高等院校精品课程教材

最优化基础理论与方法

王燕军　梁治安 编著

博学·数学系列

復旦大學出版社
www.fudanpress.com.cn

高等院校精品课程教材

最优化基础理论与方法

王燕军　梁治安　编著

博学·数学系列

复旦大学出版社
www.fudanpress.com.cn

图书在版编目(CIP)数据

最优化基础理论与方法/王燕军,梁治安编著. —上海:复旦大学出版社,2011.9
(复旦博学·数学系列)
ISBN 978-7-309-08373-6

Ⅰ. 最…　Ⅱ. ①王…②梁…　Ⅲ. 最佳化-数学理论　Ⅳ. 0224

中国版本图书馆 CIP 数据核字(2011)第 166335 号

最优化基础理论与方法
王燕军　梁治安　编著
责任编辑/范仁梅

复旦大学出版社有限公司出版发行
上海市国权路 579 号　邮编:200433
网址:fupnet@fudanpress.com　http://www.fudanpress.com
门市零售:86-21-65642857　团体订购:86-21-65118853
外埠邮购:86-21-65109143
上海崇明南海印刷厂

开本 787×960　1/16　印张 9.25　字数 163 千
2011 年 9 月第 1 版第 1 次印刷

ISBN 978-7-309-08373-6/O·477
定价: 21.00 元

“博学而笃志，切问而近思。”

（《论语》）

博晓古今，可立一家之说；

学贯中西，或成经国之才。

内容提要

本书对非线性最优化的理论、算法及相关技术做了比较系统的介绍.在内容的选取方面，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.另外，也尽可能地增加一些数值例子或经济管理方面的应用实例.全书共分7章.第一章主要介绍最优化的基础理论；第二章介绍无约束最优化问题的最优性条件以及线性搜索技术；第三章主要介绍无约束最优化算法，主要有最速下降法、Newton法、共轭梯度法；第四章主要讨论约束优化问题的最优性条件；第五章介绍二次规划的求解算法；第六章介绍一般非线性约束最优化问题的罚函数法；第七章给出两种特殊规划：几何规划和多目标规划，并给出一些应用实例.

本书可作为高等院校计算数学、应用数学、工程、经济、金融等各专业的教材，也可供有关工程技术人员和研究人员参考.

前　言

最优化问题是在有限种或无限种可行方案(决策)中选择最优的方案(决策),它是数学的一个重要的分支.自 20 世纪 50 年代后,随着实际生产的需要和计算机的发展,逐渐形成了最优化理论,以及相应的求解方法——最优化方法.目前这些方法已经在工农业、国防、交通、经济、金融、能源、通信等众多领域中得到了广泛的应用.最优化技术已经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基础理论和方法作一个比较全面的介绍.作为财经类院校的教材,在内容的选取方面,尽可能避免过分复杂的理论分析,以适应不同专业、不同层次的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,而略去烦琐的证明;但对于从事理论研究与进行方法设计的读者来说,这些理论分析是相当重要而又富于启发性的.另外,在本书中也尽力增加了一些应用实例.

全书共分 7 章.第一章主要介绍最优化相关的基础理论;第二章介绍无约束优化问题的最优性条件以及线性搜索技术;第三章主要介绍无约束最优化算法,主要有最速下降法、Newton 法、共轭梯度法;第四章主要讨论约束优化问题的最优性条件;第五章介绍了二次规划的求解算法;第六章介绍了一般非线性约束最优化问题的罚函数法;第七章给出了两种特殊规划:几何规划和多目标规划,并给出了一些应用实例.

本书可作为计算数学、应用数学、工程、经济、金融等各专业高年级本科生或研究生的教学用书或辅导用书,也可供从事优化方面的科技工作者和工程技术人员学习和参考.

限于作者水平,可能在内容的取材、结构的编排以及课程的讲法上存在不妥之处,敬请广大读者对书中的不妥与错误之处批评指正!

作者

2011 年 3 月

目　　录

第一章　最优化基础

在实际生活和许多学科中，我们总是会遇到最优化问题. 读者在学习初等数学时已经开始解一些最优化问题，如在周长一定的矩形中，正方形的面积最大. 用初等数学的方法即可证明这个优化问题，或者在学了高等数学后，也可以用 Lagrange 乘子的方法证明这个问题. 本章将介绍最优化的一般模型和分类，并介绍最优化理论和方法涉及的一些微积分和代数的基础知识.

§1.1　最优化问题的分类与应用实例

最优化问题数学模型的一般形式为

$$\begin{cases} \min f(\boldsymbol{x}); \\ \text{s.t.} \quad c_i(\boldsymbol{x}) = 0,\ i \in E = \{1,\ 2,\ \cdots,\ l\}, \\ \qquad\ \ c_i(\boldsymbol{x}) \leqslant 0,\ i \in I = \{l+1,\ l+2,\ \cdots,\ l+m\}, \\ \qquad\ \ \boldsymbol{x} \in \mathbf{R}^n. \end{cases} \tag{1.1.1}$$

这里 min 表示求极小值；s. t. 是 subject to 的意思，表示受限制于…；$\boldsymbol{x}$ 是 n 维向量，其分量是 x_1，…，x_n. 在问题(1.1.1)中称 $f(\boldsymbol{x})$ 为目标函数(objective function)，称 $c_i(\boldsymbol{x})$ $(i \in E \cup I)$ 为约束函数(constraint functions). 若求极大值，可以将目标函数写成 $\min(-f(\boldsymbol{x}))$，不等式约束 $c_i(\boldsymbol{x}) \geqslant 0$ 可以写成 $-c_i(\boldsymbol{x}) \leqslant 0$.

若记集合

$$D = \{\boldsymbol{x} \mid c_i(\boldsymbol{x}) = 0,\ i \in E;\ c_i(\boldsymbol{x}) \leqslant 0,\ i \in I,\ \boldsymbol{x} \in \mathbf{R}^n\},$$

则称 D 为问题(1.1.1)的可行域，若 $\boldsymbol{x} \in D$，则称 $\boldsymbol{x}$ 为可行点.

下面我们对最优化问题(1.1.1)进行简单的分类.

(1) 根据可行域 D 划分：若 $D = \mathbf{R}^n$，即 $\boldsymbol{x}$ 是自由变量，则称问题(1.1.1)为无约束优化问题；否则，$D \subset \mathbf{R}^n$，称问题(1.1.1)为约束优化问题.

(2) 根据函数的性质划分：若目标函数和约束函数都是线性的，则称问题

(1.1.1)为线性规划;若目标函数和约束函数中至少有一个是非线性的,则称问题(1.1.1)为非线性规划.进一步,若目标函数是二次函数,$c_i(\boldsymbol{x})$ $(i=1,\cdots,l+m)$ 是线性函数,则称问题(1.1.1)为二次规划.

(3) 根据可行域的性质划分:若可行域内点的个数是有限的,则称问题(1.1.1)为离散最优化问题;若可行域含有无穷多个点,且可行域内的点可以连续变化,则称(1.1.1)为连续最优化问题.对于离散优化问题,若变量均为整数,则称其为整数规划问题;若部分变量为整数,而另一部分变量连续变化,则称其为混合整数规划问题.

(4) 根据函数的向量性质划分:若目标函数为向量函数,则称问题(1.1.1)为多目标规划问题;若目标函数为数量函数,则称问题(1.1.1)为单目标规划问题.

(5) 根据规划问题有关信息的确定性划分:若目标函数或约束函数具有随机性,也就是问题的表述形式随时间的变化而变化,具有不确定性,则这样的优化问题称为随机规划;如果优化问题的变量(函数)具有模糊性,则这样的优化问题称为模糊规划;如果目标函数和可行域都是确定的,则这样的规划问题称为确定性规划问题.本书讨论确定性规划问题.

为了进一步说明最优化问题,下面我们举一些优化问题实例.

例 1.1.1 组合投资问题.

假设有一笔资金 a 亿元,准备投资于 n 种证券.已知第 i 种证券的期望收益率为 r_i,证券收益率之间的协方差矩阵为 A.假设投资到各种证券的资金向量为 $\boldsymbol{x}=\{x_1,x_2,\cdots,x_n\}^{\mathrm{T}}$ (称为投资组合 Portfolio),且期望收益不低于事先指定的数 r_0,如何根据 Markowitz 的投资组合理论,建立最优化模型.

解 这项投资的风险是 $V(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}$,期望收益是 $R(\boldsymbol{x})=r_1x_1+r_2x_2+\cdots+r_nx_n$. 我们有如下最优化数学模型,即投资组合模型:

$$\begin{cases}\min V(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\,;\\ \text{s. t.}\quad r_1x_1+r_2x_2+\cdots+r_nx_n\geqslant r_0\,,\\ \qquad\quad \displaystyle\sum_{i=1}^{n}x_i=a,\ x_i\geqslant 0,\ i=1,\cdots,n.\end{cases}$$

例 1.1.2 投资问题.

假定国家的下一个五年计划内用于发展某种工业的总投资为 b 亿元,可供选择兴建的项目共有 n 个.已知第 j 个项目的投资为 a_j 亿元,可得收益为 c_j 亿元,问:应如何进行投资才能使盈利率(即单位投资可得到的收益)为最高?

解 令决策变量为 x_j,则 x_j 应满足条件

$$x_j(x_j-1)=0.$$

同时 x_j 应满足约束条件 $\sum_{j=1}^{n} a_j x_j \leqslant b$. 目标函数是要求盈利率

$$f(x_1, x_2, \cdots, x_n) = \frac{\sum_{j=1}^{n} c_j x_j}{\sum_{j=1}^{n} a_j x_j}$$

达到最大. 这是一个整数规划问题.

例 1.1.3 厂址选择问题.

设有 n 个市场, 第 j 个市场位置为 (p_j, q_j), 它对某种货物的需要量为 $b_j (j = 1, \cdots, n)$. 现计划建立 m 个仓库, 第 i 个仓库的存储容量为 $a_i (i = 1, \cdots, m)$. 试确定仓库的位置, 使各仓库对各市场的运输量与路程乘积之和为最小.

解 设第 i 个仓库的位置为 (x_i, y_i) $(i = 1, \cdots, m)$, 第 i 个仓库到第 j 个市场的货物供应量为 $z_{ij} (i = 1, \cdots, m; j = 1, \cdots, n)$, 则第 i 个仓库到第 j 个市场的距离为

$$d_{ij} = \sqrt{(x_i - p_j)^2 + (y_i - q_j)^2},$$

目标函数为

$$\sum_{i=1}^{m} \sum_{j=1}^{n} z_{ij} d_{ij} = \sum_{i=1}^{m} \sum_{j=1}^{n} z_{ij} \sqrt{(x_i - p_j)^2 + (y_i - q_j)^2},$$

约束条件如下:

(1) 每个仓库向各市场提供的货物量之和不能超过它的存储容量;

(2) 每个市场从各仓库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, 问题的数学模型为

$$\begin{cases} \min f(\boldsymbol{x}) = \sum_{i=1}^{m} \sum_{j=1}^{n} z_{ij} d_{ij} = \sum_{i=1}^{m} \sum_{j=1}^{n} z_{ij} \sqrt{(x_i - p_j)^2 + (y_i - q_j)^2}; \\ \text{s. t.} \quad \sum_{j=1}^{n} z_{ij} \leqslant a_i, \ i = 1, \cdots, m, \\ \qquad \sum_{i=1}^{m} z_{ij} = b_j, \ j = 1, \cdots, n, \\ \qquad z_{ij} \geqslant 0, \ i = 1, \cdots, m; \ j = 1, \cdots, n. \end{cases}$$

例 1.1.4 曲线拟合问题.

在科学实验、工程设计和管理工作中, 经常会遇到下述问题: 通过实验或实测得到 n 组数据 (t_i, y_i), 它们可视为平面上的 n 个点, 期望确定一组参数 $\boldsymbol{x} =$

$(x_1, \cdots, x_m)^T$，使曲线 $y=\phi(x, t)$ 最佳逼近这 n 个点. 通常把这一问题归结为如下的优化问题：

$$\min f(\boldsymbol{x})=\sum_{i=1}^{n}[\phi(\boldsymbol{x}, t_i)-y_i]^2.$$

设 $\boldsymbol{x}^*$ 为问题的最优解，则 $y=\phi(\boldsymbol{x}^*, t)$ 即为通过最小二乘法对 n 组数据 (t_i, y_i)的拟合曲线.

例 1.1.5　运输问题.

设有 $V(\mathrm{m}^3)$的砂、石要由甲地运输到乙地，运输前需要先装入一个有底无盖并在底部有滑行器的木箱中，砂、石运到乙地后，从箱中倒出，再继续用空箱装运，不论箱子大小，每装运一箱，需 0.1 元，箱底和两端的材料费为每平方米 20 元，箱子两侧材料费为每平方米 5 元，箱底的两个滑行器与箱子同长，材料费为每米 2.5 元，问：木箱的长、宽、高应各为多少米，才能使运费与箱子的成本费的总和为最小？

解　设木箱的长、宽、高分别为 x_1, x_2, x_3，运费与成本费的总和为 w，则上述问题可归结为如下的优化模型：

$$\begin{cases}\min w(\boldsymbol{x})=\dfrac{0.1V}{x_1x_2x_3}+20x_1x_2+10x_1x_3+40x_2x_3+5x_1;\\ \text{s.t.}\quad x_i>0,\ i=1, 2, 3.\end{cases}$$

在此问题中，若要求箱子底与两侧使用废料来做，而废料只有 4 m^2，其他与上述问题相同，则这时问题归结为

$$\begin{cases}\min w(\boldsymbol{x})=\dfrac{0.1V}{x_1x_2x_3}+40x_2x_3+5x_1;\\ \text{s.t.}\quad \dfrac{1}{2}x_1x_3+\dfrac{1}{4}x_1x_2\leqslant 1,\\ \qquad\quad x_i>0,\ i=1, 2, 3.\end{cases}$$

§1.2　线性代数知识

在考虑最优化问题时，需要用到点列和极限的概念，而点是代数中线性空间的元素，极限又依赖于点与点之间的距离，距离是欧氏空间的概念，所以矩阵是一种工具，在最优化问题的表述和求解过程中自然是必不可少的. 下面重点介绍这 3 个概念.

1. 线性空间

设 V 是一个非空集合，$\mathbf{R}$ 是实数域，在 V 上定义加法，即对任意 $\boldsymbol{x}, \boldsymbol{y}\in V$，存在唯一的元素 $\boldsymbol{z}\in V$ 与它们对应，记作 $\boldsymbol{z}=\boldsymbol{x}+\boldsymbol{y}$，称为 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的和；对于任意

$\lambda \in \mathbf{R}$ 和任意 $\boldsymbol{x} \in V$, 存在唯一的元素 $\boldsymbol{y} \in V$ 与它们对应,记作 $\boldsymbol{y} = \lambda\boldsymbol{x}$, 称为 λ 与 $\boldsymbol{x}$ 的数量乘法. 若对任意 $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$ 和任意的 $\lambda, \mu \in \mathbf{R}$, 两种运算满足下面 8 个条件:

(1) $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$;

(2) $(\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z} = \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z})$;

(3) 存在一个元素,记作 $\mathbf{0}$,称为零元,使得 $\mathbf{0} + \boldsymbol{x} = \boldsymbol{x}$;

(4) 存在 $-\boldsymbol{x} \in V$,称为 $\boldsymbol{x}$ 的负元,使得 $\boldsymbol{x} + (-\boldsymbol{x}) = \mathbf{0}$;

(5) $1\boldsymbol{x} = \boldsymbol{x}$;

(6) $\lambda(\mu\boldsymbol{x}) = (\lambda\mu)\boldsymbol{x}$;

(7) $(\lambda + \mu)\boldsymbol{x} = \lambda\boldsymbol{x} + \mu\boldsymbol{x}$;

(8) $\lambda(\boldsymbol{x} + \boldsymbol{y}) = \lambda\boldsymbol{x} + \lambda\boldsymbol{y}$;

则称 V 为 $\mathbf{R}$ 上的线性空间,其中 V 的元素 $\boldsymbol{x}$ 称为向量.

通常我们考虑的线性空间是 n 维(列)向量空间 $\mathbf{R}^n$,记 n 维列向量为

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}.$$

若 $\boldsymbol{x}, \boldsymbol{y}$ 是线性空间 $\mathbf{R}^n$ 中的两个向量,通常的加法定义为

$$\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, x_2 + y_2, \cdots, x_n + y_n)^{\mathrm{T}},$$

若 λ 是一个标量,通常的数乘定义为

$$\lambda\boldsymbol{x} = (\lambda x_1, \lambda x_2, \cdots, \lambda x_n)^{\mathrm{T}}.$$

设 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_m$ 为 $\mathbf{R}^n$ 中 m 个向量,称

$$\boldsymbol{x} = \lambda_1\boldsymbol{\alpha}_1 + \cdots + \lambda_m\boldsymbol{\alpha}_m$$

为 $\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_m$ 的线性组合.

如果存在不全为零的 $\lambda_i (i = 1, 2, \cdots, m)$, 使得

$$\lambda_1\boldsymbol{\alpha}_1 + \cdots + \lambda_m\boldsymbol{\alpha}_m = \mathbf{0},$$

则称 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_m$ 线性相关,否则称 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_m$ 线性无关,也就是说,若

$$\lambda_1\boldsymbol{\alpha}_1 + \cdots + \lambda_m\boldsymbol{\alpha}_m = \mathbf{0},$$

则必有 $\lambda_i = 0 \ (i = 1, \cdots, m)$.

如果线性空间 $\mathbf{R}^n$ 的一个非空子集 W 对于 $\mathbf{R}^n$ 的加法和数乘运算也构成一个线性空间,则称它为 $\mathbf{R}^n$ 的子空间. 如果 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_m \in W$ 线性无关,而 W 中的任何向量均可以表示成它们的线性组合,则 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_m$ 称为 W 的一组基,且 $W = \{\lambda_1\boldsymbol{\alpha}_1 + \lambda_2\boldsymbol{\alpha}_2 + \cdots + \lambda_m\boldsymbol{\alpha}_m \mid \lambda_1, \lambda_2, \cdots, \lambda_m \in \mathbf{R}\}$. 若 $x \in \mathbf{R}^n$,则称

$x+W$ 为 $\mathbf{R}^n$ 的仿射集.

2. Euclid 空间(欧氏空间)

所谓欧氏空间,就是在线性空间上定义一个度量.对于 n 维欧氏空间(记为 $\mathbf{R}^n$),向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的内积定义为

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{n} x_i y_i = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{y}.$$

由上式,容易得到内积具有如下性质:

(1) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geqslant 0$,且$\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$ 的充分必要条件是 $\boldsymbol{x} = \mathbf{0}$;

(2) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$;

(3) $\langle \lambda\boldsymbol{x} + \mu\boldsymbol{y}, \boldsymbol{z} \rangle = \lambda\langle \boldsymbol{x}, \boldsymbol{z} \rangle + \mu\langle \boldsymbol{y}, \boldsymbol{z} \rangle$.

n 维欧氏空间 $\mathbf{R}^n$ 上的范数定义为

$$\|\boldsymbol{x}\| = \Big(\sum_{i=1}^{n} x_i^2\Big)^{\frac{1}{2}} = (\langle \boldsymbol{x}, \boldsymbol{x} \rangle)^{\frac{1}{2}},$$

即通常意义下的距离,或称为 l_2 范数.

范数具有如下性质:

(1) $\|\boldsymbol{x}\| \geqslant 0$,且 $\|\boldsymbol{x}\| = 0$ 的充分必要条件是 $\boldsymbol{x} = \mathbf{0}$;

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

(3) $\|\boldsymbol{x}+\boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ (三角不等式),以及Cauchy-Schwarz不等式

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leqslant \|\boldsymbol{x}\| \, \|\boldsymbol{y}\|,$$

且等号成立的充分必要条件是:$\boldsymbol{x}$ 与 $\boldsymbol{y}$ 共线,即存在 λ,使得

$$\boldsymbol{x} = \lambda\boldsymbol{y}.$$

将该不等式写成分量形式为

$$\Big|\sum_{i=1}^{n} x_i y_i\Big| \leqslant \Big(\sum_{i=1}^{n} x_i^2\Big)^{\frac{1}{2}} \Big(\sum_{i=1}^{n} y_i^2\Big)^{\frac{1}{2}}.$$

3. 矩阵

称

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

为 $m\times n$ 矩阵，记为 $\boldsymbol{A}=\boldsymbol{A}_{m\times n}$. 实数域上所有 $m\times n$ 矩阵组成的集合记作 $\mathbf{R}^{m\times n}$.

对于矩阵 $\boldsymbol{A}$，可以进行分块，即

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{pmatrix},$$

其中 $\boldsymbol{A}_{11}$ 为 $m_1\times n_1$ 矩阵，$\boldsymbol{A}_{12}$ 为 $m_1\times n_2$ 矩阵，$\boldsymbol{A}_{21}$ 为 $m_2\times n_1$ 矩阵，$\boldsymbol{A}_{22}$ 为 $m_2\times n_2$ 矩阵，且 $m_1+m_2=m$, $n_1+n_2=n$. 这里特别提到两种特殊的分块矩阵，按列分块

$$\boldsymbol{A}=(\boldsymbol{P}_1,\ \boldsymbol{P}_2,\ \cdots,\ \boldsymbol{P}_n),\ \boldsymbol{P}_j=\begin{pmatrix}a_{1j}\\ a_{2j}\\ \vdots\\ a_{mj}\end{pmatrix},$$

和按行分块

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{a}_1^{\mathrm{T}}\\ \boldsymbol{a}_2^{\mathrm{T}}\\ \vdots\\ \boldsymbol{a}_m^{\mathrm{T}}\end{pmatrix},\ \boldsymbol{a}_i^{\mathrm{T}}=(a_{i1},\ a_{i2},\ \cdots,\ a_{in}).$$

矩阵 $\boldsymbol{A}$ 的秩记为 $\mathrm{rank}(\boldsymbol{A})$. 当

$$\mathrm{rank}(\boldsymbol{A})=\min\{m,\ n\}$$

时，称矩阵 $\boldsymbol{A}$ 是满秩的，又若 $m<n$, 则称 $\boldsymbol{A}$ 为行满秩；若 $m>n$, 则称 $\boldsymbol{A}$ 为列满秩，若 $m=n$, 则矩阵 $\boldsymbol{A}$ 为 n 阶非奇异方阵. 元素 $a_{ij}\,(i=1,\ 2,\ \cdots,\ m;\ j=1,\ 2,\ \cdots,\ n)$ 均为零的 $m\times n$ 矩阵为零矩阵，记为 $\boldsymbol{O}$.

方阵非奇异的充分必要条件是：$\det(\boldsymbol{A})\neq 0$.

若 $\boldsymbol{A}$ 为 $n\times n$ 矩阵，且 $\boldsymbol{A}$ 满足

$$\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A},$$

则称 $\boldsymbol{A}$ 为对称矩阵；若对于一切 $\boldsymbol{x}\neq\boldsymbol{0}$, 均有

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}>0,$$

则称 $\boldsymbol{A}$ 为正定矩阵；若对一切 $\boldsymbol{x}$，均有

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\geqslant 0,$$

则称 $\boldsymbol{A}$ 为半正定矩阵.

§1.3 多元函数分析

分析多元函数在一点附近的特性，它在该点处的一阶微分和二阶微分是两个重要的工具. 它们在该点处的线性近似和二次近似对于考虑这个函数在该点处的最优性条件是非常有用的.

定义 1.3.1 设 n 元函数 $f(x)$ 对自变量 $\boldsymbol{x}=(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 的各分量 x_i 的偏导数 $\frac{\partial f(x)}{\partial x_i}$ $(i=1, \cdots, n)$ 都存在，则称函数 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}$ 处一阶可导，并称向量

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

为函数 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}$ 处的一阶导数或梯度，记 $g(\boldsymbol{x})=\nabla f(\boldsymbol{x})$ ($g(\boldsymbol{x})$ 为列向量).

定义 1.3.2 设 n 元函数 $f(\boldsymbol{x})$ 对自变量 $\boldsymbol{x}=(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 的各分量 x_i 的二阶偏导数 $\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}$ $(i=1, \cdots, n;\ j=1, \cdots, n)$ 都存在，则称函数 $f(\boldsymbol{x})$ 在点 $\boldsymbol{x}$ 处二阶可导，并称矩阵

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

为 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}$ 处的二阶导数或 Hesse 矩阵，记为 $\nabla^2 f(\boldsymbol{x})$，即

$$\nabla^2 f(\boldsymbol{x})=\left(\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}\right)_{n\times n}.$$

有时记为 $G(\boldsymbol{x})$.

若 $f(\boldsymbol{x})$ 对 $\boldsymbol{x}$ 各变元的所有二阶偏导数都连续，则

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}=\frac{\partial^2 f(\boldsymbol{x})}{\partial x_j \partial x_i},$$

此时 $\nabla^2 f(\boldsymbol{x})$ 为对称矩阵.

例 1.3.1 设 $\boldsymbol{A}\in\mathbf{R}^{n\times n}$ 为对称矩阵，$\boldsymbol{b}\in\mathbf{R}^n$，$c\in\mathbf{R}$，求：