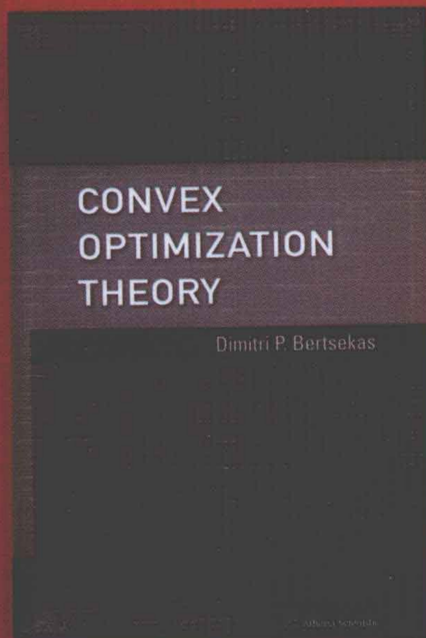
清华版双语教学用书



（影印版）

凸优化理论

Convex Optimization Theory

Dimitri P. Bertsekas 著

清华大学出版社

清华版双语教学用书

（影印版）

凸优化理论

Convex Optimization Theory

Dimitri P. Bertsekas 著

清华大学出版社
北京

图书在版编目（CIP）数据

凸优化理论：英文/（美）博赛克斯（Bertsekas，D. P.）著. --北京：清华大学出版社，2011. 1

书名原文：Convex Optimization Theory

（清华版双语教学用书）

ISBN 978-7-302-23760-0

Ⅰ. ①凸…　Ⅱ. ①博…　Ⅲ. ①凸分析—双语教学—教材—英文　Ⅳ. O174. 13

中国版本图书馆 CIP 数据核字（2010）第 168142 号

责任编辑：王一玲
责任印制：王秀菊
出版发行：清华大学出版社　　**地　址**：北京清华大学学研大厦 A 座
http://www.tup.com.cn　　**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：清华大学印刷厂
装 订 者：北京市密云县京文制本装订厂
经　销：全国新华书店
开　本：160×240　**印　张**：26
版　次：2011 年 1 月第 1 版　　**印　次**：2011 年 1 月第1 次印刷
印　数：1～3000
定　价：49.00 元

产品编号：035001-01

凸规划(Convex Optimization Theory)

影印版序言

优化理论与应用是非常经典但依然非常活跃的研究领域，涉及几乎所有的理工和管理学科以及计量社会科学学科，是系统工程、运筹学、计量经济学等学科的理论基础。凸优化是优化理论十分重要的分支，是本书讨论的重点。

凸优化是指目标函数为凸函数、约束集为凸集合的约束优化问题。凸优化具有重要的工程应用背景，求解凸优化问题的方法通常也是一般非线性规划方法的重要基础。

本书是凸优化理论与方法的重要专著和教材，主要内容分为两部分：凸分析和凸问题的对偶优化理论。本书先从基本线性代数和实分析理论出发，比较详尽地讨论了凸理论和凸分析，为求解凸优化问题建立了足够的基础。本书在引入了凸优化的基本概念后，着重讨论了对偶优化理论。本书从比较独特的几何问题角度——最小共同点和最大相交点问题——引入了对偶理论框架，讨论对偶性和对偶优化解的存在性等问题。在此统一对偶理论框架下，本书讨论了多种优化问题如线性规划、凸规划、最小最大等问题的对偶性和对偶优化理论，并讨论了当目标函数非光滑时的次梯度和最优性条件。

本书的重要特点是自成体系，所需要的基础知识除理工科本科线性代数和少量实分析基本概念和理论外，并不需要一般优化理论如线性规划、非线性规划等作为基础。所以本书既适用作研究生的教材，也可作为优化理论与方法研究者的参考书。

本书作者德梅萃·博赛克斯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。博赛克斯教授是一位多产作者，著有 14 本专著和教科书。本书是作者在优化理论与方法的系列专著和教科书中的一本，自成体系又相互对应。

博赛克斯教授于2009年访问中国，在清华大学、西安交通大学等单位讲学数周，讲学内容包括了本书的一部分。博赛克斯教授关心和支持中国相关领域的研究以及中美两国间的学术合作，近年来同包括笔者在内的国内学者多有合作。本书影印本在国内出版是国内学术界和工程界的一大幸事，笔者愿借此祝贺博赛克斯教授。

管晓宏

2010年10月

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 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 fourteen books, several of which are used as textbooks in MIT classes, including "Nonlinear Programming," "Dynamic Programming and Optimal Control," "Data Networks," "Introduction to Probability," 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 2000 Greek National Award for Operations Research, and the 2001 ACC John R. Ragazzini Education Award. In 2001, he was elected to the United States National Academy of Engineering.

ATHENA SCIENTIFIC

OPTIMIZATION AND COMPUTATION SERIES

1. Convex Optimization Theory, by Dimitri P. Bertsekas, 2009, ISBN 978-1-886529-31-1, 256 pages
2. Introduction to Probability, 2nd Edition, by Dimitri P. Bertsekas and John N. Tsitsiklis, 2008, ISBN 978-1-886529-23-6, 544 pages
3. Dynamic Programming and Optimal Control, Two-Volume Set, by Dimitri P. Bertsekas, 2007, ISBN 1-886529-08-6, 1020 pages
4. Convex Analysis and Optimization, by Dimitri P. Bertsekas, with Angelia Nedić and Asuman E. Ozdaglar, 2003, ISBN 1-886529-45-0, 560 pages
5. Nonlinear Programming, 2nd Edition, by Dimitri P. Bertsekas, 1999, ISBN 1-886529-00-0, 791 pages
6. Network Optimization: Continuous and Discrete Models, by Dimitri P. Bertsekas, 1998, ISBN 1-886529-02-7, 608 pages
7. Network Flows and Monotropic Optimization, by R. Tyrrell Rockafellar, 1998, ISBN 1-886529-06-X, 634 pages
8. Introduction to Linear Optimization, by Dimitris Bertsimas and John N. Tsitsiklis, 1997, ISBN 1-886529-19-1, 608 pages
9. Parallel and Distributed Computation: Numerical Methods, by Dimitri P. Bertsekas and John N. Tsitsiklis, 1997, ISBN 1-886529-01-9, 718 pages
10. Neuro-Dynamic Programming, by Dimitri P. Bertsekas and John N. Tsitsiklis, 1996, ISBN 1-886529-10-8, 512 pages
11. Constrained Optimization and Lagrange Multiplier Methods, by Dimitri P. Bertsekas, 1996, ISBN 1-886529-04-3, 410 pages
12. Stochastic Optimal Control: The Discrete-Time Case, by Dimitri P. Bertsekas and Steven E. Shreve, 1996, ISBN 1-886529-03-5, 330 pages

Preface

This book aims at an accessible, concise, and intuitive exposition of two related subjects that find broad practical application:

(a) Convex analysis, particularly as it relates to optimization.

(b) Duality theory for optimization and minimax problems, mainly within a convexity framework.

The focus on optimization is to derive conditions for existence of primal and dual optimal solutions for constrained problems such as

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \qquad g_j(x) \le 0, \quad j = 1, \ldots, r. \end{aligned}$$

Other types of optimization problems, such as those arising in Fenchel duality, are also part of our scope. The focus on minimax is to derive conditions guaranteeing the equality

$$\inf_{x \in X} \sup_{z \in Z} \phi(x, z) = \sup_{z \in Z} \inf_{x \in X} \phi(x, z),$$

and the attainment of the "inf" and the "sup."

The treatment of convexity theory is fairly detailed. It touches upon nearly all major aspects of the subject, and it is sufficient for the development of the core analytical issues of convex optimization. The mathematical prerequisites are a first course in linear algebra and a first course in real analysis. A summary of the relevant material is provided in an appendix. Prior knowledge of linear and nonlinear optimization theory is not assumed, although it will undoubtedly be helpful in providing context and perspective. Other than this modest background, the development is self-contained, with rigorous proofs provided throughout.

We have aimed at a unified development of the strongest possible forms of duality with the most economic use of convexity theory. To this end, our analysis often departs from the lines of Rockafellar's classic 1970 book and other books that followed the Fenchel/Rockafellar formalism. For example, we treat differently closed set intersection theory and preservation of closure under linear transformations (Sections 1.4.2 and 1.4.3); we develop subdifferential calculus by using constrained optimization duality (Section 5.4.2); and we do not rely on concepts such as infimal convolution, image, polar sets and functions, bifunctions, and conjugate saddle functions. Perhaps our greatest departure is in duality theory itself: similar to Fenchel/Rockafellar, our development rests on Legendre/Fenchel conjugacy ideas, but is far more geometrical and visually intuitive.

Our duality framework is based on two simple geometrical problems: the *min common point problem* and the *max crossing point problem*. The salient feature of the min common/max crossing (MC/MC) framework is its highly visual geometry, through which all the core issues of duality theory become apparent and can be analyzed in a unified way. Our approach is to obtain a handful of broadly applicable theorems within the MC/MC framework, and then specialize them to particular types of problems (constrained optimization, Fenchel duality, minimax problems, etc). We address all duality questions (existence of duality gap, existence of dual optimal solutions, structure of the dual optimal solution set), and other issues (subdifferential theory, theorems of the alternative, duality gap estimates) in this way.

Fundamentally, the MC/MC framework is closely connected to the conjugacy framework, and owes its power and generality to this connection. However, the two frameworks offer complementary starting points for analysis and provide alternative views of the geometric foundation of duality: conjugacy emphasizes functional/algebraic descriptions, while MC/MC emphasizes set/epigraph descriptions. The MC/MC framework is simpler, and seems better suited for visualizing and investigating questions of strong duality and existence of dual optimal solutions. The conjugacy framework, with its emphasis on functional descriptions, is more suitable when mathematical operations on convex functions are involved, and the calculus of conjugate functions can be brought to bear for analysis or computation.

The book evolved from the earlier book of the author [BNO03] on the subject (coauthored with A. Nedić and A. Ozdaglar), but has different character and objectives. The 2003 book was quite extensive, was structured (at least in part) as a research monograph, and aimed to bridge the gap between convex and nonconvex optimization using concepts of nonsmooth analysis. By contrast, the present book is organized differently, has the character of a textbook, and concentrates exclusively on convex optimization. Despite the differences, the two books have similar style and level of mathematical sophistication, and share some material.

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

Chapter 1: This chapter develops all of the convex analysis tools that are needed for the development of duality theory in subsequent chapters. It covers basic algebraic concepts such as convex hulls and hyperplanes, and topological concepts such as relative interior, closure, preservation of closedness under linear transformations, and hyperplane separation. In addition, it develops subjects of special interest in duality and optimization, such as recession cones and conjugate functions.

Chapter 2: This chapter covers polyhedral convexity concepts: extreme points, the Farkas and Minkowski-Weyl theorems, and some of their applications in linear programming. It is not needed for the developments of subsequent chapters, and may be skipped at first reading.

Chapter 3: This chapter focuses on basic optimization concepts: types of minima, existence of solutions, and a few topics of special interest for duality theory, such as partial minimization and minimax theory.

Chapter 4: This chapter introduces the MC/MC duality framework. It discusses its connection with conjugacy theory, and it charts its applications to constrained optimization and minimax problems. It then develops broadly applicable theorems relating to strong duality and existence of dual optimal solutions.

Chapter 5: This chapter specializes the duality theorems of Chapter 4 to important contexts relating to linear programming, convex programming, and minimax theory. It also uses these theorems as an aid for the development of additional convex analysis tools, such as a powerful nonlinear version of Farkas' Lemma, subdifferential theory, and theorems of the alternative. A final section is devoted to nonconvex problems and estimates of the duality gap, with special focus on separable problems.

In aiming for brevity, I have omitted a number of topics that an instructor may wish for. One such omission is applications to specially structured problems; the book by Boyd and Vanderbergue [BoV04], as well as my book on parallel and distributed computation with John Tsitsiklis [BeT89] cover this material extensively (both books are available on line).

Another important omission is computational methods. However, I have written a long supplementary sixth chapter (over 100 pages), which covers the most popular convex optimization algorithms (and some new ones), and can be downloaded from the book's web page

http://www.athenasc.com/convexduality.html.

This chapter, together with a more comprehensive treatment of convex analysis, optimization, duality, and algorithms will be part of a more extensive textbook that I am currently writing. Until that time, the chapter will serve instructors who wish to cover convex optimization algorithms in addition to duality (as I do in my M.I.T. course). This is a "living" chapter that will be periodically updated. Its current contents are as follows:

Chapter 6 on Algorithms: 6.1. Problem Structures and Computational Approaches; 6.2. Algorithmic Descent; 6.3. Subgradient Methods; 6.4. Polyhedral Approximation Methods; 6.5. Proximal and Bundle Methods; 6.6. Dual Proximal Point Algorithms; 6.7. Interior Point Methods; 6.8. Approximate Subgradient Methods; 6.9. Optimal Algorithms and Complexity.

While I did not provide exercises in the text, I have supplied a substantial number of exercises (with detailed solutions) at the book's web page. The reader/instructor may also use the end-of-chapter problems (a total of 175) given in [BNO03], which have similar style and notation to the present book. Statements and detailed solutions of these problems can be downloaded from the book's web page and are also available on line at

http://www.athenasc.com/convexity.html.

The book may be used as a text for a theoretical convex optimization course; I have taught several variants of such a course at MIT and elsewhere over the last ten years. It may also be used as a supplementary source for nonlinear programming classes, and as a theoretical foundation for classes focused on convex optimization models (rather than theory).

The book has been structured so that the reader/instructor can use the material selectively. For example, the polyhedral convexity material of Chapter 2 can be omitted in its entirety, as it is not used in Chapters 3-5. Similarly, the material on minimax theory (Sections 3.4, 4.2.5, and 5.5) may be omitted; and if this is done, Sections 3.3 and 5.3.4, which use the tools of partial minimization, may be omitted. Also, Sections 5.4-5.7 are "terminal" and may each be omitted without effect on other sections.

A "minimal" self-contained selection, which I have used in my nonlinear programming class at MIT (together with the supplementary web-based Chapter 6 on algorithms), consists of the following:

- Chapter 1, except for Sections 1.3.3 and 1.4.1.
- Section 3.1.
- Chapter 4, except for Section 4.2.5.
- Chapter 5, except for Sections 5.2, 5.3.4, and 5.5-5.7.

This selection focuses on nonlinear convex optimization, and excludes all the material relating to polyhedral convexity and minimax theory.

I would like to express my thanks to several colleagues for their contributions to the book. My collaboration with Angelia Nedić and Asuman Ozdaglar on our 2003 book was important in laying the foundations of the present book. Huizhen (Janey) Yu read carefully early drafts of portions of the book, and offered several insightful suggestions. Paul Tseng contributed substantially through our joint research on set intersection theory, given in part in Section 1.4.2 (this research was motivated by earlier collaboration with Angelia Nedić). Feedback from students and colleagues, including Dimitris Bisias, Vivek Borkar, John Tsitsiklis, Mengdi Wang, and Yunjian Xu, is highly appreciated. Finally, I wish to thank the many outstanding students in my classes, who have been a continuing source of motivation and inspiration.

Contents

1

Basic Concepts of Convex Analysis

Contents

Convex sets and functions are very useful in optimization models, and have a rich structure that is convenient for analysis and algorithms. Much of this structure can be traced to a few fundamental properties. For example, each closed convex set can be described in terms of the hyperplanes that support the set, each point on the boundary of a convex set can be approached through the relative interior of the set, and each halfline belonging to a closed convex set still belongs to the set when translated to start at any point in the set.

Yet, despite their favorable structure, convex sets and their analysis are not free of anomalies and exceptional behavior, which cause serious difficulties in theory and applications. For example, contrary to affine and compact sets, some basic operations such as linear transformation and vector sum may not preserve the closedness of closed convex sets. This in turn complicates the treatment of some fundamental optimization issues, including the existence of optimal solutions and duality.

For this reason, it is important to be rigorous in the development of convexity theory and its applications. Our aim in this first chapter is to establish the foundations for this development, with a special emphasis on issues that are relevant to optimization.

1.1 CONVEX SETS AND FUNCTIONS

We introduce in this chapter some of the basic notions relating to convex sets and functions. This material permeates all subsequent developments in this book. Appendix A provides the definitions, notational conventions, and results from linear algebra and real analysis that we will need. We first define convex sets (cf. Fig. 1.1.1).

Definition 1.1.1: A subset C of $\Re^n$ is called *convex* if

$$\alpha x + (1-\alpha)y \in C, \qquad \forall\, x, y \in C,\ \forall\, \alpha \in [0,1].$$

Note that the empty set is by convention considered to be convex. Generally, when referring to a convex set, it will usually be apparent from the context whether this set can be empty, but we will often be specific in order to minimize ambiguities. The following proposition gives some operations that preserve convexity.

Proposition 1.1.1:

(a) The intersection $\cap_{i\in I} C_i$ of any collection $\{C_i \mid i \in I\}$ of convex sets is convex.

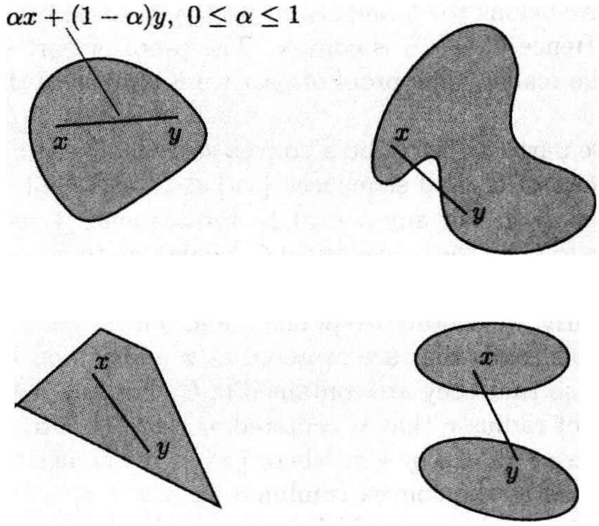


Figure 1.1.1. Illustration of the definition of a convex set. For convexity, linear interpolation between any two points of the set must yield points that lie within the set. Thus the sets on the left are convex, but the sets on the right are not.

(b) The vector sum $C_1 + C_2$ of two convex sets C_1 and C_2 is convex.

(c) The set λC is convex for any convex set C and scalar λ. Furthermore, if C is a convex set and λ_1, λ_2 are positive scalars,

$$(\lambda_1 + \lambda_2)C = \lambda_1 C + \lambda_2 C.$$

(d) The closure and the interior of a convex set are convex.

(e) The image and the inverse image of a convex set under an affine function are convex.

Proof: The proof is straightforward using the definition of convexity. To prove part (a), we take two points x and y from $\cap_{i\in I} C_i$, and we use the convexity of C_i to argue that the line segment connecting x and y belongs to all the sets C_i, and hence, to their intersection.

Similarly, to prove part (b), we take two points of $C_1 + C_2$, which we represent as $x_1 + x_2$ and $y_1 + y_2$, with $x_1, y_1 \in C_1$ and $x_2, y_2 \in C_2$. For any $\alpha \in [0, 1]$, we have

$$\alpha(x_1 + x_2) + (1-\alpha)(y_1 + y_2) = \big(\alpha x_1 + (1-\alpha)y_1\big) + \big(\alpha x_2 + (1-\alpha)y_2\big).$$

By convexity of C_1 and C_2, the vectors in the two parentheses of the righthand side above belong to C_1 and C_2, respectively, so that their sum belongs to $C_1 + C_2$. Hence $C_1 + C_2$ is convex. The proof of part (c) is left as an exercise for the reader. The proof of part (e) is similar to the proof of part (b).

To prove part (d), let C be a convex set. Choose two points x and y from the closure of C, and sequences $\{x_k\} \subset C$ and $\{y_k\} \subset C$, such that $x_k \to x$ and $y_k \to y$. For any $\alpha \in [0,1]$, the sequence $\{\alpha x_k + (1-\alpha)y_k\}$, which belongs to C by the convexity of C, converges to $\alpha x + (1-\alpha)y$. Hence $\alpha x + (1-\alpha)y$ belongs to the closure of C, showing that the closure of C is convex. Similarly, we choose two points x and y from the interior of C, and we consider open balls that are centered at x and y, and have sufficiently small radius r so that they are contained in C. For any $\alpha \in [0,1]$, consider the open ball of radius r that is centered at $\alpha x + (1-\alpha)y$. Any point in this ball, say $\alpha x + (1-\alpha)y + z$, where $\|z\| < r$, belongs to C, because it can be expressed as the convex combination $\alpha(x+z) + (1-\alpha)(y+z)$ of the vectors $x+z$ and $y+z$, which belong to C. Hence the interior of C contains $\alpha x + (1-\alpha)y$ and is therefore convex. **Q.E.D.**

Special Convex Sets

We will often consider some special sets, which we now introduce. A *hyperplane* is a set specified by a single linear equation, i.e., a set of the form $\{x \mid a'x = b\}$, where a is a nonzero vector and b is a scalar. A *halfspace* is a set specified by a single linear inequality, i.e., a set of the form $\{x \mid a'x \le b\}$, where a is a nonzero vector and b is a scalar. It is clearly closed and convex. A set is said to be *polyhedral* if it is nonempty and it is the intersection of a finite number of halfspaces, i.e., if it has the form

$$\{x \mid a_j'x \le b_j,\ j = 1,\ldots,r\},$$

where $a_1,\ldots,a_r$ and $b_1,\ldots,b_r$ are some vectors in $\Re^n$ and scalars, respectively. A polyhedral set is convex and closed, being the intersection of halfspaces [cf. Prop. 1.1.1(a)].

A set C is said to be a *cone* if for all $x \in C$ and $\lambda > 0$, we have $\lambda x \in C$. A cone need not be convex and need not contain the origin, although the origin always lies in the closure of a nonempty cone (see Fig. 1.1.2). A *polyhedral cone* is a set of the form

$$C = \{x \mid a_j'x \le 0,\ j = 1,\ldots,r\},$$

where $a_1,\ldots,a_r$ are some vectors in $\Re^n$. A subspace is a special case of a polyhedral cone, which is in turn a special case of a polyhedral set.