

国外数学名著系列

(影印版)6

Jorge Nocedal Stephen J. Wright

Numerical Optimization

数值最优化

国外数学名著系列(影印版) 6

Numerical Optimization

数值最优化

Jorge Nocedal Stephen J. Wright

科学出版社

北京

图字:01-2005-6747

Jorge Nocedal, Stephen J. Wright: Numerical Optimization
© 1999 Springer Science+Business Media, Inc.

This reprint has been authorized by Springer-Verlag (Berlin/Heidelberg/New York) for sale in the People's Republic of China only and not for export therefrom.

本书英文影印版由德国施普林格出版公司授权出版。未经出版者书面许可,不得以任何方式复制或抄袭本书的任何部分。本书仅限在中华人民共和国销售,不得出口。版权所有,翻印必究。

图书在版编目(CIP)数据

数值最优化=Numerical Optimization/(美)劳斯特(Nocedal, J.)等著. —影印版. —北京:科学出版社,2006

(国外数学名著系列)

ISBN 7-03-016675-2

I. 数… II. 劳… III. 最优化算法—研究生—教材—英文 IV. 0224

中国版本图书馆 CIP 数据核字(2005)第 154405 号

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

中国科学院印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2006年1月第 一 版 开本: B5(720×1000)

2006年1月第一次印刷 印张: 41 1/2

印数: 1—3 000 字数: 779 000

定价: 98.00 元

(如有印装质量问题,我社负责调换(科印))

《国外数学名著系列》(影印版)专家委员会

(按姓氏笔画排序)

丁伟岳 王 元 石钟慈 冯克勤 严加安 李邦河
李大潜 张伟平 张继平 杨 乐 姜伯驹 郭 雷

项目策划

向安全 林 鹏 王春香 吕 虹 范庆奎 王 璐

执行编辑

范庆奎

《国外数学名著系列》(影印版)序

要使我国的数学事业更好地发展起来,需要数学家淡泊名利并付出更艰苦地努力。另一方面,我们也要从客观上为数学家创造更有利的发展数学事业的外部环境,这主要是加强对数学事业的支持与投资力度,使数学家有较好的工作与生活条件,其中也包括改善与加强数学的出版工作。

从出版方面来讲,除了较好较快地出版我们自己的成果外,引进国外的先进出版物无疑也是十分重要与必不可少的。从数学来说,施普林格(Springer)出版社至今仍然是世界上最具权威的出版社。科学出版社影印一批他们出版的好的新书,使我国广大数学家能以较低的价格购买,特别是在边远地区工作的数学家能普遍见到这些书,无疑是对推动我国数学的科研与教学十分有益的事。

这次科学出版社购买了版权,一次影印了23本施普林格出版社出版的数学书,就是一件好事,也是值得继续做下去的事情。大体上分一下,这23本书中,包括基础数学书5本,应用数学书6本与计算数学书12本,其中有些书也具有交叉性质。这些书都是很新的,2000年以后出版的占绝大部分,共计16本,其余的也是1990年以后出版的。这些书可以使读者较快地了解数学某方面的前沿,例如基础数学中的数论、代数与拓扑三本,都是由该领域大数学家编著的“数学百科全书”的分册。对从事这方面研究的数学家了解该领域的前沿与全貌很有帮助。按照学科的特点,基础数学类的书以“经典”为主,应用和计算数学类的书以“前沿”为主。这些书的作者多数是国际知名的大数学家,例如《拓扑学》一书的作者诺维科夫是俄罗斯科学院的院士,曾获“菲尔兹奖”和“沃尔夫数学奖”。这些大数学家的著作无疑将会对我国的科研人员起到非常好的指导作用。

当然,23本书只能涵盖数学的一部分,所以,这项工作还应该继续做下去。更进一步,有些读者面较广的好书还应该翻译成中文出版,使之有更大的读者群。

总之,我对科学出版社影印施普林格出版社的部分数学著作这一举措表示热烈的支持,并盼望这一工作取得更大的成绩。

王元

2005年12月3日

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 Sotnick, 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 Stephen J. Wright
Evanston, IL Argonne, IL

Contents

Preface	v
1 Introduction	1
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	332
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	350
12.5 Other Constraint Qualifications	351
12.6 A Geometric Viewpoint	354
Notes and References	357
Exercises	358
13 Linear Programming: The Simplex Method	362
Linear Programming	364
13.1 Optimality and Duality	366
Optimality Conditions	366
The Dual Problem	367
13.2 Geometry of the Feasible Set	370
Basic Feasible Points	370
Vertices of the Feasible Polytope	372
13.3 The Simplex Method	374
Outline of the Method	374
Finite Termination of the Simplex Method	377
A Single Step of the Method	378
13.4 Linear Algebra in the Simplex Method	379
13.5 Other (Important) Details	383
Pricing and Selection of the Entering Index	383

Starting the Simplex Method	386
Degenerate Steps and Cycling	389
13.6 Where Does the Simplex Method Fit?	391
Notes and References	392
Exercises	393
14 Linear Programming: Interior-Point Methods	394
14.1 Primal–Dual Methods	396
Outline	396
The Central Path	399
A Primal–Dual Framework	401
Path-Following Methods	402
14.2 A Practical Primal–Dual Algorithm	404
Solving the Linear Systems	408
14.3 Other Primal–Dual Algorithms and Extensions	409
Other Path-Following Methods	409
Potential-Reduction Methods	409
Extensions	410
14.4 Analysis of Algorithm 14.2	411
Notes and References	416
Exercises	417
15 Fundamentals of Algorithms for Nonlinear Constrained Optimization	420
Initial Study of a Problem	422
15.1 Categorizing Optimization Algorithms	423
15.2 Elimination of Variables	426
Simple Elimination for Linear Constraints	427
General Reduction Strategies for Linear Constraints	430
The Effect of Inequality Constraints	434
15.3 Measuring Progress: Merit Functions	434
Notes and References	437
Exercises	438
16 Quadratic Programming	440
An Example: Portfolio Optimization	442
16.1 Equality–Constrained Quadratic Programs	443
Properties of Equality-Constrained QPs	444
16.2 Solving the KKT System	447
Direct Solution of the KKT System	448
Range-Space Method	449
Null-Space Method	450
A Method Based on Conjugacy	452

16.3	Inequality-Constrained Problems	453
	Optimality Conditions for Inequality-Constrained Problems	454
	Degeneracy	455
16.4	Active-Set Methods for Convex QP	457
	Specification of the Active-Set Method for Convex QP	461
	An Example	463
	Further Remarks on the Active-Set Method	465
	Finite Termination of the Convex QP Algorithm	466
	Updating Factorizations	467
16.5	Active-Set Methods for Indefinite QP	470
	Illustration	472
	Choice of Starting Point	474
	Failure of the Active-Set Method	475
	Detecting Indefiniteness Using the LBL^T Factorization	475
16.6	The Gradient-Projection Method	476
	Cauchy Point Computation	477
	Subspace Minimization	480
16.7	Interior-Point Methods	481
	Extensions and Comparison with Active-Set Methods	484
16.8	Duality	484
	Notes and References	485
	Exercises	486
17	Penalty, Barrier, and Augmented Lagrangian Methods	490
17.1	The Quadratic Penalty Method	492
	Motivation	492
	Algorithmic Framework	494
	Convergence of the Quadratic Penalty Function	495
17.2	The Logarithmic Barrier Method	500
	Properties of Logarithmic Barrier Functions	500
	Algorithms Based on the Log-Barrier Function	505
	Properties of the Log-Barrier Function and Framework 17.2	507
	Handling Equality Constraints	509
	Relationship to Primal-Dual Methods	510
17.3	Exact Penalty Functions	512
17.4	Augmented Lagrangian Method	513
	Motivation and Algorithm Framework	513
	Extension to Inequality Constraints	516
	Properties of the Augmented Lagrangian	518
	Practical Implementation	521
17.5	Sequential Linearly Constrained Methods	523
	Notes and References	525