大学数学科学丛书　12

矩阵不等式

（第 二 版）

王松桂　吴密霞　贾忠贞　编著

科 学 出 版 社

北　京

内 容 简 介

本书系统地论述了矩阵论中的各种不等式. 全书共分九章. 第 1 章是矩阵论的预备知识; 第 2~8 章分别讨论了有关秩、行列式、特征值、条件数、迹、偏序和受控等方面的不等式; 第 9 章给出了矩阵不等式在线性统计中的几个应用; 最后两个附录收集了数量、函数和概率统计中常用的不等式.

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

图书在版编目(CIP)数据

矩阵不等式/王松桂, 吴密霞, 贾忠贞编著. —2 版. —北京: 科学出版社, 2006.5

(大学数学科学丛书; 12)

ISBN 7-03-016494-6

I.矩… II.①王… ②吴… ③贾… III.矩阵-不等式 IV.O151.25

中国版本图书馆 CIP 数据核字(2005)第 137711 号

责任编辑: 吕 虹 赵彦超／责任校对: 赵桂芬
责任印制: 安春生／封面设计: 王 浩

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

中国科学院印刷厂 印刷
科学出版社发行 各地新华书店经销

*

2006 年 5 月第 二 版 开本: B5(720×1000)
2006 年 5 月第一次印刷 印张: 18 1/4
印数:1—4 000 字数: 336 000

定价: 28.00 元

(如有印装质量问题, 我社负责调换〈环伟〉)

《大学数学科学丛书》序

按照恩格斯的说法，数学是研究现实世界中数量关系和空间形式的科学．从恩格斯那时到现在，尽管数学的内涵已经大大拓展了，人们对现实世界中的数量关系和空间形式的认识和理解已今非昔比，数学科学已构成包括纯粹数学及应用数学内含的众多分支学科和许多新兴交叉学科的庞大的科学体系，但恩格斯的这一说法仍然是对数学的一个中肯而又相对来说易于为公众了解和接受的概括，科学地反映了数学这一学科的内涵．正由于忽略了物质的具体形态和属性、纯粹从数量关系和空间形式的角度来研究现实世界，数学表现出高度抽象性和应用广泛性的特点，具有特殊的公共基础地位，其重要性得到普遍的认同．

整个数学的发展史是和人类物质文明和精神文明的发展史交融在一起的．作为一种先进的文化，数学不仅在人类文明的进程中一直起着积极的推动作用，而且是人类文明的一个重要的支柱．数学教育对于启迪心智、增进素质、提高全人类文明程度的必要性和重要性已得到空前普遍的重视．数学教育本质是一种素质教育；学习数学，不仅要学到许多重要的数学概念、方法和结论，更要着重领会到数学的精神实质和思想方法．在大学学习高等数学的阶段，更应该自觉地去意识并努力体现这一点．

作为面向大学本科生和研究生以及有关教师的教材，教学参考书或课外读物的系列，本丛书将努力贯彻加强基础、面向前沿、突出思想、关注应用和方便阅读的原则，力求为各专业的大学本科生或研究生（包括硕士生及博士生）走近数学科学、理解数学科学以及应用数学科学提供必要的指引和有力的帮助，并欢迎其中相当一些能被广大学校选用为教材，相信并希望在各方面的支持及帮助下，本丛书将会愈出愈好．

李大潜

2003 年 12 月 27 日

第二版前言

本书是矩阵不等式方面的第一部中文专著. 1994 年第一版出版以后, 深受广大读者的欢迎, 并远销海外, 成为数学工作者、高等院校有关专业教师和研究生乃至一般科技工作者手中的重要参考书或工具书.

为了更好地满足广大读者的需求, 在科学出版社的支持下, 我们对原书进行了修订. 这次修订除更正了第一版的一些错误外, 主要增加了近年来许多重要的新结果. 主要包括：Wielandt 不等式的矩阵形式 (7.7 节)；凸函数的矩阵不等式 (7.8 节)；Hadamard 乘积 (7.9 节)；Log- 弱受控不等式 (8.7 节) 以及一些关于矩阵行列式、迹和矩阵幂的迹的不等式 (分别添加在 3.2 节、6.3 节和 6.6 节适当部分), 总计增加了三十多个不等式. 这些增补内容的初稿, 由吴密霞博士完成, 最后由王松桂定稿.

尽管我们在进行修订时, 尽了最大的努力, 限于水平, 不当乃至谬误之处, 在所难免, 诚请广大读者指正.

作　者

2005 年 7 月

第一版序

正如作者在本书前言中所说, 撰写一部系统地、全面地论述矩阵论中各种不等式的专著, 是他们的一个夙愿. 经过十多个春秋的努力, 这个夙愿终于成了现实. 我作为他们的一名同事和同行, 特别是目睹了他们为创作这一专著而辛苦耕耘的过程, 感到由衷的高兴.

"十年辛苦不寻常", 这确不是一部泛泛之作. 首先, 作者之一王松桂教授自 20 世纪 70 年代后期起即从事线性统计的研究, 多年来在国内外刊物上发表了一系列的论文和专著. 矩阵不等式是进行这种研究的一个基本工具. 所以, 本书的写作有他多年研究工作的素养和经验作为背景. 其次, 作者 10 年来访问了欧美各国许多著名的学术研究中心, 与国际上线性统计领域内卓有成就的同行进行了广泛的合作和切磋, 收集了许多最新的、国内不易见到的材料, 这些都大大提高了本书的学术价值和实用价值.

本书收集的材料很多, 这是一个显著的优点. 因为在通常的矩阵论教本和专著中, 对矩阵不等式这个题材多未作系统介绍, 许多结果散布在大量文献中, 使用和查找不便. 如今有了本书, 一卷在手, 可免除许多查找翻检之劳. 可贵的是, 内容虽多, 但作者努力做到了多而不乱, 多而不偏. 众多的材料按其性质组织成一些专题, 重点突出, 系统性强. 作者注意了所选材料在应用上的意义, 把它作为取舍的一项重要标准, 相信凡是细读过本书的人, 都会有这种感觉.

作者之一王松桂教授是一位统计学家. 因此可以理解, 作者在材料选择上重视其在统计学上的应用, 这是本书的一个特色. 这个特色, 加上材料收集之丰富, 使本书必将成为线性统计方面的工作者案头必备的工具, 也是统计和相近专业的研究生、大学生有用的参考读物. 这当然不是说本书的读者仅限于这些人. 相反, 由于矩阵这个工具在数学、自然科学、工程技术乃至社会科学中的重要性, 众多的读者将会发现, 他们都能从本书中找到一些对自己工作有用的东西. 我想, 这也是作者自己的愿望.

陈希孺

第一版前言

关于不等式, 已经出版了若干部英文专著, 其中最有影响的是 Hardy，Littlewood 和 Polya(1934, 1953) 的 "Inequalities", Beckenback 和 Bellman(1961) 的 "Inequalities" 以及 Marshall 和 Olkin(1979) 的 "Inequalities：Theory of Majorization and its Applications". 这些书或以数量和函数的不等式为主要讨论对象, 或从某一特定方面研究一类数量或矩阵的不等式. 随着矩阵理论的迅速发展及其在自然科学、工程技术和社会经济等领域的广泛应用, 关于矩阵不等式的新结果层出不穷, 它们或是经典不等式的改进和推广, 或是完全新型的不等式, 或是应用的深入或拓广. 这些结果都散见在各种刊物或著作之中, 对理论研究者和使用矩阵工具的广大科技工作者带来诸多不便. 多年来, 作者有一个夙愿, 就是广泛收集、整理各种涉及矩阵的不等式, 撰写一部系统、全面地论述矩阵论中的各种不等式的专著. 目前在国内外尚未见有这样内容的著作出版.

自 1987 年以来, 我们就开始在教学实践、科学研究和广泛的国内外学术交流中收集资料, 特别在跟欧洲朋友们的学术磋商中, 获益匪浅. 随着工作的深入, 我们发现有关矩阵不等式的文献之多出乎预料, 于是我们不得不在资料筛选上花费很多精力, 同时我们又特别注意一些最新结果 (本书文献截止至 1992 年 10 月), 力求使本书尽可能反映这一方向的全貌.

全书共分九章, 矩阵论中的各种不等式按秩、行列式、特征值、条件数、迹、偏序和受控等内容分类, 每一类构成一章. 有些结果就其内容讲, 既可排在这一章, 也可排在另一章. 对这样的内容, 为读者查阅方便, 我们把它们排在两处, 但只在其中一处给出证明. 为便于读者使用并使本书内容大体上做到自成体系, 在第 1 章系统叙述了矩阵论的预备知识, 多数结果给出了证明. 为了显示矩阵不等式在线性统计中的广泛应用, 第 9 章举了几个例子. 限于本书的性质和篇幅, 当然举例不可能是全面的. 最后, 附录 1 和附录 2 罗列了常见的数量、函数不等式以及概率统计中的重要不等式, 以便于读者参考.

在这里我们要特别感谢很多的朋友. 他们是芬兰 Tampere 大学 Liski 博士和 Puntanen 博士, 瑞典 Umea 大学 Kulldorff 教授, 瑞士 Berne 大学 Riedwyl 教授, 波兰 Poznan 大学 Baksalary 教授, 英国 Manchester 大学 Farebrother 教授, 加拿大 Ottawa 大学邵军教授和 J. N. K. Rao 教授, 加拿大 Waterloo 大学吴建福教授, 美国 Chicago 大学刁锦寰教授以及 Colorado 州立大学 Srivastava 教授等. 本书许多资料是我们于 1988~1989 和 1990~1991 年先后访问这些地方时收集的, 这些朋友的热情帮助和支持对本书的写作来说是非常重要的. 这里还要提到第一作者的女

儿王晓京和王晓天, 感谢她们在美国读书期间帮助我们复印了一些国内不易找到的文献. 作者还要感谢严利清和杨亚宁同志为本书出版所提供的帮助. 我们还要借此机会, 向安徽教育出版社杨晓原同志表示谢意, 感谢他为本书的出版所给予的自始至终的热情支持、大力帮助和有效合作.

最后, 在本书出版之际, 作者还要对我们的老师陈希孺教授多年来的谆谆教诲、指导、关心和帮助表示诚挚的谢意.

本书前三章由贾忠贞执笔, 后六章由王松桂执笔, 最后由王松桂定稿.

限于作者的水平, 本书不妥乃至谬误之处在所难免, 恳请国内同行和广大读者不吝赐教.

王松桂　贾忠贞

符 号 表

A'	矩阵 A 的转置
$\bar{A}$	矩阵 A 的共轭
A^*	矩阵 A 的共轭转置 (即 $\bar{A}'$)
A^{-1}	矩阵 A 的逆矩阵
A^-	矩阵 A 的广义逆矩阵
A^+	矩阵 A 的 Moore-Penrose 广义逆
$A \geqslant 0$	表示 A 为半正定阵 (实对称阵或 Hermite 阵)
$A > 0$	表示 A 为正定阵 (实对称阵或 Hermite 阵)
$A \otimes B$	A 与 B 的 Kronecker 乘积
$A \circ B$	A 与 B 的 Hadamard 乘积
$\|A\|$	矩阵 A 的任一范数
$\|A\|_F$	矩阵 A 的欧氏范数, 即 Frobenius 范数
$\|A\|_2$	矩阵 A 的谱范数
$A^{(k)}$	矩阵 A 的 k 阶复合阵
$A^{1/2}$	半正定阵 A 的半正定平方根
$\boldsymbol{C}^n$	所有 n 维复向量的全体
$\det A$	方阵 A 的行列式
D	$\{x \in R^n : x_1 \geqslant \cdots \geqslant x_n\}$
$\mathrm{Im}(z)$	z 的虚部
$k(A)$	矩阵 A 的条件数
$\mathscr{M}(A)$	矩阵 A 的列向量张成的子空间
$r(A)$	矩阵 A 的秩
R^n	所有 n 维实向量的全体
R^n_+	$\{x \in R^n : x_i \geqslant 0, i = 1, \cdots, n\}$
$\mathrm{Re}(z)$	z 的实部
$\mathrm{tr}A$	矩阵 A 的迹
$x < y$	向量 x 受控于 y
$x <_\omega y$	向量 x 弱受控于 y
$\|X\|$	向量 X 的欧氏长度
$\bar{z}$	z 的共轭复数
$\|z\|$	z 的模
$\lambda(A)$	方阵 A 的特征值
$\sigma(A)$	矩阵 A 的奇异值

目　　录

第 1 章　矩阵论的预备知识

本书的目的是系统地论述有关矩阵的各种不等式. 因此, 当写作本书时, 假定读者已经具备了一般线性代数教科书中矩阵论的知识. 但是, 为了读者阅读上的方便和叙述简洁, 在这一章, 将扼要地给出本书讨论中要用到的一些重要结论和一般文献中不易查到的事实. 因为本章前六节内容偏重于基础, 为了节省篇幅, 略去了部分证明, 读者可从许以超 (1965)、蒋尔雄 (1978) 等找到它们的证明. 而对其余各节, 所有结论都给出了较详细的证明.

§1.1　线 性 空 间

我们用 C^n 和 R^n 分别表示全体 n 维复向量和实向量组成的线性空间.

设 S 为 C^n 的一个子空间, $a_1, \cdots, a_k$ 为 S 的一组基. 记 $A = (a_1, \cdots, a_k)$ 是一个 $n \times k$ 矩阵. 则 S 可以表为

$$S = \{x\colon\ x = At, t \in C^k\},$$

即 S 是由 $a_1, \cdots, a_k$ 的所有线性组合生成的线性子空间, 称为 A 的列向量张成的子空间, 简称为 A 的列空间, 记为 $\mathscr{M}(A)$. 若用 $\dim(S)$ 表示 S 的维数, 则从矩阵秩和空间维数的定义可以推出

$$\dim \mathscr{M}(A) = r(A), \tag{1.1.1}$$

这里 $r(A)$ 表示 A 的秩. 注意, S 是零维子空间当且仅当 $S = \{0\}$, 即单个零向量构成的子空间.

设 S_1 和 S_2 为两个子空间, 那么它们的和

$$S_1 + S_2 = \{x + y\colon\ x \in S_1, y \in S_2\}$$

以及它们的交

$$S_1 \cap S_2 = \{x\colon\ x \in S_1 \text{ 且 } x \in S_2\}$$

都是子空间, 分别称为 S_1 与 S_2 的和空间、交空间. 设 A 和 B 为两个具有相同行数的矩阵, 则容易证明

$$\mathscr{M}(A \vdots B) = \mathscr{M}(A) + \mathscr{M}(B). \tag{1.1.2}$$

下面的定理给出了和空间与交空间维数之间的关系.

定理 1.1.1　设 S_1 和 S_2 为两个子空间, 则

$$\dim(S_1+S_2)=\dim(S_1)+\dim(S_2)-\dim(S_1\cap S_2)\,. \tag{1.1.3}$$

证明　记 $p_i=\dim S_i,\ i=1,\ 2.\ r=\dim(S_1\cap S_2)$, 设 $a_1,\cdots,a_r$ 为 $S_1\cap S_2$ 的基, 将 $a_1,\cdots,a_r$ 扩充为 $a_1,\cdots,a_r,b_1,\cdots,b_{p_1-r}$, 使其构成 S_1 的一组基. 同样地, 将 $a_1,\cdots,a_r$ 扩充为 $a_1,\cdots,a_r,c_1,\cdots,c_{p_2-r}$, 使其为 S_2 的一组基. 那么, 向量组

$$a_1,\cdots,a_r,b_1,\cdots,b_{p_1-r},c_1,\cdots,c_{p_2-r}$$

构成了 S_1+S_2 的基, 这就证明了结论.

特别地, 若 $S_1\cap S_2=\{0\}$, 则称 S_1+S_2 为 S_1 与 S_2 的直和, 记为 $S_1\oplus S_2$, 直和具有下列性质.

定理 1.1.2　(1) $\dim(S_1\oplus S_2)=\dim S_1+\dim S_2$;

(2) S_1+S_2 为直和 $\Longleftrightarrow S_1+S_2$ 中的任一向量能唯一地表成 S_1 与 $\boldsymbol{S}_2$ 中向量之和.

对于 C^n 中的任意两个向量 a 和 b, 它们的内积 (a,b) 定义为

$$(a,b)=b^*a,$$

这里 b^* 表示向量 b 的转置共轭向量. 当 $a,b\in R^n$ 时, $(a,b)=b'a$. 定义了内积的线性空间 R^n 和 C^n 分别称为欧氏空间和酉空间.

在内积空间中, 向量 a 的长度定义为

$$\|a\|=(a,a)^{1/2}.$$

当 $(a,b)=0$ 时, 称 a 与 b 正交, 记为 $a\perp b$. 一组向量 $a_1,\cdots,a_m$, 若

$$(a_i,a_j)=\begin{cases}1,\ i=j,\\ 0,\ i\neq j,\end{cases}$$

则称 $a_1,\cdots,a_m$ 为标准正交向量组. 如果对一切 $b\in S$, 总有 $a\perp b$, 则称 a 与子空间 S 正交, 记为 $a\perp S$. 根据子空间的定义, 容易验证, 向量集合

$$S^{\perp}=\{x\colon\ x\perp S\}$$

也是一个子空间, 称为 S 的正交补空间.

正交补空间具有下列性质.

定理 1.1.3　(1) $S\oplus S^{\perp}=C^n$;

(2) $S = (S^{\perp})^{\perp}$;

(3) $S_1 \subset S_2 \Longleftrightarrow S_2^{\perp} \subset S_1^{\perp}$;

(4) $(S_1 + S_2)^{\perp} = S_1^{\perp} \cap S_2^{\perp}$;

(5) $(S_1 \cap S_2)^{\perp} = S_1^{\perp} + S_2^{\perp}$.

对于一个 $m \times k$ 矩阵 A, 若 $m \times t$ 矩阵 B 的秩为 t, 且满足 $\mathscr{M}(B) = \mathscr{M}(A)^{\perp}$, 则以后把 B 记为 $A^{\perp}$. 于是 $\mathscr{M}(A^{\perp}) = \mathscr{M}(A)^{\perp}$.

在本节一开头, 我们定义了矩阵 A 的列空间 $\mathscr{M}(A)$, 现在引进与 A 相联系的另一个子空间. 众所周知, 齐次线性方程组 $Ax = 0$ 的解的全体构成一个子空间, 我们称它为 A 的零空间或矩阵 A 的核, 记为 $N(A)$, 即

$$N(A) = \{x\colon\ Ax = 0\}.$$

若用 A^{-} 表示 $A_{n\times k}$ 的广义逆，即满足条件 $AXA = A$ 的任一矩阵 X(详见 §1.7), 则有

$$N(A) = \mathscr{M}(I_k - A^{-}A), \tag{1.1.4}$$

这里 I_k 表示 k 阶单位阵.

因为 $I - A^{-}A$ 和 $A^{-}A$ 都是幂等阵, 利用幂等阵的迹等于它的秩, 我们有 $r(I_k - A^{-}A) = k - \mathrm{tr}(A^{-}A) = k - r(A)$, 这里 $\mathrm{tr}(\cdot)$ 表示方阵的迹, 也就是对角线元素之和, 结合 (1.1.1), 我们有下面的结论.

定理 1.1.4 对任意 $n \times k$ 矩阵 A,

$$\dim \mathscr{M}(A) + \dim N(A) = k.$$

§1.2 特征值与特征向量

特征值与特征向量是矩阵论中两个重要的基本概念, 无论在理论研究或工程技术领域, 都有着广泛的应用.

定义 1.2.1 设 A 为 n 阶方阵, 若数 λ 和非零向量 x 满足

$$Ax = \lambda x, \tag{1.2.1}$$

则称 λ 为 A 的特征值, x 为 A 的对应于 λ 的特征向量.

线性方程组 (1.2.1) 可改写为 $(\lambda I_n - A)x = 0$. 此方程组有非零解当且仅当系数行列式 $\det(\lambda I_n - A) = 0$. 因此, λ 为 A 的特征值当且仅当 $\det(\lambda I_n - A) = 0$. 记 $\varphi(\lambda) = \det(\lambda I_n - A)$, 它是 λ 的 n 次多项式, 称为 A 的特征多项式. 于是, λ 为 A 的特征值当且仅当它是特征多项式 $\varphi(\lambda)$ 的根. 正是这个原因, 特征值也常常称为特征根.

对于特征多项式, 我们有下面的重要结论.

定理 1.2.1　设 n 阶方阵 $A=(a_{ij})$ 的特征多项式展开为

$$\varphi(\lambda)=\det(\lambda I_n-A)=\lambda^n+c_1\lambda^{n-1}+\cdots+c_{n-1}\lambda+c_n, \tag{1.2.2}$$

则它的系数

$$c_k=(-1)^k\sum_{1\leqslant i_1<\cdots<i_k\leqslant n}\det \mathrm{A}(i_1,\cdots,i_k),\quad k=1,\cdots,n,$$

这里 $A(i_1,\cdots,i_k)$ 表示由 A 的第 $i_1,\cdots,i_k$ 行、列元素组成的 k 阶子阵, 称为 A 的主子阵. 而和号 $\sum\limits_{1\leqslant i_1<\cdots<i_k\leqslant n}$ 表示对所有可能的 1 至 n 中的整数 $i_1,\cdots,i_k$ 求和, 特别

$$c_1=-\mathrm{tr}A, \tag{1.2.3}$$

$$c_n=-\det A, \tag{1.2.4}$$

这里 $\mathrm{tr}A=\sum\limits_{i=1}^{n}a_{ii}$, 称为 A 的迹.

记 $\lambda_1(A),\cdots,\lambda_n(A)$ 为 A 的 n 个特征值, 则

$$\varphi(\lambda)=\prod_{k=1}^{n}(\lambda-\lambda_k(A)). \tag{1.2.5}$$

比较 (1.2.2) 和 (1.2.5), 结合 (1.2.3) 和 (1.2.4), 我们得到

推论 1.2.1　对任意的 n 阶方阵 A, 有

$$\mathrm{tr}A=\sum_{k=1}^{n}\lambda_k(A),$$

$$\det A=\prod_{k=1}^{n}\lambda_k(A).$$

利用特征值的定义和特征多项式, 容易证明特征值的下述性质.

定理 1.2.2　设 A 为 n 阶方阵, 则

(1) A' 和 A 有相同的 (包括重数) 特征值;

(2) A^* 的特征值是 A 的特征值的共轭, 即

$$\lambda(A^*)=\bar{\lambda}(A),$$

这里 $\lambda(A)$ 表示 A 的特征值, 而 $\bar{\lambda}(A)$ 表示共轭复数;

(3) 若 A 为可逆阵, 则

$$\lambda(A^{-1}) = \frac{1}{\lambda(A)}.$$

进一步, 若 A 的特征值全为实数, 且排列为 $\lambda_1(A) \geqslant \cdots \geqslant \lambda_n(A)$, 则

$$\lambda_k(A^{-1}) = \frac{1}{\lambda_{n-k+1}(A)}, \quad k = 1, 2, \cdots, n.$$

在特征值的讨论中, 矩阵相似是一个重要的概念.

定义 1.2.2 设 A 和 B 为 $n \times n$ 阵. 若存在可逆阵 P, 使得

$$B = P^{-1}AP,$$

则称 B 和 A 相似, 记为 $B \sim A$.

容易验证, 矩阵相似具有下列性质：

(1) 自反性：$A \sim A$;

(2) 对称性：若 $A \sim B$, 则 $B \sim A$;

(3) 传递性：若 $A \sim B$, $B \sim C$, 则 $A \sim C$.

设 $B = P^{-1}AP$, 并且 $\varphi_B(\lambda)$ 表示 B 的特征多项式. 则

$$\begin{aligned}
\varphi_B(\lambda) &= \det(\lambda I_n - P^{-1}AP) = \det(P^{-1}(\lambda I_n - A)P) \\
&= \det P^{-1} \cdot \det(\lambda I_n - A) \cdot \det P \\
&= \det P^{-1}P \cdot \det(\lambda I_n - A) \\
&= \det(\lambda I_n - A) = \varphi_A(\lambda).
\end{aligned}$$

于是, 相似矩阵 B 与 A 有相同的特征多项式, 所以它们有相同的特征值.

定理 1.2.3 相似矩阵有相同的特征值.

推论 1.2.2 相似矩阵有相同的迹和行列式.

从理论的角度讲, 如果一个方阵能相似于对角阵, 则很容易求到它的特征值与特征向量. 设 A 为 n 阶方阵, 若 $\varPhi = (\varphi_1, \cdots, \varphi_n)$ 为可逆阵, 使得

$$\varPhi^{-1}A\varPhi = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix},$$

等价地

$$A\varPhi = \varPhi \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix},$$

也就是

$$A\varphi_k = \lambda_k \varphi_k, \quad k = 1, \cdots, n.$$

于是 $\lambda_1, \cdots, \lambda_n$ 就是 A 的特征值, $\varphi_1, \cdots, \varphi_n$ 就是对应的特征向量, 并且这些特征向量线性无关. 所以, 如果一个方阵 A 能够相似于一个对角阵, 则该对角阵的对角元就是 A 的特征值, A 相似于对角阵时所用的可逆阵的列向量就是 A 的 n 个线性无关的特征向量. 反过来, 若 A 有 n 个线性无关的特征向量, 则以它们为列向量的矩阵, 就能使 A 相似于对角阵. 于是, 我们有

定理 1.2.4　n 阶方阵 A 相似于对角阵, 当且仅当 A 有 n 个线性无关的特征向量.

因为属于不同特征值的特征向量总是线性无关的, 所以

推论 1.2.3　若方阵 A 的特征值互不相同, 则 A 一定相似于对角阵.

众所周知, 并不是每个方阵一定能够相似于对角阵. 能够相似于对角阵的最重要的矩阵类是实对称阵和 Hermite 阵, 这将在下面两节详细讨论, 这里我们讨论一般情况.

设 λ 为 n 阶方阵 A 的一个特征值, 任何满足 $Ax = \lambda x$ 的非零向量 x, 都是 A 的对应于 λ 的特征向量. 易见, A 的对应于 λ 的全体特征向量是线性方程组 $(\lambda I_n - A)x = 0$ 的全部非零解. 从线性方程组解的理论知, 如果添上零向量, 这些解就构成一个线性子空间, 记为 S_λ, 称为 A 的对应于特征值 λ 的特征子空间, 根据上节引进的零空间的定义, 有

$$S_\lambda = N(\lambda I_n - A).$$

特征子空间 S_λ 的维数称为特征值 λ 的几何重数, 因此几何重数就是对应于 λ 的所有特征向量中线性无关的最大个数. 如果一个 n 阶方阵 A, 它的所有特征值的几何重数之和为 n, 则它一定可以相似于对角阵, 遗憾的是, 这个假设并不总是成立的.

若 λ 为 A 的 m 重特征值, 也就是说, 它是特征多项式的 m 重根, 则称 m 为 A 的代数重数. 关于代数重数和几何重数, 我们有如下定理.

定理 1.2.5　对任一特征值, 几何重数总不超过代数重数.

证明　设 A 为 n 阶方阵, λ_0 为其任一特征值. 记它的几何重数为 g, 代数重数为 m, 则存在 g 个对应于 λ_0 的线性无关的特征向量 $\varphi_1, \varphi_2, \cdots, \varphi_g$, 将其扩充为 C^n 的一组基

$$\varphi_1, \cdots, \varphi_q, \varphi_{g+1}, \cdots, \varphi_n. \tag{1.2.6}$$

记 $\Phi = (\varphi_1, \cdots, \varphi_g, \varphi_{g+1}, \cdots, \varphi_n)$, 则

$$A\Phi = (\lambda_0 \varphi_1, \cdots, \lambda_0 \varphi_g, A\varphi_{g+1}, \cdots, A\varphi_n).$$

因为向量组 (1.2.6) 是 C^n 的基. 于是 $A\varphi_{g+1}, \cdots, A\varphi_n$ 均可由它们线性表出.

设

$$A\varphi_j=\sum_{k=1}^{n}b_{kj}\varphi_k,\quad j=g+1,\cdots,n,$$

并记

$$B=\begin{pmatrix}\lambda_0 I_g & B_1\\ 0 & B_2\end{pmatrix},$$

这里 $B_1=(b_{jk}),\ j=1,\cdots,g;k=g+1,\cdots,n$. 而 $B_2=(b_{jk}),j=g+1,\cdots,n;k=g+1,\cdots,n$, 则有 $A\varPhi=\varPhi B$, 也就是 $A\sim B$. 根据定理 1.2.3 前面的证明, A 与 B 有相同的特征多项式, 但 B 的特征多项式

$$\varphi_B(\lambda)=\det(\lambda I_n-B)=(\lambda-\lambda_0)^g\det(\lambda I_{n-g}-B_2).$$

于是, λ_0 的代数重数 m 至少是 g, 这就证明了 $g\leqslant m$. 证毕.

因为一个 n 阶方阵的所有特征值的代数重数之和总是等于 n, 所以综合定理 1.2.5 和前面的讨论, 我们证明了下面的定理.

定理 1.2.6　方阵 A 相似于对角阵, 当且仅当它的每一个特征值的几何重数总是等于它的代数重数.

在后面的讨论中, 我们常常要碰到两个矩阵乘积的特征值问题, 关于这一方面, 我们有下面的重要定理.

定理 1.2.7　(1) 设 A 和 B 分别为 $m\times n$ 和 $n\times m$ 阵, 则 AB 与 BA 有相同的非零特征值 (包括重数);

(2) 设 A、B 为 n 阶方阵, 且至少一个可逆, 则 AB 与 BA 相似.

证明　(1) 设 $\lambda\neq 0$, 则

$$\begin{pmatrix}I_m & -A\\ 0 & I_n\end{pmatrix}\begin{pmatrix}\lambda I_m & A\\ B & I_n\end{pmatrix}=\begin{pmatrix}\lambda I_m-AB & 0\\ B & I_n\end{pmatrix},$$

$$\begin{pmatrix}\lambda I_m & A\\ B & I_n\end{pmatrix}\begin{pmatrix}I_m & -\dfrac{1}{\lambda}A\\ 0 & I_n\end{pmatrix}=\begin{pmatrix}\lambda I_m & 0\\ B & I_n-\dfrac{1}{\lambda}BA\end{pmatrix}.$$

对上面两式两边分别取行列式, 得到

$$\det(\lambda I_m-AB)=\lambda^m\det\left(I_n-\frac{1}{\lambda}BA\right)=\lambda^{m-n}\det(\lambda I_n-BA).\tag{1.2.7}$$

得证

(2) 是下面事实的直接结果.

$$AB=\begin{cases}A(BA)A^{-1}, & 若A可逆,\\ B^{-1}(BA)B, & 若B可逆.\end{cases}$$

在 (1.2.7) 中, 命 $\lambda = 1$, 便得到下面的事实.

推论 1.2.4　设 A 和 B 分别为 $m \times n$ 和 $n \times m$ 矩阵, 则

$$\det(I_m - AB) = \det(I_n - BA).$$

§1.3　实对称阵

设 $A = (a_{ij})$ 的元素 a_{ij} 全是实数, 且 $A' = A$, 则称 A 为实对称阵. 实对称阵具有许多重要性质.

定理 1.3.1　设 A 为 n 阶实对称阵. 则

(1) A 的所有特征值都是实数;

(2) 存在正交阵 $\varPhi$, 使得

$$\varPhi' A \varPhi = \mathrm{diag}(\lambda_1, \cdots, \lambda_n), \tag{1.3.1}$$

即实对称阵一定正交相似于对角阵.

注 1　从 (1.3.1) 知

$$A \sim \mathrm{diag}(\lambda_1, \cdots, \lambda_n),$$

所以, 由上节的讨论知, $\lambda_1, \cdots, \lambda_n$ 为 A 的特征值.

记 $\varPhi = (\varphi_1, \cdots, \varphi_n)$, 从 (1.3.1) 可推出

$$A\varPhi = \varPhi \mathrm{diag}(\lambda_1, \cdots, \lambda_n),$$

也就是 $A\varphi_j = \lambda_j \varphi_j$, 所以 $\varphi_1, \cdots, \varphi_n$ 为 A 的 n 个标准正交化特征向量.

注 2　记 $\varLambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$, 则 (1.3.1) 可改写为

$$A = \varPhi \varLambda \varPhi' = \sum_{j=1}^{n} \lambda_j \varphi_j \varphi_j', \tag{1.3.2}$$

此式称为 A 的谱分解.

从 (1.3.1), 我们很容易得到下面的推论.

推论 1.3.1　设 A 为 n 阶实对称阵, x 为 $n \times 1$ 实向量, 则存在正交阵 $\varPhi$, 使得正交变换 $y = \varPhi' x$ 把二次型 $x'Ax$ 化为平方和

$$x'Ax = y'\varLambda y = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2, \tag{1.3.3}$$

这里 $\varLambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$.

在实对称阵中, 半正定阵是一个特别重要的矩阵类.

设 A 为 n 阶实对称阵, 若对任意 $x \in R^n, x'Ax \geqslant 0$, 则称 A 为半正定阵, 记为 $A \geqslant 0$. 若进一步, $x'Ax = 0$ 当且仅当 $x = 0$, 则称 A 为正定阵. 记为 $A > 0$. 因此在以后的讨论中, 若无特殊声明, 当我们使用记号 $A \geqslant 0$ 或 $A > 0$ 时, 总假定 A 是对称的. 下面的两个定理分别表征了半正定阵和正定阵, 它们是非常有用的.

定理 1.3.2 设 A 为 n 阶实对称阵, 则 $A \geqslant 0$ 当且仅当下列条件之一成立.

(1) A 的所有特征值都是非负的;

(2) 存在对称阵 B, 使得 $A = B^2$;

(3) 存在 $t \times n$ 矩阵 B, 这里 $t = r(A)$, 使得 $A = B'B$;

(4) 对任意矩阵 P, $P'AP \geqslant 0$;

(5) A 的所有主子式为非负, 这里主子式定义为主子阵 $A(i_1, \cdots, i_k)$ 的行列式, 也就是 A 的位于第 $i_1, \cdots, i_k$ 行、列交叉处元素组成的 k 阶方阵的行列式.

注 3 对半正定阵 A, 满足 (2) 的矩阵 B 是很多的. 下面构造出一个半正定的 B. 设 $A \geqslant 0, \lambda_1, \cdots, \lambda_n$ 为其特征值. $\varphi_1, \cdots, \varphi_n$ 为其对应的特征向量, 且 $\varPhi = (\varphi_1, \cdots, \varphi_n)$ 为正交阵, 则由定理 1.3.1 之 (2). 有

$$
\begin{aligned}
A &= \varPhi \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \varPhi' \end{pmatrix} \varPhi' \\
&= \varPhi \begin{pmatrix} \lambda_1^{1/2} & & \\ & \ddots & \\ & & \lambda_n^{1/2} \end{pmatrix} \varPhi' \varPhi \begin{pmatrix} \lambda_1^{1/2} & & \\ & \ddots & \\ & & \lambda_n^{1/2} \end{pmatrix} \varPhi' \\
&= B^2,
\end{aligned}
$$

其中 $B = \varPhi \begin{pmatrix} \lambda_1^{1/2} & & \\ & \ddots & \\ & & \lambda_n^{1/2} \end{pmatrix} \varPhi' \geqslant 0.$

以后我们把 B 记为 $A^{1/2}$, 称为 A 的平方根.

定理 1.3.3 设 A 为 n 阶实对称阵, 则 $A > 0$ 当且仅当下列条件之一成立.

(1) A 的所有特征值都是正数;

(2) 存在可逆对称阵 B, 使得 $A = B^2$;

(3) 存在可逆阵 B, 使得 $A = B'B$;

(4) 对任意可逆阵 $P, P'AP > 0$;

(5) A 的所有主子式为正数;

(6) A 的所有顺序主子式为正数, 这里 k 阶顺序主子式为主子阵 $A(1, 2, \cdots, k)$ 的行列式.

在本节的剩余部分, 我们讨论两个实对称阵同时对角化的问题.

定理 1.3.4　设 A, B 为两个 n 阶实对称阵, 且 $B > 0$. 则存在可逆阵 Q, 使得

$$A = Q'\Lambda Q, \tag{1.3.4}$$

$$B = Q'Q, \tag{1.3.5}$$

这里 $\Lambda = \operatorname{diag}(\lambda_1, \cdots, \lambda_n), \lambda_j (j = 1, \cdots, n)$ 为 AB^{-1} 的特征值.

证明　因 $B > 0$, 故 $B^{-1} > 0$, 记 $B^{-1/2} = (B^{-1})^{1/2} > 0$. 因为 $B^{-1/2}AB^{-1/2}$ 仍为实对称阵, 依定理 1.3.1, 存在正交阵 Φ, 使得 $\Phi'(B^{-1/2}AB^{-1/2})\Phi = \Lambda = \operatorname{diag}(\lambda_1, \cdots, \lambda_n)$. 记 $Q = \Phi' B^{1/2}$, 即得 (1.3.4) 和 (1.3.5), 这里 $\lambda_1, \cdots, \lambda_n$ 为 $B^{-1/2}AB^{-1/2}$ 的特征值. 根据定理 1.2.7(2), $B^{-1/2}AB^{-1/2}$ 与 AB^{-1} 相似, 所以, 它们有相同的特征值, 定理得证.

为了证明下面一个重要定理, 我们需要如下引理.

引理 1.3.1　设 A 为 n 阶实对称阵, x 为任一 n 维非零向量. 则一定存在 A 的特征向量 φ, 使得

$$\varphi \in \mathscr{M}(x, Ax, A^2x, \cdots). \tag{1.3.6}$$

证明　因为向量组 $\{x, Ax, A^2x, \cdots\}$ 总是线性相关的, 故一定存在 k, 使得

$$A^k x + c_{k-1}A^{k-1}x + \cdots + c_0 x = 0. \tag{1.3.7}$$

假设 (1.3.7) 中的 k 是具有这一性质的最小 k.

记

$$f(y) = y^k + c_{k-1}y^{k-1} + \cdots + c_0.$$

依代数学基本定理, $f(y) = 0$ 有 k 个根 $\mu_1, \cdots, \mu_k$. 于是 $f(y)$ 可分解为

$$f(y) = (y - \mu_1)(y - \mu_2)\cdots(y - \mu_k).$$

用 $f(A)$ 表示将上式中 y 换上 A, μ_i 换上 $\mu_i I_n$ 所得到的矩阵, 则 (1.3.7) 变形为

$$\begin{aligned} f(A)x &= (A - \mu_1 I)(A - \mu_2 I)\cdots(A - \mu_k I)x \\ &= (A - \mu_1 I)\varphi = 0, \end{aligned} \tag{1.3.8}$$

其中

$$\varphi = (A - \mu_2 I)\cdots(A - \mu_k I)x \neq 0. \tag{1.3.9}$$

注意, $\varphi \neq 0$ 是由 k 的最小性推出的. (1.3.8) 表明, $A\varphi = \mu_1\varphi$, 即 φ 是 A 的对应于特征值 μ_1 的特征向量. 从 (1.3.9) 得

$$\varphi = A^{k-1}x + \left(-\sum_{j=2}^{k}\mu_j\right)A^{k-2}x + \cdots + (-1)^{k-1}\prod_{j=2}^{k}\mu_j x.$$

这就证明了

$$\varphi \in \mathscr{M}(A^{k-1}x, A^{k-2}x, \cdots, x).$$

引理得证.

定理 1.3.5　设 A, B 为 n 阶实对称阵, 则存在正交阵 Q, 使得 $Q'AQ$ 与 $Q'BQ$ 为对角阵, 当且仅当 $AB = BA$.

证明　必要性的证明是容易的, 下面证明充分性. 设 φ_1 为 B 的对应于特征值 λ 的特征向量, 于是, $B\varphi_1 = \lambda\varphi_1$, 用 A^k 左乘两边, 利用 $AB = BA$ 得

$$A^k B\varphi_1 = \lambda A^k \varphi_1 = BA^k\varphi_1, \quad k = 1, 2, \cdots,$$

于是

$$B(A^k\varphi_1) = \lambda(A^k\varphi_1), \quad k = 1, 2, \cdots.$$

这表明：$\varphi_1, A\varphi_1, A^2\varphi_1, \cdots$ 都是 B 的对应于同一个特征值 λ 的特征向量. 依引理 1.3.1, 存在 A 的一个特征向量 q_1, 使得

$$q_1 \in \mathscr{M}(\varphi_1, A\varphi_1, A^2\varphi_1, \cdots).$$

因为 $\mathscr{M}(\varphi_1, A\varphi_1, A^2\varphi_1, \cdots)$ 是 B 的一个特征子空间, 于是 q_1 是 A, B 的一个公共特征向量.

选择 B 的另一个特征向量 φ_2, 且满足 $\varphi_2 \perp q_1$. 用 φ_2 代替 φ_1, 重复上面的讨论, 我们可以找到 A 与 B 的另一个公共特征向量 q_2, 且 $q_2 \perp q_1$. 继续这个过程, 我们可以找到 A 与 B 的 n 个公共特征向量 $q_1, \cdots, q_n$ 满足 $q_j' q_k = 0$. 不妨假设它们的长度都等于 1, 于是 $Q = (q_1, \cdots, q_n)$ 就是所求的正交阵. 定理得证.

注 4　从本定理知, 可交换的两个 n 阶实对称阵具有 n 个两两正交的公共特征向量, 但它们的特征值不必相同.

定理 1.3.6　设 A, B 为两个 n 阶实半正定阵, 则存在可逆阵 Q, 使得 $Q'AQ$ 和 $Q'BQ$ 皆为对角阵.

证明　设 $t = r(A + B)$. 则存在可逆阵 P, 使得

$$P'(A+B)P = \begin{pmatrix} I_t & 0 \\ 0 & 0 \end{pmatrix}. \tag{1.3.10}$$

将 $P'BP$ 分块为

$$P'BP = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}, \tag{1.3.11}$$

其中 M_{11} 为 $t \times t$ 矩阵, 因为 $P'(A+B)P \geqslant P'BP$(这里 $C \geqslant D$ 表示 $C - D \geqslant 0$), 于是从 (1.3.10) 和 (1.3.11) 可以推出

$$M_{ij} = 0, \qquad \text{当 } i + j \geqslant 3 \text{ 时},$$

即

$$P'BP = \begin{pmatrix} M_{11} & 0 \\ 0 & 0 \end{pmatrix}. \tag{1.3.12}$$

注意到 M_{11} 仍为对称阵, 应用定理 1.3.1(2) 知, 存在正交阵 Q_1, 使得

$$Q_1^{-}M_{11}Q_1 = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_t \end{pmatrix}.$$

记

$$Q = P\begin{pmatrix} Q_1 & 0 \\ 0 & I_{n-t} \end{pmatrix},$$

则 Q 为可逆阵, 且有

$$\begin{aligned} Q'BQ &= \mathrm{diag}(\lambda_1, \cdots, \lambda_t, 0, \cdots, 0), \\ Q'AQ &= \mathrm{diag}(1-\lambda_1, \cdots, 1-\lambda_t, 0, \cdots, 0). \end{aligned}$$

定理证毕.

§1.4　Hermite 阵

设 $A = a_{ij}$ 为 n 阶方阵, 记 $A^* = \bar{A}'$, 即取共轭同时又转置. 若 $A^* = A$, 则称 A 是一个 Hermite 阵. 当 A 为实矩阵时, Hermite 阵就是实对称阵.

Hermite 阵具有许多类似于实对称阵的重要性质.

定理 1.4.1　设 A 为 n 阶 Hermite 阵, 则

(1) A 的所有特征值都是实数;

(2) 存在一个酉阵 U, 即 U 满足 $U^*U = I_n$, 使得

$$U^*AU = \mathrm{diag}(\lambda_1, \cdots, \lambda_n), \tag{1.4.1}$$

即 Hermite 阵一定酉相似于对角阵.

注 1　从 (1.4.1) 知

$$A \sim \mathrm{diag}(\lambda_1, \cdots, \lambda_n),$$

所以, $\lambda_1, \cdots, \lambda_n$ 为 A 的特征值, 若记 $U = (u_1, \cdots, u_n)$, 则从 (1.4.1) 可推出

$$AU = U\mathrm{diag}(\lambda_1, \cdots, \lambda_n),$$

即 $Au_j = \lambda_j u_j, j = 1, \cdots, n$. 所以, $u_1, \cdots, u_n$ 为 A 的 n 个标准正交化特征向量.

注 2　记 $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$, 则 (1.4.1) 可改写为

$$A = U\Lambda U^* = \sum_{j=1}^{n} \lambda_j u_j u_j^*,$$

称为 A 的谱分解.

由 (1.4.1) 容易得到如下推论.

推论 1.4.1　设 A 为 n 阶 Hermite 阵, x 为 $n \times 1$ 向量. 那么存在酉阵 U, 使得酉变换 $y = U^*x$ 把二次型 x^*Ax 化为平方和

$$x^*Ax = y^*\Lambda y = \lambda_1|y_1|^2 + \cdots + \lambda_n|y_n|^2,$$

其中 $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$.

因此, 完全类似于上节, 我们可以定义正定和半正定的 Hermite 阵.

设 A 为 n 阶 Hermite 阵, 若对任意 $x \in C^n, x^*Ax \geqslant 0$, 则称 A 为半正定的, 记为 $A \geqslant 0$. 若进一步, $x^*Ax = 0$ 当且仅当 $x = 0$, 则称 A 为正定的, 记为 $A > 0$.

类似于定理 1.3.2 和 1.3.3, 我们有

定理 1.4.2　设 A 为 n 阶 Hermite 阵, 则 $A \geqslant 0$ 当且仅当下列条件之一成立.

(1) A 的所有特征值为非负;

(2) 存在一个 Hermite 阵 B, 使得 $A = B^2$;

(3) 存在 $t \times n$ 矩阵 B, 其中 $t = r(A)$, 使得 $A = B^*B$;

(4) 对任一复方阵 P, $P^*AP \geqslant 0$;

(5) A 的所有主子式为非负.

定理 1.4.3　设 A 为 n 阶 Hermite 阵, 则 $A > 0$ 当且仅当下列条件之一成立.

(1) A 的所有特征值为正数;

(2) 存在一个可逆 Hermite 阵 B, 使得 $A = B^2$;

(3) 存在可逆复方阵 B, 使得 $A = B^*B$;

(4) 对任一可逆复方阵 P, $P^*AP > 0$;

(5) A 的所有主子式为正数;

(6) A 的所有顺序主子式为正数.

注 3　对半正定的 Hermite 阵 A, 也存在着半正定的平方根阵 $A^{1/2}$, 使得 $(A^{1/2})^2 = A$. 事实上, 根据定理 1.4.1, 存在酉阵 U, 使得 $A = U\Lambda U^*$, 这里 $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n), \lambda_i \geqslant 0, i = 1, \cdots, n$ 为 A 的特征值, 记 $\Lambda^{1/2} = \mathrm{diag}(\lambda_1^{1/2}, \cdots, \lambda_n^{1/2})$, 则 A 可以改写为

$$A = U\Lambda^{1/2}U^*U\Lambda^{1/2}U^* = B^2,$$

其中 $B = U\Lambda^{1/2}U^* \geqslant 0$, 就是 A 的平方根阵 $A^{1/2}$. 显然当 $A > 0$ 时, $B = A^{1/2} > 0$.

类似于定理 1.3.4~1.3.6, 我们有如下几个定理. 上节的证明过程照搬过来, 将其中的“正交阵”、“实对称阵”和“Q'”分别改为“酉阵”、“Hermite 阵”和“Q^*”.

定理 1.4.4　设 A, B 为两个 n 阶 Hermite 阵, 且 $B > 0$. 则存在可逆阵 Q, 使得

$$A = Q^* \Lambda Q,$$
$$B = Q^* Q,$$

其中 $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n), \lambda_j (j = 1, \cdots, n)$ 为 AB^{-1} 的特征值.

定理 1.4.5　设 A, B 为两个 n 阶 Hermite 阵. 则存在酉阵 U, 使得 U^*AU 和 U^*BU 为对角阵, 当且仅当 $AB = BA$.

定理 1.4.6　设 A, B 为两个同阶半正定 Hermite 阵, 则存在可逆阵 Q, 使得 Q^*AQ 与 Q^*BQ 皆为对角阵.

§1.5　矩 阵 分 解

所谓矩阵分解, 就是将一个矩阵写成从某种意义上讲比较简单或对它的性质比较熟悉的若干矩阵的乘积. 例如, 在 §1.3 中, 实对称阵 A 被分解为 $A = \Phi \Lambda \Phi'$, 这里 $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$ 是对角阵, 有比较简单的形式, 而 Φ 是一个正交阵, 具有许多重要性质. 其中最重要的一条性质是不改变向量长度, 也就是 $\|\Phi x\| = \|x\|$. 实对称阵的这种分解, 对研究实对称阵的性质带来很大方便.

本节我们叙述一般矩阵的几种重要分解.

定理 1.5.1 (Schmidt 三角化分解)　设 A 为 $m \times n$ 复矩阵, $r(A) = n$. 则存在 $n \times n$ 上三角阵 R 和 $m \times n$ 的矩阵 Q, $Q^*Q = I_n$, 使得 $A = QR$.

证明　设 $A = (a_1, \cdots, a_n)$, 因为 $a_1, \cdots, a_n$ 线性无关, 对它们施用 Schmidt 正交化程序, 即得所要结论.

定理 1.5.2 (秩分解)　设 A 为 $m \times n$ 复方阵, 则存在两个可逆阵 $P_{m \times m}$ 和 $Q_{n \times n}$, 使得

$$A = P \begin{pmatrix} I_t & 0 \\ 0 & 0 \end{pmatrix} Q, \tag{1.5.1}$$

这里 $t = r(A)$.

证明要点如下. 矩阵 A 经过行向量的如下三种初等变换：

(1) 交换两行的位置;

(2) 用一个非零数去乘某一行;

(3) 将某一行乘一个数加到另一行,

可化为行约简梯形, 成为满足下列四条的矩阵：

(1) 非零行的第一个非零元素为 1, 以下称为主导元素 1;

(2) 各列中, 若含有主导元素 1, 则其余元素全为零;

(3) 所有元素为零的行排在矩阵的下部;

(4) 主导元素 1 排成阶梯形, 即若 t 个行是非零行, 且第 j 行的主导元素 1 位于第 k_i 列, $j=1,\cdots,t$, 则 $k_1<k_2<\cdots<k_t$.

因为对矩阵 A 施以一系列三种行初等变换相当于用一个可逆阵左乘矩阵A, 于是上面我们证明了, 存在可逆阵 B, 使得 BA 成行约简梯形. 对行约简梯形 BA 再施以列的三种初等变换, 这又相当于用一个可逆阵 C 右乘 BA, 此时 BA 化为

$$BAC=\begin{pmatrix} I_t & 0, \\ 0 & 0 \end{pmatrix},$$

因为 B, C 皆可逆, 取 $P=B^{-1}, Q=C^{-1}$, 得到 (1.5.1).

注 1 秩分解 (1.5.1) 也称为 A 的等价标准形.

定理 1.5.3 (满秩分解) 设 A 为 $m\times n$ 复阵, $r(A)=t$. 则存在 $m\times t$ 和 $t\times n$ 且秩为 t 的矩阵 B 和 C, 使得

$$A=BC. \tag{1.5.2}$$

证明 应用 A 的秩分解 (1.5.1), 将 P 和 Q 分块为

$$P=(P_1 \vdots P_2), \quad Q=\begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix},$$

这里 P_1 和 Q_1 分别为 $m\times t$ 和 $t\times n$ 矩阵, 则 $A=P_1Q_1$. 明所欲证.

定理 1.5.4 (奇异值分解) 设 A 为 $m\times n$ 秩为 t 的复矩阵, 则存在两个酉阵 $U_{m\times m}$ 和 $V_{n\times n}$, 使得

$$A=U\begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}V^*, \tag{1.5.3}$$

其中 $\Sigma=\mathrm{diag}(\sigma_1,\cdots,\sigma_t), \sigma_i>0, i=1,\cdots,t$. $\lambda_1=\sigma_1^2,\cdots,\lambda_t=\sigma_t^2$ 为 A^*A 的非零特征值.

证明 设 $u_1,\cdots,u_t$ 为 A^*A 的对应于非零特征值 $\sigma_1^2,\cdots,\sigma_t^2$ 的标准正交化特征向量. 记 $v_j=\dfrac{1}{\sigma_j}A^*u_j, j=1,\cdots,t$. 容易验证, $v_1,\cdots,v_t$ 是标准正交化的.

将 $u_1,\cdots,u_t$ 和 $v_1,\cdots,v_t$ 分别扩充为 C^m 和 C^n 的标准正交基

$$u_1,\cdots,u_t,u_{t+1},\cdots,u_m; \tag{1.5.4}$$

$$v_1,\cdots,v_t,v_{t+1},\cdots,v_n, \tag{1.5.5}$$

并记以向量组 (1.5.4) 和 (1.5.5) 分别为列向量的矩阵为 U 和 V, 则 $U^*U=I_m, V^*V$

$= I_n$, 且

$$\begin{aligned} A &= UU^*A \\ &= (u_1u_1^* + \cdots + u_tu_t^*)A \\ &= \sigma_1 u_1 v_1^* + \cdots + \sigma_t u_t v_t^* \\ &= U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^*. \end{aligned}$$

证毕.

注 2 $\sigma_1 = \lambda_1^{1/2}, \cdots, \sigma_t = \lambda_t^{1/2}$ 称为 A 的奇异值. 有时, 为方便计, 也常把 A^*A 的所有特征值的平方根 $\lambda_1^{1/2}, \cdots, \lambda_n^{1/2}$ 称为 A 的奇异值. 若 $r(A) = t$, 则后面的 $n-t$ 个奇异值 $\lambda_{++1}^{1/2} = \cdots = \lambda_n^{1/2} = 0$.

注 3 由奇异值分解 (1.5.3) 可推得

$$AA^* = U \begin{pmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{pmatrix} U^*,$$

$$A^*A = V \begin{pmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{pmatrix} V^*.$$

这表明 U 的 m 个列向量是 AA^* 的标准正交化特征向量, 而 V 的 n 个列向量是 A^*A 的标准正交化特征向量.

特别地, 当 $AA^* = A^*A$ 时, 我们可以选取 $U = V$.

设 A 为复方阵, 若 $AA^* = A^*A$, 则称 A 为复规范阵. 根据上面的讨论, 对复规范阵 A, 奇异值分解 (1.5.3) 中, 可取 $U = V$. 这就证明了如下定理.

定理 1.5.5 设 A 为 n 阶复规范阵, 则存在酉阵 U, 使得

$$A = U^* \Lambda U,$$

其中 $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n), \lambda_1, \cdots, \lambda_n$ 为 A 的特征值.

定理 1.5.6 (Jordan 分解) 设 A 为 n 阶方阵, 则存在复的可逆方阵 P, 使得

$$A = P\mathrm{diag}(\Lambda_1, \Lambda_2, \cdots, \Lambda_k)P^{-1},$$

这里

$$\Lambda_j = \begin{pmatrix} \lambda_j & 1 & & 0 & \cdots & 0 \\ & & & & \ddots & \vdots \\ & \lambda_j & & 1 & & 0 \\ & & \ddots & & \ddots & \\ & & & \ddots & & 1 \\ & & & & \ddots & \lambda_j \end{pmatrix}, \quad j = 1, \cdots, k$$

为对应于 A 的特征值 λ_j 的 Jordan 块.

定理 1.5.7 (Schur 分解) 设 A 为 $n\times n$ 复方阵, 则存在酉阵 U, 使得 $A=ULU^*$, 其中 L 为上 (下) 三角阵, L 的对角元为 A 的特征值. 若 A 是 $n\times n$ 实方阵, 且它的特征值也为实数, 则 U 为正交阵.

证明 设 A 的特征值为 $\lambda_1,\cdots,\lambda_n, x_1\in C^n, x_1$ 为对应于 λ_1 的单位特征向量. 将 x_1 扩充为 C^n 的一组基：$x_1,y_2,\cdots,y_n$, 应用 Schmidt 正交化程序，将这些向量变换为 C^n 的一组标准正交基 $x_1,z_2,\cdots,z_n$. 定义 n 阶酉阵 $U_1=(x_1,z_2,\cdots,z_n)$, 则有

$$U_1^*AU_1=\begin{pmatrix}\lambda_1 & * \\ 0 & A_1\end{pmatrix},$$

其中 A_1 为 $n-1$ 阶方阵, 它的特征值为 $\lambda_2,\cdots,\lambda_n$. 设 $x_2\in C^{n-1}$, 且是 A_1 的对应于特征值 λ_2 的单位特征向量. 重复上面的方法可以得到 $n-1$ 阶酉阵 U_2, 使得

$$U_2^*A_1U_2=\begin{pmatrix}\lambda_2 & * \\ 0 & A_2\end{pmatrix}.$$

定义

$$\widetilde{U}_2=\begin{pmatrix}1 & 0 \\ 0 & U_2\end{pmatrix},$$

则 $\widetilde{U}_2$ 和 $U_1\widetilde{U}_2$ 都是酉阵, 且使

$$\widetilde{U}_2^*U_1^*AU_1\widetilde{U}_2=\left(\begin{array}{cc:c}\lambda_1 & * & * \\ 0 & \lambda_2 & \\ \hdashline 0 & & A_2\end{array}\right),$$

这里 A_2 为 $n-2$ 阶方阵, 它的特征值为 $\lambda_3,\cdots,\lambda_n$. 重复上述过程, 我们会构造出 n 阶酉阵 $U_1,\widetilde{U}_2,\cdots,\widetilde{U}_{n-1}$. 记 $U=U_1\widetilde{U}_2\cdots\widetilde{U}_{n-1}$, 使有 U^*AU 为上三角阵.

如果我们首先对 A^* 作分解：$A^*=ULU^*$, 这里 U 为酉阵, L 为上三角阵, 则 $A=UL^*U^*$, 此时 L^* 就是下三角阵了.

最后, 若 A 的元素和所有特征值都是实数, 那么它的所有特征向量都可取为实向量, 此时酉阵 U 便为正交阵. 定理证毕.

§1.6 矩阵的范数

为了度量一个向量的“大小”, 在 §1.1 我们引进了向量的长度. 对于一个 $m\times n$ 矩阵, 我们自然也可以把它看成 mn 维向量, 按向量的办法定义它的“长度”. 但是, 下面我们用更一般的途径引进度量矩阵“大小”的量, 这就是矩阵的范数.

定义 1.6.1 设 A 为 $m \times n$ 复矩阵, 定义非负实函数 $\|A\|$. 若它满足下列性质:

(1) 正定性: $\|A\| \geqslant 0$, 且 $\|A\| = 0 \Longleftrightarrow A = 0$;

(2) 齐次性: $\|\alpha A\| = |\alpha| \|A\|, \alpha$ 为一数;

(3) 三角不等式: $\|A + B\| \leqslant \|A\| + \|B\|$,

则称 $\|A\|$ 为 A 的范数. 若对任意的 $A_{m \times n}$ 和 $B_{n \times l}$, 有

$$\|AB\| \leqslant \|A\| \|B\|, \tag{1.6.1}$$

则称该范数是相容的.

例 1.6.1 (欧氏范数) 设 $A = (a_{ij})_{m \times n}$. 定义

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}. \tag{1.6.2}$$

则它是一种范数, 称为矩阵 A 的 Euclid 范数, 简称欧氏范数, 或称 Frobenius 范数. $\|A\|_F$ 也可改写为如下形式

$$\|A\|_F = (\mathrm{tr} A^* A)^{1/2}, \tag{1.6.3}$$

从此式可以看出, 这是把 A 按列先后排列成 $mn \times 1$ 向量后, 取其欧氏长度得到的.

因为

$$\begin{aligned} \|AB\|_F^2 &= \mathrm{tr}(B^* A^* AB) \\ &\leqslant \lambda_1(A^* A) \mathrm{tr}(B^* B) \leqslant (\mathrm{tr} A^* A)(\mathrm{tr} B^* B) \\ &= \|A\|_F^2 \|B\|_F^2, \end{aligned}$$

其中 $\lambda_1(D)$ 表示 D 的最大特征值, 所以, 欧氏范数是相容的.

例 1.6.2 (谱范数)

$$\|A\|_2 = \lambda_1^{1/2}(A^* A),$$

即矩阵 A 的谱范数是 $A^* A$ 的最大特征值的平方根. 谱范数也是相容范数. 事实上

$$\begin{aligned} \|AB\|_2^2 &= \lambda_1(B^* A^* AB) \\ &\leqslant \lambda_1(B^* \cdot \lambda_1(A^* A) I \cdot B) \\ &\leqslant \lambda_1(A^* A) \cdot \lambda_1(B^* B) \\ &= \|A\|_2^2 \|B\|_2^2, \end{aligned}$$

这里, 在两个不等号处应用了特征值的单调性: 若 $A \geqslant 0, B \geqslant 0, A \geqslant B$, 则 $\lambda_1(A) \geqslant \lambda_1(B)$(证明见推论 4.4.1).