JUZHENBUDENGSHI

矩阵不等式

燕子宗　余瑞艳　熊勤学　编著

同济大学出版社
TONGJI UNIVERSITY PRESS

内 容 提 要

本书主要讲述了矩阵不等式的重要结果和重要方法.作者强调思想方法,选择了重要的结果和技巧作为素材,注重对矩阵不等式的新思想和新方法的归纳和整理,内容丰富,具有一定深度,反映了矩阵不等式最新研究成果.

全书共分 14 章.第 1 章介绍矩阵论预备知识,第 2 到 14 章分别讨论了 Cauchy-Bunyakovsky-Schwarz 型不等式及其逆形式、控制不等式、Schur 补理论、投影方法、特征值的估计、矩阵单调函数、变分方法、凸性方法、Kantorovich 型矩阵不等式、算子不等式,数值域和幂有界算子.本书重点讨论了 Cauchy-Bunyakovsky-Schwarz 型不等式及其逆,凸性方法构造矩阵不等式以及矩阵单调性等内容,对最近的数值域和幂有界算子等前沿问题也给予了充分关注.全书表达简洁流畅,读者可以在较短时间内了解和掌握矩阵不等式的主要内容和主要方法.

本书读者对象为高等院校高年级本科生、研究生,有关专业的教师、数学工作者及有关工程技术人员.

图书在版编目(CIP)数据

矩阵不等式/燕子宗,余瑞艳,熊勤学编著.-- 上海:同济大学出版社,2012.5

ISBN 978-7-5608-4620-0

Ⅰ.①矩… Ⅱ.①燕… ②余… ③熊… Ⅲ.①矩阵—不等式 Ⅳ.①O151.21

中国版本图书馆 CIP 数据核字(2011)第 150612 号

矩阵不等式

燕子宗 余瑞艳 熊勤学 编著

策划编辑 杨 闯 **责任编辑** 李小敏 **责任校对** 徐春莲 **封面设计** 潘向蓁

出版发行 同济大学出版社 www.tongjipress.com.cn

(地址:上海市四平路 1239 号 邮编:200092 电话:021-65985622)

经 销 全国各地新华书店

印 刷 同济大学印刷厂

开 本 787 mm×1092 mm 1/16

印 张 16.75

字 数 418 000

版 次 2012 年 5 月第 1 版 2012 年 5 月第 1 次印刷

书 号 ISBN 978-7-5608-4620-0

定 价 49.00 元

前　　言

在数学科学几乎所有的分支中，不等式都起着非常重要的作用，关于它的研究一直是非常活跃且极富吸引力的. 作为不等式重要的组成部分，矩阵不等式内容也十分丰富. 为了比较系统地再现这方面的基本理论和应用成果，本书对此进行了广泛收集，并归纳整理成为论述矩阵不等式的一本专著.

本书基本涵盖了矩阵不等式的主要思想和方法，着重介绍一些基础的和重要的不等式. 全书共分 14 章，章节内容大体按照内容和方法划分. 第 1 章介绍矩阵论预备知识，第 2 到 14 章分别讨论了 CBS 型不等式及其逆形式、特征值的估计、Schur 补理论、投影方法、控制不等式、矩阵单调函数、变分方法、凸性方法、Kantorovich 型矩阵不等式、算子不等式、数值域和幂有界算子等内容.

本书用较大篇幅介绍了 CBS 不等式的各种形式及其推广，开辟 4 章进行专题研究，并在其他章节也有所讨论. CBS 不等式全称为 Cauchy-Bunyakovsky-Schwarz 不等式，通常也称为 Cauchy 不等式或 Cauchy-Schwarz 不等式. 该不等式及其逆形式在现代数学中扮演着极端重要的角色，被广泛应用于 Hilbert 空间理论、概率统计、经典实分析和复分析、数值分析、微分方程的定性理论及相关应用领域当中. 例如以 Kantorovich 不等式为代表的矩阵不等式在优化、统计和控制等领域得到了广泛应用. 20 世纪 90 年代，Marshall 和 Olkin 成功地将其推广到矩阵形式，形成了一个研究该不等式的热潮，其中以 Mond 和 Pečerić为代表给出的凸性方法为构造新的矩阵不等式开辟了一条有效途径. 结合 Schur 补理论，获得了许多非常有用的结果.

本书其他部分包括了特征值的估计、矩阵凸性和单调性、幂有界算子以及最近兴起的联合数值域等内容. 特征值的估计内容已包含在经典矩阵分析内容中，Courant-Fischer 变分定理和 Poincaré 分隔定理是该领域的重要成果. 借助控制不等式这一有利工具，本书比较全面地介绍了特征值估计的经典结果.

矩阵的 Löwner 序是不等式的一个重要分支，近年来学者为此作了一些卓有成效的工作，本书给予充分关注，比较系统地介绍了矩阵凸性和矩阵单调性质及其相应结果.

幂有界算子在常系数线性偏微分方程组初值问题的差分格式中发挥重要作用，主要结果之一 Kreiss 矩阵定理是最基本的，并由此给出了差分格式稳定性的一系列充分必要条件.

虽然数值域概念早在 20 世纪初就被 Toeplitz 提出，但是直到 80 年代才被引起重视，并在泛函分析和量子物理等多个领域得到广泛应用，形成一个新的研究热点. 本书对由此发展起来的联合数值域及其应用进行了初步讨论.

概括起来,本书具有如下特色:

强调不等式之间的内在联系,例如对 CBS 不等式及其逆的各种变形分成为 4 章分别讨论.

各种不等式的等价性尽量用简练语言阐述,许多等价性命题用一个定理来描述,且定理的证明力求采用简单的证明方法.

章节编排不拘泥于形式,一般按照所采用的方法来编排,部分也按照内容来编排.内容多而不乱,搭配合理.

注重多领域矩阵不等式的收集与整理.对统计、优化、数值分析、控制和图论中的一些矩阵不等式也给予了充分关注,对推导过于繁杂的不等式则未列入.

反映了矩阵不等式最新研究成果.本书大部分内容是 20 世纪 80 年代以后的,较多关注了近年来该领域的研究动向.

全书起点较低,终点较高.它既适宜于一般读者,又适宜于需要深入了解矩阵不等式的科研人员.

内容基本自闭,避免读者查阅资料的烦恼.

在资料收集和整理过程中,作者发现矩阵不等式的离散形式内容也非常丰富.由于它们的来源不尽相同,许多与矩阵不等式只是存在形式上的差别.本书并没有强行把所有离散形式都统一成矩阵形式,而是选择保留了一部分离散不等式,以及它们的特有方法和重要结果.除非特别需要,对相关的积分不等式较少涉及.

本书无法包含矩阵不等式的全部内容.由于个人的喜好,我们遴选了一些感兴趣的不等式,能否得到读者的认同有待于实践检验.鉴于作者水平,不妥之处也在所难免,恳请国内同行和广大读者不吝赐教.

本书的出版得到了国家自然科学基金(项目编号 30870460)、湖北省教育厅科学研究项目(项目编号 D20111305, Q20101309)和长江大学博士启动基金的支持,特别是得到了李传仁教授的大力支持,在此表示感谢.

作者

2011 年 8 月于古城荆州

长江大学

目　　录

第 1 章　预备知识

本书目的在于论述矩阵不等式的主要结果和方法. 为便于查阅，本章简要介绍所需要的预备知识和相关结果，并假定读者已经具备线性代数基础.

1.1　范数与内积

向量范数是用来刻画向量大小的一种度量，起源于实数的绝对值和向量长度等概念. 本书主要涉及两个线性空间，即复线性空间 $\mathbf{C}^n$ 和实线性空间 $\mathbf{R}^n$. 下面给出线性空间范数的定义.

定义 1.1.1　设 V 是数域 P 上的线性空间，$\|\cdot\|:V\to P$ 称为 V 上的**范数**，如果它满足如下性质：

(1) 对任意 $\boldsymbol{x}\in V$，有 $\|\boldsymbol{x}\|\geqslant 0$，且 $\|\boldsymbol{x}\|=0$ 当且仅当 $\boldsymbol{x}=\mathbf{0}$；

(2) 对任意 $\boldsymbol{x}\in V$ 和 $\alpha\in P$，有 $\|\alpha\boldsymbol{x}\|=|\alpha|\,\|\boldsymbol{x}\|$；

(3) 对任意 $\boldsymbol{x},\ \boldsymbol{y}\in V$，有 $\|\boldsymbol{x}+\boldsymbol{y}\|\leqslant\|\boldsymbol{x}\|+\|\boldsymbol{y}\|$.

定义了范数的线性空间 V 叫做线性赋范空间，而定义了范数的 $\mathbf{C}^n$ 和 $\mathbf{R}^n$ 分别叫做复线性赋范空间和实线性赋范空间.

定义 1.1.1 的三条性质分别称为非负性、正齐次性和次可加性，条件(3)也称为三角不等式. 由定义 1.1.1 容易导出

$$\|\mathbf{0}\|=0,\quad \|-\boldsymbol{x}\|=\|\boldsymbol{x}\|,\quad \big|\,\|\boldsymbol{x}\|-\|\boldsymbol{y}\|\,\big|\leqslant\|\boldsymbol{x}-\boldsymbol{y}\|.$$

引理 1.1.1(Young)　设 $a\geqslant 0,\ b\geqslant 0$，则

$$a^tb^{1-t}\leqslant ta+(1-t)b,\quad \forall\, t\in(0,\ 1). \tag{1.1}$$

证　令 $x=\dfrac{a}{b}$，然后利用 $f(x)=x^t-tx-(1-t)$ 的单调性即可证明. 证毕.

引理 1.1.2(Hölder)　设 $p,\ q\geqslant 1$ 满足 $\dfrac{1}{p}+\dfrac{1}{q}=1,\ a_i,\ b_i\geqslant 0\ (i=1,\ 2,\ \cdots,\ n)$. 则

$$\sum_{i=1}^{n}a_ib_i\leqslant\Big(\sum_{i=1}^{n}a_i^p\Big)^{\frac{1}{p}}\Big(\sum_{i=1}^{n}b_i^q\Big)^{\frac{1}{q}}. \tag{1.2}$$

证　在式(1.1)中令 $t=\dfrac{1}{p},\ a=a_i^p,\ b=b_i^q$，则

$$a_ib_i \leqslant \frac{1}{p}a_i^p + \frac{1}{q}b_i^q. \tag{1.3}$$

式(1.3)两边对 i 求和,得到

$$\sum_{i=1}^n a_ib_i \leqslant \frac{1}{p}\sum_{i=1}^n a_i^p + \frac{1}{q}\sum_{i=1}^n b_i^q. \tag{1.4}$$

再将式(1.4)中的 a_i 和 b_i 分别用

$$\frac{a_i}{\left(\sum_{i=1}^n a_i^p\right)^{\frac{1}{p}}} \text{和} \frac{b_i}{\left(\sum_{i=1}^n b_i^q\right)^{\frac{1}{q}}}$$

替换即可得到所需的结论. 证毕.

当 $p=2$ 时,式(1.2) 就是众所周知的 Cauchy-Schwarz 不等式. 本书称它为 CBS 不等式,将在后面予以重点关注.

例 1.1.1 设 $\boldsymbol{x}=(x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in \mathbf{C}^n$,则

$$\|\boldsymbol{x}\|_p = \left\{\sum_{i=1}^n |x_i|^p\right\}^{\frac{1}{p}}, \quad 1 \leqslant p < +\infty \tag{1.5}$$

是 $\mathbf{C}^n$ 上的范数,称为 Hölder 范数. 事实上,因为

$$\sum_{i=1}^n |x_i|(|x_i|+|y_i|)^{p-1} \leqslant \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}\left(\sum_{i=1}^n (|x_i|+|y_i|)^{(p-1)q}\right)^{\frac{1}{q}},$$

$$\sum_{i=1}^n |y_i|(|x_i|+|y_i|)^{p-1} \leqslant \left(\sum_{i=1}^n |y_i|^p\right)^{\frac{1}{p}}\left(\sum_{i=1}^n (|x_i|+|y_i|)^{(p-1)q}\right)^{\frac{1}{q}}$$

以及 $(p-1)q=p$,于是

$$\begin{aligned}\sum_{i=1}^n |x_i+y_i|^p &\leqslant \sum_{i=1}^n (|x_i|+|y_i|)^p \\ &= \sum_{i=1}^n |x_i|(|x_i|+|y_i|)^{p-1} + \sum_{i=1}^n |y_i|(|x_i|+|y_i|)^{p-1} \\ &\leqslant \left(\left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^n |y_i|^p\right)^{\frac{1}{p}}\right)\left(\sum_{i=1}^n (|x_i|+|y_i|)^p\right)^{\frac{1}{q}}.\end{aligned}$$

从而成立不等式

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

即关于范数的三角不等式 $\|\boldsymbol{x}+\boldsymbol{y}\|_p \leqslant \|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p$ 成立. 另外两个条件验证是简单的.

当 $p=2$ 时,$\|\boldsymbol{x}\|_2$ 就是通常意义的 Eculid 范数,也是本书用得最多的一种向量范数. 在不标明的情况下,都是指这种范数. 另外两种常用范数为 $p=1$ 和 $p=\infty$ 情形.

$$\|\boldsymbol{x}\|_1 = |x_1|+|x_2|+\cdots+|x_n|, \tag{1.7}$$

$$\|x\|_2=\sqrt{|x_1|^2+|x_2|^2+\cdots+|x_n|^2}, \tag{1.8}$$

$$\|x\|_\infty=\max_i |x_i|. \tag{1.9}$$

由于 $\|x\|_\infty^p=(\max_i |x_i|)^p\leqslant\sum_{i=1}^n |x_i|^p\leqslant n\max_i |x_i|^p=n\|x\|_\infty^p$,因此$\lim_{p\to\infty}\|x\|_p=\|x\|_\infty$. 此外,当 $0<t\leqslant s$ 时,

$$\frac{\left\{\sum_{i=1}^n |x_i|^s\right\}^{\frac{1}{s}}}{\left\{\sum_{j=1}^n |x_j|^t\right\}^{\frac{1}{t}}}=\left\{\sum_{i=1}^n \frac{|x_i|^s}{\left\{\sum_{j=1}^n |x_j|^t\right\}^{\frac{s}{t}}}\right\}^{\frac{1}{s}}=\left\{\sum_{i=1}^n\left\{\frac{|x_i|^t}{\sum_{j=1}^n |x_j|^t}\right\}^{\frac{s}{t}}\right\}^{\frac{1}{s}}$$

$$\leqslant\left\{\sum_{i=1}^n \frac{|x_i|^t}{\sum_{j=1}^n |x_j|^t}\right\}^{\frac{1}{s}}=1,$$

即 $\|x\|_s\leqslant\|x\|_t$.

把 $\mathbf{C}^n$ 上的范数概念推广到 $\mathbf{C}^{m\times n}$,则得到矩阵范数,一些特殊的矩阵范数随后给出. 经常使用的是算子范数和酉不变范数. 设 $A\in\mathbf{R}^{n\times n}$,其算子范数定义为

$$\|A\|=\max_{x\neq 0}\left\{\frac{\|Ax\|}{\|x\|}\right\}, \tag{1.10}$$

其中 $\|x\|$ 为某一向量范数. 容易验证式(1.10)满足范数三个条件. 由 $\|\cdot\|_1$, $\|\cdot\|_\infty$, $\|\cdot\|_2$ 诱导的三种常用矩阵范数为

$$\|A\|_1=\max_j\sum_{i=1}^n |a_{ij}|, \tag{1.11}$$

$$\|A\|_\infty=\max_i\sum_{j=1}^n |a_{ij}|, \tag{1.12}$$

$$\|A\|_2=\sqrt{\rho(A^*A)}, \tag{1.13}$$

这里 $\rho(A^*A)$ 表示矩阵 A^*A 的谱半径,即矩阵特征值最大模. 对于矩阵 A 的谱范数,容易证明

$$\|A^*\|_2=\|A^{\mathrm{T}}\|_2=\|A\|_2,$$

$$\|A^*A\|_2=\|A\|_2^2,$$

$$\|A\|_2=\max\{|y^*Ax| \mid \|x\|_2=1, \|y\|_2=1\}.$$

该范数也是本书用得最多的一种矩阵范数.

一个矩阵范数 $\|\cdot\|$ 称为酉不变范数,如果对于任意酉矩阵 $U\in\mathbf{C}^{m\times m}$ 和 $V\in\mathbf{C}^{n\times n}$,有 $\|UAV\|=\|A\|$,且当矩阵 $A=xy^*$ 时,$\|A\|=\|x\|_2\|y\|_2$.

$\|\cdot\|_2$ 是酉不变范数,而 $\|\cdot\|_1$ 和 $\|\cdot\|_\infty$ 不是. 酉不变范数是一种重要范数,关于它的进一步研究可以参阅詹兴致的《矩阵论》(2006).

定义 1.1.2 设V是数域P上的线性空间,二元映射$(\boldsymbol{x},\ \boldsymbol{y}):V\times V\to P$称为**内积**,如果满足

(1) $(\boldsymbol{x},\ \boldsymbol{x})\geqslant 0$; $(\boldsymbol{x},\ \boldsymbol{x})=0$ 当且仅当 $\boldsymbol{x}=\boldsymbol{0}$;

(2) $(\boldsymbol{x},\ \boldsymbol{y})=\overline{(\boldsymbol{y},\ \boldsymbol{x})}$, $\forall\,\boldsymbol{x},\ \boldsymbol{y}\in V$;

(3) $(\lambda\boldsymbol{x},\ \boldsymbol{y})=\bar{\lambda}(\boldsymbol{x},\ \boldsymbol{y})$, $\forall\lambda\in P$, $\forall\,\boldsymbol{x},\ \boldsymbol{y}\in V$;

(4) $(\boldsymbol{x}+\boldsymbol{y},\ \boldsymbol{z})\leqslant(\boldsymbol{x},\ \boldsymbol{z})+(\boldsymbol{y},\ \boldsymbol{z})$, $\forall\,\boldsymbol{x},\ \boldsymbol{y}\in V$.

定义了内积的线性空间称为内积空间,完备的内积空间又称为 Hilbert 空间. 两个常用的内积空间为复内积空间 $\mathbf{C}^n$ 和实内积空间 $\mathbf{R}^n$.

例如,对于 $\boldsymbol{x}=(x_1,\ x_2,\ \cdots,\ x_n)^{\mathrm{T}}\in\mathbf{C}^n$, $\boldsymbol{y}=(y_1,\ y_2,\ \cdots,\ y_n)^{\mathrm{T}}\in\mathbf{C}^n$,可以定义复欧氏空间内积为

$$(\boldsymbol{x},\ \boldsymbol{y})=\sum_{i=1}^{n}\bar{x}_i y_i. \tag{1.14}$$

又如,对于 $\boldsymbol{A}=(a_{ij})$, $\boldsymbol{B}=(b_{ij})\in\mathbf{C}^{m\times n}$,定义 Frobenius 内积为

$$(\boldsymbol{A},\ \boldsymbol{B})=\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}^*)=\sum_{i=1}^{m}\sum_{j=1}^{n}\bar{a}_{ij}b_{ij}. \tag{1.15}$$

这里 tr(·)代表矩阵的迹,即矩阵所有对角元素之和.

由内积$(\boldsymbol{x},\ \boldsymbol{y})_{\boldsymbol{A}}=(\boldsymbol{A}\boldsymbol{x},\ \boldsymbol{y})$ 导出的范数

$$\|\boldsymbol{x}\|_{\boldsymbol{A}}=\sqrt{(\boldsymbol{A}\boldsymbol{x},\ \boldsymbol{x})} \tag{1.16}$$

称为椭球范数,其中 $\boldsymbol{A}$ 是正定 Hermite 矩阵. 若 $\boldsymbol{A}$ 是单位矩阵,它代表普通的欧氏范数. 又如在空间 $\mathbf{C}^{m\times n}$中,由 Frobenius 内积导出的范数

$$\|\boldsymbol{A}\|_F=\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2} \tag{1.17}$$

称为 Frobenius 范数,它是酉不变的.

借助内积概念和 Cauchy-Schwarz 不等式,可以把几何空间中向量的长度、角度和距离等概念引入到线性空间中,从而获得线性空间中的几何拓扑结构. 该不等式允许定义两向量 $\boldsymbol{a}$, $\boldsymbol{b}$ 的夹角余弦

$$\cos\phi=\frac{(\boldsymbol{a},\ \boldsymbol{b})}{\|\boldsymbol{a}\|\ \|\boldsymbol{b}\|}. \tag{1.18}$$

若$(\boldsymbol{a},\ \boldsymbol{b})=0$,则称 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 垂直,记作 $\boldsymbol{a}\perp\boldsymbol{b}$. 若 $\boldsymbol{a}$ 与子空间M 中每个向量均垂直,则称 $\boldsymbol{a}$ 与M 垂直,记作 $\boldsymbol{a}\perp M$. 所有与 M 垂直的向量构成一个线性子空间,记作 $M^{\perp}$.

引理 1.1.3(勾股定理) 设 $\boldsymbol{a}$, $\boldsymbol{b}\in\mathbf{C}^n$. 若 $\boldsymbol{a}\perp\boldsymbol{b}$,则 $\|\boldsymbol{a}+\boldsymbol{b}\|^2=\|\boldsymbol{a}\|^2+\|\boldsymbol{b}\|^2$; 反之亦然.

证 由于

$$\begin{aligned}\|\boldsymbol{a}+\boldsymbol{b}\|^2&=(\boldsymbol{a}+\boldsymbol{b},\ \boldsymbol{a}+\boldsymbol{b})\\&=(\boldsymbol{a},\ \boldsymbol{a}+\boldsymbol{b})+(\boldsymbol{b},\ \boldsymbol{a}+\boldsymbol{b})\\&=(\boldsymbol{a},\ \boldsymbol{a})+(\boldsymbol{a},\ \boldsymbol{b})+(\boldsymbol{b},\ \boldsymbol{a})+(\boldsymbol{b},\ \boldsymbol{b}),\end{aligned}$$

若 $\boldsymbol{a} \perp \boldsymbol{b}$,则$(\boldsymbol{a}, \boldsymbol{b}) = 0$ 等价于 $\|\boldsymbol{a}+\boldsymbol{b}\|^2 = \|\boldsymbol{a}\|^2 + \|\boldsymbol{b}\|^2$. 证毕.

定理 1.1.1(投影定理)　设 $\boldsymbol{a} \in \mathbf{C}^n$, $\boldsymbol{b} \in M \subset \mathbf{C}^n$, M 是 $\mathbf{C}^n$ 的子空间. 若$(\boldsymbol{a}-\boldsymbol{b}) \perp M$,则

$$\|\boldsymbol{a}-\boldsymbol{b}\| \leqslant \|\boldsymbol{a}-\boldsymbol{x}\|, \quad \forall \boldsymbol{x} \in M. \tag{1.19}$$

证　由 $\boldsymbol{a}-\boldsymbol{x} = (\boldsymbol{a}-\boldsymbol{b}) + (\boldsymbol{b}-\boldsymbol{x})$ 和 $\boldsymbol{b}-\boldsymbol{x} \in M$ 及假设知$(\boldsymbol{a}-\boldsymbol{b}) \perp (\boldsymbol{b}-\boldsymbol{x})$. 利用引理 1.1.3 得

$$\|\boldsymbol{a}-\boldsymbol{x}\|^2 = \|\boldsymbol{a}-\boldsymbol{b}\|^2 + \|\boldsymbol{b}-\boldsymbol{x}\|^2 \geqslant \|\boldsymbol{a}-\boldsymbol{b}\|^2. \tag{1.20}$$

于是结论成立. 证毕.

定理 1.1.1 中的向量 $\boldsymbol{b}$ 被称为向量 $\boldsymbol{a}$ 在空间 M 中的正交投影,类似地可以根据内积 $(\cdot, \cdot)_A$ 定义 A-正交投影.

1.2　奇异值分解

将一个矩阵分解为比较简单或者具有某种特性的几个矩阵的和与积,能反映矩阵的数值特性,如矩阵的秩、特征值和奇异值等,为数值计算和理论分析提供依据. 本节简要给出本书需要的几种分解.

定理 1.2.1　设 $\boldsymbol{A} \in \mathbf{C}^{m\times n}$ 是列满秩矩阵,则存在 $n\times n$ 阶上三角矩阵 $\boldsymbol{R}$ 和 $m \times m$ 阶酉矩阵 $\boldsymbol{U}$,使得

$$\boldsymbol{A} = \boldsymbol{U}\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{O}\end{pmatrix}. \tag{1.21}$$

证　只要对 $\boldsymbol{A}$ 的列向量实施 Gram-Schmidt 正交化过程即可(参见公式(2.32)). 证毕.

定理 1.2.2(Schur 分解)　设 $\boldsymbol{A} \in \mathbf{C}_r^{m\times n}$,则存在 $m\times m$ 阶酉矩阵 $\boldsymbol{U}$, $n\times n$ 阶酉矩阵 $\boldsymbol{V}$ 及 r 阶下三角矩阵 $\boldsymbol{L}$,使得

$$\boldsymbol{A} = \boldsymbol{U}\begin{pmatrix}\boldsymbol{L} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}\boldsymbol{V}. \tag{1.22}$$

证　由于 $\boldsymbol{A}$ 的秩为 r,则 $\boldsymbol{A}$ 有 r 个列向量线性无关. 不妨设 $\boldsymbol{A}$ 的前 r 个列向量线性无关,其余列向量均是它们的线性组合,即

$$(a_{r+1}, \cdots, a_n) = (a_1, \cdots, a_r)\boldsymbol{B},$$

这里 $\boldsymbol{B} \in \mathbf{C}^{r\times(n-r)}$. 于是 $\boldsymbol{A} = (a_1, \cdots, a_r)(I_r, \boldsymbol{B})$. 由定理 1.2.1 知,存在酉矩阵 $\boldsymbol{U} \in \mathbf{C}^{m\times m}$ 和上三角矩阵 $\boldsymbol{R} \in \mathbf{C}^{r\times r}$ 使得

$$(a_1, \cdots, a_r) = \boldsymbol{U}\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{O}\end{pmatrix}. \tag{1.23}$$

于是

$$A = U\begin{pmatrix} R \\ O \end{pmatrix}(I_r,\ B) = U\begin{pmatrix} R & RB \\ O & O \end{pmatrix}. \tag{1.24}$$

再对行满秩矩阵(R:RB)执行同样操作,则得到所需的结论. 证毕.

定理 1.2.3(满秩分解) 设 $A \in C_r^{m\times n}$,则存在 $B \in C_r^{m\times r}$ 和 $D \in C_r^{r\times n}$,使得 $A = BD$.

证 在分解式(1.22)中,将 U, V 分别分块为

$$U = (U_1,\ U_2), \quad V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix},$$

要求与分解式(1.22)中间矩阵分块乘积相容. 于是

$$A = (U_1,\ U_2)\begin{pmatrix} L & O \\ O & O \end{pmatrix}\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = U_1 L V_1.$$

再分别取 $B = U_1$ 和 $D = LV_1$ 即可. 证毕.

定理 1.2.4(奇异值分解) 设 $A \in C_r^{m\times n}$. 则存在 $m\times m$ 阶酉矩阵 U, $n\times n$ 阶酉矩阵 V 及 r 阶对角矩阵 A_σ 使得

$$A = U\begin{pmatrix} A_\sigma & O \\ O & O \end{pmatrix}V, \tag{1.25}$$

其中,$A_\sigma = \mathrm{diag}(\sigma_1,\ \cdots,\ \sigma_r)$ 的主对角元素 $\sigma_i > 0$ $(i = 1,\ 2,\ \cdots,\ r)$ 称为矩阵 A 的奇异值.

证 设 $u_1,\ \cdots,\ u_r$ 为 A^*A 对应于 $\sigma_1^2,\ \cdots,\ \sigma_r^2$ 的标准正交的特征向量. 令 $v_i = \dfrac{1}{\sigma_i}A^*u_i$, $i = 1,\ 2,\ \cdots,\ r$. 将其分别扩张成为 C^m 和 C^n 中的标准正交基,构造两个酉矩阵 $U \in C^{m\times m}$ 和 $V \in C^{n\times n}$,则有

$$A = UU^*A = (u_1u_1^* + \cdots + u_mu_m^*)A = \sigma_1 u_1 v_1^* + \cdots + \sigma_r u_r v_r^* = U\begin{pmatrix} A_\sigma & O \\ O & O \end{pmatrix}V.$$

于是结论成立. 证毕.

定理 1.2.5(极分解) 设 $A \in C^{n\times n}$,则存在半正定矩阵 P, Q 和酉矩阵 U,使得 $A = PU = UQ$.

证 由定理 1.2.4 知,存在酉矩阵 U_1, V_1 和非负对角矩阵 A_σ 使得 $A = U_1A_\sigma V_1$. 于是 $A = U_1A_\sigma U_1^* \cdot U_1V_1 = U_1V_1 \cdot V_1^*A_\sigma V_1$. 分别令 $P = U_1A_\sigma U_1^*$, $Q = V_1^*A_\sigma V_1$ 和 $U = U_1V_1$ 即可. 证毕.

1.3 Hemite 矩阵

设 A 是 n 阶复方阵,若 A 的共轭转置等于它自身,则称 A 是 Hermite 矩阵,记作 $A\in H^n$. 若 A 半正定或正定,则记为 $A\in H_+^n$ 或 $A\in H_{++}^n$. 当 A 为实矩阵时,Hermite 矩阵对应于实对称矩阵. 由于 Hermite 矩阵的有关性质类似于实对称矩阵,这里仅罗列复矩阵有关结论而不加证明.

定理 1.3.1 设 $\boldsymbol{A}\in\boldsymbol{H}^n$. 则

(1) 对任意的 $\boldsymbol{x},\boldsymbol{y}\in\mathbf{C}^n$, $(\boldsymbol{Ax},\boldsymbol{y})=(\boldsymbol{x},\boldsymbol{Ay})$;

(2) $\boldsymbol{A}$ 的特征值是实数;

(3) $\boldsymbol{A}$ 的不同特征值的特征向量必正交.

定理 1.3.2 设 $\boldsymbol{A}\in\boldsymbol{H}^n$,则存在 $n\times n$ 阶酉矩阵 $\boldsymbol{U}$ 及 n 阶对角矩阵 $\boldsymbol{\Lambda}$,使得 $\boldsymbol{A}=\boldsymbol{U\Lambda U}^*$.

对于正定矩阵 $\boldsymbol{A}\in\boldsymbol{H}_{++}^n$,有以下结果.

定理 1.3.3 设 $\boldsymbol{A}\in\boldsymbol{H}_{++}^n$,则

(1) $\boldsymbol{A}$ 的特征值是正数;

(2) 存在可逆 Hermite 矩阵 $\boldsymbol{B}$ 使得 $\boldsymbol{A}=\boldsymbol{B}^2$;

(3) 存在可逆复方阵 $\boldsymbol{B}$ 使得 $\boldsymbol{A}=\boldsymbol{B}^*\boldsymbol{B}$;

(4) 对任意可逆复方阵 $\boldsymbol{P}$,矩阵 $\boldsymbol{P}^*\boldsymbol{AP}$ 正定;

(5) $\boldsymbol{A}$ 的所有主子矩阵正定.

对于半正定矩阵 $\boldsymbol{A}\in\boldsymbol{H}_{+}^n$,也有上述相似结果,在此不再赘述.

引理 1.3.1 设 $\boldsymbol{A}\in\boldsymbol{H}^n$,非零向量 $\boldsymbol{x}\in\mathbf{C}^n$,那么一定存在 $\boldsymbol{A}$ 的特征向量 $\boldsymbol{\xi}$ 使得它属于向量 $\boldsymbol{x},\boldsymbol{Ax},\boldsymbol{A}^2\boldsymbol{x},\cdots$ 所张成的子空间内.

证 由于向量组 $\boldsymbol{x},\boldsymbol{Ax},\boldsymbol{A}^2\boldsymbol{x},\cdots$ 是线性相关的,则存在最小自然数 $k(\leqslant n)$ 使得

$$\boldsymbol{A}^k\boldsymbol{x}+a_1\boldsymbol{A}^{k-1}\boldsymbol{x}+\cdots+a_k\boldsymbol{x}=\boldsymbol{0}.$$

将上式左边分解因式得到

$$(\boldsymbol{A}-\lambda_1\boldsymbol{I})\cdots(\boldsymbol{A}-\lambda_k\boldsymbol{I})\boldsymbol{x}=\boldsymbol{0}.$$

记 $\boldsymbol{\xi}=(\boldsymbol{A}-\lambda_2\boldsymbol{I})\cdots(\boldsymbol{A}-\lambda_k\boldsymbol{I})\boldsymbol{x}$,则 $\boldsymbol{\xi}\neq\boldsymbol{0}$ 且 $\boldsymbol{A\xi}=\lambda_1\boldsymbol{x}$. 证毕.

定理 1.3.4 设 $\boldsymbol{A},\boldsymbol{B}\in\boldsymbol{H}^n$,存在酉矩阵 $\boldsymbol{U}$ 使得 $\boldsymbol{U}^*\boldsymbol{AU}$ 和 $\boldsymbol{U}^*\boldsymbol{BU}$ 同为对角阵,当且仅当 $\boldsymbol{A},\boldsymbol{B}$ 乘积可交换,即 $\boldsymbol{AB}=\boldsymbol{BA}$.

证 这里只要证明必要性. 设 $\boldsymbol{\eta}_1$ 为 $\boldsymbol{B}$ 对应于 μ_1 的特征向量,即 $\boldsymbol{B\eta}_1=\mu_1\boldsymbol{\eta}_1$. 用 $\boldsymbol{A}^k$ 左乘两边,利用交换性得到

$$\boldsymbol{B}(\boldsymbol{A}^k\boldsymbol{\eta}_1)=\mu_1(\boldsymbol{A}^k\boldsymbol{\eta}_1),\quad k=1,2,\cdots.$$

由引理 1.3.1,存在 $\boldsymbol{A}$ 的特征向量 $\boldsymbol{\xi}_1$ 使得它属于向量 $\boldsymbol{\eta}_1,\boldsymbol{A\eta}_1,\boldsymbol{A}^2\boldsymbol{\eta}_1,\cdots$所张成的子空间内. 于是 $\boldsymbol{\xi}_1$ 是 $\boldsymbol{A},\boldsymbol{B}$ 的一个公共特征向量.

选择 $\boldsymbol{B}$ 的另外特征向量 $\boldsymbol{\eta}_2\perp\boldsymbol{\xi}_1$,重复上述过程得到 $\boldsymbol{\xi}_2$ 是 $\boldsymbol{A},\boldsymbol{B}$ 的公共特征向量,且 $\boldsymbol{\xi}_2\perp\boldsymbol{\xi}_1$,直到找完所有相互正交的公共特征向量 $\boldsymbol{\xi}_1,\cdots,\boldsymbol{\xi}_n$ 为止. 将其标准化得到酉矩阵 $\boldsymbol{U}$ 即可. 证毕.

定理 1.3.5 设 $\boldsymbol{A},\boldsymbol{B}\in\boldsymbol{H}^n$ 且 $\boldsymbol{B}$ 正定,则存在可逆矩阵 $\boldsymbol{P}\in\mathbf{C}^{n\times n}$ 使得

$$\boldsymbol{A}=\boldsymbol{P}^*\boldsymbol{\Lambda P},\quad \boldsymbol{B}=\boldsymbol{P}^*\boldsymbol{P},\tag{1.26}$$

这里 $\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$ 的对角元为矩阵 $\boldsymbol{AB}^{-1}$ 的特征值.

证 由于 $\boldsymbol{B}$ 正定,则 $\boldsymbol{B}^{-1}$ 也正定. 由于矩阵 $\boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{AB}^{-\frac{1}{2}}$ 仍然为 Hermite 矩阵,则存在酉矩阵 $\boldsymbol{U}$ 使得 $\boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{AB}^{-\frac{1}{2}}=\boldsymbol{U\Lambda U}^*$,$\boldsymbol{\Lambda}$ 为对角矩阵. 取 $\boldsymbol{P}=\boldsymbol{U}^*\boldsymbol{B}^{\frac{1}{2}}$ 即可. 注意到矩阵 $\boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{AB}^{-\frac{1}{2}}$ 与 $\boldsymbol{AB}^{-1}$ 相似,故 $\boldsymbol{\Lambda}$ 的对角元为矩阵 $\boldsymbol{AB}^{-1}$ 的特征值. 证毕.

定理 1.3.6 设 $A, B \in H_+^n$,则存在可逆矩阵 Q 使得 Q^*AQ 和 Q^*BQ 同为对角阵.

证 设 $A+B$ 的秩为 r. 则存在可逆矩阵 P 使得

$$P^*(A+B)P = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}. \tag{1.27}$$

由于 A 半正定,则 $P^*(A+B)P \geqslant P^*BP$. 将 P^*BP 进行类似于式(1.27)的分块,据 B 的半正定性知

$$P^*BP = \begin{bmatrix} M & O \\ O & O \end{bmatrix}.$$

应用定理 1.3.2 得到 $M = Q_1 \Lambda Q_1^*$,这里 Λ 为对角矩阵,Q_1 为酉矩阵. 取

$$Q = P\begin{bmatrix} Q_1 & O \\ O & I_{n-r} \end{bmatrix}$$

则可满足要求. 证毕.

定理 1.3.7(谱分解) 设 $A \in \mathbf{C}^{n\times n}$ 与对角矩阵相似,则 A 可以分解为一列幂等矩阵的乘积的加权和,即

$$A = \sum_{i=1}^{n} \lambda_i A_i, \tag{1.28}$$

其中 λ_i 为 A 的特征值,$A_i^2 = A_i$, $i = 1, 2, \cdots, n$.

证 由于 A 与对角矩阵相似,则存在可逆矩阵 P 使得

$$A = P\mathrm{diag}(\lambda_1, \cdots, \lambda_n)P^{-1}. \tag{1.29}$$

记 $P = (v_1, \cdots, v_n)$,$(P^{-1})^{\mathrm{T}} = (u_1, \cdots, u_n)$,将其代入到式(1.29)中得到

$$A = (v_1, \cdots, v_n)\mathrm{diag}(\lambda_1, \cdots, \lambda_n)\begin{bmatrix} u_1^{\mathrm{T}} \\ \vdots \\ u_n^{\mathrm{T}} \end{bmatrix} = \sum_{i=1}^{n} \lambda_i v_i u_i^{\mathrm{T}}. \tag{1.30}$$

令 $A_i = v_i u_i^{\mathrm{T}}$. 由 $P^{-1}P = I$,容易知道

$$A_i A_j = v_i u_i^{\mathrm{T}} v_j u_j^{\mathrm{T}} = (u_i^{\mathrm{T}} v_j) \cdot v_i u_j^{\mathrm{T}} = A_i \delta_{ij},$$

这里 δ_{ij} 是 Kronecker 记号. 于是 $A_i^2 = A_i$ 是幂等矩阵,且任意两个不同的 A_i 与 A_j 正交. 证毕.

特别地,对于 Hermite 矩阵,有下面的谱分解

$$A = \sum_{i=1}^{n} \lambda_i u_i u_i^{\mathrm{T}}, \tag{1.31}$$

其中,$\lambda_i (i = 1, 2, \cdots, n)$ 为 A 的特征值,$u_i (i = 1, 2, \cdots, n)$ 为相应的单位特征向量. 如果以 $\sqrt{\lambda_i} u_i$ 代替 u_i,则上述分解式可以省略 λ_i.

1.4 广义逆

20世纪20年代Moore利用投影矩阵定义了矩阵的广义逆，但是直到1955年Penrose利用四个矩阵方程给出了广义逆的新定义后，相关理论与应用才得到足够重视，并迅速在微分方程、数值线性代数、测量学、统计学、经济学和数学规划等领域得到广泛应用，其重要性逐渐为人们所认识. 近年来，广义逆矩阵理论和计算方法得到了迅速发展，已经成为矩阵理论研究和应用不可或缺的工具.

Penrose(1955)证明了如下命题：对任意矩阵$\boldsymbol{A}\in\mathbf{C}^{m\times n}$，存在唯一的矩阵$\boldsymbol{G}\in\mathbf{C}^{n\times m}$，同时满足四个条件：

$$\boldsymbol{AGA}=\boldsymbol{A};\quad \boldsymbol{GAG}=\boldsymbol{G};\quad (\boldsymbol{GA})^*=\boldsymbol{GA};\quad (\boldsymbol{AG})^*=\boldsymbol{AG}. \tag{1.32}$$

称同时满足式(1.32)四个条件的矩阵$\boldsymbol{G}$为$\boldsymbol{A}$的Moore-Penrose逆(M-P逆)，记作$\boldsymbol{A}^+$. 若$\boldsymbol{G}$仅仅满足第一个条件，则称$\boldsymbol{G}$为$\boldsymbol{A}$的广义逆，记作$\boldsymbol{A}^-$. 本书仅仅用到这样两种广义逆.

矩阵$\boldsymbol{A}$的值空间和零空间分别定义为

$$R(\boldsymbol{A})=\{\boldsymbol{Ax}\mid \boldsymbol{x}\in\mathbf{C}^n\},\quad N(\boldsymbol{A})=\{\boldsymbol{x}\in\mathbf{C}^n\mid \boldsymbol{Ax}=\boldsymbol{0}\}.$$

定理 1.4.1 设$\boldsymbol{A}\in\mathbf{C}^{m\times n}$. 则$\boldsymbol{G}\in\mathbf{C}^{n\times m}$为$\boldsymbol{A}$的广义逆的充要条件是，对任意的$\boldsymbol{b}\in R(\boldsymbol{A})$，有$\boldsymbol{AGb}=\boldsymbol{b}$.

证 充分性. 任取$\boldsymbol{u}\in\mathbf{C}^n$，则$\boldsymbol{b}=\boldsymbol{Au}\in R(\boldsymbol{A})$. 由$\boldsymbol{AGb}=\boldsymbol{b}$知

$$\boldsymbol{AGAu}=\boldsymbol{AGb}=\boldsymbol{b}=\boldsymbol{Au}.$$

由$\boldsymbol{u}$的任意性知$\boldsymbol{AGA}=\boldsymbol{A}$，即$\boldsymbol{G}$是$\boldsymbol{A}$的广义逆.

必要性. 任取$\boldsymbol{b}\in R(\boldsymbol{A})$，则存在$\boldsymbol{x}\in\mathbf{C}^n$使得$\boldsymbol{Ax}=\boldsymbol{b}$. 由$\boldsymbol{AGA}=\boldsymbol{A}$可得$\boldsymbol{AGb}=\boldsymbol{AGAx}=\boldsymbol{Ax}=\boldsymbol{b}$. 证毕.

定理 1.4.2 设$\boldsymbol{A}\in\mathbf{C}^{m\times n}$，$\boldsymbol{A}^-$为$\boldsymbol{A}$的给定广义逆. 则$\boldsymbol{A}$的任意广义逆$\boldsymbol{G}$都有如下形式：

(1) $\boldsymbol{G}=\boldsymbol{A}^-+\boldsymbol{U}-\boldsymbol{A}^-\boldsymbol{AUAA}^-$；

(2) $\boldsymbol{G}=\boldsymbol{A}^-+\boldsymbol{V}(\boldsymbol{I}_m-\boldsymbol{AA}^-)+(\boldsymbol{I}_n-\boldsymbol{A}^-\boldsymbol{A})\boldsymbol{U}$，

这里$\boldsymbol{U},\boldsymbol{V}\in\mathbf{C}^{n\times m}$为任意矩阵.

证 容易验证对于如上给出的两种形式的$\boldsymbol{G}$都满足$\boldsymbol{AGA}=\boldsymbol{A}$. 反之，若$\boldsymbol{G}$满足$\boldsymbol{AGA}=\boldsymbol{A}$，则

$$\boldsymbol{G}=\boldsymbol{A}^-+\boldsymbol{G}-\boldsymbol{A}^--\boldsymbol{A}^-\boldsymbol{A}(\boldsymbol{G}-\boldsymbol{A}^-)\boldsymbol{AA}^-.$$

取$\boldsymbol{U}=\boldsymbol{G}-\boldsymbol{A}^-$，则$\boldsymbol{G}$具有形式(1). 为证明(2)，只要令$\boldsymbol{V}=\boldsymbol{G}-\boldsymbol{A}^-$和$\boldsymbol{U}=\boldsymbol{GAA}^-$即可.

定理 1.4.3 设$\boldsymbol{A}\in\mathbf{C}^{m\times n}$，$\boldsymbol{A}^-$为$\boldsymbol{A}$的广义逆. 则

(1) $(\boldsymbol{A}^{\mathrm{T}})^-=(\boldsymbol{A}^-)^{\mathrm{T}}$，$(\boldsymbol{A}^*)^-=(\boldsymbol{A}^-)^*$；

(2) $(\boldsymbol{AA}^-)^2=\boldsymbol{AA}^-$，$(\boldsymbol{A}^-\boldsymbol{A})^2=\boldsymbol{A}^-\boldsymbol{A}$，$\mathrm{rank}(\boldsymbol{A})=\mathrm{rank}(\boldsymbol{AA}^-)=\mathrm{rank}(\boldsymbol{A}^-\boldsymbol{A})$；

(3) 若记$\lambda^-=\begin{cases}0, & \lambda=0;\\ \lambda^{-1}, & \lambda\neq 0,\end{cases}$则$\lambda^-\boldsymbol{A}^-$是$\lambda\boldsymbol{A}$的广义逆；