近代优化方法

西 安 交 通 大 学 数 学 研 究 生 教 学 丛 书

徐成贤　陈志平　李乃成 编著

科 学 出 版 社

西安交通大学数学研究生教学丛书

近代优化方法

徐成贤　陈志平　李乃成　编著

科学出版社

2002

内 容 简 介

本书对非线性最优化的算法及相关技术和理论作了比较系统介绍．全书共分七章，第一章讨论以最优性条件为主要内容的最优化基本理论；第二章介绍构成各种最优化算法基本要素的常用数值技术，包括线性方程组求解，矩阵分解与矩阵修正，线性搜索技术及信赖域子问题的求解；第三章至第五章介绍无约束最优化算法，主要有解中小规模最优化问题的拟牛顿方法，大规模优化问题的共轭梯度法，有限内存拟牛顿法，利用非线性最小二乘问题的特殊结构的高斯-牛顿类算法；第六、第七章介绍约束最优化问题的算法，其中第六章主要涉及线性约束优化问题以消去法为主体的可行点算法，第七章介绍一般非线性约束最优化问题的算法，包括罚函数法、乘子法、可行方向法与SQP方法等．

本书可作为计算数学、应用数学、工程领域各专业，文科某些专业如金融、经济等专业的研究生、高年级本科生教学或辅导用书，也用作为从事优化技术应用的工程技术人员的参考用书．

图书在版编目（CIP）数据

近代优化方法/徐成贤、陈志平、李乃成编著．—北京：科学出版社，2002

（西安交通大学数学研究生教学丛书）

ISBN 7-03-010175-8

Ⅰ．近… Ⅱ．①徐… ②陈… ③李… Ⅲ．非线性-最优化算法-研究生-教学参考资料 Ⅳ．O242．23

中国版本图书馆CIP数据核字（2002）第009614号

科学出版社 出版

北京东黄城根北街16号

邮政编码：100717

http://www.sciencep.com

源海印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2002年9月第 一 版　开本：B5（720×1000）

2002年9月第一次印刷　印张：20 3/4

印数：1—3 000　字数：399 000

定价：29.00元

（如有印装质量问题，我社负责调换〈新欣〉）

前　言

最优化问题是在有限种或无限种可行方案（决策）中挑选最优的方案（决策）. 随着高新技术、计算机及信息技术的飞速发展，最优化在工农业、国防、交通、金融、能源、通信等众多领域的应用越来越广泛，问题的规模越来越大、复杂性越来越高. 最优化技术已成为大多数专业领域必修或选修的基本技术课程.

本书旨在对非线性光滑最优化的方法和理论作一个比较系统、全面的介绍. 全书的内容在确保最优化基本理论和方法的基础上，力求尽可能多地包含近期新的研究成果，但由于最优化的内容如此广泛，新的成果如此丰富，做好这一点非作者的能力所及. 内容的选取尽可能避免过分复杂的理论分析，以适应不同专业、不同层次技术人员对最优化技术的需求. 某些定理的证明或理论分析仅在于论证所述方法的基本特性，对此不感兴趣的读者只需了解有关结论，而略去那些烦琐的证明. 对于理论研究与进行方法设计的科学技术工作者，这些理论分析是相当重要而又富于启发性的. 全书在叙述上力求深入浅出，便于读者理解.

作为一个优化技术应用的工作者，仅仅了解一些算法是不够的. 经验的积累需要不断的实践，对不同算法在计算机上的实现，比较它们的性能，了解算法实现过程中可能出现的问题将是十分必要的，也是十分有益的. 建议有条件的读者对不同的算法进行数值实验或应用于实际问题，定将受益匪浅.

阅读本书，读者只需要有多元微积分与数值代数方面的基本知识.

本书是在对西安交通大学的硕士研究生多次授课的基础上成书的. 研究生在讨论中的意见进一步完善了本书.

作者特别感谢西安交通大学理学院万萍女士在本书成书全过程中所给予的关注和帮助. 本书的责任编辑科学出版社的林鹏与杨波对本书的写作给予了热情的指导与帮助，在此一并表示谢意.

限于作者水平，敬请广大读者对书中的不妥与错误之处批评指正，不吝赐教.

作　者

2001.7.10

西安

目　　录

第一章　最优化基础

本章首先介绍非线性规划的数学模型及其基本概念;然后着重介绍多元函数分析与非线性规划最优解所满足的最优性条件;最后对非线性规划的求解方法作一般性的描述,并简要介绍求解方法产生的迭代序列应具备的性质及其收敛速度.这一章的内容是以后各章的理论基础.

§1.1　最优化问题的数学模型与分类

本节介绍最优化问题数学模型的三种表示形式及最优化问题的几种分类.

最优化问题的数学模型一般形式为

$$\begin{cases}\min f(x), \\ \text{s.t. } c_i(x)=0, & i=1,2,\cdots,m', \\ \qquad c_i(x)\geqslant 0, & i=m'+1,\cdots,m.\end{cases} \tag{1.1.1}$$

其中 $x=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in R^n$, $f:R^n\to R^1$, $c_i:R^n\to R^1(i=1,2,\cdots,m)$. x 称为优化变量或决策变量,$f(x)$称为目标函数,$c_i(x)$称为约束函数. $c_i(x)=0$ $(i=1,2,\cdots,m')$称为等式约束,$c_i(x)\geqslant 0$ $(i=m+1,\cdots,m)$称为不等式约束. $c_i(x)=0$ 和 $c_i(x)\geqslant 0$ 称为约束条件. s. t. 是 subject to 的缩写,表示优化变量 x 受约束条件的限制.

有时为书写方便,记指标集 $E=\{1,2,\cdots,m'\}$, $I=\{m'+1,\cdots,m\}$,则模型(1.1.1)可以写为

$$\begin{cases}\min f(x), \\ \text{s.t. } c_i(x)=0, & i\in E, \\ \qquad c_i(x)\geqslant 0, & i\in I.\end{cases} \tag{1.1.1$'$}$$

若记集合 $D=\{x\mid c_i(x)=0, i=1,2,\cdots,m,\ c_i(x)\geqslant 0, i=m'+1,\cdots,m, x\in R^n\}$,则模型(1.1.1)又可写为

$$\min_{x\in D} f(x). \tag{1.1.1$''$}$$

称 D 为约束集合、约束区域或可行域.若 $x\in D$,则称 x 为可行点或可行解.

若模型(1.1.1)中不含约束条件,即 $D=R^n$,称模型(1.1.1)为无约束最优化问题,简记为

$$\min f(x), \qquad x\in R^n. \tag{1.1.2}$$

无约束问题是在空间 R^n 上寻求使目标函数 $f(x)$ 达到极小或最小的点 x^*.

若模型(1.1.1)中含有约束条件,则称模型(1.1.1)为约束最优化问题,约束问题是在约束集 D 上[换句话说也就是在约束条件 $c_i(x)=0$ $(i=1,2,\cdots,m')$, $c_i(x)\geqslant 0$ $(i=m'+1,\cdots,m)$ 的限制下]寻求使目标函数 $f(x)$ 达到极小或最小的点 x^*.

对约束问题而言,若模型(1.1.1)中只含有等式约束,无不等式约束,则称模型(1.1.1)为等式约束问题;若模型(1.1.1)中无等式约束,只含有不等式约束,则称模型(1.1.1)为不等式约束问题;若模型(1.1.1)既含等式约束,又含不等式约束,则称模型(1.1.1)为混合约束问题.

若模型(1.1.1)中的目标函数 $f(x)$ 和约束函数 $c_i(x)(i=1,2,\cdots,m)$ 都是线性函数,则称模型(1.1.1)为线性规划;若 $f(x)$ 和 $c_i(x)(i=1,2,\cdots,m)$ 中至少有一个是非线性函数,则称模型(1.1.1)为非线性规划.特别,当目标函数 $f(x)$ 是二次函数, $c_i(x)(i=1,2,\cdots,m)$ 都是线性函数时,则称模型(1.1.1)为二次规划.二次规划是一种特殊的非线性规划.

本书只讨论非线性规划的理论与求解方法,不涉及线性规划问题.对线性规划问题感兴趣的读者可参阅有关论著.

最后说明一点,我们只讨论 $f(x)$ 的极小化问题.对于 $f(x)$ 的极大化问题,由于 $f(x)$ 的极大点对应于 $-f(x)$ 的极小点,可令 $\bar{f}(x)=-f(x)$,便可将 $f(x)$ 的极大化问题转化为函数 $\bar{f}(x)$ 的极小化问题.

§1.2 多元函数分析

求解非线性规划(1.1.1)的绝大多数方法是在假定目标函数和约束函数连续可微的条件下提出的.本节介绍 n 元函数的一阶、二阶导数和泰勒(Taylor)展开式及相关理论,为以后章节的学习打下基础.

1.2.1 n 元函数的梯度、海森矩阵及向量函数的雅可比矩阵

定义 1.2.1 设 n 元函数 $f(x)$ 对自变量 $x=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 各分量 x_i 的偏导数

$$\frac{\partial f(x)}{\partial x_i}, \qquad i=1,2,\cdots,n$$

都存在,则称函数 $f(x)$ 在 x 处一阶可导,并称向量

$$\nabla f(x)=\left(\frac{\partial f(x)}{\partial x_1},\frac{\partial f(x)}{\partial x_2},\cdots,\frac{\partial f(x)}{\partial x_n}\right)^{\mathrm{T}} \tag{1.2.1}$$

为函数 $f(x)$ 在 x 处的一阶导数或梯度.为书写简便,也将 $\nabla f(x)$ 记为 $g(x)$,

将$\nabla f(x)$在$x^{(k)}$处的值$\nabla f(x^{(k)})$记为$g^{(k)}$.

定义 1.2.2 设n元函数$f(x)$对自变量$x=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$各分量的二阶偏导数

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \qquad i=1,2,\cdots,n, \quad j=1,2,\cdots,n$$

都存在,则称函数$f(x)$在点x处二阶可导,并称矩阵

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \tag{1.2.2}$$

为$f(x)$在x处的二阶导数矩阵或海森(Hesse)阵,有时将$\nabla^2 f(x)$记为$G(x)$或$H(x)$,将$\nabla^2 f(x^{(k)})$记为G_k.

若$f(x)$对x各变元的所有二阶偏导数都连续,则

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}, \qquad i=1,2,\cdots,n,\ j=1,2,\cdots,n.$$

这时$\nabla^2 f(x)$是一个对称矩阵.

例 1.2.1 设$A \in R^{n\times n}$是对称矩阵,$b\in R^n$,$c\in R^1$,求二次函数

$$f(x) = \frac{1}{2}x^{\mathrm{T}}Ax + b^{\mathrm{T}}x + c$$

在任意点x处的梯度和海森矩阵.

解 设$A=(a_{ij})_{n\times n}$, $x=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$,$b=(b_1,b_2,\cdots,b_n)^{\mathrm{T}}$,于是

$$\begin{aligned} f(x) &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j + \sum_{i=1}^{n} b_i x_i + c \\ &= \frac{1}{2}\sum_{i=1}^{n} x_i \sum_{j=1}^{n} a_{ij}x_j + \sum_{i=1}^{n} b_i x_i + c. \end{aligned}$$

对$f(x)$关于x_k求偏导得

$$\begin{aligned} \frac{\partial f(x)}{\partial x_k} &= \frac{1}{2}\left[\sum_{j=1}^{n} a_{kj}x_j + \sum_{j=1}^{n} a_{jk}x_j\right] + b_k \\ &= \sum_{j=1}^{n} a_{kj}x_j + b_k, \qquad (\text{由 } A \text{ 是对称矩阵知 } a_{jk} = a_{kj}) \end{aligned}$$

$$k = 1,2,\cdots,n.$$

故

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \sum_{j=1}^{n} a_{2j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{nj}x_j \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

$$= Ax + b.$$

再对$\frac{\partial f(x)}{\partial x_i}$关于$x_j$求偏导,得

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = a_{ij}, \qquad i = 1,2,\cdots,n, \quad j = 1,2,\cdots,n,$$

所以

$$\nabla^2 f(x) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = A.$$

定义 1.2.3 设向量函数 $F(x)=(f_1(x),f_2(x),\cdots,f_m(x))^{\mathrm{T}}$ 的各分量函数 $f(x)(i=1,2,\cdots,m)$对自变量 $x=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 各分量的偏导数

$$\frac{\partial f_i(x)}{\partial x_j}, \qquad i = 1,2,\cdots,m, \quad j = 1,2,\cdots,m$$

都存在,则称 $F(x)$在点 x 处一阶可导,并称矩阵

$$F'(x) = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_2} & \cdots & \frac{\partial f_2(x)}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \frac{\partial f_m(x)}{\partial x_2} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix} \tag{1.2.3}$$

为向量函数 $F(x)$在 x 处的雅可比(Jacobi)矩阵.

例 1.2.2 设

$$F(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} = \begin{bmatrix} 3x_1 + \mathrm{e}^{x_2}x_3 \\ x_1^3 + x_2^2\sin x_3 \end{bmatrix},$$

求 $F(x)$在点 $\bar{x}=(1,0,\pi)^{\mathrm{T}}$ 处的雅可比矩阵.

解 由式(1.2.3)得

$$F'(x) = \begin{bmatrix} 3 & e^{x_2}x_3 & e^{x_2} \\ 3x_1^2 & 2x_2\sin x_3 & x_2^2\cos x_3 \end{bmatrix}.$$

代入 $\bar{x}=(1,0,\pi)^{\mathrm{T}}$,得

$$F'(\bar{x}) = \begin{bmatrix} 3 & \pi & 1 \\ 3 & 0 & 0 \end{bmatrix}.$$

1.2.2 *n* 元函数的泰勒展开式

n 元函数的泰勒展开式在非线性规划问题的研究中具有非常重要的作用,下面,我们借助于一元函数的一阶、二阶导数,导出 n 元函数的泰勒展开式.

首先,根据多元复合函数的求导法则,导出一元函数

$$\varphi(\alpha) = f(x+\alpha d), \quad \alpha \in R^1, \quad x,d \in R^n$$

的一阶、二阶导数.

记

$$\begin{aligned} u &= x + \alpha d = (x_1+\alpha d_1, x_2+\alpha d_2, \cdots, x_n+\alpha d_n)^{\mathrm{T}} \\ &= (u_1, u_2, \cdots, u_n)^{\mathrm{T}}, \end{aligned}$$

则

$$\begin{aligned} \varphi'(\alpha) &= \frac{\partial f(u)}{\partial u_1}\frac{\mathrm{d}u_1}{\mathrm{d}\alpha} + \frac{\partial f(u)}{\partial u_2}\frac{\mathrm{d}u_2}{\mathrm{d}\alpha} + \cdots + \frac{\partial f(u)}{\partial u_n}\frac{\mathrm{d}u_n}{\mathrm{d}\alpha} \\ &= \frac{\partial f(u)}{\partial u_1}d_1 + \frac{\partial f(u)}{\partial u_2}d_2 + \cdots + \frac{\partial f(u)}{\partial u_n}d_n \\ &= \nabla f(u)^{\mathrm{T}} d \\ &= \nabla f(x+\alpha d)^{\mathrm{T}} d, \end{aligned} \tag{1.2.4}$$

$$\begin{aligned} \varphi''(\alpha) = & \left(\frac{\partial^2 f(u)}{\partial u_1^2}d_1 + \frac{\partial^2 f(u)}{\partial u_1 \partial u_2}d_2 + \cdots + \frac{\partial^2 f(u)}{\partial u_1 \partial u_n}d_n\right)d_1 \\ & + \left(\frac{\partial^2 f(u)}{\partial u_2 \partial u_1}d_1 + \frac{\partial^2 f(u)}{\partial u_2^2}d_2 + \cdots + \frac{\partial^2 f(u)}{\partial u_2 \partial u_n}d_n\right)d_2 \\ & + \cdots \\ & + \left(\frac{\partial^2 f(u)}{\partial u_n \partial u_1}d_1 + \frac{\partial^2 f(u)}{\partial u_n \partial u_2}d_2 + \cdots + \frac{\partial^2 f(u)}{\partial u_n^2}d_n\right)d_n \\ = & (d_1, d_2, \cdots, d_n) \begin{bmatrix} \frac{\partial^2 f(u)}{\partial u_1^2} & \frac{\partial^2 f(u)}{\partial u_1 \partial u_2} & \cdots & \frac{\partial^2 f(u)}{\partial u_1 \partial u_n} \\ \frac{\partial^2 f(u)}{\partial u_2 \partial u_1} & \frac{\partial^2 f(u)}{\partial u_2^2} & \cdots & \frac{\partial^2 f(u)}{\partial u_2 \partial u_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f(u)}{\partial u_n \partial u_1} & \frac{\partial^2 f(u)}{\partial u_n^2 \partial u_2^2} & \cdots & \frac{\partial^2 f(u)}{\partial u_n^2} \end{bmatrix} \begin{bmatrix} d_1 \\ d_1 \\ \vdots \\ d_1 \end{bmatrix} \end{aligned}$$

$$= d^{\mathrm{T}}\nabla^2 f(x+\alpha d)d. \tag{1.2.5}$$

定理 1.2.4 (1)设 $f:R^n\to R^1$,若 $f(x)$在点 x^* 的某邻域 $N(x^*)$内一阶连续可微,则存在 $\theta\in(0,1)$使

$$f(x) = f(x^*) + \nabla f(\xi)^{\mathrm{T}}(x-x^*), x\in N(x^*), \tag{1.2.6}$$

其中 $\xi = x^* + \theta(x-x^*)$.

(2)设 $f:R^n\to R^1$,若 $f(x)$在点 x^* 的某邻域 $N(x^*)$内一阶连续可微,则

$$f(x) = f(x^*) + \nabla f(x^*)^{\mathrm{T}}(x-x^*) + o(\|x-x^*\|),\ x\in N(x^*). \tag{1.2.7}$$

(3)设 $f:R^n\to R^1$,若 $f(x)$在 x^* 的某邻域 $N(x^*)$内二阶连续可微,则存在 $\theta\in(0,1)$使

$$\begin{aligned} f(x) &= f(x^*) + \nabla f(x^*)^{\mathrm{T}}(x-x^*) \\ &\quad + \frac{1}{2}(x-x^*)^{\mathrm{T}}\nabla^2 f(\xi)(x-x^*),\ x\in N(x^*), \end{aligned} \tag{1.2.8}$$

其中 $\xi = x^* + \theta(x-x^*)$.

(4)设 $f:R^n\to R^1$,若 $f(x)$在点 x^* 的某邻域 $N(x^*)$内二阶连续可微,则

$$\begin{aligned} f(x) &= f(x^*) + \nabla f(x^*)^{\mathrm{T}}(x-x^*) + \frac{1}{2}(x-x^*)^{\mathrm{T}} \\ &\quad \cdot \nabla^2 f(x^*)(x-x^*) + o(\|x-x^*\|^2),\ x\in N(x^*). \end{aligned} \tag{1.2.9}$$

证明 结论(1),(2)留给读者自证,现证结论(3)和(4).

(3)设 $\varphi(\alpha) = f(x^*+\alpha d)$,其中 $d = x-x^*$,由一元函数的泰勒公式有

$$\varphi(\alpha) = \varphi(0) + \varphi'(0)\alpha + \frac{1}{2}\varphi''(\theta)\alpha^2,$$

其中 $0<\theta<1$.取 $\alpha=1$,得

$$\varphi(1) = \varphi(0) + \varphi'(0) + \frac{1}{2}\varphi''(\theta). \tag{1.2.10}$$

显然 $\varphi(1) = f(x)$,$\varphi(0) = f(x^*)$.由(1.2.4)和(1.2.5)式知

$$\begin{aligned} \varphi'(0) &= \nabla f(x^*)^{\mathrm{T}}(x-x^*), \\ \varphi''(\theta) &= (x-x^*)^{\mathrm{T}}\nabla^2 f(x^*+\theta(x-x^*))(x-x^*). \end{aligned}$$

将以上各式代入(1.2.10)式,便得(1.2.8)式.

(4)设 $\varphi(\alpha) = f(x^*+\alpha d)$,其中 $\alpha = \|x-x^*\|$,$d = \dfrac{x-x^*}{\|x-x^*\|}$.

由一元函数的泰勒公式有

$$\varphi(\alpha) = \varphi(0) + \varphi'(0)\alpha + \frac{1}{2}\varphi''(0)\alpha^2 + o(\alpha^2). \tag{1.2.11}$$

同样

$$\varphi(\alpha)=f(x),\varphi(0)=f(x^*),$$
$$\varphi'(0)\alpha=\nabla f(x^*)^{\mathrm{T}}(x-x^*),$$
$$\varphi''(0)\alpha^2=(x-x^*)^{\mathrm{T}}\nabla^2 f(x^*)(x-x^*).$$

将以上各式代进(1.2.11)式,即得(1.2.9)式. #

在(1.2.7)和(1.2.9)式中,略去高阶无穷小量后,则有近似关系式

$$f(x)\approx f(x^*)+\nabla f(x^*)^{\mathrm{T}}(x-x^*), \tag{1.2.12}$$

$$f(x)\approx f(x^*)+\nabla f(x^*)^{\mathrm{T}}(x-x^*)+\frac{1}{2}(x-x^*)^{\mathrm{T}}\nabla^2 f(x^*)(x-x^*). \tag{1.2.13}$$

通常把(1.1.12)式和(1.1.13)式的右边分别称为函数 $f(x)$在 x^*处的线性近似(函数)和二次近似(函数).

定理 1.2.5 设 $f:R^n\to R^1$,若 $f(x)$在 x^*的某邻域 $N(x^*)$内一阶连续可微,则

$$f(x)=f(x^*)+\int_0^1\nabla f(x^*+\alpha(x-x^*))^{\mathrm{T}}(x-x^*)\mathrm{d}\alpha,\ x\in N(x^*). \tag{1.2.14}$$

证明 设 $\varphi(\alpha)=f(x^*+\alpha d)$,其中 $d=x-x^*$,由公式(1.2.4)有

$$\begin{aligned}\varphi'(\alpha)&=\nabla f(x^*+\alpha d)^{\mathrm{T}}d\\&=\nabla f(x^*+\alpha(x-x^*))^{\mathrm{T}}(x-x^*).\end{aligned}$$

对上式两端在区间[0,1]上关于 α 积分,得

$$\varphi(1)-\varphi(0)=\int_0^1\varphi'(\alpha)\mathrm{d}\alpha=\int_0^1\nabla f(x^*+\alpha(x-x^*))^{\mathrm{T}}(x-x^*)\mathrm{d}\alpha.$$

注意到 $\varphi(1)=f(x),\varphi(0)=f(x^*)$,上式即(1.2.14)式. #

称(1.2.14)式为中值定理的积分形式.

若 $f(x)$在 x^*的某邻域内二阶连续可微时,则有积分形式的泰勒公式:

$$f(x)=f(x^*)+\nabla f(x^*)^{\mathrm{T}}(x-x^*)+\int_0^1\left[\int_0^t(x-x^*)^{\mathrm{T}}\nabla^2 f(x^*+\alpha(x-x^*))(x-x^*)\mathrm{d}\alpha\right]\mathrm{d}t. \tag{1.2.15}$$

若 $f(x)$在 R^n 上二阶连续可微时,则有另一种积分形式的泰勒公式.

定理 1.2.6 设 $f:R^n\to R^1$,若 $f(x)$在 R^n 上二阶连续可微,则对任意的向量 $x,d\in R^n$ 和任意实数 α 有

$$f(x+\alpha d)=f(x)+\alpha\nabla f(x)^{\mathrm{T}}d+\alpha^2\int_0^1(1-t)[d^{\mathrm{T}}\nabla^2 f(x+t\alpha d)d]\mathrm{d}t. \tag{1.2.16}$$

证明 令 $\varphi(t)=f(x+t\alpha d)$,则 $\varphi'(t)=\alpha\nabla f(x+t\alpha d)^{\mathrm{T}}d,\varphi''(t)=$

$\alpha^2 d^{\mathrm{T}}\nabla^2 f(x+t\alpha d)d$，由 $\varphi(1)-\varphi(0)=\int_0^1\varphi'(t)\mathrm{d}t$ 知

$$
\begin{aligned}
f(x+\alpha d)-f(x) &= \int_0^1[\alpha\nabla f(x+t\alpha d)^{\mathrm{T}}d]\mathrm{d}t \\
&= -\int_0^1[\alpha\nabla f(x+t\alpha d)^{\mathrm{T}}d]\mathrm{d}(1-t) \\
&= -[(1-t)\alpha\nabla f(x+t\alpha d)^{\mathrm{T}}d]\big|_0^1 \\
&\quad +\int_0^1(1-t)\mathrm{d}[\alpha\nabla f(x+t\alpha d)^{\mathrm{T}}d] \\
&= \alpha\nabla f(x)^{\mathrm{T}}d+\alpha^2\int_0^1[(1-t)d^{\mathrm{T}}\nabla^2 f(x+t\alpha d)d]\mathrm{d}t.
\end{aligned}
$$

#

值得注意的是向量函数没有相应于(1.2.6)式的中值定理，但可将中值定理的积分形式(1.2.14)式推广到向量函数.事实上，若 $F(x)=(f_1(x), f_2(x),\cdots,f_m(x))^{\mathrm{T}}$ 的每个分量函数 $f_i(x)(i=1,2,\cdots,m)$ 在 x^* 的某邻域内一阶连续可微，对每个分量函数 $f_i(x)$，利用(1.2.14)式可得

$$
F(x)=F(x^*)+\int_0^1 F'(x^*+\alpha(x-x^*))(x-x^*)\mathrm{d}\alpha. \tag{1.2.17}
$$

如上所述，对于梯度 $\nabla f(x)$ 是没有中值定理的，但对 $\nabla f(x)$ 的每个分量应用一阶泰勒公式(1.2.7)，并忽略其高阶项，便可得到如下的近似式

$$
\nabla f(x)\approx\nabla f(x^*)+\nabla^2 f(x^*)(x-x^*). \tag{1.2.18}
$$

1.2.3 凸集与凸函数

凸集与凸函数在最优化方法的理论分析中起着重要的作用，在本节我们给出凸集与凸函数的定义和基本性质.

定义1.2.7 设集合 $D\subset R^n$，如果对于任意的 $x,y\in D$ 与任意的 $\alpha\in[0,1]$，有

$$
\alpha x+(1-\alpha)y\in D,
$$

则称集合 D 是凸集.

有关凸集的性质由下面的定理给出.

定理1.2.8 设 $D_1,D_2\subset R^n$ 是凸集，$\alpha\in R^1$，则

(1) $D_1\cap D_2=\{x\mid x\in D_1,x\in D_2\}$ 是凸集.

(2) $\alpha D_1=\{\alpha x\mid x\in D_1\}$ 是凸集.

(3) $D_1+D_2=\{x+y\mid x\in D_1,y\in D_2\}$ 是凸集.

(4) $D_1-D_2=\{x-y\mid x\in D_1,y\in D_2\}$ 是凸集.

这个定理的证明可由凸集的定义直接得出,留给读者作为练习.

定理 1.2.9 设 D 是凸集,则对任意的 $x^{(i)} \in D$ 和数α_i 都有 $\sum_{i=1}^{m} \alpha_i x^{(i)} \in D$(其中 $\alpha_i \geqslant 0 \quad i = 1,2,\cdots,m$,且$\sum_{i=1}^{m} \alpha_i = 1$).

证明 用数学归纳法

当 $m=2$ 时,由凸集的定义,结论显然成立.

假定 $m=k$ 时结论成立,现证当 $m=k+1$ 时结论也成立.

$$\begin{aligned}\sum_{i=1}^{k+1} \alpha_i x^{(i)} &= \sum_{i=1}^{k} \alpha_i x^{(i)} + \alpha_{k+1} x^{(k+1)} \qquad \left(\alpha_i \geqslant 0, \quad \sum_{i=1}^{k+1} \alpha_i = 1\right) \\ &= (1-\alpha_{k+1}) \sum_{i=1}^{k} \frac{\alpha_i}{1-\alpha_{k+1}} x^{(i)} + \alpha_{k+1} x^{(k+1)}.\end{aligned}$$

由 $1 = \sum_{i=1}^{k+1} \alpha_i = \sum_{i=1}^{k} \alpha_i + \alpha_{k+1}$,知$\sum_{i=1}^{k} \alpha_i = 1 - \alpha_{k+1}$. 于是$\sum_{i=1}^{k} \frac{\alpha_i}{1-\alpha_{k+1}} = 1$, $\frac{\alpha_i}{1-\alpha_{k+1}} \geqslant 0$. 记 $x' = \sum_{i=1}^{k} \frac{\alpha_i}{1-\alpha_{k+1}} x^{(i)}$,由 $m = k$ 时结论成立的假定,知 $x' \in D$. 再由凸集的定义知

$$\sum_{i=1}^{k+1} \alpha_i x^{(i)} = (1-\alpha_{k+1}) x' + \alpha_{k+1} x^{(k+1)} \in D. \qquad \#$$

定义 1.2.10 设 $f: D \subset R^n \to R^1$, D 是非空凸集. 如果对任意的 $x, y \in D$ 和任意的 $\alpha \in (0,1)$ 有

$$f(\alpha x + (1-\alpha) y) \leqslant \alpha f(x) + (1-\alpha) f(y). \tag{1.2.19}$$

则称 $f(x)$是凸集 D 上的凸函数. 若 $D = R^n$,则称 $f(x)$是凸函数. 在上式中若在 $x \neq y$ 时严格不等式成立,则称 $f(x)$是凸集 D 上的严格函数. 若 $D = R^n$,则称 $f(x)$是严格凸函数.

若 $-f(x)$是凸集 D 上的(严格)凸函数,则称 $f(x)$是凸集 D 上的(严格)凹函数.

定义 1.2.11 设 $f: D \subset R^n \to R^1$, D 是非空凸集,如果存在一个常数 $\beta > 0$,对任意的 $x, y \in D$ 和任意的$\alpha \in (0,1)$有

$$\alpha f(x) + (1-\alpha) f(y) - f(\alpha x + (1-\alpha) y) \geqslant (1-\alpha)\alpha\beta \| x - y \|^2, \tag{1.2.20}$$

则称 $f(x)$在凸集 D 上是一致凸函数.

显然,凸集 D 上的一致凸函数必是凸集D 上的严格凸函数,自然也是 D 上的凸函数.

定理 1.2.12 设 $f_i(x)(i=1,2,\cdots,m)$是非空凸集 D 上的凸函数,则

(1) $f(x)=\sum_{i=1}^{m}\alpha_i f_i(x)$(数 $\alpha_i\geqslant 0, i=1,2,\cdots,m$)也是凸集 D 上的凸函数.

(2) $f(x)=\max\limits_{1\leqslant i\leqslant m}\{f_i(x)\}$ 也是凸集 D 上的凸函数.

证明 结论(1)可直接由凸函数的定义得到,现证结论(2).

设 $\forall x,y\in D, \forall \alpha\in(0,1)$,

$$\begin{aligned} f(\alpha x+(1-\alpha)y) &= \max_{1\leqslant i\leqslant m}\{f_i(\alpha x+(1-\alpha)y)\} \\ &\leqslant \max_{1\leqslant i\leqslant m}\{\alpha f_i(x)+(1-\alpha)f_i(y)\} \\ &\leqslant \alpha\max_{1\leqslant i\leqslant m}\{f_i(x)\}+(1-\alpha)\max_{1\leqslant i\leqslant m}\{f_i(y)\} \\ &= \alpha f(x)+(1-\alpha)f(y). \end{aligned}$$ #

对于一元函数,当 $m=2$ 时,结论(2)的几何直观是明显的,如图 1.2.1 所示.

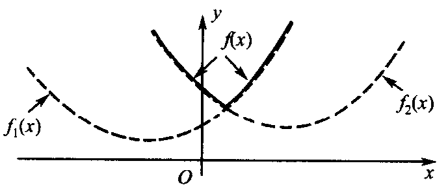

图 1.2.1

定理 1.2.13 设 $f(x)$ 是凸集 D 上的凸函数,对任意的 $x^{(i)}\in D$ $(i=1,2,\cdots,m)$ 和 $\alpha_i\geqslant 0$ 且 $\sum_{i=1}^{m}\alpha_i=1$,有

$$f\left(\sum_{i=1}^{m}\alpha_i x^{(i)}\right)\leqslant\sum_{i=1}^{m}\alpha_i f(x^{(i)}).$$

证明 用归纳法.

当 $m=2$ 时,由凸函数的定义知结论成立.

假定当 $m=k$ 时结论成立,即

$$f\left(\sum_{i=1}^{k}\beta_i x^{(i)}\right)\leqslant\sum_{i=1}^{k}\beta_i f(x^{(i)})$$

$$\left(\beta_i\geqslant 0,\quad \sum_{i=1}^{k}\beta_i=1\right).$$

现证当 $m=k+1$ 时结论也成立.

$$\begin{aligned} f\left(\sum_{i=1}^{k+1}\alpha_i x^{(i)}\right) &= f\left(\sum_{i=1}^{k}\alpha_i x^{(i)}+\alpha_{k+1}x^{(k+1)}\right) \\ &= f\left((1-\alpha_{k+1})\sum_{i=1}^{k}\frac{\alpha_i}{1-\alpha_{k+1}}x^{(i)}+\alpha_{k+1}x^{k+1}\right), \end{aligned}$$

记 $\beta_i=\sum_{i=1}^{k}\frac{\alpha_i}{1-\alpha_{k+1}}$,显然 $\beta_i\geqslant 0$ 且 $\sum_{i=1}^{k}\beta_i=1$,记 $x'=\sum_{i=1}^{k}\frac{\alpha_i}{1-\alpha_{k+1}}x^{(i)}=$

$\sum_{i=1}^{k}\beta_i x^{(i)}$，由定理 1.2.9 知 $x'\in D$，所以

$$\text{上式}\leqslant(1-\alpha_{k+1})f\Big(\sum_{i=1}^{k}\beta_i x^{(i)}\Big)+\alpha_{k+1}f(x^{(k+1)})$$

$$\leqslant(1-\alpha_{k+1})\sum_{i=1}^{k}\beta_i f(x^{(i)})+\alpha_{k+1}f(x^{(k+1)})\text{（利用 }m=k\text{ 成立的假设）}$$

$$=\sum_{i=1}^{k}\alpha_i f(x^{(i)})+\alpha_{k+1}f(x^{(k+1)})$$

$$=\sum_{i=1}^{k+1}\alpha_i f(x^{(i)}).$$

#

下面给出几个凸函数的判别定理.

定理 1.2.14 $f(x)$为凸函数的充要条件是对任意的 $x,y\in R^n$，一元函数 $\varphi(\alpha)=f(x+\alpha y)$是关于 α 的凸函数.

证明 必要性 设 $\lambda_1\geqslant0,\lambda_2\geqslant0$，且 $\lambda_1+\lambda_2=1$，由 $\varphi(\alpha)$的定义和 $f(x)$的凸性有

$$\begin{aligned}\varphi(\lambda_1\alpha_1+\lambda_2\alpha_2)&=f(x+(\lambda_1\alpha_1+\lambda_2\alpha_2)y)\\&=f(\lambda_1x+\lambda_2x+(\lambda_1\alpha_1+\lambda_2\alpha_2)y)\\&=f(\lambda_1(x+\alpha_1y)+\lambda_2(x+\alpha_2y))\\&\leqslant\lambda_1f(x+\alpha_1y)+\lambda_2f(x+\alpha_2y)\\&=\lambda_1\varphi(\alpha_1)+\lambda_2\varphi(\alpha_2).\end{aligned}$$

由凸函数的定义知 $\varphi(\alpha)$是凸函数.

充分性 任取 $x,y\in R^n$，设 $z^1=x+\alpha_1y,z^2=x+\alpha_2y$，

$$\begin{aligned}f(\lambda_1z^1+\lambda_2z^2)&=f(\lambda_1(x+\alpha_1y)+\lambda_2(x+\alpha_2y))\\&=f(x+(\lambda_1\alpha_1+\lambda_2\alpha_2)y)\\&=\varphi(\lambda_1\alpha_1+\lambda_2\alpha_2)\\&\leqslant\lambda_1\varphi(\alpha_1)+\lambda_2\varphi(\alpha_2)\\&=\lambda_1f(x+\alpha_1y)+\lambda_2f(x+\alpha_2y)\\&=\lambda_1f(z^1)+\lambda_2f(z^2).\end{aligned}$$

故知 $f(x)$是凸函数.

#

定理 1.2.15 设 $D\subset R^n$ 是非空开凸集，$f:D\subset R^n\to R^1$，且 $f(x)$在 D 上一阶连续可微，则

(1) $f(x)$是 D 上的凸函数的充要条件是

$$f(y)-f(x)\geqslant\nabla f(x)^{\mathrm{T}}(y-x),\qquad\forall x,y\in D.\quad(1.2.21)$$

(2) $f(x)$是 D 上严格凸函数的充要条件是