E.G. Birgin, J.M. Martínez, Improving ultimate convergence of an augmented Lagrangian method. Optim. Method. Softw. **23**, 177–195 (2008)
23. P. Boggs, J. Tolle, Sequential quadratic programming. Acta Numer. **4**, 1–51 (1995)
24. B.T. Boggs, J.W. Tolle, A.J. Kearsley, A truncated SQP algorithm for large scale nonlinear programming problems, in *Advances in Optimization and Numerical Analysis*, ed. by S. Gomez, J.-P. Hennart. Math. Appl. vol 275 (Kluwer Academic, Dordrecht, 1994), pp. 69–77
25. J.F. Bonnans, Asymptotic admissibility of the unit stepsize in exact penalty methods. SIAM J. Control Optim. **27**, 631–641 (1989)
26. J.F. Bonnans, Local analysis of Newton-type methods for variational inequalities and nonlinear programming. Appl. Math. Optim. **29**, 161–186 (1994)
27. J.F. Bonnans, A. Shapiro, *Perturbation Analysis of Optimization Problems* (Springer, New York, 2000)

28. J.F. Bonnans, A. Sulem, Pseudopower expansion of solutions of generalized equations and constrained optimization. Math. Program. **70**, 123–148 (1995)
29. J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, C. Sagastizábal, *Numerical Optimization: Theoretical and Practical Aspects*, 2nd edn. (Springer, Berlin, 2006)
30. R.S. Burachik, A.N. Iusem, *Set-Valued Mappings and Enlargements of Monotone Operators*. (Springer, New York, 2008)
31. R.S. Burachik, A.N. Iusem, B.F. Svaiter, Enlargement of monotone operators with applications to variational inequalities. Set-Valued Anal. **5**, 159–180 (1997)
32. R.S. Burachik, C.A. Sagastizábal, B.F. Svaiter, ε-Enlargements of maximal monotone operators: theory and applications, in *Reformulation – Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, ed. by M. Fukushima, L. Qi (Kluwer Academic, 1999), pp. 25–44
33. R.H. Byrd, R.B. Schnabel, G.A. Shultz, A trust region algorithm for nonlinearly constrained optimization. SIAM J. Numer. Anal. **24**, 1152–1170 (1987)
34. R.H. Byrd, N.I.M. Gould, J. Nocedal, R.A. Waltz, An algorithm for nonlinear optimization using linear programming and equality constrained subproblems. Math. Program. **100**, 27–48 (2004)
35. R. Byrd, M. Marazzi, J. Nocedal, On the convergence of Newton iterations to non-stationary points. Math. Program. **99**, 127–148 (2004)
36. R.H. Byrd, N.I.M. Gould, J. Nocedal, R.A. Waltz, On the convergence of successive linear-quadratic programming algorithms. SIAM J. Optim. **16**, 471–489 (2006)
37. R.H. Byrd, F.E. Curtis, J. Nocedal, An inexact SQP method for equality constrained optimization. SIAM J. Optim. **19**, 351–369 (2008)
38. R.H. Byrd, J. Nocedal, R.A. Waltz, Steering exact penalty methods for nonlinear programming. Optim. Method. Softw. **23**, 197–213 (2008)
39. R.H. Byrd, F.E. Curtis, J. Nocedal, An inexact Newton method for nonconvex equality constrained optimization. Math. Program. **122**, 273–299 (2010)
40. Y. Chen, M. Florian, The nonlinear bilevel programming problem: formulations, regularity and optimality conditions. Optimization **32**, 193–209 (1995)
41. C. Chin, R. Fletcher, On the global convergence of an SLP-filter algorithm that takes EQP steps. Math. Program. **96**, 161–177 (2003)
42. M.R. Celis, J.E. Dennis, R.A. Tapia, A trust region algorithm for nonlinear equality constrained optimization, in *Numerical Optimization*, ed. by P.T. Boggs, R.H. Byrd, R.B. Schnabel (SIAM, Philadelphia, 1985), pp. 71–82
43. F.H. Clarke, On the inverse function theorem. Pacisic J. Math. **64**, 97–102 (1976)

44. F.H. Clarke, *Optimization and Nonsmooth Analysis* (Wiley, New York, 1983)
45. A.R. Conn, N.I.M. Gould, Ph.L. Toint, *Trust-Region Methods* (SIAM, Philadelphia, 2000)
46. R.W. Cottle, J.-S. Pang, R.E. Stone, *The Linear Complementarity Problem* (SIAM, Philadelphia, 2009)
47. F.E. Curtis, J. Nocedal, A. Wächter, A matrix-free algorithm for equality constrained optimization problems with rank-deficient Jacobians. SIAM J. Optim. **20**, 1224–1249 (2009)
48. H. Dan, N. Yamashita, M. Fukushima, A superlinearly convergent algorithm for the monotone nonlinear complementarity problem without uniqueness and nondegeneracy conditions. Math. Oper. Res. **27**, 743–754 (2002)
49. A.N. Daryina, A.F. Izmailov, M.V. Solodov, A class of active-set newton methods for mixed complementarity problems. SIAM J. Optim. **15**, 409–429 (2004)
50. A.N. Daryina, A.F. Izmailov, M.V. Solodov, Mixed complementarity problems: regularity, error bounds, and Newton-type methods. Comput. Mathem. Mathem. Phys. **44**, 45–61 (2004)
51. A.N. Daryina, A.F. Izmailov, M.V. Solodov, Numerical results for a globalized active-set Newton method for mixed complementarity problems. Comput. Appl. Math. **24**, 293–316 (2005)
52. T. De Luca, F. Facchinei, C. Kanzow, A semismooth equation approach to the solution of nonlinear complementarity problems. Math. Program. **75**, 407–439 (1996)
53. T. De Luca, F. Facchinei, C. Kanzow, A theoretical and numerical comparison of some semismooth algorithms for complementarity problems. Comput. Optim. Appl. **16**, 173–205 (2000)
54. G. Debreu, Definite and semidefinite quadratic forms. Econometrica **20**, 295–300 (1952)
55. R.S. Dembo, S.C. Eisenstat, T. Steihaug, Inexact Newton methods. SIAM J. Numer. Anal. **19**, 400–408 (1982)
56. S. Dempe, *Foundations of Bilevel Programming* (Kluwer, Dordrecht, 2002)
57. J.E. Dennis, J.J. Moré, A characterization of superlinear convergence and its applications to quasi-Newton methods. Math. Comp. **28**, 549–560 (1974)
58. J.E. Dennis, R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (SIAM, Philadelphia, 1996)
59. J.E. Dennis, M. El-Alem, M.C. Maciel, A global convergence theory for general trust-region-based algorithms for equality constrained optimization. SIAM J. Optim. **7**, 177–207 (1997)
60. S.P. Dirkse, M.C. Ferris, The PATH solver: a non-monotone stabilization scheme for mixed complementarity problems. Optim. Method. Softw. **5**, 123–156 (1995)

61. A.L. Dontchev, R.T. Rockafellar, Characterizations of strong regularity for variational inequalities over polyhedral convex sets. SIAM J. Optim. **6**, 1087–1105 (1996)
62. A.L. Dontchev, R.T. Rockafellar, *Implicit Functions and Solution Mappings* (Springer, New York, 2009)
63. A.L. Dontchev, R.T. Rockafellar, Newton's method for generalized equations: a sequential implicit function theorem. Math. Program. **123**, 139–159 (2010)
64. H.G. Eggleston, *Convexity* (Cambridge University Press, Cambridge, 1958)
65. L.C. Evans, R.F. Gariepy, *Measure Theory and Fine Properties of Functions* (CRC Press, Boca Raton, 1992)
66. M. Fabian, D. Preiss, On the Clarke's generalized Jacobian, in *Proceedings of the 14th Winter School on Abstract Analysis*, Circolo Matematico di Palermo, Palermo, 1987. Rendiconti del Circolo Matematico di Palermo, Ser. II, Number 14, pp. 305–307
67. F. Facchinei, Minimization of SC^1 functions and the Maratos effect. Oper. Res. Lett. **17**, 131–137 (1995)
68. F. Facchinei, J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems* (Springer, New York, 2003)
69. F. Facchinei, J. Soares, A new merit function for nonlinear complementarity problems and a related algorithm. SIAM J. Optim. **7**, 225–247 (1997)
70. F. Facchinei, A. Fischer, C. Kanzow, Regularity properties of a semismooth reformulation of variational inequalities. SIAM J. Optim. **8**, 850–869 (1998)
71. F. Facchinei, A. Fischer, C. Kanzow, On the accurate identifcation of active constraints. SIAM J. Optim. **9**, 14–32 (1999)
72. F. Facchinei, A. Fischer, C. Kanzow, J.-M. Peng, A simply constrained optimization reformulation of KKT systems arising from variational inequalities. Appl. Math. Optim. **40**, 19–37 (1999)
73. D. Fernández, A quasi-Newton strategy for the sSQP method for variational inequality and optimization problems. Math. Program. **137**, 199–223 (2013)
74. D. Fernández, M.V. Solodov, On local convergence of sequential quadratically-constrained quadratic-programming type methods, with an extension to variational problems. Comput. Optim. Appl. **39**, 143–160 (2008)
75. D. Fernández, M. Solodov, Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. Math. Program. **125**, 47–73 (2010)
76. D. Fernández, M. Solodov, Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition. SIAM J. Optim. **22**, 384–407 (2012)

77. D. Fernández, A.F. Izmailov, M.V. Solodov, Sharp primal superlinear convergence results for some Newtonian methods for constrained optimization. SIAM J. Optim. **20**, 3312–3334 (2010)
78. D. Fernández, E.A. Pilotta, G.A. Torres, An inexact restoration strategy for the globalization of the sSQP method. Comput. Optim. Appl. **54**, 595–617 (2013)
79. M.C. Ferris, F. Tin-Loi, On the solution of a minimum weight elastoplastic problem involving displacement and complementarity constraints. Comput. Method. Appl. Mech. Eng. **174**, 108–120 (1999)
80. M.C. Ferris, C. Kanzow, T.S. Munson, Feasible descent algorithms for mixed complementarity problems. Math. Program. **86**, 475–497 (1999)
81. P. Finsler, Über das vorkommen definiter und semidefiniter formen und scharen quadratischer formen. Comment. Math. Helv. **94**, 188–192 (1937)
82. A. Fischer, A special Newton-type optimization method. Optimization **24**, 296–284 (1992)
83. A. Fischer, Solution of monotone complementarity problems with locally Lipschitzian function. Math. Program. **76**, 513–532 (1997)
84. A. Fischer, Modified Wilson's method for nonlinear programs with nonunique multipliers. Math. Oper. Res. **24**, 699–727 (1999)
85. A. Fischer, Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Program. **94**, 91–124 (2002)
86. A. Fischer, A. Friedlander, A new line search inexact restoration approach for nonlinear programming. Comput. Optim. Appl. **46**, 333–346 (2010)
87. R. Fletcher, A model algorithm for composite non-differentiable optimization problems. Math. Program. **17**, 67–76 (1982)
88. R. Fletcher, Second order corrections for non-differentiable optimization, in *Numerical Analysis*, ed. by D. Griffiths (Springer, Berlin, 1982), pp. 85–114
89. R. Fletcher, *Practical Methods of Optimization*, 2nd edn. (Wiley, New York, 1987)
90. R. Fletcher, E.S. de la Maza, Nonlinear programming and nonsmooth optimization by successive linear programming. Math. Program. **43**, 235–256 (1989)
91. R. Fletcher, S. Leyffer, Nonlinear programming without a penalty function. Math. Program. **91**, 239–269 (2002)
92. R. Fletcher, S. Leyffer, Solving mathematical programs with equilibrium constraints as nonlinear programs. Optim. Method. Softw. **19**, 15–40 (2004)
93. R. Fletcher, N. Gould, S. Leyffer, P. Toint, A. Wächter, Global convergence of trust-region and SQP-filter algorithms for general nonlinear programming. SIAM J. Optim. **13**, 635–659 (2002)
94. R. Fletcher, S. Leyffer, P.L. Toint, On the global convergence of a filter-SQP algorithm. SIAM J. Optim. **13**, 44–59 (2002)

95. R. Fletcher, S. Leyffer, D. Ralph, S. Scholtes, Local convergence of SQP methods for mathematical programs with equilibrium constraints. SIAM J. Optim. **17**, 259–286 (2006)
96. M.P. Friedlander, M.A. Saunders, A globally convergent linearly constrained Lagrangian method for nonlinear optimization. SIAM J. Optim. **15**, 863–897 (2005)
97. M. Fukushima, Z.-Q. Luo, P. Tseng, A sequential quadratically constrained quadratic programming method for differentiable convex minimization. SIAM J. Optim. **13**, 1098–1119 (2003)
98. J. Gauvin, A necessary and sufficient regularity conditions to have bounded multipliers in nonconvex programming. Math. Program. **12**, 136–138 (1977)
99. P.E. Gill, D.P. Robinson, A globally convergent stabilized SQP Method. SIAM J. Optim. **23**, 1983–2010 (2013)
100. P.E. Gill, W. Murray, M.H. Wright, *Numerical Linear Algebra and Optimization*, vol 1 (Addison Wesley, Redwood City, 1991)
101. P.E. Gill, W. Murray, M.A. Saunders, M.H. Wright, Some theoretical properties of an augmented Lagrangian merit function, in *Advances in Optimization and Parallel Computing*, ed. by P.M. Pardalos (North–Holland, Amsterdam, 1992), pp. 101–128
102. P.E. Gill, W. Murray, M.A. Saunders, SNOPT: an SQP algorithm for large-scale constrained optimization. SIAM Rev. **47**, 99–131 (2005)
103. G.H. Golub, C.F. Van Loan, *Matrix Computations*, 3rd edn. (The Johns Hopkins University Press, Baltimore, 1996)
104. M. Golubitsky, D.G. Schaeffer, *Singularities and Groups in Bifurcation Theory*, vol 1. (Springer, New York, 1985)
105. C.C. Gonzaga, E.W. Karas, M. Vanti, A globally convergent filter method for nonlinear programming. SIAM J. Optim. **14**, 646–669 (2003)
106. N.I.M. Gould, Some reflections on the current state of active-set and interior-point methods for constrained optimization. Numerical Analysis Group Internal Report 2003-1. Computational Science and Engineering Department, Rutherford Appleton Laboratory, Oxfordshire, 2003
107. N.I.M. Gould, D.P. Robinson, A second derivative SQP method: global convergence and practical issues. SIAM J. Optim. **20**, 2023–2048 (2010)
108. N.I.M. Gould, D.P. Robinson, A second derivative SQP method: local convergence and practical issues. SIAM J. Optim. **20**, 2049–2079 (2010)
109. N.I.M. Gould, D. Orban, Ph.L. Toint, Numerical methods for large-scale nonlinear optimization. Acta Numer. **14**, 299–361 (2005)
110. L. Grippo, F. Lampariello, S. Lucidi, A nonmonotone line search technique for Newton's method. SIAM J. Numer. Anal. **23**, 707–716 (1986)
111. O. Güler, F. Gürtuna, O. Shevchenko, Duality in quasi-Newton methods and new variational characterizations of the DFP and BFGS updates. Optim. Method. Softw. **24**, 45–62 (2009)
112. W.W. Hager, Stabilized sequential quadratic programming. Comput. Optim. Appl. **12**, 253–273 (1999)

113. W.W. Hager, M.S. Gowda, Stability in the presence of degeneracy and error estimation. Math. Program. **85**, 181–192 (1999)
114. S.-P. Han, A globally convergent method for nonlinear programming. J. Optim. Theory Appl. **22**, 297–309 (1977)
115. J. Han, D. Sun, Superlinear convergence of approximate Newton methods for LC^1 optimization problems without strict complementarity, in *Recent Advances in Nonsmooth Optimization*, vol 58, ed. by D.-Z. Du, L. Qi, R.S. Womersley (World Scientific, Singapore, 1993), pp. 353–367
116. J. Herskovits, *A Two-Stage Feasible Direction Algorithm Including Variable Metric Techniques for Nonlinear Optimization Problems.* Rapport de Recherche 118 (INRIA, Rocqencourt, 1982)
117. J. Herskovits, A two-stage feasible directions algorithm for nonlinear constrained optimization. Math. Program. **36**, 19–38 (1986)
118. J. Herskovits, Feasible direction interior-point technique for nonlinear optimization. J. Optim. Theory Appl. **99**, 121–146 (1998)
119. J.-B. Hiriart-Urruty, Mean-value theorems in nonsmooth analysis. Numer. Func. Anal. Optim. **2**, 1–30 (1980)
120. J.-B. Hiriart-Urruty, J.-J. Strodiot, V.H. Nguyen, Generalized Hessian matrix and second-order optimality conditions for problems with $C^{1,1}$ data. Appl. Math. Optim. **11**, 43–56 (1984)
121. T. Hoheisel, C. Kanzow, First- and second-order optimality conditions for mathematical programs with vanishing constraints. Appl. Math. **52**, 495–514 (2007)
122. T. Hoheisel, C. Kanzow, Stationarity conditions for mathematical programs with vanishing constraints using weak constraint qualifications. J. Math. Anal. Appl. **337**, 292–310 (2008)
123. T. Hoheisel, C. Kanzow, On the Abadie and Guignard constraint qualifications for mathematical programs with vanishing constraints. Optimization **58**, 431–448 (2009)
124. Z.-H. Huang, D. Sun, G. Zhao, A smoothing Newton-type algorithm of stronger convergence for the quadratically constrained convex quadratic programming. Comput. Optim. Appl. **35**, 199–237 (2006)
125. C.-M. Ip, J. Kyparisis, Local convergence of quasi-Newton methods for B-differentiable equations. Math. Program. **56**, 71–89 (1992)
126. A.F. Izmailov, On the Lagrange methods for finding degenerate solutions of constrained optimization problems. Comput. Math. Math. Phys. **36**, 423–429 (1996)
127. A.F. Izmailov, Mathematical programs with complementarity constraints: regularity, optimality conditions, and sensitivity. Comput. Math. Math. Phys. **44**, 1145–1164 (2004)
128. A.F. Izmailov, On the analytical and numerical stability of critical Lagrange multipliers. Comput. Math. Math. Phys. **45**, 930–946 (2005)
129. A.F. Izmailov, Sensitivity of solutions to systems of optimality conditions under the violation of constraint qualifications. Comput. Math. Math. Phys. **47**, 533–554 (2007)

130. A.F. Izmailov, Solution sensitivity for Karush–Kuhn–Tucker systems with nonunique Lagrange multipliers. Optimization **59**, 747–775 (2010)
131. A.F. Izmailov, Strongly regular nonsmooth generalized equations. Math. Program. (2013). doi:10.1007/s10107-013-0717-1
132. A.F. Izmailov, M.M. Golishnikov, Newton-type methods for constrained optimization with nonregular constraints. Comput. Math. Math. Phys. **46**, 1299–1319 (2006)
133. A.F. Izmailov, A.S. Kurennoy, Partial Clarke generalized Jacobian and other generalized differentials, in *Theoretical and Applied Problems of Nonlinear Analysis*, ed. by V.A. Bereznyov (Computing Center RAS, Moscow, 2010, in Russian), pp. 77–90
134. A.F. Izmailov, A.S. Kurennoy, On regularity conditions for complementarity problems. Comput. Optim. Appl. (2013). doi:10.1007/s10589-013-9604-1
135. A.F. Izmailov, A.L. Pogosyan, Optimality conditions and Newton-type methods for mathematical programs with vanishing constraints. Comput. Math. Math. Phys. **49**, 1128–1140 (2009)
136. A.F. Izmailov, A.L. Pogosyan, A semismooth sequential quadratic programming method for lifted mathematical programs with vanishing constraints. Comput. Math. Math. Phys. **51**, 919–941 (2011)
137. A.F. Izmailov, A.L. Pogosyan, Active-set Newton methods for mathematical programs with vanishing constraints. Comput. Optim. Appl. **53**, 425–452 (2012)
138. A.F. Izmailov, M.V. Solodov, Superlinearly convergent algorithms for solving singular equations and smooth reformulations of complementarity problems. SIAM J. Optim. **13**, 386–405 (2002)
139. A.F. Izmailov, M.V. Solodov, The theory of 2-regularity for mappings with Lipschitzian derivatives and its applications to optimality conditions. Math. Oper. Res. **27**, 614–635 (2002)
140. A.F. Izmailov, M.V. Solodov, Karush–Kuhn–Tucker systems: regularity conditions, error bounds and a class of Newton-type methods. Math. Program. **95**, 631–650 (2003)
141. A.F. Izmailov, M.V. Solodov, Newton-type methods for optimization problems without constraint qualifications. SIAM J. Optim. **15**, 210–228 (2004)
142. A.F. Izmailov, M.V. Solodov, A note on solution sensitivity for Karush–Kuhn–Tucker systems. Math. Meth. Oper. Res. **61**, 347–363 (2005)
143. A.F. Izmailov, M.V. Solodov, An active-set Newton method for mathematical programs with complementarity constraints. SIAM J. Optim. **19**, 1003–1027 (2008)
144. A.F. Izmailov, M.V. Solodov, Examples of dual behaviour of Newton-type methods on optimization problems with degenerate constraints. Comput. Optim. Appl. **42**, 231–264 (2009)

145. A.F. Izmailov, M.V. Solodov, Mathematical programs with vanishing constraints: optimality conditions, sensitivity, and a relaxation method. J. Optim. Theory Appl. **142**, 501–532 (2009)
146. A.F. Izmailov, M.V. Solodov, On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions. Math. Program. **117**, 271–304 (2009)
147. A.F. Izmailov, M.V. Solodov, A truncated SQP method based on inexact interior-point solutions of subproblems. SIAM J. Optim. **20**, 2584–2613 (2010)
148. A.F. Izmailov, M.V. Solodov, Inexact Josephy–Newton framework for generalized equations and its applications to local analysis of Newtonian methods for constrained optimization. Comput. Optim. Appl. **46**, 347–368 (2010)
149. A.F. Izmailov, M.V. Solodov, On attraction of linearly constrained Lagrangian methods and of stabilized and quasi-Newton SQP methods to critical multipliers. Math. Program. **126**, 231–257 (2011)
150. A.F. Izmailov, M.V. Solodov, Stabilized SQP revisited. Math. Program. **133**, 93–120 (2012)
151. A.F. Izmailov, A.A. Tret'yakov, *Factor-Analysis of Nonlinear Mappings* (Nauka, Moscow, 1994) (in Russian)
152. A.F. Izmailov, A.A. Tret'yakov, *2-Regular Solutions of Nonlinear Problems* (Nauka, Moscow, 1999) (in Russian)
153. A.F. Izmailov, E.I. Uskov, Attraction of Newton method to critical Lagrange multipliers: fully quadratic case (2013). Available at http://www.optimization-online.org/DB_HTML/2013/02/3761.html
154. A.F. Izmailov, A.L. Pogosyan, M.V. Solodov, Semismooth SQP method for equality-constrained optimization problems with an application to the lifted reformulation of mathematical programs with complementarity constraints. Optim. Method. Softw. **26**, 847–872 (2011)
155. A.F. Izmailov, A.S. Kurennoy, M.V. Solodov, A note on upper Lipschitz stability, error bounds, and critical multipliers for Lipschitz-continuous KKT systems. Math. Program. **142**, 591–604 (2013)
156. A.F. Izmailov, A.L. Pogosyan, M.V. Solodov, Semismooth Newton method for the lifted reformulation of mathematical programs with complementarity constraints. Comput. Optim. Appl. **51**, 199–221 (2012)
157. A.F. Izmailov, M.V.Solodov, E.I. Uskov, Augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints. SIAM J. Optim. **22**, 1579–1606 (2012)
158. A.F. Izmailov, A.S. Kurennoy, M.V. Solodov, The Josephy–Newton method for semismooth generalized equations and semismooth SQP for optimization. Set-Valued Variational Anal. **21**, 17–45 (2013)

159. A.F. Izmailov, A.S. Kurennoy, M.V. Solodov, Composite-step constrained optimization methods interpreted via the perturbed sequential quadratic programming framework (2013). IMPA preprint A2405
160. H. Jäger, E.W. Sachs, Global convergence of inexact reduced SQP methods. Optim. Method. Softw. **7**, 83–110 (1997)
161. N.H. Josephy, Newton's method for generalized equations. Technical Summary Report No. 1965. Mathematics Research Center, University of Wisconsin, Madison, 1979
162. N.H. Josephy, Quasi-Newton methods for generalized equations. Technical Summary Report No. 1966. Mathematics Research Center, University of Wisconsin, Madison, 1979
163. C. Kanzow, Strictly feasible equation-based methods for mixed complementarity problems. Numer. Math. **89**, 135–160 (2001)
164. C. Kanzow, M. Fukushima, Solving box constrained variational inequalities by using the natural residual with D-gap function globalization. Oper. Res. Lett. **23**, 45–51 (1998)
165. E. Karas, A. Ribeiro, C. Sagastizábal, M. Solodov, A bundle-filter method for nonsmooth convex constrained optimization. Math. Program. **116**, 297–320 (2009)
166. W. Karush, Minima of functions of several variables with inequalities as side constraints. Technical Report (master's thesis), Department of Mathematics, University of Chicago, 1939
167. T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1984)
168. D. Klatte, Upper Lipschitz behaviour of solutions to perturbed $C^{1,1}$ programs. Math. Program. **88**, 285–311 (2000)
169. D. Klatte, B. Kummer, *Nonsmooth Equations in Optimization: Regularity, Calculus, Methods and Applications* (Kluwer Academic, Dordrecht, 2002)
170. D. Klatte, K. Tammer, On the second order sufficient conditions to perturbed $C^{1,1}$ optimization problems. Optimization **19**, 169–180 (1988)
171. M. Kojima, Strongly stable stationary solutions in nonlinear programs, in *Analysis and Computation of Fixed Points*, ed. by S.M. Robinson (Academic, NewYork, 1980), pp. 93–138
172. M. Kojima, S. Shindo, Extensions of Newton and quasi-Newton methods to systems of PC^1 equations. J. Oper. Res. Soc. Japan **29**, 352–374 (1986)
173. S. Kruk, H. Wolkowicz, Sequential, quadratically constrained, quadratic programming for general nonlinear programming, in *Handbook of Semidefinite Programming*, ed. by H. Wolkowicz, R. Saigal, L. Vandenberghe (Kluwer Academic, Dordrecht, 2000), pp. 563–575
174. H.W. Kuhn, A.W. Tucker, Non-linear programming. in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, ed. by J. Neyman (University of California Press, Berkeley, 1951), pp. 481–493

175. B. Kummer, Newton's method for nondifferentiable functions, in *Advances in Mathematical Optimization*, vol 45, ed. by J. Guddat, B. Bank, H. Hollatz, P. Kall, D. Klatte, B. Kummer, K. Lommalzsch, L. Tammer, M. Vlach, K. Zimmerman (Akademie-Verlag, Berlin, 1988), pp. 114–125
176. B. Kummer, Newton's method based on generalized derivatives for nonsmooth functions, in *Advances in Optimization*, ed. by W. Oettli, D. Pallaschke (Springer, Berlin, 1992), pp. 171–194
177. F. Leibfritz, E.W. Sachs, Inexact SQP interior point methods and large scale optimal control problems. SIAM J. Control Optim. **38**, 272–293 (1999)
178. A. Levy, Solution sensitivity from general principles. SIAM J. Control Optim. **40**, 1–38 (2001)
179. A.S. Lewis, S.J. Wright, Identifying activity. SIAM J. Optim. **21**, 597–614 (2011)
180. D.-H. Li, L. Qi, Stabilized SQP method via linear equations. Applied Mathematics Technical Reptort AMR00/5. University of New South Wales, Sydney, 2000
181. G.H. Lin, M. Fukushima, Hybrid approach with active set identification for mathematical programs with complementarity constraints. J. Optim. Theory Appl. **128**, 1–28 (2006)
182. M.S. Lobo, L. Vandenberghe, S. Boyd, H. Lebret, Applications of second-order cone programming. Linear Algebra Appl. **284**, 193–228 (1998)
183. P.A. Lotito, L.A. Parente, M.V. Solodov, A class of variable metric decomposition methods for monotone variational inclusions. J. Convex Anal. **16**, 857–880 (2009)
184. Z.-Q. Luo, J.-S. Pang, D. Ralph, *Mathematical Programs with Equilibrium Constraints* (Cambridge University Press, Cambridge, 1996)
185. MacMPEC. (2012) http://wiki.mcs.anl.gov/leyffer/index.php/MacMPEC
186. O.L. Mangasarian, *Nonlinear Programming* (SIAM, Philadelphia, 1994)
187. O.L. Mangasarian, S. Fromovitz, The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. J. Math. Anal. Appl. **17**, 37–47 (1967)
188. N. Maratos, Exact penalty function algorithms for finite dimensional and control optimization problems. Ph.D. thesis, University of London, UK, 1978
189. P. Marcotte, J.-P. Dussault, A note on a globally convergent Newton method for solving monotone variational inequalities. Oper. Res. Lett. **6**, 35–42 (1987)
190. B. Martinet, Regularisation d'inequations variationelles par approximations successives. Revue Française d'Informatique et de Recherche Opérationelle **4**, 154–159 (1970)
191. J.M. Martínez, Practical quasi-Newton methods for solving nonlinear systems. J. Comput. Appl. Math. **124**, 97–121 (2000)

192. J.M. Martínez, Inexact restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming. J. Optim. Theory Appl. **111**, 39–58 (2001)
193. J.M. Martínez, E.A. Pilotta, Inexact restoration algorithms for constrained optimization. J. Optim. Theory Appl. **104**, 135–163 (2000)
194. J.M. Martínez, E.A. Pilotta, Inexact restoration methods for nonlinear programming: advances and perspectives, in *Optimization and Control with Applications*, ed. by L.Q. Qi, K.L. Teo, X.Q. Yang (Springer, New York, 2005), pp. 271–292
195. J.M. Martínez, L. Qi, Inexact Newton methods for solving nonsmooth equations. J. Comput. Appl. Math. **60**, 127–145 (1995)
196. W.F. Mascarenhas, Newton's iterates can converge to non-stationary points. Math. Program. **112**, 327–334 (2008)
197. D.Q. Mayne, E. Polak, Feasible direction algorithms for optimization problems with equality and inequality constraints. Math. Program. **11**, 67–80 (1976)
198. R.M. McLeod, Mean value theorems for vector valued functions. Proc. Edinburgh Math. Soc. **14**, 197–209 (1965)
199. R. Mifflin, Semismooth and semiconvex functions in constrained optimization. SIAM J. Control Optim. **15**, 959–972 (1977)
200. J.L. Morales, J. Nocedal, Y. Wu, A sequential quadratic programming algorithm with an additional equality constrained phase. IMA J. Numer. Anal. **32**, 553–579 (2012)
201. B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation* (Springer, Berlin, 2006)
202. J.J. Moré, D.C. Sorensen, Computing a trust region step. SIAM J. Sci. Stat. Comput. **4**, 553–572 (1983)
203. E.M.E. Mostafa, L.N. Vicente, S.J. Wright, Numerical behavior of a stabilized SQP method for degenerate NLP problems, in *Global Optimization and Constraint Satisfaction*, ed. by C. Bliek, C. Jermann, A. Neumaier. Lecture Notes in Computer Science, vol 2861 (Springer, Berlin, 2003), pp. 123–141
204. W. Murray, F.J. Prieto, A sequential quadratic programming algorithm using an incomplete solution of the subproblem. SIAM J. Optim. **5**, 590–640 (1995)
205. B.A. Murtagh, M.A. Saunders, A projected Lagrangian algorithm and its implementation for sparse nonlinear constraints. Math. Program. Study. **16**, 84–117 (1982)
206. B.A. Murtagh, M.A. Saunders, MINOS 5.0 user's guide. Technical Report SOL 83.20. Stanford University, 1983
207. Y.E. Nesterov, A.S. Nemirovskii, *Interior Point Polynomial Methods in Convex Programming: Theory and Applications* (SIAM Publications, Philadelphia, 1993)
208. J. Nocedal, S.J. Wright, *Numerical Optimization*, 2nd edn. (Springer, New York, 2006)

209. C. Oberlin, S.J. Wright, An accelerated Newton method for equations with semismooth Jacobians and nonlinear complementarity problems. Math. Program. **117**, 355–386 (2009)
210. E.O. Omojokun, Trust region algorithms for optimization with nonlinear equality and inequality constraints. Ph.D. thesis, Department of Computer Science, University of Colorado at Boulder, U.S.A., 1989
211. J.V. Outrata, M. Kocvara, J. Zowe, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications and Numerical Results* (Kluwer Academic, Dordrecht, 1998)
212. J.-S. Pang, S.A. Gabriel, NE/SQP: a robust algorithm for the nonlinear complementarity problem. Math. Program. **60**, 295–338 (1993)
213. J.-S. Pang, L. Qi, Nonsmooth equations: motivation and algorithms. SIAM J. Optim. **3**, 443–465 (1993)
214. E.R. Panier, A.L. Tits, J. Herskovits, A QP-free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization. SIAM J. Control Optim. **26**, 788–811 (1988)
215. V.M. Panin, A second-order method for discrete min-max problem. USSR Comput. Math. Math. Phys. **19**, 90–100 (1979)
216. L.A. Parente, P.A. Lotito, M.V. Solodov, A class of inexact variable metric proximal point algorithms. SIAM J. Optim. **19**, 240–260 (2008)
217. J.-M. Peng, M. Fukushima, A hybrid Newton method for solving the variational inequality problem via the D-gap function. Math. Program. **86**, 367–386 (1999)
218. J.-M. Peng, C. Kanzow, M. Fukushima, A hybrid Josephy-Newton method for solving box constrained variational inequality problem via the D-gap function. Optim. Method. Softw. **10**, 687–710 (1999)
219. M.J.D. Powell, A fast algorithm for nonlinearly constrained optimization calculations, in *Numerical Analysis Dundee 1977*, ed. by G.A. Watson. Lecture Notes in Mathematics, vol 630 (Springer, Berlin, 1978), pp. 144–157
220. M.J.D. Powell, Y. Yuan, A recursive quadratic programming algorithm that uses differentiable exact penalty function. Math. Program. **35**, 265–278 (1986)
221. L. Qi, LC^1 functions and LC^1 optimization problems. Technical Report AMR 91/21. School of Mathematics, The University of New South Wales, Sydney, 1991
222. L. Qi, Convergence analysis of some algorithms for solving nonsmooth equations. Math. Oper. Res. **18**, 227–244 (1993)
223. L. Qi, Superlinearly convergent approximate Newton methods for LC^1 optimization problems. Math. Program. **64**, 277–294 (1994)
224. L. Qi, H. Jiang, Semismooth Karush–Kuhn–Tucker equations and convergence analysis of Newton and quasi-Newton methods for solving these equations. Math. Oper. Res. **22**, 301–325 (1997)
225. L. Qi, J. Sun, A nonsmooth version of Newton's method. Math. Program. **58**, 353–367 (1993)

226. H.-D. Qi, L. Qi, A new QP-free, globally convergent, locally superlinearlly convergent algorithm for inequality constrained optimization. SIAM J. Optim. **11**, 113–132 (2000)
227. H. Rademacher, Über partielle und totale differenzierbarkeit I. Math. Ann. **89**, 340–359 (1919)
228. D. Ralph, Global convergence of damped Newtons method for nonsmooth equations, via the path search. Math. Oper. Res. **19**, 352–389 (1994)
229. D. Ralph, Sequential quadratic programming for mathematical programs with linear complementarity constraints, in *Computational Techniques and Applications: CTAC95*, ed. by R.L. May, A.K. Easton (World Scientific, Singapore, 1996), pp. 663–668
230. A.R. Ribeiro, E.W. Karas, C.C. Gonzaga, Global convergence of filter methods for nonlinear programming. SIAM J. Optim. **14**, 646–669 (2008)
231. S.M. Robinson, A quadratically convergent algorithm for general nonlinear programming problems. Math. Program. **3**, 145–156 (1972)
232. S.M. Robinson, Perturbed Kuhn–Tucker points and rates of convergence for a class of nonlinear-programming algorithms. Math. Program. **7**, 1–16 (1974)
233. S.M. Robinson, Stability theory for systems of inequalities, Part II. Differentiable nonlinear systems, SIAM J. Numer. Anal. **13**, 497–513 (1976)
234. S.M. Robinson, Strongly regular generalized equations. Math. Oper. Res. **5**, 43–62 (1980)
235. R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, 1970)
236. R.T. Rockafellar, Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
237. R.T. Rockafellar, Computational schemes for solving large-scale problems in extended linear-quadratic programming. Math. Program. **48**, 447–474 (1990)
238. R.T. Rockafellar, J.B. Wets, Generalized linear-quadratic problems of deterministic and stochastic optimal control in discrete time. SIAM J. Control Optim. **28**, 810–922 (1990)
239. R.T. Rockafellar, R.J.-B. Wets, *Variational Analysis* (Springer, Berlin, 1998)
240. H. Scheel, S. Scholtes, Mathematical programs with complementarity constraints: stationarity, optimality and sensitivity. Math. Oper. Res. **25**, 1–22 (2000)
241. S. Scholtes, M. Stöhr, How stringent is the linear independence assumption for mathematical programs with complementarity constraints? Math. Oper. Res. **26**, 851–863 (2001)
242. E. Schröder, Über unendlich viele algorithmen zur auflösung der gleichungen. Math. Annalen. **2**, 317–365 (1870)

243. D. Serre, *Matrices: Theory and Applications*, 2nd edn. (Springer, New York, 2010)
244. A. Shapiro, On concepts of directional differentiability. J. Optim. Theory Appl. **66**, 477–487 (1990)
245. A. Shapiro, Sensitivity analysis of generalized equations. J. Math. Sci. **115**, 2554–2565 (2003)
246. M.V. Solodov, A class of globally convergent algorithms for pseudomonotone variational inequalities, in *Complementarity: Applications, Algorithms and Extensions*, ed. by M.C. Ferris, O.L. Mangasarian, J.-S. Pang (Kluwer Academic, Dordrecht, 2001), pp. 297–315
247. M.V. Solodov, A class of decomposition methods for convex optimization and monotone variational inclusions via the hybrid inexact proximal point framework. Optim. Method. Softw. **19**, 557–575 (2004)
248. M.V. Solodov, On the sequential quadratically constrained quadratic programming methods. Math. Oper. Res. **29**, 64–79 (2004)
249. M.V. Solodov, Global convergence of an SQP method without boundedness assumptions on any of the iterative sequences. Math. Program. **118**, 1–12 (2009)
250. M.V. Solodov, Constraint qualifications, in *Wiley Encyclopedia of Operations Research and Management Science*, ed. by J.J. Cochran (Wiley, London, 2010)
251. M.V. Solodov, B.F. Svaiter, A globally convergent inexact Newton method for systems of monotone equations, in *Reformulation – Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, ed. by M. Fukushima, L. Qi (Kluwer Academic, Dordrecht, 1999), pp. 355–369
252. M.V. Solodov, B.F. Svaiter, A hybrid approximate extragradient–proximal point algorithm using the enlargement of a maximal monotone operator. Set-Valued Anal. **7**, 323–345 (1999)
253. M.V. Solodov, B.F. Svaiter, A hybrid projection–proximal point algorithm. J. Convex Anal. **6**, 59–70 (1999)
254. M.V. Solodov, B.F. Svaiter, A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem SIAM J. Optim. **10**, 605–625 (2000)
255. M.V. Solodov, B.F. Svaiter, Error bounds for proximal point subproblems and associated inexact proximal point algorithms. Math. Program. **88**, 371–389 (2000)
256. M.V. Solodov, B.F. Svaiter, A unified framework for some inexact proximal point algorithms. Numer. Func. Anal. Optim. **22**, 1013–1035 (2001)
257. M.V. Solodov, B.F. Svaiter, A new proximal-based globalization strategy for the Josephy-Newton method for variational inequalities. Optim. Method. Softw. **17**, 965–983 (2002)
258. O. Stein, Lifting mathematical programs with complementarity constraints. Math. Program. **131**, 71–94 (2012)

259. K. Taji, M. Fukushima, T. Ibaraki, A globally convergent Newton method for solving strongly monotone variational inequalities. Math. Program. **58**, 369–383 (1993)
260. A.L. Tits, A. Wächter, S. Bakhtiari, T.J. Urban, C.T. Lawrence, A primal-dual interior-point method for nonlinear programming with strong global and local convergence properties. SIAM J. Optim. **14**, 173–199 (2003)
261. L. Trefethen, D. Bau, *Numerical Linear Algebra* (SIAM, Philadelphia, 1997)
262. K. Ueda, N. Yamashita, Convergence properties of the regularized Newton method for the unconstrained nonconvex optimization. Appl. Math. Optim. **62**, 27–46 (2010)
263. A. Vardi, A trust region algorithm for equality constrained minimization: convergence properties and implementation. SIAM J. Numer. Anal. **22**, 575–591 (1985)
264. A. Wächter, L. Biegler, Line search filter methods for nonlinear programming: local convergence. SIAM J. Optim. **16**, 32–48 (2005)
265. A. Wächter, L. Biegler, Line search filter methods for nonlinear programming: motivation and global convergence. SIAM J. Optim. **16**, 1–31 (2005)
266. E.J. Wiest, E. Polak, A generalized quadratic programming-based phase-I–phase-II method for inequality-constrained optimization. Appl. Math. Optim. **26**, 223–252 (1992)
267. K.A. Williamson, A robust trust region algorithm for nonlinear programming. Ph.D. thesis, Department of Math. Sciences, Rice University, Houston, 1990
268. R.B. Wilson, A simplicial algorithm for concave programming. Ph.D. thesis, Graduate School of Business Administration, Harvard University, U.S.A., 1963
269. S.J. Wright, Superlinear convergence of a stabilized SQP method to a degenerate solution. Comput. Optim. Appl. **11**, 253–275 (1998)
270. S.J. Wright, Modifying SQP for degenerate problems. SIAM J. Optim. **13**, 470–497 (2002)
271. S.J. Wright, Constraint identification and algorithm stabilization for degenerate nonlinear programs. Math. Program. **95**, 137–160 (2003)
272. S.J. Wright, An algorithm for degenerate nonlinear programming with rapid local convergence. SIAM J. Optim. **15**, 673–696 (2005)
273. Y.-F. Yang , D.-H. Li, L. Qi, A feasible sequential linear equation method for inequality constrained optimization. SIAM J. Optim. **13**, 1222–1244 (2003)
274. Y. Yuan, On the convergence of a new trust region algorithm. Numer. Math. **70**, 515–539 (1995)

Index

A
active constraint 2

B
B-differential 44
  partial 51

C
canonical perturbations 26
composite-step approach 417
condition
  complementarity 11
  Dennis–Moré 72, 94, 145, 240, 284, 300
  lower-level strict complementarity condition 525, 540
  quadratic growth 7, 10, 13
  strict complementarity 12, 142
  upper-level strict complementarity condition 519
cone
  contingent 3
  critical 12, 32
  normal 15
constraint qualification 1
  linearity 2
  linear independence, LICQ 2
  Mangasarian–Fromovitz, MFCQ 2
  Slater 2
  strict Mangasarian–Fromovitz, SMFCQ 2
convergence rate
  geometric 546
  linear 546
  quadratic 547
  superlinear 547
correction
  Powell's 213
  second-order 264, 393

D
descent direction 84
directional derivative 48
dog-leg strategy 111

E
elastic mode 386
equation
  generalized 15
  quasi-Newton (secant) 73, 100
error bound 23, 23, 33, 39, 41, 50, 54, 296
exact penalization 369

F
f-iteration 419
feasible set 4, 15
filter strategy 419
first-order approximation 320
forbidden region 419
forcing sequence 69
full-space approach 416
function
  coercive 6
  complementarity 148
  Fischer–Burmeister 148
  forcing 219, 242, 249, 254
  merit 104
  natural residual 148
  objective 4
  penalty 368

G
generalized Jacobian 44
  partial 51

A.F. Izmailov and M.V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-3-319-04247-3,

H

Hessian
  generalized 58
  reduced 215

L

l_1-penalty 368
l_∞-penalty 368
Lagrange multiplier 9, 11, 16
  basic 521
  critical 31
  critical with respect to an index set 463
  noncritical 31
  noncritical with respect to an index set 464
  normal 459
Lagrangian 8, 10
  augmented 216
lifted reformulation of MPCC 529
linesearch, linesearch rule 85
  Armijo 86
  exact 85
  Goldstein 88
  Wolfe 88

M

mapping
  base 15
  BD-regular 51
  CD-regular 51
  coercive 17
  continuously invertible 321
  directionally differentiable 49
  field 15
  gradient 81
  Lipschitz-continuous 43
  Lipschitz-continuously invertible 321
  locally Lipschitz-continuous 43
  locally Lipschitz-continuous with respect to a point 43
  natural residual 22
  piecewise affine 135
  piecewise linear 135
  piecewise smooth 134
  semismooth 52
  strictly differentiable 48
  strongly semismooth 52
Maratos effect 389
mathematical program
  with complementarity constraints, MPCC 511
  with equilibrium constraints, MPEC 512
  with vanishing constraints, MPVC 538
method
  active-set 194
  active-set Gauss–Newton 198
  active-set LCL 224
  active-set Newton for MPCC 526
  active-set Newton–Lagrange 224
  Broyden 73
  Broyden–Fletcher–Goldfarb–Shanno, BFGS 102
  Davidon–Fletcher–Powell, DFP 101
  descent 84
  Gauss–Newton 77
  Josephy–Newton 140
  Levenberg–Marquardt 107
  linearly constrained Lagrangian, LCL 217, 251
  Newton 62, 82
  Newton–Lagrange 206
  outer-phase 328
  perturbed Josephy–Newton 144
  perturbed Newton 65
  perturbed Newton–Lagrange 209
  perturbed semismooth Josephy–Newton 185
  perturbed SQP 234
  piecewise Newton 134
  piecewise SQP 524
  proximal point 335
  quasi-Josephy–Newton 145
  quasi-Newton 72, 94
  quasi-Newton–Lagrange 212
  quasi-Newton semismooth SQP 300
  quasi-Newton SQP 240
  reduced Hessian 215
  secant 74
  semismooth Gauss–Newton 132
  semismooth Josephy–Newton 185
  semismooth Levenberg–Marquardt 132
  semismooth Newton 125, 136, 150, 166
  semismooth Newton–Lagrange for the lifted MPCC 533
  semismooth quasi-Josephy–Newton 186
  semismooth quasi-Newton 128
  semismooth SQP 288
  sequential quadratic programming, SQP 228
  sequential quadratically constrained quadratic programming, SQCQP 262

stabilized Newton method 480
stabilized SQP 476
steepest descent 85
truncated LCL 219, 254
truncated Newton 69
truncated Newton–Lagrange 208
truncated semismooth Newton 129
truncated SQP 249
trust-region 107
MPCC-Lagrangian 514
MPCC-linear independence constraint qualification, MPCC-LICQ 514
MPCC-multiplier 514
MPCC-second-order sufficient condition, MPCC-SOSC 517

O
optimality condition
piecewise SOSC 518
second-order necessary, SONC 7, 9, 12, 57
second-order sufficient, SOSC 7, 9, 12, 58, 59
strong second-order sufficient, SSOSC 21

P
P-matrix 20
P_0-matrix 314
parameter
penalty 216, 368
proximal 336
regularization 104, 106, 336
Pareto domination rule 419
path continuation property 322
penalty 368
point
Cauchy 111
critical 7
stationary 6, 8, 9, 11
strongly stationary 514
weakly stationary 514
principle
Fermat 7
Lagrange 9
problem
branch (piece) 515
constrained optimization 5
equality-constrained 5
feasible 4
linear programming 5
mathematical programming 5
mixed complementarity 16
nonlinear complementarity 16
optimization 4
quadratic programming 5
unconstrained optimization 5
variational 15
purification procedure 249

R
relative error 336
relaxed nonlinear programming problem, RNLP 514

S
Schur complement 20
solution
b-regular 152
BD-regular 182
CD-regular 182
global of optimization problem 5
hemistable 140
local of optimization problem 5
quasi-regular 168
semistable 21
strongly regular 19
strongly regular with respect to a set 182
weakly regular 195, 200
step
normal 417
tangential 417
stepsize parameter 84
system
Karush–Kuhn–Tucker, KKT 11, 16
Lagrange optimality 9

T
theorem
chain rule 47
Dennis–Moré 94
implicit function 18, 19
inverse function 51
Karush–Kuhn–Tucker 10
mean-value 46
Rademacher 43
Robinson's stability 3
Weierstrass 5
tightened nonlinear programming problem, TNLP 514

V
variational inequality 16
ε-enlargement 337
ψ-iteration 419