现代数学基础

6 矩阵论

■ 詹兴致

$$A^{(n-1)^2+1}>0$$

高等教育出版社
HIGHER EDUCATION PRESS

内容提要

本书基于作者在北京大学和华东师范大学的讲稿而写成，主要讲述矩阵的分析和组合性质．作者强调思想方法，选择了具有基本重要性的概念、结果和证明技巧作为本书素材．和同类书相比，本书起点较高，具有一定的深度，内容比较全面，并反映了最新的研究成果．内容包括：张量积与复合矩阵、Hermite矩阵和优超关系、奇异值和酉不变范数、矩阵扰动、非负矩阵、符号模式、矩阵的应用．本书表达简洁流畅．读者可以在较短的时间内了解和掌握矩阵论的基本知识．附录列出了一些未解决的矩阵问题，相信有兴趣的读者会继续钻研．

本书是高等院校理工科研究生的教学用书，也可供高年级本科生以及研究人员使用参考．

图书在版编目（CIP）数据

矩阵论 / 詹兴致．—北京：高等教育出版社，2008.6

ISBN 978-7-04-024465-6

Ⅰ.矩…　Ⅱ.詹…　Ⅲ.矩阵-理论-高等学校-教材
Ⅳ.0151.21

中国版本图书馆 CIP 数据核字（2008）第 070221 号

策划编辑 赵天夫　**责任编辑** 赵天夫　**封面设计** 张　楠　**责任印制** 韩　刚

出版发行	高等教育出版社	购书热线	010-58581118
社　址	北京市西城区德外大街 4 号	免费咨询	800-810-0598
邮政编码	100011	网　址	http://www.hep.edu.cn
总　机	010-58581000		http://www.hep.com.cn
		网上订购	http://www.landraco.com
经　销	蓝色畅想图书发行有限公司		http://www.landraco.com.cn
印　刷	北京中科印刷有限公司	畅想教育	http://www.widedu.com
开　本	787×1092　1/16	版　次	2008 年 6 月第 1 版
印　张	10	印　次	2008 年 6 月第 1 次印刷
字　数	220 000	定　价	26.00 元

物料号 24465-00

序 言

本书基于 2002 年春季学期我在北京大学讲授 “矩阵分析” 课以及 2007 年秋季学期我在华东师范大学讲授 “矩阵论” 课的讲稿, 因此它主要是一本理工科研究生的教学用书, 它也可以作为研究人员和高年级本科生的参考书.

矩阵论在科学与工程计算、控制论、系统论、信息论、信号处理、理论计算机科学、经济学、组合与图论、运筹学、统计学、概率论、数学物理、动力系统等领域均有应用. 矩阵论一方面是有用的工具, 另一方面也是目前一个活跃而广阔的研究领域. 本书主要讲述矩阵的分析性质和组合性质.

我选择了那些具有基本重要性的概念、结果和证明技巧作为本书的材料. 在我眼里, 这里的定理和证明至少满足下面三个条件中的一条: 1) 重要的; 2) 巧妙的; 3) 漂亮的. 我总是强调思想方法. 作为教学用书, 本书的大部分内容当然是经典的, 但也有一些非常新的, 例如在 2007 年证明的结果. 第七章 “符号模式” 以及最后的 “矩阵的应用” 那个短章所包含的内容也是比较新的. 附录是一些未解决的矩阵问题.

习题是本书重要的组成部分, 它们大部分选自研究性的论文因而具有相当的难度. 读者如果不会做这些习题完全不必气馁, 但我还是建议读者先自己想一段时间, 如果仍然没有思路再去查原文. 学数学的最好方法是 “做数学”. 我希望本书不仅提供知识, 而且也能激发进一步的研究.

感谢高等教育出版社的张小萍女士和赵天夫先生热情地邀请我在高等教育出版社出版本书! 与他们的合作十分愉快. 谢谢我的研究生裴培同学不辞辛苦地完成了本书的打字工作! 本书的写作得到国家自然科学基金 (项目号 10571060) 和华东师范大学研究生重点教材基金 (项目号 2007JC02) 的资助, 一并致谢! 当然, 我还要感谢妻子杨蕾对我工作的支持!

读者如果发现本书的任何错误 (数学的、打印的、表达上的) 或者有建议, 请好心地告诉我.

詹兴致

zhan@math.ecnu.edu.cn

2008 年 3 月于华东师范大学, 上海

目　录

第一章　预备知识

本章给出的概念和结果的大部分在本书后面要用到, 其余部分读者在未来的工作中可能会有用. 我们也建立一些通用的记号.

我们将考虑复数当然也包括实数矩阵, *方阵*是指行数和列数相同的矩阵, 而*长方矩阵*是指行数和列数可能不相同的矩阵. 用 $M_{m,n}$ 表示 m 行 n 列复矩阵的集合, 用 M_n 表示 n 阶复方阵的集合, 这里字母 M 暗示矩阵 (matrix). $\mathbb{C}^n$ 表示具有 n 个复数分量的列向量的集合, $\mathbb{R}^n$表示具有 n 个实数分量的列向量的集合. 我们总是把 $\mathbb{C}^n$ 中的向量看作是列向量以便可以把它们放在矩阵的右边相乘. 行向量可以用列向量的转置或共轭转置得到. 在运算中, 我们将行向量和列向量都看成是特殊的矩阵, 即只有一行或一列的矩阵.

设 $A = (a_{ij}) \in M_{m,n}$, a_{ij} 是矩阵 A 的元素, 用 A^T表示 A 的转置, 用 A^*表示 A 的共轭转置, 即 $A^* = (\overline{A})^T \in M_{n,m}$, $\overline{A}$ 表示把 A 的每个元素取共轭得到的矩阵. 为简便起见, 本书中的零矩阵均用 0 表示, 读者应该能从上下文区别 0 的含义.

§1.1 特殊矩阵类

设 $A \in M_n$, 若 $A^*A = AA^*$, 则 A 称为*正规矩阵*. 若 $A^* = A$, 则 A 称为 *Hermite 矩阵*. 若 $A^* = -A$, 则 A 称为*反 Hermite 矩阵*. 我们总是用 I 表示单位矩阵, 即对角元素全为 1 的对角矩阵, 其阶数一般可以从上下文看出. 若 $A^*A = I$, 则 A 称为*酉矩阵*. 因此酉矩阵就是满足条件 $A^{-1} = A^*$ 的矩阵 A. 显然, Hermite 矩阵, 反 Hermite 矩阵, 酉矩阵都是正规矩阵; 实 Hermite 矩阵就是实对称矩阵, 实酉矩阵就是实正交矩阵. 用 $\mathrm{diag}(d_1, \cdots, d_n)$ 表示对角元素为 $d_1, \cdots, d_n$ 的对角矩阵. 一个方阵全体特征值的集合称为该方阵的*谱*.

定理 1.1 (谱分解) 每个正规矩阵都酉相似于一个对角矩阵, 即若 $A \in M_n$ 正规, 则存在酉矩阵 $U \in M_n$ 满足

$$A = U\mathrm{diag}(\lambda_1, \cdots, \lambda_n)U^*. \tag{1.1}$$

(1.1) 中的 $\lambda_1, \cdots, \lambda_n$ 显然是 A 的特征值.

我们将在 §1.5 中用 Schur 酉三角化定理来证明这个定理.

用 $\langle \cdot, \cdot \rangle$ 记 $\mathbb{C}^n$ 上的标准内积. 若 $x = (x_1, \cdots, x_n)^T, y = (y_1, \cdots, y_n)^T \in \mathbb{C}^n$, 则 $\langle x, y \rangle = \sum\limits_{j=1}^{n} x_j \overline{y}_j = y^* x$.

$A \in M_n$ 称为半正定矩阵, 若

$$\langle Ax, x \rangle \geqslant 0, \quad 对任何\ x \in \mathbb{C}^n. \tag{1.2}$$

$A \in M_n$ 称为正定矩阵, 若

$$\langle Ax, x \rangle > 0, \quad 对任何\ 0 \neq x \in \mathbb{C}^n. \tag{1.3}$$

正定矩阵就是可逆的半正定矩阵.

对 $A \in M_n$ 和 $x, y \in \mathbb{C}^n$ 我们有如下的极化恒等式

$$4\langle Ax, y \rangle = \sum_{k=0}^{3} i^k \langle A(x + i^k y), x + i^k y \rangle,$$
$$4\langle x, Ay \rangle = \sum_{k=0}^{3} i^k \langle x + i^k y, A(x + i^k y) \rangle,$$

这里 $i = \sqrt{-1}$. 这两个极化恒等式清楚地表明: 如果 $A \in M_n$ 满足对任何 $x \in \mathbb{C}^n$, $\langle Ax, x \rangle$ 为实数, 则 A 为 Hermite 矩阵. 特别, 定义式子 (1.2) 蕴涵半正定矩阵首先是 Hermite 矩阵. 实际上, 半正定矩阵就是所有特征值都非负的 Hermite 矩阵, 而正定矩阵就是所有特征值都为正的 Hermite 矩阵. 若 $A \in M_n$ 半正定, 则对任何 $B \in M_{n,k}$, B^*AB 半正定; 若 $A \in M_n$ 正定, 则对任何可逆阵 $B \in M_n$, B^*AB 正定.

相似于某个对角矩阵的矩阵称为可对角化矩阵. 从 Jordan 标准形可知, 若 $A \in M_n$ 的 n 个特征值互不相同则 A 可对角化.

$A = (a_{ij}) \in M_n$ 称为上三角矩阵, 若 $a_{ij} = 0$ 对所有 $i > j$ 成立, 即这时对角线下方的元素都是 0.

$A = (a_{ij}) \in M_n$ 称为下三角矩阵, 若 $a_{ij} = 0$ 对所有 $i < j$ 成立, 即这时对角线上方的元素都是 0.

$A = (a_{ij}) \in M_n$ 称为 Hessenberg 矩阵, 若 $a_{ij} = 0$ 对所有 $i > j + 1$ 成立.

我们说矩阵 $A = (a_{ij})$ 具有上带宽 q, 如果 $a_{ij} = 0$ 对于满足 $j - i > q$ 的所有下标 i, j 成立; 说矩阵 A 具有下带宽 p, 如果 $a_{ij} = 0$ 对于满足 $i - j > p$ 的所有下标 i, j 成立. 例如, 下三角矩阵具有上带宽 0, Hessenberg 矩阵具有下带宽 1. 矩

阵 $A=(a_{ij})\in M_n$ 称为带状矩阵, 如果 A 具有上带宽 $q\leqslant n-2$ 或者具有下带宽 $p\leqslant n-2$.

一个矩阵称为稀疏矩阵, 如果它具有比较多的零元素. 可见, 稀疏矩阵不是一个准确的概念.

$A=(a_{ij})\in M_{m,n}$ 称为 0–1 矩阵, 若每个元素 $a_{ij}\in\{0,1\}$. 每一行每一列都恰好有一个 1 的 0–1 方阵称为置换矩阵.

$A=(a_{ij})\in M_n$ 称为 Toeplitz 矩阵, 若存在数

$$a_{-n+1},\cdots,a_{-1},a_0,a_1,\cdots,a_{n-1}$$

使得 $a_{ij}=a_{j-i}$. 因此一个 Toeplitz 矩阵是有如下形式的矩阵

$$\begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-1} \\ a_{-1} & a_0 & a_1 & \cdots & a_{n-2} \\ a_{-2} & a_{-1} & a_0 & \cdots & a_{n-3} \\ \vdots & & \vdots & & \vdots \\ a_{-n+1} & a_{-n+2} & a_{-n+3} & \cdots & a_0 \end{pmatrix}.$$

$A=(a_{ij})\in M_n$ 称为 Hankel 矩阵, 若存在数 $a_1,\cdots,a_{2n-1}$ 使得 $a_{ij}=a_{i+j-1}$. 因此一个 Hankel 矩阵是有如下形式的矩阵

$$\begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_n \\ a_2 & a_3 & a_4 & \cdots & a_{n+1} \\ a_3 & a_4 & a_5 & \cdots & a_{n+2} \\ \vdots & \vdots & \vdots & & \vdots \\ a_n & a_{n+1} & a_{n+2} & \cdots & a_{2n-1} \end{pmatrix}.$$

形如

$$A=\begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_n \\ a_n & a_1 & a_2 & \cdots & a_{n-1} \\ a_{n-1} & a_n & a_1 & \cdots & a_{n-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_2 & a_3 & a_4 & \cdots & a_1 \end{pmatrix} \tag{1.4}$$

的矩阵称为循环矩阵, 其中 A 由第一行决定, 从第二行开始每行都由前一行的元素向前循环移动一个位置而得到. 记 (1.4) 中的矩阵为 $\mathrm{Circ}(a_1,a_2,a_3,\cdots,a_n)$. $P=\mathrm{Circ}(0,1,0,\cdots,0)$ 称为基本循环矩阵. 注意 P 是个置换矩阵. 我们有

$$A=\sum_{k=0}^{n-1}a_{k+1}P^k. \tag{1.5}$$

P 的特征多项式为 λ^n-1, 因此它的特征值为 $z^j, j=0,1,\cdots,n-1, z=e^{\frac{2\pi}{n}i}, i=\sqrt{-1}$. 记 $x_j=\frac{1}{\sqrt{n}}(1,z^j,z^{2j},\cdots,z^{(n-1)j})^T$, 则 $x_0,x_1,\cdots,x_{n-1}$ 为 P 的一组标准正交特征向量. 记 $U=(x_0,x_1,\cdots,x_{n-1})$, 则 U 为酉矩阵, 且

$$P=U\mathrm{diag}(1,z,z^2,\cdots,z^{n-1})U^*. \tag{1.6}$$

记

$$f(t)=\sum_{k=0}^{n-1}a_{k+1}t^k.$$

(1.5) 和 (1.6) 给出

$$A=U\mathrm{diag}(f(1),f(z),f(z^2),\cdots,f(z^{n-1}))U^*. \tag{1.7}$$

(1.7) 表明: 我们可以用一个固定的酉矩阵将所有循环矩阵酉对角化.

§1.2 特征多项式

定理 1.2 设 $E_k(A)$ 表示 $A\in M_n$ 的所有 k 阶主子式之和. 则 A 的特征多项式为

$$f_A(t)=t^n-E_1(A)t^{n-1}+E_2(A)t^{n-2}+\cdots+(-1)^nE_n(A).$$

这个定理表明: 特征值连续地依赖于矩阵的元素, 或者说, 特征值是矩阵的连续函数. 关于这一事实的定量结果见第五章的定理 5.2.

§1.3 谱映射定理

定理 1.3 设 $f(t)$ 是个复系数多项式, 设 $A\in M_n$ 的特征值为 $\lambda_1,\cdots,\lambda_n$. 则 $f(A)$ 的特征值为 $f(\lambda_1),\cdots,f(\lambda_n)$.

利用 A 的 Jordan 标准形, 这个结果是显然的.

§1.4 范数

设 $F=\mathbb{C}$ 或 $\mathbb{R}$, V 是 F 上的一个向量空间. 用符号 $\forall$ 表示“任意”. 一个函数 $\|\cdot\|: V\to\mathbb{R}$ 若满足下面三条性质就称为 V 上的一个范数:

(1) (正定性) $\|x\|\geqslant 0, \forall x\in V$, 并且 $\|x\|=0$ 当且仅当 $x=0$;
(2) (正齐次性) $\|\alpha x\|=|\alpha|\|x\|$, $\forall\alpha\in F, \forall x\in V$;
(3) (三角不等式) $\|x+y\|\leqslant\|x\|+\|y\|$, $\forall x,y\in V$.

例如, 设 $p \geqslant 1$, 对于 $x = (x_1, \cdots, x_n)^T \in \mathbb{C}^n$ 定义 $\|x\|_p = (\sum\limits_{i=1}^{n} |x_i|^p)^{1/p}$, 则 $\|\cdot\|_p$ 是 $\mathbb{C}^n$ 上的一个范数, 称为 l_p 范数. l_2 范数就是欧氏范数. 矩阵空间 $M_{m,n}$ 上的欧氏范数称为矩阵的 Frobenius 范数:

$$\|A\|_F = (\mathrm{tr} A^* A)^{1/2} = \left(\sum_{i,j} |a_{ij}|^2 \right)^{1/2}, \quad A = (a_{ij}) \in M_{m,n},$$

这里 tr 表示迹.

设 $A = (a_{ij}) \in M_{m,n}$, A 的行和范数定义为:

$$\|A\|_r = \max_{1 \leqslant i \leqslant m} \sum_{j=1}^{n} |a_{ij}|.$$

A 的列和范数定义为:

$$\|A\|_c = \max_{1 \leqslant j \leqslant n} \sum_{i=1}^{m} |a_{ij}|.$$

一个矩阵 $A \in M_{m,n}$ 可以看作是从 $\mathbb{C}^n$ 到 $\mathbb{C}^m$ 的线性映射: $x \mapsto Ax$. 设 $\|\cdot\|_\alpha$ 是 $\mathbb{C}^m$ 上的范数, $\|\cdot\|_\beta$ 是 $\mathbb{C}^n$ 上的范数. 对 $A \in M_{m,n}$, 定义

$$\|A\| = \max_{0 \neq x \in \mathbb{C}^n} \frac{\|Ax\|_\alpha}{\|x\|_\beta} = \max_{x \in \mathbb{C}^n, \|x\|_\beta = 1} \|Ax\|_\alpha. \tag{1.8}$$

容易验证 $\|\cdot\|$ 是个范数, 称为由 $\|\cdot\|_\alpha$ 和$\|\cdot\|_\beta$ 诱导的算子范数. 由范数的定义可知, 任何范数都满足 $|\|x\| - \|y\|| \leqslant \|x - y\|$, 所以范数是连续函数. Weierstrass 定理说: 紧集上的实值连续函数可以取到最大值, 也可以取到最小值. 因此在 (1.8) 中我们可写 max, 而不必写 sup .

M_n 上的范数 $\|\cdot\|$ 称为是次可乘的, 若

$$\|AB\| \leqslant \|A\| \|B\|, \quad \forall A, B \in M_n.$$

易知 M_n 上的任何算子范数都是次可乘的.

向量的欧氏范数诱导的 $M_{m,n}$ 上的算子范数称为谱范数, 记作 $\|\cdot\|_\infty$:

$$\|A\|_\infty = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \max\{\|Ax\|_2 \mid \|x\|_2 = 1, x \in \mathbb{C}^n\}.$$

容易证明, 若 B 是 A 的子矩阵, 则 $\|B\|_\infty \leqslant \|A\|_\infty$. 这一事实我们以后将会有更好的理解. 另一个容易看出的有用结论是: 把一个矩阵左乘或右乘酉矩阵不改变其谱范数.

范数用于度量向量的大小、向量之间的距离, 也用于定义序列的收敛性. 设 $\|\cdot\|$ 是向量空间 V 上的范数. 序列 $\{x_j\}_{j=1}^{\infty} \subset V$ 收敛到向量 x 的意思是 $\lim\limits_{j \to \infty} \|x_j - x\| = 0$.

我们需要各种各样的范数, 因为对于特定的问题, 可能一种范数会比另一种范数用起来更方便. 另一方面, 下面的定理说明有限维向量空间上的所有范数都是等价的, 因此在判断序列的收敛性时不同的范数会给出同样的结论.

定理 1.4　设 V 是复或实的有限维向量空间, $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 是 V 上的范数. 则存在正数 c 和 d 满足

$$c\|x\|_\beta \leqslant \|x\|_\alpha \leqslant d\|x\|_\beta, \quad \forall x \in V.$$

证明　$S = \{x \in V \mid \|x\|_\beta = 1\}$ 是个紧集. 由 Weierstrass 定理, 实值连续函数 $f(x) = \|x\|_\alpha$ 在 S 上可以取到最小值 c 和最大值 d. 由范数正定性知 $c > 0, d > 0$.$\forall x \in V, x \neq 0$, 有 $x/\|x\|_\beta \in S$. 从而

$$c \leqslant \left\| \frac{x}{\|x\|_\beta} \right\|_\alpha \leqslant d,$$

即

$$c\|x\|_\beta \leqslant \|x\|_\alpha \leqslant d\|x\|_\beta.$$

对于 $x = 0$, 这两个不等式明显成立. 这就证明了它们对 $\forall x \in V$ 成立.　□

§1.5　矩阵分解

定理 1.5 (Schur 酉三角化)　每个复方阵都酉相似于某个上三角矩阵.

证明　设 $A \in M_n$, 对阶数 n 作归纳法. 当 $n = 1$ 时结论平凡地成立. 假设结论对 $n-1$ 阶矩阵成立, 现在考虑 n 阶的 A. 任取 A 的一个特征值 λ 和一个相应的单位特征向量 x_1, 将 x_1 扩充成 $\mathbb{C}^n$ 的一个标准正交基: $x_1, x_2, \cdots, x_n$, 令 $U_1 = (x_1, \cdots, x_n)$, 则 U_1 是酉矩阵, 且

$$U_1^* A U_1 = \begin{pmatrix} \lambda & y^* \\ 0 & A_1 \end{pmatrix},$$

其中 $A_1 \in M_{n-1}$. 由归纳假设, 存在酉矩阵 $U_2 \in M_{n-1}$ 使得 $U_2^* A_1 U_2$ 为上三角矩阵. 令 $U = U_1 \operatorname{diag}(1, U_2) \in M_n$, 则 U 为酉矩阵, 且

$$U^* A U = \begin{pmatrix} \lambda & y^* U_2 \\ 0 & U_2^* A_1 U_2 \end{pmatrix}$$

为上三角矩阵.　□

显然, 以上定理中上三角矩阵的对角元素就是所给方阵的特征值. 从证明过程可知, 我们可以将特征值按事先任意指定的顺序放在上三角矩阵的对角线上. 定理中的上三角矩阵也可以换成下三角矩阵, 考虑转置即可.

定理 1.1 的证明 根据定理 1.5, 存在酉矩阵 U 和上三角矩阵 T 满足 $A = UTU^*$. 从 $A^*A = AA^*$ 得 $T^*T = TT^*$. 逐次比较 T^*T 和 TT^* 的第 i 个对角元素, $i = 1, 2, \cdots, n$, 知 T 的第 i 行的非对角元素全部为零. 因此, T 是对角矩阵. □

设 $A \in M_{m,n}$ 若 $m \geqslant n$, 则半正定阵 A^*A 的特征值的非负平方根称为 A 的奇异值, 若 $m < n$, 则半正定阵 AA^* 的特征值的非负平方根称为 A 的奇异值. 我们区分 $m \geqslant n$ 和 $m < n$ 两种情况是为了避免讨论那些明显的零特征值. 对角矩阵的记号 $\mathrm{diag}(s_1, \cdots, s_p)$ 也可表示长方矩阵, 其行数列数可以从上下文看出.

定理 1.6 (奇异值分解) 设 $A \in M_{m,n}$, 则存在酉矩阵 $U \in M_m$ 和酉矩阵 $V \in M_n$ 使得

$$UAV = \mathrm{diag}(s_1, \cdots, s_p),$$

这里 $s_1 \geqslant \cdots \geqslant s_p \geqslant 0$, $p = \min(m, n)$.

证明 用 $\|\cdot\|$ 表示向量的欧氏范数或者矩阵的谱范数.

取 $x \in \mathbb{C}^n, y \in \mathbb{C}^m$ 满足 $\|x\| = \|y\| = 1$ 及 $Ax = s_1 y, s_1 = \|A\|$. 这样的 x 和 y 是存在的. 若 $A = 0$ 则 $s_1 = 0$, x 和 y 可取任意的单位向量. 若 $A \neq 0$ 则有 $x \in \mathbb{C}^n, \|x\| = 1$ 满足 $\|Ax\| = \|A\|$, 取 $y = Ax/\|A\|$ 即可.

取 U_1, V_1 使得 $U_2 = (y, U_1) \in M_m, V_2 = (x, V_1) \in M_n$ 且 U_2, V_2 为酉矩阵. 则

$$U_2^* A V_2 = \begin{pmatrix} s_1 & w^* \\ 0 & B \end{pmatrix},$$

其中 $B \in M_{m-1,n-1}$. 根据谱范数的性质, 有

$$s_1 = \|U_2^* A V_2\| \geqslant \|(s_1, w^*)\| = (s_1^2 + w^* w)^{1/2},$$

这表明 $w = 0$. 于是

$$U_2^* A V_2 = \begin{pmatrix} s_1 & 0 \\ 0 & B \end{pmatrix},$$

将以上推理用于 B 可知, 存在酉矩阵 $U_3 \in M_{m-1}$ 和酉矩阵 $V_3 \in M_{n-1}$, 使得

$$U_3^* B V_3 = \mathrm{diag}(s_2, C),$$

其中 $s_2 = \|B\| \leqslant s_1, C \in M_{m-2,n-2}$. 令 $U_4 = \mathrm{diag}(1, U_3), V_4 = \mathrm{diag}(1, V_3)$, 则 U_4 和 V_4 都是酉矩阵, 且

$$U_4^* U_2^* A V_2 V_4 = \mathrm{diag}(s_1, s_2, C).$$

下面再对 C 重复以上推理, 逐次做下去, 最终我们可以通过将 A 左乘一串酉矩阵及右乘一串酉矩阵把 A 化成对角矩阵, 且对角元素非负. 又因为有限个酉矩阵的乘积还是酉矩阵, 这就完成了证明. □

推论 1.7 (极分解) 设 $A \in M_n$, 则存在半正定矩阵 P, Q 和酉矩阵 U, V 满足 $A = PU = VQ$.

证明 由奇异值分解定理, 存在酉矩阵 W, Z 满足

$$A = WDZ, \quad D = \mathrm{diag}(s_1, \cdots, s_n), \quad s_1 \geqslant \cdots \geqslant s_n \geqslant 0,$$

于是

$$A = WDW^* \cdot WZ = WZ \cdot Z^*DZ.$$

令 $P = WDW^*, U = WZ, V = WZ, Q = Z^*DZ$ 即可. □

推论 1.8 (满秩分解) 设 $A \in M_{m,n}$, 其秩 rank $A = r$. 则有 $F \in M_{m,r}, G \in M_{r,n}$ 满足 $A = FG$.

证明 由奇异值分解定理, 存在酉矩阵 U, V 使得

$$A = U \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} V,$$

其中 $D = \mathrm{diag}(s_1, \cdots, s_r)$, $s_1 \geqslant \cdots \geqslant s_r > 0$.

记 $\sqrt{D} = \mathrm{diag}(\sqrt{s_1}, \cdots, \sqrt{s_r})$, 取 $F = U \begin{pmatrix} \sqrt{D} \\ 0 \end{pmatrix}, G = \begin{pmatrix} \sqrt{D}, 0 \end{pmatrix} V$ 即可. □

定理 1.9 (QR 分解) 设 $A \in M_n$, 则存在酉矩阵 $Q \in M_n$ 和上三角矩阵 $R \in M_n$ 使得 $A = QR$. 若 A 是实矩阵, 则 Q 和 R 都可以取作实矩阵, 此时 Q 为实正交矩阵.

证明 用线性代数中的 Gram-Schmidt 正交化过程. □

§1.6 数值范围

用 $\|x\|$ 表示 $x \in \mathbb{C}^n$ 的欧氏范数, 矩阵 $A \in M_n$ 的数值范围定义为集合

$$W(A) = \{x^*Ax \mid \|x\| = 1, x \in \mathbb{C}^n\}.$$

数值范围有用的原因之一在于 $W(A)$ 包含 A 的谱. $W(A)$ 是复平面上的有界闭凸集. $W(A)$ 作为紧集 $\{x \in \mathbb{C}^n \mid \|x\| = 1\}$ 在连续映射 $x \mapsto x^*Ax$ 下的像, 当然是紧集 (即有界闭, 在 $\mathbb{C}$ 中). 凸性则不明显.

定理 1.10 (Toeplitz-Hausdorff) 对任何 $A \in M_n$, $W(A)$ 是凸集.

证明 (P.R. Halmos) 只要证明对 $\mathbb{C}$ 中任意的直线 L, $W(A) \cap L$ 是连通集. 将 $\mathbb{C}$ 等同于 $\mathbb{R}^2$, 设 L 的方程为 $as + bt + c = 0$, 其中 $a, b, c \in \mathbb{R}$ 给定, (s, t) 为点的坐标. 设 $i = \sqrt{-1}$, 考虑 A 的笛卡儿分解 $A = G + iH$, 其中 $G = (A + A^*)/2$

和 $H=(A-A^*)/(2i)$ 都是 Hermite 矩阵. 记 $S=\{x\in\mathbb{C}^n \mid \|x\|=1\}$, 于是 $W(A)=\{(x^*Gx,x^*Hx)\mid x\in S\}$. 而

$$(x^*Gx,x^*Hx)\in L\Leftrightarrow x^*(aG+bH+cI)x=0.$$

记 $T=aG+bH+cI$, $E=\{x\in S\mid x^*Tx=0\}$. 定义映射

$$\phi:\mathbb{C}^n\to\mathbb{R}^2,\quad \phi(x)=(x^*Gx,x^*Hx).$$

则 $W(A)\cap L=\phi(E)$. 因为 ϕ 是连续函数, 而连通集在连续映射下的像是连通的, 我们只要证明 E 是连通的就可以完成证明.

我们将证明 E 是道路连通的. 设 $x,y\in E$. 若 x,y 线性相关, 因为 $\|x\|=\|y\|=1$, 则存在 $\theta\in\mathbb{R}$ 使得 $y=e^{i\theta}x$. 这里 e 是自然对数的底. 取

$$z(\alpha)=e^{i\alpha\theta}x,\quad 0\leqslant\alpha\leqslant 1,$$

则 $z(\alpha)$ 是 E 中从 x 到 y 的一条路. 下面假设 x 和 y 线性无关. 选取 $\beta\in\mathbb{R}$ 满足

$$e^{i\beta}y^*Tx\in i\mathbb{R}.$$

记 $x_0=e^{i\beta}x$, 则 $y^*Tx_0\in i\mathbb{R}$. 上面已证: E 中存在从 x 到 x_0 的一条路, 所以我们只要证明 E 中存在从 x_0 到 y 的路即可. 定义

$$u(\alpha)=(1-\alpha)x_0+\alpha y,\quad 0\leqslant\alpha\leqslant 1,$$

x_0 和 y 线性无关保证了 $u(\alpha)\neq 0$. 注意到 $x_0,y\in E$, 实部 $\mathrm{Re}(y^*Tx_0)=0$, T 为 Hermite 矩阵, 我们有

$$u(\alpha)^*Tu(\alpha)=(1-\alpha)^2x_0^*Tx_0+2(1-\alpha)\alpha\mathrm{Re}(y^*Tx_0)+\alpha^2y^*Ty=0.$$

因此令$z(\alpha)=u(\alpha)/\|u(\alpha)\|$, 则 $z(\alpha)$, $0\leqslant\alpha\leqslant 1$ 是 E 中从 x_0 到 y 的一条路. □

设 $\lambda_1,\cdots,\lambda_n$ 是 $A\in M_n$ 的特征值, 记 $\sigma(A)=\{\lambda_1,\cdots,\lambda_n\}$. 用 $\mathrm{Co}\Omega$ 表示集合 Ω 的凸包. 注意到数值范围是酉相似不变的, 即若 $U\in M_n$ 为酉矩阵, 则 $W(A)=W(U^*AU)$, 下面的结果是显然的.

定理 1.11 设 A 为正规矩阵, 则 $W(A)=\mathrm{Co}\sigma(\mathrm{A})$.

设 $A\in M_n$. $w(A)=\max\{|z|\mid z\in W(A)\}=\max\{|x^*Ax|:\|x\|=1,x\in\mathbb{C}^n\}$ 称为 A 的数值半径. 容易证明数值半径 $w(\cdot)$ 是 M_n 上的一个范数.

§1.7 多项式的伙伴矩阵

首项系数为 1 的复系数多项式

$$p(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$$

的伙伴矩阵定义为

$$C(p) = \begin{pmatrix} -a_{n-1} & -a_{n-2} & \cdots & -a_1 & -a_0 \\ 1 & 0 & & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & & 1 & 0 \end{pmatrix}. \tag{1.9}$$

不难证明: $p(z)$ 是 $C(p)$ 的特征多项式, $p(z)$ 也是 $C(p)$ 的极小多项式.

§1.8 广义逆

定理 1.12 设 $A \in M_{m,n}$, 则存在唯一的矩阵 $X \in M_{n,m}$ 满足以下四个方程:

$$\begin{aligned} &(1)\ AXA = A, && (2)\ XAX = X, \\ &(3)\ (AX)^* = AX, && (4)\ (XA)^* = XA. \end{aligned}$$

证明 设 A 的奇异值分解为

$$A = U \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} V,$$

其中 U, V 为酉矩阵, D 是对角元素全为正数的对角方阵. 令

$$X = V^* \begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^*,$$

则 $X \in M_{n,m}$ 而且 X 满足这四个方程. 下面证明唯一性. 设 $X, Y \in M_{n,m}$ 都满足这四个方程, 则

$$\begin{aligned} X &= XAX = A^*X^*X = A^*Y^*A^*X^*X = YAA^*X^*X \\ &= YAXAX = YAX = YAYAX = YY^*A^*AX \\ &= YY^*A^*X^*A^* = YY^*A^* = YAY = Y. \end{aligned}$$ □

满足以上定理中的四个方程的唯一矩阵 X 称为 A 的 Moore-Penrose *逆*, 记作 $A^\dagger$.

满足这四个方程中的一个或几个的矩阵也称为 A 的*广义逆*, 例如, 满足 (1) 的 X 称为 A 的一个 $\{1\}$-*逆*; 满足 (1) 和 (4) 的 X 称为 A 的$\{1,4\}$-*逆*; 等等. 除了 Moore-Penrose 广义逆, 其他的广义逆一般不是唯一的, 是一个集合.

§1.9 拓扑思想的应用

本节我们举两个例子说明拓扑思想在研究矩阵问题时的应用. 这两个结果是熟知的, 要点在于证明方法.

定理 1.13 (Cayley-Hamilton) 设 $f(t)$ 是 $A \in M_n$ 的特征多项式, 则 $f(A) = 0$.

证明 (Rosoff [96]) 先假设 A 是可对角化矩阵, A 的特征值是 $\lambda_1, \cdots, \lambda_n$, 则 $f(t) = \prod\limits_{j=1}^{n}(t - \lambda_j)$, 且存在可逆矩阵 T 使得 $A = T^{-1}\mathrm{diag}(\lambda_1, \cdots, \lambda_n)T$. 显然有

$$\begin{aligned} f(A) &= T^{-1}f(\mathrm{diag}(\lambda_1, \cdots, \lambda_n))T \\ &= T^{-1}\mathrm{diag}(f(\lambda_1), \cdots, f(\lambda_n))T \\ &= 0. \end{aligned}$$

因为可对角化矩阵在 M_n 中是稠密的, 对于任何 $A \in M_n$, 存在可对角化矩阵序列 $\{A_j\}_{j=1}^{\infty}$ 满足 $\lim\limits_{j\to\infty} A_j = A$. 记 $X \in M_n$ 的特征多项式为 $f_X(t)$. 因为特征多项式是矩阵的连续函数, $\lim\limits_{j\to\infty} f_{A_j}(t) = f_A(t)$. 利用上面已证的结论, 有

$$f(A) = \lim_{j\to\infty} f_{A_j}(A_j) = 0. \qquad \square$$

定理 1.14 设 $A \in M_{m,n}, B \in M_{n,m}$ 则 AB 和 BA 的非零特征值 (包括重数) 相同.

证明 在 $m \neq n$ 时, 对 A 和 B 添加适当数目的零行和零列使 A 和 B 变成同阶的方阵. 只需证 $m = n$ 的情形. 先假设 A 可逆, 此时 AB 和 BA 相似: $A^{-1}(AB)A = BA$, 从而它们的特征值相同.

因为可逆矩阵在 M_n 中稠密, 对于任何 $A \in M_n$, 存在可逆矩阵序列$\{A_j\}_{j=1}^{\infty}$ 满足 $\lim\limits_{j\to\infty} A_j = A$. 于是 $\lim\limits_{j\to\infty} A_jB = AB$. 用 $\sigma(\cdot)$ 记特征值的集合. 注意到特征值是矩阵的连续函数, 利用上面已证的结论, 有

$$\sigma(AB) = \lim_{j\to\infty} \sigma(A_jB) = \lim_{j\to\infty} \sigma(BA_j) = \sigma(BA). \qquad \square$$

用微分拓扑研究矩阵问题的一个例子是 [33].

§1.10 参考书和杂志

关于矩阵论的基本参考书有 [10, 13, 15, 18, 28, 29, 43, 47, 58, 59, 75, 78, 87, 104, 118, 124]. 本书的部分材料取自于 [15, 28, 43, 58, 59, 78, 87, 118].

关于矩阵的主要杂志有 Linear Algebra and Its Applications, SIAM Journal on Matrix Analysis and Applications, Linear and Multilinear Algebra, Operators and Matrices.

习 题

1. 设 $a_1, \cdots, a_n$ 为正实数. 证明矩阵 $\left(\frac{1}{a_i+a_j}\right)_{n\times n}$ 半正定.

2. (Oldenburger [89]) 设 $A \in M_n$, $\rho(A)$ 表示 A 的谱半径, 即 A 的特征值的模的最大者. 证明 $\lim\limits_{k\to\infty} A^k = 0$ 当且仅当 $\rho(A) < 1$.

3. 证明数值半径 $w(\cdot)$ 是 M_n 上的一个范数.

4. 证明数值半径 $w(\cdot)$ 和谱范数 $\|\cdot\|_\infty$ 满足如下关系:

$$\frac{1}{2}\|A\|_\infty \leqslant w(A) \leqslant \|A\|_\infty, \quad A \in M_n.$$

5. (Gelfand) 设 $A \in M_n$, 证明 $\rho(A) = \lim\limits_{k\to\infty} \|A^k\|_\infty^{1/k}$.

6. 设 $A \in M_{m,n}$, $B \in M_{n,m}$. 证明 $\begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix}$ 和 $\begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix}$ 相似, 从而给出定理 1.14 的另一个证明.

7. 设 $A_j \in M_n, j = 1, \cdots, m, m > n$, 且 $\sum\limits_{j=1}^m A_j$ 非奇异 (即可逆). 证明: 存在 $S \subseteq \{1, 2, \cdots, m\}$ 满足 $|S| \leqslant n$ 且 $\sum\limits_{j\in S} A_j$ 非奇异.

8. 证明每个复方阵都酉相似于某个对角元素全部相等的矩阵.

9. 证明对任意的复方阵 A, $\rho(A) \leqslant w(A) \leqslant \|A\|_\infty$.

10. 矩阵 $A = (a_{ij}) \in M_n$ 称为*严格对角占优*, 如果

$$|a_{ii}| > \sum_{j\neq i} |a_{ij}|, \quad i = 1, \cdots, n.$$

证明: 严格对角占优矩阵是可逆的.

11. (Gersgorin 圆盘定理) 用 $\sigma(A)$ 表示 $A = (a_{ij}) \in M_n$ 的特征值的集合. 记

$$D_i = \{z \in \mathbb{C} : |z - a_{ii}| \leqslant \sum_{j\neq i} |a_{ij}|\}, \quad i = 1, \cdots, n.$$

证明

$$\sigma(A) \subseteq \bigcup_{i=1}^n D_i,$$

并且如果这些圆盘 D_i 中有 k 个与其余 $n-k$ 个不相交, 则这 k 个圆盘的并集恰好含有 A 的 k 个特征值.

12. (Sherman-Morrison-Woodbury 公式) 设 $A \in M_n$ 可逆, $B, C \in M_{n,k}$ 使得 $I + C^*A^{-1}B$ 可逆, 其中 I 是单位阵. 证明 $A + BC^*$ 可逆且

$$(A + BC^*)^{-1} = A^{-1} - A^{-1}B(I + C^*A^{-1}B)^{-1}C^*A^{-1}.$$

13. (Li-Poon [76]) 证明: 每个实方阵都可以写成 4 个实正交矩阵的线性组合, 即若 A 是个实方阵, 则存在实正交矩阵 Q_i 和实数 r_i, $i = 1, 2, 3, 4$ 使得

$$A = r_1Q_1 + r_2Q_2 + r_3Q_3 + r_4Q_4.$$

14. ([117]) 如果映射 $f : M_n \to M_n$ 按某个固定的模式将 M_n 中的每个矩阵的元素重排, 则称 f 是一个*置换算子*. 怎样的置换算子保持矩阵的特征值不变? 保持秩不变?

第二章　张量积与复合矩阵

张量积和复合矩阵都是从已有的矩阵构造新矩阵的方法, 它们是矩阵论中的重要工具. 这两个概念可以对任意环上的矩阵定义, 但我们这里只叙述复矩阵的情形.

§2.1 张量积的定义及基本性质

设 $A=(a_{ij})\in M_{m,n}, B\in M_{s,t}$. A 和 B 的张量积记作 $A\otimes B$, 定义为下面的分块矩阵:

$$A\otimes B=\begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}\in M_{ms,nt}.$$

张量积又称为 Kronecker 积. 下面的性质是显然的.

(i) $(\alpha A)\otimes B=A\otimes(\alpha B)=\alpha(A\otimes B)$,对 $\alpha\in\mathbb{C}, A\in M_{m,n}, B\in M_{s,t}$.

(ii) $(A\otimes B)^T=A^T\otimes B^T$,对$A\in M_{m,n}, B\in M_{s,t}$.

(iii) $(A\otimes B)^*=A^*\otimes B^*$, 对 $A\in M_{m,n}, B\in M_{s,t}$.

(iv) $(A\otimes B)\otimes C=A\otimes(B\otimes C)$, 对 $A\in M_{m,n}, B\in M_{s,t}, C\in M_{p,q}$.

(v) $A\otimes(B+C)=(A\otimes B)+(A\otimes C)$, 对 $A\in M_{m,n}, B, C\in M_{s,t}$.

(vi) $(A+B)\otimes C=(A\otimes C)+(B\otimes C)$, 对 $A, B\in M_{m,n}, C\in M_{s,t}$.

(vii) $A\otimes B=0$ 当且仅当 $A=0$ 或 $B=0$.

(viii) 设 $A\in M_m, B\in M_n$. 若 A, B 对称, 则 $A\otimes B$ 对称.

(ix) 设 $A\in M_m, B\in M_n$. 若 A, B 为 Hermite 矩阵, 则 $A\otimes B$ 是 Hermite 矩阵.

引理 2.1 设 $A \in M_{m,n}, B \in M_{s,t}, C \in M_{n,k}$, $D \in M_{t,r}$. 则

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD).$$

证明 设 $A = (a_{ih}), C = (c_{hj})$, 则 $A \otimes B = (a_{ih}B)$, $C \otimes D = (c_{hj}D)$. 按照分块矩阵的乘法规则, $(A \otimes B)(C \otimes D)$ 的 (i, j) 块是

$$\sum_{h=1}^{n} a_{ih}Bc_{hj}D = \left(\sum_{h=1}^{n} a_{ih}c_{hj}\right) BD.$$

而这正是 AC 的 (i, j) 位置的元素乘 BD, 即 $(AC) \otimes (BD)$ 的 (i, j) 块. □

我们总是用 $\sigma(G)$ 记 $G \in M_n$ 的特征值 $\lambda_1, \cdots, \lambda_n$ 的集合: $\sigma(G) = \{\lambda_1, \cdots, \lambda_n\}$. 用 $\mathrm{sv}(G)$ 记 G 的奇异值 $s_1, \cdots, s_n$ 的集合: $\mathrm{sv}(G) = \{s_1, \cdots, s_n\}$.

定理 2.2 设 $A \in M_m, B \in M_n$.

(i) 若 A, B 可逆, 则 $A \otimes B$ 也可逆, 且 $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

(ii) 若 A, B 为正规矩阵, 则 $A \otimes B$ 也为正规矩阵.

(iii) 若 A, B 为酉矩阵, 则 $A \otimes B$ 也为酉矩阵.

(iv) 若 $\lambda \in \sigma(A)$, x 是对应的特征向量, $\mu \in \sigma(B), y$ 是对应的特征向量, 则

$$\lambda\mu \in \sigma(A \otimes B),\ x \otimes y \text{ 是对应的特征向量.}$$

(v) 若 $\sigma(A) = \{\lambda_1, \cdots, \lambda_m\}$, $\sigma(B) = \{\mu_1, \cdots, \mu_n\}$, 则

$$\sigma(A \otimes B) = \{\lambda_i\mu_j \mid i = 1, \cdots, m, j = 1, \cdots, n\}.$$

(vi) $\det(A \otimes B) = (\det A)^n(\det B)^m$.

(vii) 若 $\mathrm{sv}(A) = \{s_1, \cdots, s_m\}, \mathrm{sv}(B) = \{t_1, \cdots, t_n\}$, 则

$$\mathrm{sv}(A \otimes B) = \{s_it_j \mid i = 1, \cdots, m, j = 1, \cdots, n\}.$$

(viii) $\mathrm{rank}(A \otimes B) = (\mathrm{rank}A)(\mathrm{rank}B)$.

证明 (i) $(A \otimes B)(A^{-1} \otimes B^{-1}) = (AA^{-1}) \otimes (BB^{-1}) = I \otimes I = I$.

(ii)

$$\begin{aligned}(A \otimes B)(A \otimes B)^* &= (A \otimes B)(A^* \otimes B^*) = (AA^*) \otimes (BB^*)\\ &= (A^*A) \otimes (B^*B) = (A^* \otimes B^*)(A \otimes B)\\ &= (A \otimes B)^*(A \otimes B).\end{aligned}$$

(iii)

$$\begin{aligned}(A \otimes B)(A \otimes B)^* &= (A \otimes B)(A^* \otimes B^*) = (AA^*) \otimes (BB^*)\\ &= I \otimes I = I.\end{aligned}$$

(iv) $Ax=\lambda x, By=\mu y, x\neq 0,\ y\neq 0\Rightarrow$

$$\begin{aligned} x\otimes y\neq 0 \text{ 且 } (A\otimes B)(x\otimes y) &= (Ax)\otimes(By)=(\lambda x)\otimes(\mu y)\\ &=\lambda\mu(x\otimes y). \end{aligned}$$

(v) 据 Jordan 标准形或 Schur 酉三角化定理, 存在可逆阵 S 和 T 使得 $S^{-1}AS=R$ 和 $T^{-1}BT=G$ 为上三角矩阵, R 和 G 的对角元素分别为 $\lambda_1,\cdots,\lambda_m$ 和 $\mu_1,\cdots,\mu_n$, 且

$$(S\otimes T)^{-1}(A\otimes B)(S\otimes T)=(S^{-1}AS)\otimes(T^{-1}BT)=R\otimes G.$$

可见 $A\otimes B$ 相似于上三角矩阵 $R\otimes G$, 而 $R\otimes G$ 的对角元素是

$$\lambda_i\mu_j,\quad i=1,\cdots,m, j=1,\cdots,n.$$

(vi) 从 (v) 立即得到, 因为行列式等于特征值之积.

(vii) 设 $A=U\Sigma V$ 和 $B=W\Gamma Q$ 为奇异值分解, U,V,W,Q 为酉矩阵, $\Sigma=\operatorname{diag}(s_1,\cdots,s_m),\Gamma=\operatorname{diag}(t_1,\cdots,t_n)$. 则

$$A\otimes B=(U\Sigma V)\otimes(W\Gamma Q)=(U\otimes W)(\Sigma\otimes\Gamma)(V\otimes Q).$$

因为 $U\otimes W$ 和 $V\otimes Q$ 为酉矩阵, $\Sigma\otimes\Gamma$ 为对角矩阵且对角元素是 $s_it_j, i=1,\cdots,m, j=1,\cdots,n$. 我们有 $\operatorname{sv}(A\otimes B)=\{s_it_j\mid i=1,\cdots,m, j=1,\cdots,n\}$.

(viii) 从 (vii) 立即得到, 因为秩等于非零奇异值的个数. □

设 $f(x,y)=\sum\alpha_{st}x^sy^t$ 是一个二元多项式, 完全类似于 (v) 的证明, 我们有

$$\sigma(\Sigma\alpha_{st}A^s\otimes B^t)=\{f(\lambda_i,\mu_j)\mid i=1,\cdots,m, j=1,\cdots,n\}. \tag{2.1}$$

设 $A=(a_{ij}), B=(b_{ij})\in M_{m,n}$, 则 A 和 B 的 Hadamard 积定义为它们对应元素相乘, 并记作 $A\circ B$:

$$A\circ B=(a_{ij}b_{ij})\in M_{m,n}.$$

引理 2.3 设 $A,B\in M_n$, 则 $A\circ B$ 是 $A\otimes B$ 的位于 $1,n+2,2n+3,\cdots,n^2$ 行和列的主子矩阵.

证明 设 e_i 是 $\mathbb{R}^n$ 的第 i 个标准基向量, 即 e_i 的第 i 个分量为 1 其余分量为 0. 令

$$E=(e_1\otimes e_1,\cdots,e_n\otimes e_n).$$

设 $A=(a_{ij}), B=(b_{ij})$, 则

$$\begin{aligned} a_{ij}b_{ij}=(e_i^TAe_j)\otimes(e_i^TBe_j) &= (e_i\otimes e_i)^T(A\otimes B)(e_j\otimes e_j)\\ &= e_i^T[E^T(A\otimes B)E]e_j. \end{aligned}$$

可见 $A\circ B=E^T(A\otimes B)E$. $\square$

定理 2.4 (Schur) 设 $A,B\in M_n$. 若 A,B 半正定, 则 $A\circ B$ 半正定. 若 A,B 正定, 则 $A\circ B$ 正定.

证明 假设 A,B 半正定, 则 $A\otimes B$ 是 Hermite 矩阵, 并且由定理 2.2(v) 知 $A\otimes B$ 半正定. 由引理 2.3, $A\circ B$ 是 $A\otimes B$ 的一个主子矩阵, 所以 $A\circ B$ 半正定. 第二个结论可类似地证明. $\square$

给定矩阵 A, 我们用 vec$A$ 表示把 A 的各列依次连接起来所得的向量: 设 $A=(a_1,a_2,\cdots,a_n)\in M_{m,n}$, 则

$$\mathrm{vec}A=\begin{pmatrix}a_1\\a_2\\\vdots\\a_n\end{pmatrix}.$$

vec 运算有如下性质:

引理 2.5 (i) 设 $A\in M_{m,n},B\in M_{n,k},C\in M_{k,t}$, 则

$$\mathrm{vec}(ABC)=(C^T\otimes A)\mathrm{vec}B.$$

(ii) 存在一个只依赖于 m,n 的 mn 阶置换阵 $P(m,n)$ 使得

$$\mathrm{vec}X^T=P(m,n)\mathrm{vec}X,\ \text{对任何}\ X\in M_{m,n}\ \text{成立}. \tag{2.2}$$

证明 (i) 对于矩阵 G, 用 G_p 表示 G 的第 p 列. 设 $C=(c_{ij})$, 则

$$\begin{aligned}(ABC)_p=ABC_p&=A(\sum_{i=1}^k c_{ip}B_i)\\&=(c_{1p}A,c_{2p}A,\cdots,c_{kp}A)\mathrm{vec}B\\&=[(C_p)^T\otimes A]\mathrm{vec}B.\end{aligned}$$

因此

$$\mathrm{vec}(ABC)=\begin{pmatrix}(C_1)^T\otimes A\\\vdots\\(C_t)^T\otimes A\end{pmatrix}\mathrm{vec}B=(C^T\otimes A)\mathrm{vec}B.$$

(ii) 设 $E_{ij}\in M_{m,n}$ 是 (i,j) 位置的元素为 1 其他元素为 0 的矩阵. 定义

$$P(m,n)=\sum_{i=1}^m\sum_{j=1}^n E_{ij}\otimes E_{ij}^T. \tag{2.3}$$

因为 $X^T=\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}E_{ij}^TXE_{ij}^T$,

$$\begin{aligned}\mathrm{vec}(X^T)&=\sum_{i=1}^{m}\sum_{j=1}^{n}\mathrm{vec}(E_{ij}^TXE_{ij}^T)\\&=\sum_{i=1}^{m}\sum_{j=1}^{n}(E_{ij}\otimes E_{ij}^T)\mathrm{vec}X\\&=P(m,n)\mathrm{vec}X.\end{aligned}$$

显然 $P(m,n)^T=P(n,m)$, 且 $P(m,n)$ 为 0–1 矩阵. 对任意 $X\in M_{m,n}$ 有

$$\begin{aligned}\mathrm{vec}X=\mathrm{vec}(X^T)^T&=P(n,m)\mathrm{vec}(X^T)\\&=P(n,m)P(m,n)\mathrm{vec}X.\end{aligned}$$

所以 $P(n,m)P(m,n)=I_{mn}, P(n,m)=P(m,n)^T=P(m,n)^{-1}$. 这表明 $P(m,n)$ 为置换矩阵. □

因为 $\mathrm{vec}X\mapsto\mathrm{vec}(X^T)$ 是从 $\mathbb{C}^{mn}$ 到 $\mathbb{C}^{mn}$ 的线性映射, 由线性代数的一个基本原理知, 满足条件 2.2 的矩阵 $P(m,n)$ 是唯一的.

定理 2.6 设 $P(m,s)$ 和 $P(n,t)$ 由 (2.3) 定义, 则

$$B\otimes A=P(m,s)^T(A\otimes B)P(n,t) \tag{2.4}$$

对任何 $A\in M_{m,n},B\in M_{s,t}$ 成立. 当 $A\in M_n$, $B\in M_t$ 时,

$$B\otimes A=P(n,t)^T(A\otimes B)P(n,t). \tag{2.5}$$

证明 对 $X\in M_{n,t}$ 记 $Y=AXB^T$, 由引理 2.5 得

$$\mathrm{vec}Y=\mathrm{vec}(AXB^T)=(B\otimes A)\mathrm{vec}X, \tag{2.6}$$

$$\mathrm{vec}Y^T=\mathrm{vec}(BX^TA^T)=(A\otimes B)\mathrm{vec}X^T,$$

$$P(m,s)\mathrm{vec}Y=(A\otimes B)P(n,t)\mathrm{vec}X,$$

$$\mathrm{vec}Y=P(m,s)^T(A\otimes B)P(n,t)\mathrm{vec}X. \tag{2.7}$$

因为 (2.6) 和 (2.7) 对任何 X 成立, 比较这两个式子就得到 (2.4). (2.5) 是(2.4) 的特例. □

(2.5) 说明, 当 A 和 B 都是方阵时, $B\otimes A$ 与 $A\otimes B$ 置换相似.

§2.2 线性矩阵方程

设 $A_i \in M_{m,n}, B_i \in M_{s,t}, i = 1, \cdots, k,\ C \in M_{m,t}$ 给定. 考虑关于未知矩阵 $X \in M_{n,s}$ 的线性方程

$$A_1XB_1 + A_2XB_2 + \cdots + A_kXB_k = C. \tag{2.8}$$

对 (2.8) 两边作 vec 运算, 我们看到这个矩阵方程等价于线性方程组

$$(B_1^T \otimes A_1 + B_2^T \otimes A_2 + \cdots + B_k^T \otimes A_k)\text{vec}X = \text{vec}C.$$

下面研究 (2.8) 的一个特例. 设 $A \in M_m,\ B \in M_n, C \in M_{m,n}$ 给定, 关于未知矩阵 $X \in M_{m,n}$ 的方程

$$AX - XB = C \tag{2.9}$$

称为 Sylvester 方程. 这个方程可以写成线性方程组

$$(I_n \otimes A - B^T \otimes I_m)\text{vec}X = \text{vec}C. \tag{2.10}$$

定理 2.7 矩阵方程 (2.9) 有唯一解当且仅当 A 和 B 没有公共特征值.

证明 设 $\sigma(A) = \{\lambda_1, \cdots, \lambda_m\}, \sigma(B) = \{\mu_1, \cdots, \mu_n\}$. 则线性方程组(2.10) 的系数矩阵的特征值是

$$\sigma(I_n \otimes A - B^T \otimes I_m) = \{\lambda_i - \mu_j \mid i = 1, \cdots, m,\ j = 1, \cdots, n\}.$$ □

定理 2.8 (Roth [97]) 方程 (2.9) 有解当且仅当

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \text{和} \begin{pmatrix} A & C \\ 0 & B \end{pmatrix} \tag{2.11}$$

相似.

证明 (Flanders-Wimmer [45]) 若方程有解 X, 则

$$\begin{pmatrix} I & X \\ 0 & I \end{pmatrix}^{-1} \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & C \\ 0 & B \end{pmatrix}.$$

反过来, 设 (2.11) 中的两个矩阵相似. 定义两个线性变换 $f_i : M_{m+n} \to M_{m+n}$, $i = 1, 2$,

$$f_1(G) = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} G - G \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix},$$

$$f_2(G) = \begin{pmatrix} A & C \\ 0 & B \end{pmatrix} G - G \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}.$$

由相似性假设, $\dim\ker f_1 = \dim\ker f_2$. 直接计算知

$$\ker f_1 = \left\{ \begin{pmatrix} P & Q \\ R & S \end{pmatrix} \middle| \begin{matrix} AP = PA,\ AQ = QB, \\ BR = RA,\ BS = SB \end{matrix} \right\},$$

$$\ker f_2 = \left\{ \begin{pmatrix} P & Q \\ R & S \end{pmatrix} \middle| \begin{matrix} AP + CR = PA, & AQ + CS = QB, \\ BR = RA, & BS = SB \end{matrix} \right\}.$$

为了证明方程 (2.9) 有解, 我们只要说明 $\ker f_2$ 中有一个形如

$$\begin{pmatrix} P & Q \\ 0 & -I \end{pmatrix}$$

的矩阵, 此时 $AQ - C = QB$. 记

$$V = \{(R, S) \mid BR = RA, BS = SB\}.$$

这是 $M_{n,m+n}$ 的一个子空间. 定义 $g_i : \ker f_i \to V, i = 1, 2$,

$$g_i\left(\begin{pmatrix} P & Q \\ R & S \end{pmatrix}\right) = (R, S).$$

则

$$\ker g_1 = \ker g_2 = \left\{ \begin{pmatrix} P & Q \\ 0 & 0 \end{pmatrix} \middle| AP = PA, AQ = QB \right\}.$$

我们将证明 $\mathrm{Im} g_1 = \mathrm{Im} g_2$. 显然 $\mathrm{Im} g_1 = V$, 因为若 $BR = RA, BS = SB$ 则

$$\begin{pmatrix} 0 & 0 \\ R & S \end{pmatrix} \in \ker f_1, \quad g_1\left(\begin{pmatrix} 0 & 0 \\ R & S \end{pmatrix}\right) = (R, S).$$

所以 $\mathrm{Im} g_2 \subseteq \mathrm{Im} g_1$. 但由维数定理

$$\dim\ker g_i + \dim \mathrm{Im} g_i = \dim\ker f_i, \quad i = 1, 2.$$

因此 $\dim \mathrm{Im} g_1 = \dim \mathrm{Im} g_2$. 再从 $\mathrm{Im} g_2 \subseteq \mathrm{Im} g_1$ 得 $\mathrm{Im} g_2 = \mathrm{Im} g_1$. 显然,

$$(0, -I) \in V = \mathrm{Im} g_1 = \mathrm{Im} g_2.$$

所以, 存在

$$\begin{pmatrix} P & Q \\ R & S \end{pmatrix} \in \ker f_2 \quad 满足 \quad g_2\left(\begin{pmatrix} P & Q \\ R & S \end{pmatrix}\right) = (R, S) = (0, -I),$$

即 $\begin{pmatrix} P & Q \\ 0 & -I \end{pmatrix} \in \ker f_2$. □

定理 2.9 (Bhatia-Davis-McIntosh [21]) 设 $A, B, C \in M_n$. 如果存在正数 r 使得

$$\sigma(B) \subseteq \{z \in \mathbb{C} : |z| < r\}, \quad \sigma(A) \subseteq \{z \in \mathbb{C} : |z| > r\},$$

则关于 X 的方程 $AX - XB = C$ 的解是

$$X = \sum_{k=0}^{\infty} A^{-k-1} C B^k. \tag{2.12}$$

证明 (Bhatia [15]) 只需要证明这个级数收敛, 然后容易验证它是方程的解.

选取 $r_1 < r < r_2$ 使得

$$\sigma(B) \subseteq \{z \in \mathbb{C} : |z| < r_1\}, \quad \sigma(A) \cap \{z \in \mathbb{C} : |z| \leqslant r_2\} = \varnothing.$$

则 $\sigma(A^{-1}) \subseteq \{z \in \mathbb{C} : |z| < r_2^{-1}\}$. 用 $\|\cdot\|$ 表示矩阵的谱范数. 根据 Gelfand 的谱半径公式 (第一章习题 5), 存在正整数 N 使得当 $k \geqslant N$ 时

$$\|B^k\| \leqslant r_1^k, \quad \|A^{-k}\| \leqslant r_2^{-k}.$$

因此, 对于 $k \geqslant N$,

$$\|A^{-k-1} C B^k\| \leqslant (r_1/r_2)^k \|A^{-1} C\|.$$

所以 (2.12) 中的级数收敛. □

M_n 上的范数 $\|\cdot\|$ 称为酉不变的, 若 $\|UAV\| = \|A\|$ 对任意的 $A \in M_n$ 和任意的酉矩阵 $U, V \in M_n$ 成立. 每个酉不变范数 $\|\cdot\|$ 都满足性质

$$\|ABC\| \leqslant \|A\|_\infty \|C\|_\infty \|B\|, \quad \forall A, B, C \in M_n,$$

见第四章习题 11.

定理 2.10 (Bhatia-Davis-McIntosh [21]) 设 $A, B, C \in M_n$, A, B 为正规矩阵. 如果存在复数 a 和正数 r, d 使得

$$\sigma(B) \subseteq \{z \in \mathbb{C} : |z - a| \leqslant r\}, \quad \sigma(A) \subseteq \{z \in \mathbb{C} : |z - a| \geqslant r + d\},$$

则对于任何酉不变范数 $\|\cdot\|$, 方程 $AX - XB = C$ 的解 X 满足

$$d\|X\| \leqslant \|C\|.$$

证明 通过将 A, B 分别换成 $A - aI$, $B - aI$, 我们可以假设 $a = 0$. 应用定理 2.9 知方程的解有级数表示 (2.12). 注意到对于正规矩阵, 谱半径等于谱范数, 我们有

$$\begin{aligned}
\|X\| &\leqslant \sum_{k=0}^{\infty} \|A^{-1}\|_\infty^{k+1} \|B\|_\infty^k \|C\| \\
&\leqslant \|C\| \sum_{k=0}^{\infty} (r + d)^{-k-1} r^k \\
&= \|C\|/d.
\end{aligned}$$

□

一个复方阵 A 称为是正稳定的, 如果 A 的每个特征值的实部都是正数.

定理 2.11 (Lyapunov) 设 $A \in M_n$ 正稳定, $P \in M_n$ 正定. 则方程

$$AX + XA^* = P \tag{2.13}$$

有唯一解 X 并且这个解是正定的.

证明 注意到方程 (2.13) 是 Sylvester 方程 (2.9) 的特殊情形. 因为 A 正稳定, $\sigma(A) \bigcap \sigma(-A^*) = \varnothing$. 根据定理 2.7, 方程 (2.13) 有唯一解 X. 在 (2.13) 两边取共轭转置得 $AX^* + X^*A^* = P$. 这说明 X^* 也是方程的一个解. 由解的唯一性知 $X^* = X$, 即 X 是 Hermite 矩阵. 为证明 X 正定, 我们还需要证明 X 的特征值都是正数.

我们先证明存在一个正定矩阵 $Y \in M_n$ 使得 $AY + YA^*$ 正定. 考虑分解 $T^{-1}AT = J$, 其中 J 是 A 的 Jordan 标准形. 选取一个正数 ϵ 使得 ϵ 小于 A 的每个特征值的实部, 令 $D = \mathrm{diag}(1, \epsilon, \epsilon^2, \cdots, \epsilon^{n-1})$. 则矩阵

$$D^{-1}JD + (D^{-1}JD)^* \tag{2.14}$$

实对称、每个对角元素为正数、对角占优. 用圆盘定理 (第一章习题 11) 知这样的矩阵正定. 将 $J = T^{-1}AT$ 代入 (2.14) 并记 $G = TD$, 则 (2.14) 成为 $G^{-1}AG + (G^{-1}AG)^*$. 用可逆阵 G 对这个矩阵作合同变换并且令 $Y = GG^*$ 我们就得到正定矩阵 $AY + YA^*$. 显然 Y 正定. 记

$$AY + YA^* = Q.$$

考虑矩阵族

$$X(t) = tX + (1-t)Y, \quad t \in [0, 1].$$

对于每个 t, $X(t)$ 都是 Hermite 矩阵因而其特征值为实数. $X(0) = Y$ 的特征值都是正数. 假设 $X(1) = X$ 至少有一个特征值不是正的. 因为 $X(t)$ 的特征值连续地依赖于 t, 存在 $t_0 \in (0, 1]$ 使得 $X(t_0)$ 有一个特征值为零, 即 $X(t_0)$ 奇异. 从而 $AX(t_0)$ 奇异. 但是 $AX(t_0)$ 的实部

$$\begin{aligned}\mathrm{Re}[AX(t_0)] &= [AX(t_0) + (AX(t_0))^*]/2 \\ &= [t_0P + (1-t_0)Q]/2\end{aligned}$$

是一个正定矩阵, 因此 $AX(t_0)$ 是正稳定矩阵从而非奇异. 矛盾. 所以 $X = X(1)$ 的特征值都是正数. □

(2.13) 中的矩阵方程称为 Lyapunov 方程.

§2.3 Frobenius-König 定理

本节给出矩阵的几个关于行列与对角线关系的组合性质. 说它们是组合性质因为我们只关心矩阵的元素是否为 0. 实际上我们也可以把一个矩阵的元素分为两类, 一类称为*白元素*, 另一类称为*黑元素*.

设 $A \in M_{m,n}$. 我们把 A 的一行或一列称为一条*线*. A 的两两不共线的非零元的最大个数称为 A 的*项秩*, 记作 $\tau(A)$. 一个线的集合称为*覆盖* A, 若该集合的线包含了 A 的所有非零元素. 一个线的集合若能覆盖 A, 就称这个集合为 A 的一个*覆盖*. 能覆 A 的线的最少条数称为 A 的*线秩*, 记作 $\delta(A)$. A 的项秩和线秩都不会超过 $\min\{m,n\}$.

定理 2.12 (König) 对任意的 $A \in M_{m,n}, \delta(A) = \tau(A)$.

证明 对 A 的线数 $m+n$ 作数学归纳法. 当 $m=1$ 或 $n=1$ 时, 结论成立.

下面设 $m \geqslant 2, n \geqslant 2$ 并且假设定理对于线数 $< m+n$ 的矩阵正确. 按定义显然有 $\delta(A) \geqslant \tau(A)$. 现在来证 $\delta(A) \leqslant \tau(A)$.

A 的一个覆盖称为*未满覆盖*, 如果它不含有 A 的所有行也不含有 A 的所有列. 我们分两种情况:

情形 (1): A 没有未满覆盖. 此时 $\delta(A) = \min\{m,n\}$. 设 $A = (a_{ij}), a_{rs} \neq 0$. 记划去 A 的第 r 行和第 s 列后所得矩阵为 $A' \in M_{m-1,n-1}$, 则 $\delta(A') = \delta(A) - 1$. 首先, $\delta(A') \leqslant \delta(A) - 1 = \min\{m-1, n-1\}$. 其次, 假如 $\delta(A') \leqslant \delta(A) - 2 = \min\{m-2, n-2\}$, 则将 A' 的具有 $\delta(A')$ 条线的覆盖加上划掉的第 r 行和第 s 列将得到 A 的一个未满覆盖, 这与 A 没有未满覆盖矛盾. 由归纳假设, $\tau(A') = \delta(A') = \delta(A) - 1$, 添上元素 a_{rs} 我们得 $\delta(A) \leqslant \tau(A)$.

情形 (2): A 有一个未满覆盖. 此时 A 有一个覆盖由 p 行和 q 列组成, $p + q = \delta(A), p < m, q < n$. 置换行和列不会改变项秩和线秩. 置换 A 的行和列使那 p 行成为头 p 行且使那 q 列成为头 q 列, A 变成了

$$\begin{pmatrix} * & B \\ C & 0 \end{pmatrix},$$

其中 $B \in M_{p,n-q}, C \in M_{m-p,q}$. 我们有 $\delta(B) = p$, $\delta(C) = q$, 因为只要有一个不成立就会与条件 $p + q = \delta(A)$, $p < m, q < n$ 矛盾. 对 B 和 C 用归纳假设, 得 $\tau(B) = p$, $\tau(C) = q$. 因此 $\tau(A) \geqslant \tau(B) + \tau(C) = p + q = \delta(A)$. □

对于正整数 $m \leqslant n$, 用 $S_{m,n}$ 表示从 $\{1,2,\cdots,m\}$ 到 $\{1,2,\cdots,n\}$ 的单射的集合. 设 $A = (a_{ij}) \in M_{m,n}, m \leqslant n$. 若 $\sigma \in S_{m,n}$, 则 $a_{1\sigma(1)}, a_{2\sigma(2)}, \cdots, a_{m\sigma(m)}$ 称为 A 的

一条广义对角线, 简称为对角线. A 的积和式定义为

$$\mathrm{per} A = \sum_{\sigma \in S_{m,n}} \prod_{i=1}^{m} a_{i\sigma(i)}.$$

一个 r 行 s 列的矩阵称为一个 $r \times s$ 的矩阵.

定理 2.13 (Frobenius-König) 设 $A \in M_{m,n}, m \leqslant n$, 则 A 的每条对角线都至少含有一个零元素当且仅当 A 有一个 $r \times s$ 的零子矩阵, $r+s=n+1$.

证明 假设 A 的每条对角线都至少含有一个零元素, 则由定理 2.12 知 $\delta(A) = \tau(A) \leqslant m-1$. 因此 A 可以被 $m-1$ 条线覆盖, 划掉那 $m-1$ 条线我们就得到 A 的一个零子矩阵, 其行列之和 $= m+n-(m-1) = n+1$.

反过来, 若 A 有一个 $r\times s$ 的零子矩阵, $r+s=n+1$, 则 A 可以被$m+n-(n+1) = m-1$ 条线覆盖, 因此 $\tau(A) = \delta(A) \leqslant m-1$. 从而$A$ 的每条对角线都至少含有一个零元. □

每个元素都是非负实数的矩阵称为非负矩阵. 显然定理 2.13 可以等价地叙述为

定理 2.14 (Frobenius-König) 设 $A \in M_{m,n}$ 为非负矩阵, $m \leqslant n$. 则 $\mathrm{perA} = 0$ 当且仅当 A 有一个 $r \times s$ 的零子矩阵, $r+s=n+1$.

定理 2.13 可以推广为下面的结果.

定理 2.15 (König) 设 $A \in M_{m,n}, m \leqslant n$. 则 A 的每条对角线都至少含有 k 个零元素当且仅当 A 有一个 $r \times s$ 的零子矩阵, $r+s=n+k$.

证明 (Aharoni [1]) $k=1$ 的情形就是定理 2.13. 现在设 $k \geqslant 2$, 设 J 是 $m \times (k-1)$ 的元素全为 1 的矩阵. 令 $B=(A,J)$. 则 B 是 $m \times (n+k-1)$ 的. 假设 A 的每条对角线都至少含有 k 个零元素, 则 B 的每条对角线都至少含有一个零元素. 根据定理 2.13, B 有一个 $r \times s$ 的零子矩阵, $r+s=(n+k-1)+1=n+k$. 这个零子阵只可能在 A 中.

反过来, 若 A 有一个 $r \times s$ 的零子矩阵, $r+s=n+k=(n+k-1)+1$. 根据定理 2.13, B 的每条对角线都至少含有一个零元素, 这推出 A 的每条对角线都至少含有 k 个零元素. 实际上, 假如 A 的某条对角线只含有 t 个零元素, $t \leqslant k-1$, 则这条对角线上的 $m-t$ 个非零元素与 J 中适当选取的 t 个元素合起来成为 B 的一条不含零元的对角线. □

§2.4 复合矩阵

设 $A \in M_{m,n}, 1 \leqslant i_1 < \cdots < i_s \leqslant m, 1 \leqslant j_1 < \cdots < j_t \leqslant n$. 我们用 $A[i_1,\cdots,i_s \mid j_1,\cdots,j_t]$ 记 A 中位于 $i_1,\cdots,i_s$ 行和 $j_1,\cdots,j_t$ 列的元素组成的 $s \times t$ 的子矩阵.

用 $A(i_1,\cdots,i_s \mid j_1,\cdots,j_t)$ 记从 A 中划去 $i_1,\cdots,i_s$ 行和 $j_1,\cdots,j_t$ 列后所得到的 $(m-s)\times(n-t)$ 的子矩阵. 用 $A[i_1,\cdots,i_s \mid j_1,\cdots,j_t)$ 记从 $A[i_1,\cdots,i_s \mid 1,\cdots,n]$ 中划去 $j_1,\cdots,j_t$ 列后所得到的 $s\times(n-t)$ 的子矩阵. 用 $A(i_1,\cdots,i_s \mid j_1,\cdots,j_t]$ 记从 $A[1,\cdots,m \mid j_1,\cdots,j_t]$ 中划去 $i_1,\cdots,i_s$ 行后所得到的 $(m-s)\times t$ 的子矩阵.

对于正整数 $k\leqslant n$, 定义指标集合

$$\Gamma(k,n)=\{(i_1,\cdots,i_k) \mid 1\leqslant i_1<\cdots<i_k\leqslant n\}.$$

为了简明, 我们只对方阵定义复合矩阵, 本书只用到这种情形. 长方阵的复合矩阵可以类似地定义. 将 $\Gamma(k,n)$ 中的元素按字典顺序排列为 $\alpha_1,\alpha_2\cdots,\alpha_{\binom{n}{k}}$, 例如 $\Gamma(2,4)$ 的元素排列为 $(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)$. 对于 $A\in M_n, 1\leqslant k\leqslant n$, A 的 k 级复合矩阵, 记作 $C_k(A)$, 是一个 $\binom{n}{k}$ 阶矩阵, $C_k(A)$ 的 (i,j) 位置的元素定义为 $\det A[\alpha_i \mid \alpha_j], 1\leqslant i,j\leqslant\binom{n}{k}$.

例如对于 $A\in M_3$,

$$C_2(A)=\begin{pmatrix}\det A[1,2 \mid 1,2] & \det A[1,2 \mid 1,3] & \det A[1,2 \mid 2,3]\\ \det A[1,3 \mid 1,2] & \det A[1,3 \mid 1,3] & \det A[1,3 \mid 2,3]\\ \det A[2,3 \mid 1,2] & \det A[2,3 \mid 1,3] & \det A[2,3 \mid 2,3]\end{pmatrix}.$$

引理 2.6 设 $A\in M_n$ 是上三角矩阵, 则 $C_k(A)$ 也是上三角矩阵, 并且 $C_k(A)$ 的对角元素是从 A 的对角元素中取 k 个作乘积所得的所有乘积.

证明 对于 $\binom{n}{k}\geqslant i>j\geqslant 1$, 设 $\alpha_i=(i_1,\cdots,i_{s-1},i_s,\cdots,i_k), \alpha_j=(i_1,\cdots,i_{s-1},j_s,\cdots,j_k), i_s>j_s$ 则 $A[\alpha_i \mid \alpha_j]$ 有一个 $p\times q$ 的零子矩阵, $p+q=k+1$:

$$(A[\alpha_i \mid \alpha_j])[s,s+1,\cdots,k \mid 1,2,\cdots,s]=0.$$

由 Frobenius-König 定理, $C_k(A)(i,j)=\det A[\alpha_i \mid \alpha_j]=0$, 所以 $C_k(A)$ 为上三角矩阵. 对于 $1\leqslant i\leqslant\begin{pmatrix}n\\k\end{pmatrix}$, $A[\alpha_i \mid \alpha_i]$ 为上三角矩阵, 对角元素是 $a_{i_1,i_1},\cdots,a_{i_k,i_k}$, 因此,

$$C_k(A)(i,i)=\det A[\alpha_i \mid \alpha_i]=\prod_{t=1}^{k}a_{i_t,i_t}. \qquad \square$$

类似地可证明, 若 A 是下三角矩阵, 则 $C_k(A)$ 也是下三角矩阵. 从而若 A 是对角矩阵, 则 $C_k(A)$ 也是对角矩阵. 复合矩阵有下面的性质. 我们用 $\sigma(A)$ 表示 A 的所有特征值的集合, $\mathrm{sv}(A)$ 表示 A 的所有奇异值的集合. 当 $r<k$ 时约定 $\binom{r}{k}=0$.

定理 2.16 设 $A,B\in M_n, 1\leqslant k\leqslant n$. 则

(i) $(C_k(A))^T=C_k(A^T), (C_k(A))^*=C_k(A^*)$.

(ii) $C_k(AB)=C_k(A)C_k(B)$.

(iii) 若 A 可逆, 则 $C_k(A)$ 也可逆, 且 $(C_k(A))^{-1}=C_k(A^{-1})$.

(iv) 若 $\sigma(A)=\{\lambda_1,\cdots,\lambda_n\}$, 则 $\sigma(C_k(A))=\{\lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_k}\mid 1\leqslant i_1<i_2<\cdots<i_k\leqslant n\}$.

(v) 若 $\mathrm{sv}(A)=\{s_1,\cdots,s_n\}$, 则 $\mathrm{sv}(C_k(A))=\{s_{i_1}s_{i_2}\cdots s_{i_k}\mid 1\leqslant i_1<i_2<\cdots<i_k\leqslant n\}$.

(vi) 若 $\mathrm{rank}(A)=r$, 则 $\mathrm{rank}(C_k(A))=\binom{r}{k}$.

(vii) $\det C_k(A)=(\det A)^{\binom{n-1}{k-1}}$.

(viii)若 A 是正规, Hermite, 酉, 半正定, 正定, 对称矩阵, 则 $C_k(A)$ 也是.

证明　(i) 显然.

(ii) 由 Binet-Cauchy 公式, 有

$$
\begin{aligned}
[C_k(AB)](i,j)&=\det(AB)[\alpha_i\mid\alpha_j]\\
&=\sum_{\beta\in\Gamma(k,n)}\det A[\alpha_i\mid\beta]\det B[\beta\mid\alpha_j]=[C_k(A)C_k(B)](i,j).
\end{aligned}
$$

(iii) $C_k(A)C_k(A^{-1})=C_k(AA^{-1})=C_k(I_n)=I_{\binom{n}{k}}$.

(iv) 根据 Jordan 标准形或 Schur 定理, 存在可逆矩阵 G 使得 $A=G^{-1}RG$, 其中 R 为上三角矩阵, 对角元素是 $\lambda_1,\cdots,\lambda_n$. 用 (ii) 和 (iii) 得

$$C_k(A)=C_k(G^{-1})C_k(R)C_k(G)=[C_k(G)]^{-1}C_k(R)C_k(G).$$

由引理 2.6, $C_k(R)$ 为上三角矩阵, 其对角元素是 $\lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_k},1\leqslant i_1<i_2<\cdots<i_k\leqslant n$. 于是

$$\sigma(C_k(A))=\sigma(C_k(R))=\{\lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_k}\mid 1\leqslant i_1<i_2<\cdots<i_k\leqslant n\}.$$

(v) 由 (i) 和 (ii) 得

$$[C_k(A)]^*C_k(A)=C_k(A^*A).$$

将 (iv) 用于这个等式即得 (v).

(vi) 这条性质可由 (v) 得到, 因为秩等于非零奇异值的个数.

(vii) 由 (iv) 得到.

(viii) 由上述某些性质得到. □

后面我们将看到复合矩阵的应用.

习　题

1. 对于怎样的 $A\in M_m,B\in M_n,A\otimes B=I$?

2. 给出定理 2.4 的另一个证明.

3. 设 $A,B\in M_n,A$ 正定, B 半正定且对角元素都是正数, 则 $A\circ B$ 正定.

4. 设 $A=\mathrm{diag}(A_1,\cdots,A_k)\in M_n$, 其中 $A_i\in M_{n_i}$ 且 $\sigma(A_i)\bigcap\sigma(A_j)=\varnothing, i\neq j$. 若 $B\in M_n$ 且 $AB=BA$, 则 $B=\mathrm{diag}(B_1,\cdots,B_k)\in M_n$, 其中 $B_i\in M_{n_i}$.

5. 设 $A\in M_m, B\in M_n, C\in M_{m,n}$. 若 $\sigma(A)\bigcap\sigma(B)=\varnothing$, 则

$$\begin{pmatrix} A & C \\ 0 & B \end{pmatrix} \text{和} \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \text{相似.}$$

6. (Embry [34]) 我们说两个矩阵 X, Y 可交换是指乘法可交换, 即 $XY=YX$. 设 $A,B\in M_n$ 满足 $\sigma(A)\bigcap\sigma(B)=\varnothing$. 如果 $C\in M_n$, C 与 $A+B$ 可交换并且 C 与 AB 可交换, 则 C 与 A 和 B 都可交换.

7. (Marcus-Ree [83]) 一个非负方阵称为是双随机的, 若它的每行元素之和等于 1, 且它的每列元素之和也等于 1. 设 $A=(a_{ij})$ 为 n 阶双随机矩阵. 则存在 $1,2,\cdots,n$ 的一个排列 σ 使得对每个 $i=1,\cdots,n$,

$$a_{i\sigma(i)} \geqslant \begin{cases} \dfrac{1}{k(k+1)}, & n=2k, \\ \dfrac{1}{(k+1)^2}, & n=2k+1. \end{cases}$$

8. 设 $k\leqslant m\leqslant n$. 怎样的矩阵 $A\in M_{m,n}$ 的每条对角线恰好含有 k 个零元素?

9. 记 $m=\begin{pmatrix} n \\ k \end{pmatrix}$. 复合矩阵映射 $C_k(\cdot): M_n\to M_m$ 是单射吗? 是满射吗?

第三章　Hermite 矩阵和优超关系

Hermite 矩阵是实对称矩阵在复矩阵情形的推广, 它们的特征值有很好的性质.

两个实向量之间的优超关系出现在许多地方, 例如 Hermite 矩阵的对角元素被它的特征值优超. 优超理论也是推导不等式的一个有力工具, 例如它使我们对于 Hadamard 关于半正定矩阵行列式的不等式有更好的理解.

用 H_n 表示 n 阶 Hermite 矩阵的集合, 这是 $\mathbb{R}$ 上的一个向量空间.

§3.1 Hermite 矩阵的特征值

对于 $A \in H_n$, 我们总是将它的特征值按降序排列并记为 $\lambda_1(A) \geqslant \lambda_2(A) \geqslant \cdots \geqslant \lambda_n(A)$.

定理 3.1 (极小极大原理, Courant-Fischer)　设 $A \in H_n, 1 \leqslant k \leqslant n$. 则

$$\lambda_k(A) = \max_{\substack{S\subseteq\mathbb{C}^n \\ \dim S=k}} \min_{\substack{x\in S \\ \|x\|=1}} x^*Ax = \min_{\substack{S\subseteq\mathbb{C}^n \\ \dim S=n-k+1}} \max_{\substack{x\in S \\ \|x\|=1}} x^*Ax,$$

这里 S 是 $\mathbb{C}^n$ 的子空间, $\|\cdot\|$ 是欧氏范数.

证明　记 $\lambda_j(A)$ 为 $\lambda_j, j=1,\cdots,n$. 设 $\dim S = k, u_1,\cdots,u_n$ 是对应于 $\lambda_1,\cdots,\lambda_n$ 的 A 的标准正交的特征向量, 记 $T = \mathrm{span}\{u_k,\cdots,u_n\}$. 则

$$\dim S + \dim T = n+1.$$

所以

$$\begin{aligned}\dim(S\bigcap T) &= \dim S + \dim T - \dim(S+T) \\ &\geqslant n+1-n = 1.\end{aligned}$$

任取一个单位向量 $x \in S \bigcap T$. 设

$$x = \sum_{j=k}^{n} \xi_j u_j, \quad \sum_{j=k}^{n} |\xi_j|^2 = 1.$$

则

$$x^* A x = \sum_{j=k}^{n} |\xi_j|^2 \lambda_j \leqslant \sum_{j=k}^{n} |\xi_j|^2 \lambda_k = \lambda_k.$$

因此

$$\min_{\substack{x \in S \\ \|x\|=1}} x^* A x \leqslant \lambda_k.$$

另一方面, 取 $S_0 = \operatorname{span}\{u_1, \cdots, u_k\}$, 有

$$\min_{\substack{x \in S_0 \\ \|x\|=1}} x^* A x = \lambda_k.$$

这就证明了

$$\lambda_k = \max_{\substack{S \subseteq \mathbb{C}^n \\ \dim S = k}} \min_{\substack{x \in S \\ \|x\|=1}} x^* A x.$$

将第一个等式中的 A 换成 $-A$ 我们可以得到第二个等式. □

我们将定理 3.1 中的第一个等式称为极小极大表示, 将第二个等式称为极大极小表示. 定理 3.1 的两个重要特殊情形是

$$\lambda_1(A) = \max_{\substack{x \in \mathbb{C}^n \\ \|x\|=1}} x^* A x, \quad \lambda_n(A) = \min_{\substack{x \in \mathbb{C}^n \\ \|x\|=1}} x^* A x.$$

定理 3.2 (Cauchy 分隔定理) 设 $A \in H_n$ 的特征值是 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. B 是 A 的一个 m 阶主子矩阵, B 的特征值是 $\mu_1 \geqslant \cdots \geqslant \mu_m$. 则

$$\lambda_j \geqslant \mu_j \geqslant \lambda_{j+n-m}, \quad j = 1, 2, \cdots, m.$$

证明 若有必要, 对 A 作置换相似, 不妨设

$$A = \begin{pmatrix} B & C \\ C^* & D \end{pmatrix}.$$

由极小极大原理, 存在子空间 $S \subseteq \mathbb{C}^n, \dim S = j$ 满足

$$\mu_j = \min_{\substack{x \in S \\ \|x\|=1}} x^* B x.$$

对于 $x \in \mathbb{C}^m$, 记 $\widetilde{x} = \begin{pmatrix} x \\ 0 \end{pmatrix} \in \mathbb{C}^n$. 令 $\widetilde{S} = \{\widetilde{x} \mid x \in S\}$, 则 $x^*Bx = \widetilde{x}^*A\widetilde{x}$. 我们有

$$\mu_j = \min_{\widetilde{x} \in \widetilde{S}} \widetilde{x}^*A\widetilde{x} \leqslant \max_{\substack{T \subseteq \mathbb{C}^n \\ \dim T = j}} \min_{\substack{y \in T \\ \|y\|=1}} y^*Ay = \lambda_j.$$

应用这个结论于 $-A, -B$, 注意到

$$\begin{aligned} -\lambda_i(A) &= \lambda_{n-i+1}(-A) \ , \quad 1 \leqslant i \leqslant n, \\ -\lambda_j(B) &= \lambda_{m-j+1}(-B) \ , \quad 1 \leqslant j \leqslant m, \end{aligned}$$

取 $i = j + n - m$, 有

$$-\lambda_{j+n-m}(A) \geqslant -\lambda_j(B), \quad \text{i.e.,} \quad \mu_j = \lambda_j(B) \geqslant \lambda_{j+n-m}(A) = \lambda_{j+n-m}.$$ □

定理 3.3 (Weyl) 设 $A, B \in H_n$, 则对于 $1 \leqslant j \leqslant n$,

$$\max_{r+s=j+n} \{\lambda_r(A) + \lambda_s(B)\} \leqslant \lambda_j(A+B) \leqslant \min_{r+s=j+1} \{\lambda_r(A) + \lambda_s(B)\}.$$

证明 先用极小极大表示证明第一个不等式. 设 $r + s = j + n$, 存在子空间 $R, S \subseteq \mathbb{C}^n$ 满足 $\dim R = r, \dim S = s$,

$$\lambda_r(A) = \min_{\substack{x \in R \\ \|x\|=1}} x^*Ax, \quad \lambda_s(B) = \min_{\substack{x \in S \\ \|x\|=1}} x^*Bx.$$

因为

$$\begin{aligned} \dim(R \bigcap S) &= \dim R + \dim S - \dim(R+S) \\ &\geqslant r + s - n = j, \end{aligned}$$

存在子空间 $T_0 \subseteq R \bigcap S$ 满足 $\dim T_0 = j$. 于是

$$\begin{aligned} \lambda_j(A+B) &= \max_{\substack{T \subseteq \mathbb{C}^n \\ \dim T = j}} \min_{\substack{x \in T \\ \|x\|=1}} x^*(A+B)x \\ &\geqslant \min_{\substack{x \in T_0 \\ \|x\|=1}} (x^*Ax + x^*Bx) \\ &\geqslant \min_{\substack{x \in T_0 \\ \|x\|=1}} x^*Ax + \min_{\substack{x \in T_0 \\ \|x\|=1}} x^*Bx \\ &\geqslant \min_{\substack{x \in R \\ \|x\|=1}} x^*Ax + \min_{\substack{x \in S \\ \|x\|=1}} x^*Bx \\ &= \lambda_r(A) + \lambda_s(B). \end{aligned}$$

应用第一个不等式于 $\lambda_{n-j+1}(-A-B)$ 就得到了第二个不等式. □

下面的推论描述一个 Hermite 矩阵变成另一个 Hermite 矩阵时特征值变化的估计, 这类结果称为扰动定理.

推论 3.4 (Weyl)　设 $A,B\in H_n$, 则

$$\max_{1\leqslant j\leqslant n}|\lambda_j(A)-\lambda_j(B)|\leqslant\|A-B\|_\infty.$$

证明　由定理 3.3,

$$\lambda_j(A)=\lambda_j(B+A-B)\leqslant\lambda_j(B)+\lambda_1(A-B).$$

所以

$$\lambda_j(A)-\lambda_j(B)\leqslant\lambda_1(A-B)\leqslant\|A-B\|_\infty.$$

交换 A 与 B 的角色, 有

$$\lambda_j(B)-\lambda_j(A)\leqslant\|A-B\|_\infty.$$

从而

$$|\lambda_j(A)-\lambda_j(B)|\leqslant\|A-B\|_\infty.$$ □

对于 $A,B\in H_n$, 我们用记号 $A\leqslant B$ 或 $B\geqslant A$ 表示 $B-A$ 半正定. 显然 " $\leqslant$ " 和 " $\geqslant$ " 定义了 H_n 上的两个偏序, 统称为 Löwner 偏序. 特别, $B\geqslant 0$ 表示 B 半正定.

下面的结果是定理 3.3 的直接推论 $(A=B+(A-B))$, 也可以从极小极大表示看出.

推论 3.5 (Weyl 单调性原理)　设 $A,B\in H_n$, 若 $A\geqslant B$, 则

$$\lambda_j(A)\geqslant\lambda_j(B),\quad j=1,\cdots,n.$$

设 $A=U^*\mathrm{diag}(\lambda_1,\cdots,\lambda_n)U$ 为 $A\in H_n$ 的谱分解. U 为酉矩阵, 且 $\lambda_1\geqslant\cdots\geqslant\lambda_p\geqslant 0>\lambda_{p+1}\geqslant\cdots\geqslant\lambda_n$. 记

$$A_+=U^*\mathrm{diag}(\lambda_1,\cdots,\lambda_p,0,\cdots,0)U,$$

$$A_-=U^*\mathrm{diag}(0,\cdots,0,-\lambda_{p+1},\cdots,-\lambda_n)U.$$

则 $A_+\geqslant 0, A_-\geqslant 0$ 且 $A=A_+-A_-$. 这称为 A 的 Jordan 分解.

定理 3.6　设 $A\in H_n$ 有分解 $A=P-Q$, 其中 $P\geqslant 0,Q\geqslant 0$. 则

$$\lambda_j(A_+)\leqslant\lambda_j(P),\quad\lambda_j(A_-)\leqslant\lambda_j(Q),\quad 1\leqslant j\leqslant n.$$

证明

$$A=A_+-A_-=P-Q,\quad P=A+Q\geqslant A.$$

所以

$$\lambda_j(P)\geqslant\lambda_j(A),\quad\forall j.$$

设 $\lambda_1(A) \geqslant \cdots \geqslant \lambda_k(A) \geqslant 0 > \lambda_{k+1}(A) \geqslant \cdots \geqslant \lambda_n(A)$. 则

$$\lambda_j(A) = \lambda_j(A_+), \quad j = 1, \cdots, k; \quad \lambda_i(A_+) = 0, \quad i = k+1, \cdots, n.$$

于是

$$\lambda_j(P) \geqslant \lambda_j(A_+), \quad \forall j.$$

由 $Q = P - A \geqslant -A$ 得

$$\lambda_j(Q) \geqslant \lambda_j(-A), \quad \forall j.$$

但是

$$\lambda_j(-A) = \lambda_j(A_-), \quad j = 1, \cdots, n-k; \quad \lambda_i(A_-) = 0, \quad i = n-k+1, \cdots, n.$$

所以

$$\lambda_j(Q) \geqslant \lambda_j(A_-), \quad \forall j.$$ □

下面的两个结果再次说明极值表示十分有用.

引理 3.7 (Ky Fan) 设 $A \in H_n, 1 \leqslant k \leqslant n$. 则

$$\sum_{j=1}^{k} \lambda_j(A) = \max_{U^*U=I_k} \mathrm{tr} U^*AU,$$

$$\sum_{j=1}^{k} \lambda_{n-j+1}(A) = \min_{U^*U=I_k} \mathrm{tr} U^*AU,$$

其中 $U \in M_{n,k}$, I_k 是 k 阶单位矩阵.

证明 设 $U \in M_{n,k}$ 满足 $U^*U = I_k$, 则存在矩阵 $V \in M_{n,n-k}$ 使得 $W = (U, V)$ 为酉矩阵. 注意到 U^*AU 是与 A 酉相似的矩阵 W^*AW 的主子矩阵, 用 Cauchy 分隔定理于 W^*AW. □

定理 3.8 (Ky Fan) 设 $A, B \in H_n, 1 \leqslant k \leqslant n$. 则

$$\sum_{j=1}^{k} \lambda_j(A+B) \leqslant \sum_{j=1}^{k} \lambda_j(A) + \sum_{j=1}^{k} \lambda_j(B).$$

$k = n$ 时两边相等.

证明 用引理 3.7. □

§3.2 优超关系

为了记号的简洁, 在本章的多数情形下我们将 $\mathbb{R}^n$ 中的向量看成是行向量, 但是当它们和矩阵相乘时又被看成是列向量, 这种差别可以从上下文容易看出, 不会引起混淆.

我们将 $x=(x_1,x_2,\cdots,x_n)\in\mathbb{R}^n$ 的分量重新排序为 $x_{[1]}\geqslant x_{[2]}\geqslant\cdots\geqslant x_{[n]}$.

定义 设 $x=(x_1,x_2,\cdots,x_n), y=(y_1,y_2,\cdots,y_n)\in\mathbb{R}^n$. 若

$$\sum_{i=1}^{k}x_{[i]}\leqslant\sum_{i=1}^{k}y_{[i]},\quad k=1,2,\cdots,n,$$

则称 x 被 y 弱优超, 记为 $x\prec_w y$. 若 $x\prec_w y$ 并且 $\sum\limits_{i=1}^{n}x_i=\sum\limits_{i=1}^{n}y_i$, 则称 x 被 y 优超, 记为 $x\prec y$.

例如, 若每个 $a_i\geqslant 0, \sum\limits_{i=1}^{n}a_i=1$, 则

$$\left(\frac{1}{n},\cdots,\frac{1}{n}\right)\prec(a_1,\cdots,a_n)\prec(1,0,\cdots,0).$$

下面我们给出优超的一个十分有用的刻画. 每个元素都是非负实数的矩阵称为非负矩阵. 一个非负方阵如果每行元素之和及每列元素之和都等于 1, 则称为双随机矩阵. 设 $u\in\mathbb{R}^n$ 的每个分量都是 1, 则 n 阶矩阵 A 的每行元素之和及每列元素之和都等于 1 这个条件可表示为 $Au=u,\ u^TA=u^T$. 由此我们立刻知道有限多个双随机矩阵的乘积还是双随机矩阵. 下面定理中的向量看作是列向量.

定理 3.9 (Hardy-Littlewood-Pólya) 设 $x,y\in\mathbb{R}^n$. 则 $x\prec y$ 当且仅当存在双随机矩阵 A 使得 $x=Ay$.

证明 假设存在双随机矩阵 A 满足 $x=Ay$, 我们证明 $x\prec y$. 设 $x=(x_1,x_2,\cdots,x_n)^T, y=(y_1,y_2,\cdots,y_n)^T\in\mathbb{R}^n, A=(a_{ij})$. 选取置换矩阵 P,Q 使得 Px 和 Qy 的分量都是从大到小排列的. 我们有 $Px=(PAQ^T)Qy$ 而 PAQ^T 也是双随机矩阵. 因此不妨设 $x_1\geqslant\cdots\geqslant x_n, y_1\geqslant\cdots\geqslant y_n$. 对于任意的 $1\leqslant k\leqslant n$, 有

$$\sum_{i=1}^{k}x_i=\sum_{i=1}^{k}\sum_{j=1}^{n}a_{ij}y_j.$$

记 $t_j=\sum\limits_{i=1}^{k}a_{ij}$, 则 $0\leqslant t_j\leqslant 1,\ \sum\limits_{j=1}^{n}t_j=k$. 我们有

$$\begin{aligned}\sum_{i=1}^{k}x_i-\sum_{i=1}^{k}y_i&=\sum_{j=1}^{n}t_jy_j-\sum_{j=1}^{k}y_j\\&=\sum_{j=1}^{k}(t_j-1)(y_j-y_k)+\sum_{j=k+1}^{n}t_j(y_j-y_k)\\&\leqslant 0.\end{aligned}$$

而当 $k=n$ 时, 上面的不等号变成了等号, 因为那时每个 $t_j=1$, 这就证明了 $x\prec y$.

反过来, 假设 $x\prec y$. 我们证明存在双随机矩阵 A 使得 $x=Ay$. 对 n 用数学归纳法. $n=1$ 时结论平凡地成立. 假设结论对 $\mathbb{R}^{n-1}$ 中的向量成立, 现在考虑 $x=(x_1,x_2,\cdots,x_n)^T$, $y=(y_1,y_2,\cdots,y_n)^T\in\mathbb{R}^n$ 的情形. 不失一般性, 假设 $x_1\geqslant\cdots\geqslant x_n, y_1\geqslant\cdots\geqslant y_n$. 由 $x\prec y$ 知 $y_n\leqslant x_1\leqslant y_1$. 存在 k 使得 $y_k\leqslant x_1\leqslant y_{k-1}$. 因此存在 $0\leqslant t\leqslant 1$ 满足 $x_1=ty_1+(1-t)y_k$. 记

$$x'=(x_2,\cdots,x_n)^T,$$

$$y'=(y_2,\cdots,y_{k-1},(1-t)y_1+ty_k,y_{k+1},\cdots,y_n)^T=(y_2',\cdots,y_n')^T.$$

我们来验证 $x'\prec y'$. 因为 $y_1\geqslant\cdots\geqslant y_{k-1}\geqslant x_1\geqslant x_2\geqslant\cdots\geqslant x_n$, 对于 $2\leqslant m\leqslant k-1$,

$$\sum_{j=2}^m x_j\leqslant\sum_{j=2}^m y_j.$$

对于 $k\leqslant m\leqslant n$,

$$\begin{aligned}\sum_{j=2}^m x_j=\sum_{j=1}^m x_j-x_1&\leqslant\sum_{j=1}^m y_j-x_1\\&=\sum_{j=1}^m y_j-ty_1-(1-t)y_k\\&=\sum_{j=2}^{k-1} y_j+[(1-t)y_1+ty_k]+\sum_{j=k+1}^m y_j\\&=\sum_{j=2}^m y_j',\end{aligned}$$

而当 $m=n$ 时, 上面不等式成为等式, 因为 $x\prec y$. 所以, $x'\prec y'$. 由归纳假设, 存在 $n-1$ 阶双随机矩阵 B 使得 $x'=By'$. 定义 n 阶矩阵 $G=(g_{ij})$ 为 $g_{11}=g_{kk}=t$, $g_{1k}=g_{k1}=1-t, g_{ii}=1, i\neq 1,k, 1\leqslant i\leqslant n$, 所有其他的 $g_{ij}=0$. 则 G 为双随机矩阵. 令 $A=\mathrm{diag}(1,B)G$, 则 A 为双随机矩阵, 且满足 $x=Ay$. □

定理 3.10 (Schur) 设 Hermite 矩阵 A 的对角元素为 $d_1,\cdots,d_n$, 特征值为 $\lambda_1,\cdots,\lambda_n$. 则

$$(d_1,\cdots,d_n)\prec(\lambda_1,\cdots,\lambda_n). \tag{3.1}$$

证明 设 $A=U^*\mathrm{diag}(\lambda_1,\cdots,\lambda_n)U$ 为 A 的谱分解, $U=(u_{ij})$ 为酉矩阵. 则

$$d_i=\sum_{j=1}^n\lambda_j|u_{ji}|^2,\quad 1\leqslant i\leqslant n.$$

令 $S=(|u_{ij}|^2)_{n\times n}$, 则 S 为双随机矩阵, 且

$$\begin{pmatrix}d_1\\\vdots\\d_n\end{pmatrix}=S\begin{pmatrix}\lambda_1\\\vdots\\\lambda_n\end{pmatrix}.$$

由 Hardy-Littlewood-Pólya 定理, $(d_1,\cdots,d_n)\prec(\lambda_1,\cdots,\lambda_n)$. □

下面的定理说明: 定理 (3.10) 的逆命题也成立.

定理 3.11 (Horn [56]) *若 $2n$ 个实数 $d_i,\lambda_i,i=1,\cdots,n$ 满足 (3.1), 则存在一个 n 阶实对称矩阵以 $d_1,\cdots,d_n$ 为对角元素并且以 $\lambda_1,\cdots,\lambda_n$ 为特征值.*

证明 (Chan-Li [30]) 不妨设

$$d_1\geqslant d_2\geqslant\cdots\geqslant d_n,\quad \lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n.$$

令 $D=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$, 只需要证明存在一个正交阵 Q 使得 Q^TDQ 的对角元素为 $d_1,\cdots,d_n$. 我们对 n 用归纳法.

$n=1$ 时, 没有什么需要证明的. 在 $n=2$ 的情形, 若 $\lambda_1=\lambda_2$, 则有 $d_1=d_2$, 结论平凡地成立; 否则 $\lambda_1>\lambda_2$, 令

$$Q=(\lambda_1-\lambda_2)^{-1/2}\begin{pmatrix}\sqrt{\lambda_1-d_2} & -\sqrt{d_2-\lambda_2}\\ \sqrt{d_2-\lambda_2} & \sqrt{\lambda_1-d_2}\end{pmatrix}.$$

则 Q 为正交矩阵且 $Q^T\mathrm{diag}(\lambda_1,\lambda_2)Q$ 的对角元素为 d_1,d_2.

假设结论对于 $\mathbb{R}^{n-1}$ 中的向量成立, 现在考虑 $\mathbb{R}^n$ 中的向量, $n\geqslant 3$. 因为 $\lambda_1\geqslant d_1\geqslant\lambda_n$, 存在 j, $2\leqslant j\leqslant n$, 使得 $\lambda_{j-1}\geqslant d_1\geqslant\lambda_j$. 将已证明的 $n=2$ 时的结论用于 $(d_1,\lambda_1+\lambda_j-d_1)\prec(\lambda_1,\lambda_j)$, 存在 2 阶正交矩阵 Q_1 使得 $Q_1^T\mathrm{diag}(\lambda_1,\lambda_j)Q_1$ 的对角元素为 $d_1,\lambda_1+\lambda_j-d_1$. 令 $Q_2=\mathrm{diag}(Q_1,I_{n-2})$, 则 Q_2 为正交矩阵, 使得

$$Q_2^T\mathrm{diag}(\lambda_1,\lambda_j,\lambda_2,\cdots,\lambda_{j-1},\lambda_{j+1},\cdots,\lambda_n)Q_2=\begin{pmatrix}d_1 & a^T\\ a & D_1\end{pmatrix},$$

其中 $a\in\mathbb{R}^{n-1}$, $D_1=\mathrm{diag}(\lambda_1+\lambda_j-d_1,\lambda_2,\cdots,\lambda_{j-1},\lambda_{j+1},\cdots,\lambda_n)$. 我们断言

$$(d_2,\cdots,d_n)\prec(\lambda_1+\lambda_j-d_1,\lambda_2,\cdots,\lambda_{j-1},\lambda_{j+1},\cdots,\lambda_n).$$

事实上, 因为 $d_1\leqslant\lambda_{j-1}\leqslant\lambda_{j-2}\leqslant\cdots\leqslant\lambda_1$, 当 $2\leqslant k\leqslant j-1$ 时

$$\sum_{i=2}^k d_i\leqslant(k-1)d_1\leqslant\sum_{i=2}^k\lambda_i.$$

当 $j\leqslant k\leqslant n$ 时,

$$\begin{aligned}\sum_{i=2}^k d_i&=\sum_{i=1}^k d_i-d_1\\&\leqslant\sum_{i=1}^k\lambda_i-d_1\\&=(\lambda_1+\lambda_j-d_1)+\lambda_2+\cdots+\lambda_{j-1}+\lambda_{j+1}+\cdots+\lambda_k.\end{aligned}$$

当 $k=n$ 时, 上面的不等式变成等式. 这就证明了上述优超关系. 由归纳假设, 存在 $n-1$ 阶正交矩阵 Q_3 使得 $Q_3^T D_1 Q_3$ 的对角元素为 $d_2,\cdots,d_n$. 令 $Q=Q_2\mathrm{diag}(1,Q_3)$, 则 Q 为正交矩阵, 使得 $Q^T\mathrm{diag}(\lambda_1,\lambda_j,\lambda_2,\cdots,\lambda_{j-1},\lambda_{j+1},\cdots,\lambda_n)Q$ 的对角元素为 d_1, $\cdots,d_n$. □

定理 3.10 和定理 3.11 合起来表明: (3.1) 式中的优超关系是一个一般 Hermite 矩阵的对角元和特征值的全部关系.

下面我们寻求对优超的更多理解. 设 K 是 $\mathbb{R}^d$ 中的一个凸集. 点 $x\in K$ 称为是 K 的一个极点, 如果 $y,z\in K, 0<t<1$, $x=ty+(1-t)z$ 蕴涵 $x=y=z$. K 的所有极点的集合记作 $\mathrm{ex}(K)$. 我们用 $\mathrm{Co}(S)$ 记集合 $S\subseteq\mathbb{R}^d$ 的凸包, 即 S 中元素的凸组合的集合. 极点的重要性可以从下面的定理看出, 它的证明见 ([11], p.52).

定理 3.12 (Krein-Milman) 设 $K\subseteq\mathbb{R}^d$ 是一个紧凸集. 则 K 有极点且 $K=\mathrm{Co}(\mathrm{ex}(K))$.

n 阶双随机矩阵的集合 Ω_n 显然是一个紧凸集. 用 Π_n 记 n 阶置换矩阵的集合. 置换矩阵看起来是最简单的双随机矩阵.

定理 3.13 (Birkhoff) $\mathrm{ex}(\Omega_n)=\Pi_n$. 每个双随机矩阵都是置换矩阵的凸组合.

证明 设 $P\in\Pi_n$, 若 $P=tA+(1-t)B, 0<t<1, A,B\in\Omega_n$, 则因为 A,B 非负, A,B 的每行每列至少有一个正元素, 并且这些正元素处于和 P 中元素 1 相同的位置. 但 $A,B\in\Omega_n$, 所以 A,B 都是置换矩阵, 且 $A=B=P$. 因此 $P\in\mathrm{ex}(\Omega_n)$. 这证明了 $\Pi_n\subseteq\mathrm{ex}(\Omega_n)$.

下面证明每个双随机矩阵都是置换矩阵的凸组合. 我们对正元素的个数作归纳. 若 $G\in\Omega_n$, 则 G 至少有 n 个正元素. 若 G 恰好有 n 个正元素, 则 G 是个置换矩阵.

我们先证明每个双随机矩阵都至少有一条对角线上面的每个元素都是正元素 (称为正对角线). 设 $G\in\Omega_n$, 若 G 无零元素, 结论成立. 假设 G 有一个 r 行 s 列的 0 子阵. 若有必要, 交换某些行某些列, 不妨设

$$G=\begin{pmatrix}0 & C\\ D & E\end{pmatrix}.$$

这里 $0\in M_{r,s}$ 是个 0 矩阵. 因为 C 的每行元素之和及 D 的每列元素之和都是 1, 而 G 的全部元素之和为 n, 所以 $r+s\leqslant n$. 由 Frobenius-König 定理 (定理 2.13), G 有一条正对角线.

假设具有不超过 $k-1$ 个正元素的双随机矩阵都是置换矩阵的凸组合, 并且设 $G\in\Omega_n$ 有 k 个正元素. 取定 G 的一条正对角线, 设 a 是这条对角线上的最小正元素. 若 $G\notin\Pi_n$, 则 $a<1$. 令 Q 是 1 分布在 G 中取定的正对角线上的置换矩阵且令

$$T=(G-aQ)/(1-a).$$

则 $T \in \Omega_n$, 且 T 的正元的个数$\leqslant k-1$, 由归纳假设, T 是双随机矩阵的凸组合, 从而 G 也是双随机矩阵的凸组合, 因为 $G=(1-a)T+aQ$.

设 $W \in \Omega_n \backslash \Pi_n$, 则存在数 t_i 和置换矩阵 P_i 满足 $W=\sum\limits_{i=1}^m t_iP_i$, $0<t_i<1, i=1,\cdots,m$, $\sum\limits_{i=1}^m t_i=1$, $m \geqslant 2$. 令 $F=(\sum\limits_{j=2}^m t_jP_j)/(1-t_1)$, 则 $F \in \Omega_n, W = t_1P_1+(1-t_1)F, P_1 \neq W$. 所以 $W \notin \mathrm{ex}(\Omega_n)$. 这就证明了 $\mathrm{ex}(\Omega_n) \subseteq \Pi_n$. 开始已证 $\Pi_n \subseteq \mathrm{ex}(\Omega_n)$, 因此 $\mathrm{ex}(\Omega_n)=\Pi_n$. □

设 $x, y \in \mathbb{R}^n$. 我们称 y 是 x 的置换, 若 y 的分量是 x 的分量的重排, 即存在置换矩阵 P 使得 $y=Px$. 将 Hardy-Littlewood-Pólya 定理与 Birkhoff 定理合起来就得到下面的结果.

定理 3.14 (Rado [94]) 设 $x, y \in \mathbb{R}^n$. $x \prec y$ 当且仅当 x 是 y 的置换的凸组合.

y 的置换有 $n!$ 个, 对于给定的 x, 那 $n!$ 个 y 的置换都需要吗? 下面的结果回答这个问题.

定理 3.15 ([119]) 设 $x, y \in \mathbb{R}^n$. $x \prec y$ 当且仅当 x 是至多 n 个 y 的置换的凸组合.

定理 3.15 中的个数 n 是最小可能的, 这只需要考虑下面的优超: $e_1+\cdots+e_n \prec ne_1$, 这里 e_i 是 $\mathbb{R}^n$ 的第 i 个标准基向量.

现在给出弱优超的刻画, 为此先做些准备. 一个非负方阵称为双次随机的, 如果它的每行元素之和及每列元素之和都不超过 1. 下面的简单事实使我们可以利用双随机矩阵的性质.

引理 3.16 一个双随机矩阵的方子阵是双次随机矩阵. 反过来, 设 A 是 n 阶双次随机矩阵, 令 $R=\mathrm{diag}(r_1,\cdots,r_n), C=\mathrm{diag}(c_1,\cdots,c_n)$. 其中 r_i 是 A 的第 i 行元素之和, c_i 是 A 的第 i 列元素之和, 则

$$\begin{pmatrix} A & I-R \\ I-C & A^T \end{pmatrix}$$

是个双随机矩阵.

一个方阵称为是部分置换矩阵如果它的每行至多只有一个非零元素, 每列至多只有一个非零元素, 而那些非零元素 (如果有的话) 全部是1. 用引理 3.16 和定理 3.13 立刻得到下面的结果.

定理 3.17 用 Γ_n 记 n 阶双次随机矩阵的集合 (显然是个凸集), 用 PP_n 记 n 阶部分置换矩阵的集合. 则 $\mathrm{ex}(\Gamma_n)=PP_n$, 且每个双次随机矩阵都是部分置换矩阵的凸组合.

我们总是用 $M_{m,n}(S)$ 记元素属于集合 S 的 m 行 n 列矩阵的集合, $M_n(S)=M_{n,n}(S)$. 对于 $A=(a_{ij}),B=(b_{ij})\in M_{m,n}(\mathbb{R})$. 我们用记号 $A\leqslant_e B$ 表示 $a_{ij}\leqslant b_{ij}$ 对所有 $1\leqslant i\leqslant m,1\leqslant j\leqslant n$ 成立, 下标 e 暗示以元素的方式. 对于向量, 因为没有歧义, 我们将省略下标 e, 即对 $x,y\in\mathbb{R}^n,x\leqslant y$ 表示 x 的每个分量都不超过 y 的每个对应分量. 注意到我们可以将一个部分置换矩阵中的某些 0 改成 1 而得到一个置换矩阵. 下面的推论可从定理 3.17 轻易得到.

推论 3.18 一个非负方阵 A 是双次随机的当且仅当存在一个双随机矩阵 B 满足 $A\leqslant_e B$.

记 $\mathbb{R}_+^n=\{x\in\mathbb{R}^n\mid x=(x_1,\cdots,x_n),x_i\geqslant 0,1\leqslant i\leqslant n\}$.

定理 3.19 (i) 设 $x,y\in\mathbb{R}_+^n$. 则 $x\prec_w y$ 当且仅当存在双次随机矩阵 A 使得 $x=Ay$.

(ii) 设 $x,y\in\mathbb{R}^n$. 则 $x\prec_w y$ 当且仅当存在 $u\in\mathbb{R}^n$ 使得 $x\leqslant u$ 且 $u\prec y$.

证明 (i) 设 $x,y\in\mathbb{R}_+^n$, 且 $x=Ay,A$ 双次随机. 由推论 3.18 存在双随机矩阵 B 满足 $A\leqslant_e B$. 于是 $x=Ay\leqslant By\prec y$. 显然有 $x\prec_w y$.

反过来设 $x,y\in\mathbb{R}_+^n$, 且 $x\prec_w y$, 我们将证明存在双次随机矩阵 A 满足 $x=Ay$. 若 $x=0$, 取 $A=0$ 即可; 若 $x\prec y$, 定理 3.9 保证存在一个满足条件的双随机矩阵 A. 现在假设不是这两种情况. 设 r 是 x 的最小正分量, 记$s=\sum\limits_{i=1}^n y_i-\sum\limits_{i=1}^n x_i$. 由假设, $s>0$. 选取一个正整数 m 满足 $r\geqslant s/m$. 设 $e\in\mathbb{R}^m$ 的分量全为 1. 构造

$$x'=\begin{pmatrix}x\\ \frac{s}{m}e\end{pmatrix},\quad y'=\begin{pmatrix}y\\ 0\end{pmatrix}\in\mathbb{R}^{n+m}.$$

则 $x'\prec y'$. 于是存在 $n+m$ 阶的双随机矩阵 G 满足 $x'=Gy'$. 令 A 是 G 左上角的 n 阶子阵, 则 A 双次随机, 且 $x=Ay$.

(ii) 若 $x\leqslant u$ 且 $u\prec y$ 显然有 $x\prec_w y$. 反过来, 设 $x\prec_w y$. 选取一个正数 t 使得 $x+te$ 和 $y+te$ 的分量都非负, 这里 $e\in\mathbb{R}^n$ 的分量都为 1. 我们仍然有 $x+te\prec_w y+te$. 由 (i), 存在双次随机矩阵 A 使得 $x+te=A(y+te)$. 据推论 3.18, 存在双随机矩阵 B 满足 $A\leqslant_e B$. 那么, $x+te\leqslant B(y+te)=By+te$, 因此 $x\leqslant By$. 取 $u=By$. 则 $x\leqslant u,u\prec y$. □

下面的两个定理非常有用, 这里假定实值函数 $f(t)$, $g(t)$ 定义在某个包含 $x=(x_1,\cdots,x_n)$ 和 $y=(y_1,\cdots,y_n)$ 的分量的区间上.

定理 3.20 设 $f(t)$ 是个凸函数, 则

$$x\prec y \text{ 蕴涵 } (f(x_1),\cdots,f(x_n))\prec_w (f(y_1),\cdots,f(y_n)).$$

证明 设 $x \prec y$. 则存在双随机矩阵 $A=(a_{ij})$ 满足 $x=Ay$. 于是

$$x_i=\sum_{j=1}^{n} a_{ij}y_j, \quad 1 \leqslant i \leqslant n.$$

因此

$$f(x_i) \leqslant \sum_{j=1}^{n} a_{ij}f(y_j), \quad 1 \leqslant i \leqslant n.$$

这组不等式可以写成

$$\begin{pmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{pmatrix} \leqslant A \begin{pmatrix} f(y_1) \\ \vdots \\ f(y_n) \end{pmatrix}.$$

据定理 3.19(ii) 得$(f(x_1),\cdots,f(x_n)) \prec_w (f(y_1),\cdots,f(y_n))$. □

定理 3.21 设 $g(t)$ 是递增的凸函数, 则

$$x \prec_w y \text{ 蕴涵 } (g(x_1),\cdots,g(x_n)) \prec_w (g(y_1),\cdots,g(y_n)).$$

证明 设 $x \prec_w y$. 据定理 3.19(ii) 存在向量 u 满足 $x \leqslant u, u \prec y$. 设 $u=(u_1,\cdots,u_n)$. 因 g 递增

$$(g(x_1),\cdots,g(x_n)) \leqslant (g(u_1),\cdots,g(u_n)). \tag{3.2}$$

又因为 g 是凸函数, 应用定理 3.20 于 $u \prec y$ 得

$$(g(u_1),\cdots,g(u_n)) \prec_w (g(y_1),\cdots,g(y_n)). \tag{3.3}$$

(3.2) 和 (3.3) 合起来给出

$$(g(x_1),\cdots,g(x_n)) \prec_w (g(y_1),\cdots,g(y_n)).$$ □

下面结果的证明方法是应用优超原理的一个例子.

定理 3.22 设半正定矩阵 A 的对角元素为 $d_1 \geqslant \cdots \geqslant d_n$, 特征值为 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$, 则

$$\prod_{i=k}^{n} d_i \geqslant \prod_{i=k}^{n} \lambda_i, \quad k=1,2,\cdots,n. \tag{3.4}$$

证明 因为每个半正定矩阵都可以用一列正定矩阵来逼近, 例如 $\lim\limits_{m\to\infty}(\frac{1}{m}I+A)=A$, 而对角元素和特征值都是关于矩阵连续的, 我们只需要对 A 正定的情形证明结论. 此时每个 d_i 和每个 λ_i 都是正数, $i=1,\cdots,n$.

由定理 3.10,

$$(d_1,\cdots,d_n) \prec (\lambda_1,\cdots,\lambda_n).$$

$f(t)=-\log t$ 是 $(0,\infty)$ 上的凸函数. 应用定理 3.20 于 $f(t)$ 和上述优超关系就得到 (3.4). □

(3.4) 的特殊情形 $k=1$ 可写为

$$\prod_{i=1}^{n} d_i \geqslant \det A. \tag{3.5}$$

(3.5) 称为 Hadamard 不等式.

对于 $x=(x_1,x_2,\cdots,x_n)\in\mathbb{R}_+^n$ 和 $y=(y_1,y_2,\cdots,y_n)\in\mathbb{R}_+^n$, 若

$$\prod_{i=1}^{k} x_{[i]} \leqslant \prod_{i=1}^{k} y_{[i]}, \quad k=1,2,\cdots,n,$$

则称 x 被 y 弱对数优超, 记作 $x\prec_{\mathrm{wlog}} y$. 若 $x\prec_{\mathrm{wlog}} y$ 且

$$\prod_{i=1}^{n} x_i = \prod_{i=1}^{n} y_i,$$

则称 x 被 y 对数优超, 记作 $x\prec_{\log} y$.

下面的结果说明: 弱对数优超强于弱优超.

定理 3.23 设 $x,y\in\mathbb{R}_+^n$, 则

$$x\prec_{\mathrm{wlog}} y \text{ 蕴涵 } x\prec_w y.$$

证明 由连续性只需证明 $x=(x_1,x_2,\cdots,x_n)$ 和 $y=(y_1,y_2,\cdots,y_n)$ 的分量全都是正数的情形. 从 $x\prec_{w\log} y$ 得

$$(\log x_1,\cdots,\log x_n)\prec_w(\log y_1,\cdots,\log y_n).$$

应用定理 3.21 于递增的凸函数 $g(t)=e^t$ 和上述的弱优超关系得 $x\prec_w y$. □

论述优超的两个基本的参考文献是 [85, 3].

§3.3 关于半正定矩阵的不等式

关于半正定矩阵的不等式很多, 可以参看 [15, 58, 118, 18]. 本节强调证明方法.

设 $f(t)$ 是定义在实区间 Ω 上的实值连续函数, Hermite 矩阵 H 的特征值包含于 Ω. 设 $H=U\mathrm{diag}(\lambda_1,\cdots,\lambda_n)U^*$ 为 H 的谱分解, 其中 U 为酉矩阵. 则 H 的函数运算定义为

$$f(H)=U\mathrm{diag}(f(\lambda_1),\cdots,f(\lambda_n))U^*. \tag{3.6}$$

这个定义是合适的, 即 $f(H)$ 不依赖于 H 的特定的谱分解. 这可以从以下两点看出.

首先, (3.6) 与通常的多项式运算一致: 若 $f(t)=\sum_{j=0}^{k} c_j t^j$, 则 $f(H)=\sum_{j=0}^{k} c_j H^j$; 其次我们可以将 Ω 取成一个有限闭区间, 而 Weierstrass 逼近定理说有限闭区间上的每个连续函数都可以被一列多项式一致逼近.

用 I 表示单位阵. 矩阵 C 称为是*压缩矩阵*, 若 $\|C\|_\infty \leqslant 1$, 或等价地 $C^*C \leqslant I$. 设 $\rho(A)$ 是 A 的谱半径, 则 $\rho(A) \leqslant \|A\|_\infty$. 对于同阶复方阵 A, B, AB 和 BA 的特征值相同, 从而 $\rho(AB)=\rho(BA)$.

定理 3.24 (Löwner-Heinz) 若 $A \geqslant B \geqslant 0, 0 \leqslant r \leqslant 1$, 则

$$A^r \geqslant B^r. \tag{3.7}$$

证明 由连续性, 只需对 A 正定来证明结论. 现设 A 正定. 设

$$\triangle = \{r \in [0,1] \mid r \text{ 使 (3.7) 成立.}\}$$

显然 $0,1 \in \triangle$, 且 $\triangle$ 是闭集. 下面证明 $\triangle$ 是凸集, 从而 $\triangle=[0,1]$, 这样就证明了定理. 设 $s,t \in \triangle$. 则

$$A^{-s/2}B^sA^{-s/2} \leqslant I, \quad A^{-t/2}B^tA^{-t/2} \leqslant I,$$

等价地,

$$\|B^{s/2}A^{-s/2}\|_\infty \leqslant 1, \quad \|B^{t/2}A^{-t/2}\|_\infty \leqslant 1.$$

于是

$$\begin{aligned}
\|A^{-(s+t)/4}B^{(s+t)/2}A^{-(s+t)/4}\|_\infty &= \rho(A^{-(s+t)/4}B^{(s+t)/2}A^{-(s+t)/4}) \\
&= \rho(A^{-s/2}B^{(s+t)/2}A^{-t/2}) \\
&\leqslant \|A^{-s/2}B^{(s+t)/2}A^{-t/2}\|_\infty \\
&= \|(B^{s/2}A^{-s/2})^*(B^{t/2}A^{-t/2})\|_\infty \\
&\leqslant \|B^{s/2}A^{-s/2}\|_\infty \|B^{t/2}A^{-t/2}\|_\infty \\
&\leqslant 1.
\end{aligned}$$

因此, $A^{-(s+t)/4}B^{(s+t)/2}A^{-(s+t)/4} \leqslant I$, 从而 $B^{(s+t)/2} \leqslant A^{(s+t)/2}$, 即 $(s+t)/2 \in \triangle$. 这证明了 $\triangle$ 的凸性. □

定理 3.24 对于任何 $r>1$ 都是不对的 ([118], Thm 1.2). 例如

$$A=\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad B=\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad A^2-B^2=\begin{pmatrix} 4 & 3 \\ 3 & 2 \end{pmatrix}$$

表明 $A \geqslant B \geqslant 0 \nRightarrow A^2 \geqslant B^2$.

设 $f(t)$ 是定义在实区间 Ω 上的实值连续函数. $f(t)$ 称为是算子单调的, 若

$$A \leqslant B \text{ 蕴涵 } f(A) \leqslant f(B)$$

对于任意阶数的特征值属于 Ω 的任何 Hermite 阵 A, B 成立. 定理 3.24 表明: 当 $0 \leqslant r \leqslant 1$ 时, $f(t) = t^r$ 是 $[0, \infty)$ 上的算子单调函数.

我们有时用 $\{x_i\}_{i=1}^n$ 记 $(x_1, x_2, \cdots, x_n)$. 如果矩阵 A 的特征值全部为实数, 我们总是用 $\lambda_1(A) \geqslant \lambda_2(A) \geqslant \cdots$ 记 A 的特征值.

定理 3.25 (Wang-Gong [109]) 设 A, B 为 n 阶半正定矩阵, $0 < s < t$. 则

$$\{\lambda_i^{1/s}(A^s B^s)\}_{i=1}^n \prec_{\log} \{\lambda_i^{1/t}(A^t B^t)\}_{i=1}^n. \tag{3.8}$$

证明 我们先证明一个结论: 若 E 和 F 是同阶半正定矩阵且 $\lambda_1(EF) \leqslant 1$, 则对于 $0 \leqslant r \leqslant 1$, 有 $\lambda_1(E^r F^r) \leqslant 1$. 由连续性, 不妨设 $E > 0$. 则

$$\begin{aligned}\lambda_1(EF) \leqslant 1 &\Rightarrow \lambda_1(E^{\frac{1}{2}} F E^{\frac{1}{2}}) \leqslant 1 \Rightarrow E^{\frac{1}{2}} F E^{\frac{1}{2}} \leqslant I \\ &\Rightarrow F \leqslant E^{-1} \Rightarrow F^r \leqslant E^{-r} \Rightarrow E^{\frac{r}{2}} F^r E^{\frac{r}{2}} \leqslant I \\ &\Rightarrow \lambda_1(E^{\frac{r}{2}} F^r E^{\frac{r}{2}}) \leqslant 1 \Rightarrow \lambda_1(E^r F^r) \leqslant 1.\end{aligned}$$

这里我们用到了定理 3.24.

上面的结论蕴涵如下的结论: 若 G, H 为同阶半正定矩阵且 $0 \leqslant r \leqslant 1$, 则

$$\lambda_1(G^r H^r) \leqslant \lambda_1^r(GH). \tag{3.9}$$

由连续性不妨设 $G > 0, H > 0$. 设 $\lambda_1(GH) = u^2$, 则 $u \neq 0$, 且 $\lambda_1(\frac{G}{u}\frac{H}{u}) = 1$. 于是 $\lambda_1((\frac{G}{u})^r(\frac{H}{u})^r) \leqslant 1$, 即 $\lambda_1(G^r H^r) \leqslant u^{2r} = \lambda_1^r(GH)$. 这就证明了 (3.9).

作变量替换, (3.9) 蕴涵

$$\lambda_1^{1/s}(G^s H^s) \leqslant \lambda_1^{1/t}(G^t H^t) \tag{3.10}$$

对任何 $0 < s < t$ 成立.

现在考虑复合矩阵. 对 $1 \leqslant k \leqslant n$, 在 (3.10) 中令 $G = C_k(A), H = C_k(B)$ 就得到

$$\left[\prod_{i=1}^k \lambda_i(A^s B^s)\right]^{1/s} \leqslant \left[\prod_{i=1}^k \lambda_i(A^t B^t)\right]^{1/t} \tag{3.11}$$

而且当 $k = n$ 时 (3.11) 成为等式: 两边都等于 $\det(AB)$. 这就证明了 (3.8). □

推论 3.26 (Lieb-Thirring 不等式) 若 A, B 为同阶半正定矩阵, m 为正整数, 则

$$\operatorname{tr}(AB)^m \leqslant \operatorname{tr} A^m B^m. \tag{3.12}$$

证明 由定理 3.25 得

$$\{\lambda_j^m(AB)\} \prec_{\log} \{\lambda_j(A^m B^m)\}. \tag{3.13}$$

因为弱对数优超蕴涵弱优超 (定理 3.23), (3.13) 给出

$$\{\lambda_j^m(AB)\} \prec_w \{\lambda_j(A^m B^m)\}.$$

而 $\lambda_j^m(AB) = \lambda_j[(AB)^m)]$, 所以

$$\{\lambda_j[(AB)^m]\} \prec_w \{\lambda_j(A^m B^m)\}.$$

特别有 (3.12) 成立. □

不等式 (3.12) 看起来非常初等, 但不用优超来证明它似乎是困难的.

分块矩阵是矩阵论里的重要技巧, 它的效果类似于张量积和复合矩阵. 下面给出一个例子 (定理 3.29).

引理 3.27 *若 $A > 0$, 则*

$$\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \geqslant 0$$

当且仅当 Schur 补 $C - B^*A^{-1}B \geqslant 0$. 若 $C > 0$, 则

$$\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \geqslant 0$$

当且仅当 Schur 补 $A - BC^{-1}B^* \geqslant 0$.

证明 只需考虑合同变换

$$\begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix}^* \begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & C - B^*A^{-1}B \end{pmatrix},$$

$$\begin{pmatrix} I & 0 \\ -C^{-1}B^* & I \end{pmatrix}^* \begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \begin{pmatrix} I & 0 \\ -C^{-1}B^* & I \end{pmatrix} = \begin{pmatrix} A - BC^{-1}B^* & 0 \\ 0 & C \end{pmatrix}.$$ □

另一个判断分块矩阵半正定的标准是下面的引理.

引理 3.28

$$\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \geqslant 0 \tag{3.14}$$

当且仅当 $A \geqslant 0, C \geqslant 0$, 并且存在压缩矩阵 W 使得 $B = A^{1/2}WC^{1/2}$.

证明 易知

$$\begin{pmatrix} I & X \\ X^* & I \end{pmatrix} \geqslant 0$$

当且仅当 X 是压缩矩阵.

若 $A \geqslant 0, C \geqslant 0$, 且 $B = A^{1/2}WC^{1/2}$, W 是压缩矩阵, 则

$$\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} = \begin{pmatrix} A^{1/2} & 0 \\ 0 & C^{1/2} \end{pmatrix}^* \begin{pmatrix} I & W \\ W^* & I \end{pmatrix} \begin{pmatrix} A^{1/2} & 0 \\ 0 & C^{1/2} \end{pmatrix} \geqslant 0.$$

反过来设 (3.14) 成立. 先考虑 $A > 0, C > 0$ 的情形. 此时

$$\begin{pmatrix} I & A^{-1/2}BC^{-1/2} \\ (A^{-1/2}BC^{-1/2})^* & I \end{pmatrix} = \begin{pmatrix} A^{-1/2} & 0 \\ 0 & C^{-1/2} \end{pmatrix}^* \begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \begin{pmatrix} A^{-1/2} & 0 \\ 0 & C^{-1/2} \end{pmatrix} \geqslant 0.$$

于是, $W = A^{-1/2}BC^{-1/2}$ 是压缩矩阵, 且 $B = A^{1/2}WC^{1/2}$.

一般情形, 有

$$\begin{pmatrix} A + m^{-1}I & B \\ B^* & C + m^{-1}I \end{pmatrix} \geqslant 0$$

对任何正整数 m 成立. 由上面已证的结论, 对每个 m 存在压缩矩阵 W_m 使得

$$B = (A + m^{-1}I)^{1/2}W_m(C + m^{-1}I)^{1/2}. \tag{3.15}$$

设 $W_m \in M_{n,r}$. 因为 $M_{n,r}$ 是有限维空间, 谱范数的单位球 $\{x \in M_{n,r} \mid \|x\|_\infty \leqslant 1\}$ 是紧集. 所以 $\{W_m\}_{m=1}^\infty$ 有一个收敛的子序列 $\{W_{m_k}\}_{k=1}^\infty$, $\lim\limits_{k\to\infty} W_{m_k} = W$. 显然 W 是压缩矩阵. 在 (3.15) 中令 $k \to \infty$ 得 $B = A^{1/2}WC^{1/2}$. □

我们用 $\circ_{j=1}^k A_j$ 表示 $A_1 \circ A_2 \circ \cdots \circ A_k$.

定理 3.29 设 $A_i \in M_n$ 正定, $X_i \in M_{n,m}$ 任意, $i = 1, \cdots, k$. 则

$$\circ_{i=1}^k (X_i^* A_i^{-1} X_i) \geqslant (\circ_{i=1}^k X_i)^* (\circ_{i=1}^k A_i)^{-1} (\circ_{i=1}^k X_i), \tag{3.16}$$

$$\sum_{i=1}^k X_i^* A_i^{-1} X_i \geqslant \left(\sum_{i=1}^k X_i\right)^* \left(\sum_{i=1}^k A_i\right)^{-1} \left(\sum_{i=1}^k X_i\right). \tag{3.17}$$

证明 由引理 3.27,

$$\begin{pmatrix} A_i & X_i \\ (X_i)^* & X_i^* A_i^{-1} X_i \end{pmatrix} \geqslant 0, \quad i = 1, \cdots, k.$$

应用 Schur 乘积定理得

$$\begin{pmatrix} \circ_{i=1}^k A_i & \circ_{i=1}^k X_i \\ (\circ_{i=1}^k X_i)^* & \circ_{i=1}^k (X_i^* A_i^{-1} X_i) \end{pmatrix} \geqslant 0. \tag{3.18}$$

再从另一个方向用引理 3.27 于 (3.18) 就得到 (3.16).

不等式 (3.17) 可以类似地证明. □

不等式 (3.17) 证明于 [52]. 不等式 (3.16) 独立地证明于 [110] 和 [114].

在定理 3.29 中分别取 $X_i = I, i = 1, \cdots, k$ 和 $A_i = I, i = 1, \cdots, k$ 我们得到下面的推论.

推论 3.30 设 $A_i \in M_n$ 正定, $X_i \in M_{n,m}$ 任意, $i = 1, \cdots, k$, 则

$$\circ_{i=1}^{k} A_i^{-1} \geqslant (\circ_{i=1}^{k} A_i)^{-1}, \tag{3.19}$$

$$\circ_{i=1}^{k} (X_i^* X_i) \geqslant (\circ_{i=1}^{k} X_i)^* (\circ_{i=1}^{k} X_i),$$

$$\sum_{i=1}^{k} A_i^{-1} \geqslant k^2 \left(\sum_{i=1}^{k} A_i \right)^{-1},$$

$$k \sum_{i=1}^{k} X_i^* X_i \geqslant \left(\sum_{i=1}^{k} X_i \right)^* \left(\sum_{i=1}^{k} X_i \right).$$

在 (3.19) 中记 $A_1 = A$, 取 $k = 2, A_2 = A^{-1}$, 得 $A \circ A^{-1} \geqslant (A \circ A^{-1})^{-1}$. 等价地, 对于正定的 A, 有

$$A \circ A^{-1} \geqslant I. \tag{3.20}$$

(3.20) 是 M. Fiedler 的著名的不等式.

习 题

1. 设 $A \in M_n$. 证明若 $AA^* = A^2$ 则 $A^* = A$.

2. 设 $A \in M_n, B \in M_{r,t}$ 是 A 的一个子矩阵. 则它们的奇异值满足

$$s_j(B) \leqslant s_j(A), \quad j = 1, \cdots, \min\{r, t\}.$$

3. (Aronszajn) 设

$$C = \begin{pmatrix} A & X \\ X^* & B \end{pmatrix}$$

为 Hermite 矩阵, $C \in M_n, A \in M_k$. 设 A, B, C 的特征值分别为 $\alpha_1 \geqslant \cdots \geqslant \alpha_k$, $\beta_1 \geqslant \cdots \geqslant \beta_{n-k}, \gamma_1 \geqslant \cdots \geqslant \gamma_n$. 则对于满足 $i + j - 1 \leqslant n$ 的任意的 i, j,

$$\gamma_{i+j-1} + \gamma_n \leqslant \alpha_i + \beta_j.$$

4. 设 $x, y, u \in \mathbb{R}^n$ 的分量都是递减的. 证明

(i) 若 $x \prec y$ 则 $\langle x, u \rangle \leqslant \langle y, u \rangle$.

(ii) 若 $x \prec_w y$ 且 $u \in \mathbb{R}_+^n$, 则 $\langle x, u \rangle \leqslant \langle y, u \rangle$.

5. 不用 Weierstrass 定理, 直接证明 Hermite 矩阵的函数运算 (3.6) 与特定的谱分解无关.

6. 设 $A, B \in M_n$, A 是正定矩阵, B 是 Hermite 矩阵. 则

$$A + B \text{ 正定当且仅当 } \lambda_j(A^{-1}B) > -1, \quad j = 1, \cdots, n.$$

7. 设 $A \in M_n$ 正定, $1 \leqslant k \leqslant n$. 则

$$\prod_{j=1}^{k} \lambda_j(A) = \max_{U^*U=I_k} \det U^*AU,$$

$$\prod_{j=1}^{k} \lambda_{n-j+1}(A) = \min_{U^*U=I_k} \det U^*AU.$$

其中 $U \in M_{n,k}$.

8. 证明每个半正定矩阵都有唯一的半正定平方根, 即若 $A \geqslant 0$, 则存在唯一的 $B \geqslant 0$ 满足 $B^2 = A$.

9. 用公式

$$t^r = \frac{\sin r\pi}{\pi} \int_0^\infty \frac{s^{r-1}t}{s+t} ds \quad (0 < r < 1)$$

证明定理 3.24.

10. 设 A, B 是同阶半正定矩阵, $0 \leqslant s \leqslant 1$. 证明 $\|A^sB^s\|_\infty \leqslant \|AB\|_\infty^s$.

11. (Ky Fan) 对于 $A \in M_n$, 记 $\mathrm{Re}A = (A + A^*)/2$. 证明

$$\mathrm{Re}\lambda(A) \prec \lambda(\mathrm{Re}A),$$

其中 $\lambda(A)$ 表示 A 的特征值作成的向量, $\mathrm{Re}\lambda(A)$ 表示取 A 的特征值的实部所得向量.

12. (Webster [111]) 设 $A = (a_{ij})$ 是有 k 个正元素的 n 阶双随机矩阵. 证明, 存在 $1, 2, \cdots, n$ 的一个排列 σ 使得

$$\sum_{i=1}^{n} \frac{1}{a_{i\sigma(i)}} \leqslant k.$$

13. (Cayley 变换) 记 $i = \sqrt{-1}$. 若 A 为 Hermite 矩阵, 则

$$\phi(A) = (A - iI)(A + iI)^{-1}$$

是个酉矩阵.

14. 用 Hadamard 不等式 (3.5) 证明下面的不等式 (也称为 Hadamard 不等式): 设 $A = (a_1, \cdots, a_n) \in M_n$, 则

$$|\det A| \leqslant \prod_{i=1}^{n} \|a_i\|,$$

其中 $\|\cdot\|$ 表示列向量的欧氏范数.

15. ([121]) 设 $S_n[a, b]$ 表示所有元素属于给定的区间 $[a, b]$ 的 n 阶实对称矩阵的集合. 对于 $j = 1, n$ 确定 $\max\{\lambda_j(A) \mid A \in S_n[a, b]\}$ 和 $\min\{\lambda_j(A) \mid A \in S_n[a, b]\}$, 以及分别取到最大值和最小值的矩阵.

第四章 奇异值和酉不变范数

奇异值和酉不变范数是矩阵分析中两个紧密相连的重要的论题. 为简明扼要我们只考虑方阵. 从方阵到长方矩阵的推广是显然的, 而且关于长方矩阵的奇异值和酉不变范数的问题通常都可以通过对长方矩阵添加一些零行或零列转化成方阵的情形.

§4.1 奇异值

$A \in M_n$ 的奇异值定义为 A^*A 的特征值的非负平方根. 记 $|A| = (A^*A)^{1/2}$, 称为 A 的绝对值. A 的奇异值就是 $|A|$ 的特征值. 我们总是用 $s_1(A) \geqslant s_2(A) \geqslant \cdots \geqslant s_n(A)$ 表示 $A \in M_n$ 的奇异值并记 $s(A) = (s_1(A), \cdots, s_n(A))$. 显然奇异值是酉不变的: 对于任意 $A \in M_n$ 及任意酉矩阵 $U, V \in M_n, s(UAV) = s(A)$. 因此, 正规矩阵的奇异值就是它的特征值的模. 特别, 对于半正定矩阵, 奇异值和特征值是一样的.

用 $\|x\|$ 表示 $x \in \mathbb{C}^n$ 的欧氏范数. 从关于 Hermite 矩阵的特征值的 Courant-Fischer 定理我们立刻得到

引理 4.1 设 $A \in M_n$. 则对于 $k = 1, \cdots, n$,

$$s_k(A) = \max_{\substack{\Omega \subseteq \mathbb{C}^n \\ \dim \Omega = k}} \min_{\substack{x \in \Omega \\ \|x\|=1}} \|Ax\| = \min_{\substack{\Omega \subseteq \mathbb{C}^n \\ \dim \Omega = n-k+1}} \max_{\substack{x \in \Omega \\ \|x\|=1}} \|Ax\|.$$

从这个引理我们看到, 谱范数 $\|A\|_\infty$ 等于最大奇异值 $s_1(A)$.

引理 4.2 设 $A \in M_n$. 则对于 $k = 1, \cdots, n$,

$$s_k(A) = \min\{\|A - G\|_\infty \mid \text{rank} G \leqslant k-1, G \in M_n\}.$$

证明 由引理 4.1,

$$s_k(A)=\min_{\substack{\Omega\subseteq\mathbb{C}^n\\ \dim\Omega=n-k+1}}\max_{\substack{x\in\Omega\\ \|x\|=1}}\|Ax\|.$$

设 $\mathrm{rank}G\leqslant k-1$, 则 $\dim\ker(G)\geqslant n-k+1$. 任取一个子空间 $\Omega_0\subseteq\ker(G)$ 满足 $\dim\Omega_0=n-k+1$. 于是

$$s_k(A)\leqslant\max_{\substack{x\in\Omega_0\\ \|x\|=1}}\|Ax\|=\max_{\substack{x\in\Omega_0\\ \|x\|=1}}\|(A-G)x\|\leqslant\|A-G\|_\infty.$$

另一方面, 设 $A=U\mathrm{diag}(s_1(A),\cdots,s_n(A))V$ 为 A 的奇异值分解. 令

$$G_0=U\mathrm{diag}(s_1,\cdots,s_{k-1},0,\cdots,0)V.$$

则 $\mathrm{rank}G_0\leqslant k-1$, 且 $s_k(A)=\|A-G_0\|_\infty$. 这就证明了结论. □

定理 4.3 (Ky Fan) 设 $A,B\in M_n, 1\leqslant i,j\leqslant n, i+j-1\leqslant n$, 则

$$s_{i+j-1}(A+B)\leqslant s_i(A)+s_j(B), \tag{4.1}$$

$$s_{i+j-1}(AB)\leqslant s_i(A)s_j(B). \tag{4.2}$$

证明 应用引理 4.2, 存在 $G,H\in M_n$ 满足

$$\begin{aligned}&\mathrm{rank}G\leqslant i-1, && \mathrm{rank}H\leqslant j-1,\\ &s_i(A)=\|A-G\|_\infty, && s_j(B)=\|B-H\|_\infty.\end{aligned}$$

则 $\mathrm{rank}(G+H)\leqslant(i+j-1)-1$. 所以

$$\begin{aligned}s_{i+j-1}(A+B)&\leqslant\|A+B-(G+H)\|_\infty\\ &\leqslant\|A-G\|_\infty+\|B-H\|_\infty\\ &=s_i(A)+s_j(B).\end{aligned}$$

又

$$\begin{aligned}\mathrm{rank}[AH+G(B-H)]&\leqslant\mathrm{rank}(AH)+\mathrm{rank}[G(B-H)]\\ &\leqslant(i+j-1)-1,\end{aligned}$$

因此

$$\begin{aligned}s_{i+j-1}(AB)&\leqslant\|AB-[AH+G(B-H)]\|_\infty\\ &=\|(A-G)(B-H)\|_\infty\\ &\leqslant\|A-G\|_\infty\|B-H\|_\infty\\ &=s_i(A)s_j(B).\end{aligned}$$

□

(4.1) 和 (4.2) 分别是 $\|A+B\|_\infty \leqslant \|A\|_\infty + \|B\|_\infty$ 和 $\|AB\|_\infty \leqslant \|A\|_\infty \|B\|_\infty$ 的推广.

推论 4.4 设 $A, B \in M_n$, 且 $\mathrm{rank}B \leqslant k$. 则

$$s_i(A) \geqslant s_{i+k}(A+B), \quad i = 1, \cdots, n-k.$$

证明 条件 $\mathrm{rank}B \leqslant k$ 蕴涵 $s_{k+1}(B) = 0$. 由(4.1),

$$s_{i+k}(A+B) \leqslant s_i(A) + s_{k+1}(B) = s_i(A).$$ □

定理 4.5 设 B 是 $A \in M_n$ 的一个 $r \times t$ 的子矩阵, 则

$$s_i(A) \geqslant s_i(B) \geqslant s_{i+2n-r-t}(A),$$

其中第一个不等式中 i 的范围是 $1 \leqslant i \leqslant \min\{r,t\}$, 第二个不等式中 i 的范围是 $1 \leqslant i \leqslant r+t-n$.

证明 因为交换行和列都不改变奇异值, 不妨设 B 位于 A 的左上角:

$$A = \begin{pmatrix} B & C \\ D & E \end{pmatrix}, \quad 从而\ A^*A = \begin{pmatrix} B^*B + D^*D & * \\ * & * \end{pmatrix}.$$

用 Cauchy 分隔定理和 Weyl 单调性定理,

$$s_i^2(A) = \lambda_i(A^*A) \geqslant \lambda_i(B^*B + D^*D) \geqslant \lambda_i(B^*B) = s_i^2(B).$$

注意到 $\mathrm{rank}(D^*D) \leqslant n-r$, 用推论 4.4 和 Cauchy 分隔定理, 有

$$\begin{aligned} s_i^2(B) = \lambda_i(B^*B) &\geqslant \lambda_{i+n-r}(B^*B + D^*D) \\ &\geqslant \lambda_{i+n-r+n-t}(A^*A) \\ &= s_{i+2n-r-t}^2(A). \end{aligned}$$ □

定理 4.6 (Horn [55]) 设 $A, B \in M_n$, 则

$$s(AB) \prec_{\log} \{s_i(A)s_i(B)\}_{i=1}^n.$$

证明 用复合矩阵. 对 $1 \leqslant k \leqslant n$,

$$\begin{aligned} \prod_{i=1}^k s_i(AB) &= s_1[C_k(AB)] = \|C_k(AB)\|_\infty \\ &= \|C_k(A)C_k(B)\|_\infty \\ &\leqslant \|C_k(A)\|_\infty \|C_k(B)\|_\infty \\ &= s_1[C_k(A)]s_1[C_k(B)] \\ &= \prod_{i=1}^k s_i(A) \prod_{i=1}^k s_i(B) \\ &= \prod_{i=1}^k s_i(A)s_i(B). \end{aligned}$$

$$\prod_{i=1}^{n} s_i(AB) = |\det(AB)| = |\det(A)||\det(B)| = \prod_{i=1}^{n} s_i(A)s_i(B). \qquad \square$$

因为对于分量非负的向量, 弱对数优超蕴涵弱优超 (定理 3.23), 定理 4.6 给出下面的结果.

推论 4.7 设 $A, B \in M_n$, 则

$$s(AB) \prec_w \{s_i(A)s_i(B)\}_{i=1}^n.$$

下面的引理有时候可以把奇异值的问题转化为 Hermite 矩阵的特征值的问题.

引理 4.8 若 $A \in M_n$ 的奇异值是 $s_1, \cdots, s_n$, 则

$$\varphi(A) = \begin{pmatrix} 0 & A^* \\ A & 0 \end{pmatrix}$$

的特征值是 $s_1, \cdots, s_n, -s_n, \cdots, -s_1$.

证明 设 $U^*AV = \mathrm{diag}(s_1, \cdots, s_n)$ 为奇异值分解, U, V 为酉矩阵. 令

$$Q = \frac{1}{\sqrt{2}} \begin{pmatrix} V & V \\ U & -U \end{pmatrix}.$$

则 Q 为酉矩阵, 且

$$Q^*\varphi(A)Q = \mathrm{diag}(s_1, \cdots, s_n, -s_1, \cdots, -s_n). \qquad \square$$

显然, 映射 $\varphi : M_n \to M_{2n}$ 是可加的: $\varphi(A+B) = \varphi(A) + \varphi(B)$.

定理 4.9 设 $A, B \in M_n$, 则

$$s(A+B) \prec_w s(A) + s(B).$$

证明 设 $1 \leqslant k \leqslant n$, 应用引理 4.8 和定理 3.8, 有

$$\begin{aligned}
\sum_{i=1}^{k} s_i(A+B) &= \sum_{i=1}^{k} \lambda_i[\varphi(A+B)] = \sum_{i=1}^{k} \lambda_i[\varphi(A) + \varphi(B)] \\
&\leqslant \sum_{i=1}^{k} \lambda_i[\varphi(A)] + \sum_{i=1}^{k} \lambda_i[\varphi(B)] \\
&= \sum_{i=1}^{k} s_i(A) + \sum_{i=1}^{k} s_i(B) \\
&= \sum_{i=1}^{k} [s_i(A) + s_i(B)]. \qquad \square
\end{aligned}$$

定理 4.10 (Weyl) 设 $A \in M_n$ 的特征值为 $\lambda_1, \cdots, \lambda_n$. 则

$$\{|\lambda_i|\}_{i=1}^n \prec_{\log} s(A). \tag{4.3}$$

证明 不妨设 $|\lambda_1| \geqslant |\lambda_2| \geqslant \cdots \geqslant |\lambda_n|$. 用 $\rho(\cdot)$ 表示谱半径. 应用复合矩阵, 对 $1 \leqslant k \leqslant n$, 有

$$\begin{aligned}\prod_{i=1}^{k} |\lambda_i| = \rho(C_k(A)) &\leqslant \|C_k(A)\|_\infty \\ &= s_1[C_k(A)] \\ &= \prod_{i=1}^{k} s_i(A).\end{aligned}$$

$\prod_{i=1}^{n} |\lambda_i| = |\det A| = \prod_{i=1}^{n} s_i(A)$. 这就证明了 (4.3). □

推论 4.11 设 $A \in M_n$ 的特征值为 $\lambda_1, \cdots, \lambda_n$. 则

$$\{|\lambda_i|\}_{i=1}^{n} \prec_w s(A).$$

特别

$$|\mathrm{tr}A| \leqslant \sum_{i=1}^{n} s_i(A).$$

下面的结果说明定理 4.10 的逆命题也成立.

定理 4.12 (Horn [57]) 给定复数 $\lambda_1, \cdots, \lambda_n$ 和非负实数 $s_1, \cdots, s_n$. 若

$$\{|\lambda_i|\}_{i=1}^{n} \prec_{\log} \{s_i\}_{i=1}^{n}.$$

则存在一个 n 阶复矩阵以 $\lambda_1, \cdots, \lambda_n$ 为特征值并且以 $s_1, \cdots, s_n$ 为奇异值.

定理 4.10 和定理 4.12 合起来说明对数优超关系 (4.3) 是一般复方阵的特征值和奇异值所满足的全部关系.

下面来证明一个有一定深度的奇异值不等式 (定理 4.17), 为此需要一些引理.

引理 4.13 (Löwner) (证明见 [15]) 设 $f(t)$ 是 $[0, \infty)$ 上的算子单调函数, 则存在 $[0, \infty)$ 上的一个正测度 μ 使得

$$f(t) = \alpha + \beta t + \int_0^\infty \frac{st}{s+t}\, d\mu(s), \tag{4.4}$$

其中 α 是个实数, $\beta \geqslant 0$.

于是对于半正定矩阵 A 有

$$f(A) = \alpha I + \beta A + \int_0^\infty (sA)(sI + A)^{-1}\, d\mu(s).$$

带有参数的矩阵的定积分按元素方式理解, 即若 $G(s) = (g_{ij}(s))_{n\times n}$, 则

$$\int_0^\infty G(s)d\mu(s) = \left(\int_0^\infty g_{ij}(s)d\mu(s)\right)_{n\times n}.$$

引理 4.14 (Audenaert [8]) 设 $A, B \in M_n$ 半正定, $f(t)$ 是 $[0,\infty)$ 上的算子单调函数, 则

$$Af(A) + Bf(B) \geqslant \left(\frac{A+B}{2}\right)^{1/2} [f(A) + f(B)] \left(\frac{A+B}{2}\right)^{1/2} \tag{4.5}$$

证明 由推论 3.30,

$$(A+I)^{-1} + (B+I)^{-1} \geqslant 4(2I + A + B)^{-1}. \tag{4.6}$$

记

$$C_k = A^k(A+I)^{-1} + B^k(B+I)^{-1}, \quad M = (A+B)/2.$$

(4.6) 可写成

$$C_0 \geqslant 2(I+M)^{-1}.$$

因此,

$$C_0 + M^{1/2}C_0M^{1/2} \geqslant 2(I+M)^{-1} + 2M^{1/2}(I+M)^{-1}M^{1/2} = 2I. \tag{4.7}$$

注意到 $C_k + C_{k+1} = A^k + B^k$, 特别 $C_0 + C_1 = 2I$. (4.7) 可写成

$$M^{1/2}(2I - C_1)M^{1/2} \geqslant C_1. \tag{4.8}$$

因为 $C_1 + C_2 = 2M$, (4.8) 等价于 $C_2 \geqslant M^{1/2}C_1M^{1/2}$, 即

$$A^2(A+I)^{-1} + B^2(B+I)^{-1} \geqslant \frac{1}{2}(A+B)^{1/2}[A(A+I)^{-1} + B(B+I)^{-1}](A+B)^{1/2}.$$

对于正数 s, 在上面的不等式中分别将 A, B 换成 $s^{-1}A, s^{-1}B$, 然后再将两边同乘以 s^2, 得

$$sA^2(A+sI)^{-1}+sB^2(B+sI)^{-1} \geqslant \frac{1}{2}(A+B)^{1/2}[sA(A+sI)^{-1}+sB(B+sI)^{-1}](A+B)^{1/2}. \tag{4.9}$$

因为 $(A+B)^2 \leqslant 2(A^2+B^2)$, 对 $\alpha \in \mathbb{R}, \beta \geqslant 0$, 有

$$A(\alpha I + \beta A) + B(\alpha I + \beta B) \geqslant \frac{1}{2}(A+B)^{1/2}[2\alpha I + \beta(A+B)](A+B)^{1/2}. \tag{4.10}$$

因为 f 在 $[0,\infty)$ 上算子单调, 它有积分表达式(4.4), 对 (4.9) 两边关于 $\mu(s)$ 求积分再和(4.10) 相加就得到了 (4.5). □

应用 Weyl 单调性原理及 $\sigma(XY) = \sigma(YX)$, 引理 4.14 给出下面的推论.

推论 4.15 (Audenaert [8]) 设 $A, B \in M_n$ 半正定, $f(t)$ 是 $[0,\infty)$ 上的算子单调函数, 则

$$\lambda_j(Af(A) + Bf(B)) \geqslant \frac{1}{2}\lambda_j[(A+B)(f(A)+f(B))], \quad j = 1, \cdots, n.$$

引理 4.15 (Tao [106]) 设 $M, N, K \in M_n$. 若

$$Z = \begin{pmatrix} M & K \\ K^* & N \end{pmatrix} \geqslant 0,$$

则

$$\lambda_j(Z) \geqslant 2s_j(K), \quad j = 1, \cdots, n.$$

证明 记 $Q = \begin{pmatrix} 0 & K \\ K^* & 0 \end{pmatrix}$, 则

$$0 \leqslant \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \begin{pmatrix} M & K \\ K^* & N \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} = \begin{pmatrix} M & -K \\ -K^* & N \end{pmatrix} = Z - 2Q.$$

所以 $Z \geqslant 2Q$. 由 Weyl 单调性原理及引理 4.8, 得

$$\lambda_j(Z) \geqslant 2\lambda_j(Q) = 2s_j(K), \quad j = 1, \cdots, n.$$ □

定理 4.17 (Audenaert [8]) 设 $A, B \in M_n$ 半正定, $0 \leqslant t \leqslant 1$. 则

$$s_j(A^t B^{1-t} + A^{1-t} B^t) \leqslant s_j(A + B), \quad j = 1, \cdots, n. \tag{4.11}$$

证明 Löwner-Heinz 不等式 (定理 3.24) 表明: 若 $0 \leqslant r \leqslant 1$, 则 $f(x) = x^r$ 是 $[0, \infty)$ 上的算子单调函数. 应用推论 4.15 得

$$\begin{aligned} \lambda_j(A^{r+1} + B^{r+1}) &\geqslant \frac{1}{2}\lambda_j[(A+B)(A^r + B^r)] \\ &= \frac{1}{2}\lambda_j \left[\begin{pmatrix} A^{r/2} \\ B^{r/2} \end{pmatrix} (A+B) \left(A^{r/2}, B^{r/2} \right) \right] \qquad (4.12) \\ &= \frac{1}{2}\lambda_j \left[\begin{pmatrix} A^{1/2} \\ B^{1/2} \end{pmatrix} (A^r + B^r) \left(A^{1/2}, \ B^{1/2} \right) \right]. \qquad (4.13) \end{aligned}$$

应用引理 4.15 于 (4.12) 得

$$\begin{aligned} \lambda_j(A^{r+1} + B^{r+1}) &\geqslant s_j[A^{r/2}(A+B)B^{r/2}] \\ &= s_j(A^{1+r/2}B^{r/2} + A^{r/2}B^{1+r/2}). \end{aligned} \tag{4.14}$$

在不等式 (4.14) 中分别将 A, B 换成 $A^{1/(r+1)}, B^{1/(r+1)}$ 就得到 (4.11) 对应于 $t = (1 + r/2)/(1 + r)$ 的情形. 因为 $0 \leqslant r \leqslant 1, 3/4 \leqslant t \leqslant 1$. 将 t 换成 $1 - t$ (4.11) 不变, 所以 (4.11) 对于 $0 \leqslant t \leqslant 1/4$ 和 $3/4 \leqslant t \leqslant 1$ 都成立. 用同样的方法从 (4.13) 开始我们可以证明 (4.11) 对 $t = (1/2 + r)/(1 + r)$ 成立, 此时 t 的变化范围是余下的情形 $1/4 \leqslant t \leqslant 3/4$. 这就完成了证明. □

利用矩阵的极分解, 不等式 (4.11) 的 $t = 1/2$ 的情形可以写成

推论 4.18 (Bhatia-Kittaneh [23]) 设 $A, B \in M_n$. 则

$$2s_j(AB^*) \leqslant s_j(A^*A + B^*B), \quad j = 1, \cdots, n.$$

定理 4.17 是在 [116] 中提出的一个猜想, 陶云星 [106] 证明了 $t = 1/4$ 和 $t = 3/4$ 的情形.

§4.2 对称规度函数

对于 $x = (x_1, \cdots, x_n) \in \mathbb{C}^n$. 记 $|x| = (|x_1|, \cdots, |x_n|)$. 对于 $x = (x_1, \cdots, x_n), y = (y_1, \cdots, y_n) \in \mathbb{R}^n$, $x \leqslant y$ 表示 $x_i \leqslant y_i, i = 1, \cdots, n$.

$\mathbb{C}^n$ 上的一个范数称为绝对的, 若

$$\| |x| \| = \|x\|, \quad \forall x \in \mathbb{C}^n;$$

$\|\cdot\|$ 称为单调的, 若

$$x, y \in \mathbb{C}^n, \quad |x| \leqslant |y| \Rightarrow \|x\| \leqslant \|y\|.$$

引理 4.19 $\mathbb{C}^n$ 上的一个范数是单调的当且仅当它是绝对的.

证明 单调显然蕴涵绝对. 反过来, 设 $\|\cdot\|$ 是绝对的. 要证单调性只要证对于 $x = (x_1, \cdots, x_n)$, $y = (y_1, \cdots, y_n) \in \mathbb{C}^n$, 若 $x_i = t_i y_i, 0 \leqslant t_i \leqslant 1, i = 1, \cdots, n$ 则 $\|x\| \leqslant \|y\|$. 先证明除了某个 t_k 外其余的 t_i 都等于 1 的情形. 记 $a = \frac{1+t}{2}, b = \frac{1-t}{2}$, 则 $a + b = 1, a - b = t$.

$$\begin{aligned}
\|(y_1, \cdots, y_{k-1}, ty_k, y_{k+1}, \cdots, y_n)\| &= \|a(y_1, \cdots, y_n) + b(y_1, \cdots, y_{k-1}, \\
&\quad -y_k, y_{k+1}, \cdots, y_n)\| \\
&\leqslant a\|(y_1, \cdots, y_n)\| + b\|(y_1, \cdots, y_{k-1}, \\
&\quad -y_k, y_{k+1}, \cdots, y_n)\| \\
&= \|(y_1, \cdots, y_n)\|.
\end{aligned}$$

逐次应用这一特殊情形就可以证明一般情形. □

对于 $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$ 和 $\sigma \in S_n$, 记 $x_\sigma = (x_{\sigma(1)}, \cdots, x_{\sigma(n)})$.

一个映射 $\Phi : \mathbb{R}^n \to \mathbb{R}_+$ 称为对称规度函数, 若

(i) Φ 是一个绝对范数;

(ii) $\Phi(x_\sigma) = \Phi(x), \forall x \in \mathbb{R}^n, \forall \sigma \in S_n$.

例如, $\mathbb{R}^n$ 上的 l_p 范数 $(p \geqslant 1)$:

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

和 Fan k-范数

$$\Phi_k(x) = \max\left\{\sum_{j=1}^{k} |x_{i_j}| : 1 \leqslant i_1 < \cdots < i_k \leqslant n\right\}$$

都是对称规度函数, 这里 $x = (x_1, \cdots, x_n)$.

根据引理 4.19, 对称规度函数是单调的.

定理 4.20　设 $x, y \in \mathbb{R}_+^n$, 且 $x \prec_w y$. 若 Φ 是个对称规度函数, 则 $\Phi(x) \leqslant \Phi(y)$.

证明　由定理 3.19(ii), 存在 $u \in \mathbb{R}^n$ 使得 $x \leqslant u$, 且 $u \prec y$. 根据 Rado 定理 (定理 3.14), u 是 y 的置换的凸组合. 于是

$$u = \sum_{\sigma \in S_n} t_\sigma y_\sigma, \quad 0 \leqslant t_\sigma \leqslant 1, \quad \sum_\sigma t_\sigma = 1.$$

利用范数 Φ 的单调性、置换不变性得

$$\begin{aligned}\Phi(x) &\leqslant \Phi(u) \\ &= \Phi\Big(\sum_{\sigma \in S_n} t_\sigma y_\sigma\Big) \\ &\leqslant \sum_{\sigma \in S_n} t_\sigma \Phi(y_\sigma) \\ &= \sum_{\sigma \in S_n} t_\sigma \Phi(y) \\ &= \Phi(y).\end{aligned}$$

□

§4.3 酉不变范数

M_n 上的范数 $\|\cdot\|$ 称为酉不变的, 若

$$\|UAV\| = \|A\|$$

对任意的 $A \in M_n$ 和任意的酉矩阵 $U, V \in M_n$ 成立. 显然谱范数 $\|\cdot\|_\infty$ 和 Frobenius 范数$\|\cdot\|_F$ 都是酉不变的.

根据奇异值分解定理, 每个酉不变范数都是矩阵的奇异值的一个函数. 是怎样的函数呢? 下面我们将给出刻画.

记 $\mathbb{R}_+^n \downarrow = \{(x_1, \cdots, x_n) \mid x_1 \geqslant \cdots \geqslant x_n \geqslant 0, x_i \in \mathbb{R}\}$. 将 $y \in \mathbb{R}^n$ 的分量按递减的顺序重排为 $y_{[1]} \geqslant \cdots \geqslant y_{[n]}$, 记 $y \downarrow = (y_{[1]}, \cdots, y_{[n]})$.

$\mathbb{R}^n$ 上的一个对称规度函数由它在 $\mathbb{R}_+^n \downarrow$ 上的取值唯一确定, 因此, 若定义在 $\mathbb{R}_+^n \downarrow$ 上的一个函数能够延拓为 $\mathbb{R}^n$ 上的对称规度函数, 则延拓是唯一的. 对 $\phi : \mathbb{R}_+^n \downarrow \to \mathbb{R}$, 定义 $\widehat{\phi} : \mathbb{R}^n \to \mathbb{R}$, 为

$$\widehat{\phi}(x) = \phi(|x| \downarrow), \quad \forall x \in \mathbb{R}^n.$$

则 $\widehat{\phi}$ 是 ϕ 的延拓: 对于 $x\in\mathbb{R}_+^n\downarrow$, $\widehat{\phi}(x)=\phi(x)$.

我们总是将 $A\in M_n$ 的奇异值按降序排列为 $s_1\geqslant\cdots\geqslant s_n$ 并记 $s(A)=(s_1,\cdots,s_n)$.

定理 4.21 (von Neumann) 给定 $\mathbb{R}_+^n\downarrow$ 上的实值函数 ϕ, 定义 M_n 上的函数

$$\|A\|_\phi=\phi(s(A)).$$

则 $\|\cdot\|_\phi$ 是酉不变范数当且仅当 $\widehat{\phi}$ 是对称规度函数.

证明 对于 $x=(x_1,\cdots,x_n)\in\mathbb{C}^n$ 用 $\operatorname{diag}(x)$ 记对角矩阵 $\operatorname{diag}(x_1,\cdots,x_n)$.

假设 $\|\cdot\|_\phi$ 是酉不变范数. $\widehat{\phi}$ 的正定性, 正齐次性, 绝对性, 以及 $\widehat{\phi}(x_\sigma)=\widehat{\phi}(x),\forall x\in\mathbb{R}^n,\forall\sigma\in S_n$ 都是显然的. 为了证明 $\widehat{\phi}$ 是对称规度函数我们只需要验证 $\widehat{\phi}$ 满足三角不等式.

$$\begin{aligned}\widehat{\phi}(x+y)&=\phi(|x+y|\downarrow)\\&=\|\operatorname{diag}(x+y)\|_\phi\\&=\|\operatorname{diag}(x)+\operatorname{diag}(y)\|_\phi\\&\leqslant\|\operatorname{diag}(x)\|_\phi+\|\operatorname{diag}(y)\|_\phi\\&=\phi(|x|\downarrow)+\phi(|y|\downarrow)\\&=\widehat{\phi}(x)+\widehat{\phi}(y).\end{aligned}$$

反过来, 假设 $\widehat{\phi}$ 是对称规度函数. $\forall A,B\in M_n$, 由定理 4.9 有 $s(A+B)\prec_w s(A)+s(B)$. 应用定理 4.20 有

$$\begin{aligned}\|A+B\|_\phi&=\phi(s(A+B))\\&\leqslant\phi(s(A)+s(B))\\&\leqslant\phi(s(A))+\phi(s(B))\\&=\|A\|_\phi+\|B\|_\phi.\end{aligned}$$

$\|\cdot\|_\phi$ 显然满足酉不变范数的其他定义性质, 所以 $\|\cdot\|_\phi$ 是酉不变范数. □

给定 $\phi:\mathbb{R}_+^n\downarrow\to\mathbb{R}$, 容易验证 $\widehat{\phi}$ 是对称规度函数当且仅当 ϕ 满足以下性质:

(i) 正定性: $\phi(x)\geqslant 0,\phi(x)=0\Leftrightarrow x=0$.

(ii) 正齐次性: $\phi(\gamma x)=\gamma\phi(x),\ \gamma\geqslant 0$.

(iii) 三角不等式: $\phi(x+y)\leqslant\phi(x)+\phi(y)$.

(iv) 关于弱优超单调: $x,y\in\mathbb{R}_+^n\downarrow,x\prec_w y\Rightarrow\phi(x)\leqslant\phi(y)$.

定理 4.21 的一个常用推论是: 如果 $\Phi:\mathbb{R}^n\to\mathbb{R}_+$ 是对称规度函数, 则 $\|A\|_\Phi=\Phi(s(A))$ 定义 M_n 上的一个酉不变范数. 让我们应用这个推论来得到几类酉不变范

数. 设 $p \geqslant 1$, $\mathbb{R}^n$ 上的 l_p 范数所对应的 M_n 上的范数

$$\|A\|_p = \left[\sum_{j=1}^{n} s_j^p(A)\right]^{1/p}$$

称为 Schatten p-范数. 设 $1 \leqslant k \leqslant n$,

$$\|A\|_{(k)} = \sum_{j=1}^{k} s_j(A)$$

称为 Fan k-范数. 其中, $\|\cdot\|_{(1)} = \|\cdot\|_\infty$ 是谱范数, $\|\cdot\|_{(n)} = \|\cdot\|_1$ 称为迹范数. 注意, $\|\cdot\|_2 = \|\cdot\|_F$.

给定 $\gamma = (\gamma_1, \cdots, \gamma_n), \gamma_1 \geqslant \cdots \geqslant \gamma_n \geqslant 0, \gamma_1 > 0$, 定义

$$\|A\|_\gamma = \sum_{j=1}^{n} \gamma_j s_j(A), \quad A \in M_n.$$

这是一个酉不变范数.

设 $V = \mathbb{C}^n$ 或 $\mathbb{R}^n$, $\|\cdot\|$ 是 V 上的一个范数. $\|\cdot\|$ 的对偶范数 $\|\cdot\|^D$ 定义为

$$\|x\|^D = \max\{|\langle x, y\rangle| : \|y\| = 1, y \in V\}.$$

引理 4.22 (对偶原理)([58], Thm 5.5.14) 设 $\|\cdot\|$ 是 $\mathbb{C}^n$ 或 $\mathbb{R}^n$ 上的一个范数, 则 $(\|\cdot\|^D)^D = \|\cdot\|$.

M_n 上的范数 $\|\cdot\|$ 关于 Frobenius 内积 $\langle A, B\rangle = \mathrm{tr}AB^*$ 的对偶范数定义为

$$\|A\|^D = \max\{|\mathrm{tr}AB^*| : \|B\| = 1, B \in M_n\}.$$

注意到 $\mathrm{vec} : M_n \to \mathbb{C}^{n^2}$ 是一个同构, 且 $\langle A, B\rangle = \langle \mathrm{vec}A, \mathrm{vec}B\rangle$, 由上述关于向量范数的对偶原理我们得到矩阵范数的对偶原理: $(\|\cdot\|^D)^D = \|\cdot\|$. 根据推论 4.11 和推论 4.7 我们得到: 若 $\|\cdot\|$ 是 M_n 上的酉不变范数, 则

$$\|A\| = \max\left\{\sum_{j=1}^{n} s_j(A)s_j(B) : \|B\|^D = 1, B \in M_n\right\}. \tag{4.15}$$

容易验证: 一个酉不变范数的对偶范数还是酉不变范数. 利用对偶原理和 (4.15) 我们得到

引理 4.23 设 $\|\cdot\|$ 是 M_n 上的酉不变范数, 记 $\Gamma = \{s(X) : \|X\|^D = 1, X \in M_n\}$. 则对于任何 $A \in M_n$,

$$\|A\| = \max\{\|A\|_\gamma : \gamma \in \Gamma\}. \tag{4.16}$$

定理 4.24 (Fan 支配原理) 设 $A, B \in M_n$. 若

$$\|A\|_{(k)} \leqslant \|B\|_{(k)}, \quad k = 1, \cdots, n,$$

则

$$\|A\| \leqslant \|B\|$$

对任何酉不变范数成立.

证明 利用 (4.16) 和等式

$$\|A\|_\gamma = \sum_{j=1}^{n-1} (\gamma_j - \gamma_{j+1}) \|A\|_{(j)} + \gamma_n \|A\|_{(n)},$$

其中 $\gamma = (\gamma_1, \cdots, \gamma_n)$, $\gamma_1 \geqslant \cdots \geqslant \gamma_n \geqslant 0$, $\gamma_1 > 0$. □

Fan 支配原理可以等价地叙述为: $\|A\| \leqslant \|B\|$ 对任何酉不变范数成立当且仅当 $s(A) \prec_w s(B)$.

引理 4.25 设 $T \in M_n$. 则对于 $k = 1, \cdots, n$

$$\|T\|_{(k)} = \min\{\|X\|_1 + k\|Y\|_\infty : T = X + Y, X, Y \in M_n\}.$$

证明 若 $T = X + Y$, 则

$$\|T\|_{(k)} \leqslant \|X\|_{(k)} + \|Y\|_{(k)} \leqslant \|X\|_1 + k\|Y\|_\infty.$$

另一方面, 设 $T = U\mathrm{diag}(s_1, \cdots, s_n)V$ 为奇异值分解, U, V 为酉矩阵, $s_1 \geqslant \cdots \geqslant s_n \geqslant 0$, 则

$$\begin{aligned} X &= U\mathrm{diag}(s_1 - s_k, s_2 - s_k, \cdots, s_k - s_k, 0, \cdots, 0)V, \\ Y &= U\mathrm{diag}(s_k, \cdots, s_k, s_{k+1}, \cdots, s_n)V \end{aligned}$$

满足 $T = X + Y$ 和

$$\|T\|_{(k)} = \|X\|_1 + k\|Y\|_\infty.$$ □

引理 4.26 设 $A, B \in M_n$. 则

$$\|\mathrm{diag}(s(A) - s(B))\|_1 \leqslant \|A - B\|_1. \tag{4.17}$$

证明 设 $G, H \in M_m$ 为 Hermite 矩阵, 我们将证明

$$\sum_{j=1}^{m} |\lambda_j(G) - \lambda_j(H)| \leqslant \|G - H\|_1. \tag{4.18}$$

用 $G - H$ 的 Jordan 分解我们有 $\|G - H\|_1 = \mathrm{tr}(G - H)_+ + \mathrm{tr}(G - H)_-$.

令 $C = G + (G-H)_- = H + (G-H)_+$, 则 $C \geqslant G, C \geqslant H$. 由 Weyl 单调性原理, $\lambda_j(C) \geqslant \lambda_j(G)$, $\lambda_j(C) \geqslant \lambda_j(H), j = 1, \cdots, m$. 于是

$$|\lambda_j(G) - \lambda_j(H)| \leqslant \lambda_j(2C) - \lambda_j(G) - \lambda_j(H).$$

从而

$$\sum_{j=1}^{m} |\lambda_j(G) - \lambda_j(H)| \leqslant \mathrm{tr}(2C - G - H) = \|G - H\|_1.$$

在 (4.18) 中令

$$G = \begin{pmatrix} 0 & A^* \\ A & 0 \end{pmatrix}, \quad H = \begin{pmatrix} 0 & B^* \\ B & 0 \end{pmatrix},$$

并利用引理 4.8 就得到了 (4.17). □

下面是一个关于奇异值扰动的结果.

定理 4.27 (Mirsky [88])　设 $A, B \in M_n$. 则

$$\|\mathrm{diag}(s(A) - s(B))\| \leqslant \|A - B\| \tag{4.19}$$

对任何酉不变范数成立.

证明　由 Fan 支配原理, 只需要证明 (4.19) 对于 Fan k-范数成立, $k = 1, \cdots, n$. 定理 4.3 表明 (4.19) 对于谱范数 $\|\cdot\|_\infty$ 成立 (在不等式 (4.1) 中取 $j = 1$), 而引理 4.26 说明 (4.19) 对于迹范数 $\|\cdot\|_1$ 成立. 下面利用这两个特殊情形和引理 4.25 来证明 (4.19) 对于 Fan$k$-范数成立.

任取一个满足 $1 \leqslant k \leqslant n$ 的固定的 k. 存在 $X, Y \in M_n$ 使得

$$A - B = X + Y, \quad \|A - B\|_{(k)} = \|X\|_1 + k\|Y\|_\infty.$$

从等式

$$\mathrm{diag}(s(A) - s(B)) = \mathrm{diag}(s(X+B) - s(B)) + \mathrm{diag}(s(A) - s(X+B))$$

出发, 我们有

$$\begin{aligned}
\|\mathrm{diag}(s(A) - s(B))\|_{(k)} &\leqslant \|\mathrm{diag}(s(X+B) - s(B))\|_1 + k\|\mathrm{diag}(s(A) - s(X+B))\|_\infty \\
&\leqslant \|X + B - B\|_1 + k\|A - X - B\|_\infty \\
&= \|X\|_1 + k\|Y\|_\infty \\
&= \|A - B\|_{(k)}.
\end{aligned}$$

□

对于 Hermite 矩阵 $G \in M_n$, 我们总是记 $\lambda(G) = (\lambda_1(G), \cdots, \lambda_n(G))$, 其中 $\lambda_1(G) \geqslant \cdots \geqslant \lambda_n(G))$ 是 G 的从大到小排列的特征值.

定理 4.28 (Lidskii [81]) 设 $G, H \in M_n$ 为 Hermite 矩阵, 则

$$\lambda(G) - \lambda(H) \prec \lambda(G - H).$$

证明 对于任何半正定矩阵 P, $s(P) = \lambda(P)$. Mirsky 定理可以等价地写成

$$|s(A) - s(B)| \prec_w s(A - B).$$

选取实数 a, b 满足

$$G + aI \geqslant H + bI \geqslant 0,$$

并令 $A = G + aI$, $B = H + bI$ 得

$$\{|\lambda_j(G) - \lambda_j(H) + a - b|\} \prec_w \{\lambda_j(G - H) + a - b\}.$$

将这个关系与显然的事实

$$\{\lambda_j(G) - \lambda_j(H) + a - b\} \prec_w \{|\lambda_j(G) - \lambda_j(H) + a - b|\}$$

合起来得到

$$\{\lambda_j(G) - \lambda_j(H) + a - b\} \prec_w \{\lambda_j(G - H) + a - b\}.$$

所以 $\lambda(G) - \lambda(H) \prec_w \lambda(G - H)$. 最后再用 $\mathrm{tr}G - \mathrm{tr}H = \mathrm{tr}(G - H)$ 就得到要证的结果. □

现在来讨论一个酉不变范数和数值半径的大小关系.

引理 4.29 (Toeplitz) 设 $A \in M_n$. 则

$$w(A) \leqslant \|A\|_\infty \leqslant 2w(A).$$

证明 第一个不等式显然, 只需要证第二个不等式. 记

$$A_1 = (A + A^*)/2, A_2 = (A - A^*)/2.$$

则 A_1 和 A_2 都是正规矩阵且 $A = A_1 + A_2$. 因为正规矩阵的谱范数等于数值半径,

$$\begin{aligned} \|A\|_\infty = \|A_1 + A_2\|_\infty &\leqslant \|A_1\|_\infty + \|A_2\|_\infty \\ &= w(A_1) + w(A_2) \\ &\leqslant w(A) + w(A^*) \\ &= 2w(A). \end{aligned}$$ □

引理 4.30 (Marcus-Sandy [84]) 设 $A \in M_n$. 则

$$w(A) \leqslant \|A\|_1 \leqslant nw(A).$$

证明 第一个不等式显然, 只需要证第二个不等式. 设 $A = UP$ 为极分解, U 为酉矩阵, P 半正定. 于是 $U^*A = P$. 设 $U^* = VDV^*$ 为谱分解, V 为酉矩阵, $D = \mathrm{diag}(d_1, \cdots, d_n), |d_i| = 1, i = 1, \cdots, n$. 我们得到

$$DV^*AV = V^*PV.$$

记 $V = (v_1, \cdots, v_n), v_i \in \mathbb{C}^n$, 则 v_i 是单位向量且

$$\begin{aligned}\|A\|_1 = \mathrm{tr}P &= \mathrm{tr}V^*PV \\ &= \mathrm{tr}DV^*AV \\ &= |\sum_{i=1}^n d_i v_i^* A v_i| \\ &\leqslant \sum_{i=1}^n |d_i v_i^* A v_i| \\ &= \sum_{i=1}^n |v_i^* A v_i| \\ &\leqslant nw(A).\end{aligned}$$

□

用 E_{ij} 记 (i, j) 元素为 1 其他元素都为 0 的 n 阶矩阵.

定理 4.31 (Johnson-Li [67]) 设 $\|\cdot\|$ 是 M_n 上的一个酉不变范数, 记

$$G = \begin{cases} 2\sum_{i=1}^{n/2} E_{2i-1,2i}, & \text{当 } n \text{ 为偶数}, \\ 2\sum_{i=1}^{(n-1)/2} E_{2i-1,2i} + E_{n,n}, & \text{当 } n \text{ 为奇数}. \end{cases}$$

则对于任何 $A \in M_n$,

$$\|E_{11}\|w(A) \leqslant \|A\| \leqslant \|G\|w(A). \tag{4.20}$$

证明 当 $A = 0$ 时, (4.20) 成立. 假设 $A \neq 0$, 令 $A' = A/w(A)$. 由引理 4.29, $1 \leqslant s_1(A') \leqslant 2$; 由引理 4.30, $\sum_{i=1}^n s_i(A') \leqslant n$. 而当 n 为偶数时, $s(G)$ 含有 $n/2$ 个 2 和 $n/2$ 个 0; 当 n 为奇数时, $s(G)$ 含有 $(n-1)/2$ 个 2, 1 个 1, $(n-1)/2$ 个 0. 所以

$$s(E_{11}) \prec_w s(A') \prec_w s(G).$$

由 Fan 支配原理, (4.20) 对于任何酉不变范数成立. □

(4.20) 中的两个不等式都是精确的: 第一个不等式在 $A = E_{11}$ 时为等式; 第二个不等式在 $A = G$ 时为等式.

推论 4.32 (Johnson-Li [67]) 设 $p \geqslant 1, 1 \leqslant k \leqslant n, A \in M_n$. 则

$$w(A) \leqslant \|A\|_p \leqslant \alpha w(A),$$

$$w(A) \leqslant \|A\|_{(k)} \leqslant \beta w(A),$$

其中

$$\alpha = \begin{cases} (2^{p-1}n)^{1/p}, & \text{当 } n \text{ 为偶数}, \\ (2^{p-1}(n-1)+1)^{1/p}, & \text{当 } n \text{ 为奇数}, \end{cases}$$

$$\beta = \begin{cases} 2k, & \text{当 } k \leqslant n/2, \\ n, & \text{当 } k > n/2. \end{cases}$$

证明 应用定理 4.31, 计算 $\|G\|_p$ 和 $\|G\|_{(k)}$. □

§4.4 矩阵的笛卡儿分解

本节中 $i=\sqrt{-1}$. 每个矩阵 $T \in M_n$ 都可以唯一地写成 $T = A + iB$, 其中 A, B 为 Hermite 矩阵:

$$A = \frac{T+T^*}{2}, \quad B = \frac{T-T^*}{2i}.$$

这称为矩阵 T 的笛卡儿分解. A 和 B 分别称为 T 的实部和虚部. 我们将要研究 T 的奇异值和 A, B 的特征值之间的关系. 记 T 的奇异值为 $s_1 \geqslant \cdots \geqslant s_n$, A, B 的特征值分别为 α_j 和 β_j 并且是按模从大到小排列的:

$$|\alpha_1| \geqslant \cdots \geqslant |\alpha_n|, \quad |\beta_1| \geqslant \cdots \geqslant |\beta_n|.$$

我们需要下面的引理.

引理 4.33 设 G, $H \in M_n$ 为 Hermite 矩阵, 则

$$\{\lambda_j(G) + \lambda_{n-j+1}(H)\} \prec \lambda(G+H) \prec \lambda(G) + \lambda(H).$$

这个引理中的第一个优超关系是 Lidskii 定理 (定理 4.28) 的变形, 第二个关系是 Fan 的定理 3.8.

定理 4.34 (Ando-Bhatia [4]) 下面的优超关系成立:

$$\{|\alpha_j + i\beta_{n-j+1}|^2\} \prec \{s_j^2\}, \tag{4.21}$$

$$\{(s_j^2 + s_{n-j+1}^2)/2\} \prec \{|\alpha_j + i\beta_j|^2\}. \tag{4.22}$$

证明 在引理 4.33 中令 $G = A^2$, $H = B^2$ 我们得

$$\{|\alpha_j + i\beta_{n-j+1}|^2\} \prec \lambda(A^2+B^2) \prec \{|\alpha_j + i\beta_j|^2\}. \tag{4.23}$$

注意

$$A^2 + B^2 = (T^*T + TT^*)/2,$$

$$\lambda_j(T^*T) = \lambda_j(TT^*) = s_j^2.$$

再次用引理 4.33, 令 $G=T^*T/2$, $H=TT^*/2$, 得

$$\{(s_j^2+s_{n-j+1}^2)/2\}\prec\lambda(A^2+B^2)\prec\{s_j^2\}. \tag{4.24}$$

将 (4.23) 和 (4.24) 合起来就得到了 (4.21) 和 (4.22). □

下面我们来推导几个关于 Schatten p-范数的不等式.

定理 4.35 (Bhatia-Kittaneh [24]) 若 $2\leqslant p\leqslant\infty$, 则

$$2^{2/p-1}(\|A\|_p^2+\|B\|_p^2)\leqslant\|T\|_p^2\leqslant 2^{1-2/p}(\|A\|_p^2+\|B\|_p^2); \tag{4.25}$$

若 $1\leqslant p\leqslant 2$, 则

$$2^{2/p-1}(\|A\|_p^2+\|B\|_p^2)\geqslant\|T\|_p^2\geqslant 2^{1-2/p}(\|A\|_p^2+\|B\|_p^2). \tag{4.26}$$

证明 当 $p\geqslant 2$ 时, $f(t)=t^{p/2}$ 是 $[0,\infty)$ 上的凸函数, 当 $1\leqslant p\leqslant 2$ 时 $g(t)=-t^{p/2}$ 是 $[0,\infty)$ 上的凸函数. 应用定理 3.20, (4.22) 得

$$\{2^{-p/2}(s_j^2+s_{n-j+1}^2)^{p/2}\}\prec_w\{|\alpha_j+i\beta_j|^p\},\quad p\geqslant 2,$$

$$\{-2^{-p/2}(s_j^2+s_{n-j+1}^2)^{p/2}\}\prec_w\{-|\alpha_j+i\beta_j|^p\},\quad 1\leqslant p\leqslant 2.$$

特别, 我们有

$$\sum_{j=1}^n(s_j^2+s_{n-j+1}^2)^{p/2}\leqslant 2^{p/2}\sum_{j=1}^n|\alpha_j+i\beta_j|^p,\quad p\geqslant 2, \tag{4.27}$$

$$\sum_{j=1}^n(s_j^2+s_{n-j+1}^2)^{p/2}\geqslant 2^{p/2}\sum_{j=1}^n|\alpha_j+i\beta_j|^p,\quad 1\leqslant p\leqslant 2. \tag{4.28}$$

因为对于固定的非负实数 a, b, 函数 $t\mapsto(a^t+b^t)^{1/t}$ 在 $(0,\infty)$ 上递减,

$$s_j^p+s_{n-j+1}^p\leqslant(s_j^2+s_{n-i+1}^2)^{p/2},\quad p\geqslant 2,$$

而当 $1\leqslant p\leqslant 2$ 时这个不等式反向. 因此, 从 (4.27) 和 (4.28) 得

$$\sum_{j=1}^n s_j^p\leqslant 2^{p/2-1}\sum_{j=1}^n|\alpha_j+i\beta_j|^p,\quad p\geqslant 2, \tag{4.29}$$

$$\sum_{j=1}^n s_j^p\geqslant 2^{p/2-1}\sum_{j=1}^n|\alpha_j+i\beta_j|^p,\quad 1\leqslant p\leqslant 2. \tag{4.30}$$

最后应用关于非负实数序列 $\{x_j\}$, $\{y_j\}$ 的 Minkowski 不等式

$$\left(\sum_j(x_j+y_j)^r\right)^{1/r}\leqslant\left(\sum_j x_j^r\right)^{1/r}+\left(\sum_j y_j^r\right)^{1/r},\quad r\geqslant 1,$$

$$\left(\sum_j(x_j+y_j)^r\right)^{1/r}\geqslant\left(\sum_j x_j^r\right)^{1/r}+\left(\sum_j y_j^r\right)^{1/r},\quad 0<r\leqslant 1$$

于不等式 (4.29) 和 (4.30) 我们就得到了 (4.25) 和 (4.26) 中右边的不等式. (4.25) 和 (4.26) 中左边的不等式可以用类似的方法从优超关系 (4.21) 推导出来. □

从优超关系 (4.24) 我们可以得到下面的

定理 4.36 (Bhatia-Kittaneh [24]) 若 $2 \leqslant p \leqslant \infty$, 则

$$\|(A^2+B^2)^{1/2}\|_p \leqslant \|T\|_p \leqslant 2^{1/2-1/p}\|(A^2+B^2)^{1/2}\|_p;$$

若 $1 \leqslant p \leqslant 2$, 则

$$\|(A^2+B^2)^{1/2}\|_p \geqslant \|T\|_p \geqslant 2^{1/2-1/p}\|(A^2+B^2)^{1/2}\|_p.$$

定理 4.37 (Queiro-Duarte [92]) 下面的行列式不等式成立:

$$|\det T| \leqslant \prod_{j=1}^{n} |\alpha_j + i\beta_{n-j+1}|.$$

证明 (Ando-Bhatia [4]) 因为 M_n 中的可逆矩阵是稠密的, 我们可以假设 T 是可逆的. 那么每个 $s_j > 0$ 并且 (4.21) 蕴涵对于每个 j, $|\alpha_j+i\beta_{n-j+1}| > 0$. 应用 $(0,\infty)$ 上的凸函数 $f(t) = -\frac{1}{2}\log t$ 于优超关系 (4.21) 就完成了证明. □

笛卡儿分解中实部和虚部都为半正定矩阵的情形以及只有实部半正定的情形分别在 [25], [26] 中被研究过.

习 题

1. (Fan-Hoffman [41]). 设 $A \in M_n$, 记 $\mathrm{Re}A = (A+A^*)/2$. 则

$$\lambda_j(\mathrm{Re}A) \leqslant s_j(A), \quad j = 1, \cdots, n.$$

2. (Thompson [108]). 设 $A, B \in M_n$, 则存在酉矩阵 $U, V \in M_n$ 满足

$$|A+B| \leqslant U|A|U^* + V|B|V^*.$$

3. $G \in M_n$ 称为一个秩 k 部分等距矩阵, 若

$$s_1(G) = \cdots = s_k(G) = 1, \quad s_{k+1}(G) = \cdots = s_n(G) = 0.$$

证明对 $X \in M_n$,

$$\sum_{j=1}^{k} s_j(X) = \max\{|\mathrm{tr}(XG)| : G\text{是个秩 } k \text{ 部分等距矩阵}, G \in M_n\}.$$

再用这个表达式证明定理 4.9.

4. 设 $A=(a_{ij})\in M_n$, 则

$$(|a_{11}|,|a_{22}|,\cdots,|a_{nn}|)\prec_w s(A).$$

5. 设 $A,B\in M_n$, 则

$$s_j(AB)\leqslant \|A\|_\infty s_j(B),\quad s_j(AB)\leqslant \|B\|_\infty s_j(A),\quad j=1,\cdots,n.$$

6. ([115]) 设 $A,B\in M_n$ 半正定, 则

$$s_j(A-B)\leqslant s_j\left(\begin{pmatrix}A & 0\\ 0 & B\end{pmatrix}\right),\quad j=1,\cdots,n.$$

7. ([120]) 设 $A_0\in M_n$ 正定, $A_i\in M_n$ 半正定, $i=1,\cdots,k$, 则

$$\mathrm{tr}\sum_{j=1}^{k}\left(\sum_{i=0}^{j}A_i\right)^{-2}A_j<\mathrm{tr}A_0^{-1}.$$

8. 设 p,q 为正实数, 满足 $\frac{1}{p}+\frac{1}{q}=1$, 设 $x,y\in\mathbb{R}_+^n$, 则对 $\mathbb{R}^n$ 上的任何对称规度函数 φ 有

$$\varphi(x\circ y)\leqslant[\varphi(x^p)]^{1/p}[\varphi(y^q)]^{1/q},$$

其中 x^p 表示将 x 的每个分量取 p 次方所得的向量.

9. 设 $\|\cdot\|$ 是 M_n 上的酉不变范数, 则 $\|\cdot\|$ 次可乘当且仅当

$$\|\mathrm{diag}(1,0,\cdots,0)\|\geqslant 1.$$

10. 设 $A,B\in M_n$ 并且 AB 为 Hermite 矩阵, 则对任何酉不变范数

$$\|AB\|\leqslant\|\mathrm{Re}(BA)\|.$$

11. M_n 上的范数 $\|\cdot\|$ 称为是对称的, 若

$$\|ABC\|\leqslant\|A\|_\infty\|C\|_\infty\|B\|,\quad\forall A,B,C\in M_n.$$

证明: $\|\cdot\|$ 对称当且仅当 $\|\cdot\|$ 是酉不变的.

12. 设 p,q 为正实数, 满足 $\frac{1}{p}+\frac{1}{q}=1$, 则对 $A,B\in M_n$ 和酉不变范数有

$$\|AB\|\leqslant\||A|^p\|^{1/p}\||B|^q\|^{1/q}.$$

13. (Bhatia-Davis [20]) 设 $A,B,X\in M_n$, 则

$$\|AXB^*\|\leqslant\frac{1}{2}\|A^*AX+XB^*B\|$$

对任何酉不变范数成立.

14. 设 $A,B\in M_n$, 则对 M_{2n} 上的任何酉不变范数有

$$\frac{1}{2}\left\|\begin{pmatrix}A+B & 0\\ 0 & A+B\end{pmatrix}\right\|\leqslant\left\|\begin{pmatrix}A & 0\\ 0 & B\end{pmatrix}\right\|\leqslant\left\|\begin{pmatrix}|A|+|B| & 0\\ 0 & 0\end{pmatrix}\right\|.$$

15. (Fan-Hoffman [41]) 设 $A, H \in M_n$, 其中 H 为 Hermite 矩阵, 则

$$\|A - \mathrm{Re}A\| \leqslant \|A - H\|$$

对任何酉不变范数成立.

16. (Fan-Hoffman [41]) 设 $A \in M_n$, $A = UP$ 为极分解, U 为酉矩阵, P 为半正定阵. 若 $W \in M_n$ 为酉矩阵, 则

$$\|A - U\| \leqslant \|A - W\| \leqslant \|A + U\|$$

对任何酉不变范数成立.

17. (Ando-Zhan [5]) 设 $A, B \in M_n$ 半正定, $\|\cdot\|$ 是一个酉不变范数. 则

$$\|(A+B)^r\| \leqslant \|A^r + B^r\| \quad (0 < r \leqslant 1),$$

$$\|(A+B)^r\| \geqslant \|A^r + B^r\| \quad (1 \leqslant r < \infty).$$

第五章 矩阵扰动

设 $f:\Omega\to\Delta$ 是一个映射, 关于 f 的扰动问题是: 当变元的取值从 A 变成 B 时, $f(A)$ 和 $f(B)$ 的差别有多大? 扰动问题不仅来源于工程技术, 也出现在某些纯粹数学问题中. 如果 Ω 是由矩阵组成的集合, 我们就得到矩阵扰动. 例如, 推论 3.4 是关于 Hermite 矩阵的特征值的扰动定理, 而定理 4.27 是关于一般矩阵的奇异值的扰动定理. 关于矩阵扰动的两本专著是 [17] 和 [104]. 这里说的扰动并不意味着一定是小的扰动, 而是指一般的变化.

§5.1 特征值

在本节我们将研究在以下几种情形下特征值的变化: (1) A,B 为一般矩阵; (2) A,B 为正规矩阵; (3) A 为正规矩阵, B 为一般矩阵; (4) A 为 Hermite 矩阵, B 为反 Hermite 矩阵.

设 $\sigma(A)=\{\lambda_1,\cdots,\lambda_n\}$ 和 $\sigma(B)=\{\mu_1,\cdots,\mu_n\}$ 分别是 $A,B\in M_n$ 的特征值的集合, A 和 B 的谱的最佳匹配距离定义为

$$d(\sigma(A),\sigma(B))=\min_{\tau\in S_n}\max_{1\leqslant i\leqslant n}|\lambda_i-\mu_{\tau(i)}|,$$

其中 S_n 表示 $1,2,\cdots,n$ 的所有置换的集合.

引理 5.1 ([22]) *设 Γ 是复平面上以 a,b 为端点的一条连续曲线, $\lambda_1,\cdots,\lambda_n$ 是复平面上的点. 则 Γ 上存在一点 λ 满足*

$$\prod_{i=1}^{n}|\lambda-\lambda_i|\geqslant 2^{1-2n}|b-a|^n. \tag{5.1}$$

证明 设 L 是通过 a,b 的直线, S 是连接 a,b 的线段:

$$S=\{z\in\mathbb{C}|z=a+t(b-a),\ 0\leqslant t\leqslant 1\}.$$

对于平面上任意一点 z, 用 z' 记 z 在 L 上的投影. 则对 $\forall z,w,|z-w|\geqslant|z'-w'|$. 对于给定的点 λ_i, 设 $\lambda_i'=a+t_i(b-a)$, $t_i\in\mathbb{R}$, $i=1,\cdots,n$. L 上的每一点 z 都可以写成 $z=a+t(b-a)$, $t\in\mathbb{R}$. 于是

$$\prod_{i=1}^n|z-\lambda_i'|=\prod_{i=1}^n|(t-t_i)(b-a)|=|b-a|^n\prod_{i=1}^n|t-t_i|. \tag{5.2}$$

Chebyshev 的一个经典结果 ([95], p.31) 说: 如果 p 是个次数为 n 首项系数为 1 的多项式, 则

$$\max_{0\leqslant t\leqslant 1}|p(t)|\geqslant 2^{1-2n}.$$

所以, 从 (5.2) 我们知道存在 $z\in S$ 使得

$$\prod_{i=1}^n|z-\lambda_i'|\geqslant 2^{1-2n}|b-a|^n. \tag{5.3}$$

因为 Γ 是连接 a 和 b 的连续曲线, 存在 $\lambda\in\Gamma$ 满足 $\lambda'=z$. 又因为 $|\lambda-\lambda_i|\geqslant|z-\lambda_i'|$, $i=1,\cdots,n$, 从 (5.3) 我们就得到(5.1). □

定理 5.2 (Bhatia-Elsner-Krause [22]) 用 $\|\cdot\|$ 表示矩阵的谱范数, 设 $A,B\in M_n$. 则

$$d(\sigma(A),\sigma(B))\leqslant 4(\|A\|+\|B\|)^{1-1/n}\|A-B\|^{1/n}. \tag{5.4}$$

证明 记

$$\Omega=\left\{z\in\mathbb{C}\mid \text{ 存在 } t,\ 0\leqslant t\leqslant 1,\ \text{使得 } z\in\sigma\left((1-t)A+tB\right)\right\}.$$

设 Ω' 是 Ω 的任意一个连通分支, 则由同伦理论 [90] 知: Ω' 含有 A 的特征值的个数与 Ω' 含有 B 的特征值的个数相等. 所以, 为了证定理, 我们只需证明: 如果 $a,b\in\Omega'$, 则 $|a-b|$ 不超过 (5.4) 中右边的数.

不妨设 $\|A\|\leqslant\|B\|$. 设 $\lambda_1,\cdots,\lambda_n$ 是 A 的特征值. 据引理 5.1, 存在 $\lambda\in\Omega'$ 使得

$$|\det(\lambda I-A)|=\prod_{i=1}^n|\lambda-\lambda_i|\geqslant 2^{1-2n}|b-a|^n. \tag{5.5}$$

设 $X,Y\in M_n$. 如果 $\mu\in\sigma(Y)$, 我们断言

$$|\det(\mu I-X)|\leqslant\|X-Y\|\left(\|X\|+\|Y\|\right)^{n-1}. \tag{5.6}$$

用 $e_j=(0,\cdots,0,1,0,\cdots,0)^T$ 记 $\mathbb{R}^n$ 的第 j 个标准基向量. 设 U^*YU 为 Y 的 Schur 酉上三角分解, U 为酉矩阵, 使得 $(U^*YU)(1,1)=\mu$. 于是, $(U^*YU)e_1=\mu e_1$. 通过

将 X 和 Y 分别换成 U^*XU 和 U^*YU (这不改变要证的结论 (5.6)), 我们不妨设 $Ye_1 = \mu e_1$. 现在我们也用 $\|\cdot\|$ 表示向量的欧氏范数. 注意到

$$\|(\mu I - X)e_1\| = \|(Y - X)e_1\| \leqslant \|X - Y\|,$$
$$\|(\mu I - X)e_j\| \leqslant |\mu| + \|X\| \leqslant \|X\| + \|Y\|, \quad j = 2, \cdots, n.$$

应用 Hadamard 不等式 (第三章, 习题 14), 有

$$|\det(\mu I - X)| \leqslant \prod_{i=1}^{n} \|(\mu I - X)e_i\| \leqslant \|X - Y\|(\|X\| + \|Y\|)^{n-1}.$$

这就证明了 (5.6).

在 (5.6) 中取 $X = A$, $Y = (1-t)A + tB$, $\mu = \lambda$, 这里的 $t \in [0,1]$ 使得 $\lambda \in \sigma((1-t)A + tB)$, 我们得到

$$|\det(\lambda I - A)| \leqslant \|A - B\|(\|A\| + \|B\|)^{n-1}. \tag{5.7}$$

这里我们用到了 $\|A\| \leqslant \|B\|$ 的假定. 将 (5.5) 和 (5.7) 结合起来得到

$$\begin{aligned} |a - b| &\leqslant 2^{\frac{2n-1}{n}} \|A - B\|^{1/n}(\|A\| + \|B\|)^{1-1/n} \\ &\leqslant 4\|A - B\|^{1/n}(\|A\| + \|B\|)^{1-1/n}. \end{aligned}$$

这就完成了证明. □

定理 5.2 是 "特征值是矩阵的连续函数" 这一事实的定量结果, 从 (5.4) 显然有当 $B \to A$ 时, $\sigma(B) \to \sigma(A)$. 这个定理及其证明中的想法是 Phillips 的论文 [91] 的改进. 用类似的想法可以得到关于多项式的根的扰动结果, 见 [22].

现在我们考虑正规矩阵的特征值扰动.

定理 5.3 (Hoffman-Wielandt[54]) 设正规矩阵 $A, B \in M_n$ 的特征值分别为 $\lambda_1, \cdots, \lambda_n$ 和 $\mu_1, \cdots, \mu_n$. 则存在 $\tau \in S_n$ 使得

$$\left(\sum_{i=1}^{n} \left| \lambda_i - \mu_{\tau(i)} \right|^2 \right)^{1/2} \leqslant \|A - B\|_F. \tag{5.8}$$

证明 考虑谱分解

$$A = UD_1U^*, \quad B = VD_2V^*,$$

其中 U, V 为酉矩阵, $D_1 = \operatorname{diag}(\lambda_1, \cdots, \lambda_n), D_2 = \operatorname{diag}(\mu_1, \cdots, \mu_n)$. 利用 Frobenius 范数的酉不变性我们有

$$\begin{aligned} \|A - B\|_F &= \|U(D_1U^*V - U^*VD_2)V^*\|_F \\ &= \|D_1W - WD_2\|_F, \end{aligned}$$

其中 $W = U^*V = (w_{ij})$ 为酉矩阵. 记

$$G = \left(|\lambda_i - \mu_j|^2\right), \quad Z = \left(|w_{ij}|^2\right) \in M_n,$$

并且用 $e \in \mathbb{R}^n$ 表示分量全为 1 的列向量. 则

$$\begin{aligned}\|A - B\|_F^2 &= \sum_{i,j=1}^{n} |\lambda_i - \mu_j|^2 \, |w_{ij}|^2 \\ &= e^T (G \circ Z) e.\end{aligned}$$

因为 Z 为双随机矩阵, 由 Birkhoff 定理 (定理 3.13), Z 是置换矩阵的凸组合:

$$Z = \sum_{i=1}^{n!} t_i P_i, \quad t_i \geqslant 0, \quad \sum_i t_i = 1, \quad P_i \text{ 为置换矩阵}.$$

设

$$e^T (G \circ P_k) e = \min_{1 \leqslant i \leqslant n!} e^T (G \circ P_i) e,$$

并且设 P_k 对应于 $\tau \in S_n$, 则

$$\begin{aligned}\|A - B\|_F^2 &= e^T (G \circ Z) e \\ &= \sum_{i=1}^{n!} t_i e^T (G \circ P_i) e \\ &\geqslant \sum_{i=1}^{n!} t_i e^T (G \circ P_k) e \\ &= e^T (G \circ P_k) e \\ &= \sum_{i=1}^{n} \left|\lambda_i - \mu_{\tau(i)}\right|^2.\end{aligned}$$

这就证明了 (5.8). □

下面考虑 A 正规, B 一般的情形.

引理 5.4 ([105]) 设 $A = (a_{ij}) \in M_n$ 正规. 则

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (j - i)|a_{ij}|^2 = \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} (i - j)|a_{ij}|^2. \tag{5.9}$$

证明 (Krause, 见 [105] 中的说明) 将 A 分块为

$$A = \begin{pmatrix} A_k & B_k \\ C_k & D_k \end{pmatrix}, \quad A_k \in M_k.$$

从 $AA^* = A^*A$ 得

$$A_kA_k^* + B_kB_k^* = A_k^*A_k + C_k^*C_k, \quad k = 1, 2, \cdots, n-1.$$

两边取迹得

$$\|B_k\|_F^2 = \|C_k\|_F^2, \quad k = 1, 2, \cdots, n-1.$$

将这些等式相加得

$$\sum_{k=1}^{n-1} \|B_k\|_F^2 = \sum_{k=1}^{n-1} \|C_k\|_F^2.$$

这个等式等价于 (5.9). □

对于 $X \in M_n$, 记分解 $X = L(X) + D(X) + U(X)$, 其中 $L(X)$ 为严格下三角矩阵, $D(X)$ 为对角矩阵, $U(X)$ 为严格上三角矩阵.

引理 5.5 (Sun [105]) 设 $A \in M_n$ 正规. 则

$$\|U(A)\|_F \leqslant \sqrt{n-1}\|L(A)\|_F, \quad \|L(A)\|_F \leqslant \sqrt{n-1}\|U(A)\|_F.$$

证明 设 $A = (a_{ij})$. 应用引理 5.4, 有

$$\begin{aligned}
\|U(A)\|_F^2 &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} |a_{ij}|^2 \leqslant \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (j-i)|a_{ij}|^2 \\
&= \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} (i-j)|a_{ij}|^2 \\
&\leqslant (n-1) \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} |a_{ij}|^2 = (n-1)\|L(A)\|_F^2.
\end{aligned}$$

第二个不等式可以类似地证明. □

定理 5.6 (Sun [105]) 设 $A, B \in M_n$, A 正规, A 和 B 的特征值分别为 $\lambda_1, \cdots, \lambda_n$ 和 $\mu_1, \cdots, \mu_n$. 则存在 $\tau \in S_n$ 使得

$$\left(\sum_{i=1}^{n} \left| \lambda_i - \mu_{\tau(i)} \right|^2 \right)^{1/2} \leqslant \sqrt{n}\|A - B\|_F. \tag{5.10}$$

证明 设 $U^*BU = M + T$ 为 B 的 Schur 酉上三角分解, U 为酉矩阵, $M = \mathrm{diag}(\mu_1, \cdots, \mu_n)$, T 严格上三角. 通过将 A 和 B 分别换成 U^*AU 和 U^*BU, 不妨设 $B = M + T$. 记 $E = A - B$, 则

$$A - M = E + T, \quad T = U(A) - U(E), \quad L(A) = L(E). \tag{5.11}$$

因为 A 和 M 为正规矩阵, 由 Hoffman-Wielandt 定理 (定理 5.3), 存在 $\tau \in S_n$ 使得

$$\left(\sum_{i=1}^{n} \left|\lambda_i - \mu_{\tau(i)}\right|^2\right)^{1/2} \leqslant \|A - M\|_F = \|E + T\|_F. \tag{5.12}$$

利用 (5.11) 和引理 5.5, 有

$$\begin{aligned}
\|E+T\|_F^2 &= \|E + U(A) - U(E)\|_F^2 \\
&= \|L(E) + D(E) + U(A)\|_F^2 \\
&= \|L(E)\|_F^2 + \|D(E)\|_F^2 + \|U(A)\|_F^2 \\
&\leqslant \|L(E)\|_F^2 + \|D(E)\|_F^2 + (n-1)\|L(A)\|_F^2 \\
&= \|L(E)\|_F^2 + \|D(E)\|_F^2 + (n-1)\|L(E)\|_F^2 \\
&\leqslant n\|E\|_F^2.
\end{aligned}$$

将这个不等式与 (5.12) 结合起来就得到了 (5.10). □

例子

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

表明: 不等式 (5.10) 中的常数 $\sqrt{n}$ 是最好的, 即不能再小了.

因为 $\|X\|_F \leqslant \sqrt{n}\|X\|_\infty$ 对任何 $X \in M_n$ 成立, 从定理 5.6 我们得到如下推论.

推论 5.7 (Sun [105]) 设 $A, B \in M_n$, A 正规. 则

$$d(\sigma(A), \sigma(B)) \leqslant n\|A - B\|_\infty.$$

最后我们考虑一个 Hermite 矩阵变成一个反 Hermite 矩阵的情形. 把 $G \in M_n$ 的特征值 λ_j 按模的大小排序: $|\lambda_1| \geqslant \cdots \geqslant |\lambda_n|$ (模相等的特征值的顺序可以任意指定) 并记 $\mathrm{Eig}(G) = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$. 我们有下面的结果.

定理 5.8 设 $H \in M_n$ 为 Hermite 矩阵, $S \in M_n$ 为反 Hermite 矩阵, 则

$$\|\mathrm{Eig}(H) - \mathrm{Eig}(S)\| \leqslant \sqrt{2}\|H - S\| \tag{5.13}$$

对任何酉不变范数成立.

考虑迹范数和例子

$$H = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

我们看到 (5.13) 中的常数 $\sqrt{2}$ 是最好的.

对于 $x=(x_j)\in\mathbb{C}^n$, 记 $|x|=(|x_1|,\cdots,|x_n|)$. 从这里至本节末尾记号 $i=\sqrt{-1}$. 因为 S 为反 Hermite 矩阵当且仅当 iS 为 Hermite 矩阵, 定理 5.8 等价于如下关于笛卡儿分解的结果.

定理 5.9　设 $T=A+iB\in M_n$, A,B 为 Hermite 矩阵. 设 α_j 和 β_j 分别为 A 和 B 的特征值, 满足 $|\alpha_1|\geqslant\cdots\geqslant|\alpha_n|$ 及 $|\beta_1|\geqslant\cdots\geqslant|\beta_n|$, 则

$$\|\mathrm{diag}(\alpha_1+i\beta_1,\cdots,\alpha_n+i\beta_n)\|\leqslant\sqrt{2}\,\|T\| \tag{5.14}$$

对于任何酉不变范数成立.

证明　因为一个 Hermite 矩阵的奇异值就是它的特征值的绝对值,

$$s(\mathrm{diag}(\alpha_1+i\beta_1,\cdots,\alpha_n+i\beta_n))=|s(A)+is(B)|.$$

由 Fan 支配原理 (定理 4.24), 只需要证明 (5.14) 对于所有 Fan k-范数成立, $k=1,2,\cdots,n$. 我们先证明 (5.14) 对于迹范数 $\|\cdot\|_{(n)}$ 和谱范数 $\|\cdot\|_{(1)}$ 成立. 第四章不等式 (4.30) 之 $p=1$ 的情形表明 (5.14) 对迹范数成立. 从

$$A=\frac{T+T^*}{2},\quad B=\frac{T-T^*}{2i}.$$

我们有 $|\alpha_1|=\|A\|_{(1)}\leqslant\|T\|_{(1)},|\beta_1|=\|B\|_{(1)}\leqslant\|T\|_{(1)}$, 从而$|\alpha_1+i\beta_1|\leqslant\sqrt{2}\|T\|_{(1)}$. 所以 (5.14) 对谱范数成立.

现在取定一个满足 $1\leqslant k\leqslant n$ 的 k. 由引理 4.25 存在 $X,Y\in M_n$ 满足 $T=X+Y$ 以及

$$\|T\|_{(k)}=\|X\|_{(n)}+k\|Y\|_{(1)}. \tag{5.15}$$

设 $X=C+iD$ 和 $Y=E+iF$ 是 X 和 Y 的笛卡儿分解, 即 C,D,E,F 为 Hermite 矩阵. 于是 $T=C+E+i(D+F)$. 因为笛卡儿分解是唯一的, 我们有

$$A=C+E,\quad B=D+F. \tag{5.16}$$

从 (5.14) 已经证明了的两种情形我们得

$$\sqrt{2}\|X\|_{(n)}\geqslant\Phi_{(n)}(|s(C)+is(D)|), \tag{5.17}$$

$$\sqrt{2}\|Y\|_{(1)}\geqslant\Phi_{(1)}(|s(E)+is(F)|), \tag{5.18}$$

这里 $\Phi_{(k)}$ 表示向量的 Fan k-范数. 将 (5.15), (5.17) 和 (5.18) 合起来得

$$\begin{aligned}\sqrt{2}\|T\|_{(k)}&\geqslant\Phi_{(n)}(|s(C)+is(D)|)+k\Phi_{(1)}(|s(E)+is(F)|),\\&\geqslant\sum_{j=1}^{k}|s_j(C)+is_j(D)|+k|s_1(E)+is_1(F)|.\end{aligned}$$

因此

$$\sqrt{2}\|T\|_{(k)} \geqslant \sum_{j=1}^{k} |s_j(C) + is_j(D)| + k|s_1(E) + is_1(F)|. \tag{5.19}$$

定理 4.27 关于谱范数的情形说: 对于任意 $W, Z \in M_n$,

$$\max_j |s_j(W) - s_j(Z)| \leqslant s_1(W - Z).$$

所以, 对于 $1 \leqslant j \leqslant n$,

$$s_j(C+E) \leqslant s_j(C) + s_1(E), \quad s_j(D+F) \leqslant s_j(D) + s_1(F). \tag{5.20}$$

用 (5.20) 和 (5.16) 我们来计算,

$$\begin{aligned}
&\sum_{j=1}^{k} |s_j(C) + is_j(D)| + k|s_1(E) + is_1(F)| \\
&= \sum_{j=1}^{k} \left\{ |s_j(C) + is_j(D)| + |s_1(E) + is_1(F)| \right\} \\
&\geqslant \sum_{j=1}^{k} \left| [s_j(C) + s_1(E)] + i[s_j(D) + s_1(F)] \right| \\
&\geqslant \sum_{j=1}^{k} \left| s_j(C+E) + is_j(D+F) \right| \\
&= \Phi_{(k)}\left(|s(A) + is(B)| \right).
\end{aligned}$$

所以

$$\sum_{j=1}^{k} \left| s_j(C) + is_j(D) \right| + k\left| s_1(E) + is_1(F) \right| \geqslant \Phi_{(k)}\left(|s(A) + is(B)| \right). \tag{5.21}$$

将 (5.19) 与 (5.21) 合起来得到

$$\sqrt{2}\|T\|_{(k)} \geqslant \Phi_{(k)}\left(|s(A) + is(B)| \right), \quad k = 1, 2, \cdots, n.$$

这就完成了证明. □

定理 5.9 是 Ando 和 Bhatia 的猜想 [4], 它在 [113] 中被证明. 该定理的等价形式定理 5.8 是 Bhatia 的书 ([17], p.119) 里的猜想.

§5.2 极分解

每个矩阵 $A \in M_n$ 都有极分解 $A = UP$, 其中 U 是酉矩阵, P 半正定. 半正定因子 P 是唯一的: $P = |A| = (A^*A)^{1/2}$. 若 A 可逆, 则酉因子 U 也是唯一的:

$U = A|A|^{-1}$. 本节我们讨论形如 $A = UP$ 的极分解的扰动, 另一个形式的极分解 $A = QV$, V 是酉矩阵, Q 半正定, 是类似的. 也可以通过考虑 A^* 将第二种形式转化成第一种形式. 用 $s_j(A)$ 表示 $A \in M_n$ 的第 j 大的奇异值.

定理 5.10 (Li [79]) 设 $A, B \in M_n$ 可逆, 它们的极分解是 $A = UP$, $B = VQ$, 其中 U, V 为酉矩阵, P, Q 半正定. 则

$$\|U - V\| \leqslant \frac{2}{s_n(A) + s_n(B)} \|A - B\| \tag{5.22}$$

对任何酉不变范数成立.

证明 设 $\|\cdot\|$ 是一个酉不变范数. 令

$$Y = P - U^*VQ, \quad Z = Q - V^*UP.$$

则

$$\|Y\| = \|U(P - U^*VQ)\| = \|A - B\|, \tag{5.23}$$

$$\|Z^*\| = \|Z\| = \|V(Q - V^*UP)\| = \|A - B\|, \tag{5.24}$$

$$P(I - U^*V) - (I - U^*V)(-Q) = Y + Z^*. \tag{5.25}$$

因为 P 和 Q 的特征值分别就是 A 和 B 的奇异值, P 和 $-Q$ 的谱的距离

$$\operatorname{dist}(\sigma(P), \sigma(-Q)) = s_n(A) + s_n(B).$$

应用定理 2.10 于关于 $I - U^*V$ 的 Sylvester 方程 (5.25) 我们得

$$\begin{aligned} \|I - U^*V\| &\leqslant \frac{1}{s_n(A) + s_n(B)} \|Y + Z^*\| \\ &\leqslant \frac{\|Y\| + \|Z^*\|}{s_n(A) + s_n(B)}. \end{aligned} \tag{5.26}$$

最后利用 $\|U - V\| = \|U(I - U^*V)\| = \|I - U^*V\|$, (5.23), (5.24), 我们就从 (5.26) 得到 (5.22). □

下面我们来研究极分解中半正定因子的扰动.

引理 5.11 (Kittaneh [70]) 设 $A, B \in M_n$ 为正规矩阵, 则

$$\big\||A| - |B|\big\|_F \leqslant \|A - B\|_F. \tag{5.27}$$

证明 设

$$A = U\operatorname{diag}(\lambda_1, \cdots, \lambda_n)U^*, \quad B = V\operatorname{diag}(\mu_1, \cdots, \mu_n)V^*$$

为谱分解, U,V 为酉矩阵. 则

$$|A| = U\mathrm{diag}\big(|\lambda_1|,\cdots,|\lambda_n|\big)U^*, \quad |B| = V\mathrm{diag}\big(|\mu_1|,\cdots,|\mu_n|\big)V^*.$$

记 $W=(w_{ij})=U^*V$, 则

$$\begin{aligned}
\big\||A|-|B|\big\|_F &= \big\|U\big[\mathrm{diag}\big(|\lambda_1|,\cdots,|\lambda_n|\big)W - W\mathrm{diag}\big(|\mu_1|,\cdots,|\mu_n|\big)\big]V^*\big\|_F \\
&= \big\|\mathrm{diag}\big(|\lambda_1|,\cdots,|\lambda_n|\big)W - W\mathrm{diag}\big(|\mu_1|,\cdots,|\mu_n|\big)\big\|_F \\
&= \left\{\sum_{i,j=1}^n \big||\lambda_i|-|\mu_j|\big|^2\cdot|w_{ij}|^2\right\}^{1/2} \\
&\leqslant \left\{\sum_{i,j=1}^n |\lambda_i-\mu_j|^2\cdot|w_{ij}|^2\right\}^{1/2} \\
&= \left\{\sum_{i,j=1}^n |(\lambda_i-\mu_j)w_{ij}|^2\right\}^{1/2} \\
&= \|A-B\|_F.
\end{aligned}$$

□

定理 5.12 (Kittaneh [71]) 设 $A,B\in M_n$, 则

$$\big\||A|-|B|\big\|_F^2 + \big\||A^*|-|B^*|\big\|_F^2 \leqslant 2\|A-B\|_F^2. \tag{5.28}$$

证明 将不等式 (5.27) 中的 A,B 分别换成

$$\begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix}$$

就得到 (5.28). 注意, 这是两个 Hermite 矩阵. □

从不等式 (5.28) 我们得到下面的推论.

推论 5.13 (Araki-Yamagami [6]) 设 $A,B\in M_n$, 则

$$\big\||A|-|B|\big\|_F \leqslant \sqrt{2}\|A-B\|_F. \tag{5.29}$$

下面的例子表明: (5.29) 中的常数 $\sqrt{2}$ 不能再小了. 取

$$A=\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B=\begin{pmatrix} 1 & \epsilon \\ 0 & 0 \end{pmatrix},$$

则

$$|A| = A, \quad |B| = \frac{1}{\sqrt{1+\epsilon^2}}\begin{pmatrix} 1 & \epsilon \\ \epsilon & \epsilon^2 \end{pmatrix}.$$

当 $\epsilon\to 0$ 时, $\big\||A|-|B|\big\|_F/\|A-B\|_F \to \sqrt{2}$.

§5.3 矩阵的带状部分

本节取自 Bhatia 的论文 [16]. 设 $A=(a_{ij})\in M_n$, $1\leqslant k\leqslant n-1$. A 的第 k 条上对角线是指元素序列 $a_{1,k+1},a_{2,k+2},\cdots,a_{n-k,n}$ 即满足 $j-i=k$ 的元素 a_{ij}; A 的第 k 条下对角线是指元素序列 $a_{k+1,1},a_{k+2,2},\cdots,a_{n,n-k}$ 即满足 $i-j=k$ 的元素 a_{ij}. A 的主对角线当然是指元素序列 $a_{11},a_{22},\cdots,a_{nn}$. 用 $D_0(A)$ 表示保留 A 的主对角线而将 A 的非对角元素换成 0 所得的矩阵, $D_k(A)$ 表示保留 A 的第 k 条上对角线而将 A 的其他元素换成 0 所得的矩阵, $D_{-k}(A)$ 表示保留 A 的第 k 条下对角线而将 A 的其他元素换成 0 所得的矩阵.

记 $i=\sqrt{-1},\omega=e^{2\pi i/n},U=\mathrm{diag}(1,\omega,\omega^2,\cdots,\omega^{n-1})$, 则 U 为酉矩阵并且有

$$D_0(A)=\frac{1}{n}\sum_{j=0}^{n-1}U^jAU^{*j}. \tag{5.30}$$

M_n 上的一个范数 $\|\cdot\|$ 称为是弱酉不变的, 如果 $\|VAV^*\|=\|A\|$ 对任何 $A\in M_n$ 和任何酉矩阵 $V\in M_n$ 成立. 显然, 每个酉不变范数都是弱酉不变的. 数值半径 $w(\cdot)$ 是弱酉不变的, 但不是酉不变的.

以下若无其他说明总是设 $\|\cdot\|$ 是 M_n 上的一个弱酉不变范数. 从表达式 (5.30) 得

$$\|D_0(A)\|\leqslant\frac{1}{n}\sum_{j=0}^{n-1}\|U^jAU^{*j}\|=\|A\|.$$

(5.30) 中表达主对角线的想法可以用积分推广到上、下对角线. 对于实数 θ, 记 $U_\theta=\mathrm{diag}(e^{i\theta},e^{i2\theta},\cdots,e^{in\theta})$. 则 U_θ 为酉矩阵且 $U_\theta AU_\theta^*$ 的 (r,s) 处的元素为 $e^{i(r-s)\theta}a_{rs}$. 所以我们有

$$D_k(A)=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{ik\theta}U_\theta AU_\theta^*\mathrm{d}\theta. \tag{5.31}$$

当 $k=0$ 时这给出了 $D_0(A)$ 的另一个表达式. 从 (5.31) 我们得到

$$\|D_k(A)\|\leqslant\|A\|,\quad 1-n\leqslant k\leqslant n-1.$$

利用 (5.31) 我们有

$$D_k(A)+D_{-k}(A)=\frac{1}{2\pi}\int_{-\pi}^{\pi}(2\cos k\theta)U_\theta AU_\theta^*\mathrm{d}\theta.$$

因此

$$\|D_k(A)+D_{-k}(A)\|\leqslant\frac{1}{2\pi}\int_{-\pi}^{\pi}|2\cos k\theta|\mathrm{d}\theta\|A\|.$$

求出这个积分我们得到

$$\|D_k(A)+D_{-k}(A)\|\leqslant\frac{4}{\pi}\|A\|. \tag{5.32}$$

设 $T_3(A) = D_{-1}(A) + D_0(A) + D_1(A)$, A 的三对角部分. 用同样的推理我们得到

$$\|T_3(A)\| \leqslant \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|1+2\cos\theta\right|\mathrm{d}\theta\|A\|.$$

算出这个积分得

$$\|T_3(A)\| \leqslant \left(\frac{1}{3}+\frac{2\sqrt{3}}{\pi}\right)\|A\|. \tag{5.33}$$

更一般地, 考虑带状矩阵

$$T_{2k+1}(A) = \sum_{j=-k}^{k} D_j(A), \quad 1 \leqslant k \leqslant n-1.$$

类似地, 我们得

$$\|T_{2k+1}(A)\| \leqslant L_k\|A\|. \tag{5.34}$$

其中 L_k 是 Lebesgue 常数:

$$L_k = \frac{1}{2\pi}\int_{-\pi}^{\pi}\Big|\sum_{j=-k}^{k} e^{ij\theta}\Big|\mathrm{d}\theta.$$

已知 $L_k \leqslant \log k + \log\pi + \frac{2}{\pi}(1+\frac{1}{2k})$, 并且 $L_k = \frac{4}{\pi^2}\log k + O(1)$.

设 $\Delta: M_k \to M_k, \Delta(B)$ 保留 B 的上三角部分不变而将 B 的主对角线以下的元素都换成 0. 对于 $B \in M_k$, 考虑 $A = \begin{pmatrix} 0 & B^* \\ B & 0 \end{pmatrix}$. 则有

$$T_{2k+1}(A) = \begin{pmatrix} 0 & \Delta(B)^* \\ \Delta(B) & 0 \end{pmatrix}.$$

注意到 B 的每个奇异值在 A 的奇异值中出现两次, 利用 (5.34) 并应用 Fan 支配原理, 对于任意酉不变范数 $\|\cdot\|$ 我们得到

$$\|\Delta(B)\| \leqslant L_k\|B\|.$$

现在我们来说明 (5.32) 和 (5.33) 中的常数对于迹范数是渐近最优的, 即要使这两个不等式对任意阶数的矩阵都成立, 则那两个常数不能再小了.

设 $A = J_n$, n 阶全 1 矩阵, 记 $B = D_1(A) + D_{-1}(A)$. 则 B 的特征值是 $2\cos[(j\pi)/(n+1)]$. 于是

$$\frac{\|B\|_1}{\|A\|_1} = \frac{1}{n}\sum_{j=1}^{n}\left|2\cos\frac{j\pi}{n+1}\right|. \tag{5.35}$$

设 $f(\theta) = |2\cos\theta|$. 和式

$$\frac{1}{n+1}\sum_{j=1}^{n+1}\left|2\cos\frac{j\pi}{n+1}\right|$$

是函数 $\pi^{-1}f(\theta)$ 在区间 $[0,\pi]$ 上的一个 Riemann 和 (将 $[0,\pi]$ $n+1$ 等分). 当 $n\to\infty$ 时, 这个和以及 (5.35) 中的和趋近于同一个极限

$$\frac{1}{\pi}\int_0^{\pi}|2\cos\theta|\mathrm{d}\theta=\frac{1}{2\pi}\int_{-\pi}^{\pi}|2\cos\theta|\mathrm{d}\theta=\frac{4}{\pi}.$$

这个例子也能说明 (5.33) 中的常数对于迹范数是渐近最优的.

习 题

1. $A\in M_n$ 称为正交投影矩阵如果 A 是 Hermite 矩阵且幂等: $A^*=A=A^2$. 证明: 若 $A,B\in M_n$ 为正交投影矩阵, 则 $\|A-B\|_\infty\leqslant 1$.

2. 用 $\mathrm{Im}A$ 表示 $A\in M_n$ 的像空间: $\mathrm{Im}A=\{Ax\mid x\in\mathbb{C}^n\}$. 设 $A,B\in M_n$ 为正交投影矩阵, 满足 $\|A-B\|_\infty<1$. 证明 $\dim\mathrm{Im}A=\dim\mathrm{Im}B$.

3. (Bhatia-Davis [19]) 设 $A,B\in M_n$ 为酉矩阵, 则

$$d(\sigma(A),\sigma(B))\leqslant\|A-B\|_\infty.$$

4. (G.M. Krause) 令

$$\lambda_1=1,\quad \lambda_2=\frac{4+5\sqrt{3}i}{13},\quad \lambda_3=\frac{-1+2\sqrt{3}i}{13},\quad v=\left(\sqrt{\frac{5}{8}},\frac{1}{2},\sqrt{\frac{1}{8}}\right)^T$$

再令 $A=\mathrm{diag}(\lambda_1,\lambda_2,\lambda_3)$, $U=I-2vv^T$, $B=-U^*AU$. 则 U 为酉矩阵, A,B 为正规矩阵. 验证

$$d(\sigma(A),\sigma(B))=\sqrt{\frac{28}{13}},\quad \|A-B\|_\infty=\sqrt{\frac{27}{13}}.$$

于是, 对于这一对正规矩阵 A,B,

$$d(\sigma(A),\sigma(B))>\|A-B\|_\infty.$$

5. (Friedland [46]) 给定 $A\in M_n,\lambda_i\in\mathbb{C},i=1,\cdots,n$. 证明: 存在对角矩阵 $D\in M_n$ 使得 $\sigma(A+D)=\{\lambda_1,\lambda_2,\cdots,\lambda_n\}$, 并且满足上述条件的对角矩阵 D 只有有限多个.

第六章　非负矩阵

非负矩阵自然地出现在许多地方, 例如经济学. 在本章我们将研究非负矩阵的谱性质及组合性质, 并考虑几类特殊的非负矩阵. 非负矩阵的一个迷人之处在于它们的分析性质与组合性质之间的关系 (非本原指标是个例子).

本章之内, 对于 $A=(a_{ij}), B=(b_{ij})\in M_{m,n}(\mathbb{R})$, 记号 $A\geqslant B$ 或者 $B\leqslant A$ 的意思是 $a_{ij}\geqslant b_{ij}$ 对 $i=1,\cdots,m, j=1,\cdots,n$ 都成立. 因此 $A\geqslant 0$ 表示 A 是非负矩阵. $A>0$ 表示 A 是正矩阵, 即每个元素都是正实数的矩阵. 这些记号在 $n=1$ 的情形表示实向量的相应内容. 于是, 非负向量和正向量是指它们的每个分量分别都是非负实数和都是正实数. $\mathbb{C}^n$ 中的元素我们看成列向量, 以便放在矩阵的右边相乘. 记 $|A|=(|a_{ij}|)\in M_{m,n}$. $\mathbb{R}_+^n=\left\{x\mid x=(x_1,\cdots,x_n)^T\in\mathbb{R}^n, x_i\geqslant 0,\ i=1,\cdots,n\right\}$. 对于 $z\in\mathbb{R}^n$, 我们总是用 z_i 记 z 的第 i 个分量.

§6.1 Perron-Frobenius 理论

1907 年 O. Perron 发现了正矩阵的谱的一些特别有趣的性质. G. Frobenius 在 1908—1912 年间将 Perron 的结果推广到不可约非负阵的情形, 并且得到了新的进一步的结果. Perron-Frobenius 理论有多种证明, 我们这里用 Wielandt [112] 的优美方法.

两个矩阵 X 和 Y 称为是置换相似的, 若存在一个置换矩阵 P 满足 $P^TXP=Y$. 设 $A\in M_n$. A 称为是可约的, 若 A 置换相似于一个形如

$$\begin{pmatrix} B & 0 \\ C & D \end{pmatrix}$$

的矩阵, 其中 B 和 D 为方阵. 若 A 不是可约的, 则称 A 不可约. 显然, $A \in M_n$ 可约当且仅当存在 $1, \cdots, n$ 的一个排列 $i_1, \cdots, i_n$ 和整数 $1 \leqslant s \leqslant n-1$ 使得 $A[i_1, \cdots, i_s \mid i_{s+1}, \cdots, i_n] = 0$. 按定义, 任何 1 阶矩阵都是不可约的.

引理 6.1 设 A 是 n 阶不可约非负矩阵, $n \geqslant 2$, $y \in \mathbb{R}_+^n$ 且 y 有至少一个分量为 0. 则 $(I+A)y$ 的正分量的个数大于 y 的正分量的个数.

证明 设 y 中恰好有 k 个正分量, $1 \leqslant k \leqslant n-1$. 设$P$ 是置换矩阵使得 $x = Py$ 的头 k 个分量为正而其余分量为 0. 因为 $A \geqslant 0, (I+A)y = y + Ay$ 的零分量的个数不超过 $n-k$. 假设这个个数等于 $n-k$. 则有 $y_i = 0 \Rightarrow (Ay)_i = 0$. 即 $(Py)_i = 0 \Rightarrow (PAy)_i = 0$, 于是 $(PAP^Tx)_i = 0, i = k+1, \cdots, n$. 设 $B = PAP^T = (b_{ij})$. 则当 $k+1 \leqslant i \leqslant n$ 时,

$$(Bx)_i = \sum_{j=1}^{n} b_{ij}x_j = \sum_{j=1}^{k} b_{ij}x_j = 0.$$

但当 $1 \leqslant j \leqslant k$ 时, $x_j > 0$. 所以 $b_{ij} = 0, k+1 \leqslant i \leqslant n, 1 \leqslant j \leqslant k$, 从而 A 可约, 矛盾. 因此 $(I+A)y$ 的零分量的个数 $< n-k$. 即它的正分量的个数 $> k$. □

从引理 6.1 我们立刻得到下面的结论.

引理 6.2 设 A 是 n 阶不可约非负矩阵, $y \in \mathbb{R}_+^n \backslash \{0\}$. 则 $(I+A)^{n-1}y > 0$.

引理 6.3 设 $n \geqslant 2$. 则 n 阶非负矩阵 A 不可约当且仅当 $(I+A)^{n-1} > 0$.

证明 用引理 6.2, 考虑 $(I+A)^{n-1}e_j$, 这里e_j 是 $\mathbb{R}^n$ 的第 j 个标准基向量. □

引理 6.4 一个不可约非负阵的非负特征向量是正向量.

证明 设 A 不可约非负, $Ax = \lambda x, x \geqslant 0, x \neq 0$. 显然 $\lambda \geqslant 0$. 我们有

$$(I+A)x = (1+\lambda)x.$$

因此 $(I+A)x$ 和 x 有相同个数的正分量. 由引理 6.1, $x > 0$. □

我们用 $A(i,j)$ 记矩阵 A 的 (i,j) 位置的元素.

引理 6.5 设 $n \geqslant 2$. n 阶非负矩阵 A 不可约当且仅当对于每对 $(i,j), 1 \leqslant i, j \leqslant n$, 都存在正整数 k 使得 $A^k(i,j) > 0$.

证明 假设 A 不可约. 由引理 6.3, $(I+A)^{n-1} > 0$. 记 $B = (I+A)^{n-1}A$. 显然 $B > 0$. 设

$$B = A^n + c_{n-1}A^{n-1} + \cdots + c_2A^2 + c_1A.$$

对于任何 $1 \leqslant i, j \leqslant n$,

$$B(i,j) = A^n(i,j) + c_{n-1}A^{n-1}(i,j) + \cdots + c_2A^2(i,j) + c_1A(i,j) > 0.$$

因此存在正整数 k 使得 $A^k(i,j)>0$.

假设 A 可约. 则存在置换矩阵 P 使得

$$P^TAP=\begin{pmatrix}B & 0\\ C & D\end{pmatrix},$$

其中 B 为 m 阶方阵, 我们有

$$P^TA^kP=(P^TAP)^k=\begin{pmatrix}B^k & 0\\ * & D^k\end{pmatrix}.$$

可见对于 $1\leqslant i\leqslant m, m+1\leqslant j\leqslant n$, 和任意 k, $(P^TA^kP)(i,j)=0$. □

设 A 是一个 n 阶非负矩阵. A 的 Collatz-Wielandt 函数 $f_A:\mathbb{R}_+^n\backslash\{0\}\to\mathbb{R}_+$ 定义为

$$f_A(x)=\min_{x_i\neq 0}\frac{(Ax)_i}{x_i}.$$

引理 6.6 设 $A\in M_n$ 非负不可约. 则
(i) $f_A(tx)=f_A(x), \forall\, t>0$.
(ii) $f_A(x)=\max\{\rho\mid Ax-\rho x\geqslant 0, \rho\in\mathbb{R}\}$.
(iii) 设 $x\in\mathbb{R}_+^n\backslash\{0\}$, 记 $y=(I+A)^{n-1}x$, 则 $f_A(y)\geqslant f_A(x)$.

证明 (i) 和 (ii) 显然.

(iii) 我们有 $Ax-f_A(x)x\geqslant 0$. 在这个不等式的两边左乘以 $(I+A)^{n-1}$, 并利用 A 与 $(I+A)^{n-1}$ 乘法可交换的性质, 得

$$A(I+A)^{n-1}x-f_A(x)(I+A)^{n-1}x\geqslant 0.$$

即

$$Ay-f_A(x)y\geqslant 0.$$

再用 (ii) 就得到 $f_A(y)\geqslant f_A(x)$. □

容易证明: f_A 是有界函数, 实际上, f_A 非负并且不超过 A 的最大行和.

记 $\Omega_n=\left\{x\in\mathbb{R}_+^n\mid \sum\limits_{i=1}^n x_i=1\right\}$. 引理 6.6(i) 说明, 我们只需要在 Ω_n 上研究 f_A. 显然 Ω_n 是个紧集, 但 f_A 可能在 Ω_n 的边界点上不连续, 例如, 设

$$A=\begin{pmatrix}2 & 2 & 1\\ 2 & 2 & 1\\ 0 & 2 & 1\end{pmatrix},\quad x(\varepsilon)=(1,0,\varepsilon)/(1+\varepsilon),\quad \varepsilon>0.$$

则 $f_A(x(\varepsilon))=1$. 但

$$f_A(x(0))=2\neq 1=\lim_{\varepsilon\to 0}f_A(x(\varepsilon)).$$

引理 6.7 设 $A \in M_n$ 非负不可约. 则 f_A 在 $\mathbb{R}_+^n \backslash \{0\}$ 上可以取到最大值.

证明 记

$$\Gamma = (I+A)^{n-1}\Omega_n = \left\{ y \mid y = (I+A)^{n-1}x, x \in \Omega_n \right\}.$$

则 Γ 是个紧集, 且由引理 6.2 知 Γ 中的向量都是正向量. 又显然 f_A 在 Γ 上连续, 由 Weierstrass 定理, f_A 在某个点 $y^0 \in \Gamma$ 取到 f_A 在 Γ 上的最大值. 记 $z^0 = y^0 / \sum\limits_{i=1}^n y_i^0 \in \Omega_n$. $\forall x \in \Omega_n$, 记 $y = (I+A)^{n-1}x$. 利用引理 6.6 (iii) 和 (i) 得

$$f_A(x) \leqslant f_A(y) \leqslant f_A(y^0) = f_A(z^0).$$

这就证明了 f_A 在 z^0 点取到它在 Ω_n 上的最大值.

$\forall z \in \mathbb{R}_+^n \backslash \{0\}$, 利用引理 6.6(i) 有

$$f_A(z) = f_A\left(z \Big/ \sum_{i=1}^n z_i \right) \leqslant f_A(z^0).$$

可见 f_A 在 z^0 取到它在 $\mathbb{R}_+^n \backslash \{0\}$ 上的最大值. □

设 $\lambda_1, \cdots, \lambda_n$ 为 $A \in M_n$ 的特征值, A 的谱半径 $\rho(A)$ 定义为

$$\rho(A) = \max \left\{ |\lambda_i| : i = 1, \cdots, n \right\}.$$

设 λ 为 $A \in M_n$ 的一个特征值. λ 的特征子空间的维数称为 λ 的几何重数; λ 作为 A 的特征多项式的根的重数称为 λ 的代数重数. 如果 λ 的代数重数为 1, 则称 λ 是 A 的一个单特征值.

定理 6.8 (Perron-Frobenius) 设 A 是一个 n 阶不可约非负矩阵, $n \geqslant 2$, 则下面的结论成立.

(i) $\rho(A) > 0$, 且 $\rho(A)$ 是 A 的一个单特征值.

(ii) A 有一个对应于 $\rho(A)$ 的正特征向量.

(iii) A 的每个非负特征向量都对应于特征值 $\rho(A)$.

证明 由引理 6.7, 存在 $x^0 \in \mathbb{R}_+^n \backslash \{0\}$ 满足

$$f_A(x^0) \geqslant f_A(x), \quad \forall x \in \mathbb{R}_+^n \backslash \{0\}.$$

记 $r = f_A(x^0)$. 取 $u = (1, 1, \cdots, 1)^T$. 因为 $A = (a_{ij})$ 没有零行,

$$r \geqslant f_A(u) = \min_i \sum_{j=1}^n a_{ij} > 0.$$

下面来证明 r 是 A 的一个特征值. 我们有

$$Ax^0 - rx^0 \geqslant 0. \tag{6.1}$$

假设 $Ax^0 - rx^0 \neq 0$. 由引理 6.2,

$$(I+A)^{n-1}(Ax^0 - rx^0) > 0,$$

即

$$Ay^0 - ry^0 > 0, \tag{6.2}$$

其中 $y^0 = (I+A)^{n-1}x^0 > 0$. 因为 (6.2) 是个严格的不等式, 存在正数 ε 使得

$$Ay^0 - (r+\varepsilon)y^0 \geqslant 0.$$

由引理 6.6 (ii), $f_A(y^0) \geqslant r + \varepsilon > r$, 这与 $r = f_A(x^0)$ 的最大性矛盾. 所以, (6.1) 是等式, 从而 r 是 A 的一个特征值, x^0 是对应于 r 的非负特征向量. 由引理 6.4, x^0 是正向量.

设 λ 是 A 的任何一个特征值, x 是对应的特征向量: $Ax = \lambda x$. 则

$$|\lambda||x| \leqslant A|x|.$$

于是 $|\lambda| \leqslant f_A(|x|) \leqslant r$. 这表明 $r = \rho(A)$.

现在来证明 $\rho(A)$ 是单特征值. 我们先证明 $\rho(A)$ 的几何重数是 1. 设

$$Ay = \rho(A)y, \quad 0 \neq y \in \mathbb{C}^n.$$

则

$$A|y| \geqslant \rho(A)|y|. \tag{6.3}$$

上面的证明过程表明 (6.3) 是等式, 且 $|y| > 0$. 可见, A 的对应于 $\rho(A)$ 的特征向量不含零分量. 设 y 和 z 是对应于 $\rho(A)$ 的特征向量, 则 $|y| > 0, |z| > 0, z_1y - y_1z$ 属于 $\rho(A)$ 的特征空间. 因为 $z_1y - y_1z$ 的第一个分量为 0, 它们不可能是 $\rho(A)$ 的特征向量, 所以 $z_1y - y_1z = 0$, y 和 z 线性相关. 这就证明了 $\rho(A)$ 的几何重数是 1.

为了证明 $\rho(A) = r$ 是特征多项式 $\Delta(\lambda) = \det(\lambda I - A)$ 的单根, 只需要证导数 $\Delta'(r) \neq 0$. 若 $X = (x_{ij})$ 的每个元素都是 λ 的可微函数, 则

$$\frac{d}{d\lambda}(\det X) = \sum_{i,j=1}^{n}(-1)^{i+j}\det(X(i \mid j))\frac{d}{d\lambda}x_{ij}. \tag{6.4}$$

用 $\mathrm{adj}(Z)$ 记矩阵 Z 的伴随矩阵. 我们有

$$\Delta'(\lambda) = \sum_{i=1}^{n}\det[(\lambda I - A)(i \mid i)] = \mathrm{tr}[\mathrm{adj}(\lambda I - A)].$$

记 $B(r) = \mathrm{adj}(rI - A)$, 则 $\Delta'(r) = \mathrm{tr}B(r)$,

$$(rI - A)B(r) = \det(rI - A)I = 0. \tag{6.5}$$

因为 r 的几何重数是 1, $\operatorname{rank}(rI-A)=n-1$, 于是 $B(r)\neq 0$. 设 b 是 $B(r)$ 的任意一个非零列. 由 (6.5) 得 $(rI-A)b=0$, 可见 b 是 A 的对应于 r 的特征向量. 但 A 有一个对应于 r 的正特征向量, 且因为 r 的几何重数是 1, 所以 b 是那个正特征向量的一个常数倍, 从而 $b>0$ 或 $b<0$. 这证明了 $B(r)$ 的每一列要么是零列, 要么是正向量, 要么是负向量. 考虑转置 $[B(r)]^T=\operatorname{adj}(rI-A^T), r=\rho(A)=\rho(A^T)$. 上面的结论应用于 $[B(r)]^T$ 的列说明 $B(r)$ 的每一行要么是零行, 要么是正的行向量, 要么是负的行向量. 所以

$$B(r)>0 \quad \text{或者} \quad B(r)<0.$$

从而 $\Delta'(r)=\operatorname{tr}B(r)\neq 0$. 这就证明了 $\rho(A)$ 是单特征值.

我们已经证明了 (i) 和 (ii). 现在来证 (iii). 设 $y>0$ 是 A^T 对应于 $\rho(A)$ 的特征向量, 又设 z 是 A 的任意一个非负特征向量:

$$Az=\mu z.$$

则 $\mu y^Tz=y^TAz=\rho(A)y^Tz$. 因为 $y^Tz>0$, 我们得到 $\mu=\rho(A)$. □

定理 6.8 (iii) 可以等价地叙述成: 在 A 的所有特征值中, 只有 $\rho(A)$ 有非负特征向量 (由引理 6.4, A 的非负特征向量实际上都是正向量). 定理 6.8 的证明还确立了下面的结果.

定理 6.9 设 A 是一个 n 阶不可约非负矩阵, $n\geqslant 2$, 则

$$\rho(A)=\max\left\{f_A(x)\mid x\in\mathbb{R}_+^n\backslash\{0\}\right\}.$$

若 $x\in\mathbb{R}_+^n\backslash\{0\}, f_A(x)=\rho(A)$, 则 $x>0$ 且 x 是对应于 $\rho(A)$ 的一个特征向量.

利用连续性, 我们可以得到如下关于一般非负矩阵的结果.

定理 6.10 设 A 是一个非负方阵, 则 $\rho(A)$ 是 A 的特征值, 并且 A 有一个对应于 $\rho(A)$ 的非负特征向量.

证明 设 A 的阶数为 n. 定理在 $n=1$ 的情形平凡地成立. 下面设 $n\geqslant 2$. 用 J 表示元素全为 1 的 n 阶矩阵.

对于正整数 k, 记 $A_k=A+\frac{1}{k}J$, 这是个正矩阵. 由定理 6.8, $\rho(A_k)$ 是 A_k 的特征值, 并且 A_k 在 $\Omega_n=\left\{x\in\mathbb{R}_+^n\backslash\{0\}\mid \sum\limits_{i=1}^n x_i=1\right\}$ 中有唯一一个对应于 $\rho(A_k)$ 的特征向量 x^k:

$$A_kx^k=\rho(A_k)x^k. \tag{6.6}$$

因为向量序列 $\{x^k\}$ 有界, 由 Bolzano-Weierstrass 定理, $\{x^k\}$ 有一个收敛子序列 $\{x^{k_i}\}$: $\lim\limits_{i\to\infty}x^{k_i}=x$. 显然 $x\in\Omega_n$. 从 (6.6) 得

$$A_{k_i}x^{k_i}=\rho(A_{k_i})x^{k_i}. \tag{6.7}$$

注意到 $A_{k_i} \to A, \rho(A_{k_i}) \to \rho(A)$, 在 (6.7) 中令 $i \to \infty$ 得到 $Ax = \rho(A)x$. □

设 A 是个非负方阵, 我们将 $\rho(A)$ 称为 A 的 Perron 根, 对应于 $\rho(A)$ 的一个非负特征向量称为 A 的一个 Perron 向量.

用 r_i 和 c_j 分别记 $A = (a_{ij}) \in M_n$ 的第 i 个行和与第 j 个列和:

$$r_i = \sum_{k=1}^{n} a_{ik}, \quad c_j = \sum_{k=1}^{n} a_{kj}.$$

关于 Perron 根的上下界估计有很多结果, 下面是最基本的一个.

定理 6.11 设 A 是一个 n 阶非负矩阵, 则

$$\min_{1\leqslant i\leqslant n} r_i \leqslant \rho(A) \leqslant \max_{1\leqslant i\leqslant n} r_i, \tag{6.8}$$

$$\min_{1\leqslant i\leqslant n} c_i \leqslant \rho(A) \leqslant \max_{1\leqslant i\leqslant n} c_i. \tag{6.9}$$

若 A 不可约, 则 (6.8) 中有一个不等式成为等式当且仅当 A 的所有行和相等, 而 (6.9) 中有一个不等式成为等式当且仅当 A 的所有列和相等.

证明 设 $A = (a_{ij})$, x 是 A^T 的一个 Perron 向量. 因为 $\rho(A^T) = \rho(A)$, 从 $A^T x = \rho(A)x$ 得

$$\rho(A)x_i = \sum_{k=1}^{n} a_{ki}x_k, \quad i = 1, \cdots, n.$$

将这些等式加起来得到 $\rho(A)\sum\limits_{i=1}^{n} x_i = \sum\limits_{k=1}^{n} r_k x_k$, 即

$$\rho(A) = \frac{\sum\limits_{k=1}^{n} r_k x_k}{\sum\limits_{k=1}^{n} x_k}. \tag{6.10}$$

从 (6.10) 我们得到 (6.8). 若 A 不可约, 则有 $x > 0$. 此时根据 (6.10) 知 (6.8) 中有一个是等式当且仅当 $r_1 = \cdots = r_k$. 在 (6.8) 中将 A 换成 A^T 我们就得到了关于列和的相应结论. □

推论 6.12 设 A 为 n 阶非负矩阵, $x \in \mathbb{R}^n$ 为正向量, 则

$$\min_{1\leqslant i\leqslant n} \frac{(Ax)_i}{x_i} \leqslant \rho(A) \leqslant \max_{1\leqslant i\leqslant n} \frac{(Ax)_i}{x_i}.$$

证明 设 $x = (x_1, \cdots, x_n)^T$. 记对角矩阵 $D = \operatorname{diag}(x_1, \cdots, x_n)$. 应用定理 6.11 于矩阵 $D^{-1}AD$ 并利用 $\rho(D^{-1}AD) = \rho(A)$. □

从推论 6.12 我们立刻得到下面的结果.

推论 6.13 设 A 为 n 阶非负矩阵. 若存在实数 c, d 和正向量 $x \in \mathbb{R}^n$ 满足

$$cx \leqslant Ax \leqslant dx,$$

则 $c \leqslant \rho(A) \leqslant d$.

下面定理中等式成立的条件很不显然, 而且有用.

定理 6.14 (Wielandt [112]) 设 $A, B \in M_n, n \geqslant 2, A$ 非负不可约且 $|B| \leqslant A$. 则对于 B 的任何特征值 λ 有

$$|\lambda| \leqslant \rho(A). \tag{6.11}$$

(6.11) 成为等式当且仅当

$$B = e^{i\theta} DAD^{-1}, \tag{6.12}$$

其中 $\rho(A)e^{i\theta} = \lambda, D$ 为对角酉矩阵.

证明 设

$$Bx = \lambda x. \tag{6.13}$$

其中 $0 \neq x \in \mathbb{C}^n$. 则 $|B||x| \geqslant |\lambda||x|$. 但 $A \geqslant |B|$, 所以

$$A|x| \geqslant |B||x| \geqslant |\lambda||x|. \tag{6.14}$$

由引理 6.6(ii) 和定理 6.9 得

$$|\lambda| \leqslant f_A(|x|) \leqslant \rho(A). \tag{6.15}$$

这证明了 (6.11). 若 (6.12) 成立且 $\rho(A)e^{i\theta} = \lambda$, 显然 (6.11) 成为等式. 反过来, 假设 (6.11) 成为等式, 则有实数 θ 使得 $\lambda = \rho(A)e^{i\theta}$. 那么 (6.15) 给出 $f_A(|x|) = \rho(A)$. 由定理 6.9, $|x| > 0$ 且 $|x|$ 是 A 的一个 Perron 向量. 因此, (6.14) 成为

$$A|x| = |B||x| = |\lambda||x|. \tag{6.16}$$

所以 $(A - |B|)|x| = 0$, 但 $A - |B| \geqslant 0, |x| > 0$, 我们得到

$$A = |B|. \tag{6.17}$$

令 $D = \mathrm{diag}(x_1/|x_1|, x_2/|x_2|, \cdots, x_n/|x_n|), G = e^{-i\theta}D^{-1}BD$. 则 D 为对角酉矩阵. 从 (6.13) 得

$$BD|x| = \lambda D|x| = \rho(A)e^{i\theta}D|x|.$$

从而 $G|x| = \rho(A)|x|$. 再用 (6.16) 得

$$G|x| = A|x|. \tag{6.18}$$

(6.17) 和 (6.18) 合起来给出 $G|x| = |B||x|$, 但由 G 的定义有 $|G| = |B|$, 所以

$$(|G| - G)|x| = 0. \tag{6.19}$$

对于复数 $\alpha_1, \cdots, \alpha_k$, 若 $\sum_{j=1}^{k} |\alpha_j| = \sum_{j=1}^{k} \alpha_j$ 则 $\alpha_j = |\alpha_j|, j = 1, \cdots, k$. 因此, (6.19) 给出 $G = |G| = |B| = A$, 从而 $B = e^{i\theta} DAD^{-1}$. □

根据谱半径的连续性我们从定理 6.14 得到下面的推论.

推论 6.15 设 $A, B \in M_n$, A 非负且 $|B| \leqslant A$. 则 $\rho(B) \leqslant \rho(A)$.

推论 6.16 设 $A, B \in M_n$, A 不可约, $0 \leqslant B \leqslant A, B \neq A$. 则 $\rho(B) < \rho(A)$.

证明 $n = 1$ 的情形显然. 下面设 $n \geqslant 2$. 由推论 6.15, $\rho(B) \leqslant \rho(A)$. 假如 $\rho(B) = \rho(A)$, 则由定理 6.14 知 $B = |B| = A$, 这与 $B \neq A$ 矛盾. 所以 $\rho(B) < \rho(A)$. □

设 G 是矩阵 A 的一个子矩阵. 若 $G \neq A$, 则称 G 是 A 的一个真子矩阵.

推论 6.17 (Frobenius) 设 G 是非负方阵 A 的一个主子矩阵, 则 $\rho(G) \leqslant \rho(A)$. 若 A 不可约且 G 是 A 的一个真子矩阵, 则 $\rho(G) < \rho(A)$.

证明 设 $G = A[\alpha], \alpha = (i_1, \cdots, i_k)$. $1 \leqslant i_1 < \cdots < i_k \leqslant n$, n 为 A 的阶数. 令 B 是满足 $B[\alpha] = G$ 并且在这个主子阵之外的元素全为 0 的 n 阶矩阵, 则 $0 \leqslant B \leqslant A$. 由推论 6.15 得 $\rho(G) = \rho(B) \leqslant \rho(A)$. 若 G 是 A 的真子矩阵且 A 不可约, 则 $B \neq A$. 由推论 6.16 得 $\rho(G) = \rho(B) < \rho(A)$. □

设 A 是不可约的非负方阵. A 的模等于谱半径 $\rho(A)$ 的特征值的个数称为 A 的非本原指数. 若 A 的非本原指数等于 1, 则称 A 为本原矩阵; 若 A 的非本原指数大于 1, 则称 A 为非本原矩阵.

定理 6.18 (Frobenius) 设不可约非负方阵 A 的非本原指数为 t, 设 $\lambda_1, \cdots, \lambda_t$ 是 A 的模等于谱半径 $\rho(A)$ 的特征值. 则 $\lambda_1, \cdots, \lambda_t$ 是 $\rho(A)^t$ 的互不相同的 t 次方根.

证明 记 $r = \rho(A)$, 设 $\lambda_j = re^{i\theta_j}, j = 1, \cdots, t$. 在定理 6.14 中令 $B = A$, $\lambda = \lambda_j$, 则 (6.11) 成为等式. 所以

$$A = e^{i\theta_j} D_j A D_j^{-1}, \quad j = 1, \cdots, t. \tag{6.20}$$

可见 A 与 $e^{i\theta_j} A$ 相似. 因为 r 是 A 的单特征值, 每个 $e^{i\theta_j} r = \lambda_j$ 也是 $e^{i\theta_j} A$ 的单特征值, 从而是 A 的单特征值. 从 (6.20) 得

$$A = e^{i\theta_j} D_j (e^{i\theta_k} D_k A D_k^{-1}) D_j^{-1} = e^{i(\theta_j + \theta_k)} (D_j D_k) A (D_j D_k)^{-1}.$$

所以 A 与 $e^{i(\theta_j+\theta_k)}A$ 相似, 从而 $re^{i(\theta_j+\theta_k)}$ 是 A 的一个特征值. 于是, $e^{i(\theta_j+\theta_k)}$ 是 $e^{i\theta_1},\cdots,e^{i\theta_t}$ 中的一个, 这表明, t 个不同的数 $e^{i\theta_1},\cdots,e^{i\theta_t}$ 关于乘法是封闭的, 因而它们是 t 次单位根. □

定理 6.19 (Frobenius) 非本原指数为 t 的不可约非负方阵的谱在旋转角度 $2\pi/t$ 后不变, 但在旋转比$2\pi/t$ 更小的正的角度后改变.

证明 定理 6.18 的证明显示, 若不可约非负方阵 A 的非本原指数为 t, 则 A 与 $e^{i2\pi/t}A$ 相似, 所以 A 的谱在旋转 $2\pi/t$ 后不变. 另一方面, 定理 6.18 表明, 把 A 的谱在旋转任何比 $2\pi/t$ 更小的正的角度, 那些模等于 $\rho(A)$ 的 t 个特征值构成的集合已经改变了. □

定理 6.20 (Frobenius) 设不可约非负方阵 A 的非本原指数为 t, 设

$$\lambda^n + a_1\lambda^{n_1} + a_2\lambda^{n_2} + \cdots + a_k\lambda^{n_k}$$

为 A 的特征多项式, 其中 $n > n_1 > n_2 > \cdots > n_k$, 每个 $a_j \neq 0, j = 1,\cdots,k$. 则

$$t = \gcd(n-n_1, n_1-n_2, \cdots, n_{k-1}-n_k).$$

证明 设 m 是个正整数. 则 A 和 $e^{i2\pi/m}A$ 的特征多项式分别是

$$f(\lambda) = \lambda^n + a_1\lambda^{n_1} + a_2\lambda^{n_2} + \cdots + a_k\lambda^{n_k},$$

$$g(\lambda) = \lambda^n + a_1z^{n-n_1}\lambda^{n_1} + \cdots + a_kz^{n-n_k}\lambda^{n_k},$$

其中 $z = e^{i2\pi/m}$. 于是, A 和 $e^{i2\pi/m}A$ 的谱相同 $\Leftrightarrow f(\lambda) = g(\lambda) \Leftrightarrow z^{n-n_j} = 1, j = 1,\cdots,k \Leftrightarrow m \mid n-n_j, j = 1,\cdots,k$. 由定理 6.19, t 是使 A 和 $e^{i2\pi/m}A$ 的谱相同的最大整数 m. 所以,

$$t = \gcd(n-n_1, n-n_2, \cdots, n-n_k) = \gcd(n-n_1, n_1-n_2, \cdots, n_{k-1}-n_k). \quad \square$$

推论 6.21 设 A 非负不可约. 若 A 的迹为正, 则 A 本原.

证明 A 的特征多项式形如 $\lambda^n + a_1\lambda^{n-1} + \cdots, a_1 \neq 0$. 所以 A 的非本原指数 $= \gcd(n-(n-1),\cdots) = 1$. □

一个实数的平方非负, 但一个实矩阵的平方可能含有负元素. 一个实矩阵的平方最多可能含有多少个负元呢? 下面来回答这个问题.

引理 6.22 (Eschenbach-Li [37]) 设 $n \geqslant 2$. 若 n 阶实矩阵 A 满足 $A^2 \leqslant 0$, 则 A^2 可约.

证明 假设 A^2 不可约, 令 $B = -A^2$, 则 B 非负不可约. 由定理 6.8 (Perron-Frobenius), $\rho(B)$ 是 B 的一个正的单特征值. 因此, $-\rho(B)$ 是 A^2 的一个单特征值. A

有一个特征值 λ 满足 $\lambda^2=-\rho(B)<0$. 显然, λ 是纯虚数. 但 A 是实矩阵, 其非实的特征值成共轭对出现, 所以 $\overline{\lambda}$ 也是 A 的特征值. 而 $\overline{\lambda}^2=-\rho(B)$. 这样 $-\rho(B)$ 作为 A^2 的特征值的代数重数至少为 2, 这与它是单特征值矛盾. □

引理 6.23 (DeMarr-Steger [32]) 没有实矩阵的平方只含有负元素.

证明 设 A 是 n 阶实矩阵. $n=1$ 时结论成立. 下面考虑 $n\geqslant 2$. 假设 A^2 只含有负元素, 则由引理 6.22, A^2 可约, 从而 A^2 含有至少 $n-1$ 个 0 元素, 矛盾. □

定理 6.24 (Eschenbach-Li [37]) 一个 n 阶实矩阵的平方至多含有 n^2-1 个负元素, 且这个上界可以达到.

证明 根据引理 6.23, 我们只需要对每个 n 给出一个n 阶实矩阵 A, 其平方有 n^2-1 个负元素. $n=1$ 的情形平凡; $n=2$, 取

$$A=\begin{pmatrix}0&2\\-1&-1/2\end{pmatrix};$$

用 $u_k\in\mathbb{R}^k$ 记分量全为 1 的列向量, 用 J_k 记元素全为 1 的 k 阶矩阵. $n\geqslant 3$ 时, 取

$$A=\begin{pmatrix}0&u_{n-2}^T&n\\-u_{n-2}&0&u_{n-2}\\-1&-u_{n-2}^T&-(n-1)/n\end{pmatrix},$$

则

$$A^2=\begin{pmatrix}-(2n-2)&-nu_{n-2}^T&-1\\-u_{n-2}&-2J_{n-2}&-(n+\frac{n-1}{n})u_{n-2}\\\alpha&-\frac{1}{n}u_{n-2}^T&-[(2n-2)-(\frac{n-1}{n})^2]\end{pmatrix},$$

其中 $\alpha=(n-2)+(n-1)/n>0$. □

§6.2 矩阵与图

设 V 是个有限集合, $E\subseteq V^2$, 则集合对 $D=(V,E)$ 称为一个有向图. V 中的元素称为顶点, E 中的元素称为弧. 显然一条弧是由两个顶点组成的有序对. 设 $\alpha=(a,b)$ 是一条弧, 则顶点 a 和 b 称为 α 的端点, a 称为 α 的始点, b 称为 α 的终点. 形如 (a,a) 的弧称为环. V 所含顶点的个数称为 D 的阶数. 以顶点 a 为始点的弧的条数称为 a 的出次数, 以顶点 a 为终点的弧的条数称为 a 的进次数.

有向图 D 中首尾相连的一串弧

$$(a_0,a_1),(a_1,a_2),\cdots,(a_{m-1},a_m)$$

称为一条有向途径 (directed walk). 这条有向途径也可以记作

$$a_0\to a_1\to a_2\to\cdots\to a_{m-1}\to a_m. \tag{6.21}$$

一条有向途径所含弧的条数称为这条途径的长度. 例如上面这条途径的长为 m. 顶点 a_0 和 a_m 分别称为有向途径 (6.21) 的始点和终点, $a_1,\cdots,a_{m-1}$ 称为内部顶点. 始点和终点重合的有向途径称为闭有向途径. 除了始点和终点相同外所有顶点都互不相同的闭有向途径称为一个圈 (cycle). 长度为偶 (奇) 数的圈称为偶 (奇) 圈.

所有的弧都不相同的有向途径称为有向迹 (directed trail). 所有的顶点都不相同的有向途径称为有向路 (directed path).

设 $D=(V,E)$ 和 $D'=(V',E')$ 都是有向图. 若 $V'\subseteq V$, $E'\subseteq E$, 则称 D' 是 D 的子图. 设 $W\subseteq V$. W诱导的子图 $D(W)$ 以 W 为顶点集合并且含有 E 中所有始点和终点都属于 W 的弧.

有向图 $D=(V,E)$ 的两个顶点 a 和 b 称为是强连通的, 如果有一条从 a 到 b 的有向途径, 也有一条从 b 到 a 的有向途径. 每个顶点都认为是和自己强连通的. 强连通关系显然是顶点之间的一个等价关系, 因而将顶点集 V 划分成不相交的子集:

$$V=V_1\cup V_2\cup\cdots\cup V_k.$$

这些子集诱导的子图 $D(V_1),\cdots,D(V_k)$ 称为 D 的强连通分支. 若 D 只含有一个强连通分支, 则称 D 是强连通的. 显然, D 是强连通的当且仅当对于 D 的任意两个顶点 i,j 都有从 i 到 j 的有向途径.

设 $A=(a_{ij})$ 是 n 阶矩阵, A 的有向图 $D(A)=(V,E)$ 定义如下: $V=\{1,2,\cdots,n\}$, $(i,j)\in E$ 当且仅当 $a_{ij}\neq 0$. 反过来, 给定一个 n 阶有向图 $D=(V,E)$, 我们可以定义一个 n 阶 0–1 矩阵 $A=(a_{ij})$: 若 $(i,j)\in E$, 定义 $a_{ij}=1$; 若 $(i,j)\notin E$, 定义 $a_{ij}=0$. 显然 $D(A)=D$. A 称为 D 的邻接矩阵.

设 $A=(a_{ij})$ 是 n 阶非负矩阵, k 为一个正整数, 则 $A^k(s,t)>0$ 当且仅当 $D(A)$ 中有一条从 s 到 t、长为 k 的有向途径. 这可以从表达式

$$A^k(s,t)=\sum_{1\leqslant i_1,\cdots,i_{k-1}\leqslant n} a_{si_1}a_{i_1i_2}\cdots a_{i_{k-1}t}$$

看出.

因为一个矩阵是否可约只与 0 元素有关, 引理 6.5 可重新写成如下的结论:

引理 6.25　*方阵 A 不可约当且仅当它的有向图 $D(A)$ 强连通.*

设 V 是个有限集合, E 是由某些形如 $\{a,b\}$ 的无序元素对组成的集合, 其中 $a,b\in V, a\neq b$, 则 (V,E) 称为一个简单无向图, 简称为图. V 中的元素称为顶点, E 中的元素称为边. 图的其他概念和有向图的对应概念类似地定义. 若 $\{a,b\}$ 是一条边, 则称 a 和 b 是邻接的, 并且 a 和 b 互称为邻居. 顶点 a 的度数定义为 a 的邻居的个数.

类似于有向图, 一个对角元素全为 0 的对称矩阵的零–非零结构可以用一个图来描述. 反过来, 一个图的邻接矩阵是一个对角元素全为 0 的对称 0–1 矩阵.

§6.3 本原与非本原矩阵

我们这里只给出最基本的结果. 注意, 本原与非本原矩阵首先都是不可约非负矩阵.

定理 6.26 (Frobenius) 设 A 是阶数至少为 2 的非负方阵, 则 A 本原当且仅当存在某个正整数 m 使得 A^m 为正矩阵.

证明 (Marcus-Minc [82]) 假设 $A^m > 0$, 则 A 不可约. 因为如果 A 可约, 则 A 置换相似于一个形如

$$\begin{pmatrix} B & 0 \\ C & D \end{pmatrix}$$

的矩阵, 从而 A^m 置换相似于形如

$$\begin{pmatrix} B^m & 0 \\ * & D^m \end{pmatrix}$$

的矩阵, 这样 A^m 不是正矩阵. 设 A 的非本原指数为 k, 由谱映射定理知, A^m 的非本原指数也等于 k. 但 $A^m > 0$, 由推论 6.21, A^m 本原, 即 $k = 1$, 所以 A 本原.

反过来, 设 A 本原. 通过考虑 $A/\rho(A)$, 不妨设 $\rho(A) = 1$. 我们用 $G \oplus H$ 表示分块对角矩阵 $\mathrm{diag}(G, H)$. 设

$$S^{-1}AS = 1 \oplus B$$

为 A 的 Jordan 标准形, 则我们有以下结论:

(i) $\rho(B) < 1$, 因此 $\lim\limits_{p\to\infty} B^p = 0$.
(ii) S 的第一列是 A 的对应于 Perron 根 1 的特征向量, 因而它没有零分量.
(iii) S^{-1} 的第一行是 A^T 的对应于 Perron 根 1 的特征向量, 因而它没有零分量.

现在

$$\begin{aligned} \lim_{p\to\infty} A^p &= \lim_{p\to\infty} [S(1 \oplus B)S^{-1}]^p \\ &= \lim_{p\to\infty} S(1 \oplus B^p)S^{-1} \\ &= S(1 \oplus 0)S^{-1} \end{aligned}$$

是个非负矩阵. 但

$$[S(1 \oplus 0)S^{-1}](i,j) = S(i,1)S^{-1}(1,j) \neq 0,$$

所以 $\lim\limits_{p\to\infty} A^p > 0$.

因此, 对于充分大的正整数 m 有 $A^m > 0$. □

定理 6.27 (Wielandt [112]) 设 A 是一个 n 阶本原矩阵, 则 $A^{(n-1)^2+1} > 0$.

证明 (Sedlacek [100]) 因为 A 本原, A 不可约, 它的有向图 $D(A)$ 强连通. 特别, $D(A)$ 有圈. 考虑$D(A)$ 中长度最短的圈. 不妨设 $C: 1 \to 2 \to \cdots \to m \to 1$ 是 $D(A)$ 中一个长度最短的圈, 如果不是这样, 可以对 A 作置换相似. 这对于$D(A)$ 的效果是将顶点重新编号. 于是 $A^m(i,i) > 0,\ i = 1, \cdots, m$.

对于任何顶点 i, 存在一条以 i 为始点长度不超过 $n-m$ 的有向途径到达顶点 $1, 2, \cdots, m$ 中的某一个. 若有必要, 沿着圈 C 前进, 我们总可以将这条有向途径的长度变成 $n-m$. 于是, 存在 $k \in \{1, 2, \cdots, m\}$ 并且存在一条以 i 为始点以 k 为终点的长度为 $n-m$ 的有向途径 w_1.

A^m 也本原. 现在考虑 A^m 的有向图 $D(A^m)$. 在这个图中, 任给顶点 j, 有一条从顶点 k 出发长度不超过 $n-1$ 的到达 j 的有向途径. 因为 $k \to k$ 是 $D(A^m)$ 中的环, 若有必要反复利用这个环我们可以得到一条从 k 到 j 的长度为 $n-1$ 的有向途径. 如果 $s \to t$ 是 $D(A^m)$ 中的一条弧, 则在 $D(A)$ 中有一条从 s 到 t 的长度为 m 的有向途径. 所以, 在 $D(A)$ 中存在一条从 k 到 j 的长度为 $m(n-1)$ 的有向途径 w_2.

将 w_1 和 w_2 连起来, 我们得到一条 $D(A)$ 中的从 i 到 j 的长度为

$$n - m + m(n-1) = m(n-2) + n$$

的有向途径. 因为顶点 i 和 j 是任意的, $A^{m(n-2)+n} > 0$. 我们断言 $m \leqslant n-1$. 否则 $m = n$, 此时 A 只有 $A(1,2), A(2,3), \cdots, A(n-1,n), A(n,1)$, 这 n 个非零元素, 这与 A 本原矛盾. 所以 $m(n-2)+n \leqslant (n-1)(n-2)+n = (n-1)^2+1$. 最后, 注意到对于一个非负方阵 B, 若有正整数 p 使得 $B^p > 0$, 则对于任何整数 $q \geqslant p$ 都有 $B^q > 0$. 从 $A^{m(n-2)+n} > 0$ 我们得到 $A^{(n-1)^2+1} > 0$. □

对于本原矩阵 A, 满足 $A^p > 0$ 的最小正整数 p 称为 A 的指数. 定理 6.27 表明, n 阶本原矩阵的指数不超过 $(n-1)^2+1$. 这个上界是可以达到的. 考虑

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 1 & 0 & \cdots & 0 \end{pmatrix}.$$

A 是将 n 阶基本循环矩阵 $(n,2)$ 位置的 0 换成 1 所得, 用 $e_i \in \mathbb{R}^n$ 记第 i 个分量为 1 其余分量为 0 的向量. 则

$$Ae_1 = e_n, \quad Ae_2 = e_1 + e_n, \quad Ae_3 = e_2, \quad \cdots, \quad Ae_n = e_{n-1}.$$

令 $B = A^{n-1}$. 容易验证,

$$Be_1 = e_2, \quad Be_2 = e_2 + e_3, \quad Be_3 = e_3 + e_4, \quad \cdots, \quad Be_n = e_n + e_1.$$

所以, $B^{n-1}=A^{(n-1)^2}$ 在 $(1,1)$ 位置有一个 0 元素, 而 $AB^{n-1}=A^{(n-1)^2+1}>0$.

给定正整数 n, 并非 1 和 $(n-1)^2+1$ 之间的每个整数都是某个 n 阶本原矩阵的指数. n 阶本原矩阵的所有可能的指数已经完全确定, 见 [125] 以及那里的参考文献.

邵嘉裕教授 [102] 确定了对称本原矩阵所有可能的指数, 他的定理是: 一个正整数 k 是某个 n 阶对称本原矩阵的指数当且仅当 $1\leqslant k\leqslant 2n-2$ 并且 k 不是区间 $[n,2n-2]$ 中的奇数.

定理 6.28 (Frobenius) 设不可约非负方阵 A 的非本原指数为 $k\geqslant 2$. 则 A 置换相似于如下形式的矩阵:

$$\begin{pmatrix} 0 & A_{12} & 0 & \cdots & 0 \\ 0 & 0 & A_{23} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & A_{k-1,k} \\ A_{k1} & 0 & 0 & \cdots & 0 \end{pmatrix}, \tag{6.22}$$

其中对角线上的零块都是方阵.

证明 (Wielandt [112]) 记 $r=\rho(A)$ 为 A 的 Perron 根. 根据定理 6.18 的证明, 存在一个对角酉矩阵 D 使得

$$A=e^{i2\pi/k}DAD^{-1}. \tag{6.23}$$

设 $D=\operatorname{diag}(d_1,\cdots,d_n)$. 将 D 换成 $d_1^{-1}D, DAD^{-1}$ 不变. 所以我们可以假设 $d_1=1$. 反复应用 (6.23) 我们有

$$\begin{aligned} A &= e^{i2\pi/k}D(e^{i2\pi/k}DAD^{-1})D^{-1} \\ &= e^{i2(2\pi/k)}D^2AD^{-2} \\ &= \cdots \\ &= D^kAD^{-k}, \end{aligned}$$

这里 $D^{-m}=(D^{-1})^m$. 于是, $A=D^{-k}AD^k$. 设 z 是 A 的一个 Perron 向量, 则

$$rz=Az=D^{-k}AD^kz.$$

因此 $A(D^kz)=rD^kz$. 可见 D^kz 是 A 的对应于 r 的一个特征向量. 但是 r 的特征子空间是 1 维的, 所以 D^kz 是 z 的一个常数倍. 又因为 $z>0$ 且 D^k 的第一个对角元素为 1, 我们得到 $D^k=I$. 可见 D 的对角元素都是 k 次单位根. 我们用 $\oplus_{i=1}^{s}B_j$ 表示分块对角矩阵 $\operatorname{diag}(B_1,\cdots,B_s), I_{n_j}$ 表示 n_j 阶单位矩阵. 选取置换矩阵 P 使得

$$P^TDP=\oplus_{j=1}^{s}e^{im_j2\pi/k}I_{n_j},$$

其中

$$0 = m_1 < m_2 < \cdots < m_s \leqslant k-1. \tag{6.24}$$

按 P^TDP 的分块方式对 P^TAP 分块:

$$P^TAP = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1s} \\ A_{21} & A_{22} & \cdots & A_{2s} \\ \vdots & \vdots & & \vdots \\ A_{s1} & A_{s2} & \cdots & A_{ss} \end{pmatrix},$$

其中 A_{pq} 是 $n_p \times n_q$ 的. 比较等式 $P^TAP = e^{i2\pi/k}(P^TDP)(P^TAP)(P^TD^{-1}P)$ 两边的块矩阵得

$$A_{pq} = e^{i(1+m_p-m_q)2\pi/k}A_{pq},$$

因此对于每一对 (p,q) 要么

$$A_{pq} = 0, \tag{6.25}$$

要么

$$m_q - m_p \equiv 1 \pmod k. \tag{6.26}$$

因为 A 不可约, 对于每个 p 都有某个 q 使得 $A_{pq} \neq 0$, 从而 (6.26) 成立; 对于每个 q 都有某个 p 使得 $A_{pq} \neq 0$, 从而 (6.26) 成立.

若 $p=1$ 同余式 (6.26) 成为

$$m_q \equiv 1 \pmod k. \tag{6.27}$$

考虑到条件 (6.24),(6.27) 只有 $m_2 = 1$ 这一个解. 因此对于所有的 $q \neq 2, A_{1q} = 0$. 对于 $p=2$, (6.26) 成为

$$m_q - m_2 \equiv 1 \pmod k \quad \text{即} \quad m_q \equiv 2 \pmod k.$$

这个同余式只有 $m_3 = 2$ 这一个解. 继续上述推理, 我们得到 $m_{p+1} = p$, 且 $A_{pq} = 0, \forall q \neq p+1, p = 1,2,\cdots,s-1$.

最后考虑 $p=s$ 的情形. 此时 (6.25) 和 (6.26) 表明: 对每个 q 要么 $A_{sq} = 0$ 要么 $m_q - m_s \equiv 1 \pmod k$ 即 $m_q \equiv s \pmod k$. 我们有 $A_{s1} \neq 0$, 因为其他 A_{p1} 都是零矩阵. 所以,

$$0 = m_1 \equiv s \pmod k. \tag{6.28}$$

由 $m_s = s-1$ 及 (6.24) 知 $1 \leqslant s \leqslant k$. 因此从 (6.28) 我们得到 $s=k$. 从 (6.24) 我们得到

$$m_q \not\equiv s \pmod k, \quad \forall q \neq 1.$$

因此 $A_{sq} = 0, \forall q \neq 1$. □

矩阵 (6.22) 称为非本原矩阵 A 的 Frobenius 标准形, 它十分有用.

§6.4 几类特殊的非负矩阵

我们先介绍全面非负矩阵. 全面非负矩阵的理论起源于 Gantmacher 和 Krein 1937 年先驱性的工作, 现在这类矩阵出现在许多领域.

一个矩阵 A 称为全面非负的, 如果 A 的所有子行列式都非负; A 称为全面正的, 如果 A 的所有子行列式都是正的. 我们只研究方阵的情形, 并且只讲几条最基本的性质. 更全面的论述请见 [2], [48], [69].

一个全面非负矩阵 A 称为是振荡的, 如果存在一个正整数 m 使得 A^m 是全面正的. 显然振荡矩阵是本原矩阵.

定理 6.29 一个振荡矩阵的所有特征值是互不相同的正实数.

证明 设 $\lambda_1, \cdots, \lambda_n$ 是 n 阶振荡矩阵 A 的特征值, $\lambda_1 > |\lambda_2| \geqslant \cdots \geqslant |\lambda_n|$. 根据定理 2.16(iv), 复合矩阵 $C_r(A)$ 的特征值为

$$\lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_r}, \quad 1 \leqslant i_1 < i_2 < \cdots < i_r \leqslant n.$$

因为存在一个正整数 m 使得 A^m 是全面正矩阵, $C_r(A^m) = [C_r(A)]^m$ 为正矩阵, 从而 $C_r(A)$ 为本原矩阵. 所以

$$\rho[C_r(A)] = \lambda_1\lambda_2\cdots\lambda_r > 0, \quad r = 1, 2, \cdots, n.$$

这就推出 $\lambda_1, \cdots, \lambda_n$ 全为正实数. 对于 $2 \leqslant r \leqslant n-1$, $C_r(A)$ 为本原矩阵这个事实还给出

$$\rho[C_r(A)] = \lambda_1\lambda_2\cdots\lambda_{r-1}\lambda_r > \lambda_1\lambda_2\cdots\lambda_{r-1}\lambda_{r+1}.$$

因此 $\lambda_r > \lambda_{r+1}, r = 2, \cdots, n-1$. 再由已知的 $\lambda_1 > \lambda_2$ 就得到

$$\lambda_1 > \lambda_2 > \cdots > \lambda_n. \qquad \square$$

定理 6.30 一个全面非负方阵的所有特征值都是非负实数.

证明 [2] 中的 Theorem 2.7 说: n 阶全面正矩阵的集合在 n 阶全面非负矩阵的集合中是稠密的, 即任给一个全面非负矩阵 A, 存在一列全面正矩阵 A_j 满足 $\lim_{j\to\infty} A_j = A$. 再用定理 6.29 以及特征值是矩阵的连续函数这个事实. $\square$

全面非负矩阵有一个十分有用的双对角分解, 即可以写成一些双对角矩阵的乘积, 见 [39]. 例如用双对角分解可证明下面简洁但十分不显然的判别准则 [40] (在 [2] 中也有一个证明).

定理 6.31 (Gantmacher-Krein) 一个 n 阶全面非负矩阵 $A = (a_{ij})$ 是振荡的当且仅当 A 是可逆的并且

$$a_{i,i+1} > 0, \quad a_{i+1,i} > 0, \quad i = 1, 2, \cdots, n-1.$$

下面介绍 M-矩阵. 据说 M 暗指 Minkowski. 这类矩阵是由 Ostrowski 在 1937—1938 年开始研究的. M-矩阵的很多性质由 Ostrowski, Fan, Fiedler 和 Ptak 独立地发现. M-矩阵并不是非负矩阵, 但它们与非负阵密切相关.

一个实方阵 A 称为是一个M-矩阵, 如果 A 可以写成

$$A = cI - B,$$

其中 B 是一个非负阵且 $c \geqslant \rho(B)$.

引理 6.32 *M-矩阵的主子矩阵是 M-矩阵.*

证明 用推论 6.17. □

一个实方阵 A 称为是一个 Z-矩阵, 如果 A 的所有非对角元素都小于或等于 0. 显然 M-矩阵是 Z-矩阵. 每个 Z-矩阵 A 都可以写成 $A = cI - B$ 的形式, 其中 c 是实数, B 为非负阵. 实际上设 α 是 A 的对角元素中最大者, 取 $c \geqslant \alpha$, $B = cI - A$ 即可.

定理 6.33 *设 A 是个 Z-矩阵. 则下列条件等价:*

(i) A 是 M-矩阵.

(ii) A 的每个特征值的实部都非负.

(iii) A 的所有实特征值都非负.

(iv) A 的所有主子式都非负.

证明 证明路线是 (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i) $\Rightarrow$ (iv) $\Rightarrow$ (i).

(i)$\Rightarrow$(ii). 设 λ 是 M-矩阵 $A = cI - B$ 的一个特征值, $B \geqslant 0, c \geqslant \rho(B)$. 因为 $c - \lambda$ 是 B 的一个特征值,

$$c \geqslant \rho(B) \geqslant |c - \lambda| \geqslant c - \mathrm{Re}\lambda.$$

所以 $\mathrm{Re}\lambda \geqslant 0$.

(ii)$\Rightarrow$(iii). 显然.

(iii)$\Rightarrow$(i). 设 $A = (a_{ij})_{n\times n}$, 令 $a = \max\{a_{ii} \mid 1 \leqslant i \leqslant n\}, B = aI - A$. 则 $B \geqslant 0, A = aI - B$. 因为 $a - \rho(B)$ 是 A 的一个实特征值, $a - \rho(B) \geqslant 0$, 即 $a \geqslant \rho(B)$. 所以 A 为 M-矩阵.

(i)$\Rightarrow$(iv). 设 G 是 A 的任意一个主子矩阵, 由引理 6.32, G 是 M-矩阵. 上面已证 (i)$\Rightarrow$(iii), 所以 G 的所有实特征值都非负. 又因为 G 为实阵, G 的非实特征值成共轭对出现. 而 $\det G$ 等于 G 的所有特征值之积. 于是 $\det G \geqslant 0$.

(iv)$\Rightarrow$(i). 设 $A = (a_{ij})_{n\times n}$, 令 $a = \max\{a_{ii} \mid 1 \leqslant i \leqslant n\}, B = aI - A$. 则 $B \geqslant 0, A = aI - B$. 若 A 的所有主子式都等于 0, 则 A 的所有特征值都等于 0, 因此 $\rho(A) = 0$, 而 a_{ii} 是一阶主子式, 所以每个 $a_{ii} = 0$, 从而 $a = 0$. 于是 $B = -A$. 此时 $a = 0 = \rho(B)$. 可见 A 是 M-矩阵. 下面假设 A 的所有主子式都非负, 并且至少有一

个主子式为正. 用 $E_i(A)$ 记 A 的特征值的第 i 个初等对称多项式, $E_i(A)$ 等于 A 的所有 i 阶主子式之和. 那么, $E_i(A) \geqslant 0, 1 \leqslant i \leqslant n$, 且至少有一个 i 使这个不等式是严格的. 对于任意正数 p,

$$\begin{aligned}\det[(p+a)I - B] &= \det[pI + (aI - B)] \\ &= \det[pI + A] \\ &= \sum_{k=0}^{n} p^{n-k} E_k(A) \\ &> 0,\end{aligned}$$

其中 $E_0(A) = 1$. 这表明对于任意正数 $p, p+a$ 不可能是 B 的特征值. 所以 $a \geqslant \rho(B)$, 从而 A 是 M-矩阵. □

定理 6.34 设 A 是 Z-矩阵. 则下列条件等价:

(i) A 是可逆的 M-矩阵.

(ii) A 可逆, 且 $A^{-1} \geqslant 0$.

(iii) 存在列向量 $x > 0$, 使得 $Ax > 0$.

(iv) A 的每个特征值的实部都是正数.

(v) A 的所有实特征值都是正数.

(vi) A 的所有主子式都大于零.

(vii) A 的所有顺序主子式都大于零.

证明 证明路线是 (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) $\Rightarrow$ (v) $\Rightarrow$ (i) $\Rightarrow$ (vi) $\Rightarrow$ (vii) $\Rightarrow$ (ii).

(i)$\Rightarrow$(ii). 设 $A = cI - B, B \geqslant 0, c > \rho(B)$. 因为 A 可逆且 $\rho(B)$ 是 B 的一个特征值, c 不会等于 $\rho(B)$. 于是 $\rho(B/c) < 1$, $\lim\limits_{m\to\infty} (\frac{B}{c})^m = 0$,

$$A^{-1} = c^{-1}\left(I - \frac{B}{c}\right)^{-1} = c^{-1}\sum_{m=0}^{\infty}\left(\frac{B}{c}\right)^m \geqslant 0.$$

(ii)$\Rightarrow$(iii). 用 e 记分量全为 1 的列向量, 令 $x = A^{-1}e$. 则由 $A^{-1} \geqslant 0$ 及 A^{-1} 没有零行知 $x > 0$ 且 $Ax = e > 0$.

(iii)$\Rightarrow$(iv). 每个 Z-矩阵 A 都可以写成 $A = sI - B$ 的形式, 其中 $s > 0, B \geqslant 0$. 这只需要把正数 s 取得大于 A 的所有对角元素即可. 现在有向量 $x > 0$ 使得 $Ax > 0$. 于是 $sx > Bx$. 设 y 是 B^T 的一个 Perron 向量, 则

$$sy^Tx > y^TBx = \rho(B^T)y^Tx = \rho(B)y^Tx.$$

但 $y^Tx > 0$, 所以 $s > \rho(B)$. 设 λ 是 A 的任意一个特征值, 则 $s - \lambda$ 是 B 的一个特征值, 从而

$$s > \rho(B) \geqslant |s - \lambda| \geqslant s - \mathrm{Re}\lambda.$$

这给出 $\mathrm{Re}\lambda > 0$.

(iv)⇒(v). 显然.

(v)⇒(i). 与定理 6.33 的证明中的 (iii)⇒(i) 类似.

(i)⇒(vi). 与定理 6.33 的证明中的 (i)⇒(iv) 类似.

(vi)⇒(vii). 显然.

(vii)⇒(ii). 对矩阵的阶数 n 作归纳法. $n=1$ 时, $A=(a_{11}), a_{11}>0$. 结论成立. 假设结论对 $n-1$ 阶的 Z-矩阵成立, 现在假定 n 阶 Z-矩阵 A 的顺序主子式都大于零. 记

$$A=\begin{pmatrix} A_{n-1} & x \\ y^T & a \end{pmatrix}, \quad G=\begin{pmatrix} A_{n-1}^{-1} & 0 \\ -y^T A_{n-1}^{-1} & 1 \end{pmatrix},$$

则

$$GA=\begin{pmatrix} I_{n-1} & A_{n-1}^{-1}x \\ 0 & \omega \end{pmatrix},$$

其中 $\omega=a-y^T A_{n-1}^{-1}x=\det(GA)=\det A_{n-1}^{-1}\det A>0$. 再令

$$F=\begin{pmatrix} I_{n-1} & -\omega^{-1}A_{n-1}^{-1}x \\ 0 & \omega^{-1} \end{pmatrix}.$$

简单的计算表明 $FGA=I_n$. 所以 $A^{-1}=FG$. 因为 $x\leqslant 0$, $y\leqslant 0, \omega>0$ 并且由归纳假设 $A_{n-1}^{-1}\geqslant 0$ 我们看到 $G\geqslant 0, F\geqslant 0$, 所以 $A^{-1}=FG\geqslant 0$. □

定理 6.35 设 A 是个 n 阶 Z-矩阵. 若存在正向量 $x\in\mathbb{R}^n$ 使得 $Ax\geqslant 0$, 则 A 为 M-矩阵.

证明 因为 A 是 Z-矩阵, A 可以写成 $A=cI-B$, 其中 c 为实数, B 为非负阵. 条件 $Ax\geqslant 0$ 给出 $cx\geqslant Bx$. 由推论 6.13 得 $c\geqslant\rho(B)$. 所以 A 是 M-矩阵. □

定理 6.36 n 阶不可约奇异 M-矩阵的秩等于 $n-1$.

证明 A 可以写成 $A=cI-B$, 其中 c 为实数, B 为非负阵, $c\geqslant\rho(B)$. 因为 A 奇异, c 是 B 的一个特征值, 所以 $\rho(B)\leqslant c\leqslant\rho(B)$. 于是 $c=\rho(B)$. A 不可约推出 B 不可约. 由 Perron-Frobenius 定理, $c=\rho(B)$ 是 B 的单重特征值. 因此 A 的秩是 $n-1$. □

习 题

1. 怎样的非负方阵可逆并且其逆也非负?

2. 设 A 是个非负方阵且存在一个正整数 p 使得 $A^p>0$, 则对所有正整数 $q\geqslant p, A^q>0$.

3. 设 λ 是一个复数. 证明: 存在非负方阵 A 使得 λ 是 A 的一个特征值.

4. 设 A 是个不可约非负方阵, $0\leqslant t\leqslant 1$, 则

$$\rho[tA+(1-t)A^T]\geqslant\rho(A).$$

5. (Levinger, 1970) 设 A 是个不可约非负方阵, 则函数

$$f(t) = \rho[tA + (1-t)A^T]$$

在 $[0, 1/2]$ 上递增, 在 $[1/2, 1]$ 上递减.

6. 设 A 是个非负本原方阵, 则

$$\lim_{k\to\infty} [\rho(A)^{-1}A]^k = xy^T,$$

其中 x 和 y 分别是 A 和 A^T 的 Perron 根, 满足 $x^Ty = 1$.

7. 设 A 是个非负幂零矩阵, 即存在正整数 p 使得 $A^p = 0$. 则 A 置换相似于一个上三角矩阵.

8. 设 A 是个不可约奇异 M-矩阵, 则存在正向量 x 满足 $Ax = 0$.

9. (Hopf, 见 [98]) 将 n 阶正矩阵 $A = (a_{ij})$ 的特征值按模由大到小排列为 $\rho(A) > |\lambda_2| \geqslant \cdots \geqslant |\lambda_n|$, 并记

$$\alpha = \max\{a_{ij} \mid 1 \leqslant i, j \leqslant n\}, \quad \beta = \min\{a_{ij} \mid 1 \leqslant i, j \leqslant n\}.$$

则

$$\frac{|\lambda_2|}{\rho(A)} \leqslant \frac{\alpha - \beta}{\alpha + \beta}.$$

10. ([42], [122]) 非本原指数为 k 的 n 阶不可约非负矩阵的正元素的个数可能是哪些数呢?

11. (Gasca-Pena [49]) 一个 n 阶可逆矩阵 A 是全面非负阵当且仅当对每个 $1 \leqslant k \leqslant n$,

$$\det A[1, 2, \cdots, k] > 0,$$

$$\det A[\alpha \mid 1, 2, \cdots, k] \geqslant 0, \quad \det A[1, 2, \cdots, k \mid \alpha] \geqslant 0, \quad \forall \alpha \in Q_{k,n}.$$

12. 设 A 是个 n 阶振荡矩阵, 则 A^{n-1} 是全面正矩阵.

13. (Sinkhorn [103]) 设 A 是一个方的正矩阵, 则存在对角元素为正数的两个对角矩阵 D_1 和 D_2 使得 D_1AD_2 为双随机矩阵. 关于这个结果的不同证明及推广可见 [9] 及其参考文献.

14. (Shao [101]) 设非负方阵 A 具有 (6.22) 的形式并且 A 没有零行也没有零列. 证明: A 不可约且非本原指数为 k 当且仅当乘积

$$A_{12}A_{23}\cdots A_{k-1,k}A_{k1}$$

是本原矩阵.

15. (Hu-Li-Zhan [61]) 秩为 k 的 n 阶对称 0–1 矩阵中 1 的个数可能是哪些数呢?

第七章　符号模式

符号模式的基本想法是只根据实矩阵的元素的符号 $(+,-,0)$ 来判断矩阵的性质, 例如, 一个 6 阶实矩阵 A 的符号如果是

$$\operatorname{sgn}(A)=\begin{pmatrix}0&+&0&+&0&-\\-&0&+&0&0&0\\0&-&0&+&0&0\\-&0&-&0&-&0\\0&0&0&0&0&+\\+&0&0&0&0&0\end{pmatrix},$$

则我们断定 A 的所有特征值都不是实数. 符号模式矩阵起源于 P.A. Samuelson (诺贝尔经济学奖得主) 在 1947 年出版的书《经济分析基础》[99], 他在书中指出:"人们需要只凭矩阵元素的符号来解答某些在经济学及其他领域出现的问题." 符号模式已经被人们从实矩阵推广到了复矩阵. 有两种方式来研究复矩阵的符号模式, 它们基于复数的两种表达方式: $z=a+ib, z=re^{i\theta}$. 一种是考虑复矩阵元素的实部和虚部的符号, 另一种是考虑非零元素的幅角. 本章除了最后一节我们只讲实矩阵的符号模式. 现在关于符号模式已有大量的工作, 但另一方面还有许多问题没有解决. 关于符号模式的两个基本参考文献是 [29],[51].

符号模式矩阵 (简称*符号模式*) 就是元素取自集合 $\{+,-,0\}$ 的矩阵. 对于一个实矩阵 B, 用 $\operatorname{sgn}(B)$ 表示元素是 B 的对应元素的符号的符号模式矩阵, $\operatorname{sgn}(B)$ 称为 B 的*符号模式*. 设 A 是 $m\times n$ 的符号模式矩阵, A 的*符号模式类*定义为

$$Q(A)=\{B\in M_{m,n}(\mathbb{R})\mid \operatorname{sgn}(B)=A\}.$$

而一个实矩阵 E 的符号模式类定义为 $Q(E)=Q(\mathrm{sgn}E)$. 设 P 是关于实矩阵的一个性质, 我们称一个符号模式 A 要求 P, 若 $Q(A)$ 中的每个矩阵都有性质 P; 称 A 允许 P, 若 $Q(A)$ 中存在一个矩阵具有性质 P. 这是两类主要的关于符号模式的问题.

设 $A=(a_{ij})$ 是一个 n 阶符号模式矩阵或者实矩阵. 形如

$$\gamma=a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1}$$

的非零元素的一个形式乘积 (即有顺序地放在一起) 称为长度为 k 的简单圈, 如果这些下标 $i_1,\cdots,i_k$ 互不相同. 有限个简单圈的形式乘积

$$\gamma=\gamma_1\gamma_2\cdots\gamma_m$$

称为一个复合圈, 这里 $m\geqslant 2$, 并且任何两个 γ_i 和 γ_j 的下标集不相交, $1\leqslant i<j\leqslant m$. γ 的长度定义为 $\gamma_1,\gamma_2,\cdots,\gamma_m$ 的长度之和. 简单圈和复合圈统称为圈. 长度为 k 的圈称为 k-圈. 长度为偶数 (奇数) 的圈称为偶 (奇) 圈. 符号模式矩阵的一个圈称为正 (负) 的, 如果它含有偶数 (奇数) 个 $-$ 号; 实矩阵的一个圈称为正 (负)的, 如果它含有偶数 (奇数) 个负元素. 矩阵 A 的一个圈对应着 A 的由圈的下标所指明的主子式的标准展开式中的一项, 除了可能有符号的差别. 设下标 $i_1,\cdots,i_k$ 互不相同, 作为置换的圈 $(i_1i_2\cdots i_k)$ 的符号是 $(-1)^{k-1}$. 所以若简单圈 γ_i 的长度是 l_i, 则复合圈 $\gamma=\gamma_1\cdots\gamma_m$ 的置换符号定义为 $\mathrm{sgn}(\gamma)=(-1)^{\sum\limits_{i=1}^{m}(l_i-1)}$. γ 中的元素相乘放上它的置换符号就是行列式展开式中的一项. 我们用符号 $\pi(\gamma)$ 表示将一个实矩阵的圈 γ 上的元素相乘所得到的数. A 的非零元素的一个形式乘积

$$a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_{k+1}}$$

称为是长度为 k 的一条路, 如果下标 $i_1,\cdots,i_{k+1}$ 互不相同.

n 阶符号模式矩阵 $A=(a_{ij})$ 的有向图 $D(A)$ 以 $\{1,\cdots,n\}$ 为顶点集, (s,t) 是一条弧当且仅当 $a_{st}\neq 0$. 显然 A 的简单圈对应 $D(A)$ 的圈, A 的路对应 $D(A)$ 的路.

一个有向图称为二部图, 如果它的顶点集可以分成不相交的两部分 V_1 和 V_2, 并且它的任意一条弧 (s,t) 都满足 $s\in V_1,t\in V_2$ 或者 $s\in V_2,t\in V_1$. 若符号模式矩阵 A 的有向图 $D(A)$ 是二部图, 则称 A 是二部矩阵. 显然, 一个二部矩阵置换相似于形如

$$\begin{pmatrix} 0 & A_1 \\ A_2 & 0 \end{pmatrix}$$

的矩阵, 这里两个零块都是方阵. 容易看出, 二部图只可能有偶圈, 不太明显的是对于强连通图反过来也对.

引理 7.1 强连通的有向图 D 是二部图当且仅当 D 的每个圈都是偶圈.

证明 由定义立刻看出二部图的圈都是偶圈.

反过来假设 D 的每个圈都是偶圈. 从顶点 i 到顶点 j 的距离定义为从 i 到 j 的最短路的长度. 我们定义从 i 到 i 的距离为零. 取定 D 的任意一个顶点 a. 设 D 的顶点集为 V. 定义

$$X = \{x \in V \mid \text{从 } a \text{ 到 } x \text{ 的距离为偶数}\},$$

$$Y = \{y \in V \mid \text{从 } a \text{ 到 } y \text{ 的距离为奇数}\}.$$

则 X, Y 非空 $(a \in X)$, $X \cap Y = \varnothing$ 且 $V = X \cup Y$. 设 p, q 是 X 中任意的两个顶点, 我们将证明 (p, q) 不是 D 的弧. 设

$$a \to \cdots \to p \quad \text{和} \quad a \to \cdots \to q$$

分别是从 a 到 p 和从 a 到 q 的最短的路. D 中的每条闭途径的长度都是偶数, 因为一条闭途径可以分解成有限个圈. 设 $q \to \cdots \to a$ 是 q 到 a 的一条路, 则这条路的长度是偶数, 因为否则闭途径 $a \to \cdots \to q \to \cdots \to a$ 的长度为奇数. 注意到 $a \to \cdots \to p$ 的长度为偶数, 我们断言 (p, q) 不是 D 的弧, 因为否则闭途径 $a \to \cdots \to p \to q \to \cdots \to a$ 的长度为奇数.

类似地我们可以证明若 $p, q \in Y$, 则 (p, q) 不是 D 的弧. 所以 D 是二部图. □

引理 7.1 中的强连通条件不能去掉. 例如, 设有向图 D 的顶点集 $V = \{1, 2, 3\}$, 弧的集合 $E = \{(1,2), (2,1), (2,3), (1,3)\}$, 则 D 只有一个圈 $1 \to 2 \to 1$, 是个偶圈, 但 D 不是二部图.

§7.1 符号非奇异模式

设 A 是符号模式方阵或者是个实方阵. A 称为符号非奇异的, 若 $Q(A)$ 中的每个实矩阵都是非奇异的, 即可逆的.

定理 7.2 **(Samuelson [99], Lancaster [74], Bassett, Maybee and Quirk [12])** 设 A 是符号模式方阵, $B \in Q(A)$. 则 A 符号非奇异当且仅当 B 的行列式的标准展开式中至少有一项非零, 并且所有非零项的符号都相同.

证明 定理中的条件显然是充分的. 下面假设 A 符号非奇异, 则 $\det B \neq 0$, B 的行列式展开式中至少有一项非零. 设 $B = (b_{ij})_{n\times n}$ 且 $b_{1j_1}b_{2j_2}\cdots b_{nj_n} \neq 0, j_1, \cdots, j_n$ 是 $1, \cdots, n$ 的一个置换. 设 $X_\epsilon = (x_{ij})_{n\times n}, x_{kj_k} = 1, k = 1, 2, \cdots, n, X_\epsilon$ 的其他元素都等于给定的正数 ϵ, 则 $Y_\epsilon = B \circ X_\epsilon \in Q(A)$, $Y_1 = B$, $\det Y_\epsilon$ 是 ϵ 的连续函数. 那么对于 $\forall \epsilon > 0$, $\det Y_\epsilon$ 和 $\det B = \det Y_1$ 的符号相同, 因为否则由介值定理存在某个 ϵ 使 $\det Y_\epsilon = 0$. 另一方面, 当 ϵ 充分小时, $\det Y_\epsilon$ 的符号与 $\text{sign}(j_1, j_2, \cdots, j_n) b_{1j_1}b_{2j_2}\cdots b_{nj_n}$ 的符号相同. 于是当 ϵ 充分小时,

$$\text{sign}(\det B) = \text{sign}\,(\det Y_\epsilon) = \text{sign}\left[\text{sign}(j_1, j_2, \cdots, j_n) b_{1j_1}b_{2j_2}\cdots b_{nj_n}\right].$$

这就证明了 B 的行列式展开式中每个非零项的符号都与 $\det B$ 的符号相同. □

置换一个符号模式矩阵 A 的某些行或者某些列, 以及把 A 的某一行或某一列的符号全部改变, 这些都不改变 A 是否符号非奇异. 所以, 在下面的定理中我们假设 A 的对角元素都是 $-$ 号并不失一般性.

定理 7.3 (Bassett-Maybee-Quirk [12]) 设符号模式方阵 A 的对角元素都是 $-$ 号. 则 A 符号非奇异当且仅当 A 的每个简单圈都是负的.

证明 设 $A=(a_{ij})_{n\times n}$, 假设 A 符号非奇异并设 $\gamma=a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1}$ 是 A 的一个简单圈. 任取 $B=(b_{ij})\in Q(A)$. 根据定理 7.2, B 的行列式的展开式中的每一个非零项的符号都是 $(-1)^n$, 因为全部对角元素所对应的那一项的符号是 $(-1)^n$. 设 γ 中有 q 个 $-$ 号. 设 $\{j_1,j_2,\cdots,j_{n-k}\}=\{1,2,\cdots,n\}\backslash\{i_1,i_2,\cdots,i_k\}$. 考虑 B 的行列式展开式中的一项

$$(-1)^{k-1}b_{i_1i_2}b_{i_2i_3}\cdots b_{i_ki_1}b_{j_1j_1}b_{j_2j_2}\cdots b_{j_{n-k}j_{n-k}}$$

的符号我们得到

$$(-1)^{k-1}(-1)^q(-1)^{n-k}=(-1)^n,\quad \text{即}(-1)^q=-1.$$

所以 q 为奇数, 从而 γ 是负的.

反过来, 假设 A 的每个简单圈都是负的, 即 B 的每个简单圈都是负的. B 的行列式中的每一个非零项都具有形式

$$(-1)^{\sum\limits_{i=1}^{m}(l_i-1)}\pi(\gamma_1\gamma_2\cdots\gamma_m),\tag{7.1}$$

这里 γ_i 是 B 的长度为 l_i 的简单圈, $\sum\limits_{i=1}^{m}l_i=n$. 因为每个 γ_i 是负的, (7.1) 中这一项的符号为

$$(-1)^{\sum\limits_{i=1}^{m}(l_i-1)}(-1)^m=(-1)^n.$$

这说明 B 的行列式展开式中每个非零项的符号都相同. 又按题设, B 的对角元素所对应的那一项不等于 0, 据定理 7.2, A 符号非奇异. □

下面的结果说明: 一个符号非奇异模式必定含有很多零元素.

定理 7.4 (Gibson [50], Thomassen [107]) 设 $n\geqslant 3$. 一个 n 阶符号非奇异模式矩阵有至少 $(n-1)(n-2)/2$ 个零元素.

比符号非奇异更一般的是具有固定秩的模式, 见 [73], [53].

§7.2 特征值

本节取自 Eschenbach 和 Johnson 的论文 [36]. 设 A 是符号模式方阵或者实方阵, 则存在置换矩阵 P 使得

$$PAP^T = \begin{pmatrix} A_{11} & 0 & \cdots & 0 \\ A_{21} & A_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ A_{t1} & A_{t2} & \cdots & A_{tt} \end{pmatrix},$$

其中 A_{ii} 是不可约方阵, $i = 1, \cdots, t$. 这里, 置换矩阵和符号模式矩阵相乘定义为通常的矩阵乘法, 并且对 $\forall a \in \{+, -, 0\}$, $1 \cdot a = a \cdot 1 = a$, $0 \cdot a = a \cdot 0 = 0$, $a + 0 = 0 + a = a$. $A_{11}, A_{22}, \cdots, A_{tt}$ 称为 A 的不可约分支, A 为实方阵, 则 A 的特征值 $\sigma(A) = \bigcup\limits_{i=1}^{t} \sigma(A_{ii})$. 所以, 在讨论特征值问题时, 我们只需要讨论矩阵的不可约分支, 换句话说, 我们可以假定矩阵是不可约的.

我们说一个符号模式方阵 $A = (a_{ij})$ 是一个树符号模式, 若 A 满足 1) A 是组合对称的, 即 $a_{ij} \neq 0$ 当且仅当 $a_{ji} \neq 0$; 2) $D(A)$ 强连通; 3) $D(A)$ 中简单圈的长度不超过 2. 我们可以用无向图来描述组合对称矩阵. 一个无圈的连通无向图称为一棵树. 一个 n 阶图是一棵树当且仅当它是连通的并且它恰好有 $n-1$ 条边. 因此 A 是树符号模式意味着 A 组合对称并且 A 的无向图 $G(A)$ 是一棵树.

圈的性质在后面的证明中将起关键作用. 设 n 阶不可约符号模式方阵 $A = (a_{ij})$ 有一个简单 k-圈 γ. 给定正数 ϵ, 定义一个实矩阵 $B_\gamma(\epsilon) = (b_{ij}(\epsilon)) \in Q(A)$:

$$|b_{ij}(\epsilon)| = \begin{cases} 1, & \text{若 } a_{ij} \text{ 在 } \gamma \text{ 上}, \\ \epsilon, & \text{若 } a_{ij} \text{ 不在 } \gamma \text{ 上}, a_{ij} \neq 0, \\ 0, & \text{若 } a_{ij} = 0. \end{cases} \tag{7.2}$$

定义

$$B_\gamma(0) = \lim_{\epsilon \to 0} B_\gamma(\epsilon). \tag{7.3}$$

则 $B_\gamma(0)$ 只有对应于 γ 的元素为 ± 1, 其他元素都为 0. $B_\gamma(0)$ 的 k 个非 0 特征值是 1 或 -1 的 k 次根, 因而它们是 k 个单重根. 若 γ 是负的偶圈, 则 $B_\gamma(0)$ 有 k 个不同的非实特征值. 因为特征值是矩阵元素的连续函数, 当 $\epsilon > 0$ 充分小时, $B_\gamma(\epsilon)$ 也有 k 个非实的特征值. 若 γ 是正的偶圈, 则 $B_\gamma(0)$ 有实特征值 1 和 -1. 因为实矩阵的非实特征值成共轭对出现, 当 $\epsilon > 0$ 充分小时, $B_\gamma(\epsilon)$ 也有两个靠近 1 和 -1 的单重实特征值.

定理 7.5 一个符号模式方阵 A 要求所有特征值都是非实数当且仅当 A 的每个不可约分支满足下面三个条件:

(i) 是二部矩阵.

(ii) 所有简单圈都是负的.

(iii) 是符号非奇异的.

证明 不妨设 A 不可约, A 的阶数为 n. 先假设 A 满足定理中的三个条件. 任取 $B \in Q(A)$, 用 $E_k(B)$ 记 B 的所有 k 阶主子式之和. 我们将证明当 k 为奇数时 $E_k(B) = 0$, 当 k 为偶数时 $E_k(B) \geqslant 0$, n 为偶数, $E_n(B) > 0$, 从而 B 的特征多项式为

$$f(x) = x^n + E_2(B)x^{n-2} + E_4(B)x^{n-4} + \cdots + E_n(B).$$

对任何实数 $x, f(x) > 0$. 因此 B 没有实特征值.

用 $B[\alpha]$ 记 B 的由指标集 α 指定的主子矩阵, 记 $|\alpha| = p$. 因为 B 是二部矩阵, B 只有偶圈. 从而当 p 为奇数时, $B[\alpha]$ 的每条对角线都含有 0 元素, 这样 $\det B[\alpha] = 0$, 所以 $E_p(B) = 0$. 当 p 为偶数时, 假设 $E_p(B) \neq 0$, 则存在 α, $|\alpha| = p$, 使得 $\det B[\alpha] \neq 0$. $B[\alpha]$ 中的每条由非 0 元素组成的对角线都对应于偶圈的乘积. 设 γ 是 $B[\alpha]$ 中的一个 p-圈. 若 γ 含有偶数个偶简单圈, 则 $\operatorname{sgn}(\gamma)\pi(\gamma) > 0$. 若 γ 含有奇数个偶简单圈, $\operatorname{sgn}(\gamma) = -1$, 我们也有 $\operatorname{sgn}(\gamma)\pi(\gamma) > 0$. 所以 $\det B[\alpha] > 0$, 从而 $E_p(B) > 0$.

条件 (iii) 表明 $\det B = E_n(B) \neq 0$. 根据前一段已证的结果我们断定 n 为偶数, 且 $E_n(B) > 0$.

反过来假设 A 要求所有特征值非实, 我们来说明如果 A 违反那三个条件之一, 都有 $B \in Q(A)$ 具有实特征值. 假设 A 不满足 (i), 由引理 7.1, A 有一个奇的简单圈 γ. 则 (7.3) 中的 $B_\gamma(0)$ 有一个单重特征值 1 或 -1. 当 $\epsilon > 0$ 充分小时, (7.2) 所定义的的 $B_\gamma(\epsilon)$ 也有一个靠近 1 或 -1 的实特征值. 假设 A 不满足 (ii), 则 A 有一个正的简单圈 γ. 若 γ 为奇圈, 上面已证 $B_\gamma(\epsilon)$ 有实特征值. 若 γ 为偶圈, 因为 γ 是正的, $B_\gamma(\epsilon)$ 有两个分别靠近 1 和 -1 的实特征值. 最后假设 A 不满足 (iii), 则存在 $B \in Q(A), B$ 是奇异的, 从而 B 有实特征值 0. □

对于符号模式矩阵"要求"和"允许"是对偶的: A 允许性质 $\overline{P}$ (性质 P 的否定) 当且仅当 A 不要求 P. 从定理 7.5 我们得到下面的推论.

推论 7.6 符号模式方阵 A 允许实特征值当且仅当 A 有一个不可约分支满足下面三个条件中的至少一个:

(i) 有一个奇简单圈.

(ii) 有一个正的偶简单圈.

(iii) 不是符号非奇异的.

定理 7.7 符号模式方阵 A 要求所有特征值为实数当且仅当 A 的每个不可约分支是树符号模式并且每个简单 2-圈都是正的.

证明 不妨设 A 不可约. 假设 A 满足定理中的条件, 不难证明, 对于任何

$B \in Q(A)$, 存在可逆的实对角矩阵 D 使得 $D^{-1}BD$ 为对称矩阵, 因而 B 的特征值全为实数.

反过来, 假设 A 不满足定理中的条件, 我们将证明存在 $B \in Q(A)$, 具有非实的特征值. 若 A 有一个负的简单的 2-圈 γ, 则当 $\epsilon > 0$ 充分小时, (7.2) 中的 $B_\gamma(\epsilon)$ 有两个靠近 $\sqrt{-1}$ 和 $-\sqrt{-1}$ 的非实特征值. 若 A 有一个长度为 $l \geqslant 3$ 的简单圈 γ, 则当 l 为奇数时, $B_\gamma(\epsilon)$ 有 $l-1$ 个非实特征值; 当 l 为偶数 γ 为正时, $B_\gamma(\epsilon)$ 有 $l-2$ 个非实特征值; 当 l 为偶数 γ 为负时, $B_\gamma(\epsilon)$ 有 l 个非实特征值. 最后若 A 不是组合对称的, 即存在 $a_{ij} \neq 0$ 但是 $a_{ji} = 0$, 因为 $D(A)$ 强连通, 有一条从 j 到 i 的路. 这条路与弧 (i,j) 合起来构成一个长度至少为 3 的简单圈. 上面已证明, 在这种情形, 存在 $B \in Q(A)$ 具有非实特征值. □

从定理 7.7 我们得到下面的推论.

推论 7.8 符号模式方阵 A 允许非实的特征值当且仅当 A 有一个负的简单 2-圈或者有一个长度至少为 3 的简单圈.

定理 7.9 符号模式方阵 A 要求所有的特征值为纯虚数当且仅当

(i) A 的每个对角元素都是 0, 并且

(ii) A 的每个不可约分支是树符号模式, 且每个2-圈都是负的.

证明 类似定理 7.7 的证明. □

定理 7.10 符号模式方阵 A 要求所有的特征值都不是正实数当且仅当 A 没有正的简单圈.

证明 不妨设 A 不可约, A 是 n 阶的. 假设 A 没有正的简单圈. 类似定理 7.5 的证明, 我们可以证明: 对任何 $B \in Q(A)$, 当 i 为奇数时 $E_i(B) \leqslant 0$, 当 i 为偶数时 $E_i(B) \geqslant 0$, B 的特征多项式是

$$f(x) = x^n - E_1(B)x^{n-1} + E_2(B)x^{n-2} + \cdots + (-1)^n E_n(B).$$

可见对任何正实数 x, $f(x) > 0$. 因此 B 没有正实特征值.

反过来假设 A 有正的简单圈 γ, 当 $\epsilon > 0$ 充分小时, $B_\gamma(\epsilon)$ 有一个实特征值靠近 1. □

从定理 7.10 我们得到下面的推论.

推论 7.11 符号模式方阵 A 允许一个正的实特征值当且仅当 A 有一个正的简单圈.

§7.3 符号稳定模式

一个复方阵 A 称为是稳定的 (半稳定的) 如果 A 的所有特征值的实部都是负数

(非正数). 稳定与半稳定矩阵用于刻画生态学中出现的某些线性微分方程组的解的稳定平衡状态.

一个符号模式方阵 A 称为是符号稳定的 (符号半稳定的), 如果 $Q(A)$ 中的每个矩阵都是稳定的 (半稳定的).

定理 7.12 (Quirk-Ruppert [93]) 设 $A=(a_{ij})$ 是不可约符号模式方阵. 则 A 符号半稳定当且仅当 A 满足下面三个条件:

(i) A 的每个对角元素非正.

(ii) A 的每个简单 2-圈 (如果有的话) 都是负的.

(iii) A 是树符号模式.

证明 假定 A 是符号半稳定的. 如果某个 $a_{ii}=+$, 取$B=(b_{ij})\in Q(A), b_{ii}=1$, B 的其他非 0 元素的绝对值等于正数 ϵ. 则当 $\epsilon>0$ 充分小时, B 有一个特征值 λ 充分靠近 1, 从而 $\mathrm{Re}\lambda>0$, 这与 B 半稳定矛盾. 所以 (i) 成立. 如果 A 有一个简单 2-圈 γ 是正的, 则当 $\epsilon>0$ 充分小时, (7.2) 中的 $B_\gamma(\epsilon)$ 有一个特征值充分靠近 1, 从而其实部为正, 矛盾. 所以 (ii) 成立. 如果 A 有一个长度为 $l\geqslant 3$ 的简单圈 γ, 则 (7.3) 中的 $B_\gamma(0)$ 的特征多项式为 $f(x)=x^{n-l}(x^l\pm 1)$, 这里的 "$\pm$" 号取决于 γ 是负的或正的. 所以当 $l\geqslant 3$ 时, $B_\gamma(0)$ 总有实部为正的特征值, 从而当 $\epsilon>0$ 充分小时, $B_\gamma(\epsilon)$ 有实部为正的特征值, 矛盾. 这证明了 A 的每个简单圈的长度都 $\leqslant 2$. 设 $a_{ij}\neq 0$, 即 $D(A)$ 有弧 (i,j). 因为 $D(A)$ 强连通, 有一条从 j 到 i 的路 P. 将 P 与弧 (i,j) 合起来我们将得到一个简单圈. 因为上面已证明 A 的每个简单圈的长度都 $\leqslant 2, P$ 的长度只能是 1, 即 (j,i) 是 $D(A)$ 的一条弧, 所以 $a_{ji}\neq 0$, 这证明了 A 是组合对称的, 又已知 A 是强连通的, 以及已证 A 的每个简单圈的长度都 $\leqslant 2$, 所以 A 是树符号模式.

反过来假设 A 满足那三个条件. 任取 $B\in Q(A)$, 存在实对角矩阵 E 使得 $E^{-1}BE=D+S$, 其中 D 为对角元素 $\leqslant 0$ 的对角矩阵, S 是一个反对称矩阵. 设 λ 是 B 的任意一个特征值, x 是对应的特征向量: $Bx=\lambda x$. 令 $y=E^{-1}x$, 则

$$\lambda y^*y=y^*Dy+y^*Sy.$$

因为 $y^*y>0, y^*Dy\leqslant 0, \mathrm{Re}y^*Sy=0$. 我们得到 $\mathrm{Re}\lambda\leqslant 0$. 这证明了 B 半稳定. 所以 A 是符号半稳定模式. □

论文 [64] 刻画了符号稳定模式 (也可在 [29] 中看到), 稳定模式的条件比半稳定模式的条件更复杂一些.

§7.4 逆正符号模式

设 A 是个符号模式矩阵, 记号 $A\geqslant 0$ 表示 A 的每个元素都是 $+$ 或 0, $A>0$ 表示 A 的每个元素都是 $+$; $A\leqslant 0$ 表示 A 的每个元素都是 $-$ 或 0, $A<0$ 表示 A 的

每个元素都是 $-$. 对于实矩阵 B, $B \geqslant 0$, $B > 0$, $B \leqslant 0$, $B < 0$ 分别表示按元素方式 B 为非负矩阵, 正矩阵, 非正矩阵, 负矩阵.

设 A 是个符号模式方阵, A 称为是*逆正的*, 如果存在一个实可逆矩阵 $B \in Q(A)$ 满足 $B^{-1} > 0$; A 称为是*逆非负的*, 如果存在一个实可逆矩阵 $B \in Q(A)$ 满足 $B^{-1} \geqslant 0$. 本节我们刻画逆正的符号模式, 即所有可逆的正矩阵的逆矩阵的符号模式.

两个矩阵 A, B 称为*置换等价的*, 如果存在置换矩阵 P, Q 使得 $PAQ = B$. 方阵 A 称为是*部分可分的*, 如果 A 置换等价于形如

$$\begin{pmatrix} B & 0 \\ C & D \end{pmatrix}$$

的矩阵, 其中 B 和 D 为非空 (即阶数至少为 1) 方阵. 因此, 一个 n 阶矩阵 A 部分可分当且仅当 A 有一个行数与列数之和为 n 的零子矩阵. A 称为*完全不可分的*, 如果 A 不是部分可分的. 我们将需要下面的引理, 它可以用 Frobenius-Konig 定理证明.

引理 7.13 *设矩阵 A 完全不可分, 则存在置换阵 P 使得 PA 的对角元素都不为零.*

一个 0–1 方阵 C 称为*圈矩阵*, 如果 C 的有向图 $D(C)$ 由一个圈以及可能的一些孤立的顶点组成. 注意, 按我们的定义, 图的圈都是简单圈. 因为一个不可约矩阵的有向图是强连通的, 而在一个强连通的有向图中, 每条弧都在某个圈上, 所以我们有下面的引理.

引理 7.14 *设 A 是个不可约非负方阵, 则存在圈矩阵 $C_1, \cdots, C_m$ 使得*

$$\sum_{i=1}^{m} C_i \in Q(A).$$

对于符号模式矩阵 S, $-S$ 表示将 S 中的 $+$ 元素变成 $-$, $-$ 元素变成 $+$, 0 元素不变所得矩阵; S^+ 表示将 S 中的 $-$ 元素换成 0 并保持 $+$ 元素和 0 元素不变所得矩阵. 用 e 表示分量全为 1 的列向量.

一个部分可分矩阵若可逆, 其逆矩阵必含有零元素. 所以, 逆正符号模式一定是完全不可分的.

定理 7.15 (Fiedler-Grone [44]) *设 S 是一个 n 阶完全不可分的符号模式矩阵, 则下列条件等价:*

(i) *S 是逆正的.*

(ii) *S 不置换等价于形如*

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

的矩阵, 其中 $S_{12} \geqslant 0, S_{21} \leqslant 0$ 并且这两个矩阵至少有一个是非空的.

(iii) 矩阵

$$\begin{pmatrix} 0 & S \\ -S^T & 0 \end{pmatrix}^+$$

不可约.

(iv) 存在实矩阵 $A \in Q(S)$ 满足 $Ae = A^Te = 0$.

(v) 存在实矩阵 $A \in Q(S)$ 满足 $Ae = A^Te = 0$ 并且 $\operatorname{rank}A = n-1$.

证明 (i)⇒(ii). 假设 S 满足 (i) 但不满足 (ii). 存在 $A \in Q(S)$ 及置换阵 P, W 使得

$$P^TAW^T = \begin{array}{cc} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} & \begin{matrix} r \\ s \end{matrix}, \\ \begin{matrix} t & \;\; k \end{matrix} & \end{array}$$

其中

$$A_{12} \geqslant 0, \quad A_{21} \leqslant 0, \quad A^{-1} > 0. \tag{7.4}$$

作分块

$$WA^{-1}P = \begin{array}{cc} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} & \begin{matrix} t \\ k \end{matrix}. \\ \begin{matrix} r & \;\; s \end{matrix} & \end{array}$$

容易验证: P^TAW^T 的分块行和分块列都不能为空, 即确实要出现.

用 B_{21} 左乘

$$A_{11}B_{12} + A_{12}B_{22} = 0,$$

用 B_{12} 右乘

$$B_{21}A_{11} + B_{22}A_{21} = 0,$$

然后两式相减得

$$B_{21}A_{12}B_{22} - B_{22}A_{21}B_{12} = 0. \tag{7.5}$$

将 (7.4) 和 (7.5) 合起来我们得到 $A_{12} = 0$ 且 $A_{21} = 0$. 但这与 A 是完全不可分的矛盾.

(ii)⇒(iii). 假设矩阵

$$Z = \begin{pmatrix} 0 & S \\ -S^T & 0 \end{pmatrix}^+$$

可约. 则 Z 有一个子矩阵 $Z[i_1, \cdots, i_p \mid j_1, \cdots, j_q] = 0$, $1 \leqslant p \leqslant 2n-1, \{i_1, \cdots, i_p\} \cap \{j_1, \cdots, j_q\} = \varnothing, \{i_1, \cdots, i_p\} \cup \{j_1, \cdots, j_q\} = \{1, 2, \cdots, 2n\}$. 若有必要置换 S 的行并置换 S 的列, 我们可以假设行 $i_1, \cdots, i_p$ 是 Z 的头 n 行中的最后 k 行和 Z 的最后 l 行, $k + l = p$. 作分块

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix},$$

其中 S_{22} 是 $k \times l$ 的. 此时

$$Z = \begin{pmatrix} 0 & 0 & S_{11} & S_{12} \\ 0 & 0 & S_{21} & S_{22} \\ -S_{11}^T & -S_{21}^T & 0 & 0 \\ -S_{12}^T & -S_{22}^T & 0 & 0 \end{pmatrix}^+$$

于是 $S_{21}^+ = 0, (-S_{12}^T)^+ = 0$, 即 $S_{21} \leqslant 0, S_{12} \geqslant 0. S_{12}$ 和 S_{21} 不可能两个都为空, 否则 S_{22} 是 0×0 的或 $n \times n$ 的, 从而 $p = k + l = 0$ 或 $2n$, 这与 $1 \leqslant p \leqslant 2n - 1$ 矛盾. 这证明了条件 (ii) 不成立.

(iii)⇒(iv). 因为矩阵

$$Z = \begin{pmatrix} 0 & S \\ -S^T & 0 \end{pmatrix}^+$$

不可约, 由引理 7.14, 存在圈矩阵 $C_1, \cdots, C_m$ 使得

$$\sum_{i=1}^m C_i \in Q(Z).$$

将 C_i 作与 Z 一致的分块:

$$C_i = \begin{pmatrix} 0 & C_{i1} \\ C_{i2} & 0 \end{pmatrix}, \quad i = 1, \cdots, m.$$

由圈矩阵的定义, C_i 的第 r 行有一个 1 当且仅当 C_i 的第 r 列也恰好有一个 1. 因此,

$$\left(C_{i1} - C_{i2}^T\right) e = 0, \quad \left(C_{i1}^T - C_{i2}\right) e = 0.$$

令 $A = \sum\limits_{i=1}^m C_{i1} - \sum\limits_{i=1}^m C_{i2}^T$, 则 $Ae = A^T e = 0$. 因为 $\sum\limits_{i=1}^m C_i \in Q(Z)$ 并且 Z 的任何处于对称位置 (s, t) 和 (t, s) 的一对元素中至少有一个为 $0, C_i$ 也是这样, $i = 1, \cdots, m$. 再次用 $\sum\limits_{i=1}^m C_i \in Q(Z)$ 我们得到 $A \in Q(S)$. 所以, A 满足 (iv).

(iv)⇒(v). 设 $A \in Q(S)$, $Ae = A^T e = 0$. 由引理 7.13, 不失一般性我们可以设 S 的从而 A 的对角元素都不为零. 设 B 是将 A 的对角元素都变成 0 而其他元素保持不变所得的矩阵, $|B|$ 是将 B 的元素取绝对值所得的矩阵. 则 $|B|$ 不可约. 由引理 7.14, 存在圈矩阵 $C_1, \cdots, C_m$ 使得

$$\sum_{i=1}^m C_i \in Q(|B|).$$

注意, 因为 $|B|$ 的对角元素都是零, 每个 C_i 的对角元素也都是零. 设 n 阶圈矩阵 C 的有向图 $D(C)$ 的圈是 $i_1 \to i_2 \to \cdots \to i_k \to i_1$, 定义 n 阶对角矩阵 $\Lambda(C) = (d_{ij})$

为 $d_{ii}=1$, 若 $i=i_t, t=1,\cdots,k$, 在所有其他情形 $d_{ij}=0$. 记

$$M=\sum_{i=1}^{m}\Lambda(C_i)-\sum_{i=1}^{m}C_i.$$

则 M 的非对角元素非正且

$$Me=M^Te=0.$$

由第六章定理 6.35, M 是个 M-矩阵. 存在正数 ξ, 使得对 $\forall\varepsilon\in[0,\xi]$

$$A+\epsilon M\in Q(S).$$

我们有

$$(A+\epsilon M)e=(A+\epsilon M)^Te=0,\quad \forall\epsilon.$$

因此 $\operatorname{rank}(A+\epsilon M)\leqslant n-1$. 因为 M 是个不可约的奇异的 n 阶 M-矩阵, 由第六章定理 6.36, $\operatorname{rank}M=n-1$. 这样 M 有一个 $n-1$ 阶的非奇异子矩阵 M_1. 设 A_1 是 A 的对应于 M_1 所处位置的子矩阵. 因为至多有 $n-1$ 个 ϵ 的值使 $\det(A_1+\epsilon M_1)=0$, 存在 $\epsilon_0\in[0,\xi]$ 使 $\det(A_1+\epsilon_0 M_1)\neq 0$. 从而 $\operatorname{rank}(A+\epsilon_0 M)=n-1$.

(v)⇒(i). 设 $A\in Q(S), Ae=A^Te=0$, $\operatorname{rank}A=n-1$. 由 $\operatorname{rank}A=n-1$ 及 $Ae=0$ 知 A 的伴随矩阵 $\operatorname{adj}A$ 的每一列都是 e 的一个常数倍, 因为 $A(\operatorname{adj}A)=0$. 同样地, 由 $\operatorname{rank}A=n-1$ 及 $A^Te=0$ 知 $\operatorname{adj}A$ 的每一行都是 e^T 的一个常数倍. 所以存在一个实数 α 使

$$\operatorname{adj}A=\alpha J\neq 0,$$

J 是每个元素都等于 1 的矩阵. 若有必要置换 A 的行且置换 A 的列, 我们可以假设 A 的 $(1,1)$ 位置的元素不为零. 用 E_{11} 记 $(1,1)$ 位置的元素为 1 其他位置的元素为 0 的 n 阶方阵. 则

$$\det(A+\epsilon E_{11})=\epsilon\alpha.$$

所以对 $\epsilon\neq 0$,

$$(A+\epsilon E_{11})^{-1}=\frac{1}{\epsilon\alpha}\operatorname{adj}(A+\epsilon E_{11}).$$

对于充分小的正数 ϵ, 我们有 $(A+\epsilon E_{11})^{-1}>0$ 且 $A+\epsilon E_{11}\in Q(A)=Q(S)$. 这证明了 S 是逆正的. □

上面证明中的 (iv)⇒(v) 属于 [14]. 定理 7.15 是 [66] 中的主要结果的推广.

逆非负符号模式在 [65] 中被刻画了. 对于一个完全不可分符号模式方阵 S, S 逆非负当且仅当 S 逆正. 部分可分的情形要稍微复杂一些.

关于符号模式矩阵有许多未解决的问题, 例如, 怎样的模式要求或者允许可对角化?

§7.5 Jordan 标准形的组合刻画

本节取自论文 [27]. 线性代数的中心结果之一是 Jordan 标准形定理. 该定理说: 每个复方阵 A 都相似于一个 Jordan 矩阵

$$J(A) = \mathrm{diag}(J_{n_1}(\lambda_1), \cdots, J_{n_k}(\lambda_k)),$$

其中

$$J_{n_i}(\lambda_i) = \begin{pmatrix} \lambda_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda_i & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_i \end{pmatrix}$$

是一个 n_i 阶的 Jordan 块, $i = 1, \cdots, k$. 这个 Jordan 标准形 $J(A)$ 在不计较对角 Jordan 块的顺序时是唯一的. 我们将用 $J(A)$ 表示 A 的任何一个 Jordan 标准形. Jordan 标准形在相似变换下不变; 实际上, 两个矩阵相似当且仅当它们有一个共同的 Jordan 标准形. 用 $\mathcal{S}(A)$ 记与 A 相似的所有复矩阵的集合.

我们将证明: 对于任意复矩阵 A, 在 $\mathcal{S}(A)$ 中, $J(A)$ 具有最多个数的非对角零元素, 并且我们将确定 $\mathcal{S}(A)$ 中所有取到这个最多个数的矩阵.

在本节我们总是用 $\phi(A)$ 表示矩阵 A 的非对角位置的非零元素的个数. 下面这个关于纯粹组合性质的引理是我们分析问题的基础.

引理 7.16 ([27]) *设 n, k 为正整数,$1 \leqslant k \leqslant n$. 如果 n 阶矩阵 A 满足 $\phi(A) \leqslant n-k$, 则存在置换矩阵 P 使得*

$$P^T AP = \mathrm{diag}(A_1, A_2, \cdots, A_k),$$

其中 A_j 是方阵, $j = 1, \cdots, k$.

证明 设 $A = (a_{ij})$. 我们用 A 来定义一个图 G, 它的顶点集是 $V = \{1, 2, \cdots, n\}$, 在顶点 i 和 j 之间有一条边当且仅当 $i \neq j$ 并且 $a_{ij} \neq 0$ 或者 $a_{ji} \neq 0$. 设这个图 G 有 p 条边, 则 $p \leqslant \phi(A) \leqslant n-k$. 我们按某种顺序将 G 的边命名为 $e_1, e_2, \cdots, e_p$. 用 G_i 表示以 V 为顶点集以 $\{e_1, e_2, \cdots, e_i\}$ 为边的图, $i = 0, 1, \cdots, p$. 则 G_0 没有边, $G_p = G$, G_i 是从 G_{i-1} 添加边 e_i 而得到的, $i = 1, 2, \cdots, p$. 于是 G_i 的连通分支比 G_{i-1} 的连通分支至多少一个, 因为一条边至多连接两个连通分支. 因为 G_0 有 n 个 (平凡的) 连通分支, G_p 至少有 $n-p \geqslant n-(n-k) = k$ 个连通分支 $C_1, C_2, \cdots, C_{n-p}$. 设 C'_k 是连通分支 C_j, $j \geqslant k$ 的并集. 则 A 的对应于 $C_1, \cdots, C_{k-1}, C'_k$ 的主子矩阵 $A_1, \cdots, A_{k-1}, A_k$ 就满足引理的结论. □

我们将需要下面的结果.

引理 7.17 ([27]) 设

$$A=\begin{pmatrix} a & x^T \\ 0 & B \end{pmatrix}$$

是一个 n 阶复矩阵, 其中 B 的阶数是 $n-1$. 如果 $J(A)$ 只有一个 Jordan 块, 则 $J(B)$ 只有一个 Jordan 块.

证明 $J(A)$ 的唯一 Jordan 块是 $J_n(a)$. 假设引理的结论不成立, 那么 $J(B)$ 有至少两个 Jordan 块. 存在非奇异矩阵 W 使得 $W^{-1}BW=\operatorname{diag}(B_1,B_2)$, 其中 B_1 和 B_2 是方阵, 阶数分别是 r, s, $1\leqslant r,s\leqslant n-2$, $r+s=n-1$. 记 $E=\operatorname{diag}(1,W)$, 我们得

$$E^{-1}AE=\begin{pmatrix} a & y_1^T & y_2^T \\ 0 & B_1 & 0 \\ 0 & 0 & B_2 \end{pmatrix}=H, \tag{7.6}$$

其中 $(y_1^T,y_2^T)=x^TW$ 且 $y_1\in\mathbb{C}^r$, $y_2\in\mathbb{C}^s$. 因为 H 相似于 A, $J(H)=J_n(a)$. 用 I 表示单位矩阵, 其阶数可从上下文看出. 从 (7.6) 我们得

$$(H-aI)^{n-1}=\begin{pmatrix} 0 & y_1^T(B_1-aI)^{n-2} & y_2^T(B_2-aI)^{n-2} \\ 0 & (B_1-aI)^{n-1} & 0 \\ 0 & 0 & (B_2-aI)^{n-1} \end{pmatrix}. \tag{7.7}$$

这是一个矛盾, 因为 $(J_n(a)-aI)^{n-1}\neq 0$ 以及 H 相似于 $J_n(a)$ 蕴涵 $(H-aI)^{n-1}\neq 0$. 而另一方面, (7.7) 右边的矩阵是零矩阵. 事实上, 因为 B_1 和 B_2 的特征值都等于 a, 且 B_1 和 B_2 的阶数至多为 $n-2$,

$$(B_1-aI)^{n-2}=0,\quad (B_2-aI)^{n-2}=0.$$

所以 $J(B)$ 只有一个 Jordan 块. □

一个复方阵称为单项矩阵, 如果它的每行和每列有且只有一个非零元素. 用 Γ_n 表示 n 阶单项矩阵的集合. 那么, $M\in\Gamma_n$ 当且仅当存在置换矩阵 P 和非奇异对角矩阵 D 满足 $M=PD$, 当且仅当存在置换矩阵 Q 和非奇异对角矩阵 E 满足 $M=EQ$. 显然, Γ_n 关于乘法作成一个群. 我们现在来证明在 $\mathcal{S}(A)$ 中, $J(A)$ 具有最大的非对角稀疏性.

定理 7.18 (Brualdi-Pei-Zhan [27]) 设 A 是个复方阵, $B\in\mathcal{S}(A)$. 则

$$\phi(B)\geqslant\phi(J(A)). \tag{7.8}$$

(7.8) 中的等号成立当且仅当存在一个单项矩阵 M 使得

$$M^{-1}BM=J(A). \tag{7.9}$$

证明 设 A 和 B 的阶数为 n 并且 $J(A)$ 有 k 个 Jordan 块, 因而 $\phi(J(A)) = n-k$. 我们用 Π_n 记 n 阶置换矩阵的集合. 假设 $\phi(B) < n-k$, 即 $\phi(B) \leqslant n-(k+1)$. 注意 $2 \leqslant k+1 \leqslant n$. 由引理 7.16, 存在一个 $P \in \Pi_n$ 使得 $P^TBP = \mathrm{diag}(B_1, \cdots, B_{k+1})$, 其中每个 B_j 都是方阵. 这意味着 $J(B)$ 至少有 $k+1$ 个 Jordan 块. 这是一个矛盾, 因为 $B \in \mathcal{S}(A)$ 表明 $J(B)$ 和 $J(A)$ 有相同的 Jordan 块, 特别, 它们的 Jordan 块的个数都是 k. 所以, 假设错误, 我们有 $\phi(B) \geqslant n-k$.

下面我们考虑等号成立的情形:

$$\phi(B) = \phi(J(A)). \tag{7.10}$$

条件 (7.9) 显然蕴涵等式 (7.10). 反过来, 假设 B 满足 (7.10). 我们将证明存在 M 满足 (7.9). 我们首先证明在 $J(A)$ 只有一个 Jordan 块的情形结论成立, 此时 $J(A) = J_n(a)$, $\phi(J(A)) = n-1$. 我们对 n 用归纳法. $n=1$ 的情形平凡, 我们现在设 $n \geqslant 2$. 假设结论对于所有 $n-1$ 阶的 Jordan 标准形只有一个 Jordan 块的矩阵成立. 因为 $\phi(B) = \phi(J(A)) = n-1$, B 有一列不含有非对角位置的非零元素. 于是, 存在 $P \in \Pi_n$ 使得

$$P^TBP = \begin{pmatrix} a & x^T \\ 0 & B_1 \end{pmatrix}, \tag{7.11}$$

其中 $x \in \mathbb{C}^{n-1}$, B_1 的阶数为 $n-1$. 因为 $J(B)$ 只有一个 Jordan 块, $x \neq 0$. 当 $n=2$ 时, x 只有一个 (非零) 分量. 如果当 $n \geqslant 3$ 时 x 有不止一个非零分量, 那么 $\phi(B_1) \leqslant n-3$, 因为 $\phi(B) = n-1$. 由引理 7.16, 存在一个 $Q \in \Pi_{n-1}$ 使得 $Q^TB_1Q = \mathrm{diag}(C_1, C_2)$. 因此 $J(B_1)$ 至少有两个 Jordan 块. 根据引理 7.17 这是不可能的, 因为 (7.11) 中的矩阵的 Jordan 标准形只有一个 Jordan 块. 所以, x 恰好有一个非零分量. 现在 $\phi(B_1) = n-2$ 并且根据引理 7.17, $J(B_1)$ 只有一个 Jordan 块. 由归纳假设, 存在一个 $M_1 \in \Gamma_{n-1}$ 使得 $M_1^{-1}B_1M_1 = J_{n-1}(a)$. 令 $M_2 = \mathrm{diag}(1, M_1)$, 从 (7.11) 我们得到

$$M_2^{-1}P^TBPM_2 = \begin{pmatrix} a & y^T \\ 0 & J_{n-1}(a) \end{pmatrix} = H, \tag{7.12}$$

其中 $y^T = x^TM_1$ 恰好有一个非零分量. 因为 H 相似于 B, $J(H) = J_n(a)$. 注意 $J_{n-1}(0)^{n-1} = 0$. 因此由 (7.12) 我们有

$$0 \neq (H-aI)^{n-1} = \begin{pmatrix} 0 & y^TJ_{n-1}(0)^{n-2} \\ 0 & 0 \end{pmatrix}.$$

于是 $0 \neq y^TJ_{n-1}(0)^{n-2}$. 这推出 y 的第一个分量非零, 因为 $J_{n-1}(0)^{n-2}$ 的唯一非零元素是 1, 在位置 $(1, n-1)$. 因此 $y = (y_1, 0, \cdots, 0)^T$, $y_1 \neq 0$. 现在令 $M = PM_2\mathrm{diag}(y_1, I_{n-1})$, 从 (7.12) 我们得 $M^{-1}BM = J_n(a)$. 这证明了当 $J(A)$ 只有一个 Jordan 块时结论成立.

下面假设 $J(A)$ 有 k 个 Jordan 块, $k \geqslant 2$:

$$J(A) = \mathrm{diag}(J_{n_1}(a_1), \cdots, J_{n_k}(a_k)).$$

则 $\phi(B) = \phi(J(A)) = n - k$. 由引理 7.16, 存在一个 $Q_1 \in \Pi_n$ 使得

$$Q_1^T B Q_1 = \mathrm{diag}(B_1, \cdots, B_k),$$

其中每个 B_j 是方阵. 因为 B 相似于 A, $J(Q_1^T B Q_1)$ 和 $J(A)$ 有相同的 Jordan 块 $J_{n_i}(a_i)$, $i = 1, \cdots, k$. 因此每个 $J(B_i)$ 只有一个 Jordan 块, $1 \leqslant i \leqslant k$, 并且有一个 $1, \cdots, k$ 的置换 γ 使得 $J(B_{\gamma(i)}) = J_{n_i}(a_i)$, $1 \leqslant i \leqslant k$. 令 $H_i = B_{\gamma(i)}$, 我们看到: 存在一个 $Q_2 \in \Pi_n$ 使得

$$Q_2^T Q_1^T B Q_1 Q_2 = \mathrm{diag}(H_1, \cdots, H_k). \tag{7.13}$$

因为 H_i 的阶数是 n_i 并且 $J(H_i)$ 只有一个 Jordan 块, 引理 7.16 推出

$$\phi(H_i) \geqslant n_i - 1, \quad i = 1, \cdots, k.$$

我们有

$$n - k = \sum_{i=1}^{k} \phi(H_i) \geqslant \sum_{i=1}^{k} (n_i - 1) = n - k.$$

所以 $\phi(H_i) = n_i - 1 = \phi(J(H_i))$ $i = 1, \cdots, k$. 应用已经证明了的 Jordan 标准形只含有一个 Jordan 块的矩阵的情形于 H_i, 我们推断: 存在一个 $M_i \in \Gamma_{n_i}$ 使得

$$M_i^{-1} H_i M_i = J_{n_i}(a_i), \quad i = 1, \cdots, k. \tag{7.14}$$

令 $M = Q_1 Q_2 \mathrm{diag}(M_1, \cdots, M_k)$, 则 $M \in \Gamma_n$. 由 (7.13) 和 (7.14) 我们有

$$M^{-1} B M = \mathrm{diag}(J_{n_1}(a_1), \cdots, J_{n_k}(a_k)) = J(A).$$

这就完成了证明. □

定理 7.18 表明: 如果不计较置换相似, $J(A)$ 是 $\mathcal{S}(A)$ 中非对角位置零元素的个数达到最大的唯一的零–非零模式. 这给出了 Jordan 标准形的一个组合刻画.

习 题

1. (Maybee [86]) 设 A 是一个树符号模式. 证明: (i) 若 A 的每个简单 2-圈都是正的, 则对于任何 $B \in Q(A)$, 存在可逆的实对角矩阵 D 使得 $D^{-1}AD$ 为对称矩阵; (ii) 若 A 的每个对角元素为 0 且 A 的每一个 2-圈都是负的, 则对于任何 $B \in Q(A)$, 存在可逆的实对角矩阵 D 使得 $D^{-1}AD$ 为反对称矩阵.

2. 证明引理 7.13.

3. 一个 n 阶符号模式方阵 A 称为是*谱任意模式*, 如果每个首一的 n 次实多项式都是 $Q(A)$ 中某个矩阵的特征多项式. 研究谱任意模式, 可以看 [51] 中的相关文献.

4. 怎样的符号模式方阵要求所有特征值都互不相同呢? 这个问题人们还没有很好地理解, 见 [80].

5. 元素属于 $\{0, *\}$ 的矩阵称为*零模式矩阵*. 设 A 是零模式矩阵, 用 $Q_F(A)$ 记元素属于域 F 的具有零模式 A 的矩阵的集合, 即若 $B \in Q_F(A)$, $B = (b_{ij})$, $A = (a_{ij})$, 则 $b_{ij} = 0$ 当且仅当 $a_{ij} = 0$. 设 F 的元素不少于 3 个. 证明: $Q_F(A)$ 中的每个矩阵非奇异当且仅当 A 置换等价于一个对角元素非零的上三角矩阵.

6. 举例说明: 存在那样的实方阵 A, A 的零元素的个数大于 A 的 Jordan 标准形的零元素的个数.

第八章　矩阵的应用

我们举例说明矩阵在数学内部其他领域的应用. 矩阵在数学之外的应用通常更加广为人知.

§8.1 图论

"友谊定理" 说: 如果在一群人中任何两个人都恰好有一个共同的朋友, 那么有一个人是每个人的朋友. "恰好有一个" 就是有且只有 1 个.

这个结果用图 (无向图) 论的语言可以叙述成下面的定理.

定理 8.1 (Erdös-Rényi-Sós [35])　*如果在图 G 中任何两个不同的顶点都恰好有一个共同的邻居, 那么 G 有一个顶点与其他所有顶点都邻接.*

证明　G 至少有 3 个顶点, 否则 G 只有一个顶点从而定理平凡地成立. 我们首先证明两个顶点 x, y 若不邻接则它们的度数相同. 用 $N(x)$ 记与 x 邻接的顶点的集合. 定义映射 $f: N(x) \to N(y)$, 对于 $z \in N(x)$, $f(z)$ 定义为 z 与 y 共同的邻居. $f(z) \neq x$, 因为 x 与 y 不邻接. 显然 f 为双射, 这就证明了 x 和 y 的度数相同.

假设定理的结论不成立. 设有一个顶点的度数为 $k > 1$, 我们将证明所有顶点的度数都为 k. 设 Ω 是度数为 k 的顶点的集合, Γ 是度数不为 k 的顶点的集合. 假如 Γ 不为空集, 则由第一段的结论, Ω 中的每个顶点都与 Γ 中的每个顶点相邻接. 如果 Ω 或 Γ 是单点集, 那么那个顶点与 G 的所有其他顶点相邻接, 这与假设矛盾. 所以 Ω 和 Γ 都不是单点集, 但是在此情形, Ω 中有两个顶点在 Γ 中有两个共同的邻居, 这与定理的假设矛盾. 因此, Γ 为空集, G 的所有顶点的度数都是 k, 即 G 是 k-正则的.

现在证明 G 的顶点的个数 $n = k(k-1)+1$. 我们用两种方法来算 G 的长度为 2

的路的条数 t. 一方面由假设 $t=\binom{n}{2}$. 另一方面对于每个顶点 v, 恰好有 $\binom{k}{2}$ 条长度为 2 的路以 v 为中间点, 因此 $t=n\cdot\binom{k}{2}$. 由 $\binom{n}{2}=n\cdot\binom{k}{2}$ 得 $n=k(k-1)+1$.

$k=1,2$ 给出 $n=1,3$, 这两种情形定理显然成立. 下面假设 $k\geqslant 3$.

设 A 是 G 的邻接矩阵, 用 J 记元素全为 1 的 n 阶矩阵, 则 $\mathrm{tr}A=0$. 由定理的条件及 G 是 k-正则的事实得

$$A^2=(k-1)I+J. \tag{8.1}$$

因为 J 的特征值是 n (1 重) 和 0 ($n-1$ 重), A^2 的特征值是 $k-1+n=k^2$ (1 重) 和 $k-1$ ($n-1$ 重). 由谱映射定理, A 的特征值是 k (1 重) 和 $\pm\sqrt{k-1}$. 设 A 有 r 个特征值等于 $\sqrt{k-1}$, 有 s 个特征值等于 $-\sqrt{k-1}$. 则由 A 的特征值之和等于它的迹得

$$k+r\sqrt{k-1}-s\sqrt{k-1}=0.$$

所以, $r\neq s$ 且

$$\sqrt{k-1}=\frac{k}{s-r}.$$

可见 $\sqrt{k-1}=h$ 是有理数, 从而是整数 ($\sqrt{m}$ 为有理数 $\Rightarrow$ $\sqrt{m}$ 为整数). 从

$$h(s-r)=k=h^2+1$$

知 h 整除 h^2+1. 于是 h 整除 $(h^2+1)-h^2=1$. 这就得出 $h=1$ 因而 $k=2$. 这与假设 $k\geqslant 3$ 矛盾. 这就证明了假设定理的结论不成立是错的. □

这个证明的前半部分取自 [62], 后半部分取自 [35].

满足友谊定理条件的图显然只有一种, 称为风车图, 它们由一些有公共顶点的三角形组成.

§8.2 数论

本节取自 Li 和 Lutzer 的论文 [77], 除了末尾的例子. 类似的想法也出现在 [38, 68] 中.

一个复数称为代数数, 如果它是某个系数为有理数的非零多项式的根. 显然, 如果 α 是个代数数, 则存在唯一的一个首项系数为 1 的不可约有理系数多项式 $p(x)$ 以 α 为根. α 的次数定义为 $p(x)$ 的次数. 代数数论中的一个基本事实是下面的定理.

定理 8.2 代数数的集合构成一个域.

在通常的数论书和代数书中, 这个定理都是用域扩张或者模来证明的. 现在我们用矩阵给出一个明确的构造性的证明.

定理 8.2 的证明 我们只需要证明如果 α, β 为代数数且 $\beta \neq 0$, 则 $\alpha \pm \beta$, $\alpha\beta, \alpha/\beta$ 都是代数数. 设 $p(x)$ 和 $q(x)$ 分别是 m 次和 n 次的首项系数为 1 的有理系数多项式使得 $p(\alpha) = 0, q(\beta) = 0$, 设 A 和 B 分别是 $p(x)$ 和 $q(x)$ 的伙伴矩阵. 则 α 和 β 分别是 A 和 B 的特征值. 由张量积的性质, $\alpha+\beta, \alpha-\beta$, $\alpha\beta$ 分别是 $A \otimes I_n + I_m \otimes B, A \otimes I_n - I_m \otimes B, A \otimes B$ 的特征值, 从而分别是这三个矩阵的特征多项式的根. 因为这三个矩阵是有理矩阵, 它们的特征多项式都是有理系数多项式. 因为 $\alpha/\beta = \alpha \cdot \beta^{-1}$, 我们只需要证明 β^{-1} 是代数数.

设 $q(x) = x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0$. 不妨设 $b_0 \neq 0$. 否则对于某个适当的正整数 k 考虑 $q(x)/x^k$. 令 $f(x) = b_0 x^n + b_1 x^{n-1} + \cdots + b_{n-1}x + 1$. 直接验证可知 $f(\beta^{-1}) = 0$. □

从上面的证明我们得到下面的推论.

推论 8.3 设 α 和 β 是次数分别为 m 和 n 的代数数且 $\beta \neq 0$, 则代数数 $\alpha \pm \beta$, $\alpha\beta, \alpha/\beta$ 的次数都不超过 mn.

一个复数称为代数整数, 如果它是某个首项系数为 1 的整系数多项式的根. 定理 8.2 的证明也给出下面定理的证明.

定理 8.4 代数整数的集合构成一个环.

书 [63] 的第 257 页说:“ 我们知道 $\sqrt[3]{2} + \sqrt{3}$ 是代数数, 但是如何找到一个明确的有理系数多项式 f 满足 $f(\sqrt[3]{2} + \sqrt{3}) = 0$ 当然是不完全明显的.” 利用上面的矩阵证明, 我们很容易找到这样一个 f. $\sqrt[3]{2}$ 是 $g(x) = x^3 - 2$ 的根, $\sqrt{3}$ 是 $h(x) = x^2 - 3$ 的根, $g(x)$ 和 $h(x)$ 的伙伴矩阵分别是

$$A = \begin{pmatrix} 0 & 0 & 2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 3 \\ 1 & 0 \end{pmatrix}.$$

则矩阵 $A \otimes I_2 + I_3 \otimes B$ 的特征多项式

$$f(x) = x^6 - 9x^4 - 4x^3 + 27x^2 - 36x - 23$$

就满足 $f(\sqrt[3]{2} + \sqrt{3}) = 0$.

§8.3 代数

定理 8.5 (Hilbert 零点定理) 设 F 是个代数闭域, L 是 $F[x_1, \cdots, x_n]$ 的一个真理想, 则存在 $(a_1, \cdots, a_n) \in F^n$ 使得 $f(a_1, \cdots, a_n) = 0$ 对所有 $f \in L$ 成立.

零点定理对于代数几何有基本的重要性. 它有多个证明, 这里我们给出一个用到矩阵的最近发现的证明.

引理 8.6 (Noether 规范化引理) 设 F 是个无限域, f 是 $F[x_1,\cdots,x_n]$ 中的一个次数为 d 的多项式, $n\geqslant 2, d\geqslant 1$, 则存在 $\lambda_1,\cdots,\lambda_{n-1}\in F$ 使得在多项式

$$f(x_1+\lambda_1 x_n,\cdots,x_{n-1}+\lambda_{n-1}x_n,x_n) \tag{8.2}$$

中 x_n^d 的系数不为零.

证明 设 f_d 是 f 的 d 次齐次部分, 则在多项式 (8.2) 中, x_n^d 的系数是 $f_d(\lambda_1,\cdots,\lambda_{n-1},1)$. 因为 $f_d(x_1,\cdots,x_{n-1},1)$ 是 $F[x_1,\cdots,x_{n-1}]$ 中的非零多项式并且 F 是无限的, 存在 F^{n-1} 中的一点 $(\lambda_1,\cdots,\lambda_{n-1})$ 使得 $f_d(\lambda_1,\cdots,\lambda_{n-1},1)\neq 0$, 这可以对变量的个数作归纳证明. □

代数闭域一定是无限域. 事实上, 设 $K=\{a_0,a_1,\cdots,a_m\}$ 是个有限域, $a_1\neq 0$, 则多项式 $f(x)=a_1+\prod\limits_{j=0}^{m}(x-a_j)$ 在 K 中没有根.

定理 8.5 的证明 (Arrondo [7]) 我们假设 $L\neq\{0\}$, 否则结论平凡地成立. 对 n 作归纳. $n=1$ 的情形是显然的, 因为 $F[x]$ 的任何一个非零真理想 L 都是由一个非常量的多项式生成, 而这样一个生成元在 F 中有一个根 a, 因为 F 是代数闭的. 所以, $f(a)=0$ 对 $\forall f\in L$ 成立.

现在设 $n\geqslant 2$, 并且假设定理对 $n-1$ 个变量的多项式环成立. 根据引理 8.6, 若有必要作变量替换并且把多项式乘以一个非零元, 我们可以假设 L 含有一个多项式 g 具有形式

$$g=g_0+g_1x_n+\cdots+g_{k-1}x_n^{k-1}+x_n^k,$$

其中 $g_j\in F[x_1,\cdots,x_{n-1}], j=0,1,\cdots,k-1$. 用 L' 记 L 中不含有变量 x_n 的多项式的集合, 则 L' 是 $F[x_1,\cdots,x_{n-1}]$ 的一个真理想. 由归纳假设, 存在一点 $(a_1,\cdots,a_{n-1})\in F^{n-1}$, 使得 L' 中的每个多项式在这点取值为 0. 我们断言

$$G=\{f(a_1,\cdots,a_{n-1},x_n)\mid f\in L\}$$

是 $F[x_n]$ 的一个真理想. G 显然是 $F[x_n]$ 的理想, 假设 G 不是真理想, 则有 $f\in L$ 使得 $f(a_1,\cdots,a_{n-1},x_n)=1$. 将 f 写成

$$f=f_0+f_1x_n+\cdots+f_dx_n^d,$$

其中所有 $f_i\in F[x_1,\cdots,x_{n-1}]$,

$$f_0(a_1,\cdots,a_{n-1})=1,\quad f_1(a_1,\cdots,a_{n-1})=\cdots=f_d(a_1,\cdots,a_{n-1})=0.$$

考虑 $k+d$ 阶矩阵

$$A(x_1,\cdots,x_{n-1})=\begin{pmatrix} f_0 & f_1 & & f_d & 0 & 0 & & 0 \\ 0 & f_0 & & f_{d-1} & f_d & 0 & & 0 \\ & & \ddots & & & & & \\ 0 & & 0 & f_0 & f_1 & & f_{d-1} & f_d \\ g_0 & g_1 & & g_{k-1} & 1 & 0 & \cdots & 0 \\ 0 & g_0 & & g_{k-2} & g_{k-1} & 1 & & 0 \\ & & \ddots & & & & & \\ 0 & & 0 & g_0 & g_1 & & g_{k-1} & 1 \end{pmatrix},$$

其中 A 的对角线上有 k 个 f_0, d 个 1. A 的行列式

$$R(x_1,\cdots,x_{n-1})=\det A(x_1,\cdots,x_{n-1})$$

就是 f 和 g 关于 x_n 的结式. 熟知且易验证 R 是 f 和 g 的线性组合, 即存在 $u,w\in F[x_1,\cdots,x_{n-1}]$ 满足 $R=uf+wg$. 于是, $R\in L$, 从而 $R\in L'$. 因为 $A(a_1,\cdots,a_{n-1})$ 是一个对角元素全为 1 的下三角矩阵, $R(a_1,\cdots,a_{n-1})=1$. 这与 $R\in L'$ 矛盾, 因为 L' 中的每个多项式在点 $(a_1,\cdots,a_{n-1})$ 取值都是 0. 所以 G 是 $F[x_n]$ 的一个真理想.

作为 $F[x_n]$ 的真理想, G 由某个多项式 $h(x_n)$ 生成, $\deg h\geqslant 1$ 或者 $h=0$. 因为 F 是代数闭的, 无论哪种情况都有 $a_n\in F$ 使得 $h(a_n)=0$. 因此对所有 $f\in L, f(a_1,\cdots,a_{n-1},a_n)=0$. □

§8.4 多项式

本节取自 Kittaneh 的论文 [72]. 我们想估计复系数多项式

$$p(z)=z^n+a_{n-1}z^{n-1}+\cdots+a_1z+a_0$$

的根的范围, 这里 $n\geqslant 2$. Carmichael 和 Mason 关于 $p(z)$ 的根 z 的界是

$$|z|\leqslant\left(1+\sum_{i=0}^{n-1}|a_i|^2\right)^{1/2} \tag{8.3}$$

我们将利用伴随矩阵的奇异值来改进这个界. $p(z)$ 的伴随矩阵是

$$C(p)=\begin{pmatrix} -a_{n-1} & -a_{n-2} & \cdots & -a_1 & -a_0 \\ 1 & 0 & & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ & & & & \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

设 $z_1,\cdots,z_n$ 是 $p(z)$ 的根, $|z_1|\geqslant|z_2|\geqslant\cdots\geqslant|z_n|$. 则 $z_1,\cdots,z_n$ 是 $C(p)$ 的特征值. 记 $C(p)$ 的奇异值为 $s_1\geqslant\cdots\geqslant s_n$. 根据 Weyl 定理 (定理 4.10), 有

$$\prod_{i=1}^{k}|z_i|\leqslant\prod_{i=1}^{k}s_i,\quad k=1,\cdots,n, \tag{8.4}$$

$$\prod_{i=k}^{n}|z_i|\geqslant\prod_{i=k}^{n}s_i,\quad k=1,\cdots,n. \tag{8.5}$$

下面我们来求出 $C(p)$ 的奇异值 $s_i,i=1,\cdots,n$ (这个方法属于 Roger A. Horn). 记

$$\gamma=\sum_{i=0}^{n-1}|a_i|^2.$$

则

$$I-C(p)C(p)^*=\begin{pmatrix}1-\gamma & a_{n-1} & a_{n-2} & \cdots & a_1\\ \bar{a}_{n-1} & 0 & 0 & & 0\\ \bar{a}_{n-2} & 0 & 0 & & 0\\ & & & \cdots & \cdots\\ \bar{a}_1 & 0 & 0 & & 0\end{pmatrix}. \tag{8.6}$$

(8.6) 中的 Hermite 矩阵至多有两列线性无关, 所以它的秩至多为 2, 从而它有至少 $n-2$ 个零特征值. 于是, $C(p)C(p)^*$ 有至少 $n-2$ 个特征值等于 1. 设 $C(p)C(p)^*$ 的其余两个特征值为 λ,μ. 我们有

$$n-2+\lambda+\mu=\mathrm{tr}C(p)C(p)^*=\|C(p)\|_F^2=n-1+\gamma,$$

$$\lambda\mu\cdot\underbrace{1\cdots1}_{n-2}=\lambda\mu=\det C(p)C(p)^*=|\det C(p)|^2=|a_0|^2.$$

可见 λ 和 μ 是方程 $x^2-(\gamma+1)x+|a_0|^2=0$ 的根. 这样我们就得到了

引理 8.7　$C(p)$ 的奇异值是

$$s_1=\left\{\frac{\gamma+1+\left[(\gamma+1)^2-4|a_0|^2\right]^{1/2}}{2}\right\}^{1/2},$$

$$s_n=\left\{\frac{\gamma+1-\left[(\gamma+1)^2-4|a_0|^2\right]^{1/2}}{2}\right\}^{1/2},$$

$$s_i=1,\quad i=2,\cdots,n-1.$$

将 (8.4), (8.5) 和引理 8.7 合起来我们得到

定理 8.8

$$\prod_{i=1}^{k}|z_i| \leqslant s_1, \quad k=1,\cdots,n-1,$$
$$\prod_{i=k}^{n}|z_i| \geqslant s_n, \quad k=2,\cdots,n.$$

显然, 定理 8.8 的特殊情形 $|z_1| \leqslant s_1$ (从而 $|z_i| \leqslant s_1, i=1,\cdots,n$) 优于界 (8.3). 利用 (8.4) 和引理 8.7 并且应用定理 3.23 我们得到

定理 8.9 对任何正数 t 有

$$\sum_{i=1}^{k}|z_i|^t \leqslant s_1^t + k - 1, \quad k=1,\cdots,n-1.$$

§8.5 有限几何

本节取自 Houck 和 Paul 的论文 [60], 但这些结果属于 de Bruijn 和 Erdös [31]. 论文 [60] 的新颖之处在于它利用矩阵给出的简洁证明.

设 S 是一个有限集合, S 中的元素称为点, 设 $S_1, S_2, \cdots, S_m$ 为 S 的子集, 每个 S_i 含有至少两个点, 这些子集 S_i 称为线. 我们说这些点和线构成一个有限几何, 如果它们满足下面三条公理:

公理 1 任给两点, 存在一条线包含这两点.

公理 2 两条线相交于至多一个点.

公理 3 存在至少两条线.

公理 1 和公理 2 合起来表明: 任给两点, 存在唯一的一条线包含它们. 给定有限几何 $G=(S; S_1,\cdots,S_m)$, 设 $S=\{p_1,p_2,\cdots,p_n\}$. 点 p_i 的次数 d_i 定义为包含 p_i 的线的条数. 用 k_j 记线 S_j 所含点的个数. G 的关联矩阵 $A=(a_{ij})_{n\times m}$ 是个 0–1 矩阵, 定义为

$$a_{ij} = \begin{cases} 1, & \text{若点 } p_i \text{ 在线 } S_j \text{ 上}, \\ 0, & \text{否则}. \end{cases}$$

A 的行对应点而列对应线. 由定义,

$$d_i = \sum_{j=1}^{m} a_{ij} \geqslant 2, \quad k_j = \sum_{i=1}^{n} a_{ij}.$$

定理 8.10 每个有限几何的线的条数不少于点的个数.

证明　用上面的记号, 我们要证明 $m \geqslant n$. 用 J 表示所有元素都是 1 的方阵, 其阶数由上下文确定. 我们有

$$AA^T = \operatorname{diag}(d_1 - 1, d_2 - 1, \cdots, d_n - 1) + J.$$

因为 $d_i \geqslant 2, i = 1, \cdots, n, AA^T$ 等于一个正定矩阵加一个半正定矩阵因而是正定矩阵. 所以

$$n = \operatorname{rank} AA^T = \operatorname{rank} A.$$

但 A 是 $n \times m$ 的, 故 $m \geqslant n$. □

现在我们来考虑线数等于点数的情形.

定理 8.11　若一个有限几何的线数等于点数, 则任意两条线都相交.

证明　沿用上面的记号. 现在 $m = n$. 记 $B = J - A^T$. 则

$$AB = AJ - AA^T = \begin{pmatrix} 0 & d_1 - 1 & d_1 - 1 & \cdots & d_1 - 1 \\ d_2 - 1 & 0 & d_2 - 1 & \cdots & d_2 - 1 \\ d_3 - 1 & d_3 - 1 & 0 & \cdots & d_3 - 1 \\ & & & & \\ d_n - 1 & d_n - 1 & d_n - 1 & \cdots & 0 \end{pmatrix},$$

$$\det(AB) = \prod_{i=1}^{n} (d_i - 1) \det(J - I) \neq 0.$$

所以 $\det B \neq 0$. 从而 B 有一条广义对角线上面的每个元素都不为 0. 因为 B 是 0–1 矩阵, 那条广义对角线上面的元素全为 1. 于是 A 有一条广义对角线上面的元素全为 0. 可见存在置换矩阵 P 使得 PA 的主对角元素全为 0. 因此将点重新命名我们可以假定 A 的主对角元全为 0. 此时, 点 p_i 不在线 S_i 上, 而线 S_i 上的每个点产生一条通过 p_i 的线, 所以 $k_i \leqslant d_i, i = 1, \cdots, n$. 但是, $\sum\limits_{i=1}^{n} k_i = \sum\limits_{i,j=1}^{n} a_{ij} = \sum\limits_{i=1}^{n} d_i$, 我们得到 $k_i = d_i, i = 1, \cdots, n$.

我们来数不同的两个点组成的无序对的个数. 注意到每对不同的点都位于唯一的一条线上, 有

$$\binom{n}{2} = \sum_{i=1}^{n} \binom{k_i}{2} = \sum_{i=1}^{n} \binom{d_i}{2}$$

$$= \operatorname{card}\{(i,j) \mid 1 \leqslant i < j \leqslant n, S_i \cap S_j \neq \varnothing\},$$

其中 card 表示集合的势. 这就证明了任意两条线都必须相交. □

考虑下面关于有限几何的公理:

公理 4 存在 4 个点, 其中任何 3 个点都不共线.

假设有限几何 G 的线数和点数都是 n, 我们分两种情况:

(i) 若 G 不满足公理 4, 则由定理 8.11 知存在一条线 $S_{j_0}, k_{j_0} = n-1$, 其他 $k_j = 2$. 此时若有必要对点和线重新编号, 我们有 $S = \{p_1, \cdots, p_n\}, S_1 = \{p_1, \cdots, p_{n-1}\}$, $S_2 = \{p_1, p_n\}, \cdots, S_n = \{p_{n-1}, p_n\}$.

(ii) 若 G 满足公理 4, 则利用定理 8.11 可知任给两条线, 存在一点不在这两条线上. 由这个性质及定理 8.11 我们推断

$$k_1 = k_2 = \cdots = k_n = d_1 = d_2 = \cdots = d_n.$$

记这个公共的数为 k, 则由定理 8.11 的证明知 $n = k(k-1)+1$. 在此情形, 每条线上有 k 个点, 每个点位于 k 条线上.

附录 未解决的问题

这个附录取自文章 [123], 最后一节作了修改. 附录中提到的参考文献见附录末尾的第 137 页.

Sometimes solutions to challenging matrix problems can reveal connections between different parts of mathematics. Two examples of this phenomenon are the proof of the van der Waerden conjecture on permanents (see [47] or [69]) and the recent proof of Horn's conjecture on eigenvalues of sums of Hermitian matrices (see [11] and [32]). Difficult matrix problems can also expose limits to the strength of existing mathematical tools.

We will describe the history and current state of some open problems in matrix theory, which we arrange chronologically in the following sections.

1. Existence of Hadamard matrices

A Hadamard matrix is a square matrix with entries equal to ± 1 whose rows and hence columns are mutually orthogonal. In other words, a Hadamard matrix of order n is a $\{1,-1\}$-matrix A satisfying

$$AA^T = nI,$$

where I is the identity matrix. In 1867 Sylvester proposed a recurrent method for construction of Hadamard matrices of order 2^k. In 1893 Hadamard proved his famous determinantal inequality for a positive semidefinite matrix A :

$$\det A \leqslant h(A),$$

where $h(A)$ is the product of the diagonal entries of A. It follows from this inequality that if $A = (a_{ij})$ is a real matrix of order n with $|a_{ij}| \leqslant 1$ then

$$|\det A| \leqslant n^{n/2};$$

equality occurs if and only if A is a Hadamard matrix. This result gives rise to the term "Hadamard matrix". In 1898 Scarpis proved that if $p \equiv 3 \pmod 4$ or $p \equiv 1 \pmod 4$ is a prime number then there is a Hadamard matrix of order $p+1$ and $p+3$ respectively.

In 1933 Paley stated that the order n ($n \geqslant 4$) of any Hadamard matrix is divisible by 4. This is easy to prove. The converse has been a long-standing conjecture.

Conjecture 1 *For every positive integer n, there exists a Hadamard matrix of order $4n$.*

Conjecture 1 has been proved for $4n = 2^k m$ with $m^2 \leqslant 2^k$. According to [68], the smallest unknown case is now $4n = 668$. See [34, 57, 58, 63, 64].

Hadamard matrices have applications in information theory and combinatorial designs. See [1].

Let $k \leqslant n$ be positive integers. A square matrix A of order n with entries in $\{0, -1, 1\}$ is called a *weighted matrix with weight k* if

$$AA^T = kI.$$

Geramita and Wallis posed the following more general conjecture in 1976 [33].

Conjecture 2 *If $k \leqslant n$ are positive integers with $n \equiv 0 \pmod 4$, then there exists a weighted matrix of order n with weight k.*

Note that Conjecture 1 corresponds to the case $k = n$ of Conjecture 2.

2. Characterization of the eigenvalues of nonnegative matrices

In 1937 Kolmogorov asked the question: When is a given complex number an eigenvalue of some (entrywise) nonnegative matrix? The answer is: Every complex number is an eigenvalue of some nonnegative matrix [52, p.166]. Suleimanova [62] extended Kolmogorov's question in 1949 to the following problem which is called the *nonnegative inverse eigenvalue problem.*

Problem 3 *Determine necessary and sufficient conditions for a set of n complex numbers to be the eigenvalues of a nonnegative matrix of order n.*

Problem 3 is open for $n \geqslant 4$. The case $n = 2$ is easy while the case $n = 3$ is due to Loewy and London [48].

In the same paper [62] Suleimanova also considered the following *real nonnegative inverse eigenvalue problem* and gave a sufficient condition.

Problem 4 *Determine necessary and sufficient conditions for a set of n real numbers to be the eigenvalues of a nonnegative matrix of order n.*

Problem 4 is open for $n \geqslant 5$. In 1974 Fiedler [29] posed the following *symmetric nonnegative inverse eigenvalue problem.*

Problem 5 *Determine necessary and sufficient conditions for a set of n real numbers to be the eigenvalues of a symmetric nonnegative matrix of order n.*

Problem 5 is open for $n \geqslant 5$. There are some necessary conditions and many sufficient conditions for these three problems. See the survey paper [27] and the book [52, Chapter VII].

3. The permanental dominance conjecture

Let S_n denote the symmetric group on $\{1, 2, \cdots, n\}$ and M_n denote the set of complex matrices of order n. Suppose G is a subgroup of S_n and χ is a character of G. The *generalized matrix function* $d_\chi : M_n \to C$ is defined by

$$d_\chi(A) = \sum_{\sigma \in G} \chi(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)},$$

where $A = (a_{ij})$. Incidental to his work on group representation theory, Schur introduced this notion. For $G = S_n$, if χ is the alternating character then d_χ is the determinant while if χ is the principal character then d_χ is the permanent

$$\mathrm{per}A = \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i\sigma(i)}.$$

When χ is the principal character of $G = \{e\}$ where e is the identity permutation in S_n, d_χ is Hadamard's function $h(A)$.

In 1907 Fischer proved that if the matrix

$$A = \begin{pmatrix} A_1 & B \\ B^* & A_2 \end{pmatrix}$$

is positive semidefinite with A_1 and A_2 square, then

$$\det A \leqslant (\det A_1)(\det A_2).$$

Hadamard's inequality follows from this inequality immediately. In 1918 Schur obtained the following generalization of Fischer's inequality:

$$\chi(e)\det A \leqslant d_\chi(A)$$

for positive semidefinite A. Let G be a subgroup of S_n and let χ be an irreducible character of G. The normalized generalized matrix function is defined as

$$\bar{d}_\chi(A) = d_\chi(A)/\chi(e).$$

Since any character of G is a sum of irreducible characters, Schur's inequality is equivalent to

$$\det A \leqslant \bar{d}_\chi(A)$$

for positive semidefinite A. In 1963, M. Marcus proved the permanental analog of Hadamard's inequality

$$\mathrm{per}A \geqslant h(A)$$

and E.H. Lieb proved the permanental analog of Fischer's inequality

$$\mathrm{per}A \geqslant (\mathrm{per}A_1)(\mathrm{per}A_2)$$

three years later, where A is positive semidefinite. These results naturally led to the following conjecture which was first published by Lieb [45] in 1966:

Conjecture 6 (The permanental dominance conjecture) *Suppose G is a subgroup of S_n and χ is an irreducible character of G. Then for any positive semidefinite matrix A of order n,*

$$\mathrm{per}A \geqslant \bar{d}_\chi(A).$$

A lot of work has been done on this conjecture. It has been confirmed for every irreducible character of S_n with $n \leqslant 13$. The reader is referred to [22, section 3] and the references therein for more details and recent progress.

We order the elements of S_n lexicographically to obtain a sequence L_n. For $A = (a_{ij}) \in M_n$ the *Schur power* of A, denoted by $\Pi(A)$, is the matrix of order $n!$ whose rows and columns are indexed by L_n and whose (σ,τ)-entry is $\prod_{i=1}^n a_{\sigma(i),\tau(i)}$. Since $\Pi(A)$ is a principal submatrix of $\otimes^n A$, if A is positive semidefinite then so is $\Pi(A)$. It is not difficult to see that both $\mathrm{per}A$ and $\det A$ are eigenvalues of $\Pi(A)$. A result of Schur asserts that if A is positive semidefinite then $\det A$ is the smallest eigenvalue of $\Pi(A)$. In 1966, Soules [61] posed the following

Conjecture 7 (The "permanent on top" conjecture) *If the matrix A is positive semidefinite, then* $\mathrm{per}A$ *is the largest eigenvalue of* $\Pi(A)$.

Conjecture 7, if true, implies Conjecture 6.

4. The Marcus-de Oliveira conjecture

Let S_n denote the symmetric group on $\{1, 2, \cdots, n\}$ and $\mathrm{co}\Omega$ denote the convex hull of a set Ω in the complex plane. In 1973 Marcus [50] and in 1982 de Oliveira [56] independently made the following

Conjecture 8 *Let A, B be normal complex matrices of order n with eigenvalues $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$ respectively. Then*

$$\det(A+B) \in \mathrm{co}\left\{\prod_{i=1}^n \left(x_i + y_{\sigma(i)}\right) : \sigma \in S_n\right\}.$$

It is known that Conjecture 8 is true in many special cases, e.g., (1) A, B are Hermitian [30]; (2) all the eigenvalues have the same modulus, $|x_1| = \cdots = |x_n| = |y_1| = \cdots = |y_n|$ [8]; (3) $A + B$ is singular [26]. See [6, 7] for more verified cases.

5. Permanents of Hadamard matrices

In 1974 Wang [65] posed the following

Question 9 *Can the permanent of a Hadamard matrix of order n vanish for $n > 2$?*

Wanless [66] showed that the answer is negative for $2 < n < 32$.

6. The Bessis-Moussa-Villani trace conjecture

In 1975, while studying partition functions of quantum mechanical systems, Bessis, Moussa and Villani [10] formulated the conjecture that if A, B are Hermitian matrices of the same order with B positive semidefinite then the function

$$f(t) = \operatorname{Tr} \exp(A - tB)$$

is the Laplace transform of a positive measure on $[0, \infty)$, where t is a real variable and Tr means trace. Recently, Lieb and Seiringer [46] has proved that this conjecture is equivalent to the following

Conjecture 10 (Bessis-Moussa-Villani) *Let A, B be positive semidefinite matrices of order n and let k be a positive integer. Then the polynomial $p(t) = \operatorname{Tr}(A + tB)^k$ has all nonnegative coefficients.*

The following cases of this conjecture are proved: (1) $k \leqslant 5$ and all n; (2) $n = 2$ and any k. See [38] and the references therein for many partial results and recent advances.

7. The S-matrix conjecture

An S-matrix of order n is a 0-1 matrix formed by taking a Hadamard matrix of order $n + 1$ in which the entries in the first row and column are 1, changing 1's to 0's and -1's to 1's, and deleting the first row and column. Let $\|\cdot\|_F$ denote the Frobenius norm. In 1976 Sloane and Harwit [60] made the following conjecture. See also [37, p.59].

Conjecture 11 *If A is a nonsingular matrix of order n all of whose entries are in the interval* [0, 1], *then*

$$\|A^{-1}\|_F \geqslant \frac{2n}{n+1}.$$

Equality holds if and only if A is an S-matrix.

I do not know of any results on Conjecture 11. This problem arose from weighing designs in optics and statistics.

8. Ryser's conjecture on minimum values of permanents

Let Λ_n^k be the set of those $n \times n$ 0-1 matrices with each row and column having exactly k 1's. In 1978 Ryser (in [53]) posed the following

Conjecture 12 *If Λ_v^k contains incidence matrices of (v, k, λ)-designs, then the permanent takes its minimum in Λ_v^k at one of these incidence matrices.*

Wanless [67] showed by computer enumeration that this conjecture is true for $v \leqslant 12$.

9. Foregger's conjecture on minimum values of permanents

A fully indecomposable square matrix A is called *nearly decomposable* if whenever a nonzero entry of A is replaced with a 0, the resulting matrix is partly decomposable. In 1980 Foregger [31] made the following

Conjecture 13 *If A is a nearly decomposable doubly stochastic matrix of order n, then*

$$\mathrm{per} A \geqslant 2^{1-n}.$$

Note that this lower bound can be attained at $A = (I + P)/2$ where P is the permutation matrix corresponding to the permutation cycle $(1234 \cdots n)$. Foregger [31] proved the cases $2 \leqslant n \leqslant 9$.

10. Dittert's conjecture on permanents

Let K_n be the set of $n \times n$ nonnegative matrices with the sum of their entries equal to n. Define the function ϕ on $n \times n$ matrices by

$$\phi(A) = \prod_{i=1}^{n} r_i + \prod_{j=1}^{n} c_j - \mathrm{per} A,$$

where $r_1, \cdots, r_n$ and $c_1, \cdots, c_n$ are the row and column sums of A respectively. In 1983 Dittert (in [54]) posed the following

Conjecture 14

$$\max\{\phi(A) : A \in K_n\} = 2 - \frac{n!}{n^n},$$

and the maximum is attained only for the matrix with each entry equal to $1/n$.

Sinkhorn [59] proved the case $n = 2$ and Hwang [40] proved the case $n = 3$.

11. The Brualdi-Li conjecture on tournament matrices

A tournament matrix is a square 0-1 matrix A satisfying $A + A^T = J - I$ where J is the all ones matrix. Such matrices arise from the results of round robin competitions.

An $n\times n$ tournament matrix A is called *regular* if each of the row sums of A is $(n-1)/2$. For even n, an $n \times n$ tournament matrix is called *almost regular* if half of its row sums are $(n-2)/2$ and the other half are $n/2$. It is known [19] that for odd n the regular tournament matrices maximize the Perron root over the class of $n \times n$ tournament matrices. For even n, it is not known which tournament matrices maximize the Perron root. Let U_k be the strictly upper triangular matrix of order k with ones above the main diagonal. In 1983, Brualdi and Li [20] made the following

Conjecture 15 *For even n, the matrix*

$$\begin{pmatrix} U_{n/2} & U_{n/2}^T \\ U_{n/2}^T + I & U_{n/2} \end{pmatrix}$$

maximizes the Perron root over the class of tournament matrices of order n.

See [28] and the references therein for some partial results. Kirkland [43] has proved that for all sufficiently large even orders n the maximizers are almost regular.

12. A possible generalization of the Perron-Frobenius theorem

Let $A = (A_{ij})_{n\times n}$ be a block matrix of order nm, where each A_{ij} is a positive semidefinite matrix of order m. Let us call such matrices *block positive semidefinite (BPSD)*. Note that when $m = 1$, A is a nonnegative matrix, while when $n = 1$, A is a positive semidefinite matrix. Thus BPSD matrices interpolate two familiar classes of matrices. Both nonnegative matrices and positive semidefinite matrices have the Perron-Frobenius property: The spectral radius is an eigenvalue.

Numerical experiments show that some BPSD matrices have the Perron-Frobenius property while some others do not. In 1988, Roger A. Horn posed the following problem in an unpublished note.

Problem 16 *Let A be a BPSD matrix. Give necessary and/or sufficient conditions on A such that the spectral radius $\rho(A)$ is an eigenvalue of A. More generally, study the properties of the eigenvalues and eigenvectors of BPSD matrices.*

13. The Grone-Merris conjecture on Laplacian spectra

Let G be a graph of order n and let $d(G) = (d_1, \cdots, d_n)$ be the degree sequence of G, where $d_1, \cdots, d_n$ are the degrees of the vertices of G. Let $A(G)$ be the adjacency matrix of G and denote $D(G) = \mathrm{diag}(d_1, \cdots, d_n)$. Then the matrix $L(G) = D(G) - A(G)$ is called the *Laplacian matrix* of G, and $s(G) = (\lambda_1, \cdots, \lambda_n)$ is called the *Laplacian spectrum,* where $\lambda_1, \cdots, \lambda_n$ are the eigenvalues of $L(G)$.

For a sequence $x = (x_1, \cdots, x_n)$ with nonnegative integer components, the *conju-*

gate sequence of x is $x^\star = (x_1^\star, \cdots, x_n^\star)$, where

$$x_j^\star = |\{i : x_i \geqslant j\}|.$$

Denote by $d^\star(G)$ the conjugate sequence of the degree sequence $d(G)$. Use $x \prec y$ to mean that x is majorized by y. In 1994, Grone and Merris [35] made the following

Conjecture 17 *Let G be a connected graph. Then $s(G) \prec d^\star(G)$.*

If we arrange the components of $s(G)$ and $d^\star(G)$ in decreasing order, then it is known that

$$\lambda_1 \leqslant d_1^\star, \quad \lambda_1 + \lambda_2 \leqslant d_1^\star + d_2^\star.$$

14. The CP-rank conjecture

An $n \times n$ real matrix A is called *completely positive* (CP) if, for some m, there exists an $n \times m$ nonnegative matrix B such that $A = BB^T$. The smallest such m is called the *CP-rank* of A. CP matrices have applications in block designs. In 1994, Drew, Johnson and Loewy [25] posed the following

Conjecture 18 *If A is a CP matrix of order $n \geqslant 4$, then*

$$\text{CP-rank}\,(A) \leqslant \lfloor n^2/4 \rfloor.$$

This conjecture is true in the following cases: (1) $n = 4$ [51]; (2) $n = 5$ and A has at least one zero entry [49]; (3) the graph of A does not contain an odd cycle of length greater than 4 [24, 9]. It is known [25, p.309] that for each $n \geqslant 4$ the conjectured upper bound $\lfloor n^2/4 \rfloor$ can be attained.

15. Bhatia-Kittaneh's question on singular values

Denote by $s_1(X) \geqslant s_2(X) \geqslant \cdots$ the ordered singular values of a complex matrix X. In 2000, Bhatia and Kittaneh [17] asked the following

Question 19 *Let A, B be positive semidefinite matrices of order n. Is it true that*

$$s_j^{1/2}(AB) \leqslant \frac{1}{2} s_j(A+B), \quad j = 1, 2, \cdots, n?$$

The case $n = 2$ is known to be true [17]. Since the square function $f(t) = t^2$ is operator convex on R, this inequality is stronger than the known inequality

$$2s_j(XY^*) \leqslant s_j(X^*X + Y^*Y), \quad j = 1, 2, \cdots, n$$

for any complex matrices X, Y of order n due to the same authors [18].

16. Convergence of the iterated Aluthge transforms

Every square complex matrix A has the polar decomposition $A = UP$ where U is unitary and P is positive semidefinite. The *Aluthge transform* of A is

$$\Delta(A) = P^{1/2}UP^{1/2}.$$

Though the unitary factor in the polar decomposition is not unique when A is singular, the Aluthge transform is well defined, that is, it does not depend on the choice made for the unitary factor. If A is normal, then $\Delta(A) = A$. For $0 < \lambda < 1$, the λ-*Aluthge transform*

$$\Delta_\lambda(A) = P^\lambda U P^{1-\lambda}$$

is also well defined. Note that $\Delta = \Delta_{1/2}$. Let $B(H)$ be the algebra of bounded linear operators on a Hilbert space H. The Aluthge transform can also be defined for operators in $B(H)$. In 2000, Jung, Ko and Pearcy [41] conjectured that for any $T \in B(H)$, the sequence $\{\Delta^m(T)\}_{m=1}^\infty$ is norm convergent to an operator. Here $\Delta^1(T) = \Delta(T)$ and $\Delta^m(T) = \Delta(\Delta^{m-1}(T))$, $m = 2, 3, \cdots$. However Cho, Jung and Lee [23] showed that this conjecture is false for infinite dimensional Hilbert spaces. So there remains the possibility that it holds in finite dimensions:

Conjecture 20 *Let A be a square complex matrix. Then the sequence $\{\Delta^m(A)\}_{m=1}^\infty$ converges.*

Note that $\Delta(A)$ and A have the same eigenvalues. It is known ([41, Prop. 1.10], [42, Prop. 3.1], [2, Thm 1]) that if the Aluthge sequence of a matrix converges, then the limit matrix is normal. These two properties make the Aluthge transform more interesting.

Ando and Yamazaki [4] verified Conjecture 20 when A is of order 2. See [42] for some special cases. Huang and Tam [39] proved that if the nonzero eigenvalues of A have distinct moduli, then the λ-Aluthge sequence $\{\Delta_\lambda^m(A)\}_{m=1}^\infty$ converges. Huang and Tam [39] also posed the following

Conjecture 21 *For any square complex matrix A and $0 < \lambda < 1$,*

$$\|A^*A - AA^*\|_F \geqslant \|\Delta_\lambda(A)^*\Delta_\lambda(A) - \Delta_\lambda(A)\Delta_\lambda(A)^*\|_F.$$

17. Expressing real matrices as linear combinations of orthogonal matrices

In 2002, Li and Poon [44] proved that every square real matrix is a linear combination of 4 orthogonal matrices, i.e., given a square real matrix A, there exist real orthogonal matrices Q_i and real numbers r_i, $i = 1, 2, 3, 4$ (depending on A, of course) such that

$$A = r_1Q_1 + r_2Q_2 + r_3Q_3 + r_4Q_4.$$

They asked the following

Question 22 *Is the number* 4 *of the terms in the above expression least possible?*

18. Sign patterns

Research on sign patterns of matrices is active now and there are many open problems in that field. See [21] and [36].

Let $f(A)$ be the number of positive entries of a nonnegative matrix A. In a talk at the 12th ILAS conference (Regina, Canada, June 26-29, 2005) I posed the following

Problem 23 *Characterize those sign patterns of square nonnegative matrices A such that the sequence $\{f(A^k)\}_{k=1}^{\infty}$ is nondecreasing.*

Sidak observed in 1964 that there exists a primitive nonnegative matrix A of order 9 satisfying

$$18 = f(A) > f(A^2) = 16.$$

This is the motivation for Problem 23.

We may consider the same problem with "nondecreasing" replaced by "nonincreasing". Perhaps the first step is to study the case when A is irreducible.

19. Monotonicity of a geometric mean of positive definite matrices

The following geometric mean of three or more positive definite matrices has recently been defined in [55] and [15] independently using a geometric approach. See also [12, 16]. Here we give only the basic idea. Denote by P_n the set of positive definite matrices of order n. We consider P_n as a differentiable manifold. The distance $\delta(A, B)$ between A, $B \in P_n$ is the infimum of lengths of curves in P_n that connect A to B. It can be proved that $\delta(A, B) = \|\log(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})\|_F$. Given $A_i \in P_n$, $i = 1, 2, \cdots, k$, there is a unique matrix in P_n, denoted $G(A_1, \cdots, A_k)$, that minimizes the function

$$f(X) = \sum_{i=1}^{k} \delta^2(A_i, X).$$

Then $G(A_1, \cdots, A_k)$ is called the *geometric mean* of $A_1, \cdots, A_k$. This geometric mean is symmetric, invariant under congruence, and continuous. Let " $\leqslant$ " be the Loewner partial order. In 2006, Bhatia and Holbrook [15, p.616] made the following

Conjecture 24 *The mean G is monotone with respect to its arguments, i.e., if A_i, $B_i \in P_n$ satisfy $A_i \leqslant B_i$, $i = 1, \cdots, k$, then*

$$G(A_1, \cdots, A_k) \leqslant G(B_1, \cdots, B_k).$$

Another geometric mean of three or more positive definite matrices has been defined in [3].

20. Eigenvalues of real symmetric matrices

Let $S_n[a,b]$ denote the set of $n \times n$ real symmetric matrices whose entries are in the interval $[a, b]$. For an $n \times n$ real symmetric matrix A, we always denote the eigenvalues of A in decreasing order by $\lambda_1(A) \geqslant \cdots \geqslant \lambda_n(A)$. The *spread* of an $n \times n$ real symmetric matrix A is $s(A) = \lambda_1(A) - \lambda_n(A)$.

The following two problems were posed in [70].

Problem 25 *For a given j with $2 \leqslant j \leqslant n-1$, determine*

$$\max\{\lambda_j(A) : A \in S_n[a,b]\},$$

$$\min\{\lambda_j(A) : A \in S_n[a,b]\}$$

and determine the matrices that attain the maximum and the matrices that attain the minimum.

The cases $j = 1, n$ are solved in [70].

Problem 26 *Determine*

$$\max\{s(A) : A \in S_n[a,b]\}$$

and determine the matrices that attain the maximum.

The special case when $a = -b$ is solved in [70].

21. Sharp constants in spectral variation

Let α_j and β_j, $j = 1, \cdots, n$, be the eigenvalues of $n \times n$ complex matrices A and B, respectively, and denote

$$\mathrm{Eig}A = \{\alpha_1, \cdots, \alpha_n\}, \quad \mathrm{Eig}B = \{\beta_1, \cdots, \beta_n\}.$$

The *optimal matching distance* between the spectra of A and B is

$$d(\mathrm{Eig}A,\ \mathrm{Eig}B) = \min_{\sigma} \max_{1 \leqslant j \leqslant n} |\alpha_j - \beta_{\sigma(j)}|$$

where σ varies over all permutations of the indices $\{1, 2, \cdots, n\}$.

Let $\|\cdot\|$ be the spectral norm. It is known [13] that there exists a number c with $1 < c < 3$ such that

$$d(\mathrm{Eig}A,\ \mathrm{Eig}B) \leqslant c\|A - B\|$$

for any normal matrices A, B of any order. See [14] for the very interesting history of this result.

Bhatia [13, pp.154–155] posed the following natural

Problem 27 *Determine the best possible constant c such that*

$$d(\text{Eig}A,\ \text{Eig}B) \leqslant c\|A - B\|$$

for any normal matrices A, B *of any order.*

There are several other such constants whose exact values are not known [13].

22. Singular values of Heinz means

Let A, B be positive semidefinite matrices of order n. For $0 \leqslant t \leqslant 1$, the *Heinz mean* of A and B is

$$H_t(A, B) = (A^t B^{1-t} + A^{1-t} B^t)/2.$$

Note that $H_t = H_{1-t}$, and $H_0 = H_1$ is the arithmetic mean.

Denote the ordered singular values of an $n \times n$ complex matrix X by $s_1(X) \geqslant s_2(X) \geqslant \cdots \geqslant s_n(X)$. It was conjectured in [71] that

$$s_j(A^t B^{1-t} + A^{1-t} B^t) \leqslant s_j(A + B), \quad j = 1, \cdots, n$$

and this has recently been proved by Audenaert [5]. Now I have the following further

Problem 28 *Let* A, B *be positive semidefinite matrices of order* n. *For a fixed* j *with* $1 \leqslant j \leqslant n$, *define the function*

$$f(t) = s_j(A^t B^{1-t} + A^{1-t} B^t), \quad t \in [0,\ 1].$$

Determine the sub-intervals of $[0, 1]$ *on which* $f(t)$ *is monotone. In particular, determine the values of* t *at which* $f(t)$ *attains its minimum value.*

The above inequality says that $f(t)$ attains its maximum value at $t = 0, 1$. An example of A, B of order 3 is given in [5, p.280] such that for the second largest singular value, $f(t)$ does not attain its minimum value at $t = 1/2$.

Acknowledgment. I am grateful to Professors T. Ando, R. Bhatia, R.A. Brualdi, S. Friedland, F. Hiai, R.A. Horn, C.-K. Li, T.-Y. Tam for their kind suggestions about the possible content of this survey.

References

[1] S.S. Agaian, *Hadamard Matrices and Their Applications,* LNM 1168, Springer, 1985.

[2] T. Ando, Aluthge transforms and the convex hull of the eigenvalues of a matrix, *Linear and Multilinear Algebra,* 52(2004), 281–292.

[3] T. Ando, C.-K. Li and R. Mathias, Geometric means, *Linear Algebra Appl.,* 385(2004), 305–334.

[4] T. Ando and T. Yamazaki, The iterated Aluthge transforms of a 2-by-2 matrix converge, *Linear Algebra Appl.,* 375(2003), 299–309.

[5] K.M.R. Audenaert, A singular value inequality for Heinz means, *Linear Algebra Appl.,* 422(2007), 279–283.

[6] N. Bebiano, New developments on the Marcus-Oliveira conjecture, *Linear Algebra Appl.,* 197/198 (1994), 793–803.

[7] N. Bebiano, A. Kovacec and J. da Provideneia, The validity of the Marcus-de Oliveira conjecture for essentially Hermitian matrices, *Linear Algebra Appl.,* 197/198(1994), 411–427.

[8] N. Bebiano and J. da Provideneia, Some remarks on a conjecture of de Oliveira, *Linear Algebra Appl.,* 102(1988), 241–246.

[9] A. Berman and N. Shaked-Monderer, *Completely Positive Matrices,* World Scientific, 2003.

[10] D. Bessis, P. Moussa and M. Villani, Monotonic converging variational approximations to the functional integrals in quantum statistical mechanics, *J. Math. Phys.,* 16(1975), no.11, 2318–2325.

[11] R. Bhatia, Linear algebra to quantum cohomology: the story of Alfred Horn's inequalities, *Amer. Math. Monthly,* 108(2001), no.4, 289–318.

[12] R. Bhatia, *Positive Definite Matrices,* Princeton University Press, 2007.

[13] R. Bhatia, *Perturbation Bounds for Matrix Eigenvalues,* Reprint of the 1987 original with supplements, SIAM, Philadelphia, 2007.

[14] R. Bhatia, Spectral variation, normal matrices, and Finsler geometry, *Math. Intelligencer,* 29(2007), 41–46.

[15] R. Bhatia and J. Holbrook, Riemannian geometry and matrix geometric means, *Linear Algebra Appl.,* 413(2006), 594–618.

[16] R. Bhatia and J. Holbrook, Noncommutative geometric means, *Math. Intelligencer,* 28(2006), 32–39.

[17] R. Bhatia and F. Kittaneh, Notes on matrix arithmetic-geometric mean inequalities, *Linear Algebra Appl.,* 308(2000), 203–211.

[18] R. Bhatia and F. Kittaneh, On the singular values of a product of operators, *SIAM J. Matrix Anal. Appl.,* 11(1990), 272–277.

[19] A. Brauer and I.C. Gentry, On the characteristic roots of tournament matrices, *Bull. Amer. Math. Soc.,* 74(1968), 1133–1135.

[20] R.A. Brualdi and Q. Li, Problem 31, *Discrete Math.,* 43(1983), 329–330.

[21] R.A. Brualdi and B.L. Shader, *Matrices of Sign-Solvable Linear Systems*, Cambridge University Press, Cambridge, 1995.

[22] G.-S. Cheon and I.M. Wanless, An update on Minc's survey of open problems involving permanents, *Linear Algebra Appl.,* 403(2005), 314–342

[23] M. Cho, I.B. Jung and W.Y. Lee, On Aluthge transform of p-hyponormal operators, *Integral Equations Operator Theory,* 53(2005), 321–329.

[24] J.H. Drew and C.R. Johnson, The no long odd cycle theorem for completely positive matrices, in *Random Discrete Structures,* Springer, New York, 1996, pp. 103–115.

[25] J.H. Drew, C.R. Johnson, and R. Loewy, Completely positive matrices associated with M-matrices, *Linear and Multilinear Algebra,* 37(1994), 303–310.

[26] S.W. Drury and B. Cload, On the determinantal conjecture of Marcus and de Oliveira, *Linear Algebra Appl.,* 177(1992), 105–109.

[27] P.D. Egleston, T.D. Lenker and S.K. Narayan, The nonnegative inverse eigenvalue problem, *Linear Algebra Appl.,* 379(2004), 475–490.

[28] C. Eschenbach, F. Hall, R. Hemasinha, S.J. Kirkland, Z. Li, B.L. Shader, J.L. Stuart and J.R. Weaver, On almost regular tournament matrices, *Linear Algebra Appl.,* 306(2000), 103–121.

[29] M. Fiedler, Eigenvalues of nonnegative symmetric matrices, *Linear Algebra Appl.,* 9(1974), 119–142.

[30] M. Fiedler, Bounds for the determinant of the sum of two Hermitian matrices, *Proc. Amer. Math. Soc.,* 30(1971), 27–31.

[31] T.H. Foregger, On the minimum value of the permanent of a nearly decomposable doubly stochastic matrix, *Linear Algebra Appl.,* 32(1980), 75–85.

[32] W. Fulton, Eigenvalues, invariant factors, highest weights, and Schubert calculus, *Bull. Amer. Math. Soc. (N.S.),* 37(2000), no.3, 209–249.

[33] A.V. Geramita, J.M. Deramita and J.S. Wallis, Orthogonal designs, *Queen's Math. Preprint,* N 1973-37, 1976.

[34] A.V. Geramita and J. Seberry, *Orthogonal Designs,* Lecture Notes in Pure and Applied Mathematics, Vol. 45, Marcel Dekker, New York, 1979

[35] R. Grone and R. Merris, The Laplacian spectrum of a graph II, *SIAM J. Discrete Math.,* 7(1994), no.2, 221–229.

[36] F.J. Hall and Z. Li, Sign pattern matrices, Chapter 33 of the book, L. Hogben (editor), *Handbook of Linear Algebra,* CRC Press, 2006.

[37] M. Harwit and N.J.A. Sloane, *Hadamard Transform Optics,* Academic, New York, 1979.

[38] C.J. Hillar, Advances on the Bessis-Moussa-Villani trace conjecture, *Linear Algebra Appl.,* 426(2007), 130–142.

[39] H. Huang and T.-Y. Tam, On the convergence of Aluthge sequence, *Operators and Matrices,* 1(2007), 121–142.

[40] S.G. Hwang, On a conjecture of E. Dittert, *Linear Algebra Appl.,* 95(1987), 161–169.

[41] I.B. Jung, E. Ko and C. Pearcy, Aluthge transforms of operators, *Integral Equations Operator Theory,* 37(2000), 437–448.

[42] I.B. Jung, E. Ko and C. Pearcy, The iterated Aluthge transform of an operator, *Integral Equations Operator Theory,* 45(2003), 375–387.

[43] S. Kirkland, Perron vector bounds for a tournament matrix with applications to a conjecture of Brualdi and Li, *Linear Algebra Appl.,* 262(1997), 209–227.

[44] C.-K. Li and E. Poon, Additive decomposition of real matrices, *Linear and Multilinear Algebra,* 50(2002), 321–326.

[45] E.H. Lieb, Proofs of some conjectures on permanents, *J. Math. & Mech.,* 16(1966), 127–134.

[46] E.H. Lieb and R. Seiringer, Equivalent forms of the Bessis-Moussa-Villani conjecture, *J. Stat. Phys.,* 115(2004), 185–190.

[47] J.H. van Lint, The van der Waerden conjecture: two proofs in one year, *Math. Intelligencer,* 4(1982), 72–77.

[48] R. Loewy and D.D. London, A note on an inverse problem for nonnegative matrices, *Linear and Multilinear Algebra,* 6(1978), 83–90.

[49] R. Loewy and B.-S. Tam, CP rank of completely positive matrices of order 5, *Linear Algebra Appl.,* 363(2003), 161–176.

[50] M. Marcus, Derivations, Plücker relations and the numerical range, *Indiana Univ. Math. J.,* 22(1973), 1137–1149.

[51] J.E. Maxfield and H. Minc, On the matrix equation $XX' = A$, *Proc. Edinburgh Math. Soc.,* 13(II)(1962), 125–129.

[52] H. Minc, *Nonnegative Matrices,* John Wiley and Sons, New York, 1988.

[53] H. Minc, *Permanents,* Addison-Wesley, Reading, MA, 1978.

[54] H. Minc, Theory of permanents 1978-1981, *Linear and Multilinear Algebra,* 12(1983), 227–263.

[55] M. Moakher, A differential geometric approach to the geometric mean of symmetric positive definite matrices, *SIAM J. Matrix Anal. Appl.,* 26(2005), no.3, 735–747.

[56] G.N. de Oliveira, Normal matrices (research problem), *Linear and Multilinear Algebra,* 12(1982), 153–154.

[57] K. Sawada, A Hadamard matrix of order 268, *Graphs Combin.,* 1(1985), 185–187.

[58] J. Seberry and M. Yamada, Hadamard matrices, sequences, and block designs, in *Contemporary Design Theory,* eds. J.H. Dinitz and D.R. Stinson, Wiley, New York, 431–560.

[59] R. Sinkhorn, A problem related to the van der Waerden permanent theorem, *Linear and Multilinear Algebra,* 16(1984), 167-173.

[60] N.J.A. Sloane and M. Harwit, Masks for Hadamard transform optics, and weighing designs, *Appl. Optics,* 15(1976), 107–114.

[61] G. Soules, *Matrix functions and the Laplace expansion theorem,* PhD Dissertation, Univ. Calif. Santa Barbara, July 1966.

[62] K.R. Suleimanova, Stochastic matrices with real eigenvalues, Soviet Math Dokl., 66(1949), 343-345, (in Russian).

[63] J.S. Wallis, On the existence of Hadamard matrices, *J. Combin. Theory A,* 21(1976), 188–195.

[64] W.D. Wallis, A.P. Street and J.S. Wallis, *Combinatorics: Room Squares, Sum-Free Sets, Hadamard Matrices,* LNM 292, Springer, 1972.

[65] E.T.H. Wang, On permanents of $(1,-1)$-matrices, *Israel J. Math.,* 18(1974), 353–361.

[66] I.M. Wanless, Permanents of matrices of signed ones, *Linear and Multilinear Algebra,* 52(2005), no.6, 427–433.

[67] I.M. Wanless, On Minc's sixth conjecture, *Linear and Multilinear Algebra,* 55(2007), no.1, 57–63.

[68] E.W. Weisstein, Hadamard matrix, From *MathWorld-A Wolfram Web Resource,* http://mathworld.wolfram.com/HadamardMatrix.html

[69] X. Zhan, *Matrix Inequalities,* LNM 1790, Springer, 2002.

[70] X. Zhan, Extremal eigenvalues of real symmetric matrices with entries in an interval, *SIAM J. Matrix Anal. Appl.* 27(2006), no.3, 851–860.

[71] X. Zhan, Some research problems on the Hadamard product and singular values of matrices, *Linear and Multilinear Algebra,* 47(2000), 191–194.

参考文献

[1] R. Aharoni, *On a theorem of Dénes König*, Linear and Multilinear Algebra, 4 (1976), 31–32.

[2] T. Ando, *Totally positive matrices*, Linear Algebra Appl., 90 (1987), 165–219.

[3] T. Ando, *Majorization, doubly stochastic matrices and comparison of eigenvalues*, Linear Algebra Appl., 118 (1989), 163–248.

[4] T. Ando and R. Bhatia, *Eigenvalue inequalities associated with the Cartesian decomposition*, Linear and Multilinear Algebra, 22(1987), 133–147.

[5] T. Ando and X. Zhan, *Norm inequalities related to operator monotone functions*, Math. Ann., 315(1999), 771–780.

[6] H. Araki and S. Yamagami, *An inequality for the Hilbert-Schmidt norm*, Commun. Math. Phys., 81(1981), 89–98.

[7] E. Arrondo, *Another elementary proof of the Nullstellensatz*, Amer. Math. Monthly, 113 (2006), no.2, 169–171.

[8] K.M.R. Audenaert, *A singular value inequality for Heinz means*, Linear Algebra Appl., 422 (2007), 279–283.

[9] R. Bapat, *D_1AD_2 theorems for multidimensional matrices*, Linear Algebra Appl., 48 (1982), 437–442.

[10] R.B. Bapat and T.E.S. Raghavan, *Nonnegative Matrices and Applications*, Cambridge University Press, 1997.

[11] A. Barvinok, *A Course in Convexity*, The American Mathematical Society, 2002.

[12] L. Bassett, J. Maybee and J. Quirk, *Qualitative economics and the scope of the correspondence principle*, Econometrica, 36 (1968), 544–563.

[13] A. Berman and R.J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, SIAM, Philadelphia, 1994.

[14] A. Berman and B.D. Saunders, *Matrices with zero line sums and maximal rank*, Linear Algebra Appl., 40 (1981), 229–235.

[15] R. Bhatia, *Matrix Analysis*, GTM169, Springer, New York, 1997.

[16] R. Bhatia, *Pinching, trimming, truncating, and averaging of matrices*, Amer. Math. Monthly, 107(2000), 602–608.

[17] R. Bhatia, *Perturbation Bounds for Matrix Eigenvalues*, First Edition, Longman, Harlow, 1987; Second Edition, SIAM, Philadelphia, 2007.

[18] R. Bhatia, *Positive Definite Matrices*, Princeton University Press, 2007.

[19] R. Bhatia and C. Davis, *A bound for the spectral variation of a unitary operator*, Linear and Multilinear Algebra, 15(1984), 71–76.

[20] R. Bhatia and C. Davis, *More matrix forms of the arithmetic-geometric mean inequality*, SIAM J. Matrix Anal. Appl., 14 (1993), 132–136.

[21] R. Bhatia, C. Davis and A. McIntosh, *Perturbation of spectral subspaces and solution of linear operator equations*, Linear Algebra Appl., 52/53 (1983), 45–67.

[22] R. Bhatia, L. Elsner and G. Krause, *Bounds for the variation of the roots of a polynomial and the eigenvalues of a matrix*, Linear Algebra Appl., 142(1990), 195–209.

[23] R. Bhatia and F. Kittaneh, *On the singular values of a product of operators*, SIAM J. Matrix Anal. Appl., 11 (1990), 272–277.

[24] R. Bhatia and F. Kittaneh, *Cartesian decompositions and Schatten norms,* Linear Algebra Appl., 318(2000), 109–116.

[25] R. Bhatia and X. Zhan, *Compact operators whose real and imaginary parts are positive,* Proc. Amer. Math. Soc., 129(2001), no.8, 2277–2281.

[26] R. Bhatia and X. Zhan, *Norm inequalities for operators with positive real part,* J. Operator Theory, 50(2003), 67–76.

[27] R.A. Brualdi, P. Pei and X. Zhan, *An extremal sparsity property of the Jordan canonical form*, Linear Algebra Appl., 2008, to appear.

[28] R.A. Brualdi and H.J. Ryser, *Combinatorial Matrix Theory*, Cambridge University Press, 1991.

[29] R.A. Brualdi and B.L. Shader, *Matrices of Sign-Solvable Linear Systems*, Cambridge University Press, Cambridge, 1995.

[30] N.N. Chan and K.H. Li, *Diagonal elements and eigenvalues of a real symmetric matrix*, J. Math. Anal. Appl., 91 (1983), 562–566.

[31] N.G. de Bruijn and P. Erdös, *On a combinatorial problem*, Indag. Math., 10 (1948), 421–423.

[32] R. Demarr and A. Steger, *On elements with negative squares*, Proc. Amer. Math. Soc. 31 (1972), 57–60.

[33] F. Ding and X. Zhan, *On the unitary orbit of complex matrices*, SIAM J. Matrix Anal. Appl., 23(2001), no.2, 511–516.

[34] M.R. Embry, *A connection between commutativity and separation of spectra of linear operators*, Acta Sci. Math. (Szeged), 32(1971), 235–237.

[35] P. Erdős, A. Rényi and V. Sós, *On a problem of graph theory*, Studia Sci. Math., 1 (1966), 215–235.

[36] C.A. Eschenbach and C.R. Johnson, *Sign patterns that require real, nonreal or pure imaginary eigenvalues*, Linear and Multilinear Algebra, 29 (1991), no. 3–4, 299–311.

[37] C.A. Eschenbach and Z. Li, *How many negative entries can A^2 have?*, Linear Algebra Appl., 254 (1997), 99–117.

[38] S. Fallat, *Algebraic integers and the tensor product of matrices*, Crux Mathematicorum, 22 (1996), 341–343.

[39] S. Fallat, *Bidiagonal factorizations of totally nonnegative matrices*, Amer. Math. Monthly, 108 (2001), 697–712.

[40] S. Fallat, *A remark on oscillatory matrices*, Linear Algebra Appl., 393 (2004), 139–147.

[41] K. Fan and A.J. Hoffman, *Some metric inequalities in the space of matrices*, Proc. Amer. Math. Soc., 6 (1955), 111–116.

[42] M. Fang and A. Wang, *Possible numbers of positive entries of imprimitive nonnegative matrices*, Linear and Multilinear Algebra, 55 (2007), no.4, 399–404.

[43] M. Fiedler, *Special Matrices and Their Applications in Numerical Mathematics,* Martinus Nijhoff Publishers, Dordrecht, 1986.

[44] M. Fiedler and R. Grone, *Characterizations of sign patterns of inverse-positive matrices*, Linear Algebra Appl., 40 (1981), 237–245.

[45] H. Flanders and H.K. Wimmer, *On the matrix equation $AX-XB=C$ and $AX-YB=C$*, SIAM J. Appl. Math., 32 (1977), 707–710.

[46] S. Friedland, *Matrices with prescribed off-diagonal elements*, Israel J. Math., 11(1972), 184–189.

[47] F.R. Gantmacher, *The Theory of Matrices*, vol. I and II, Chelsea, New York, 1959.

[48] M. Gasca and C.A. Micchelli, *Total Positivity and its Applications*, Mathematics and its Applications, vol.359, Kluwer Academic, Dordrecht, 1996.

[49] M. Gasca and J.M. Pena, *Total positivity, QR-factorization and Neville elimination*, SIAM J. Matrix Anal. Appl., 14 (1993), 1132–1140.

[50] P.M. Gibson, *Conversion of the permanent into the determinant*, Proc. Amer. Math. Soc., 27 (1971), no.3, 471–476.

[51] F.J. Hall and Z. Li, *Sign pattern matrices*, Chapter 33 of the book, L. Hogben (editor), Handbook of Linear Algebra, CRC Press, 2006.

[52] E.V. Haynsworth, *Applications of an inequality for the Schur complement*, Proc. Amer. Math. Soc., 24 (1970), 512–516.

[53] D. Hershkowitz and H. Schneider, *Ranks of zero patterns and sign patterns*, Linear and Multilinear Algebra, 34 (1993), no.1, 3–19.

[54] A.J. Hoffman and H. Wielandt, *The variation of the spectrum of a normal matrix*, Duke Math. J., 20 (1953), 37–39.

[55] A. Horn, *On the singular values of a product of completely continuous operators*, Proc.

National Acad. Sciences (U.S.) 36 (1950), 374–375.

[56] A. Horn, *Doubly stochastic matrices and the diagonal of a rotation matrix*, Amer. J. Math., 76 (1954), 620–630.

[57] A. Horn, *On the eigenvalues of a matrix with prescribed singular values*, Proc. Amer. Math. Soc., 5 (1954), 4–7.

[58] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

[59] R.A. Horn and C.R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, 1991.

[60] D.J. Houck and M.E. Paul, *On a theorem of de Bruijn and Erdös*, Linear Algebra Appl., 23 (1979), 157–165.

[61] Q. Hu, Y. Li and X. Zhan, *Possible numbers of ones in 0-1 matrices with a given rank,* Linear and Multilinear Algebra, 53(2005), no.6, 435–443.

[62] C. Huneke, *The friendship theorem*, Amer. Math. Monthly, 109 (2002), no.2, 192–194.

[63] I.M. Isaacs, *Algebra: A Graduate Course*, Wadsworth Inc., 1994.

[64] C. Jeffries, V. Klee and P. vanden Driessche, *When is a matrix sign stable?*, Canad. J. Math., 29 (1977), 315–326.

[65] C.R. Johnson, *Sign patterns of inverse nonnegative matrices*, Linear Algebra Appl., 55 (1983), 69–80.

[66] C.R. Johnson, F.T. Leighton and H.A. Robinson, *Sign patterns of inverse-positive matrices*, Linear Algebra Appl., 24 (1979), 75–83.

[67] C.R. Johnson and C.-K. Li, *Inequalities relating unitarily invariant norms and the numerical radius*, Linear and Multilinear Algebra, 23 (1988), 183–191.

[68] D. Kalman, R. Mena and S. Shahriari, *Variations on an irrational theme-geometry, dynamics, algebra*, Math. Mag. 70 (1997), no.2, 93–104.

[69] S. Karlin, *Total Positivity*, vol.I, Stanford University Press, Stanford, 1968.

[70] F. Kittaneh, *On Lipschitz functions of normal operators*, Proc. Amer. Math. Soc., 94(1985), 416–418.

[71] F. Kittaneh, *Inequalities for the Schatten p-norm IV*, Commun. Math. Phys., 106(1986), 581–585.

[72] F. Kittaneh, *Singular values of companion matrices and bounds on zeros of polynomials*, SIAM J. Matrix Anal. Appl., 16 (1995), no.1, 333–340.

[73] V. Klee, R. Ladner and R. Manber, *Signsolvability revisited*, Linear Algebra Appl., 59 (1984), 131–157.

[74] K. Lancaster, *The scope of qualitative economics*, Rev. Econ. Studies, 29 (1962), 99–132.

[75] P. Lancaster and M. Tismenetsky, *The Theory of Matrices,* Second edition, Academic Press, 1985.

[76] C.-K. Li and E. Poon, *Additive decomposition of real matrices*, Linear and Multilinear Algebra, 50(2002), 321–326.

[77] C.-K. Li and D. Lutzer, *The arithmetic of algebraic numbers: an elementary approach*, College Math. J., 35 (2004), 307–309.

[78] 李乔, 矩阵论八讲, 上海科技出版社, 1988.

[79] R.C. Li, *New perturbation bounds for the unitary polar factor*, SIAM J. Matrix Anal. Appl., 16(1995), 327–332.

[80] Z. Li and L. Harris, *Sign patterns that require all distinct eigenvalues*, JP J. Alg. Num. Theory Appl., 2 (2002), 161–179.

[81] V.B. Lidskii, *On the proper values of a sum and product of symmetric matrices,* Dokl. Akad Nauk SSSR., 75(1950), 769–772.

[82] M. Marcus and H. Minc, *On two theorems of Frobenius*, Pacific J. Math., 60 (1975), 149–151.

[83] M. Marcus and R. Ree, *Diagonals of doubly stochastic matrices*, Quart. J. Math. Oxford. Ser.(2), 10(1959), 296–302.

[84] M. Marcus and M. Sandy, *Singular values and numerical radii*, Linear and Multilinear Algebra, 18 (1985), 337–353.

[85] A.W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, 1979.

[86] J.S. Maybee, *Matrices of class* J_2, J. of Research of the National Bureau Standards, 71B(1967), 215–224.

[87] H. Minc, *Nonnegative Matrices*, John Wiley and Sons, New York, 1988.

[88] L. Mirsky, *Symmetric gauge functions and unitarily invariant norms*, Quart. J. Math. Oxford ser. (2), 11 (1960), 50–59.

[89] R. Oldenburger, *Infinite powers of matrices and characteristic roots*, Duke Math. J., 6 (1940), 357–361.

[90] A.M. Ostrowski, *Solution of Equations in Euclidean and Banach Spaces*, 3rd ed., Academic, New York, 1973.

[91] D. Phillips, *Improving spectral-variation bounds with Chebyshev polynomials*, Linear Algebra Appl., 133(1990), 165–173.

[92] J.F. Queiro and A.L. Duarte, *On the Cartesian decomposition of a matrix,* Linear and Multilinear Algebra, 18(1985), 77–85.

[93] J. Quirk and R. Ruppert, *Qualitative economics and the stability of equilibrium*, Rev. Economic Studies, 32 (1965), 311–326.

[94] R. Rado, *An inequality*, J. London Math. Soc. 27 (1952), 1–6.

[95] T.J. Rivlin, *An Introduction to the Approximation of Functions*, Dover, New York, 1981.

[96] J.A. Rosoff, *A topological proof of the Cayley-Hamilton theorem,* Missouri J. Math. Sci., 7(1995), no.2, 63–67.

[97] W. Roth, *The equations* $AX - YB = C$ *and* $AX - XB = C$ *in Matrices*, Proc. Amer. Math. Soc., 3 (1952), 392–396.

[98] U. Rothblum and C. Tan, *Upper bounds on the maximum modulus of subdominant eigenvalues of nonnegative matrices*, Linear Algebra Appl., 66 (1985), 45–86.

[99] P.A. Samuelson, *Foundations of Economic Analysis*, Harvard University Press, Cambridge, MA, 1947.

[100] I. Sedlacek, *O incidencnich maticich orientovych grafu*, Casop. Pest. Mat. 84 (1959), 303–316.

[101] J.-Y. Shao, *Matrices permutation equivalent to primitive matrices*, Linear Algebra Appl., 65(1985), 225–247.

[102] J.-Y. Shao, *The exponent set of symmetric primitive matrices*, Scientia Sinica (Ser. A), 30(1987), no.4, 348–358.

[103] R. Sinkhorn, *A relationship between arbitrary positive matrices and doubly stochastic matrices*, Ann. Math. Statist, 35 (1964), 876–879.

[104] G.W. Stewart and J.G. Sun, *Matrix Perturbation Theory,* Academic Press, 1990.

[105] J.G. Sun, *On the variation of the spectrum of a normal matrix*, Linear Algebra Appl., 246(1996), 215–223.

[106] Y. Tao, *More results on singular value inequalities of matrices*, Linear Algebra Appl., 416 (2006), 724–729.

[107] C. Thomassen, *Sign-nonsingular matrices and even cycles in directed graphs*, Linear Algebra Appl., 75 (1986), 27–41.

[108] R.C. Thompson, *Convex and concave functions of singular values of matrix sums*, Pacific J. Math., 66 (1976), 285–290.

[109] B. Wang and M. Gong, *Some eigenvalue inequalities for positive semidefinite matrix power products*, Linear Algebra Appl., 184 (1993), 249–260.

[110] B. Wang and F. Zhang, *Schur complements and matrix inequalities of Hadamard products*, Linear and Multilinear Algebra, 43 (1997), 315–326.

[111] R.J. Webster, *The hamonic means of diagonals of doubly stochastic matrices*, Linear and Multilinear Algebra 13 (1983), no.4, 367–369.

[112] H. Wielandt, *Unzerlegbare nicht-negative matrizen*, Math. Z. 52 (1950), 642–648.

[113] X. Zhan, *Norm inequalities for Cartesian decompositions*, Linear Algebra Appl., 286(1999), 297–301.

[114] X. Zhan, *Inequalities involving Hadamard products and unitarily invariant norms*, Adv. in Math. (China), 27 (1998), 416–422.

[115] X. Zhan, *Singular values of differences of positive semidefinite matrices*, SIAM J. Matrix Anal. Appl., 22 (2000), 819–823.

[116] X. Zhan, *Some research problems on the Hadamard product and singular values of matrices*, Linear and Multilinear Algebra, 47 (2000), 191–194.

[117] X. Zhan, *Linear preservers that permute the entries of a matrix,* Amer. Math. Monthly, 108(2001), no.7, 643–645.

[118] X. Zhan, *Matrix Inequalities*, LNM 1790, Springer, Berlin, 2002.

[119] X. Zhan, *The Sharp Rado theorem for majorizations*, Amer. Math. Monthly, 110 (2003), 152–153.

[120] X. Zhan, *On some matrix inequalities*, Linear Algebra Appl., 376 (2004), 299–303.

[121] X. Zhan, *Extremal eigenvalues of real symmetric matrices with entries in an interval*, SIAM J. Matrix Anal. Appl., 27(2006), no.3, 851–860.

[122] X. Zhan, *Extremal numbers of positive entries of imprimitive nonnegative matrices*, Linear Algebra Appl., 424 (2007), 132–138.

[123] X. Zhan, *Open problems in matrix theory,* in Proceedings of the 4th International Congress of Chinese Mathematicians, Vol. I, edited by L. Ji, K. Liu, L. Yang and S.-T. Yau, Higher Education Press, Beijing, 2008, 367–382.

[124] F. Zhang, *Matrix Theory: Basic Results and Techniques*, Springer, New York, 1999.

[125] K.M. Zhang, *On Lewin and Vitek's conjecture about the exponent set of primitive matrices*, Linear Algebra Appl., 96 (1987), 101–108.

名词索引