国际知名大学原版教材

——信息技术学科与电气工程学科系列 30

Convex Analysis and Optimization

凸分析与优化

Dimitri P. Bertsekas with
Angelia Nedić and Asuman E. Ozdaglar

清华大学出版社

国际知名大学原版教材——信息技术学科与电气工程学科系列

Convex Analysis and Optimization

凸分析与优化

Dimitri P. Bertsekas with
Angelia Nedić and Asuman E. Ozdaglar

清华大学出版社
北京

图书在版编目（CIP）数据

凸分析与优化/（　）伯特塞卡斯著．—影印本．—北京：清华大学出版社，2006.2
（国际知名大学原版教材．信息技术学科与电气工程学科系列）
ISBN 7-302-12328-4

Ⅰ．凸…　Ⅱ．伯…　Ⅲ．凸分析—高等学校—教材—英文　Ⅳ．O174.13

中国版本图书馆 CIP 数据核字（2006）第 000811 号

出 版 者：清华大学出版社　　**地　　址**：北京清华大学学研大厦
http://www.tup.com.cn　　**邮　　编**：100084
社 总 机：010-62770175　　**客户服务**：010-62776969
责任编辑：王一玲
印 刷 者：清华大学印刷厂
装 订 者：三河市新茂装订有限公司
发 行 者：新华书店总店北京发行所
开　　本：175×245　**印张**：35.75
版　　次：2006 年 2 月第 1 版　2006 年 2 月第 1 次印刷
书　　号：ISBN 7-302-12328-4/O·513
印　　数：1～3000
定　　价：65.00 元

国际知名大学原版教材
——信息技术学科与电气工程学科系列

出 版 说 明

郑大钟
清华大学信息科学与技术学院

当前，在我国的高等学校中，教学内容和课程体系的改革已经成为教学改革中的一个非常突出的问题，而为数不少的课程教材中普遍存在的“课程体系老化，内容落伍时代，本研层次不清”的现象又是其中的急需改变的一个重要方面。同时，随着科教兴国方针的贯彻落实，要求我们进一步转变观念扩大视野，使教学过程适应以信息技术为先导的技术革命和我国社会主义市场经济体制的需要，加快教学过程的国际化进程。在这方面，系统地研究和借鉴国外知名大学的相关教材，将会对推进我们的课程改革和推进我国大学教学的国际化进程，乃至对我们一些重点大学建设国际一流大学的努力，都将具有重要的借鉴推动作用。正是基于这种背景，我们决定在国内推出信息技术学科和电气工程学科国外知名大学原版系列教材。

本系列教材的组编将遵循如下的几点基本原则。(1) 书目的范围限于信息技术学科和电气工程学科所属专业的技术基础课和主要的专业课。(2) 教材的范围选自于具有较大影响且为国外知名大学所采用的教材。(3) 教材属于在近 5 年内所出版的新书或新版书。(4) 教材适合于作为我国大学相应课程的教材或主要教学参考书。(5) 每本列选的教材都须经过国内相应领域的资深专家审看和推荐。(6) 教材的形式直接以英文原版形式印刷出版。

本系列教材将按分期分批的方式组织出版。为了便于使用本系列教材的相关教师和学生从学科和教学的角度对其在体系和内容上的特点和特色有所了解，在每本教材中都附有我们所约请的相关领域资深教授撰写的影印版序言。此外，出于多样化的考虑，对于某些基本类型的课程，我们还同时列选了多于一本的不同体系、不同风格和不同层次的教材，以供不同要求和不同学时的同类课程的选用。

本系列教材的读者对象为信息技术学科和电气工程学科所属各专业的本科生，同时兼顾其他工程学科专业的本科生或研究生。本系列教材，既可采用作为相应课程的教材或教学参考书，也可提供作为工作于各个技术领域的工程师和技术人员的自学读物。

组编这套国外知名大学原版系列教材是一个尝试。不管是书目确定的合理性，教材选择的恰当性，还是评论看法的确切性，都有待于通过使用和实践来检验。感谢使用本系列教材的广大教师和学生的支持。期望广大读者提出意见和建议。

Convex Analysis and Optimization

影印版序

本书针对最优化问题介绍凸分析方法。第 1 章介绍凸集、凸函数、上境图、凸包、仿射包、相对内点、回收锥等凸分析的基本概念及其相关性质；第 2 章讨论凸性在最优化问题中的基本作用，介绍最优解集的存在性定理、投影定理、凸集分离定理、极小公共点与极大交叉点对偶问题以及一般性的极小极大定理和鞍点定理；第 3 章讨论凸集为多面体的情况，介绍线性 Farkas 引理、凸多面体的 Minkowski-Weyl 表示定理、线性规划的基本定理、凸多面体的极小极大定理以及非线性 Farkas 引理；第 4 章介绍方向导数、次梯度、次微分、切锥、法锥等基本概念及其相关性质，给出 Danskin 定理和抽象可行集描述的约束优化问题最优性条件；第 5 章讨论由抽象集合与等式和不等式一起构成的约束优化问题最优性条件，介绍 Lagrange 乘子、广义 Fritz John 条件以及各种常用约束品性；第 6 章讨论 Lagrange 对偶问题，介绍几何乘子、Lagrange 对偶定理、Lagrange 对偶问题的鞍点定理以及用几何乘子描述的广义 Fritz John 条件；第 7 章讨论共轭对偶问题，介绍 Fenchel 对偶定理；第 8 章讨论对偶计算方法，给出多种基于对偶理论求解优化问题的具体算法。这些内容涵盖了凸分析与经典优化理论所有重要的结论。

尽管同时涉及凸分析和最优化问题的教材很多，但像本书这样完整地介绍这方面内容的却不多见。与其他同类书籍相比，本书至少有三个特点。第一，在凸分析方面，系统而全面地介绍了有关的概念、性质及其在解决最优化问题中的作用；第二，在最优化理论方面，将抽象集合加入到通常仅由等式和不等式定义的可行集中，利用凸分析方法给出了最具一般性的结果，例如，第 5 章用切锥描述的广义最优性条件；第三，采用大量的几何图形说明数学概念和命题，将直观的几何解释和详尽而严格的数学推导结合在一起，为读者阅读和掌握本书内容提供了极大的便利。

本书主要作者 Dimitri P. Bertsekas 是美国麻省理工学院电气工程和计算机科学系的资深教授，他是“动态规划与随机控制”、“约束优化与 Lagrange 乘子方法”、“非线性规划”、“连续和离散模型的网络优化”、“离散时间随机最优控制”、“并行和分布计算中的数值方法”等十余部教科书的主要作者，这些教科书的大部分被用作麻省理工学院的研究生或本科生教材，本书就是其中之一。

阅读本书仅需要线性代数和数学分析的基本知识。通过学习本书，可以了解凸分析和优化领域的主要结果，掌握有关理论的本质内容，提高分析和解决最优化问题的能力。因此，所有涉足最优化与系统分析领域的理论研究人员和实际工作者均

可从学习或阅读本书中获得益处。此外，本书也可用作高年级大学生或研究生学习凸分析方法和最优化理论的教材或辅助材料。

王书宁　教授
清华大学自动化系
智能与网络化系统研究中心

ATHENA SCIENTIFIC

OPTIMIZATION AND COMPUTATION SERIES

1. Convex Analysis and Optimization, by Dimitri P. Bertsekas, with Angelia Nedić and Asuman E. Ozdaglar, 2003, ISBN 1-886529-45-0, 560 pages
2. Introduction to Probability, by Dimitri P. Bertsekas and John N. Tsitsiklis, 2002, ISBN 1-886529-40-X, 430 pages
3. Dynamic Programming and Optimal Control, Two-Volume Set (2nd Edition), by Dimitri P. Bertsekas, 2001, ISBN 1-886529-08-6, 840 pages
4. Nonlinear Programming, 2nd Edition, by Dimitri P. Bertsekas, 1999, ISBN 1-886529-00-0, 791 pages
5. Network Optimization: Continuous and Discrete Models by Dimitri P. Bertsekas, 1998, ISBN 1-886529-02-7, 608 pages
6. Network Flows and Monotropic Optimization by R. Tyrrell Rockafellar, 1998, ISBN 1-886529-06-X, 634 pages
7. Introduction to Linear Optimization by Dimitris Bertsimas and John N. Tsitsiklis, 1997, ISBN 1-886529-19-1, 608 pages
8. Parallel and Distributed Computation: Numerical Methods by Dimitri P. Bertsekas and John N. Tsitsiklis, 1997, ISBN 1-886529-01-9, 718 pages
9. Neuro-Dynamic Programming, by Dimitri P. Bertsekas and John N. Tsitsiklis, 1996, ISBN 1-886529-10-8, 512 pages
10. Constrained Optimization and Lagrange Multiplier Methods, by Dimitri P. Bertsekas, 1996, ISBN 1-886529-04-3, 410 pages
11. Stochastic Optimal Control: The Discrete-Time Case by Dimitri P. Bertsekas and Steven E. Shreve, 1996, ISBN 1-886529-03-5, 330 pages

Preface

The knowledge at which geometry aims is the knowledge of the eternal

(Plato, Republic, VII, 52)

This book focuses on the theory of convex sets and functions, and its connections with a number of topics that span a broad range from continuous to discrete optimization. These topics include Lagrange multiplier theory, Lagrangian and conjugate/Fenchel duality, minimax theory, and nondifferentiable optimization.

The book evolved from a set of lecture notes for a graduate course at M.I.T. It is widely recognized that, aside from being an eminently useful subject in engineering, operations research, and economics, convexity is an excellent vehicle for assimilating some of the basic concepts of real analysis within an intuitive geometrical setting. Unfortunately, the subject's coverage in academic curricula is scant and incidental. We believe that at least part of the reason is the shortage of textbooks that are suitable for classroom instruction, particularly for nonmathematics majors. We have therefore tried to make convex analysis accessible to a broader audience by emphasizing its geometrical character, while maintaining mathematical rigor. We have included as many insightful illustrations as possible, and we have used geometric visualization as a principal tool for maintaining the students' interest in mathematical proofs.

Our treatment of convexity theory is quite comprehensive, with all major aspects of the subject receiving substantial treatment. The mathematical prerequisites are a course in linear algebra and a course in real analysis in finite dimensional spaces (which is the exclusive setting of the book). A summary of this material, without proofs, is provided in Section 1.1.

The coverage of the theory has been significantly extended in the exercises, which represent a major component of the book. Detailed solutions of all the exercises (nearly 200 pages) are internet-posted in the book's www page

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

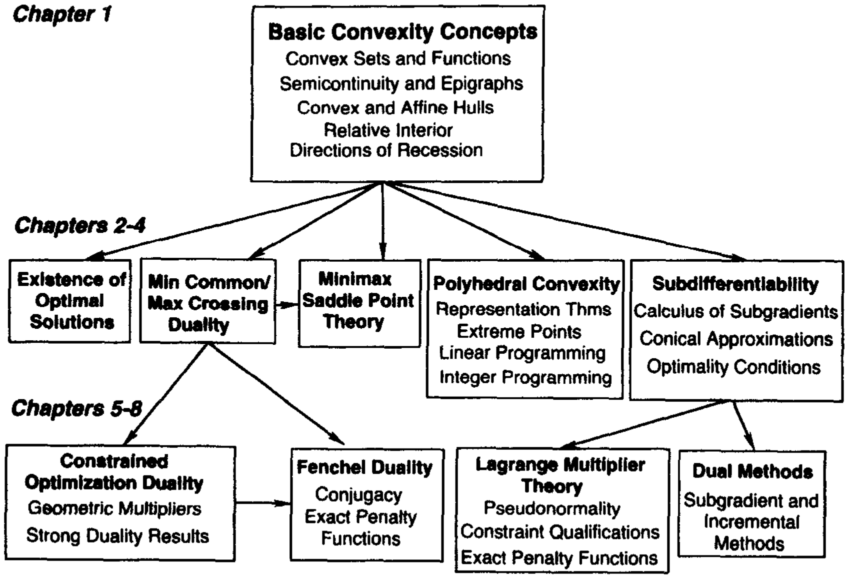
Some of the exercises may be attempted by the reader without looking at the solutions, while others are challenging but may be solved by the advanced reader with the assistance of hints. Still other exercises represent substantial theoretical results, and in some cases include new and unpublished research. Readers and instructors should decide for themselves how to make best use of the internet-posted solutions.

An important part of our approach has been to maintain a close link between the theoretical treatment of convexity and its application to optimization. For example, in Chapter 2, after the development of some of the basic facts about convexity, we discuss some of their applications to optimization and saddle point theory; in Chapter 3, after the discussion of polyhedral convexity, we discuss its application in linear and integer programming; and in Chapter 4, after the discussion of subgradients, we discuss their use in optimality conditions. We follow this style in the remaining chapters, although having developed in Chapters 1-4 most of the needed convexity theory, the discussion in the subsequent chapters is more heavily weighted towards optimization.

The chart of the opposite page illustrates the main topics covered in the book, and their interrelations. At the top level, we have the most basic concepts of convexity theory, which are covered in Chapter 1. At the middle level, we have fundamental topics of optimization, such as existence and characterization of solutions, and minimax theory, together with some supporting convexity concepts such as hyperplane separation, polyhedral sets, and subdifferentiability (Chapters 2-4). At the lowest level, we have the core issues of convex optimization: Lagrange multipliers, Lagrange and Fenchel duality, and numerical dual optimization (Chapters 5-8).

An instructor who wishes to teach a course from the book has a choice between several different plans. One possibility is to cover in detail just the first four chapters, perhaps augmented with some selected sections from the remainder of the book, such as the first section of Chapter 7, which deals with conjugate convex functions. The idea here is to concentrate on convex analysis and illustrate its application to minimax theory through the minimax theorems of Chapters 2 and 3, and to constrained optimization theory through the Nonlinear Farkas' Lemma of Chapter 3 and the optimality conditions of Chapter 4. An alternative plan is to cover Chapters 1-4 in less detail in order to allow some time for Lagrange multiplier theory and computational methods. Other plans may also be devised, possibly including some applications or some additional theoretical topics of the instructor's choice.

While the subject of the book is classical, the treatment of several of its important topics is new and in some cases relies on new research. In



particular, our new lines of analysis include:

(a) A unified development of minimax theory and constrained optimization duality as special cases of the duality between two simple geometrical problems: the min common point problem and the max crossing point problem. Here, by minimax theory, we mean the analysis relating to the minimax 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." By constrained optimization theory, we mean the analysis of problems such as

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

and issues such as the existence of optimal solutions and Lagrange multipliers, and the absence of a duality gap [equality of the optimal value of the above problem and the optimal value of an associated dual problem, obtained by assigning multipliers to the inequality constraints $g_j(x)\le 0$].

(b) A unification of conditions for existence of solutions of convex optimization problems, conditions for the minimax equality to hold, and conditions for the absence of a duality gap in constrained optimization. This unification is based on conditions guaranteeing that a nested family of closed convex sets has a nonempty intersection.

(c) A unification of the major constraint qualifications that guarantee the existence of Lagrange multipliers for nonconvex constrained optimization. This unification is achieved through the notion of constraint pseudonormality, which is motivated by an enhanced form of the Fritz John necessary optimality conditions.

(d) The development of incremental subgradient methods for dual optimization, and the analysis of their advantages over classical subgradient methods.

We provide some orientation by informally summarizing the main ideas of each of the above topics.

Min Common/Max Crossing Duality

In this book, duality theory is captured in two easily visualized problems: the min common point problem and the max crossing point problem, introduced in Chapter 2. Fundamentally, these problems revolve around the existence of *nonvertical* supporting hyperplanes to convex sets that are unbounded from above along the vertical axis. When properly specialized, this turns out to be the critical issue in constrained optimization duality and saddle point/minimax theory, under standard convexity and/or concavity assumptions.

The salient feature of the min common/max crossing framework is its simple geometry, in the context of which the fundamental constraint qualifications needed for strong duality theorems are visually apparent, and admit straightforward proofs. This allows the development of duality theory in a unified way: first within the min common/max crossing framework in Chapters 2 and 3, and then by specialization, to saddle point and minimax theory in Chapters 2 and 3, and to optimization duality in Chapter 6. All of the major duality theorems discussed in this book are derived in this way, including the principal Lagrange multiplier and Fenchel duality theorems for convex programming, and the von Neuman Theorem for zero sum games.

From an instructional point of view, it is particularly desirable to unify constrained optimization duality and saddle point/minimax theory (under convexity/concavity assumptions). Their connection is well known, but it is hard to understand beyond a superficial level, because there is not enough overlap between the two theories to develop one in terms of the other. In our approach, rather than trying to build a closer connection between constrained optimization duality and saddle point/minimax theory, we show how they both stem from a common geometrical root: the min common/max crossing duality.

We note that the constructions involved in the min common and max crossing problems arise in the theories of subgradients, conjugate convex functions, and duality. As such they are implicit in several earlier analyses; in fact they have been employed for visualization purposes in the first author's nonlinear programming textbook [Ber99]. However, the two problems have not been used as a unifying theoretical framework for constrained optimization duality, saddle point theory, or other contexts, except implicitly through the theory of conjugate convex functions, and the complicated and specialized machinery of conjugate saddle functions. Pedagogically, it appears desirable to postpone the introduction of conjugacy theory until it is needed for the limited purposes of Fenchel duality (Chapter 7), and to bypass altogether conjugate saddle function theory, which is what we have done.

Existence of Solutions and Strong Duality

We show that under convexity assumptions, several fundamental issues in optimization are intimately related. In particular, we give a unified analysis of conditions for optimal solutions to exist, for the minimax equality to hold, and for the absence of a duality gap in constrained optimization.

To provide a sense of the main idea, we note that given a constrained optimization problem, lower semicontinuity of the cost function and compactness of the constraint set guarantee the existence of an optimal solution (the Weierstrass Theorem). On the other hand, the same conditions plus convexity of the cost and constraint functions guarantee not only the existence of an optimal solution, but also the absence of a duality gap. This is not a coincidence, because as it turns out, the conditions for both cases critically rely on the same fundamental properties of compact sets, namely that the intersection of a nested family of nonempty compact sets is nonempty and compact, and that the projections of compact sets on any subspace are compact.

In our analysis, we extend this line of reasoning under a variety of assumptions relating to convexity, directions of recession, polyhedral sets, and special types of sets specified by quadratic and other types of inequalities. The assumptions are used to establish results asserting that the intersection of a nested family of closed convex sets is nonempty, and that the function $f(x) = \inf_z F(x, z)$, obtained by partial minimization of a convex function F, is lower semicontinuous. These results are translated in turn to a broad variety of conditions that guarantee the existence of optimal solutions, the minimax equality, and the absence of a duality gap.

Pseudonormality and Lagrange Multipliers

In Chapter 5, we discuss Lagrange multiplier theory in the context of optimization of a smooth cost function, subject to smooth equality and inequality constraints, as well as an additional set constraint. Our treatment of Lagrange multipliers is new, and aims to generalize, unify, and streamline the theory of constraint qualifications.

The starting point for our development is an enhanced set of necessary conditions of the Fritz John type, that are sharper than the classical Karush-Kuhn-Tucker conditions (they include extra conditions, which may narrow down the field of candidate local minima). They are also more general in that they apply when there is an abstract (possibly nonconvex) set constraint, in addition to the equality and inequality constraints. To achieve this level of generality, we bring to bear notions of nonsmooth analysis, and we find that the notion of regularity of the abstract constraint set provides the critical distinction between problems that do and do not admit a satisfactory theory.

Fundamentally, Lagrange multiplier theory should aim to identify the essential constraint structure that guarantees the existence of Lagrange multipliers. For smooth problems with equality and inequality constraints, but no abstract set constraint, this essential structure is captured by the classical notion of quasiregularity (the tangent cone at a given feasible point is equal to the cone of first order feasible variations). However, in the presence of an additional set constraint, the notion of quasiregularity breaks down as a viable unification vehicle. Our development introduces the notion of pseudonormality as a substitute for quasiregularity for the case of an abstract set constraint. Pseudonormality unifies and expands the major constraint qualifications, and simplifies the proofs of Lagrange multiplier theorems. In the case of equality constraints only, pseudonormality is implied by either one of two alternative constraint qualifications: the linear independendence of the constraint gradients and the linearity of the constraint functions. In fact, in this case, pseudonormality is not much different than the union of these two constraint qualifications. However, pseudonormality is a meaningful unifying property even in the case of an additional set constraint, where the classical proof arguments based on quasiregularity fail. Pseudonormality also provides the connecting link between constraint qualifications and the theory of exact penalty functions.

An interesting byproduct of our analysis is a taxonomy of different types of Lagrange multipliers for problems with nonunique Lagrange multipliers. Under some convexity assumptions, we show that if there exists at least one Lagrange multiplier vector, there exists at least one of a special type, called informative, which has nice sensitivity properties. The nonzero components of such a multiplier vector identify the constraints that need to be violated in order to improve the optimal cost function value. Furthermore, a particular informative Lagrange multiplier vector characterizes the direction of steepest rate of improvement of the cost function for a given level of the norm of the constraint violation. Along that direction, the equality and inequality constraints are violated consistently with the signs of the corresponding multipliers.

The theory of enhanced Fritz John conditions and pseudonormality are extended in Chapter 6 to the case of a convex programming problem, without assuming the existence of an optimal solution or the absence of a duality gap. They form the basis for a new line of analysis for asserting the existence of informative multipliers under the standard constraint qualifications.

Incremental Subgradient Methods

In Chapter 8, we discuss one of the most important uses of duality: the numerical solution of dual problems, often in the context of discrete optimization and the method of branch-and-bound. These dual problems are often nondifferentiable and have special structure. Subgradient methods have been among the most popular for the solution of these problems, but they often suffer from slow convergence.

We introduce incremental subgradient methods, which aim to accelerate the convergence by exploiting the additive structure that a dual problem often inherits from properties of its primal problem, such as separability. In particular, for the common case where the dual function is the sum of a large number of component functions, incremental methods consist of a sequence of incremental steps, each involving a single component of the dual function, rather than the sum of all components.

Our analysis aims to identify effective variants of incremental methods, and to quantify their advantages over the standard subgradient methods. An important question is the selection of the order in which the components are selected for iteration. A particularly interesting variant uses randomization of the order to resolve a worst-case complexity bottleneck associated with the natural deterministic order. According to both analysis and experiment, this randomized variant performs substantially better than the standard subgradient methods for large scale problems that typically arise in the context of duality. The randomized variant is also particularly well-suited for parallel, possibly asynchronous, implementation, and is the only available method, to our knowledge, that can be used efficiently within this context.

We are thankful to a few persons for their contributions to the book. Several colleagues contributed information, suggestions, and insights. We would like to single out Paul Tseng, who was extraordinarily helpful by proofreading portions of the book, and collaborating with us on several research topics, including the Fritz John theory of Sections 5.7 and 6.6. We would also like to thank Xin Chen and Janey Yu, who gave us valuable feedback and some specific suggestions. Finally, we wish to express our appreciation for the stimulating environment at M.I.T., which provided an excellent setting for this work.

Dimitri P. Bertsekas, dimitrib@mit.edu
Angelia Nedić, angelia.nedich@alphatech.com
Asuman E. Ozdaglar, asuman@mit.edu

Contents