21世纪高等院校教材

最优化方法
及其Matlab程序设计

马昌凤 编著

科学出版社
www.sciencep.com

21 世纪高等院校教材

最优化方法
及其 Matlab 程序设计

马昌凤　编著

科学出版社
北京

内 容 简 介

本书较系统地介绍了非线性最优化问题的基本理论和算法，以及主要算法的 Matlab 程序设计. 主要内容包括（精确或非精确）线搜索技术、最速下降法与（修正）牛顿法、共轭梯度法、拟牛顿法、信赖域方法、非线性最小二乘问题的解法、约束优化问题的最优性条件、罚函数法、可行方向法、二次规划问题的解法、序列二次规划法等. 设计的 Matlab 程序有精确线搜索的 0.618 法和抛物线法、非精确线搜索的 Armijo 准则、最速下降法、牛顿法、再开始共轭梯度法、BFGS 算法、DFP 算法、Broyden 族方法、信赖域方法、求解非线性最小二乘问题的 L-M 算法、解约束优化问题的乘子法、求解二次规划的有效集法、SQP 子问题的光滑牛顿法以及求解约束优化问题的 SQP 方法等. 此外，本书配有丰富的例题和习题，并在附录介绍了 Matlab 优化工具箱的使用方法. 本书既注重计算方法的实用性，又注意保持理论分析的严谨性，强调数值方法的思想和原理在计算机上的实现. 读者只需具备微积分、线性代数和 Matlab 程序设计方面的初步知识即可学习本书.

本书可供数学与应用数学、信息与计算科学专业的本科生，应用数学、计算数学、运筹学与控制论专业的研究生，理工科相关专业的研究生，对最优化理论与算法感兴趣的教师及科技工作者阅读.

图书在版编目(CIP)数据

最优化方法及其 Matlab 程序设计/马昌凤编著. —北京：科学出版社，2010

(21 世纪高等院校教材)

ISBN 978-7-03-028921-6

I. ①最… II. ①马… III. ①最优化算法-计算机辅助计算-软件包，MATLAB-高等学校-教材 IV. ①O242.23

中国版本图书馆 CIP 数据核字(2010)第 174850 号

责任编辑：姚莉丽 房 阳 / 责任校对：陈玉凤

责任印制：张克忠 / 封面设计：陈 敬

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

铭浩彩色印装有限公司 印刷

科学出版社发行 各地新华书店经销

*

2010 年 8 月第 一 版 开本：B5(720 × 1000)

2010 年 8 月第一次印刷 印张：14 3/4

印数：1—3 000 字数：290 000

定价：32.00 元（含光盘）

（如有印装质量问题，我社负责调换）

前　言

运筹学的理论与方法广泛应用于工业与农业、交通与运输、国防与建筑, 以及通信与管理等各个部门和领域, 它主要解决最优计划、最优分配、最优决策以及最佳设计和最佳管理等最优化问题. 本书所介绍的最优化方法又称为数学规划, 是运筹学的一个重要分支, 也是计算数学和应用数学的一个重要组成部分.

本书系统地介绍了非线性优化的理论与方法, 及其 Matlab 程序设计, 适合数学与应用数学、信息与计算科学专业的本科生, 应用数学、计算数学、运筹学与控制论专业的研究生, 理工科相关专业的研究生, 对最优化理论与算法感兴趣的教师及科技工作者阅读. 读者只需具备微积分、线性代数和 Matlab 程序设计方面的初步知识.

本书的主要内容包括最优化理论基础、(精确或非精确) 线搜索技术、最速下降法与 (修正) 牛顿法、共轭梯度法、拟牛顿法、信赖域方法、非线性最小二乘问题的解法、(约束优化问题的) 最优性条件、罚函数法、可行方向法、二次规划问题的解法、序列二次规划法等. 设计的 Matlab 程序有精确线搜索的 0.618 法和抛物线法、非精确线搜索的 Armijo 准则、最速下降法、牛顿法、再开始共轭梯度法、对称秩 1 算法、BFGS 算法、DFP 算法、Broyden 族方法、信赖域方法、求解非线性最小二乘问题的 L-M 算法、解约束优化问题的乘子法、求解二次规划的有效集法、牛顿–拉格朗日算法、SQP 子问题的光滑牛顿法以及求解约束优化问题的 SQP 方法等. 此外, 本书配有丰富的例题和习题, 并在附录介绍了 Matlab 优化工具箱的使用方法. 本书既注重计算方法的实用性, 又注意保持理论分析的严谨性, 强调数值方法的思想和原理在计算机上的实现.

本书具有如下特点:

1. 介绍非线性优化中最重要、最基础的理论与方法, 它们是研究各种复杂的最优化问题的基础和工具.

2. 最优化方法与程序设计相结合, 采用当前最流行的数学软件 Matlab 编制了主要优化算法的程序. 所有 Matlab 程序都在计算机上经过调试和运行, 简洁而不失准确性.

3. 书中所给的每一程序之后都有相应的计算实例. 这不仅能帮助学生理解程序中所包含的最优化理论知识, 而且对培养学生处理数值最优化问题的能力也大有裨益.

4. 每章都配备了一定数量的习题, 帮助学生理解和巩固所学的知识.

本书的编写和出版得到了“福建省高校服务海西建设重点项目 —— 基于数学的信息化技术研究”经费的资助以及福建省自然科学基金项目 (编号: 2009J01002) 的部分资助, 在此, 作者表示由衷的感谢. 作者还要感谢福建师范大学教务处以及数学与计算机科学学院给予的帮助和支持. 此外, 感谢夫人刘菊庄女士给予的理解和支持.

由于作者水平有限, 加之时间仓促, 书中的疏漏和不足之处在所难免, 敬请专家和读者批评指正.

作　者

2009 年 12 月

目　　录

第 1 章　最优化理论基础

1.1　最优化问题的数学模型

通俗地说，最优化问题就是求一个多元函数在某个给定集合上的极值. 几乎所有类型的最优化问题都可以用下面的数学模型来描述:

$$\begin{aligned} \min \quad & f(x), \\ \text{s.t.} \quad & x \in K, \end{aligned} \tag{1.1}$$

其中, K 为某个给定的集合 (称为可行集或可行域), $f(x)$ 为定义在集合 K 上的实值函数. 此外, 在模型 (1.1) 中, x 通常称为决策变量, s.t. 是 subject to (受限于) 的缩写.

人们通常按照可行集的性质对最优化问题 (1.1) 进行一个大致的分类.

(1) 线性规划和非线性规划：可行集是有限维空间中的一个子集.

(2) 组合优化或网络规划：可行集中的元素是有限的.

(3) 动态规划：可行集是一个依赖于时间的决策序列.

(4) 最优控制：可行集是无穷维空间中的一个连续子集.

本书主要考虑在工程设计中有着重要应用的非线性规划, 其数学模型为

$$\min \quad f(x), \tag{1.2}$$

$$\text{s.t.} \quad h_i(x) = 0,\ i = 1, \cdots, l, \tag{1.3}$$

$$g_i(x) \geqslant 0,\ i = 1, \cdots, m, \tag{1.4}$$

其中, $f(x)$, $h_i(x)$ $(i = 1, \cdots, l)$ 及 $g_i(x)$ $(i = 1, \cdots, m)$ 都是定义在 $\mathbb{R}^n$ 上连续可微的多元实值函数, 并且至少有一个是非线性的. 记

$$E = \{i | h_i(x) = 0\}, \quad I = \{i | g_i(x) \geqslant 0\}. \tag{1.5}$$

若指标集 $E \cup I = \varnothing$, 则称之为无约束优化问题; 否则称为约束优化问题. 特别地, 把 $E \neq \varnothing$ 且 $I = \varnothing$ 的优化问题称为等式约束优化问题; 而把 $I \neq \varnothing$ 且 $E = \varnothing$ 的优化问题称为不等式约束优化问题. $f(x)$ 称为目标函数, $h_i(x)$, $g_j(x)$ $(i = 1, \cdots, l; j = 1, \cdots, m)$ 称为约束函数. 此外, 通常把目标函数为二次函数, 而约束函数都为线性函数的优化问题称为二次规划, 而目标函数和约束函数都为线性函数的优化问题称为线性规划.

1.2 向量和矩阵范数

在算法的收敛性分析中, 需要用到向量和矩阵范数的概念及其有关理论.

设 $\mathbb{R}^n$ 表示实 n 维向量空间, $\mathbb{R}^{n\times n}$ 表示实 n 阶矩阵全体所组成的线性空间. 在这两个空间中, 分别定义向量和矩阵的范数.

向量 $x\in\mathbb{R}^n$ 的范数 $\|x\|$ 是一个非负数, 它必须满足以下条件:

(1) $\|x\|\geqslant 0$, $\|x\|=0 \Longleftrightarrow x=0$;

(2) $\|\lambda x\|=|\lambda|\|x\|$, $\lambda\in\mathbb{R}$;

(3) $\|x+y\|\leqslant\|x\|+\|y\|$.

向量 $x=(x_1,\cdots,x_n)^{\mathrm{T}}$ 的 p 范数定义为

$$\|x\|_p=\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}. \tag{1.6}$$

常用的向量范数有如下几个:

1 范数: $\|x\|_1=\sum\limits_{i=1}^{n}|x_i|$;

2 范数: $\|x\|_2=\left(\sum\limits_{i=1}^{n}|x_i|^2\right)^{\frac{1}{2}}$;

∞ 范数: $\|x\|_\infty=\max\limits_{1\leqslant i\leqslant n}|x_i|$.

矩阵 $A\in\mathbb{R}^{n\times n}$ 的范数是一个非负实数, 它除了满足跟向量范数相似的三条性质之外, 还必须具备如下乘法性质:

(4) $\|AB\|\leqslant\|A\|\|B\|, A,B\in\mathbb{R}^{n\times n}$.

如果一矩阵范数 $\|\cdot\|_\mu$ 相对于某向量范数 $\|\cdot\|$ 满足下面的不等式:

(5) $\|Ax\|\leqslant\|A\|_\mu\|x\|,\quad x\in\mathbb{R}^n$,

则称矩阵范数 $\|\cdot\|_\mu$ 和向量范数 $\|\cdot\|$ 是相容的. 进一步, 若存在 $x\neq 0$, 使得成立

$$\|A\|_\mu=\max_{x\neq 0}\frac{\|Ax\|}{\|x\|}=\max_{\|x\|=1}\|Ax\|,\quad A\in\mathbb{R}^{n\times n}, \tag{1.7}$$

则称矩阵范数 $\|\cdot\|_\mu$ 为由向量范数 $\|\cdot\|$ 诱导出来的算子范数, 简称为算子范数, 有时也称为从属于向量范数 $\|\cdot\|$ 的矩阵范数. 此时, 向量范数和算子范数通常采用相同的符号 $\|\cdot\|$.

不难验证, 从属于向量范数 $\|x\|_\infty$, $\|x\|_1$, $\|x\|_2$ 的矩阵范数分别为

$$\|A\|_\infty=\max_{1\leqslant i\leqslant n}\sum_{j=1}^{n}|a_{ij}|,$$

$$\|A\|_1 = \max_{1\leqslant j\leqslant n} \sum_{i=1}^{n} |a_{ij}|,$$

$$\|A\|_2 = \max\{\sqrt{\lambda} \mid \lambda \in \lambda(A^{\mathrm{T}}A)\},$$

它们分别称为行和范数、列和范数和谱范数.

本书在讨论各种迭代算法的收敛性时, 通常采用谱范数和按下述方式定义的F范数:

$$\|A\|_F = \left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2\right)^{1/2} = \sqrt{\mathrm{tr}(A^{\mathrm{T}}A)}. \tag{1.8}$$

现在来讨论向量序列和矩阵序列的收敛性. 已经知道, 若 $\{x^{(k)}\}_{k=1}^{\infty} \subset \mathbb{R}^n$, 则

$$\lim_{k\to\infty} x^{(k)} = x \Longleftrightarrow \lim_{k\to\infty} x_i^{(k)} = x_i, i = 1, \cdots, n.$$

类似地, 若 $\{A^{(k)}\}_{k=1}^{\infty} \subset \mathbb{R}^{n\times n}$, 则

$$\lim_{k\to\infty} A^{(k)} = A \Longleftrightarrow \lim_{k\to\infty} a_{ij}^{(k)} = a_{ij}, i,\ j = 1, \cdots, n.$$

为了利用范数来描述上述极限, 必须建立向量范数的等价定理以及矩阵范数的等价定理.

定理 1.1 (1) 设 $\|\cdot\|$ 和 $\|\cdot\|'$ 是定义在 $\mathbb{R}^n$ 上的两个向量范数, 则存在两个正数 c_1, c_2, 对所有 $x \in \mathbb{R}^n$ 均成立

$$c_1\|x\| \leqslant \|x\|' \leqslant c_2\|x\|;$$

(2) 设 $\|\cdot\|$ 和 $\|\cdot\|'$ 是定义在 $\mathbb{R}^{n\times n}$ 上的两个矩阵范数, 则存在两个正数 m_1, m_2, 对所有 $A \in \mathbb{R}^{n\times}$ 均成立

$$m_1\|A\| \leqslant \|A\|' \leqslant m_2\|A\|.$$

下面利用范数的概念来等价地定义向量序列和矩阵序列的收敛性.

定理 1.2 (1) 设 $\{x^{(k)}\}$ 为 n 维向量序列, $\|\cdot\|$ 为定义在 $\mathbb{R}^n$ 上的向量范数, 则

$$\lim_{k\to\infty} x^{(k)} = x \Longleftrightarrow \lim_{k\to\infty} \|x^{(k)} - x\| = 0;$$

(2) 设 $\{A^{(k)}\}$ 为 $n\times n$ 矩阵序列, $\|\cdot\|$ 为定义在 $\mathbb{R}^{n\times n}$ 上的向量范数, 则

$$\lim_{k\to\infty} A^{(k)} = A \Longleftrightarrow \lim_{k\to\infty} \|A^{(k)} - A\| = 0.$$

1.3 函数的可微性与展开

本节主要介绍后文需要经常用到的 n 元函数的一阶和二阶导数以及泰勒展开式.

定义 1.1 设有 n 元实函数 $f(x)$, 其中自变量 $x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{R}^n$. 称向量

$$\nabla f(x)=\left(\frac{\partial f(x)}{\partial x_1},\frac{\partial f(x)}{\partial x_2},\cdots,\frac{\partial f(x)}{\partial x_n}\right)^{\mathrm{T}} \tag{1.9}$$

为 $f(x)$ 在 x 处的一阶导数或梯度. 称矩阵

$$\nabla^2 f(x)=\begin{bmatrix} \dfrac{\partial^2 f(x)}{\partial x_1^2} & \dfrac{\partial^2 f(x)}{\partial x_1\partial x_2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_1\partial x_n} \\ \dfrac{\partial^2 f(x)}{\partial x_2\partial x_1} & \dfrac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_2\partial x_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial^2 f(x)}{\partial x_n\partial x_1} & \dfrac{\partial^2 f(x)}{\partial x_n\partial x_2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \tag{1.10}$$

为 $f(x)$ 在 x 处的二阶导数或 Hesse 矩阵. 若梯度 $\nabla f(x)$ 的每个分量函数在 x 都连续, 则称 f 在 x 一阶连续可微. 若 Hesse 矩阵 $\nabla^2 f(x)$ 的各个分量函数都连续, 则称 f 在 x 二阶连续可微.

若 f 在开集 D 的每一点都连续可微, 则称 f 在 D 上一阶连续可微. 若 f 在开集 D 的每一点都二阶连续可微, 则称 f 在 D 上二阶连续可微.

由定义 1.1 不难发现, 若 f 在 x 二阶连续可微, 则

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

即 Hesse 矩阵 $\nabla^2 f(x)$ 是对称阵.

例 1.1 设二次函数

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

其中 $c\in\mathbb{R}^n$, $H\in\mathbb{R}^{n\times n}$ 是对称阵. 那么, 不难计算它在 x 的梯度及 Hesse 矩阵为

$$\nabla f(x)=c+Hx,\quad \nabla^2 f(x)=H.$$

例 1.2(泰勒展开) 设函数 $f:\mathbb{R}^n\to\mathbb{R}$ 连续可微, 那么

$$\begin{aligned} f(x+h)&=f(x)+\int_0^1\nabla f(x+\tau h)^{\mathrm{T}}h\mathrm{d}\tau \\ &=f(x)+\nabla f(x+\xi h)^{\mathrm{T}}h\quad(\xi\in(0,1)) \\ &=f(x)+\nabla f(x)^{\mathrm{T}}h+o(\|h\|). \end{aligned}$$

进一步, 若函数 f 是二次连续可微的, 则有

$$\begin{aligned} f(x+h) &= f(x) + \nabla f(x)^{\mathrm{T}} h + \int_0^1 (1-\tau) h^{\mathrm{T}} \nabla^2 f(x+\tau h) h \mathrm{d}\tau \\ &= f(x) + \nabla f(x)^{\mathrm{T}} h + \frac{1}{2} h^{\mathrm{T}} \nabla^2 f(x+\xi h) h \quad (\xi \in (0,1)) \\ &= f(x) + \nabla f(x)^{\mathrm{T}} h + \frac{1}{2} h^{\mathrm{T}} \nabla^2 f(x) h + o(\|h\|^2) \end{aligned}$$

及

$$\begin{aligned} \nabla f(x+h) &= \nabla f(x) + \int_0^1 \nabla^2 f(x+\tau h)^{\mathrm{T}} h \mathrm{d}\tau \\ &= \nabla f(x) + \nabla^2 f(x+\xi h)^{\mathrm{T}} h \quad (\xi \in (0,1)) \\ &= \nabla f(x) + \nabla^2 f(x)^{\mathrm{T}} h + o(\|h\|). \end{aligned}$$

下面简单介绍一下向量值函数的可微性及中值定理. 设有向量值函数 $F = (F_1, F_2, \cdots, F_m)^{\mathrm{T}} : \mathbb{R}^n \to \mathbb{R}^m$. 若每个分量函数 F_i 都是 (连续) 可微的, 则称 F 为 (连续) 可微的. 向量值函数 F 在 x 的导数 $F' \in \mathbb{R}^{m \times n}$ 是指它在 x 的 Jacobi 矩阵, 记为 $F'(x)$ 或 $J_F(x)$, 即

$$F'(x) := J_F(x) := \begin{bmatrix} \dfrac{\partial F_1(x)}{\partial x_1} & \dfrac{\partial F_1(x)}{\partial x_2} & \cdots & \dfrac{\partial F_1(x)}{\partial x_n} \\ \dfrac{\partial F_2(x)}{\partial x_1} & \dfrac{\partial F_2(x)}{\partial x_2} & \cdots & \dfrac{\partial F_2(x)}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial F_m(x)}{\partial x_1} & \dfrac{\partial F_m(x)}{\partial x_2} & \cdots & \dfrac{\partial F_m(x)}{\partial x_n} \end{bmatrix}.$$

考虑到标量函数的梯度定义, 有时也把向量函数 F 的 Jacobi 矩阵的转置称为 F 在 x 的梯度, 记为

$$\nabla F(x) = J_F(x)^{\mathrm{T}} = \big(\nabla F_1(x), \nabla F_2(x), \cdots, \nabla F_m(x)\big).$$

不难发现, 例 1.2 中关于多元实函数的一些结论可以推广到向量值函数的情形. 例如, 若向量值函数 $F : \mathbb{R}^n \to \mathbb{R}^m$ 是连续可微的, 则对于任意的 x, $h \in \mathbb{R}^n$ 有

$$F(x+h) = F(x) + \int_0^1 \nabla F(x+\tau h)^{\mathrm{T}} h \mathrm{d}\tau.$$

对于向量值函数 F, 也可以定义 Lipschitz 连续性的概念.

定义 1.2 设向量值函数 $F : \mathbb{R}^n \to \mathbb{R}^m$, $x \in \mathbb{R}^n$, 称 F 在 x 为 Lipschitz 连续的, 是指存在常数 $L > 0$, 使得对任意的 $y \in \mathbb{R}^n$, 满足

$$\|F(x) - F(y)\| \leqslant L \|x - y\|, \tag{1.11}$$

其中, L 称为 Lipschitz 常数. 若式 (1.11) 对任意的 x, $y \in \mathbb{R}^n$ 都成立, 则称 F 在 $\mathbb{R}^n$ 内为 Lipschitz 连续的.

在迭代法的收敛性分析中, 有时需要用到向量值函数的"中值定理", 现引述如下:

定理 1.3　设向量值函数 $F:\mathbb{R}^n \to \mathbb{R}^m$ 连续可微, 那么

(1) 对任意的 $x, y \in \mathbb{R}^n$ 有

$$\|F(x)-F(y)\| \leqslant \sup_{0\leqslant t\leqslant 1}\|F'(y+t(x-y))\|\,\|x-y\|;$$

(2) 对任意的 $x, y, z \in \mathbb{R}^n$ 有

$$\|F(y)-F(z)-F'(x)(y-z)\| \leqslant \sup_{0\leqslant t\leqslant 1}\|F'(z+t(y-z))-F'(x)\|\,\|x-y\|.$$

由定理 1.3 (2) 可推得下面的结论.

推论 1.1　设向量值函数 $F:\mathbb{R}^n \to \mathbb{R}^m$ 是连续可微的, 并且其 Jacobi 矩阵映射是 Lipschitz 连续的, 即存在常数 $L>0$, 使得

$$\|F'(u)-F'(v)\| \leqslant L\|u-v\|, \quad \forall\, u, v \in \mathbb{R}^n, \tag{1.12}$$

则对任意的 $x, h \in \mathbb{R}^n$ 有

$$\|F(x+h)-F(x)-F'(x)h\| \leqslant \frac{1}{2}L\|h\|^2. \tag{1.13}$$

1.4　凸集与凸函数

本节介绍凸集、锥和多面体的有关概念.

定义 1.3　设集合 $D \subset \mathbb{R}^n$. 称集合 D 为凸集, 是指对任意的 $x, y \in D$ 及任意的实数 $\lambda \in [0,1]$, 都有 $\lambda x + (1-\lambda)y \in D$.

由定义 1.3 不难知道凸集的几何意义, 即对非空集合 $D \subset \mathbb{R}^n$, 若连接其中任意两点的线段仍属于该集合, 则称该集合 D 为凸集.

不难证明凸集的下列基本性质:

引理 1.1　设 D, D_1, D_2 是凸集, α 是一实数, 那么

(1) $\alpha D = \{y | y = \alpha x,\ x \in D\}$ 是凸集;

(2) 交集 $D_1 \cap D_2$ 是凸集;

(3) 和集 $D_1 + D_2 = \{z | z = x + y,\ x \in D_1,\ y \in D_2\}$ 也是凸集.

例 1.3　n 维欧氏空间中 l 个点的凸组合是一个凸集, 即集合

$$\left\{x = \sum_{i=1}^{l}\alpha_i x_i \,\middle|\, x_i \in \mathbb{R}^n,\ \alpha_i \geqslant 0,\ \sum_{i=1}^{l}\alpha_i = 1\right\}$$

是凸集.

例 1.4 n 维欧氏空间中的超平面 $H=\{x|c^{\mathrm{T}}x=\alpha\}$ 是一个凸集, 其中 $\alpha\in\mathbb{R}$, $c\in\mathbb{R}^n\backslash\{0\}$ 是超平面的法向量. 此外, 下面的 4 个半空间:

(1) 正的闭半空间 $H^+=\{x|c^{\mathrm{T}}x\geqslant\alpha\}$;

(2) 负的闭半空间 $H^+=\{x|c^{\mathrm{T}}x\leqslant\alpha\}$;

(3) 正的开半空间 $H^+=\{x|c^{\mathrm{T}}x>\alpha\}$;

(4) 负的开半空间 $H^+=\{x|c^{\mathrm{T}}x<\alpha\}$

都是凸集.

例 1.5 以 $x^0\in\mathbb{R}^n$ 为起点, $d\in\mathbb{R}^n\backslash\{0\}$ 为方向的射线

$$r(x^0;d):=\{x\in\mathbb{R}^n\,|\,x=x^0+\alpha d,\ \alpha\geqslant 0\}$$

是凸集.

定义 1.4 集合 $D\subset\mathbb{R}^n$ 的凸包 (convex hull) 是指所有包含 D 的凸集的交集, 记为

$$\mathrm{conv}(D):=\bigcap_{C\supseteq D}C,$$

其中 C 为凸集.

下面给出锥和凸锥的定义.

定义 1.5 设非空集合 $C\subset\mathbb{R}^n$. 若对任意的 $x\in C$ 和任意的实数 $\lambda>0$ 有 $\lambda x\in C$, 则称 C 为一个锥 (cone). 若 C 同时也是凸集, 则称 C 为一个凸锥 (convex cone). 此外, 对于锥 C, 若 $0\in C$, 则称 C 为一个尖锥 (pointed cone). 相应地, 包含 0 的凸锥称为尖凸锥.

例 1.6 多面体 $\{x\in\mathbb{R}^n|Ax\geqslant 0\}$ 是一个尖凸锥, 通常称之为多面锥 (polyhedral cone).

例 1.7 集合

$$\mathbb{R}^n_+:=\{x\in\mathbb{R}^n|x_i\geqslant 0,\ i=1,2,\cdots,n\}$$

是一个尖凸锥, 通常称之为非负锥 (nonnegative cone). 相应地, 凸锥

$$\mathbb{R}^n_{++}:=\{x\in\mathbb{R}^n|x_i>0,\ i=1,2,\cdots,n\}$$

称为正锥 (positive cone).

有了凸集的概念之后, 就可以定义凸集上所谓的凸函数.

定义 1.6 设函数 $f:D\subset\mathbb{R}^n\to\mathbb{R}$, 其中 D 为凸集.

(1) 称 f 为 D 上的凸函数, 是指对任意的 $x,y\in D$ 及任意的实数 $\lambda\in[0,1]$, 都有

$$f(\lambda x+(1-\lambda)y)\leqslant\lambda f(x)+(1-\lambda)f(y);$$

(2) 称 f 为 D 上的严格凸函数, 是指对任意的 $x, y \in D$, $x \neq y$ 及任意的实数 $\lambda \in [0,1]$, 都有

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y);$$

(3) 称 f 为 D 上的一致凸函数, 是指存在常数 $\gamma > 0$, 使得对任意的 $x, y \in D$ 及任意的实数 $\lambda \in [0,1]$, 都有

$$f(\lambda x + (1-\lambda)y) + \frac{1}{2}\lambda(1-\lambda)\gamma\|x-y\|^2 \leqslant \lambda f(x) + (1-\lambda)f(y).$$

凸函数具有下列基本性质:

命题 1.1　设 f, f_1, f_2 都是凸集 D 上的凸函数, $c_1, c_2 \in \mathbb{R}_+$, $\alpha \in \mathbb{R}$, 则有

(1) $cf_1(x) + c_2 f_2(x)$ 也是 D 上的凸函数;

(2) 水平集

$$\mathcal{L}(f,\alpha) = \{x | x \in D,\ f(x) \leqslant \alpha\}$$

是凸集.

凸集和凸函数在优化理论中起着举足轻重的作用, 但是利用凸函数的定义来判断一个函数是否具有凸性并非一件容易的事. 如果函数是一阶或二阶连续可微的, 则可利用函数的梯度或 Hesse 矩阵来判别或验证函数的凸性要相对容易一些. 下面给出几个判别定理.

定理 1.4　设 f 在凸集 $D \subset \mathbb{R}^n$ 上一阶连续可微, 则

(1) f 在 D 上为凸函数的充要条件是

$$f(x) \geqslant f(x^*) + \nabla f(x^*)^{\mathrm{T}}(x - x^*), \quad \forall\, x^*, x \in D; \tag{1.14}$$

(2) f 在 D 上严格凸的充要条件是当 $x \neq y$ 时, 成立

$$f(x) > f(x^*) + \nabla f(x^*)^{\mathrm{T}}(x - x^*), \quad \forall\, x^*, x \in D; \tag{1.15}$$

(3) f 在 D 上一致凸的充要条件是存在常数 $c > 0$, 使得对任意的 $x^*, x \in D$, 成立

$$f(x) \geqslant f(x^*) + \nabla f(x^*)^{\mathrm{T}}(x - x^*) + c\|x - x^*\|^2. \tag{1.16}$$

已经知道, 在一元函数中, 若 $f(x)$ 在区间 (a,b) 上二阶可微且 $f''(x) \geqslant 0\ (>0)$, 则 $f(x)$ 在 (a,b) 内凸 (严格凸). 对于二阶连续可微的多元函数 $f: D \subset \mathbb{R}^n \to \mathbb{R}$, 也可以由其二阶导数 (Hesse 矩阵) 给出凸性的一个近乎完整的表述.

定义 1.7　设 n 元实函数 f 在凸集 D 上是二阶连续可微的. 若对一切 $h \in \mathbb{R}^n$ 有 $h^{\mathrm{T}}\nabla^2 f(x)h \geqslant 0$, 则称 $\nabla^2 f$ 在点 x 处为半正定的. 若对一切 $0 \neq h \in \mathbb{R}^n$ 有 $h^{\mathrm{T}}\nabla^2 f(x)h > 0$, 则称 $\nabla^2 f$ 在点 x 处为正定的. 进一步, 若存在常数 $c > 0$, 使得对任意的 $h \in \mathbb{R}^n$, $x \in D$ 有 $h^{\mathrm{T}}\nabla^2 f(x)h \geqslant c\|h\|^2$, 则称 $\nabla^2 f$ 在 D 上为一致正定的.