

UNITEXT

La Matematica per il 3+2

Volume 152

Editor-in-Chief

Alfio Quarteroni, Politecnico di Milano, Milan, Italy

École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Series Editors

Luigi Ambrosio, Scuola Normale Superiore, Pisa, Italy

Paolo Biscari, Politecnico di Milano, Milan, Italy

Ciro Ciliberto, Università di Roma "Tor Vergata", Rome, Italy

Camillo De Lellis, Institute for Advanced Study, Princeton, New Jersey, USA

Victor Panaretos, Institute of Mathematics, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Lorenzo Rosasco, DIBRIS, Università degli Studi di Genova, Genova, Italy

Center for Brains Mind and Machines, Massachusetts Institute of Technology,
Cambridge, Massachusetts, US

Istituto Italiano di Tecnologia, Genova, Italy

The **UNITEXT - La Matematica per il 3+2** series is designed for undergraduate and graduate academic courses, and also includes advanced textbooks at a research level.

Originally released in Italian, the series now publishes textbooks in English addressed to students in mathematics worldwide.

Some of the most successful books in the series have evolved through several editions, adapting to the evolution of teaching curricula.

Submissions must include at least 3 sample chapters, a table of contents, and a preface outlining the aims and scope of the book, how the book fits in with the current literature, and which courses the book is suitable for.

For any further information, please contact the Editor at Springer: francesca.bonadei@springer.com

THE SERIES IS INDEXED IN SCOPUS

UNITEXT is glad to announce a new series of free webinars and interviews handled by the Board members, who rotate in order to interview top experts in their field.

Access this link to subscribe to the events:

<https://cassyni.com/events/TPQ2UgkCbJvvz5QbkcWXo3>

Luigi Grippo • Marco Sciandrone

Introduction to Methods for Nonlinear Optimization

 Springer

Luigi Grippo
Rome, Italy

Marco Sciandrone
Department of Computer, Control and
Management Engineering
Sapienza University of Rome
Rome, Italy

ISSN 2038-5714

UNITEXT

ISSN 2038-5722

La Matematica per il 3+2

ISBN 978-3-031-26789-5

<https://doi.org/10.1007/978-3-031-26790-1>

ISSN 2532-3318 (electronic)

ISSN 2038-5757 (electronic)

ISBN 978-3-031-26790-1 (eBook)

© Springer Nature Switzerland AG 2023

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Optimization (or mathematical programming) concerns the study of “decision problems”, where the aim is to determine the minimum or the maximum points of a real function (the objective) in a prefixed set (the feasible set). Many real problems arising in the fields of economics, engineering, chemistry, physics, statistics, computer science, management science, operations research and mathematics itself can be formulated as optimization problems.

This book is an introduction to the important area of nonlinear optimization, known also as nonlinear programming, that studies optimization problems where some of the problem functions can be nonlinear. In particular, we restrict our study to continuous nonlinear optimization problems, defined on finite dimensional real spaces, under the assumption that the problem functions are smooth. Even with these restrictions, nonlinear programming has many relevant applications in the solution of nonlinear systems of equations and inequalities, in the construction and the analysis of mathematical models, in the solution of (almost all) engineering problems, in the computation of the control laws of discrete-time dynamical systems. At present, nonlinear optimization techniques represent also a basic ingredient of *machine learning* methods, for the computation of the parameters of the learning system on the basis of the available data.

The book contains three blocks of chapters concerning:

- Basic theory and optimality conditions (Chaps. 1–7)
- Unconstrained and constrained algorithms (Chaps. 8–23)
- Advanced topics (Chaps. 24–26).

The first two blocks consist of short chapters, and each of these chapters may correspond, in the authors’ experience, to the notes for one or two lessons (with each lesson of approximately 2 hours). Some more space is given to the matter of the third block. In each chapter, we add a few exercises and a terminal section on notes and references. As the main emphasis of the book is on solution methods, we believe that the best instructive exercise could be the realization of a computer code, related to the methods illustrated in the book, under the guidance of the teacher.

The basic concepts and results of linear algebra, of calculus on R^n and of convex analysis, are reported in Chaps. 27, 28, and 29. In particular, in Chap. 29 on convex analysis, we give also the proof of all the results used in the book, since, in many cases, this chapter must be included in a course. Then, the only prerequisite is the content of basic courses on linear algebra and calculus.

Most of the basic theoretical results reported in the text are formally proved in order to allow the students, potential readers of the book, to acquire a sound background in optimization. As regards “classical” optimization subjects, we attempted to give the simplest proofs available in the literature or adapted from the best known sources. Particular attention is devoted to the study of global convergence properties of the algorithms, since the definition of globalization techniques represents one of the major contributions of optimization to computational methods.

The book contains also more specialized topics, which are receiving an increasing interest in the last years, like those concerning derivative-free methods, non-monotone methods, decomposition algorithms, which have been, and currently are, research topics of the authors, and are not usually reported in introductory textbooks.

The material given in the book can be used for constructing different courses, both at an (advanced) undergraduate level and at a graduate level.

A short course (say of 30 hours) can be based on elements of convex analysis (Chap. 29), basic definitions and existence results (Chap. 2), optimality conditions (at least Chaps. 3, 4, and 5). This material can be combined, for instance, with a basic course on linear programming. Note, however, that linear programming is not a prerequisite and that the basic theory of linear programming can also be derived, as special case, from the optimality conditions of nonlinear programming.

A course of 60 hours for advanced undergraduate students could be based, for instance, on the same material indicated above, with the addition of of a selection of unconstrained algorithms that includes at least Chaps. 8, 9, 10, and 11.

Similarly, a course of 60 hours for graduated students could be focused on a selection of classes of methods presented in the text, both for unconstrained and constrained optimization problems. The choice of the arguments will depend on the type of master degree and on the background and scientific interests of the potential readers of the book.

Some important topics like, for instance, sparse optimization, incremental and online algorithms, complexity analysis, non-smooth optimization, global optimization, cone programming, semidefinite programming have been leaved out for space limitations. We believe, however, that the study of the material reported in the book can be a useful prerequisite for further extensions. The basic techniques introduced here can also be useful in the study of other sectors of decision sciences like variational inequalities, multi objective optimization and game theory.

Finally we remark that most of the material on unconstrained methods is extracted from the book (in Italian):

L. Grippo and M. Sciandrone, *Metodi di ottimizzazione non vincolata*, Springer-Verlag Italia, Milano, 2011, which contains also some of the proofs omitted here.

Contents

1	Introduction	1
1.1	Optimization Problems	1
1.2	Classification of Optimization Problems	3
1.3	Continuous Optimization Problems Over R^n	4
1.3.1	Classification	4
1.3.2	Examples of Continuous Problems	6
1.4	Appendix: Examples of Problems in Different Spaces	13
1.4.1	An Infinite-Dimensional Optimization Problem: Control of a Dynamical System	14
1.4.2	An Integer Linear Programming Problem: Scheduling of Sport Tournaments	14
1.4.3	A Mixed Integer Programming Problem: Clustering	17
1.5	Notes and References	19
2	Fundamental Definitions and Basic Existence Results	21
2.1	Local and Global Minimum Points	21
2.2	Minimizers in Convex Programming	24
2.3	Minimizers in Generalized Convex Problems	26
2.4	Minimizers in Concave Programming	29
2.5	Equivalent Transformations	30
2.5.1	Equivalence Criterion	30
2.5.2	Some Basic Examples	31
2.5.3	Fractional Programming	33
2.5.4	Goal Programming	34
2.6	Existence Conditions	36
2.6.1	Existence in Unconstrained Optimization	37
2.6.2	Existence in Constrained Optimization	44
2.6.3	Feasibility Problems	47
2.7	Exercises	49
2.8	Notes and References	51

3	Optimality Conditions for Unconstrained Problems in R^n	53
3.1	Descent Directions	53
3.2	Optimality Conditions	56
3.3	Optimality Conditions in the Convex Case	60
3.4	Exercises	66
3.5	Notes and References	67
4	Optimality Conditions for Problems with Convex Feasible Set	69
4.1	Optimality Conditions Based on Feasible Directions	69
4.2	The Case of Feasible Set Defined by Linear Constraints	71
4.2.1	Feasible Directions for Problems with Box Constraints	74
4.3	Optimality Conditions over a Convex Set	76
4.4	Projection on a Convex Set	80
4.5	Exercises	86
4.6	Notes and References	86
5	Optimality Conditions for Nonlinear Programming	87
5.1	Problem Formulation and Notation	87
5.2	Fritz John Conditions	88
5.3	Constraint Qualifications and KKT Conditions	93
5.4	Sufficient Conditions in the Convex Case	100
5.5	Second Order Optimality Conditions	101
5.6	Linearly Constrained Problems	109
5.6.1	Problems with Non Negativity Constraints	110
5.6.2	Problems with Box Constraints	111
5.6.3	Problems with Simplex Constraints	112
5.6.4	Quadratic Programming	113
5.6.5	Linear Programming	114
5.7	Exercises	115
5.8	Notes and References	117
6	Duality Theory	119
6.1	The Dual Problem	119
6.2	Lagrangian Duality	121
6.2.1	Basic Concepts, Weak and Strong Lagrangian Duality	121
6.2.2	Saddle Point Optimality Conditions	125
6.3	Wolfe's Dual	126
6.3.1	Wolfe's Dual in Quadratic Convex Programming Problems with Linear Inequality Constraints	130
6.4	Appendix	132
6.5	Exercises	134
6.6	Notes and References	135

- 7 Optimality Conditions Based on Theorems of the Alternative** 137
 - 7.1 Introduction 137
 - 7.2 Optimality Conditions for Problems with Inequality Constraints 138
 - 7.3 Optimality Conditions Based on Tangent Directions 141
 - 7.4 Optimality Conditions for Problems with Equality and Inequality Constraints 142
 - 7.5 Fritz John Conditions Established with Motzkin Theorem 147
 - 7.5.1 KKT Conditions from FJ Conditions 148
 - 7.6 Appendix 150
 - 7.6.1 Linear Inequalities and Theorems of the Alternative 150
 - 7.6.2 Non Homogeneous Farkas Lemma and Optimality for LP 162
 - 7.7 Notes and References 165
- 8 Basic Concepts on Optimization Algorithms** 167
 - 8.1 Structure of Optimization Algorithms 167
 - 8.2 Existence and Uniqueness of Limit Points 169
 - 8.3 Convergence Towards Critical Points 170
 - 8.4 Local and Global Convergence 171
 - 8.5 Convergence Rate 172
 - 8.6 Basic Concepts of Complexity Analysis 174
 - 8.7 Notes and References 175
- 9 Unconstrained Optimization Algorithms** 177
 - 9.1 Preliminary Concepts 177
 - 9.2 Limit Points 178
 - 9.3 Convergence Towards Stationary Points 179
 - 9.4 Classification of Unconstrained Algorithms 180
 - 9.5 Convergence of Line Search Based Methods 181
 - 9.6 Exercises 185
 - 9.7 Notes and References 186
- 10 Line Search Methods** 187
 - 10.1 Basic Features and Classification of Line Searches 187
 - 10.2 Armijo’s Method 191
 - 10.3 Step-Length Extension and Goldstein Conditions 198
 - 10.4 Wolfe Conditions 201
 - 10.5 Derivative-Free Line Searches 204
 - 10.5.1 Backtracking Armijo-Type Derivative-Free Algorithms 205
 - 10.5.2 Derivative-Free Linesearches with Step Extension 208
 - 10.6 Appendix A: Algorithms Employing Wolfe Conditions 213
 - 10.7 Appendix B: Implementation of Line Searches 216
 - 10.7.1 Initial Interval 217
 - 10.7.2 Initial Estimate of the Step-Size 217

10.7.3	Interpolation Formulas	219
10.7.4	Application of Interpolation Formulas	224
10.7.5	Stopping Criteria and Failures	225
10.8	Exercises	226
10.9	Notes and References	227
11	Gradient Method	229
11.1	The Steepest Descent Direction	229
11.2	The Gradient Method	230
11.3	Gradient Method with Constant Step-Size	232
11.4	Convergence Rate	234
11.5	Finite Convergence in the Quadratic Case	238
11.6	Complexity of the Steepest Descent Method	239
11.7	Modified Gradient Methods	243
11.8	Exercises	246
11.9	Notes and References	247
12	Conjugate Direction Methods	249
12.1	The Conjugate Direction Method	249
12.2	The Conjugate Gradient Method (CGM): The Quadratic Case ...	252
12.3	Convergence Rate	258
12.4	Preconditioning	259
12.5	The CGM in the Non Quadratic Case	262
12.6	Fletcher-Reeves Method	265
12.7	Method of Polyak-Polak-Ribiére (PPR)	266
12.8	Appendix: The CG Method When the Hessian Matrix is Positive Semidefinite	271
12.9	Exercises	273
12.10	Notes and References	273
13	Newton's Method	275
13.1	The Pure Newton Iteration	275
13.2	Local Convergence and Convergence Rate	277
13.3	Shamanskii Method	280
13.4	Globalization of Newton's Method for Minimization	283
13.5	Hybrid Methods	289
13.6	Modification Methods	291
13.7	Exercises	294
13.8	Notes and References	295
14	Trust Region Methods	297
14.1	The General Trust Region Framework and Convergence Results	297
14.2	Classification of Solution Methods for the Trust Region Subproblem	301
14.3	The Cauchy Step-Based Method	302
14.4	The Dogleg Method	304

- 14.5 The Conjugate Gradient Method of Steihaug 306
- 14.6 Necessary and Sufficient Optimality Conditions for the Trust Region Subproblem 308
- 14.7 Trust Region Approach to Globalizing Newton’s Method 311
- 14.8 Complexity Issues and Adaptive Regularized Methods 316
- 14.9 Appendix: Proofs of Convergence 318
- 14.10 Exercises 323
- 14.11 Notes and References 324
- 15 Quasi-Newton Methods 325**
 - 15.1 Preliminaries 325
 - 15.2 Rank 1 Formulae 328
 - 15.3 Rank Two Formulae and the BFGS Method 329
 - 15.4 Global Convergence of the BFGS Method 334
 - 15.5 Convergence Rate 335
 - 15.6 Exercises 336
 - 15.7 Notes and References 336
- 16 Methods for Nonlinear Equations 337**
 - 16.1 Preliminaries 337
 - 16.2 Newton-Type Methods 339
 - 16.3 Broyden’s Method 342
 - 16.4 Residual Based Methods 344
 - 16.5 Notes and References 344
- 17 Methods for Least Squares Problems 345**
 - 17.1 Least Squares Problems 345
 - 17.2 Linear Least Squares Problems 347
 - 17.3 Methods for Nonlinear Least Squares Problems 349
 - 17.4 Gauss-Newton Method 350
 - 17.5 Levenberg-Marquardt Method 355
 - 17.6 Recursive Linear Least Squares Algorithm 358
 - 17.7 Some Notes on Incremental Methods for Nonlinear Problems ... 360
 - 17.8 Exercises 362
 - 17.9 Notes and References 363
- 18 Methods for Large-Scale Optimization 365**
 - 18.1 Solution Strategies for Large Scale Problems 365
 - 18.2 Truncated Newton Method 366
 - 18.3 Globally Convergent Truncated Newton Methods 367
 - 18.3.1 Preliminaries 367
 - 18.3.2 A Truncated Newton Method Based on a Line Search 368
 - 18.4 Quasi-Newton Methods for Large-Scale Optimization 375
 - 18.4.1 Preliminaries 375
 - 18.4.2 Memoryless Quasi-Newton Methods 376
 - 18.4.3 Limited-Memory Quasi-Newton Methods 376

18.5	Exercises	380
18.6	Notes and References	380
19	Derivative-Free Methods for Unconstrained Optimization	383
19.1	Motivation and Classification of Derivative-Free Methods	383
19.2	The Coordinate Method	385
	19.2.1 Preliminary Concepts	385
	19.2.2 Coordinate Method with Simple Decrease	386
19.3	The Hooke-Jeeves Method	392
19.4	The Nelder-Mead Method	394
19.5	Linesearch-Based Methods	396
	19.5.1 Basic Assumptions and Convergence Conditions	398
	19.5.2 Globalization of Direct Search Methods Through Line Searches	403
19.6	Approximation of Derivatives and Implicit Filtering	405
	19.6.1 Simplex Gradient	405
	19.6.2 Implicit Filtering	408
	19.6.3 Combination with Coordinate Search	409
19.7	Model-Based Methods	409
19.8	Exercises	411
19.9	Notes and References	411
20	Methods for Problems with Convex Feasible Set	413
20.1	Problems with Convex Feasible Set	413
20.2	Line Search Along a Feasible Direction	415
20.3	The Frank-Wolfe Method (Conditional Gradient Method)	419
20.4	Gradient Projection Method	421
20.5	A Derivative-Free Method for Box Constrained Problems	425
20.6	Concave Programming for Minimizing the Zero-Norm Over Polyhedral Sets	434
20.7	Frank-Wolfe Method Applied to Concave Programming Problems	437
20.8	Exercises	439
20.9	Notes and References	439
21	Penalty and Augmented Lagrangian Methods	441
21.1	Basic Concepts	441
21.2	Sequential Penalty Functions	444
21.3	Augmented Lagrangian Functions	452
21.4	Non Differentiable Exact Penalty Functions	459
21.5	Continuously Differentiable Exact Penalty Functions	459
21.6	Exact Shifted Barrier Functions	462
21.7	Exact Augmented Lagrangian Functions	471
21.8	Exercises	478
21.9	Notes and References	478

- 22 SQP Methods** 481
 - 22.1 Newton Type Methods for the Equality Constrained Case 481
 - 22.2 Extension to Inequality Constrained Problems 484
 - 22.3 EQP Methods 485
 - 22.4 Quasi-Newton Modification 488
 - 22.5 Solution of the Quadratic Programming Subproblem 489
 - 22.6 Globalization Strategies 490
 - 22.6.1 Nondifferentiable Merit Functions 491
 - 22.6.2 Smooth Merit Functions 494
 - 22.7 Notes and References 496
- 23 Introduction to Interior Point Methods** 497
 - 23.1 Basic Definitions and Barrier Methods 497
 - 23.2 Interior Point Methods for Linear Programming 503
 - 23.2.1 Definitions and Notation 503
 - 23.2.2 A Summary of Basic LP Theory 504
 - 23.2.3 Main Classes of IPMs for LP and Basic Assumptions 505
 - 23.2.4 Feasible Path-Following Methods 509
 - 23.2.5 Infeasible Methods 519
 - 23.3 Extensions 521
 - 23.3.1 Quadratic Programming 521
 - 23.3.2 IPMs for Nonconvex Problems 522
 - 23.4 Appendix: Regularity and Optimality 524
 - 23.5 Notes and References 526
- 24 Nonmonotone Methods** 529
 - 24.1 Motivation and Basic Concepts 529
 - 24.2 Convergence Issues in Nonmonotone Methods 531
 - 24.3 Reference Values and Preliminary Results 532
 - 24.4 Armijo-Type Nonmonotone Line Searches 538
 - 24.5 Nonmonotone Armijo-Goldstein and Parabolic Searches 541
 - 24.6 Nonmonotone Derivative-Free Linesearches 546
 - 24.7 Watchdog Techniques 552
 - 24.8 Nonmonotone Globalization of Newton’s Method 557
 - 24.9 Nonmonotone Inexact Newton-Type Methods for Nonlinear Equations 560
 - 24.10 Notes and References 570
- 25 Spectral Gradient Methods** 573
 - 25.1 The BB Method in the Quadratic Case 573
 - 25.2 Nonmonotone Globalization of the BB Method 578
 - 25.3 Spectral Gradient Methods for Nonlinear Equations 580
 - 25.4 Projected Spectral Gradient Method for Minimization on Convex Sets 585

25.5	Exercises	586
25.6	Notes and References	586
26	Decomposition Methods	589
26.1	Motivation and Examples	589
26.2	Classes of Decomposition Methods and Basic Notation	592
26.3	Block Gauss-Seidel (GS) Method and Extensions	595
26.3.1	The Basic Scheme of the GS Method	595
26.3.2	A Line Search Algorithm in the Component Space	596
26.3.3	Limit Points of the GS Algorithm	599
26.3.4	The Two-Block GS Method	601
26.3.5	Convergence of the GS Method Under Generalized Convexity Assumptions	602
26.3.6	Proximal-Point Modification of the GS Method	607
26.4	Block Descent Methods	609
26.4.1	A Basic Unconstrained Block Descent Algorithm	609
26.4.2	A Basic Constrained Block Descent Algorithm	612
26.5	The Gauss-Southwell Algorithm	615
26.6	Decomposition with Variable and Overlapping Blocks	619
26.7	The Jacobi Method	624
26.8	Algorithms for Problems with a Single Linear Equality Constraint and Box Constraints	626
26.9	Sequential Minimal Optimization (SMO) Algorithms	627
26.10	Appendix: Proof of Proposition 26.16	633
26.11	Exercises	637
26.12	Notes and References	637
27	Basic Concepts of Linear Algebra and Analysis	639
27.1	The Space R^n as Linear Space	639
27.2	Matrices and Systems of Linear Equalities	641
27.3	Norm, Metric, Topology, Scalar Product Over R^n	644
27.4	Notation and Results on Real Matrices	651
27.4.1	Determinant, Minors, Trace	651
27.4.2	Norm of a Matrix	652
27.4.3	Orthogonal Matrices	654
27.4.4	Eigenvalues of Symmetric Matrices	654
27.4.5	Singular Value Decomposition	656
27.5	Quadratic Forms	658
28	Differentiation in R^n	663
28.1	First Order Derivatives of a Real Function	663
28.2	Differentiation of a Vector Function	665
28.3	Second Order Derivatives of a Real Function	667
28.4	The Mean Value Theorem and Taylor's Theorem	669
28.5	Derivatives of Composite Functions	671
28.6	Examples	672

29 Introduction to Convex Analysis	681
29.1 Convex Sets	681
29.2 Convex Functions	689
29.3 Composition of Convex Functions	692
29.4 Convexity of Differentiable Functions	695
29.5 Monotonicity Conditions on ∇f	700
29.6 Basic Notions of Generalized Convexity	703
References	709
Index	719