Convex Optimization Algorithms

Dimitri P. Bertsekas

Massachusetts Institute of Technology

WWW site for book information and orders

http://www.athenasc.com

Athena Scientific, Belmont, Massachusetts

Athena Scientific
Post Office Box 805
Nashua, NH 03061-0805
U.S.A.

Email: info@athenasc.com
WWW: http://www.athenasc.com

Publisher's Cataloging-in-Publication Data

Bertsekas, Dimitri P.
Convex Optimization Algorithms
Includes bibliographical references and index
1. Nonlinear Programming 2. Mathematical Optimization. I. Title.
T57.8.B475 2015 519.703
Library of Congress Control Number: 2002092168

ISBN-10: 1-886529-28-0, ISBN-13: 978-1-886529-28-1

Contents

ATHENA SCIENTIFIC

OPTIMIZATION AND COMPUTATION SERIES

1. Convex Optimization Algorithms, by Dimitri P. Bertsekas, 2015, ISBN 978-1-886529-28-1, 576 pages
2. Abstract Dynamic Programming, by Dimitri P. Bertsekas, 2013, ISBN 978-1-886529-42-7, 256 pages
3. Dynamic Programming and Optimal Control, Two-Volume Set, by Dimitri P. Bertsekas, 2012, ISBN 1-886529-08-6, 1020 pages
4. Convex Optimization Theory, by Dimitri P. Bertsekas, 2009, ISBN 978-1-886529-31-1, 256 pages
5. Introduction to Probability, 2nd Edition, by Dimitri P. Bertsekas and John N. Tsitsiklis, 2008, ISBN 978-1-886529-23-6, 544 pages
6. Convex Analysis and Optimization, by Dimitri P. Bertsekas, Angelia Nedić, and Asuman E. Ozdaglar, 2003, ISBN 1-886529-45-0, 560 pages
7. Nonlinear Programming, 2nd Edition, by Dimitri P. Bertsekas, 1999, ISBN 1-886529-00-0, 791 pages
8. Network Optimization: Continuous and Discrete Models, by Dimitri P. Bertsekas, 1998, ISBN 1-886529-02-7, 608 pages
9. Network Flows and Monotropic Optimization, by R. Tyrrell Rockafellar, 1998, ISBN 1-886529-06-X, 634 pages
10. Introduction to Linear Optimization, by Dimitris Bertsimas and John N. Tsitsiklis, 1997, ISBN 1-886529-19-1, 608 pages
11. Parallel and Distributed Computation: Numerical Methods, by Dimitri P. Bertsekas and John N. Tsitsiklis, 1997, ISBN 1-886529-01-9, 718 pages
12. Neuro-Dynamic Programming, by Dimitri P. Bertsekas and John N. Tsitsiklis, 1996, ISBN 1-886529-10-8, 512 pages
13. Constrained Optimization and Lagrange Multiplier Methods, by Dimitri P. Bertsekas, 1996, ISBN 1-886529-04-3, 410 pages
14. Stochastic Optimal Control: The Discrete-Time Case, by Dimitri P. Bertsekas and Steven E. Shreve, 1996, ISBN 1-886529-03-5, 330 pages

ABOUT THE AUTHOR

Dimitri Bertsekas studied Mechanical and Electrical Engineering at the National Technical University of Athens, Greece, and obtained his Ph.D. in system science from the Massachusetts Institute of Technology. He has held faculty positions with the Engineering-Economic Systems Department, Stanford University, and the Electrical Engineering Department of the University of Illinois, Urbana. Since 1979 he has been teaching at the Electrical Engineering and Computer Science Department of the Massachusetts Institute of Technology (M.I.T.), where he is currently the McAfee Professor of Engineering.

His teaching and research spans several fields, including deterministic optimization, dynamic programming and stochastic control, large-scale and distributed computation, and data communication networks. He has authored or coauthored numerous research papers and sixteen books, several of which are currently used as textbooks in MIT classes, including "Nonlinear Programming," "Dynamic Programming and Optimal Control," "Data Networks," "Introduction to Probability," "Convex Optimization Theory," as well as the present book. He often consults with private industry and has held editorial positions in several journals.

Professor Bertsekas was awarded the INFORMS 1997 Prize for Research Excellence in the Interface Between Operations Research and Computer Science for his book "Neuro-Dynamic Programming" (co-authored with John Tsitsiklis), the 2001 AACC John R. Ragazzini Education Award, the 2009 INFORMS Expository Writing Award, the 2014 AACC Richard Bellman Heritage Award for "contributions to the foundations of deterministic and stochastic optimization-based methods in systems and control," and the 2014 Khachiyan Prize for "life-time accomplishments in optimization." In 2001, he was elected to the United States National Academy of Engineering for "pioneering contributions to fundamental research, practice and education of optimization/control theory, and especially its application to data communication networks."

Preface

There is no royal way to geometry
(Euclid to king Ptolemy of Alexandria)

Interest in convex optimization has become intense due to widespread applications in fields such as large-scale resource allocation, signal processing, and machine learning. This book aims at an up-to-date and accessible development of algorithms for solving convex optimization problems.

The book complements the author's 2009 "Convex Optimization Theory" book, but can be read independently. The latter book focuses on convexity theory and optimization duality, while the present book focuses on algorithmic issues. The two books share mathematical prerequisites, notation, and style, and together cover the entire finite-dimensional convex optimization field. Both books rely on rigorous mathematical analysis, but also aim at an intuitive exposition that makes use of visualization where possible. This is facilitated by the extensive use of analytical and algorithmic concepts of duality, which by nature lend themselves to geometrical interpretation.

To enhance readability, the statements of definitions and results of the "theory book" are reproduced without proofs in Appendix B. Moreover, some of the theory needed for the present book, has been replicated and/or adapted to its algorithmic nature. For example the theory of subgradients for real-valued convex functions is fully developed in Chapter 3. Thus the reader who is already familiar with the analytical foundations of convex optimization need not consult the "theory book" except for the purpose of studying the proofs of some specific results.

The book covers almost all the major classes of convex optimization algorithms. Principal among these are gradient, subgradient, polyhedral approximation, proximal, and interior point methods. Most of these methods rely on convexity (but not necessarily differentiability) in the cost and constraint functions, and are often connected in various ways to duality. I have provided numerous examples describing in detail applications to specially structured problems. The reader may also find a wealth of analysis and discussion of applications in books on large-scale convex optimization, network optimization, parallel and distributed computation, signal processing, and machine learning.

The chapter-by-chapter description of the book follows:

Chapter 1: Here we provide a broad overview of some important classes of convex optimization problems, and their principal characteristics. Several problem structures are discussed, often arising from Lagrange duality theory and Fenchel duality theory, together with its special case, conic duality. Some additional structures involving a large number of additive terms in the cost, or a large number of constraints are also discussed, together with their applications in machine learning and large-scale resource allocation.

Chapter 2: Here we provide an overview of algorithmic approaches, focusing primarily on algorithms for differentiable optimization, and we discuss their differences from their nondifferentiable convex optimization counterparts. We also highlight the main ideas of the two principal algorithmic approaches of this book, iterative descent and approximation, and we illustrate their application with specific algorithms, reserving detailed analysis for subsequent chapters.

Chapter 3: Here we discuss subgradient methods for minimizing a convex cost function over a convex constraint set. The cost function may be nondifferentiable, as is often the case in the context of duality and machine learning applications. These methods are based on the idea of reduction of distance to the optimal set, and include variations aimed at algorithmic efficiency, such as ϵ-subgradient and incremental subgradient methods.

Chapter 4: Here we discuss polyhedral approximation methods for minimizing a convex function over a convex constraint set. The two main approaches here are outer linearization (also called the cutting plane approach) and inner linearization (also called the simplicial decomposition approach). We show how these two approaches are intimately connected by conjugacy and duality, and we generalize our framework for polyhedral approximation to the case where the cost function is a sum of two or more convex component functions.

Chapter 5: Here we focus on proximal algorithms for minimizing a convex function over a convex constraint set. At each iteration of the basic proximal method, we solve an approximation to the original problem. However, unlike the preceding chapter, the approximation is not polyhedral, but rather it is based on quadratic regularization, i.e., adding a quadratic term to the cost function, which is appropriately adjusted at each iteration. We discuss several variations of the basic algorithm. Some of these include combinations with the polyhedral approximation methods of the preceding chapter, yielding the class of bundle methods. Others are obtained via duality from the basic proximal algorithm, including the augmented Lagrangian method (also called method of multipliers) for constrained optimization. Finally, we discuss extensions of the proximal algorithm for finding a zero of a maximal monotone operator, and a major special case: the alternating direction method of multipliers, which is well suited for taking advantage of the structure of several types of large-scale problems.

Chapter 6: Here we discuss a variety of algorithmic topics that supplement our discussion of the descent and approximation methods of the preceding chapters. We first discuss gradient projection methods and variations with extrapolation that have good complexity properties, including Nesterov's optimal complexity algorithm. These were developed for differentiable problems, and can be extended to the nondifferentiable case by means of a smoothing scheme. Then we discuss a number of combinations of gradient, subgradient, and proximal methods that are well suited for specially structured problems. We pay special attention to incremental versions for the case where the cost function consists of the sum of a large number of component terms. We also describe additional methods, such as the classical block coordinate descent approach, the proximal algorithm with a nonquadratic regularization term, and the ϵ-descent method. We close the chapter with a discussion of interior point methods.

Our lines of analysis are largely based on differential calculus-type ideas, which are central in nonlinear programming, and on concepts of hyperplane separation, conjugacy, and duality, which are central in convex analysis. A traditional use of duality is to establish the equivalence and the connections between a pair of primal and dual problems, which may in turn enhance insight and enlarge the set of options for analysis and computation. The book makes heavy use of this type of problem duality, but also emphasizes a qualitatively different, algorithm-oriented type of duality that is largely based on conjugacy. In particular, some fundamental algorithmic operations turn out to be dual to each other, and whenever they arise in various algorithms they admit dual implementations, often with significant gains in insight and computational convenience. Some important examples are the duality between the subdifferentials of a convex function and its conjugate, the duality of a proximal operation using a convex function and an augmented Lagrangian minimization using its conjugate, and the duality between outer linearization of a convex function and inner linearization of its conjugate. Several interesting algorithms in Chapters 4-6 admit dual implementations based on these pairs of operations.

The book contains a fair number of exercises, many of them supplementing the algorithmic development and analysis. In addition a large number of theoretical exercises (with carefully written solutions) for the "theory book," together with other related material, can be obtained from the book's web page http://www.athenasc.com/convexalgorithms.html, and the author's web page http://web.mit.edu/dimitrib/www/home.html. The MIT OpenCourseWare site http://ocw.mit.edu/index.htm, also provides lecture slides and other relevant material.

The mathematical prerequisites for the book are a first course in linear algebra and a first course in real analysis. A summary of the relevant material is provided in Appendix A. Prior exposure to linear and nonlinear optimization algorithms is not assumed, although it will undoubtedly be helpful in providing context and perspective. Other than this background, the development is self-contained, with proofs provided throughout.

The present book, in conjunction with its "theory" counterpart may be used as a text for a one-semester or two-quarter convex optimization course; I have taught several variants of such a course at MIT and elsewhere over the last fifteen years. Still the book may not provide all of the convex optimization material an instructor may wish for, and it may need to be supplemented by works that aim primarily at specific types of convex optimization models, or address more comprehensively computational complexity issues. I have added representative citations for such works, which, however, are far from complete in view of the explosive growth of the literature on the subject.

The book may also be used as a supplementary source for nonlinear programming classes that are primarily focused on classical differentiable nonconvex optimization material (Kuhn-Tucker theory, Newton-like and conjugate direction methods, interior point, penalty, and augmented Lagrangian methods). For such courses, it may provide a nondifferentiable convex optimization component.

I was fortunate to have several outstanding collaborators in my research on various aspects of convex optimization: Vivek Borkar, Jon Eckstein, Eli Gafni, Xavier Luque, Angelia Nedić, Asuman Ozdaglar, John Tsitsiklis, Mengdi Wang, and Huizhen (Janey) Yu. Substantial portions of our joint research have found their way into the book. In addition, I am grateful for interactions and suggestions I received from several colleagues, including Leon Bottou, Steve Boyd, Tom Luo, Steve Wright, and particularly Mark Schmidt and Lin Xiao who read with care major portions of the book. I am also very thankful for the valuable proofreading of parts of the book by Mengdi Wang and Huizhen (Janey) Yu, and particularly by Ivan Pejcic who went through most of the book with a keen eye. I developed the book through convex optimization classes at MIT over a fifteen-year period, and I want to express appreciation for my students who provided continuing motivation and inspiration.

Finally, I would like to mention Paul Tseng, a major contributor to numerous topics in this book, who was my close friend and research collaborator on optimization algorithms for many years, and whom we unfortunately lost while he was still at his prime. I am dedicating the book to his memory.

Dimitri P. Bertsekas
dimitrib@mit.edu
January 2015

This monograph presents the main complexity theorems in convex optimization and their corresponding algorithms. Starting from the fundamental theory of black-box optimization, the material progresses towards recent advances in structural optimization and stochastic optimization. Our presentation of black-box optimization, strongly influenced by Nesterov's seminal book and Nemirovski's lecture notes, includes the analysis of cutting plane methods, as well as (accelerated) gradient descent schemes. Convex Optimization. Stephen Boyd Department of Electrical Engineering Stanford University Lieven Vandenberghe Electrical Engineering Department University of California, Los Angeles.Â This book is about convex optimization, a special class of mathematical optimiza-tion problems, which includes least-squares and linear programming problems. It is well known that least-squares and linear programming problems have a fairly complete theory, arise in a variety of applications, and can be solved numerically very eciently. Surprisingly, algorithms for convex optimization have also been used to design counting problems over discrete objects such as matroids. Simultaneously, algorithms for convex optimization have become central to many modern machine learning applications. The demand for al-gorithms for convex optimization, driven by larger and increasingly complex input instances, has also signicantly pushed the state of the art of convex op-timization itself.