通识教育平台数学课程系列教材

SHUZHIZUIYOUHUA

数值最优化

李董辉 童小娇 万 中 编

$$\min_{0\leqslant t\leqslant t_{max}} f(x^{(k)}+td^{(k)})=f(x^{(k)}+t_kd^{(k)})$$

$$\begin{aligned} \min\quad & f(x)=x_1^2+x_2^2-x_1x_2-2x_1-3x_2 \\ \text{s.t.}\quad & g_1(x)=-x_1-x_2+2\geqslant 0, \\ & g_2(x)=-x_1-5x_2+5\geqslant 0, \\ & g_3(x)=x_1\geqslant 0, \\ & g_4(x)=x_2\geqslant 0. \end{aligned}$$

$$\nabla f(x)=(2x_1-x_2-2,-x_1+2x_2-3)^{\mathrm{T}}.$$

$$\begin{aligned} \min\quad & \nabla f(x^0)^{\mathrm{T}}d=-2d_1-3d_2 \\ & d_1\geqslant 0, d_2\geqslant 0, \\ & \|d\|_\infty\leqslant 1. \end{aligned}$$

科学出版社
www.sciencep.com

内 容 简 介

本书较为系统地介绍了最优化领域中比较成熟的基本理论与方法.基本理论包括最优化问题解的必要条件和充分条件,以及各种算法的收敛性理论.介绍的算法有:求解无约束问题的最速下降法、Newton 法、拟 Newton 法、共轭梯度法、信赖域算法和直接法;求解约束问题的罚函数法、乘子法、可行方向法、序列二次规划算法和信赖域算法等.此外,本书还介绍了线性规划的基本理论与单纯形算法以及求解二次规划的有效算法,求解全局最优化问题的几种常用算法.作为基本工具,本书在附录中简要介绍了求解线性方程组的常用直接法和迭代法以及 Matlab 初步知识.

本书可作为数学类本科各专业和工程类研究生最优化课程的教材.书中的许多章节内容相对独立,使用者可根据需要灵活取舍.本书也可作为工程技术人员的参考书.

图书在版编目(CIP)数据

数值最优化/李董辉等编.—北京:科学出版社,2005.5

(通识教育平台数学课程系列教材)

ISBN 7-03-015312-X

Ⅰ.数…　Ⅱ.李…　Ⅲ.最优化算法-教材　Ⅳ.O242.23

中国版本图书馆 CIP 数据核字(2005)第 027361 号

责任编辑:杨瑰玉

责任印制:高　嵘/封面设计:李梦佳

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

武汉大学出版社印刷总厂印刷

科学出版社发行　各地新华书店经销

*

2005 年 5 月第 一 版　开本:A5(890×1240)

2005 年 5 月第一次印刷　印张:9

印数:1～3 000　字数:278 000

定价:16.00 元

(如有印装质量问题,我社负责调换)

前 言

最优化是一门应用性很强的学科，其应用领域涉及各类工程、军事、生产、管理、经济等.最优化方法已成为许多工程技术人员、管理工作者和研究人员的必备工具.

本书较为系统地介绍最优化的基本理论与算法，侧重于介绍最优化领域中比较成熟的算法及其收敛性理论.同时，结合作者的研究成果，介绍求解无约束问题的Newton 法和拟Newton 法的部分改进工作.

本书既注重算法的理论基础，同时也突出对算法的直观解释.每一个算法之后都配有例题.算法的收敛性理论介绍安排在各算法例题之后，便于不同的使用对象对内容进行取舍.书中注 * 号的部分为选修内容.

本书各章节间的联系如下：第一章是准备工作，主要介绍有关最优化的基本概念以及以后各章需用到的数学基础知识，是全书的基础.其他各章分别介绍求解无约束问题和约束问题的基本理论和各种常用算法.第二章介绍无约束问题解的最优性条件和求解无约束问题的下降算法的一般步骤及其收敛性质，它是第三、四、五章的基础.第八章介绍约束问题解的最优性条件，它是第十、十一、十二和十三章的基础.第六、七、八章相对独立.最后，在第十四章介绍求解全局最优化问题的几种常用算法.学习该章时需用到第九章关于线性规划的知识.

本书的第一章至第五章以及第七章至第十二章由李董辉撰写，第六章和第十三章由童小娇撰写，第十四章由万中撰写.曾金平教授对本书提出了宝贵的修改意见和建议，并为本书撰写了附录，特在此表示感谢.同时，感谢编者的学生杨春旭、陈娟、程万友、张雷、易芳、曹二保、谢锐、谢水连、安晓敏、沈志伙、胡华丽等对本书的认真校对工作.

本书的编写得到了教育部博士点专项基金的资助，在出版过程中得到了刘楚中教授的支持和帮助，在此表示衷心感谢.

由于编者水平有限，书中难免有错误之处，欢迎广大读者批评指正.

编　者

2005 年 3 月于岳麓山下

目　录

第一章 引 言

§1.1 最优化问题概述

最优化问题的数学模型如下：

$$\min f(x), \quad x\in D\subseteq R^n, \tag{1.1}$$

其中，函数 f 是定义在 R^n 上的实值函数. 我们称函数 f 为问题(1.1)的目标函数，集合 D 为问题的可行域，可行域中的点为可行点. 如果将函数 f 视为生产中的某种成本，集合 D 视为生产过程中的各种可采用的方案所构成的集合，则上述最优化问题可通俗地理解为在众多可行的方案中寻求最佳方案. 最优化有广泛的应用，如：在生产管理中如何最大限度地降低成本；如何合理下料最节省材料；如何设计运输方案使得运输费用最少等. 在实际工作中，最优化还有另一种含义，即使获得的效益最大. 此时，相应的模型为：

$$\max f(x), \quad x\in D\subseteq R^n, \tag{1.2}$$

其中，f 为某种效益函数. 注意到 $\max f=-\min(-f)$，问题(1.2)可转化为形如(1.1)的等价问题.

为了对最优化问题有一个直观的了解，我们考察几个实际问题的数学模型.

例 1.1.1 *（食谱问题）设市场上可以买到 n 种不同的食品，每种食品含有 m 种营养成分. 设每单位的 j 种食品含 i 种营养成分的数量为 a_{ij}，$i=1,\cdots,m$，$j=1,\cdots,n$. 第 j 种食品的单位价格为 c_j，$j=1,\cdots,n$. 再设每人每天对第 i 种营养成分的需求量为 b_i，$i=1,\cdots,m$. 试确定在保证营养需求条件下的经济食谱.*

解 设购入第 j 种食品的数目为 x_j，$j=1,\cdots,n$. 则总开支为：

$$f(x)\triangleq\sum_{j=1}^{n}c_jx_j.$$

所获得的第 i 种营养成分总量为

$$g_i(x) \triangleq \sum_{j=1}^{n} a_{ij} x_j.$$

我们的目标是使得函数 f 达到最小. 所满足的条件为：$g_i(x) \geqslant b_i, i=1, \cdots, m$. 此外，购入食品的数目 x_j 不能为负数. 因此，我们得到如下数学模型：

$$\begin{aligned} \min & \sum_{j=1}^{n} c_j x_j \\ \text{s.t.} & \sum_{j=1}^{n} a_{ij} x_j \geqslant b_i, \quad i=1, \cdots, m, \\ & x_j \geqslant 0, \quad j=1, \cdots, n, \end{aligned}$$

其中，min 是minimizing（极小化）的简称，s. t. 是subject to（受约束于）的简称.

例 1.1.2 （数据拟合问题）设有观测数据$(x_k, y_k), k=1, \cdots, 5$，其值由表 1.1 给出：

表 1.1 观测数据

k	1	2	3	4	5
x_k	2	4	5	8	9
y_k	2.01	2.98	3.50	5.02	5.47

试用一个简单的函数关系拟合这些数据.

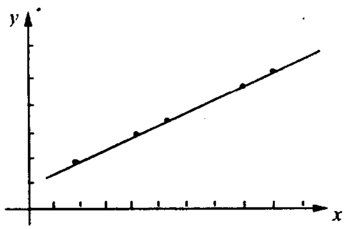

图 1.1 线性拟合

解 将观测数据点在直角坐标系上标出，得图 1.1.

由图 1.1 不难发现，(x_k, y_k)大致位于某一条直线的附近. 因此，x 与y 间的函数关系近似于线性函数关系. 考虑用直线

$$y = ax + b \tag{1.3}$$

对这些点进行拟合，即确定a, b 的值，使得点(x_k, y_k)通过或靠近上面的直线. 不难发现，对于上述数据点，不存在a, b，使得方程(1.3)对所有的$(x_k, y_k), k=1, \cdots, 5$ 都满足. 因此，我们求a, b，使得函数

$$f(a, b) = \sum_{i=1}^{5} |y_i - (ax_i + b)|$$

或

$$f(a,b)=\sum_{i=1}^{5}[y_i-(ax_i+b)]^2$$

达到最小值，即上面的数据拟合问题可通过如下的极小值问题来描述：

$$\min f(a,b),\quad (a,b)^{\mathrm{T}}\in R^2.$$

在问题(1.1)中，若$D=R^n$，则称问题(1.1)为无约束最优化问题(简称无约束问题)；否则称问题(1.1)为约束最优化问题(简称约束问题). 上面的例1.1.1是约束问题，而例1.1.2是无约束问题.

无约束问题通常记为：

$$\min f(x),\quad x\in R^n. \tag{1.4}$$

约束问题的一般形式为：

$$\begin{aligned} \min\quad & f(x) \\ \text{s.t.}\quad & g_i(x)\geqslant 0, i\in I=\{1,\cdots,m_1\}, \\ & h_j(x)=0, j\in E=\{m_1+1,\cdots,m\}, \end{aligned} \tag{1.5}$$

其中$g_i,h_j:R^n\to R, i\in I, j\in E$. 约束问题(1.5)中，函数$g_i, i\in I, h_j, j\in E$称为约束函数. 不等式组$g_i(x)\geqslant 0, i\in I$和等式组$h_j(x)=0, j\in E$分别称为不等式约束条件和等式约束条件，统称为约束条件. 显然，问题(1.5)的可行域D为

$$D=\{x\,|\,g_i(x)\geqslant 0, i\in I, h_j(x)=0, j\in E\}.$$

最优化的一个主要研究内容就是求问题(1.1)的解. 最优化问题的解分为局部最优解和全局(整体)最优解. 其定义如下：

定义 1.1.1 *设点$x^*\in D$. 若存在x^*的一个邻域$U_\delta(x^*)\triangleq\{x\in R^n\,|\,\|x-x^*\|<\delta\}$，使得如下不等式成立：*

$$f(x^*)\leqslant f(x),\quad \forall x\in D\cap U_\delta(x^*), \tag{1.6}$$

则称x^*是问题(1.1)的一个局部最优解，或简称为问题(1.1)的一个最优解. 若不等式(1.6)对所有$x\in D\cap U_\delta(x^*)\backslash\{x^*\}$成立严格不等式，则称$x^*$是问题(1.1)的一个严格局部最优解. 若不等式

$$f(x^*)\leqslant f(x),\quad \forall x\in D \tag{1.7}$$

成立，则称x^*是问题(1.1)的一个全局(整体)最优解. 若不等式(1.7)对所有$x\in D\backslash\{x^*\}$成立严格不等式，则称x^*是问题(1.1)的一个严格全局(整体)最优解.

约束最优化问题中有两类重要的问题. 当目标函数f和约束函数g_i，

$i\in I, h_j, j\in E$ 都是线性函数时，约束问题(1.5)称为线性规划. 当目标函数 f 是二次函数且约束函数 $g_i, i\in I, h_j, j\in E$ 都是线性函数时，约束问题(1.5)称为二次规划. 此外，若函数 f 是 R^n 中的凸函数且 D 是凸集时，问题(1.1)称为凸规划. 凸集和凸函数的概念我们将在下一节中介绍.

在结束本节之前，我们回顾一下多元函数的 Taylor 展开式以及向量和矩阵范数.

设 $f: R^n\to R$ 二次连续可微. 我们用 $\nabla f(x)$ 和 $\nabla^2 f(x)$ 分别表示 f 在 x 处的梯度向量和 Hessian 矩阵，即

$$\nabla f(x)=\left(\frac{\partial f(x)}{\partial x_1},\cdots,\frac{\partial f(x)}{\partial x_n}\right)^{\mathrm{T}},\quad \nabla^2 f(x)=\begin{bmatrix}\dfrac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_1\partial x_n}\\ \vdots & & \vdots\\ \dfrac{\partial^2 f(x)}{\partial x_n\partial x_1} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_n^2}\end{bmatrix}.$$

多元函数的一阶 Taylor 展开式(中值定理)如下：

$$\begin{aligned}f(x)&=f(y)+\nabla f(y+\theta(x-y))^{\mathrm{T}}(x-y)\\&=f(y)+\nabla f(y)^{\mathrm{T}}(x-y)+o(\|x-y\|),\end{aligned}$$

其中 $\theta\in(0,1)$.

多元函数的二阶 Taylor 展开式如下：

$$\begin{aligned}f(x)&=f(y)+\nabla f(y)^{\mathrm{T}}(x-y)+\frac{1}{2}(x-y)^{\mathrm{T}}\nabla^2 f(y+\theta(x-y))(x-y)\\&=f(y)+\nabla f(y)^{\mathrm{T}}(x-y)+\frac{1}{2}(x-y)^{\mathrm{T}}\nabla^2 f(y)(x-y)\\&\quad+o(\|x-y\|^2),\end{aligned}$$

其中 $\theta\in(0,1)$.

向量值函数有类似的中值定理. 设 $F=(F_1,\cdots,F_m)^{\mathrm{T}}: R^n\to R^m$ 连续可微. $F'(x)$ 表示 F 在 x 处的 Jacobi 矩阵，则有

$$\begin{aligned}F(x)&=F(y)+\int_0^1 F'(y+\tau(x-y))\mathrm{d}\tau(x-y)=F(y)+F'(y)(x-y)\\&\quad+o(\|x-y\|).\end{aligned}$$

如无特别说明，本书所用到的向量范数均为Euclid 范数，即对 $x\in R^n$，$\|x\|=(x^{\mathrm{T}}x)^{1/2}$. 对矩阵 $A\in R^{n\times n}$，$\|A\|$ 表示由向量范数 $\|\cdot\|$ 导出的范数，即

$$\|A\|=\max_{\|x\|=1}\|Ax\|=\max\left\{\frac{\|Ax\|}{\|x\|}\,\middle|\, x\in R^n, x\neq 0\right\}.$$

我们以 $\|\cdot\|_F$ 表示矩阵的Frobenius 范数，即

$$\|A\|_F=\Big(\sum_{i,j=1}^{n}a_{ij}^2\Big)^{1/2},\quad \forall A=(a_{ij})\in R^{n\times n}.$$

§1.2 凸集和凸函数

本节介绍凸集、凸函数的定义及其基本性质.

§1.2.1 凸集

定义 1.2.1 若集合 $S\subseteq R^n$ 满足：

$$\alpha x+(1-\alpha)y\in S,\quad \forall x,y\in S,\forall\alpha\in[0,1], \tag{1.8}$$

则称 S 是 R^n 中的凸集.

从几何的角度，凸集 S 可解释为：若 S 包含点 x,y，则它包含了 x 与 y 的连线，如图1.2 所示.

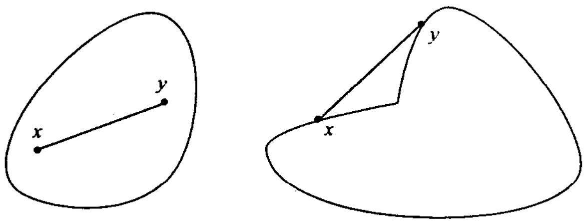

图1.2 凸集与非凸集

例1.2.1 设$r>0$，则R^n 中的球$S\triangleq\{x\in R^n\mid\|x\|\leqslant r\}$是凸集. 设$a\in R^n$ 是给定的向量，$b\in R$ 是一个常数，则 R^n 中超平面 $\pi\triangleq\{x\in R^n\mid a^Tx=b\}$ 是一个凸集. 设 $b\in R^m$，$a_i\in R^n$，$i=1,\cdots,m$ 是给定的向量，则多面体 $\pi_1\triangleq\{x\in R^n\mid a_i^Tx\leqslant b_i,i=1,\cdots,m\}$ 和 $\pi_2\triangleq\{x\in R^n\mid a_i^Tx=b_i,x_i\geqslant 0,i=1,\cdots,m\}$ 都是凸集.

容易证明，R^n 中的凸集有如下性质：

1. 若 $S\subseteq R^n$ 是凸集，则对任意 $\alpha\in R$，集合

$$\alpha S\triangleq\{\alpha x\mid x\in S\}$$

是 R^n 中的凸集.

2. 若$S_1,S_2\subseteq R^n$是凸集,则集合$S_1\cap S_2$,

$$S_1+S_2\triangleq\{x=x^{(1)}+x^{(2)}\mid x^{(1)}\in S_1,x\backslash+(2)\in S_2\}$$

和

$$S_1-S_2\triangleq\{x=x^{(1)}-x^{(2)}\mid x^{(1)}\in S_1,x^{(2)}\in S_2\}$$

都是R^n中的凸集.

定义1.2.2 设$C\subseteq R^n$. 若

$$\lambda x\in C,\quad \forall\lambda\in R,\lambda\geqslant 0,\forall x\in C,$$

则称C是R^n中的一个锥. 若C是锥且为凸集,则称它是R^n中的一个凸锥.

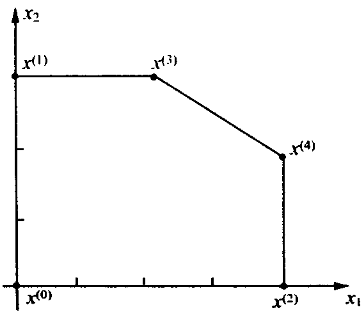

图1.3 凸集的顶点

R^n中凸锥是一种重要的凸集. 不难证明,如果在例1.2.1中取$b=0$,则例中的π_1和π_2都是凸锥.

定义1.2.3 设$S\subseteq R^n$是闭凸集,$x\in S$. 若不存在$x^{(1)},x^{(2)}\in S$以及数$\alpha\in(0,1)$,使得$x=\alpha x^{(1)}+(1-\alpha)x^{(2)}$,则称$x$是凸集$S$的一个顶点或极点,即$x\in S$是顶点的充要条件是$x$不能表示为$S$中两个不同点的凸组合.

例1.2.2 凸集

$$S=\{(x_1,x_2)^{\mathrm{T}}\mid x_1+2x_2\leqslant 8,0\leqslant x_1\leqslant 4,0\leqslant x_2\leqslant 3\}$$

中有顶点: $x^{(0)}=(0,0)^{\mathrm{T}},x^{(1)}=(0,3)^{\mathrm{T}},x^{(2)}=(4,0)^{\mathrm{T}},x^{(3)}=(2,3)^{\mathrm{T}},x^{(4)}=(4,2)^{\mathrm{T}}$. 如图1.3所示.

凸集可以有无限个顶点. 如单位圆

$$S=\{x\in R^2\mid \|x\|\leqslant 1\}$$

的边界上的任意点都是顶点.

定义1.2.4 设$S\subseteq R^n$是闭凸集,$d\in R^n$为非零向量. 若对任意$x\in S$,均有

$$\{x+\alpha d\mid \alpha\geqslant 0\}\subseteq S,$$

则称d是S的一个方向. 若S的方向d不能表示为S的其他两个不同方向的正线性组合,则称它为S的一个极方向.

由上面的定义易知,有界集合没有方向. 因此,例1.2.2中的S没有

方向.

例 1.2.3 凸集

$$S=\{x\in R^2 \mid x_1-4x_2\leqslant 0,3x_1-x_2\geqslant 0\}$$

有两个极方向:

$$d^{(1)}=(4,1)^{\mathrm{T}} \quad 和 \quad d^{(2)}=(1,3)^{\mathrm{T}}.$$

$d^{(1)}$和$d^{(2)}$的任何非负线性组合都是S的方向. 如图1.4所示.

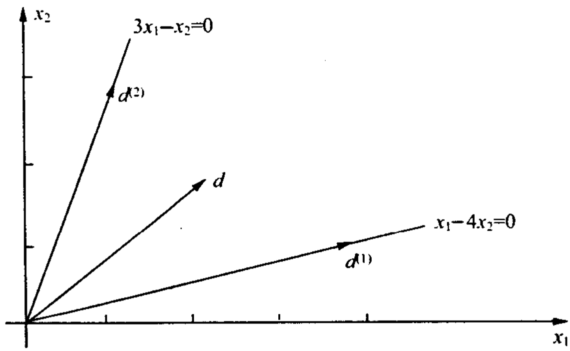

图1.4 凸集的方向与极方向

例 1.2.4 设$A\in R^{m\times n}$,$b\in R^m$,

$$S=\{x \mid Ax=b,x\geqslant 0\}.$$

则$d\in R^n$是S的一个方向的充要条件是

$$d\geqslant 0 \quad 且 \quad Ad=0.$$

证明 由定义,$d\in R^n$是S的方向的充要条件是:对任何$x\in S$,有

$$\{x+\alpha d \mid \alpha\geqslant 0\}\subseteq S,$$

即

$$A(x+\alpha d)=b, \quad x+\alpha d\geqslant 0, \quad \forall x\in S,\forall \alpha\geqslant 0.$$

从而,$Ad=0$,$d\geqslant 0$. **证毕**

定理 1.2.1 (表示定理)设$A\in R^{m\times n}$,$b\in R^m$,

$$S=\{x \mid Ax=b,x\geqslant 0\}.$$

则

(1) 凸集S有有限个顶点$x^{(1)},x^{(2)},\cdots,x^{(r)}$.

(2) S有极方向的充要条件是S无界. 而且,若S无界,则存在有限个极方向$d^{(1)},\cdots,d^{(t)}$.

(3) $x\in S$ 的充要条件是：存在非负数$\alpha_i\in R, i=1,\cdots,r$ 和非负数$\beta_j\in R, j=1,2,\cdots,t$，使得

$$x=\sum_{i=1}^{r}\alpha_i x^{(i)}+\sum_{j=1}^{t}\beta_j d^{(j)}.$$

$$\sum_{i=1}^{r}\alpha_i=1.$$

§1.2.2 凸函数

凸函数的定义如下：

定义 1.2.5 设$S\subseteq R^n$ 是凸集. 若函数 $f:R^n\to R$ 满足：

$$f(\alpha x+(1-\alpha)y)\leqslant \alpha f(x)+(1-\alpha)f(y),\quad \forall x,y\in S,\forall \alpha\in[0,1], \tag{1.9}$$

则称f 是S 上的凸函数. 若不等式(1.9)对所有$x\neq y$ 和$\alpha\in(0,1)$成立严格不等式，则称 f 是 S 上的严格凸函数. 若存在常数 $m>0$，使得不等式

$$f(\alpha x+(1-\alpha)y)\leqslant \alpha f(x)+(1-\alpha)f(y)-m\alpha(1-\alpha)\|x-y\|^2 \tag{1.10}$$

对所有 $x,y\in S$ 以及所有 $\alpha\in[0,1]$成立，则称 f 是 S 上的一致凸函数(或强凸函数).

由定义 1.2.5 不难看出一致凸函数是严格凸函数，严格凸函数是凸函数.

凸函数的几何解释为：函数的图像上的任意两点确定的弦在其图像的上方. 如图 1.5 所示.

不难证明，R^n 中的凸函数有如下性质：

1. 设 $f_1,f_2:R^n\to R$ 是凸函数，则对任意 $\alpha_1\geqslant 0,\alpha_2\geqslant 0$，函数 $\alpha_1 f_1+\alpha_2 f_2$ 是凸函数.

2. 设$f_i:R^n\to R, i=1,\cdots,m$ 是凸函数，则对任意$\alpha_i\geqslant 0, i=1,\cdots,m$，函数$\sum_{i=1}^{m}\alpha_i f_i$ 是凸函数.

下面的定理给出了凸函数的几个等价性条件.

定理 1.2.2 设函数 $f:R^n\to R$ 二次连续可微. 则下面的命题等价：

(1) 函数 f 是凸函数.

(2) 对任意$x,y\in R^n$，一元函数$\phi(t)\triangleq f(tx+(1-t)y)$是$[0,1]$上的凸

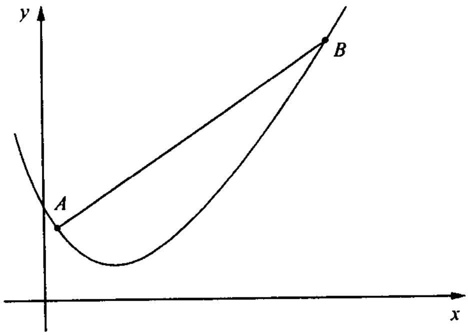

图 1.5　凸函数

函数.

(3) 对任意的 $x,y\in R^n$，下面的不等式成立：

$$f(x)-f(y)\geqslant \nabla f(y)^{\mathrm{T}}(x-y). \tag{1.11}$$

(4) 梯度函数 ∇f 单调，即：

$$(\nabla f(x)-\nabla f(y))^{\mathrm{T}}(x-y)\geqslant 0,\quad \forall x,y\in R^n. \tag{1.12}$$

(5) 对所有 $x\in R^n$，$\nabla^2 f(x)$ 半正定.

证明　先证明命题(1)与命题(2)等价. 设命题(1)成立. 对给定的 x，$y\in R^n$，$t_1,t_2\in[0,1]$，$\alpha\in[0,1]$，有

$$\begin{aligned}\phi(\alpha t_1+(1-\alpha)t_2)&=f((\alpha t_1+(1-\alpha)t_2)x+(1-(\alpha t_1+(1-\alpha)t_2))y)\\&=f(\alpha(t_1x+(1-t_1)y)+(1-\alpha)(t_2x+(1-t_2)y)\\&\leqslant\alpha\phi(t_1)+(1-\alpha)\phi(t_2),\end{aligned}$$

即命题(2)成立.

反之，设命题(2)成立. 对任意 $x,y\in R^n$ 及 $\alpha\in(0,1]$，有

$$\begin{aligned}f(\alpha x+(1-\alpha)y)&=\phi(\alpha\cdot 1+(1-\alpha)\cdot 0)\\&\leqslant\alpha\phi(1)+(1-\alpha)\phi(0)=\alpha f(x)+(1-\alpha)f(y),\end{aligned}$$

即命题(1)成立.

下面按如下顺序证明命题(1)，(3)，(4)，(5)等价：命题(1)⇒命题(3)⇒命题(4)⇒命题(5)⇒命题(1).

命题(1)⇒命题(3). 对任意 $x,y\in R^n$ 及 $\alpha\in(0,1]$，由 f 的凸性，有

$$f(\alpha x+(1-\alpha)y)\leqslant\alpha f(x)+(1-\alpha)f(y).$$

由此可得：

$$f(x)-f(y)\geqslant\frac{f(y+\alpha(x-y))-f(y)}{\alpha}.$$

上式对 $\alpha\to0^+$ 取极限即得(1.11).

命题(3)⇒命题(4). 对任意 $x,y\in R^n$，由命题(3)有

$$f(x)-f(y)\geqslant\nabla f(y)^{\mathrm T}(x-y),\quad f(y)-f(x)\geqslant\nabla f(x)^{\mathrm T}(y-x).$$

上面的两式相加即得(1.12).

命题(4)⇒命题(5). 对任意 $x,p\in R^n$ 及 $t\in(0,1)$，令 $y=x+tp$. 由(1.12)得

$$t(\nabla f(x+tp)-\nabla f(x))^{\mathrm T}p\geqslant0.$$

上式两端同除以 t^2 后再对 $t\to0^+$ 取极限，得 $p^{\mathrm T}\nabla^2f(x)p\geqslant0$，即命题(5)成立.

命题(5)⇒命题(1). 对任意 $x,y\in R^n$ 及 $\alpha\in[0,1]$，令 $z=\alpha x+(1-\alpha)y$. 由 Taylor 展开并利用命题(5)，可得：

$$f(x)\geqslant f(z)+\nabla f(z)^{\mathrm T}(x-z),$$
$$f(y)\geqslant f(z)+\nabla f(z)^{\mathrm T}(y-z).$$

上面两式分别乘以 α 和 $1-\alpha$ 后相加，注意到 z 的定义即得命题(1). **证毕**

定义 1.2.6 若函数 $-f$ 是凸函数，则称函数 f 是凹函数.

类似于定理1.2.2，不难建立凹函数相应的等价性定理.

定理 1.2.3 设函数 $f:R^n\to R$ 二次连续可微，则下面的命题等价：

(1) 函数 f 是凹函数.

(2) 对任意 $x,y\in R^n$，一元函数 $\phi(t)\triangleq f(tx+(1-t)y)$ 是 $[0,1]$ 上的凹函数.

(3) 下面的不等式成立：

$$f(x)-f(y)\leqslant\nabla f(y)^{\mathrm T}(x-y),\quad \forall x,y\in R^n.$$

(4) 下面的不等式成立：

$$(\nabla f(x)-\nabla f(y))^{\mathrm T}(x-y)\leqslant0,\quad \forall x,y\in R^n.$$

(5) 对所有 $x\in R^n$，$\nabla^2f(x)$ 半负定.

定理 1.2.4 设 $S\subseteq R^n$ 是凸集，则 $f:S\to R$ 是凸函数的充要条件是 f 的上图像

$$P(f)\triangleq\{(x,\alpha)\mid x\in S,\alpha\in R,f(x)\leqslant\alpha\}$$

是 R^{n+1} 中的凸集.

证明 必要性. 设 $f:S\to R$ 是凸函数. 对任意 $(x^{(1)},\alpha_1),(x^{(2)},\alpha_2)\in$