Springer Series in Operations Research

Editors:
Peter Glynn Stephen M. Robinson

Springer
New York
Berlin
Heidelberg
Barcelona
Hong Kong
London
Milan
Paris
Singapore
Tokyo

Springer Series in Operations Research

Altiok: Performance Analysis of Manufacturing Systems
Birge and Louveaux: Introduction to Stochastic Programming
Bonnans and Shapiro: Perturbation Analysis of Optimization Problems
Bramel and Simchi-Levi: The Logic of Logistics: Theory, Algorithms, and Applications for Logistics Management
Dantzig and Thapa: Linear Programming 1: Introduction
Drezner (Editor): Facility Location: A Survey of Applications and Methods
Fishman: Discrete-Event Simulation: Modeling, Programming, and Analysis
Fishman: Monte Carlo: Concepts, Algorithms, and Applications
Nocedal and Wright: Numerical Optimization
Olson: Decision Aids for Selection Problems
Yao (Editor): Stochastic Modeling and Analysis of Manufacturing Systems

Jorge Nocedal Stephen J. Wright

Numerical Optimization

With 85 Illustrations

Springer

Jorge Nocedal
ECE Department
Northwestern University
Evanston, IL 60208-3118
USA

Stephen J. Wright
Mathematics and Computer
Science Division
Argonne National Laboratory
9700 South Cass Avenue
Argonne, IL 60439-4844
USA

Series Editors:
Peter Glynn
Department of Operations Research
Stanford University
Stanford, CA 94305
USA

Stephen M. Robinson
Department of Industrial Engineering
University of Wisconsin–Madison
1513 University Avenue
Madison, WI 53706-1572
USA

Cover illustration is from *Pre-Hispanic Mexican Stamp Designs* by Frederick V. Field, courtesy of Dover Publications, Inc.

Library of Congress Cataloging-in-Publication Data
Nocedal, Jorge.
Numerical optimization / Jorge Nocedal, Stephen J. Wright.
p. cm. — (Springer series in operations research)
Includes bibliographical references and index.
ISBN 0-387-98793-2 (hardcover)
1. Mathematical optimization. I. Wright, Stephen J., 1960– .
II. Title. III. Series.
QA402.5.N62 1999
519.3—dc21 99–13263

© 1999 Springer-Verlag New York, Inc.
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

ISBN 0-387-98793-2 Springer-Verlag New York Berlin Heidelberg SPIN 10764949

To Our Parents:

Raúl and Concepción Peter and Berenice

Preface

This is a book for people interested in solving optimization problems. Because of the wide (and growing) use of optimization in science, engineering, economics, and industry, it is essential for students and practitioners alike to develop an understanding of optimization algorithms. Knowledge of the capabilities and limitations of these algorithms leads to a better understanding of their impact on various applications, and points the way to future research on improving and extending optimization algorithms and software. Our goal in this book is to give a comprehensive description of the most powerful, state-of-the-art, techniques for solving continuous optimization problems. By presenting the motivating ideas for each algorithm, we try to stimulate the reader's intuition and make the technical details easier to follow. Formal mathematical requirements are kept to a minimum.

Because of our focus on continuous problems, we have omitted discussion of important optimization topics such as discrete and stochastic optimization. However, there are a great many applications that can be formulated as continuous optimization problems; for instance,

finding the optimal trajectory for an aircraft or a robot arm;

identifying the seismic properties of a piece of the earth's crust by fitting a model of the region under study to a set of readings from a network of recording stations;

designing a portfolio of investments to maximize expected return while maintaining an acceptable level of risk;

controlling a chemical process or a mechanical device to optimize performance or meet standards of robustness;

computing the optimal shape of an automobile or aircraft component.

Every year optimization algorithms are being called on to handle problems that are much larger and complex than in the past. Accordingly, the book emphasizes large-scale optimization techniques, such as interior-point methods, inexact Newton methods, limited-memory methods, and the role of partially separable functions and automatic differentiation. It treats important topics such as trust-region methods and sequential quadratic programming more thoroughly than existing texts, and includes comprehensive discussion of such "core curriculum" topics as constrained optimization theory, Newton and quasi-Newton methods, nonlinear least squares and nonlinear equations, the simplex method, and penalty and barrier methods for nonlinear programming.

THE AUDIENCE

We intend that this book will be used in graduate-level courses in optimization, as offered in engineering, operations research, computer science, and mathematics departments. There is enough material here for a two-semester (or three-quarter) sequence of courses. We hope, too, that this book will be used by practitioners in engineering, basic science, and industry, and our presentation style is intended to facilitate self-study. Since the book treats a number of new algorithms and ideas that have not been described in earlier textbooks, we hope that this book will also be a useful reference for optimization researchers.

Prerequisites for this book include some knowledge of linear algebra (including numerical linear algebra) and the standard sequence of calculus courses. To make the book as self-contained as possible, we have summarized much of the relevant material from these areas in the Appendix. Our experience in teaching engineering students has shown us that the material is best assimilated when combined with computer programming projects in which the student gains a good feeling for the algorithms—their complexity, memory demands, and elegance—and for the applications. In most chapters we provide simple computer exercises that require only minimal programming proficiency.

EMPHASIS AND WRITING STYLE

We have used a conversational style to motivate the ideas and present the numerical algorithms. Rather than being as concise as possible, our aim is to make the discussion flow in a natural way. As a result, the book is comparatively long, but we believe that it can be read relatively rapidly. The instructor can assign substantial reading assignments from the text and focus in class only on the main ideas.

A typical chapter begins with a nonrigorous discussion of the topic at hand, including figures and diagrams and excluding technical details as far as possible. In subsequent sections, the algorithms are motivated and discussed, and then stated explicitly. The major theoretical results are stated, and in many cases proved, in a rigorous fashion. These proofs can be skipped by readers who wish to avoid technical details.

The practice of optimization depends not only on efficient and robust algorithms, but also on good modeling techniques, careful interpretation of results, and user-friendly software. In this book we discuss the various aspects of the optimization process—modeling, optimality conditions, algorithms, implementation, and interpretation of results—but not with equal weight. Examples throughout the book show how practical problems are formulated as optimization problems, but our treatment of modeling is light and serves mainly to set the stage for algorithmic developments. We refer the reader to Dantzig [63] and Fourer, Gay, and Kernighan [92] for more comprehensive discussion of this issue. Our treatment of optimality conditions is thorough but not exhaustive; some concepts are discussed more extensively in Mangasarian [154] and Clarke [42]. As mentioned above, we are quite comprehensive in discussing optimization algorithms.

TOPICS NOT COVERED

We omit some important topics, such as network optimization, integer programming, stochastic programming, nonsmooth optimization, and global optimization. Network and integer optimization are described in some excellent texts: for instance, Ahuja, Magnanti, and Orlin [1] in the case of network optimization and Nemhauser and Wolsey [179], Papadimitriou and Steiglitz [190], and Wolsey [249] in the case of integer programming. Books on stochastic optimization are only now appearing; we mention those of Kall and Wallace [139], Birge and Louveaux [11]. Nonsmooth optimization comes in many flavors. The relatively simple structures that arise in robust data fitting (which is sometimes based on the ℓ_1 norm) are treated by Osborne [187] and Fletcher [83]. The latter book also discusses algorithms for nonsmooth penalty functions that arise in constrained optimization; we discuss these briefly, too, in Chapter 18. A more analytical treatment of nonsmooth optimization is given by Hiriart-Urruty and Lemaréchal [137]. We omit detailed treatment of some important topics that are the focus of intense current research, including interior-point methods for nonlinear programming and algorithms for complementarity problems.

ADDITIONAL RESOURCE

The material in the book is complemented by an online resource called the NEOS Guide, which can be found on the World-Wide Web at

`http://www.mcs.anl.gov/otc/Guide/`

The Guide contains information about most areas of optimization, and presents a number of case studies that describe applications of various optimization algorithms to real-world prob-

lems such as portfolio optimization and optimal dieting. Some of this material is interactive in nature and has been used extensively for class exercises.

For the most part, we have omitted detailed discussions of specific software packages, and refer the reader to Moré and Wright [173] or to the Software Guide section of the NEOS Guide, which can be found at

`http://www.mcs.anl.gov/otc/Guide/SoftwareGuide/`

Users of optimization software refer in great numbers to this web site, which is being constantly updated to reflect new packages and changes to existing software.

ACKNOWLEDGMENTS

We are most grateful to the following colleagues for their input and feedback on various sections of this work: Chris Bischof, Richard Byrd, George Corliss, Bob Fourer, David Gay, Jean-Charles Gilbert, Phillip Gill, Jean-Pierre Goux, Don Goldfarb, Nick Gould, Andreas Griewank, Matthias Heinkenschloss, Marcelo Marazzi, Hans Mittelmann, Jorge Moré, Will Naylor, Michael Overton, Bob Plemmons, Hugo Scolnik, David Stewart, Philippe Toint, Luis Vicente, Andreas Waechter, and Ya-xiang Yuan. We thank Guanghui Liu, who provided help with many of the exercises, and Jill Lavelle who assisted us in preparing the figures. We also express our gratitude to our sponsors at the Department of Energy and the National Science Foundation, who have strongly supported our research efforts in optimization over the years.

One of us (JN) would like to express his deep gratitude to Richard Byrd, who has taught him so much about optimization and who has helped him in very many ways throughout the course of his career.

FINAL REMARK

In the preface to his 1987 book [83], Roger Fletcher described the field of optimization as a "fascinating blend of theory and computation, heuristics and rigor." The ever-growing realm of applications and the explosion in computing power is driving optimization research in new and exciting directions, and the ingredients identified by Fletcher will continue to play important roles for many years to come.

Jorge Nocedal
Evanston, IL

Stephen J. Wright
Argonne, IL

Contents

Preface **vii**

1 Introduction **xxi**
Mathematical Formulation 2
Example: A Transportation Problem 4
Continuous versus Discrete Optimization 4
Constrained and Unconstrained Optimization 6
Global and Local Optimization 6
Stochastic and Deterministic Optimization 7
Optimization Algorithms 7
Convexity 8
Notes and References 9

2 Fundamentals of Unconstrained Optimization **10**
2.1 What Is a Solution? 13
Recognizing a Local Minimum 15
Nonsmooth Problems 18

2.2 Overview of Algorithms . . . 19
Two Strategies: Line Search and Trust Region . . . 19
Search Directions for Line Search Methods . . . 21
Models for Trust-Region Methods . . . 26
Scaling . . . 27
Rates of Convergence . . . 28
R-Rates of Convergence . . . 29
Notes and References . . . 30
Exercises . . . 30

3 Line Search Methods **34**
3.1 Step Length . . . 36
The Wolfe Conditions . . . 37
The Goldstein Conditions . . . 41
Sufficient Decrease and Backtracking . . . 41
3.2 Convergence of Line Search Methods . . . 43
3.3 Rate of Convergence . . . 46
Convergence Rate of Steepest Descent . . . 47
Quasi-Newton Methods . . . 49
Newton's Method . . . 51
Coordinate Descent Methods . . . 53
3.4 Step-Length Selection Algorithms . . . 55
Interpolation . . . 56
The Initial Step Length . . . 58
A Line Search Algorithm for the Wolfe Conditions . . . 58
Notes and References . . . 61
Exercises . . . 62

4 Trust-Region Methods **64**
Outline of the Algorithm . . . 67
4.1 The Cauchy Point and Related Algorithms . . . 69
The Cauchy Point . . . 69
Improving on the Cauchy Point . . . 70
The Dogleg Method . . . 71
Two-Dimensional Subspace Minimization . . . 74
Steihaug's Approach . . . 75
4.2 Using Nearly Exact Solutions to the Subproblem . . . 77
Characterizing Exact Solutions . . . 77
Calculating Nearly Exact Solutions . . . 78
The Hard Case . . . 82
Proof of Theorem 4.3 . . . 84
4.3 Global Convergence . . . 87

Reduction Obtained by the Cauchy Point . . . 87
Convergence to Stationary Points . . . 89
Convergence of Algorithms Based on Nearly Exact Solutions . . . 93
4.4 Other Enhancements . . . 94
Scaling . . . 94
Non-Euclidean Trust Regions . . . 96
Notes and References . . . 97
Exercises . . . 97

5 Conjugate Gradient Methods 100
5.1 The Linear Conjugate Gradient Method . . . 102
Conjugate Direction Methods . . . 102
Basic Properties of the Conjugate Gradient Method . . . 107
A Practical Form of the Conjugate Gradient Method . . . 111
Rate of Convergence . . . 112
Preconditioning . . . 118
Practical Preconditioners . . . 119
5.2 Nonlinear Conjugate Gradient Methods . . . 120
The Fletcher–Reeves Method . . . 120
The Polak–Ribière Method . . . 121
Quadratic Termination and Restarts . . . 122
Numerical Performance . . . 124
Behavior of the Fletcher–Reeves Method . . . 124
Global Convergence . . . 127
Notes and References . . . 131
Exercises . . . 132

6 Practical Newton Methods 134
6.1 Inexact Newton Steps . . . 136
6.2 Line Search Newton Methods . . . 139
Line Search Newton–CG Method . . . 139
Modified Newton's Method . . . 141
6.3 Hessian Modifications . . . 142
Eigenvalue Modification . . . 143
Adding a Multiple of the Identity . . . 144
Modified Cholesky Factorization . . . 145
Gershgorin Modification . . . 150
Modified Symmetric Indefinite Factorization . . . 151
6.4 Trust-Region Newton Methods . . . 154
Newton–Dogleg and Subspace-Minimization Methods . . . 154
Accurate Solution of the Trust-Region Problem . . . 155
Trust-Region Newton–CG Method . . . 156

Preconditioning the Newton–CG Method 157
Local Convergence of Trust-Region Newton Methods 159
Notes and References 162
Exercises 162

7 Calculating Derivatives 164
7.1 Finite-Difference Derivative Approximations 166
Approximating the Gradient 166
Approximating a Sparse Jacobian 169
Approximating the Hessian 173
Approximating a Sparse Hessian 174
7.2 Automatic Differentiation 176
An Example 177
The Forward Mode 178
The Reverse Mode 179
Vector Functions and Partial Separability 183
Calculating Jacobians of Vector Functions 184
Calculating Hessians: Forward Mode 185
Calculating Hessians: Reverse Mode 187
Current Limitations 188
Notes and References 189
Exercises 189

8 Quasi-Newton Methods 192
8.1 The BFGS Method 194
Properties of the BFGS Method 199
Implementation 200
8.2 The SR1 Method 202
Properties of SR1 Updating 205
8.3 The Broyden Class 207
Properties of the Broyden Class 209
8.4 Convergence Analysis 211
Global Convergence of the BFGS Method 211
Superlinear Convergence of BFGS 214
Convergence Analysis of the SR1 Method 218
Notes and References 219
Exercises 220

9 Large-Scale Quasi-Newton and Partially Separable Optimization 222
9.1 Limited-Memory BFGS 224
Relationship with Conjugate Gradient Methods 227
9.2 General Limited-Memory Updating 229

Compact Representation of BFGS Updating . . . 230
SR1 Matrices . . . 232
Unrolling the Update . . . 232
9.3 Sparse Quasi-Newton Updates . . . 233
9.4 Partially Separable Functions . . . 235
A Simple Example . . . 236
Internal Variables . . . 237
9.5 Invariant Subspaces and Partial Separability . . . 240
Sparsity vs. Partial Separability . . . 242
Group Partial Separability . . . 243
9.6 Algorithms for Partially Separable Functions . . . 244
Exploiting Partial Separability in Newton's Method . . . 244
Quasi-Newton Methods for Partially Separable Functions . . . 245
Notes and References . . . 247
Exercises . . . 248

10 Nonlinear Least-Squares Problems 250
10.1 Background . . . 253
Modeling, Regression, Statistics . . . 253
Linear Least-Squares Problems . . . 256
10.2 Algorithms for Nonlinear Least-Squares Problems . . . 259
The Gauss–Newton Method . . . 259
The Levenberg–Marquardt Method . . . 262
Implementation of the Levenberg–Marquardt Method . . . 264
Large-Residual Problems . . . 266
Large-Scale Problems . . . 269
10.3 Orthogonal Distance Regression . . . 271
Notes and References . . . 273
Exercises . . . 274

11 Nonlinear Equations 276
11.1 Local Algorithms . . . 281
Newton's Method for Nonlinear Equations . . . 281
Inexact Newton Methods . . . 284
Broyden's Method . . . 286
Tensor Methods . . . 290
11.2 Practical Methods . . . 292
Merit Functions . . . 292
Line Search Methods . . . 294
Trust-Region Methods . . . 298
11.3 Continuation/Homotopy Methods . . . 304
Motivation . . . 304

Practical Continuation Methods . . . 306
Notes and References . . . 310
Exercises . . . 311

12 Theory of Constrained Optimization 314
Local and Global Solutions . . . 316
Smoothness . . . 317
12.1 Examples . . . 319
A Single Equality Constraint . . . 319
A Single Inequality Constraint . . . 321
Two Inequality Constraints . . . 324
12.2 First-Order Optimality Conditions . . . 327
Statement of First-Order Necessary Conditions . . . 327
Sensitivity . . . 330
12.3 Derivation of the First-Order Conditions . . . 331
Feasible Sequences . . . 331
Characterizing Limiting Directions: Constraint Qualifications . . . 336
Introducing Lagrange Multipliers . . . 339
Proof of Theorem 12.1 . . . 341
12.4 Second-Order Conditions . . . 342
Second-Order Conditions and Projected Hessians . . . 348
Convex Programs . . . 349
12.5 Other Constraint Qualifications . . . 350
12.6 A Geometric Viewpoint . . . 353
Notes and References . . . 356
Exercises . . . 357

13 Linear Programming: The Simplex Method 360
Linear Programming . . . 362
13.1 Optimality and Duality . . . 364
Optimality Conditions . . . 364
The Dual Problem . . . 365
13.2 Geometry of the Feasible Set . . . 368
Basic Feasible Points . . . 368
Vertices of the Feasible Polytope . . . 370
13.3 The Simplex Method . . . 372
Outline of the Method . . . 372
Finite Termination of the Simplex Method . . . 374
A Single Step of the Method . . . 376
13.4 Linear Algebra in the Simplex Method . . . 377
13.5 Other (Important) Details . . . 381
Pricing and Selection of the Entering Index . . . 381

Starting the Simplex Method . . . 384
Degenerate Steps and Cycling . . . 387
13.6 Where Does the Simplex Method Fit? . . . 389
Notes and References . . . 390
Exercises . . . 391

14 Linear Programming: Interior-Point Methods 392
14.1 Primal–Dual Methods . . . 394
Outline . . . 394
The Central Path . . . 397
A Primal–Dual Framework . . . 399
Path-Following Methods . . . 400
14.2 A Practical Primal–Dual Algorithm . . . 402
Solving the Linear Systems . . . 406
14.3 Other Primal–Dual Algorithms and Extensions . . . 407
Other Path-Following Methods . . . 407
Potential-Reduction Methods . . . 407
Extensions . . . 408
14.4 Analysis of Algorithm 14.2 . . . 409
Notes and References . . . 414
Exercises . . . 415

15 Fundamentals of Algorithms for Nonlinear Constrained Optimization 418
Initial Study of a Problem . . . 420
15.1 Categorizing Optimization Algorithms . . . 422
15.2 Elimination of Variables . . . 424
Simple Elimination for Linear Constraints . . . 426
General Reduction Strategies for Linear Constraints . . . 429
The Effect of Inequality Constraints . . . 431
15.3 Measuring Progress: Merit Functions . . . 432
Notes and References . . . 436
Exercises . . . 436

16 Quadratic Programming 438
An Example: Portfolio Optimization . . . 440
16.1 Equality–Constrained Quadratic Programs . . . 441
Properties of Equality-Constrained QPs . . . 442
16.2 Solving the KKT System . . . 445
Direct Solution of the KKT System . . . 446
Range-Space Method . . . 447
Null-Space Method . . . 448
A Method Based on Conjugacy . . . 450

16.3 Inequality-Constrained Problems . . . 451
Optimality Conditions for Inequality-Constrained Problems . . . 452
Degeneracy . . . 453
16.4 Active-Set Methods for Convex QP . . . 455
Specification of the Active-Set Method for Convex QP . . . 460
An Example . . . 461
Further Remarks on the Active-Set Method . . . 463
Finite Termination of the Convex QP Algorithm . . . 464
Updating Factorizations . . . 465
16.5 Active-Set Methods for Indefinite QP . . . 468
Illustration . . . 470
Choice of Starting Point . . . 472
Failure of the Active-Set Method . . . 473
Detecting Indefiniteness Using the LBL^T Factorization . . . 473
16.6 The Gradient–Projection Method . . . 474
Cauchy Point Computation . . . 475
Subspace Minimization . . . 478
16.7 Interior-Point Methods . . . 479
Extensions and Comparison with Active-Set Methods . . . 482
16.8 Duality . . . 482
Notes and References . . . 483
Exercises . . . 484

17 Penalty, Barrier, and Augmented Lagrangian Methods 488
17.1 The Quadratic Penalty Method . . . 490
Motivation . . . 490
Algorithmic Framework . . . 492
Convergence of the Quadratic Penalty Function . . . 493
17.2 The Logarithmic Barrier Method . . . 498
Properties of Logarithmic Barrier Functions . . . 498
Algorithms Based on the Log-Barrier Function . . . 503
Properties of the Log-Barrier Function and Framework 17.2 . . . 505
Handling Equality Constraints . . . 507
Relationship to Primal–Dual Methods . . . 508
17.3 Exact Penalty Functions . . . 510
17.4 Augmented Lagrangian Method . . . 511
Motivation and Algorithm Framework . . . 512
Extension to Inequality Constraints . . . 514
Properties of the Augmented Lagrangian . . . 517
Practical Implementation . . . 520
17.5 Sequential Linearly Constrained Methods . . . 522
Notes and References . . . 523

Exercises . . . 524

18 Sequential Quadratic Programming 526
18.1 Local SQP Method . . . 528
SQP Framework . . . 529
Inequality Constraints . . . 531
IQP vs. EQP . . . 531
18.2 Preview of Practical SQP Methods . . . 532
18.3 Step Computation . . . 534
Equality Constraints . . . 534
Inequality Constraints . . . 536
18.4 The Hessian of the Quadratic Model . . . 537
Full Quasi-Newton Approximations . . . 538
Hessian of Augmented Lagrangian . . . 539
Reduced-Hessian Approximations . . . 540
18.5 Merit Functions and Descent . . . 542
18.6 A Line Search SQP Method . . . 545
18.7 Reduced-Hessian SQP Methods . . . 546
Some Properties of Reduced-Hessian Methods . . . 547
Update Criteria for Reduced-Hessian Updating . . . 548
Changes of Bases . . . 549
A Practical Reduced-Hessian Method . . . 550
18.8 Trust-Region SQP Methods . . . 551
Approach I: Shifting the Constraints . . . 553
Approach II: Two Elliptical Constraints . . . 554
Approach III: Sℓ_1QP (Sequential ℓ_1 Quadratic Programming) . . . 555
18.9 A Practical Trust-Region SQP Algorithm . . . 558
18.10 Rate of Convergence . . . 561
Convergence Rate of Reduced-Hessian Methods . . . 563
18.11 The Maratos Effect . . . 565
Second-Order Correction . . . 568
Watchdog (Nonmonotone) Strategy . . . 569
Notes and References . . . 571
Exercises . . . 572

A Background Material 574
A.1 Elements of Analysis, Geometry, Topology . . . 575
Topology of the Euclidean Space $\mathbb{R}^n$ . . . 575
Continuity and Limits . . . 578
Derivatives . . . 579
Directional Derivatives . . . 581
Mean Value Theorem . . . 582

Implicit Function Theorem 583
Geometry of Feasible Sets 584
Order Notation 589
Root-Finding for Scalar Equations 590
A.2 Elements of Linear Algebra 591
Vectors and Matrices 591
Norms 592
Subspaces 595
Eigenvalues, Eigenvectors, and the Singular-Value Decomposition 596
Determinant and Trace 597
Matrix Factorizations: Cholesky, LU, QR 598
Sherman–Morrison–Woodbury Formula 603
Interlacing Eigenvalue Theorem 603
Error Analysis and Floating-Point Arithmetic 604
Conditioning and Stability 606

References **609**

Index **623**