

Springer Undergraduate Texts in Mathematics and Technology

Series editors

H. Holden, Norwegian University of Science and Technology,
Trondheim, Norway

Keri A. Kornelson, University of Oklahoma, Norman, OK, USA

Editorial Board

Lisa Goldberg, University of California, Berkeley, CA, USA

Armin Iske, University of Hamburg, Germany

Palle E. T. Jorgensen, The University of Iowa, Iowa City, IA, USA

Springer Undergraduate Texts in Mathematics and Technology (SUMAT) publishes textbooks aimed primarily at the undergraduate. Each text is designed principally for students who are considering careers either in the mathematical sciences or in technology-based areas such as engineering, finance, information technology and computer science, bioscience and medicine, optimization or industry. Texts aim to be accessible introductions to a wide range of core mathematical disciplines and their practical, real-world applications; and are fashioned both for course use and for independent study.

More information about this series at <http://www.springer.com/series/7438>

Francisco J. Aragón · Miguel A. Goberna ·
Marco A. López · Margarita M. L. Rodríguez

Nonlinear Optimization

Francisco J. Aragón
Department of Mathematics
University of Alicante
Alicante, Spain

Miguel A. Goberna
Department of Mathematics
University of Alicante
Alicante, Spain

Marco A. López
Department of Mathematics
University of Alicante
Alicante, Spain

Margarita M. L. Rodríguez
Department of Mathematics
University of Alicante
Alicante, Spain

ISSN 1867-5506 ISSN 1867-5514 (electronic)
Springer Undergraduate Texts in Mathematics and Technology
ISBN 978-3-030-11183-0 ISBN 978-3-030-11184-7 (eBook)
<https://doi.org/10.1007/978-3-030-11184-7>

Library of Congress Control Number: 2018966836

Mathematics Subject Classification (2010): 65kxx, 97uxx, 47N10

© Springer Nature Switzerland AG 2019, corrected publication 2019

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, express 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

This book is aimed at upper-level undergraduate students of mathematics and statistics, and graduate students of industrial engineering. We assume that the readers are familiar with real analysis and linear algebra (subjects that are briefly revisited in Chapter 1), and we decline to use sophisticated tools from convex analysis (such as the Rockafellar subdifferential or the Fenchel–Moreau conjugate function) and nonsmooth analysis (such as the Clarke and Mordukhovich subdifferentials). It is not easy to select, among the vast amount of knowledge on the subject, the suitable contents for a textbook on nonlinear optimization. We have included in this book some basic theoretical results (in particular, optimality and duality theorems for different problems) and a representative selection of the many available numerical methods for problems with or without constraints, some of them not using derivatives. Both theoretical results and numerical methods are illustrated with applications to a diverse range of fields.

The contents are organized in two parts that can be used as textbooks for undergraduate courses on convex and nonconvex optimization, respectively. In fact, this book is based on the contents of two undergraduate courses taught at the University of Alicante. These two courses are preceded by a course on linear programming, which is based on the book [42]. Nevertheless, this knowledge is not assumed in the present book. With the exception of Chapter 1, which is preliminary for both parts, Part I and Part II are essentially independent. The sections and subsections marked with an asterisk (*) are mathematically harder than the rest and can be skipped when either the students have a limited background or when the available time is insufficient.

Part I, composed of Chapters 2–4, focuses on the analytic calculation of global minima (required in engineering, operations research, statistics, and economy) and local minima (required in natural sciences, where the equilibrium situations correspond to local minima—or even critical points—of certain functions). Chapters 1 and 2 contain the basic ingredients for the calculus of local minima (optimality conditions for differentiable functions) and global minima (coercivity and convexity) for unconstrained optimization problems. Chapter 3 provides closed formulas for unconstrained optimization problems arising in different fields, while

Chapter 4 deals with unconstrained and constrained convex optimization problems for which local and global minima coincide. For the sake of simplicity, the sections devoted to constrained continuous optimization preclude linear equations and make use of a unique constraint qualification (the Slater condition).

Part II is focused on the numerical calculation of local optima in problems whose solutions cannot be analytically obtained, and it consists of two chapters. Chapter 5 deals with the standard algorithms for unconstrained optimization problems, such as the steepest descent method, Newton's method and variants (trust regions, Gauss–Newton, Levenberg–Marquardt), and other gradient-based methods such as those using conjugate directions (conjugate gradient and quasi-Newton). The local or global convergence of such methods is discussed in detail, as well as the respective rates of convergence (linear, superlinear, and quadratic). This chapter also includes an introduction to methods that do not use derivatives. Finally, Chapter 6 presents, in the first part, an introduction to the so-called penalty and barrier methods. These procedures address the numerical solution of constrained optimization problems by applying the algorithms described in Chapter 5, and they are based on the idea of converting a constrained problem in a sequence of unconstrained ones. The second part of this final chapter is devoted to the optimality conditions for constrained optimization problems, with equality and/or inequality restrictions. The relevance of the so-called constraint qualifications is emphasized as they yield necessary optimality conditions, which can be used as stopping rules for the algorithms. Finally, we show how the optimality conditions give rise to the sequential quadratic programming methods, when Newton's method is applied for solving the associated system of equations.

Definitions, proofs, and numerical methods are illustrated with figures. Moreover, all chapters contain collections of exercises, and the solutions of those exercises that our students usually find harder are provided in a separate appendix at the end of the book.

This is a book on nonlinear optimization for undergraduate students, so it is not intended to provide an exhaustive overview of the available optimality theory and optimization methods. Readers wishing to get a deeper knowledge of the topics included in this book are kindly invited to consult the list of references provided at the end, especially the books [5, 8, 11, 28, 42, 52, 54, 67, 70, 75], whose treatment of different subjects has inspired the approach adopted in some sections. Readers interested in the history of optimization, or in its key role in operations research, may find sources of information in the books [39, 41, 47, 62].

The top five reasons to use this textbook are:

- It only assumes basic knowledge of differential and matrix calculus. All the main concepts, algorithms, and proofs are illustrated with highly explanatory and appealing figures.
- It pays special attention to model building and validation of real problems, and emphasizes the practical advantages of obtaining good reformulations of optimization problems. It shows important applications of optimization models to natural and social sciences, engineering, and data science.

- It provides rigorous optimality conditions. The existence and uniqueness of optimal solutions is analyzed via coercivity and convexity. Dual problems are introduced in order to get lower bounds, sensitivity information (“what if” questions), and stopping rules for primal-dual algorithms.
- It provides an accurate description of the main numerical approaches to solve nonlinear optimization problems. These numerical methods have been chosen with the aim of covering an ample range of techniques. Intentionally, the algorithms described are not implemented in a specific software, as experience shows that this type of choice is highly ephemeral and mainly depends on personal preferences. Nonetheless, five assignments for laboratory sessions of two hours have been included in Section 5.8.
- It has been thoroughly tested. Its preliminary versions have been used for many years as class notes for various undergraduate courses on optimization theory and methods. The exercises have been carefully selected to push the students to a deeper understanding of the main topics in each chapter. Further, a detailed solution to those exercises that our students consider harder is provided, to allow independent study.

We would like to thank our families for their patience and constant support. We are grateful to the anonymous referees and to our editors at Springer, Elizabeth Loew and Razia Amzad, whose suggestions helped us improve this book. We are indebted to Bernardo Cascales, who proposed us to write the initial version of this book in Spanish, and to Jonathan Borwein, who encouraged us to translate the manuscript into English. Our gratitude also goes to all the students of the Mathematics Degree at the University of Alicante who gave us critical remarks on the preliminary versions of this book. Finally, we are thankful to the funding agencies MINECO (Spain) and ERDF (EU) for their financial support along the years. In particular, this book was written as part of the outreach activities of the grant MTM2014-59179-C2-1-P and the research contract RYC-2013-13327 of the first author.

Alicante, Spain
November 2018

Francisco J. Aragón
Miguel A. Goberna
Marco A. López
Margarita M. L. Rodríguez

The original version of this book was inadvertently published with incorrect reference numbers in the frontmatter. The corrections to this book can be found at https://doi.org/10.1007/978-3-030-11184-7_7

Contents

1 Preliminaries	1
1.1 Optimization Models	1
1.1.1 A Brief History of Optimization	1
1.1.2 Building Up Optimization Models	4
1.1.3 Planning Blending Operations	4
1.1.4 Optimization Problems	7
1.1.5 Reformulating Optimization Problems	10
1.2 Basic Calculus Tools	17
1.2.1 Complements of Differential Calculus	17
1.2.2 Complements of Matrix Calculus	27
1.3 Optimality Conditions at Interior Points	33
1.4 Coercivity	38
1.4.1 Coercive Functions	39
1.4.2 The Fundamental Theorem of Algebra [★]	45
1.5 Exercises	48
Part I Analytical Optimization	
2 Convexity	55
2.1 Convex Sets	55
2.2 Convex Functions of One Variable	62
2.3 Convex Functions of Several Variables	72
2.4 Special Convex Functions	82
2.4.1 Strictly Convex Functions	82
2.4.2 Strongly Convex Functions	83
2.5 Exercises	86
3 Unconstrained Optimization	91
3.1 Unconstrained Quadratic Optimization	92
3.2 Least Squares Regression	94

3.2.1	Linear Regression	95
3.2.2	Polynomial Regression	98
3.3	Least Squares Solutions to Overdetermined Linear Systems	101
3.4	Optimizing Without Derivatives	103
3.4.1	Jensen's Inequalities	103
3.4.2	The Geometric-Arithmetic Inequality	107
3.4.3	Duality in Unconstrained Geometric Optimization*	109
3.5	Exercises	113
4	Convex Optimization	117
4.1	The Optimal Set	118
4.2	Unconstrained Convex Optimization	122
4.2.1	The Fermat–Steiner Problem	122
4.2.2	The Fixed-Point Method of Weiszfeld	130
4.2.3	An Application to Statistical Inference	131
4.3	Linearly Constrained Convex Optimization	134
4.3.1	Optimality Conditions	134
4.3.2	Quadratic Optimization	141
4.3.3	Some Closed Formulas	150
4.4	Arbitrarily Constrained Convex Optimization*	153
4.4.1	Sensitivity Analysis	153
4.4.2	Optimality Conditions	161
4.4.3	Lagrange Duality	168
4.4.4	Wolfe Duality	170
4.5	A Glimpse on Conic Optimization*	172
4.6	Exercises	176
 Part II Numerical Optimization		
5	Unconstrained Optimization Algorithms	183
5.1	Line Search Methods	183
5.1.1	The Family of Gradient Methods	186
5.1.2	The Step size	186
5.2	Convergence of the Line Search Methods	193
5.2.1	Rate of Convergence	197
5.2.2	Convergence Rate of Gradient Methods for Quadratic Forms	201
5.2.3	Gradient Algorithms	207
5.3	Newton's Method	209
5.3.1	Trust Region Methods	212
5.3.2	Least Squares Problems	214
5.4	Newton's Method for Solving Equations	216

- 5.5 Conjugate Direction Methods 218
 - 5.5.1 Conjugate Gradient Method 222
 - 5.5.2 Quasi-Newton Methods 226
- 5.6 Derivative-Free Optimization Methods 232
 - 5.6.1 Coordinate Descent Algorithms 233
 - 5.6.2 Nelder and Mead Method 234
 - 5.6.3 Directional Direct Search Methods Using Positive Spanning Sets 236
- 5.7 Exercises 241
- 5.8 Computer Exercises 246
 - 5.8.1 Assignment 1 246
 - 5.8.2 Assignment 2 248
 - 5.8.3 Assignment 3 249
 - 5.8.4 Assignment 4 250
 - 5.8.5 Assignment 5 251
- 6 Constrained Optimization 253**
 - 6.1 Penalty and Barrier Methods 253
 - 6.1.1 Penalty Methods 254
 - 6.1.2 Methods Using Exterior Penalty Functions 256
 - 6.1.3 Barrier Methods 260
 - 6.1.4 A Logarithmic Barrier Approach to Linear Programming 264
 - 6.2 Problems with Equality Constraints 266
 - 6.3 Problems with Inequality Constraints 276
 - 6.3.1 Karush–Kuhn–Tucker Optimality Conditions 276
 - 6.3.2 Other Constraint Qualifications[★] 281
 - 6.3.3 Fritz John Optimality Conditions[★] 288
 - 6.4 Problems with Equality and Inequality Constraints 289
 - 6.4.1 Second-Order Optimality Conditions[★] 292
 - 6.5 Sequential Quadratic Programming Methods[★] 297
 - 6.6 Concluding Remarks[★] 302
 - 6.7 Exercises 305
- Correction to: Nonlinear Optimization C1**
- Solutions to Selected Exercises 311**
- References 343**
- Index 347**

Nomenclature

$]a, b[$	Open interval between the real numbers a and b
$[a, b]$	Closed interval between the real numbers a and b
\mathbb{R}_+	Set of nonnegative real numbers
\mathbb{R}_{++}	Set of positive real numbers
$\overline{\mathbb{R}}$	Extended real line, page 154
$\min X$	Minimum of $X \subset \mathbb{R}$
$\max X$	Maximum of $X \subset \mathbb{R}$
$\inf X$	Infimum of $X \subset \mathbb{R}$
$\sup X$	Supremum of $X \subset \mathbb{R}$
I_n	$n \times n$ identity matrix
A_k	$k \times k$ submatrix of A which results of taking the first k rows and columns
Δ_k	k th director principal minor of A
A^T	Transpose matrix of A
A^{-1}	Inverse matrix of A
$\det A$	Determinant of the matrix A
$\text{diag}(\lambda_1, \dots, \lambda_n)$	Diagonal matrix with diagonal elements $\lambda_1, \dots, \lambda_n$
$\rho(A)$	Spectral radius of the matrix A , page 31
$\text{cond}(A)$	Condition number of the matrix A , page 31
\mathcal{S}_n	Family of symmetric $n \times n$ real matrices
\mathcal{S}_q^+	Cone of positive semidefinite symmetric matrices
$0_{n \times n}$	$n \times n$ null matrix
0_n	Zero vector of \mathbb{R}^n
e_i	i th vector of the standard basis
1_n	Vector of all ones in \mathbb{R}^n
$\ x\ _1$	Manhattan (or ℓ_1) norm of x , page 15
$\ x\ _\infty$	Chebyshev (or ℓ_∞) norm of x , page 15
$\ x\ $	Euclidean (or ℓ_2) norm of x , page 15
\mathbb{B}	Euclidean closed unit ball
$x^T y$	Scalar product of the vectors x and y

$\{x_k\}$	Sequence in \mathbb{R}^n
\limsup	Upper limit of a sequence
$ X $	Cardinality of X
$\text{int } X$	Topological interior of X
$\text{cl } X$	Closure of X
$\text{bd } X$	Boundary of X
$\text{span } X$	Linear hull of X
$\text{conv } X$	Convex hull of X , page 55
$\text{cone } X$	Conical (convex) hull of X , page 58
K°	(Negative) polar cone of the cone K , page 135
K_p^m	Second-order cone, page 172
$f'(x)$	Derivative of f at x
$\frac{\partial f(x)}{\partial x_i}$	Partial derivative of f with respect to x_i at x
$\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$	Second-order partial derivative of f with respect to x_i and to x_j at x
$f'(x; u)$	Directional derivative of f at x in the direction of u
$f'_+(a)$	Right derivative of f at a , page 62
$f'_-(a)$	Left derivative of f at a , page 62
$\nabla f(x)$	Gradient of f at x , page 19
$\nabla^2 f(x)$	Gradient matrix of f at x , page 26
$J_f(x)$	Hessian matrix of f at x , page 24
$\text{dom } f$	Jacobian matrix of f at x , page 26
$\text{gph } f$	Domain of the function f , pages 17 and 154
$\text{epi } f$	Graph of the function f , page 17
$S_\lambda(f)$	Epigraph of the function f , page 72
$\text{argmin } f$	Sublevel set λ of f , page 38
$o(g(x))$	Set of points in which f reaches its minimum value
v_+	Landau notation, page 24
p_+	Positive part of the vector v , page 120
\mathcal{F}	Positive part function, page 121
ϑ	Feasible set multifunction, page 154
$\mathcal{C}(X)$	Value function, page 154
$\mathcal{C}^m(X)$	Linear space of continuous real-valued functions on X
F	Linear space of real functions on X with continuous partial derivatives of order m
F^*	Feasible set of the problem P , page 8
$v(P)$	Optimal set of the problem P , page 8
$I(\bar{x})$	Optimal value of the problem P , page 8
$A(\bar{x})$	Set of active indices at \bar{x} , page 134
$D(\bar{x})$	Active cone at \bar{x} , page 135
	Feasible direction cone at \bar{x} , page 134