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

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
$B(x, \delta)$	$= \{y \in \mathbf{R}^n \mid \|y - x\| \le \delta\}$; 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°	$= \{\xi \in \mathbf{R}^n \mid \langle \xi, x\rangle \le 0\ \forall\, x \in C\}$; 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 $g_i(x)\le 0$, $i=1, \ldots, m$, 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

$\operatorname{dist}(x,\,S)$	$= \inf_{y \in S} \\|y - x\\|$; 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
$\partial F(x)$	$= \operatorname{conv} \partial_B F(x)$; 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 $(x,\,y)$
$\partial_x F(x,\,y)$	$= \operatorname{conv}(\partial_B)_x F(x,\,y)$; 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 $(x,\,\lambda,\,\mu)$
$\operatorname{dom} \Psi$	$= \{x \in \mathbf{R}^n \mid \Psi(x) \neq \emptyset\}$; the domain of a set-valued mapping Ψ defined on $\mathbf{R}^n$