

重庆大学教材建设基金资助

矩阵理论及其应用

李 新 何传江 主编

重 庆 大 学 出 版 社

内 容 提 要

本书主要内容分成两部分,第一部分包括第 1 章、第 2 章、第 3 章内容,这部分作为《线性代数》的衔接与补充,主要讲了线性空间、内积空间、线性变换。第二部分包括第 4 章到第 9 章,这一部分是考虑到当前各工科学科研究生的实际需要而选择的内容,主要包括:范数理论及其应用;矩阵分析及其应用;矩阵分解;广义矩阵及其应用;特征值的估计及广义特征值;矩阵的 kronecker 积等。

图书在版编目(CIP)数据

矩阵理论及其应用/李新主编. —重庆:重庆大学出版社, 2005.8
ISBN 7-5624-3411-5
. 矩... . 李... . 矩阵—理论—高等学校—
教材 . 0151.21

中国版本图书馆 CIP 数据核字(2005)第 064077 号

矩阵理论及其应用

李 新 何传江 主 编
责任编辑:周 立 版式设计:周 立
责任校对:任卓惠 责任印制:秦 梅

*

重庆大学出版社出版发行
出版人:张鸽盛
社址:重庆市沙坪坝正街 174 号重庆大学(A 区)内
邮编:400030
电话:(023) 65102378 65105781
传真:(023) 65103686 65105565
网址: <http://www.cqup.com.cn>
邮箱: fxk@cqup.com.cn (市场营销部)
全国新华书店经销
重庆师范大学印刷厂印刷

*

开本:787×1092 1/16 印张:14.5 字数:362千
2005年8月第1版 2005年8月第1次印刷
印数:1—3 000
ISBN 7-5624-3411-5 定价:19.50 元

本书如有印刷、装订等质量问题,本社负责调换
版权所有,请勿擅自翻印和用本书
制作各类出版物及配套用书,违者必究。

前言

用矩阵的理论与方法来处理现代工程技术中的各种问题已越来越普遍。在工程技术中引进矩阵理论不仅使理论的表达极为简捷,而且对理论的实质刻画也更为深刻,这一点是不容置疑的,更由于计算机和计算方法的普及发展,不仅为矩阵理论的应用开辟了广阔的前景,也使工程技术的研​​究发生新的变化,开拓了崭新的研究途径,例如系统工程,优化方法,稳定性理论等,无不与矩阵理论发生紧密结合。因此矩阵的理论与方法已成为研究现代工程技术的数学基础。本书是为了提高工科研究生的理论分析能力和科学实践能力以适应研究工作需要而编写的。

考虑到研究生已具备的数学基础,本书起点放在已学习40学时工程数学《线性代数》的基础上,结合各学科研究生教学和科研工作的需要,选编了此书,作为工科研究生数学选修课的参考教材之一。此书在出版之前曾编成讲义,在重庆大学硕士研究生数学选修课中试讲过9年。感谢重庆大学教材建设基金的资助,此书才得以顺利出版。

本书内容分为两部分。第一部分包括第1章、第2章、第3章内容,这一部分作为《线性代数》的衔接与补充,为学习第二部分内容打下必要的基础。它们有线性空间,内积空间、线性变换,考虑到教学效果,在介绍约当标准型时采用的是——矩阵方法,而约当标准型是后面矩阵函数的重要基础。第二部分包括第4章至第9章,这一部分是考虑到当前各工科学科研究生的实际需要而选择的内容。学时数在50~60学时的除讲授第1篇以外,第2篇可参考讲授第4章,第5章,第6章的第三节:矩阵的最大秩分解,第7章,第8章的第一、二、三节,其中“*”号内容可不讲。

各章后面都配有一定量的习题,全书最后也配有习题答案或提示。

本书在编写过程中,得到重庆大学研究生院及重庆大学数理学院领导的大力支持,特别承蒙重庆大学李平渊教授、张心明老师的审核,提出了许多宝贵意见。本书由李新与何传江主编。

由于编者水平有限,经验不足,对于书中谬误和不当之处,如蒙赐教,不胜感激。

编 者

2005年8月

符号说明

A, B, C, V, R, K, \dots	集合
$a, b, c, \dots, \mu k, \dots$	元素或数
A, B, C, X, \dots	矩阵
$\alpha, \beta, \gamma, \delta, \epsilon, \zeta, \eta, \dots$	向量
A, B, E, T, \dots	线性变换
K	一般数域(集)
C	复数域(集)
R	实数域(集)
Q	有理数域(集)
Z	整数集
N	自然数集
$f: X \rightarrow Y$	f 是集合 X 到集合 Y 的映射
$f(x) = y$	给定 x , 有惟一的 y 与之对应
R^n	n 维实向量空间
$R^{m \times n}$	$m \times n$ 阶实矩阵空间
Re	复数 z 的实部
Im	复数 z 的虚部
A^T	矩阵 A 的转置
A^H	矩阵 A 的转置共轭
A^*	矩阵 A 的伴随矩阵
$R(A)$	矩阵 A 的值域(列空间)
$N(A)$	矩阵 A 的核(零空间)
$r(A)$	矩阵 A 的秩(或注 $\operatorname{rank} A$)
$(A)_{ij}$	矩阵 A 的 i 行 j 列的元素(或记 a_{ij})
$\rho(A)$	矩阵 A 的谱半径
$\det A$	方阵 A 的行列式(或记 $ A $)
$\operatorname{tr} A$	方阵 A 的迹, 即 A 的主对角元素之和
$\ A\ $	矩阵 A 的范数
$\ A\ = \max_{\ x\ =1} \ Ax\ $	所有向量 x 范数为 1 的使向量 Ax 范数最大者
A^+	矩阵 A 的 Moore-Penrose 广义逆
A_L^{-1}	矩阵 A 的左逆
A_R^{-1}	矩阵 A 的右逆
$A: V \rightarrow V$	A 是线性空间 V 到 V 的线性变换
$\operatorname{Hom}(V, V)$	线性空间 V 到 V 的线性变换全体
$R(A)$	线性变换 A 的值域
$N(A)$	线性变换 A 的核
$A \dot{\smash{,}} B$	矩阵 A 与 B 的 Kronecker 积
$A \oslash B$	矩阵 A 与 B 的 Kronecker 和

目录

第 1 篇 线性空间与线性变换

第 1 章 线性空间	2
1.1 集合与映射	2
1.2 线性空间定义及其性质	5
1.3 线性空间的基与坐标	7
1.4 基变换与坐标变换	8
1.5 线性子空间	11
1.6 子空间的交与和	13
习题 1	16
第 2 章 内积空间	18
2.1 欧氏空间	19
2.2 标准正交基与 Gram-Schmidt 过程	22
2.3 正交补与投影定理	25
2.4 酉空间	28
习题 2	30
第 3 章 线性变换	32
3.1 线性变换定义	32
3.2 线性变换的矩阵表示	35
3.3 线性变换的最简矩阵表示——相似形理论	39
3.4 Hamilton-Cayley 定理、最小多项式	55
3.5 正交变换、酉变换	60
习题 3	63

第 2 篇 矩阵理论及其应用

第 4 章 范数理论及其应用.....	68
4.1 向量范数及其性质	68
4.2 矩阵的范数	75
4.3 范数应用	83
习题 4	87
第 5 章 矩阵分析及其应用.....	89
5.1 向量和矩阵的极限	89
5.2 函数矩阵的微分和积分	94
5.3 方阵的幂级数.....	103
5.4 方阵函数.....	109
5.5 常用方阵函数的一些性质.....	119
5.6 方阵函数在微分方程组中的应用.....	123
习题 5	126
第 6 章 矩阵分解	129
6.1 Gauss 消去法与矩阵的三角分解	129
6.2 单纯矩阵的谱分解.....	135
6.3 矩阵的最大秩分解.....	137
6.4 矩阵的 QR 分解	141
6.5* 矩阵的奇异值分解	144
习题 6	147
第 7 章 广义逆矩阵及其应用	149
7.1 广义逆矩阵及其分类.....	149
7.2 广义逆矩阵 A^-	150
7.3 广义逆矩阵 A^+	156
7.4* 广义逆矩阵的通式	163
7.5 广义逆矩阵的应用.....	165
习题 7	168
第 8 章 特征值的估计及广义特征值	171
8.1 特征值的界的估计.....	171
8.2 圆盘定理.....	173
8.3 谱半径的估计.....	177
8.4* 特征值的摄动	180
8.5* 广义特征值	183
习题 8	186

第 9 章 矩阵的 kronecker 积 188

9.1 kronecker 积的基本性质 188

9.2 kronecker 积的特征值 191

9.3 kronecker 积的应用 196

习题 9 203

练习答案(仅供参考) 205

第 1 篇

线性空间与线性变换

线性空间与线性变换是学习《矩阵论》经常用到的两个极其重要的概念。本篇先简要论述这两个概念及其理论,然后再讨论两个特殊的线性空间,这就是 Euclid 空间和酉空间。所有论述是在假定读者已经具备了 n 维向量空间的理论,矩阵的初步运算,线性方程组的理论和二次型的有关知识基础上进行的。

第 1 章

线性空间

1.1 集合与映射

1.1.1 集 合

集合是数学中最基本的概念之一。所谓集合,是指作为整体看的一堆东西。例如由一些数(有限个或无限个)组成的集合,叫做数集合或数集;一个线性方程组解的全体组成一个集合,叫做解集合;一个已知半径和圆心的开圆内的所有点组成一个集合,叫做点集合或点集等等。组成集合的事物叫做这个集合的元素。一般用英文大写字母 A, B, \dots 表示集合,小写字母 a, b, c, \dots 表示元素,常用记号

$$a \in A$$

表示 a 是集合 A 的元素,读作 a 属于 A ;用记号

$$a \notin A$$

表示 a 不是集合 A 的元素,读作 a 不属于 A 。

集合有多种定义法,常见的有枚举法、隐式法、递归法。

枚举法是列出集合的全体元素。但当集合的元素有一定规律时,可只列出一部分。例如,由桌子,椅子,教室,电灯组成的集合 A ,可记为

$$A = \{ \text{桌子, 椅子, 教室, 电灯} \}$$

自然数集合 N ,可记

$$N = \{ 0, 1, 2, 3, \dots \}$$

隐式法是用描述集合的元素所具有的特征性质的方法定义集合。例如,适合方程 $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ 的全部点组成的集合 B ,可记为

$$B = \{ (x, y) \mid \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \}$$

递归定义一个集合分三步完成。如定义偶数集 E ,

- 1) 2 属于集合 E 。
- 2) 若 x, y 属于集合 E , 则 $x + y, x - y$ 都属于 E 。
- 3) 满足 1 和 2 的元素都属于集合 E 。

不包含任何元素的集合称为空集, 记为 \emptyset 。例如, 一个无解的线性方程组的解的集合就是一个空集。空集合在集合运算中所起的作用, 类似于数零在数的运算中所起的作用。

如果集合 B 的所有元素全是集合 A 的元素, 即由 $a \in B$ 可以推出 $a \in A$, 那么就称 B 为 A 的子集合, 记为

$$B \subseteq A \text{ 或 } A \supseteq B$$

例如, 全体偶数集 E 就是全体整数集 Z 的子集合。我们规定空集合 \emptyset 是任意一个集合的子集合, 按定义, 每个集合都是它自身的子集合, 于是把集合的这两个子集合称为平凡子集合, 而把它的其他子集合称为真子集。

如果集合 A 的元素与集合 B 的元素完全相同, 即 $a \in A$, 当且仅当 $a \in B$, 那么称它们相等, 记为

$$A = B$$

显然两个集合 A 与 B 如果同时满足 $A \subseteq B$ 与 $B \subseteq A$, 那么 $A = B$ 。

我们把既属于集合 A , 又属于集合 B 的全体元素所组成的集合叫做 A 与 B 的交, 记为

$$A \cap B = \{x \mid x \in A \text{ 且 } x \in B\}$$

例如, $A = \{1, 2, 3, 4\}, B = \{6, 3, 2, 1\}, A \cap B = \{1, 2, 3\}$, 两个集合的交集显然具有下面性质

$$A \cap B = A \cap A \cap B = B \cap B$$

属于集合 A 或属于集合 B 的元素全体组成的集合, 称为 A 与 B 的并, 记为

$$A \cup B = \{x \mid x \in A \text{ 或 } x \in B\}$$

A 与 B 的并显然满足关系

$$A \cup B = A \cup B \cup B = B \cup A$$

集合 A 与集合 B 的和集是指如下的集合

$$\{a + b \mid a \in A, b \in B\}$$

常用记号 $A + B$ 来表示, 于是有

$$A + B = \{a + b \mid a \in A, b \in B\}$$

应该指出, 两个集合的和集概念不同于它们并集的概念, 例如

$$\begin{aligned} \{1, 2, 3\} \cup \{2, 3, 4\} &= \{1, 2, 3, 4\} \\ \{1, 2, 3\} + \{2, 3, 4\} &= \{3, 4, 5, 6, 7\} \end{aligned}$$

某些数集(含 0, 1), 如果其中任意两个数的和、差、积、商(除数不为 0)仍在该数集中(即数集关于四则运算封闭), 那么称该数集为数域。例如, 实数集关于四则运算封闭, 因此它形成一个数域, 称其为实数域, 记为 R ; 同样, 复数集也形成一个数域, 称其为复数域, 记为 C ; 读者可以自行验证有理数集形成一个有理数域, 记为 Q 。但奇数集不能成为数域, 偶数集也不能成为数域。

1.1.2 映 射

设 X 与 Y 是两个集合。所谓集合 X 到集合 Y 的一个映射(或映照)是指一个法则(规则)

，使 X 中每一个元素 x 都有 Y 中惟一确定的元素 y 与之对应，记为

$$f: X \rightarrow Y \quad (x) = y \quad \text{或} \quad x \rightarrow y (= f(x))$$

y 称为 x 在映射 f 下的像，而 x 称为 y 在映射 f 下的源像(像源)。

X 到 X 自身的映射，有时也称为 X 到自身的一个变换。这种特殊的映射，在矩阵论中，也是经常出现的，读者应予以注意。

例如， A 是数域 K 上全体 n 阶方阵的集合，定义

$$f_1(X) = \det X, X \in A$$

则有 $f_1: A \rightarrow K$ ，即 f_1 是 A 到 K 的一个映射；如果定义

$$f_2(k) = kE, k \in K$$

这里 E 是 n 阶单位矩阵，则有 $f_2: K \rightarrow A$ 。

令 P_n 表示所有次数不超过 n 的实系数多项式集合，定义

$$(f(t))' = f'(t) \quad f(t) \in P_n$$

是 P_n 到 P_n 的一个映射(实为求导运算)。

设 f 是 X 到 Y 的一个映射，若对 Y 中每个元 y ，都有 X 中的元 x 与之对应，即 $f(x) = y$ ，称 f 是满映射。例如 f_1 ，对任意的 K 中一个 k ，显然选

$$X = \begin{pmatrix} k & & & \\ & 1 & & \\ & & \ddots & \\ & & & w \\ & & & & 1 \end{pmatrix}, \text{ 则 } \det X = k$$

故 f_1 是 A 到 K 的一个满映射。

设 f 是 X 到 Y 的一个映射，若对任意的 $x_1, x_2 \in X$ ，当 $x_1 \neq x_2$ 时，有 $f(x_1) \neq f(x_2)$ ，称 f 是单映射。例如 f_2 ，当 $k_1 \neq k_2$ 时， $k_1, k_2 \in K$ ，这时 $f_2(k_1) = k_1E \neq k_2E = f_2(k_2)$ ，故称 f_2 是 K 到 A 的一个单映射。

若 f 是既单又满的映射，称 f 是一一映射。例如定义 $f: \mathbb{R} \rightarrow \mathbb{R}$ ，对任意 $a \in \mathbb{R}$ ， $f(a) = 2a + 1$ ，这时若 $a_1 \neq a_2$ ，当然 $f(a_1) = 2a_1 + 1 \neq 2a_2 + 1 = f(a_2)$ ，并且对任意 $k \in \mathbb{R}$ ，存在 $a = \frac{k-1}{2} \in \mathbb{R}$ ，使 $f(a) = k$ 。所以 f 是一一映射。

设 f_1 和 f_2 都是集合 X 到 Y 的映射，如果对于每个元素 $x \in X$ ，都有 $f_1(x) = f_2(x)$ ，则称映射 f_1 与 f_2 相等，记为 $f_1 = f_2$ 。

设 f 是 X 到 Y 的映射， g 是 Y 到 Z 的映射，映射的乘积 $g \circ f$ 定义如下

$$(g \circ f)(x) = [g(f(x))], x \in X$$

此即相继施行映射 f 和 g 的结果， $g \circ f$ 是 X 到 Z 的映射。

设 $f: X \rightarrow Y$ ， $g: Y \rightarrow Z$ ， $h: Z \rightarrow W$ ，则可以证明映射的乘积满足结合律，但不满足交换律，即分别有

$$(h \circ g) \circ f = h \circ (g \circ f)$$

1.2 线性空间定义及其性质

线性空间是线性代数最基本的概念之一，也是学习矩阵论的重要基础。

定义 1.1 设 K 是一个数域, V 是一个非空集合, 如果 V 满足以下条件:

1) 在 V 中定义了一个加法运算, 即给定一个法则, 对任意的 $\alpha, \beta \in V$, 通过此法则, 都有惟一确定的 V 中元素 $\alpha + \beta$ 与 (α, β) 对应, 元素 $\alpha + \beta$ 叫做 α 与 β 的和, 记为 $\alpha + \beta$ (即 $+$: $V \times V \rightarrow V$ 是映射)。

$$V \times V = \{(\alpha, \beta) \mid \alpha, \beta \in V\}$$

$$(\alpha, \beta) \mapsto \alpha + \beta$$

2) 在 V 中定义了一个数乘运算, 即给定一个法则, 对任意的 $k \in K$, 任意的 $\alpha \in V$, 通过此法则, 都有惟一确定的 V 中元素 $k\alpha$ 与 k 和 α 对应, 元素 $k\alpha$ 叫做 k 与 α 的数乘, 记为 $k\alpha$ (即 \cdot : $K \times V \rightarrow V$ 是映射)。

$$(k, \alpha) \mapsto (k\alpha) = k \cdot \alpha$$

3) 加法和数乘满足以下 8 条性质

- 1° 交换律 $\alpha + \beta = \beta + \alpha$;
- 2° 结合律 $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$;
- 3° 存在零元素 0 使 $\alpha + 0 = \alpha$;
- 4° 存在负元素, 即对任一元素 α 存在元素 $-\alpha$ 使 $\alpha + (-\alpha) = 0$, 叫 $-\alpha$ 的负元;
- 5° 分配律 $k(\alpha + \beta) = k\alpha + k\beta$
- 6° 分配律 $(k + l)\alpha = k\alpha + l\alpha$
- 7° $k(l\alpha) = (kl)\alpha$
- 8° $1\alpha = \alpha$

其中 $\alpha, \beta, \gamma \in V, k, l \in K$, 称 V 为数域 K 上的线性空间, 也叫向量空间, α, β, γ 称为向量。

V 中定义的加法及数乘运算统称为 V 的线性运算。

例 1 $K^{1 \times n} = \{ (a_1, a_2, \dots, a_n) \mid a_i \in K \}$ 关于向量的加法与数乘作成数域 K 上的线性空间。

例 2 $K^{m \times n} = \{ A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \mid a_{ij} \in K \}$

关于矩阵的加法及数乘做成数域 K 上的线性空间。

例 3 $K[x]_n = \{ \text{数域 } K \text{ 上次数不超过 } n \text{ 的多项式全体, 零多项式} \}$ 关于多项式的加法及数乘做成数域 K 上的线性空间。

例 4 设 R_+ 表示所有正实数集合, 其加法和数乘各定义为

$$\alpha \oplus \beta = \alpha\beta, \quad k \odot \alpha = \alpha^k$$

证明 R_+ 是 R 上的线性空间。

证明 设 $\alpha, \beta \in R_+, k \in R$, 则有

$$\alpha \oplus (\beta \odot k) = \alpha \oplus \beta^k = \alpha\beta^k = (\alpha\beta)^k = (\alpha \oplus \beta) \odot k$$

即 R_+ 对定义的加法“ \oplus ”与数乘运算“ \odot ”是封闭的, 且有

- 1 ° $1 \cdot 1 = 1$
- 2 ° $(k \cdot 1) = k = (1 \cdot k) = k$
- 3 ° 1 是零元素 0 , 因为 $1 + 1 = 1$
- 4 ° 1 的负元素是 $\frac{1}{1}$, 因为 $\frac{1}{1} = \frac{1}{1} = 1$
- 5 ° $k \cdot (1 \cdot 1) = k \cdot 1 = (k \cdot 1) = k$
- 6 ° $(k + 1) \cdot 1 = k + 1 = (k + 1) = k + 1$
- 7 ° $k \cdot (1 \cdot 1) = k \cdot 1 = (k \cdot 1) = k$
- 8 ° $1 \cdot 1 = 1$

成立, 故 R_+ 在上述定义的运算下成为 R 上的线性空间。

从以上例子看到, 线性空间的概念是很抽象的, 正因为抽象, 它才代表广泛的含意, 这源于集合元素具有广泛的含意, 而加法、数乘也仅仅是称呼而已, 与原来对数的加法, 数乘的理解完全不同。

线性空间的概念是从现实世界中得到的, 它以极度抽象的形式出现, 在表面上掩盖了它们源于外部世界的实质。它是描述自然现象中某种量的关系的数学概念, 只是由于在数学形式上与描述平常几何空间有很大类似, 因此也叫它作“空间”, 实际上“空间”已经推广了我们平常用语的含义了。

定理 1.2.1 线性空间 V 中的零元素 0 惟一, 记为 0 ; 任一元素 a 的负元素惟一, 记为 $-a$ 。

证明 若 $0_1, 0_2$ 为线性空间 V 的零元素, 考虑和

$$0_1 = 0_1 + 0_2 = 0_2$$

这就证明了零元素的惟一性。

设 a_1, a_2 都是 a 的负元, 所以 $a + a_1 = 0 = a + a_2$, 从而有

$$a_1 = a_1 + 0 = a_1 + (a + a_2) = (a_1 + a) + a_2 = 0 + a_2 = a_2$$

这就证明了负元素的惟一性。

利用负元素, 定义 V 中的减法如下

$$a - b = a + (-b)$$

命题 1 $k \cdot 0 = 0, 0 \cdot a = 0, (-1) \cdot a = -a$

证明 因为 $k \cdot 0 = k \cdot (0 + 0) = k \cdot 0 + k \cdot 0 = k \cdot 0$

由于零元的惟一性, 所以 $k \cdot 0 = 0$ 。

$$\text{因为 } k \cdot 0 = (k + 0) \cdot 0 = k \cdot 0$$

由于零元的惟一性, 所以 $0 \cdot a = 0$ 。

$$\text{因为 } (-1) \cdot a = 1 \cdot (-a) = (1 - 1) \cdot a = 0 \cdot a = 0$$

由于负元的惟一性, 所以 $(-1) \cdot a = -a$

命题 2 若 $k \cdot a = 0$, 则必有 $k = 0$ 或 $a = 0$ 。

证明 若 $k \cdot a = 0$, 且 $k = 0$, 则结论已经成立; 若 $k \neq 0$, 则 $\frac{1}{k} \cdot (k \cdot a) = \frac{1}{k} \cdot 0 = 0$, 也即 $\frac{1}{k} \cdot (k \cdot a) = (\frac{1}{k} k) \cdot a = 1 \cdot a = a = 0$

于是, 由上述, 可得符号法则

$$\begin{aligned} -(-\alpha) &= \alpha, & -(-\alpha) &= (-\alpha) - (-\alpha) \\ (-k)\alpha &= -k\alpha, & (-k)\alpha - (-k)\alpha &= k\alpha \end{aligned}$$

1.3 线性空间的基与坐标

在不至引起混淆的前提下， $\alpha + \beta$ 简记为 $\alpha + \beta$ ， $k\alpha = k\alpha$ 。

定义 1.2 设 V 是数域 K 上的线性空间， $\alpha_1, \alpha_2, \dots, \alpha_r \in V$ ，若存在不全为 0 的数 k_1, k_2, \dots, k_r ，使

$$k_1\alpha_1 + k_2\alpha_2 + \dots + k_r\alpha_r = 0$$

成立，称 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性相关，否则称 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性无关。

例如， $\alpha = (a_1 \ a_2 \ a_3)$ 与 $\beta = (2a_1 \ 2a_2 \ 2a_3)$ 就是 $K^{1 \times 3}$ 上的两个线性相关的向量； $\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$ 与 $\begin{bmatrix} 4 & 4 \\ 8 & 8 \end{bmatrix}$ 也是 $K^{2 \times 2}$ 上的线性相关的两个向量； $1, x, x^2$ 是 $K[x]_2$ 中三个线性无关的向量，因为若 $k_1 + k_2x + k_3x^2 = 0$ ，只有 $k_1 = k_2 = k_3 = 0$ 才成立。

定义 1.3 设 V 是数域 K 上的线性空间， $\alpha_1, \dots, \alpha_n (n \geq 1)$ 是 V 中任意 n 个向量，如果它满足

$\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关；

V 中任一向量 α 都可以由 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性表示。

则 $\alpha_1, \alpha_2, \dots, \alpha_n$ 为 V 的一个基(或基底)，并称 $\alpha_1, \alpha_2, \dots, \alpha_n$ 为基向量，且称 V 是 n 维向量空间(线性空间)。

由定义可见，基不过是线性空间 V 中的极大线性无关组而已；而线性空间的维数就是其基中所含向量的个数罢了！

例 1 K^n 中的元素 $\alpha_1 = (1 \ 0 \ \dots \ 0)^T, \dots, \alpha_n = (0 \ \dots \ 0 \ 1)^T$ 就是基， K^n 是 n 维。

例 2 $K^{m \times n}$ 中的元素 E_{ij} (i 行 j 列处元素为 1，其余全是零的矩阵) $i = 1, 2, \dots, m, j = 1, 2, \dots, n$ ，是一基， $K^{m \times n}$ 是 mn 维。

例 3 $K[x]$ 是无穷维空间，因为对任何正整数 $n, 1, x, x^2, \dots, x^n$ 都是线性无关的。

例 4 $K[x]_n$ 中的 $1, x, x^2, \dots, x^n$ 是一个基，维数为 $n + 1$ 。

例 5 1.2 节例 5 中的线性空间 R_+ ，任意选 $\alpha \in R_+, \alpha \neq 1$ ，则 α 是线性无关的。另选 $\beta \in R_+, \beta \neq 1$ ，设

$$k_1 \cdot \alpha + k_2 \cdot \beta = 0$$

即 $k_1 \cdot k_2 = 1$

两边取对数 $k_1 = \frac{\ln \alpha}{\ln \beta} k_2$

所以 α, β 是线性相关的，故可选 β 作 R_+ 的基，对任意 $\alpha \in R_+$
 $\alpha = \log_2 \alpha \cdot \beta$

因此 R_+ 是 1 维的。

需要指出，一个线性空间的基是不惟一的。

定理 1.3.1 设 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 n 维线性空间 V 的一个基底，对任意 $\alpha \in V$ ，则 α 可以惟一

表示为 $\alpha_1, \alpha_2, \dots, \alpha_n$ 的线性组合。

证明 设 $\alpha = k_1 \alpha_1 + k_2 \alpha_2 + \dots + k_n \alpha_n$
若还有 $\alpha = k'_1 \alpha_1 + k'_2 \alpha_2 + \dots + k'_n \alpha_n$
这两式相减, 得

$$(k_1 - k'_1) \alpha_1 + (k_2 - k'_2) \alpha_2 + \dots + (k_n - k'_n) \alpha_n = 0$$

因为 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关, 所以 $k_1 = k'_1, \dots, k_n = k'_n$ 。

定义 1.4 设 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 n 维线性空间 V 的一个基底, 对任意 $\alpha \in V$, 若 $\alpha = k_1 \alpha_1 + k_2 \alpha_2 + \dots + k_n \alpha_n$, 称 (k_1, k_2, \dots, k_n) 是 α 在基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 之下的坐标。

注意, 就是同一个向量 α 在不同基下的坐标一般是不相同的, 例如 $\alpha = (4 \ 3) = 4(1 \ 0) + 3(0 \ 1) = (1 \ 0) + 3(1 \ 1), (1 \ 0), (0 \ 1)$ 与 $(1 \ 0), (1 \ 1)$ 都可以作 K^2 的基底, 那么 α 在 $(1 \ 0), (0 \ 1)$ 之下的坐标是 $(4 \ 3)$, 而在 $(1 \ 0), (1 \ 1)$ 之下的坐标是 $(1 \ 3)$ 。

例 6 可以验证复数域 C 是自身上的关于数的加法、乘法作成的线性空间, 且维数 $=1$, 因为 " $\alpha \in C, \alpha = \alpha \cdot 1$ 。复数域 C 也可以看成实数域上的线性空间, 且维数 $=2$, 因为 " $\alpha \in C, \alpha = a + bi, a, b$ 是实数。复数域 C 也可看成有理数域 Q 上的线性空间, 此时 C 是无穷维的, 因为 $1, i, i^2, \dots, i^n, \dots$ 是线性无关的。

此例说明, 作为一个集合 V , 它形成一个线性空间时的维数与它所在的数域是紧密相关的。

1.4 基变换与坐标变换

1.4.1 基变换公式

在 n 维线性空间 V 中, 任意 n 个线性无关的向量都可以作它的基。对不同的基, 同一个向量的坐标一般是不同的。现在讨论当基改变时, 向量的坐标如何变化? 先介绍由 V 的一基底改变成另一基底的过渡矩阵的概念。

设 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 V 的一旧基, $\beta_1, \beta_2, \dots, \beta_n$ 是 V 的一新基, 则由基定义, 可得

$$\begin{aligned} \beta_1 &= c_{11} \alpha_1 + c_{21} \alpha_2 + \dots + c_{n1} \alpha_n \\ \beta_2 &= c_{12} \alpha_1 + c_{22} \alpha_2 + \dots + c_{n2} \alpha_n \\ &\vdots \\ \beta_n &= c_{1n} \alpha_1 + c_{2n} \alpha_2 + \dots + c_{nn} \alpha_n \end{aligned}$$

或者写成(形式上)

$$(\beta_1, \beta_2, \dots, \beta_n) = (\alpha_1, \alpha_2, \dots, \alpha_n) \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{pmatrix} \tag{1.1}$$

这里矩阵 $(\alpha_1, \alpha_2, \dots, \alpha_n)$ 的元素是向量, 矩阵 C 的元素是数, 这种不同类型的矩阵间的运算, 称为矩阵的形式运算, 一般无意义。不过在特殊情况下, 可使这种约定的算法不会出问题。

矩阵 $C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \cdots & \cdots & & \cdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$

称为由旧基变为新基的过渡矩阵, 式(1.1) 称为基变换公式。

可以证明过渡矩阵 C 是非奇异矩阵。

1.4.2 坐标变换公式

设 V , 在基 $\alpha_1, \dots, \alpha_n$ 之下的坐标为 (x_1, x_2, \dots, x_n) , 在基 β_1, \dots, β_n 之下的坐标为 (y_1, \dots, y_n) , 即

$$\alpha = x_1 \alpha_1 + \cdots + x_n \alpha_n = y_1 \beta_1 + \cdots + y_n \beta_n$$

采用形式写法, 并使用式(1.1) 便有

$$(y_1 \cdots y_n) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_n = (x_1 \cdots x_n) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_n = (x_1 \cdots x_n) C \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_n$$

由于 α 在基 β_1, \dots, β_n 之下的坐标是惟一的, 因此有

$$\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_n = C \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_n \tag{1.2}$$

或者

$$\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_n = C^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_n \tag{1.3}$$

式(1.2) 与式(1.3) 给出在基变换公式(1.1) 之下的向量坐标变换公式。

例 1 在 R^n 中, 设 $\alpha_1 = (1\ 0\ \cdots\ 0)$, $\alpha_2 = (0\ 1\ 0\ \cdots\ 0)$, \dots , $\alpha_n = (0\ \cdots\ 0\ 1)$, $\beta_1 = (1, 0, \dots, 0)$, $\beta_2 = (1, 1, 0, \dots, 0)$, \dots , $\beta_n = (1, \dots, 1)$, 已知向量 α 在 $\alpha_1, \dots, \alpha_n$ 之下坐标为 $(x_1 \cdots x_n)$, 求在 β_1, \dots, β_n 之下的坐标。

解 由(1.1) 有

$$(y_1 \cdots y_n) = (x_1 \cdots x_n) \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

于是得过渡矩阵 $C = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \cdots & \cdots & & \cdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$

不难求得 $C^{-1} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \dots & \dots & \dots & & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & -1 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}$

由公式(1.3)知，在 $(\alpha_1, \dots, \alpha_n)$ 之下的坐标 $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = C^{-1} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

也就是 $x_1 = x_1 - x_2, \dots, x_{n-1} = x_{n-1} - x_n, x_n = x_n$

例2 求 V_4 的基 $\alpha_1 = (1, 2, -1, 0), \alpha_2 = (1, -1, 1, 1), \alpha_3 = (-1, 2, 1, 1), \alpha_4 = (-1, -1, 0, 1)$

变为新基 $\beta_1 = (2, 1, 0, 1) \quad \beta_2 = (0, 1, 2, 2)$
 $\beta_3 = (-2, 1, 1, 2) \quad \beta_4 = (1, 3, 1, 2)$

时向量坐标的变换公式。

解 因为 $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (\beta_1, \beta_2, \beta_3, \beta_4) \begin{pmatrix} 1 & 1 & -1 & -1 \\ 2 & -1 & 2 & -1 \\ -1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix}$
 $= (\beta_1, \beta_2, \beta_3, \beta_4) A$
 $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (\beta_1, \beta_2, \beta_3, \beta_4) \begin{pmatrix} 2 & 0 & -2 & 1 \\ 1 & 1 & 1 & 3 \\ 0 & 2 & 1 & 1 \\ 1 & 2 & 2 & 2 \end{pmatrix}$
 $= (\beta_1, \beta_2, \beta_3, \beta_4) B$

所以有 $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (\beta_1, \beta_2, \beta_3, \beta_4) A^{-1} B$

于是得由旧基改变为新基的过渡矩阵是

$$C = A^{-1} B$$

由公式(1.3)可得向量的坐标变换公式为

$$(x_1, \dots, x_4) = (\alpha_1, \dots, \alpha_4) (C^{-1})^T$$

由计算可得

$$\begin{aligned} (C^{-1})^T &= (B^{-1} A)^T = A^T (B^T)^{-1} \\ &= \begin{pmatrix} 1 & 2 & -1 & 0 \\ 1 & -1 & 1 & 1 \\ -1 & 2 & 1 & 1 \\ -1 & -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4/13 & 2/13 & -3/13 & -1/13 \\ -6/13 & -3/13 & -2/13 & 8/13 \\ -8/13 & 9/13 & -7/13 & 2/13 \\ 11/13 & -1/13 & 8/13 & -6/13 \end{pmatrix} \\ &= \begin{pmatrix} 0 & -1 & 0 & 1 \\ 1 & 1 & 0 & -1 \\ -1 & 0 & 0 & 1 \\ 1 & 0 & 1 & -1 \end{pmatrix} \end{aligned}$$

例 2 集合 $C = \{ (a_{ij})_{n \times n} \mid a_{ij} \in K, \sum_{i=1}^n a_{ii} = 0 \}$ 容易验证 C 关于矩阵的加法和数乘也是 $K^{n \times n}$ 的一个线性子空间。

例 3 集合 $M = \{ (a_{ij})_{n \times n} \mid a_{ij} \in K, a_{ji} = a_{ij} \}$ 容易验证 M 也做成 $K^{n \times n}$ 的一个线性子空间。

由于线性子空间也是线性空间, 因此, 前面引入的关于维数、基和坐标等概念, 亦可以应用到线性子空间中去。

由于零子空间不含线性无关的向量, 因此它没有基, 并规定其维数为零。
因为线性子空间中不可能比在整个线性空间中有更多数目的线性无关的向量, 所以任何一个线性子空间 W 的维数不能大于整个线性空间 V 的维数, 即有

$$\dim W \leq \dim V \tag{1.4}$$

例如, n 元齐次线性方程组 $AX = 0$, 系数矩阵 A 的秩 $= r (1 \leq r < n)$ 时, 其解空间的维数等于 $n - r$, 小于 R^n 的维数, 前面例 3 中 M 的维数等于 $\frac{n(n+1)}{2}$, 例 2 中 C 的维数为 $n^2 - 1$ 。

1.5.2 线性子空间的生成问题

设 $\alpha_1, \alpha_2, \dots, \alpha_r$ 是数域 K 上的线性空间 V 的一组向量, 其所有可能的线性组合的集合

$$V_1 = \{ k_1 \alpha_1 + \dots + k_r \alpha_r \mid k_i \in K \}$$

是非空的, 容易验证 V_1 对 V 的线性运算封闭, 因而 V_1 是 V 的一个线性子空间。这个子空间称为由 $\alpha_1, \alpha_2, \dots, \alpha_r$ 生成(或张成)的子空间, 记为

$$L(\alpha_1, \dots, \alpha_r) = \{ k_1 \alpha_1 + \dots + k_r \alpha_r \mid k_i \in K \} \tag{1.5}$$

反过来, 在有限维线性空间 V 中, 它的任何一个线性子空间 W 都可以写成式(1.5)形式。事实上, 由于 W 也是有限维的, 设 $\alpha_1, \dots, \alpha_r$ 是 W 的一个基, 那么

$$W = \{ k_1 \alpha_1 + \dots + k_r \alpha_r \mid k_i \in K \} = L(\alpha_1, \dots, \alpha_r)$$

就是 r 维子空间。特别地, 零子空间就是由零元素生成的子空间 $L(0)$ 。

矩阵的值域和核空间(零空间)的理论, 在线性最小二乘法问题和广义逆矩阵的讨论中都占有重要地位, 现定义如下。

定义 1.6 令 $A = (a_{ij})_{m \times n} \in R^{m \times n}$, 以 $\alpha_i (i = 1, 2, \dots, n)$ 表示 A 的第 i 个列向量, 称子空间 $L(\alpha_1, \dots, \alpha_n)$ 为矩阵的值域(列空间), 记为

$$R(A) = L(\alpha_1, \dots, \alpha_n) \tag{1.6}$$

由前面的论述及矩阵秩的概念可知 $R(A) \in R^m$, 且有

$$\text{rank}(A) = \dim R(A)$$

同样可以定义 A^T 的值域

$$R(A^T) = L(\tilde{\alpha}_1, \dots, \tilde{\alpha}_m) \in R^n \tag{1.7}$$

且有 $\dim R(A) = \dim R(A^T) = r(A)$

定义 1.7 令 $A = (a_{ij})_{m \times n} \in R^{m \times n}$, 称集合 $\{ \alpha \mid A\alpha = 0, \alpha \in R^n \}$ 为 A 的核空间(或核、零空间), 记为 $N(A)$, 即

用 $\text{rank}(A)$ 表示 A 的秩, 或记 $r(A)$

$$N(A) = \{ \mid A = 0, \quad R^n \} \tag{1.8}$$

显然 $N(A)$ 是齐次线性方程组 $A = 0$ 的解空间, 它是 R^n 的一个子空间。A 的核空间的维数称为 A 的零度, 记为 $n(A)$, 即

$$n(A) = \dim N(A) = n - r(A)$$

例 4 已知 $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, 求 A 的秩及零度。

解 显然有 $x_1 + x_2 - x_3 = 0$, 即 A 的三个列向量线性相关。但 A 的前两个列向量线性无关, 故 $r(A) = 2, n(A) = 3 - 2 = 1, \dim R(A^T) = \dim R(A) = r(A) = 2$, 而 $n(A^T) = 2 - 2 = 0$

一般而言, 设 $A = (a_{ij})_{m \times n}$, 则下面一般公式

$$\text{rank}(A) + n(A) = n \tag{1.9}$$

$$n(A) - n(A^T) = n - m \tag{1.10}$$

事实上线性方程组 $A^T X = 0$ 的解空间维数 $= n(A^T) = m - \text{rank}(A^T)$, 所以 $n(A) - n(A^T) = n - \text{rank}(A) - m + \text{rank}(A^T) = n - m$ 。

定理 1.5.2 设 V_1 是数或 K 上 n 维线性空间 V 的一个 m 维子空间, $\alpha_1, \dots, \alpha_m$ 是 V_1 的一个基, 则这个基必可扩充为 V 的一个基。换言之, 在 V 中必可找到 $n - m$ 个向量 $\alpha_{m+1}, \dots, \alpha_n$, 使 $\alpha_1, \dots, \alpha_m, \alpha_{m+1}, \dots, \alpha_n$ 是 V 的一个基。

证明 对维数差 $n - m$ 用数学归纳法。

当 $n - m = 0$ 时, 定理显然成立, 因为 $\alpha_1, \dots, \alpha_n$ 已是 V 的基。设 $n - m = k$ 时, 定理成立, 考虑 $n - m = k + 1$ 的情形。

既然 $\alpha_1, \dots, \alpha_m$ 还不是 V 的基, 因为它们线性无关, 由基的定义, 在 V 中至少还有一个向量 α_{m+1} 不能被 $\alpha_1, \dots, \alpha_m$ 线性表示, 那么 $\alpha_1, \dots, \alpha_m, \alpha_{m+1}$ 必定线性无关(因为若 $\alpha_1, \dots, \alpha_m$ 线性无关, 而 $\alpha_1, \dots, \alpha_m, \alpha_{m+1}$ 线性相关, 则 α_{m+1} 一定被 $\alpha_1, \dots, \alpha_m$ 线性表示)。于是子空间 $V_2 = L(\alpha_1, \dots, \alpha_m, \alpha_{m+1})$ 是 $m + 1$ 维的, 因为 $n - \text{维}(V_2) = n - (m + 1) = k$, 由归纳假设, 子空间 V_2 的基 $\alpha_1, \dots, \alpha_{m+1}$ 可以扩为 V 的基 $\alpha_1, \dots, \alpha_m, \alpha_{m+1}, \dots, \alpha_n$, 证毕。

1.6 子空间的交与和

此节讨论的问题是子空间的运算生成的子空间问题。

1.6.1 子空间的交

定理 1.6.1 如果 V_1, V_2 是数域 K 上线性空间 V 的两个子空间, 那么它们的交 $V_1 \cap V_2$ 也是 V 的子空间。

证明: 因为 $0 \in V_1, 0 \in V_2$, 所以 $0 \in V_1 \cap V_2$, 故 $V_1 \cap V_2$ 非空, 若 $\alpha \in V_1 \cap V_2$, 由于 $\alpha \in V_1, \alpha \in V_2$, 而 V_1, V_2 都是子空间, 所以 $k\alpha \in V_1, k\alpha \in V_2$, 故 $k\alpha \in V_1 \cap V_2$ 。又因 $\alpha \in V_1, \alpha \in V_2$, 所以 $\alpha + \beta \in V_1 \cap V_2$ 。由定理 1.5.1 知 $V_1 \cap V_2$ 是 V 的线性子空间。

由集合的交的定义可知, 子空间的交满足交换律及结合律, 即有

$$\begin{aligned} V_1 \cap V_2 &= V_2 \cap V_1 \\ (V_1 \cap V_2) \cap V_3 &= V_1 \cap (V_2 \cap V_3) \end{aligned}$$

1.6.2 子空间的和

定义 1.8 设 V_1, V_2 都是数域 K 上线性空间 V 的子空间, $\alpha \in V_1, \beta \in V_2$, 则所有 $\alpha + \beta$ 这样元素的集合叫做 V_1 与 V_2 的和(或和空间), 记为 $V_1 + V_2$

$$V_1 + V_2 = \{ \alpha + \beta \mid \alpha \in V_1, \beta \in V_2 \}$$

定理 1.6.2 如果 V_1, V_2 都是数域 K 上线性空间 V 的子空间, 那么它们的和 $V_1 + V_2$ 也是 V 的子空间。

证明 显然 $V_1 + V_2$ 非空, 又对任意 $\alpha, \beta \in V_1 + V_2$ 。因为 $\alpha = \alpha_1 + \alpha_2, \beta = \beta_1 + \beta_2, \alpha_1, \alpha_2 \in V_1, \beta_1, \beta_2 \in V_2$, 则 $\alpha + \beta = \alpha_1 + \beta_1 + \alpha_2 + \beta_2$, 所以 $\alpha + \beta = (\alpha_1 + \beta_1) + (\alpha_2 + \beta_2) \in V_1 + V_2$ 。又 $k\alpha = k\alpha_1 + k\alpha_2 \in V_1 + V_2$, 故 $V_1 + V_2$ 是 V 的子空间。

由于 V 中元之间加法满足交换律与结合律, 所以子空间和运算满足交换律和结合律, 即

$$\begin{aligned} V_1 + V_2 &= V_2 + V_1 \\ (V_1 + V_2) + V_3 &= V_1 + (V_2 + V_3) \end{aligned}$$

例如, 在线性空间 R^3 中, V_1, V_2 表示两条过原点的不同直线 L_1, L_2 上所有向量形成的子空间, 则 $V_1 \cap V_2 = \{0\}$ 是只含原点的零子空间, $V_1 + V_2$ 就是由 L_1 与 L_2 所决定的平面上全体向量做成的子空间。

显然有 $V_i \cap (V_1 + V_2) = V_i, V_i \subset V_1 + V_2, i = 1, 2$

1.6.3 两子空间的交与和的维数公式

定理 1.6.3(维数公式) 如果 V_1, V_2 是 V 的子空间, 那么有下面公式

$$\dim V_1 + \dim V_2 = \dim(V_1 + V_2) + \dim(V_1 \cap V_2) \tag{1.11}$$

证明 设 $\dim V_1 = n_1, \dim V_2 = n_2, \dim(V_1 \cap V_2) = r$, 设 $\alpha_1, \dots, \alpha_r$ 是 $V_1 \cap V_2$ 的基, 由扩基定理, 可将它们分别扩为 V_1, V_2 的基

$$\begin{aligned} \alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_{n_1-r} \\ \alpha_1, \dots, \alpha_r, \gamma_1, \dots, \gamma_{n_2-r} \end{aligned}$$

证明 $\alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_{n_1-r}, \gamma_1, \dots, \gamma_{n_2-r}$ 是 $V_1 + V_2$ 的基。这样一来, $\dim(V_1 + V_2) = n_1 + n_2 - r$, 从而公式(1.11)成立。

因为 V_1 中任意元都可以由 $\alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_{n_1-r}$ 线性表示, 当然也可以由 $\alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_{n_1-r}, \gamma_1, \dots, \gamma_{n_2-r}$ 线性表示, 同理 V_2 中的任意元也可以由 $\alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_{n_1-r}, \gamma_1, \dots, \gamma_{n_2-r}$ 线性表示, 所以 $V_1 + V_2 = L(\alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_{n_1-r}, \gamma_1, \dots, \gamma_{n_2-r})$

还需证明这 $n_1 + n_2 - r$ 个向量线性无关。设

$$k_1 \alpha_1 + \dots + k_r \alpha_r + l_1 \beta_1 + \dots + l_{n_1-r} \beta_{n_1-r} + h_1 \gamma_1 + \dots + h_{n_2-r} \gamma_{n_2-r} = 0$$

设 $h_1 \gamma_1 + \dots + h_{n_2-r} \gamma_{n_2-r} = -k_1 \alpha_1 - \dots - k_r \alpha_r - l_1 \beta_1 - \dots - l_{n_1-r} \beta_{n_1-r}$

所以 $\gamma_1, \dots, \gamma_{n_2-r} \in V_1 \cap V_2$, 所以 $h_1 \gamma_1 + \dots + h_{n_2-r} \gamma_{n_2-r} = 0$

则有 $k_1 \alpha_1 + \dots + k_r \alpha_r + l_1 \beta_1 + \dots + l_{n_1-r} \beta_{n_1-r} = 0, \alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_{n_1-r}$ 是 V_1 的基, 因此它们线性无关, 所以有

$$\begin{aligned} k_1 = \dots = k_r = h_1 = \dots = h_{n_2-r} &= 0 \\ k_1 = \dots = k_r = l_1 = \dots = l_{n_1-r} &= 0 \end{aligned}$$

从而也有

这就证明了 $\alpha_1, \dots, \alpha_r, \alpha_{r+1}, \dots, \alpha_{n_1-r}, \alpha_{r+1}, \dots, \alpha_{n_2-r}$ 线性无关, 证毕。

推论 1 维 $V_1 + V_2 = \text{维}(V_1 + V_2)$
推论 2 若 V_1, V_2 都是 n 维线性空间 V 的子空间, 且 $\text{维 } V_1 + \text{维 } V_2 > n$, 则 $\text{维}(V_1 \cap V_2) = \{0\}$

例 1 设 V 是线性空间, $\alpha_1, \dots, \alpha_r, \alpha_{r+1}, \dots, \alpha_{r+s} \in V$
证明 $L(\alpha_1, \dots, \alpha_r) + L(\alpha_{r+1}, \dots, \alpha_{r+s}) = L(\alpha_1, \dots, \alpha_r, \alpha_{r+1}, \dots, \alpha_{r+s})$
证明 只需证明集合相等。对任意 $\alpha \in L(\alpha_1, \dots, \alpha_r) + L(\alpha_{r+1}, \dots, \alpha_{r+s})$, 由定理 $\alpha = \beta + \gamma$, $\beta \in L(\alpha_1, \dots, \alpha_r)$, 所以 $\beta = k_1 \alpha_1 + \dots + k_r \alpha_r$, $\gamma \in L(\alpha_{r+1}, \dots, \alpha_{r+s})$, 所以 $\gamma = l_1 \alpha_{r+1} + \dots + l_s \alpha_{r+s}$, 则 $\alpha = k_1 \alpha_1 + \dots + k_r \alpha_r + l_1 \alpha_{r+1} + \dots + l_s \alpha_{r+s} \in L(\alpha_1, \dots, \alpha_r, \alpha_{r+1}, \dots, \alpha_{r+s})$ 。
反过来对任意 $\alpha \in L(\alpha_1, \dots, \alpha_r, \alpha_{r+1}, \dots, \alpha_{r+s})$, 那么 $\alpha = k_1 \alpha_1 + \dots + k_r \alpha_r + l_1 \alpha_{r+1} + \dots + l_s \alpha_{r+s}$, 所以 $\alpha = \beta + \gamma$, 其中 $\beta = k_1 \alpha_1 + \dots + k_r \alpha_r \in L(\alpha_1, \dots, \alpha_r)$, $\gamma = l_1 \alpha_{r+1} + \dots + l_s \alpha_{r+s} \in L(\alpha_{r+1}, \dots, \alpha_{r+s})$, 则 $\alpha \in L(\alpha_1, \dots, \alpha_r) + L(\alpha_{r+1}, \dots, \alpha_{r+s})$, 所以 $L(\alpha_1, \dots, \alpha_r) + L(\alpha_{r+1}, \dots, \alpha_{r+s}) = L(\alpha_1, \dots, \alpha_r, \alpha_{r+1}, \dots, \alpha_{r+s})$ 。

例 2 设 $\alpha_1 = (1, 0, 1, 1), \alpha_2 = (1, 1, -1, 0)$
 $\alpha_3 = (4, 5, -1, -1), \alpha_4 = (4, 3, -2, 1), \alpha_5 = (1, 1, 0, 0)$
求 $L(\alpha_1, \alpha_2)$ 与 $L(\alpha_3, \alpha_4, \alpha_5)$ 的和与交的基与维数。

解 令 $V_1 = L(\alpha_1, \alpha_2)$, 显然维 $V_1 = 2$, 基是 α_1, α_2 。
令 $V_2 = L(\alpha_3, \alpha_4, \alpha_5)$, 显然维 $V_2 = 3$, 基是 $\alpha_3, \alpha_4, \alpha_5$ 。

由例 1 可知, $V_1 + V_2 = L(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$, 求和空间 $V_1 + V_2$ 的基。

$$\begin{pmatrix} 1 & 1 & 4 & 4 & 1 \\ 0 & 1 & 5 & 3 & 1 \\ 1 & -1 & -1 & -2 & 0 \\ 1 & 0 & -1 & 1 & 0 \end{pmatrix} \xrightarrow{\text{行初等变换}} \begin{pmatrix} 1 & 1 & 4 & 4 & 1 \\ 0 & 1 & 5 & 3 & 1 \\ 0 & -2 & -5 & -6 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

所以 $V_1 + V_2$ 的维数 $= 3$, $\alpha_1, \alpha_2, \alpha_3$ 可作 $V_1 + V_2$ 的基。由维数公式可知 $\text{维}(V_1 \cap V_2) = \text{维 } V_1 + \text{维 } V_2 - \text{维}(V_1 + V_2) = 2 + 3 - 3 = 2$ 。

设 $\alpha \in V_1 \cap V_2$, 则 $\alpha \in V_1$, 所以 $\alpha = k_1 \alpha_1 + k_2 \alpha_2$, 又 $\alpha \in V_2$, 所以 $\alpha = l_1 \alpha_3 + l_2 \alpha_4 + l_3 \alpha_5$, 那么

$$\begin{cases} k_1 \alpha_1 + k_2 \alpha_2 - l_1 \alpha_3 - l_2 \alpha_4 - l_3 \alpha_5 = 0 \\ k_1 + k_2 - 4l_1 - 4l_2 - l_3 = 0 \\ k_2 - 5l_1 - 3l_2 - l_3 = 0 \\ k_1 - k_2 + l_1 + 2l_2 = 0 \\ k_1 + l_1 - l_2 = 0 \end{cases}$$

解此齐次性方程组, 得基础解系 $(1, 3, 0, 1, 0), (1, 0, -1, 0, 5)$ 。于是 $\alpha = \alpha_1 + 3\alpha_2 = \alpha_3$, $\alpha = \alpha_1 = -\alpha_3 + 5\alpha_5$, 由于 $\text{维}(V_1 \cap V_2) = 2$, 故可选 2 个线性无关向量就行了, 所以 α_3, α_5 可作基。

有时我们特别关心 $V_1 + V_2$ 的元的分解表达是否惟一, 例如
 $V_1 = L(\alpha_1, \alpha_2), \alpha_1 = (1, 0, 0), \alpha_2 = (1, 1, 1), V_2 = L(\alpha_3, \alpha_4), \alpha_3 = (0, 0, 1), \alpha_4 = (3, 1, 2)$
则 $V_1 + V_2$ 中的零元
 $0 = (0, 0, 0) = (0, 0, 0) + (0, 0, 0) = (2\alpha_1 + \alpha_2) - (\alpha_2 - \alpha_1)$ 这说明零元的分解表达式不惟一。
定义 1.9 如果 $V_1 + V_2$ 中每个元的分解表达式惟一, 即" $\alpha \in V_1 + V_2$ 只有惟一。 $\alpha = \alpha_1 + \alpha_2$, 称 $V_1 + V_2$ 是直和, 记 $V_1 \oplus V_2$ 。

定理 1.6.4 $V_1 + V_2$ 是直和的充分必要条件是 $V_1 \cap V_2 = \{0\}$

证明 $V_1 + V_2$ 是直和, 则零元分解式惟一, 即 $0 = 0 + 0$ 。设 $\alpha \in V_1 \cap V_2$, $\alpha = \alpha_1 + \alpha_2 = 0 + 0$, 其中 $\alpha_1 \in V_1, \alpha_2 \in V_2$, 故 $\alpha_1 = -\alpha_2 \in V_1 + V_2$, 所以, $\alpha = 0$, 即 $V_1 \cap V_2 = \{0\}$ 。

反之设 $V_1 \cap V_2 = \{0\}$, 若 $V_1 + V_2$ 不是直和, 那么至少有一个元的分解表达式不惟一, 设 $\alpha = \alpha_1 + \alpha_2 = \beta_1 + \beta_2$, 其中 $\alpha_1, \beta_1 \in V_1, \alpha_2, \beta_2 \in V_2$, 则 $(\alpha_1 - \beta_1) + (\alpha_2 - \beta_2) = 0$, 即 $\alpha_1 - \beta_1 = \beta_2 - \alpha_2$, 所以 $\alpha_1 - \beta_1 \in V_1 \cap V_2, \beta_2 - \alpha_2 \in V_1 \cap V_2$, 故 $\alpha_1 = \beta_1, \alpha_2 = \beta_2$, 证毕。

推论 设 $V_1 = L(\alpha_1, \dots, \alpha_r), V_2 = L(\beta_1, \dots, \beta_s)$, $\alpha_1, \dots, \alpha_r$ 与 β_1, \dots, β_s 分别是 V_1 与 V_2 的基, 则 $V_1 + V_2$ 是直和的充分必要条件是 $\alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_s$ 线性无关。

子空间的直和概念及定理 1.6.4 可以推广到多个子空间的情形: 设 $V_i (i = 1, 2, \dots, s)$ 是线性空间 V 的子空间。如果和 $\sum_{i=1}^s V_i$ 中每向量 α 的分解式 $\alpha = \alpha_1 + \dots + \alpha_s, \alpha_i \in V_i (i = 1, 2, \dots, s)$ 是惟一的, 则称该和为直和, 记为

$$V_1 \oplus V_2 \oplus \dots \oplus V_s = \sum_{i=1}^s V_i$$

同时, $\sum_{i=1}^s V_i$ 是直和的充要条件是 $V_i \cap \sum_{j=1, j \neq i}^s V_j = \{0\}, i = 1, 2, \dots, s$, 或

$$\dim(V_1 + \dots + V_s) = \dim V_1 + \dots + \dim V_s$$

习 题 1

1. 设 S_1, S_2 是两个集合, 且 $S_1 \cap S_2 = \emptyset$, 证明 $S_1 \cup S_2 = S_1 \oplus S_2 = S_2$
2. 判别数集 $K = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ 是否为数域。
3. 判别下列集合对所指运算是否构成实数域 \mathbb{R} 上的线性空间。
(1) 所有 n 阶实对称矩阵, 对矩阵的加法及数乘矩阵运算;
(2) 所有 n 阶实反对称阵, 对矩阵的加法及数乘矩阵运算;
(3) 所有 n 阶实可逆阵, 对矩阵的加法及数乘矩阵运算;
(4) 平面上全体向量的集合, 于对通常向量的加法及如下定义的数乘运算 $k \cdot \alpha = 0$ 。
4. 求习题 3 中各线性空间的维数, 并各选出它们的一组基。
5. 证明在实函数空间中, $1, \cos^2 t, \cos 2t$ 是线性相关的。
6. 求线性方程组
$$\begin{cases} x_1 + x_2 - 3x_3 - x_4 = 0 \\ 3x_1 - x_2 - 3x_3 + 4x_4 = 0 \\ x_1 + 5x_2 - 9x_3 - 8x_4 = 0 \end{cases}$$
 的解空间的维数与基。
7. 求 \mathbb{R}^3 中向量 $\alpha = (3, 7, 1)$ 对基 $\alpha_1 = (1, 3, 5), \alpha_2 = (6, 3, 2), \alpha_3 = (3, 1, 0)$ 的坐标。
8. 求 $\mathbb{P}[t]_2$ 中向量 $1 + t + t^2$ 对基 $1, t - 1, (t - 2)(t - 1)$ 的坐标。
9. 求 \mathbb{R}^3 中向量 α 对两个基 $\alpha_1 = (1, 2, 1), \alpha_2 = (2, 3, 3), \alpha_3 = (3, 7, 1), \beta_1 = (3, 1, 4),$

$\alpha_2 = (5, 2, 1)$, $\alpha_3 = (1, 1, -6)$ 的不同坐标间的关系。

10. 在 R^4 中有两个基 $\alpha_1 = \alpha_1, \alpha_2 = \alpha_2, \alpha_3 = \alpha_3, \alpha_4 = \alpha_4, \beta_1 = (2, 1, -1, 1), \beta_2 = (0, 3, 1, 0), \beta_3 = (5, 3, 2, 1), \beta_4 = (6, 6, 1, 3)$ 。

- (1) 求由基 $\alpha_1, \dots, \alpha_4$ 到基 β_1, \dots, β_4 的过渡矩阵;
- (2) 求向量 $(1, 0, 1, 0)$ 在基 $\alpha_1, \dots, \alpha_4$ 之下的坐标;
- (3) 求对两个基有相同坐标的非零向量。

11. 假定 $\alpha_1, \alpha_2, \alpha_3$ 是 R^3 的一个基, 试求由 $\beta_1 = \alpha_1 - 2\alpha_2 + 3\alpha_3, \beta_2 = 2\alpha_1 + 3\alpha_2 + 2\alpha_3, \beta_3 = 4\alpha_1 + 13\alpha_2$ 生成的子空间 $L(\alpha_1, \alpha_2, \alpha_3)$ 的基。

12. 求 R^4 的子空间 $V_1 = \{ (a_1, a_2, a_3, a_4) \mid a_1 - a_2 + a_3 - a_4 = 0 \}$
 $V_2 = \{ (a_1, a_2, a_3, a_4) \mid a_1 + a_2 + a_3 + a_4 = 0 \}$

的交 $V_1 \cap V_2$ 的基。

13. 在 R^n 中, 分量满足下列条件的全体向量的集合能否形成子空间?

- (1) $\alpha_1 + \alpha_2 + \dots + \alpha_n = 0$
- (2) $\alpha_1 + \alpha_2 + \dots + \alpha_n = 1$ 。

14. 设 $\alpha_1, \dots, \alpha_n$ 是 n 维线性空间 V 的一组基, A 是 $n \times s$ 矩阵

$$(\alpha_1, \dots, \alpha_s) = (\alpha_1, \dots, \alpha_n) A$$

证明: $L(\alpha_1, \dots, \alpha_s)$ 的维数 = A 的秩。

第 2 章

内 积 空 间

上一章用纯代数方法给出了空间中一些有几何色彩的概念: 空间、子空间、基底、维数、坐标等。但是反映欧氏几何的一些本质的概念, 如向量的长度, 向量之间夹角, 向量的正交性, 标准正交基等都还没有提出。如何在抽象的线性空间中引入上述类似的定义并演绎出相应的结果, 就是本章的任务。我们知道, 在解析几何中, 上述概念都与向量的点积(以后改称内积)有直接的联系。譬如, 向量 $\alpha = (a_1 \ a_2 \ a_3)$ 与 $\beta = (b_1 \ b_2 \ b_3)$ 的点积定义为

$$(\alpha, \beta) = a_1 b_1 + a_2 b_2 + a_3 b_3 = \alpha^T \beta$$

而向量 α 的长度(以后改称范数)定义为

$$|\alpha| = \sqrt{a_1^2 + a_2^2 + a_3^2} = \sqrt{(\alpha, \alpha)}$$

非零向量 α 与 β 之间的夹角余弦为

$$\cos \angle(\alpha, \beta) = \frac{(\alpha, \beta)}{|\alpha| |\beta|} = \frac{a_1 b_1 + a_2 b_2 + a_3 b_3}{\sqrt{a_1^2 + a_2^2 + a_3^2} \cdot \sqrt{b_1^2 + b_2^2 + b_3^2}}$$

特别是当 $(\alpha, \alpha) = 0$ 时, 规定 $|\alpha| = 0$, 还有点 α, β 之间的距离定义为

$$d(\alpha, \beta) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2} = |\alpha - \beta|$$

这表明, 一旦有了点积的定义, 长度、夹角、距离等概念便逐个诱导出来。所以首先想到要在抽象的线性空间中引入向量内积的概念, 为此, 需要考查 R^3 中向量点积的基本性质。显然, 两个向量的点积是一个标量(即实数), 所以可以把它看成定义在 R^3 上的二元实值函数, 且满足下述的性质:

- 1. 正定性 $(\alpha, \alpha) \geq 0$, 当且仅当 $\alpha = 0$ 时, $(\alpha, \alpha) = 0$;
- 2. 对称性 $(\alpha, \beta) = (\beta, \alpha)$;
- 3. 可加性 $(\alpha + \beta, \gamma) = (\alpha, \gamma) + (\beta, \gamma)$;
- 4. 齐次性 $(k\alpha, \beta) = k(\alpha, \beta)$ 。

当然点积还有其他性质。数学家经过相当长时间的反复研究、归纳, 认为这些性质是最根本的, 可以把它当作公理体系在抽象的线性空间中定义内积。

2.1 欧氏空间

2.1.1 欧氏空间定义

定义 2.1 设 V 是实数域 R 上的线性空间, 如果 V 中任意两个向量 α, β 都按某一确定的法则对应于惟一确定的实数, 记作 (α, β) , 并且 (α, β) 满足

- 对任意的 $\alpha, \beta \in V$, 有 $(\alpha, \beta) = (\beta, \alpha)$
- 对任意的 $\alpha, \beta, \gamma \in V$, 有 $(\alpha + \beta, \gamma) = (\alpha, \gamma) + (\beta, \gamma)$
- 对任意的 $k \in R, \alpha \in V$ 有 $(k\alpha, \beta) = k(\alpha, \beta)$
- 对任意的 $\alpha \in V$, 有 $(\alpha, \alpha) \geq 0$, 当且仅当 $\alpha = 0$ 时, $(\alpha, \alpha) = 0$

则称 (α, β) 为向量 α 与 β 的内积。定义了内积的实线性空间 V 称为欧几里德 (Enclide) 空间, 简称为欧氏空间。

例 1 在 n 维向量空间 R^n 中, 对任意的向量 $\alpha = (a_1, a_2, \dots, a_n), \beta = (b_1, b_2, \dots, b_n)$ 定义

$$(\alpha, \beta) = a_1 b_1 + a_2 b_2 + \dots + a_n b_n \tag{2.1}$$

显然这样定义的 (α, β) 满足定义 2.1 中的 4 个条件, 于是 R^n 关于这个内积做成一个欧氏空间。

很明显, 在 $n = 3$ 时, 式 (2.1) 就是三维几何空间中向量在直角坐标系下的点积表达式。

例 2 在 $n \times n$ 维空间 $R^{n \times n}$ 中, 对任意向量 $A = (a_{ij})_{n \times n}, B = (b_{ij})_{n \times n}$ 定义

$$(A, B) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} b_{ij} \tag{2.2}$$

则有

$$(A, B) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} b_{ij} = \text{Tr}(AB^T)$$

从而有

(1) 对任意的 $A, B \in R^{n \times n}$, 有 $(A, B) = \text{Tr}(AB^T) = \text{Tr}(BA^T) = (B, A)$

(2) 对任意的 $A, B, C \in R^{n \times n}$, 有

$$\begin{aligned} (A + B, C) &= \text{Tr}((A + B)C^T) = \text{Tr}(AC^T + BC^T) \\ &= \text{Tr}(AC^T) + \text{Tr}(BC^T) = (A, C) + (B, C) \end{aligned}$$

(3) 对任意的 $k \in R, A, B \in R^{n \times n}$, 有

$$\begin{aligned} (kA, B) &= \text{Tr}((kA)B^T) = \text{Tr}(kAB^T) \\ &= k\text{Tr}(AB^T) = k(A, B) \end{aligned}$$

(4) 对任意的 $A = (a_{ij})_{n \times n}$, 有

$$(A, A) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} a_{ij} \geq 0$$

$(A, A) = 0$ 的必要充分条件是 $a_{ij} = 0 \quad (i, j = 1, 2, \dots, n)$, 即 $A = 0$ 。

因此 (2.2) 式满足定义 2.1 的 4 个条件, 于是 $R^{n \times n}$ 对这个内积作成是一个欧氏空间。

例 3 在闭区间 $[a, b]$ 上一切 x 的连续实函数集合 $C[a, b]$, 对于通常函数的加法和实数

与函数的乘法运算构成一实线性空间, 对任意的 $f(x), g(x) \in C[a, b]$, 定义

$$(f(x), g(x)) = \int_a^b f(x)g(x)dx \tag{2.3}$$

容易验证(2.3) 式满足定义 2.1 中的 4 个条件。于是 $C[a, b]$ 关于上述内积作成一个欧氏空间。

对于一个实线性空间 V , 可以用不同的方法来规定内积(见习题 1)。对于这些不同的内积, V 做成不同的欧氏空间。例如, 在 R^n 中, 还可以定义 $(\alpha, \beta) = a_1b_1 + 2a_2b_2 + \dots + na_nb_n$, 容易验证这样定义的运算也满足定义 2.1 中的 4 个条件。不过我们约定, 今后凡称 R^n 为欧氏空间时, 其内积都按(2.1) 式规定(除了特别声明外) 的运算。

必须指出, 在线性空间中定义内积与向量的加法和数乘运算彼此是无关的, 因此无论内积如何规定都不影响该向量空间的维数, 从而可知欧氏空间的子空间仍是欧氏空间。

下面讨论欧氏空间的向量内积性质

1) 对任意的 $\alpha \in V$ (V 是欧氏空间) 则

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

事实上

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

特别是, 若 α 对任何向量 $\beta \in V$, 都有 $(\alpha, \beta) = 0$, 那么对 $\beta = \alpha$ 时也有 $(\alpha, \alpha) = (\alpha, \alpha) = 0$, 必有 $\alpha = 0$ 。

2) 对任意的 $\alpha, \beta \in V$, 恒有

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

事实上

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

类似地, 对任意的 $\alpha, \beta \in V, k \in R$, 有

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

3) 对欧氏空间中任意向量 $\alpha_1, \alpha_2, \dots, \alpha_s, \beta_1, \beta_2, \dots, \beta_t$ 及实数 $k_1, \dots, k_s, l_1, \dots, l_t$, 用数学归纳法不难证明

$$(\sum_{i=1}^s k_i \alpha_i, \sum_{j=1}^t l_j \beta_j) = \sum_{i=1}^s \sum_{j=1}^t k_i l_j (\alpha_i, \beta_j)$$

下面利用欧氏空间中内积的概念引进向量的长度和夹角的定义。对于欧氏空间的任意向量 α 来说 (α, α) 总是一个非负实数, 因此要定义向量的长度。

2.1.2 向量的长度

定义 2.2 设 α 是欧氏空间的一个向量, 非负实数 (α, α) 的算术根 $\sqrt{(\alpha, \alpha)}$ 称为 α 的长度, 并记为 $|\alpha|$ 。

向量的长度具有下述性质, 即

定理 2.1.1 设 V 是欧氏空间, 则对任意 $\alpha, \beta \in V$ 及 $k \in R$, 有

- (1) 齐次性 $|k\alpha| = |k||\alpha|$
- (2) 非负性 $|\alpha| \geq 0$ 当且仅当 $\alpha = 0$ 时, $|\alpha| = 0$
- (3) Canchy-Schwarz 不等式:

$$(\alpha, \alpha)^2 = (\alpha, \alpha)(\alpha, \alpha)$$

当且仅当 α 与 α 线性相关时, 等式成立。

证明 性质(1)(2)显然成立, 下面只证性质(3)

如果 α, β 线性相关, 那么 $\beta=0$ 或 $\beta=k\alpha$, 不论哪一种情况都有

$$(\alpha, \beta)^2 = (\alpha, \alpha)(\beta, \beta)$$

如果 α, β 线性无关, 那么对任何实数 k , 都有 $\alpha - k\beta \neq 0$, 于是

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

按内积性质有 $(\alpha, \alpha)k^2 - 2(\alpha, \beta)k + (\beta, \beta) > 0$

这是一个关于 k 的二次三项式, 对任意的实数 k 皆大于 0, 故其判别式一定为负, 即

$$= 4(\alpha, \beta)^2 - 4(\alpha, \alpha)(\beta, \beta) < 0$$

从而有

$$(\alpha, \beta)^2 < (\alpha, \alpha)(\beta, \beta)$$

反过来, 若对于欧氏空间中的向量 α, β , 有

$$(\alpha, \beta)^2 = (\alpha, \alpha)(\beta, \beta)$$

也就是说, 二次三项式 $(\alpha, \alpha)x^2 - 2(\alpha, \beta)x + (\beta, \beta)$ 的判别式

$$= 4(\alpha, \beta)^2 - 4(\alpha, \alpha)(\beta, \beta) = 0$$

那么一定有实数 $x=k$, 使

$$(\alpha, \beta)k^2 - 2(\alpha, \beta)k + (\beta, \beta) = 0$$

此即

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

从而有 $\alpha - k\beta = 0$, 即 α, β 线性相关, 于是定理成立。

在欧氏空间 R^n 中, $\alpha = (a_1, \dots, a_n), \beta = (b_1, \dots, b_n)$, 就有

$$|a_1b_1 + a_2b_2 + \dots + a_nb_n| \leq \sqrt{a_1^2 + \dots + a_n^2} \cdot \sqrt{b_1^2 + \dots + b_n^2}$$

在欧氏空间 $C[a, b]$ 中, $f(x), g(x) \in C[a, b]$, 就有

$$\left| \int_a^b f(x)g(x)dx \right| \leq \left[\int_a^b f^2(x)dx \right]^{\frac{1}{2}} \left[\int_a^b g^2(x)dx \right]^{\frac{1}{2}}$$

它们都是数学史上著名的不等式。

长度为 1 的向量称为单位向量。如果 $\alpha \neq 0$, 则有

$$\left| \frac{\alpha}{|\alpha|} \right| = \frac{1}{|\alpha|} |\alpha| = 1$$

即 $\frac{\alpha}{|\alpha|}$ 是单位向量。利用非 0 向量 α 的长度的倒数 $\frac{1}{|\alpha|}$ 去乘向量 α , 使它化为单位向量的方法

称为把向量 α 单位化。

定义 2.3 设 α, β 是欧氏空间中任意两个非零向量, 称

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

为 α 与 β 的夹角。注意这样定义的两个非零向量的夹角总介于 0 与 π 之间。零向量与其他向量的夹角认为是不确定的。

如果欧氏空间中两向量 α, β 的内积 $(\alpha, \beta) = 0$, 则称 α 与 β 正交, 记作 $\alpha \perp \beta$, 显然, 零向

量与任意向量都正交。

利用 Cauchy-Schwarz 不等式及正交概念可得如下命题。

命题 1 $|x| + |y| \geq |x + y|$ (三角不等式)。

证明 因为 $|x + y|^2 = (x + y, x + y) = |x|^2 + 2(x, y) + |y|^2$
 $|x|^2 + 2|x||y| + |y|^2$
 $= (|x| + |y|)^2$

所以 $|x| + |y| \geq |x + y|$

命题 2 $|x - y| \geq |x| - |y|$, $|x - y| \geq |y| - |x|$

证明 可令 $z = (x - y)$ +

由命题 1

$$|x - y| = |(x - y) + (y - y)| \geq |x - y| + |y - y|$$

移项即得 $|x - y| \geq |x| - |y|$

因为 $|x - y| = |(-y + x)| = |-y + x| \geq |-y| + |x| = |y| - |x|$

命题 3 若 x 与 y 正交, 则 $|x + y|^2 = |x|^2 + |y|^2$ (勾股定理)

证明 因为 $(x, y) = 0$, 所以 $(x, y) = 0$, 由命题 1 的证明过程可知结论正确。

2.2 标准正交基与 Gram-Schmidt 过程

2.2.1 标准正交基

在空间解析几何中, 常取两两正交的单位向量 i, j, k 来线性表出空间的所有向量, 这给问题的研究带来极大的方便。那么, 在抽象的欧氏空间, 能否也找到这样的一组向量, 它们又具备哪些优点, 下面就来回答这个问题。

定义 2.4 在欧氏空间 V 中, 一组不含零向量的向量组 $\alpha_1, \dots, \alpha_s$, 如果两两正交, 则称之为一个正交向量组。

定理 2.2.1 正交向量组是线性无关的。

证明 设 $\alpha_1, \dots, \alpha_s$ 是正交向量组, $k_1, \dots, k_s \in \mathbb{R}$, 使得

$$k_1 \alpha_1 + \dots + k_s \alpha_s = 0$$

则对任意的 $\alpha_i (i = 1, 2, \dots, s)$, 有

$$(k_1 \alpha_1 + \dots + k_s \alpha_s, \alpha_i) = 0$$

另一方面, 由内积性质及两两正交性, 有

$$(k_1 \alpha_1 + \dots + k_s \alpha_s, \alpha_i) = k_i (\alpha_i, \alpha_i)$$

所以只能 $k_i = 0 (i = 1, 2, \dots, s)$

故 $\alpha_1, \dots, \alpha_s$ 线性无关。

这个定理说明, 在 n 维欧氏空间中, 两两正交的非零向量不能超过 n 个。这个事实的几何意义很清楚, 例如, 在平面上就找不到三个两两正交的向量。另一方面, 这个定理还说明, n 个两两正交的非零向量组可以作为 n 维欧氏空间的基。

定义 2.5 在 n 维欧氏空间中, 由 n 个向量组成的正交向量组称为正交基, 由单位向量组

成的正交基叫标准正交基。

例如 R^n 中的 $\alpha_1 = (1, 0, \dots)$, $\alpha_2 = (0, 1, 0 \dots 0)$, $\alpha_n = (0, \dots, 0, 1)$ 就是一个标准正交基。
很显然, n 维欧氏空间中, $\alpha_1, \dots, \alpha_n$ 是标准正交基的充分必要条件是

$$(\alpha_i, \alpha_j) = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \quad i, j = 1, 2, \dots, n$$

那么一般的 n 维欧氏空间中是否有标准正交基呢? 若存在, 又怎样求出它呢? 下面的定理回答了这个问题。

2.2.2 求标准正交基的 Schmide 方法

定理 2.2.2 n 维欧氏空间中必有标准正交基。

证明 证明是构造性的, 从 n 维欧氏空间的任意一组给定的基出发, 通过 Gram-Schmidt 正交化方法找出一个标准正交基。

设 $\alpha_1, \dots, \alpha_n$ 是 n 维欧氏空间 V 的一组基。首先令

$$\beta_1 = \alpha_1, \quad \alpha_1^0 = \frac{1}{|\alpha_1|}$$

α_1^0 是单位向量, 然后欲寻求一个向量 β_2 , 使 $\beta_2 \perp \alpha_1^0$, 故可令

$$\beta_2 = \alpha_2 + k_{21} \alpha_1^0$$

由于 $\beta_2 \perp \alpha_1^0$, 所以

$$(\beta_2, \alpha_1^0) = (\alpha_2, \alpha_1^0) + k_{21} = 0$$

得 $k_{21} = -(\alpha_2, \alpha_1^0)$

从而有 $\beta_2 = \alpha_2 - (\alpha_2, \alpha_1^0) \alpha_1^0, \quad \alpha_2^0 = \frac{\beta_2}{|\beta_2|}$

由于 $\beta_3 \perp \alpha_1^0, \beta_3 \perp \alpha_2^0$, 所以

$$(\beta_3, \alpha_1^0) = (\alpha_3, \alpha_1^0) + k_{31} = 0$$

$$(\beta_3, \alpha_2^0) = (\alpha_3, \alpha_2^0) + k_{32} = 0$$

得 $k_{31} = -(\alpha_3, \alpha_1^0), \quad k_{32} = -(\alpha_3, \alpha_2^0)$

从而有 $\beta_3 = \alpha_3 - (\alpha_3, \alpha_1^0) \alpha_1^0 - (\alpha_3, \alpha_2^0) \alpha_2^0, \quad \alpha_3^0 = \frac{\beta_3}{|\beta_3|}$

到此为止已经做出三个两两正交的单位向量 $\alpha_1^0, \alpha_2^0, \alpha_3^0$ 。继续这样进行下去, 假若已经求出 m 个单位正交向量 $\alpha_1^0, \alpha_2^0, \dots, \alpha_m^0$, 为求第 $m+1$ 个单位正交向量 α_{m+1}^0 , 可设

$$\beta_{m+1} = \alpha_{m+1} + k_{m+11} \alpha_1^0 + k_{m+12} \alpha_2^0 + \dots + k_{m+1m} \alpha_m^0$$

由于 $\beta_{m+1} \perp \alpha_i^0 (i = 1, 2, \dots, m)$, 所以

$$(\beta_{m+1}, \alpha_i^0) = (\alpha_{m+1}, \alpha_i^0) + k_{m+1i} = 0$$

得 $k_{m+1i} = -(\alpha_{m+1}, \alpha_i^0) (i = 1, 2, \dots, m)$

从而有 $\beta_{m+1} = \alpha_{m+1} - (\alpha_{m+1}, \alpha_1^0) \alpha_1^0 - \dots - (\alpha_{m+1}, \alpha_m^0) \alpha_m^0$

$$\alpha_{m+1}^0 = \frac{\beta_{m+1}}{|\beta_{m+1}|}$$

当 $m + 1 = n$ 时, 就得到了欧氏空间中的一个标准正交基 $\alpha_1^0, \dots, \alpha_n^0$ 。

在求 $\alpha_1^0, \dots, \alpha_n^0$ 过程中有两步骤, 先求正交向量组 $\alpha_1, \dots, \alpha_n$, 同时也进行了 $\alpha_1^0 = \frac{\alpha_1}{|\alpha_1|}, \dots, \alpha_n^0 = \frac{\alpha_n}{|\alpha_n|}$ 。前者叫正交化过程, 后者叫标准化过程, 整个过程叫 Gram-Schmidt 方法, 一般教科书是先全部正交化, 然后再全部标准化。这里穿插进行, 是为了后面容易得到矩阵的 QR 分解式。

从定理的证明过程, 可得到

$$\begin{aligned} \alpha_1 &= |\alpha_1| \alpha_1^0 \\ \alpha_2 &= (\alpha_2, \alpha_1^0) \alpha_1^0 + |\alpha_2| \alpha_2^0 \\ &\dots \\ \alpha_n &= (\alpha_n, \alpha_1^0) \alpha_1^0 + (\alpha_n, \alpha_2^0) \alpha_2^0 + \dots + (\alpha_n, \alpha_{n-1}^0) \alpha_{n-1}^0 + |\alpha_n| \alpha_n^0 \end{aligned}$$

用矩阵的记法, 便有

$$(\alpha_1, \alpha_2, \dots, \alpha_n) = (\alpha_1^0, \alpha_2^0, \dots, \alpha_n^0) \begin{pmatrix} |\alpha_1| & (\alpha_2, \alpha_1^0) & \dots & (\alpha_n, \alpha_1^0) \\ 0 & |\alpha_2| & \dots & (\alpha_n, \alpha_2^0) \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & |\alpha_n| \end{pmatrix} \quad (2.4)$$

上式右端的上三角阵, 由于 $|\alpha_i| \neq 0$, 故是一个满秩阵, 记为 R , 它是两个基之间的变换阵(过渡矩阵)。

例 1 把向量组 $\alpha_1 = (1, 1, 0, 0), \alpha_2 = (1, 0, 1, 0), \alpha_3 = (-1, 0, 0, 1), \alpha_4 = (1, -1, -1, 1)$ 正交化、标准化。

解 令 $\alpha_1^0 = \frac{\alpha_1}{|\alpha_1|} = \frac{1}{\sqrt{2}}(1, 1, 0, 0)$

$$\begin{aligned} \alpha_2 &= \alpha_2 - (\alpha_2, \alpha_1^0) \alpha_1^0 = (1, 0, 1, 0) - \frac{1}{2}(1, 1, 0, 0) \\ &= (\frac{1}{2}, -\frac{1}{2}, 1, 0), \quad \alpha_2^0 = \frac{\sqrt{2}}{\sqrt{3}}(\frac{1}{2}, -\frac{1}{2}, 1, 0) \\ \alpha_3 &= \alpha_3 - (\alpha_3, \alpha_1^0) \alpha_1^0 - (\alpha_3, \alpha_2^0) \alpha_2^0 \\ &= (-1, 0, 0, 1) + \frac{1}{2}(1, 1, 0, 0) + \frac{1}{3}(\frac{1}{2}, -\frac{1}{2}, 1, 0) \\ &= (-\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 1), \quad \alpha_3^0 = (-\frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{3}{\sqrt{12}}) \\ \alpha_4 &= \alpha_4 - (\alpha_4, \alpha_1^0) \alpha_1^0 - (\alpha_4, \alpha_2^0) \alpha_2^0 - (\alpha_4, \alpha_3^0) \alpha_3^0 \\ \alpha_4^0 &= \alpha_4, \quad \alpha_4^0 = (\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}) \end{aligned}$$

例 2 $1, x, x^2, x^3$ 是 $R[x]_3$ 中一个基, 在内积 $\int_{-1}^1 f(x)g(x)dx$ 之下把它们正交化、单位化。

解 设 $\alpha_1 = 1, \alpha_2 = x, \alpha_3 = x^2, \alpha_4 = x^3$, 则

$$\begin{aligned} |e_1| &= \left(\int_{-1}^1 dx \right)^{\frac{1}{2}} = \sqrt{2} \quad e_1^0 = \frac{1}{\sqrt{2}} \\ e_2 &= e_2^0 - (e_2^0, e_1^0) e_1^0 = x - \frac{1}{\sqrt{2}} \int_{-1}^1 \frac{1}{\sqrt{2}} x dx = x \\ e_2^0 &= \frac{e_2}{|e_2|} = \frac{x}{\left(\int_{-1}^1 x^2 dx \right)^{\frac{1}{2}}} = \frac{\sqrt{3}}{\sqrt{2}} x \\ e_3 &= e_3^0 - (e_3^0, e_1^0) e_1^0 - (e_3^0, e_2^0) e_2^0 = x^2 - \frac{1}{3} \\ e_3^0 &= \frac{e_3}{|e_3|} = \frac{\sqrt{10}}{4} (3x^2 - 1) \\ e_4 &= e_4^0 - (e_4^0, e_1^0) e_1^0 - (e_4^0, e_2^0) e_2^0 - (e_4^0, e_3^0) e_3^0 = x^3 - \frac{3}{5}x \\ e_4^0 &= \frac{e_4}{|e_4|} = \frac{\sqrt{14}}{4} (5x^3 - 3x) \end{aligned}$$

在 n 维欧氏空间中取定一个标准正交基 e_1, \dots, e_n , 则对任意 $\alpha, \beta \in V$, 在这个标准正交基下的坐标及内积 (α, β) 表示式特别简单。

设 $\alpha = a_1 e_1 + \dots + a_n e_n$, 则

$$(\alpha, e_j) = \left(\sum_{i=1}^n a_i e_i, e_j \right) = a_j$$

因此

$$\alpha = (\alpha, e_1) e_1 + \dots + (\alpha, e_n) e_n \tag{2.5}$$

再设 $\beta = b_1 e_1 + \dots + b_n e_n$, 则

$$\begin{aligned} (\alpha, \beta) &= \left(\sum_{i=1}^n a_i e_i, \sum_{j=1}^n b_j e_j \right) \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i b_j (e_i, e_j) = \sum_{i=1}^n a_i b_i \end{aligned} \tag{2.6}$$

这说明, 对任意的 n 维欧氏空间, 在标准正交基下两向量的内积等于这两个向量在这基下的对应坐标乘积之和!

2.3 正交补与投影定理

2.3.1 正交补

定义 2.6 设 W 是欧氏空间 V 的一个非空子集, α 是 V 中一个向量, 若 $(\alpha, \beta) = 0, \forall \beta \in W$, 都有, 称 α 与子集 W 正交, 记 $\alpha \perp W$ 。对于 V 中两个非空子集 W_1, W_2 , 如果任取 $\alpha \in W_1$, 任取 $\beta \in W_2$, 都有 $(\alpha, \beta) = 0$, 那么称 W_1 与 W_2 互相正交, 记为 $W_1 \perp W_2$ 。关于子集 W , 可做集合

$$W^\perp = \{ \alpha \in V, \alpha \perp W \}$$

下面证明 W^\perp 是 V 的一个子空间。

由于 $0 \in W$, 所以 W 是非空集, 若 $\alpha \in W$, 则对任意 $\beta \in W$ 有

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

即 $\alpha + \beta \in W$, 对任意 $k \in \mathbb{R}$, 有

$$(k\alpha, \alpha) = k(\alpha, \alpha) = k \cdot 0 = 0$$

即 $k\alpha \in W$, 这就证明了 W 是 V 的一个子空间。

定义 2.7 设 W_1, W_2 是欧氏空间 V 的两个子空间, 且满足

- (1) $V = W_1 + W_2$
- (2) $W_1 \perp W_2$

则称 V 有一个正交直和分解, 并把 W_2 叫做 W_1 的正交补, 记之为 $W_2 = W_1^\perp$, 从而有 $V = W_1 + W_1^\perp$ 。

从上述定义的条件 2 很容易推出 $W_1 \cap W_2 = \{0\}$, 所以 $W_1 + W_2$ 是直和, 又因为 $W_1 \perp W_2$, 故叫正交直和。

定理 2.3.1 设 W 是 n 维欧氏空间 V 的任意一个子空间, 则一定存在 W 的正交补 W^\perp , 使 $V = W + W^\perp$ 。

证明 若 $W = \{0\}$, 则 $W^\perp = V$
若 $W = V$, 则 $W^\perp = \{0\}$

设 W 是非平凡子空间, 设维 $W = k, 1 \leq k < n$, 不妨设 $\alpha_1, \dots, \alpha_k$ 是 W 的标准正交基。适当扩充, 可得 V 的标准正交基 $\alpha_1, \dots, \alpha_k, \alpha_{k+1}, \dots, \alpha_n$, 令 $W^\perp = L(\alpha_{k+1}, \dots, \alpha_n)$ 。因为 $W = L(\alpha_1, \dots, \alpha_k)$, 故 " $\alpha \in W$, 有 $\alpha = l_1 \alpha_1 + \dots + l_k \alpha_k$, 对任意 $\beta \in W^\perp$, 有 $\beta = h_{k+1} \alpha_{k+1} + \dots + h_n \alpha_n$ 。所以 $(\alpha, \beta) = 0$, 故 W^\perp 恰为 W 的直交补, 且 $V = W + W^\perp$ 。

由这个定理可以知道, " $\alpha \in V, \alpha = \alpha_W + \alpha_{W^\perp}, \alpha_W \in W, \alpha_{W^\perp} \in W^\perp$ 称 α_W 为 沿着 W 在 W 上的正交投影, α_{W^\perp} 为 沿着 W 在 W^\perp 上的正交投影。

2.3.2 投影定理

定义 2.8 设 W 为欧氏空间 V 中的非空子集, $\alpha \in V$, 若 $\alpha_0 \in W$, 满足下述等式

$$|\alpha - \alpha_0| = \inf_{\alpha \in W} |\alpha - \alpha|$$

称 α_0 是 α 在 W 上的最佳逼近。

定理 2.3.2(最佳逼近特征) 设 W 为欧氏空间 V 的子空间, $\alpha \in V$, 则 $\alpha_0 \in W$ 为 α 在 W 上的最佳逼近的充要条件是

$$\text{若 } \alpha = \alpha_W + \alpha_{W^\perp}, \text{ 有 } \alpha_W = \alpha_0$$

证明 设 α 在 W 上的最佳逼近是 α_0 , 即 $|\alpha - \alpha_0| = \inf_{\alpha \in W} |\alpha - \alpha|$ 。若 $\alpha_0 \neq \alpha_W$, 因为 $\alpha_0 \in W$, 所以 $\alpha_0 - \alpha_W \in W$, 且 $\alpha_0 - \alpha_W \neq 0$, 那么由于 $\alpha - \alpha_W \in W^\perp$, 所以根据勾股定理

$$\begin{aligned} |\alpha - \alpha_0|^2 &= |\alpha - \alpha_W + \alpha_W - \alpha_0|^2 \\ &= |\alpha - \alpha_W|^2 + |\alpha_W - \alpha_0|^2 > |\alpha - \alpha_W|^2 \end{aligned}$$

所以 $|\alpha - \alpha_0| > |\alpha - \alpha_W|$

这矛盾于 α_0 是最佳逼近的定义, 故 $\alpha_0 = \alpha_W$ 。

反过来, 若 $\alpha_0 = \alpha_W$, 那么对任意 $\alpha \in W$, 都有

$$\begin{aligned} \|x - w\|^2 &= \|x - w + w - w\|^2 \\ &= \|x - w\|^2 + \|w - w\|^2 = \|x - w\|^2 \end{aligned}$$

这表明 w 就是 x 在 W 上的最佳逼近。

如何求这个最佳逼近呢？

定理 2.3.3 设 $\alpha_1, \dots, \alpha_s$ 是 W 的任意一组基, $x \in V$, 则向量 $k_1\alpha_1 + \dots + k_s\alpha_s$ 是 x 在 W 上的最佳逼近的充分必要条件是 k_1, \dots, k_s 是线性方程组 $\sum_{i=1}^s (\alpha_j, \alpha_i) x_i = (\alpha_j, x), j = 1, 2, \dots, s$ 的解。

证明 若向量 $k_1\alpha_1 + \dots + k_s\alpha_s$ 是 x 在 W 上的最佳逼近, 那么 $x = w + w^\perp, w = k_1\alpha_1 + \dots + k_s\alpha_s$, 因 $w^\perp \perp W$, 所以

即 $(x - w, \alpha_j) = 0 \quad (j = 1, 2, \dots, s)$

即 $(w, \alpha_j) = (\alpha_j, x) \quad (j = 1, 2, \dots, s)$

即 $\sum_{i=1}^s (\alpha_j, \alpha_i) k_i = (\alpha_j, x) \quad (j = 1, 2, \dots, s)$

反过来, 因为 $\alpha_1, \dots, \alpha_s$ 线性无关, 所以矩阵

$$G(\alpha_1, \dots, \alpha_s) = \begin{pmatrix} (\alpha_1, \alpha_1) & (\alpha_1, \alpha_2) & \dots & (\alpha_1, \alpha_s) \\ (\alpha_2, \alpha_1) & (\alpha_2, \alpha_2) & \dots & (\alpha_2, \alpha_s) \\ \dots & \dots & \dots & \dots \\ (\alpha_s, \alpha_1) & (\alpha_s, \alpha_2) & \dots & (\alpha_s, \alpha_s) \end{pmatrix}$$

是正定矩阵, 则线性方程组

$$\begin{cases} (\alpha_1, \alpha_1) x_1 + \dots + (\alpha_1, \alpha_s) x_s = (\alpha_1, x) \\ (\alpha_2, \alpha_1) x_1 + \dots + (\alpha_2, \alpha_s) x_s = (\alpha_2, x) \\ \dots \\ (\alpha_s, \alpha_1) x_1 + \dots + (\alpha_s, \alpha_s) x_s = (\alpha_s, x) \end{cases}$$

有惟一解 k_1, \dots, k_s , 令 $w_0 = k_1\alpha_1 + \dots + k_s\alpha_s$, 也就有

也即 $(w_0, \alpha_j) = (\alpha_j, x) \quad j = 1, 2, \dots, s$

也即 $(x - w_0, \alpha_j) = 0 \quad j = 1, 2, \dots, s$

这说明 $x - w_0$ 与 W 的基正交, 从而 $x - w_0 \perp W$, 所以 $x - w_0 \in W^\perp$, 于是

$$x = w_0 + (x - w_0) \quad w_0 \in W, \quad x - w_0 \in W^\perp$$

由直和分解的惟一性 $w_0 = w$

这就证得 $w_0 = k_1\alpha_1 + \dots + k_s\alpha_s$ 是 x 在 W 上的最佳逼近。

定理证明过程中的矩阵 $G(\alpha_1, \dots, \alpha_s)$ 称格拉姆(Gram)矩阵。若 $\alpha_1, \dots, \alpha_s$ 是 W 的标准正交基, 则 $G(\alpha_1, \dots, \alpha_s) = E$, 于是要求向量 x 在 W 上的最佳逼近, 只需计算出 $(\alpha_i, x) \quad i = 1, 2, \dots, s$, 令

$$w = (\alpha_1, x)\alpha_1 + \dots + (\alpha_s, x)\alpha_s$$

则成。

读者不要小看这个定理, 它是很多优化问题, 特别是最小二乘问题中常用的一条基本定理, 借助它能解决大量的工程实际问题。

(4) 对任意 V , 有 $(\alpha, \alpha) \geq 0$, 当且仅当 $\alpha = 0$ 时, 实数 $(\alpha, \alpha) = 0$ 。 (α, α) 称为向量 α 与 α 的内积, 称 V 为一个酉空间(或复内积空间), 酉空间与欧氏空间统称为内积空间。

注意: 这里定义的内积与欧氏空间中定义的内积仅条件(1)不同, 原因是这里的 (α, β) 一般是复数。但根据条件(1), (α, α) 就是实数了, 这样条件(4)才有意义。据此, 称

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

为向量 α 的长度。

若 $(\alpha, \beta) = 0$, 还是称 α 与 β 正交, 记 $\alpha \perp \beta$ 。不过若 $(\alpha, \alpha) = 0$, 一般在酉空间中就不再定义 α 与 β 的夹角了(有的书上定义向量 α 与 β 的夹角 Q 为, $\cos^2 Q = \frac{(\alpha, \beta)(\beta, \alpha)}{|\alpha|^2 |\beta|^2}, (\alpha, \alpha) \neq 0, (\beta, \beta) \neq 0$)。

例 1 在 n 维复线性空间 C^n 中, 对任意向量 $\alpha = (a_1, \dots, a_n), \beta = (b_1, \dots, b_n), b_i, a_i \in C$, 规定

$$(\alpha, \beta) = \overline{a_1} b_1 + \overline{a_2} b_2 + \dots + \overline{a_n} b_n$$

不难验证它满足定义 2.8 中的 4 条, 故 (α, β) 是内积, 从而 C^n 是一个酉空间。

由内积定义, 可以得到下列结果。

- (1) $(\alpha, k\beta) = \overline{k}(\alpha, \beta)$
- (2) $(\alpha, \beta + \gamma) = (\alpha, \beta) + (\alpha, \gamma)$
- (3) $(0, \beta) = (\alpha, 0) = 0$
- (4) $(\sum_{i=1}^s k_i \alpha_i, \sum_{j=1}^l l_j \beta_j) = \sum_{i=1}^s \sum_{j=1}^l \overline{k_i} l_j (\alpha_i, \beta_j)$
- (5) Cauchy-Schwarz 不等式

$$|(\alpha, \beta)| \leq \sqrt{(\alpha, \alpha)} \sqrt{(\beta, \beta)}$$

证明性质 1, 2, 3, 4 都很显然, 下面只证性质 5。

当 $\alpha = 0$ 时, 显然结论成立。

现在设 $\alpha \neq 0$, 对任意复数 c , 恒有

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

根据内积性质, 有

$$(\alpha, \alpha) - \overline{c}(\alpha, \beta) - c(\beta, \alpha) + c\overline{c}(\beta, \beta) \geq 0$$

取 $c = \frac{(\alpha, \beta)}{(\beta, \beta)}$, 代入上式, 有

$$(\alpha, \alpha) - \frac{(\alpha, \beta)(\beta, \alpha)}{(\beta, \beta)} \geq 0$$

所以

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

在 n 维酉空间中, 同样可以定义正交基和标准正交基的概念。关于标准正交基, 也有下列性质

- (6) 任意线性无关的向量组可以用 Schmidt 方法正交化;
- (7) 任一非零酉空间都存在正交基和标准正交基;
- (8) 任一 n 维酉空间均可分解为其子空间 V_1 与 V_1^\perp 的直和。

习 题 2

1. 设 V 是实数域 R 上 n 维向量空间, $\alpha_1, \dots, \alpha_n$ 是 V 的一个基, 对任意向量 $\alpha, \beta \in V$,

$$\begin{aligned}\alpha &= x_1 \alpha_1 + x_2 \alpha_2 + \dots + x_n \alpha_n \\ \beta &= y_1 \alpha_1 + y_2 \alpha_2 + \dots + y_n \alpha_n\end{aligned}$$

定义 (α, β) 所对应的实数 (x, y) 为

$$(x, y) = x_1 y_1 + 2x_2 y_2 + \dots + nx_n y_n$$

证明: (α, β) 是 α 与 β 的内积, 从而 V 对这个内积成为一个欧氏空间。

2. 设 $A = (a_{ij})_{n \times n}$ 是实对称正定阵, 对向量空间 R^n 中任意两向量

$$\alpha = (a_1, a_2, \dots, a_n) \quad \beta = (b_1, b_2, \dots, b_n)$$

定义 (α, β) 所对应的实数 (x, y) 为

$$(x, y) = A^{-1}xy^T$$

证明: (α, β) 是 α 与 β 的内积, 从而 R^n 对这个内积成为一个欧氏空间, 并写出在这个欧氏空间 R^n 中的 Cauchy-Schwarz 不等式。

3. 证明: 在欧氏空间 V 中, 对任意向量 $\alpha, \beta \in V$, 恒有

$$\begin{aligned}(1) & \quad |\alpha + \beta|^2 + |\alpha - \beta|^2 = 2|\alpha|^2 + 2|\beta|^2 \\ (2) & \quad (\alpha, \beta) = \frac{1}{4}|\alpha + \beta|^2 - \frac{1}{4}|\alpha - \beta|^2\end{aligned}$$

4. 在欧氏空间 R^4 中找出两个单位向量 α_1, α_2 , 使它们同时与下列向量

$$\begin{aligned}\alpha &= (2, 1, -4, 0) \\ \beta &= (-1, -1, 2, 2) \\ \gamma &= (3, 2, 5, 4)\end{aligned}$$

中每个都正交。

5. 设 $\alpha_1, \dots, \alpha_n$ 是 n 维欧氏空间 V 的一组基, 证明:

- (1) 若 $\alpha \in V$, 且 $(\alpha, \alpha_i) = 0, i = 1, 2, \dots, n$, 则 $\alpha = 0$;
- (2) 若 $\alpha_1, \alpha_2 \in V$, 使对任意 $\alpha \in V$, 有 $(\alpha, \alpha_1) = (\alpha, \alpha_2)$, 则 $\alpha_1 = \alpha_2$ 。

6. 设 $\alpha_1, \dots, \alpha_n$ 是欧氏空间中 n 个向量, 格莱姆行列式

$$G(\alpha_1, \dots, \alpha_n) = \begin{vmatrix} (\alpha_1, \alpha_1) & (\alpha_1, \alpha_2) & \dots & (\alpha_1, \alpha_n) \\ (\alpha_2, \alpha_1) & (\alpha_2, \alpha_2) & \dots & (\alpha_2, \alpha_n) \\ \dots & \dots & \dots & \dots \\ (\alpha_n, \alpha_1) & (\alpha_n, \alpha_2) & \dots & (\alpha_n, \alpha_n) \end{vmatrix}$$

证明: $\alpha_1, \dots, \alpha_n$ 线性相关的必要充分条件是 $G(\alpha_1, \dots, \alpha_n) = 0$

7. 设 $\alpha_1, \alpha_2, \alpha_3$ 是欧氏空间 R^3 的一组标准正交基, 证明

$$\begin{aligned}\alpha_1 &= \frac{1}{3}(2\alpha_1 + 2\alpha_2 - \alpha_3) \\ \alpha_2 &= \frac{1}{3}(2\alpha_1 - 2\alpha_2 + 2\alpha_3)\end{aligned}$$

$$\alpha_3 = \frac{1}{3}(\alpha_1 - 2\alpha_2 - 2\alpha_3)$$

也是一组标准正交基。

8. 设 $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ 是五维欧氏空间 V 的一组标准正交基, $V_1 = L(\alpha_1, \alpha_2, \alpha_3)$, 其中 $\beta_1 = \alpha_1 + \alpha_5, \beta_2 = \alpha_1 - \alpha_2 + \alpha_4, \beta_3 = 2\alpha_1 + \alpha_2 + \alpha_3$, 求 V_1 的一组标准正交基。

9. 已知 $\alpha_1 = (0, 2, 1, 0), \alpha_2 = (1, -1, 0, 0), \alpha_3 = (1, 2, 0, -1), \alpha_4 = (1, 0, 0, 1)$ 是 R^4 的一个基, 对这个基用 Gram-Schmidt 方法求一标准正交基。

10. 设 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 n 维欧氏空间 V 的一标准正交基, $\beta_1, \dots, \beta_n \in V$, 且

$$\beta_i = a_{i1}\alpha_1 + a_{i2}\alpha_2 + \dots + a_{in}\alpha_n \quad i = 1, 2, \dots, n$$

证明: $\beta_1, \beta_2, \dots, \beta_n$ 的 Gram 行列式

$$G(\beta_1, \beta_2, \dots, \beta_n) = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}^2$$

11. 设 $\alpha_1, \dots, \alpha_n$ 是 n 维欧氏空间的一个基, V 中任意向量

$$\beta = x_1\alpha_1 + x_2\alpha_2 + \dots + x_n\alpha_n$$

的坐标 $x_i = |\beta| \cos Q_i$, Q_i 表示 β 与 α_i 的夹角, 证明: $\alpha_1, \dots, \alpha_n$ 是 V 的一个标准正交基。

12. 证明: 对任意实数 a_1, a_2, \dots, a_n , 恒有不等式

$$\sum_{i=1}^n |a_i| \leq \sqrt{n(a_1^2 + a_2^2 + \dots + a_n^2)}$$

成立。

13. 设 $\alpha_1, \dots, \alpha_m$ 是 n 维欧氏空间 V 的一个标准正交向量组, 证明: 对任意向量 $\beta \in V$, 恒有不等式

$$\sum_{i=1}^m (\beta, \alpha_i)^2 \leq |\beta|^2$$

成立。

14. 设 V 是 n 维欧氏空间, α_0 是 V 中固定向量, 证明:

- (1) $W = \{ \alpha \in V, (\alpha, \alpha_0) = 0 \}$ 是 V 的一个子空间;
- (2) $\dim W = n - 1$ 。

第 3 章

线性变换

线性变换, 有时又叫线性映射或线性算子, 它是研究线性空间之间联系的一种最简单同时也是最重要的变换, 它是矩阵论中最有用、最重要内容之一, 在这一章中将对线性变换作一般性介绍, 主要借助于矩阵来表示并刻划它们的各种性质。

3.1 线性变换定义

定义 3.1 设 A 是数域 K 上线性空间 V 到 V 的一个变换, 如果满足

- (1) $A(\alpha + \beta) = A(\alpha) + A(\beta)$, $\alpha, \beta \in V$
- (2) $A(k\alpha) = kA(\alpha)$ $k \in K, \alpha \in V$

则称 A 是 V 上的一个线性变换(或称线性算子)。

根据上面定义, 也可以定义

$$A(k\alpha + l\beta) = kA(\alpha) + lA(\beta), \quad \alpha, \beta \in V, k, l \in K$$

作为 A 是 V 的一个线性变换的充分必要条件。

例 1 设 V 是任意一个数域 K 上的线性空间, 定义如下映射

$$E: V \rightarrow V$$

易验证 E 是 V 上的一个线性变换, 以后称之为恒等变换;

$$O: V \rightarrow V$$
$$0$$

易验证 O 是 V 上的一个线性变换, 以后称之为零变换;

$$K: V \rightarrow V$$
$$k\alpha = k\alpha, \quad k \in K, k \text{ 为某固定数}$$

易验证 K 是 V 上的一个线性变换, 以后称之为数乘变换。

例 2 设 $R[x]$ 是所有实系数多项式构成的线性空间, 定义

$$D: R[x] \rightarrow R[x]$$

$$f(x) = f'(x)$$

易验证 D 是 $R[x]$ 上的一个线性变换。

例 3 给定 $A = (a_{ij})_{n \times n}$, $a_{ij} \in R$, 定义

$$A : R^n \rightarrow R^n$$
$$A$$

易验证 A 是 R^n 上一个线性变换。

例 4 $R[a, b]$ 表示 $[a, b]$ 上实连续函数构成的线性空间, 定义

$$T : R[a, b] \rightarrow R[a, b]$$
$$f(x) \mapsto \int_0^x f(t) dt$$

易证明 T 是 $R[a, b]$ 上的一个线性变换。

下面介绍线性变换的几个简单性质(设 T 是线性空间 V 上的线性变换)

(1) $T(0) = 0$ $T(-\alpha) = -T(\alpha)$

因为 $T(0) = T(0 \cdot \alpha) = 0 \cdot T(\alpha) = 0$,

$$T(-\alpha) = T((-1) \cdot \alpha) = -1 \cdot T(\alpha) = -T(\alpha)$$

(2) $T(\sum_{i=1}^s k_i \alpha_i) = \sum_{i=1}^s k_i T(\alpha_i)$

这从定义很容易得知, 从这两条性质可知线性变换把线性相关的向量还是变为线性相关的向量。值得注意的是线性无关的向量在线性变换之下的像未必线性无关, 例如零变换。

下面介绍线性变换的值域与核。

定义 3.2 设 T 是线性空间 V 上的线性变换, 集合

$$R(T) = \{T(\alpha) \mid \alpha \in V\}$$

称之为 T 的值域, 也就是 V 在 T 之下的像的全体集合。集合

$$N(T) = \{\alpha \in V \mid T(\alpha) = 0\}$$

称之为 T 的核域, 也就是像为零元的那些源像的全体集合。

定理 3.1.1 T 是线性空间 V 的一个线性变换, 则 $R(T)$ 和 $N(T)$ 都是 V 的线性子空间。

证明 对任意 $\alpha, \beta \in R(T)$, 则存在 $\gamma, \delta \in V$, 使 $T(\gamma) = \alpha$, $T(\delta) = \beta$, 于是

$$\alpha + \beta = T(\gamma) + T(\delta) = T(\gamma + \delta) \in R(T)。$$

$$k\alpha = kT(\gamma) = T(k\gamma) \in R(T),$$

故 $R(T)$ 是子空间。

对任意 $\alpha \in N(T)$, 则 $T(\alpha) = 0$, $T(k\alpha) = 0$, 于是 $T(\alpha + \beta) = T(\alpha) + T(\beta) = 0$, 所以 $\alpha + \beta \in N(T)$ 且 $T(k\alpha) = kT(\alpha) = k \cdot 0 = 0$, 所以 $k\alpha \in N(T)$, 故 $N(T)$ 是子空间, 证毕。

定理 3.1.2 若 V 是 n 维线性空间, T 是 V 上的一个线性变换, 则

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

证明, 设 $\alpha_1, \dots, \alpha_r$ 是 $N(T)$ 的基, 当然 $T\alpha_i = 0, i = 1, 2, \dots, r$ 扩 $\alpha_1, \dots, \alpha_r, \alpha_{r+1}, \dots, \alpha_n$ 作为 V 的一组基, 下面证明 $T\alpha_{r+1}, \dots, T\alpha_n$ 就是 $R(T)$ 的一个基。因为任取 $\alpha \in V$, 有

$$\alpha = k_1\alpha_1 + \dots + k_r\alpha_r + k_{r+1}\alpha_{r+1} + \dots + k_n\alpha_n$$
$$T(\alpha) = k_{r+1}T\alpha_{r+1} + \dots + k_nT\alpha_n$$

所以

$$R(T) = l(T_{r+1}, \dots, T_n)$$

设
$$l_{r+1}T_{r+1} + \dots + l_nT_n = 0$$

即
$$T(l_{r+1}T_{r+1} + \dots + l_nT_n) = 0$$

这说明 $l_{r+1}T_{r+1} + \dots + l_nT_n \in N(T)$, 故存在数 l_1, \dots, l_r , 使

$$l_{r+1}T_{r+1} + \dots + l_nT_n = l_1T_1 + \dots + l_rT_r$$

即
$$l_1T_1 + \dots + l_rT_r - l_{r+1}T_{r+1} - \dots - l_nT_n = 0$$

由于 T_1, \dots, T_n 线性无关, 所以

$$l_1 = \dots = l_n = 0$$

于是 T_{r+1}, \dots, T_n 线性无关。

例如在 $R[x]_n$ 中, 微分变换 D 的值域 $R(D) = R[x]_{n-1}$, 而 $N(D) = R$, 当然 $\dim R(D) + \dim N(D) = n + 1$ 。

称 $\dim R(D)$ 为线性变换 D 的秩, $\dim N(D)$ 为线性变换 D 的零度。

例 5 设 $A : R^4 \rightarrow R^4$
 $A = (a_1, a_2, a_3, a_4) \quad A(\alpha) = (a_1 + a_2 - 3a_3 - a_4, 3a_1 - a_2 - 3a_3 + 4a_4, 0, 0)$

A 是 R^4 上的一个线性变换, 求 A 的秩及零度。

解 A 的像子空间 $R(A)$ 显然是由形如 $(\alpha_1, \alpha_2, 0, 0)$ 的向量构成的子空间, 故 $\dim R(A) = 2$, A 的核子空间 $N(A)$ 是线性方程组

$$\begin{cases} \alpha_1 + \alpha_2 - 3\alpha_3 - \alpha_4 = 0 \\ 3\alpha_1 - \alpha_2 - 3\alpha_3 + 4\alpha_4 = 0 \end{cases}$$

的解空间, 即形如 $k(6, 6, 4, 0) + l(-3, 7, 0, 4)$ 的元素构成的子空间, 当然 $\dim N(A) = 2, \dim R(A) = 2$ 。

定义 3.3 设 A_1, A_2 都线性空间 V 上的线性变换, 若对任意 $\alpha \in V$, 都有 $A_1\alpha = A_2\alpha$, 称 A_1 与 A_2 相等, 记 $A_1 = A_2$ 。

定义 3.4 设 A, B 都是线性空间 V 上的线性变换, 定义 A 和 B 的和

$$(A + B)(\alpha) = A(\alpha) + B(\alpha) \quad \alpha \in V$$

易证 $A + B$ 仍是 V 上的线性变换, 因为对任意 $\alpha \in V, k, l \in K$ 。

$$\begin{aligned} (A + B)(k\alpha + l\beta) &= A(k\alpha + l\beta) + B(k\alpha + l\beta) \\ &= kA\alpha + lA\beta + kB\alpha + kB\beta \\ &= k(A + B)(\alpha) + l(A + B)(\beta) \end{aligned}$$

定义 3.5 设 A 是线性空间 V 上的线性变换, 定义数 k 乘以线性变换 A

$$(kA)(\alpha) = kA(\alpha) \quad \alpha \in V$$

易证数 k 乘以线性变换 kA 也是线性变换。

让 $\text{Hom}(V, V)$ 表示线性空间 V 上的全体线性变换的集合, 容易验证集合 $\text{Hom}(V, V)$ 关于上述定义加法和数乘作成线性空间(请读者自行验证上述的加法与数乘是满足 8 条规则的)。

定义 3.6 设 A, B 都是线性空间 V 上的线性变换, 定义 A 与 B 的积

$$(AB)(\alpha) = A[B(\alpha)] \quad \alpha \in V$$

即 AB 是先施行 B , 再施行 A 的变换, 并且 AB 也是线性变换, 因为对任意 $\alpha \in V, k, l \in K$ 。

$$\begin{aligned} (A+B)(k+1) &= A[B(k+1)] \\ &= A[kB + 1B] \\ &= kA[B] + 1A[B] \\ &= k(A+B) + 1(A+B) \end{aligned}$$

这就证明了 $A+B$ 也是线性变换, 不仅如此, 线性变换的乘法还有

$$\begin{aligned} (AB)C &= A(BC) \\ A(B+C) &= AB + AC \\ (B+C)A &= BA + CA \end{aligned}$$

线性变换的乘法与矩阵乘法一样, 不满足交换律。例如 A 表示在 R^2 (平面) 中把向量逆时针旋转 $\frac{\pi}{2}$ 角, B 表示把向量向 x 轴投影。则 $AB(1,0) = (0,1)$, $BA(1,0) = (0,0)$ 说明 $AB \neq BA$ 。但是对于恒等变换 E , 任何的线性变换 A , 都有

$$AE(x) = A(x) = EA(x)$$

可见恒等变换就像单位阵在乘法中的作用一样。如果线性变换 A 是一一变换, 那么存在 A 的逆变换 A^{-1}

$$\begin{aligned} A^{-1}: V &\rightarrow V \\ A^{-1}(y) &= x, \text{ 若 } A(x) = y \end{aligned}$$

于是 $AA^{-1}(x) = x$, $A^{-1}A(x) = x$, 故 $AA^{-1} = E$ 。可以证明逆变换 A^{-1} 也是线性变换。

$$\begin{aligned} \text{记 } T^n &= \underbrace{T \cdot T \cdots T}_n, \text{ 若 } T \text{ 可逆, } T^{-n} = \underbrace{T^{-1} \cdots T^{-1}}_n \\ T^0 &= E \end{aligned}$$

因此可以建立线性变换的指数法则如下

$$\begin{aligned} T^{m+n} &= T^m \cdot T^n, (T^m)^n = T^{mn} \\ T^{-n} &= (T^{-1})^n \end{aligned}$$

设多项式 $f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$, $a_i \in K$, T 是数域 K 上线性空间 V 的线性变换, 则由线性变换运算可知

$$f(T) = a_n T^n + a_{n-1} T^{n-1} + \dots + a_1 T + a_0 E$$

也是 V 的一个线性变换, 特别地, 若

$$\begin{aligned} \text{那么 } h(x) &= f(x)g(x), \quad q(x) = f(x) + g(x) \\ h(T) &= f(T)g(T) = g(T)f(T) \\ q(T) &= f(T) + g(T) \end{aligned}$$

3.2 线性变换的矩阵表示

已经把线性空间的向量用坐标表示出来, 这里还要把线性变换用矩阵表示出来, 使线性变换与数发生联系, 从而可把比较抽象的线性变换用具体的矩阵来处理。

事实上, 给定一个线性变换, 并不需要定义出 V 中的每个元的像是什么, 而只要定义出基

元的像, 就可以得到每一个元的像。

设 $\alpha_1, \dots, \alpha_n$ 是线性空间 V 的一个基, V 上的线性变换 A 在这基之下的像 $A\alpha_1, \dots, A\alpha_n$ 它们在这组基下有线性表达式

$$\begin{aligned} A\alpha_1 &= a_{11}\alpha_1 + a_{21}\alpha_2 + \dots + a_{n1}\alpha_n \\ A\alpha_2 &= a_{12}\alpha_1 + a_{22}\alpha_2 + \dots + a_{n2}\alpha_n \\ &\vdots \\ A\alpha_n &= a_{1n}\alpha_1 + a_{2n}\alpha_2 + \dots + a_{nn}\alpha_n \end{aligned} \tag{3.1}$$

采用矩阵形式运算规则, (3.1) 可表示为

$$(A\alpha_1, \dots, A\alpha_n) = (\alpha_1, \alpha_2, \dots, \alpha_n) \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \tag{3.2}$$

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \tag{3.3}$$

可简记为 $A(\alpha_1, \dots, \alpha_n) = (\alpha_1, \dots, \alpha_n)A$, 矩阵 A 的第 i 列是 $A\alpha_i$ 在基 $\alpha_1, \dots, \alpha_n$ 之下的坐标。

定义 3.7 (3.2) 式中的矩阵 A 称为线性变换 A 在基 $\alpha_1, \dots, \alpha_n$ 之下的矩阵表示。
(3.3) 式说明, 线性空间 V 上的一个线性变换 A 在给定的基下可以惟一确定一个矩阵 A , 反过来, 可以证明, 任意给定一个矩阵 A , 可以惟一确定一个线性变换 A , 使它们满足(3.3)式。这也是线性变换的可随意构造性。不过注意的是, 给定一个线性变换 A , 它们在不同的基之下矩阵表示一般是不相同的, 后面要讨论。

例 1 设 $\alpha_1, \dots, \alpha_m$ 是 $n(n > m)$ 维线性空间 V 的子空间 W 的基, 把 $\alpha_1, \dots, \alpha_m$ 扩为 V 的基 $\alpha_1, \dots, \alpha_m, \alpha_{m+1}, \dots, \alpha_n$, 定义变换 A

$$\begin{aligned} A(\alpha_i) &= \alpha_i \quad i = 1, 2, \dots, m \\ A(\alpha_j) &= 0 \quad j = m + 1, \dots, n \end{aligned}$$

易证 A 是 V 上的线性变换, 称 A 是对子空间 W 的一个投影, 因为对任意 $\alpha \in V, \alpha = k_1\alpha_1 + \dots + k_m\alpha_m + k_{m+1}\alpha_{m+1} + \dots + k_n\alpha_n$

$$A(\alpha) = k_1\alpha_1 + \dots + k_m\alpha_m \in W$$

求出 A 在基 $\alpha_1, \dots, \alpha_n$ 之下的矩阵表示 A

解 因为

$$(A\alpha_1, \dots, A\alpha_n) = (\alpha_1, \dots, \alpha_n) \begin{pmatrix} 1 & \dots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 1 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & 0 \end{pmatrix}$$

所以矩阵表示 $A = \begin{bmatrix} E_m & 0 \\ 0 & 0 \end{bmatrix}$

例 2 在 R^3 中取基 $\alpha_1 = (1, 0, 0)$, $\alpha_2 = (0, 1, 0)$, $\alpha_3 = (0, 0, 1)$, 设 A 是绕 Ox 轴由 Oy 轴向 Oz 轴旋转 90° 的线性变换。

- (1) 求 A 在基 $\alpha_1, \alpha_2, \alpha_3$ 之下的矩阵 A ;
- (2) 求 A 在基 $\alpha_3, \alpha_1, \alpha_2$ 之下的矩阵 B ;
- (3) 求 A 在基 $\alpha_1, 2\alpha_2, \alpha_2 + \alpha_3$ 下的矩阵 C 。

解 因为 $A\alpha_1 = \alpha_1$ $A\alpha_2 = \alpha_3$ $A\alpha_3 = -\alpha_2$
所以

$$A(\alpha_1, \alpha_2, \alpha_3) = (\alpha_1, \alpha_2, \alpha_3) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$
$$A(\alpha_3, \alpha_1, \alpha_2) = (\alpha_3, \alpha_1, \alpha_2) \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}$$
$$A(\alpha_1, 2\alpha_2, \alpha_2 + \alpha_3) = (\alpha_1, 2\alpha_2, \alpha_2 + \alpha_3) \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & -1 \\ 0 & 2 & 1 \end{pmatrix}$$

故

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

例 3 求线性变换 $D: R[x]_n \rightarrow R[x]_n$
 $f(x) \mapsto f'(x)$

关于基底 $1, x, x^2, \dots, x^n$ 的矩阵表示 D 。

解 因为 $D(1) = 0, D(x) = 1, \dots, D(x^n) = nx^{n-1}$, 所以

$$D(1 \ x \ x^2 \ \dots \ x^n) = (1 \ x \ x^2 \ \dots \ x^n) \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 2 & & \dots \\ \dots & \dots & \dots & & n \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

故矩阵表示 $D = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 2 & & \dots \\ \dots & \dots & \dots & & n \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}$

很显然, 任意 n 维线性空间 V 上的恒等变换 E , 数乘变换 K , 零变换 0 在 V 中任意基 $\alpha_1, \dots, \alpha_n$ 之下的矩阵表示不变, 分别是单位矩阵 E , 数量矩阵 kE , 零矩阵 0 , 所以也把恒等变换称为单位变换。

命题 设 A 是数域 K 上 n 维线性空间 V 的一个线性变换, $\alpha_1, \dots, \alpha_n$ 是 V 的一个基底, 且

$$A(\alpha_1, \dots, \alpha_n) = (\alpha_1, \dots, \alpha_n)A$$

则(1) $R(A)=L(A_1, \dots, A_n)$
(2) A 的秩 $= A$ 的秩
(3) $N(A)=\{x=k_1x_1+\dots+k_nx_n \mid A\begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix}=0\}$

证明(1) 因对任意 $x \in R(A)$, 存在 V , 使 $Ax =$
因 $Ax = k_1x_1 + \dots + k_nx_n$
所以 $Ax = k_1A_1 + \dots + k_nA_n$
故 $Ax \in L(A_1, \dots, A_n)$
所以 $R(A) \subseteq L(A_1, \dots, A_n)$
反过来, 设 $x \in L(A_1, \dots, A_n)$, 所以
 $x = k_1A_1 + \dots + k_nA_n$
 $x = A(k_1x_1 + \dots + k_nx_n) \in R(A)$
故 $R(A) = L(A_1, \dots, A_n)$

(2) 因为 $R(A) = L(A_1, \dots, A_n)$
所以 $\dim R(A) =$ 向量组 A_1, \dots, A_n 的秩
又 $A(x_1, \dots, x_n) = (x_1, \dots, x_n)A$
由习题 1 的第 14 题结论 $\dim R(A) = A$ 的秩

(3) 若 $x = k_1x_1 + \dots + k_nx_n \in N(A)$
那么 $Ax = A(x_1, \dots, x_n)\begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix} = (x_1, \dots, x_n)A\begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix} = 0$
由坐标的惟一性, 所以 $A\begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix} = 0$

也即 A 的核空间 $N(A)$ 中的元 x 在 x_1, \dots, x_n 之下的坐标恰是线性方程组 $Ax = 0$ 的解。
证毕。

定理 3.2.1 设 x_1, \dots, x_n 是数域 K 上 n 维线性空间 V 的一个基, 线性变换 A, B 在该基下按(3.3)式依次用矩阵表示 A, B 。则有

(1) $(A+B)(x_1, \dots, x_n) = (x_1, \dots, x_n)(A+B)$
(2) $(kA)(x_1, \dots, x_n) = (x_1, \dots, x_n)(kA)$
(3) $(AB)(x_1, \dots, x_n) = (x_1, \dots, x_n)(AB)$
(4)若 A 可逆, 则 $A^{-1}(x_1, \dots, x_n) = (x_1, \dots, x_n)A^{-1}$ 。

证明 因为 $A(x_1, \dots, x_n) = (x_1, \dots, x_n)A$
 $B(x_1, \dots, x_n) = (x_1, \dots, x_n)B$
所以 (1) $(A+B)(x_1, \dots, x_n) = ((A+B)x_1, \dots, (A+B)x_n)$
 $= (Ax_1+Bx_1, \dots, Ax_n+Bx_n)$
 $= (Ax_1, \dots, Ax_n) + (Bx_1, \dots, Bx_n)$

$$\begin{aligned} &= (\alpha_1, \dots, \alpha_n) A + (\alpha_1, \dots, \alpha_n) B \\ &= (\alpha_1, \dots, \alpha_n) (A + B) \end{aligned}$$

(2)、(3) 留给读者证明。

(4) 若 A 可逆, A^{-1} 是 A 的逆变换, 设

$$A^{-1}(\alpha_1, \dots, \alpha_n) = (\beta_1, \dots, \beta_n) B$$

于是

$$\begin{aligned} A^{-1} A (\alpha_1, \dots, \alpha_n) &= A^{-1} (\beta_1, \dots, \beta_n) A \\ &= (\beta_1, \dots, \beta_n) B A \end{aligned}$$

由于 $A^{-1} A = E$, 而 E 在 $\alpha_1, \dots, \alpha_n$ 之下矩阵表示为 E , 由于满足 (3.3) 式的 A 与 A 是互相惟一确定的, 故 $BA = E$, 所以 $B = A^{-1}$ 。

定义 3.8 若映射 φ 是数域 K 上线性空间 V 到 V 的一一映射, 且 φ 还满足

- (1) $\varphi(\alpha + \beta) = \varphi(\alpha) + \varphi(\beta)$, $\alpha, \beta \in V$
- (2) $\varphi(k \cdot \alpha) = k \varphi(\alpha)$, $\alpha \in V, k \in K$

那么称 φ 是同构映射(同胚映射), 称 V 与 V 是同构的, 记 $V \cong V$ 。

定理 3.2.2 设 V 是数域 K 上的 n 维线性空间。 $\alpha_1, \dots, \alpha_n$ 是 V 中的一组基。

定义 $\varphi: \text{Hom}(V, V) \rightarrow K^{n \times n}$

$$\varphi(A) = (A) = A, \text{ 其中 } A(\alpha_1, \dots, \alpha_n) = (\beta_1, \dots, \beta_n) A$$

则 φ 是一个同构映射, 从而 $\text{Hom}(V, V) \cong K^{n \times n}$ 。

证明 这由 (3.3) 式的互相惟一确定性定理 3.2.1 是很显然的。

定理 3.2.2 深刻划了 $\text{Hom}(V, V)$ 与 $K^{n \times n}$ 的内在本质是完全相同的, 因此完全有理由用矩阵的性质来代替线性变换的性质。

不过, 关于上面的讨论, 都是假定在固定了线性空间 V 的一组基 $\alpha_1, \dots, \alpha_n$ 之下完成的, 但是 V 中的基并不惟一, 一个线性变换 A 在不同基下的矩阵表示一般是不同的, 那么这些不同基下的矩阵表示之间有什么关系? 我们能否找到 V 的一组基, 使得某线性变换在这组基下的矩阵表示最简单?

3.3 线性变换的最简矩阵表示——相似形理论

3.3.1 一般数域上矩阵相似最简形

这一节为的是解决上一节提出的两个问题, 内容较多, 有的定理就只引述, 不证明。

定义 3.9 设 A, B 皆为 n 级方阵, 若存在 n 级可逆阵 C , 使 $C^{-1}AC = B$, 称 A 与 B 相似, 记 $A \sim B$ 。

定理 3.3.1 n 维线性空间 V 上的线性变换 A 在不同的基下的矩阵表示是相似的。反过来, 相似矩阵可以看做某一线性变换在不同基下的矩阵表示。

证明: 设 $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n$ 分别是 V 的两个基底, 它们之间有关系

$$(\beta_1, \dots, \beta_n) = (\alpha_1, \dots, \alpha_n) C \tag{3.4}$$

设线性变换 A 在这两组基下的矩阵表示分别为 A, B 。即

$$A(\alpha_1, \dots, \alpha_n) = (\alpha_1, \dots, \alpha_n)A \tag{3.5}$$

$$A(\beta_1, \dots, \beta_n) = (\beta_1, \dots, \beta_n)B \tag{3.6}$$

把(3.4)式代入(3.6)式, 得

$$A(\alpha_1, \dots, \alpha_n)C = (\alpha_1, \dots, \alpha_n)CB$$

两边右乘以 C^{-1} 。得

$$A(\alpha_1, \dots, \alpha_n) = (\alpha_1, \dots, \alpha_n)CBC^{-1} \tag{3.7}$$

对比(3.5)式与(3.7)式, 由矩阵表示的惟一性, 所以

$$A = CBC^{-1}$$

反过来, 若 $B = C^{-1}AC$, 任选定 $\alpha_1, \dots, \alpha_n$ 是 V 的一基, 再令

$$(\beta_1, \dots, \beta_n) = (\alpha_1, \dots, \alpha_n)C, \text{ 且 } A(\alpha_1, \dots, \alpha_n) = (\alpha_1, \dots, \alpha_n)A$$

那么由前面的证明过程知道, A 在新基 β_1, \dots, β_n 之下矩阵表示为 B 。

先讨论一下相似矩阵的性质, 以后要用。

- (1) 反射性 $A \sim A$;
- (2) 对称性 $A \sim B$ 则 $B \sim A$;
- (3) 传递性 $A \sim B, B \sim C$ 则 $A \sim C$;
- (4) 可加性 若 $B_1 = C^{-1}A_1C, B_2 = C^{-1}A_2C$
则 $B_1 + B_2 \sim A_1 + A_2$
- (5) 可乘性 若 $B_1 = C^{-1}A_1C, B_2 = C^{-1}A_2C$
则 $B_1 B_2 \sim A_1 A_2$
- (6) 数乘性 若 $A \sim B$ 则 $kA \sim kB$
- (7) 若 $A \sim B, f(x) \in K[x]$ 则
 $f(A) \sim f(B)$

以上性质用相似定义来证明很容易, 读者可自己验证。

例 1 设 V 是数域 K 上二维线性空间, $\alpha_1, \alpha_2, \beta_1, \beta_2$ 分别是 V 的两组基, A 是 V 上的线性变换, 且

$$(\alpha_1, \alpha_2) = (\beta_1, \beta_2) \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}, A(\alpha_1, \alpha_2) = (\alpha_1, \alpha_2) \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}$$

- (1) 求 A 在 β_1, β_2 之下的矩阵表示 B ;
- (2) 求 B^k ;
- (3) 求 A^k 。

解 (1) 设 $A = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}, C = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$

因为 $B = C^{-1}AC$

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

所以

$$B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

(2) $B^k = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^k = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$

(3) $A^k = (CBC^{-1})^k = CB^kC^{-1}$
 $= \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1+k & k \\ -k & 1-k \end{bmatrix}$

定理 3.3.1 回答了上一节提出的第一个问题, 并且第二个问题转化为如何寻求矩阵 A 的最简单相似矩阵 B, 又怎样判断 B 为最简?

最简单的矩阵零阵、单位矩阵、数量矩阵分别为特殊变换零变换、单位变换、数乘变换的矩阵表示, 那么一般的线性变换可能相似的最简矩阵只能是对角矩阵。下面讨论, 哪些线性变换能用对角矩阵作为其矩阵表示。

设 $\alpha_1, \dots, \alpha_n$ 是 V 的一组给定的基, A 是 V 的线性变换

$A(\alpha_1, \dots, \alpha_n) = (\alpha_1, \dots, \alpha_n)A$

先看若 A 有对角矩阵 Λ 作为其矩阵表示, 应满足什么必要条件。即存在基 $\alpha_1, \dots, \alpha_n$, 使

$$A(\alpha_1, \dots, \alpha_n) = (\alpha_1, \dots, \alpha_n) \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_n \end{pmatrix}$$

即 $A\alpha_i = \lambda_i \alpha_i \quad i = 1, 2, \dots, n \tag{3.8}$

就是说 A 把基 $\alpha_1, \dots, \alpha_n$ 分别变为与自己平行的向量。

那么, 存在可逆阵 P , 使 $P^{-1}AP = \Lambda = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_n \end{pmatrix}$

即 $A(\alpha_1 \ \alpha_2 \ \dots \ \alpha_n) = (\alpha_1 \ \alpha_2 \ \dots \ \alpha_n) \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_n \end{pmatrix}$

其中 $P = (\alpha_1, \alpha_2, \dots, \alpha_n)$, $\alpha_1, \dots, \alpha_n$ 线性无关。

也即 $A\alpha_i = \lambda_i \alpha_i \quad i = 1, 2, \dots, n \tag{3.9}$

由(3.9) 式 $(A - \lambda_i E)\alpha_i = 0 \quad i = 1, 2, \dots, n \tag{3.10}$

说明线性方程组 $(A - \lambda_i E)x = 0$ 有非 0 解 α_i , 当然 $|A - \lambda_i E| = 0 \tag{3.11}$

反过来, 再看充分条件, 对矩阵表示 A , 若存在 n 个线性无关的列向量 $\alpha_1, \dots, \alpha_n$ 及数 $\lambda_1, \lambda_2, \dots, \lambda_n$, 使

$A\alpha_i = \lambda_i \alpha_i \quad i = 1, 2, \dots, n$
成立。那么就有 $P^{-1}AP = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_n \end{pmatrix}$, 其中 $P = (\alpha_1, \dots, \alpha_n)$

于是可构造新基 $(\alpha_1, \dots, \alpha_n) = (\beta_1, \dots, \beta_n)P$

由定理 3.3.1 知道 $A(\alpha_1, \dots, \alpha_n) = (\alpha_1, \dots, \alpha_n) \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_n \end{pmatrix}$

在上面论述中的(3.8)、(3.9)、(3.10)、(3.11) 式都很重要。

定义 3.10 设 A 是数域 K 上 n 维线性空间 V 的线性变换, 若对于 $\lambda_0 \in K$, 存在非 0 向量 α_0 , 使得

$$A\alpha_0 = \lambda_0\alpha_0 \tag{3.12}$$

称 λ_0 是 A 的特征值, α_0 是 A 的关于 λ_0 的特征向量。

定义 3.11 设 $A \in K^{n \times n}$, 若对于 $\lambda_0 \in K$, 存在非 0 向量 $\alpha_0 \in K^n$ 使

$$A\alpha_0 = \lambda_0\alpha_0 \tag{3.13}$$

称 λ_0 是 A 的特征值, α_0 是 A 的 λ_0 对应的特征向量。

由定义 3.11 及前一页的推导不难证明下述定理

定理 3.3.2 n 级方阵 A 能相似于对角阵 Λ 的必要充分条件是 A 有 n 个线性无关的特征向量 $\alpha_1, \dots, \alpha_n$ 。

读者自证。

定理 3.3.3 线性变换 A 能有对角矩阵表示 $= \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_n \end{pmatrix}$ 的必要充分条件是

A 有 n 个线性无关的特征向量 $\alpha_1, \dots, \alpha_n$ 。

由前一页的推导可知, 若 $A(\alpha_1, \dots, \alpha_n) = (\alpha_1, \dots, \alpha_n) \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_n \end{pmatrix}$,

且 $P^{-1}AP = \Lambda, P = (\alpha_1, \alpha_2, \dots, \alpha_n)$
则 $\alpha_i = (\alpha_1, \dots, \alpha_n)^{-1} \alpha_i \quad i = 1, 2, \dots, n$

定义 3.12 设 $A \in K^{n \times n}$, λ 是一个文字。矩阵 $E - \lambda A$ 称为 A 的特征矩阵; 方程 $|E - \lambda A| = 0$ 称为 A 的特征方程, 多项式 $|E - \lambda A|$ 叫 A 的特征多项式, 记 $f_A(\lambda)$ 。即 $f_A(\lambda) = |E - \lambda A|$, 事实上 $f_A(\lambda)$ 的根就是 A 的特征值或称特征根。

n 级方阵 A 在复数域 C 上一定有 n 个特征根(重根按重数计), 这是代数基本定理的结论, 不证明。但在一般数域 K 上, 根的情况较复杂。

- 下面给出求线性变换 A 的特征值与特征向量的方法步骤:
- (1) 任选线性空间 V 的一组基 $\alpha_1, \dots, \alpha_n$, 求出 A 在 $\alpha_1, \dots, \alpha_n$ 下矩阵表示 A ;
 - (2) 求出特征多项式 $f_A(\lambda) = |E - \lambda A|$ 在数域 K 中的全部特征根 $\lambda_1, \dots, \lambda_k$; 当然它们也是 A 的特征值;
 - (3) 对每一个特征根 λ_i , 求解线性方程组

(λ_iE - A) X = 0 (3.14)

若 ξ_1, \dots, ξ_{l_i} 是 (3.14) 式的基础解系, 那它们都是 A 的对应于 λ_i 的线性无关的特征向量;

(4) $\xi_t = (\xi_{t1}, \dots, \xi_{tn})^T$ ($t=1, 2, \dots, l_i$) 就是线性变换 A 的对应于 λ_i 的线性无关的特征向量;

(5) 若 $\xi_1, \dots, \xi_{l_1}, \dots, \xi_{l_k}$ 是 n 个线性无关的特征向量那么它们可作 V 的基, 且 A 在此基下的矩阵表示是对角阵

$$= \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_k \end{bmatrix}_{n \times n}$$

例 2 已知 $K[x]_n$ 中微分变换 D 在基 $1, x, x^2, \dots, x^n$ 之下的矩阵表示是

$$D = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 2 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & n \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

求 D 的特征值与特征向量。

解 求 $|E - D| = 0$ 的全部根。

$$\begin{vmatrix} \lambda - 1 & 0 & \dots & 0 \\ 0 & \lambda - 2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \lambda - n \end{vmatrix} = 0$$

所以 $\lambda^{n+1} = 0$

即 D 只有 0 作为其特征根 (n+1 重), 则 D 也只有 0 作为其特征值。

把特征值 $\lambda = 0$ 代入线性方程组 $(E - D)X = 0$

即
$$\begin{cases} -x_2 = 0 \\ -2x_3 = 0 \\ \dots \\ -nX_{n+1} = 0 \end{cases}$$

得基础解系 $(1, 0, 0, \dots, 0)$, 此为 D 的特征向量。

于是 D 也只有常数 $k(k \neq 0)$, 作为其特征向量。

例 3 求平面上逆时针旋转 Q 角的线性变换 T_Q 的特征值与特征向量。

解 选平面 R^2 的基 $\alpha_1 = (1, 0)$, $\alpha_2 = (0, 1)$, 线性变换 T_Q 在 α_1, α_2 之下的矩阵表示为

$$\begin{bmatrix} \cos Q & -\sin Q \\ \sin Q & \cos Q \end{bmatrix}$$

它的特征多项式为

$$\begin{vmatrix} -\cos Q & \sin Q \\ -\sin Q & -\cos Q \end{vmatrix} = \lambda^2 - 2\cos Q + 1$$

当 $Q \neq k\pi$ 时, 这个多项式没有实根, 因此, 当 $Q \neq k\pi$ 时, T_Q 没有特征值, 从几何上看, 这个结论是很明显的。

例 4 设线性变换 A 在基 $\alpha_1, \alpha_2, \alpha_3$ 之下的矩阵表示是

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

求 A 的特征值与特征向量。

解 因为特征多项式为

$$|E - A| = \begin{vmatrix} -1 & -2 & -2 \\ -2 & -1 & -2 \\ -2 & -2 & -1 \end{vmatrix} = (\lambda + 1)^2(\lambda - 5)$$

所以特征值是 $\lambda_1 = -1, \lambda_2 = -1, \lambda_3 = 5$ 。

求对应于 $\lambda = -1$ 的特征向量

$$(\lambda - E - A)x = 0, \text{ 即 } \begin{cases} -2x_1 - 2x_2 - 2x_3 = 0 \\ -2x_1 - 2x_2 - 2x_3 = 0 \\ -2x_1 - 2x_2 - 2x_3 = 0 \end{cases}$$

基础解系

$$\alpha_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \alpha_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

因此属于 -1 的两个线性无关的特征向量 $\alpha_1 = \alpha_1 - \alpha_3, \alpha_2 = \alpha_2 - \alpha_3$, 求对应于 $\lambda = 5$ 的特征向量

$$(5E - A)x = 0 \quad \text{即} \quad \begin{cases} 4x_1 - 2x_2 - 2x_3 = 0 \\ -2x_1 + 4x_2 - 2x_3 = 0 \\ -2x_1 - 2x_2 + 4x_3 = 0 \end{cases}$$

基础解系

$$\alpha_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

因此属于 5 的线性无关的特征向量 $\alpha_3 = \alpha_1 + \alpha_2 + \alpha_3$

例 5 设矩阵 $A = \begin{pmatrix} 3 & 1 & 0 \\ -4 & -1 & 0 \\ 4 & -8 & -2 \end{pmatrix}$

求 A 的特征值与特征向量。

解 $|E - A| = \begin{vmatrix} -3 & -1 & 0 \\ 4 & +1 & 0 \\ -4 & 8 & +2 \end{vmatrix} = (\lambda + 2)(\lambda - 1)^2$

所以 A 的特征值 $\lambda_1 = 1, \lambda_2 = 1, \lambda_3 = -2$ 。

求 $\lambda = 1$ 的特征向量

$$(E - A) = 0 \quad \text{即} \quad \begin{cases} -2x_1 - x_2 = 0 \\ 4x_1 + 2x_2 = 0 \\ -4x_1 + 8x_2 + 3x_3 = 0 \end{cases}$$

解得基础解系

$$\alpha_1 = \begin{pmatrix} 3 \\ -6 \\ 20 \end{pmatrix}$$

对应于 1 的 A 的特征向量是 $k \alpha_1$ (k 为任意非 0 数)。

求 $\lambda = -2$ 的特征向量 $(-2E - A) = 0$ 即
$$\begin{cases} -5x_1 - x_2 = 0 \\ 4x_1 - x_2 = 0 \\ -4x_1 + 8x_2 = 0 \end{cases}$$

解得基础解系

$$\alpha_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

对应于 -2 的 A 的特征向量是 $k \alpha_2$ ($k \neq 0$)。

当然我们不必担心 V 的基选得不同, 而使得线性变换 A 的特征值不同, 有如下定理。

定理 3.3.4 相似矩阵有完全相同的特征值。

注意: 定理 3.3.4 的逆定理不成立。例如 $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

以上一些例题可以看出, 确有一些矩阵可化为对角矩阵, 而也有一些矩阵不可能化为对角矩阵。如果一个矩阵可以化为对角矩阵, 其对角元素就是矩阵的全部特征值, 因此对角矩阵除了次序以外是惟一确定的, 但是相应的过渡矩阵却不惟一。事实上, 每一个特征向量都可相差一个非 0 常数; 而且对应于一个特征值的所有特征向量再添上 0 向量构成 K^n 的线性子空间 (如果 A 是 n 级方阵的话), 记为 V_λ 。即 $V_\lambda = \{ \alpha \mid (\lambda E - A) \alpha = 0 \}$ 。为了进一步研究矩阵化为对角矩阵的问题, 有下列定理。

定理 3.3.5 如果 $\alpha_1, \dots, \alpha_m$ 是 A 的分别属于互不相同的特征值 $\lambda_1, \dots, \lambda_m$ 的特征向量, 则 $\alpha_1, \dots, \alpha_m$ 线性无关。

证明(略)

推论: 若 n 级矩阵 A 有 n 个互不相同的特征值, 则 A 可相似于对角阵。

如果 A 有重根, 那么情况较复杂, 进一步有以下定理。

定理 3.3.6 设 $\lambda_1, \dots, \lambda_k$ 是 A 的互不相同的特征值, $\alpha_{11}, \dots, \alpha_{1r_1}$ 是 λ_1 对应的 r_1 个线性无关特征向量, $\dots, \alpha_{k1}, \dots, \alpha_{kr_k}$ 是 λ_k 对应的 r_k 个线性无关的特征值向量, 则 $\alpha_{11}, \dots, \alpha_{1r_1}, \dots, \alpha_{k1}, \dots, \alpha_{kr_k}$ 是线性无关的。

证明(略)

定理 3.3.7 λ_0 是 A 的一个 k 重特征值, 则对应于 λ_0 的 A 的线性无关特征向量的最多个数 k 。

证明(略)

推论 一个矩阵 A 在复数域 C 上能化为对角阵的充要条件是对应于 A 的每一个特征值的特征向量的线性无关最大个数等于该特征值的重根重数。

不过, 对于实对称矩阵 $A(A = A)$, 总能相似于对角阵, 即下面定理。

定理 3.3.8 任意实对称矩阵一定相似于对角阵, 且它的不同特征值对应的特征向量总是正交的。

关于这些结论可在线性代数中查阅, 此处不一一证明了。

下面要讨论的是: 在复数域 C 上任意一个 n 级方阵 A 总能相似于一个若当形矩阵, 那就是 A 在复数域上的最简单相似矩阵。

3.3.2 复数域上矩阵相似最简形——若当形矩阵

为了讨论这一问题, 还需要简单介绍一下变量 λ 的多项式矩阵。

以 λ 的多项式为元素的矩阵, 如

$$A(\lambda) = \begin{bmatrix} a_{11}(\lambda) & \dots & a_{1n}(\lambda) \\ \vdots & & \vdots \\ a_{m1}(\lambda) & \dots & a_{mn}(\lambda) \end{bmatrix} \quad a_{ij}(\lambda) \in K[\lambda]$$

$K[\lambda]$ 表示系数是数域 k 上的多项式全体

叫做多项式矩阵, 又称 λ -矩阵, 显然 λ -矩阵任何子式计算出来, 都是一个 λ 的多项式。

定义 3.13 若 λ -矩阵 $A(\lambda)$ 中有一个 $r(r \geq 1)$ 级子式不恒为 0, 而所有的 $r+1$ 级子式(若有的话)恒为 0, 则称 $A(\lambda)$ 的秩为 r , 记为 $\text{rank}A(\lambda) = r$ 简记为 $r(A) = r$ 。

例如 $A_1(\lambda) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, A_2(\lambda) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$

则 $r(A_1) = 1, r(A_2) = 2$ 。

定义 3.14 设 $A(\lambda)$ 是一个 n 级 λ -方阵, 若存在 n 级 λ -方阵 $B(\lambda)$, 使 $A(\lambda)B(\lambda) = B(\lambda)A(\lambda) = E$, 称 $A(\lambda)$ 是可逆 λ -矩阵, 并称 $B(\lambda)$ 是 $A(\lambda)$ 的逆阵, 记为 $A^{-1}(\lambda)$ 。

容易证明, 若 $A(\lambda)$ 可逆, 则 $A^{-1}(\lambda)$ 是惟一的。

在数字矩阵中, 满秩矩阵就是可逆阵, 但满秩的 λ -矩阵未必可逆, 例如, $A(\lambda) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, 就没有逆阵。

定理 3.3.9 一个 n 级 λ -方阵 $A(\lambda)$ 可逆的充要条件是行列式 $|A(\lambda)|$ 为非 0 常数。

证明 若 $A(\lambda)$ 可逆, 那么有 $B(\lambda)$, 使

$$A(\lambda)B(\lambda) = B(\lambda)A(\lambda) = E$$

两边取行列式 $|A(\lambda)||B(\lambda)| = 1$

因为 $|A(\lambda)|, |B(\lambda)|$ 都是 λ 多项式, 由次数性质, $|A(\lambda)|$ 与 $|B(\lambda)|$ 都是零次多项式, 即非 0 常数。

反过来, 若 $|A(\lambda)|$ 是一个非 0 常数, 那么令 $B(\lambda) = \frac{1}{|A(\lambda)|}A^*(\lambda)$, 其中 $A^*(\lambda)$ 表示 $A(\lambda)$ 的伴随矩阵, 这是一个 λ -矩阵, 而

$$A(\lambda) \cdot \frac{1}{|A(\lambda)|}A^*(\lambda) = \frac{1}{|A(\lambda)|}A^*(\lambda) \cdot A(\lambda) = E$$

$a_{ij}(\lambda) \in k[\lambda]$ 设 $a_{ij}(\lambda) = a_n \lambda^n + a_{n-1} \lambda^{n-1} + \dots + a_1 \lambda + a_0$
若 $a_n \neq 0$, 称多项式 $a_{ij}(\lambda)$ 的次数为 n , 记 a_{ij}
若 $a_n = 1$, 称 $a_{ij}(\lambda)$ 是首一多项式

所以 $A(\lambda)$ 是可逆的, 且 $A^{-1}(\lambda) = \frac{1}{|A(\lambda)|} A^*(\lambda)$ 。

定义 3.15 以下三种变换称为 λ -矩阵的初等变换

(1) 互换 $A(\lambda)$ 的 i, j 两行(列), 相当于左(右)乘以初等 λ -矩阵;

$$P(i, j) = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

(2) $A(\lambda)$ 的第 i 行(列)扩大 k 倍($k \neq 0$), 相当于左(右)乘以初等 λ -矩阵;

$$p(i(k)) = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & k & \\ & & & \ddots \end{pmatrix};$$

(3) 把 $A(\lambda)$ 的 i 行的 $f(\lambda)$ 倍(f 是一个多项式) 加到第 j 行上, 相当于左乘以初等 λ -矩阵;

$$P(j + i(f(\lambda))) = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \ddots \end{pmatrix}$$

若 $A(\lambda)$ 经有限次初等变换变为 $B(\lambda)$, 称 $A(\lambda)$ 与 $B(\lambda)$ 等价, 记作 $A(\lambda) \sim B(\lambda)$ 。
可以验证等价关系满足

- (1) 自反性: $A(\lambda) \sim A(\lambda)$;
- (2) 对称性: $A(\lambda) \sim B(\lambda)$ 则 $B(\lambda) \sim A(\lambda)$;
- (3) 传递性: $A(\lambda) \sim B(\lambda), B(\lambda) \sim C(\lambda)$ 则 $A(\lambda) \sim C(\lambda)$ 。

定理 3.3.10 两个 $m \times n$ 阶的 λ -矩阵 $A(\lambda), B(\lambda)$ 等价的充要条件是存在可逆 m 阶 $P(\lambda)$ 阵及 n 阶 $Q(\lambda)$ 阵, 使

$$P(\lambda) A(\lambda) Q(\lambda) = B(\lambda)$$

这从初等变换的定义很容易得知, 因为初等 λ -矩阵皆可逆, 且逆也是初等 λ -矩阵。

定理 3.3.11 设 A, B 是两个数字方阵, 则 $A \sim B$ 的充分必要条件是 $E - A \sim E - B$ 。

证明 若 $A \sim B$, 则存在可逆阵 P , 使 $P^{-1}AP = B$ 。

那么 $E - B = E - P^{-1}AP = P^{-1}(E - A)P$

由定理 3.3.10, $E - B \sim E - A$ 。

(充分性证明从略。)

这个定理的作用在于沟通了数字矩阵的相似关系与它们特征矩阵等价的关系, 可以通过研究特征矩阵等价关系来研究相似关系。

定理 3.3.12 任意一个秩为 r 的 $m \times n$ 阶 λ -矩阵 $A(\lambda)$ 都等价于一个对角形 λ -矩阵

$$\begin{pmatrix} d_1(\lambda) & & 0 & \cdots & 0 \\ & & & & \\ & 0 & & d_r(\lambda) & \cdots & 0 \\ & 0 & \cdots & \cdots & \cdots & 0 \\ & 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}$$

其中 $d_i(\lambda)$ ($i=1, \dots, r$) 都是首 1 多项式, 且 $d_i(\lambda) \mid d_{i+1}(\lambda)$, $i=1, 2, \dots, r-1$, 此对角形矩阵叫 $A(\lambda)$ 的 smith 标准形。(注: $d_i(\lambda) \mid d_{i+1}(\lambda)$ 表示存在多项式 $g(\lambda)$, 使 $d_{i+1}(\lambda) = d_i(\lambda)g(\lambda)$)

本定理的证明较繁, 略去证明, 我们以例子说明如何用初等变换将 $A(\lambda)$ 化为它的 smith 标准形。

例 6 用初等变换化 λ -矩阵

$$A(\lambda) = \begin{pmatrix} 1 - \lambda^2 & & \\ & -\lambda^2 & \\ 1 + \lambda^2 & \lambda^2 & -\lambda^2 \end{pmatrix}$$

为 smith 标准形。

解 $A(\lambda) \xrightarrow{P(1+3)} \begin{pmatrix} 1 - \lambda^2 & & \\ 0 & -\lambda^2 & \\ 1 & \lambda^2 & -\lambda^2 \end{pmatrix} \xrightarrow{P(3-1)} \begin{pmatrix} 1 - \lambda^2 & & \\ 0 & -\lambda^2 & \\ 0 & 0 & -\lambda^2 - \lambda^2 \end{pmatrix}$

$\xrightarrow{P(3+2)} \begin{pmatrix} 1 - \lambda^2 & & +\lambda^2 \\ 0 & 0 & \\ 0 & 0 & -\lambda^2 - \lambda^2 \end{pmatrix} \xrightarrow{P(1-2(\lambda))} \begin{pmatrix} 1 & 0 & +\lambda^2 \\ 0 & 0 & \\ 0 & 0 & -\lambda^2 - \lambda^2 \end{pmatrix}$

$\xrightarrow{P(1+3)} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & \\ 0 & 0 & (-\lambda^2 + 1) \end{pmatrix}$

定义 3.16 设 $\text{rank}A(\lambda) = r$ ($r \geq 1$), 则对正整数 k ($1 \leq k \leq r$), $A(\lambda)$ 中必有非 0 的 k 阶子式。 $A(\lambda)$ 的全部 k 阶子式的最高公因式(首 1 多项式) 记为 $D_k(\lambda)$, 称之为 $A(\lambda)$ 的 k 阶行列式因子。

显然当 $r = 1$ 时, $A(\lambda)$ 有 r 个行列式因子 $D_1(\lambda), D_2(\lambda), \dots, D_r(\lambda)$ 。

定理 3.3.13 若 $A(\lambda) \sim B(\lambda)$, 则 $A(\lambda)$ 与 $B(\lambda)$ 有相同的秩及各级行列式因子。

证明(略)

所以 $A(\lambda)$ 经有限次初等变换成为 smith 标准形 $B(\lambda)$, 即

$$A(\lambda) - B(\lambda) = \begin{pmatrix} d_1(\lambda) & & & & 0 \\ & d_2(\lambda) & & & \\ & & \ddots & & \\ & & & d_r(\lambda) & \\ & & & & 0 \\ 0 & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix}$$

因为 $d_1(\lambda), \dots, d_r(\lambda)$ 皆为多项式, $d_i(\lambda) \mid d_{i+1}(\lambda) \quad i = 1, 2, \dots, r - 1$ 。
那么 $D_1(\lambda) = d_1(\lambda), D_2(\lambda) = d_1(\lambda) d_2(\lambda), \dots, D_r(\lambda) = d_1(\lambda) \dots d_r(\lambda)$, 即
$$d_1(\lambda) = D_1(\lambda), d_2(\lambda) = \frac{D_2(\lambda)}{D_1(\lambda)}, \dots, d_r(\lambda) = \frac{D_r(\lambda)}{D_{r-1}(\lambda)}$$

 $d_1(\lambda), d_2(\lambda), \dots, d_r(\lambda)$ 又称为 $A(\lambda)$ 的不变因子, 因此 $A(\lambda)$ 的 smith 标准形是惟一的。
我们还可以用求行列式因子的方法来求不变因子, 从而得到 smith 标准形。

例 7 设 $A(\lambda) = \begin{pmatrix} 1 - \lambda^2 & & \\ & -\lambda^2 & \\ 1 + \lambda^2 & \lambda^2 & -\lambda^2 \end{pmatrix}$

求 $A(\lambda)$ 的 Smith 标准形。

解 因为 $a_{11}(\lambda)$ 与 $a_{21}(\lambda)$ 互素, 故 $D_1(\lambda) = 1$

又因为 $\begin{vmatrix} 1 - \lambda^2 \end{vmatrix} = (1 - \lambda^2)$, $\begin{vmatrix} -\lambda^2 \end{vmatrix} = -\lambda^2(1 - \lambda^2)$

且显然 $-\lambda^2$ 是每个 2 阶子式的公因子, 故 $D_2(\lambda) = -\lambda^2$ 。

$$D_3(\lambda) = |A(\lambda)| = -\lambda^2(\lambda^2 + 1)$$

那么 $A(\lambda)$ 的 Smith 标准形为

$$\begin{pmatrix} 1 & & \\ & \lambda^2 & \\ & & (\lambda^2 + 1) \end{pmatrix}$$

一般可以将初等变换与行列式因子结合使用求 Smith 标准形。

定义 3.17 设 $A(\lambda)$ 的不变因子 $d_1(\lambda), d_2(\lambda), \dots, d_r(\lambda)$, 这些不变因子在复数域分解为一次因式方幂

$$\begin{aligned} d_1(\lambda) &= (\lambda - \lambda_1)^{l_{11}} (\lambda - \lambda_2)^{l_{12}} \dots (\lambda - \lambda_t)^{l_{1t}} \\ d_2(\lambda) &= (\lambda - \lambda_1)^{l_{21}} (\lambda - \lambda_2)^{l_{22}} \dots (\lambda - \lambda_t)^{l_{2t}} \\ &\dots\dots\dots \\ d_r(\lambda) &= (\lambda - \lambda_1)^{l_{r1}} (\lambda - \lambda_2)^{l_{r2}} \dots (\lambda - \lambda_t)^{l_{rt}} \end{aligned}$$

其中 $l_{ij} \geq 0 (i = 1, \dots, r; j = 1, 2, \dots, t)$ 称其中 $l_{ij} > 0$ 的一切 $(\lambda - \lambda_j)^{l_{ij}}$ 为 $A(\lambda)$ 的初等因子。对于数字矩阵 $A, E - \lambda A$ 的初等因子就称为 A 的初等因子。

应当注意, $A(\lambda)$ 的初等因子有可能有相同的。如例 7 的 $A(\lambda)$ 的初等因子是 $-\lambda^2, (\lambda^2 + 1)$ 。

注: 我们称 n 阶方阵 A 的特征矩阵 $E - A$ 的初等因子为矩阵 A 的初等因子。

例 8 设 $A \in K^{12 \times 12}$, 且

$$E - A = \begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & (\lambda - 1)^2 & & & \\ & & & (\lambda - 1)^2(\lambda + 1) & & \\ & & & & (\lambda - 1)^2(\lambda + 1)(\lambda - i)^2(\lambda + i)^2 & \\ & & & & & \ddots \end{pmatrix}_{12 \times 12}$$

求 A 的全部初等因子。

解 A 的全部初等因子有

$$(\lambda + 1), (\lambda - 1)^2, (\lambda - 1)^2, (\lambda - 1)^2, (\lambda + 1), (\lambda - i)^2, (\lambda + i)^2$$

反过来若知道 A 的全部初等因子及 A 的阶数, 就可以写出 A 的 Smith 标准形来。

下面给出一个总结性的定理:

定理 3.3.14 $A(\lambda)$ 与 $B(\lambda)$ 都是 $m \times n$ 阶 λ -矩阵

则以下命题等价。

- 1° $A(\lambda) \sim B(\lambda)$
- 2° $A(\lambda), B(\lambda)$ 有相同的 Smith 标准形;
- 3° $A(\lambda), B(\lambda)$ 有相同的各级行列式因子;
- 4° $A(\lambda), B(\lambda)$ 有相同的不变因子;
- 5° $A(\lambda), B(\lambda)$ 有相同的秩及初等因子;
- 6° 存在可逆矩阵 $P(\lambda), Q(\lambda)$, 使 $P(\lambda)A(\lambda)Q(\lambda) = B(\lambda)$ 。

注意, 若 $A(\lambda)$ 与 $B(\lambda)$ 都是 $m \times n$ 阶 λ -矩阵, 且初等因子相同, 是不能得到 $A(\lambda) \sim B(\lambda)$ 的。例如

$$A(\lambda) = \begin{bmatrix} \lambda + 1 & 0 \\ 0 & \lambda - 1 \end{bmatrix}, B(\lambda) = \begin{bmatrix} 1 & 0 \\ 0 & (\lambda + 1)(\lambda - 1) \end{bmatrix}, B(\lambda) = \begin{bmatrix} \lambda^2 - 1 & 0 \\ 0 & 0 \end{bmatrix}$$

则 $A(\lambda)$ 与 $B(\lambda)$ 是不等价的。

例 9 求 n 阶 λ -矩阵

$$A(\lambda) = \begin{pmatrix} \lambda - \alpha_1 & & & \\ b_1 & \lambda - \alpha_2 & & \\ & & \ddots & \\ & & & \lambda - \alpha_{n-1} & \\ & & & & b_{n-1} & \lambda - \alpha_n \end{pmatrix}$$

的初等因子, 这里 b_1, b_2, \dots, b_{n-1} 都是非 0 常数。

解 显然 $D_n(\lambda) = (\lambda - \alpha_n)^n$, 又在 $A(\lambda)$ 中划去第 1 行和第 n 列, 得到一个右下角的 $n - 1$ 阶非 0 子式 $= b_1 \dots b_{n-1}$, 所以 $D_{n-1}(\lambda) = 1$ 。因此由 $D_{i-1}(\lambda) \mid D_i(\lambda)$, 知 $D_{n-1}(\lambda) = 1, \dots, D_1(\lambda) = 1$, 从而得:

$$d_1(\lambda) = 1, \dots, d_{n-1}(\lambda) = 1, d_n(\lambda) = (\lambda - \alpha_n)^n, \text{ 可见 } A(\lambda) \text{ 的初等因子只有一个 } (\lambda - \alpha_n)^n。$$

以上讲了许多概念: Smith 标准形、初等因子、行列式因子、不变因子及定理 3.3.14, 知道它们都是初等变换的不变量, 这些与讨论的若当标准形有什么联系呢?

定义 3.18 形式为

$$J(\lambda_0, t) = \begin{pmatrix} \lambda_0 & & & \\ 1 & \lambda_0 & & \\ & 1 & \ddots & \\ & & \ddots & \lambda_0 \\ & & & 1 & \lambda_0 \end{pmatrix}_{t \times t}$$

的矩阵称为若当(Jordan)块, 其中 λ_0 是复数。由若干个若当块组成的准对角矩阵称为若当形矩阵。

例如

$$\begin{pmatrix} 2 & & & \\ & 1 & 2 & \\ & & 1 & 2 \\ & & & 1 & 2 \end{pmatrix} \begin{pmatrix} 0 & & & \\ 1 & 0 & & \\ & 1 & 0 & \\ & & 1 & 0 \end{pmatrix} \begin{pmatrix} i & & \\ & 1 & i \end{pmatrix}$$

都是若当块, 而

$$\begin{pmatrix} 1 & & & & \\ & 1 & 1 & & \\ & & & 4 & \\ & & & & i \\ & & & & 1 & i \\ & & & & & 1 & i \end{pmatrix}$$

就是一个若当形矩阵(没有填上数的地方全是 0)。

一阶若当块就是一阶矩阵, 因此若当形矩阵包括对角阵。

1) 若当块 $J(\lambda_0, t)$ 的初等因子。

$$E - J(\lambda_0, t) = \begin{pmatrix} -\lambda_0 & & & \\ & -\lambda_0 & & \\ & 1 & -\lambda_0 & \\ & & \ddots & \ddots \\ & & & 1 & -\lambda_0 \end{pmatrix}_{t \times t}$$

由例 9 这一个 t -矩阵的初等因子是 $(-\lambda_0)^t$, 也即是 $J(\lambda_0, t)$ 的初等因子, 反过来, 给出一个一次因式幂 $(-\lambda_0)^t$, 可以惟一写出一个若当块 $J(\lambda_0, t)$, 使 $J(\lambda_0, t)$ 的初等因子就是 $(-\lambda_0)^t$ 。

2) 若当形矩阵的初等因子。

在讨论此问题之前先给一个定理, 证明(略)。

定理 3.3.15 若 $A(\lambda)$ 呈分块对角形

$$A(\lambda) = \begin{pmatrix} A_1(\lambda) & & 0 \\ & A_2(\lambda) & \\ 0 & & A_r(\lambda) \end{pmatrix}$$

$A_1(\lambda), \dots, A_r(\lambda)$ 都是 n_i -矩阵, 则 $A_1(\lambda), \dots, A_r(\lambda)$ 的全部初等因子就是 $A(\lambda)$ 的全部初等

因子。

设若当形矩阵 J

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_s \end{pmatrix}$$

其中 $J_i = J_i(\lambda_i, t_i), (i = 1, 2, \dots, s),$

那么它的特征矩阵 $E - J$ 。

$$E - J = \begin{pmatrix} \lambda_1 - \lambda & & & & \\ & \lambda_1 - \lambda & & & \\ & & \ddots & & \\ & & & \lambda_1 - \lambda & \\ & & & & \ddots \\ & & & & & \lambda_2 - \lambda & \\ & & & & & & \ddots \\ & & & & & & & \lambda_2 - \lambda & \\ & & & & & & & & \ddots \\ & & & & & & & & & \lambda_s - \lambda & \\ & & & & & & & & & & \ddots \\ & & & & & & & & & & & \lambda_s - \lambda \end{pmatrix}$$
$$= \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & (\lambda_1 - \lambda)^{t_1} & & \\ & & & 1 & \\ & & & & \ddots \\ & & & & & 1 & \\ & & & & & & (\lambda_2 - \lambda)^{t_2} \\ & & & & & & & \ddots \\ & & & & & & & & 1 \\ & & & & & & & & & \ddots \\ & & & & & & & & & & 1 \\ & & & & & & & & & & & (\lambda_s - \lambda)^{t_s} \end{pmatrix}$$

由定理 3.3.15 可知 J 的全部初等因子就是 $(\lambda - \lambda_1)^{t_1}, \dots, (\lambda - \lambda_s)^{t_s}$ 。反过来, 若给定这些一次因式方幂 $(\lambda - \lambda_1)^{t_1}, \dots, (\lambda - \lambda_s)^{t_s}$, 可以作出若当形矩阵 J , 若不计较若当块的顺序, 可以认为 J 是被这些一次因式方幂惟一确定的。

现在可以得到此节最后的一个定理

定理 3.3.16 复数域上任意一个 n 级方阵 A , 总是可以相似于一个若当形矩阵 J , 若不计较若当块的顺序, J 还是惟一的。

证明 设 A 的全部初等因子就是 $(\lambda - \lambda_1)^{t_1}, \dots, (\lambda - \lambda_s)^{t_s}$ 。那么可以构造若当形矩阵 J

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_s \end{pmatrix}$$

其中 $J_i = J(\lambda_i, t_i)$, 当然 J 是以 $(\lambda - \lambda_1)^{t_1}, \dots, (\lambda - \lambda_s)^{t_s}$ 作为其全部初等因子。注意到 $E - A$ 与 $E - J$ 的秩都是 $t_1 + \dots + t_s$, 由定理 3.3.14 可知, 故 $E - A \sim E - J$, 又由定理 3.3.11 $A \sim J$ 。

例 10 求矩阵

$$A = \begin{pmatrix} 0 & 3 & 3 \\ -1 & 8 & 6 \\ 2 & -14 & -10 \end{pmatrix}$$

的若当标准形 J 。

解 因为 $E - A = \begin{pmatrix} 1 & -3 & -3 \\ 1 & -8 & -6 \\ -2 & 14 & +10 \end{pmatrix}$

显然 $D_1(\lambda) = 1$, 也可计算出 $D_2(\lambda) = 1, D_3(\lambda) = (\lambda + 1)^2$

所以 $d_1(\lambda) = 1, d_2(\lambda) = 1, d_3(\lambda) = (\lambda + 1)^2$, 初等因子为 $\lambda, (\lambda + 1)^2$, A 的若当标准形 J

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

例 11 求矩阵

$$A = \begin{pmatrix} 4 & 6 & 0 \\ -3 & -5 & 0 \\ -3 & -6 & 1 \end{pmatrix}$$

的若当标准形 J 。

解 $E - A = \begin{pmatrix} -4 & -6 & 0 \\ 3 & +5 & 0 \\ 3 & 6 & -1 \end{pmatrix} \xrightarrow{(3-2)} \begin{pmatrix} -4 & -6 & 0 \\ 3 & +5 & 0 \\ 0 & 1 & -1 \end{pmatrix}$

$$\xrightarrow{(2+3)} \begin{pmatrix} -4 & -6 & 0 \\ 3 & +5 & 0 \\ 0 & 0 & -1 \end{pmatrix} \xrightarrow{(1+2)} \begin{pmatrix} -1 & -1 & 0 \\ 3 & +5 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$
$$\xrightarrow{} \begin{pmatrix} -1 & 0 & 0 \\ 3 & +2 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

显然 $D_1(\lambda) = 1, D_2(\lambda) = (\lambda - 1), D_3(\lambda) = (\lambda + 2)(\lambda - 1)^2$ 。

所以 A 的初等因子 $(\lambda - 1), (\lambda - 1), (\lambda + 2)$, 那么若当形 J

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

不难看出, 当且仅当初等因子全为一次式时, A 相似于对角阵。
下面介绍一个命题, 不用求行列式因子及不变因子, 可直接求初等因子。
命题 若 n -矩阵 $A(\lambda)$

$$A(\lambda) = \begin{pmatrix} f_1(\lambda) & & & \\ & f_r(\lambda) & & \\ & & 0 & \\ & & & \ddots & \\ & & & & 0 \end{pmatrix}$$

$f_1(\lambda), \dots, f_r(\lambda)$ 皆是 λ 的多项式, 那么 $f_1(\lambda), \dots, f_r(\lambda)$ 的所有一次因式方幂就是 $A(\lambda)$ 的全部初等因子。

例 12 设 $A = \begin{pmatrix} 3 & 0 & 8 \\ 3 & -1 & 6 \\ -2 & 0 & -5 \end{pmatrix}$

求 A 的若当标准形 J 及变换阵 P , 使 $P^{-1}AP = J$ 。

解 $E - A = \begin{pmatrix} -3 & 0 & -8 \\ -3 & +1 & -6 \\ 2 & 0 & +5 \end{pmatrix} \xrightarrow{p(1-2)} \begin{pmatrix} -3 & -1 & -2 \\ -3 & +1 & -6 \\ 2 & 0 & +5 \end{pmatrix}$

$$\xrightarrow{p(1+2)} \begin{pmatrix} -1 & -1 & -2 \\ -2 & +1 & -6 \\ 2 & 0 & +5 \end{pmatrix}$$
$$\xrightarrow{p(2+1(2)) \atop (3+1(2))} \begin{pmatrix} -1 & -1 & -2 \\ 0 & -2+2+3 & -2-2 \\ 0 & -2-2 & +1 \end{pmatrix}$$

显然 $D_1(\lambda) = 1, D_2(\lambda) = \lambda + 1, D_3(\lambda) = (\lambda + 1)^3$, 故初等因子为 $\lambda + 1, (\lambda + 1)^2$, 所以 A 的若当标准形 J

$$J = \begin{pmatrix} -1 & & \\ & -1 & \\ & & 1 & -1 \end{pmatrix}$$

设 $P = (\alpha_1 \ \alpha_2 \ \alpha_3)$, 由 $A(\alpha_1 \ \alpha_2 \ \alpha_3) = (\alpha_1 \ \alpha_2 \ \alpha_3) \begin{pmatrix} -1 & & \\ & -1 & \\ & & 1 & -1 \end{pmatrix}$

可知 $A\alpha_i = -\alpha_i \ (i=1, 3), \quad (A+E)\alpha_2 = \alpha_3$

α_1, α_3 皆为 -1 对应的特征向量, 求出特征向量 $\alpha_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \alpha_3 = \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}$ 。为使方程组

$(A + E) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix}$ 相容, 可令 $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = k_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + k_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + k_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$
即解方程组

$$\begin{cases} 4x_1 + 8x_3 = 2k_2 \\ 3x_1 + 6x_3 = k_1 \\ -2x_1 - 4x_3 = -k_2 \end{cases}, \quad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$\left(\begin{array}{ccc|c} 4 & 0 & 8 & 2k_2 \\ 3 & 0 & 6 & k_1 \\ -2 & 0 & -4 & -k_2 \end{array} \right) \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & 2 & \frac{1}{2}k_2 \\ 0 & 0 & 0 & -\frac{3}{2}k_2 + k_1 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

所以可选 $k_2 = 2, k_1 = 3$, 即选 $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \\ -2 \end{pmatrix}$, 则方程组有解。

解得 $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, 即 $P = \begin{pmatrix} 0 & 4 & 1 \\ 1 & 3 & 0 \\ 0 & -2 & 0 \end{pmatrix}$

3.4 Hamlilton-Cayley 定理、最小多项式

本节内容为矩阵函数的基础。
定义 3.19 设 A 是 n 阶方阵, 若存在多项式 $f(\lambda)$, 使得 $f(A) = 0$, 即 $f(A)$ 是零矩阵, 称 $f(\lambda)$ 是矩阵 A 的零化多项式。

下面指出两点:
1) 对任何 n 阶方阵 A , 都存在零化多项式。因为线性空间 $K^{n \times n}$ 是 n^2 维的, 故 $E, A, A^2, \dots, A^{n^2}$ 必定线性相关。故存在不全为 0 的数 $k_0, k_1, k_2, \dots, k_{n^2}$, 使

$$k_0 E + k_1 A + k_2 A^2 + \dots + k_{n^2} A^{n^2} = 0$$

即多项式

$$f(\lambda) = k_0 + k_1 \lambda + k_2 \lambda^2 + \dots + k_{n^2} \lambda^{n^2}$$

是 A 的零化多项式。

2) 任何矩阵的零化多项式不惟一。因为若 $f(\lambda)$ 是 A 的零化多项式, 则 $f(\lambda)g(\lambda)$ 也是 A 的零化多项式, 这里的 $g(\lambda)$ 可以是任意的非零多项式。

定理(Hamlilton-Cayley 定理) 设

$$f(\lambda) = |\lambda E - A| = \lambda^n + a_1 \lambda^{n-1} + \dots + a_{n-1} \lambda + a_n$$

则 $f(A) = A^n + a_1 A^{n-1} + \dots + a_n E = 0$

证明 因为特征矩阵 $\lambda E - A$ 的伴随矩阵的元素是由 $\lambda E - A$ 的 $n - 1$ 阶子式作成的元素 $b_{ij}(\lambda)$ 这是一个次数不超过 $n - 1$ 次的多项式, 故伴随矩阵 $B(\lambda)$ 是一个 $(n - 1)$ -矩阵, 从而有 $B_i \in K^{n \times n}, (i = 1, 2, \dots, n)$, 使得

$$B(\lambda) = B_0 + B_1 \lambda + \dots + B_{n-1} \lambda^{n-1}$$

那么

$$\begin{aligned} & (E - A)(B_0 + B_1 + \dots + B_{n-1}) \\ &= B_{n-1} + (B_{n-2} - AB_{n-1}) + \dots + (B_0 - AB_1) - AB_0 \tag{*} \\ &= |E - A|E \\ &= (x^n + a_1 x^{n-1} + \dots + a_{n-1}x + a_n)E \tag{**} \end{aligned}$$

对比(*)与(**)

$$\begin{aligned} B_{n-1} &= E \\ B_{n-2} - AB_{n-1} &= a_1 E \\ &\dots \\ B_0 - AB_1 &= a_{n-1} E \\ -AB_0 &= a_n E \end{aligned}$$

以 $A^n, A^{n-1}, A^{n-2}, \dots, A, E$ 自上至下依次左乘等式两端, 然后相加便得

$$0 = A^n + a_1 A^{n-1} + a_2 A^{n-2} + \dots + a_{n-1} A + a_n E$$

此即 $f(A) = 0$

证毕。

根据 Hamilton-Cayley 定理, 对于 n 阶方阵 A , 当 $k \geq n$ 时, 计算 A^k 可以用小于 n 的 A 的方幂来表示, 从而简化矩阵运算。

例 1 设

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

试计算 $g(A) = A^7 - A^5 - 19A^4 + 28A^3 + 6A - 4E$

解 $f(x) = |E - A| = x^3 - 4x^2 + 5x - 2$

令 $g(x) = x^7 - x^5 - 19x^4 + 28x^3 + 6x - 4$

用 $f(x)$ 去除 $g(x)$, 得

$$g(x) = f(x)(x^4 + 4x^3 + 10x^2 + 3x - 2) + (-3x^2 + 22x - 8)$$

那 $g(A) = f(A)(A^4 + 4A^3 + 10A^2 + 3A - 2E) - 3A^2 + 22A - 8E$

由 Hamilton-Caley 定理 $f(A) = 0$, 于是

$$\begin{aligned} g(A) &= -3A^2 + 22A - 8E \\ &= \begin{pmatrix} -19 & 16 & 0 \\ -64 & 43 & 0 \\ 19 & -3 & 24 \end{pmatrix} \end{aligned}$$

上面已经指出任何 n 阶方阵 A 的零化多项式是不惟一的, 我们希望找到一个次数最低的零化多项式, 例如对于

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

$f(x) = |E - A| = (x - 2)^2$ 是 A 的零化式项式, 而且 $m(x) = x - 2$ 也是 A 的零化多项式。但是次数比 $m(x)$ 低的多项式是零次多项式 $c(x) = c(0)$

$$(A) = cE - 0$$

因此 $m(\lambda)$ 是 A 的一切零化多项式中次数最低的零化多项式。

定义 3.20 在 n 阶方阵 A 的所有零化多项式中, 次数最低的首一多项式, 称为 A 的最小多项式(Minimal Polynomaial), 记为 $m(\lambda)$

由 Hamilton-Cayley 定理可知, 任何 n 阶方阵 A 的最小多项式是存在的, 并且次数不超过 n 。

下面先讨论最小多项式的性质, 再介绍最小多项式的求法。

定理 3.4.1 n 阶方阵 A 的任意零化多项式都可以被 A 的最小多项式整除。

证明 设 $m(\lambda)$ 是 A 的最小多项式, $g(\lambda)$ 是 A 的任意一个零化多项式, 那么

$$g(\lambda) = m(\lambda)q(\lambda) + r(\lambda), r(\lambda) = 0 \text{ 或 } r < m$$

则
$$g(A) = m(A)q(A) + r(A)$$

由于 $g(A) = 0, m(A) = 0$, 所以 $r(A) = 0$

如果 $r(\lambda) \neq 0$, 说明 $r(\lambda)$ 是比 $m(\lambda)$ 次数还低的零化多项式, 矛盾。

因此 $r(\lambda) = 0$, 说明 $m(\lambda) | g(\lambda)$ 。

证毕。

定理 3.4.2 A 的最小多项式是惟一的。

证明 若 $m(\lambda), m^*(\lambda)$ 都是 A 的最小多项式, 那么由定理 3.4.1, 就应该

$$m(\lambda) | m^*(\lambda), m^*(\lambda) | m(\lambda)$$

即

$$m(\lambda) = m^*(\lambda)q(\lambda), m^*(\lambda) = m(\lambda)p(\lambda)$$

于是
$$m(\lambda) = m(\lambda)p(\lambda)q(\lambda)$$

由多项式乘法的次数性质, $p(\lambda)q(\lambda)$ 是零次多项式 C , 且因 $m(\lambda)$ 与 $m^*(\lambda)$ 都是首 1 的, 所以

$$m(\lambda) = m^*(\lambda)$$

证毕。

定理 3.4.3 A 的最小多项式的根是 A 的特征根, 反过来, A 的特征根必是 A 的最小多项式的根。

证明 设 $m(\lambda)$ 是 A 的最小多项式, $f(\lambda)$ 是 A 的特征多项式, 若 λ_0 是 $m(\lambda)$ 的根, 那么 $m(\lambda_0) = 0$, 由定理 3.4.1

$$f(\lambda) = m(\lambda)q(\lambda)$$

所以
$$f(\lambda_0) = m(\lambda_0)q(\lambda_0) = 0$$

说明 λ_0 是 A 的特征根。

反过来, 设 λ_0 是 A 的特征根, α 是 A 的对应于 λ_0 的特征向量, 则有 $A\alpha = \lambda_0\alpha$, 又因为 $A^k\alpha = \lambda_0^k\alpha$, 所以

$$m(A)\alpha = m(\lambda_0)\alpha = 0$$

由于 $\alpha \neq 0$, 所以 $m(\lambda_0) = 0$, 说明 λ_0 是 $m(\lambda)$ 的根。

证毕。

由定理 3.4.3, 可以得到如下结论:

若
$$f(\lambda) = |\lambda E - A| = (\lambda - \lambda_1)^{l_1}(\lambda - \lambda_2)^{l_2} \dots (\lambda - \lambda_s)^{l_s}$$

则 $m(\lambda) = (\lambda - \lambda_1)^{t_1} (\lambda - \lambda_2)^{t_2} \dots (\lambda - \lambda_s)^{t_s}$
其中 $1 \leq t_i \leq l_i \quad (i = 1, 2, \dots, s)$.

例 2 分别求矩阵

$$A = \begin{pmatrix} 7 & 4 & -1 \\ 4 & 7 & -1 \\ -4 & -4 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 3 & 1 & 0 & 0 \\ -4 & -1 & 0 & 0 \\ 7 & 1 & 2 & 1 \\ -7 & -6 & -1 & 0 \end{pmatrix}$$

的最小多项式。

解 先求 A 的最小多项式

$$f(\lambda) = |\lambda E - A| = \begin{vmatrix} \lambda - 7 & -4 & 1 \\ -4 & \lambda - 7 & 1 \\ 4 & 4 & \lambda - 4 \end{vmatrix} = (\lambda - 3)^2 (\lambda - 12)$$

由上述结论, $m(\lambda)$ 可能是 $(\lambda - 3)(\lambda - 12)$, 也可能就是 $f(\lambda)$, 只有这两种情况, 经验算

$$(\lambda E - A)(\lambda E - 12E) = \begin{pmatrix} 4 & 4 & -1 \\ 4 & 4 & -1 \\ -4 & -4 & 1 \end{pmatrix} \begin{pmatrix} -5 & 4 & -1 \\ 4 & -5 & -1 \\ -4 & -4 & -8 \end{pmatrix} = 0$$

于是 $m(\lambda) = (\lambda - 3)(\lambda - 12)$

再求 B 的最小多项式

$$f(\lambda) = |\lambda E - B| = \begin{vmatrix} \lambda - 3 & -1 & 0 & 0 \\ 4 & \lambda + 1 & 0 & 0 \\ -7 & -1 & \lambda - 2 & -1 \\ 7 & 6 & 1 & \lambda \end{vmatrix} = (\lambda - 1)^4$$

B 的最小多项式 $m(\lambda)$, 可能是 $(\lambda - 1)^4, (\lambda - 1)^3, (\lambda - 1)^2, (\lambda - 1)$, 四种情况。
经验算

$$(\lambda E - B)^3 \neq 0$$

所以 B 的最小多项式是 $(\lambda - 1)^4$ 。

下面再介绍一种求最小多项式的方法。

定理 3.4.4 n 阶方阵 A 的最小多项式 $m(\lambda) = d_n(\lambda)$ (第 n 个不变因子)

证明 (略)

关于求 A 的第 n 个不变因子的方法:

方法 1: 将 $(\lambda E - A)$ 用初等变换化为 Smith 标准形, 最后一个不变因子就是 $d_n(\lambda)$ 。

方法 2: 将 $(\lambda E - A)$ 用初等变换适当变形, 使其好求行列式因子 $D_i(\lambda)$, 则

$$d_n(\lambda) = \frac{D_n(\lambda)}{D_{n-1}(\lambda)}$$

方法 3: 若 $(\lambda E - A)$ 容易变为对角阵, 先求初等因子, 则所有互异初等因子指数最高者之积就是 $d_n(\lambda)$ 。

例 3 求矩阵

$$A = \begin{pmatrix} a & & & \\ 1 & a & & \\ & w & w & \\ & & w & w \\ & & & 1 & a \end{pmatrix}_{n \times n}$$

的最小多项式。

解 因为 $f(\lambda) = |E - A| = (\lambda - a)^n = D_n(\lambda)$ 而 $E - A$ 的 $n - 1$ 级行列式 $D_{n-1}(\lambda) = 1$, 所以 $m(\lambda) = (\lambda - a)^n$ 。

例 4 求矩阵

$$\begin{pmatrix} 2 & 1 & 0 \\ -4 & -2 & 0 \\ 2 & 1 & 0 \end{pmatrix}$$

的最小多项式。

解 把 A 的特征矩阵化为标准形

$$(\lambda E - A) = \begin{pmatrix} \lambda - 2 & -1 & 0 \\ 4 & \lambda + 2 & 0 \\ -2 & -1 & \lambda \end{pmatrix} \sim \begin{pmatrix} 1 & & \\ & \lambda^2 & \\ & & \lambda^2 \end{pmatrix}$$

所以以 A 的最小多项式 $m(\lambda) = \lambda^2$ 。

例 5 设

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

证明 当 $n \geq 3$ 时 $A^n = A^{n-2} + A^2 - E$, 并利用这个关系式计算 A^{100} 。

证明 因为

$$(\lambda E - A) = \begin{pmatrix} \lambda - 1 & 0 & 0 \\ -1 & \lambda & 1 \\ 0 & -1 & \lambda \end{pmatrix} \sim \begin{pmatrix} 1 & & \\ & \lambda^2 - 1 & \\ & & (\lambda - 1)^2(\lambda + 1) \end{pmatrix}$$

所以最小多项式是 $m(\lambda) = (\lambda^2 - 1)(\lambda + 1)$, 即

$$A^3 - A^2 - A + E = 0$$

所以当 $n = 3$ 时 $A^3 = A + A^2 - E$

假设 $n = k$ 时 $A^k = A^{k-2} + A^2 - E$

对于 $n = k + 1$

$$\begin{aligned} A^{k+1} &= A^{k-1} + A^3 - A \\ &= A^{k-1} + A + A^2 - E - A \\ &= A^{k-1} + A^2 - E \end{aligned}$$

结论成立, 故当 $n \geq 3$ 时, $A^n = A^{n-2} + A^2 - E$, 再计算 A^{100}

$$\begin{aligned} A^{100} &= A^{98} + A^2 - E = A^{96} + A^2 - E + A^2 - E \\ &= A^{96} + 2(A^2 - E) = A^{94} + 3(A^2 - E) \end{aligned}$$

$$\begin{aligned} &= \dots = A^2 + 49(A^2 - E) \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 50 & 1 & 0 \\ 50 & 0 & 1 \end{pmatrix} \end{aligned}$$

定理 3.4.5 n 阶复方阵 A 相似于对角矩阵的必要充分条件是 A 的最小多项式 $m(\lambda)$ 无重根。

证明 因为 A 相似于对角矩阵的必要充分条件是 A 的初等因式全是一次的, 而 A 的初等因子全为一次式的必要充分条件是 A 的最后一个不变因子 $d_n(\lambda)$ 全是不同一次因子乘积。

例 6 证明非零的幂零矩阵 A (即 $A^k = 0$) 不能与对角矩阵相似。

证明 因为 $A^k = 0$, 所以 A 的零化多项式是 λ^k , 那么 A 的最小多项式必是 λ^l , 且 $l > 1$, 这是因为 A 是非 0 矩阵。说明最小多项式 $m(\lambda) = \lambda^l$ 有重根 0。故 A 不能与对角矩阵相似。

例 7 设 $A^2 = A$, 且 $\text{rank}(A) = r$, 证明 $|A + E| = 2^r$ 。

证明 因为 $A^2 = A$, 所以 A 的最小多项式 $m(\lambda) = \lambda^2 - \lambda$ 没有重根, 故 A 能相似于对角矩阵。

且 A 只有 0 和 1 作其特征值, 由 $\text{rank}(A) = r$, 故

$$A \sim \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} =$$

即存在可逆阵 P , 使 $P^{-1}AP =$
所以

$$\begin{aligned} |A + E| &= |P| |P^{-1}AP + P^{-1}P| |P^{-1}| \\ &= |I + E| = 2^r \end{aligned}$$

3.5 正交变换、酉变换

作为本章的结束, 还介绍在欧氏空间中(酉空间中)很重要的一类变换——正交变换(酉变换)。在通常的几何空间中的旋转、镜面反射就是这一类变换。

定义 3.21 设 A 是欧氏空间(酉空间) V 上的线性变换, 如果对任意向量 $\alpha, \beta \in V$, 恒有

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

则称 A 是 V 的一个正交变换(酉变换)

正交变换即保持模长不变。

例 1 设 $\alpha_1 = (1\ 0)$, $\alpha_2 = (0\ 1)$, 线性变换 A 在 α_1, α_2 之下的矩阵表示为

$$\begin{pmatrix} \cos Q & -\sin Q \\ \sin Q & \cos Q \end{pmatrix}$$

对任意 $\alpha \in R^2$, 设 $\alpha = (x\ y)$, $A(\alpha) = (x'\ y')$, 则

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos Q & -\sin Q \\ \sin Q & \cos Q \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

所以

$$\begin{aligned} (A \quad A) &= (A)^T (A) \\ &= (x \ y) \begin{pmatrix} \cos Q & \sin Q \\ -\sin Q & \cos Q \end{pmatrix} \begin{pmatrix} \cos Q & -\sin Q \\ \sin Q & \cos Q \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= (x \ y) \begin{pmatrix} x \\ y \end{pmatrix} = (\quad , \quad) \end{aligned}$$

所以 A 是 R^2 上的正交变换

下面讨论正交变换(酉变换)的性质:

定义 n 阶复方阵 A, 若 $\overline{A}^T = A^{-1}$, 称 A 为酉阵, 若 $\overline{A}^T = A$ 称为 Hermite(埃米特) 矩阵。

定理 3.5.1 设 A 是 n 维欧氏空间(酉空间) V 的线性变换, 则下列提法是等价的:

- A 是正交变换(酉变换);
- 对任意向量 $\alpha, \beta \in V$, 恒有 $(A\alpha, A\beta) = (\alpha, \beta)$;
- 若 $\alpha_1, \dots, \alpha_n$ 是 V 的一个标准正交基, 则 $A\alpha_1, \dots, A\alpha_n$ 也是 V 的一个标准正交基;
- A 在 V 的任意一组标准正交基下的矩阵表示是正交矩阵(酉阵)。

证明 由 推导 。因为 A 是 V 上的正交变换, 由定义对任意 $\alpha, \beta \in V$

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

由于 $(A\alpha, A\alpha) = (\alpha, \alpha)$ $(A\beta, A\beta) = (\beta, \beta)$

所以 $(A\alpha, A\beta) = (\alpha, \beta)$

的提法即保持内积不变。

注意: 若 A 是酉空间 V 上的酉变换, 对任意 $t \in C$, 因为

$$(A(\alpha + t\beta), A(\alpha + t\beta)) = (\alpha + t\beta, \alpha + t\beta)$$

按照酉空间内积, 可化简得

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

取 $t=1$ 知道 $(A\alpha, A\beta)$ 与 (α, β) 的实部相同,

取 $t=i$ 知道 $(A\alpha, A\beta)$ 与 (α, β) 的虚部相同,

故也有

$$(A\alpha \quad A\beta) = (\alpha, \beta)$$

由 推导 因为对任意的 $i, j, (i=1, 2, \dots, j=1, 2, \dots, n)$

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

这就证明了 $A\alpha_1, \dots, A\alpha_n$ 也是一个标准正交基。

推导 设 $\alpha_1, \dots, \alpha_n$ 是 V 的任意一个标准正交基, V 上的线性变换 A 在此基下的矩阵表示为 A, 即

$$A(\alpha_1, \dots, \alpha_n) = (\alpha_1, \dots, \alpha_n) \begin{pmatrix} a_{11} & a_{12} \dots a_{1n} \\ a_{21} & a_{22} \dots a_{2n} \\ \dots & \dots \\ a_{n1} & a_{n2} \dots a_{nn} \end{pmatrix}$$

令 $A = (\alpha_1, \dots, \alpha_n)$, 其中 $\alpha_i^T = (a_{1i}, a_{2i}, \dots, a_{ni})$ ($i = 1, 2, \dots, n$)。注意到 α_i 恰是 A 在标准正交基 β_1, \dots, β_n 之下的坐标, 由于 β_1, \dots, β_n 也是标准正交基, 由式(2.6)得

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

所以 A 是正交阵(酉阵, 因为 $(\alpha_i, \alpha_j) = \overline{(\alpha_j, \alpha_i)}$)。

推导 设 β_1, \dots, β_n 是 V 的一个标准正交基, 因为 A 在 β_1, \dots, β_n 之下的矩阵表示是

正交阵 A , 那么对任意 $\alpha \in V$, 设 $\alpha = (\alpha_1, \dots, \alpha_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, 则

$$A \alpha = (\alpha_1, \dots, \alpha_n) A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

所以, 由第 2 章的式(2.6)

$$\begin{aligned} (A \alpha, A \alpha) &= \left(A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right)^T \left(A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right) \\ &= \sum_{i=1}^n x_i^2 = (\alpha, \alpha) \end{aligned}$$

这就证明了 A 是正交变换。

推论 正交变换保持向量间夹角不变。

1. A 是 n 阶正交矩阵(酉阵)的必要充分条件为 $A^T A = E$ 即 $A^T = A^{-1}$ ($A^H A = E$, 即 $A^H = A^{-1}$) 也即 A 的行向量组和列向量组分别都作成 R^n (C^n) 的标准正交基。

2. A 是 n 阶正交矩阵(酉阵), 则 A^T, A^*, A^{-1} 皆为正交矩阵(酉阵)。若 B 也是正交矩阵(酉阵)则 AB 也是正交矩阵(酉阵)。

3. 正交矩阵(酉阵)的特征值的模为 1。

4. 正交矩阵(酉阵) A 的行列式的值 $|A| = \pm 1$ 。

证明 3 设 λ 是酉阵 A 的特征值, 则 $|\lambda| = 1$, 设 α 是 λ 对应的特征向量, 则

$$A^{-1} A \alpha = A^{-1} \lambda \alpha = \lambda A^{-1} \alpha$$

所以 $A^{-1} \alpha = \frac{1}{\lambda} \alpha$

再由等式 $A \alpha = \lambda \alpha$ 两边取共轭转置, 得 $\overline{\lambda} \alpha^H = \alpha^H A^H$

再以 α 右乘等式两端, 得

$$\overline{\lambda} \alpha^H \alpha = \alpha^H A^H \alpha = \alpha^H 1 \alpha$$

$$(\overline{\lambda} - 1) \alpha^H \alpha = 0$$

因为 $\alpha \neq 0$, 所以 $\alpha^H \alpha > 0$, 于是

$$\overline{\lambda} - 1 = 0$$

所以 $\overline{\quad} = 1$, 即 $|\quad| = 1$

习 题 3

1. 判别下面所定义的变换, 哪些是线性变换, 哪些不是?
- (1) 在线性空间 V 中, $A \quad = \quad + \quad$, 其中 $\quad V$ 是一个固定的向量;
 - (2) 在线性空间 V 中, $A \quad = \quad$, 其中 $\quad V$ 是一个固定的向量;
 - (3) 在 R^3 中, $A(x_1, x_2, x_3) = (x_1^2, x_2 + x_3, x_3^2)$;
 - (4) 在 R^3 中, $A(x_1, x_2, x_3) = (2x_1 - x_2, x_2 + x_3, x_1)$;
 - (5) 在 $K[x]$ 中, $A f(x) = f(x + 1)$;
 - (6) 在 $K[x]$ 中, $A f(x) = f(x_0)$, 其中 $x_0 \in K$ 是一固定的数;
 - (7) 把复数域看做复数域上的线性空间, $A \quad = \quad$;
 - (8) 在 $R^{n \times n}$ 中, $A(X) = BXC$, 其中 $B, C \in R^{n \times n}$ 是两个固定的矩阵。
2. 设 $\alpha_1, \dots, \alpha_n$ 是线性空间 V 的一个基, A 是 V 上的线性变换, 证明 A 可逆当且仅当 $A \alpha_1, \dots, A \alpha_n$ 线性无关。

3. 找下列线性变换在所指定基下的矩阵。
- (1) 第 1 题中(4)的变换 A 在基 $\alpha_1 = (1, 0, 0), \alpha_2 = (0, 1, 0), \alpha_3 = (0, 0, 1)$ 下的矩阵;
 - (2) 在 $K[x]_n$ 中, 设变换 $A f(x) = f(x + 1) - f(x)$, A 在基 $\alpha_0 = 1, \alpha_1 = x, \alpha_2 = \frac{x(x - 1)}{2!}, \dots, \alpha_n = \frac{x(x - 1) \dots (x - n + 1)}{n!}$

下的矩阵;

(3) 6 个函数

$$\begin{aligned} \alpha_1 &= e^{ax} \cos bx, & \alpha_2 &= e^{ax} \sin bx \\ \alpha_3 &= x e^{ax} \cos bx, & \alpha_4 &= x e^{ax} \sin bx \\ \alpha_5 &= \frac{1}{2} x^2 e^{ax} \cos bx, & \alpha_6 &= \frac{1}{2} x^2 e^{ax} \sin bx \end{aligned}$$

的所有实系数线性组合构成实数域上一个六维线性空间, 求微分变换 D 在 $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6$ 下的矩阵

(4) 已知 R^3 中线性变换 A 在基 $\alpha_1 = (-1, 1, 1), \alpha_2 = (1, 0, -1), \alpha_3 = (0, 1, 1)$ 下的矩阵是

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

求 A 在基 $\alpha_1 = (1, 0, 0), \alpha_2 = (0, 1, 0), \alpha_3 = (0, 0, 1)$ 之下的矩阵。

4. 设 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 是四维线性空间 V 的一个基, 已知线性变换 A 在这组基下的矩阵为
- $$\begin{pmatrix} 1 & 0 & 2 & 1 \\ -1 & 2 & 1 & 3 \\ 1 & 2 & 5 & 5 \\ 2 & -2 & 1 & -2 \end{pmatrix}$$

- (1) 求 A 在基 $\alpha_1 = \alpha_1 - 2\alpha_2 + \alpha_4, \alpha_2 = 3\alpha_2 - \alpha_3 - \alpha_4, \alpha_3 = \alpha_3 + \alpha_4, \alpha_4 = 2\alpha_4$ 下的矩阵;
- (2) 求 $R(A)$ 及 $N(A)$;
- (3) 在 $N(A)$ 中选一组基, 把它扩为 V 的一个基, 并求 A 在这组基下的矩阵;
- (4) 在 $R(A)$ 中选一组基, 把它扩为 V 的一个基, 并求 A 在这组基下的矩阵。
5. 求复数域上线性空间 V 的线性变换 A 的特征向量与特征值, 已知 A 在一组基下的矩阵为:

(1) $A = \begin{bmatrix} 3 & 4 \\ 5 & 2 \end{bmatrix}$ (2) $A = \begin{bmatrix} 3 & 1 & 0 \\ -4 & -1 & 0 \\ 4 & -8 & -2 \end{bmatrix}$

(3) $A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$

6. 在上题中哪些变换的矩阵可以在适当的基下变成对角形? 在可以化成对角形的情况下, 写出相应基变换的过渡阵 P 。

7. 设 A 是线性空间 V 上的可逆线性变换。
- (1) 证明: A 的特征值一定不为 0;
- (2) 证明: 若 λ 是 A 的特征值, 则 $\frac{1}{\lambda}$ 是 A^{-1} 的特征值。
8. 设 λ_1, λ_2 是线性变换 A 的两个不同特征值, α_1, α_2 是分别属于 λ_1, λ_2 的特征向量。证明: $\alpha_1 + \alpha_2$ 不是 A 的特征向量。
9. A 是数域 K 上 n 维线性空间 V 的线性变换, 且 $A^2 = A$
- 证明: (1) $N(A) = \{ \alpha - A\alpha \mid \alpha \in V \}$
- (2) $V = N(A) \oplus R(A)$
10. 设 A, B 皆是 n 维线性空间 V 的线性变换, 且 $AB = BA$ 。若 A 有 n 个互异特征值, $\lambda_1, \dots, \lambda_n$ 。证明: B 一定有对角矩阵表示。

11. 化下列 n -矩阵成标准形:

(1) $\begin{bmatrix} \lambda^2 + 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & (\lambda + 1)^2 \end{bmatrix}$ (2) $\begin{bmatrix} 0 & 0 & 0 & \lambda^2 \\ 0 & 0 & \lambda^2 - 1 & 0 \\ 0 & (\lambda - 1)^2 & 0 & 0 \\ \lambda^2 - 1 & 0 & 0 & 0 \end{bmatrix}$

(3) $\begin{bmatrix} 2 & 3 & 0 & 1 \\ 4 & 3 + 6 & 0 & +2 & 2 \\ 0 & 6 & 2 & 0 \\ -1 & 0 & -1 & 0 & 0 \\ 3 & -3 & 1 & 2 & -2 & 0 & 0 \end{bmatrix}$

12. 证明: n 阶 λ -矩阵

$$A(\lambda) = \begin{pmatrix} & 0 & 0 & \cdots & 0 & a_n \\ -1 & & 0 & \cdots & 0 & a_{n-1} \\ & & & \cdots & & \\ 0 & 0 & 0 & \cdots & & a_2 \\ 0 & 0 & 0 & \cdots & -1 & +a_1 \end{pmatrix}$$

$n-1$ 个

的不变因子是 $1, 1, \dots, 1, f(\lambda)$, 其中 $f(\lambda) = \lambda^n + a_1 \lambda^{n-1} + \dots + a_{n-1} \lambda + a_n$ 。

13. 设 A 是数域 K 上 n 阶方阵, 证明: A 与 A^T 相似。

14. 设 $A = \begin{pmatrix} & 0 & 0 \\ 1 & & 0 \\ 0 & 1 & \end{pmatrix}$, 求 A^k 。

15. 求下列矩阵的若当标准形 J 。

$$A = \begin{pmatrix} 3 & 7 & -3 \\ -2 & -5 & 2 \\ -4 & -10 & 3 \end{pmatrix} \qquad B = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & -1 & 1 \\ 0 & 0 & 0 & 2 & 2 & -1 \\ 0 & 0 & 0 & 1 & 2 & -1 \end{pmatrix}$$
$$C = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ & & & \cdots & & \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

16. 设

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

求 A 的若当标准形 J , 并求相似变换阵 P , 使 $P^{-1}AP = J$ 。

17. 设 A 是 3×3 矩阵, 它的特征值为 $\lambda_1 = 2, \lambda_2 = 1, \lambda_3 = 3$ 相应的特征向量为

$$\alpha_1 = (1, 2, 2)^T, \alpha_2 = (2, -2, 1)^T, \alpha_3 = (-2, -1, 2)^T$$

试求 A 。

18. 设 $A = \begin{bmatrix} 1 & -1 \\ 2 & 5 \end{bmatrix}$

证明: $2A^4 - 12A^3 + 19A^2 - 29A + 37E$ 可逆, 并将其逆矩阵表示为 A 的多项式。

19. 设 α 是欧氏空间 V 中的单位向量, $\alpha \perp V$, 定义线性变换

$$A(\alpha) = -2(\alpha, \alpha)\alpha$$

证明 A 是正交变换, 常称这变换为镜面反射。

20. 设 T 是 R^3 中线性变换, $T(x, y, z) = (y, z, x)$, 证明 T 是正交变换。

21. n 阶方阵 A 是幂么矩阵($A^k = E$), 证明 A 与对角矩阵相似。

22.
$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & -1 & -1 & 0 \\ 1 & 1 & 1 & 0 \\ 2 & 2 & 2 & 0 \end{pmatrix}$$

计算 A 的最小多项式, 并计算 A^{100} 。

23. 设

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

计算 $2A^8 - 3A^5 + A^4 + A^2 - 4E$ 。

24. 若 R^4 上线性变换 A 关于 R^4 的标准正交基 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 的矩阵表示为

$$A = \begin{pmatrix} 5 & -1 & -2 & 0 \\ -1 & 5 & 0 & -2 \\ -2 & 0 & 5 & -1 \\ 0 & -2 & -1 & 5 \end{pmatrix}$$

现有正交矩阵

$$Q = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

把标准正交基 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 变为 $\beta_i = (\alpha_1, \alpha_2, \alpha_3, \alpha_4) Q^{-1} \quad i=1,2,3,4$ 其中 $Q = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$, 证明 A 关于后一基的矩阵表示为一对角阵。并求 A 的特征值、特征向量, 也求出 A 的特征值、特征向量。

25. 设 $A \in C^{n \times n}$, 证明: 存在酉阵 U , 使 $U^H A U = T$ (T 为上三角阵)。
26. 设 $A \in C^{n \times n}$, 证明: 存在酉阵 U , 使 $U^H A U = \Lambda$ (Λ 为对角矩阵) 的必要充分条件是 $A^H A = A A^H$ (这样的 A 叫正规矩阵)。

第 2 篇

矩阵理论及其应用

在工程技术中引进矩阵理论, 不仅使表达极为简捷, 而且对工程技术理论的实质刻画也极为深刻。例如系统工程、优化方法、稳定性理论等, 无不与矩阵理论发生紧密结合。因此, 矩阵的理论与方法已成为研究现代工程技术的数学基础。

第 4 章

范数理论及其应用

在计算数学中, 特别在数值代数中, 研究数值方法的收敛性稳定性及误差分析等问题, 范数理论显得十分重要, 本书以后各章讨论中也要经常使用范数。本章论述范数的理论及性质。对于向量范数, 着重讨论常用的 n 维向量空间 C^n 的情况, 至于矩阵范数, 将着重讨论矩阵空间 $C^{n \times n}$ 的情况。

4.1 向量范数及其性质

4.1.1 向量范数的概念及 P-范数

设给定了 n 维向量空间 R^n 中的向量序列 $\{x^{(k)}\}$, 其中 $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})$, $k = 1, 2, 3, \dots$ 。如果每一个分量 $x_i^{(k)}$ 当 $k \rightarrow \infty$ 时都有极限 x_i , 即

$$\lim_{k \rightarrow \infty} x_i^{(k)} = x_i, i = 1, 2, \dots, n$$

记 $x = (x_1, x_2, \dots, x_n)$, 则称向量序列 $\{x^{(k)}\}$ 有极限 x , 或称 $x^{(k)}$ 收敛于 x , 简称 $\{x^{(k)}\}$ 收敛, 记为

$$\lim_{k \rightarrow \infty} x^{(k)} = x, \text{ 或 } x^{(k)} \rightarrow x$$

不收敛的向量序列称为是发散的。例如向量序列

$$x^{(k)} = \begin{pmatrix} \frac{1}{2^k} \\ \frac{\sin k}{k} \end{pmatrix}, \quad k = 1, 2, 3, \dots$$

是收敛的。因为当 $k \rightarrow \infty$ 时, $\frac{1}{2^k} \rightarrow 0, \frac{\sin k}{k} \rightarrow 0$, 所以

$$\lim_{k \rightarrow \infty} x^{(k)} = \begin{pmatrix} \lim_{k \rightarrow \infty} \frac{1}{2^k} \\ \lim_{k \rightarrow \infty} \frac{\sin k}{k} \end{pmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

而向量序列

$$^{(k)} = \begin{pmatrix} \frac{1}{2^1} \\ \vdots \\ \frac{1}{2^k} \\ \vdots \\ \frac{1}{2^k} \end{pmatrix}, \quad k = 1, 2, 3, \dots$$

是发散的。因为 $\sum_{i=1}^k \frac{1}{2^i} = \frac{1}{2} \cdot \frac{1 - (\frac{1}{2})^{k+1}}{1 - \frac{1}{2}} \rightarrow 1$, 而 $\sum_{i=1}^k \frac{1}{i} \rightarrow \infty$ 。

显然, 如果向量序列 $\{^{(k)}\}$ 收敛到向量 α , 则向量序列 $\{^{(k)} - \alpha\} = \{\alpha_1^{(k)} - \alpha_1, \alpha_2^{(k)} - \alpha_2, \dots, \alpha_n^{(k)} - \alpha_n\}$ 一定收敛到零向量 $(0, 0, \dots, 0)$, 反之亦然。于是当 $\lim_{k \rightarrow \infty} ^{(k)} = \alpha$ 时, 则向量 $(\alpha_1^{(k)} - \alpha_1, \dots, \alpha_n^{(k)} - \alpha_n)$ 的欧氏长度 $|^{(k)} - \alpha| = \sqrt{(\alpha_1^{(k)} - \alpha_1)^2 + \dots + (\alpha_n^{(k)} - \alpha_n)^2}$ 收敛于零; 反之若有一向量序列的欧氏长度收敛于零, 则它的每一个分量一定收敛于零, 从而该向量序列收敛于零向量。

由上述可见, 向量的长度可用来刻画收敛的性质。但是对一般的线性空间 V , 如何定义向量的长度呢? 这就是所谓范数的概念。范数是比长度更为广泛的概念, 现定义于下。

定义 4.1 如果 V 是数域 K 上的线性空间, 且对于 V 的任一向量 α , 对应一个实值函数 $|\alpha|$, 它满足以下三个条件。

- 1. 非负性 当 $\alpha \neq 0$ 时 $|\alpha| > 0$; 当 $\alpha = 0$ 时 $|\alpha| = 0$;
- 2. 齐次性 $|a\alpha| = |a| |\alpha|$, $a \in K, \alpha \in V$;
- 3. 三角不等式 $|\alpha + \beta| \leq |\alpha| + |\beta|$, $\alpha, \beta \in V$

则称 $|\alpha|$ 为 V 上的范数 (norm)。

例 1 在 n 维复空间 C^n 上, 复向量 $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ 的长度

$$|\alpha| = \sqrt{|\alpha_1|^2 + |\alpha_2|^2 + \dots + |\alpha_n|^2}$$

就是一种范数。

为了说明这里的 $|\alpha|$ 是范数, 只需验证它满足范数三个条件就行了。事实上有

- 1. 对于 $|\alpha| = \sqrt{|\alpha_1|^2 + \dots + |\alpha_n|^2}$, 当 $\alpha \neq 0$ 时, 显然 $|\alpha| > 0$; 当 $\alpha = 0$ 时, 则 $|\alpha| = \sqrt{0^2 + \dots + 0^2} = 0$ 。
- 2. 对任意的复数 a , 因为

$$a\alpha = (a\alpha_1, a\alpha_2, \dots, a\alpha_n)$$

所以

$$\begin{aligned} |a\alpha| &= \sqrt{|a\alpha_1|^2 + |a\alpha_2|^2 + \dots + |a\alpha_n|^2} \\ &= |a| \sqrt{|\alpha_1|^2 + |\alpha_2|^2 + \dots + |\alpha_n|^2} \\ &= |a| |\alpha| \end{aligned}$$

- 3. 对于任意两个复向量 $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$, $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ 有

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

所以

$$\begin{aligned} |\alpha + \beta| &= \sqrt{|\alpha_1 + \beta_1|^2 + |\alpha_2 + \beta_2|^2 + \dots + |\alpha_n + \beta_n|^2} \\ &= \sqrt{|\alpha_1|^2 + |\alpha_2|^2 + \dots + |\alpha_n|^2 + |\beta_1|^2 + |\beta_2|^2 + \dots + |\beta_n|^2} \end{aligned}$$

$$\begin{aligned} &= \sqrt{|x_1|^2 + |x_2|^2 + \dots + |x_n|^2} \\ + \quad &^2 = |x_1 + y_1|^2 + |x_2 + y_2|^2 + \dots + |x_n + y_n|^2 \\ &= |x_1|^2 + \dots + |x_n|^2 + \sum_{i=1}^n \overline{x_i} y_i + \sum_{i=1}^n x_i \overline{y_i} + |y_1|^2 + \dots + |y_n|^2 \\ &= \quad^2 + 2\operatorname{Re}(\quad, \quad) + \quad^2 \\ &\quad^2 + 2 \quad + \quad^2 \\ &= (\quad + \quad)^2 \end{aligned}$$

这就证明了 $\quad = \sqrt{|x_1|^2 + \dots + |x_n|^2}$ 是 C^n 上的一种范数。通常称这种范数为 2-范数或欧氏范数), 记为 \quad_2 , (或 \quad_E)。即

$$\quad_2 = \sqrt{|x_1|^2 + |x_2|^2 + \dots + |x_n|^2} \tag{4.1}$$

式(4.1)对欧氏空间 R^n 亦成立, 这只要把复数域 C 改为实数域就行了。

可以证明 \quad_2 还满足不等式。

$$\quad - \quad \quad - \tag{4.2}$$

这里 \quad, \quad 是 C^n 中任意向量。

事实上, 因为有

$$= \quad - \quad + \quad + \quad X -$$

所以

$$- \quad -$$

同理可得

$$- \quad X \quad -$$

联合上面两个不等式便得式(4.2)。

如果用 $-$ 来换式(4.2)中的 \quad , 不等式(4.2)变成

$$\quad - \quad \quad + \tag{4.3}$$

不等式(4.2), 式(4.3)在 R^2 中有明确的几何意义, 即它们表示任一三角形两边长度之差不大于第三边的长度(图 4.1)。

向量 \quad 和 \quad 的差 $\quad - \quad$, 其范数就是 \quad 和 \quad 两个终点的距离。而 \quad 和 \quad 本身也表示 \quad 和 \quad 的终点到它们的始点(原点)的距离, 所以也可以用距离来解释范数。

例 2 证明 $\quad = \max_i |x_i|$ 是 C^n 上的一种范数, 这里 $\quad = (x_1, x_2, \dots, x_n) \in C^n$

证 当 $\quad = 0$ 时, 有 $\quad = \max |x_i| > 0$; 当 $\quad = 0$ 时, 显然 $\quad = 0$ 。

又对任意 $a \in C$, 有

$$a \quad = \max |a x_i| = |a| \max |x_i| = |a| \quad$$

对 C^n 的任意两个向量 $\quad = (x_1, x_2, \dots, x_n), \quad = (y_1, y_2, \dots, y_n)$ 有

$$\begin{aligned} + \quad &= \max_i |x_i + y_i| \leq \max_i (|x_i| + |y_i|) \\ &= \max_i |x_i| + \max_i |y_i| = \quad + \quad \end{aligned}$$

以上所论表明 $\quad = \max |x_i|$ 确实是 C^n 上的一种范数, 通常称其为 \quad -范数, 记为 \quad 。于是有

$$\quad = \max_i |x_i| \tag{4.4}$$

例 3 证明 $\quad = \sum_{i=1}^n |x_i|$ 也是 C^n 上的一种范数, 其中 $\quad = (x_1, x_2, \dots, x_n) \in C^n$ 。

证 当 $\alpha = 0$ 时, 显然 $\|\alpha x\| = 0$, 当 $\alpha \neq 0$ 时, 由于 x 的每一个分量都是零, 故 $\|\alpha x\| = 0$ 。

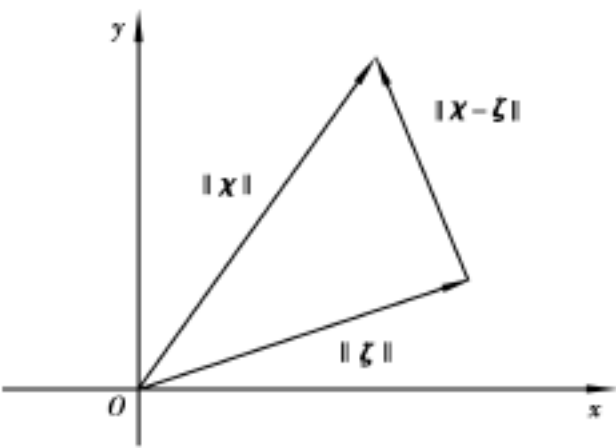


图 4.1

又对于任意 $a \in C$, 有

$$\|a\|_1 = \sum_{i=1}^n |a_i| = \sum_{i=1}^n |a_i| = \|a\|_1$$

对于任意两个向量 $x, y \in C^n$, 有

$$\begin{aligned} \|x+y\|_1 &= \sum_{i=1}^n |x_i + y_i| \leq \sum_{i=1}^n (|x_i| + |y_i|) \\ &= \sum_{i=1}^n |x_i| + \sum_{i=1}^n |y_i| = \|x\|_1 + \|y\|_1 \end{aligned}$$

于是由定义 4.1 知 $\|x\|_1 = \sum_{i=1}^n |x_i|$ 是 C^n 上一种范数。通常称其为 1-范数, 即

$$\|x\|_1 = \sum_{i=1}^n |x_i| \tag{4.5}$$

为了易于理解, 用度量 R^2 中两点 P, Q 的距离(长度)的大小对例 1, 例 2, 例 3 中的三种范数加以说明(图 4.2)。

为了度量 P, Q 的距离, 除了使用欧氏长度 $\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$ 来度量外, 当然还可以使用 PR 及 RQ 中最长一边的长度

$$\max\{|x_1 - y_1|, |x_2 - y_2|\}$$

来度量它, 或者以 PR 和 RQ 两边长度之和

$$|x_1 - y_1| + |x_2 - y_2|$$

去度量。这三种长度相当于平面情况的向量范数 $\|x\|_\infty$, $\|x\|_1$ 和 $\|x\|_2$ 。

由例 1 ~3 可知, 在一个线性空间中, 可以定义多种向量范数, 实际上可以定义无限多种范数。例如对于不小于 1 的任意实数 p 及 $x = (x_1, x_2, \dots, x_n) \in C^n$ 可以证明实值函数

$$\left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}, \quad 1 \leq p < +\infty$$

满足定义 4.1 的三个条件。事实上, 该函数显然具有非负性及齐次性, 下证明它还满足三角不等式, 即要证明下面不等式

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

成立, 其中 $x = (x_1, x_2, \dots, x_n) \in \mathbb{C}^n$ 。

因为

$$\sum_{i=1}^n |x_i + y_i|^p = \sum_{i=1}^n |x_i + y_i|^{p-1} |x_i + y_i| = \sum_{i=1}^n |x_i + y_i|^{p-1} |x_i| + \sum_{i=1}^n |x_i + y_i|^{p-1} |y_i|$$

再对它使用下面的 Holder 不等式

$$\begin{aligned} \sum_{i=1}^n |x_i + y_i| &\leq \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^n |y_i|^q\right)^{\frac{1}{q}} \\ \frac{1}{p} + \frac{1}{q} &= 1, \quad p, q > 1 \end{aligned} \tag{4.6}$$

于是便有:

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

两端同除以 $\left(\sum_{i=1}^n |x_i + y_i|^{(p-1) \cdot \frac{p}{p-1}}\right)^{\frac{p-1}{p}}$ 便得所要不等式。

称函数 $\left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}$ 为向量 x 的 p -范数, 记为 $\|x\|_p$, 于是有

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}} \tag{4.7}$$

在式(4.7)中, 让 $p=1$ 便得 $\|x\|_1$; $p=2$ 便得 $\|x\|_2$ 。并且还有

$$\|x\|_p = \lim_{p \rightarrow \infty} \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}$$

事实上, 如果 $|x_1|, |x_2|, \dots, |x_n|$ 中最大的一个 $|x_k| > 0$, 那么有

$$\|x\|_p = \max_i |x_i| = |x_k|$$

于是

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

因为

$$\sum_{i=1}^n \left|\frac{x_i}{x_k}\right|^p \leq n \left|\frac{x_k}{x_k}\right|^p = n$$

所以有

$$\|x\|_p \leq |x_k| n^{\frac{1}{p}}$$

令 $p \rightarrow \infty$ 知

$$\lim_{p \rightarrow \infty} n^{\frac{1}{p}} = 1, \text{ 从而有 } \lim_{p \rightarrow \infty} \|x\|_p = |x_k| = \|x\|_\infty$$

式(4.7)表明它为计算数学中常用的三种范数 $\|x\|_1, \|x\|_2, \|x\|_\infty$ 的统一表达式。下

面再举一些例子, 以示范数求法。

例 4 在区间[a b]上定义的实连续函数集合, 关于通常函数的加法及实数与函数的乘法构成 R 上一线性空间。可以验证

$$\begin{aligned} \|f(t)\|_1 &= \int_a^b |f(t)| dt; \\ \|f(t)\|_p &= \left[\int_a^b |f(t)|^p dt \right]^{\frac{1}{p}}, 1 \leq p < +\infty; \\ \|f(t)\|_\infty &= \max_{t \in [a,b]} |f(t)|. \end{aligned}$$

都满足范数定义的三个条件, 因此, 它们都分别是该线性空间上的范数。

例 5 计算 C^4 中向量 $x = (3i, 0, -4i, -12)$ 的 1-范数, 2-范数及 ∞ -范数, 这里 $i = \sqrt{-1}$ 。

解

$$\begin{aligned} \|x\|_1 &= \sum_{i=1}^n |x_i| = |3i| + |0| + |-4i| + |-12| = 19 \\ \|x\|_2 &= \left(|3i|^2 + |0|^2 + |-4i|^2 + |-12|^2 \right)^{\frac{1}{2}} = 13 \\ \|x\|_\infty &= \max_i |x_i| = 12 \end{aligned}$$

由例 5 可见, 一个向量的范数, 依其范数定义不同而大小不同。

4.1.2 n 维线性空间 V 上的向量范数等价性

前面已经指出, 在数域 K 上的线性空间 V, 特别是 C^n 上可以定义各种各样的向量范数, 其数值大小一般不同如例 5 所示。但是在各种向量范数之间存在有下述定理所论的重要关系。

定理 4.1.1 设 $\|\cdot\|_1$ 和 $\|\cdot\|_2$ 为有限维线性空间 V 的任意两种向量范数, 它们不限于 P-范数, 则总存在两个与向量 x 无关的正常数 c_1 和 c_2 , 使下面不等式成立

$$c_1 \|x\|_1 \leq \|x\|_2 \leq c_2 \|x\|_1, \forall x \in V \tag{4.8}$$

证明 如果范数 $\|\cdot\|_1$ 和范数 $\|\cdot\|_2$ 都与固定范数譬如 $\|\cdot\|_2$ 满足形式如式(4.8)的关系, c_1, c_2 及 c_1, c_2 使

$$\begin{aligned} c_1 \|x\|_1 &\leq \|x\|_2 \leq c_2 \|x\|_1 \\ c_1 \|x\|_1 &\leq \|x\|_2 \leq c_2 \|x\|_1 \end{aligned}$$

成立, 则显然有 $c_1 c_1 \leq c_2 c_2$ 。令 $c_1 = c_1 c_1, c_2 = c_2 c_2$ 便得不等式(4.8)。因此只要对 $n=2$ 证明不等式(4.8)成立就行了。设 V 是 n 维的, e_1, e_2, \dots, e_n 是它的一个基, 于是 V 中的任意向量 x , 有

$$\begin{aligned} x &= e_1 x_1 + e_2 x_2 + \dots + e_n x_n \\ \|x\|_1 &= |x_1| + |x_2| + \dots + |x_n| \end{aligned}$$

从而 $\|x\|_1$ 可以视为 n 个变量 x_1, x_2, \dots, x_n 的函数, 记为

$$f(x_1, x_2, \dots, x_n) = \|x\|_1$$

易证 $f(x_1, \dots, x_n)$ 是连续函数。事实上, 若令 $y = (y_1, y_2, \dots, y_n)$

$$\begin{aligned} \|f(x_1, \dots, x_n) - f(y_1, \dots, y_n)\| &= \left| \|x\|_1 - \|y\|_1 \right| \\ &= \left| (x_1 - y_1) + (x_2 - y_2) + \dots + (x_n - y_n) \right| \\ &\leq |x_1 - y_1| + |x_2 - y_2| + \dots + |x_n - y_n| \end{aligned}$$

由于 x_1, x_2, \dots, x_n 是固定数, 因此当 x_i 充分接近 x_i 时, (x_1, \dots, x_n) 与 (x_1, \dots, x_n) 充分接近。这就说明了 (x_1, x_2, \dots, x_n) 是连续函数。

根据连续函数性质, 可知在有界闭集

$$|x_1|^2 + |x_2|^2 + \dots + |x_n|^2 = 1$$

上(即欧氏空间 R^n 的单位球上), 函数 (x_1, \dots, x_n) 可达到最大值 c_2 及最小值 c_1 。因为在上面的有界闭集中 x_i 不可能全为零, 所以 $c_1 > 0$ 。记向量

$$x = \frac{1}{\sqrt{2}}e_1 + \frac{2}{\sqrt{2}}e_2 + \dots + \frac{n}{\sqrt{2}}e_n$$

其中 $\sqrt{2} = \sqrt{\sum_{i=1}^n |x_i|^2}$, 则其分量满足

$$\left(\frac{|x_1|}{\sqrt{2}}\right)^2 + \left(\frac{|x_2|}{\sqrt{2}}\right)^2 + \dots + \left(\frac{|x_n|}{\sqrt{2}}\right)^2 = 1$$

因此 x 在上面的单位圆上。从而有

$$0 < c_1 = \left(\frac{1}{\sqrt{2}}, \frac{2}{\sqrt{2}}, \dots, \frac{n}{\sqrt{2}}\right) \leq c_2$$

但 $\frac{1}{\sqrt{2}} = \frac{1}{\sqrt{2}}$, 故

$$c_1 = \frac{1}{\sqrt{2}} \leq c_2$$

即 $\frac{c_1}{\sqrt{2}} \leq \frac{c_2}{\sqrt{2}}$
推论 设 x_1, x_2, \dots, x_n 和 y_1, y_2, \dots, y_n 都是 R^n ($p=1, 2, \dots$), 则下面两个不等式成立

$$\frac{x_1}{\sqrt{2}} \leq \frac{y_1}{\sqrt{n}}$$

上面两式表明, 对某一向量而言, 如果它的某一种范数小(或大), 那么它的另两种范数也小(或大)。

定义 4.2 满足不等式(4.8)的两种范数称为等价的。

于是定理 4.1.1 可叙述为: 有限维线性空间上的不同范数是等价的。

例 6 设 A 是一个 n 阶对称正定矩阵, 列向量 $x \in R^n$ 则函数

$$|x|_A = (x^T A x)^{\frac{1}{2}} \tag{4.9}$$

也是一种向量范数(加权范数或椭圆范数)。

证明 因为 A 正定, 所以当 $x=0$ 时, $|x|_A=0$; 当 $x \neq 0$ 时, $|x|_A > 0$, 即 $|x|_A$ 具有非负性。又对任何数 $a \in R$, 有

$$|ax|_A = \sqrt{(ax)^T A (ax)} = \sqrt{a^2 x^T A x} = |a| |x|_A$$

所以 $|x|_A$ 具有齐次性。关于它的三角不等式可推证如下。

因为 A 正定, 所以存在非奇异阵 P , 使 $P^T A P = E$, 从而 $A = (P^T)^{-1} P^{-1} = B^T B$, 这里 $B = P^{-1}$ 。于是

若 A 是 Hermite 正定阵, $x \in C^n$, 该例也成立。

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)} = \sqrt{(\|B\|_2)^T (\|B\|_2)} = \|B\|_2$$

从而

$$\begin{aligned} \|A+B\|_2 &= \|B\|_2 + \|A\|_2 \\ \|B\|_2 + \|A\|_2 &= \|A+B\|_2 \end{aligned}$$

利用向量范数的等价性容易证明下面的定理。

定理 4.1.2 C^n 中的向量序列 $\{x^{(k)}\} = (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})$, $k = 1, 2, \dots$ 收敛到向量 $x = (x_1, x_2, \dots, x_n)$ 的充要条件是对任一种范数 $\|\cdot\|$, 序列 $\|x^{(k)} - x\|$ 收敛到零。

证明 利用范数的等价性, 只需证明对一种范数成立就行, 为此取 $\|\cdot\| = \|\cdot\|_1$ 。

因为 $\lim_{k \rightarrow \infty} \|x^{(k)} - x\|_1 = 0$
的充分条件为 $\lim_{k \rightarrow \infty} x_i^{(k)} = x_i \quad i = 1, 2, \dots, n$

的充要条件为 $\lim_{k \rightarrow \infty} |x_i^{(k)} - x_i| = 0 \quad i = 1, 2, \dots, n$

的充要条件为 $\lim_{k \rightarrow \infty} \max_i |x_i^{(k)} - x_i| = 0$

也即范数序列 $\|x^{(k)} - x\|$ 收敛于 0。

定理 4.1.2 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收敛问题时, 却表现出明显的一致性。这就是说, 如果向量序列 $\{x^{(k)}\}$ 对某一范数 $\|\cdot\|$ 收敛, 且极限为 x , 则对其他范数这个序列仍然收敛, 并且具有相同的极限 x 。

4.2 矩阵的范数

这一节, 我们将向量范数的概念推广到矩阵空间 $C^{n \times n}$ 上去, 同时还要讨论几种具体的矩阵范数, 以兹应用。

4.2.1 矩阵范数的定义与性质

由于一个 $n \times n$ 的矩阵可以看成是一个拉直了的 $n \times n$ 维向量, 因此可以按定义向量范数的方法来定义矩阵范数, 但矩阵之间还有乘法运算, 因此, 对于 $n \times n$ 矩阵 A , 定义其范数如下。

定义 4.3 设 $A, B \in C^{n \times n}$, $a \in C$, 按某一法则在 $C^{n \times n}$ 上定义一个 A 的实值函数, 记为 $\|A\|$, 它满足以下 4 个条件

1. 非负性 如果 $A \neq 0$, 则 $\|A\| > 0$
如果 $A = 0$, 则 $\|A\| = 0$;
2. 齐次性 对任意的 $a \in C$, $\|aA\| = |a| \|A\|$;
3. 三角不等式 $\|A+B\| \leq \|A\| + \|B\|$
4. 相容性 $\|AB\| \leq \|A\| \|B\|$

则称 $\|A\|$ 为矩阵范数或乘积范数。

对于 $C^{m \times n}$ 中的矩阵 A ($m \neq n$ 时), 只要第 4 条的 $\|AB\| \leq \|A\| \|B\|$ 要求 AB 有意义就行。

例 1 已知 $A \in C^{n \times n}$, $A = (a_{ij})_{n \times n}$, 证明下面两个函数

$$A_{m1} = \max_{i,j} |a_{ij}|, \quad A_m = n \cdot \max_{i,j} |a_{ij}|$$

都是矩阵 A 的范数。

证明 只需验证它们依次满足定义 4.3 中的 4 个条件就行了。事实上,对于函数 A_{m1} 而言,它显然具有非负性与齐次性,现仅就三角不等式与相容性加以验证于下。

$$\begin{aligned} A+B_{m1} &= \max_{i,j=1}^n |a_{ij} + b_{ij}| = \max_{i,j=1}^n (|a_{ij}| + |b_{ij}|) \\ &= \max_{i,j=1}^n |a_{ij}| + \max_{i,j=1}^n |b_{ij}| = A_{m1} + B_{m1} \\ AB_{m1} &= \max_{i=1}^n \left| \sum_{j=1}^n a_{ij}b_{ij} \right| \\ &= \max_{i=1}^n (|a_{i1}| |b_{1j}| + |a_{i2}| |b_{2j}| + \dots + |a_{in}| |b_{nj}|) \\ &= \max_{i=1}^n (|a_{i1}| + |a_{i2}| + \dots + |a_{in}|) \cdot \max_{i=1}^n (|b_{1i}| + |b_{2i}| + \dots + |b_{ni}|) \\ &= \max_{i,j=1}^n |a_{ij}| \cdot \max_{i,j=1}^n |b_{ij}| \\ &= A_{m1} \cdot B_{m1} \end{aligned}$$

从而证实了 A_{m1} 是方阵 A 的一种范数。对于 A_m ,前 3 条都是很容易的,下面只验证第 4 条满足。

$$\begin{aligned} AB_m &= n \max_{i,j} |a_{i1}b_{1j} + \dots + a_{in}b_{nj}| \\ &= n \max_{i,j} (|a_{i1}| |b_{1j}| + \dots + |a_{in}| |b_{nj}|) \\ &= n^2 |a_{i_0k}| |b_{kj_0}| \quad i_0, j_0 \text{ 就是使上式取最大值的 } i_0, j_0 \\ &= n \max_{i,j} |a_{ij}| \cdot n \max_{i,j} |b_{ij}| \end{aligned}$$

这就说明 A_m 是方阵 A 的一种范数。

如同向量范数的情况一样,矩阵范数也是多种多样的。但是,在数值方法中进行某种估计时,遇到的多数情况是:矩阵范数与向量范数混合一起使用,而矩阵经常是作为两个线性空间上的线性映射(变换)出现的。因此考虑一些矩阵范数,应使它们与向量范数联系起来,这就是下面将讨论的矩阵范数与向量范数相容的概念。

定义 4.4 如果任意向量 C^n 及任意 n 级方阵 $A \in C^{n \times n}$,对于给定的向量范数 $\| \cdot \|$ 和矩阵范数 A 满足不等式

$$\|Ax\| \leq A \|x\| \tag{4.10}$$

则称矩阵范数 A 和向量范数 $\| \cdot \|$ 相容。

例 2 设 $A=(a_{ij})_{n \times n} \in C^{n \times n}$,定义函数

$$A_F = \left[\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \right]^{\frac{1}{2}} = (\text{tr } A^T A)^{\frac{1}{2}} \tag{4.11}$$

试验证它是一种矩阵范数。

证明 因为 A_F 可以视为 $n \times n$ 维向量 A 的 2-范数。若记 a_j 为 A 的第 j 列($j=1, 2,$

..., n), 则式(4.11) 可改写为

$$\|A\|_F^2 = (\|A_1\|_2^2 + \|A_2\|_2^2 + \dots + \|A_n\|_2^2)$$

所以 $\|A\|_F$ 具有非负性, 齐次性, 及三角不等式是明显的。

下面先证明 $\|A\|_F$ 与 $\|A\|_2$ 是相容的, 对任意 $X = (x_1, x_2, \dots, x_n)^T \in C^n$ 。然后再证明函数 $\|A\|_F$ 满足定义 4.3 中的第 4 条件。

把矩阵 A 按行分块, $A = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{pmatrix}$, 其中 $A_i = (a_{i1} \ a_{i2} \ \dots \ a_{in})$

则 $A X = \begin{pmatrix} (A_1^T, \overline{}) \\ \vdots \\ (A_n^T, \overline{}) \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$

其中 $(A_i^T, \overline{}) = a_{i1} x_1 + a_{i2} x_2 + \dots + a_{in} x_n$

由 Cauchy-by 不等式, 有

于是

$$\begin{aligned} \|A X\|_2^2 &= \sum_{i=1}^n |(A_i^T, \overline{})|^2 = \sum_{i=1}^n \|A_i\|_2^2 \|\overline{}\|_2^2 \\ &= \sum_{i=1}^n \|A_i\|_2^2 = \|A\|_F^2 \end{aligned}$$

这就证明了式(4.11)

记 B 的第 j 列为 $B_j, j=1, 2, \dots, n, B = (B_1, B_2, \dots, B_n)$ 。

$$\begin{aligned} \|AB\|_F^2 &= \|(A_1, A_2, \dots, A_n) B\|_F^2 \\ &= \|A_1 B_1\|_2^2 + \|A_2 B_1\|_2^2 + \dots + \|A_n B_1\|_2^2 \\ &\quad + \|A_1 B_2\|_2^2 + \|A_2 B_2\|_2^2 + \dots + \|A_n B_2\|_2^2 \\ &= \|A\|_F^2 (\|B_1\|_2^2 + \|B_2\|_2^2 + \dots + \|B_n\|_2^2) \\ &= \|A\|_F^2 \cdot \|B\|_F^2 \end{aligned}$$

这就证明了 $\|AB\|_F = \|A\|_F \cdot \|B\|_F$

范数(4.12) 又称为 Frobenius 范数, 或简称 F-范数。为了与例 7 的两种范数一致起见, 这一范数亦可用 $\|A\|_{m2}$ 表示。

$\|A\|_F$ 有一优点, 现以定理给出于下。

定理 4.2.1 设 $A \in C^{n \times n}, P, Q \in C^{n \times n}, P, Q$ 皆为酉矩阵, 则

$$\|PA\|_F = \|A\|_F = \|AQ\|_F \tag{4.12}$$

(在 A 是 n 级实方阵时, P、Q 都是正交矩阵)。

证明 由式(4.11) 及 $P^T P = E, Q Q^T = E$

$$\|PA\|_F^2 = \text{tr}[(\overline{PA})^T PA] = \text{tr}(A^H A) = \|A\|_F^2$$

$$\begin{aligned} \|AQ\|_F^2 &= \|(AQ)^H\|_F^2 = \text{tr}[(Q^H A^H)^H (AQ)^H] \\ &= \text{tr}(AA^H) = \|A\|_F^2 \end{aligned}$$

推论 与 A 酉(正交)相似的矩阵 F-范数是相同的。既若 $B = Q^H A Q$, 则 $\|B\|_F = \|A\|_F$, 其中 Q 是酉阵。

定理 4.2.2 方阵 $A \in C^{n \times n}$ 的任一种范数是 A 的元素的连续函数(证明从略)。
定理 4.2.3 对于 $C^{n \times n}$ 中任意两种方阵范数 $\| \cdot \|_a$ 与 $\| \cdot \|_b$ 必存在 $k_2, k_1 > 0$, 使

$$k_1 \|A\|_b \leq \|A\|_a \leq k_2 \|A\|_b$$

对于 $C^{n \times n}$ 中一切方阵 A 都成立(证明从略)。

4.2.2 几种常用的矩阵范数

现在给出一种规定矩阵范数的具体方法, 使矩阵范数与向量的已知范数相容。若已给定向量范数 $\| \cdot \|$, 为使

也即
$$\|Ax\| \leq \|A\| \|x\|$$

即
$$\|A\| = \max_{\|x\|=1} \|Ax\|$$

其中 $\|x\|=1$ 表示 C^n 中范数为 1 的任意向量。根据向量范数是其分量的连续函数这个性质, 则对每个矩阵 A 来讲, $\|Ax\|$ 的最大值都是可以达到的。即是说, 总可以找到这样的向量 $x_0 \neq 0, \|x_0\|=1$, 使 $\|Ax_0\| = \max_{\|x\|=1} \|Ax\|$, 于是定义

$$\|A\| = \max_{\|x\|=1} \|Ax\|$$

下面证明, 按这样定义的矩阵范数, 满足定义 4.3 中的 4 条及与向量范数的相容性。

- 1. 非负性 $\|A\| \geq 0$, 显然 $\|A\| = \max_{\|x\|=1} \|Ax\| \geq 0$
 $\|A\| = 0$, 存在 $x_0 \neq 0$, 使 $\|Ax_0\| = 0$, 令 $x = \frac{x_0}{\|x_0\|}$
那么 $\|x\|=1$, 且 $\|Ax\| = 0$, 所以 $\|A\| = 0$
当然 $\|A\| = \max_{\|x\|=1} \|Ax\| \geq \|Ax\| \geq 0$
- 2. 齐次性 $\|aA\| = \max_{\|x\|=1} \|aAx\| = \max_{\|x\|=1} |a| \|Ax\| = |a| \|A\|$
- 3. 与向量范数相容性 " $\| \cdot \|$ 为 C^n 中范数, 令 $x = \frac{Ax}{\|Ax\|}$, 则 $\|x\|=1$, 于是
$$\|Ax\| = \|A(\frac{Ax}{\|Ax\|})\| = \frac{1}{\|Ax\|} \|A(Ax)\| = \frac{1}{\|Ax\|} \|A^2x\| \leq \frac{1}{\|Ax\|} \|A\| \|A^2x\|$$

有的书定义 $\|A\| = \sup_{\|x\|=1} \|Ax\|$

= A

4. 三角不等式 对矩阵 A + B, 总存在 θ_0 , 使

$A + B = \max_{\theta=1} (A + B) = (A + B)_{\theta_0}$

$A_{\theta_0} + B_{\theta_0}$

$(A + B)_{\theta_0} = A_{\theta_0} + B_{\theta_0}$

5. 相容性 $AB = \max_{\theta=1} AB = AB_{\theta_0}$

其中 $\theta_0 = 1$, 且使 AB 达到最大值者。

所以 $AB = A_{\theta_0} B_{\theta_0} = A_{\theta_0} B$

定义 4.5 设 α 是 C^n 的一个向量范数, 对任何 $A \in C^{n \times n}$, 则

$A_{\alpha} = \max_{\alpha=1} A_{\alpha}$

是一个与 α 相容的方阵范数, 称此方阵范数为从属于向量范数 α 的算子范数。矩阵算子范数的计算归结为函数的约束极值。当向量的范数是适当时, 是有可能计算出来的。一般情况下, 从分析的角度来看, 一定存在(连续函数在有界闭集上可达到极大极小值), 但计算不是很容易的。下面给出几种特殊情形的算例。

定理 4.2.4 设 $A \in C^{n \times n}$, $\alpha \in C^n$, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$ 则从属于向量 α 的三种范数 $\alpha_1, \alpha_2, \alpha_{\infty}$ 的矩阵算子范数分别是

1. $A_{\alpha_1} = \max_{j=1}^n |a_{ij}|$ (4.13)

2. $A_{\alpha_2} = \sqrt{\lambda_1}$ 为 $A^H A$ 的最大特征值 (4.14)

3. $A_{\alpha_{\infty}} = \max_{i=1}^n |a_{ij}|$ (4.15)

证明 1 因为 $A_{\alpha_1} = \max_{\alpha=1} A_{\alpha_1}$

考查所有 $\alpha_1 = 1$ 的 $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$, $|\alpha_1| + \dots + |\alpha_n| = 1$

$$\begin{aligned} A_{\alpha_1} &= \max_{i=1}^n |a_{i1} \alpha_1 + a_{i2} \alpha_2 + \dots + a_{in} \alpha_n| \\ &= \max_{i=1}^n \left| \sum_{j=1}^n a_{ij} \alpha_j \right| = \max_{j=1}^n \left| \sum_{i=1}^n a_{ij} \alpha_j \right| \\ &= \max_{j=1}^n (|\alpha_j| \sum_{i=1}^n |a_{ij}|) = \max_{j=1}^n \left(\sum_{i=1}^n |a_{ij}| \right) \max_{j=1}^n |\alpha_j| \\ &= \max_{j=1}^n \sum_{i=1}^n |a_{ij}| \end{aligned} \tag{*}$$

另外对每个确定的 A, $\sum_{j=1}^n |a_{ij}|$ 是一个确定的值, 它是下列 n 个数

$\sum_{i=1}^n |a_{i1}|, \sum_{i=1}^n |a_{i2}|, \dots, \sum_{i=1}^n |a_{in}|$

中最大一个, 不妨设 $\sum_{j=1}^n |a_{ij}| = \sum_{i=1}^n |a_{ij0}|$ 。

那么可以取 $\alpha_0 = (\alpha_{10}, \alpha_{20}, \dots, \alpha_{n0})^T$, 其中 $\alpha_{j0} = 1, \alpha_{i0} = 0, i_0 \neq j_0$,
当然 $\alpha_{00} = 1$, 且

$$A_{00} = \max_{i=1}^n |\alpha_{ij0}| = \max_j \max_{i=1}^n |\alpha_{ij}|$$

因此 $A_{11} = \max_{i=1}^n A_{i1} \quad A_{00} = \max_j \max_{i=1}^n |\alpha_{ij}| \quad (*)$

综合式(*)与式(*), $A_{11} = \max_j \max_{i=1}^n |\alpha_{ij}|$

同样的推理可证 $A_{22} = \max_i \max_{j=1}^n |\alpha_{ij}|$

2. 因为 $A_{22} = \max_{2=1}^2 A_{22} = \max_{2=1}^2 (A^H A)_{22}$
 $\max_{2=1}^2 \alpha_{22}^{-T} A^H A \alpha_{22}$

由于 $A^H A$ 是厄米特阵, 那么复二次型 $f(\alpha)$
 $f(\alpha) = (\alpha^H)^T A \alpha \quad 0 \leq \alpha^H \alpha = 1$

那么 $f(\alpha)$ 是正定或半正定二次型, $A^H A$ 的特征值非负, 设为
 $\lambda_1, \lambda_2, \dots, \lambda_n \geq 0$

是 $A^H A$ 的 n 个特征值, 对应的标准正交特征向量为

$$\alpha_1, \alpha_2, \dots, \alpha_n$$

即有 $A^H A \alpha_i = \lambda_i \alpha_i, (\alpha_i, \alpha_j) = \delta_{ij}$

作酉阵 $U = (\alpha_1, \alpha_2, \dots, \alpha_n)$

便有 $A^H A = U D U^H$, 其中 $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$

从而有

$$A_{22} = \alpha_{22}^{-T} U D U^H \alpha_{22}$$

令 $\alpha = \alpha_{22}^{-T}$ 则 $\alpha^H \alpha = \alpha_{22}^{-T} \alpha_{22} = 1$, 故可得

$$A_{22} = \alpha^{-T} D \alpha = \sum_{i=1}^n \lambda_i |\alpha_i|^2 \leq \lambda_1$$

其中用到 $\alpha^H \alpha = \sum_{i=1}^n |\alpha_i|^2 = 1$

由于上述 α 的任意性, 当然 $\lambda_2 = 1$, 便得

$$A_{22} = \max_{2=1}^2 A_{22} = \sqrt{\lambda_1}$$

另一方面, 特别取 $\alpha_0 = \alpha_1$, 那么

$$A_{00} = A_{11} = \alpha_1^{-T} A^H A \alpha_1 = \lambda_1 \alpha_1^{-T} \alpha_1 = \lambda_1$$

所以 $A_{22} = \max_{2=1}^2 A_{22} = \sqrt{\lambda_1}$

例3^[4] 试在二维欧氏空间 R^2 中, 作出从属于向量 $\alpha = (\alpha_1, \alpha_2)^T \in R^2$ 的三种范数 1-范数, 2-范数, ∞ -范数的矩阵

$$A = \begin{bmatrix} 3 & 2 \\ -1 & 0 \end{bmatrix}$$

的三种常用范数 $\|A\|_1$, $\|A\|_2$, $\|A\|_\infty$ 的几何图形。

解 我们用实线段表示 $\|x\|_1 = 1$, 虚线段表示向量 $y = Ax = (3x_1 + 2x_2, -x_1)^T = (y_1, y_2)^T$, 则对于三种向量范数和矩阵范数而言, 其几何图形依次如图 4.3, 图 4.4, 图 4.5 所示。

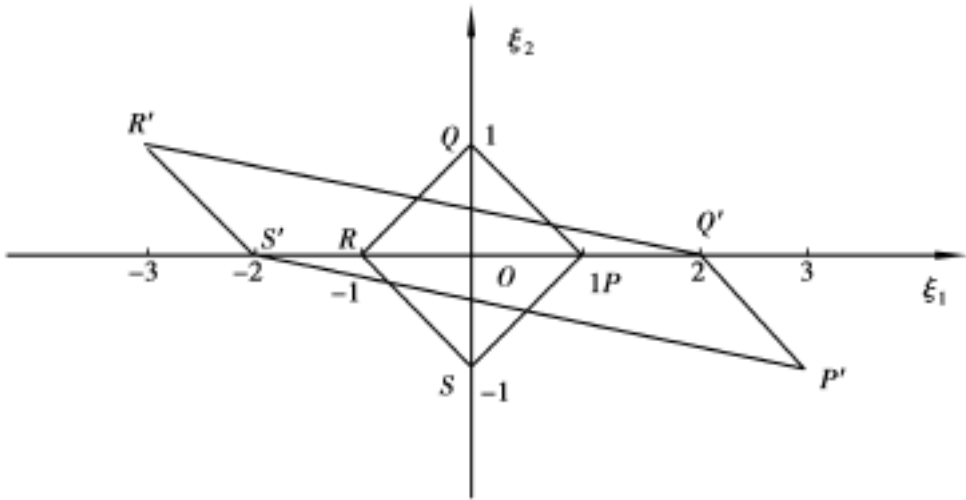


图 4.3

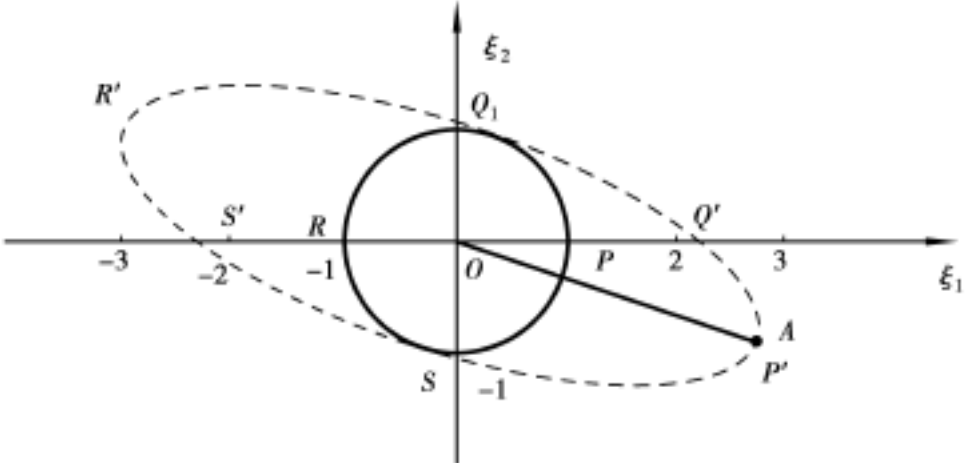


图 4.4

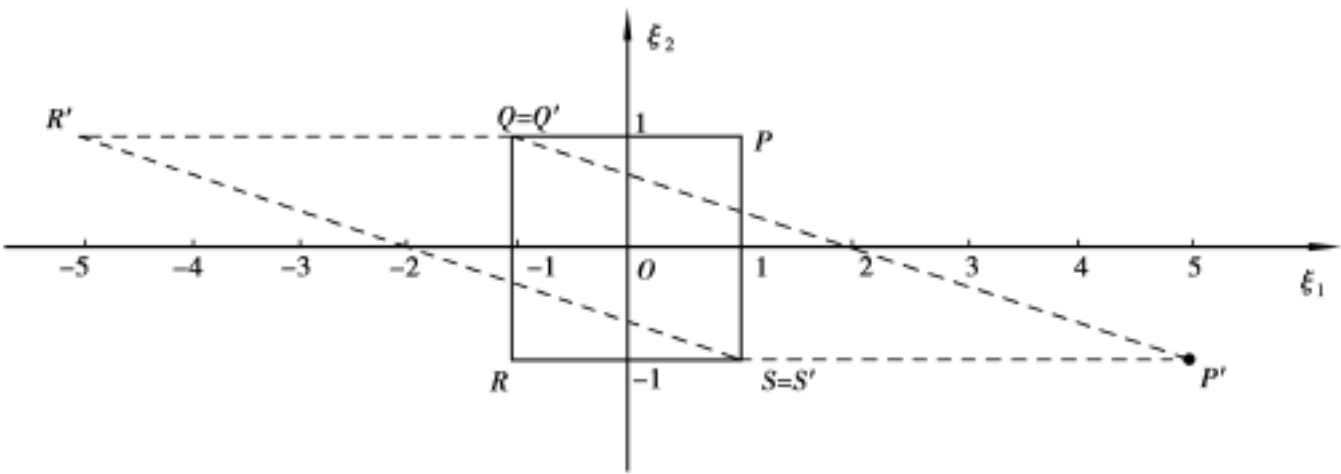


图 4.5

在图 4.3 中, 显然 $\|x\|_1 = |x_1| + |x_2| = 1$ 的图形是由正菱形 PQRS 的边界直线构成, 而 $y = Ax$ 的图形是由平行四边形 P'Q'R'S' 的边界线段构成。这里的向量 OP, OQ, OR, OS 依次是向量 OP, OQ, OR, OS 在线性变换 $y = Ax$ 下的像。于是有

$$\begin{aligned} \|A\|_1 &= \max_{\|x\|_1=1} \|Ax\|_1 = \max_{\|x\|_1=1} (|y_1| + |y_2|) \\ &= \max_{\|x\|_1=1} |y_1| + \max_{\|x\|_1=1} |y_2| \end{aligned}$$

$$=|SP|+|OS|=3+1=4$$

另一方面直接用公式(4.14) 直接可得

$$A_1=\max_{j=1}^2|a_{ij}|=\max(4,2)=4$$

从而有

$$A_1=\max_{i=1}^2A_1=\max_{j=1}^2|a_{ij}|=4$$

这就是说明线段 SP 与 OS 的长度之和等于 A₁。

在图 4.4 中, 单位圆表示 $x_1^2+x_2^2=1$ 的图形。该单位圆经线性变换 $x=A_1x_1$, 即在

$$\begin{cases} x_1=x_1-x_2 \\ x_2=\frac{1}{2}x_1+\frac{3}{2}x_2 \end{cases}$$

之下的像是图 4.4 中的虚线所示的椭圆, 其方程为

$$x_1^2+x_2^2=(x_1-x_2)^2+(\frac{1}{2}x_1+\frac{3}{2}x_2)^2=1$$

即
$$x_1^2+6x_1x_1+13x_2^2-4=0$$

而 A₂ 由椭圆的长半轴 OA 给出, 即

$$A_2=\max_{i=1}^2A_2=OA=\sqrt{7+3\sqrt{5}}=3.7$$

在图 4.5 中, 正方形 PQRS 的边界线段给出了 $\max_i|x_i|=1$, 即 $|x_1|=1$ 或 $|x_2|=1$

的图形。平行四边形 P₁Q₁R₁S₁的边界线段给出 $x=A_1x_1$ 的图形, 这时 \bar{OP} , \bar{OQ} , \bar{OR} , \bar{OS} , 依次是 \bar{OP} , \bar{OQ} , \bar{OR} , \bar{OS} 在 $x=A_1x_1$ 之下的像。于是有

$$A_1=\max_{i=1}^2A_1=\max_{i=1}^2(\max_j|a_{ij}|) \\ =\text{点 P 的横坐标}=5$$

另一方面, 由公式(4.16) 直接有

$$A_1=\max_{i=1}^2|a_{ij}|=\max(5,1)=5$$

从而有

$$A_1=\max_{i=1}^2A_1=\max_{i=1}^2\max_{j=1}^2|a_{ij}|=5$$

同样, 对任意给定的方阵范数 $\|A\|$, 也可以定义与之相容的向量范数, 这就是下面的定理。

定理 4.2.5 对任意方阵范数 $\|A\|$, $A\in C^{n\times n}$, 必在 C^n 上存在与之相容的向量范数 $\|x\|$, $\|x\|\in C^n$, $\|x\|\geq 0$ 。

证明 证明 $\|x\|$ 满足范数定义及相容性要求。

- 1. 非负性 $\|x\|\in C^n$, $\|x\|\geq 0$ 都有 $\|x\|\geq 0$
- 2. 齐次性 $k\in C$, $\|kx\|=|k|\|x\|=|k|\|x\|$
- 3. 三角不等式 $\|x+y\|\leq\|x\|+\|y\|$

$$\begin{aligned} + &= (+)^T = + + \\ &+ + = + \end{aligned}$$

4. 相容性 $A = A^T A^T = A$

与 A_1, A 比较, A_2 在计算上不如前两种方便, 但 A_2 具有一些性质, 使它在理论上有许多用处, 其中一些性质包含在下面定理中。先介绍一个符号 $\rho(A)$ 。

定义 4.6 若 $n \times n$ 矩阵 A 的全部特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 则称

$$\rho(A) = \max_i |\lambda_i|$$

为方阵 A 的谱半径。

定理 4.2.6 设 $A \in C^{n \times n}$ 则

- (1) $A_2 = A^H_2 = A^T_2 = \overline{A}_2$
- (2) $A^H A_2 = A A^H_2 = A_2^2$
- (3) 对任何 n 阶酉阵 U 及 V , 都有

$$U A_2 = A V_2 = U A V_2 = A_2$$

证明(1) 先证 $A_2 = A^H_2$, 即证 $(A^H A) = (A A^H)$ 。

设 λ 是 $A^H A$ 的任一特征值, 且 $\lambda \neq 0$, 满足

$$A^H A = \lambda I$$

若 $\lambda = 0$, 则 $A^H A$ 不满秩, 当然 $A A^H$ 也是不满秩的, 故存在非零向量 x^* , 使 $A A^H x^* = 0$, 说明 $\lambda = 0$ 也是 $A A^H$ 的特征值。

若 $\lambda \neq 0$, 则 $A^H A = \lambda I$

从而 $\lambda = A_2$

故 $A A^H = A A^H A = A = A =$

所以 λ 也是 $A A^H$ 的特征值。

同理可证明 $A A^H$ 的任一特征值也是 $A^H A$ 的特征值, 显然

$$A_2^2 = (A^H A) = (A A^H) = A^H_2^2$$

此外由

$$|E - (A^T)^H(A^T)| = |E - (A A^H)^T| = |E - A A^H|$$

立即得

$$A^T_2 = A^H_2 = A_2$$

证明(2) 由

$$\begin{aligned} A^H A_2^2 &= [(A^H A)^H(A^H A)] \\ &= [(A^H A)^2] = [(A^H A)]^2 \end{aligned}$$

立即有 $A^H A_2 = (A^H A) = A_2^2$

而 $A A^H_2 = (A^H)^H A^H_2 = A^H_2^2 = A_2^2$

证明(3) $U A_2^2 = [(U A)^H(U A)] = (A^H A) = A_2^2$

$$A V_2^2 = [(A V)^H(A V)] = (V^H A^H A V) = (A^H A) = A_2^2$$

当然也就有

$$U A V_2 = U A_2 = A_2$$

顺便强调一下, 在矩阵空间 $C^{n \times n}$ 上的各种范数都是等价的。因而在考虑收敛性问题时,

取哪一种范数, 所得收敛和发散的结论是一致的。

4.3 范数应用

4.3.1 近似逆矩阵的误差

在实际问题的计算中, 数字一般都带有某种误差, 数字矩阵 $A = (a_{ij})$ 的每个元素 a_{ij} , 通常也带有误差 δ_{ij} , 即准确矩阵为

$$A + \Delta A = (a_{ij}) + (\delta_{ij}) = (a_{ij} + \delta_{ij})$$

这里 $\Delta A = (\delta_{ij})$ 称为摄动矩阵, 现在讨论以下问题:

- (1) 若 A 可逆时, A 与 ΔA 满足什么条件时 $A + \Delta A$ 也可逆。
- (2) 当 $A + \Delta A$ 也可逆时, A^{-1} 与 $(A + \Delta A)^{-1}$ 的近似程度如何估计。

为解决上述问题, 先证明如下定理。

定理 4.3.1 若 $\|A\|_a < 1$, $\|A\|_a$ 是与向量范数 $\|x\|_a$ 相容的算子范数, 则 $E - A$ 可逆, 且

$$\|(E - A)^{-1}\|_a = \frac{1}{1 - \|A\|_a} \tag{4.16}$$

证明 设 $x \in C^n, x \neq 0$ 。

$$\begin{aligned} \|(E - A)x\|_a &= \|x - Ax\|_a = \|x\|_a - \|Ax\|_a \\ &= \|x\|_a (1 - \|A\|_a) > 0 \end{aligned}$$

这说明方程组 $(E - A)x = 0$ 只有零解。故 $E - A$ 可逆。又

$$\begin{aligned} (E - A)(E - A)^{-1} &= E \\ (E - A)^{-1} &= E + A(E - A)^{-1} \end{aligned}$$

所以

$$\begin{aligned} \|(E - A)^{-1}\|_a &= \|E\|_a + \|A(E - A)^{-1}\|_a \\ \|(E - A)^{-1}\|_a &= \frac{\|E\|_a}{1 - \|A\|_a} \end{aligned}$$

注意到对何算子范数, $\|E\|_a = \max_{a=1} \|E\|_a = 1$, 所以

$$\|(E - A)^{-1}\|_a = \frac{1}{1 - \|A\|_a}$$

定理 4.3.2 设 A 为非奇异矩阵, ΔA 为摄动矩阵, 且 $\|A^{-1}\Delta A\|_a < 1$, 则

- (1) $A + \Delta A$ 为非奇异矩阵
- (2) $(A + \Delta A)^{-1}$ 可表示成

$$(A + \Delta A)^{-1} = (E + F)A^{-1} \tag{4.17}$$

其中

$$F = \frac{A^{-1}\Delta A}{1 - \|A^{-1}\Delta A\|_a}$$

$$(3) \frac{A^{-1} - (A +)^{-1}}{A^{-1}} \frac{A^{-1}}{1 - A^{-1}} \quad (4.18)$$

证明(1) 因为

$$A + = A(E + A^{-1})$$

$$A^{-1} < 1$$

由于

根据定理 4.3.1 $E + A^{-1}$ 可逆, 从而 $A +$ 可逆。

证明(2) 因为 $(A +)^{-1} = (E + A^{-1})^{-1} A^{-1}$

$$= [E + (E + A^{-1})^{-1} - E] A^{-1}$$

令

$$F = (E + A^{-1})^{-1} - E$$

则有

$$(A +)^{-1} = (E + F) A^{-1}$$

又

$$F = (E + A^{-1})^{-1} - E = - A^{-1} (E + A^{-1})^{-1}$$

$$F A^{-1} (E + A^{-1})^{-1}$$

$$\frac{A^{-1}}{1 - A^{-1}}$$

证明(3) 由于 $(A +)^{-1} = (E + F) A^{-1}$

所以

$$A^{-1} - (A +)^{-1} = - F A^{-1}$$

$$A^{-1} - (A +)^{-1} \quad F \quad A^{-1}$$

从而

$$\frac{A^{-1} - (A +)^{-1}}{A^{-1}} \frac{A^{-1}}{1 - A^{-1}}$$

推论 1 设 $K(A) = A \quad A^{-1}$, 若 $A^{-1} < 1$, 则

$$F \quad \frac{K(A) \quad A}{1 - K(A) \quad A}$$

且

$$\frac{A^{-1} - (A +)^{-1}}{A^{-1}} \frac{K(A) \quad A}{1 - K(A) \quad A}$$

证明 因为

$$A^{-1} \quad A^{-1} \quad = \quad A^{-1} \quad A \cdot \frac{A}{A}$$

$$= K(A) \quad A$$

则有

$$F \quad \frac{A^{-1}}{1 - A^{-1}} \frac{K(A) \quad A}{1 - K(A) \quad A}$$

显然也有

$$\frac{A^{-1} - (A + \Delta A)^{-1}}{A^{-1}} = \frac{K(A) \frac{\|\Delta A\|_a}{\|A\|_a}}{1 - K(A) \frac{\|\Delta A\|_a}{\|A\|_a}}$$

从推论 1 可以看出, 若 $K(A)$ 值愈大, 则 A^{-1} 与 $(A + \Delta A)^{-1}$ 的相对误差

$$\frac{A^{-1} - (A + \Delta A)^{-1}}{A^{-1}}$$

就愈大。 $K(A)$ 刻画了矩阵 A 摄动前后其逆之间的相对误差界, 称 $K(A)$ 为方阵 A 的条件数, $K(A)$ 较大的方阵 A 通常称为病态的, $K(A)$ 较小的矩阵 A 称为良态的。

推论 2 若 A 是酉阵, 且 $\|\Delta A\|_2 < 1$ 则有

$$\frac{\|A^{-1} - (A + \Delta A)^{-1}\|_2}{\|A^{-1}\|_2} \leq \frac{\|\Delta A\|_2}{1 - \|\Delta A\|_2}$$

证明 因为 $A^{-1} = A^H$, 且

$$\|A^{-1}\|_2 = \|A^H\|_2 = \|A\|_2 = \sqrt{(A^H A)} = \sqrt{(E)} = 1$$

故

$$\|A^{-1}\|_2 = \|A^{-1}\|_2 = 1$$

由定理 4.3.2

$$\frac{\|A^{-1} - (A + \Delta A)^{-1}\|_2}{\|A^{-1}\|_2} \leq \frac{\|\Delta A\|_2}{1 - \|\Delta A\|_2}$$

当然也有

4.3.2 线性方程组的摄动

下面讨论线性方程组

$$Ax = b, A \in C^{n \times n}, b \in C^n$$

当系数矩阵 A 满秩且有摄动 ΔA , 从而使方程组变成

$$(A + \Delta A)x = b$$

或当 b 有摄动变为 $b + \Delta b$, 即方程组变成

$$Ax = b + \Delta b$$

时, 方程组解的误差估计问题。

设 $\|\cdot\|_a$ 是与向量范数 $\|\cdot\|_a$ 相容的算子范数, 则有以下定理

定理 4.3.3 设 A 可逆, $\|A\|_a > 0$, 若 $\|\Delta A\|_a, \|\Delta b\|_a$ 分别满足

$$\|A\|_a = \|A\|_a, \|A\|_a^* = \|A\|_a^*$$

则有

$$\frac{\|x - x^*\|_a}{\|x\|_a} \leq K(A) \frac{\|b - b^*\|_a}{\|b\|_a} \tag{4.19}$$

证明 因 $b - b^* = A^{-1}(Ax - Ax^*)$

故

$$\|x - x^*\|_a = A^{-1}(\|x - x^*\|_a) = A^{-1} \|x - x^*\|_a$$

又

$$\|x - x^*\|_a = A \|x - x^*\|_a$$

或

$$\|x - x^*\|_a = \frac{\|x - x^*\|_a}{A}$$

从而

$$\|x - x^*\|_a = \frac{\|x - x^*\|_a}{A^{-1} A} = \frac{\|x - x^*\|_a}{A^{-1} A} = K(A) \|x - x^*\|_a$$

定理 4.3.4 设 A 可逆, $\|A^{-1}\|_a < 1$, 若方程组 $Ax = b$ 的系数矩阵 A 有摄动 E , 且 $\|A^{-1}\|_a \|E\|_a < 1$, 而 x^* 满足

$$(A + E)x = b$$

则

$$\|x - x^*\|_a = \frac{K(A) \|E\|_a}{1 - K(A) \|E\|_a} \|x^*\|_a$$

证明 由定理 4.3.2 知, 当 $\|A^{-1}\|_a \|E\|_a < 1$ 时, $A + E$ 可逆, 故方程组有惟一解, 并记为 x^* 。又因

故有

$$x - x^* = (A + E)^{-1} (Ax - (A + E)x^*)$$
$$\|x - x^*\|_a = \|(A + E)^{-1}\|_a \|Ax - (A + E)x^*\|_a$$
$$\|x - x^*\|_a = \|(A + E)^{-1}\|_a \|Ax - Ax^* - Ex^*\|_a$$
$$\|x - x^*\|_a = \|(A + E)^{-1}\|_a (\|Ax - Ax^*\|_a + \|Ex^*\|_a)$$
$$\|x - x^*\|_a = \|(A + E)^{-1}\|_a (\|A\|_a \|x - x^*\|_a + \|E\|_a \|x^*\|_a)$$
$$\|x - x^*\|_a = \frac{\|(A + E)^{-1}\|_a \|E\|_a \|x^*\|_a}{1 - \|(A + E)^{-1}\|_a \|A\|_a \|x - x^*\|_a}$$

以上两个定理分别给出了由系数矩阵的相对摄动 $\frac{\|E\|_a}{\|A\|_a}$ 及 $\|A^{-1}\|_a$ 的相对摄动 $\frac{\|A^{-1}\|_a}{\|A\|_a}$ 引起的解的相对摄动的估计式。

习 题 4

- 1. 证明: 对任意 C^n , 有
 - (1) $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$
 - (2) $\|x\|_1 = \sum_{i=1}^n |x_i|$
 - (3) $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$
- 2. 证明: 在 R^n 中当且仅当 α, β 线性相关而且 $\alpha^T \beta = 0$ 时才有 $\|\alpha + \beta\|_2^2 = \|\alpha\|_2^2 + \|\beta\|_2^2$

3. 对任何 $\alpha, \beta \in C^n$, 总有

$$\|\alpha + \beta\|_H^2 = \frac{1}{2} \left[\|\alpha\|_H^2 + \|\beta\|_H^2 + 2 \operatorname{Re}(\alpha, \beta) \right]$$

4. 设 $\|\cdot\|_a, \|\cdot\|_b$ 是 C^n 上两个向量范数, a_1, a_2 是两个正实数, 证明

- (1) $\max\{\|\cdot\|_a, \|\cdot\|_b\} = \|\cdot\|_c$
(2) $a_1 \|\cdot\|_a + a_2 \|\cdot\|_b = \|\cdot\|_d$

都是 C^n 中的向量范数。

5. 设 a_1, a_2, \dots, a_n 是正实数, 证明: 对于任意的 $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n) \in C^n$

$$\|\alpha\| = \left(\sum_{i=1}^n a_i |\alpha_i|^2 \right)^{\frac{1}{2}}$$

是 C^n 中的向量范数。

6. 设 $\|\cdot\|_a$ 是 $C^{n \times n}$ 的方阵范数, B, C 都是 n 阶可逆矩阵, 且 $\|B^{-1}\|_a$ 及 $\|C^{-1}\|_a$ 都小于或等于 1, 证明对任何 $A \in C^{n \times n}$

$$\|A\|_b = \|BAC\|_a$$

定义了 $C^{n \times n}$ 上的一个方阵范数。

7. 设 $\|\cdot\|_a$ 是 $C^{n \times n}$ 的方阵范数, D 是 n 阶可逆矩阵, 证明对任何 $A \in C^{n \times n}$

$$\|A\|_b = \|D^{-1}AD\|_a$$

是 $C^{n \times n}$ 的方阵范数。

8. 对任意算子范数 $\|\cdot\|$, 证明

- (1) $\|E\| = 1$, E 为 n 阶单位矩阵
(2) 若 A 可逆, 则 $\|A^{-1}\| = \|A\|^{-1}$

9. 若 U 为正交矩阵, 证明

- (1) $\|U\|_2 = 1$
(2) 若 $A \in R^{n \times n}$, 则 $\|A\|_2 = \|AU\|_2$

10. 证明:

$$\frac{1}{\sqrt{n}} \|A\|_F = \|A\|_2 = \|A\|_F$$

11. 设 $\lambda \in C^{n \times n}$, λ 是 A 的特征量, 证明

$$\frac{1}{\|\lambda^{-1}\|_2} \leq \|\lambda\|_2 \leq \|A\|_2$$

12. 设 $A, B \in R^{n \times n}$, $\|\cdot\|$ 为 $R^{n \times n}$ 的算子范数, $K(A)$ 表示 A 的条件数, 证明

$$K(AB) = K(A) \cdot K(B)$$

13. 若 A 为正交矩阵, 对于方阵的 2-范数, 证明: $K(A) = 1$ 。

14. (1) 设矩阵

$$A = \begin{pmatrix} 1 & -4 & -1 & -4 \\ 2 & 0 & 5 & -4 \\ -1 & 1 & -2 & 3 \\ -1 & 4 & -1 & 6 \end{pmatrix}$$

求 $\|A\|_1$ 和 $\|A\|_\infty$ 。

(2) 设矩阵

$$B = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$$

求 $\|B\|_2$, $\|B\|_\infty$, $\|B\|_1$ 。

15. 已给 $A = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 0.5 \\ 0.2 & 0.3 \end{pmatrix}$, 估计 $\frac{\|A^{-1} - (A+B)^{-1}\|}{\|A^{-1}\|}$

第 5 章

矩阵分析及其应用

在线性代数课程中, 主要讨论矩阵的代数运算, 完全没有涉及本章将要介绍的矩阵分析的理论。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而形成的。其内容丰富, 是研究数值方法, 其他数学分枝以及许多科技问题的重要工具。本章先讨论矩阵序列的极限运算, 然后论述矩阵序列和矩阵级数的收敛定理, 引入矩阵的微分和积分的概念及其性质, 由矩阵幂级数引入矩阵函数的概念, 诸如 $e^A, \sin A, \cos A$ 等。

5.1 向量和矩阵的极限

5.1.1 向量序列极限

在上一章, 我们已经知道向量极限的概念, 为统一起见, 重新叙述一下。
定义 5.1 设给定 C^n 中向量序列 $\{x^{(k)}\}$, 其中 $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})$, 如果每一个分量 $x_i^{(k)} (i = 1, 2, \dots, n)$ 当 $k \rightarrow \infty$ 时都有极限 x_i , 即

$$\lim_{k \rightarrow \infty} x_i^{(k)} = x_i \quad i = 1, 2, \dots, n$$

则称向量序列 $\{x^{(k)}\}$ 有极限 $x_0 = (x_1, x_2, \dots, x_n)$, 或称 $\{x^{(k)}\}$ 收敛于 x_0 , 记为

$$\lim_{k \rightarrow \infty} x^{(k)} = x_0$$

或简记为

若在 $k \rightarrow \infty$ 时, 数列 $\{x_i^{(k)}\} (i = 1, 2, \dots, n)$ 中, 至少有一个极限不存在, 便称向量序列 $\{x^{(k)}\}$ 是发散的。

由向量序列极限定义易知, 向量序列 $\{x^{(k)}\}$ 的收敛于 x_0 的充分必要条件是向量序列 $\{x^{(k)} - x_0\}$ 收敛于零向量。

并且在第 4.1 节里证明了定理 4.1.2, C^n 中向量序列 $\{x^{(k)}\}$ 收敛于向量 $x_0 = (x_1, x_2, \dots, x_n)$ 的充分必要条件是对任意一种向量范数 $\| \cdot \|$, 序列 $\|x^{(k)} - x_0\|$ 收敛于零。

5.1.2 方阵序列极限

定义 5.2 对于复方阵序列 $\{A_k\}$, 其中

$$A_k = \begin{pmatrix} a_{11}^{(k)} & a_{12}^{(k)} & \dots & a_{1n}^{(k)} \\ a_{21}^{(k)} & a_{22}^{(k)} & \dots & a_{2n}^{(k)} \\ \dots & \dots & \dots & \dots \\ a_{n1}^{(k)} & a_{n2}^{(k)} & \dots & a_{nn}^{(k)} \end{pmatrix}$$

若在 $k \rightarrow \infty$ 时, n^2 个复数序列 $\{a_{ij}^{(k)}\}$ 都分别收敛于 a_{ij} , 即

$$\lim_{k \rightarrow \infty} a_{ij}^{(k)} = a_{ij} \quad (i, j = 1, 2, \dots, n)$$

则称方阵序列 $\{A_k\}$ 收敛于方阵

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

并称 A 是 $\{A_k\}$ 在 $k \rightarrow \infty$ 时的极限, 记作

$$\lim_{k \rightarrow \infty} A_k = A$$

即

$$\begin{aligned} \lim_{k \rightarrow \infty} A_k &= \lim_{k \rightarrow \infty} \begin{pmatrix} a_{11}^{(k)} & a_{12}^{(k)} & \dots & a_{1n}^{(k)} \\ a_{21}^{(k)} & a_{22}^{(k)} & \dots & a_{2n}^{(k)} \\ \dots & \dots & \dots & \dots \\ a_{n1}^{(k)} & a_{n2}^{(k)} & \dots & a_{nn}^{(k)} \end{pmatrix} \\ &= \begin{pmatrix} \lim_{k \rightarrow \infty} a_{11}^{(k)} & \lim_{k \rightarrow \infty} a_{12}^{(k)} & \dots & \lim_{k \rightarrow \infty} a_{1n}^{(k)} \\ \lim_{k \rightarrow \infty} a_{21}^{(k)} & \lim_{k \rightarrow \infty} a_{22}^{(k)} & \dots & \lim_{k \rightarrow \infty} a_{2n}^{(k)} \\ \dots & \dots & \dots & \dots \\ \lim_{k \rightarrow \infty} a_{n1}^{(k)} & \lim_{k \rightarrow \infty} a_{n2}^{(k)} & \dots & \lim_{k \rightarrow \infty} a_{nn}^{(k)} \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \end{aligned}$$

若 $k \rightarrow \infty$ 时, $\{A_k\}$ 不收敛, 则称方阵序列 $\{A_k\}$ 是发散的。

由上述定义可知: 方阵序列收敛的充要条件是 n^2 个数值序列 $\{a_{ij}^{(k)}\} (i, j = 1, 2, \dots, n)$ 都收敛; 发散的充分必要条件是这 n^2 个数值序列中至少有一个是发散的。

方阵序列 $\{A_k\}$ 的收敛性虽然由定义归结为 n^2 个数值序列的收敛性, 但借助于方阵范数就可与 n^2 个数值序列的收敛问题等价。

定理 5.1.1 $C^{n \times n}$ 中的矩阵序列 $\{A_k\}$ 收敛于方阵 A 的充分必要条件是, 对任意一种方阵

范数 $\| \cdot \|_m$, 序列 $\{ \| A_k - A \|_m \}$ 收敛于 0

证明 由上一章定理 4.2.3, 对于方阵范数 $\| \cdot \|_m$, 存在常数 $k_2 \geq k_1 > 0$, 使

$$\begin{aligned} k_1 \| B \|_m &\leq \| B \|_1 \leq k_2 \| B \|_m \\ \text{对一切 } B \in C^{n \times n} \text{ 都成立, } \| B \|_m &= n \max_{i,j} |b_{ij}|, \text{ 从而} \\ k_1 \| A_k - A \|_m &\leq \| A_k - A \|_1 \leq k_2 \| A_k - A \|_m \end{aligned} \tag{5.1}$$

也成立, 又由于方阵序列 $\{ A_k \}$ 收敛于 A , 当且仅当

$$\lim_k a_{ij}^{(k)} = a_{ij} \quad i, j = 1, 2, \dots, n$$

当且仅当
$$\lim_k | a_{ij}^{(k)} - a_{ij} | = 0 \quad i, j = 1, 2, \dots, n$$

当且仅当
$$\lim_k n \max_{i,j} | a_{ij}^{(k)} - a_{ij} | = 0$$

当且仅当
$$\lim_k \| A_k - A \|_m = 0$$

由(5.1) 式, 当且仅当

$$\lim_k \| A_k - A \|_m = 0$$

$C^{n \times n}$ 中收敛的方阵序列有如下性质:

(1) 若 $\lim_k A_k = A$, 则对 $C^{n \times n}$ 中任意方阵范数 $\| \cdot \|$, $\{ A_k \}$ 有界。

(2) 若 $\lim_k A_k = A, \lim_k B_k = B$

且 $\lim_k a_k = a, \lim_k b_k = b, \{ a_k \}, \{ b_k \}$ 是数列,

则
$$\lim_k (a_k A_k + b_k B_k) = aA + bB \tag{5.2}$$

(3) 若 $\lim_k A_k = A, \lim_k B_k = B$

则
$$\lim_k A_k B_k = \lim_k A_k \cdot \lim_k B_k = AB \tag{5.3}$$

(4) $\lim_k A_k = A$, 且 A_k^{-1} 及 A^{-1} 都存在, 则

$$\lim_k A_k^{-1} = A^{-1} \tag{5.4}$$

证明 (1) 因为存在常数 $k_2 \geq k_1 > 0$, 使

$$k_1 \| A \|_m \leq \| A \|_1 \leq k_2 \| A \|_m$$

而 $\| A \|_m = n \max_{i,j} |a_{ij}|$ 是一个固定的常数, 又

$$\| A_k - A \|_m = \| A_k - A + A - A \|_m = \| A_k - A \|_m + \| A - A \|_m$$

所以

$$\lim_k \| A_k - A \|_m = \lim_k \| A_k - A \|_m + \| A - A \|_m = 0$$

当然 $\| A_k - A \|_m \leq k_2 \| A_k - A \|_1 \leq k_2 \| A - A \|_1 + 1$ (充分大的 k)

(2) 由于对任意方阵范数 $\| \cdot \|$ 有

$$\| a_k A_k + b_k B_k - aA - bB \|$$

$$= \| a_k A_k - a_k A + a_k A - aA \| + \| b_k B_k - b_k B + b_k B - bB \|$$

$$\leq | a_k | \| A_k - A \| + | a_k - a | \| A \| + | b_k | \| B_k - B \| + | b_k - b | \| B \|$$

注意到 $\{ a_k \}, \{ b_k \}$ 有界, $| a_k - a | \rightarrow 0, \| A_k - A \| \rightarrow 0, | b_k - b | \rightarrow 0, \| B_k - B \| \rightarrow 0$, 所以上述不等式右端, 当 $k \rightarrow \infty$ 时的极限是零, 从而

$$\| a_k A_k + b_k B_k - aA - bB \| \rightarrow 0$$

由定理 5.1.1 有

$$\lim_k (a_k A_k + b_k B_k) = aA + bB$$

(3)事实上,对 $C^{n \times n}$ 的任何一种范数 $\|\cdot\|$, 有

$$\begin{aligned} \|A_k B_k - AB\| &= \|A_k B_k - A_k B + A_k B - AB\| \\ &\leq \|A_k\| \|B_k - B\| + \|A_k - A\| \|B\| \end{aligned}$$

由(1), A_k 有界,又 $B_k - B \rightarrow 0, A_k - A \rightarrow 0$, 上述不等式右端,当 $k \rightarrow \infty$ 时的极限是零,从而

$$A_k B_k - AB \rightarrow 0$$

由定理 5.1.1, 有

$$\lim_k A_k B_k = AB$$

由性质 3, 显然可得

$$\lim_k P A_k Q = PAQ$$

其中 P, Q 为确定数值矩阵。

(4)留作证明。

为了证明方阵序列的两个收敛定理, 下面先计算 Jordan 块的正整数次幂。
设

$$J(\lambda, t) = \begin{pmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \\ & 1 & & & \\ & & \ddots & & \\ & & & \lambda & \\ & & & & \lambda \end{pmatrix}_{t \times t}$$

是 t 阶 Jordan 块, 并记 t 阶矩阵

$$H = \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \\ & 1 & & \\ & & \ddots & \\ & & & 1 & \\ & & & & 0 \end{pmatrix}_{t \times t}$$

则有 $J(\lambda, t) = \lambda E + H$ (E 是 t 阶单位矩阵)

由于

$$H^m = \begin{pmatrix} \vdots & 0 & & \\ \vdots & \vdots & & \\ \vdots & 0 & & \\ \vdots & 1 & & 0 \\ & & \ddots & \\ & & & 1 & 0 & \dots & 0 \end{pmatrix}_{t \times t}$$

及 $H^t = 0$, 便有

$$J^k(\lambda, t) = (\lambda E + H)^k$$

$$\begin{aligned} &= ({}_0E)^k + C_k^1({}_0E)^{k-1}H + C_k^2({}_0E)^{k-2}H^2 + \dots + C_k^{t-1}({}_0E)^{k-t+1}H^{t-1} \\ &= {}_0^kE + k{}_0^{k-1}H + C_k^2{}_0^{k-2}H^2 + \dots + C_k^{t-1}({}_0E)^{k-t+1}H^{t-1} \\ &= \begin{pmatrix} {}_0^k & k{}_0^{k-1} & \frac{k(k-1)}{2!}{}_0^{k-2} & \dots & C_k^{t-1}{}_0^{k-t+1} \\ & {}_0^k & k{}_0^{k-1} & \dots & \\ & & {}_0^k & \dots & \\ & & & \dots & \\ & & & & C_k^2{}_0^{k-2} \\ & & & & k{}_0^{k-1} \\ & & & & {}_0^k \end{pmatrix}^T_{t \times t} \end{aligned}$$

同理, 对数值变量 x , 由

$$J({}_0, t)x = {}_0xE + xH$$

可得 $[J({}_0, t)x]^k = ({}_0xE + xH)^k$

$$= \begin{pmatrix} ({}_0X)^k & [({}_0X)^k]_0 & \frac{1}{2!}[({}_0X^k)]_0 & \dots & \frac{1}{(t-1)!}[({}_0X)^k]_0^{(t-1)} \\ & ({}_0X)^k & [({}_0X)^k]_0 & \dots & \\ & & ({}_0X)^k & \frac{1}{2!}[({}_0X^k)]_0 & \\ & & & [({}_0X)^k]_0 & \\ & & & & ({}_0X)^k \end{pmatrix}^T_{t \times t}$$

定理 5. 1. 2 矩阵 $A \in C^{n \times n}$ 的方幂 $E, A, A^2, \dots, A^k, \dots$ 所构成的矩阵序列 $\{A^k\}$ 收敛于零矩阵的充分必要条件是 A 的特征值的模都小于 1。

证明 设 A 的 Jordan 标准形为

$$\begin{aligned} J &= \begin{pmatrix} J_1(n_1) & & \\ & J_2(n_2) & \\ & & \ddots \\ & & & J_p(n_p) \end{pmatrix} \\ &= \text{diag}[J_1(n_1), J_2(n_2), \dots, J_p(n_p)] \end{aligned}$$

这里

$$J_i(n_i) = \begin{pmatrix} i & & \\ 1 & i & \\ & & \ddots \\ & & & 1 & i \end{pmatrix}_{t_i \times t_i}$$

为 t_i 阶 Jordan 块, 也就是说, 存在可逆矩阵 T , 使

$$A = TJT^{-1}$$

所以

$$A^k = TJ^kT^{-1}$$

故 $A^k = 0$ 的充分必要条件是 $J^k = 0$, 而由

$$J^k = \text{diag}[J_1^k(\lambda_1), J_2^k(\lambda_2), \dots, J_p^k(\lambda_p)]$$

显然可知, $J^k = 0$ 的充分必要条件是

$$J_i^k(\lambda_i) = 0 \quad (i = 1, 2, \dots, p)$$

但

$$J_i^k(\lambda_i) = \begin{pmatrix} \lambda_i^k & k\lambda_i^{k-1} & \dots & \frac{k(k-1)\dots(k-t_i+1)}{(t_i-1)!}\lambda_i^{k-t_i+1} \\ & \lambda_i^k & k\lambda_i^{k-1} & \\ & & \lambda_i^k & k\lambda_i^{k-1} \\ & & & \lambda_i^k \end{pmatrix}^T$$

因此, 当 $|\lambda_i| < 1$, 且 $k \rightarrow \infty$ 时

$$\left(\frac{k}{i}\right)^{(s)} = 0 \quad (s = 0, 1, 2, \dots, t_{i-1})$$

故有

$$A^k = 0$$

反之, 若有某一特征值 $|\lambda_i| = 1$, 则

$$\lambda_i^k \not\rightarrow 0$$

当然

$$A^k \not\rightarrow 0$$

命题 设 $\|A\|_a$ 是相容于向量范数 $\|\cdot\|_a$ 的方阵范数, 则 $\lim_{k \rightarrow \infty} \|A^k\|_a = 0$ 。

证明 设 λ_0 是方阵 A 的任一特征值, x_0 是 λ_0 对应的特征向量, 则

$$Ax_0 = \lambda_0 x_0$$

那么

$$\|x_0\|_a = \left\| \frac{1}{\|\lambda_0\|_a} Ax_0 \right\|_a = \frac{\|Ax_0\|_a}{\|\lambda_0\|_a} = \frac{\|A\|_a \|x_0\|_a}{\|\lambda_0\|_a}$$

定理 5.1.3 $A^k = 0$ 的充分必要条件是至少存在一种方阵范数 $\|\cdot\|$, 使 $\|A\| < 1$ 。

5.2 函数矩阵的微分和积分

本节所研究的内容: 是用矩阵来描述微积分中的若干结果, 它在实用上很方便。我们着重讨论在工程实际中常见的三个问题, 即是函数矩阵关于自变量的微分和积分; 纯量函数关于矩阵的微分; 向量函数关于向量的微分。

5.2.1 函数矩阵对自变量的微分与积分

定义 5.3 设矩阵

$$A(z) = \begin{pmatrix} a_{11}(z) & a_{12}(z) & \dots & a_{1n}(z) \\ a_{21}(z) & a_{22}(z) & \dots & a_{2n}(z) \\ \hline a_{n1}(z) & a_{n2}(z) & \dots & a_{nn}(z) \end{pmatrix}$$

其中每个元素 $a_{ij}(z)$ 都是复变量函数, 称 $A(z)$ 为函数矩阵。

定义 5.4 若函数矩阵 $A(z) = (a_{ij}(z))_{m \times n}$ 的每个元素 $a_{ij}(z)$ 都是复变量 z 的函数, 且都在 $z=z_0$ 或变量 z 的某个区域 D 上可微, 则称此函数矩阵 $A(z)$ 在 $z=z_0$ 或区域 D 上是可微的, 并规定 $A(z)$ 对 z 的导数为

$$\begin{aligned} \frac{d}{dz}A(z) &= \left(\frac{d}{dz}a_{ij}(z) \right)_{m \times n} \\ &= \begin{pmatrix} \frac{d}{dz}a_{11}(z) & \frac{d}{dz}a_{12}(z) & \dots & \frac{d}{dz}a_{1n}(z) \\ \frac{d}{dz}a_{21}(z) & \frac{d}{dz}a_{22}(z) & \dots & \frac{d}{dz}a_{2n}(z) \\ \hline \frac{d}{dz}a_{m1}(z) & \frac{d}{dz}a_{m2}(z) & \dots & \frac{d}{dz}a_{mn}(z) \end{pmatrix} \end{aligned}$$

单元函数矩阵有以下性质:

性质 1 若函数矩阵 $A(z), B(z)$ 是可微的, 则它们的和也可微, 且

$$\frac{d}{dz}[A(z) + B(z)] = \frac{d}{dz}A(z) + \frac{d}{dz}B(z) \tag{5.5}$$

性质 2 设函数矩阵 $A(z), B(z)$ 分别是 $m \times n$ 及 $n \times s$ 阶矩阵, 且 $A(z), B(z)$ 都可微,

$$\frac{d}{dz}[A(z) B(z)] = \left[\frac{d}{dz}A(z) \right] B(z) + A(z) \left[\frac{d}{dz}B(z) \right] \tag{5.6}$$

性质 3 设函数矩阵 $A(u) = (a_{ij}(u))_{m \times n}$ 及变量 z 的函数 $u = f(z)$ 都可微, 则

$$\frac{d}{dz}A[f(z)] = \frac{d}{du}A(u) \cdot \frac{d}{dz}f(z) \tag{5.7}$$

性质 4 若 n 阶函数矩阵 $A(z)$ 可逆, 且 $A(z)$ 及其逆阵 $A^{-1}(z)$ 都可微, 则

$$\frac{d}{dz}A^{-1}(z) = -A^{-1}(z) \left[\frac{d}{dz}A(z) \right] A^{-1}(z) \tag{5.8}$$

证明性质 2 设 $A(z) B(z) = C(z)$

$$\left[\frac{d}{dz}A(z) \right] B(z) = H(z)$$

$$A(z) \frac{d}{dz}B(z) = D(z)$$

元素 $c_{ij}(z) = \sum_{k=1}^n a_{ik}(z) b_{kj}(z) \quad i = 1, \dots, m \quad j = 1, \dots, s$

那么

$$\begin{aligned} \frac{d}{dz}c_{ij}(z) &= \sum_{k=1}^n \frac{d}{dz}[a_{ik}(z) b_{kj}(z)] \\ &= \sum_{k=1}^n \left[\frac{d}{dz}a_{ik}(z) \right] b_{kj}(z) + \sum_{k=1}^n a_{ik}(z) \frac{d}{dz}b_{kj}(z) \end{aligned}$$

$$= h_{ij}(z) + d_{ij}(z)$$

其中 $c_{ij}(z)$, $h_{ij}(z)$, $d_{ij}(z)$ 分别是函数矩阵 $C(z)$, $H(z)$, $D(z)$ 的元素, 可得

$$\begin{aligned} \frac{d}{dz}[A(z)B(z)] &= \frac{d}{dz}C(z) \\ &= H(z) + D(z) \\ &= \left[\frac{d}{dz}A(z) \right]B(z) + A(z) \frac{d}{dz}B(z) \end{aligned}$$

这里需注意的是, 由于矩阵的乘法不满足交换律, 所以上式中乘积的顺序一般是不能交换的。

若 K 是一个常数矩阵, 则有

$$\begin{aligned} \frac{d}{dz}[K \cdot A(z)] &= K \frac{d}{dz}A(z) \\ \frac{d}{dz}[A(z) \cdot K] &= \left[\frac{d}{dz}A(z) \right] \cdot K \end{aligned}$$

这两个式子也不能交换顺序。又如

$$\begin{aligned} \frac{d}{dz}[A^2(z)] &= \left[\frac{d}{dz}A(z) \right]A(z) + A(z) \frac{d}{dz}A(z) \\ &= 2A(z) \frac{d}{dz}A(z) \end{aligned}$$

证明性质 3, 因为

$$\frac{d}{dz}a_{ij}[f(z)] = \frac{d}{du}a_{ij}(u) \frac{d}{dz}f(z)$$

由此立刻得出

$$\frac{d}{dz}A[f(z)] = \frac{d}{du}A(u) \frac{d}{dz}f(z)$$

证明性质 4, 因为

$$A^{-1}(z) \cdot A(z) = E$$

所以

$$\begin{aligned} \frac{d}{dz}[A^{-1}A(z)] &= \frac{dA^{-1}(z)}{dz}A(z) + A^{-1}(z) \frac{d}{dz}A(z) = 0 \\ \frac{d}{dz}[A^{-1}(z)] &= -A^{-1}(z) \left[\frac{d}{dz}A(z) \right]A^{-1}(z) \end{aligned}$$

例 1 求二次型 tA 对变量 t 的导数。

其中

$$= \begin{pmatrix} x_1(t) \\ x_2(t) \\ \dots \\ x_n(t) \end{pmatrix}; A = (a_{ij})_{n \times n}, a_{ij} = a_j$$

$$\begin{aligned} \text{解 } \frac{d}{dt}[^tA] &= \left[\frac{d}{dt}^t \right]A + ^tA \frac{d}{dt} \\ &= ^tA^T \frac{d}{dt} + ^tA \frac{d}{dt} = 2^tA \frac{d}{dt} \end{aligned}$$

类似地定义函数矩阵的积分。
定义 5.5 若函数矩阵 $A(t) = (a_{ij}(t))$ 的每个元素都是实变数 t 的函数, 且都在 $[a, b]$ 上可积, 则称函数矩阵 $A(t)$ 在 $[a, b]$ 上是可积的, 并规定

$$\int_a^b A(t) dt = \begin{pmatrix} \int_a^b a_{11}(t) dt & \dots & \int_a^b a_{1n}(t) dt \\ \int_a^b a_{m1}(t) dt & \dots & \int_a^b a_{mn}(t) dt \end{pmatrix}$$

为 $A(t)$ 在 $[a, b]$ 上的定积分, 而

$$\int A(t) dt = \begin{pmatrix} \int a_{11}(t) dt & \dots & \int a_{1n}(t) dt \\ \int a_{m1}(t) dt & \dots & \int a_{mn}(t) dt \end{pmatrix}$$

称为 $A(t)$ 的不定积分。

易证矩阵的积分有以下性质:
性质 1 对任何函数 $A(t)$, 有

$$\int A^T(t) dt = \left[\int A(t) dt \right]^T \tag{5.9}$$

性质 2 对函数矩阵 $A(t), B(t)$ 及 $a, b \in \mathbb{R}$, 有

$$\int [aA(t) + bB(t)] dt = a \int A(t) dt + b \int B(t) dt \tag{5.10}$$

性质 3 对函数矩阵 $A(t)$ 及常矩阵 B, C , 有

$$\int A(t) B dt = \left[\int A(t) dt \right] B \tag{5.11}$$

$$C \int A(t) dt = C \int A(t) dt \tag{5.12}$$

性质 4 对于函数矩阵 $A(t), B(t)$, 有

$$\int \left[A(t) \frac{d}{dt} B(t) \right] dt = A(t) B(t) - \int \left[\frac{d}{dt} A(t) \right] B(t) dt \tag{5.13}$$

下面只证性质 4, 设

$$A(t) = (a_{ij}(t))_{m \times n}, B(t) = (b_{ij}(t))_{n \times s}$$

则 $A(t) \cdot \frac{d}{dt} B(t)$ 的 i 行 j 列元素为

$$\sum_{k=1}^n a_{ik}(t) \frac{d}{dt} b_{kj}(t)$$

所以

$$\begin{aligned} & \int \sum_{k=1}^n a_{ik}(t) \frac{d}{dt} b_{kj}(t) dt \\ &= \sum_{k=1}^n \int a_{ik}(t) \frac{d}{dt} b_{kj}(t) dt \\ &= \sum_{k=1}^n \{ a_{ik}(t) b_{kj}(t) - \int \left[\frac{d}{dt} a_{ik}(t) \right] b_{kj}(t) dt \} \end{aligned}$$

$$= \sum_{k=1}^n a_{ik}(t) b_{kj}(t) - \sum_{k=1}^n \left\{ \frac{d}{dt} a_{ik}(t) \right\} b_{kj}(t) dt$$

故
$$A(t) \frac{d}{dt} B(t) dt = A(t) B(t) - \left\{ \frac{d}{dt} A(t) \right\} B(t) dt$$

5.2.2* 复合矩阵 A(X) 对矩阵 X 的微分法

设有 m× n 阶矩阵

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{pmatrix}$$

而 p× q 阶矩阵 A(X) 是以 X 为变元的函数矩阵

$$A(X) = \begin{pmatrix} a_{11}(X) & a_{12}(X) & \cdots & a_{1q}(X) \\ a_{21}(X) & a_{22}(X) & \cdots & a_{2q}(X) \\ \cdots & \cdots & \cdots & \cdots \\ a_{p1}(X) & a_{p2}(X) & \cdots & a_{pq}(X) \end{pmatrix}$$

现在讨论 A(X) 对 X 的微分法

定义 5.6 设 $X = (x_{ij})$ 是 m× n 阶矩阵, A(X) 是以 X 为变元的 p× q 阶矩阵, 则称 np× mq 阶矩阵

$$\begin{pmatrix} \frac{A(X)}{x_{11}} & \frac{A(X)}{x_{21}} & \cdots & \frac{A(X)}{x_{m1}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{A(X)}{x_{1n}} & \frac{A(X)}{x_{2n}} & \cdots & \frac{A(X)}{x_{mn}} \end{pmatrix}$$

为函数矩阵 A(X) 对矩阵 X 的导数, 记为

$$\frac{d}{dX} A(X) = \begin{pmatrix} \frac{A(X)}{x_{11}} & \cdots & \frac{A(X)}{x_{m1}} \\ \cdots & \cdots & \cdots \\ \frac{A(X)}{x_{1n}} & \cdots & \frac{A(X)}{x_{mn}} \end{pmatrix}$$

其中
$$\frac{A(X)}{x_{ij}} = \begin{pmatrix} \frac{a_{11}(X)}{x_{ij}} & \cdots & \frac{a_{1q}(X)}{x_{ij}} \\ \cdots & \cdots & \cdots \\ \frac{a_{p1}(X)}{x_{ij}} & \cdots & \frac{a_{pq}(X)}{x_{ij}} \end{pmatrix}$$

定义的 X 及 A(X) 可以是一阶矩阵(标量), 也可以是向量或矩阵, 所以矩阵对于矩阵的导数的一般运算公式是比较复杂的, 现仅讨论几种常用的特殊情况。

(1) A(X) 是 1 级方阵函数 f(X), 对矩阵 X 的导数:

$$\frac{d}{dX} f(X) = \begin{pmatrix} \frac{f}{x_{11}} & \cdots & \frac{f}{x_{m1}} \\ \frac{f}{x_{1n}} & \cdots & \frac{f}{x_{mn}} \end{pmatrix}$$

是 $n \times m$ 阶矩阵, 特别当 $X = (x_1, \dots, x_n)$

$$\frac{d}{dX} f(X) = \begin{pmatrix} \frac{f}{x_1} \\ \vdots \\ \frac{f}{x_n} \end{pmatrix}$$

$$\frac{d}{dX^T} f(X) = \left[\frac{f}{x_1}, \dots, \frac{f}{x_n} \right]$$

函数 $f(X)$ 对 X 的导数有以下基本性质:

性质 1 $\frac{d}{dX^T} f(X) = \left[\frac{d}{dX} f(X) \right]^T$ (5.14)

性质 2 对数值函数 $f(X), g(X)$ 及任意常数

$$\frac{d}{dX} [af(X) + bg(X)] = a \frac{d}{dX} f(X) + b \frac{d}{dX} g(X) \tag{5.15}$$

性质 3 $\frac{d}{dX} [f(X) \cdot g(X)] = g(X) \frac{d}{dX} f(X) + f(X) \frac{d}{dX} g(X)$ (5.16)

下面只证性质 3, 注意到 $\frac{d}{dX} f(X)$ 及 $\frac{d}{dX} g(X)$ 的 i 行 j 列元素分别是 $\frac{f}{x_{ji}}, \frac{g}{x_{ji}}$, 而

$\frac{d}{dX} [f(X) \cdot g(X)]$ 的 i 行 j 列处元素是

$$\frac{d}{dX} [f(X) \cdot g(X)]_{ji} = \frac{f}{x_{ji}} g(X) + f(X) \frac{g}{x_{ji}}$$

故有

$$\frac{d}{dX} [f(X) \cdot g(X)] = \left[\frac{d}{dX} f(X) \right] g(X) + f(X) \frac{d}{dX} g(X)$$

特殊情况下, 当 $f(X) = g(X)$ 时

$$\frac{d}{dX} f^2(X) = 2f(X) \frac{d}{dX} f(X)$$

一般地有

$$\frac{d}{dX} f^n(X) = nf^{n-1}(X) \frac{d}{dX} f(X)$$

例 2 设 $A = (a_{ij}) \in C^{m \times n}, X = (x_{ij}) \in C^{n \times m}$, 证明

$$\frac{d}{dX} \text{tr}(AX) = \frac{d}{dX} \text{tr}(XA) = A$$

$$\frac{d}{dX} \text{tr}(BX^T) = \frac{d}{dX} \text{tr}(XB^T) = B^T$$

证明 先证

$$\frac{d}{dX} \text{tr}(AX) = A$$

因为 $\text{tr}(AX) = \sum_{i=1}^n a_{1i}x_{i1} + \sum_{i=1}^n a_{2i}x_{i2} + \dots + \sum_{i=1}^n a_{ni}x_{in}$

$$\frac{d}{dx_{ij}} \text{tr}(AX) = a_{ji}$$

所以 $\frac{d}{dX} \text{tr}(AX) = A$

再由 $\text{tr}(AX) = \text{tr}(XA)$
 $\text{tr}(XB^T) = \text{tr}(XB^T)^T = \text{tr}(BX^T)$

立即有

$$\frac{d}{dX} \text{tr}(XA) = A$$

$$\frac{d}{dX} \text{tr}(XB^T) = \frac{d}{dX} \text{tr}(BX^T) = B^T$$

特别当 $X \in C^{n \times n}$ 时, 有

$$\frac{d}{dX} \text{tr}X = \frac{d}{dX} \text{tr}(XE) = E$$

$$\frac{d}{dX} \text{tr}X^T = \frac{d}{dX} \text{tr}(EX^T) = E$$

这里 E 是 n 阶单位矩阵。

例 3 设 A, B 是 $m \times n$ 阶常数矩阵, X 为 $n \times m$ 阶矩阵, 证明

$$\frac{d}{dX} \text{tr}(AXBX) = \frac{d}{dX} \text{tr}(XAXB) = AXB + BXA$$

证明 因为

$$\begin{aligned} \frac{d}{dx_{ij}} \text{tr}(AXBX) &= \text{tr} \frac{d}{dx_{ij}}(AXBX) \\ &= \text{tr} \left\{ \frac{AX}{x_{ij}}BX + AX \frac{BX}{x_{ij}} \right\} \\ &= \text{tr} \{ AE_{ij}BX + AXBE_{ij} \} \\ &= \text{tr}(AE_{ij} \cdot BX) + \text{tr}(AXBE_{ij}) \\ &= \text{tr}(BX \cdot AE_{ij}) + \text{tr}(AXBE_{ij}) \\ &= (BXA)_{ji} + (AXB)_{ji} \end{aligned}$$

$(BXA)_{ji}$ 表示 BXA 的第 j 行 i 列处元素, 所以

$$\frac{d}{dX} \text{tr}(AXBX) = BXA + AXB$$

同理可证

$$\frac{d}{dX} \text{tr}(XAXB) = BXA + AXB$$

例 4 设 X 为 n 阶方阵, k 为正整数, 证明: 当 $\det X \neq 0$ 时, 有

$$\frac{d}{dX}(\det X^k) = k\det X^k \cdot X^{-1}$$

证明 把 $\det X$ 按第 j 行展开得

$$\det X = x_{j1} X_{j1} + x_{j2} X_{j2} + \dots + x_{jn} X_{jn}$$

这里 X_{jk} 是 x_{jk} 的代数余子式, 由此有

$$\begin{aligned} \frac{\det X}{x_{ji}} &= X_{ji} \\ \text{即 } \frac{d}{dX} \det X &= \begin{pmatrix} X_{11} & X_{21} & \dots & X_{n1} \\ X_{12} & X_{22} & \dots & X_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & \dots & X_{nn} \end{pmatrix} = X^* \\ &= (\det X) \cdot X^{-1} \end{aligned}$$

从而

$$\begin{aligned} \frac{d}{dX}(\det X^k) &= \frac{d}{dX}(\det X)^k \\ &= k(\det X)^{k-1} \frac{d}{dX}(\det X) \\ &= k\det X^k \cdot X^{-1} \end{aligned}$$

(2) $A(X)$ 是向量函数 $f(\quad)$, 对向量 \quad 的导数;

$$f(\quad) = \begin{pmatrix} f_1(\quad) \\ \vdots \\ f_m(\quad) \end{pmatrix}, \quad \quad = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

则由定义 5.6 有

$$\frac{d}{d} f(\quad) = \begin{pmatrix} \frac{f_1}{x_1} & \dots & \frac{f_1}{x_2} & \frac{f_1}{x_n} \\ \frac{f_2}{x_1} & \dots & \frac{f_2}{x_2} & \frac{f_2}{x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{f_m}{x_1} & \dots & \frac{f_m}{x_2} & \frac{f_m}{x_n} \end{pmatrix}$$

通常称此矩阵为 Jacobi 矩阵。

例 5 设 $\quad = (x_1, \dots, x_n)^T$, 求 $\frac{d}{d}$

$$\begin{aligned} \text{解 } \frac{d}{d} &= \left(\frac{1}{x_1}, \dots, \frac{1}{x_n} \right) = \begin{pmatrix} \frac{1}{x_1} & \dots & \frac{1}{x_n} \\ \frac{1}{x_2} & \dots & \frac{1}{x_n} \\ \vdots & \ddots & \vdots \\ \frac{1}{x_n} & \dots & \frac{1}{x_n} \end{pmatrix} \\ &= \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} = E \end{aligned}$$

同样由定义 5.6 可以得出

$$\frac{d}{d} f^T(), \quad \frac{d}{d^T} f(), \quad \frac{d}{d^T} f^T()$$

关于向量对向量的导数 $\frac{d}{d} f()$ 有以下性质:

性质 1 $\frac{d}{d^T} f^T() = \left[\frac{d}{d} f() \right]^T$ (5.17)

性质 2 对于任意 m 维列向量 $f(), g()$ 及任意常数 a, b 有

$$\frac{d}{d} [a f() + b g()] = a \frac{d}{d} f() + b \frac{d}{d} g()$$
 (5.18)

性质 3 对于 $s \times m$ 阶常数矩阵 A 有

$$\frac{d}{d} [A \cdot f()] = A \cdot \frac{d}{d} f()$$

性质 4 若 $f(), g()$ 都是 m 维列向量, 则

$$\frac{d}{d} [f^T() \cdot g()] = f^T() \frac{d g()}{d} + g^T() \frac{d f()}{d}$$

性质 1, 2, 易于证明, 下面只证性质 3, 4。

设 $A = \begin{pmatrix} a_{11} \dots a_{1m} \\ \dots \\ a_{s1} \dots a_{sm} \end{pmatrix}$

则 $A \cdot f() = \begin{pmatrix} a_{11} f_1() + a_{12} f_2() + \dots + a_{1m} f_m() \\ \dots \\ a_{s1} f_1() + a_{s2} f_2() + \dots + a_{sm} f_m() \end{pmatrix}$

所以

$$\begin{aligned} \frac{d}{d} [A \cdot f()] &= \begin{pmatrix} \frac{d}{d} [a_{11} f_1() + a_{12} f_2() + \dots + a_{1m} f_m()] \\ \dots \\ \frac{d}{d} [a_{s1} f_1() + a_{s2} f_2() + \dots + a_{sm} f_m()] \end{pmatrix} \\ &= \begin{pmatrix} a_{11} \dots a_{1m} \\ \dots \\ a_{s1} \dots a_{sm} \end{pmatrix} \begin{pmatrix} \frac{d f_1()}{d} \dots \frac{d f_m()}{d} \\ \dots \\ \frac{d f_m()}{d} \dots \frac{d f_m()}{d} \end{pmatrix} \\ &= A \cdot \frac{d}{d} f() \end{aligned}$$

故性质 3 成立, 对于性质 4, 因为

$$f(\quad) = \begin{pmatrix} f_1(\quad) \\ \vdots \\ f_m(\quad) \end{pmatrix} \quad g(\quad) = \begin{pmatrix} g_1(\quad) \\ \vdots \\ g_m(\quad) \end{pmatrix} \quad \frac{df}{d} = \begin{pmatrix} \frac{df_1}{d} \\ \vdots \\ \frac{df_m}{d} \end{pmatrix}$$

又

$$f^T \cdot g = f_1(\quad) g_1(\quad) + \dots + f_m(\quad) g_m(\quad)$$

所以

$$\frac{d}{d}[f^T \cdot g] = \frac{d}{d}[f_1(\quad) g_1(\quad) + \dots + f_m(\quad) g_m(\quad)]$$

由(5.15) 式及(5.16) 式, 上式为

$$\begin{aligned} & f_1(\quad) \frac{dg_1(\quad)}{d} + \dots + f_m(\quad) \frac{dg_m(\quad)}{d} + \\ & g_1(\quad) \frac{df_1(\quad)}{d} + \dots + g_m(\quad) \frac{df_m(\quad)}{d} \\ &= (f_1(\quad), \dots, f_m(\quad)) \begin{pmatrix} \frac{dg_1(\quad)}{d} \\ \vdots \\ \frac{dg_m(\quad)}{d} \end{pmatrix} + (g_1(\quad), \dots, g_m(\quad)) \begin{pmatrix} \frac{df_1(\quad)}{d} \\ \vdots \\ \frac{df_m(\quad)}{d} \end{pmatrix} \\ &= f^T \frac{dg(\quad)}{d} + g^T \frac{df(\quad)}{d} \end{aligned}$$

例 6 设 A 为 n 阶常数矩阵, 求 $f(\quad) = {}^T A$ 对 \quad 的导数。
解

$$\begin{aligned} \frac{d}{d}({}^T A) &= {}^T \frac{dA}{d} + (A)^T \frac{d}{d} \\ &= {}^T A + {}^T A^T \end{aligned}$$

特别当 A 为实对称矩阵时, 则有

$$\frac{d}{d}({}^T A) = 2 {}^T A$$

5.3 方阵的幂级数

5.3.1 方阵级数收敛定义及判断

定义 5.7 对于任意的 n 阶复方阵序列 $\{A_k\}$, 称 $\sum_{k=0}^N A_k$ 为方阵级数, 记 $S_N = \sum_{k=0}^N A_k$, 若方阵序列 $\{S_N\}$ 收敛且

$$\lim_N S_N = S$$

则称方阵级数 $\sum_{k=0}^{\infty} A_k$ 收敛, 且称该方阵级数的和为 S , 记

$$S = \sum_{k=0}^{\infty} A_k$$

不收敛的方阵级数称为发散。

如果记方阵 A_k 第 i 行 j 列元素为 $(A_k)_{ij}$, 方阵 S 的 i 行 j 列元素为 $(S)_{ij}$, 则由定义 5.7 及定义 5.2 知, 方阵级数 $\sum_{k=0}^{\infty} A_k$ 收敛于 S 的充分必要条件是, 对方阵序列 $\{S_N\}$

$$\lim_{N \rightarrow \infty} S_N = S, S_N = \sum_{k=0}^N A_k$$

的充分必要条件是 n^2 个数值序列 $\{(S_N)_{ij}\}$

$$\lim_{N \rightarrow \infty} (S_N)_{ij} = (S)_{ij} \quad \begin{matrix} i = 1, 2, \dots, n \\ j = 1, 2, \dots, n \end{matrix}$$

也即 n^2 个数值级数

$$\sum_{k=0}^{\infty} (A_k)_{ij} \quad (i, j = 1, 2, \dots, n)$$

都收敛, 当且仅当这 n^2 个数值级数中至少有一个发散时, 方阵级数 $\sum_{k=0}^{\infty} A_k$ 发散。

定义 5.8 若方阵级数 $\sum_{k=0}^{\infty} A_k$ 所对应的 n^2 个数值级数

$$\sum_{k=0}^{\infty} (A_k)_{ij} \quad (i, j = 1, 2, \dots, n)$$

都绝对收敛, 则称该方阵级数绝对收敛。

由数项级数, 方阵级数的收敛及绝对收敛定义和数项级数收敛及绝对收敛的性质可以得到下述方阵级数的收敛和绝对收敛性质。

性质 1 若方阵级数 $\sum_{k=0}^{\infty} A_k$ 绝对收敛, 则它一定收敛, 并且任意调换各项的次序所得的新级数仍收敛, 且其和不变。

性质 2 方阵级数绝对收敛的充分必要条件是对于任意一种方阵范数 $\|\cdot\|$, 级数 $\sum_{k=0}^{\infty} \|A_k\|$ 收敛。

性质 3 对于确定的数字矩阵 $P, Q \in C^{n \times n}$, 若方阵级数 $\sum_{k=0}^{\infty} A_k$ 收敛(绝对收敛), 则方阵级数 $\sum_{k=0}^{\infty} PA_kQ$ 也收敛(绝对收敛), 且

$$\sum_{k=0}^{\infty} PA_kQ = P \left(\sum_{k=0}^{\infty} A_k \right) Q$$

证明性质 2 若 $\sum_{k=0}^{\infty} A_k$ 收敛, 由于对任意的方阵范数 $\|\cdot\|$ 及方阵范数 $\|\cdot\|_1$, 必存在 $k_2 > k_1 > 0$, 使

$$\|A_k\|_1 \leq k_1 \quad \|A_k\| \leq k_2 \quad A_k \geq 0$$

对任意 $A_k \in C^{n \times n}$ 都成立, 所以由比较法则, 级数 $\sum_{k=0}^{\infty} A_{k+1}$ 也收敛, 但

$$\|A_{k+1}\| = \max_{j=1, \dots, n} \sum_{i=1}^n |(A_k)_{ij}|$$

因此 $\sum_{k=0}^{\infty} \|(A_k)_{ij}\| \leq \|A_{k+1}\| \quad (i, j = 1, \dots, n)$

由比较法则可知级数

$$\sum_{k=0}^{\infty} |(A_k)_{ij}| \quad (i, j = 1, 2, \dots, n)$$

都收敛, 从而方阵级数

$$\sum_{k=0}^{\infty} A_k$$

绝对收敛。

反之, 若方阵级数

$$\sum_{k=0}^{\infty} A_k$$

绝对收敛, 则级数

$$\sum_{k=0}^{\infty} |(A_k)_{ij}| \quad (i, j = 1, 2, \dots, n)$$

都收敛, 当然 n^2 个收敛级数的和仍收敛, 故级数

$$\sum_{k=0}^{\infty} \left(\sum_{i=1}^n \sum_{j=1}^n |(A_k)_{ij}| \right)$$

也收敛, 而

$$\|A_{k+1}\| = \max_{j=1, \dots, n} \sum_{i=1}^n |(A_k)_{ij}| = \sum_{j=1}^n \max_{i=1}^n |(A_k)_{ij}|$$

故由比较法则知级数

$$\sum_{k=0}^{\infty} \|A_{k+1}\|$$

收敛, 由范数等价性知级数

$$\sum_{k=0}^{\infty} A_k$$

也收敛。

证明 3 若 $\sum_{k=0}^{\infty} A_k$ 收敛, 令

$$S_N = \sum_{k=0}^N A_k \quad S = \sum_{k=0}^{\infty} A_k$$

则有

$$\lim_{N \rightarrow \infty} S_N = S$$

从而

$$\lim_{N \rightarrow \infty} \sum_{k=0}^N P A_k Q = \lim_{N \rightarrow \infty} P \left(\sum_{k=0}^N A_k \right) Q$$

$$= P(\lim_{N \rightarrow \infty} \sum_{k=0}^N A_k) Q = PSQ$$

即

$$\sum_{k=0}^{\infty} PA_k Q$$

收敛, 且 $\sum_{k=0}^{\infty} PA_k Q = P(\sum_{k=0}^{\infty} A_k) Q$ (显然, 对于 $P \in C^{s \times n}, Q \in C^{n \times s}$ 上述结论仍成立)。

若 $\sum_{k=0}^{\infty} A_k$ 绝对收敛, 由性质 2, 对任意方阵范数 $\|\cdot\|$, 级数 $\sum_{k=0}^{\infty} A_k$ 收敛, 由于

$$\|PA_k Q\| = \|P\| \|A_k\| \|Q\| = \|P\| \|Q\| \|A_k\|$$

所以级数 $\sum_{k=0}^{\infty} PA_k Q$ 收敛, 从而 $\sum_{k=0}^{\infty} PA_k Q$ 绝对收敛。

5.3.2 方阵幂级数收敛定义及判断

下面讨论一类重要的方阵级数——方阵幂级数。

对于任意方阵 $A \in C^{n \times n}$ 及复数域上数列 $\{a_k\}$, 称方阵级数

$$\sum_{k=0}^{\infty} a_k A^k$$

为方阵 A 的幂级数。

类似前面定义, $A \in C^{n \times n}$, 方阵幂级数 $\sum_{k=0}^{\infty} a_k A^k$ 收敛充分必要条件为 n^2 个数项级数

$\sum_{k=0}^{\infty} a_k (A^k)_{ij}$ 收敛, $i, j = 1, 2, \dots, n$; 方阵幂级数 $\sum_{k=0}^{\infty} a_k A^k$ 绝对收敛的充分必要条件为对任意一种

方阵范数 $\|A\|$, 级数 $\sum_{k=0}^{\infty} \|a_k A^k\|$ 收敛。

为了得到判断方阵幂级数收敛和发散的方法, 先证明下面定理。

定理 5.3.1 设 $A \in C^{n \times n}$ 的谱半径为 $\rho(A)$, 则对任意给定的 $\epsilon > 0$, 总存在一方阵范数 $\|A\|_*$, 使

$$\|A\|_* < \rho(A) + \epsilon$$

证明 对于 $A \in C^{n \times n}$, 必有可逆阵 P , 使

$$P^{-1}AP = J = \begin{pmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_s & & \\ & & & \ddots & \\ & & & & \lambda_s \end{pmatrix}^T$$

为 Jordan 标准形, $\lambda_1, \dots, \lambda_s$ 是 A 的特征值, 故

$$\|\lambda_i\| \leq \rho(A) \quad (i = 1, \dots, s)$$

令 $D = \text{diag}(1, \sqrt{2}/2, \sqrt{2}^2/4, \dots, \sqrt{2}^{n-1}/2^{n-1})$

$$D^{-1}JD = D^{-1}P^{-1}APD = \begin{pmatrix} 1 & & & & \\ & \sqrt{2}/2 & & & \\ & & \sqrt{2}^2/4 & & \\ & & & \ddots & \\ & & & & \sqrt{2}^{n-1}/2^{n-1} \end{pmatrix}^T$$

若对任意的 $B \in C^{n \times n}$, 定义

$$B_* = D^{-1}P^{-1}BPD$$

易证 B_* 是 $C^{n \times n}$ 中的方阵范数, 对此方阵范数有

$$A_* = D^{-1}P^{-1}APD = D^{-1}JD$$

$$\max_i |a_{ii}| + \frac{1}{2} < (A) +$$

定理 5.3.2 设复变幂级数 $\sum_{k=0}^{\infty} a_k z^k$ 的收敛半径为 R , $A \in C^{n \times n}$ 的谱半径为 (A) , 则

(1) 当 $(A) < R$ 时, 方阵幂级数 $\sum_{k=0}^{\infty} a_k A^k$ 绝对收敛;

(2) 当 $(A) > R$ 时, 方阵幂级数 $\sum_{k=0}^{\infty} a_k A^k$ 发散。

证明 (1) 因为 $(A) < R$, 故存在 $\epsilon > 0$, 使 $(A) + \epsilon < R$

那么级数

$$\sum_{k=0}^{\infty} |a_k| ((A) + \epsilon)^k$$

收敛, 对于上述 $\epsilon > 0$, 由定理 5.3.1, 存在一方阵范数 $\|\cdot\|_*$, 使

$$A_* < (A) + \epsilon$$

所以 $\|a_k A^k\|_* = |a_k| \|A^k\|_* = |a_k| \|A\|_*^k < |a_k| ((A) + \epsilon)^k$

故级数 $\sum_{k=0}^{\infty} a_k A^k$ 收敛, 由性质 2 知级数

$$\sum_{k=0}^{\infty} a_k A^k$$

绝对收敛。

(2) 若 $(A) > R$, 设 $A = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$ 且

$$(A) = |\lambda_i|, \quad |\lambda_i|^H = 1$$

若方阵级数

$$\sum_{k=0}^{\infty} a_k A^k$$

收敛, 则级数

$$\begin{aligned} \sum_{k=0}^{\infty} a_k A^k &= \sum_{k=0}^{\infty} a_k \sum_{i=1}^n \lambda_i^k e_i e_i^H \\ &= \sum_{k=0}^{\infty} \sum_{i=1}^n a_k \lambda_i^k e_i e_i^H \\ &= \sum_{i=1}^n \left(\sum_{k=0}^{\infty} a_k \lambda_i^k \right) e_i e_i^H \end{aligned}$$

也收敛, 这与假设矛盾, 故当 $\rho(A) > R$ 时, 方阵级数

$$\sum_{k=0}^{\infty} a_k A^k$$

发散。

推论 1 若复变幂级数

$$\sum_{k=0}^{\infty} a_k (z - z_0)^k$$

的收敛半径是 R , 则对于 $A \in \mathbb{C}^{n \times n}$, 当其特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$

$$|\lambda_i - z_0| < R \quad (i = 1, 2, \dots, n)$$

时, 方阵幂级数

$$\sum_{k=0}^{\infty} a_k (A - z_0 E)^k$$

绝对收敛, 若有某一 λ_j , 使

$$|\lambda_j - z_0| > R$$

则上述方阵幂级数发散。

证明 令 $B = A - z_0 E$

则 $\rho(B) = \max_j \{ |\lambda_j - z_0| \}$

及

$$\sum_{k=0}^{\infty} a_k (A - z_0 E)^k = \sum_{k=0}^{\infty} a_k B^k$$

由定理 5.3.2 知, 当 $\rho(B) = \max_i |\lambda_i - z_0| < R$ 时该级数收敛。当有某 λ_j , 使 $|\lambda_j - z_0| > R$ 时, 便有 $\rho(A - z_0 E) > R$, 故该级数发散。

推论 2 若复变幂级数

$$\sum_{k=0}^{\infty} a_k z^k$$

在整个复平面都收敛, 则对任意的 $A \in \mathbb{C}^{n \times n}$, 方阵幂级数

$$\sum_{k=0}^{\infty} a_k A^k$$

都收敛。

5.4 方阵函数

5.4.1 方阵函数 $f(A)$ 定义

最简单的方阵函数是矩阵多项式 $B = f(A) = a_0E + a_1A + \dots + a_nA^n$, 其中 $A \in C^{n \times n}$, $a_i \in C$ 。
在上节讨论的基础上, 本节利用复变幂级数的和函数定义方阵幂级数和函数——方阵函数。

在复变函数论中已知

$$\begin{aligned} e^z &= 1 + z + \frac{1}{2!}z^2 + \dots + \frac{1}{n!}z^n + \dots \\ \sin z &= z - \frac{1}{3!}z^3 + \frac{1}{5!}z^5 - \frac{1}{7!}z^7 + \dots \\ \cos z &= 1 - \frac{1}{2!}z^2 + \frac{1}{4!}z^4 - \frac{1}{6!}z^6 + \dots \end{aligned}$$

在整个复平面上都成立, 这样由定理 5.3.2 知, 对任意方阵 $A \in C^{n \times n}$, 方阵幂级数

$$\begin{aligned} E + A + \frac{1}{2!}A^2 + \dots + \frac{1}{n!}A^n + \dots \\ A - \frac{1}{3!}A^3 + \frac{1}{5!}A^5 - \frac{1}{7!}A^7 + \dots \\ E - \frac{1}{2!}A^2 + \frac{1}{4!}A^4 - \frac{1}{6!}A^6 + \dots \end{aligned}$$

都收敛, 即它们的和都存在, 分别用 e^A , $\sin A$, $\cos A$ 表示其和, 即

$$\begin{aligned} e^A &= E + A + \frac{1}{2!}A^2 + \dots + \frac{1}{n!}A^n + \dots \\ \sin A &= A - \frac{1}{3!}A^3 + \frac{1}{5!}A^5 - \frac{1}{7!}A^7 + \dots \\ \cos A &= E - \frac{1}{2!}A^2 + \frac{1}{4!}A^4 - \frac{1}{6!}A^6 + \dots \end{aligned}$$

并分别称为方阵 A 的指数函数, 正弦函数及余弦函数。同样由

$$\ln(1+z) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} z^k \quad |z| < 1$$

及
$$(1+z)^a = \sum_{k=0}^{\infty} \frac{a(a-1)\dots(a-k+1)}{k!} z^k \quad |z| < 1$$

可以定义方阵函数 $\ln(E+A)$ 及 $(E+A)^a$ (a 为任意实数):

$$\begin{aligned} \ln(E+A) &= \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!} A^k \\ (E+A)^a &= \sum_{k=0}^{\infty} \frac{a(a-1)\dots(a-k+1)}{k!} A^k \end{aligned}$$

但这两个方阵函数只对 $\rho(A) < 1$ 的方阵 A 有意义。

5.4.2 用方阵 A 的若当形计算方阵函数 f(A)

问题是对于给定的方阵 $A \in C^{n \times n}$, 如何求出方阵函数 $e^A, \sin A, \cos A$ 等。解决这一问题的方法有多种, 下面先介绍利用方阵 A 的若当标准形计算方阵函数, 为此先证明两个有关定理。

定理 5.4.1 若方阵 $X \in C^{n \times n}$ 的幂级数 $\sum_{k=0}^{\infty} a_k X^k$ 收敛, 并记

$$f(X) = \sum_{k=0}^{\infty} a_k X^k$$

则当 $X = \text{diag}(X_1, X_2, \dots, X_t)$ 时, 有

$$\begin{aligned} f(X) &= f(\text{diag}(X_1, X_2, \dots, X_t)) \\ &= \text{diag}(f(X_1), f(X_2), \dots, f(X_t)) \end{aligned}$$

证明 $f(X) = f(\text{diag}(X_1, X_2, \dots, X_t))$

$$\begin{aligned} &= \lim_{N \rightarrow \infty} \text{diag}(a_0 X_1^N, a_0 X_2^N, \dots, a_0 X_t^N) \\ &= \text{diag}(\lim_{N \rightarrow \infty} a_0 X_1^N, \lim_{N \rightarrow \infty} a_0 X_2^N, \dots, \lim_{N \rightarrow \infty} a_0 X_t^N) \\ &= \text{diag}(f(X_1), f(X_2), \dots, f(X_t)) \end{aligned}$$

定理 5.4.2 任给收敛半径为 R 的复变幂级数

$$f(z) = \sum_{k=0}^{\infty} a_k z^k$$

及一 n 阶 Jordan 块

$$J_0 = \begin{pmatrix} 0 & & & \\ 1 & 0 & & \\ & 1 & \ddots & \\ & & \ddots & 0 \end{pmatrix}_{n \times n}$$

则当 $|z_0| < R$ 时, 级数

$$\sum_{k=0}^{\infty} a_k J_0^k$$

绝对收敛, 且

$$\begin{aligned} f(J_0) &= \sum_{k=0}^{\infty} a_k J_0^k \\ &= \begin{pmatrix} f(z_0) & & & \\ f'(z_0) & f(z_0) & & \\ \frac{1}{2!} f''(z_0) & f'(z_0) & f(z_0) & \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{(n-1)!} f^{(n-1)}(z_0) & \dots & \frac{1}{2!} f''(z_0) & f'(z_0) & f(z_0) \end{pmatrix} \end{aligned}$$

证明 记

$$f_N(z) = \sum_{k=0}^N a_k z^k, f_N(J_0) = \sum_{k=0}^N a_k J_0^k$$

由于

$$f_N(J_0) = \begin{pmatrix} f_N(0) & & & & \\ f_N(0) & f_N(0) & & & \\ \frac{1}{2!} f_N(0) & f_N(0) & & & \\ \dots & & & & \\ \frac{1}{(n-1)!} f_N^{(n-1)}(0) & \dots & \frac{1}{2!} f_N(0) & f_N(0) & f_N(0) \end{pmatrix}$$

这里

$$\begin{cases} f_N(0) = \sum_{k=0}^N a_k 0^k \\ f_N(0) = \sum_{k=0}^N k a_k 0^{k-1} \\ f_N(0) = \sum_{k=0}^N k(k-1) a_k 0^{k-2} \\ \dots\dots\dots \\ f_N^{(n-1)}(0) = \sum_{k=0}^N k(k-1)\dots(k-n+2) a_k 0^{k-n+1} \end{cases} \tag{5.19}$$

注意到 $|z| < R$ 时, 级数 $\sum_{k=0}^{\infty} a_k z^k$ 绝对收敛, 且收敛于 $f(z)$, 故 $f(z)$ 为 z 的解析函数, 它在 $|z| < R$ 内任意次可微, 且

$$f^{(1)}(z) = \sum_{k=0}^{\infty} k(k-1)\dots(k-1+1) a_k z^{k-1} \quad (1 = 1, 2, \dots)$$

在 $|z| < R$ 内都成立, 从而(5.19) 式中各式在 $N \rightarrow \infty$ 的极限都存在, 且

$$\begin{aligned} \lim_N f_N(0) &= \lim_N \sum_{k=0}^N a_k 0^k = f(0) \\ \lim_N f_N(0) &= \lim_N \sum_{k=0}^N k a_k 0^{k-1} = f'(0) \\ \dots\dots\dots \\ \lim_N f_N^{(n-1)}(0) &= \lim_N \sum_{k=0}^N k(k-1)\dots(k-n+2) a_k 0^{k-n+1} = f^{(n-1)}(0) \end{aligned}$$

于是有

$$f(J_0) = \sum_{k=0}^{\infty} a_k J_0^k = \lim_N \sum_{k=0}^N a_k J_0^k = \lim_N f_N(J_0)$$

$$\begin{aligned} &= \lim_N \begin{bmatrix} f_N(\lambda_0) & & & \\ f_N(\lambda_0) & f_N(\lambda_0) & & W \\ \frac{1}{2!} f_N(\lambda_0) & f_N(\lambda_0) & & W & W \\ \dots & & W & W & W \\ \frac{1}{(n-1)!} f_N^{(n-1)}(\lambda_0) & \dots & \frac{1}{2!} f_N(\lambda_0) & f_N(\lambda_0) & f_N(\lambda_0) \end{bmatrix} \\ &= \begin{bmatrix} f(\lambda_0) & & & \\ f(\lambda_0) & f(\lambda_0) & & W \\ \frac{1}{2!} f(\lambda_0) & f(\lambda_0) & & W & W \\ \dots & & W & W & W \\ \frac{1}{(n-1)!} f^{(n-1)}(\lambda_0) & \dots & \frac{1}{2!} f(\lambda_0) & f(\lambda_0) & f(\lambda_0) \end{bmatrix} \end{aligned}$$

从而定理得证。
有了上面的准备, 现在讨论如何用 A 的标准形计算方阵函数 f(A)。
(1) 当 A 与对角形相似时, 即存在可逆阵 P, 使

$$A = P[\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)]P^{-1}$$

则对于复变幂级数

$$f(z) = \sum_{k=0}^{\infty} a_k z^k \quad |z| < R$$

当 $\rho(A) < R$ 时, 矩阵幂级数

$$f(A) = \sum_{k=0}^{\infty} a_k A^k$$

收敛, 且

$$\begin{aligned} f(A) &= f(P \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) P^{-1}) \\ &= \sum_{k=0}^{\infty} a_k [P \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) P^{-1}]^k \\ &= P \{ \sum_{k=0}^{\infty} a_k [\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)]^k \} P^{-1} \\ &= P[\text{diag}(\sum_{k=0}^{\infty} a_k \lambda_1^k, \sum_{k=0}^{\infty} a_k \lambda_2^k, \dots, \sum_{k=0}^{\infty} a_k \lambda_n^k)] P^{-1} \\ &= P[\text{diag}(f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n))] P^{-1} \end{aligned}$$

分别以 $e^A, \sin A, \cos A$ 代入, 便有

$$e^A = P[\text{diag}(e^{\lambda_1}, e^{\lambda_2}, \dots, e^{\lambda_n})]P^{-1} \tag{5.20}$$

$$\sin A = P[\text{diag}(\sin \lambda_1, \sin \lambda_2, \dots, \sin \lambda_n)]P^{-1} \tag{5.21}$$

$$\cos A = P[\text{diag}(\cos \lambda_1, \cos \lambda_2, \dots, \cos \lambda_n)]P^{-1} \tag{5.22}$$

从而

$$e^A \text{ 的特征值为 } e^{\lambda_1}, e^{\lambda_2}, \dots, e^{\lambda_n}$$

sinA 的特征值为 $\sin \lambda_1, \sin \lambda_2, \dots, \sin \lambda_n$
cosA 的特征值为 $\cos \lambda_1, \cos \lambda_2, \dots, \cos \lambda_n$

例 1 设 $A = \begin{pmatrix} 0 & 1 \\ 0 & -2 \end{pmatrix}$
求 $e^A, \sin A, \cos A$ 。
解 因为

$$|E - A| = (\lambda + 2)$$

所以 A 的特征值是 $\lambda_1 = 0, \lambda_2 = -2$ 。

对 $\lambda_1 = 0$, 求得特征向量 $\alpha_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

对 $\lambda_2 = -2$, 求得特征向量 $\alpha_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$

从而求得

$$P = \begin{pmatrix} 1 & 1 \\ 0 & -2 \end{pmatrix} \quad P^{-1} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & -\frac{1}{2} \end{pmatrix}$$

因此

$$e^A = \begin{pmatrix} 1 & 1 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^{-2} \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2}(1 - e^{-2}) \\ 0 & e^{-2} \end{pmatrix}$$

$$\sin A = \begin{pmatrix} 1 & 1 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} \sin 0 & 0 \\ 0 & \sin(-2) \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} 0 & -\frac{1}{2}\sin 2 \\ 0 & \sin 2 \end{pmatrix}$$

$$\cos A = \begin{pmatrix} 1 & 1 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} \cos 0 & 0 \\ 0 & \cos(-2) \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2}(1 - \cos 2) \\ 0 & \cos 2 \end{pmatrix}$$

在实用上遇到的方阵函数往往不是常数矩阵 A 的函数, 而是变量 t 的函数矩阵 At 的函数, 即尚须计算方阵函数 $e^{At}, \sin At, \cos At$, 计算方法也与上面一样, 即

$$f(At) = f(P \operatorname{diag}(\lambda_1 t, \lambda_2 t, \dots, \lambda_n t) P^{-1})$$

$$\begin{aligned} &= \sum_{k=0}^{\infty} \frac{a_k}{k!} [P \operatorname{diag}(\lambda_1 t, \lambda_2 t, \dots, \lambda_n t) P^{-1}]^k \\ &= P \left\{ \sum_{k=0}^{\infty} \frac{a_k}{k!} [\operatorname{diag}(\lambda_1 t, \lambda_2 t, \dots, \lambda_n t)]^k \right\} P^{-1} \\ &= P \operatorname{diag}(f(\lambda_1 t), f(\lambda_2 t), \dots, f(\lambda_n t)) P^{-1} \end{aligned}$$

从而

$$e^{At} = P \operatorname{diag}(e^{\lambda_1 t}, e^{\lambda_2 t}, \dots, e^{\lambda_n t}) P^{-1} \tag{5.23}$$

sinAt = Pdiag(sin_1t, sin_2t, ..., sin_nt) P^{-1} (5.24)

cosAt = Pdiag(cos_1t, cos_2t, ..., cos_nt) P^{-1} (5.25)

(2) 当 A 不与对角形矩阵相似时, 这时存在可逆阵 P, 使

A = P \begin{pmatrix} J_1(m_1) & & \\ & J_2(m_2) & \\ & & \ddots \\ & & & J_s(m_s) \end{pmatrix} P^{-1}

其中

J_i(m_i) = \begin{pmatrix} \lambda_i & & & \\ 1 & \lambda_i & & \\ & \ddots & \ddots & \\ & & 1 & \lambda_i \end{pmatrix}_{m_i \times m_i}

为 m_i 阶 Jordan 块, 且 m_1 + m_2 + ... + m_s = n, 由定理 5.4.1 及定理 5.4.2

f(A) = P f(diag(J_1(m_1), J_2(m_2), ..., J_s(m_s))) P^{-1} = P diag(f(J_1), f(J_2), ..., f(J_s)) P^{-1}

及

f(J_i) = \begin{pmatrix} f(\lambda_i) & & & \\ f'(\lambda_i) & f(\lambda_i) & & \\ \frac{1}{2!} f''(\lambda_i) & f'(\lambda_i) & f(\lambda_i) & \\ \vdots & \vdots & \vdots & \ddots \\ \frac{1}{(m_i-1)!} f^{(m_i-1)}(\lambda_i) & \dots & \frac{1}{2!} f''(\lambda_i) & f'(\lambda_i) & f(\lambda_i) \end{pmatrix}

便可计算出 f(A)

例 2 设

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

求 e^A。

解 因为 A 的特征矩阵

(E - A) = \begin{pmatrix} \lambda & -1 & 0 \\ 0 & \lambda & -1 \\ -2 & -3 & \lambda \end{pmatrix} = \begin{pmatrix} \lambda & & \\ & \lambda & \\ & & (\lambda - 2)(\lambda + 1)^2 \end{pmatrix}

故 A 的 Jordan 标准形为

J = \begin{pmatrix} 2 & & \\ & -1 & \\ & 1 & -1 \end{pmatrix}

下面求 P, 使 P^{-1}AP = J 或 AP = PJ, 令

P = (p_1, p_2, p_3)

则

$$A(\alpha_1, \alpha_2, \alpha_3) = (\alpha_1, \alpha_2, \alpha_3) \begin{pmatrix} 2 & & \\ & -1 & \\ & 1 & -1 \end{pmatrix}$$
$$(A\alpha_1, A\alpha_2, A\alpha_3) = (2\alpha_1, -\alpha_2 + \alpha_3, -\alpha_3)$$

可得下列方程组

$$(2E - A)\alpha_1 = 0$$
$$(E + A)\alpha_2 = \alpha_3$$
$$(E + A)\alpha_3 = 0$$

解出

$$\alpha_1 = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} \quad \alpha_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \quad \alpha_3 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

可求出

$$P = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 0 & -1 \\ 4 & -1 & 1 \end{pmatrix} \quad P^{-1} = \frac{1}{9} \begin{pmatrix} 1 & 2 & 1 \\ 6 & 3 & -3 \\ 2 & -5 & 2 \end{pmatrix}$$

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

那么

$$e^{At} = \frac{1}{9} \begin{pmatrix} 1 & 1 & 1 \\ 2 & 0 & -1 \\ 4 & -1 & 1 \end{pmatrix} \begin{pmatrix} e^{2t} & & \\ & e^{-t} & \\ & e^{-t} & e^{-t} \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 6 & 3 & -3 \\ 2 & -5 & 2 \end{pmatrix}$$
$$= \frac{1}{9} \begin{pmatrix} e^{2t} + 14e^{-t} & 2e^{2t} + e^{-t} & e^{2t} - 4e^{-t} \\ 2e^{2t} - 8e^{-t} & 4e^{2t} + 2e^{-t} & 2e^{2t} + e^{-t} \\ 4e^{2t} + 2e^{-t} & 8e^{2t} - 5e^{-t} & 4e^{2t} + 2e^{-t} \end{pmatrix}$$

因为

$$At = \begin{pmatrix} 0 & t & 0 \\ 0 & 0 & t \\ 2t & 3t & 0 \end{pmatrix} = P \begin{pmatrix} 2t & & \\ & -t & \\ & t & -t \end{pmatrix} P^{-1}$$

而

$$J_1(2)t = 2t \quad J_2(-1)t = \begin{pmatrix} -t & \\ t & -t \end{pmatrix}$$

所以

$$e^{J_1(2)t} = e^{2t} \quad e^{J_2(-1)t} = \begin{pmatrix} e^{-t} & \\ te^{-t} & e^{-t} \end{pmatrix}$$
$$e^{At} = \frac{1}{9} \begin{pmatrix} 1 & 1 & 1 \\ 2 & 0 & -1 \\ 4 & -1 & 1 \end{pmatrix} \begin{pmatrix} e^{2t} & & \\ & e^{-t} & \\ & te^{-t} & e^{-t} \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 6 & 3 & -3 \\ 2 & -5 & 2 \end{pmatrix}$$

$$= \frac{1}{9} \begin{bmatrix} e^{2t} + (8 + 6t)e^{-t} & 2e^{2t} + (3t - 2)e^{-t} & e^{2t} - (1 + 3t)e^{-t} \\ 2e^{2t} - (2 + 6t)e^{-t} & 4e^{2t} + (5 - 3t)e^{-t} & 2e^{2t} + (3t - 2)e^{-t} \\ 4e^{2t} + (6t - 4)e^{-t} & 8e^{2t} + (3t - 8)e^{-t} & 4e^{2t} + (5 - 3t)e^{-t} \end{bmatrix}$$

5.4.3 用 $f(z)$ 在 A 上的谱值计算方阵函数 $f(A)$

为了避免计算 P 及 J , 现在另觅途径。首先让我们分析一下, 如何对 $f(A) = \sum_{k=0}^{\infty} a_k A^k$ 加以变形, 以达到求出 $f(A)$ 的目的。

设 $h(\lambda)$ 是有限次多项式, $m(\lambda)$ 是方阵 A 的最小多项式(不妨设 $\deg[m(\lambda)] = t$), 用 $m(\lambda)$ 去除 $h(\lambda)$, 其商为 $g(\lambda)$, 余式为 $r(\lambda)$, 便有

$$h(\lambda) = m(\lambda)g(\lambda) + r(\lambda)$$

这里 $\deg[r(\lambda)] \leq t - 1$, 或 $r(\lambda) = 0$, 由 $m(A) = 0$, 有

$$h(A) = m(A)g(A) + r(A)$$

上式说明方阵 A 的任意一个多项式 $h(A)$ 总可以表示为 A 的次数不超过 $t - 1$ 的多项式 $r(A)$ (其中 t 是 A 的最小多项式 $m(\lambda)$ 的次数)。具体地说, 方阵 A 的任何有限次多项式 $h(A)$ 都可以被 E, A, \dots, A^{t-1} 线性表示, 且 E, A, \dots, A^{t-1} 是线性无关的(否则与 $m(\lambda)$ 是 A 的最小多项式矛盾), 所以上述表示法中的 $r(A)$ 还是惟一的。

既然方阵函数 $f(A) = \sum_{k=0}^{\infty} a_k A^k$ 的幂级数表达式收敛, 而 $A^k (k \geq t)$ 又可表示为 A 的次数不超过 $t - 1$ 的多项式, 则可望把 $f(A)$ 表示成次数不超过 $t - 1$ 的方阵多项式 $T(A)$ 。先给定义。

定义 5.9 设 n 阶方阵 A 的最小多项式为

$$m(\lambda) = (\lambda - \lambda_1)^{t_1} (\lambda - \lambda_2)^{t_2} \dots (\lambda - \lambda_s)^{t_s}$$

其中 $\lambda_1, \lambda_2, \dots, \lambda_s$ 是 A 的互不相同的特征根。如果复函数 $f(z)$ 及其各阶导数 $f^{(i)}(z)$ 在 $z = \lambda_i (i = 1, 2, \dots, s)$ 处的导数值, 即

$$f^{(i)}(\lambda_i) = \left. \frac{d^i f(z)}{dz^i} \right|_{z=\lambda_i} \quad \left[\begin{matrix} i = 1, 2, \dots, s \\ l = 0, 1, \dots, t_i - 1 \end{matrix} \right]$$

均为有限值, 便称函数 $f(z)$ 在方阵 A 的谱上给定, 并称这些值为 $f(z)$ 在 A 上的谱值。

定理 5.4.3 设 $A \in C^{n \times n}$ 的最小多项式为

$$m(\lambda) = (\lambda - \lambda_1)^{t_1} (\lambda - \lambda_2)^{t_2} \dots (\lambda - \lambda_s)^{t_s}$$

其中 $t_1 + t_2 + \dots + t_s = t$, $\lambda_i \neq \lambda_j (i \neq j, i, j = 1, 2, \dots, s)$, 而方阵函数 $f(A)$ 是收敛的方阵幂级数 $\sum_{k=0}^{\infty} a_k A^k$ 的和函数, 即

$$f(A) = \sum_{k=0}^{\infty} a_k A^k$$

设 $T(\lambda) = b_0 + b_1 \lambda + \dots + b_{t-1} \lambda^{t-1}$, 使 $f^{(i)}(\lambda_i) = T^{(i)}(\lambda_i) \quad \left[\begin{matrix} i = 1, 2, \dots, s \\ l = 0, 1, \dots, t_i - 1 \end{matrix} \right]$

则
$$T(A) = f(A) = \sum_{k=0}^{\infty} a_k A^k$$

证明略

因为

$$\begin{vmatrix} -2 & -1 & -4 \\ 0 & -2 & 0 \\ 0 & -3 & -1 \end{vmatrix} = (-2)^2(-1) - 1 = -5, \quad \lambda_1 = 2, \quad \lambda_2 = 1$$

容易验证 $(A - 2E)(A - E) = 0$, 所以 A 的最小多项式

$$m(\lambda) = (\lambda - 2)^2(\lambda - 1)$$

可设

$$e^{Az} = b_0(z)E + b_1(z)A + b_2(z)A^2$$

由此可得线性方程组

$$\begin{cases} b_0(z) + b_1(z) \lambda_1 + b_2(z) \lambda_1^2 = e^{\lambda_1 z} \\ b_1(z) + 2b_2(z) \lambda_1 = z e^{\lambda_1 z} \\ b_0(z) + b_1(z) \lambda_2 + b_2(z) \lambda_2^2 = e^{\lambda_2 z} \end{cases} \quad \text{或} \quad \begin{cases} b_0(z) + 2b_1(z) + 4b_2(z) = e^{2z} \\ b_1(z) + 4b_2(z) = z e^{2z} \\ b_0(z) + b_1(z) + b_2(z) = e^z \end{cases}$$

解出此方程组, 得

$$\begin{aligned} b_0 &= 4e^z - 3e^{2z} + 2ze^{2z} \\ b_1 &= -4e^z + 4e^{2z} - 3ze^{2z} \\ b_2 &= e^z - e^{2z} + ze^{2z} \end{aligned}$$

所以

$$\begin{aligned} e^{Az} &= (4e^z - 3e^{2z} + 2ze^{2z})E + (-4e^z + 4e^{2z} - 3ze^{2z})A + (e^z - e^{2z} + ze^{2z})A^2 \\ &= \begin{pmatrix} e^{2z} & 12e^z - 12e^{2z} + 13ze^{2z} & -4e^z + 4e^{2z} \\ 0 & e^{2z} & 0 \\ 0 & -3e^z + 3e^{2z} & e^z \end{pmatrix} \end{aligned}$$

类似地, 可设

$$\sin Az = b_0(z)E + b_1(z)A + b_2(z)A^2$$

可得方程组

$$\begin{cases} b_0(z) + 2b_1(z) + 4b_2(z) = \sin 2z \\ b_1(z) + 4b_2(z) = z \cos 2z \\ b_0(z) + b_1(z) + b_2(z) = \sin z \end{cases}$$

解出此方程组, 得

$$\begin{aligned} b_0 &= 4\sin z - 3\sin 2z + 2z \cos 2z \\ b_1 &= -4\sin z + 4\sin 2z - 3z \cos 2z \\ b_2 &= \sin z - \sin 2z + z \cos 2z \end{aligned}$$

于是

$$\sin Az = \begin{pmatrix} \sin 2z & 12\sin z - 12\sin 2z + 13z \sin 2z & -4\sin z + 4\sin 2z \\ 0 & \sin 2z & 0 \\ 0 & -3\sin z + 3\sin 2z & \sin z \end{pmatrix}$$

例 5 设 $A = \begin{pmatrix} 5 & -4 \\ 4 & -3 \end{pmatrix}$ 计算 A^{100} 。

解 设 $f(\lambda) = \lambda^{100}$ $f(A) = A^{100}$
因为

$$|E - A| = (\lambda - 1)^2, \lambda_1 = 1$$

由于 $\lambda - 1$ 不是 A 的化零多项式, 所以 A 的最小多项式为

$$m(\lambda) = (\lambda - 1)^2$$

故可设

$$A^{100} = a_0E + a_1A$$

由此可得方程组

$$\begin{cases} a_0 + a_1 = 1 \\ a_1 = 100 \end{cases}$$

解得 $a_1 = 100, a_0 = -99$

于是

$$A^{100} = -99E + 100A = \begin{pmatrix} 401 & -400 \\ 400 & -399 \end{pmatrix}$$

例 6 设 $A = \begin{pmatrix} -2 & 2 & -2 & 4 \\ -1 & 2 & -1 & 1 \\ 0 & 0 & 1 & 0 \\ -2 & 1 & -1 & 4 \end{pmatrix}$ 计算 $\cos A$ 。

解 设 $f(\lambda) = \cos \lambda, f(A) = \cos A$, 经过计算可知 A 的最小多项式

$$m(\lambda) = (\lambda - 1)^2(\lambda - 2), \lambda_1 = 1, \lambda_2 = 2$$

可设 $T(\lambda) = b_0 + b_1\lambda + b_2\lambda^2$

可得方程组

$$\begin{cases} b_0 + b_1 + b_2 = \cos 1 = -1 \\ b_1 + 2b_2 = -\sin 1 = 0 \\ b_0 + 2b_1 + 4b_2 = \cos 2 = 1 \end{cases}$$

解出 $b_0 = 1, b_1 = -4, b_2 = 2$

于是

$$\cos A = E - 4A + 2A^2 = \begin{pmatrix} -3 & 0 & 0 & 4 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ -2 & 0 & 0 & 3 \end{pmatrix}$$

5.5 常用方阵函数的一些性质

本节主要讨论常用方阵函数 $e^A, \sin A, \cos A$ 的一些性质。虽然这些矩阵函数有些性质与普通复变函数一样, 但由于矩阵相乘的不可交换性, 从而使得方阵函数的某些性质与一般复变函数的性质也有一些不同的地方, 必须注意这些不同点。对于方阵函数, 不可随便套用普通函

数的性质。

性质 1 对任意的 $A \in C^{n \times n}$, 都有

$$\frac{d}{dz} e^{Az} = A \cdot e^{Az} = e^{Az} \cdot A \tag{5.29}$$

证明 因为

$$\begin{aligned} e^{Az} &= \sum_{k=0}^{\infty} \frac{1}{k!} (Az)^k \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} A^k z^k \\ (e^{Az})_{ij} &= \sum_{k=0}^{\infty} \frac{1}{k!} (A^k)_{ij} z^k \end{aligned}$$

对任何 z 值都是收敛的, 因而可以逐项微分

$$\begin{aligned} \frac{d}{dz} e^{Az} &= \sum_{k=1}^{\infty} \frac{1}{(k-1)!} A^k z^{k-1} \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} A^{k+1} z^k \\ &= \begin{cases} A \sum_{k=0}^{\infty} \frac{1}{k!} A^k z^k = Ae^{Az} \\ (\sum_{k=0}^{\infty} \frac{1}{k!} A^k z^k) \cdot A = e^{Az} \cdot A \end{cases} \end{aligned}$$

性质 2 若 $A, B \in C^{n \times n}$, 且 $AB = BA$, 则

$$e^{Az} \cdot B = Be^{Az} \tag{5.30}$$

证明

$$\begin{aligned} e^{Az} \cdot B &= (\sum_{k=1}^{\infty} \frac{1}{k!} A^k z^k) \cdot B \\ &= \sum_{k=1}^{\infty} \frac{1}{k!} z^k A^k B \\ &= \sum_{k=1}^{\infty} \frac{1}{k!} z^k BA^k \\ &= B(\sum_{k=1}^{\infty} \frac{1}{k!} z^k A^k) \\ &= Be^{Az} \end{aligned}$$

性质 3 若 $A, B \in C^{n \times n}$, 且 $AB = BA$, 则

$$e^A \cdot e^B = e^B \cdot e^A = e^{A+B} \tag{5.31}$$

证明 令

$$C(z) = e^{(A+B)z} \cdot e^{-Az} \cdot e^{-Bz}$$

由性质 1 和性质 2 有

$$\frac{d}{dz} C(z) = (A+B)e^{(A+B)z} \cdot e^{-Az} \cdot e^{-Bz} +$$

$$\begin{aligned} & e^{(A+B)z} \cdot (-A) e^{-Az} \cdot e^{-Bz} + e^{(A+B)z} \cdot e^{-Az} \cdot (-B) e^{-Bz} \\ &= (A+B) e^{(A+B)z} \cdot e^{-Az} \cdot e^{-Bz} - A e^{(A+B)z} \cdot e^{-Az} e^{-Bz} - B e^{(A+B)z} \cdot e^{-Az} \cdot e^{-Bz} \\ &= (A+B-A-B) e^{(A+B)z} \cdot e^{-Az} \cdot e^{-Bz} = 0 \end{aligned}$$

故有

$$\frac{d}{dz}C(z)_{ij} = 0$$

这说明 $C(z)$ 的各元素与 z 无关, 为常数, 所以

$$C(1) = C(0)$$

注意到对于零方阵 0

$$e^0 = \text{diag}(e^0, e^0, \dots e^0) = E$$

所以

$$C(0) = e^{(A+B)0} \cdot e^{-A0} \cdot e^{-B0} = E$$

从而

$$C(1) = e^{(A+B)} \cdot e^{-A} \cdot e^{-B} = E \tag{5.32}$$

若取 $B = -A$, 由(5.30) 式

$$e^{-A} \cdot e^A = E$$

从而对任意方阵 A , 有

$$(e^A)^{-1} = e^{-A}$$

这样就有

$$e^{A+B} e^{-A} e^{-B} = e^{A+B} (e^A)^{-1} (e^B)^{-1} = E$$

也即

$$e^{A+B} = e^B \cdot e^A$$

而

$$e^{A+B} = e^{B+A} = e^A e^B$$

所以

$$e^{A+B} = e^A \cdot e^B = e^B \cdot e^A$$

由上面的证明过程可得如下两个结果

$$(1) e^0 = E \tag{5.33}$$

对任意的 $A \in C^{n \times n}$, e^A 都可逆, 且

$$(2) (e^A)^{-1} = e^{-A} \tag{5.34}$$

这里需注意的是: 当 $AB \neq BA$ 时, 性质 3: $e^{A+B} = e^A \cdot e^B = e^B \cdot e^A$ 不一定成立

例如 令 $A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, B = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$

易知 $A^2 = A, B^2 = B$ 。从而可得

$$A = A^2 = A^3 = \dots \quad B = B^2 = B^3 = \dots$$

于是得

$$e^A = E + (e - 1)A = \begin{pmatrix} e & e - 1 \\ 0 & 1 \end{pmatrix}$$

$$e^B = E + (e - 1)B = \begin{pmatrix} e & 1 - e \\ 0 & 1 \end{pmatrix}$$

因此

$$e^A \cdot e^B = \begin{pmatrix} e^2 & -(e - 1)^2 \\ 0 & 1 \end{pmatrix}, e^B \cdot e^A = \begin{pmatrix} e^2 & (e - 1)^2 \\ 0 & 1 \end{pmatrix}$$

又由 $A + B = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$, 可得 $(A + B)^2 = 2(A + B)$, $(A + B)^k = 2^{k-1}(A + B)$ ($k = 1, 2, \dots$), 由此容易推出

$$e^{A+B} = E + \frac{1}{2}(e^2 - 1)(A + B) = \begin{pmatrix} e^2 & 0 \\ 0 & 1 \end{pmatrix}$$

可见 $e^A \cdot e^B, e^B \cdot e^A$ 及 e^{A+B} 互不相等。

利用绝对收敛级数的性质, 容易推得

性质 4 对任意 $A \in C^{n \times n}$, 有

$$e^{iA} = \cos A + i \sin A \tag{5.35}$$

$$\cos A = \frac{1}{2}(e^{iA} + e^{-iA}) \tag{5.36}$$

$$\sin A = \frac{1}{2i}(e^{iA} - e^{-iA}) \tag{5.37}$$

$$\cos(-A) = \cos A \tag{5.38}$$

$$\sin(-A) = -\sin A \tag{5.39}$$

性质 5 对任意的 $A, B \in C^{n \times n}$, 且 $AB = BA$, 有

$$\cos(A + B) = \cos A \cos B - \sin A \sin B \tag{5.40}$$

$$\sin(A + B) = \sin A \cos B + \cos A \sin B \tag{5.41}$$

性质 6 对任意 $A \in C^{n \times n}$

$$\sin^2 A + \cos^2 A = E \tag{5.42}$$

性质 7 对任意 $A \in C^{n \times n}$

$$\sin(A + 2E) = \sin A \tag{5.43}$$

$$\cos(A + 2E) = \cos A \tag{5.44}$$

由前面性质 3, 我们知道对任意矩阵 A, e^A 总是可逆的, 但 $\sin A$ 与 $\cos A$ 却不一定可逆, 尽管 A 是可逆的。例如给定的 A 的特征值有 $\frac{\pi}{2}$, 且 A 的特征值都不为 0, 则 $\sin A$ 必有一特征值 $\sin \frac{\pi}{2} = 0$, 故 $\sin A$ 不可逆, 同理当 A 有一特征值 $= \frac{\pi}{2}$ 时, $\cos A$ 也不可逆。

$$\begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0) = \left. \frac{d}{dz} \right|_{z=0} = \begin{pmatrix} x_1(0) \\ x_2(0) \\ \cdots \\ x_n(0) \end{pmatrix}, \quad \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0) = \left. \frac{d^2}{dz^2} \right|_{z=0} = \begin{pmatrix} x_1(0) \\ x_2(0) \\ \cdots \\ x_n(0) \end{pmatrix}$$

但由

$$\begin{aligned} \frac{d}{dz} &= A \\ \frac{d^2}{dz^2} &= A^2 \\ \frac{d^3}{dz^3} &= A^3 \end{aligned}$$

及

$$\begin{aligned} \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0) &= A \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0) \\ \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0) &= A^2 \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0) \\ &\dots\dots\dots \end{aligned}$$

从而

$$\begin{aligned} &= \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0) + \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0) z + \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0) \frac{z^2}{2!} + \dots \\ &= \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0) + A \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0) z + A^2 \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0) \frac{z^2}{2!} + \dots \\ &= e^{Az} \cdot \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0) \end{aligned}$$

这说明微分方程组

$$\frac{d}{dz} = A$$

满足初始条件

$$\begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0) = \begin{pmatrix} c_1 \\ c_2 \\ \cdots \\ c_n \end{pmatrix}$$

的解必有

$$= e^{Az} \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0)$$

的形式。可以证明它确实是解, 由于

$$\begin{aligned} \frac{d}{dz} &= \frac{d}{dz} (e^{Az} \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0)) \\ &= (\frac{d}{dz} (e^{Az})) \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0) + e^{Az} [\frac{d}{dz} \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0)] \\ &= A e^{Az} \cdot \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix} (0) = A \end{aligned}$$

这样便证明了如下的定理

定理 5.6.1 满足初始条件

$$\begin{pmatrix} 0 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \dots \\ c_n \end{pmatrix}$$

的一阶线性常系数齐次微分方程组

$$\frac{d}{dz} = A$$

有且有惟一解 $= e^{Az} \cdot \begin{pmatrix} 0 \end{pmatrix}$ 。

对于初值问题

$$\begin{cases} \frac{d}{dz} = A \\ |_{z=z_0} = \begin{pmatrix} z_0 \end{pmatrix} \end{cases}$$

可将 $x_i(z) (i = 1, 2, \dots, n)$ 在 $z = z_0$ 处展成幂级数。

$$x_i(z) = x_i(z_0) + x_i(z_0)(z - z_0) + x_i(z_0)(z - z_0)^2/2! + \dots$$

便可得到该初值问题的惟一解

$$\begin{pmatrix} z \end{pmatrix} = e^{A(z-z_0)} \begin{pmatrix} z_0 \end{pmatrix}$$

由此看出, 解线性常系数齐次微分方程组, 实际上只需求出 e^{Az} 或 $e^{A(z-z_0)}$ 即可。

5.6.2 解一阶线性常系数非齐次方程组

定理 5.6.2 线性常系数非齐次微分方程组

$$\begin{cases} \frac{d}{dz} = A + \begin{pmatrix} z \end{pmatrix} \\ |_{z=z_0} = \begin{pmatrix} z_0 \end{pmatrix} \end{cases}$$

的解为

$$= e^{A(z-z_0)} \begin{pmatrix} z_0 \end{pmatrix} + \int_{z_0}^z e^{-A(z-z)} \begin{pmatrix} \end{pmatrix} d$$

其中

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}, \quad \begin{pmatrix} z \end{pmatrix} = \begin{pmatrix} x_1(z) \\ x_2(z) \\ \dots \\ x_n(z) \end{pmatrix} \quad \begin{pmatrix} z \end{pmatrix} = \begin{pmatrix} F_1(z) \\ F_2(z) \\ \dots \\ F_n(z) \end{pmatrix} \quad \begin{pmatrix} z_0 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \dots \\ c_n \end{pmatrix}$$

证明 可将方程组改写成

$$\frac{d}{dz} - A = \begin{pmatrix} z \end{pmatrix}$$

两端同乘以 e^{-Az}

$$e^{-Az} \cdot \left(\frac{d}{dz} - A \right) = e^{-Az} \cdot \begin{pmatrix} z \end{pmatrix}$$

在 $[z_0, z]$ 上对上式积分, 得

$$e^{-Az} - e^{-Az_0} \begin{pmatrix} z_0 \end{pmatrix} = \int_{z_0}^z e^{-A(z-z)} \begin{pmatrix} \end{pmatrix} d$$

所以

$$= e^{A(z-z_0)}(z_0) + \int_{z_0}^z e^{-A(z-z)}(z) dz$$

习 题 5

- 1. 构造一个 2×2 阶的可逆矩阵序列, 其收敛于不可逆的极限。
- 2. 证明: 当且仅当 $A = E$ 时, 有 $\lim_n A^n = E$ 。
- 3. 计算下面级数的和

$$\sum_{k=0}^{\infty} \begin{pmatrix} 0.2 & 0.7 \\ 0.3 & 0.6 \end{pmatrix}^k$$

- 4. 证明: 若 $A \in C^{n \times n}$, $\rho(A) < 1$, 则 $\sum_{k=0}^{\infty} kA^k = A(E - A)^{-2}$
- 5. 计算

$$\sum_{k=0}^{\infty} \frac{k}{10^k} \begin{pmatrix} 1 & 2 \\ 8 & 1 \end{pmatrix}^k$$

- 6. 设函数矩阵

$$A(t) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$$

求 $\frac{d}{dt}A(t)$, $\frac{d}{dt}[\det A(t)]$, $\det[\frac{d}{dt}A(t)]$, $\frac{d}{dt}A^{-1}(t)$ 。

- 7. 设函数矩阵

$$A(t) = \begin{pmatrix} \sin t & -\cos t \\ -\cos t & \sin t \end{pmatrix}$$

求 $\int_0^t A(t) dt$, $\frac{d}{dt} \int_0^{t^2} A(t) dt$ 。

- 8. 设函数矩阵

$$A(t) = \begin{pmatrix} e^{2t} & te^t & 1 \\ e^{-t} & 2e^{2t} & 0 \\ 3t & 0 & 0 \end{pmatrix}$$

求 $\int_0^t A(t) dt$, $\int_0^t A(t) dt$ 。

- 9. 假设下列各式中的乘积都有意义, 证明:

(1) $\frac{d}{dX} \text{tr}(AXB) = BA$

(2) $\frac{d}{dX} \text{tr}(AX^T B) = (BA)^T$

(3) $\frac{d}{dX} \text{tr}(X^2) = 2X$

- (4) $\frac{d}{dX} \text{tr}(X^T X) = \frac{d}{dX} \text{tr}(XX^T) = 2X^T$
- (5) $\frac{d}{dX} \text{tr}(AXBX^T) = \frac{d}{dX} \text{tr}(BX^T AX) = BX^T A + B^T X^T A^T$
10. 设 α, β, γ 都是维数相同的列向量, 且 α, β 都是与 γ 无关的固定向量, 证明

$$\frac{d}{d}(\alpha - \beta)^T \cdot \gamma = \gamma^T$$

11. 证明, 对任何矩阵 A 都有
- (1) $\sin A \cos A = \cos A \sin A$
- (2) $\sin^2 A + \cos^2 A = E$
- (3) $e^{A+2iE} = e^A$
- (4) $\sin(A+2E) = \sin A$
12. 证明公式

$$e^{\begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix}} = \begin{pmatrix} \cos a & \sin a \\ -\sin a & \cos a \end{pmatrix}$$

13. 设 $A = \begin{pmatrix} & \\ & \end{pmatrix}$, 利用上题结果求 e^A 。
14. 设矩阵

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

- 求 $A^{100} + 3A^{23} + A^{20}$ 。
15. 设矩阵

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

- 求 A^{1000} 。
16. 已知四阶矩阵 A 的特征值是 $1, -1, 0, 0$ 。求 $\sin A, \cos A$ 。
17. 设 $f(x) = \frac{1}{x}$, 求 $f(A)$, 其中

(1) $A = \begin{pmatrix} 2 & & & \\ 1 & 2 & & \\ & 1 & 2 & \\ & & 1 & 2 \end{pmatrix}$ (2) $A = \begin{pmatrix} 2 & & & \\ 1 & 2 & & \\ & & 3 & \\ & & & 1 \end{pmatrix}$

18. 设矩阵 $A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ & 1 & 2 & 3 \\ & & 1 & 2 \\ & & & 1 \end{pmatrix}$

- 求 \sqrt{A} 。
19. 对下列方阵 A , 求方阵函数 e^{Az} 。

(1) $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{pmatrix}$

(2) $A = \begin{pmatrix} 3 & & & \\ & -2 & & \\ & 1 & -2 & \\ & & 1 & -2 \end{pmatrix}$

(3) $A = \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix}$

20. 求线性常系数齐次微分方程组

$$\begin{cases} \frac{dx_1}{dz} = -7x_1 - 7x_2 + 5x_3 \\ \frac{dx_2}{dz} = -8x_1 - 8x_2 - 5x_3 \\ \frac{dx_3}{dz} = -5x_2 \end{cases}$$

满足初始条件 $x_1(0) = 3, x_2(0) = -2, x_3(0) = 1$ 的解。

21. 求常系数非齐次微分方程组

$$\begin{cases} \frac{dx_1}{dz} = -6x_1 + x_2 + 2u(z) \\ \frac{dx_2}{dz} = -11x_1 + x_3 + 6u(z) \\ \frac{dx_3}{dz} = -6x_1 + 2u(z) \end{cases}$$

满足初始条件 $x_1(0) = 2, x_2(0) = 4, x_3(0) = -2$ 的解, 其中

$$u(z) = \begin{cases} 1 & z \geq 0 \\ 0 & z < 0 \end{cases}$$

第 6 章

矩 阵 分 解

本章首先讨论以 Gauss 消去法为根据导出的矩阵的各种分解,特别是三角(或 LU)分解。然后论述与之可以媲美的是 20 世纪 60 年代以后由 Given 与 Householder 变换发展起来的矩阵的 QR 分解。所有这些分解在《计算数学》中都已扮演着十分重要的角色,尤其是以 QR 分解所建立的 QR 方法,已对《数值线性代数》理论在近代的发展起了关键作用。最后介绍在广义逆矩阵等理论中,经常遇到的矩阵的满秩分解和奇异值分解,它与 QR 方法都是求解各类最小二乘问题和最优化问题等的主要数学工具。

6.1 Gauss 消去法与矩阵的三角分解

6.1.1 Gauss 消去法的矩阵表示

读者已经学习过解 n 元线性方程组

[illegible]

的 Gauss 主元素消去法, 写成矩阵形式, 即

$$A = \quad (6.2)$$

其中 $A = (a_{ij})_{n \times n}$, $X = (x_1, x_2, \dots, x_n)^T$, $b = (b_1, b_2, \dots, b_n)^T$ 。这种方法的基本思想是化系数矩阵 A 为上三角阵, 或化增广矩阵 $(A|b)$ 为上阶梯形矩阵以求其解。这种消去法有三种形式, 即按自然顺序(按主对角元的顺序)选主元素法, 按列选主元素法以及总体选主元素法。这些方法各有千秋, 不可偏废。

为了建立矩阵的三角分解理论, 我们使用矩阵理论描写以上所说的消元法的消元过程, 并假定化 A 为上三角矩阵的过程未用列交换, 即采用按自然顺序选主元进行消元。

设 $A^{(0)} = A$, 其元素 $a_{ij}^{(0)} = a_{ij}$ ($i, j = 1, 2, \dots, n$)。记 A 的 k 阶顺序主子式为

$$\Delta_k = \det A \begin{pmatrix} 1 & 2 \dots k \\ 1 & 2 \dots k \end{pmatrix} \quad k = 1, 2, \dots, n - 1 \tag{6.3}$$

且假定 $a_{11}^{(0)} \neq 0$, 则可以作一系列行初等变换, 即第一行的 $(-\frac{a_{i1}}{a_{11}})$ 倍加到第 i 行上, ($i = 2, 3, \dots, n$), 可以把第一列除 a_{11} 以外的元变为 0, 即令

$$L_1^{-1} = \begin{pmatrix} 1 & & & \\ -\frac{a_{21}}{a_{11}^{(0)}} & 1 & & \\ \dots & & \ddots & \\ -\frac{a_{n1}}{a_{11}^{(0)}} & \dots & \dots & 1 \end{pmatrix}$$

去左乘以 $A^{(0)}$, 于是有

$$L_1^{-1} A^{(0)} = \begin{pmatrix} a_{11}^{(0)} & a_{12}^{(0)} & \dots a_{1n}^{(0)} \\ 0 & a_{22}^{(1)} & \dots a_{2n}^{(1)} \\ 0 & a_{n2}^{(1)} & \dots a_{nn}^{(1)} \end{pmatrix} = A^{(1)} \tag{6.4}$$

即

$$A^{(0)} = L_1 A^{(1)}$$

其中

$$L_1 = \begin{pmatrix} 1 & & & \\ \frac{a_{21}^{(0)}}{a_{11}^{(0)}} & 1 & & \\ \dots & & \ddots & \\ \frac{a_{n1}^{(0)}}{a_{11}^{(0)}} & \dots & 0 & 1 \end{pmatrix} \tag{6.5}$$

因为上述的初等变换不改变矩阵 A 的行列式的值, 所以由 $A^{(1)}$ 得 A 的二阶顺序主子式是

$$\Delta_2 = a_{11}^{(0)} a_{22}^{(1)} \tag{6.6}$$

设 $a_{22}^{(1)} \neq 0$, 如此继续做下去, 直到第 $r - 1$ 步, 得到

$$A^{(r-1)} = \begin{pmatrix} a_{11}^{(0)} & \dots & a_{1r-1}^{(0)} & a_{1r}^{(0)} & \dots & a_{1n}^{(0)} \\ 0 & a_{22}^{(1)} & \dots & & & \\ & & a_{r-1\ r-1}^{(r-2)} & a_{r-1\ r}^{(r-2)} & \dots & a_{r-1\ n}^{(r-2)} \\ & & 0 & a_{rr}^{(r-1)} & \dots & a_{rn}^{(r-1)} \\ \dots & \dots & \dots & \dots & & \dots \\ 0 & & 0 & a_{nr}^{(r-1)} & \dots & a_{nn}^{(r-1)} \end{pmatrix}$$

对于第 r 步, 假定 $a_{rr}^{(r-1)} \neq 0$, 同样令

$$L_r = \begin{pmatrix} 1 & & & & \\ & w & & & \\ & & 1 & & \\ & & & 1 & \\ & & & \frac{a_{(r+1)r}^{(r-1)}}{a_{rr}^{(r-1)}} & 1 \\ \dots & w & & \frac{a_{nr}^{(r-1)}}{a_{rr}^{(r-1)}} & 1 \end{pmatrix} \tag{6.7}$$

于是有

$$L_r^{-1} A^{(r-1)} = \begin{pmatrix} a_{11}^{(0)} & \dots & a_{1r}^{(0)} & a_{1r+1}^{(0)} & \dots & a_{1n}^{(0)} \\ 0 & w & \dots & & & \\ \dots & & a_{rr}^{(r-1)} & a_{r,r+1}^{(r-1)} & \dots & a_{rn}^{(r-1)} \\ \dots & & 0 & a_{r+1,r+1}^{(r)} & \dots & a_{(r+1)n}^{(r)} \\ \dots & & 0 & \dots & & \\ \dots & & \dots & \dots & & \\ 0 & 0 & a_{n(r+1)}^{(r)} & \dots & a_{nn}^{(r)} \end{pmatrix} = A^{(r)} \tag{6.8}$$

$A^{(r)}$ 的前 r 个列,主元素 $a_{ii}^{(i-1)}$ ($i = 1, 2, \dots, r$) 以下的元素全为 0, (6.8) 式还可以写为

$$A^{(r-1)} = L_r A^{(r)} \tag{6.9}$$

由 $A^{(r)}$ 易得 A 的 r 阶顺序主子式是

$$r = a_{11}^{(0)} a_{22}^{(1)} \dots a_{rr}^{(r-1)} \tag{6.10}$$

这种对 A 的元素进行的消元过程叫 Gauss 消元过程。如果它可以一直进行下去,则最后在第 $n - 1$ 步便有

$$n-1 = a_{11}^{(0)} a_{22}^{(1)} \dots a_{n-1,n-1}^{(n-2)} \tag{6.11}$$

而且由于假定 $a_{n-1,n-1}^{(n-2)} \neq 0$, 从而可得上三角阵

$$A^{(n-1)} = \begin{pmatrix} a_{11}^{(0)} & a_{12}^{(0)} & \dots & a_{1n}^{(0)} \\ & a_{22}^{(1)} & \dots & a_{2n}^{(1)} \\ & & w & \dots \\ & & & a_{nn}^{(n-1)} \end{pmatrix} \tag{6.12}$$

Gauss 消元过程能够进行到底当且仅当 $a_{11}^{(0)}, a_{22}^{(1)}, \dots, a_{n-1,n-1}^{(n-2)}$ 都不为 0, 但根据归纳法和式 (6.10), 这个条件相当于

$$r \neq 0, r = 1, 2, \dots, n - 1 \tag{6.13}$$

由于 Gauss 顺序消元过程的特点是未用列的互换, 由此要求条件(6.13) 是合理的。

6.1.2 矩阵的三角(LU) 分解

当条件(6.13) 满足时, 由式(6.9) 有

$$A = A^{(0)} = L_1 A^{(1)} = L_1 L_2 A^{(2)}$$

= ... = L_1 L_2 ... L_{n-1} A^{(n-1)}

使用矩阵乘法,可以得到

L = L_1 L_2 ... L_{n-1} = \begin{pmatrix} 1 & & & \\ a_{21}^{(0)} & 1 & & \\ a_{31}^{(0)} & & \ddots & \\ \vdots & & & W & W \\ a_{ln}^{(0)} & & & & a_{nn-1}^{(n-2)} \\ a_{11}^{(0)} & \dots & & a_{n-1,n-1}^{(n-2)} & 1 \end{pmatrix}

这是一个对角元素都是 1 的下三角矩阵,称为单位下三角矩阵。

若令

A^{(n-1)} = U

则得

A = LU

于是 A 分解成一个单位下三角矩阵与一个上三角矩阵 U 的乘积。

定义 6.1 如果方阵 A 可分解成一个下三角矩阵 L 和一个上三角矩阵 U 的乘积,则称 A 可做三角分解或 LU 分解。

下面研究方阵的三角分解的存在性和惟一性问题。

首先指出,一个方阵的 LU 分解并不惟一。这是因为如果 A = LU 是 A 的一个三角分解。令 D 是对角元素都不为 0 的对角矩阵,则 A = LU = LDD^{-1}U = \hat{L} \hat{U}。由于上(下)三角矩阵的乘积仍是上(下)三角矩阵,因此 \hat{L} = LD, \hat{U} = D^{-1}U 也分别是下,上三角矩阵。从而 A = \hat{L} \hat{U}也是 A 的一个三角分解。这说明,一般的矩阵的三角分解不是惟一的。有下面定理。

定理 6.1.1 设 A = (a_{ij}) 是 n 阶方阵,则当且仅当 A 的顺序主子式 \Delta_k = \det A \begin{pmatrix} 1 & 2 \dots k \\ 1 & 2 \dots k \end{pmatrix} \neq 0 (k=1,2,...,n-1) 时,A 可以惟一地分解为

A = LDU

其中 L,U 分别是单位下,上三角矩阵,D 是对角矩阵

D = \text{diag}(d_1, d_2, ..., d_n)

其中 d_k = \frac{\Delta_k}{\Delta_{k-1}} \quad k = 1,2,...,n, \quad d_0 = 1

证明略。

推论 n 阶非奇异方阵 A 有三角分解 A = LU 的充分必要条件是 A 的顺序主子式 \Delta_k = \det A \begin{pmatrix} 1 & 2 \dots k \\ 1 & 2 \dots k \end{pmatrix} \neq 0, (k=1,2,...,n-1)。

证明略。

例 1 求矩阵 A = \begin{pmatrix} 2 & -1 & 3 \\ 1 & 2 & 1 \\ 2 & 4 & 2 \end{pmatrix}

的 LDU 分解。

解 因为 $a_{11}=2, a_{22}=5, a_{33}=0$, 所以 A 有惟一的 LDU 分解, 由公式 (6.3) 有

$$L_1^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

再由公式 (6.4) 得

$$L_1^{-1} A^{(0)} = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 & 3 \\ 1 & 2 & 1 \\ 2 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 2 & -1 & 3 \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 5 & -1 \end{pmatrix} = A^{(1)}$$

即 $A^{(0)} = L_1 A^{(1)}$

再由 $A^{(1)}$, 做 L_2^{-1}

$$L_2^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix}$$

使用 (6.8)

$$\begin{aligned} L_2^{-1} A^{(1)} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 & 3 \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 5 & -1 \end{pmatrix} = \begin{pmatrix} 2 & -1 & 3 \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 & 0 \\ 0 & \frac{5}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & -\frac{1}{5} \\ 0 & 0 & 1 \end{pmatrix} = A^{(2)} \end{aligned}$$

于是得 $A^{(0)} = A$ 的 LDU 分解是

$$\begin{aligned} A &= L_1 L_2 A^{(2)} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & \frac{5}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & -\frac{1}{5} \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & \frac{5}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & -\frac{1}{5} \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

矩阵 A 的 LDU 分解与 LU 两种分解都需要假设 A 的前 $n-1$ 阶顺序主子式非零, 如果这个条件不满足, 可以给 A 左(或右)乘以置换矩阵 P , 就可以把 A 的行(或列)的次序重新排列, 使之满足这个条件。从而就有如下的带行交换的矩阵分解定理。

定理 6.1.2 设 A 是 n 阶非奇异矩阵, 则存在置换矩阵 P , 使 PA 的 n 个顺序主子式非零。

该定理的证明可在《计算方法》的书中找到。

推论 设 A 是 n 阶非奇异矩阵, 则存在置换矩阵 P , 使

$$PA = L\tilde{U} = LDU$$

其中 L, U 分别是下、上单位三角阵, \tilde{U} 是上三角阵, D 是对角阵。

如果方程组(6.2)的系数矩阵 A 非奇异, 且 $a_k \neq 0, (k=1, 2, \dots, n-1)$, 则由定理 6.1.1 的推论知, 存在有 A 的三角分解 $A=LU$ 。于是便得与(6.2)同解的具有以下三角矩阵为系数矩阵的两个方程组

$$\begin{cases} Ux = b \\ Lx = Ux \end{cases} \quad (6.14)$$

由方程组(6.14)先解出 Ux , 再代入第一个方程组解出 x 。这就是解线性方程组(6.2)的三角分解法。

如果方程组(6.2)中 A 的顺序主子式不满足全不为零的条件时, 可按定理 6.1.2 的推论考虑与其同解的方程组

$$PA = P \quad (6.15)$$

于是我们仍可用三角分解法(或 Gauss 消去法)来求其解。

6.1.3 分块矩阵的拟 LU 与拟 LDU 分解

将矩阵 A 分解成两个拟三角矩阵和一个拟对角矩阵的乘积, 这种分解式无疑对处理高阶方阵分解问题带来方便, 并能减少计算工作量, 这里我们只限于参加运算的矩阵分成 2×2 块的情况。但反复使用所得结果, 很容易推出 3×3 块或更多的情况下的公式。

设 $A \in R^{n \times n}$, 把 A 分块为

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{matrix} \} n_1 \\ \} n_2 \end{matrix} \quad n_1 + n_2 = n$$

首先, 如果 A_{11} 非奇异, 则用非奇异下三角矩阵

$$\begin{pmatrix} E_{n_1} & 0 \\ -A_{21}A_{11}^{-1} & E_{n_2} \end{pmatrix}$$

左乘 A 得

$$\begin{pmatrix} E_{n_1} & 0 \\ -A_{21}A_{11}^{-1} & E_{n_2} \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{pmatrix} \quad (6.16)$$

这相当于对 A 进行 n_1 个倍加初等行变换, 故得 A 的拟 LU 分解式

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} E_{n_1} & 0 \\ A_{21}A_{11}^{-1} & E_{n_2} \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{pmatrix} \quad (6.17)$$

更进一步, 由 A_{11} 非奇异, 可得

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} E_{n_1} & 0 \\ A_{21}A_{11}^{-1} & E_{n_2} \end{pmatrix} \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{pmatrix} \begin{pmatrix} E_{n_1} & A_{11}^{-1}A_{12} \\ 0 & E_{n_2} \end{pmatrix} \quad (6.18)$$

(6.18) 式就是分块矩阵 A 的拟 LDU 分解。

如果 A_{22} 是非奇异矩阵, 与推导(6.18) 式过程一样可得分块矩阵 A 的另一个拟 LDU 分解

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} E_{n_1} & A_{12} A_{22}^{-1} \\ 0 & E_{n_2} \end{pmatrix} \begin{pmatrix} A_{11} - A_{12} A_{22}^{-1} A_{21} & 0 \\ 0 & A_{22} \end{pmatrix} \begin{pmatrix} E_{n_1} & 0 \\ A_{22}^{-1} A_{21} & E_{n_2} \end{pmatrix} \quad (6.19)$$

从(6.18) 式和(6.19) 式可以看出, 当 A_{11} 可逆(或 A_{22} 可逆时), A 非奇异的充分必要条件是 $A_{22} - A_{21} A_{11}^{-1} A_{12}$ (或 $A_{11} - A_{12} A_{22}^{-1} A_{21}$) 非奇异。于是有以下结论:

$$\begin{aligned} \det A &= \det A_{11} \cdot \det(A_{22} - A_{21} A_{11}^{-1} A_{12}), \det A_{11} \neq 0 \\ \det A &= \det A_{22} \cdot \det(A_{11} - A_{12} A_{22}^{-1} A_{21}), \det A_{22} \neq 0 \end{aligned}$$

例 2 设 $A \in R^{m \times n}, B \in R^{n \times m}$, 则有

$$\det(E_m + AB) = \det(E_n + BA)$$

特别是对于 $R^{1 \times n}, R^{n \times 1}$, 有 $\det(E_n + \quad) = 1 + \quad$

证明 分别用(6.18) 式与(6.19) 式来计算下面的 $m + n$ 阶方阵

$$\begin{pmatrix} E_n & B \\ -A & E_m \end{pmatrix}$$

的行列式, 便有

$$\det \begin{pmatrix} E_n & B \\ -A & E_m \end{pmatrix} = \det(E_n + BA) = \det(E_m + AB)$$

类似地又有

$$\det \begin{pmatrix} E_1 & \\ - & E_m \end{pmatrix} = \det(E_n + \quad) = \det(1 + \quad) = 1 + \quad$$

例 2 的结果在系统理论中经常碰到。

6.2 单纯矩阵的谱分解

定义 6.2 若 n 级方阵 A 相似于对角阵, 称 A 为单纯矩阵。

定理 6.2.1 设 n 阶方阵 A 是单纯矩阵, $\lambda_1, \dots, \lambda_k \in C$ 是 A 的互异特征根, m_1, \dots, m_k 分别是 $\lambda_1, \dots, \lambda_k$ 的重数。则存在 $S_1, \dots, S_k \in C^{n \times n}$, 使得

$$(1) A = \sum_{i=1}^k S_i \quad (6.20)$$

$$(2) S_i S_j = \begin{cases} S_i & i = j \\ 0 & i \neq j \end{cases}, \quad i, j = 1, 2, \dots, k$$

$$(3) \sum_{i=1}^k S_i = E_n$$

$$(4) S_i A = A S_i = \lambda_i S_i \quad i = 1, 2, \dots, k$$

$$(5) \text{rank} S_i = m_i \quad i = 1, 2, \dots, k$$

(6) 满足以上性质的 S_1, \dots, S_k 是惟一的。

证明 (1) 设 $A = T T^{-1}$ 其中 T 是以 A 的特征值为对角元的对角矩阵, T 与 T^{-1} 按相应的互异特征值分块为

令
$$T = (T_1 T_2 \dots T_k)$$

$$T^{-1} = \begin{pmatrix} \tilde{T}_1^T \\ \tilde{T}_2^T \\ \dots \\ \tilde{T}_k^T \end{pmatrix}$$

由 T^{-1} 的结构知

$$\tilde{T}_i^T T_j = \begin{cases} E_{m_i} & i = j \\ O_{m_i \times m_j} & i \neq j \end{cases}$$

于是
$$A = T T^{-1} = (T_1 T_2 \dots T_k) \begin{pmatrix} \tilde{T}_1^T \\ \tilde{T}_2^T \\ \dots \\ \tilde{T}_k^T \end{pmatrix}$$

$$= \sum_{i=1}^k T_i \tilde{T}_i^T = \sum_{i=1}^k S_i$$

式中 $S_i = T_i \tilde{T}_i^T$ ($i = 1, 2, \dots, k$)。

(6.20) 式称为单纯矩阵 A 的谱分解, S_1, \dots, S_k 称为 A 的谱族。

(2) $S_i S_j = T_i \tilde{T}_i^T T_j \tilde{T}_j^T$

$i = j$ 时 上式 $= T_i E_{m_i} \tilde{T}_i^T = S_i$

$i \neq j$ 时 上式 $= T_i 0 \tilde{T}_j^T = 0$

(3)
$$\sum_{i=1}^k S_i = T_1 \tilde{T}_1^T + T_2 \tilde{T}_2^T + \dots + T_k \tilde{T}_k^T = (T_1 T_2 \dots T_k) \begin{pmatrix} \tilde{T}_1^T \\ \dots \\ \tilde{T}_k^T \end{pmatrix} = E_n$$

(4) $S_i A = S_i \sum_{l=1}^k S_l = \sum_{l=1}^k S_i S_l = \sum_{l=1}^k S_l = A S_i$

(5) 由 $S_i = T_i \tilde{T}_i^T$

T_i 是秩为 m_i 的 $n \times m_i$ 的矩阵, \tilde{T}_i^T 是秩为 m_i 的 $m_i \times n$ 的矩阵, 所以存在 n 级可逆阵 P , Q , 使 $PT_i = \begin{pmatrix} E_{m_i} \\ 0 \end{pmatrix}$, $\tilde{T}_i^T Q = (E_{m_i}, 0)$, 即 $PS_i Q = \begin{pmatrix} E_{m_i} & 0 \\ 0 & 0 \end{pmatrix}$

于是 $\text{rank}S_i = m_i \quad i = 1, 2, \dots, k$

(6) 若另有 $\tilde{S}_1, \dots, \tilde{S}_k$ 也满足上述要求, 由

$$\begin{aligned} S_i \tilde{S}_j &= S_i (A \tilde{S}_j) = S_i (S_i A) \tilde{S}_j \\ &= (AS_i) \tilde{S}_j = S_i \tilde{S}_j \end{aligned}$$

故 $(S_i - S_j) \tilde{S}_j = 0$ 当 $i \neq j$ 时, $S_i \tilde{S}_j = 0$
于是

$$\begin{aligned} S_i &= S_i E_n = S_i \sum_{j=1}^k \tilde{S}_j = S_i \tilde{S}_i \\ &= \left(\sum_{j=1}^k S_j \right) \tilde{S}_i = E_n \tilde{S}_i = \tilde{S}_i \end{aligned}$$

即 A 的谱族惟一。

推论 设 $f(\lambda) = a_0 + a_1 \lambda + \dots + a_m \lambda^m$, A 是 n 阶单纯矩阵, 且 A 的谱分解为 $A = \sum_{i=1}^k S_i$, 则

$$f(A) = \sum_{i=1}^k f(S_i) S_i$$

6.3 矩阵的最大秩分解

上面介绍的是 n 阶方阵的分解, 下面介绍长矩阵的分解。
本节将给出矩阵 $A \in C^{m \times n}$ 分解为两个与 A 同秩的因子积的具体方法, 并讨论不同分解之间的关系。它们在广义逆矩阵的讨论中, 将是十分重要的。

定理 6.3.1 设矩阵 $A = (a_{ij})_{m \times n}$, 且 $\text{rank}A = r \leq \min(m, n)$, 则经过有限次初等行变换可把 A 化成

$$\tilde{A}_r = \left[\begin{array}{cccc|cccc} & (k_1) & & (k_2) & & (k_r) & & \\ 0 \dots 0 & 1 & * \dots * & 0 & * \dots * & 0 & * \dots * & \\ 0 \dots 0 & 0 & 0 \dots 0 & 1 & * \dots * & 0 & * \dots * & \\ \hline 0 & 0 & 0 & 0 \dots 0 & 0 & 0 \dots 0 & 1 & * \dots * \\ 0 & 0 & 0 & 0 \dots 0 & 0 & 0 \dots 0 & 0 & 0 \dots 0 \\ \hline 0 & 0 & 0 & 0 \dots 0 & 0 & 0 \dots 0 & 0 & 0 \dots 0 \end{array} \right] \left. \begin{array}{l} \\ \\ \\ \\ \end{array} \right\} \begin{array}{l} r \text{ 行} \\ \\ \\ m - r \text{ 行} \end{array}$$

其中 $1 \leq k_1 < k_2 < \dots < k_r \leq n$, $*$ 号的元素可以不是零, \tilde{A}_r 中第 k_i 个列向量为 \tilde{S}_i

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \text{ 第 } i \text{ 个分量为 } 1, (i = 1, 2, \dots, r)$$

证明 设 $\text{rank}A = r - 1$

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \hline a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = (\alpha_1, \alpha_2, \dots, \alpha_n)$$

如果 A 的第一个非 0 列向量是

$$\alpha_{k_1} = \begin{pmatrix} a_{1k_1} \\ a_{2k_1} \\ \vdots \\ a_{mk_1} \end{pmatrix} \neq 0$$

不妨设 $a_{1k_1} \neq 0$ (通过互换行, 把非零元对调到第一行即成), 把第一行的 $(-\frac{a_{jk_1}}{a_{1k_1}})$ 倍加到第 j 行上, 再把第一行乘以 $1/a_{1k_1}$ 可得

$$\tilde{A}_1 = \begin{pmatrix} & (k_1) & \\ 0 \dots 0 & 1 & * & \dots * \\ 0 \dots 0 & 0 & * & \dots * \\ \hline 0 \dots 0 & 0 & * & \dots * \end{pmatrix} \quad \text{令 } A_1 = \begin{pmatrix} * & \dots * \\ \vdots & \\ * & \dots * \end{pmatrix}_{(m-1) \times (n-k_1)}$$

若 $A_1 = 0$, 则定理得证。若 $A_1 \neq 0$, 可再用上述方法将 \tilde{A}_1 化为

$$\tilde{A}_2 = \begin{pmatrix} & (k_1) & & (k_2) & \\ 0 \dots 0 & 1 & * & \dots * & 0 & * & \dots * \\ 0 \dots 0 & 0 & 0 \dots 0 & 1 & * & \dots * \\ 0 \dots 0 & 0 & 0 \dots 0 & 0 & * & \dots * \\ \hline 0 \dots 0 & 0 & 0 \dots 0 & 0 & * & \dots * \end{pmatrix}$$

如此进行下去, 经过有限次初等行变换后就能把 A 变为 \tilde{A}_r 。

引理 设列分块矩阵 $A = (\alpha_1, \dots, \alpha_n)$ 经过一次行初等变换化为 $A_1 = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_n)$, 那么 $k_1 \alpha_1 + \dots + k_n \alpha_n = 0$ 的充分必要条件是 $k_1 \tilde{\alpha}_1 + \dots + k_n \tilde{\alpha}_n = 0$, 其中 k_1, \dots, k_n 是数。

定理 6.3.2 设 $A = (a_{ij})_{m \times n}$, $\text{rank}A = r = \min(m, n)$ 则可将 A 做满秩分解(或称 A 的最大秩分解)

A = CD

其中 C 是 m× r 阶矩阵, D 是 r× n 阶矩阵, 且 rankC = rankD = r。

证明 由定理 6.3.1, A 经过有限次行初等变换可化为 \tilde{A}_r , 则 \tilde{A}_r 中的第 k_1, k_2, \dots, k_r 列向量 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性无关, 因此对 \tilde{A}_r 中任意一列向量, 有

~_j = \begin{pmatrix} l_{1j} \\ l_{rj} \\ 0 \\ \dots \\ 0 \end{pmatrix} = l_{1j-k_1} + l_{2j-k_2} + \dots + l_{rj-k_r} \quad j = 1, 2, \dots, n

于是由引理, 对矩阵 A 的第 j 个列向量有 $\alpha_j = l_{1j-k_1} + l_{2j-k_2} + \dots + l_{rj-k_r}$ 取 $C = (\alpha_{k_1}, \alpha_{k_2}, \dots, \alpha_{k_r}) \quad rankC = r$

取 \tilde{A}_r 中的前 r 行作为 D, 即

D = \begin{pmatrix} 0\dots0 & 1 & * & \dots * & 0 & * & \dots * & 0 & * & \dots * \\ 0\dots0 & 0 & 0\dots0 & 1 & * & \dots * & 0 & * & \dots * \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0\dots0 & 0 & 0\dots0 & 0 & 0\dots0 & 1 & * & \dots * \end{pmatrix}_{r \times n}

当然 rankD = r, 且 D 中第 j 个列向量

j = \begin{pmatrix} l{1j} \\ \dots \\ l_{rj} \end{pmatrix} \quad j = 1, 2, \dots, n

于是

C_j = (\alpha_{k_1}, \alpha_{k_2}, \dots, \alpha_{k_r}) \begin{pmatrix} l_{1j} \\ \dots \\ l_{rj} \end{pmatrix} = l_{1j-k_1} + l_{2j-k_2} + \dots + l_{rj-k_r} = \alpha_j

所以

CD = C(\alpha_1, \alpha_2, \dots, \alpha_n) = (C\alpha_1, C\alpha_2, \dots, C\alpha_n) = (\alpha_1, \alpha_2, \dots, \alpha_n) = A

当然, 如果 rankA = k = min(m, n), 则 A 已经是最大秩分解了。

例 1 求矩阵

A = \begin{pmatrix} 2 & 0 & 1 & 4 \\ 0 & 1 & 0 & 2 \\ 2 & -1 & 1 & 2 \end{pmatrix}

的最大秩分解。

解 对 A 进行初等行变换

A = \begin{pmatrix} 2 & 0 & 1 & 4 \\ 0 & 1 & 0 & 2 \\ 2 & -1 & 1 & 2 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 0 & 1 & 4 \\ 0 & 1 & 0 & 2 \\ 0 & -1 & 0 & -2 \end{pmatrix}

$$\begin{pmatrix} 2 & 0 & 1 & 4 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & \frac{1}{2} & 2 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \tilde{A}_2$$

所以取 A 的前 2 列做 C

$$C = \begin{pmatrix} 2 & 0 \\ 0 & 1 \\ 2 & -1 \end{pmatrix} \quad \text{rank} C = 2$$

取 \tilde{A}_2 的前 2 行做 D ,

$$D = \begin{pmatrix} 1 & 0 & \frac{1}{2} & 2 \\ 0 & 1 & 0 & 2 \end{pmatrix} \quad \text{rank} D = 2$$

而且 $A = CD$

上面两个定理对初等列变换也成立, 因此也可用初等列变换来求 A 的满秩分解。

例 2 将例 1 中的 A 用初等列变换求 A 的最大秩分解。

$$A = \begin{pmatrix} 2 & 0 & 1 & 4 \\ 0 & 1 & 0 & 2 \\ 2 & -1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & 4 \\ 0 & 1 & 0 & 2 \\ 1 & -1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 \\ 1 & -1 & 0 & -2 \end{pmatrix} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{pmatrix} = A_r$$

所以 A 的前两行是线性无关的, 取 A 的前 2 行做

$$\tilde{D} = \begin{pmatrix} 2 & 0 & 1 & 4 \\ 0 & 1 & 0 & 2 \end{pmatrix} \quad \text{rank} D = 2$$

取 A_r 的前 2 列做 \tilde{C}

$$\tilde{C} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{pmatrix} \quad \text{rank} C = 2$$

且 $A = \tilde{C}\tilde{D}$

由此可见矩阵 A 的最大秩分解不是惟一的, 但最大秩分解之间, 有如下的关系。

定理 6.3.3 设 $A = (a_{ij})_{m \times n}$, 且 $\text{rank} A = r = \min(m, n)$ 。若 $A = CD = \tilde{C}\tilde{D}$ 均为 A 的最大秩分解, 则

存在 r 阶可逆阵 Q , 使得

$$C = \tilde{C}Q, \quad D = Q^{-1}\tilde{D} \quad (6.21)$$

$$D^H(DD^H)^{-1}(C^HC)^{-1}C^H = \tilde{D}^H(\tilde{D}\tilde{D}^H)^{-1}(\tilde{C}^H\tilde{C})^{-1}\tilde{C}^H \quad (6.22)$$

证明 注意到 $\text{rank}(AA^H) = \text{rank}(A^HA) = \text{rank}(A)$ 所以 DD^H 及 C^HC 都是 r 阶可逆方阵

由于 $CD = \tilde{C}\tilde{D}$

就有 $CDD^H = \tilde{C}\tilde{D}\tilde{D}^H$

$$C = \tilde{C}\tilde{D}\tilde{D}^H(DD^H)^{-1} \tag{6.23}$$

令 $\tilde{D}\tilde{D}^H(DD^H)^{-1} = Q_1$, 即得

$$C = \tilde{C}Q_1 \tag{6.24}$$

同理可得

$$D = Q_2\tilde{D} \tag{6.25}$$

其中 $Q_2 = (C^HC)^{-1}C^H\tilde{C}$

将(6.24)式、(6.25)式代入 $CD = \tilde{C}\tilde{D}$, 得

$$\tilde{C}Q_1Q_2\tilde{D} = \tilde{C}\tilde{D}$$
$$\tilde{C}^H\tilde{C}Q_1Q_2\tilde{D}\tilde{D}^H = \tilde{C}^H\tilde{C}\tilde{D}\tilde{D}^H$$

由于 $\tilde{C}^H\tilde{C}$ 与 $\tilde{D}\tilde{D}^H$ 皆可逆, 故

$$Q_1Q_2 = E$$

故令 $Q_1 = Q \quad Q_2 = Q^{-1}$, 即

$$C = \tilde{C}Q \quad D = Q^{-1}\tilde{D}$$

由(6.21)式 $C = \tilde{C}Q \quad D = Q^{-1}\tilde{D}$

那么 $D^H(DD^H)^{-1}(C^HC)^{-1}C^H$

$$\begin{aligned} &= (Q^{-1}\tilde{D})^H\{Q^{-1}\tilde{D}(Q^{-1}\tilde{D})^H\}^{-1}\{(\tilde{C}Q)^H\tilde{C}Q\}^{-1}(\tilde{C}Q)^H \\ &= \tilde{D}^H(Q^H)^{-1}Q^H(\tilde{D}\tilde{D}^H)^{-1}QQ^{-1}(\tilde{C}^H\tilde{C})^{-1}(Q^H)^{-1}Q^H\tilde{C}^H \\ &= \tilde{D}^H(\tilde{D}\tilde{D}^H)^{-1}(\tilde{C}^H\tilde{C})^{-1}\tilde{C}^H \end{aligned}$$

(6.22) 式得证。

(6.22) 式表明, 矩阵 A 的最大秩分解虽不惟一, 但由最大秩分解所做出的这种形式的乘积

$D^H(DD^H)^{-1}(C^HC)^{-1}C^H$

是相同的。这个乘积表达式正是后面将讨论的矩阵 A 的广义逆矩阵中的 Moore-Penrose 广义逆。

6.4 矩阵的 QR 分解

矩阵的 QR 分解, 在解决最小二乘问题, 特征值计算、广义逆矩阵的计算方面, 都是十分重要

142

要的,从而成为有名的 QR 算术。这里只给出分解定理。

定理 6.4.1 设 $A \in C^{m \times n}$, $\text{rank} A = r$, 则 A 可分解为

$$A = QR$$

其中 $Q \in C^{m \times r}$, 且 $Q^H Q = E_r$, $R \in C^{r \times n}$, $\text{rank} R = r$

矩阵 A 的这种分解称为 QR 分解。

证明 设 $A = CD$ 是 A 的最大秩分解, 对于 C 的 r 个线性无关的列向量用 Gram-Schmidt 标准正交化方法

$$C = (c_1 \ c_2 \ \dots \ c_r)$$

注意到第 2 章的(2.4) 式子

$$(c_1 \ c_2 \ \dots \ c_r) = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 1 & 2 & \dots & r \end{pmatrix} \begin{pmatrix} k_{11} & k_{12} & \dots & k_{1r} \\ 0 & k_{22} & \dots & k_{2r} \\ \dots & \dots & & \dots \\ 0 & 0 & \dots & k_{rr} \end{pmatrix}$$

其中 e_1, \dots, e_r 是两两正交的单位向量, k_{11}, \dots, k_{rr} 皆大于 0。

$$C = QK \quad K = \begin{pmatrix} k_{11} & k_{12} & \dots & k_{1r} \\ & k_{22} & \dots & k_{2r} \\ & & \ddots & \\ & & & k_{rr} \end{pmatrix}$$

$$Q = (e_1, \dots, e_r), Q^H Q = E_r$$

令 $KD = R$, 有

$$A = QR$$

由上述定理 6.4.1 知, A 的 QR 分解是一种特殊的最大秩分解。

推论 1 若 $A \in C^{m \times n}$, $\text{rank} A = r$, 则 A 可惟一分解为

$$A = QR$$

其中 $Q \in C^{m \times r}$, 且 $Q^H Q = E_r$, $R \in C^{r \times n}$ 为具正对角元素的上三角阵。

证明 将 A 的列向量正交化、标准化, 可得

$$A = QR$$

其中 Q, R 皆满足命题条件, 下证惟一性, 设 A 有两种上述分解

$$A = QR = Q_1 R_1$$

则

$$Q = Q_1 R_1 R^{-1} = Q_1 K$$

其中 $K = R_1 R^{-1}$ 是具正对角元素的 r 级上三角阵

由于

$$Q^H Q = K^H Q_1^H Q_1 K = K^H K = E_r$$

这就说明 K 是酉阵, 则 $K = E$

所以

$$R_1 = R \quad Q_1 = Q$$

推论 2 若 $A \in C^{r \times n}$, $\text{rank} A = r$, 则 A 可惟一地分解为

$$A = LQ$$

其中 $Q \in C^{r \times n}$, $QQ^H = E_r$, $L \in C^{r \times r}$ 为具有正对角元素的下三角阵。

证明 $A^H \in C^{n \times r}$, 由推论 1, A^H 有分解

$$A^H = Q_1 R_1$$

其中 $Q_1 \in C^{n \times r}$, 且 $Q_1^H Q_1 = E_r$, $R_1 \in C^{r \times r}$ 为具正对角元的上三角阵。那么

$$A = R_1^H Q_1^H$$

令 $R_1^H = L, Q_1^H = Q$ 正为所求。

由以上结果知, A 的 QR 分解中的矩阵 Q , 显然是次酉阵。

例 1 用 QR 方法解线性方程组

$A =$

其中

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} A = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \\ 2 & 3 & 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 1 \end{pmatrix}$$

解 设 $A = (\alpha_1 \alpha_2 \alpha_3)$

$$\alpha_1 = (1 \ 1 \ 1 \ 2)^T \quad \alpha_2 = (1 \ 2 \ 1 \ 3)^T \quad \alpha_3 = (2 \ 1 \ 3 \ 3)^T$$

将 $\alpha_1, \alpha_2, \alpha_3$ 标准正交化为

$$\begin{aligned} \beta_1 &= \left[\frac{1}{\sqrt{7}} \quad \frac{1}{\sqrt{7}} \quad \frac{1}{\sqrt{7}} \quad \frac{2}{\sqrt{7}} \right]^T \\ \beta_2 &= \left[-\frac{3}{\sqrt{35}} \quad \frac{4}{\sqrt{35}} \quad -\frac{3}{\sqrt{35}} \quad \frac{1}{\sqrt{35}} \right]^T \\ \beta_3 &= \left[\frac{-2}{\sqrt{15}} \quad \frac{1}{\sqrt{15}} \quad \frac{3}{\sqrt{15}} \quad \frac{-1}{\sqrt{15}} \right]^T \end{aligned}$$

于是令

$$Q = \begin{pmatrix} \frac{1}{\sqrt{7}} & -\frac{3}{\sqrt{35}} & \frac{-2}{\sqrt{15}} \\ \frac{1}{\sqrt{7}} & \frac{4}{\sqrt{35}} & \frac{1}{\sqrt{15}} \\ \frac{1}{\sqrt{7}} & -\frac{3}{\sqrt{35}} & \frac{3}{\sqrt{15}} \\ \frac{2}{\sqrt{7}} & \frac{1}{\sqrt{35}} & \frac{-1}{\sqrt{15}} \end{pmatrix}$$

得

$$R = Q^H A = \begin{pmatrix} \sqrt{7} & \frac{10}{\sqrt{7}} & \frac{12}{\sqrt{7}} \\ 0 & \frac{5}{\sqrt{35}} & -\frac{8}{\sqrt{35}} \\ 0 & 0 & \frac{3}{\sqrt{15}} \end{pmatrix}$$

$$R^{-1} = \begin{pmatrix} \frac{1}{\sqrt{7}} & \frac{-10}{\sqrt{35}} & -\frac{140}{7\sqrt{15}} \\ 0 & \frac{7}{\sqrt{35}} & \frac{8}{\sqrt{15}} \\ 0 & 0 & \frac{5}{\sqrt{15}} \end{pmatrix}$$

所以 $A =$ 即 $QR =$,那么
 $= R^{-1} Q^H = (-1, 4, 1)^T$

6.5* 矩阵的奇异值分解

矩阵的约当标准形的重要性已为人们所公认,但它有两点局限,一是它只是方阵的一种分解;二是虽然约当标准形是一种特殊的下三角阵,但它仍不能像对角阵那样方便。人们的研究突破了这两点,获得了新的矩阵的奇异值分解,它在现代矩阵理论中的重要性是不言而喻的。近年来,古典控制中的频率法,正是由于有了矩阵奇异值分解的帮助而得到了新的发展,这里只给出分解定理。

定理 6.5.1 设 $A \in C^{m \times n}$, 则
a) $A^H A, AA^H$ 的特征值均为非负实数。
b) $A^H A$ 与 AA^H 的非零特征值相同。
证明 a) $A^H A, AA^H$ 均为埃尔米特阵,故它们特征值皆为实数,设 λ 是 $A^H A$ 的特征值,是 $A^H A$ 对应于 λ 的特征向量,故

$$A^H A \lambda = \lambda A^H A$$

又 $0 \leq (A^H A) = (\lambda, A^H A) = (\lambda) = (\lambda)$
而 $(\lambda) > 0$, 故 $\lambda \geq 0$

同理可证明 AA^H 的特征值为非负实数。
b) $A \in C^{m \times n}$, 不妨设 $m \geq n$, $\text{rank} A = r$, 那么存在可逆阵 P, Q , 使

$$PAQ = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$$

令 $Q^{-1} A^H P^{-1} = \begin{pmatrix} B_1 & B_2 \\ B_3 & B_4 \end{pmatrix}_{n \times m}$ B_1 是 r 级方阵

那么 $PAQQ^{-1} A^H P^{-1} = PAA^H P^{-1} = \begin{pmatrix} B_1 & B_2 \\ 0 & 0 \end{pmatrix}_{m \times m}$
 $Q^{-1} A^H P^{-1} PAQ = Q^{-1} A^H A Q = \begin{pmatrix} B_1 & 0 \\ B_3 & 0 \end{pmatrix}_{n \times n}$

所以

$$|E_m - AA^H| = |E_m - \begin{pmatrix} B_1 & B_2 \\ 0 & 0 \end{pmatrix}| = {}^{m-r} |E_r - B_1|$$

$$|E_n - A^H A| = |E_n - \begin{pmatrix} B_1 & 0 \\ B_3 & 0 \end{pmatrix}| = \begin{vmatrix} \lambda_1 - B_1 & \\ & \lambda_2 - B_1 \end{vmatrix}$$

$$|E_m - AA^H| = \begin{vmatrix} \lambda_1 - B_1 & \\ & \lambda_2 - B_1 \end{vmatrix} = |E_n - A^H A|$$

这个等式说明 AA^H 与 $A^H A$ 有完全相同的非零特征值。

定义 6.2 设 $A \in C^{m \times n}$, $\text{rank} A = r$, $A^H A$ 的特征值为

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > \lambda_{r+1} = \dots = \lambda_n = 0$$

则称 $\sigma_i = \sqrt{\lambda_i}$, ($i = 1, 2, \dots, r$) 为矩阵 A 的正奇异值($\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}$ 称为奇异值)。

由此定义可知, A 与 A^H 有相同的正奇值。

例 1 设

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

求 A 的正奇异值。

解 由于

$$AA^H = \begin{pmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

显然, AA^H 的正特征值为 5, 所以 A 的正奇值为 $\sqrt{5}$

定义 6.3 设 $A, B \in C^{m \times n}$, 若存在 m 阶酉阵 S 及 n 阶酉矩阵 T , 使得

$$A = SBT$$

则称 A 与 B 酉等价。

定理 6.5.2 若 $A, B \in C^{m \times n}$ 是酉等价, 则 A 与 B 有相同的正奇值。

证明 因为 A, B 酉等价, 所以存在 m 阶及 n 阶酉阵 S, T 使

$$A = SBT$$

那么

$$AA^H = SBT T^H B^H S^H = SBB^H S^H$$

这说明 AA^H 与 BB^H 是酉相似的, 它们有相同的特征值, 故 A 与 B 有相同的正奇异值。

定理 6.5.3 设 $A \in C^{m \times n}$, $\text{rank} A = r$, 则存在 m 阶酉矩阵 V 及 n 阶酉矩阵 U , 使得

$$A = V \begin{pmatrix} \sigma_1 & & 0 \\ & \sigma_r & \\ 0 & & 0 \end{pmatrix} U^H \tag{6.26}$$

其中 $\sigma_r = \text{diag}(\sigma_1 \sigma_2 \dots \sigma_r)$, $\sigma_1 \sigma_2 \dots \sigma_r > 0$, 是 A 的正奇异值。

证明 由于 $A^H A$ 是正定或半正定的 Hermite 阵, 且

$$\text{rank} A^H A = \text{rank} A = r$$

所以存在 n 阶酉阵 U , 使

$$U^H A^H A U = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_r & & \\ & & & \\ & & & \sigma_r \\ & & & & 0 \end{pmatrix}$$

其中 $\lambda_1, \lambda_2, \dots, \lambda_r > 0$ 是 $A^H A$ 的 r 个正特征值。

令 $U = (\alpha_1 \dots \alpha_r \alpha_{r+1} \dots \alpha_n) = (U_1, U_2)$

由 $A^H A$ 的 n 个单位正交特征向量构成

$$\begin{aligned} A^H A \alpha_i &= \lambda_i \alpha_i \quad i = 1, \dots, r \\ A^H A \alpha_j &= 0 \quad j = r + 1, \dots, n \end{aligned}$$

其中 $U_1 = (\alpha_1, \dots, \alpha_r), U_2 = (\alpha_{r+1}, \dots, \alpha_n)$

$$\alpha_i = \frac{1}{\sqrt{\lambda_i}} A \beta_i \quad i = 1, 2, \dots, r$$
$$\Lambda_r = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r)$$

则有

$$\begin{aligned} U_1^H A^H A U_1 &= \begin{pmatrix} \alpha_1^H \\ \vdots \\ \alpha_r^H \end{pmatrix} (A^H A \alpha_1, \dots, A^H A \alpha_r) \\ &= \begin{pmatrix} \alpha_1^H \\ \vdots \\ \alpha_r^H \end{pmatrix} (\lambda_1 \alpha_1, \dots, \lambda_r \alpha_r) \\ &= \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{pmatrix} = \Lambda_r \end{aligned}$$

因为 Λ_r 是 r 阶实可逆阵, 故有

$$\Lambda_r^{-1} U_1^H A^H A U_1 \Lambda_r^{-1} = (A U_1 \Lambda_r^{-1})^H (A U_1 \Lambda_r^{-1}) = E_r$$

令 $V_1 = A U_1 \Lambda_r^{-1}$

则 $V_1^H V_1 = E_r$

把 V_1 的 r 个列向 $\alpha_1, \dots, \alpha_r$ 扩充为 C^m 的标准正交基

$$\alpha_1, \dots, \alpha_r, \alpha_{r+1}, \dots, \alpha_m$$

并令 $V_1 = (\alpha_1, \dots, \alpha_r) \quad V_2 = (\alpha_{r+1}, \dots, \alpha_m)$

$$V = (V_1, V_2)$$

则 V 是 m 阶酉阵, 并注意到

$$U_2^H A^H A U_2 = U_2^H (A^H A \alpha_{r+1}, \dots, A^H A \alpha_n) = 0$$

所以 $A U_2 = 0$

于是

$$\begin{aligned} V \begin{pmatrix} \Lambda_r & 0 \\ 0 & 0 \end{pmatrix} U^H &= (V_1, V_2) \begin{pmatrix} \Lambda_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} U_1^H \\ U_2^H \end{pmatrix} \\ &= V_1 \Lambda_r U_1^H = A U_1 \Lambda_r^{-1} \Lambda_r U_1^H \\ &= A U_1 U_1^H = A U_1 U_1^H + A U_2 U_2^H \\ &= A (U_1 U_1^H + U_2 U_2^H) = A U U^H = A \end{aligned}$$

这就证明了

$$A = V \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & 0 \end{pmatrix} U^H$$

上式称为 A 的奇异值分解, 它表明 A 与一个长方形对角阵酉等价。
在定理的证明中还得到一个与 A 的奇异值有关的最大秩分解

其中

$$\begin{aligned} A &= CD & C &= V_1 \\ C &= AU_1 & D &= \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & 0 \end{pmatrix} U_1^H \end{aligned}$$

例 2 求例 1 中矩阵 A 的奇值分解

解 由例 1 已知 A 的正奇值 $\sigma_1 = \sqrt{5}$, 所以 $\sigma_2 = 0$

因为 $A = \begin{pmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$ 所以 $A^H A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$

则 $A^H A$ 的特征值 $\lambda_1 = 5, \lambda_2 = 0$
对应的单位特征向量分别为

$$\begin{aligned} u_1 &= \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad u_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix} \\ \text{则 } U &= (u_1, u_2) = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} \\ V_1 &= AU, \quad V_1^{-1} = \begin{pmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \frac{1}{\sqrt{5}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \\ \text{故可取 } V_2 &= \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

于是

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{5} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}$$

习 题 6

1. 用三角分解法求解线性方程组
- $$\begin{cases} x_1 + x_2 + 2x_3 + 3x_4 = 1 \\ 3x_1 - x_2 - x_3 - 2x_4 = -4 \\ 2x_1 + 3x_2 - x_3 - x_4 = -6 \\ x_1 + 2x_2 + 3x_3 - x_4 = -4 \end{cases}$$
2. 求单纯矩阵 $A = \begin{pmatrix} -29 & 6 & 18 \\ -20 & 5 & 12 \\ -40 & 8 & 25 \end{pmatrix}$

的谱分解。

3. 求下列矩阵的最大秩分解

$$(1) \quad A = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 1 & 1 \\ -2 & 3 & 2 & 5 \end{pmatrix} \quad (2) \quad B = \begin{pmatrix} 1 & -1 & 1 & 1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix}$$

4. 求 3 题所给矩阵 B 的 QR 分解。

5. 求 3 题所给矩阵 A, B 的奇异值分解。

第7章

广义逆矩阵及其应用

当 n 阶方阵 A 可逆时, 线性方程组

$$Ax = b \tag{1}$$

的解存在, 且惟一, 即 $x = A^{-1}b$ 。这一事实可否推广到一般的线性方程组

$$A_{m \times n} x = b \tag{2}$$

上去, 使得其解 $x = G_{n \times m} b$?

1920 年摩勒(E. N. Moore) 首先引进了广义逆矩阵这一概念, 其后 30 年未能引起人们的重视。到了 1950 年后, 电子计算机的出现, 推动了计算科学的发展, 1953 年彭罗司(R. Penrose) 以更明确的形式给出 Moore 广义逆的定义之后, 广义逆阵的研究才进入了一个新的时期。由于广义逆矩阵在数理统计、系统理论、优化计算和控制论等许多领域中的重要应用逐渐为人们所认识, 因而大大推动了对广义逆矩阵的理论与应用的研究, 使得这一学科得到迅速发展。本章着重介绍应用最多的两类广义逆矩阵 A^- 和 A^+ 。

7.1 广义逆矩阵及其分类

定义 7.1 设 A 是任意的 $m \times n$ 阶矩阵, 如果存在 $n \times m$ 阶矩阵 G 满足 Penrose-Moore 方程

(1)

$AGA = A;$

(2)

$GAG = G;$

(3)

$(GA)^H = GA;$

(4)

$(AG)^H = AG。$

的全部或一部分, 则称 G 为 A 的广义逆矩阵。

按照这一定义, 广义逆矩阵可以分为满足一个方程的广义逆矩阵, 满足其中两个或三个、四个方程的广义逆矩阵, 一共有

$$C_4^1 + C_4^2 + C_4^3 + C_4^4 = 15$$

即 15 类广义逆矩阵。

如果 G 是满足第 i 个方程的广义逆矩阵, 就记为

$$G = A^{(i)} \quad (i = 1, 2, 3, 4)$$

如果 G 是满足第 i, j 两个方程, 或满足第 i, j, k 三个方程的, 记

$$G = A^{(i, j)} \text{ 或 } G = A^{(i, j, k)} \quad (i, j, k = 1, 2, 3, 4)$$

满足四个方程的广义逆矩阵记为

$$G = A^{(1, 2, 3, 4)}$$

下面我们将会看到, 除了 $A^{(1, 2, 3, 4)}$ 是惟一确定之外(并记 $A^+ = A^{(1, 2, 3, 4)}$), 其余各类广义逆矩阵都不是惟一确定的, 每一种广义逆矩阵都包含着一类矩阵, 为了表示这种情况, 把满足前面所述相应条件的一切广义逆矩阵分别记为

$$A\{i\}, A\{i, j\}, A\{i, j, k\}$$

从而相应的有 $A^{(i)} A\{i\}, A^{(i, j)} A\{i, j\}, A^{(i, j, k)} A\{i, j, k\}$

在上述 15 种广义逆矩阵中, 应用较多的是以下 5 种。

- (1) $A\{1\}$, 其中任意一个固定的广义逆矩阵记为 A^- ;
- (2) $A\{1, 2\}$, 其中任意一个固定的广义逆矩阵记为 A_r^- ;
- (3) $A\{1, 3\}$, 其中任意一个固定的广义逆矩阵记为 A_m^- ;
- (4) $A\{1, 4\}$, 其中任意一个固定的广义逆矩阵记为 A_l^- ;
- (5) A^+

因为 A^+ 满足全部四个方程, 显然有

$$A^+ A\{1\}, A^+ A\{1, 2\}, A^+ A\{1, 3\}, A^+ A\{1, 4\}$$

因此 A^+ 在广义逆阵中占有十分重要的位置。

7.2 广义逆矩阵 A^-

广义逆矩阵 A^- 起源于线性方程组

$$A_{m \times n} x = b \tag{2}$$

的求解问题。我们知道方程组(2)有解的充分必要条件是

$$\text{rank}(A) = \text{rank}(A, b)$$

自然会想到它的解是否也能用某个 $n \times m$ 矩阵 G 表示为 $x = Gb$ 呢? 如果它的解能表示为 $x = Gb$, 那么矩阵 G 如何求, 以及 G 有什么性质?

首先由方程组的理论可以知道, 若方程组(2)有解的话, 解一般不是惟一的, 因此表示解 $x = Gb$ 的矩阵 G 如果存在的话, 显然在一般情况下也不是惟一的。

我们把使方程组(2)有解的一切 $C^{m \times 1}$ 所成的集合记为 $R(A)$, 即

$$R(A) = \{ b \mid \text{rank} A = \text{rank}(A, b), \quad b \in C^{m \times 1} \}$$

则对任何的 $b \in R(A)$, 必存在 C^n 使

$$Ax = b$$

反过来, 对任意 $b \in C^n, A \in R(A)$ 。

定理 7.2.1 对任意的 $b \in R(A)$, 存在 $n \times m$ 阶矩阵 G 使 $x = Gb$ 都是方程组 $Ax = b$ 的解的充分必要条件是 G 满足

$$AGA = A$$

证明 先证必要性。若对一切 $R(A), G$ 都是 $A =$ 的解, 由于对任意的 C^n, A
 $R(A)$, 那么 GA 是 $A = A$ 的一个解, 则

$$AGA = A$$

注意到 的任意性, 故有

$$AGA = A$$

再证充分性。对任意的 $R(A)$, 设 C^n 是方程组 $A =$ 的解, 那么

$$A = \tag{*}$$

又因为 $AGA = A$ 所以(*)式两端左乘以 AG

$$AGA = AG = A =$$

说明 G 是 $A =$ 的解。

下面给出求一般 A^{-} 的方法

定理 7.2.2 设 $A \in C^{m \times n}$, 且 $\text{rank}(A) = r$, 存在 m 阶可逆阵 P 及 n 阶可逆阵 Q , 使

$$PAQ = \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix}$$

则 $n \times m$ 阶矩阵 G 使得 $AGA = A$ 的充分必要条件是

$$G = Q \begin{bmatrix} E_r & G_{12} \\ G_{21} & G_{22} \end{bmatrix} P$$

其中 G_{12}, G_{21}, G_{22} 分别是 $r \times (m - r), (n - r) \times r, (n - r) \times (m - r)$ 阶任意矩阵。

证明 先证必要性, 由条件有 m 阶及 n 阶可逆阵 P, Q , 使

$$PAQ = \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix}$$

那么

$$A = P^{-1} \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix} Q^{-1}$$

根据 G 应满足的 $AGA = A$, 有

$$P^{-1} \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix} Q^{-1} GP^{-1} \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix} Q^{-1} = P^{-1} \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix} Q^{-1}$$

$$\begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix} Q^{-1} GP^{-1} \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix}$$

再令

$$Q^{-1} GP^{-1} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}$$

分块如题设要求, 代入上式

$$\begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} G_{11} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix}$$

所以 $G_{11} = E_r$, 于是有

$$Q^{-1}GP^{-1} = \begin{bmatrix} E_r & G_{12} \\ G_{21} & G_{22} \end{bmatrix}$$

得到

$$G = Q \begin{bmatrix} E_r & G_{12} \\ G_{21} & G_{22} \end{bmatrix} P$$

再证充分性, 由于

$$G = Q \begin{bmatrix} E_r & G_{12} \\ G_{21} & G_{22} \end{bmatrix} P, \quad A = P^{-1} \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix} Q^{-1}$$

则

$$\begin{aligned} AGA &= P^{-1} \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix} Q^{-1} Q \begin{bmatrix} E_r & G_{12} \\ G_{21} & G_{22} \end{bmatrix} P P^{-1} \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix} Q^{-1} \\ &= P^{-1} \begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix} Q^{-1} = A \end{aligned}$$

由定理 7.2.2 知, 对任意的 $m \times n$ 阶矩阵 A , 它的 A^- 总是存在的, 并可表示为

$$A^- = Q \begin{bmatrix} E_r & G_{12} \\ G_{21} & G_{22} \end{bmatrix} P \tag{7.1}$$

由 G_{12}, G_{21}, G_{22} 的任意性知 A^- 不惟一, 因此 A^- 的全体所成的集合 $A\{1\}$ 是非空集合, 当然若 A 是可逆方阵, 则 $A^- = A^{-1}$, 此时由公式(7.1)得到的 A^{-1} 当然也惟一。

由 A^- 的定义可以导出以下性质:

- (1) 对任意的 $m \times n$ 阶矩阵 A , $\text{rank}(A^-) = \text{rank}(A)$;
- (2) $(A^-)^H = (A^H)^-, (A^-)^T = (A^T)^-$;
- (3) 若 $m = n = r$, 则 $A\{1\}$ 只有惟一元素 A^{-1} ;
- (4) AA^-, A^-A 均为幂等矩阵, 并且

$$\text{rank}(A) = \text{rank}(AA^-) = \text{rank}(A^-A)$$

- (5) $\text{rank}(A) = n$ (此时称为列满秩阵或高矩阵)

的充分必要条件是 $A^-A = E_n$, 此时 $A^- = (A^H A)^{-1} A^H$ 称 A 的一个左逆, 记为 A_L^{-1} ;

- (6) $\text{rank}(A) = m$ (此时称为行满秩阵)。

的充分条件是 $AA^- = E_m$, 此时 $A^- = A^H (AA^H)^{-1}$ 称 A 的一个右逆, 记为 A_R^{-1} ;

- (7) 对任意非 0 复数 λ , $B = \lambda A$, 则 $B^- = \frac{1}{\lambda} A^-$ 。

证明 性质(1) ~ (4) 留作习题。

证明性质(5)充分性, 若 $A^-A = E_n$ 则

$$n = \text{rank}(E_n) = \text{rank}(A) = \text{rank}(A^-A) = n$$

所以 $\text{rank} A = n$

必要性, 若 $\text{rank} A = n$, 则存在 m 阶及 n 阶可逆阵 P, Q , 使

$$PAQ = \begin{bmatrix} E_n \\ 0 \end{bmatrix} \quad \text{或} \quad A = P^{-1} \begin{bmatrix} E_n \\ 0 \end{bmatrix} Q^{-1}$$

由定理 7.2.2 的(7.1) 式知

$$G = Q(E_n \quad G_{12}) P$$

G 即 A^- , 于是有 $A^- A = E_n$

由于

$$\text{rank}(A^H A) = \text{rank}(A) = n$$

所以 $A^H A$ 是可逆阵, 那么

$$(A^H A)^{-1} A^H A = E_n$$

所以, 可取

$$A^- = A_L^{-1} = (A^H A)^{-1} A^H \tag{7.2}$$

证明性质(6) 仿照上面可证, AA^H 可逆, 有

$$AA^H (AA^H)^{-1} = E_m$$

所以, 可取

$$A^- = A_R^{-1} = A^H (AA^H)^{-1} \tag{7.3}$$

证明性质(7) 可以直接验证其成立。

下面讨论 A^- 的计算方法。

1 ° 满秩分解法

对任意 $m \times n$ 矩阵 A , 第 6 章定理 6.3.2 知 $A = CD$, 其中 C 是 $m \times r$ 矩阵, D 是 $r \times n$ 阶矩阵, 且 $\text{rank} C = \text{rank} D = \text{rank} A = r$, 再由性质(5), (6) 可得

$$A^- = D^- C^- \tag{7.4}$$

如果 A 是实矩阵, 由(7.2), (7.3) 及(7.4) 式可得

$$A^- = D_R^{-1} C_L^{-1} = D^T (DD^T)^{-1} (C^T C)^{-1} C^T \tag{7.5}$$

例 1 设

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 2 \\ 2 & 4 & 0 \end{bmatrix}$$

试用满秩分解求广义逆矩阵 A^- 。

解 容易得到 A 的满秩分解

$$A = CD$$

其中 $C = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 2 & 0 \end{bmatrix}, D = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

$$D_R^{-1} = D^T (DD^T)^{-1} = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}^{-1} = \frac{1}{5} \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 5 \end{bmatrix}$$

$$C_L^{-1} = (C^T C)^{-1} C^T = \begin{bmatrix} 5 & 0 \\ 0 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 2 & 0 \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 4 & 0 & 8 \\ 0 & 10 & 0 \end{bmatrix}$$

所以

$$A^- = D_R^{-1} C_L^{-1} = \frac{1}{100} \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 4 & 0 & 8 \\ 0 & 10 & 0 \end{bmatrix} = \frac{1}{100} \begin{bmatrix} 4 & 0 & 8 \\ 8 & 0 & 16 \\ 0 & 50 & 0 \end{bmatrix}$$

应当指出: 用满秩分解求广义逆 A^- 比较麻烦, 但用此法所求的广义逆正是重要的广义逆-Moore-Penrose 广义逆 A^+ 。

2 ° 初等行、列变换法

设 $m \times n$ 阶矩阵 A 的秩 $=r$, 对 A 做初等行、列变换后, 总可以把 A 写成如下分块矩阵形式, 即存在 m 阶及 n 阶可逆阵 P, Q , 使

$$PAQ = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}$$

其中 A_1 是 $r \times r$ 阶的可逆矩阵, 且 A_2, A_3, A_4 是满足 $A_4 = A_3 A_1^{-1} A_2$ 适当阶数的矩阵, 不难验证

$$A^- = Q \begin{bmatrix} A_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} P \tag{7.6}$$

此时的 A^- 具有反射性, 即 $A^- AA^- = A^- \quad A\{1, 2\}$

例 2 设矩阵

$$A = \begin{bmatrix} 0 & 0 & 2 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

试用初等行、列变换法求 A^- 。

解 因为 $\text{rank}A=2$, 把 A 的第一、三两列对换, 得

$$PAQ = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

其中 $P = E_4 \quad Q = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad A_1 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \quad A_1^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}$

$A_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad A_3 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad A_4 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

所以

$$A^- = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} E_4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \end{bmatrix}$$

3 ° 初等行变换法

设 $m \times n$ 阶矩阵 A 的秩等于 r , 经过初等行变换总可以把 A 变为如下分块矩阵形式, 即存在 m 阶可逆阵 P , 使

$$PA = \begin{bmatrix} C \\ B \end{bmatrix}$$

其中 C 是 $r \times n$ 阶行满秩矩阵, B 是 $(m - r) \times n$ 阶矩阵, 且满足

$$BC^{-}C = B$$

不难验证

$$A^{-} = (C^{-} \ 0)P \tag{7.7}$$

此时的 A^{-} 具反射性, 即 $A^{-}AA^{-} = A^{-} \quad A\{1,2\}$

例 3 设矩阵

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

求广义逆矩阵 A^{-} 。

解 因为 $\text{rank}A = 2$, 对 A 作初等行变换, 则有

$$PA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 0 \end{bmatrix}, \quad P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$
$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad B = (0 \ 2 \ 0)$$

由(7.3) 式

$$C^{-} = C_R^{-1} = C^T(CC^T)^{-1}$$
$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

且满足 $BC^{-}C = B$ 。

再由公式(7.7) 得

$$A^{-} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

4 ° 初等列变换法

设 $m \times n$ 阶矩阵 A 的秩等于 r , 经过初等列变换总可以把 A 变成如下分块形式, 即存在 n 阶可逆阵 Q , 使

$$AQ = (C \ B)$$

其中 C 是 $m \times r$ 阶的列满秩矩阵, B 是 $m \times (n - r)$ 阶矩阵, 且满足

$$CC^{-}B = B$$

不难验证

$$A^{-} = Q \begin{bmatrix} C^{-} \\ 0 \end{bmatrix} \tag{7.8}$$

此时的 A^{-} 具反射性, 即 $A^{-}AA^{-} = A^{-} \quad A\{1,2\}$ 。

例 4 设矩阵

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

试用初等变换法求 A^- 。

解 因为 $\text{rank}A=2$, 将 A 的第一、三列对换, 得

$$\begin{aligned}AQ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad Q = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \\C &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \\C^- &= C_L^{-1} = (C^T C)^{-1} C^T = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \\&= \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}\end{aligned}$$

且满足 $CC_L^{-1}B = B$ 。

再由公式(7.8)得

$$\begin{aligned}A^- &= Q \begin{bmatrix} C^- \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix} \\&= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}\end{aligned}$$

7.3 广义逆矩阵 A^+

定理 7.3.1 对任意的 $m \times n$ 阶矩阵 A , A^+ 存在且惟一。

证明 先证存在性

当 $A=0$ 时, 显然 $A^+=0$

设 $\text{rank}A=r-1$, 则 A 具有最大秩分解

$$A = CD$$

其中 C 是 $m \times r$ 阶矩阵, D 是 $r \times n$ 阶矩阵, 且 $\text{rank}C = \text{rank}D = r$, 那么 $C^H C$ 与 DD^H 都是 $r \times r$ 阶可逆矩阵, 下面证明

$$G = D^- C^- = D^H (DD^H)^{-1} (C^H C)^{-1} C^H$$

就是 A^+ , 只需证明 G 满足 Penrose-Moore 四个方程即可。由 7.2 节 A^- 的性质(5), (6) 知 $C^- C = E_r, DD^- = E_r$ 。

(1) $AGA = CDD^- C^- CD = CD = A$

(2) $GAG = D^- C^- CDD^- C^- = D^- C^- = G$

$$\begin{aligned} (3) (AG)^H &= (CDD^H C^H)^H = (CC^H)^H \\ &= [C(C^H C)^{-1} C^H]^H \\ &= C(C^H C)^{-1} C^H = AG \\ (4) (GA)^H &= (D^H C^H CD)^H = (D^H D)^H \\ &= [D^H (DD^H)^{-1} D]^H \\ &= D^H (DD^H)^{-1} D = GA \end{aligned}$$

因此 G 确实是 A^+ , 这证明了 A^+ 的存在性

再证惟一性, 设 A_1^+, A_2^+ 都是 A 的 Penrose-Moore 广义逆, 当然 A_1^+ 与 A_2^+ 都满足四个方程, 从而

$$\begin{aligned} A_1^+ &= A_1^+ AA_1^+ = A_1^+ (AA_1^+)^H = A_1^+ (A_1^+)^H A^H \\ &= A_1^+ (A_1^+)^H (AA_2^+ A)^H = A_1^+ (A_1^+)^H A^H (A_2^+)^H A^H \\ &= A_1^+ (AA_1^+)^H (AA_2^+)^H = A_1^+ AA_1^+ AA_2^+ \\ &= A_1^+ AA_2^+ \\ A_2^+ &= A_2^+ AA_2^+ = (A_2^+ A)^H A_2^+ = A^H (A_2^+)^H A_2^+ \\ &= (AA_1^+ A)^H (A_2^+)^H A_2^+ = A^H (A_1^+)^H A^H (A_2^+)^H A_2^+ \\ &= (A_1^+ A)^H (A_2^+ A)^H A_2^+ = A_1^+ AA_2^+ AA_2^+ \\ &= A_1^+ AA_2^+ \end{aligned}$$

所以 $A_1^+ = A_2^+$

定理 7.3.2 对任何 $m \times n$ 阶矩阵 A, 都有

$$\begin{aligned} (1) (A^+)^+ &= A \\ (2) (A^H)^+ &= (A^+)^H \\ (3) (A^T)^+ &= (A^+)^T \\ (4) A^+ &= (A^H A)^+ A^H = A^H (AA^H)^+ \end{aligned}$$

证明 (1)(2)(3) 留作习题, 证明(4)。

设 $\text{rank} A = r, r=0$ 结论显然成立, 设 $r \geq 1$, 那么 A 有最大秩分解

$$\begin{aligned} A &= CD \\ A^H A &= D^H (C^H CD) = C_1 D_1 \end{aligned}$$

其中 $C_1 = D^H$ 是 $n \times r$ 阶列满秩阵, $D_1 = C^H CD$, 由于 $C^H C$ 是 r 阶可逆阵, 故 D_1 是 $r \times r$ 阶行满秩阵, 则 $C_1 D_1$ 是 $A^H A$ 的最大秩分解, 由定理 7.3.1

$$\begin{aligned} (A^H A)^+ &= D_1^- C_1^- = D_1^H (D_1 D_1^H)^{-1} (C_1^H C_1)^{-1} C_1^H \\ &= D^H C^H C (C^H C D D^H C^H C)^{-1} (D D^H)^{-1} D \\ &= D^H C^H C (C^H C)^{-1} (D D^H)^{-1} (C^H C)^{-1} (D D^H)^{-1} D \\ &= D^H (D D^H)^{-1} (C^H C)^{-1} (D D^H)^{-1} D \end{aligned}$$

所以

$$\begin{aligned} (A^H A)^+ A^H &= D^H (D D^H)^{-1} (C^H C)^{-1} (D D^H)^{-1} D D^H C^H \\ &= D^H (D D^H)^{-1} (C^H C)^{-1} C^H \\ &= D^- C^- = A^+ \end{aligned}$$

同理可证 $A^+ = A^H (AA^H)^+$

定理 7.3.3 对任何矩阵 A , A 是 $m \times n$ 阶矩阵, 都有

$$(1) (A^H A)^+ = A^+ (A^H)^+, (AA^H)^+ = (A^H)^+ A^+$$

$$(2) (A^H A)^+ = A^+ (AA^H)^+ A = A^H (AA^H)^+ (A^H)^+$$

(3) 若 $A^H = A$, 则有

$$(A^2)^+ = (A^+)^2$$

$$A^2 (A^2)^+ = (A^2)^+ A^2 = AA^+$$

$$A^+ A^2 = A^2 A^+$$

$$AA^+ = A^+ A$$

$$(4) AA^+ = (AA^H)(AA^H)^+ = (AA^H)^+(AA^H)$$

$$A^+ A = (A^H A)(A^H A)^+ = (A^H A)^+(A^H A)$$

证明 只证(1), 其余留作习题。设 $G = A^+ (A^H)^+$ 以下验证 G 对于 $A^H A$ 满足定义 7.1 的四个方程。

$$(A^H A)G(A^H A) = A^H AA^+ (A^H)^+ A^H A = A^H AA^+ (AA^+)^H A$$

$$= A^H AA^+ AA^+ A = A^H A$$

$$G(A^H A)G = A^+ (A^H)^+ A^H AA^+ (A^H)^+ = A^+ (AA^+)^H AA^+ (A^H)^+$$

$$= A^+ AA^+ AA^+ (A^H)^+ = A^+ (A^H)^+$$

$$[G(A^H A)]^H = [A^+ (A^H)^+ A^H A]^H = [A^+ (AA^+)^H A]^H$$

$$= [A^+ AA^+ A]^H = (A^+ A)^H = A^+ A$$

$$= A^+ (AA^+)^H A = A^+ (A^H)^+ A^H A = GA^H A$$

$$[(A^H A)G]^H = [A^H AA^+ (A^+)^H]^H = [A^H (AA^+)^H (A^+)^H]^H$$

$$= A^+ AA^+ A = (A^+ A)^H (A^+ A)^H = A^H (A^+)^H A^H (A^+)^H$$

$$= A^H (AA^+)^H (A^+)^H = A^H AA^+ (A^H)^+ = A^H AG$$

于是按定义及惟一性有 $G = (A^H A)^+ = A^+ (A^H)^+$

下面介绍 A^+ 的计算方法

1° 最大秩分解法

设 $m \times n$ 阶矩阵 A 的秩 $= r$, A 的最大秩分解为

$$A = CD$$

其中 C 是 $m \times r$ 阶矩阵, D 是 $r \times n$ 阶矩阵, 且 $\text{rank} C = \text{rank} D = \text{rank} A = r$, 则

$$A^+ = D^H (DD^H)^{-1} (C^H C)^{-1} C^H \quad (7.9)$$

特别当 $\text{rank} A = m$ 时(行满秩阵)

$$A^+ = A^H (AA^H)^{-1} \quad (7.10)$$

当 $\text{rank} A = n$ 时(列满秩阵)

$$A^+ = (A^H A)^{-1} A^H \quad (7.11)$$

例 1 设矩阵 $A = \begin{bmatrix} 1 & -1 \\ 2 & -2 \\ 4 & -4 \end{bmatrix}$

求 A^+ 。

解 令 $B = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$ $C = (1 \quad -1)$

则 $A = BC$
是 A 的最大秩分解, 而

$$C^H (CC^H)^{-1} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} 2^{-1} = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$$

$$(B^H B)^{-1} B^H = \frac{1}{21} (1 \ 2 \ 4) = \frac{1}{21} (1 \ 2 \ 4)$$

所以

$$\begin{aligned} A^+ &= C^H (CC^H)^{-1} (B^H B)^{-1} B^H = \frac{1}{42} \begin{bmatrix} 1 \\ -1 \end{bmatrix} [1 \ 2 \ 4] \\ &= \frac{1}{42} \begin{bmatrix} 1 & 2 & 4 \\ -1 & -2 & -4 \end{bmatrix} \end{aligned}$$

2 ° 奇异值分解法
设 $m \times n$ 阶矩阵 A 的秩等于 r , A 的奇异值分解为

$$A = V \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & 0 \end{bmatrix} U^H$$

其中 $\sigma_r = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_r})$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$ 是 $A^H A$ 的正特征值, $U = \{ u_1 \dots u_r \dots u_{r+1} \dots u_n \}$ 是 $A^H A$ 的 n 个单位正交特征向量, $U = (U_1, U_2)$ $U_1 = (u_1, \dots, u_r)$ 则

$$A^+ = U_1^{-2} U_1^H A^H \tag{7.12}$$

这是因为 $A = V_1 \sigma_r U_1^H$ 是与奇异值有关的最大秩分解, 其中 $V_1 = A U_1^{-1}$ 于是

$$\begin{aligned} A^+ &= (\sigma_r U_1^H)^H [\sigma_r U_1^H (\sigma_r U_1^H)^H]^{-1} (V_1^H V_1)^{-1} V_1^H \\ &= U_1^{-1} (\sigma_r^2)^{-1} V_1^H \end{aligned}$$

由于 $V_1 = A U_1^{-1}$, 则可得到

$$\begin{aligned} A^+ &= U_1^{-1} (A U_1^{-1})^H \\ &= U_1^{-1} \sigma_r^{-1} U_1^H A^H \\ &= U_1^{-2} U_1^H A^H \end{aligned}$$

例 2 设

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

用奇异值分解法求 A^+ 。

解 $A^H A = \begin{bmatrix} -1 & 2 \\ 0 & 0 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 \\ 2 & 0 & -2 \end{bmatrix} = \begin{bmatrix} 5 & 0 & -5 \\ 0 & 0 & 0 \\ -5 & 0 & 5 \end{bmatrix}$

$$|E - A^H A| = \lambda^2 (\lambda - 10)$$

因此 $A^H A$ 的特征值 $\lambda_1 = 10, \lambda_2 = \lambda_3 = 0$

求出对应于 $\lambda_1 = 10$ 的单位特征向量 $U_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix} = U_1$

所以

$$\begin{aligned} A^+ &= U_1 \lambda_1^{-2} U_1^H A^H = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \frac{1}{10} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 0 & 0 \\ 1 & -2 \end{bmatrix} \\ &= \frac{1}{10} \begin{bmatrix} -1 & 2 \\ 0 & 0 \\ 1 & -2 \end{bmatrix} \end{aligned}$$

3 ° 谱分解法

设 $m \times n$ 阶矩阵 A 的秩等于 r , $A^H A$ 的 S 个互异特征值为 $\lambda_1, \dots, \lambda_S, \lambda_1, \dots, \lambda_S$ 是它们分别的重数, $\lambda_1 + \dots + \lambda_S = n$, $A^H A$ 的谱分解为

$$A^H A = \sum_{i=1}^S \lambda_i S_i$$

则

$$A^+ = (A^H A)^+ A^H = \sum_{i=1}^S \mu_i \frac{p_i(A^H A)}{p_i(\lambda_i)} A^H \tag{7.13}$$

其中 $p_i(\lambda) = (\lambda - \lambda_1) \dots (\lambda - \lambda_{i-1})(\lambda - \lambda_{i+1}) \dots (\lambda - \lambda_S)$

$$\mu_i = \frac{1}{\lambda_i} \quad (i = 1, 2, \dots, S, \text{ 若 } \lambda_i = 0, \text{ 则 } \mu_i = 0)$$

证明略。

例 3 设矩阵

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

利用谱分解法求 A^+ 。

解 先求 $A^H A$ 的特征值。

$$A^H A = \begin{bmatrix} 6 & 2 & -2 \\ 2 & 2 & -2 \\ -2 & -2 & 2 \end{bmatrix}$$

$$|E - A^H A| = (\lambda - 2)(\lambda - 8)$$

$A^H A$ 的特征值为 $\lambda_1 = 2, \lambda_2 = 8, \lambda_3 = 0$, 则 $\mu_1 = \frac{1}{2}, \mu_2 = \frac{1}{8}, \mu_3 = 0$, 再构造二次多项式

$$\begin{aligned} p_1(\lambda) &= (\lambda - 2)(\lambda - 3) \\ p_2(\lambda) &= (\lambda - 1)(\lambda - 3) \\ p_3(\lambda) &= (\lambda - 1)(\lambda - 2) \end{aligned}$$

分别求出

$$p_1(\lambda_1) = -12 \qquad p_2(\lambda_2) = 48$$

因为 $\mu = 0$, 故无需求 $p_3(\lambda_3)$, 又求出

$$\begin{aligned} p_1(A^H A) &= (A^H A - 8E) A^H A \\ p_2(A^H A) &= (A^H A - 2E) A^H A \end{aligned}$$

最后求得 A^+

$$\begin{aligned} A^+ &= \sum_{i=1}^3 \mu \frac{p_i(A^H A)}{p_i(\lambda_i)} A^H \\ &= \left[\frac{1}{2} \frac{A^H A - 8E}{-12} A^H A + \frac{1}{8} \frac{A^H A - 2E}{48} A^H A \right] A^H \\ &= \frac{1}{8 \times 48} [-16A^H A + 128E + A^H A - 2E] A^H A A^H \\ &= \frac{1}{8 \times 16} (42E - 5A^H A) A^H A A^H \\ &= \frac{1}{8 \times 8} \begin{bmatrix} 6 & -5 & 5 \\ -5 & 16 & 5 \\ 5 & 5 & 16 \end{bmatrix} A^H A A^H \\ &= \frac{1}{8} \begin{bmatrix} 2 & -2 & 2 & 2 \\ -1 & 3 & -1 & 1 \\ 1 & -3 & 1 & -1 \end{bmatrix} \end{aligned}$$

为了介绍求 A^+ 的第 4 种方法, 先证明一个引理
引理 对任意的 n 阶 Hermite 矩阵 B , 恒有

$$\begin{aligned} BB^+ &= \lim_{\epsilon \rightarrow 0} (B + \epsilon E)^{-1} B \\ &= \lim_{\epsilon \rightarrow 0} B (B + \epsilon E)^{-1} \end{aligned}$$

证明 因 B 是 Hermite 阵, 故 B 的 n 个特征值 $\lambda_1, \dots, \lambda_n$ 皆为实数, $B + \epsilon E$ 的特征值为 $\lambda_1 + \epsilon, \dots, \lambda_n + \epsilon$, 所以对充分小的 $|\epsilon| > 0$, 可使 $\lambda_i + \epsilon \neq 0, (i = 1, 2, \dots, n)$, 从而 $|B + \epsilon E| \neq 0$, 故 $B + \epsilon E$ 可逆。设

$$B = U U^H \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n), U \text{ 是酉阵。}$$

由定理 7.3.3 $B = BB^+ B = BBB^+$

$$\begin{aligned} (B + \epsilon E)^{-1} B &= (B + \epsilon E)^{-1} B^2 B^+ \\ &= U (\Lambda + \epsilon E)^{-1} U^H U^{-2} U^H B^+ \\ &= U (\Lambda + \epsilon E)^{-1} \Lambda^2 U^H B^+ \end{aligned}$$

于是

$$\lim_{\epsilon \rightarrow 0} (B + \epsilon E)^{-1} B = \lim_{\epsilon \rightarrow 0} [U (\Lambda + \epsilon E)^{-1} \Lambda^2 U^H B^+]$$

$$\begin{aligned} &= U \left[\lim_{\epsilon \rightarrow 0} \text{diag} \left[\frac{\epsilon^2}{\epsilon^2 + 1}, \dots, \frac{\epsilon^2}{\epsilon^2 + n} \right] \right] U^H B^+ \\ &= U U^H B^+ = B B^+ \end{aligned}$$

4 ° 极限法

设 A 是 m× n 阶矩阵, 则

$$\begin{aligned} A^+ &= \lim_{\epsilon \rightarrow 0} (A^H A + \epsilon E)^{-1} A^H \\ &= \lim_{\epsilon \rightarrow 0} A^H (A A^H + \epsilon E)^{-1} \end{aligned} \tag{7.14}$$

证明 因为

$$A^H = A^H (A^H)^+ A^H = A^H (A^+)^H A^H = A^H (A A^+)^H = A^H A A^+$$

由引理

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} (A^H A + \epsilon E)^{-1} A^H &= \lim_{\epsilon \rightarrow 0} (A^H A + \epsilon E)^{-1} A^H A A^+ \\ &= A^H A (A^H A)^+ A^+ \\ &= A^+ A A^+ \quad (\text{由定理 7.3.3(4)}) \\ &= A^+ \end{aligned}$$

例 4 设

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

用极限法求 A⁺

解 因为

$$\begin{aligned} A^H A &= \begin{bmatrix} 5 & 0 & -5 \\ 0 & 0 & 0 \\ -5 & 0 & 5 \end{bmatrix} \\ A^H A + \epsilon E &= \begin{bmatrix} 5 + \epsilon & 0 & -5 \\ 0 & \epsilon & 0 \\ -5 & 0 & 5 + \epsilon \end{bmatrix} \\ (A^H A + \epsilon E)^{-1} A^H &= \frac{1}{(10 + \epsilon)} \begin{bmatrix} 5 + \epsilon & 0 & 5 \\ 0 & +10 & 0 \\ 5 & 0 & 5 + \epsilon \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 0 & 0 \\ 1 & -2 \end{bmatrix} \\ &= \frac{1}{10 + \epsilon} \begin{bmatrix} -1 & 2 \\ 0 & 0 \\ 1 & -2 \end{bmatrix} \end{aligned}$$

因此

$$A^+ = \lim_{\epsilon \rightarrow 0} (A^H A + \epsilon E)^{-1} A^H = \frac{1}{10} \begin{bmatrix} -1 & 2 \\ 0 & 0 \\ 1 & -2 \end{bmatrix}$$

7.4^{*} 广义逆矩阵的通式

上一节已经详细地讨论了 A^+ 的求法。因为 A^+ 满足 Penrose-Moore 四个方程, 所以 $A^+ \in A\{1\}, A^+ \in A\{1, 2\}, A^+ \in A\{1, 3\}, A^+ \in A\{1, 4\}$ 。因此只要能由各类广义逆矩阵中某一固定的广义逆矩阵建立各类广义逆矩阵的通式, 那么求任一广义逆矩阵 A^-, A_r^-, A_m^-, A_l^- 也就不困难了。本节就要建立这一通式。

定理 7.4.1 $A \in C^{m \times n}, A^- \in A\{1\}$, 则 $A\{1\}$ 的一种通式为

$$G = A^- + V - A^- AVAA^- \tag{7.15}$$

其中 V 是任意一个 $n \times m$ 的矩阵。

证明 先证明 $G \in A\{1\}$, 事实上

$$\begin{aligned}AGA &= A(A^- + V - A^- AVAA^-)A \\&= AA^-A + AVA - AA^- AVAA^-A = A\end{aligned}$$

此即 $G \in A\{1\}$ 。

再设对任意 $G \in A\{1\}$

$$G = A^- + (G - A^-) = A^- + A(G - A^-)AA^-$$

令 $V = G - A^-$, 即得

$$G = A^- + V - A^- AVAA^-$$

即任何的 $G \in A\{1\}$, 都可以表示为(7.15)形式。

定理 7.4.2 设 $m \times n$ 阶矩阵 A 的秩为 $r \geq 1, A = CD$ 是 A 的最大秩分解, 如果 $\hat{C}C = E_r, D\hat{D} = E_r$, 则 $A\{1, 2\}$ 的一种通式为

$$G = \hat{D} \cdot \hat{C} \tag{7.16}$$

证明 由 A^- 的性质(5)(6), 这样的 \hat{C}, \hat{D} 是存在的, 根据(7.2)式、(7.3)式即可求出(但不惟一), 从而对任意的矩阵 $A, G \in A\{1, 2\}$ 是存在的, 不难验证

$$\begin{aligned}AGA &= CD\hat{D}\hat{C}CD = CD = A \\GAG &= \hat{D}\hat{C}CD\hat{D}\hat{C} = \hat{C}\hat{D} = G\end{aligned}$$

故 $G = \hat{D} \cdot \hat{C} \in A\{1, 2\}$

反之, 对任意的 $G \in A\{1, 2\}$, 则有

$$\begin{aligned}AGA &= A \\CDGCD &= CD\end{aligned}$$

两边左乘 \hat{C} , 右乘 \hat{D}

$$\begin{aligned}\hat{C}CDGCD\hat{D} &= \hat{C}CD\hat{D} \\DGC &= E_r\end{aligned}$$

上式表明 $DG = \hat{C}_1, GC = \hat{D}_1$, 使 $\hat{C}_1C = D\hat{D}_1 = E_r$

并且

$$\hat{D}_1 \hat{C}_1 = GCDG = GAG = G$$

这表明 G 可以表示为(7.16)式形式。

定理 7.4.3 设 A 是 $m \times n$ 阶矩阵, 某 $A_m^- \in A\{1, 3\}$, 则 $A\{1, 3\}$ 的一种通式为

$$G = A_m^- + Z(E - AA_m^-) \quad (7.17)$$

其中 Z 是任意的 $n \times m$ 阶矩阵。

证明 由(7.17)式

$$\begin{aligned} AGA &= A[A_m^- + Z(E - AA_m^-)]A \\ &= AA_m^-A + AZA - AZAA_m^-A = A \\ (GA)^H &= A^H(A_m^-)^H + A^H(E - AA_m^-)^HZ^H \\ &= (A_m^-A)^H + A^HZ^H - A^H(A_m^-)^HA^HZ^H \\ &= A_m^-A = [A_m^- + Z(E - AA_m^-)]A = GA \end{aligned}$$

所以 $G \in A\{1, 3\}$ 。

反之, 对任意 $G \in A\{1, 3\}$, 令 $Z = G - A_m^-$, 注意到

$$\begin{aligned} A_m^-A &= A_m^-AGA = (A_m^-A)^H(GA)^H = A^H(A_m^-)^HA^HG^H \\ &= (AA_m^-A)^HG^H = A^HG^H = (GA)^H = GA \end{aligned}$$

则

$$G = A_m^- + Z - (G - A_m^-)AA_m^- = A_m^- + Z(E - AA_m^-)$$

这说明 G 可以表示为(7.17)式的形式。

定理 7.4.4 设 A 是 $m \times n$ 阶矩阵, 某 $A_l^- \in A\{1, 4\}$, 则 $A\{1, 4\}$ 的一种式通为

$$G = A_l^- + (E - A_l^-A)Z \quad (7.18)$$

其中 Z 为任意的 $n \times m$ 阶矩阵。

证明 由(7.18)式

$$\begin{aligned} AGA &= A[A_l^- + (E - A_l^-A)Z]A \\ &= AA_l^-A + AZA - AA_l^-AZA = A \\ (AG)^H &= [AA_l^- + A(E - A_l^-A)Z]^H \\ &= AA_l^- + Z^HA^H - Z^HA^H \\ &= AA_l^- = A[A_l^- + (E - A_l^-A)Z] = AG \end{aligned}$$

说明 $G \in A\{1, 4\}$ 。

反过来, 对任意 $G \in A\{1, 4\}$, 令 $Z = G - A_l^-$, 并注意到

$$\begin{aligned} AA_l^- &= AGAA_l^- = (AG)^H(AA_l^-)^H = G^HA^H(A_l^-)^HA^H \\ &= G^H(AA_l^-A)^H = G^HA^H = AG \end{aligned}$$

则

$$\begin{aligned} G &= A_l^- + Z = A_l^- + Z - A_l^-A(G - A_l^-) \\ &= A_l^- + (E - A_l^-A)Z \end{aligned}$$

这说明 G 可以表示成(7.18)式的形式。

7.5 广义逆矩阵的应用

7.5.1 相容方程组的通解

已知相容方程组

$$A \begin{matrix} \\ \\ \end{matrix} = \begin{matrix} \\ \\ \end{matrix} \quad A \in C^{m \times n}, \quad C^m \tag{7.19}$$

的通解可由它的一个特解及相应齐次线性方程组

$$A \begin{matrix} \\ \\ \end{matrix} = 0 \tag{7.20}$$

的通解之和给出, 而(7.19)的一个特解可由 A 的任意一广义逆 A^- 立即得到, 所以只需讨论(7.20)方程组的通解。

定理 7.5.1 齐次线性方程组 $A \begin{matrix} \\ \\ \end{matrix} = 0$ 的通解是

$$\begin{matrix} \\ \\ \end{matrix} = (E - A^- A) \begin{matrix} \\ \\ \end{matrix}$$

其中 $\begin{matrix} \\ \\ \end{matrix}$ 是任意 n 维向量。

证明 先验证, 对任意的 n 维向量

$$A \begin{matrix} \\ \\ \end{matrix} = A(E - A^- A) \begin{matrix} \\ \\ \end{matrix} = (A - AA^- A) \begin{matrix} \\ \\ \end{matrix} = 0$$

说明 $\begin{matrix} \\ \\ \end{matrix} = (E - A^- A) \begin{matrix} \\ \\ \end{matrix}$ 是解。

反之, 对 $A \begin{matrix} \\ \\ \end{matrix} = 0$ 的任意解 $\begin{matrix} \\ \\ \end{matrix}_0$, 有

$$\begin{matrix} \\ \\ \end{matrix}_0 = \begin{matrix} \\ \\ \end{matrix}_0 - A^- A \begin{matrix} \\ \\ \end{matrix}_0 = (E - A^- A) \begin{matrix} \\ \\ \end{matrix}_0$$

故 $A \begin{matrix} \\ \\ \end{matrix} = 0$ 的任意解都可写成 $(E - A^- A) \begin{matrix} \\ \\ \end{matrix}$ 的形式, 故

$$\begin{matrix} \\ \\ \end{matrix} = (E - A^- A) \begin{matrix} \\ \\ \end{matrix}$$

是 $A \begin{matrix} \\ \\ \end{matrix} = 0$ 的通解。

定理 7.5.2 相容方程组

$$A \begin{matrix} \\ \\ \end{matrix} = \begin{matrix} \\ \\ \end{matrix} \quad (A \in C^{m \times n}, \quad C^m)$$

的通解是

$$\begin{matrix} \\ \\ \end{matrix} = A^- \begin{matrix} \\ \\ \end{matrix} + (E - A^- A) \begin{matrix} \\ \\ \end{matrix}$$

这里 $\begin{matrix} \\ \\ \end{matrix}$ 是任意的 n 维向量。

7.5.2 相容方程组的最小范数解

对于给定的相容方程组

$$A \begin{matrix} \\ \\ \end{matrix} = \begin{matrix} \\ \\ \end{matrix} \quad (A \in C^{m \times n}, \quad C^m)$$

其解一般是不惟一的, 而在一些实际中需要在它的一切解中求出 2-范数最小的解, 这样的解称为该方程组的最小范数解。

定理 7.5.3 对任意的 $G \in A\{1, 3\}$, $G \begin{matrix} \\ \\ \end{matrix}$ 都是相容方程组 $A \begin{matrix} \\ \\ \end{matrix} = \begin{matrix} \\ \\ \end{matrix}$ 的最小范数解, 且 $A \begin{matrix} \\ \\ \end{matrix}$ 的最小范数解是惟一的。

证明 因为线性方程组 $A \begin{matrix} \\ \\ \end{matrix} = \begin{matrix} \\ \\ \end{matrix}$ 的任意解 $\begin{matrix} \\ \\ \end{matrix}$ 可以表示为

$$\begin{matrix} \\ \\ \end{matrix} = G \begin{matrix} \\ \\ \end{matrix} + (E - GA) \begin{matrix} \\ \\ \end{matrix}$$

的形式, 取 $G_0 = A_0$, 则

$$\begin{aligned} \frac{1}{2} \|G_0\|^2 &= \|G_0\|^2 + \|(E - GA_0)G_0\|^2 \\ &= [G_0 + (E - GA_0)G_0]^H [G_0 + (E - GA_0)G_0] \\ &= \|G_0\|^2 + \|(E - GA_0)G_0\|^2 + (G_0)^H (E - GA_0) \\ &\quad + (E - GA_0)^H (G_0) \end{aligned}$$

而 $(G_0)^H (E - GA_0) = (GA_0)^H (E - GA_0)$

$$\begin{aligned} &= A_0^H (GA_0)^H (E - GA_0) \\ &= A_0^H (GA_0) (E - GA_0) \\ &= A_0^H (GA_0 - GA_0) = 0 \end{aligned}$$

$$\begin{aligned} (E - GA_0)^H (G_0) &= (E - GA_0)^H (GA_0) \\ &= (GA_0 - GA_0)^H = 0 \end{aligned}$$

所以 $\frac{1}{2} \|G_0\|^2 = \|G_0\|^2 + \|(E - GA_0)G_0\|^2$

由此知 G_0 是 $A =$ 的最小范数解。

设 G_0 也是 $A =$ 的最小范数解, 则必有

$$\|G_0\|^2 = \|G\|^2$$

必有 $G_0 \in C^n$, 使

$$G_0 = G + (E - GA)G_0$$

由上述证明知 $(E - GA)G_0 = 0$

从而有 $(E - GA)G_0 = 0$

所以 $G_0 = G$

定理 7.5.4 若 $D \in C^{n \times m}$ 对一切 C^m 都使 D 是相容方程组 $A =$ 的最小范数解, 则 $D \in A\{1,3\}$ 。

证明 设 $A = (a_1, a_2, \dots, a_n)$

这里 a_1, a_2, \dots, a_n 分别是 A 的列向量, 当然 $a_i \in R(A)$ 。设 $G \in A\{1,3\}$, 则 D_i, G_i 都是方程组

$$Ax = a_i \quad (i = 1, 2, \dots, n)$$

的最小范数解, 由最小范数解的惟一性, 知

$$D_i = G_i \quad (i = 1, \dots, n)$$

从而有

$$DA = GA$$

注意到 $G \in A\{1,3\}$

$$\begin{aligned} ADA &= AGA = A \\ (DA)^H &= (GA)^H = GA = DA \end{aligned}$$

所以 $D \in A\{1,3\}$ 。

7.5.3 不相容线性方程组 $A =$ 的最小二乘解

在许多实际问题中, 例如数据处理问题或与正态分布有关的统计问题等所得到的线性方

程组 $A = b$, 往往是不相容的, 当然不能求得满足方程组的 x , 即它没有通常意义下的解, 但是在实际问题中, 往往要求出 $x \in C^n$, 使

$$\|Ax - b\|_2$$

为最小, 并把满足这个要求的 x 作为不相容线性方程组 $A = b$ 的一个近似解, 称为最小二乘解, 但需注意的是最小二乘解并不是 $A = b$ 的解。

定理 7.5.5 对任何 $G \in A\{1,4\}$, 则 $x = Gb$ 是不相容方程组 $A = b$ 的最小二乘解。

证明 对任意的 $x \in C^n$

$$\begin{aligned} \|Ax - b\|_2^2 &= \|Ax - AGx + AGx - b\|_2^2 \\ &= (AGx - b + A - AG)x^H (AGx - b + A - AG) \\ &= \|AGx - b\|_2^2 + \|A - AG\|_2^2 + (AGx - b)^H (A - AG) \\ &\quad + (A - AG)^H (AGx - b) \end{aligned}$$

而

$$\begin{aligned} (AGx - b)^H (A - AG) &= x^H (AG)^H A - x^H (AG)^H AG - b^H A + b^H AG \\ &= x^H A - x^H A - b^H AG + b^H AG = 0 \\ (A - AG)^H (AGx - b) &= [(AGx - b)^H (A - AG)]^H = 0 \end{aligned}$$

所以

$$\|Ax - b\|_2^2 = \|AGx - b\|_2^2 + \|A - AG\|_2^2 \quad \|AGx - b\|_2^2$$

说明 $x = Gb$ 就是 $A = b$ 的最小二乘解。

一般说来, 最小二乘解不惟一, 为求通式, 先介绍下面定理。

定理 7.5.6 $x \in C^n$ 是不相容方程组 $A = b$ 的最小二乘解的充分必要条件是 x 为方程组

$$A = AG$$

的解, 这里 $G \in A\{1,4\}$

证明 因为 x 是不相容方程组 $A = b$ 的最小二乘解, 对 $G \in A\{1,4\}$ 由定理 7.5.5 知

$$\|Ax - b\|_2^2 = \|AGx - b\|_2^2$$

而

$$\|A - AG\|_2^2 = 0$$

所以

$$A = AG$$

反过来, 若 x 是方程组

$$A = AG$$

的解, 则必有

$$A = AG$$

$$\|Ax - b\|_2^2 = \|AGx - b\|_2^2$$

这说明 x 是 $A = b$ 的最小二乘解。

定理 7.5.7 不相容方程组 $A = b$ 的最小二乘解的通式是

$$x = Gb + (E - A^+A)x_0 \quad (G \in A\{1,4\})$$

证明 由定理 7.5.6 知 x 是方程组 $A = b$ 的最小二乘解的充分必要条件为 x 是方程组

$$A = AG$$

的解, 也即有

$$A(x - Gb) = 0$$

由定理 7.5.1 知齐次线性方程组 $A(x - Gb) = 0$ 的通解为

$$-G = (E - A^+A)$$

这里 η 是任意 n 维列向量, 也即

$$= G + (E - A^+A)\eta$$

7.5.4 不相容线性方程组的极小最小二乘解

因为最小二乘解一般不是惟一的, 通常把它们中 2-范数最小的一个称为不相容方程组 $Ax = b$ 的极小最小二乘解(最佳逼近解)。

由定理 7.5.6 知道, x 是不相容方程组 $Ax = b$ 的最小二乘解的充分必要条件是 x 是相容方程组

$$Ax = AG$$

的解, 而该方程组的最小范数解是

$$x = D(AG)$$

这里 $D \in A\{1, 3\}$, 从而 $x = DAG$ 就是极小最小二乘解。而且这样的解是惟一的。

注意到

$$DAG = A^+$$

(证明留作练习), 故有

定理 7.5.8 不相容方程组 $Ax = b$ 必有惟一的极小最小二乘解

$$x = A^+b$$

综上所述, 因为 $A^+ \in A\{1\}$, $A^+ \in A\{1, 3\}$, $A^+ \in A\{1, 4\}$ 所以对于方程组可得下表:

$Ax = b$ 相容时 $x = A^+b + (E - A^+A)\eta$ 是其通解 $x = A^+b$ 是其最小范数解	$Ax = b$ 不相容时 $x = A^+b + (E - A^+A)\eta$ 是其最小二乘解通式 $x = A^+b$ 是其极小最小二乘解
--	--

习 题 7

1. 证明 A^+ 的性质:
(1) 对于任意的 $m \times n$ 阶矩阵 A , $\text{rank} A^+ = \text{rank} A$;
(2) $(A^+)^H = (A^H)^+$, $(A^+)^T = (A^T)^+$;
(3) 若 $m = n = r$, 则 $A\{1\}$ 只有惟一元素 A^{-1} ;
(4) AA^+ , A^+A 均为幂等矩阵, 且
$$\text{rank} AA^+ = \text{rank} A^+A = \text{rank} A$$
2. 设 A 是 n 阶方阵, 证明: 总有可逆的 A^+ 存在。
3. 设 P 是 m 阶可逆阵, Q 是 n 阶可逆阵, 且 $B = PAQ$, 证明: $Q^{-1}A^+P^{-1} \in B\{1\}$ 。
4. 设 $A \in C^{m \times n}$, $G, D \in A\{1\}$, 证明:
$$GAD \in A\{1, 2\}$$
5. 设 $A^2 = A = A^H$, 证明: $A = A^+$ 。
6. 若 $A = BC$ 是 A 的最大秩分解, 证明

$A^+ = C^+ B^+$

7. 求下列矩阵

$A_1 = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \\ 1 & 0 & 2 \end{bmatrix}$

$A_2 = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$

的 $A_1^{(1)}, A_2^{(1)}$ 。

8. 证明 A^+ 的性质

- (1) $(A^+)^+ = A$
- (2) $(A^H)^+ = (A^+)^H$
- (3) $(A^T)^+ = (A^+)^T$

9. 设 U 是 m 阶酉阵, V 是 n 阶酉阵, $A \in C^{m \times n}$

证明: $(UAV^H)^+ = VA^+U^H$

10. 证明: 线性方程组

$A = (A \in C^{m \times n}, C^m)$

有解的充分必要条件是 $AA^+ = I_n$ 。

11. 证明: 方程组

$A^H A = A^H (A \in C^{m \times n}, C^m)$

总是相容的。

12. 用各种不同的方法求 A^+

(1) $A = \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 2 & 4 \end{bmatrix}$

(2) $A = \begin{bmatrix} i & 0 \\ 1 & i \\ 0 & 1 \end{bmatrix}$

(3) $A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$

(4) $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & 0 \\ 2 & 1 & -1 \end{bmatrix}$

13. 已知

$A = \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$

求线性方程组 $Ax = b$ 的通解及极小范数解。

14. 已知

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 2 & 4 \end{bmatrix} \qquad b = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

求不相容线性方程组 $Ax = b$ 的最小二乘解和极小最小二乘解。

第 8 章

特征值的估计及广义特征值

n 阶方阵 A 的 n 个特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 对应复平面上 n 个点, 在没有求出 $\lambda_1, \lambda_2, \dots, \lambda_n$ 的情况下, 对它们的位置给出一个大致的范围, 这就是特征值的估计问题。

特征值的估计无论在数学本身或在应用科学中都很重要。例如在线性方程组的迭代法中讨论解的敛散性时, 就要估计线性方程组的系数矩阵的特征值是否都在复平面上以原点为中心的圆内; 在微分方程的稳定性理论中, 要估计矩阵的特征值是否有负的实部, 即表示特征值的点是否在复平面的左半平面上; 在差分方法的稳定性理论、自动控制理论中都要求估计矩阵的特征值是否在复平面上的某一确定的区域内。另一方面计算矩阵的特征值并非易事。而在许多问题中也只要求对特征值的模或表示特征值的点之位置作出大致的估计, 而无需求出特征值, 这时特征值的估计问题就更加重要了。

8.1 特征值的界的估计

这一节介绍特征值的界的 n 个重要不等式。为了方便, 对于给定的 $A \in C^{n \times n}$, 令

$$B = (b_{ij}) = \frac{1}{2}(A + A^H)$$
$$C = (c_{ij}) = \frac{1}{2}(A - A^H)$$

显然 B 和 C 分别是 Hermite 阵和反 Hermite 阵, 此外还假定 A, B, C 的特征值分别为 $\{\lambda_1, \lambda_2, \dots, \lambda_n\}, \{\mu_1, \mu_2, \dots, \mu_n\}, \{i\nu_1, i\nu_2, \dots, i\nu_n\}$, 且满足 $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|, \mu_1 \geq \mu_2 \geq \dots \geq \mu_n, \nu_1 \geq \nu_2 \geq \dots \geq \nu_n$ 。

定理 8.1.1 (Schur 不等式, 1909 年) 若 $A = (a_{ij}) \in C^{n \times n}$ 的特征值为 $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$, 则有

$$\sum_{i=1}^n |\lambda_i|^2 \leq \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \tag{8.1}$$

其中等号当且仅当 A 是正规矩阵时成立。

证明 因为 A 酉阵相似于上三角阵, 即存在酉阵 U 及上三角阵 T , 使

$$U^H A U = T \quad U^H A^H U = T^H$$

从而有

$$U^H A A^H U = T T^H$$

$$\text{tr}(A A^H) = \text{tr}(U^H A A^H U) = \text{tr}(T T^H)$$

注意到 T 的主对角元就是 A 的特征值, 便有

$$\sum_{i=1}^n |t_{ii}|^2 = \sum_{i=1}^n |t_{ii}|^2 \quad \sum_{i=1}^n \sum_{j=1}^n |t_{ij}|^2 = \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2$$

上述不等式中的等号成立当且仅当 $t_{ij} = 0 (i \neq j)$, 即当且仅当 A 酉相似于对角阵时成立, 又根据第 3 章习题 26 结论当且仅当 A 为正规矩阵。

推论 8.1.1 A 如定理 8.1.1 所设, 则有

$$1. \quad |t_{ii}| \leq \max_{i,j} |a_{ij}| \quad (8.2)$$

$$2. \quad |\text{Re}(t_{ii})| \leq \max_{i,j} |b_{ij}| \quad (8.3)$$

$$3. \quad |\text{Im}(t_{ii})| \leq \max_{i,j} |c_{ij}| \quad (8.4)$$

证明 根据定理 8.1.1 的证明可得

$$U^H B U = U^H (A + A^H) U / 2 = (T + T^H) / 2$$

$$U^H C U = U^H (A - A^H) U / 2 = (T - T^H) / 2$$

故有

$$\sum_{i=1}^n \left| \frac{t_{ii} + \overline{t_{ii}}}{2} \right|^2 + \sum_{j=1}^n \sum_{i=1}^{j-1} \frac{|t_{ij}|^2}{2} = \sum_{i=1}^n \sum_{j=1}^n |b_{ij}|^2$$

$$\sum_{i=1}^n \left| \frac{t_{ii} - \overline{t_{ii}}}{2} \right|^2 + \sum_{j=1}^n \sum_{i=1}^{j-1} \frac{|t_{ij}|^2}{2} = \sum_{i=1}^n \sum_{j=1}^n |c_{ij}|^2$$

亦即

$$\sum_{i=1}^n |\text{Re}(t_{ii})|^2 + \sum_{i=1}^n \sum_{j=1}^n |b_{ij}|^2 \leq n^2 \max_{i,j} |b_{ij}|^2$$

$$\sum_{i=1}^n |\text{Im}(t_{ii})|^2 + \sum_{i=1}^n \sum_{j=1}^n |c_{ij}|^2 \leq n^2 \max_{i,j} |c_{ij}|^2$$

又由 Schur 不等式

$$\sum_{i=1}^n |t_{ii}|^2 \leq \sum_{j=1}^n \sum_{i=1}^n |a_{ij}|^2 \leq n^2 \max_{i,j} |a_{ij}|^2$$

由以上三个不等式显然可推知推论 8.1.2 成立。

推论 8.1.2 若 A 是实矩阵, 则更有

$$|\text{Im}(t_{ii})| \leq \sqrt{\frac{n(n-1)}{2}} \max_{i,j} |c_{ij}| \quad (8.5)$$

证明 根据上述推论, 已有不等式

$$\sum_{i=1}^n |\text{Im}(t_{ii})|^2 \leq \sum_{i=1}^n \sum_{j=1}^n |c_{ij}|^2$$

注意到当 A 是实矩阵时, $C = (A - A^H) / 2$ 的主对角元 $c_{ii} = 0 (i = 1, 2, \dots, n)$ 故上可改写为

$$\sum_{i=1}^n |\text{Im}(t_{ii})|^2 \leq \sum_{i \neq j} |c_{ij}|^2 \leq n(n-1) \max_{i,j} |c_{ij}|^2$$

再注意到 A 是实方阵, 它的特征多项式是实系数的, 那么对于 A 的复特征值是共轭成对出现。不妨设共有 s 对复特征值, 这时上面的不等式的左端为

$$\sum_{i=1}^n |\operatorname{Im}(\lambda_i)|^2 = 2 \sum_{i=1}^s |\operatorname{Im}(\lambda_i)|^2$$

从而有

$$\begin{aligned} 2 \sum_{i=1}^s |\operatorname{Im}(\lambda_i)|^2 &\leq n(n-1) \max_{i,j} |c_{ij}|^2 \\ \sum_{i=1}^s |\operatorname{Im}(\lambda_i)|^2 &\leq \frac{n(n-1)}{2} \max_{i,j} |c_{ij}|^2 \end{aligned}$$

这样便证得

$$|\operatorname{Im}(\lambda_i)| \leq \sqrt{\frac{n(n-1)}{2}} \max_{i,j} |c_{ij}|$$

由推论 8.1.1 可以得到结论: Hermite 矩阵的特征值皆实数, 这是因为 $c_{ij} = 0$, 故

$$|\operatorname{Im}(\lambda_i)| = 0$$

且反 Hermite 矩阵的特征值皆虚数, 这是因为 $b_{ij} = 0$, 故

$$|\operatorname{Re}(\lambda_i)| = 0$$

当然由推论 8.1.2 可以得到结论: 实对称阵的特征值皆实数。

例 1 估计矩阵 $A = \begin{bmatrix} 1 & -0.8 \\ 0.5 & 0 \end{bmatrix}$ 的特征值上界。

解 应用推论 8.1.2

$$\max_{i,j} |a_{ij}| = 1 \quad \max_{i,j} |b_{ij}| = 1 \quad \max_{i,j} |c_{ij}| = 0.65$$

从而

$$|\lambda| \leq 2, \quad |\operatorname{Re}(\lambda)| \leq 2, \quad |\operatorname{Im}(\lambda)| \leq 1.3$$

若用推论 8.1.3 估计 $\operatorname{Im}(\lambda)$, 则

$$|\operatorname{Im}(\lambda)| \leq \sqrt{\frac{2 \times 1}{2}} \times 0.65 = 0.65$$

实际上 A 的两个特征值 $\lambda_{1,2} = \frac{1 \pm \sqrt{0.6i}}{2}$, 从而 $|\lambda| = 0.63$, $|\operatorname{Re}(\lambda)| = 0.5$, $|\operatorname{Im}(\lambda)| = 0.39$ 。

例 1 表明, 在估计实矩阵的特征值的虚部上界时, 推论 8.1.3 的结果优于推论 8.1.2 的结果。

8.2 圆盘定理

上一节对矩阵 A 的特征值的界做了大致的估计, 本节将对矩阵 A 的特征值在复平面上的位置做出更准确的估计, 这就是圆盘定理, 在此定理之前先给出一个定理。

定理 8.2.1 如果 $A \in \mathbb{C}^{n \times n}$, $A = (a_{ij})$ 是行对角占优的, 即满足

$$|a_{ii}| > \sum_{j=1, j \neq i}^n |a_{ij}| \quad (i = 1, 2, \dots, n)$$

则 A 是非奇异的。

证明 反证法。若 $\det(A) = 0$, 则线性方程组 $AX = 0$ 有非 0 解 $X = (x_1, x_2, \dots, x_n)^T$, 设

$$|x_k| = \max_i |x_i|$$

则显然 $x_k \neq 0$ 。那么便有

$$\begin{aligned} |a_{kk}| |x_k| &= \left| - \sum_{j \neq k}^n a_{kj} x_j \right| \leq \sum_{j \neq k}^n |a_{kj}| |x_j| \\ &\leq |x_k| \sum_{j \neq k}^n |a_{kj}| \end{aligned}$$

从而推出 $|a_{kk}| \leq \sum_{j \neq k}^n |a_{kj}|$, 这与假设矛盾。

推论 8.2.1 如果 $A \in C^{n \times n}$ 是列对角占优, 即满足

$$|a_{jj}| > \sum_{i=1, i \neq j}^n |a_{ij}| \quad (j = 1, 2, \dots, n)$$

则 A 是非奇异的。

定理 8.2.2 (Gerschgorin 定理) 设 $A \in C^{n \times n}$, $A = (a_{ij})$ 则它的 n 个特征值都落在复平面上的 n 个圆盘

$$D_i(A) = \{z \mid |z - a_{ii}| \leq R_i\}, i = 1, 2, \dots, n \tag{8.6}$$

的并集 $(\bigcup_{i=1}^n D_i(A))$ 中, 其中 $R_i = \sum_{j=1, j \neq i}^n |a_{ij}|$ 。

证明 对于 A 的任一特征值 λ_i , 都有

$$| \lambda_i E - A | = 0$$

根据定理 8.2.1, 矩阵 $(\lambda_i E - A)$ 一定不是行对角占优的, 即至少存在一个 k, 使得

$$| \lambda_i - a_{kk} | \leq R_k$$

成立, 这表明 $\lambda_i \in D_k(A) \subset \bigcup_{i=1}^n D_i(A)$ 。

例 1 估计矩阵

$$A = \begin{bmatrix} 1 & -1/2 & -1/2 & 0 \\ -1/2 & 3/2 & i & 0 \\ 0 & -i/2 & 5 & i/2 \\ -1 & 0 & 0 & 5i \end{bmatrix}$$

的特征值分布范围。

解 由 A 可知

$$a_{11} = 1, a_{22} = 3/2, a_{33} = 5, a_{44} = 5i$$

由定理 8.2.2 指定的四个圆盘为

$$\begin{aligned} D_1(A) &= \{z \mid |z - 1| \leq 1\} \\ D_2(A) &= \{z \mid |z - \frac{3}{2}| \leq \frac{3}{2}\} \end{aligned}$$

$$D_3(A) = \{z \mid |z - 5| = 1\}$$

$$D_4(A) = \{z \mid |z - 5i| = 1\}$$

A 的 4 个特征值都落在复平面上由图 8.1 所确定的四个圆盘的并集中。

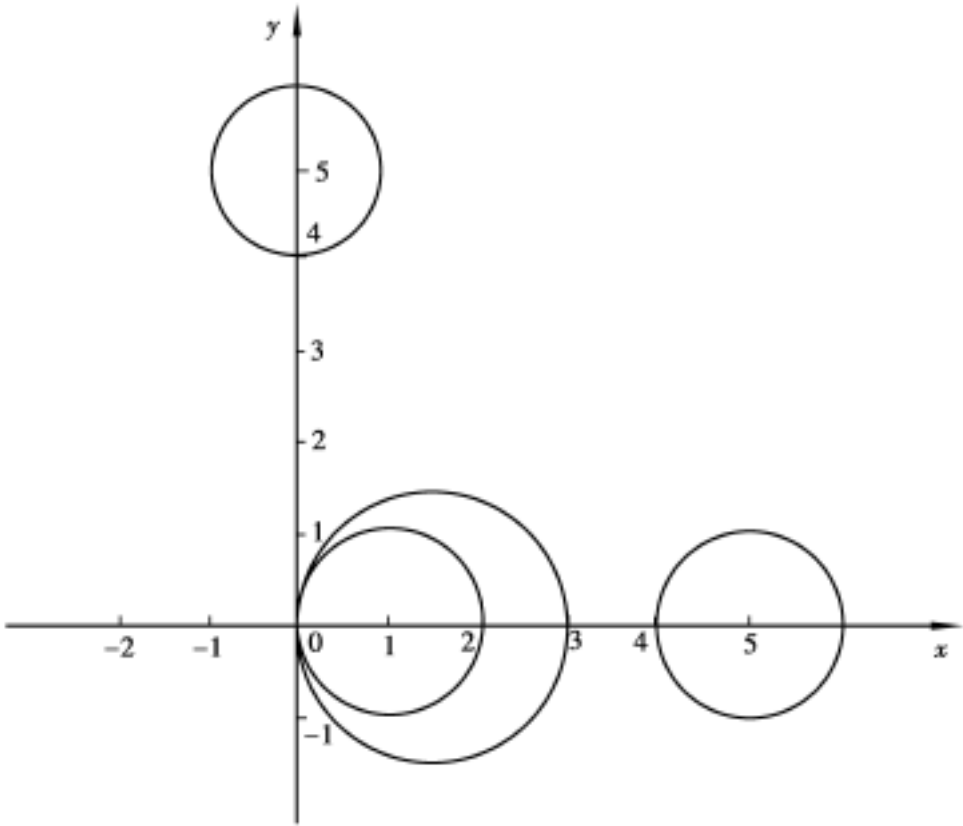


图 8.1

在例 1 中 D_1 与 D_2 重叠在一起(一般地可能两个圆相交), D_1 和 D_2 的并集就构成一个连通域, 通常称相交的几个圆盘所构成的连通域为并集 $\bigcup_{i=1}^n D_i$ 的一个连通部分, 独立的一个圆盘也称为并集的一个连通部分。例 1 中的并集是由 3 个连通部分组成。定理 8.2.2 只说明矩阵 A 的全部特征值在复平面上 n 个圆盘的并集中, 并没有说清楚那个连通部分有几个特征值。下面的圆盘定理 2 给出更详细的结果。

定理 8.2.3(圆盘定理 2) 设 G 是 A 的 m 个圆盘组成的任一连通部分, 则在 G 中必有且只有 A 的 m 个特征值(主对角线上有相同元素时须重复计算, 有相同特征值时 also 须按重数重复计算)。

上述定理说明, 孤立的盘中含有且只含有 A 的一个特征值, 两个圆盘构成的连通部分含有 A 的 2 个特征值, 但不保证每一个圆盘中都有 A 的 1 个特征值。

例 2 估计矩阵

$$A = \begin{bmatrix} 0 & -0.4 \\ 0.9 & 1 \end{bmatrix}$$

的特征值的分布范围。

解 由定理 8.2.2, A 的特征值分布在复平面上下面两个圆盘中

$$|z - 1| = 0.9, \quad |z| = 0.4$$

如图 8.2 所示, 这里两个圆盘构成一个连通部分。

求出 A 的特征值

$$|E - A| = \lambda^2 - \lambda + 0.36$$

A 的两个特征值为

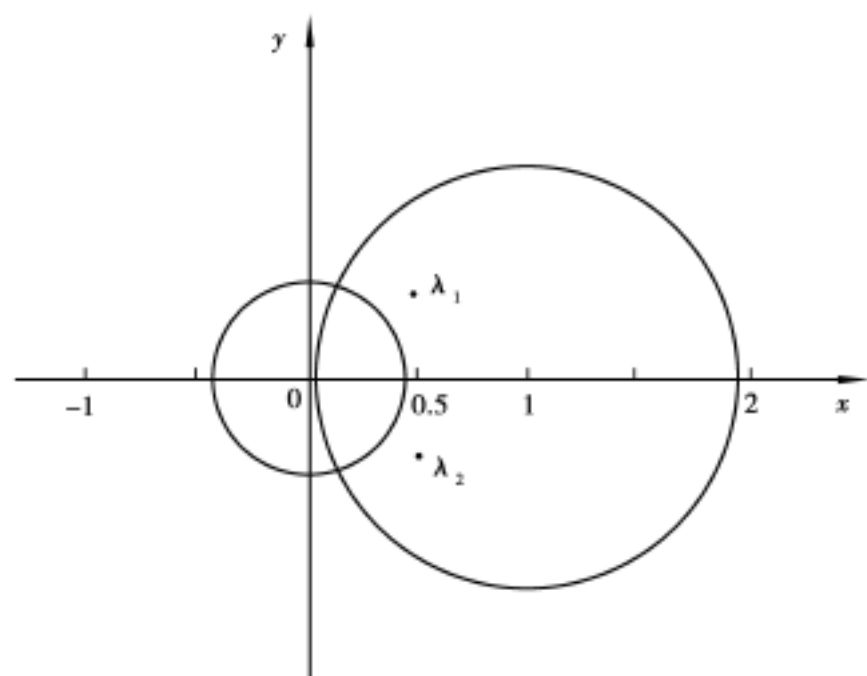


图 8.2

$$\lambda_1 = \frac{1 + \sqrt{0.44}i}{2}$$
$$\lambda_2 = \frac{1 - \sqrt{0.44}i}{2}$$

那么 λ_1 与 λ_2 的模为

$$|\lambda_1| = |\lambda_2| = 0.6 > 0.4$$

可见两个特征值都不在圆盘 $|z| \leq 0.4$ 中,而同落在圆盘 $|z - 1| \leq 0.9$ 中。

如果 A 的 n 个圆盘两两互不相交,则 A 有 n 个互异的特征值,从而有

推论 8.2.2 如果方阵 $A \in \mathbb{C}^{n \times n}$ 有 n 个两两互不相交的圆盘,则 A 相似于对角阵。

推论 8.2.3 如果方阵 $A \in \mathbb{R}^{n \times n}$ 有 n 个两两互不相交的圆盘,则 A 的特征值皆为实数。

证明 实方阵 A 的圆盘的圆心都位于实轴上。由于这些圆盘两两互不相交,所以有 n 个互异特征值分含于每一个圆盘中。实方阵 A 的特征多项式是实系数,它若有复根(虚部不为 0)的话,必是共轭成对的,这与每个圆盘中只有一个特征值矛盾。

为了进一步改善圆盘定理估计特征值的准确性,引进下面的定理。

定理 8.2.4 设 A 是 n 级方阵, b_1, b_2, \dots, b_n 是一组正数,记

$$r_i = \max_{j=1, \dots, n} |a_{ij}|, G_i = \{z \mid |z - a_{ii}| \leq r_i\} \quad (8.7)$$
$$(i = 1, 2, \dots, n)$$

则 A 的特征值都在并集 $\bigcup_{i=1}^n G_i$ 中。

证明 记矩阵 $B = \text{diag}(b_1, b_2, \dots, b_n)$, 因 $b_i > 0 (i = 1, 2, \dots, n)$, 故 B 是可逆阵,且

$$D = B^{-1}AB = \begin{bmatrix} b_1^{-1} & & & \\ & b_2^{-1} & & \\ & & \ddots & \\ & & & b_n^{-1} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} b_1 & & & \\ & b_2 & & \\ & & \ddots & \\ & & & b_n \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} \frac{b_2}{b_1} & \cdots & a_{1n} \frac{b_n}{b_1} \\ a_{21} \frac{b_1}{b_2} & a_{22} & \cdots & a_{2n} \frac{b_n}{b_2} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} \frac{b_1}{b_n} & a_{n2} \frac{b_2}{b_n} & \cdots & a_{nn} \end{bmatrix}$$

因为 D 与 A 相似, 故有相同的特征值, 而 D 的特征值由定理 8.2.2 知应在并集 $\bigcup_{i=1}^n G_i$ 中。

例 3 估计矩阵

$$A = \begin{bmatrix} 1 & 0.02 & 0.11 \\ 0.01 & 0.9 & 0.14 \\ 0.02 & 0.01 & 0.5 \end{bmatrix}$$

的特征值的分布范围。

解 由定理 8.2.2 及定理 8.2.3, A 的特征值分布在下面三个圆盘

$$\begin{aligned} D_1: & |z - 1| = 0.13 \\ D_2: & |z - 0.9| = 0.15 \\ D_3: & |z - 0.5| = 0.03 \end{aligned}$$

所组成的并集中, 其中 D_1 与 D_2 组成一个连通部分, D_3 是孤立的一个连通部分, 那么在 $D_1 \cup D_2$ 中有 A 的两个特征值, 在 D_3 中有一个 A 的特征值。为了得到更精确的结果, 选正实数 $b_1 = b_2 = 1, b_3 = \frac{1}{10}$, 做矩阵 $B = \text{diag}(1, 1, \frac{1}{10})$, 则

$$D = B^{-1}AB = \begin{bmatrix} 1 & 0.02 & 0.011 \\ 0.01 & 0.9 & 0.014 \\ 0.2 & 0.1 & 0.5 \end{bmatrix}$$

与 A 相似, 而 D 的特征值分布在下面三个圆盘

$$\begin{aligned} G_1: & |z - 1| = 0.031 \\ G_2: & |z - 0.9| = 0.024 \\ G_3: & |z - 0.5| = 0.3 \end{aligned}$$

的并集中。圆盘 G_1, G_2, G_3 是分隔开的, 特别 $D_3 \cap G_3 = \emptyset$, 所以 A 的特征值分布在三个孤立的圆盘 G_1, G_2, D_3 中。

再给一个推论。

推论 8.2.4 设 A 是 n 阶实方阵, 且

$$a_{ii} > \sum_{j=1, j \neq i}^n |a_{ij}| \quad i = 1, 2, \dots, n$$

则 A 的所有特征值都分布在复平面的右半平面。

证明 只证明 A 的每个特征值 λ 的实部 $\text{Re}(\lambda) > 0$ 即可。

由题设条件矩阵 A 是行对角占优矩阵, 因此 A 是非奇异矩阵。设 λ 是 A 的任意一个特征

值 $\lambda = a + bi$, 当然 $b \neq 0$, 由定理 8.2.2, 至少存在某 $k, 1 \leq k \leq n$, 使

$$| \lambda - a_{kk} | = \left| \sum_{j=1, j \neq k}^n a_{kj} \right| < a_{kk}$$

即

$$| a + bi - a_{kk} | < a_{kk}$$

故

$$(a - a_{kk})^2 + b^2 < a_{kk}^2$$

从而有

$$0 < a^2 + b^2 < 2aa_{kk}$$

因此 $\operatorname{Re}(\lambda) = a > 0$ 。

8.3 谱半径的估计

$n \times n$ 矩阵 A 的谱半径的概念及其估计, 对于矩阵的幂级数 $\sum_{k=0}^{\infty} a_k A^k$ 的敛散性的研究以及矩阵迭代敛散性问题的讨论都起着重要的作用。本节在谱半径概念的基础上, 讨论谱半径的估计。

如前所述, 若以 $\lambda_1, \lambda_2, \dots, \lambda_n$ 表示 $n \times n$ 矩阵 A 的 n 个特征值, 则称

$$\rho(A) = \max_i \{ |\lambda_i|, i = 1, 2, \dots, n \}$$

为矩阵 A 的谱半径。

从几何上说, 复平面上以原点为圆心, $\rho(A)$ 为半径的圆 $|z| = \rho(A)$ 恰好包含 A 的全部特征值, 或者说 $\rho(A)$ 是复平面上以原点为圆心能包含 A 的全部特征值的圆的半径之最小值。

需注意, 谱半径 $\rho(A)$ 一般不是矩阵范数。

例如

$$A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \quad \rho(A) = 0$$

但 $\rho(A) = 0$, 因此 $\rho(A)$ 未必满足矩阵范数正定性条件。

又例如

$$A = \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

易求得 $\rho(A) = \rho(B) = 1$, 而

$$A + B = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}$$

不难求得 $\rho(A + B) = 4$, 从而有

$$\rho(A + B) = 4 > 2 = \rho(A) + \rho(B)$$

因此 $\rho(A + B)$ 未必有三角不等式。

下面介绍估计 $\rho(A)$ 的方法。

在前面第 5 章曾介绍一个结论: 对于矩阵的算子范数 $\|A\|_a$, 总有 $\rho(A) \leq \|A\|_a$ 。事实

上这个结论对矩阵的任意范数 $\|\cdot\|$ 都成立的。

定理 8.3.1 设 A 是 $n \times n$ 矩阵, 则对任意一种矩阵范数 $\|A\|$, 都有

$$\|A\| \geq \rho(A) \tag{8.8}$$

证明 设 $\|\cdot\|$ 是 n 阶矩阵的任意一种范数, 由第 4 章定理 4.2.5 知必存在 C^n 上与之相容的向量范数 $\|x\|_a = \|x\|_1$, x 是 C^n 中固定非零向量。

设 λ 是 A 的任意一个特征值, x 是对应的 A 的特征向量, 那么

$$Ax = \lambda x$$

则 $\|Ax\|_a = \|\lambda x\|_a = |\lambda| \|x\|_a$

所以

$$\|Ax\|_a = |\lambda| \|x\|_a$$

当然 $\|Ax\|_a \leq \|A\| \|x\|_a$

推论 8.3.1 设 A 是 $n \times n$ 矩阵, $A = (a_{ij})$ 则

$$(1) \|A\|_1 = \max_{j=1,2,\dots,n} \sum_{i=1}^n |a_{ij}| \tag{8.9}$$

$$(2) \|A\|_\infty = \max_{i=1,2,\dots,n} \sum_{j=1}^n |a_{ij}| \tag{8.10}$$

$$(3) \|A\| = \min\{\|A\|_1, \|A\|_\infty\} \tag{8.11}$$

推论 8.3.2 设 A 是 $n \times n$ 矩阵, $A = (a_{ij})$, b_1, b_2, \dots, b_n 是 n 个正实数, 记

$$r_1 = \max_{i=1,2,\dots,n} \sum_{j=1}^n |a_{ij}| \frac{b_j}{b_i}$$

$$r_2 = \max_{j=1,2,\dots,n} \sum_{i=1}^n |a_{ij}| \frac{b_j}{b_i}$$

则 $\|A\| = \min\{r_1, r_2\}$

证明 做矩阵 B

$$B = \text{diag}(b_1, b_2, \dots, b_n)$$

则 B 为可逆矩阵, 且

$$B^{-1} = \text{diag}(b_1^{-1}, b_2^{-1}, \dots, b_n^{-1})$$

从而有

$$B^{-1}AB = \begin{bmatrix} a_{11} & a_{12} \frac{b_2}{b_1} & \dots & a_{1n} \frac{b_n}{b_1} \\ a_{21} \frac{b_1}{b_2} & a_{22} & \dots & a_{2n} \frac{b_n}{b_2} \\ \dots & \dots & \dots & \dots \\ a_{n1} \frac{b_1}{b_n} & a_{n2} \frac{b_2}{b_n} & \dots & a_{nn} \end{bmatrix}$$

因为相似矩阵有相同的特征值, 从而上述矩阵与 A 有相同的谱半径, 而

$$\|B^{-1}AB\|_1 = r_1, \|B^{-1}AB\|_\infty = r_2$$

由推论 8.3.1 即得

$$\|A\| = \|B^{-1}AB\| = \min\{r_1, r_2\}$$

定理 8.3.2 设 $A = (a_{ij})$ 是 $n \times n$ 的行(列)对角占优矩阵, 则

$$(A) = \max_i \{2|a_{ii}|, i = 1, 2, \dots, n\}$$

证明 因为 A 是行对角占优, 则

$$\begin{aligned} |a_{ii}| &> \sum_{j=1, j \neq i}^n |a_{ij}| \quad i = 1, 2, \dots, n \\ A &= \max_i \sum_{j=1}^n |a_{ij}|, \quad i = 1, 2, \dots, n \\ &= \max_i \{ |a_{ii}| + \sum_{j=1, j \neq i}^n |a_{ij}|, i = 1, 2, \dots, n \} \\ &= \max_i \{2|a_{ii}|, i = 1, 2, \dots, n\} \end{aligned}$$

对于应用广泛的正规矩阵谱半径的估计, 可以得到更精确的结果。

定理 8.3.3 设 A 是 $n \times n$ 的正规矩阵, 则

$$(A) = A_2 \tag{8.12}$$

证明 设正规矩阵 A 的 n 个特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 且其模满足不等式

$$|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$$

因此 $(A) = |\lambda_1|$ 。

又因 A 是正规矩阵, 故必存在酉矩阵 P , 使

$$P^H A P = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

及

$$P^H A^H P = \text{diag}(\overline{\lambda_1}, \overline{\lambda_2}, \dots, \overline{\lambda_n})$$

从而有

$$P^H A^H A P = \text{diag}(\overline{\lambda_1} \lambda_1, \overline{\lambda_2} \lambda_2, \dots, \overline{\lambda_n} \lambda_n)$$

所以 $A^H A$ 的最大特征值为 $\overline{\lambda_1} \lambda_1 = |\lambda_1|^2$, 根据矩阵算子范数的求法, 有

$$A_2 = \sqrt{|\lambda_1|^2} = |\lambda_1| = (A)$$

推论 8.3.3 若 $A = A^H$, 则 $(A) = A_2$ (8.13)

推论 8.3.4 若 $A^{-1} = A^H$, 则 $(A) = 1$ (8.14)

这时因为 $P^H A^H A P = P^H E P = E$ 。

推论 8.3.5 设 $A \in C^{n \times n}$, U 为 $n \times n$ 酉矩阵, 则

$$(AU) = A_2 \tag{8.15}$$

8.4* 特征值的摄动

8.4.1 对角矩阵摄动特征值估计

设 $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, 摄动矩阵 $E = (e_{ij}) \in C^{n \times n}$, 根据圆盘定理 1, 摄动后的矩阵 $D + E$ 的所有特征值包含在如下的 n 个圆盘中

$$D_i: |z - \lambda_i - \sum_{j=1}^n |a_{ij}|| \quad i = 1, 2, \dots, n$$

而上述圆盘又包括在如下圆盘中

$$G_i: |z - \lambda_i| \leq \sum_{j=1}^n |a_{ij}| \quad i = 1, \dots, n$$

所以对应矩阵 $D + \Delta$ 的特征值 λ , 存在 D 的某个特征值 λ_i , 使

$$|\lambda - \lambda_i| \leq \sum_{j=1}^n |a_{ij}| \tag{8.16}$$

对角矩阵的特征值的摄动的研究是很简单的。但这种估计不能直接推广到一般的矩阵上, 不过可以将它们推广到可以对角化矩阵上。

8.4.2 可对角化矩阵摄动特征值估计

定理 8.4.1 设 A 是 n 阶可对角化矩阵, 且 $P^{-1}AP = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ 。令 $\Delta = (a_{ij})_{n \times n} \in C^{n \times n}$, 若 λ 是 $A + \Delta$ 的特征值, 则存在 A 的某个特征值 λ_i , 使

$$|\lambda - \lambda_i| \leq \|P\| \|P^{-1}\| \|\Delta\| = K(P) \|\Delta\|$$

其中 $K(P) = \|P\| \|P^{-1}\|$ 是矩阵 P 关于矩阵范数 $\|\cdot\|$ 的条件数。

证明 因为 $A + \Delta$ 与 $P^{-1}(A + \Delta)P = \Lambda + P^{-1}\Delta P$ 有相同的特征值, 其中 $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n) = P^{-1}AP$, 由前面的讨论知道, 对于 $A + \Delta$ 的特征值 λ , 存在 A 的某个特征值 λ_i , 使

$$\begin{aligned} |\lambda - \lambda_i| &\leq \|P^{-1}\Delta P\| \\ &\leq \|P^{-1}\| \|\Delta\| \|P\| \\ &= K(P) \|\Delta\| \end{aligned} \tag{8.17}$$

定理 8.4.2 设 A 是 n 阶可对角化矩阵, 即有 $P^{-1}AP = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, $\Delta = (a_{ij}) \in C^{n \times n}$ 。 $\|\cdot\|$ 是这样的矩阵范数, 对任一个对角阵 $D = \text{diag}(d_1, d_2, \dots, d_n)$ 皆有 $\|D\| = \max_i |d_i|$ 。若 λ 是 $A + \Delta$ 的特征值, 则存在 A 的某个特征值 λ_i , 使

$$\min_i |\lambda - \lambda_i| \leq K(P) \|\Delta\| \tag{8.18}$$

其中 $K(\cdot)$ 是关于矩阵范数 $\|\cdot\|$ 的条件数。

证明 因为 $A + \Delta$ 与 $\Lambda + P^{-1}\Delta P$ 有相同的特征值, 故可以考虑后一矩阵。若 λ 是 $\Lambda + P^{-1}\Delta P$ 的特征值, 则 $(E - \lambda E - P^{-1}\Delta P)$ 奇异, 若 $(E - \lambda E)$ 奇异则 A 有某个特征值 λ_i , 使 $\lambda = \lambda_i$, (8.18) 式成立; 若 $(E - \lambda E)$ 非奇异, 那么矩阵

$$(E - \lambda E)^{-1}(E - \lambda E - P^{-1}\Delta P) = E - (E - \lambda E)^{-1}P^{-1}\Delta P$$

是奇异的, 则 1 是 $(E - \lambda E)^{-1}P^{-1}\Delta P$ 的一个特征值, 可推出 $\|(E - \lambda E)^{-1}P^{-1}\Delta P\| \geq 1$, 又因为

$$\|(E - \lambda E)^{-1}P^{-1}\Delta P\| \leq \|(E - \lambda E)^{-1}\| \|P^{-1}\| \|\Delta\| \|P\|$$

所以有

$$1 \leq \max_i \{|\lambda - \lambda_i|^{-1}\} \|\Delta\| K(P)$$

即

$$\min_i | \lambda_i - \lambda_i^0 | \leq K(P)$$

矩阵范数 $\|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_\infty$ 都属于定理 8.4.2 所规定的矩阵范数的例子, 读者可以找出不符合定理所规定的矩阵范数例子。

在实际应用中, 矩阵 $A = (a_{ij})_{n \times n}$ 内的数 a_{ij} 往往带有误差, 记该误差为 δ_{ij} , 称 $E = (\delta_{ij})_{n \times n}$ 为矩阵 A 的摄动矩阵。 E 的存在对计算 A 的特征值产生了影响, 其影响的大小由公式

$$\frac{|\lambda_i - \lambda_i^0|}{|\lambda_i|} \leq K(P) \tag{8.19}$$

估计。不等式的左边表示因摄动矩阵 E 而产生的特征值的相对误差, 条件数 $K(P)$ 是该相对误差的上界, 其中 P, λ_i 如定理 8.4.2 所示含义。

推论 8.4.1 设 $A \in C^{n \times n}$ 是正规矩阵, $\lambda_1, \lambda_2, \dots, \lambda_n$ 是 A 的特征值, $\lambda \in C^{n \times n}$, 若 λ 是 $A + E$ 的特征值, 则存在 A 的某个特征值 λ_i 使

$$|\lambda - \lambda_i| \leq \|E\|_2 \tag{8.20}$$

证明 因为 A 是正规矩阵, 故 A 可以酉相似于对角阵, 即 $U^H A U = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, 而

$$K_2(U) = U^H U = I = 1$$

所以存在 A 的某个特征值 λ_i , 使

$$|\lambda - \lambda_i| \leq \|E\|_2$$

值得指出的是, 这里并不要求摄动矩阵 E 和摄动后的矩阵 $A + E$ 是正规的。

8.4.3 不可对角化矩阵摄动特征值估计

下面讨论 A 不是可对角化矩阵的情形。这时还不知道是否能得到像(8.18)式那样简单的不等式, 但可导出单特征值受摄动影响的结果。

定义 8.1 设 λ 是矩阵 A 的特征值, 如果 $U^H A = \lambda U^H, U \neq 0$, 则称 U 是 A 的对应于 λ 的左特征向量; 而 $A V = \lambda V, V \neq 0$ 所确定的特征向量 V 称为右特征向量。

引理 设 $A \in C^{n \times n}$, λ 是 A 的单特征值(即一重特征根)。 U 和 $V \in C^n$ 分别是相应于 λ 的右特征向量与左特征向量, 则 $U^H V = 1$ 。

证明 设 λ 是 A 的单特征值, 那么存在酉阵 U , 使

$$U^H A U = \begin{bmatrix} \lambda & * \\ 0 & B \end{bmatrix}$$

其中 $B \in C^{(n-1) \times (n-1)}$, 当然 λ 不是 B 的特征值。因而

$$U^H A U e_1 = \begin{bmatrix} \lambda & * \\ 0 & B \end{bmatrix} e_1 = \lambda e_1 = \begin{bmatrix} \lambda \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

其中 $e_1 = (1 \ 0 \ \dots \ 0)^T$, 说明 e_1 是 $U^H A U$ 相应于 λ 的特征向量, 又设 V 是矩阵

$$(U^H A U)^H = U^H A^H U = \begin{bmatrix} 0 \\ * & B^H \end{bmatrix}$$

对应于特征值 λ 的特征向量, 即 $U^H A^H U = \lambda$, 则 λ 的第一个分量不为零。事实上, 假设 λ 的

第一个分量为零, 即 $\lambda = \begin{bmatrix} 0 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix}$, 那么 $\begin{bmatrix} \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix}$ 是 B^H 的对应于 λ 的特征向量, 导致 λ 是 B 的特征值。

这与 λ 为 A 的单特征值的条件矛盾。因此 $\lambda_1 \neq 0, (U_1)^H U_1 = \lambda_1 \neq 0$ 。 U 和 U_1 分别是 A 的对应于 λ 的左特征向量和右特征向量。由于 λ 是单特征值, 所以 λ 的左特征子空间和右特征子空间都是一维的, 于是 λ 的左、右特征向量可分别表示为 $U = k_1 U_1, U_1 = k_2 U$ 其中 k_1, k_2 是任意非零常数。因而推出

$$\lambda_1 = 0$$

注意, 引理中关于“ λ 是单特征值 ”的条件是必要的。例如

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

有二重特征值 $\lambda = 1, U = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, U_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 分别是 A 的关于 $\lambda = 1$ 的左、右特征向量, 但 $U^H U_1 = 0$ 。

设 λ 是 A 的单特征值, U 是对应于 λ 的右单位特征向量, $U^H U = 1$; U_1 是对应于 λ 的左特征向量, 使 $U_1^H U_1 = 1$ 。又设含参变量 t 的函数矩阵 $A(t) (A(0) = A)$ 在 $t=0$ 及其邻域内可微。则 $A(t)$ 的特征值是 $\lambda(t)$, 且 $\lambda(0) = \lambda$, 其对应于 $\lambda(t)$ 的右单位特征向量是 $U(t), U^H(t) U(t) = 1$; 对应于 $\lambda(t)$ 的左特征向量是 $U_1(t), U_1^H(t) U_1(t) = 1$ 。显然, 在 $t=0$ 及其邻域内 $U(t)$ 连续可微, $U(0) = U, U_1(0) = U_1$ 。

将恒等式 $U^H(t) U(t) = 1$ 两边对 t 求导, 得到

$$\dot{U}^H(t) U(t) + U^H(t) \dot{U}(t) = 0$$

因 $A(t) U(t) = \lambda(t) U(t)$, 故 $U^H(t) A(t) U(t) = \lambda(t) U^H(t) U(t) = \lambda(t)$, 再将此式两端对 t 求导, 使得

$$\begin{aligned} \dot{\lambda}(t) &= \dot{U}^H(t) A(t) U(t) + U^H(t) \dot{A}(t) U(t) + U^H(t) A(t) \dot{U}(t) \\ &= \lambda(t) [\dot{U}^H(t) U(t) + U^H(t) \dot{U}(t)] + U^H(t) \dot{A}(t) U(t) \\ &= U^H(t) \dot{A}(t) U(t) \end{aligned}$$

令 $t=0$, 得到

$$\dot{\lambda}(0) = U^H(0) \dot{A}(0) U(0) = U^H \dot{A}(0) \tag{8.21}$$

其中 $U^H U = 1, U_1^H U_1 = 1$ 。

设 U 和 U_1 分别是对应于 λ 的右特征向量和左特征向量(它们不一定是单位向量), 则在 (8.21) 式中, 以 $U / \sqrt{U^H U}$ 代替 U , 以 $\sqrt{U_1^H U_1} U_1 / U_1^H$ 代替 U_1^H , 便得到等式

$$\dot{\lambda}(0) = U^H \dot{A}(0) U$$

于是可得下述定理。

定理 8.4.3 设含参变量 t 的函数矩阵 $A(t)$ 在 $t=0$ 及其邻域内可微, $A(0) = A$, 且 λ 是 A 的单特征值, x 和 y 分别是 A 的对应于 λ 的右特征向量和左特征向量。假设对充分小的 t , $\lambda(t)$ 是 $A(t)$ 的特征值, $x(0) = x$, 则

$$\lambda'(0) = \frac{y^H \dot{A}(0) x}{y^H x} \tag{8.22}$$

(8.22) 式表示 A 的单特征值 λ 相对小参数 t 的变化率。

例如, 设 λ 是 $A \in C^{n \times n}$ 的单特征值, $A(t) = A + tE$, t 是 A 的小摄动, 那么在 $t=0$ 处有

$$\frac{d\lambda}{dt} = \frac{y^H E x}{y^H x}$$

又如要考虑 $A = (a_{ij})$ 的单特征值相对 a_{ij} 元的变化率, 可以令 $A(t) = A + tE_{ij}$, 其中 E_{ij} 是 i 行 j 列处元素为 1, 其余全为 0 的方阵。那么在 $t=0$ 处有

$$\frac{d\lambda}{da_{ij}} = \frac{y_i x_j}{y^H x} \quad 1 \leq i, j \leq n$$

其中 x_j 是 x 的第 j 个分量 \bar{y}_i 是 y 的第 i 个分量的共轭复数。

8.5* 广义特征值

8.5.1 广义特征值, 特征向量定义

在力学、物理学及系统工程理论中, 常会遇到广义特征值问题。

定义 8.2 设 A, B 都 $n \times n$ 阶矩阵, 如果对于一个复数 λ , 存在非 0(列) 向量 x , 使

$$(A - \lambda B)x = 0 \tag{8.23}$$

则称复数 λ 为矩阵 A 相对于矩阵 B 的特征值, 或称 λ 为 A 与 B 确定的广义特征值, 非零向量 x 称为与 λ 相应的广义特征向量。

通常形如(8.23)式的特征值问题称为矩阵 A 相对于矩阵 B 的广义特征值问题。显然在 $B = E$ 时, (8.23) 式就化为矩阵 A 的一般特征值问题。因此广义特征值问题是一般特征值问题的推广。

一般矩阵 A 和 B 的广义特征值问题, 情况比较复杂。如果 B 是满秩阵时, (8.23) 式就可化为

$$B^{-1}A x = \lambda x \tag{8.24}$$

这样就把广义特征值问题(8.23)式, 化为矩阵 $B^{-1}A$ 的一般特征值问题解决。

8.5.2 Hermite 矩阵的广义特征向量的性质

在许多科技问题中, A, B 都是 Hermite 矩阵即 $A^H = A, B^H = B$, 且 B 正定。如果用(8.24)式讨论特征值问题, 虽然, A, B, B^{-1} 都是 Hermite 矩阵, 但 $B^{-1}A$ 一般不是 Hermite 矩阵, 这样就不能直接利用 Hermite 矩阵特征值问题的结论。通常采取以下方法处理。

由于 B 正定, 所以存在满秩阵 P , 使

$$B = P^H P$$

于是(8.23) 式化为

$$A = P^H P \tag{8.25}$$

若记 $\alpha = P$ 则 $P^{-1} = \alpha^{-1}$ 代入(8.25) 式得

$$\begin{aligned} A P^{-1} &= P^H \\ (P^{-1})^H A P^{-1} &= \end{aligned}$$

若记 $T = (P^{-1})^H A P^{-1}$, 就有

$$T = \tag{8.26}$$

显然 $T^H = T$, 即 T 为 Hermite 矩阵, 从而广义特征值问题(8.23) 式就化为等价的 Hermite 矩阵 T 的一般特征值问题(8.26) 式。即求数 λ , 非零向量 α , 满足 $T \alpha = \lambda \alpha$ 。由于 T 是 Hermite 矩阵, 所以广义特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 都是实数, 并且存在由 n 个单位特征向量构成的标准正交向量系 $\alpha_1, \alpha_2, \dots, \alpha_n$, 有

$$\alpha_i^H \alpha_j = \alpha_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

由于 $\alpha_i = P \beta_i$, 从而有

$$\alpha_i^H \alpha_j = (P \beta_i)^H (P \beta_j) = \beta_i^H B \beta_j$$

所以

$$\beta_i^H B \beta_j = \alpha_{ij} \quad (i = 1, 2, \dots, n)$$

这时称 $\alpha_1, \alpha_2, \dots, \alpha_n$ 为 B 共轭向量系。

定理 8.5.1 设 n 阶方阵 $A^H = A, B^H = B$, 且 B 是正定的, 则 B 共轭向量系 $\alpha_1, \alpha_2, \dots, \alpha_n$ 具有以下性质:

- (1) $\alpha_i \neq 0 \quad (i = 1, 2, \dots, n)$;
- (2) $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关;
- (3) α_i 与 α_i 满足方程

$$A \alpha_i = \lambda_i B \alpha_i \quad (i = 1, 2, \dots, n)$$

- (4) 若记 $Q = (\alpha_1, \alpha_2, \dots, \alpha_n)$, 则

$$Q^H B Q = E, Q^H A Q = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

证明 (1) 设 α_i 是 Hermite 矩阵 $T = (P^{-1})^H A P^{-1}$ 的特征值 λ_i 所对应的特征向量, 那么由 $\lambda_i \neq 0$ 及 $\alpha_i = P \beta_i$ (其中 P 可逆), 可知

$$\beta_i = P^{-1} \alpha_i \neq 0 \quad (i = 1, 2, \dots, n)$$

- (2) 设有 $k_1, k_2, \dots, k_n \in C$, 使得 B 共轭向量系 $\alpha_1, \alpha_2, \dots, \alpha_n$ 的线性组合等于零向量, 即

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

用 $\alpha_i^H B$ 左乘等式两端, 得

$$k_1 \alpha_i^H B \alpha_1 + k_2 \alpha_i^H B \alpha_2 + \dots + k_i \alpha_i^H B \alpha_i + \dots + k_n \alpha_i^H B \alpha_n = 0$$

注意到

$$\alpha_i^H B \alpha_j = \alpha_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

从而有

$$k_i \mathbf{B}_i^H \mathbf{B}_i = k_i = 0 \quad (i = 1, 2, \dots, n)$$

因此 $\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_n$ 线性无关。

(3) 设有满秩矩阵 P , 使 $B = P^H P$, 而 $T = (P^{-1})^H A P^{-1}$ 。 λ_i 是 T 的一个特征值, 对应的特征向量为 \mathbf{v}_i 即

$$T \mathbf{v}_i = \lambda_i \mathbf{v}_i$$

从而有

$$\begin{aligned} (P^{-1})^H A P^{-1} \mathbf{v}_i &= \lambda_i \mathbf{v}_i \\ A P^{-1} \mathbf{v}_i &= \lambda_i P^H \mathbf{v}_i \\ A P^{-1} \mathbf{v}_i &= \lambda_i P^H P P^{-1} \mathbf{v}_i \end{aligned}$$

由 $\mathbf{v}_i = P^{-1} \mathbf{u}_i$ 及 $B = P^H P$, 得

$$A \mathbf{u}_i = \lambda_i B \mathbf{u}_i \quad (i = 1, 2, \dots, n)$$

所以 λ_i 是广义特征值问题(8.23)式的广义特征值, 而 $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ 是对应的广义特征向量, 且 $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ 线性无关。

(4) 设 $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ 是 B 共轭向量系, 由于它们线性无关, 则矩阵

$$Q = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n)$$

是满秩的 $n \times n$ 矩阵, 又

$$\mathbf{B}_i^H \mathbf{B}_j = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad i, j = 1, 2, \dots, n$$

故

$$\begin{aligned} Q^H B Q &= \begin{bmatrix} \mathbf{B}_1^H \\ \mathbf{B}_2^H \\ \vdots \\ \mathbf{B}_n^H \end{bmatrix} B (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) \\ &= \begin{bmatrix} \mathbf{B}_1^H B \mathbf{u}_1 & \mathbf{B}_1^H B \mathbf{u}_2 & \dots & \mathbf{B}_1^H B \mathbf{u}_n \\ \mathbf{B}_2^H B \mathbf{u}_1 & \mathbf{B}_2^H B \mathbf{u}_2 & \dots & \mathbf{B}_2^H B \mathbf{u}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_n^H B \mathbf{u}_1 & \mathbf{B}_n^H B \mathbf{u}_2 & \dots & \mathbf{B}_n^H B \mathbf{u}_n \end{bmatrix} \\ &= \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & W & \\ 0 & & & 1 \end{bmatrix} \\ Q^H A Q &= \begin{bmatrix} \mathbf{B}_1^H \\ \mathbf{B}_2^H \\ \vdots \\ \mathbf{B}_n^H \end{bmatrix} (A \mathbf{u}_1, A \mathbf{u}_2, \dots, A \mathbf{u}_n) \end{aligned}$$

$$\begin{aligned} &= \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_n \end{bmatrix} (B_1, B_2, \dots, B_n) \\ &= \begin{bmatrix} H_1 B_1 & H_2 B_1 & \dots & H_n B_1 \\ H_1 B_2 & H_2 B_2 & \dots & H_n B_2 \\ \vdots & \vdots & \ddots & \vdots \\ H_1 B_n & H_2 B_n & \dots & H_n B_n \end{bmatrix} \\ &= \begin{bmatrix} 1 & & & 0 \\ & 2 & & \\ & & W & \\ 0 & & & n \end{bmatrix} \end{aligned}$$

习 题 8

- 1. 证明: 实对称矩阵 A 的所有特征值在区间 $[a, b]$ 上的充分必要条件是对任何 $\epsilon_0 < a$, $A - \epsilon_0 E$ 是正定矩阵; 而对任何 $\epsilon_0 > b$, $A - \epsilon_0 E$ 是负定矩阵。
- 2. 设 A, B 都是实对称矩阵, A 的一切特征值在区间 $[a, b]$ 上, B 的一切特征值在区间 $[c, d]$ 上, 试利用 1 题的结论证明: $A + B$ 的特征值必在区间 $[a + c, b + d]$ 上。
- 3. 设 P 是酉矩阵, $A = \text{diag}(a_1, a_2, \dots, a_n)$, 试证明: PA 的特征值 μ 恒满足不等式

$$m \leq |\mu| \leq M$$

其中 $m = \min_i |a_i|, i = 1, 2, \dots, n, M = \max_i |a_i|, i = 1, 2, \dots, n$ 。

(提示: 记属于 μ 的特征向量 $x \neq 0$, 即 $PAx = \mu x$, 注意到 $|\mu|^2 = x^H P^H P A^H A x = x^H A^H A x$, 可推得 $m \leq |\mu| \leq M$)。

- 4. 用圆盘定理估计矩阵

$$A = \begin{bmatrix} 1 & 0.1 & 0.2 & 0.3 \\ 0.5 & 3 & 0.1 & 0.2 \\ 1 & 0.3 & -1 & 0.5 \\ 0.2 & -0.3 & -0.1 & -4 \end{bmatrix}$$

的特征值的分布范围, 并在复平面作出示意图。

- 5. 用圆盘定理估计矩阵

$$\begin{bmatrix} i & 0.03 & 0.4 \\ 0.02 & 0 & 1 \\ 0.001 & 0.05 & 2 \end{bmatrix}$$

的特征值的分布范围。然后适当选择一组正实数 b_1, b_2, b_3 , 利用推论 8.3.1 对 A 的特征值作更精确的估计(要求 A 的三个圆盘互不相交)。

6. 以 2×2 矩阵为例, 举例说明: 圆盘定理中两个圆盘构成连通部分, 可以在每个圆中各有一个特征值。

7. 证明矩阵

$$A = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{6} & \frac{1}{6} & \frac{3}{6} & \frac{1}{6} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{3}{7} \end{bmatrix}$$

的谱半径 $\rho(A) < 1$ 。

8. 证明矩阵

$$B = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{6} & \frac{1}{6} & \frac{3}{6} & \frac{1}{6} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{4}{7} \end{bmatrix}$$

的谱半径 $\rho(B) = 1$ 。

9. 设 λ 和 μ 是 A 的不同特征值, x 是对应于 λ 的右特征向量, y 是对应于 μ 的左特征向量, 证明: x 和 y 是正交的。

10. 设 A 是任意 n 级方阵, 证明: $(A^H A) = (A A^H) = A^H A$

第 9 章

矩阵的 Kronecker 积

矩阵的 Kronecker 积, 又称为矩阵的直积或张量积, 最初起源于群论, 物理上用来研究粒子理论。现在它已成功地应用到矩阵论的各个领域中。本章将介绍 Kronecker 积的基本性质, 并用它来求解线性矩阵方程和微分方程。

9.1 Kronecker 积的基本性质

9.1.1 Kronecker 积定义

定义 9.1 设 $A = (a_{ij}) \in C^{m \times n}$, $B = (b_{ij}) \in C^{r \times s}$, A 和 B 的 kronecker 积定义为

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \dots & \dots & & \dots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{bmatrix} \tag{9.1}$$

以后把 Kronecker 积简称为 K 积。 $A \otimes B$ 是 $mr \times ns$ 矩阵, 共有 mn 个子块, 第 (i, j) 位置上是 $r \times s$ 型的子矩阵块 $a_{ij}B$, K 积 $A \otimes B$ 对任何类型的矩阵 A 与 B 都可以定义, 仅就此点看, 它就好比一般的矩阵乘积要优越。

例如, 设

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

则

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{bmatrix}$$

$$= \left[\begin{array}{cc|cc} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ \hline a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{array} \right]$$

显然 K 积的交换律不成立,从 K 积的定义可得到下列结果:

- (1) $E \otimes A = \text{diag}(A \otimes A \dots A)$
- (2) $A \otimes E = \begin{bmatrix} a_{11}E & a_{12}E & \dots & a_{1n}E \\ a_{21}E & a_{22}E & \dots & a_{2n}E \\ \dots & \dots & \dots & \dots \\ a_{m1}E & a_{m2}E & \dots & a_{mn}E \end{bmatrix}$
- (3) $E_{mn} = E_m \otimes E_n = E_n \otimes E_m$
- (4) $O_{mn} = O_m \otimes O_n, O_{mn}$ 是 mn 阶零矩阵

9.1.2 Kronecker 积的性质

矩阵乘法的某些性质对 K 积同样成立,但有些性质有变动。K 积有许多好的性质是通常的矩阵乘积所没有的,正因为如此,K 积在应用上有它的独到之处。

定理 9.1.1 设 μ 是任一常数,则

$$(\mu A) \otimes B = A \otimes (\mu B) = \mu(A \otimes B) \tag{9.2}$$

证明 $A \otimes (\mu B)$ 的第 (i, j) 位置上的矩阵子块是 $a_{ij}(\mu B) = \mu(a_{ij}B)$, 它正是 $\mu(A \otimes B)$ 及 $\mu A \otimes B$ 的第 (i, j) 位置上的矩阵子块。

定理 9.1.2 K 积对加法是可分配的,即

$$(A + B) \otimes C = A \otimes C + B \otimes C \tag{9.3}$$

$$A \otimes (B + C) = A \otimes B + A \otimes C \tag{9.4}$$

证明: 只证(9.3) 式

$(A + B) \otimes C$ 的第 (i, j) 位置上的矩阵子块是

$$(a_{ij} + b_{ij})C$$

而矩阵 $A \otimes C + B \otimes C$ 的第 (i, j) 位置上是 $a_{ij}C + b_{ij}C, (a_{ij} + b_{ij})C = a_{ij}C + b_{ij}C$ 。

定理 9.1.3 K 积有结合律,即

$$A \otimes (B \otimes C) = (A \otimes B) \otimes C \tag{9.5}$$

留给读者证明。

定理 9.1.4 K 积的转置规则是

$$(A \otimes B)^T = A^T \otimes B^T \tag{9.6}$$

证明 $(A \otimes B)^T$ 的第 (i, j) 位置上的矩阵子块是 $a_{ji}B^T$, 它与 $A^T \otimes B^T$ 的第 (i, j) 位置上的矩阵子块相等。

定理 9.1.5 K 积混合积的规则是

$$(A \otimes B)(C \otimes D) = AC \otimes BD \tag{9.7}$$

证明 (9.7) 式的左端第 (i, j) 位置的子块是 $A \otimes B$ 的第 i 行子块与 $C \otimes D$ 的第 j 列子块对

应相乘之和 $\sum_k a_{ik}c_{kj}BD$ 。而(9.7) 式的右端第 (i,j) 位置的子块是 AC 的第 (i,j) 位置元素与 BD 之积, 即 $\sum_k a_{ik}c_{kj}BD$, , 故(9.7) 式成立。

例 1 设有两个线性变换

其中 $A \in C^{m \times n}, B \in C^{r \times s}$, 令 $\mu = \begin{pmatrix} A & B \end{pmatrix}, \nu = \begin{pmatrix} A \\ B \end{pmatrix}$, 求 μ 与 ν 之间的线性变换。

解 应用定理 9.1.5, 得

$$\begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix}^{-1}$$

于是 μ 与 ν 之间的线性变换是

$$\mu = \begin{pmatrix} A & B \end{pmatrix} \nu^{-1}$$

定理 9.1.6 K 积的求逆规则是, 设 $A \in C^{m \times m}, B \in C^{n \times n}$ 可逆, 则

$$\begin{pmatrix} A & B \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & B^{-1} \end{pmatrix} \tag{9.8}$$

证明 由(9.7) 式, 得

$$\begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} A^{-1} & B^{-1} \end{pmatrix} = AA^{-1} \quad BB^{-1} = E_{mn}$$

定理 9.1.7 设 $f(z)$ 是任意一个解析函数, $A \in C^{n \times n}$, 则

$$f \begin{pmatrix} E_m & A \end{pmatrix} = E_m \quad f(A) \tag{9.9}$$

$$f \begin{pmatrix} A & E_m \end{pmatrix} = f(A) \quad E_m \tag{9.10}$$

证明 因为 $f(z)$ 解析, 故它可表示为收敛的幂级数: $f(A) = \sum_{k=0}^{\infty} a_k A^k$, 利用 Cayley-Hamilton 定理将 $f(A)$ 化为 $n-1$ 次矩阵多项式

$$f(A) = \sum_{k=0}^{n-1} c_k A^k$$

则得

$$\begin{aligned} f \begin{pmatrix} E_m & A \end{pmatrix} &= \sum_{k=0}^{n-1} c_k \begin{pmatrix} E_m & A \end{pmatrix}^k \\ &= \sum_{k=0}^{n-1} E_m \quad c_k A^k \\ &= E_m \sum_{k=0}^{n-1} c_k A^k = E_m \quad f(A) \end{aligned}$$

类似证明(9.10) 式。

上述定理的一个重要应用是用于指数函数 $f(z) = e^z$ 上, 得

$$e^{\begin{pmatrix} E_m & A \end{pmatrix}} = E_m \quad e^A \tag{9.11}$$

$$e^{\begin{pmatrix} A & E_m \end{pmatrix}} = e^A \quad E_m \tag{9.12}$$

定理 9.1.8 设 $A \in C^{m \times m}, B \in C^{n \times n}$, 则

$$T_r \begin{pmatrix} A & B \end{pmatrix} = T_r(A) T_r(B) \tag{9.13}$$

证明 $T_r \begin{pmatrix} A & B \end{pmatrix} = T_r(a_{11}B) + T_r(a_{22}B) + \dots + T_r(a_{mm}B)$
 $= a_{11} T_r B + a_{22} T_r B + \dots + a_{mm} T_r B$
 $= T_r(A) T_r(B)$

对于一般的矩阵乘积

$$T_r(AB) = T_r(A) T_r(B)$$

9.2 Kronecker 积的特征值

9.2.1 介绍 Vec 算符

设 $m \times n$ 型矩阵

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

记 A 的列向量为 $A_{c1}, A_{c2}, \dots, A_{cn}$, 记 A 的行向量的转置所成的列向量为 $A_{1r}, A_{2r}, \dots, A_{mr}$, 即

$$\begin{aligned} A_{cj} &= \begin{bmatrix} a_{1j} \\ a_{2j} \\ \cdots \\ a_{mj} \end{bmatrix} \quad j = 1, 2, \dots, n \\ A_{ir} &= \begin{bmatrix} a_{i1} \\ a_{i2} \\ \cdots \\ a_{in} \end{bmatrix} \quad i = 1, 2, \dots, m \end{aligned} \tag{9.14}$$

于是有

$$A = (A_{c1}, A_{c2}, \dots, A_{cn}) = (A_{1r}, A_{2r}, \dots, A_{mr})^T \tag{9.15}$$

现在引进一个算符—— V_{ec} 算符, 称它为向量化算符, 把它和 K 积配合起来很有用。

定义 9.2 设 $m \times n$ 型矩阵 $A = (A_{c1}, A_{c2}, \dots, A_{cn})$, V_{ec} 算符定义为

$$V_{ec}A = \begin{bmatrix} A_{c1} \\ A_{c2} \\ \cdots \\ A_{cn} \end{bmatrix} \tag{9.16}$$

从定义看, $V_{ec}A$ 是 mn 维列向量。 V_{ec} 算符作用在矩阵 A 上, 就是把 A 的列向量按照在 A 中排列的次序排成列向量。 V_{ec} 算符是线性的, 即对任意常 a, b , 有

$$V_{ec}(aA + bB) = aV_{ec}A + bV_{ec}B \tag{9.17}$$

再则, 如果矩阵 $A_1, A_2, \dots, A_k \in C^{m \times n}$ 作为线性空间中的向量是线性无关的, 则 $V_{ec}A_1, V_{ec}A_2, \dots, V_{ec}A_k$ 作为 C^{mn} 中的向量也是线性无关的, 反之亦真。下面定理 9.2.1 说明了 V_{ec} 算符和 K 积的密切关系。

9.2.2 Vec 算符与 K 积的关系

定理 9.2.1 设 $A \in C^{m \times n}, D \in C^{n \times r}, B \in C^{r \times s}$, 则

$$V_{ec}(ADB) = (B^T \quad A) V_{ec}D \tag{9.18}$$

证明 现仅就 $A, D, B \in C^{n \times n}$ 的情况证明, 对于一般情况可以类似证明。
对 $j = 1, 2, \dots, n$ 列向量 $(ADB)_{cj}$ 能表示为

$$(ADB)_{cj} = \sum_{k=1}^n (AD)_{ck} b_{kj} = \sum_{k=1}^n (b_{kj}A) D_{ck}$$

因此

$$\begin{aligned} (ADB)_{cj} &= [b_{1j}A, b_{2j}A, \dots, b_{nj}A] V_{ec}D \\ &= [(B_{cj})^T \quad A] V_{ec}D = [(B^T)_{jr} \quad A] V_{ec}D \end{aligned}$$

例 1 已知矩阵方程为

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 & x_3 \\ x_2 & x_4 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

这可写成

$$AXE = C$$

因 $V_{ec}(AXE) = (E \quad A) V_{ec}X = V_{ec}C$

故原方程可化为下述方程

$$\begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ 0 & 0 & a_{11} & a_{12} \\ 0 & 0 & a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} c_{11} \\ c_{21} \\ c_{12} \\ c_{22} \end{bmatrix}$$

- 推论 9.2.1 设 $A \in C^{m \times m}, B \in C^{n \times n}, D \in C^{m \times n}$, 则
- (1) $V_{ec}(AD) = (E_n \quad A) V_{ec}D$ (9.19)
 - (2) $V_{ec}(DB) = (B^T \quad E_m) V_{ec}D$ (9.20)
 - (3) $V_{ec}(AD + DB) = [(E_n \quad A) + (B^T \quad E_m)] V_{ec}D$ (9.21)
- 证明 在(9.18)式中, 令 $B = E_n$ 得(1), 令 $A = E_m$ 得(2), 根据 V_{ec} 的线性性质可得(3)。

9.2.3 K 积 $A \circ B$ 的特征值

设二元多项式

$$f(x, y) = \sum_{i,j=0}^l c_{ij} x^i y^j \tag{9.22}$$

若 $A \in C^{m \times m}, B \in C^{n \times n}$, 那么

$$f(A, B) = \sum_{i,j=0}^l c_{ij} A^i B^j \in C^{mn \times mn} \tag{9.23}$$

例如 $f(x, y) = 2x + xy^3 = 2x^1 y^0 + x^1 y^3$, 则

$$f(A, B) = 2A \quad E_n + A \quad B^3$$

下面定理给出了 A, B 的特征值和矩阵 $f(A, B)$ 的特征值之间的关系。

定理 9.2.2 设 $A \in C^{m \times m}$ 的特征值是 $\lambda_1, \lambda_2, \dots, \lambda_m$, $B \in C^{n \times n}$ 的特征值是 $\mu_1, \mu_2, \dots, \mu_n$, 则 $f(A, B)$ 的特征值是 $f(\lambda_i, \mu_j)$, $i = 1, 2, \dots, m, j = 1, 2, \dots, n$ 。

证明 设 $P^{-1}AP = J_1, Q^{-1}BQ = J_2$, J_1 和 J_2 是 Jordan 标准形。 J_1^i 是以 $\lambda_1^i, \lambda_2^i, \dots, \lambda_m^i$ 为主

对角元的上三角矩阵, J_2^j 是以 $\mu_1^j, \mu_2^j, \dots, \mu_n^j$ 为主对角元的上三角矩阵。从而 $J_1^i J_2^j$ 也是上三角矩阵。容易验证它的主对角元是 $\mu_l^i \mu_h^j, l=1, 2, \dots, m, h=1, 2, \dots, n$ 。因此矩阵 $f(J_1, J_2)$ 是上三角矩阵, 且有主对角元 $f(\mu_l^i, \mu_h^j)$, 从而 $f(\mu_l^i, \mu_h^j)$ 是 $f(J_1, J_2)$ 的特征值, 其中 $l=1, 2, \dots, m; h=1, 2, \dots, n$ 。

剩下的要证明 $f(J_1, J_2)$ 和 $f(A, B)$ 有相同特征值, 由 (9.7) 式

$$\begin{aligned} J_1^i J_2^j &= P^{-1} A^i P Q^{-1} B^j Q \\ &= (P^{-1} Q^{-1})(A^i B^j)(P Q) \\ &= (P Q)^{-1}(A^i B^j)(P Q) \end{aligned}$$

所以 $f(J_1, J_2) = (P Q)^{-1} f(A, B) (P Q)$

这证明了 $f(J_1, J_2)$ 与 $f(A, B)$ 是相似的, 故它们有相同的特征值。

推论 9.2.2 设 μ_1, \dots, μ_m 是 $A \in C^{m \times m}$ 的特征值 $\alpha_1, \dots, \alpha_m$ 是相应的特征向量; μ_1, \dots, μ_n 是 $B \in C^{n \times n}$ 的特征值, β_1, \dots, β_n 是相应的特征向量。则 $A \otimes B$ 的 mn 个特征值是 $\mu_l \mu_h$, 对应的特征向量是 $\alpha_l \otimes \beta_h, l=1, 2, \dots, m, h=1, 2, \dots, n$ 。

证明 因为

$$\begin{aligned} (A \otimes B)(\alpha_l \otimes \beta_h) &= A \alpha_l \otimes B \beta_h \\ &= \mu_l \alpha_l \otimes \mu_h \beta_h \\ &= \mu_l \mu_h \alpha_l \otimes \beta_h \end{aligned}$$

9.2.4 K 积的行列式与秩及拟交换性

定理 9.2.3 (K 积的行列式) 设 $A \in C^{m \times m}, B \in C^{n \times n}$, 则

$$\det(A \otimes B) = (\det A)^n (\det B)^m \tag{9.24}$$

证明 设 $\mu_1, \dots, \mu_m, \nu_1, \dots, \nu_n$ 分别是 A 和 B 的特征值, 则

$$\det A = \mu_1 \dots \mu_m, \det B = \nu_1 \dots \nu_n$$

因此

$$\begin{aligned} \det(A \otimes B) &= \prod_{l=1}^m \prod_{h=1}^n (\mu_l \nu_h) \\ &= \left(\prod_{h=1}^n \nu_h \right) \left(\prod_{h=1}^n \nu_h \right) \dots \left(\prod_{h=1}^n \nu_h \right) \\ &= \prod_{l=1}^m \left(\prod_{h=1}^n \nu_h \right)^m \\ &= (\det A)^n (\det B)^m \end{aligned}$$

定理 9.2.4(K 积的秩) 设 $A \in C^{m \times m}, B \in C^{n \times n}$, 则

$$\text{rank}(A \otimes B) = \text{rank}(A) \text{rank}(B) \tag{9.25}$$

证明 设 $\text{rank}(A) = s, \text{rank}(B) = t$

故存在可逆阵 P, Q, K, H 使

$A = P A_1 Q, B = K B_1 H$, 其中 A_1, B_1 分别为 A, B 的标准型

所以

$$\begin{aligned} A \otimes B &= (P A_1 Q) \otimes (K B_1 H) \\ &= (P \otimes K)(A_1 \otimes B_1)(Q \otimes H) \end{aligned}$$

而 $A_1 \otimes B_1$ 的对角线上 st 个 1, 秩 $= st$

所以 $\text{rank}(A \ B) = \text{rank}(A_1 \ B_1) = st$

K 积与普通矩阵乘法一样, 没有交换律, 但有所谓的“拟交换性”。

定理 9.2.5 (K 积的拟交换性) 设 $A \in C^{m \times n}, B \in C^{r \times s}$, 则存在置换矩阵 $U_1 \in C^{mr \times mr}, U_2 \in C^{ns \times ns}$ 使得

$$A \ B = U_1 (B \ A) U_2^{-1} \tag{9.26}$$

证明 设 $X = (x_{ij}) \in C^{m \times n}, E_{ij} \in C^{m \times n}$, 其中 E_{ij} 的行 i 行 j 列处元素为 1, 其余元素全为 0。于是

$$\begin{aligned} X &= \sum_{i,j} x_{ij} E_{ij} \\ X^T &= \sum_{i,j} x_{ij} E_{ij}^T \\ V_{ec} X^T &= \sum_{i,j} x_{ij} V_{ec} E_{ij}^T \\ &= (V_{ec} E_{11}^T, \dots, V_{ec} E_{m1}^T, \dots, V_{ec} E_{1n}^T, \dots, V_{ec} E_{mn}^T) \begin{bmatrix} X_{11} \\ \vdots \\ X_{m1} \\ \vdots \\ X_{1n} \\ \vdots \\ X_{mn} \end{bmatrix} \\ &= (V_{ec} E_{11}^T, \dots, V_{ec} E_{m1}^T, \dots, V_{ec} E_{1n}^T, \dots, V_{ec} E_{mn}^T) V_{ec} X \end{aligned}$$

联系 $V_{ec} X^T$ 和 $V_{ec} X$ 之间的置换矩阵

$$U = (V_{ec} E_{11}^T, \dots, V_{ec} E_{m1}^T, \dots, V_{ec} E_{1n}^T, \dots, V_{ec} E_{mn}^T) \tag{9.27}$$

即
$$V_{ec} X^T = U V_{ec} X$$

现令 $AXB^T = Y$, 转置后 $BX^T A^T = Y^T$, 从而由定理 9.2.1

$$V_{ec} Y = (B \ A) V_{ec} X \tag{9.28}$$

$$V_{ec} Y^T = (A \ B) V_{ec} X^T \tag{9.29}$$

由前一段证明知, 存在 mr 的阶置换阵 U_1 及 ns 阶置换阵 U_2 , 使

$$V_{ec} Y^T = U_1 V_{ec} Y \tag{9.30}$$

$$V_{ec} X^T = U_2 V_{ec} X \tag{9.31}$$

将(9.30)式与(9.31)式代入(9.28)式, (9.29)式, 得到

$$U_1 (B \ A) V_{ec} X = (A \ B) U_2 V_{ec} X$$

由于 X 是任意矩阵, 所以

$$A \ B = U_1 (B \ A) U_2^{-1}$$

必须指出, 置换矩阵 U_1 和 U_2 只取决于 A 与 B 的行数和列数, U_1 和 U_2 的阶数分别是 $A \ B$ 的行数(mr)和列数(ns)。

例 2 设

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

求 U_1 和 U_2 , 使

$$A \otimes B = U_1 (B \otimes A) U_2^{-1}$$

解 $A \otimes B$ 是 4×6 型矩阵, 故 $U_1 \in C^{4 \times 4}, U_2 \in C^{6 \times 6}$, 由 (9.27) 式, 得到

$$\begin{aligned} U_1 &= (V_{ec} E_{11}^T \quad V_{ec} E_{21}^T \quad V_{ec} E_{12}^T \quad V_{ec} E_{22}^T) \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ U_2 &= (V_{ec} E_{11}^T \quad V_{ec} E_{21}^T \quad V_{ec} E_{12}^T \quad V_{ec} E_{22}^T \quad V_{ec} E_{13}^T \quad V_{ec} E_{23}^T) \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

9.2.5 Kronecker 和

定义 9.3 设 $A \in C^{m \times m}, B \in C^{n \times n}$, A 和 B 的 Kronecker 和定义为

$$A \otimes B = A \otimes E_n + E_m \otimes B \tag{9.32}$$

简称为 K 和。

定理 9.2.6 设 $\lambda_1, \dots, \lambda_m$ 是 $A \in C^{m \times m}$ 的特征值, $\alpha_1, \dots, \alpha_m$ 是对应的特征向量, μ_1, \dots, μ_n 是 $B \in C^{n \times n}$ 的特征值, β_1, \dots, β_n 是对应的特征向量, 则 $\lambda_i + \mu_h$ 是 $A \otimes B$ 的特征值, $\alpha_i \otimes \beta_h$ 是对应的特征向量。 $i=1, 2, \dots, m, h=1, 2, \dots, n$ 。

证明 事实上

$$\begin{aligned} (A \otimes B)(\alpha_i \otimes \beta_h) &= (A \otimes E_n)(\alpha_i \otimes \beta_h) + (E_m \otimes B)(\alpha_i \otimes \beta_h) \\ &= A \alpha_i \otimes \beta_h + \alpha_i \otimes B \beta_h \\ &= \lambda_i \alpha_i \otimes \beta_h + \alpha_i \otimes \mu_h \beta_h \\ &= (\lambda_i + \mu_h) \alpha_i \otimes \beta_h \end{aligned}$$

例 3 设

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}$$

则 A 的特征值是 $\lambda_1 = 1, \lambda_2 = 2$, 相应特征向量是 $\alpha_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \alpha_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, B 的特征值是 $\mu = 1, \mu = -1$, 相应的特征向量是 $\beta_1 = (1, 1)^T, \beta_2 = (0, 1)^T$, 因此 A 与 B 的 K 和为

$$A \otimes B = \begin{bmatrix} 2 & 0 & -1 & 0 \\ 2 & 0 & 0 & -1 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 2 & 1 \end{bmatrix}$$

$A \times B$ 的特征多项式是

$$\det(E - A \times B) = (\lambda - 1)(\lambda - 2)(\lambda - 3)$$

故 $A \times B$ 的特征值与特征向量分别是

$$\begin{aligned} \lambda_1 &= 0 = \lambda_1 + \mu, & \lambda_1 &= 0, & \lambda_1 &= 0, & \lambda_1 &= 0, & \lambda_1 &= 0, \\ \lambda_2 &= 1 = \lambda_2 + \mu, & \lambda_2 &= 1, & \lambda_2 &= 1, & \lambda_2 &= 1, & \lambda_2 &= 1, \\ \lambda_3 &= 2 = \lambda_3 + \mu, & \lambda_3 &= 2, & \lambda_3 &= 2, & \lambda_3 &= 2, & \lambda_3 &= 2, \\ \lambda_4 &= 3 = \lambda_4 + \mu, & \lambda_4 &= 3, & \lambda_4 &= 3, & \lambda_4 &= 3, & \lambda_4 &= 3, \end{aligned}$$

对一般的矩阵和 $A + B$, 当 $AB = BA$ 时, 有

$$e^{A+B} = e^A e^B = e^B e^A$$

$$\sin(A + B) = \sin A \cos B + \cos A \sin B$$

当 $AB \neq BA$ 时, 上面等式不成立。但对 K 和 $A \times B$ 而言, 并不要求 A 与 B 可交换。

定理 9.2.7 对任何矩阵 $A \in C^{m \times m}, B \in C^{n \times n}$, 恒有

$$e^{A \times B} = e^A \times e^B = e^B \times e^A \tag{9.33}$$

证明 因为

$$(A \times E_n)(E_m \times B) = A \times B = (E_m \times B)(A \times E_n)$$

故 $A \times E_n$ 和 $E_m \times B$ 可交换。于是

$$\begin{aligned} e^{A \times B} &= e^{(A \times E_n + E_m \times B)} \\ &= e^{A \times E_n} e^{E_m \times B} \\ &= (e^A \times E_n)(E_m \times e^B) \\ &= e^A \times e^B \end{aligned}$$

请读者证明下式

$$\sin(A \times B) = \sin A \times \cos B + \cos A \times \sin B \tag{9.34}$$

9.3 Kronecker 积的应用

K 积已在许多领域中得到了应用, 包括统计、经济、系统控制等。本节, 我们只作初步介绍。

和矩阵乘积的微分公式

$$\frac{d}{dt}(AB) = \frac{dA}{dt}B + A \frac{dB}{dt}$$

相似, 有矩阵的 K 积微分公式

$$\frac{d}{dt}(A \times B) = \frac{dA}{dt} \times B + A \times \frac{dB}{dt}$$

这是因为 $A \times B$ 的第 (i, j) 位置上的矩阵子块对 t 求导数 $\frac{d}{dt}(a_{ij}B) = \frac{da_{ij}}{dt}B + a_{ij}\frac{dB}{dt}$, 它正是矩阵

$\frac{dA}{dt} \times B + A \times \frac{dB}{dt}$ 的第 (i, j) 位置上的子块。

9.3.1 利用 K 积解矩阵方程

问题一 设 $A \in C^{m \times m}, B \in C^{n \times n}, D \in C^{m \times n}$, 解矩阵方程

$$AX + XB = D \tag{9.35}$$

利用定理 9.2.1 及其推论, 方程(9.35)可化为

$$(E_n \otimes A + B^T \otimes E_m) V_{ec} X = V_{ec} D$$

或写为

$$(B^T \otimes A) V_{ec} X = V_{ec} D$$

令 $G = B^T \otimes A, X = V_{ec} X, D = V_{ec} D$, 则方程(9.35)可化为如下标准形式

$$GX = D \tag{9.36}$$

方程(9.36) 或(9.35) 有惟一解的充要条件是 G 为非奇异的, 即 G 的所有特征值不为零。根据定理 9.2.6, G 的特征值是 $\lambda_i + \mu_h$, 其中 λ_i 和 μ_h 分别是 A 和 B 的特征值。因此方程(9.35)有惟一解的充要条件是 $\lambda_i + \mu_h \neq 0$, 即 A 和 $(-B)$ 没有公共的特征值。

若 A 和 $(-B)$ 有公共的特征值, 那么方程(9.35)有解的充要条件是 $\text{rank} G = \text{rank}(G, D)$ 。

例 1 解矩阵方程 $AX + XB = D$, 其中

$$\begin{aligned} (1) \quad & A = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix} \quad B = \begin{bmatrix} -3 & 4 \\ 1 & 0 \end{bmatrix} \quad D = \begin{bmatrix} 1 & 3 \\ -2 & 2 \end{bmatrix} \\ (2) \quad & A = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix} \quad B = \begin{bmatrix} -3 & 4 \\ 0 & -1 \end{bmatrix} \quad D = \begin{bmatrix} 0 & 5 \\ 2 & -9 \end{bmatrix} \end{aligned}$$

解 (1) 令 $X = \begin{bmatrix} x_1 & x_3 \\ x_2 & x_4 \end{bmatrix}$, 将方程化为(9.37) 式形式

$$\begin{bmatrix} -2 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 4 & 0 & 1 & -1 \\ 0 & 4 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 3 \\ 2 \end{bmatrix}$$

A 的特征值为 $\lambda_1 = 1, \lambda_2 = 2$, $(-B)$ 的特征值为 $\mu = 4, \mu = -1$, 它们没有公共的特征值, 所以方程有惟一解

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

(2) 这时 A 和 $(-B)$ 有公共的特征值 $\lambda_1 = \mu = 1$, 将方程化为如下形式

$$\begin{bmatrix} -2 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 4 & 0 & 0 & -1 \\ 0 & 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 5 \\ -9 \end{bmatrix}$$

因 $\text{rank} G = \text{rank}(G, D) = 3$, 故方程有解

$$X = \begin{bmatrix} 1 & 0 \\ -2 & -1 \end{bmatrix} + k \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

其中 k 是任意常数。

问题二 设 $A, X \in C^{n \times n}$, μ 是常数, 解矩阵方程

$$AX - XA = \mu X \tag{9.37}$$

利用定理 9.2.1 及其推论, 方程(9.37)可化为

$$H = \mu I \tag{9.38}$$

此处 $H = E_n - A - A^T \in E_n$, $X = V_{ec} X$ 。齐次方程(9.38)有非平凡解 $X \neq 0$ 的充要条件是 $\det(\mu E_n - H) = 0$, 即 μ 是 H 的特征值。所以方程(9.37)有非 0 解 X 的充要条件是 $\mu = \lambda_1 - \lambda_h, 1 \leq h \leq n$, $\lambda_1, \dots, \lambda_n$ 是 A 的特征值。

例 2 解矩阵方程

$$\begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 & x_3 \\ x_2 & x_4 \end{bmatrix} - \begin{bmatrix} x_1 & x_3 \\ x_2 & x_4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} = -2 \begin{bmatrix} x_1 & x_3 \\ x_2 & x_4 \end{bmatrix}$$

解 方程化为 $H = \mu I$ 的形式, 得

$$\begin{bmatrix} 0 & 0 & -2 & 0 \\ 2 & 2 & 0 & -2 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = -2 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

A 的特征值是 $\lambda_1 = 1, \lambda_2 = 3$, 且因 $\mu = \lambda_1 - \lambda_2 = -2$, 故方程有非零解

$$X = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$$

问题三 设 $A, B, D, X \in C^{n \times n}$ 解矩阵方程

$$AXB = D \tag{9.39}$$

利用定理 9.2.1 及其推论, 把方程化为

$$H = 0 \tag{9.40}$$

其中 $H = B^T A$, $X = V_{ec} X$, $D = V_{ec} D$, 因此可用解问题一的方法来解决此方程。

现应用上述方法解更一般的矩阵方程

$$A_1 X B_1 + A_2 X B_2 + \dots + A_r X B_r = D \tag{9.41}$$

方程两端施以 V_{ec} 运算后化为

$$H = 0 \tag{9.42}$$

其中 $H = B_1^T A_1 + B_2^T A_2 + \dots + B_r^T A_r$, $X = V_{ec} X$, $D = V_{ec} D$ 。

例 3 解矩阵方程

$$A_1 X B_1 + A_2 X B_2 = D$$

其中

$$\begin{aligned} A_1 &= \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix} & B_1 &= \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} & A_2 &= \begin{bmatrix} 0 & 1 \\ -2 & -1 \end{bmatrix} \\ B_2 &= \begin{bmatrix} 0 & 2 \\ -1 & 3 \end{bmatrix} & D &= \begin{bmatrix} 4 & -6 \\ 0 & 8 \end{bmatrix} & X &= \begin{bmatrix} x_1 & x_3 \\ x_2 & x_4 \end{bmatrix} \end{aligned}$$

解 $H = B_1^T A_1 + B_2^T A_2$

$$\begin{aligned} &= \begin{bmatrix} 2 & 2 & -2 & -3 \\ 2 & -1 & 0 & 2 \\ 0 & 2 & 2 & 5 \\ -4 & -2 & -4 & -4 \end{bmatrix} \\ &= (x_1 \ x_2 \ x_3 \ x_4)^T, \quad = (4 \ 0 \ -6 \ 8)^T \end{aligned}$$

解方程 $H =$ 得到 $=H^{-1} = (1 \ -1 \ -2 \ 0)^T$, 故原方程的解

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

9.3.2 利用 K 积解矩阵微分方程

问题四 设 $A \in C^{m \times m}, B \in C^{n \times n}, X = X(t)$, 解矩阵微分方程的初值问题

$$\frac{d}{dt}X = AX + XB \tag{9.43}$$

$$X(0) = X_0 \tag{9.44}$$

解 方程两边施以 V_{ec} 运算后化为

$$\begin{cases} \dot{} = G \\ \phantom{\dot{}} (0) = \end{cases} \tag{9.45}$$

其中 $G = E_n \quad A + B^T \quad E_m, \quad = V_{ec}X, \quad = V_{ec}X_0$, 根据 5.7 节理论, 得到方程(9.45) 的解为

$$\begin{aligned} V_{ec}X &= \exp\{ (E_n \quad A + B^T \quad E_m) t \} V_{ec}X_0 \\ &= [\exp(E_n \quad A) t] [\exp(B^T \quad E_m) t] V_{ex}X_0 \\ &= [E_n \quad e^{At}] [e^{BTt} \quad E_m] V_{ec}X_0 \end{aligned}$$

此处 $\exp() = e^{()}$, 在定理 9.2.1 的推论(2) 中, 令 $X = A$, 得

$$V_{ec}(AB) = (B^T \quad E_m) V_{ec}A$$

又 $[\exp(B^T t)]^T = \exp(Bt)$, 于是有

$$(e^{BTt} \quad E_m) V_{ec}X_0 = V_{ec}(X_0 e^{Bt})$$

因此

$$V_{ec}X = V_{ec}(e^{At} X_0 e^{Bt})$$

故原初值问题的解为

$$X = e^{At} X_0 e^{Bt}$$

例 4 解矩阵微分方程的初值问题

$$\begin{cases} \frac{dx}{dt} = AX + XB \\ X(0) = X_0 \end{cases}$$

其中

$$\begin{aligned} A &= \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix} & B &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\ X &= \begin{bmatrix} x_1 & x_3 \\ x_2 & x_4 \end{bmatrix} & X_0 &= \begin{bmatrix} -2 & 0 \\ 1 & 1 \end{bmatrix} \end{aligned}$$

解 易求得

$$e^{At} = \begin{bmatrix} e^t & e^t - e^{2t} \\ 0 & e^{2t} \end{bmatrix} \quad e^{Bt} = \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}$$

故方程组的解是

$$X = e^{At} X_0 e^{Bt} = \begin{bmatrix} -e^{2t} - e^{3t} & 1 - e^t \\ e^{3t} & e^t \end{bmatrix}$$

问题五 在控制理论中, 线性系统的动态方程为

$$\frac{d}{dt} = A + B \mu \quad (\text{状态方程}) \tag{9.47}$$

$$= C \quad (\text{输出方程}) \tag{9.48}$$

其中 A, B, C 分别是 $n \times n, n \times p, q \times n$ 型常数矩阵, μ (输入), (状态), (输出) 分别是 p, n, q 维列向量。设系统的输出反馈为

$$\mu = K \tag{9.49}$$

其中 K 是待定的常数矩阵(起控制作用)。问题是要确定 K, 使闭环系统具有预先指定的特征值。

解 把(9.49) 式代入方程(9.47), 得到闭环系统方程

$$\begin{cases} \frac{d}{dt} = (A + BKC) \\ = C \end{cases} \tag{9.50}$$

于是, 上述问题归结为: 给定矩阵 A, B 和 C, 要求确定矩阵 K, 使预先指定的数 $\lambda_1, \lambda_2, \dots, \lambda_n$ 是多项式

$$\det[E - (A + BKC)] = \lambda^n + a_{n-1} \lambda^{n-1} + \dots + a_1 \lambda + a_0 \tag{9.51}$$

的根。

令 $H = A + BKC$, 则 $BKC = H - A = Q$, 两边施以 V_{ec} 运算后得

$$(C^T \quad B) V_{ec} K = V_{ec} Q$$

简记为

$$P = \tag{9.52}$$

其中 $P = C^T \quad B, \quad = V_{ec} K, \quad = V_{ec} Q$, 设 $\text{rank} P = r$, 则存在可逆阵 T, 使

$$TP = \begin{bmatrix} P_1 \\ 0 \end{bmatrix}$$

其中 P_1 是 r 行的行满秩矩阵。用 T 左乘以(9.52) 式两端, 得

$$TP = T$$

或

$$\begin{bmatrix} P_1 \\ 0 \end{bmatrix} = \begin{bmatrix} \mu \\ v \end{bmatrix} \tag{9.53}$$

方程(9.53) 有解的充要条件是 $v = 0$, 此条件依赖于矩阵 H。

- 矩阵 H 一般选取:
- (1) 对角矩阵
 - (2) 上(下)三角矩阵
 - (3) 相伴矩阵

相伴矩阵 H 是指:对多项式

$$H(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0$$

则
$$H = \begin{bmatrix} 0 & 1 & & \\ & 0 & w & \\ & & w & 1 \\ & & & w \\ -a_0 & -a_1 & \dots & -a_{n-1} \end{bmatrix}$$

称为 $H(\lambda)$ 的相伴矩阵。

例 5 求反馈秩矩阵 K , 使系统

$$\begin{cases} \dot{x} = \begin{bmatrix} 0 & 1 & 0 \\ 3 & 3 & 1 \\ 2 & -3 & 2 \end{bmatrix} x + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} u \\ y = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} x \end{cases}$$

的闭环系统有特征值 $\lambda_1 = -1, \lambda_2 = -2, \lambda_3 = -3$ 。

解 令 $H = A + BKC, Q = H - A = BKC$

因 B 的第一行全为 0, 故 Q 的第一行也全为 0, 故可选 H 为相伴矩阵。以 $-1, -2, -3$ 作为特征值的闭环系统的特征多项式为

$$(\lambda + 1)(\lambda + 2)(\lambda + 3) = \lambda^3 + 6\lambda^2 + 11\lambda + 6$$

故相应的相伴矩阵是

$$H = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix}$$

计算后, 得

$$Q = \begin{bmatrix} 0 & 0 & 0 \\ -3 & -3 & 0 \\ -8 & -8 & -8 \end{bmatrix}$$

$$P = C^T \quad B = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

现求解方程 $P = \dots$, 其中 $\dots = V_{ec} K, \dots = V_{ec} Q$ 。施行行的初等变换, 则增广矩阵 $(P \dots)$ 化为

$$A[-\int_0^{+\infty} Z(t) dt] + [-\int_0^{+\infty} Z(t) dt] B = Z(0) = D$$

这说明

$$X = -\int_0^{+\infty} Z(t) dt = -\int_0^{+\infty} e^{At} D e^{Bt} dt$$

是方程(9.35) 的解。

作为上述定理的推广, 我们证明在稳定性理论中有重要作用的下述定理。

定理 9.3.2 在矩阵方程

$$A^H X + X A = - D \tag{9.57}$$

中, 如果 $A \in C^{m \times m}$ 的所有特征值实部都是负的, D 是 Hermite 半正定(正定) 矩阵, 则方程有唯一的 Hermite 半正定(正定) 矩阵解

$$X = \int_0^{+\infty} e^{A^H t} D e^{A t} dt \tag{9.58}$$

证明 在定理 9.3.1 中, 把 A 换成 A^H , B 换成 A , D 换成 $(-D)$, 就得到式(9.58)。剩下只需证明它是 Hermite 半正定(正定) 的。

因 D 是 Hermite 半正定(正定的), 故 $D = D_1^H D_1$, 从而矩阵 $e^{A^H t} D e^{A t} = (D_1 e^{A t})^H (D_1 e^{A t})$ 是 Hermite 半正定(正定) 的, 因此积分 $\int_0^{+\infty} e^{A^H t} D e^{A t} dt$ 亦然。

习 题 9

1. 设 C^m, C^n , 证明:
$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^T = \begin{pmatrix} A^T & C^T \\ B^T & D^T \end{pmatrix} = \begin{pmatrix} A^T & B^T \\ C^T & D^T \end{pmatrix}^T$$
2. 设 $A = (A_{c1}, A_{c2}, \dots, A_{cm}), \{ \varphi_j \}$ 是 C^n 的标准基, 证明
$$A = \sum_{j=1}^n (A_{c_j} \varphi_j^T) \quad A^T = \sum_{j=1}^n (A_{c_j}^T \varphi_j)$$
3. 已知 $X^2 = 1, X^T = 1$
证明: $X = 1$ 。
4. 设 $A \in C^{m \times m}$ 和 $B \in C^{n \times n}$ 是酉矩阵
证明: $A \otimes B$ 也是酉矩阵。
5. 设 $A \in C^{m \times m}$ 和 $B \in C^{n \times n}$ 是正规矩阵
证明: $A \otimes B$ 和 $B \otimes A$ 都是正规矩阵; 如果有酉矩阵 $U_1 \in C^{m \times m}$ 和 $U_2 \in C^{n \times n}$, 使
$$U_1^H A U_1 = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$$
$$U_2^H B U_2 = M = \text{diag}(\mu_1, \mu_2, \dots, \mu_n)$$

证明: $A \otimes B$ 酉相似准对角阵 $\text{diag}(\lambda_1 M, \dots, \lambda_m M), B \otimes A$ 酉相似准角阵 $\text{diag}(\mu_1 \Lambda, \mu_2 \Lambda, \dots, \mu_n \Lambda)$ 。
6. 证明: (1) 对角矩阵的 K 积仍是对角矩阵; (2) 上(下)三角矩阵的 K 积仍是上(下)三角矩阵。
7. 如果 A 和 B 是对角化矩阵, 证明: $A \otimes B$ 也是可对角化矩阵。

8. 设 $B \in \mathbb{C}^{n \times n}$ 的特征值是 $\lambda_1, \lambda_2, \dots, \lambda_n$ (不必不相同), $m \times m$ 阶矩阵

$$A = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \hline 1 & 1 & \dots & 1 \end{bmatrix}$$

求 $A + B$ 矩阵的特征值。

9. 如果 $A \in \mathbb{R}^{m \times m}$ 和 $B \in \mathbb{R}^{n \times n}$ 是正定矩阵, 证明: $A + B$ 也是正定矩阵。

10. 若 $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$ 是 \mathbb{C}^m 的基, $\{\beta_1, \beta_2, \dots, \beta_n\}$ 是 \mathbb{C}^n 的基, 证明: $\alpha_j^T \beta_i, 1 \leq j \leq n, 1 \leq i \leq m$, 构成 $\mathbb{C}^{m \times n}$ 的基。

11. 令 $\alpha \in \mathbb{C}^m, \mu \in \mathbb{C}^n$ 。证明: 由等式 $\mu = \alpha$ 推出存在某个数 C , 使 $\mu = C\alpha, \mu =$

练习答案(仅供参考)

习 题 1

3. (1) 是。 (2) 是。
(3) 不是。因为若 A, B 皆是可逆矩阵, 但 A + B 不一定可逆。
(4) 不是。因为若 0, 则 1 · = 0 。
4. (1) 中的线性空间维数是 n(n + 1) / 2 维。
(2) 中的线性空间维数是 n(n - 1) / 2 维。
5. 因为 cos2t = 2cos²t - 1
6. $\alpha_1 = (6, 6, 4, 0), \alpha_2 = (-3, 7, 0, 4)$ 为解空间的基, 维数 = 2。
7. 对基 $\alpha_1, \alpha_2, \alpha_3$ 的坐标是(33, - 82, 154) 。
8. $1 + t + t^2$ 对基, 1, t - 1, (t - 2) (t - 1) 的坐标是(3, 4, 1) 。
9. 设($\alpha_1, \alpha_2, \alpha_3$) = ($\beta_1, \beta_2, \beta_3$) A

任意向量 $= (\alpha_1, \alpha_2, \alpha_3) \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = (\beta_1, \beta_2, \beta_3) \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix}$

那么有 $= (\alpha_1, \alpha_2, \alpha_3) A \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix}$

所以 $\begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = A \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix}$, 其中 $A = \begin{bmatrix} -27 & -71 & -41 \\ 9 & 20 & 9 \\ 4 & 12 & 8 \end{bmatrix}$

10. (1) $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (\alpha_1, \alpha_2, \alpha_3, \alpha_4) \begin{bmatrix} 2 & 0 & 5 & 6 \\ 1 & 3 & 3 & 6 \\ -1 & 1 & 2 & 1 \\ 1 & 0 & 1 & 3 \end{bmatrix}$
- (2) $(1, 0, 1, 0)$ 在 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 之下的坐标 $(-5/9, -8/27, 1/3, 2/27)$ 。
- (3) 对两个基有相同坐标的非零向量为 $(1, 1, 1, -1)$ 。
11. 因为 $2\alpha_1 - 3\alpha_2 + \alpha_3 = 0$, 故 $L(\alpha_1, \alpha_2, \alpha_3)$ 的基可选 α_1, α_2 。
12. $V_1 = V_2$ 的基 $\alpha_1 = (-1, 0, 1, 0)$ $\alpha_2 = (0, -1, 0, 1)$ 。
13. (1) 是一个子空间。
(2) 不是。
14. 即证 $\alpha_1, \dots, \alpha_s$ 的极大线性无关组所含向量的个数 $= \text{rank}(A)$ 。设 $\text{rank}(A) = r$, 不妨设 A 的前 r 列线性无关, 那么可以证得 $\alpha_1, \dots, \alpha_r$ 线性无关, 又证任添一个 $\alpha_j (j = r + 1, \dots, s)$, 使 $\alpha_1, \dots, \alpha_r, \alpha_j$ 线性相关, 即可证得结论。

习 题 2

4. $\pm \frac{1}{57}(34, -44, 6, -11)$
5. 提示: 设 $\alpha = k_1 \alpha_1 + \dots + k_n \alpha_n, (\alpha, \alpha) = (k_1 \alpha_1 + \dots + k_n \alpha_n, \alpha) = 0$
8. $\alpha_1 = \frac{\sqrt{2}}{2}(\alpha_1 + \alpha_5)$
 $\alpha_2 = \frac{\sqrt{10}}{10}(\alpha_1 - 2\alpha_2 + 2\alpha_4 - \alpha_5)$
 $\alpha_3 = \frac{1}{2}(\alpha_1 + \alpha_2 + \alpha_3 - \alpha_5)$, 为 V 的一组标准正交基。
9. $\alpha_1^0 = \frac{1}{\sqrt{5}}(0, 2, 1, 0)$
 $\alpha_2^0 = \frac{5}{\sqrt{30}}(1, -\frac{1}{5}, \frac{2}{5}, 0)$
 $\alpha_3^0 = \frac{2}{\sqrt{10}}(\frac{1}{2}, \frac{1}{2}, -1, 1)$
 $\alpha_4^0 = \frac{15}{\sqrt{60}}(-\frac{2}{15}, -\frac{2}{15}, \frac{4}{15}, \frac{6}{15})$
10. 提示: $(\alpha_i, \alpha_j) = (\sum_{k=1}^n a_{ik} \alpha_k, \sum_{t=1}^n a_{jt} \alpha_t)$
 $= \sum_{k=1}^n \sum_{t=1}^n a_{ik} a_{jt} (\alpha_k, \alpha_t)$
 $= \sum_{k=1}^n a_{ik} a_{jk}$

11. 提示: $x_i = 0 \cdot x_1 + \dots + 1 \cdot x_i + \dots + 0 \cdot x_n$
$$x_i = 1 = |x_i|, x_j = |x_i| \cos Q_{ij} = 0$$
12. 提示: 选 $\alpha = (\alpha_1, \dots, \alpha_n), \alpha_i = \begin{cases} 1 & a_i \geq 0 \\ -1 & a_i < 0 \end{cases}$ 利用 Cauchy-Schwarz 不等式。
13. 提示: 扩 $\alpha_1, \dots, \alpha_m$ 为 V 的标准正交基 $\alpha_1, \dots, \alpha_m, \alpha_{m+1}, \dots, \alpha_n$, 设 $\alpha = k_1 \alpha_1 + \dots + k_n \alpha_n$
$$\sum_{i=1}^m (\alpha, \alpha_i)^2 + \sum_{i=m+1}^n (\alpha, \alpha_i)^2 = |\alpha|^2$$

习 题 3

1. 1), 2), 3), 7) 不是线性变换
4), 5), 6), 8) 是线性变换。

3. 1) $A = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ 4) $A = \begin{bmatrix} -1 & 1 & -2 \\ 2 & 2 & 0 \\ 3 & 0 & 2 \end{bmatrix}$

2) $A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \hline 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$ 3) $A = \begin{bmatrix} a & b & 1 & 0 & 0 & 0 \\ -b & a & 0 & 1 & 0 & 0 \\ 0 & 0 & a & b & 1 & 0 \\ 0 & 0 & -b & a & 0 & 1 \\ 0 & 0 & 0 & 0 & a & b \\ 0 & 0 & 0 & 0 & -b & a \end{bmatrix}$

4. 1) $A = \frac{1}{3} \begin{bmatrix} 6 & -9 & 9 & 6 \\ 2 & -4 & 10 & 10 \\ 8 & -16 & 40 & 40 \\ 0 & 3 & -21 & -24 \end{bmatrix}$

2) $R(A) = L(\alpha_1, \dots, \alpha_4) = L(\alpha_1, \alpha_2)$
 $N(A) = \{ \alpha \mid A\alpha = 0, \alpha \in V \}$
$$= L(-2\alpha_1 - \frac{3}{2}\alpha_2 + \alpha_3, -\alpha_1 - 2\alpha_2 + \alpha_4)$$

3) 由 $\alpha_1 = -2\alpha_1 - \frac{3}{2}\alpha_2 + \alpha_3, \alpha_2 = -\alpha_1 - 2\alpha_2 + \alpha_4$, 知 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 是 V 的一个基, A 在此基下的矩阵

$$B = \begin{bmatrix} 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & 5 & 2 \\ 0 & 0 & 9/2 & 1 \end{bmatrix}$$

4) 由 $A\alpha_1 = \alpha_1 - \alpha_2 + \alpha_3 + 2\alpha_4, A\alpha_2 = 2\alpha_2 + 2\alpha_3 - 2\alpha_4$, 易知 $A\alpha_1, A\alpha_2, \alpha_3, \alpha_4$ 是 V 的一个基, A 在此基下的矩阵为

$$D = \begin{bmatrix} 5 & 2 & 2 & 1 \\ 9/2 & 1 & 3/2 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

5. 1) 特征值 $\lambda_1 = 7, \lambda_2 = -2$
对应的线性无关特征向量 $\alpha_1 = \alpha_1 + \alpha_2, \alpha_2 = 4\alpha_1 - 5\alpha_2$
- 2) 特征值 $\lambda_1 = \lambda_2 = 1, \lambda_3 = -2$
对应的线性无关特征向量 $\alpha_1 = 3\alpha_1 - 6\alpha_2 + 20\alpha_3, \alpha_3 = \alpha_3$
- 3) 特征值 $\lambda_1 = \lambda_2 = \lambda_3 = 2, \lambda_4 = -2$
对应的线性无关的特征向量 $\alpha_1 = \alpha_1 + \alpha_2, \alpha_2 = \alpha_1 + \alpha_3, \alpha_3 = \alpha_1 + \alpha_4, \alpha_4 = \alpha_1 - \alpha_2 - \alpha_3 - \alpha_4$

6. 在上题中, 第 1, 3 两个线性变换的矩阵可以在适当的基下变成对角形。

第 1) 题 $(\alpha_1, \alpha_2) = (\alpha_1, \alpha_2) \begin{bmatrix} 1 & 4 \\ 1 & -5 \end{bmatrix}$

第 3) 题 $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (\alpha_1, \alpha_2, \alpha_3, \alpha_4) \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$

7. 提示: 设 A 是线性变换 A 在某基下的矩阵表示, 若 $\lambda = 0$ 是其特征值, 则 $|A| = 0$ 。
8. 提示: 用反证法, 若 $\alpha_1 + \alpha_2$ 是 A 的特征向量, 那么有数 λ , 使 $A(\alpha_1 + \alpha_2) = \lambda(\alpha_1 + \alpha_2)$ 。
9. 提示: 证明集合 $N(A) = \{ \alpha - A\alpha \mid \alpha \in V \}$, 再证 $N(A) \cap R(A) = \{0\}$ 。
10. 提示: 对 A 的 n 个互异特征值 $\lambda_1, \dots, \lambda_n$ 及对应的特征向量 $\alpha_1, \dots, \alpha_n$, 有 $L(\alpha_i) = A\alpha_i$, $i = 1, 2, \dots, n$, 由于 $AB = BA$, 可得 $AB\alpha_i = BA\alpha_i = \lambda_i B\alpha_i$, 所以 $B\alpha_i \in L(\alpha_i)$, 则 $B\alpha_i = k_i \alpha_i, i = 1, \dots, n$ 。

11. 1) $\begin{bmatrix} 1 & & & \\ & (\lambda + 1) & & \\ & & (\lambda + 1)^2 & \\ & & & \ddots \end{bmatrix}$

2) $\begin{bmatrix} 1 & & & \\ & (\lambda - 1) & & \\ & & (\lambda - 1)^2 & \\ & & & \ddots \end{bmatrix}$

3) $\begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & (\lambda - 1) \\ & & & & \ddots \end{bmatrix}$

4) $\begin{bmatrix} \lambda^k & \lambda^{k-1} & \dots & c_k^2 & \lambda^{k-2} \\ & \lambda^k & & \lambda^{k-1} & \\ & & \lambda^k & & \\ & & & \lambda^k & \end{bmatrix}^T$

15. A 的若当标准形矩阵 $J_1 = \begin{bmatrix} 1 & & \\ & i & \\ & & -i \end{bmatrix}$

B 的若当标准形矩阵 $J_2 = \begin{bmatrix} -1 & & & & \\ & -1 & & & \\ & & 2 & & \\ & & & 1 & \\ & & & 1 & 1 \\ & & & & 1 & 1 \end{bmatrix}$

C 的若当标准形矩阵 $J_3 = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 2 & & \\ & & & W & \\ & & & & n-1 \end{bmatrix}$

其中 $\epsilon_1, \dots, \epsilon_{n-1}$ 是 $x^n - 1 = 0$ 的 $n-1$ 个虚 n 次单位根。

16. $J = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}, \quad P = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 0 & -1 & 0 \end{bmatrix}$

17. $A = \frac{1}{9} \begin{bmatrix} 14 & -2 & -14 \\ -2 & -1 & -16 \\ -14 & -16 & 5 \end{bmatrix}$

18. 由 $|E - A| = \lambda^2 - 6\lambda + 7 = (\lambda - 2)(\lambda - 3)$, 所以 $f(\lambda) = \lambda^4 - 12\lambda^3 + 19\lambda^2 - 29\lambda + 37$
 $= (\lambda - 2)q(\lambda) + (\lambda + 2)$

所以 $[f(A)]^{-1} = (A + 2E)^{-1} = \frac{1}{23} \begin{bmatrix} 7 & 1 \\ -2 & 3 \end{bmatrix} = -\frac{1}{23}A + \frac{8}{23}E$

20. 提示。找 T 在基 $\epsilon_1 = (1, 0, 0), \epsilon_2 = (0, 1, 0), \epsilon_3 = (0, 0, 1)$ 之下的矩阵表示。

21. 提示: 证明 A 的最小多项式无重根。

22. A 的最小多项式 $\lambda^2(\lambda - 1)$, 则 $A^3 = A^2$, 所以

$$A^{100} = A^{97} \cdot A^2 = \dots = A^2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix}$$

23. $\begin{bmatrix} -3 & 48 & -26 \\ 0 & 95 & -61 \\ 0 & -61 & 34 \end{bmatrix}$

24. A 的特征值 $\lambda_1 = 2, \lambda_2 = 4, \lambda_3 = 6, \lambda_4 = 8$, 它们分别对应的特征向量就是 Q 的列向量 $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$, 其中

$$Q^{-1}AQ = \begin{bmatrix} 2 & & & \\ & 4 & & \\ & & 6 & \\ & & & 8 \end{bmatrix}$$

A 的特征值与 A 相同, 特征向量是

$$\alpha_i = (\alpha_{i1}, \alpha_{i2}, \alpha_{i3}, \alpha_{i4})^T, i = 1, 2, 3, 4$$

- 25. 对 A 的阶数 n 用数学归纳法。
- 26. 根据 25 题结论。

习 题 4

- 1. (1) 因为 $\|x\|_2^2 = |x_1|^2 + \dots + |x_n|^2$ $\|x\|_1^2 = (|x_1| + \dots + |x_n|)^2$
又因为习题 2 第 12 题结论
 $(|x_1| + \dots + |x_n|)^2 \leq n(|x_1|^2 + \dots + |x_n|^2)$
(2) 因为 $\max_i |x_i| \leq \frac{1}{n}(|x_1| + \dots + |x_n|) \leq \max_i |x_i|$
(3) 因为 $\max_i |x_i| \leq (\sum_{i=1}^n |x_i|^2)^{\frac{1}{2}} = \sqrt{n} \max_i |x_i|$
- 2. α, β 相关 $\alpha = k\beta$
 $\alpha^T \beta = 0 \quad k = 0$ 。
- 3. 由范数定义可直接得到。
- 4. 只证三角不等式
(1) 不妨设 $\max(|a| + |b|, |a| + |b|) = |a| + |b|$
则 $|a| + |b| = \max(|a|, |b|) + \min(|a|, |b|)$
 $= \max(|a|, |b|) + \max(|a|, |b|)$
 $= |c| + |c|$
(2) 略
- 5. 提示: 令 $\alpha = \sqrt{a_1}, \dots, \sqrt{a_n}$
只需证明 $\|\alpha\|_2 = \sqrt{2}$
- 6. 提示: 只需验证矩阵范数的条件满足, 做一点适当变形, 对范数 $\|\cdot\|_a$ 而言。
- 7. 提示: 如第 6 题。
- 8. 因为 $\|E\|_1 = \max_i \|E\|_1 = \|E\|_1 = 1$
- 9. 因为 $\|AU\|_2^2 = (AU)^T (AU) = [AU(AU)^T] = (AA^T) = \|A\|_2^2$
- 10. 因为 $\|A\|_F^2 = \text{tr}(A^H A) = \lambda_1 + \lambda_2 + \dots + \lambda_n$
其中 $\lambda_1, \lambda_2, \dots, \lambda_n$ 是 $A^H A$ 的特征值, λ_1 为最大,
又 $\|A\|_2^2 = \lambda_1(A^H A) = \lambda_1$
故 $\|A\|_F^2 \leq n \|A\|_2^2$

故 $\begin{pmatrix} 2 \\ 2 \end{pmatrix} = A \begin{pmatrix} 2 \\ 2 \end{pmatrix}$ 是 A 的特征向量

11. 因为 $\begin{pmatrix} 2 \\ 2 \end{pmatrix} = A \begin{pmatrix} 2 \\ 2 \end{pmatrix}$ 是 A 的特征向量

所以

$$A \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

又因 $A^{-1} = \frac{1}{\dots}$

12. 因为 $K(AB) = AB \cdot (AB)^{-1} = A \cdot B \cdot B^{-1} = A^{-1}$

13. 注意 9 题结论。

14. (1) $A^{-1} = 17$ $A = 12$

$$(2) \quad B^{-1} = \frac{3 + \sqrt{5}}{2} \quad B = 2 \quad B^{-1} = 2$$

15. $A^{-1} = 0.28$ 可用定理 4.3.2 估计

$$\frac{A^{-1} - (A +)^{-1}}{A^{-1}} = \frac{0.28}{1 - 0.28}$$

习 题 5

1. 例如 $A_k = \begin{bmatrix} e^{\frac{1}{k}} & 0 \\ 0 & \frac{1}{k} \end{bmatrix}$

2. 若 $\lim_n A^n = E$ 证明 $A = E$

设 $A = PJP^{-1}$ $A^n = PJ^nP^{-1}$

即证

$$\lim_n J^n = E \quad \text{有} \quad J = E$$

注意 J_i^n 的构造, J_i 必须是 1 级方阵, 且 $J_i^n = 1$ 只能 $J_i = 1$ 。

3. 因为当 $|z| < 1$ 时 $\sum_{k=0}^{\infty} z^k = \frac{1}{1 - z}$

故当 $\rho(A) < 1$ 时 $\sum_{k=0}^{\infty} A^k = (E - A)^{-1}$

$$\sum_{k=0}^{\infty} \begin{bmatrix} 0.2 & 0.7 \\ 0.3 & 0.6 \end{bmatrix}^k = \begin{bmatrix} 0.8 & -0.7 \\ -0.3 & 0.4 \end{bmatrix}^{-1} = \frac{1}{0.11} \begin{bmatrix} 0.4 & 0.7 \\ 0.3 & 0.8 \end{bmatrix}$$

4. 提示: 因为当 $|z| < 1$ 时

$$\sum_{k=0}^{\infty} (k+1)z^k = (1 - z)^{-2}$$

故 $\sum_{k=0}^{\infty} kA^k = A(E - A)^{-2}$

5. 利用 4 题结论 $A = \frac{1}{10} \begin{bmatrix} 1 & 2 \\ 8 & 1 \end{bmatrix}$

- 6. 略
- 7. 略
- 8. 略

9. (1) $\frac{d\text{tr}(AXB)}{dX}$ 的第 j 行 i 列处元素是

$$\begin{aligned}\frac{d}{dx_{ij}}\text{tr}(AXB) &= \text{tr} \frac{d}{dx_{ij}}(AXB) \\ &= \text{tr}\left[\frac{A}{x_{ij}}XB + AX\frac{B}{x_{ij}}\right] \\ &= \text{tr}(AE_{ij}B) + \text{tr}(BAE_{ij}) \\ &= (BA)_{ji}\end{aligned}$$

- (2) 类似(1)。
- (3) 由例 3 令 $A=B=E$ 即得结果。
- (4) 由例 4 令 $A=E$ 即得结果。
- (5) 参考例 3 的证法。

10. $\frac{d}{d}\begin{pmatrix} - & \end{pmatrix}^T = \frac{d}{d}\begin{pmatrix} ^T & - & ^T \end{pmatrix}$
 $= \frac{d}{d}\begin{pmatrix} ^T \end{pmatrix} = ^T$

- 11. 由性质 4, 性质 5 易得。
- 12. 因为 $|E - A| = ^2 + a^2$
故有 $A^2 + a^2E = 0$

由
$$\begin{aligned}e^A &= E + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \dots \\ &= E - \frac{a^2}{2!}E + \frac{1}{4!}a^4E + \dots + (-1)^k \frac{1}{(2k)!}a^{2k}E + \dots \\ &\quad + \frac{A}{a}\left(a - \frac{a^3}{3!} + \frac{a^5}{5!} + \dots + (-1)^k \frac{1}{(2k+1)!}a^{2k+1} + \dots\right) \\ &= \cos a \cdot E + \frac{1}{a}\sin a \cdot A \\ &= \begin{bmatrix} \cos a & \sin a \\ -\sin a & \cos a \end{bmatrix}\end{aligned}$$

13. $A = E + A \quad A = \begin{bmatrix} 0 & \\ - & 0 \end{bmatrix}$
$$e^A = e \begin{bmatrix} \cos & \sin \\ -\sin & \cos \end{bmatrix}$$

14. $\begin{bmatrix} -184 & 189 \\ -189 & 194 \end{bmatrix}$

15. $A^{1000} = a_0E + a_1A$

其中 $a_0 = \frac{5}{4} - \frac{1}{4}5^{1000} \quad a_1 = \frac{1}{4}(5^{1000} - 1)$

16. 因不知若当标准型, 故只能用特征多项式得

$$\begin{aligned} A^4 - \frac{1}{2} A^2 &= 0 \\ \sin A &= A - \frac{1}{3!} A^3 + \frac{1}{5!} A^5 - \frac{1}{7!} A^7 + \dots \\ &= A - \frac{1}{3!} A^3 + \frac{1}{5!} A^5 - \frac{1}{7!} A^7 + \dots \\ &= A + A^3 \left(-\frac{1}{3!} + \frac{1}{5!} A^2 - \frac{1}{7!} A^4 + \dots \right) \\ &= A + \frac{A^3}{3} \left(-1 + \frac{1}{5} A^2 - \frac{1}{7} A^4 + \dots \right) \\ &= A + \frac{A^3}{3} (\sin A - A) \\ &= A - \frac{1}{3} A^3 \end{aligned}$$

$$\cos A = E - \frac{2}{2} A^2$$

17. (1) $f(A) = A^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 \\ -\frac{1}{4} & \frac{1}{2} & 0 & 0 \\ \frac{1}{8} & -\frac{1}{4} & \frac{1}{2} & 0 \\ -\frac{1}{16} & \frac{1}{8} & -\frac{1}{4} & \frac{1}{2} \end{bmatrix}$

(2) $f(A) = A^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 \\ -\frac{1}{4} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

18. $\sqrt{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

19. (1) $e^{Az} = a_0(z) E + a_1(z) A + a_2(z) A^2$
其中 $a_0(z) = 3e^{-z} - 3e^{-2z} + e^{-3z}$
 $a_1(z) = \frac{5}{2}e^{-z} - 4e^{-2z} - \frac{3}{2}e^{-3z}$
 $a_2(z) = -\frac{1}{2}e^{-z} - e^{-2z} - \frac{1}{2}e^{-3z}$

(2)
$$e^{Az} = \begin{bmatrix} e^{3z} & 0 & 0 & 0 \\ 0 & e^{-2i} & 0 & 0 \\ 0 & ze^{-2z} & e^{-2z} & 0 \\ 0 & \frac{1}{2!}z^2e^{-2z} & ze^{-2z} & e^{-2z} \end{bmatrix}$$

(3)
$$e^{At} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} e^{-z} & 0 \\ 0 & e^{-2z} \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2e^{-z} - e^{-2z} & e^{-t} - e^{-2t} \\ -2e^{-t} + 2e^{-2t} & -e^{-z} + 2e^{-2z} \end{bmatrix}$$

20.
$$(t) = e^{At} \begin{bmatrix} 17e^{5t} + 9e^{-5t} + 4e^{-15t} \\ -17e^{5t} - 9e^{-5t} + 6e^{-15t} \\ 17e^{5t} - 9e^{-5t} + 2e^{-15t} \end{bmatrix}$$

21.
$$(t) = e^{At} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \int_0^t e^{A(t-\tau)} Bu(\tau) d\tau$$

其中
$$A = \begin{bmatrix} -6 & 1 & 0 \\ -11 & 0 & 1 \\ -6 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 6 \\ 2 \end{bmatrix}$$

$$(t) = \begin{bmatrix} e^{-t} & e^{-2t} & e^{-3t} \\ 5e^{-t} & 4e^{-2t} & 3e^{-3t} \\ 6e^{-t} & 3e^{-2t} & 2e^{-3t} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -4 & 2 & -1 \\ \frac{9}{2} & -\frac{3}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} X_1(0) \\ X_2(0) \\ X_3(0) \end{bmatrix}$$

$$+ \begin{bmatrix} 1 & 1 & 1 \\ 5 & 4 & 3 \\ 6 & 3 & 2 \end{bmatrix} \begin{bmatrix} e^{-t} - 1 \\ 1 - e^{-2t} \\ \frac{1}{3}(1 - e^{-3t}) \end{bmatrix}$$

习 题 6

1. $A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ 2 & -\frac{1}{4} & -\frac{27}{4} & 0 \\ 1 & -\frac{1}{4} & -\frac{3}{4} & -\frac{3}{4} \end{bmatrix}$, 当 $A = LU$ $\begin{cases} U = \\ L = \end{cases}$

其中 $L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ 2 & -\frac{1}{4} & -\frac{27}{4} & 0 \\ 1 & -\frac{1}{4} & -\frac{3}{4} & -\frac{3}{4} \end{bmatrix}$ $U = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 0 & -4 & -7 & -11 \\ 0 & 0 & 1 & \frac{13}{9} \\ 0 & 0 & 0 & \frac{68}{9} \end{bmatrix}$

$= (-1, -1, 0, 1)^T$ $= (1, -7, \frac{13}{9}, \frac{68}{9})^T$

2. A 的特征值 1, 1, - 1

$$\text{变换阵 } T = \begin{bmatrix} 1 & 3 & 3 \\ 5 & 0 & 2 \\ 0 & 5 & 4 \end{bmatrix}, \quad T^{-1}AT = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

$$S_1 = \frac{1}{5} \begin{bmatrix} 1 & 3 \\ 5 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} -10 & 3 & 6 \\ -20 & 4 & 13 \end{bmatrix} = \begin{bmatrix} -14 & 3 & 9 \\ -10 & 3 & 6 \\ -20 & 4 & 13 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix} (5 \quad -1 \quad -3) = \begin{bmatrix} 15 & -3 & -9 \\ 10 & -2 & -6 \\ 20 & -4 & -12 \end{bmatrix}$$

$$A = S_1 - S_2$$

$$3. A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ -2 & 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -\frac{5}{3} \\ 0 & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & \frac{1}{3} \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ -1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 \end{bmatrix}$$

$$4. B = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & -2 & -2 \end{bmatrix}$$

5. $B^H B$ 的特征值 4, 12, 0, 0

$$\text{对应的单位特征向量阵 } U = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 0 & 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 0 & 1/\sqrt{3} & 0 & 2/\sqrt{6} \end{bmatrix}$$

$$V_1 = BU_{1 \times 2}^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$
 扩 V_1 为 $V = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$

$$B = V \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2\sqrt{3} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} U^H$$

习 题 7

1. (1) 略 (2) 略
- (3) 因为 $AGA = A$ 中 A 是可逆的, 所以
- $$GA = E \quad G = A^{-1}$$
- (4) 略
2. 因为 A 为 n 阶方阵, 故 G 也是 n 阶方阵, 所以 令 $G = Q \begin{bmatrix} E_r & 0 \\ 0 & E_{n-r} \end{bmatrix} P$ 便得所求之一
3. 直接验证
4. 直接验证
5. 直接验证
6. 因为 B 是列满秩, 故

$$B^+ = (B^H B)^{-1} B^H$$
$$C^+ = C^H (CC^H)^{-1}$$
$$A^+ = C^H (CC^H)^{-1} (B^H B)^{-1} B^H = C^+ B^+$$

同理

而

7. (1)
$$\begin{bmatrix} 1 & 0 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 1 \\ 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

故
$$A_1^{(1)} = Q \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} P$$

$$\begin{aligned}
 &= \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \\
 &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\
 A_2^{(2)} &= \begin{bmatrix} 1 & 0 & -1 & -1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & -2 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}
 \end{aligned}$$

8. 略

9. 直接验证

10. 由定理 7.2.1 可得

11. 因为 $\text{rank}(A^H A) = \text{rank}(A)$

故 $\text{rank}(A^H A) = \text{rank}(A^H A \dots A^H) = \text{rank}[A^H (A \dots)] = \text{rank}(A)$

$$12. (1) A^+ = \frac{1}{25} \begin{bmatrix} 1 & 0 & 2 \\ 2 & 0 & 4 \end{bmatrix}$$

$$(2) A^+ = \frac{1}{3} \begin{bmatrix} -2i & 1 & -i \\ 1 & -i & 2 \end{bmatrix}$$

$$(3) A^+ = \frac{1}{3} \begin{bmatrix} 2 & -3 \\ 1 & 0 \\ -1 & 3 \end{bmatrix}$$

$$(4) A^+ = \frac{1}{8} \begin{bmatrix} 2 & -2 & 2 & 2 \\ -1 & 3 & -1 & 1 \\ 1 & -3 & 1 & -1 \end{bmatrix}$$

$$13. A^+ = \frac{1}{25} \begin{bmatrix} 1 & 0 & 2 \\ 2 & 0 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \frac{1}{25} \begin{bmatrix} 5 \\ 10 \end{bmatrix} \text{是极小范数解}$$

$$= \frac{1}{5} \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \frac{1}{5} \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \text{是通解。}$$

$$14. A^+ = \frac{1}{25} \begin{bmatrix} 1 & 0 & 2 \\ 2 & 0 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$= \frac{1}{5} \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \text{是通解。}$$

习 题 8

1. 因为 λ_0 是 A 的特征值的必要充分条件是 $\lambda - \lambda_0$ 是 $A - \lambda_0 E$ 的特值。所以

$$\begin{vmatrix} a - \lambda_0 & b \\ b & a - \lambda_0 \end{vmatrix} = (a - \lambda_0)^2 - b^2 = 0$$

那么当 $\lambda_0 < a$ 时 $0 < a - \lambda_0$

所以 $0 < \lambda - \lambda_0$

则 $A - \lambda_0 E$ 为正定矩阵。

同理证明当 $\lambda_0 > b$, $A - \lambda_0 E$ 是负定矩阵

2. 对任意的 $\lambda_0 < a + c$, 证明 $A + B - \lambda_0 E$ 正定就行。因为对任意的 $\lambda_0 < a + c$, 可找 a_0, c_0 使

$$\lambda_0 = a_0 + c_0, \text{ 且 } a_0 < a, c_0 < c$$

则 $A - a_0 E, B - c_0 E$ 皆正定

则 $A + B - \lambda_0 E$ 正定。

同理可证对任意 $\lambda_0 > b + d$, $A + B - \lambda_0 E$ 负定

由 1 题结论可得本题结论。

3. 设属于 μ 的特征向量 $x = (x_1, \dots, x_n)^T$

即
$$PA = \mu x$$
$$x^H A^H A x = x^H A^H P^H P A x = |\mu|^2 x^H x$$

选择单位长度的特征向量 x , 则

$$x^H A^H A x = |\mu|^2$$

另一方面

$$x^H A^H A x = (\bar{x}_1, \dots, \bar{x}_n) \begin{bmatrix} |a_1|^2 & & \\ & |a_2|^2 & \\ & & \ddots \\ & & & |a_n|^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$
$$= |a_1|^2 |x_1|^2 + \dots + |a_n|^2 |x_n|^2$$

所以有 $\min\{|a_i|\} \leq |\mu| \leq \max\{|a_i|\}$

4. 略

5. $D_1: |z - i| = 0.43$

$D_2: |z| = 1.02$

$D_3: |z - 2| = 0.051$

选正数 $b_1 = b_2 = 1, b_3 = 0.1$

$G_1: |z - i| = 0.07$

$G_2: |z| = 0.12$

$G_3: |z - 2| = 0.051$

6. 略

7. 盖尔圆是 $D_1: |z - \frac{1}{4}| = \frac{3}{4}$

$D_2: |z - \frac{2}{5}| = \frac{3}{5}$

$D_3: |z - \frac{3}{6}| = \frac{3}{6}$

$D_4: |z - \frac{3}{7}| = \frac{3}{7}$

总之 $|z| = 1$
所以 $(A) = 1$

选正数 $b_1 = b_2 = b_3 = 1, b_4 = \frac{1}{1.1}$

则

$$\begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4.4} \\ \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5.5} \\ \frac{1}{6} & \frac{1}{6} & \frac{3}{6} & \frac{1}{6.6} \\ \frac{1.1}{7} & \frac{1.1}{7} & \frac{1.1}{7} & \frac{3}{7} \end{bmatrix}$$

可得 $|z| < 1$
故 $(A) < 1$

8. 因为 $|E - A| = 0$
所以 1 是 A 的特征值, 又由盖尔圆定理 $(A) = 1$ 所以 $(A) = 1$ 。

9. 因为 $A = \mu^H A = \mu^H$

那么 $\mu^H A = \mu^H$

$\mu^H = \mu^H$

$(\mu^H - \mu^H) = 0$

由于 μ^H 所以 $\mu^H = 0$

习 题 9

- 1. 略
- 2. 因为 $A = (A_{c1}, A_{c2}, \dots, A_{cn})$
而 $A_{cj}^T = (0, \dots, 0, A_{cj}, 0, \dots, 0)$
- 3. 因为 $X_2 = 1, Y_2 = 1$
所以它们的特征值有 $\max | \lambda_i | = 1, \max | \mu | = 1$
 $(X \ Y)^T (X \ Y) = X^T X \ Y^T Y$ 的特征值 λ_i, μ 也有 $\max | \lambda_i \mu | = 1$
- 4. 因为 $(A \ B)^H (A \ B) = (A^H \ B^H) (A \ B) = A^H A \ B^H B$

$E_m \quad E_n = E_{mn}$

5. 因为 $(A \quad B)^H (A \quad B) = A^H A \quad B^H B$
 $= AA^H \quad BB^H$
 $= (A \quad B) (A \quad B)^H$

因为 $U_1^H AU_1 = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & m \end{bmatrix}$

$U_2^H BU_2 = \begin{bmatrix} \mu & & \\ & \ddots & \\ & & \mu \end{bmatrix} = M$

故 $(U_1 \quad U_2)^H (A \quad B) (U_1 \quad U_2)$
 $= U_1^H AU_1 \quad U_2^H BU_2$
 $= \begin{bmatrix} 1 & & \\ & \ddots & \\ & & m \end{bmatrix} M = \begin{bmatrix} 1M & & \\ & \ddots & \\ & & mM \end{bmatrix}$

6. 略

7. 因为 $P^{-1}AP = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}, \quad Q^{-1}BQ = \begin{bmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_m \end{bmatrix}$
 $\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_m \end{bmatrix} = P^{-1}AP \quad Q^{-1}BQ$
 $= (P \quad Q)^{-1} (A \quad B) (P \quad Q)$

8. 因为 $|E - A| = (-\lambda)^{m-1}$
故 A 的特征值 $m, 0(m-1 \text{ 重})$

则 A B 的特征值是 $m_1, m_2, \dots, m_n, 0, \dots, 0$

9. 因为 A, B 皆为正定阵, $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_n$ 皆大于 0。

所以 A B 的特征值 $\lambda_i \mu_j$ 皆为正, $i=1, \dots, m$
 $j=1, \dots, n$

10. 注意到 $A \quad B = 0 \quad A=0$ 或 $B=0$

那么, $(a_1 \quad \dots \quad a_n)^T (b_1 \quad \dots \quad b_m)^T$
 $= \sum_{j=1}^m \sum_{i=1}^n a_i b_j$ (*)

若(*)式右边 = 0, 则 $a_1 \quad \dots \quad a_n = 0$
或 $b_1 \quad \dots \quad b_m = 0$

由于 $\lambda_1, \dots, \lambda_n$ 线性无关, μ_1, \dots, μ_m 线性无关

那么 $a_i b_j = 0 \quad i=1, \dots, n, \quad j=1, \dots, m$

所以 $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$

.....
 $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}, \dots, \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ 是 $C^{m \times n}$ 的一组基

11. 设 $\begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \quad \mu = \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ n \end{bmatrix}$

则
$$\mu = (x_1 \mu, \dots, x_1 \mu, \dots, x_m \mu, \dots, x_m \mu)^T$$
$$= (y_{1-1}, \dots, y_{1-n}, \dots, y_{m-1}, \dots, y_{m-n})^T$$

所以 $x_i \mu = y_{i-j} \quad \frac{x_i}{y_i} = \frac{j}{\mu} =$