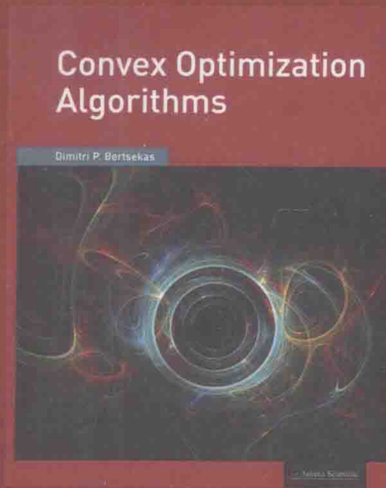
清华版双语教学用书



凸优化算法

Convex Optimization Algorithms

Dimitri P. Bertsekas 著

清华大学出版社

清华版双语教学用书

凸优化算法

Convex Optimization Algorithms

Dimitri P. Bertsekas 著

清华大学出版社
北京

图书在版编目(CIP)数据

凸优化算法=Convex Optimization Algorithms/(美)博塞卡斯(Bertsekas,D. P.)著. —北京 :清华大学出版社,2016
清华版双语教学用书
ISBN 978-7-302-43070-4

Ⅰ. ①凸… Ⅱ. ①博… Ⅲ. ①凸分析—最优化算法—双语教学—教材—英、汉 Ⅳ. ①O174.13 ②O242.23

中国版本图书馆 CIP 数据核字(2016)第 036014 号

责任编辑:王一玲
封面设计:常雪影
责任印制:刘海龙

出版发行:清华大学出版社
网　　址:http://www.tup.com.cn, http://www.wqbook.com
地　　址:北京清华大学学研大厦 A 座　　**邮　　编:**100084
社 总 机:010-62770175　　**邮　　购:**010-62786544
投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn
质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn
印 刷 者:三河市君旺印务有限公司
装 订 者:三河市新茂装订有限公司
经　　销:全国新华书店
开　　本:155mm×235mm　　**印　　张:**36.25　　**字　　数:**623 千字
版　　次:2016 年 5 月第 1 版　　**印　　次:**2016 年 5 月第1 次印刷
印　　数:1～2000
定　　价:89.00 元

产品编号:066025-01

影印版序

随着大规模资源分配、信号处理、机器学习等应用领域的快速发展，凸优化近来正引起人们日益浓厚的兴趣。本书力图给大家较为全面通俗地介绍求解大规模凸优化问题的最新算法。

本书是作者2009年出版的Convex Optimization Theory一书(原版影印本及中译本：《凸优化理论》，分别于2010年和2015年由清华大学出版社出版)的补充。不过，本书也可以单独阅读。《凸优化理论》一书侧重在凸性理论和基于对偶性的优化方面，而本书则侧重于凸优化的算法方面。本书是从《凸优化理论》原来的一章扩展而来。两本书所需要的数学基础相同，合起来内容比较完整地涵盖了有限维凸优化领域的几乎全部知识。两本书的一个共同特色是在坚持严格的数学分析基础上，十分注重对概念的直观展示。

本书几乎囊括了所有主流的凸优化算法。包括梯度法、次梯度法、多面体逼近法、邻近法和内点法等。这些方法通常依赖于代价函数和约束条件的凸性(而不一定依赖于其可微性)，并与对偶性有着直接或间接的联系。作者针对具体问题的特定结构，给出了大量的例题，来充分展示算法的应用。各章的内容如下：第1章，凸优化模型概述；第2章，优化算法概述；第3章，次梯度算法；第4章，多面体逼近算法；第5章，邻近算法；第6章，其他算法问题。本书的一个特色是在强调问题之间的对偶性的同时，也十分重视建立在共轭概念上的算法之间的对偶性，这常常能为选择合适的算法实现方式提供新的灵感和计算上的便利。

本书均取材于作者过去15年在美国麻省理工学院的凸优化方面课堂教学的内容。本书和《凸优化理论》这两本书合起来可以作为一个学期的凸优化课程的教材；本书也可以用作非线性规划课程的补充材料。因为通常传统的非线性规划课程侧重于可微但非凸的内容，如Kuhn－Tucker理论、牛顿法、共轭方向法、内点法、罚函数法和增广的拉格朗日法等。

本书作者德梅萃·博塞克斯(Dimitri P. Bertsekas)教授是优化理论的国际著名学者、美国国家工程院院士，现任美国麻省理工学院电气工程

与计算机科学系教授，曾在斯坦福大学工程经济系和伊利诺伊大学电气工程系任教，在优化理论、控制工程、通信工程、计算机科学等领域有丰富的科研教学经验，成果丰硕。博塞克斯教授是一位多产作者，著有 14 本专著和教科书。

赵千川　教授

2016 年 1 月于清华大学

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," the 2014 Khachiyan Prize for "life-time accomplishments in optimization," and the SIAM/MOS 2015 George B. Dantzig Prize for "original research, which by its originality, breadth, and scope, is having a major impact on the field of mathematical programming." 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

Contents

1
Convex Optimization Models: An Overview

Contents

In this chapter we provide an overview of some broad classes of convex optimization models. Our primary focus will be on large challenging problems, often connected in some way to duality. We will consider two types of duality. The first is *Lagrange duality* for constrained optimization, which is obtained by assigning dual variables to the constraints. The second is *Fenchel duality* together with its special case, conic duality, which involves a cost function that is the sum of two convex function components. Both of these duality structures arise often in applications, and in Sections 1.1 and 1.2 we provide an overview, and discuss some examples.†

In Sections 1.3 and 1.4, we discuss additional model structures involving a large number of additive terms in the cost, or a large number of constraints. These types of problems also arise often in the context of duality, as well as in other contexts such as machine learning and signal processing with large amounts of data. In Section 1.5, we discuss the exact penalty function technique, whereby we can transform a convex constrained optimization problem to an equivalent unconstrained problem.

1.1 LAGRANGE DUALITY

We start our overview of Lagrange duality with the basic case of nonlinear inequality constraints, and then consider extensions involving linear inequality and equality constraints. Consider the problem‡

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in X, \quad g(x) \leq 0, \end{array} \tag{1.1}$$

where X is a nonempty set,

$$g(x) = \big(g_1(x), \ldots, g_r(x)\big)',$$

and $f : X \mapsto \Re$ and $g_j : X \mapsto \Re$, $j = 1, \ldots, r$, are given functions. We refer to this as the *primal problem*, and we denote its optimal value by f^*. A vector x satisfying the constraints of the problem is referred to as *feasible*. The *dual* of problem (1.1) is given by

$$\begin{array}{ll} \text{maximize} & q(\mu) \\ \text{subject to} & \mu \in \Re^r, \end{array} \tag{1.2}$$

† Consistent with its overview character, this chapter contains few proofs, and refers frequently to the literature, and to Appendix B, which contains a full list of definitions and propositions (without proofs) relating to nonalgorithmic aspects of convex optimization. This list reflects and summarizes the content of the author's "Convex Optimization Theory" book [Ber09]. The proposition numbers of [Ber09] have been preserved, so all omitted proofs of propositions in Appendix B can be readily accessed from [Ber09].

‡ Appendix A contains an overview of the mathematical notation, terminology, and results from linear algebra and real analysis that we will be using.

where the dual function q is

$$q(\mu) = \begin{cases} \inf_{x\in X} L(x,\mu) & \text{if } \mu \geq 0, \\ -\infty & \text{otherwise,} \end{cases}$$

and L is the Lagrangian function defined by

$$L(x,\mu) = f(x) + \mu' g(x), \qquad x \in X,\ \mu \in \Re^r;$$

(cf. Section 5.3 of Appendix B).

Note that the dual function is extended real-valued, and that the effective constraint set of the dual problem is

$$\left\{ \mu \geq 0 \;\middle|\; \inf_{x\in X} L(x,\mu) > -\infty \right\}.$$

The optimal value of the dual problem is denoted by q^*.

The *weak duality* relation, $q^* \leq f^*$, always holds. It is easily shown by writing for all $\mu \geq 0$, and $x \in X$ with $g(x) \leq 0$,

$$q(\mu) = \inf_{z\in X} L(z,\mu) \leq L(x,\mu) = f(x) + \sum_{j=1}^{r} \mu_j g_j(x) \leq f(x),$$

so that

$$q^* = \sup_{\mu\in\Re^r} q(\mu) = \sup_{\mu\geq 0} q(\mu) \leq \inf_{x\in X,\, g(x)\leq 0} f(x) = f^*.$$

We state this formally as follows (cf. Prop. 4.1.2 in Appendix B).

Proposition 1.1.1: (Weak Duality Theorem) Consider problem (1.1). For any feasible solution x and any $\mu \in \Re^r$, we have $q(\mu) \leq f(x)$. Moreover, $q^* \leq f^*$.

When $q^* = f^*$, we say that *strong duality* holds. The following proposition gives necessary and sufficient conditions for strong duality, and primal and dual optimality (see Prop. 5.3.2 in Appendix B).

Proposition 1.1.2: (Optimality Conditions) Consider problem (1.1). There holds $q^* = f^*$, and (x^*, μ^*) are a primal and dual optimal solution pair if and only if x^* is feasible, $\mu^* \geq 0$, and

$$x^* \in \arg\min_{x\in X} L(x,\mu^*), \qquad \mu_j^* g_j(x^*) = 0, \quad j = 1,\ldots,r.$$