SOUTHEAST
UNIVERSITY PRESS

东南大学出版社

最优化方法基本教程

盛昭瀚
曹　忻

A COURSE IN OPTIMIZATION

Shen Zhaohan　Cao Xin

最优化方法基本教程

盛昭瀚　　曹　忻

东南大学出版社

(苏)新登字第012号

最优化方法基本教程
盛昭瀚　曹　忻编

东南大学出版社出版
南京四牌楼2号
江苏省新华书店发行　东南大学印刷厂印刷
开本850×1168毫米　1/32　印张14　字数361千字
1992年1月第1版　1992年1月第1次印刷
印数：1—1000册

ISBN7—81023—532—X

O·51　　　　定价：4.25元

内 容 简 介

本书以实分析和线性代数为出发点，系统讨论最优化的基本理论和常用算法，重点为对各种实用算法的介绍并适当介绍有影响的新研究成果。书中收有较多例题，书末有习题。

本书为工科和管理学科有关专业硕士研究生教学用书，也可供科技人员、管理工作者以及高年级大学生参考。

出 版 说 明

研究生教育是培养高层次专门人才的一条重要途径。通过研究生阶段的教学，应使研究生在本门学科上掌握坚实的基础理论和系统的专门知识，并具有从事科学研究工作或独立担负专门技术工作的能力。编辑出版能够体现学校研究生教育特色、有较高学术水平的研究生教材，是研究生教育的重要基础工作之一。

一本好的研究生教材，应当富有教育性、系统性、启迪性、学术性和新颖性。这即是说，研究生教材必须符合教学的基本规律，注意理论联系实际；必须系统阐明本门学科所必要的基础理论和专门知识，注意突出基本原理和基本内容；必须着眼于研究生能力的培养，注意启发他们的创造性思维；必须体现较高的学术水平，注意有足够的理论深度；必须充分反映国内外的最新研究动态，注意当代科学技术发展的前沿。所有这些，既是对研究生教材的要求，也是我们组织出版研究生教材所要遵循的原则。当然，使研究生教材能对本学科领域的科研和工程技术人员有较高的参考价值，也是我们追求的目标。

现在出版的教材虽然是作者多年研究生教学的实践与研究的结晶，从选题、审定到编辑出版，我们也都经过了细致认真的工作，但要使一本研究生教材能满足大家之所求，决非易事。限于我们的水平和经验，难免有失当和错误之处，尚祈读者不吝指正。

东南大学研究生院

东南大学出版社

1990年10月

前　　言

近10年来，我们承担了《运筹学》、《数学规划与网络理论》课程的教学工作，对象以东南大学（原南京工学院）自动化研究所与管理学院的控制理论及应用专业的硕士研究生为主，同时还有电力系统、道路交通等专业的研究生及进修教师。

根据多年的教学实践，在原有两轮讲义的基础上，我们整理、编写了这本《最优化方法基本教程》。

在完成这一工作的过程中，我们注意到：

第一、虽然数学规划是运筹学理论与方法中极为活跃的一个分支，但其基本理论分析与常用算法还是相对稳定的，以这些内容为基本素材，较简洁、清晰地介绍给有关专业的工学硕士研究生、大学生及工程技术人员，仍然是重要和有意义的。

第二、对数学规划实用算法及其基本性质的介绍是本书的重点，这是突出应用性的一种安排，正因为如此，本书的书名定为《最优化方法基本教程》。

第三、考虑到工学硕士研究生、大学生和工程技术人员的实际需要与专业基础，我们认为，本书以实分析为出发点是比较适宜的。

第四、在介绍比较基础和经典的内容的同时，本书也适当地介绍了一些有影响的新领域和新方法，这对读者了解最优化方法的发展趋势与动向是有意义的。

第五、为提高教学效果，每章都安排了适量的习题，限于篇幅，详细的习题解答未能编入。

本书授课的参考时数为：第一、二章，16学时；第三、四章，22学时；第五章，8学时；第六章，10学时；第七章，6学时；第八章，6学时。加上考试，共约70学时左右。在对书中内容作若干取舍之后，也可用于40～60学时授课。

承蒙南京大学胡宣达副教授、河海大学刘伟斌副教授详细审阅了本书原稿，并提出了许多宝贵意见，在此谨致衷心的谢意；我们还非常感谢东南大学徐南荣教授、邱成悌副教授、朱凯强副教授、沈理彪同志在本书写作过程中所给予的支持。

对于本书的不足之处，恳请批评指正。

编　者

1990年4月

目　录

第一章　概　论

本章简要介绍数学规划的几个基本概念，如数学规划问题的基本模型和类型，数学规划的最优解以及实际求解数学规划问题算法的基本概念。

1.1　数学规划的基本概念

一、数学规划模型与基本分类

数学规划，就是寻找 n 元单值函数 $f(\boldsymbol{x})$ 的极值点。变量 $\boldsymbol{x}\in R^n$ 可能没有任何限制也可能受约束于有限多个等式或不等式。因此，数学规划的一般模型可表示成

$$
\begin{aligned}
&(\mathrm{MP}) \qquad \min f(\boldsymbol{x}) \qquad (\max f(\boldsymbol{x}))\\
&\qquad\qquad g_i(\boldsymbol{x})\geqslant 0 \qquad i=1,\ \cdots,\ m \qquad\qquad (1\text{-}1)\\
&\qquad\qquad h_j(\boldsymbol{x})=0 \qquad j=1,\ \cdots,\ p
\end{aligned}
$$

这里的 $f(\boldsymbol{x})$ 称为**目标函数**，$g_i(\boldsymbol{x})\geqslant 0$（$i=1,\ \cdots,\ m$），$h_j(\boldsymbol{x})=0$（$j=1,\ \cdots,\ p$）称为**约束条件**，而 $g_i(\boldsymbol{x})$，$h_j(\boldsymbol{x})$ 称为**约束函数**。

应注意：

（1）对同样的自变量变化范围，

$$\min f(\boldsymbol{x})=\max(-f(\boldsymbol{x}))$$

（2）一个等式约束条件 $h_j(\boldsymbol{x})=0$ 等价于两个不等式约束条件 $h_j(\boldsymbol{x})\geqslant 0$ 与 $-h_j(\boldsymbol{x})\geqslant 0$，因此，式（1-1）还可以表示成统一

形式

$$\text{(MP)}\quad \min f(\boldsymbol{x}) \qquad\qquad (1\text{-}2)$$
$$g_i(\boldsymbol{x})\geqslant 0 \quad i=1,\ \cdots,\ m$$

或

$$\text{(MP)}\quad \min_{\boldsymbol{x}\in D} f(\boldsymbol{x}) \qquad\qquad (1\text{-}3)$$
$$D \stackrel{\text{def}}{=} \{\boldsymbol{x}\,|\,g_i(\boldsymbol{x})\geqslant 0,\ i=1,\ \cdots,\ m\}$$

根据模型式（1-2）中有无约束条件，数学规划问题首先可以分为两大类型：

（1）如果对变量无任何约束条件，即研究

$$\text{(MP)}\quad \min_{\boldsymbol{x}\in R^n} f(\boldsymbol{x})$$

则称为**无约束规划问题**；

（2）如果对变量$\boldsymbol{x}$有一定的约束要求，即研究形如式（1-2）的问题，则称为**有约束规划问题**。

另外，从目标函数与约束函数的函数类型看：

（1）如果$f(\boldsymbol{x})$与诸$g_i(\boldsymbol{x})$均为线性函数，则称此时的规划为**线性规划**；

（2）如果$f(\boldsymbol{x})$与诸$g_i(\boldsymbol{x})$中至少有一个为非线性函数，则称其为**非线性规划**；

（3）如果线性规划中某些参数为随机变量，则称其为**随机线性规划**。

此外，出于对一些特殊的$f(\boldsymbol{x})$与$g_i(\boldsymbol{x})$的考虑，还有**几何规划**、**二次规划**等概念，或者在线性规划模型基础上派生出的**目标规划**问题，或由纯量目标函数发展到向量目标函数而出现的**多目标规划**问题以及用于解决多阶段决策问题的**动态规划**等。总之，随着科学技术的不断发展，人们需要解决的规划问题越来越复杂，要求越来越高。相应地，反映在数学规划研究领域，数学规

划问题的类型也越来越丰富。

这里需要指出的是，虽然微分学基本理论已经解决了一般无约束问题极值点的必要和充分条件，且对定义在开集上的连续可微等约束问题，可以用 Lagrange 乘子法求解。但事实表明，至少在以下两个方面，数学规划基础理论和方法的研究已把传统的函数极值分析工作推向了一个崭新的阶段：

第一，规划问题中不等式约束，特别是非严格不等式约束的出现，使得对极值点条件的分析大大复杂化，需要通过新的概念和思想来解决这一问题；

第二、现代科学技术与生产的实际需要，电子计算机技术的飞速发展，迫切要求在数学规划研究中发展各种基于迭代运算的数值方法以获得高精度的极值点的近似解。

二、最优解的基本概念

为不失普遍性，现讨论

$$(\mathrm{MP}) \qquad \min_{x \in D} f(\boldsymbol{x})$$

$$D \stackrel{\mathrm{def}}{=\!=} \{\boldsymbol{x} \mid g_i(\boldsymbol{x}) \geqslant 0, \ i=1, \ \cdots, \ m\}$$

定义 1-1-1 称$D \subset R^n$为(MP)的**可行域（容许域）**，并称D中的$\boldsymbol{x}$为(MP)的**可行解（容许解）**。

定义 1-1-2 若$\exists \boldsymbol{x}^* \in D$，并对$\forall \boldsymbol{x} \in D$有

$$f(\boldsymbol{x}) \geqslant f(\boldsymbol{x}^*)$$

则称此$\boldsymbol{x}^*$为(MP)的**全局最优解**(**全局极小值点**)。如这里的“$\geqslant$”改为“$>$”，则上述定义名前冠以“**严格的**”，$\boldsymbol{x}^*$组成的集称为（MP）的**最优解集**。

定义 1-1-3 若$\exists \boldsymbol{x}^* \in D$与$\boldsymbol{x}^*$的$\varepsilon$—邻域

$$N_\varepsilon(\boldsymbol{x}^*) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}^*\| \leqslant \varepsilon, \ \varepsilon > 0\} \tag{1-4}$$

并对 $\forall \boldsymbol{x}\in D\cap N_\varepsilon(\boldsymbol{x}^*)$有

$$f(\boldsymbol{x})\geqslant f(\boldsymbol{x}^*)$$

则称$\boldsymbol{x}^*$为(MP)的**局部最优解（局部极小值点）**。

显然，(MP) 的任意全局最优解必为其局部最优解。另外，不难证明下面的定理。

定理 1-1-1 若 (MP)的 $f(\boldsymbol{x})$，$g_i(\boldsymbol{x})(i=1, \cdots, m)$为 R^n 上的连续函数，则其可行域与最优解集均为闭集。

定义 1-1-4 设 $\boldsymbol{x}\in D$, $\boldsymbol{h}\in R^n$，若$\exists\delta>0$，使对 $\forall\alpha\in(0, \delta)$，向量$\boldsymbol{x}+\alpha\boldsymbol{h}$均在$D$的内部，则称$\boldsymbol{h}$为点$\boldsymbol{x}$的一个**可行方向(容许方向）**。

由方向导数的概念知，若点$\boldsymbol{x}^*$为$f(\boldsymbol{x})$的极小点，$\boldsymbol{h}$为$\boldsymbol{x}^*$的任一可行方向，则有

$$\frac{\partial f}{\partial \boldsymbol{h}}=\boldsymbol{h}^T\nabla f(\boldsymbol{x}^*)\geqslant 0 \tag{1-5}$$

反之，根据下面的 Taylor 公式

$$f(\boldsymbol{x}^*+\Delta\boldsymbol{x})=f(\boldsymbol{x}^*)+\Delta\boldsymbol{x}^T\nabla f(\boldsymbol{x}^*)+\frac{1}{2}\Delta\boldsymbol{x}^T\nabla^2 f(\boldsymbol{x}^*)\Delta\boldsymbol{x}+o(\|\Delta\boldsymbol{x}\|^2) \tag{1-6}$$

易得：

定理 1-1-2 设(MP)式(1-3)中，$\boldsymbol{x}^*\in D$, $f\in c^2(D)$，对于$\boldsymbol{x}^*$的任一可行方向$\boldsymbol{h}$，均有 $\boldsymbol{h}^T\nabla f(\boldsymbol{x}^*)\geqslant 0$ 且$\nabla^2 f(\boldsymbol{x}^*)>0$，则$\boldsymbol{x}^*$为(MP)的严格局部最优解。

1.2 凸规划及其基本性质

一、凸集

定义 1-2-1 设 $C\subset R^n$，若对任意的$\boldsymbol{x}^{(1)}$，$\boldsymbol{x}^{(2)}\in C$，以及$\lambda\in[0, 1]$，恒有

$$\lambda x^{(1)}+(1-\lambda)x^{(2)}\in C \tag{1-7}$$

则称集C为**凸集**，并称连接$x^{(1)}$与$x^{(2)}$的线段$\lambda x^{(1)}+(1-\lambda)x^{(2)}$为$x^{(1)}$与$x^{(2)}$的**凸组合**。

例 1 空集ϕ、单点集以及全空间R^n均为凸集。

例 2 设$a\neq 0$为已知向量，b为已知数，集

$$C=\{x\,|\,a^Tx=b,\ x\in R^n\} \tag{1-8}$$

为凸集，并称这样生成的集C为**超平面**。

现给出凸集的某些简单的代数性质与几何性质。

定义 1-2-2 设$x^{(i)}\in R^n$ $(i=1,\ \cdots,\ p)$，$\lambda_i\geqslant 0$，$\sum_{i=1}^{p}\lambda_i=1$，则称

$$x=\sum_{i=1}^{p}\lambda_i x^{(i)} \tag{1-9}$$

为$x^{(i)}$ $(i=1,\ \cdots,\ p)$的**凸组合**。

定理 1-2-1 集$C\subset R^n$为凸集的充要条件为对$\forall s>1$，$x^{(i)}\in C$ $(i=1,\ \cdots,\ s)$的任意凸组合仍在C中。

证明 由$s=2$知充分性显然，往证必要性。首先对$s=2$定理显然正确，现设它对$s\geqslant 2$正确，令

$$x=\sum_{i=1}^{s+1}\lambda_i x^{(i)} \qquad \lambda_{s+1}\neq 0 \tag{1-10}$$

由归纳假设知

$$z\overset{\text{def}}{=\!=}\frac{\lambda_1}{1-\lambda_{s+1}}x^{(1)}+\cdots+\frac{\lambda_s}{1-\lambda_{s+1}}x^{(s)} \tag{1-11}$$

在C中，于是

$$x=(1-\lambda_{s+1})z+\lambda_{s+1}x^{(s+1)}\in C \tag{1-12}$$

另外，易证：

定理 1-2-2 任意多个凸集的交仍为凸集。

定义 1-2-3 包含任意集合$X\subset R^n$的所有凸集之交称为X的**凸包**，记为$C_0(X)$。

由定理 1-2-2，显然凸包为凸集，并且不难看出，$C_0(X)$实际上是R^n中包含X的最小凸集（见图 1-1）。进一步地，能够证明$C_0(X)$是由X中任意有限个点的所有凸组合生成的集，即

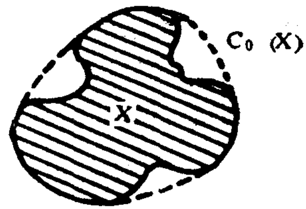


图 1-1

$$C_0(X)=\left\{\sum_{i=1}^{s}\lambda_i\boldsymbol{x}^{(i)}\,\middle|\,\boldsymbol{x}^{(i)}\in X,\ \lambda_i\geqslant 0,\ \sum_{i=1}^{s}\lambda_i=1,\ s\text{为任意正整数}\right\}\tag{1-13}$$

定义 1-2-4 设$X=\{\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$为$R^n$中$n+1$个点组成的集合，若向量组$\boldsymbol{x}_i-\boldsymbol{x}_0(i=1, \cdots, n)$线性无关，则称此$X$的凸包$C_0(X)$为**单纯形**，并称$\boldsymbol{x}_1, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$为单纯形的**顶点**。

二、凸函数

定义 1-2-5 设$f(\boldsymbol{x})$为定义在非空集$X\subset R^n$上的实函数，若对$\forall\boldsymbol{x}, \boldsymbol{y}\in X$，以及数$\lambda\in[0, 1]$，恒有

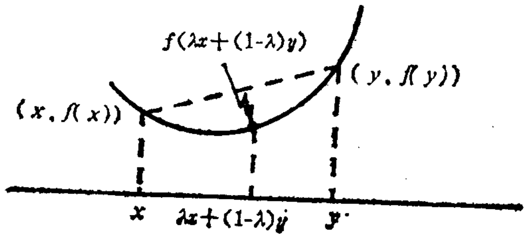
$$f(\lambda\boldsymbol{x}+(1-\lambda)\boldsymbol{y})\leqslant\lambda f(\boldsymbol{x})+(1-\lambda)f(\boldsymbol{y})\tag{1-14}$$

则称$f(\boldsymbol{x})$为X上的**凸函数**。若式(1-14)中的“$\leqslant$”为“$<$”，则称$f(\boldsymbol{x})$为X上的**严格凸函数**。

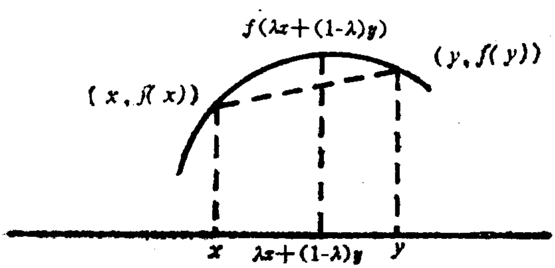
定义 1-2-6 若$-f(\boldsymbol{x})$为X上的凸函数（严格凸函数），则称$f(\boldsymbol{x})$为X上的**凹函数（严格凹函数）**。

对一元函数$f(x)$，由于$\lambda f(x)+(1-\lambda)f(y)$在几何上表示过点$(x, f(x))$与$(y, f(y))$的弦线段，而$f(\lambda x+(1-\lambda)y)$则表示

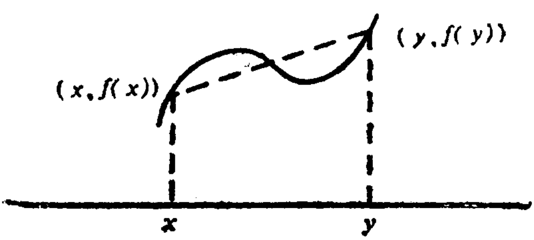
在点 $\lambda x+(1-\lambda)y$处 f 的函数值，因此，凸函数定义表明，该函数任意两点间的曲线弧总是位于这两点间的弦线段之下，而对凹函数来说，正好相反。



凸函数



凹函数



非凸非凹函数

图 1-2

例 3 下面几个函数都是凸函数

(1) $f(x)=x^2-2x$

(2) $f(\boldsymbol{x})=x_1^4+2x_1^2+3x_3^2-2x_2x_3-4x_1 \qquad \boldsymbol{x}\in R^3$

例 4 特别地，线性函数 $f(\boldsymbol{x})=\boldsymbol{c}^T\boldsymbol{x}+b$ 既为凸函数又为凹函数。

可以证明凸函数有如下基本性质：

(1) 设 f 与 g 为 X 上凸函数，则 $f+g$ 也为 X 上的凸函数；

(2) 设 f 为凸函数，λ 为非负数，则 λf 也是凸函数。更一般地，X 上凸函数 f_1，f_2，…，f_k 的非负线性组合即 $\sum_{i=1}^{k}\lambda_i f_i$ ($\lambda_i\geqslant 0$，$i=1$，…，k) 亦为凸函数。

(3) 设 f 为 R^n 上凸函数，$\boldsymbol{x}^{(i)}$ ($i=1$，…，s) 为 R^n 中的点，$\lambda_i\geqslant 0$ ($i=1$，…，s) 且 $\sum_{i=1}^{s}\lambda_i=1$，则

$$f(\lambda_1\boldsymbol{x}^{(1)}+\cdots+\lambda_s\boldsymbol{x}^{(s)})$$
$$\leqslant\lambda_1 f(\boldsymbol{x}^{(1)})+\cdots+\lambda_s f(\boldsymbol{x}^{(s)}) \qquad (1\text{-}15)$$

定理 1-2-3 设 $X\subset R^n$ 为非空开凸集，$f(\boldsymbol{x})$ 在 X 上可微，则 $f(\boldsymbol{x})$ 在 X 上为凸函数（严格凸函数）的充要条件为对 $\forall\boldsymbol{x}\in X$，$\boldsymbol{y}\in X$，恒有

$$f(\boldsymbol{y})+\nabla^T f(\boldsymbol{y})(\boldsymbol{x}-\boldsymbol{y})\leqslant f(\boldsymbol{x}) \qquad (1\text{-}16)$$

$$(f(\boldsymbol{y})+\nabla^T f(\boldsymbol{y})(\boldsymbol{x}-\boldsymbol{y})<f(\boldsymbol{x}))$$

证明 必要性。若 $f(\boldsymbol{x})$ 为 X 上的凸函数，且 $\boldsymbol{x}$，$\boldsymbol{y}\in X$，$\lambda\in[0,1]$，则

$$f(\lambda\boldsymbol{x}+(1-\lambda)\boldsymbol{y}\leqslant\lambda f(\boldsymbol{x})+(1-\lambda)f(\boldsymbol{y})$$
$$=f(\boldsymbol{y})+\lambda(f(\boldsymbol{x})-f(\boldsymbol{y})) \qquad (1\text{-}17)$$

又因为 $f(\lambda\boldsymbol{x}+(1-\lambda)\boldsymbol{y})=f(\boldsymbol{y}+\lambda(\boldsymbol{x}-\boldsymbol{y}))$，故