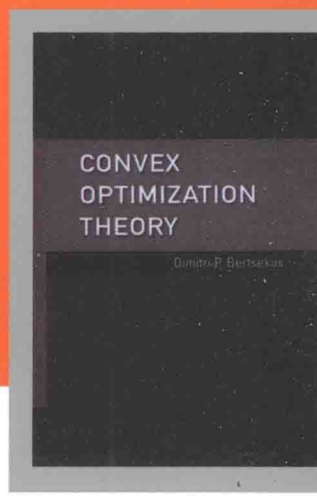
信息技术和电气工程学科国际知名教材中译本系列

Convex Optimization Theory

凸优化理论



[美] Dimitri P. Bertsekas 著

赵千川 王梦迪 译

清华大学出版社

信息技术和电气工程学科国际知名教材中译本系列

Convex Optimization Theory

凸优化理论

[美] Dimitri P. Bertsekas 著
赵千川 王梦迪 译

清华大学出版社
北京

北京市版权局著作权合同登记号　图字：01-2009-6309

Authorized translation from the English language edition, entitled Convex Optimization Theory.
ISBN 978-1886529311 by Dimitri P. Bertsekas, published by Athena Scientific, copyright ©2009.

图书在版编目(CIP)数据

凸优化理论/(美)博塞克斯(Bertsekas, D.P.)著；赵千川，王梦迪译. --北京：清华大学出版社，2015
信息技术和电气工程学科国际知名教材中译本系列
ISBN 978-7-302-39956-8

Ⅰ. ①凸…　Ⅱ. ①博…　②赵…　③王…　Ⅲ. ①凸分析　Ⅳ. ①O174.13

中国版本图书馆 CIP 数据核字(2015)第 086761 号

责任编辑：王一玲
封面设计：常雪影
责任校对：焦丽丽
责任印制：李红英

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
课 件 下 载：http://www.tup.com.cn，010-62795954
印 装 者：三河市金元印装有限公司
经　销：全国新华书店
开　本：153mm×234mm　**印　张**：15.5　**字　数**：283 千字
版　次：2015 年 11 月第 1 版　**印　次**：2015 年 11 月第1 次印刷
印　数：1～1500
定　价：49.00 元

产品编号：035002-01

《信息技术和电气工程学科国际知名教材中译本系列》

出 版 说 明

三年多以前，2000 年 10 月，为了系统地参考和借鉴国外知名相关大学教材，推进我国大学的课程改革和我国大学教学的国际化进程，清华大学出版社策划、出版了《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》，至今已经出版了 30 多种，深受高等院校信息技术与电气工程及相关学科师生和其他科技人员的欢迎和好评，在学术界和教育界产生了积极的影响. 现在这个系列中的大部分教材都已经重印，并曾获得《2001 年引进版优秀畅销丛书奖》. 在此期间，我们曾收到来自各地高校师生的很多反映，期望我们选择这个系列中的一些较为基础性和较为前沿性的教材译成中译本出版，以为更广大的院校师生和科技人员所选用. 正是基于这种背景和考虑，清华大学出版社决定进一步推出《信息技术和电气工程学科国际知名教材中译本系列》.

这套国际知名教材中译本系列所选书目的范围，限于信息技术和电气工程学科所属各专业的技术基础课和主要专业课. 教材原版本除了选自《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》外，还将精选其他具有较大影响的国外知名的相关领域教材或教学参考书. 教材内容适于作为我国普通高等院校相应课程的教材或主要教学参考书.

本国际知名教材中译本系列按分期分批的方式组织出版. 为了便于使用这套国际知名教材中译本教材系列的相关师生和科技人员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，在每种中译本教材中都附有我们约请的相关领域资深教授撰写的推荐说明，其中的一些直接取自于《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》中的影印版序.

本国际知名教材中译本系列的读者对象为信息技术和电气工程学科所属各专业的本科生或研究生，同时兼顾其他工程学科专业的本科生或研究生. 既可采用作为相应课程的教材，也可作为相应课程的教学参考书. 此外，

本国际知名教材中译本系列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.

感谢使用本国际知名教材中译本系列的广大师生和科技人员的支持. 期望广大读者提出意见和建议.

郑大钟 教授
清华大学信息科学技术学院
2003 年 12 月

致中国读者

欣闻本人的著作被译成中文和读者见面了. 在过去的几年间，我有幸几次造访中国，因此也和非常活跃的中国学术界熟悉起来，结交了不少在中国的大学里教书的朋友.

中国年轻学生的禀赋和进取心的确令人印象深刻. 他们当中有几位成为我的博士生和研究伙伴是我的运气. 因此我高兴中国学者现在可以用自己的母语来阅读我的著作《凸优化理论》了. 在此向译者表示由衷的感谢. 做个小广告，我的其他几本著作《非线性规划》、《网络优化》和《概率导论》也已经有中译本了.

Dimitri P. Bertsekas

2015.5.12

作 者 简 介

Dimitri P. Bertsekas 毕业于希腊雅典国立技术大学，主修机械与电气工程专业，在麻省理工学院系统科学专业获得博士学位. 他曾经在斯坦福大学工程与经济系统系、伊利诺伊大学香槟分校电气工程系任教. 从 1979 年起，他在麻省理工学院电气工程与计算机科学系任教，目前是 McAfee 工程讲座教授.

他的教学科研领域包括：确定性优化、动态规划与随机控制、大规模及分布式计算以及数据通信网络. 他发表和合著了大量研究论文，出版专著 14 本，其中部分专著被麻省理工学院作为教材使用，包括《非线性规划》、《动态规划与最优控制》、《数据网络》、《概率论入门》以及本书. 他经常为企业进行咨询，并为若干学术期刊做编辑工作.

由于在他的著作《神经元动态规划》(与 John Tsitsiklis 合著) 中反映出的在运筹学与计算机科学结合方面的出色研究成果，Bertsekas 教授获得了 1997 年的 INFORMS 奖. 他还因运筹学研究获得过 2000 年度希腊国家奖章和 2001 年 ACC John R. Ragazzini 教育奖. 2001 年, 他当选为美国工程院院士.

译 者 序

凸优化理论是非线性规划研究领域的核心成果，也是研究一般非线性规划问题的理论基础.

本书力图以简洁的篇幅，介绍凸优化的一个完整理论分析框架. 凸优化理论的基石在于对偶. 优化问题的对偶有多种形式，而对偶的本质在于闭的凸集有两种等价的描述方式：用该集合包含的所有点的并集来描述，或用超平面描述，也即凸闭集等于所有包含它的闭半空间的交集. 作者选取了最小公共点/最大相交点的几何框架 (简称为 MC/MC 框架) 作为凸优化问题的对偶性分析的基础框架. 相比于基于函数共轭性的代数框架，MC/MC 框架更适用于直观地分析和理解各种重要的优化问题，也更适合初学者学习和理解凸优化理论.

本书的主要内容包括：凸分析的基本概念、多面体凸性、凸优化的基本概念、对偶原理的几何框架、对偶性在优化中的运用等.

与凸优化相关的国内外许多其他的优秀教材相比，本书的一个突出特色在于：以低维空间的例子贯穿全书，直观地解释对偶性等抽象的概念. 这在启发读者开展更深入的理论研究方面，有独特的价值.

本书可以作为高年级本科生、研究生“运筹学优化类”课程的教材或相关研究人员的参考书.

原著作者美国工程院院士 Dimitri P. Bertsekas 教授，有极高的学术造诣和学术声誉，在学术专著和教材的写作方面取得了公认的成就. 限于译者水平，中译本错误纰漏之处还恳请读者指评指正.

译 者

2015 年 1 月

前　　言

本书的目标是给出以下两个主题的易懂、简洁和直观的展示.

(a) 凸分析，特别是与优化的联系.

(b) 优化与最小最大问题的对偶理论，特别是在凸性框架中的情形.

它们是在广泛的实际应用中相关的两个主题.

优化的重点在于推导出约束问题存在原始和对偶最优解的条件. 约束问题的例子是

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

其他类型的优化问题，包括从 Fenchel 对偶性产生的问题，也属于我们考虑的范围. 最小最大问题的重点是推导保证等式

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

成立，以及下确界 "inf" 和上确界 "sup" 可取到的条件.

凸性的理论内容介绍得比较详细. 囊括了这个领域几乎所有重要的方面，对于凸优化中核心的分析问题的展开是足够了. 数学预备知识是线性代数和实分析的入门知识. 附录中包含了用到的有关知识的总结. 除了这些少量背景外，本书的内容是自足的，严格的证明会贯穿全书. 线性和非线性优化理论的先修知识不是必需的，尽管作为背景知识无疑它们是有帮助的.

我们的目标是尽量发挥凸性理论在以一种统一的方式建立最强的对偶性方面的作用. 为此，我们的分析常会偏离 Rockafellar 1970 年的经典著作的思路，而是遵从 Fenchel/Rockafellar 的框架. 例如，我们采用不同的方式来处理闭集相交理论和线性变换下闭包的保持 (1.4.2 和 1.4.3 节)；我们用约束优化情形下的对偶性来发展次微分运算 (5.4.2 节)；此外，我们没有

使用下确界卷积 (infimal convolution)、函数图像 (image)、极性集合和函数 (polar sets and functions)、双函数 (bifunctions) 和共轭鞍点函数 (conjugate saddle functions) 等概念. 类似于 Fenchel/Rockafellar，我们的理论体系是基于 Legendre/Fenchel 共轭的思想，不过相比之下，在几何和可视化方面要来得直观得多.

我们的对偶框架是基于两个简单的几何问题：最小公共点问题 (min common point problem) 和最大相交点问题 (max crossing point problem). 最小公共点/最大相交点 (MC/MC) 框架的突出优点在于其几何上的直观性. 借助这个框架，对偶性理论的核心问题都变得显然，并且可以采用统一的方法处理. 我们的方法是先在 MC/MC 框架里得到许多广泛可用的定理，然后把它们用于解决特定问题 (约束优化、Fenchel 对偶性、最小最大问题等) 上. 我们处理所有对偶性问题 (对偶间隙的存在性、对偶最优解的存在性、对偶最优解集合的结构) 和其他问题 (次微分理论、择一定理、对偶间隙估计) 都按照这样的思路.

从根本上说，MC/MC 框架与共轭性框架存在着密切的联系. 也正因为如此，MC/MC 框架在理论上很有用也有一般性. 不过，这两个框架在分析对偶性和提供几何解释上扮演者互补的角色：共轭性强调函数/代数描述，而 MC/MC 强调集合/上图描述. MC/MC 框架更简单，而且看起来更适合可视化和研究强对偶性和对偶最优解的存在性问题. 共轭性框架，由于强调函数的描述，更适合凸函数的数学运算比较复杂，而且共轭函数的计算可以用于分析和计算.

本书源自作者早期的著作 [BNO03](与 A. Nedić 和 A. Ozdaglar 合著)，但具有不同的特点. 2003 年的书内容很多，从结构上更像是学术专著，目标是利用非光滑分析的概念建立凸的和非凸的优化问题之间的联系. 本书的组织与此不同，本书集中于介绍凸优化问题. 尽管有这些区别，两本书在写作风格、数学基础和某些内容上还是有共同之处.

本书各章的内容如下:

第 1 章: 本章给出后续各章描述对偶性理论所需要的全部凸分析工具. 会介绍基本的代数概念，如凸锥、超平面、拓扑概念，如相对内点、闭包、线性变换下闭性的保持，超平面分离. 另外，本章还会给出与对偶和优化相关的特定概念，如回收锥和共轭函数.

第 2 章: 本章介绍多面体凸性概念: 顶点、Farkas 和 Minkowski-Weyl定理及其在线性规划中的应用. 在后续章节中不会用到，首次阅读时可以跳过.

第 3 章: 本章集中在优化的基本概念上：极小值的类型、解的存在性和对偶理论专题，如部分最小化和最小最大理论.

第 4 章: 本章介绍 MC/MC 对偶框架. 我们会讨论它和共轭理论之间的联系，及在约束优化和最小最大问题上的应用. 本章最后给出与强对偶性和对偶最优解存在性有关的应用广泛的定理.

第 5 章: 本章把第 4 章的对偶定理应用到线性规划、凸规划和最小最大理论等专题上. 我们还应用这些定理作为进一步发展凸分析工具的辅助. 这些工具包括强有力的 Farkas 引理的非线性版本、次微分理论、择一定理. 最后一节主要侧重于可分问题，给出非凸问题和对偶间隙的估计.

为了简洁起见，我们略去了教师们可能会感兴趣的一些话题. 例如把理论应用到特定结构的问题; Boyd 和 Vanderbergue 的著作 [BoV04]，以及我和 John Tsitsiklis 合著的关于并行与分布式计算的著作 [BeT89] 包含了这方面的许多材料 (这两本书都可在线访问).

另外一个忽略的重要部分是计算方法. 不过，我补充了一个很长的第 6 章 (超过 100 页)，其中有最常见的凸优化算法 (和一些新算法)，并且可以从本书的网站下载 (http://www.athenasc.com/convexduality.html).

本章和更全面的凸分析、优化、对偶性以及算法等内容一起，将成为作者正在编著的教材的一部分. 到那时，本章将在对偶性之外，为教师们提供凸优化算法内容 (如作者在麻省理工学院所做的). 本章是一个定期更新的“活” 的章节. 它的当前内容是: **算法方面的第 6 章:** 6.1. 问题结构与计算方法; 6.2. 算法中的递减性; 6.3. 次梯度法; 6.4. 多面体近似方法; 6.5. 邻近性和 Bundle 方法; 6.6. 对偶邻近点算法; 6.7. 内点法; 6.8. 近似次梯度法; 6.9. 最优算法和复杂性.

虽然作者没有在书中提供习题，但是在本书的网站上提供了大量的习题 (并附有详细的解答). 读者/教师也可以使用 [BNO03] 中给出的章节后习题 (共 175 道). 这些习题的风格和符号与本书类似. 习题解答在本书的网站可以下载，也可以在线获取 (http://www.athenasc.com/convexity.html).

本书可以作为凸优化理论课程的教材，作者在过去十年在麻省理工学院和其他场所教授过类似课程. 本书也可以作为非线性规划课程的补充材料，或者作为凸优化模型 (而不是理论) 方面课程的理论基础.

本书的组织使得读者/教师可以选择性地使用其中的内容. 例如，第 2 章多面体凸性的材料完全可以略去，因为第 3~5 章完全不涉及这部分内容. 类似地，最小最大理论 (3.4，4.2.5 和 5.5 节) 可以被略去；并且如果是

这样，那么 3.3 和 5.3.4 节这些使用部分最小化工具的内容也可以略去. 另外，5.4~5.7 节处在“末端”，都可以略去而不会影响其他章节. 如作者在麻省理工学院的“非线性规划”课程 (加上网站上补充的关于算法的第 6 章) 上所做的，一种“最小的”选项包含以下内容：

- 第 1 章，除去 1.3.3 和 1.4.1 节.
- 第 3.1 节.
- 第 4 章，除去 4.2.5 节.
- 第 5 章，除去 5.2，5.3.4 和 5.5~5.7 节.

这种组合侧重于非线性凸优化，而完全不涉及多面体凸性和最小最大理论.

作者感谢同事们对本书的贡献. 作者与 Angelia Nedić 和 Asuman Ozdaglar 在他们的 2003 年的著作上的合作为本书打下了基础. Huizhen (Janey) Yu 仔细阅读了本书部分内容的早期书稿，并给出了一些很有启发的建议. Paul Tseng 通过与作者在集合相交理论方面的合作研究，对本书做出了实质性的贡献. 部分体现在 1.4.2 节 (这项研究受到与 Angelia Nedić 的早期合作的启发). 非常感谢 Dimitris Bisias，Vivek Borkar，John Tsitsiklis，Mengdi Wang 和 Yunjian Xu 等学生和同事提供的反馈信息. 最后，作者希望感谢课堂上的许多优秀学生所不断提供的动力和灵感.

目　录

第 1 章　凸分析的基本概念

凸集和凸函数在优化模型中非常有用，是一种便于分析和算法设计的内涵丰富的结构. 这个结构的主体可以归结为几条基本性质. 例如，每个闭的凸集合都可以被支撑该集合的超平面所描述; 凸集边界上的每个点都可以通过该集合的相对内点集来逼近，以及包含于闭凸集的每条半直线当被平移到该集合中的任意一个点发出的时候仍然包含于该集合.

不过，尽管有这些好的性质，凸集及其分析并非完全没有理论和应用上难以处理的异常和例外情况. 例如，不同于仿射的紧集，像线性变换和向量和这样的某些基本运算并不保持闭凸集的闭性不变. 这会使得某些优化问题的处理，包括最优解的存在性和对偶性，变得复杂起来.

因此，有必要认真对待凸集的理论和应用的学习. 第 1 章的目标是建立凸集学习的基础，特别是要强调与优化有关的问题.

1.1　凸集与凸函数

本章将介绍凸集合与凸函数相关的基本概念，这些内容将贯穿本书所有的后续章节. 附录 A 列举了本书将用到的线性代数和实分析的定义、符号和性质. 首先我们给出凸集合的定义如下 (见图 1.1.1).

定义 1.1.1　$\Re^n$ 的子集 C 被称为**凸集**，如果其满足

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

依惯例我们认为空集是凸的. 通常根据问题的背景，我们可容易地判定某特定凸集是否为非空. 然而多数情况下，我们会尽量说明集合是否为非空，从而降低模糊性. 命题 1.1.1 给出了一些保持集合凸性不变的集合变换.

命题 1.1.1　(a) 任意多个凸集 $\{C_i \mid i \in I\}$ 的交集 $\cap_{i\in I} C_i$ 是凸集.

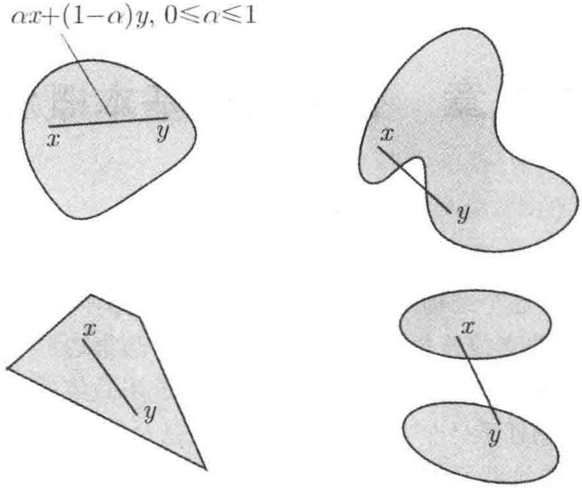


图 1.1.1 凸集的定义. 凸集中任意两点的连线线段都包含在集合内部，因此左图中的集合是凸集，而右图中的不是.

(b) 任意两个凸集 C_1 与 C_2 的向量和 C_1+C_2 是凸集.

(c) 对任意凸集 C 和标量 λ，集合 λC 是凸集. 另外，如 λ_1, λ_2 为正标量，则以下集合是凸的，

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

(d) 凸集的闭包 (closure) 与内点集 (interior) 是凸集.

(e) 凸集在仿射函数下的象和原象是凸集.

证明 证明的思路是直接利用凸集的定义. 在 (a) 中，我们在交集 $\cap_{i\in I}C_i$ 中任取两点 x，y. 由于每个 C_i 都是凸集，x 和 y 间的线段被每个 C_i 所包含, 因而也属于它们的交集.

类似地在 (b) 中，任取 C_1+C_2 中的两点，可以用 x_1+x_2 和 y_1+y_2 表示，其中 $x_1,y_1\in C_1$ 且 $x_2,y_2\in C_2$. 对任意 $\alpha\in[0,1]$ 有如下关系

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

由于 C_1 和 C_2 分别是凸集，上式右侧中两个小括号代表的向量分别属于 C_1 和 C_2，而它们的向量和属于 C_1+C_2. 因此根据定义 C_1+C_2 是凸集. 对 (c) 的证明留给读者作为练习. 对 (e) 可用类似 (b) 的方法来证明.

为证明 (d)，考虑某凸集合 C，以及 C 的闭包中任取的两点 x 与 y. 根据闭包的性质可得，在 C 中存在序列 $\{x_k\}\subset C$ 和 $\{y_k\}\subset C$ 分别收敛