博士丛书

最优化理论与方法

袁亚湘 孙文瑜 著

科学出版社

博 士 丛 书

最优化理论与方法

袁亚湘 孙文瑜 著

科 学 出 版 社

1997

内 容 简 介

本书全面、系统地介绍了无约束最优化、约束最优化和非光滑最优化的理论和计算方法，它包括了近年来国际上关于优化研究的最新成果.

本书可作研究生教材，可供从事计算数学、应用数学、运筹学和计算技术的科研人员参考.

博 士 丛 书
最优化理论与方法
袁亚湘 孙文瑜 著
责任编辑 林 鹏 徐宇星
科学出版社出版
北京东黄城根北街 16 号
邮政编码：100717
中国科学院印刷厂印刷
新华书店北京发行所发行 各地新华书店经售
*
1997 年 1 月第 一 版 开本：850 × 1168 1/32
1997 年 1 月第一次印刷 印张：20 1/2
印数：1—1 600 字数：538 000
ISBN 7-03-005413-x/O·864
定价：41.00 元

序

环顾当今世界，国家的发达，民族的振兴，无一例外地离不开科学技术的推动作用。年轻博士们历来是科技队伍中最活跃、最富创造性的生力军。他们的科研成果是学科发展强有力的推动力量，是体现一个国家高层次教育水平和科研水平的窗口。为了系统地反映年轻博士们的科研成果，促使他们的快速成长，加强国际国内的学术交流，在老一辈科学家的热心支持下，科学出版社决定出版一套《博士丛书》。

我们指导思想是突出本丛书的学术性、创造性、新颖性、先进性和代表性，使之成为所有青年博士平等竞争的学术舞台和优秀科研成果的缩影。

这套丛书以专著为主，并适时组织编写介绍学科最新进展的综述性著作。它将覆盖自然科学各个领域，是一套充分体现我国青年学者科研成果和特色的丛书。

丛书编委会将在由著名科学家组成的专家委员会指导下开展编辑工作。本丛书得到了国家自然科学基金委员会和全国博士后管理协调委员会的特别资助。在此我们深表谢意。

《博士丛书》编委会
一九九三年十月

《博士丛书》专家委员会

前　　言

最优化是一门应用相当广泛的学科，它讨论决策问题的最佳选择之特性，构造寻求最佳解的计算方法，研究这些计算方法的理论性质及实际计算表现. 伴随着计算机的高速发展和优化计算方法的进步，规模越来越大的优化问题得到解决. 因为最优化问题广泛见于经济计划、工程设计、生产管理、交通运输、国防等重要领域，它已受到政府部门、科研机构和产业部门的高度重视.

本书全面、系统地介绍了最优化理论和方法，详细论述了无约束最优化、约束最优化和非光滑最优化的最优性条件、求解方法以及各类求解方法的特点. 作者在本书拟稿时曾打算用一章来介绍线性规划，后发现要想仅用一章系统地介绍线性规划是远远不够的，故本书未对线性规划作介绍. 感兴趣的读者可参阅 Dantzig (1963), Chvàtal (1983), Walsch (1985).

本书的特点之一是内容新，它介绍了近些年来国际上关于最优化研究的许多新的成果. 书中的不少内容是作者在优化科研中取得的结果，例如关于信赖域法、非精确牛顿法、自调比变尺度法、非二次模型方法、非拟牛顿法以及逐步二次规划方面的结果. 本书的另一个特点是理论性强，它深入地探讨了许多算法的收敛性，给出了大量的全局收敛性和局部收敛性结果.

本书可作为研究生教材，也可作为科研人员以及从事实际应用的工程技术人员的参考书.

本书的一至七章由南京大学孙文瑜撰写，作者感谢 J. Stoer, E. Spedicato, Liqun Qi 和胡毓达等教授的支持. 作者的一些研究生对书稿提过很好的建议，也在此致谢.

八至十四章由中国科学院袁亚湘撰写. 作者在此感谢 M.J.D. Powell 和 $\boxed{\text{冯康}}$ 先生、石钟慈教授的关心和鼓励. 作者的学生陈

新对部分书稿进行了认真的校对，也一并表示感谢.

北京航空航天大学王日爽教授对全书手稿进行了认真审阅，并提出了宝贵的修改意见，作者谨向他致以衷心的感谢.

本书的出版得到了中国科学院出版基金的资助，在此表示感谢.

由于水平有限，书中难免有不妥和错误之处，欢迎读者批评指正.

作　者

1995. 12. 30

目　　录

第一章　引　论

§1.1　引　言

最优化理论与方法是一门应用性很强的年轻学科. 它研究某些数学上定义的问题的最优解，即对于给出的实际问题，从众多的方案中选出最优方案.

虽然最优化可以追溯到十分古老的极值问题，然而，它成为一门独立的学科是在本世纪 40 年代末，是在 1947 年 Dantzig 提出求解一般线性规划问题的单纯形法之后. 现在，解线性规划、非线性规划以及随机规划、非光滑规划、多目标规划、几何规划、整数规划等各种最优化问题的理论的研究发展迅速，新方法不断出现，实际应用日益广泛. 在电子计算机的推动下，最优化理论与方法在经济计划、工程设计、生产管理、交通运输等方面得到了广泛应用，成为一门十分活跃的学科.

最优化问题的一般形式为

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } x \in X, \end{aligned} \tag{1.1.1}$$

其中 $x \in R^n$ 是决策变量， $f(x)$ 为目标函数， $X \subset R^n$ 为约束集或可行域. 特别地，如果约束集 $X = R^n$, 则最优化问题 (1.1.1) 称为无约束最优化问题：

$$\min_{x \in R^n} f(x). \tag{1.1.2}$$

约束最优化问题通常写为

$$\begin{aligned}
&\min f(x) \\
&\text{s.t. } c_i(x) = 0, \quad i \in E, \\
&\qquad c_i(x) \geqslant 0, \quad i \in I,
\end{aligned} \tag{1.1.3}$$

这里 E 和 I 分别是等式约束的指标集和不等式约束的指标集，$c_i(x)$ 是约束函数. 当目标函数和约束函数均为线性函数时，问题称为线性规划. 当目标函数和约束函数中至少有一个是变量 x 的非线性函数时，问题称为非线性规划. 此外，根据决策变量、目标函数和要求的不同，最优化还分成了整数规划、动态规划、网络规划、非光滑规划、随机规划、几何规划、多目标规划等若干分支. 本书主要研究求解无约束最优化问题 (1.1.2) 和约束最优化问题 (1.1.3) 的理论和方法，其中第三章至第七章研究无约束最优化问题，第八章至第十三章研究约束最优化问题，第十四章研究非光滑优化问题.

§1.2 数 学 基 础

本节介绍在最优化理论与方法中常常用到的一些数学基础知识.

1.2.1 范数

定义 1.2.1 映射 $\|\cdot\|: R^n \to R$ 称为 R^n 上的半范数，当且仅当它具有下列性质：

(i) $\|x\| \geqslant 0, \forall x \in R^n$.

(ii) $\|\alpha x\| = |\alpha|\|x\|, \forall \alpha \in R, x \in R^n$.

(iii) $\|x + y\| \leqslant \|x\| + \|y\|, \forall x, y \in R^n$.

此外，除了上述三个性质外，如果映射还满足

(iv) $\|x\| = 0 \Leftrightarrow x = 0$,

则 $\|\cdot\|$ 称为 R^n 上的范数.

设 $x=(x_1,x_2,\cdots,x_n)^T\in R^n$, 常用的向量范数为

$$\|x\|_\infty=\max_i|x_i|,\quad (l_\infty\text{ 范数}) \tag{1.2.1}$$

$$\|x\|_1=\sum_{i=1}^{n}|x_i|,\quad (l_1\text{ 范数}) \tag{1.2.2}$$

$$\|x\|_2=\left(\sum_{i=1}^{n}x_i^2\right)^{1/2},\quad (l_2\text{ 范数}) \tag{1.2.3}$$

这些都是 l_p 范数的特例. 一般地，对于 $1\leqslant p<\infty$, l_p 向量范数定义为

$$\|x\|_p=\left(\sum_{i=1}^{n}|x_i|^p\right)^{1/p},\quad (l_p\text{ 范数}). \tag{1.2.4}$$

类似于向量范数的定义，可以定义矩阵范数. 设 $A\in R^{n\times n}$, 其诱导矩阵范数定义为

$$\|A\|=\max_{x\neq 0}\left\{\frac{\|Ax\|}{\|x\|}\right\}, \tag{1.2.5}$$

其中 $\|x\|$ 是某一向量范数. 特别，l_1 诱导矩阵范数或列和范数:

$$\|A\|_1=\max_j\{\|a_{\cdot j}\|_1\}=\max_j\sum_{i=1}^{n}|a_{ij}|, \tag{1.2.6}$$

l_∞ 诱导矩阵范数或行和范数:

$$\|A\|_\infty=\max_i\{\|a_{i\cdot}\|_1\}=\max_i\sum_{j=1}^{n}|a_{ij}|, \tag{1.2.7}$$

l_2 诱导矩阵范数或谱范数:

$$\|A\|_2=(\lambda_{A^TA})^{1/2}, \tag{1.2.8}$$

这里 λ_{A^TA} 表示 A^TA 的最大特征值，$a_{\cdot j}$ 表示 A 的第 j 列，$a_{i\cdot}$ 表示 A 的第 i 行. 显然，

$$\|A^{-1}\|=\frac{1}{\min\limits_{x\neq 0}\dfrac{\|Ax\|}{\|x\|}}.$$

此外，对于诱导矩阵范数，我们总有 $\|I\|=1$.

经常采用的矩阵范数还有 Frobenius 范数，它定义为

$$\|A\|_F=\left(\sum_{j=1}^{n}\sum_{i=1}^{n}|a_{ij}|^2\right)^{1/2}=[\operatorname{tr}(A^TA)]^{1/2}, \tag{1.2.9}$$

加权 Frobenius 范数和加权 l_2 范数分别定义为

$$\|A\|_{M,F}=\|MAM\|_F,\quad \|A\|_{M,2}=\|MAM\|_2, \tag{1.2.9a}$$

其中 M 是 $n\times n$ 对称正定矩阵.

如果某个范数 $\|\cdot\|$ 满足

$$\|AB\|\leqslant\|A\|\,\|B\|, \tag{1.2.10}$$

则称范数 $\|\cdot\|$ 满足相容性条件. 容易看出，诱导 p- 范数和 Frobenius 范数满足相容性条件，并且有

$$\|AB\|_F\leqslant\min\{\|A\|_2\|B\|_F,\|A\|_F\|B\|_2\}. \tag{1.2.10a}$$

此外，椭球向量范数也是常用的向量范数. 设 $x\in R^n$, $A\in R^{n\times n}$ 是对称正定矩阵，向量 x 的椭球范数定义为

$$\|x\|_A=(x^TAx)^{1/2}. \tag{1.2.11}$$

直交变换下不变的矩阵范数也是一类重要的矩阵范数. 设 U 为 n 阶直交矩阵，若

$$\|UA\|=\|A\|,$$

则称 $\|\cdot\|$ 为直交不变矩阵范数. 显然，谱范数和 Frobenius 范数是直交不变范数.

关于范数的等价性，我们有

定义 1.2.2 设 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 是 R^n 上任意两个范数. 如果存在 $\mu_1,\mu_2>0$, 使得

$$\mu_1\|x\|_\alpha\leqslant\|x\|_\beta\leqslant\mu_2\|x\|_\alpha,\quad \forall x\in R^n, \tag{1.2.12}$$

则称范数 $\|\cdot\|_\alpha$ 和范数 $\|\cdot\|_\beta$ 是等价的.

特别，我们有

$$\|x\|_2 \leqslant \|x\|_1 \leqslant \sqrt{n}\|x\|_2, \tag{1.2.13}$$

$$\|x\|_\infty \leqslant \|x\|_2 \leqslant \sqrt{n}\|x\|_\infty, \tag{1.2.14}$$

$$\|x\|_\infty \leqslant \|x\|_1 \leqslant n\|x\|_\infty, \tag{1.2.15}$$

$$\|x\|_\infty \leqslant \|x\|_2 \leqslant \|x\|_1, \tag{1.2.16}$$

$$\sqrt{\lambda}\|x\|_2 \leqslant \|x\|_A \leqslant \sqrt{\Lambda}\|x\|_2, \tag{1.2.17}$$

其中 λ 和 Λ 分别为 A 的最小和最大特征值.

设 $\{x_k\}$ 是向量序列，如果

$$\lim_{k\to\infty} \|x_k - x^*\| = 0, \tag{1.2.18}$$

则称序列 $\{x_k\}$ 依范数收敛到 x^*, 简称收敛到 x^*.

在 R^n 中，如果序列 $\{x_k\}$ 满足

$$\lim_{m,l\to\infty} \|x_m - x_l\| = 0,$$

则 $\{x_k\}$ 称为 Cauchy 序列. 这就是说，对任给的 $\varepsilon > 0$, 存在整数 N_ε, 使得每当 $m, l > N_\varepsilon$ 时，就有

$$\|x_m - x_l\| < \varepsilon$$

成立. 在 R^n 中，序列 $\{x_k\}$ 收敛，当且仅当 $\{x_k\}$ 是 Cauchy 序列.

关于范数的几个重要不等式是：

(1) Cauchy-Schwarz 不等式：

$$|x^T y| \leqslant \|x\|\,\|y\|,$$

当且仅当 x 和 y 线性相关时，等式成立.

(2) 设 A 是 $n\times n$ 正定矩阵，则

$$|x^TAy|\leqslant\|x\|_A\|y\|_A,$$

当且仅当 x 和 y 线性相关时，等式成立.

(3) 设 A 是 $n\times n$ 正定矩阵，则

$$|x^Ty|\leqslant\|x\|_A\|y\|_{A^{-1}},$$

当且仅当 x 和 $A^{-1}y$ 线性相关时，等式成立.

(4) Young 不等式：假定 p 和 q 都是大于 1 的实数，$\dfrac{1}{p}+\dfrac{1}{q}=1$，如果 x 和 y 是实数，则

$$xy\leqslant\frac{x^p}{p}+\frac{y^q}{q},$$

当且仅当 $x^p=y^q$ 时，等式成立.

证明 令 $s=x^p,t=y^q$，由算术-几何不等式，

$$xy=s^{1/p}t^{1/q}\leqslant\frac{s}{p}+\frac{t}{q}=\frac{x^p}{p}+\frac{y^q}{q}.$$

此外，当且仅当 $s=t$，即 $x^p=y^q$ 时，等式成立. □

(5) Hölder 不等式：

$$|x^Ty|\leqslant\|x\|_p\|y\|_q=\left(\sum_{i=1}^n|x_i|^p\right)^{1/p}\left(\sum_{i=1}^n|y_i|^q\right)^{1/q},$$

其中 p 和 q 都大于 1 且满足 $\dfrac{1}{p}+\dfrac{1}{q}=1$.

证明 若 $x=0$ 或 $y=0$，则不等式显然成立. 若 x 和 y 均不为零，记

$$\|x\|_p=\left(\sum_{i=1}^n|x_i|^p\right)^{1/p},\quad\|y\|_q=\left(\sum_{i=1}^n|y_i|^q\right)^{1/q},$$