大学知识提前学丛书

最优化理论与方法

ZUIYOUHUALILUNYUFANGFA

傅英定 成孝予 唐应辉

在科学研究、工程技术、经济管理工作中，经常需要从多种方案中挑选一种最佳方案，挑选最佳方案的方法就是最优化方法，其理论基础就是最优化理论。掌握了此理论与方法，就可以将大量的实际问题按其内在的规律抽象为某种形式的数学模型，然后利用计算机帮助寻找和判断最佳方案或最优参数……

电子科技大学出版社

大学知识提前学丛书

最优化理论与方法

傅英定　成孝予　唐应辉

电子科技大学出版社

最优化理论与方法

傅英定　成孝予　唐应辉

出　　版:电子科技大学出版社(成都建设北路二段四号)
责任编辑:赵宝如
发　　行:电子科技大学出版社
印　　刷:北京市朝教印刷厂
开　　本:850mm×1168mm　1/32　　印张:12.875　　字数:323 千字
版　　次:1996 年 9 月第一版
印　　次:2005 年 10 月第二次印刷
书　　号:ISBN 7—81043—601—5/O・46
定　　价:32.50 元

前　言

在科学研究，工程技术，经济管理工作中，经常需要从多种方案中挑选一种最佳方案，挑选最佳方案的方法就是最优化方法，其理论基础就是最优化理论。

最优化理论与方法是随着电子计算机的广泛应用而迅速发展起来的一门新的数学分支。它与许多传统数学分支最明显的区别是与计算机关系紧密，同时具有很强的实用性。由于现代科学技术，工程设计与管理问题的日益复杂化，仅靠传统的方法与手段已难以解决。而最优化理论与方法则在许多实际问题与计算机之间架设了一座桥梁，掌握了这种理论与方法，就可以将大量的实际问题按其内在的规律抽象为某种形式的数学模型，然后利用计算机帮助寻找和判断最佳方案或最优参数。实践表明，最优化理论与方法已经在科学研究、工程设计、经济管理中发挥着越来越大的作用并且产生了直接的、巨大的经济效益，它已成为当代科技工作者、工程师和管理决策人员必备的知识。

本书内容包括最优化基础、线性规划、对偶线性规划、无约束最优化方法、约束优化方法、直接搜索的方向加速法、多目标优化、动态规则等内容。除第一章内容为以后各章的基础外，其余各章内容基本独立，读者可根据需要学习全部或部分章节。本书着重介绍当前最优化方法中理论上成熟、并且应用性较强的内容。对于个别证明比较繁难，而又必须引用的理论性结果，我们采用直接引用，指明出处的方法，不在书中详细证明。对于具体的计算方法，我们一般先介绍算法的基本思想，再给出计算步骤或框图，希望能对读者掌握与理解算法的要旨有所帮助。阅读本书的预备知识是微积分与线性代数。

本书是在原讲义《最优化方法及其应用》的基础上，按照国家教委审核批准的《工学硕士研究生最优化方法课程教学基本要求》补充、修改而成的。中国科学院院士林为干教授，范朝勋高级工程师，电子科技大学唐小我教授，对原讲义给予了高度评价。本书由四川联合大学王荫清教授、电子科技大学李正良教授、孙世清教授审稿，他们对书稿提出了许多宝贵意见，并一致认为这是一本适合工科研究生，工科高年级学生和应用数学专业学生使用的好教材。在此，我们向他们并向所有关心、支持该书编写与出版工作的同志表示衷心的感谢。

本书由傅英定主编，各章的编者为：成孝予（一、二、三章）、傅英定（四、五、六章）、唐应辉（七、八章）。

由于编者水平有限，书中难免有错误或不妥之处，恳请读者批评、指正。

编者

目　录

第一章　最优化问题与凸分析基础

在日常生活中，无论做什么事情，总是有多种方案可供选择，并且可能出现多种不同的结果。我们在做这些事情的时候，总是自觉或不自觉地选择一种最优方案，以期达到最优的结果。在现代工程技术与经济管理中，我们有意识地追求最优方案以达到最优结果。这种追求最优方案以达到最优结果的学科就是最优化。寻求最优方案的方法就是最优化方法，这种方法的理论基础就是最优化理论，而凸分析又是最优化理论的基础之一。

§1　最优化问题

所谓最优化问题，用数学语言来说，就是求一个一元函数或多元函数的极值。在微积分中，我们曾经接触过一些比较简单的极值问题。下面通过具体例子来看看什么是最优化问题。

§1.1　最优化问题的例子

例 1　已知热敏电阻的阻值 R 与温度 t 的函数关系为 $R=x_1\exp\left(\dfrac{x_2}{t+x_3}\right)$，$x_1,x_2,x_3$ 为待定参数。通过实验，测得在温度为 t_i 时，阻值为 R_i，得到一组数据 $(t_1,R_1),(t_2,R_2),\cdots,(t_m,R_m)$，问应怎样根据这一组测量数据来确定参数 x_1,x_2,x_3？

当我们将 x_1,x_2,x_3 确定后，就确定了 R 对于 t 的一个函数关系，这个函数在几何上对应于一条平面曲线。但是，这条曲线未必刚好通过那 m 个测量点，一般都要产生偏差，而这种偏差当然越

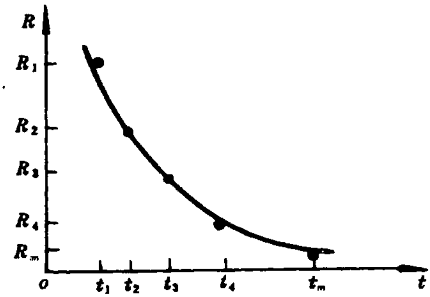

图 1-1

小越好。我们用所有测量点沿铅直方向到曲线距离的平方和来描述这种偏差，则此问题的数学模型为

$$\min \sum_{i=1}^{m} \left[R_i - x_1 \exp\left(\frac{x_2}{t_i + x_3} \right) \right]^2$$

其示意图如图 1-1。

例 2 （运输问题）已知某省煤炭有 m 个产地 $A_1, A_2, \cdots, A_m$，其产量分别为 $a_1, a_2, \cdots, a_m$（吨）。有 n 个销地 $B_1, B_2, \cdots, B_n$，其销量分别为 $b_1, b_2, \cdots, b_n$（吨）。假定产销平衡，即

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$$

由 A_i 到 B_j 的运费为 c_{ij}（元/吨），$(i=1,2,\cdots,m, j=1,2,\cdots,n)$。问在保障供给的条件下，由每个产地到每个销地的运输量为多少吨时，总运费最少。

解 设由 A_i 到 B_j 的运输量为 x_{ij}（吨），则总运费为

$$\sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}$$

其中，x_{ij}应满足以下条件

$$\sum_{j=1}^{n} x_{ij} = a_i \quad (i = 1,2,\cdots,n)$$

$$\sum_{i=1}^{n} x_{ij} = b_j \quad (j = 1,2,\cdots,n)$$

用数学式子来描述，可得以上问题的数学模型为

$$\begin{cases} \min \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \\ \text{s.t.} \quad \sum_{j=1}^{n} x_{ij} = a_i \quad (i = 1,2,\cdots,m) \\ \qquad \sum_{i=1}^{m} x_{ij} = b_j \quad (j = 1,2,\cdots,n) \\ \qquad\qquad x_{ij} \geqslant 0 \end{cases}$$

这里“s.t.”是英文“subject to”的缩写，意为“满足于”。s.t. 后面的式子称为约束条件。

例 3　信号发生仪中正弦波形逼近的优化设计

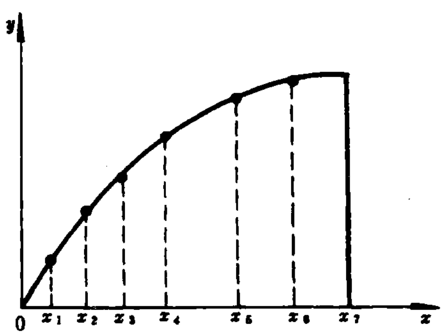

图　1-2

这是一个实际的电路设计问题，要求用折线近似地代替正弦

曲线,并要求在$(0,\frac{\pi}{2})$内确定 6 个点 $x_1,x_2,\cdots,x_6$,使得将$(0,0)$,$(x_1,\sin x_1),\cdots,(x_6,\sin x_6),(\frac{\pi}{2},1)$等点连接所得折线代替$[0,\frac{\pi}{2}]$上的正弦曲线时失真度最小。在数学上,就是使该折线与正弦曲线之间所围成的平面图形面积最小。

正弦曲线与 $x=\frac{\pi}{2}$及 x 轴所围图形的面积为

$$\int_0^{\frac{\pi}{2}}\sin x\,dx = 1$$

折线与 $x=\frac{\pi}{2}$及 x 轴所围图形的面积为

$$\frac{1}{2}\sum_{i=1}^{7}(x_i - x_{i-1})(\sin x_i + \sin x_{i-1})$$

其中,$x_0=0,x_7=\frac{\pi}{2}$。

于是,正弦曲线与折线所围图形的面积为

$$f(x_1,x_2,\cdots,x_6) = 1 - \frac{1}{2}\sum_{i=1}^{7}(x_i - x_{i-1})(\sin x_i + \sin x_{i-1})$$

该问题的数学模型为

$$\begin{cases} \min f(x_1,x_2,\cdots,x_6) \\ \text{s.t. } 0 = x_0 < x_1 < \cdots < x_7 = \dfrac{\pi}{2} \end{cases}$$

或者

$$\begin{cases} \min f(X) \\ \text{s.t.} \quad g_i(X) = x_i - x_{i-1} > 0 \\ \qquad x_i \in \left(0,\dfrac{\pi}{2}\right) \quad (i = 1,2,\cdots,6) \end{cases}$$

其中,$X=(x_1,x_2,\cdots,x_6)^T$。

§1.2　最优化问题的数学模型

以上三个不同类型的最优化问题的共同特点是,求变量 x_1,

$x_2,\cdots,x_n$ 的值，使某函数 $f(x_1,x_2,\cdots,x_n)$的值达到最小，通常将 $f(x_1,x_2,\cdots,x_n)$称为目标函数。

1. 最优化问题的一般形式

$$\begin{cases} \min f(x_1,x_2,\cdots,x_n) \\ \text{s.t. } g_i(x_1,x_2,\cdots,x_n) \geqslant 0 \quad (i=1,2,\cdots,m) \\ \qquad h_j(x_1,x_2,\cdots,x_n) = 0 \quad (j=1,2,\cdots,k) \end{cases} \tag{1-1-1}$$

如果具体问题是求 $\max f(x_1,x_2,\cdots,x_n)$，则可令 $\varphi(x_1,x_2,\cdots,x_n) = -f(x_1,x_2,\cdots,x_n)$，于是最大值问题就转化为最小值问题 $\min \varphi(x_1,x_2,\cdots,x_n)$。

如果约束条件中有 $s_i(x_1,x_2,\cdots,x_n)\leqslant 0$，则可令 $g_i(x_1,x_2,\cdots,x_n)=-s_i(x_1,x_2,\cdots,x_n)$，于是，原来的“$\leqslant$”就转化为“$\geqslant$”。

所以，一般的最优化问题都可以表示为(1-1-1)的形式。其中，$g_i(x_1,x_2,\cdots,x_n)\geqslant 0$ 称为不等式约束，$h_j(x_1,x_2,\cdots,x_n)=0$ 称为等式约束。

2. 最优化问题的向量表达式

为了叙述与讨论的方便，我们也可将(1-1-1)式记为下面的向量形式

$$\begin{cases} \min f(X) \\ \text{s.t. } G(X) \geqslant 0 \\ \qquad H(X) = 0 \end{cases} \tag{1-1-2}$$

其中，$X=(x_1,x_2,\cdots,x_n)^T$

$G(X)=[g_1(X),g_2(X),\cdots,g_m(X)]^T$

$H(X)=[h_1(X),h_2(X),\cdots,h_k(X)]^T$

§1.3 最优化问题的分类

1. 与时间的关系

如果所论及的问题与时间无关，则称为静态问题，否则称为动态问题。

2. 是否有约束条件

有约束的最优化问题称为有约束问题，如例 2，例 3。否则称为无约束问题，如例 1。

3. 函数类型

若目标函数 $f(X)$ 与约束条件中的函数 $g_i(X)$，$h_j(X)$ 都是线性函数，则称此最优化问题为线性规划，如例 2。否则称为非线性规划问题，如例 1，例 3。

§2 梯度与 Hesse 矩阵

§2.1 等值线

二维最优化问题具有明显的几何意义，并且可以将这种几何意义推广到 n 维空间中去，这对于理解最优化理论和掌握最优化方法都是有益的，为此，我们在这里提出等值线的概念及其简单性质。

一般说来，二元函数 $z=f(x,y)$ 在 R^3 中表示一个曲面，该曲面被平面 $y=c$（c 是常数）所截得的曲线 L 的方程为

$$\begin{cases} z = f(x,y) \\ z = c \end{cases}$$

曲线 L 在 xoy 平面上的投影是一条平面曲线 L ，（如图 1-3 所示）。L 在 xoy 平面上的方程是

$$f(x,y) = c$$

对于 L^* 上的所有点 (x,y)，$f(x,y)$ 的值都等于 c，我们将 L 称为函数 $z=f(x,y)$ 的等值线。

在高维（$n\geqslant3$）空间中，使目标函数值取同一常数的点集 $\{X \mid f(X)=c, c \text{ 为一常数}\}$ 称为 $f(X)$ 的等值线（或等值面）。

一般情况下，一个函数的等值线族可具有图 1-4 所示的形状。

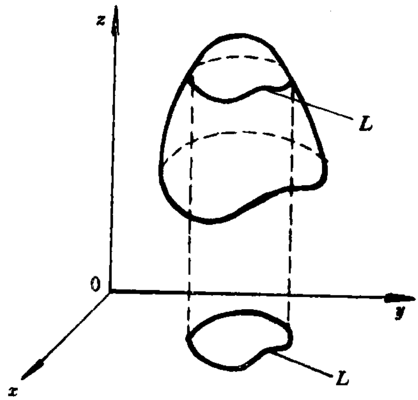

图 1-3

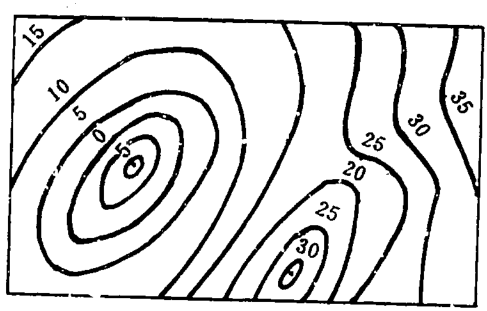

图 1-4

在通常情况下，若目标函数 $f(X)$ 是连续的单值函数，则其等值线具有以下性质：

(1)不同值的等值线不相交；

(2)除极值点所在的等值线外，等值线不会中断；

(3)等值线稠密的地方，目标函数值变化较快，等值线稀疏的地方，目标函数值变化较慢；

(4)在极值点附近，等值线近似地呈现为同心椭圆族。

§2.2 n 元函数的可微性与梯度

1. 可微与梯度的定义

我们以后所要讨论的目标函数多为 n 元可微实值函数。下面给出相应的符号与定义。

$f: D\subset R^n\to R^1$，表示 f 是定义在 n 维空间 R 的子集 D 上的 n 元实值函数。

n 维向量 $X\in D$ 通常表示列向量，为书写方便记为 $X=(x_1,x_2,\cdots,x_n)^T$。

$$\Delta X=(\Delta x_1,\Delta x_2,\cdots,\Delta x_n)^T$$
$$=(x_1-x_1^0,x_2-x_2^0,\cdots,x_n-x_n^0)^T$$
$$X^0=(x_1^0,x_2^0,\cdots,x_n^0)$$
$$f(X)=f(x_1,x_2,\cdots,x_n)$$

定义 1 设 $f: D\subset R^n\to R^1$，且 $X^0\in D$，若存在 n 维向量 L，对于任意 n 维向量 P，都有

$$\lim_{\|P\|\to 0}\frac{f(X^0+P)-f(X^0)-L^TP}{\|P\|}=0 \qquad (1\text{-}2\text{-}1)$$

则称 $f(X)$在 X^0 处可微。

若令

$$\frac{f(X^0+P)-f(X^0)-L^TP}{\|P\|}=\alpha$$

则 $\lim\limits_{\|P\|\to 0}\alpha=0$。于是，(1-2-1)式与下式等价，

$$\begin{aligned}f(X^0+P)-f(X^0)&=L^TP+\alpha(\|P\|)\\&=L^TP+0(\|P\|)\end{aligned} \qquad (1\text{-}2\text{-}2)$$

其中，$0(\|P\|)$是比 $\|P\|$ 高阶的无穷小。

下面的定理 1 给出了定义 1 中 L 的表达式。

定理 1　若 $f(X)$在 X^0 处可微，则 $f(X)$在 X^0 关于各变量的一阶偏导数存在，且定义 1 中的

$$L=\left(\frac{\partial f(X^0)}{\partial x_1},\frac{\partial f(X^0)}{\partial x_2},\cdots,\frac{\partial f(X^0)}{\partial x_n}\right)^T$$

证　设 $L=(l_1,l_2,\cdots,l_n)^T$

$e^i=(0,\cdots,0,1,0,\cdots,0)$，1 是 e 的第 i 个分量。依次取 $P=p_ie^i,(i=1,2,\cdots,n)$，则由(1-2-2)式可得

$$\begin{aligned}f(X^0+P)-f(X^0)&=f(X^0+p_ie^i)-f(X^0)\\&=l_ip_i+0(|p_i|),(i=1,2,\cdots,n)\end{aligned}$$

$$\lim_{\to 0}\frac{f(X^0+p_ie^i)-f(X^0)}{p_i}=l_i,(i=1,2,\cdots,n)$$

此式的左端即$\frac{\partial f(X^0)}{\partial x}$，所以，$\frac{\partial f(X^0)}{\partial x_i}=l_i$，

$$L=\left(\frac{\partial f(X^0)}{\partial x_1},\frac{\partial f(X^0)}{\partial x_2},\cdots,\frac{\partial f(X^0)}{\partial x_n}\right)^T$$

定义 2　依次以 $f(X)$的 n 个偏导数为分量的向量称为 $f(X)$的梯度。记为

$$\nabla f(X)=\left(\frac{\partial f(X)}{\partial x_1},\frac{\partial f(X)}{\partial x_2},\cdots,\frac{\partial f(X)}{\partial x_n}\right)^T$$

由定理 1 可知，L 就是 $f(X)$在 X^0 处的梯度。

$\nabla f(X)$也称为函数 $f(X)$关于向量 X 的一阶导数。

如果 $f(X)$在 X^0 可微，用$\nabla f(X^0)$代替 L，且记 $X=X^0+P$，即 $P=X-X^0$，则(1-2-2)式可记为

$$f(X)-f(X^0)=\nabla f(X^0)^T(X-X^0)+0(\|X-X^0\|) \tag{1-2-2'}$$

2. 梯度的性质

(1)若$\nabla f(X^0)\neq 0$，则$\nabla f(X^0)$必与过 X^0 点的等值线“垂直”。