

研究生教学用书
公共基础课系列

矩 阵 论

(第二版)

杨 明 刘先忠

华中科技大学出版社

图书在版编目(CIP)数据

矩阵论(第二版)/杨明 刘先忠
武汉:华中科技大学出版社, 2005年月
ISBN 7-5609--

I. 矩…
II. 杨…
III.
IV.

矩阵论(第二版)

杨 明 刘先忠

责任编辑:李 德
责任校对:刘 飞

封面设计:潘 群
责任监印:

出版发行:华中科技大学出版社
武昌喻家山 邮编:430074 电话:(027)87545012

经 销:新华书店湖北发行所

录 排:华大数码设计室
印 刷:

开本: × 1/16 印张: 字数:
版次:2005年月第1版 印次:2005年月第1次印刷 印数:1—5 000
ISBN 7-5609-- 定价:元

(本书若有印装质量问题,请向出版社发行部调换)

内 容 简 介

本教材适用于工学硕士和工程硕士研究生数学基础课——矩阵论。全书共分 7 章,主要内容为线性空间与线性变换、Jordan 标准形、矩阵分解、矩阵的广义逆、矩阵分析、矩阵的 Kronecker 积与 Hadamard 积和非负矩阵介绍。为工学硕士研究生的应用研究提供所需的数学工具。为他们的继续学习提供必需的数学基础。

本书适用于 50 学时左右的矩阵论课程的教学使用,也可作为同类课程的教学参考书。

Abstract

Matrix Theory is an important component of postgraduate mathematics, particularly for postgraduate students majoring in the scientific and engineering. This book is a text book for MA program course—matrix theory. The topics been covered in the seven chapters are general vector space and linear transformations, Jordan Canonical matrix of matrices. matrix decomposition, generalized inverse of matrix, matrix analyses, the Kronecker product and Hadamard product and introduction to nonnegative matrices. The text contains all the topics recommended by the Ministry of Education.

This book is suitable for 50 lectures.

前　　言

本书是为工学硕士研究生数学基础课“矩阵论”编写的教材。全书由线性空间和线性变换、Jordan 标准形、矩阵分解、矩阵的广义逆、矩阵分析和非负矩阵介绍等6 章构成，其内容符合教育部工学硕士研究生矩阵论教学基本要求。

在科学技术和工程应用中，矩阵理论的重要性和应用的广泛性是众所周知的。计算机的广泛使用和 MATLAB, MAPLE 等数学计算软件的迅猛普及为矩阵理论提供了更为广阔的发展和应用前景。本书注重将线性空间、线性变换和内积、赋范空间的各种问题对应到数值向量空间 \mathbb{R}^n 上，强调抽象内容的矩阵处理技巧，使问题的描述形式和处理方法简洁，可有效地利用矩阵这一数学工具，同时为将功能强大的矩阵计算数学软件有效地使用到各类问题上奠定了基础，便于提高教学效率。本书的内容取舍力求理论体系简明清晰，适合工学硕士后续学习和研究应用的需要，深浅适度。本书的推荐教学学时为 50 学时。

本书是在华中科技大学(原华中理工大学)数学系为硕士研究生开设矩阵论课程的长期教学实践中发展形成的。第 1、2、3 章由杨明编写，第 4、5、6 章由刘先忠编写。本书的编写得到了华中科技大学研究生院的教改项目支持，得到了华中科技大学数学系和课程教学组老师们的关心和协助，在此表示衷心的感谢。作者对华中科技大学出版社和本书编辑为本书出版所做的工作表示感谢。

本书内容虽经多次讲授，随着科学技术的发展仍然需要不断的完善，其问题和不足，敬请同行和读者不吝指正。

作　　者

2003 年 9 月于华中科技大学

第二版前言

本书的第二版是在第一版的基础上,根据教学实践和工科硕士研究生学习和应用中对矩阵理论的需要,从便于研究生阅读自学的角度做了一些扩充和修改而成.

本版全书的内容为7章:线性空间和线性变换、Jordan标准形、矩阵分解、矩阵的广义逆、矩阵分析、矩阵的 Kronecker 积和 Hadamard 积、非负矩阵介绍. 其中第1、2、3、6章由杨明编写,第4、5、7章由刘先忠编写.

本版的扩充主要是在第1章中增加了从 m 维线性空间 $V_m(F)$ 到 n 维线性空间 $V_n(F)$ 的线性变换; 在第3章和第5章中分别增加了极分解和数值范围等部分内容; 增写了第6章: 矩阵的 Kronecker 积和 Hadamard 积. 在各章节增补了部分内容、例题以介绍矩阵理论在一些特殊矩阵上的研究成果.

本版对矩阵的奇异值分解根据现代应用的发展做了较大的修改。另外,重新修订了原书的部分证明和习题.

本书适用于50学时左右的教学,可以根据学生的线性代数基础做相应的内容取舍. 其中部分内容可供研究生自学和应用查阅.

本书第一版在应用过程中,得到了华中科技大学数学系课程教学组教师、研究生院教学顾问组余明书教授和修矩阵论课程的研究生的宝贵意见和建议,在本版中,作者考虑了这些意见和建议,在此对他们表示衷心地感谢. 同时作者对华中科技大学出版社编辑们为本书出版所做的工作表示感谢.

本书虽经多次修改和讲授,但不足之处一定不少,尚祈读者不吝指正.

作 者

2004年11月于华中科技大学

目 录

第1章 线性空间与线性变换	(1)
1.1 线性空间	(1)
一、线性空间的概念	(1)
二、线性空间的基与维数	(2)
三、坐标	(3)
四、基变换与坐标变换	(5)
五、子空间	(8)
1.2 内积空间	(13)
一、欧氏空间与酉空间	(13)
二、标准正交基	(16)
1.3 线性变换	(19)
一、线性变换	(19)
二、线性变换的矩阵	(21)
三、不变子空间	(24)
四、正交变换与酉变换	(25)
五、线性空间 $V_n(F)$ 到线性空间 $V_m(F)$ 的线性变换	(28)
习题一	(31)
第2章 Jordan 标准形介绍	(35)
2.1 线性变换的对角矩阵表示	(35)
一、线性变换的特征值与特征向量	(35)
二、线性变换矩阵的对角化	(38)
2.2 Jordan 矩阵介绍	(40)
一、Jordan 矩阵	(40)
二、Jordan 标准形的求法	(42)
2.3 最小多项式	(48)
一、矩阵多项式	(49)
二、方阵的化零多项式	(52)
三、最小多项式	(53)
习题二	(57)
第3章 矩阵的分解	(60)

3.1 常见的矩阵标准形与分解.....	(60)
一、矩阵的三角分解.....	(61)
二、矩阵的满秩分解.....	(66)
三、可对角化矩阵的谱分解.....	(70)
3.2 Schur 分解与正规矩阵	(74)
一、Schur 分解	(74)
二、正规矩阵.....	(77)
3.3 矩阵的奇异值分解.....	(81)
一、矩阵的奇异值及其性质.....	(82)
二、矩阵的奇异值分解.....	(83)
三、矩阵 A 的奇异值分解与线性变换 T_A	(88)
四、方阵的极分解.....	(89)
习题三	(91)
第4章 矩阵的广义逆	(93)
4.1 矩阵的左逆与右逆.....	(93)
一、满秩矩阵与单侧逆.....	(93)
二、单侧逆与解线性方程组.....	(94)
4.2 广义逆矩阵.....	(95)
一、减号广义逆.....	(95)
二、Moore-Penrose 广义逆(加号广义逆)	(98)
4.3 投影变换	(101)
一、投影变换与投影矩阵	(101)
二、正交投影变换与正交投影矩阵	(103)
4.4 最佳的最小二乘解	(105)
习题四	(108)
第5章 矩阵分析	(109)
5.1 向量范数	(109)
一、向量范数的概念	(109)
二、向量范数的连续性与等价性	(110)
5.2 矩阵范数	(111)
一、矩阵范数的概念	(112)
二、诱导范数	(114)
5.3 向量序列和矩阵序列的极限	(115)
5.4 矩阵幂级数	(118)
一、谱半径	(118)

二、数值范围	(120)
三、矩阵幂级数	(122)
5.5 矩阵函数	(125)
一、矩阵函数的定义与性质	(125)
二、矩阵函数的求法	(127)
5.6 函数矩阵的微分与积分	(130)
5.7 矩阵函数的应用	(131)
一、一阶线性常系数齐次微分方程组	(132)
二、一阶线性常系数非齐次微分方程组	(133)
习题五	(134)
第6章 矩阵的 Kronecker 积与 Hadamard 积	(136)
6.1 Kronecker 积与 Hadamard 积的定义	(136)
6.2 Kronecker 积与 Hadamard 积的性质	(140)
6.3 矩阵的向量化算子与 Kronecker 积	(145)
习题六	(150)
*第7章 非负矩阵介绍	(151)
7.1 非负矩阵	(151)
7.2 正矩阵	(154)
7.3 素矩阵	(158)
7.4 M 矩阵	(160)
习题答案与提示	(163)
参考书目	(172)

本书使用符号一览表

$V_n(F)$	数域 F 上的 n 维线性空间
F^n	以数域 F 中的数为分量的 n 维列向量集合
$F^{m \times n}$	以数域 F 中的数为元素的 $m \times n$ 阶矩阵集合
$P_n[x]$	次数不超过 $n-1$ 次的实系数多项式集合
$L\{\alpha_1, \alpha_2, \dots, \alpha_n\}$	由线性空间中向量 $\alpha_1, \alpha_2, \dots, \alpha_n$ 生成的子空间
W^\perp	W 的正交补子空间
$W_1 \oplus W_2$	子空间 W_1 与 W_2 的直和
$ k $	复数 k 的模长
$\ \alpha\ $	向量 α 的长度(范数): $\ \alpha\ ^2 = \sum_{i=1}^n \alpha_i ^2$
e_i	第 i 个分量为 1, 其余分量为 0 的基本单位向量
$[A]_{ij}$	矩阵的第 i 行, 第 j 列位置上的元素或子块
E_{ij}	第 i 行, 第 j 列位置上元素为 1, 其余元素为 0 的矩阵
A^H	矩阵 A 的共轭转置矩阵, $A^H = (\bar{A})^T$
$\det(A)$	矩阵 A 的行列式
$\text{rank}(A)$	矩阵 A 的秩
$R(A)$	矩阵 A 的列空间
$N(A)$	矩阵 A 的零空间
$\text{tr}(A)$	矩阵 A 的迹 $\text{tr}(A) = \sum_{i=1}^n a_{ii}$
V_λ	特征值 λ 对应的特征子空间
$m_A(\lambda)$	矩阵 A 的最小多项式
A_R^{-1}	矩阵 A 的右逆
A_L^{-1}	矩阵 A 的左逆
$A\{1\}$	矩阵 A 的减号逆或 $\{1\}$ 逆
A^+	矩阵 A 的 Moore-Penrose 广义逆
$\rho(A)$	矩阵 A 的谱半径
$\ X\ _p$	向量 X 的 p 范数 $\ X\ _p = (\sum_i x_i ^p)^{\frac{1}{p}}$
$\ A\ _F$	矩阵 A 的 Frobenius 范数:

$$\|A\|_F = \left[\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \right]^{\frac{1}{2}} = [\text{tr}(A^H A)]^{\frac{1}{2}}$$

$\text{diag}(J_1, J_2, \dots, J_m)$ 以 J_1, J_2, \dots, J_m 为对角线上子块的对角阵

$A \otimes B$ 矩阵 A 与 B 的 Kronecker 积

$A \circ B$ 矩阵 A 与 B 的 Hadamard 积

Vec 向量化算子

$A > 0$ A 为正矩阵

$A \geq 0$ A 为非负矩阵

第1章 线性空间与线性变换

矩阵是处理有限维空间形式和数量关系的重要工具,线性空间与线性变换是其中的基本研究对象.这一章在线性代数的基础上,推广向量空间 \mathbb{R}^n ,一般地建立线性空间的概念,进而定义线性变换,介绍其中的基本理论及其矩阵方法.

1.1 线性空间

我们从线性代数中了解的向量空间 \mathbb{R}^n ,从数学的角度看,它涉及一个集合和一个数域,定义有集合中元素的加法运算和数域中数与集合中元素的数乘运算,我们由此出发来推广建立线性空间的概念.

一、线性空间的概念

定义 1.1 设 V 是一个以 $\alpha, \beta, \gamma, \dots$ 为元素的非空集合, F 是一个数域. 在其中定义两种运算,一种叫加法: $\forall \alpha, \beta \in V, \alpha + \beta \in V$; 另一种叫数量乘法: $\forall k \in F, \alpha \in V, k\alpha \in V$, 并且满足下面八条运算法则:

- (1) 加法交换律: $\alpha + \beta = \beta + \alpha$;
- (2) 加法结合律: $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$;
- (3) V 中存在零元素: $\exists \alpha_0 \in V, \forall \alpha \in V, \alpha + \alpha_0 = \alpha$, 记 $\alpha_0 = 0$;
- (4) 负元素存在: $\forall \alpha \in V, \exists \beta \in V$, 使 $\alpha + \beta = 0$, 记 $\beta = -\alpha$;
- (5) 数乘结合律: $(kl)\alpha = k(l\alpha)$;
- (6) 存在 $1 \in F, 1 \cdot \alpha = \alpha$;
- (7) 分配律: $(k+l)\alpha = k\alpha + l\alpha$;
- (8) 分配律: $k(\alpha + \beta) = k\alpha + k\beta$,

则称 V 为数域 F 上的线性空间. V 中元素称为向量. F 为实(复)数域时, 称 V 是实(复)线性空间.

例 1 对给定的数域 F , 集合 $F^n = \{(x_1 \quad x_2 \quad \cdots \quad x_n)^T \mid x_i \in F\}$, 对通常的向量的加法和数乘运算, F^n 为域 F 上的线性空间. 当 F 为实数域 \mathbb{R} 和复数域 \mathbb{C} 时, \mathbb{R}^n 和 \mathbb{C}^n 是它的两种具体形式.

例 2 $V = F^{m \times n} = \{A = (a_{ij})_{m \times n} \mid a_{ij} \in F\}$, 它在矩阵的加法与数乘运算下构成数域 F 上的线性空间, 称为矩阵空间, 其中 $\mathbb{R}^{n \times n}$ 为由一切 $m \times n$ 实矩阵构成的实矩阵空间.

例 3 实数域 \mathbb{R} 上次数不超过 $n-1$ 次的关于文字 x 的一切多项式和零多项式所构

成的集合

$$P_n[x] = \left\{ \sum_{i=0}^{n-1} a_i x^i \mid a_i \in \mathbb{R} \right\},$$

在通常多项式加法与数乘多项式运算下构成线性空间,称为多项式空间 $P_n[x]$.

值得指出的是次数等于 $n-1$ 的多项式集合 $V = \left\{ \sum_{i=0}^{n-1} a_i x^i \mid a_i \in \mathbb{R}, a_{n-1} \neq 0 \right\}$ 不是线性空间,因为两个次数为 $n-1$ 的多项式相加后所得多项式次数不一定为 $n-1$,从而有可能不在 V 中,这也称为 V 对加法不封闭. 定义 1 要求线性空间对加法和数量乘法这两种运算均是封闭的.

例 4 $V = C[a, b] = \{f(x) \mid f(x) \text{ 是区间 } [a, b] \text{ 上实连续函数}\}$,对于函数的加法与数乘运算构成实数域上的线性空间.

从上述线性空间例子中可以看到,许多常见的研究对象都可以在线性空间中作为向量来研究. 另外应理解加法和数乘分别是 V 中的一个二元运算和数域 F 和 V 中元素间的运算,要求运算满足定义 1 中的八条性质,它们已不再局限在数的加法、乘法的概念中.

例如取集合为正数集合 \mathbb{R}^+ , F 为实数域 \mathbb{R} , 加法 \oplus 和数乘. 如下定义

$$\begin{aligned} \oplus: \forall a, b \in \mathbb{R}^+ \quad a \oplus b &= ab, \\ \circ: \forall k \in \mathbb{R}, a \in \mathbb{R}^+ \quad k \circ a &= a^k. \end{aligned}$$

在此运算下, \mathbb{R}^+ 是 \mathbb{R} 上的一个线性空间. 其中加法零元素是 \mathbb{R}^+ 中的数 1, \mathbb{R}^+ 中元素 a 的负元素是 a^{-1} .

定理 1.1 线性空间 V 有如下性质:

- (1) V 中的零元素惟一;
- (2) V 中任一元素的负元素惟一;
- (3) 设 0 为数零, 0 为 V 中零向量, 则
 - (i) $0 \cdot \alpha = 0$.
 - (ii) $k \cdot 0 = 0, k \in F$.
 - (iii) 若 $k \cdot \alpha = 0$, 则一定有 $k = 0$ 或 $\alpha = 0$.
 - (iv) $(-1)\alpha = -\alpha$.

证明留给读者.

二、线性空间的基与维数

线性空间中的基与维数是依赖于向量的线性相关与线性无关的概念来定义的. 线性空间 V 作为一个向量集合,其中向量的线性相关与线性无关的定义与线性代数中给出的定义完全类似,因而在线性空间中,极大线性无关组、等价等概念在形式上与向量空间 \mathbb{R}^n 中的定义一样. 与上述概念相关的性质与结果也可平移到线性空间中. 在此不

再叙述.

定义 1.2 设 V 是线性空间, 若存在一组线性无关的向量 $\alpha_1, \alpha_2, \dots, \alpha_n$, 使空间中任一向量可由它们线性表示, 则称向量组 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 为 V 的一组基. 基所含向量个数为 V 的维数, 记为 $\dim V = n$, $n < +\infty$ 或者 $n = +\infty$.

例 5 向量组 $\{e_1 = (1 \ 0 \ 0 \ \cdots \ 0)^T, e_2 = (0 \ 1 \ 0 \ \cdots \ 0)^T, \dots, e_n = (0 \ 0 \ \cdots \ 0 \ 1)^T\}$ 是 F^n 的一组基, 所以 $\dim F^n = n$.

例 6 求矩阵空间 $R^{2 \times 2}$ 的维数与一组基.

解 任取矩阵 $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, 有

$$A = a_{11} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + a_{12} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + a_{21} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + a_{22} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

因此 $R^{2 \times 2}$ 中任何一个向量都可写成向量组

$$E_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad E_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad E_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad E_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

的线性组合. 又任取数 k_i , 由

$$k_1 E_{11} + k_2 E_{12} + k_3 E_{21} + k_4 E_{22} = \begin{pmatrix} k_1 & k_2 \\ k_3 & k_4 \end{pmatrix} = 0,$$

得 $k_i = 0$, $i = 1, 2, 3, 4$, 故 $E_{11}, E_{12}, E_{21}, E_{22}$ 线性无关. 由定义 2 知 $\{E_{11}, E_{12}, E_{21}, E_{22}\}$ 是 $R^{2 \times 2}$ 的一组基, $\dim R^{2 \times 2} = 4$.

类似地, $\{E_{ij}, i = 1, 2, \dots, m; j = 1, 2, \dots, n\}$ 是矩阵空间 $R^{m \times n}$ 的一组基,

$$\dim R^{m \times n} = m \times n.$$

例 7 向量组 $\{1, x, x^2, \dots, x^{n-1}\}$ 是线性空间 $P_n[x]$ 的一组基, $\dim P_n[x] = n$.

以上例子中, 空间维数 $\dim V$ 都为有限数, 这样的空间称为有限维线性空间. 若 $\dim V$ 不是有限数, 则称 V 为无限维线性空间. 由于有限维空间与无限维空间在研究方法上的较大差异, 这里我们只讨论有限维空间. 约定记号 $V_n(F)$ 表示 V 是数域 F 上的 n 维线性空间.

由于基就是向量集合 V 的极大线性无关组, 从而线性空间的基也不是惟一的.

定理 1.2 n 维线性空间中任意 n 个线性无关的向量构成的向量组都是空间的基.

三、坐标

在线性空间 $V_n(F)$ 中, 设 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 是一组基, 则 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 线性无关, $\forall \beta \in V_n(F)$, $\{\alpha_1, \alpha_2, \dots, \alpha_n, \beta\}$ 线性相关, 故 β 可由 $\alpha_1, \alpha_2, \dots, \alpha_n$ 惟一地线性表示. 因此有

定义 1.3 设 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是线性空间 $V_n(F)$ 的一组基, $\forall \beta \in V$,

$$\beta = \sum_{i=1}^n x_i \alpha_i = (\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad (1.1)$$

则称数 x_1, x_2, \dots, x_n 是 β 在基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 下的坐标, (1.1) 式中向量 $(x_1 \quad x_2 \quad \cdots \quad x_n)^T$ 为 β 的坐标向量, 也简称为坐标.

(1.1) 式中第二个等式是借助于矩阵的形式运算来表示的, 显然 $(\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n)$ 在一般的向量意义下不一定为矩阵, 但这种表示会给今后矩阵处理带来很多便利.

例 8 求 $R^{2 \times 2}$ 中向量 $\begin{pmatrix} 3 & 1 \\ 4 & 5 \end{pmatrix}$ 在基 $\{E_{11}, E_{12}, E_{21}, E_{22}\}$ 下的坐标.

解 $\begin{pmatrix} 3 & 1 \\ 4 & 5 \end{pmatrix} = 3E_{11} + E_{12} + 4E_{21} + 5E_{22} = (E_{11} \quad E_{12} \quad E_{21} \quad E_{22}) \begin{pmatrix} 3 \\ 1 \\ 4 \\ 5 \end{pmatrix}$, 故该向量在所给基下

坐标为 $(3 \ 1 \ 4 \ 5)^T$. 一般地 $R^{2 \times 2}$ 中向量 $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ 在所给基 $\{E_{ij}\}$ 下坐标为 $(a_{11} \quad a_{12} \quad a_{21} \quad a_{22})^T$.

例 9 已知 $\{1, x, x^2, x^3\}$ 和 $\{1, (x-1), (x-1)^2, (x-1)^3\}$ 是线性空间 $P_4[x]$ 的两组基, 求 $P_4[x]$ 中向量 $f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$ 在这两组基下的坐标.

解 因为 $f(x) = (1 \quad x \quad x^2 \quad x^3) \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix}$,

所以 $f(x)$ 在基 $\{1, x, x^2, x^3\}$ 下坐标为 $(a_0 \quad a_1 \quad a_2 \quad a_3)^T$.

又由 Taylor 公式

$$\begin{aligned} f(x) &= f(1) + f'(1)(x - 1) + \frac{f''(1)}{2}(x - 1)^2 + \frac{f'''(1)}{3!}(x - 1)^3 \\ &= (1 \ (x - 1) \ (x - 1)^2 \ (x - 1)^3) \begin{pmatrix} f(1) \\ f'(1) \\ \frac{1}{2} f''(1) \\ \frac{1}{3!} f'''(1) \end{pmatrix}, \end{aligned}$$

故 $f(x)$ 在基 $\{1, (x-1), (x-1)^2, (x-1)^3\}$ 下的坐标为 $\left[f(1) \quad f'(1) \quad \frac{f''(1)}{2!} \quad \frac{f'''(1)}{3!} \right]^T$.

从以上例子和(1.1)式可以看到,不论 $V_n(F)$ 为何种具体的线性空间,当在 $V_n(F)$ 中取定一组基时, $V_n(F)$ 中向量在该基下的坐标都是线性空间 F^n 中的向量.正是这一特点,奠定了可以用数量矩阵和 R^n 中向量来研究一般的线性空间中有关问题的基础.一般同一个向量在不同基下的坐标是不同的.

在线性空间 $V_n(F)$ 中取定一组基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$, $\forall \beta \in V_n(F)$, 取坐标作为对应关系, β 惟一地对应于 F^n 中一个向量 X (β 的坐标).反之, $\forall X \in F^n$, $(\alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_n)X$ 就是 X 所对应的 $V_n(F)$ 中的向量.因此,坐标关系建立了线性空间 $V_n(F)$ 和 F^n 的一一对应关系 σ ,显然 σ 满足

$$\begin{aligned}\sigma(\alpha + \beta) &= \sigma(\alpha) + \sigma(\beta), \\ \sigma(k\alpha) &= k\sigma(\alpha).\end{aligned}$$

在对应关系 σ 下,数域 F 上任何一个 n 维线性空间 $V_n(F)$ 都和 n 维线性空间 F^n 同构.

设 $V_n(F)$ 中向量 β_i 的坐标为 X_i , $i = 1, 2, \dots, m$, 则 β_i 的线性组合 $\sum_{i=1}^m k_i \beta_i$ 的坐标是 $\sum_{i=1}^m k_i X_i$, 又零向量坐标为0, 所以

$$\sum_{i=1}^m k_i \beta_i = 0 \Leftrightarrow \sum_{i=1}^m k_i X_i = 0.$$

该结果可叙述为下述定理.

定理 1.3 设 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 是 n 维线性空间 $V_n(F)$ 的一组基, $V_n(F)$ 中向量 β_i 在该基下坐标为 X_i , $i = 1, 2, 3, \dots, m$, 则 $V_n(F)$ 中向量组 $\{\beta_1, \beta_2, \dots, \beta_m\}$ 线性相关的充分必要条件是其坐标向量组 $\{X_1, X_2, \dots, X_m\}$ 是 F^n 中的线性相关组.

定理 1.3 说明由坐标建立的 $V_n(F)$ 和 F^n 之间的一一对应关系保持线性关系不变.若不计较向量的具体形式,仅就线性关系而言, $V_n(F)$ 中有关问题都可归结为我们所熟悉的线性空间 F^n 中的相应问题,可应用熟悉的方法和已建立的理论来解决.

例 10 讨论 $P_4[x]$ 中向量: $f_1 = 1 + 2x + 4x^3$, $f_2 = x + x^2 + 4x^3$, $f_3 = 1 + x - 3x^2$, $f_4 = -2x^2 + x^3$ 的线性相关性.

解 在 $P_4[x]$ 中取基 $\{1, x, x^2, x^3\}$, 则向量组对应的坐标 $X_1 = (1 \ 2 \ 0 \ 4)^T$, $X_2 = (0 \ 1 \ 1 \ 4)^T$, $X_3 = (1 \ 1 \ -3 \ 0)^T$, $X_4 = (0 \ 0 \ -2 \ 1)^T$,

$$A = (X_1 \ X_2 \ X_3 \ X_4) = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 2 & 1 & 1 & 0 \\ 0 & 1 & -3 & -2 \\ 4 & 4 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & -2 & -2 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$\text{rank}(A) = 4$, 因而 X_1, X_2, X_3, X_4 线性无关, 即 $\{f_1, f_2, f_3, f_4\}$ 线性无关.

四、基变换与坐标变换

从定理 1.2 可以知道,线性空间 $V_n(F)$ 的基不是惟一的,同一向量在不同基下的坐

标一般也不相同,这里讨论 $V_n(F)$ 中不同基之间的关系和同一向量在不同基下坐标的关系.

设 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 和 $\{\beta_1, \beta_2, \dots, \beta_n\}$ 是 $V_n(F)$ 中的两组基,则由基的定义

$$\beta_i = (\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n) \begin{pmatrix} c_{1i} \\ c_{2i} \\ \vdots \\ c_{ni} \end{pmatrix} \text{ 记为 } \{\alpha_1, \alpha_2, \dots, \alpha_n\} C_i, \quad (1.2)$$

$$i = 1, 2, \dots, n.$$

由矩阵分块运算,(1.2)式中 n 个式子可表示为

$$(\beta_1 \quad \beta_2 \quad \cdots \quad \beta_n) = (\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n) C, \quad (1.3)$$

其中 $C = (C_1 \quad C_2 \quad \cdots \quad C_n) \in F^{n \times n}$.

定义 1.4 设 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}, \{\beta_1, \beta_2, \dots, \beta_n\}$ 是 n 维线性空间 $V_n(F)$ 的两组基,若有矩阵 $C \in F^{n \times n}$,使

$$(\beta_1 \quad \beta_2 \quad \cdots \quad \beta_n) = (\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n) C,$$

则称矩阵 C 是从基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 到基 $\{\beta_1, \beta_2, \dots, \beta_n\}$ 的过渡矩阵(基变换矩阵).

C 作为过渡矩阵,它一定是可逆矩阵,它的逆 C^{-1} 是从 $\{\beta_1, \beta_2, \dots, \beta_n\}$ 到 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 的过渡矩阵.从(1.2)式可知,在构成上,矩阵 C 的第 i 列是 β_i 在基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 下的坐标 C_i .

设向量 $\alpha \in V_n(F)$, α 在两组基下坐标分别为 X 和 Y ,则有

$$\alpha = (\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n) X, \quad (1.4)$$

$$\alpha = (\beta_1 \quad \beta_2 \quad \cdots \quad \beta_n) Y. \quad (1.5)$$

因此

$$\alpha = (\beta_1 \quad \beta_2 \quad \cdots \quad \beta_n) Y = (\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n) C Y. \quad (1.6)$$

比较(1.4)式与(1.6)式,有

$$X = C Y.$$

定理 1.4 设线性空间 $V_n(F)$ 的一组基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 到另一组基 $\{\beta_1, \beta_2, \dots, \beta_n\}$ 的过渡矩阵为 C , $V_n(F)$ 中向量 α 在两组基下坐标分别为 X, Y ,则有

$$X = C Y. \quad (1.7)$$

例 11 设 R^3 的两组基为 $\alpha_1 = (1 \quad 0 \quad -1)^T, \alpha_2 = (2 \quad 1 \quad 1)^T, \alpha_3 = (1 \quad 1 \quad 1)^T$ 和 $\beta_1 = (0 \quad 1 \quad 1)^T, \beta_2 = (-1 \quad 1 \quad 0)^T, \beta_3 = (1 \quad 2 \quad 1)^T$.

(1) 求从基 $\{\alpha_1, \alpha_2, \alpha_3\}$ 到基 $\{\beta_1, \beta_2, \beta_3\}$ 的过渡矩阵 C .

(2) 求向量 $\alpha = \alpha_1 + 2\alpha_2 - 3\alpha_3$ 在基 $\beta_1, \beta_2, \beta_3$ 下的坐标.

解 (1)由定义 1.4,过渡矩阵 C 是使等式:($\beta_1 \quad \beta_2 \quad \beta_3$) $= (\alpha_1 \quad \alpha_2 \quad \alpha_3) C$ 成立的矩阵.对线性空间 R^3 而言,这是矩阵等式:

$$\begin{pmatrix} 0 & -1 & 1 \\ 1 & 1 & 2 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ -1 & 1 & 1 \end{pmatrix} C,$$

从中可求得

$$C = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ -1 & 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & -1 & 1 \\ 1 & 1 & 2 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ -1 & -3 & -2 \\ 2 & 4 & 4 \end{pmatrix}.$$

(2) 由 $\alpha = \alpha_1 + 2\alpha_2 - 3\alpha_3$, 得 α 在基 $\alpha_1, \alpha_2, \alpha_3$ 下坐标 $X = (1 \ 2 \ -3)^T$, 由 $X = CY$, 得

$$Y = C^{-1}X = \begin{pmatrix} 0 & 1 & 1 \\ -1 & -3 & -2 \\ 2 & 4 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -7 \\ -1 \\ 3 \end{pmatrix}.$$

例12 从例10可知 $f_1 = 1 + 2x + 4x^3, f_2 = x + x^2 + 4x^3, f_3 = 1 + x - 3x^2, f_4 = -2x^2 + x^3$ 也是线性空间 $P_4[x]$ 的一组基, 求空间的基 $\{1, x, x^2, x^3\}$ 到基 $\{f_1, f_2, f_3, f_4\}$ 的过渡矩阵 C , 并求向量 $f = 1 + x + x^2 + x^3$ 在基 $\{f_1, f_2, f_3, f_4\}$ 下的坐标 Y .

解 因为易求得 f_i 在基 $\{1, x, x^2, x^3\}$ 下的坐标 C_i , 从而易得从基 $\{1, x, x^2, x^3\}$ 到基 $\{f_1, f_2, f_3, f_4\}$ 的过渡矩阵 C ,

$$C = (C_1 \ C_2 \ C_3 \ C_4) = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 2 & 1 & 1 & 0 \\ 0 & 1 & -3 & -2 \\ 4 & 4 & 0 & 1 \end{pmatrix},$$

所以

$$C^{-1} = \begin{pmatrix} 6 & -\frac{9}{2} & \frac{1}{2} & 1 \\ -7 & \frac{11}{2} & -\frac{1}{2} & -1 \\ -5 & \frac{9}{2} & -\frac{1}{2} & -1 \\ 4 & -4 & 0 & 1 \end{pmatrix}.$$

又

$$f = 1 + x + x^2 + x^3 = (1 \ x \ x^2 \ x^3) \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$= (f_1 \ f_2 \ f_3 \ f_4) C^{-1} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

所以

$$Y = C^{-1} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ -1 & 3 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}.$$

五、子空间

线性空间 $V_n(F)$ 中 V 作为集合, 对于集合中子集, 会涉及交与并等运算关系, 这里讨论对于这些关系的有关性质, 我们首先给出子空间的概念.

定义 1.5 设 $V_n(F)$ 为线性空间, W 是 V 的非空子集合. 若 W 的元素关于 V 中加法与数乘向量法运算也构成线性空间, 则称 W 是 V 的一个子空间.

例 13 任何线性空间有两个平凡子空间: 一个是它自身 $V \subseteq V$, 另一个是 $W = \{0\}$, 称为零元素空间.

子集的包含关系使得 $V_n(F)$ 的一个子集合是否为子空间的判别比较方便.

定理 1.5 设 W 是线性空间 $V_n(F)$ 的非空子集合, 则 W 是 $V_n(F)$ 的子空间的充分必要条件是

- (1) 若 $\alpha, \beta \in W$, 则 $\alpha + \beta \in W$;
- (2) 若 $\alpha \in W$, $k \in F$, 则 $k\alpha \in W$.

证明 必要性是显然的, 只证充分性.

设 W 满足(1)与(2), 则只需验证定义 1.1 中八条运算法则也满足.

因为 $k\alpha \in W$, 取 $k=0$, 则 $0\alpha = 0 \in W$. 又取 $k=-1$, $-\alpha = (-1)\alpha \in W$, 即 W 中存在零元素和一个元素的负元素. 又因为 $W \subseteq V$, 对 V 中加法与数乘, 定义 1.1 其余 6 条法则对 W 中元素进行运算时必须满足, 故由定义 1.1, W 是线性空间, 从而是 $V_n(F)$ 的子空间.

□

例 14 在线性空间 $R^{n \times n}$ 中取集合.

$$W_1 = \{A \mid A \in R^{n \times n}, A^T = A\};$$

$$W_2 = \{B \mid B \in R^{n \times n}, |B| \neq 0\}.$$

讨论 W_1 与 W_2 是否为 $R^{n \times n}$ 的子空间.

解 由于 $\forall A_1, A_2 \in W_1$, 有

$$(A_1 + A_2)^T = A_1^T + A_2^T = A_1 + A_2,$$

$$(kA_1)^T = kA_1^T = kA_1,$$

因此 $A_1 + A_2, kA_1 \in W_1$, 由定理 1.5, W_1 是 $R^{n \times n}$ 的子空间.

因为 $0 \notin W_2$, 所以 W_2 不是 $R^{n \times n}$ 的子空间.

例 15 设 $V(F)$ 是线性空间, $\alpha_1, \alpha_2, \dots, \alpha_m$ 是 V 中一组向量, 则由它们一切线性组合构成的集合:

$$L\{\alpha_1, \alpha_2, \dots, \alpha_m\} = \{\alpha \mid \alpha = \sum_{i=1}^m k_i \alpha_i, k_i \in F\}$$

是 V 的一个子空间, 称为由 $\alpha_1, \alpha_2, \dots, \alpha_m$ 生成的子空间.

证明 $\forall \alpha, \beta \in L\{\alpha_1, \alpha_2, \dots, \alpha_m\}$, 即

$$\alpha = \sum_{i=1}^m k_i \alpha_i, \quad \beta = \sum_{i=1}^m l_i \alpha_i.$$

则

$$\alpha + \beta = \sum_{i=1}^m (k_i + l_i) \alpha_i \in L\{\alpha_1, \alpha_2, \dots, \alpha_m\},$$

$$k\alpha = \sum_{i=1}^m (kk_i) \alpha_i \in L\{\alpha_1, \alpha_2, \dots, \alpha_m\},$$

故 $L\{\alpha_1, \alpha_2, \dots, \alpha_m\}$ 为 V 的子空间.

值得指出的是, 任何一个线性空间 $V_n(F)$, 若 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是它的一组基, 则 $V_n(F)$ 可表示为

$$V_n(F) = L\{\alpha_1, \alpha_2, \dots, \alpha_n\}.$$

例 16 对一个矩阵 $A \in F^{m \times n}$, 可得到两个与 A 相关的子空间:

$$N(A) = \{X \mid AX = 0\} \subseteq F^n,$$

$$R(A) = L\{A_1, A_2, \dots, A_n\} \subseteq F^m, \quad (1.8)$$

其中 $A_i, i=1, 2, \dots, n$ 是矩阵 A 的 n 个列向量. 子空间 $N(A)$ 称为矩阵 A 的零空间; $R(A)$ 称为矩阵 A 的列空间.

以子空间作为子集, 有子集的交: $W_1 \cap W_2$, 并: $W_1 \cup W_2$ 等运算, 对它们有如下定理.

定理 1.6 设 W_1, W_2 是线性空间 V 的子空间, 则有

(1) W_1 与 W_2 的交集 $W_1 \cap W_2 = \{\alpha \mid \alpha \in W_1 \text{ 且 } \alpha \in W_2\}$ 是 V 的子空间, 称为 W_1 与 W_2 的交空间.

(2) W_1 与 W_2 的和

$$W_1 + W_2 = \{\alpha \mid \alpha = \alpha_1 + \alpha_2, \alpha_1 \in W_1, \alpha_2 \in W_2\} \quad (1.9)$$

是 V 的子空间, 称为 W_1 与 W_2 的和空间.

证明 读者易验证(1), 在此只证明(2). 由定义 $W_1 + W_2 \subseteq V$, 而且非空. $\forall \alpha, \beta \in W_1 + W_2$, 则有 $\alpha_i, \beta_i \in W_i, i=1, 2$. 由

$$\alpha = \alpha_1 + \alpha_2, \quad \beta = \beta_1 + \beta_2,$$

$$\alpha + \beta = \alpha_1 + \alpha_2 + \beta_1 + \beta_2 = (\alpha_1 + \beta_1) + (\alpha_2 + \beta_2),$$

$$k\alpha = k\alpha_1 + k\alpha_2,$$

因 W_i 为子空间, 则 $\alpha_i + \beta_i \in W_i, \alpha_i, \beta_i \in W_i$, 所以 $\alpha + \beta, k\alpha \in W_1 + W_2$, 即 $W_1 + W_2$ 为 V 的子空间. \square

例 17 设有 \mathbb{R}^3 的子空间 $W_1 = L\{e_1\}$, $W_2 = L\{e_2\}$, 求 $W_1 + W_2$. 其中 $e_1 = (1 \ 0 \ 0)^T$,

$$e_2 = (0 \ 1 \ 0)^T.$$

解 取 $\alpha \in W_1 + W_2$, 则存在 $\alpha_i \in W_i, i=1,2$,

$$\alpha = \alpha_1 + \alpha_2,$$

又 $\alpha_1 \in W_1$, 故 $\alpha_1 = k_1 e_1$. 同理, $\alpha_2 \in W_2, \alpha_2 = k_2 e_2$,

因此 $\alpha = \alpha_1 + \alpha_2 = k_1 e_1 + k_2 e_2$, 即 $W_1 + W_2 \subseteq L\{e_1, e_2\}$.

又容易证明

$$L\{e_1, e_2\} \subseteq W_1 + W_2,$$

故有

$$W_1 + W_2 = L\{e_1, e_2\}.$$

在几何上, W_1 是 \mathbb{R}^3 中的 OX 轴, W_2 是 \mathbb{R}^3 中的 OY 轴, $W_1 + W_2$ 是 \mathbb{R}^3 的 XOY 平面.

W_1 与 W_2 的并集 $W_1 \cup W_2 \subseteq W_1 + W_2$, 但本例中 $W_1 \cup W_2$ 不是子空间, 这是由于 $e_1 \in W_1 \cup W_2, e_2 \in W_1 \cup W_2, e_1 + e_2 = (1 \ 1 \ 0)^T \notin W_1 \cup W_2$ 的缘故.

一般的, 若 $W_1 = L\{\alpha_1, \alpha_2, \dots, \alpha_s\}, W_2 = L\{\beta_1, \beta_2, \dots, \beta_t\}$, 则有

$$W_1 + W_2 = L\{\alpha_1, \alpha_2, \dots, \alpha_s, \beta_1, \beta_2, \dots, \beta_t\}. \quad (1.10)$$

对 V 的子空间 $W_1, W_2, W_1 \cap W_2, W_1 + W_2$, 成立的包含关系是

$$W_1 \cap W_2 \subseteq \frac{W_1}{W_2} \subseteq W_1 + W_2 \subseteq V, \quad (1.11)$$

这自然有 $\dim(W_1 \cap W_2) \leq \dim W_1 \leq \dim(W_1 + W_2) \leq \dim V$, 它们的维数成立如下关系.

定理 1.7 设 W_1 和 W_2 是线性空间 V 的子空间, 则有如下维数公式:

$$\dim W_1 + \dim W_2 = \dim(W_1 + W_2) + \dim(W_1 \cap W_2). \quad (1.12)$$

证明 设 $\dim(W_1 \cap W_2) = r, \dim W_1 = s_1, \dim W_2 = s_2, W_1 \cap W_2$ 基为 $\{\alpha_1, \alpha_2, \dots, \alpha_r\}$, 由(1.11)式, 它们可分别扩充为:

$$W_1 \text{ 的基 } |\alpha_1, \alpha_2, \dots, \alpha_r, \beta_{r+1}, \dots, \beta_{s_1}|,$$

$$W_2 \text{ 的基 } |\alpha_1, \alpha_2, \dots, \alpha_r, \gamma_{r+1}, \dots, \gamma_{s_2}|,$$

$$W_1 = L\{\alpha_1, \alpha_2, \dots, \alpha_r, \beta_{r+1}, \dots, \beta_{s_1}\},$$

则

$$W_2 = L\{\alpha_1, \alpha_2, \dots, \alpha_r, \gamma_{r+1}, \dots, \gamma_{s_2}\},$$

$$W_1 + W_2 = L\{\alpha_1, \alpha_2, \dots, \alpha_r, \beta_{r+1}, \dots, \beta_{s_1}, \gamma_{r+1}, \dots, \gamma_{s_2}\}.$$

下面证明 $\{\alpha_1, \dots, \alpha_r, \beta_{r+1}, \dots, \beta_{s_1}, \gamma_{r+1}, \dots, \gamma_{s_2}\}$ 为线性无关组.

任取数 k_i, q_i, p_i , 使

$$\sum_{i=1}^r k_i \alpha_i + \sum_{i=r+1}^{s_1} q_i \beta_i + \sum_{i=r+1}^{s_2} p_i \gamma_i = 0. \quad (1.13)$$

将上式改写为

$$-\sum_{i=r+1}^{s_1} q_i \beta_i = \sum_{i=1}^r k_i \alpha_i + \sum_{i=r+1}^{s_2} p_i \gamma_i,$$

等式左边是 $W_1 \cap W_2$ 中的向量 $\{\beta_1, \beta_2, \dots, \beta_{s_1}\}$ 的线性组合, 所以有

$$-\sum_{i=r+1}^{s_1} q_i \beta_i \in W_1 \cap W_2.$$

借助于 $W_1 \cap W_2$ 的基,该向量可以表示为

$$-\sum_{i=r+1}^{s_1} q_i \beta_i = \sum_{i=1}^r n_i \alpha_i,$$

$$\text{将它写为线性组合} \quad \sum_{i=1}^r n_i \alpha_i + \sum_{i=r+1}^{s_1} q_i \beta_i = 0.$$

由 $\{\alpha_1, \dots, \alpha_r, \beta_{r+1}, \dots, \beta_{s_1}\}$ 线性无关,有 $q_i = 0, i = r+1, \dots, s_1$. 代入(1.13)式后得

$$\sum_{i=1}^r k_i \alpha_i + \sum_{i=r+1}^{s_2} p_i \gamma_i = 0,$$

注意其中 $\alpha_1, \dots, \alpha_r, \gamma_{r+1}, \dots, \gamma_{s_2}$ 为 W_2 的基,于是 $k_i = 0 (i = 1, 2, \dots, r), p_i = 0 (i = r+1, \dots, s_2)$,故 $\{\alpha_1, \dots, \alpha_r, \beta_{r+1}, \dots, \beta_{s_1}, \gamma_{r+1}, \dots, \gamma_{s_2}\}$ 线性无关, $\dim(W_1 + W_2) = r + (s_1 - r) + (s_2 - r) = s_1 + s_2 - r$,即(1.12)式得证. \square

从(1.12)式知,若 $W_1 \cap W_2 \neq \{0\}$,则有 $\dim(W_1 + W_2) < \dim W_1 + \dim W_2$,这时 $\forall \alpha \in W_1 + W_2, \alpha = x_1 + x_2, x_i \in W_i, i = 1, 2$,往往产生其表达式中 x_1 与 x_2 不是惟一的问题.

例如, $W_1 = L\left\{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\right\}, W_2 = \left\{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}\right\}$,有 $\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \in W_1 \cap W_2$,即 $W_1 \cap W_2 \neq \{0\}$.

这时 $0 \in W_1 + W_2$ 可有两种表达式 $0 = 0 + 0$ 和 $0 = (2 \ 1 \ 0)^T - (2 \ 1 \ 0)^T$.

定义 1.6 设 W_1 和 W_2 是线性空间 V 的子空间, $W = W_1 + W_2$,如果 $W_1 \cap W_2 = \{0\}$,则称 W 是 W_1 与 W_2 的直和子空间. 记为

$$W = W_1 \oplus W_2.$$

对直和子空间,读者易证有如下等价条件.

定理 1.8 设 W_1 与 W_2 是 V 的子空间, $W = W_1 + W_2$,则成立以下等价条件

(1) $W = W_1 \oplus W_2$;

(2) $\forall X \in W, X$ 有惟一的表示式 $X = X_1 + X_2$,其中 $X_1 \in W_1, X_2 \in W_2$;

(3) W 中零向量表达式是惟一的,即只要 $0 = X_1 + X_2, X_1 \in W_1, X_2 \in W_2$,就有 $X_1 = 0, X_2 = 0$;

(4) 维数公式 $\dim W = \dim W_1 + \dim W_2$.

证明 证(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (1)即可证明等价性.

(1) \Rightarrow (2) 设 $W = W_1 \oplus W_2$,由直和的定义, $W_1 \cap W_2 = \{0\}$. 如果 $X \in W, X$ 有两种表示式,

$$X = X_1 + X_2, \quad x_i \in W_i, \quad i = 1, 2,$$

$$X = Y_1 + Y_2, \quad y_i \in W_i, \quad i = 1, 2,$$

则两式相减,有 $0 = (X_1 - Y_1) + (X_2 - Y_2)$,即

$$X_1 - Y_1 = -(X_2 - Y_2).$$

又因为 $(X_1 - Y_1) \in W_1, (X_2 - Y_2) \in W_2$,从而有

$$(X_1 - Y_1), (X_2 - Y_2) \in W_1 \cap W_2,$$

由已知 $X_1 = Y_1, X_2 = Y_2$, 即 $\forall X \in W$, 即形如 $X = X_1 + X_2$ 的表示式惟一.

(2) \Rightarrow (3) 设 $\forall X \in W, X = X_1 + X_2, x_i \in W_i, i=1,2$ 的表示式是惟一的.

又因为 $0 \in W$, 且已知有 $0 = 0 + 0, 0 \in W_i, i=1,2$, 故当 $0 = X_1 + X_2$ 时, 由表示的惟一性, 有

$$X_1 = 0, \quad X_2 = 0.$$

(3) \Rightarrow (4) 已知 W 中零向量有(3)中的表示惟一性.

任取 $X \in W_1 \cap W_2$, 则 X 可表示为

$$X = X + 0, \quad X \in W_1, \quad 0 \in W_2,$$

$$X = 0 + X, \quad 0 \in W_1, \quad X \in W_2,$$

两式相减, 即有 $0 = X - X$. 由 0 的惟一表示知 $x = 0$, 即 $W_1 \cap W_2 = \{0\}$. 从而 $\dim(W_1 \cap W_2) = 0$. 由维数公式有,

$$\dim W = \dim W_1 + \dim W_2.$$

(4) \Rightarrow (1) 当 $\dim W = \dim W_1 + \dim W_2$ 时, 由维数公式可推出

$$\dim(W_1 \cap W_2) = \dim W - (\dim W_1 + \dim W_2) = 0.$$

又注意到 $\dim(W_1 \cap W_2) = 0$ 当且仅当 $W_1 \cap W_2 = \{0\}$. 故 $W_1 \cap W_2 = \{0\}$, 由定义:

$$W = W_1 + W_2 = W_1 \oplus W_2. \quad \square$$

例 18 设 I_r 表示 r 阶单位矩阵, 对 n 阶方阵 $A = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}, B = \begin{pmatrix} 0 & 0 \\ 0 & I_{n-r} \end{pmatrix}$. 它们的列空间为 $R(A), R(B)$, 证明: $R = R(A) \oplus R(B)$.

证明 $\forall x = (a_1 \ a_2 \ \cdots \ a_n)^T \in R$, 有 $x = (a_1 \ a_2 \ \cdots \ a_r \ 0 \ \cdots \ 0)^T + (0 \ \cdots \ 0 \ a_{r+1} \ \cdots \ a_n)^T$, 由于

$$(a_1 \ \cdots \ a_r \ 0 \ \cdots \ 0)^T \in R(A),$$

$$(0 \ \cdots \ 0 \ a_{r+1} \ \cdots \ a_n)^T \in R(B),$$

所以 $R \subseteq R(A) + R(B)$, 又显然 $R \supseteq R(A) + R(B)$, 从而

$$R = R(A) + R(B).$$

又

$$\dim R(A) + \dim R(B) = r + (n - r) = n = \dim R,$$

由定理 1.8

$$R = R(A) \oplus R(B).$$

对 n 维空间 V 的任何子空间 W , 设 $\alpha_1, \dots, \alpha_r$ 为 W 的基, $r < n$, 把它们扩充为 V 的基

$$\{\alpha_1, \alpha_2, \dots, \alpha_r, \beta_{r+1}, \dots, \beta_n\},$$

设

$$U = L\{\beta_{r+1}, \dots, \beta_n\},$$

则成立

$$V = W \oplus U.$$

我们称 U 是 W 的直和补子空间. 对 V 的任何子空间 W , $\dim W = r < n$, 都存在直和补子空间 $U \neq \{0\}$.

上述定义的子空间的交、和、直和等概念可推广到 k 个子空间的情形

$$\begin{aligned}\bigcap_{i=1}^k W_i &= W_1 \cap W_2 \cap \cdots \cap W_k = \{x \mid x \in W_i, i = 1, 2, \dots, k\}, \\ \sum_{i=1}^k W_i &= W_1 + W_2 + \cdots + W_k = \{x_1 + x_2 + \cdots + x_k \mid x_i \in W_i, i = 1, 2, \dots, k\}.\end{aligned}$$

1.2 内积空间

在实际应用中,线性空间 $V_n(F)$ 中的许多问题会涉及诸如向量的长度、向量之间的夹角、正交等与度量有关的问题.这一节,我们将把几何空间 \mathbb{R}^3 中的数量积推广到线性空间 $V_n(F)$ 中,定义内积的概念,并由此建立有关的度量关系.

一、欧氏空间与酉空间

定义1.7 对数域 F 上的 n 维线性空间 $V_n(F)$, 定义的一个从 $V_n(F)$ 中向量到数域 F 的二元运算, 记为 (α, β) , 即 $(\alpha, \beta) : V_n(F) \rightarrow F$, 如果满足

(1) 对称性: $(\alpha, \beta) = (\overline{\beta}, \alpha)$, 其中 $(\overline{\beta}, \alpha)$ 表示复数 (β, α) 的共轭;

(2) 线性性: $(k\alpha, \beta) = k(\alpha, \beta)$,

$$(\alpha_1 + \alpha_2, \beta) = (\alpha_1, \beta) + (\alpha_2, \beta);$$

(3) 正定性: $(\alpha, \alpha) \geq 0$, $(\alpha, \alpha) = 0$ 的充要条件是 $\alpha = 0$,

则称 (α, β) 是 $V_n(F)$ 的一个内积, 并称其中定义了内积的线性空间 $[V_n(F); (\alpha, \beta)]$ 为内积空间.

如果 $V_n(F)$ 是实数域 \mathbb{R} 上的线性空间 $V_n(\mathbb{R})$, 则 $(\alpha, \beta) \in \mathbb{R}$ 为实内积, 对称性相应为 $(\alpha, \beta) = (\beta, \alpha)$, $[V_n(\mathbb{R}); (\alpha, \beta)]$ 为欧氏空间 (Euclidean Space).

同理 $V_n(F)$ 是复数域 \mathbb{C} 上的线性空间 $V_n(\mathbb{C})$ 时, $(\alpha, \beta) \in \mathbb{C}$ 为复内积, 称 $[V_n(\mathbb{C}); (\alpha, \beta)]$ 为酉空间 (Unitary Space).

例19 下列线性空间对所定义的内积为欧氏空间:

(1) $[\mathbb{R}^n; (\alpha, \beta) = \alpha^T \beta]$, 其中 $\alpha = (x_1 \ x_2 \ \cdots \ x_n)^T$, $\beta = (y_1 \ y_2 \ \cdots \ y_n)^T$, $(\alpha, \beta) = \sum_{i=1}^n x_i y_i$. 它是几何空间 \mathbb{R}^3 中数量积的自然推广, 习惯上仍用 \mathbb{R}^n 表示欧氏空间 $[\mathbb{R}^n; (\alpha, \beta) = \alpha^T \beta]$.

(2) $[\mathbb{R}^{n \times n}; (A, B) = \text{tr}(AB^T)]$.

对此定义的内积 $\text{tr}(AB^T)$, 易证对称性与线性性.

$\forall A \neq 0$, (AA^T) 为正定或半正定矩阵, 所以 $\text{tr}(AA^T) \geq 0$, 且 $\text{tr}(AA^T) = 0 \Leftrightarrow A = 0$, 所以正定性满足.

(3) $[P_n[x]; (f(x), g(x)) = \int_0^1 f(x) \cdot g(x) dx]$.

读者易证 $\int_0^1 f \cdot g dx$ 为实内积.

对我们讨论过的线性空间 $C[a, b]$, 也可类似地定义内积 $(f(x), g(x)) = \int_a^b f(x) \cdot g(x) dx$, 但 $C[a, b]$ 是无穷维线性空间, 在此不考虑.

值得注意的是, 在同一个线性空间上, 可以定义不同的内积, 例如对 \mathbb{R}^n , 取一个给定的正定矩阵 $A_{n \times n}$, 可以定义内积 $(\alpha, \beta) = \alpha^T A \beta$. 这是一个双线性型, 其中 $(\alpha, \alpha) = \alpha^T A \alpha$ 就是大家熟悉的二次型. $[\mathbb{R}^n, (\alpha, \beta) = \alpha^T A \beta]$ 与 $[\mathbb{R}^n, (\alpha, \beta) = \alpha^T \beta]$ 就是不一样的欧氏空间. 当然, $\alpha^T \beta$ 可看做是 $\alpha^T A \beta$ 中 $A = I$ 的特例.

在复数空间中, 经常会用到一个矩阵共轭的概念. 矩阵 $A = (a_{ij}) \in \mathbb{C}^{n \times n}$, A 的共轭是在 A 中对每一个元素取其共轭复数后得到的矩阵. 用 \bar{A} 记 A 的共轭矩阵, 则 $\bar{A} = (\bar{a}_{ij}) \in \mathbb{C}^{n \times n}$. A 的共轭转置矩阵记为 A^H , $A^H = (\bar{a}_{ij})^T$.

例 20 常见的酉空间:

(1) $[\mathbb{C}^n; (\alpha, \beta) = \beta^H \alpha]$, 其中 β^H 表示对向量 β 取共轭转置.

设 $\alpha = (x_1 \ \cdots \ x_n)^T$, $\beta = (y_1 \ \cdots \ y_n)^T$, 则

$$(\alpha, \beta) = \beta^H \alpha = x_1 \bar{y}_1 + x_2 \bar{y}_2 + \cdots + x_n \bar{y}_n,$$

其中 $\beta^H = (\bar{\beta})^T$.

(2) $[\mathbb{C}^n; (A, B) = \text{tr}(B^H A)]$.

设 $A = (a_{ij})_{n \times n}$, $B = (b_{ij})_{n \times n}$, 则

$$(A, B) = \text{tr}(B^H A) = \sum_{i=1}^n \sum_{j=1}^n a_{ji} \bar{b}_{ji}.$$

在内积空间 $[V_n(F); (\alpha, \beta)]$ 中, 可类似于几何空间引入向量长度、夹角等概念.

定义 1.8 设 $[V_n(F); (\alpha, \beta)]$ 为内积空间, 称

$$\|\alpha\| = \sqrt{(\alpha, \alpha)} \quad (1.14)$$

为向量 α 的长度, 若 $\|\alpha\| = 1$, 则称 α 为单位向量.

在欧氏空间中, 我们也称由内积定义的向量长度 $\|\alpha\|$ 为 α 的欧几里得范数.

定理 1.9(Cauchy 不等式) 设 $[V_n(F); (\alpha, \beta)]$ 是内积空间, 则对空间中任意向量 $\alpha, \beta \in V_n(F)$, 都有

$$|(\alpha, \beta)|^2 \leq (\alpha, \alpha)(\beta, \beta),$$

其中等式成立的充要条件是 α 与 β 线性相关.

证明 只证明实数域的情形. 若 $\alpha = 0$ 或 $\beta = 0$, 则结论显然成立. 设 $\beta \neq 0$, 对任意数 k , 取向量 $\alpha - k\beta$, 则有

$$(\alpha - k\beta, \alpha - k\beta) \geq 0,$$

$$\text{即 } (\alpha, \alpha) + |k|^2 (\beta, \beta) - 2k(\alpha, \beta) \geq 0.$$

由 $\beta \neq 0$, 可取到数

$$k = \frac{(\alpha, \beta)}{(\beta, \beta)},$$

代入上式,即得

$$(\alpha, \beta)^2 \leq (\alpha, \alpha) \cdot (\beta, \beta),$$

从而有

$$|(\alpha, \beta)|^2 \leq (\alpha, \alpha) \cdot (\beta, \beta).$$

等式成立的充要条件是

$$\alpha - k\beta = 0,$$

即 α 与 β 线性相关.

内积空间中的 Cauchy-Schwarz 不等式是数学中非常重要的不等式之一. 当内积取具体的形式时,它可得出相应具体的不等式. 例如

$$C^*, (\alpha, \beta) = \beta^H \alpha \Rightarrow \left| \sum_{i=1}^n x_i y_i \right|^2 \leq \sum_{i=1}^n |x_i|^2 \cdot \sum_{i=1}^n |y_i|^2.$$

$$C^{n \times n}, (A, B) = \text{tr}(B^H A) \Rightarrow |\text{tr}(B^H A)|^2 \leq \text{tr}(A^H A) \cdot \text{tr}(B^H B)$$

用向量长度 $\|\alpha\| = \sqrt{(\alpha, \alpha)}$. $\|\beta\| = \sqrt{(\beta, \beta)}$, Cauchy-Schwarz 不等式可写为

$$|(\alpha, \beta)| \leq \|\alpha\| \cdot \|\beta\|. \quad \square$$

例 21 证明内积空间 $[V_n(F); (\alpha, \beta)]$ 中, $\forall \alpha, \beta \in V_n(F)$, 成立三角不等式

$$\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|.$$

证明 因为,由 Cauchy 不等式

$$\begin{aligned} \|\alpha + \beta\|^2 &= (\alpha + \beta, \alpha + \beta) = (\alpha, \alpha) + (\beta, \beta) + (\alpha, \beta) + (\beta, \alpha) \\ &\leq \|\alpha\|^2 + \|\beta\|^2 + 2\|\alpha\|\|\beta\| \\ &= (\|\alpha\| + \|\beta\|)^2, \end{aligned}$$

所以

$$\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|.$$

由此可得出向量长度 α 的一些简单性质:

- (1) $\|k\alpha\| = \|k\| \cdot \|\alpha\|;$
- (2) $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|;$

(3) $\forall \alpha \neq 0, \alpha^0 = \frac{\alpha}{\|\alpha\|}$, $\|\alpha^0\| = 1$, 取 $\alpha^0 = \frac{\alpha}{\|\alpha\|}$ 的过程称为向量 α 的标准化.

从 Cauchy 不等式,可得出

$$\left| \frac{(\alpha, \beta)}{\|\alpha\|\|\beta\|} \right| \leq 1,$$

从而可在欧氏空间中定义向量 α 与 β 之间的夹角.

□

在欧氏空间中,非零向量 α 与 β 的夹角定义为

$$\theta = \arccos \frac{(\alpha, \beta)}{\|\alpha\|\|\beta\|}.$$

若 $(\alpha, \beta) = 0$, 则 α 与 β 是正交的(垂直的).

定义 1.9 在内积空间中,若向量 α, β 满足

$$(\alpha, \beta) = 0,$$

则称向量 α 与 β 是正交的.

对内积空间中的一个向量组 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$, 若 $(\alpha_i, \alpha_j) = 0, i \neq j$, 则称 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 为正交向量组. 进一步, 若满足

$$(\alpha_i, \alpha_j) = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$$

则称向量组 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 为标准正交的向量组.

定理 1.10 不含零向量的正交向量组是线性无关的.

证明 设 $\{\alpha_1, \dots, \alpha_n\}$ 是内积空间 $V_n(F)$ 中的正交向量组, $\alpha_i \neq 0, i=1, 2, \dots, n$. 设有数 $k_i, i=1, 2, \dots, n$, 使

$$k_1\alpha_1 + k_2\alpha_2 + \dots + k_n\alpha_n = 0.$$

取 α_j , 则有

$$\left(\sum_{i=1}^n k_i \alpha_i, \alpha_j \right) = k_j (\alpha_j, \alpha_j) = 0.$$

因 $(\alpha_j, \alpha_j) \neq 0$, 故必有 $k_j = 0, j=1, 2, \dots, n$, 即 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 线性无关. \square

定理 1.10 说明, $V_n(F)$ 中不含零向量的正交向量组至多含 n 个向量, 而且由 n 个非零向量构成的正交组就是 $V_n(F)$ 的基.

二、标准正交基

在内积空间 $V_n(F)$ 中, 取一组基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$, 则 $\forall \alpha, \beta \in V_n(F)$, 分别有坐标 $X, Y \in F^n$:

$$\alpha = (\alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_n)X = \sum_{i=1}^n x_i \alpha_i,$$

$$\beta = (\alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_n)Y = \sum_{j=1}^n y_j \alpha_j,$$

则

$$(\alpha, \beta) = \sum_{i=1}^n \sum_{j=1}^n x_i \bar{y}_j (\alpha_i, \alpha_j). \quad (1.15)$$

设矩阵 $A = ((\alpha_i, \alpha_j))_{n \times n}$, 即矩阵 A 的元素 $a_{ij} = (\alpha_i, \alpha_j)$, 则 (1.15) 式可用矩阵运算表示为

$$(\alpha, \beta) = Y^H A X \quad (1.16)$$

因此, 给定 $A, V_n(F)$ 中向量的内积就可转化为矩阵的运算. 由 A 中元素构成可知, $A^H = A$, 即 A 为 Hermite 矩阵, 又 $\alpha \neq 0$ 时, 其坐标 $X \neq 0$, 有 $(\alpha, \alpha) = X^H A X > 0$, 即 A 为正定的 Hermite 阵. 我们称 A 为内积空间 $[V_n(F); (\alpha, \beta)]$ 的度量矩阵(或 Grame 矩阵).

如果 $A = I$, 则 $(\alpha, \beta) = Y^H X$. $A = I$ 相当于

$$(\alpha_i, \alpha_j) = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

定义 1.10 在内积空间 $[V_n(F); (\alpha, \beta)]$ 中, 若一组基 $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$ 满足条件

$$(\varepsilon_i, \varepsilon_j) = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

则称 $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$ 为 $V_n(F)$ 的标准正交基.

为讨论标准正交基的求法, 我们先给出如下定理:

定理 1.11 (Gram-Schmidt 正交化方法) 设 $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$ 是内积空间 $[V_n(F); (\alpha, \beta)]$ 中线性无关的向量组, 则由如下方法:

$$\begin{aligned}\beta_1 &= \alpha_1, \\ \beta_k &= \alpha_k - \sum_{i=1}^{k-1} \frac{(\alpha_k, \beta_i)}{(\beta_i, \beta_i)} \beta_i, \quad k = 2, 3, \dots, m,\end{aligned}\tag{1.17}$$

所得向量组 $\{\beta_1, \beta_2, \dots, \beta_m\}$ 是正交向量组.

证明 对向量组的个数进行归纳证明.

$n=2$ 时, 由(1.17)式

$$\begin{aligned}(\beta_1, \beta_2) &= (\beta_1, \alpha_2) - \overline{\frac{(\alpha_2, \beta_1)}{(\beta_1, \beta_1)} (\beta_1, \beta_1)} \\ &= (\beta_1, \alpha_2) - \overline{(\alpha_2, \beta_1)} = (\beta_1, \alpha_2) - (\beta_1, \alpha_2) = 0,\end{aligned}$$

即 β_1 与 β_2 正交.

设 $n=m-1$ 时, 由(1.17)式所得向量组 $\{\beta_1, \beta_2, \dots, \beta_{m-1}\}$ 正交.

$$\begin{aligned}n=m \text{ 时}, \quad \beta_m &= \alpha_m - \sum_{i=1}^{m-1} \frac{(\alpha_m, \beta_i)}{(\beta_i, \beta_i)} \beta_i, \quad \forall \beta_i, i < m, \\ (\beta_m, \beta_j) &= (\alpha_m, \beta_j) - \sum_{i=1}^{m-1} \frac{(\alpha_m, \beta_i)}{(\beta_i, \beta_i)} (\beta_i, \beta_j).\end{aligned}$$

由归纳假设 $(\beta_i, \beta_j) = 0, i \neq j, i, j < m,$

$$\text{则有 } (\beta_m, \beta_j) = (\alpha_m, \beta_j) - \overline{\frac{(\alpha_m, \beta_i)}{(\beta_i, \beta_i)} (\beta_i, \beta_j)} = 0.$$

所以由归纳法, $\{\beta_1, \beta_2, \dots, \beta_m\}$ 为正交向量组. \square

从 Schmidt 过程可以看到, 向量组之间有关系:

$$L\{\alpha_1, \alpha_2, \dots, \alpha_m\} = L\{\beta_1, \beta_2, \dots, \beta_m\}. \tag{1.18}$$

当 $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$ 是一组基时, 用 Schmidt 过程便可得到正交的基 $\{\beta_1, \beta_2, \dots, \beta_m\}$, 然后再标准化为

$$\varepsilon_i = \frac{\beta_i}{\|\beta_i\|}, \quad i = 1, 2, \dots, n,$$

就可得 $V_n(F)$ 中的标准正交基 $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$. 因此, 内积空间 $[V_n(F); (\alpha, \beta)]$ 中的标准正交基是存在的.

把正交化方法和标准化方法结合在一起, 可得从一组基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 得到标准正交基的方法

$$\left\{ \begin{array}{l} \beta_1 = \alpha_1, \varepsilon_1 = \frac{\beta_1}{\|\beta_1\|}, \\ \beta_2 = \alpha_2 - (\alpha_2, \varepsilon_1)\varepsilon_1, \varepsilon_2 = \frac{\beta_2}{\|\beta_2\|}, \\ \dots \\ \beta_n = \alpha_n - \sum_{i=1}^{n-1} (\alpha_n, \varepsilon_i)\varepsilon_i, \varepsilon_n = \frac{\beta_n}{\|\beta_n\|}. \end{array} \right.$$

用矩阵运算可表示为

$$(\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n) = (\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_n) \begin{pmatrix} \|\beta_1\| & (\alpha_2, \varepsilon_1) & \cdots & (\alpha_n, \varepsilon_1) \\ & \|\beta_2\| & \cdots & (\alpha_n, \varepsilon_2) \\ & & \ddots & \vdots \\ & & & \|\beta_n\| \end{pmatrix}. \quad (1.19)$$

标准正交基是内积空间中十分方便的基. $\forall \alpha, \beta \in V_n(F), \alpha = (\varepsilon_1 \cdots \varepsilon_n)X, \beta = (\varepsilon_1 \cdots \varepsilon_n)Y$, 则由前面讨论可知:

当 $V_n(F)$ 为欧氏空间 $V_n(R)$ 时, $(\alpha, \beta) = X^T Y = Y^T X$,

当 $V_n(F)$ 为酉空间 $V_n(C)$ 时, $(\alpha, \beta) = Y^H X$,

它们分别转换为了 R^n 和 C^n 空间中最常见的内积计算. 而且当坐标 $X = (x_1 \ x_2 \ \cdots \ x_n)^T$ 时,

$$x_i = (\alpha, \varepsilon_i), i = 1, 2, \dots, n.$$

例 22 内积空间 $[R^n; (\alpha, \beta) = \alpha^T \beta]$ 中, 基 $e_1 = (1 \ 0 \ 0 \ \cdots \ 0)^T, e_2 = (0 \ 1 \ 0 \ \cdots \ 0)^T, \dots, e_n = (0 \ 0 \ \cdots \ 0 \ 1)^T$ 是标准正交基. $\{e_1, e_2, \dots, e_n\}$ 也是 $[C^n; (\alpha, \beta) = \beta^H \alpha]$ 的标准正交基.

$\{E_{ij}, i=1, 2, \dots, n; j=1, 2, \dots, n\}$ 是内积空间 $[R^{n \times n}; (A, B) = \text{tr}(B^T A)]$ 和 $[C^{n \times n}; (A, B) = \text{tr}(B^H A)]$ 的标准正交基.

例 23 设 U 为内积空间 $[V_n(F); (\alpha, \beta)]$ 的一个子空间, 定义 $V_n(F)$ 的一个子集

$$U^\perp = \{\alpha | \alpha \in V_n(F), \forall \beta \in U, (\alpha, \beta) = 0\}, \quad (1.20)$$

证明 (1) U^\perp 是 $V_n(F)$ 的子空间;

$$(2) V_n(F) = U \oplus U^\perp.$$

证明 (1) 显然 $0 \in U^\perp$, 所以 $U^\perp \neq \emptyset$.

又 $\forall \alpha_1, \alpha_2 \in U^\perp$, 对 $k_1 \alpha_1 + k_2 \alpha_2$ 和 $\forall \beta \in U$,

$$(k_1 \alpha_1 + k_2 \alpha_2, \beta) = k_1 (\alpha_1, \beta) + k_2 (\alpha_2, \beta) = 0,$$

因此 $k_1 \alpha_1 + k_2 \alpha_2 \in U^\perp$.

由定理 1.5, U^\perp 是 $V_n(F)$ 的子空间.

(2) 设 $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$ 是 U 的标准正交基, 把它扩充为 $V_n(F)$ 的标准正交基 $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n, \eta_{n+1}, \dots, \eta_r\}$ 则

$$\begin{aligned}V_n &= L\{\alpha_1, \alpha_2, \dots, \alpha_r, \eta_{r+1}, \dots, \eta_p\}, \\U &= L\{\alpha_1, \alpha_2, \dots, \alpha_r\}, \\U^\perp &= L\{\eta_{r+1}, \dots, \eta_p\}.\end{aligned}$$

所以

$$V_n = U + U^\perp.$$

又 $\forall \xi \in U \cap U^\perp$, 有 $\xi \in U$, 而且 $\xi \in U^\perp$, 从而 $(\xi, \xi) = 0$, 即 $\xi = 0$, 即 $V_n = U \oplus U^\perp$.

我们称由(1.21)式所定义的子空间 U^\perp 为 U 的正交补子空间.

1.3 线性变换

线性变换讨论的是线性空间上的一类对应关系, 它也是数学及工程技术中研究问题的重要工具. 这一节建立线性变换的概念, 重点是借助于矩阵来刻画线性变换的性质.

一、线性变换

定义 1.11 设 $V_n(F)$ 是一个线性空间, 若有 $V_n(F)$ 上的对应关系 T , 使 $\forall \alpha \in V_n(F)$, 都有确定的向量 $\alpha' = T(\alpha) \in V_n(F)$ 与之对应, 则称 T 为 $V_n(F)$ 上一个变换. 又若 T 对线性空间中的线性运算, 满足

$$(1) \quad \forall \alpha, \beta \in V_n(F), T(\alpha + \beta) = T(\alpha) + T(\beta);$$

$$(2) \quad \forall k \in F, \forall \alpha \in V_n(F), T(k\alpha) = kT(\alpha),$$

则称 T 是线性空间 $V_n(F)$ 上的一个线性变换.

为了表示方便, (1)与(2)可合写为

$$T(k_1\alpha_1 + k_2\alpha_2) = k_1T(\alpha_1) + k_2T(\alpha_2). \quad (1.21)$$

若 $\alpha' = T(\alpha)$, 则称 α' 为 α 在 T 下的像, α 为 α' 的原像.

例 24 线性空间 $V_n(F)$ 上的相似变换 $T: \forall \alpha \in V_n(F), T(\alpha) = \lambda\alpha$ 是线性变换, 其中 λ 为给定的数, $\lambda \in F$.

当 $\lambda = 0$ 时, $\forall \alpha \in V_n(F), T(\alpha) = 0$, 称 T 为零变换;

当 $\lambda = 1$ 时, $\forall \alpha \in V_n(F), T(\alpha) = \alpha$, 称 T 为恒等变换.

例 25 微分变换 $\frac{d}{dx}$ 是线性空间 $P_n[x]$ 上的线性变换.

应注意积分变换 $J_x: J_x(p(x)) = \int_a^x p(t)dt$ 不是 $P_n[x]$ 上的线性变换. 尽管积分运算具有线性性, 但对 $x^{n-1}, J_x(x^{n-1}) = \frac{1}{n}(x^n - a^n) \notin P_n[x]$. J_x 是线性空间 $C[a, b]$ 上的线性变换.

例 26 在线性空间 F^n 中, 定义变换 T_A 如下:

$$\forall X \in F^n, \quad T_A(X) = AX,$$

其中 $A = (a_{ij}) \in F^{n \times n}$ 是一个给定的方阵. 证明 T_A 是 F^n 上的线性变换.

证明 因为 $\forall X \in F^n, AX \in F^n$, 所以 T_A 是 F^n 上的变换.

又 $\forall X_1, X_2 \in F^n$ 及数 $k_1, k_2 \in F$,

$$\begin{aligned} T_A(k_1X_1 + k_2X_2) &= A(k_1X_1 + k_2X_2) = k_1AX_1 + k_2AX_2 \\ &= k_1T_A(X_1) + k_2T_A(X_2), \end{aligned}$$

故 T_A 是 F^n 上的线性变换.

由定义读者可自己验证 V_n 上线性变换 T 具有如下简单性质:

$$(1) T(0) = 0;$$

$$(2) T(-\alpha) = -T(\alpha);$$

$$(3) T \sum_{i=1}^r k_i \alpha_i = \sum_{i=1}^r k_i T(\alpha_i);$$

(4) 若 $\{\alpha_1, \alpha_2, \dots, \alpha_s\}$ 是线性相关的向量组, 则 $\{T(\alpha_1), T(\alpha_2), \dots, T(\alpha_s)\}$ 也是线性相关的向量组.

性质(4)说明线性变换把线性相关组仍变为线性相关组, 但一般线性变换不能保持线性无关性不变, 如零变换就把线性无关组变成线性相关组.

定理 1.12 设 T 是线性空间 $V_n(F)$ 上的线性变换, 则有:

(1) $R(T) = \{\beta \mid \exists \alpha \in V_n(F), \text{使 } \beta = T(\alpha)\}$ 是 $V_n(F)$ 的子空间, 称为 T 的像空间.

(2) $N(T) = \{\alpha \mid T(\alpha) = 0\}$ 是 $V_n(F)$ 的子空间, 称为 T 的零空间.

证明 (1) $\forall k_1, k_2 \in F, \beta_1, \beta_2 \in R(T)$, 有

$$k_1\beta_1 + k_2\beta_2 = k_1T(\alpha_1) + k_2T(\alpha_2) = T(k_1\alpha_1 + k_2\alpha_2),$$

所以, $k_1\beta_1 + k_2\beta_2 \in R(T)$; 又 $R(T) \subseteq V$, 且非空, 所以 $R(T)$ 为 $V_n(F)$ 的子空间. \square

同理可证 $N(T)$ 也是 $V_n(F)$ 的非空子集, 对线性运算也封闭, 从而为子空间.

我们称子空间 $R(T)$ 的维数 $\dim R(T)$ 为线性变换 T 的秩, $\dim N(T)$ 为 T 的零度.

例 27 求例 26 中给出线性变换 T_A 的像空间与零空间.

解 由 T_A 的定义

$$R(T_A) = \{Y \mid Y = AX, X \in F^n\},$$

设 A 可按列分块为 $(A_1 \ A_2 \ \cdots \ A_n)$, $X = (x_1 \ x_2 \ \cdots \ x_n)^T$, 对 $\forall Y \in R(T_A)$, 则

$$Y = AX = \sum_{i=1}^n x_i A_i.$$

所以 $R(T_A) \subseteq L\{A_1, A_2, \dots, A_n\}$. 又显然 $R(T_A) \supseteq L\{A_1, A_2, \dots, A_n\}$,

所以 $R(T_A) = L\{A_1, A_2, \dots, A_n\} = R(A)$,

即 T_A 的像空间 $R(T_A)$ 就是前面讨论过的矩阵 A 的列空间.

同理可得 $N(T_A) = N(A)$, 即 $N(T_A)$ 为矩阵 A 的零空间.

另外, 设 T 是线性空间 $V_n(F)$ 上的线性变换, $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 是 $V_n(F)$ 的基, 即 $V_n(F) = L\{\alpha_1, \alpha_2, \dots, \alpha_n\}$, 则有 $R(T) = L\{T(\alpha_1), T(\alpha_2), \dots, T(\alpha_n)\}$. (证明留作习题)

当线性空间 $V_n(F)$ 上定义有多个线性变换时, 实际应用中常常会用它们构成新的变换.

定义 1.12 设 T_1 与 T_2 都是线性空间 $V_n(F)$ 上的线性变换, 定义如下运算:

(1) 变换的乘积 $T_1 \cdot T_2$:

$$\forall \alpha \in V_n(F), \quad (T_1 \cdot T_2)(\alpha) = T_1(T_2(\alpha));$$

(2) 变换的加法 $T_1 + T_2$:

$$\forall \alpha \in V_n(F), \quad (T_1 + T_2)(\alpha) = T_1(\alpha) + T_2(\alpha);$$

(3) 数乘变换 kT :

$$\forall \alpha \in V_n(F), \quad k \in F, \quad (kT)(\alpha) = kT(\alpha);$$

(4) 可逆变换: 对变换 T_1 , 如果存在变换 T_2 , 使

$$T_1 \cdot T_2 = T_2 \cdot T_1 = I \text{ (恒等变换),}$$

则称 T_1 为可逆变换, T_2 是 T_1 的逆变换, 记为 $T_2 = T_1^{-1}$.

可以证明, 上述线性变换运算的结果仍然是 $V_n(F)$ 上的线性变换.

对线性变换 T , 当 n 个 T 相乘时, 我们用 T 的 n 次幂来表示, 即

$$T^n = \underset{n \text{ 个}}{\overbrace{T \cdot T \cdot T \cdots T}}.$$

二、线性变换的矩阵

设 T 为线性空间 $V_n(F)$ 上的线性变换, 取 $V_n(F)$ 的一组基为 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$, $\forall \alpha \in V_n(F)$,

$$\alpha = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n,$$

所以

$$T(\alpha) = x_1T(\alpha_1) + x_2T(\alpha_2) + \cdots + x_nT(\alpha_n)$$

这表明 $T(\alpha)$ 是由基 $\{\alpha_i\}$ 的像 $\{T(\alpha_i)\}$ 决定的. 进一步 $T(\alpha_i) \in V_n(F)$, 它可由基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 表示为

$$\begin{aligned} T(\alpha_1) &= a_{11}\alpha_1 + a_{21}\alpha_2 + \cdots + a_{n1}\alpha_n, \\ T(\alpha_2) &= a_{12}\alpha_1 + a_{22}\alpha_2 + \cdots + a_{n2}\alpha_n, \\ &\vdots \\ T(\alpha_n) &= a_{1n}\alpha_1 + a_{2n}\alpha_2 + \cdots + a_{nn}\alpha_n, \end{aligned} \tag{1.22}$$

可用矩阵表示为

$$(T(\alpha_1) \quad T(\alpha_2) \quad \cdots \quad T(\alpha_n)) = (\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n)A,$$

其中 $A = (a_{ij})_{n \times n}$. 我们记

$$(T(\alpha_1) \quad T(\alpha_2) \quad \cdots \quad T(\alpha_n)) = T(\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n).$$

定义 1.13 设 T 为 $V_n(F)$ 上线性变换, $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 为 $V_n(F)$ 的基, 若存在 n 阶方阵 $A \in F^{n \times n}$, 使

$$T(\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n) = (\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n)A, \tag{1.23}$$

则称 A 为 T 在基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 下的矩阵.

从(1.21)式可知, 变换矩阵 A 的第 i 列是 α_i 的像 $T(\alpha_i)$ 在基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 下的坐标, 坐标的惟一性决定了 A 的惟一性. 反之给定一个 A , 由(1.24)式可完全确定基中向量 α_i 的像 $T(\alpha_i)$. 从而也就完全确定了一个线性变换 T . 因此在选定的基下, 线性变换 T 与 n 阶方阵 A 之间是一一对应的,

$$T \xleftarrow{|\alpha_1, \alpha_2, \dots, \alpha_n|} A.$$

又 $\forall \alpha \in V^n$, 设 α 与 $T(\alpha)$ 在基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 下坐标分别是 X 与 Y , 即

$$\alpha = (\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n)X, \quad T(\alpha) = (\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n)Y.$$

$$\begin{aligned} \text{由 } T(\alpha) &= T[(\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n)X] = T(\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n)X \\ &= (\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n)AX, \end{aligned}$$

得

$$Y = AX. \quad (1.24)$$

这就是变换在给定基下的坐标式. 在给定基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 下, 确定向量 α 的像 $T(\alpha)$, 只需确定它的坐标 Y . 式(1.24)说明不论 T 为何种线性变换, 在给定基下求得变换矩阵 A 后, 从原像的坐标 X , 求像的坐标 Y , 总是归结为线性变换 $T_A: F^n \rightarrow F^n$ 的像 $Y = T_A(X)$. 这就是用矩阵处理线性变换的方便之处. 下面将看到, 线性变换的运算与它们的矩阵的相应运算相对应.

定理 1.13 设 T_1, T_2 是线性空间 $V^n(F)$ 上的两个线性变换, 对基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$, 它们的矩阵分别是 A_1 和 A_2 , 则在基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 下,

- (1) $T_1 + T_2$ 的矩阵为 $(A_1 + A_2)$;
- (2) $T_1 T_2$ 的矩阵为 $A_1 A_2$;
- (3) kT_1 的矩阵为 kA_1 ;
- (4) T_1 为可逆变换的充分必要条件是 A_1 为可逆矩阵, 且 T_1^{-1} 的矩阵为 A_1^{-1} .

证明 只选证(2)与(4), 其余留作练习.

- (2) 由已知条件, 有

$$T_1(\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n) = (\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n)A_1,$$

$$T_2(\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n) = (\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n)A_2.$$

$$\begin{aligned} \text{由此 } T_1 T_2(\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n) &= T_1(T_2(\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n)) \\ &= T_1(\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n)A_2 = (\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n)A_1 A_2. \end{aligned}$$

因此, 在基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 下线性变换 $T_1 T_2$ 的矩阵是 $A_1 A_2$.

(4) T_1 可逆当且仅当 $\exists T_2$, 使 $T_1 T_2 = I$, 又恒等变换 I 在任何基下的矩阵为单位矩阵 I , 由(2)有等式 $A_1 A_2 = I$, 从而 T_1 可逆当且仅当 A_1 可逆, T_1^{-1} 的矩阵是 A_1^{-1} . \square

线性变换的矩阵是与空间中选定的基相联系的, 一般基改变, 同一个线性变换就会有不同的矩阵. 下面讨论不同基下矩阵之间的关系.

设 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 和 $\{\beta_1, \beta_2, \dots, \beta_n\}$ 是 V^n 的两组基, 它们之间的关系为

$$(\beta_1 \quad \beta_2 \quad \cdots \quad \beta_n) = (\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n)C,$$

这里 C 称过渡矩阵.

设 T 在两组基下变换矩阵分别是 A 和 B , 即

$$T(\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n) = (\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n)A,$$

$$T(\beta_1 \quad \beta_2 \quad \cdots \quad \beta_n) = (\beta_1 \quad \beta_2 \quad \cdots \quad \beta_n)B,$$

$$\begin{aligned} \text{则 } T(\beta_1 \quad \beta_2 \quad \cdots \quad \beta_n) &= T[(\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n)C] = (\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n)AC \\ &= (\beta_1 \quad \beta_2 \quad \cdots \quad \beta_n)C^{-1}AC. \end{aligned}$$

与前面比较, 有

$$B = C^{-1}AC.$$

定理 1.14 设线性空间 $V_n(F)$ 的基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 到基 $\{\beta_1, \beta_2, \dots, \beta_n\}$ 的过渡矩阵为 C . 又 $V_n(F)$ 上线性变换 T 在上述两组基下矩阵分别为 A 和 B , 则有

$$B = C^{-1}AC. \quad (1.25)$$

定理 1.14 说明线性空间上同一线性变换在不同基下的矩阵是相似的. 反之, 如果两个矩阵相似: $B = C^{-1}AC$, 又已知 A 是 T 在某一组基下的矩阵, 则一定可找到另一组基, 使 T 在该基下矩阵为 B , C 正是两组基之间的过渡矩阵.

例 28 设 \mathbb{R}^3 上线性变换 T 为

$$T((x_1 \quad x_2 \quad x_3)^T) = (x_1 + 2x_2 + x_3 \quad x_2 - x_3 \quad x_1 + x_3)^T,$$

求 T 在基

$$\alpha_1 = (1 \quad 0 \quad 1)^T, \quad \alpha_2 = (0 \quad 1 \quad 1)^T, \quad \alpha_3 = (1 \quad -1 \quad 1)^T$$

下的矩阵 B .

解 在自然基 e_1, e_2, e_3 下, 线性变换 T 的坐标关系式为:

$$Y = \begin{pmatrix} x_1 + 2x_2 + x_3 \\ x_2 - x_3 \\ x_1 + x_3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

根据由变换的坐标式 $Y = AX$ 得 T 在自然基下矩阵

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix},$$

又从 $(\alpha_1 \quad \alpha_2 \quad \alpha_3) = (e_1 \quad e_2 \quad e_3)C$ 得过渡矩阵

$$C = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix}, \quad C^{-1} = \begin{pmatrix} 2 & 1 & -1 \\ -1 & 0 & 1 \\ -1 & -1 & 1 \end{pmatrix},$$

所以

$$B = C^{-1}AC = \begin{pmatrix} 1 & 5 & 4 \\ 0 & -2 & 2 \\ 1 & -2 & 4 \end{pmatrix}.$$

例 29 设单位向量 $u = \begin{pmatrix} \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \end{pmatrix}^T$, 定义 \mathbb{R}^3 上正交投影变换 P 为

$$P(x) = x - (x, u)u,$$

求 P 在自然基 e_1, e_2, e_3 下的矩阵 A .

解 由(1.23)式, A 的第 i 列是 e_i 的像 $P(e_i)$ 在自然基 e_1, e_2, e_3 下坐标, 即为 $P(e_i)$, 又

$$P(e_1) = e_1 - (e_1, u)u = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \frac{2}{9} \begin{pmatrix} 2 \\ -2 \\ -1 \end{pmatrix} = \frac{1}{9} \begin{pmatrix} 5 \\ 4 \\ 2 \end{pmatrix},$$

$$P(e_2) = e_2 - (e_2, u)u = \frac{1}{9} \begin{pmatrix} 4 \\ 5 \\ -2 \end{pmatrix}, \quad P(e_3) = \frac{1}{9} \begin{pmatrix} 2 \\ -2 \\ 8 \end{pmatrix}.$$

所以 $A = (P(e_1) \ P(e_2) \ P(e_3)) = \frac{1}{9} \begin{pmatrix} 5 & 4 & 2 \\ 4 & 5 & -2 \\ 2 & -2 & 8 \end{pmatrix}.$

三、不变子空间

为了分析线性变换的矩阵化简与空间分解之间的联系, 我们借助于不变子空间的概念为基础来展开讨论.

定义 1.14 设 T 是线性空间 $V_n(F)$ 上的线性变换, W 是 $V_n(F)$ 的子空间, 如果 $\forall \alpha \in W$, 有 $T(\alpha) \in W$, 即值域 $T(W) \subseteq W$, 则称 W 是 T 的不变子空间.

例 30 对 \mathbb{R}^3 上正交投影 $P(x) = x - (x, u)u$, u 为单位向量, \mathbb{R}^3 的子空间 $W = L\{u\}$ 和 $W^\perp = \{x \mid (x, u) = 0\}$ 都是 P 的不变子空间.

证明 $\forall X \in L\{u\}$, 则

$$X = ku, \quad P(X) = ku - (ku, u)u = ku - ku = 0 \in L\{u\},$$

所以, $L\{u\}$ 是 P 的不变子空间.

又 $\forall X \in W^\perp$,

$$P(X) = X - (X, u)u = X \in W^\perp,$$

所以 W^\perp 也是 P 的不变子空间.

由单位向量 u 定义的正交投影 P 在 $L\{W\}$ 上是零变换, 在 W^\perp 上为恒等变换.

一般, 如果 W 是 $V_n(F)$ 的子空间, 由于 $T(W) \subseteq W$, 因此 T 可看做是线性空间 W 上的线性变换.

设 $W = L\{\alpha_1, \alpha_2, \dots, \alpha_r\}$, 则 $T(W) = L\{T(\alpha_1), T(\alpha_2), \dots, T(\alpha_r)\}$, $T(W) \subseteq W$ 的充分必要条件是 $T(\alpha_i) \in W$, 即 W 是 T 的不变子空间的充分必要条件是 $T(\alpha_i) \in W$, $i=1, 2, \dots, r$.

设 $\alpha_1, \alpha_2, \dots, \alpha_r$ 是 W 的基, U 为 W 的直和补子空间, 且 $U = L\{\beta_{r+1}, \beta_{r+2}, \dots, \beta_s\}$, 则

有等式

$$V_n(F) = W \oplus U.$$

$\{\alpha_1, \alpha_2, \dots, \alpha_r, \beta_{r+1}, \dots, \beta_s\}$ 为 $V_n(F)$ 的一组基, 当 W 是 T 的不变子空间时, 有

$$T(\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_r \quad \beta_{r+1} \quad \cdots \quad \beta_s) = (\alpha_1 \quad \cdots \quad \alpha_r \quad \beta_{r+1} \quad \cdots \quad \beta_s) \begin{pmatrix} A_1 & A_2 \\ 0 & A_3 \end{pmatrix},$$

其中 A_1 是 T 作为 W 上线性变换时, 在基 $\{\alpha_1, \alpha_2, \dots, \alpha_r\}$ 下的矩阵. 进一步, 如果 U 也是 T 的不变子空间, 则

$$T(\alpha_1 \quad \cdots \quad \alpha_r \quad \beta_{r+1} \quad \cdots \quad \beta_s) = (\alpha_1 \quad \cdots \quad \alpha_r \quad \beta_{r+1} \quad \cdots \quad \beta_s) \begin{pmatrix} A_1 & 0 \\ 0 & A_3 \end{pmatrix}.$$

如果

$$V_n(F) = W_1 \oplus W_2 \oplus \cdots \oplus W_s, \quad W_i := L\{\xi_i\},$$

且 W_i 均为 T 的不变子空间, 就有

$$T(\xi_1 \quad \xi_2 \quad \cdots \quad \xi_s) = (\xi_1 \quad \xi_2 \quad \cdots \quad \xi_s) \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_s \end{pmatrix}. \quad (1.26)$$

线性变换的矩阵分解为准对角形是与空间分解为不变子空间的直和相应的问题.

一方面, 如果 $V = W_1 \oplus W_2 \oplus \cdots \oplus W_s$, W_i 为 T 的不变子空间, 取 W_i 的基

$$\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{in_i}, \quad i = 1, 2, \dots, s,$$

用它们构成空间的基, T 在这组基下的矩阵为准对角矩阵

$$A = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_s \end{pmatrix}.$$

其中 A_i 是 T 在不变子空间 W_i 上的矩阵, 即

$$T(\alpha_{i1} \quad \alpha_{i2} \quad \cdots \quad \alpha_{in_i}) = (\alpha_{i1} \quad \alpha_{i2} \quad \cdots \quad \alpha_{in_i})A_i, \quad i = 1, 2, \dots, s.$$

另一方面, 如在基 $\{\alpha_1, \alpha_2, \dots, \alpha_s\}$ 下矩阵为上述准对角形, A_i 为 n_i 阶矩阵. 按 n_i 依次把基分成 s 组:

$$\{\alpha_{11}, \alpha_{12}, \dots, \alpha_{1n_1}; \alpha_{21}, \alpha_{22}, \dots, \alpha_{2n_2}; \dots; \alpha_{s1}, \alpha_{s2}, \dots, \alpha_{sn_s}\},$$

则 $W_i := L\{\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{in_i}\}$ 就是 T 的不变子空间, $i = 1, 2, \dots, s$.

四、正交变换与酉变换

正交变换与酉变换是内积空间 $[V_n(F); (\alpha, \beta)]$ 中应用广泛的重要变换, 这里我们定义正交变换, 并给出它们的主要性质.

定义 1.15 设 T 为内积空间 $[V_n(F); (\alpha, \beta)]$ 上的线性变换, 如果 T 不改变向量的

内积, 即 $\forall \alpha, \beta \in [V_n(F); (\alpha, \beta)]$, 都有

$$(T(\alpha), T(\beta)) = (\alpha, \beta),$$

则称 T 为内积空间上的正交变换. 当空间为欧氏空间时称 T 为正交变换; 若空间为酉空间, 则称 T 为酉变换. 正交(酉)变换在标准正交基下的矩阵称为正交(酉)矩阵.

正交(酉)变换的定义有几种等价的形式, 我们把它们归纳在下面的定理中.

定理 1.15 设 T 是内积空间上的线性变换, 则下列命题等价:

- (1) T 是正交(酉)变换;
- (2) T 保持向量长度不变;
- (3) T 把空间 $V_n(F)$ 的标准正交基变换为标准正交基;
- (4) 正交变换关于任一标准正交基的矩阵 C 满足 $C^T C = CC^T = I$; 酉变换关于任一标准正交基的矩阵 U 满足 $U^H U = UU^H = I$.

证明 采用循环证明方法完成等价性证明.

$$(1) \Rightarrow (2) \text{ 由 (1) } \forall \alpha, \beta \in V_n(F), (T(\alpha), T(\beta)) = (\alpha, \beta),$$

$$\text{则有 } (T(\alpha), T(\alpha)) = (\alpha, \alpha),$$

$$\text{即 } \forall \alpha \in V_n(F), \|\alpha\| = \|T(\alpha)\|, \text{ 命题(2)得证.}$$

$$(2) \Rightarrow (3) \text{ 由 (2) } \forall \alpha, \beta \in V_n(F) \text{ 有}$$

$$(T(\alpha + \beta), T(\alpha + \beta)) = (\alpha + \beta, \alpha + \beta),$$

由内积定义, 再利用 $(T(\alpha), T(\alpha)) = (\alpha, \alpha)$, $(T(\beta), T(\beta)) = (\beta, \beta)$ 消去等式两边相同的项, 得

$$(T(\alpha), T(\beta)) + (T(\beta), T(\alpha)) = (\alpha, \beta) + (\beta, \alpha), \quad (1.27)$$

从而

$$(T(\alpha), T(\beta)) + \overline{(T(\alpha), T(\beta))} = (\alpha, \beta) + \overline{(\alpha, \beta)},$$

即

$$\operatorname{Re}(T(\alpha), T(\beta)) = \operatorname{Re}(\alpha, \beta).$$

又

$$T(|\alpha + \beta|) = |\alpha + \beta|,$$

得到

$$(iT(\alpha), T(\beta)) + \overline{(iT(\alpha), T(\beta))} = (i\alpha, \beta) + \overline{(i\alpha, \beta)},$$

由已知结果

$$\operatorname{Re}[i(T(\alpha), T(\beta))] = \operatorname{Re}[i(\alpha, \beta)],$$

利用已得结果, 就有

$$\operatorname{Im}(T(\alpha), T(\beta)) = \operatorname{Im}(\alpha, \beta).$$

从而, 从式(1.28)得到 $\forall \alpha, \beta \in V_n(F); (T(\alpha), T(\beta)) = (\alpha, \beta)$.

再取 $V_n(F)$ 的标准正交基 $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$, 在 T 下它们的像为 $\{T(\varepsilon_1), T(\varepsilon_2), \dots, T(\varepsilon_n)\}$

$$(T(\varepsilon_i), T(\varepsilon_j)) = (\varepsilon_i, \varepsilon_j) = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$$

所以 $\{T(\varepsilon_1), T(\varepsilon_2), \dots, T(\varepsilon_n)\}$ 也是 $V_n(F)$ 的标准正交基, 即(3)得证.

$(3) \Rightarrow (4)$ 设 $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$ 为欧氏空间 $V_n(R)$ 的标准正交基.

$$T(\varepsilon_1 \ \varepsilon_2 \ \cdots \ \varepsilon_n) = (\varepsilon_1 \ \varepsilon_2 \ \cdots \ \varepsilon_n)C$$

把矩阵 C 按列分块为 $(C_1 \ C_2 \ \cdots \ C_n)$, 则

$$T(\varepsilon_i) = (\varepsilon_1 \ \varepsilon_2 \ \cdots \ \varepsilon_n)C_i,$$

从而 $(T(\varepsilon_i), T(\varepsilon_j)) = C_i^T C_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$

所以

$$C^T C = I_n.$$

当 $V_n(F)$ 为酉空间时, 注意到内积的共轭对称, 即可证 $U^H U = I_n$.

(4) \Rightarrow (1) 取欧氏空间标准正交基 $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$, 由(4)

$$T(\varepsilon_1 \ \varepsilon_2 \ \cdots \ \varepsilon_n) = (\varepsilon_1 \ \varepsilon_2 \ \cdots \ \varepsilon_n)C; \quad C^T C = I_n,$$

$$\forall \alpha \in V_n(R), \alpha = (\varepsilon_1 \ \varepsilon_2 \ \cdots \ \varepsilon_n)X, \beta = (\varepsilon_1 \ \varepsilon_2 \ \cdots \ \varepsilon_n)Y,$$

就有

$$T(\alpha) = T[(\varepsilon_1 \ \varepsilon_2 \ \cdots \ \varepsilon_n)X] = (\varepsilon_1 \ \varepsilon_2 \ \cdots \ \varepsilon_n)CX,$$

$$T(\beta) = (\varepsilon_1 \ \varepsilon_2 \ \cdots \ \varepsilon_n)CY,$$

$$(T(\alpha), T(\beta)) = (CX)^T CY = X^T Y = (\alpha, \beta),$$

即 T 为正交变换.

□

同理可证酉变换情形.

在标准正交基下, 正交(酉)变换是与正交(酉)矩阵 $C(U)$ 联系在一起的, 正交矩阵是在线性代数中介绍过的矩阵, 为了应用方便, 我们把它们的性质一并列出, 证明留做练习.

定理 1.16 正交矩阵 C 和酉矩阵 U 有如下性质:

(1) 正交矩阵的行列式为 ± 1 ; 酉矩阵的行列式的模长为 1.

(2) $C^{-1} = C^T; U^{-1} = U^H$.

(3) 正交(酉)矩阵的逆矩阵与乘积仍然是正交(酉)矩阵.

(4) n 阶正交(酉)矩阵的列和行向量组是欧氏(酉)空间 $R^n(C)$ 中的标准正交基.

作为例子, 我们给出常见的正交变换.

例 31 R^2 上绕原点逆时针旋转 θ 角的线性变换 T_θ 为正交变换.

解 由解析几何知, $T e_1 = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}, T e_2 = \begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix}$,

所以 $T(e_1, e_2) = (e_1, e_2) \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$.

矩阵 $\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$ 是正交矩阵, 所以 T_θ 为正交变换.

例 32 如图 1.1 所示, 空间 R^3 上绕过原点的直线 L 旋转 θ 角的变换 $T_{L\theta}$ 为正交变换.

解 设 u_1 为直线 L 的方向矢量, u_1 为单位矢量. 令 $W = L\{u_1\}$, W 的正交补子空间记为 $W^\perp = L\{u_2, u_3\}$, $\{u_2, u_3\}$ 为 W^\perp 的标准正交基. 所以, $\{u_1, u_2, u_3\}$ 为 R^3 的一组标准正交基.

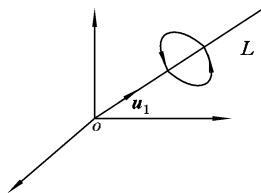


图 1.1

$\mathbb{R}^3 = W \oplus W^\perp$,
由 $T_{L\theta}$ 的定义, W 与 W^\perp 均为 $T_{L\theta}$ 的不变子空间.

在 W 上, $T_{L\theta}$ 为恒等变换, 在 $\{u_1\}$ 下矩阵为(1). 在 W^\perp 上,
 $T_{L\theta}$ 为平面 $L\{u_2, u_3\}$ 上的逆时针旋转. 在 $\{u_2, u_3\}$ 下矩阵为

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

由空间分解可知 $T_{L\theta}$ 在基 $\{u_1, u_2, u_3\}$ 下矩阵为

$$C = \begin{pmatrix} 1 & & \\ & \ddots & \\ \cos\theta & -\sin\theta & \\ \sin\theta & \cos\theta & \end{pmatrix},$$

C 为正交矩阵. 由定理 1.15, $T_{L\theta}$ 为正交变换.

例 33 设 u_1 为 \mathbb{R}^n 的单位矢量, 线性变换

$$S: \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad \forall X \in \mathbb{R}^n,$$

$$S(X) = X - 2(X, u_1)u_1,$$

称 S 为 \mathbb{R}^n 上的镜像变换, 证明 S 为正交变换.

证明 由 $S(X)$ 的定义, 设 $W = L\{u_1\}$, 则 $W^\perp = L\{u_2, u_3, \dots, u_n\}$, 其中 $\{u_1, u_2, \dots, u_n\}$ 为 \mathbb{R}^n 的标准正交基, 则 $\mathbb{R}^n = W \oplus W^\perp$,

$$S(u_1) = -u_1, \quad S(u_i) = u_i, \quad i = 2, 3, \dots, n.$$

所以 W 和 W^\perp 都是 S 的不变子空间.

$$\forall \alpha \in W, \quad S(\alpha) = -\alpha; \quad \forall \alpha \in W^\perp, \quad S(\alpha) = \alpha,$$

所以, 在标准正交基 $\{u_1, u_2, \dots, u_n\}$ 下, S 的矩阵为 $\begin{pmatrix} -1 & & \\ & \ddots & \\ & & I_{n-1} \end{pmatrix}$ 是正交矩阵, 即 S 为正交变换.

我们称 S 为对子空间 W^\perp 的镜像变换. 当 $n=3$ 时, \mathbb{R}^3 中 S 是相对于以 u_1 为法矢量的平面的镜像变换, 它在 $L\{u_1\}$ 上是 $\lambda=-1$ 的相似变换, 在 W^\perp 上为恒等变换.

五、线性空间 $V_n(F)$ 到线性空间 $V_m(F)$ 的线性变换

前面讨论的线性变换是建立在同一个线性空间上的线性变换. 在实际应用中, 许多问题会涉及从一个线性空间到另一个线性空间的线性变换. 这一部分将把线性变换推广到同一数域上的任意两个空间上, 建立相应概念及矩阵处理.

定义 1.16 设 $V_n(F)$ 和 $V_m(F)$ 是同一个数域 F 上的两个线性空间, 变换 $T: V_n(F) \rightarrow V_m(F)$, 即 $\forall \alpha \in V_n(F), T(\alpha) \in V_m(F)$. 如果变换 T 满足条件:

$\forall \alpha, \beta \in V_n(F)$, 数 $k \in F$, 有

$$T(\alpha + \beta) = T(\alpha) + T(\beta),$$

$$T(k\alpha) = kT(\alpha).$$

则称 T 是从 $V_n(F)$ 到 $V_m(F)$ 的一个线性变换.

例 34 设 $A_{m \times n}$ 是空间 $F^{n \times n}$ 中任意一个给定的矩阵, 定义 $T_A: F^n \rightarrow F^m$ 如下:

$$\forall X \in F^n, \quad T_A(X) = AX,$$

则 T_A 是从 F^n 到 F^m 的一个线性变换.

例 35 定义 $T: R^{n \times n} \rightarrow R$, 使 $\forall A \in R^{n \times n}, T(A) = \det(A)$ (A 的行列式), 证明 T 不是一个线性变换.

证明 因为 $\forall A, B \in R^{n \times n}$, 一般有:

$$\det(A + B) \neq \det(A) + \det(B),$$

从而

$$T(A + B) \neq T(A) + T(B),$$

故 T 不是从 $R^{n \times n}$ 到 R 的线性变换.

可类似于 $V_n(F)$ 上的线性变换的讨论来建立 $V_n(F)$ 到 $V_m(F)$ 的线性变换的变换矩阵.

设 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 是 $V_n(F)$ 的一组基, $\{\xi_1, \xi_2, \dots, \xi_m\}$ 是 $V_m(F)$ 的一组基, 则由 $\forall \alpha_i \in V_n(F), T(\alpha_i) \in V_m(F)$, 可得到表示式

$$\left\{ \begin{array}{l} T(\alpha_1) = a_{11}\xi_1 + a_{12}\xi_2 + \dots + a_{1n}\xi_n, \\ T(\alpha_2) = a_{21}\xi_1 + a_{22}\xi_2 + \dots + a_{2n}\xi_n, \\ \vdots \\ T(\alpha_n) = a_{n1}\xi_1 + a_{n2}\xi_2 + \dots + a_{nn}\xi_n. \end{array} \right. \quad (1.28)$$

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \in F^{m \times n},$$

取矩阵

式(1.28)可写为

$$T(\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n) = (\xi_1 \ \xi_2 \ \cdots \ \xi_n)A.$$

定义 A 为 T 的变换矩阵. 当 T 是 $V_n(F)$ 到 $V_m(F)$ 的线性变换时, A 是数域 F 上的一个 $m \times n$ 阶矩阵.

设 $T: V_n(F) \rightarrow V_m(F)$ 是线性变换, 同样可定义两个子空间: 像空间 $R(T)$ 和零空间 $N(T)$ 如下:

$$R(T) = \{T(\alpha) \mid \alpha \in V_n(F)\} \subseteq V_m(F),$$

$$N(T) = \{\alpha \in V_n(F) \mid T(\alpha) = 0\} \subseteq V_n(F).$$

它们的维数具有如下关系.

定理 1.17 设 T 是 n 维线性空间 $V_n(F)$ 到 m 维线性空间 $V_m(F)$ 的线性变换, 则

$$\dim R(T) + \dim N(T) = n.$$

证明 设 $\dim N(T) = t$.

取 $N(T)$ 的基 $\{\alpha_1, \alpha_2, \dots, \alpha_t\} \subseteq V_n(F)$, 将它们扩充为空间 $V_n(F)$ 的基: $\{\alpha_1, \alpha_2, \dots, \alpha_t, \beta_{t+1}, \dots, \beta_n\}$, 它们在变换 T 下的像集合为

$$\begin{aligned} & \{T(\alpha_1), T(\alpha_2), \dots, T(\alpha_t), T(\beta_{t+1}), \dots, T(\beta_n)\} \\ &= \{0, 0, \dots, T(\beta_{t+1}), T(\beta_{t+2}), \dots, T(\beta_n)\} \subseteq V_m(F). \end{aligned}$$

$\forall \eta \in R(T)$, $\exists \alpha \in V_n(F)$, $\eta = T(\alpha)$, 用基表示为

$$\alpha = \sum_{i=1}^t k_i \alpha_i + \sum_{j=t+1}^n k_j \beta_j,$$

从而 $\eta = T(\alpha) = \sum_{i=1}^t k_i T(\alpha_i) + \sum_{j=t+1}^n k_j T(\beta_j) = \sum_{j=t+1}^n k_j T(\beta_j)$,

即 $R(T) = L\{T(\beta_{t+1}), T(\beta_{t+2}), \dots, T(\beta_n)\}.$

又任取数 l_i , 使

$$\sum_{j=t+1}^n l_j T(\beta_j) = 0.$$

由线性变换的定义, 上式即为 $T\left(\sum_{j=t+1}^n l_j \beta_j\right) = 0$. 又由线性变换的性质, 得

$$\sum_{j=t+1}^n l_j \beta_j = 0.$$

而 $\{\beta_j\}_{j=t+1}^n$ 线性无关. 从而 $l_j = 0, j = t+1, \dots, n$, 故 $\{T(\beta_{t+1}), \dots, T(\beta_n)\}$ 线性无关. 它们就是空间 $R(T)$ 的基. 因此, $\dim R(T) = n - t$,

即

$$\dim R(T) + \dim N(T) = n.$$

□

类似于 $V_n(F)$ 上线性变换的讨论, 可以建立 $T: V_n(F) \rightarrow V_m(F)$, 在不同基下变换矩阵的关系式.

设在两个线性空间中, 不同的基分别是

$$V_n(F): \begin{cases} \{\alpha_1, \alpha_2, \dots, \alpha_n\} \\ \{\beta_1, \beta_2, \dots, \beta_n\} \end{cases}; \quad V_m(F): \begin{cases} \{\xi_1, \xi_2, \dots, \xi_m\} \\ \{\eta_1, \eta_2, \dots, \eta_m\} \end{cases}.$$

过渡矩阵分别为 $P_{n \times n}$ 和 $Q_{m \times m}$ 即

$$\begin{aligned} (\beta_1 & \beta_2 \cdots \beta_n) = (\alpha_1 \alpha_2 \cdots \alpha_n)P, \\ (\eta_1 & \eta_2 \cdots \eta_m)Q = (\xi_1 \xi_2 \cdots \xi_m). \end{aligned}$$

又

$$T(\alpha_1 \alpha_2 \cdots \alpha_n) = (\xi_1 \xi_2 \cdots \xi_m)A_{m \times n},$$

$$T(\beta_1 \beta_2 \cdots \beta_n) = (\eta_1 \eta_2 \cdots \eta_m)B_{m \times n},$$

则因为

$$\begin{aligned} T(\beta_1 \beta_2 \cdots \beta_n) &= T[(\alpha_1 \alpha_2 \cdots \alpha_n)P] = T(\alpha_1 \alpha_2 \cdots \alpha_n)P \\ &= (\xi_1 \xi_2 \cdots \xi_m)AP = (\eta_1 \eta_2 \cdots \eta_m)QAP. \end{aligned}$$

因此,

$$B = QAP.$$

从以上结果可以看到, 空间 $V_n(F)$ 到 $V_m(F)$ 的线性变换在不同基下的矩阵是等价的.

习 题 一

1. 判断下列集合对指定的运算是否构成 \mathbb{R} 上线性空间.

$$(1) V_1 = \{A = (a_{ij})_{n \times n} \mid \sum_{i=1}^n a_{ii} = 0\}; \text{ 对矩阵加法和数乘运算.}$$

$$(2) V_2 = \{A \mid A \in \mathbb{R}, A^T = -A\}; \text{ 对矩阵加法和数乘运算.}$$

$$(3) V_3 = \mathbb{R}^3; \text{ 对 } \mathbb{R}^3 \text{ 中向量加法和如下定义的数乘向量: } \forall \alpha \in \mathbb{R}^3, k \in \mathbb{R}, k\alpha = 0.$$

$$(4) V_4 = \{f(x) \mid f(x) \geq 0\}, \text{ 通常的函数加法与数乘函数运算.}$$

2. 求线性空间 $W = \{A \in \mathbb{R}^{n \times n} \mid A^T = A\}$ 的维数和一组基.

3. 判断下列子集哪些是 \mathbb{R}^3 的子空间.

$$(1) W_1 = \{(x_1 \ x_2 \ x_3)^T \mid x_1 + x_2 = 0\};$$

$$(2) W_2 = \{(x_1 \ x_2 \ x_3)^T \mid x_1 + x_2 + x_3 = 1\};$$

$$(3) W_3 = \{(2x_1 + x_2 \ x_1 - x_3 \ 2x_1 + 3x_2)^T \mid x_i \in \mathbb{R}\};$$

$$(4) W_4 = \{(x_1 \ x_2 \ x_3)^T \int_0^1 (x_1 S^2 + x_2 S + x_3) dS = 0\}.$$

4. 如果 U_1, U_2 都是线性空间 V 的子空间, 若 $\dim U_1 = \dim U_2$, 而且 $U_1 \subseteq U_2$, 证明: $U_1 = U_2$.

5. 设 $A \in \mathbb{R}^{n \times n}$, 证明 $\dim R(A) + \dim N(A) = n$.

$$6. \text{ 设 } A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 3 \\ 3 & 1 & 5 \end{pmatrix}, \text{ 讨论向量 } \alpha = (2 \ 3 \ 4)^T \text{ 是否在 } R(A) \text{ 中.}$$

7. 设 $A \in \mathbb{R}^{n \times n}$, 证明下列各条等价:

(1) A 是可逆矩阵;

(2) $N(A) = \{0\}$;

(3) $R(A) = \mathbb{R}^n$.

8. 讨论线性空间 $P_4[x]$ 中向量 $p_1 = x^3 + x^2 + x + 1, p_2 = 2x^3 - x^2 + 3x, p_3 = 4x^3 + x^2 + 5x + 2$ 的线性相关性.

$$9. \text{ 设 } A = \begin{pmatrix} 1 & -1 & 3 & 0 \\ -2 & 1 & -2 & 1 \\ -1 & -1 & 5 & 2 \end{pmatrix}, \text{ 求矩阵 } A \text{ 的列空间 } R(A) \text{ 和零空间 } N(A).$$

$$10. \text{ 设 } \alpha_1 = (1 \ 2 \ 1 \ 0)^T, \alpha_2 = (-1 \ 1 \ 1 \ 1)^T, \beta_1 = (2 \ -1 \ 0 \ 1)^T, \beta_2 = (1 \ -1 \ 3 \ 7)^T, W_1 = L\{\alpha_1, \alpha_2\}, W_2 = L\{\beta_1, \beta_2\}, \text{ 求 } W_1 \cap W_2 \text{ 和 } W_1 + W_2.$$

11. 在 $\mathbb{R}^{2 \times 2}$ 中, 已知两组基为

$$E_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad E_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad E_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

$$G_1 = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \quad G_2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad G_3 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad G_4 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix},$$

求基 $\{E_i\}$ 到基 $\{G_i\}$ 的过渡矩阵 C , 并求矩阵 $\begin{pmatrix} 0 & 1 \\ 2 & -3 \end{pmatrix}$ 在基 $\{G_i\}$ 下的坐标 X .

12. 设 $V_3(F)$ 中两组基 $\{\alpha_1, \alpha_2, \alpha_3\}$ 和 $\{\beta_1, \beta_2, \beta_3\}$. 已知 $\beta_1 = \alpha_1 - \alpha_2, \beta_2 = 2\alpha_1 + 3\alpha_2 + 2\alpha_3, \beta_3 = \alpha_1 + 3\alpha_2 + 2\alpha_3$.

(1) 求从基 $\{\alpha_i\}$ 到基 $\{\beta_i\}$ 的过渡矩阵 C ;

(2) 求 $\alpha = 2\beta_1 - \beta_2 + 3\beta_3$ 在基 $\{\alpha_1, \alpha_2, \alpha_3\}$ 下的坐标 X .

13. 设 $A_{m \times n}$ 和 $B_{n \times m}$ 满足 $AB = 0$, 证明

(1) $R(B) \subseteq N(A)$,

(2) $R(A^T) \subseteq N(B^T)$.

14. 在内积空间 $[V_n(F); (\alpha, \beta)]$ 中. 设 $\alpha, \beta, \gamma \in V_n(F), k \in F$, 证明

$$(\alpha, k\beta) = k(\alpha, \beta); \quad (\alpha, \beta + \gamma) = (\alpha, \beta) + (\alpha, \gamma).$$

15. 设 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 是 n 维线性空间 $V_n(R)$ 中的基, $\forall \alpha = \sum_{i=1}^n x_i \alpha_i, \beta = \sum_{i=1}^n y_i \alpha_i \in V_n(R)$, 定义到 $V_n(R)$ 实数集 R 的二元运算:

$$\mu(\alpha, \beta) = x_1 y_1 + 2x_2 y_2 + \dots + nx_n y_n,$$

证明 $\mu(\alpha, \beta)$ 是 $V_n(R)$ 上的内积.

16. 在线性空间 R^n 中, $\forall X, Y \in R^n$, 定义内积

$$(X, Y) = Y^T A X,$$

其中 A 为一个给定的实对称的正定矩阵, $A \in R^{n \times n}$, 试给出 Cauchy 不等式关于这种内积的具体形式.

17. 在欧氏空间 $V_n(R)$ 中, 证明向量组 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 线性无关的充要条件是

$$\left| \begin{array}{cccc} (\alpha_1, \alpha_1) & (\alpha_1, \alpha_2) & \cdots & (\alpha_1, \alpha_n) \\ (\alpha_2, \alpha_1) & (\alpha_2, \alpha_2) & \cdots & (\alpha_2, \alpha_n) \\ | & | & & | \\ (\alpha_n, \alpha_1) & (\alpha_n, \alpha_2) & \cdots & (\alpha_n, \alpha_n) \end{array} \right| \neq 0.$$

18. 设 $\{\varepsilon, \varepsilon, \dots, \varepsilon\}$ 是 $V_5(F)$ 的标准正交基, 又 $\alpha_1 = \varepsilon + \varepsilon, \alpha_2 = \varepsilon - \varepsilon + \varepsilon, \alpha_3 = 2\varepsilon + \varepsilon + \varepsilon$, 求 $W = L\{\alpha_1, \alpha_2, \alpha_3\}$ 的标准正交基 $\{\beta_1, \beta_2, \beta_3\}$.

19. 设 α, β 是内积空间 $[V_n(F), (\alpha, \beta)]$ 中彼此正交的向量, 证明

$$\|\alpha + \beta\|^2 = \|\alpha\|^2 + \|\beta\|^2.$$

20. 在欧氏空间 R^4 中, 求子空间

$W = L\{(1 \ 1 \ -1 \ 1)^T, (1 \ -1 \ -1 \ 1)^T\}$ 的正交补子空间 W^\perp .

21. 设 W_1 与 W_2 是欧氏空间的两个子空间, 证明

$$(W_1 + W_2)^\perp = W_1^\perp \cap W_2^\perp.$$

$$(W_1 \cap W_2)^\perp = W_1^\perp + W_2^\perp.$$

22. 设 W_1 与 W_2 是内积空间 $[V_n(F); (\alpha, \beta)]$ 的两个子空间.

证明 $W_1 \subseteq W_2 \Rightarrow W_2^\perp \subseteq W_1^\perp$.

23. 设矩阵 $A, B \in C^{n \times n}$, 证明子空间 $R(A)$ 与 $R(B)$ 是正交子空间, 当且仅当 $A^H B = 0$.

24. 判断下列变换哪些是线性变换

(1) \mathbb{R}^2 中: $T(x_1 \ x_2)^T = (x_1 + 1 \ x_2^2)^T$;

(2) \mathbb{R}^3 中: $T(x_1 \ x_2 \ x_3)^T = (x_1 + x_2 \ x_1 - x_2 \ 2x_3)^T$;

(3) $\mathbb{R}^{n \times n}$ 中: A 为给定的 n 阶方阵, $\forall X \in \mathbb{R}^{n \times n}$, $T(X) = AX + A$;

(4) $\mathbb{R}^{2 \times 2}$ 中: $T(A) = A^*$, A^* 为 A 的伴随矩阵.

25. 在 \mathbb{R}^3 中, 定义线性变换 $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$, $\forall \alpha = (x_1 \ x_2 \ x_3)^T$, $T(\alpha) = (2x_1 - x_2 \ 3x_2 - 2x_3 \ 2x_1 - x_2 + x_3)^T$, 求 T 在基 $e_1 = (1 \ 0 \ 0)^T$, $e_2 = (0 \ 1 \ 0)^T$, $e_3 = (0 \ 0 \ 1)^T$ 下的矩阵 A .

26. 在 $\mathbb{R}^{2 \times 2}$ 中定义线性变换, $\forall A \in \mathbb{R}^{2 \times 2}$,

$$T_1(A) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} A, \quad T_2(A) = A \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad T_3(A) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} A \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

其中 $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 是 $\mathbb{R}^{2 \times 2}$ 中给定的矩阵, 求

(1) 求 T_i 在基 $\{E_{11}, E_{12}, E_{21}, E_{22}\}$ 下的矩阵 X_i , $i=1, 2, 3$.

(2) 求 $T_1 + T_2$ 和 $T_1 \cdot T_2$ 在基 $\{E_{11}, E_{12}, E_{21}, E_{22}\}$ 下的矩阵.

27. 设线性变换 $T: V_n(F) \rightarrow V_n(F)$, 证明 T 是可逆变换的充要条件是 $N(T) = \{0\}$.

28. 设 \mathbb{R}^3 中线性变换 T 为: $\forall X = (x_1 \ x_2 \ x_3)^T$, $T(X) = (0 \ x_1 \ x_2)^T$,

(1) 求 T 的像子空间 $R(T)$ 和零空间 $N(T)$.

(2) 求 $R(T)$ 和 $N(T)$ 的维数与基.

29. 设 $\{\alpha_1, \alpha_2, \dots, \alpha_k\}$ 和 $\{\beta_1, \beta_2, \dots, \beta_k\}$ 是 n 维线性空间 $V_n(F)$ 的两个线性无关组. 证明一定存在 $V_n(F)$ 上可逆线性变换 T , 使

$$T(\alpha_i) = \beta_i, \quad i = 1, 2, \dots, k.$$

30. 设 T 为 \mathbb{R}^2 上线性变换, $\forall \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$, $T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}$,

求 T 的两个不变子空间 V_1 和 V_2 , 使 $\mathbb{R}^2 = V_1 \oplus V_2$.

31. 设 T 是欧氏空间 \mathbb{R}^3 上线性变换, 对 \mathbb{R}^3 中单位矢量 u , $\forall x \in \mathbb{R}^3$,

$$T(x) = x - (1 - k)(x, u)u,$$

(1) 求 T 的不变子空间;

(2) 把 \mathbb{R}^3 分解为不变子空间的直和, 并求相应的矩阵分解.

32. 设欧氏空间 \mathbb{R}^2 上压缩变换 T 定义为:

$$T(X) = X - \frac{1}{2}(X, e_2)e_2, \quad \forall X \in \mathbb{R}^2, \quad e_2 = (0 \ 1)^T,$$

求 \mathbb{R}^2 上单位圆: $x_1^2 + x_2^2 = 1$, 经过 T 变换后的图形.

33. 设欧氏空间上向量 $u = \begin{pmatrix} \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \end{pmatrix}^\top$, \mathbb{R}^3 上正交投影 P 如下定义

$$P(X) = X - (X, u)u,$$

求 P 的不变子空间, 把 \mathbb{R}^3 分解成不变子空间的直和, 并求相应矩阵分解.

34. 设 $\{\varepsilon_1, \varepsilon_2, \varepsilon_3\}$ 是欧氏空间 \mathbb{R}^3 中一个标准正交基, 试求一个正交变换 $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$, 使

$$T(\varepsilon_1) = \frac{2}{3}\varepsilon_1 + \frac{2}{3}\varepsilon_2 - \frac{1}{3}\varepsilon_3,$$

$$T(\varepsilon_2) = \frac{2}{3}\varepsilon_1 - \frac{1}{3}\varepsilon_2 + \frac{2}{3}\varepsilon_3.$$

第 2 章 Jordan 标准形介绍

这一章的讨论源于如何选择线性空间的基,使线性变换在该基下矩阵具有尽可能简单的形式这一问题.该问题在矩阵讨论上相应于矩阵相似变换化简.这一章将介绍矩阵相似的相关理论及其应用,并简要介绍 Jordan 标准形求法,介绍一些特殊矩阵的特征值和相似问题.

2.1 线性变换的对角矩阵表示

1.3 小节的讨论已为这里将要讨论的求一组基,使线性变换的矩阵形式最简奠定了基础.首先考虑线性变换矩阵为对角矩阵所需要的条件.

一、线性变换的特征值与特征向量

设 T 为线性空间 $V_n(F)$ 上的线性变换,由式(1.27)可知, T 在某组基 $\{\xi_1, \xi_2, \dots, \xi_n\}$

下矩阵为对角矩阵
$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$
 的充分必要条件是:基中向量 ξ_i 满足,

$$T(\xi_i) = \lambda_i \xi_i, \quad i=1, 2, \dots, n.$$

定义 2.1 设 T 为线性空间 V_n 上的线性变换,如果存在 $\xi \in V_n(F)$ 和数 $\lambda \in F$, $\xi \neq 0$,使得 $T(\xi) = \lambda \xi$,则称数 λ 为 T 的特征值,向量 ξ 为线性变换 T 的对应于特征值 λ 的特征向量.

为分析 λ 和 ξ 的求法,设对线性变换 T ,已知 T 在某一组基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 下的矩阵为 A (A 不一定为对角矩阵),设 ξ 是 T 的关于 λ 的特征向量,则有

$$\begin{aligned} \xi &= (\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n)X, \\ T(\xi) &= (\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n)AX, \end{aligned}$$

因此

$$T(\xi) = \lambda \xi \Leftrightarrow AX = \lambda X. \quad (2.1)$$

这说明 λ 是矩阵 A 的特征值, X 是矩阵 A 关于 λ 的特征向量.

定理 2.1 设 $V_n(F)$ 上线性变换 T 在基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 下矩阵为 A ,则 A 的特征值 λ 就是变换 T 的特征值;若 X 是 A 的特征向量,则 $\xi = (\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n)X$ 就是 T 的特征向量.

式(2.1)及其分析过程已基本上给出了定理 2.1 的证明.还应指出的是,矩阵 A 是

和基相关的,若 T 在另一组基下矩阵为 B ,从 1.3 小节知道 B 相似于 A ,即 $B = P^{-1}AP$ 或 $PB = AP$.

$$AX = \lambda X \Leftrightarrow B(P^{-1}X) = \lambda(P^{-1}X),$$

因此 B 与 A 的特征值是一样的,特征向量不一样.这说明 T 的特征值是由 T 决定的,和基与变换矩阵的选择无关.

应用定理 2.1,我们可以从 T 的一个变换矩阵 A 求得 T 的特征值与特征向量.计算步骤可归纳如下:

- (1) 选择 $V_n(F)$ 的基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$,求线性变换 T 关于该基的矩阵 A .
- (2) 求 A 的特征值:先求 A 的特征多项式: $f(\lambda) = |\lambda I - A|$, $f(\lambda) = 0$ 的根 $\lambda_1, \lambda_2, \dots, \lambda_n$ 即为 A 的全部特征值.
- (3) 求矩阵 A 关于 λ_i 的特征向量 X_i ,即 $(\lambda_i I - A)X_i = 0$ 的非零解,它们给出 T 的特征值 λ_i 对应的特征向量关于基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 的坐标.

从线性代数的学习中,大家是熟悉一个矩阵 A 的特征值与特征向量的求法与性质的.一般我们也称 A 的特征多项式 $f(\lambda) = |\lambda I - A|$ 为线性变换 T 的特征多项式.同样,在复数域上, $f(\lambda) = 0$ 有 n 个根,导出 n 维线性空间上的线性变换 T 有 n 个特征值.

下面,我们从空间的角度讨论线性变换 T 的特征向量的性质.

定义 2.2 设 λ 为线性变换 T 的特征值, $\xi_1, \xi_2, \dots, \xi_s$ 是 T 对应于 λ 的特征向量的极大线性无关组,则称子空间 $V_\lambda = L\{\xi_1, \xi_2, \dots, \xi_s\}$ 为 T 关于 λ 的特征子空间.

应该注意到 $V_\lambda = \{0\}$ 才给出线性变换 T 的全部特征向量.即 $\forall \alpha \in V_\lambda, \alpha \neq 0$ 时, α 就是 T 关于 λ 的特征向量, $V_\lambda = N(T - \lambda I)$.如果线性变换 T 有 s 个互异的特征值: $\lambda_1, \lambda_2, \dots, \lambda_s$, 它们就会对应 s 个特征子空间, $V_{\lambda_1}, V_{\lambda_2}, \dots, V_{\lambda_s}$. 特征子空间具有如下性质:

定理 2.2 设 $\lambda_1, \lambda_2, \dots, \lambda_s$ 是 $V_n(F)$ 上线性变换 T 的 s 个互异的特征值. V_{λ_i} 是 λ_i 的特征子空间, $i = 1, 2, \dots, s$, 则有

- (1) V_{λ_i} 是 T 的不变子空间;
- (2) $\lambda_i \neq \lambda_j$ 时, $V_{\lambda_i} \cap V_{\lambda_j} = \{0\}$;
- (3) 若 λ_i 是 T 的 k_i 重特征值, 则 $\dim V_{\lambda_i} \leq k_i$.

证明 (1) 由 V_{λ_i} 的定义, $\forall \alpha \in V_{\lambda_i}, T(\alpha) = \lambda_i \alpha \in V_{\lambda_i}$, 所以 $T(V_{\lambda_i}) \subseteq V_{\lambda_i}$, 即 V_{λ_i} 是 T 的不变子空间.

(2) $\forall \alpha \in V_{\lambda_i} \cap V_{\lambda_j}$, 则有 $\alpha \in V_{\lambda_i}$ 且 $\alpha \in V_{\lambda_j}$. 因此有, $T(\alpha) = \lambda_i \alpha$ 而且 $T(\alpha) = \lambda_j \alpha$. 两式相减得 $(\lambda_i - \lambda_j)\alpha = 0$, 又 $\lambda_i \neq \lambda_j$, 所以 $\alpha = 0$, 即

$$V_{\lambda_i} \cap V_{\lambda_j} = \{0\}, \quad \lambda_i \neq \lambda_j.$$

(3) 设 T 的矩阵为 A , 则 T 的特征多项式

$$f(\lambda) = |\lambda I - A| = (\lambda - \lambda_1)^{k_1}(\lambda - \lambda_2)^{k_2} \cdots (\lambda - \lambda_s)^{k_s},$$

$\lambda_i \neq \lambda_j, i \neq j$, 即 λ_i 为 T 的 k_i 重特征值. 反设 $\dim V_{\lambda_i} = t_i > k_i$, 则可取 V_{λ_i} 的基 $\{\xi_{i1}, \xi_{i2}, \dots, \xi_{it_i}\}$,

并把它扩充为 $V_n(F)$ 的基 $\{\xi_{i1}, \xi_{i2}, \dots, \xi_{it_i}, \eta_1, \eta_2, \dots, \eta_{n-t_i}\}$. $V_n = V_{\lambda_i} \oplus U$, 其中 $U = L\{\eta_1, \eta_2, \dots, \eta_{n-t_i}\}$. 由(1) V_{λ_i} 为 T 的不变子空间, 且 $T(\xi_i) = \lambda_i \xi_i$, 可得 T 在基 $\{\xi_{i1}, \xi_{i2}, \dots, \xi_{it_i}, \eta_1, \eta_2, \dots, \eta_{n-t_i}\}$ 下矩阵

$$B = \begin{pmatrix} \lambda_i I_{t_i} & C \\ 0 & D \end{pmatrix},$$

其中 I_{t_i} 为 t_i 阶单位矩阵, $\lambda_i I_{t_i}$ 是 T 作为 V_{λ_i} 上线性变换在基 $\{\xi_{i1}, \xi_{i2}, \dots, \xi_{it_i}\}$ 下的变换矩阵.

由 1.3 小节的讨论, B 和 A 相似, B 应与 A 有相同的特征多项式, 即

$$f(\lambda) = |\lambda I - A| = |\lambda I - B|.$$

又 $|\lambda I - B| = (\lambda - \lambda_i)^{t_i} q(\lambda)$, 这说明 λ_i 至少是 B 的 t_i 重根, 而 $t_i > k_i$, 与所设 $f(\lambda)$ 式矛盾. 因此反设不成立, 即 $\dim V_{\lambda_i} \leq k_i$.

定理 2.2 说明 V_{λ_i} 与 V_{λ_j} 满足 $V_{\lambda_1} + V_{\lambda_2} + \dots + V_{\lambda_s} = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \dots \oplus V_{\lambda_s} \subseteq V_n(F)$,

$$\sum_{i=1}^s \dim V_{\lambda_i} \leq \dim V_n(F) = n. \text{ 如果 } \sum_{i=1}^s \dim V_{\lambda_i} = n, \text{ 则有}$$

$$V_{\lambda_1} \oplus V_{\lambda_2} \oplus \dots \oplus V_{\lambda_s} = V_n(F).$$

例 1 求 $P_4[x]$ 上微分变换 $\frac{d}{dx}$ 的特征值与特征向量.

解 取 $P_4[x]$ 中基 $\{1, x, x^2, x^3\}$, 则 $\frac{d}{dx}$ 在该基下矩阵

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$|\lambda I - A| = \lambda^4$, 所以线性变换 $\frac{d}{dx}$ 的特征值

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0.$$

又从 $(0 \cdot I - A)X = 0$ 求得 A 关于 $\lambda = 0$ 的特征向量 $X = k \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, k \neq 0$, 所以 $\frac{d}{dx}$ 的特

征向量 $\xi = k \in P_4[x]$. ($k \neq 0$).

例 2 设矩阵 $A \in C^{n \times n}$, $B \in C^{m \times m}$, 证明 AB 和 BA 有相同的非零特征值.

证明 若 $m = n$, 且 A 为非奇异矩阵, 由于

$$A^{-1}(AB)A = BA,$$

因此, AB 与 BA 相似, 有相同的特征值.

对一般的情形, $AB \in C^{n \times m}$, $BA \in C^{m \times n}$. 取非奇异矩阵:

$$\begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix}, \quad \text{则} \quad \begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix}^{-1} = \begin{pmatrix} I_m & -A \\ 0 & I_n \end{pmatrix},$$

$$\text{又} \quad \begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix}^{-1} \begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix} \begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix},$$

$$\text{所以} \quad \left| \lambda I_{(m+n)} - \begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix} \right| = \left| \lambda I_{(m+n)} - \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix} \right|,$$

$$\text{从而} \quad \lambda^n |I - AB| = \lambda^m |I - BA|.$$

$$\text{对 } \lambda \neq 0, \text{ 有} \quad |I - AB| = 0 \Leftrightarrow |I - BA| = 0,$$

即 AB 与 BA 有一样的非零特征值, 且代数重数也相同.

二、线性变换矩阵的对角化

一个线性变换在空间不同基下的矩阵构成一个相似类. 设 $V_n(F)$ 上线性变换 T 在某组基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 下矩阵为 A , 则对 T 能否有基使它在其下矩阵为对角矩阵的问题, 等价于矩阵 A 能否相似于对角矩阵的问题. 方阵相似于对角矩阵的条件是大家已知的问题, 我们不再重复. 这里主要从变换 T 和空间分解的角度给出线性变换 T 有对角矩阵表示的充分必要条件.

定理 2.3 线性变换 T 有对角矩阵表示的充分必要条件是 T 有 n 个线性无关的特征向量.

证明 必要性: 设有基 $\{\xi_1, \xi_2, \dots, \xi_n\}$, 使 T 的矩阵为对角阵, 则有

$$T(\xi_1 \ \ \xi_2 \ \ \cdots \ \ \xi_n) = (\xi_1 \ \ \xi_2 \ \ \cdots \ \ \xi_n) \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}.$$

它等价于

$$T(\xi_i) = \lambda_i \xi_i, \quad i = 1, 2, \dots, n,$$

所以 λ_i 为 T 的特征值, $\{\xi_1, \xi_2, \dots, \xi_n\}$ 为 T 的 n 个线性无关的特征向量.

充分性: 若 T 有 n 个线性无关的特征向量 $\xi_1, \xi_2, \dots, \xi_n$, 则可取 $\{\xi_1, \xi_2, \dots, \xi_n\}$ 为线性空间 $V_n(F)$ 的基, T 在该基下的矩阵即为对角矩阵

$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix},$$

其中 λ_i 为 ξ_i 所对应的特征值.

从前面的讨论还可知, 特征子空间 V_{λ_i} 是 T 的不变子空间, 特别地, 线性变换 T 在不变子空间 V_{λ_i} 上矩阵为对角矩阵 $\lambda_i I_{n_i}$. 从空间分解角度, 我们有下列定理.

定理 2.4 线性变换 T 有对角矩阵表示的充分必要条件是：

$$V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_s} = V_n(F). \quad (2.2)$$

从定理 2.3 可以推出，(2.2) 式成立的充分必要条件是若 λ_i 是 T 的 k_i 重特征值，则必有 $\dim V_{\lambda_i} = k_i, i=1, 2, \dots, n$. 特别地，若 λ 为 T 的单特征值，则必有 $\dim V_{\lambda_i} = 1$.

推论 若线性变换 T 有 n 个互异的特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ ，则必有

$$V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_n} = V_n(F),$$

从而 T 有对角线上元素互异的对角矩阵表示.

更进一步，若 V_i 为 T 的一维不变子空间，则 $\forall \alpha \in V_i, T(\alpha) \in V_i$ ，所以 $T(\alpha) = \lambda_i \alpha$ ，即 T 在 V_i 上的矩阵是一阶对角矩阵，因此有：

推论 $V_n(F)$ 上线性变换有对角矩阵表示的充分必要条件是 $V_n(F)$ 可分解成 T 的一维不变子空间的直和.

例 3 证明 $P_4[x]$ 上的微分变换 $\frac{d}{dx}$ 没有对角矩阵表示.

证明 从例题 1 可知 $\lambda=0$ 是 $\frac{d}{dx}$ 的四重特征值. 而 $\dim V_{\lambda=0} = 1 < 4$ ，所以 T 没有对角矩阵表示.

当线性变换 T 有对角矩阵表示时，我们也常简称为线性变换 T 可对角化.

例 4 n 阶方阵 A ，若满足条件 $A^2 = A$ ，则称 A 为幂等矩阵；若满足 $A^2 = I$ ，则称 A 为乘方矩阵. 证明

(1) A 为幂等矩阵的充要条件是 A 相似于对角矩阵 $\begin{pmatrix} I_r & \\ & 0 \end{pmatrix}$ ，其中 r 为矩阵 A 的秩.

(2) A 为乘方矩阵的充要条件是 A 相似于对角矩阵 $\begin{pmatrix} I_s & \\ & -I_t \end{pmatrix}$ ，其中 $s+t=n$.

证明 题目的充分性是显然的，下证必要性.

(1) 设 λ 是幂等矩阵 A 的特征值，则从 $A^2 = A$ 可导出 λ 满足条件 $\lambda^2 = \lambda$ ，因此 $\lambda=1$ 或 0.

又 $\dim V_{\lambda=1} = n - \text{rank}(I-A)$ ， $\dim V_{\lambda=0} = n - \text{rank}(A)$.

从 $A^2 = A \Rightarrow A(I-A) = 0 \Rightarrow \text{rank}(A) + \text{rank}(I-A) \leq n$ ，

又 $A + (I-A) = I \Rightarrow \text{rank}(A) + \text{rank}(I-A) \geq \text{rank}(I) = n$.

上两个不等式导出结论

$$\text{rank}(A) + \text{rank}(I-A) = n.$$

从而 $\dim V_{\lambda=1} + \dim V_{\lambda=0} = n - [\text{rank}(I-A) + \text{rank}(A)] = n$ ，

亦即 A 有 n 个线性无关的特征向量. 从而 A 相似于对角矩阵 $\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$ ， $\lambda=1$ 或

0. 从 $\text{rank}(A) = \text{rank} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$ 得, 当 $\text{rank}(A) = r$ 时, A 相似于对角矩阵 $\begin{pmatrix} I_r & \\ 0 & \end{pmatrix}$.

(2) 从 $A^2 = I$ 可得出 A 的特征值 λ 满足等式: $\lambda^2 = 1$, 从而 $\lambda = 1$ 或 -1 .

又 $A^2 = I$ 时, A 为可逆矩阵, $\text{rank}(A) = n$.

$$\text{又 } A^2 = I \Rightarrow (I - A)(I + A) = 0 \Rightarrow \text{rank}(I - A) + \text{rank}(I + A) \leq n,$$

$$(I - A) + (I + A) = 2I \Rightarrow \text{rank}(I - A) + \text{rank}(I + A) \geq n,$$

从而

$$\text{rank}(I - A) + \text{rank}(I + A) = n.$$

$$\text{因此, } \dim V_{\lambda=1} + \dim V_{\lambda=-1} = n - \text{rank}(I - A) + n - \text{rank}(I + A) = n,$$

故 A 可相似于对角矩阵

$$\begin{pmatrix} I_s & \\ & -I_t \end{pmatrix},$$

其中

$$s + t = \dim V_{\lambda=1} + \dim V_{\lambda=-1} = n.$$

2.2 Jordan 矩阵介绍

从上节讨论我们已看到不是所有的线性变换都有对角矩阵表示. 因此, 对角矩阵不能像人们最初所希望的那样, 用以作为矩阵相似标准形. 当矩阵不能相似于对角矩阵时, 我们希望能找到形式尽可能简单一些的矩阵, 使任何方阵都能相似于这种矩阵, 这就是这里将要介绍的 Jordan 矩阵.

一、Jordan 矩阵

定义 2.3 形如

$$J(\lambda) = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix} \quad (2.3)$$

的 r 阶方阵称为一个 r 阶 Jordan 块. 由若干个 Jordan 块 $J_i(\lambda)$ 构成的准对角矩阵

$$J = \begin{pmatrix} J_1(\lambda_1) & & & \\ & J_2(\lambda_2) & & \\ & & \ddots & \\ & & & J_m(\lambda_m) \end{pmatrix}, \quad (2.4)$$

称为 Jordan 矩阵.

例如

$$J = \begin{pmatrix} 5 & 1 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix} = \begin{pmatrix} J_1(5) & & & & \\ & J_2(2) & & & \\ & & J_3(2) & & \end{pmatrix}$$

就是一个 5 阶 Jordan 矩阵, 它的三个 Jordan 块分别是

$$J_1(5) = \begin{pmatrix} 5 & 1 \\ & 5 \end{pmatrix}; \quad J_2(2) = \begin{pmatrix} 2 & 1 \\ & 2 \end{pmatrix}; \quad J_3(2) = \begin{pmatrix} 2 & 1 \\ & 2 \end{pmatrix}.$$

下面两个矩阵也是 Jordan 矩阵

$$J = \begin{pmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & 0 & \\ & & & 0 \end{pmatrix}; \quad J = \begin{pmatrix} 3 & & & \\ & 4 & & \\ & & 5 & \\ & & & 0 \end{pmatrix}.$$

前一个是含一个 Jordan 块的 Jordan 矩阵, 后一个矩阵由三个一阶 Jordan 块构成, 其中 $J_1(3) = (3), J_2(4) = (4), J_3(5) = (5)$.

Jordan 矩阵是准对角矩阵, 呈上三角阵形, 主对角线上是它的全部特征值. Jordan 块的特点是主对角线上元素相等, 紧邻上方元素 $a_{i,i+1} = 1$, 其余元素为 0. Jordan 矩阵中的非零元素只位于主对角线及其上方紧邻元素中, $a_{i,i+1} = 1$ 或 0. Jordan 块都是一阶的 Jordan 矩阵就是对角矩阵.

事实上, Jordan 矩阵就是相似标准形, 也就是说若在复有限维空间中考虑. 每个线性变换 T 都有一个 Jordan 矩阵表示, 每个复方阵都相似于一个 Jordan 矩阵. 讲清 Jordan 标准形的理论是需要相当多的篇幅的, 而与我们应用相关的常常是 Jordan 标准形的具体形式. 因此, 我们避开理论上的证明, 只给出下面的基本定理.

定理 2.5 在复数域上, 每个 n 阶方阵 A 都相似于一个 Jordan 矩阵, 即存在可逆矩阵 P , 使得

$$P^{-1}AP = J_A = \begin{pmatrix} J_1(\lambda_1) & & & \\ & J_2(\lambda_2) & & \\ & & \ddots & \\ & & & J_s(\lambda_s) \end{pmatrix},$$

$$\text{其中 } J_i(\lambda_i) = \begin{pmatrix} J_{i1}(\lambda_i) & & & \\ & J_{i2}(\lambda_i) & & \\ & & \ddots & \\ & & & J_{it_i}(\lambda_i) \end{pmatrix} \in C^{k_i \times k_i},$$

$J_{ij}(\lambda)$ 为 n_j 阶 Jordan 块, $\sum_{j=1}^{l_i} n_j = k_i$. $J_i(\lambda)$ 是 $k_i \times k_i$ 阶 Jordan 矩阵, $\sum_{i=1}^s k_i = n$. 若不计较 Jordan 块的排列次序, 则每个方阵的 Jordan 标准形 J_A 是惟一的.

定理 2.5 是 Jordan 标准形的存在定理, 它说明 n 维复线性空间上的线性变换 T , 都存在一组基, 使 T 在该基下矩阵为 Jordan 矩阵, 从而 Jordan 矩阵就是线性变换矩阵的最简形式.

二、Jordan 标准形的求法

对方阵 A , 我们需要确定它的 Jordan 矩阵 J_A 中 Jordan 块的结构和求出可逆矩阵 P , 使 $P^{-1}AP = J_A$. 求 J_A 和 P 的方法有多种, 这里介绍以存在定理为前提的分析确定法.

设 A 为 n 阶方阵, 由 A 相似于 J_A 可知, J_A 主对角线上元素就是 A 的全部特征值.

设 A 的特征多项式为

$$|A - \lambda I| = (\lambda - \lambda_1)^{k_1}(\lambda - \lambda_2)^{k_2} \cdots (\lambda - \lambda_s)^{k_s},$$

其中 λ 是 A 的 k_i 重特征值. $\sum_{i=1}^s k_i = n$, $\lambda_1, \lambda_2, \dots, \lambda_s$ 互异, 所以

$$J_A = \begin{pmatrix} J_1(\lambda_1) & & & \\ & J_2(\lambda_2) & & \\ & & \ddots & \\ & & & J_s(\lambda_s) \end{pmatrix}. \quad (2.5)$$

$J_i(\lambda)$ 是主对角元素为 λ 的 k_i 阶 Jordan 矩阵. 它是同一特征值 λ_i 对应的 Jordan 块放在一起得到的 Jordan 矩阵.

把相似变换的可逆矩阵 P 依(2.5)式所示 J_A 的结构, 相应取 k_1 列, k_2 列, \dots, k_s 列分块为 $P = (p_1 \ p_2 \ \cdots \ p_s)$, $p_i \in C^{n \times k_i}$, $i = 1, 2, \dots, s$, $AP = P J_A$ 可具体表示为:

$$A(p_1 \ p_2 \ \cdots \ p_s) = (p_1 \ p_2 \ \cdots \ p_s) \begin{pmatrix} J_1(\lambda_1) & & & \\ & J_2(\lambda_2) & & \\ & & \ddots & \\ & & & J_s(\lambda_s) \end{pmatrix},$$

即 $(Ap_1 \ Ap_2 \ \cdots \ Ap_s) = (p_1 J_1(\lambda_1) \ p_2 J_2(\lambda_2) \ \cdots \ p_s J_s(\lambda_s))$,

从而有 $Ap_i = p_i J_i(\lambda_i)$, $i = 1, 2, \dots, s$.

为表述简便, 不妨取 $Ap_1 = p_1 J_1(\lambda_1)$ 为代表来分析. 设

$$J_1(\lambda_1) = \begin{pmatrix} J_{11}(\lambda_1) & & & \\ & J_{12}(\lambda_1) & & \\ & & \ddots & \\ & & & J_{1t}(\lambda_1) \end{pmatrix},$$

其中 J_{1i} 为 n_i 阶 Jordan 块, $\sum_{i=1}^t n_i = k_1$, 由此可知

$$J_{1i}(\lambda_1) = \begin{pmatrix} \lambda_1 & 1 & & & \\ & \lambda_1 & 1 & & \\ & & \ddots & & \\ & & & \ddots & 1 \\ & & & & \lambda_1 \end{pmatrix}_{n_i \times n_i}. \quad (2.7)$$

再把 p_1 依 n_1 列, n_2 列, \cdots , n_t 列分块,

$$p_1 = (p_1^{(1)} \quad p_2^{(1)} \quad \cdots \quad p_t^{(1)}),$$

从(2.6)式有

$$Ap_j^{(1)} = p_j^{(1)} J_{1j}(\lambda_1), \quad j = 1, 2, \cdots, t. \quad (2.8)$$

设 $p_j^{(1)} = (\alpha_1 \quad \beta_2 \quad \cdots \quad \beta_{n_j})$, 结合(2.7)式, (2.8)式化为

$$A(\alpha_1 \quad \beta_2 \quad \cdots \quad \beta_{n_j}) = (\alpha_1 \quad \beta_2 \quad \cdots \quad \beta_{n_j}) \begin{pmatrix} \lambda_1 & 1 & & & \\ & \lambda_1 & 1 & & \\ & & \ddots & & \\ & & & \ddots & 1 \\ & & & & \lambda_1 \end{pmatrix}_{n_j \times n_j}.$$

该矩阵等式等价于由 n_j 个方程组成的方程组

$$\left\{ \begin{array}{l} (A - \lambda_1 I)\alpha_1 = 0, \\ (A - \lambda_1 I)\beta_2 = \alpha_1, \\ (A - \lambda_1 I)\beta_3 = \beta_2, \\ \vdots \\ (A - \lambda_1 I)\beta_{n_j} = \beta_{n_{j-1}}. \end{array} \right. \quad (2.9)$$

从(2.9)式可以求得一组向量 $\{\alpha_1, \beta_2, \cdots, \beta_{n_j}\}$, 我们称它为 Jordan 链. 链中第一个向量 α_1 是 A 关于 λ_1 的特征向量, $\beta_2, \cdots, \beta_{n_j}$ 称为广义特征向量. 它的长度 n_j 就是 $J_{1j}(\lambda_1)$ 的阶数. (2.9)式给出的是一个递归的过程. 从 α_1 求 β_2 , 从 β_2 求 β_3 等等. 该过程到线性方程组 $(A - \lambda_1 I)\beta_{n_j+1} = \beta_{n_j}$ 无解(不相容)时终止, 便得到了 Jordan 链和 $J_{1j}(\lambda_1)$ 的阶数 n_j .

由(2.8)式, $J_1(\lambda)$ 有多少个 Jordan 块 $J_{1j}(\lambda_1)$, 就有多少条 Jordan 链, 也就会有多少个线性无关的特征向量 $\alpha_i, i = 1, 2, \cdots, t$. 事实上是, 矩阵 A 关于 λ 的线性无关的特征向量的个数决定 $J_1(\lambda)$ 中 Jordan 块的个数.

分析过程逆回去, 便可确定 P 的所有列向量和 J_A 的构造了. 我们把计算步骤归纳如下:

(1) 求 A 的特征多项式

$$|\lambda I - A| = (\lambda - \lambda_1)^{k_1} (\lambda - \lambda_2)^{k_2} \cdots (\lambda - \lambda_s)^{k_s},$$

$\lambda_1, \lambda_2, \dots, \lambda_r$ 互异, 从而 λ_i 为 A 的 k_i 重特征值, 其代数重数 k_i 决定 Jordan 矩阵 $J_i(\lambda)$ 的阶数为 k_i .

(2) 对 λ_i , 由 $(A - \lambda_i I)X = 0$, 求 A 的线性无关的特征向量 $\alpha_1, \alpha_2, \dots, \alpha_{k_i}$. λ_i 的几何重数 $\dim V_{\lambda_i} = t_i$ 决定 $J_i(\lambda)$ 中有 t_i 个 Jordan 块.

(3) 若 λ_i 的代数重数等于几何重数: $k_i = \dim V_{\lambda_i}$, λ_i 对应的 Jordan 矩阵为 k_i 阶对角矩阵.

若 $\dim V_{\lambda_i} = t_i < k_i$, 则在 V_{λ_i} 中选择适当特征向量 α_i . 由(2.9)式求 Jordan 链 $\alpha_i, \beta_1, \dots, \beta_{t_i}$, 确定 $J_i(\lambda)$ 中 Jordan 块 $J_{ij}(\lambda)$ 的阶数 n_j , 从而得到了 J_A 的结构.

(4) 所有 Jordan 链构成矩阵 P , 必有

$$P^{-1}AP = J_A.$$

例 5 设

$$A = \begin{pmatrix} -3 & 3 & -2 \\ -7 & 6 & -3 \\ 1 & -1 & 2 \end{pmatrix},$$

求可逆矩阵 P 和 Jordan 矩阵 J_A , 使 $P^{-1}AP = J_A$.

解 由 $|A - \lambda I| = (\lambda - 1)(\lambda - 2)^2 = 0$ 得 $\lambda_1 = 1, \lambda_2 = \lambda_3 = 2$, 所以

$$J_A = \begin{pmatrix} J_1(1) \\ & J_2(2) \end{pmatrix},$$

其中 $J_1(1) = (1)$ 已被确定, $J_2(2) = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$ 或 $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.

从 $(A - I)X = 0$, 求得 $\lambda_1 = 1$ 对应的特征向量

$$\alpha_1 = (1 \ 2 \ 1)^T,$$

从 $(A - 2I)X = 0$, 求得 $\lambda_2 = 2$ 对应线性无关的特征向量, $\alpha_2 = (-1 \ -1 \ 1)^T$, 只有一个, 所以可确定 $J_2(2)$ 只由一个 Jordan 块构成, 即

$$J_2(2) = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}.$$

由(2.9)式求解

$$(A - 2I)\beta = \alpha_2,$$

取一个解 $\beta = (-1 \ -2 \ 0)^T$, 得到所需的一个广义特征向量 β .

$$\text{故 } P = (\alpha_1 \ \alpha_2 \ \beta) = \begin{pmatrix} 1 & -1 & -1 \\ 2 & -1 & -2 \\ 1 & 1 & 0 \end{pmatrix},$$

$$J_A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

例6 设

$$A = \begin{pmatrix} 2 & 1 & 0 & -1 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix},$$

求可逆矩阵 P 和 Jordan 矩阵 J_A , 使 $P^{-1}AP = J_A$.

解 $|\lambda I - A| = (\lambda - 2)^4$, 所以 $\lambda = 2$ 为 A 的四重特征值, $\lambda = 2$ 的代数重数为 4, 故 $J_A = J(2)$ 是一个对角线元素为 2 的四阶 Jordan 矩阵.

从 $(A - 2I)X = 0$, 即

$$\begin{pmatrix} 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} X = 0$$

求得两个线性无关的特征向量:

$$\alpha_1 = (1 \ 0 \ 0 \ 0)^T, \quad \alpha_2 = (0 \ 0 \ 1 \ 0)^T,$$

因此 $\lambda = 2$ 的几何重数 $\dim V_{\lambda=2} = 2$, 这决定 $J_A = J(2)$ 含两个 Jordan 块, (不计较块的次序) 它只有两种可能:

$$J_A = \left[\begin{array}{c|ccc} 2 & & & \\ \hline & 2 & 1 & & \\ & & 2 & 1 & \\ & & & 2 & \\ & & & & 2 \end{array} \right] \quad \text{或} \quad J_A = \left[\begin{array}{cc|cc} 2 & 1 & & \\ & 2 & & \\ \hline & & 2 & 1 \\ & & & 2 \end{array} \right].$$

取 α_1 , 所以 $(A - 2I)\beta = \alpha_1$ 的增广矩阵为

$$(A - 2I | \alpha_1) = \left[\begin{array}{cccc|c} 0 & 1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right],$$

对应的方程组是相容的, 由此解得 $\beta_1 = (0 \ 1 \ 0 \ 0)^T$, 但 $(A - 2I)\beta_2 = \alpha_2$ 的增广矩阵

$$(A - 2I | \beta_2) = \left[\begin{array}{cccc|c} 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

所对应的方程组不相容(即无解), 所以 $\{\alpha_1, \beta_2\}$ 为一长为 2 的 Jordan 链, 它对应一个二阶 Jordan 块.

再由 $(A - 2I)\eta = \alpha_2$, 即

$$\begin{pmatrix} 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \eta = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix},$$

得另一个广义特征向量 $\eta_2 = (0 \ 1 \ 0 \ 1)^T$, 由此第2个Jordan链为 $\{\alpha_2, \eta_2\}$. 因为得到 η_2 , 故 $\{\alpha_1, \beta_2, \alpha_2, \eta_2\}$ 已够组成 P 所需的四个列向量. 不再讨论 α_2 打头的Jordan链是否断开的问题, 事实上, $(A - 2I)\eta_2 = \eta_2$ 也是不相容的, 故 α_2 打头只有长度为2的Jordan链. 由此

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \quad J_A = \begin{pmatrix} 2 & 1 & & \\ & 2 & & \\ & & 2 & 1 \\ & & & 2 \end{pmatrix}.$$

例7 设 $P_3[x]$ 上线性变换 T 在基 $\{1, x, x^2\}$ 下的矩阵为 A , 求 $P_3[x]$ 的基 $\{\xi_1, \xi_2, \xi_3\}$, 使 T 在此基下的矩阵为 Jordan 矩阵. 其中

$$A = \begin{pmatrix} 2 & -1 & -1 \\ 2 & -1 & -2 \\ -1 & 1 & 2 \end{pmatrix}.$$

解 $|A - \lambda I| = (\lambda - 1)^3$, 因此 $\lambda = 1$ 的代数重数为3. 又 $(A - I)X = 0$ 为

$$\begin{pmatrix} -1 & 1 & 1 \\ -2 & 2 & 2 \\ 1 & -1 & -1 \end{pmatrix} X = 0,$$

得 $X = k_1(1 \ 1 \ 0)^T + k_2(1 \ 0 \ 1)^T$, 所以, 有两个线性无关的特征向量

$$\alpha_1 = (1 \ 1 \ 0)^T, \alpha_2 = (1 \ 0 \ 1)^T.$$

$\lambda = 1$ 的几何重数 $\dim V_{\lambda=1} = 2$, J_A 含2个Jordan块, 已可确定 J_A 形如

$$J_A = \begin{pmatrix} 1 & & \\ & 1 & 1 \\ & & 1 \end{pmatrix},$$

为求一个广义特征向量, 先取 α_1 , 方程 $(A - I)\beta = \alpha_1$ 的增广矩阵为

$$\begin{pmatrix} -1 & 1 & 1 & | & 1 \\ -2 & 2 & 2 & | & 1 \\ 1 & -1 & -1 & | & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & -1 & -1 & | & 1 \\ 0 & 0 & 0 & | & -1 \\ 0 & 0 & 0 & | & 1 \end{pmatrix},$$

所以 $(A - I)\beta = \alpha_1$ 不相容. 再取 α_2 , 方程 $(A - I)\beta = \alpha_2$ 也不相容, 这说明要求的Jordan链第一个特征向量选择不当. 为此, 在空间 $V_{\lambda=1}$ 中取一般形式的特征向量 $\alpha = k_1(1 \ 1 \ 0)^T + k_2(1 \ 0 \ 1)^T = (k_1 + k_2 \ k_1 \ k_2)^T$,

$$(A - I|\alpha) = \begin{pmatrix} 1 & -1 & -1 & | & k_1 + k_2 \\ 2 & -2 & -2 & | & k_1 \\ -1 & 1 & 1 & | & k_2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & -1 & -1 & | & k_1 + k_2 \\ 0 & 0 & 0 & | & k_1 + 2k_2 \\ 0 & 0 & 0 & | & 0 \end{pmatrix}.$$

按相容性要求: $k_1 + 2k_2 = 0$, 取 $k_1 = 2, k_2 = -1$, 得 $\alpha_3 = (1 \ 2 \ -1)^T$, $(A - I)\beta = \alpha_3$ 必相容, 从中求得

$$\beta = (1 \ 0 \ 0)^T.$$

因为 α_2 与 α_3 线性无关, 故可取

$$P = (\alpha_2 \ \alpha_3 \ \beta) = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 0 \\ 1 & -1 & 0 \end{pmatrix}; \text{ 相应 } J_A = \left[\begin{array}{c|ccc} 1 & & & \\ \hline & 1 & & \\ & & 1 & 1 \\ & & & 1 \end{array} \right].$$

该题中 α_1 与 α_3 也线性无关, 上述 P 也可取为 $P = (\alpha_1 \ \alpha_3 \ \beta)$, 所以可逆矩阵 P 不惟一.

应注意 P 的列向量构成应和 J 相应. 这题如取 $P = (\alpha_3 \ \beta \ \alpha_2)$ 则相应 $J_A = \left[\begin{array}{c|cc} 1 & 1 & \\ \hline & 1 & \\ & & 1 \end{array} \right]$. 由于 Jordan 矩阵中 Jordan 块次序不计, $\left[\begin{array}{c|cc} 1 & 1 & \\ \hline & 1 & 1 \\ & & 1 \end{array} \right]$ 和 $\left[\begin{array}{c|cc} 1 & 1 & \\ \hline & 1 & \\ & & 1 \end{array} \right]$ 都是 A 的

Jordan 标准形.

由 1.3 节式(1.26)知, 所求可逆矩阵 P 是从基 $\{1, x, x^2\}$ 到基 $\{\xi_1, \xi_2, \xi_3\}$ 的过渡矩阵, 即

$$(\xi_1 \ \xi_2 \ \xi_3) = (1 \ x \ x^2)P = (1 \ x \ x^2) \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 0 \\ 1 & -1 & 0 \end{pmatrix},$$

因此

$$\{\xi_1, \xi_2, \xi_3\} = \{1+x^2, 1+2x-x^2, 1\},$$

T 在 $\{\xi_1, \xi_2, \xi_3\}$ 下矩阵为

$$\begin{pmatrix} 1 & & \\ & 1 & 1 \\ & & 1 \end{pmatrix}.$$

例 8 证明对任意方阵 $A \in C^{n \times n}$, A 相似于 A^T .

证明 设 A 的 Jordan 标准形 J_A 具有形式

$$J_A = \begin{pmatrix} J_1(\lambda_1) & & & \\ & J_2(\lambda_2) & & \\ & & \ddots & \\ & & & J_m(\lambda_m) \end{pmatrix},$$

其中 $J_i(\lambda)$ 为 k_i 阶 Jordan 块, 则

$$J_A^T = \begin{pmatrix} J_1^T(\lambda_1) & & & \\ & J_2^T(\lambda_2) & & \\ & & \ddots & \\ & & & J_m^T(\lambda_m) \end{pmatrix}.$$

注意到

$$J_i(\lambda) = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \ddots & \ddots \\ & & & \lambda & 1 \end{pmatrix} \stackrel{\text{列}}{=} (\lambda e_1, e_1 + \lambda e_2, \dots, e_{k_i-1} + \lambda e_{k_i}),$$

$$J_i^T(\lambda) = \begin{pmatrix} \lambda & & & \\ 1 & \lambda & & \\ & 1 & \ddots & \\ & & \ddots & \lambda \\ & & & 1 \end{pmatrix} \stackrel{\text{列}}{=} (\lambda e_1 + e_2, \lambda e_2 + e_3, \dots, \lambda e_{k_i}),$$

取矩阵 $S_i \in \mathbb{C}^{k_i \times k_i}$,

$$S_i = \begin{pmatrix} & & & 1 \\ & & 1 & \\ & \ddots & & \\ 1 & & & \end{pmatrix}_{k_i \times k_i} \stackrel{\text{列}}{=} (e_{k_i}, e_{k_i-1}, \dots, e_1),$$

S_i 被称为逆向单位矩阵, $S_i^{-1} = S_i$.

对 Jordan 块 $J_i(\lambda)$, 用矩阵的分块运算,

$$\begin{aligned} S_i^{-1} J_i(\lambda) S_i &= S_i (J_i(\lambda) e_{k_i}, \dots, J_i(\lambda) e_2, J_i(\lambda) e_1) \\ &= S_i (e_{k_i-1} + \lambda e_{k_i}, e_{k_i-2} + \lambda e_{k_i-1}, \dots, e_1 + \lambda e_2, \lambda e_1) \\ &= (S_i e_{k_i-1} + \lambda S_i e_{k_i}, S_i e_{k_i-2} + \lambda S_i e_{k_i-1}, \dots, S_i e_1 + \lambda S_i e_2, \lambda S_i e_1) \\ &= (\lambda e_1 + e_2, \lambda e_2 + e_3, \dots, \lambda e_{k_i-1} + e_{k_i}, \lambda e_{k_i}) \\ &= J_i^T(\lambda). \end{aligned}$$

因此, 取可逆矩阵

$$S = \begin{pmatrix} S_1 & & & \\ & S_2 & & \\ & & \ddots & \\ & & & S_m \end{pmatrix},$$

则

$$S^{-1} J_A S = J_A^T,$$

即 A 的 Jordan 矩阵 J_A 相似于 J_A^T .

又当 $A = P J_A P^{-1}$ 时, $A^T = (P^T)^{-1} J_A^T P^T$.

由相似的传递性, 可以得到结论: A 相似于 A^T .

2.3 最小多项式

一个方阵 $A_{n \times n}$ 的 Jordan 标准形 J_A , 在和矩阵 A 相似的一切矩阵构成的相似类中, 是

形式最简单的. A 和 J_A 又都具有相似类的共性(相似不变性), 如 $|A| = |J_A|$; $\text{rank}(A) = \text{rank}(J_A)$; $|M - A| = |M - J_A|$ 等等. 因此, 在讨论相关于 A 的一些问题时, 常常利用 J_A 的简单形式, 先就 J_A 进行讨论, 得到 A 的相应结论, 这就是 Jordan 化方法. 这一节用 Jordan 化方法先讨论 A 的矩阵多项式的计算问题. 由此证明 Cayley 定理, 再给出矩阵 A 的最小多项式及其性质.

一、矩阵多项式

定义 2.4 设 $A \in F^{n \times n}$, $a_i \in F$, $g(\lambda) = a_m \lambda^m + a_{m-1} \lambda^{m-1} + \cdots + a_1 \lambda + a_0$ 是一个多项式, 则称矩阵 $g(A) = a_m A^m + a_{m-1} A^{m-1} + \cdots + a_1 A + a_0 I$ 为 A 的矩阵多项式.

应注意, $A \in F^{n \times n}$, $a_i \in F$, $i = 0, 1, 2, \dots, m$ 时, $g(A) \in F^{n \times n}$, 即 $g(A)$ 也是一个方阵, 容易看到 A 和 $g(A)$ 有如下关系.

定理 2.6 设 $A \in F^{n \times n}$, $g(A)$ 是 A 的矩阵多项式, 则有如下结果:

- (1) 若 λ_0 是 A 的特征值, 则 $g(\lambda_0)$ 是 $g(A)$ 的特征值.
- (2) 如果 A 相似于 B : $P^{-1}AP = B$, 则 $g(A)$ 相似于 $g(B)$: $P^{-1}g(A)P = g(B)$.
- (3) 如果 A 为准对角矩阵, 则 $g(A)$ 也是准对角矩阵. 而且

若 $A = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{pmatrix}$, A_i 为方子块,

则 $g(A) = \begin{pmatrix} g(A_1) & & & \\ & g(A_2) & & \\ & & \ddots & \\ & & & g(A_k) \end{pmatrix}$.

证明 只证(2), 其余留做练习.

已知 $P^{-1}AP = B$, 则

$$\begin{aligned} P^{-1}g(A)P &= P^{-1}(a_m A^m + a_{m-1} A^{m-1} + \cdots + a_1 A + a_0 I)P \\ &= a_m (P^{-1}AP)^m + a_{m-1} (P^{-1}AP)^{m-1} + \cdots + a_1 P^{-1}AP + a_0 I \\ &= g(P^{-1}AP) \\ &= g(B). \end{aligned}$$

(2) 得证. □

下面, 我们讨论 $g(A)$ 的计算问题, 即已知方阵 A 和多项式 $g(\lambda)$, 如何计算方阵 $g(A)$ 的问题. 设

$$g(\lambda) = a_m \lambda^m + a_{m-1} \lambda^{m-1} + \cdots + a_1 \lambda + a_0.$$

为计算 $g(A)$, 用 Jordan 化方法. 设

$$A = P J_A P^{-1}; \quad J_A = \begin{pmatrix} J_1(\lambda_1) & & & \\ & J_2(\lambda_2) & & \\ & & \ddots & \\ & & & J_s(\lambda_s) \end{pmatrix},$$

其中 $J_1(\lambda_1), J_2(\lambda_2), \dots, J_s(\lambda_s)$ 为 J_A 的全部 Jordan 块.

由定理 2.7 知

$$g(A) = Pg(J_A)P^{-1} = P \begin{pmatrix} g(J_1(\lambda_1)) & & & \\ & g(J_2(\lambda_2)) & & \\ & & \ddots & \\ & & & g(J_s(\lambda_s)) \end{pmatrix} P^{-1},$$

因此计算 $g(A)$ 的问题转化为对 Jordan 块 $J_i(\lambda_i)$ 计算 $g(J_i(\lambda_i))$ 的问题.

取一个 r 阶 Jordan 块为代表作分析.

$$\text{设 } J(\lambda) = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda \end{pmatrix}_{r \times r} = \lambda I + U_r,$$

首先有

$$J(\lambda)^k = (\lambda I + U_r)^k = \lambda^k I + C_k^1 \lambda^{k-1} U + C_k^2 \lambda^{k-2} U^2 + \dots + U^k, \quad (2.10)$$

$$\text{利用 } U_r^k = \begin{cases} \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 & 0 \cdots 0 \\ & \ddots & & \ddots & \ddots & \\ 0 & \ddots & \ddots & \ddots & & 1 \\ & \ddots & \ddots & \ddots & 0 & \\ & & \ddots & \ddots & 0 & \\ & & & & 0 & \end{pmatrix}, & k < r, \\ 0, & k \geq r, \end{cases} \quad (2.11)$$

其中

$$C_k^j = \frac{k!}{j!(k-j)!}.$$

(2.10)式表示成矩阵形式是

$$J(\lambda)^k = \begin{cases} \begin{pmatrix} \lambda^k & C_k^1 \lambda^{k-1} & \cdots & 1 & 0 \\ & \lambda^k & C_k^1 \lambda^{k-1} & \ddots & 0 \\ & & \ddots & \ddots & 1 \\ & & & \ddots & C_k^1 \lambda^{k-1} \\ & & & & \lambda^k \end{pmatrix}, & k < r, \\ \begin{pmatrix} \lambda^k & C_k^1 \lambda^{k-1} & \cdots & C_k^{r-1} \lambda^{k-r+1} \\ & \lambda^k & C_k^1 \lambda^{k-1} & \vdots \\ & & \ddots & C_k^1 \lambda^{k-1} \\ & & & \lambda^k \end{pmatrix}, & k \geq r. \end{cases}$$

代入 $g(J(\lambda))$, 并合并 U 的相同幂次项得

在 $m \geq r$ 时,

$$g(J(\lambda)) = \sum_{i=0}^{r-1} \left(\sum_{k=i}^m a_k C_k^i \lambda^{k-i} \right) U^i.$$

$$\begin{aligned} \text{又 } \sum_{k=i}^m a_k C_k^i \lambda^{k-i} &= \sum_{k=i}^m \frac{a_k k!}{(k-i)! i!} \lambda^{k-i} = \frac{1}{i!} \sum_{k=i}^m a_k \frac{d^i}{d\lambda^i}(\lambda^k) \\ &= \frac{1}{i!} g^{(i)}(\lambda), \end{aligned}$$

$$g(J(\lambda)) = \sum_{i=0}^{r-1} \frac{g^{(i)}(\lambda)}{i!} U^i = \begin{pmatrix} g(\lambda) & g'(\lambda) & \cdots & \frac{g^{(r-1)}(\lambda)}{(r-1)!} \\ & g(\lambda) & \ddots & | \\ & & \ddots & g'(\lambda) \\ & & & g(\lambda) \end{pmatrix}; \quad (2.12)$$

当 $m < r$ 时,

$$g(J(\lambda)) = \begin{pmatrix} g(\lambda) & g'(\lambda) & \cdots & \frac{g^m(\lambda)}{m!} & 0 & \cdots & 0 \\ & g(\lambda) & & & \ddots & & 0 \\ & & \ddots & & & \frac{g^{(m)}(\lambda)}{m!} & \\ & & & \ddots & & & | \\ & & & & g'(\lambda) & & \\ & & & & g(\lambda) & & \end{pmatrix}.$$

二、方阵的化零多项式

对 n 阶方阵 A , 若存在多项式 $g(\lambda)$, 使矩阵 $g(A) = 0$, 则称 $g(\lambda)$ 为矩阵 A 的化零多项式.

我们感兴趣的是方阵 A 的化零多项式是否存在? Cayley-Hamilton 定理回答这个问题.

定理 2.7 (Cayley-Hamilton) 设 $A \in F^{n \times n}$, 则方阵 A 的特征多项式就是 A 的化零多项式, 即若

$$f(\lambda) = |\lambda I - A| = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0,$$

则有 $f(A) = 0$.

证明 用 Jordan 化方法证.

设 $f(\lambda) = (\lambda - \lambda_1)^{r_1}(\lambda - \lambda_2)^{r_2} \cdots (\lambda - \lambda_s)^{r_s}$, 其中 $\sum_{i=1}^s r_i = n$, $\lambda_1, \lambda_2, \dots, \lambda_s$ 互不相同.

$$\text{设 } A = P \begin{pmatrix} J_1(\lambda_1) & & & \\ & J_2(\lambda_2) & & \\ & & \ddots & \\ & & & J_m(\lambda_m) \end{pmatrix} P^{-1},$$

$J_i(\lambda_i)$ 是对角线元素为 λ_i 的 t_i 阶 Jordan 块, $\lambda_1, \dots, \lambda_m$ 不必互异, $i = 1, 2, \dots, m$,

$$\text{则 } f(A) = P \begin{pmatrix} f(J_1(\lambda_1)) & & & \\ & f(J_2(\lambda_2)) & & \\ & & \ddots & \\ & & & f(J_m(\lambda_m)) \end{pmatrix} P^{-1},$$

$$f(J_i(\lambda_i)) = \begin{pmatrix} f(\lambda_i) & f'(\lambda_i) & \frac{1}{2!}f''(\lambda_i) & \cdots & \frac{1}{t_i-1}f^{(t_i-1)}(\lambda_i) \\ f(\lambda_i) & & & & \vdots \\ & \ddots & & & \vdots \\ & & & & f'(\lambda_i) \\ & & & & f(\lambda_i) \end{pmatrix}_{t_i \times t_i}.$$

我们只看 $f(J_i(\lambda_i))$ 的第 1 行元素, 由于 $f(\lambda) = |\lambda I - A| = (\lambda - \lambda)^{r_i} g(\lambda)$ 中含因子 $(\lambda - \lambda)^{r_i}$, $t_i \leq r_i$, 所以 $t_i - 1 < r_i$, 从而 $f(\lambda), f'(\lambda), \dots, f^{(t_i-1)}(\lambda)$ 中均含有因子 $(\lambda - \lambda)$,

因此 $f(\lambda) = f'(\lambda) = \cdots = f^{(t_i-1)}(\lambda) = 0$,

即 $f(J_i(\lambda_i))$ 第一行元素全为 0, 由 $f(J_i(\lambda_i))$ 的结构有

$$f(J_i(\lambda_i)) = 0, \quad i = 1, 2, \dots, m.$$

所以

$$f(A) = 0.$$

□

$f(\lambda)$ 为 n 次多项式, $f(A) = 0$ 使得 A^n 可以表示成 A^k ($k < n$) 的幂次项的组合:

$$A^n = - (a_{n-1} A^{n-1} + a_{n-2} A^{n-2} + \cdots + a_1 A + a_0 I). \quad (2.13)$$

例 9 设 $A = \begin{pmatrix} 2 & 1 & 0 & -1 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}$, 求 A^6 .

解 方法 1:

由例 4

$$A = P \begin{pmatrix} 2 & 1 & & \\ & 2 & & \\ & & 2 & 1 \\ & & & 2 \end{pmatrix} P^{-1}, \quad \text{其中 } P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

所以

$$A^6 = P \begin{pmatrix} \begin{pmatrix} 2 & 1 \\ & 2 \end{pmatrix}^6 & & \\ & \begin{pmatrix} 2 & 1 \\ & 2 \end{pmatrix}^6 & \\ & & \begin{pmatrix} 2 & 1 \\ & 2 \end{pmatrix}^6 \end{pmatrix} P^{-1}.$$

令 $g(\lambda) = \lambda^6$, 则

$$\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}^6 = g \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} g(2) & g'(2) \\ 0 & g(2) \end{pmatrix} = \begin{pmatrix} 2^6 & 6 \times 2^5 \\ & 2^6 \end{pmatrix},$$

$$\text{所以 } A^6 = P \begin{pmatrix} 2^6 & 6 \times 2^5 & 0 & 0 \\ 0 & 2^6 & 0 & 0 \\ 0 & 0 & 2^6 & 6 \times 2^5 \\ 0 & 0 & 0 & 2^6 \end{pmatrix} P^{-1} = \begin{pmatrix} 64 & 192 & 0 & -192 \\ 0 & 64 & 0 & 0 \\ 0 & 0 & 64 & 192 \\ 0 & 0 & 0 & 64 \end{pmatrix}.$$

方法 2:

$$f(\lambda) = |\lambda I - A| = (\lambda - 2)^4,$$

$$g(\lambda) = \lambda^6 = (\lambda^2 - 8\lambda + 40)f(\lambda) + 160\lambda^3 - 720\lambda^2 - 1152\lambda - 640,$$

$$g(A) = 160A^3 - 720A^2 - 1152A - 640I = \begin{pmatrix} 64 & 192 & 0 & -192 \\ 0 & 64 & 0 & 0 \\ 0 & 0 & 64 & 192 \\ 0 & 0 & 0 & 64 \end{pmatrix}.$$

三、最小多项式

对 $A \in F^{n \times n}$, Cayley 定理指出它有化零多项式 $f(\lambda) = |\lambda I - A|$, 事实上, A 有很多化零多项式, 任取多项式 $h(\lambda)$, $\varrho(\lambda) = h(\lambda)f(\lambda)$ 都是 A 的化零多项式. 若 A 是 $V_n(F)$ 上线性

变换 T 的矩阵, 对 A 的化零多项式 $f(\lambda)$, 相应地用多项式得到的线性变换 $f(T)$ 就是零变换: $f(T) = 0$. 所以线性变换 T 的化零多项式也是存在的, 而且

$$\text{线性变换 } f(T) = 0 \Leftrightarrow \text{矩阵 } f(A) = 0 \quad (2.14)$$

线性变换 T 的(或者说矩阵 A 的)最小多项式 $m_T(\lambda)$ 是 T 的所有化零多项式中次数最低, 首项系数是 1 的多项式. 这里我们讨论最小多项式的性质与结构, 并从最小多项式结构讨论线性变换 T 有对角矩阵(T 可对角化)的充分必要条件.

定义 2.5 设 T 是线性空间 $V_n(T)$ 上的线性变换, $m_T(\lambda)$ 是一个关于文字 λ 的多项式, 如果 $m_T(\lambda)$ 满足

- (1) $m_T(\lambda)$ 最高次项系数为 1;
- (2) $m_T(\lambda)$ 是 T 的一个化零多项式, 即 $m_T(T) = 0$;
- (3) $m_T(\lambda)$ 是 T 的化零多项式中次数最低的多项式,

则称 $m_T(\lambda)$ 是 T 的最小多项式.

定义 2.5 中条件(3)可以推出 T 的最小多项式 $m_T(\lambda)$ 整除 T 的一切化零多项式, 即:

如果多项式 $\varphi(\lambda)$, 使 $\varphi(T) = 0$, 则 $m_T(\lambda) | \varphi(\lambda)$. 否则, $m_T(\lambda) \nmid \varphi(\lambda)$, 则有次数低于 $m_T(\lambda)$ 的多项式 $h(\lambda)$, 使

$$\varphi(\lambda) = m_T(\lambda)g(\lambda) + h(\lambda), \quad \varphi(T) = 0 \Rightarrow h(T) = 0,$$

与 $m_T(\lambda)$ 次数最低矛盾.

又由(2.14)式, 若 $m_T(\lambda)$ 是 T 的最小多项式, 则 $m_T(\lambda)$ 也是 T 的矩阵 A 的最小多项式, 又

$$m_T(P^{-1}AP) = P^{-1}m_T(A)P,$$

所以, T 与它的矩阵 A 有相同的化零多项式和最小多项式. 相似矩阵也有相同的化零多项式与最小多项式. 因而研究 A 和研究 T 的最小多项式有相同的结论.

定理 2.8 T 的特征多项式 $f(\lambda)$ 与最小多项式 $m_T(\lambda)$ 有相同的根(重数不计), 即若

$$f(\lambda) = |\lambda I - A| = (\lambda - \lambda_1)^{r_1}(\lambda - \lambda_2)^{r_2} \cdots (\lambda - \lambda_s)^{r_s},$$

则

$$m_T(\lambda) = (\lambda - \lambda_1)^{t_1}(\lambda - \lambda_2)^{t_2} \cdots (\lambda - \lambda_s)^{t_s}, \quad (2.15)$$

$$1 \leq t_i \leq r_i, \quad i = 1, 2, \dots, s.$$

证明 由 $m_T(\lambda) | f(\lambda)$, $t_i \leq r_i$ 是显然的, 所以, 证明只证 $t_i \geq 1$ 即可.

反设 $t_i = 0$, 即 λ_i 不是最小多项式 $m_T(\lambda)$ 的根, 取 T 的 Jordan 矩阵

$$J = \begin{pmatrix} J_1(\lambda_1) & & & \\ & J_2(\lambda_2) & & \\ & & \ddots & \\ & & & J_m(\lambda_m) \end{pmatrix},$$

其中, $J_i(\lambda_i)$ 为 Jordan 块, $\lambda_1, \lambda_2, \dots, \lambda_m \in \{\lambda_1, \lambda_2, \dots, \lambda_s\}$, 而 $m_A(A) = 0 \Leftrightarrow m_A(J_i(\lambda_i)) = 0$, 则由于 $m_T(\lambda_i) \neq 0$, 而使

$$m_T(J_i(\lambda_i)) = \begin{pmatrix} m_T(\lambda_1) & & & \\ & m_T(\lambda_2) & & \\ & & \ddots & \\ & & & m_T(\lambda_i) \end{pmatrix} \neq 0.$$

所以 $m_T(J) \neq 0$, 从而 $m_T(T) \neq 0$. 与 $m_T(\lambda)$ 是 T 的最小多项式矛盾, 所以 $r_i \geq 1$. \square

从证明中可以看到, 最小多项式 $m_T(\lambda)$ 必须满足

$$m_T(J) = \begin{pmatrix} m_T(J_1(\lambda_1)) & & & \\ & m_T(J_2(\lambda_2)) & & \\ & & \ddots & \\ & & & m_T(J_s(\lambda_s)) \end{pmatrix} = 0.$$

这启发我们更进一步来根据 Jordan 矩阵 J 的关于 λ_i 的 Jordan 块的阶数来确定式(2.15)给出的 $m_T(\lambda)$ 结构中的 r_i .

定理 2.9 设变换 T 的特征多项式为

$$f(\lambda) = (\lambda - \lambda_1)^{r_1}(\lambda - \lambda_2)^{r_2} \cdots (\lambda - \lambda_s)^{r_s},$$

又 T 的 Jordan 标准形中关于特征值 λ_i 的 Jordan 块的最高阶数为 \bar{n}_i , 则 T 的最小多项式

$$m_T(\lambda) = (\lambda - \lambda_1)^{\bar{n}_1}(\lambda - \lambda_2)^{\bar{n}_2} \cdots (\lambda - \lambda_s)^{\bar{n}_s}. \quad (2.16)$$

证明 取 $g(\lambda) = (\lambda - \lambda_1)^{\bar{n}_1}(\lambda - \lambda_2)^{\bar{n}_2} \cdots (\lambda - \lambda_s)^{\bar{n}_s}$, 设 T 的 Jordan 标准形为

$$J = \begin{pmatrix} J_1(\lambda_1) & & & \\ & J_2(\lambda_2) & & \\ & & \ddots & \\ & & & J_s(\lambda_s) \end{pmatrix},$$

$$\text{其中 } J_i(\lambda) = \begin{pmatrix} J_{i1}(\lambda_i) & & & \\ & J_{i2}(\lambda_i) & & \\ & & \ddots & \\ & & & J_{ii}(\lambda_i) \end{pmatrix},$$

$J_{ij}(\lambda)$ 关于 λ_i 的 Jordan 块为 n_{ij} 阶.

$$\bar{n}_i = \max \{n_{i1}, n_{i2}, \dots, n_{ii}\}, \quad i = 1, 2, \dots, s.$$

不妨设 $J_{ii}(\lambda_i)$ 是 \bar{n}_i 阶的 Jordan 块,

$$\text{则 } g(J) = \begin{pmatrix} g(J_1(\lambda_1)) & & & \\ & g(J_2(\lambda_2)) & & \\ & & \ddots & \\ & & & g(J_s(\lambda_s)) \end{pmatrix},$$

$$g(J_i(\lambda)) = \begin{pmatrix} g(J_{ii}(\lambda)) & & & \\ & g(J_{i2}(\lambda)) & & \\ & & \ddots & \\ & & & g(J_{it_i}(\lambda)) \end{pmatrix},$$

$$g(J_{ii}(\lambda)) = \begin{pmatrix} g(\lambda) & g'(\lambda) & \cdots & \frac{1}{(n_i - 1)!} g^{(n_i - 1)}(\lambda) \\ g(\lambda) & \ddots & & | \\ \vdots & & g'(\lambda) & \\ & & & g(\lambda) \end{pmatrix}.$$

由于 $g(\lambda)$ 含因子 $(\lambda - \lambda)^{n_i}$,

$$\text{所以 } g(\lambda) = g'(\lambda) = \cdots = g^{(n_i - 1)}(\lambda) = 0.$$

由此 $g(J_{ii}(\lambda)) = 0$, 因而也有 $g(J_{ij}(\lambda)) = 0, j = 2, 3, \dots, t_i, i = 1, 2, \dots, s$. 这说明: $g(J) = 0$, 即 $g(\lambda)$ 是 T 的化零多项式. 设又有多项式

$$h(\lambda) = (\lambda - \lambda_1)^{k_1} (\lambda - \lambda_2)^{k_2} \cdots (\lambda - \lambda_s)^{k_s},$$

其中 $1 \leq k_i \leq n_i$, 若 $h(\lambda)$ 的次数低于 $g(\lambda)$, 则一定 $\exists i$, 使 $k_i < n_i$, 对这样的 $h(\lambda)$ 可以看到, 由于

$$h^{(n_i - 1)}(\lambda) \neq 0 \quad (n_i - 1 \geq k_i),$$

从而 $h(\lambda)$ 不是 T 的化零多项式, 因此 $g(\lambda)$ 是含因子 $(\lambda - \lambda_i), i = 1, 2, \dots, s$ 的多项式中次数最低的 T 的化零多项式. 由定义知, $g(\lambda)$ 为最小多项式, 即

$$m_T(\lambda) = g(\lambda). \quad \square$$

作为定理 2.9 的推论, 若 $m_T(\lambda)$ 为一次因子之积, 即 $n_i = 1, i = 1, 2, \dots, s$, 则 $J_i(\lambda)$ 必为对角阵, 反之亦然. 因此有

定理 2.10 线性变换 T 可以对角化的充分必要条件是 T 的最小多项式 $m_T(\lambda)$ 是一次因子的乘积, 即

$$m_T(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_s).$$

定理 2.10 在矩阵上相应的叙述为, 方阵 A 可对角化的充要条件是 A 的最小多项式 $m_A(\lambda)$ 为一次因子之积. 这样我们就得到了线性变换 T (矩阵 A) 可对角化的又一个充分必要条件.

例 10 设 R^3 上线性变换 T 在基 $\{e_1, e_2, e_3\}$ 下矩阵为 $A, A = \begin{pmatrix} -3 & 3 & -2 \\ -7 & 6 & -3 \\ 1 & -1 & 2 \end{pmatrix}$, 求 T

的最小多项式 $m_T(\lambda)$.

解

$$|\lambda I - A| = (\lambda - 1)(\lambda - 2)^2,$$

由定理 2.8

$$m_T(\lambda) = (\lambda - 1)(\lambda - 2),$$

或者

$$m_T(\lambda) = (\lambda - 1)(\lambda - 2)^2.$$

由于 $(A - I)(A - 2I) = \begin{pmatrix} -4 & 3 & -2 \\ -7 & 5 & -3 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} -5 & 3 & -2 \\ -7 & 4 & -3 \\ 1 & -1 & 0 \end{pmatrix}$

 $= \begin{pmatrix} -3 & 2 & -1 \\ -3 & 2 & -1 \\ 3 & -2 & 1 \end{pmatrix} \neq 0,$

可断定

$$m_A(\lambda) = (\lambda - 1)(\lambda - 2)^2.$$

例 11 设 $A = \begin{pmatrix} 0 & 0 & 0 & 4 \\ 1 & 0 & 0 & -4 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 4 \end{pmatrix}$, 求矩阵 A 的最小多项式 $m_A(\lambda)$.

解

$$|M - A| = (\lambda - 1)(\lambda + 1)(\lambda - 2)^2.$$

由于 $\lambda_1 = 1, \lambda_2 = -1$ 为 $f(\lambda)$ 的单根, 因此相应 A 的 Jordan 标准形中关于 λ_1, λ_2 的 Jordan 块有 $n_1 = n_2 = 1$.

对 $\lambda_3 = \lambda_4 = 2, \text{rank}(A - 2I) = 3$, 所以 $\lambda = 2$ 对应 $n - \text{rank}(A - 2I) = 4 - 3 = 1$ 个 Jordan 块, 只能为 $\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$, 所以 $n_3 = 2$, 即有

$$J_A = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & 2 & 1 \\ & & & 2 \end{pmatrix}.$$

由定理 2.9, 得 $m_A(\lambda) = (\lambda - 1)(\lambda + 1)(\lambda - 2)^2$.

本例中矩阵 A 是不可对角化的.

例 12 设 $g(x)$ 是一个多项式, λ 是方阵 A 的特征值. 证明如果 $g(A) = 0$, 则有 $g(\lambda) = 0$, 即 A 的特征值是 A 的任何一个化零多项式的根.

证明 如果 $g(A) = 0$, 则由最小多项式的定义, 有 $m_A(x) | g(x)$, 从而存在多项式 $h(x)$, 使

$$g(x) = m_A(x) \cdot h(x).$$

当 λ 是 A 的特征值时, 由定理 2.9, 有 $m_A(\lambda) = 0$, 从而

$$g(\lambda) = m_A(\lambda) \cdot h(\lambda) = 0.$$

习 题 二

- 设 $A, B \in C^{n \times n}$ 是可逆矩阵, 但 $AB - BA$ 不可逆. 证明 1 是 $A^{-1}B^{-1}AB$ 的特征值.

2. 设 S 是逆向单位矩阵, 即 $S = \begin{pmatrix} & & & 1 \\ & \ddots & & \\ & & 1 & \\ 1 & & & \end{pmatrix}$, 证明 $S^{-1} = S$, 取 $A \in C^{4 \times 4}, S \in C^{4 \times 4}$,

计算 $S^{-1}AS$.

3. 设欧氏空间 R^4 上线性变换 T 在标准正交基 $\{e_1, e_2, e_3, e_4\}$ 下矩阵

$$A = \begin{pmatrix} 5 & -1 & -2 & 0 \\ -1 & 5 & 0 & -2 \\ -2 & 0 & 5 & -1 \\ 0 & -2 & -1 & 5 \end{pmatrix}.$$

(1) 求 T 的特征值与特征向量.

(2) 证明 $\forall \alpha, \beta \in R^4, (T(\alpha), \beta) = (\alpha, T(\beta))$.

(3) 设有正交矩阵 $Q = \frac{1}{2} \begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$,

证明 T 在基 $\{Qe_1, Qe_2, Qe_3, Qe_4\}$ 下的矩阵是对角矩阵.

4. 设矩阵 $A \in R^{n \times n}$ 是幂等矩阵, $A^2 = A$, $\text{rank}(A) = r$.

(1) 求 A 的特征值.

(2) 证明 R^n 可分解为 A 的特征子空间的直和.

5. 设 T 是线性空间 $V_3(C)$ 上线性变换, T 在 $V_3(C)$ 的基 $\{\alpha_1, \alpha_2, \alpha_3\}$ 下矩阵

$$A = \begin{pmatrix} 3 & 1 & 0 \\ -4 & -1 & 0 \\ 4 & -8 & -2 \end{pmatrix}, \text{求 } T \text{ 的特征值与特征子空间.}$$

6. 设矩阵 $A, B \in C^{n \times n}$ 可交换, 即 $AB = BA$, 证明 A 的特征子空间一定是 B 的不变子空间.

7. 设矩阵 $A \in C^{n \times n}$, 如果 A 满足 $A \neq I, A^k = I$ (k 为大于 1 的正整数), 证明 A 可对角化.

8. 设 $\varepsilon \neq 0$, 矩阵 A, B, C 分别为 $A = \begin{pmatrix} a & \varepsilon & 0 \\ 0 & a & \varepsilon \\ 0 & 0 & a \end{pmatrix}; B = \begin{pmatrix} a & 1 & 0 \\ 0 & a & 1 \\ \varepsilon & 0 & a \end{pmatrix}; C = \begin{pmatrix} a & 1 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{pmatrix}$. 讨

论 A, B, C 中哪些矩阵具有相似关系.

9. 设 a, b, c 为参数, 矩阵 $A = \begin{pmatrix} 2 & 0 & 0 \\ a & 2 & 0 \\ b & c & -1 \end{pmatrix}$,

(1) 写出 A 的所有可能的 Jordan 矩阵:

(2) 给出 A 可对角化的条件.

10. 设矩阵 $A \in \mathbb{R}^{n \times n}$, $\lambda \in \mathbb{R}$ 是 A 的一个实特征值, 则一定存在一个实向量 $X \in \mathbb{R}^n$ 是 A 的对应于 λ 的特征向量.

11. 设方阵 $A, B \in \mathbb{C}^{n \times n}$ 可以对角化. 证明 A 与 B 可以同时对角化, 即存在同一个可逆阵 S , 使 $S^{-1}AS, S^{-1}BS$ 为可逆矩阵的充要条件是 A 与 B 乘法可交换, 即 $AB = BA$.

12. 求下列矩阵的 Jordan 标准形 J_A 和矩阵 P , 使 $P^{-1}AP = J_A$.

$$(1) A = \begin{pmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ 1 & 0 & 2 \end{pmatrix}; \quad (2) A = \begin{pmatrix} 2 & 6 & -15 \\ 1 & 1 & -5 \\ 1 & 2 & -6 \end{pmatrix}; \quad (3) A = \begin{pmatrix} 3 & 1 & 0 & 0 \\ -4 & -1 & 0 & 0 \\ 6 & 2 & 0 & -1 \\ -2 & 0 & 1 & 2 \end{pmatrix}.$$

13. 设 $V_3(\mathbb{R})$ 上线性变换 T 在基 $\{\alpha_1, \alpha_2, \alpha_3\}$ 下矩阵为 A , $A = \begin{pmatrix} 3 & 1 & 1 \\ 0 & 4 & 0 \\ -1 & 1 & 5 \end{pmatrix}$,

(1) 求 $V_3(\mathbb{R})$ 的基 $\{\beta_1, \beta_2, \beta_3\}$, 使 T 在基 $\{\beta_i\}$ 下矩阵为 Jordan 矩阵;

(2) 求 T 的最小多项式.

14. 分别求习题 8 中矩阵的最小多项式 $m_A(\lambda)$.

15. 设 $V_4(\mathbb{R})$ 上线性变换 T 的最小多项式 $m_T(\lambda) = (\lambda - 2)(\lambda + 3)^2$. 讨论 T 的 Jordan 矩阵 J 的可能形式.

16. 设矩阵 $A \in \mathbb{C}^{n \times n}$ 满足等式: $A^3 - 2A^2 - 5A + 6 = 0$, 证明 A 可对角化.

17. 设 $g(\lambda) = \lambda^7 + 2\lambda^5 + \lambda^4 - 3\lambda^2 + \lambda - 5$, 对习题 8(1) 中矩阵 $A \in \mathbb{R}^{3 \times 3}$, 用 Jordan 化方法求 $g(A)$.

18. 设 A 是 n 阶可逆矩阵, 则存在次数不超过 $n-1$ 的多项式 $q_1(t)$, 使 $A^{-1} = q_1(A)$.

19. 设矩阵 A 的特征多项式 $|J-A| = (\lambda - 2)^2(\lambda - 3)^2$.

(1) 给出 A 的所有可能的最小多项式;

(2) 给出 A 的所有可能的 Jordan 矩阵.

20. 设 T 为 $V_n(F)$ 上线性变换, $W \subseteq V_n(F)$ 是 T 的一个不变子空间, 且 $W \neq \{0\}$, 证明 W 中至少有一个 T 的特征向量.

第3章 矩阵的分解

为了理论分析,计算方法和应用的需要,我们常常要寻求一个矩阵在不同意义下的分解形式:把矩阵分解为几个矩阵的乘积或者是若干矩阵的和.这种分解往往能反映出原矩阵的某些数值特征,又能提供分析问题所需的简化形式.这一章将介绍矩阵的几种基本分解形式.

3.1 常见的矩阵标准形与分解

从分解的角度看,我们已讨论过的矩阵的几种标准形,实际上已给出了相关分解形式.如等价标准形:

$A \in C^{n \times n}$,存在可逆矩阵 $P \in C^{n \times n}$, $Q \in C^{n \times n}$,使

$$A = P \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Q.$$

相似标准形:

$A \in C^{n \times n}$,存在可逆矩阵 $P \in C^{n \times n}$,使

$$A = P \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} P^{-1} \quad \text{或} \quad A = P J_A P^{-1}.$$

$A \in R^{n \times n}$, A 为对称矩阵,则存在可逆矩阵 P ,使

$$P^T A P = \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{pmatrix} \quad \text{或} \quad P^T A P = \begin{pmatrix} I_p & & & \\ & I_q & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}.$$

特别地,对 A ,存在正交矩阵 $C(C^T = C^{-1})$,使

$$C^T A C = C^{-1} A C = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}.$$

等价标准形分解能使我们得到与 A 等秩的最简单的矩阵 $\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$,相似分解则使我

们得到对角线元素和 A 特征值相等的形式简单的对角矩阵或 Jordan 矩阵, 它们在理论分析和实际应用中都起到了重要作用.

根据实际问题的需要, 我们还要借助于已有标准形的相关讨论得到不同的分解, 这一节我们将讨论三角分解、满秩分解和谱分解.

一、矩阵的三角分解

这里讨论的是把一个方阵 A 分解为三角形矩阵乘积的问题. 介绍 LU 分解和 LDV 分解.

定义 3.1 设 $A \in F^{n \times n}$.

(1) 若 $L, U \in F^{n \times n}$ 分别是下三角矩阵和上三角矩阵, $A = LU$, 则称 A 可作 LU 分解.

(2) 若 $L, V \in F^{n \times n}$ 分别是对角线元素为 1 的下三角矩阵和上三角矩阵, D 为对角矩阵. $A = LDV$, 则称 A 可作 LDV 分解.

用 Gauss 消元法, 一个方阵总可以用行初等变换化为上三角形矩阵. 若只用第 i 行乘数 k 加到第 j 行 ($i < j$) 型初等变换能把 A 化为上三角矩阵 U , 则有下三角形可逆矩阵 P , 使 $PA = U$, 从而有 LU 分解: $A = P^{-1}U$.

例 1 设 $A = \begin{pmatrix} 2 & 2 & 3 \\ 4 & 7 & 7 \\ -2 & 4 & 5 \end{pmatrix}$, 求 A 的 LU 分解和 LDV 分解.

解 为求 P , 对下面矩阵作如下行初等变换:

$$(A | I_3) = \left[\begin{array}{ccc|ccc} 2 & 2 & 3 & 1 & 0 & 0 \\ 4 & 7 & 7 & 0 & 1 & 0 \\ -2 & 4 & 5 & 0 & 0 & 1 \end{array} \right] \xrightarrow{\text{行 1 } \times 2, \text{ 行 2 } - 2 \times \text{行 1}, \text{ 行 3 } + 2 \times \text{行 1}} \left[\begin{array}{ccc|ccc} 2 & 2 & 3 & 1 & 0 & 0 \\ 0 & 3 & 1 & -2 & 1 & 0 \\ 0 & 6 & 8 & 1 & 0 & 1 \end{array} \right] \xrightarrow{\text{行 2 } \times 2, \text{ 行 3 } - 2 \times \text{行 2}} \left[\begin{array}{ccc|ccc} 2 & 2 & 3 & 1 & 0 & 0 \\ 0 & 3 & 1 & -2 & 1 & 0 \\ 0 & 0 & 6 & 5 & -2 & 1 \end{array} \right],$$

因此

$$P = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 5 & -2 & 1 \end{pmatrix}, \quad PA = \begin{pmatrix} 2 & 2 & 3 \\ 0 & 3 & 1 \\ 0 & 0 & 6 \end{pmatrix}.$$

令

$$L = P^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 2 & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 2 & 2 & 3 \\ 0 & 3 & 1 \\ 0 & 0 & 6 \end{pmatrix},$$

则

$$A = L \begin{pmatrix} 2 & 2 & 3 \\ 0 & 3 & 1 \\ 0 & 0 & 6 \end{pmatrix} = LU.$$

再利用初等变换, 有

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & & \\ & 3 & \\ & & 6 \end{pmatrix} \begin{pmatrix} 1 & 1 & \frac{3}{2} \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 1 \end{pmatrix},$$

就得到

$$A = LDV,$$

$$\text{其中 } L = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 2 & 1 \end{pmatrix}, D = \begin{pmatrix} 2 & & \\ & 3 & \\ & & 6 \end{pmatrix}, V = \begin{pmatrix} 1 & 1 & \frac{3}{2} \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 1 \end{pmatrix}.$$

从初等变换,大家知道方阵的 LU 分解、 LDV 分解一般不是惟一的.下面讨论方阵的 LU 和 LDV 分解的存在性和惟一性.先讨论 LDV 分解的条件.

定理3.1 设 $A = (a_{ij})_{n \times n} \in F^{n \times n}$, 则 A 有惟一的 LDV 分解 $A = LDV$ 的充分必要条件是 A 的顺序主子式

$$\Delta_k = \begin{vmatrix} a_{11} & \cdots & a_{1k} \\ a_{21} & \cdots & a_{2k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{vmatrix} \neq 0, \quad k = 1, 2, \dots, n-1; \quad \Delta_0 = 1,$$

$$\text{其中 } D = \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{pmatrix}, \quad d_k = \frac{\Delta_k}{\Delta_{k-1}}; \quad k = 1, 2, \dots, n.$$

证明 充分性: 对 A 的阶数 n 进行归纳证明

$$n=1, \quad A = (a_{11}) = (1)(a_{11})(1) = L_1 D_1 V_1,$$

所以定理对 $n=1$ 成立, 设定理对 $n-1$ 成立, 即

$$A = (a_{ij})_{(n-1) \times (n-1)} = L_{n-1} D_{n-1} V_{n-1},$$

则对 n , 将 A 分块为

$$A = \begin{pmatrix} A_{n-1} & \tau_n \\ u_n^T & a_{nn} \end{pmatrix},$$

其中 $\tau_n = (a_{1n} \ a_{2n} \ \cdots \ a_{n-1n})^T$, $u_n = (a_{n1} \ a_{n2} \ \cdots \ a_{nn})^T$.

$$\text{设 } \begin{pmatrix} A_{n-1} & \tau_n \\ u_n^T & a_{nn} \end{pmatrix} = \begin{pmatrix} L_{n-1} & 0 \\ b_n^T & 1 \end{pmatrix} \begin{pmatrix} D_{n-1} & \\ & d_n \end{pmatrix} \begin{pmatrix} V_{n-1} & v_n \\ 0 & 1 \end{pmatrix},$$

比较两边, 则有

$$A_{n-1} = L_{n-1} D_{n-1} V_{n-1}, \tag{3.1}$$

$$\tau_n = L_{n-1} D_{n-1} v_n, \tag{3.2}$$

$$u_n^T = b_n^T D_{n-1} V_{n-1}, \tag{3.3}$$

$$a_{nn} = \bar{l}^T D_{n-1} v_n + d_n. \quad (3.4)$$

由归纳假设,(3.1)式成立. 由 $\Delta_k \neq 0$, $L_{n-1}D_{n-1}$ 非奇异, $D_{n-1}V_{n-1}$ 非奇异, 从而由(3.2)式和(3.3)式可惟一确定 v_n 和 \bar{l}^T . 又从(3.4)式可惟一求得 d_n , 所以 $A = LDV$ 分解是存在而且惟一的.

又由归纳证明过程, A 的 k 阶顺序主子式

$$\begin{aligned}\Delta_1 &= |A_1| = |L_1 D_1 V_1| = |D_1|, \\ \Delta_2 &= |A_2| = |L_2 D_2 V_2| = |D_2| = d_2 |D_1|, \\ &\dots \\ \Delta_k &= |A_k| = |L_k D_k V_k| = |D_k| = d_k |D_{k-1}|,\end{aligned}$$

所以

$$d_k = \frac{\Delta_k}{\Delta_{k-1}}, k = 1, 2, \dots, n.$$

必要性: 设 A 有惟一的 LDV 分解:

$$A = LDV,$$

把它们写成矩阵分块形

$$\begin{pmatrix} A_{n-1} & \tau \\ u^T & a_{nn} \end{pmatrix} = \begin{pmatrix} L_{n-1} & 0 \\ \bar{l}^T & 1 \end{pmatrix} \begin{pmatrix} D_{n-1} & \\ & d_n \end{pmatrix} \begin{pmatrix} V_{n-1} & v \\ 0 & 1 \end{pmatrix},$$

则比较两边便有(3.1)式~(3.4)式成立.

如 $\Delta_{n-1} = |A_{n-1}| = 0$, 则由(3.1)式有

$$|D_{n-1}| = |A_{n-1}| = 0,$$

于是 $|L_{n-1}D_{n-1}| = |D_{n-1}| = 0$, 即 $L_{n-1}D_{n-1}$ 为奇异矩阵, $\text{rank}(L_{n-1}D_{n-1}) < n-1$, 则(3.2)式的解 v 不惟一, 与 A 的 LDV 分解的惟一性相矛盾, 因此 $\Delta_n \neq 0$.

应该注意到的是, 定理3.1的证明已给出了计算 A 的 LDV 分解的递归过程:

取 A 的一阶主子式, 作 LDV 分解

$$A_1 = (1)(a_{11})(1), L_1 = (1), D_1 = (a_{11}), V_1 = (1),$$

用(3.1)式~(3.4)式确定 v_1, l_1 .

$$\begin{aligned}\text{从而 } L_2 &= \begin{pmatrix} L_1 & 0 \\ l_1 & 1 \end{pmatrix}, \quad D_2 = \begin{pmatrix} D_1 & \\ & d_2 \end{pmatrix}, V_2 = \begin{pmatrix} V_1 & v_1 \\ 0 & 1 \end{pmatrix}, \\ A_2 &= L_2 D_2 V_2.\end{aligned}$$

然后重复使用(3.1)式~(3.4)式得到 A 的顺序主子式的 LDV 分解.

$$A_k = L_k D_k V_k, \quad k = 1, 2, \dots, n.$$

$k = n$ 时, 即完成 A 的 LDV 分解. □

当 A 为非奇异矩阵时, 从 $A = LDV$ 分解出发

$$A = LDV = (LD)V = L(DV).$$

由于 LD 仍是下三角矩阵, DV 仍是上三角矩阵, 就可以得到 A 的两种 LU 分解. 对一般的 n 阶方阵 A , 它的 LU 分解有如下定理.

定理 3.2 设 $A = (a_{ij}) \in F^{n \times n}$, $\text{rank}(A) = k$ ($k \leq n$), 如果 A 的顺序主子式

$$\Delta_j \neq 0, \quad j = 1, 2, \dots, k,$$

则 A 有 LU 分解.

证明 设 A_{11} 为 A 的 k 阶主子矩阵, 将 A 分块为:

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

则 A_{11} 为可逆矩阵, 且各阶主子式非 0. 由定理 3.1, A_{11} 有 LU 分解 $A_{11} = L_{11}U_{11}$, 其中 L_{11} 和 U_{11} 均为可逆矩阵.

又因为 $\text{rank}(A) = k$, 在所设条件下, A 的前 k 行线性无关, 后 $(n-k)$ 行是前 k 行的线性组合, 即存在矩阵 $B \in F^{(n-k) \times k}$,

$$A_{21} = BA_{11}, \quad A_{22} = BA_{12}.$$

取

$$L_{21} = A_{21}U_{11}^{-1}, \quad U_{12} = L_{11}^{-1}A_{12},$$

令 L_{22} 与 U_{22} 分别是下三角和上三角阵, 满足条件 $L_{22}U_{22} = 0$, (例如取 L_{22} 为对角阵, 取 $U_{22} = 0$) 则可以得到下三角矩阵 L 和上三角矩阵 U 如下:

$$L = \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix}, \quad U = \begin{pmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{pmatrix}.$$

注意: $L_{11}U_{12} = L_{11}(L_{11}^{-1}A_{12}) = A_{12}$, $L_{21}U_{11} = A_{21}U_{11}^{-1}U_{11} = A_{21}$,

$$L_{21}U_{12} + L_{22}U_{22} = A_{21}U_{11}^{-1}L_{11}^{-1}A_{12} + 0 = BA_{11}A_{11}^{-1}A_{12} = BA_{12} = A_{22},$$

从而 $LU = \begin{pmatrix} L_{11}U_{11} & L_{11}U_{12} \\ L_{21}U_{11} & L_{21}U_{12} + L_{22}U_{22} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$,

即 $A = LU$, A 有 LU 分解. □

值得指出的是, 不是所有的矩阵都有 LU 分解的. 例如 $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, A 是可逆矩阵.

如果 A 有 LU 分解, 则应有

$$A = LU = \begin{pmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{pmatrix}.$$

这将导出 $l_{11}u_{11} = 0$, 即 $l_{11} = 0$ 或 $u_{11} = 0$. 这说明 L 与 U 中至少有一个是奇异矩阵, 与 $A = LU$ 非奇异矛盾, 从而 A 没有 LU 分解.

另外, 定理 3.2 中 $\Delta_j \neq 0$ 是 A 有 LU 分解的充分条件, 并不必要. 例如:

$$A = \begin{pmatrix} 0 & 0 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

所以 A 有 LU 分解, 但 $\Delta_1 = 0$.

推论 可逆矩阵 $A \in F^{n \times n}$, 有 LU 分解的充分必要条件是 A 的顺序主子式

$$\Delta_k \neq 0, \quad k = 1, 2, \dots, n - 1.$$

例2 设 $A = \begin{pmatrix} 1 & 2 & 3 & -1 \\ 2 & -1 & 9 & -7 \\ -3 & 4 & -3 & 19 \\ 4 & -2 & 6 & -21 \end{pmatrix}$, 求 A 的 LDV 分解.

解 $A_1 = (1 \ 1 \ 1)$, 由(3.1)式~(3.4)式, 得

$$\nu_2 = \tau_2 = 2, \quad l_2 = u_2 = 2, \quad d_2 = -5,$$

$$\text{所以 } A_2 = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}, \quad L_2 = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 1 & & \\ & -5 \end{pmatrix},$$

$$V_2 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \quad A_2 = L_2 D_2 V_2, \quad A_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & -1 & 9 \\ -3 & 4 & -3 \end{pmatrix},$$

同理求得

$$L_3 = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -3 & -2 & 1 \end{pmatrix}, \quad D_3 = \begin{pmatrix} 1 & & \\ & -5 & \\ & & 12 \end{pmatrix}, \quad V_3 = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & -\frac{3}{5} \\ 0 & 0 & 1 \end{pmatrix},$$

$A_4 = A$, 从(3.1)式~(3.4)式求得:

$$l_4 = (4 \ 2 \ -1 \ 1)^T, \quad d_4 = 1, \quad \nu_4 = (-1 \ 1 \ \frac{1}{2} \ 1)^T.$$

$$\text{所以 } L_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -3 & -2 & 1 & 0 \\ 4 & 2 & -1 & 1 \end{pmatrix}, \quad D_4 = \begin{pmatrix} 1 & & & \\ & -5 & & \\ & & 12 & \\ & & & -1 \end{pmatrix},$$

$$V_4 = \begin{pmatrix} 1 & 2 & 3 & -1 \\ 0 & 1 & -\frac{3}{5} & 1 \\ 0 & 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad A = L_4 D_4 V_4.$$

从Gauss消元法知, 当系数矩阵 A 为三角形矩阵时, 线性方程组 $AX = b$ 的求解很容易. 因此, 对一般的矩阵 $A_{m \times n}$, 它的三角分解的一个很自然的应用就是用于求解线性方程组 $AX = b$.

设 A 的 LU 分解为 $A = LU$, 则

$$AX = b \Leftrightarrow LUX = b \Leftrightarrow \begin{cases} LY = b, \\ UX = Y. \end{cases} \quad (3.5)$$

$$(3.6)$$

(3.5)式和(3.6)式都是系数矩阵为三角形矩阵, 是易于用回代法求解的线性方程组. 先

用自上往下的回代法求解(3.5)式得 Y . 再代入求解(3.6)式(用自下往上的回代法), 即可得到原方程组 $AX = b$ 的解 X .

例3 设 $A = \begin{pmatrix} 1 & -3 & 7 \\ 2 & 4 & -3 \\ -3 & 7 & 2 \end{pmatrix}$, $b = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}$, 用 LU 分解求解线性方程组 $AX = b$.

解 A 为三阶方阵, 易求得 A 的 LU 分解为

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -3 & -1/5 & 1 \end{pmatrix} \begin{pmatrix} 1 & -3 & 7 \\ 0 & 10 & -17 \\ 0 & 0 & 98/5 \end{pmatrix},$$

L U

$$\begin{cases} y_1 = 2, \\ 2y_1 + y_2 = -1, \\ -3y_1 - \frac{1}{5}y_2 + y_3 = 3. \end{cases}$$

则 $LY = b$ 的形式为

自上往下用回代法, 可得

$$y_1 = 2, \quad y_2 = -1 - 2y_1 = -5, \quad y_3 = 3 + 3y_1 + \frac{1}{5}y_2 = 8.$$

又 $UX = Y$ 的具体形式为

$$\begin{cases} x_1 - 3x_2 + 7x_3 = 2, \\ 10x_2 - 17x_3 = -5, \\ \frac{98}{5}x_3 = 8. \end{cases}$$

自下往上用回代法可得

$$x_3 = \frac{20}{49}, \quad x_2 = \frac{1}{10}(-5 + 17x_3) = \frac{19}{98}, \quad x_1 = -\frac{27}{98}.$$

这样就求得线性方程组 $AX = b$ 的解向量为

$$X = \begin{pmatrix} -\frac{27}{98} & \frac{19}{98} & \frac{20}{49} \end{pmatrix}.$$

二、矩阵的满秩分解

定义 3.2 设 $A \in F^{m \times n}$, $\text{rank}(A) = r$, 若存在秩为 r 的矩阵 $B \in F^{m \times r}$, $C \in F^{r \times n}$, 使

$$A = BC, \tag{3.7}$$

则称(3.7)式为矩阵 A 的满秩分解.

在 A 的满秩分解中, B 和 C 分别为列满秩和行满秩的矩阵.

定理 3.3 对任何非零矩阵 $A \in F^{m \times n}$, 都存在满秩分解.

证明 设 $\text{rank}(A) = r$. 由等价标准形知道存在可逆矩阵 $P \in F^{m \times m}$, $O \in F^{n \times n}$. 使得

$$PAQ = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix},$$

即

$$A = P^{-1} \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Q^{-1}.$$

分块为

$$P^{-1} = (B \mid B_1), \quad Q^{-1} = \begin{pmatrix} C \\ \cdots \\ C_1 \end{pmatrix}.$$

B 为 P^{-1} 的前 r 列组成的矩阵, C 为 Q^{-1} 的前 r 行组成的矩阵, 则 $B \in F^{m \times r}$, $C \in F^{r \times n}$, 且 $\text{rank}(B) = \text{rank}(C) = r$.

$$A = P^{-1} \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Q^{-1} = (B \mid B_1) \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} C \\ \cdots \\ C_1 \end{pmatrix} = BC.$$

矩阵 A 的满秩分解一般不是唯一的. 求 A 的满秩分解有多种方法, 下面介绍常用的几种方法.

方法1 它来自于定理 3.2 的证明, B, C 分别为 P^{-1}, Q^{-1} 的前 r 列和前 r 行.

从线性代数已知求 P 和 Q 的方法如下:

$$\begin{pmatrix} A & I_m \\ I_n & 0 \end{pmatrix} \xrightarrow{\text{初等变换}} \left[\begin{array}{cc|c} I_r & 0 & P \\ 0 & 0 & \cdots \\ \hline Q & & 0 \end{array} \right].$$

再求 P^{-1}, Q^{-1} 即可得到 B, C . 但该方法计算量太大, 一般不使用. 下面介绍简化计算量的一种方法.

方法2 若只对 A 做行初等变换, 可得到阶梯形矩阵: $\begin{pmatrix} C \\ \cdots \\ 0 \end{pmatrix}$, 其中 $\text{rank}(C) = \text{rank}(A) = r$, 因此有可逆矩阵 P , 使

$$PA = \begin{pmatrix} C \\ \cdots \\ 0 \end{pmatrix},$$

从而

$$A = P^{-1} \begin{pmatrix} C \\ \cdots \\ 0 \end{pmatrix} = (B \mid B_1) \begin{pmatrix} C \\ \cdots \\ 0 \end{pmatrix} = BC.$$

方法是

$$(A \mid I_m) \xrightarrow{\text{行变换}} \begin{pmatrix} C & | \\ \cdots & | \\ 0 & | \\ P & | \end{pmatrix},$$

B 为 P^{-1} 的前 r 列, C 是 A 化为阶梯形中的非零行. $A = BC$.

例 4 设 $A = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 2 & 2 \\ 1 & 0 & 1 \end{pmatrix}$, 求 A 的满秩分解.

$$\text{解 } (A | I_3) = \left[\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 2 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{array} \right] \xrightarrow{\text{列初等变换}} \left[\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & -1 & \frac{1}{2} & 1 \end{array} \right],$$

$$\text{解得 } C = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix}, P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ -1 & \frac{1}{2} & 1 \end{pmatrix}, P^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & -1 & 1 \end{pmatrix},$$

$$\text{所以 } B = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 1 & -1 \end{pmatrix}, A = BC = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix}.$$

与方法 2 对称地有一个列初等变换的结果.

$$\left[\begin{array}{c|cc} A & \\ \hline I_n & \end{array} \right] \xrightarrow{\text{列初等变换}} \left[\begin{array}{c|cc} B & 0 \\ \hline Q & \end{array} \right],$$

C 为 Q^{-1} 的前 r 行, B 为把 A 用列变换化成的阶梯形中的非零列.

如例 4 中

$$\left[\begin{array}{c|cc} A & \\ \hline I_3 & \end{array} \right] = \left[\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 2 & 0 & 2 & 0 \\ 1 & 0 & 1 & 1 & -1 & 0 \\ \hline 1 & 0 & 0 & 1 & -1 & -1 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] \xrightarrow{\text{列变换}} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 2 & 0 \\ 1 & -1 & 0 & 1 & -1 & 0 \\ \hline 1 & -1 & -1 & 1 & -1 & -1 \\ 0 & 1 & -1 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right],$$

$$\text{因此 } B = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 1 & -1 \end{pmatrix}, Q = \begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}, Q^{-1} = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

$$\text{从而 } C = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix}, A = BC = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix}.$$

方法 3 为分析方法 3, 我们首先考虑这样的情形: $A \in F^{m \times n}$, 设 $\text{rank}(A) = r$, 而且 A 的前 r 列线性无关, 则它们是 A 的列向量的极大无关组 $\{\alpha_1, \alpha_2, \dots, \alpha_r\}$, 设 $A_1 = (\alpha_1 \ \alpha_2 \ \dots \ \alpha_r)$, 则 $\text{rank}(A_1) = r$, $A_1 \in F^{m \times r}$. 又 A 的后 $n-r$ 列 $\{\alpha_{r+1}, \alpha_{r+2}, \dots, \alpha_n\}$ 可表示为列向量

极大无关组的线性组合,设

$$A_2 = (\alpha_{r+1} \quad \alpha_{r+2} \quad \cdots \quad \alpha_n),$$

则

$$A_2 = A_1 S,$$

其中

$$S_{r \times (n-r)} = (X_{r+1} \quad X_{r+2} \quad \cdots \quad X_n),$$

X_j 满足

$$\alpha_j = (\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_r) X_j, \quad j = r+1, r+2, \dots, n,$$

因此

$$A = (A_1 \mid A_2) = (A_1 \mid A_1 S),$$

即

$$A = A_1 (I_r \mid S),$$

则 $B = A_1, C = (I_r \mid S)$ 即为满秩分解.

方法 3 的关键是要知道 A 的列向量组的极大线性无关组,用它们构成矩阵 B . 还需要知道其余列向量关于其极大无关组的线性组合. 在线性代数中,用行初等变换把矩阵 A 化为 Hermite 标准形(简化阶梯形)的有关结论已帮助我们解决了这一问题.

Hermite 标准形是阶梯形中每一行第一个非零元素为 1,而且该元素所在的列中其他元素为 0 的特殊的一种.

当 A 的前 r 列是列向量的极大无关组时, $(I_r \mid S)$ 即为 A 的简化阶梯形. 如果 A 的秩为 r ,但是前 r 列不满足线性无关,则可调动列的位置使前 r 列线性无关,这相当于右乘一个可逆矩阵 P ,使前 r 列线性无关,结合(3.8)式,即有

$$AP = A_1 (I_r \mid S),$$

所以

$$A = A_1 (I_r \mid S) P^{-1},$$

仍有满秩分解:

$$B = A_1, C = (I_r \mid S) P^{-1}.$$

应注意矩阵 P 是由初等矩阵中调换两列的 C_{ij} 型初等矩阵的乘积构成的,从而 P^{-1} 仍为 C_{ij} 型初等矩阵的乘积. 这意味着 $C = (I_r \mid S) P^{-1}$ 是矩阵 $(I_r \mid S)$ 的列向量调整后的结果,其中 I_r 列 $\{e_1, e_2, \dots, e_r\}$ 所在的位置就是 A 的极大无关组中列向量所在的位置. 我们不再重证线性代数中的已有结论,只归纳方法 3 的步骤如下:

(1)用行初等变换把 A 化为 Hermite 标准形.

(2)依 Hermite 标准形中向量 e_i 所在的列的位置第 j_i 列,相应取出 A 的第 j_i 列 α_{j_i} ,得到 A 的列向量极大无关组 $\{\alpha_{j_1}, \alpha_{j_2}, \dots, \alpha_{j_r}\}, B = (\alpha_{j_1} \quad \alpha_{j_2} \quad \cdots \quad \alpha_{j_r})$.

(3) A 的 Hermite 标准形中非零行构成矩阵 C . 得到 A 的满秩分解: $A = BC$.

例 5 用方法 3 求例 4 中 A 的满秩分解.

解 用行初等变换化 A 为 Hermite 标准形,

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 2 & 2 \\ 1 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 2 \\ 0 & 2 & 2 \\ 0 & -1 & -1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

则可知: $\text{rank}(A) = 2, A$ 的前两列线性无关,取出 A 的前两列构成 B . 因此

$$B = \begin{pmatrix} 1 & 1 \\ 0 & 2 \\ 1 & 0 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, A = BC.$$

例 6 设 $A = \begin{pmatrix} 0 & 1 & 0 & -1 & 5 & 6 \\ 0 & 2 & 0 & 0 & 0 & -14 \\ 2 & -1 & 2 & -4 & 0 & 1 \\ -2 & 1 & -2 & 2 & 10 & 25 \end{pmatrix}$, 求 A 的满秩分解.

解 把 A 用行初等变换化为简化阶梯形

$$A = \begin{pmatrix} 0 & 1 & 0 & -1 & 5 & 6 \\ 0 & 2 & 0 & 0 & 0 & -14 \\ 2 & -1 & 2 & -4 & 0 & 1 \\ -2 & 1 & -2 & 2 & 10 & 25 \end{pmatrix} \xrightarrow{\text{行变换}} \begin{pmatrix} 1 & 0 & 1 & 0 & -10 & -29 \\ 0 & 1 & 0 & 0 & 0 & -7 \\ 0 & 0 & 0 & 1 & -5 & -13 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

由此可知, $\text{rank}(A) = 3$; A 的 Hermite 标准形中 $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, $e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$, $e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$ 分别在第 1, 2, 4 列, 即 A 的第 1, 2, 4 列线性无关, 把它们取作矩阵 B .

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

C 为简化阶梯形中非零行:

$$C = \begin{pmatrix} 1 & 0 & 1 & 0 & -10 & -29 \\ 0 & 1 & 0 & 0 & 0 & -7 \\ 0 & 0 & 0 & 1 & -5 & -13 \end{pmatrix}, \quad A = BC.$$

三、可对角化矩阵的谱分解

对方阵 $A \in F^{n \times n}$, 设 $\lambda_1, \lambda_2, \dots, \lambda_s$ 为矩阵 A 的 n 个特征值. A 互异的特征值集合 $\{\lambda_1, \lambda_2, \dots, \lambda_s\}$ 称为矩阵 A 的谱. 矩阵的谱分解是讨论矩阵可相似于对角形时, 依 A 的谱或特征值把矩阵 A 分解为矩阵和的形式的一种分解. 从分解中我们可得到矩阵可相似于对角矩阵的又一个充分必要条件.

设 A 的谱为 $\{\lambda_1, \lambda_2, \dots, \lambda_s\}$, 其中 λ_i 为 A 的 r_i 重特征值 ($i=1, 2, \dots, s$), 因此

$$\sum_{i=1}^s r_i = n.$$

当 A 可相似于对角形时, 则有可逆矩阵 P , 使

$$A = P \begin{pmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_1 & & \\ & & & \ddots & \\ & & & & \lambda_2 \\ & & & & & \ddots \\ & & & & & & \lambda_s \\ & & & & & & & \ddots \\ & & & & & & & & \lambda_s \end{pmatrix} P^{-1}, \quad (3.9)$$

首先分解对角矩阵,

$$\begin{aligned} & \begin{pmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_1 & & \\ & & & \ddots & \\ & & & & \lambda_2 \\ & & & & & \ddots \\ & & & & & & \lambda_s \\ & & & & & & & \ddots \\ & & & & & & & & \lambda_s \end{pmatrix} \\ &= \lambda_1 \begin{pmatrix} I_{r_1} & & & \\ 0 & \ddots & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 & & & \\ & I_{r_2} & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} + \cdots + \lambda_s \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & \ddots & \\ & & & I_{r_s} \end{pmatrix} \\ &= \sum_{i=1}^s \lambda_i \begin{pmatrix} 0 & & & \\ & \ddots & & \\ & & I_{r_i} & \\ & & & 0 \\ & & & & 0 \end{pmatrix}, \\ \text{令 } Q_1 &= \begin{pmatrix} I_{r_1} & & & \\ 0 & \ddots & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}, \quad Q_2 = \begin{pmatrix} 0 & & & \\ & I_{r_2} & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}, \dots, \quad Q_s = \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & \ddots & \\ & & & I_{r_s} \end{pmatrix}, \end{aligned}$$

则 Q_i 满足以下性质：

- (1) $\sum_{i=1}^s Q_i = I_n$;
- (2) $Q_i^2 = Q_i, i = 1, 2, \dots, s$;
- (3) $Q_i \cdot Q_j = 0, i \neq j$.

代入(3.9)式, 则有

$$A = P \left(\sum_{i=1}^s \lambda_i Q_i \right) P^{-1} = \sum_{i=1}^s \lambda_i (P Q_i P^{-1}).$$

令 $P_i = P Q_i P^{-1}$, 则 P_i 具有以下性质：

- (1) $\sum_{i=1}^s P_i = I_n$;
- (2) $P_i^2 = P_i, i = 1, 2, \dots, s$; (3.10)
- (3) $P_i \cdot P_j = 0, i \neq j$.

$$A = \sum_{i=1}^s \lambda_i P_i. \quad (3.11)$$

(3.11)式就是一个可对角化矩阵 A 的谱分解, 即可对角化矩阵可分解为 s 个方阵 P_i 的加权和. 在后面章节, 我们将会看到 P_i 是投影矩阵, 这里我们只讨论 P_i 的性质.

定理3.4 方阵 $P \in F^{n \times n}$, 若满足 $P^2 = P$, 则称 P 为幂等矩阵. 幂等矩阵 P 具有如下性质:

- (1) P^H 和 $(I - P)$ 仍为幂等矩阵.
- (2) P 的特征值为 1 或者是 0, 而且 P 可相似于对角矩阵.
- (3) $F^n = N(P) \oplus R(P)$.

证明 只证(2)和(3), (1)是易于验证的.

(2) 设 λ 为 P 的特征值, 则存在向量 $X \neq 0$ 使 $PX = \lambda X$,

两边左乘 P , 得 $P^2 X = \lambda^2 X$,

由 $P^2 = P$, 可得等式 $P^2 - P = 0$.

取 $g(\lambda) = \lambda^2 - \lambda = \lambda(\lambda - 1)$, 则 $g(\lambda)$ 为 P 的化零多项式, 从而 P 的最小多项式 $m_p(\lambda)$ 满足 $m_p(\lambda) | g(\lambda)$, 从而 $m_p(\lambda)$ 为一次因子之积. 由定理 2.11, P 可相似于对角形.

(3) 首先证明 P 的列空间 $R(P)$ 等于 P 的特征值 $\lambda = 1$ 对应的特征子空间 $V_{\lambda=1}$.

$\forall X \in R(P)$, 则 $\exists X'$, 使 $X = PX'$,

则 $PX = P(PX') = P^2 X' = PX' = X$, 所以 $X \in V_{\lambda=1}$, 即 $R(P) \subseteq V_{\lambda=1}$.

又 $\forall X \in V_{\lambda=1}$, $X = PX$, 说明 $X \in R(P)$, 即 $R(P) \supseteq V_{\lambda=1}$. 因此

$$R(P) = V_{\lambda=1}.$$

同理可证 P 的零空间 $N(P)$ 等于 P 的特征值 $\lambda = 0$ 对应的特征子空间 $V_{\lambda=0}$, 即

$$N(P) = V_{\lambda=0},$$

由证明(2)有

$$F^n = N(P) \oplus R(P).$$

定理3.5(可对角化矩阵的谱分解) 设 $A \in C^{n \times n}$, A 的谱为 $\{\lambda_1, \lambda_2, \dots, \lambda_s\}$, 则 A 可对角化的充分必要条件是 A 有如下分解式

$$A = \sum_{i=1}^s \lambda_i P_i,$$

其中方阵 $P_i \in C^{n \times n}$, 满足如下条件:

$$(1) P_i^2 = P_i, i = 1, 2, \dots, s;$$

$$(2) P_i P_j = 0, i \neq j;$$

$$(3) \sum_{i=1}^s P_i = I_n.$$

证明 前面的分析已给出了必要性的证明, 这里只证充分性.

$\forall X \in C$, 则由(3)

$$X = I_n X = (\sum_{i=1}^s P_i) X = \sum_{i=1}^s (P_i X); \quad (3.12)$$

又对 $P_i X$, 由(2)和(1),

$$\begin{aligned} A(P_i X) &= (\sum_{i=1}^s \lambda_i P_i) P_i X = \sum_{i=1}^s \lambda_i (P_i \cdot P_i) X \\ &= \lambda_i P_i^2 X = \lambda_i (P_i X). \end{aligned}$$

从而 $P_i X \in V_{\lambda_i}$, 即 $P_i X \neq 0$ 时, 它为 A 关于特征值 λ_i 的特征向量.

由(3.12)式说明 C 可分解为特征子空间的直和. 从而

$$C = V_{\lambda_1} + V_{\lambda_2} + \dots + V_{\lambda_s} = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \dots \oplus V_{\lambda_s},$$

由定理2.4, A 可相似于对角形.

当方阵 $A \in F^{n \times n}$, $A^H = A$ 时, A 为 Hermite 矩阵 ($F = R$ 时, $A^H = A^T$, A 为对称矩阵). Hermite 矩阵是可对角化的矩阵, 从而可用上述谱分解将其分解为矩阵的和. 下面的定理将显示当 A 为半正定的 Hermite 矩阵时, A 可分解为半正定矩阵的和.

定理3.6 设 $A \in F^{n \times n}$ 是半正定的 Hermite 矩阵. $\text{rank}(A) = k$, 则 A 可被分解为下列矩阵的和

$$A = v_1 v_1^H + v_2 v_2^H + \dots + v_k v_k^H,$$

其中 $v_i \in F^n$, $\{v_1, v_2, \dots, v_k\}$ 是空间 F^n 中非零的正交向量组.

证明 由已知 $\text{rank}(A) = r$, 且 A 为 Hermite 矩阵, 可得 A 的特征值, $\lambda \geq 0$, 不妨设 $\lambda_j > 0$, $j = 1, 2, \dots, k$; $\lambda_j = 0$, $j = k+1, \dots, n$. 且有酉矩阵 U , 使 $A = U \Lambda U^H$.

令 $U = \{u_1 \ u_2 \ \dots \ u_n\}$, 则可将 A 写为

$$\begin{aligned}
 A = (u_1 & u_2 & \cdots & u_n) &= \begin{pmatrix} \sqrt{-\lambda_1} & & & \\ & \sqrt{-\lambda_2} & & \\ & & \ddots & \\ & & & \sqrt{-\lambda_k} \\ & & & 0 \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix} \begin{pmatrix} u_1^H \\ u_2^H \\ \vdots \\ u_n^H \end{pmatrix} \\
 &= v_1 v_1^H + v_2 v_2^H + \cdots + v_k v_k^H.
 \end{aligned}$$

其中 $v_i = \sqrt{-\lambda_i} u_i$, $i = 1, 2, \dots, k$, 从而 $\{v_1, v_2, \dots, v_k\}$ 为正交的向量组. \square

3.2 Schur 分解与正规矩阵

在内积空间中,酉相似(正交相似)比一般可逆矩阵下的相似有更多的优点. 在欧氏空间中,一个实对称矩阵 A ($A^T = A$) 一定可正交相似于对角形: 即存在正交矩阵 C , 使

$$C^T A C = C^{-1} A C = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}.$$

相应于酉空间,一个 Hermite 矩阵 A ($A^H = A$) 一定可酉相似于对角形: 即存在酉矩阵 U , 使

$$U^H A U = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}.$$

一般方阵是可以相似于它的 Jordan 标准形——一个三角矩阵. 这一节讨论一个方阵是否能酉相似于上三角形矩阵, 以及矩阵满足什么条件可酉相似于对角矩阵的问题.

一、Schur 分解

Schur 分解将给出任何一个方阵都可酉相似于一个三角形矩阵的结果. 为建立这个结果, 我们先讨论 UR 分解.

定理3.7 设 $A \in C^{n \times n}$ 为可逆矩阵, 则存在酉矩阵 $U \in C^{n \times n}$ 和主对角线上元素皆为正的上三角矩阵

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \\ & & & r_{nn} \end{pmatrix}, \quad r_{ii} > 0; i = 1, 2, \dots, n,$$

使

$$A = UR.$$

证明 因为 A 可逆, 则 A 的列向量组 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 是空间 C^n 的一组基. 可对它施行 Schmidt 正交化过程, 得到 C^n 空间中的标准正交基 $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$.

由 1.2 节中的(1.20)式, 可有矩阵关系式

$$(\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n) = (\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_n) \begin{pmatrix} \|\beta_1\| & (\alpha_2, \varepsilon_1) & \cdots & (\alpha_n, \varepsilon_1) \\ & \|\beta_2\| & \cdots & (\alpha_n, \varepsilon_2) \\ & & \ddots & | \\ & & & \|\beta_n\| \end{pmatrix}, \quad (3.13)$$

令 $U = (\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_n)$, 则 $U \in C^{n \times n}$ 为酉矩阵.

$$R = \begin{pmatrix} \|\beta_1\| & (\alpha_2, \varepsilon_1) & \cdots & (\alpha_n, \varepsilon_1) \\ & \|\beta_2\| & \cdots & (\alpha_n, \varepsilon_2) \\ & & \ddots & | \\ & & & \|\beta_n\| \end{pmatrix},$$

则 R 为上三角矩阵, 且主对角线元素 $r_{ii} = \|\beta_i\| > 0, i = 1, 2, \dots, n$, (3.13) 式就是

$$A = UR.$$

□

例 7 设 $A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$, 求 A 的 UR 分解.

解 A 是一个可逆矩阵, A 的列向量为

$$A_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

用 Schmidt 正交化过程, 可从 A 的列向量得到 R^3 中标准正交的向量组

$$u_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad u_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}, \quad u_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}.$$

$$\begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

因此

$$U = (u_1 \ u_2 \ u_3) = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

U 为酉矩阵, 由定理 3.7, $R = U^H A$, 即

$$R = U^H A = \begin{pmatrix} \sqrt{-3} & \frac{2}{\sqrt{-3}} & \frac{1}{\sqrt{-3}} \\ 0 & \frac{2}{\sqrt{-6}} & \frac{1}{\sqrt{-6}} \\ 0 & 0 & \frac{1}{\sqrt{-2}} \end{pmatrix}.$$

得到 A 的 UR 分解

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{-3}} & -\frac{2}{\sqrt{-6}} & 0 \\ \frac{1}{\sqrt{-3}} & \frac{2}{\sqrt{-6}} & -\frac{1}{\sqrt{-2}} \\ \frac{1}{\sqrt{-3}} & \frac{2}{\sqrt{-6}} & \frac{1}{\sqrt{-2}} \end{pmatrix} \begin{pmatrix} \sqrt{-3} & \frac{2}{\sqrt{-3}} & \frac{1}{\sqrt{-3}} \\ 0 & \frac{2}{\sqrt{-6}} & \frac{1}{\sqrt{-6}} \\ 0 & 0 & \frac{1}{\sqrt{-2}} \end{pmatrix}.$$

定理 3.7 是对可逆方阵给出的 UR 分解, 这一结果可以推广到列满秩的矩阵. 定理 3.8 给出列满秩矩阵的 QR 分解.

定理 3.8 设 $A \in C^{n \times k}$ 是一个列满秩的矩阵, 即 $\text{rank}(A) = k$, 则 A 可被分解为

$$A = QR,$$

其中 $Q \in C^{n \times k}$, Q 的列向量是 A 的列空间的标准正交基. $R \in C^{k \times k}$ 是一个可逆的上三角矩阵.

证明 当 A 为列满秩时, 有 $m \geq k$. 将 A 的列向量扩充为空间 C^n 的基, 得到可逆矩阵 $(A \mid A_1) \in C^{n \times m}$, 由定理 3.7, 有 UR 分解

$$(A \mid A_1) = UR.$$

将酉矩阵 $U \in C^{n \times m}$ 分块为 $U = (Q \mid Q_1)$, $Q \in C^{n \times k}$.

将上三角矩阵 R 分块为 $R = \begin{pmatrix} R_1 & R_2 \\ \cdots & \cdots \\ 0 & R_3 \end{pmatrix}$,

其中 $R_1 \in C^{k \times k}$ 是上三角矩阵, 由于

$$(A \mid A_1) = (Q \mid Q_1) \begin{pmatrix} R_1 & R_2 \\ \cdots & \cdots \\ 0 & R_3 \end{pmatrix} = (QR_1 \mid QR_2 + Q_1 R_3),$$

有

$$A = QR_1.$$

$A = QR_1$ 即为满足题意的 QR 分解.

定理 3.9 (Schur 分解) 设 $A \in C^{n \times n}$, 则存在酉矩阵 U 和上三角矩阵 T , 使得

$$U^H A U = T = \begin{pmatrix} \lambda_1 & t_{12} & \cdots & t_{1n} \\ & \lambda_2 & \cdots & t_{2n} \\ & & \ddots & \vdots \\ & & & \lambda_n \end{pmatrix},$$

其中 λ 为矩阵 A 的特征值, $i=1, 2, \dots, n$.

证明 因为 $A \in C^{n \times n}$, 故 A 可相似于 Jordan 标准形,

$$A = PJP^{-1}.$$

又 $P \in C^{n \times n}$ 为可逆矩阵, 则由定理 3.7, P 有 UR 分解 $P = UR$, 则

$$A = PJP^{-1} = URJR^{-1}U^H.$$

令 $T = RJR^{-1}$, 则 T 是一个上三角形矩阵, 即有

$$U^H A U = T.$$

由相似矩阵特征值相等, 上三角形矩阵 T 的主对角线元素是 A 的全部特征值: $\lambda_1, \lambda_2, \dots, \lambda_n$.

Schur 分解一方面指出方阵 A 可酉相似于三角矩阵, 另一方面, $A = UTU^H$ 的证明过程, 也给出矩阵 A 的 Schur 分解.

Schur 分解有重要的理论意义, 可以用于证明许多定理.

例 8 设 $A = (a_{ij}) \in F^{n \times n}$, 证明 $\forall \varepsilon > 0$, 存在矩阵 $A(\varepsilon) = (a_{ij}(\varepsilon)) \in F^{n \times n}$, $A(\varepsilon)$ 有 n 个互异的特征值, 而且使得

$$\sum_{i,j=1}^n |a_{ij} - a_{ij}(\varepsilon)|^p < \varepsilon.$$

证明 对方阵 A , 由 Schur 分解, 存在酉矩阵 U , 使

$$U^H A U = T = \begin{pmatrix} t_{11} & & & \\ & t_{22} & * & \\ & & \ddots & \\ & & & t_{nn} \end{pmatrix}.$$

$$E = \begin{pmatrix} l_1 & & & \\ & l_2 & & \\ & & \ddots & \\ & & & l_n \end{pmatrix},$$

取 其中 l_i 满足性质:

$$|l_i| < \left(\frac{\varepsilon}{n} \right)^{\frac{1}{2}},$$

而且使 $t_{11} + l_1, t_{22} + l_2, \dots, t_{nn} + l_n$ 是 n 个互不相同的数, 从而 $T + E$ 有 n 个互异的特征值 $t_{ii} + l_i, i=1, 2, \dots, n$.

令 $A(\varepsilon) = A + U^H E U = U^H (T + E) U$, 则 $A(\varepsilon)$ 有 n 个互异的特征值, $A - A(\varepsilon) = -UEU^H$, 而且

$$\sum_{i,j} |a_{ij} - a_{ij}(\varepsilon)|^p = \sum_{i=1}^n |l_i|^p < n \frac{\varepsilon}{n} = \varepsilon. \quad \square$$

二、正规矩阵

定义 3.3 若矩阵 $A \in C^{n \times n}$ 满足 $A^H A = AA^H$. 则称 A 是一个正规矩阵 (normal

matrix).

正规矩阵是一类常见的矩阵,对实矩阵 A ,正规条件为: $A^T A = AA^T$.

例9 下列矩阵都是正规矩阵

(1) 对角矩阵;

(2) 对称与反对称矩阵: $A^T = A$, $A^H = -A$;

(3) Hermite 与反 Hermite 矩阵: $A^H = A$, $A^H = -A$;

(4) 正交矩阵与酉矩阵, $A^T A = AA^T = I_n$, $A^H A = AA^H = I_n$.

验证留做练习.

例10 设 A 为正规矩阵, B 酉相似于 A , 证明 B 也是正规矩阵.

证明 设 U 为酉矩阵, 使 $B = U^H AU$, 则

$$B^H B = U^H A^H U U^H A U = U^H A^H A U,$$

又

$$B B^H = U^H A U U^H A^H U = U^H A A^H U,$$

由 A 为正规矩阵, $A^H A = AA^H$, 得

$$B^H B = B B^H.$$

由定义, B 为正规矩阵.

例9 说明正规矩阵具有酉相似的不变性. 正规矩阵的基本特性如下所述.

定理3.10 $A \in C^{n \times n}$ 是正规矩阵的充分必要条件是 A 酉相似于对角矩阵, 即存在酉矩阵 $U \in C^{n \times n}$,

$$\text{使 } U^H A U = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}. \quad (3.14)$$

证明 必要性: 若 A 满足 $A^H A = AA^H$, 则有 Schur 分解, 存在酉矩阵 U , 使 $A = UTU^H$, 从例9 推出上三角形矩阵 T 满足: $T^H T = TT^H$, 由

$$T = \begin{pmatrix} \lambda_1 & t_{12} & \cdots & t_{1n} \\ & \lambda_2 & \cdots & t_{2n} \\ & & \ddots & \vdots \\ & & & \lambda_n \end{pmatrix},$$

比较 TT^H 与 $T^H T$ 的第 i 行, 第 i 列元素:

$$\begin{cases} (TT^H)_{ii} = \sum_{j=1}^n |t_{ij}|^2 + |\lambda_i|^2, \\ (T^H T)_{ii} = \sum_{j=1}^n |t_{ji}|^2 + |\lambda_i|^2, \end{cases} \quad i = 1, 2, \dots, n.$$

由 $(T^H T)_{ii} = (TT^H)_{ii}$ 得出: $t_{ij} = 0$ ($i \neq j$), 因此 T 为对角矩阵:

$$T = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}.$$

充分性: 因为对角矩阵是正规矩阵, 又 A 酉相似于正规矩阵, 由例 9, A 是正规矩阵. \square

推论 $A \in C^{n \times n}$ 是正规矩阵的充分必要条件是 A 有 n 个线性无关的特征向量构成空间 C^n 的标准正交基.

正规矩阵酉相似于对角形, 能使很多问题的证明得到简化.

例 11 证明 Hermite 矩阵的特征值是实数, 而且属于不同特征值的特征向量是正交的.

证明 设 A 为 Hermite 矩阵, 则由于 A 是正规矩阵, 所以存在酉矩阵 U , 使

$$U^H A U = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix},$$

作为 Hermite 矩阵, $A^H = A$, 即有

$$\begin{pmatrix} \bar{\lambda}_1 & & \\ & \bar{\lambda}_2 & \\ & & \ddots \\ & & & \bar{\lambda}_n \end{pmatrix} = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_n \end{pmatrix},$$

即 $\bar{\lambda}_i = \lambda_i, i = 1, 2, \dots, n$, 因此 λ_i 为实数.

又设 A 的谱为 $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$, 由 A 正规,

$$C^n = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_n},$$

设 C^n 的标准正交基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 是 A 的 n 个特征向量, 则 $V_{\lambda_1} = L\{\alpha_1, \alpha_2, \dots, \alpha_{r_1}\}$, $V_{\lambda_2} = L\{\alpha_{r_1+1}, \dots, \alpha_{r_1+r_2}\}$, \dots , $V_{\lambda_s} = L\{\alpha_{r_1+r_2+\dots+r_{s-1}+1}, \dots, \alpha_n\}$, 因此当 $\lambda_i \neq \lambda_j$ 时, V_{λ_i} 与 V_{λ_j} 为彼此正交的子空间, 这说明 A 关于不同特征值的特征向量是正交的.

例 12 证明酉矩阵的特征值的模长为 1, 即分布在复平面的单位圆上.

证明 设 A 为酉矩阵, 则 A 正规, 即存在酉矩阵 U , 使

$$A = U \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_n \end{pmatrix} U^H,$$

则 $A^H A = I_n \Leftrightarrow U \begin{pmatrix} |\lambda_1|^2 & & \\ & |\lambda_2|^2 & \\ & & \ddots \\ & & & |\lambda_n|^2 \end{pmatrix} U^H = I_n,$

即

$$|\lambda_i|^2 = 1, i=1, 2, \dots, n.$$

下面我们讨论正规矩阵的谱分解, 给出一个矩阵是正规矩阵的另一个充分必要条件.

定理 3.11 (正规矩阵的谱分解) 设 $A \in C^{n \times n}$, A 的谱为 $\{\lambda_1, \lambda_2, \dots, \lambda_s\}$, $s \leq n$, 则 A 是正规矩阵的充分必要条件是 A 有如下谱分解

$$A = \sum_{i=1}^s \lambda_i P_i, \quad (3.15)$$

其中 $P_i \in C^{n \times n}$, 满足如下条件

$$(1) P_i^2 = P_i, P_i^H = P_i, i=1, 2, \dots, s; \quad (3.16)$$

$$(2) P_i \cdot P_j = 0, i \neq j; \quad (3.17)$$

$$(3) I_n = \sum_{i=1}^s P_i. \quad (3.18)$$

证明 充分性: 由 P_i 满足的性质

$$AA^H = (\sum_{i=1}^s \lambda_i P_i)(\sum_{j=1}^s \lambda_j P_j^H) = \sum_{i=1}^s |\lambda_i|^2 P_i = A^H A,$$

所以 A 为正规矩阵.

必要性: 设 λ_i 为 A 的 r_i 重根, $\sum_{i=1}^s r_i = n$. 由 A 为正规矩阵, 存在酉矩阵 U , 使得

$$\begin{aligned} A &= U \begin{pmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_1 & \\ & & & \lambda_2 & \\ & & & & \ddots & \\ & & & & & \lambda_2 & \\ & & & & & & \ddots & \\ & & & & & & & \lambda_s & \\ & & & & & & & & \ddots & \\ & & & & & & & & & \lambda_s & \end{pmatrix}_{r_1}^{r_s} U^H \\ &= U \begin{pmatrix} \lambda_1 I_{r_1} & & & \\ & \lambda_2 I_{r_2} & & \\ & & \ddots & \\ & & & \lambda_s I_{r_s} \end{pmatrix} U^H = U \sum_{i=1}^s \lambda_i \begin{pmatrix} 0 & & & \\ & \ddots & & \\ & & I_{r_i} & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} U^H \end{aligned}$$

$$= \sum_{i=1}^s \lambda U \begin{pmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & I_{r_i} \\ & & & & 0 \\ & & & & & \ddots \\ & & & & & & 0 \end{pmatrix} U^H.$$

$$\text{令 } P_i = U \begin{pmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & I_{r_i} \\ & & & & 0 \\ & & & & & \ddots \\ & & & & & & 0 \end{pmatrix} U^H,$$

则 P_i 满足(3.16)式~(3.18)式,而且

$$A = \sum_{i=1}^s \lambda P_i.$$

在正规矩阵的谱分解中,矩阵 P_i 不仅是幂等矩阵,而且是 Hermite 矩阵,在后面章节,我们将会看到 P_i 不仅是投影,而且是正交投影矩阵.

3.3 矩阵的奇异值分解

从 3.1 节提到的矩阵的等价标准形,已知一个秩为 r 的矩阵 $A_{m \times n}$,存在可逆矩阵 $P_{m \times m}, Q_{n \times n}$,使

$$A = P \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Q.$$

这一节讨论矩阵“酉等价”的问题.将证明对一个秩为 r 的矩阵 $A_{m \times n}$,有酉矩阵 $U_{m \times m}, V_{n \times n}$,使

$$A = U \Sigma V^H,$$

矩阵 $\Sigma_{n \times n} = \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix}$, Δ 是一个可逆的对角矩阵.这一重要结果就是矩阵的奇异值分解.

矩阵的奇异值分解是在线性动态系统的辨识,最佳逼近问题,实验数据处理,数字图像存储中应用广泛的一种分解.这节给出矩阵 A 的奇异值概念和奇异值的分解的有关结果.

一、矩阵的奇异值及其性质

一个矩阵 $A \in C^{n \times n}$ 的奇异值是与矩阵 $A^H A$ 和 AA^H 相关联的概念, 在建立奇异值的概念之前, 我们先讨论矩阵 $A^H A$ 和 AA^H 的有关性质.

当 $A \in C^{n \times n}$ 时, $A^H A \in C^{n \times n}$, $AA^H \in C^{n \times n}$ 都是 Hermite 矩阵, 从而都是正规矩阵.

定理 3.12 设 $A \in C^{n \times n}$, 则矩阵 $A^H A \in C^{n \times n}$ 和矩阵 $AA^H \in C^{n \times n}$ 具有如下性质

(1) $\text{rank}(A) = \text{rank}(A^H A) = \text{rank}(AA^H)$.

(2) $A^H A$ 和 AA^H 的非零特征值相等.

(3) $A^H A$ 与 AA^H 都是半正定矩阵, 当 $\text{rank}(A) = n$ 时, $A^H A$ 为正定矩阵, 当 $\text{rank}(A) = m$ 时, AA^H 为正定矩阵.

证明 这些基本上都是线性代数或前面内容中已建立了的结论, 我们只证(3), 其余留做练习.

取复二次型 $f = X^H A^H AX$, 则 $\forall X \neq 0, f = X^H A^H AX = (AX)^H (AX) = (AX, AX) \geq 0$, 又 $\text{rank}(A) = n$ 时, 线性方程组 $AX = 0$ 只有零解, 故 $\forall X \neq 0$, 有 $AX \neq 0$, 从而

$$f = (AX, AX) > 0,$$

即 f 正定, 亦即矩阵 $(A^H A)$ 正定.

同理可证 AA^H 的相应结果.

当 AA^H 和 $A^H A$ 半正定时, 可推出 AA^H 和 $A^H A$ 的特征值 $\lambda \geq 0$; 当它们正定时, $\lambda > 0$, $i = 1, 2, \dots, n$. 由此我们可定义矩阵 A 的奇异值. \square

定义 3.4 对于 $A \in C^{n \times n}$, $\text{rank}(A) = r$, 矩阵 $A^H A$ 的特征值为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$, $\lambda_{r+1} = \lambda_{r+2} = \dots = \lambda_n = 0$, 称正数 $\sigma_i = \sqrt{\lambda_i}$ ($i = 1, 2, \dots, r$) 为矩阵 A 的奇异值, 简称 A 的奇值.

由定义 3.4 知, 矩阵 $A \in C^{n \times n}$ 有 r 个奇异值, 它们的个数等于矩阵 A 的秩 r . 值得指出的是, 对一般的 $n \times n$ 矩阵 A , A 的特征值和 A 的秩则没有这样的结论.

例如

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

A 秩为 3, 但 A 的特征值全部为 0, 或者说 A 的非零特征值的个数为零.

定理 3.13 矩阵 A 的奇异值具有如下性质:

(1) $A \in C^{n \times n}$ 为正规矩阵时, A 的奇异值为 A 的特征值的模, $|\lambda_i|$, $i = 1, 2, \dots, n$.

(2) $A \in C^{n \times n}$ 为正定的 Hermite 矩阵时, A 的奇异值等于 A 的特征值.

(3) 若存在酉矩阵 $U \in C^{n \times n}$, $V \in C^{n \times n}$, 矩阵 $B \in C^{n \times n}$, 使 $UAV = B$, 则称 A 和 B 酉等价. 酉等价的矩阵 A 和 B 有相同的奇异值.

证明 (1) A 为正规矩阵, 有酉矩阵 $U \in C^{n \times n}$, 使

$$A = U \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} U^H,$$

所以 $A^H A = U \begin{pmatrix} |\lambda_1|^2 & & & \\ & |\lambda_2|^2 & & \\ & & \ddots & \\ & & & |\lambda_n|^2 \end{pmatrix} U^H,$

即 $A^H A$ 的特征值为 $|\lambda_i|^2$, 从而 A 的奇异值

$$\sigma_i = \sqrt{|\lambda_i|^2} = |\lambda_i|, \quad i = 1, 2, \dots, n.$$

(2) A 为正定 Hermite 矩阵时, A 的特征值为实数, 而且 $\lambda_i > 0, i = 1, 2, \dots, n$. 又 A 为正规矩阵. 由(1), A 的奇异值 $\sigma_i = |\lambda_i| = \lambda_i, i = 1, 2, \dots, n$.

(3) 设 $UAV = B$, 则

$$B^H B = V^H A^H U^H U A V = V^H A^H A V \quad (3.19)$$

(3.19)式说明 A 和 B 酉等价时, $A^H A$ 和 $B^H B$ 酉相似, 从而 $A^H A$ 和 $B^H B$ 有相同的特征值, 从而 A 和 B 有相同的奇异值. \square

二、矩阵的奇异值分解

定理 3.14 设矩阵 $A \in C^{n \times n}$, $\text{rank}(A) = r$. $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ 是矩阵 A 的奇异值, 则存在酉矩阵 $U \in C^{n \times n}$, $V \in C^{n \times n}$, 分块矩阵 $\Sigma = \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix} \in C^{n \times n}$, 使

$$A = U \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix} V^H,$$

其中

$$\Delta = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{pmatrix}.$$

证明 已知 $\text{rank}(A^H A) = \text{rank}(A) = r$, 设 $A^H A$ 的 n 个特征值按大小排列为

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0, \lambda_{r+1} = \dots = \lambda_n = 0.$$

对正规矩阵 $A^H A$, 存在酉矩阵 $V \in C^{n \times n}$, 使

$$V^H A^H A V = \begin{pmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_n & \\ & & 0 & \ddots \\ & & & 0 \end{pmatrix} = \begin{pmatrix} \Delta^2 & 0 \\ 0 & 0 \end{pmatrix}_{n \times n},$$

将 V 按列分块为 $V = (v_1, v_2, \dots, v_n)$, 它的 n 个列 $\{v_1, v_2, \dots, v_n\}$ 是对应于特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 的标准正交的特征向量.

为了得到酉矩阵 U , 首先考查 C^* 中的向量组 $\{Av_1, Av_2, \dots, Av_r\}$,

$$\begin{aligned} (Av_i, Av_j) &= (Av_j)^H (Av_i) = v_j^H A^H A v_i = v_j^H \lambda v_i \\ &= \lambda v_i^H v_j = 0, \quad i \neq j, \end{aligned}$$

所以 $\{Av_1, Av_2, \dots, Av_r\}$ 是 C^* 中的正交向量组.

又

$$\|Av_i\|^2 = v_i^H A^H A v_i = \lambda v_i^H v_i = \sigma_i^2,$$

所以

$$\|Av_i\| = \sigma_i.$$

令

$$u_i = \frac{1}{\sigma_i} Av_i, \quad i = 1, 2, \dots, r,$$

则得到 C^* 中标准正交的向量组 $\{u_1, u_2, \dots, u_r\}$. 把它扩充为 C^* 中的标准正交基 $\{u_1, \dots, u_r, \dots, u_m\}$, 令

$$U = (u_1 \ u_2 \ \cdots \ u_r \ \cdots \ u_m),$$

就得到了酉矩阵 $U \in C^{n \times m}$.

已知 $Av_i = \sigma_i u_i, i = 1, 2, \dots, r, Av_i = 0, i = r+1, \dots, n$,

从而 $AV = A(v_1 \ v_2 \ \cdots \ v_n) = (Av_1 \ Av_2 \ \cdots \ Av_r \ 0 \ \cdots \ 0)$

$$\begin{aligned} &= (\sigma_1 u_1 \ \sigma_2 u_2 \ \cdots \ \sigma_r u_r \ 0 \ \cdots \ 0) \\ &= (u_1 \ u_2 \ \cdots \ u_m) \begin{pmatrix} \sigma_1 & \cdots & 0 & | & 0 \\ \vdots & \ddots & \vdots & | & 0 \\ 0 & \cdots & \sigma_r & | & 0 \\ \hline 0 & & & | & 0 \end{pmatrix} = U\Sigma. \end{aligned}$$

有

$$AV = U\Sigma, \quad \text{即 } A = U\Sigma V^H.$$

例 13 求矩阵 $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ 和 $B = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ 的奇异值分解.

解 (1) $A^H A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix}$, $|M - A^H A| = (\lambda - 3)(\lambda - 1)\lambda$,

所以 $\lambda_1 = 3, \lambda_2 = 1, \lambda_3 = 0, \text{rank}(A) = 2$. 由此

$$\sigma_1 = \sqrt{-3}, \quad \sigma_2 = 1, \quad \sigma_3 = 0, \quad \Delta = \begin{pmatrix} \sqrt{-3} & 0 \\ 0 & 1 \end{pmatrix}.$$

又 $A^H A$ 的特征值 $\lambda_1 = 3, \lambda_2 = 1, \lambda_3 = 0$ 对应的特征向量分别是

$$\alpha_1 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \quad \alpha_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad \alpha_3 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}.$$

由于 $A^H A$ 是 Hermite 矩阵, $\alpha_1, \alpha_2, \alpha_3$ 是正交的, 只需将其标准化, 得正交矩阵

$$V = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \end{pmatrix}.$$

$$\text{取 } u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{\sqrt{-3}} A v_1 = \begin{pmatrix} \frac{1}{\sqrt{-2}} \\ \frac{1}{\sqrt{-2}} \\ 0 \end{pmatrix}, \quad u_2 = \frac{1}{\sigma_2} A v_2 = \begin{pmatrix} \frac{1}{\sqrt{-2}} \\ -\frac{1}{\sqrt{-2}} \\ 0 \end{pmatrix},$$

$$\text{则 } U_1 = (u_1 \ u_2) = \begin{pmatrix} \frac{1}{\sqrt{-2}} & \frac{1}{\sqrt{-2}} \\ \frac{1}{\sqrt{-2}} & -\frac{1}{\sqrt{-2}} \\ 0 & 0 \end{pmatrix},$$

$$U_1^\perp = \{\beta \mid u_1^\top \beta = 0, \ u_2^\top \beta = 0\}.$$

$$\text{取子空间 } U_1^\perp \text{ 的基 } \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \text{ 得 } U_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

$$\text{从而 } U = (U_1 \ | \ U_2) = \begin{pmatrix} \frac{1}{\sqrt{-2}} & \frac{1}{\sqrt{-2}} & 0 \\ \frac{1}{\sqrt{-2}} & -\frac{1}{\sqrt{-2}} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

$$A \text{ 的奇异值分解 } A = U \begin{pmatrix} \sqrt{-3} & & \\ & 1 & \\ & & 0 \end{pmatrix} V^H.$$

$$(2) B^H B = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

可计算得 $B^H B$ 的特征值 $\lambda_1 = 2, \lambda_2 = 1, \lambda_3 = 0$. 它们相应的标准正交的特征向量

$$v_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}.$$

A 的奇异值 $\sigma_1 = \sqrt{2}, \sigma_2 = 1, \sigma_3 = 0, \text{rank}(A) = 2$,

$$\text{由此 } V = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

计算求 U :

$$u_1 = \frac{1}{\sigma_1} B v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix},$$

$$u_2 = \frac{1}{\sigma_2} B v_2 = \frac{1}{1} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

u_1 与 u_2 已是 \mathbb{R}^2 的标准正交基. 因此, 有

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

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{pmatrix} = U \Sigma V^T.$$

在 A 的奇异值分解中, 矩阵 V 的列被称为 A 的右奇异向量. V 的前 r 列是 $A^H A$ 的 r 个非零特征值所对应的特征向量, 将它们取为矩阵 V_1 , 则 $V = (V_1 \mid V_2)$. 矩阵 U 的列被称为 A 的左奇异向量. 将 U 从前 r 列处分块为 $U = (U_1 \mid U_2)$, 由分块运算, 有

$$U^H A V = \begin{pmatrix} U_1^H \\ U_2^H \end{pmatrix} A (V_1 \quad V_2) = \begin{pmatrix} U_1^H A V_1 & U_1^H A V_2 \\ U_2^H A V_1 & U_2^H A V_2 \end{pmatrix} = \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix}$$

从而

$$A V_2 = 0, \quad A V_1 = u_1 \Delta.$$

因此,有下列结果:

- (1) V_2 的列向量是空间 $N(A)$ 的一个标准正交基,
- (2) V_1 的列向量是空间 $N^\perp(A)$ 的一个标准正交基,
- (3) U_1 的列向量是空间 $R(A)$ 的一个标准正交基,
- (4) U_2 的列向量是空间 $R^\perp(A)$ 的一个标准正交基.

在 A 的奇异值分解中,酉矩阵 U 和 V 不是惟一的. A 的奇异值分解给出矩阵 A 的许多重要信息.

更进一步, $U = (u_1 \quad u_2 \quad \cdots \quad u_n), V = (v_1 \quad v_2 \quad \cdots \quad v_n)$ 可借助于奇异值分解,将 A 表示为:

$$A = U \Sigma V^H = (u_1 \quad u_2 \quad \cdots \quad u_n) \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \\ \hline & 0 & & \\ \hline & & & 0 \end{pmatrix} \begin{pmatrix} v_1^H \\ v_2^H \\ \vdots \\ v_n^H \end{pmatrix}$$

$$= \sigma_1 u_1 v_1^H + \sigma_2 u_2 v_2^H + \cdots + \sigma_r u_r v_r^H.$$

归纳这一结果,有如下定理.

定理 3.15 设 $A \in C^{n \times n}$, A 的奇异值为 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, 令 u_1, u_2, \dots, u_r 是相应于奇异值的左奇异向量, v_1, v_2, \dots, v_r 是相应于奇异值的右奇异向量, 则

$$A = \sigma_1 u_1 v_1^H + \sigma_2 u_2 v_2^H + \cdots + \sigma_r u_r v_r^H. \quad (3.20)$$

□

式(3.20)给出的形式被称为矩阵 A 的奇异值展开式. 对一个数 $k \leq r$, 略去 A 的一些小的奇异值对应的项, 取矩阵 A_k ,

$$A_k = \sigma_1 u_1 v_1^H + \sigma_2 u_2 v_2^H + \cdots + \sigma_k u_k v_k^H,$$

则 A_k 是一个秩为 k 的 $m \times n$ 阶矩阵. 在第5章将会看到, A_k 是在所有的秩为 k 的 $m \times n$ 阶矩阵中, 从 Frobenius 范数的意义上, 与矩阵 A 距离最近的一个矩阵. 这在实际中应用广泛. 例如, 在图像数字化技术中, 一幅图片可转换成一个 $m \times n$ 阶像素矩阵来存储, 存储量是 $m \times n$ 个数. 如果利用矩阵 A 的奇异值展开式, 则只要存储 A 的奇异值 σ_i , 奇异向量 u_i, v_i 的分量, 总计 $r(m + n + 1)$ 个数. 取 $m = n = 1000, r = 100$ 作一个比较,

$$m \times n = 10^6,$$

$$r(m + n + 1) = 100(1000 + 1000 + 1) = 200100.$$

取 A 的奇异值展开式, 存储量较 A 的元素情形减少了 80%. 另外, 可取 $k < r$. 用 A_k 来逼近

A ,能达到既压缩数据的存储量,又保持图像不失真的目的.

三、矩阵 A 的奇异值分解与线性变换 T_A

设 A 是一个 $m \times n$ 阶复矩阵,即 $A \in \mathbb{C}^{m \times n}$,则由 $\beta = T_A(\alpha) = A\alpha$ 可以定义线性变换

$$T_A: \mathbb{C}^n \rightarrow \mathbb{C}^m.$$

设矩阵 A 有奇异值分解, $A = U\Sigma V^H$, 则将矩阵 V ($\in \mathbb{C}^{n \times n}$) 的列向量 $\{v_1, v_2, \dots, v_n\}$ 取作空间 \mathbb{C}^n 的标准正交基; 将矩阵 U ($\in \mathbb{C}^{m \times m}$) 的列向量 $\{u_1, u_2, \dots, u_m\}$ 取作空间 \mathbb{C}^m 的标准正交基, 则在所取的基下, 线性变换 T_A 对应的变换矩阵就是 Σ .

设 $\alpha \in \mathbb{C}^n$, α 在基 $\{v_1, v_2, \dots, v_n\}$ 下坐标向量为 $X = (x_1 \ x_2 \ \dots \ x_n)^T$, 即 $\alpha = VX$. 那么 α 在线性变换 T_A 下的像 β 具有形式:

$$\beta = T_A(\alpha) = A\alpha = (U\Sigma V^H)(VX) = U(\Sigma X) = U \begin{pmatrix} \sigma_1 X_1 \\ \sigma_2 X_2 \\ \vdots \\ \sigma_r X_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

所以, α 的像 $\beta = T_A(\alpha)$ 在 \mathbb{C}^m 中基 $\{u_1, u_2, \dots, u_m\}$ 下的坐标是

$$Y = \Sigma X = (\sigma_1 x_1 \ \sigma_2 x_2 \ \dots \ \sigma_r x_r \ 0 \ \dots \ 0)^T.$$

从中可以看到, 当 $\text{rank}(A) = r$ 时, 在取定的基下, 线性变换 $T_A(\alpha)$ 的作用是将原像坐标中的前 r 个分量分别乘以 A 的奇异值 $\sigma_1, \sigma_2, \dots, \sigma_r$ 后 $(n-r)$ 个分量化为零. 如果原像坐标满足条件:

$$\sum_{i=1}^n x_i^2 = 1,$$

则像坐标满足条件

$$\left(\frac{y_1}{\sigma_1} \right)^2 + \left(\frac{y_2}{\sigma_2} \right)^2 + \dots + \left(\frac{y_r}{\sigma_r} \right)^2 \leq 1.$$

在 $\text{rank}(A) = r = n$ 时, 等式成立. 因此, 有如下定理.

定理3.16 设 $A = U\Sigma V^H$ 是 $m \times n$ 阶实矩阵 A 的奇异值分解, $\text{rank}(A) = r$, 则 \mathbb{R}^r 中的单位球面在线性变换 T_A 下的像集合是:

- (1) 若 $r = n$, 则像集合是 \mathbb{R}^n 中的椭球面;
- (2) 若 $r < n$, 则像集合是 \mathbb{R}^n 中的椭球体.

例14 设 $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$, 求 \mathbb{R}^3 中单位球面在线性变换 $T_A: X \rightarrow AX$ 下像的几何图

形.

解 由例 13, 矩阵 A 有如下奇异值分解

$$A = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{-3} & & \\ & 1 & \\ & & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \end{pmatrix}.$$

$\text{rank}(A) = 2 < 3 = n$, 用定理 3.16, 单位球面的像满足不等式

$$\left(\frac{y_1}{\sqrt{3}}\right)^2 + \left(\frac{y_2}{1}\right)^2 \leq 1 \quad \text{或} \quad \frac{y_1^2}{3} + y_2^2 \leq 1,$$

即单位球的像是实心椭圆

$$\frac{y_1^2}{3} + y_2^2 \leq 1.$$

四、方阵的极分解

借助于矩阵的奇异值分解, 对方阵 $A \in C^{n \times n}$, 可以导出另一种在理论和应用中都很重要的矩阵分解——极分解(polar decomposition).

定理 3.17 设 $A \in C^{n \times n}$, $\text{rank}(A) = r$, 则 A 可以被分解为

$$A = PQ,$$

其中 P 是秩为 r 的 $n \times n$ 阶半正定矩阵, Q 是 $n \times n$ 阶的酉矩阵. 特别地, 若 $r = n$, 则 P 为正定矩阵.

证明 $A \in C^{n \times n}$, 有 A 的奇异值分解

$$A = U\Sigma V^H = U\Sigma U^H UV^H = (U\Sigma U^H)(UV^H).$$

令 $P = U\Sigma U^H$, 则 P 是 $n \times n$ 阶的 Hermite 矩阵. 又 P 酉相似于对角矩阵 Σ , 因此 P 的秩为 r , P 以 A 的奇异值为其非负的特征值, 从而 P 是半正定矩阵. 特别当 $r = n$ 时, A 为可逆矩阵, A 的奇异值皆非零, 故 P 的 n 个特征值大于零, 即 P 为正定矩阵.

令 $Q = UV^H$, 则 Q 为酉矩阵的乘积, 从而 Q 也是酉矩阵, 从而有分解: $A = PQ$. \square

A 的形如定理 3.17 的分解形式 $A = PQ$ 称为极分解, 是因为对非零复数 $z = a + bi \in C$, 可被表示为

$$z = re^{i\theta}, (r, \theta) \text{ 是 } z \text{ 的极坐标,}$$

其中

$$r = |z| > 0, |e^{i\theta}| = 1.$$

由此 z 被分解为一个拉伸因子 r 和一个旋转因子 $e^{i\theta}$ (将复数绕数轴逆时针转一个 θ 角) 的乘积.

在矩阵的极分解 $A = PO$ 中, P 的正定性相应 z 的非负性. 酉矩阵 O , $|b| = 1$ 相应于

一个旋转. 从而得名极分解.

例15 令 $A = \begin{pmatrix} \sqrt{-3} & 2 \\ 0 & \sqrt{-3} \end{pmatrix}$, 求矩阵 A 的极分解, 讨论 \mathbb{R}^2 上线性变换 $T_A: Y = AX$ 的几何意义.

解 $AA^T = \begin{pmatrix} 7 & 2\sqrt{-3} \\ 2\sqrt{-3} & 3 \end{pmatrix}$ 的特征值为 9 和 1, 所以 A 的奇异值 $\sigma_1 = 3, \sigma_2 = 1$. 易求得 A 的奇异值分解

$$A = \begin{pmatrix} \frac{\sqrt{-3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{-3}}{2} \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{-3}}{2} \\ -\frac{\sqrt{-3}}{2} & \frac{1}{2} \end{pmatrix},$$

$$P = U\Sigma U^H = \begin{pmatrix} \frac{\sqrt{-3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{-3}}{2} \end{pmatrix} \begin{pmatrix} 3 & \\ & 1 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{-3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{-3}}{2} \end{pmatrix} = \begin{pmatrix} \frac{5}{2} & \frac{\sqrt{-3}}{2} \\ \frac{\sqrt{-3}}{2} & \frac{5}{2} \end{pmatrix},$$

$$Q = UV^T = \begin{pmatrix} \frac{\sqrt{-3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{-3}}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{-3}}{2} \\ -\frac{\sqrt{-3}}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{-3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{-3}}{2} \end{pmatrix}.$$

A 的极分解为 $A = PQ$.

线性变换 T_A , 在 A 的极分解下,

$$Y = AX = P(QX).$$

Q 可表示为

$$Q = \begin{pmatrix} \cos \frac{\pi}{6} & \sin \frac{\pi}{6} \\ -\sin \frac{\pi}{6} & \cos \frac{\pi}{6} \end{pmatrix}.$$

从第一章知线性变换 QX 是一个绕坐标原点, 顺时针转动一个 $\frac{\pi}{6}$ 的旋转变换.

而 P 是一个实对称矩阵, 特征值 $\lambda_1 = 3, \lambda_2 = 1$, 相应于它们的标准正交特征向量为

$$u_1 = \begin{pmatrix} \frac{\sqrt{-3}}{2} \\ \frac{1}{2} \end{pmatrix}^T, \quad u_2 = \begin{pmatrix} -\frac{1}{2} \\ \frac{\sqrt{-3}}{2} \end{pmatrix}^T.$$

因为 $\{u_1, u_2\}$ 是 \mathbb{R}^2 中的一个基, 所以 $\forall X \in \mathbb{R}^2$,

$$X = k_1 u_1 + k_2 u_2,$$

$$\begin{aligned} PX &= k_1 P u_1 + k_2 P u_2 = \lambda_1 k_1 u_1 + \lambda_2 k_2 u_2 \\ &= 3k_1 u_1 + k_2 u_2. \end{aligned}$$

由此可知 PX 是一个拉伸变换, 将向量沿 u_1 方向拉伸 3 倍, 沿 u_2 方向不变.

归纳其上,可见线性变换 $Y=AX$ 在极分解下,可分解成为对应一个旋转加上一个拉伸变换的综合作用.

习 题 三

1. 设 $A=\begin{pmatrix} 2 & -1 & 2 \\ 4 & -4 & 6 \\ 6 & -7 & 8 \end{pmatrix}$,

(1) 求 A 的 LDV 分解.

(2) 设 $b=(1 \ 0 \ 2)^T$,用 LDV 分解求解方程组 $AX=b$.

2. $A=\begin{pmatrix} 5 & 2 & -4 & 0 \\ 2 & 1 & -2 & 1 \\ -4 & 2 & 5 & 0 \\ 0 & 1 & 0 & 2 \end{pmatrix}$,求矩阵 A 的 LDV 分解.

3. 设下列矩阵的满秩分解

(1) $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$; (2) $\begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 1 & -1 \\ 1 & 0 & 2 & 1 \end{pmatrix}$; (3) $\begin{pmatrix} 0 & 0 & 1 \\ 2 & 1 & 1 \\ 2i & i & 0 \end{pmatrix}$, $i=\sqrt{-1}$.

4. 设 $A \in C^{n \times n}$, A 可分块表示为 $A=\begin{pmatrix} X & Y \\ Z & W \end{pmatrix}$,其中 $X \in C^{r \times r}$, $\text{rank}(A)=\text{rank}(X)=r$, $W=ZX^{-1}Y$,证明 A 有下列形式的满秩分解:

$$A=\begin{pmatrix} X \\ Z \end{pmatrix}(I_r-X^{-1}Y), \quad A=\begin{pmatrix} I_r \\ ZX^{-1} \end{pmatrix}(X-Y).$$

5. 求矩阵 $A=\begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}$ 的谱分解.

6. 求矩阵 $A=\begin{pmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{pmatrix}$ 的谱分解.

7. 证明一个正规矩阵若是三角阵,则一定是对角阵.

8. 设 $A \in C^{n \times n}$ 为正规矩阵,证明

(1) A^H, A^k (k 为正整数)也是正规矩阵.

(2) 对多项式 $f(\lambda), f(A)$ 也是正规矩阵.

9. 判断下列矩阵是否为正规矩阵

(1) $A=\begin{pmatrix} -1 & i & 0 \\ -i & 0 & -i \\ 0 & -i & -1 \end{pmatrix}$; (2) $A=\begin{pmatrix} 0 & i & 1 \\ -i & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$.

10. 证明反对称矩阵 $A \in \mathbb{R}^{n \times n}$ ($A^T = -A$) 和反Hermite 矩阵 $B \in \mathbb{C}^{n \times n}$ ($B^H = -B$) 的特征值是零或纯虚数.

11. 证明如果一个实对称矩阵 A 主对角线元素 $a_{ii} > 0, i=1, 2, \dots, n$, 则 A 至少有一个正特征值.

12. 设 $A, B \in \mathbb{C}^{n \times n}$, A 和 B 的特征值分别是 $\lambda_1, \lambda_2, \dots, \lambda_n$ 和 $\mu_1, \mu_2, \dots, \mu_n$, 用 Schur 分解证明:

如果 $AB = BA$, 则 $A + B$ 的特征值为 $\lambda_1 + \mu_1, \lambda_2 + \mu_2, \dots, \lambda_n + \mu_n$.

13. 求下列矩阵的奇异值分解

$$(1) A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}. \quad (2) A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

14. 设 $A \in \mathbb{C}^{n \times n}$ 的奇值是 $\sigma_1, \sigma_2, \dots, \sigma_n$, 证明 $|A| = \pm \prod_{i=1}^n \sigma_i$, 即 A 的行列式的模为 $\prod_{i=1}^n \sigma_i$.

15. 设 $A \in \mathbb{C}^{n \times n}$, λ 是 A 的特征值, σ 是 A 的奇异值. 证明

$$\sum_i |\lambda_i|^2 \leq \text{tr}(A^H A) = \text{tr}(AA^H) = \sum_i \sigma_i^2$$

等式成立的充要条件是 A 为正规矩阵.

第 4 章 矩阵的广义逆

逆矩阵的概念只是对非奇异矩阵才有意义.但是在实际问题中,遇到的矩阵不一定是方阵,即便是方阵也不一定是非奇异矩阵,这就产生了推广逆矩阵的概念问题,产生了各种各样的矩阵广义逆.本章就其最重要的 Moore-Penrose 广义逆及其应用做一简单介绍.

4.1 矩阵的左逆与右逆

设 A 是 n 阶矩阵, A 可逆当且仅当存在 n 阶矩阵 B , 使得

$$AB = BA = I,$$

当 A 可逆时, 其逆惟一, 记为 A^{-1} .

现在, 我们把方阵的逆矩阵概念推广到 $m \times n$ 矩阵上, 定义一种单侧逆.

一、满秩矩阵与单侧逆

定义 4.1 设 $A \in \mathbf{C}^{m \times n}$, 若存在矩阵 $B \in \mathbf{C}^{n \times m}$, 使得

$$BA = I_n,$$

则称 A 是左可逆的, 称 B 为 A 的一个左逆矩阵, 记为 A_L^{-1} .

若存在矩阵 $C \in \mathbf{C}^{n \times m}$, 使得

$$AC = I_m,$$

则称 A 是右可逆的, 称 C 为 A 的一个右逆矩阵, 记为 A_R^{-1} .

下面给出矩阵左可逆与右可逆的几个等价条件.

定理 4.1 设 $A \in \mathbf{C}^{m \times n}$, 则下列条件是等价的:

- (1) A 是左可逆的;
- (2) A 的零空间 $N(A) = \{\mathbf{0}\}$;
- (3) $m \geq n$, $\text{rank}(A) = n$, 即 A 是列满秩的;
- (4) $A^H A$ 是可逆的.

我们采用循环证法来证明.

证明 (1) \Rightarrow (2), 设 A 是左可逆的, 则存在 $B \in \mathbf{C}^{n \times m}$, 使得 $BA = I_n$. $\forall x \in N(A)$, $Ax = \mathbf{0}$, 于是 $x = I_n x = BAx = \mathbf{0}$, 证得 $N(A) = \{\mathbf{0}\}$.

(2) \Rightarrow (3), 设 $N(A) = \{\mathbf{0}\}$, 由维数公式

$$\text{rank}(A) = \dim R(A) = n - \dim N(A) = n$$

知 A 是列满秩的, 自然有 $m \geq n$.

(3) \Rightarrow (4), 设 $\text{rank}(A) = n$, 由 $\text{rank}(A^H A) = \text{rank}(A) = n$, 知 $A^H A$ 是可逆的.

(4) \Rightarrow (1), 设 $A^H A$ 是可逆的, 由 $(A^H A)^{-1} A^H A = I_n$ 知, $(A^H A)^{-1} A^H$ 是 A 的一个左逆矩阵. \square

对于右可逆, 有类似的结果.

定理 4.2 设 $A \in \mathbf{C}^{m \times n}$, 则下列条件是等价的:

- (1) A 是右可逆的;
- (2) A 的列空间 $R(A) = \mathbf{C}^m$;
- (3) $m \leq n$, $\text{rank}(A) = m$, 即 A 是行满秩的;
- (4) $A A^H$ 是可逆的.

其证明留给读者. 由 $A A^H (A A^H)^{-1} = I_m$ 知 $A^H (A A^H)^{-1}$ 是 A 的一个右逆矩阵.

例 1 矩阵 $A = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \end{pmatrix}$ 是右可逆的, 不是左可逆的. 由于

$$\begin{pmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{5} \\ C_{31} & C_{32} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

注意到右逆最后一行元素是完全任意的, 故存在无穷多个右逆矩阵.

一般地, 一个矩阵左可逆未必右可逆, 而且左逆矩阵和右逆矩阵都不是惟一的.

二、单侧逆与解线性方程组

定理 4.3 设 $A \in \mathbf{C}^{m \times n}$ 是左可逆的, $B \in \mathbf{C}^{n \times m}$ 是 A 的一个左逆矩阵, 则线性方程组 $AX = b$ 有形如 $X = Bb$ 解的充分必要条件是

$$(I_m - AB)b = 0.$$

若上式成立, 则方程组有惟一解

$$X = (A^H A)^{-1} A^H b.$$

证明 设方程组 $AX = b$ 有解 X_0 , 则

$$b = AX_0 = A(BA)X_0 = ABb,$$

从而 $(I_m - AB)b = 0$. 反过来, 若 $(I_m - AB)b = 0$, 则 $ABb = b$, 从而 $X_0 = Bb$ 是方程组 $AX = b$ 的解.

当方程组有解时, 因为 A 左可逆, 所以 $\text{rank}(A) = n$, 从而方程组 $AX = b$ 有惟一解. 由 $AX = b$, 有

$$(A^H A)^{-1} A^H AX = (A^H A)^{-1} A^H b,$$

即 $X = (A^H A)^{-1} A^H b$ 为 $AX = b$ 的惟一解. \square

定理 4.4 设 $A \in \mathbf{C}^{m \times n}$ 是右可逆的, 则线性方程组 $AX = b$ 对任何 $b \in \mathbf{C}^m$ 都有解, 且对 A 的任意一个右逆矩阵 A_R^{-1} , $X = A_R^{-1}b$ 是其解. 特别地, $X = A^H(AA^H)^{-1}b$ 是方程组 $AX = b$ 的一个解.

证明 因 A 右可逆, 则 $AA_R^{-1} = I_m$. 对任何 $b \in \mathbf{C}^m$, 都有

$$AA_R^{-1}b = I_m b = b,$$

即 $X = A_R^{-1}b$ 是方程组 $AX = b$ 的解. \square

4.2 广义逆矩阵

有各种各样的矩阵广义逆, 本节仅介绍减号广义逆与加号广义逆(或 M-P 广义逆), 重点介绍应用广泛的加号广义逆.

一、减号广义逆

定义 4.2 设 $A \in \mathbf{C}^{m \times n}$, 若存在矩阵 $G \in \mathbf{C}^{n \times m}$, 使得

$$AGA = A,$$

则称 G 为 A 的一个减号广义逆或 $\{1\}$ -逆.

A 的全部减号广义逆的集合记为 $A\{1\}$, $A\{1\}$ 的元素用 A_1^- , A_2^- , … 表示.

设 $A = \mathbf{0} \in \mathbf{C}^{m \times n}$, 则 $A\{1\} = \mathbf{C}^{n \times m}$.

一般地, 对什么样的矩阵 $A \in \mathbf{C}^{m \times n}$, 它有减号广义逆? 怎样求减号广义逆? 这是首先需要回答的问题.

定理 4.5 设 $A \in \mathbf{C}^{m \times n}$, $\text{rank}(A) = r$, 若存在可逆阵 $P \in \mathbf{C}^{m \times m}$ 和 $Q \in \mathbf{C}^{n \times n}$, 使得

$$PAQ = \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

则 $G \in A\{1\}$ 的充分必要条件是

$$G = Q \begin{pmatrix} I_r & U \\ V & W \end{pmatrix} P,$$

其中 $U \in \mathbf{C}^{r \times (m-r)}$, $V \in \mathbf{C}^{(n-r) \times r}$, $W \in \mathbf{C}^{(n-r) \times (m-r)}$ 是任意的.

证明 充分性直接验证便知, 下证必要性. 设 G 是 A 的任意一个减号广义逆, 则 $AGA = A$. 令

$$G = Q \begin{pmatrix} X & U \\ V & W \end{pmatrix} P,$$

代入 $AGA = A$ 中, 比较即得 $X = I_r$, 故

$$G = Q \begin{pmatrix} I_r & U \\ V & W \end{pmatrix} P. \quad \square$$

例 2 设

$$\mathbf{A} = \begin{pmatrix} 0 & -1 & 3 & 0 \\ 2 & -4 & 1 & 5 \\ -4 & 5 & 7 & -10 \end{pmatrix},$$

求 \mathbf{A} 的减号广义逆.

$$\text{解 } \left[\begin{array}{cc|ccc} \mathbf{A} & \mathbf{I}^s \\ \mathbf{I} & \mathbf{0} \end{array} \right] = \left(\begin{array}{cccc|ccc} 0 & -1 & 3 & 0 & 1 & 0 & 0 \\ 2 & -4 & 1 & 5 & 0 & 1 & 0 \\ -4 & 5 & 7 & -10 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array} \right) \rightarrow \left(\begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & -2 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -3 & 2 & 1 \\ \hline 1 & 0 & \frac{11}{2} & -\frac{5}{2} & 0 & 0 & 0 \\ 0 & 1 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array} \right),$$

$$\text{于是 } \mathbf{P} = \begin{pmatrix} -2 & \frac{1}{2} & 0 \\ -1 & 0 & 0 \\ -3 & 2 & 1 \end{pmatrix}, \mathbf{Q} = \begin{pmatrix} 1 & 0 & \frac{11}{2} & -\frac{5}{2} \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

所以 \mathbf{A} 的减号广义逆为

$$\mathbf{G} = \mathbf{Q} \begin{pmatrix} \mathbf{I}_2 & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{pmatrix} \mathbf{P},$$

其中

$$\mathbf{U} \in \mathbf{C}^{2 \times 1}, \mathbf{V} \in \mathbf{C}^{2 \times 2}, \mathbf{W} \in \mathbf{C}^{2 \times 1}.$$

由定理 4.5 知道,任意矩阵都有减号广义逆,进一步有下列定理.

定理 4.6 设 $\mathbf{A} \in \mathbf{C}^{m \times n}$, 则 \mathbf{A} 的减号广义逆惟一的必要充分条件是 $m=n$, 且 \mathbf{A}^{-1} 存在.

证明 先证充分性. 当 $m=n$ 且 \mathbf{A}^{-1} 存在时, $\mathbf{A}^{-1} \in \mathbf{A}\{1\}$. $\forall \mathbf{G} \in \mathbf{A}\{1\}$, 因为 $\mathbf{AGA} = \mathbf{A}$, 用 \mathbf{A}^{-1} 左右乘等式两边得 $\mathbf{G} = \mathbf{A}^{-1}$, 故减号广义逆惟一.

再证必要性. 设 $\text{rank}(\mathbf{A}) = r$, 当 $r=0$ 时, $\mathbf{A}=\mathbf{0}$, 此时 $\mathbf{A}\{1\} = \mathbf{C}^{n \times m}$ 与假设不符, 所以 $r>0$.

当 $m \neq n$ 时, 由定理 4.5 知 $\mathbf{A}\{1\} = \left\{ \mathbf{Q} \begin{pmatrix} \mathbf{I}_r & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{pmatrix} \mathbf{P} \right\}$ 不是单元素集.

若 \mathbf{A}^{-1} 不存在, 则 $r < n$. 此时 $\mathbf{A}\{1\} = \left\{ \mathbf{Q} \begin{pmatrix} \mathbf{I}_r & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{pmatrix} \mathbf{P} \right\}$ 也不是单元素集.

综上 \mathbf{A}^{-1} 存在. □

定理 4.7 设 $\mathbf{A} \in \mathbf{C}^{m \times n}$, $\lambda \in \mathbf{C}$, 则 \mathbf{A}^- 满足以下性质:

(1) $\text{rank}(A) \leq \text{rank}(A^-)$;

(2) AA^- 与 A^-A 都是幂等矩阵, 且 $\text{rank}(A) = \text{rank}(AA^-) = \text{rank}(A^-A)$;

(3) $R(AA^-) = R(A)$, $N(A^-A) = N(A)$.

证明 (1) $\text{rank}(A) = \text{rank}(AA^-A) \leq \text{rank}(A^-)$.

(2) 因为 $AA^-A = A$, 所以 $(AA^-)^2 = AA^-$, $(A^-A)^2 = A^-A$, 即 AA^- 与 A^-A 都是幂等矩阵.

而 $\text{rank}(A) = \text{rank}(AA^-A) \leq \text{rank}(AA^-) \leq \text{rank} A$, 所以 $\text{rank}(A) = \text{rank}(AA^-)$.

同理可证 $\text{rank}(A) = \text{rank}(A^-A)$.

(3) 显然 $R(AA^-) \subseteq R(A)$, $N(A) \subseteq N(A^-A)$. 下证 $R(A) \subseteq R(AA^-)$, $N(A^-A) \subseteq N(A)$.

对任意 $y \in R(A)$, 存在 $x \in \mathbf{C}^n$, 使得 $y = Ax$, 于是

$$y = Ax = (AA^-A)x = AA^-(Ax) \in R(AA^-),$$

故 $R(A) \subseteq R(AA^-)$. □

对任意 $x \in N(A^-A)$, 有 $A^-Ax = \mathbf{0}$, 于是

$$Ax = (AA^-A)x = A(A^-Ax) = A\mathbf{0} = \mathbf{0},$$

故 $N(A^-A) \subseteq N(A)$. □

定理 4.8 设 $A \in \mathbf{C}^{m \times n}$, $A^- \in A\{1\}$, 则当方程组 $Ax = b$ 有解时, 其通解可表示为

$$x = A^-b + (I_n - A^-A)z,$$

这里 z 是任意的 n 维列向量.

* **证明** 先看齐次线性方程组 $Ax = \mathbf{0}$, 由

$$A = AA^-A,$$

对任意 $z \in \mathbf{C}^n$, 有

$$A(I_n - A^-A)z = \mathbf{0},$$

即对任意 $z \in \mathbf{C}^n$, $(I_n - A^-A)z$ 一定是 $Ax = \mathbf{0}$ 的解.

要证明 $Ax = \mathbf{0}$ 的解均可表示成这种形式, 只须证明 $I_n - A^-A$ 的秩为 $n - r$, 其中 r 是 A 的秩. 由定理 4.5

$$A^- = Q \begin{pmatrix} I_r & U \\ V & W \end{pmatrix} P,$$

$$I_n - A^-A = I_n - Q \begin{pmatrix} I_r & U \\ V & W \end{pmatrix} PA,$$

而

$$\text{rank}(I_n - A^-A) = \text{秩}(Q^{-1}(I_n - A^-A)Q)$$

$$= \text{rank}\left(I_n - \left[\begin{array}{cc|cc} I_r & U & I_r & \mathbf{0} \\ V & W & \mathbf{0} & \mathbf{0} \end{array} \right] \right)$$

$$= \text{rank}\left(\begin{array}{cc} \mathbf{0} & \mathbf{0} \\ -V & I_{n-r} \end{array}\right) = n - r.$$

对非齐次线性方程组 $Ax=b$, 若它有解, 则由减号广义逆的定义知 A^-b 必为解, 从而通解为

$$x = A^-b + (I_n - A^-A)z, \quad z \in \mathbf{C}^n.$$

□

二、Moore-Penrose 广义逆(加号广义逆)

定义 4.3 设 $A \in \mathbf{C}^{m \times n}$, 若存在矩阵 $G \in \mathbf{C}^{n \times m}$, 使得

- (1) $AGA=A$;
- (2) $GAG=G$;
- (3) $(AG)^H=AG$;
- (4) $(GA)^H=GA$,

则称 G 为 A 的 Moore-Penrose 广义逆或加号广义逆, 简称为 A 的 M-P 逆. A 的任意 M-P 逆记为 A^+ .

由定义直接验算

若 $A=\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$, 则 $\begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \end{pmatrix}$ 是其 M-P 逆.

若 $B=\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, 则 $\begin{pmatrix} \frac{1}{2}, \frac{1}{2} \end{pmatrix}$ 是其 M-P 逆.

若 $C=\begin{pmatrix} D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$, D 为可逆矩阵, 则 $\begin{pmatrix} D^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$ 是其 M-P 逆. 特别地, 若 C 可逆, 则 $C^{-1}=C^+$.

对于任意矩阵 $A \in \mathbf{C}^{m \times n}$, 是否存在 M-P 逆 A^+ ? 若存在, 其 M-P 逆是否惟一? 怎样求 M-P 逆 A^+ ? 这是需要首先回答的问题.

定理 4.9 若矩阵 $A \in \mathbf{C}^{m \times n}$ 存在 M-P 广义逆, 则 A 的 M-P 逆是惟一的.

证明 设 G_1, G_2 都是 A 的 M-P 逆, 则 G_1 与 G_2 均满足 M-P 逆的定义中的四个条件, 于是

$$\begin{aligned} G_1 &= (G_1 A)G_1 = (G_1 A)^H G_1 = A^H G_1^H G_1 \\ &= (A^H G_2^H A^H)G_1^H G_1 = (G_2 A)^H (G_1 A)^H G_1 \\ &= G_2 AG_1 AG_1 = G_2 AG_1, \\ G_2 &= G_2 (AG_2) = G_2 (AG_2)^H = G_2 G_2^H A^H \\ &= G_2 G_2^H (A^H G_1^H A^H) = G_2 (AG_2)^H (AG_1)^H \\ &= G_2 AG_2 AG_1 = G_2 AG_1, \end{aligned}$$

故

$$G_1 = G_2.$$

□

下面证明, 对任意 $A \in \mathbf{C}^{m \times n}$, 都有 M-P 逆存在, 并提供实际计算 A^+ 的一个有效方

法.

定理 4.10 任意矩阵 $A \in \mathbf{C}^{m \times n}$ 都存在 M-P 广义逆 A^+ . 设 $\text{rank}(A) = r$, A 的一个满秩分解为

$$A = BC, B \in \mathbf{C}^{m \times r}, C \in \mathbf{C}^{r \times n}, \quad \text{rank}(B) = \text{rank}(C) = r,$$

则

$$A^+ = C^H (CC^H)^{-1} (B^H B)^{-1} B^H.$$

证明 因 $\text{rank}(A) = \text{rank}(B) = \text{rank}(C) = r$, 由定理 4.1, 定理 4.2 知 $B^H B$ 与 CC^H 都可逆. 令 $G = C^H (CC^H)^{-1} (B^H B)^{-1} B^H$, 直接验证知 G 满足 M-P 广义逆定义中的四个条件, 故 G 是 A 的 M-P 广义逆. 因 A 的 M-P 广义逆惟一, 故

$$A^+ = G = C^H (CC^H)^{-1} (B^H B)^{-1} B^H. \quad \square$$

例 3 求矩阵 $A = \begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{pmatrix}$ 的 M-P 逆 A^+ .

解 首先求得 A 的满秩分解为

$$A = BC = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix},$$

故

$$\begin{aligned} A^+ &= C^H (CC^H)^{-1} (B^H B)^{-1} B^H \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 2 & -4 \\ -4 & 20 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 4 \end{pmatrix} \\ &= \frac{1}{18} \begin{pmatrix} 5 & 2 & -1 \\ 1 & 1 & 1 \\ -4 & -1 & 2 \\ 6 & 3 & 0 \end{pmatrix}. \end{aligned}$$

除了利用 A 的满秩分解计算 A^+ 之外, 还可利用 A 的奇异值分解结果, 很方便地计算 A 的 M-P 广义逆 A^+ .

定理 4.11 设矩阵 $A \in \mathbf{C}^{m \times n}$, $\text{rank}(A) = r$, A 的奇异值分解为

$$A = U \begin{pmatrix} \Delta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} V^H,$$

则

$$A^+ = V \begin{pmatrix} \Delta^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} U^H.$$

其中 Δ 为对角线由 A 的正奇异值所构成的对角矩阵, $\Delta \in \mathbf{C}^{r \times r}$.

证明 直接验证上式满足 M-P 逆定义中的四个条件.

例 4 求矩阵 $A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ 的 M-P 逆 A^+ .

解 首先求得 A 的奇异值分解为

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

故 $A^+ = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$

下面我们来讨论 M-P 逆的一些性质.首先 M-P 逆 A^+ 与通常的逆矩阵 A^{-1} 有如下区别:

(1) $(AB)^+ = B^+ A^+$ 不成立.例如取

$$A = (1 \quad 1), \quad B = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

则 $A^+ = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad B^+ = (1 \quad 0),$

$$(AB)^+ = (1)^+ = (1), \quad B^+ A^+ = \left(\frac{1}{2} \right).$$

(2) $(A^+)^k = (A^k)^+$ (k 为正整数) 不成立.例如取

$$A = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix},$$

则 $A^2 = A, (A^2)^+ = A^+ = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}, (A^+)^2 = \frac{1}{4} \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}.$

但是, M-P 逆也与通常的逆有一些相同或相似的性质.

定理 4.12 设 $A \in \mathbf{C}^{m \times n}$, $\lambda \in \mathbf{C}^n$, 则 A^+ 满足以下性质:

(1) $(A^+)^+ = A;$

(2) $(A^+)^H = (A^H)^+;$

(3) $(\lambda A)^+ = \lambda^+ A^+$, 其中 $\lambda^+ = \begin{cases} \frac{1}{\lambda}, & \lambda \neq 0, \\ 0, & \lambda = 0. \end{cases}$

(4) 若 $A \in \mathbf{C}^{m \times n}$ 是列满秩的, 则

$$\mathbf{A}^+ = (\mathbf{A}^\text{H} \mathbf{A})^{-1} \mathbf{A}^\text{H};$$

若 \mathbf{A} 是行满秩的, 则

$$\mathbf{A}^+ = \mathbf{A}^\text{H} (\mathbf{A} \mathbf{A}^\text{H})^{-1}.$$

(5) 若 \mathbf{A} 有满秩分解 $\mathbf{A} = \mathbf{B}\mathbf{C}$, 则 $\mathbf{A}^+ = \mathbf{C}^+ \mathbf{B}^+$.

证明 (1),(2),(3)由定义即得.

(4) 若 \mathbf{A} 是列满秩的, 则 \mathbf{A} 的满秩分解为 $\mathbf{A} = \mathbf{A}\mathbf{I}_n$, 即取 $\mathbf{B} = \mathbf{A}$, $\mathbf{C} = \mathbf{I}_n$, 则

$$\mathbf{A}^+ = (\mathbf{A}^\text{H} \mathbf{A})^{-1} \mathbf{A}^\text{H}.$$

若 \mathbf{A} 是行满秩的, 同理可证

$$\mathbf{A}^+ = \mathbf{A}^\text{H} (\mathbf{A} \mathbf{A}^\text{H})^{-1}.$$

(5) 若 \mathbf{A} 有满秩分解 $\mathbf{A} = \mathbf{B}\mathbf{C}$, 则 \mathbf{B} 列满秩, \mathbf{C} 行满秩, 由(4)知

$$\mathbf{B}^+ = (\mathbf{B}^\text{H} \mathbf{B})^{-1} \mathbf{B}^\text{H}, \quad \mathbf{C}^+ = \mathbf{C}^\text{H} (\mathbf{C} \mathbf{C}^\text{H})^{-1},$$

故

$$\mathbf{A}^+ = \mathbf{C}^+ \mathbf{B}^+.$$

□

例 5 设 $\mathbf{A} \in \mathbf{C}^{m \times n}$, 证明

$$(1) \operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}^+);$$

$$(2) \operatorname{rank}(\mathbf{A}^+ \mathbf{A}) = \operatorname{rank}(\mathbf{A} \mathbf{A}^+) = \operatorname{rank}(\mathbf{A}).$$

证明 (1) 由 $\mathbf{A} = \mathbf{A}\mathbf{A}^+ \mathbf{A}$ 知, $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}\mathbf{A}^+ \mathbf{A}) \leqslant \operatorname{rank}(\mathbf{A}^+)$; 由 $\mathbf{A}^+ = \mathbf{A}^+ \mathbf{A} \mathbf{A}^+$ 知, $\operatorname{rank}(\mathbf{A}^+) = \operatorname{rank}(\mathbf{A}^+ \mathbf{A} \mathbf{A}^+) \leqslant \operatorname{rank}(\mathbf{A})$, 故

$$\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}^+).$$

(2) 一方面, $\operatorname{rank}(\mathbf{A}^+ \mathbf{A}) \leqslant \operatorname{rank}(\mathbf{A})$; 另一方面

$$\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A} \mathbf{A}^+ \mathbf{A}) \leqslant \operatorname{rank}(\mathbf{A}^+ \mathbf{A}),$$

故

$$\operatorname{rank}(\mathbf{A}^+ \mathbf{A}) = \operatorname{rank}(\mathbf{A}).$$

同理

$$\operatorname{rank}(\mathbf{A} \mathbf{A}^+) = \operatorname{rank}(\mathbf{A}).$$

4.3 投影变换

投影变换是研究广义逆矩阵和最小二乘问题的重要工具, 本节将较系统地讨论投影变换和正交投影变换.

一、投影变换与投影矩阵

定义 4.4 设 $\mathbf{C}^n = L \oplus M$, $x = y + z$, $y \in L$, $z \in M$.

如果线性变换 $\sigma: \mathbf{C}^n \rightarrow \mathbf{C}^n$ 满足 $\sigma(x) = y$, 则称 σ 是从 \mathbf{C}^n 沿子空间 M 到子空间 L 上的投影变换, 投影变换在 \mathbf{C}^n 空间的一组基下的矩阵称为投影矩阵.

投影变换 σ 把 \mathbf{C}^n 映射成子空间 L , 故子空间 L 称为投影子空间, 显然, 它就是 σ 的像空间 $R(\sigma)$, 子空间 M 是投影变换的核空间 $N(\sigma)$, 这时 \mathbf{C}^n 空间的直和分解为

$$\mathbf{C}^n = R(\sigma) \oplus N(\sigma).$$

定理 4.13 \mathbf{C}^n 空间上的线性变换 σ 是投影变换的充分必要条件是 σ 是幂等变换, 即 $\sigma^2 = \sigma$.

证明 必要性: 设 σ 是 \mathbf{C}^n 空间沿 M 到 L 上的投影变换, 则 $\forall \mathbf{x} \in \mathbf{C}^n$, 存在 $\mathbf{y} \in L, \mathbf{z} \in M$, 使得

$$\mathbf{x} = \mathbf{y} + \mathbf{z}, \quad \sigma(\mathbf{x}) = \mathbf{y},$$

于是

$$\sigma^2(\mathbf{x}) = \sigma[\sigma(\mathbf{x})] = \sigma(\mathbf{y}) = \mathbf{y} = \sigma(\mathbf{x}),$$

故

$$\sigma^2 = \sigma.$$

充分性: 首先证明

$$\mathbf{C}^n = R(\sigma) + N(\sigma),$$

$\forall \mathbf{x} \in \mathbf{C}^n$, 有 $\mathbf{x} = \sigma(\mathbf{x}) + [\mathbf{x} - \sigma(\mathbf{x})]$, 注意到 $\sigma^2 = \sigma$, 有 $\sigma[\mathbf{x} - \sigma(\mathbf{x})] = \sigma(\mathbf{x}) - \sigma^2(\mathbf{x}) = \mathbf{0}$, 于是 $\sigma(\mathbf{x}) \in R(\sigma)$, $\mathbf{x} - \sigma(\mathbf{x}) \in N(\sigma)$, 故

$$\mathbf{C}^n = R(\sigma) + N(\sigma).$$

其次, $\forall \mathbf{x} \in R(\sigma) \cap N(\sigma)$, 因 $\mathbf{x} \in R(\sigma)$, 故存在 $\mathbf{y} \in \mathbf{C}^n$, 使得 $\mathbf{x} = \sigma(\mathbf{y})$. 又因 $\mathbf{x} \in N(\sigma)$, 故 $\sigma(\mathbf{x}) = \mathbf{0}$, 于是

$$\mathbf{0} = \sigma(\mathbf{x}) = \sigma^2(\mathbf{y}) = \sigma(\mathbf{y}) = \mathbf{x},$$

故

$$R(\sigma) \cap N(\sigma) = \{\mathbf{0}\}.$$

于是

$$\mathbf{C}^n = R(\sigma) \oplus N(\sigma).$$

此时, $\forall \mathbf{x} \in \mathbf{C}^n$, 存在 $\mathbf{y} \in R(\sigma), \mathbf{z} \in N(\sigma)$, 使得 $\mathbf{x} = \mathbf{y} + \mathbf{z}$, $\mathbf{y} = \sigma(\mathbf{x})$, $\sigma(\mathbf{z}) = \mathbf{0}$, 故 $\sigma(\mathbf{x}) = \sigma(\mathbf{y}) = \sigma^2(\mathbf{x}) = \sigma(\mathbf{x}) = \mathbf{y}$. 这便证明了 σ 是 \mathbf{C}^n 空间沿 $N(\sigma)$ 到 $R(\sigma)$ 上的投影变换. \square

推论 \mathbf{C}^n 空间上的线性变换 σ 是投影变换的充分必要条件是 σ 关于某组基下的矩阵 A 为幂等矩阵, 即 $A^2 = A$.

因此, 投影矩阵亦即幂等矩阵.

投影矩阵 A 可按如下方法求得.

假定 $\dim L = r$, 则 $\dim M = n - r$. 在子空间 L 和 M 中分别取定基底

$$\{y_1, y_2, \dots, y_r\}, \quad \{z_1, z_2, \dots, z_{n-r}\},$$

于是 $\{y_1, \dots, y_r, z_1, \dots, z_{n-r}\}$ 便构成 \mathbf{C}^n 的基底. 根据投影矩阵的性质有

$$Ay_i = y_i \quad (i = 1, 2, \dots, r),$$

$$Az_j = 0 \quad (j = 1, 2, \dots, n - r).$$

作分块矩阵

$$\mathbf{B} = (y_1 \ y_2 \ \cdots \ y_r), \quad \mathbf{C} = (z_1 \ z_2 \ \cdots \ z_{n-r}),$$

于是

$$\mathbf{A}(\mathbf{B} \mid \mathbf{C}) = (\mathbf{B}, \mathbf{0}).$$

由于 $(\mathbf{B} \mid \mathbf{C})$ 是 n 阶可逆阵, 因此投影矩阵 A 为

$$\mathbf{A} = (\mathbf{B} \mid \mathbf{0})(\mathbf{B} \mid \mathbf{C})^{-1}.$$

例 6 设 L 是由向量 $(1 \ 0)^T$ 所生成的子空间, M 是由向量 $(1 \ -1)^T$ 所生成的子空间, 则 \mathbf{R}^2 沿子空间 M 到子空间 L 上的投影矩阵为

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

二、正交投影变换与正交投影矩阵

投影变换的一个子类——正交投影变换,具有更为良好的性质.

定义 4.5 设 σ 是 \mathbf{C}^n 空间的投影变换,

$$\mathbf{C}^n = R(\sigma) \oplus N(\sigma),$$

如果 $R(\sigma)$ 的正交补子空间 $R(\sigma)^\perp = N(\sigma)$, 则称 σ 是 \mathbf{C}^n 空间的正交投影变换. 正交投影变换在 \mathbf{C}^n 空间的一组基下的矩阵称为正交投影矩阵.

定理 4.14 \mathbf{C}^n 空间上的线性变换 σ 是正交投影变换的充分必要条件是 σ 关于某组基下的矩阵 A 为幂等的 Hermite 矩阵, 即 $A^2 = A, A^H = A$.

证明 必要性: 设 A 是线性变换 σ 在某组基下的矩阵, 由定理 4.13 只需证 $A^H = A$.

由 $\left\{ \begin{array}{l} \mathbf{C}^n = R(\sigma) \oplus N(\sigma), \\ R(\sigma)^\perp = N(\sigma), \end{array} \right.$

有 $\left\{ \begin{array}{l} \mathbf{C}^n = R(A) \oplus N(A), \\ R(A)^\perp = N(A), \end{array} \right.$

又令 $x \in R(A), y \in N(A^H)$, 由 $A^2 = A$ 有 $x = Ax$, 于是

$$(x, y) = (Ax, y) = (x, A^H y) = (x, \mathbf{0}) = 0,$$

所以 $N(A^H) \subseteq R(A)^\perp$.

另一方面, 对任意 $y \in R(A)^\perp$, 有

$$(A^H y, A^H y) = (y, AA^H y) = 0.$$

从而 $A^H y = \mathbf{0}$, 即 $y \in N(A^H)$, 故 $R(A)^\perp = N(A^H)$. 于是

$$\left\{ \begin{array}{l} \mathbf{C}^n = R(A) \oplus N(A^H), \\ R(A)^\perp = N(A^H). \end{array} \right.$$

由正交补的惟一性推得 $N(A) = N(A^H)$, 同理, 由 $(A^H)^2 = A^H$, 推得

$$\left\{ \begin{array}{l} \mathbf{C}^n = R(A^H) \oplus N(A), \\ R(A^H)^\perp = N(A), \end{array} \right.$$

故 $R(A) = R(A^H)$.

$\forall x \in \mathbf{C}^n, x = y + z, y \in R(A) = R(A^H), z \in N(A) = N(A^H)$, 由

$$Ax = Ay + Az = Ay = y,$$

$$A^H x = A^H y + A^H z = A^H y = y,$$

证得 $A^H = A$.

充分性: 因 $A^2 = A$, 所以 σ 是 \mathbf{C}^n 空间的投影变换, 且

$$\mathbf{C}^n = R(A) \oplus N(A),$$

$\forall x \in R(A), y \in N(A)$, 有

$(\mathbf{x}, \mathbf{y}) = (\mathbf{Ax}, \mathbf{y}) = (\mathbf{A}^H \mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{Ay}) = (\mathbf{x}, \mathbf{0}) = 0$,
证得 $N(\mathbf{A}) \subseteq R(\mathbf{A})^\perp$. 又 $\forall \mathbf{x} \in R(\mathbf{A})^\perp$,

$$(\mathbf{Ax}, \mathbf{Ax}) = (\mathbf{x}, \mathbf{A}^H \mathbf{Ax}) = (\mathbf{x}, \mathbf{Ax}) = 0,$$

所以 $\mathbf{Ax} = \mathbf{0}$, 即 $\mathbf{x} \in N(\mathbf{A})$, 故 $R(\mathbf{A})^\perp = N(\mathbf{A})$. \square

作为正交投影矩阵的例子, 我们考察矩阵 \mathbf{AA}^+ 和 $\mathbf{A}^+ \mathbf{A}$.

定理 4.15 设 $\mathbf{A} \in \mathbf{C}^{m \times n}$, 则

$$(1) (\mathbf{A}^+ \mathbf{A})^2 = \mathbf{A}^+ \mathbf{A}, (\mathbf{A}^+ \mathbf{A})^H = \mathbf{A}^+ \mathbf{A};$$

$$(2) \begin{cases} \mathbf{C}^n = R(\mathbf{A}^+) \oplus N(\mathbf{A}), \\ R(\mathbf{A}^+)^\perp = N(\mathbf{A}); \end{cases}$$

$$(3) (\mathbf{AA}^+)^2 = \mathbf{AA}^+, (\mathbf{AA}^+)^H = \mathbf{AA}^+;$$

$$(4) \begin{cases} \mathbf{C}^m = R(\mathbf{A}) \oplus N(\mathbf{A}^+), \\ R(\mathbf{A})^\perp = N(\mathbf{A}^+). \end{cases}$$

证明 由 \mathbf{A}^+ 的定义可直接证明(1)和(3)成立. 下面证明(2)成立.

由定理 4.14 及(1)知, $\mathbf{A}^+ \mathbf{A}$ 是正交投影矩阵, 故

$$\begin{cases} \mathbf{C}^n = R(\mathbf{A}^+ \mathbf{A}) \oplus N(\mathbf{A}^+ \mathbf{A}), \\ R(\mathbf{A}^+ \mathbf{A})^\perp = N(\mathbf{A}^+ \mathbf{A}). \end{cases}$$

下证 $R(\mathbf{A}^+ \mathbf{A}) = R(\mathbf{A}^+)$, $N(\mathbf{A}^+ \mathbf{A}) = N(\mathbf{A})$.

显然, $R(\mathbf{A}^+ \mathbf{A}) \subseteq R(\mathbf{A}^+)$, $N(\mathbf{A}) \subseteq N(\mathbf{A}^+ \mathbf{A})$, 只需证 $R(\mathbf{A}^+) \subseteq R(\mathbf{A}^+ \mathbf{A})$, $N(\mathbf{A}^+ \mathbf{A}) \subseteq N(\mathbf{A})$.

$\forall \mathbf{y} \in R(\mathbf{A}^+)$, 存在 \mathbf{x} 使 $\mathbf{y} = \mathbf{A}^+ \mathbf{x} = \mathbf{A}^+ \mathbf{A}(\mathbf{A}^+ \mathbf{x})$, 故 $R(\mathbf{A}^+) \subseteq R(\mathbf{A}^+ \mathbf{A})$.

$\forall \mathbf{x} \in N(\mathbf{A}^+ \mathbf{A})$, $\mathbf{A}^+ \mathbf{Ax} = \mathbf{0}$, $\mathbf{Ax} = \mathbf{AA}^+ \mathbf{Ax} = \mathbf{0}$, 故 $N(\mathbf{A}^+ \mathbf{A}) \subseteq N(\mathbf{A})$.

同理可证(4)成立. \square

定理 4.16 设 W 是 \mathbf{C}^n 的子空间, $\mathbf{x}_0 \in \mathbf{C}^n$, $\mathbf{x}_0 \in W$. 如果 σ 是 \mathbf{C}^n 空间向 W 的正交投影变换, 则 $\sigma(\mathbf{x}_0)$ 是 W 中离 \mathbf{x}_0 最近的向量, 即对欧几里得范数 $\|\cdot\|$, 都有

$$\|\sigma(\mathbf{x}_0) - \mathbf{x}_0\| \leq \|\mathbf{y} - \mathbf{x}_0\|, \forall \mathbf{y} \in W.$$

证明 因 σ 是 \mathbf{C}^n 空间向 W 的正交投影变换, 所以

$$\mathbf{C}^n = W \oplus W^\perp, \quad W = R(\sigma), \quad W^\perp = N(\sigma),$$

$\forall \mathbf{y} \in W$, 由于 $[\mathbf{y} - \sigma(\mathbf{x}_0)] \in W$, $[\sigma(\mathbf{x}_0) - \mathbf{x}_0] \in W^\perp$, 因此

$$\begin{aligned} \|\mathbf{y} - \mathbf{x}_0\|^2 &= \|[\mathbf{y} - \sigma(\mathbf{x}_0)] + [\sigma(\mathbf{x}_0) - \mathbf{x}_0]\|^2 \\ &= \|\mathbf{y} - \sigma(\mathbf{x}_0)\|^2 + \|\sigma(\mathbf{x}_0) - \mathbf{x}_0\|^2 \\ &\geq \|\sigma(\mathbf{x}_0) - \mathbf{x}_0\|^2, \end{aligned}$$

故

$$\|\sigma(\mathbf{x}_0) - \mathbf{x}_0\| \leq \|\mathbf{y} - \mathbf{x}_0\|, \forall \mathbf{y} \in W. \quad \square$$

如果把 \mathbf{x}_0 视为 \mathbf{C}^n 中的“点”, 则 $\sigma(\mathbf{x}_0)$ 可以看作是 \mathbf{x}_0 在 W 中由正交投影得到的“垂足”.

正交投影矩阵 \mathbf{A} 可按如下方法求得:

假定 $\dim L = r$, 则 $\dim L^\perp = n - r$. 在子空间 L 和 L^\perp 中分别取定基底

$$\{y_1, y_2, \dots, y_r\}, \quad \{z_1, z_2, \dots, z_{n-r}\},$$

作分块矩阵

$$\mathbf{B} = \begin{pmatrix} y_1 & y_2 & \cdots & y_r \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} z_1 & z_2 & \cdots & z_{n-r} \end{pmatrix},$$

则 $\mathbf{B}^H \mathbf{C} = \mathbf{0}$. 由投影矩阵的求法有

$$\begin{aligned} \mathbf{A} &= (\mathbf{B} | \mathbf{0})(\mathbf{B} | \mathbf{C})^{-1} = (\mathbf{B} | \mathbf{0})((\mathbf{B} | \mathbf{C})^H(\mathbf{B} | \mathbf{C}))^{-1}(\mathbf{B} | \mathbf{C})^H \\ &= (\mathbf{B} | \mathbf{0}) \begin{pmatrix} \mathbf{B}^H \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^H \mathbf{C} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{B}^H \\ \mathbf{C}^H \end{pmatrix} = (\mathbf{B} | \mathbf{0}) \begin{pmatrix} (\mathbf{B}^H \mathbf{B})^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{C}^H \mathbf{C})^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{B}^H \\ \mathbf{C}^H \end{pmatrix} \\ &= \mathbf{B}(\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}. \end{aligned}$$

例 7 考虑 \mathbf{R}^3 中由向量 $\alpha = (1 \ 2 \ 0)^T$ 和 $\beta = (0 \ 1 \ 1)^T$ 所生成的子空间 L , 求正交投影矩阵 \mathbf{A} 和向量 $x = (1 \ 2 \ 3)^T$ 沿 L^\perp 到 L 的投影.

解 因为

$$\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{B}^H \mathbf{B} = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix},$$

$$(\mathbf{B}^H \mathbf{B})^{-1} = \frac{1}{6} \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix},$$

所以 $\mathbf{A} = \mathbf{B}(\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H = \frac{1}{6} \begin{pmatrix} 2 & 2 & -2 \\ 2 & 5 & 1 \\ -2 & 1 & 5 \end{pmatrix}.$

x 沿 L^\perp 到 L 的投影为

$$\mathbf{Ax} = \begin{pmatrix} 0 & \frac{5}{2} & \frac{5}{2} \end{pmatrix}^T.$$

4.4 最佳的最小二乘解

设 $A \in \mathbf{C}^{m \times n}$, $b \in \mathbf{C}^m$, 线性方程组 $Ax = b$ 有解当且仅当 $b \in R(A)$. 如果 $b \in R(A)$, 称方程组 $Ax = b$ 是相容的, 否则, 称其是不相容的. 对不相容的方程组, 希望求其近似解 u , 使得对欧几里得范数 $\|\cdot\|_2$, 误差 $\|Au - b\|_2$ 达到极小.

定义 4.6 设 $A \in \mathbf{C}^{m \times n}$, $b \in \mathbf{C}^m$, 如果存在 $u \in \mathbf{C}^n$, 使得

$$\|Au - b\|_2 \leq \|Ax - b\|_2, \forall x \in \mathbf{C}^n,$$

则称 u 是方程组 $Ax = b$ 的一个最小二乘解.

设 x_0 是 $Ax = b$ 的最小二乘解, 如果对于 $Ax = b$ 的每一个最小二乘解 u , 都有

$$\|x_0\|_2 \leq \|u\|_2,$$

则称 x_0 是最佳的最小二乘解(或称按范数最小的最小二乘解, 或称最佳逼近解).

定理 4.17 设 $A \in \mathbf{C}^{m \times n}$, $b \in \mathbf{C}^m$, 则 $x_0 = A^+ b$ 是线性方程组 $Ax = b$ 的最佳的最小二乘解.

证明 由定理 4.15 知道, AA^+ 是 \mathbf{C}^m 空间向 $R(A)$ 上的一个正交投影变换所对应的矩阵, 再由定理 4.16, 有

$$\|A(A^+ b) - b\| \leq \|Ax - b\|, x \in \mathbf{C}^n,$$

故 $x_0 = A^+ b$ 是 $Ax = b$ 的最小二乘解.

又由定理 4.15, 有

$$\begin{cases} \mathbf{C}^m = R(A) \oplus N(A^+), \\ R(A)^\perp = N(A^+), \end{cases}$$

每个 $b \in \mathbf{C}^m$ 可惟一地分解为

$$b = b_1 + b_2, \quad b_1 \in R(A), \quad b_2 \in N(A^+),$$

$\forall x \in \mathbf{C}^n$, 有

$$\begin{aligned} \|Ax - b\|^2 &= \|(Ax - b_1) + (-b_2)\|^2 \\ &= \|Ax - b_1\|^2 + \|b_2\|^2, \end{aligned}$$

因此, u 是 $Ax = b$ 的最小二乘解当且仅当 u 是 $Ax = b_1$ 的解.

设 u 是 $Ax = b$ 的任一个最小二乘解, 因 x_0 也是 $Ax = b$ 的最小二乘解, 故 $A(x_0 - u) = Ax_0 - Au = b_1 - b_1 = \mathbf{0}$, 从而 $(x_0 - u) \in N(A)$, 又因 $x_0 = A^+ b \in R(A^+)$, $R(A^+)^\perp = N(A)$, 故 x_0 与 $(x_0 - u)$ 正交. 因而

$$\|u\|^2 = \|x_0 + (u - x_0)\|^2 = \|x_0\|^2 + \|u - x_0\|^2 \geq \|x_0\|^2,$$

证得 $x_0 = A^+ b$ 是 $Ax = b$ 的最佳的最小二乘解. \square

推论 设矩阵方程 $AX = B$, 其中 $A \in \mathbf{C}^{m \times n}$, $B \in \mathbf{C}^{m \times k}$, 则 $X_0 = A^+ B$ 是 $AX = B$ 的最佳的最小二乘解.

例 8 求不相容的线性方程组

$$\begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 2 \end{pmatrix}$$

最佳的最小二乘解.

解 由 4.2 例 3 知系数矩阵 A 的 M-P 逆为

$$A^+ = \frac{1}{18} \begin{pmatrix} 5 & 2 & -1 \\ 1 & 1 & 1 \\ -4 & -1 & 2 \\ 6 & 3 & 0 \end{pmatrix},$$

于是方程组的最佳的最小二乘解为

$$x_0 = A^+ b = \frac{1}{18} (20 \quad 7 \quad -13 \quad 27)^T.$$

应用最佳最小二乘解, 可以解决在实际中常常要求寻找经验公式的问题. 设由观察

得到关于物理量 x 和 y 的一组数据为

$$(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m).$$

人们希望通过这些数据, 找出函数 $y=f(x)$, 使得它能最好的反映 x 和 y 之间的依赖关系.

所谓最好的反映, 是指以 $\delta_i = f(x_i) - y_i$ ($i=1, 2, \dots, m$) 作为分量的误差向量 $\delta = (\delta_1 \ \ \delta_2 \ \ \dots \ \ \delta_m)^T$ 按欧几里得范数 $\|\delta\|_2$ 最小, 称如此得到的函数 $y=f(x)$ 的图形是拟合数据 (x_i, y_i) ($i=1, 2, \dots, m$) 的最佳拟合曲线.

例 9 设有一组实验数据:

$$(1, 2), (2, 3), (3, 5), (4, 7).$$

从数据点的走向看接近一条直线, 实验者希望使用直线 $y=\beta_0 + \beta_1 x$ 拟合数据点, 求最佳拟合直线.

解 把实验数据代入 $y=\beta_0 + \beta_1 x$, 得到

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 5 \\ 7 \end{pmatrix},$$

因系数矩阵 A 是列满秩的, $A^+ = (A^H A)^{-1} A^H$, 求得

$$A^+ = \frac{1}{20} \begin{pmatrix} 20 & 10 & 0 & -10 \\ -6 & -2 & 2 & 6 \end{pmatrix},$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = A^+ b = \begin{pmatrix} 0 \\ 1.7 \end{pmatrix}.$$

所以最佳拟合直线为 $y=1.7x$, 其误差为

$$\|\delta\|_2 = \|Ab - b\|_2 = \sqrt{0.3}.$$

例 10 设一个质点运动的轨道是椭圆, 观察到的位置坐标是

$$(1, 1), (0, 2), (-1, 1), (-1, -2),$$

点在同一平面上, 求拟合观察点的最佳标准椭圆方程.

解 把观察点的坐标代入椭圆方程 $\beta_0 x^2 + \beta_1 y^2 = 1$, 得到

$$\begin{pmatrix} 1 & 1 \\ 0 & 4 \\ 1 & 1 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

因系数矩阵 A 是列满秩的, $A^+ = (A^H A)^{-1} A^H$, 求得

$$A^+ = \frac{1}{66} \begin{pmatrix} 28 & -24 & 28 & 10 \\ -3 & 12 & -3 & 6 \end{pmatrix},$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \mathbf{A}^+ \mathbf{b} = \frac{1}{11} \begin{pmatrix} 7 \\ 2 \end{pmatrix}.$$

所以拟合观察点的最佳标准椭圆方程为

$$\frac{7}{11}x^2 + \frac{2}{11}y^2 = 1,$$

其误差为

$$\|\delta\|_2 = \|\mathbf{A}\beta - \mathbf{b}\|_2 = 0.5.$$

习 题 四

1. 求矩阵 $\mathbf{A} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \\ 2 & -1 \end{pmatrix}$ 的一个左逆.
2. 求矩阵 $\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ 的一个右逆.
3. 设矩阵 $\mathbf{A} \in \mathbf{C}^{m \times m}$ 和 $\mathbf{C} \in \mathbf{C}^{n \times n}$ 可逆, 证明:
 - (1) 若 $\mathbf{B} \in \mathbf{C}^{m \times n}$ 是左可逆的, 则 \mathbf{ABC} 左可逆;
 - (2) 若 $\mathbf{B} \in \mathbf{C}^{m \times n}$ 是右可逆的, 则 \mathbf{ABC} 右可逆.
4. 设 $\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{Q}$, 其中 \mathbf{P} 列满秩, \mathbf{Q} 行满秩, 证明 $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{B})$.
5. 求矩阵 $\mathbf{A} = \begin{pmatrix} 1 & -1 & 2 \\ 2 & 2 & 3 \end{pmatrix}$ 的减号广义逆.
6. 设 $\mathbf{A} \in \mathbf{C}^{m \times n}$, $\mathbf{G} \in \mathbf{C}^{n \times m}$, 子空间 T 满足 $\mathbf{C}^n = N(\mathbf{A}) \oplus T$, 且 $\forall \mathbf{x} \in T, \mathbf{G}\mathbf{Ax} = \mathbf{x}$, 证明 $\mathbf{G} \in \mathbf{A}\{1\}$.
 7. 求矩阵 $\mathbf{A} = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 0 & -2 \\ -1 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix}$ 的 M-P 逆 \mathbf{A}^+ .
 8. 设 \mathbf{A} 是幂等 Hermite 矩阵, 证明 $\mathbf{A}^+ = \mathbf{A}$.
 9. 设 \mathbf{A} 的满秩分解为 $\mathbf{A} = \mathbf{BC}$, 证明 $\mathbf{Ax} = \mathbf{0}$ 的充分必要条件是 $\mathbf{Cx} = \mathbf{0}$.
 10. 设 \mathbf{A}, \mathbf{B} 为投影矩阵, 证明 $\mathbf{A} + \mathbf{B}$ 仍为投影矩阵当且仅当 $\mathbf{AB} = \mathbf{BA} = \mathbf{0}$.
 11. 设 \mathbf{A}, \mathbf{B} 为投影矩阵, 证明 $\mathbf{A} - \mathbf{B}$ 仍为投影矩阵当且仅当 $\mathbf{AB} = \mathbf{BA} = \mathbf{B}$.
 12. 证明 $\forall \mathbf{A} \in \mathbf{C}^{m \times n}$, 恒有

$$N(\mathbf{A}^+) = N(\mathbf{A}^H),$$

$$R(\mathbf{A}^+) = R(\mathbf{A}^H).$$
 13. 设 $\mathbf{A} \in \mathbf{C}^{m \times n}$, $\mathbf{B} \in \mathbf{C}^{n \times m}$, $\mathbf{BA} = \mathbf{I}_n$, $\text{rank}(\mathbf{AB}) = n$ ($m \geq n$),
 - (1) 求 \mathbf{AB} 的全部特征值;

(2) 证明 $R(\mathbf{AB}) = R(\mathbf{A})$, $N(\mathbf{AB}) = N(\mathbf{B})$.

14. 求线性方程组 $\begin{pmatrix} 0 & 2 & 0 \\ 1 & 0 & 2 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ 的最佳的最小二乘解.

第 5 章 矩阵分析

矩阵分析理论的建立,同数学分析一样,也是以极限理论为基础的,其内容丰富,是研究数值分析方法和其他数学分支以及许多工程问题的重要工具.本章在引入向量和矩阵的范数及极限的基础上,借助矩阵幂级数来定义矩阵函数,同时讨论函数矩阵的微分和积分,最后介绍它们在求解微分方程组中的应用.

5.1 向量范数

在欧氏空间与酉空间中,我们通过向量的内积定义了向量的长度 $\| \mathbf{x} \| = \sqrt{(\mathbf{x}, \mathbf{x})}$. 对于一般的线性空间,我们能否引入一个类似长度而又比其含义更广泛的概念呢? 为此,我们引入向量范数的概念.

一、向量范数的概念

定义 5.1 设 V 是数域 F 上的线性空间,且对于 V 的任一个向量 \mathbf{x} ,对应一个非负实数 $\| \mathbf{x} \|$,满足以下条件:

(1) 正定性: $\| \mathbf{x} \| \geq 0$, $\| \mathbf{x} \| = 0$ 当且仅当 $\mathbf{x} = \mathbf{0}$;

(2) 齐次性: $\| a\mathbf{x} \| = |a| \cdot \| \mathbf{x} \|$, $a \in F$;

(3) 三角不等式: 对任意 $\mathbf{x}, \mathbf{y} \in V$,都有 $\| \mathbf{x} + \mathbf{y} \| \leq \| \mathbf{x} \| + \| \mathbf{y} \|$.

则称 $\| \mathbf{x} \|$ 为向量 \mathbf{x} 的范数, $[V; \| \cdot \|]$ 为赋范空间.

例 1 在 n 维酉空间 \mathbf{C}^n 中, 复向量 $\mathbf{x} = (x_1 \quad x_2 \quad \cdots \quad x_n)^T$ 的长度

$$\| \mathbf{x} \| = \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2}$$

就是一种向量范数. 通常称这种范数为 2-范数, 记作 $\| \mathbf{x} \|_2$.

例 2 证明 $\| \mathbf{x} \| = \max_i |x_i|$ 是 \mathbf{C}^n 上的一种向量范数. 其中 $\mathbf{x} = (x_1 \quad x_2 \quad \cdots \quad x_n)^T \in \mathbf{C}^n$.

证明 当 $\mathbf{x} \neq \mathbf{0}$ 时, x_1, x_2, \dots, x_n 不全为零, 故 $\| \mathbf{x} \| = \max_i |x_i| > 0$; $\| \mathbf{x} \| = 0$, $\max_i |x_i| = 0$, 当且仅当 $x_i = 0$ ($i = 1, 2, \dots, n$), 即 $\mathbf{x} = \mathbf{0}$.

又对任意 $a \in \mathbf{C}^n$, 有

$$\| a\mathbf{x} \| = \max_i |ax_i| = |a| \max_i |x_i| = |a| \| \mathbf{x} \|.$$

最后, 我们来证明三角不等式也满足. 对任意 $\mathbf{y} = (y_1 \quad y_2 \quad \cdots \quad y_n)^T \in \mathbf{C}^n$, 有

$$\| \mathbf{x} + \mathbf{y} \| = \max_i |x_i + y_i| \leq \max_i |x_i| + \max_i |y_i|$$

$$= \| \mathbf{x} \| + \| \mathbf{y} \|.$$

故 $\| \mathbf{x} \| = \max_i |x_i|$ 是 \mathbf{C}^n 上的一种向量范数, 称这种范数为 ∞ -范数, 记做 $\| \mathbf{x} \|_\infty$.

同样, 可以证明 $\| \mathbf{x} \| = \sum_{i=1}^n |x_i|$ 也是 \mathbf{C}^n 的一种向量范数, 称这种范数为 1-范数, 记做 $\| \mathbf{x} \|_1$ (证明留给读者).

一般地, 可以证明 $\| \mathbf{x} \| = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$ ($1 \leq p < \infty$) 也是 \mathbf{C}^n 的一种向量范数, 称这种范数为 p -范数, 记做 $\| \mathbf{x} \|_p$, 其证明相当繁琐, 这里不作介绍.

例 3 设 A 是任意一个 n 阶正定矩阵, $\mathbf{x} \in \mathbf{R}^n$ 是一个 n 元列向量. 证明

$$\| \mathbf{x} \|_A = (\mathbf{x}^T A \mathbf{x})^{1/2}$$

是一种向量范数, 称为加权范数或椭圆范数.

证明 因为 A 正定, 所以 $\| \mathbf{x} \|_A = (\mathbf{x}^T A \mathbf{x})^{1/2} \geq 0$, 且 $\| \mathbf{x} \|_A = 0$ 当且仅当 $\mathbf{x} = \mathbf{0}$.

又对任意 $a \in \mathbf{R}$, 有

$$\begin{aligned} \| a\mathbf{x} \|_A &= ((a\mathbf{x})^T A (a\mathbf{x}))^{1/2} = (a^2 \mathbf{x}^T A \mathbf{x})^{1/2} \\ &= |a| \| \mathbf{x} \|_A. \end{aligned}$$

最后, 我们来证明三角不等式也满足. 因为 A 正定, 所以存在可逆矩阵 P , 使 $P^T A P = I$, 从而

$$A = (P^T)^{-1} P^{-1} = (P^{-1})^T P^{-1} = B^T B,$$

其中 $B = P^{-1}$ 为可逆矩阵, 于是

$$\| \mathbf{x} \|_A = (\mathbf{x}^T A \mathbf{x})^{1/2} = (\mathbf{x}^T B^T B \mathbf{x})^{1/2} = [(\mathbf{Bx})^T \mathbf{Bx}]^{1/2} = \| \mathbf{Bx} \|_2,$$

从而

$$\begin{aligned} \| \mathbf{x} + \mathbf{y} \|_A &= \| \mathbf{B}(\mathbf{x} + \mathbf{y}) \|_2 = \| \mathbf{Bx} + \mathbf{By} \|_2 \\ &\leq \| \mathbf{Bx} \|_2 + \| \mathbf{By} \|_2 = \| \mathbf{x} \|_A + \| \mathbf{y} \|_A. \end{aligned}$$

二、向量范数的连续性与等价性

下面介绍向量范数的两个性质, 第一个性质是向量范数是向量坐标的连续函数.

定理 5.1 设 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 \mathbf{C}^n 空间的一组基, $\| \mathbf{x} \|$ 是 \mathbf{C}^n 空间的任意一个向量范数, 则对任意 $\epsilon > 0$, $\alpha = \sum_{i=1}^n a_i \alpha_i$, $\beta = \sum_{i=1}^n b_i \alpha_i \in \mathbf{C}^n$, 存在 $\delta > 0$, 当 $|a_i - b_i| < \delta$ ($i = 1, 2, \dots, n$) 时, 有

$$|\| \alpha \| - \| \beta \| | < \epsilon.$$

证明 令 $M = \max_i \{ \| \alpha_i \| \}$, 则 $M > 0$, 选取 δ , 使得 $0 < \delta < \frac{\epsilon}{nM}$. 对任意向量 $\alpha, \beta \in \mathbf{C}^n$, 当 $|a_i - b_i| < \delta$ ($i = 1, 2, \dots, n$) 时, 有

$$|\| \alpha \| - \| \beta \| | \leq \| \alpha - \beta \| \leq \sum_{i=1}^n |a_i - b_i| \| \alpha_i \|$$

$$\leq M \sum_{i=1}^n |a_i - b_i| < \epsilon.$$

□

由前面的讨论可见,在有限维线性空间(如 \mathbf{R}^n 及 \mathbf{C}^n)上可以引入各种各样的向量范数. 虽然范数的种数可无穷多,但下面我们可以证明这些范数都是彼此等价的.

定义 5.2 设 $\|\mathbf{x}\|^{(1)}$ 与 $\|\mathbf{x}\|^{(2)}$ 是线性空间 V 上定义的两种向量范数. 如果存在正数 c_1 与 c_2 , 使得

$$c_1 \|\mathbf{x}\|^{(2)} \leq \|\mathbf{x}\|^{(1)} \leq c_2 \|\mathbf{x}\|^{(2)}, \forall \mathbf{x} \in V,$$

则称这两个向量范数是等价的.

定理 5.2 有限维线性空间的任意两种向量范数都是等价的.

证明 因为 n 维线性空间 V 和 \mathbf{C}^n 空间是同构的, 所以只就 \mathbf{C}^n 空间来证明.

设 $\|\mathbf{x}\|^{(1)}$ 和 $\|\mathbf{x}\|^{(2)}$ 是 \mathbf{C}^n 空间的任意两种向量范数. 当 $\mathbf{x} = \mathbf{0}$ 时, 不等式显然满足. 下面证明对 $\mathbf{x} \neq \mathbf{0}$ 不等式也满足. 在 \mathbf{C}^n 空间中, 单位球壳

$$S_1 = \{\mathbf{x} \in \mathbf{C}^n \mid \|\mathbf{x}\|^{(1)} = 1\},$$

$$S_2 = \{\mathbf{x} \in \mathbf{C}^n \mid \|\mathbf{x}\|^{(2)} = 1\}$$

是有界闭集. $\|\mathbf{x}\|^{(1)}$ 与 $\|\mathbf{x}\|^{(2)}$ 都是 S_1 与 S_2 上的连续函数, 于是可取到最大值, 设

$$t_1 = \max_{\mathbf{x} \in S_1} \{\|\mathbf{x}\|^{(2)}\},$$

$$t_2 = \max_{\mathbf{x} \in S_2} \{\|\mathbf{x}\|^{(1)}\},$$

则 $t_1, t_2 > 0$. 对任意 $\mathbf{x} \neq \mathbf{0}$, 有 $\frac{\mathbf{x}}{\|\mathbf{x}\|^{(1)}} \in S_1$, $\frac{\mathbf{x}}{\|\mathbf{x}\|^{(2)}} \in S_2$, 于是

$$t_1 \geq \left\| \frac{\mathbf{x}}{\|\mathbf{x}\|^{(1)}} \right\|^{(2)} = \frac{\|\mathbf{x}\|^{(2)}}{\|\mathbf{x}\|^{(1)}},$$

$$t_2 \geq \left\| \frac{\mathbf{x}}{\|\mathbf{x}\|^{(2)}} \right\|^{(1)} = \frac{\|\mathbf{x}\|^{(1)}}{\|\mathbf{x}\|^{(2)}}.$$

取 $c_1 = t_2$, $c_2 = 1/t_1$, 便有

$$c_1 \|\mathbf{x}\|^{(2)} \leq \|\mathbf{x}\|^{(1)} \leq c_2 \|\mathbf{x}\|^{(2)}.$$

□

注意: 在无限维线性空间中, 两个向量范数是可以不等价的.

5.2 矩阵范数

矩阵空间 $\mathbf{C}^{m \times n}$ 是一个 mn 维的线性空间, 将 $m \times n$ 矩阵 A 看作线性空间 $\mathbf{C}^{m \times n}$ 中的向量, 可以按向量范数的办法来定义范数. 但是, 矩阵之间还有乘法运算, 它应该在定义矩阵范数时予以体现. 本书只涉及方阵的范数, 至于不是方阵的范数可类似地定义.

一、矩阵范数的概念

定义 5.3 在 $F^{n \times n}$ 上定义一个非负实值函数, 使得对任意矩阵 $A \in F^{n \times n}$, 对应一个

非负实数 $\|A\|$, 满足以下四个条件:

- (1) 正定性: $\|A\| \geq 0$, $\|A\| = 0$ 当且仅当 $A = \mathbf{0}$;
- (2) 齐次性: $\|aA\| = |a|\|A\|$, $a \in F$;
- (3) 三角不等式: 对任意 $A, B \in F^{n \times n}$, 都有 $\|A+B\| \leq \|A\| + \|B\|$;
- (4) 相容性: 对任意 $A, B \in F^{n \times n}$, 都有 $\|AB\| \leq \|A\|\|B\|$,

则称 $\|A\|$ 为矩阵 A 的范数.

例 4 证明: 对任意 $A = (a_{ij}) \in \mathbf{C}^{n \times n}$, $\|A\| = \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|$ 是矩阵范数.

证明 正定性与齐次性容易验证, 下证三角不等式与相容性满足.

设 $A = (a_{ij}), B = (b_{ij}) \in \mathbf{C}^{n \times n}$, 则

$$\begin{aligned}\|A+B\| &= \sum_{i=1}^n \sum_{j=1}^n |a_{ij} + b_{ij}| \leq \sum_{i=1}^n \sum_{j=1}^n (|a_{ij}| + |b_{ij}|) \\ &= \sum_{i=1}^n \sum_{j=1}^n |a_{ij}| + \sum_{i=1}^n \sum_{j=1}^n |b_{ij}| = \|A\| + \|B\|, \\ \|AB\| &= \sum_{i=1}^n \sum_{j=1}^n |a_{i1}b_{1j} + \dots + a_{in}b_{nj}| \\ &\leq \sum_{i=1}^n \sum_{j=1}^n (|a_{i1}b_{1j}| + \dots + |a_{in}b_{nj}|) \\ &\leq \sum_{i=1}^n (|a_{i1}| + \dots + |a_{in}|) \cdot \sum_{j=1}^n (|b_{1j}| + \dots + |b_{nj}|) \\ &= \sum_{i=1}^n \sum_{j=1}^n |a_{ij}| \cdot \sum_{i=1}^n \sum_{j=1}^n |b_{ij}| = \|A\|\|B\|.\end{aligned}$$

例 5 设 $A = (a_{ij}) \in \mathbf{C}^{n \times n}$, 证明

$$\|A\|_F = \left(\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2} = [\operatorname{tr}(A^H A)]^{1/2}$$

是 $\mathbf{C}^{n \times n}$ 上的矩阵范数. 称其为 Frobenius 范数, 简称为 F-范数.

证明 正定性与齐次性易证, 下证三角不等式与相容性满足.

设 $A = (a_{ij}), B = (b_{ij}) \in \mathbf{C}^{n \times n}$, 则

$$\begin{aligned}\|A+B\|_F &= \left(\sum_{i=1}^n \sum_{j=1}^n |a_{ij} + b_{ij}|^2 \right)^{1/2} \\ &\leq \left(\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2} + \left(\sum_{i=1}^n \sum_{j=1}^n |b_{ij}|^2 \right)^{1/2} \\ &= \|A\|_F + \|B\|_F; \\ \|AB\|_F^2 &= \sum_{i=1}^n \sum_{j=1}^n \left(\left| \sum_{k=1}^n a_{ik}b_{kj} \right|^2 \right)\end{aligned}$$

$$\begin{aligned}
&\leq \sum_{i=1}^n \sum_{j=1}^n \left[\left(\sum_{k=1}^n |a_{ik}|^2 \right) \left(\sum_{k=1}^n |b_{kj}|^2 \right) \right] \\
&= \left(\sum_{i=1}^n \sum_{k=1}^n |a_{ik}|^2 \right) \left(\sum_{j=1}^n \sum_{k=1}^n |b_{kj}|^2 \right) \\
&= \|A\|_F^2 \|B\|_F^2.
\end{aligned}$$

例 6 设 $A \in \mathbf{C}^{m \times n}$, $\sigma_1, \sigma_2, \dots, \sigma_r$ 是矩阵 A 的奇异值, 证明

$$\|A\|_F^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2.$$

证明 首先注意到由特征值的定义, $A^H A$ 的非零特征值为 $\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2$.

从而

$$\text{tr}(A^H A) = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2.$$

由定义

$$\|A\|_F^2 = \text{tr}(A^H A) = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2.$$

例 7 设 $m \times n$ 阶矩阵 A 的秩为 r , A 的奇异值展开式为

$$A = U \Sigma V^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_r u_r v_r^T.$$

令

$$M = \{S \mid S \in \mathbf{C}^{m \times n}, \text{rank}(S) \leq k\}, k \leq r,$$

取

$$A_k = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_k u_k v_k^T, (k \leq r)$$

证明

$$\|A - A_k\|_F = \min_{S \in M} \|A - S\|_F.$$

证明 首先可注意到

$$\|A - A_k\|_F = (\sigma_{k+1}^2 + \sigma_{k+2}^2 + \dots + \sigma_r^2)^{\frac{1}{2}}.$$

设 $x \in M$, 且 $\|A - x\|_F = \min_{S \in M} \|A - S\|_F$, 由 x 的定义

$$\|A - x\|_F \leq \|A - A_k\|_F = (\sigma_{k+1}^2 + \dots + \sigma_r^2)^{\frac{1}{2}} \quad (1)$$

又设 x 的奇异值分解为

$$x = Q^T \begin{pmatrix} \Omega_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} P = Q^T \Omega P,$$

将 A 表示为

$$A = Q Q^T A P P^T = Q B P^T,$$

其中

$$B = Q^T A P,$$

则

$$\|A - x\|_F = \|Q^T (B - \Omega) P\|_F = \|B - \Omega\|_F.$$

将 B 分块方式相应分块为

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},$$

$$\begin{aligned}
\|A - x\|_F^2 &= \|B - \Omega\|_F^2 \\
&= \|B_{11} - \Omega_k\|_F^2 + \|B_{12}\|_F^2 + \|B_{21}\|_F^2 + \|B_{22}\|_F^2,
\end{aligned}$$

由 x 的定义, 必有 $B_{12} = 0, B_{22} = 0, B_{11} = \Omega_k$. 否则可取

$$z = \begin{pmatrix} \Omega_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \in M.$$

$\|A - z\|_F^2 = \|B_{22}\|_F^2 \leq \|A - x\|_F^2$, 与 x 的定义矛盾. 因此
 $\|A - x\|_F = \|B_{22}\|_F$.

设 B_{22} 的奇异值分解为 $B_{22} = U_1^T \Lambda V_1$, 令

$$U_2 = \begin{pmatrix} I_k & \mathbf{0} \\ \mathbf{0} & U_1 \end{pmatrix}, \quad V_2 = \begin{pmatrix} I_k & \mathbf{0} \\ \mathbf{0} & V_1 \end{pmatrix},$$

则

$$U_2^T B V_2 = U_2^T Q^T A P V_2 = \begin{pmatrix} \Omega_k \\ \Lambda \end{pmatrix},$$

从而

$$A = (Q U_2) \begin{pmatrix} \Omega_k & \mathbf{0} \\ \mathbf{0} & \Lambda \end{pmatrix} (P V_2)^T.$$

矩阵 Λ 的主对角线上元素是矩阵 A 的奇异值, 所以

$$\|A - x\|_F = \|B_{22}\|_F = \|\Lambda\|_F \geq (\sigma_{k+1}^2 + \dots + \sigma_r^2)^{\frac{1}{2}} \quad \textcircled{2}$$

由式①和式②, 可得

$$\|A - x\|_F = (\sigma_{k+1}^2 + \dots + \sigma_r^2)^{\frac{1}{2}} = \|A - A_k\|_F,$$

即

$$\|A - A_k\|_F = \min_{S \in \mathcal{M}} \|A - S\|_F. \quad \square$$

二、诱导范数

如同向量范数的情况一样, 矩阵也可以有各种各样的范数, 而且在大多数情况下, 需要涉及矩阵 $A \in \mathbf{C}^{m \times n}$ 和向量 $X \in \mathbf{C}^{n \times 1}$ 的运算 AX . 因此, 矩阵范数常和向量范数混合在一起使用. 考虑一些矩阵范数时, 应该使它能与向量范数联系起来. 这可由矩阵范数与向量范数相容的概念来实现. 下面引入这个概念.

定义 5.4 设 $\|x\|$ 是向量范数, $\|A\|$ 是矩阵范数, 若

$$\|Ax\| \leq \|A\| \cdot \|x\|,$$

则称矩阵范数 $\|A\|$ 与向量范数 $\|x\|$ 是相容的.

矩阵的 F-范数与向量的 2-范数是相容的. 事实上, 因为 $\|A\|_F = (\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2)^{1/2}$,

$\|x\|_2 = (\sum_{i=1}^n |x_i|^2)^{1/2}$, 所以

$$\begin{aligned} \|Ax\|_2^2 &= \sum_{i=1}^n \left(\left| \sum_{j=1}^n a_{ij} x_j \right|^2 \right) \leq \sum_{i=1}^n \left(\left(\sum_{j=1}^n |a_{ij}|^2 \right) \left(\sum_{j=1}^n |x_j|^2 \right) \right) \\ &= \left(\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \right) \left(\sum_{j=1}^n |x_j|^2 \right) \\ &= \|A\|_F^2 \|x\|_2^2, \end{aligned}$$

于是

$$\|Ax\|_2 \leq \|A\|_F \|x\|_2.$$

定理 5.3 设 $\|x\|$ 是向量范数, 则

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

是与向量范数 $\| \mathbf{x} \|$ 相容的矩阵范数. 称其为由向量范数 $\| \mathbf{x} \|$ 所诱导的诱导范数.

证明 正定性与齐次性显然.

根据向量范数的三角不等式, 有

$$\begin{aligned}\| \mathbf{A} + \mathbf{B} \| &= \max_{\mathbf{x} \neq 0} \left\{ \frac{\| (\mathbf{A} + \mathbf{B}) \mathbf{x} \|}{\| \mathbf{x} \|} \right\} = \max_{\mathbf{x} \neq 0} \left\{ \frac{\| \mathbf{Ax} + \mathbf{Bx} \|}{\| \mathbf{x} \|} \right\} \\ &\leq \max_{\mathbf{x} \neq 0} \left\{ \frac{\| \mathbf{Ax} \| + \| \mathbf{Bx} \|}{\| \mathbf{x} \|} \right\} \\ &\leq \max_{\mathbf{x} \neq 0} \left\{ \frac{\| \mathbf{Ax} \|}{\| \mathbf{x} \|} \right\} + \max_{\mathbf{x} \neq 0} \left\{ \frac{\| \mathbf{Bx} \|}{\| \mathbf{x} \|} \right\} = \| \mathbf{A} \| + \| \mathbf{B} \|.\end{aligned}$$

现证矩阵范数的相容性. 设 $\mathbf{B} \neq \mathbf{0}$, 则

$$\begin{aligned}\| \mathbf{AB} \| &= \max_{\mathbf{x} \neq 0} \left\{ \frac{\| \mathbf{ABx} \|}{\| \mathbf{x} \|} \right\} = \max_{\mathbf{x} \neq 0} \left\{ \frac{\| \mathbf{A}(\mathbf{Bx}) \|}{\| \mathbf{Bx} \|} \cdot \frac{\| \mathbf{Bx} \|}{\| \mathbf{x} \|} \right\} \\ &\leq \max_{\mathbf{x} \neq 0} \left\{ \frac{\| \mathbf{A}(\mathbf{Bx}) \|}{\| \mathbf{Bx} \|} \right\} \cdot \max_{\mathbf{x} \neq 0} \left\{ \frac{\| \mathbf{Bx} \|}{\| \mathbf{x} \|} \right\} \\ &\leq \| \mathbf{A} \| \cdot \| \mathbf{B} \|.\end{aligned}$$

$\mathbf{B} = \mathbf{0}$ 或存在 \mathbf{x} , 使 $\mathbf{Bx} = \mathbf{0}$, 上面推导仍能成立, 请读者思考. 最后, 我们来证明 $\| \mathbf{A} \|$ 与 $\| \mathbf{x} \|$ 是相容的. 因为

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

所以

$$\| \mathbf{Ax} \| \leq \| \mathbf{A} \| \cdot \| \mathbf{x} \|.$$

□

由 $\| \mathbf{x} \|_p$ 所诱导的矩阵范数称为矩阵 p -范数. 常用的 p -范数为 $\| \mathbf{A} \|_1$, $\| \mathbf{A} \|_2$ 与 $\| \mathbf{A} \|_\infty$. 我们不加证明的给出这三种范数的计算方法.

- (1) $\| \mathbf{A} \|_1 = \max_j (\sum_{i=1}^n |a_{ij}|)$, 称为列和范数;
- (2) $\| \mathbf{A} \|_2 = \sqrt{\lambda}$, λ 为 $\mathbf{A}^\text{T} \mathbf{A}$ 的最大特征值, 称为谱范数;
- (3) $\| \mathbf{A} \|_\infty = \max_i (\sum_{j=1}^n |a_{ij}|)$, 称为行和范数.

5.3 向量序列和矩阵序列的极限

定义 5.5 设 $\mathbf{x}^{(k)} = (x_1^{(k)} \quad x_2^{(k)} \quad \cdots \quad x_n^{(k)})^\text{T}$, $k=1, 2, \dots$ 是 \mathbf{C}^n 空间的一个向量序列, 如果当 $k \rightarrow +\infty$ 时, 它的 n 个分量数列都收敛, 即

$$\lim_{k \rightarrow \infty} x_i^{(k)} = a_i, \quad i = 1, 2, \dots, n,$$

则称向量序列 $\{\mathbf{x}^{(k)}\}$ 是按分量收敛的. 向量 $\alpha = (a_1 \quad a_2 \quad \cdots \quad a_n)^\text{T}$ 是它的极限, 记为 $\lim_{k \rightarrow \infty} \mathbf{x}^{(k)} = \alpha$ 或 $\mathbf{x}^{(k)} \rightarrow \alpha$. 如果至少有一个分量数列是发散的, 则称向量序列是发散的.

例如向量序列

$$\mathbf{x}^{(k)} = \begin{pmatrix} \frac{1}{2^k} \\ \frac{\sin k}{k} \end{pmatrix}, \quad k = 1, 2, \dots$$

是收敛的,因为 $k \rightarrow \infty$ 时, $\frac{1}{2^k} \rightarrow 0$, $\frac{\sin k}{k} \rightarrow 0$, 所以

$$\lim_{k \rightarrow \infty} \mathbf{x}^{(k)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

而向量序列

$$\mathbf{x}^{(k)} = \begin{pmatrix} 1 - \frac{1}{k} \\ \sin k \end{pmatrix}, \quad k = 1, 2, \dots$$

是发散的,因为 $\lim_{k \rightarrow \infty} \sin k$ 不存在.

定义 5.6 设 $\{\mathbf{x}^{(k)}\}$ 是 \mathbf{C}^n 空间的一个向量序列, $\|\mathbf{x}\|$ 是 \mathbf{C}^n 空间的一个向量范数. 如果存在向量 $\alpha \in \mathbf{C}^n$, 当 $k \rightarrow \infty$ 时, $\|\mathbf{x}^{(k)} - \alpha\| \rightarrow 0$, 则称向量序列按向量范数收敛于 α .

向量序列按分量收敛与按向量范数收敛有何关系呢? 我们有下面的定理.

定理 5.4 设 $\{\mathbf{x}^k\}$ 是 \mathbf{C}^n 空间的一个向量序列, 它按分量收敛的充分必要条件是它按 \mathbf{C}^n 空间的任意一个向量范数收敛.

证明 利用范数的等价性, 易知只要对于一种范数来证明. 为此取 $\|\mathbf{x}\|_\infty$ 来证明.

设 $\mathbf{x}^{(k)} = (x_1^{(k)} \quad x_2^{(k)} \quad \cdots \quad x_n^{(k)})^\top$, $\alpha = (a_1 \quad a_2 \quad \cdots \quad a_n)^\top$, 则

$$\|\mathbf{x}^{(k)} - \alpha\|_\infty \rightarrow 0$$

当且仅当

$$\max_i |x_i^{(k)} - a_i| \rightarrow 0,$$

当且仅当对每个 i ($i=1, 2, \dots, n$), 有

$$|x_i^{(k)} - a_i| \rightarrow 0,$$

此即

$$\mathbf{x}^{(k)} \rightarrow \alpha. \quad \square$$

定理 5.4 表明, 尽管不同的向量范数可能具有不同的大小, 然而在各种范数下考虑向量序列的敛散性问题时, 却表现出明显的一致性. 这就是说, 如果向量序列 $\{\mathbf{x}^{(k)}\}$ 对某一种范数收敛, 且极限为 α , 则对其他范数这个序列仍然收敛, 并且具有相同的极限 α .

对于矩阵范数, 我们有类似的结果.

定义 5.7 设 $\mathbf{A}^{(k)} = (a_{ij}^{(k)}) \in \mathbf{C}^{n \times n}$, $\{\mathbf{A}^{(k)}\}$ 是一个矩阵序列. 如果当 $k \rightarrow \infty$ 时, 它的 n^2 个数列 $\{a_{ij}^{(k)}\}$ 都收敛, 即

$$\lim_{k \rightarrow \infty} a_{ij}^{(k)} = a_{ij}, \quad i, j = 1, 2, \dots, n,$$

则称矩阵序列 $\{\mathbf{A}^{(k)}\}$ 按元素数列收敛. 矩阵 $\mathbf{A} = (a_{ij}) \in \mathbf{C}^{n \times n}$ 是它的极限, 记为 $\lim_{k \rightarrow \infty} \mathbf{A}^{(k)} = \mathbf{A}$ 或 $\mathbf{A}^{(k)} \rightarrow \mathbf{A}$. 如果至少有一个元素数列是发散的, 则称该矩阵序列发散.

例如:

$$\begin{pmatrix} \left(1 + \frac{1}{k}\right)^k & 1 + \frac{1}{k} \\ -1 & \frac{(-1)^k}{k} \end{pmatrix} \rightarrow \begin{pmatrix} e & 1 \\ -1 & 0 \end{pmatrix}.$$

与数列收敛的性质类似,矩阵序列收敛满足以下性质:

(1) 设 $\mathbf{A}^{(k)} \rightarrow \mathbf{A}, \mathbf{B}^{(k)} \rightarrow \mathbf{B}$, 则

$$\alpha \mathbf{A}^{(k)} + \beta \mathbf{B}^{(k)} \rightarrow \alpha \mathbf{A} + \beta \mathbf{B}, \quad \alpha, \beta \in \mathbf{C}.$$

(2) 设 $\mathbf{A}^{(k)} \rightarrow \mathbf{A}, \mathbf{B}^{(k)} \rightarrow \mathbf{B}$, 则

$$\mathbf{A}^{(k)} \mathbf{B}^{(k)} \rightarrow \mathbf{AB}.$$

(3) 设 $\mathbf{A}^{(k)}$ 与 \mathbf{A} 都是可逆矩阵,且 $\mathbf{A}^{(k)} \rightarrow \mathbf{A}$, 则

$$(\mathbf{A}^{(k)})^{-1} \rightarrow \mathbf{A}^{-1}.$$

证明 只证明(2)、(3). 设 $\mathbf{A}^{(k)} = (a_{ij}^{(k)}), \mathbf{B}^{(k)} = (b_{ij}^{(k)}), \mathbf{A} = (a_{ij}), \mathbf{B} = (b_{ij}) \in \mathbf{C}^{n \times n}$.

(2) 由题设,有

$$\begin{aligned} a_{ij}^{(k)} &\rightarrow a_{ij} & i = 1, 2, \dots, n \\ b_{ij}^{(k)} &\rightarrow b_{ij} & j = 1, 2, \dots, n \end{aligned}$$

于是矩阵 $\mathbf{A}^{(k)} \mathbf{B}^{(k)}$ 的第 i 行第 j 列的元素为

$$\sum_{t=1}^n a_{it}^{(k)} b_{tj}^{(k)} \rightarrow \sum_{t=1}^n a_{it} b_{tj},$$

故

$$\mathbf{A}^{(k)} \mathbf{B}^{(k)} \rightarrow \mathbf{AB}.$$

(3) 因为 $(\mathbf{A}^{(k)})^{-1} = \frac{1}{|\mathbf{A}^{(k)}|} (\mathbf{A}^{(k)})^*$,

这里 $(\mathbf{A}^{(k)})^*$ 是 $\mathbf{A}^{(k)}$ 的伴随矩阵,它的元素与 $|\mathbf{A}^{(k)}|$ 的元素均为 $\mathbf{A}^{(k)}$ 的元素的多项式,于是

$$(\mathbf{A}^{(k)})^* \rightarrow \mathbf{A}^*, \quad |\mathbf{A}^{(k)}| \rightarrow |\mathbf{A}|,$$

所以有 $(\mathbf{A}^{(k)})^{-1} = \frac{1}{|\mathbf{A}^{(k)}|} (\mathbf{A}^{(k)})^* \rightarrow \frac{1}{|\mathbf{A}|} \mathbf{A}^* = \mathbf{A}^{-1}$.

值得注意的是性质(3)中 $\mathbf{A}^{(k)}$ 与 \mathbf{A} 都可逆时才有 $(\mathbf{A}^{(k)})^{-1} \rightarrow \mathbf{A}^{-1}$. 如仅有 $\mathbf{A}^{(k)}$ 可逆,不能保证其极限 \mathbf{A} 可逆. 例如取

$$\mathbf{A}^{(k)} = \begin{pmatrix} 1 + \frac{1}{k} & 1 \\ 1 & 1 + \frac{1}{k} \end{pmatrix},$$

则对任意 $k=1, 2, \dots$ 由 $|\mathbf{A}^{(k)}| = \left(1 + \frac{1}{k}\right)^2 - 1 \neq 0$ 知 $\mathbf{A}^{(k)}$ 皆可逆. 但

$$\lim_{k \rightarrow \infty} \mathbf{A}^{(k)} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \mathbf{A},$$

其极限 \mathbf{A} 不可逆.

定义 5.8 设 $\{A^{(k)}\}$ 是 $\mathbf{C}^{n \times n}$ 空间的一个矩阵序列, $\|A\|$ 是 $\mathbf{C}^{n \times n}$ 的一个矩阵范数. 如果存在矩阵 $A \in \mathbf{C}^{n \times n}$, 当 $k \rightarrow \infty$ 时, $\|A^{(k)} - A\| \rightarrow 0$, 则称矩阵序列按矩阵范数收敛于 A .

仿照定理 5.4 的证明, 有

定理 5.5 设 $\{A^{(k)}\}$ 是 $\mathbf{C}^{n \times n}$ 的一个矩阵序列, 它按元素数列收敛的充分必要条件是它按 $\mathbf{C}^{n \times n}$ 的任意一个矩阵范数收敛.

定义 5.9 矩阵序列 $\{A^{(k)}\}$ 称为有界的, 如果存在常数 $M > 0$, 使得对所有 k 都有

$$|a_j^{(k)}| \leq M \quad (i = 1, 2, \dots, m; \quad j = 1, 2, \dots, n).$$

在数学分析中已经知道, 有界数列必有收敛的子数列. 对于矩阵序列也有: 有界的矩阵序列 $\{A^{(k)}\}$ 必有收敛的子序列.

这一结论可以由数列的相应结论推出.

在矩阵序列中, 最常见的是由一个方阵的幂构成的序列, 关于这样的矩阵序列有以下的概念和结果.

定义 5.10 设 A 为方阵, 且当 $k \rightarrow \infty$ 时有 $A^k \rightarrow 0$, 则称 A 为收敛矩阵.

例 8 对 n 阶方阵 A 的方幂所作成的矩阵序列 $\{A^k\}$, 如果对某一种矩阵范数有 $\|A\| < 1$, 证明 $\lim_{k \rightarrow \infty} A^k = \mathbf{0}$, 即 A 为收敛矩阵.

证明 由矩阵范数的相容性, 有 $\|A^k\| \leq \|A\|^k$, 又 $\|A\| < 1$, 所以 $\lim_{k \rightarrow \infty} \|A^k\| = 0$, 由定理 5.5, 有 $\lim_{k \rightarrow \infty} A^k = \mathbf{0}$.

5.4 矩阵幂级数

幂级数的理论在数学分析中占重要地位. 在建立矩阵分析的理论时, 也特别重视讨论矩阵幂级数, 因为它是建立矩阵函数的依据. 在讨论矩阵幂级数时, 自然要讨论其敛散性. 矩阵幂级数的敛散性不仅与其收敛半径有关, 而且与其谱半径有关.

一、谱半径

定义 5.11 设矩阵 $A \in \mathbf{C}^{n \times n}$ 的全部特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 则称 $\rho(A) = \max_i |\lambda_i|$ 为 A 的谱半径.

由定义知, A 的全部特征值分布在复平面上以原点为中心, $\rho(A)$ 为半径的圆盘上.

例 9 设 $A \in \mathbf{C}^{n \times n}$, 则

$$\rho(A^k) = (\rho(A))^k.$$

证明 设 A 的全部特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 则 A^k 的全部特征值为 $\lambda_1^k, \lambda_2^k, \dots, \lambda_n^k$, 于是有

$$\rho(A^k) = \max_i |\lambda_i^k| = (\max_i |\lambda_i|)^k = (\rho(A))^k.$$

例 10 $A^k \rightarrow 0$ ($k \rightarrow \infty$) 的充分必要条件是 $\rho(A) < 1$.

证明 设 A 的 Jordan 标准形为 J , 则存在可逆矩阵 P , 使得

$$A = PJP^{-1}.$$

于是

$$A^k = PJ^k P^{-1}.$$

由此可见, $A^k \rightarrow 0$ 的充分必要条件是 $J^k \rightarrow 0$. 注意到

$$J^k = \begin{pmatrix} J_1^k(\lambda_1) & & & \\ & J_2^k(\lambda_2) & & \\ & & \ddots & \\ & & & J_s^k(\lambda_s) \end{pmatrix},$$

其中 $\lambda_1, \lambda_2, \dots, \lambda_s$ 是 A 的特征值. 又 $J^k \rightarrow 0$ 的充分必要条件是 $J_i^k(\lambda_i) \rightarrow 0$ ($i=1, 2, \dots, s$), 而

$$J_i^k(\lambda_i) = \begin{pmatrix} \lambda_i^k & C_k^1 \lambda_i^{k-1} & \cdots & C_k^{n_i-1} \lambda_i^{k-n_i+1} \\ & \lambda_i^k & \ddots & \cdots \\ & & \ddots & C_k^1 \lambda_i^{k-1} \\ & & & \lambda_i^{k_i} \end{pmatrix}.$$

当 $\rho(A) < 1$ 时, $|\lambda_i| < 1$ ($i=1, 2, \dots, s$), 于是

$$C_k^l \lambda_i^{k-l} \rightarrow 0 \quad (l=0, 1, 2, \dots, n_i - 1)$$

故

$$J_i^k(\lambda_i) \rightarrow 0.$$

反之, 若有某个特征值 $|\lambda_i| \geq 1$, 则 λ_i^k 不趋于零, 此时 $J_i^k(\lambda_i)$ 不趋于零, 从而 J^k 不趋于零.

综上, $A^k \rightarrow 0$ 的充要条件是 $\rho(A) < 1$.

下面讨论谱半径与矩阵范数之间的关系.

定理 5.6 设 $A \in \mathbf{C}^{n \times n}$, 则对 $\mathbf{C}^{n \times n}$ 上任意一种矩阵范数 $\|A\|$, 都有

$$\rho(A) \leq \|A\|,$$

即 A 的谱半径是 A 的任意一种矩阵范数的下界.

证明 设 λ 是 A 的任意一个特征值, $x \neq 0$ 是 λ 对应的特征向量, 作一个 n 阶方阵

$$B = (x, 0, \dots, 0) \neq 0,$$

则由 $Ax = \lambda x$, 有 $AB = \lambda B$. 于是

$$|\lambda| \cdot \|B\| = \|\lambda B\| = \|AB\| \leq \|A\| \cdot \|B\|,$$

又 $B \neq 0$, $\|B\| > 0$, 所以 $|\lambda| \leq \|A\|$, 从而 $\rho(A) \leq \|A\|$. □

定理 5.7 设 $A \in \mathbf{C}^{n \times n}$, 则对任意 $\epsilon > 0$, 存在某种矩阵范数 $\|A\|_*$, 使得

$$\|A\|_* \leq \rho(A) + \epsilon.$$

如果在圆周 $S = \{z \in \mathbf{C} \mid |z| = \rho(A)\}$ 上的特征值 λ 对应的 Jordan 子块都是一阶的, 则

$$\rho(A) = \|A\|_*.$$

证明 对 $A \in \mathbf{C}^{n \times n}$, 存在可逆阵 P , 使得

$$P^{-1}AP = J = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_m \end{pmatrix},$$

其中 $J_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & 1 & \\ & & \ddots & \ddots \\ & & & 1 \\ & & & \lambda_i \end{pmatrix}, \quad i = 1, 2, \dots, m.$

对任意 $\varepsilon > 0$, 令 $D = \begin{pmatrix} 1 & & & \\ & \varepsilon & & \\ & & \ddots & \\ & & & \varepsilon^{n-1} \end{pmatrix}$,

作变换 $J \rightarrow D^{-1}JD$, 则每个 Jordan 子块化为

$$\begin{pmatrix} \lambda & \varepsilon & & \\ \lambda_i & \varepsilon & & \\ & \ddots & \ddots & \\ & & & \varepsilon \\ & & & \lambda_i \end{pmatrix},$$

于是

$$\|(PD)^{-1}APD\|_{\infty} = \|D^{-1}JD\|_{\infty} \leqslant \rho(A) + \varepsilon.$$

现在定义 $\|A\|_* = \|(PD)^{-1}APD\|_{\infty}$, 易证它是 $\mathbf{C}^{n \times n}$ 上的一种矩阵范数, 从而证得 $\|A\|_* \leqslant \rho(A) + \varepsilon$.

设 A 的特征值的顺序为

$$\rho(A) = |\lambda_1| = \dots = |\lambda_r| > |\lambda_{r+1}| \geqslant \dots \geqslant |\lambda_n|,$$

因为 $\lambda_1, \dots, \lambda_r$ 对应的 Jordan 子块都是一阶的, 所以只要选取 $\varepsilon = \rho(A) - |\lambda_{r+1}| > 0$, 便有

$$\|A\|_* = \|D^{-1}JD\|_{\infty} = \rho(A).$$

□

定理 5.7 说明 A 的谱半径是 A 的所有矩阵范数的下确界.

二、数值范围

一个 n 阶复方阵 A 的数值范围(或值域)定义为

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

例如, 取

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

则 $W(\mathbf{A})$ 是闭区间 $[0, 1]$. 取

$$\mathbf{B} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix},$$

则 $W(\mathbf{B})$ 是一个焦点在 $(0, 0)$ 和 $(1, 0)$, 长半轴为 $\sqrt{2}$, 短半轴为 1 的椭圆盘.

关于数值范围, 有下面的 Toeplitz-Hausdorff 定理.

定理 5.8 (Toeplitz-Hausdorff) 一个方阵的数值范围是复平面的一个凸的紧子集.

其证明从略.

对于数值范围 $W(\mathbf{A})$, 定义

$$w(\mathbf{A}) = \sup\{|z| \mid z \in W(\mathbf{A})\},$$

称为 \mathbf{A} 的数值半径, 即

$$w(\mathbf{A}) = \sup_{\|\mathbf{x}\|=1} |\mathbf{x}^* \mathbf{A} \mathbf{x}|.$$

由定义立即知道, 对任意 $\mathbf{x} \in \mathbf{C}^n$, 都有

$$|\mathbf{x}^* \mathbf{A} \mathbf{x}| \leq w(\mathbf{A}) \|\mathbf{x}\|.$$

我们现在来比较一下数值半径 $w(\mathbf{A})$ 与谱半径 $\rho(\mathbf{A})$. 对方阵 \mathbf{A} 的最大的奇异值 $\sigma_{\max}(\mathbf{A})$, 容易看到

$$\sigma_{\max}(\mathbf{A}) = \sqrt{\rho(\mathbf{A}^* \mathbf{A})} = \sup_{\|\mathbf{x}\|=1} \|\mathbf{A} \mathbf{x}\| = \sup_{\mathbf{x} \neq 0} \frac{\|\mathbf{A} \mathbf{x}\|}{\|\mathbf{x}\|}.$$

于是对任意 $\mathbf{x} \in \mathbf{C}^n$, 总有

$$\|\mathbf{A} \mathbf{x}\| \leq \sigma_{\max}(\mathbf{A}) \|\mathbf{x}\|.$$

谱半径 $\rho(\mathbf{A})$, 数值半径 $w(\mathbf{A})$ 与最大的奇异值 $\sigma_{\max}(\mathbf{A})$ (也称为谱模) 之间有如下关系.

定理 5.9 设 \mathbf{A} 是一个复方阵, 则

$$\rho(\mathbf{A}) \leq w(\mathbf{A}) \leq \sigma_{\max}(\mathbf{A}) \leq 2w(\mathbf{A}).$$

* 证明 令 λ 是 \mathbf{A} 的满足 $\rho(\mathbf{A}) = |\lambda|$ 的一个特征值, \mathbf{u} 是 \mathbf{A} 的属于特征值 λ 的一个单位特征向量, 则

$$\rho(\mathbf{A}) = |\mathbf{u}^* \mathbf{A} \mathbf{u}| \leq w(\mathbf{A}).$$

由 Cauchy-Schwarz 不等式

$$|\mathbf{x}^* \mathbf{A} \mathbf{x}| = |(\mathbf{A} \mathbf{x}, \mathbf{x})| \leq \|\mathbf{A} \mathbf{x}\| \|\mathbf{x}\|,$$

有

$$w(\mathbf{A}) \leq \sigma_{\max}(\mathbf{A}).$$

最后我们证明

$$\sigma_{\max}(\mathbf{A}) \leq 2w(\mathbf{A}).$$

因为

$$\begin{aligned} 4(\mathbf{A} \mathbf{x}, \mathbf{y}) &= (\mathbf{A}(\mathbf{x} + \mathbf{y}), \mathbf{x} + \mathbf{y}) - (\mathbf{A}(\mathbf{x} - \mathbf{y}), \mathbf{x} - \mathbf{y}) \\ &\quad + i(\mathbf{A}(\mathbf{x} + i\mathbf{y}), \mathbf{x} + i\mathbf{y}) - i(\mathbf{A}(\mathbf{x} - i\mathbf{y}), \mathbf{x} - i\mathbf{y}), \end{aligned}$$

再利用 $|x^* Ax| \leq w(A) \|x\|^2$, 有

$$\begin{aligned} |(Ax, y)| &\leq w(A)(\|x+y\|^2 + \|x-y\|^2 + \|x+iy\|^2 + \|x-iy\|^2) \\ &= 4w(A)(\|x\|^2 + \|y\|^2). \end{aligned}$$

因此, 对任意 $x, y \in \mathbf{C}^n$, $\|x\| = \|y\| = 1$, 有

$$|(Ax, y)| \leq 2w(A).$$

注意到 $\sigma_{\max}(A) = \sup_{\|x\|=1} |(Ax, y)|$, 有

$$\sigma_{\max}(A) \leq 2w(A).$$

三、矩阵幂级数

定义 5.12 设 $A \in \mathbf{C}^{n \times n}$, $a_k \in \mathbf{C}$, $k=0, 1, 2, \dots$ 称

$$a_0 I + a_1 A + a_2 A^2 + \dots + a_k A^k + \dots$$

为矩阵 A 的幂级数, 记为 $\sum_{k=0}^{\infty} a_k A^k$.

定义 5.13 矩阵幂级数 $\sum_{k=0}^{\infty} a_k A^k$ 的前 $N+1$ 项的和 $S_N(A) = \sum_{k=0}^N a_k A^k$ 称为矩阵幂级数的部分和. 若矩阵幂级数 $\sum_{k=0}^{\infty} a_k A^k$ 的部分和序列 $\{S_N(A)\}$ 收敛, 则称 $\sum_{k=0}^{\infty} a_k A^k$ 收敛; 否则, 称其为发散. 若 $\lim_{N \rightarrow \infty} S_N(A) = S$, 则称 S 为 $\sum_{k=0}^{\infty} a_k A^k$ 的和矩阵.

给定复变量 z 的幂级数 $\sum_{k=0}^{\infty} a_k z^k$, 究竟有哪些方阵 A 使 $\sum_{k=0}^{\infty} a_k A^k$, 收敛? 这个问题既与 $\sum_{k=0}^{\infty} a_k z^k$ 的收敛半径有关, 又与 A 的谱半径有关.

定理 5.10 若复变量 z 的幂级数 $\sum_{k=0}^{\infty} a_k z^k$ 的收敛半径为 R , 而方阵 $A \in \mathbf{C}^{n \times n}$ 的谱半径为 $\rho(A)$, 则

(1) 当 $\rho(A) < R$ 时, 矩阵幂级数 $\sum_{k=0}^{\infty} a_k A^k$ 收敛;

(2) 当 $\rho(A) > R$ 时, 矩阵幂级数 $\sum_{k=0}^{\infty} a_k A^k$ 发散.

证明 对 $A \in \mathbf{C}^{n \times n}$, 存在可逆阵 P , 使得

$$A = PJP^{-1} = P \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_m \end{pmatrix} P^{-1},$$

其中

$$J_i = \begin{pmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & & 1 \\ & & & & \lambda_i \end{pmatrix}, \quad i = 1, 2, \dots, m.$$

$$\begin{aligned} S_N(\mathbf{A}) &= \sum_{k=0}^N a_k \mathbf{A}^k = \mathbf{P} S_N(\mathbf{J}) \mathbf{P}^{-1} \\ &= \mathbf{P} \begin{pmatrix} S_N(\mathbf{J}_1) & & & \\ & \ddots & & \\ & & S_N(\mathbf{J}_m) & \end{pmatrix} \mathbf{P}^{-1}, \end{aligned}$$

因此,矩阵序列 $\{S_N(\mathbf{A})\}$ 收敛当且仅当 m 个矩阵序列 $\{S_N(\mathbf{J}_i)\}$ 收敛, $i=1, 2, \dots, m$. 而

$$S_N(\mathbf{J}_i) = \begin{pmatrix} S_N(\lambda_i) & S'_N(\lambda_i) & \frac{1}{2} S''_N(\lambda_i) & \cdots & \frac{1}{(n_i - 1)!} S^{(n_i-1)}_N(\lambda_i) \\ & \ddots & & & \cdots \\ S_N(\lambda_i) & S'_N(\lambda_i) & & \frac{1}{2} S''_N(\lambda_i) & \\ & \ddots & & & S'_N(\lambda_i) \\ & & & & S_N(\lambda_i) \end{pmatrix},$$

其中 $S_N^{(t)}(\lambda)$ 表示 $S_N(\lambda)$ 在 $\lambda=\lambda_i$ 处的 t 阶导数, n_i 是 \mathbf{J}_i 的阶数.

(1) 若 $\rho(\mathbf{A}) < R$, 则 $|\lambda_i| < R$, 此时

$$\{S_N(\lambda_i)\}, \{S'_N(\lambda_i)\}, \dots, \{S^{(n_i-1)}_N(\lambda_i)\}$$

皆收敛, 从而 $\{S_N(\mathbf{J}_i)\}$ 收敛, 故 $\sum_{k=0}^{\infty} a_k \mathbf{A}^k$ 收敛.

(2) 若 $\rho(\mathbf{A}) > R$, 则存在某个特征值 λ_0 , 使 $|\lambda_0| > R$, 于是幂级数 $\sum_{k=0}^{\infty} a_k \lambda_0^k$ 发散, 从而

$\{S_N(\mathbf{J}_i)\}$ 发散, 故 $\sum_{k=0}^{\infty} a_k \mathbf{A}^k$ 发散.

例 11 讨论矩阵幂级数

$$\sum_{k=0}^{\infty} \mathbf{A}^k = \mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^k + \cdots$$

的敛散性. 当它收敛时, 求它的和矩阵.

解 复变量 z 的幂级数 $\sum_{k=0}^{\infty} z^k$ 的收敛半径 $R=1$, 故当 $\rho(\mathbf{A}) < 1$, 即 \mathbf{A} 的所有特征值的模小于 1 时, 矩阵幂级数 $\sum_{k=0}^{\infty} \mathbf{A}^k$ 收敛.

当 $\rho(\mathbf{A}) < 1$ 时, 1 不是 \mathbf{A} 的特征值, $|\mathbf{I} - \mathbf{A}| \neq 0$, 故 $\mathbf{I} - \mathbf{A}$ 可逆. 又

$$S_N(\mathbf{A}) = \mathbf{I} + \mathbf{A} + \cdots + \mathbf{A}^N,$$

$$\mathbf{I} - \mathbf{A}^{N+1} = (\mathbf{I} - \mathbf{A})(\mathbf{I} + \mathbf{A} + \cdots + \mathbf{A}^N) = (\mathbf{I} - \mathbf{A})S_N(\mathbf{A}),$$

于是

$$\lim_{N \rightarrow \infty} (\mathbf{I} - \mathbf{A})S_N(\mathbf{A}) = \lim_{N \rightarrow \infty} (\mathbf{I} - \mathbf{A}^{N+1}) = \mathbf{I},$$

故

$$S = \lim_{N \rightarrow \infty} S_N(\mathbf{A}) = (\mathbf{I} - \mathbf{A})^{-1}.$$

当计算 \mathbf{A} 的特征值比较困难时,由定理 5.6 知 \mathbf{A} 的每个范数都是谱半径 $\rho(\mathbf{A})$ 的上界,只要能找到一种特殊的矩阵范数 $\|\mathbf{A}\|$,使 $\|\mathbf{A}\| < R$,便可断定该矩阵幂级数是收敛的.

例如,对矩阵

$$\mathbf{A} = \begin{pmatrix} 0.2 & 0.1 & 0.2 \\ 0.5 & 0.5 & 0.4 \\ 0.1 & 0.3 & 0.2 \end{pmatrix},$$

由 $\|\mathbf{A}\|_1 = 0.9$, $\rho(\mathbf{A}) \leq \|\mathbf{A}\|_1 < 1$ 知 $\sum_{k=0}^{\infty} \mathbf{A}^k$ 收敛.

当矩阵的谱半径 $\rho(\mathbf{A})$ 等于幂级数的收敛半径 R 时,定理 5.10 的判定方法失效,须单独考虑.

例 12 讨论矩阵幂级数 $\sum_{k=1}^{\infty} \frac{1}{k^2} \mathbf{A}^k$ 的敛散性,其中

$$\mathbf{A} = \begin{pmatrix} 1 & 4 \\ -1 & -3 \end{pmatrix}.$$

解 矩阵 \mathbf{A} 的特征值为 $\lambda_1 = \lambda_2 = -1$, $\rho(\mathbf{A}) = 1$, 收敛半径 $R = 1$. 不能按定理 5.10 的判定方法来判断. 直接计算知 \mathbf{A} 不能对角化,于是,存在可逆阵 \mathbf{P} ,使得

$$\mathbf{P}^{-1} \mathbf{A} \mathbf{P} = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}.$$

而

$$\mathbf{A}^k = \mathbf{P} \begin{pmatrix} (-1)^k & (-1)^{k-1} k \\ 0 & (-1)^k \end{pmatrix} \mathbf{P}^{-1},$$

$$\sum_{k=1}^{\infty} \frac{1}{k^2} \mathbf{A}^k = \mathbf{P} \begin{pmatrix} \sum_{k=1}^{\infty} \frac{(-1)^k}{k^2} & \sum_{k=1}^{\infty} \frac{(-1)^{k-1} k}{k} \\ 0 & \sum_{k=1}^{\infty} \frac{(-1)^k}{k^2} \end{pmatrix} \mathbf{P}^{-1},$$

由于 $\sum_{k=1}^{\infty} \frac{(-1)^k}{k^2}$ 与 $\sum_{k=1}^{\infty} \frac{(-1)^{k-1} k}{k}$ 均收敛,因此 $\sum_{k=1}^{\infty} \frac{1}{k^2} \mathbf{A}^k$ 收敛.

5.5 矩阵函数

矩阵函数的概念与通常的函数概念一样,它是以 n 阶矩阵为自变量和函数值(因变量)的一种函数.本节利用矩阵幂级数定义矩阵函数,并讨论矩阵函数的有关性质与计算方法.

一、矩阵函数的定义与性质

定义 5.14 设 $f(z)$ 是复变量的解析函数, $f(z) = \sum_{k=0}^{\infty} a_k z^k$ 的收敛半径为 R . 如果矩阵 $A \in \mathbf{C}^{n \times n}$ 的谱半径 $\rho(A) < R$, 则称

$$f(A) = \sum_{k=0}^{\infty} a_k A^k$$

为 A 的矩阵函数.

例如, 函数

$$\begin{aligned} e^z &= 1 + \frac{z}{1} + \frac{z^2}{2} + \cdots + \frac{z^k}{k} + \cdots \\ \cos z &= 1 - \frac{z^2}{2} + \frac{z^4}{4} - \cdots + (-1)^k \frac{z^{2k}}{(2k)} + \cdots \\ \sin z &= z - \frac{z^3}{3} + \frac{z^5}{5} - \cdots + (-1)^k \frac{z^{2k+1}}{(2k+1)} + \cdots \end{aligned}$$

在整个复平面上都是收敛的, 于是无论 $A \in \mathbf{C}^{n \times n}$ 是任何矩阵, 总有

$$\begin{aligned} e^A &= \sum_{k=0}^{\infty} \frac{1}{k} A^k, \quad A \in \mathbf{C}^{n \times n}; \\ \cos A &= \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)} A^{2k}, \quad A \in \mathbf{C}^{n \times n}; \\ \sin A &= \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)} A^{2k+1}, \quad A \in \mathbf{C}^{n \times n} \end{aligned}$$

分别称之为矩阵 A 的指数函数、余弦函数及正弦函数. 同样地, 由

$$\begin{aligned} \frac{1}{1-z} &= 1 + z + z^2 + \cdots + z^k + \cdots \\ \ln(1+z) &= z - \frac{z^2}{2} + \frac{z^3}{3} - \cdots + (-1)^{k-1} \frac{z^k}{k} + \cdots \end{aligned}$$

在复平面 $|z| < 1$ 内是收敛的, 于是对 $A \in \mathbf{C}^{n \times n}$, 当 $\rho(A) < 1$ 时, 总有

$$\begin{aligned} (I-A)^{-1} &= \sum_{k=0}^{\infty} A^k, \quad \rho(A) < 1; \\ \ln(I+A) &= \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} A^k, \quad \rho(A) < 1. \end{aligned}$$

值得注意的是, 在数学分析中指数函数具有运算规则 $e^{z_1} e^{z_2} = e^{z_2} e^{z_1} = e^{z_1 + z_2}$, 但是在矩阵分析中 $e^A e^B = e^B e^A = e^{A+B}$ 一般不再成立. 例如, 令

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

则 $A^2 = A, B^2 = B$. 从而

$$\mathbf{A} = \mathbf{A}^2 = \mathbf{A}^3 = \dots, \quad \mathbf{B} = \mathbf{B}^2 = \mathbf{B}^3 = \dots$$

于是

$$\mathbf{e}^{\mathbf{A}} = \mathbf{I} + (\mathrm{e} - 1)\mathbf{A} = \begin{pmatrix} \mathrm{e} & \mathrm{e} - 1 \\ 0 & 1 \end{pmatrix},$$

$$\mathbf{e}^{\mathbf{B}} = \mathbf{I} + (\mathrm{e} - 1)\mathbf{B} = \begin{pmatrix} \mathrm{e} & 1 - \mathrm{e} \\ 0 & 1 \end{pmatrix},$$

$$\mathbf{e}^{\mathbf{A}} \mathbf{e}^{\mathbf{B}} = \begin{pmatrix} \mathrm{e}^2 & -(\mathrm{e} - 1)^2 \\ 0 & 1 \end{pmatrix},$$

$$\mathbf{e}^{\mathbf{B}} \mathbf{e}^{\mathbf{A}} = \begin{pmatrix} \mathrm{e}^2 & (\mathrm{e} - 1)^2 \\ 0 & 1 \end{pmatrix}.$$

又由

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix},$$

可得 $(\mathbf{A} + \mathbf{B})^2 = 2(\mathbf{A} + \mathbf{B})$, 于是 $(\mathbf{A} + \mathbf{B})^k = 2^{k-1}(\mathbf{A} + \mathbf{B})$, $k = 1, 2, \dots$ 由此容易推出

$$\mathbf{e}^{\mathbf{A} + \mathbf{B}} = \mathbf{I} + \frac{1}{2}(\mathrm{e}^2 - 1)(\mathbf{A} + \mathbf{B}) = \begin{pmatrix} \mathrm{e}^2 & 0 \\ 0 & 1 \end{pmatrix}.$$

可见 $\mathbf{e}^{\mathbf{A}} \mathbf{e}^{\mathbf{B}}$, $\mathbf{e}^{\mathbf{B}} \mathbf{e}^{\mathbf{A}}$ 及 $\mathbf{e}^{\mathbf{A} + \mathbf{B}}$ 两两不等.

如果矩阵 \mathbf{A} 和 \mathbf{B} 可交换, 则有 $\mathbf{e}^{\mathbf{A}} \mathbf{e}^{\mathbf{B}} = \mathbf{e}^{\mathbf{B}} \mathbf{e}^{\mathbf{A}} = \mathbf{e}^{\mathbf{A} + \mathbf{B}}$.

事实上,

$$\begin{aligned} \mathbf{e}^{\mathbf{A}} \mathbf{e}^{\mathbf{B}} &= \left(\sum_{k=0}^{\infty} \frac{1}{k} \mathbf{A}^k \right) \left(\sum_{k=0}^{\infty} \frac{1}{k} \mathbf{B}^k \right) \\ &= \mathbf{I} + (\mathbf{A} + \mathbf{B}) + \left(\frac{1}{2} \mathbf{A}^2 + \mathbf{AB} + \frac{1}{2} \mathbf{B}^2 \right) + \dots \end{aligned}$$

$$\begin{aligned} \mathbf{e}^{\mathbf{A} + \mathbf{B}} &= \sum_{k=0}^{\infty} \frac{1}{k} (\mathbf{A} + \mathbf{B})^k \\ &= \mathbf{I} + (\mathbf{A} + \mathbf{B}) + \frac{1}{2} (\mathbf{A} + \mathbf{B})^2 + \dots \end{aligned}$$

因为 $\mathbf{AB} = \mathbf{BA}$, 所以 $\mathbf{e}^{\mathbf{A}} \mathbf{e}^{\mathbf{B}} = \mathbf{e}^{\mathbf{A} + \mathbf{B}}$, 同理可证 $\mathbf{e}^{\mathbf{B}} \mathbf{e}^{\mathbf{A}} = \mathbf{e}^{\mathbf{A} + \mathbf{B}}$.

如果矩阵 \mathbf{A} 和 \mathbf{B} 可交换, 那么许多三角恒等式也可以推广到矩阵上, 例如

$$\sin(\mathbf{A} + \mathbf{B}) = \sin \mathbf{A} \cos \mathbf{B} + \cos \mathbf{A} \sin \mathbf{B},$$

$$\sin 2\mathbf{A} = 2 \sin \mathbf{A} \cos \mathbf{A}.$$

二、矩阵函数的求法

在实际应用中常常是给定了解析函数 $f(z)$ 和矩阵 $\mathbf{A} \in \mathbf{C}^{n \times n}$, 要求计算矩阵 $f(\mathbf{A}) \in \mathbf{C}^{n \times n}$. 下面介绍两种有效的计算方法. 在计算 $f(\mathbf{A}) = \sum_{k=0}^{\infty} a_k \mathbf{A}^k$ 时, 总假定 \mathbf{A} 的谱半径小于幂级数的收敛半径.

1. Jordan 标准形法

定理 5.11 设 $f(z)$ 是复变量 z 的解析函数, $\mathbf{A} \in \mathbf{C}^{n \times n}$, 且存在可逆阵 \mathbf{P} , 使得

$$A = PJP^{-1} = P\text{diag}(J_1, J_2, \dots, J_m)P^{-1},$$

$$\text{则 } f(A) = Pf(J)P^{-1} = P\text{diag}(f(J_1), f(J_2), \dots, f(J_m))P^{-1}.$$

其中

$$f(J_i) = \begin{pmatrix} f(\lambda_i) & f'(\lambda_i) & \frac{1}{2} f''(\lambda_i) & \cdots & \frac{1}{(n_i-1)} f^{(n_i-1)}(\lambda_i) \\ f(\lambda_i) & f'(\lambda_i) & & & \ddots \\ & & \ddots & & \cdots \\ & & & \frac{1}{2} f''(\lambda_i) & \\ & & & f'(\lambda_i) & \\ & & & f(\lambda_i) & \end{pmatrix}_{n_i \times n_i},$$

其证明只需将定理 5.10 中有关等式取极限即得.

推论 若 A 的特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 则 $f(A)$ 的特征值为 $f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n)$.

例 13 已知矩阵

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

计算 e^A 和 $\sin A$.

解 A 的 Jordan 标准形为

$$J = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix},$$

变换矩阵 P 和 P^{-1} 分别为

$$P = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix}, \quad P^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix},$$

$$\text{且 } f(J) = \begin{pmatrix} f(2) & f'(2) & 0 \\ 0 & f(2) & 0 \\ 0 & 0 & f(2) \end{pmatrix},$$

故

$$\begin{aligned} f(A) &= Pf(J)P^{-1} \\ &= \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} f(2) & f'(2) & 0 \\ 0 & f(2) & 0 \\ 0 & 0 & f(2) \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} f(2) & 0 & 0 \\ f'(2) & f(2) - f'(2) & f'(2) \\ f'(2) & -f'(2) & f(2) + f'(2) \end{pmatrix}. \end{aligned}$$

当 $f(z) = e^z$ 时, $f(2) = e^2$, $f'(2) = e^2$, 故

$$e^A = \begin{pmatrix} e^2 & 0 & 0 \\ e^2 & 0 & e^2 \\ e^2 & -e^2 & 2e^2 \end{pmatrix}.$$

当 $f(z) = \sin z$ 时, $f(2) = \sin 2$, $f'(2) = \cos 2$, 故

$$\sin A = \begin{pmatrix} \sin 2 & 0 & 0 \\ \cos 2 & \sin 2 - \cos 2 & \cos 2 \\ \cos 2 & -\cos 2 & \sin 2 + \cos 2 \end{pmatrix}.$$

定理 5.11 提供了计算矩阵函数的一种方法, 用这种方法须先算出 Jordan 标准形 J 和变换矩阵 P . 从所举例子看到, 对一个比较简单的矩阵 A , 即使只要求出最简单的矩阵函数 e^A 、 $\sin A$, 其过程也还是比较繁琐的. 下面介绍的最小多项式法, 比起上述方法来要简便些. 但是, 这一方法的理论推导相当繁琐, 所以这里不作推导, 我们只将结果介绍给读者.

2. 最小多项式法

设 $m_A(\lambda)$ 是 n 阶矩阵 A 的最小多项式, 它的次数为 m , 若 $f(\lambda)$ 是 $l(l \geq m)$ 次多项式, 以 $m_A(\lambda)$ 去除 $f(\lambda)$ 即得

$$f(\lambda) = m_A(\lambda)q(\lambda) + r(\lambda),$$

余式 $r(\lambda) = 0$ 或 $r(\lambda)$ 的次数低于 $m_A(\lambda)$ 的次数. 因此

$$f(A) = m_A(A)q(A) + r(A) = r(A).$$

由此可见, 次数大于等于 m 的任意矩阵多项式 $f(A)$ 都可以化为次数小于等于 $m-1$ 的 A 的多项式 $r(A)$ 来计算.

把这一思想推广到由矩阵幂级数确定的矩阵函数 $f(A)$ 上, 我们有如下的定理.

定理 5.12 设 n 阶矩阵 A 的最小多项式为

$$m_A(\lambda) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_s)^{n_s},$$

其中 $\lambda_1, \lambda_2, \dots, \lambda_s$ 为 A 的所有不同特征值, $\sum_{i=1}^s n_i = m$, $f(\lambda)$ 是复变量 λ 的解析函数, 令

$$g(\lambda) = c_0 + c_1 \lambda + \cdots + c_{m-1} \lambda^{m-1},$$

则 $f(A) = g(A)$ 的充分必要条件是

$$g^{(j)}(\lambda_i) = f^{(j)}(\lambda_i), \quad i = 1, 2, \dots, s, \quad j = 0, 1, 2, \dots, n_i - 1. \quad (5.1)$$

我们用例子说明这一方法.

例 14 用定理 5.12 提供的方法, 对例 10 的矩阵 A 计算 e^A , $\sin A$, e^{At} .

解 A 的最小多项式为 $m_A(\lambda) = (\lambda - 2)^2$, 令

$$\begin{aligned} g(\lambda) &= c_0 + c_1 \lambda, \\ \begin{cases} f(2) = c_0 + 2c_1, \\ f'(2) = c_1. \end{cases} \end{aligned}$$

则

解得 $c_0 = f(2) - 2f'(2)$, $c_1 = f'(2)$, 故

$$f(\mathbf{A}) = c_0 \mathbf{I} + c_1 \mathbf{A} = \begin{pmatrix} f(2) & 0 & 0 \\ f'(2) & f(2) - f'(2) & f'(2) \\ f'(2) & -f'(2) & f(2) + f'(2) \end{pmatrix}.$$

当 $f(z) = e^z$ 时, $f(2) = e^2$, $f'(2) = e^2$, 故

$$e^{\mathbf{A}} = \begin{pmatrix} e^2 & 0 & 0 \\ e^2 & 0 & e^2 \\ e^2 & -e^2 & 2e^2 \end{pmatrix}.$$

当 $f(z) = \sin z$ 时, $f(2) = \sin 2$, $f'(2) = \cos 2$, 故

$$\sin \mathbf{A} = \begin{pmatrix} \sin 2 & 0 & 0 \\ \cos 2 & \sin 2 - \cos 2 & \cos 2 \\ \cos 2 & -\cos 2 & \sin 2 + \cos 2 \end{pmatrix}.$$

当 $f(z) = e^{iz}$ 时, $f(2) = e^{2i}$, $f'(2) = te^{2i}$, 故

$$e^{A_t} = \begin{pmatrix} e^{2t} & 0 & 0 \\ te^{2t} & e^{2t} - te^{2t} & te^{2t} \\ te^{2t} & -te^{2t} & e^{2t} + te^{2t} \end{pmatrix}.$$

例 14 中, 求 $e^{\mathbf{A}}$ 应注意 t 只是一个参数, 求导时 t 看成一个常量.

例 15 已知矩阵 $\mathbf{A} = \begin{pmatrix} 0 & 0 & -2 \\ 0 & 1 & 0 \\ 1 & 0 & 3 \end{pmatrix}$, 计算 $e^{\mathbf{A}_t}$.

解 \mathbf{A} 的最小多项式为 $m_{\mathbf{A}}(\lambda) = (\lambda - 1)(\lambda - 2)$, 令 $g(\lambda) = c_0 + c_1 \lambda$, 则

$$\begin{cases} c_0 + c_1 = e^t, \\ c_0 + 2c_1 = e^{2t}. \end{cases}$$

$$c_0 = 2e^t - e^{2t}, \quad c_1 = e^{2t} - e^t,$$

$$\text{故 } e^{\mathbf{A}_t} = c_0 \mathbf{I} + c_1 \mathbf{A} = \begin{pmatrix} 2e^t - e^{2t} & 0 & 2e^t - 2e^{2t} \\ 0 & e^t & 0 \\ e^{2t} - e^t & 0 & 2e^{2t} - e^t \end{pmatrix}.$$

当计算最小多项式比较困难时, 我们也可以用特征多项式来代替最小多项式. 读者可以对例 15 的矩阵 \mathbf{A} , 计算 $e^{\mathbf{A}_t}$.

5.6 函数矩阵的微分与积分

在实际应用中, 矩阵函数与函数矩阵的微分、积分常常是同时出现的, 因此我们学习了矩阵函数的计算以后, 还须学习函数矩阵的微分与积分. 比起前者来, 后者要简单、容易得多, 它只是一般微积分概念法则的一种形式的推广.

现在考虑矩阵元素是实变量 t 的实函数的矩阵

$$\mathbf{A}(t) = \begin{pmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t) \\ a_{21}(t) & a_{22}(t) & \cdots & a_{2n}(t) \\ \cdots & \cdots & & \cdots \\ a_{m1}(t) & a_{m2}(t) & \cdots & a_{mn}(t) \end{pmatrix},$$

$\mathbf{A}(t)$ 中所有的元素 $a_{ij}(t)$ 定义在同一区间 $[a, b]$ 上.

$\mathbf{A}(t)$ 在区间 $[a, b]$ 上有界, 连续, 可微, 可积是指所有的 $a_{ij}(t)$ 在 $[a, b]$ 上有界, 连续, 可微, 可积. 例如函数矩阵 $\mathbf{A}(t)$ 的微分与积分定义为

$$\frac{d}{dt}\mathbf{A}(t) = \left[\frac{d}{dt}a_{ij}(t) \right]_{m \times n},$$

$$\int_{t_0}^t \mathbf{A}(\tau) d\tau = \left[\int_{t_0}^t a_{ij}(\tau) d\tau \right]_{m \times n}.$$

函数矩阵的微分满足以下性质:

- (1) $\frac{d}{dt}(\mathbf{A}(t) + \mathbf{B}(t)) = \frac{d}{dt}\mathbf{A}(t) + \frac{d}{dt}\mathbf{B}(t);$
- (2) $\frac{d}{dt}(k(t)\mathbf{A}(t)) = \frac{d k(t)}{dt}\mathbf{A}(t) + k(t)\frac{d}{dt}\mathbf{A}(t)$, 其中 $k(t)$ 为 t 的可微函数;
- (3) $\frac{d}{dt}(\mathbf{A}(t)\mathbf{B}(t)) = \frac{d}{dt}\mathbf{A}(t) \cdot \mathbf{B}(t) + \mathbf{A}(t) \cdot \frac{d}{dt}\mathbf{B}(t);$
- (4) 若 $\mathbf{A}(t)$ 与 $\mathbf{A}^{-1}(t)$ 都可微, 则

$$\frac{d}{dt}\mathbf{A}^{-1}(t) = -\mathbf{A}^{-1}(t) \left(\frac{d}{dt}\mathbf{A}(t) \right) \mathbf{A}^{-1}(t).$$

一般说来, $\frac{d}{dt}\mathbf{A}^2(t) \neq 2\mathbf{A}(t)\frac{d}{dt}\mathbf{A}(t)$, 因为 $\mathbf{A}(t)$ 和 $\frac{d}{dt}\mathbf{A}(t)$ 一般不可交换.

例 16 设 n 阶矩阵 \mathbf{A} 与 t 无关, 证明:

- (1) $\frac{d}{dt}e^{\mathbf{A}} = \mathbf{A}e^{\mathbf{A}} = e^{\mathbf{A}}\mathbf{A};$
- (2) $\frac{d}{dt}\cos(t\mathbf{A}) = -\mathbf{A}\sin t\mathbf{A} = -(\sin t\mathbf{A})\mathbf{A};$
- (3) $\frac{d}{dt}(\sin t\mathbf{A}) = \mathbf{A}\cos t\mathbf{A} = (\cos t\mathbf{A})\mathbf{A}.$

证明 只证(1), 其余留给读者.

注意到 $e^{\mathbf{A}}$ 的第 i 行第 j 列的元素为

$$(e^{\mathbf{A}})_{ij} = \sum_{k=0}^{\infty} \frac{1}{k!} t^k (\mathbf{A}^k)_{ij},$$

上式右边是 t 的幂级数, 不论 t 的取值, 它总是收敛的. 因此, 可以对其逐项微分

$$\frac{d}{dt}(e^{\mathbf{A}})_{ij} = \sum_{k=1}^{\infty} \frac{1}{(k-1)!} t^{k-1} (\mathbf{A}^k)_{ij},$$

于是

$$\frac{d}{dt} e^A = \sum_{k=1}^{\infty} \frac{1}{(k-1)!} t^{k-1} A^k = A \sum_{k=1}^{\infty} \frac{1}{(k-1)!} t^{k-1} A^{k-1} = A e^A.$$

同样,

$$\frac{d}{dt} e^A = \left[\sum_{k=1}^{\infty} \frac{1}{(k-1)!} t^{k-1} A^{k-1} \right] A = e^A A.$$

函数矩阵的积分满足以下性质:

$$(1) \int [a\mathbf{A}(t) + b\mathbf{B}(t)] dt = a \int \mathbf{A}(t) dt + b \int \mathbf{B}(t) dt (a, b \text{ 为任意实数});$$

$$(2) \int \mathbf{C} \cdot \mathbf{A}(t) dt = \mathbf{C} \int \mathbf{A}(t) dt (\mathbf{C} \text{ 为常数矩阵});$$

$$(3) \int \mathbf{A}(t) \mathbf{B}'(t) dt = \mathbf{A}(t) \mathbf{B}(t) - \int \mathbf{A}'(t) \mathbf{B}(t) dt (\mathbf{B}'(t) \text{ 表示 } \mathbf{B}(t) \text{ 的导数}).$$

上面所列性质都可直接验证,这里不再给出.顺便指出,上述函数矩阵的微分、积分概念,还可以作些推广,比如可定义 $\mathbf{A}(t)$ 的广义积分与拉氏变换等.但正如以上所讲的情况一样,这仅仅是一种约定,实质性的东西并不多,如在应用中遇到,可按所给出的定义去理解,这里不再赘述.

5.7 矩阵函数的应用

在线性控制系统中,常常涉及求解线性微分方程组的问题.矩阵函数在其中有着重要的应用.我们要求解形如:

$$\begin{cases} \dot{x}_1(t) = a_{11}(t)x_1(t) + \dots + a_{1n}(t)x_n(t) + f_1(t), \\ \dot{x}_2(t) = a_{21}(t)x_1(t) + \dots + a_{2n}(t)x_n(t) + f_2(t), \\ \dots \\ \dot{x}_n(t) = a_{n1}(t)x_1(t) + \dots + a_{nn}(t)x_n(t) + f_n(t) \end{cases}$$

的一阶线性常微分方程组,在应用中,牵涉得最多的是求解微分方程组的初值问题,即寻求一组函数 $x_1(t), x_2(t), \dots, x_n(t)$,使它们满足方程组和下列一组初始条件:

$$x_1(t_0) = c_1, x_2(t_0) = c_2, \dots, x_n(t_0) = c_n.$$

为了应用矩阵函数求解上述微分方程的初值问题,首先用矩阵表示方程组和初始条件.令 $\mathbf{A}(t) = (a_{ij}(t))_{n \times n}, \mathbf{f}(t) = (f_1(t) \quad f_2(t) \quad \dots \quad f_n(t))^T, \mathbf{x}(t) = (x_1(t) \quad x_2(t) \quad \dots \quad x_n(t))^T, \mathbf{x}(t_0) = (c_1 \quad c_2 \quad \dots \quad c_n)^T = \mathbf{C}_{n \times 1}$.于是,方程组和初始条件的矩阵形式是

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{f}(t), \\ \mathbf{x}(t_0) = \mathbf{C}_{n \times 1}. \end{cases}$$

设所考虑的时间区间为 $[t_0, +\infty)$,若在该区间上 $\mathbf{f}(t) = \mathbf{0}$,称方程组是齐次的,否

则称为非齐次的;如果系数矩阵 $A(t)=A$ 是常数矩阵,称方程组是常系数的,否则称为变系数的.本节只讨论一阶常系数微分方程组的求解问题.

一、一阶线性常系数齐次微分方程组

定理 5.13 一阶线性常系数齐次微分方程组

$$\begin{cases} \dot{\mathbf{x}}(t) = A\mathbf{x}(t), \\ \mathbf{x}(t_0) = \mathbf{C}_{n \times 1} \end{cases}$$

的解为

$$\mathbf{x}(t) = e^{A(t-t_0)} \mathbf{x}(t_0).$$

证明 根据函数矩阵的微分法则

$$\begin{aligned} \frac{d}{dt}(e^{-At}\mathbf{x}(t)) &= e^{-At}(-A)\mathbf{x}(t) + e^{-At}\dot{\mathbf{x}}(t) \\ &= e^{-At}(\dot{\mathbf{x}}(t) - A\mathbf{x}(t)) = \mathbf{0}, \end{aligned}$$

将上式两端在 $[t_0, t]$ 上积分,得

$$e^{-At}\mathbf{x}(t) - e^{-At_0}\mathbf{x}(t_0) = \mathbf{0},$$

所以

$$\mathbf{x}(t) = e^{A(t-t_0)} \mathbf{x}(t_0).$$

例 17 设 $A = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & 3 \end{pmatrix}$, 求微分方程组 $\dot{\mathbf{x}}(t) = A\mathbf{x}(t)$ 满足初始条件 $\mathbf{x}(0) = (1 \ 1 \ 1)^T$ 的解.

解 由定理 5.13 知微分方程组满足初始条件的解为

$$\mathbf{x}(t) = e^{At} \mathbf{x}(0),$$

在例 11 中已经求出

$$e^{At} = e^{2t} \begin{pmatrix} 1 & 0 & 0 \\ t & 1-t & t \\ t & -t & 1+t \end{pmatrix},$$

所以解为

$$\mathbf{x}(t) = (e^{2t} \ (1+t)e^{2t} \ (1+t)e^{2t})^T.$$

二、一阶线性常系数非齐次微分方程组

定理 5.14 一阶线性常系数非齐次线性方程组

$$\begin{cases} \dot{\mathbf{x}}(t) = A\mathbf{x}(t) + \mathbf{f}(t), \\ \mathbf{x}(t_0) = \mathbf{C} \end{cases}$$

的解为

$$\mathbf{x}(t) = e^{A(t-t_0)} \mathbf{x}(t_0) + e^{At} \int_{t_0}^t e^{-A\tau} \mathbf{f}(\tau) d\tau$$

$$= e^{A(t-t_0)} \mathbf{x}(t_0) + \int_{t_0}^t e^{A(t-\tau)} \mathbf{f}(\tau) d\tau.$$

证明 其证明过程和定理 5.13 的类似. 根据函数矩阵的微分法则, 有

$$\begin{aligned} \frac{d}{dt}(e^{-At}\mathbf{x}(t)) &= e^{-At}(-A)\mathbf{x}(t) + e^{-At}\dot{\mathbf{x}}(t) \\ &= e^{-At}(\dot{\mathbf{x}}(t) - A\mathbf{x}(t)) \\ &= e^{-At}\mathbf{f}(t), \end{aligned}$$

将上式两端在 $[t_0, t]$ 上积分, 得

$$e^{-At}\mathbf{x}(t) - e^{-At_0}\mathbf{x}(t_0) = \int_{t_0}^t e^{-A\tau}\mathbf{f}(\tau)d\tau.$$

于是有

$$\mathbf{x}(t) = e^{A(t-t_0)}\mathbf{x}(t_0) + \int_{t_0}^t e^{A(t-\tau)}\mathbf{f}(\tau)d\tau.$$

例 18 设

$$\mathbf{A} = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & 3 \end{pmatrix},$$

$\mathbf{f}(t) = (e^{2t} \quad e^{2t} \quad 0)^T$. 求微分方程组 $\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{f}(t)$ 满足初始条件 $\mathbf{x}(0) = (-1 \quad 1 \quad 0)^T$ 的解.

解 由定理 5.14 知微分方程组满足初始条件的解为

$$\mathbf{x}(t) = e^{At}\mathbf{x}(0) + \int_0^t e^{A(t-\tau)}\mathbf{f}(\tau)d\tau,$$

在例 14 中已经求出

$$e^{At} = e^{2t} \begin{pmatrix} 1 & 0 & 0 \\ t & 1-t & t \\ t & -t & 1+t \end{pmatrix}.$$

$$\text{计算} \quad e^{-A\tau}\mathbf{f}(\tau) = e^{-2\tau} \begin{pmatrix} 1 & 0 & 0 \\ -\tau & 1+\tau & -\tau \\ -\tau & \tau & 1-\tau \end{pmatrix} \begin{pmatrix} e^{2\tau} \\ e^{2\tau} \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix},$$

$$\int_0^t e^{-A\tau}\mathbf{f}(\tau)d\tau = \begin{pmatrix} t \\ t \\ 0 \end{pmatrix},$$

所以解为

$$\mathbf{x}(t) = e^{At} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + e^{At} \begin{pmatrix} t \\ t \\ 0 \end{pmatrix} = e^{2t} \begin{pmatrix} 1 & 0 & 0 \\ t & 1-t & t \\ t & -t & 1+t \end{pmatrix} \begin{pmatrix} t-1 \\ t+1 \\ 0 \end{pmatrix} = \begin{pmatrix} (t-1)e^{2t} \\ (1-t)e^{2t} \\ -2te^{2t} \end{pmatrix}.$$

习 题 五

1. 设 a_1, a_2, \dots, a_n 都是正实数, $\mathbf{x} = (x_1 \quad x_2 \quad \dots \quad x_n)^T \in \mathbf{R}^n$, 证明由 $\|\mathbf{x}\| = (\sum_{i=1}^n a_i x_i^2)^{1/2}$ 定义的非负实值函数是 \mathbf{R}^n 空间的一个向量范数.

2. 证明: 若 $\mathbf{x} \in \mathbf{C}^n$, 则

- (1) $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{n} \|\mathbf{x}\|_2$;
- (2) $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1 \leq n \|\mathbf{x}\|_\infty$;
- (3) $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{n} \|\mathbf{x}\|_\infty$.

3. 已知矩阵

$$\mathbf{A} = \begin{pmatrix} -1 & 0 & 2 & i \\ 3 & 5 & -1 & 0 \\ 1 & 2 & 0 & -1 \\ 7 & -i & 2 & -4 \end{pmatrix}, \quad i^2 = -1.$$

计算 $\|\mathbf{A}\|_1$ 和 $\|\mathbf{A}\|_\infty$. 如果 $\mathbf{x} = (-1 \quad 2 \quad 0 \quad -i)^T$, 计算 $\|\mathbf{Ax}\|_1$ 和 $\|\mathbf{Ax}\|_\infty$.

4. 设 \mathbf{U} 是酉矩阵, 证明:

- (1) $\|\mathbf{U}\|_2 = 1$;
- (2) $\|\mathbf{U}\mathbf{A}\|_2 = \|\mathbf{A}\|_2$;
- (3) $\|\mathbf{U}^H \mathbf{A} \mathbf{U}\|_2 = \|\mathbf{A}\|_2$.

5. 设 $\|\mathbf{A}\|_p$ 是由向量范数 $\|\mathbf{x}\|_p$ 诱导的矩阵范数, \mathbf{A} 可逆, 证明:

- (1) $\|\mathbf{A}^{-1}\|_p \geq \|\mathbf{A}\|_p^{-1}$;
- (2) $\|\mathbf{A}^{-1}\|_p^{-1} = \min_{\mathbf{x} \neq 0} \frac{\|\mathbf{Ax}\|_p}{\|\mathbf{x}\|_p}$.

6. 设矩阵序列 $\{\mathbf{A}^{(k)}\}$ 收敛, 其极限为 \mathbf{A} . 证明: 对任意 n 阶可逆阵 \mathbf{P} , 矩阵序列 $\{\mathbf{P}^{-1} \mathbf{A}^{(k)} \mathbf{P}\}$ 也收敛, 且极限为 $\mathbf{P}^{-1} \mathbf{A} \mathbf{P}$.

7. 设三阶方阵 \mathbf{A} 的特征值为 $\lambda_1 = \lambda_2 = 1, \lambda_3 = \frac{1}{2}$, 对应的特征向量为 $\alpha_1 = (1 \quad 0 \quad 0)^T, \alpha_2 = (1 \quad 1 \quad 0)^T, \alpha_3 = (1 \quad 1 \quad 1)^T$, 求 $\lim_{k \rightarrow \infty} \mathbf{A}^k$.

8. 设 $\mathbf{A} = \begin{pmatrix} 0 & c & c \\ c & 0 & c \\ c & c & 0 \end{pmatrix}$, 讨论 c 取何值时, $\sum_{k=0}^{\infty} \mathbf{A}^k$ 收敛.

9. 设 λ 为矩阵 $\mathbf{A} \in \mathbf{C}^{n \times n}$ 的特征值. 证明: 对任意正整数 k 及矩阵范数 $\|\mathbf{A}\|$, 都有

$$|\lambda| \leq \|\mathbf{A}^k\|^{\frac{1}{k}}.$$

10. 设 \mathbf{A} 是正规矩阵. 证明:

$$\rho(\mathbf{A}) = \|\mathbf{A}\|_2.$$

11. 求矩阵幂级数 $\sum_{k=0}^{\infty} \mathbf{A}^k$ 的和矩阵. 其中

$$\mathbf{A} = \begin{pmatrix} 0.1 & 0.7 \\ 0.3 & 0.6 \end{pmatrix}.$$

12. 讨论矩阵幂级数 $\sum_{k=1}^{\infty} \frac{1}{k^2} \mathbf{A}^k$ 的敛散性. 其中

$$\mathbf{A} = \begin{pmatrix} -2 & 1 \\ -1 & 0 \end{pmatrix}.$$

13. 证明: 矩阵幂级数的敛散性是相似不变性.

14. 设 $\mathbf{A} = \begin{pmatrix} 2 & 1 & 4 \\ 0 & 2 & 0 \\ 0 & 3 & 1 \end{pmatrix}$, 计算 $e^{At}, \sin At$.

15. 设 $\mathbf{A} = \begin{pmatrix} 1 & 4 & 10 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$, 计算 e^{At} .

16. 设 $\mathbf{A} = \begin{pmatrix} 3 & 0 & 8 \\ 3 & -1 & 6 \\ -2 & 0 & -5 \end{pmatrix}$, 求微分方程组 $\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t)$ 满足初始条件 $\mathbf{x}(0) =$

$(1 \ 1 \ 1)^T$ 的解.

17. 设 $\mathbf{A} = \begin{pmatrix} -2 & 1 & 0 \\ -4 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}$, $\mathbf{f}(t) = (1 \ 2 \ e^t - 1)^T$. 求微分方程组 $\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{f}(t)$ 满足初始条件 $\mathbf{x}(0) = (1 \ 1 \ -1)^T$ 的解.

第 6 章 矩阵的 Kronecker 积 与 Hadamard 积

已有的知识背景已足以使我们认识到了矩阵乘法的重要性. 实际上, 在矩阵的普通乘法以外, 还可以定义其他的乘法. 这一章将要介绍的 Kronecker 积与 Hadamard 积也是重要的矩阵乘法, 它们有很好的物理、数学等学科研究问题的背景, 已被成功地应用在许多领域中.

这一章将介绍 Kronecker 积与 Hadamard 积的定义、基本性质及其结合矩阵的向量化算子建立起来的基本理论结果, 同时也简要介绍了它们在求解矩阵方程中的应用例子.

6.1 Kronecker 积与 Hadamard 积的定义

与矩阵的普通乘法 \mathbf{AB} 要求 $\mathbf{A} \in F^{m \times k}$, $\mathbf{B} \in F^{k \times n}$ 不同, Kronecker 积是在任意的两个矩阵之间可以定义的运算, 而 Hadamard 积则是在两个同阶的矩阵之间定义的简单运算.

定义 6.1 设 $\mathbf{A} = (a_{ij}) \in \mathbf{C}^{m \times n}$, $\mathbf{B} = (b_{ij}) \in \mathbf{C}^{s \times t}$, 则 \mathbf{A} 与 \mathbf{B} 的 Kronecker 积 $\mathbf{A} \otimes \mathbf{B}$ 定义为

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11} \mathbf{B} & a_{12} \mathbf{B} & \cdots & a_{1n} \mathbf{B} \\ a_{21} \mathbf{B} & a_{22} \mathbf{B} & \cdots & a_{2n} \mathbf{B} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} \mathbf{B} & a_{m2} \mathbf{B} & \cdots & a_{mn} \mathbf{B} \end{pmatrix} \in \mathbf{C}^{ms \times nt} \quad (6.1)$$

如果 \mathbf{A} 与 \mathbf{B} 是同阶矩阵, 即 $\mathbf{A} = (a_{ij}) \in \mathbf{C}^{m \times n}$, $\mathbf{B} = (b_{ij}) \in \mathbf{C}^{m \times n}$, 则 \mathbf{A} 与 \mathbf{B} 的 Hadamard 积 $\mathbf{A} \circ \mathbf{B}$ 定义为

$$\mathbf{A} \circ \mathbf{B} = (a_{ij} b_{ij}) \in \mathbf{C}^{m \times n}. \quad (6.2)$$

从定义看矩阵的 Kronecker 积被表示为矩阵的分块运算, 即 $\mathbf{A} \otimes \mathbf{B}$ 是一个分块矩阵, 每一个子块是数乘矩阵运算 ($a_{ij} \mathbf{B}$). 矩阵的 Kronecker 积也称为直积或张量积. 矩阵 \mathbf{A} 与 \mathbf{B} 的 Hadamard 积是简单地将 \mathbf{A} 与 \mathbf{B} 对应的元素相乘. 矩阵的 Hadamard 也称为 Schur 积.

例 1 设 $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} -1 & 5 \\ 4 & -3 \end{pmatrix}$, 计算 $\mathbf{A} \otimes \mathbf{B}$, $\mathbf{B} \otimes \mathbf{A}$, $\mathbf{A} \circ \mathbf{B}$ 和 $\mathbf{B} \circ \mathbf{A}$.

$$\text{解 } A \otimes B = \begin{pmatrix} 1B & 2B \\ 3B & 4B \end{pmatrix} = \left(\begin{array}{cc|cc} -1 & 5 & -2 & 10 \\ 4 & -3 & 8 & -6 \\ \hline -3 & 15 & -4 & 20 \\ 12 & -9 & 16 & -12 \end{array} \right),$$

$$B \otimes A = \begin{pmatrix} -A & 5A \\ 4A & -3A \end{pmatrix} = \left(\begin{array}{cc|cc} -1 & -2 & 5 & 10 \\ -3 & -4 & 15 & 20 \\ \hline 4 & 8 & -3 & -6 \\ 12 & 16 & -9 & -12 \end{array} \right),$$

$$A^\circ B = \begin{pmatrix} -1 & 10 \\ 12 & -12 \end{pmatrix}, \quad B^\circ A = \begin{pmatrix} -1 & 10 \\ 12 & -12 \end{pmatrix}.$$

从例 1 可以看到, $A^\circ B = B^\circ A$. 一般地, 由于 $a_{ij} b_{ij} = b_{ij} a_{ij}$, 因此成立 $A^\circ B = B^\circ A$, 即 Hadamard 积是有交换律的, 它克服了普通乘法不可交换($AB \neq BA$)的问题. 但 $A \otimes B \neq B \otimes A$, 即 Kronecker 积仍不具有交换律, 这一点和普通乘法一样. 但观察 $A \otimes B$ 与 $B \otimes A$,

$$A \otimes B \xrightarrow{c_2 \leftrightarrow c_3} \begin{pmatrix} -1 & -2 & 5 & 10 \\ 4 & 8 & -3 & -6 \\ -3 & -4 & 15 & 20 \\ 12 & 16 & 9 & -12 \end{pmatrix} \xrightarrow{r_2 \leftrightarrow r_3} \begin{pmatrix} -1 & -2 & 5 & 10 \\ -3 & -4 & 15 & 20 \\ 4 & 8 & -3 & -6 \\ 12 & 16 & -9 & -12 \end{pmatrix} = B \otimes A,$$

即互换 $A \otimes B$ 中第 2 列与第 3 列, 再互换第 2 行与第 3 行就得到了 $B \otimes A$. 后面将证明: 一般地, 只需要简单地重排 $A \otimes B$ 的某些行和列, 就可以得到 $B \otimes A$. 从这一点上可以说 Kronecker 积部分地克服了普通乘法不可交换的缺点.

如果 A, B 是向量, $A = (a_1 \ a_2 \ \cdots \ a_n)^T \in \mathbf{C}^n$, $B = (b_1 \ b_2 \ \cdots \ b_n)^T \in \mathbf{C}^n$, 则有

$$A \otimes B = (a_1 b_1 \ a_1 b_2 \ \cdots \ a_1 b_n \ \cdots \ a_n b_1 \ a_n b_2 \ \cdots \ a_n b_n)^T \in \mathbf{C}^{nn},$$

$$A^\circ B = (a_1 b_1 \ a_2 b_2 \ \cdots \ a_n b_n)^T \in \mathbf{C}^n,$$

即向量的 Kronecker 积与 Hadamard 积仍是向量.

易于验证下列简单性质成立.

$$(1) \mathbf{0} \otimes \mathbf{0} = \mathbf{0}, \quad \mathbf{0}^\circ \mathbf{0} = \mathbf{0}.$$

$$(2) I \otimes I = I, \quad I^\circ I = I.$$

(3) 如果 A 是对角矩阵, 则 $A \otimes B$ 是准对角矩阵, $A^\circ B$ 是对角矩阵,

$$A \otimes B = \begin{pmatrix} a_{11} B & & & \\ & a_{22} B & & \\ & & \ddots & \\ & & & a_{nn} B \end{pmatrix}, \quad A^\circ B = \begin{pmatrix} a_{11} b_{11} & & & \\ & a_{22} b_{22} & & \\ & & \ddots & \\ & & & a_{nn} b_{nn} \end{pmatrix}.$$

特别地, 两个对角矩阵的 Kronecker 积和 Hadamard 积仍是对角矩阵.

更进一步, 可以证明 Kronecker 积的结合律, 分配律等性质成立.

定理 6.1 设 A, B, C 是矩阵, k 是数, 则下列性质成立

- (1) $(kA) \otimes B = A \otimes (kB) = k(A \otimes B)$;
- (2) $A \otimes (B + C) = A \otimes B + A \otimes C; (B + C) \otimes A = B \otimes A + C \otimes A$;
- (3) $(A \otimes B) \otimes C = A \otimes (B \otimes C)$;
- (4) $(A \otimes B)^H = A^H \otimes B^H$.

证明 因为 Kronecker 积是用分块矩阵定义的, 所以用分块矩阵性质, 证明等式两边子矩阵相等即可.

- (1) $(kA) \otimes B = (ka_{ij}) \otimes B = (ka_{ij} B) = (a_{ij} (kB)) = A \otimes (kB)$;
- (2) $A \otimes (B + C) = (a_{ij} (B + C)) = (a_{ij} B + a_{ij} C) = (a_{ij} B) + (a_{ij} C)$
 $= A \otimes B + A \otimes C$;

(3) 设 $A = (a_{ij}) \in \mathbf{C}^{m \times n}$, $B = (b_{ij}) \in \mathbf{C}^{s \times t}$,

则 $(A \otimes B) \otimes C = (a_{ij} B) \otimes C = (a_{ij} b_{st} C)$.

又 $A \otimes (B \otimes C) = A \otimes (b_{st} C) = (a_{ij} b_{st} C)$,

因此 $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.

(4) $(A \otimes B)^H = (a_{ij} B)^H$, 所以矩阵 $(A \otimes B)^H$ 的第 i 行第 j 列位置上的子块为:

$$((A \otimes B)^H)_{ij} = (\bar{a}_{ji} B^H).$$

又 $A^H \otimes B^H = ((A^H)_{ij} B^H) = (\bar{a}_{ji} B^H)$,

因此 $(A \otimes B)^H = A^H \otimes B^H$. □

下面定理将有助于处理 Kronecker 积和矩阵普通乘法的关系.

定理 6.2 设矩阵 A, B, C, D 是使下列运算有意义的矩阵, 则有

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

证明 设 $(A \otimes B)_{ik}$ 表示矩阵依 B 阶数分块中的第 i 行、第 k 列子块矩阵, 同理于 $(C \otimes D)_{kj}$.

因此, 用分块矩阵的乘法, 可以有下列表子块表示式:

$$\begin{aligned} ((A \otimes B)(C \otimes D))_{ij} &= \sum_{k=1}^n (A \otimes B)_{ik} (C \otimes D)_{kj} = \sum_{k=1}^n (\mathbf{a}_{ik} B)(\mathbf{c}_{kj} D) \\ &= \sum_{k=1}^n \mathbf{a}_{ik} \mathbf{c}_{kj} (\mathbf{B}\mathbf{D}) = (AC)_{ij} \mathbf{B}\mathbf{D}, \end{aligned}$$

从而 $(A \otimes B)(C \otimes D) = (AC) \otimes (\mathbf{B}\mathbf{D})$. □

应用定理 6.2

$$A \otimes B = (IA) \otimes (BI) = (I \otimes B)(A \otimes I).$$

推论 设 $A \in F^{m \times m}$, $B \in F^{n \times n}$, 则

$$A \otimes B = (I_m \otimes B)(A \otimes I_n).$$

定理 6.2 给出了 Kronecker 积与矩阵乘法的关系式, 应用它可以得到许多涉及 Kronecker 积运算得到的矩阵的重要性质.

例 2 证明如果方阵 A, B 皆可逆, 则矩阵 $A \otimes B$ 可逆, 而且 $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

证明 由已知, 存在 A^{-1} 和 B^{-1} . 由定理 6.2,

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

由矩阵可逆的定义, $A \otimes B$ 可逆, 而且

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$

从定义可以看到, 矩阵的 Hadamard 积是比较简单的运算, 易于验证下列性质成立.

- (1) $(kA)^\circ B = A^\circ (kB)$;
- (2) $A^\circ (B+C) = A^\circ B + A^\circ C$;
- (3) $(A^\circ B)^\circ C = A^\circ (B^\circ C)$;
- (4) $(A^\circ B)^H = A^H \circ B^H$.

另外, $A^\circ B$ 的第 i 行、第 j 列的元素 $(A^\circ B)_{ij} = a_j b_{ij}$, 一定是 $A \otimes B$ 的元素之一, 在例 1 中, 观察 $A \otimes B$ 和 $A^\circ B$, 可以发现如果在 $A \otimes B$ 中取第 1 行, 第 1 列; 第 4 行, 第 4 列, 由位于这些行与列交叉处元素构成的子矩阵就是 $A^\circ B$. 该事实在一般的方阵中也成立.

定理 6.3 设 $A, B \in F^{n \times n}$, 集合 $S = \{1, n+2, 2n+3, 3n+4, \dots, (n-1)n+n = n^2\}$, 则 Hardamard 积 $A^\circ B$ 是在 Kronecker 积 $A \otimes B$ 中同时取 S 中的数对应的行和列得到的子矩阵, 记为 $A \otimes B|S| = A^\circ B$.

证明 首先应用空间 F^n 中的基本单位向量 e_i , 可以将矩阵 A 和 B 的元素分别表示为

$$a_{ij} = e_i^T A e_j, \quad b_{ij} = e_i^T B e_j,$$

从而有

$$a_{ij} b_{ij} = (e_i^T A e_j) \otimes (e_i^T B e_j) = (e_i \otimes e_i)^T (A \otimes B) (e_j \otimes e_j).$$

以向量 $e_i \otimes e_i$ 作为第 i 列得矩阵 E ,

$$E = (e_1 \otimes e_1 \quad e_2 \otimes e_2 \quad \dots \quad e_n \otimes e_n) \in F^{n^2 \times n},$$

则

$$\begin{aligned} E^T (A \otimes B) E &= \begin{pmatrix} (e_1 \otimes e_1)^T \\ (e_2 \otimes e_2)^T \\ \dots \\ (e_n \otimes e_n)^T \end{pmatrix} (A \otimes B) ((e_1 \otimes e_1) \quad (e_2 \otimes e_2) \quad \dots \quad (e_n \otimes e_n)) \\ &= \begin{pmatrix} (e_1 \otimes e_1)^T (A \otimes B) (e_1 \otimes e_1) & (e_1 \otimes e_1)^T (A \otimes B) (e_2 \otimes e_2) & \dots & (e_1 \otimes e_1)^T (A \otimes B) (e_n \otimes e_n) \\ (e_2 \otimes e_2)^T (A \otimes B) (e_1 \otimes e_1) & (e_2 \otimes e_2)^T (A \otimes B) (e_2 \otimes e_2) & \dots & (e_2 \otimes e_2)^T (A \otimes B) (e_n \otimes e_n) \\ \dots & \dots & \dots & \dots \\ (e_n \otimes e_n)^T (A \otimes B) (e_1 \otimes e_1) & (e_n \otimes e_n)^T (A \otimes B) (e_2 \otimes e_2) & \dots & (e_n \otimes e_n)^T (A \otimes B) (e_n \otimes e_n) \end{pmatrix} \end{aligned}$$

$$= \begin{pmatrix} a_{11} b_{11} & a_{12} b_{12} & \cdots & a_{1n} b_{1n} \\ a_{21} b_{21} & a_{22} b_{22} & \cdots & a_{2n} b_{2n} \\ \cdots & \cdots & & \cdots \\ a_{n1} b_{n1} & a_{n2} b_{n2} & \cdots & a_{nn} b_{nn} \end{pmatrix} = A \circ B.$$

由于矩阵 E 的 n 个列向量分别是空间 F^{n^2} 中的基本单位向量 $e_1, e_{n+2}, e_{2n+3}, e_{3n+4}, \dots, e_{n^2}$, 故 $E^T (A \otimes B) E$ 的作用是取 S 中数对应的行和列, 即取第 j 行、第 j 列, $j \in S$, 因此 $A \circ B$ 是(取 $A \otimes B$ 中第 j 行与第 j 列对所有的 $j \in S$)得到的子矩阵. \square

6.2 Kronecker 积与 Hadamard 积的性质

在上一节中, 已经证明了 Kronecker 积与 Hadamard 积的一些简单性质. 对矩阵 A 和 B 这两种积运算以后的结果, $A \otimes B$ 与 $A \circ B$ 仍然是矩阵. 这一节主要从矩阵的角度, 讨论矩阵 $A \otimes B$ 和 $A \circ B$ 的一些性质, 它们将有助于在矩阵研究和方法中处理 $A \otimes B$ 和 $A \circ B$.

定理 6.4 设 A 和 B 是使下列运算有意义的矩阵, 对 A 与 B 的 Kronecker 积矩阵 $A \otimes B$, 下列性质成立.

(1) 当 A, B 分别可逆时, $A \otimes B$ 为逆矩阵, 而且有

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1};$$

(2) 当方阵 $A \in F^{m \times m}$, $B \in F^{n \times n}$ 时, 方阵 $A \otimes B$ 的行列式

$$|A \otimes B| = |B \otimes A| = |A|^n |B|^m;$$

(3) 若 A 和 B 都是 Hermite 矩阵, 则 $A \otimes B$ 是 Hermite 矩阵;

(4) 若 A 与 B 都是酉矩阵, 则 $A \otimes B$ 是酉矩阵.

证明 (1)已在例 2 中给出了证明.

(2)对方阵 A , 存在可逆矩阵 P 和 Jordan 矩阵 J_A , 使 $A = P J_A P^{-1}$, 其中

$$J = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}, \quad \text{且已知 } |A| = \prod_{i=1}^n \lambda_i.$$

由定理 5.2 及其推论

$$\begin{aligned} A \otimes B &= (I_m \otimes B)(A \otimes I_n) = (I_m \otimes B)((P J_A P^{-1}) \otimes I_n) \\ &= (I_m \otimes B)(P \otimes I)(J_A \otimes I_n)(P \otimes I)^{-1}, \end{aligned}$$

从而有

$$|A \otimes B| = |I_m \otimes B| |J_A \otimes I_n| = \begin{vmatrix} B & & & \\ & B & & \\ & & \ddots & \\ & & & B \end{vmatrix} \begin{vmatrix} \lambda_1 I & & & \\ & \lambda_2 I & & \\ & & \ddots & \\ & & & \lambda_n I \end{vmatrix}$$

$$= |B|^m \left(\prod_{i=1}^n \lambda_i \right)^n = |A|^n |B|^m.$$

用同样方法, 显然

$$|A \otimes B| = |B \otimes A| = |A|^n |B|^m.$$

(3) 已知 Hermite 矩阵 A, B 满足条件 $A^H = A; B^H = B$,

$$\text{则 } (A \otimes B)^H = A^H \otimes B^H = A \otimes B.$$

从而, $A \otimes B$ 也是 Hermite 矩阵.

(4) 已知酉矩阵 A, B 满足条件

$$A^H A = A A^H = I, \quad B^H B = B B^H = I,$$

$$\text{则 } (A \otimes B)^H (A \otimes B) = (A^H \otimes B^H)(A \otimes B) = (A^H A) \otimes (B^H B)$$

$$= I \otimes I = I.$$

同理可验证

$$(A \otimes B)(A \otimes B)^H = I,$$

从而 $A \otimes B$ 也是酉矩阵. □

从定理 6.4 可以看到, 矩阵的 Kronecker 积有一些很方便的性质, 它们较矩阵的普通乘法优越一些. 如性质(3)与矩阵的普通乘法对照, 当 A 和 B 是 Hermite 矩阵时, AB 是 Hermite 矩阵是需要 A 与 B 乘法可交换这一条件的.

在矩阵的等价和矩阵相似上, 可以在 Kronecker 积中建立如下结果.

定理 6.5 设 A, B 使下列运算有意义, 则有

(1) 设 A, B 为同阶矩阵, 且 A 等价于 B , 则对任意单位矩阵 I , $(A \otimes I)$ 等价于 $(B \otimes I)$.

(2) 设方阵 $A \in F^{m \times m}, B \in F^{n \times n}$, 如果 A 相似于 J_A , B 相似于 J_B , 则 $A \otimes B$ 相似于 $J_A \otimes J_B$.

证明 (1) 由等价的定义, 存在可逆矩阵 P, Q , 使 $PAQ = B$, 则由定理 6.3 与定理 6.4, 对任意单位矩阵 I , $P \otimes I$ 和 $Q \otimes I$ 仍然是可逆矩阵. 而且

$$(P \otimes I)(A \otimes I)(Q \otimes I) = ((PA) \otimes I)(Q \otimes I) = (PAQ) \otimes I = B \otimes I,$$

所以, 矩阵 $(A \otimes I)$ 与矩阵 $(B \otimes I)$ 等价.

(2) 设 P, Q 为可逆阵, $P \in F^{m \times m}, Q \in F^{n \times n}$, 使

$$P^{-1}AP = J_A, \quad Q^{-1}BQ = J_B,$$

则 $P \otimes Q$ 仍然是可逆矩阵

$$(P \otimes Q)^{-1}(A \otimes B)(P \otimes Q) = ((P^{-1}A) \otimes (Q^{-1}B))(P \otimes Q)$$

$$= (\mathbf{P}^{-1} \mathbf{A} \mathbf{P}) \otimes \mathbf{Q}^{-1} \mathbf{B} \mathbf{Q} = \mathbf{J}_A \otimes \mathbf{J}_B.$$

□

在矩阵的普通乘法中, 方阵 \mathbf{A}, \mathbf{B} 的特征值与它们的乘积 (\mathbf{AB}) 的特征值之间是没有相应的乘法关系的. 在 Kronecker 积中, 下面的定理将证明相应的乘法关系成立.

定理 6.6 设方阵 $\mathbf{A} \in F^{m \times m}$, \mathbf{A} 的特征值是 λ_i , 相应的特征向量是 $\mathbf{x}_i, i=1, 2, \dots, m$. 方阵 $\mathbf{B} \in F^{n \times n}$, \mathbf{B} 的特征值是 μ_j , 相应的特征向量是 $\mathbf{y}_j, j=1, 2, \dots, n$, 则

(1) $\mathbf{A} \otimes \mathbf{B}$ 的特征值是 $\lambda_i \mu_j$, 对应的特征向量是 $\mathbf{x}_i \otimes \mathbf{y}_j, i=1, 2, \dots, m; j=1, 2, \dots, n$.

(2) $\mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{B}$ 的特征值是 $\lambda_i + \mu_j$, 对应的特征向量是 $\mathbf{x}_i \otimes \mathbf{y}_j, i=1, 2, \dots, m; j=1, 2, \dots, n$.

证明 (1) $(\mathbf{A} \otimes \mathbf{B})(\mathbf{x}_i \otimes \mathbf{y}_j) = (\mathbf{Ax}_i) \otimes (\mathbf{By}_j) = (\lambda_i \mathbf{x}_i) \otimes (\mu_j \mathbf{y}_j) = \lambda_i \mu_j (\mathbf{x}_i \otimes \mathbf{y}_j)$.

$$\begin{aligned} (2) (\mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{B})(\mathbf{x}_i \otimes \mathbf{y}_j) &= (\mathbf{A} \otimes \mathbf{I}_n)(\mathbf{x}_i \otimes \mathbf{y}_j) + (\mathbf{I}_m \otimes \mathbf{B})(\mathbf{x}_i \otimes \mathbf{y}_j) \\ &= (\mathbf{Ax}_i \otimes \mathbf{y}_j) + (\mathbf{x}_i \otimes \mathbf{By}_j) = \lambda_i (\mathbf{x}_i \otimes \mathbf{y}_j) + \mu_j (\mathbf{x}_i \otimes \mathbf{y}_j) \\ &= (\lambda_i + \mu_j) (\mathbf{x}_i \otimes \mathbf{y}_j). \end{aligned}$$

因此, $(\lambda_i + \mu_j)$ 是 $\mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{B}$ 的特征值, $\mathbf{x}_i \otimes \mathbf{y}_j$ 是与之对应的特征向量. □

在矩阵讨论中, $(\mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{B})$ 称为方阵 \mathbf{A} 和 \mathbf{B} 的 Kronecker 和, 记为

$$\mathbf{A} \oplus \mathbf{B} = \mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{B}. \quad (6.3)$$

当 \mathbf{A}, \mathbf{B} 可对角化时, 即存在可逆矩阵 \mathbf{P}, \mathbf{Q} , 使 $\mathbf{P}^{-1} \mathbf{A} \mathbf{P} = \Lambda_A, \mathbf{Q}^{-1} \mathbf{B} \mathbf{Q} = \Lambda_B$, 其中

$$\Lambda_A = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix}, \quad \Lambda_B = \begin{bmatrix} \mu_1 & & & \\ & \mu_2 & & \\ & & \ddots & \\ & & & \mu_n \end{bmatrix}.$$

由定理 6.5

$$(\mathbf{P} \otimes \mathbf{Q})^{-1} (\mathbf{A} \otimes \mathbf{B}) (\mathbf{P} \otimes \mathbf{Q}) = \Lambda_A \otimes \Lambda_B.$$

$$(\mathbf{P} \otimes \mathbf{Q})^{-1} (\mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{B}) (\mathbf{P} \otimes \mathbf{Q}) = \Lambda_A \otimes \mathbf{I} + \mathbf{I}_m \otimes \Lambda_B.$$

从而, $\mathbf{A} \otimes \mathbf{B}, \mathbf{A} \oplus \mathbf{B}$ 都可对角化.

对方阵 $\mathbf{A} \in F^{m \times m}, \mathbf{B} \in F^{n \times n}$, 矩阵多项式

$$P(\mathbf{A}, \mathbf{B}) = \sum_{i,j=0}^T c_{ij} \mathbf{A}^i \otimes \mathbf{B}^j \in F^{mn \times mn}$$

可以将矩阵多项式 $P(\mathbf{A}, \mathbf{B})$ 对应于一个关于不可交换的变元 x, y 的多项式 $P(x, y)$.

如 $P(x, y) = x^2 - 2xy + x - y^2$.

则 $P(\mathbf{A}, \mathbf{B}) = \mathbf{A}^2 \otimes \mathbf{I} - 2\mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{B}^2$.

在这种对应下, 可建立 Kronecker 积的特征值的一个更为一般的结果.

定理 6.7 设方阵 $\mathbf{A} \in F^{m \times m}$ 的特征值是 $\lambda_1, \lambda_2, \dots, \lambda_m$, 矩阵 $\mathbf{B} \in F^{n \times n}$ 的特征值是 $\mu_1, \mu_2, \dots, \mu_n$, 则矩阵

$$P(\mathbf{A}, \mathbf{B}) = \sum_{i,j=0}^T c_{ij} \mathbf{A}^i \otimes \mathbf{B}^j$$

的特征值是

$$P(\lambda_r, \mu_s) = \sum_{i,j=0}^T c_{ij} \lambda_r^i \mu_s^j, \quad r = 1, 2, \dots, m, \quad s = 1, 2, \dots, n.$$

证明 设 $\mathbf{A} = \mathbf{P} \mathbf{J}_A \mathbf{P}^{-1}$, $\mathbf{B} = \mathbf{Q} \mathbf{J}_B \mathbf{Q}^{-1}$, 其中 $\mathbf{J}_A, \mathbf{J}_B$ 分别是 \mathbf{A} 和 \mathbf{B} 的 Jordan 矩阵, 则有乘幂结果

$$\mathbf{A}^i = \mathbf{P} \mathbf{J}_A^i \mathbf{P}^{-1}, \quad \mathbf{B}^j = \mathbf{Q} \mathbf{J}_B^j \mathbf{Q}^{-1},$$

从而

$$\mathbf{A}^i \otimes \mathbf{B}^j = (\mathbf{P} \otimes \mathbf{Q})(\mathbf{J}_A^i \otimes \mathbf{J}_B^j)(\mathbf{P} \otimes \mathbf{Q})^{-1},$$

进而有

$$P(\mathbf{A}, \mathbf{B}) = (\mathbf{P} \otimes \mathbf{Q}) P(\mathbf{J}_A, \mathbf{J}_B) (\mathbf{P} \otimes \mathbf{Q})^{-1}.$$

已知 \mathbf{A}, \mathbf{B} 的特征值分别是 λ_r 和 μ_s 时, \mathbf{J}_A^i 和 \mathbf{J}_B^j 的特征值分别是 λ_r^i 和 μ_s^j .

由定理 6.6, $\mathbf{J}_A^i \otimes \mathbf{J}_B^j$ 的特征值为 $\lambda_r^i \mu_s^j$, 从而矩阵多项式 $P(\mathbf{J}_A, \mathbf{J}_B)$ 的特征值是 $P(\lambda_r, \mu_s)$ 又从上述相似结果, $P(\mathbf{A}, \mathbf{B})$ 的特征值是

$$P(\lambda_r, \mu_s) = \sum_{i,j} c_{ij} \lambda_r^i \mu_s^j, \quad 1 \leq r \leq m, \quad 1 \leq s \leq n. \quad \square$$

$\mathbf{A} \otimes \mathbf{B}$ 与 $\mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{B}$ 可视为矩阵多项式 $P(\mathbf{A}, \mathbf{B})$ 的特例:

$$P(x, y) = xy \text{ 时, } P(\mathbf{A}, \mathbf{B}) = \mathbf{A} \otimes \mathbf{B},$$

$$P(x, y) = x + y \text{ 时, } P(\mathbf{A}, \mathbf{B}) = \mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{B},$$

从而定理 6.7 的结果比定理 6.6 更一般.

定理 6.8 设 $f(z)$ 是解析函数, $\mathbf{A} \in F^{n \times n}$, 且 $f(\mathbf{A})$ 存在, 则

$$f(\mathbf{I}_m \otimes \mathbf{A}) = \mathbf{I}_m \otimes f(\mathbf{A}),$$

$$f(\mathbf{A} \otimes \mathbf{I}_m) = f(\mathbf{A}) \otimes \mathbf{I}_m.$$

证明 当 $f(z)$ 为解析函数, $f(\mathbf{A})$ 收敛时, 从 6.5 小节已知 $f(\mathbf{A})$ 可表示为一个次数不超过 $n-1$ 的矩阵多项式, 即

$$\begin{aligned} f(\mathbf{A}) &= \sum_{k=0}^{n-1} a_k \mathbf{A}^k, \\ f(\mathbf{I}_m \otimes \mathbf{A}) &= \sum_{k=0}^{n-1} a_k (\mathbf{I}_m \otimes \mathbf{A})^k = \sum_{k=0}^{n-1} a_k (\mathbf{I}_m \otimes \mathbf{A}^k) = \sum_{k=0}^{n-1} (\mathbf{I}_m \otimes a_k \mathbf{A}^k) \\ &= \mathbf{I}_m \otimes \sum_{k=0}^{n-1} a_k \mathbf{A}^k = \mathbf{I}_m \otimes f(\mathbf{A}). \end{aligned}$$

类似地,

$$\begin{aligned} f(\mathbf{A} \otimes \mathbf{I}_n) &= \sum_{k=0}^{n-1} a_k (\mathbf{A} \otimes \mathbf{I}_n)^k = \sum_{k=0}^{n-1} a_k (\mathbf{A}^k \otimes \mathbf{I}_n) \\ &= \left(\sum_{k=0}^{n-1} a_k \mathbf{A}^k \right) \otimes \mathbf{I}_n = f(\mathbf{A}) \otimes \mathbf{I}_n. \end{aligned}$$

定理得证. \square

定理 6.8 的一个重要应用是对 $f(z) = e^z$,

$$\begin{aligned} e^{I_m \otimes A} &= I_m \otimes e^A, \\ e^{A \otimes I_m} &= e^A \otimes I_m. \end{aligned}$$

例 3 设 $A = \begin{pmatrix} 3 & -1 \\ 0 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 2 & 0 \\ 1 & -1 \end{pmatrix}$,

(1) 求 $A \otimes B$ 的特征值与特征向量;

(2) 求 $A \otimes I_2 + I_2 \otimes B$ 的特征值与特征向量.

解 易求得 A 的特征值 $\lambda_1 = 3, \lambda_2 = 1$, 对应的特征向量 $x_1 = (1 \ 0)^T, x_2 = (1 \ 3)^T$.

B 的特征值 $\mu_1 = 2, \mu_2 = -1$, 对应的特征向量

$$y_1 = (3 \ -1)^T, \quad y_2 = (0 \ 1)^T.$$

$$(1) A \otimes B = \left(\begin{array}{cc|cc} 6 & 0 & -2 & 0 \\ 3 & -3 & -1 & 1 \\ \hline 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & -1 \end{array} \right),$$

$A \otimes B$ 的特征值及其对应特征向量分别是:

$$\lambda_1 \mu_1 = 6; \quad x_1 \otimes y_1 = (3 \ 1 \ 0 \ 0)^T,$$

$$\lambda_1 \mu_2 = -3; \quad x_1 \otimes y_2 = (0 \ 1 \ 0 \ 0)^T,$$

$$\lambda_2 \mu_1 = 2; \quad x_2 \otimes y_1 = (3 \ 1 \ 9 \ 3)^T,$$

$$\lambda_2 \mu_2 = -1; \quad x_2 \otimes y_2 = (0 \ 1 \ 0 \ 3)^T.$$

$$\begin{aligned} (2) A \otimes I_2 + I_2 \otimes B &= \left(\begin{array}{cc|cc} 3 & 0 & -1 & 0 \\ 0 & 3 & 0 & -1 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right) + \left(\begin{array}{cc|cc} 2 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ \hline 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & -1 \end{array} \right) \\ &= \left(\begin{array}{cc|cc} 5 & 0 & -1 & 0 \\ 1 & 2 & 0 & -1 \\ \hline 0 & 0 & 3 & 0 \\ 0 & 0 & 1 & 0 \end{array} \right). \end{aligned}$$

$A \otimes I_2 + I_2 \otimes B$ 的特征值及对应的特征向量为

$$\lambda_1 + \mu_1 = 5; \quad x_1 \otimes y_1 = (3 \ 1 \ 0 \ 0)^T,$$

$$\lambda_1 + \mu_2 = 2; \quad x_1 \otimes y_2 = (0 \ 1 \ 0 \ 0)^T,$$

$$\lambda_2 + \mu_1 = 3; \quad x_2 \otimes y_1 = (3 \ 1 \ 9 \ 3)^T,$$

$$\lambda_2 + \mu_2 = 0; \quad x_2 \otimes y_2 = (0 \ 1 \ 0 \ 3)^T.$$

本题中 A, B 可相似对角化, 即 J_A 与 J_B 为对角矩阵. 因此, $A \otimes B$ 与 $A \otimes I_2 + I_2 \otimes B$ 的 Jordan 标准形分别是:

$$J_{A \otimes B} = J_A \otimes J_B, \quad J_{(A \otimes I_2 + I_2 \otimes B)} = J_A \otimes I_2 + I_2 \otimes J_B.$$

从定理 6.8 还可以推出, 如果 A, B 是正定矩阵, 则 $A \otimes B$ 与 $(A \otimes I_n + I_m \otimes B)$ 都是正定矩阵. Hadamard 积也具有保有正定性这一性质. Schur 积定理将证明这一点. Schur 积定理也是建立矩阵不等式和处理与正定、半正定矩阵相关的问题中常用的重要定理.

定理 6.9(Schur 积定理) 设 $A, B \in F^{n \times n}$, 如果 A, B 是半正定(正定)矩阵, 则 $A^\circ B$ 也是半正定(正定)矩阵.

证明 设 A, B 分别是秩为 k 和 l 的半正定矩阵, 则由定理 5.6, A, B 可分别表示为

$$A = v_1 v_1^H + v_2 v_2^H + \cdots + v_k v_k^H; \quad B = w_1 w_1^H + w_2 w_2^H + \cdots + w_l w_l^H.$$

由此, 可将 Hadamard 积表示为

$$A^\circ B = \sum_{i,j=1}^{k,l} u_{ij} \cdot u_{ij}^H,$$

其中

$$u_{ij} = v_i \circ w_j.$$

而 $u_{ij} u_{ij}^H$ 是半正定矩阵, 从而其和 $A^\circ B$ 是半正定矩阵.

当 A, B 是正定矩阵时, 有 $k=l=n$, $\{v_i\}$ 和 $\{w_i\}$ 是空间 F^n 的正交基. 由上证明 $A^\circ B$ 半正定, 要证 $A^\circ B$ 正定, 仅证 $A^\circ B$ 非奇异即可.

反证 $A^\circ B$ 是奇异矩阵, 则 $\exists x \in F^n, x \neq \mathbf{0}$, 但 $(A^\circ B)x = \mathbf{0}$, 从而

$$x^H (A^\circ B)x = \sum_{i,j=1}^{k,l} x^H (u_{ij} u_{ij}^H) x = \sum_{i,j=1}^{k,l} |u_{ij}^H x|^2 = 0.$$

这导出

$$u_{ij}^H x = \mathbf{0} \quad \text{或} \quad x^H (v_i \circ w_j) = \mathbf{0}, \quad \forall i, j.$$

$$\text{又} \quad x^H (v_i \circ w_j) = (x \circ \bar{v}_i)^H w_j = \mathbf{0}, \quad \forall i, j,$$

这说明 $(x \circ \bar{v}_i)$ 与 F^n 中正交基 $\{w_j\}$ 中每一向量正交,

$$\text{因此有} \quad x \circ \bar{v}_i = \mathbf{0}, \quad i = 1, 2, \dots, n.$$

$$\text{注意} \quad x \circ \bar{v}_i = \mathbf{0} \Leftrightarrow v_i^H x = \mathbf{0}, \quad i = 1, 2, \dots, n.$$

这导出 x 与空间 F^n 的正交基 $\{v_i\}$ 正交, 有 $x = \mathbf{0}$, 与 $x \neq \mathbf{0}$ 矛盾, 从而 $A^\circ B$ 非奇异, 即 $A^\circ B$ 是正定矩阵.

6.3 矩阵的向量化算子与 Kronecker 积

在许多理论和应用问题的分析中, 处理数值向量的问题往往比处理数值矩阵的问题有更多的方法上的方便之处. 例如一个方程 $Ax = B$, 当 A 为奇异矩阵时, 若 x 和 B 是向量, 则容易用线性方程组的方法求解. 当 x 和 B 是矩阵时, 其中的一个方法是在 x 的列向量 x_i 和 B 的列向量 b_i 之间建立方程 $Ax_i = b_i, i = 1, 2, \dots$ 来求解. 而对形如 $Ax + xA = B$ 的方程, 就难以直接用列向量将问题转化为线性方程组了. 这一节介绍的矩阵的向量化算子是一种将矩阵表示为向量的方法, 将它结合于 Kronecker 积, 可用于解决上述

方程组的问题和解决其他的问题.

定义 6.2 设 $A \in F^{m \times n}$ 是 $m \times n$ 阶矩阵, $A = (A_1, A_2, \dots, A_n)$, 其中 $A_i \in F^m$ 是 A 的第 i 列, 则 A 的向量算子 $\text{Vec}(A)$, 定义为:

$$\text{Vec}(A) = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{pmatrix} \in F^{mn}.$$

$\text{Vec}(A)$ 在 A 上的作用是依 A 的列的顺序将 A 转化为一个列向量. 例如

$$\text{Vec} \begin{pmatrix} 2 & 1 & 4 & 3 \\ 0 & -1 & 2 & 7 \\ 5 & 4 & 3 & 1 \end{pmatrix} = [2 \ 0 \ 5 \ 1 \ -1 \ 4 \ 4 \ 2 \ 3 \ 3 \ 7 \ 1]^T.$$

用定义, 易于证明下列简单性质:

- (1) 当 A 是 $m \times n$ 矩阵时, $\text{Vec}(A)$ 是 $mn \times 1$ 向量;
- (2) 当 x 是向量时, $\text{Vec}(x) = \text{Vec}(x^T) = x$;
- (3) 当 x 是 $m \times 1$ 向量, y 是 $n \times 1$ 向量时, $\text{Vec}(xy^T) = x \otimes y$;
- (4) Vec 是线性算子, 即 $A, B \in F^{m \times n}$, a, b 是数,

$$\text{Vec}(aA + bB) = a\text{Vec}(A) + b\text{Vec}(B);$$

(5) 若 A_1, A_2, \dots, A_k 是空间 $F^{m \times n}$ 中的线性无关向量组, 则 $\text{Vec}(A_1), \text{Vec}(A_2), \dots, \text{Vec}(A_k)$ 是空间 F^{mn} 中的线性无关组.

更进一步有下列定理.

定理 6.10 设矩阵 $A \in F^{m \times k}, B \in F^{k \times s}, C \in F^{s \times n}$, 则

$$\text{Vec}(ABC) = (C^T \otimes A)\text{Vec}(B). \quad (6.4)$$

证明 设用列向量分别将 B, C 表示为

$$B = (B_1 \ B_2 \ \cdots \ B_s), \quad C = (C_1 \ C_2 \ \cdots \ C_n),$$

$$\begin{aligned} \text{则 } ABC &= A(B_1 \ B_2 \ \cdots \ B_s)C = A \left(\sum_{j=1}^s C_{1j} B_j \ \sum_{j=1}^s C_{2j} B_j \ \cdots \ \sum_{j=1}^s C_{nj} B_j \right) \\ &= \left[A \sum_{j=1}^s C_{1j} B_j \ A \sum_{j=1}^s C_{2j} B_j \ \cdots \ A \sum_{j=1}^s C_{nj} B_j \right] \\ &= (AC_1^T \text{Vec}(B) \ AC_2^T \text{Vec}(B) \ \cdots \ AC_n^T \text{Vec}(B)). \end{aligned}$$

$$\text{Vec}(ABC) = \text{Vec}(AC_1^T \text{Vec}(B) \ AC_2^T \text{Vec}(B) \ \cdots \ AC_n^T \text{Vec}(B))$$

$$= \begin{pmatrix} AC_1^T \text{Vec}(B) \\ AC_2^T \text{Vec}(B) \\ \cdots \\ AC_n^T \text{Vec}(B) \end{pmatrix} = \begin{pmatrix} C_{11}A & C_{21}A & \cdots & C_{s1}A \\ C_{12}A & C_{22}A & \cdots & C_{s2}A \\ \cdots & \cdots & \cdots & \cdots \\ C_{1n}A & C_{2n}A & \cdots & C_{sn}A \end{pmatrix} \text{Vec}(B)$$

$$= (\mathbf{C}^T \otimes \mathbf{A}) \text{Vec}(\mathbf{B}).$$

定理得证. □

推论 设 $\mathbf{A} \in F^{m \times k}$, $\mathbf{X} \in F^{k \times s}$, $\mathbf{C} \in F^{s \times n}$, 则有

- (1) $\text{Vec}(\mathbf{AX}) = (\mathbf{I}_n \otimes \mathbf{A}) \text{Vec}(\mathbf{X})$;
- (2) $\text{Vec}(\mathbf{XC}) = (\mathbf{C}^T \otimes \mathbf{I}_k) \text{Vec}(\mathbf{X})$.

例 4 设 $\mathbf{A} = \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} -3 & 4 \\ 1 & 0 \end{pmatrix}$, $\mathbf{D} = \begin{pmatrix} 1 & 3 \\ -2 & 2 \end{pmatrix}$, 求解矩阵方程 $\mathbf{AX} + \mathbf{XB} = \mathbf{D}$.

D.

解 用向量化算子 Vec 作用在等式两边, 有

$$(\mathbf{I}_2 \otimes \mathbf{A} + \mathbf{B}^T \otimes \mathbf{I}_2) \text{Vec}(\mathbf{X}) = \text{Vec}(\mathbf{D}),$$

即

$$\begin{pmatrix} -2 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 4 & 0 & 1 & -1 \\ 0 & 4 & 0 & 2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 3 \\ 2 \end{pmatrix}. \quad (6.5)$$

易求得二阶方阵 \mathbf{A}, \mathbf{B} 的特征值分别为 $\lambda_1 = 1, \lambda_2 = 2; \mu_1 = -4, \mu_2 = 1$. 从而上述线性方程组(6.5)的系数矩阵 $(\mathbf{I}_2 \otimes \mathbf{A} + \mathbf{B}^T \otimes \mathbf{I}_2)$ 的特征值为 $\lambda_i + \mu_j \neq 0, i=1, 2, j=1, 2$, 即系数矩阵非奇异, 故线性方程组(6.5)有惟一解

$$\text{Vec}(\mathbf{X}) = (0 \ 1 \ 2 \ -1)^T.$$

因此, 原方程组的解矩阵 $\mathbf{X} = \begin{pmatrix} 0 & 2 \\ 1 & -1 \end{pmatrix}$.

例 5 设 $\mathbf{A}_1 = \begin{pmatrix} 2 & 2 \\ 2 & -1 \end{pmatrix}$, $\mathbf{A}_2 = \begin{pmatrix} 0 & 1 \\ -2 & -1 \end{pmatrix}$, $\mathbf{B}_1 = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$, $\mathbf{B}_2 = \begin{pmatrix} 0 & 2 \\ -1 & 3 \end{pmatrix}$, $\mathbf{D} = \begin{pmatrix} 4 & -6 \\ 0 & 8 \end{pmatrix}$, 求解矩阵方程 $\mathbf{A}_1 \mathbf{ZB}_1 + \mathbf{A}_2 \mathbf{ZB}_2 = \mathbf{D}$.

解 用向量化算子 Vec 作用方程两边得

$$(\mathbf{B}_1^T \otimes \mathbf{A}_1 + \mathbf{B}_2^T \otimes \mathbf{A}_2) \text{Vec}(\mathbf{Z}) = \text{Vec}(\mathbf{D}),$$

即

$$\begin{pmatrix} 2 & 2 & -2 & -3 \\ 2 & -1 & 0 & 2 \\ 0 & 2 & 2 & 5 \\ -4 & -2 & -4 & -4 \end{pmatrix} \text{Vec}(\mathbf{Z}) = \begin{pmatrix} 4 \\ 0 \\ -6 \\ 8 \end{pmatrix}.$$

解得该方程组的惟一解为

$$\text{Vec}(\mathbf{Z}) = (1 \ -1 \ -2 \ 0)^T,$$

从而

$$\mathbf{Z} = \begin{pmatrix} 1 & -2 \\ -1 & 0 \end{pmatrix}.$$

从例 4 和例 5 可以看到, 用算子 Vec 结合 Kronecker 积能求解用矩阵普通乘法不易求

解的矩阵方程. 它们的另一个应用是求解微分方程组.

例 6 求解微分方程组

$$\begin{cases} \dot{\mathbf{X}}(t) = \mathbf{A}\mathbf{X}(t) + \mathbf{X}(t)\mathbf{B}, \\ \mathbf{X}(0) = \mathbf{C}, \end{cases}$$

其中

$$\mathbf{A} \in \mathbb{C}^{m \times m}, \mathbf{B} \in \mathbb{C}^{n \times n}, \mathbf{X} \in \mathbb{C}^{m \times n}.$$

解 用向量化算子作用方程两边, 有

$$\begin{aligned} \text{Vec}(\dot{\mathbf{X}}(t)) &= (\mathbf{I}_n \otimes \mathbf{A} + \mathbf{B}^T \otimes \mathbf{I}_m) \text{Vec}(\mathbf{X}(t)) \\ \text{Vec}(\mathbf{X}(0)) &= \text{Vec}(\mathbf{C}). \end{aligned}$$

令

$$\mathbf{Y}(t) = \text{Vec}(\mathbf{X}(t)), \quad \mathbf{C}_1 = \text{Vec}(\mathbf{C}),$$

$$\mathbf{G} = \mathbf{I}_n \otimes \mathbf{A} + \mathbf{B}^T \otimes \mathbf{I}_m,$$

则原方程组等价于

$$\begin{cases} \dot{\mathbf{Y}}(t) = \mathbf{G}\mathbf{Y}(t), \\ \mathbf{Y}(0) = \mathbf{C}_1. \end{cases}$$

用定理 5.11, 有 $\mathbf{Y}(t) = e^t \mathbf{C}_1$, 即

$$\begin{aligned} \text{Vec}(\mathbf{X}(t)) &= \exp\{(\mathbf{I}_n \otimes \mathbf{A} + \mathbf{B}^T \otimes \mathbf{I}_m)t\} \text{Vec}(\mathbf{C}) \\ &= \exp\{(\mathbf{I}_n \otimes \mathbf{A})t\} \cdot \exp\{(\mathbf{B}^T \otimes \mathbf{I}_m)t\} \text{Vec}(\mathbf{C}) \\ &= (\mathbf{I}_n \otimes \exp(\mathbf{A}t))(\exp(\mathbf{B}^T t) \otimes \mathbf{I}_m) \text{Vec}(\mathbf{C}). \end{aligned}$$

又 $(\exp(\mathbf{B}^T t) \otimes \mathbf{I}_n) \text{Vec}(\mathbf{C}) = \text{Vec}(\mathbf{C} \exp(\mathbf{B}t))$,

从而 $\text{Vec}(\mathbf{X}(t)) = (\mathbf{I}_n \otimes \exp(\mathbf{A}t)) \cdot \text{Vec}(\mathbf{C} \exp(\mathbf{B}t)) = \text{Vec}(\exp(\mathbf{A}t) \mathbf{C} \exp(\mathbf{B}t))$.

因此, 方程组的解为

$$\mathbf{X}(t) = \exp(\mathbf{A}t) \cdot \mathbf{C} \exp(\mathbf{B}t).$$

在 6.1 的例 1 中, 我们曾指出 $\mathbf{A} \otimes \mathbf{B} \neq \mathbf{B} \otimes \mathbf{A}$, 但交换 $\mathbf{A} \otimes \mathbf{B}$ 的某些行和列, 可得到 $\mathbf{B} \otimes \mathbf{A}$. 下面将证明可找到矩阵 \mathbf{P} , 使

$$\mathbf{P}^T (\mathbf{A} \otimes \mathbf{B}) \mathbf{P} = \mathbf{B} \otimes \mathbf{A}.$$

定义 6.3 令 $\mathbf{E}_{ij} \in F^{m \times n}$ 表示第 i 行, 第 j 列元素为 1, 其余元素为 0 的矩阵, 则方阵 $\mathbf{K}_{mn} \in F^{mn \times mn}$ 定义为

$$\mathbf{K}_{mn} = \sum_{i=1}^m \sum_{j=1}^n \mathbf{E}_{ij}^T \otimes \mathbf{E}_{ij}^T.$$

在矩阵理论中, \mathbf{K}_{mn} 称为交换矩阵 (commutation matrix).

定理 6.11 矩阵 \mathbf{K}_{mn} 有下列性质

(1) $\mathbf{K}_{mn} = \mathbf{K}_{nm}$;

(2) $\mathbf{K}_{1n} = \mathbf{K}_{n1} = \mathbf{I}_n$;

(3) $\mathbf{K}_{mn} = \sum_{j=1}^n (\mathbf{e}_j^T \otimes \mathbf{I}_m \otimes \mathbf{e}_j)$.

证明 (1) $\mathbf{K}_{mn}^T = \left[\sum_{i=1}^m \sum_{j=1}^n \mathbf{E}_{ij} \otimes \mathbf{E}_{ij}^T \right]^T = \sum_{i=1}^m \sum_{j=1}^n (\mathbf{E}_{ij} \otimes \mathbf{E}_{ij}^T)^T$

$$= \sum_{i=1}^m \sum_{j=1}^n (\mathbf{E}_{ij}^T \otimes \mathbf{E}_{ij}) = \sum_{i=1}^m \sum_{j=1}^n (\mathbf{E}_{ij} \otimes \mathbf{E}_{ij}^T) = \mathbf{K}_{mn}.$$

(2) 注意到 $\mathbf{E}_{ij} \otimes \mathbf{E}_{ij}^T = \mathbf{e}_j^T \otimes \mathbf{e}_j$,

$$\text{所以 } \mathbf{K}_{1n} = \sum_{j=1}^n (\mathbf{E}_{1j} \otimes \mathbf{E}_{1j}^T) = \sum_{j=1}^n (\mathbf{e}_j^T \otimes \mathbf{e}_j) = \mathbf{I}_n.$$

$$\text{由(1)} \quad \mathbf{K}_{n1} = \mathbf{K}_{1n}^T = \mathbf{I}_n^T = \mathbf{I}_n.$$

(3) 注意到 $\mathbf{E}_{ij} = \mathbf{d}_i \mathbf{e}_j^T$, 其中 $\mathbf{d}_i \in F^m$, $\mathbf{e}_j \in F^n$ 均是空间中的基本单位向量(第 $i(j)$ 个分量为 1, 其余分量为 0), 所以

$$\begin{aligned} \mathbf{K}_{mn} &= \sum_{i=1}^m \sum_{j=1}^n ((\mathbf{d}_i \mathbf{e}_j^T) \otimes (\mathbf{d}_i \mathbf{e}_j^T)^T) = \sum_{i=1}^m \sum_{j=1}^n ((\mathbf{d}_i \mathbf{e}_j^T) \otimes (\mathbf{e}_j \mathbf{d}_i^T)) \\ &= \sum_{i=1}^m \sum_{j=1}^n ((\mathbf{e}_j^T \otimes \mathbf{d}_i) \otimes (\mathbf{d}_i^T \otimes \mathbf{e}_j)) = \sum_{j=1}^n \left(\mathbf{e}_j^T \otimes \left(\sum_{i=1}^m \mathbf{d}_i \otimes \mathbf{d}_i^T \right) \otimes \mathbf{e}_j \right) \\ &= \sum_{j=1}^n (\mathbf{e}_j^T \otimes \mathbf{I}_m \otimes \mathbf{e}_j). \end{aligned}$$

□

定理 6.12 设 $\mathbf{A} = (a_{ij}) \in F^{m \times n}$, 则

$$\text{Vec}(\mathbf{A}) = \mathbf{K}_{mn}^T \text{Vec}(\mathbf{A}^T).$$

证明 由于矩阵 $\{\mathbf{E}_{ij}, 1 \leq i \leq m, 1 \leq j \leq n\}$ 是空间 $F^{m \times n}$ 的基, 又 \mathbf{A} 的元素 a_{ij} 可表示为

$$a_{ij} = \mathbf{d}_i^T \mathbf{A} \mathbf{e}_j = \mathbf{e}_j^T \mathbf{A}^T \mathbf{d}_i,$$

$$\begin{aligned} \text{所以 } \mathbf{A} &= \sum_{j=1}^n \sum_{i=1}^m a_{ij} \mathbf{E}_{ij} = \sum_{j=1}^n \sum_{i=1}^m (\mathbf{d}_i^T \mathbf{A} \mathbf{e}_j) (\mathbf{d}_i \mathbf{e}_j^T) \\ &= \sum_{j=1}^n \sum_{i=1}^m \mathbf{d}_i (\mathbf{e}_j^T \mathbf{A}^T \mathbf{d}_i) \mathbf{e}_j^T. \end{aligned}$$

$$\begin{aligned} \text{从而 } \text{Vec}(\mathbf{A}) &= \text{Vec} \left(\sum_{j=1}^n \sum_{i=1}^m \mathbf{d}_i (\mathbf{e}_j^T \mathbf{A}^T \mathbf{d}_i) \mathbf{e}_j^T \right) = \text{Vec} \left(\sum_{j=1}^n \sum_{i=1}^m (\mathbf{d}_i \mathbf{e}_j^T) \mathbf{A}^T (\mathbf{d}_i \mathbf{e}_j^T) \right) \\ &= \text{Vec} \left(\sum_{j=1}^n \sum_{i=1}^m (\mathbf{E}_{ij} \mathbf{A}^T \mathbf{E}_{ij}) \right) = \sum_{j=1}^n \sum_{i=1}^m \text{Vec}(\mathbf{E}_{ij} \mathbf{A}^T \mathbf{E}_{ij}) \\ &= \sum_{j=1}^n \sum_{i=1}^m (\mathbf{E}_{ij}^T \otimes \mathbf{E}_{ij}) \text{Vec}(\mathbf{A}^T) = \mathbf{K}_{mn}^T \text{Vec}(\mathbf{A}^T). \end{aligned}$$

□

定理 6.13 设矩阵 $\mathbf{A} \in F^{m \times p}$, $\mathbf{B} \in F^{n \times q}$, 则

$$\mathbf{K}_{mn} (\mathbf{B} \otimes \mathbf{A}) \mathbf{K}_{pq}^T = \mathbf{A} \otimes \mathbf{B}.$$

证明 设 \mathbf{C} 为 $F^{q \times p}$ 空间中的任意矩阵, 则

$$\begin{aligned} \mathbf{K}_{mn} (\mathbf{B} \otimes \mathbf{A}) \mathbf{K}_{pq}^T (\text{Vec}(\mathbf{C})) &= \mathbf{K}_{mn} (\mathbf{B} \otimes \mathbf{A}) \text{Vec}(\mathbf{C}^T) = \mathbf{K}_{mn} \text{Vec}(\mathbf{AC}^T \mathbf{B}^T) \\ &= \text{Vec}((\mathbf{AC}^T \mathbf{B}^T)^T) = \text{Vec}(\mathbf{BCA}^T) \\ &= (\mathbf{A} \otimes \mathbf{B}) \text{Vec}(\mathbf{C}), \end{aligned}$$

即对任意 $C \in F^{p \times q}$, 有

$$\mathbf{K}_{mn}(\mathbf{B} \otimes \mathbf{A}) \mathbf{K}_{pq}^T(\text{Vec}(\mathbf{C})) = (\mathbf{A} \otimes \mathbf{B})(\text{Vec}(\mathbf{C}))$$

即

$$\mathbf{K}_{mn}(\mathbf{B} \otimes \mathbf{A}). \quad \square$$

习 题 六

1. 设 \mathbf{A}, \mathbf{B} 为使下列运算有意义的矩阵, 证明

$$(1) (\mathbf{A} \otimes \mathbf{B})^k = \mathbf{A}^k \otimes \mathbf{B}^k;$$

$$(2) \text{rank}(\mathbf{A} \otimes \mathbf{B}) = \text{rank}(\mathbf{A}) \cdot \text{rank}(\mathbf{B});$$

$$(3) \mathbf{A} \otimes \mathbf{B} = \mathbf{0} \text{ 当且仅当 } \mathbf{A} = \mathbf{0} \text{ 或 } \mathbf{B} = \mathbf{0}.$$

2. 设 $\mathbf{A} \in F^{n \times n}$ 为非奇异矩阵, 证明行列式 $|\mathbf{A} \otimes \mathbf{A}^{-1}| = |\mathbf{A}^{-1} \otimes \mathbf{A}| = 1$.

3. 设 $\mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{v}$ 为酉空间 $(\mathbf{C}^n | (\mathbf{x}, \mathbf{y}) = \mathbf{y}^H \mathbf{x})$ 中的元素, 证明

$$(\mathbf{x}, \mathbf{y})(\mathbf{u}, \mathbf{v}) = (\mathbf{x} \otimes \mathbf{u}, \mathbf{y} \otimes \mathbf{v}).$$

4. 设 \mathbf{A} 和 \mathbf{B} 为方阵, 证明

$$\text{tr}(\mathbf{A} \otimes \mathbf{B}) = \text{tr}(\mathbf{A}) \cdot \text{tr}(\mathbf{B}).$$

5. 设 $\mathbf{H} \in \mathbf{R}^{n \times n}$, 若 \mathbf{H} 满足条件(1) \mathbf{H} 的元素只取土 1, (2) $\mathbf{H}^T \mathbf{H} = n\mathbf{I}$, 则 \mathbf{H} 称为 Hadamard 矩阵. 证明若 $\mathbf{H}_1 \in \mathbf{R}^{n_1 \times n_1}$, $\mathbf{H}_2 \in \mathbf{R}^{n_2 \times n_2}$ 都是 Hadamard 矩阵, 则 $\mathbf{H}_1 \otimes \mathbf{H}_2$ 仍是 Hadamard 矩阵.

6. 设 $\mathbf{A} = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 2 & 1 & 0 \\ 3 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$, 求矩阵 \mathbf{E} , 使 $\mathbf{E}^T(\mathbf{A} \otimes \mathbf{B})\mathbf{E} = \mathbf{A} \circ \mathbf{B}$.

7. 举例说明当矩阵 \mathbf{A} 和 \mathbf{B} 的秩均为正数时, 矩阵 $\mathbf{A} \circ \mathbf{B}$ 的秩可以是 0.

8. 证明 $\text{rank}(\mathbf{A} \circ \mathbf{B}) \geq \text{rank}(\mathbf{A}) \cdot \text{rank}(\mathbf{B})$.

9. 设 $\mathbf{A} \in \mathbf{C}^{m \times m}$ 和 $\mathbf{B} \in \mathbf{C}^{n \times n}$ 是方的矩阵, 线性空间 $\mathbf{C}^{m \times n}$ 上线性变换 T_1 和 T_2 定义为

$$\forall X \in \mathbf{C}^{m \times n}, \quad T_1(X) = \mathbf{A}X\mathbf{B},$$

$$\forall X \in \mathbf{C}^{m \times n}, \quad T_2(X) = \mathbf{A}X + \mathbf{X}\mathbf{B}.$$

证明线性变换 T_1 和 T_2 关于空间 $\mathbf{C}^{m \times n}$ 中基 $\{\mathbf{E}_{ij}, i=1, 2, \dots, m; j=1, 2, \dots, n\}$ 的变换矩阵分别为 $\mathbf{B}^T \otimes \mathbf{A}$ 和 $(\mathbf{I}_n \otimes \mathbf{A} + \mathbf{B}^T \otimes \mathbf{I}_m)$.

10. 设 $\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} -3 & 4 \\ 0 & -1 \end{bmatrix}$, $\mathbf{D} = \begin{bmatrix} 0 & 5 \\ 2 & -9 \end{bmatrix}$, 求解关于 X 的矩阵方程:

$$\mathbf{AX} + \mathbf{XB} = \mathbf{D}.$$

11. 已知 $\mathbf{A} \in \mathbf{C}^{n \times n}$, μ 是常数, 讨论矩阵方程 $\mathbf{AX} - \mathbf{XA} = \mu\mathbf{X}$ 有惟一解的条件.

12. 设矩阵 $\mathbf{A} \in \mathbf{C}^{m \times m}$, $\mathbf{B} \in \mathbf{C}^{n \times n}$, $\mathbf{D} \in \mathbf{C}^{m \times n}$, 证明谱半径 $\rho(\mathbf{A}) \cdot \rho(\mathbf{B}) < 1$ 时, 方程:

$$\mathbf{X} = \mathbf{AXB} + \mathbf{D}$$

的解为

$$\mathbf{X} = \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{CB}^k.$$

13. 给定函数矩阵 $\mathbf{A}(t), \mathbf{B}(t)$, 设 $\mathbf{C}(t) = \mathbf{A}(t) \otimes \mathbf{B}(t)$, 证明 $\frac{d\mathbf{C}(t)}{dt} = \frac{d\mathbf{A}(t)}{dt} \otimes \mathbf{B} + \mathbf{A}$

$$(t) \otimes \frac{d\mathbf{B}(t)}{dt}.$$

* 第 7 章 非负矩阵介绍

在经济数学、概率论以及系统稳定性分析等方面,常遇到一类特殊的矩阵,即所谓非负矩阵.这些矩阵的元素皆是非负的,故有某些一般矩阵所没有的性质.

7.1 非负矩阵

定义 7.1 设 $A \in \mathbf{R}^{m \times n}$. 如果 A 的所有元素皆是非负的, 则称 A 为非负矩阵. 如果 A 的所有元素皆为正, 则称 A 为正矩阵.

对于矩阵 $A = (a_{ij})$, $B = (b_{ij}) \in \mathbf{R}^{m \times n}$, 我们约定:

$$\begin{aligned} A \geqslant B & \text{ 表示 } a_{ij} \geqslant b_{ij}, \quad (i=1, 2, \dots, m; j=1, 2, \dots, n). \\ A > B & \text{ 表示 } a_{ij} > b_{ij}, \end{aligned}$$

按照上面约定, $A \geqslant \mathbf{0}$ 表示 A 是非负矩阵, $A > \mathbf{0}$ 表示 A 是正矩阵.

非负矩阵满足以下性质:

(1) 设 $A, B \in \mathbf{R}^{m \times n}$. 如果 $A \geqslant \mathbf{0}$, $B \geqslant \mathbf{0}$, 则对任意非负实数 a, b , 有

$$aA + bB \geqslant \mathbf{0};$$

(2) 设 $A \in \mathbf{R}^{m \times n}$, $B \in \mathbf{R}^{n \times k}$. 如果 $A \geqslant \mathbf{0}$, $B \geqslant \mathbf{0}$, 则

$$AB \geqslant \mathbf{0};$$

(3) 设 $A \in \mathbf{R}^{m \times n}$, 则 $A \geqslant \mathbf{0}$ 当且仅当对任意 $x \in \mathbf{R}^n$, $x \geqslant \mathbf{0}$, 都有 $Ax \geqslant \mathbf{0}$;

(4) 设 $A, B \in \mathbf{R}^{n \times n}$. 如果 $A \geqslant B \geqslant \mathbf{0}$, 则对任意正整数 k , 有

$$A^k \geqslant B^k.$$

证明 只证明(4), 对 k 用数学归纳法.

当 $k=1$ 时, 结论成立.

假设结论对 $k-1$ 成立, 即

$$A^{k-1} \geqslant B^{k-1},$$

所以

$$A^{k-1} - B^{k-1} \geqslant \mathbf{0}.$$

又 $A \geqslant B \geqslant \mathbf{0}$, 结合(2), 有

$$\begin{aligned} A^k - B^k &= A^k - AB^{k-1} + AB^{k-1} - B^k \\ &= A(A^{k-1} - B^{k-1}) + (A - B)B^{k-1} \geqslant \mathbf{0}, \end{aligned}$$

$$A^k \geqslant B^k.$$

值得注意的是, 如果 $A \geqslant \mathbf{0}$ 且 $A \neq \mathbf{0}$, 则未必有 $A > \mathbf{0}$.

在第 5 章, 我们证明了任意一种矩阵范数都是它的谱半径的上界. 对于非负矩阵,

故

有更深刻的结果. 我们先给出一个一般的结果.

引理 7.1 设 $A \in \mathbf{R}^{n \times n}$, $\|A\|$ 是 $\mathbf{R}^{n \times n}$ 上的任意一种矩阵范数, 则

$$\rho(A) = \lim_{k \rightarrow \infty} \|A^k\|^{\frac{1}{k}}.$$

证明 由第 5 章例 7, 对任意正整数 k , 有

$$\rho(A^k) = [\rho(A)]^k,$$

又由定理 5.6, 有

$$\rho(A^k) \leq \|A^k\|,$$

所以

$$[\rho(A)]^k \leq \|A^k\|,$$

于是

$$\rho(A) \leq \|A^k\|^{\frac{1}{k}},$$

取极限, 得

$$\rho(A) \leq \lim_{k \rightarrow \infty} \|A^k\|^{\frac{1}{k}}.$$

另一方面, 对任意 $\epsilon > 0$, 令 $B = \frac{1}{\rho(A) + \epsilon} A$, 则 $\rho(B) < 1$. 因而 $\lim_{k \rightarrow \infty} B^k = \mathbf{0}$. 于是存在正整数 N , 当 $k > N$ 时, 有

$$\|B^k\| < 1.$$

即

$$\frac{1}{[\rho(A) + \epsilon]^k} \|A^k\| < 1,$$

于是

$$\|A^k\|^{\frac{1}{k}} < \rho(A) + \epsilon,$$

取极限, 再由 ϵ 的任意性, 有

$$\lim_{k \rightarrow \infty} \|A^k\|^{\frac{1}{k}} \leq \rho(A),$$

故

$$\rho(A) = \lim_{k \rightarrow \infty} \|A^k\|^{\frac{1}{k}}.$$

定理 7.1 设 $A, B \in \mathbf{R}^{n \times n}$, 如果 $A \geq B \geq \mathbf{0}$, 则

$$\rho(A) \geq \rho(B).$$

证明 因为 $A \geq B \geq \mathbf{0}$, 所以对任意正整数 k , 有

$$A^k \geq B^k \geq \mathbf{0},$$

于是

$$\|A^k\|_1 \geq \|B^k\|_1,$$

故

$$\|A^k\|_1^{\frac{1}{k}} \geq \|B^k\|_1^{\frac{1}{k}},$$

取极限, 有

$$\lim_{k \rightarrow \infty} \|A^k\|_1^{\frac{1}{k}} \geq \lim_{k \rightarrow \infty} \|B^k\|_1^{\frac{1}{k}}.$$

由引理 7.1, 有

$$\rho(A) \geq \rho(B).$$

□

推论 设 $A \in \mathbf{R}^{n \times n}$ 是非负矩阵, A_k 为 A 的任意 k 阶主子式, 则

$$\rho(A) \geq \rho(A_k).$$

特别地, 有 $\rho(A) \geq a_{ii}$, $i=1, 2, \dots, n$.

证明 将 A 中 A_k 以外的所有元素换为零而得到的矩阵记为 B , 则

$$A \geq B \geq \mathbf{0},$$

由定理 7.1, 有

$$\rho(\mathbf{A}) \geq \rho(\mathbf{B}) = \rho(\mathbf{A}_k).$$

引理 7.2 设 $\mathbf{A}=(a_{ij}) \in \mathbf{R}^{n \times n}$ 是非负矩阵, 则

(1) 当 \mathbf{A} 的各个行和是同一常数 a 时, 有

$$\rho(\mathbf{A}) = \|\mathbf{A}\|_\infty = a.$$

(2) 当 \mathbf{A} 的各个列和是同一常数 a 时, 有

$$\rho(\mathbf{A}) = \|\mathbf{A}\|_1 = a.$$

证明 (1) 因为

$$a = \sum_{j=1}^n |a_{ij}| = \sum_{j=1}^n a_{ij},$$

所以 $\|\mathbf{A}\|_\infty = a$. 由定理 5.6, 有

$$\rho(\mathbf{A}) \leq a,$$

另一方面, 令 $\mathbf{x}=(1 \ 1 \ \cdots \ 1)^T$, 则 $\mathbf{Ax}=\mathbf{ax}$, 即 a 是 \mathbf{A} 的一个特征值, 故

$$\rho(\mathbf{A}) \geq a.$$

(2) 考虑 \mathbf{A}^T , 因为 $\rho(\mathbf{A})=\rho(\mathbf{A}^T)$, 且 \mathbf{A}^T 各个行和为同一常数 a , 由(1)知 $\rho(\mathbf{A}^T)=\|\mathbf{A}^T\|_\infty$. 但 $\|\mathbf{A}^T\|_\infty=\|\mathbf{A}\|_1$, 故

$$\rho(\mathbf{A}) = \|\mathbf{A}\|_1 = a.$$

定理 7.2 (Frobenius 定理) 设 $\mathbf{A} \in \mathbf{R}^{n \times n}$ 是非负矩阵, 则

$$(1) \quad \min_i \sum_{j=1}^n a_{ij} \leq \rho(\mathbf{A}) \leq \max_i \sum_{j=1}^n a_{ij}, \quad (7.1)$$

$$\min_j \sum_{i=1}^n a_{ij} \leq \rho(\mathbf{A}) \leq \max_j \sum_{i=1}^n a_{ij}; \quad (7.2)$$

(2) 对任意 n 个正数 x_1, x_2, \dots, x_n , 有

$$\min_i \frac{1}{x_i} \sum_{j=1}^n a_{ij} x_j \leq \rho(\mathbf{A}) \leq \max_i \frac{1}{x_i} \sum_{j=1}^n a_{ij} x_j, \quad (7.3)$$

$$\min_j x_j \sum_{i=1}^n \frac{a_{ij}}{x_i} \leq \rho(\mathbf{A}) \leq \max_j x_j \sum_{i=1}^n \frac{a_{ij}}{x_i}; \quad (7.4)$$

(3) 当 $\mathbf{Ax}=\lambda\mathbf{x}$, 且 $\mathbf{x}=(x_1 \ x_2 \ \cdots \ x_n)^T > 0$ 时, $\lambda=\rho(\mathbf{A})$, 即非负矩阵 \mathbf{A} 的正特征向量对应的特征值就是 \mathbf{A} 的谱半径.

证明 (1) 令

$$a = \min_i \sum_{j=1}^n a_{ij}, b_{ij} = a a_{ij} (\sum_{j=1}^n a_{ij})^{-1}, \quad \mathbf{B} = (b_{ij}),$$

则

$$\mathbf{B} \geq \mathbf{0}, a_{ij} \geq b_{ij} \quad (i, j = 1, 2, \dots, n),$$

于是

$$\mathbf{A} \geq \mathbf{B} \geq \mathbf{0}.$$

又

$$\sum_{j=1}^n b_{ij} = a \quad (i = 1, 2, \dots, n),$$

由引理 7.2 有 $\rho(\mathbf{B}) = a$, 再由定理 6.1, 有

$$\rho(\mathbf{A}) \geq \rho(\mathbf{B}) = a = \min_i \sum_{j=1}^n a_{ij}.$$

同理可证 $\rho(\mathbf{A}) \leq \max_i \sum_{j=1}^n a_{ij}$, 这便证得(7.1)式.

在(7.1)式中, 用 \mathbf{A}^T 替换 \mathbf{A} , 并注意到 $\rho(\mathbf{A}) = \rho(\mathbf{A}^T)$, 便得(7.2)式.

(2) 令 $\mathbf{S} = \text{diag}(x_1 \quad x_2 \quad \cdots \quad x_n)$, 其中 $x_i > 0, i = 1, 2, \dots, n$, 则当 $\mathbf{A} \geq \mathbf{0}$ 时, $\mathbf{S}^{-1} \mathbf{A} \mathbf{S} = (x_i^{-1} a_{ij} x_j) \geq \mathbf{0}$, 且

$$\rho(\mathbf{A}) = \rho(\mathbf{S}^{-1} \mathbf{A} \mathbf{S}).$$

将(7.1)式与(7.2)式应用到 $\mathbf{S}^{-1} \mathbf{A} \mathbf{S}$ 及其转置矩阵上, 便得到(7.3)式与(7.4)式.

(3) 由 $\mathbf{A} \geq \mathbf{0}, \mathbf{x} > \mathbf{0}$ 且 $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, 有 $\lambda \geq 0$, 且

$$\lambda = x_i^{-1} \sum_{j=1}^n a_{ij} x_j \quad (i = 1, 2, \dots, n),$$

由(7.3)式, 有 $\lambda = \min_i x_i^{-1} \sum_{j=1}^n a_{ij} x_j \leq \rho(\mathbf{A}) \leq \max_i x_i^{-1} \sum_{j=1}^n a_{ij} x_j = \lambda$,

故

$$\lambda = \rho(\mathbf{A}).$$

□

7.2 正 矩 阵

正矩阵中最重要的结果是本节将要介绍的配朗(Perron)定理. 作为准备, 我们先介绍 Wielandt 引理.

定理 7.3(Wielandt 引理) 设 $\mathbf{A} \in \mathbf{R}^{n \times n}$ 是正矩阵, 则

- (1) \mathbf{A} 的谱半径 $\rho(\mathbf{A})$ 是 \mathbf{A} 的一个正特征值, 且有正的特征向量;
- (2) \mathbf{A} 的所有其他特征值的模都小于 $\rho(\mathbf{A})$.

证明 (1) 设 λ 是 \mathbf{A} 的模最大的特征值, 即 $|\lambda| = \rho(\mathbf{A})$, $\mathbf{x} = (x_1 \quad x_2 \quad \cdots \quad x_n)^T$ 是对应的特征向量, 记 $|\mathbf{x}| = (|x_1| \quad |x_2| \quad \cdots \quad |x_n|)^T$, 下面证明 $|\mathbf{x}|$ 是 \mathbf{A} 的关于特征值 $|\lambda| = \rho(\mathbf{A})$ 的正特征向量.

因为 $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, 所以

$$\lambda x_i = \sum_{j=1}^n a_{ij} x_j \quad (i = 1, 2, \dots, n),$$

于是 $\rho(\mathbf{A}) |x_i| = |\lambda x_i| \leq \sum_{j=1}^n a_{ij} |x_j| \quad (i = 1, 2, \dots, n)$,

由此可得到

$$\rho(\mathbf{A}) |\mathbf{x}| \leq \mathbf{A} |\mathbf{x}|,$$

亦即

$$(\mathbf{A} - \rho(\mathbf{A}) \mathbf{I}) |\mathbf{x}| \geq \mathbf{0},$$

如果 $(\mathbf{A} - \rho(\mathbf{A}) \mathbf{I}) |\mathbf{x}| \neq \mathbf{0}$, 令 $\mathbf{A} - \rho(\mathbf{A}) \mathbf{I} |\mathbf{x}| = \mathbf{z}$, 则 \mathbf{z} 为非零的非负向量, 因为 \mathbf{A} 为正矩阵, 所以 $\mathbf{A}\mathbf{z} > \mathbf{0}$. 又由于 $\mathbf{A} |\mathbf{x}| > \mathbf{0}$, 因此存在数 $\epsilon > 0$, 使得 $\mathbf{A}\mathbf{z} \geq \epsilon \mathbf{A} |\mathbf{x}|$. 由

$$(A - \rho(A)I) |x| = z,$$

可得到

$$A(A - \rho(A)I) |x| = Az,$$

于是

$$\begin{aligned} A^2 |x| &= Az + \rho(A)A |x| \geq \epsilon A |x| + \rho(A)A |x| \\ &= (\epsilon + \rho(A))A |x|, \end{aligned}$$

故

$$(\epsilon + \rho(A))^{-1} A^2 |x| \geq A |x|.$$

再令 $B = (\epsilon + \rho(A))^{-1} A$, 便有

$$BA |x| \geq A |x|,$$

$$B^2 A |x| \geq BA |x| \geq A |x|,$$

依此类推, 对任意正整数 k , 都有

$$B^k A |x| \geq A |x|.$$

又 B 的谱半径为 $(\epsilon + \rho(A))^{-1} \rho(A) < 1$, 故 $\lim_{k \rightarrow \infty} B^k = \mathbf{0}$, 从而 $A |x| \leq \mathbf{0}$ 矛盾, 故 $z = \mathbf{0}$, 即 $(A - \rho(A)I) |x| = \mathbf{0}$, 这就证明了(1).

(2) 设除了 $|\lambda| = \rho(A)$ 以外, 还有 $|u| = \rho(A)$, 且 u 对应的特征向量为 $\mathbf{y} = (y_1 \ y_2 \ \dots \ y_n)^T$. 由(1)有

$$|Ay| = |u\mathbf{y}| = |u| \cdot |\mathbf{y}| = A|\mathbf{y}|,$$

于是

$$|\sum_{j=1}^n a_{ij} y_j| = \sum_{j=1}^n |a_{ij}| |y_j|, \quad i = 1, 2, \dots, n.$$

这表明三角不等式中等式必须成立, 因此复数 $a_{ij} y_j$ 有相同的幅角 θ . 但所有 a_{ij} 都是正实数, 于是所有 y_j 有相同的幅角 θ , 即有

$$y_j = |y_j| e^{i\theta} \quad (j = 1, 2, \dots, n),$$

于是

$$\mathbf{y} = e^{i\theta} |\mathbf{y}|.$$

但 $|\mathbf{y}|$ 是 A 的对应于特征值 $|u| = \rho(A)$ 的特征向量, 所以 \mathbf{y} 也是 A 的对应于特征值 $\rho(A)$ 的特征向量. 而 \mathbf{y} 又是 A 的对应于特征值 u 的特征向量, 且不同特征值不可能有相同的特征向量, 故

$$u = \rho(A).$$

这样便证明了 A 的不同于 $\rho(A)$ 的特征值的模一定小于 $\rho(A)$. \square

定理 7.4 设 $A \in \mathbb{R}^{n \times n}$ 是正矩阵, 则矩阵 $B = (\rho(A))^{-1} A$ 是幂收敛的, 即 $\{B^k\}$ 收敛.

证明 显然 $\rho(B) = 1$. 由定理 7.3 知 1 是 B 的一个特征值, 且其他特征值的模都小于 1. 由 Jordan 分解知, 存在可逆阵 P , 使得 $P^{-1}AP = J$. 如能证明特征值 1 对应的 Jordan 块是一阶的, 则 B 便是幂收敛的.

设 \mathbf{y} 是 B 对应于特征值 1 的正特征向量, 即 $B\mathbf{y} = \mathbf{y} > \mathbf{0}$, 则对任意正整数 k , 都有

$$B^k \mathbf{y} = \mathbf{y}.$$

令

$$y_s = \max_i y_i, \quad y_t = \min_i y_i,$$

则

$$y_s \geq y_i = \sum_{j=1}^n (B^k)_{ij} y_j \geq (B^k)_{ii} y_j \geq (B^k)_{ii} y_t.$$

其中 $(\mathbf{B}^k)_{ij}$ 表示 \mathbf{B}^k 的第 i 行第 j 列的元素. 于是

$$\mathbf{0} < (\mathbf{B}^k)_{ij} \leqslant y_i / y_j.$$

这就证明了 \mathbf{B} 是幂有界的, 这意味着 \mathbf{B} 的 Jordan 标准形中, 对应于特征值 1 的 Jordan 块只能是一阶的.

定理 7.5 正矩阵 A 的特征值 $\rho(A)$ 是一个单根.

证明 注意到 $\rho(A)$ 作为 A 的特征值的重数等于 1 作为 $\mathbf{B} = (\rho(A))^{-1} A$ 的特征值的重数, 所以我们对 \mathbf{B} 来讨论.

由定理 7.4 知, \mathbf{B} 的对应于特征值 1 的 Jordan 块都是一阶的, 所以 \mathbf{B} 的 Jordan 标准形为

$$\mathbf{J} = \begin{pmatrix} \mathbf{I}_{n_0} & & & \\ & \mathbf{J}_{n_1}(\lambda_1) & & \\ & & \ddots & \\ & & & \mathbf{J}_{n_t}(\lambda_t) \end{pmatrix},$$

其中 n_i 为 1 作为 \mathbf{B} 的特征值的重数, 且 $|\lambda| < 1 (i=1, 2, \dots, t)$. 下证 $n_0 = 1$, 由

$$\mathbf{J} - \mathbf{I} = \begin{pmatrix} \mathbf{0} & & & \\ & \mathbf{J}_{n_1}(\lambda_1 - 1) & & \\ & & \ddots & \\ & & & \mathbf{J}_{n_t}(\lambda_t - 1) \end{pmatrix},$$

因为 $|\lambda| < 1$, 所以 $\mathbf{J}_{n_i}(\lambda - 1) (i=1, 2, \dots, t)$ 为满秩矩阵, 故

$$n_0 = \dim N(\mathbf{J} - \mathbf{I}) = \dim N(\mathbf{B} - \mathbf{I}),$$

设 $x > \mathbf{0}$ 是 \mathbf{B} 关于特征值 1 的特征向量, 则 $\mathbf{B}x = x$, 即 $(\mathbf{B} - \mathbf{I})x = \mathbf{0}$, 故 $x \in N(\mathbf{B} - \mathbf{I})$, 于是 $n_0 = \dim N(\mathbf{B} - \mathbf{I}) \geq 1$.

如果 $n_0 > 1$, 则存在与 x 线性无关的向量 $y \in \mathbf{R}^n$, 满足 $\mathbf{B}y = y$, 令

$$\max_i(y_i/x_i) = y_j/x_j = a,$$

则

$$ax \geqslant y,$$

因为 x, y 线性无关, 所以 $ax - y \neq \mathbf{0}$, 于是

$$\mathbf{B}(ax - y) > \mathbf{0},$$

$$a\mathbf{B}x - \mathbf{B}y > \mathbf{0},$$

又因为 $\mathbf{B}x = x, \mathbf{B}y = y$, 所以

$$ax - y > \mathbf{0}.$$

考虑其分量, 便有

$$ax_i > y_i \quad (i = 1, 2, \dots, n),$$

$$a > y_i/x_i,$$

这与 a 的取法矛盾, 故 $n_0 = 1$. □

将定理 7.3 与定理 7.5 合在一起, 得到经典的 Perron 定理.

定理 7.6(Perron 定理) 设 $A \in \mathbf{R}^{n \times n}$ 是正矩阵, 则 $\rho(A)$ 是 A 的一个单特征值, 并且对应于 $\rho(A)$ 有正的特征向量, 而 A 的其他的不等于 $\rho(A)$ 的特征值的模都小于 $\rho(A)$.

1912 年, Frobenius 把上述定理推广到不可约的非负矩阵上, 结论更为圆满, 所以历史上又叫做 Perron–Frobenius 定理, 这里不作介绍.

Perron 定理有许多应用, 一个优美而有效的应用是: 利用占优非负矩阵的谱半径和主对角元得到了矩阵 A 的特征值包含区域.

定理 7.7(Ky Fan 定理) 设 $A = (a_{ij}) \in \mathbf{C}^{n \times n}$, λ 为 A 的任意特征值, 非负矩阵 $B = (b_{ij}) \in \mathbf{R}^{n \times n}$ 满足

$$b_{ij} \geq |a_{ij}| \quad (i, j = 1, 2, \dots, n),$$

则存在某个 i , 使得

$$|\lambda - a_{ii}| \leq \rho(B) - b_{ii}.$$

证明 若 B 是正矩阵, 由 Perron 定理, 存在正向量 $x = (x_1 \ x_2 \ \cdots \ x_n)^T > \mathbf{0}$, 使得 $Bx = \rho(B)x$.

令 $S = \text{diag}(x_1 \ x_2 \ \cdots \ x_n)$, $C = (c_{ij}) = S^{-1}AS$,

由盖尔斯果林圆盘定理, 存在某个 i 使得

$$|\lambda - c_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |c_{ij}|.$$

由 $C = S^{-1}AS$, 有

$$c_{ij} = x_i^{-1} a_{ij} x_j,$$

所以 $|\lambda - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^n x_i^{-1} |a_{ij}| x_j \leq \sum_{\substack{j=1 \\ j \neq i}}^n x_i^{-1} b_{ij} x_j = \sum_{j=1}^n x_i^{-1} b_{ij} x_j - b_{ii}.$

由 $Bx = \rho(B)x$, 有

$$\sum_{j=1}^n b_{ij} x_j = \rho(B) x_i,$$

故

$$|\lambda - a_{ii}| \leq \rho(B) - b_{ii}.$$

若 B 不是正矩阵, 令

$$b_{ij}^{(k)} = b_{ij} + \frac{1}{k} \quad (i, j = 1, 2, \dots, n),$$

则 $B_k = (b_{ij}^{(k)})$ 便是正矩阵. 由前面的证明, 对任意正整数 k , 都存在某个 i_k , 使得

$$|\lambda - a_{i_k i_k}| \leq \rho(B_k) - b_{i_k i_k}^{(k)}.$$

由于 $1 \leq i \leq n$, 所以当 $k = 1, 2, \dots$ 时, i_k 必有重复. 事实上, i_k 将无限次地取 1 到 n 间的某个正整数 i , 把取 i 的 i_k 的下脚标 k 依次记为 $k_1, k_2, \dots, k_t, \dots$ 于是有

$$|\lambda - a_{ii}| \leq \rho(B_{k_t}) - b_{ii}^{(k_t)},$$

又注意到矩阵的特征值及其模为矩阵元素的连续函数, 故

$$\lim_{t \rightarrow \infty} \rho(B_{k_t}) = \rho(B), \quad \lim_{t \rightarrow \infty} b_{ii}^{(k_t)} = b_{ii}.$$

从而证明了

$$|\lambda - a_{ii}| \leq \rho(\mathbf{B}) - b_{ii}.$$

7.3 素 矩 阵

上一节我们简单地介绍了正矩阵的某些性质,现在讨论更为一般的非负矩阵,即素矩阵.将正矩阵的某些性质推广到素矩阵上.

定义 7.2 设 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 是非负矩阵.如果存在正整数 m ,使得 \mathbf{A}^m 为正矩阵,则称 \mathbf{A} 是素矩阵或本原矩阵或幂正矩阵.

显然正矩阵是素矩阵,但反过来不对.如取

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix},$$

则 \mathbf{A} 不是正矩阵,但 $\mathbf{A}^4 > \mathbf{0}$,所以 \mathbf{A} 是素矩阵.

由定义不难验证,素矩阵满足以下性质:

- (1) 如果 \mathbf{A} 是素矩阵,则 \mathbf{A}^T 是素矩阵;
- (2) 如果 \mathbf{A} 是素矩阵,则对任意正整数 k , \mathbf{A}^k 是素矩阵;
- (3) 如果 \mathbf{A} 是素矩阵,则对任意非负矩阵 \mathbf{B} , $\mathbf{A} + \mathbf{B}$ 是素矩阵.

证明 只证明(3),其他的留给读者.

因为 \mathbf{A} 是素矩阵,所以存在正整数 m ,使得 $\mathbf{A}^m > \mathbf{0}$,而

$$(\mathbf{A} + \mathbf{B})^m = \mathbf{A}^m + \mathbf{C},$$

其中 $\mathbf{C} \geq \mathbf{0}$,故 $(\mathbf{A} + \mathbf{B})^m > \mathbf{0}$,即 $\mathbf{A} + \mathbf{B}$ 是素矩阵. □

下面我们将正矩阵的 Perron 定理推广到素矩阵上.

定理 7.8 设 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 为素矩阵,则

- (1) $\rho(\mathbf{A})$ 为 \mathbf{A} 的具有正特征向量的特征值;
- (2) \mathbf{A} 的特征值 $\rho(\mathbf{A})$ 是一个单根;
- (3) \mathbf{A} 的其他特征值的模小于 $\rho(\mathbf{A})$.

证明 (1)由定义,存在正整数 m ,使得 $\mathbf{A}^m > \mathbf{0}$,由定理 6.3,存在 $\mathbf{x} > \mathbf{0}$,使得

$$\mathbf{A}^m \mathbf{x} = \rho(\mathbf{A}^m) \mathbf{x}.$$

$$\bar{\mathbf{x}} = \sum_{i=0}^{m-1} (\rho(\mathbf{A}))^{-i} \mathbf{A}^i \mathbf{x},$$

则 $\bar{\mathbf{x}} > \mathbf{0}$,且

$$\begin{aligned} \mathbf{A} \bar{\mathbf{x}} &= \sum_{i=0}^{m-1} (\rho(\mathbf{A}))^{-i} \mathbf{A}^{i+1} \mathbf{x} \\ &= \sum_{i=0}^{m-2} (\rho(\mathbf{A}))^{-i} \mathbf{A}^{i+1} \mathbf{x} + (\rho(\mathbf{A}))^{1-m} \mathbf{A}^m \mathbf{x} \end{aligned}$$

$$= \sum_{i=1}^{m-1} (\rho(\mathbf{A}))^{1-i} \mathbf{A}^i \mathbf{x} + (\rho(\mathbf{A}))^{1-m} \rho(\mathbf{A}^m) \mathbf{x},$$

又对任意正整数 k , 有 $\rho(\mathbf{A}^k) = (\rho(\mathbf{A}))^k$, 因此

$$\begin{aligned}\mathbf{A}\bar{\mathbf{x}} &= \rho(\mathbf{A}) \left(\sum_{i=1}^{m-1} (\rho(\mathbf{A}))^{-i} \mathbf{x} + \mathbf{x} \right) \\ &= \rho(\mathbf{A}) \sum_{i=0}^{m-1} (\rho(\mathbf{A}))^{-i} \mathbf{x} = \rho(\mathbf{A}) \bar{\mathbf{x}}.\end{aligned}$$

这说明 $\rho(\mathbf{A})$ 是 \mathbf{A} 的具有正特征向量 $\bar{\mathbf{x}}$ 的特征值.

又由 $\mathbf{A}\bar{\mathbf{x}} = \rho(\mathbf{A})\bar{\mathbf{x}}$, 有

$$\mathbf{A}^m \bar{\mathbf{x}} = (\rho(\mathbf{A}))^m \bar{\mathbf{x}} = \rho(\mathbf{A}^m) \bar{\mathbf{x}}.$$

这说明 $\mathbf{x}, \bar{\mathbf{x}}$ 都是 \mathbf{A}^m 的对应于单特征值 $\rho(\mathbf{A}^m)$ 的特征向量, 所以 \mathbf{x} 与 $\bar{\mathbf{x}}$ 是线性相关的, 从而 \mathbf{x} 也是 \mathbf{A} 的对应于特征值 $\rho(\mathbf{A})$ 的正特征向量.

(2) 注意到 $\rho(\mathbf{A})$ 作为 \mathbf{A} 的特征值的重数(不超过 $(\rho(\mathbf{A}))^m$)作为 \mathbf{A}^m 的特征值的重数, 而 $(\rho(\mathbf{A}))^m = \rho(\mathbf{A}^m)$ 是正矩阵 \mathbf{A}^m 的单重特征值, 所以 $l=1$.

(3) 如果 \mathbf{A} 有特征值 λ 满足 $|\lambda| = \rho(\mathbf{A})$, 则存在非零向量 \mathbf{y} , 使得 $\mathbf{A}\mathbf{y} = \lambda\mathbf{y}$. 于是

$$\mathbf{A}^m \mathbf{y} = \lambda^m \mathbf{y},$$

$$|\lambda^m| = |\lambda|^m = (\rho(\mathbf{A}))^m = \rho(\mathbf{A}^m).$$

又 λ^m 是 \mathbf{A}^m 的特征值, \mathbf{A}^m 是正矩阵, 故 $\lambda^m = \rho(\mathbf{A}^m)$. 于是

$$\mathbf{A}^m \mathbf{y} = \rho(\mathbf{A}^m) \mathbf{y}.$$

这说明 \mathbf{x}, \mathbf{y} 都是 \mathbf{A}^m 对应于特征值 $\rho(\mathbf{A}^m)$ 的特征向量, 但 $\rho(\mathbf{A}^m)$ 是 \mathbf{A}^m 的单特征值, 所以 \mathbf{y} 与 \mathbf{x} 线性相关. 由(1)知 \mathbf{x} 是 \mathbf{A} 的对应于特征值 $\rho(\mathbf{A})$ 的特征向量, 所以 \mathbf{y} 也是 \mathbf{A} 的对应于特征值 $\rho(\mathbf{A})$ 的特征向量, 故必有 $\lambda = \rho(\mathbf{A})$.

给定非负矩阵 \mathbf{A} , 按照定义如要判定 \mathbf{A} 是否为素矩阵, 则要计算 \mathbf{A}^m . 自然要问: 使得 $\mathbf{A}^m > \mathbf{0}$ 的正整数 m 有什么特性? 为此, 我们引入素数指数的概念.

定义 7.3 素矩阵 \mathbf{A} 的使得 $\mathbf{A}^m > \mathbf{0}$ 的最小正整数 m 称为 \mathbf{A} 的素数指数或本原指数, 记为 $\gamma(\mathbf{A})$.

我们不加证明地给出下面关于素数指数 $\gamma(\mathbf{A})$ 的一些上界.

定理 7.9 设 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 是素矩阵, 且对某个正整数 h 和矩阵

$$\mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^h,$$

至少有 $d (> 0)$ 个正对角元, 则有

$$\gamma(\mathbf{A}) \leqslant n - d + (n-1)h.$$

将此定理用于本节开头的非负矩阵

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix},$$

取 $h=1$, 则 $d=1$, 而 $n=3$, 所以

$$\gamma(\mathbf{A}) \leqslant 3 - 1 + (3 - 1) = 4.$$

推论 设 $\mathbf{A} \in \mathbf{R}^{n \times n}$ 是素矩阵, 并且 a_{ij} 与 a_{ji} 同时为正或同时为零, 则

$$\gamma(\mathbf{A}) \leqslant 2(n-1).$$

证明 首先可推出 \mathbf{A}^2 的对角元全为正, 从而取 $h=2$, 此时 $d=n$. 由定理 7.9, 有

$$\gamma(\mathbf{A}) \leqslant n - n + 2(n-1) = 2(n-1).$$

定理 7.10 设 $\mathbf{A} \in \mathbf{R}^{n \times n}$ 是素矩阵, 则

$$\gamma(\mathbf{A}) \leqslant (n-1)^2 + 1.$$

7.4 M 矩 阵

本节所要介绍的 M 矩阵, 经常出现在诸如偏微分方程的有限差分法和有限元素法、经济学中的投入产出法、运筹学中的线性余问题以及概率统计的 Markov 过程等不同的科技领域中. M 矩阵这个术语是 Ostrowski 在 1937 年首先提出的, 它有许多等价的定义.

定义 7.4 设 $\mathbf{A}=s\mathbf{I}-\mathbf{B}$ 为 n 阶实矩阵, 且 $\mathbf{B} \geqslant \mathbf{0}, s > 0$, 那么, 若 $s \geqslant \rho(\mathbf{B})$, 则称 \mathbf{A} 为 M 矩阵; 若 $s > \rho(\mathbf{B})$, 则称 \mathbf{A} 为非奇异 M 矩阵.

记 $\mathbf{Z}^{n \times n} = \{\mathbf{A} = (a_{ij}) \in \mathbf{R}^{n \times n}, \text{ 当 } i \neq j \text{ 时 } a_{ij} \leqslant 0; i, j = 1, 2, \dots, n\}.$

定理 7.11 设 $\mathbf{A} = (a_{ij}) \in \mathbf{R}^{n \times n}$, 则以下各条等价:

- (1) \mathbf{A} 的所有主子式大于 0;
- (2) 若对角矩阵 $\mathbf{D} \geqslant \mathbf{0}$, 则 $\mathbf{A} + \mathbf{D}$ 的所有主子式不为零;
- (3) $\forall \mathbf{0} \neq \mathbf{x} = (x_1 \quad x_2 \quad \cdots \quad x_n)^T \in \mathbf{R}^n$, $\mathbf{y} = \mathbf{Ax} = (y_1 \quad y_2 \quad \cdots \quad y_n)^T$, 则存在 i , 使 $x_i y_i > 0$;
- (4) $\forall \mathbf{0} \neq \mathbf{x} \in \mathbf{R}^n$, 存在正对角矩阵 \mathbf{D}_x , 使 $\mathbf{x}^T \mathbf{D}_x \mathbf{Ax} > 0$;
- (5) $\forall \mathbf{0} \neq \mathbf{x} \in \mathbf{R}^n$, 存在非负对角矩阵 \mathbf{H}_x , 使 $\mathbf{x}^T \mathbf{H}_x \mathbf{Ax} > 0$;
- (6) \mathbf{A} 的每一个主子矩阵的任意实特征值都大于 0;
- (7) $\forall \mathbf{0} \neq \mathbf{x} \in \mathbf{R}^n$, 且 $\mathbf{z} = \mathbf{A}^T \mathbf{x} = (z_1 \quad z_2 \quad \cdots \quad z_n)^T$, 存在 k , 使得 $\mathbf{x}_k \mathbf{z}_k > 0$;
- (8) 对每个 n 阶符号矩阵 \mathbf{S} (即 \mathbf{S} 为对角矩阵, 且对角元素为 1 或 -1), 由 $\mathbf{SA}^T \mathbf{Sz} \leqslant \mathbf{0}, \mathbf{z} \geqslant \mathbf{0}$ 可推出 $\mathbf{z} = \mathbf{0}$.

证明 (1) \Rightarrow (2) 对 $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_n) \geqslant \mathbf{0}$, 当把 $\mathbf{A} + \mathbf{D}$ 的任意一个 k 阶主子式拆成 2^k 个行列式之和时, 由于每一个行列式均非负, 且至少有一个为正, 故此 k 阶主子式大于零(当然不等于零).

(2) \Rightarrow (3) 设存在 $\mathbf{x} = (x_1 \quad x_2 \quad \cdots \quad x_n)^T \neq \mathbf{0}$ 使 $\mathbf{y} = \mathbf{Ax} = (y_1 \quad y_2 \quad \cdots \quad y_n)^T$, 对一切 i , 均有 $x_i y_i \leqslant 0$.

设 \mathbf{x} 的分量中不为零的元是 x_{i_1}, \dots, x_{i_m} , 其中 $i_1 < \dots < i_m$, 令 $\mathbf{B} = \mathbf{A} \begin{pmatrix} i_1 & \cdots & i_m \\ \cdots & \cdots & \cdots \\ i_1 & \cdots & i_m \end{pmatrix}$ 为 \mathbf{A} 的相应的主子矩阵, $\mathbf{z} = (x_{i_1} \ \cdots \ x_{i_m})^T$, $\mathbf{u} = (y_{i_1} \ \cdots \ y_{i_m})^T$, 则 $\mathbf{u} = \mathbf{Bz}$.

因为 $x_i y_i \leq 0$, 所以存在非负的对角矩阵 $\mathbf{D} = \text{diag}(d_{i_1} \ \cdots \ d_{i_m})$, 使 $\mathbf{u} = -\mathbf{Dz}$, 从而 $(\mathbf{B} + \mathbf{D})\mathbf{z} = \mathbf{0}$. 由(2)的假设知 $|\mathbf{B} + \mathbf{D}| \neq 0$, 故 $\mathbf{z} = \mathbf{0}$, 矛盾.

(3) \Rightarrow (4) 取 $\mathbf{x}^T = (x_1 \ \cdots \ x_n)^T$, $\mathbf{y} = \mathbf{Ax} = (y_1 \ \cdots \ y_n)^T$, 存在 i , 使 $x_i y_i > 0$. 从而存在充分小的 $\epsilon > 0$, 使 $x_i y_i + \epsilon \sum_{j \neq i} x_j y_j > 0$. 令 $\mathbf{D} = \text{diag}(\epsilon \ \cdots \ \epsilon \ 1 \ \epsilon \ \cdots \ \epsilon)$ (其中 1 位于第 i 个), 则 $\mathbf{D} > \mathbf{0}$, $\mathbf{x}^T \mathbf{D} \mathbf{Ax} = \mathbf{x}^T \mathbf{Dy} = x_i y_i + \epsilon \sum_{j \neq i} x_j y_j > 0$.

(4) \Rightarrow (5) 显然.

(5) \Rightarrow (6) 设 $\mathbf{B} = \mathbf{A} \begin{pmatrix} i_1 & \cdots & i_k \\ \cdots & \cdots & \cdots \\ i_1 & \cdots & i_k \end{pmatrix}$ 为 \mathbf{A} 的任一主子阵, 其中 $i_1 < \cdots < i_k$. 设 λ 是 \mathbf{B}

的任一实特征值, α 是对应的实特征向量, $\mathbf{B}\alpha = \lambda\alpha$, $\alpha = (x_{i_1} \ \cdots \ x_{i_k})^T$. 令 $\mathbf{y} = (y_1 \ \cdots \ y_n)^T$, 其中 $y_{i_j} = x_{i_j}$, $j = 1, 2, \dots, k$. 其余 y_i 都取零, 则 $\mathbf{y} \neq \mathbf{0}$. 由(5)的假设, 存在 $\mathbf{H} \geq \mathbf{0}$, 使 $\mathbf{y}^T \mathbf{H} \mathbf{Ay} > 0$, 令 $\mathbf{S} = \text{diag}(h_{i_1} \ \cdots \ h_{i_k})$, 则

$$0 < \mathbf{y}^T \mathbf{H} \mathbf{Ay} = \alpha^T \mathbf{S} \mathbf{B} \alpha = \alpha^T \mathbf{S} \lambda \alpha = \lambda \alpha^T \mathbf{S} \alpha$$

但 $\mathbf{S} \geq \mathbf{0}$, 所以 $\alpha^T \mathbf{S} \alpha \geq 0$, 因而 $\lambda > 0$.

(6) \Rightarrow (1) 设 $\mathbf{B} = \mathbf{A} \begin{pmatrix} i_1 & \cdots & i_k \\ \cdots & \cdots & \cdots \\ i_1 & \cdots & i_k \end{pmatrix}$ 为 \mathbf{A} 的任一主子阵, 则由 \mathbf{B} 的复特征值成对, 实

特征值都大于零知 $|\mathbf{B}| > 0$.

(7) \Rightarrow (3) 由于(1)与(3)等价, \mathbf{A} 的所有主子式大于 0, 又等价于 \mathbf{A}^T 的所有主子式大于零.

(7) \Rightarrow (8) 若存在符号矩阵 \mathbf{S} , 使得 $\mathbf{u} \geq \mathbf{0}$, $\mathbf{SA}^T \mathbf{Su} \leq \mathbf{0}$, $\mathbf{u} \neq \mathbf{0}$, 则

$$\mathbf{u}^T \mathbf{SA}^T \mathbf{Su} \leq 0.$$

令 $\mathbf{Su} = \mathbf{y} = (y_1 \ \cdots \ y_n)^T$, 由 $\mathbf{u} \neq \mathbf{0}$, \mathbf{S} 可逆知 $\mathbf{y} \neq \mathbf{0}$. 再令 $\mathbf{A}^T \mathbf{y} = \mathbf{z} = (z_1 \ \cdots \ z_n)^T$, 由(1)有

$$\mathbf{0} \geq (\mathbf{Su})^T \mathbf{A}^T \mathbf{Su} = \mathbf{y}^T \mathbf{z}.$$

所以 $y_i z_i \leq 0$, $i = 1, 2, \dots, n$. 这与(7)的假设矛盾.

(8) \Rightarrow (7) 若存在 $\mathbf{y} = (y_1 \ \cdots \ y_n)^T \neq \mathbf{0}$, 使得 $\mathbf{z} = \mathbf{A}^T \mathbf{y} = (z_1 \ \cdots \ z_n)^T$, 而 $y_k z_k \leq 0$, $k = 1, 2, \dots, n$. 可选取符号矩阵 \mathbf{S} 使得 $\mathbf{u} = \mathbf{Sy} \geq \mathbf{0}$ 且 $\mathbf{Sz} \leq \mathbf{0}$, 而 $\mathbf{S}^2 = \mathbf{I}$.

$$\mathbf{SA}^T \mathbf{Su} = \mathbf{SA}^T \mathbf{S}(\mathbf{Sy}) = \mathbf{SA}^T \mathbf{y} = \mathbf{Sz} \leq \mathbf{0}, \quad \mathbf{u} \geq \mathbf{0}.$$

□

但 $\mathbf{u} \neq \mathbf{0}$, 这与(8)的假设矛盾.

定理 7.12 设 A 是 n 阶 M 矩阵, 则下面几条等价:

(1) A 是非奇异 M 矩阵;

(2) 存在 $P, Q \in \mathbb{R}^{n \times n}$, 使得 $A = P - Q, P^{-1} \geqslant \mathbf{0}, Q \geqslant \mathbf{0}, \rho(P^{-1}Q) < 1$;

(3) A 可逆, 且 $A^{-1} \geqslant \mathbf{0}$;

(4) 存在 $x > \mathbf{0}$, 使 $Ax > \mathbf{0}$;

(5) A 的任一特征值的实部大于 0 .

证明 (1) \Rightarrow (2) 设 $A = sE - B, s > \rho(B) \geqslant 0, B \geqslant \mathbf{0}$. 令 $P = sE, Q = B$, 则 $P^{-1} = \frac{1}{s}E \geqslant \mathbf{0}, Q \geqslant \mathbf{0}$. 于是 $A = P - Q$, 且 $\rho(P^{-1}Q) = \rho\left(\frac{1}{s}Q\right) = \frac{1}{s}\rho(Q) < 1$.

(2) \Rightarrow (3) 由假设, $A = P - Q, P^{-1} \geqslant \mathbf{0}, Q \geqslant \mathbf{0}, \rho(P^{-1}Q) < 1$. 令 $C = P^{-1}Q$, 则 $A = P - Q = P(E - C)$. 因为 $\rho(C) < 1$, 所以就有 $|E - C| \neq 0$, 从而 $|A| \neq 0$, 故 A 可逆, 且

$$A^{-1} = (E - C)^{-1}P^{-1} = (E + C + C^2 + \dots)P^{-1}.$$

由于 $P^{-1} \geqslant \mathbf{0}, Q \geqslant \mathbf{0}$, 因此 $C = P^{-1}Q \geqslant \mathbf{0}$. 从而 $A^{-1} \geqslant \mathbf{0}$.

(3) \Rightarrow (4) 令 $x = A^{-1}e$, 其中 $e' = (1 \dots 1)$. 由于 $A^{-1} \geqslant \mathbf{0}$, 因此 $x = A^{-1}e \geqslant \mathbf{0}, Ax = e > \mathbf{0}$.

若 $x > \mathbf{0}$, 则结论成立. 否则 x 的分量中有为 0 者时, 取充分小 $\epsilon > 0$, 使 $0 < Ax + \epsilon Ae = A(x + \epsilon e)$, 这时 $x + \epsilon e > \mathbf{0}$ 即为所求.

(4) \Rightarrow (5) 由于 A 是 M 矩阵, 则 $A = sE - B, s > 0, B \geqslant \mathbf{0}$. 若存在 $x > \mathbf{0}$, 使 $Ax > \mathbf{0}$, 则 $(sE - B)x > \mathbf{0}$, 即

$$sx > Bx.$$

设 B' 相应于特征值 $\rho(B')$ 的一个特征向量为 γ , 即 $B'\gamma = \rho(B)\gamma$. 在(1)式两端左乘以 γ' 得 $s\gamma'x > \gamma'Bx = (B'\gamma)'x = (\rho(B)\gamma)'x = \rho(B)\gamma'x$, 所以 $s > \rho(B)$.

但 $A = sE - B$. 设 $\lambda = a + bi$ 为 A 的任一特征值, α 是相应的一个特征向量, 那么由 $A\alpha = \lambda\alpha$ 得

$$(sE - B)\alpha = \lambda\alpha, \quad B\alpha = (\lambda - s)\alpha.$$

从而 $\lambda - s$ 为 B 的特征值, 且 $|\lambda - s| \leqslant \rho(B) < s$, $|(a - s) + bi| < s$, 所以 $a > 0$.

(5) \Rightarrow (1) 设 $A = sE - B, B \geqslant \mathbf{0}$, 则 $s - \rho(B)$ 是 A 的特征值. 而 $s - \rho(B)$ 是实数, 所以 $s - \rho(B) > 0$, 故 A 是非奇异 M 矩阵. \square

习题答案与提示

习 题 一

1. (1),(2) 是; (3),(4) 不是

$$2. (1) \dim W = \frac{n^2 + n}{2}; \text{ 一组基: } \left\{ \begin{pmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}, \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} \right\}$$

$$\left\{ \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & & \\ 0 & \ddots & & \\ \cdots & & & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & & \\ 1 & 0 & \ddots & \\ \cdots & & & 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 & 0 & \cdots & 1 \\ 0 & 0 & & \\ \ddots & & & \\ \cdots & & & 0 \end{pmatrix} \right\}.$$

3. (1) 是; (2) 不是; (3) 是; (4) 是

8. 线性相关.

$$9. R(\mathbf{A}) = E \left\{ \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix} \right\}, N(\mathbf{A}) = L\{(1 4 1 0)^T, (1 1 0 1)^T\}.$$

10. $W_1 \cap W_2 = L\{\alpha_3 - 4\alpha_2\} = L\{3\beta_1 - \beta_2\}, W_1 + W_2 = L\{\alpha_1, \alpha_2, \beta_1\}.$

$$11. \mathbf{C} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 0 \\ -1 \\ -2 \\ 3 \end{pmatrix}.$$

$$12. \mathbf{C} = \begin{pmatrix} 1 & 2 & 1 \\ -1 & 3 & 3 \\ 0 & 2 & 2 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 3 \\ 4 \\ 4 \end{pmatrix}.$$

$$18. \beta_1 = \frac{1}{\sqrt{2}}(\varepsilon_1 + \varepsilon_2), \beta_2 = \frac{1}{\sqrt{10}}(\varepsilon_1 - 2\varepsilon_2 + 2\varepsilon_3 - \varepsilon_5), \beta_3 = \frac{1}{2}(\varepsilon_1 + \varepsilon_2 + \varepsilon_3 - \varepsilon_5).$$

$$20. W^\perp = L\{(1 0 1 0)^T, (-1 0 0 1)^T\}.$$

24. (1) 不是; (2) 是; (3) 不是; (4) 是.

$$25. \mathbf{A} = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 3 & -2 \\ 2 & -1 & 1 \end{pmatrix}.$$

$$26. (1) \mathbf{X}_1 = \begin{pmatrix} a & 0 & b & 0 \\ 0 & a & 0 & b \\ c & 0 & d & 0 \\ 0 & c & 0 & d \end{pmatrix}; \mathbf{X}_2 = \begin{pmatrix} a & c & 0 & 0 \\ b & d & 0 & 0 \\ 0 & 0 & a & c \\ 0 & 0 & b & d \end{pmatrix}; \mathbf{X}_3 = \mathbf{X}_1 \cdot \mathbf{X}_2.$$

(2) $T_1 + T_2$ 矩阵: $\mathbf{X}_1 + \mathbf{X}_2$; $T_1 \cdot T_2$ 矩阵: \mathbf{X}_3 .

$$28. (1) R(T) = L\{T(\mathbf{e}_1), T(\mathbf{e}_2), T(\mathbf{e}_3)\} = L\left\{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right\}, N(T) = L\left\{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right\}.$$

(2) $\dim R(T) = 2$, 基 $\{\mathbf{e}_1, \mathbf{e}_2\}$; $\dim N(T) = 1$, 基 $\{\mathbf{e}_3\}$.

$$30. V_1 = L\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}, V_2 = L\left\{\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right\}.$$

31. (1) 不变子空间 $L(\mathbf{u})$ 和 \mathbf{u} 的正交补子空间 \mathbf{u}^\perp , $R^3 = L(\mathbf{u}) \oplus \mathbf{u}^\perp$,

$$(2) T \xrightarrow{\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}} \begin{pmatrix} k & & \\ & 1 & \\ & & 1 \end{pmatrix}; \mathbf{u} = \mathbf{u}_1, \mathbf{u}^\perp = L\{\mathbf{u}_2, \mathbf{u}_3\}, \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} \text{ 为标准正交基.}$$

$$32. \text{椭圆 } y_1^2 + 4y_2^2 = 1.$$

$$33. R^3 = L(\mathbf{u}) \oplus L\left\{\left(\frac{1}{3}, \frac{2}{3}, -\frac{2}{3}\right)^T, \left(\frac{2}{3}, \frac{1}{3}, \frac{2}{3}\right)^T\right\}, P \xrightarrow{\mathbf{u}_1 \mathbf{u}_2 \mathbf{u}_3} \begin{pmatrix} 0 & & \\ & 1 & \\ & & 1 \end{pmatrix}.$$

$$34. T(\mathbf{e}_3) = -\frac{1}{3}\mathbf{e}_1 + \frac{2}{3}\mathbf{e}_2 + \frac{2}{3}\mathbf{e}_3.$$

习 题 二

1. 提示: 由题意, 存在 $\mathbf{X} \neq \mathbf{0}$, 使 $(\mathbf{AB} - \mathbf{BA})\mathbf{X} = \mathbf{0}$, 证明该 \mathbf{X} 是 $A^{-1}B^{-1}\mathbf{AB}$ 关于 $\lambda = 1$ 的特征向量.

2. (1) 证明 $\mathbf{SS} = \mathbf{I}$, 从而 $\mathbf{S}^{-1} = \mathbf{S}$.

$$(2) \mathbf{S}^{-1} \mathbf{AS} = \begin{pmatrix} a_{14} & a_{13} & a_{12} & a_{11} \\ a_{24} & a_{23} & a_{22} & a_{21} \\ a_{34} & a_{33} & a_{32} & a_{31} \\ a_{44} & a_{43} & a_{42} & a_{41} \end{pmatrix}$$

3. $\lambda_1 = 2, \lambda_2 = 4, \lambda_3 = 6, \lambda_4 = 8$

$$\alpha_1 = (1 \ 1 \ 1 \ 1)^T, \quad \alpha_2 = (1 \ -1 \ 1 \ -1), \\ \alpha_3 = (1 \ 1 \ -1 \ -1)^T, \quad \alpha_4 = (1 \ -1 \ -1 \ 1)^T.$$

4. (1) $\lambda=1$ 或者 $\lambda=0$.

(2) A 可对角化.

5. $\lambda_1 = \lambda_2 = 1, V_{\lambda_1} = L\{3\alpha_1 - 6\alpha_2 + 20\alpha_3\}, \lambda_3 = -2, V_{\lambda_3} = L\{\alpha_3\}$.

8. A 和 C 相似, B 和 C 不相似.

9. (1) $\begin{pmatrix} 2 & & & \\ & 2 & & \\ & & -1 \end{pmatrix}, \begin{pmatrix} 2 & 1 & & \\ & 2 & & \\ & & -1 \end{pmatrix}.$

(2) $a=0; b$ 与 c 可取任何常数.

12. (1) $J_A = \begin{pmatrix} 2 & & & \\ & 1 & 1 & \\ & & 1 \end{pmatrix}; P = \begin{pmatrix} 0 & 1 & -1 \\ 0 & 2 & -1 \\ 1 & -1 & 0 \end{pmatrix}.$

(2) $J_A = \begin{pmatrix} -1 & & & \\ & -1 & 1 & \\ & & -1 \end{pmatrix}; P = \begin{pmatrix} 2 & 3 & 1 \\ -1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$

(3) $J_A = \begin{pmatrix} 1 & 1 & & \\ & 1 & & \\ & & 1 & 1 \\ & & & 1 \end{pmatrix}; P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 2 & 0 & -1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$

13. (1) $\begin{cases} \beta_1 = \alpha_1 + \alpha_2 \\ \beta_2 = \alpha_1 + \alpha_3; \\ \beta_3 = -\alpha_2 \end{cases} J_A = \begin{pmatrix} 4 & & \\ & 4 & 1 \\ & & 4 \end{pmatrix}.$

(2) $m_T(\lambda) = (\lambda - 4)^2$.

14. (1) $m_A(\lambda) = (\lambda - 2)(\lambda - 1)^2; (2) m_A(\lambda) = (\lambda + 1)^2;$

(3) $m_A(\lambda) = (\lambda - 1)^2$.

15. $J = \begin{pmatrix} 2 & & & \\ & 2 & & \\ & & -3 & 1 \\ & & -3 & -3 \end{pmatrix} \text{ 或 } \begin{pmatrix} 2 & & & \\ & -3 & & \\ & & -3 & 1 \\ & & & -3 \end{pmatrix}.$

17. $g(A) = P \begin{pmatrix} g(2) & & \\ & g(1) & g'(1) \\ & & g(1) \end{pmatrix} P^{-1}, P = \begin{pmatrix} 0 & 1 & -1 \\ 0 & 2 & -1 \\ 1 & -1 & 0 \end{pmatrix}.$

19. (1) $m_A(\lambda) = (\lambda - 2)(\lambda - 3), m_A(\lambda) = (\lambda - 2)^2(\lambda - 3),$

$$m_A(\lambda) = (\lambda - 2)(\lambda - 3)^2, \quad m_A(\lambda) = (\lambda - 2)^2(\lambda - 3)^2.$$

$$(2) J_A = \begin{pmatrix} 2 & & & \\ & 2 & & \\ & & 3 & \\ & & & 3 \end{pmatrix}, \begin{pmatrix} 2 & 1 & & \\ & 2 & & \\ & & 3 & \\ & & & 3 \end{pmatrix},$$

$$\begin{pmatrix} 2 & & & \\ & 2 & & \\ & & 3 & 1 \\ & & & 3 \end{pmatrix}, \begin{pmatrix} 2 & 1 & & \\ & 2 & & \\ & & 3 & 1 \\ & & & 3 \end{pmatrix}.$$

20. 提示: 分解 $V_n(F) = W \oplus U$, T 的矩阵为 $\begin{pmatrix} A & B \\ \mathbf{0} & C \end{pmatrix}$,

求 A 的一个特征值 λ 对应的特征向量.

习 题 三

$$2. A = \begin{pmatrix} 1 & & & \\ \frac{2}{\sqrt{5}} & 1 & & \\ \frac{-4}{5} & -2 & 1 & \\ 0 & 5 & 2 & 1 \end{pmatrix} \begin{pmatrix} 5 & & & \\ \frac{1}{5} & & & \\ & 1 & & \\ & & -7 & \end{pmatrix} \begin{pmatrix} 1 & \frac{2}{5} & -\frac{4}{5} & 0 \\ 1 & -2 & 5 & \\ 1 & 2 & & \\ 1 & & & \end{pmatrix}.$$

$$3. (1) A = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} (1 \ 1 \ 1).$$

$$(2) A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & -2 & -1 & 1 \end{pmatrix}.$$

$$(3) A = \begin{pmatrix} 0 & 1 \\ 2 & 1 \\ 2i & 0 \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{2} & 0 \end{pmatrix}.$$

$$5. A = P \begin{pmatrix} 3 & \\ & 0 \end{pmatrix} P^{-1} + P \begin{pmatrix} 0 & \\ & -1 \end{pmatrix} P^{-1} = 3 \begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ 1 & \frac{1}{2} \end{pmatrix} - \begin{pmatrix} \frac{1}{2} & -\frac{1}{4} \\ -1 & \frac{1}{2} \end{pmatrix},$$

$$P = \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix}.$$

$$6. \quad A = P \begin{pmatrix} -2 & & \\ & -2 & \\ & & 0 \end{pmatrix} P^{-1} + P \begin{pmatrix} 0 & & \\ & 0 & \\ & & 4 \end{pmatrix} P^{-1}, \quad P = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

$$13. (1) \quad A = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

$$(2) \quad A = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

习 题 四

$$1. \quad \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 1 \end{pmatrix}.$$

$$2. \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}.$$

3. (1) $C^{-1}B_L^{-1}A^{-1}$ 是 ABC 的左逆.

(2) $C^{-1}B_R^{-1}A^{-1}$ 是 ABC 的右逆.

4. 由 $A = PBQ$ 知, $\text{rank}(A) \leq \text{rank}(B)$. 又因 P 列满秩, Q 行满秩, 所以 $B = P_L^{-1}AQ_R^{-1}$, 于是 $\text{rank}(B) \leq \text{rank}(A)$.

$$5. \quad A\{1\} = \left\{ Q \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ a & b \end{pmatrix} P, a, b \text{ 为任意数} \right\}, \quad P = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad Q = \begin{pmatrix} -3 & -2 & -7 \\ 0 & 0 & 1 \\ 2 & 1 & 4 \end{pmatrix}.$$

6. $\forall z \in \mathbb{C}^n, z = z_1 + z_2, z_1 \in T, z_2 \in N(A)$, 于是 $AGAz = AGAz_1 + AGAz_2 = AGAz_1 = Az_1 = Az$. 故 $AGA = A$.

$$7. \quad A^+ = \frac{1}{20} \begin{pmatrix} 2 & 3 & -1 & 1 \\ 8 & 2 & 6 & -6 \\ -2 & -3 & 1 & -1 \end{pmatrix}.$$

8. 验证 A 满足 M-P 逆定义中的四个条件.

9. $Ax = 0$, 即 $BCx = 0$, 由 B 列满秩知 $Cx = 0$.

10. 必要性: 由条件可推得, $AB + BA = 0$, 将此式左乘 A 得: $AB + ABA = 0$; 右乘 A , 得: $ABA + BA = 0$, 于是 $AB = BA$, 故 $AB = 0$.

11. 必要性:由条件可推得, $2B=AB+BA$, 左乘 A 得 $2AB=AB+ABA$, 即 $AB=ABA$, 右乘 A 得 $2BA=BA+ABA$, 即 $BA=ABA$, 故 $AB=BA$, 从而 $B=AB=BA$.

12. $\forall x \in N(A^+)$ $\Rightarrow A^+x = 0 \Rightarrow AA^+x = 0 \Rightarrow (AA^+)^Hx = 0 \Rightarrow (A^+)^HA^Hx = 0 \Rightarrow A^H(A^+)^HA^Hx = 0 \Rightarrow A^Hx = 0 \Rightarrow x \in N(A^H)$,

$\forall x \in N(A^H) \Rightarrow A^Hx = 0 \Rightarrow (A^+)^HA^Hx = 0 \Rightarrow (AA^+)^Hx = 0 \Rightarrow AA^+x = 0 \Rightarrow AA^+Ax = 0 \Rightarrow Ax = 0 \Rightarrow x \in N(A)$.

$\forall y \in R(A^+)$, $\exists x$, 使 $y = A^+x = A^+AA^+x = (A^+A)^H A^+x = A^H(A^+)^HA^+x \in R(A^H)$, 从而 $R(A^+) \subseteq R(A^H)$, 反过来的包含关系同理可证.

13. (1) $(AB)^2 = ABAB = AB$, 所以 AB 的特征值为 1 或 0, 又 $\text{rank}(AB) = n$, 所以 AB 的特征值有 n 个 1, $m-n$ 个 0.

(2) $R(AB) \subseteq R(A)$, $N(B) \subseteq N(AB)$ 易证. 注意到

$$\forall y \in R(A), \exists x, \text{ 使 } y = Ax = ABx \in R(AB),$$

$$\forall x \in N(AB), \text{ 有 } ABx = 0 \Rightarrow BABx = 0 \Rightarrow Bx = 0,$$

即得证.

$$14. A^+ = \frac{1}{5} \begin{pmatrix} 0 & 1 & 0 \\ 2 & 0 & 1 \\ 0 & 2 & 0 \end{pmatrix}, x = A^+ b = \begin{pmatrix} \frac{1}{5} & \frac{3}{5} & \frac{2}{5} \end{pmatrix}^T.$$

习 题 五

1. 只验证三角不等式成立. 在 \mathbf{R}^n 中, 定义

$$(x, y) = \sum_{i=1}^n a_i x_i y_i,$$

验证它是 \mathbf{R}^n 的一个内积, 于是三角不等式成立.

2. 直接验证.

$$3. \|A\|_1 = 12, \|A\|_\infty = 14;$$

$$\|Ax\|_1 = 9 + \sqrt{10} + \sqrt{53}, \|Ax\|_\infty = \sqrt{53}.$$

$$4. (1) \|U\|_2 = \max_{x \neq 0} \frac{\|Ux\|_2}{\|x\|_2} = \max_{x \neq 0} \frac{((Ux)^H Ux)^{1/2}}{(x^H x)^{1/2}}$$

$$= \max_{x \neq 0} \frac{(x^H U^H Ux)^{1/2}}{(x^H x)^{1/2}} = 1.$$

(2),(3) 类似可证.

5. (1) 因为

$$1 = \|I\|_p = \|AA^{-1}\|_p \leqslant \|A\|_p \|A^{-1}\|_p,$$

所以

$$\|A^{-1}\|_p \geqslant \|A\|_p^{-1}.$$

(2) 因为

$$\begin{aligned}\|\mathbf{A}^{-1}\|_p &= \max_{x \neq 0} \frac{\|\mathbf{A}^{-1}\mathbf{x}\|_p}{\|\mathbf{x}\|_p} = \frac{\|\mathbf{A}^{-1}\mathbf{x} = \mathbf{y}\|_p}{\|\mathbf{x}\|_p} \max_{y \neq 0} \frac{\|\mathbf{y}\|_p}{\|\mathbf{Ay}\|_p} \\ &= \left(\min_{y \neq 0} \frac{\|\mathbf{Ay}\|_p}{\|\mathbf{y}\|_p} \right)^{-1},\end{aligned}$$

所以

$$\|\mathbf{A}^{-1}\|_p^{-1} = \min_{x \neq 0} \frac{\|\mathbf{Ax}\|_p}{\|\mathbf{x}\|_p}.$$

6. 因 $\{\mathbf{A}^{(k)}\}$ 收敛于 \mathbf{A} , 则对任意矩阵范数 $\|\cdot\|$, 有

$$\|\mathbf{A}^{(k)} - \mathbf{A}\| \rightarrow 0,$$

又

$$\begin{aligned}\|\mathbf{P}^{-1}\mathbf{A}^{(k)}\mathbf{P} - \mathbf{P}^{-1}\mathbf{AP}\| &= \|\mathbf{P}^{-1}(\mathbf{A}^{(k)} - \mathbf{A})\mathbf{P}\| \\ &\leqslant \|\mathbf{P}^{-1}\| \cdot \|\mathbf{A}^{(k)} - \mathbf{A}\| \cdot \|\mathbf{P}\| \rightarrow 0,\end{aligned}$$

所以

$$\lim_{k \rightarrow \infty} \mathbf{P}^{-1}\mathbf{A}^{(k)}\mathbf{P} = \mathbf{P}^{-1}\mathbf{AP}.$$

$$7. \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

$$8. |c| < \frac{1}{2}.$$

9. $|\lambda| \leqslant \rho(\mathbf{A}) \Rightarrow |\lambda^k| = |\lambda|^k \leqslant (\rho(\mathbf{A}))^k = \rho(\mathbf{A}^k)$. 又 $\rho(\mathbf{A}^k) \leqslant \|\mathbf{A}^k\|$, 所以 $|\lambda| \leqslant \|\mathbf{A}^k\|^{\frac{1}{k}}$.

10. 设 \mathbf{A} 的特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 因为 \mathbf{A} 是正规阵, 所以存在酉矩阵 \mathbf{U} , 使得

$$\mathbf{U}^H \mathbf{A} \mathbf{U} = \text{diag}(\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_n),$$

于是 $\mathbf{U}^H \mathbf{A}^H \mathbf{A} \mathbf{U} = (\mathbf{U}^H \mathbf{A} \mathbf{U})^H \cdot \mathbf{U}^H \mathbf{A} \mathbf{U} = \text{diag}(|\lambda_1|^2 \quad |\lambda_2|^2 \quad \cdots \quad |\lambda_n|^2)$,

即 $\mathbf{A}^H \mathbf{A}$ 的特征值为 $|\lambda_1|^2, |\lambda_2|^2, \dots, |\lambda_n|^2$, 从而

$$\|\mathbf{A}\|_2 = \mathbf{A}^H \mathbf{A} \text{ 的最大特征值的平方根} = \rho(\mathbf{A}).$$

$$11. \frac{2}{3} \begin{pmatrix} 4 & 7 \\ 3 & 9 \end{pmatrix}.$$

12. 收敛.

13. $\mathbf{B} = \mathbf{P}^{-1}\mathbf{AP}, \mathbf{B}^k = \mathbf{P}^{-1}\mathbf{A}^k\mathbf{P}$. 验证 $\sum_{k=0}^{\infty} \mathbf{B}^k$ 收敛 $\Leftrightarrow \sum_{k=0}^{\infty} \mathbf{A}^k$ 收敛.

$$14. e^{At} = \begin{pmatrix} e^{2t} & 12e^t - 12e^{2t} + 13te^{2t} & -4e^t + 4e^{2t} \\ 0 & e^{2t} & 0 \\ 0 & -3e^t + 3e^{2t} & e^t \end{pmatrix},$$

$$\sin At = \begin{pmatrix} \sin 2t & 12 \sin t - 12 \sin 2t + 13t \cos 2t & -4 \sin t + 4 \sin 2t \\ 0 & \sin 2t & 0 \\ 0 & -3 \sin t + 3 \sin 2t & \sin t \end{pmatrix}.$$

$$15. \ e^{At} = \begin{pmatrix} e^t & -4e^t + 4e^{2t} & -5e^t + 5e^{3t} \\ 0 & e^{2t} & 0 \\ 0 & 0 & e^{3t} \end{pmatrix}.$$

$$16. \ e^{At} = e^{-t} \begin{pmatrix} 1+4t & 0 & 8t \\ 3t & 1 & 6t \\ -2t & 0 & 1-4t \end{pmatrix}, \ x(t) = ((1+12t)e^{-t} \quad (1+9t)e^{-t} \quad (1-6t)e^{-t})^T.$$

$$17. \ e^{At} = \begin{pmatrix} 1-2t & t & 0 \\ -4t & 2t+1 & 0 \\ 1+2t-e^t & e^t-t-1 & e^t \end{pmatrix}, \ x(t) = (1 \quad 1 \quad (t-1)e^t)^T.$$

习 题 六

1. (1) 使用定理 5.2, 对 k 用归纳法证.

(2) 设 $\text{rank}(A)=r, \text{rank}(B)=s, \exists P_1, P_2, Q_1, Q_2$, 有

$$P_1 A Q_1 = \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad P_2 B Q_2 = \begin{pmatrix} I_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

则 $(P_1 \otimes P_2)(A \otimes B)(Q_1 \otimes Q_2) = (P_1 A Q_1) \otimes (P_2 B Q_2)$

$$= \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \otimes \begin{pmatrix} I_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

$$\text{rank}(A \otimes B) = \text{rank} \left[\begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \otimes \begin{pmatrix} I_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \right] = r \cdot s.$$

(3) $\text{rank}(A \otimes B) = \mathbf{0} \Leftrightarrow A \otimes B = \mathbf{0}$.

由(2) $\text{rank}(A \otimes B) = \text{rank}(A) \text{rank}(B) = 0 \Rightarrow$ 或 $\text{rank}(A) = 0 \Leftrightarrow A = \mathbf{0}$,
 $\text{rank}(B) = 0 \Leftrightarrow B = \mathbf{0}$.

3. $(x \otimes u, y \otimes v) = (y \otimes v)^H (x \otimes u) = (y^H x) \otimes (v^H u) = (y^H x) \cdot (v^H u)$. 内积为数.

$$\begin{aligned} 4. \ \text{tr}(A \otimes B) &= \text{tr}(a_{11} B) + \text{tr}(a_{22} B) + \cdots + \text{tr}(a_{nn} B) \\ &= a_{11} \text{tr}(B) + a_{22} \text{tr}(B) + \cdots + a_{nn} \text{tr}(B) \\ &= (a_{11} + a_{22} + \cdots + a_{nn}) \text{tr}(B) = \text{tr}(A) \cdot \text{tr}(B). \end{aligned}$$

5. 直接验证 $H_1 \otimes H_2$ 满足 Hadamard 定义的条件.

$$6. \quad \mathbf{E} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

8. 用定理 6.3 结合题 1(2) 证.

9. 注意 $\text{Vec}(T_1(\mathbf{X}))$ 是 $T_1(\mathbf{X})$ 在 $\{\mathbf{E}_{ij}, i=1, 2, \dots, m; j=1, 2, \dots, n\}$ 下的坐标, 又

$$T_1(\mathbf{X}) = \mathbf{A}\mathbf{X}\mathbf{B} \Rightarrow \text{Vec}(T_1(\mathbf{X})) = \mathbf{B}^T \otimes \mathbf{A}\text{Vec}(\mathbf{X}),$$

由线性变换的坐标式, 可知 T_1 的变换矩阵为 $\mathbf{B}^T \otimes \mathbf{A}$. 同理可证 $T_2(\mathbf{X})$.

$$10. \quad \mathbf{X} = \begin{pmatrix} 1 & k \\ -2 & -1 \end{pmatrix}, k \text{ 为任意常数.}$$

11. $\forall r, s, \mu \neq \lambda - \lambda_i, \lambda_i$ 为 \mathbf{A} 的特征值.

12. 因为 $(\mathbf{I}_n^T \otimes \mathbf{I}_m - \mathbf{B}^T \otimes \mathbf{A}) \text{Vec} \mathbf{X} = \text{Vec} \mathbf{D}$, 当 $\rho(\mathbf{A}) \cdot \rho(\mathbf{B}) < 1$ 时 $\Rightarrow \rho(\mathbf{B}^T \otimes \mathbf{A}) < 1$, $(\mathbf{I}_{mn} - \mathbf{B}^T \otimes \mathbf{A})$ 可逆, 且

$$\text{Vec} \mathbf{X} = (\mathbf{I}_{mn} - \mathbf{B}^T \otimes \mathbf{A})^{-1} \text{Vec} \mathbf{C} = \sum_{k=0}^{\infty} (\mathbf{B}^T \otimes \mathbf{A})^k \text{Vec}(\mathbf{C}) = \text{Vec} \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{C} \mathbf{B}^k.$$

$$13. \quad (\mathbf{C}(t))_{ij} = (\mathbf{A}(t) \otimes \mathbf{B}(t))_{ij} = a_{ij}(t) \mathbf{B}(t)$$

$$\begin{aligned} \text{利用} \quad \frac{d}{dt}(a_{ij}(t) \mathbf{B}(t)) &= \frac{da_{ij}(t)}{dt} \mathbf{B}(t) + a_{ij}(t) \frac{d\mathbf{B}(t)}{dt} \\ &= \left[\frac{d\mathbf{A}(t)}{dt} \otimes \mathbf{B}(t) + \mathbf{A} \times \frac{d\mathbf{B}(t)}{dt} \right]_{ij}. \end{aligned}$$

参 考 书 目

- 1 余鄂西.矩阵论.北京:高等教育出版社,1995
- 2 蒋正新等.矩阵理论及其应用.北京:北京航空学院出版社,1988
- 3 程云鹏.矩阵论.西安:西北工业大学出版社,1999
- 4 北京大学数学系.高等代数.北京:高等教育出版社,1987
- 5 史荣昌.矩阵分析.北京:北京理工大学出版社,1996
- 6 罗家洪.矩阵分析引论.广州:华南理工大学出版社,2000
- 7 李乔.矩阵论八讲.上海:上海科学技术出版社,1988
- 8 Fuzhen Zhang, Matrix Theory, Springer, 1999