ZUI YOUHUA JISUAN FANGFA

最优化
计算方法

张光澄
黄世莹
侯泽华 编

成都科技大学出版社

最优化计算方法

张光澄　黄世莹　侯泽华

成都科技大学出版社

内 容 简 介

本书系统地讲述最优化的基本方法和理论，并将其用于阐述工程中的问题。它可作为应用数学专业以及工科其它专业高年级学生和工科研究生的教材，也可供科技人员自学参考。

最优化计算方法

张光澄 黄世玺 侯泽华 编著

成都科技大学出版社出版发行

四 川 省 新 华 书 店 经 销

成都科 技 大 学 印 刷 厂 印 刷

开本 787×1092毫米 1/32 印张：8.0625

1989年9月第1版 1989年9月第1次印刷

印数：1— 字数： 175千字

ISBN 7-0355-X/0.34(课)

定价：1.81元

前　言

最优化方法是对极值问题进行数值分析的手段。极值问题历史悠久而其内容不断地换新，它来源于科学和技术的各个方面，问题的深度和广度也随着各个不同阶段的科学技术水平而有所发展。

早在微积分理论大体完成之前，人们就已经研究过一些简单的极小问题。例如：发射炮弹时仰角多大才能使射程最远？光源离桌面多高时照度最大？等等。

微积分的发展受到极值问题的很大推动。如微分学一开始就被应用于求极大或极小，直到现在，微积分的标准教科书中还有相当篇幅进行极值问题的讨论。

变分学则研究更复杂类型的极值问题，即泛函的极值问题。这里，自变量不是一个或多个变数，而是函数。从历史上看，这类问题直接来自物理学的需要。如光走短程线；肥皂泡张成最小曲面；微观粒子最稳定的状态处于最小能级等等。可以这样讲，许多物理规律都可以通过变分原理概括为极值问题，并且使物理学进入了一个新的阶段。

近二、三十年来，由于科学技术的需要，以及电子计算技术的发展，为求解各种极值问题的最优化方法和理论提供了雄厚的基础和有效手段。最优化的应用范围愈来愈广，几乎在设计、操作、工业过程、生产装置分析乃至生产计划之类的有关问题中，最后都归结为确定某个目标函数取最优值——极大或者极小。由于最优化方法是寻找取得最好效果的条件的方法，所以具有十分重要的现实意义。

本书系统地讲述最优化的基本方法及理论，它可以作为应用数学专业以及工科其它专业高年级学生和研究生的教材，也可供科技人员自学参考。

全书共分十章。第一章基础知识包括凸集，凸函数和多元函数极值的概念，还引入无约束极值的最优性条件和凸函数的极值性质。第二章概述最优化问题的分类和下降算法的特征，并介绍算法的收敛阶，收敛准则，收敛条件等概念。第三章一维寻查是最优化算法的基础。第四、五两章介绍无约束最优化的解析方法。第六章介绍平方和函数的极小算法，是全书相对独立的部分。第七章讨论约束极值的最优性条件与鞍点问题。第八章至第十章介绍约束最优化的计算方法。

由于作者水平有限，缺点和错误在所难免，敬请读者批评指正。

作者

1989年3月

目　　录

前言……………………………………………………3

第一章　基础知识…………………………………………1

§ 1　凸集及其基本性质……………………………1

§ 2　极值的最优性条件……………………………9

§ 3　凸函数及其极值性质…………………………22

第二章　最优化方法概述………………………………26

§ 1　最优化问题的提法及分类……………………26

§ 2　最优化问题举例………………………………27

§ 3　无约束问题算法综述…………………………33

第三章　一维搜索（寻查）……………………………47

§ 1　搜索（寻查）区间的确定……………………49

§ 2　二分法…………………………………………52

§ 3　直接方法………………………………………54

第四章　Newton方法及其改进…………………………67

§ 1　Newton算法及其局限性………………………67

§ 2　Newton算法的改进……………………………71

§ 3　特征值法（Greenstadt方法）…………………73

§ 4　Newton算法的Gill和Murray修正方案…76

第五章　共轭方向法……………………………………82

§ 1　共轭方向………………………………………82

§ 2　共轭方向法……………………………………85

§ 3　共轭梯度法……………………………………94

第六章　非线性最小二乘法……105
§ 1　非线性最小二乘法问题……105
§ 2　Gauss-Newton算法（简称G-N算法）…111
§ 3　修正的G-N算法（Hartley 方法）……122
§ 4　Levenberg-Marquarat算法(简称L-M算法)……125
第七章　约束极值的最优性条件与鞍点问题……140
§ 1　可行方向和起作用约束的概念……140
§ 2　最优性条件……148
§ 3　鞍点问题……163
第八章　可行方向法……169
§ 1　Zoutendijk可行方向法……169
§ 2　投影梯度法……187
§ 3　简约梯度法……208
第九章　罚函数法和障碍函数法……219
§ 1　罚函数法……219
§ 2　障碍函数法……229
第十章　割平面方法……240
参考文献……251

第一章　基础知识

本章扼要介绍最优化方法的理论基础，即凸集和凸函数的概念及基本性质，并提出本书研究的对象和基本方法.

§1　凸集及其基本性质

定义1　非空集合$S\subset R^n$称为**凸集**，若对于任意x_1、$x_2\in S$和任意实数$\lambda\in[0,1]$，使联线

$$\lambda x_1+(1-\lambda)x_2\in S.$$

二维情形如图1-1所示.

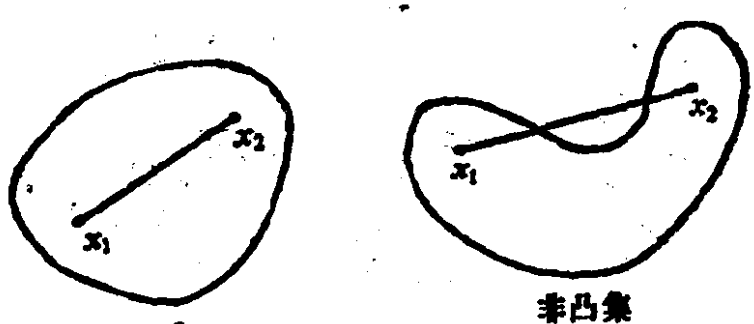


图　1-1

显然，平面上的圆域是凸集合，R^n中以$x^*\in R^n$为心的球$S=\{x\mid\|x-x^*\|<\delta\}$是凸集合.

从定义出发，可以得到凸集的几个简单性质：

设S_1、S_2为凸集，α为一定数，则如下集合

$$\alpha S_1=\{y\mid y=\alpha x,\ x\in S_1\};$$

$$S_1+S_2=\{x\mid x=x_1+x_2,\ x_1\in S_1,\ x_2\in S_2\};$$

$$S_1 \cap S_2 = \{x | x \in S_1, \ x \in S_2\}$$

均为凸集.

定义2 设$S \subset R^n$，称集合$H(S)$为S的**凸包**，若$x \in H(S)$，当且仅当

$$x = \sum_{j=1}^{k} \lambda_j x_j .$$

其中$\sum_{j=1}^{k} \lambda_j = 1$，$\lambda_j \geqslant 0$ $(j=1, \cdots, k)$ k为正整数，$x_j \in S$ $(j=1, \cdots, k)$.

二维情形如图1-2所示.

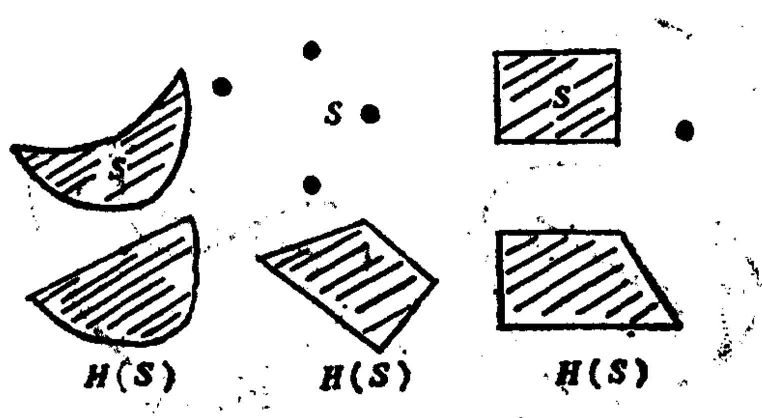
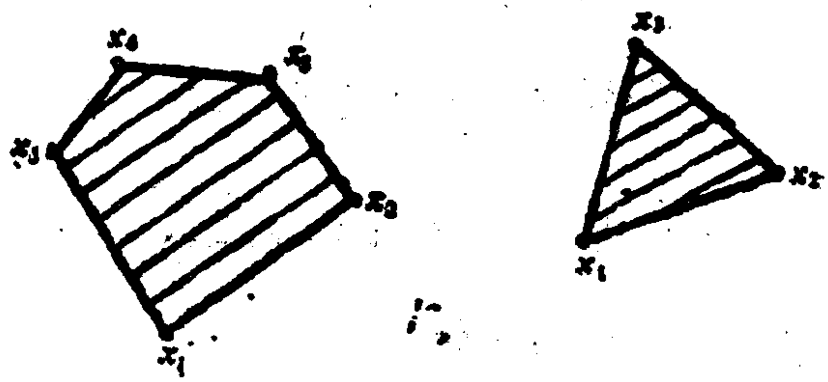


图 1-2

从图1-2看到，$H(S)$是所有包含S的最小凸集.

定义3 R^n中有限个点$x_1, x_2, \cdots, x_{k+1}$所成的凸包$H(x_1, \cdots, x_{k+1})$称为凸多面体.如$x_2 - x_1, x_3 - x_1, \cdots, x_{k+1} - x_1$是线性独立的，则称$H(x_1, \cdots, x_{k+1})$是以$x_1, \cdots, x_{k+1}$为顶点的**单纯形**.

二维情形如图1-3所示.



凸多面体　　　　　单纯形

图 1-3

注意到在R^n中线性独立的向量的最大个数为n，因此，在R^n中不存在多于$n+1$个顶点的单纯形.

下面介绍凸集的几个重要性质：

定理1（投影定理）设S为非空、闭集，$y_0 \in S$，则有$x_0 \in S$使

$$\|x_0 - y_0\| = \inf_{x \in S} \|x - y_0\|$$

（如图1-4）.

证明：设S为有限点集，定理显然成立．若S为无限点集，由下确界定义，对于$k=1, 2, \cdots$有

$$\inf_{x \in S} \|x - y_0\| \leqslant \|x_k - y_0\|$$

$$\leqslant \inf_{x \in S} \|x - y_0\| + \frac{1}{k},$$

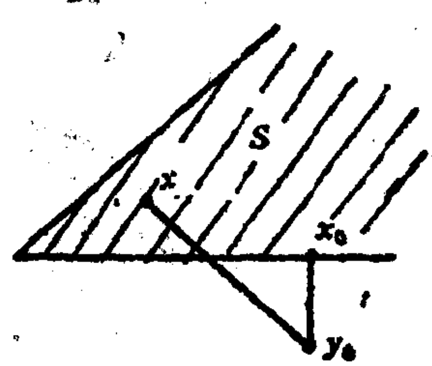


图 1-4

而　$\|x_k\| = \|x_k - y_0 + y_0\|$

$$\leqslant \|x_k - y_0\| + \|y_0\|$$

$$\leqslant \inf_{x\in S}\|x - y_0\| + \frac{1}{k} + \|y_0\|$$

$$\leqslant \inf_{x\in S}\|x - y_0\| + 1 + \|y_0\| \xlongequal{\text{令}} c,$$

故$\{x_k\}$有界，必存在子序列$\{x_{k_i}\}$使

$$\lim_{k_i\to\infty} x_{k_i} = x_0.$$

故S闭，于是$x_0\in S$. 又由

$$\inf_{x\in S}\|x - y_0\| \leqslant \|x_{k_i} - y_0\| \leqslant \inf_{x\in S}\|x - y_0\| + \frac{1}{k_i}.$$

令$k_i\to\infty$有

$$\inf\|x - y_0\| \leqslant \|x_0 - y_0\| \leqslant \inf\|x - y_0\|,$$

所以
$$\|x_0 - y_0\| = \inf_{x\in S}\|x - y_0\|.$$

定理2（分离平面定理）设S非空闭、凸，$y_0\in S$，则存在非零向量p及常数α，使

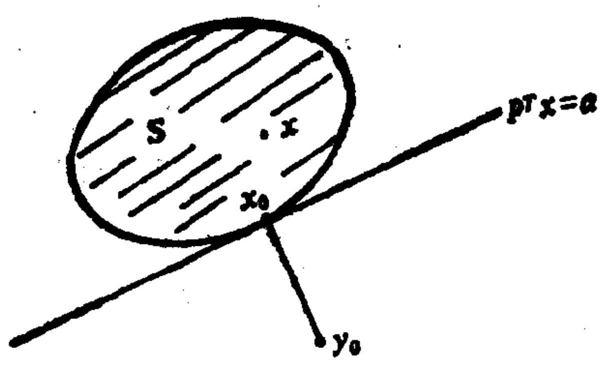


图 1-5

$$p^T x \geqslant \alpha,$$

$$p^T y_0 < \alpha. \qquad \text{（如图1-5）}$$

证明：由定理1知，存在$x_0 \in S$使

$$\|x_0 - y_0\| = \inf_{x \in S} \|x - y_0\| > 0,$$

由S凸，对任给入$\in$（0，1）总有

$$\lambda x + (1-\lambda)x_0 \in S, \qquad x \in S,$$

故 $$\|\lambda x + (1-\lambda)x_0 - y_0\| \geqslant \|x_0 - y_0\|,$$

即 $$\|\lambda x + (1-\lambda)x_0 - y_0\|^2 \geqslant \|x_0 - y_0\|^2,$$

而 $$\|\lambda x + (1-\lambda)x_0 - y_0\|^2 = \|\lambda(x - x_0) + (x_0 - y_0)\|^2$$

$$= \lambda^2 \|x - x_0\|^2 + 2\lambda(x_0 - y_0)^T(x - x_0) + \|x_0 - y_0\|^2,$$

因此 $$\lambda^2 \|x - x_0\|^2 + 2\lambda(x_0 - y_0)^T(x - x_0) \geqslant 0,$$

即 $$\lambda \|x - x_0\|^2 + 2(x_0 - y_0)^T(x - x_0) \geqslant 0.$$

因为 $\lambda \in$（0，1），取λ充分小，总有

$$2(x_0 - y_0)^T(x - x_0) \geqslant 0,$$

即 $$(x_0 - y_0)^T(x - x_0) \geqslant 0,$$

令 $p = x_0 - y_0$，$\alpha = p^T x_0$，

则有 $$p^T x \geqslant \alpha;$$

又 $$\alpha - p^T y_0 = p^T x_0 - p^T y_0 = p^T(x_0 - y_0)$$

$$= (x_0 - y_0)^T(x_0 - y_0)$$

$$= \|x_0 - y_0\|^2 > 0,$$

所以 $$p^T y_0 < \alpha.$$

定理3（Farkas引理）

设A为给定的$m \times n$实矩阵，$b \in R^n$，则$A^T y = b$，$y \geqslant 0$有解的充要条件是存在向量X使

$$\begin{cases} AX \geqslant 0, \\ b^T X \geqslant 0 \end{cases}$$

同时满足．

证明：必要性．设有$y\geqslant 0$使 $A^Ty=b$，则对一切满足$AX\geqslant 0$的向量X，有 $(AX)^Ty\geqslant 0$，

即
$$X^TA^Ty\geqslant 0,$$
从而
$$X^Tb\geqslant 0.$$

充分性．设$S=\{A^Ty\,|\,y\geqslant 0\}$，易证 S 为闭凸集．现证明，若$\begin{cases}AX\geqslant 0,\\ b^TX\geqslant 0\end{cases}$同时成立，则 $A^Ty=b$，$y\geqslant 0$ 有解．事实上，只需证明$b\in S$．

用反证法，设$b\notin S$，由定理2可知，必存在非零向量p及常数α，使对任意$\hat{y}\in S$都有

$$p^T\hat{y}\geqslant\alpha>p^Tb, \qquad (1\text{-}1)$$

易见$o\in S$，因而有　$p^Tb<0$．

下面将证明，对任意$\hat{y}\in S$，又有$p^T\hat{y}\geqslant 0$．　　(1-2)

用反证法。事实上，若存在$\overline{y}\in S$ 使$p^T\overline{y}<0$．因为对任意$k>0$，总有$k\overline{y}\in S$，特别取

$$k>\frac{p^Tb}{p^T\overline{y}}(>0)$$

时，则得到

$$p^T(k\overline{y})<p^Tb, \qquad (1\text{-}3)$$

显见，(1-1) 式与 (1-3) 式矛盾，故 (1-2) 式得证，因此有

$$p^Ty\geqslant 0>p^Tb.$$

若取 $y=e_i=(0, \cdots, 0, 1, 0, \cdots 0)^T$,

$$\underbrace{}_{\text{第}i\text{个分量}}$$

显然，此时$y\geqslant 0$.记

$$A=\begin{pmatrix} a_1^T \\ \vdots \\ a_i^T \\ \vdots \\ a_m^T \end{pmatrix}, \quad \text{故} A^T y = A^T e_i = a_i \in S,$$

由(1-2)式知必有

$$p^T a_i \geqslant 0, \quad i=1, 2, \cdots, n,$$

或
$$a_i^T p \geqslant 0, \quad i=1, 2, \cdots, n,$$

即
$$Ap \geqslant 0.$$

但由于$p^T b<0$，取$X=p$，这便导致与充分性假设相矛盾。故必有$b\in S$，即存在y满足

$$A^T y=b, \quad y\geqslant 0.$$

定理4（择一定理）设矩阵$A_{m\times n}$，$B_{p\times n}$，则

$$Ax>0,$$
$$Bx\geqslant 0$$

不同解的充要条件是存在$\overline{u}\in R^m$，$\overline{V}\in R^p$（$\overline{u}\geqslant 0, \overline{V}\geqslant 0$，且$\overline{u}\neq 0$）使

$$A^T u+B^T V=0.$$

证明：显然，

$$\begin{cases} Ax>0, \\ Bx\geqslant 0 \end{cases}$$

等价于
$$\begin{cases} Ax\geqslant(\alpha, \cdots, \alpha)^T\in R^m, \quad \alpha>0, \\ Bx\geqslant 0. \end{cases}$$

记 $$L=(1,\ \cdots,\ 1)^T\in R^m,$$

令 $$\overline{A}=\begin{pmatrix}A & -L\\ B & 0\end{pmatrix},\quad \overline{b}=(0,\cdots,0,-1)^T\in R^{n+1}$$

易证 $$\begin{cases}Ax\geqslant \alpha L, & \alpha>0\\ Bx\geqslant 0\end{cases}$$

等价于 $$\begin{cases}(x^T,\ \alpha)\overline{A}^T\geqslant 0,\\ (x^T,\ \alpha)\overline{b}<0,\end{cases}$$

即 $$\begin{cases}\overline{A}\begin{pmatrix}x\\ \alpha\end{pmatrix}\geqslant 0,\\ \overline{b}\begin{pmatrix}x\\ \alpha\end{pmatrix}<0.\end{cases}$$

由Farkas引理，上式不同解的充要条件是

存在 $$y=\begin{pmatrix}\overline{u}\\ \overline{v}\end{pmatrix}\geqslant 0,$$

使 $$\overline{A}^T\begin{pmatrix}\overline{u}\\ \overline{v}\end{pmatrix}=\overline{b},\ \overline{u}\in R^m,\ \overline{v}\in R^p,$$

即 $$\begin{pmatrix}A^T & B^T\\ -L^T & 0\end{pmatrix}\begin{pmatrix}\overline{u}\\ \overline{v}\end{pmatrix}=\overline{b}.$$

故得 $$\begin{cases}A^T\overline{u}+B^T\overline{v}=0,\\ -L^T\overline{u}=-1.\end{cases}$$

由 $-L^T\overline{u}=-1$ 得

$$\sum_{i=1}^{m}\overline{u}_i=1,\quad \overline{u}_i\geqslant 0,$$

即 $$\overline{u}\geqslant 0\ \text{且}\ \overline{u}\neq 0,$$

故得 $$A^T\overline{u}+B^T\overline{v}=0,$$

$$\overline{u}\geqslant 0,\ \overline{v}\geqslant 0\quad 且\ \overline{u}\neq 0.$$

综上有 $$\begin{cases} Ax>0, \\ Bx\geqslant 0 \end{cases}$$

不同解的充要条件是存在 $\overline{u}$，$\overline{v}$，使

$$A^T u+B^T\overline{v}=0\quad(\overline{u},\overline{v}\geqslant 0,\overline{u}\neq 0).$$

凸集的上述性质是第九章中研究约束极值最优性条件的基础.

§2 极值的最优性条件

关于一元函数极值的概念在微积分中已有叙述，本节针对多元函数的情形给出各类极小点的定义及局部极小点的判定条件（最优性条件），同时介绍凸函数及其极值性质，进而得到局部极小与全局极小之间的关系.

2·1 多元函数极值概念

定义4 对于任意给定的实数 $\delta>0$，满足不等式 $\|x-x_0\|<\delta$ 的 x 的集合称为**点 x_0 的邻域**，记为 $N(x_0,\ \delta)=\{x\mid\|x-x_0\|<\delta\}$.

定义5 设 f：$S\subset R^n\to R$，若存在一个数 $\delta>0$，对 $\forall x\in N(x^*,\ \delta)\cap S$ 都有

$$f(x^*)\leqslant f(x),$$

则称 x^* 为 $f(x)$ 的**局部极小点**.

若对 $\forall x\in N(x^*,\ \delta)\cap S$，但 $x\neq x^*$ 有 $f(x^*)<f(x)$，则称 x^* 为 $f(x)$ 的严格局部极小点.

定义6 设f：$S\subset R^n\to R^1$，若对$\forall x\in S$都有

$$f(x^*)\leqslant f(x),$$

则称x^*为$f(x)$在S上的**全局极小点**.

若对$\forall x\in S$，但$x\neq x^*$，有$f(x^*)<f(x)$，则称x^*为$f(x)$在S上的严格全局极小点.

由以上定义看到，x^*是局部极小点，是指在邻域$N(x^*,\delta)$中$f(x)$在x^*取得最小的值；而x^*是全局极小点，是指在定义域S中$f(x)$在x^*取得最小值.全局极小点可能在某个局部极小点处取得，也可能在S的边界上取得.

完全类似地可以定义函数局部极大点和全局极大点.

还应该注意到，并非每一个实值函数都有极值，例如定义在R^n上的非零线性函数就没有极值.因此，还需进一步判断$f(x)$在S上是否存在极值的问题，如果存在，又如何判别是极小还是极大.

2·2 梯度与Hesse矩阵

首先引入多元函数可微的概念，在此基础上形成的函数梯度及Hesse矩阵的概念是研究局部极值点的重要工具.

定义7 设f：$S\subset R^n\to R^1$，$x_0\in S$，称$f(x)$在x_0处可微，若存在向量$L\in R^n$，对于任意的$P\in R^n$，使

$$\lim_{h\to 0}\frac{f(x_0+P)-f(x_0)-L^TP}{\|P\|}=0 \qquad (2\text{-}1)$$

成立.

易看出，(2-1)式等价于关系式

$$f(x_0+P)=f(x_0)+L^TP+o(\|P\|) \qquad (2\text{-}2)$$

因此，我们称关于P的线性函数L^TP为函数增量$f(x_0+P)-f(x_0)$的线性主部，即$f(x)$在x_0处的微分.下面的定理给出(2-1)式中L的表示式.