清华大学研究生公共课教材——数学系列

最优化理论与算法（第2版）

陈宝林　编著

清华大学出版社

清华大学研究生公共课教材——数学系列

最优化理论与算法（第2版）

陈宝林　编著

清华大学出版社
北京

内容简介

本书是陈宝林教授在多年实践基础上编著的.书中包括线性规划单纯形方法、对偶理论、灵敏度分析、运输问题、内点算法、非线性规划 K-T 条件、无约束最优化方法、约束最优化方法、整数规划和动态规划等内容.本书含有大量经典的和新近的算法,有比较系统的理论分析,实用性比较强;定理的证明和算法的推导主要以数学分析和线性代数为基础,比较简单易学.本书可以作为运筹学类课程的教学参考书,也可供应用数学工作者和工程技术人员参考.

图书在版编目(CIP)数据

最优化理论与算法/陈宝林编著.—2版.—北京:清华大学出版社,2005.10
(清华大学研究生公共课教材.数学系列)
ISBN 7-302-11376-9

Ⅰ.最… Ⅱ.陈… Ⅲ.①最优化理论—研究生—教材 ②最优化算法—研究生—教材 Ⅳ.O242.23

中国版本图书馆 CIP 数据核字(2005)第 077650 号

出 版 者:清华大学出版社
http://www.tup.com.cn
社 总 机:010-62770175
地 址:北京清华大学学研大厦
邮 编:100084
客户服务:010-62776969
组稿编辑:刘 颖
文稿编辑:王海燕
印 刷 者:清华大学印刷厂
装 订 者:三河市新茂装订有限公司
发 行 者:新华书店总店北京发行所
开 本:185×230 **印张**:30 **字数**:636 千字
版 次:2005 年 10 月第 2 版 2005 年 10 月第 1 次印刷
书 号:ISBN 7-302-11376-9/O·478
印 数:1~4000
定 价:38.00 元

第 1 版前言

《最优化理论与算法》是为高等院校开设“线性规划与非线性规划”课程提供的教材.

本书包括凸集凸函数、线性规划和非线性规划三方面的内容.有完整的理论系统,关于凸集凸函数的一些基本定理、线性规划的原理、最优性条件、对偶理论及算法收敛性定理等都做了适度的介绍.书中不仅有大量的实用算法,还介绍了一些最新研究成果,Karmarkar 算法就是一例.为了使具有大学本科程度的读者能够自学,定理的证明和算法的推导主要以数学分析和线性代数为基础,尽可能少地涉及更为高深的知识.本书内容比较丰富,算法比较齐全,有一定的理论深度,层次清晰,叙述简明,便于应用,可作为高等院校教学参考书,也可供应用数学工作者和工程技术人员参考.

本书在编写过程中,得到郑乐宁同志的大力支持,他审阅了原稿,提出了宝贵意见.谭泽光、祁力群和施妙根等同志也给予了满腔热情的帮助.在此表示衷心感谢.

由于作者水平有限,缺点和错误在所难免,敬请批评指教.

作　者

1989 年 8 月

第 2 版前言

本书自 1989 年出版以来，被一些高等学校选作教学参考书，作者本人也在研究生学位课“最优化方法”和“运筹学”的教学中使用了本教材. 经多年教学实践，收到比较满意的效果，总体反映良好，但也发现一些有待改进之处. 为了改进教材的不足，拓宽使用范围，更好地适应教学和自学的需要，作者认真听取关心教材建设的专家和读者的建议，决定再版.

第 2 版教材保持第 1 版的理论体系和写作特点. 增加了基本数学概念介绍、强互补松弛定理、含参数线性规划、运输问题、线性规划路径跟踪法、信赖域方法、二次规划路径跟踪法、整数规划、动态规划等内容. 删除一些原有算法，改写了部分章节. 与第 1 版相比，本版教材算法更加丰富，理论有所深入，在一定程度上反映出近些年运筹学一些分支的新进展.

本书由预备知识、线性规划、非线性规划、整数规划和动态规划等五部分组成. 使用本教材时，可根据需要决定取舍. 一般来讲，要求较多的专业，可用 64 学时讲授去掉带 * 号章节后的全部内容；要求较少的专业，可用 32 学时讲授线性规划和动态规划部分；标有 * 号的章节可酌情选用.

责任编辑刘颖为本书付出了辛勤劳动，部分插图是清华大学建筑设计研究院陈若光所绘，在此向两位年轻专家表示衷心感谢.

作　者

2005 年 5 月

目　录

第 1 章　引言 …… 1
1.1　学科简述 …… 1
1.2　线性与非线性规划问题 …… 2
*1.3　几个数学概念 …… 5
1.4　凸集和凸函数 …… 10
习题 …… 23

第 2 章　线性规划的基本性质 …… 26
2.1　标准形式及图解法 …… 26
2.2　基本性质 …… 28
习题 …… 35

第 3 章　单纯形方法 …… 37
3.1　单纯形方法原理 …… 37
3.2　两阶段法与大 M 法 …… 50
3.3　退化情形 …… 66
3.4　修正单纯形法 …… 74
*3.5　变量有界的情形 …… 85
*3.6　分解算法 …… 94
习题 …… 118

第 4 章　对偶原理及灵敏度分析 …… 122
4.1　线性规划中的对偶理论 …… 122
4.2　对偶单纯形法 …… 133
4.3　原始-对偶算法 …… 143
4.4　灵敏度分析 …… 149
*4.5　含参数线性规划 …… 157
习题 …… 163

第 5 章　运输问题 …… 167
5.1　运输问题的数学模型与基本性质 …… 167
5.2　表上作业法 …… 170
5.3　产销不平衡运输问题 …… 177
习题 …… 178

第 6 章　线性规划的内点算法 …… 180
* 6.1　Karmarkar 算法 …… 180
* 6.2　内点法 …… 193
6.3　路径跟踪法 …… 196

第 7 章　最优性条件 …… 203
7.1　无约束问题的极值条件 …… 203
7.2　约束极值问题的最优性条件 …… 206
* 7.3　对偶及鞍点问题 …… 232
习题 …… 243

* **第 8 章　算法** …… 246
8.1　算法概念 …… 246
8.2　算法收敛问题 …… 250
习题 …… 253

第 9 章　一维搜索 …… 254
9.1　一维搜索概念 …… 254
9.2　试探法 …… 256
9.3　函数逼近法 …… 265
习题 …… 280

第 10 章　使用导数的最优化方法 …… 281
10.1　最速下降法 …… 281
10.2　牛顿法 …… 287
10.3　共轭梯度法 …… 291
10.4　拟牛顿法 …… 306
10.5　信赖域方法 …… 315

10.6　最小二乘法 …… 322
习题 …… 328

第 11 章　无约束最优化的直接方法 …… 332
11.1　模式搜索法 …… 332
11.2　Rosenbrock 方法 …… 337
11.3　单纯形搜索法 …… 343
11.4　Powell 方法 …… 349
习题 …… 358

第 12 章　可行方向法 …… 360
12.1　Zoutendijk 可行方向法 …… 360
12.2　Rosen 梯度投影法 …… 371
*12.3　既约梯度法 …… 379
12.4　Frank-Wolfe 方法 …… 388
习题 …… 392

第 13 章　惩罚函数法 …… 394
13.1　外点罚函数法 …… 394
13.2　内点罚函数法 …… 401
*13.3　乘子法 …… 405
习题 …… 413

第 14 章　二次规划 …… 415
14.1　Lagrange 方法 …… 415
14.2　起作用集方法 …… 417
14.3　Lemke 方法 …… 422
14.4　路径跟踪法 …… 426
习题 …… 431

*第 15 章　整数规划简介 …… 432
15.1　分支定界法 …… 432
15.2　割平面法 …… 436
15.3　0-1 规划的隐数法 …… 439

15.4 指派问题…………………………………………………………………… 444
习题………………………………………………………………………………… 450

第 16 章 动态规划简介 …………………………………………………… 452
16.1 动态规划的一些基本概念…………………………………………… 452
16.2 动态规划的基本定理和基本方程…………………………………… 454
16.3 逆推解法和顺推解法………………………………………………… 456
16.4 动态规划与静态规划的关系………………………………………… 459
16.5 函数迭代法…………………………………………………………… 463
习题………………………………………………………………………………… 466

参考文献………………………………………………………………………… 467

第1章 引　言

1.1 学科简述

最优化理论与算法是一个重要的数学分支,它所研究的问题是讨论在众多的方案中什么样的方案最优以及怎样找出最优方案.这类问题普遍存在.例如,工程设计中怎样选择设计参数,使得设计方案既满足设计要求又能降低成本;资源分配中,怎样分配有限资源,使得分配方案既能满足各方面的基本要求,又能获得好的经济效益;生产计划安排中,选择怎样的计划方案才能提高产值和利润;原料配比问题中,怎样确定各种成分的比例,才能提高质量,降低成本;城建规划中,怎样安排工厂、机关、学校、商店、医院、住户和其他单位的合理布局,才能方便群众,有利于城市各行各业的发展;农田规划中,怎样安排各种农作物的合理布局,才能保持高产稳产,发挥地区优势;军事指挥中,怎样确定最佳作战方案,才能有效地消灭敌人,保存自己,有利于战争的全局;在人类活动的各个领域中,诸如此类,不胜枚举.最优化这一数学分支,正是为这些问题的解决,提供理论基础和求解方法,它是一门应用广泛、实用性强的学科.

最优化是个古老的课题.长期以来,人们对最优化问题进行着探讨和研究.早在17世纪,英国科学家 Newton 发明微积分的时代,就已提出极值问题,后来又出现 Lagrange 乘数法.1847年法国数学家 Cauchy 研究了函数值沿什么方向下降最快的问题,提出最速下降法.1939年前苏联数学家 Л. В. Канторович 提出了解决下料问题和运输问题这两种线性规划问题的求解方法.人们关于最优化问题的研究工作,随着历史的发展不断深入.但是,任何科学的进步,都受到历史条件的限制,直至20世纪30年代,最优化这个古老课题并未形成独立的有系统的学科.

20世纪40年代以来,由于生产和科学研究突飞猛进地发展,特别是电子计算机日益广泛应用,使最优化问题的研究不仅成为一种迫切需要,而且有了求解的有力工具.因此最优化理论和算法迅速发展起来,形成一个新的学科.至今已出现线性规划、整数规划、非线性规划、几何规划、动态规划、随机规划、网络流等许多分支.最优化理论和算法在实际应用中正在发挥越来越大的作用.

1.2 线性与非线性规划问题

线性与非线性规划有着广泛的实际背景，许多实际问题抽象成数学模型后，可归结为求解这类问题，本书重点介绍线性与非线性规划. 下面先来研究几个例题.

例 1.2.1 生产计划问题

设某工厂用 4 种资源生产 3 种产品，每单位第 j 种产品需要第 i 种资源的数量为 a_{ij}，可获利润为 c_j，第 i 种资源总消耗量不能超过 b_i，由于市场限制，第 j 种产品的产量不超过 d_j，试问如何安排生产才能使总利润最大?

解析 下面分析怎样建立数学模型. 设 3 种产品的产量分别为 x_1, x_2, x_3，这是决策变量，目标函数是总利润 $c_1x_1+c_2x_2+c_3x_3$，约束条件有资源限制 $a_{i1}x_1+a_{i2}x_2+a_{i3}x_3 \leqslant b_i$ $(i=1,2,3,4)$，市场销量限制，$x_j \leqslant d_j (j=1,2,3)$，及产量非负限制 $x_j \geqslant 0 (j=1,2,3)$. 问题概括为，在一组约束条件下，确定一个最优生产方案 $\boldsymbol{x}^*=(x_1^*, x_2^*, x_3^*)$，使目标函数值最大. 数学模型如下：

$$
\begin{aligned}
\max \quad & \sum_{j=1}^{3} c_j x_j \\
\text{s.t.} \quad & \sum_{j=1}^{3} a_{ij} x_j \leqslant b_i, \quad && i=1,\cdots,4, \\
& x_j \leqslant d_j, && j=1,2,3, \\
& x_j \geqslant 0, && j=1,2,3,
\end{aligned}
$$

其中 max 表示 maximize，读作“极大化”，s. t. 表示 subject to，读作“约束条件是”.

例 1.2.2 食谱问题

设市场上可买到 n 种不同的食品，第 j 种食品单位售价为 c_j. 每种食品含有 m 种基本营养成分，第 j 种食品每一个单位含第 i 种营养成分为 a_{ij}. 又设每人每天对第 i 种营养成分的需要量不少于 b_i. 试确定在保证营养要求条件下的最经济食谱.

解析 建立食谱问题的数学模型. 设每人每天需要各种食品的数量分别为 $x_1, x_2, \cdots, x_n$. 我们的目标是使伙食费用最少，即使 $c_1x_1+c_2x_2+\cdots+c_nx_n$ 最小. 条件是保证用餐者对各种营养成分的基本需要，即满足 $a_{i1}x_1+a_{i2}x_2+\cdots+a_{in}x_n \geqslant b_i (i=1,2,\cdots,m)$. 数学模型是

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

其中 min 表示 minimize,读作“极小化”.

例 1.2.3 结构设计问题

以两个构件组成的对称桁架为例(参见图 1.2.1).

已知桁架的跨度 $2L$,高度 x_2 的上限 H,承受负荷 $2P$,钢管的厚度 T,材料比重 ρ,纵向弹性模量 E 及容许应力 σ_y. 试确定钢管的平均直径 x_1 及桁架的高度 x_2,使桁架的重量最小.

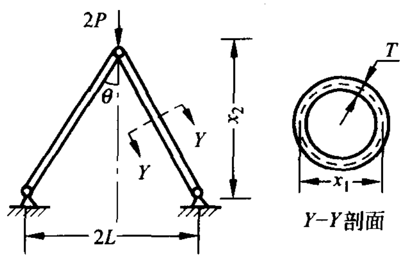

图 1.2.1

解析 桁架的重量

$$G = 2\pi\rho T x_1 (L^2 + x_2^2)^{\frac{1}{2}},$$

它是平均直径 x_1 和高度 x_2 的函数. x_1 和 x_2 的选择不是任意的,必须满足以下几个条件:

(1) 由于空间限制,要求 x_2 不能超过高度上限 H,即

$$x_2 \leqslant H.$$

(2) 钢管上的压应力不能超过材料的容许应力 σ_y. 在负荷 $2P$ 作用下,钢管承受的压力为

$$F = \frac{P}{\cos\theta} = \frac{P(L^2 + x_2^2)^{\frac{1}{2}}}{x_2},$$

钢管的横截面面积

$$S \approx \pi T x_1,$$

由此可知,钢管上的压应力为

$$\sigma(x_1, x_2) = \frac{P(L^2 + x_2^2)^{\frac{1}{2}}}{\pi T x_1 x_2},$$

因此要求

$$\frac{P(L^2 + x_2^2)^{\frac{1}{2}}}{\pi T x_1 x_2} \leqslant \sigma_y.$$

(3) 参数的选择还必须保证在负荷 $2P$ 的作用下钢管不发生弯曲,这就要求压应力不超过临界应力 σ_l. 临界应力可由 Euler 公式算出:

$$\sigma_l = \frac{\pi^2 E(x_1^2 + T^2)}{8(L^2 + x_2^2)},$$

其中 E 是已知的弹性模量. 按此要求应有

$$\frac{P(L^2 + x_2^2)^{\frac{1}{2}}}{\pi T x_1 x_2} \leqslant \frac{\pi^2 E(x_1^2 + T^2)}{8(L^2 + x_2^2)}.$$

根据以上分析,桁架的最优设计问题,就是求重量函数 G 在上述 3 个约束条件下的极小

点问题.它的数学模型是

$$
\begin{aligned}
\min \quad & 2\pi\rho T x_1 (L^2 + x_2^2)^{\frac{1}{2}} \\
\text{s.t.} \quad & x_2 \leqslant H, \\
& \frac{P(L^2 + x_2^2)^{\frac{1}{2}}}{\pi T x_1 x_2} \leqslant \sigma_y, \\
& \frac{P(L^2 + x_2^2)^{\frac{1}{2}}}{\pi T x_1 x_2} \leqslant \frac{\pi^2 E(x_1^2 + T^2)}{8(L^2 + x_2^2)}, \\
& x_1, x_2 \geqslant 0.
\end{aligned}
$$

例 1.2.4 选址问题

设有 n 个市场,第 j 个市场的位置为(a_j,b_j),对某种货物的需要量为 $q_j(j=1,\cdots,n)$.现计划建立 m 个货栈,第 i 个货栈的容量为 $c_i(i=1,\cdots,m)$.试确定货栈的位置,使各货栈到各市场的运输量与路程乘积之和最小.

解析 现在来建立数学模型.设第 i 个货栈的位置为$(x_i,y_i)(i=1,\cdots,m)$.第 i 个货栈供给第 j 个市场的货物量为 $W_{ij}(i=1,\cdots,m;j=1,\cdots,n)$.第 i 个货栈到第 j 个市场的距离为 d_{ij},一般定义为

$$
d_{ij} = \sqrt{(x_i - a_j)^2 + (y_i - b_j)^2} \tag{1.2.1}
$$

或

$$
d_{ij} = |x_i - a_j| + |y_i - b_j|, \tag{1.2.2}
$$

我们的目标是使运输量与路程乘积之和最小,如果距离按(1.2.1)式定义,就是使

$$
\sum_{i=1}^{m}\sum_{j=1}^{n} W_{ij} \sqrt{(x_i - a_j)^2 + (y_i - b_j)^2}
$$

最小.约束条件是:

(1) 每个货栈向各市场提供的货物量之和不能超过它的容量;

(2) 每个市场从各货栈得到的货物量之和应等于它的需要量;

(3) 运输量不能为负数.

因此,问题的数学模型如下:

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{m}\sum_{j=1}^{n} W_{ij}\sqrt{(x_i - a_j)^2 + (y_i - b_j)^2} \\
\text{s.t.} \quad & \sum_{j=1}^{n} W_{ij} \leqslant c_i, \quad i = 1,\cdots,m, \\
& \sum_{i=1}^{m} W_{ij} = q_j, \quad j = 1,\cdots,n, \\
& W_{ij} \geqslant 0, \quad i = 1,\cdots,m, j = 1,\cdots,n.
\end{aligned}
$$

在上述例 1.2.1 和例 1.2.2 的数学模型中，目标函数和约束函数都是线性的，称之为**线性规划问题**；而例 1.2.3 和例 1.2.4 的数学模型中含有非线性函数，因此称为**非线性规划问题**.

在线性规划与非线性规划中，满足约束条件的点称为**可行点**，全体可行点组成的集合称为**可行集**或**可行域**. 如果一个问题的可行集是整个空间. 那么此问题就称为**无约束问题**.

下面给出最优解概念.

定义 1.2.1 设 $f(\boldsymbol{x})$为目标函数，S 为可行域，$\bar{\boldsymbol{x}} \in S$，若对每个 $\boldsymbol{x}\in S$，成立 $f(\boldsymbol{x})\geqslant f(\bar{\boldsymbol{x}})$，则称 $\bar{\boldsymbol{x}}$ 为 $f(\boldsymbol{x})$在 S 上的全局极小点.

定义 1.2.2 设 $f(\boldsymbol{x})$为目标函数，S 为可行域，若存在 $\bar{\boldsymbol{x}}\in S$ 的 $\varepsilon>0$ 邻域 $N(\bar{\boldsymbol{x}},\varepsilon)=\{\boldsymbol{x}\mid \|\boldsymbol{x}-\bar{\boldsymbol{x}}\|<\varepsilon\}$，使得对每个 $\boldsymbol{x}\in S\cap N(\bar{\boldsymbol{x}},\varepsilon)$成立 $f(\boldsymbol{x})\geqslant f(\bar{\boldsymbol{x}})$，则称 $\bar{\boldsymbol{x}}$ 为 $f(\boldsymbol{x})$在 S 上的一个局部极小点.

对于极大化问题，可类似地定义全局极大点和局部极大点，这里不再叙述.

根据上述定义，全局极小点也是局部极小点，而局部极小点不一定是全局极小点. 但是对于某些特殊情形，如将在后面介绍的凸规划，局部极小点也是全局极小点.

*1.3 几个数学概念

1.3.1 向量范数和矩阵范数

定义 1.3.1 若实值函数 $\|\cdot\|:\mathbb{R}^n\to\mathbb{R}$ 满足下列条件：

(1) $\|\boldsymbol{x}\|\geqslant 0,\ \forall\, \boldsymbol{x}\in\mathbb{R}^n$；$\|\boldsymbol{x}\|=0$ 当且仅当 $\boldsymbol{x}=\boldsymbol{0}$；

(2) $\|\alpha\boldsymbol{x}\|=|\alpha|\,\|\boldsymbol{x}\|,\ \forall\,\alpha\in\mathbb{R},\ \boldsymbol{x}\in\mathbb{R}^n$；

(3) $\|\boldsymbol{x}+\boldsymbol{y}\|\leqslant\|\boldsymbol{x}\|+\|\boldsymbol{y}\|,\ \forall\,\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$.

则称 $\|\cdot\|$ 为向量范数. 其中$\mathbb{R}^n$表示 n 维向量空间.

设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in\mathbb{R}^n$，常用的向量范数有 L_1 范数，L_2 范数和 L_∞ 范数，分别为

$$\|\boldsymbol{x}\|_1=\sum_{j=1}^{n}|x_j|,$$

$$\|\boldsymbol{x}\|_2=\left(\sum_{j=1}^{n}x_j^2\right)^{\frac{1}{2}},$$

$$\|\boldsymbol{x}\|_\infty=\max_j|x_j|.$$

一般地，对于 $1\leqslant p<\infty$，L_p 范数为

$$\|\boldsymbol{x}\|_p=\left(\sum_{j=1}^{n}|x_j|^p\right)^{\frac{1}{p}}.$$

关于范数的等价性，有下列定义.

定义 1.3.2 设 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 是 $\mathbb{R}^n$ 上任意两个范数，如果存在正数 c_1 和 c_2，使得对每个 $x\in\mathbb{R}^n$ 成立 $c_1\|x\|_\alpha\leqslant\|x\|_\beta\leqslant c_2\|x\|_\alpha$，则称范数 $\|x\|_\alpha$ 和范数 $\|x\|_\beta$ 等价.

在 $\mathbb{R}^n$ 中任何两种范数均等价.

这里应指出，上述向量范数中，$\|x\|_2$ 称为 Euclid 范数，如无特殊指明，后面将用 $\mathbb{R}^n$ 表示 n 维 Euclid 空间.

关于矩阵范数，定义如下.

定义 1.3.3 设 $\boldsymbol{A}$ 为 $m\times n$ 矩阵，$\|\cdot\|_\alpha$ 是 $\mathbb{R}^m$ 上向量范数，$\|\cdot\|_\beta$ 是 $\mathbb{R}^n$ 上向量范数，定义矩阵范数 $\|\boldsymbol{A}\|=\max\limits_{\|x\|_\beta=1}\|\boldsymbol{A}x\|_\alpha$.

根据矩阵范数定义，对于单位矩阵 $\boldsymbol{I}$，总有 $\|\boldsymbol{I}\|=1$. 关于矩阵范数有下列结论.

定理 1.3.1 矩阵范数具有下列性质：

(1) $\|\boldsymbol{A}x\|_\alpha\leqslant\|\boldsymbol{A}\|\,\|x\|_\beta$；

(2) $\|\lambda\boldsymbol{A}\|=|\lambda|\,\|\boldsymbol{A}\|$；

(3) $\|\boldsymbol{A}+\boldsymbol{B}\|\leqslant\|\boldsymbol{A}\|+\|\boldsymbol{B}\|$；

(4) $\|\boldsymbol{AD}\|\leqslant\|\boldsymbol{A}\|\,\|\boldsymbol{D}\|$.

其中 $\boldsymbol{A},\boldsymbol{B}$ 为 $m\times n$ 矩阵，$\boldsymbol{D}$ 是 $n\times p$ 矩阵，λ 为实数，$x\in\mathbb{R}^n$.

设矩阵 $\boldsymbol{A}=(a_{ij})_{m\times n}$，下面给出 3 种常用的矩阵范数，分别记作 $\|\boldsymbol{A}\|_1$，$\|\boldsymbol{A}\|_2$，$\|\boldsymbol{A}\|_\infty$：

(1) $\|\boldsymbol{A}\|_1=\max\limits_j\sum\limits_{i=1}^m|a_{ij}|$；

(2) $\|\boldsymbol{A}\|_2=\sqrt{\lambda_{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}}}$；

(3) $\|\boldsymbol{A}\|_\infty=\max\limits_i\sum\limits_{j=1}^n|a_{ij}|$.

其中 $\lambda_{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}}$ 表示 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的最大特征值. $\|\boldsymbol{A}\|_2$ 称为 $\boldsymbol{A}$ 的谱范数.

1.3.2 序列的极限

定义 1.3.4 设 $\{x^{(k)}\}$ 是 $\mathbb{R}^n$ 中一个向量序列，$\bar{x}\in\mathbb{R}^n$，如果对每个任给的 $\varepsilon>0$ 存在正整数 K_ε，使得当 $k>K_\varepsilon$ 时就有 $\|x^{(k)}-\bar{x}\|<\varepsilon$，则称序列收敛到 $\bar{x}$，或称序列以 $\bar{x}$ 为极限，记作 $\lim\limits_{k\to\infty}x^{(k)}=\bar{x}$.

按此定义，序列若存在极限，则任何子序列有相同的极限，即序列的极限是惟一的.

定义 1.3.5 设 $\{x^{(k)}\}$ 是 $\mathbb{R}^n$ 中一个向量序列，如果存在一个子序列 $\{x^{(k_j)}\}$，使 $\lim\limits_{k_j\to\infty}x^{(k_j)}=\hat{x}$，则称 $\hat{x}$ 是序列 $\{x^{(k)}\}$ 的一个聚点.

根据定义易知，如果无穷序列有界，即存在正数 M，使得对所有 k 均有 $\|x^{(k)}\|\leqslant M$，

则这个序列必有聚点.

定义 1.3.6 设$\{\boldsymbol{x}^{(k)}\}$是$\mathbb{R}^n$中一个向量序列,如果对任意给定的$\varepsilon>0$,总存在正整数K_ε,使得当$m,l>K_\varepsilon$时,就有$\|\boldsymbol{x}^{(m)}-\boldsymbol{x}^{(l)}\|<\varepsilon$,则$\{\boldsymbol{x}^{(k)}\}$称为 Cauchy 序列.

在$\mathbb{R}^n$中,Cauchy 序列有极限.

定理 1.3.2 设$\{\boldsymbol{x}^{(j)}\}\subset\mathbb{R}^n$为 Cauchy 序列,则$\{\boldsymbol{x}^{(j)}\}$的聚点必为极限点.(证明从略)

后面算法介绍和理论分析中,还常涉及闭集、开集、紧集等概念.定义如下:设S为$\mathbb{R}^n$中一个集合,如果S中每个收敛序列的极限均属于S,则称S为闭集.如果对每一点$\hat{\boldsymbol{x}}\in S$,存在正数$\varepsilon$,使得$\hat{\boldsymbol{x}}$的$\varepsilon$邻域$N(\hat{\boldsymbol{x}},\varepsilon)=\left\{\boldsymbol{x}\mid\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|<\varepsilon\right\}\subset S$,则称$S$为开集.如果$S$是有界闭集,则称$S$为紧集.

1.3.3 梯度、Hesse 矩阵、Taylor 展开式

设集合$S\subset\mathbb{R}^n$非空,$f(\boldsymbol{x})$为定义在S上的实函数.如果f在每一点$\boldsymbol{x}\in S$连续,则称f在S上连续,记作$f\in C(S)$.再设S为开集,如果在每一点$\boldsymbol{x}\in S$,对所有$j=1,\cdots,n$,偏导数$\dfrac{\partial f(\boldsymbol{x})}{\partial x_j}$存在且连续,则称$f$在开集$S$上连续可微,记作$f\in C^1(S)$.如果在每一点$\boldsymbol{x}\in S$,对所有$i=1,\cdots,n$和$j=1,\cdots,n$,二阶偏导数$\dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_i\partial x_j}$存在且连续,则称$f$在开集$S$上二次连续可微,记作$f\in C^2(S)$.

函数f在$\boldsymbol{x}$处的梯度为n维列向量:

$$\nabla f(\boldsymbol{x})=\left[\frac{\partial f(\boldsymbol{x})}{\partial x_1},\frac{\partial f(\boldsymbol{x})}{\partial x_2},\cdots,\frac{\partial f(\boldsymbol{x})}{\partial x_n}\right]^{\mathrm{T}}. \tag{1.3.1}$$

f在$\boldsymbol{x}$处的 Hesse 矩阵为$n\times n$矩阵$\nabla^2 f(\boldsymbol{x})$,第$i$行第$j$列元素为

$$[\nabla^2 f(\boldsymbol{x})]_{ij}=\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i\partial x_j},\qquad 1\leqslant i,j\leqslant n. \tag{1.3.2}$$

当$f(\boldsymbol{x})$为二次函数时,梯度及 Hesse 矩阵很容易求得.二次函数可以写成下列形式:

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

其中$\boldsymbol{A}$是n阶对称矩阵,$\boldsymbol{b}$是n维列向量,c是常数.函数$f(\boldsymbol{x})$在$\boldsymbol{x}$处的梯度$\nabla f(\boldsymbol{x})=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}$,Hesse 矩阵$\nabla^2 f(\boldsymbol{x})=\boldsymbol{A}$.

假设在开集$S\subset\mathbb{R}^n$上$f\in C^1(S)$,给定点$\bar{\boldsymbol{x}}\in S$,则$f$在点$\bar{\boldsymbol{x}}$的一阶 Taylor 展开式为

$$f(\boldsymbol{x})=f(\bar{\boldsymbol{x}})+\nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}})+o(\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|),$$

其中$o(\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|)$当$\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|\to 0$时,关于$\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|$是高阶无穷小量.

假设在开集$S\subset\mathbb{R}^n$上$f\in C^2(S)$,则f在$\bar{\boldsymbol{x}}\in S$的二阶 Taylor 展开式为

$$f(\boldsymbol{x})=f(\bar{\boldsymbol{x}})+\nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{x}})+\frac{1}{2}(\boldsymbol{x}-\bar{\boldsymbol{x}})^{\mathrm{T}}\nabla^2 f(\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}})+o(\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|^2),$$

其中 $o(\|x-\bar{x}\|^2)$ 当 $\|x-\bar{x}\|^2\to 0$ 时，关于 $\|x-\bar{x}\|^2$ 是高阶无穷小量.

1.3.4 Jacobi 矩阵、链式法则和隐函数存在定理

1. Jacobi 矩阵

考虑向量值函数

$$h(\boldsymbol{x}) = (h_1(\boldsymbol{x}),h_2(\boldsymbol{x}),\cdots,h_m(\boldsymbol{x}))^{\mathrm{T}},$$

其中每个分量 $h_i(\boldsymbol{x})$ 为 n 元实值函数，假设对所有 i,j 偏导数 $\dfrac{\partial h_i(\boldsymbol{x})}{\partial x_j}$ 存在. $\boldsymbol{h}$ 在点 $\boldsymbol{x}$ 的 Jacobi 矩阵为

$$\begin{bmatrix} \dfrac{\partial h_1(\boldsymbol{x})}{\partial x_1} & \dfrac{\partial h_1(\boldsymbol{x})}{\partial x_2} & \cdots & \dfrac{\partial h_1(\boldsymbol{x})}{\partial x_n} \\ \dfrac{\partial h_2(\boldsymbol{x})}{\partial x_1} & \dfrac{\partial h_2(\boldsymbol{x})}{\partial x_2} & \cdots & \dfrac{\partial h_2(\boldsymbol{x})}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial h_m(\boldsymbol{x})}{\partial x_1} & \dfrac{\partial h_m(\boldsymbol{x})}{\partial x_2} & \cdots & \dfrac{\partial h_m(\boldsymbol{x})}{\partial x_n} \end{bmatrix}, \tag{1.3.3}$$

这个矩阵称为向量值函数 $\boldsymbol{h}$ 在 $\boldsymbol{x}$ 的导数，记作 $\boldsymbol{h}'(\boldsymbol{x})$ 或 $\nabla \boldsymbol{h}(\boldsymbol{x})^{\mathrm{T}}$，其中 $\nabla \boldsymbol{h}(\boldsymbol{x})=(\nabla h_1(\boldsymbol{x}), \nabla h_2(\boldsymbol{x}),\cdots,\nabla h_m(\boldsymbol{x}))$.

例 1.3.1 设有向量值函数

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(x_1,x_2) = \begin{bmatrix} \sin x_1 + \cos x_2 \\ e^{2x_1+x_2} \\ 2x_1^2 + x_1 x_2 \end{bmatrix},$$

则 $f(\boldsymbol{x})$ 在任一点 (x_1,x_2) 的 Jacobi 矩阵，即导数为

$$\boldsymbol{f}'(\boldsymbol{x}) = \begin{bmatrix} \cos x_1 & -\sin x_2 \\ 2e^{2x_1+x_2} & e^{2x_1+x_2} \\ 4x_1+x_2 & x_1 \end{bmatrix}.$$

2. 链式法则

设有复合函数 $\boldsymbol{h}(\boldsymbol{x})=\boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x}))$，其中向量值函数 $\boldsymbol{f}(\boldsymbol{g})$ 和 $\boldsymbol{g}(\boldsymbol{x})$ 均可微，$\boldsymbol{x}\in D^n\subset\mathbb{R}^n$，$\boldsymbol{g}: D^n\to D_1^m$，$\boldsymbol{f}: D_2^m\to\mathbb{R}^k$，其中 $D_1^m\subset D_2^m$，$\boldsymbol{h}: D^n\to\mathbb{R}^k$. 根据复合函数求导数的链式法则，必有

$$\boldsymbol{h}'(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{g}(\boldsymbol{x}))\boldsymbol{g}'(\boldsymbol{x}), \qquad \boldsymbol{x}\in D^n, \tag{1.3.4}$$

其中 $\boldsymbol{f}'$ 和 $\boldsymbol{g}'$ 分别为 $k\times m$ 和 $m\times n$ 矩阵，$\boldsymbol{h}'$ 为 $k\times n$ 矩阵. 若记 $\nabla \boldsymbol{f}=(\nabla \boldsymbol{f}_1,\nabla \boldsymbol{f}_2,\cdots,\nabla \boldsymbol{f}_k)$，$\nabla \boldsymbol{g}=(\nabla \boldsymbol{g}_1,\nabla \boldsymbol{g}_2,\cdots,\nabla \boldsymbol{g}_m)$，由于 $\boldsymbol{h}'=\nabla \boldsymbol{h}^{\mathrm{T}}$，$\boldsymbol{f}'=\nabla \boldsymbol{f}^{\mathrm{T}}$ 和 $\boldsymbol{g}'=\nabla \boldsymbol{g}^{\mathrm{T}}$，可将(1.3.4)式改写为