最优化理论与算法

陈宝林

清华大学出版社

最优化理论与算法

陈 宝 林

清 华 大 学 出 版 社

内 容 简 介

本书包括凸集和凸函数，线性规划原理和方法，对偶理论，最优性条件，无约束最优化方法和约束最优化方法等内容。书中介绍了一些最新研究成果，如Karmarkar算法等，内容比较丰富，算法比较齐全，实用性比较强。定理的证明和算法的推导主要以数学分析和线性代数为基础，简明易学。可作为《线性规划与非线性规划》课程的教学参考书，也可供应用数学工作者和工程技术人员参考。

最优化理论与算法

陈宝林

*

清华大学出版社出版

北京 清华园

商务印书馆上海印刷厂印刷

新华书店北京发行所发行

*

开本: 850×1168 1/32 印张: 17 字数: 440 千字

1989 年 8 月第 1 版 1989 年 8 月第 1 次印刷

印数: 0001-6000

ISBN 7-302-00424-2/O·77

定价: 4.00元

前　言

«最优化理论与算法»是为高等院校开设«线性规划与非线性规划»课程提供的教材。

本书包括凸集凸函数,线性规划和非线性规划三方面的内容。有完整的理论系统,关于凸集凸函数的一些基本定理,线性规划的原理,最优性条件,对偶理论及算法收敛性定理等都做了适度的介绍。书中不仅有大量的实用算法,而且介绍了一些最新研究成果,Karmarkar 算法就是一例。为了使具有大学本科程度的读者能够自学,定理的证明和算法的推导主要以数学分析和线性代数为基础,尽可能少涉及更为高深的知识。本书内容比较丰富,算法比较齐全,有一定的理论深度,层次清晰,叙述简明,便于应用,可作为高等院校教学参考书,也可供应用数学工作者和工程技术人员参考。

本书在编写过程中,得到郑乐宁同志的大力支持,他审阅了原稿,提出了宝贵意见。谭泽光,祁力群和施妙根等同志也给以满腔热情的帮助。借此表示衷心感谢。

由于编者水平有限,缺点和错误在所难免,敬请批评指教。

编　者

目　录

第1章 引　言

§1.1 学科简述

最优化理论与算法是一个重要的数学分支，它所研究的问题是讨论在众多的方案中什么样的方案最优以及怎样找出最优方案。这类问题普遍存在。例如，工程设计中怎样选择设计参数，使得设计方案既满足设计要求又能降低成本；资源分配中，怎样分配有限资源，使得分配方案既能满足各方面的基本要求，又能获得好的经济效益；生产计划安排中，选择怎样的计划方案才能提高产值和利润；原料配比问题中，怎样确定各种成分的比例，才能提高质量，降低成本；城建规划中，怎样安排工厂、机关、学校、商店、医院、住户和其他单位的合理布局，才能方便群众，有利于城市各行各业的发展；农田规划中，怎样安排各种农作物的合理布局，才能保持高产稳产，发挥地区优势；军事指挥中，怎样确定最佳作战方案，才能有效地消灭敌人，保存自己，有利于战争的全局；在人类活动的各个领域中，诸如此类，不胜枚举。最优化这一数学分支，正是为这些问题的解决，提供理论基础和求解方法，它是一门应用广泛、实用性强的学科。

最优化是个古老的课题。长期以来，人们对最优化问题进行探讨和研究。早在17世纪，英国的伟大科学家Newton发明微积分的时代，已经提出极值问题，后来又出现Lagrangian乘数法。1847年法国数学家Cauchy研究了函数值沿什么方向下降最快的问题，提出最速下降法。1939年苏联数学家Л. В. Канторович提出了解决下料问题和运输问题这两种线性规划问题的求解方法。

人们关于最优化问题的研究工作，随着历史的发展不断深入。但是，任何科学的进步，都受到历史条件的限制，直至本世纪三十年代，最优化这个古老课题并未形成独立的有系统的学科。

四十年代以来，由于生产和科学研究突飞猛进地发展，特别是电子计算机日益广泛应用，使最优化问题的研究不仅成为一种迫切需要，而且有了求解的有力工具。因此最优化理论和算法迅速发展起来，形成一个新的学科。至今已出现线性规划、整数规划、非线性规划、几何规划、动态规划、随机规划、网络流等许多分支。最优化理论和算法在实际应用中正在发挥越来越大的作用。

§1.2 线性规划与非线性规划问题

本书重点介绍线性规划与非线性规划，它们是最优化理论的重要分支，也是最基本的部分，许多实际问题抽象成数学模型后，可以归结为求解这类问题。下面先来研究几个例题。

例 1.2.1 农业生产计划安排问题

某村计划在100公顷土地上种植 A、B、C 三种农作物，可提供的劳力、粪肥和化肥等资源的数量，种植每公顷作物所需这三种资源的数量以及能获得的利润如表 1-1：

表 1-1

	用工	粪肥(吨)	化肥(千克)	利润(元)
A	450	35	350	1500
B	600	25	400	1200
C	900	30	300	1800
可提供资源	63000	3300	33000	

其中一个劳动力干一天为1个工。试确定三种农作物的种植面积，使得总利润最大。

下面分析怎样建立数学模型。

设农作物 A、B、C 的种植面积分别为 x_1、x_2 和 x_3 公顷。显然，总利润的表达式是

$$1500x_1+1200x_2+1800x_3 \tag{1.2.1}$$

确定种植面积时，资源限制有 4 种：

1．土地限制，总的种植面积为 100 公顷，即

$$x_1+x_2+x_3=100 \tag{1.2.2}$$

2．劳力限制，种植三种农作物用工之和不能超过允许值 63000 个工，即

$$450x_1+600x_2+900x_3\leqslant 63000 \tag{1.2.3}$$

3．粪肥限制，即

$$35x_1+25x_2+30x_3\leqslant 3300 \tag{1.2.4}$$

4．化肥限制，即

$$350x_1+400x_2+300x_3\leqslant 33000 \tag{1.2.5}$$

此外，种植面积不能为负数，即

$$x_1,\ x_2,\ x_3\geqslant 0 \tag{1.2.6}$$

综上所述，我们的问题就是在条件(1.2.2)至(1.2.6)的限制下，求最优解 x_1、x_2、x_3，使(1.2.1)最大。因此，问题的数学模型写作

$$\begin{aligned}
\max\quad & 1500x_1+1200x_2+1800x_3\\
\text{s.t}\quad & x_1+x_2+x_3=100\\
& 450x_1+600x_2+900x_3\leqslant 63000\\
& 35x_1+25x_2+30x_3\leqslant 3300\\
& 350x_1+400x_2+300x_3\leqslant 33000\\
& x_1,\ x_2,\ x_3\geqslant 0
\end{aligned}$$

其中，max 是 maximize 的简写，读作"极大化"。s. t 是 Subject to 的简写，读作"受限制于"或"约束条件是"。(1.2.1)称为目标函数，式(1.2.2)至(1.2.6)称为约束条件。

例 1.2.2 食谱问题

设市场上可买到 n 种不同的食品，第 j 种食品单位售价为 c_j。每种食品含有 m 种基本营养成分，第 j 种食品的每一个单位所含第 i 种营养成分为 a_{ij}。又设每人每天对第 i 种营养成分的需要量不少于 b_i。试确定在保证营养要求条件下的最经济食谱。

下面建立食谱问题的数学模型。设每人每天需要各种食品的数量分别为 x_1, x_2, $\cdots$, x_n。我们的目标是使伙食费用最少，即使

$$c_1x_1+c_2x_2+\cdots+c_nx_n$$

最小。条件是保证用膳者对各种营养成分的基本需要，即满足

$$a_{i1}x_1+a_{i2}x_2+\cdots+a_{in}x_n\geqslant b_i \qquad i=1, \cdots, m$$

因此，数学模型是

$$\begin{aligned}
\min \quad & c_1x_1+c_2x_2+\cdots+c_nx_n \\
\text{s. t} \quad & a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n\geqslant b_1 \\
& a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n\geqslant b_2 \\
& \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
& a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n\geqslant b_m \\
& x_1, x_2, \cdots, x_n\geqslant 0
\end{aligned}$$

其中 min 是 minimize 的简写，读作“极小化”。

例 1.2.3 结构设计问题

以两个构件组成的对称桁架为例(参看图 1.2.1)。

已知桁架的跨度 $2L$，高度 x_2 的上限 H，承受负荷 $2P$，钢管的厚度 T，材料比重 ρ，纵向弹性模数 E 及容许应力 σ_y。试确定钢管的平均直径 x_1 及桁架的高度 x_2，使桁架的重量最小。

桁架的重量

$$G=2\pi\rho Tx_1(L^2+x_2^2)^{\frac{1}{2}} \qquad (1.2.7)$$

它是平均直径 x_1 和高度 x_2 的函数。x_1 和 x_2 的选择不是任意的，必须满足以下几个条件：

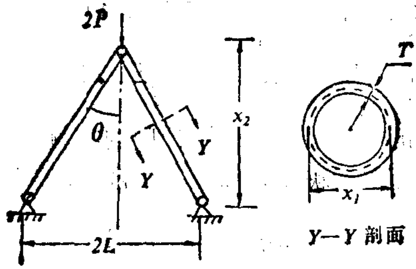


图 1.2.1

1．由于空间限制，要求 x_2 不能超过高度上限 H，即

$$x_2 \leqslant H \tag{1.2.8}$$

2．钢管上的压应力不能超过材料的容许应力 σ_y。在负荷 $2P$ 作用下，钢管承受的压力为

$$F = \frac{P}{\cos\theta} = \frac{P(L^2 + x_2^2)^{\frac{1}{2}}}{x_2}$$

钢管的横截面面积

$$S \approx \pi T x_1$$

由此可知，钢管上的压应力为

$$\sigma(x_1, x_2) = \frac{P(L^2 + x_2^2)^{\frac{1}{2}}}{\pi T x_1 x_2} \tag{1.2.9}$$

因此要求

$$\frac{P(L^2 + x_2^2)^{\frac{1}{2}}}{\pi T x_1 x_2} \leqslant \sigma_y \tag{1.2.10}$$

3．参数的选择还必须保证在负荷 $2P$ 的作用下钢管不发生弯曲，这就要求压应力不超过临界应力 σ_l。临界应力可由欧拉公式算出：

$$\sigma_l = \frac{\pi^2 E(x_1^2 + T^2)}{8(L^2 + x_2^2)} \tag{1.2.11}$$

其中 E 是已知的弹性模数。按此要求应有

$$\frac{P(L^2+x_2^2)^{\frac{1}{2}}}{\pi T x_1 x_2} \leqslant \frac{\pi^2 E(x_1^2+T^2)}{8(L^2+x_2^2)} \qquad (1.2.12)$$

根据以上分析，桁架的最优设计问题，就是求重量函数 G 在上述 3 个约束条件下的极小点问题。它的数学模型是：

$$\begin{aligned} \min \quad & 2\pi\rho T x_1(L^2+x_2^2)^{\frac{1}{2}} \\ \text{s.t} \quad & x_2 \leqslant H \\ & \frac{P(L^2+x_2^2)^{\frac{1}{2}}}{\pi T x_1 x_2} \leqslant \sigma_y \\ & \frac{P(L^2+x_2^2)^{\frac{1}{2}}}{\pi T x_1 x_2} \leqslant \frac{\pi^2 E(x_1^2+T^2)}{8(L^2+x_2^2)} \\ & x_1,\ x_2 \geqslant 0 \end{aligned}$$

例 1.2.4 选址问题

设有 n 个市场，第 j 个市场的位置为 $(a_j,\ b_j)$，对某种货物的需要量为 q_j，$j=1,\ \cdots,\ n$。现计划建立 m 个货栈，第 i 个货栈的容量为 c_i，$i=1,\ \cdots,\ m$。试确定货栈的位置，使各货栈到各市场的运输量与路程乘积之和最小。

现在来建立数学模型。设第 i 个货栈的位置为 $(x_i,\ y_i)$，$i=1,\ \cdots,\ m$。第 i 个货栈供给第 j 个市场的货物量为 W_{ij}，$i=1,\ \cdots,\ m$，$j=1,\ \cdots,\ n$。第 i 个货栈到第 j 个市场的距离为 d_{ij}，一般定义为

$$d_{ij}=\sqrt{(x_i-a_j)^2+(y_i-b_j)^2} \qquad (1.2.13)$$

或

$$d_{ij}=|x_i-a_j|+|y_i-b_j| \qquad (1.2.14)$$

我们的目标是使运输量与路程乘积之和最小，如果距离按式 (1.2.13) 定义，就是使

$$\sum_{i=1}^{m}\sum_{j=1}^{n} W_{ij}\sqrt{(x_i-a_j)^2+(y_i-b_j)^2} \qquad (1.2.15)$$

最小。约束条件是：

1. 每个货栈向各市场提供的货物量之和不能超过它的容量；

2. 每个市场从各货栈得到的货物量之和应等于它的需要量；

3. 运输量不能为负数。

因此，问题的数学模型如下：

$$\min \quad \sum_{i=1}^{m}\sum_{j=1}^{n} W_{ij}\sqrt{(x_i-a_j)^2+(y_i-b_j)^2}$$

$$\text{s.t} \quad \sum_{j=1}^{n} W_{ij}\leqslant c_i \qquad i=1, \cdots, m$$

$$\sum_{i=1}^{m} W_{ij}=q_j \qquad j=1, \cdots, n$$

$$W_{ij}\geqslant 0 \quad i=1, \cdots, m, \ j=1, \cdots, n$$

在上述例1.2.1和例1.2.2的数学模型中，目标函数和约束函数都是线性的，称之为**线性规划问题**；而例1.2.3和例1.2.4的数学模型中含有非线性函数，因此称为**非线性规划问题**。

在线性规划与非线性规划中，满足约束条件的点称为**可行点**，全体可行点组成的集合称为**可行集**或**可行域**。如果一个问题的可行集是整个空间，那么此问题就称为**无约束问题**。

下面给出最优化问题的最优解概念。

定义1.2.1 设$f(x)$为目标函数，S为可行域，$\bar{x}\in S$，若对每一个$x\in S$，成立$f(x)\geqslant f(\bar{x})$，则称$\bar{x}$为极小化问题$\min f(x)$，$x\in S$的**最优解**(**整体最优解**)。

定义1.2.2 设$f(x)$为目标函数，S为可行域，若存在$\bar{x}$的ε邻域

$$N_\varepsilon(\bar{x})=\{x \mid \|x-\bar{x}\|<\varepsilon, \ \varepsilon>0\}$$

使得对每个$x\in S\cap N_\varepsilon(\bar{x})$，成立$f(x)\geqslant f(\bar{x})$，则称$\bar{x}$为极小化问题$\min f(x)$，$x\in S$的**局部最优解**。

对于极大化问题，可类似地定义整体最优解和局部最优解，这

里不再叙述。

根据上述定义，显然，整体最优解也是局部最优解，而局部最优解不一定是整体最优解。但是对于某些特殊情形，如第2章将要介绍的凸规划，局部最优解也是整体最优解。

第2章　凸集与凸函数

凸集和凸函数是线性规划和非线性规划都要涉及到的基本概念。关于凸集和凸函数的一些定理在最优化问题的理论证明及算法研究中具有十分重要的作用。因此这一章关于凸集和凸函数的介绍，将为后面关于最优化理论与算法的研究提供必不可少的基础。

本书对凸集和凸函数只作一般性介绍，要想对这方面的知识有更深入的了解和研究，可参看文献[1]、[2]和[3]。

§2.1　凸　集

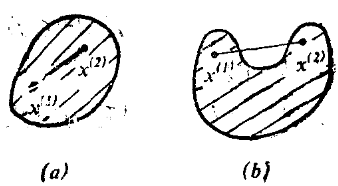
2.1.1　凸集的概念

定义 2.1.1　设 S 为 n 维欧氏空间 E^n 中一个集合。若对 S 中任意两点，连结它们的线段仍属于 S；换言之，对 S 中任意两点 $x^{(1)}$, $x^{(2)}$ 及每个实数 $\lambda\in[0, 1]$，都有

$$\lambda x^{(1)}+(1-\lambda)x^{(2)}\in S$$

则称 S 为凸集。

$\lambda x^{(1)}+(1-\lambda)x^{(2)}$ 称为 $x^{(1)}$ 和 $x^{(2)}$ 的凸组合。图 2.1.1 中，(a) 为凸集，(b) 非凸集。



(a)　　(b)

图　2.1.1

例 2.1.1 集合 $H=\{x|p^Tx=\alpha\}$为凸集，其中，p 为 n 维列向量，α 为实数。

由于对任意两点 $x^{(1)}$, $x^{(2)}\in H$ 及每个实数 $\lambda\in[0, 1]$，都有

$$p^T[\lambda x^{(1)}+(1-\lambda)x^{(2)}]=\alpha$$

因此 $$\lambda x^{(1)}+(1-\lambda)x^{(2)}\in H$$

根据定义 2.1.1 知 H 为凸集。

例 2.1.1 中定义的集合 H 称为 E^n 中的**超平面**，**故超平面为凸集**。

例 2.1.2 集合 $H^-=\{x|p^Tx\leqslant\alpha\}$为凸集。

这是因为对任意的 $x^{(1)}$, $x^{(2)}\in H^-$及每一个数 $\lambda\in[0, 1]$，都有

$$p^T[\lambda x^{(1)}+(1-\lambda)x^{(2)}]=\lambda p^Tx^{(1)}+(1-\lambda)p^Tx^{(2)}\leqslant\alpha$$

所以 $\lambda x^{(1)}+(1-\lambda)x^{(2)}\in H^-$。根据定义 2.1.1 知 H^- 为凸集。

集合 $H^-=\{x|p^Tx\leqslant\alpha\}$称为**半空间**，**故半空间为凸集**。

例 2.1.3 集合 $L=\{x|x=x^{(0)}+\lambda d, \lambda\geqslant0\}$为凸集。其中 d 是给定的非零向量，$x^{(0)}$为定点。

这是因为对任意两点 $x^{(1)}$, $x^{(2)}\in L$ 及每一个数 $\lambda\in[0, 1]$，必有 $x^{(1)}=x^{(0)}+\lambda_1d$, $x^{(2)}=x^{(0)}+\lambda_2d$, λ_1 和 λ_2 是$[0, 1]$中某两个数，从而得出

$$\begin{aligned}&\lambda x^{(1)}+(1-\lambda)x^{(2)}\\&=\lambda(x^{(0)}+\lambda_1d)+(1-\lambda)(x^{(0)}+\lambda_2d)\\&=x^{(0)}+[\lambda\lambda_1+(1-\lambda)\lambda_2]d\end{aligned}$$

以及 $$\lambda\lambda_1+(1-\lambda)\lambda_2\geqslant0$$

因此有 $$\lambda x^{(1)}+(1-\lambda)x^{(2)}\in L$$

根据定义 2.1.1 知 L 为凸集。

集合 $L=\{x|x=x^{(0)}+\lambda d, \lambda\geqslant0\}$称为**射线**，$x^{(0)}$ 为**射线的顶点**。**故射线为凸集**。

运用定义 2.1.1 不难验证下列命题：