

研究生数学教学系列 (工科类)

3

矩阵论

南京航空航天大学

戴华 编著

GM

科学出版社

研究生数学教学系列(工科类)

矩 阵 论

南京航空航天大学

戴 华 编著

科 学 出 版 社

2 0 0 1

内 容 简 介

本书较全面、系统地介绍了矩阵理论的基本理论、方法和某些应用。全书共分 10 章,分别介绍了线性空间与内积空间、线性映射与线性变换、 λ 矩阵与 Jordan 标准形、初等矩阵与矩阵因子分解、Hermite 矩阵与正定矩阵、范数理论与扰动分析、矩阵函数与矩阵值函数、广义逆矩阵与线性方程组、Kronecker 积与线性矩阵方程、非负矩阵与 M 矩阵等内容。本书内容丰富、论述严谨,各章后面配有一定数量的习题,有利于读者学习和巩固。

本书可作为理工院校硕士研究生和高年级本科生的教材,也可作为有关专业的教师和工程技术人员的参考书。

图书在版编目(CIP)数据

矩阵论/戴华编著. —北京:科学出版社,2001.8

(研究生数学教学系列·工科类)

ISBN 7-03-009673-8

I. 矩… II. 戴… III. 矩阵-理论-研究生-教材 IV. O151.21

中国版本图书馆 CIP 数据核字(2001)第 051680 号

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

新蕾印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2001 年 8 月第 一 版 开本: 720×1000 1/16

2001 年 8 月第一次印刷 印张: 18 3/4

印数: 1—5 000 字数: 334 000

定价: 28.00 元

(如有印装质量问题,我社负责调换〈新欣〉)

前 言

作为数学的一个重要分支,矩阵理论具有极为丰富的内容.作为一种基本的工具,矩阵理论在数学学科以及其他科学技术领域,如数值分析、最优化理论、概率统计、运筹学、控制理论、力学、电学、信息科学与技术、管理科学与工程等学科都有十分重要的应用.因此,学习和掌握矩阵的基本理论和方法,对于工科研究生来说是必不可少的.

本书的主要内容曾作为南京航空航天大学工科研究生的必修课教材讲授多年,经过不断地充实更新,并对内容的安排作了精心的处理.撰写时,力求使本书具有一定的理论深度,且注重广度适中、深入浅出、简洁易懂、便于自学.

全书共十章,较全面、系统地介绍了矩阵的基本理论、方法和某些应用.第一与第二章是线性代数的基础理论,主要介绍线性空间与内积空间、线性映射与线性变换等基本概念和性质.这部分内容的熟练掌握和深刻理解,对后面内容的学习有着很大的影响.第三至第五章是矩阵分解理论,主要介绍矩阵的 Jordan 标准形、初等矩阵与矩阵因子分解、Hermite 矩阵与正定矩阵.这些内容是矩阵理论研究、矩阵计算及其应用中不可缺少的工具和手段.第六至第八章介绍了范数理论、矩阵函数与矩阵值函数、广义逆矩阵及其应用.第九与第十章主要介绍了矩阵 Kronecker 积、线性矩阵方程与矩阵最佳逼近、非负矩阵等基本理论及其应用.每章配有一定数量的习题,以供读者练习.带 * 号的内容可用于选讲或自学.

作者感谢东南大学应用数学系陈建龙教授,他仔细审阅了全部书稿,并提出了不少有益的建议.感谢南京航空航天大学研究生院、教务处对本书的出版给予的大力支持.同时感谢熊春茹副编审和吕虹编审为本书的出版付出的辛勤劳动.

限于水平,书中不妥之处,敬请读者指正.

编 者

2000.12

目 录

第一章 线性空间与内积空间	1
§ 1.1 预备知识:集合、映射与数域	1
1.1.1 集合及其运算	1
1.1.2 二元关系与等价关系	2
1.1.3 映射	3
1.1.4 数域与代数运算	6
§ 1.2 线性空间	7
1.2.1 线性空间及其基本性质	7
1.2.2 向量的线性相关性	9
1.2.3 线性空间的维数	12
§ 1.3 基与坐标	13
§ 1.4 线性子空间	17
1.4.1 线性子空间的概念	17
1.4.2 子空间的交与和	20
1.4.3 子空间的直和	23
§ 1.5 线性空间的同构	24
§ 1.6 内积空间	27
1.6.1 内积空间及其基本性质	27
1.6.2 标准正交基与 Gram-Schmidt 正交化方法	32
1.6.3 正交补与投影定理	35
习 题	41
第二章 线性映射与线性变换	45
§ 2.1 线性映射及其矩阵表示	45
2.1.1 线性映射的定义及其性质	45
2.1.2 线性映射的运算	48
2.1.3 线性映射的矩阵表示	50
§ 2.2 线性映射的值域与核	55
§ 2.3 线性变换	57
§ 2.4 特征值和特征向量	60
§ 2.5 矩阵的相似对角形	68
§ 2.6 线性变换的不变子空间	70

§ 2.7 酉(正交)变换与酉(正交)矩阵	75
习 题	77
第三章 λ 矩阵与矩阵的 Jordan 标准形	82
§ 3.1 一元多项式	82
§ 3.2 λ 矩阵及其在相抵下的标准形	85
3.2.1 λ 矩阵的基本概念	85
3.2.2 λ 矩阵的初等变换与相抵	87
3.2.3 λ 矩阵在相抵下的标准形	89
§ 3.3 λ 矩阵的行列式因子和初等因子	93
§ 3.4 矩阵相似的条件	102
§ 3.5 矩阵的 Jordan 标准形	104
§ 3.6 Cayley-Hamilton 定理与最小多项式	110
习 题	114
第四章 矩阵的因子分解	117
§ 4.1 初等矩阵	117
4.1.1 初等矩阵	117
4.1.2 初等下三角矩阵	118
4.1.3 Householder 矩阵	119
§ 4.2 满秩分解	120
§ 4.3 三角分解	125
§ 4.4 QR 分解	130
§ 4.5 Schur 定理与正规矩阵	133
§ 4.6 奇异值分解	139
习 题	142
第五章 Hermite 矩阵与正定矩阵	146
§ 5.1 Hermite 矩阵与 Hermite 二次型	146
5.1.1 Hermite 矩阵	146
5.1.2 矩阵的惯性	147
5.1.3 Hermite 二次型	149
§ 5.2 Hermite 正定(非负定)矩阵	151
§ 5.3 矩阵不等式	159
* § 5.4 Hermite 矩阵的特征值	162
习 题	166
第六章 范数与极限	169
§ 6.1 向量范数	169
§ 6.2 矩阵范数	175

6.2.1 基本概念	175
6.2.2 相容矩阵范数	176
6.2.3 算子范数	177
§ 6.3 矩阵序列与矩阵级数	182
6.3.1 矩阵序列的极限	183
6.3.2 矩阵级数	185
§ 6.4 矩阵扰动分析	189
6.4.1 矩阵逆的扰动分析	190
6.4.2 线性方程组解的扰动分析	192
6.4.3 矩阵特征值的扰动分析*	193
习 题	198
第七章 矩阵函数与矩阵值函数	201
§ 7.1 矩阵函数	201
7.1.1 矩阵函数的幂级数表示	201
7.1.2 矩阵函数的另一种定义	206
§ 7.2 矩阵值函数	210
7.2.1 矩阵值函数	211
7.2.2 矩阵值函数的分析运算	212
§ 7.3 矩阵值函数在微分方程组中的应用	217
§ 7.4 特征对的灵敏度分析*	222
习 题	230
第八章 广义逆矩阵	233
§ 8.1 广义逆矩阵的概念	233
§ 8.2 广义逆矩阵 A^+ 与线性方程组的解	234
§ 8.3 极小范数广义逆 A_m 与线性方程组的极小范数解	236
§ 8.4 最小二乘广义逆 A_l 与矛盾方程组的最小二乘解	239
§ 8.5 广义逆矩阵 A^+ 与线性方程组的极小最小二乘解	242
习 题	246
第九章 Kronecker 积与线性矩阵方程	248
§ 9.1 矩阵的 Kronecker 积	248
§ 9.2 矩阵的拉直与线性矩阵方程	252
9.2.1 矩阵的拉直	252
9.2.2 线性矩阵方程	253
§ 9.3 矩阵方程 $AXB = C$ 与矩阵最佳逼近问题	254
9.3.1 矩阵方程 $AXB = C$	254
9.3.2 带约束的矩阵最佳逼近问题	256

§ 9.4 矩阵方程 $AX = B$ 的 Hermite 解与矩阵最佳逼近问题	258
§ 9.5 矩阵方程 $AX + XB = C$ 和 $X - AXB = C^*$	262
9.5.1 矩阵方程 $AX + XB = C$	262
9.5.2 矩阵方程 $X - AXB = C$	263
习 题	264
第十章 非负矩阵	266
§ 10.1 非负矩阵与正矩阵	266
§ 10.2 素矩阵与不可约非负矩阵	274
10.2.1 素矩阵	274
10.2.2 不可约非负矩阵	276
§ 10.3 随机矩阵	278
§ 10.4 M 矩阵	282
习 题	287
参考文献	289

第一章 线性空间与内积空间

本章概述线性空间与内积空间的基本概念和基本理论. 这些概念是通常几何空间概念的推广和抽象. 在近代数学发展中, 这些概念和理论已渗透到数学的各个分支. 本章内容是学习本书的基础.

§ 1.1 预备知识: 集合、映射与数域

1.1.1 集合及其运算

我们常常会遇到或处理一类对象或某种整体. 例如, 实数的全体 \mathbf{R} , 复数的全体 \mathbf{C} , 闭区间 $[a, b]$ 上定义的连续函数全体 $C[a, b]$, 等. 所涉及的这些整体性对象, 在数学上用集合(简称集)这一抽象概念概括.

集合是近代数学的最基本概念之一, 它是由具有某种性质所确定的事物的总体. 根据这种性质可以辨别任一事物属于或不属于这个集合. 属于这个集合的事物称为这个集合的元素. 通常用大写字母 A, B, C, \dots 表示集合, 而用小写字母 a, b, c, \dots 表示集合的元素. 若 a 为集合 A 的元素, 则称 a 属于 A , 记为 $a \in A$; 若 a 不是集合 A 的元素, 则称 a 不属于 A , 记为 $a \notin A$.

表示一个集合通常有两种方法. 一种是列举法, 即把一个集合的元素都列举出来. 例如, 集合 A 由元素 a_1, a_2, a_3 组成, 则记 $A = \{a_1, a_2, a_3\}$; 另一种是概括法, 即把这个集合的元素所具有的特征性质表示出来. 具有这种特征的元素在这个集合中, 而不具有这种特征的元素就不在这个集合中. 如果这种特征可以用一个关于元素 x 的命题 $P(x)$ ($P(x)$ 是对元素 x 的说明, 它可以是说明语句, 也可以是数学表达式) 表示, 则这个集合 A 表示为 $A = \{x | P(x)\}$. 当 x 具有这种特征时, 命题 $P(x)$ 为真; x 不具有这种特征时, 命题 $P(x)$ 为假, 则 A 是具有性质 $P(x)$ 的元素 x 所成的集合. 例如, 适合方程 $x^2 + y^2 = r^2$ 的全部点 (x, y) 的集合 A 可表示为 $A = \{(x, y) | x^2 + y^2 = r^2\}$.

设 A, B 是两个集合, 如果集合 A 的元素都是集合 B 的元素, 则称 A 为 B 的子集, 或称 B 包含 A , 记为 $B \supseteq A$ 或 $A \subseteq B$. 如果 $A \subseteq B$ 且 $B \subseteq A$, 则称集合 A 与 B 相等, 记为 $A = B$.

含有有限个元素的集合称为有限集; 否则称为无限集. 不含任何元素的集合称为空集, 记为 \emptyset . 为了方便, 我们规定空集是任意集合的子集.

定义 1.1.1 设 A, B 是两个集合, 由属于 A 或者属于 B 的所有元素组成的集合称为 A 与 B 的并集, 记为 $A \cup B$, 即 $A \cup B = \{x | x \in A \text{ 或 } x \in B\}$;

由既属于 A 又属于 B 的所有元素作成的集合称为 A 与 B 的交集, 记为 $A \cap B$, 即 $A \cap B = \{x | x \in A \text{ 且 } x \in B\}$.

例如, 设 $A = \{1, 2, 3\}$, $B = \{2, 3, 4, 5\}$, 则

$$A \cup B = \{1, 2, 3, 4, 5\}, A \cap B = \{2, 3\}$$

由集合的交与并运算的定义, 显然有

$$A \subseteq A \cup B, B \subseteq A \cup B, A \cup A = A, A \cup \emptyset = A$$

$$A \cap B \subseteq A, A \cap B \subseteq B, A \cap A = A, A \cap \emptyset = \emptyset$$

并且集合的交与并运算满足以下运算规律.

定理 1.1.1 设 A, B, C 是 3 个集合, 则

- (1) $A \cup B = B \cup A, A \cap B = B \cap A$;
- (2) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C), A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$;
- (3) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C), A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

证明 这里我们仅证明(3)中第二个等式, 其余证明留给读者.

如果 $x \in A \cap (B \cup C)$, 则 $x \in A$ 且 $x \in B \cup C$, 即 $x \in A$ 且 $x \in B$ 或 $x \in C$. 于是有 $x \in A \cap B$ 或 $x \in A \cap C$, 即 $x \in (A \cap B) \cup (A \cap C)$. 从而有

$$A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$$

另一方面, 如果 $x \in (A \cap B) \cup (A \cap C)$, 则 $x \in A \cap B$ 或 $x \in A \cap C$, 即 $x \in A$ 且 $x \in B$ 或 $x \in A$ 且 $x \in C$. 于是有 $x \in A$ 且 $x \in B \cup C$, 即 $x \in A \cap (B \cup C)$. 从而有

$$(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$$

因此

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad \square$$

1.1.2 二元关系与等价关系

定义 1.1.2 设 A, B 是两个非空集合, 元素对的集合 $\{(a, b) | a \in A, b \in B\}$ 称为 A 与 B 的 **Descartes 积**, 记作 $A \times B$, 即 $A \times B = \{(a, b) | a \in A, b \in B\}$.

例如, 设 $A = \{1, 2, 3\}$, $B = \{a, b\}$, 则

$$A \times B = \{(1, a), (1, b), (2, a), (2, b), (3, a), (3, b)\}$$

定义 1.1.3 设 A, B 是两个集合, $A \times B$ 的子集 R 称为 $A \times B$ 中的一个二元关系, 即对任意 $a \in A, b \in B$, 如果 $(a, b) \in R$, 则称 a 与 b 有关系 R , 记

为 aRb . 特别地, $A \times A$ 中的二元关系简称为 A 上的二元关系.

例如, 设 $A = \{\text{一群学生}\}$, 两学生 $a, b \in A$, 如果 a 和 b 年龄相同, 就记为 aRb , 则 R 是 A 上的一个二元关系.

定义 1.1.4 如果集合 A 上的一个二元关系 R 满足

- (1) 自反性: 对任意 $a \in A$, 有 aRa ;
- (2) 对称性: 对任意 $a, b \in A$, 如果 aRb , 则 bRa ;
- (3) 传递性: 对任意 $a, b, c \in A$, 如果 aRb, bRc , 则 aRc ,

则称 R 是 A 上的一个等价关系.

例如, 实数集 \mathbf{R} 上的相等“=”是 \mathbf{R} 上的等价关系.

定义 1.1.5 设 R 是 A 上的一个等价关系, $a \in A$, 称 $[a] = \{x \mid x \in A, xRa\}$ 为 a 关于 R 的等价类. A 的所有元素关于 R 的等价类集合 $A/R = \{[a] \mid a \in A\}$ 称为 A 关于 R 的商集.

定义 1.1.6 设每个 $B_i (i \in I)$ 都是集合 A 的非空子集, 如果 $A = \bigcup_{i \in I} B_i$, 并且对任意 $i, j \in I$, 当 $i \neq j$ 时有 $B_i \cap B_j = \emptyset$, 则称 $\{B_i\}$ 是 A 的一个分类.

如下定理建立了集合的分类与等价关系之间的联系.

定理 1.1.2 (1) 集合 A 上的每个等价关系 R 都决定 A 的一个分类.

(2) 集合 A 的每个分类都决定 A 上的一个等价关系.

证明 (1) 如果 R 是 A 上的等价关系, 则 A/R 给出了 A 的一个分类.

(2) 如果 $\{B_i\}$ 是 A 的一个分类, 令

$$R = \{(x, y) \mid \text{存在 } B_i, \text{使得 } x \in B_i, y \in B_i\}$$

则 R 是 A 上的一个等价关系.

1.1.3 映射

映射是函数概念的推广, 它描述了两个集合的元素之间的关系, 是数学中最基本的概念和工具.

定义 1.1.7 设 A, B 是两个非空集合, 如果存在一个 A 到 B 的对应法则 f , 使得对 A 中的每一个元素 x 都有 B 中惟一的一个元素 y 与之对应, 则称 f 是 A 到 B 的一个映射, 记为 $y = f(x)$. 元素 $y \in B$ 称为元素 $x \in A$ 在映射 f 下的像, 称 x 为 y 的原像. 集合 A 称为映射 f 的定义域. 当 A 中元素 x 改变时, x 在映射 f 下的像的全体作成 B 的一个子集, 称为映射 f 的值域, 记为 $R(f)$.

通常用记号

$$f: A \rightarrow B$$

抽象地表示 f 是 A 到 B 的一个映射. 而用记号

$$f: x \rightarrow f(x)$$

表示映射 f 所规定的元素之间的具体对应关系.

例 1.1.1 设 $A = \{a, b, c\}$, $B = \{1, 2, 3, 4\}$. 规定

$$f: a \rightarrow 1$$

$$b \rightarrow 2$$

$$c \rightarrow 3$$

则 f 是 A 到 B 的一个映射.

例 1.1.2 设 A 是整数集, B 是自然数集. 规定

$$f: n \rightarrow |n| + 1, \quad n \in A$$

则 f 是 A 到 B 的一个映射.

例 1.1.3 设 A 是一个集合. 规定

$$I_A: a \rightarrow a, \quad a \in A$$

则 I_A 是 A 到自身的一个映射. 这种映射称为集合 A 上的恒等映射或单位映射.

定义 1.1.8 设 f 是集合 A 到集合 B 的一个映射, 如果对任意 $a, b \in A$, 当 $a \neq b$ 时有 $f(a) \neq f(b)$, 则称 f 是 A 到 B 内的一一映射或称 f 是 A 到 B 的单映射; 如果对任意 $b \in B$ 都有一个 $a \in A$ 使得 $f(a) = b$ (即 $R(f) = B$), 则称 f 是 A 到 B 上的映射或称 f 是 A 到 B 的满映射; 如果映射 f 既是单映射又是满映射, 则称 f 是 A 到 B 上的一一映射或称 f 是 A 到 B 的双映射.

例 1.1.1 中的映射是一个单映射, 例 1.1.2 中的映射是一个满映射, 而例 1.1.3 中的映射是一个双映射.

设 f_1 是集合 A_1 到 B_1 的一个映射, f_2 是集合 A_2 到 B_2 的一个映射, 如果 $A_1 = A_2$, $B_1 = B_2$, 并且对任意 $a \in A_1$ 有 $f_1(a) = f_2(a)$, 则称映射 f_1 和 f_2 相等, 记为 $f_1 = f_2$.

定义 1.1.9 设 A, B, C 是三个非空集合, 并设有两个映射 $f_1: A \rightarrow B$, $f_2: B \rightarrow C$, 由 f_1, f_2 确定的 A 到 C 的映射 $f_3: a \rightarrow f_2(f_1(a)), a \in A$, 称为映射 f_1 和 f_2 的乘积, 记为 $f_3 = f_2 \cdot f_1$.

一般说来, 映射 f_1 和 f_2 的乘积不具有交换律, 即 $f_2 \cdot f_1 \neq f_1 \cdot f_2$. 事实上, 当 $A \neq C$ 时, 映射 $f_1 \cdot f_2$ 未必有意义. 即使 $A = B = C$, 映射 $f_2 \cdot f_1$ 与 $f_1 \cdot f_2$ 都有意义, 但 $f_2 \cdot f_1$ 和 $f_1 \cdot f_2$ 也未必相等. 例如, 当 $A = B = C = \{0, 1\}$, 定义映射 f_1, f_2 如下

$$\begin{cases} f_1(0) = 1, \\ f_1(1) = 0, \end{cases} \quad \begin{cases} f_2(0) = 0 \\ f_2(1) = 0 \end{cases}$$

则容易验证 $f_2 \cdot f_1 \neq f_1 \cdot f_2$. 映射的乘积适合结合律.

定理 1.1.3 设有映射 $f_1: A \rightarrow B$, $f_2: B \rightarrow C$, $f_3: C \rightarrow D$, 则有

$$(1) f_3 \cdot (f_2 \cdot f_1) = (f_3 \cdot f_2) \cdot f_1;$$

$$(2) f_1 \cdot I_A = I_B \cdot f_1 = f_1.$$

证明 显然, $f_3 \cdot (f_2 \cdot f_1)$ 与 $(f_3 \cdot f_2) \cdot f_1$ 都是 A 到 D 的映射. 对任意 $a \in A$, 有

$$[f_3 \cdot (f_2 \cdot f_1)](a) = f_3[(f_2 \cdot f_1)(a)] = f_3[f_2(f_1(a))]$$

$$[(f_3 \cdot f_2) \cdot f_1](a) = (f_3 \cdot f_2)(f_1(a)) = f_3[f_2(f_1(a))]$$

因此 $f_3 \cdot (f_2 \cdot f_1) = (f_3 \cdot f_2) \cdot f_1$. 类似可证(2). \square

定义 1.1.10 设有映射 $f: A \rightarrow B$, 如果存在映射 $g: B \rightarrow A$ 使得

$$g \cdot f = I_A, \quad f \cdot g = I_B$$

其中 I_A, I_B 分别是 A 与 B 上的恒等映射, 则称 g 为 f 的逆映射, 记为 f^{-1} . 如果映射 f 有逆映射 f^{-1} , 则称 f 为可逆映射.

定理 1.1.4 设映射 $f: A \rightarrow B$ 是可逆的, 则 f 的逆映射 f^{-1} 是惟一的.

证明 设映射 $g: B \rightarrow A$ 和 $h: B \rightarrow A$ 均为 f 的逆映射, 则 $g \cdot f = I_A$, $f \cdot g = I_B$, $h \cdot f = I_A$, $f \cdot h = I_B$. 于是由定理 1.1.3 有

$$h = h \cdot I_B = h \cdot (f \cdot g) = (h \cdot f) \cdot g = I_A \cdot g = g \quad \square$$

定理 1.1.5 映射 $f: A \rightarrow B$ 是可逆映射的充分必要条件是 f 为 A 到 B 上的一一映射.

证明 设 $f: A \rightarrow B$ 是可逆映射, 则存在 f 的逆映射 $f^{-1}: B \rightarrow A$. 对任意 $a_1, a_2 \in A$, 如果 $f(a_1) = f(a_2)$, 则有

$$(f^{-1} \cdot f)(a_1) = (f^{-1} \cdot f)(a_2)$$

从而 $a_1 = a_2$. 因此 f 是 A 到 B 内的一一映射.

对任意 $b \in B$, 若 $f^{-1}(b) = a$, 则 $f(a) = f(f^{-1}(b)) = (f \cdot f^{-1})(b) = b$. 因此 f 是 A 到 B 的一个满映射. 于是, f 是 A 到 B 上的一一映射.

反过来, 若 f 是 A 到 B 上的一一映射, 则对任意 $b \in B$, 存在 $a \in A$ 使得 $f(a) = b$. 定义映射 g 为

$$g: B \rightarrow A, g(b) = a$$

则

$$(f \cdot g)(b) = f(g(b)) = f(a) = b$$

即 $f \cdot g = I_B$, 并且对任意 $a \in A$,

$$(g \cdot f)(a) = g(f(a)) = g(b) = a$$

即 $g \cdot f = I_A$. 由定义 1.1.10 知, $g = f^{-1}$ 为 f 的逆映射. \square

定义 1.1.11 设 A 是一个非空集合, A 到自身的映射称为 A 的变换; A 到自身的双映射称为 A 的一一变换; 如果 A 是有限集, A 的一一变换称为 A 的置换.

1.1.4 数域与代数运算

讨论数学问题时, 经常涉及到某一个数集中数的运算规则及其性质. 为了准确地分析和研究数学问题, 引进数域的概念.

定义 1.1.12 设 P 是包含 0 和 1 在内的数集, 如果 P 中任意两个数的和、差、积、商(除数不为 0)仍是 P 中的数, 则称 P 为一个数域.

显然, 有理数集 \mathbf{Q} 、实数集 \mathbf{R} 和复数集 \mathbf{C} 都是数域, 分别称为有理数域、实数域和复数域. 但是, 数域不限于有理数域 \mathbf{Q} 、实数域 \mathbf{R} 和复数域 \mathbf{C} . 例如, 数集 $\mathbf{Q}(\sqrt{2}) = \{a + b\sqrt{2} \mid a, b \in \mathbf{Q}\}$ 也是一个数域.

我们经常要研究带有运算的集合, 为此, 这里介绍代数运算的概念.

定义 1.1.13 设 A, B, C 是三个非空集合, $A \times B$ 到 C 的映射称为 A 与 B 到 C 的一个代数运算. 特别地, $A \times A$ 到 C 的映射称为 A 到 C 的代数运算; $A \times A$ 到 A 的映射称为 A 的代数运算或 A 的二元运算, 也称集合 A 对代数运算是封闭的.

由定义可知, 一个代数运算是一个特殊的映射. 如果有 A 与 B 到 C 的一个代数运算, 记为 “ \circ ”, 则由定义, 对任意 $a \in A, b \in B$, 经过代数运算 \circ , 得惟一的 $c \in C$, 即

$$\circ: (a, b) \rightarrow c$$

记为 $c = a \circ b$.

例如, 在有理数集 \mathbf{Q} 中, 普通的加法与乘法都是 \mathbf{Q} 的代数运算.

例 1.1.4 设 \mathbf{R}^n 是实数域 \mathbf{R} 上的一个 n 维向量空间, 对 $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$,

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \in \mathbf{R}^n, \text{ 则}$$

$$\circ: (x, y) \rightarrow x \circ y = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

是 \mathbf{R}^n 的一个代数运算, 通常称为 \mathbf{R}^n 上的加法运算, 记为“+”. 对 $k \in \mathbf{R}, x \in \mathbf{R}^n$, 则

$$\ast: (k, x) \rightarrow k \ast x = \begin{bmatrix} kx_1 \\ kx_2 \\ \vdots \\ kx_n \end{bmatrix}$$

是 $\mathbf{R} \times \mathbf{R}^n$ 到 \mathbf{R}^n 的一个代数运算, 通常称为 \mathbf{R}^n 上的数乘运算, 记为“ \cdot ”.

§ 1.2 线性空间

1.2.1 线性空间及其基本性质

线性空间是线性代数最基本的概念之一, 它是向量空间的自然推广.

定义 1.2.1 设 V 是一个非空集合, P 是一个数域. 在 V 上定义了一种代数运算, 称为加法, 记为“+”; 定义了 P 与 V 到 V 的一种代数运算, 称为数量乘法 (简称数乘), 记为“ \cdot ”. 如果加法与数量乘法满足如下规则:

- (1) $\alpha + \beta = \beta + \alpha$;
- (2) $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$;
- (3) 在 V 中有一个元素 0 (称为零元素), 对于 V 中任一元素 α 都有 $\alpha + 0 = \alpha$;
- (4) 对 V 中任一元素 α , 都有 V 中的元素 β 使得 $\alpha + \beta = 0$ (β 称为 α 的负元素, 记为 $-\alpha$);
- (5) $1 \cdot \alpha = \alpha$;
- (6) $k \cdot (m \cdot \alpha) = (km) \cdot \alpha$;
- (7) $(k + m) \cdot \alpha = k \cdot \alpha + m \cdot \alpha$;

$$(8) k \cdot (\alpha + \beta) = k \cdot \alpha + k \cdot \beta,$$

其中 k, m 是 \mathbf{P} 中的任意数, α, β, γ 是 V 中的任意元素, 则称 V 为数域 \mathbf{P} 上的线性空间.

线性空间中的元素也称为向量, 这里所指的向量比线性代数中 n 维向量的含义更为广泛. 线性空间有时也称为向量空间.

例 1.2.1 分量属于数域 \mathbf{P} 的全体 n 维向量作成的集合, 按 n 维向量的加法和数与 n 维向量的数量乘法构成数域 \mathbf{P} 上的线性空间, 记为 \mathbf{P}^n .

例 1.2.2 元素属于数域 \mathbf{P} 的全体 $m \times n$ 矩阵作成的集合, 按矩阵的加法和数与矩阵的数量乘法构成数域 \mathbf{P} 上的线性空间, 记为 $\mathbf{P}^{m \times n}$.

例 1.2.3 数域 \mathbf{P} 上一元多项式的全体作成的集合 $\mathbf{P}[x]$, 按通常的多项式加法和数与多项式的乘法构成数域 \mathbf{P} 上的线性空间. 数域 \mathbf{P} 上次数小于 n 的一元多项式再添上零多项式也构成数域 \mathbf{P} 上的线性空间, 记为 $\mathbf{P}[x]_n$.

例 1.2.4 区间 $[a, b]$ 上全体连续实函数作成的集合, 按函数的加法和数与函数的数量乘法构成实数域 \mathbf{R} 上的线性空间, 记为 $\mathbf{C}[a, b]$.

直接由定义 1.2.1 可以证明线性空间的一些基本性质.

定理 1.2.1 设 V 是数域 \mathbf{P} 上的线性空间, 则

- (1) V 中零元素是惟一的;
- (2) V 中任一元素 α 的负元素是惟一的;
- (3) $0 \cdot \alpha = 0, k \cdot 0 = 0, (-1) \cdot \alpha = -\alpha$;
- (4) 如果 $k \cdot \alpha = 0$, 那么 $k = 0$ 或者 $\alpha = 0$.

证明 这里仅证明(2), 其余的证明留给读者去完成.

假设 α 有两个负元素 β 与 γ , 则

$$\alpha + \beta = 0, \alpha + \gamma = 0$$

从而

$$\beta = \beta + 0 = \beta + (\alpha + \gamma) = (\beta + \alpha) + \gamma = 0 + \gamma = \gamma \quad \square$$

利用负元素, 我们定义线性空间中的减法

$$\alpha - \beta = \alpha + (-\beta)$$

为了书写方便, 数量乘法“ \cdot ”也按普通乘积的表示法直接连写.

现在我们给出线性空间中凸集的概念.

定义 1.2.2 设 V 是数域 \mathbf{P} 上的线性空间, W 是 V 的一个非空子集, 如果对 W 中任意两个向量 α, β 以及任意 $0 \leq \lambda \leq 1$, 都有 $\lambda\alpha + (1-\lambda)\beta \in W$, 则称 W 是凸集.

1.2.2 向量的线性相关性

定义 1.2.3 设 V 是数域 P 上的线性空间, $\alpha_1, \alpha_2, \dots, \alpha_r (r \geq 1)$ 是 V 中一组向量, k_1, k_2, \dots, k_r 是数域 P 中的数. 如果 V 中向量 α 可以表示为

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

则称 α 可由 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性表示, 或称 α 是 $\alpha_1, \alpha_2, \dots, \alpha_r$ 的线性组合.

定义 1.2.4 设 $\alpha_1, \alpha_2, \dots, \alpha_r$ 与 $\beta_1, \beta_2, \dots, \beta_s$ 是线性空间 V 中两个向量组, 如果 $\alpha_1, \alpha_2, \dots, \alpha_r$ 中每个向量都可由向量组 $\beta_1, \beta_2, \dots, \beta_s$ 线性表示, 则称向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 可以由向量组 $\beta_1, \beta_2, \dots, \beta_s$ 线性表示. 如果向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 与向量组 $\beta_1, \beta_2, \dots, \beta_s$ 可以互相线性表示, 则称向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 与向量组 $\beta_1, \beta_2, \dots, \beta_s$ 是等价的.

容易证明向量组之间的等价具有如下性质:

- (1) 自反性: 每一个向量组都与它自身等价;
- (2) 对称性: 如果向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 与 $\beta_1, \beta_2, \dots, \beta_s$ 等价, 那么向量组 $\beta_1, \beta_2, \dots, \beta_s$ 也与 $\alpha_1, \alpha_2, \dots, \alpha_r$ 等价;
- (3) 传递性: 若向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 与 $\beta_1, \beta_2, \dots, \beta_s$ 等价, 且向量组 $\beta_1, \beta_2, \dots, \beta_s$ 与 $\gamma_1, \gamma_2, \dots, \gamma_t$ 等价, 则向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 与 $\gamma_1, \gamma_2, \dots, \gamma_t$ 等价.

定义 1.2.5 设 V 是数域 P 上的线性空间, $\alpha_1, \alpha_2, \dots, \alpha_r (r \geq 1)$ 是 V 中一组向量, 如果存在 r 个不全为零的数 $k_1, k_2, \dots, k_r \in P$, 使得

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

则称 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性相关; 如果向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 不线性相关, 就称为线性无关.

由定义 1.2.5 可得向量组线性相关定义的另一个说法.

定理 1.2.2 设 V 为数域 P 上的线性空间, V 中一个向量 α 线性相关的充分必要条件是 $\alpha = 0$; V 中一组向量 $\alpha_1, \alpha_2, \dots, \alpha_r (r \geq 2)$ 线性相关的充分必要条件是其中有一个向量是其余向量的线性组合.

证明 如果一个向量 α 线性相关, 由定义 1.2.5 可知, 有 $k \neq 0$, 使

$$k\alpha = 0$$

由定理 1.2.1(4) 知 $\alpha = 0$.

反之, 若 $\alpha = 0$, 则对任意数 $k \neq 0$ 都有 $k\alpha = 0$. 由定义 1.2.5 知, 向量 α 线性相关.

如果向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性相关, 则存在不全为零的数 k_1, k_2, \dots, k_r , 使得

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

因为 k_1, k_2, \dots, k_r 不全为零, 不妨设 $k_r \neq 0$, 于是上式可改写为

$$\alpha_r = -\frac{k_1}{k_r}\alpha_1 - \frac{k_2}{k_r}\alpha_2 - \cdots - \frac{k_{r-1}}{k_r}\alpha_{r-1}$$

即向量 α_r 是其余向量 $\alpha_1, \alpha_2, \dots, \alpha_{r-1}$ 的线性组合.

反过来, 如果向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 中有一个向量是其余向量的线性组合, 譬如说

$$\alpha_r = l_1\alpha_1 + l_2\alpha_2 + \cdots + l_{r-1}\alpha_{r-1}$$

上式可改写为

$$l_1\alpha_1 + l_2\alpha_2 + \cdots + l_{r-1}\alpha_{r-1} + (-1)\alpha_r = 0$$

因为 $l_1, \dots, l_{r-1}, -1$ 不全为零, 由定义 1.2.5 知, 向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性相关. \square

例 1.2.5 实数域 \mathbf{R} 上线性空间 $\mathbf{R}^{2 \times 2}$ 的一组向量(矩阵)

$$\mathbf{E}_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \mathbf{E}_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \mathbf{E}_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \mathbf{E}_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

是线性无关的.

事实上, 如果

$$k_1\mathbf{E}_{11} + k_2\mathbf{E}_{12} + k_3\mathbf{E}_{21} + k_4\mathbf{E}_{22} = 0$$

即

$$\begin{pmatrix} k_1 & k_2 \\ k_3 & k_4 \end{pmatrix} = 0$$

则 $k_1 = k_2 = k_3 = k_4 = 0$. 因此, 满足 $k_1\mathbf{E}_{11} + k_2\mathbf{E}_{12} + k_3\mathbf{E}_{21} + k_4\mathbf{E}_{22} = 0$ 的 k_1, k_2, k_3, k_4 只能全为零. 于是 $\mathbf{E}_{11}, \mathbf{E}_{12}, \mathbf{E}_{21}, \mathbf{E}_{22}$ 线性无关.

容易证明: 线性空间 $\mathbf{R}^{2 \times 2}$ 的一组向量(矩阵) $\alpha_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \alpha_2 =$

$$\begin{pmatrix} -1 & 0 \\ 1 & 1 \end{pmatrix}, \alpha_3 = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \text{ 线性相关.}$$

定理 1.2.3 设 V 为数域 \mathbf{P} 上的线性空间, 如果 V 中向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性无关, 并且可由向量组 $\beta_1, \beta_2, \dots, \beta_s$ 线性表示, 则 $r \leq s$.

证明 采用反证法. 假设 $r > s$, 因为向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 可由向量组

$\beta_1, \beta_2, \dots, \beta_s$ 线性表示, 即

$$\alpha_i = \sum_{j=1}^s a_{ji} \beta_j, \quad i = 1, \dots, r$$

作线性组合

$$x_1 \alpha_1 + x_2 \alpha_2 + \dots + x_r \alpha_r = \sum_{i=1}^r x_i \sum_{j=1}^s a_{ji} \beta_j = \sum_{j=1}^s \left(\sum_{i=1}^r a_{ji} x_i \right) \beta_j$$

考虑齐次线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1r}x_r = 0 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2r}x_r = 0 \\ \dots\dots\dots \\ a_{s1}x_1 + a_{s2}x_2 + \dots + a_{sr}x_r = 0 \end{cases}$$

因为上述齐次线性方程组未知数 x_1, x_2, \dots, x_r 的个数 r 大于方程的个数 s , 从而有非零解 x_1, x_2, \dots, x_r , 即我们可找到不全为零的数 x_1, x_2, \dots, x_r 使得

$$x_1 \alpha_1 + x_2 \alpha_2 + \dots + x_r \alpha_r = 0$$

因此, 向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性相关. 这与 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性无关矛盾, 于是 $r \leq s$. \square

由定理 1.2.3 直接可得如下结论.

推论 1.2.1 两个等价的线性无关向量组必含有相同个数的向量.

定理 1.2.4 设线性空间 V 中向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性无关, 而向量组 $\alpha_1, \alpha_2, \dots, \alpha_r, \beta$ 线性相关, 则 β 可由 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性表示, 并且表示法是惟一的.

证明 因为向量组 $\alpha_1, \alpha_2, \dots, \alpha_r, \beta$ 线性相关, 则存在不全为零的数 $k_1, k_2, \dots, k_r, k_{r+1}$ 使

$$k_1 \alpha_1 + k_2 \alpha_2 + \dots + k_r \alpha_r + k_{r+1} \beta = 0$$

并且 $k_{r+1} \neq 0$; 否则向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性相关, 这与条件矛盾. 从而

$$\beta = -\frac{k_1}{k_{r+1}} \alpha_1 - \frac{k_2}{k_{r+1}} \alpha_2 - \dots - \frac{k_r}{k_{r+1}} \alpha_r$$

即 β 可由 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性表示.

假设 β 可由 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性表示为

$$\beta = k_1 \alpha_1 + k_2 \alpha_2 + \dots + k_r \alpha_r = l_1 \alpha_1 + l_2 \alpha_2 + \dots + l_r \alpha_r$$

则

$$(k_1 - l_1)\alpha_1 + (k_2 - l_2)\alpha_2 + \cdots + (k_r - l_r)\alpha_r = 0$$

因为向量组 $\alpha_1, \alpha_2, \cdots, \alpha_r$ 线性无关, 从而 $k_i - l_i = 0 (i = 1, \cdots, r)$. 因此, β 可惟一地表示为 $\alpha_1, \alpha_2, \cdots, \alpha_r$ 的线性组合: \square

定义 1.2.6 设 $\alpha_1, \alpha_2, \cdots, \alpha_s$ 是线性空间 V 中一组向量, 如果 $\alpha_1, \alpha_2, \cdots, \alpha_s$ 中存在 r 个线性无关的向量 $\alpha_{i_1}, \alpha_{i_2}, \cdots, \alpha_{i_r} (1 \leq i_j \leq s, j = 1, \cdots, r)$, 并且 $\alpha_1, \alpha_2, \cdots, \alpha_s$ 中任一向量都可由向量组 $\alpha_{i_1}, \alpha_{i_2}, \cdots, \alpha_{i_r}$ 线性表示, 则称向量组 $\alpha_{i_1}, \alpha_{i_2}, \cdots, \alpha_{i_r}$ 为向量组 $\alpha_1, \alpha_2, \cdots, \alpha_s$ 的极大线性无关组, 数 r 称为向量组 $\alpha_1, \alpha_2, \cdots, \alpha_s$ 的秩, 记为 $\text{rank}\{\alpha_1, \alpha_2, \cdots, \alpha_s\} = r$.

一般说来, 向量组的极大线性无关组不惟一, 但是每一个极大线性无关组都与向量组本身等价. 由等价的传递性可知, 一个向量组的任意两个极大线性无关组都是等价的, 并且任意两个等价向量组的极大线性无关组也等价. 由推论 1.2.1 知, 一个向量组的极大线性无关组都含有相同个数的向量, 即向量组的秩是惟一的, 并且等价的向量组具有相同的秩.

1.2.3 线性空间的维数

现在引入线性空间的维数概念, 它是线性空间的一个重要属性.

定义 1.2.7 如果线性空间 V 中有 n 个线性无关的向量, 但是没有更多数目的线性无关向量, 则称 V 是 n 维的, 记为 $\dim(V) = n$; 如果在 V 中可以找到任意多个线性无关的向量, 则称 V 是无限维的.

按照这个定义, 数域 \mathbf{P} 上的线性空间 \mathbf{P}^n 是 n 维的. 数域 \mathbf{P} 上一元多项式的全体作成的线性空间 $\mathbf{P}[x]$ 是无限维的, 因为对于任意的正整数 N , $\mathbf{P}[x]$ 中都有 N 个线性无关的向量 $1, x, \cdots, x^{N-1}$.

无限维线性空间与有限维线性空间有比较大的差别. 本书主要讨论有限维线性空间.

定理 1.2.5 如果在线性空间 V 中有 n 个线性无关的向量 $\alpha_1, \alpha_2, \cdots, \alpha_n$, 并且 V 中任一向量都可由 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 线性表示, 则 $\dim(V) = n$.

证明 因为 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 线性无关, 所以 V 至少是 n 维的. 为了证明 $\dim(V) = n$, 只需证明 V 中任意 $n+1$ 个向量必线性相关即可.

设 $\beta_1, \beta_2, \cdots, \beta_{n+1}$ 是 V 中任意 $n+1$ 个向量, 则它们可由 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 线性表示. 假如 $\beta_1, \beta_2, \cdots, \beta_{n+1}$ 线性无关, 由定理 1.2.3 知 $n+1 \leq n$, 于是得出矛盾. \square

例 1.2.6 数域 \mathbf{P} 上线性空间 $\mathbf{P}^{m \times n}$ 的一组向量(矩阵) $\{E_{ij}\}$,

$$E_{ij} = \begin{bmatrix} & & & 0 & & & \\ & & & \vdots & & & \\ & & & 0 & & & \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ & & & 0 & & & \\ & & & \vdots & & & \\ & & & 0 & & & \end{bmatrix} \begin{matrix} \text{第 } i \text{ 行,} \\ \\ \\ i = 1, \cdots, m; j = 1, \cdots, n \\ \\ \\ \text{第 } j \text{ 列} \end{matrix}$$

是线性无关的,并且 $\mathbf{P}^{m \times n}$ 中任一矩阵 $\mathbf{A} = (\alpha_{ij})$ 可表示为

$$\mathbf{A} = \sum_{i=1}^m \sum_{j=1}^n \alpha_{ij} \mathbf{E}_{ij}$$

由定理 1.2.5 知 $\dim(\mathbf{P}^{m \times n}) = mn$.

§ 1.3 基与坐标

在解析几何中,为了研究向量的性质,引入坐标是一个重要的步骤. 在有限维线性空间中,坐标同样是一个有力的工具.

定义 1.3.1 设 V 为数域 \mathbf{P} 上的 n 维线性空间, V 中 n 个线性无关的向量 $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ 称为 V 的一组基. 设 α 是 V 中任一向量,则 α 可惟一地表示为基 $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ 的线性组合

$$\alpha = x_1 \epsilon_1 + x_2 \epsilon_2 + \cdots + x_n \epsilon_n \quad (1.3.1)$$

其中系数 x_1, x_2, \cdots, x_n 称为 α 在基 $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ 下的坐标,记为 (x_1, x_2, \cdots, x_n) 或 $(x_1, x_2, \cdots, x_n)^\top$.

经常将(1.3.1)改写为

$$\alpha = x_1 \epsilon_1 + x_2 \epsilon_2 + \cdots + x_n \epsilon_n = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n) \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (1.3.2)$$

其中 $(\epsilon_1, \epsilon_2, \cdots, \epsilon_n)$ 可视为 $1 \times n$ 分块矩阵,其元素 $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ 不是数字而是向量.

例如,在例 1.2.6 中向量(矩阵)组 $\{E_{ij} \mid (i=1, \cdots, m, j=1, \cdots, n)\}$ 是线性空间 $\mathbf{P}^{m \times n}$ 的一组基.

定理 1.3.1 在 n 维线性空间 V 中,任意一个线性无关的向量组 $\alpha_1, \alpha_2, \cdots, \alpha_r$ 都可以扩充成 V 的一组基.

证明 如果 $r=n$, 则 $\alpha_1, \alpha_2, \cdots, \alpha_r$ 是 V 的一组基. 下面设 $r < n$. 此时 V 中必有一个向量 β_1 不能由 $\alpha_1, \alpha_2, \cdots, \alpha_r$ 线性表示; 否则由定理 1.2.5 有 $\dim(V)=r$, 这与 $\dim(V)=n$ 矛盾. 因此, $\alpha_1, \alpha_2, \cdots, \alpha_r, \beta_1$ 线性无关. 如果 $r+1=n$, 则 $\alpha_1, \alpha_2, \cdots, \alpha_r, \beta_1$ 是 V 的一组基. 如果 $r+1 < n$, 则 V 中必有一个向量 β_2 不能由 $\alpha_1, \alpha_2, \cdots, \alpha_r, \beta_1$ 线性表示, 从而 $\alpha_1, \alpha_2, \cdots, \alpha_r, \beta_1, \beta_2$ 线性无关. 依次下去, 可得线性无关的向量组

$$\alpha_1, \alpha_2, \cdots, \alpha_r, \beta_1, \beta_2, \cdots, \beta_p$$

其中 $r+p=n$. 这个向量组就是 V 的一组基. \square

例 1.3.1 在 n 维线性空间 \mathbf{P}^n 中,

$$\epsilon_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \epsilon_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \cdots, \epsilon_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

是 \mathbf{P}^n 的一组基. 对任一向量 $x = (x_1, x_2, \cdots, x_n)^T \in \mathbf{P}^n$ 都有

$$x = x_1 \epsilon_1 + x_2 \epsilon_2 + \cdots + x_n \epsilon_n = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n) \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

所以 $(x_1, x_2, \cdots, x_n)^T$ 就是向量 x 在这组基下的坐标.

例 1.3.2 在线性空间 $\mathbf{P}[x]_n$ 中,

$$\epsilon_1 = 1, \epsilon_2 = x, \epsilon_3 = x^2, \cdots, \epsilon_n = x^{n-1}$$

是 $\mathbf{P}[x]_n$ 中 n 个线性无关的向量, 并且数域 \mathbf{P} 上任一次数小于 n 的多项式都可由 $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ 线性表示. 因此 $\dim(\mathbf{P}[x]_n) = n$, 并且 $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ 是 $\mathbf{P}[x]_n$ 的一组基. 在这组基下, 多项式 $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1}$ 的坐标为 $(a_0, a_1, \cdots, a_{n-1})^T$.

容易验证: $\epsilon'_1 = 1, \epsilon'_2 = x - a, \epsilon'_3 = (x - a)^2, \cdots, \epsilon'_n = (x - a)^{n-1}$ 也是 $\mathbf{P}[x]_n$ 的一组基.

由 Taylor 展开公式

$$f(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots$$

$$+ \frac{f^{(n-1)}(a)}{(n-1)!}(x-a)^{n-1}$$

因此,多项式 $f(x)$ 在基 $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$ 下的坐标为 $\left(f(a), f'(a), \frac{f''(a)}{2!}, \dots, \frac{f^{(n-1)}(a)}{(n-1)!}\right)^T$.

在 n 维线性空间 V 中,任意 n 个线性无关的向量都可取作 V 的基. 例 1.3.2 说明同一个向量在不同基下的坐标一般是不同的. 下面研究向量的坐标是如何随着基的改变而变化的. 建立 V 中任一向量在不同基下坐标间的关系.

设 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 与 $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$ 是 n 维线性空间 V 的两组基,它们之间有如下关系

$$\begin{cases} \epsilon'_1 = t_{11}\epsilon_1 + t_{21}\epsilon_2 + \dots + t_{n1}\epsilon_n \\ \epsilon'_2 = t_{12}\epsilon_1 + t_{22}\epsilon_2 + \dots + t_{n2}\epsilon_n \\ \dots\dots\dots \\ \epsilon'_n = t_{1n}\epsilon_1 + t_{2n}\epsilon_2 + \dots + t_{nn}\epsilon_n \end{cases} \quad (1.3.3)$$

关系式(1.3.3)用矩阵记号可表示为

$$(\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n) = (\epsilon_1, \epsilon_2, \dots, \epsilon_n) \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ \vdots & \vdots & & \vdots \\ t_{n1} & t_{n2} & \dots & t_{nn} \end{pmatrix} \quad (1.3.4)$$

$$n \text{ 阶矩阵 } T = \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ \vdots & \vdots & & \vdots \\ t_{n1} & t_{n2} & \dots & t_{nn} \end{pmatrix} \text{ 称为由基 } \epsilon_1, \epsilon_2, \dots, \epsilon_n \text{ 到基 } \epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$$

的过渡矩阵.

容易证明:过渡矩阵 T 是可逆的.

现在建立 V 中任一向量在不同基下坐标间的关系.

设向量 $\alpha \in V$ 在基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 与 $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$ 下的坐标分别为 $(x_1, x_2, \dots, x_n)^T$ 与 $(x'_1, x'_2, \dots, x'_n)^T$, 即

$$\alpha = (\epsilon_1, \epsilon_2, \dots, \epsilon_n) \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \alpha = (\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n) \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{bmatrix}$$

于是

$$(\epsilon_1, \epsilon_2, \dots, \epsilon_n) \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = (\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n) \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{bmatrix}$$

将(1.3.4)代入上式右端得

$$(\epsilon_1, \epsilon_2, \dots, \epsilon_n) \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n) \mathbf{T} \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{bmatrix}$$

因为 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 线性无关, 所以

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{T} \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{bmatrix} \quad \text{或} \quad \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{bmatrix} = \mathbf{T}^{-1} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (1.3.5)$$

(1.3.5)给出了在基变换(1.3.4)下, 向量的坐标变换公式.

例 1.3.3 在线性空间 $\mathbf{R}^{2 \times 2}$ 中,

$$\epsilon_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \epsilon_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \epsilon_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \epsilon_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

和

$$\epsilon'_1 = \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix}, \epsilon'_2 = \begin{bmatrix} 0 & 3 \\ -1 & 4 \end{bmatrix}, \epsilon'_3 = \begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix}, \epsilon'_4 = \begin{bmatrix} 1 & -3 \\ 0 & 2 \end{bmatrix}$$

是两组基, 求由基 $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$ 到基 $\epsilon'_1, \epsilon'_2, \epsilon'_3, \epsilon'_4$ 的过渡矩阵, 并求矩阵 \mathbf{A}

$$= \begin{pmatrix} -1 & 3 \\ 0 & 2 \end{pmatrix} \text{ 在这两组基下的坐标.}$$

解 因为

$$\begin{cases} \varepsilon'_1 = -\varepsilon_1 + 0\varepsilon_2 + 0\varepsilon_3 + 2\varepsilon_4 \\ \varepsilon'_2 = 0\varepsilon_1 + 3\varepsilon_2 - \varepsilon_3 + 4\varepsilon_4 \\ \varepsilon'_3 = 2\varepsilon_1 + 0\varepsilon_2 + \varepsilon_3 + 0\varepsilon_4 \\ \varepsilon'_4 = \varepsilon_1 - 3\varepsilon_2 + 0\varepsilon_3 + 2\varepsilon_4 \end{cases}$$

即

$$(\varepsilon'_1, \varepsilon'_2, \varepsilon'_3, \varepsilon'_4) = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4) \begin{pmatrix} -1 & 0 & 2 & 1 \\ 0 & 3 & 0 & -3 \\ 0 & -1 & 1 & 0 \\ 2 & 4 & 0 & 2 \end{pmatrix}$$

则由基 $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$ 到基 $\varepsilon'_1, \varepsilon'_2, \varepsilon'_3, \varepsilon'_4$ 的过渡矩阵为 $T =$

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

矩阵 A 在基 $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$ 下的坐标为 $(-1, 3, 0, 2)^T$, 而 A 在基 $\varepsilon'_1, \varepsilon'_2,$

$$\varepsilon'_3, \varepsilon'_4 \text{ 下的坐标则为 } T^{-1} \begin{bmatrix} -1 \\ 3 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1}{3} \\ \frac{1}{3} \\ -\frac{2}{3} \end{bmatrix}.$$

§ 1.4 线性子空间

1.4.1 线性子空间的概念

我们知道 $P[x]_n$ 是线性空间 $P[x]$ 的一部分, 同时对于 $P[x]$ 的运算构成线性空间, 我们称 $P[x]_n$ 为 $P[x]$ 的线性子空间. 一般地, 我们不仅要研究整个线性空间的结构, 而且要研究它的线性子空间. 一方面线性子空间本身有它的应用, 另一方面通过研究线性子空间可以更深刻地揭示整个线性空间的

结构.

定义 1.4.1 设 W 是数域 P 上线性空间 V 的非空子集, 如果 W 对于 V 的两种运算也构成数域 P 上的线性空间, 则称 W 为 V 的一个线性子空间(简称子空间).

如何判断线性空间 V 的一个非空子集 W 是否构成 V 的子空间? 我们有如下定理.

定理 1.4.1 数域 P 上线性空间 V 的非空子集 W 是 V 的一个线性子空间当且仅当 W 对于 V 的两种运算封闭, 即

(1) 如果 $\alpha, \beta \in W$, 则 $\alpha + \beta \in W$;

(2) 如果 $k \in P, \alpha \in W$, 则 $k\alpha \in W$.

证明 必要性由定义 1.4.1 直接得出. 下面证明充分性.

由已知条件, V 的两种运算都是 W 的运算. 因为 V 是线性空间, 而 W 是 V 的子集, 所以 W 的两种运算满足线性空间定义(见定义 1.2.1)中的规则 (1), (2), (5), (6), (7), (8). 因为 $0 \in P, \alpha \in W$, 所以 $0\alpha \in W$. 由于 V 是线性空间, 由定理 1.2.1 知 $0\alpha = 0$, 从而 $0 \in W$. 于是 V 的零元素就是 W 的零元素. 因为 $-1 \in P, \alpha \in W$, 所以 $(-1)\alpha \in W$. 由于 V 是线性空间, 由定理 1.2.1 有 $(-1)\alpha = -\alpha$, 从而 $-\alpha \in W$. 由于 $\alpha + (-1)\alpha = 0$, 于是 W 中每个元素 α 在 V 中的负元素 $-\alpha$ 也是 α 在 W 中的负元素. 因此 W 是 V 的一个线性子空间. \square

线性空间 V 的每个子空间都是凸集.

因为线性子空间也是一个线性空间, 所以线性空间的维数、基、坐标等概念也可以应用到线性子空间上.

例 1.4.1 在线性空间 V 中, 由 V 的零向量所组成的子集 $\{0\}$ 是 V 的一个子空间, 称为零子空间; 线性空间 V 本身也是 V 的一个子空间. 零子空间和线性空间 V 都称为 V 的平凡子空间, V 的其他子空间称为非平凡子空间.

例 1.4.2 设 $A \in \mathbb{R}^{m \times n}$, 齐次线性方程组

$$Ax = 0$$

的全部解向量构成 n 维线性空间 \mathbb{R}^n 的一个子空间. 这个子空间称为齐次线性方程组的解空间或矩阵 A 的零空间(核), 记为 $N(A)$ 或 $\text{Ker}(A)$. 因为解空间的基就是齐次线性方程组的基础解系, 所以 $\dim(N(A)) = n - \text{rank}(A)$.

由于在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向量, 于是我们得到如下结论.

定理 1.4.2 如果 W 是线性空间 V 的一个子空间, 则 $\dim(W) \leq \dim(V)$.

由定理 1.4.2 和定理 1.3.1 直接可得以下结论.

定理 1.4.3 若 W 是有限维线性空间 V 的子空间, 则 W 的一组基可扩充成 V 的一组基.

现在我们给出由线性空间 V 的一组向量构造 V 的子空间的方法.

设 $\alpha_1, \alpha_2, \dots, \alpha_s$ 是数域 \mathbf{P} 上线性空间 V 的一组向量, 这个向量组的所有线性组合作成的集合记为 W , 即

$$W = \{k_1\alpha_1 + k_2\alpha_2 + \dots + k_s\alpha_s \mid k_i \in \mathbf{P}, i = 1, \dots, s\}$$

显然, W 是 V 的非空子集, 并且由定理 1.4.1 知 W 是 V 的子空间. 我们称 W 是由向量 $\alpha_1, \alpha_2, \dots, \alpha_s$ 生成(或张成)的子空间, 记为 $L(\alpha_1, \alpha_2, \dots, \alpha_s)$ 或 $\text{span}(\alpha_1, \alpha_2, \dots, \alpha_s)$.

在有限维线性空间 V 中, 任何一个子空间 W 都可以用上述方法得到. 事实上, 设 W 是 V 的任一子空间, W 也是有限维的. 在 W 中取一组基 $\alpha_1, \alpha_2, \dots, \alpha_r$, 则 $W = \text{span}(\alpha_1, \alpha_2, \dots, \alpha_r)$.

关于由向量组所生成的子空间我们有如下常用的重要结果.

定理 1.4.4 设 $\alpha_1, \dots, \alpha_s$ 与 β_1, \dots, β_t 是线性空间 V 中两个向量组, 则

(1) $\text{span}(\alpha_1, \dots, \alpha_s) = \text{span}(\beta_1, \dots, \beta_t)$ 的充分必要条件是 $\alpha_1, \dots, \alpha_s$ 与 β_1, \dots, β_t 等价;

(2) $\dim(\text{span}(\alpha_1, \dots, \alpha_s)) = \text{rank}(\alpha_1, \dots, \alpha_s)$, 并且 $\text{span}(\alpha_1, \dots, \alpha_s)$ 的基是向量组 $\alpha_1, \dots, \alpha_s$ 的一个极大线性无关组.

证明 (1) 若 $\text{span}(\alpha_1, \dots, \alpha_s) = \text{span}(\beta_1, \dots, \beta_t)$, 则 $\alpha_i (i = 1, \dots, s)$ 可以由 β_1, \dots, β_t 线性表示; $\beta_j (j = 1, \dots, t)$ 可以由 $\alpha_1, \dots, \alpha_s$ 线性表示. 因此 $\alpha_1, \dots, \alpha_s$ 与 β_1, \dots, β_t 等价.

反过来, 如果 $\alpha_1, \dots, \alpha_s$ 与 β_1, \dots, β_t 等价, 则可以由 $\alpha_1, \dots, \alpha_s$ 线性表示的向量都可以由 β_1, \dots, β_t 线性表示; 反之亦然. 因而 $\text{span}(\alpha_1, \dots, \alpha_s) = \text{span}(\beta_1, \dots, \beta_t)$.

(2) 设向量组 $\alpha_1, \dots, \alpha_s$ 的秩是 r , 并且 $\alpha_1, \dots, \alpha_r (r \leq s)$ 是 $\alpha_1, \dots, \alpha_s$ 的一个极大线性无关组. 因为 $\alpha_1, \dots, \alpha_r$ 与 $\alpha_1, \dots, \alpha_s$ 等价, 所以 $\text{span}(\alpha_1, \dots, \alpha_s) = \text{span}(\alpha_1, \dots, \alpha_r)$. 由定理 1.2.5 知 $\text{span}(\alpha_1, \dots, \alpha_r)$ 的维数是 r , 并且 $\alpha_1, \dots, \alpha_r$ 是 $\text{span}(\alpha_1, \dots, \alpha_r)$ 的一组基. \square

例 1.4.3 设 $A \in \mathbf{R}^{m \times n}$, 记 $A = [\alpha_1, \alpha_2, \dots, \alpha_n]$, 其中 $\alpha_i \in \mathbf{R}^m (i = 1, \dots, n)$, 则 $\text{span}(\alpha_1, \dots, \alpha_n)$ 是 m 维线性空间 \mathbf{R}^m 的一个子空间, 称为矩阵 A 的列空间(值域), 记为 $R(A)$ 或 $\text{span}(A)$. $R(A)$ 可表成

$$R(A) = \{y \mid y = Ax, x \in \mathbf{R}^n\}$$

由定理 1.4.4 知 $\dim(R(A)) = \text{rank}(A)$, 并且 $\dim(N(A)) + \dim(R(A)) =$

n .

1.4.2 子空间的交与和

我们知道,集合有交、并运算.线性空间的子空间作为集合,也可以求交集和并集.那么两个子空间的交集、并集是否仍是子空间?

定理 1.4.5 设 V_1, V_2 是数域 \mathbf{P} 上线性空间 V 的两个子空间,则它们的交 $V_1 \cap V_2$ 也是 V 的子空间.

证明 因为 $0 \in V_1 \cap V_2$, 所以 $V_1 \cap V_2$ 是非空的. 设 $\alpha, \beta \in V_1 \cap V_2$, 则 $\alpha, \beta \in V_1$, 且 $\alpha, \beta \in V_2$. 因为 V_1, V_2 是 V 的子空间, 则对任意的数 $k \in \mathbf{P}$ 有 $\alpha + \beta \in V_1, k\alpha \in V_1$ 且 $\alpha + \beta \in V_2, k\alpha \in V_2$, 从而 $\alpha + \beta \in V_1 \cap V_2, k\alpha \in V_1 \cap V_2$. 因此, $V_1 \cap V_2$ 是 V 的子空间. \square

线性空间 V 的两个子空间 V_1 与 V_2 的并集一般不是 V 的子空间. 例如, 设 V 是以原点 O 为起点的所有向量组成的 3 维实线性空间, V_1, V_2 是通过原点的两个不同的平面, 它们都是 V 的子空间. 由于 $V_1 \cup V_2$ 对加法不封闭, 因此 $V_1 \cup V_2$ 不是 V 的子空间. 如果我们要构造一个包含 $V_1 \cup V_2$ 的子空间, 则这个子空间应当包含 V_1 中的任一向量 α_1 与 V_2 中的任一向量 α_2 的和. 由此受到启发, 我们引进如下定义.

定义 1.4.2 设 V_1, V_2 是数域 \mathbf{P} 上线性空间 V 的两个子空间, 则集合

$$\{\alpha_1 + \alpha_2 \mid \alpha_1 \in V_1, \alpha_2 \in V_2\}$$

称为 V_1 与 V_2 的和, 记为 $V_1 + V_2$.

定理 1.4.6 设 V_1, V_2 是数域 \mathbf{P} 上线性空间 V 的两个子空间, 则它们的和 $V_1 + V_2$ 也是 V 的子空间.

证明 因为 $0 \in V_1 + V_2$, 所以 $V_1 + V_2$ 非空. 对任意两个向量 $\alpha, \beta \in V_1 + V_2$ 及任意数 $k \in \mathbf{P}$, 则

$$\alpha = \alpha_1 + \alpha_2, \quad \beta = \beta_1 + \beta_2, \quad \alpha_1, \beta_1 \in V_1, \alpha_2, \beta_2 \in V_2$$

于是

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

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

由于 V_1, V_2 是 V 的子空间, 所以 $\alpha_1 + \beta_1 \in V_1, \alpha_2 + \beta_2 \in V_2, k\alpha_1 \in V_1, k\alpha_2 \in V_2$, 从而 $\alpha + \beta \in V_1 + V_2, k\alpha \in V_1 + V_2$. 因此 $V_1 + V_2$ 是 V 的子空间. \square

显然 $V_1 \cup V_2 \subseteq V_1 + V_2$, 所以 $V_1 + V_2$ 是包含 $V_1 \cup V_2$ 的子空间.

由子空间的交与和的定义可知,子空间的交与和适合下列运算规则:

(i) 交换律: $V_1 \cap V_2 = V_2 \cap V_1$;

$$V_1 + V_2 = V_2 + V_1.$$

(ii) 结合律: $(V_1 \cap V_2) \cap V_3 = V_1 \cap (V_2 \cap V_3)$;

$$(V_1 + V_2) + V_3 = V_1 + (V_2 + V_3).$$

由结合律,我们可定义多个子空间的交与和

$$V_1 \cap V_2 \cap \cdots \cap V_s = \bigcap_{i=1}^s V_i$$

$$V_1 + V_2 + \cdots + V_s = \sum_{i=1}^s V_i$$

用数学归纳法容易证明: $\bigcap_{i=1}^s V_i$ 和 $\sum_{i=1}^s V_i$ 都是 V 的子空间.

例 1.4.4 设 $\alpha_1, \cdots, \alpha_s$ 与 β_1, \cdots, β_t 是数域 \mathbf{P} 上线性空间 V 的两个向量组, 则

$$\text{span}(\alpha_1, \cdots, \alpha_s) + \text{span}(\beta_1, \cdots, \beta_t) = \text{span}(\alpha_1, \cdots, \alpha_s, \beta_1, \cdots, \beta_t)$$

证明 由子空间和的定义, 有

$$\begin{aligned} & \text{span}(\alpha_1, \cdots, \alpha_s) + \text{span}(\beta_1, \cdots, \beta_t) \\ &= \{(k_1\alpha_1 + \cdots + k_s\alpha_s) + (l_1\beta_1 + \cdots + l_t\beta_t) \mid k_i, l_j \in \mathbf{P}\} \\ &= \text{span}(\alpha_1, \cdots, \alpha_s, \beta_1, \cdots, \beta_t). \end{aligned}$$

□

例 1.4.5 设

$$\alpha_1 = \begin{bmatrix} 2 \\ 1 \\ 3 \\ 1 \end{bmatrix}, \alpha_2 = \begin{bmatrix} -1 \\ 1 \\ -3 \\ 1 \end{bmatrix}, \beta_1 = \begin{bmatrix} 4 \\ 5 \\ 3 \\ -1 \end{bmatrix}, \beta_2 = \begin{bmatrix} 1 \\ 5 \\ -3 \\ 1 \end{bmatrix}$$

$$V_1 = \text{span}(\alpha_1, \alpha_2), V_2 = \text{span}(\beta_1, \beta_2).$$

求 $V_1 \cap V_2, V_1 + V_2$ 的基与维数.

解 设 $\alpha \in V_1 \cap V_2$, 则 $\alpha \in V_1, \alpha \in V_2$, 于是

$$\alpha = k_1\alpha_1 + k_2\alpha_2 = k_3\beta_1 + k_4\beta_2$$

从而 $k_1 = 0, k_2 = \frac{5}{3}k_4, k_3 = -\frac{2}{3}k_4, \alpha = k_4\left(-\frac{5}{3}, \frac{5}{3}, -5, \frac{5}{3}\right)^T$. 因此 $V_1 \cap$

V_2 的基为 $\left(-\frac{5}{3}, \frac{5}{3}, -5, \frac{5}{3}\right)^T$, $\dim(V_1 \cap V_2) = 1$.

由例 1.4.4 知 $V_1 + V_2 = \text{span}(\alpha_1, \alpha_2, \beta_1, \beta_2)$. 因为 $\alpha_1, \alpha_2, \beta_1$ 是 $\alpha_1, \alpha_2, \beta_1, \beta_2$ 的极大线性无关组, 所以 $\alpha_1, \alpha_2, \beta_1$ 是 $V_1 + V_2$ 的基, $\dim(V_1 + V_2) = 3$, 并且 $\dim(V_1 + V_2) + \dim(V_1 \cap V_2) = \dim(V_1) + \dim(V_2)$.

定理 1.4.7(维数公式) 设 V_1, V_2 是数域 \mathbf{P} 上线性空间 V 的两个有限维子空间, 则 $V_1 \cap V_2$ 与 $V_1 + V_2$ 都是有限维的, 并且

$$\dim(V_1) + \dim(V_2) = \dim(V_1 + V_2) + \dim(V_1 \cap V_2)$$

证明 因为 V_1 是有限维的, 而 $V_1 \cap V_2$ 是 V_1 的子空间, 所以 $V_1 \cap V_2$ 也是有限维的. 设 $\dim(V_1) = n_1, \dim(V_2) = n_2, \dim(V_1 \cap V_2) = m$. 取 $V_1 \cap V_2$ 的一组基 $\alpha_1, \dots, \alpha_m$, 把它扩充成 V_1 的一组基 $\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_{n_1-m}$, 并且把 $\alpha_1, \dots, \alpha_m$ 也扩充成 V_2 的一组基 $\alpha_1, \dots, \alpha_m, \gamma_1, \dots, \gamma_{n_2-m}$, 则

$$V_1 = \text{span}(\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_{n_1-m}), \quad V_2 = \text{span}(\alpha_1, \dots, \alpha_m, \gamma_1, \dots, \gamma_{n_2-m})$$

并且 $V_1 + V_2 = \text{span}(\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_{n_1-m}, \gamma_1, \dots, \gamma_{n_2-m})$.

考虑等式

$$k_1 \alpha_1 + \dots + k_m \alpha_m + p_1 \beta_1 + \dots + p_{n_1-m} \beta_{n_1-m} + q_1 \gamma_1 + \dots + q_{n_2-m} \gamma_{n_2-m} = 0$$

令

$$\begin{aligned} \alpha &= k_1 \alpha_1 + \dots + k_m \alpha_m + p_1 \beta_1 + \dots + p_{n_1-m} \beta_{n_1-m} \\ &= -q_1 \gamma_1 - \dots - q_{n_2-m} \gamma_{n_2-m} \end{aligned}$$

由第一个等式知 $\alpha \in V_1$; 而由第二个等式有 $\alpha \in V_2$, 于是 $\alpha \in V_1 \cap V_2$. 因此 α 可以由 $\alpha_1, \dots, \alpha_m$ 线性表示, 即 $\alpha = l_1 \alpha_1 + \dots + l_m \alpha_m$, 则

$$l_1 \alpha_1 + \dots + l_m \alpha_m + q_1 \gamma_1 + \dots + q_{n_2-m} \gamma_{n_2-m} = 0$$

因为 $\alpha_1, \dots, \alpha_m, \gamma_1, \dots, \gamma_{n_2-m}$ 线性无关, 则得 $l_1 = \dots = l_m = q_1 = \dots = q_{n_2-m} = 0$, 因而 $\alpha = 0$. 从而有

$$k_1 \alpha_1 + \dots + k_m \alpha_m + p_1 \beta_1 + \dots + p_{n_1-m} \beta_{n_1-m} = 0$$

由于 $\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_{n_1-m}$ 线性无关, 得 $k_1 = \dots = k_m = p_1 = \dots = p_{n_1-m} = 0$. 这就证明了 $\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_{n_1-m}, \gamma_1, \dots, \gamma_{n_2-m}$ 线性无关. 由定理 1.4.4 知 $\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_{n_1-m}, \gamma_1, \dots, \gamma_{n_2-m}$ 是 $V_1 + V_2$ 的一组基, 因此 $V_1 + V_2$

是有限维的, 且 $\dim(V_1 + V_2) = n_1 + n_2 - m$. \square

由维数公式可知, 线性空间的两个有限维子空间之和的维数往往小于这两个子空间的维数之和. 两个子空间之和的维数等于它们维数之和的情形是特别重要的, 这样的两个子空间之和称为直和.

1.4.3 子空间的直和

子空间的直和是子空间之和的一个重要的特殊情形.

定义 1.4.3 设 V_1, V_2 是数域 \mathbf{P} 上线性空间 V 的两个子空间, 如果和 $V_1 + V_2$ 中每个向量 α 可惟一地表示成

$$\alpha = \alpha_1 + \alpha_2, \quad \alpha_1 \in V_1, \alpha_2 \in V_2$$

则称和 $V_1 + V_2$ 为直和, 记为 $V_1 \dot{+} V_2$.

定理 1.4.8 设 V_1, V_2 是数域 \mathbf{P} 上线性空间 V 的两个子空间, 则下面的叙述是等价的.

- (1) 和 $V_1 + V_2$ 是直和;
- (2) 和 $V_1 + V_2$ 中零向量的表示法惟一, 即若 $\alpha_1 + \alpha_2 = 0$ ($\alpha_1 \in V_1, \alpha_2 \in V_2$), 则 $\alpha_1 = 0, \alpha_2 = 0$;
- (3) $V_1 \cap V_2 = \{0\}$;
- (4) $\dim(V_1 + V_2) = \dim(V_1) + \dim(V_2)$.

证明 (1) \Rightarrow (2) 是显然的.

现在证 (2) \Rightarrow (3). 任取 $\alpha \in V_1 \cap V_2$, 因为零向量可表示为

$$0 = \alpha + (-\alpha), \alpha \in V_1, -\alpha \in V_2$$

由 (2) 得 $\alpha = 0$. 因此 $V_1 \cap V_2 = \{0\}$.

利用维数公式, 由 (3) 可直接导出 (4).

最后证 (4) \Rightarrow (1). 由维数公式和 (4) 知 $\dim(V_1 \cap V_2) = 0$, 即 $V_1 \cap V_2 = \{0\}$. 设 $\alpha \in V_1 + V_2$ 有两个分解式

$$\alpha = \alpha_1 + \alpha_2 = \beta_1 + \beta_2, \alpha_1, \beta_1 \in V_1, \alpha_2, \beta_2 \in V_2$$

则

$$\alpha_1 - \beta_1 = -(\alpha_2 - \beta_2)$$

其中 $\alpha_1 - \beta_1 \in V_1, \alpha_2 - \beta_2 \in V_2$, 从而 $\alpha_1 - \beta_1, \alpha_2 - \beta_2 \in V_1 \cap V_2$. 于是 $\alpha_1 = \beta_1, \alpha_2 = \beta_2$, 即向量 α 的分解式是惟一的. 因此和 $V_1 + V_2$ 是直和. \square

定理 1.4.9 设 U 是数域 \mathbf{P} 上有限维线性空间 V 的一个子空间, 则存在 V 的一个子空间 W 使得 $V = U \dot{+} W$.

证明 因为 V 是有限维的,不妨设 $\dim(V)=n$,则 U 也是有限维的. 假定 $\dim(U)=m$,取 U 的一组基 $\alpha_1, \alpha_2, \dots, \alpha_m$,把它扩充成 V 的一组基 $\alpha_1, \alpha_2, \dots, \alpha_m, \alpha_{m+1}, \dots, \alpha_n$. 令 $W = \text{span}(\alpha_{m+1}, \dots, \alpha_n)$,则 W 即满足要求. \square

子空间的直和概念可以推广到多个子空间的情形.

定义 1.4.4 设 V_1, V_2, \dots, V_s 是数域 \mathbf{P} 上线性空间 V 的 s 个子空间, 如果和 $V_1 + V_2 + \dots + V_s$ 中每个向量 α 可惟一地表示成

$$\alpha = \alpha_1 + \alpha_2 + \dots + \alpha_s, \quad \alpha_i \in V_i, \quad i = 1, \dots, s$$

则称和 $V_1 + V_2 + \dots + V_s$ 为直和,记为 $V_1 \dot{+} V_2 \dot{+} \dots \dot{+} V_s$.

类似于定理 1.4.8,我们有

定理 1.4.10 设 V_1, \dots, V_s 是数域 \mathbf{P} 上线性空间 V 的 s 个子空间,则下面的叙述是等价的.

- (1) 和 $V_1 + V_2 + \dots + V_s$ 是直和;
- (2) 和 $V_1 + V_2 + \dots + V_s$ 中零向量的表示法惟一;
- (3) $V_i \cap \sum_{j \neq i} V_j = \{0\}$;
- (4) $\dim(V_1 + V_2 + \dots + V_s) = \dim(V_1) + \dim(V_2) + \dots + \dim(V_s)$.

这个定理的证明和定理 1.4.8 基本相同,这里不再重复.

§ 1.5 线性空间的同构

设 V 是数域 \mathbf{P} 上的 n 维线性空间, $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 是 V 的一组基. 对 V 中任一向量 α ,它可惟一地表示为 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 的线性组合,即

$$\alpha = x_1 \epsilon_1 + x_2 \epsilon_2 + \dots + x_n \epsilon_n$$

令

$$\sigma: V \rightarrow \mathbf{P}^n$$

$$\alpha = x_1 \epsilon_1 + x_2 \epsilon_2 + \dots + x_n \epsilon_n \rightarrow x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

则 σ 是 V 到 \mathbf{P}^n 上的一一映射. 这个映射的重要性表现在它保持运算关系不变.

事实上,对 $k \in \mathbf{P}$ 及 V 中向量 β ,有

$$\beta = y_1 \epsilon_1 + y_2 \epsilon_2 + \cdots + y_n \epsilon_n$$

$$\alpha + \beta = (x_1 + y_1) \epsilon_1 + (x_2 + y_2) \epsilon_2 + \cdots + (x_n + y_n) \epsilon_n$$

$$k\alpha = kx_1 \epsilon_1 + kx_2 \epsilon_2 + \cdots + kx_n \epsilon_n$$

因为

$$\sigma(\alpha) = x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \sigma(\beta) = y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

则

$$\sigma(\alpha + \beta) = x + y = \sigma(\alpha) + \sigma(\beta)$$

$$\sigma(k\alpha) = kx = k\sigma(\alpha)$$

这说明在向量用坐标表示之后,向量的运算就可以归结为它们坐标的运算,因而线性空间 V 的讨论也就可以归结为 \mathbf{P}^n 的讨论. 为了确切地说明这个事实,需要介绍同构的概念.

定义 1.5.1 设 V 与 V' 都是数域 \mathbf{P} 上的线性空间,如果存在 V 到 V' 上的一一映射 σ 满足

$$(1) \sigma(\alpha + \beta) = \sigma(\alpha) + \sigma(\beta);$$

$$(2) \sigma(k\alpha) = k\sigma(\alpha),$$

其中 α, β 是 V 中任意向量, k 是数域 \mathbf{P} 中任意数,则称 σ 为 V 到 V' 的**同构映射**,并且称 V 与 V' 是**同构**的.

前面的讨论说明在数域 \mathbf{P} 上 n 维线性空间 V 中取定一组基以后,向量与它的坐标之间的对应就是 V 到 \mathbf{P}^n 的一个同构映射. 因此数域 \mathbf{P} 上任一 n 维线性空间都与 \mathbf{P}^n 同构.

定理 1.5.1 设 V 与 V' 是数域 \mathbf{P} 上的同构线性空间, σ 为 V 到 V' 的同构映射,则

$$(1) \sigma(0) = 0', 0' \text{ 是 } V' \text{ 的零向量};$$

$$(2) \text{ 对任意 } \alpha \in V, \sigma(-\alpha) = -\sigma(\alpha);$$

$$(3) \text{ 如果 } \alpha_1, \cdots, \alpha_m \text{ 是 } V \text{ 的一个向量组, } k_1, \cdots, k_m \in \mathbf{P}, \text{ 则}$$

$$\sigma(k_1 \alpha_1 + \cdots + k_m \alpha_m) = k_1 \sigma(\alpha_1) + \cdots + k_m \sigma(\alpha_m)$$

(4) V 中向量组 $\alpha_1, \cdots, \alpha_m$ 线性相关当且仅当它们的像 $\sigma(\alpha_1), \cdots, \sigma(\alpha_m)$ 是 V' 中线性相关的向量组;

(5) 如果 V 是 n 维的, $\epsilon_1, \dots, \epsilon_n$ 是 V 的一组基, 则 V' 也是 n 维的, 并且 $\sigma(\epsilon_1), \dots, \sigma(\epsilon_n)$ 是 V' 的一组基.

证明 (1) 任取 $\alpha' \in V'$, 因为 σ 是满映射, 所以存在 $\alpha \in V$ 使得 $\sigma(\alpha) = \alpha'$. 于是

$$\alpha' + \sigma(0) = \sigma(\alpha) + \sigma(0) = \sigma(\alpha + 0) = \sigma(\alpha) = \alpha'$$

这表明 $\sigma(0)$ 是 V' 的零向量.

(2) 因为 $\sigma(\alpha) + \sigma(-\alpha) = \sigma(\alpha + (-\alpha)) = \sigma(0) = 0'$, 所以 $\sigma(-\alpha) = \sigma(\alpha)$.

(3) 由定义 1.5.1 即得.

(4) 如果向量组 $\alpha_1, \dots, \alpha_m$ 线性相关, 则存在不全为零的数 $k_1, \dots, k_m \in \mathbf{P}$ 使得

$$k_1\alpha_1 + \dots + k_m\alpha_m = 0$$

由(1)和(3)得

$$k_1\sigma(\alpha_1) + \dots + k_m\sigma(\alpha_m) = 0'$$

所以 $\sigma(\alpha_1), \dots, \sigma(\alpha_m)$ 线性相关.

反过来, 如果 $\sigma(\alpha_1), \dots, \sigma(\alpha_m)$ 线性相关, 则存在不全为零的数 $k_1, \dots, k_m \in \mathbf{P}$ 使得

$$k_1\sigma(\alpha_1) + \dots + k_m\sigma(\alpha_m) = 0'$$

即

$$\sigma(k_1\alpha_1 + \dots + k_m\alpha_m) = 0'$$

因为 σ 是一一映射, 所以 $k_1\alpha_1 + \dots + k_m\alpha_m = 0$, 从而 $\alpha_1, \alpha_2, \dots, \alpha_m$ 线性相关.

(5) 由(4)知 $\sigma(\epsilon_1), \dots, \sigma(\epsilon_n)$ 是 V' 的线性无关向量组. 对任意 $\alpha' \in V'$, 因为 σ 是满映射, 所以存在 $\alpha \in V$ 使得 $\sigma(\alpha) = \alpha'$. 因为 $\alpha = x_1\epsilon_1 + \dots + x_n\epsilon_n$, 则

$$\alpha' = \sigma(x_1\epsilon_1 + \dots + x_n\epsilon_n) = x_1\sigma(\epsilon_1) + \dots + x_n\sigma(\epsilon_n)$$

由定理 1.2.5 知 V' 是 n 维的, 并且 $\sigma(\epsilon_1), \dots, \sigma(\epsilon_n)$ 是 V' 的一组基. \square

定理 1.5.2 数域 \mathbf{P} 上的两个有限维线性空间 V 与 V' 同构的充分必要条件是它们的维数相同.

证明 必要性由定理 1.5.1(5) 即得. 下面证充分性. 设 $\dim(V) = \dim(V')$. 在 V 与 V' 中分别取一组基 $\epsilon_1, \dots, \epsilon_n$ 与 $\epsilon'_1, \dots, \epsilon'_n$, 令

$$\sigma: V \rightarrow V'$$

$$\alpha = x_1 \epsilon_1 + \cdots + x_n \epsilon_n \rightarrow \alpha' = x_1 \epsilon'_1 + \cdots + x_n \epsilon'_n$$

则 σ 是 V 到 V' 上的一一映射.

设 $\beta = y_1 \epsilon_1 + \cdots + y_n \epsilon_n \in V$, 对 $k \in \mathbf{P}$, 直接计算可得

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

因此 σ 是 V 到 V' 的一个同构映射, 从而 V 与 V' 同构. \square

在线性空间的讨论中, 如果只涉及线性空间在线性运算下的代数性质, 那么同构的线性空间具有相同的性质. 特别地, 每一个数域 \mathbf{P} 上的 n 维线性空间都与 n 维向量空间 \mathbf{P}^n 同构. 因此, 关于 n 维向量空间 \mathbf{P}^n 的一些结论在一般的线性空间中也成立.

§ 1.6 内积空间

迄今为止, 我们对线性空间的讨论主要是围绕着向量之间的加法和数量乘法进行的. 与几何空间相比, 向量的度量性质如长度、夹角等, 在前几节的讨论中还没有得到反映. 但是向量的度量性质在实际应用中是非常重要的, 因此有必要引入度量的概念.

1.6.1 内积空间及其基本性质

在解析几何中引进了向量的内积概念后, 向量的长度、两个向量之间的夹角等度量性质都可以用内积来表示. 受此启发, 我们首先在一般的线性空间中定义内积运算, 导出内积空间的概念, 然后引进长度、角度等度量概念.

定义 1.6.1 设 V 是数域 \mathbf{P} 上的线性空间, V 到 \mathbf{P} 的一个代数运算记为 (α, β) . 如果 (α, β) 满足下列条件:

- (1) $(\alpha, \beta) = \overline{(\beta, \alpha)}$;
- (2) $(\alpha + \beta, \gamma) = (\alpha, \gamma) + (\beta, \gamma)$;
- (3) $(k\alpha, \beta) = k(\alpha, \beta)$;
- (4) $(\alpha, \alpha) \geq 0$, 当且仅当 $\alpha = 0$ 时 $(\alpha, \alpha) = 0$,

其中 k 是数域 \mathbf{P} 中的任意数, α, β, γ 是 V 中的任意元素, 则称 (α, β) 为 α 与 β 的内积. 定义了内积的线性空间 V 称为内积空间. 特别地, 称实数域 \mathbf{R} 上的内积空间 V 为 Euclid 空间 (简称为欧氏空间); 称复数域 \mathbf{C} 上的内积空间 V 为酉空间.

由定义 1.6.1 不难导出, 在内积空间中有

$$(1) (\alpha, \beta + \gamma) = (\alpha, \beta) + (\alpha, \gamma);$$

$$(2) (\alpha, k\beta) = k(\alpha, \beta);$$

$$(3) (\alpha, 0) = (0, \alpha) = 0.$$

例 1.6.1 在实数域 \mathbf{R} 上 n 维线性空间 \mathbf{R}^n 中, 对向量

$$x = (x_1, x_2, \dots, x_n)^T, y = (y_1, y_2, \dots, y_n)^T$$

定义内积

$$(x, y) = y^T x = \sum_{i=1}^n x_i y_i \quad (1.6.1)$$

则 \mathbf{R}^n 成为一个欧氏空间, 仍用 \mathbf{R}^n 表示这个欧氏空间.

例 1.6.2 在复数域 \mathbf{C} 上 n 维线性空间 \mathbf{C}^n 中, 对向量

$$x = (x_1, x_2, \dots, x_n)^T, y = (y_1, y_2, \dots, y_n)^T$$

定义内积

$$(x, y) = y^H x = \sum_{i=1}^n x_i \bar{y}_i \quad (1.6.2)$$

其中 $y^H = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_n)$, 则 \mathbf{C}^n 成为一个酉空间, 仍用 \mathbf{C}^n 表示这个酉空间.

以上两个例子中的内积称为 \mathbf{R}^n 或 \mathbf{C}^n 上的标准内积. 以后如不特别声明, \mathbf{R}^n 或 \mathbf{C}^n 上的内积均指标准内积.

例 1.6.3 在线性空间 $\mathbf{R}^{m \times n}$ 中, 对矩阵 $A, B \in \mathbf{R}^{m \times n}$, 定义

$$(A, B) = \text{tr}(B^T A) \quad (1.6.3)$$

其中 $\text{tr}(D)$ 表示方阵 D 的迹 (即方阵 D 的对角元之和). 容易证明 (A, B) 是 $\mathbf{R}^{m \times n}$ 上的内积, $\mathbf{R}^{m \times n}$ 是欧氏空间.

例 1.6.4 在线性空间 $C[a, b]$ 中, 对 $f(x), g(x) \in C[a, b]$, 定义

$$(f, g) = \int_a^b f(x)g(x)dx \quad (1.6.4)$$

则 (f, g) 是 $C[a, b]$ 上的内积, $C[a, b]$ 成为欧氏空间.

现在把几何空间中向量的长度、夹角等概念推广到内积空间.

定义 1.6.2 设 V 是内积空间, V 中向量 α 的长度定义为 $\|\alpha\| = \sqrt{(\alpha, \alpha)}$.

长度为 1 的向量称为单位向量. 如果 $\alpha \neq 0$, 则 $\frac{\alpha}{\|\alpha\|}$ 是一个单位向量.

这样定义的向量长度与几何空间中向量的长度是一致的.

定理 1.6.1 设 V 是数域 P 上的内积空间, 则向量长度 $\|\alpha\|$ 具有如下性质:

(1) $\|\alpha\| \geq 0$, 当且仅当 $\alpha=0$ 时 $\|\alpha\|=0$;

(2) 对任意 $k \in \mathbf{P}$, 有 $\|k\alpha\| = |k| \|\alpha\|$;

(3) 对任意 $\alpha, \beta \in V$, 有

$$\|\alpha + \beta\|^2 + \|\alpha - \beta\|^2 = 2(\|\alpha\|^2 + \|\beta\|^2) \quad (1.6.5)$$

(4) 对任意 $\alpha, \beta \in V$, 有 $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$;

(5) 对任意 $\alpha, \beta \in V$, 有

$$|(\alpha, \beta)| \leq \|\alpha\| \|\beta\| \quad (1.6.6)$$

并且等号成立的充分必要条件是 α, β 线性相关.

证明 (1)和(2)显然成立. 下面证(3), (4)和(5).

先证明(3). 将长度用内积表示, 即得

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

其次证明(5). 当 $\beta=0$ 时, (1.6.6)显然成立. 以下设 $\beta \neq 0$, 对任意 $t \in \mathbf{P}$, $\alpha + t\beta \in V$, 则

$$0 \leq (\alpha + t\beta, \alpha + t\beta) = (\alpha, \alpha) + t(\beta, \alpha) + t(\alpha, \beta) + |t|^2(\beta, \beta)$$

令 $t = -\frac{(\alpha, \beta)}{(\beta, \beta)}$, 代入上式得

$$(\alpha, \alpha) - \frac{|(\alpha, \beta)|^2}{(\beta, \beta)} \geq 0$$

于是不等式(1.6.6)成立.

当 α, β 线性相关时, (1.6.6)中等号显然成立. 如果 α, β 线性无关, 则对任意 $t \in \mathbf{P}$, $\alpha + t\beta \neq 0$, 从而

$$(\alpha + t\beta, \alpha + t\beta) > 0$$

取 $t = -\frac{(\alpha, \beta)}{(\beta, \beta)}$, 有 $|(\alpha, \beta)|^2 < (\alpha, \alpha)(\beta, \beta) = \|\alpha\|^2 \|\beta\|^2$, 这与(1.6.6)等号成立矛盾. 因此 α, β 线性相关.

最后证明(4). 对任意 $\alpha, \beta \in V$, 有

$$\begin{aligned} \|\alpha + \beta\|^2 &= (\alpha + \beta, \alpha + \beta) = (\alpha, \alpha) + (\alpha, \beta) + (\beta, \alpha) + (\beta, \beta) \quad (1.6.7) \\ &\leq \|\alpha\|^2 + 2|(\alpha, \beta)| + \|\beta\|^2 \\ &\leq \|\alpha\|^2 + 2\|\alpha\| \|\beta\| + \|\beta\|^2 = (\|\alpha\| + \|\beta\|)^2 \end{aligned}$$

由此即得(4). \square

不等式(1.6.6)有十分重要的应用. 例如它在欧氏空间 \mathbf{R}^n 中的形式为

$$\left| \sum_{i=1}^n x_i y_i \right| \leq \left(\sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^n y_j^2 \right)^{\frac{1}{2}} \quad (1.6.8)$$

不等式(1.6.8)称为 **Cauchy 不等式**.

定义 1.6.3 设 V 是内积空间, V 中向量 α 与 β 之间的距离定义为

$$d(\alpha, \beta) = \|\alpha - \beta\| \quad (1.6.9)$$

并称 $d(\alpha, \beta) = \|\alpha - \beta\|$ 是由长度导出的距离.

由定理 1.6.1 知, 这样定义的距离满足如下三个基本条件:

- (1) $d(\alpha, \beta) \geq 0$, 并且 $d(\alpha, \beta) = 0$ 的充分必要条件是 $\alpha = \beta$;
- (2) $d(\alpha, \beta) = d(\beta, \alpha)$;
- (3) $d(\alpha, \gamma) \leq d(\alpha, \beta) + d(\beta, \gamma)$,

其中 α, β, γ 是内积空间 V 中的任意向量. 因此, 内积空间按此距离成为度量空间, 从而可以在其中讨论极限、开集、闭集等.

由不等式(1.6.6), 可以定义欧氏空间中两个向量之间的夹角.

定义 1.6.4 设 α, β 是欧氏空间中两个非零向量, 它们之间的夹角 $\langle \alpha, \beta \rangle$ 定义为

$$\langle \alpha, \beta \rangle = \arccos \frac{(\alpha, \beta)}{\|\alpha\| \|\beta\|}, \quad 0 \leq \langle \alpha, \beta \rangle \leq \pi \quad (1.6.10)$$

定义 1.6.5 设 α, β 是内积空间中两个向量, 如果 $(\alpha, \beta) = 0$, 则称 α 与 β 正交, 记为 $\alpha \perp \beta$.

由定义 1.6.5 及内积的性质知, 零向量与任何向量都正交, 并且只有零向量与自身正交.

如果 α 与 β 正交, 则由(1.6.7)即得“勾股定理”

$$\|\alpha + \beta\|^2 = \|\alpha\|^2 + \|\beta\|^2 \quad (1.6.11)$$

设 V 是数域 \mathbf{P} 上的 n 维内积空间, $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 是 V 的一组基. 对任意 $\alpha, \beta \in V$, 有

$$\alpha = x_1 \epsilon_1 + x_2 \epsilon_2 + \dots + x_n \epsilon_n, \quad \beta = y_1 \epsilon_1 + y_2 \epsilon_2 + \dots + y_n \epsilon_n$$

则 α 与 β 的内积

$$(\alpha, \beta) = \left(\sum_{i=1}^n x_i \epsilon_i, \sum_{j=1}^n y_j \epsilon_j \right) = \sum_{i=1}^n \sum_{j=1}^n (\epsilon_i, \epsilon_j) x_i y_j$$

令

$$a_{ij} = (\epsilon_i, \epsilon_j), \quad i, j = 1, \dots, n$$

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

称矩阵 A 为基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 的度量矩阵. 显然 $a_{ij} = \bar{a}_{ji} (i, j = 1, 2, \dots, n)$, 并且

$$(\alpha, \beta) = y^H A x \quad (1.6.12)$$

定理 1.6.2 设 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 是数域 P 上 n 维内积空间 V 的一组基, 则它的度量矩阵 A 非奇异.

证明 假若基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 的度量矩阵 A 奇异, 则齐次线性方程组

$$Ax = 0$$

有非零解 $x = (x_1, x_2, \dots, x_n)^T \in P^n$. 令

$$\alpha = x_1 \epsilon_1 + x_2 \epsilon_2 + \cdots + x_n \epsilon_n$$

则 $\alpha \neq 0$, 但 $(\alpha, \alpha) = x^H A x = 0$, 这与 $(\alpha, \alpha) > 0$ 矛盾. 因此 A 非奇异. \square

定义 1.6.6 设 $A \in C^{n \times n}$, 用 \bar{A} 表示以 A 的元素的共轭复数为元素组成的矩阵, $A^H = (\bar{A})^T$ 称为 A 的共轭转置矩阵.

容易验证矩阵的共轭转置运算具有下列性质:

- (1) $A^H = (\bar{A}^T)$;
- (2) $(A + B)^H = A^H + B^H$;
- (3) $(kA)^H = \bar{k}A^H$;
- (4) $(AB)^H = B^H A^H$;
- (5) $(A^H)^H = A$;
- (6) 如果 A 可逆, 则 $(A^H)^{-1} = (A^{-1})^H$.

定义 1.6.7 设 $A \in C^{n \times n}$, 如果 $A^H = A$, 则称 A 为 Hermite 矩阵; 如果 $A^H = -A$, 则称 A 为反 Hermite 矩阵.

实对称矩阵是 Hermite 矩阵, 有限维内积空间的度量矩阵是 Hermite 矩阵.

对于内积空间不同的基, 它们的度量矩阵一般是不同的, 下面的定理给出了它们之间的关系.

定理 1.6.3 设 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 与 $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$ 是数域 P 上 n 维内积空间 V 的两组基, 它们的度量矩阵分别为 A 和 B , 并且基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 到基 $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$ 的过渡矩阵为 P , 则 $B = P^H A P$.

证明作为练习留给读者.

定义 1.6.8 设 $A, B \in \mathbb{C}^{n \times n}$, 如果存在 n 阶非奇异矩阵 P , 使得

$$B = P^H A P$$

则称 A 与 B 是相合的.

方阵之间的相合关系是等价关系. 由定理 1.6.3 可知, 不同基的度量矩阵是相合的.

1.6.2 标准正交基与 Gram-Schmidt 正交化方法

定义 1.6.9 设 $\alpha_1, \alpha_2, \dots, \alpha_m$ 是内积空间 V 中的非零向量组, 如果它们两两正交, 则称 $\alpha_1, \alpha_2, \dots, \alpha_m$ 是正交向量组. 如果正交向量组 $\alpha_1, \alpha_2, \dots, \alpha_m$ 中的每一个向量都是单位向量, 则称 $\alpha_1, \alpha_2, \dots, \alpha_m$ 是标准正交向量组.

由定义可知, 向量组 $\alpha_1, \alpha_2, \dots, \alpha_m$ 是正交向量组的充分必要条件是

$$(\alpha_i, \alpha_j) = 0, \quad i \neq j$$

定理 1.6.4 设 $\alpha_1, \alpha_2, \dots, \alpha_m$ 是内积空间 V 中的非零正交向量组, 则 $\alpha_1, \alpha_2, \dots, \alpha_m$ 线性无关.

证明 若

$$k_1 \alpha_1 + k_2 \alpha_2 + \dots + k_m \alpha_m = 0$$

上式两边与 α_j 作内积, 有

$$(k_1 \alpha_1 + k_2 \alpha_2 + \dots + k_m \alpha_m, \alpha_j) = \sum_{i=1}^m k_i (\alpha_i, \alpha_j) = 0$$

因为 $(\alpha_i, \alpha_j) = 0 (i \neq j)$, 则上式简化为

$$k_j (\alpha_j, \alpha_j) = 0$$

由于 $(\alpha_j, \alpha_j) \neq 0$, 故 $k_j = 0$. 因此 $\alpha_1, \alpha_2, \dots, \alpha_m$ 线性无关. \square

定理 1.6.5 设 $\alpha_1, \alpha_2, \dots, \alpha_m$ 是内积空间 V 中的一个向量组, 则 $\alpha_1, \alpha_2, \dots, \alpha_m$ 线性无关的充分必要条件是矩阵

$$G(\alpha_1, \alpha_2, \dots, \alpha_m) = \begin{bmatrix} (\alpha_1, \alpha_1) & (\alpha_1, \alpha_2) & \cdots & (\alpha_1, \alpha_m) \\ (\alpha_2, \alpha_1) & (\alpha_2, \alpha_2) & \cdots & (\alpha_2, \alpha_m) \\ \vdots & \vdots & & \vdots \\ (\alpha_m, \alpha_1) & (\alpha_m, \alpha_2) & \cdots & (\alpha_m, \alpha_m) \end{bmatrix}$$

非奇异.

证明 若 $\alpha_1, \alpha_2, \dots, \alpha_m$ 线性无关, 则 $\alpha_1, \alpha_2, \dots, \alpha_m$ 是 $\text{span}(\alpha_1, \alpha_2, \dots, \alpha_m)$ 的一组基, 由定理 1.6.2 知 $\alpha_1, \alpha_2, \dots, \alpha_m$ 的度量矩阵 $G(\alpha_1, \alpha_2, \dots, \alpha_m)$ 非奇异.

反过来, 设 $G(\alpha_1, \alpha_2, \dots, \alpha_m)$ 非奇异. 假若 $\alpha_1, \alpha_2, \dots, \alpha_m$ 线性相关, 则 $\alpha_1, \alpha_2, \dots, \alpha_m$ 中必有一个向量是其余向量的线性组合. 不妨设 α_m 是 $\alpha_1, \alpha_2, \dots, \alpha_{m-1}$ 的线性组合 $\alpha_m = \sum_{i=1}^{m-1} l_i \alpha_i$, 则

$$\begin{aligned} |G(\alpha_1, \alpha_2, \dots, \alpha_m)| &= \begin{vmatrix} (\alpha_1, \alpha_1) & (\alpha_1, \alpha_2) & \cdots & (\alpha_1, \alpha_{m-1}) & (\alpha_1, \sum_{i=1}^{m-1} l_i \alpha_i) \\ (\alpha_2, \alpha_1) & (\alpha_2, \alpha_2) & \cdots & (\alpha_2, \alpha_{m-1}) & (\alpha_2, \sum_{i=1}^{m-1} l_i \alpha_i) \\ \vdots & \vdots & & \vdots & \vdots \\ (\alpha_m, \alpha_1) & (\alpha_m, \alpha_2) & \cdots & (\alpha_m, \alpha_{m-1}) & (\alpha_m, \sum_{i=1}^{m-1} l_i \alpha_i) \end{vmatrix} \\ &= \sum_{i=1}^{m-1} \bar{l}_i \begin{vmatrix} (\alpha_1, \alpha_1) & (\alpha_1, \alpha_2) & \cdots & (\alpha_1, \alpha_{m-1}) & (\alpha_1, \alpha_i) \\ (\alpha_2, \alpha_1) & (\alpha_2, \alpha_2) & \cdots & (\alpha_2, \alpha_{m-1}) & (\alpha_2, \alpha_i) \\ \vdots & \vdots & & \vdots & \vdots \\ (\alpha_m, \alpha_1) & (\alpha_m, \alpha_2) & \cdots & (\alpha_m, \alpha_{m-1}) & (\alpha_m, \alpha_i) \end{vmatrix} = 0 \end{aligned}$$

这与 $|G(\alpha_1, \alpha_2, \dots, \alpha_m)| \neq 0$ 矛盾. 故 $\alpha_1, \alpha_2, \dots, \alpha_m$ 线性无关. \square

矩阵 $G(\alpha_1, \alpha_2, \dots, \alpha_m)$ 称为向量组 $\alpha_1, \alpha_2, \dots, \alpha_m$ 的 **Gram 矩阵**.

由定理 1.6.4 可知, 在 n 维内积空间中两两正交的非零向量不超过 n 个.

定义 1.6.10 在 n 维内积空间中, 由 n 个正交向量组成的基称为**正交基**. 由 n 个标准正交向量组成的基称为**标准正交基**.

显然, $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 是 n 维内积空间的标准正交基当且仅当

$$(\epsilon_i, \epsilon_j) = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

即标准正交基的度量矩阵是单位矩阵.

从内积空间的一组基出发, 可以构造它的标准正交基.

定理 1.6.6 设 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 n 维内积空间 V 中的一组基, 则可以找到 V 的一组标准正交基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$, 使得

$$\text{span}(\alpha_1, \alpha_2, \cdots, \alpha_i) = \text{span}(\epsilon_1, \epsilon_2, \cdots, \epsilon_i), \quad i = 1, 2, \cdots, n$$

证明 我们由 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 逐个地求出标准正交向量组 $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$.
令

$$\epsilon_1 = \frac{\alpha_1}{\|\alpha_1\|}$$

则 ϵ_1 是单位向量, 并且 $\text{span}(\alpha_1) = \text{span}(\epsilon_1)$.

一般地, 假定已经求出标准正交向量 $\epsilon_1, \epsilon_2, \cdots, \epsilon_m$, 并且

$$\text{span}(\alpha_1, \alpha_2, \cdots, \alpha_i) = \text{span}(\epsilon_1, \epsilon_2, \cdots, \epsilon_i), \quad i = 1, 2, \cdots, m$$

下一步求 ϵ_{m+1} .

作向量

$$\beta_{m+1} = \alpha_{m+1} - k_1 \epsilon_1 - k_2 \epsilon_2 - \cdots - k_m \epsilon_m$$

其中 k_1, k_2, \cdots, k_m 是待定常数. 选取

$$k_i = (\alpha_{m+1}, \epsilon_i), \quad i = 1, 2, \cdots, m$$

则 $(\beta_{m+1}, \epsilon_i) = 0 (i = 1, 2, \cdots, m)$. 因为 $\alpha_1, \alpha_2, \cdots, \alpha_{m+1}$ 线性无关, $\text{span}(\alpha_1, \alpha_2, \cdots, \alpha_m) = \text{span}(\epsilon_1, \epsilon_2, \cdots, \epsilon_m)$, 于是 $\beta_{m+1} \neq 0$, 并且 $\epsilon_1, \epsilon_2, \cdots, \epsilon_m, \beta_{m+1}$ 是正交向量组. 令

$$\epsilon_{m+1} = \frac{\beta_{m+1}}{\|\beta_{m+1}\|}$$

则 $\epsilon_1, \epsilon_2, \cdots, \epsilon_m, \epsilon_{m+1}$ 是标准正交向量组, 并且

$$\text{span}(\alpha_1, \alpha_2, \cdots, \alpha_m, \alpha_{m+1}) = \text{span}(\epsilon_1, \epsilon_2, \cdots, \epsilon_m, \epsilon_{m+1})$$

由归纳法原理, 定理 1.6.6 得证. \square

定理 1.6.6 的证明中把一组线性无关向量变成标准正交向量组的方法通常称为 **Gram-Schmidt 正交化方法**.

由定理 1.4.3 和定理 1.6.6, 可得如下定理.

定理 1.6.7 设 $\epsilon_1, \epsilon_2, \cdots, \epsilon_m$ 是 n 维内积空间 V 中的一组标准正交向量, 则可将它扩充为 V 的一组标准正交基 $\epsilon_1, \epsilon_2, \cdots, \epsilon_m, \epsilon_{m+1}, \cdots, \epsilon_n$.

例 1.6.5 在欧氏空间 \mathbb{R}^4 中, 把向量组

$$\alpha_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \alpha_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \alpha_3 = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \alpha_4 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}$$

化成标准正交向量组.

解 利用 Gram-Schmidt 正交化方法,可依次求出

$$\epsilon_1 = \frac{\alpha_1}{\|\alpha_1\|} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0\right)^T$$

$$\beta_2 = \alpha_2 - (\alpha_2, \epsilon_1)\epsilon_1 = \left(\frac{1}{2}, -\frac{1}{2}, 1, 0\right)^T$$

$$\epsilon_2 = \frac{\beta_2}{\|\beta_2\|} = \left(\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, 0\right)^T$$

$$\beta_3 = \alpha_3 - (\alpha_3, \epsilon_1)\epsilon_1 - (\alpha_3, \epsilon_2)\epsilon_2 = \left(-\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 1\right)^T$$

$$\epsilon_3 = \frac{\beta_3}{\|\beta_3\|} = \left(-\frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{3}{\sqrt{12}}\right)^T$$

$$\beta_4 = \alpha_4 - (\alpha_4, \epsilon_1)\epsilon_1 - (\alpha_4, \epsilon_2)\epsilon_2 - (\alpha_4, \epsilon_3)\epsilon_3 = (1, -1, -1, 1)^T$$

$$\epsilon_4 = \frac{\beta_4}{\|\beta_4\|} = \left(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right)^T$$

则 $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$ 是标准正交向量组.

在标准正交基下,向量的坐标和内积有特别简单的表达式. 设 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 是 n 维内积空间 V 的一组标准正交基,则对任意 $\alpha \in V$ 都有

$$\alpha = (\alpha, \epsilon_1)\epsilon_1 + (\alpha, \epsilon_2)\epsilon_2 + \cdots + (\alpha, \epsilon_n)\epsilon_n \quad (1.6.13)$$

并且对任意 $\alpha, \beta \in V$, 如果

$$\alpha = x_1\epsilon_1 + x_2\epsilon_2 + \cdots + x_n\epsilon_n, \beta = y_1\epsilon_1 + y_2\epsilon_2 + \cdots + y_n\epsilon_n \quad (1.6.14)$$

则

$$(\alpha, \beta) = \sum_{i=1}^n x_i \tilde{y}_i \quad (1.6.15)$$

该表达式是解析几何中向量内积表达式的推广.

1.6.3 正交补与投影定理

定义 1.6.11 设 V_1, V_2 是内积空间 V 中两个子空间. 向量 $\alpha \in V$, 如果对任意 $\beta \in V_1$ 都有 $(\alpha, \beta) = 0$, 则称 α 与子空间 V_1 正交, 记为 $\alpha \perp V_1$. 如果对任意 $\alpha \in V_1, \beta \in V_2$ 都有 $(\alpha, \beta) = 0$, 则称子空间 V_1 与 V_2 正交, 记为 V_1

$\perp V_2$.

定理 1.6.8 设 V_1, V_2 是内积空间 V 中两个子空间, 若 V_1 与 V_2 正交, 则和 $V_1 + V_2$ 是直和.

证明 对任意 $\alpha \in V_1 \cap V_2$, 则 $\alpha \in V_1$. 因为对任意 $\beta \in V_2$ 都有 $(\alpha, \beta) = 0$, 而 $\alpha \in V_2$, 故 $(\alpha, \alpha) = 0$. 于是 $\alpha = 0$, 从而 $V_1 \cap V_2 = \{0\}$, 由定理 1.4.8 知 $V_1 + V_2$ 是直和. \square

定义 1.6.12 设 V_1, V_2 是内积空间 V 中两个子空间, 如果 V_1 与 V_2 正交, 则和 $V_1 + V_2$ 称为 V_1 与 V_2 的正交和, 记为 $V_1 \oplus V_2$.

定义 1.6.13 设 V_1 是内积空间 V 的一个子空间, V 中所有与 V_1 正交的向量所作成的集合记为 V_1^\perp , 即 $V_1^\perp = \{\alpha \in V \mid \alpha \perp V_1\}$, 则称 V_1^\perp 为 V_1 的正交补.

由定义容易验证 V_1^\perp 是 V 的一个子空间.

定理 1.6.9 设 V_1 是内积空间 V 的一个有限维子空间, 则存在 V_1 的惟一正交补 V_1^\perp 使得 $V = V_1 \oplus V_1^\perp$.

证明 设 $\dim(V_1) = m$, 并且 $\epsilon_1, \epsilon_2, \dots, \epsilon_m$ 是 V_1 的一组标准正交基. 对任意 $\alpha \in V$, 令

$$\alpha_1 = (\alpha, \epsilon_1)\epsilon_1 + (\alpha, \epsilon_2)\epsilon_2 + \dots + (\alpha, \epsilon_m)\epsilon_m, \alpha_2 = \alpha - \alpha_1$$

则 $\alpha_1 \in V_1$, 且

$$\begin{aligned} (\alpha_2, \epsilon_i) &= (\alpha, \epsilon_i) - (\alpha_1, \epsilon_i) = (\alpha, \epsilon_i) - \left(\sum_{j=1}^m (\alpha, \epsilon_j)\epsilon_j, \epsilon_i \right) \\ &= (\alpha, \epsilon_i) - (\alpha, \epsilon_i)(\epsilon_i, \epsilon_i) = 0, \quad i = 1, 2, \dots, m \end{aligned}$$

故 α_2 与 V_1 中每个向量都正交, 即 $\alpha_2 \perp V_1$, 从而 $\alpha_2 \in V_1^\perp$.

因为 $\alpha = \alpha_1 + \alpha_2$, 所以 $V = V_1 + V_1^\perp$, 并且由定理 1.6.8 知 $V = V_1 \oplus V_1^\perp$.

再证惟一性. 设 V_2, V_3 都是 V_1 的正交补, 则

$$V = V_1 \oplus V_2, \quad V = V_1 \oplus V_3$$

对任意 $\alpha \in V_2$, 有

$$\alpha = \alpha_1 + \alpha_3, \quad \alpha_1 \in V_1, \alpha_3 \in V_3$$

因为 $\alpha \perp \alpha_1$, 则

$$(\alpha, \alpha_1) = (\alpha_1, \alpha_1) + (\alpha_3, \alpha_1) = (\alpha_1, \alpha_1) = 0$$

于是 $\alpha_1 = 0, \alpha \in V_3$, 从而 $V_2 \subset V_3$. 同理可证 $V_3 \subset V_2$. 因此 $V_2 = V_3$. \square

定义 1.6.14 设 V_1 是内积空间 V 的一个子空间, $\alpha \in V$, 如果有 $\alpha_1 \in V_1, \alpha_2 \perp V_1$ 使得

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

则称 α_1 是 α 在 V_1 上的正交(直交)投影.

定理 1.6.10(投影定理) 设 V_1 是内积空间 V 的有限维子空间, 则对任意 $\alpha \in V$, α 在 V_1 上的正交投影存在并且惟一.

证明 由定理 1.6.9 知, 存在 V_1 的惟一正交补 V_1^\perp 使得 $V = V_1 \oplus V_1^\perp$, 则 V 中任一向量 α 可惟一地表示为

$$\alpha = \alpha_1 + \alpha_2, \alpha_1 \in V_1, \alpha_2 \in V_1^\perp \quad (1.6.16)$$

于是, α_1 是 α 在 V_1 上的正交投影并且是惟一的. \square

定义 1.6.15 设 V_1 是内积空间 V 的一个非空子集, $\alpha \in V$ 是给定的向量, 如果存在 $\alpha_1 \in V_1$ 满足如下等式

$$\|\alpha - \alpha_1\| = \inf_{\beta \in V_1} \|\alpha - \beta\| = d(\alpha, V_1) \quad (1.6.17)$$

则称 α_1 为 α 在 V_1 上的最佳逼近.

定理 1.6.11 设 V_1 是内积空间 V 的一个子空间, $\alpha \in V$ 是给定的向量, 则 $\alpha_1 \in V_1$ 为 α 在 V_1 上的最佳逼近的充分必要条件是 $\alpha - \alpha_1 \perp V_1$.

证明 必要性. 采用反证法. 设 $\alpha_1 \in V_1$ 为 α 在 V_1 上的最佳逼近, 但 $\alpha - \alpha_1$ 不正交于 V_1 , 则在 V_1 中至少有一向量 $\beta \neq 0$ 且 $\|\beta\| = 1$, 使得 $(\alpha - \alpha_1, \beta) = \delta \neq 0$. 令 $\gamma = \alpha_1 + \delta\beta$, 则 $\gamma \in V_1$, 且有

$$\|\alpha - \gamma\|^2 = (\alpha - \alpha_1 - \delta\beta, \alpha - \alpha_1 - \delta\beta) = \|\alpha - \alpha_1\|^2 - |\delta|^2$$

因为 $|\delta|^2 > 0$, 所以 $\|\alpha - \gamma\| < \|\alpha - \alpha_1\|$. 这与 $\alpha_1 \in V_1$ 是 α 在 V_1 上的最佳逼近相矛盾, 故 $\alpha - \alpha_1 \perp V_1$.

充分性. 如果 $\alpha_1 \in V_1$ 且 $\alpha - \alpha_1 \perp V_1$, 则对任意 $\beta \in V_1$, 有

$$\begin{aligned} \|\alpha - \beta\|^2 &= \|(\alpha - \alpha_1) + (\alpha_1 - \beta)\|^2 \\ &= \|\alpha - \alpha_1\|^2 + \|\alpha_1 - \beta\|^2 \geq \|\alpha - \alpha_1\|^2 \end{aligned} \quad (1.6.18)$$

上式表明 $\alpha_1 \in V_1$ 是 α 在 V_1 上的最佳逼近. \square

定理 1.6.11 给出了最佳逼近的特征, 但没有解决最佳逼近的存在性问题. 下面的两个定理将给出最佳逼近存在的充分条件.

定理 1.6.12 设 V_1 是内积空间 V 的一个 m 维子空间, 则 V 中任一向

量 α 在 V_1 上都有惟一的最佳逼近, 并且 α 在 V_1 上的最佳逼近是 α 在 V_1 上的正交投影.

证明 由定理 1.6.10 知, V 中任一向量 α 可惟一地表示为

$$\alpha = \alpha_1 + \alpha_2, \alpha_1 \in V_1, \alpha_2 \perp V_1$$

其中 α_1 是 α 在 V_1 上的正交投影. 因为 $\alpha_2 = \alpha - \alpha_1 \perp V_1$, 则由定理 1.6.11 知 α_1 是 α 在 V_1 上的最佳逼近. 又因为 α 在 V_1 上的正交投影 α_1 是惟一的, 因此 α 在 V_1 上的最佳逼近是惟一的. \square

下面讨论内积空间 V 中向量 α 在 m 维子空间 V_1 上最佳逼近的构造.

设 $\alpha_1, \alpha_2, \dots, \alpha_m$ 是 V_1 的一组基, 则对 $\beta \in V_1$ 有

$$\beta = x_1 \alpha_1 + x_2 \alpha_2 + \dots + x_m \alpha_m \quad (1.6.19)$$

根据定理 1.6.11, β 是向量 α 在 V_1 上的最佳逼近当且仅当 $\alpha - \beta \perp V_1$, 即

$$\alpha - \beta \perp \alpha_i, \quad i = 1, 2, \dots, m$$

从而得到

$$(\alpha - \beta, \alpha_i) = (\alpha, \alpha_i) - \sum_{j=1}^m x_j (\alpha_j, \alpha_i) = 0, \quad i = 1, 2, \dots, m$$

于是得到方程组

$$Ax = b \quad (1.6.20)$$

其中 $A = (G(\alpha_1, \alpha_2, \dots, \alpha_m))^T$, $x = (x_1, x_2, \dots, x_m)^T$, $b = ((\alpha, \alpha_1), (\alpha, \alpha_2), \dots, (\alpha, \alpha_m))^T$. 由定理 1.6.5 知矩阵 A 非奇异, 所以方程组 (1.6.20) 有惟一解 x , 从而得 α 在 V_1 上的惟一最佳逼近 $\beta = x_1 \alpha_1 + x_2 \alpha_2 + \dots + x_m \alpha_m$.

特别地, 当 $\alpha_1, \alpha_2, \dots, \alpha_m$ 是 V_1 的一组标准正交基时, 其 Gram 矩阵是单位矩阵, 则 α 在 V_1 上的惟一最佳逼近为 $\beta = (\alpha, \alpha_1) \alpha_1 + (\alpha, \alpha_2) \alpha_2 + \dots + (\alpha, \alpha_m) \alpha_m$.

*** 定理 1.6.13 (最佳逼近定理)** 设 V_1 是 n 维内积空间 V 的一个闭凸集, 则 V 中任一向量 α 在 V_1 上都有惟一的最佳逼近.

证明 由 $d(\alpha, V_1)$ 的定义知, 在 V_1 中存在点列 $\{\beta_k\}$ 使得

$$d(\alpha, V_1) \leq \|\alpha - \beta_k\| \leq d(\alpha, V_1) + \frac{1}{k}, \quad k = 1, 2, \dots \quad (1.6.21)$$

对任意正整数 m 和 n , 由 (1.6.5) 得

$$2\left\|\frac{\beta_m - \beta_n}{2}\right\|^2 = \|\beta_m - \alpha\|^2 + \|\beta_n - \alpha\|^2 - 2\left\|\frac{\beta_m + \beta_n}{2} - \alpha\right\|^2$$

因为 V_1 是凸集, 则 $\frac{\beta_m + \beta_n}{2} \in V_1$, $\left\|\frac{\beta_m + \beta_n}{2} - \alpha\right\| \geq d(\alpha, V_1)$. 从而由上式得

$$0 \leq 2\left\|\frac{\beta_m - \beta_n}{2}\right\|^2 \leq \|\beta_m - \alpha\|^2 + \|\beta_n - \alpha\|^2 - 2[d(\alpha, V_1)]^2$$

令 $m, n \rightarrow \infty$, 则 $\lim_{m, n \rightarrow \infty} \|\beta_m - \beta_n\| = 0$. 因为 V 是 n 维内积空间, 并且 V_1 是闭集, 所以存在 $\alpha_1 \in V_1$ 使得 $\beta_k \rightarrow \alpha_1 (k \rightarrow \infty)$. 于是 $\|\alpha - \alpha_1\| = \lim_{k \rightarrow \infty} \|\alpha - \beta_k\| = d(\alpha, V_1)$.

如果 V_1 中还有 α_0 使 $\|\alpha - \alpha_0\| = d(\alpha, V_1)$, 则由(1.6.5)得

$$\begin{aligned} 2\left\|\frac{\alpha_1 - \alpha_0}{2}\right\|^2 &= \|\alpha_1 - \alpha\|^2 + \|\alpha_0 - \alpha\|^2 - 2\left\|\frac{\alpha_1 + \alpha_0}{2} - \alpha\right\|^2 \\ &\leq 2[d(\alpha, V_1)]^2 - 2[d(\alpha, V_1)]^2 = 0 \end{aligned}$$

于是 $\alpha_0 = \alpha_1$, 即在 V_1 中使 $\|\alpha - \alpha_1\| = d(\alpha, V_1)$ 的向量 α_1 是惟一的. \square

定理 1.6.11 至定理 1.6.13 具有非常重要的应用价值. 下面我们看两个例子.

例 1.6.6(最小二乘问题) 在许多实际观测数据的处理问题中, 如果已知量 y 与量 x_1, x_2, \dots, x_n 之间呈线性关系

$$y = \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_n x_n \quad (1.6.22)$$

但不知道线性系数 $\lambda_1, \lambda_2, \dots, \lambda_n$. 为了确定这些系数, 通常做 $m \geq n$ 次试验, 得到 m 组观测数据

$$\begin{array}{cccc} 1 & 2 & \cdots & m \\ \hline x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(m)} \\ x_2^{(1)} & x_2^{(2)} & \cdots & x_2^{(m)} \\ \vdots & \vdots & & \vdots \\ x_n^{(1)} & x_n^{(2)} & \cdots & x_n^{(m)} \\ \hline y^{(1)} & y^{(2)} & \cdots & y^{(m)} \end{array}$$

按如下意义确定系数: 求 $\lambda_1, \lambda_2, \dots, \lambda_n$ 使得

$$\min_{\lambda_i \in \mathbf{P}} \sum_{j=1}^m \left| y^{(j)} - \sum_{i=1}^n c_i x_i^{(j)} \right|^2 = \sum_{j=1}^m \left| y^{(j)} - \sum_{i=1}^n \lambda_i x_i^{(j)} \right|^2 \quad (1.6.23)$$

记

$$a_i = \begin{bmatrix} x_i^{(1)} \\ x_i^{(2)} \\ \vdots \\ x_i^{(m)} \end{bmatrix}, i = 1, 2, \dots, n, b = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}, c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}, \lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix}$$

则(1.6.23)化为

$$\min_{c \in \mathbf{P}^n} \|b - \sum_{i=1}^n c_i a_i\|^2 = \|b - \sum_{i=1}^n \lambda_i a_i\|^2 \quad (1.6.24)$$

这个问题可看成是求 \mathbf{P}^m 中向量 b 在 $\text{span}(a_1, a_2, \dots, a_n)$ 上的最佳逼近. 如果记

$$A = [a_1, a_2, \dots, a_n]$$

由定理 1.6.11 知系数 $\lambda_1, \lambda_2, \dots, \lambda_n$ 满足

$$A^H A \lambda = A^H b \quad (1.6.25)$$

例 1.6.7 (函数的最佳平方逼近问题) 设 $f(x) \in C[a, b]$, $\varphi_1(x), \varphi_2(x), \dots, \varphi_n(x)$ 是 $C[a, b]$ 中线性无关的函数组, 求系数 a_1, a_2, \dots, a_n 使函数

$p(x) = \sum_{i=1}^n a_i \varphi_i(x)$ 逼近 $f(x)$ 时 $\int_a^b [f(x) - p(x)]^2 dx$ 最小.

解 因为 $C[a, b]$ 按例 1.6.4 定义的内积成为欧氏空间. 令

$$V_1 = \text{span}(\varphi_1(x), \varphi_2(x), \dots, \varphi_n(x))$$

则 V_1 是 $C[a, b]$ 的一个 n 维子空间,

$$\|f - p\|^2 = \int_a^b [f(x) - p(x)]^2 dx$$

于是问题转化为求 $C[a, b]$ 中向量 $f(x)$ 在 n 维子空间 V_1 上的最佳逼近 $p(x)$. 由定理 1.6.11 和定理 1.6.12 知, 这个问题的解存在且惟一, 并且 a_1, a_2, \dots, a_n 是如下线性方程组的解:

$$\begin{bmatrix} (\varphi_1, \varphi_1) & (\varphi_1, \varphi_2) & \cdots & (\varphi_1, \varphi_n) \\ (\varphi_2, \varphi_1) & (\varphi_2, \varphi_2) & \cdots & (\varphi_2, \varphi_n) \\ \vdots & \vdots & & \vdots \\ (\varphi_n, \varphi_1) & (\varphi_n, \varphi_2) & \cdots & (\varphi_n, \varphi_n) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} (f, \varphi_1) \\ (f, \varphi_2) \\ \vdots \\ (f, \varphi_n) \end{bmatrix} \quad (1.6.26)$$

这个例子是函数逼近论和数值分析中的基本问题.

习 题

1. 设 $A \subseteq B$, 求 $A \cup B, A \cap B$.
2. 设 A, B, C 是三个集合, 证明:
 - (1) $A \cap (B \cap C) = (A \cap B) \cap C$;
 - (2) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
3. 设 \mathbf{Z} 是整数集, 规定 \mathbf{Z} 的元素之间的关系 \mathbf{R}

$$aRb \Leftrightarrow n \mid (a - b), n > 0$$

证明 \mathbf{R} 是 \mathbf{Z} 上的一个等价关系.

4. 设 $f: A \rightarrow B, g: B \rightarrow C$. 证明: 若 f 和 g 都是单(满)映射, 则 $g \circ f$ 也是单(满)映射.
5. 设 $f: A \rightarrow B, g: B \rightarrow C$. 证明: 若 f 和 g 都是可逆映射, 则 $g \circ f$ 也是可逆映射, 并且有 $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

6. 检验以下集合对于所指的线性运算是否构成实数域 \mathbf{R} 上的线性空间:

- (1) 实数域 \mathbf{R} 上的全体 n 阶对称(反对称)矩阵, 对矩阵的加法和数量乘法;
- (2) 平面上不平行于某一向量的全体向量所组成的集合, 对向量的加法和数量乘法;
- (3) 实数域 \mathbf{R} 上次数等于 $n (n \geq 1)$ 的多项式全体, 对多项式的加法和数量乘法;
- (4) 全体实数对 $\{(a, b) \mid a, b \in \mathbf{R}\}$, 对于如下定义的加法 \oplus 和数量乘法 \circ :

$$(a_1, b_1) \oplus (a_2, b_2) = (a_1 + a_2, b_1 + b_2 + a_1 a_2)$$

$$k \circ (a_1, b_1) = (ka_1, kb_1 + \frac{k(k-1)}{2} a_1^2)$$

(5) 设 \mathbf{R}^+ 表示全体正实数, 加法 \oplus 和数量乘法 \circ 定义为

$$a \oplus b = ab$$

$$k \circ a = a^k$$

其中 $a, b \in \mathbf{R}^+, k \in \mathbf{R}$.

7. 设 V 是数域 \mathbf{P} 上的线性空间, 证明:

- (1) V 中零元素是惟一的;
- (2) $0\alpha = 0, k0 = 0, (-1)\alpha = -\alpha$;
- (3) 如果 $k\alpha = 0$, 那么 $k = 0$ 或者 $\alpha = 0$;
- (4) $k(\alpha - \beta) = k\alpha - k\beta$.

8. 试证: 在 $\mathbf{R}^{2 \times 2}$ 中矩阵

$$\alpha_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \alpha_2 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \alpha_3 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \quad \alpha_4 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

线性无关.

9. 设 V 为数域 \mathbf{P} 上的线性空间, 如果 V 中向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 可由向量组 $\beta_1, \beta_2, \dots, \beta_s$ 线性表示, 证明: $\text{rank}(\alpha_1, \alpha_2, \dots, \alpha_r) \leq \text{rank}(\beta_1, \beta_2, \dots, \beta_s)$.

10. 求下列线性空间的维数与一组基:

(1) $\mathbf{R}^{n \times n}$ 中全体对称(反对称)矩阵构成的实数域 \mathbf{R} 上的线性空间;

(2) 第6题(5)中的线性空间 \mathbf{R}^4 .

11. 在 \mathbf{R}^4 中, 求向量 $\alpha = (1, 2, 1, 1)^T$ 在基

$$\alpha_1 = (1, 1, 1, 1)^T, \alpha_2 = (1, 1, -1, -1)^T, \\ \alpha_3 = (1, -1, 1, -1)^T, \alpha_4 = (1, -1, -1, 1)^T$$

下的坐标.

12. 在 $\mathbf{R}^{2 \times 2}$ 中, 求矩阵 $A = \begin{pmatrix} 1 & 2 \\ -2 & 0 \end{pmatrix}$ 在基

$$\alpha_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \alpha_2 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \alpha_3 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \alpha_4 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

下的坐标.

13. 设 $\mathbf{R}[x]_n$ 是所有次数小于 n 的实系数多项式组成的线性空间, 求多项式 $f(x) = 3 + 2x^{n-1}$ 在基 $1, (x-1), (x-1)^2, \dots, (x-1)^{n-1}$ 下的坐标.

14. 在 \mathbf{R}^4 中, 求基

$$\alpha_1 = (1, 0, 0, 0)^T, \alpha_2 = (0, 1, 0, 0)^T, \alpha_3 = (0, 0, 1, 0)^T, \alpha_4 = (0, 0, 0, 1)^T$$

到基

$$\beta_1 = (2, 1, -1, 1)^T, \beta_2 = (0, 3, 1, 0)^T, \beta_3 = (5, 3, 2, 1)^T, \beta_4 = (6, 6, 1, 3)^T$$

的过渡矩阵, 确定向量 $\xi = (x_1, x_2, x_3, x_4)^T$ 在基 $\beta_1, \beta_2, \beta_3, \beta_4$ 下的坐标, 并求一非零向量, 使它在这两组基下有相同的坐标.

15. 试判定下列各子集哪些为线性空间 \mathbf{R}^n 的子空间:

(1) $W_1 = \{x = (x_1, x_2, \dots, x_n)^T \in \mathbf{R}^n \mid x_1 + x_2 + \dots + x_n = 0\};$

(2) $W_2 = \{x = (x_1, x_2, \dots, x_n)^T \in \mathbf{R}^n \mid x_1 x_2 \dots x_n = 0\};$

(3) $W_3 = \{x = (x_1, x_2, \dots, x_n)^T \in \mathbf{R}^n \mid x_1 + x_2 + \dots + x_n = 1\};$

(4) $W_4 = \{x = (x_1, x_2, \dots, x_n)^T \in \mathbf{R}^n \mid x_1^2 + x_2^2 + \dots + x_n^2 = 0\}.$

16. 设 $\mathbf{R}[x]$ 表示实数域上一元多项式全体构成的线性空间, 问下列子集是否构成 $\mathbf{R}[x]$ 的子空间:

(1) $\{p(x) \mid p(1) = 0\};$

(2) $\{p(x) \mid p(x) \text{ 的常数项为零}\};$

(3) $\{p(x) \mid p(-x) = p(x)\};$

(4) $\{p(x) \mid p(-x) = -p(x)\}.$

17. 在 \mathbf{R}^5 中, 求如下齐次线性方程组

$$\begin{cases} x_1 + x_2 - 3x_4 - x_5 = 0 \\ x_1 - x_2 + 2x_3 - x_4 = 0 \\ 4x_1 - 2x_2 + 6x_3 + 3x_4 - 4x_5 = 0 \\ 2x_1 + 4x_2 - 2x_3 + 4x_4 - 7x_5 = 0 \end{cases}$$

确定的解空间的维数和基.

18. 设 $A \in \mathbf{P}^{n \times n}$, $\mathbf{P}^{n \times n}$ 中全体与 A 可交换的矩阵记为 $W = \{X \in \mathbf{P}^{n \times n} \mid AX = XA\}$.

(1) 证明: W 是 $\mathbf{P}^{n \times n}$ 的一个子空间;

(2) 当 $A = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & n \end{bmatrix}$ 时, 求 W 的维数和一组基.

19. 求下列由向量 $\{\alpha_i\}$ 生成的子空间与由向量 $\{\beta_j\}$ 生成的子空间的交与和的维数和一组基:

$$(1) \begin{cases} \alpha_1 = (1, 2, 1, 0)^T, \\ \alpha_2 = (-1, 1, 1, 1)^T, \end{cases} \quad \begin{cases} \beta_1 = (2, -1, 0, 1)^T \\ \beta_2 = (1, -1, 3, 7)^T \end{cases}$$

$$(2) \begin{cases} \alpha_1 = (1, 0, 2, 1)^T, \\ \alpha_2 = (2, 0, 1, -1)^T, \\ \alpha_3 = (3, 0, 3, 0)^T, \end{cases} \quad \begin{cases} \beta_1 = (1, 1, 0, 1)^T \\ \beta_2 = (4, 1, 3, 1)^T \end{cases}$$

20. 证明: 如果 W 是有限维线性空间 V 的子空间, 且 $\dim(W) = \dim(V)$, 则 $W = V$.

21. 设 V_1 与 V_2 分别是齐次线性方程组

$$x_1 + x_2 + \cdots + x_n = 0$$

和

$$x_1 = x_2 = \cdots = x_n$$

的解空间, 证明: $\mathbf{R}^n = V_1 \perp V_2$.

22. 设 S, K 分别是 n 阶实对称矩阵和反对称矩阵的全体, 证明 S, K 均为线性空间 $\mathbf{R}^{n \times n}$ 的子空间, 并且 $\mathbf{R}^{n \times n} = S \perp K$.

23. 设 V_1, V_2, \dots, V_s 是数域 \mathbf{P} 上线性空间 V 的 s 个子空间, 证明下面的叙述等价:

(1) 和 $V_1 + V_2 + \cdots + V_s$ 是直和;

(2) 和 $V_1 \perp V_2 \perp \cdots \perp V_s$ 中零向量的表示法惟一;

(3) $V_i \cap \sum_{j \neq i} V_j = \{0\}$;

(4) $\dim(V_1 + V_2 + \cdots + V_s) = \dim(V_1) + \dim(V_2) + \cdots + \dim(V_s)$.

24. 设

$$V = \left\{ \begin{bmatrix} a & a+b \\ c & c \end{bmatrix} \mid a, b, c \in \mathbf{R} \right\}$$

作出线性空间 V 到 \mathbf{R}^3 的同构对应.

25. 在 \mathbf{R}^4 中求一单位向量与 $(1, 1, -1, 1)^T, (1, -1, -1, 1)^T, (2, 1, 1, 3)^T$ 都正交.

26. 设 $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5$ 是 \mathbf{R}^5 中一组标准正交基, $V = \text{span}\{\alpha_1, \alpha_2, \alpha_3\}$, 其中 $\alpha_1 = \varepsilon_1 + \varepsilon_5, \alpha_2 = \varepsilon_1 - \varepsilon_2 + \varepsilon_4, \alpha_3 = 2\varepsilon_1 + \varepsilon_2 + \varepsilon_3$, 求 V 的一组标准正交基.

27. 用 Gram-Schmidt 正交化方法, 将内积空间 V 的给定子集 S 正交化, 再找出 V 的标准正交基, 并求给定向量 α 在标准正交基下的坐标表达式:

(1) $V = \mathbf{R}^4, S = \{(1, 2, 2, -1)^T, (1, 1, -5, 3)^T, (3, 2, 8, -7)^T\}, \alpha = (3, 1, 1, -3)^T;$

(2) $V = \mathbf{R}^4, S = \{(2, 1, 3, -1)^T, (1, 4, 3, -3)^T, (1, 1, -6, 0)^T\}, \alpha = (2, 1, 3, -1)^T;$

(3) $V = \mathbf{R}[x],$ 定义内积为 $(f, g) = \int_{-1}^1 f(t)g(t)dt, S = \{1, x, x^2\}, \alpha = 1 + x.$

28. 设 $\alpha_1 = (1, 0, 2, 1)^T, \alpha_2 = (2, 1, 2, 3)^T, \alpha_3 = (0, 1, -2, 1)^T, V_1 = \text{span}(\alpha_1, \alpha_2, \alpha_3),$ 则 V_1 是欧氏空间 \mathbf{R}^4 的子空间, 求 V_1 的正交补 V_1^\perp 的一组基.

第二章 线性映射与线性变换

第一章研究了一个集合到另一个集合的映射和数域 \mathbf{P} 上线性空间的结构. 在许多数学问题和实际问题中起着重要作用的是线性空间到线性空间的映射, 并且这些映射有一个共同点, 即保持加法和数量乘法两种运算, 我们称这样的映射为线性映射. 本章我们讨论线性空间到线性空间的线性映射, 着重讨论线性空间 V 到自身的线性映射—线性变换, 并建立它们和矩阵之间的联系.

§ 2.1 线性映射及其矩阵表示

2.1.1 线性映射的定义及其性质

定义 2.1.1 设 V_1, V_2 是数域 \mathbf{P} 的两个线性空间, \mathcal{A} 是 V_1 到 V_2 的一个映射, 如果对 V_1 中任意两个向量 α, β 和任意数 $k \in \mathbf{P}$, 都有

$$\mathcal{A}(\alpha + \beta) = \mathcal{A}(\alpha) + \mathcal{A}(\beta)$$

$$\mathcal{A}(k\alpha) = k\mathcal{A}(\alpha)$$

则称 \mathcal{A} 是 V_1 到 V_2 的线性映射或线性算子.

与第一章第 5 节给出的线性空间 V_1 到 V_2 的同构映射相比, 线性映射比同构映射少了单映射和满映射这两条要求. 因此线性映射比同构映射更广泛. 线性空间 V_1 到 V_2 的线性映射也称为同态映射.

例 2.1.1 将线性空间 V_1 中每一个向量映射成线性空间 V_2 中零向量的映射是一个线性映射, 称为零映射, 记为 \mathcal{O} , 即

$$\mathcal{O}(\alpha) = 0, \forall \alpha \in V_1$$

例 2.1.2 线性空间 V 到自身的恒等映射是一个线性映射, 记为 \mathcal{I}_V , 即

$$\mathcal{I}_V(\alpha) = \alpha, \forall \alpha \in V$$

例 2.1.3 任意给定数 $k \in \mathbf{P}$, 数域 \mathbf{P} 上线性空间 V 到自身的一个映射

$$k(\alpha) = k\alpha, \forall \alpha \in V$$

是一个线性映射, 称为 V 上的由数 k 决定的数乘映射.

例 2.1.4 设 \mathcal{A} 是线性空间 V_1 到 V_2 的线性映射, 定义 V_1 到 V_2 的映

射 $-\mathcal{A}$ 如下:

$$(-\mathcal{A})(\alpha) = -\mathcal{A}(\alpha), \forall \alpha \in V_1$$

则 $-\mathcal{A}$ 是线性空间 V_1 到 V_2 的线性映射, 称为 \mathcal{A} 的负映射.

类似于定理 1.5.1, 线性映射具有如下性质.

定理 2.1.1 设 \mathcal{A} 是线性空间 V_1 到 V_2 的线性映射, 则

- (1) $\mathcal{A}(0) = 0$;
- (2) $\mathcal{A}(-\alpha) = -\mathcal{A}(\alpha), \forall \alpha \in V_1$;
- (3) 如果 $\alpha_1, \alpha_2, \dots, \alpha_m$ 是 V_1 的一组向量, $k_1, k_2, \dots, k_m \in \mathbf{P}$, 有

$$\mathcal{A}(k_1\alpha_1 + k_2\alpha_2 + \dots + k_m\alpha_m) = k_1\mathcal{A}(\alpha_1) + k_2\mathcal{A}(\alpha_2) + \dots + k_m\mathcal{A}(\alpha_m)$$

(4) 如果 $\alpha_1, \alpha_2, \dots, \alpha_m$ 是 V_1 的一组线性相关向量, 则 $\mathcal{A}(\alpha_1), \mathcal{A}(\alpha_2), \dots, \mathcal{A}(\alpha_m)$ 是 V_2 中的一组线性相关向量; 并且当且仅当 \mathcal{A} 是一一映射时, V_1 中线性无关向量组的像是 V_2 中的线性无关向量组.

证明 (1)至(3)的证明与定理 1.5.1 相同, 这里仅证(4).

如果 $\alpha_1, \alpha_2, \dots, \alpha_m$ 线性相关, 则存在不全为零的数 $k_1, k_2, \dots, k_m \in \mathbf{P}$ 使得

$$k_1\alpha_1 + k_2\alpha_2 + \dots + k_m\alpha_m = 0$$

由(1)和(3)有

$$\begin{aligned} & k_1\mathcal{A}(\alpha_1) + k_2\mathcal{A}(\alpha_2) + \dots + k_m\mathcal{A}(\alpha_m) \\ &= \mathcal{A}(k_1\alpha_1 + k_2\alpha_2 + \dots + k_m\alpha_m) = \mathcal{A}(0) = 0 \end{aligned}$$

故 $\mathcal{A}(\alpha_1), \mathcal{A}(\alpha_2), \dots, \mathcal{A}(\alpha_m)$ 线性相关.

如果线性映射 \mathcal{A} 是一一映射, 并且 $\alpha_1, \alpha_2, \dots, \alpha_m$ 是 V_1 中线性无关向量组, 则对任一组不全为零的数 $k_1, k_2, \dots, k_m \in \mathbf{P}$, 有 $k_1\alpha_1 + k_2\alpha_2 + \dots + k_m\alpha_m \neq 0$. 从而

$$k_1\mathcal{A}(\alpha_1) + k_2\mathcal{A}(\alpha_2) + \dots + k_m\mathcal{A}(\alpha_m) = \mathcal{A}(k_1\alpha_1 + k_2\alpha_2 + \dots + k_m\alpha_m) \neq 0$$

上式说明 $\mathcal{A}(\alpha_1), \mathcal{A}(\alpha_2), \dots, \mathcal{A}(\alpha_m)$ 线性无关.

反过来, 假如线性映射 \mathcal{A} 不是一一映射, 则存在 $\alpha, \beta \in V_1$ 且 $\alpha \neq \beta$, 但 $\mathcal{A}(\alpha) = \mathcal{A}(\beta)$, 即 $\alpha - \beta \neq 0$ 而 $\mathcal{A}(\alpha - \beta) = 0$. 这说明 V_1 中线性无关向量 $\alpha - \beta$ 的像 $\mathcal{A}(\alpha - \beta)$ 是 V_2 中线性相关的向量, 这与条件矛盾, 因此 \mathcal{A} 是一一映射. \square

如果 V_1 是 n 维线性空间, $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 是 V_1 的一组基, 则对任意 $\alpha \in V_1$ 有

$$\alpha = x_1 \epsilon_1 + x_2 \epsilon_2 + \cdots + x_n \epsilon_n$$

α 在线性映射 \mathcal{A} 下的像为

$$\mathcal{A}(\alpha) = x_1 \mathcal{A}(\epsilon_1) + x_2 \mathcal{A}(\epsilon_2) + \cdots + x_n \mathcal{A}(\epsilon_n)$$

这说明如果知道 V_1 的一组基在线性映射 \mathcal{A} 下的像, 则 V_1 中每一个向量在 \mathcal{A} 下的像也就确定了.

定理 2.1.2 设 \mathcal{A}, \mathcal{B} 是 n 维线性空间 V_1 到线性空间 V_2 的两个线性映射, 如果 $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ 是 V_1 的一组基, 并且 $\mathcal{A}(\epsilon_i) = \mathcal{B}(\epsilon_i) (i=1, 2, \cdots, n)$, 则 $\mathcal{A} = \mathcal{B}$.

证明 对任意 $\alpha \in V_1$, 有

$$\alpha = \sum_{i=1}^n x_i \epsilon_i$$

因为

$$\begin{aligned} \mathcal{A}(\alpha) &= \mathcal{A}\left(\sum_{i=1}^n x_i \epsilon_i\right) = \sum_{i=1}^n x_i \mathcal{A}(\epsilon_i) \\ &= \sum_{i=1}^n x_i \mathcal{B}(\epsilon_i) = \mathcal{B}\left(\sum_{i=1}^n x_i \epsilon_i\right) = \mathcal{B}(\alpha) \end{aligned}$$

所以 $\mathcal{A} = \mathcal{B}$. \square

定理 2.1.2 说明线性映射由基像组惟一确定.

定理 2.1.3 设 $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ 是 n 维线性空间 V_1 的一组基, $\alpha_1, \alpha_2, \cdots, \alpha_n$ 是线性空间 V_2 的任意 n 个向量, 则存在 V_1 到 V_2 的惟一线性映射 \mathcal{A} , 使得

$$\mathcal{A}(\epsilon_i) = \alpha_i, \quad i = 1, 2, \cdots, n$$

证明 对任意 $\alpha \in V_1$, 有

$$\alpha = x_1 \epsilon_1 + x_2 \epsilon_2 + \cdots + x_n \epsilon_n$$

令

$$\mathcal{A}(\alpha) = x_1 \alpha_1 + x_2 \alpha_2 + \cdots + x_n \alpha_n$$

因为 $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ 是 V_1 的一组基, 所以上式定义了 V_1 到 V_2 的一个映射 \mathcal{A} , 并且容易证明 \mathcal{A} 是线性映射.

因为

$$\epsilon_i = 0 \cdot \epsilon_1 + \cdots + 0 \cdot \epsilon_{i-1} + 1 \cdot \epsilon_i + 0 \cdot \epsilon_{i+1} + \cdots + 0 \cdot \epsilon_n$$

所以

$$\begin{aligned}\mathcal{A}(\varepsilon_i) &= 0 \cdot \alpha_1 + \cdots + 0 \cdot \alpha_{i-1} + 1 \cdot \alpha_i + 0 \cdot \alpha_{i+1} \\ &\quad + \cdots + 0 \cdot \alpha_n = \alpha_i (i = 1, 2, \cdots, n)\end{aligned}$$

由定理 2.1.2 知, 线性映射 \mathcal{A} 是惟一的. \square

2.1.2 线性映射的运算

设 V_1, V_2, V_3 都是数域 \mathbf{P} 上的线性空间, 我们把 V_1 到 V_2 的所有线性映射组成的集合记为 $\mathcal{L}(V_1, V_2)$. 类似地, $\mathcal{L}(V_2, V_3)$ 和 $\mathcal{L}(V_1, V_3)$ 分别表示 V_2 到 V_3 和 V_1 到 V_3 的所有线性映射组成的集合.

因为映射有乘法运算, 所以线性映射也有乘法运算. 由于线性空间有加法运算, 因此可以定义线性映射的加法运算.

设 $\mathcal{A}, \mathcal{B} \in \mathcal{L}(V_1, V_2)$, 定义它们的和 $\mathcal{A} + \mathcal{B}$ 为

$$(\mathcal{A} + \mathcal{B})(\alpha) = \mathcal{A}(\alpha) + \mathcal{B}(\alpha), \forall \alpha \in V_1$$

定理 2.1.4 (1) 设 $\mathcal{A}, \mathcal{B} \in \mathcal{L}(V_1, V_2)$, 则 $\mathcal{A} + \mathcal{B} \in \mathcal{L}(V_1, V_2)$;

(2) 设 $\mathcal{A} \in \mathcal{L}(V_1, V_2)$, $\mathcal{B} \in \mathcal{L}(V_2, V_3)$, 则 $\mathcal{B}\mathcal{A} \in \mathcal{L}(V_1, V_3)$.

证明 (1) 对任意 $\alpha, \beta \in V_1, k \in \mathbf{P}$, 有

$$\begin{aligned}(\mathcal{A} + \mathcal{B})(\alpha + \beta) &= \mathcal{A}(\alpha + \beta) + \mathcal{B}(\alpha + \beta) \\ &= \mathcal{A}(\alpha) + \mathcal{A}(\beta) + \mathcal{B}(\alpha) + \mathcal{B}(\beta) \\ &= (\mathcal{A}(\alpha) + \mathcal{B}(\alpha)) + (\mathcal{A}(\beta) + \mathcal{B}(\beta)) \\ &= (\mathcal{A} + \mathcal{B})(\alpha) + (\mathcal{A} + \mathcal{B})(\beta) \\ (\mathcal{A} + \mathcal{B})(k\alpha) &= \mathcal{A}(k\alpha) + \mathcal{B}(k\alpha) = k\mathcal{A}(\alpha) + k\mathcal{B}(\alpha) \\ &= k(\mathcal{A}(\alpha) + \mathcal{B}(\alpha)) = k(\mathcal{A} + \mathcal{B})(\alpha)\end{aligned}$$

因此 $\mathcal{A} + \mathcal{B}$ 是 V_1 到 V_2 的线性映射.

(2) 对任意 $\alpha, \beta \in V_1, k \in \mathbf{P}$, 有

$$\begin{aligned}(\mathcal{B}\mathcal{A})(\alpha + \beta) &= \mathcal{B}(\mathcal{A}(\alpha + \beta)) = \mathcal{B}(\mathcal{A}(\alpha) + \mathcal{A}(\beta)) \\ &= \mathcal{B}(\mathcal{A}(\alpha)) + \mathcal{B}(\mathcal{A}(\beta)) = (\mathcal{B}\mathcal{A})(\alpha) + (\mathcal{B}\mathcal{A})(\beta) \\ (\mathcal{B}\mathcal{A})(k\alpha) &= \mathcal{B}(\mathcal{A}(k\alpha)) = \mathcal{B}(k\mathcal{A}(\alpha)) \\ &= k(\mathcal{B}(\mathcal{A}(\alpha))) = k(\mathcal{B}\mathcal{A})(\alpha)\end{aligned}$$

则 $\mathcal{B}\mathcal{A}$ 是 V_1 到 V_3 的线性映射. \square

容易验证线性映射的加法适合交换律和结合律. 零映射 O 具有性质

$$\mathcal{A} + O = \mathcal{A}, \quad \forall \mathcal{A} \in \mathcal{L}(V_1, V_2)$$

并且对每一个 $\mathcal{A} \in \mathcal{L}(V_1, V_2)$, 它的负映射 $-\mathcal{A} \in \mathcal{L}(V_1, V_2)$ 满足

$$\mathcal{A} + (-\mathcal{A}) = O$$

设 $\mathcal{A}, \mathcal{B} \in \mathcal{L}(V_1, V_2)$, 线性映射的减法定义为

$$\mathcal{A} - \mathcal{B} = \mathcal{A} + (-\mathcal{B})$$

由于映射的乘法适合结合律, 因此线性映射的乘法也适合结合律.

利用线性映射的乘法和数乘映射可以定义线性映射的数量乘法. 设 $\mathcal{A} \in \mathcal{L}(V_1, V_2)$, $k \in \mathbf{P}$, 定义 k 与 \mathcal{A} 的数量乘积 $k\mathcal{A}$ 为

$$k\mathcal{A} = k\mathcal{A}$$

即

$$(k\mathcal{A})(\alpha) = k(\mathcal{A}(\alpha)) = k\mathcal{A}(\alpha), \quad \forall \alpha \in V_1$$

由定理 2.1.4 知, $k\mathcal{A} \in \mathcal{L}(V_1, V_2)$.

容易验证线性映射的数量乘法满足以下规则

$$1 \cdot \mathcal{A} = \mathcal{A}$$

$$k(m\mathcal{A}) = (km)\mathcal{A}$$

$$(k+m)\mathcal{A} = k\mathcal{A} + m\mathcal{A}$$

$$k(\mathcal{A} + \mathcal{B}) = k\mathcal{A} + k\mathcal{B}$$

对于线性映射, 我们已经定义乘法、加法和数量乘法 3 种运算. 由加法和数量乘法的性质知, $\mathcal{L}(V_1, V_2)$ 对上面定义的加法和数量乘法构成数域 \mathbf{P} 上的线性空间.

定理 2.1.5 设 $\mathcal{A} \in \mathcal{L}(V_1, V_2)$, 如果 \mathcal{A} 是可逆映射, 则 $\mathcal{A}^{-1} \in \mathcal{L}(V_2, V_1)$.

证明 因为 \mathcal{A} 可逆, 则

$$\mathcal{A}\mathcal{A}^{-1} = \mathcal{I}_{V_2}, \quad \mathcal{A}^{-1}\mathcal{A} = \mathcal{I}_{V_1}$$

于是对任意 $\alpha, \beta \in V_2, k \in \mathbf{P}$, 有

$$\begin{aligned} \mathcal{A}^{-1}(\alpha + \beta) &= \mathcal{A}^{-1}[(\mathcal{A}\mathcal{A}^{-1})(\alpha) + (\mathcal{A}\mathcal{A}^{-1})(\beta)] \\ &= \mathcal{A}^{-1}[\mathcal{A}(\mathcal{A}^{-1}(\alpha)) + \mathcal{A}(\mathcal{A}^{-1}(\beta))] \end{aligned}$$

$$\begin{aligned}
 &= \mathcal{A}^{-1}[\mathcal{A}(\mathcal{A}^{-1}(\alpha) + \mathcal{A}^{-1}(\beta))] \\
 &= (\mathcal{A}^{-1}\mathcal{A})(\mathcal{A}^{-1}(\alpha) + \mathcal{A}^{-1}(\beta)) \\
 &= \mathcal{A}^{-1}(\alpha) + \mathcal{A}^{-1}(\beta)
 \end{aligned}$$

$$\begin{aligned}
 \mathcal{A}^{-1}(k\alpha) &= \mathcal{A}^{-1}[k((\mathcal{A}\mathcal{A}^{-1})(\alpha))] = \mathcal{A}^{-1}[k(\mathcal{A}(\mathcal{A}^{-1}(\alpha)))] \\
 &= \mathcal{A}^{-1}[\mathcal{A}(k\mathcal{A}^{-1}(\alpha))] = (\mathcal{A}^{-1}\mathcal{A})(k\mathcal{A}^{-1}(\alpha)) = k\mathcal{A}^{-1}(\alpha)
 \end{aligned}$$

因此 $\mathcal{A}^{-1} \in \mathcal{L}(V_2, V_1)$. \square

由定理 1.1.5 知 \mathcal{A} 是线性空间 V_1 到 V_2 的可逆线性映射当且仅当 \mathcal{A} 是 V_1 到 V_2 的同构映射. 因此由定理 1.5.2 知, 有限维线性空间 V_1 到 V_2 的可逆线性映射存在的充分必要条件是 $\dim(V_1) = \dim(V_2)$.

2.1.3 线性映射的矩阵表示

为了利用矩阵来研究线性映射, 我们来建立线性映射与矩阵的关系.

设 V_1 是数域 \mathbf{P} 上的 n 维线性空间, $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 是 V_1 的一组基, V_2 是数域 \mathbf{P} 上的 m 维线性空间, $\eta_1, \eta_2, \dots, \eta_m$ 为 V_2 的一组基, \mathcal{A} 是 V_1 到 V_2 的一个线性映射.

由定理 2.1.2 知, 线性映射 \mathcal{A} 完全被它在 V_1 的基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 上的作用所决定, 即被 V_2 中的向量组 $\mathcal{A}(\epsilon_1), \mathcal{A}(\epsilon_2), \dots, \mathcal{A}(\epsilon_n)$ 所决定, 而 $\mathcal{A}(\epsilon_1), \mathcal{A}(\epsilon_2), \dots, \mathcal{A}(\epsilon_n)$ 完全被它们在基 $\eta_1, \eta_2, \dots, \eta_m$ 下的坐标所决定. 设

$$\begin{cases} \mathcal{A}(\epsilon_1) = a_{11}\eta_1 + a_{21}\eta_2 + \dots + a_{m1}\eta_m \\ \mathcal{A}(\epsilon_2) = a_{12}\eta_1 + a_{22}\eta_2 + \dots + a_{m2}\eta_m \\ \dots\dots\dots \\ \mathcal{A}(\epsilon_n) = a_{1n}\eta_1 + a_{2n}\eta_2 + \dots + a_{mn}\eta_m \end{cases} \quad (2.1.1)$$

上式可形式地记为

$$\mathcal{A}(\epsilon_1, \epsilon_2, \dots, \epsilon_n) = (\mathcal{A}(\epsilon_1), \mathcal{A}(\epsilon_2), \dots, \mathcal{A}(\epsilon_n)) = (\eta_1, \eta_2, \dots, \eta_m)A \quad (2.1.2)$$

其中

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

矩阵 A 称为线性映射 \mathcal{A} 在 V_1 的基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 和 V_2 的基 $\eta_1, \eta_2, \dots, \eta_m$ 下的矩阵

显然矩阵 A 由线性映射 \mathcal{A} 惟一确定; 反过来, 若给定 $m \times n$ 矩阵 A , 则由 (2.1.2) 可确定基像组 $\mathcal{A}(\epsilon_1), \mathcal{A}(\epsilon_2), \dots, \mathcal{A}(\epsilon_n)$, 由定理 2.1.3 知线性映射 \mathcal{A} 就完全确定. 这就是说, 在给定基的情况下, 线性空间 V_1 到 V_2 的线性映射 \mathcal{A} 与 $m \times n$ 矩阵 A 一一对应, 并且这种对应保持加法和数量乘法两种运算.

定理 2.1.6 设 V_1 是数域 P 上的 n 维线性空间, $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 是 V_1 的一组基, V_2 是数域 P 上的 m 维线性空间, $\eta_1, \eta_2, \dots, \eta_m$ 是 V_2 的一组基, 则 V_1 到 V_2 的每一个线性映射与它在基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 和基 $\eta_1, \eta_2, \dots, \eta_m$ 下的矩阵之间的对应 σ 是线性空间 $\mathcal{L}(V_1, V_2)$ 到 $P^{m \times n}$ 的同构映射, 从而 $\mathcal{L}(V_1, V_2)$ 与 $P^{m \times n}$ 同构.

证明 在 V_1 中取定一组基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 和 V_2 中取定一组基 $\eta_1, \eta_2, \dots, \eta_m$ 以后, 对 V_1 到 V_2 的每一个线性映射 \mathcal{A} 都有惟一确定的 $m \times n$ 矩阵 A 与之对应. 因此这个对应给出了 $\mathcal{L}(V_1, V_2)$ 到 $P^{m \times n}$ 的一个映射 σ . 上面的讨论说明 σ 是 $\mathcal{L}(V_1, V_2)$ 到 $P^{m \times n}$ 上的一一映射. 下面证明 σ 保持加法和数量乘法运算.

任取 $\mathcal{A}, \mathcal{B} \in \mathcal{L}(V_1, V_2)$, 设 $\sigma(\mathcal{A}) = A, \sigma(\mathcal{B}) = B$, 即

$$\mathcal{A}(\epsilon_1, \epsilon_2, \dots, \epsilon_n) = (\mathcal{A}(\epsilon_1), \mathcal{A}(\epsilon_2), \dots, \mathcal{A}(\epsilon_n)) = (\eta_1, \eta_2, \dots, \eta_m)A$$

$$\mathcal{B}(\epsilon_1, \epsilon_2, \dots, \epsilon_n) = (\mathcal{B}(\epsilon_1), \mathcal{B}(\epsilon_2), \dots, \mathcal{B}(\epsilon_n)) = (\eta_1, \eta_2, \dots, \eta_m)B$$

则

$$\begin{aligned} (\mathcal{A} + \mathcal{B})(\epsilon_1, \dots, \epsilon_n) &= ((\mathcal{A} + \mathcal{B})(\epsilon_1), \dots, (\mathcal{A} + \mathcal{B})(\epsilon_n)) \\ &= (\mathcal{A}(\epsilon_1) + \mathcal{B}(\epsilon_1), \dots, \mathcal{A}(\epsilon_n) + \mathcal{B}(\epsilon_n)) \\ &= (\mathcal{A}(\epsilon_1), \dots, \mathcal{A}(\epsilon_n)) + (\mathcal{B}(\epsilon_1), \dots, \mathcal{B}(\epsilon_n)) \\ &= (\eta_1, \dots, \eta_m)A + (\eta_1, \dots, \eta_m)B \\ &= (\eta_1, \dots, \eta_m)(A + B) \end{aligned}$$

即 $\mathcal{A} + \mathcal{B}$ 在基 $\epsilon_1, \dots, \epsilon_n$ 和基 η_1, \dots, η_m 下的矩阵是 $A + B$, 从而

$$\sigma(\mathcal{A} + \mathcal{B}) = A + B = \sigma(\mathcal{A}) + \sigma(\mathcal{B})$$

对任意 $k \in P$, 有

$$\begin{aligned} (k\mathcal{A})(\epsilon_1, \dots, \epsilon_n) &= ((k\mathcal{A})(\epsilon_1), \dots, (k\mathcal{A})(\epsilon_n)) \\ &= (k(\mathcal{A}(\epsilon_1)), \dots, k(\mathcal{A}(\epsilon_n))) \end{aligned}$$

$$\begin{aligned}
&= \left(\sum_{i=1}^m (ka_{i1}) \eta_i, \cdots, \sum_{i=1}^m (ka_{im}) \eta_i \right) \\
&= (\eta_1, \cdots, \eta_m)(kA)
\end{aligned}$$

即 $k\mathcal{A}$ 在基 $\varepsilon_1, \cdots, \varepsilon_n$ 和基 η_1, \cdots, η_m 下的矩阵是 kA , 从而

$$\sigma(k\mathcal{A}) = kA = k\sigma(\mathcal{A})$$

因此 σ 是 $\mathcal{L}(V_1, V_2)$ 到 $\mathbf{P}^{m \times n}$ 的同构映射, 从而 $\mathcal{L}(V_1, V_2)$ 与 $\mathbf{P}^{m \times n}$ 同构. \square

如果 $\dim(V_1) = n, \dim(V_2) = m$, 则由定理 1.5.2 知, 线性空间 $\mathcal{L}(V_1, V_2)$ 是有限维的, 并且 $\dim(\mathcal{L}(V_1, V_2)) = mn$. $\mathcal{L}(V_1, V_2)$ 中的一个线性映射可用一个 $m \times n$ 矩阵代表.

定理 2.1.7 设 \mathcal{A} 是 n 维线性空间 V_1 到 m 维线性空间 V_2 的一个线性映射, 并且 \mathcal{A} 在 V_1 的基 $\varepsilon_1, \cdots, \varepsilon_n$ 和 V_2 的基 η_1, \cdots, η_m 下的矩阵为 A . 对任意

$\alpha \in V_1$, 若 $\alpha = \sum_{i=1}^n x_i \varepsilon_i, \mathcal{A}(\alpha) = \sum_{i=1}^m y_i \eta_i$, 则

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad (2.1.3)$$

证明 对任意 $\alpha \in V_1$, 若

$$\alpha = \sum_{i=1}^n x_i \varepsilon_i$$

则

$$\begin{aligned}
\mathcal{A}(\alpha) &= \sum_{i=1}^n x_i \mathcal{A}(\varepsilon_i) = (\mathcal{A}(\varepsilon_1), \cdots, \mathcal{A}(\varepsilon_n)) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\
&= (\eta_1, \cdots, \eta_m) A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}
\end{aligned}$$

即 $A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ 是 $\mathcal{A}(\alpha)$ 在基 η_1, \cdots, η_m 下的坐标, 由于向量在同一组基下的坐标

是惟一的, 于是得 (2.1.3). \square

例 2.1.5 设 $\mathbf{R}[x]_n$ 表示实数域 \mathbf{R} 上次数小于 n 的一元多项式再添上零多项式构成的线性空间. 求 $\mathbf{R}[x]_n$ 到 $\mathbf{R}[x]_{n-1}$ 的线性映射 \mathcal{D}

$$\mathcal{D}(f(x)) = f'(x), \forall f(x) \in \mathbf{R}[x]_n$$

在 $\mathbf{R}[x]_n$ 的基 $1, x, \cdots, x^{n-1}$ 与 $\mathbf{R}[x]_{n-1}$ 的基 $1, x, \cdots, x^{n-2}$ 下的矩阵 D .

解 在 $\mathbf{R}[x]_n$ 中取基 $\epsilon_1 = 1, \epsilon_2 = x, \cdots, \epsilon_n = x^{n-1}$, 在 $\mathbf{R}[x]_{n-1}$ 中取基 $\eta_1 = 1, \eta_2 = x, \cdots, \eta_{n-1} = x^{n-2}$, 则

$$\begin{cases} \mathcal{D}(\epsilon_1) = 0 = 0\eta_1 + 0\eta_2 + \cdots + 0\eta_{n-1} \\ \mathcal{D}(\epsilon_2) = 1 = \eta_1 + 0\eta_2 + \cdots + 0\eta_{n-1} \\ \mathcal{D}(\epsilon_3) = 2x = 0\eta_1 + 2\eta_2 + 0\eta_3 + \cdots + 0\eta_{n-1} \\ \cdots \cdots \cdots \\ \mathcal{D}(\epsilon_n) = (n-1)x^{n-2} = 0\eta_1 + \cdots + 0\eta_{n-2} + (n-1)\eta_{n-1} \end{cases}$$

即

$$\mathcal{D}(\epsilon_1, \cdots, \epsilon_n) = (\eta_1, \cdots, \eta_{n-1}) \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & n-2 & 0 \\ 0 & 0 & 0 & \cdots & 0 & n-1 \end{pmatrix}_{(n-1) \times n}$$

于是 \mathcal{D} 在基 $1, x, \cdots, x^{n-1}$ 与 $1, x, \cdots, x^{n-2}$ 下的矩阵为

$$D = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & n-1 \end{pmatrix}_{(n-1) \times n}$$

如果在 $\mathbf{R}[x]_{n-1}$ 中取基 $\eta'_1 = 1, \eta'_2 = 2x, \cdots, \eta'_{n-1} = (n-1)x^{n-2}$, 则 \mathcal{D} 在基 $1, x, \cdots, x^{n-1}$ 与基 $1, 2x, \cdots, (n-1)x^{n-2}$ 下的矩阵为

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}_{(n-1) \times n}$$

由此可见,同一个线性映射在不同基下的矩阵一般是不同的.那么线性映射在不同基下的矩阵之间有什么关系?下面的定理将回答这个问题.

定理 2.1.8 设 \mathcal{A} 是 n 维线性空间 V_1 到 m 维线性空间 V_2 的一个线性映射, $\epsilon_1, \dots, \epsilon_n$ 和 $\epsilon'_1, \dots, \epsilon'_n$ 是 V_1 的两组基,由 $\epsilon_1, \dots, \epsilon_n$ 到 $\epsilon'_1, \dots, \epsilon'_n$ 的过渡矩阵为 Q , η_1, \dots, η_m 和 η'_1, \dots, η'_m 是 V_2 的两组基,由 η_1, \dots, η_m 到 η'_1, \dots, η'_m 的过渡矩阵为 P , \mathcal{A} 在基 $\epsilon_1, \dots, \epsilon_n$ 与基 η_1, \dots, η_m 下的矩阵为 A ,而在基 $\epsilon'_1, \dots, \epsilon'_n$ 与 η'_1, \dots, η'_m 下的矩阵为 B ,则

$$B = P^{-1}AQ$$

证明 因为

$$\mathcal{A}(\epsilon_1, \dots, \epsilon_n) = (\eta_1, \dots, \eta_m)A$$

$$\mathcal{A}(\epsilon'_1, \dots, \epsilon'_n) = (\eta'_1, \dots, \eta'_m)B$$

$$(\epsilon'_1, \dots, \epsilon'_n) = (\epsilon_1, \dots, \epsilon_n)Q$$

$$(\eta'_1, \dots, \eta'_m) = (\eta_1, \dots, \eta_m)P$$

并且 P 非奇异,则

$$\begin{aligned}\mathcal{A}(\epsilon'_1, \dots, \epsilon'_n) &= \mathcal{A}(\epsilon_1, \dots, \epsilon_n)Q \\ &= (\eta_1, \dots, \eta_m)AQ \\ &= (\eta'_1, \dots, \eta'_m)P^{-1}AQ\end{aligned}$$

因为线性映射 \mathcal{A} 的矩阵由基惟一确定,所以

$$B = P^{-1}AQ \quad \square$$

定义 2.1.2 设 $A, B \in P^{m \times n}$, 如果存在数域 P 上的 m 阶非奇异矩阵 P 和 n 阶非奇异矩阵 Q 使得

$$B = PAQ$$

则称 A 与 B 相抵(等价).

容易验证相抵关系是 $P^{m \times n}$ 上的一个等价关系.

由定理 2.1.8 和定义 2.1.2 可知,线性映射在不同基下的矩阵是相抵的.反过来,有下面定理.

定理 2.1.9 设 $A, B \in P^{m \times n}$, 如果 A 与 B 相抵,则它们可作为 n 维线性空间 V_1 到 m 维线性空间 V_2 的同一线性映射在两对基下所对应的矩阵.

证明 对 $A \in P^{m \times n}$, 则 A 可看作 V_1 到 V_2 的线性映射 \mathcal{A} 在 V_1 的基 $\epsilon_1, \dots, \epsilon_n$ 与 V_2 的基 η_1, \dots, η_m 下的矩阵. 因为 A 与 B 相抵,则存在 m 阶可

逆矩阵 P 和 n 阶可逆矩阵 Q 使得 $B = PAQ$. 令

$$(\epsilon_1, \cdots, \epsilon_n) = (\epsilon_1, \cdots, \epsilon_n)Q$$

$$(\eta_1, \cdots, \eta_m) = (\eta_1, \cdots, \eta_m)P^{-1}$$

则 $\epsilon_1, \cdots, \epsilon_n$ 和 η_1, \cdots, η_m 分别是 V_1 和 V_2 的基, 且线性映射 \mathcal{A} 在这对基下的矩阵恰为 B . \square

对一个线性映射, 能否选择一对基, 使它在这对基下的矩阵尽可能简单? 下面来解决这个问题.

由线性代数知道, 矩阵的初等变换与初等矩阵相对应, 并且一个矩阵可逆的充分必要条件是它能表示成一些初等矩阵的乘积. 因此 $m \times n$ 矩阵 A 与 B 相抵的充分必要条件是 A 可以经有限次初等变换变成 B .

设 $A \in \mathbf{P}^{m \times n}$, 并且 $\text{rank}(A) = r \geq 1$, 则通过一系列初等变换可将 A 化为

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \quad (2.1.4)$$

其中 I_r 是 r 阶单位矩阵, (2.1.4) 称为矩阵 A 的相抵标准形. 如果 $\text{rank}(A) = 0$, 则 A 的相抵标准形是零矩阵. 因此我们有如下结论.

定理 2.1.10 设 $A \in \mathbf{P}^{m \times n}$ 且 $\text{rank}(A) = r \geq 1$, 则存在 m 阶可逆矩阵 P 和 n 阶可逆矩阵 Q 使得

$$A = P \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} Q \quad (2.1.5)$$

由定理 2.1.8, 定理 2.1.9 和定理 2.1.10 可知, 对一个线性映射 \mathcal{A} 可以选取一对适当的基, 使它在这对基下的矩阵具有最简单的形式 (2.1.4).

定理 2.1.11 设 $A, B \in \mathbf{P}^{m \times n}$, 则 A 与 B 相抵的充分必要条件是它们有相同的秩.

证明 必要性. 设 A 与 B 相抵, 则 A 可经有限次初等变换化为 B , 因为初等变换不改变矩阵的秩, 所以 $\text{rank}(A) = \text{rank}(B)$.

充分性. 设 $\text{rank}(A) = \text{rank}(B) = r$. 如果 $r = 0$, 则 $A = B = 0$, 它们显然相抵. 若 $r \geq 1$, 则由定理 2.1.10 得, A 与 B 都相抵于矩阵 (2.1.4). 由相抵的对称性和传递性知 A 与 B 相抵. \square

§ 2.2 线性映射的值域与核

本节讨论伴随线性映射的两类重要子空间: 线性映射的值域与核.

定义 2.2.1 设 \mathcal{A} 是数域 \mathbf{P} 上线性空间 V_1 到 V_2 的线性映射, 令

$$R(\mathcal{A}) = I_m(\mathcal{A}) = \{\mathcal{A}(\alpha) \mid \alpha \in V_1\}$$

$$\text{Ker}(\mathcal{A}) = N(\mathcal{A}) = \{\alpha \in V_1 \mid \mathcal{A}(\alpha) = 0\}$$

称 $R(\mathcal{A})$ 是线性映射 \mathcal{A} 的**值域**; 而称 $\text{Ker}(\mathcal{A})$ 是线性映射 \mathcal{A} 的**核**.

定理 2.2.1 设 \mathcal{A} 是线性空间 V_1 到 V_2 的一个线性映射, 则

- (1) $R(\mathcal{A})$ 是 V_2 的一个子空间;
- (2) $\text{Ker}(\mathcal{A})$ 是 V_1 的一个子空间.

证明 (1) 显然 $R(\mathcal{A})$ 是 V_2 的非空子集. 对任意 $\mathcal{A}(\alpha), \mathcal{A}(\beta) \in R(\mathcal{A})$, $k \in \mathbf{P}$, 有

$$\mathcal{A}(\alpha) + \mathcal{A}(\beta) = \mathcal{A}(\alpha + \beta) \in R(\mathcal{A})$$

$$k\mathcal{A}(\alpha) = \mathcal{A}(k\alpha) \in R(\mathcal{A})$$

由定理 1.4.1 知, $R(\mathcal{A})$ 是 V_2 的一个子空间.

(2) 因为 $\mathcal{A}(0) = 0$, 所以 $\text{Ker}(\mathcal{A})$ 是 V_1 的非空子集. 对任意 $\alpha, \beta \in \text{Ker}(\mathcal{A})$, $k \in \mathbf{P}$, 有

$$\mathcal{A}(\alpha + \beta) = \mathcal{A}(\alpha) + \mathcal{A}(\beta) = 0$$

$$\mathcal{A}(k\alpha) = k\mathcal{A}(\alpha) = k0 = 0$$

则 $\alpha + \beta, k\alpha \in \text{Ker}(\mathcal{A})$. 因此 $\text{Ker}(\mathcal{A})$ 是 V_1 的一个子空间. \square

$\dim(R(\mathcal{A}))$ 称为 \mathcal{A} 的**秩**, 记为 $\text{rank}(\mathcal{A})$; $\dim(\text{Ker}(\mathcal{A}))$ 称为 \mathcal{A} 的**零度**.

定理 2.2.2 设 \mathcal{A} 是 n 维线性空间 V_1 到 m 维线性空间 V_2 的一个线性映射, $\epsilon_1, \dots, \epsilon_n$ 和 η_1, \dots, η_m 分别是 V_1 与 V_2 的基, \mathcal{A} 在这对基下的矩阵是 A , 则

$$(1) R(\mathcal{A}) = \text{span}(\mathcal{A}(\epsilon_1), \dots, \mathcal{A}(\epsilon_n));$$

$$(2) \text{rank}(\mathcal{A}) = \text{rank}(A);$$

$$(3) \dim(R(\mathcal{A})) + \dim(\text{Ker}(\mathcal{A})) = n.$$

证明 (1) 对任意 $\alpha \in V_1$, 有

$$\alpha = \sum_{i=1}^n x_i \epsilon_i$$

则

$$\mathcal{A}(\alpha) = \sum_{i=1}^n x_i \mathcal{A}(\epsilon_i)$$

故 $R(\mathcal{A}) = \text{span}(\mathcal{A}(\epsilon_1), \mathcal{A}(\epsilon_2), \dots, \mathcal{A}(\epsilon_n))$.

(2) 由(1)和定理 1.4.4, 有

$$\text{rank}(\mathcal{A}) = \text{rank}(\mathcal{A}(\epsilon_1), \mathcal{A}(\epsilon_2), \dots, \mathcal{A}(\epsilon_n))$$

另一方面, 矩阵 A 是由 $\mathcal{A}(\epsilon_1), \mathcal{A}(\epsilon_2), \dots, \mathcal{A}(\epsilon_n)$ 的坐标按列排成的. 如果在 V_2 中取定基 η_1, \dots, η_m , 把 V_2 的每一个向量与它的坐标对应起来, 则 V_2 与 \mathbf{P}^m 同构. 从而由定理 1.5.1 知, 基像组 $\mathcal{A}(\epsilon_1), \dots, \mathcal{A}(\epsilon_n)$ 与它们的坐标组 (即矩阵 A 的列向量组) 有相同的秩.

(3) 设 $\dim(\text{Ker}(\mathcal{A})) = r$. 在 $\text{Ker}(\mathcal{A})$ 中取一组基 $\alpha_1, \dots, \alpha_r$, 把它扩充成 V_1 的基 $\alpha_1, \dots, \alpha_r, \alpha_{r+1}, \dots, \alpha_n$, 则

$$\begin{aligned} R(\mathcal{A}) &= \text{span}(\mathcal{A}(\alpha_1), \dots, \mathcal{A}(\alpha_r), \mathcal{A}(\alpha_{r+1}), \dots, \mathcal{A}(\alpha_n)) \\ &= \text{span}(\mathcal{A}(\alpha_{r+1}), \dots, \mathcal{A}(\alpha_n)) \end{aligned}$$

现证明 $\mathcal{A}(\alpha_{r+1}), \dots, \mathcal{A}(\alpha_n)$ 线性无关.

事实上, 设

$$\sum_{j=r+1}^n k_j \mathcal{A}(\alpha_j) = 0$$

则 $\mathcal{A}(\sum_{j=r+1}^n k_j \alpha_j) = 0$, 从而 $\sum_{j=r+1}^n k_j \alpha_j \in \text{Ker}(\mathcal{A})$. 因此

$$\sum_{j=r+1}^n k_j \alpha_j = \sum_{i=1}^r k_i \alpha_i$$

因为 $\alpha_1, \dots, \alpha_n$ 线性无关, 所以 $k_i = 0 (i=1, \dots, r)$. 因此 $\mathcal{A}(\alpha_{r+1}), \dots, \mathcal{A}(\alpha_n)$ 线性无关. 于是

$$\dim(R(\mathcal{A})) = n - r$$

即

$$\dim(R(\mathcal{A})) + \dim(\text{Ker}(\mathcal{A})) = n. \quad \square$$

§ 2.3 线性变换

定义 2.3.1 设 V 是数域 \mathbf{P} 上的线性空间, V 到自身的线性映射称为 V 上的**线性变换**.

线性空间 V 到自身的数乘映射 k 称为数 k 决定的数乘变换. 当 $k=1$ 时, 得恒等变换; 当 $k=0$ 时, 得零变换.

因为线性变换是一类特殊的线性映射, 所以有关线性映射的性质, 对线性变换完全适用.

对线性空间 V 上所有线性变换组成的集合 $\mathcal{L}(V, V)$, 可在 $\mathcal{L}(V, V)$ 上定义线性变换的加法、乘法和数量乘法, 并且加法和乘法满足如下的运算规律:

- (1) 加法适合交换律和结合律;
- (2) 乘法适合结合律;
- (3) 乘法对加法有左、右分配律.

$\mathcal{L}(V, V)$ 关于线性变换的加法和数量乘法构成数域 P 上的线性空间.

设 \mathcal{A} 是线性空间 V 上的线性变换, 如果存在 V 的变换 \mathcal{B} 使得

$$\mathcal{A}\mathcal{B} = \mathcal{B}\mathcal{A} = \mathcal{I}_V$$

其中 \mathcal{I}_V (简记为 \mathcal{I}) 是 V 上的恒等变换, 则称 \mathcal{A} 是可逆的, 并称 \mathcal{B} 为 \mathcal{A} 的逆变换, 记作 \mathcal{A}^{-1} .

类似于定理 2.1.5, 可以证明: 如果线性变换 \mathcal{A} 是可逆的, 则 \mathcal{A}^{-1} 也是线性变换.

因为线性变换的乘法满足结合律, 所以可定义线性变换的幂:

$$\mathcal{A}^0 = \mathcal{I}, \mathcal{A}^n \equiv \underbrace{\mathcal{A}\mathcal{A}\cdots\mathcal{A}}_n$$

并且容易推出指数法则:

$$\mathcal{A}^m \cdot \mathcal{A}^n = \mathcal{A}^{m+n}, (\mathcal{A}^n)^m = \mathcal{A}^{nm}, \quad m, n \text{ 是非负整数}$$

如果 \mathcal{A} 可逆, 可定义 \mathcal{A} 的负整数幂:

$$\mathcal{A}^{-m} \equiv (\mathcal{A}^{-1})^m, \quad m \text{ 是正整数}$$

因为映射的乘法不适合交换律, 所以对线性变换 \mathcal{A} 和 \mathcal{B} , 一般说来

$$(\mathcal{A}\mathcal{B})^n \neq \mathcal{A}^n \mathcal{B}^n, \quad n \text{ 是正整数}$$

设 $f(x) = a_m x^m + a_{m-1} x^{m-1} + \cdots + a_1 x + a_0 \in P[x]$, \mathcal{A} 是 V 上的一个线性变换. 令

$$f(\mathcal{A}) \equiv a_m \mathcal{A}^m + a_{m-1} \mathcal{A}^{m-1} + \cdots + a_1 \mathcal{A} + a_0 \mathcal{I}$$

则 $f(\mathcal{A})$ 是 V 上的一个线性变换, 称为线性变换 \mathcal{A} 的多项式.

现在我们来讨论 n 维线性空间 V 上的线性变换与矩阵的关系. 设 \mathcal{A} 是线性空间 V 上的一个线性变换, 我们把上一节关于线性映射与矩阵的关系应用到 V 上的线性变换 \mathcal{A} . 这时, 只需在 V 中取一组基 $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$, 基向量的像 $\mathcal{A}(\epsilon_1), \mathcal{A}(\epsilon_2), \cdots, \mathcal{A}(\epsilon_n)$ 仍可用基线性表示

$$\begin{cases} \mathcal{A}(\epsilon_1) = a_{11}\epsilon_1 + a_{21}\epsilon_2 + \cdots + a_{n1}\epsilon_n \\ \mathcal{A}(\epsilon_2) = a_{12}\epsilon_1 + a_{22}\epsilon_2 + \cdots + a_{n2}\epsilon_n \\ \cdots \cdots \cdots \\ \mathcal{A}(\epsilon_n) = a_{1n}\epsilon_1 + a_{2n}\epsilon_2 + \cdots + a_{nn}\epsilon_n \end{cases}$$

或表示成

$$\mathcal{A}(\epsilon_1, \epsilon_2, \cdots, \epsilon_n) = (\mathcal{A}(\epsilon_1), \mathcal{A}(\epsilon_2), \cdots, \mathcal{A}(\epsilon_n)) = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n)A \quad (2.3.1)$$

其中

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \quad (2.3.2)$$

矩阵 A 称为线性变换 \mathcal{A} 在基 $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ 下的矩阵.

类似于定理 2.1.6, 有如下结论.

定理 2.3.1 设 V 是数域 \mathbf{P} 上的 n 维线性空间, $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ 是 V 的一组基, 则 V 上的每一个线性变换与它在基 $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ 下的矩阵之间的对应 σ 是线性空间 $\mathcal{L}(V, V)$ 到 $\mathbf{P}^{n \times n}$ 的同构映射, 从而 $\mathcal{L}(V, V)$ 与 $\mathbf{P}^{n \times n}$ 同构, 并且

- (1) 线性变换的乘积对应于矩阵的乘法;
- (2) 可逆的线性变换与可逆矩阵对应, 且逆变换对应于逆矩阵.

证明 第一部分的证明类似于定理 2.1.6, 下面证明(1)与(2).

设 \mathcal{A}, \mathcal{B} 是 V 上的两个线性变换, 它们在基 $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ 下的矩阵分别为 A, B , 即

$$\mathcal{A}(\epsilon_1, \epsilon_2, \cdots, \epsilon_n) = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n)A$$

$$\mathcal{B}(\epsilon_1, \epsilon_2, \cdots, \epsilon_n) = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n)B$$

则

$$\begin{aligned} (\mathcal{A}\mathcal{B})(\epsilon_1, \epsilon_2, \cdots, \epsilon_n) &= \mathcal{A}(\mathcal{B}(\epsilon_1, \epsilon_2, \cdots, \epsilon_n)) \\ &= \mathcal{A}[(\epsilon_1, \epsilon_2, \cdots, \epsilon_n)B] \\ &= [\mathcal{A}(\epsilon_1, \epsilon_2, \cdots, \epsilon_n)]B \end{aligned}$$

$$= (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) \mathbf{A} \mathbf{B}$$

于是线性变换 $\mathcal{A}\mathcal{B}$ 在基 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 下的矩阵为 $\mathbf{A}\mathbf{B}$.

因为恒等变换对应于单位矩阵, 所以等式

$$\mathcal{A}\mathcal{B} = \mathcal{B}\mathcal{A} = \mathcal{I}$$

与等式

$$\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A} = \mathbf{I}$$

相对应, 从而可逆线性变换与可逆矩阵对应, 且逆变换对应逆矩阵. \square

类似于定理 2.1.8, 可证下面定理.

定理 2.3.2 设 n 维线性空间 V 上线性变换 \mathcal{A} 在基 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 和 $\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_n$ 下的矩阵分别为 \mathbf{A} 和 \mathbf{B} , 由基 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 到基 $\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_n$ 的过渡矩阵为 \mathbf{P} , 则

$$\mathbf{B} = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$$

定义 2.3.2 设 $\mathbf{A}, \mathbf{B} \in \mathbf{P}^{n \times n}$, 如果存在可逆矩阵 $\mathbf{P} \in \mathbf{P}^{n \times n}$ 使得

$$\mathbf{B} = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$$

则称 \mathbf{A} 与 \mathbf{B} 相似.

容易证明相似关系是 $\mathbf{P}^{n \times n}$ 上的一个等价关系.

定理 2.3.2 说明, 线性变换在不同基下所对应的矩阵是相似的. 反过来, 类似于定理 2.1.9, 可以证明两个相似的矩阵要看作同一线性变换在两组基下所对应的矩阵.

线性变换理论要研究的一个重要问题是选取线性空间的一组基使线性变换在这组基下的矩阵具有较简单的形式. 我们将在本章的第 4 节至第 6 节和下一章讨论这个问题.

§ 2.4 特征值和特征向量

线性变换的特征值和特征向量是重要的概念, 它们不仅对线性变换的研究具有基本的重要性, 而且在物理、力学和工程技术中具有实际的意义.

定义 2.4.1 设 \mathcal{A} 是数域 \mathbf{P} 上线性空间 V 的一个线性变换, 如果存在 $\lambda \in \mathbf{P}$ 以及非零向量 $\alpha \in V$ 使得

$$\mathcal{A}(\alpha) = \lambda \alpha \quad (2.4.1)$$

则称 λ 为 \mathcal{A} 的特征值, 并称 α 为 \mathcal{A} 的属于 (或对应于) 特征值 λ 的特征向量.

下面讨论特征值和特征向量的求法.

设 V 是数域 \mathbf{P} 上的 n 维线性空间, $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 是 V 的一组基, 线性变换 \mathcal{A} 在这组基下的矩阵为 A . 如果 λ 是 \mathcal{A} 的特征值, α 是相应的特征向量, 则

$$\alpha = (\epsilon_1, \epsilon_2, \dots, \epsilon_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad (2.4.2)$$

把它代入(2.4.1), 并利用定理 2.1.7 得

$$(\epsilon_1, \epsilon_2, \dots, \epsilon_n) A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n) (\lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix})$$

由于 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 线性无关, 所以

$$A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad (2.4.3)$$

这说明特征向量 α 的坐标 $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ 满足齐次线性方程组

$$(\lambda I - A)x = 0 \quad (2.4.4)$$

因为 $\alpha \neq 0$, 所以 $x \neq 0$, 即齐次线性方程组(2.4.4)有非零解. 方程组(2.4.4)有非零解的充分必要条件是它的系数矩阵行列式为零, 即

$$|\lambda I - A| = 0 \quad (2.4.5)$$

定义 2.4.2 设 A 是数域 \mathbf{P} 上的 n 阶矩阵, λ 是一个文字, 矩阵 $\lambda I - A$ 称为 A 的特征矩阵, 其行列式 $|\lambda I - A|$ 称为 A 的特征多项式. 方程 $|\lambda I - A| = 0$ 称为 A 的特征方程, 它的根称为 A 的特征根(或特征值). 以 A 的特征值

λ 代入齐次线性方程组 (2.4.4) 所得的非零解 x 称为 A 对应于 λ 的特征向量.

如果 λ 是线性变换 \mathcal{A} 的特征值, 则 λ 是矩阵 A 的特征值; 反过来, 如果 λ 是矩阵 A 的特征值, 即 $|\lambda I - A| = 0$, 则齐次线性方程组 (2.4.4) 有非零解 x , 从而非零向量

$$\alpha = x_1 \epsilon_1 + x_2 \epsilon_2 + \cdots + x_n \epsilon_n$$

满足 (2.4.1), 即 λ 是线性变换 \mathcal{A} 的特征值, α 是属于特征值 λ 的一个特征向量. 因此, 线性变换的特征值、特征向量的性质可由矩阵的特征值、特征向量的性质得到.

例 2.4.1 设线性变换 \mathcal{A} 在基 $\epsilon_1, \epsilon_2, \epsilon_3$ 下的矩阵是 $A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$,

求 \mathcal{A} 的特征值与特征向量.

解 矩阵 A 的特征多项式为

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

所以矩阵 A (即线性变换 \mathcal{A}) 的特征值是 $\lambda_1 = -1$ (二重) 和 $\lambda_2 = 3$.

对应于特征值 $\lambda_1 = -1$, 齐次线性方程组 $(\lambda_1 I - A)x = 0$ 的基础解系为

$$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

因此, 线性变换 \mathcal{A} 属于特征值 -1 的两个线性无关的特征向量为 $\alpha_1 = \epsilon_1 - \epsilon_2, \alpha_2 = \epsilon_3$.

对应于特征值 $\lambda_2 = 3$, 齐次线性方程组 $(\lambda_2 I - A)x = 0$ 的基础解系为

$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, 因此线性变换 \mathcal{A} 属于特征值 3 的一个线性无关的特征向量是 $\alpha_3 = \epsilon_1 + \epsilon_2$.

例 2.4.2 在 n 维线性空间 $R[x]_n$ 中, 线性变换

$$\mathcal{Q}(f(x)) = f'(x)$$

在基 $1, x, \frac{x^2}{2!}, \dots, \frac{x^{n-1}}{(n-1)!}$ 下的矩阵是

$$D = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

D 的特征多项式为

$$|\lambda I - D| = \begin{vmatrix} \lambda & -1 & 0 & \cdots & 0 \\ 0 & \lambda & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & -1 \\ 0 & 0 & 0 & \cdots & \lambda \end{vmatrix} = \lambda^n$$

所以矩阵 D (即线性变换 \mathcal{D}) 的特征值是 $\lambda = 0$ (n 重), 相应的齐次线性方程组 $(\lambda I - D)x = 0$ 的基础解系为 $(1, 0, 0, \dots, 0)^T$. 因此, 线性变换 \mathcal{D} 属于特征值 0 的线性无关特征向量是任一非零常数.

对线性空间 V 上线性变换 \mathcal{A} 的任一特征值 λ , 所有满足

$$\mathcal{A}(\alpha) = \lambda\alpha$$

的向量 α 所组成的集合, 也就是 \mathcal{A} 的属于特征值 λ 的全部特征向量再添上零向量所组成的集合, 记为 V_λ , 即

$$V_\lambda = \{\alpha \mid \mathcal{A}(\alpha) = \lambda\alpha, \alpha \in V\}$$

对矩阵 $A \in \mathbb{C}^{n \times n}$, λ 是 A 的一个特征值, 我们也记

$$V_\lambda = \{x \mid Ax = \lambda x, x \in \mathbb{C}^n\}$$

则 V_λ 是 V (或 \mathbb{C}^n) 的一个子空间, 称 V_λ 为 \mathcal{A} (或 A) 的属于 λ 的特征子空间. 显然 $\dim(V_\lambda)$ 就是属于 λ 的线性无关特征向量的最大数目, 称 $\dim(V_\lambda)$ 为特征值 λ 的几何重数.

对 n 阶矩阵 $A = (a_{ij})$, 由行列式的性质可得 A 的特征多项式是 λ 的 n 次多项式. 确切地, 有如下结果.

定理 2.4.1 设 $A = (a_{ij}) \in \mathbb{P}^{n \times n}$, 则

$$|\lambda I - A| = \lambda^n + \sum_{k=1}^n (-1)^k b_k \lambda^{n-k} \quad (2.4.6)$$

其中 $b_k (k=1, 2, \dots, n)$ 是 A 的所有 k 阶主子式之和, 特别地

$$b_1 = \text{tr}(A), b_n = |A| \quad (2.4.7)$$

证明 记 $I = [e_1, e_2, \dots, e_n]$, $A = [a_1, a_2, \dots, a_n]$, 其中 e_i 和 a_i 分别是 I 和 A 的第 i 列, 则

$$|\lambda I - A| = |\lambda e_1 - a_1, \lambda e_2 - a_2, \dots, \lambda e_n - a_n|$$

利用行列式的性质, 将上式右端拆成每列是 λe_i 或 $-a_i$ 的行列式, 例如

$$\begin{aligned} |\lambda I - A| &= |\lambda e_1, \lambda e_2 - a_2, \dots, \lambda e_n - a_n| \\ &+ |[-a_1, \lambda e_2 - a_2, \dots, \lambda e_n - a_n]| \end{aligned}$$

于是有

$$\begin{aligned} |\lambda I - A| &= \lambda^n |e_1, e_2, \dots, e_n| + \lambda^{n-1} \sum_{i=1}^n |e_1, \dots, e_{i-1}, a_i, e_{i+1}, \dots, e_n| \\ &+ \dots + (-1)^k \lambda^{n-k} \sum_{\substack{1 \leq i_1 < \dots < i_k \leq n \\ 2 \leq k \leq n-1}} |[\dots, a_{i_1}, \dots, a_{i_k}, \dots]| + \dots + (-1)^n |A| \end{aligned}$$

行列式 $|[\dots, a_{i_1}, \dots, a_{i_k}, \dots]|$ 中第 i_1 列, \dots , 第 i_k 列依次是 a_{i_1}, \dots, a_{i_k} , 其余的列是单位矩阵 I 的相应列. 例如 $n=5, k=3, i_1=1, i_2=2, i_3=4$, 则

$$|a_1, a_2, e_3, a_4, e_5| = \begin{vmatrix} a_{11} & a_{12} & 0 & a_{14} & 0 \\ a_{21} & a_{22} & 0 & a_{24} & 0 \\ a_{31} & a_{32} & 1 & a_{34} & 0 \\ a_{41} & a_{42} & 0 & a_{44} & 0 \\ a_{51} & a_{52} & 0 & a_{54} & 1 \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{14} \\ a_{21} & a_{22} & a_{24} \\ a_{41} & a_{42} & a_{44} \end{vmatrix}$$

是 A 的一个 3 阶主子式. 因此, b_3 是关于 $1 \leq i_1 < i_2 < i_3 \leq 5$ 求和, 即 A 的所有 3 阶主子式之和.

对一般 n 阶矩阵 A , 同理可得 $|[\dots, a_{i_1}, \dots, a_{i_k}, \dots]|$ 是 A 的一个 k 阶主子式, 因此 b_k 是 A 的所有 k 阶主子式之和. \square

由代数基本定理(n 次多项式方程在复数域内有且仅有 n 个根(重根按重数计算))知, n 阶矩阵 A 在复数域内恰有 n 个特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$, 其中 λ_i 作为特征方程的根的重数, 称为 λ_i 的代数重数, 记为 $m_{\lambda_i}(A)$. 矩阵 A 的特

征值的全体称为 \mathbf{A} 的谱, 记为 $\lambda(\mathbf{A})$. 矩阵 \mathbf{A} 的特征值的最大模称为 \mathbf{A} 的谱半径, 记为 $\rho(\mathbf{A})$.

由根与系数的关系以及定理 2.4.1, 直接得到下面定理.

定理 2.4.2 设 $\mathbf{A} = (a_{ij}) \in \mathbb{C}^{n \times n}$, $\lambda_1, \lambda_2, \dots, \lambda_n$ 是 \mathbf{A} 的特征值, 则

$$\sum_{i=1}^n \lambda_i = \text{tr}(\mathbf{A}) \quad (2.4.8)$$

$$\prod_{i=1}^n \lambda_i = |\mathbf{A}|. \quad (2.4.9)$$

例 2.4.3 设

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_n \\ 1 & 0 & 0 & \cdots & 0 & -a_{n-1} \\ 0 & 1 & 0 & \cdots & 0 & -a_{n-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -a_1 \end{pmatrix}$$

矩阵 \mathbf{A} 在控制论中称为友矩阵或相伴矩阵, 求 \mathbf{A} 的特征多项式.

解 记

$$d_i = \begin{vmatrix} \lambda & 0 & 0 & \cdots & a_i \\ -1 & \lambda & 0 & \cdots & a_{i-1} \\ 0 & 1 & \lambda & \cdots & a_{i-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & -1 & \lambda + a_1 \end{vmatrix}, \quad i \geq 1, d_0 = 1$$

对 d_i 按第 1 行展开, 有

$$d_i = \lambda d_{i-1} + a_i, i \geq 1$$

由上式逐次递推得

$$\begin{aligned} d_n &= |\lambda \mathbf{I} - \mathbf{A}| = \lambda d_{n-1} + a_n = \lambda(\lambda d_{n-2} + a_{n-1}) + a_n \\ &= \lambda^2(\lambda d_{n-3} + a_{n-2}) + a_{n-1}\lambda + a_n \\ &= \lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + \cdots + a_{n-1}\lambda + a_n \end{aligned}$$

例 2.4.4 设 $x, y \in \mathbb{C}^n$, $\mathbf{A} = xy^H$, 求 \mathbf{A} 的特征多项式.

解 因为 $\text{rank}(A) \leq \text{rank}(x) \leq 1$, 所以 A 的所有 $k (k \geq 2)$ 阶子式全为零, 而 A 的一阶主子式之和为 $\text{tr}(xy^H) = y^H x$. 由定理 2.4.1, 得

$$|\lambda I - xy^H| = \lambda^n - y^H x \lambda^{n-1} = \lambda^{n-1}(\lambda - y^H x)$$

在 n 维线性空间 V 中取定一组基之后, 线性变换的特征值就是它在这组基下矩阵的特征值. 随着基的不同, 线性变换的矩阵一般是不同的, 但这些矩阵是相似的. 对于相似矩阵有如下定理.

定理 2.4.3 如果 n 阶矩阵 A 与 B 相似, 则

- (1) A 与 B 有相同的特征多项式;
- (2) A 与 B 有相同的特征值;
- (3) $\text{tr}(A) = \text{tr}(B)$.

证明 (1) 因为 A 与 B 相似, 所以存在可逆矩阵 P 使 $B = P^{-1}AP$, 于是

$$\begin{aligned} |\lambda I - B| &= |\lambda I - P^{-1}AP| = |P^{-1}(\lambda I - A)P| \\ &= |P^{-1}| |\lambda I - A| |P| = |\lambda I - A| \end{aligned}$$

由(1)即得(2), 而由(2)及定理 2.4.2 可直接得(3). \square

例 2.4.5 设 $A \in \mathbb{C}^{m \times n}$, $B \in \mathbb{C}^{n \times m}$, 试证:

$$\lambda^n |\lambda I_m - AB| = \lambda^m |\lambda I_n - BA|$$

证明 容易验证

$$\begin{bmatrix} I_m & -A \\ 0 & I_n \end{bmatrix} \begin{bmatrix} AB & 0 \\ B & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ B & BA \end{bmatrix} \begin{bmatrix} I_m & -A \\ 0 & I_n \end{bmatrix}$$

因为 $\begin{bmatrix} I_m & -A \\ 0 & I_n \end{bmatrix}$ 可逆, 则由上式知 $\begin{bmatrix} AB & 0 \\ B & 0 \end{bmatrix}$ 与 $\begin{bmatrix} 0 & 0 \\ B & BA \end{bmatrix}$ 相似, 从而由定理 2.4.3 有

$$\left| \lambda I_{m+n} - \begin{bmatrix} AB & 0 \\ B & 0 \end{bmatrix} \right| = \left| \lambda I_{m+n} - \begin{bmatrix} 0 & 0 \\ B & BA \end{bmatrix} \right|$$

即

$$\left| \begin{bmatrix} \lambda I_m - AB & 0 \\ -B & \lambda I_n - BA \end{bmatrix} \right| = \left| \begin{bmatrix} \lambda I_m & 0 \\ -B & \lambda I_n - BA \end{bmatrix} \right|$$

从而

$$\lambda^n |\lambda I_m - AB| = \lambda^m |\lambda I_n - BA|$$

由例 2.4.5 可知, m 阶方阵 AB 与 n 阶方阵 BA 具有相同的非零特征值, 从而由定理 2.4.3 有 $\text{tr}(AB) = \text{tr}(BA)$. 特别地, 若 A, B 为同阶方阵, 则 AB 与 BA 具有相同的特征值.

关于线性变换或矩阵的特征向量有如下结论.

定理 2.4.4 设 $\lambda_1, \lambda_2, \dots, \lambda_r$ 是线性变换 \mathcal{A} (或矩阵 A) 的 r 个互不相同的特征值, $\alpha_i (i=1, 2, \dots, r)$ 是对应于 λ_i 的特征向量, 则 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性无关.

证明 对特征值的个数 r 作数学归纳法. 当 $r=1$ 时, 由于特征向量非零, 所以单个特征向量必线性无关.

假设属于 r 个不同特征值的特征向量线性无关, 现证明属于 $r+1$ 个不同特征值 $\lambda_1, \lambda_2, \dots, \lambda_r, \lambda_{r+1}$ 的特征向量 $\alpha_1, \dots, \alpha_r, \alpha_{r+1}$ 也线性无关.

设有关系式

$$k_1 \alpha_1 + \dots + k_r \alpha_r + k_{r+1} \alpha_{r+1} = 0 \quad (2.4.10)$$

上式两边同乘 λ_{r+1} , 有

$$k_1 \lambda_{r+1} \alpha_1 + \dots + k_r \lambda_{r+1} \alpha_r + k_{r+1} \lambda_{r+1} \alpha_{r+1} = 0$$

(2.4.10) 两边同时作用 \mathcal{A} , 得

$$k_1 \lambda_1 \alpha_1 + \dots + k_r \lambda_r \alpha_r + k_{r+1} \lambda_{r+1} \alpha_{r+1} = 0$$

则得

$$k_1 (\lambda_{r+1} - \lambda_1) \alpha_1 + \dots + k_r (\lambda_{r+1} - \lambda_r) \alpha_r = 0$$

由归纳假设, $\alpha_1, \dots, \alpha_r$ 线性无关, 则

$$k_i (\lambda_{r+1} - \lambda_i) = 0, i = 1, 2, \dots, r$$

因为 $\lambda_{r+1} - \lambda_i \neq 0 (i \leq r)$, 所以 $k_i = 0 (i = 1, 2, \dots, r)$, 从而 (2.4.10) 变成

$$k_{r+1} \alpha_{r+1} = 0$$

又因为 $\alpha_{r+1} \neq 0$, 所以 $k_{r+1} = 0$, 因此 $\alpha_1, \dots, \alpha_r, \alpha_{r+1}$ 线性无关. \square

定理 2.4.4 还可推广为下面的定理.

定理 2.4.5 设 $\lambda_1, \dots, \lambda_r$ 是线性变换 \mathcal{A} (或矩阵 A) 的 r 个互异特征值, $\alpha_1^{(1)}, \dots, \alpha_{k_1}^{(1)}$ 是属于特征值 λ_1 的 k_1 个线性无关特征向量 ($i = 1, \dots, r$), 则 $\alpha_1^{(1)}, \dots, \alpha_{k_1}^{(1)}, \alpha_1^{(2)}, \dots, \alpha_{k_2}^{(2)}, \dots, \alpha_1^{(r)}, \dots, \alpha_{k_r}^{(r)}$ 线性无关.

这个定理的证明与定理 2.4.4 的证明类似, 也是对 r 作数学归纳法.

线性变换 \mathcal{A} (或矩阵 A) 的任一特征值 λ_i 的几何重数 $\dim(V_{\lambda_i})$ 与代数重

数 $m_{\lambda_i}(\mathcal{A})$ 之间有如下关系.

定理 2.4.6 线性变换 \mathcal{A} 的任一特征值的几何重数不超过它的代数重数.

证明 设 \mathcal{A} 是 n 维线性空间 V 上的线性变换, λ_i 是 \mathcal{A} 的任一特征值, 且 $\dim(V_{\lambda_i}) = k$. 取 V_{λ_i} 的一组基 $\alpha_1, \dots, \alpha_k$, 把它扩充成 V 的一组基 $\alpha_1, \dots, \alpha_k, \alpha_{k+1}, \dots, \alpha_n$, 则

$$\begin{cases} \mathcal{A}(\alpha_j) = \lambda_i \alpha_j, & j = 1, 2, \dots, k \\ \mathcal{A}(\alpha_{k+1}) = a_{1,k+1} \alpha_1 + \dots + a_{k,k+1} \alpha_k + a_{k+1,k+1} \alpha_{k+1} + \dots + a_{n,k+1} \alpha_n \\ \dots\dots\dots \\ \mathcal{A}(\alpha_n) = a_{1n} \alpha_1 + \dots + a_{kn} \alpha_k + a_{k+1,n} \alpha_{k+1} + \dots + a_{nn} \alpha_n \end{cases}$$

则 \mathcal{A} 在基 $\alpha_1, \dots, \alpha_k, \alpha_{k+1}, \dots, \alpha_n$ 下的矩阵为

$$\mathbf{A} = \begin{pmatrix} \lambda_i & & & a_{1,k+1} & \dots & a_{1n} \\ & \ddots & & \vdots & & \vdots \\ & & \lambda_i & a_{k,k+1} & \dots & a_{kn} \\ & & & a_{k+1,k+1} & \dots & a_{k+1,n} \\ & 0 & & \vdots & & \vdots \\ & & & a_{n,k+1} & \dots & a_{nn} \end{pmatrix}$$

从而, \mathcal{A} 的特征多项式为

$$f(\lambda) = |\lambda \mathbf{I} - \mathbf{A}| = (\lambda - \lambda_i)^k |\lambda \mathbf{I} - \mathbf{A}_{22}|$$

其中

$$\mathbf{A}_{22} = \begin{pmatrix} a_{k+1,k+1} & \dots & a_{k+1,n} \\ \vdots & & \vdots \\ a_{n,k+1} & \dots & a_{nn} \end{pmatrix}$$

故 λ_i 作为特征方程的根其重数不小于 k , 即线性变换 \mathcal{A} 的特征值 λ_i 的几何重数不超过其代数重数. \square

§ 2.5 矩阵的相似对角形

对角矩阵是矩阵中最简单的一种. 本节研究对于数域 \mathbf{P} 上 n 维线性空间

V 上一个线性变换 \mathcal{A} , 是否存在一组基使得 \mathcal{A} 在该组基下的矩阵是对角矩阵.

定义 2.5.1 设 \mathcal{A} 是数域 P 上 n 维线性空间 V 上的一个线性变换, 如果 V 中存在一组基, 使得 \mathcal{A} 在这组基下的矩阵是对角矩阵, 则称 \mathcal{A} 是可对角化的.

由于线性变换 \mathcal{A} 在不同基下的矩阵是相似的, 所以 \mathcal{A} 是否可对角化, 用矩阵语言可叙述为: 对矩阵 $A \in P^{n \times n}$, 是否存在 n 阶可逆矩阵 P , 使得

$$P^{-1}AP = A = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ & & \ddots \\ 0 & & & \lambda_n \end{pmatrix} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

定义 2.5.2 如果 n 阶矩阵 A 与对角矩阵相似, 则称矩阵 A 是可对角化的.

定理 2.5.1 数域 P 上 n 维线性空间 V 上的一个线性变换 \mathcal{A} 可对角化的充分必要条件是 \mathcal{A} 有 n 个线性无关的特征向量.

证明 设 \mathcal{A} 在基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 下的矩阵是对角矩阵 $\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, 则

$$\mathcal{A}(\epsilon_i) = \lambda_i \epsilon_i, \quad i = 1, 2, \dots, n$$

于是 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 是 \mathcal{A} 的 n 个线性无关特征向量.

反过来, 如果 \mathcal{A} 有 n 个线性无关特征向量 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$, 则取 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 为 V 的一组基, 并且 \mathcal{A} 在这组基下的矩阵是对角矩阵. \square

类似地, 有下面的定理.

定理 2.5.2 矩阵 $A \in P^{n \times n}$ 可对角化的充分必要条件是 A 有 n 个线性无关的特征向量.

由定理 2.4.4 和定理 2.5.1 直接得.

定理 2.5.3 如果数域 P 上 n 维线性空间 V 上的线性变换 \mathcal{A} (或 n 阶矩阵 A) 有 n 个不同的特征值, 则线性变换 \mathcal{A} (或矩阵 A) 是可对角化的.

定理 2.5.4 设数域 P 上 n 维线性空间 V 上线性变换 \mathcal{A} 的互异特征值为 $\lambda_1, \lambda_2, \dots, \lambda_r$, 则 \mathcal{A} 可对角化的充分必要条件是

$$V = V_{\lambda_1} \dot{+} V_{\lambda_2} \dot{+} \dots \dot{+} V_{\lambda_r}$$

证明 必要性. 设 \mathcal{A} 可对角化, 在 V_{λ_i} 取一组基 $\alpha_1^{(i)}, \dots, \alpha_{k_i}^{(i)} (i=1, 2, \dots, r)$, 则由定理 2.4.5 知, $\alpha_1^{(1)}, \dots, \alpha_{k_1}^{(1)}, \dots, \alpha_1^{(r)}, \dots, \alpha_{k_r}^{(r)}$ 是 \mathcal{A} 的最大线性无关特

征向量组. 因为 \mathcal{A} 可对角化, 由定理 2.5.1 知, $k_1 + k_2 + \cdots + k_r = n$. 从而 $\alpha_1^{(1)}, \cdots, \alpha_{k_1}^{(1)}, \cdots, \alpha_1^{(r)}, \cdots, \alpha_{k_r}^{(r)}$ 是 V 的一组基. 因此 $V = V_{\lambda_1} \dot{+} V_{\lambda_2} \dot{+} \cdots \dot{+} V_{\lambda_r}$.

充分性. 设 $V = V_{\lambda_1} \dot{+} V_{\lambda_2} \dot{+} \cdots \dot{+} V_{\lambda_r}$, 在 $V_{\lambda_i} (i=1, 2, \cdots, r)$ 中取一组基 $\alpha_1^{(i)}, \cdots, \alpha_{k_i}^{(i)}$, 则 $\alpha_1^{(1)}, \cdots, \alpha_{k_1}^{(1)}, \cdots, \alpha_1^{(r)}, \cdots, \alpha_{k_r}^{(r)}$ 是 $V_{\lambda_1} \dot{+} V_{\lambda_2} \dot{+} \cdots \dot{+} V_{\lambda_r}$ 的一组基, 即 V 的一组基, 这说明 V 中存在由 \mathcal{A} 的特征向量组成的一组基, 从而 \mathcal{A} 可对角化. \square

定理 2.5.5 数域 P 上 n 维线性空间 V 上的线性变换 \mathcal{A} 可对角化的充分必要条件是 \mathcal{A} 的每一个特征值的几何重数等于代数重数.

证明 设 \mathcal{A} 的互异特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_r$, λ_i 的代数重数为 m_{λ_i} , 几何重数为 $k_i (i=1, 2, \cdots, r)$, 则

$$m_{\lambda_1} + m_{\lambda_2} + \cdots + m_{\lambda_r} = n$$

如果 $m_{\lambda_i} = k_i (i=1, 2, \cdots, r)$, 则由定理 2.4.5 和定理 2.5.1 知, \mathcal{A} 是可对角化的.

反过来, 如果 \mathcal{A} 是可对角化的, 则由定理 2.5.1 知

$$k_1 + k_2 + \cdots + k_r = n$$

由定理 2.4.6 有

$$k_i \leq m_{\lambda_i}, \quad i = 1, 2, \cdots, r$$

从而

$$n = k_1 + k_2 + \cdots + k_r \leq m_{\lambda_1} + m_{\lambda_2} + \cdots + m_{\lambda_r} = n$$

故得 $k_i = m_{\lambda_i} (i=1, 2, \cdots, r)$. \square

对于例 2.4.1 中的 3 维线性空间 V 上的线性变换 \mathcal{A} , 它有 3 个线性无关的特征向量, 因此 \mathcal{A} 可对角化. 对于例 2.4.2 中 n 维线性空间 $\mathbf{R}[x]_n$ 上的线性变换 \mathcal{B} , 其零特征值的几何重数小于代数重数, 因此 \mathcal{B} 是不可对角化的.

§ 2.6 线性变换的不变子空间

这一节讨论线性变换与子空间的关系, 介绍线性变换的不变子空间概念, 由此说明线性变换的矩阵化简与线性变换的内在联系.

定义 2.6.1 设 \mathcal{A} 是数域 P 上线性空间 V 的线性变换, W 是 V 的子空间, 如果对任意 $\alpha \in W$, 都有 $\mathcal{A}(\alpha) \in W$, 则称 W 是 \mathcal{A} 的不变子空间.

例如, 线性空间 V 的任何一个子空间都是数乘变换的不变子空间, 这是

因为子空间对于数量乘法是封闭的.

例 2.6.1 整个线性空间 V 和零子空间 $\{0\}$, 对于每个线性变换 \mathcal{A} 而言都是 \mathcal{A} 的不变子空间. 称 V 和 $\{0\}$ 为 \mathcal{A} 的平凡不变子空间.

类似于线性映射的值域与核, 对线性空间 V 上的线性变换 \mathcal{A} , 可定义 \mathcal{A} 的值域 $R(\mathcal{A})$ 和核 $\text{Ker}(\mathcal{A})$ 如下:

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

$$\text{Ker}(\mathcal{A}) = \{\alpha \in V \mid \mathcal{A}(\alpha) = 0\}$$

由定理 2.2.1 知, 线性变换 \mathcal{A} 的值域 $R(\mathcal{A})$ 和核 $\text{Ker}(\mathcal{A})$ 都是 V 的子空间.

例 2.6.2 试证线性空间 V 上线性变换 \mathcal{A} 的值域 $R(\mathcal{A})$ 与核 $\text{Ker}(\mathcal{A})$ 以及 \mathcal{A} 的特征子空间都是 \mathcal{A} 的不变子空间.

证明 任取 $\alpha \in R(\mathcal{A})$, 因为 $\mathcal{A}(\alpha) \in R(\mathcal{A})$, 所以 $R(\mathcal{A})$ 是 \mathcal{A} 的不变子空间.

任取 $\alpha \in \text{Ker}(\mathcal{A})$, 因为 $\mathcal{A}(\alpha) \in \text{Ker}(\mathcal{A})$, 所以 $\text{Ker}(\mathcal{A})$ 是 \mathcal{A} 的不变子空间.

设 V_λ 是 \mathcal{A} 的任一特征子空间, 任取 $\alpha \in V_\lambda$, 因为 $\mathcal{A}(\alpha) = \lambda\alpha \in V_\lambda$, 所以 V_λ 是 \mathcal{A} 的不变子空间.

定理 2.6.1 线性变换 \mathcal{A} 的不变子空间的和与交都是 \mathcal{A} 的不变子空间.

证明 设 V_1, V_2, \dots, V_m 都是 \mathcal{A} 的不变子空间. 在 $V_1 + V_2 + \dots + V_m$ 中任取一向量 $\alpha_1 + \alpha_2 + \dots + \alpha_m$, 其中 $\alpha_i \in V_i (i=1, 2, \dots, m)$, 则

$$\mathcal{A}(\alpha_1 + \alpha_2 + \dots + \alpha_m) = \mathcal{A}(\alpha_1) + \mathcal{A}(\alpha_2) + \dots + \mathcal{A}(\alpha_m)$$

因为 $\mathcal{A}(\alpha_i) \in V_i (i=1, 2, \dots, m)$, 所以 $\mathcal{A}(\alpha_1 + \alpha_2 + \dots + \alpha_m) \in V_1 + V_2 + \dots + V_m$. 因此 $\sum_{i=1}^m V_i$ 是 \mathcal{A} 的不变子空间.

任取 $\alpha \in \bigcap_{i=1}^m V_i$, 则 $\alpha \in V_i (i=1, 2, \dots, m)$. 因为 $\mathcal{A}(\alpha) \in V_i (i=1, 2, \dots, m)$. 所以 $\mathcal{A}(\alpha) \in \bigcap_{i=1}^m V_i$. 因此 $\bigcap_{i=1}^m V_i$ 是 \mathcal{A} 的不变子空间. \square

下面给出线性空间 V 的有限维子空间 W 是 \mathcal{A} 的不变子空间的一个判别法则.

定理 2.6.2 设线性空间 V 的子空间 $W = \text{span}(\alpha_1, \dots, \alpha_m)$, 则 W 是线性变换 \mathcal{A} 的不变子空间的充分必要条件是 $\mathcal{A}(\alpha_i) \in W (i=1, 2, \dots, m)$.

证明 必要性是显然的. 现在来证充分性. 对任意 $\alpha \in W$, 则

$$\alpha = \sum_{i=1}^m k_i \alpha_i$$

从而 $\mathcal{A}(\alpha) = \sum_{i=1}^m k_i \mathcal{A}(\alpha_i) \in W$, 所以 W 是 \mathcal{A} 的不变子空间. \square

线性变换 \mathcal{A} 的一维不变子空间与 \mathcal{A} 的特征向量有密切的关系.

定理 2.6.3 设 \mathcal{A} 是线性空间 V 上的线性变换, 如果 W 是 \mathcal{A} 的一维不变子空间, 则 W 中任何一个非零向量都是 \mathcal{A} 的特征向量; 反之, 若 α 是 \mathcal{A} 的一个特征向量, 则 $\text{span}(\alpha)$ 是 \mathcal{A} 的一维不变子空间.

证明 设 W 是 \mathcal{A} 的一维不变子空间, 对任意 $\alpha \in W$ 且 $\alpha \neq 0$, 则 α 是 W 的一个基. 因为 W 是 \mathcal{A} 的不变子空间, 所以 $\mathcal{A}(\alpha) \in W$, 从而存在 $\lambda \in \mathbf{P}$ 使 $\mathcal{A}(\alpha) = \lambda\alpha$. 因此 α 是 \mathcal{A} 的特征向量.

反过来, 设 α 是 \mathcal{A} 的一个特征向量, 即 $\mathcal{A}(\alpha) = \lambda\alpha$. 在 $\text{span}(\alpha)$ 中任取一个向量 $k\alpha$, 有

$$\mathcal{A}(k\alpha) = k\mathcal{A}(\alpha) = k(\lambda\alpha) = (k\lambda)\alpha \in \text{span}(\alpha)$$

所以 $\text{span}(\alpha)$ 是 \mathcal{A} 的一维不变子空间. \square

现在我们利用线性变换 \mathcal{A} 的不变子空间来研究 \mathcal{A} 的矩阵化简.

定理 2.6.4 设 \mathcal{A} 是数域 \mathbf{P} 上 n 维线性空间 V 的线性变换, 则 \mathcal{A} 可对角化的充分必要条件是 V 可以分解成 \mathcal{A} 的一维不变子空间的直和.

证明 必要性. 设 \mathcal{A} 可对角化, 则由定理 2.5.1 知, \mathcal{A} 有 n 个线性无关特征向量 $\alpha_1, \alpha_2, \dots, \alpha_n$, 它们组成 V 的一组基, 因此

$$V = \text{span}(\alpha_1) \dot{+} \text{span}(\alpha_2) \dot{+} \dots \dot{+} \text{span}(\alpha_n)$$

由定理 2.6.3 知, $\text{span}(\alpha_i) (i=1, 2, \dots, n)$ 是 \mathcal{A} 的一维不变子空间.

充分性. 设 V 可分解成 \mathcal{A} 的一维不变子空间 $W_i (i=1, 2, \dots, n)$ 的直和.

$$V = W_1 \dot{+} W_2 \dot{+} \dots \dot{+} W_n$$

在 W_i 中取一个基 α_i , 则由定理 2.6.3 知, $\alpha_i (i=1, 2, \dots, n)$ 是 \mathcal{A} 的特征向量, 所以 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 V 的一组基, 由定理 2.5.1 知, \mathcal{A} 可对角化. \square

定理 2.6.5 设 \mathcal{A} 是数域 \mathbf{P} 上 n 维线性空间 V 的线性变换, 则 \mathcal{A} 在 V 的一组基下的矩阵为形如 $\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$ 的块上三角矩阵的充分必要条件是 \mathcal{A} 有非平凡的不变子空间.

证明 必要性. 设 \mathcal{A} 在 V 的一组基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 下的矩阵 $A = (a_{ij})$ 为如下块上三角矩阵

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \quad (2.6.1)$$

其中 $A_{11} \in \mathbf{P}^{m \times m} (0 < m < n)$. 由 $\mathcal{A}(\alpha_1, \alpha_2, \dots, \alpha_n) = (\alpha_1, \alpha_2, \dots, \alpha_n)A$, 则对

$1 \leq j \leq m$, 有

$$\mathcal{A}(\alpha_j) = a_{1j}\alpha_1 + a_{2j}\alpha_2 + \cdots + a_{mj}\alpha_m \quad (2.6.2)$$

令 $W = \text{span}(\alpha_1, \alpha_2, \cdots, \alpha_m)$, 则 W 是 V 的一个 m 维子空间, 并且由 (2.6.2) 可见, $\mathcal{A}(\alpha_j) \in W (j=1, 2, \cdots, m)$. 因此由定理 2.6.2 知, W 是 \mathcal{A} 的 m 维不变子空间. 因为 $0 < m < n$, 所以 W 是 \mathcal{A} 的非平凡不变子空间.

充分性. 设 W 是 \mathcal{A} 的非平凡不变子空间, 且 $\dim(W) = m$, 则 $0 < m < n$. 在 W 中取一组基 $\alpha_1, \alpha_2, \cdots, \alpha_m$, 把它扩充成 V 的一组基 $\alpha_1, \cdots, \alpha_m, \alpha_{m+1}, \cdots, \alpha_n$. 因为 $\mathcal{A}(\alpha_j) \in W (j=1, 2, \cdots, m)$, 则

$$\left\{ \begin{array}{l} \mathcal{A}(\alpha_1) = a_{11}\alpha_1 + a_{21}\alpha_2 + \cdots + a_{m1}\alpha_m \\ \cdots \cdots \cdots \\ \mathcal{A}(\alpha_m) = a_{1m}\alpha_1 + a_{2m}\alpha_2 + \cdots + a_{mm}\alpha_m \\ \mathcal{A}(\alpha_{m+1}) = a_{1,m+1}\alpha_1 + a_{2,m+1}\alpha_2 + \cdots \\ \quad + a_{m,m+1}\alpha_m + a_{m+1,m+1}\alpha_{m+1} + \cdots + a_{n,m+1}\alpha_n \\ \cdots \cdots \cdots \\ \mathcal{A}(\alpha_n) = a_{1n}\alpha_1 + a_{2n}\alpha_2 + \cdots \\ \quad + a_{m,n}\alpha_m + a_{m+1,n}\alpha_{m+1} + \cdots + a_{nn}\alpha_n \end{array} \right.$$

因此

$$\mathcal{A}(\alpha_1, \alpha_2, \cdots, \alpha_n) = (\alpha_1, \alpha_2, \cdots, \alpha_n) \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ 0 & \mathbf{A}_{22} \end{bmatrix}$$

其中

$$\mathbf{A}_{11} = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mm} \end{bmatrix}, \mathbf{A}_{12} = \begin{bmatrix} a_{1,m+1} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m,m+1} & \cdots & a_{m,n} \end{bmatrix}$$

$$\mathbf{A}_{22} = \begin{bmatrix} a_{m+1,m+1} & \cdots & a_{m+1,n} \\ \vdots & & \vdots \\ a_{n,m+1} & \cdots & a_{nn} \end{bmatrix}$$

即 \mathcal{A} 在基 $\alpha_1, \cdots, \alpha_m, \alpha_{m+1}, \cdots, \alpha_n$ 下的矩阵是形如 (2.6.1) 的块上三角矩阵. \square

定理 2.6.6 设 \mathcal{A} 是数域 \mathbf{P} 上 n 维线性空间 V 上的线性变换, 则 \mathcal{A} 在 V

的一组基下的矩阵为块对角矩阵的充分必要条件是 V 能分解成 \mathcal{A} 的若干个非平凡不变子空间的直和.

证明 必要性. 设 \mathcal{A} 在 V 的一组基 $\alpha_1^{(1)}, \dots, \alpha_{k_1}^{(1)}, \dots, \alpha_1^{(r)}, \dots, \alpha_{k_r}^{(r)}$ ($k_1 + k_2 + \dots + k_r = n$) 下的矩阵是块对角矩阵

$$\begin{pmatrix} \mathbf{A}_1 & & & \\ & \mathbf{A}_2 & & \\ & & \ddots & \\ & & & \mathbf{A}_r \end{pmatrix}$$

其中 \mathbf{A}_i ($i=1, \dots, r$) 是 k_i 阶矩阵. 令

$$W_i = \text{span}(\alpha_1^{(i)}, \dots, \alpha_{k_i}^{(i)}), \quad i = 1, 2, \dots, r$$

则

$$\mathcal{A}(\alpha_1^{(i)}, \dots, \alpha_{k_i}^{(i)}) = (\alpha_1^{(i)}, \dots, \alpha_{k_i}^{(i)}) \mathbf{A}_i, \quad i = 1, 2, \dots, r \quad (2.6.3)$$

因此 $\mathcal{A}(\alpha_j^{(i)}) \in W_i$ ($j=1, \dots, k_i$), 从而 W_i 是 \mathcal{A} 的不变子空间, 并且

$$V = W_1 \dot{+} W_2 \dot{+} \dots \dot{+} W_r$$

充分性. 设 V 可表示为 \mathcal{A} 的 r 个非平凡不变子空间 W_1, W_2, \dots, W_r 的直和, 即

$$V = W_1 \dot{+} W_2 \dot{+} \dots \dot{+} W_r$$

在 W_i ($i=1, \dots, r$) 中取一组基 $\alpha_1^{(i)}, \dots, \alpha_{k_i}^{(i)}$, 则 $\alpha_1^{(1)}, \dots, \alpha_{k_1}^{(1)}, \dots, \alpha_1^{(r)}, \dots, \alpha_{k_r}^{(r)}$ 是 V 的一组基. 因为 W_i 是 \mathcal{A} 的不变子空间, 所以

$$\mathcal{A}(\alpha_1^{(i)}, \dots, \alpha_{k_i}^{(i)}) = (\alpha_1^{(i)}, \dots, \alpha_{k_i}^{(i)}) \mathbf{A}_i, \quad i = 1, \dots, r \quad (2.6.4)$$

其中 \mathbf{A}_i 是 k_i 阶矩阵. 于是

$$\begin{aligned} & \mathcal{A}(\alpha_1^{(1)}, \dots, \alpha_{k_1}^{(1)}, \dots, \alpha_1^{(r)}, \dots, \alpha_{k_r}^{(r)}) \\ &= (\alpha_1^{(1)}, \dots, \alpha_{k_1}^{(1)}, \dots, \alpha_1^{(r)}, \dots, \alpha_{k_r}^{(r)}) \begin{pmatrix} \mathbf{A}_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \mathbf{A}_r \end{pmatrix} \quad \square \end{aligned}$$

由此可见, 线性变换在一组基下的矩阵为块对角矩阵与线性空间分解为不变子空间的直和是相当的.

下一章我们将利用矩阵在相似变换下的标准形研究线性变换的结构.

§ 2.7 酉(正交)变换与酉(正交)矩阵

本节首先介绍酉(正交)矩阵的概念及其基本性质,然后讨论内积空间中保持内积的线性变换,称之为酉(正交)变换.

定义 2.7.1 如果 n 阶实矩阵 A 满足

$$A^T A = A A^T = I \quad (2.7.1)$$

则称 A 为正交矩阵. 如果 n 阶复矩阵 A 满足

$$A^H A = A A^H = I \quad (2.7.2)$$

则称 A 为酉矩阵.

根据定义容易验证:如果 A, B 是正交矩阵,则

- (1) $A^{-1} = A^T$, 且 A^T 也是正交矩阵;
- (2) A 非奇异且 $|A| = \pm 1$;
- (3) AB 仍是正交矩阵.

对酉矩阵也有类似的结论.

定理 2.7.1 设 A 是 n 阶矩阵,则 A 是酉(正交)矩阵的充分必要条件是 A 的 n 个列向量是 $C^n(R^n)$ 的标准正交向量组.

证明留给读者完成.

在内积空间中经常需要讨论与内积有关的线性变换.

定义 2.7.2 设 \mathcal{A} 是 n 维酉(欧氏)空间 V 的线性变换,如果对任意 $\alpha, \beta \in V$ 都有

$$(\mathcal{A}(\alpha), \mathcal{A}(\beta)) = (\alpha, \beta) \quad (2.7.3)$$

则称 \mathcal{A} 是 V 的酉(正交)变换.

定理 2.7.2 设 \mathcal{A} 是 n 维酉(欧氏)空间 V 的线性变换,则下列命题等价:

- (1) \mathcal{A} 是酉(正交)变换;
- (2) $\|\mathcal{A}(\alpha)\| = \|\alpha\|, \forall \alpha \in V$;
- (3) 如果 $\epsilon_1, \dots, \epsilon_n$ 是 V 的一组标准正交基,则 $\mathcal{A}(\epsilon_1), \dots, \mathcal{A}(\epsilon_n)$ 也是 V 的一组标准正交基;
- (4) \mathcal{A} 在 V 的任意一组标准正交基下的矩阵是酉(正交)矩阵.

证明 (1) \Rightarrow (2)由定义 2.7.2 即得.

(2) \Rightarrow (1)对任意 $\alpha, \beta \in V$, 由(2)有

$$(\mathcal{A}(\alpha + \beta), \mathcal{A}(\alpha + \beta)) = (\alpha + \beta, \alpha + \beta)$$

$$(\mathcal{A}(\alpha + i\beta), \mathcal{A}(\alpha + i\beta)) = (\alpha + i\beta, \alpha + i\beta), i = \sqrt{-1}$$

因为 \mathcal{A} 是线性变换, 则利用内积性质展开上面两式, 得

$$(\mathcal{A}(\alpha), \mathcal{A}(\beta)) + (\mathcal{A}(\beta), \mathcal{A}(\alpha)) = (\alpha, \beta) + (\beta, \alpha)$$

$$(\mathcal{A}(\alpha), \mathcal{A}(\beta)) - (\mathcal{A}(\beta), \mathcal{A}(\alpha)) = (\alpha, \beta) - (\beta, \alpha)$$

两式相加, 得

$$(\mathcal{A}(\alpha), \mathcal{A}(\beta)) = (\alpha, \beta)$$

即 \mathcal{A} 是酉(正交)变换.

(1) \Rightarrow (3) 如果 $\epsilon_1, \dots, \epsilon_n$ 是 V 的一组标准正交基, 且 \mathcal{A} 是酉(正交)变换, 则

$$(\mathcal{A}(\epsilon_i), \mathcal{A}(\epsilon_j)) = (\epsilon_i, \epsilon_j) = \delta_{ij}$$

故 $\mathcal{A}(\epsilon_1), \dots, \mathcal{A}(\epsilon_n)$ 是 V 的一组标准正交基.

(3) \Rightarrow (1) 如果 $\epsilon_1, \dots, \epsilon_n$ 与 $\mathcal{A}(\epsilon_1), \dots, \mathcal{A}(\epsilon_n)$ 都是 V 的标准正交基, 则对任意 $\alpha, \beta \in V$, 有

$$\alpha = \sum_{i=1}^n x_i \epsilon_i, \quad \beta = \sum_{i=1}^n y_i \epsilon_i$$

则

$$(\mathcal{A}(\alpha), \mathcal{A}(\beta)) = \left(\sum_{i=1}^n x_i \mathcal{A}(\epsilon_i), \sum_{j=1}^n y_j \mathcal{A}(\epsilon_j) \right) = \sum_{i=1}^n x_i \overline{y_i} = (\alpha, \beta)$$

即 \mathcal{A} 是酉(正交)变换.

(3) \Rightarrow (4) 设 $\epsilon_1, \dots, \epsilon_n$ 和 $\mathcal{A}(\epsilon_1), \dots, \mathcal{A}(\epsilon_n)$ 是 V 的两组标准正交基, \mathcal{A} 在基 $\epsilon_1, \dots, \epsilon_n$ 下的矩阵为 $A = (a_{ij})$, 则

$$(\mathcal{A}(\epsilon_1), \dots, \mathcal{A}(\epsilon_n)) = (\epsilon_1, \dots, \epsilon_n)A \quad (2.7.4)$$

于是, 对 $i, j = 1, 2, \dots, n$, 有

$$\begin{aligned} \delta_{ij} &= (\mathcal{A}(\epsilon_i), \mathcal{A}(\epsilon_j)) = \left(\sum_{k=1}^n a_{ki} \epsilon_k, \sum_{l=1}^n a_{lj} \epsilon_l \right) \\ &= \sum_{k=1}^n \sum_{l=1}^n a_{ki} \overline{a_{lj}} (\epsilon_k, \epsilon_l) = \sum_{k=1}^n a_{ki} \overline{a_{kj}} \end{aligned} \quad (2.7.5)$$

即 A 的列向量是标准正交向量组, 由定理 2.7.1 知, A 为酉(正交)矩阵.

(4) \Rightarrow (3) 设 $\epsilon_1, \dots, \epsilon_n$ 是 V 的标准正交基且 A 是酉(正交)矩阵, 则由

(2.7.4)和(2.7.5)即得 $\mathcal{A}(\varepsilon_1), \dots, \mathcal{A}(\varepsilon_n)$ 是 V 的标准正交量. \square

因为酉(正交)变换在标准正交基下的矩阵是酉(正交)矩阵,而酉(正交)矩阵是可逆的,所以酉(正交)变换是可逆的,并且酉(正交)变换的逆变换仍是酉(正交)变换,酉(正交)变换的乘积仍是酉(正交)变换.

例 2.7.1 设 $\theta \in \mathbf{R}$

$$\mathbf{R}(i, j) = \begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & \cos\theta & \cdots & \cdots & \cdots & \sin\theta & \cdots & \cdots & \cdots \\ & & & \vdots & 1 & & & \vdots & & & \\ & & & \vdots & & \ddots & & \vdots & & & \\ & & & \vdots & & & 1 & & & & \\ & & & -\sin\theta & \cdots & \cdots & \cdots & \cos\theta & \cdots & \cdots & \cdots \\ & & & & & & & & 1 & & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & 1 \end{pmatrix} \begin{array}{l} \\ \\ \\ \text{第 } i \text{ 行} \\ \\ \\ \\ \text{第 } j \text{ 行} \\ \\ \end{array}$$

则 $\mathbf{R}(i, j)$ 是正交矩阵,并称为 Givens 矩阵.

对 $x = (x_1, \dots, x_n)^T \in \mathbf{R}^n$, 若 $y = \mathbf{R}(i, j)x = (y_1, \dots, y_n)^T$, 则

$$\begin{cases} y_k = x_k, & k \neq i, j \\ y_i = \cos\theta x_i + \sin\theta x_j \\ y_j = -\sin\theta x_i + \cos\theta x_j \end{cases}$$

如果 $x_i^2 + x_j^2 \neq 0$, 选取

$$\cos\theta = \frac{x_i}{\sqrt{x_i^2 + x_j^2}}, \quad \sin\theta = \frac{x_j}{\sqrt{x_i^2 + x_j^2}}$$

则 $y_i = (x_i^2 + x_j^2)^{\frac{1}{2}}, y_j = 0$.

Givens 矩阵是矩阵计算的重要工具之一.

习 题

1. 设 \mathcal{L} 是 \mathbf{R}^4 到 \mathbf{R}^3 的线性映射, \mathbf{R}^4 中基 e_1, e_2, e_3, e_4 的像分别为

$$\begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 5 \end{pmatrix}, \begin{pmatrix} 5 \\ 3 \\ -13 \end{pmatrix}, \begin{pmatrix} 4 \\ -3 \\ 11 \end{pmatrix}$$

(1) 在 \mathbf{R}^3 中取一组基 $\alpha_1, \alpha_2, \alpha_3$, 写出 \mathcal{A} 在基 e_1, e_2, e_3, e_4 和 $\alpha_1, \alpha_2, \alpha_3$ 下的矩阵;

(2) 求 \mathcal{A} 的值域和核, 并确定其维数.

2. 设 \mathcal{A} 是 n 维线性空间 V_1 到 m 维线性空间 V_2 的线性映射. 证明: 可以适当选取 V_1 和 V_2 的基使 \mathcal{A} 在这对基下的矩阵为

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

3. 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, $\mathcal{T}(\xi) = \xi + \alpha$, 其中 $\alpha \in V$ 是一固定向量;

(2) 在 \mathcal{R}^3 , $\mathcal{T}(x_1, x_2, x_3) = (x_1^2, x_2 + x_3, x_3^2)$;

(3) 在 \mathbf{R}^3 中, $\mathcal{T}(x_1, x_2, x_3) = (2x_1 - x_2, x_2 + x_3, x_1)$;

(4) 在 $\mathbf{R}^{n \times n}$ 中, $\mathcal{T}(X) = BXC$, 其中 $B, C \in \mathbf{R}^{n \times n}$ 是两个固定的矩阵;

(5) 在 $\mathbf{R}[x]$ 中, $\mathcal{T}[f(x)] = f(x+a)$, 其中 a 是 \mathbf{R} 中一个固定的数;

(6) 把复数域 \mathbf{C} 分别看作实数域 \mathbf{R} 和复数域 \mathbf{C} 上的线性空间, $\mathcal{T}(z) = \bar{z}$.

4. 在 \mathbf{R}^2 中, $\mathcal{T}_1(x_1, x_2) = (x_2, -x_1)$, $\mathcal{T}_2(x_1, x_2) = (x_1, -x_2)$, 证明 \mathcal{T}_1 和 \mathcal{T}_2 是 \mathbf{R}^2 的两个线性变换, 并求 $\mathcal{T}_1 + \mathcal{T}_2$, $\mathcal{T}_1\mathcal{T}_2$ 和 $\mathcal{T}_2\mathcal{T}_1$.

5. 设 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 是 n 维线性空间 V 的一组基, \mathcal{A} 是 V 的线性变换. 证明: \mathcal{A} 可逆的充分必要条件是 $\mathcal{A}(\varepsilon_1), \mathcal{A}(\varepsilon_2), \dots, \mathcal{A}(\varepsilon_n)$ 线性无关.

6. 求下列线性变换在指定基下的矩阵:

(1) 第3题(3)中线性变换 \mathcal{T} 在基 $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$ 下的矩阵;

(2) 已知 \mathbf{R}^3 中线性变换 \mathcal{T} 在基 $\eta_1 = (-1, 1, 1)^T$, $\eta_2 = (1, 0, -1)^T$, $\eta_3 = (0, 1, 1)^T$ 下的矩阵为

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

求 \mathcal{T} 在基 $e_1 = (1, 0, 0)^T$, $e_2 = (0, 1, 0)^T$, $e_3 = (0, 0, 1)^T$ 下的矩阵;

(3) 在 \mathbf{R}^3 中, $\eta_1 = (-1, 0, 2)^T$, $\eta_2 = (0, 1, 1)^T$, $\eta_3 = (3, -1, 0)^T$, 定义线性变换 \mathcal{T} 为

$$\mathcal{T}(\eta_1) = (-5, 0, 3)^T$$

$$\mathcal{T}(\eta_2) = (0, -1, 6)^T$$

$$\mathcal{T}(\eta_3) = (-5, -1, 9)^T$$

求 \mathcal{T} 在基 $e_1 = (1, 0, 0)^T$, $e_2 = (0, 1, 0)^T$, $e_3 = (0, 0, 1)^T$ 下的矩阵;

(4) 在 $\mathbf{R}^{2 \times 2}$ 中定义线性变换 \mathcal{T} 为

$$\mathcal{A}(\mathbf{X}) = \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix} \mathbf{X} \begin{pmatrix} 3 & 0 \\ 1 & 1 \end{pmatrix}$$

求 \mathcal{A} 在基

$$\mathbf{E}_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \mathbf{E}_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \mathbf{E}_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \mathbf{E}_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

下的矩阵.

7. 设 \mathcal{A} 是 n 维线性空间 V 的线性变换, 如果有向量 ξ 使得 $\mathcal{A}^{n-1}(\xi) \neq 0$, 但 $\mathcal{A}^n(\xi) = 0$, 证明:

- (1) $\xi, \mathcal{A}(\xi), \dots, \mathcal{A}^{n-1}(\xi)$ 线性无关;
- (2) \mathcal{A} 在某组基下的矩阵为

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

8. 给定线性空间 \mathbf{R}^3 的两组基

$$\epsilon_1 = (1, 0, 1)^T, \epsilon_2 = (2, 1, 0)^T, \epsilon_3 = (1, 1, 1)^T$$

$$\eta_1 = (1, 2, -1)^T, \eta_2 = (2, 2, -1)^T, \eta_3 = (2, -1, -1)^T$$

定义 \mathbf{R}^3 的线性变换 \mathcal{A} 为 $\mathcal{A}(\epsilon_i) = \eta_i (i=1, 2, 3)$. 试求:

- (1) 从基 $\epsilon_1, \epsilon_2, \epsilon_3$ 到基 η_1, η_2, η_3 的过渡矩阵;
- (2) \mathcal{A} 在基 $\epsilon_1, \epsilon_2, \epsilon_3$ 下的矩阵;
- (3) \mathcal{A} 在基 η_1, η_2, η_3 下的矩阵.

9. 设 $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$ 是 4 维线性空间 V 的一组基, 线性变换 \mathcal{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 \mathcal{A} 在基 $\eta_1 = \epsilon_1 - 2\epsilon_2 + \epsilon_4, \eta_2 = 3\epsilon_2 - \epsilon_3 - \epsilon_4, \eta_3 = \epsilon_3 + \epsilon_4, \eta_4 = 2\epsilon_4$ 下的矩阵;
- (2) 求 \mathcal{A} 的值域与核.

10. 在 n 维线性空间 $\mathbf{R}[x]_n$ 中, 定义线性变换 $\mathcal{Q}(f(x)) = f'(x)$, 其中 $f(x) \in \mathbf{R}[x]_n$. 求 \mathcal{Q} 的值域与核.

11. 设 \mathcal{A}, \mathcal{B} 是 n 维线性空间 V 的线性变换, 证明:

$$\text{rank}(\mathcal{A}\mathcal{B}) \geq \text{rank}(\mathcal{A}) + \text{rank}(\mathcal{B}) - n$$

12. 设 \mathcal{A}, \mathcal{B} 是 n 维线性空间 V 的线性变换, 并且 $\mathcal{A}^2 = \mathcal{A}, \mathcal{B}^2 = \mathcal{B}$. 证明:

- (1) \mathcal{A} 和 \mathcal{B} 有相同值域的充分必要条件是 $\mathcal{A}\mathcal{B} = \mathcal{B}, \mathcal{B}\mathcal{A} = \mathcal{A}$;

(2) \mathcal{A} 和 \mathcal{B} 有相同核的充分必要条件是 $\mathcal{A}\mathcal{B} = \mathcal{A}$, $\mathcal{B}\mathcal{A} = \mathcal{B}$.

13. 设 \mathcal{A} 是线性空间 \mathbf{R}^3 的线性变换, 它在 \mathbf{R}^3 中基 $\alpha_1, \alpha_2, \alpha_3$ 下的矩阵表示是

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

(1) 求 \mathcal{A} 在基 $\beta_1 = \alpha_1, \beta_2 = \alpha_1 + \alpha_2, \beta_3 = \alpha_1 + \alpha_2 + \alpha_3$ 下的矩阵表示;

(2) 求 \mathcal{A} 的特征值与特征向量.

14. 设线性空间 \mathbf{R}^3 的线性变换 \mathcal{A} 定义为

$$\mathcal{A} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2x_1 + x_3 \\ 2x_2 \\ 3x_3 \end{pmatrix}$$

求线性变换 \mathcal{A} 的特征值与特征向量.

15. 求如下矩阵的特征值和特征向量:

$$(1) \begin{pmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{pmatrix}; \quad (2) \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}; \quad (3) \begin{pmatrix} 0 & 2 & 1 \\ -2 & 0 & 3 \\ -1 & -3 & 0 \end{pmatrix}$$

16. 设 A 是 n 阶矩阵, 证明:

(1) 如果 A 是 Hermite 矩阵, 则 A 的特征值均为实数;

(2) 如果 A 是反 Hermite 矩阵, 则 A 的特征值为零或纯虚数.

17. 设 A, B 是 n 阶矩阵, 如果 A 与 B 相似, 证明:

(1) 对任意正整数 k , A^k 与 B^k 相似;

(2) 对任意多项式 $p(\lambda)$, $p(A)$ 与 $p(B)$ 相似;

(3) A^T 与 B^T 相似.

18. 设 A, B 是 n 阶矩阵, 如果 A, B 中有一个可逆, 证明 AB 与 BA 相似.

19. 设 A, B 是 n 阶矩阵, C, D 是 m 阶矩阵. 如果 A 与 B 相似, C 与 D 相似, 证明

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

20. 在第 15 题中, 哪些矩阵是可对角化的? 对可对角化的矩阵, 求可逆矩阵 P 使得 $P^{-1}AP$ 是对角矩阵.

21. 设 A 是 n 阶非零矩阵, 如果存在某一正整数 k 使 $A^k = 0$, 证明 A 不可能相似于对角矩阵.

22. 设 A, B 是 n 阶矩阵, 证明:

(1) $\text{tr}(A^k) = \sum_{i=1}^n \lambda_i^k$, λ_i 是 A 的特征值;

(2) $\text{tr}[(AB)^k] = \text{tr}[(BA)^k]$, $k = 1, 2, \dots$.

23. 设 \mathcal{A} 是 n 维线性空间 V 的线性变换, 证明:

- (1) \mathcal{A} 可逆的充分必要条件是 \mathcal{A} 的特征值均非零;
 (2) 如果 λ 是可逆线性变换 \mathcal{A} 的特征值, 则 λ^{-1} 是 \mathcal{A}^{-1} 的特征值.

24. 设 \mathcal{A}, \mathcal{B} 是复数域上 n 维线性空间 V 的线性变换, 并且 $\mathcal{A}\mathcal{B} = \mathcal{B}\mathcal{A}$. 证明:

- (1) 如果 λ 是 \mathcal{A} 的一个特征值, 则 V_λ 是 \mathcal{B} 的不变子空间;
 (2) \mathcal{A} 与 \mathcal{B} 至少有一个公共的特征向量.

25. 设 A, B 是 n 阶矩阵, 并且 $AB = BA$. 证明

- (1) 如果 A 有 n 个互异的特征值, 则 B 相似于对角矩阵;
 (2) 如果 A 与 B 均为可对角化矩阵, 则存在可逆矩阵 P 使得 $P^{-1}AP$ 与 $P^{-1}BP$ 同时为对角矩阵.

26. 设 $\alpha_1, \alpha_2, \alpha_3$ 是 3 维欧氏空间 V 的一组标准正交基, 求 V 的一个正交变换 \mathcal{A} 使得

$$\begin{cases} \mathcal{A}(\alpha_1) = \frac{2}{3}\alpha_1 + \frac{2}{3}\alpha_2 - \frac{1}{3}\alpha_3 \\ \mathcal{A}(\alpha_2) = \frac{2}{3}\alpha_1 - \frac{1}{3}\alpha_2 + \frac{2}{3}\alpha_3 \end{cases}$$

27. 设 \mathcal{A} 是欧氏空间 V 的一个正交变换, V_1 是 \mathcal{A} 的不变子空间. 证明: V_1 的正交补 V_1^\perp 也是 \mathcal{A} 的不变子空间.

28. 证明: 正交(酉)矩阵的特征值的模为 1.

第三章 λ 矩阵与矩阵的 Jordan 标准形

本章讨论矩阵在相似变换下的标准形问题. 首先介绍一元多项式与 λ 矩阵的基本理论, 然后利用 λ 矩阵的理论导出矩阵的 Jordan 标准形, 并利用矩阵的 Jordan 标准形研究线性变换的结构.

λ 矩阵的理论和矩阵的 Jordan 标准形不仅在矩阵理论和矩阵计算中起着十分重要的作用, 而且在力学、控制理论等学科具有广泛的应用.

§ 3.1 一元多项式

在初等代数中已学过一元多项式, 这里简要介绍一元多项式的基本理论.

定义 3.1.1 设 P 是数域, n 是一非负整数, λ 是一个文字, 形式表达式

$$f(\lambda) = a_n \lambda^n + a_{n-1} \lambda^{n-1} + \cdots + a_1 \lambda + a_0 \quad (3.1.1)$$

其中 $a_i \in P (i=0, 1, \cdots, n)$, 称为数域 P 上的一元多项式, 数域 P 上一元多项式的全体记为 $P[\lambda]$.

如果 $a_n \neq 0$, 则称 $a_n \lambda^n$ 为多项式 $f(\lambda)$ 的首项, n 称为 $f(\lambda)$ 的次数, 记为 $\partial(f(\lambda)) = n$, a_n 称为 $f(\lambda)$ 的首项系数. 若 $a_n = 1$, 则称 $f(\lambda)$ 为首项系数为 1 的多项式. 系数全为零的多项式称为零多项式, 记为 0. 零多项式是惟一不定义次数的多项式.

对两个多项式

$$f(\lambda) = \sum_{i=0}^n a_n \lambda^{n-i}, \quad g(\lambda) = \sum_{i=0}^m b_{m-i} \lambda^{m-i} \quad (3.1.2)$$

如果 $m = n$, 且 $b_j = a_j (j=0, 1, \cdots, n)$, 则称 $f(\lambda)$ 与 $g(\lambda)$ 相等, 记为 $f(\lambda) = g(\lambda)$.

在初等代数中, 两个多项式可以相加、相减和相乘. 对形式表达式 (3.1.1), 可类似地引入这些运算. 对 (3.1.2) 中两个多项式 $f(\lambda)$ 与 $g(\lambda)$, 在表示它们的和时, 若 $n \geq m$, 为了方便起见, 在 $g(\lambda)$ 中令 $b_n = b_{n-1} = \cdots = b_{m+1} = 0$, 则 $f(\lambda)$ 与 $g(\lambda)$ 的加减为

$$f(\lambda) \pm g(\lambda) = \sum_{i=0}^n (a_{n-i} \pm b_{n-i}) \lambda^{n-i}$$

而 $f(\lambda)$ 与 $g(\lambda)$ 的乘积为

$$f(\lambda)g(\lambda) = \sum_{i=0}^{m+n} c_{m+n-i} \lambda^{m+n-i}$$

其中 $c_k = \sum_{i+j=k} a_i b_j$.

数域 \mathbf{P} 上的两个多项式经过加、减、乘运算后, 所得结果仍然是数域 \mathbf{P} 上的多项式, 并且这些运算满足如下规律.

$$(1) \partial(f(\lambda) \pm g(\lambda)) \leq \max\{\partial(f(\lambda)), \partial(g(\lambda))\},$$

$$\partial(f(\lambda)g(\lambda)) = \partial(f(\lambda)) + \partial(g(\lambda)).$$

(2) 加法和乘法适合交换律

$$f(\lambda) + g(\lambda) = g(\lambda) + f(\lambda),$$

$$f(\lambda)g(\lambda) = g(\lambda)f(\lambda).$$

(3) 加法和乘法适合结合律

$$(f(\lambda) + g(\lambda)) + h(\lambda) = f(\lambda) + (g(\lambda) + h(\lambda)),$$

$$(f(\lambda)g(\lambda))h(\lambda) = f(\lambda)(g(\lambda)h(\lambda)).$$

(4) 乘法对加法适合分配律

$$f(\lambda)(g(\lambda) + h(\lambda)) = f(\lambda)g(\lambda) + f(\lambda)h(\lambda).$$

(5) 乘法消去律

如果 $f(\lambda)g(\lambda) = f(\lambda)h(\lambda)$ 且 $f(\lambda) \neq 0$, 则有 $g(\lambda) = h(\lambda)$.

一般说来, 两个多项式不能随便作除法, 但有如下定理.

定理 3.1.1 设 $f(\lambda), g(\lambda) \in \mathbf{P}[\lambda]$ 且 $g(\lambda) \neq 0$, 则存在惟一的多项式 $q(\lambda), r(\lambda) \in \mathbf{P}[\lambda]$ 使得

$$f(\lambda) = q(\lambda)g(\lambda) + r(\lambda) \quad (3.1.3)$$

其中 $r(\lambda) = 0$ 或 $r(\lambda) \neq 0$ 但 $\partial(r(\lambda)) < \partial(g(\lambda))$.

证明 先证存在性. 如果 $f(\lambda) = 0$, 则取 $q(\lambda) = r(\lambda) = 0$ 即可. 现设 $f(\lambda) \neq 0$, 记

$$f(\lambda) = a_n \lambda^n + a_{n-1} \lambda^{n-1} + \cdots + a_1 \lambda + a_0,$$

$$g(\lambda) = b_m \lambda^m + b_{m-1} \lambda^{m-1} + \cdots + b_1 \lambda + b_0$$

其中 $a_n b_m \neq 0$. 对 $f(\lambda)$ 的次数 n 作数学归纳法.

当 $n < m$ 时, 则取 $q(\lambda) = 0, r(\lambda) = f(\lambda)$ 即可.

当 $n \geq m$ 时, 令

$$f_1(\lambda) = f(\lambda) - b_m^{-1} a_n \lambda^{n-m} g(\lambda)$$

则 $\partial(f_1(\lambda)) \leq n-1$. 由归纳假设知, 存在多项式 $q_1(\lambda), r(\lambda) \in \mathbf{P}[\lambda]$ 使得

$$f_1(\lambda) = q_1(\lambda)g(\lambda) + r(\lambda)$$

其中 $r(\lambda) = 0$ 或 $r(\lambda) \neq 0$ 但 $\partial(r(\lambda)) < \partial(g(\lambda))$. 取 $q(\lambda) = b_m^{-1}a_n\lambda^{n-m} + q_1(\lambda)$ 即证明了存在性.

下面证惟一性. 设另有多项式 $q_0(\lambda), r_0(\lambda) \in \mathbf{P}[\lambda]$ 使得

$$f(\lambda) = q_0(\lambda)g(\lambda) + r_0(\lambda)$$

其中 $r_0(\lambda) = 0$ 或 $r_0(\lambda) \neq 0$ 但 $\partial(r_0(\lambda)) < \partial(g(\lambda))$. 于是

$$(q(\lambda) - q_0(\lambda))g(\lambda) = r_0(\lambda) - r(\lambda)$$

如果 $q(\lambda) \neq q_0(\lambda)$, 则由 $g(\lambda) \neq 0$, 有 $r_0(\lambda) - r(\lambda) \neq 0$, 从而

$$\partial(q(\lambda) - q_0(\lambda)) + \partial(g(\lambda)) = \partial(r_0(\lambda) - r(\lambda))$$

这与 $\partial(g(\lambda)) > \partial(r_0(\lambda) - r(\lambda))$ 矛盾. 因此

$$q(\lambda) = q_0(\lambda), r_0(\lambda) = r(\lambda). \quad \square$$

(3.1.3) 中的 $q(\lambda)$ 称为 $g(\lambda)$ 除 $f(\lambda)$ 的商, 而 $r(\lambda)$ 称为 $g(\lambda)$ 除 $f(\lambda)$ 的余式.

定义 3.1.2 设 $f(\lambda), g(\lambda) \in \mathbf{P}[\lambda]$, 如果存在 $h(\lambda) \in \mathbf{P}[\lambda]$ 使得

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

则称多项式 $g(\lambda)$ 整除 $f(\lambda)$, 记为 $g(\lambda) | f(\lambda)$.

如果 $g(\lambda) | f(\lambda)$, 则称 $g(\lambda)$ 是 $f(\lambda)$ 的因式, $f(\lambda)$ 是 $g(\lambda)$ 的倍式.

由定理 3.1.1 直接得如下结论.

定理 3.1.2 设 $f(\lambda), g(\lambda) \in \mathbf{P}[\lambda]$ 且 $g(\lambda) \neq 0$, 则 $g(\lambda) | f(\lambda)$ 的充分必要条件是 $g(\lambda)$ 除 $f(\lambda)$ 的余式为零.

多项式的整除具有以下一些常用的性质.

- (1) 如果 $g(\lambda) | f(\lambda)$ 且 $f(\lambda) | g(\lambda)$, 则 $f(\lambda) = cg(\lambda)$, 其中 c 是常数.
- (2) 如果 $g(\lambda) | h(\lambda)$, $h(\lambda) | f(\lambda)$, 则 $g(\lambda) | f(\lambda)$.
- (3) 如果 $g(\lambda) | f_i(\lambda) (i = 1, \dots, m)$, 则 $g(\lambda) | (u_1(\lambda)f_1(\lambda) + \dots + u_m(\lambda)f_m(\lambda))$, 其中 $u_i(\lambda) \in \mathbf{P}[\lambda] (i = 1, \dots, m)$.

这些性质的证明留给读者.

设 $f(\lambda), g(\lambda) \in \mathbf{P}[\lambda]$, 如果多项式 $h(\lambda)$ 既是 $f(\lambda)$ 的因式, 又是 $g(\lambda)$ 的因式, 则称 $h(\lambda)$ 是 $f(\lambda)$ 与 $g(\lambda)$ 的公因式.

定义 3.1.3 设 $f(\lambda), g(\lambda) \in \mathbf{P}[\lambda]$, 如果存在 $d(\lambda) \in \mathbf{P}[\lambda]$ 满足

- (1) $d(\lambda)$ 是 $f(\lambda)$ 与 $g(\lambda)$ 的公因式;

(2) $f(\lambda)$ 与 $g(\lambda)$ 的任一公因式都是 $d(\lambda)$ 的因式,
则称 $d(\lambda)$ 是 $f(\lambda)$ 与 $g(\lambda)$ 的一个**最大公因式**,并用 $(f(\lambda), g(\lambda))$ 表示 $f(\lambda)$ 与 $g(\lambda)$ 的首项系数为 1 的最大公因式.

可以证明: $f(\lambda)$ 与 $g(\lambda)$ 的首项系数为 1 的最大公因式存在并且惟一.

以后,多项式的最大公因式均指首项系数为 1 的最大公因式. 如果 $(f(\lambda), g(\lambda)) = 1$, 则称多项式 $f(\lambda)$ 与 $g(\lambda)$ **互素**.

如果多项式 $\varphi(\lambda)$ 既是 $f(\lambda)$ 的倍式, 又是 $g(\lambda)$ 的倍式, 则称 $\varphi(\lambda)$ 是 $f(\lambda)$ 与 $g(\lambda)$ 的**公倍式**.

定义 3.1.4 设 $f(\lambda), g(\lambda) \in \mathbf{P}[\lambda]$, 如果存在 $\psi(\lambda) \in \mathbf{P}[\lambda]$ 满足

(1) $\psi(\lambda)$ 是 $f(\lambda)$ 与 $g(\lambda)$ 的公倍式;

(2) $f(\lambda)$ 与 $g(\lambda)$ 的任一公倍式都是 $\psi(\lambda)$ 的倍式,

则称 $\psi(\lambda)$ 是 $f(\lambda)$ 与 $g(\lambda)$ 的一个**最小公倍式**, 用 $[f(\lambda), g(\lambda)]$ 表示 $f(\lambda)$ 与 $g(\lambda)$ 的首项系数为 1 的最小公倍式.

上述两个多项式的最大公因式和最小公倍式的概念可以推广到多个多项式的情形.

§ 3.2 λ 矩阵及其在相抵下的标准形

3.2.1 λ 矩阵的基本概念

定义 3.2.1 设 $a_{ij}(\lambda) (i=1, 2, \dots, m, j=1, 2, \dots, n)$ 是数域 \mathbf{P} 上的多项式, 以 $a_{ij}(\lambda)$ 为元素的 $m \times n$ 矩阵

$$\mathbf{A}(\lambda) = \begin{pmatrix} a_{11}(\lambda) & a_{12}(\lambda) & \cdots & a_{1n}(\lambda) \\ a_{21}(\lambda) & a_{22}(\lambda) & \cdots & a_{2n}(\lambda) \\ \vdots & \vdots & & \vdots \\ a_{m1}(\lambda) & a_{m2}(\lambda) & \cdots & a_{mn}(\lambda) \end{pmatrix}$$

称为**多项式矩阵**或 **λ 矩阵**, 多项式 $a_{ij}(\lambda) (i=1, 2, \dots, m, j=1, 2, \dots, n)$ 中的最高次数称为 $\mathbf{A}(\lambda)$ 的**次数**, 数域 \mathbf{P} 上 $m \times n$ λ 矩阵的全体记为 $\mathbf{P}[\lambda]^{m \times n}$.

显然, 数字矩阵是 λ 矩阵的特例. 数字矩阵 \mathbf{A} 的特征矩阵 $\lambda \mathbf{I} - \mathbf{A}$ 就是 1 次 λ 矩阵.

如果 $m \times n$ λ 矩阵 $\mathbf{A}(\lambda)$ 的次数为 k , 则 $\mathbf{A}(\lambda)$ 可表示为

$$\mathbf{A}(\lambda) = \mathbf{A}_k \lambda^k + \mathbf{A}_{k-1} \lambda^{k-1} + \cdots + \mathbf{A}_1 \lambda + \mathbf{A}_0$$

其中 $\mathbf{A}_i (i=0, 1, \dots, k)$ 是 $m \times n$ 数字矩阵, 并且 $\mathbf{A}_k \neq 0$. 例如

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

如果另一个 $m \times n$ λ 矩阵 $B(\lambda)$ 表示为

$$B(\lambda) = B_l \lambda^l + B_{l-1} \lambda^{l-1} + \cdots + B_1 \lambda + B_0$$

则当且仅当 $k=l$, $A_j = B_j$ ($j=0, 1, \cdots, k$), $A(\lambda)$ 与 $B(\lambda)$ 相等, 记为 $A(\lambda) = B(\lambda)$.

由于 λ 的多项式可作加法、减法、乘法三种运算, 并且它们与数的运算有相同的运算规律; 而矩阵的加法、减法、乘法和数量乘法的定义仅用到其元素的加法、减法、乘法. 因此, 我们可以同样定义 λ 矩阵的加法、减法、乘法和数量乘法, 并且 λ 矩阵的这些运算同数字矩阵的加法、减法、乘法和数量乘法具有相同的运算规律.

矩阵行列式的定义也仅用到其元素的加法与乘法, 因此, 同样可以定义一个 n 阶 λ 矩阵的行列式. 一般说来, λ 矩阵的行列式是 λ 的多项式. λ 矩阵的行列式与数字矩阵的行列式有相同的性质. 例如, 对两个 n 阶 λ 矩阵 $A(\lambda)$ 与 $B(\lambda)$, 有

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

有了 λ 矩阵行列式的概念, 可以同样定义 λ 矩阵的子式、代数余子式.

定义 3.2.2 设 $A(\lambda) \in P[\lambda]^{m \times n}$, 如果 $A(\lambda)$ 中有一个 r ($1 \leq r \leq \min\{m, n\}$) 阶子式不为零, 而所有 $r+1$ 阶子式 (如果有的话) 全为零, 则称 $A(\lambda)$ 的秩为 r , 记为 $\text{rank}(A(\lambda)) = r$.

例 3.2.1 设 A 是 n 阶数字矩阵, 则 $|\lambda I - A|$ 是 λ 的 n 次多项式. 因此 A 的特征矩阵 $\lambda I - A$ 的秩为 n , 即 $\lambda I - A$ 总是满秩的.

定义 3.2.3 设 $A(\lambda) \in P[\lambda]^{n \times n}$, 如果存在一个 n 阶 λ 矩阵 $B(\lambda)$ 使得

$$A(\lambda)B(\lambda) = B(\lambda)A(\lambda) = I \quad (3.2.1)$$

则称 λ 矩阵 $A(\lambda)$ 是可逆的, 并称 $B(\lambda)$ 为 $A(\lambda)$ 的逆矩阵, 记作 $A(\lambda)^{-1}$.

容易证明: 如果 n 阶 λ 矩阵 $A(\lambda)$ 可逆, 则它的逆矩阵是惟一的.

定理 3.2.1 设 $A(\lambda) \in P[\lambda]^{n \times n}$, 则 $A(\lambda)$ 可逆的充分必要条件是 $|A(\lambda)|$ 是非零常数.

证明 必要性. 设 $A(\lambda)$ 可逆, 则存在 n 阶 λ 矩阵 $B(\lambda)$ 满足 (3.2.1), 从而

$$|A(\lambda)| |B(\lambda)| = 1$$

因为 $|A(\lambda)|$ 与 $|B(\lambda)|$ 都是 λ 的多项式, 则由上式可知 $|A(\lambda)|$ 与 $|B(\lambda)|$ 都是零次多项式, 故 $|A(\lambda)|$ 是非零常数.

充分性. 设 $|A(\lambda)| = d$ 是非零常数, $A(\lambda)^*$ 是 $A(\lambda)$ 的伴随矩阵, 则 $\frac{1}{d}A(\lambda)^*$ 是一个 n 阶 λ 矩阵, 并且

$$A(\lambda) \frac{1}{d}A(\lambda)^* = \frac{1}{d}A(\lambda)^* A(\lambda) = I$$

因此 $A(\lambda)$ 可逆, 并且 $A(\lambda)^{-1} = \frac{1}{d}A(\lambda)^*$. \square

3.2.2 λ 矩阵的初等变换与相抵

与数字矩阵类似, 对于 λ 矩阵, 也可进行初等变换.

定义 3.2.4 下列三种变换称为 λ 矩阵的初等变换:

- (1) λ 矩阵的两行(列)互换位置;
- (2) λ 矩阵的某一行(列)乘以非零常数 k ;
- (3) λ 矩阵的某一行(列)的 $\varphi(\lambda)$ 倍加到另一行(列), 其中 $\varphi(\lambda)$ 是 λ 的多项式.

对单位矩阵施行上述三种初等变换便得相应的一种 λ 矩阵的初等矩阵 $P(i, j), P(i(k)), P(i, j(\varphi))$, 即

$$P(i, j) = \begin{pmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & 0 & \cdots & 1 & \cdots & \cdots \\ & & & & 1 & & & \\ & & \vdots & & & \ddots & & \\ & & & & & & 1 & \\ & & 1 & \cdots & 0 & \cdots & \cdots & \cdots \\ & & & & & 1 & & \\ & & & & & & \ddots & \\ & & & & & & & 1 \end{pmatrix} \begin{array}{l} \\ \\ \\ \text{第 } i \text{ 行} \\ \\ \\ \text{第 } j \text{ 行} \\ \\ \end{array}$$

$$P(i(k)) = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & k & \cdots & \cdots & \cdots \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \quad \text{第 } i \text{ 行}$$

$$P(i, j(\varphi)) = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & \cdots & \varphi(\lambda) & \cdots & \cdots \\ & & & \ddots & \vdots & & \\ & & & & 1 & \cdots & \cdots \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \quad \begin{array}{l} \text{第 } i \text{ 行} \\ \text{第 } j \text{ 行} \end{array}$$

与数字矩阵的情形完全一样, 对一个 $m \times n$ λ 矩阵 $A(\lambda)$ 作一次初等行变换相当于在 $A(\lambda)$ 左边乘上相应的 m 阶初等矩阵; 对 $A(\lambda)$ 作一次初等列变换相当于在 $A(\lambda)$ 的右边乘上相应的 n 阶初等矩阵, 并且初等变换都是可逆的.

容易证明: 初等矩阵都是可逆的, 并且

$$P(i, j)^{-1} = P(i, j), P(i(k))^{-1} = P(i(k^{-1})), P(i, j(\varphi))^{-1} = P(i, j(-\varphi)).$$

为了方便起见, 我们用下列记号表示初等变换:

$[i, j]$ 表示第 i, j 行(列)互换位置;

$[i(k)]$ 表示用非零常数 k 乘第 i 行(列);

$[i + j(\varphi)]$ 表示将第 j 行(列)的 $\varphi(\lambda)$ 倍加到第 i 行(列).

定义 3.2.5 设 $A(\lambda), B(\lambda) \in P[\lambda]^{m \times n}$, 如果 $A(\lambda)$ 经过有限次的初等变换化为 $B(\lambda)$, 则称 λ 矩阵 $A(\lambda)$ 与 $B(\lambda)$ 相抵, 记为 $A(\lambda) \cong B(\lambda)$.

由初等变换的可逆性可知, 相抵是 λ 矩阵之间的一种等价关系.

利用初等变换与初等矩阵的对应关系可得.

定理 3.2.2 设 $A(\lambda), B(\lambda) \in P[\lambda]^{m \times n}$, 则 $A(\lambda)$ 与 $B(\lambda)$ 相抵的充分必要条件为存在 m 阶初等矩阵 $P_1(\lambda), \dots, P_l(\lambda)$ 与 n 阶初等矩阵 $Q_1(\lambda), \dots, Q_r(\lambda)$ 使得

$$A(\lambda) = P_l(\lambda) \cdots P_1(\lambda) B(\lambda) Q_1(\lambda) \cdots Q_r(\lambda) \quad (3.2.2)$$

与数字矩阵不同, 具有相同秩的两个 λ 矩阵未必相抵. 例如

$$A(\lambda) = \begin{pmatrix} \lambda & 2 \\ 0 & \lambda \end{pmatrix}, B(\lambda) = \begin{pmatrix} \lambda & -2 \\ \lambda & 2 \end{pmatrix}$$

因为 $|A(\lambda)| = \lambda^2$, $|B(\lambda)| = 4\lambda$, 所以 $A(\lambda)$ 与 $B(\lambda)$ 的秩均为 2. 因为初等变换是可逆的, 则由定理 3.2.2 知, 两个相抵的 λ -方阵的行列式只能相差一个非零常数, 故 $A(\lambda)$ 与 $B(\lambda)$ 不相抵. 因此, 秩相等不是 λ 矩阵相抵的充分条件.

3.2.3 λ 矩阵在相抵下的标准形

现在我们讨论 λ 矩阵在初等变换下的标准形. 为此, 先证明一个引理.

引理 3.2.1 设 λ 矩阵 $A(\lambda) = (a_{ij}(\lambda))$ 的左上角元素 $a_{11}(\lambda) \neq 0$, 并且 $A(\lambda)$ 中至少有一个元素不能被 $a_{11}(\lambda)$ 整除, 则存在一个与 $A(\lambda)$ 相抵的 λ 矩阵 $B(\lambda) = (b_{ij}(\lambda))$ 使得 $b_{11}(\lambda) \neq 0$ 且 $\partial(b_{11}(\lambda)) < \partial(a_{11}(\lambda))$.

证明 根据 $A(\lambda)$ 中不能被 $a_{11}(\lambda)$ 整除的元素所在的位置, 分三种情形来讨论.

(1) 若在 $A(\lambda)$ 的第一列中有一个元素 $a_{i1}(\lambda)$ 不能被 $a_{11}(\lambda)$ 整除, 则由定理 3.1.1 知, 存在多项式 $q(\lambda)$ 和 $r(\lambda)$ 使得

$$a_{i1}(\lambda) = q(\lambda)a_{11}(\lambda) + r(\lambda)$$

其中 $r(\lambda) \neq 0$ 且 $\partial(r(\lambda)) < \partial(a_{11}(\lambda))$. 对 $A(\lambda)$ 作两次初等行变换, 首先将 $A(\lambda)$ 第 1 行的 $-q(\lambda)$ 倍加到第 i 行, 这时第 i 行第 1 列位置的元素是 $r(\lambda)$; 然后将第 1 行与第 i 行互换即得所要求的 λ 矩阵 $B(\lambda)$.

(2) 在 $A(\lambda)$ 的第一行中有一个元素 $a_{1j}(\lambda)$ 不能被 $a_{11}(\lambda)$ 整除, 这种情形的证明与 (1) 类似.

(3) $A(\lambda)$ 的第一行与第一列中的元素都能被 $a_{11}(\lambda)$ 整除, 但 $A(\lambda)$ 中有一个元素 $a_{ij}(\lambda)$ ($i > 1, j > 1$) 不能被 $a_{11}(\lambda)$ 整除. 因为 $a_{11}(\lambda) | a_{1j}(\lambda)$, 所以存在一个多项式 $\varphi(\lambda)$ 使得 $a_{1j}(\lambda) = \varphi(\lambda)a_{11}(\lambda)$. 对 $A(\lambda)$ 作两次初等列变换, 首先将 $A(\lambda)$ 第 1 列的 $-\varphi(\lambda)$ 倍加到第 j 列, 这时第 1 行第 j 列位置的元素是 0, 第 i 行第 j 列位置的元素变为 $a_{ij}(\lambda) - \varphi(\lambda)a_{i1}(\lambda)$; 然后把第 j 列的 1 倍加到第 1 列, 此时第 1 行第 1 列位置的元素仍是 $a_{11}(\lambda)$, 而第 i 行第 1 列位置的元素变为 $a_{ij}(\lambda) + [1 - \varphi(\lambda)]a_{i1}(\lambda)$, 它不能被 $a_{11}(\lambda)$ 整除. 这就化为已经证明的情形 (1). \square

定理 3.2.3 设 $A(\lambda) = (a_{ij}(\lambda)) \in P[\lambda]^{m \times n}$, 且 $\text{rank}(A(\lambda)) = r$, 则 $A(\lambda)$ 相抵于如下“对角形”矩阵

$$\begin{pmatrix} d_1(\lambda) & & & & \\ & d_2(\lambda) & & & \\ & & \ddots & & \\ & & & d_r(\lambda) & \\ & & & & 0 \\ & & & & & \ddots \\ & & & & & & 0 \end{pmatrix}_{m \times n} \quad (3.2.3)$$

其中 $d_i(\lambda) (i=1, \dots, r)$ 是首项系数为 1 的多项式, 并且 $d_i(\lambda) \mid d_{i+1}(\lambda) (i=1, \dots, r-1)$.

证明 若 $r=0$, 则 $A(\lambda)$ 为零矩阵, 结论显然成立. 现设 $r>0$, 且 $A(\lambda) = (a_{ij}(\lambda))$ 的左上角元素 $a_{11}(\lambda) \neq 0$, 否则可通过行、列交换做到这一点. 由引理 3.1.1 知, 对 $A(\lambda)$ 进行一系列初等变换可得一个与 $A(\lambda)$ 相抵的 λ 矩阵 $B(\lambda) = (b_{ij}(\lambda))$, 并且 $b_{11}(\lambda)$ 是首项系数为 1 的多项式, $b_{11}(\lambda)$ 整除 $B(\lambda)$ 的全部元素, 即有

$$b_{ij}(\lambda) = q_{ij}(\lambda)b_{11}(\lambda), i=1, \dots, m; j=1, \dots, n$$

则可对 $B(\lambda)$ 作一系列初等变换, 使得第 1 行、第 1 列除对角元 $b_{11}(\lambda)$ 外全为零, 即

$$B(\lambda) \cong \begin{pmatrix} d_1(\lambda) & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & & \\ & & A_1(\lambda) & \\ 0 & & & \end{pmatrix}$$

其中 $d_1(\lambda) = b_{11}(\lambda)$, $A_1(\lambda)$ 是 $(m-1) \times (n-1)$ 矩阵. 因为 $A_1(\lambda)$ 的元素是 $B(\lambda)$ 中元素的组合, 而 $b_{11}(\lambda)$ (即 $d_1(\lambda)$) 整除 $B(\lambda)$ 的所有元素, 所以 $d_1(\lambda)$ 整除 $A_1(\lambda)$ 的所有元素.

如果 $A_1(\lambda) \neq 0$, 则对 $A_1(\lambda)$ 重复上述过程, 进而把矩阵化成

$$\begin{pmatrix} d_1(\lambda) & 0 & \cdots & \cdots & 0 \\ 0 & d_2(\lambda) & 0 & \cdots & 0 \\ \vdots & 0 & & & \\ \vdots & \vdots & & A_2(\lambda) & \\ 0 & 0 & & & \end{pmatrix}$$

其中 $d_1(\lambda), d_2(\lambda)$ 都是首项系数为 1 的多项式, 并且 $d_1(\lambda) | d_2(\lambda), d_2(\lambda)$ 整除 $A_2(\lambda)$ 的全部元素. 继续上述过程, 最后把 $A(\lambda)$ 化成所要求的形式. \square

定理 3.2.3 中的“对角形”矩阵 (3.2.3) 称为 λ 矩阵 $A(\lambda)$ 在相抵下的标准形或 Smith 标准形.

定义 3.2.6 λ 矩阵 $A(\lambda) \in P[\lambda]^{m \times n}$ 的 Smith 标准形“主对角线”上非零元 $d_1(\lambda), d_2(\lambda), \dots, d_r(\lambda)$ 称为 $A(\lambda)$ 的不变因子.

例 3.2.2 用初等变换把 λ 矩阵

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

化为标准形.

解

$$\begin{aligned} A(\lambda) &\xrightarrow{\substack{[1+3(1)] \\ [3+1(-1)]}} \begin{pmatrix} 1 & \lambda^2 & \lambda \\ 0 & \lambda & -\lambda \\ 1 & \lambda^2 & -\lambda^2 \end{pmatrix} \xrightarrow{[3+1(-1)]} \begin{pmatrix} 1 & \lambda^2 & \lambda \\ 0 & \lambda & -\lambda \\ 0 & 0 & -\lambda^2 - \lambda \end{pmatrix} \\ &\xrightarrow{\substack{[2+1(-\lambda^2)] \\ [3+1(-\lambda)]}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & -\lambda^2 - \lambda \end{pmatrix} \xrightarrow{\substack{[3(-1)] \\ [3+2(1)]}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda^2 + \lambda \end{pmatrix} \end{aligned}$$

例 3.2.3 用初等变换将 λ 矩阵

$$A(\lambda) = \begin{pmatrix} \lambda - a & -1 & 0 & 0 \\ 0 & \lambda - a & -1 & 0 \\ 0 & 0 & \lambda - a & -1 \\ 0 & 0 & 0 & \lambda - a \end{pmatrix}$$

化为标准形.

解

$$\begin{aligned}
 A(\lambda) &\xrightarrow{[1,2]} \begin{pmatrix} -1 & \lambda-a & 0 & 0 \\ \lambda-a & 0 & -1 & 0 \\ 0 & 0 & \lambda-a & -1 \\ 0 & 0 & 0 & \lambda-a \end{pmatrix} \\
 &\xrightarrow{[2+1(\lambda-a)]} \begin{pmatrix} -1 & \lambda-a & 0 & 0 \\ 0 & (\lambda-a)^2 & -1 & 0 \\ 0 & 0 & \lambda-a & -1 \\ 0 & 0 & 0 & \lambda-a \end{pmatrix} \\
 &\xrightarrow{\begin{matrix} [1(-1)] \\ [2+1(-\lambda+a)] \end{matrix}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & (\lambda-a)^2 & -1 & 0 \\ 0 & 0 & \lambda-a & -1 \\ 0 & 0 & 0 & \lambda-a \end{pmatrix} \\
 &\xrightarrow{[2,3]} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & (\lambda-a)^2 & 0 \\ 0 & \lambda-a & 0 & -1 \\ 0 & 0 & 0 & \lambda-a \end{pmatrix} \\
 &\xrightarrow{[3+2(\lambda-a)]} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & (\lambda-a)^2 & 0 \\ 0 & 0 & (\lambda-a)^3 & -1 \\ 0 & 0 & 0 & \lambda-a \end{pmatrix} \\
 &\xrightarrow{\begin{matrix} [3+2((\lambda-a)^2)] \\ [2(-1)] \end{matrix}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & (\lambda-a)^3 & -1 \\ 0 & 0 & 0 & \lambda-a \end{pmatrix}
 \end{aligned}$$

$$\begin{aligned}
 & \xrightarrow{[3,4]} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & (\lambda-a)^3 \\ 0 & 0 & \lambda-a & 0 \end{pmatrix} \xrightarrow{[4+3(\lambda-a)]} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & (\lambda-a)^3 \\ 0 & 0 & 0 & (\lambda-a)^4 \end{pmatrix} \\
 & \xrightarrow{\begin{matrix} [4+3((\lambda-a)^3)] \\ [3(-1)] \end{matrix}} \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & (\lambda-a)^4 \end{pmatrix}
 \end{aligned}$$

一般地

$$\begin{pmatrix} \lambda-a & & & \\ & \lambda-a & & \\ & & \ddots & \\ & & & \lambda-a \end{pmatrix}_{m \times m} \cong \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & (\lambda-a)^m \end{pmatrix}_{m \times m}$$

§ 3.3 λ 矩阵的行列式因子和初等因子

本节讨论 λ 矩阵 Smith 标准形的惟一性, 并给出两个 λ 矩阵相抵的条件. 为此, 需要引进 λ 矩阵的行列式因子.

定义 3.3.1 设 $A(\lambda) \in P[\lambda]^{m \times n}$ 且 $\text{rank}(A(\lambda)) = r$, 对于正整数 k ($1 \leq k \leq r$), $A(\lambda)$ 的全部 k 阶子式的最大公因式称为 $A(\lambda)$ 的 k 阶行列式因子, 记为 $D_k(\lambda)$.

例 3.3.1 求

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

的各阶行列式因子.

解 由于 $(1-\lambda, \lambda) = 1$, 所以 $D_1(\lambda) = 1$. 又

$$\begin{vmatrix} -\lambda + 1 & \lambda^2 \\ \lambda & \lambda \end{vmatrix} = \lambda(-\lambda^2 - \lambda + 1) = \varphi_1(\lambda),$$

$$\begin{vmatrix} -\lambda + 1 & \lambda^2 \\ \lambda^2 + 1 & \lambda^2 \end{vmatrix} = \lambda^3(-\lambda - 1) = \varphi_2(\lambda)$$

故 $(\varphi_1(\lambda), \varphi_2(\lambda)) = \lambda$. 其余的二阶子式(还有 7 个)都包含因子 λ , 所以 $D_2(\lambda) = \lambda$.

最后, 由于 $\det(A(\lambda)) = -\lambda^3 - \lambda^2$, 所以 $D_3(\lambda) = \lambda^3 + \lambda^2$.

行列式因子的重要性在于它在初等变换下是不变的.

定理 3.3.1 相抵的 λ 矩阵具有相同的秩和相同的各阶行列式因子.

证明 只要证明 λ 矩阵经过一次初等变换后, 其秩与行列式因子不变.

设 λ 矩阵 $A(\lambda)$ 经过一次初等行变换变成 $B(\lambda)$, $f(\lambda)$ 与 $g(\lambda)$ 分别是 $A(\lambda)$ 与 $B(\lambda)$ 的 k 阶行列式因子. 针对 3 种初等变换来证明 $f(\lambda) = g(\lambda)$.

(1) 交换 $A(\lambda)$ 的某两行得到 $B(\lambda)$. 这时 $B(\lambda)$ 的每个 k 阶子式或者等于 $A(\lambda)$ 的某个 k 阶子式, 或者是 $A(\lambda)$ 的某个 k 阶子式的 -1 倍. 因此 $f(\lambda)$ 是 $B(\lambda)$ 的 k 阶子式的公因式, 从而 $f(\lambda) | g(\lambda)$.

(2) 用非零数 α 乘 $A(\lambda)$ 的某一行得到 $B(\lambda)$. 这时 $B(\lambda)$ 的每个 k 阶子式或者等于 $A(\lambda)$ 的某个 k 阶子式, 或者等于 $A(\lambda)$ 的某个 k 阶子式的 α 倍. 因此 $f(\lambda)$ 是 $B(\lambda)$ 的 k 阶子式的公因式, 从而 $f(\lambda) | g(\lambda)$.

(3) 将 $A(\lambda)$ 第 j 行的 $\varphi(\lambda)$ 倍加到第 i 行得到 $B(\lambda)$. 这时, $B(\lambda)$ 中那些包含第 i 行与第 j 行的 k 阶子式和那些不包含第 i 行的 k 阶子式都等于 $A(\lambda)$ 中对应的 k 阶子式; $B(\lambda)$ 中那些包含第 i 行但不包含第 j 行的 k 阶子式等于 $A(\lambda)$ 中对应的一个 k 阶子式与另一个 k 阶子式的 $\pm \varphi(\lambda)$ 倍之和, 也就是 $A(\lambda)$ 的两个 k 阶子式的组合. 因此 $f(\lambda)$ 是 $B(\lambda)$ 的 k 阶子式的公因式, 从而 $f(\lambda) | g(\lambda)$.

由初等变换的可逆性, $B(\lambda)$ 也可以经过一次初等行变换变成 $A(\lambda)$. 由上面的讨论, 同样有 $g(\lambda) | f(\lambda)$, 所以 $f(\lambda) = g(\lambda)$.

对于初等列变换, 可以完全一样地讨论. 总之, 如果 $A(\lambda)$ 经过一次初等变换变成 $B(\lambda)$, 则 $f(\lambda) = g(\lambda)$.

当 $A(\lambda)$ 的全部 k 阶子式为零时, $f(\lambda) = 0$, 则 $g(\lambda) = 0$, $B(\lambda)$ 的全部 k 阶子式也为零; 反之亦然. 因此 $A(\lambda)$ 与 $B(\lambda)$ 既有相同的行列式因子, 又有相同的秩. \square

由定理 3.3.1 知, 任意 λ 矩阵的秩和行列式因子与其 Smith 标准形的秩和行列式因子是相同的.

设 λ 矩阵 $A(\lambda)$ 的 Smith 标准形为

$$\begin{pmatrix} d_1(\lambda) & & & & & \\ & d_2(\lambda) & & & & \\ & & \ddots & & & \\ & & & d_r(\lambda) & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{pmatrix} \quad (3.3.1)$$

其中 $d_i(\lambda) (i=1, \dots, r)$ 是首项系数为 1 的多项式, 并且 $d_i(\lambda) \mid d_{i+1}(\lambda) (i=1, \dots, r-1)$.

容易求得 $A(\lambda)$ 的各阶行列式因子如下:

$$\begin{cases} D_1(\lambda) = d_1(\lambda) \\ D_2(\lambda) = d_1(\lambda)d_2(\lambda) \\ \dots\dots\dots \\ D_r(\lambda) = d_1(\lambda)d_2(\lambda)\cdots d_r(\lambda) \end{cases} \quad (3.3.2)$$

于是有

$$\begin{cases} (1) & D_1(\lambda) \mid D_2(\lambda), D_2(\lambda) \mid D_3(\lambda), \dots, D_{r-1}(\lambda) \mid D_r(\lambda) \\ (2) & d_1(\lambda) = D_1(\lambda), d_2(\lambda) = D_2(\lambda)/D_1(\lambda), \dots, d_r(\lambda) = D_r(\lambda)/D_{r-1}(\lambda) \end{cases} \quad (3.3.3)$$

从而得如下结论.

定理 3.3.2 λ 矩阵 $A(\lambda)$ 的 Smith 标准形是惟一的.

证明 因为 $A(\lambda)$ 的各阶行列式因子是惟一的, 则由 (3.3.3) 知 $A(\lambda)$ 的不变因子也是惟一的. 因此 $A(\lambda)$ 的 Smith 标准形是惟一的. \square

应用 λ 矩阵的 Smith 标准形, 可以证明如下定理.

定理 3.3.3 设 $A(\lambda), B(\lambda) \in P[\lambda]^{m \times n}$, 则 $A(\lambda)$ 与 $B(\lambda)$ 相抵的充分必要条件是它们有相同的行列式因子, 或者它们有相同的不变因子.

证明 因为 λ 矩阵的行列式因子与不变因子是互相完全确定的, 所以两个 λ 矩阵有相同的行列式因子, 则它们有相同的不变因子; 反之亦然.

必要性由定理 3.3.1 即得.

充分性. 若 λ 矩阵 $A(\lambda)$ 与 $B(\lambda)$ 有相同的不变因子, 则 $A(\lambda)$ 与 $B(\lambda)$ 和同一 Smith 标准形相抵, 因而 $A(\lambda)$ 与 $B(\lambda)$ 相抵. \square

一般说来,应用行列式因子求不变因子比较复杂,但对一些特殊的 λ 矩阵,先求行列式因子再求不变因子反而简单.

例 3.3.2 求

$$A(\lambda) = \begin{pmatrix} \lambda - a & -1 & & 0 \\ 0 & \lambda - a & \ddots & \\ \vdots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & \lambda - a \end{pmatrix}_{m \times m}$$

的行列式因子和不变因子.

解 由于 $A(\lambda)$ 的一个 $m-1$ 阶子式

$$\begin{vmatrix} -1 & & & \\ & \lambda - a & -1 & \\ & & \ddots & \ddots \\ & & & \lambda - a & -1 \end{vmatrix} = (-1)^{m-1}$$

故 $D_{m-1}(\lambda) = 1$. 由 (3.3.3) 的第一式,即行列式因子的“依次”整除性,有

$$D_1(\lambda) = D_2(\lambda) = \cdots = D_{m-2}(\lambda) = 1$$

而 $D_m(\lambda) = (\lambda - a)^m$. 因此 $A(\lambda)$ 的不变因子为

$$d_1(\lambda) = d_2(\lambda) = \cdots = d_{m-1}(\lambda) = 1, \quad d_m(\lambda) = (\lambda - a)^m$$

由此可知 $A(\lambda)$ 的标准形为

$$A(\lambda) \cong \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & (\lambda - a)^m \end{pmatrix}_{m \times m}$$

定理 3.3.4 设 $A(\lambda) \in P[\lambda]^{n \times n}$, 则 $A(\lambda)$ 可逆的充分必要条件是 $A(\lambda)$ 可表示为一系列初等矩阵的乘积.

证明 必要性. 设 $A(\lambda)$ 为一 n 阶可逆矩阵, 则由定理 3.2.1 知 $|A(\lambda)| = d \neq 0$, 从而 $A(\lambda)$ 的行列式因子为

$$D_1(\lambda) = D_2(\lambda) = \cdots = D_n(\lambda) = 1$$

于是 $A(\lambda)$ 的不变因子为

$$d_1(\lambda) = d_2(\lambda) = \cdots = d_n(\lambda) = 1$$

因此 $A(\lambda)$ 与单位矩阵相抵, 即存在一系列初等矩阵 $P_1(\lambda), \cdots, P_t(\lambda), Q_1(\lambda), \cdots, Q_r(\lambda)$ 使得

$$A(\lambda) = P_t(\lambda) \cdots P_1(\lambda) I Q_1(\lambda) \cdots Q_r(\lambda) = P_t(\lambda) \cdots P_1(\lambda) Q_1(\lambda) \cdots Q_r(\lambda)$$

充分性. 设 $A(\lambda)$ 可表示为一系列初等矩阵的乘积, 即存在一系列初等矩阵 $P_1(\lambda), \cdots, P_t(\lambda), Q_1(\lambda), \cdots, Q_r(\lambda)$ 使得

$$A(\lambda) = P_t(\lambda) \cdots P_1(\lambda) Q_1(\lambda) \cdots Q_r(\lambda)$$

则 $A(\lambda)$ 的行列式是一个非零常数. 因此由定理 3.2.1 知 $A(\lambda)$ 可逆. \square

利用定理 3.2.2 和定理 3.3.4 容易证明下面定理.

定理 3.3.5 设 $A(\lambda), B(\lambda) \in P[\lambda]^{m \times n}$, 则 $A(\lambda)$ 与 $B(\lambda)$ 相抵的充分必要条件是存在两个可逆 λ 矩阵 $P(\lambda) \in P[\lambda]^{m \times m}$ 与 $Q(\lambda) \in P[\lambda]^{n \times n}$ 使得 $B(\lambda) = P(\lambda)A(\lambda)Q(\lambda)$.

下面再引进 λ 矩阵的初等因子. 设 λ 矩阵 $A(\lambda)$ 的不变因子为 $d_1(\lambda), d_2(\lambda), \cdots, d_r(\lambda)$, 在复数域内将它们分解成一次因式的幂的乘积:

$$\begin{cases} d_1(\lambda) = (\lambda - \lambda_1)^{e_{11}} (\lambda - \lambda_2)^{e_{12}} \cdots (\lambda - \lambda_s)^{e_{1s}} \\ d_2(\lambda) = (\lambda - \lambda_1)^{e_{21}} (\lambda - \lambda_2)^{e_{22}} \cdots (\lambda - \lambda_s)^{e_{2s}} \\ \cdots \cdots \cdots \\ d_r(\lambda) = (\lambda - \lambda_1)^{e_{r1}} (\lambda - \lambda_2)^{e_{r2}} \cdots (\lambda - \lambda_s)^{e_{rs}} \end{cases} \quad (3.3.4)$$

其中 $\lambda_1, \cdots, \lambda_s$ 是互异的复数, e_{ij} 是非负整数. 因为 $d_i(\lambda) \mid d_{i+1}(\lambda) (i = 1, \cdots, r-1)$, 所以 e_{ij} 满足如下关系

$$\begin{cases} 0 \leq e_{11} \leq e_{21} \leq \cdots \leq e_{r1} \\ 0 \leq e_{12} \leq e_{22} \leq \cdots \leq e_{r2} \\ \cdots \cdots \cdots \\ 0 \leq e_{1s} \leq e_{2s} \leq \cdots \leq e_{rs} \end{cases}$$

定义 3.3.2 在 (3.3.4) 式中, 所有指数大于零的因子

$$(\lambda - \lambda_j)^{e_{ij}}, e_{ij} > 0, i = 1, \cdots, r, j = 1, \cdots, s$$

称为 λ 矩阵 $A(\lambda)$ 的初等因子.

例如, 若 λ 矩阵 $A(\lambda)$ 的不变因子为

$$\begin{cases} d_1(\lambda) = 1 \\ d_2(\lambda) = \lambda(\lambda - 1) \\ d_3(\lambda) = \lambda(\lambda - 1)^2(\lambda + 1)^2 \\ d_4(\lambda) = \lambda^2(\lambda - 1)^3(\lambda + 1)^3(\lambda - 2) \end{cases}$$

则 $A(\lambda)$ 的初等因子为 $\lambda, \lambda, \lambda^2, \lambda - 1, (\lambda - 1)^2, (\lambda - 1)^3, (\lambda + 1)^2, (\lambda + 1)^3, \lambda - 2$.

由定义 3.3.2 知, 若给定 λ 矩阵 $A(\lambda)$ 的不变因子, 则可惟一确定其初等因子; 反过来, 如果知道一个 λ 矩阵的秩和初等因子, 则也可惟一确定它的不变因子. 事实上, λ 矩阵 $A(\lambda)$ 的秩 r 确定了不变因子的个数, 同一个一次因式的方幂作成的初等因子中, 方次最高的必在 $d_r(\lambda)$ 的分解中, 方次次高的必在 $d_{r-1}(\lambda)$ 的分解中. 如此顺推下去, 可知属于同一个一次因式的方幂的初等因子在不变因子的分解式中出现的位置是惟一确定的.

例如, 若已知 5×6 λ 矩阵 $A(\lambda)$ 的秩为 4, 其初等因子为

$$\lambda, \lambda, \lambda^2, \lambda - 1, (\lambda - 1)^2, (\lambda - 1)^3, (\lambda + i)^3, (\lambda - i)^3$$

则可求得 $A(\lambda)$ 的不变因子

$$d_4(\lambda) = \lambda^2(\lambda - 1)^3(\lambda + i)^3(\lambda - i)^3$$

$$d_3(\lambda) = \lambda(\lambda - 1)^2$$

$$d_2(\lambda) = \lambda(\lambda - 1)$$

$$d_1(\lambda) = 1$$

从而 $A(\lambda)$ 的 Smith 标准形为

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

由定理 3.3.3 以及不变因子与初等因子之间的关系容易导出如下定理.

定理 3.3.6 设 $A(\lambda), B(\lambda) \in P[\lambda]^{m \times n}$, 则 $A(\lambda)$ 与 $B(\lambda)$ 相抵的充分必要条件是它们有相同的秩和相同的初等因子.

对块对角矩阵

$$A(\lambda) = \begin{bmatrix} B(\lambda) & 0 \\ 0 & C(\lambda) \end{bmatrix}$$

不能从 $B(\lambda)$ 与 $C(\lambda)$ 的不变因子求得 $A(\lambda)$ 的不变因子, 但是能从 $B(\lambda)$ 与 $C(\lambda)$ 的初等因子求得 $A(\lambda)$ 的初等因子.

定理 3.3.7 设 λ 矩阵

$$A(\lambda) = \begin{bmatrix} B(\lambda) & 0 \\ 0 & C(\lambda) \end{bmatrix} \quad (3.3.5)$$

为块对角矩阵, 则 $B(\lambda)$ 与 $C(\lambda)$ 的初等因子的全体是 $A(\lambda)$ 的全部初等因子.

证明 将 $B(\lambda)$ 与 $C(\lambda)$ 分别化为 Smith 标准形

$$B(\lambda) \cong \begin{bmatrix} b_1(\lambda) & & & & \\ & \ddots & & & \\ & & b_{r_B}(\lambda) & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}$$

$$C(\lambda) \cong \begin{bmatrix} c_1(\lambda) & & & & \\ & \ddots & & & \\ & & c_{r_C}(\lambda) & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}$$

其中 $r_B = \text{rank}(B(\lambda))$, $r_C = \text{rank}(C(\lambda))$, $b_1(\lambda), \dots, b_{r_B}(\lambda)$ 与 $c_1(\lambda), \dots, c_{r_C}(\lambda)$ 分别为 $B(\lambda)$ 与 $C(\lambda)$ 的不变因子, 则 $\text{rank}(A(\lambda)) = r = r_B + r_C$.

把 $b_i(\lambda)$ 和 $c_j(\lambda)$ 分解为不同的一次因式的方幂的乘积

$$b_i(\lambda) = (\lambda - \lambda_1)^{b_{i1}} (\lambda - \lambda_2)^{b_{i2}} \cdots (\lambda - \lambda_s)^{b_{is}}, i = 1, \dots, r_B$$

$$c_j(\lambda) = (\lambda - \lambda_1)^{c_{j1}} (\lambda - \lambda_2)^{c_{j2}} \cdots (\lambda - \lambda_s)^{c_{js}}, j = 1, \dots, r_C$$

则 $B(\lambda)$ 与 $C(\lambda)$ 的初等因子分别为

$$(\lambda - \lambda_1)^{b_{11}}, (\lambda - \lambda_2)^{b_{12}}, \dots, (\lambda - \lambda_r)^{b_{1r}}, j = 1, \dots, r_B$$

和

$$(\lambda - \lambda_1)^{c_{11}}, (\lambda - \lambda_2)^{c_{12}}, \dots, (\lambda - \lambda_r)^{c_{1r}}, j = 1, \dots, r_C$$

中非常数的多项式.

我们先证明 $B(\lambda)$ 与 $C(\lambda)$ 的初等因子是 $A(\lambda)$ 的全部初等因子. 不失一般性, 仅考虑 $B(\lambda)$ 与 $C(\lambda)$ 中只含 $\lambda - \lambda_1$ 的方幂的那些初等因子, 将 $\lambda - \lambda_1$ 的指数

$$b_{11}, b_{21}, \dots, b_{r_B 1}, c_{11}, c_{21}, \dots, c_{r_C 1}$$

按由小到大的顺序排列, 记为 $0 \leq j_1 \leq j_2 \leq \dots \leq j_r$. 由 (3.3.5) 可知, 对 $B(\lambda)$ 与 $C(\lambda)$ 进行初等变换实际上是对 $A(\lambda)$ 进行初等变换, 于是

$$\begin{aligned} A(\lambda) &\cong \begin{pmatrix} b_1(\lambda) & & & & & \\ & \ddots & & & & \\ & & b_{r_B}(\lambda) & & & \\ & & & c_1(\lambda) & & \\ & & & & \ddots & \\ & & & & & c_{r_C}(\lambda) \\ & & & & & & 0 \\ & & & & & & & \ddots \\ & & & & & & & & 0 \end{pmatrix} \\ &\cong \begin{pmatrix} (\lambda - \lambda_1)^{j_1} \varphi_1(\lambda) & & & & & \\ & (\lambda - \lambda_1)^{j_2} \varphi_2(\lambda) & & & & \\ & & \ddots & & & \\ & & & (\lambda - \lambda_1)^{j_r} \varphi_r(\lambda) & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{pmatrix} \end{aligned}$$

其中多项式 $\varphi_1(\lambda), \dots, \varphi_r(\lambda)$ 都不含因式 $\lambda - \lambda_1$.

设 $A(\lambda)$ 的行列式因子和不变因子分别为 $D_1(\lambda), D_2(\lambda), \dots, D_r(\lambda)$ 和 $d_1(\lambda), d_2(\lambda), \dots, d_r(\lambda)$, 则在行列式因子中因子 $\lambda - \lambda_1$ 的幂指数分别为 $j_1, j_1 + j_2, \dots, \sum_{i=1}^{r-1} j_i, \sum_{i=1}^r j_i$, 而由行列式因子与不变因子的关系 (3.3.3) 知, $d_1(\lambda), d_2(\lambda), \dots, d_r(\lambda)$ 中因子 $\lambda - \lambda_1$ 的幂指数分别为 $j_1, j_2, \dots, j_{r-1}, j_r$. 因此 $A(\lambda)$ 中与 $\lambda - \lambda_1$ 相应的初等因子是

$$(\lambda - \lambda_1)^{j_i}, j_i > 0, i = 1, \dots, r$$

也就是 $B(\lambda), C(\lambda)$ 中与 $\lambda - \lambda_1$ 相应的全部初等因子.

对 $\lambda - \lambda_2, \lambda - \lambda_3, \dots, \lambda - \lambda_l$ 进行类似的讨论, 可得相同结论. 于是 $B(\lambda), C(\lambda)$ 的全部初等因子都是 $A(\lambda)$ 的初等因子.

下面证明, 除 $B(\lambda), C(\lambda)$ 的初等因子外, $A(\lambda)$ 再没有其他的初等因子.

因为 $D_r(\lambda)$ 为 $A(\lambda)$ 的所有初等因子的乘积, 而

$$D_r(\lambda) = b_1(\lambda) \cdots b_{r_B}(\lambda) c_1(\lambda) \cdots c_{r_C}(\lambda)$$

如果 $(\lambda - a)^k$ 是 $A(\lambda)$ 的初等因子, 则它必包含在某个 $b_i(\lambda) (i = 1, \dots, r_B)$ 或 $c_j(\lambda) (j = 1, \dots, r_C)$ 中, 即 $A(\lambda)$ 的初等因子包含在 $B(\lambda)$ 与 $C(\lambda)$ 的初等因子中. 因此, 除 $B(\lambda)$ 与 $C(\lambda)$ 的全部初等因子外, $A(\lambda)$ 再没有别的初等因子.

□

定理 3.3.7 可推广为

定理 3.3.8 若 λ 矩阵 $A(\lambda)$ 等价于块对角矩阵

$$A(\lambda) \cong \begin{pmatrix} A_1(\lambda) & & & \\ & A_2(\lambda) & & \\ & & \ddots & \\ & & & A_t(\lambda) \end{pmatrix}$$

则 $A_1(\lambda), A_2(\lambda), \dots, A_t(\lambda)$ 各个初等因子的全体就是 $A(\lambda)$ 的全部初等因子.

对 t 应用数学归纳法, 请读者自行证明.

例 3.3.3 求 λ 矩阵

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

的初等因子,不变因子和标准形.

解 记 $A_1(\lambda) = \lambda^2 + \lambda, A_2(\lambda) = \lambda, A_3(\lambda) = \begin{bmatrix} (\lambda+1)^2 & \lambda+1 \\ -2 & \lambda-2 \end{bmatrix}$, 则

$$A(\lambda) = \begin{bmatrix} A_1(\lambda) & 0 & 0 \\ 0 & A_2(\lambda) & 0 \\ 0 & 0 & A_3(\lambda) \end{bmatrix}$$

对于 $A_3(\lambda)$, 其初等因子为 $\lambda, \lambda-1, \lambda+1$. 利用定理 3.3.8, 可得 $A(\lambda)$ 的初等因子

$$\lambda, \lambda, \lambda, \lambda-1, \lambda+1, \lambda+1$$

因为 $A(\lambda)$ 的秩为 4, 故 $A(\lambda)$ 的不变因子为

$$d_4(\lambda) = \lambda(\lambda-1)(\lambda+1), d_3(\lambda) = \lambda(\lambda+1), d_2(\lambda) = \lambda, d_1(\lambda) = 1$$

因此 $A(\lambda)$ 的 Smith 标准形为

$$A(\lambda) \cong \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda(\lambda+1) & 0 \\ 0 & 0 & 0 & \lambda(\lambda+1)(\lambda-1) \end{bmatrix}$$

§ 3.4 矩阵相似的条件

设 A 是 n 阶数字矩阵, 其特征矩阵 $\lambda I - A$ 是 λ 矩阵, 它是研究数字矩阵的重要工具. 应用特征矩阵可以给出两个 n 阶数字矩阵 A 与 B 之间相似性的判断准则. 为此, 我们先证明两个引理.

引理 3.4.1 设 A, B 是两个 n 阶矩阵, 如果存在 n 阶数字矩阵 P, Q 使得

$$\lambda I - A = P(\lambda I - B)Q \quad (3.4.1)$$

则 A 与 B 相似.

证明 比较 (3.4.1) 两边 λ 的同次幂的系数矩阵, 得

$$PQ = I, A = PBQ$$

由此 $Q = P^{-1}, A = PBQ$, 故 A 与 B 相似. \square

引理 3.4.2 设 A 是 n 阶非零数字矩阵, $U(\lambda)$ 与 $V(\lambda)$ 是 n 阶 λ 矩阵,

则存在 n 阶 λ 矩阵 $Q(\lambda)$ 与 $R(\lambda)$ 以及 n 阶数字矩阵 U_0 及 V_0 , 使得

$$U(\lambda) = (\lambda I - A)Q(\lambda) + U_0 \quad (3.4.2)$$

$$V(\lambda) = R(\lambda)(\lambda I - A) + V_0 \quad (3.4.3)$$

证明 (3.4.2) 与 (3.4.3) 的证明类似, 这里仅证 (3.4.2) 式. 把 $U(\lambda)$ 改写成

$$U(\lambda) = D_0 \lambda^m + D_1 \lambda^{m-1} + \cdots + D_{m-1} \lambda + D_m$$

其中 D_0, D_1, \cdots, D_m 都是 n 阶数字矩阵, 并且 $D_0 \neq 0$.

(1) 若 $m=0$, 则取 $Q(\lambda)=0$ 及 $U_0=D_0$, 它们满足要求, 并且 (3.4.2) 成立.

(2) 若 $m>0$, 令

$$Q(\lambda) = Q_0 \lambda^{m-1} + Q_1 \lambda^{m-2} + \cdots + Q_{m-2} \lambda + Q_{m-1}$$

其中 $Q_0, Q_1, \cdots, Q_{m-1}$ 是待定的 n 阶数字矩阵. 由

$$\begin{aligned} (\lambda I - A)Q(\lambda) &= Q_0 \lambda^m + (Q_1 - AQ_0) \lambda^{m-1} + \cdots \\ &+ (Q_k - AQ_{k-1}) \lambda^{m-k} + \cdots + (Q_{m-1} - AQ_{m-2}) \lambda - AQ_{m-1} \end{aligned}$$

取 $Q_0 = D_0, Q_1 = D_1 + AQ_0, Q_2 = D_2 + AQ_1, \cdots, Q_{m-1} = D_{m-1} + AQ_{m-2}, U_0 = D_m + AQ_{m-1}$, 则 (3.4.2) 成立. \square

定理 3.4.1 n 阶矩阵 A 与 B 相似的充分必要条件是它们的特征矩阵 $\lambda I - A$ 和 $\lambda I - B$ 相抵.

证明 必要性. 若 A 与 B 相似, 则存在非奇异矩阵 P 使 $P^{-1}AP = B$, 从而

$$P^{-1}(\lambda I - A)P = \lambda I - B$$

因为 P, P^{-1} 是可逆的 λ 矩阵, 所以上式表明 $\lambda I - A$ 与 $\lambda I - B$ 相抵.

充分性. 设 $\lambda I - A$ 与 $\lambda I - B$ 相抵, 由定理 3.3.5 知存在可逆的 λ 矩阵 $U(\lambda), V(\lambda)$ 使

$$\lambda I - A = U(\lambda)(\lambda I - B)V(\lambda) \quad (3.4.4)$$

由引理 3.4.2, 存在 λ 矩阵 $Q(\lambda)$ 与 $R(\lambda)$ 以及数字矩阵 U_0 及 V_0 , 使得

$$U(\lambda) = (\lambda I - A)Q(\lambda) + U_0 \quad (3.4.5)$$

$$V(\lambda) = R(\lambda)(\lambda I - A) + V_0 \quad (3.4.6)$$

则(3.4.4)式改写为

$$U(\lambda)^{-1}(\lambda I - A) = (\lambda I - B)V(\lambda) \quad (3.4.7)$$

$$(\lambda I - A)V(\lambda)^{-1} = U(\lambda)(\lambda I - B) \quad (3.4.8)$$

将 $V(\lambda)$ 的表达式(3.4.6)代入(3.4.7),得

$$[U(\lambda)^{-1} - (\lambda I - B)R(\lambda)](\lambda I - A) = (\lambda I - B)V_0$$

因为上式右边 λ 的次数 ≤ 1 , 所以 $U(\lambda)^{-1} - (\lambda I - B)R(\lambda)$ 是数字矩阵, 记为 T , 即

$$T = U(\lambda)^{-1} - (\lambda I - B)R(\lambda) \quad (3.4.9)$$

从而

$$T(\lambda I - A) = (\lambda I - B)V_0 \quad (3.4.10)$$

由(3.4.9), 并利用(3.4.5)和(3.4.8), 得

$$\begin{aligned} I &= U(\lambda)T + U(\lambda)(\lambda I - B)R(\lambda) \\ &= U(\lambda)T + (\lambda I - A)V(\lambda)^{-1}R(\lambda) \\ &= [(\lambda I - A)Q(\lambda) + U_0]T + (\lambda I - A)V(\lambda)^{-1}R(\lambda) \\ &= U_0T + (\lambda I - A)[Q(\lambda)T + V(\lambda)^{-1}R(\lambda)] \end{aligned}$$

上式右边第二项必为零; 否则右边 λ 的次数至少是 1, 等式不可能成立. 因此 $I = U_0T$. 从而 U_0, T 可逆, 并且 $T^{-1} = U_0$. 由(3.4.10), 得

$$\lambda I - A = U_0(\lambda I - B)V_0$$

由引理 3.4.1 知 A 与 B 相似. \square

定义 3.4.1 设 A 是 n 阶数字矩阵, 其特征矩阵 $\lambda I - A$ 的行列式因子, 不变因子和初等因子分别称为矩阵 A 的行列式因子, 不变因子和初等因子.

由定理 3.3.3 和定理 3.4.1 立即得

定理 3.4.2 n 阶矩阵 A 与 B 相似的充分必要条件是它们有相同的行列式因子, 或者它们有相同的不变因子.

由例 3.2.1, 定理 3.3.6 和定理 3.4.1 得

定理 3.4.3 n 阶矩阵 A 与 B 相似的充分必要条件是它们有相同的初等因子.

§ 3.5 矩阵的 Jordan 标准形

定义 3.5.1 形状为

$$J_i = \begin{bmatrix} \lambda_i & 1 & & 0 \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{bmatrix}_{n_i \times n_i} \quad (3.5.1)$$

的矩阵称为 **Jordan 块**, 其中 λ_i 为复数. 由若干个 Jordan 块为对角块组成的块对角矩阵称为 **Jordan 形矩阵**.

例如, 矩阵

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i & 1 & 0 \\ 0 & 0 & 0 & 0 & -i & 1 \\ 0 & 0 & 0 & 0 & 0 & -i \end{bmatrix}$$

是一个 Jordan 形矩阵.

容易验证, n_i 阶 Jordan 块 J_i 具有如下性质:

(1) J_i 具有一个 n_i 重特征值 λ_i , 对应于特征值 λ_i 仅有一个线性无关的特征向量.

(2) J_i 的乘幂有明显的表示式

$$J_i^p = \begin{bmatrix} f_p(\lambda_i) & f'_p(\lambda_i) & \frac{1}{2!}f''_p(\lambda_i) & \cdots & \cdots & \frac{1}{(n_i-1)!}f_p^{(n_i-1)}(\lambda_i) \\ & & & \ddots & & \vdots \\ & f_p(\lambda_i) & f'_p(\lambda_i) & & \ddots & \vdots \\ & & \ddots & \ddots & & \frac{1}{2!}f''_p(\lambda_i) \\ & & & \ddots & & \vdots \\ 0 & & & \ddots & f'_p(\lambda_i) & \\ & & & & f_p(\lambda_i) & \end{bmatrix}, p = 1, 2, \cdots$$

其中 $f_p(\lambda) = \lambda^p$.

(3) J_i 的不变因子为

$$d_1(\lambda) = \cdots = d_{n_i-1}(\lambda) = 1, \quad d_{n_i}(\lambda) = (\lambda - \lambda_i)^{n_i}$$

从而 J_i 的初等因子为 $(\lambda - \lambda_i)^{n_i}$.

设

$$J = \text{diag}(J_1, J_2, \cdots, J_s)$$

是 Jordan 形矩阵, 其中 J_i 为形如 (3.5.1) 的 Jordan 块. J 的特征矩阵为

$$\lambda I - J = \text{diag}(\lambda I_{n_1} - J_1, \cdots, \lambda I_{n_s} - J_s)$$

由定理 3.3.8 知 Jordan 形矩阵 J 的初等因子为

$$(\lambda - \lambda_1)^{n_1}, (\lambda - \lambda_2)^{n_2}, \cdots, (\lambda - \lambda_s)^{n_s}$$

可见, Jordan 形矩阵的全部初等因子由它的全部 Jordan 块的初等因子组成, 而 Jordan 块被它的初等因子惟一决定. 因此, Jordan 形矩阵除去其中 Jordan 块排列的次序外被它的初等因子惟一决定.

定理 3.5.1 设 $A \in \mathbb{C}^{n \times n}$, 则 A 与一个 Jordan 形矩阵相似, 并且 Jordan 形矩阵除去其中 Jordan 块的排列次序外是被矩阵 A 惟一决定的.

证明 设 A 的初等因子为

$$(\lambda - \lambda_1)^{n_1}, (\lambda - \lambda_2)^{n_2}, \cdots, (\lambda - \lambda_s)^{n_s} \quad (3.5.2)$$

其中 $\lambda_1, \cdots, \lambda_s$ 可能有相同的, n_1, \cdots, n_s 也可能有相同的. 每个初等因子 $(\lambda - \lambda_i)^{n_i}$ 对应于一个 Jordan 块

$$J_i = \begin{pmatrix} \lambda_i & 1 & & 0 \\ & \lambda_i & 1 & \\ & & \ddots & \ddots \\ & & & \ddots & 1 \\ 0 & & & & \lambda_i \end{pmatrix}_{n_i \times n_i}, \quad i = 1, \cdots, s$$

这些 Jordan 块构成一个 Jordan 形矩阵

$$J = \text{diag}(J_1, J_2, \cdots, J_s) \quad (3.5.3)$$

其初等因子也是 (3.5.2). 因为 J 与 A 有相同的初等因子, 由定理 3.4.3 知 J

与 A 相似. Jordan 形矩阵 (3.5.3) 称为矩阵 A 的 **Jordan 标准形**.

若有另一个 Jordan 形矩阵 J 与 A 相似, 则 J 与 A 有相同的初等因子. 因此, J 与 J 除去其中 Jordan 块排列的次序外是相同的, 这就证明了惟一性. \square

利用矩阵在相似变换下的 Jordan 标准形, 可得线性变换的结构.

定理 3.5.2 设 \mathcal{A} 是复数域上 n 维线性空间 V 的线性变换, 则在 V 中存在一组基使得 \mathcal{A} 在这组基下的矩阵是 Jordan 形矩阵.

证明 在 V 中任取一组基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$, 设线性变换 \mathcal{A} 在这组基下的矩阵是 A . 由定理 3.5.1 知, 存在可逆矩阵 P 使得 $P^{-1}AP = J$ 为 Jordan 形矩阵. 令

$$(\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n) = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)P$$

则线性变换 \mathcal{A} 在基 $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$ 下的矩阵是 $P^{-1}AP = J$ 为 Jordan 形矩阵. \square

如果 $n_i = 1$, 则 $J_i = \lambda_i$ 是一阶 Jordan 块. 当矩阵 A 的 Jordan 标准形中的 Jordan 块都是一阶块时, A 的 Jordan 标准形就是对角矩阵. 因为一阶 Jordan 块的初等因子是一次的, 所以对角矩阵的初等因子都是一次的. 由此得

定理 3.5.3 设 $A \in \mathbb{C}^{n \times n}$, 则 A 与一个对角矩阵相似的充分必要条件是 A 的初等因子都是一次的.

例 3.5.1 求矩阵

$$A = \begin{bmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{bmatrix}$$

的 Jordan 标准形.

解 因为

$$\lambda I - A = \begin{bmatrix} \lambda + 1 & 2 & -6 \\ 1 & \lambda & -3 \\ 1 & 1 & \lambda - 4 \end{bmatrix} \cong \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda - 1 & 0 \\ 0 & 0 & (\lambda - 1)^2 \end{bmatrix}$$

则 A 的初等因子为 $\lambda - 1, (\lambda - 1)^2$, 故 A 的 Jordan 标准形为

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

由定理 3.5.1 知, 对任意的 n 阶矩阵 A , 存在 n 阶可逆矩阵 P 使得 $P^{-1}AP = J$ 为 Jordan 标准形. 下面介绍求变换矩阵 P 的方法. 先看一下例子.

例 3.5.2 求化矩阵

$$A = \begin{pmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{pmatrix}$$

为 Jordan 标准形的变换矩阵.

解 由例 3.5.1 知, 存在 3 阶可逆矩阵 P 使得

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

记 $P = (p_1, p_2, p_3)$, 则得

$$(Ap_1, Ap_2, Ap_3) = (p_1, p_2, p_3) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

比较上式两边得

$$\begin{cases} Ap_1 = p_1 \\ Ap_2 = p_2 \\ Ap_3 = p_2 + p_3 \end{cases}$$

由此可见, p_1, p_2 是 A 的对应于特征值 1 的两个线性无关的特征向量.

从方程组

$$(I - A)x = 0$$

$$\text{可求得两个线性无关的特征向量 } \xi = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \eta = \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}.$$

可以取 $p_1 = \xi$, 但不能简单地取 $p_2 = \eta$, 因为 p_2 的选取应保证非齐次线性方程组 $(I - A)p_3 = -p_2$ 有解. 由于 ξ, η 的线性组合仍是 $(I - A)x = 0$ 的解, 因此我们选取 $p_2 = k_1\xi + k_2\eta$, 其中待定常数 k_1, k_2 只要保证 p_1 与 p_2 线性无关, 且使得 $(I - A)p_3 = -p_2$ 有解. 因为 $p_2 = k_1\xi + k_2\eta = (-k_1 + 3k_2, k_1, k_2)^T$, 所以选取 k_1, k_2 使得方程组

$$\begin{pmatrix} 2 & 2 & -6 \\ 1 & 1 & -3 \\ 1 & 1 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} k_1 - 3k_2 \\ -k_1 \\ -k_2 \end{pmatrix}$$

有解. 容易看出, 当 $k_1 = k_2$ 时方程组有解, 且其解为

$$x_1 = x_2 + 3x_3 - k_1$$

其中 k_1 是任意非零常数. 取 $k_1 = 1$, 可得 $p_2 = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$, $p_3 = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$. 于是 $P =$

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

$$P^{-1}AP = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

一般地, 设 $A \in \mathbb{C}^{n \times n}$, 则存在 n 阶可逆矩阵 P 使得

$$P^{-1}AP = J = \begin{pmatrix} J_1 & & \\ & J_2 & \\ & & \ddots \\ & & & J_s \end{pmatrix} \quad (3.5.4)$$

其中 J_i 为形如(3.5.1)的 Jordan 块. 记

$$P = (P_1, P_2, \dots, P_s) \quad (3.5.5)$$

其中 $P_i \in \mathbb{C}^{n \times n_i}$, 由(3.5.4)和(3.5.5), 得

$$(AP_1, AP_2, \dots, AP_s) = (P_1 J_1, P_2 J_2, \dots, P_s J_s)$$

比较上式两边得

$$AP_i = P_i J_i, i = 1, \dots, s \quad (3.5.6)$$

记 $P_i = (p_1^{(i)}, p_2^{(i)}, \dots, p_{n_i}^{(i)})$, 由(3.5.6)可得

$$\begin{cases} Ap_1^{(i)} = \lambda_i p_1^{(i)}, \\ Ap_2^{(i)} = \lambda_i p_2^{(i)} + p_1^{(i)}, \\ \dots\dots\dots \\ Ap_{n_i}^{(i)} = \lambda_i p_{n_i}^{(i)} + p_{n_i-1}^{(i)}, \end{cases} \quad i = 1, \dots, s \quad (3.5.7)$$

由上式可见, $p_1^{(i)}$ 是矩阵 A 对应于特征值 λ_i 的特征向量, 且由 $p_1^{(i)}$ 可依次求得 $p_2^{(i)}, \dots, p_{n_i}^{(i)}$. 由例 3.5.2 可知, 特征向量 $p_1^{(i)}$ 的选取应保证 $p_2^{(i)}$ 可以求出, 类似地 $p_2^{(i)}$ 的选取 (因为 $p_2^{(i)}$ 的选取一般不惟一, 只要适当选取一个即可) 也应保证 $p_3^{(i)}$ 可以求出, 依次类推, 并且使 $p_1^{(i)}, p_2^{(i)}, \dots, p_{n_i}^{(i)}$ 线性无关.

§ 3.6 Cayley-Hamilton 定理与最小多项式

设 A 为任意 n 阶矩阵, 其特征多项式为

$$f(\lambda) = \det(\lambda I - A) = \lambda^n + a_1 \lambda^{n-1} + a_2 \lambda^{n-2} + \dots + a_{n-1} \lambda + a_n$$

矩阵 A 与其特征多项式之间有如下重要关系.

定理 3.6.1 (Cayley-Hamilton 定理) 设 A 是 n 阶矩阵, $f(\lambda)$ 是 A 的特征多项式, 则 $f(A) = 0$.

证明 考虑特征矩阵 $\lambda I - A$ 的伴随矩阵 $(\lambda I - A)^*$, 其元素至多是 λ 的 $n-1$ 次多项式, 则 $(\lambda I - A)^*$ 可表示为

$$(\lambda I - A)^* = C_1 \lambda^{n-1} + C_2 \lambda^{n-2} + \dots + C_{n-1} \lambda + C_n$$

其中 C_1, C_2, \dots, C_n 都是 n 阶数字矩阵.

因为 $(\lambda I - A)(\lambda I - A)^* = f(\lambda)I$, 即

$$\begin{aligned} & (\lambda I - A)(C_1 \lambda^{n-1} + C_2 \lambda^{n-2} + \dots + C_{n-1} \lambda + C_n) \\ &= I \lambda^n + a_1 I \lambda^{n-1} + \dots + a_{n-1} I \lambda + a_n I \end{aligned}$$

比较两边 λ 的同次幂的系数矩阵, 得

$$C_1 = I$$

$$C_2 - AC_1 = a_1 I$$

$$C_3 - AC_2 = a_2 I$$

$$\dots$$

$$C_n - AC_{n-1} = a_{n-1} I$$

$$-AC_n = a_n I$$

用 $A^n, A^{n-1}, \dots, A, I$ 分别左乘上面各式, 再两边相加, 得

$$\begin{aligned} & A^n C_1 + A^{n-1}(C_2 - AC_1) + A^{n-2}(C_3 - AC_2) + \dots + A(C_n - AC_{n-1}) - AC_n \\ &= A^n + a_1 A^{n-1} + \dots + a_{n-1} A + a_n I = f(A) \end{aligned}$$

因为上式左边为零矩阵, 所以 $f(A) = 0$. \square

定义 3.6.1 设 A 为 n 阶矩阵, 如果存在多项式 $\varphi(\lambda)$ 使得 $\varphi(A) = 0$, 则称 $\varphi(\lambda)$ 为 A 的化零多项式.

对任意 n 阶矩阵 A , $f(\lambda)$ 是 A 的特征多项式, 由定理 3.6.1 知 $f(\lambda)$ 为 A 的化零多项式. 如果 $g(\lambda)$ 是任意多项式, 则 $g(\lambda)f(\lambda)$ 也是 A 的化零多项式. 因此, 任意 n 阶矩阵 A 的化零多项式总存在, 并且 A 的化零多项式有无穷多个.

定义 3.6.2 n 阶矩阵 A 的所有化零多项式中, 次数最低且首项系数为 1 的多项式称为 A 的最小多项式.

由定理 3.6.1 知, 任意 n 阶矩阵 A 的最小多项式存在且次数不会超过 n .

定理 3.6.2 设 A 是 n 阶矩阵, 则

(1) A 的最小多项式 $m(\lambda)$ 能整除 A 的任一化零多项式 $\varphi(\lambda)$, 特别地, $m(\lambda)$ 能整除 A 的特征多项式 $f(\lambda)$;

(2) A 的最小多项式 $m(\lambda)$ 的零点是 A 的特征值; 反之, A 的特征值是 $m(\lambda)$ 的零点;

(3) A 的最小多项式是惟一的.

证明 (1) 设 $m(\lambda)$ 是 A 的最小多项式, $\varphi(\lambda)$ 是 A 的任一化零多项式, 由定理 3.1.1 有

$$\varphi(\lambda) = q(\lambda)m(\lambda) + r(\lambda)$$

其中 $q(\lambda), r(\lambda)$ 是多项式, 并且 $r(\lambda) = 0$ 或者 $r(\lambda) \neq 0$ 但 $\partial(r(\lambda)) < \partial(m(\lambda))$. 因此 $r(\lambda) = 0$; 否则与 $m(\lambda)$ 是 A 的最小多项式矛盾. 于是 $m(\lambda) \mid \varphi(\lambda)$.

(2) 设 $f(\lambda)$ 是 A 的特征多项式, 由 (1) 知 $f(\lambda) = q(\lambda)m(\lambda)$, 其中 $q(\lambda)$ 是一个多项式. 因此 $m(\lambda) = 0$ 的根必为 $f(\lambda) = 0$ 的根, 即 A 的特征值.

反过来, 设 λ_0 是 A 的任一特征值, 相应的特征向量为 $\xi \neq 0$, 即

$$A\xi = \lambda_0 \xi$$

则

$$m(A)\xi = m(\lambda_0)\xi$$

因为 $m(A)=0, \xi \neq 0$, 所以 $m(\lambda_0)=0$, 即 λ_0 是 $m(\lambda)=0$ 的根.

(3) 设 A 有两个最小多项式 $m_1(\lambda), m_2(\lambda)$, 则它们的次数相同. 如果 $m_1(\lambda) \neq m_2(\lambda)$, 则 $m(\lambda) = m_1(\lambda) - m_2(\lambda) \neq 0$ 且 $\partial(m(\lambda)) < \partial(m_1(\lambda))$.

设 $m(\lambda)$ 的首项系数为 a , 则 $m_3(\lambda) = \frac{m(\lambda)}{a}$ 是首项系数为 1 的多项式且 $\partial(m_3(\lambda)) < \partial(m_1(\lambda))$. 由于

$$m_3(A) = \frac{1}{a}m(A) = \frac{1}{a}(m_1(A) - m_2(A)) = 0$$

于是, $m_3(\lambda)$ 是 A 的化零多项式. 这与 $m_1(\lambda), m_2(\lambda)$ 是 A 的最小多项式的假设矛盾. 因此 A 的最小多项式是惟一的. \square

定理 3.6.3 相似的矩阵具有相同的最小多项式.

证明 设 n 阶矩阵 A 与 B 相似, 则存在非奇异矩阵 P 使得

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

对任意多项式 $g(\lambda)$ 恒有

$$g(B) = P^{-1}g(A)P$$

可见, A 与 B 有相同的化零多项式, 从而它们具有相同的最小多项式. \square

例 3.6.1 求 Jordan 块

$$J_i = \begin{pmatrix} \lambda_i & 1 & & 0 \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{pmatrix}_{n_i \times n_i}$$

的最小多项式.

解 因为 J_i 的特征多项式 $f(\lambda) = (\lambda - \lambda_i)^{n_i}$, 则由定理 3.6.2 知 J_i 的最小多项式 $m(\lambda)$ 具有如下形式

$$m(\lambda) = (\lambda - \lambda_i)^k$$

其中正整数 $k \leq n_i$, 但当 $k < n_i$ 时

$$m(J_i) = (J_i - \lambda_i I)^k = \begin{bmatrix} 0 & 1 & & 0 \\ & & \ddots & \\ & & & 1 \\ 0 & & & 0 \end{bmatrix}_{n_i \times n_i}^k \neq 0$$

因此 $m(\lambda) = (\lambda - \lambda_i)^{n_i}$.

定理 3.6.4 块对角矩阵 $A = \text{diag}(A_1, \dots, A_s)$ 的最小多项式等于其诸对角块的最小多项式的最小公倍式.

证明 设 A_i 的最小多项式为 $m_i(\lambda) (i = 1, \dots, s)$. 由于对任意多项式 $\varphi(\lambda)$

$$\varphi(A) = \text{diag}(\varphi(A_1), \dots, \varphi(A_s))$$

如果 $\varphi(\lambda)$ 为 A 的化零多项式, 则 $\varphi(\lambda)$ 必为 $A_i (i = 1, \dots, s)$ 的化零多项式, 从而 $m_i(\lambda) | \varphi(\lambda) (i = 1, \dots, s)$. 因此 $\varphi(\lambda)$ 为 $m_1(\lambda), \dots, m_s(\lambda)$ 的公倍式.

反过来, 如果 $\varphi(\lambda)$ 为 $m_1(\lambda), \dots, m_s(\lambda)$ 的任一公倍式, 则 $\varphi(A_i) = 0 (i = 1, \dots, s)$, 从而 $\varphi(A) = 0$. 因此, A 的最小多项式为 $m_1(\lambda), \dots, m_s(\lambda)$ 的公倍式中次数最低者, 即它们的最小公倍式. \square

定理 3.6.5 设 $A \in \mathbb{C}^{n \times n}$, 则 A 的最小多项式为 A 的第 n 个不变因子 $d_n(\lambda)$.

证明 由定理 3.5.1 知 A 相似于 Jordan 标准形 $J = \text{diag}(J_1, \dots, J_s)$, 其中 J_i 为形如(3.5.1)的 Jordan 块. 由定理 3.4.2 和定理 3.6.3 知 A 与 J 有相同的不变因子和最小多项式. 而由定理 3.6.4 知 J 的最小多项式为 J_1, \dots, J_s 的最小多项式的最小公倍式. 因为 J_i 的最小多项式为 $(\lambda - \lambda_i)^{n_i} (i = 1, \dots, s)$, 而 $(\lambda - \lambda_1)^{n_1}, (\lambda - \lambda_2)^{n_2}, \dots, (\lambda - \lambda_s)^{n_s}$ 的最小公倍式是 J 的第 n 个不变因子 $d_n(\lambda)$. 因此 A 的最小多项式就是 A 的第 n 个不变因子 $d_n(\lambda)$. \square

由定理 3.5.3 和定理 3.6.5 可得如下定理.

定理 3.6.6 n 阶矩阵 A 相似于对角矩阵的充分必要条件是 A 的最小多项式 $m(\lambda)$ 没有重零点.

例 3.6.2 如果 n 阶矩阵 A 满足 $A^2 = A$, 则称矩阵 A 为幂等矩阵. 证明: 幂等矩阵 A 一定相似于对角矩阵.

证明 记 $\varphi(\lambda) = \lambda^2 - \lambda$, 则 $\varphi(\lambda)$ 是 A 的化零多项式. 由定理 3.6.2 知 A 的最小多项式 $m(\lambda)$ 整除 $\varphi(\lambda)$. 因为 $\varphi(\lambda) = 0$ 没有重根, 所以 $m(\lambda) = 0$ 也没有重根. 据定理 3.6.6 知 A 相似于对角矩阵. \square

习 题

1. 求 $g(\lambda)$ 除 $f(\lambda)$ 的商 $q(\lambda)$ 与余式 $r(\lambda)$:

(1) $f(\lambda) = \lambda^3 - 3\lambda^2 - \lambda - 1$, $g(\lambda) = 3\lambda^2 - 2\lambda + 1$;

(2) $f(\lambda) = 2\lambda^5 - 5\lambda^3 - 8\lambda$, $g(\lambda) = \lambda + 3$.

2. 化下列 λ 矩阵为 Smith 标准形:

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

(3) $\begin{pmatrix} \lambda + 1 & \lambda^2 + 1 & \lambda^2 \\ 3\lambda - 1 & 3\lambda^2 - 1 & \lambda^2 + 2\lambda \\ \lambda - 1 & \lambda^2 - 1 & \lambda \end{pmatrix}$; (4) $\begin{pmatrix} \lambda(\lambda + 1) & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & (\lambda + 1)^2 \end{pmatrix}$;

(5) $\begin{pmatrix} 0 & 0 & 0 & \lambda^2 \\ 0 & 0 & \lambda^2 - \lambda & 0 \\ 0 & (\lambda - 1)^2 & 0 & 0 \\ \lambda^2 - \lambda & 0 & 0 & 0 \end{pmatrix}$; (6) $\begin{pmatrix} 2\lambda & 3 & 0 & 1 & \lambda \\ 4\lambda & 3\lambda + 6 & 0 & \lambda + 2 & 2\lambda \\ 0 & 6\lambda & \lambda & 2\lambda & 0 \\ \lambda - 1 & 0 & \lambda - 1 & 0 & 0 \\ 3\lambda - 3 & 1 - \lambda & 2\lambda - 2 & 0 & 0 \end{pmatrix}$.

3. 求下列 λ 矩阵的不变因子和初等因子:

(1) $\begin{pmatrix} \lambda - 3 & -1 & 0 \\ 0 & \lambda - 3 & -1 \\ 0 & 0 & \lambda - 3 \end{pmatrix}$; (2) $\begin{pmatrix} 3\lambda^2 + 2\lambda - 3 & 2\lambda - 1 & \lambda^2 + 2\lambda - 3 \\ 4\lambda^2 + 3\lambda - 5 & 3\lambda - 2 & \lambda^2 + 3\lambda - 4 \\ \lambda^2 + \lambda - 4 & \lambda - 2 & \lambda - 1 \end{pmatrix}$;

(3) $\begin{pmatrix} \lambda & -1 & 0 & 0 \\ 0 & \lambda & -1 & 0 \\ 0 & 0 & \lambda & -1 \\ 5 & 4 & 3 & \lambda + 2 \end{pmatrix}$; (4) $\begin{pmatrix} 0 & 0 & 1 & \lambda + 2 \\ 0 & 1 & \lambda + 2 & 0 \\ 1 & \lambda + 2 & 0 & 0 \\ \lambda + 2 & 0 & 0 & 0 \end{pmatrix}$.

4. 求 λ 矩阵

$$A(\lambda) = \begin{pmatrix} \lambda & 0 & \cdots & 0 & a_n \\ -1 & \lambda & \cdots & 0 & a_{n-1} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda & a_2 \\ 0 & 0 & \cdots & -1 & \lambda + a_1 \end{pmatrix}$$

的行列式因子和不变因子.

5. 判断 λ 矩阵

$$A(\lambda) = \begin{pmatrix} 3\lambda + 1 & \lambda & 4\lambda - 1 \\ 1 - \lambda^2 & \lambda - 1 & \lambda - \lambda^2 \\ \lambda^2 + \lambda + 2 & \lambda & \lambda^2 + 2\lambda \end{pmatrix}$$

与

$$B(\lambda) = \begin{pmatrix} \lambda + 1 & \lambda - 2 & \lambda^2 - 2\lambda \\ 2\lambda & 2\lambda - 3 & \lambda^2 - 2\lambda \\ -2 & 1 & 1 \end{pmatrix}$$

是否相抵.

6. 判断矩阵 A 与 B 是否相似:

$$(1) A = \begin{pmatrix} 3 & 2 & -5 \\ 2 & 6 & -10 \\ 1 & 2 & -3 \end{pmatrix}, \quad B = \begin{pmatrix} 6 & 20 & -34 \\ 6 & 32 & -51 \\ 4 & 20 & -32 \end{pmatrix};$$

$$(2) A = \begin{pmatrix} 6 & 6 & -15 \\ 1 & 5 & -5 \\ 1 & 2 & -2 \end{pmatrix}, \quad B = \begin{pmatrix} 37 & -20 & -4 \\ 34 & -17 & -4 \\ 119 & -70 & -11 \end{pmatrix}.$$

7. 设 A 为 n 阶矩阵, 证明 A 与 A^{-1} 相似.8. 设 $\varepsilon \neq 0$, 证明:

$$(1) n \text{ 阶矩阵 } A = \begin{pmatrix} a & 1 & & \\ & a & \ddots & \\ & & \ddots & 1 \\ & & & a \end{pmatrix} \text{ 与 } B = \begin{pmatrix} a & \varepsilon & & \\ & a & \ddots & \\ & & \ddots & \varepsilon \\ & & & a \end{pmatrix} \text{ 相似};$$

$$(2) n \text{ 阶矩阵 } A = \begin{pmatrix} a & 1 & & \\ & a & \ddots & \\ & & \ddots & 1 \\ & & & a \end{pmatrix} \text{ 与 } B = \begin{pmatrix} a & 1 & & \\ & a & \ddots & \\ & & \ddots & 1 \\ \varepsilon & & & a \end{pmatrix} \text{ 不相似}.$$

9. 求下列矩阵的 Jordan 标准形:

$$(1) \begin{pmatrix} 2 & 6 & -15 \\ 1 & 1 & -5 \\ 1 & 2 & -6 \end{pmatrix}; \quad (2) \begin{pmatrix} 4 & 6 & -15 \\ 1 & 3 & -5 \\ 1 & 2 & -4 \end{pmatrix};$$

$$(3) \begin{pmatrix} 4 & -5 & 2 \\ 5 & -7 & 3 \\ 6 & -9 & 4 \end{pmatrix}; \quad (4) \begin{pmatrix} 3 & -4 & 0 & 2 \\ 4 & -5 & -2 & 4 \\ 0 & 0 & 3 & -2 \\ 0 & 0 & 2 & -1 \end{pmatrix};$$

$$(5) \begin{pmatrix} 0 & 3 & 3 \\ -1 & 8 & 6 \\ 2 & -14 & -10 \end{pmatrix}; \quad (6) \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

10. 求下列矩阵 A 的 Jordan 标准形, 并求变换矩阵 P 使得 $P^{-1}AP = J$:

$$(1) A = \begin{pmatrix} 0 & -4 & 0 \\ 1 & -4 & 0 \\ 1 & -2 & -2 \end{pmatrix}; \quad (2) A = \begin{pmatrix} 1 & -3 & 4 \\ 4 & -7 & 8 \\ 6 & -7 & 7 \end{pmatrix}.$$

11. 设矩阵

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

试计算 $2A^8 - 3A^5 + A^4 + A^2 - 4I$.

12. 证明: 如果 $\lambda_1, \lambda_2, \dots, \lambda_n$ 是 n 阶矩阵 A 的 n 个特征值, $f(\lambda)$ 是任一多项式, 则 $f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n)$ 是 n 阶矩阵 $f(A)$ 的 n 个特征值.

13. 证明: 任意可逆矩阵 A 的逆矩阵 A^{-1} 可以表示为 A 的多项式.

14. 求下列矩阵的最小多项式:

$$(1) \begin{pmatrix} 3 & 1 & 1 \\ 0 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix}; \quad (2) \begin{pmatrix} 4 & -2 & 2 \\ -5 & 7 & -5 \\ -6 & 7 & -4 \end{pmatrix};$$

15. 证明: n 阶矩阵 A 的最小多项式的次数为 n 的充分必要条件是 A 的最小多项式就是 A 的特征多项式.

16. 设 A 是 n 阶幂等矩阵. 证明 A 相似于矩阵 $\text{diag}(I_r, 0)$, 其中 $r = \text{rank}(A)$.

17. 证明: 如果 n 阶矩阵 A 满足 $A^p = I$ (p 是某个正整数), 则 A 相似于对角矩阵.

18. 证明: 如果 n 阶矩阵 A 满足 $A^2 - 5A = -6I$, 则 A 相似于对角矩阵.

第四章 矩阵的因子分解

矩阵的因子分解就是将给定的矩阵分解成特殊类型矩阵的乘积. 本章介绍一些在矩阵理论和数值计算中具有广泛应用的矩阵因子分解.

§ 4.1 初等矩阵

4.1.1 初等矩阵

在线性代数课程中, 我们已经看到初等矩阵对矩阵求逆与线性方程组的研究起着重要的作用. 本节介绍更一般形式的初等矩阵, 它是矩阵理论和矩阵计算的基本工具.

定义 4.1.1 设 $u, v \in \mathbb{C}^n$, σ 为一复数, 如下形式的矩阵

$$E(u, v, \sigma) = I - \sigma uv^H \quad (4.1.1)$$

称为初等矩阵.

定理 4.1.1 初等矩阵 $E(u, v, \sigma)$ 具有如下性质:

- (1) $\det(E(u, v, \sigma)) = 1 - \sigma v^H u$;
- (2) 如果 $\sigma v^H u \neq 1$, 则 $E(u, v, \sigma)$ 可逆, 并且其逆矩阵也是初等矩阵

$$E(u, v, \sigma)^{-1} = E(u, v, \tau) \quad (4.1.2)$$

其中 $\tau = \frac{\sigma}{\sigma v^H u - 1}$.

- (3) 对任意非零向量 $a, b \in \mathbb{C}^n$, 可适当选取 u, v 和 σ 使得

$$E(u, v, \sigma)a = b \quad (4.1.3)$$

证明 (1) 如果 $v=0$, 则(1)显然成立; 如果 $v \neq 0$, 则令 $u_1 = \frac{v}{\|v\|}$, 并在 $\text{span}(v)^\perp$ 中取一组标准正交基 u_2, \dots, u_n . 记 $U = [u_1, u_2, \dots, u_n]$, 则 U 是酉矩阵, 且

$$U^H E(u, v, \sigma) U = \begin{bmatrix} 1 - \sigma v^H u & 0 & \cdots & 0 \\ -\sigma \|v\| u_2^H u & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\sigma \|v\| u_n^H u & 0 & \cdots & 1 \end{bmatrix}$$

由上式即得 $\det(E(u, v, \sigma)) = 1 - \sigma v^H u$.

(2) 由关系式

$$E(u, v, \sigma)E(u, v, \tau) = E(u, v, \sigma + \tau - \sigma\tau v^H u)$$

及(1)可知, 当且仅当 $\sigma v^H u \neq 1$ 时, $E(u, v, \sigma)$ 可逆, 并且当 $\sigma + \tau - \sigma\tau v^H u = 0$ 时, 即 $\tau = \frac{\sigma}{\sigma v^H u - 1}$ 时, $E(u, v, \sigma)^{-1} = E(u, v, \tau)$.

(3) 只需取 u, v 和 σ 满足 $v^H u \neq 0, \sigma u = \frac{a-b}{v^H u}$ 即可. \square

线性代数中所用的初等矩阵都可以用初等矩阵 $E(u, v, \sigma)$ 表示.

例 4.1.1 初等(交换)矩阵 $P(i, j)$, 即交换单位矩阵 I 的第 i, j 两行(或列)所得的矩阵. 令 $u = v = e_i - e_j$, 其中 $e_i = (\underbrace{0, \dots, 0}_i, 1, 0, \dots, 0)^T, \sigma = 1$, 则

$$P(i, j) = E(e_i - e_j, e_i - e_j, 1) = I - (e_i - e_j)(e_i - e_j)^T$$

由定理 4.1.1 知 $\det(P(i, j)) = -1 (i \neq j)$, 并且 $P(i, j)^{-1} = P(i, j)$.

例 4.1.2 初等矩阵 $P(i(k))$, 即由单位矩阵 I 的第 i 行(列)乘以非零数 k 所得的矩阵. 令 $u = v = e_i, \sigma = 1 - k$, 则

$$P(i(k)) = E(e_i, e_i, 1 - k) = I - (1 - k)e_i e_i^T$$

由定理 4.1.1 知 $\det(P(i(k))) = k$, 并且 $P(i(k))^{-1} = P(i(\frac{1}{k}))$.

例 4.1.3 初等矩阵 $P(i, j(k))$, 即把单位矩阵第 j 行的 k 倍加到第 i 行所得的矩阵. 令 $u = e_i, v = e_j, \sigma = -k$, 则

$$P(i, j(k)) = E(e_i, e_j, -k) = I + k e_i e_j^T$$

并且 $\det(P(i, j(k))) = 1, P(i, j(k))^{-1} = P(i, j(-k))$.

下面介绍两种不同的初等矩阵.

4.1.2 初等下三角矩阵

令 $u = l_i = (0, \dots, 0, l_{i, 1, i}, \dots, l_{i, m})^T, v = e_i, \sigma = 1$, 则

$$L_i = L_i(l_i) = E(l_i, e_i, 1)$$

称为初等下三角矩阵. 即

$$L_i = L_i(l_i) = I - l_i e_i^T = \begin{pmatrix} 1 & & & & 0 \\ & \ddots & & & \\ & & 1 & & \\ & & -l_{i+1,i} & 1 & \\ & & \vdots & & \ddots \\ 0 & & \vdots & 0 & \ddots \\ & & -l_m & & 1 \end{pmatrix} \quad (4.1.4)$$

由定理 4.1.1 知 $\det(L_i) = 1$, 并且 $L_i^{-1} = E(l_i, e_i, -1) = L_i(-l_i)$.

对初等下三角矩阵(4.1.4), 当 $i < j$ 时, 有

$$L_i(l_i)L_j(l_j) = I - l_i e_i^T - l_j e_j^T = \begin{pmatrix} 1 & & & & \\ & \ddots & & & 0 \\ & & 1 & & \\ & & -l_{i+1,i} & \ddots & \\ & & \vdots & & 1 \\ 0 & & \vdots & 0 & -l_{j+1,j} & \ddots \\ & & & \vdots & \\ & & -l_m & -l_n & 0 & 1 \end{pmatrix} \quad (4.1.5)$$

用初等下三角矩阵 L_i 左乘一个矩阵 A , 等于从 A 的第 k 行减去第 i 行乘以 l_k ($k = i + 1, \dots, n$). 对于 $A = (a_{ij})$, 如果 $a_{ij} \neq 0$, 取

$$l_k = \frac{a_{ki}}{a_{ij}}, \quad k = i + 1, \dots, n \quad (4.1.6)$$

则 $L_i A$ 的第 $(i + 1, j), \dots, (n, j)$ 元素全为零. 这就是消去法的一步.

初等下三角矩阵在矩阵的满秩分解、三角分解以及用消去法求解线性方程组中起着重要的作用.

4.1.3 Householder 矩阵

在(4.1.1)中取 $u = v = w$, $\sigma = 2$, 并且 w 是单位向量, 即 $\|w\| = 1$, 初

等矩阵

$$H(w) = E(w, w, 2) = I - 2ww^H \quad (4.1.7)$$

称为 Householder 矩阵或初等 Hermite 矩阵.

定理 4.1.2 Householder 矩阵 $H(w)$ 具有如下性质:

(1) $\det(H(w)) = -1$;

(2) $H(w)^H = H(w) = H(w)^{-1}$;

(3) 设 $a, b \in \mathbb{C}^n$ 且 $a \neq b$, 则存在单位向量 w 使得 $H(w)a = b$ 的充分必要条件是

$$a^H a = b^H b, \quad a^H b = b^H a \quad (4.1.8)$$

并且若上述条件成立, 则使 $H(w)a = b$ 成立的单位向量 w 可取为

$$w = e^{i\theta}(a - b) / \|a - b\| \quad (4.1.9)$$

其中 θ 为任一实数.

证明 (1), (2) 可直接从初等矩阵的性质得出. 性质 (3) 的必要性是显然的, 以下证明其充分性.

因为 $a \neq b$, 则取 $w = e^{i\theta}(a - b) / \|a - b\|$, 有

$$H(w)a = (I - 2 \frac{(a - b)(a - b)^H}{\|a - b\|^2})a = a - \frac{2(a^H a - b^H a)(a - b)}{a^H a + b^H b - a^H b - b^H a}$$

由条件 (4.1.8) 即得 $H(w)a = b$. \square

对于 $a = (a_1, \dots, a_n)^T \neq 0$, 若令

$$\sigma = \begin{cases} \|a\|, & a_1 = 0 \\ -e^{i\arg a_1} \|a\|, & a_1 \neq 0 \end{cases} \quad (4.1.10)$$

并取

$$w = (a - \sigma e_1) / \|a - \sigma e_1\| \quad (4.1.11)$$

则有

$$H(w)a = \sigma e_1 \quad (4.1.12)$$

Householder 矩阵在矩阵的 QR 分解和矩阵计算中具有重要的应用.

§ 4.2 满秩分解

设 A 为 $m \times n$ 矩阵, 且 $\text{rank}(A) = r$. 由定理 2.1.10 知, 存在 m 阶可逆矩阵 P 和 n 阶可逆矩阵 Q 使得

$$A = P \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} Q \quad (4.2.1)$$

定理 4.2.1(满秩分解定理) 设 $m \times n$ 矩阵 A 的秩为 $r > 0$, 则存在 $m \times r$ 矩阵 B 和 $r \times n$ 矩阵 C 使得

$$A = BC \quad (4.2.2)$$

并且 $\text{rank}(B) = \text{rank}(C) = r$.

证明 方法 1. 因为

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} I_r \\ 0 \end{bmatrix} [I_r \quad 0]$$

故(4.2.1)可以改写成

$$A = P \begin{bmatrix} I_r \\ 0 \end{bmatrix} [I_r \quad 0] Q$$

令 $P \begin{bmatrix} I_r \\ 0 \end{bmatrix} = B, [I_r \quad 0] Q = C$, 使得

$$A = BC$$

其中 B 是 $m \times r$ 矩阵, 它的 r 个列是非奇异矩阵 P 的前 r 列, 因而线性无关, $\text{rank}(B) = r$. C 是 $r \times n$ 矩阵, 它的 r 个行也是线性无关的, 故 $\text{rank}(C) = r$.

方法 2. 利用初等下三角矩阵, 可以构造性地证明满秩分解(4.2.2)的存在性, 证明的过程同时给出了进行满秩分解的方法.

设 $A^{(0)} = A$ 的第 1 个非零列为第 j_1 列, 第 j_1 列的第 1 个非零元素在第 i_1 行, 则矩阵 $P(1, i_1)A^{(0)}$ 中 $(1, j_1)$ 位置上元素非零. 于是左乘以适当的初等下三角矩阵 L_1 , 可使

$$A^{(1)} = L_1 P(1, i_1) A^{(0)}$$

的第 j_1 列第 1 个元素以下全部为零, 同时 $A^{(1)}$ 的前 $j_1 - 1$ 列全部为零. 然后在 $A^{(1)}$ 中找出第 1 个在第 1 行以下有非零元素的列, 假设它是第 j_2 列, 并且在 (i_2, j_2) 位置上元素是第 j_2 列中第 1 行以下第 1 个非零元素, 则 $P(2, i_2)A^{(1)}$ 中 $(2, j_2)$ 位置上元素非零. 于是左乘以适当的初等下三角矩阵 L_2 , 可使

$$A^{(2)} = L_2 P(2, i_2) A^{(1)}$$

的第 j_2 列第 2 个元素以下全部为零, 并且 $A^{(2)}$ 的前 $j_2 - 1$ 列保持与 $A^{(1)}$ 相

同. 依此类推, 直到某一个 $A^{(r)}$, 它的第 r 行以下各行全部为零为止. 这时

$$A^{(r)} = L_r P(r, i_r) L_{r-1} P(r-1, i_{r-1}) \cdots L_1 P(1, i_1) A^{(0)}$$

记

$$A^{(r)} = \begin{bmatrix} C \\ 0 \end{bmatrix}, (L_r P(r, i_r) \cdots L_1 P(1, i_1))^{-1} = (B, B')$$

其中 B 是 $m \times r$ 矩阵, C 是 $r \times n$ 矩阵, 并且 $\text{rank}(B) = r, \text{rank}(C) = r, A = BC$. \square

矩阵 A 的分解式(4.2.2)中, B 称为列满秩矩阵, C 称为行满秩矩阵, (4.2.2)称为 A 的满秩分解.

注意 矩阵 A 的满秩分解(4.2.2)一般是不惟一的. 事实上, 对任一 r 阶非奇异矩阵 D , 若令 $B_1 = BD, C_1 = D^{-1}C$, 则显然有 $A = B_1 C_1$.

例 4.2.1 求矩阵

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

的一个满秩分解.

解 由(4.1.6)可取

$$L_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\frac{a_{21}}{a_{11}} & 1 & 0 & 0 \\ -\frac{a_{31}}{a_{11}} & 0 & 1 & 0 \\ -\frac{a_{41}}{a_{11}} & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ -2 & 0 & 0 & 1 \end{bmatrix}$$

则

$$L_1 A = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 0 & 2 & 0 & 3 \\ 0 & 2 & 0 & 3 \\ 0 & -4 & 0 & -6 \end{bmatrix}$$

取

$$L_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 2 & 0 & 1 \end{pmatrix}$$

则

$$L_2(L_1 A) = \begin{pmatrix} -1 & 0 & 1 & 2 \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

于是

$$\begin{aligned} A &= L_1^{-1} L_2^{-1} \begin{pmatrix} -1 & 0 & 1 & 2 \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & -2 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 1 & 2 \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -2 & 1 & 1 & 0 \\ 2 & -2 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 1 & 2 \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

$$\text{令 } B = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ -2 & 1 \\ 2 & -2 \end{pmatrix}, C = \begin{pmatrix} -1 & 0 & 1 & 2 \\ 0 & 2 & 0 & 3 \end{pmatrix}, \text{ 则 } A = BC.$$

例 4.2.2 求矩阵

$$A = \begin{pmatrix} 0 & 2 & 1 & -1 \\ 1 & 2 & 3 & 0 \\ -2 & -2 & -5 & -1 \end{pmatrix}$$

的一个满秩分解

解 取

$$P(1,2) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

则

$$P(1,2)A = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 1 & -1 \\ -2 & -2 & -5 & -1 \end{pmatrix}$$

令

$$L_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}$$

则

$$L_1 P(1,2)A = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 1 & -1 \\ 0 & 2 & 1 & -1 \end{pmatrix} = A^{(1)}$$

取

$$L_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$$

则

$$L_2 L_1 P(1,2)A = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} = A^{(2)}$$

因此

$$A = P(1,2)L_1^{-1}L_2^{-1}A^{(2)} = P(1,2) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$= P(1,2) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{令 } B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ -2 & 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 1 & -1 \end{bmatrix}, \text{ 则 } A = BC.$$

§ 4.3 三角分解

设 $A = (a_{ij})$ 是 n 阶矩阵, 如果 A 的对角线下(上)方的元素全为零, 即对 $i > j, a_{ij} = 0$ (对 $i < j, a_{ij} = 0$), 则称矩阵 A 为上(下)三角矩阵. 上三角矩阵和下三角矩阵统称为三角矩阵. 对角元全为 1 的上(下)三角矩阵称为单位上(下)三角矩阵.

A, B 是两个 n 阶上(下)三角矩阵, 容易验证: $A + B, AB$ 仍是上(下)三角矩阵, 并且 A 可逆的充分必要条件是 A 的对角元均非零. 当 A 可逆时, 其逆矩阵也是上(下)三角矩阵. 特别地, 两个单位上(下)三角矩阵的乘积仍是单位上(下)三角矩阵, 并且单位上(下)三角矩阵的逆矩阵也是单位上(下)三角矩阵.

设 A 是 n 阶矩阵, 如果有下三角矩阵 L 和上三角矩阵 U 使得 $A = LU$, 则称 A 能作三角分解, 并且称 $A = LU$ 为 A 的三角分解或 LU 分解.

定理 4.3.1 (LU 分解定理) 设 A 是 n 阶非奇异矩阵, 则存在唯一的单位下三角矩阵 L 和上三角矩阵 U 使得

$$A = LU \quad (4.3.1)$$

的充分必要条件是 A 的所有顺序主子式均非零, 即

$$\Delta_k = A \begin{bmatrix} 1 \cdots k \\ 1 \cdots k \end{bmatrix} \neq 0, k = 1, \cdots, n-1 \quad (4.3.2)$$

证明 必要性. 如果存在单位下三角矩阵 L 和上三角矩阵 U 使得 $A = LU$, 记

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & u_{nn} \end{bmatrix}$$

则 $|A| = |LU| = |U| = u_{11} u_{22} \cdots u_{nn}$. 因为 A 非奇异, 所以 $u_{ii} \neq 0$. 将 $A = LU$ 分块写成

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix}$$

其中 A_{11}, L_{11}, U_{11} 分别为 A, L, U 的 k 阶顺序主子矩阵, 于是

$$A_{11} = L_{11} U_{11}$$

从而 $|A_{11}| = \Delta_k = |U_{11}| = u_{11} \cdots u_{kk} \neq 0 (k = 1, 2, \cdots, n)$, 并且

$$u_{11} = a_{11}, \quad u_{kk} = \frac{\Delta_k}{\Delta_{k-1}}, k = 2, \cdots, n \quad (4.3.3)$$

充分性. 对矩阵的阶数作归纳法证明分解式(4.3.1)存在. 当矩阵的阶为 1 时结论显然成立. 设对 $n-1$ 阶矩阵有分解式(4.3.1). 对 n 阶矩阵 A , 记

$$A = \begin{bmatrix} A_{n-1} & \beta \\ \alpha & a_{nn} \end{bmatrix}$$

其中 A_{n-1} 为 A 的 $n-1$ 阶顺序主子矩阵. 根据定理的条件, A_{n-1} 是非奇异矩阵, 则有

$$\begin{bmatrix} I_{n-1} & 0 \\ -\alpha A_{n-1}^{-1} & 1 \end{bmatrix} A = \begin{bmatrix} I_{n-1} & 0 \\ -\alpha A_{n-1}^{-1} & 1 \end{bmatrix} \begin{bmatrix} A_{n-1} & \beta \\ \alpha & a_{nn} \end{bmatrix} = \begin{bmatrix} A_{n-1} & \beta \\ 0 & a_{nn} - \alpha A_{n-1}^{-1} \beta \end{bmatrix}$$

从而

$$A = \begin{bmatrix} I_{n-1} & 0 \\ \alpha A_{n-1}^{-1} & 1 \end{bmatrix} \begin{bmatrix} A_{n-1} & \beta \\ 0 & a_{nn} - \alpha A_{n-1}^{-1} \beta \end{bmatrix} \quad (4.3.4)$$

由归纳假设, 存在 $n-1$ 阶单位下三角矩阵 L_{n-1} 和上三角矩阵 U_{n-1} 使得 $A_{n-1} = L_{n-1} U_{n-1}$. 于是可得

$$\begin{aligned} A &= \begin{bmatrix} I_{n-1} & 0 \\ \alpha A_{n-1}^{-1} & 1 \end{bmatrix} \begin{bmatrix} L_{n-1} U_{n-1} & \beta \\ 0 & a_{nn} - \alpha A_{n-1}^{-1} \beta \end{bmatrix} \\ &= \begin{bmatrix} I_{n-1} & 0 \\ \alpha A_{n-1}^{-1} & 1 \end{bmatrix} \begin{bmatrix} L_{n-1} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} U_{n-1} & L_{n-1}^{-1} \beta \\ 0 & a_{nn} - \alpha A_{n-1}^{-1} \beta \end{bmatrix} \end{aligned}$$

令

$$L = \begin{pmatrix} I_{n-1} & 0 \\ \alpha A_{n-1}^{-1} & 1 \end{pmatrix} \begin{pmatrix} L_{n-1} & 0 \\ 0 & 1 \end{pmatrix}, U = \begin{pmatrix} U_{n-1} & L_{n-1}^{-1} \beta \\ 0 & a_{nn} - \alpha A_{n-1}^{-1} \beta \end{pmatrix}$$

即得 $A = LU$, 其中 L 是单位下三角矩阵, U 是上三角矩阵. 因此矩阵的阶为 n 时分解式(4.3.1)也存在.

下面证明惟一性. 如果

$$A = LU = \tilde{L} \tilde{U}$$

其中 L, \tilde{L} 为 n 阶单位下三角矩阵, U, \tilde{U} 为 n 阶可逆上三角矩阵, 则

$$\tilde{L}^{-1} L = \tilde{U} U^{-1}$$

上式左边的矩阵是单位下三角矩阵, 而右边的矩阵是上三角矩阵. 因此 $\tilde{L}^{-1} L = \tilde{U} U^{-1} = I$. 于是 $L = \tilde{L}, U = \tilde{U}$. 这就证明了惟一性. \square

因为非奇异上三角矩阵

$$= \begin{pmatrix} u_{11} & & & & \\ & u_{22} & & & \\ & & \ddots & & \\ & & & u_{n-1,n-1} & \\ & & & & u_{nn} \end{pmatrix} \begin{pmatrix} 1 & \frac{u_{12}}{u_{11}} & \frac{u_{13}}{u_{11}} & \cdots & \frac{u_{1n}}{u_{11}} \\ & 1 & \frac{u_{23}}{u_{22}} & \cdots & \frac{u_{2n}}{u_{22}} \\ & & \ddots & \ddots & \vdots \\ & & & 1 & \frac{u_{n-1,n}}{u_{n-1,n-1}} \\ & & & & 1 \end{pmatrix} \quad (4.3.5)$$

由定理 4.3.1 和(4.3.3), (4.3.5)容易得到如下定理.

定理 4.3.2 (LDU 分解定理) 设 A 是 n 阶非奇异矩阵, 则存在惟一的单位下三角矩阵 L , 对角矩阵 $D = \text{diag}(d_1, d_2, \cdots, d_n)$ 和单位上三角矩阵 U 使

得

$$A = LDU \quad (4.3.6)$$

的充分必要条件是 A 的所有顺序主子式均非零, 即 $\Delta_k \neq 0 (k = 1, \dots, n-1)$, 并且

$$d_1 = a_{11}, d_k = \frac{\Delta_k}{\Delta_{k-1}}, k = 2, \dots, n \quad (4.3.7)$$

分解式(4.3.6)称为矩阵 A 的 LDU 分解.

关于矩阵的 LU 分解或 LDU 分解的实际计算, 仿照例 4.2.1 即可实现.

由以上两个定理知, 即使矩阵 A 非奇异, A 未必能作 LU 分解和 LDU 分

解. 例如, 矩阵 $A = \begin{bmatrix} 0 & 4 & 6 \\ 0 & -3 & -5 \\ 1 & -3 & -6 \end{bmatrix}$ 非奇异, 但 A 不能作 LU 分解和 LDU 分

解. 由定理 4.2.1 的证明方法 2 可见, 适当改变非奇异矩阵 A 的行的次序, 使改变后的矩阵可以作 LU 分解. 为此, 我们先介绍排列矩阵的概念.

定义 4.3.1 设 e_i 是 n 阶单位矩阵的第 i 列 ($i = 1, 2, \dots, n$), 以 $e_{i_1}, e_{i_2}, \dots, e_{i_n}$ 为列作成的矩阵 $[e_{i_1}, e_{i_2}, \dots, e_{i_n}]$ 称为 n 阶排列矩阵, 其中 i_1, i_2, \dots, i_n 是 $1, 2, \dots, n$ 的一个排列.

显然, 初等交换矩阵 $P(i, j)$ 是特殊的排列矩阵, 并且初等交换矩阵的乘积是排列矩阵.

容易证明: 排列矩阵的转置仍为排列矩阵, 并且排列矩阵是正交矩阵, 排列矩阵的逆是排列矩阵.

以排列矩阵 $[e_{i_1}, e_{i_2}, \dots, e_{i_n}]^T$ 左乘 n 阶矩阵 A , 就是将 A 的行按照 i_1, \dots, i_n 的次序重排; 以排列矩阵 $[e_{i_1}, e_{i_2}, \dots, e_{i_n}]$ 右乘矩阵 A , 就是将 A 的列按照 i_1, \dots, i_n 的次序重排.

定理 4.3.3 设 A 是 n 阶非奇异矩阵, 则存在排列矩阵 P 使得

$$PA = \tilde{L}\tilde{U} = LDU \quad (4.3.8)$$

其中 L 是单位下三角矩阵, \tilde{U} 是上三角矩阵, U 是单位上三角矩阵, D 是对角矩阵.

证明 因为 A 非奇异, 所以 $\text{rank}(A) = r = n$. 由定理 4.2.1 的证法 2, $A^{(n-1)} = \tilde{U}$ 是上三角矩阵, 则

$$A^{(n-1)} = \tilde{U} = L_{n-1}P(n-1, i_{n-1}) \cdots L_1P(1, i_1)A$$

其中 $L_i (i=1, \dots, n-1)$ 是初等下三角矩阵, $P(j, i_j) (j=1, \dots, n-1)$ 是初等交换矩阵. 因为 $P(i, j)^{-1} = P(i, j)$, 所以

$$A = P(1, i_1)L_1^{-1} \cdots P(n-1, i_{n-1})L_{n-1}^{-1}\tilde{U}$$

令 $\tilde{L} = P(1, i_1)L_1^{-1} \cdots P(n-1, i_{n-1})L_{n-1}^{-1}$, 注意到

$$\begin{aligned} & P(n-1, i_{n-1}) \cdots P(2, i_2)P(1, i_1)\tilde{L} = (P(n-1, i_{n-1}) \\ & \cdots P(2, i_2)L_1^{-1}P(2, i_2) \cdots P(n-1, i_{n-1})) \cdot \\ & (P(n-1, i_{n-1}) \cdots P(3, i_3)L_2^{-1}P(3, i_3) \cdots P(n-1, i_{n-1})) \\ & \cdots (P(n-1, i_{n-1})L_{n-2}^{-1}P(n-1, i_{n-1}))L_{n-1}^{-1} \\ & = \tilde{L}_1\tilde{L}_2 \cdots \tilde{L}_{n-2}\tilde{L}_{n-1} \end{aligned}$$

其中 $\tilde{L}_{n-1} = L_{n-1}^{-1}$,

$$\begin{aligned} \tilde{L}_j &= P(n-1, i_{n-1}) \cdots P(j+1, i_{j+1})L_j^{-1}P(j+1, i_{j+1}) \cdots P(n-1, i_{n-1}), \\ j &= 1, \dots, n-2 \end{aligned}$$

因为初等下三角矩阵的逆仍为初等下三角矩阵, 则 $\tilde{L}_1, \tilde{L}_2, \dots, \tilde{L}_{n-1}$ 都是单位下三角矩阵. 令

$$P = P(n-1, i_{n-1}) \cdots P(1, i_1), \quad L = \tilde{L}_1\tilde{L}_2 \cdots \tilde{L}_{n-1}, \quad \tilde{U} = A^{(n-1)}$$

则 P 是排列矩阵, L 是单位下三角矩阵, \tilde{U} 是非奇异上三角矩阵, 并且

$$PA = L\tilde{U}$$

由上式及(4.3.5)即得(4.3.8)的第二个等式. \square

下面介绍矩阵的 LU 分解在求解线性方程组中的应用.

设 A 是 n 阶非奇异矩阵, b 是 n 维向量, 对线性方程组

$$Ax = b \quad (4.3.9)$$

如果 A 的顺序主子式都不等于零, 由定理 4.3.1 知 A 有三角分解 $A = LU$, 其中 L 是单位下三角矩阵, U 是上三角矩阵. 则方程组(4.3.9)等价于如下方程组

$$\begin{cases} Ly = b \\ Ux = y \end{cases} \quad (4.3.10)$$

于是先从(4.3.10)的第一组方程解出 y , 然后将 y 代入第二组方程再求 x . 这就是求解线性方程组的直接三角分解法.

如果 A 的顺序主子式中有等于零, 由定理 4.3.3, 可考虑如下方程组

$$PAx = Pb \quad (4.3.11)$$

其中 P 是适当的排列矩阵. 对此方程组应用直接三角分解法.

§ 4.4 QR 分解

本节讨论矩阵 A 的另一类分解——正交三角分解, 简称 QR 分解, 即将矩阵 A 分解为正交(酉)矩阵 Q 与上三角矩阵 R 的乘积 $A = QR$. 先讨论非奇异矩阵的 QR 分解, 然后推广到一般矩阵的情形.

定理 4.4.1 设 A 是 n 阶非奇异实(复)矩阵, 则存在正交(酉)矩阵 Q 和非奇异实(复)上三角矩阵 R 使得

$$A = QR \quad (4.4.1)$$

且除去相差一个对角元绝对值(模)全等于 1 的对角矩阵因子外分解式(4.4.1)是惟一的.

证明 记 $A = [\alpha_1, \alpha_2, \dots, \alpha_n]$, α_i 是矩阵 A 的第 i 列向量. 因为矩阵 A 非奇异, 所以向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关. 由定理 1.6.6, 应用 Gram-Schmidt 正交化方法将线性无关向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 化为标准正交向量组 q_1, q_2, \dots, q_n , 则可得

$$\begin{cases} \alpha_1 = r_{11}q_1 \\ \alpha_2 = r_{12}q_1 + r_{22}q_2 \\ \alpha_3 = r_{13}q_1 + r_{23}q_2 + r_{33}q_3 \\ \dots\dots\dots \\ \alpha_n = r_{1n}q_1 + r_{2n}q_2 + \dots\dots\dots + r_{nn}q_n \end{cases} \quad (4.4.2)$$

令

$$Q = (q_1, q_2, \dots, q_n), R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \\ 0 & r_{22} & r_{23} & \cdots & r_{2n} \\ 0 & 0 & r_{33} & \cdots & r_{3n} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & r_{nn} \end{pmatrix}$$

则 Q 为正交(酉)矩阵, R 为非奇异实(复)上三角矩阵. 由(4.4.2)式, 有 $A = QR$. 这就证明了 QR 分解的存在性.

设矩阵 A 有两个 QR 分解

$$A = QR = Q_1 R_1$$

其中 Q, Q_1 为正交(酉)矩阵, R, R_1 为非奇异上三角矩阵, 则

$$Q = Q_1 R_1 R^{-1} = Q_1 D$$

其中 $D = R_1 R^{-1}$ 为非奇异上三角矩阵. 于是

$$I = Q^H Q = (Q_1 D)^H (Q_1 D) = D^H D$$

这说明 D 为酉矩阵. 比较等式 $D^H D = D D^H = I$ 的对角元, 可导出 D 为对角矩阵, 并且对角元的模全等于 1, 于是 $R_1 = D R, Q_1 = Q D^{-1}$. \square

注意 如果在非奇异矩阵 A 的 QR 分解中规定上三角矩阵 R 的各个对角元的符号(例如全为正数), 则 A 的 QR 分解是惟一的.

定理 4.4.1 的结论可以推广如下:

定理 4.4.2 设 A 是 $m \times n$ 实(复)矩阵, 且其 n 个列向量线性无关, 则存在 m 阶正交(酉)矩阵 Q 和 n 阶非奇异实(复)上三角矩阵 R 使得

$$QA = \begin{pmatrix} R \\ 0 \end{pmatrix} \quad (4.4.3)$$

证明 利用 Householder 矩阵给出定理的一个构造性证明.

令 $A = A^{(0)} = [\alpha_1, \alpha_2, \dots, \alpha_n]$, α_i 是矩阵 A 的第 i 列. 对 α_1 , 由定理 4.1.2 知存在 Householder 矩阵 H_1 使得 $H_1 \alpha_1 = k_1 e_1$, 这里 $|k_1|^2 = \|\alpha_1\|^2 > 0$, e_1 是 m 阶单位矩阵的第 1 列. 令

$$A^{(1)} \equiv H_1 A^{(0)} = [H_1 \alpha_1, H_1 \alpha_2, \dots, H_1 \alpha_n] = [k_1 e_1, \alpha_2^{(1)}, \dots, \alpha_n^{(1)}]$$

$\alpha_i^{(1)}$ 是矩阵 $A^{(1)}$ 的第 i 列 ($i = 2, \dots, n$). 设 $\tilde{\alpha}_2^{(1)}$ 是将 $\alpha_2^{(1)}$ 的第一个分量换为 0 所得的向量, 对 $\tilde{\alpha}_2^{(1)}$ 同样有 Householder 矩阵 H_2 使得 $H_2 \tilde{\alpha}_2^{(1)} = k_2 e_2$, 其中 $k_2 \neq 0$. 如此继续, 最终可得 $A^{(n)}$, 并且 $A^{(n)}$ 的第 n 行以下的元素全为零, 即

$$H_n \cdots H_2 H_1 A^{(0)} = A^{(n)} = \begin{pmatrix} R \\ 0 \end{pmatrix}$$

其中 R 为 n 阶非奇异实(复)上三角矩阵. 令 $Q = H_n \cdots H_2 H_1$, 则 Q 为 m 阶正交(酉)矩阵, 并且 $QA = \begin{pmatrix} R \\ 0 \end{pmatrix}$. \square

记 $Q^H = [Q_1, Q_2]$, 其中 Q_1 是 $m \times n$ 矩阵. 因为 Q^H 是酉矩阵, 由 $QQ^H = I$ 可得

$$Q_1^H Q_1 = I_n \quad (4.4.4)$$

我们将满足等式(4.4.4)的 $m \times n$ 矩阵 Q_1 称为列正交规范矩阵. 于是由(4.4.3)得

$$A = Q_1 R \quad (4.4.5)$$

其中 Q_1 是 $m \times n$ 列正交规范矩阵, R 是 n 阶非奇异实(复)上三角矩阵.

对一般的 $m \times n$ 矩阵, 有如下定理.

定理 4.4.3 设 A 是 $m \times n$ 矩阵, 且 $\text{rank}(A) = r > 0$, 则存在 m 阶正交(酉)矩阵 Q 和 $r \times n$ 行满秩矩阵 R 使得

$$QA = \begin{bmatrix} R \\ 0 \end{bmatrix} \quad (4.4.6)$$

或 A 有分解

$$A = Q_1 R \quad (4.4.7)$$

其中 Q_1 是 $m \times r$ 列正交规范矩阵, R 是 $r \times n$ 行满秩矩阵.

证明 作 A 的满秩分解

$$A = BC$$

其中 B, C 分别为 $m \times r, r \times n$ 矩阵, $\text{rank}(B) = \text{rank}(C) = r$. 由定理 4.4.2, 存在 m 阶正交(酉)矩阵 Q 和 r 阶非奇异实(复)上三角矩阵 R_1 使得

$$QB = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

令 $R = R_1 C$, 则 R 是 $r \times n$ 行满秩矩阵, 并且

$$QA = (QB)C = \begin{bmatrix} R_1 \\ 0 \end{bmatrix} C = \begin{bmatrix} R \\ 0 \end{bmatrix}$$

记 $Q^H = [Q_1, Q_2]$, 其中 Q_1 是 $m \times r$ 列正交规范矩阵, 则由上式即得(4.4.7). \square

类似于定理 4.4.2 的证明, 利用 Householder 矩阵也可以给出定理 4.4.3 的另一种构造性证明.

§ 4.5 Schur 定理与正规矩阵

本节先介绍在矩阵理论中十分重要的 Schur 定理, 然后讨论它的一些应用.

定义 4.5.1 设 $A, B \in \mathbb{R}^{n \times n} (\mathbb{C}^{n \times n})$, 如果存在 n 阶正交(酉)矩阵 U 使得

$$U^T A U = U^{-1} A U = B \quad (U^H A U = U^{-1} A U = B)$$

则称 A 正交(酉)相似于 B .

定理 4.5.1 (Schur 定理) 任何一个 n 阶复矩阵 A 都酉相似于一个上三角矩阵, 即存在一个 n 阶酉矩阵 U 和一个 n 阶上三角矩阵 R 使得

$$U^H A U = R \quad (4.5.1)$$

其中 R 的对角元是 A 的特征值, 它们可以按要求的次序排列.

证明 对矩阵的阶数作归纳法. 对 1 阶矩阵结论显然成立. 假设对 $n-1$ 阶矩阵结论成立. 对 n 阶矩阵 A , 按要求的次序取 A 的第 1 个特征值 λ_1 , 相应的特征向量为 ξ_1 , 则

$$A \xi_1 = \lambda_1 \xi_1 \quad (4.5.2)$$

对非零向量 ξ_1 , 由定理 4.1.2 知, 有 Householder 矩阵 H_1 使得

$$H_1 \xi_1 = \sigma e_1 \quad (4.5.3)$$

其中 $|\sigma|^2 = \|\xi_1\|^2 > 0$. 由 (4.5.2) 可得

$$(H_1 A H_1^H)(H_1 \xi_1) = \lambda_1 H_1 \xi_1$$

由 (4.5.3), 上式化为

$$(H_1 A H_1^H) e_1 = \lambda_1 e_1$$

因此矩阵 $H_1 A H_1^H = A^{(1)}$ 具有如下形式

$$A^{(1)} = \begin{bmatrix} \lambda_1 & \beta \\ 0 & A_1 \end{bmatrix} \quad (4.5.4)$$

其中 A_1 为 $n-1$ 阶矩阵. 根据归纳法假设, 存在 $n-1$ 阶酉矩阵 U_1 和上三角矩阵 R_1 使得

$$U_1^H A_1 U_1 = R_1$$

并且 $n-1$ 阶上三角矩阵 \mathbf{R}_1 的 $n-1$ 个对角元按要求的次序排列. 令

$$\mathbf{U} = \mathbf{H}_1^H \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{U}_1 \end{bmatrix}$$

则 \mathbf{U} 是 n 阶酉矩阵, 并且

$$\begin{aligned} \mathbf{U}^H \mathbf{A} \mathbf{U} &= \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{U}_1^H \end{bmatrix} \mathbf{H}_1 \mathbf{A} \mathbf{H}_1^H \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{U}_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{U}_1^H \end{bmatrix} \begin{bmatrix} \lambda_1 & \beta \\ 0 & \mathbf{A}_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{U}_1 \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & \beta \mathbf{U}_1 \\ 0 & \mathbf{U}_1^H \mathbf{A}_1 \mathbf{U}_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \beta \mathbf{U}_1 \\ 0 & \mathbf{R}_1 \end{bmatrix} = \mathbf{R} \end{aligned}$$

可见对 n 阶矩阵结论也成立, 这就证明了定理. \square

定义 4.5.2 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 如果

$$\mathbf{A} \mathbf{A}^H = \mathbf{A}^H \mathbf{A} \quad (4.5.5)$$

则称 \mathbf{A} 为正规矩阵.

显然, 对角矩阵、Hermite 矩阵、反 Hermite 矩阵、正交(酉)矩阵都是正规矩阵.

应用定理 4.5.1 可证明如下两个定理.

定理 4.5.2 n 阶矩阵 \mathbf{A} 酉相似于一个对角矩阵的充分必要条件为 \mathbf{A} 是正规矩阵.

证明 必要性. 若矩阵 \mathbf{A} 酉相似于对角矩阵 $\mathbf{\Lambda}$, 即存在酉矩阵 \mathbf{U} 使 $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$, 则

$$\mathbf{A}^H \mathbf{A} = \mathbf{U} \mathbf{\Lambda}^H \mathbf{U}^H \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H = \mathbf{U} \mathbf{\Lambda}^H \mathbf{\Lambda} \mathbf{U}^H = \mathbf{U} \mathbf{\Lambda} \mathbf{\Lambda}^H \mathbf{U}^H = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H \mathbf{U} \mathbf{\Lambda}^H \mathbf{U}^H = \mathbf{A} \mathbf{A}^H$$

这说明 \mathbf{A} 是正规矩阵.

充分性. 由 Schur 定理知存在酉矩阵 \mathbf{U} 使得 $\mathbf{A} = \mathbf{U} \mathbf{R} \mathbf{U}^H$, 其中 \mathbf{R} 是上三角矩阵. 记

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \\ 0 & r_{22} & r_{23} & \cdots & r_{2n} \\ 0 & 0 & r_{33} & \cdots & r_{3n} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & r_{nn} \end{bmatrix}$$

因为 $\mathbf{A}^H \mathbf{A} = \mathbf{A} \mathbf{A}^H$, 所以 $\mathbf{R}^H \mathbf{R} = \mathbf{R} \mathbf{R}^H$. 比较

$$\begin{aligned}
& \begin{pmatrix} \bar{r}_{11} & 0 & \cdots & 0 \\ r_{12} & \bar{r}_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \bar{r}_{1n} & r_{2n} & \cdots & \bar{r}_{nn} \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & r_{nn} \end{pmatrix} \\
&= \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & r_{nn} \end{pmatrix} \begin{pmatrix} \bar{r}_{11} & 0 & \cdots & 0 \\ \bar{r}_{12} & \bar{r}_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \bar{r}_{1n} & \bar{r}_{2n} & \cdots & \bar{r}_{nn} \end{pmatrix}
\end{aligned}$$

两边的对角元素, 即得 $\mathbf{R} = \mathbf{A} = \text{diag}(r_{11}, r_{22}, \cdots, r_{nn})$, 则 $\mathbf{A} = \mathbf{U}\mathbf{A}\mathbf{U}^H$. \square

推论 4.5.1 若 \mathbf{A} 是 n 阶 Hermite 矩阵, 则 \mathbf{A} 必酉相似于实对角矩阵, 即存在 n 阶酉矩阵 \mathbf{U} 使得

$$\mathbf{U}^H \mathbf{A} \mathbf{U} = \mathbf{A} \quad (4.5.6)$$

其中 $\mathbf{A} = \text{diag}(\lambda_1, \cdots, \lambda_n)$, $\lambda_i (i=1, \cdots, n)$ 是 \mathbf{A} 的实特征值.

证明 由定理 4.5.2 知, 存在 n 阶酉矩阵 \mathbf{U} 使得

$$\mathbf{U}^H \mathbf{A} \mathbf{U} = \mathbf{A} = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

因为 $\mathbf{A}^H = \mathbf{A}$, 则 $\mathbf{A}^H = \mathbf{A}$. 因此, $\mathbf{A} = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ 是实对角矩阵. \square

(4.5.6) 式称为 Hermite 矩阵 \mathbf{A} 的谱分解式.

定理 4.5.3 设 \mathbf{A}, \mathbf{B} 均为 n 阶正规矩阵, 并且 $\mathbf{AB} = \mathbf{BA}$, 则存在 n 阶酉矩阵 \mathbf{U} 使得 $\mathbf{U}^H \mathbf{A} \mathbf{U}$ 与 $\mathbf{U}^H \mathbf{B} \mathbf{U}$ 同时为对角矩阵.

证明 因为 \mathbf{A} 是正规矩阵, 由定理 4.5.2 知存在酉矩阵 \mathbf{U}_1 使得

$$\mathbf{A} = \mathbf{U}_1 \mathbf{D} \mathbf{U}_1^H$$

其中 $\mathbf{D} = \text{diag}(\mathbf{D}_1, \cdots, \mathbf{D}_r)$, $\mathbf{D}_i = d_i \mathbf{I}_{n_i} (i=1, \cdots, r)$, d_1, \cdots, d_r 互不相同, 且

$$\sum_{i=1}^r n_i = n.$$

记

$$\mathbf{C} = \mathbf{U}_1^H \mathbf{B} \mathbf{U}_1 = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1r} \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \cdots & \mathbf{C}_{2r} \\ \vdots & \vdots & & \vdots \\ \mathbf{C}_{r1} & \mathbf{C}_{r2} & \cdots & \mathbf{C}_{rr} \end{pmatrix}$$

其中 $C_u \in \mathbb{C}^{n_i \times n_i}$ ($i=1, \dots, r$). 因为 B 是正规矩阵, 则 C 也是正规矩阵, 且 $B = U_1 C U_1^H$, 因为 $AB = BA$, 所以 $DC = CD$. 直接比较

$$\begin{bmatrix} d_1 I_{n_1} & & 0 \\ & \ddots & \\ 0 & & d_r I_{n_r} \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1r} \\ C_{21} & C_{22} & \cdots & C_{2r} \\ \vdots & \vdots & & \vdots \\ C_{r1} & C_{r2} & \cdots & C_{rr} \end{bmatrix} =$$

$$\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1r} \\ C_{21} & C_{22} & \cdots & C_{2r} \\ \vdots & \vdots & & \vdots \\ C_{r1} & C_{r2} & \cdots & C_{rr} \end{bmatrix} \begin{bmatrix} d_1 I_{n_1} & & 0 \\ & \ddots & \\ 0 & & d_r I_{n_r} \end{bmatrix}$$

两边对应的块即得 $C = \text{diag}(C_{11}, \dots, C_{rr})$, 其中 C_u 是 n_i 阶正规矩阵. 由定理 4.5.2 知存在 n_i 阶酉矩阵 P_i 使得 $C_u = P_i \Lambda_i P_i^H$ ($i=1, \dots, r$), 其中 Λ_i 为 n_i 阶对角矩阵, 于是

$$C = \begin{bmatrix} P_1 & & 0 \\ & P_2 & \\ & & \ddots \\ 0 & & & P_r \end{bmatrix} \begin{bmatrix} \Lambda_1 & & 0 \\ & \Lambda_2 & \\ & & \ddots \\ 0 & & & \Lambda_r \end{bmatrix} \begin{bmatrix} P_1^H & & 0 \\ & P_2^H & \\ & & \ddots \\ 0 & & & P_r^H \end{bmatrix}$$

令

$$U = U_1 \begin{bmatrix} P_1 & & 0 \\ & \ddots & \\ 0 & & P_r \end{bmatrix}$$

则 U 是 n 阶酉矩阵, 并且

$$U^H A U = \begin{bmatrix} P_1^H & & \\ & \ddots & \\ & & P_r^H \end{bmatrix} U_1^H A U_1 \begin{bmatrix} P_1 & & \\ & \ddots & \\ & & P_r \end{bmatrix}$$

$$\begin{aligned}
&= \begin{bmatrix} \mathbf{P}_1^H & & \\ & \ddots & \\ & & \mathbf{P}_r^H \end{bmatrix} \begin{bmatrix} d_1 \mathbf{I}_{n_1} & & \\ & \ddots & \\ & & d_r \mathbf{I}_{n_r} \end{bmatrix} \begin{bmatrix} \mathbf{P}_1 & & \\ & \ddots & \\ & & \mathbf{P}_r \end{bmatrix} = \begin{bmatrix} d_1 \mathbf{I}_{n_1} & & \\ & \ddots & \\ & & d_r \mathbf{I}_{n_r} \end{bmatrix} \\
\mathbf{U}^H \mathbf{B} \mathbf{U} &= \begin{bmatrix} \mathbf{P}_1^H & & \\ & \ddots & \\ & & \mathbf{P}_r^H \end{bmatrix} \begin{bmatrix} \mathbf{U}_1^H \mathbf{B} \mathbf{U}_1 & & \\ & \ddots & \\ & & \mathbf{U}_r^H \mathbf{B} \mathbf{U}_r \end{bmatrix} \begin{bmatrix} \mathbf{P}_1 & & \\ & \ddots & \\ & & \mathbf{P}_r \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{P}_1^H & & \\ & \ddots & \\ & & \mathbf{P}_r^H \end{bmatrix} \begin{bmatrix} \mathbf{C}_{11} & & \\ & \ddots & \\ & & \mathbf{C}_{rr} \end{bmatrix} \begin{bmatrix} \mathbf{P}_1 & & \\ & \ddots & \\ & & \mathbf{P}_r \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_r \end{bmatrix} \quad \square
\end{aligned}$$

当 \mathbf{A} 是 n 阶实矩阵时, 可以修改定理 4.5.1 而得到只用实运算的因子分解, 这在实际应用中是很重要的.

定理 4.5.4 任何 n 阶实矩阵 \mathbf{A} 都正交相似于一个拟上三角矩阵, 即存在一个 n 阶正交矩阵 \mathbf{Q} 和一个 n 阶拟上三角矩阵 \mathbf{R} 使得

$$\mathbf{Q}^T \mathbf{A} \mathbf{Q} = \mathbf{R} \quad (4.5.7)$$

其中 \mathbf{R} 是块上三角矩阵(或称拟上三角矩阵), 其对角块为 1 阶块或 2 阶块, 每个 1 阶块是 \mathbf{A} 的实特征值, 而每个 2 阶块的两个特征值是 \mathbf{A} 的一对共轭复特征值, 且 \mathbf{R} 的对角块可以按要求的次序排列.

证明 与定理 4.5.1 的证明类似, 对矩阵的阶数作归纳法. 若 \mathbf{A} 的特征值全为实数, 则证法与定理 4.5.1 完全一样. 因此不妨假设开始的两个特征值是一对共轭复数 $\lambda \pm i\mu$ ($\mu \neq 0$). 显然, 只要证明存在正交矩阵 \mathbf{Q}_1 使得

$$\mathbf{Q}_1^T \mathbf{A} \mathbf{Q}_1 = \begin{bmatrix} \mathbf{D}_1 & \mathbf{B}_1 \\ 0 & \mathbf{A}_1 \end{bmatrix} \quad (4.5.8)$$

其中 \mathbf{D}_1 是 2 阶实矩阵, 它的两个特征值是 $\lambda \pm i\mu$, 而 \mathbf{B}_1 是 $2 \times (n-2)$ 实矩阵, \mathbf{A}_1 是 $n-2$ 阶实矩阵.

设 \mathbf{A} 对应于特征值 $\lambda \pm i\mu$ 的特征向量为 $x \pm iy$, 即

$$\mathbf{A}(x \pm iy) = (\lambda \pm i\mu)(x \pm iy) \quad (4.5.9)$$

则实向量 x 和 y 必线性无关(否则 $x + iy$ 与 $x - iy$ 线性相关, 这与定理 2.4.4 矛盾). 上式可以改写为

$$\mathbf{A}[x, y] = [x, y] \begin{bmatrix} \lambda & \mu \\ -\mu & \lambda \end{bmatrix} = [x, y] \mathbf{A}_1 \quad (4.5.10)$$

其中 $\mathbf{A}_1 = \begin{bmatrix} \lambda & \mu \\ -\mu & \lambda \end{bmatrix}$. 由定理 4.4.2 知存在 n 阶正交矩阵 \mathbf{Q}_1 和一个 2 阶可逆上三角矩阵 \mathbf{R}_1 使得

$$\mathbf{Q}_1^T [x, y] = \begin{bmatrix} \mathbf{R}_1 \\ 0 \end{bmatrix} \quad (4.5.11)$$

由 (4.5.10), 可得

$$(\mathbf{Q}_1^T \mathbf{A} \mathbf{Q}_1)(\mathbf{Q}_1^T [x, y]) = \mathbf{Q}_1^T [x, y] \mathbf{A}_1 \quad (4.5.12)$$

令

$$\mathbf{Q}_1^T \mathbf{A} \mathbf{Q}_1 = \begin{bmatrix} \mathbf{D}_1 & \mathbf{B}_1 \\ \mathbf{C}_1 & \mathbf{A}_1 \end{bmatrix}$$

其中 \mathbf{D}_1 是 2 阶实矩阵, 则由 (4.5.12) 可得

$$\begin{bmatrix} \mathbf{D}_1 & \mathbf{B}_1 \\ \mathbf{C}_1 & \mathbf{A}_1 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_1 \\ 0 \end{bmatrix} \mathbf{A}_1$$

从而

$$\mathbf{D}_1 \mathbf{R}_1 = \mathbf{R}_1 \mathbf{A}_1, \quad \mathbf{C}_1 \mathbf{R}_1 = 0$$

因为 \mathbf{R}_1 可逆, 则

$$\mathbf{D}_1 = \mathbf{R}_1 \mathbf{A}_1 \mathbf{R}_1^{-1}, \quad \mathbf{C}_1 = 0$$

而 \mathbf{A}_1 的两个特征值恰为 \mathbf{A} 的一对共轭复特征值 $\lambda \pm i\mu$, 故 2 阶实矩阵 \mathbf{D}_1 的两个特征值也是 $\lambda \pm i\mu$. 这就证明了 (4.5.8). \square

推论 4.5.2 若 \mathbf{A} 是 n 阶实对称矩阵, 则 \mathbf{A} 正交相似于实对角矩阵, 即存在 n 阶正交矩阵 \mathbf{Q} 使得

$$\mathbf{Q}^T \mathbf{A} \mathbf{Q} = \mathbf{\Lambda} \quad (4.5.13)$$

其中 $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$, $\lambda_i (i=1, \dots, n)$ 是 \mathbf{A} 的实特征值.

证明 因为实对称矩阵是 Hermite 矩阵, 则由推论 4.5.1 知 \mathbf{A} 的特征值全是实数. 因此由定理 4.5.4 即得结论. \square

§ 4.6 奇异值分解

本节介绍矩阵的奇异值分解,它不仅是矩阵理论和矩阵计算的最基本和最重要的工具之一,而且在控制理论,系统辨识和信号处理等许多领域都有直接的应用.

为了引入矩阵的奇异值概念,我们先证明两个引理.

引理 4.6.1 设 $A \in \mathbb{C}^{m \times n}$, 则

$$\text{rank}(A^H A) = \text{rank}(A A^H) = \text{rank}(A) \quad (4.6.1)$$

证明 如果 $x \in \mathbb{C}^n$ 是齐次线性方程组 $Ax=0$ 的解,则 x 显然是齐次线性方程组 $A^H A x=0$ 的解;反过来,如果 x 是 $A^H A x=0$ 的解,则 $x^H A^H A x=0$, 即 $(Ax)^H (Ax)=0$, 于是 $Ax=0$, 这表明 x 也是 $Ax=0$ 的解. 因此齐次线性方程组 $Ax=0$ 与 $A^H A x=0$ 同解,从而 $\text{rank}(A^H A) = \text{rank}(A)$.

同样可证 $\text{rank}(A A^H) = \text{rank}(A)$. \square

引理 4.6.2 设 $A \in \mathbb{C}^{m \times n}$, 则

- (1) $A^H A$ 与 $A A^H$ 的特征值均为非负实数;
- (2) $A^H A$ 与 $A A^H$ 的非零特征值相同,并且非零特征值的个数(重特征值按重数计算)等于 $\text{rank}(A)$.

证明 (1) 设 λ 是 $A^H A$ 的任一特征值, $x \neq 0$ 为相应的特征向量, 则

$$A^H A x = \lambda x$$

因为 $A^H A$ 是 Hermite 矩阵, 所以 λ 是实数, 并且

$$\lambda x^H x = x^H A^H A x = (Ax)^H (Ax) \geq 0$$

由于 $x^H x > 0$, 因此 $\lambda \geq 0$.

类似可证 $A A^H$ 的特征值均为非负实数.

(2) 由例 2.4.5 知, $A^H A$ 与 $A A^H$ 的非零特征值相同, 并且它们非零特征值的个数等于 $\text{rank}(A^H A)$. 由引理 4.6.1 即得结论. \square

定义 4.6.1 设 $A \in \mathbb{C}^{m \times n}$, 如果存在非负实数 σ 和非零向量 $u \in \mathbb{C}^n, v \in \mathbb{C}^m$ 使得

$$A u = \sigma v, \quad A^H v = \sigma u \quad (4.6.2)$$

则称 σ 为 A 的**奇异值**, u 和 v 分别称为 A 对应于奇异值 σ 的**右奇异向量**和**左奇异向量**.

由(4.6.2)可得

$$A^H A u = \sigma A^H v = \sigma^2 u \quad (4.6.3)$$

$$AA^H v = \sigma A u = \sigma^2 v \quad (4.6.4)$$

因此 σ^2 是 $A^H A$ 的特征值,也是 AA^H 的特征值,而 u 和 v 分别是 $A^H A$ 和 AA^H 对应于特征值 σ^2 的特征向量.

设 $A \in \mathbb{C}^{m \times n}$, $\text{rank}(A) = r$, 且 $A^H A$ 的特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. 由引理 4.6.2 知, $\lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0$. 记 $k = \min\{m, n\}$, 也称 $\sigma_i = \sqrt{\lambda_i}$ ($i=1, \cdots, k$) 为 A 的奇异值, 特别地, 称 $\sigma_1, \cdots, \sigma_r$ 为 A 的正奇异值.

定理 4.6.1 若 A 是正规矩阵, 则 A 的奇异值是 A 的特征值的模.

证明 设 A 是 n 阶正规矩阵, 由定理 4.5.2 知存在 n 阶酉矩阵 U 使得

$$A = U \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) U^H$$

其中 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 是 A 的特征值, 则

$$A^H A = U \text{diag}(|\lambda_1|^2, |\lambda_2|^2, \cdots, |\lambda_n|^2) U^H$$

于是 $A^H A$ 的特征值为 $|\lambda_1|^2, |\lambda_2|^2, \cdots, |\lambda_n|^2$. 因此 A 的奇异值是 $|\lambda_1|, |\lambda_2|, \cdots, |\lambda_n|$. \square

定理 4.6.2 设 A 是 $m \times n$ 矩阵, 且 $\text{rank}(A) = r$, 则存在 m 阶酉矩阵 V 和 n 阶酉矩阵 U 值得

$$V^H A U = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \quad (4.6.5)$$

其中 $\Sigma = \text{diag}(\sigma_1, \cdots, \sigma_r)$, 且 $\sigma_1 \geq \cdots \geq \sigma_r > 0$.

证明 因为 $\text{rank}(A) = r$, 由引理 4.6.2, 可设 $A^H A$ 的特征值是

$$\sigma_1^2 \geq \cdots \geq \sigma_r^2 > 0, \sigma_{r+1}^2 = \cdots = \sigma_n^2 = 0$$

因为 $A^H A$ 是 Hermite 矩阵, 由推论 4.5.1 知存在 n 阶酉矩阵 U 使得

$$U^H A^H A U = \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$$

或

$$A^H A U = U \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} \quad (4.6.6)$$

记 $U = [U_1, U_2]$, 其中 U_1 是 $n \times r$ 矩阵. 上式可改写为

$$A^H A [U_1, U_2] = [U_1, U_2] \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} \quad (4.6.7)$$

则有

$$\mathbf{A}^H \mathbf{A} \mathbf{U}_1 = \mathbf{U}_1 \boldsymbol{\Sigma}^2, \mathbf{A}^H \mathbf{A} \mathbf{U}_2 = 0 \quad (4.6.8)$$

记 $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2]$, 其中 \mathbf{V}_1 是 $m \times r$ 矩阵, \mathbf{V}_2 是 $m \times (m-r)$ 矩阵. 令

$$\mathbf{V}_1 = \mathbf{A} \mathbf{U}_1 \boldsymbol{\Sigma}^{-1} \quad (4.6.9)$$

由(4.6.8)的第一式知, \mathbf{V}_1 是列正交规范矩阵, 即 $\mathbf{V}_1^H \mathbf{V}_1 = \mathbf{I}_r$. 取 \mathbf{V}_2 使 $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2]$ 是酉矩阵, 则

$$\mathbf{V}_2^H \mathbf{A} \mathbf{U}_1 = \mathbf{V}_2^H \mathbf{V}_1 \boldsymbol{\Sigma} = 0 \quad (4.6.10)$$

由(4.6.8)和(4.6.10)有

$$\mathbf{V}^H \mathbf{A} \mathbf{U} = \begin{bmatrix} \mathbf{V}_1^H \\ \mathbf{V}_2^H \end{bmatrix} \mathbf{A} [\mathbf{U}_1 \quad \mathbf{U}_2] = \begin{bmatrix} \mathbf{V}_1^H \mathbf{A} \mathbf{U}_1 & \mathbf{V}_1^H \mathbf{A} \mathbf{U}_2 \\ \mathbf{V}_2^H \mathbf{A} \mathbf{U}_1 & \mathbf{V}_2^H \mathbf{A} \mathbf{U}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma} & 0 \\ 0 & 0 \end{bmatrix}$$

这就得到了(4.6.5). \square

(4.6.5)称为矩阵 \mathbf{A} 的奇异值分解.

从定理 4.6.2 可见, \mathbf{U} 的列向量是 $\mathbf{A}^H \mathbf{A}$ 的标准正交特征向量, \mathbf{U} 的前 r 列向量是 $\mathbf{A}^H \mathbf{A}$ 对应于 r 个非零特征值 $\sigma_1^2, \dots, \sigma_r^2$ 的标准正交特征向量; 而 \mathbf{V} 的列向量是 $\mathbf{A} \mathbf{A}^H$ 的标准正交特征向量, 从 $\mathbf{V}_1 = \mathbf{A} \mathbf{U}_1 \boldsymbol{\Sigma}^{-1}$ 的选取, 有

$$\mathbf{A} \mathbf{A}^H \mathbf{V}_1 = \mathbf{A} \mathbf{A}^H \mathbf{A} \mathbf{U}_1 \boldsymbol{\Sigma}^{-1} = \mathbf{A} \mathbf{U}_1 \boldsymbol{\Sigma}^2 \boldsymbol{\Sigma}^{-1} = \mathbf{A} \mathbf{U}_1 \boldsymbol{\Sigma} = \mathbf{V}_1 \boldsymbol{\Sigma}^2$$

这表明 \mathbf{V} 的前 r 列向量恰是 $\mathbf{A} \mathbf{A}^H$ 对应于特征值 $\sigma_1^2, \dots, \sigma_r^2$ 的标准正交特征向量.

例 4.6.1 设

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

作出矩阵 \mathbf{A} 的奇异值分解.

解 因为

$$\mathbf{A}^H \mathbf{A} = \begin{pmatrix} 2 & 0 \\ 0 & 5 \end{pmatrix}$$

则 \mathbf{A} 的非零奇异值为 $\sqrt{2}, \sqrt{5}$. $\mathbf{A}^H \mathbf{A}$ 对应于特征值 5 和 2 的标准正交特征向量

为 $u_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $u_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, 而 AA^H 对应于特征值 5 和 2 的标准正交特征向量为

$$v_1 = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ 0 \\ \frac{2}{\sqrt{5}} \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}, AA^{11} \text{ 对应于特征值 0 的标准正交特征向量为}$$

$$v_3 = \begin{pmatrix} -\frac{2}{\sqrt{5}} \\ 0 \\ \frac{1}{\sqrt{5}} \\ 0 \end{pmatrix}, v_4 = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}. \text{ 因此 } A \text{ 的奇异值分解为}$$

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \\ 0 & 2 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} & 0 & -\frac{2}{\sqrt{5}} & 0 \\ 0 & -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{5} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^H.$$

习 题

1. 设 A 是 n 阶可逆矩阵, $u, v \in \mathbb{C}^n$. 证明: 如果 $v^H A^{-1} u \neq 1$, 则 $A - uv^H$ 非奇异, 并且有 Sherman-Morrison 公式

$$(A - uv^H)^{-1} = A^{-1} - \frac{A^{-1}uv^HA^{-1}}{v^HA^{-1}u - 1}$$

2. 设 A 是 n 阶可逆矩阵, $U, V \in \mathbb{C}^{n \times m} (n \geq m)$. 证明: 如果 $V^HA^{-1}U - I$ 非奇异, 则 $A - UV^H$ 可逆, 并且

$$(A - UV^H)^{-1} = A^{-1} + A^{-1}U(V^HA^{-1}U - I)^{-1}V^HA^{-1}$$

3. 求下列矩阵 A 的满秩分解

$$(1) A = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 1 & -1 \\ 1 & 0 & 2 & 1 \end{pmatrix}; \quad (2) A = \begin{pmatrix} 1 & -1 & 1 & 1 \\ -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \end{pmatrix};$$

$$(3) A = \begin{pmatrix} 2 & 1 & 1 & 2 \\ -3 & -1 & 2 & 0 \\ 5 & 1 & 4 & -1 \end{pmatrix}; \quad (4) A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 3 & 2 & 1 & 1 & -3 \\ 0 & 1 & 2 & 2 & 6 \\ 5 & 4 & 3 & 3 & -1 \end{pmatrix}.$$

4. 设 $A, B \in C^{m \times n}$, 证明: $\text{rank}(A+B) \leq \text{rank}(A) + \text{rank}(B)$.

5. 设 A 和 B 分别为 $m \times n$ 和 $n \times m$ 矩阵, 如果 $AB = I$, 则称 B 为 A 的右逆矩阵, A 为 B 的左逆矩阵. 证明: A 有右逆矩阵的充分必要条件是 A 为行满秩矩阵.

6. 设 B 和 C 分别为 $m \times r$ 和 $r \times n$ 矩阵, 并且 $\text{rank}(B) = \text{rank}(C) = r$. 证明: $\text{rank}(BC) = r$.

7. 设 G 是 $p \times m$ 列满秩矩阵, A 是 $m \times n$ 矩阵, H 是 $n \times k$ 行满秩矩阵, 证明: $\text{rank}(GAH) = \text{rank}(GA) = \text{rank}(AH) = \text{rank}(A)$.

8. 设 A 是可逆矩阵, 证明:

$$(1) \begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A| |D - CA^{-1}B|;$$

$$(2) \text{rank} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \text{rank}(A) + \text{rank}(D - CA^{-1}B).$$

9. 求下列矩阵的 LU 分解和 LDU 分解

$$(1) A = \begin{pmatrix} 2 & 3 & 4 \\ 1 & 1 & 9 \\ 1 & 2 & -6 \end{pmatrix}; \quad (2) A = \begin{pmatrix} 2 & 3 & 4 \\ 3 & 5 & 2 \\ 4 & 3 & 30 \end{pmatrix}.$$

10. 利用系数矩阵的 LU 分解解下列线性方程组

$$(1) \begin{cases} 2x_1 + 3x_2 + 4x_3 = 1, \\ x_1 + x_2 + 9x_3 = -7, \\ x_1 + 2x_2 - 6x_3 = 9; \end{cases} \quad (2) \begin{cases} 2x_1 + 3x_2 + 4x_3 = 0, \\ 3x_1 + 5x_2 + 2x_3 = -5, \\ 4x_1 + 3x_2 + 30x_3 = 28. \end{cases}$$

11. 设 A 是 n 阶非奇异矩阵, 证明: 存在 n 阶排列矩阵 P 使得 PA 的顺序主子式都不等于零.

12. 作出下列矩阵的 QR 分解

$$(1) A = \begin{pmatrix} 2 & 2 & 1 \\ 0 & 2 & 2 \\ 2 & 1 & 2 \end{pmatrix}; \quad (2) A = \begin{pmatrix} 1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix}.$$

13. 设 P, Q 分别为 m 阶方阵及 n 阶方阵, 证明: 若 $m+n$ 阶方阵

$$A = \begin{pmatrix} P & B \\ 0 & Q \end{pmatrix}$$

是酉矩阵, 则 P, Q 也是酉矩阵, 且 B 是零矩阵.

14. 证明: 如果 A 是上三角矩阵, 并且是酉矩阵, 则 A 是对角矩阵, 并且其对角元的模均为 1.

15. 设 V_1 为 $n \times k (k < n)$ 列正交规范矩阵, 证明: 存在 $n \times (n-k)$ 列正交规范矩阵 V_2 使得 $V = [V_1, V_2]$ 为酉矩阵.

16. 对下列矩阵

$$\begin{aligned} (1) A &= \begin{pmatrix} 2 & -2 & 0 \\ -2 & 1 & -2 \\ 0 & -2 & 0 \end{pmatrix}; & (2) A &= \begin{pmatrix} 0 & -1 & i \\ 1 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}; \\ (3) A &= \begin{pmatrix} 3 & 0 & 8 \\ 3 & -1 & 0 \\ -3 & -6 & 1 \end{pmatrix}; & (4) A &= \begin{pmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{pmatrix}. \end{aligned}$$

试求酉(正交)矩阵 U 使得 $U^H A U$ 是上三角矩阵.

17. 设 A 是 n 阶矩阵, 记 $A_1 = A$. 对 $k=1, 2, \dots$, 作 A_k 的 QR 分解

$$A_k = Q_k R_k$$

其中 Q_k 是酉矩阵, R_k 是上三角矩阵, 并计算

$$A_{k+1} = R_k Q_k$$

上述产生矩阵序列 $\{A_k\}$ 的过程称为 QR 算法. 证明:

(1) $A_k (k=1, 2, \dots)$ 酉相似于 A ;

(2) $A_k = Q^{(k)} R^{(k)}$, 其中 $Q^{(k)} = Q_1 Q_2 \cdots Q_k$, $R^{(k)} = R_k R_{k-1} \cdots R_1$.

18. 设 A, B 均为正规矩阵, 试证: A 与 B 酉相似的充分必要条件是 A 与 B 的特征值相同.

19. 设 A 是 Hermite 矩阵, 且 $A^2 = A$, 则存在酉矩阵 U 使得

$$U^H A U = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

其中 $r = \text{rank}(A)$.

20. 设 A 是 Hermite 矩阵, 且 $A^2 = I$, 则存在酉矩阵 U 使得

$$U^H A U = \begin{pmatrix} I_r & 0 \\ 0 & -I_{n-r} \end{pmatrix}$$

其中 r 是 A 的正特征值的个数.

21. 设 A 是 n 阶实矩阵, 证明: A 为正规矩阵的充分必要条件是存在 n 阶正交矩阵 Q

$$Q^T A Q = R = \text{diag}(R_1, R_2, \dots, R_s)$$

其中 $R_i (i=1, 2, \dots, s)$ 是 1 阶或 2 阶形如 $\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$ 的实矩阵.

22. 设 A 是 n 阶实矩阵, 证明: A 为正交矩阵的充分必要条件是存在 n 阶正交矩阵 Q 使得

$$Q^T A Q = T = \text{diag}(T_1, T_2, \dots, T_s, \lambda_{2s+1}, \dots, \lambda_n)$$

其中 $T_i = \begin{bmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{bmatrix}$ ($i = 1, 2, \dots, s$), $\lambda_{2s+1}, \dots, \lambda_n$ 等于 1 或 -1.

23. 设 A, B 均为 n 阶 Hermite 矩阵, 证明: 存在 n 阶酉矩阵 U 使得 $U^{H1}AU^{-1}$ 与 $U^{H1}BU$ 均为对角矩阵的充分必要条件是 $AB=BA$.

24. 作出下列矩阵的奇异值分解

$$(1) A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}; \quad (2) A = \begin{bmatrix} 0 & -1 & 1 \\ 2 & 0 & 0 \end{bmatrix}; \quad (3) A = \begin{bmatrix} -1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix}.$$

第五章 Hermite 矩阵与正定矩阵

实对称矩阵是一类十分重要的矩阵,它在力学、物理学、自动控制与工程技术中有着广泛的应用.复矩阵中的 Hermite 矩阵与实对称矩阵在其性质和证明方法上都十分相似.实对称矩阵在学习线性代数中已有所了解.本章介绍 Hermite 矩阵及其性质,讨论一类具有特殊正性的 Hermite 矩阵——正定(非负定)矩阵及其性质,在此基础上介绍矩阵不等式.

§ 5.1 Hermite 矩阵与 Hermite 二次型

5.1.1 Hermite 矩阵

由 Hermite 矩阵的定义可知, Hermite 矩阵具有如下简单性质:

- (1) 如果 A 是 Hermite 矩阵,则对正整数 k , A^k 也是 Hermite 矩阵;
- (2) 如果 A 是可逆 Hermite 矩阵,则 A^{-1} 也是 Hermite 矩阵;
- (3) 如果 A, B 是 Hermite 矩阵,则对实数 k, p , $kA + pB$ 也是 Hermite 矩阵;
- (4) 若 A, B 是 Hermite 矩阵,则 AB 是 Hermite 矩阵的充分必要条件是 $AB = BA$;
- (5) A 是 Hermite 矩阵的充分必要条件是对任意方阵 S , $S^H A S$ 是 Hermite 矩阵.

定理 5.1.1 设 $A = (a_{jk}) \in C^{n \times n}$, 则 A 是 Hermite 矩阵的充分必要条件是 对任意 $x \in C^n$, $x^H A x$ 是实数.

证明 必要性. 如果 A 是 Hermite 矩阵, 则对任意 $x \in C^n$, 因为 $x^H A x$ 是数, 所以

$$\overline{(x^H A x)} = (x^H A x)^H = x^H A^H x = x^H A x$$

因此 $x^H A x$ 是实数.

充分性. 因为对任意 $x, y \in C^n$, $x^H A x, y^H A y, (x+y)^H A (x+y)$ 都是实数, 而

$$(x+y)^H A (x+y) = x^H A x + x^H A y + y^H A x + y^H A y$$

于是对任意 $x, y \in C^n$, $x^H A y + y^H A x$ 是实数. 特别地, 令

$$x = (\underbrace{0, \cdots, 0}_j, 1, 0, \cdots, 0)^T, y = (\underbrace{0, \cdots, 0}_k, 1, 0, \cdots, 0)^T$$

则 $x^H A y + y^H A x = a_{jk} + a_{kj}$ 是实数, 这表明 a_{jk} 与 a_{kj} 的虚部值相等, 但符号相反. 令

$$x = (\underbrace{0, \cdots, 0}_j, i, \underbrace{0, \cdots, 0}_k)^T, \quad y = (\underbrace{0, \cdots, 0}_k, 1, \underbrace{0, \cdots, 0}_j)^T$$

其中 $i = \sqrt{-1}$, $x^H A y + y^H A x = -i a_{jk} + i a_{kj}$ 是实数, 则 a_{jk} 与 a_{kj} 的实部相等. 因此

$$a_{jk} = \bar{a}_{kj}, \quad j, \quad k = 1, 2, \cdots, n$$

即 A 是 Hermite 矩阵. \square

定理 5.1.2 设 A 为 n 阶 Hermite 矩阵, 则

- (1) A 的所有特征值全是实数;
- (2) A 的不同特征值所对应的特征向量是互相正交的.

证明 由推论 4.5.1 即得(1). 下面证明(2).

设 λ, μ 是 A 的两个不同特征值, 相应的特征向量分别为 x, y , 则

$$Ax = \lambda x, \quad Ay = \mu y$$

从而 $y^H Ax = \lambda y^H x$, $x^H Ay = \mu x^H y$. 因为 A 是 Hermite 矩阵, λ, μ 均为实数, 则 $y^H Ax = \mu y^H x$. 于是 $(\lambda - \mu)y^H x = 0$. 由于 $\lambda \neq \mu$, 故 x 与 y 正交. \square

由推论 4.5.1, 容易证明如下定理.

定理 5.1.3 设 $A \in C^{n \times n}$, 则 A 是 Hermite 矩阵的充分必要条件是存在酉矩阵 U 使得

$$U^H A U = \Lambda = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \quad (5.1.1)$$

其中 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 均为实数.

实对称矩阵的特征值全为实数, 相应的特征向量可取为实向量(这是与 Hermite 矩阵的不同之处). 由推论 4.5.2 可证明如下定理.

定理 5.1.4 设 $A \in R^{n \times n}$, 则 A 是实对称矩阵的充分必要条件是存在正交矩阵 Q 使得

$$Q^T A Q = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \quad (5.1.2)$$

其中 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 均为实数.

5.1.2 矩阵的惯性

定理 5.1.5 设 A 是 n 阶 Hermite 矩阵, 则 A 相合于矩阵

$$D_0 = \begin{bmatrix} I_r & 0 & 0 \\ 0 & -I_{r-s} & 0 \\ 0 & 0 & O_{n-r} \end{bmatrix} \quad (5.1.3)$$

其中 $r = \text{rank}(\mathbf{A})$, s 是 \mathbf{A} 的正特征值(重特征值按重数计算)的个数.

证明 由定理 5.1.3 知, 存在 n 阶酉矩阵 \mathbf{U} 使得

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\text{H}}$$

其中 $\mathbf{\Lambda}$ 是以 \mathbf{A} 的特征值为对角元的对角矩阵. 排列 $\mathbf{\Lambda}$ 的特征值使得

$$\mathbf{P}_1^{\text{H}}\mathbf{A}\mathbf{P}_1 = \text{diag}(\lambda_1, \dots, \lambda_s, \lambda_{s+1}, \dots, \lambda_r, 0, \dots, 0) = \mathbf{D}$$

其中 \mathbf{P}_1 是排列矩阵, $\lambda_i > 0 (i = 1, \dots, s)$, $\lambda_j < 0 (j = s+1, \dots, r)$, 令

$$\mathbf{D}_1 = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_s}, \sqrt{|\lambda_{s+1}|}, \dots, \sqrt{|\lambda_r|}, 1, \dots, 1)$$

令 $\mathbf{P} = \mathbf{U}\mathbf{P}_1\mathbf{D}_1^{-1}$, 则 \mathbf{P} 非奇异且 $\mathbf{P}^{\text{H}}\mathbf{A}\mathbf{P} = \mathbf{D}_0$, 即 \mathbf{A} 相合于 \mathbf{D}_0 . \square

式(5.1.3)中矩阵 \mathbf{D}_0 称为 n 阶 Hermite 矩阵 \mathbf{A} 的相合标准形.

设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, $\pi(\mathbf{A})$, $v(\mathbf{A})$ 和 $\delta(\mathbf{A})$ 分别表示 \mathbf{A} 的位于右半开平面、左半开平面和虚轴上的特征值(重特征值按重数计算)的个数. 记

$$\text{In}(\mathbf{A}) = \{\pi(\mathbf{A}), v(\mathbf{A}), \delta(\mathbf{A})\} \quad (5.1.4)$$

则称 $\text{In}(\mathbf{A})$ 为矩阵 \mathbf{A} 的惯性.

如果 \mathbf{A} 是 n 阶 Hermite 矩阵, 则 $\pi(\mathbf{A})$, $v(\mathbf{A})$ 和 $\delta(\mathbf{A})$ 分别表示 \mathbf{A} 的正、负和零特征值的个数(重特征值按重数计算). 因此 \mathbf{A} 非奇异的充分必要条件为 $\delta(\mathbf{A}) = 0$, 并且

$$\pi(\mathbf{A}) + v(\mathbf{A}) = \text{rank}(\mathbf{A}) \quad (5.1.5)$$

定理 5.1.6 (Sylvester 惯性定律) 设 \mathbf{A}, \mathbf{B} 均为 n 阶 Hermite 矩阵, 则 \mathbf{A} 与 \mathbf{B} 相合的充分必要条件是

$$\text{In}(\mathbf{A}) = \text{In}(\mathbf{B}) \quad (5.1.6)$$

证明 必要性. 如果 \mathbf{A} 与 \mathbf{B} 相合, 则 $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{B}) = r$. 由定理 5.1.5 知存在非奇异矩阵 \mathbf{P} 和 \mathbf{Q} 使得

$$\mathbf{P}^{\text{H}}\mathbf{A}\mathbf{P} = \text{diag}(\mathbf{I}_s, -\mathbf{I}_t, \mathbf{0}_{n-r}) = \mathbf{D}_0^{(1)}, \mathbf{Q}^{\text{H}}\mathbf{B}\mathbf{Q} = \text{diag}(\mathbf{I}_t, -\mathbf{I}_r, \mathbf{0}_{n-r}) = \mathbf{D}_0^{(2)}$$

其中 $s = \pi(\mathbf{A})$, $t = \pi(\mathbf{B})$.

下面证明 $s = t$. 因为 \mathbf{A} 与 \mathbf{B} 相合, 则存在非奇异矩阵 \mathbf{M} 使得 $\mathbf{B} = \mathbf{M}^{\text{H}}\mathbf{A}\mathbf{M}$, 从而

$$\mathbf{D}_0^{(1)} = \mathbf{P}^{\text{H}}\mathbf{A}\mathbf{P} = (\mathbf{Q}^{-1}\mathbf{M}^{-1}\mathbf{P})^{\text{H}}\mathbf{D}_0^{(2)}\mathbf{Q}^{-1}\mathbf{M}^{-1}\mathbf{P} = \mathbf{R}^{\text{H}}\mathbf{D}_0^{(2)}\mathbf{R} \quad (5.1.7)$$

其中 $\mathbf{R} = \mathbf{Q}^{-1}\mathbf{M}^{-1}\mathbf{P}$.

假若 $t < s$. 令 $x = \begin{pmatrix} y \\ 0 \end{pmatrix} \in \mathbb{C}^n$, 其中 $y = \begin{pmatrix} y_1 \\ \vdots \\ y_s \end{pmatrix} \in \mathbb{C}^s$, 且 $y \neq 0$, 则

$$x^H D_0^{(1)} x = \sum_{i=1}^r |y_i|^2 > 0 \quad (5.1.8)$$

记

$$R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}$$

其中 $R_{11} \in C^{t \times t}$. 因为 $t < s$, 则可选取 $y \neq 0$ 使得 $R_{11}y = 0$. 从而 $Rx = \begin{pmatrix} 0 \\ z \end{pmatrix}$, 其中 $z = R_{21}y = (z_1, \dots, z_{n-t})^T$. 由 (5.1.7), 有

$$x^H D_0^{(1)} x = x^H R^H D_0^{(2)} Rx = - \sum_{i=1}^{n-t} |z_i|^2 \leq 0 \quad (5.1.9)$$

式 (5.1.8) 与式 (5.1.9) 矛盾. 因此 $t \geq s$.

同样可证, $s < t$ 是不可能的. 于是 $s = t$.

充分性. 如果 $\text{In}(A) = \text{In}(B)$, 则由 (5.1.5) 知, $\text{rank}(A) = \text{rank}(B) = r$. 由定理 5.1.5 知 A 和 B 相合于同一标准形. 由相合的对称性和传递性知 A 与 B 相合. \square

5.1.3 Hermite 二次型

线性代数中已介绍过实二次型, 现在把它推广到复数情况. 下面我们讨论具有重要应用的一类复二次型——Hermite 二次型. 与实二次型相仿, 我们研究 Hermite 二次型的标准形与分类等问题.

由 n 个复变量 x_1, \dots, x_n , 系数为复数的二次齐式

$$f(x_1, \dots, x_n) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \quad (5.1.10)$$

其中 $a_{ij} = \overline{a_{ji}}$, 称为 Hermite 二次型. 记

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

则 A 为 Hermite 矩阵. 我们称矩阵 A 为 Hermite 二次型的矩阵, 并且称 A 的秩为 Hermite 二次型的秩. 于是, Hermite 二次型 (5.1.10) 可改写为

$$f(x) = x^H A x \quad (5.1.11)$$

其中 $x = (x_1, \dots, x_n)^T$. 因此, 一个 Hermite 二次型与一个 Hermite 矩阵相对应.

若作可逆线性变换 $x = Py$, 其中 P 为 n 阶可逆矩阵, $y = (y_1, \dots, y_n)^T$, 则

$$f(x) = x^H A x = y^H B y$$

其中 $B = P^H A P$. 显然 B 是 Hermite 矩阵, 且与矩阵 A 相合.

Hermite 二次型中最简单的一种是只包含平方项的二次型

$$\lambda_1 \bar{y}_1 y_1 + \lambda_2 \bar{y}_2 y_2 + \dots + \lambda_n \bar{y}_n y_n \quad (5.1.12)$$

我们称形如(5.1.12)的二次型为 **Hermite 二次型的标准形**.

用 Hermite 二次型的语言, 定理 5.1.3 可以叙述为

定理 5.1.7 对 Hermite 二次型 $f(x) = x^H A x$, 存在酉线性变换 $x = Uy$ (其中 U 是酉矩阵) 使得 Hermite 二次型 $f(x)$ 变成标准形

$$\lambda_1 \bar{y}_1 y_1 + \lambda_2 \bar{y}_2 y_2 + \dots + \lambda_n \bar{y}_n y_n$$

其中 $\lambda_1, \dots, \lambda_n$ 是 Hermite 矩阵 A 的特征值.

定理 5.1.5 也可用 Hermite 二次型的语言来叙述.

定理 5.1.8 对 Hermite 二次型 $f(x) = x^H A x$, 存在可逆线性变换 $x = Py$ 使得 Hermite 二次型 $f(x)$ 化为

$$f(x) = x^H A x = y_1 y_1 + \dots + \bar{y}_s y_s - \bar{y}_{s+1} y_{s+1} - \dots - y_r y_r \quad (5.1.13)$$

其中 $r = \text{rank}(A)$, $s = \pi(A)$.

式(5.1.13)称为 Hermite 二次型 $f(x) = x^H A x$ 的规范形, 其中 s 和 $r - s$ 分别称为 Hermite 二次型的正惯性指数和负惯性指数.

因为 Hermite 二次型 $f(x) = x^H A x$ 的正惯性指数 s 与秩 r 之间满足 $0 \leq s \leq r \leq n$, 所以 Hermite 二次型可分为五种情况:

(1) 若 $s = r = n$, 则规范形为 $x^H A x = \sum_{i=1}^n |y_i|^2$. 若 $x \neq 0$, 则 $y \neq 0$, $x^H A x > 0$.

(2) 若 $s = r < n$, 则规范形为 $x^H A x = \sum_{i=1}^r |y_i|^2$. 对任意 $x \in C^n$ 都有 $x^H A x \geq 0$.

(3) 若 $s = 0, r = n$, 则规范形为 $x^H A x = -\sum_{i=1}^n |y_i|^2$. 若 $x \neq 0$, 则 $y \neq 0$, $x^H A x < 0$.

(4) 若 $s = 0, r < n$, 则规范形为 $x^H A x = -\sum_{i=1}^r |y_i|^2$. 对任意 $x \in C^n$ 都有

$x^H A x \leq 0$.

(5) 若 $0 < s < r \leq n$, 则规范形为 $x^H A x = \sum_{i=1}^s |y_i|^2 - \sum_{i=s+1}^r |y_i|^2$. 对不同的 x , $x^H A x$ 之值可以大于 0, 小于 0 或等于 0.

根据上面的讨论, 我们可以对 Hermite 二次型进行分类.

定义 5.1.1 设 $f(x) = x^H A x$ 为 Hermite 二次型.

- (1) 如果对任意 $x \in C^n$ 且 $x \neq 0$, 都有 $x^H A x > 0$, 则称 $x^H A x$ 为正定的;
- (2) 如果对任意 $x \in C^n$, 都有 $x^H A x \geq 0$, 则称 $x^H A x$ 为半正定(非负定)的;
- (3) 如果对任意 $x \in C^n$ 且 $x \neq 0$, 都有 $x^H A x < 0$, 则称 $x^H A x$ 为负定的;
- (4) 如果对任意 $x \in C^n$, 都有 $x^H A x \leq 0$, 则称 $x^H A x$ 为半负定的;
- (5) 对不同的 $x \in C^n$, $x^H A x$ 有时为正, 有时为负, 则称 $x^H A x$ 为不定的.

根据上面的讨论, 直接得如下定理.

定理 5.1.9 对 Hermite 二次型 $f(x) = x^H A x$, 有

- (1) $x^H A x$ 正定的充分必要条件为 $s = r = n$;
- (2) $x^H A x$ 半正定的充分必要条件为 $s = r < n$;
- (3) $x^H A x$ 负定的充分必要条件为 $s = 0, r = n$;
- (4) $x^H A x$ 半负定的充分必要条件为 $s = 0, r < n$;
- (5) $x^H A x$ 不定的充分必要条件为 $0 < s < r \leq n$.

§ 5.2 Hermite 正定(非负定)矩阵

利用 Hermite 二次型的正定(非负定)可以定义 Hermite 矩阵的正定(非负定).

定义 5.2.1 设 A 是 n 阶 Hermite 矩阵, 如果对任意 $x \in C^n$ 且 $x \neq 0$, 都有 $x^H A x > 0$, 则称 A 为正定矩阵, 记作 $A > 0$; 如果对任意 $x \in C^n$, 都有 $x^H A x \geq 0$, 则称 A 为非负定(半正定)矩阵, 记作 $A \geq 0$.

在定义 5.2.1 中, 矩阵 A 是 Hermite 矩阵的假定是必要的. 在以下的讨论中, 矩阵 A 均指 Hermite 矩阵. 由定义 5.2.1 可知, 如果 A 是正定矩阵, 则它也是半正定矩阵.

正定(非负定)矩阵具有如下基本性质:

- (1) 单位矩阵 $I > 0$;
- (2) 若 $A > 0$, 数 $k > 0$, 则 $kA > 0$;
- (3) 若 $A > 0, B > 0$, 则 $A + B > 0$;
- (4) 若 $A \geq 0, B \geq 0$, 则 $A + B \geq 0$.

下面我们给出 Hermite 矩阵 A 正定(非负定)的条件.

定理 5.2.1 设 A 是 n 阶 Hermite 矩阵, 则下列命题等价:

- (1) A 是正定矩阵;
- (2) 对任意 n 阶可逆矩阵 P , $P^H A P$ 都是 Hermite 正定矩阵;
- (3) A 的 n 个特征值均为正数;
- (4) 存在 n 阶可逆矩阵 P 使得 $P^H A P = I$;
- (5) 存在 n 阶可逆矩阵 Q 使得 $A = Q^H Q$;
- (6) 存在 n 阶可逆 Hermite 矩阵 S 使得 $A = S^2$.

证明 首先按(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (5) \Rightarrow (1)进行证明.

(1) \Rightarrow (2) 对任意 n 阶可逆矩阵 P 及任意 $y \in \mathbb{C}^n$ 且 $y \neq 0$, 令 $x = Py$, 则 $x \in \mathbb{C}^n$ 且 $x \neq 0$

$$y^H (P^H A P) y = x^H A x > 0$$

故 $P^H A P$ 是 Hermite 正定矩阵.

(2) \Rightarrow (3) 对 Hermite 矩阵 A , 由定理 5.1.3 知存在酉矩阵 U 使得

$$U^H A U = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) \quad (5.2.1)$$

其中 $\lambda_1, \lambda_2, \dots, \lambda_n$ 为 A 的特征值, 由(2)知 $\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ 是正定矩阵, 则 $\lambda_1, \lambda_2, \dots, \lambda_n$ 均为正数.

(3) \Rightarrow (4) 因为 A 的特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 均为正数, 令

$$P_1 = \text{diag}\left(\frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}, \dots, \frac{1}{\sqrt{\lambda_n}}\right), \quad P = U P_1$$

则 P 是可逆矩阵, 并且由(5.2.1)有 $P^H A P = I$.

(4) \Rightarrow (5) 因为存在 n 阶可逆矩阵 P 使得 $P^H A P = I$, 则令 $Q = P^{-1}$, 有 $A = Q^H Q$.

(5) \Rightarrow (1) 因为存在 n 阶可逆矩阵 Q 使得 $A = Q^H Q$, 则对任意 $x \in \mathbb{C}^n$ 且 $x \neq 0$ 都有 $Qx \neq 0$, 从而 $x^H A x = (Qx)^H (Qx) > 0$. 故 A 是正定矩阵.

下面证明(1) \Leftrightarrow (6).

(1) \Rightarrow (6) 设 λ 为 A 的任一特征值, x 为相应的特征向量, 则

$$Ax = \lambda x$$

因为 A 是正定矩阵, 所以 $\lambda x^H x = x^H A x > 0$, 从而 $\lambda > 0$. 因此 A 的特征值均为正数. 由(5.2.1)得

$$A = U \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) U^H$$

其中 $\lambda_1, \lambda_2, \dots, \lambda_n$ 为 A 的正特征值. 令

$$S = U \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n}) U^H$$

则 S 是 n 阶可逆 Hermite 矩阵, 并且 $A = S^2$.

(6) \Rightarrow (1) 因为存在 n 阶可逆 Hermite 矩阵 S 使得 $A = S^2 = S^H S$, 类似于 (5) \Rightarrow (1) 即知 A 是正定矩阵. \square

由定理 5.2.1 和定理 2.4.2 直接可得如下推论.

推论 5.2.1 设 A 是 n 阶 Hermite 正定矩阵, 其特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 则

- (1) A^{-1} 是正定矩阵;
- (2) 如果 Q 是任一 $n \times m$ 列满秩矩阵, 则 $Q^H A Q > 0$;
- (3) $|A| > 0$;
- (4) $\operatorname{tr}(A) > \lambda_i (i = 1, 2, \dots, n)$.

例 5.2.1 设 n 阶 Hermite 矩阵 A 有如下分块

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^H & A_{22} \end{bmatrix}, \quad A_{11} \in \mathbb{C}^{k \times k}$$

则 $A > 0$ 的充分必要条件是 $A_{11} > 0$ 和 $A_{22} - A_{12}^H A_{11}^{-1} A_{12} > 0$.

证明 如果 A_{11} 非奇异, 则

$$\begin{bmatrix} I_k & 0 \\ -A_{12}^H A_{11}^{-1} & I_{n-k} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^H & A_{22} \end{bmatrix} \begin{bmatrix} I_k & -A_{11}^{-1} A_{12} \\ 0 & I_{n-k} \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} - A_{12}^H A_{11}^{-1} A_{12} \end{bmatrix}$$

即 $\begin{bmatrix} A_{11} & A_{12} \\ A_{12}^H & A_{22} \end{bmatrix}$ 相合于 $\begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} - A_{12}^H A_{11}^{-1} A_{12} \end{bmatrix}$. 因此 $A > 0$ 的充分必要条件

是 $A_{11} > 0$ 和 $A_{22} - A_{12}^H A_{11}^{-1} A_{12} > 0$. \square

类似于定理 5.2.1, 有如下结论.

定理 5.2.2 设 A 是 n 阶 Hermite 矩阵, 则下列命题等价:

- (1) A 是非负定矩阵;
- (2) 对任意 n 阶可逆矩阵 P , $P^H A P$ 是 Hermite 非负定矩阵;
- (3) A 的 n 个特征值均为非负数;
- (4) 存在 n 阶可逆矩阵 P 使得 $P^H A P = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$, 其中 $r = \operatorname{rank}(A)$;
- (5) 存在秩为 r 的矩阵 Q 使得 $A = Q^H Q$;
- (6) 存在 n 阶 Hermite 矩阵 S 使得 $A = S^2$.

读者不难自行证明这个定理.

由定理 5.2.2 和定理 2.4.2 可得如下推论.

推论 5.2.2 设 A 是 n 阶 Hermite 非负定矩阵, 其特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 则

- (1) 如果 Q 是任一 $n \times m$ 矩阵, 则 $Q^H A Q \geq 0$;
 (2) $|A| \geq 0$;
 (3) $\operatorname{tr}(A) \geq \lambda_i (i = 1, 2, \dots, n)$.

定理 5.2.3 n 阶 Hermite 矩阵 A 正定的充分必要条件是 A 的顺序主子式均为正数, 即

$$\Delta_k = |A_{11} \dots A_{kk}| > 0, \quad k = 1, \dots, n$$

证明 必要性. 首先证明 Hermite 正定矩阵 A 的顺序主子矩阵也是正定矩阵. 记

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^H & A_{22} \end{pmatrix}$$

其中 A_{11} 是 A 的 k 阶顺序主子矩阵. 对任意 $x = \begin{pmatrix} y \\ 0 \end{pmatrix} \in C^n, y \in C^k$ 且 $y \neq 0$, 则

$$0 < x^H A x = y^H A_{11} y$$

故 A_{11} 是 k 阶 Hermite 正定矩阵.

因为 A 的顺序主子矩阵都是正定矩阵, 由推论 5.2.1 知 A 的顺序主子式均为正数.

充分性. 对矩阵的阶数作归纳法. 阶数为 1 时结论显然成立. 今设阶数为 $n-1$ 时结论成立. 对 n 阶 Hermite 矩阵 A , 记

$$A = \begin{pmatrix} A_{n-1} & a \\ a^H & a_{nn} \end{pmatrix}$$

其中 A_{n-1} 为 A 的 $n-1$ 阶顺序主子矩阵. 因为 A_{n-1} 非奇异, 令

$$P = \begin{pmatrix} I_{n-1} & -A_{n-1}^{-1}a \\ 0 & 1 \end{pmatrix}$$

则

$$P^H A P = \begin{pmatrix} A_{n-1} & 0 \\ 0 & a_{nn} - a^H A_{n-1}^{-1} a \end{pmatrix} \quad (5.2.2)$$

于是 $a_{nn} - a^H A_{n-1}^{-1} a = \frac{|A|}{|A_{n-1}|} = \frac{\Delta_n}{\Delta_{n-1}} > 0$. 由归纳法假设 $A_{n-1} > 0$, 则

$$\begin{vmatrix} \mathbf{A}_{n-1} & 0 \\ 0 & a_{nn} - a^H \mathbf{A}_{n-1}^{-1} a \end{vmatrix} > 0$$

由定理 5.2.1 和(5.2.2)可知 $\mathbf{A} > 0$. 这说明阶数为 n 时结论也成立. \square

必须指出, Hermite 矩阵 \mathbf{A} 的所有顺序主子式均非负, 并不能推出 \mathbf{A} 是

非负定矩阵. 例如, 矩阵 $\mathbf{A} = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$ 是 Hermite 矩阵. 容易验证 \mathbf{A}

的所有顺序主子式均非负, 但 \mathbf{A} 不是非负定矩阵.

定理 5.2.4 n 阶 Hermite 矩阵 \mathbf{A} 正定的充分必要条件是 \mathbf{A} 的所有主子式全大于零.

证明 必要性. 对 \mathbf{A} 的任一 k 阶主子式

$$\mathbf{A} \begin{pmatrix} i_1 & \cdots & i_k \\ i_1 & \cdots & i_k \end{pmatrix} = \begin{vmatrix} a_{i_1 i_1} & a_{i_1 i_2} & \cdots & a_{i_1 i_k} \\ a_{i_2 i_1} & a_{i_2 i_2} & \cdots & a_{i_2 i_k} \\ \vdots & \vdots & & \vdots \\ a_{i_k i_1} & a_{i_k i_2} & \cdots & a_{i_k i_k} \end{vmatrix}$$

存在某个排列矩阵 \mathbf{P} , 使 $\mathbf{P}^H \mathbf{A} \mathbf{P}$ 的 k 阶顺序主子式为 $\mathbf{A} \begin{pmatrix} i_1 & \cdots & i_k \\ i_1 & \cdots & i_k \end{pmatrix}$. 因为 $\mathbf{A} >$

0 , 由定理 5.2.1 知 $\mathbf{P}^H \mathbf{A} \mathbf{P} > 0$, 从而由定理 5.2.3 有 $\mathbf{A} \begin{pmatrix} i_1 & \cdots & i_k \\ i_1 & \cdots & i_k \end{pmatrix} > 0$.

充分性由定理 5.2.3 即得. \square

定理 5.2.5 n 阶 Hermite 矩阵 \mathbf{A} 非负定的充分必要条件是 \mathbf{A} 的所有主子式均非负.

证明 必要性. 因为 Hermite 非负定矩阵 \mathbf{A} 的顺序主子矩阵都是非负定矩阵, 并且由推论 5.2.2 知, 非负定矩阵 \mathbf{A} 的顺序主子式均非负. 类似于定理 5.2.4 的必要性证明可知, \mathbf{A} 的所有主子式均非负.

充分性. 对任意 $\varepsilon > 0$, 由定理 2.4.1, 有

$$|\varepsilon \mathbf{I} + \mathbf{A}| = \varepsilon^n + \sum_{k=1}^n b_k \varepsilon^{n-k}$$

其中 $b_k (k=1, 2, \cdots, n)$ 是 \mathbf{A} 的所有 k 阶主子式之和, 则 $|\varepsilon \mathbf{I} + \mathbf{A}| > 0$. 类似可证, $\varepsilon \mathbf{I} + \mathbf{A}$ 所有顺序主子式均为正数. 因此 $\varepsilon \mathbf{I} + \mathbf{A}$ 是正定矩阵. 从而对任意 $x \in \mathbb{C}^n$ 且 $x \neq 0$ 都有

$$x^H(\epsilon I + A)x > 0$$

在上式令 $\epsilon \rightarrow 0$, 则得

$$x^H Ax \geq 0$$

故 A 是非负定矩阵. \square

例 5.2.2 若 A, C 均为 n 阶 Hermite 正定矩阵, 且 $AC = CA$, 则 AC 为正定矩阵.

证明 因为 $(AC)^H = C^H A^H = CA = AC$, 所以 AC 是 Hermite 矩阵. 又因为 $A > 0$, 由定理 5.2.1 知, 存在 n 阶可逆 Hermite 矩阵 B 使 $A = B^2$. 于是

$$B^{-1}(AC)B = BCB = B^H CB$$

则 AC 与 $B^H CB$ 具有相同的特征值. 由 $C > 0$ 及定理 5.2.1 知 $B^H CB > 0$, 故 $B^H CB$ 的特征值均为正数, 从而 AC 的特征值均为正数. 由定理 5.2.1 知 $AC > 0$. \square

同样可以证明: 若 $A > 0, C \geq 0$, 且 $AC = CA$, 则 $AC \geq 0$. 进一步地, 若 $A \geq 0, C \geq 0$, 且 $AC = CA$, 则 $AC \geq 0$.

例 5.2.3 若 $A = (a_{ij})$ 是 n 阶 Hermite 正定矩阵, 则

$$|A| \leq a_{nn} |A_{n-1}|$$

其中 A_{n-1} 为 A 的 $n-1$ 阶顺序主子矩阵, 且等号成立当且仅当 $a_{1n} = a_{2n} = \cdots = a_{n-1,n} = 0$.

证明 记

$$A = \begin{pmatrix} A_{n-1} & a \\ a^H & a_{nn} \end{pmatrix}$$

其中 A_{n-1} 为 A 的 $n-1$ 阶顺序主子矩阵, $a = (a_{1n}, a_{2n}, \cdots, a_{n-1,n})^T$. 因为 A_{n-1} 是正定矩阵, 则 A_{n-1} 可逆, 并且 A_{n-1}^{-1} 也是正定矩阵. 由于

$$\begin{pmatrix} I_{n-1} & 0 \\ -a^H A_{n-1}^{-1} & 1 \end{pmatrix} \begin{pmatrix} A_{n-1} & a \\ a^H & a_{nn} \end{pmatrix} = \begin{pmatrix} A_{n-1} & a \\ 0 & a_{nn} - a^H A_{n-1}^{-1} a \end{pmatrix}$$

故

$$|A| = (a_{nn} - a^H A_{n-1}^{-1} a) |A_{n-1}|$$

而 $a^H A_{n-1}^{-1} a \geq 0$, 当且仅当 $a = 0$ 时等号成立, 所以

$$|A| \leq a_{nn} |A_{n-1}|$$

当且仅当 $a_{1n} = a_{2n} = \cdots = a_{n-1,n} = 0$ 时等号成立. \square

对 n 阶 Hermite 正定矩阵 $A = (a_{ij})$, 反复应用例 5.2.3 的结果, 有

$$|A| \leq \prod_{i=1}^n a_{ii}$$

这就是著名的 Hadamard 不等式.

对 n 阶 Hermite 非负定矩阵 $A = (a_{ij})$, 上述 Hadamard 不等式仍成立.

事实上, 若 $A \geq 0$, 则对充分小的正数 ε , $A + \varepsilon I > 0$. 从而有

$$|A + \varepsilon I| \leq \prod_{i=1}^n (a_{ii} + \varepsilon)$$

令 $\varepsilon \rightarrow 0$, 即得 $|A| \leq \prod_{i=1}^n a_{ii}$.

定理 5.2.6 n 阶 Hermite 矩阵 A 正定的充分必要条件是存在 n 阶非奇异下三角矩阵 L 使得

$$A = LL^H \quad (5.2.3)$$

证明 必要性. 若 A 是 n 阶正定矩阵, 由定理 5.2.3 知 A 的顺序主子式全大于零. 由定理 4.3.2 知, A 有惟一的 LDU 分解

$$A = L_1 D U_1$$

其中 L_1, U_1 分别为单位下三角矩阵和单位上三角矩阵, $D = \text{diag}(d_1, d_2, \cdots, d_n)$ 且 $d_i > 0 (i = 1, 2, \cdots, n)$. 因为 $A^H = A$, 则

$$A = L_1 D U_1 = U_1^H D^H L_1^H = A^H$$

由 LDU 分解的惟一性, 有 $L_1 = U_1^H$. 从而有

$$A = L_1 D L_1^H \quad (5.2.4)$$

令

$$L = L_1 \text{diag}(\sqrt{d_1}, \cdots, \sqrt{d_n})$$

则 L 是非奇异下三角矩阵, 并且 $A = LL^H$.

充分性由定理 5.2.1 即得. \square

分解式(5.2.3)称为正定矩阵 A 的 Cholesky 分解, 它不仅应用于正定矩阵线性方程组 $Ax = b$ 的求解, 而且可以应用于解决如下的广义特征值问题.

定义 5.2.2 设 $A, B \in \mathbb{C}^{n \times n}$, 如果存在复数 λ 和非零向量 $x \in \mathbb{C}^n$ 使得

$$Ax = \lambda Bx \quad (5.2.5)$$

则称 λ 为广义特征值问题 $Ax = \lambda Bx$ 的特征值, 非零向量 x 称为对应于特征值 λ 的特征向量.

如果 B 是 n 阶非奇异矩阵, 则广义特征值问题(5.2.5)可化为如下标准特征值问题

$$B^{-1}Ax = \lambda x \quad (5.2.6)$$

在振动理论等应用问题中, A, B 均为 n 阶 Hermite 矩阵, 且 B 是正定矩阵. 在这种情况下, 广义特征值问题(5.2.5)可化为 Hermite 矩阵的特征值问题.

事实上, 因为 B 是正定矩阵, 则 B 有 Cholesky 分解

$$B = LL^H \quad (5.2.7)$$

令

$$\hat{A} = L^{-1}A(L^{-1})^H, \quad y = L^Hx$$

则(5.2.5)化为

$$\hat{A}y = \lambda y \quad (5.2.8)$$

其中 \hat{A} 是 Hermite 矩阵. 这样, 广义特征值问题 $Ax = \lambda Bx$ 就等价于 Hermite 矩阵 \hat{A} 的标准特征值问题, 而这两个问题的特征向量之间的关系为 $y = L^Hx$.

定理 5.2.7 设 A, B 均为 n 阶 Hermite 矩阵, 且 $B > 0$, 则存在非奇异矩阵 P 使得

$$P^HAP = \text{diag}(\lambda_1, \dots, \lambda_n), \quad P^HB P = I \quad (5.2.9)$$

其中 $\lambda_1, \dots, \lambda_n$ 是广义特征值问题(5.2.5)的特征值.

证明 因为 $B > 0$, 由定理 5.2.1, 存在非奇异矩阵 P_1 使得 $P_1^HB P_1 = I$, 而 $P_1^HAP_1$ 仍为 Hermite 矩阵, 则有酉矩阵 U 使

$$U^H(P_1^HAP_1)U = \text{diag}(\lambda_1, \dots, \lambda_n)$$

令 $P = P_1U$, 则 P 非奇异, 并且

$$P^HB P = U^HP_1^HB P_1U = U^HU = I, \quad P^HAP = U^HP_1^HAP_1U = \text{diag}(\lambda_1, \dots, \lambda_n)$$

由上式得

$$P^{-1}B^{-1}AP = \text{diag}(\lambda_1, \dots, \lambda_n)$$

即 $B^{-1}A$ 相似于 $\text{diag}(\lambda_1, \dots, \lambda_n)$, 则 $\lambda_1, \dots, \lambda_n$ 是矩阵 $B^{-1}A$ 的特征值, 即 $\lambda_1, \dots, \lambda_n$ 是广义特征值问题(5.2.5)的特征值. \square

§ 5.3 矩阵不等式

利用 Hermite 矩阵的非负定性可以在 Hermite 矩阵类中引进偏序,从而可讨论矩阵不等式. 当矩阵的阶数为 1 时,此偏序与实数的序是一致的,故矩阵不等式是数值不等式的推广,并且矩阵不等式在现代控制理论等学科中具有重要应用.

定义 5.3.1 设 A, B 都是 n 阶 Hermite 矩阵,如果 $A - B \geq 0$,则称 A 大于或等于 B (或称 B 小于或等于 A),记作 $A \geq B$ (或 $B \leq A$);如果 $A - B > 0$,则称 A 大于 B (或称 B 小于 A),记作 $A > B$ (或 $B < A$).

由定义 5.3.1 知,对 n 阶 Hermite 矩阵 A 与 B ,如果 $A \geq B$,则对任意 $x \in C^n$ 都有

$$x^H A x \geq x^H B x$$

反之亦然.

必须指出:(1)任意两个实数总可以比较大小.但任意两个 n 阶 Hermite 矩阵未必能“比较大小”,即并非 $A \geq B$ 或 $B \geq A$ 两者之中必有一成立.例如,对

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & 0 \\ 0 & 4 \end{bmatrix}$$

$A \geq B$ 和 $B \geq A$ 均不成立.

(2) 对任意两个实数 a 和 b ,如果 $a \geq b$,而 $a \not\geq b$,则有 $a = b$.但对两上 n ($n \geq 2$) 阶 Hermite 矩阵 A 与 B ,从 $A \geq B$ 和 $A \not\geq B$,不能推出 $A = B$.例如,对

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

可见 $A \geq B$,且 $A \not\geq B$,但 $A \neq B$.

因此,定义 5.3.1 给出的“ \geq ”是 Hermite 矩阵类中的一种偏序,本节中涉及的矩阵不等式均按此偏序理解.

例 5.3.1 设 $A = \text{diag}(a_1, a_2, \dots, a_n)$, $B = \text{diag}(b_1, b_2, \dots, b_n)$ 均为实对角矩阵,则由定义 5.3.1 可得: $A \geq B$ ($A > B$) 的充分必要条件是 $a_i \geq b_i$ ($a_i > b_i$) ($i = 1, 2, \dots, n$).

下面讨论矩阵不等式的一些性质.

定理 5.3.1 设 A, A_1, B, B_1, C 均为 n 阶 Hermite 矩阵,则

(1) $A \geq B$ ($A > B$) 的充分必要条件是 $-A \leq -B$ ($-A < -B$);

(2) $A \geq B (A > B)$ 的充分必要条件是对任意 n 阶可逆矩阵 P 都有

$$P^H A P \geq P^H B P (P^H A P > P^H B P)$$

(3) 若 $A \geq B (A > B)$, k 为正数, 则 $kA \geq kB (kA > kB)$;

(4) 若 $A \geq 0, -A \geq 0$, 则 $A = 0$;

(5) 若 $A \geq 0, B \geq 0$, 则 $A + B \geq 0$;

(6) 若 $A \geq B, B \geq C$, 则 $A \geq C$;

(7) 若 $A \geq B, A_1 \geq B_1$, 则 $A + A_1 \geq B + B_1$;

(8) 若 $A \geq 0, B > 0$, 则 $A + B > 0$;

(9) 若 $A \geq B, B > C$, 则 $A > C$;

(10) 若 $A > B$, P 为 $n \times m$ 列满秩矩阵, 则 $P^H A P > P^H B P$;

(11) 若 $A \geq B$, P 为 $n \times m$ 矩阵, 则 $P^H A P \geq P^H B P$;

(12) 若 $A > 0 (A \geq 0)$, $C > 0 (C \geq 0)$, 且 $AC = CA$, 则 $AC > 0 (AC \geq 0)$.

证明 (1)和(3)显然成立.

(2)由定义 5.3.1, 定理 5.2.1 和定理 5.2.2 即知结论成立.

(4)若 $A \geq 0$, 由定理 5.2.2 知, A 的特征值均非负. 而 $-A$ 的特征值恰为 A 的特征值的相反数, 则由 $-A \geq 0$ 可知, A 的特征值又必须全部非正. 因此 Hermite 矩阵 A 的特征值全部为零. 由定理 5.1.3 知 $A = 0$.

(5)和(8)由正定(非负定)矩阵的基本性质即得.

(6)和(7)是(5)的直接推论; 而(9)是(8)的直接推论.

(10)和(11)可分别由推论 5.2.1 和推论 5.2.2 导出.

(12)由例 5.2.2 即得. \square

定理 5.3.2 设 A, B 均为 n 阶 Hermite 矩阵, 且 $A \geq 0, B > 0$, 则

(1) $B \geq A$ 的充分必要条件是 $\rho(AB^{-1}) \leq 1$;

(2) $B > A$ 的充分必要条件是 $\rho(AB^{-1}) < 1$.

证明 (1) 由定理 5.2.7 知, 有非奇异矩阵 P 值得

$$P^H B P = I, P^H A P = A = \text{diag}(\lambda_1, \dots, \lambda_n)$$

由定理 5.3.1 知, $B \geq A$ 等价于 $I \geq A$, 而后者等价于 $\lambda_i \leq 1 (i = 1, \dots, n)$. 因为 $\lambda_1, \dots, \lambda_n$ 是矩阵 AB^{-1} 的全部特征值, 所以 $\lambda_i \leq 1 (i = 1, \dots, n)$ 等价于 $\rho(AB^{-1}) \leq 1$.

类似地可证明(2). \square

定理 5.3.3 设 A 是 n 阶 Hermite 矩阵, 则 $\lambda_{\min}(A)I \leq A \leq \lambda_{\max}(A)I$, 这时 $\lambda_{\max}(A)$ 和 $\lambda_{\min}(A)$ 分别表示 A 的最大和最小特征值.

证明 由定理 5.1.3 知, 存在酉矩阵 U 使得

$$A = U \text{diag}(\lambda_1, \dots, \lambda_n) U^H$$

其中 $\lambda_1 \geq \cdots \geq \lambda_n$ 是矩阵 A 的特征值. 则 $\lambda_{\max}(A) = \lambda_1, \lambda_{\max}(A) - \lambda_i \geq 0 (i = 1, 2, \cdots, n), \lambda_{\min}(A) = \lambda_n, \lambda_i - \lambda_{\min}(A) \geq 0 (i = 1, 2, \cdots, n)$, 并且

$$\lambda_{\max}(A)I - A = U[\lambda_{\max}(A)I - \text{diag}(\lambda_1, \cdots, \lambda_n)]U^H \geq 0$$

$$A - \lambda_{\min}(A)I = U[\text{diag}(\lambda_1, \cdots, \lambda_n) - \lambda_{\min}(A)I]U^H \geq 0$$

即 $\lambda_{\max}(A)I \geq A, A \geq \lambda_{\min}(A)I$. \square

如果 A 是 Hermite 非负定矩阵, 则由推论 5.2.2, 定理 5.3.1 和定理 5.3.3 可得

推论 5.3.1 设 A 是 Hermite 非负定矩阵, 则 $A \leq \text{tr}(A)I$.

定理 5.3.4 设 A, B 均为 n 阶 Hermite 正定矩阵, 则

(1) 若 $A \geq B > 0$, 则 $B^{-1} \geq A^{-1} > 0$;

(2) 若 $A > B > 0$, 则 $B^{-1} > A^{-1} > 0$.

证明 由定理 5.2.7 知, 有非奇异矩阵 P 使得

$$P^H B P = I, P^H A P = A = \text{diag}(\lambda_1, \cdots, \lambda_n)$$

由 $A \geq B$ 和定理 5.3.1 有 $A \geq I$, 则 $\lambda_i \geq 1 (i = 1, \cdots, n)$. 从而

$$0 < A^{-1} = \text{diag}(\lambda_1^{-1}, \cdots, \lambda_n^{-1}) \leq I$$

因为 $A^{-1} = P A^{-1} P^H, B^{-1} = P I P^H$, 再利用定理 5.3.1, 便得 $B^{-1} \geq A^{-1} > 0$.

将上面证明中的 \geq 号改为 $>$ 号, 可得 $B^{-1} > A^{-1} > 0$. \square

一般说来, 不能从 $A \geq B > 0$ 推出 $A^2 \geq B^2$. 例如

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix}, B = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

容易验证: $A > 0, B > 0$ 且 $A \geq B$. 但

$$A^2 - B^2 = \begin{bmatrix} 6 & 7 \\ 7 & 8 \end{bmatrix}$$

不是非负定的, 这说明 $A^2 \geq B^2$ 不成立. 然而, 我们有如下结果.

定理 5.3.5 设 A, B 均为 n 阶 Hermite 正定矩阵, 且 $AB = BA$, 则

(1) 若 $A \geq B$, 则 $A^2 \geq B^2$;

(2) 若 $A > B$, 则 $A^2 > B^2$.

证明 以(1)为例, (2)的证明类似. 由 $AB = BA$, 则

$$A^2 - B^2 = (A - B)(A + B) = (A + B)(A - B)$$

因为 $A - B \geq 0, A + B > 0$, 则由定理 5.3.1(12)知 $A^2 - B^2 \geq 0$, 即 $A^2 \geq B^2$. \square

注意:如果定理 5.3.5 中“ A, B 正定”的条件改成“ A, B 非负定”,则定理 5.3.5(1)仍成立.

事实上,当 $A \geq 0, B \geq 0, A \geq B, AB = BA$ 时,对任意正数 ϵ ,有

$$\begin{aligned} A + \epsilon I &> 0, B + \epsilon I > 0, A + \epsilon I \geq B + \epsilon I, (A + \epsilon I)(B + \epsilon I) \\ &= (B + \epsilon I)(A + \epsilon I) \end{aligned}$$

由定理 5.3.5 有 $(A + \epsilon I)^2 \geq (B + \epsilon I)^2$. 因此,对任意 n 维向量 x 都有

$$x^H(A + \epsilon I)^2 x \geq x^H(B + \epsilon I)^2 x$$

令 $\epsilon \rightarrow 0$,有 $x^H A^2 x \geq x^H B^2 x$,即 $A^2 \geq B^2$.

定理 5.3.6 设 A 是 $m \times n$ 行满秩矩阵, B 是 $n \times k$ 矩阵,则

$$B^H B \geq (AB)^H (AA^H)^{-1} (AB) \quad (5.3.1)$$

其中等号成立的充分必要条件是存在一个 $m \times k$ 矩阵 C 使得 $B = A^H C$.

证明 因为

$$\begin{aligned} 0 &\leq [B - A^H(AA^H)^{-1}AB]^H [B - A^H(AA^H)^{-1}AB] \\ &= B^H B - 2B^H A^H(AA^H)^{-1}AB + B^H A^H(AA^H)^{-1}AA^H(AA^H)^{-1}AB \\ &= B^H B - (AB)^H(AA^H)^{-1}AB \end{aligned}$$

则得(3.3.1),并且上式右端为零矩阵的充分必要条件是

$$B = A^H(AA^H)^{-1}AB$$

如果这个条件满足,则令 $C = (AA^H)^{-1}AB$,有 $B = A^H C$;反过来,如果存在一个 $m \times k$ 矩阵 C 使得 $B = A^H C$,则

$$A^H(AA^H)^{-1}AB = A^H(AA^H)^{-1}AA^H C = A^H C = B$$

于是定理证毕. \square

* § 5.4 Hermite 矩阵的特征值

我们首先介绍 Rayleigh 商的概念,讨论其性质,然后运用 Rayleigh 商研究 Hermite 矩阵的特征值.

定义 5.4.1 设 A 为 n 阶 Hermite 矩阵,对任意 $x \in C^n$ 且 $x \neq 0$,称

$$R(x) = \frac{x^H A x}{x^H x}, \quad x \neq 0$$

为 Hermite 矩阵 A 的 Rayleigh 商.

显然, Hermite 矩阵的 Rayleigh 商之值是实数, 并且具有如下性质.

定理 5.4.1 设 A 是 n 阶 Hermite 矩阵, 其特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, 则

- (1) $R(kx) = R(x), k \in \mathbb{C}, k \neq 0$;
- (2) $\lambda_n \leq R(x) \leq \lambda_1, x \neq 0$;
- (3) $\lambda_1 = \max_{x \neq 0} R(x), \lambda_n = \min_{x \neq 0} R(x)$.

(5.4.1)

证明 (1) 由 Rayleigh 商的定义即得. (2) 由定理 5.3.3 即得.

(3) 设 A 对应于特征值 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 的标准正交特征向量为 x_1, x_2, \cdots, x_n . 显然, $R(x_1) = \lambda_1, R(x_n) = \lambda_n$. 因此由(2)可知,

$$\max_{x \neq 0} R(x) = \lambda_1, \quad \min_{x \neq 0} R(x) = \lambda_n \quad \square$$

可以推广定理 5.4.1 而得到如下定理.

定理 5.4.2 设 A 是 n 阶 Hermite 矩阵, 其特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, 相应的标准正交特征向量为 x_1, x_2, \cdots, x_n . 记 $V_i^{(j)} = \text{span}\{x_i, x_{i+1}, \cdots, x_j\}$ ($i \leq j$), 则

$$\lambda_i = \max_{\substack{x \in V_i^{(j)} \\ x \neq 0}} R(x), \quad \lambda_j = \min_{\substack{x \in V_i^{(j)} \\ x \neq 0}} R(x) \quad (5.4.2)$$

证明 对任意 $x \in V_i^{(j)}$ 且 $x \neq 0$, 有 $x = \alpha_i x_i + \alpha_{i+1} x_{i+1} + \cdots + \alpha_j x_j$, 从而

$$\lambda_j \leq R(x) = \frac{|\alpha_i|^2 \lambda_i + |\alpha_{i+1}|^2 \lambda_{i+1} + \cdots + |\alpha_j|^2 \lambda_j}{|\alpha_i|^2 + |\alpha_{i+1}|^2 + \cdots + |\alpha_j|^2} \leq \lambda_i$$

特别地, 若 $x = x_i$, 则 $R(x) = \lambda_i$; 而 $x = x_j$ 时, 有 $R(x) = \lambda_j$. 因此

$$\lambda_i = \max_{\substack{x \in V_i^{(j)} \\ x \neq 0}} R(x), \quad \lambda_j = \min_{\substack{x \in V_i^{(j)} \\ x \neq 0}} R(x) \quad \square$$

一般地, 有如下著名的极大极小或极小极大定理.

定理 5.4.3 设 A 是 n 阶 Hermite 矩阵, 其特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, V_i 是 \mathbb{C}^n 中 i 维子空间, 则

$$\lambda_i = \max_{V_i} \min_{\substack{x \in V_i \\ x \neq 0}} R(x) \quad (5.4.3)$$

和

$$\lambda_i = \min_{V_{n-i+1}} \max_{\substack{x \in V_{n-i+1} \\ x \neq 0}} R(x) \quad (5.4.4)$$

其中 $i = 1, 2, \cdots, n$.

证明 这里仅证明(5.4.3), (5.4.4)的证明是类似的.

考虑 C^n 的任一 i 维子空间 V_i , 令 G_i 是以子空间 V_i 的一组标准正交基为列向量形成的 $n \times i$ 矩阵, 则

$$G_i^H G_i = I_i$$

设 x_1, \dots, x_n 是 A 对应于特征值 $\lambda_1, \dots, \lambda_n$ 的标准正交特征向量. 记

$$U = [x_1, \dots, x_i, x_{i+1}, \dots, x_n] \equiv [U_i, U_{n-i}]$$

其中 $U_i = [x_1, \dots, x_i]$, $U_{n-i} = [x_{i+1}, \dots, x_n]$. 由定理 1.4.7 知, i 维子空间 V_i 与由 $[x_i, U_{n-i}]$ 的列向量所生成的 $n-i+1$ 维子空间 $\text{span}[x_i, U_{n-i}]$ 必有公共的非零元, 设 \tilde{x} 就是这样一个元素. 因为 $\tilde{x} \in \text{span}[x_i, U_{n-i}]$, 则

$$\tilde{x} = \alpha_i x_i + \alpha_{i+1} x_{i+1} + \dots + \alpha_n x_n$$

从而

$$R(\tilde{x}) = \frac{|\alpha_i|^2 \lambda_i + |\alpha_{i+1}|^2 \lambda_{i+1} + \dots + |\alpha_n|^2 \lambda_n}{|\alpha_i|^2 + |\alpha_{i+1}|^2 + \dots + |\alpha_n|^2} \leq \lambda_i$$

又因为 $\tilde{x} \in V_i$, 所以对任取的 i 维子空间 V_i 更成立

$$\min_{\substack{x \in V_i \\ x \neq 0}} R(x) \leq \lambda_i$$

另一方面, 若取 $G_i = U_i$, 则

$$\min_{\substack{x \in V_i \\ x \neq 0}} R(x) \geq \lambda_i$$

因此

$$\lambda_i = \max_{V_i} \min_{\substack{x \in V_i \\ x \neq 0}} R(x) \quad \square$$

应用定理 5.4.3, 可以证明下述分隔定理.

定理 5.4.4 设 A 为 n 阶 Hermite 矩阵, $U_{n-1} \in C^{n \times (n-1)}$ 且 $U_{n-1}^H U_{n-1} = I_{n-1}$, 令

$$A' = U_{n-1}^H A U_{n-1}$$

A 与 A' 的特征值分别为 $\lambda_1(A) \geq \dots \geq \lambda_n(A)$ 与 $\lambda_1(A') \geq \dots \geq \lambda_{n-1}(A')$. 则

$$\lambda_n(A) \leq \lambda_{n-1}(A') \leq \lambda_{n-1}(A) \leq \dots \leq \lambda_2(A) \leq \lambda_1(A') \leq \lambda_1(A) \quad (5.4.5)$$

证明 由定理 5.4.3, 有

$$\lambda_i(\mathbf{A}') = \max_{V_i} \min_{\substack{x \in V_i \\ x \neq 0}} \frac{x^H \mathbf{A}' x}{x^H x} \quad (5.4.6)$$

$$= \min_{V_{(n-1)-i+1}} \max_{\substack{x \in V_{(n-1)-i+1} \\ x \neq 0}} \frac{x^H \mathbf{A}' x}{x^H x} \quad (5.4.7)$$

其中 V_i 是 \mathbf{C}^{n-1} 中 i 维子空间.

设 \tilde{V}_i 是使等式(5.4.6)成立的一个子空间, 于是

$$\begin{aligned} \lambda_i(\mathbf{A}') &= \min_{\substack{x \in V_i \\ x \neq 0}} \frac{x^H \mathbf{A}' x}{x^H x} = \min_{\substack{x \in V_i \\ x \neq 0}} \frac{(\mathbf{U}_{n-1} x)^H \mathbf{A} (\mathbf{U}_{n-1} x)}{(\mathbf{U}_{n-1} x)^H (\mathbf{U}_{n-1} x)} = \min_{\substack{y \in \mathbf{U}_{n-1} \tilde{V}_i \\ y \neq 0}} \frac{y^H \mathbf{A} y}{y^H y} \\ &\leq \max_{V_i^{(n)}} \min_{\substack{y \in V_i^{(n)} \\ y \neq 0}} \frac{y^H \mathbf{A} y}{y^H y} = \lambda_i(\mathbf{A}), \quad i = 1, \dots, n-1 \end{aligned} \quad (5.4.8)$$

其中 $V_i^{(n)}$ 是 \mathbf{C}^n 中 i 维子空间.

另一方面, 设 $\tilde{V}_{(n-1)-i+1}$ 是使等式(5.4.7)成立的一个子空间, 于是

$$\begin{aligned} \lambda_i(\mathbf{A}') &= \max_{\substack{x \in \tilde{V}_{(n-1)-i+1} \\ x \neq 0}} \frac{x^H \mathbf{A}' x}{x^H x} = \max_{\substack{x \in \tilde{V}_{(n-1)-i+1} \\ x \neq 0}} \frac{(\mathbf{U}_{n-1} x)^H \mathbf{A} (\mathbf{U}_{n-1} x)}{(\mathbf{U}_{n-1} x)^H (\mathbf{U}_{n-1} x)} \\ &= \max_{\substack{y \in \mathbf{U}_{n-1} \tilde{V}_{(n-1)-i+1} \\ y \neq 0}} \frac{y^H \mathbf{A} y}{y^H y} \geq \min_{V_{(n-1)-i+1}^{(n)}} \max_{\substack{y \in V_{(n-1)-i+1}^{(n)} \\ y \neq 0}} \frac{y^H \mathbf{A} y}{y^H y} \\ &= \lambda_{i+1}(\mathbf{A}), \quad i = 1, \dots, n-1 \end{aligned} \quad (5.4.9)$$

由(5.4.8)和(5.4.9)即得(5.4.5). \square

如果在定理 5.4.4 中 $\mathbf{U}_{n-1} = [e_1, e_2, \dots, e_{n-1}]$, 其中 $e_i (i = 1, 2, \dots, n-1)$ 是 n 阶单位矩阵的第 i 列, 则 $\mathbf{U}_{n-1}^H \mathbf{A} \mathbf{U}_{n-1} = \mathbf{A}_{n-1}$ 是 Hermite 矩阵 \mathbf{A} 的 $n-1$ 阶顺序主子矩阵. 若 \mathbf{A}_{n-1} 的特征值记为 $\mu_1 \geq \mu_2 \geq \dots \geq \mu_{n-1}$, 则由定理 5.4.4 有

$$\lambda_n(\mathbf{A}) \leq \mu_{n-1} \leq \lambda_{n-1}(\mathbf{A}) \leq \dots \leq \lambda_2(\mathbf{A}) \leq \mu_1 \leq \lambda_1(\mathbf{A}) \quad (5.4.10)$$

最后, 我们应用极大极小定理研究 Hermite 矩阵特征值的扰动问题, 即讨论 Hermite 矩阵的元素发生微小变化时相应矩阵特征值的变化范围.

定理 5.4.5 设 \mathbf{A}, \mathbf{E} 均为 n 阶 Hermite 矩阵, $\mathbf{B} = \mathbf{A} + \mathbf{E}$, 且 \mathbf{A}, \mathbf{B} 和 \mathbf{E} 的特征值分别为 $\lambda_1 \geq \dots \geq \lambda_n, \mu_1 \geq \dots \geq \mu_n$ 和 $\varepsilon_1 \geq \dots \geq \varepsilon_n$, 则

$$\lambda_i + \varepsilon_n \leq \mu_i \leq \lambda_i + \varepsilon_1, \quad i = 1, 2, \dots, n \quad (5.4.11)$$

证明 设 x_1, \dots, x_n 是 A 的对应于特征值 $\lambda_1, \dots, \lambda_n$ 的标准正交特征向量. 令

$$V_{n-i+1} = \text{span}\{x_i, x_{i+1}, \dots, x_n\}$$

则由(5.4.4)得

$$\mu_i \leq \max_{\substack{x \in V_{n-i+1} \\ x \neq 0}} \frac{x^H B x}{x^H x} \leq \max_{\substack{x \in V_{n-i+1} \\ x \neq 0}} \frac{x^H A x}{x^H x} + \max_{\substack{x \in V_{n-i+1} \\ x \neq 0}} \frac{x^H E x}{x^H x}$$

由定理 5.4.1 和定理 5.4.2, 有

$$\mu_i \leq \lambda_i + \max_{\substack{x \in V_{n-i+1} \\ x \neq 0}} \frac{x^H E x}{x^H x} \leq \lambda_i + \epsilon_1, \quad i = 1, 2, \dots, n \quad (5.4.12)$$

另一方面, 令 $D = -E$, 记 D 的特征值为 $\delta_1 \geq \dots \geq \delta_n$, 则 $A = B + D$. 由(5.4.12)有

$$\lambda_i \leq \mu_i + \delta_1, \quad i = 1, 2, \dots, n$$

其中 $\delta_1 = -\epsilon_n$. 因此(5.4.11)成立. \square

定理 5.4.6 设 A, B 均为 n 阶 Hermite 矩阵, 且 A 和 B 的特征值分别为 $\lambda_1 \geq \dots \geq \lambda_n$ 和 $\mu_1 \geq \dots \geq \mu_n$. 如果 $A \geq B$, 则

$$\lambda_i \geq \mu_i, \quad i = 1, 2, \dots, n$$

证明 记 $E = A - B$, 则 $E \geq 0$. 设 E 的特征值为 $\epsilon_1 \geq \dots \geq \epsilon_n$, 由定理 5.2.2 知 $\epsilon_n \geq 0$. 因为 $A = B + E$, 则由定理 5.4.5 有

$$\mu_i + \epsilon_1 \geq \lambda_i \geq \mu_i + \epsilon_n \geq \mu_i, \quad i = 1, 2, \dots, n \quad \square$$

如果 A, B 均为 n 阶 Hermite 非负定矩阵, 并且 $A \geq B$, 则由定理 2.4.2 和定理 5.4.6 得

$$\text{tr}(A) \geq \text{tr}(B), \quad |A| \geq |B|.$$

习 题

1. 设 A 为正规矩阵, 证明: A 为 Hermite 矩阵的充分必要条件是 A 的特征值均为实数.
2. 设 A, B 均为 n 阶 Hermite 矩阵, 证明: AB 为 Hermite 矩阵的充分必要条件是 $AB = BA$.
3. 设 A, B 均为 n 阶 Hermite 矩阵, 证明: A 与 B 相似的充分必要条件是它们的特征多项式相同.
4. 设 A 为反 Hermite 矩阵, 证明:

- (1) $I - A, I + A$ 都是可逆矩阵;
 (2) $(I - A)(I + A)^{-1}$ 是特征值不等于 -1 的酉矩阵.
 5. 设 n 阶酉矩阵 U 的特征值不等于 -1 , 试证: $I + U$ 是可逆矩阵, 并且 $H = i(I - U) \cdot (I + U)^{-1}$ 是 Hermite 矩阵. 反之, 若 H 是 Hermite 矩阵, 则 $I - iH$ 是可逆矩阵, 并且 $U = (I + iH)(I - iH)^{-1}$ 是酉矩阵.
 6. 写出如下 Hermite 二次型的矩阵, 并用酉线性替换将它们化为标准形:
 (1) $f(x_1, x_2, x_3) = i\bar{x}_1x_2 + \bar{x}_1x_3 - ix_1\bar{x}_2 + x_1\bar{x}_3$;
 (2) $f(x_1, x_2, x_3) = \bar{x}_1x_1 + i\bar{x}_1x_2 + (1+i)\bar{x}_1x_3 - i\bar{x}_2x_1 + \bar{x}_2x_3 + (1-i)\bar{x}_3x_1 + \bar{x}_3x_2 + 2\bar{x}_3x_3$.
 7. 设 A 是 $m \times n$ 矩阵, 试证:
 (1) $A^H A$ 和 AA^H 都是 Hermite 非负定矩阵;
 (2) $A^H A$ 和 AA^H 的非零特征值相同;
 (3) 如果 A 是列满秩矩阵, $A^H A$ 是 Hermite 正定矩阵.
 8. 证明: (1) 两个 Hermite 非负定矩阵之和是非负定的;
 (2) Hermite 非负定矩阵与 Hermite 正定矩阵之和是正定的.
 9. 设 A 是 Hermite 正定矩阵, 且 A 还是酉矩阵, 则 $A = I$.
 10. 设 A, B 分别为 n 阶 Hermite 正定矩阵和非负定矩阵, 并且 $AB = BA$. 证明: AB 为 Hermite 非负定矩阵.
 11. 设 A, B 均为 n 阶 Hermite 矩阵, 且 A 正定, 试证: AB 相似于实对角矩阵.
 12. 设 A, B 分别为 n 阶 Hermite 正定矩阵和非负定矩阵, 证明: $\text{tr}(AB) = 0$ 的充分必要条件是 $B = 0$.
 13. 设 A 为 n 阶 Hermite 非负定矩阵, 证明: $|I + A| \geq 0$, 并且等号成立的充分必要条件是 $A = 0$.
 14. 设 A, B 分别为 n 阶 Hermite 正定矩阵和非负定矩阵, 证明: $|A + B| \geq |A|$, 并且等号成立的充分必要条件是 $B = 0$.
 15. 设 n 阶 Hermite 正定矩阵 A 有如下分块

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^H & A_{22} \end{bmatrix}, A_{11} \in C^{k \times k}$$

证明: $|A| \leq |A_{11}| |A_{22}|$.

16. 设 A 为 n 阶 Hermite 矩阵, 试证: 存在 $t > 0$, 使得 $tI + A$ 是正定矩阵.
 17. 证明: Hermite 正定矩阵的模最大的元素一定位于主对角线上.
 18. 设 $A = (a_{ij})$ 是 n 阶实矩阵, 证明:

$$|\det(A)| \leq \left(\sum_{i=1}^n a_{i1}^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^n a_{i2}^2\right)^{\frac{1}{2}} \cdots \left(\sum_{i=1}^n a_{in}^2\right)^{\frac{1}{2}}$$

19. 设 $A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix}$, 求对角元为正数的下三角矩阵 L 使得 $A = LL^T$.
 20. 设 A 为 n 阶 Hermite 非负定(正定)矩阵, 则存在唯一的 Hermite 非负定(正定)矩

阵 S 使得 $A = S^2$, 且任何一个与 A 可交换的矩阵必与 S 可交换.

21. 设 A 为 n 阶矩阵, 证明: A 可分解(极分解)为 $A = HQ$, 其中 Q 是酉矩阵, H 是 Hermite 非负定矩阵. 如果 A 非奇异, 则 H 是 Hermite 正定矩阵, 且分解式是惟一的.

22. 设 A 为 n 阶 Hermite 正定矩阵, 对 $x, y \in \mathbb{C}^n$, 定义

$$(x, y) = y^H A x$$

证明:

(1) 在上述定义下 \mathbb{C}^n 是酉空间;

(2) $(x^H A y)^2 \leq (x^H A x)(y^H A y)$.

23. 证明: Hermite 矩阵 A 的全部特征值都在区间 $[a, b]$ 内的充分必要条件是 $aI \leq A \leq bI$.

24. 设 A, B 均为 n 阶 Hermite 正定矩阵, 证明: 若 $A \geq B$ 且 $AB = BA$, 则 $A^3 \geq B^3$.

25. 设 A, B 分别为 n 阶 Hermite 正定矩阵和非负定矩阵, 证明: 若 $A^2 \geq B^2$, 则 $A \geq B$.

第六章 范数与极限

前面几章主要研究了矩阵的代数运算,完全没有涉及矩阵的分析运算.为了建立矩阵分析等的基本概念,需要用到向量和矩阵的范数;在数值分析、数据处理等学科中范数理论起着十分重要的作用.因此,本章首先在线性空间中定义向量的范数,介绍赋范线性空间的概念;其次讨论矩阵的范数及其性质;最后介绍矩阵序列及矩阵幂级数的收敛性概念.

§ 6.1 向量范数

在许多场合,对同一线性空间中的向量需要引入非负数量作为它们“大小”的一种度量,进而比较两个向量之间的“接近”程度.对于线性空间 V ,我们可以在 V 上定义内积 (α, β) ,由内积可导出向量 $\alpha \in V$ 的“长度”或范数 $\|\alpha\| = \sqrt{(\alpha, \alpha)}$,按此范数可以度量一个向量的“大小”以及两个向量之间彼此“接近”的程度.但在实际应用中,仅使用由内积导出的范数往往会带来不便,而且由本节后面的讨论看出,范数不一定从内积导出,它只要满足一些基本的公理条件即可.

定义 6.1.1 设 V 是数域 P 上的线性空间, $\|\alpha\|$ 是以 V 中的向量 α 为自变量的非负实值函数,如果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:对任意 $k \in P, \alpha \in V$,有 $\|k\alpha\| = |k| \|\alpha\|$;
- (3) 三角不等式:对任意 $\alpha, \beta \in V$,有 $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$,

则称 $\|\alpha\|$ 为向量 α 的范数,并称定义了范数的线性空间为**赋范线性空间**.

由定义 6.1.1 容易导出,对任意 $\alpha, \beta \in V$,有

$$|\|\alpha\| - \|\beta\|| \leq \|\alpha - \beta\| \quad (6.1.1)$$

在赋范线性空间 V 中,可以由范数定义两点间的距离.对 $\alpha, \beta \in V$,定义 α 与 β 之间的距离为

$$d(\alpha, \beta) = \|\alpha - \beta\|$$

由上式规定的距离 $d(\alpha, \beta)$ 称为由范数 $\|\cdot\|$ 决定的距离.以后对每个赋范线性空间总是按上式引入距离,使之成为度量空间.

例 6.1.1 在 n 维向量空间 C^n 中,对任意的向量 $x = (x_1, \dots, x_n)^T \in$

C^n , 定义

$$\|x\|_1 = \sum_{i=1}^n |x_i| \quad (6.1.2)$$

$$\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}} \quad (6.1.3)$$

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i| \quad (6.1.4)$$

容易证明: $\|x\|_1$, $\|x\|_2$ 和 $\|x\|_\infty$ 都满足定义 6.1.1 中的三个条件. 因此 $\|x\|_1$,

$\|x\|_2$ 和 $\|x\|_\infty$ 都是 C^n 上的范数, 分别称为 1 范数, 2 范数 (或 Euclid 范数) 和 ∞ 范数.

对 $1 \leq p < +\infty$, 在 C^n 上定义

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}, \quad 1 \leq p < +\infty \quad (6.1.5)$$

则当 $p = 1$ 时, $\|x\|_p = \sum_{i=1}^n |x_i| = \|x\|_1$; 当 $p = 2$ 时, $\|x\|_p = \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}} = \|x\|_2$.

下面证明由 (6.1.5) 定义的 $\|x\|_p$ 是 C^n 上的一种向量范数.

引理 6.1.1 如果实数 $p > 1, q > 1$ 且 $\frac{1}{p} + \frac{1}{q} = 1$, 则对任意非负实数 a, b 有

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} \quad (6.1.6)$$

证明 若 $a = 0$ 或 $b = 0$, 则 (6.1.6) 显然成立. 下面考虑 a, b 均为正数的情况.

对 $x > 0, 0 < \alpha < 1$, 记 $f(x) = x^\alpha - \alpha x$. 容易验证 $f(x)$ 在 $x = 1$ 处达到最大值 $1 - \alpha$, 从而 $f(x) \leq 1 - \alpha$, 即

$$x^\alpha \leq 1 - \alpha + \alpha x$$

对任意正实数 A, B , 在上式中令 $x = \frac{A}{B}, \alpha = \frac{1}{p}, 1 - \alpha = \frac{1}{q}$, 则 $A^{\frac{1}{p}} B^{\frac{1}{q}} \leq \frac{A}{p} + \frac{B}{q}$. 由此再令 $a = A^{\frac{1}{p}}, b = B^{\frac{1}{q}}$, 即得 (6.1.6). \square

定理 6.1.1 (Hölder 不等式) 设 $x = (x_1, \dots, x_n)^T, y = (y_1, \dots, y_n)^T \in C^n$, 则

$$\sum_{i=1}^n |x_i y_i| \leq \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^n |y_i|^q \right)^{\frac{1}{q}} \quad (6.1.7)$$

其中实数 $p > 1, q > 1$ 且 $\frac{1}{p} + \frac{1}{q} = 1$.

证明 如果 $x=0$ 或 $y=0$, 则(6.1.7)显然成立. 下面设 $x \neq 0, y \neq 0$. 令

$$a = \frac{|x_i|}{\left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}}, \quad b = \frac{|y_i|}{\left(\sum_{i=1}^n |y_i|^q \right)^{\frac{1}{q}}}$$

则由(6.1.6)得

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

从而有

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

由上式即得(6.1.7). \square

定理 6.1.2 (Minkowski 不等式) 设 $x = (x_1, \dots, x_n)^T, y = (y_1, \dots, y_n)^T \in \mathbb{C}^n$, 则

$$\left\{ \sum_{i=1}^n |x_i + y_i|^p \right\}^{\frac{1}{p}} \leq \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^n |y_i|^p \right)^{\frac{1}{p}} \quad (6.1.8)$$

其中实数 $p \geq 1$.

证明 当 $p=1$ 时, (6.1.8)显然成立. 下面设 $p > 1$, 记 $q = \frac{p}{p-1}$, 则 $q >$

1 且 $\frac{1}{p} + \frac{1}{q} = 1$. 从而由(6.1.7)有

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

因此

$$\left\{ \sum_{i=1}^n |x_i + y_i|^p \right\}^{\frac{1}{p}} \leq \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^n |y_i|^p \right)^{\frac{1}{p}} \quad \square$$

定理 6.1.3 对任意向量 $x = (x_1, \dots, x_n)^T \in \mathbf{C}^n$, $1 \leq p < \infty$, 由 (6.1.5) 定义的 $\|x\|_p$ 是 \mathbf{C}^n 上的向量范数, 并且 $\lim_{p \rightarrow \infty} \|x\|_p = \|x\|_\infty$.

证明 $\|x\|_p$ 显然满足定义 6.1.1 中的条件 (1)、(2), 由定理 6.1.2 知 $\|x\|_p$ 满足定义 6.1.1 中的条件 (3). 因此 $\|x\|_p$ 是 \mathbf{C}^n 上的向量范数.

不妨设 $x \neq 0$, $|x_{i_0}| = \max_i |x_i| > 0$, 则

$$\|x\|_p = \left(\sum_{i=1}^n |x_{i_0}|^p \left| \frac{x_i}{x_{i_0}} \right|^p \right)^{\frac{1}{p}} = |x_{i_0}| \left(\sum_{i=1}^n \left| \frac{x_i}{x_{i_0}} \right|^p \right)^{\frac{1}{p}}$$

因为 $|x_{i_0}|^p \leq \sum_{i=1}^n |x_i|^p \leq n |x_{i_0}|^p$, 所以 $1 \leq \left(\sum_{i=1}^n \left| \frac{x_i}{x_{i_0}} \right|^p \right)^{\frac{1}{p}} \leq n^{\frac{1}{p}}$. 由于

$\lim_{p \rightarrow \infty} n^{\frac{1}{p}} = 1$, 则

$$\lim_{p \rightarrow \infty} \|x\|_p = |x_{i_0}| = \max_i |x_i| = \|x\|_\infty \quad \square$$

下述定理指出, 可以利用已知的向量范数去构造新范数.

定理 6.1.4 设 $\|\cdot\|_\beta$ 是 \mathbf{C}^m 上的向量范数, $A \in \mathbf{C}^{m \times n}$ 且 $\text{rank}(A) = n$, 则由

$$\|x\|_\alpha = \|Ax\|_\beta, \quad x \in \mathbf{C}^n \quad (6.1.9)$$

所定义的 $\|\cdot\|_\alpha$ 是 \mathbf{C}^n 上的向量范数.

证明 只需验证 $\|\cdot\|_\alpha$ 满足定义 6.1.1 中的三个条件.

(1) 当 $x \neq 0$ 时 $Ax \neq 0$, 从而 $\|x\|_\alpha = \|Ax\|_\beta > 0$, 并且当 $x = 0$ 时 $\|x\|_\alpha = \|A0\|_\beta = 0$;

(2) 对任意复数 k 有 $\|kx\|_\alpha = \|kAx\|_\beta = |k| \|Ax\|_\beta = |k| \|x\|_\alpha$;

(3) 对 $x, y \in \mathbf{C}^n$, 有 $\|x+y\|_\alpha = \|A(x+y)\|_\beta \leq \|Ax\|_\beta + \|Ay\|_\beta = \|x\|_\alpha + \|y\|_\alpha$. \square

定理 6.1.5 设 V 是数域 \mathbf{P} 上的 n 维线性空间, $\epsilon_1, \dots, \epsilon_n$ 为 V 的一组基, 则 V 中任一向量 α 可唯一地表示为 $\alpha = \sum_{i=1}^n x_i \epsilon_i$, $x = (x_1, \dots, x_n)^T \in \mathbf{P}^n$. 又设 $\|\cdot\|$ 是 \mathbf{P}^n 上的向量范数, 令

$$\|\alpha\|_v = \|x\| \quad (6.1.10)$$

则 $\|\alpha\|_v$ 是 V 上的向量范数.

证明 对 V 中任一向量 α , 如果 $\alpha \neq 0$, 则其坐标向量 $x \neq 0$, 从而 $\|\alpha\|_v$

$= \|x\| > 0$; 如果 $\alpha = 0$, 则其坐标向量 $x = 0$, 于是 $\|\alpha\|_v = \|x\| = 0$.

对 $k \in \mathbf{P}$, $k\alpha = \sum_{i=1}^n kx_i \epsilon_i$, 即 $k\alpha$ 的坐标向量为 $kx = (kx_1, \dots, kx_n)^T$, 故

$$\|k\alpha\|_v = \|kx\| = |k| \|x\| = |k| \|\alpha\|_v.$$

对 $\beta \in V$, 则 $\beta = \sum_{i=1}^n y_i \epsilon_i$, $y = (y_1, \dots, y_n)^T$, $\alpha + \beta$ 的坐标向量为 $x + y$.

于是

$$\|\alpha + \beta\|_v = \|x + y\| \leq \|x\| + \|y\| = \|\alpha\|_v + \|\beta\|_v$$

因此 $\|\alpha\|_v$ 是 V 上的向量范数. \square

定理 6.1.6 设 $\|\cdot\|$ 是数域 \mathbf{P} 上 n 维线性空间 V 上的任一向量范数, $\epsilon_1, \dots, \epsilon_n$ 为 V 的一组基, V 中任一向量 α 可惟一地表示为 $\alpha = \sum_{i=1}^n x_i \epsilon_i$, $x = (x_1, \dots, x_n)^T \in \mathbf{P}^n$, 则 $\|\alpha\|$ 是 x_1, \dots, x_n 的连续函数.

证明 对任意 $\epsilon > 0$ 及任意向量 $\alpha = \sum_{i=1}^n x_i \epsilon_i, \beta = \sum_{i=1}^n y_i \epsilon_i \in V$, 令 $\delta = \frac{\epsilon}{\left(\sum_{i=1}^n \|\epsilon_i\|^2\right)^{\frac{1}{2}}}$, 则当 $\|x - y\|_2 < \delta$ 时, 由 (6.1.1) 和定理 6.1.1 有

$$\begin{aligned} |\|\alpha\| - \|\beta\|| &\leq \|\alpha - \beta\| = \left\| \sum_{i=1}^n (x_i - y_i) \epsilon_i \right\| \leq \sum_{i=1}^n |x_i - y_i| \|\epsilon_i\| \\ &\leq \sqrt{\sum_{i=1}^n |x_i - y_i|^2} \sqrt{\sum_{i=1}^n \|\epsilon_i\|^2} < \epsilon \end{aligned}$$

于是 $\|\alpha\|$ 是 x_1, \dots, x_n 的连续函数. \square

由定理 6.1.3, 定理 6.1.4 和定理 6.1.5 知, 在 \mathbf{P}^n 或一般的 n 维线性空间 V 上可以定义无穷多种范数. 那么, 这些范数之间有什么关系呢?

定义 6.1.2 称 $\|\alpha\|_a, \|\alpha\|_b$ 是 n 维线性空间 V 上定义的一种向量范数, 如果存在两个与 α 无关的正常数 d_1, d_2 使得

$$d_1 \|\alpha\|_b \leq \|\alpha\|_a \leq d_2 \|\alpha\|_b, \quad \forall \alpha \in V \quad (6.1.11)$$

则称范数 $\|\alpha\|_a$ 与 $\|\alpha\|_b$ 是等价的.

容易证明: 向量范数的等价具有自反性、对称性和传递性.

定理 6.1.7 有限维线性空间 V 上的任意两个向量范数都是等价的.

证明 设 V 是数域 \mathbf{P} 上的 n 维线性空间, $\epsilon_1, \dots, \epsilon_n$ 为 V 的一组基, 则

V 中任一向量 α 可惟一地表示为 $\alpha = \sum_{i=1}^n x_i \epsilon_i, x = (x_1, \dots, x_n)^T \in \mathbf{P}^n$. 令

$$\| \alpha \|_2 = \| x \|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}}$$

由定理 6.1.5 知 $\| \alpha \|_2$ 是 V 上的一个向量范数.

设 $\| \alpha \|_a$ 是 V 上的任一向量范数. 因为向量范数的等价具有对称性和传递性, 所以我们只需证明 $\| \alpha \|_a$ 与 $\| \alpha \|_2$ 等价即可.

如果 $\alpha = 0$, 则 $\| \alpha \|_a$ 与 $\| \alpha \|_2$ 显然等价. 因此我们不妨假定 $\alpha \neq 0$. 由定理 6.1.6 知 $\| \alpha \|_a$ 是 x_1, \dots, x_n 的连续函数, 所以 $\| \alpha \|_a$ 在有界闭集

$$\varphi = \{ x = (x_1, \dots, x_n)^T \in \mathbf{P}^n \mid |x_1|^2 + |x_2|^2 + \dots + |x_n|^2 = 1 \} \quad (6.1.12)$$

上可取得最大值 M 和最小值 m . 因为如果 $x \in \varphi$, 则 $x \neq 0$, 所以 $m > 0$. 记

$$\beta = \sum_{i=1}^n \frac{x_i}{\| x \|_2} \epsilon_i$$

则 β 在基 $\epsilon_1, \dots, \epsilon_n$ 下的坐标向量为 $y = \left(\frac{x_1}{\| x \|_2}, \frac{x_2}{\| x \|_2}, \dots, \frac{x_n}{\| x \|_2} \right)^T \in \varphi$, 从而有

$$0 < m \leq \| \beta \|_a \leq M$$

但 $\beta = \frac{\alpha}{\| x \|_2}$, 故 $m \| x \|_2 \leq \| \alpha \|_a \leq M \| x \|_2$, 即

$$m \| \alpha \|_2 \leq \| \alpha \|_a \leq M \| \alpha \|_2$$

因此 $\| \alpha \|_a$ 与 $\| \alpha \|_2$ 等价. \square

定义 6.1.3 设 $\{x^{(k)}\}$ 是 C^n 中的向量序列, 其中 $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})^T$. 如果当 $k \rightarrow \infty$ 时 $x^{(k)}$ 的每一个分量 $x_i^{(k)}$ 都有极限 x_i ($i = 1, 2, \dots, n$), 则称向量序列 $\{x^{(k)}\}$ 是收敛的, 并且向量 $x = (x_1, x_2, \dots, x_n)^T$ 称为 $\{x^{(k)}\}$ 的极限, 记为

$$\lim_{k \rightarrow \infty} x^{(k)} = x$$

不收敛的向量序列称为发散的.

类似于数列的收敛性质, 读者不难证明向量序列的收敛性具有如下性质.

设 $\{x^{(k)}\}, \{y^{(k)}\}$ 是 C^n 中两个向量序列, a, b 是复常数, $A \in C^{m \times n}$. 如果 $\lim_{k \rightarrow \infty} x^{(k)} = x, \lim_{k \rightarrow \infty} y^{(k)} = y$, 则

$$(1) \lim_{k \rightarrow \infty} (ax^{(k)} + by^{(k)}) = ax + by;$$

$$(2) \lim_{k \rightarrow \infty} Ax^{(k)} = Ax.$$

定理 6.1.8 C^n 中向量序列 $\{x^{(k)}\}$ 收敛于向量 x 的充分必要条件是对

任一向量范数 $\|\cdot\|$ 数列 $\{\|x^{(k)} - x\|\}$ 收敛于 0.

证明 由范数的等价性, 只要对某一种范数证明即可. 为此取 $\|\cdot\| = \|\cdot\|_\infty$.

必要性. 如果 $\lim_{k \rightarrow \infty} x^{(k)} = x$, 则对 $i = 1, 2, \dots, n$, 有 $x_i^{(k)} - x_i \rightarrow 0 (k \rightarrow \infty)$. 故

$$\max_i |x_i^{(k)} - x_i| \rightarrow 0 \quad (k \rightarrow \infty)$$

即 $\|x^{(k)} - x\|_\infty$ 收敛于 0.

充分性. 如果 $\lim_{k \rightarrow \infty} \|x^{(k)} - x\|_\infty = 0$, 即 $\max_i |x_i^{(k)} - x_i| \rightarrow 0 (k \rightarrow \infty)$, 因为 $|x_j^{(k)} - x_j| \leq \max_i |x_i^{(k)} - x_i| (j = 1, 2, \dots, n)$, 所以对 $j = 1, 2, \dots, n$, 有

$$x_j^{(k)} - x_j \rightarrow 0 \quad (k \rightarrow \infty)$$

因此 $\lim_{k \rightarrow \infty} x^{(k)} = x$. \square

§ 6.2 矩阵范数

6.2.1 基本概念

因为 $m \times n$ 复矩阵的全体 $C^{m \times n}$ 是复数域上线性空间, 所以上节中范数的定义也适用于矩阵. 更确切地, 我们给出如下定义.

定义 6.2.1 设 $\|A\|$ 是以 $C^{m \times n}$ 中的矩阵 A 为自变量的非负实值函数, 如果它满足以下三个条件:

- (1) 非负性: 当 $A \neq 0$ 时, $\|A\| > 0$; 当 $A = 0$ 时, $\|A\| = 0$;
- (2) 齐次性: 对任意 $k \in C, A \in C^{m \times n}$, 有 $\|kA\| = |k| \|A\|$;
- (3) 三角不等式: 对任意 $A, B \in C^{m \times n}$, 有 $\|A + B\| \leq \|A\| + \|B\|$,

则称 $\|A\|$ 为 $m \times n$ 矩阵 A 的范数.

从定义 6.2.1 容易导出

$$|\|A\| - \|B\|| \leq \|A - B\| \quad (6.2.1)$$

例 6.2.1 对于 $A = (a_{ij}) \in C^{m \times n}$, 令

$$\|A\|_1 \equiv \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|$$

$$\|A\|_\infty \equiv \max_{i,j} |a_{ij}|$$

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

容易证明: $\|\cdot\|_1$, $\|\cdot\|_\infty$ 和 $\|\cdot\|_F$ 都是 $C^{m \times n}$ 上的矩阵范数, $\|A\|_F$ 称为 A 的 Frobenius 范数.

由例 1.6.3 可知, 矩阵 A 的 Frobenius 范数 $\|A\|_F$ 是 $C^{m \times n}$ 中的内积 $(A, B) = \text{tr}(B^H A)$ 所导出的矩阵范数. 因此, 矩阵 Frobenius 范数是向量 Euclid 范数的自然推广.

因为 $C^{m \times n}$ 是复数域上 mn 维线性空间, 则由定理 6.1.7 得如下矩阵范数等价性定理.

定理 6.2.1 设 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 是 $C^{m \times n}$ 上的矩阵范数, 则存在仅与 $\|\cdot\|_\alpha$, $\|\cdot\|_\beta$ 有关的正数 d_1 与 d_2 使得

$$d_1 \|A\|_\beta \leq \|A\|_\alpha \leq d_2 \|A\|_\beta, \quad \forall A \in C^{m \times n}$$

6.2.2 相容矩阵范数

我们经常遇到矩阵之间或矩阵与向量之间的乘法运算, 对 $A \in C^{m \times n}$, $B \in C^{n \times k}$, 自然希望对定义在 $C^{m \times n}$ 与 $C^{n \times k}$ 上的同一矩阵范数 $\|\cdot\|$ 有

$$\|AB\| \leq \|A\| \|B\| \quad (6.2.2)$$

满足条件(6.2.2)的矩阵范数 $\|\cdot\|$ 称为具备相容性条件.

Frobenius 范数 $\|\cdot\|_F$ 具备相容性条件, 即

$$\|AB\|_F \leq \|A\|_F \|B\|_F, \quad \forall A \in C^{m \times n}, \forall B \in C^{n \times k} \quad (6.2.3)$$

事实上, 若令 $\lambda_{\max}(A^H A)$ 表示矩阵 $A^H A$ 的最大特征值, 则

$$\begin{aligned} \|AB\|_F^2 &= \text{tr}(B^H A^H AB) \leq \lambda_{\max}(A^H A) \text{tr}(B^H B) \\ &\leq \text{tr}(A^H A) \text{tr}(B^H B) = \|A\|_F^2 \|B\|_F^2 \end{aligned}$$

读者不难证明矩阵范数 $\|\cdot\|_1$ 也具备相容性条件.

必须指出, 并非所有的矩阵范数都具备相容性条件(6.2.2). 例如, 矩阵范数 $\|\cdot\|_\infty$ 就不具备相容性条件. 事实上, 对矩阵

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

则有 $\|A\|_\infty = 1$, $\|B\|_\infty = 1$, 而 $\|AB\|_\infty = 2 > \|A\|_\infty \|B\|_\infty$.

具备相容性条件的矩阵范数在使用上特别方便. 因此, 我们给出相容矩阵范数的定义.

定义 6.2.2 设 $\|\cdot\|_\alpha$, $\|\cdot\|_\beta$ 和 $\|\cdot\|_\gamma$ 分别是 $C^{m \times n}$, $C^{n \times k}$ 和 $C^{m \times k}$ 上的矩阵范数, 如果

$$\|AB\|_\gamma \leq \|A\|_\alpha \|B\|_\beta, \quad \forall A \in C^{m \times n}, \quad \forall B \in C^{n \times k} \quad (6.2.4)$$

则称 $\|\cdot\|_\alpha$, $\|\cdot\|_\beta$ 和 $\|\cdot\|_\gamma$ 是相容的. 特别地, 如果 $C^{n \times n}$ 上的矩阵范数 $\|\cdot\|$ 满足

$$\|AB\| \leq \|A\| \|B\|, \quad \forall A, B \in C^{n \times n} \quad (6.2.5)$$

则称 $\|\cdot\|$ 是自相容的矩阵范数, 或简称为相容范数.

定义 6.2.2 包括了矩阵范数与向量范数的相容性定义. 例如, 矩阵 Frobenius 与向量 Euclid 范数是相容的, 由 (6.2.3) 即得

$$\|Ax\|_2 \leq \|A\|_F \|x\|_2, \quad \forall A \in C^{n \times n}, \quad \forall x \in C^n \quad (6.2.6)$$

定理 6.2.2 设 $\|\cdot\|$ 是 $C^{n \times n}$ 上的相容矩阵范数, 则在 C^n 上存在与 $\|\cdot\|$ 相容的向量范数.

证明 任取一非零向量 $a \in C^n$, 定义

$$\|x\|_a = \|xa^T\|, \quad x \in C^n$$

容易验证 $\|\cdot\|_a$ 是 C^n 上的向量范数, 并且

$$\|Ax\|_a = \|Axa^T\| \leq \|A\| \|xa^T\| = \|A\| \|x\|_a$$

即矩阵范数 $\|\cdot\|$ 与向量范数 $\|\cdot\|_a$ 相容. \square

定理 6.2.3 设 $\|\cdot\|$ 是 $C^{n \times n}$ 上的任一相容矩阵范数, 则对任意 $A \in C^{n \times n}$ 有

$$|\lambda_i| \leq \|A\|, \quad \forall \lambda_i \in \lambda(A) \quad (6.2.7)$$

证明 由定理 6.2.2 知在 C^n 上存在与矩阵范数 $\|\cdot\|$ 相容的向量范数 $\|\cdot\|_a$. 设 $x_i \in C^n$ 是 A 对应于特征值 λ_i 的特征向量, 则

$$Ax_i = \lambda_i x_i, \quad x_i \neq 0$$

从而由

$$|\lambda_i| \|x_i\|_a = \|\lambda_i x_i\|_a = \|Ax_i\|_a \leq \|A\| \|x_i\|_a$$

和 $\|x_i\|_a > 0$ 即得结论. \square

由定理 6.2.3 可知, 对 $A \in C^{n \times n}$ 有

$$\rho(A) \leq \|A\| \quad (6.2.8)$$

其中 $\|\cdot\|$ 是 $C^{n \times n}$ 上的任一相容矩阵范数.

6.2.3 算子范数

定理 6.2.2 说明对于任一相容矩阵范数, 必存在与之相容的向量范数. 反

过来,对给定的向量范数,能否定义与之相容的矩阵范数呢?为了解决这个问题,先介绍两个引理.

引理 6.2.1 设 $\|\cdot\|_v$ 是 C^n 上的向量范数,则点集

$$\varphi_v = \{x \in C^n \mid \|x\|_v = 1\}$$

是 C^n 中的有界闭集.

证明 由定理 6.1.7 知存在常数 $d_1 > 0, d_2 > 0$ 使得

$$d_1 \|x\|_v \leq \|x\|_2 \leq d_2 \|x\|_v, \quad \forall x \in C^n \quad (6.2.9)$$

由此可知,对任意 $x \in \varphi_v$ 有 $\|x\|_2 \leq d_2$. 因此 φ_v 在 C^n 内有界.

因为

$$\|x_k\|_v - \|x_k - x\|_v \leq \|x\|_v \leq \|x_k\|_v + \|x_k - x\|_v$$

如果 x_1, x_2, \dots 是 φ_v 中收敛于 x 的点列,则由定理 6.1.8 知 $\|x_k - x\|_v \rightarrow 0 (k \rightarrow \infty)$. 因为 $\|x_k\|_v = 1$, 则得 $\|x\|_v = 1$, 即 $x \in \varphi_v$. 因此 φ_v 是 C^n 中的有界闭集. \square

类似于定理 6.1.6, 可证如下结论.

引理 6.2.2 设 $\|\cdot\|_\mu$ 是 C^m 上的向量范数, $A \in C^{m \times n}$, 则 $\|Ax\|_\mu$ 是 $x \in C^n$ 的连续函数.

如果 $\|\cdot\|_\mu$ 和 $\|\cdot\|_v$ 分别是 C^m 和 C^n 上的两个向量范数,则由引理 6.2.1 和引理 6.2.2 知, $\|Ax\|_\mu$ 在 φ_v 上达到最大值.

定理 6.2.4 设 $\|\cdot\|_\mu$ 和 $\|\cdot\|_v$ 分别是 C^m 和 C^n 上的两个向量范数,对 $A \in C^{m \times n}$, 令

$$\|A\|_{\mu, v} = \max_{\|x\|_v = 1} \|Ax\|_\mu \quad (6.2.10)$$

则 $\|\cdot\|_{\mu, v}$ 是 $C^{m \times n}$ 上的矩阵范数, 并且 $\|\cdot\|_\mu, \|\cdot\|_v$ 和 $\|\cdot\|_{\mu, v}$ 相容.

证明 对任意 $x \in C^n$ 且 $x \neq 0$, 则 $\frac{x}{\|x\|_v} \in \varphi_v$, 从而有

$$\left\| A \left(\frac{x}{\|x\|_v} \right) \right\|_\mu \leq \|A\|_{\mu, v}$$

于是

$$\|Ax\|_\mu \leq \|A\|_{\mu, v} \|x\|_v \quad (6.2.11)$$

显然上式对 $x=0$ 也成立.

下面证明: (6.2.10) 定义的 $\|\cdot\|_{\mu, v}$ 满足矩阵范数定义 6.2.1 中的三个条件.

(1) 如果 $A \neq 0$, 则必有 e_i 使 $Ae_i \neq 0$, 于是由 (6.2.11), 有

$$0 < \|Ae_i\|_\mu \leq \|A\|_{\mu, v} \|e_i\|_v$$

因此当 $A \neq 0$ 时, $\|A\|_{\mu, v} > 0$; 显然当 $A = 0$ 时, $\|A\|_{\mu, v} = 0$.

(2) 任取 $k \in C$, 有

$$\|kA\|_{\mu, v} = \max_{\|x\|_v=1} \|kAx\|_\mu = \max_{\|x\|_v=1} |k| \|Ax\|_\mu = |k| \|A\|_{\mu, v}$$

(3) 任取 $A, B \in C^{m \times n}$, 设 x_0 满足 $\|x_0\|_v = 1$ 并且 $\|(A+B)x_0\|_\mu = \|A+B\|_{\mu, v}$, 则由 (6.2.11), 有

$$\begin{aligned} \|A+B\|_{\mu, v} &= \|(A+B)x_0\|_\mu \leq \|Ax_0\|_\mu + \|Bx_0\|_\mu \\ &\leq \|A\|_{\mu, v} \|x_0\|_v + \|B\|_{\mu, v} \|x_0\|_v = \|A\|_{\mu, v} + \|B\|_{\mu, v} \end{aligned}$$

由 (6.2.11) 可见, $\|\cdot\|_\mu$, $\|\cdot\|_v$ 和 $\|\cdot\|_{\mu, v}$ 相容. \square

根据定理 6.2.4, 可以引进如下定义.

定义 6.2.3 设 $\|\cdot\|_\mu$ 与 $\|\cdot\|_v$ 分别是 C^m 和 C^n 上的向量范数, 由 (6.2.10) 定义的非负实值函数 $\|\cdot\|_{\mu, v}$ 叫做 $C^{m \times n}$ 上的算子范数或称为由向量范数 $\|\cdot\|_\mu$ 和 $\|\cdot\|_v$ 导出的矩阵范数.

由 (6.2.11) 式可知, 算子范数与诱导它的向量范数是相容的.

关于算子范数之间的相容性, 有下述结果.

定理 6.2.5 设 $\|\cdot\|_\mu$, $\|\cdot\|_v$ 与 $\|\cdot\|_\omega$ 分别是 C^m , C^n 与 C^k 上的向量范数, 如果按照 (6.2.10) 式分别定义 $C^{m \times n}$, $C^{n \times k}$ 和 $C^{m \times k}$ 上的算子范数 $\|\cdot\|_{\mu, v}$, $\|\cdot\|_{v, \omega}$ 和 $\|\cdot\|_{\mu, \omega}$, 则

$$\|AB\|_{\mu, \omega} \leq \|A\|_{\mu, v} \|B\|_{v, \omega}, \quad \forall A \in C^{m \times n}, \forall B \in C^{n \times k} \quad (6.2.12)$$

证明 设 $x_0 \in C^k$ 并且 $\|x_0\|_\omega = 1$, $\|ABx_0\|_\mu = \|AB\|_{\mu, \omega}$, 则由 (6.2.11) 得

$$\begin{aligned} \|AB\|_{\mu, \omega} &= \|ABx_0\|_\mu = \|A(Bx_0)\|_\mu \leq \|A\|_{\mu, v} \|Bx_0\|_v \\ &\leq \|A\|_{\mu, v} \|B\|_{v, \omega} \|x_0\|_\omega = \|A\|_{\mu, v} \|B\|_{v, \omega} \quad \square \end{aligned}$$

如果 $m = n = k$, 并且 $\|\cdot\|_\mu$, $\|\cdot\|_v$ 与 $\|\cdot\|_\omega$ 为 C^n 上的同一个向量范数 $\|\cdot\|$, 则从定理 6.2.5 得到.

定理 6.2.6 设 $\|\cdot\|$ 是 C^n 上的向量范数, 则在 $C^{n \times n}$ 上由向量范数 $\|\cdot\|$ 导出的矩阵范数 $\|\cdot\|$ 是相容矩阵范数, 即

$$\|AB\| \leq \|A\| \|B\|, \quad \forall A, B \in C^{n \times n}$$

如果把 C^n 上的向量范数 $\|\cdot\|_p$ ($p=1, 2, \infty$) 限制到 C^m 上, 恰好是 C^m 上的向量范数 $\|\cdot\|_p$. 由定理 6.2.4, 可以得到 $C^{m \times n}$ 上的算子范数 $\|\cdot\|_p$

$$\|A\|_p = \max_{\|x\|_p=1} \|Ax\|_p, \quad A \in C^{m \times n} \quad (p=1, 2, \infty) \quad (6.2.13)$$

并且由定理 6.2.5 知, 这些算子范数都是相容的, 即

$$\|AB\|_p \leq \|A\|_p \|B\|_p, \quad A \in C^{m \times n}, B \in C^{n \times k} \quad (p=1, 2, \infty) \quad (6.2.14)$$

关于 $\|A\|_p$, 有下面的表示定理.

定理 6.2.7 设 $A=(a_{ij}) \in C^{m \times n}$, 则有

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}| \quad (6.2.15)$$

$$\|A\|_2 = (\lambda_{\max}(A^H A))^{\frac{1}{2}} \quad (6.2.16)$$

$$\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}| \quad (6.2.17)$$

证明 记 $A=[a_1, \dots, a_n]$, 其中 $a_j \in C^m$ ($j=1, \dots, n$), 对任意 $x=(x_1, \dots, x_n)^T \neq 0$ 有

$$\|Ax\|_1 = \left\| \sum_{j=1}^n x_j a_j \right\|_1 \leq \sum_{j=1}^n |x_j| \|a_j\|_1 \leq \max_{1 \leq j \leq n} \|a_j\|_1 \|x\|_1$$

因此

$$\|A\|_1 \leq \max_{1 \leq j \leq n} \|a_j\|_1$$

另一方面, 如果 $\max_{1 \leq j \leq n} \|a_j\|_1 = \|a_k\|_1$, 则由 $\|e_k\|_1=1$ 和

$$\|Ae_k\|_1 = \|a_k\|_1 = \max_{1 \leq j \leq n} \|a_j\|_1$$

知 $\|A\|_1 \geq \max_{1 \leq j \leq n} \|a_j\|_1$. 因此(6.2.15)成立.

同理可证(6.2.17).

下面证明(6.2.16). 对 n 阶 Hermite 矩阵 $A^H A$, 存在 n 阶酉矩阵 U 使得

$$A^H A = U \Lambda U^H$$

其中 Λ 是对角矩阵, 其对角元为 $A^H A$ 的特征值, 则

$$\begin{aligned} \|A\|_2^2 &= \max_{\|x\|_2=1} x^H A^H A x = \max_{\|x\|_2=1} (U^H x)^H \Lambda U^H x \\ &= \max_{\|y\|_2=1} y^H \Lambda y = \lambda_{\max}(A^H A). \end{aligned}$$

□

通常将 $\|A\|_1$ 称为 A 的列和范数, $\|A\|_2$ 称为 A 的谱范数, $\|A\|_\infty$ 称为 A 的行和范数.

例 6.2.2 设

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

计算 $\|A\|_1$, $\|A\|_2$, $\|A\|_\infty$ 和 $\|A\|_F$.

解 $\|A\|_1 = 5$, $\|A\|_\infty = 5$, $\|A\|_F = \sqrt{23}$.

因为 $A^H A = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 9 & 6 \\ 0 & 6 & 9 \end{pmatrix}$, 所以 $\lambda_{\max}(A^H A) = 15$. 因此 $\|A\|_2 = \sqrt{15}$.

值得指出的是, Frobenius 范数 $\|\cdot\|_F$ 是相容范数, 但不是算子范数. Frobenius 范数 $\|\cdot\|_F$ 不仅计算简单, 而且具有如下性质.

定理 6.2.8 设 $A \in C^{m \times n}$, U 和 V 分别为 m 阶和 n 阶酉矩阵, 则

$$\|UAV\|_F = \|A\|_F \quad (6.2.18)$$

证明 因为 U 和 V 均为酉矩阵, 所以

$$\begin{aligned} \|UAV\|_F^2 &= \text{tr}[(UAV)^H(UAV)] = \text{tr}(V^H A^H A V) \\ &= \text{tr}(V^{-1} A^H A V) = \text{tr}(A^H A) = \|A\|_F^2 \end{aligned}$$

因此 $\|UAV\|_F = \|A\|_F$. \square

矩阵的谱范数虽不便计算, 但有许多很好的性质, 所以在理论研究中常常使用谱范数.

定理 6.2.9 设 $A \in C^{m \times n}$, 则

$$\|A\|_2 = \max_{\substack{\|x\|_2=1 \\ \|y\|_2=1}} |y^H A x| \quad (6.2.19)$$

$$\|A^H\|_2 = \|A^T\|_2 = \|A\|_2 \quad (6.2.20)$$

$$\|A^H A\|_2 = \|A\|_2^2 \quad (6.2.21)$$

$$\|A\|_2^2 \leq \|A\|_1 \|A\|_\infty \quad (6.2.22)$$

并且对 m 阶酉矩阵 U 和 n 阶酉矩阵 V , 有

$$\|UAV\|_2 = \|A\|_2 \quad (6.2.23)$$

证明 设 $x \in C^n, y \in C^m$ 满足 $\|x\|_2 = 1, \|y\|_2 = 1$, 则由定理 1.6.1 有

$$|y^H Ax| \leq \|y\|_2 \|Ax\|_2 \leq \|A\|_2 \|x\|_2 = \|A\|_2$$

另一方面, 设 $\|x\|_2 = 1$ 并且满足 $\|Ax\|_2 = \|A\|_2$, 令 $y = \frac{Ax}{\|Ax\|_2}$,

$$|y^H Ax| = \frac{x^H A^H Ax}{\|Ax\|_2} = \|Ax\|_2 = \|A\|_2$$

因此(6.2.19)成立.

由(6.2.19), 得

$$\|A^H\|_2 = \max_{\substack{\|x\|_2=1 \\ \|y\|_2=1}} |y^H A^H x| = \max_{\substack{\|x\|_2=1 \\ \|y\|_2=1}} |x^H A y| = \|A\|_2$$

同理可证 $\|A^T\|_2 = \|A\|_2$. 因此(6.2.20)成立.

因为

$$\|A^H A\|_2 \leq \|A^H\|_2 \|A\|_2$$

由(6.2.20), 有

$$\|A^H A\|_2 \leq \|A\|_2^2$$

另一方面, 取 $x \in C^n$ 满足 $\|x\|_2 = 1$ 和 $\|Ax\|_2 = \|A\|_2$, 并在(6.2.19)中令 $y = x$, 则

$$\|A^H A\|_2 \geq x^H A^H Ax = \|Ax\|_2^2 = \|A\|_2^2$$

因此(6.2.21)成立.

由(6.2.16), 有 $\|A\|_2^2 = \lambda_{\max}(A^H A)$. 因为 $\|\cdot\|_1$ 是相容矩阵范数, 由(6.2.8)得

$$\lambda_{\max}(A^H A) \leq \|A^H A\|_1 \leq \|A^H\|_1 \|A\|_1 = \|A\|_1 \|A\|_{\infty}$$

所以(6.2.22)成立. \square

因为 U 和 V 均为酉矩阵, 由(6.2.16)有

$$\begin{aligned} \|UAV\|_2 &= (\lambda_{\max}((UAV)^H UAV))^{\frac{1}{2}} = (\lambda_{\max}(V^H A^H AV))^{\frac{1}{2}} \\ &= (\lambda_{\max}(A^H A))^{\frac{1}{2}} = \|A\|_2 \end{aligned}$$

所以(6.2.23)成立. \square

§ 6.3 矩阵序列与矩阵级数

本节利用前两节建立的范数理论讨论矩阵序列的极限和矩阵级数等分析

概念.

6.3.1 矩阵序列的极限

由于 $C^{m \times n}$ 中矩阵可以看作一个 mn 维向量, 其收敛性可以和 C^m 中的向量一样考虑. 因此, 我们可以用矩阵各个元素序列的同时收敛来规定矩阵序列的收敛性.

定义 6.3.1 设有矩阵序列 $\{A^{(k)}\}$, 其中 $A^{(k)} = (a_{ij}^{(k)}) \in C^{m \times n}$. 如果当 $k \rightarrow \infty$ 时, 矩阵 $A^{(k)}$ 的每一个元素 $a_{ij}^{(k)}$ 都有极限 a_{ij} , 即

$$\lim_{k \rightarrow \infty} a_{ij}^{(k)} = a_{ij}, \quad 1 \leq i \leq m; 1 \leq j \leq n$$

则称矩阵序列 $\{A^{(k)}\}$ 是收敛的, 并把矩阵 $A = (a_{ij}) \in C^{m \times n}$ 称为 $\{A^{(k)}\}$ 的极限, 或称 $\{A^{(k)}\}$ 收敛于 A , 记为

$$\lim_{k \rightarrow \infty} A^{(k)} = A \text{ 或 } A^{(k)} \rightarrow A \quad (k \rightarrow \infty)$$

类似于定理 6.1.8, 有

定理 6.3.1 设 $\|\cdot\|$ 是 $C^{m \times n}$ 上任一矩阵范数, $C^{m \times n}$ 中矩阵序列 $\{A^{(k)}\}$ 收敛于矩阵 A 的充分必要条件是

$$\lim_{k \rightarrow \infty} \|A^{(k)} - A\| = 0$$

关于矩阵序列的极限运算有如下性质.

(1) 设 $\{A^{(k)}\}, \{B^{(k)}\}$ 是 $C^{m \times n}$ 中矩阵序列, 且 $\lim_{k \rightarrow \infty} A^{(k)} = A, \lim_{k \rightarrow \infty} B^{(k)} = B$, 则

$$\lim_{k \rightarrow \infty} (aA^{(k)} + bB^{(k)}) = aA + bB$$

其中 $a, b \in C$ 是常数.

(2) 设 $\{A^{(k)}\}$ 与 $\{B^{(k)}\}$ 分别是 $C^{m \times n}$ 与 $C^{n \times k}$ 中矩阵序列, 并且 $\lim_{k \rightarrow \infty} A^{(k)} = A, \lim_{k \rightarrow \infty} B^{(k)} = B$, 则

$$\lim_{k \rightarrow \infty} A^{(k)} B^{(k)} = AB$$

由定义 6.3.1 和定理 6.3.1 容易验证性质(1)和(2).

现在考虑由矩阵 $A \in C^{n \times n}$ 的幂所构成的矩阵序列 $A, A^2, A^3, \dots, A^k, \dots$ 的收敛性.

定理 6.3.2 设 $A \in C^{n \times n}$, $\lim_{k \rightarrow \infty} A^k = 0$ 的充分必要条件是 $\rho(A) < 1$.

证明 对 $A \in C^{n \times n}$, 由定理 3.5.1 知, 存在 n 阶可逆矩阵 P 使得

$$P^{-1}AP = J = \text{diag}(J_1, J_2, \dots, J_s) \quad (6.3.1)$$

其中

$$J_i = \begin{pmatrix} \lambda_i & 1 & & 0 \\ & \lambda_i & 1 & \\ & & \ddots & \ddots \\ 0 & & & \lambda_i & 1 \\ & & & & \lambda_i \end{pmatrix}_{n_i \times n_i} \quad (6.3.2)$$

则

$$P^{-1}A^kP = J^k = \text{diag}(J_1^k, J_2^k, \dots, J_s^k) \quad (6.3.3)$$

因此 $\lim_{k \rightarrow \infty} A^k = 0 \Leftrightarrow \lim_{k \rightarrow \infty} J^k = 0 \Leftrightarrow \lim_{k \rightarrow \infty} J_i^k = 0 (i = 1, 2, \dots, s)$. 而

$$J_i^k = \begin{pmatrix} f_k(\lambda_i) & f'_k(\lambda_i) & \frac{1}{2!}f''_k(\lambda_i) & \cdots & \frac{1}{(n_i-1)!}f_k^{(n_i-1)}(\lambda_i) \\ & f_k(\lambda_i) & f'_k(\lambda_i) & \ddots & \vdots \\ & & & \ddots & \vdots \\ & & & & \frac{1}{2!}f''_k(\lambda_i) \\ 0 & & & & f'_k(\lambda_i) \\ & & & & f_k(\lambda_i) \end{pmatrix} \quad (6.3.4)$$

其中 $f_k(\lambda) = \lambda^k$. 因为对任一多项式 $g(\lambda)$, 当 $k \rightarrow \infty$ 时, $g(k)|\lambda_i|^k \rightarrow 0 \Leftrightarrow |\lambda_i| < 1$. 而 $|\lambda_i| < 1 (i = 1, 2, \dots, s) \Leftrightarrow \rho(A) < 1$. \square

由(6.2.8)和定理 6.3.2 即得如下结果.

定理 6.3.3 设 $A \in C^{n \times n}$, 如果存在 $C^{n \times n}$ 上的一种相容矩阵范数 $\|\cdot\|$ 使 $\|A\| < 1$, 则 $\lim_{k \rightarrow \infty} A^k = 0$.

定理 6.3.4 设 $\|\cdot\|$ 是 $C^{n \times n}$ 上的相容矩阵范数, 则对任意 $A \in C^{n \times n}$, 有

$$\rho(A) = \lim_{k \rightarrow \infty} \|A^k\|^{\frac{1}{k}}$$

证明 因为 $(\rho(A))^k = \rho(A^k) \leq \|A^k\|$, 所以对所有 $k = 1, 2, \dots$, 有 $\rho(A) \leq \|A^k\|^{\frac{1}{k}}$. 另一方面, 对任意 $\varepsilon > 0$, 矩阵 $\tilde{A} = [\rho(A) + \varepsilon]^{-1}A$ 的谱半径严格小于 1. 由定理 6.3.2 知 $\lim_{k \rightarrow \infty} \tilde{A}^k = 0$, 于是当 $k \rightarrow \infty$ 时 $\|\tilde{A}^k\| \rightarrow 0$. 因此, 存在正整数 K 使得当 $k > K$ 时 $\|\tilde{A}^k\| < 1$, 即对所有 $k > K$ 有 $\|A^k\| \leq$

$[\rho(A) + \varepsilon]^k$ 或 $\|A^k\|^{\frac{1}{k}} \leq \rho(A) + \varepsilon$. 故 $\lim_{k \rightarrow \infty} \|A^k\|^{\frac{1}{k}} = \rho(A)$. \square

6.3.2 矩阵级数

在建立矩阵函数以及表示线性微分方程组的解时,常常用到矩阵级数.特别是矩阵幂级数.下面利用矩阵序列的收敛性讨论矩阵级数的收敛性.

定义 6.3.2 设 $\{A^{(k)}\}$ 是 $C^{m \times n}$ 的矩阵序列,其中 $A^{(k)} = (a_{ij}^{(k)}) \in C^{m \times n}$, 无穷和

$$A^{(1)} + A^{(2)} + \cdots + A^{(k)} + \cdots \quad (6.3.5)$$

称为矩阵级数,记为 $\sum_{k=1}^{\infty} A^{(k)}$. 对正整数 $k \geq 1$, 记 $S^{(k)} = \sum_{i=1}^k A^{(i)}$, 称 $S^{(k)}$ 为矩阵级数 $\sum_{k=1}^{\infty} A^{(k)}$ 的部分和. 如果矩阵序列 $\{S^{(k)}\}$ 收敛, 且有极限 S , 即 $\lim_{k \rightarrow \infty} S^{(k)} = S$, 则称矩阵级数 $\sum_{k=1}^{\infty} A^{(k)}$ 收敛, 并称 S 为矩阵级数 $\sum_{k=1}^{\infty} A^{(k)}$ 的和, 记为 $\sum_{k=1}^{\infty} A^{(k)} = S$. 不收敛的矩阵级数称为发散的.

由定义 6.3.2 可知, 矩阵级数 $\sum_{k=1}^{\infty} A^{(k)}$ 收敛的充分必要条件是 mn 个数项级数 $\sum_{k=1}^{\infty} a_{ij}^{(k)} (i = 1, 2, \cdots, m, j = 1, 2, \cdots, n)$ 都收敛.

由矩阵级数的收敛性定义易知

(1) 若矩阵级数 $\sum_{k=1}^{\infty} A^{(k)}$ 收敛, 则 $\lim_{k \rightarrow \infty} A^{(k)} = 0$;

(2) 若矩阵级数 $\sum_{k=1}^{\infty} A^{(k)} = S_1, \sum_{k=1}^{\infty} B^{(k)} = S_2, a, b \in C$ 是常数, 则

$$\sum_{k=1}^{\infty} (aA^{(k)} + bB^{(k)}) = aS_1 + bS_2$$

(3) 设 $P \in C^{m \times m}, Q \in C^{m \times n}$, 若矩阵级数 $\sum_{k=1}^{\infty} A^{(k)}$ 收敛, 则 $\sum_{k=1}^{\infty} PA^{(k)}Q$ 收敛, 且

$$\sum_{k=1}^{\infty} PA^{(k)}Q = P\left(\sum_{k=1}^{\infty} A^{(k)}\right)Q \quad (6.3.6)$$

定义 6.3.3 设 $\sum_{k=1}^{\infty} A^{(k)}$ 是矩阵级数, 其中 $A^{(k)} = (a_{ij}^{(k)}) \in C^{m \times n}$, 如果 mn 个数项级数 $\sum_{k=1}^{\infty} a_{ij}^{(k)} (i = 1, 2, \cdots, m; j = 1, 2, \cdots, n)$ 都绝对收敛, 则称矩

阵级数 $\sum_{k=1}^{\infty} \mathbf{A}^{(k)}$ 绝对收敛.

定理 6.3.5 矩阵级数 $\sum_{k=1}^{\infty} \mathbf{A}^{(k)}$ (其中 $\mathbf{A} = (a_{ij}^{(k)}) \in \mathbb{C}^{m \times n}$) 绝对收敛的充分必要条件是数项级数 $\sum_{k=1}^{\infty} \|\mathbf{A}^{(k)}\|$ 收敛, 其中 $\|\cdot\|$ 是 $\mathbb{C}^{m \times n}$ 上的任一矩阵范数.

证明 由定理 6.2.1, 不妨取例 6.2.1 中的矩阵范数 $\|\cdot\|_1$. 如果 $\sum_{k=1}^{\infty} \mathbf{A}^{(k)}$ 绝对收敛, 则存在与 i, j 无关的正数 M 使得

$$\sum_{k=1}^{\infty} |a_{ij}^{(k)}| < M, \quad i = 1, 2, \dots, m; j = 1, 2, \dots, n$$

从而有

$$\sum_{k=1}^{\infty} \|\mathbf{A}^{(k)}\|_1 = \sum_{k=1}^{\infty} \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}^{(k)}| \right) < mnM$$

因此 $\sum_{k=1}^{\infty} \|\mathbf{A}^{(k)}\|_1$ 收敛.

反过来, 如果 $\sum_{k=1}^{\infty} \|\mathbf{A}^{(k)}\|_1$ 收敛, 则由 $|a_{ij}^{(k)}| \leq \|\mathbf{A}^{(k)}\|_1$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$) 知, mn 个数项级数 $\sum_{k=1}^{\infty} a_{ij}^{(k)}$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$) 都绝对收敛. 故矩阵级数 $\sum_{k=1}^{\infty} \mathbf{A}^{(k)}$ 绝对收敛. \square

与数项级数类似, 若矩阵级数 $\sum_{k=1}^{\infty} \mathbf{A}^{(k)}$ 绝对收敛, 则它一定收敛, 并且任意交换各项的求和次序所得的新级数仍收敛, 和也不改变.

对矩阵级数也有幂级数的概念.

定义 6.3.4 设 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 形如

$$\sum_{k=0}^{\infty} c_k \mathbf{A}^k = c_0 \mathbf{I} + c_1 \mathbf{A} + c_2 \mathbf{A}^2 + \cdots + c_k \mathbf{A}^k + \cdots \quad (6.3.7)$$

的矩阵级数称为矩阵幂级数.

我们对矩阵幂级数作进一步讨论, 它是研究矩阵函数的重要工具.

由定理 6.3.5 即得如下定理.

定理 6.3.6 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 如果数项级数 $\sum_{k=0}^{\infty} |c_k| \|\mathbf{A}\|^k$ 收敛, 则矩阵幂级数 $\sum_{k=0}^{\infty} c_k \mathbf{A}^k$ 绝对收敛, 其中 $\|\cdot\|$ 是 $\mathbb{C}^{n \times n}$ 上的某种相容矩阵范数.

推论 6.3.1 设 $A \in C^{m \times n}$, 如果 $C^{m \times n}$ 上的某种相容矩阵范数 $\|\cdot\|$ 使得 $\|A\|$ 在幂级数

$$\sum_{k=0}^{\infty} c_k z^k = c_0 + c_1 z + c_2 z^2 + \cdots + c_k z^k + \cdots$$

的收敛圆内, 则矩阵幂级数 $\sum_{k=0}^{\infty} c_k A^k$ 绝对收敛.

定理 6.3.7 设 $A \in C^{m \times n}$, 并且幂级数 $\sum_{k=0}^{\infty} c_k z^k$ 的收敛半径为 R . 如果 $\rho(A) < R$, 则矩阵幂级数 $\sum_{k=0}^{\infty} c_k A^k$ 绝对收敛; 如果 $\rho(A) > R$, 则矩阵幂级数 $\sum_{k=0}^{\infty} c_k A^k$ 发散.

证明 设矩阵 A 的 Jordan 标准形为 J , 即存在可逆矩阵 P 使得 (6.3.1) 成立, 并且由 (6.3.2) - (6.3.4) 可得

$$\sum_{k=0}^{\infty} c_k A^k = P \left(\sum_{k=0}^{\infty} c_k J^k \right) P^{-1} = P \operatorname{diag} \left(\sum_{k=0}^{\infty} c_k J_1^k, \cdots, \sum_{k=0}^{\infty} c_k J_s^k \right) P^{-1}$$

其中

$$\sum_{k=0}^{\infty} c_k J_i^k = \begin{pmatrix} \sum_{k=0}^{\infty} c_k \lambda_i^k & \sum_{k=1}^{\infty} c_k C_k^1 \lambda_i^{k-1} & \cdots & \sum_{k=n_i-1}^{\infty} c_k C_k^{n_i-1} \lambda_i^{k-n_i+1} \\ & \ddots & & \vdots \\ & & \ddots & \\ & & & \sum_{k=1}^{\infty} c_k C_k^1 \lambda_i^{k-1} \\ 0 & & & \sum_{k=0}^{\infty} c_k \lambda_i^k \end{pmatrix}_{n_i \times n_i}$$

$$\begin{cases} C_k = \frac{k(k-1)\cdots(k-i+1)}{i!}, & k \geq i \\ C_k = 0, & k < i \end{cases}$$

则当 $\rho(A) < R$ 时, 幂级数 $\sum_{k=0}^{\infty} c_k \lambda_i^k, \sum_{k=1}^{\infty} c_k C_k^1 \lambda_i^{k-1}, \cdots, \sum_{k=n_i-1}^{\infty} c_k C_k^{n_i-1} \lambda_i^{k-n_i+1}$ ($i=1, 2, \cdots, s$) 绝对收敛, 因此矩阵幂级数 $\sum_{k=0}^{\infty} c_k A^k$ 绝对收敛; 当 $\rho(A) > R$ 时, 则 A

有某个特征值 $|\lambda_i| > R$, 幂级数 $\sum_{k=0}^{\infty} c_k \lambda_i^k$ 发散, 故矩阵幂级数 $\sum_{k=0}^{\infty} c_k \mathbf{A}^k$ 发散. \square

推论 6.3.2 如果幂级数 $\sum_{k=0}^{\infty} c_k z^k$ 在整个平面上都收敛, 则对任意 $\mathbf{A} \in \mathbf{C}^{m \times n}$, 矩阵幂级数 $\sum_{k=0}^{\infty} c_k \mathbf{A}^k$ 收敛.

定理 6.3.8 设 $\mathbf{A} \in \mathbf{C}^{m \times n}$, 如果 $\|\mathbf{A}\| < 1$, 则矩阵 $\mathbf{I} - \mathbf{A}$ 非奇异, 并且

$$\|(\mathbf{I} - \mathbf{A})^{-1}\| \leq \frac{1}{1 - \|\mathbf{A}\|} \quad (6.3.8)$$

其中 $\|\cdot\|$ 是 $\mathbf{C}^{m \times n}$ 上的相容矩阵范数且满足 $\|\mathbf{I}\| = 1$.

证明 因为 $\|\cdot\|$ 是 $\mathbf{C}^{m \times n}$ 上的相容矩阵范数, 则由定理 6.2.2 知, 在 \mathbf{C}^n 上存在与 $\|\cdot\|$ 相容的向量范数 $\|\cdot\|_a$. 如果矩阵 $\mathbf{I} - \mathbf{A}$ 奇异, 则存在非零向量 x 使得

$$(\mathbf{I} - \mathbf{A})x = 0$$

即 $x = \mathbf{A}x$, 从而

$$\|x\|_a = \|\mathbf{A}x\|_a \leq \|\mathbf{A}\| \|x\|_a.$$

则 $\|\mathbf{A}\| \geq 1$, 这与 $\|\mathbf{A}\| < 1$ 矛盾. 因此 $\mathbf{I} - \mathbf{A}$ 非奇异.

因为 $(\mathbf{I} - \mathbf{A})(\mathbf{I} - \mathbf{A})^{-1} = \mathbf{I}$, 则

$$(\mathbf{I} - \mathbf{A})^{-1} = \mathbf{I} + \mathbf{A}(\mathbf{I} - \mathbf{A})^{-1}$$

从而

$$\|(\mathbf{I} - \mathbf{A})^{-1}\| \leq \|\mathbf{I}\| + \|\mathbf{A}(\mathbf{I} - \mathbf{A})^{-1}\| \leq 1 + \|\mathbf{A}\| \|(\mathbf{I} - \mathbf{A})^{-1}\|$$

于是由上式即得(6.3.8). \square

定理 6.3.9 设 $\mathbf{A} \in \mathbf{C}^{m \times n}$, 则矩阵幂级数 $\sum_{k=0}^{\infty} \mathbf{A}^k$ 收敛的充分必要条件是 $\rho(\mathbf{A}) < 1$, 并且其和为 $(\mathbf{I} - \mathbf{A})^{-1}$, 即

$$\sum_{k=0}^{\infty} \mathbf{A}^k = (\mathbf{I} - \mathbf{A})^{-1} \quad (6.3.9)$$

此外, 如果 $\|\mathbf{A}\| < 1$, 则有

$$\|(\mathbf{I} - \mathbf{A})^{-1} - \sum_{k=0}^m \mathbf{A}^k\| \leq \frac{\|\mathbf{A}\|^{m+1}}{1 - \|\mathbf{A}\|} \quad (6.3.10)$$

其中 $\|\cdot\|$ 是 $\mathbf{C}^{m \times n}$ 上的相容矩阵范数且满足 $\|\mathbf{I}\| = 1$.

证明 必要性. 如果 $\sum_{k=0}^{\infty} \mathbf{A}^k$ 收敛, 则 $\lim_{k \rightarrow \infty} \mathbf{A}^k = 0$. 由定理 6.3.2 知, $\rho(\mathbf{A}) < 1$.

充分性. 因为幂级数 $\sum_{k=0}^{\infty} z^k$ 的收敛半径 $R = 1$, 则由定理 6.3.7 知, 当 $\rho(\mathbf{A}) < 1$ 时, 矩阵幂级数 $\sum_{k=0}^{\infty} \mathbf{A}^k$ 收敛.

因为 $\rho(\mathbf{A}) < 1$, 所以 $\mathbf{I} - \mathbf{A}$ 非奇异, 并且 $\lim_{k \rightarrow \infty} \mathbf{A}^k = 0$. 令 $\mathbf{S}_m = \sum_{k=0}^m \mathbf{A}^k$, 则

$$\mathbf{S}_m(\mathbf{I} - \mathbf{A}) = \mathbf{I} - \mathbf{A}^{m+1}$$

从而

$$\mathbf{S}_m = (\mathbf{I} - \mathbf{A})^{-1} - \mathbf{A}^{m+1}(\mathbf{I} - \mathbf{A})^{-1} \quad (6.3.11)$$

在上式两端令 $m \rightarrow \infty$ 即得(6.3.9).

下面证明不等式(6.3.10). 由(6.3.11)得

$$(\mathbf{I} - \mathbf{A})^{-1} - \mathbf{S}_m = \mathbf{A}^{m+1}(\mathbf{I} - \mathbf{A})^{-1}$$

从而

$$\left\| (\mathbf{I} - \mathbf{A})^{-1} - \sum_{k=0}^m \mathbf{A}^k \right\| \leq \left\| \mathbf{A}^{m+1} \right\| \left\| (\mathbf{I} - \mathbf{A})^{-1} \right\| \leq \left\| \mathbf{A} \right\|^{m+1} \left\| (\mathbf{I} - \mathbf{A})^{-1} \right\| \quad (6.3.12)$$

再由(6.3.8)即得(6.3.10). \square

§ 6.4 矩阵扰动分析

为了解决科学与工程中的实际问题, 人们依据物理、力学等规律建立问题的数学模型, 并根据数学模型提出求解数学问题的数值计算方法, 然后进行程序设计, 在计算机上计算出实际需要的结果. 在数学问题的求解过程中, 通常存在两类误差影响计算结果的精度, 即数值计算方法引起的截断误差和计算环境引起的舍入误差. 为了分析这些误差对数学问题解的影响, 人们将其归结为原始数据的扰动(或摄动)对解的影响. 自然地, 我们需要研究该扰动引起了问题解的多大变化, 即问题解的稳定性.

下面看一个简单的例子.

考虑一个 2 阶线性方程组

$$\begin{pmatrix} 1 & 0.99 \\ 0.99 & 0.98 \end{pmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

可以验证, 该方程组的精确解为 $x_1 = 100, x_2 = -100$.

如果系数矩阵有一扰动 $\begin{pmatrix} 0 & 0 \\ 0 & 0.01 \end{pmatrix}$, 并且右端项也有一扰动 $\begin{bmatrix} 0 \\ 0.001 \end{bmatrix}$, 则扰动后的线性方程组为

$$\begin{pmatrix} 1 & 0.99 \\ 0.99 & 0.99 \end{pmatrix} \begin{pmatrix} x_1 + \delta x_1 \\ x_2 + \delta x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1.001 \end{pmatrix}$$

可以验证, 这个方程组的精确解为 $x_1 + \delta x_1 = -0.1, x_2 + \delta x_2 = \frac{10}{9}$.

可见, 系数矩阵和右端项的微小扰动引起了解的强烈变化.

注意到上面的例子并没有截断误差和舍入误差, 因此原始数据的扰动对问题解的影响程度取决于问题本身的固有性质. 如果原始数据的小扰动引起问题解的很大变化, 则称该问题是病态的(敏感的)或不稳定的; 否则, 称该问题是良态的(不敏感的)或稳定的.

矩阵扰动分析就是研究矩阵元素的变化对矩阵问题解的影响, 它对矩阵论和矩阵计算都具有重要意义. 矩阵扰动分析的理论及其主要结果是在最近三、四十年里得到的. 随着各种科学计算问题的深入与扩大, 矩阵扰动理论不仅会有新的进展, 而且还存在许多问题有待进一步解决. 这里简要介绍矩阵 A 的逆矩阵、以 A 为系数矩阵的线性方程组的解和矩阵特征值的扰动分析.

以下总是假定 $\|\cdot\|$ 为 $C^{m \times n}$ 上满足 $\|I\| = 1$ 的相容矩阵范数.

6.4.1 矩阵逆的扰动分析

设矩阵 $A \in C^{m \times n}$ 并且 A 非奇异, 经扰动变为 $A + E$, 其中 $E \in C^{m \times n}$ 称为扰动矩阵. 我们需要解决在什么条件下 $A + E$ 非奇异? 当 $A + E$ 非奇异时, $(A + E)^{-1}$ 与 A^{-1} 的近似程度?

定理 6.4.1 设 $A, E \in C^{m \times n}, B = A + E$. 如果 A 与 B 均非奇异, 则

$$\frac{\|B^{-1} - A^{-1}\|}{\|A^{-1}\|} \leq \|A\| \|B^{-1}\| \|E\| \quad (6.4.1)$$

证明 由

$$B^{-1} - A^{-1} = A^{-1}(A - B)B^{-1} = -A^{-1}EB^{-1} \quad (6.4.2)$$

即得

$$\|B^{-1} - A^{-1}\| \leq \|A^{-1}\| \|B^{-1}\| \|E\|$$

于是

$$\frac{\|B^{-1} - A^{-1}\|}{\|A^{-1}\|} \leq \|B^{-1}\| \|E\| = \|A\| \|B^{-1}\| \|E\| \quad \square$$

定理 6.4.2 设 $A \in C^{n \times n}$ 是非奇异矩阵, $E \in C^{n \times n}$ 满足条件

$$\|A^{-1}E\| < 1 \quad (6.4.3)$$

则 $A + E$ 非奇异, 并且有

$$\|(A + E)^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}E\|} \quad (6.4.4)$$

$$\frac{\|(A + E)^{-1} - A^{-1}\|}{\|A^{-1}\|} \leq \frac{\|A^{-1}E\|}{1 - \|A^{-1}E\|} \quad (6.4.5)$$

证明 因为 $A + E = A(I + A^{-1}E)$, 其中 $\|A^{-1}E\| < 1$, 则由定理 6.3.8 知 $I + A^{-1}E$ 非奇异, 从而 $A + E$ 也非奇异. 由于

$$(A + E)^{-1} = (I + A^{-1}E)^{-1}A^{-1} \quad (6.4.6)$$

则

$$\|(A + E)^{-1}\| \leq \|(I + A^{-1}E)^{-1}\| \|A^{-1}\|$$

由(6.3.8)得

$$\|(A + E)^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}E\|}$$

因为

$$(A + E)^{-1} - A^{-1} = [(I + A^{-1}E)^{-1} - I]A^{-1}$$

则由定理 6.3.9 有

$$\|(A + E)^{-1} - A^{-1}\| \leq \|(I + A^{-1}E)^{-1} - I\| \|A^{-1}\| \leq \frac{\|A^{-1}\| \|A^{-1}E\|}{1 - \|A^{-1}E\|}$$

由上式即得(6.4.5). \square

因为 $\|A^{-1}E\| \leq \|A^{-1}\| \|E\|$, 所以由定理 6.4.2 可得如下推论.

推论 6.4.1 设 $A \in C^{n \times n}$ 是非奇异矩阵, $E \in C^{n \times n}$ 满足条件 $\|A^{-1}\| \|E\| < 1$, 则 $A + E$ 非奇异, 并且有

$$\frac{\|(A + E)^{-1} - A^{-1}\|}{\|A^{-1}\|} \leq \frac{\kappa(A) \frac{\|E\|}{\|A\|}}{1 - \kappa(A) \frac{\|E\|}{\|A\|}} \quad (6.4.7)$$

其中 $\kappa(A) = \|A\| \|A^{-1}\|$.

估计式(6.4.7)表明, $\kappa(A)$ 反映了 A^{-1} 对于 A 的扰动的敏感性. $\kappa(A)$ 愈大, $(A + E)^{-1}$ 与 A^{-1} 的相对误差就愈大.

定义 6.4.1 设 n 阶矩阵 A 非奇异, 则称 $\kappa(A) = \|A\| \|A^{-1}\|$ 为 A

关于求逆的条件数.

由推论 6.4.1 知, 如果 $\kappa(A)$ 很大, 则矩阵 A 关于求逆是病态的.

6.4.2 线性方程组解的扰动分析

对线性方程组

$$Ax = b \quad (6.4.8)$$

如果系数矩阵 A 和右端项 b 分别有扰动 E 和 δb , 则扰动后方程组为

$$(A + E)(x + \delta x) = b + \delta b \quad (6.4.9)$$

在什么条件下扰动后的方程组(6.4.9)有惟一解? 并估计解的相对误差.

定理 6.4.3 设 $A \in C^{m \times n}$ 是非奇异矩阵, $b, \delta b \in C^m$, x 是方程组(6.4.8)的解. 如果 $E \in C^{m \times n}$ 满足条件 $\|A^{-1}\| \|E\| < 1$, 则方程组(6.4.9)有惟一解 $x + \delta x$, 并且有

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{\kappa(A)}{1 - \kappa(A) \frac{\|E\|}{\|A\|}} \left(\frac{\|E\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|} \right) \quad (6.4.10)$$

其中 $\kappa(A) = \|A\| \|A^{-1}\|$, 向量范数 $\|\cdot\|$ 与矩阵范数 $\|\cdot\|$ 相容.

证明 因为 $\|A^{-1}\| \|E\| < 1$, 由推论 6.4.1 知 $A + E$ 非奇异, 因此方程组(6.4.9)有惟一解 $x + \delta x$. 下面证明估计式(6.4.10).

由(6.4.8)得 $x = A^{-1}b$, 而从(6.4.9)有

$$\begin{aligned} \delta x &= (A + E)^{-1}(b + \delta b) - x = (A + E)^{-1}b + (A + E)^{-1}\delta b - x \\ &= (A + E)^{-1}\delta b - x + (I + A^{-1}E)^{-1}A^{-1}b \\ &= (A + E)^{-1}\delta b - [I - (I + A^{-1}E)^{-1}]x \\ &= (A + E)^{-1}\delta b - (I + A^{-1}E)^{-1}A^{-1}Ex \end{aligned}$$

由 $\|A^{-1}E\| \leq \|A^{-1}\| \|E\|$, 定理 6.3.8 和定理 6.4.2 可得

$$\begin{aligned} \|\delta x\| &\leq \frac{\|A^{-1}\| \|\delta b\|}{1 - \|A^{-1}E\|} + \frac{\|A^{-1}E\| \|x\|}{1 - \|A^{-1}E\|} \\ &\leq \frac{\|A^{-1}\| \|x\|}{1 - \|A^{-1}\| \|E\|} \left(\frac{\|\delta b\|}{\|x\|} + \|E\| \right) \end{aligned}$$

因为 $\|b\| \leq \|A\| \|x\|$, 则由上式即得

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \|E\|} \left(\frac{\|\delta b\|}{\|x\|} + \|E\| \right)$$

$$\leq \frac{\kappa(A)}{1 - \kappa(A)} \frac{\|E\|}{\|A\|} \left(\frac{\|\delta b\|}{\|b\|} + \frac{\|E\|}{\|A\|} \right) \quad \square$$

估计式(6.4.10)表明, $\kappa(A)$ 反映了线性方程组 $Ax = b$ 的解 x 的相对误差对于 A 和 b 的相对误差的依赖程度. 因此, $\kappa(A)$ 也称为求解线性方程组 $Ax = b$ 的条件数. 由定理 6.4.3 知, 如果 $\kappa(A)$ 很大, 则线性方程组 $Ax = b$ 是病态的. 在求矩阵的逆或求解线性方程组时, 可以通过变换降低矩阵的条件数, 即所谓预处理或预条件.

例如, 对线性方程组

$$\begin{bmatrix} 1 & 10^5 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 10^5 \\ 2 \end{bmatrix}$$

因为 $A = \begin{bmatrix} 1 & 10^5 \\ 1 & 1 \end{bmatrix}$, 所以 $A^{-1} = \frac{1}{10^5 - 1} \begin{bmatrix} -1 & 10^5 \\ 1 & -1 \end{bmatrix}$. 取矩阵范数 $\|\cdot\|$ 为 $\|\cdot\|_\infty$, 则 $\kappa(A) = \frac{(1 + 10^5)^2}{10^5 - 1} \approx 10^5$. 因此矩阵 A 求逆和求解方程组 $Ax = b$ 都是病态的. 如果用 10^5 除方程组的第一个方程, 得 $Bx = c$, 即

$$\begin{bmatrix} 10^{-5} & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

其中 $B = \begin{bmatrix} 10^{-5} & 1 \\ 1 & 1 \end{bmatrix}$, 则 $\kappa(B) \approx 4$, 于是方程组 $Bx = c$ 是良态的.

6.4.3 矩阵特征值的扰动分析*

现在讨论矩阵元素的变化对矩阵特征值的影响. 因为多项式的根是其系数的连续函数, 而矩阵特征多项式的系数是矩阵元素的连续函数, 因此矩阵的特征值是矩阵元素的连续函数. 一般说来, 矩阵的特征向量不是矩阵元素的连续函数, 但是单特征值对应的特征向量是矩阵元素的连续函数.

定理 6.4.4 设 $A, E \in C^{n \times n}$, 如果 λ 是 A 的单特征值, x 是 A 对应于 λ 的特征向量, 则 $A + E$ 有特征值 $\lambda(E)$ 和相应的特征向量 $x(E)$ 使得

$$\text{当 } E \rightarrow 0 \text{ 时, } \lambda(E) \rightarrow \lambda, \quad x(E) \rightarrow x$$

证明 因为矩阵的特征值是矩阵元素的连续函数, 所以存在 $A + E$ 的特征值 $\lambda(E)$ 使得当 $E \rightarrow 0$, $\lambda(E) \rightarrow \lambda$, 并且对充分小的 $\|E\|$, $\lambda(E)$ 是 $A + E$ 的单特征值.

因为 λ 是 A 的单特征值, 则由 Jordan 标准形定理知 $\text{rank}(\lambda I - A) = n -$

1, 从而存在 i 和 j 使得 $\lambda I - A$ 删去第 i 行和第 j 列所得的 $n-1$ 阶矩阵非奇异. 于是从

$$Ax = \lambda x$$

其中 $x = (x_1, \dots, x_n)^T$, 可知线性方程组

$$-\sum_{\substack{m=1 \\ m \neq j}}^n (a_{km} - \delta_{km}\lambda)x_m = (a_{kj} - \delta_{kj}\lambda)x_j, \quad k \neq i \quad (6.4.11)$$

的系数矩阵非奇异. 不失一般性, 假定 $x_j = 1$.

因为对充分小的 $\|E\|$, $\lambda(E)$ 是 $A + E$ 的单特征值, 则 $\text{rank}[\lambda(E)I - (A + E)] = n-1$. 则由定理 6.4.2 知, $\lambda(E)I - (A + E)$ 删去第 i 行和第 j 列所得的 $n-1$ 阶矩阵也非奇异, 从而线性方程组

$$-\sum_{\substack{m=1 \\ m \neq j}}^n (a_{km} + e_{km} - \delta_{km}\lambda(E))x_m(E) = (a_{kj} + e_{kj} - \delta_{kj}\lambda(E))x_j(E), \quad k \neq i$$

有惟一解 $x_1(E), \dots, x_{j-1}(E), x_{j+1}(E), \dots, x_n(E)$, 并且它们是 E 的连续函数. 如果选取 $x_j(E) = x_j = 1$, 则

当 $E \rightarrow 0$ 时, $x(E) = (x_1(E), \dots, x_{j-1}(E), x_j(E), x_{j+1}(E), \dots, x_n(E))^T \rightarrow x$

□

虽然上面的连续性结果非常重要, 但它们并没有给出因矩阵元素变化而引起特征值有多大变化的任何信息. 下面给出这方面的一些结果.

定义 6.4.2 设 $A = (a_{ij}) \in \mathbb{C}^{n \times n}$, 令

$$G_i(A) = \{z \in \mathbb{C} \mid |z - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|\}, \quad i = 1, 2, \dots, n \quad (6.4.12)$$

$G_i(A) (i=1, 2, \dots, n)$ 是复平面上的 n 个圆盘, 称为矩阵 A 的 Gerschgorin 圆盘.

定理 6.4.5 (Gerschgorin 定理) 设 $A = (a_{ij}) \in \mathbb{C}^{n \times n}$, 则

$$(1) \lambda(A) \subseteq \bigcup_{i=1}^n G_i(A);$$

(2) 如果 A 的 Gerschgorin 圆盘中有 m 个互相连通, 并且与其余的 $n-m$ 个圆盘不连通, 则此 m 个圆盘构成的连通区域中恰有 A 的 m 个特征值.

证明 (1) 设 λ 是矩阵 A 的任一特征值, 相应的特征向量为 $x = (x_1, \dots, x_n)^T$. 令 $|x_{i_0}| = \max_i |x_i|$, 则 $|x_{i_0}| > 0$. 由 $Ax = \lambda x$ 有

$$(\lambda - a_{i_0 i_0})x_{i_0} = \sum_{\substack{j=1 \\ j \neq i_0}}^n a_{i_0 j} x_j$$

从而

$$|\lambda - a_{i_0 i_0}| \leq \sum_{\substack{j=1 \\ j \neq i_0}}^n |a_{i_0 j}| \frac{|x_j|}{|x_{i_0}|} \leq \sum_{\substack{j=1 \\ j \neq i_0}}^n |a_{i_0 j}|$$

即 $\lambda \in G_{i_0}(A)$, 故 $\lambda(A) \subseteq \bigcup_{i=1}^n G_i(A)$.

(2) 记 $D = \text{diag}(a_{11}, \dots, a_{nn})$, $C = A - D$. 令

$$A(t) = D + tC, \quad 0 \leq t \leq 1$$

不失一般性, 假定(6.4.12)中的圆盘 $G_1(A), \dots, G_m(A)$ 构成一连通区域 $G_1 = \bigcup_{i=1}^m G_i(A)$, 并与其余的 $n - m$ 个圆盘分离, 即 G_1 与 $G_2 = \bigcup_{i=m+1}^n G_i(A)$ 不相交. 记

$$G_i(A(t)) = \{z \in \mathbb{C} \mid |z - a_{ii}| \leq t \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|\}, \quad i = 1, 2, \dots, n$$

$$G_1(t) = \bigcup_{i=1}^m G_i(A(t)), \quad G_2(t) = \bigcup_{i=m+1}^n G_i(A(t))$$

则 $G_i(A(t)) \subseteq G_i(A) (i = 1, 2, \dots, n)$, $G_1(t) \subseteq G_1, G_2(t) \subseteq G_2$, 且对所有 $t \in [0, 1]$, $G_1(t)$ 与 $G_2(t)$ 都不相交. 特别地, $G_1(0)$ 恰包含 $A(0)$ 的 m 个特征值 a_{11}, \dots, a_{mm} . 由(1)知对所有 $t \in [0, 1]$, $A(t)$ 的特征值都包含在 $G_1(t) \cup G_2(t)$ 中. 因为 $G_1(t)$ 与 $G_2(t)$ 不相交, 而矩阵的特征值是矩阵元素的连续函数, 则当 t 增加时 $A(t)$ 的特征值不能从 $G_1(t)$ “跳跃”到 $G_2(t)$. 因为 $G_1(0)$ 恰包含 $A(0)$ 的 m 个特征值, 所以对所有 $t \in [0, 1]$, $G_1(t)$ 恰包含 $A(t)$ 的 m 个特征值. 因此 $G_1(1)$ (即 G_1) 恰包含 $A(1)$ (即 A) 的 m 个特征值. \square

设 $A = (a_{ij}) \in \mathbb{C}^{n \times n}$, 令 $D = \text{diag}(d_1, \dots, d_n)$, 其中 d_1, \dots, d_n 均为正数, 则 $B = DAD^{-1}$ 与 A 相似, 并且 B 与 A 具有相同的对角元. 适当选取对角矩阵 D , 对 B 应用 Gerschgorin 定理有时能够得到更精确的特征值的包含区域.

例 6.4.1 设

$$A = \begin{bmatrix} 20 & 3 & 2.8 \\ 4 & 10 & 2 \\ 0.5 & 1 & 10i \end{bmatrix}$$

隔离矩阵 A 的特征值.

解 A 的 3 个 Gerschgorin 圆盘为

$$G_1(A): |z - 20| \leq 5.8, \quad G_2(A): |z - 10| \leq 6, \quad G_3(A): |z - 10i| \leq 1.5$$

因为 $G_3(A)$ 孤立, 所以 $G_3(A)$ 恰包含 A 的一个特征值, 记为 λ_3 .

选取 $D = \text{diag}(1, 1, 4)$, 则

$$B = DAD^{-1} = \begin{bmatrix} 20 & 3 & 0.7 \\ 4 & 10 & 0.5 \\ 2 & 4 & 10i \end{bmatrix}$$

B 的 3 个 Gerschgorin 圆盘为

$$G_1(B): |z - 20| \leq 3.7, \quad G_2(B): |z - 10| \leq 4.5, \quad G_3(B): |z - 10i| \leq 6$$

容易看出, 这是 3 个孤立的 Gerschgorin 圆盘. 因此每个 Gerschgorin 圆盘中恰有 B (即 A) 的一个特征值. 因为 $G_3(B)$ 中的特征值就是 $G_3(A)$ 中的特征值 λ_3 , 所以 A 的 3 个特征值分别位于 $G_1(B)$, $G_2(B)$ 和 $G_3(A)$ 中.

定理 6.4.6 设 $A, E \in C^{n \times n}$, 并且存在可逆矩阵 P 使得

$$P^{-1}AP = \text{diag}(\lambda_1, \dots, \lambda_n) \quad (6.4.13)$$

$A + E$ 有特征值 μ_1, \dots, μ_n , 则对任一 μ_j , 存在 λ_i 使得

$$|\lambda_i - \mu_j| \leq \|P^{-1}EP\|_{\infty} \quad (6.4.14)$$

此外, 如果 λ_i 是 m 重特征值, 且圆盘

$$R_i = \{z \in C \mid |z - \lambda_i| \leq \|P^{-1}EP\|_{\infty}\} \quad (6.4.15)$$

与圆盘

$$R_k = \{z \in C \mid |z - \lambda_k| \leq \|P^{-1}EP\|_{\infty}\}, \quad \lambda_k \neq \lambda_i \quad (6.4.16)$$

不相交, 则 R_i 恰包含 $A + E$ 的 m 个特征值.

证明 令 $C = P^{-1}(A + E)P$, 则 C 的特征值为 μ_1, \dots, μ_n . 记 $P^{-1}EP = (b_{ij}) \in C^{n \times n}$, 则 C 的对角元是 $\lambda_k + b_{kk}$. 由 Gerschgorin 定理知存在 i 使得

$$|\lambda_i + b_{ii} - \mu_j| \leq \sum_{\substack{k=1 \\ k \neq i}}^n |b_{ik}|$$

因此

$$|\lambda_i - \mu_j| \leq \sum_{k=1}^n |b_{ik}| \leq \|P^{-1}EP\|_{\infty}$$

从 Gerschgorin 定理的第二个结论可直接推得定理 6.4.6 的第二部分结论. \square

定理 6.4.7 设 $A, E \in E^{n \times n}$, 并且存在可逆矩阵 P 使得

$$P^{-1}AP = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n) \quad (6.4.17)$$

则对 $A + E$ 的任一特征值 μ , 有

$$\min_i |\lambda_i - \mu| \leq \|P^{-1}EP\| \quad (6.4.18)$$

其中 $\|\cdot\|$ 是满足 $\|\text{diag}(d_1, \dots, d_n)\| = \max_{1 \leq i \leq n} |d_i|$ 的任何相容矩阵范数.

证明 令 $C = P^{-1}(A + E)P = \Lambda + B$, 其中 $B = P^{-1}EP = (b_{ij}) \in C^{n \times n}$. 对 $A + E$ 的任一特征值 μ , 若 $\Lambda - \mu I$ 奇异, 则存在某个 i 使得 $\mu = \lambda_i$. 于是定理得证.

下面假定 $\Lambda - \mu I$ 非奇异. 因为 $\Lambda - \mu I + B$ 奇异, 所以存在非零向量 $y \in C^n$ 使得

$$(\Lambda - \mu I)y = -By$$

则

$$y = (\Lambda - \mu I)^{-1}By$$

从而

$$\|y\| \leq \|(\Lambda - \mu I)^{-1}\| \|B\| \|y\| = (\min_i |\lambda_i - \mu|)^{-1} \|B\| \|y\|$$

因此

$$\min_i |\lambda_i - \mu| \leq \|B\| = \|P^{-1}EP\| \quad \square$$

由(6.4.18)可得

$$\min_i |\lambda_i - \mu| \leq \|P^{-1}\| \|P\| \|E\| \quad (6.4.19)$$

上式表明, $\|P^{-1}\| \|P\|$ 反映了矩阵 A 的特征值对扰动 E 的敏感性. 若 $\|P^{-1}\| \|P\|$ 不很大, 则 A 的特征值问题是良态的.

定义 6.4.3 设 $A \in C^{n \times n}$, 并且存在可逆矩阵 P 使得 $P^{-1}AP = \text{diag}(\lambda_1, \dots, \lambda_n)$, 则称 $\|P^{-1}\| \|P\|$ 为矩阵 A 关于特征值问题的条件数, 记为 $\kappa(P)$.

如果矩阵 A 是 Hermite 矩阵, 并且其扰动矩阵 E 也是 Hermite 矩阵, 则由定理 5.4.5 可得如下结论.

定理 6.4.8 设 A, E 均为 n 阶 Hermite 矩阵, $B = A + E$, 且 A, B 和 E 的特征值分别为 $\lambda_1 \geq \dots \geq \lambda_n$, $\mu_1 \geq \dots \geq \mu_n$ 和 $\epsilon_1 \geq \dots \geq \epsilon_n$, 则

$$|\lambda_i - \mu_i| \leq \|E\|_2, \quad i = 1, 2, \dots, n \quad (6.4.20)$$

证明 因为 E 是 Hermite 矩阵, 则 $\|E\|_2 = \max\{|\epsilon_1|, |\epsilon_n|\}$. 于是, 由定理 5.4.5 即得(6.4.20). \square

由定理 6.4.8 可知, Hermite 矩阵的特征值问题是良态的.

习 题

1. 设 $\|\cdot\|$ 是赋范线性空间 V 上的向量范数, 证明: 对任意 $\alpha, \beta \in V$, 有

$$|\|\alpha\| - \|\beta\|| \leq \|\alpha - \beta\|$$

2. 在 n 维向量空间 C^n 中, 对任意向量 $x = (x_1, \dots, x_n)^T \in C^n$, 定义

$$\|x\|_1 = \sum_{i=1}^n |x_i|$$

$$\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}}$$

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$$

证明: $\|x\|_1$, $\|x\|_2$ 和 $\|x\|_\infty$ 都是 C^n 上的向量范数, 并且

$$\|x\|_\infty \leq \|x\|_1 \leq n \|x\|_\infty$$

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{n} \|x\|_2$$

$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n} \|x\|_\infty$$

3. 设 A 是 n 阶 Hermite 正定矩阵, 对向量 $x \in C^n$, 定义 $\|x\| = (x^H A x)^{\frac{1}{2}}$.

(1) 证明: $\|x\|$ 是 C^n 上的向量范数;

(2) 当 $A = \text{diag}(d_1, d_2, \dots, d_n) > 0$ 时, 具体写出 $\|x\|$ 的表达式.

4. 设 $A = (a_{ij}) \in C^{m \times n}$, 证明: $\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 = \text{tr}(A^H A)$.

5. 设 $A = (a_{ij}) \in C^{m \times n}$, 令 $\|A\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{\frac{1}{2}}$. 证明: $\|\cdot\|_F$ 是 $C^{m \times n}$ 上的矩阵范数.

6. 设 $A = (a_{ij}) \in C^{m \times n}$, 令 $\|A\| = n \max_{i,j} |a_{ij}|$. 证明: $\|\cdot\|$ 是 $C^{m \times n}$ 上的相容矩阵范数.

7. 设 $\|\cdot\|$ 是 $C^{n \times n}$ 上的相容矩阵范数, 证明:

(1) $\|I\| \geq 1$;

(2) 设 A 是 n 阶可逆矩阵, λ 为 A 的任一特征值, 则 $\|A^{-1}\|^{-1} \leq |\lambda| \leq \|A\|$.

8. 设 $\|\cdot\|_\alpha$ 是 $C^{n \times n}$ 上的相容矩阵范数, D 是可逆矩阵. 对任意矩阵 $A \in C^{n \times n}$, 令 $\|A\|_\beta = \|D^{-1} A D\|_\alpha$. 证明: $\|\cdot\|_\beta$ 是 $C^{n \times n}$ 上的相容矩阵范数.

9. 设 $\|\cdot\|_\alpha$ 是 $C^{n \times n}$ 上的相容矩阵范数, D 是可逆矩阵, 且 $\|D^{-1}\|_\alpha < 1$. 对任意矩阵 $A \in C^{n \times n}$, 令 $\|A\|_\beta = \|DA\|_\alpha$. 证明: $\|\cdot\|_\beta$ 是 $C^{n \times n}$ 上的相容矩阵范数.

10. 对下列矩阵 A , 求 $\|A\|_1$, $\|A\|_2$, $\|A\|_\infty$, $\|A\|_F$

$$(1) A = \begin{pmatrix} -1 & -1 & 4 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{pmatrix} \quad (2) A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ -2 & 1 \end{pmatrix}$$

11. 设 $\|\cdot\|$ 是 $C^{m \times n}$ 上的算子范数, 证明:

(1) $\|I\| = 1$;

(2) 设 A 是 n 阶可逆矩阵, 则 $\|A^{-1}\| \geq \|A\|^{-1}$, 并且 $\|A^{-1}\|^{-1} = \min_{x \neq 0} \frac{\|Ax\|}{\|x\|}$.

12. 设 $A = (a_{ij}) \in C^{m \times n}$, 证明:

(1) $\|A\|_2 \leq \|A\|_F \leq \sqrt{\text{rank}(A)} \|A\|_2$;

(2) $\frac{1}{n} \|A\|_\infty \leq \|A\|_1 \leq m \|A\|_\infty$;

(3) $\frac{1}{\sqrt{n}} \|A\|_\infty \leq \|A\|_2 \leq \sqrt{m} \|A\|_\infty$.

13. 设 $A = (a_{ij}) \in C^{m \times n}$, $B = (b_{ij}) \in C^{n \times p}$, 证明:

$$\|AB\|_F \leq \min\{\|A\|_F \|B\|_2, \|A\|_2 \|B\|_F\}$$

14. 设 $A = (a_{ij}) \in C^{n \times n}$, $\lambda_1, \lambda_2, \dots, \lambda_n$ 是 A 的特征值, 证明:

$$\sum_{i=1}^n |\lambda_i|^2 \leq \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2$$

并且等号成立的充分必要条件是 A 为正规矩阵.

15. 设 A 是 n 阶矩阵, 证明: $\rho(A) < 1$ 的充分必要条件是存在 n 阶 Hermite 正定矩阵 Q 使得 $Q - AQA^H > 0$.

16. 设 A 是 n 阶矩阵, 如果 A 的特征值实部均为负数, 则称 A 为稳定矩阵. 利用上题结论证明: $-A$ 为稳定矩阵的充分必要条件是存在 Hermite 正定矩阵 M 使得 $AM + MA^H > 0$.

17. 构造一个收敛的 2 阶可逆矩阵序列, 但它的极限矩阵不可逆.

18. 讨论矩阵幂级数 $\sum_{k=1}^{\infty} \frac{1}{k^2} \begin{bmatrix} 1 & 4 \\ -1 & -3 \end{bmatrix}^k$ 的收敛性.

19. 设 A 是 n 阶矩阵, 并且 $\rho(A) < 1$, 试求 $\sum_{k=1}^{\infty} kA^k$.

20. 已知 $A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$, $E = \begin{bmatrix} 0 & 0.5 \\ 0.2 & 0 \end{bmatrix}$, 试估计 $\frac{\|A^{-1} - (A+E)^{-1}\|_\infty}{\|A^{-1}\|_\infty}$ 的上界.

21. 设 A 是 n 阶可逆矩阵, x^* 是 $Ax = b$ 的精确解, \hat{x} 是 $Ax = b$ 的近似解, $r = b - A\hat{x}$, 证明

$$\frac{\|x^* - \hat{x}\|}{\|x^*\|} \leq \kappa(A) \frac{\|r\|}{\|b\|}$$

其中 $\kappa(A) = \|A\| \|A^{-1}\|$, 并且向量范数与矩阵范数相容.

22. 设 $A = (a_{ij}) \in C^{n \times n}$, 如果

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|, \quad i = 1, 2, \dots, n$$

则称 A 为严格对角占优矩阵. 证明: 严格对角占优矩阵一定非奇异.

23. 设 $A = (a_{ij}) \in C^{n \times n}$ 为严格对角占优矩阵, 并且其对角元均为正数. 证明 A 的特征值都在复平面的右半平面内 (即 $\operatorname{Re}(\lambda) > 0$).

24. 设 A 为 n 阶可对角化矩阵, $(A - \mu I)v = r$, 其中 $\|v\|_2 = 1$. 证明存在 A 的特征值 λ 使得

$$|\lambda - \mu| \leq \|P\|_2 \|P^{-1}\|_2 \|r\|_2$$

其中 P 满足 $P^{-1}AP = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$.

25. 设 A 为 n 阶非奇异矩阵, 为了用迭代法求线性方程组

$$Ax = b$$

的解 x^* , 经常把 $Ax = b$ 写成等价的形式:

$$x = Gx + g$$

其中 G 是 n 阶矩阵, g 是 n 维向量. 选取初始向量 $x^{(0)}$, 构造迭代格式

$$x^{(k+1)} = Gx^{(k)} + g \quad k = 0, 1, 2, \dots$$

证明: (1) $\{x^{(k)}\}$ 收敛于 x^* 的充分必要条件是 $\rho(G) < 1$;

(2) 如果对某一相容矩阵范数 $\|\cdot\|$ 有 $\|G\| < 1$, 则

$$\|x^* - x^{(k)}\| \leq \frac{\|G\|}{1 - \|G\|} \|x^{(k)} - x^{(k-1)}\|;$$

$$\|x^* - x^{(k)}\| \leq \frac{\|G\|^k}{1 - \|G\|} \|x^{(1)} - x^{(0)}\|.$$

第七章 矩阵函数与矩阵值函数

矩阵函数与矩阵值函数是矩阵理论的重要内容,它们在力学、控制理论、信号处理等学科中具有重要应用.本章介绍矩阵函数与矩阵值函数的概念,讨论它们的有关性质以及矩阵值函数的微分与积分,并研究它们在微分方程组中的应用.

§ 7.1 矩阵函数

矩阵函数的概念与通常的函数概念类似,所不同的是这里的自变量和因变量都是 n 阶矩阵.本节首先以定理 6.3.7 与矩阵幂级数的和为依据,给出矩阵函数的幂级数表示,并讨论矩阵指数函数和矩阵三角函数的性质.其次以 Hermite 多项式插值为基础,给出矩阵函数的多项式表示.

7.1.1 矩阵函数的幂级数表示

受高等数学或复变函数的启发,我们可以利用矩阵幂级数来定义矩阵函数.

定义 7.1.1 设 $A \in C^{n \times n}$, 一元函数 $f(z)$ 能够展开为 z 的幂级数

$$f(z) = \sum_{k=0}^{\infty} c_k z^k \quad (7.1.1)$$

并且该幂级数的收敛半径为 R . 当矩阵 A 的谱半径 $\rho(A) < R$ 时,则将收敛矩阵幂级数 $\sum_{k=0}^{\infty} c_k A^k$ 的和定义为矩阵函数,记为 $f(A)$,即

$$f(A) = \sum_{k=0}^{\infty} c_k A^k \quad (7.1.2)$$

因为当 $|z| < +\infty$ 时,有

$$e^z = 1 + z + \frac{1}{2!} z^2 + \cdots + \frac{1}{n!} z^n + \cdots$$

$$\sin z = z - \frac{1}{3!} z^3 + \frac{1}{5!} z^5 - \cdots + (-1)^n \frac{1}{(2n+1)!} z^{2n+1} + \cdots$$

$$\cos z = 1 - \frac{1}{2!} z^2 + \frac{1}{4!} z^4 - \cdots + (-1)^n \frac{1}{(2n)!} z^{2n} + \cdots$$

则由推论 6.3.2 可知, 对任意 $A \in C^{n \times n}$, 矩阵幂级数

$$I + A + \frac{1}{2!}A^2 + \cdots + \frac{1}{n!}A^n + \cdots$$

$$A - \frac{1}{3!}A^3 + \frac{1}{5!}A^5 - \cdots + (-1)^n \frac{1}{(2n+1)!}A^{2n+1} + \cdots$$

$$I - \frac{1}{2!}A^2 + \frac{1}{4!}A^4 - \cdots + (-1)^n \frac{1}{(2n)!}A^{2n} + \cdots$$

都是收敛的. 它们的和分别记为 $e^A, \sin A, \cos A$, 即

$$e^A = I + A + \frac{1}{2!}A^2 + \cdots + \frac{1}{n!}A^n + \cdots \quad (7.1.3)$$

$$\sin A = A - \frac{1}{3!}A^3 + \frac{1}{5!}A^5 - \cdots + (-1)^n \frac{1}{(2n+1)!}A^{2n+1} + \cdots \quad (7.1.4)$$

$$\cos A = I - \frac{1}{2!}A^2 + \frac{1}{4!}A^4 - \cdots + (-1)^n \frac{1}{(2n)!}A^{2n} + \cdots \quad (7.1.5)$$

通常称 e^A 为矩阵指数函数, $\sin A$ 和 $\cos A$ 为矩阵三角函数.

由 (7.1.3) ~ (7.1.5) 可以推得下面一组等式

$$\begin{cases} e^{iA} = \cos A + i \sin A \\ \cos A = \frac{1}{2}(e^{iA} + e^{-iA}) \\ \sin A = \frac{1}{2i}(e^{iA} - e^{-iA}) \\ \cos(-A) = \cos A \\ \sin(-A) = -\sin A \end{cases} \quad (7.1.6)$$

其中 $i = \sqrt{-1}$.

必须指出, 指数函数的运算规则 $e^a e^b = e^b e^a = e^{a+b}$ 对矩阵指数函数一般不再成立. 例如, 令

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

则 $A = A^2 = A^3 = \cdots, B = B^2 = B^3 = \cdots$, 并且

$$A + B = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}, \quad (A + B)^k = 2^{k-1}(A + B), \quad k \geq 1$$

于是

$$e^A = I + (e-1)A = \begin{pmatrix} e & 1-e \\ 0 & 1 \end{pmatrix}$$

$$e^B = I + (e-1)B = \begin{pmatrix} e & e-1 \\ 0 & 1 \end{pmatrix}$$

因此

$$e^A e^B = \begin{pmatrix} e^2 & (e-1)^2 \\ 0 & 1 \end{pmatrix}, \quad e^B e^A = \begin{pmatrix} e^2 & -(e-1)^2 \\ 0 & 1 \end{pmatrix}$$

$$e^{A+B} = I + \frac{1}{2}(e^2-1)(A+B) = \begin{pmatrix} e^2 & 0 \\ 0 & 1 \end{pmatrix}$$

可见 $e^A e^B, e^B e^A, e^{A+B}$ 互不相等.

定理 7.1.1 设 $A, B \in C^{n \times n}$, 如果 $AB=BA$, 则

$$e^A e^B = e^B e^A = e^{A+B} \quad (7.1.7)$$

证明 因为矩阵的加法满足交换律, 所以只需证明 $e^A e^B = e^{A+B}$ 即可.

$$\begin{aligned} e^A e^B &= (I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots)(I + B + \frac{1}{2!}B^2 + \frac{1}{3!}B^3 + \cdots) \\ &= I + (A+B) + \frac{1}{2!}(A^2 + AB + BA + B^2) \\ &\quad + \frac{1}{3!}(A^3 + 3A^2B + 3AB^2 + B^3) + \cdots \\ &= I + (A+B) + \frac{1}{2!}(A+B)^2 + \frac{1}{3!}(A+B)^3 + \cdots = e^{A+B} \quad \square \end{aligned}$$

推论 7.1.1 设 $A \in C^{n \times n}$, 则

- (1) $e^A e^{-A} = e^{-A} e^A = I, (e^A)^{-1} = e^{-A}$;
- (2) 设 m 为整数, 则 $(e^A)^m = e^{mA}$.

定理 7.1.2 设 $A, B \in C^{n \times n}$, 则

- (1) $\sin^2 A + \cos^2 A = I$;
- (2) 如果 $AB=BA$, 则

$$\begin{cases} \sin(A+B) = \sin A \cos B + \cos A \sin B \\ \sin 2A = 2 \sin A \cos A \\ \cos(A+B) = \cos A \cos B - \sin A \sin B \\ \cos 2A = \cos^2 A - \sin^2 A \end{cases} \quad (7.1.8)$$

证明 这里仅证(7.1.8)中的第一个等式,其余证明留给读者.由(7.1.6)有

$$\begin{aligned}\sin(\mathbf{A} + \mathbf{B}) &= \frac{1}{2i}(e^{i(\mathbf{A}+\mathbf{B})} - e^{-i(\mathbf{A}+\mathbf{B})}) = \frac{1}{2i}(e^{i\mathbf{A}}e^{i\mathbf{B}} - e^{-i\mathbf{A}}e^{-i\mathbf{B}}) \\ &= \frac{1}{2i}(e^{i\mathbf{A}} - e^{-i\mathbf{A}}) \frac{1}{2}(e^{i\mathbf{B}} + e^{-i\mathbf{B}}) + \frac{1}{2}(e^{i\mathbf{A}} + e^{-i\mathbf{A}}) \frac{1}{2i}(e^{i\mathbf{B}} - e^{-i\mathbf{B}}) \\ &= \sin \mathbf{A} \cos \mathbf{B} + \cos \mathbf{A} \sin \mathbf{B} \quad \square\end{aligned}$$

例 7.1.1 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 且 $\rho(\mathbf{A}) < 1$, 对函数 $f(z) = \ln(1+z)$ ($|z| < 1$), 求矩阵函数 $f(\mathbf{A})$.

解 因为当 $|z| < 1$ 时有

$$\ln(1+z) = z - \frac{1}{2}z^2 + \frac{1}{3}z^3 - \cdots + (-1)^{n+1} \frac{1}{n}z^n + \cdots$$

而 $\rho(\mathbf{A}) < 1$, 则由定理 6.3.7 可得

$$f(\mathbf{A}) = \ln(\mathbf{I} + \mathbf{A}) = \mathbf{A} - \frac{1}{2}\mathbf{A}^2 + \frac{1}{3}\mathbf{A}^3 - \cdots + (-1)^{n+1} \frac{1}{n}\mathbf{A}^n + \cdots$$

对矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 和函数 $f(z) = \sum_{k=0}^{\infty} c_k z^k$, 下面讨论如何计算矩阵函数 $f(\mathbf{A})$.

如果矩阵 \mathbf{A} 有 Jordan 标准形 \mathbf{J} , 即存在可逆矩阵 \mathbf{P} 使得

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{J} = \text{diag}(\mathbf{J}_1, \cdots, \mathbf{J}_l) \quad (7.1.9)$$

其中

$$\mathbf{J}_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ & & & \lambda_i \end{pmatrix}_{r_i \times n_i} \quad (7.1.10)$$

则

$$f(\mathbf{A}) = \sum_{k=0}^{\infty} c_k \mathbf{A}^k = \mathbf{P} \left(\sum_{k=0}^{\infty} c_k \mathbf{J}^k \right) \mathbf{P}^{-1}$$

$$= \mathbf{P} \begin{bmatrix} \sum_{k=0}^{\infty} c_k \mathbf{J}_1^k & & \\ & \ddots & \\ & & \sum_{k=0}^{\infty} c_k \mathbf{J}_r^k \end{bmatrix} \mathbf{P}^{-1} = \mathbf{P} \begin{bmatrix} f(\mathbf{J}_1) & & \\ & \ddots & \\ & & f(\mathbf{J}_r) \end{bmatrix} \mathbf{P}^{-1} \quad (7.1.11)$$

其中

$$f(\mathbf{J}_i) = \sum_{k=0}^{\infty} c_k \mathbf{J}_i^k = \begin{bmatrix} \sum_{k=0}^{\infty} c_k \lambda_i^k & \frac{1}{1!} \sum_{k=1}^{\infty} c_k C_k^1 \lambda_i^{k-1} & \cdots & \frac{1}{(n_i-1)!} \sum_{k=n_i-1}^{\infty} c_k C_k^{n_i-1} \lambda_i^{k-n_i+1} \\ & \sum_{k=0}^{\infty} c_k \lambda_i^k & \ddots & \vdots \\ & & \ddots & \frac{1}{1!} \sum_{k=1}^{\infty} c_k C_k^1 \lambda_i^{k-1} \\ & & & \sum_{k=0}^{\infty} c_k \lambda_i^k \end{bmatrix} \\ = \begin{bmatrix} f(\lambda_i) & \frac{1}{1!} f'(\lambda_i) & \cdots & \frac{1}{(n_i-1)!} f^{(n_i-1)}(\lambda_i) \\ & f(\lambda_i) & \ddots & \vdots \\ & & \ddots & \frac{1}{1!} f'(\lambda_i) \\ & & & f(\lambda_i) \end{bmatrix} \quad (7.1.12)$$

例 7.1.2 设

$$\mathbf{A} = \begin{bmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{bmatrix}$$

求 $e^{\mathbf{A}}$, e^{A_i} 和 $\sin \mathbf{A}$.

解 由例 3.5.2 知, 存在可逆矩阵 $\mathbf{P} = \begin{pmatrix} -1 & 2 & 2 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$ 使得 $\mathbf{P}^{-1} \mathbf{A} \mathbf{P} =$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

当 $f(x) = e^x$ 时, $f(1) = e$, $f'(1) = e$, 则

$$e^{\mathbf{A}} = \mathbf{P} \begin{pmatrix} e & 0 & 0 \\ 0 & e & e \\ 0 & 0 & e \end{pmatrix} \mathbf{P}^{-1} = \begin{pmatrix} -e & -2e & 6e \\ -e & 0 & 3e \\ -e & -e & 4e \end{pmatrix}$$

当 $f(x) = e^{tx}$ 时, $f(1) = e^t$, $f'(1) = te^t$, 则

$$e^{\mathbf{A}t} = \mathbf{P} \begin{pmatrix} e^t & 0 & 0 \\ 0 & e^t & te^t \\ 0 & 0 & e^t \end{pmatrix} \mathbf{P}^{-1} = \begin{pmatrix} (1-2t)e^t & -2te^t & 6te^t \\ -te^t & (1-t)e^t & 3te^t \\ -te^t & -te^t & (1+3t)e^t \end{pmatrix}$$

当 $f(x) = \sin x$ 时, $f(1) = \sin 1$, $f'(1) = \cos 1$, 则

$$\sin \mathbf{A} = \mathbf{P} \begin{pmatrix} \sin 1 & 0 & 0 \\ 0 & \sin 1 & \cos 1 \\ 0 & 0 & \sin 1 \end{pmatrix} \mathbf{P}^{-1} = \begin{pmatrix} \sin 1 - 2\cos 1 & -2\cos 1 & 6\cos 1 \\ -\cos 1 & \sin 1 - \cos 1 & 3\cos 1 \\ -\cos 1 & -\cos 1 & \sin 1 + 3\cos 1 \end{pmatrix}$$

7.1.2 矩阵函数的另一种定义

利用定理 6.3.7 和推论 6.3.2 定义矩阵函数, 其实质就是先将函数 $f(z)$ 展开成 z 的收敛幂级数, 再将 z 代以矩阵 \mathbf{A} 来定义矩阵函数 $f(\mathbf{A})$. 但这个条件比较强, 一般不易满足. 下面我们拓宽矩阵函数的定义.

对矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 假定存在 n 阶可逆矩阵 \mathbf{P} 使得

$$\mathbf{P}^{-1} \mathbf{A} \mathbf{P} = \mathbf{J} = \text{diag}(\mathbf{J}_1, \dots, \mathbf{J}_s) \quad (7.1.13)$$

其中 \mathbf{J}_i 是 (7.1.10) 定义的 Jordan 块, 则对任意多项式 $p(\lambda)$, 有

$$p(\mathbf{A}) = \mathbf{P} p(\mathbf{J}) \mathbf{P}^{-1} = \mathbf{P} \text{diag}(p(\mathbf{J}_1), \dots, p(\mathbf{J}_s)) \mathbf{P}^{-1} \quad (7.1.14)$$

其中

$$p(\mathbf{J}_i) = \begin{bmatrix} p(\lambda_i) & \frac{1}{1!}p'(\lambda_i) & \cdots & \frac{1}{(n_i-1)!}p^{(n_i-1)}(\lambda_i) \\ & p(\lambda_i) & \ddots & \vdots \\ & & \ddots & \frac{1}{1!}p'(\lambda_i) \\ & & & p(\lambda_i) \end{bmatrix}_{n_i \times n_i} \quad (7.1.15)$$

(7.1.14)表明, $p(\mathbf{A})$ 与 \mathbf{A} 的 Jordan 标准形结构以及 $p(\lambda)$ 在 \mathbf{A} 的特征值处的函数值与各阶导数值有关.

设矩阵 \mathbf{A} 的最小多项式为

$$m(\lambda) = (\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_k)^{m_k} \quad (7.1.16)$$

其中 $\lambda_1, \lambda_2, \dots, \lambda_k$ 为 \mathbf{A} 的 k 个互异特征值. 对任意函数 $f(z)$, 如果

$$f(\lambda_i), f'(\lambda_i), \dots, f^{(m_i-1)}(\lambda_i), i = 1, 2, \dots, k$$

存在, 则称函数 $f(z)$ 在 \mathbf{A} 的谱 $\lambda(\mathbf{A})$ 有定义, 并称 $f(\lambda_i), f'(\lambda_i), \dots, f^{(m_i-1)}(\lambda_i) (i=1, 2, \dots, k)$ 为 $f(z)$ 在 \mathbf{A} 的谱 $\lambda(\mathbf{A})$ 上的值.

定理 7.1.3 设 $\mathbf{A} \in \mathbb{C}^{m \times n}$, $p_1(\lambda)$ 和 $p_2(\lambda)$ 是两个多项式, 则 $p_1(\mathbf{A}) = p_2(\mathbf{A})$ 的充分必要条件是 $p_1(\lambda)$ 和 $p_2(\lambda)$ 在 \mathbf{A} 的谱 $\lambda(\mathbf{A})$ 上具有相同的值.

证明 记 $p_0(\lambda) = p_1(\lambda) - p_2(\lambda)$. 如果 $p_1(\mathbf{A}) = p_2(\mathbf{A})$, 则 $p_0(\mathbf{A}) = 0$, 从而 $p_0(\lambda)$ 是 \mathbf{A} 的一个化零多项式. 由定理 3.6.2 知存在一个多项式 $q(\lambda)$ 使得 $p_0(\lambda) = q(\lambda)m(\lambda)$. 容易计算

$$p_1^{(j)}(\lambda_i) - p_2^{(j)}(\lambda_i) = p_0^{(j)}(\lambda_i) = 0, j = 0, 1, \dots, m_i - 1; i = 1, 2, \dots, k \quad (7.1.17)$$

于是 $p_1(\lambda)$ 和 $p_2(\lambda)$ 在 \mathbf{A} 的谱 $\lambda(\mathbf{A})$ 上具有相同的值.

反之, 如果 (7.1.17) 成立, 则 $\lambda_i (i=1, 2, \dots, k)$ 是 $p_0(\lambda)$ 的至少 m_i 重零点, 从而存在一个多项式 $q(\lambda)$ 使得 $p_0(\lambda) = q(\lambda)m(\lambda)$. 因为 $m(\mathbf{A}) = 0$, 所以 $p_0(\mathbf{A}) = 0$, 即 $p_1(\mathbf{A}) = p_2(\mathbf{A})$. \square

现在利用多项式给出矩阵函数的另一种定义.

定义 7.1.2 设矩阵 $\mathbf{A} \in \mathbb{C}^{m \times n}$ 的最小多项式为 (7.1.16), 函数 $f(z)$ 在 \mathbf{A} 在谱 $\lambda(\mathbf{A})$ 上有定义, 如果存在多项式 $p(\lambda)$ 满足

$$p^{(j)}(\lambda_i) = f^{(j)}(\lambda_i), i = 1, 2, \dots, k; j = 0, 1, \dots, m_i - 1 \quad (7.1.18)$$

则定义矩阵函数 $f(\mathbf{A})$ 为

$$f(\mathbf{A}) \equiv p(\mathbf{A}) \quad (7.1.19)$$

一个自然的问题是满足条件(7.1.18)的多项式 $p(\lambda)$ 是否存在? 它与 Hermite 多项式插值问题密切相关. 对此, 我们有如下结果.

定理 7.1.4 设 $\lambda_1, \lambda_2, \dots, \lambda_k$ 是 k 个互异数, m_1, m_2, \dots, m_k 是 k 个正整数且 $m = \sum_{i=1}^k m_i$. 给定一组数

$$f_{i,0}, f_{i,1}, \dots, f_{i,m_i-1}, i = 1, 2, \dots, k$$

则存在次数小于 m 的多项式 $p(\lambda)$ 使得

$$p^{(j)}(\lambda_i) = f_{i,j}, i = 1, 2, \dots, k; j = 0, 1, \dots, m_i - 1 \quad (7.1.20)$$

证明 对 $1 \leq i \leq k$, 令 $p_i(\lambda) = \varphi_i(\lambda)\psi_i(\lambda)$, 其中

$$\varphi_i(\lambda) = \mu_{i,0} + \mu_{i,1}(\lambda - \lambda_i) + \frac{\mu_{i,2}}{2!}(\lambda - \lambda_i)^2 + \dots + \frac{\mu_{i,m_i-1}}{(m_i-1)!}(\lambda - \lambda_i)^{m_i-1}$$

$$\psi_i(\lambda) = \prod_{\substack{j=1 \\ j \neq i}}^k (\lambda - \lambda_j)^{m_j}$$

则 $p_i(\lambda)$ 是次数小于 m 的多项式, 并且对任意的 $\mu_{i,0}, \mu_{i,1}, \mu_{i,m_i-1}, p_i(\lambda)$ 满足

$$p_i(\lambda_j) = p'_i(\lambda_j) = \dots = p_i^{(m_j-1)}(\lambda_j) = 0, j \neq i \quad (7.1.21)$$

而

$$\begin{aligned} p_i^{(l)}(\lambda_i) &= (\varphi_i(\lambda)\psi_i(\lambda))^{(l)}|_{\lambda=\lambda_i} = \sum_{j=0}^l C_l^j \varphi_i^{(l-j)}(\lambda_i) \psi_i^{(j)}(\lambda_i) \\ &= \mu_{i,l} \psi_i(\lambda_i) + C_l^1 \mu_{i,l-1} \psi'_i(\lambda_i) + \dots + C_l^l \mu_{i,0} \psi_i^{(l)}(\lambda_i), l = 0, 1, \dots, m_i - 1 \end{aligned}$$

因为 $\psi_i(\lambda_i) \neq 0$, 令

$$\mu_{i,0} = \frac{f_{i,0}}{\psi_i(\lambda_i)}$$

并且对 $l = 1, 2, \dots, m_i - 1$ 可依次令 $\mu_{i,l}$ 如下

$$\mu_{i,l} = \frac{1}{\psi_i(\lambda_i)} (f_{i,l} - C_l^1 \mu_{i,l-1} \psi'_i(\lambda_i) - \dots - C_l^l \mu_{i,0} \psi_i^{(l)}(\lambda_i))$$

则多项式 $p_i(\lambda_i)$ 满足

$$p_i^{(l)}(\lambda_i) = f_{i,l}, l = 0, 1, \dots, m_i - 1 \quad (7.1.22)$$

令

$$p(\lambda) = p_1(\lambda) + \dots + p_k(\lambda)$$

则 $p(\lambda)$ 是次数小于 m 的多项式, 并且由 (7.1.21) 和 (7.1.22) 知 $p(\lambda)$ 满足条件 (7.1.20). \square

通常把满足条件 (7.1.20) 的多项式 $p(\lambda)$ 称为 **Hermite 插值多项式**.

定理 7.1.5 设矩阵 $A \in C^{n \times n}$ 是一个块对角矩阵 $A = \text{diag}(A_1, \dots, A_s)$. 如果函数 $f(z)$ 在 A 的谱 $\lambda(A)$ 上有定义, 则

$$f(A) = \text{diag}(f(A_1), \dots, f(A_s)) \quad (7.1.23)$$

证明 显然, 对任意多项式 $q(\lambda)$ 有

$$q(A) = \text{diag}(q(A_1), \dots, q(A_s))$$

如果 $p(\lambda)$ 是函数 $f(z)$ 在 A 的谱 $\lambda(A)$ 上的 Hermite 插值多项式, 则

$$f(A) - p(A) = \text{diag}(p(A_1), \dots, p(A_s))$$

因为 $A_i (1 \leq i \leq s)$ 的谱是 A 的谱的子集, 并且 $p(\lambda)$ 与 $f(z)$ 在 A 的谱 $\lambda(A)$ 上具有相同的值, 则它们在 $A_i (i = 1, 2, \dots, s)$ 的谱上也有相同的值, 从而 $f(A_i) = p(A_i) (i = 1, 2, \dots, s)$. 因此 $f(A) = \text{diag}(p(A_1), \dots, p(A_s)) = \text{diag}(f(A_1), \dots, f(A_s))$. \square

定理 7.1.6 设 $A, B \in C^{n \times n}$, 如果存在可逆矩阵 P 使得 $B = PAP^{-1}$, 并且函数 $f(z)$ 在 A 的谱 $\lambda(A)$ 上有定义, 则

$$f(B) = Pf(A)P^{-1} \quad (7.1.24)$$

证明 因为 A 和 B 相似, 则由定理 3.6.3 知 A 和 B 有相同的最小多项式. 因此, 如果 $p(\lambda)$ 是函数 $f(z)$ 在 A 的谱 $\lambda(A)$ 上的 Hermite 插值多项式, 则它也是 $f(z)$ 在 B 的谱上的 Hermite 插值多项式. 于是 $f(A) = p(A)$, $f(B) = p(B)$, $p(B) = Pp(A)P^{-1}$, 从而 $f(B) = Pp(A)P^{-1} = Pf(A)P^{-1}$. \square

定理 7.1.7 设矩阵 $A \in C^{m \times n}$ 的 Jordan 标准形为 (7.1.13), 若函数 $f(z)$ 在 A 的谱 $\lambda(A)$ 上有定义, 则

$$f(A) = Pf(J)P^{-1} = P \text{diag}(f(J_1), f(J_2), \dots, f(J_s))P^{-1} \quad (7.1.25)$$

其中

$$f(\mathbf{J}_i) = \begin{bmatrix} f(\lambda_i) & \frac{1}{1!}f'(\lambda_i) & \cdots & \frac{1}{(n_i-1)!}f^{(n_i-1)}(\lambda_i) \\ & f(\lambda_i) & \ddots & \vdots \\ & & \ddots & \frac{1}{1!}f'(\lambda_i) \\ & & & f(\lambda_i) \end{bmatrix} \quad (7.1.26)$$

并且(7.1.25)给出的矩阵函数 $f(\mathbf{A})$ 与 \mathbf{A} 的 Jordan 标准形 \mathbf{J} 中 Jordan 块的排列次序以及变换矩阵 \mathbf{P} 的选取均无关.

证明 设 $p(\lambda)$ 是函数 $f(z)$ 在 \mathbf{A} 的谱 $\lambda(\mathbf{A})$ 上的 Hermite 插值多项式, 则由定义 7.1.2 和(7.1.14)有

$$f(\mathbf{A}) = p(\mathbf{A}) = \mathbf{P} \operatorname{diag}(p(\mathbf{J}_1), \cdots, p(\mathbf{J}_s)) \mathbf{P}^{-1}$$

因为 $p(\lambda)$ 和 $f(z)$ 在 \mathbf{A} 的谱 $\lambda(\mathbf{A})$ 上具有相同的值, 则由(7.1.15)和(7.1.26)得

$$p(\mathbf{J}_i) = f(\mathbf{J}_i), i = 1, 2, \cdots, s$$

由定理 7.1.5 和定理 7.1.6, 有

$$f(\mathbf{A}) = p(\mathbf{A}) = \mathbf{P} \operatorname{diag}(f(\mathbf{J}_1), \cdots, f(\mathbf{J}_s)) \mathbf{P}^{-1} = \mathbf{P} f(\mathbf{J}) \mathbf{P}^{-1}$$

设矩阵 \mathbf{A} 的另一个 Jordan 标准形为 $\tilde{\mathbf{J}}$, 即存在可逆矩阵 $\tilde{\mathbf{P}}$ 使得

$$\tilde{\mathbf{P}}^{-1} \mathbf{A} \tilde{\mathbf{P}} = \tilde{\mathbf{J}} = \operatorname{diag}(\mathbf{J}_{p_1}, \cdots, \mathbf{J}_{p_s})$$

其中 p_1, \cdots, p_s 是 $1, \cdots, s$ 的一个排列, 则

$$\begin{aligned} f(\mathbf{A}) &= \mathbf{P} \operatorname{diag}(f(\mathbf{J}_1), \cdots, f(\mathbf{J}_s)) \mathbf{P}^{-1} = \mathbf{P} \operatorname{diag}(p(\mathbf{J}_1), \cdots, p(\mathbf{J}_s)) \mathbf{P}^{-1} \\ &= p(\mathbf{A}) = \tilde{\mathbf{P}} \operatorname{diag}(p(\mathbf{J}_{p_1}), \cdots, p(\mathbf{J}_{p_s})) \tilde{\mathbf{P}}^{-1} = \tilde{\mathbf{P}} \operatorname{diag}(f(\mathbf{J}_{p_1}), \cdots, f(\mathbf{J}_{p_s})) \tilde{\mathbf{P}}^{-1} \end{aligned}$$

因此, 矩阵函数 $f(\mathbf{A})$ 与 \mathbf{A} 的 Jordan 标准形 \mathbf{J} 中 Jordan 块的排列次序以及变换矩阵 \mathbf{P} 的选取均无关. \square

由定理 7.1.7 即得如下定理.

定理 7.1.8 设矩阵 $\mathbf{A} \in \mathbf{C}^{n \times n}$ 的特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$, 函数 $f(z)$ 在 \mathbf{A} 的谱 $\lambda(\mathbf{A})$ 上有定义, 则 $f(\mathbf{A})$ 的特征值为 $f(\lambda_1), f(\lambda_2), \cdots, f(\lambda_n)$.

§ 7.2 矩阵值函数

带有实变量 x 并取值在 $\mathbf{R}^{m \times n}$ 内的映射 $\mathbf{A}(x)$ 称为矩阵值函数. 与常数矩

阵相比,矩阵值函数有其特殊性质.本节介绍矩阵值函数的一些基本概念及其分析运算.

7.2.1 矩阵值函数

定义 7.2.1 设 $a_{ij}(x) (i=1,2,\cdots,m; j=1,2,\cdots,n)$ 都是定义在区间 (a,b) 上的实函数,则 $m \times n$ 矩阵

$$A(x) = \begin{pmatrix} a_{11}(x) & a_{12}(x) & \cdots & a_{1n}(x) \\ a_{21}(x) & a_{22}(x) & \cdots & a_{2n}(x) \\ \vdots & \vdots & & \vdots \\ a_{m1}(x) & a_{m2}(x) & \cdots & a_{mn}(x) \end{pmatrix}_{m \times n}$$

称为定义在区间 (a,b) 上的矩阵值函数.

特别地,当 $n=1$ 时,我们得到向量值函数,通常用 $\alpha(x)$ 等形式表示.

因为函数可以作加法、减法和乘法等运算,而矩阵的加法、减法、乘法和数量乘法的定义仅用到其元素的加法、减法、乘法,所以我们同样可定义矩阵值函数的加法、减法、乘法、数量乘法和转置运算,并且矩阵值函数的这些运算与常数矩阵的相应运算具有相同的运算规律.

因为矩阵行列式的定义仅涉及其元素的加法与乘法,所以同样可以定义一个 n 阶矩阵值函数的行列式、子式和代数余子式.

定义 7.2.2 区间 (a,b) 上 $m \times n$ 矩阵值函数 $A(x)$ 不恒等于零的子式的最高阶数称为 $A(x)$ 的秩,记为 $\text{rank}(A(x))$.特别地,如果 $A(x)$ 是区间 (a,b) 上 n 阶矩阵值函数,并且 $\text{rank}(A(x))=n$,则称 $A(x)$ 为满秩的.

定义 7.2.3 设 $A(x)=(a_{ij}(x))$ 是区间 (a,b) 上 n 阶矩阵值函数,如果存在 n 阶矩阵值函数 $B(x)=(b_{ij}(x))$,使得对于任何 $x \in (a,b)$ 都有

$$A(x)B(x) = B(x)A(x) = I$$

则称 $A(x)$ 在 (a,b) 上可逆,并称 $B(x)$ 为 $A(x)$ 的逆矩阵,记为 $A^{-1}(x)$.

类似于定理 3.2.1 容易证明如下结论.

定理 7.2.1 n 阶矩阵值函数 $A(x)$ 在 (a,b) 上可逆的充分必要条件是 $|A(x)|$ 在 (a,b) 上处处不为零,并且

$$A^{-1}(x) = \frac{1}{|A(x)|} \text{adj}(A(x))$$

其中

$$\operatorname{adj}(\mathbf{A}(x)) = \begin{bmatrix} A_{11}(x) & A_{21}(x) & \cdots & A_{n1}(x) \\ A_{12}(x) & A_{22}(x) & \cdots & A_{n2}(x) \\ \vdots & \vdots & & \vdots \\ A_{1n}(x) & A_{2n}(x) & \cdots & A_{nn}(x) \end{bmatrix}$$

是 $\mathbf{A}(x)$ 的伴随矩阵值函数, $A_{ij}(x)$ 是 $\mathbf{A}(x)$ 中元素 $a_{ij}(x)$ 的代数余子式.

必须指出, 如果 n 阶矩阵值函数 $\mathbf{A}(x)$ 在 (a, b) 上是可逆的, 则行列式 $|\mathbf{A}(x)|$ 在 (a, b) 上处处不为零, 因而 $\mathbf{A}(x)$ 是满秩的. 反之, 一个 n 阶满秩矩阵值函数未必是可逆的. 这是因为如果 $\mathbf{A}(x)$ 满秩, 则 $\mathbf{A}(x)$ 的行列式不恒等于零, 但不排除 $|\mathbf{A}(x)|$ 在 (a, b) 上有零点. 例如

$$\mathbf{A}(x) = \begin{bmatrix} 0 & -1 \\ x & x^2 \end{bmatrix}$$

则 $|\mathbf{A}(x)| = x$. 于是在任何区间 (a, b) 上 $\mathbf{A}(x)$ 的秩为 2, 即 $\mathbf{A}(x)$ 是满秩的. 但是 $\mathbf{A}(x)$ 在 (a, b) 上是否可逆依赖于 a, b 的取值. 当区间 (a, b) 包含原点时, $|\mathbf{A}(x)|$ 在 (a, b) 上有零点, $\mathbf{A}(x)$ 不可逆. 只有当 (a, b) 不包含原点时, $\mathbf{A}(x)$ 才是可逆的.

7.2.2 矩阵值函数的分析运算

定义 7.2.4 设 $\mathbf{A}(x) = (a_{ij}(x))$ 是区间 (a, b) 上的 $m \times n$ 矩阵值函数. 如果 $\mathbf{A}(x)$ 的所有元素 $a_{ij}(x)$ 在 $x = x_0 \in (a, b)$ 处有极限, 即

$$\lim_{x \rightarrow x_0} a_{ij}(x) = a_{ij}, i = 1, 2, \cdots, m; j = 1, 2, \cdots, n$$

其中 a_{ij} 为固定常数, 则称 $\mathbf{A}(x)$ 在 $x = x_0$ 处有极限, 且记为

$$\lim_{x \rightarrow x_0} \mathbf{A}(x) = \mathbf{A}$$

其中 $\mathbf{A} = (a_{ij}) \in \mathbf{R}^{m \times n}$; 如果 $\mathbf{A}(x)$ 的所有元素 $a_{ij}(x)$ 在 $x = x_0$ 处连续, 即

$$\lim_{x \rightarrow x_0} a_{ij}(x) = a_{ij}(x_0), i = 1, 2, \cdots, m; j = 1, 2, \cdots, n$$

则称 $\mathbf{A}(x)$ 在 $x = x_0$ 处连续, 且记为 $\lim_{x \rightarrow x_0} \mathbf{A}(x) = \mathbf{A}(x_0)$.

容易验证函数极限的运算规则对矩阵值函数也适用. 例如, 如果 $\mathbf{A}(x)$, $\mathbf{B}(x)$ 是区间 (a, b) 上的 $m \times n$ 矩阵值函数, $x_0 \in (a, b)$, 并且 $\lim_{x \rightarrow x_0} \mathbf{A}(x) = \mathbf{A}$, $\lim_{x \rightarrow x_0} \mathbf{B}(x) = \mathbf{B}$, 则

$$(1) \lim_{x \rightarrow x_0} (\mathbf{A}(x) \pm \mathbf{B}(x)) = \mathbf{A} \pm \mathbf{B};$$

$$(2) \lim_{x \rightarrow x_0} (k\mathbf{A}(x)) = k\mathbf{A}.$$

如果 $\mathbf{A}(x)$ 是区间 (a, b) 上的 $m \times n$ 矩阵值函数, $\mathbf{B}(x)$ 是区间 (a, b) 上的 $n \times q$ 矩阵值函数, $x_0 \in (a, b)$, 并且 $\lim_{x \rightarrow x_0} \mathbf{A}(x) = \mathbf{A}$, $\lim_{x \rightarrow x_0} \mathbf{B}(x) = \mathbf{B}$, 则

$$\lim_{x \rightarrow x_0} (\mathbf{A}(x)\mathbf{B}(x)) = \mathbf{AB}$$

定义 7.2.5 设 $\mathbf{A}(x) = (a_{ij}(x))$ 是区间 (a, b) 上的 $m \times n$ 矩阵值函数. 如果 $\mathbf{A}(x)$ 的所有元素 $a_{ij}(x) (i=1, 2, \dots, m; j=1, 2, \dots, n)$ 在点 $x = x_0 \in (a, b)$ 处 (或在区间 (a, b) 内) 可导, 则称矩阵值函数 $\mathbf{A}(x)$ 在 $x = x_0$ 处 (或在区间 (a, b) 内) 可导, 并且

$$\mathbf{A}'(x_0) = \left. \frac{d\mathbf{A}(x)}{dx} \right|_{x=x_0} = (a'_{ij}(x_0)) \quad (7.2.1)$$

称为 $\mathbf{A}(x)$ 在 $x = x_0$ 处的导数. 特别地, 若 $\mathbf{A}(x)$ 的所有元素 $a_{ij}(x) (i=1, 2, \dots, m, j=1, 2, \dots, n)$ 都是 $x = x_0$ 处的解析函数, 则称矩阵值函数 $\mathbf{A}(x)$ 在 $x = x_0$ 处解析. 如果 $\mathbf{A}(x)$ 在区间 (a, b) 内任一点都解析, 则称 $\mathbf{A}(x)$ 为区间 (a, b) 上的解析矩阵值函数.

矩阵值函数的导数运算具有下列性质:

设函数 $k(x)$, $m \times n$ 矩阵值函数 $\mathbf{A}(x)$, $\mathbf{B}(x)$ 和 $n \times q$ 矩阵值函数 $\mathbf{C}(x)$ 在区间 (a, b) 上均可导, 则

$$(1) \mathbf{A}(x) \text{ 是常数矩阵的充分必要条件是 } \frac{d\mathbf{A}(x)}{dx} = 0;$$

$$(2) \frac{d}{dx}(\mathbf{A}(x) + \mathbf{B}(x)) = \frac{d\mathbf{A}(x)}{dx} + \frac{d\mathbf{B}(x)}{dx};$$

$$(3) \frac{d}{dx}(k(x)\mathbf{A}(x)) = \frac{dk(x)}{dx}\mathbf{A}(x) + k(x)\frac{d\mathbf{A}(x)}{dx};$$

$$(4) \frac{d}{dx}(\mathbf{A}(x)\mathbf{C}(x)) = \frac{d\mathbf{A}(x)}{dx}\mathbf{C}(x) + \mathbf{A}(x)\frac{d\mathbf{C}(x)}{dx};$$

$$(5) \text{ 如果 } x = f(t) \text{ 是 } t \text{ 的可微函数, 则}$$

$$\frac{d}{dt}\mathbf{A}(x) = \frac{d\mathbf{A}(x)}{dx}f'(t) = f'(t)\frac{d\mathbf{A}(x)}{dx}$$

因为矩阵乘法没有交换律, 一般地, 对正整数 $m > 1$ 和可导的 n 阶矩阵值函数 $\mathbf{A}(x)$

$$\frac{d}{dx}[\mathbf{A}(x)]^m \neq m[\mathbf{A}(x)]^{m-1} \frac{d\mathbf{A}(x)}{dx}$$

定理 7.2.2 如果 n 阶矩阵值函数 $\mathbf{A}(x)$ 在 (a, b) 上可逆且可导, 则

$$\frac{d\mathbf{A}^{-1}(x)}{dx} = -\mathbf{A}^{-1}(x) \frac{d\mathbf{A}(x)}{dx} \mathbf{A}^{-1}(x) \quad (7.2.2)$$

证明 因为 $\mathbf{A}^{-1}(x)\mathbf{A}(x) = \mathbf{I}$, 所以

$$\frac{d}{dx}[\mathbf{A}^{-1}(x)\mathbf{A}(x)] = \frac{d\mathbf{A}^{-1}(x)}{dx}\mathbf{A}(x) + \mathbf{A}^{-1}(x) \frac{d\mathbf{A}(x)}{dx} = 0$$

于是

$$\frac{d\mathbf{A}^{-1}(x)}{dx} = -\mathbf{A}^{-1}(x) \frac{d\mathbf{A}(x)}{dx} \mathbf{A}^{-1}(x) \quad \square$$

矩阵值函数的导数还是一个矩阵值函数, 它可以再进行求导运算. 因此, 可以定义矩阵值函数的高阶导数

$$\frac{d^k \mathbf{A}(x)}{dx^k} = \frac{d}{dx} \left(\frac{d^{k-1} \mathbf{A}(x)}{dx^{k-1}} \right) \quad k > 1$$

定义 7.2.6 设 $\mathbf{A}(x) = (a_{ij}(x))$ 是区间 $[a, b]$ 上的 $m \times n$ 矩阵值函数. 如果 $\mathbf{A}(x)$ 的所有元素 $a_{ij}(x) (i=1, 2, \dots, m; j=1, 2, \dots, n)$ 在区间 $[a, b]$ 上可积, 则称矩阵值函数 $\mathbf{A}(x)$ 在 $[a, b]$ 上可积, 并称

$$\int_a^b \mathbf{A}(x) dx = \left(\int_a^b a_{ij}(x) dx \right) \quad (7.2.3)$$

为 $\mathbf{A}(x)$ 在 $[a, b]$ 上的积分.

容易证明矩阵值函数的积分具有如下性质:

$$(1) \int_a^b [\mathbf{A}(x) + \mathbf{B}(x)] dx = \int_a^b \mathbf{A}(x) dx + \int_a^b \mathbf{B}(x) dx;$$

$$(2) \text{ 对常数 } k \in \mathbf{R}, \text{ 有 } \int_a^b k\mathbf{A}(x) dx = k \int_a^b \mathbf{A}(x) dx;$$

(3) 对常数矩阵 \mathbf{A} 和 \mathbf{C} , 有

$$\int_a^b [\mathbf{A}\mathbf{B}(x)\mathbf{C}] dx = \mathbf{A} \left[\int_a^b \mathbf{B}(x) dx \right] \mathbf{C}$$

(4) 如果矩阵值函数 $\mathbf{A}(x)$ 在区间 $[a, b]$ 上连续, 则

$$\frac{d}{dx} \int_a^x \mathbf{A}(t) dt = \mathbf{A}(x)$$

(5) 如果矩阵值函数 $\mathbf{A}'(x)$ 在区间 $[a, b]$ 上连续, 则

$$\int_a^b \mathbf{A}'(x) dx = \mathbf{A}(b) - \mathbf{A}(a)$$

类似地, 可以考虑矩阵值函数的不定积分.

一般地,我们把带有多变量(矩阵变量) $\mathbf{X} \in \mathbf{R}^{p \times q}$ ($\mathbf{C}^{p \times q}$) 并取值在 $\mathbf{R}^{m \times n}$ (或 $\mathbf{C}^{m \times n}$) 内的映射 $F(\mathbf{X})$ 称为矩阵值函数.

同样地,可以给出矩阵值函数 $F(\mathbf{X})$ 在 $\mathbf{X} = \mathbf{X}_0$ 处的极限和连续性等概念.

定义 7.2.7 设 $F(\mathbf{X}) = (f_{ij}(\mathbf{X}))$ 是 $\mathbf{X} \in \mathbf{R}^{p \times q}$ 的 $m \times n$ 矩阵值函数, $F(\mathbf{X})$ 的所有元素 $f_{ij}(\mathbf{X})$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$) 作为 $\mathbf{X} \in \mathbf{R}^{p \times q}$ 的多元函数是可微的. 令

$$\frac{dF(\mathbf{X})}{d\mathbf{X}} \equiv \left(\frac{\partial F}{\partial x_{ij}} \right)_{m \times p \times n \times q} = \begin{pmatrix} \frac{\partial F}{\partial x_{11}} & \frac{\partial F}{\partial x_{12}} & \cdots & \frac{\partial F}{\partial x_{1q}} \\ \frac{\partial F}{\partial x_{21}} & \frac{\partial F}{\partial x_{22}} & \cdots & \frac{\partial F}{\partial x_{2q}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial F}{\partial x_{p1}} & \frac{\partial F}{\partial x_{p2}} & \cdots & \frac{\partial F}{\partial x_{pq}} \end{pmatrix} \quad (7.2.4)$$

其中 $\frac{\partial F}{\partial x_{kl}} = \left(\frac{\partial f_{ij}(\mathbf{X})}{\partial x_{kl}} \right)$ ($k = 1, 2, \dots, p; l = 1, 2, \dots, q$), 则称 $\frac{dF(\mathbf{X})}{d\mathbf{X}}$ 为 $F(\mathbf{X})$ 对 \mathbf{X} 的导数.

矩阵值函数的导数具有如下运算规则:

设 $F(\mathbf{X}), G(\mathbf{X})$ 都是 $\mathbf{X} \in \mathbf{R}^{p \times q}$ 的 $m \times n$ 矩阵值函数, 则

$$(1) \frac{d}{d\mathbf{X}} [F(\mathbf{X}) + G(\mathbf{X})] = \frac{dF(\mathbf{X})}{d\mathbf{X}} + \frac{dG(\mathbf{X})}{d\mathbf{X}};$$

$$(2) \text{ 对常数 } k, \text{ 有 } \frac{d}{d\mathbf{X}} [kF(\mathbf{X})] = k \frac{dF(\mathbf{X})}{d\mathbf{X}}.$$

特别地, 如果 $f = f(\mathbf{X})$ 作为 $\mathbf{X} \in \mathbf{R}^{p \times q}$ 的多元函数是可微的, 则 f 对矩阵 \mathbf{X} 的导数为

$$\frac{df}{d\mathbf{X}} = \left(\frac{\partial f}{\partial x_{ij}} \right)_{p \times q} \quad (7.2.5)$$

因为向量可看作矩阵的特殊情形, 如果 $f = f(x)$ 是向量 $x \in \mathbf{R}^n$ 的可微函数, 则 f 对向量 x 的导数为

$$\frac{df}{dx} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \cdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} \quad (7.2.6)$$

例 7.2.1 设 $X = (x_{ij}) \in \mathbf{R}^{n \times n}$, 求

(1) $\frac{d}{dX} \text{tr}(X)$;

(2) $\frac{d}{dX} \det(X)$.

解 (1) 因为 $\text{tr}(X) = x_{11} + x_{22} + \cdots + x_{nn}$, 则

$$\frac{\partial}{\partial x_{ij}} \text{tr}(X) = \delta_{ij}, i, j = 1, 2, \cdots, n$$

于是

$$\frac{d}{dX} \text{tr}(X) = I$$

(2) 因为 $\det(X) = \sum_{k=1}^n x_{ik} X_{ik}$, 其中 X_{ik} 是 x_{ik} 的代数余子式, 则

$$\frac{\partial}{\partial x_{ij}} \det(X) = \frac{\partial}{\partial x_{ij}} \left(\sum_{k=1}^n x_{ik} X_{ik} \right) = X_{ij}$$

从而

$$\frac{d}{dX} \det(X) = (X_{ik}) = [\text{adj}(X)]^T$$

例 7.2.2 设 A 是 n 阶实对称矩阵, $x, b \in \mathbf{R}^n$, $f(x) = \frac{1}{2} x^T A x - b^T x$,

试求 $\frac{df}{dx}$.

解 因为

$$\begin{aligned} \frac{\partial f}{\partial x_i} &= \frac{1}{2} \left[\frac{\partial x^T}{\partial x_i} A x + x^T A \frac{\partial x}{\partial x_i} \right] - b^T \frac{\partial x}{\partial x_i} \\ &= \frac{1}{2} [e_i^T A x + x^T A e_i] - b^T e_i = e_i^T (A x - b) \end{aligned}$$

所以 $\frac{df}{dx} = A x - b$.

设

$$\begin{cases} y_1 = f_1(x_1, x_2, \cdots, x_n) \\ y_2 = f_2(x_1, x_2, \cdots, x_n) \\ \vdots \\ y_m = f_m(x_1, x_2, \cdots, x_n) \end{cases} \quad (7.2.7)$$

$x_i(t) (i=1, 2, \dots, n)$ 是 t 的未知函数. 引进矩阵值函数与向量值函数, (7.3.1) 可以表示成

$$\frac{dx(t)}{dt} = A(t)x(t) + f(t) \quad (7.3.2)$$

其中

$$A(t) = \begin{bmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t) \\ a_{21}(t) & a_{22}(t) & \cdots & a_{2n}(t) \\ \vdots & \vdots & & \vdots \\ a_{n1}(t) & a_{n2}(t) & \cdots & a_{nn}(t) \end{bmatrix}, \quad x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \quad f(t) = \begin{bmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_n(t) \end{bmatrix} \quad (7.3.3)$$

方程组(7.3.1)的初始条件

$$x_1(t_0) = x_{10}, x_2(t_0) = x_{20}, \dots, x_n(t_0) = x_{n0} \quad (7.3.4)$$

可以表示成

$$x(t_0) = x_0 = (x_{10}, x_{20}, \dots, x_{n0})^T \quad (7.3.5)$$

定理 7.3.1 设 A 是 n 阶常数矩阵, 则微分方程组

$$\frac{dx}{dt} = Ax(t) \quad (7.3.6)$$

满足初始条件 $x(t_0) = x_0$ 的解为

$$x(t) = e^{A(t-t_0)} x_0 \quad (7.3.7)$$

证明 将 $x_i(t) (i=1, 2, \dots, n)$ 在 $t=t_0$ 处展开成幂级数

$$x_i(t) = x_i(t_0) + x'_i(t_0)(t-t_0) + \frac{1}{2!}x''_i(t_0)(t-t_0)^2 + \dots$$

从而有

$$x(t) = x(t_0) + x'(t_0)(t-t_0) + \frac{1}{2!}x''(t_0)(t-t_0)^2 + \dots$$

因为

$$x'(t_0) = \left. \frac{dx}{dt} \right|_{t=t_0} = Ax(t_0), \quad x''(t_0) = \left. \frac{d^2x}{dt^2} \right|_{t=t_0}$$

$$= \frac{d}{dt}(Ax) \Big|_{t=t_0} = A^2 x(t_0), \dots$$

于是

$$x(t) = \left\{ I + A(t-t_0) + \frac{1}{2!}[A(t-t_0)]^2 + \dots \right\} x(t_0) = e^{A(t-t_0)} x(t_0)$$

这说明一阶线性常系数齐次微分方程组 $\frac{dx}{dt} = Ax(t)$ 满足初始条件 $x(t_0) = x_0$ 的解必有

$$x(t) = e^{A(t-t_0)} x_0$$

的形式. 另一方面, 由于

$$\frac{dx}{dt} = \frac{d}{dt}[e^{A(t-t_0)} x_0] = \left(\frac{d}{dt} e^{A(t-t_0)} \right) x_0 = A e^{A(t-t_0)} x_0 = Ax(t)$$

因此, 微分方程组(7.3.6)满足初始条件 $x(t_0) = x_0$ 的惟一解为(7.3.7). \square

定义 7.3.1 设 A 是 n 阶常数矩阵, 如果对任意的 t_0 和 x_0 , 初值问题

$$\begin{cases} \frac{dx}{dt} = Ax(t) \\ x(t_0) = x_0 \end{cases} \quad (7.3.8)$$

的解 $x(t)$ 满足 $\lim_{t \rightarrow +\infty} x(t) = 0$, 则称微分方程组 $\frac{dx}{dt} = Ax(t)$ 的解是渐近稳定的.

微分方程组 $\frac{dx}{dt} = Ax(t)$ 解的渐近稳定性是系统与控制理论的基本问题, 对此有如下结果.

定理 7.3.2 对任意的 t_0 和 x_0 , 初值问题(7.3.8)的解 $x(t)$ 渐近稳定的充分必要条件是矩阵 A 的特征值都有负实部.

证明 必要性. 采用反证法. 假若矩阵 A 有一个特征值 $\lambda_1 = \alpha_1 + i\beta_1$ 满足 $\alpha_1 \geq 0$, 设 x_1 是对应于特征值 λ_1 的特征向量, 则

$$Ax_1 = \lambda_1 x_1$$

由定理 7.3.1 知, 初值问题

$$\begin{cases} \frac{dx}{dt} = Ax(t) \\ x(0) = x_1 \end{cases}$$

的解为

$$x(t) = e^{At}x_1 = e^{\lambda_1 t}x_1 = e^{\beta_1 t}(\cos\beta_1 t + i\sin\beta_1 t)x_1$$

因为 $\alpha_1 \geq 0$, 则 $\lim_{t \rightarrow +\infty} x(t) \neq 0$, 这与必要性的假设矛盾. 因此 A 的特征值都有负实部.

充分性. 对任意的 t_0 和 x_0 , 初值问题(7.3.8)的解为

$$x(t) = e^{A(t-t_0)}x_0$$

如果矩阵 A 的特征值都有负实部, 则

$$\lim_{t \rightarrow +\infty} e^{A(t-t_0)} = 0$$

故 $\lim_{t \rightarrow +\infty} x(t) = 0$, 即初值问题(7.3.8)的解 $x(t)$ 渐近稳定. \square

定义 7.3.2 设 A 是 n 阶矩阵, 如果 A 的特征值都有负实部, 则称 A 为稳定矩阵.

由定理 7.3.2 和定义 7.3.2 知, 初值问题(7.3.8)的解 $x(t)$ 渐近稳定的充分必要条件是矩阵 A 为稳定矩阵.

定理 7.3.3 设 A 是 n 阶常数矩阵, 则微分方程组

$$\frac{dx}{dt} = Ax(t) + f(t) \quad (7.3.9)$$

满足初始条件 $x(t_0) = x_0$ 的解为

$$x(t) = e^{A(t-t_0)}x_0 + \int_{t_0}^t e^{A(t-\tau)}f(\tau)d\tau \quad (7.3.10)$$

证明 将一阶线性常系数非齐次微分方程组 $\frac{dx}{dt} = Ax(t) + f(t)$ 改写为

$$\frac{dx}{dt} - Ax(t) = f(t)$$

方程组两边同乘 e^{-At} , 得

$$e^{-At}\left(\frac{dx}{dt} - Ax(t)\right) = e^{-At}f(t)$$

即

$$\frac{d}{dt}(e^{-At}x(t)) = e^{-At}f(t)$$

在 $[t_0, t]$ 上对上式积分, 得

$$e^{-At}x(t) - e^{-At_0}x(t_0) = \int_{t_0}^t e^{-A\tau}f(\tau)d\tau$$

因此 $x(t) = e^{A(t-t_0)}x_0 + \int_{t_0}^t e^{A(t-\tau)}f(\tau)d\tau$ 是微分方程组(7.3.9)满足条件 $x(t_0) = x_0$ 的解. \square

例 7.3.1 设 $A = \begin{bmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{bmatrix}$, 求微分方程组 $\frac{dx}{dt} = Ax(t)$ 满足初始

条件 $x(0) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ 的解.

解 由例 7.1.2 知

$$e^{-At} = \begin{bmatrix} (1-2t)e^t & -2te^t & 6te^t \\ -te^t & (1-t)e^t & 3te^t \\ -te^t & -te^t & (1+3t)e^t \end{bmatrix}$$

由定理 7.3.1 得微分方程组 $\frac{dx}{dt} = Ax(t)$ 满足初始条件 $x(0) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ 的解为

$$x(t) = e^{-At}x(0) = \begin{bmatrix} (1+2t)e^t \\ (1+t)e^t \\ (1+t)e^t \end{bmatrix}$$

例 7.3.2 设 $A = \begin{bmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{bmatrix}$, $f(t) = \begin{bmatrix} -e^t \\ 0 \\ e^t \end{bmatrix}$, 求微分方程组 $\frac{dx}{dt} =$

$Ax(t) + f(t)$ 满足初始条件 $x(0) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ 的解.

解 由定理 7.3.3 知微分方程组 $\frac{dx}{dt} = Ax(t) + f(t)$ 满足初始条件 $x(0) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ 的解为

$$x(t) = e^{-At}x(0) + \int_0^t e^{A(t-\tau)}f(\tau)d\tau$$

上式中 $e^{-At}x(0)$ 已由例 7.3.1 求出, 而

$$e^{A(t-\tau)} = \begin{bmatrix} (1-2t+2\tau)e^{t-\tau} & -2(t-\tau)e^{t-\tau} & 6(t-\tau)e^{t-\tau} \\ -(t-\tau)e^{t-\tau} & (1-t+\tau)e^{t-\tau} & 3(t-\tau)e^{t-\tau} \\ -(t-\tau)e^{t-\tau} & -(t-\tau)e^{t-\tau} & (1+3t-3\tau)e^{t-\tau} \end{bmatrix}$$

则

$$e^{A(t-\tau)}f(\tau) = e^t \begin{bmatrix} -1+8(t-\tau) \\ 4(t-\tau) \\ 4(t-\tau)+1 \end{bmatrix}$$

$$\int_0^t e^{A(t-\tau)}f(\tau)d\tau = e^t \begin{bmatrix} \int_0^t [-1+8(t-\tau)]d\tau \\ \int_0^t 4(t-\tau)d\tau \\ \int_0^t [4(t-\tau)+1]d\tau \end{bmatrix} = e^t \begin{bmatrix} 4t^2-t \\ 2t^2 \\ 2t^2+t \end{bmatrix}$$

于是

$$x(t) = \begin{bmatrix} (1+t+4t^2)e^t \\ (1+t+2t^2)e^t \\ (1+2t+2t^2)e^t \end{bmatrix}$$

以上仅介绍了矩阵值函数在一阶线性常系数齐次和非齐次微分方程组中的应用. 同样地, 在求解一阶线性变系数齐次和非齐次微分方程组时, 也需要利用矩阵值函数, 但比以上所述复杂多了. 在此不作进一步的讨论.

§ 7.4 特征对的灵敏度分析*

矩阵 A 的特征值 λ 与相应的特征向量 x 称为矩阵 A 的**特征对**, 记为 (λ, x) . 我们知道矩阵的特征值是矩阵元素的连续函数, 并且对应于单特征值的特征向量也是矩阵元素的连续函数. 本节我们进一步研究矩阵值函数特征对

的灵敏度分析. 这些结果在结构设计、结构故障诊断和结构模型修正等许多领域都具有重要的应用.

我们不加证明地介绍两个隐函数定理.

定理 7.4.1 如果复值函数 $f_i(\xi_1, \dots, \xi_m, \eta_1, \dots, \eta_n) (i = 1, 2, \dots, m)$ 在 \mathbb{C}^{m+n} 的原点的某个邻域内是 $m+n$ 个复变量的解析函数, 并且

$$f_i(0, \dots, 0) = 0 (i = 1, 2, \dots, m), \quad \left. \frac{\partial(f_1, \dots, f_m)}{\partial(\xi_1, \dots, \xi_m)} \right|_{\substack{\xi_j = 0 (j=1, 2, \dots, m) \\ \eta_j = 0 (j=1, 2, \dots, n)}} \neq 0$$

则方程组

$$f_i(\xi_1, \dots, \xi_m, \eta_1, \dots, \eta_n) = 0 \quad i = 1, \dots, m$$

在 \mathbb{C}^n 的原点的某个邻域内有惟一的解析解 $\xi_i = g_i(\eta_1, \dots, \eta_n) (i = 1, \dots, m)$, 并且当 $\eta_1 = \dots = \eta_n = 0$ 时, 有 $\xi_1 = \dots = \xi_m = 0$.

定理 7.4.2 如果实值函数 $f_i(\xi_1, \dots, \xi_m, \eta_1, \dots, \eta_n) (i = 1, 2, \dots, m)$ 在 \mathbb{R}^{m+n} 的原点的某个邻域内是 $m+n$ 个实变量的解析函数, 并且

$$f_i(0, \dots, 0) = 0 (i = 1, 2, \dots, m), \quad \left. \frac{\partial(f_1, \dots, f_m)}{\partial(\xi_1, \dots, \xi_m)} \right|_{\substack{\xi_j = 0 (j=1, 2, \dots, m) \\ \eta_j = 0 (j=1, 2, \dots, n)}} \neq 0$$

则方程组

$$f_i(\xi_1, \dots, \xi_m, \eta_1, \dots, \eta_n) = 0, i = 1, \dots, m$$

在 \mathbb{R}^n 的原点的某个邻域内有惟一的解析解 $\xi_i = g_i(\eta_1, \dots, \eta_n) (i = 1, \dots, m)$, 并且当 $\eta_1 = \dots = \eta_n = 0$ 时, 有 $\xi_1 = \dots = \xi_m = 0$.

定理 7.4.1 和定理 7.4.2 是隐函数理论中的基本定理, 其证明可以在多复变函数和多元函数的参考书中找到, 可参见文[14]和[15].

设 $c = (c_1, \dots, c_m)^T \in \mathbb{C}^m$ (或 \mathbb{R}^m), $A(c) = (a_{ij}(c)) \in \mathbb{C}^{n \times n}$ (或 $\mathbb{R}^{n \times n}$) 是 \mathbb{C}^m (或 \mathbb{R}^m) 的原点的某个邻域 $N(0)$ 内的解析(或实解析)矩阵值函数. 设 λ 是 $A(0)$ 的一个特征值, 则存在非零向量 $x, y \in \mathbb{C}^n$ (或 \mathbb{R}^n) 使得

$$A(0)x = \lambda x, \quad y^T A(0) = \lambda y^T \quad (7.4.1)$$

向量 x 和 y 分别称为 $A(0)$ 对应于特征值 λ 的右和左特征向量.

定理 7.4.3 设 $c \in \mathbb{C}^m$, $A(c) \in \mathbb{C}^{n \times n}$ 是 \mathbb{C}^m 的原点的某个邻域 $N(0)$ 内的解析矩阵值函数, λ_1 是 $A(0)$ 的一个单特征值, x_1, y_1 是 $A(0)$ 对应于特征值 λ_1 的右和左特征向量, 并且 $\|x_1\|_2 = 1, y_1^T x_1 = 1$, 则

(1) $A(c)$ 存在一个单特征值 $\lambda_1(c)$, 使得 $\lambda_1(c)$ 是 \mathbb{C}^m 的原点的某个邻域 N_0 内的解析函数, 并且 $\lambda_1(0) = \lambda_1$;

(2) $A(c)$ 对应于单特征值 $\lambda_1(c)$ 的右特征向量 $x_1(c)$ 和左特征向量 $y_1(c)$ 可定义为 N_0 内的解析向量值函数, 且 $x_1(0) = x_1, y_1(0) = y_1$.

证明 因为 λ_1 是 $A(0)$ 的一个单特征值, x_1 是 $A(0)$ 对应于特征值 λ_1 的单位特征向量, 则存在矩阵 $X_2 \in \mathbb{C}^{n \times (n-1)}$ 使得

$$X = [x_1, X_2] \quad (7.4.2)$$

是非奇异矩阵, 并且

$$X^{-1}A(0)X = \begin{bmatrix} \lambda_1 & 0 \\ 0 & A_2 \end{bmatrix}, \lambda_1 \notin \lambda(A_2) \quad (7.4.3)$$

令

$$Y = (X^{-1})^T = [z_1, Y_2] \quad (7.4.4)$$

其中 $z_1 \in \mathbb{C}^n, Y_2 \in \mathbb{C}^{n \times (n-1)}$, 则

$$Y^T X = I \quad (7.4.5)$$

从而 $z_1 = (X^{-1})^T e_1$ 满足 $z_1^T x_1 = 1$ 并且 $z_1^T A(0) = \lambda_1 z_1^T$. 因为对应于单特征值的特征向量除相差一个非零常数外是惟一的, 所以 $z_1 = y_1$. 于是

$$Y = (X^{-1})^T = [y_1, Y_2] \quad (7.4.6)$$

由(7.4.3)和(7.4.6)得

$$Y^T A(0)X = \begin{bmatrix} \lambda_1 & 0 \\ 0 & A_2 \end{bmatrix}, \lambda_1 \notin \lambda(A_2) \quad (7.4.7)$$

令

$$\tilde{A}(c) = Y^T A(c)X = \begin{bmatrix} \tilde{a}_{11}(c) & \tilde{a}_{12}^T(c) \\ \tilde{a}_{21}(c) & \tilde{A}_{22}(c) \end{bmatrix}, \tilde{a}_{11}(c) \in \mathbb{C}, \tilde{a}_{21}(c) \in \mathbb{C}^{n-1} \quad (7.4.8)$$

考虑向量值函数

$$(f_1(z, c), \dots, f_{n-1}(z, c))^T = \tilde{a}_{21}(c) + [\tilde{A}_{22}(c) - \tilde{a}_{11}(c)I]z - z\tilde{a}_{12}^T(c)z \quad (7.4.9)$$

其中 $z = (\xi_1, \dots, \xi_{n-1})^T \in \mathbb{C}^{n-1}$.

对 $z \in \mathbb{C}^{n-1}$ 和 $c \in N(0)$, $(f_1(z, c), \dots, f_{n-1}(z, c))^T$ 是解析的, 并且由

(7.4.7)–(7.4.9)知 $f_i(0,0)=0 (i=1,2,\cdots,n-1)$,

$$\left. \frac{\partial(f_1, \cdots, f_{n-1})}{\partial(\xi_1, \cdots, \xi_{n-1})} \right|_{\substack{z=0 \\ c=0}} = \det(\mathbf{A}_2 - \lambda_1 \mathbf{I}) \neq 0$$

由定理 7.4.1 知, 方程组

$$f_i(z, c) = 0 \quad i = 1, 2, \cdots, n-1 \quad (7.4.10)$$

在 \mathbf{C}^m 的原点的某个邻域 $N_1(0) (\subseteq N(0))$ 内有惟一解析解 $z = z(c)$, 并且 $z(0)=0$.

由(7.4.8)–(7.4.10)可见, $z(c)$ 满足

$$\tilde{\mathbf{A}}(c) \begin{bmatrix} 1 \\ z(c) \end{bmatrix} = (\tilde{a}_{11}(c) + \tilde{a}_{12}^T(c)z(c)) \begin{bmatrix} 1 \\ z(c) \end{bmatrix}, c \in N_1(0) \quad (7.4.11)$$

令

$$\lambda_1(c) = \tilde{a}_{11}(c) + \tilde{a}_{12}^T(c)z(c), x_1(c) = x_1 + \mathbf{X}_2 z(c) \quad (7.4.12)$$

由(7.4.8), (7.4.11)和(7.4.12), 有

$$\mathbf{A}(c)x_1(c) = \lambda_1(c)x_1(c), c \in N_1(0) \quad (7.4.13)$$

由(7.4.12)可见, $\lambda_1(c)$ 和 $x_1(c)$ 在 $N_1(0)$ 内解析, 并且满足

$$\lambda_1(0) = \lambda_1, \quad x_1(0) = x_1 \quad (7.4.14)$$

因为 λ_1 是 $\mathbf{A}(0)$ 的一个单特征值, 而矩阵的特征值是矩阵元素的连续函数, 所以只要 $\|c\|_2$ 充分小, $\lambda_1(c)$ 是 $\mathbf{A}(c)$ 的一个单特征值. 可以假定邻域 $N_1(0)$ 充分小使得对 $c \in N_1(0)$, $\lambda_1(c)$ 是 $\mathbf{A}(c)$ 的一个单特征值.

类似地, 我们引入一个向量值函数

$$(g_1(w, c), \cdots, g_{n-1}(w, c)) = \tilde{a}_{12}^T(c) + w^T[\tilde{\mathbf{A}}_{22}(c) - \tilde{a}_{11}(c)\mathbf{I}] - w^T \tilde{a}_{21}(c)w^T \quad (7.4.15)$$

其中 $w = (w_1, \cdots, w_{n-1})^T \in \mathbf{C}^{n-1}$.

显然, 对 $w \in \mathbf{C}^{n-1}$ 和 $c \in N(0)$, $g_i(w, c) (i=1, 2, \cdots, n-1)$ 是解析的, 并且由(7.4.7), (7.4.8)和(7.4.15)知 $g_i(0,0)=0 (i=1, 2, \cdots, n-1)$,

$$\left. \frac{\partial(g_1, \cdots, g_{n-1})}{\partial(w_1, \cdots, w_{n-1})} \right|_{\substack{w=0 \\ c=0}} = \det(\mathbf{A}_2 - \lambda_1 \mathbf{I}) \neq 0$$

由定理 7.4.1 知, 方程组

$$g_i(w, c) = 0, i = 1, 2, \dots, n-1 \quad (7.4.16)$$

在 \mathbb{C}^m 的原点的某个邻域 $N_2(0) (\subseteq N(0))$ 内有惟一的解析解 $w = w(c)$, 并且 $w(0) = 0$.

由(7.4.8), (7.4.15)和(7.4.16)可见, $w(c)$ 满足

$$\begin{bmatrix} 1 \\ w(c) \end{bmatrix}^T \tilde{A}(c) = [\tilde{a}_{11}(c) + w^T(c)\tilde{a}_{21}(c)] \begin{bmatrix} 1 \\ w(c) \end{bmatrix}^T, c \in N_2(0) \quad (7.4.17)$$

令

$$\lambda(c) = \tilde{a}_{11}(c) + w^T(c)\tilde{a}_{21}(c), \quad y_1(c) = y_1 + Y_2 w(c) \quad (7.4.18)$$

则由(7.4.8), (7.4.17)和(7.4.18), 有

$$y_1^T(c)A(c) = \lambda(c)y_1^T(c), c \in N_2(0)$$

即 $y_1(c)$ 是 $A(c)$ 对应于特征值 $\lambda(c)$ 的左特征向量.

由(7.4.18)知, $\lambda(c)$ 和 $y_1(c)$ 是 $N_2(0)$ 内的解析函数, 并且满足 $\lambda(0) = \lambda_1, y_1(0) = y_1$. 因为矩阵的特征值是矩阵元素的连续函数, 所以只要 $\|c\|_2$ 充分小, $\lambda(c)$ 是 $A(c)$ 的一个单特征值并且 $\lambda(c) = \lambda_1(c)$. 因此存在 \mathbb{C}^m 的原点的一个邻域 $N_0 \subseteq N_1(0) \cap N_2(0)$ 使得

$$\lambda(c) = \lambda_1(c), \quad |w^T(c)z(c)| < 1, \quad c \in N_0$$

于是解析向量值函数 $y_1(c)$ 满足

$$y_1(0) = y_1, y_1^T(c)x_1(c) \neq 0, y_1^T(c)A(c) = \lambda_1(c)y_1^T(c), c \in N_0 \quad (7.4.19) \square$$

定理 7.4.4 假设定理 7.4.3 的条件成立, 对由(7.4.12)和(7.4.18)定义的单特征值 $\lambda_1(c)$ 和相应的右和左特征向量 $x_1(c)$ 和 $y_1(c)$ 有

$$\left(\frac{\partial \lambda_1(c)}{\partial c_i} \right)_{c=0} = y_1^T \left(\frac{\partial A(c)}{\partial c_i} \right)_{c=0} x_1 \quad (7.4.20)$$

$$\left(\frac{\partial x_1(c)}{\partial c_i} \right)_{c=0} = X_2 [\lambda_1 I - Y_2^T A(0) X_2]^{-1} Y_2^T \left(\frac{\partial A(c)}{\partial c_i} \right)_{c=0} x_1 \quad (7.4.21)$$

$$\left(\frac{\partial y_1(c)}{\partial c_i}\right)_{c=0}^T \approx y_1^T \left(\frac{\partial \mathbf{A}(c)}{\partial c_i}\right)_{|c=0} \mathbf{X}_2 [\lambda_1 \mathbf{I} - \mathbf{Y}_2^T \mathbf{A}(0) \mathbf{X}_2]^{-1} \mathbf{Y}_2^T \quad (7.4.22)$$

其中 $i=1, 2, \dots, n$, \mathbf{X}_2 和 \mathbf{Y}_2 由(7.4.2)和(7.4.6)定义.

证明 由定理 7.4.3

$$\lambda_1(c) = \frac{y_1^T(c) \mathbf{A}(c) x_1(c)}{y_1^T(c) x_1(c)}, c \in N_0 \quad (7.4.23)$$

由(7.4.13)和(7.4.19), 有

$$\frac{\partial \lambda_1(c)}{\partial c_i} = -\frac{y_1^T(c) \frac{\partial \mathbf{A}(c)}{\partial c_i} x_1(c)}{y_1^T(c) x_1(c)}, i = 1, 2, \dots, m, c \in N_0 \quad (7.4.24)$$

由(7.4.14)和(7.4.19)即得(7.4.20).

由(7.4.13), 有

$$[\lambda_1(c) \mathbf{I} - \mathbf{A}(c)] \frac{\partial x_1(c)}{\partial c_i} = \left(\frac{\partial \mathbf{A}(c)}{\partial c_i} - \frac{\partial \lambda_1(c)}{\partial c_i} \mathbf{I} \right) x_1(c), c \in N_0 \quad (7.4.25)$$

并且由(7.4.5), (7.4.7)和(7.4.12), 得

$$\begin{aligned} [\lambda_1 \mathbf{I} - \mathbf{A}(0)] \left(\frac{\partial x_1(c)}{\partial c_i} \right)_{|c=0} &= \left[\lambda_1 \mathbf{I} - (\mathbf{Y}^T)^{-1} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \mathbf{A}_2 \end{pmatrix} \mathbf{X}^{-1} \right] \mathbf{X} \begin{bmatrix} 0 \\ \left(\frac{\partial z(c)}{\partial c_i} \right)_{|c=0} \end{bmatrix} \\ &= \mathbf{X} \begin{bmatrix} 0 & 0 \\ 0 & \lambda_1 \mathbf{I} - \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} 0 \\ \left(\frac{\partial z(c)}{\partial c_i} \right)_{|c=0} \end{bmatrix} \end{aligned} \quad (7.4.26)$$

因为 $\lambda_1 \in \lambda(\mathbf{A}_2)$, 则由(7.4.5), (7.4.25)和(7.4.26), 有

$$\begin{bmatrix} 0 \\ \left(\frac{\partial z(c)}{\partial c_i} \right)_{|c=0} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & (\lambda_1 \mathbf{I} - \mathbf{A}_2)^{-1} \end{bmatrix} \mathbf{Y}^T \left(\left(\frac{\partial \mathbf{A}(c)}{\partial c_i} \right)_{c=0} - \left(\frac{\partial \lambda_1(c)}{\partial c_i} \right)_{c=0} \mathbf{I} \right) x_1$$

于是由(7.4.5)和上式得

$$\left(\frac{\partial x_1(c)}{\partial c_i}\right)_{|c=0} = \mathbf{X} \begin{bmatrix} 0 \\ \left(\frac{\partial z(c)}{\partial c_i}\right)_{|c=0} \end{bmatrix} = \mathbf{X}_2(\lambda_1 \mathbf{I} - \mathbf{A}_2)^{-1} \mathbf{Y}_2^T \left(\frac{\partial \mathbf{A}(c)}{\partial c_i}\right)_{|c=0} x_1$$

由(7.4.2), (7.4.6)和(7.4.7), 有 $\mathbf{A}_2 = \mathbf{Y}_2^T \mathbf{A}(0) \mathbf{X}_2$, 从而由上式即得(7.4.21).

由(7.4.19), 有

$$\left(\frac{\partial y_1(c)}{\partial c_i}\right)^T [\lambda_1(c) \mathbf{I} - \mathbf{A}(c)] = y_1^T(c) \left[\frac{\partial \mathbf{A}(c)}{\partial c_i} - \frac{\partial \lambda_1(c)}{\partial c_i} \mathbf{I} \right], c \in N_0 \quad (7.4.27)$$

并且由(7.4.5), (7.4.7)和(7.4.18)得

$$\left(\frac{\partial y_1(c)}{\partial c_i}\right)_{|c=0}^T [\lambda_1 \mathbf{I} - \mathbf{A}(0)] = \left[\left(\frac{\partial w(c)}{\partial c_i}\right)_{|c=0} \right]^T \begin{bmatrix} 0 & 0 \\ 0 & \lambda_1 \mathbf{I} - \mathbf{A}_2 \end{bmatrix} \mathbf{Y}^T \quad (7.4.28)$$

因为 $\lambda_1 \in \lambda(\mathbf{A}_2)$, 则由(7.4.5), (7.4.27)和(7.4.28)有

$$\begin{bmatrix} 0 \\ \left(\frac{\partial w(c)}{\partial c_i}\right)_{|c=0} \end{bmatrix}^T = y_1^T \left[\left(\frac{\partial \mathbf{A}(c)}{\partial c_i}\right)_{|c=0} - \left(\frac{\partial \lambda_1(c)}{\partial c_i}\right)_{|c=0} \mathbf{I} \right] \mathbf{X} \begin{bmatrix} 0 & 0 \\ 0 & (\lambda_1 \mathbf{I} - \mathbf{A}_2)^{-1} \end{bmatrix}$$

于是

$$\left(\frac{\partial y_1(c)}{\partial c_i}\right)_{|c=0}^T = \begin{bmatrix} 0 \\ \left(\frac{\partial w(c)}{\partial c_i}\right)_{|c=0} \end{bmatrix}^T \mathbf{Y}^T = y_1^T \left(\frac{\partial \mathbf{A}(c)}{\partial c_i}\right)_{|c=0} \mathbf{X}_2 (\lambda_1 \mathbf{I} - \mathbf{A}_2)^{-1} \mathbf{Y}_2^T$$

由 $\mathbf{A}_2 = \mathbf{Y}_2^T \mathbf{A}(0) \mathbf{X}_2$ 和上式即得(7.4.22). \square

应用定理 7.4.2, 可以证明如下定理.

定理 7.4.5 设 $c \in \mathbf{R}^m$, $\mathbf{A}(c)$ 是 n 阶实对称矩阵, 并且 $\mathbf{A}(c)$ 是 \mathbf{R}^m 的原点的某个邻域 $N(0)$ 内的实解析矩阵值函数, λ_1 是 $\mathbf{A}(0)$ 的一个单特征值, x_1 是相应的特征向量并且满足

$$\mathbf{A}(0)x_1 = \lambda_1 x_1, \quad \|x_1\|_2 = 1$$

则

(1) $\mathbf{A}(c)$ 存在一个单特征值 $\lambda_1(c)$, 它是 \mathbf{R}^m 的原点的某个邻域 N_0 内的

实解析函数, 并且 $\lambda_1(0) = \lambda_1$;

(2) $\mathbf{A}(c)$ 对应于特征值 $\lambda_1(c)$ 的特征向量 $x_1(c)$ 可定义为 N_0 内的一个实解析向量值函数, 并且 $x_1(0) = x_1$.

证明 因为 λ_1 是 $\mathbf{A}(0)$ 的单特征值, x_1 是相应的单位特征向量, 则存在 $\mathbf{X}_2 \in \mathbf{R}^{n \times (n-1)}$ 使得 $\mathbf{X} = [x_1, \mathbf{X}_2]$ 是一个正交矩阵, 并且

$$\mathbf{X}^T \mathbf{A}(0) \mathbf{X} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \mathbf{A}_2 \end{bmatrix}, \lambda_1 \in \lambda(\mathbf{A}_2) \quad (7.4.29)$$

令

$$\tilde{\mathbf{A}}(c) = \mathbf{X}^T \mathbf{A}(c) \mathbf{X} = \begin{bmatrix} \tilde{a}_{11}(c) & \tilde{a}_{21}^T(c) \\ \tilde{a}_{21}(c) & \tilde{\mathbf{A}}_{22}(c) \end{bmatrix}, \quad \tilde{a}_{11}(c) \in \mathbf{R}, \tilde{a}_{21}(c) \in \mathbf{R}^{n-1}$$

类似于定理 7.4.3 的证明, 可证在 \mathbf{R}^n 的原点的某个邻域 N_0 内存在一个实解析向量值函数 $z(c) \in \mathbf{R}^{n-1}$ 使得 $z(0) = 0$, 并且对 $c \in N_0$ 有 $\|z(c)\|_2 < 1$, 且实解析函数

$$\lambda_1(c) = \tilde{a}_{11}(c) + \tilde{a}_{21}^T(c)z(c), \quad x_1(c) = x_1 + \mathbf{X}_2 z(c) \quad (7.4.30)$$

满足 $\lambda_1(0) = \lambda_1$, $x_1(0) = x_1$ 以及

$$\mathbf{A}(c)x_1(c) = \lambda_1(c)x_1(c), \quad c \in N_0 \quad \square$$

由定理 7.4.5, 用证明定理 7.4.4 同样的方法, 可得如下定理.

定理 7.4.6 假定定理 7.4.5 的条件成立, 则对由 (7.4.30) 定义的单特征值 $\lambda_1(c)$ 和相应的特征向量 $x_1(c)$, 有

$$\left(\frac{\partial \lambda_1(c)}{\partial c_i} \right)_{|c=0} = x_1^T \left(\frac{\partial \mathbf{A}(c)}{\partial c_i} \right)_{|c=0} x_1 \quad (7.4.31)$$

$$\left(\frac{\partial x_1(c)}{\partial c_i} \right)_{|c=0} = \mathbf{X}_2 [\lambda_1 \mathbf{I} - \mathbf{X}_2^T \mathbf{A}(0) \mathbf{X}_2]^{-1} \mathbf{X}_2^T \left(\frac{\partial \mathbf{A}(c)}{\partial c_i} \right)_{|c=0} x_1 \quad (7.4.32)$$

其中 $i = 1, 2, \dots, m$, $\mathbf{X} = [x_1, \mathbf{X}_2] \in \mathbf{R}^{n \times n}$ 是使 (7.4.29) 成立的正交矩阵.

例 7.4.1 设 $\mathbf{A}(c) = \begin{bmatrix} c_1 & c_2 \\ c_2 & -c_1 \end{bmatrix}$, $c = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \in \mathbf{R}^2$. $\mathbf{A}(c)$ 是实对称的解析矩阵值函数, $\mathbf{A}(0)$ 具有 2 重特征值 0. $\mathbf{A}(c)$ 的特征值为

$$\lambda_1(c) = \sqrt{c_1^2 + c_2^2}, \lambda_2(c) = -\sqrt{c_1^2 + c_2^2}$$

容易证明,函数 $\lambda_1(c)$ 和 $\lambda_2(c)$ 在 $c=0$ 处不可微,并且 $A(c)$ 的对应于特征值 $\lambda_1(c)$ 和 $\lambda_2(c)$ 的特征向量在 $c=0$ 处不连续.

例 7.4.1 说明当矩阵值函数 $A(c) \in \mathbb{C}^{n \times n}$ (或 $\mathbb{R}^{n \times n}$) 在 $c=0$ 处有重特征值 λ 时, $A(c)$ 与 λ 对应的特征值 $\lambda(c)$ 在 $c=0$ 处一般不可微,并且 $A(c)$ 对应于 $\lambda(c)$ 的特征向量一般不能定义为在 $c=0$ 处的连续函数.

习 题

1. 对下列矩阵 A , 计算 e^A , e^{At} 和 $\sin At$

$$(1) A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad (2) A = \begin{pmatrix} 2 & 2 & 1 \\ -2 & 6 & 1 \\ 0 & 0 & 4 \end{pmatrix}$$

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

2. 设 A 是 n 阶矩阵, 证明:

- (1) $e^{iA} = \cos A + i \sin A$;
- (2) $e^{2\pi i I + A} = e^A$;
- (3) $\sin(2\pi I + A) = \sin A$;
- (4) $\sin^2 A + \cos^2 A = I$;
- (5) $|e^A| = e^{\operatorname{tr}(A)}$;
- (6) $\|e^A\| \leq e^{\|A\|}$.

其中 $i = \sqrt{-1}$, $\|\cdot\|$ 是相容矩阵范数.

3. 求下列三类 n 阶矩阵 A 的矩阵函数 e^A , $\cos A$, $\sin A$:

(1) A 满足 $A^2 = A$; (2) A 满足 $A^2 = I$; (3) A 满足 $A^2 = 0$.

4. 证明: 若 A 是实反对称 (反 Hermite) 矩阵, 则 e^A 为正交 (酉) 矩阵.

5. 证明: 若 A 是 Hermite 矩阵, 则 e^{iA} 为酉矩阵.

6. 设 A 是 n 阶矩阵, 证明: 若 A 的所有特征值的实部都小于零, 则 $\lim_{t \rightarrow +\infty} e^{At} = 0$.

7. 若 A 为可逆矩阵, $f(A)$ 有定义, 那么 $f(A^{-1})$ 是否有定义? 试举例说明.

8. 计算下列矩阵函数:

$$(1) A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \text{求 } A^{\frac{1}{2}} \text{ 和 } \ln A;$$

$$(2) A = \begin{pmatrix} 0 & -1 \\ 4 & 4 \end{pmatrix}, \text{求 } \arcsin \frac{A}{4};$$

$$(3) A = \begin{pmatrix} 16 & 8 \\ 8 & 4 \end{pmatrix}, \text{求 } (I + A)^{-1} \text{ 及 } A^{\frac{1}{2}}.$$

9. 设矩阵值函数

$$A(t) = \begin{pmatrix} \sin t & \cos t & t \\ \frac{\sin t}{t} & e^t & t^2 \\ 1 & 0 & t^3 \end{pmatrix}$$

其中 $t \neq 0$, 计算 $\lim_{t \rightarrow 0} A(t)$, $\frac{dA(t)}{dt}$, $\frac{d^2 A(t)}{dt^2}$, $\frac{d|A(t)|}{dt}$, $\left| \frac{dA(t)}{dt} \right|$.

10. 设 A 是 n 阶矩阵, 证明: $\frac{d}{dt} e^{At} = A e^{At} = e^{At} A$.

11. 设 $A(t) = (a_{ij}(t))$ 是 n 阶矩阵值函数. 举例说明, 对正整数 $m > 1$, 关系式

$$\frac{d}{dt} [A(t)]^m = m[A(t)]^{m-1} \frac{dA(t)}{dt}$$

一般不成立. 在什么条件下上式成立?

12. 设矩阵值函数

$$A(x) = \begin{pmatrix} e^{2x} & xe^x & x^2 \\ e^{-x} & 2e^{2x} & 0 \\ 3x & 0 & 0 \end{pmatrix}$$

计算 $\int_0^1 A(x) dx$ 和 $\frac{d}{dx} \left(\int_0^2 A(t) dt \right)$.

13. 设 $\varphi(X)$ 是 $m \times n$ 矩阵变量 $X = (x_{ij})$ 的数值函数, 并且 $X = X(t) = (x_{ij}(t))$ 是变量 t 的矩阵值函数. 证明:

$$\frac{d\varphi(X)}{dt} = \text{tr} \left(\frac{dX}{dt} \left(\frac{d\varphi(X)}{dX} \right)^T \right)$$

14. 设 $A(t) = (a_{ij}(t))$ 是 n 阶非奇异矩阵值函数. 证明

$$(1) \frac{d \text{tr}(A(t))}{dt} = \text{tr} \left(\frac{dA(t)}{dt} \right);$$

$$(2) \frac{d|A(t)|}{dt} = |A(t)| \text{tr} \left(A(t)^{-1} \frac{dA(t)}{dt} \right);$$

$$(3) \frac{d \ln(|A(t)|)}{dt} = \text{tr} \left(A(t)^{-1} \frac{dA(t)}{dt} \right).$$

15. 设 A 是 $n \times m$ 常数矩阵, B 是 m 阶常数矩阵, X 是 $m \times n$ 矩阵变量. 证明:

$$(1) \frac{d}{dX} \text{tr}(XX^T) = \frac{d}{dX} \text{tr}(X^T X) = 2X;$$

$$(2) \frac{d}{dX} \text{tr}(AX) = \frac{d}{dX} \text{tr}(X^T A^T) = A^T;$$

$$(3) \frac{d}{dX} \text{tr}(X^T BX) = (B + B^T)X.$$

16. 求解如下初值问题:

$$\begin{cases} \frac{dx}{dt} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} x(t) \\ x|_{t=0} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{cases}$$

17. 求如下微分方程组 $\frac{dx}{dt} = Ax + f(t)$ 满足初始条件 $x(0) = b$ 的解:

$$(1) A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}, \quad f(t) = \begin{bmatrix} e^{-t} \\ 4e^{-t} \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \end{bmatrix};$$

$$(2) A = \begin{bmatrix} 2 & 2 & 1 \\ -2 & 6 & 1 \\ 0 & 0 & 4 \end{bmatrix}, \quad f(t) = \begin{bmatrix} 2 \\ -1 \\ t \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}.$$

第八章 广义逆矩阵

在线性代数中,如果 A 是 n 阶非奇异矩阵,则 A 存在惟一的逆矩阵 A^{-1} . 如果线性方程组 $Ax = b$ 的系数矩阵非奇异,则该方程组存在惟一解 $x = A^{-1}b$. 但是,在许多实际问题中所遇到的矩阵 A 往往是奇异方阵或长方阵,并且线性方程组 $Ax = b$ 可能是矛盾方程组,这时应该如何将该方程组在某种意义下的解通过矩阵 A 的某种逆加以表示呢? 这就促使人们设法将矩阵逆的概念、理论和方法推广到奇异方阵或长方阵的情形.

1920 年 E. H. Moore 首先提出了广义逆矩阵的概念,但其后 30 年并未引起人们的重视. 直到 1955 年, R. Penrose 利用四个矩阵方程给出广义逆矩阵的新的更简便实用的定义之后,广义逆矩阵的研究才进入了一个新的时期,其理论和应用得到了迅速发展,已成为矩阵论的一个重要分支. 广义逆矩阵在数理统计、最优化理论、控制理论、系统识别和数字图像处理等许多领域都具有重要应用.

本章着重介绍几种常用的广义逆矩阵及其在解线性方程组中的应用. 我们仅限于对实矩阵进行讨论. 类似地,对复矩阵也有相应的结果.

§ 8.1 广义逆矩阵的概念

对 $A \in \mathbb{R}^{m \times n}$, Penrose 以简便实用的形式给出了矩阵 A 的广义逆定义,并陈述了四个条件,称为 Penrose 方程.

- (1) $AGA = A$;
- (2) $GAG = G$;
- (3) $(AG)^T = AG$;
- (4) $(GA)^T = GA$.

定义 8.1.1 对任意 $m \times n$ 矩阵 A , 如果存在某个 $n \times m$ 矩阵 G , 满足 Penrose 方程的一部分或全部, 则称 G 为 A 的广义逆矩阵.

如果广义逆矩阵 G 满足第 i 个条件, 则把 G 记作 $A^{(i)}$, 并把这类矩阵的全体记作 $A\{i\}$, 于是 $A^{(i)} \in A\{i\}$. 类似地, 把满足第 i, j 两个条件的广义逆矩阵 G 记作 $A^{(i,j)}$; 满足第 i, j, k 三个条件的广义逆矩阵 G 记作 $A^{(i,j,k)}$; 满足全部 4 个条件的广义逆矩阵 G 记作 $A^{(1,2,3,4)}$. 相应地分别有 $A\{i, j\}, A\{i, j, k\}, A\{1, 2, 3, 4\}$.

由定义 8.1.1 可知, 满足 1 个、2 个、3 个、4 个 Penrose 方程的广义逆矩阵

共有 15 种,但应用较多的是 $A^{(1)}$, $A^{(1,3)}$, $A^{(1,4)}$ 和 $A^{(1,2,3,4)}$ 四种广义逆,分别记为 A^- , A_l^- , A_m^- 和 A^+ ,并称 A^- 为减号逆或 g -逆, A_l^- 为最小二乘广义逆, A_m^- 为极小范数广义逆, A^+ 为加号逆或 Moore-Penrose 广义逆.下面分别介绍 4 种广义逆以及它们与线性方程组 $Ax=b$ 的关系.

§ 8.2 广义逆矩阵 A^- 与线性方程组的解

关于 $m \times n$ 矩阵 A 的广义逆矩阵 A^- 的存在性及其计算,有如下结果.

定理 8.2.1 设 A 是 $m \times n$ 矩阵且 $\text{rank}(A) = r \geq 1$, 如果存在非奇异矩阵 P 和 Q 使得

$$PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \quad (8.2.1)$$

则 $G \in A\{1\}$ 的充分必要条件是

$$G = Q \begin{bmatrix} I_r & K \\ L & M \end{bmatrix} P \quad (8.2.2)$$

其中 $K \in \mathbb{R}^{r \times (m-r)}$, $L \in \mathbb{R}^{(n-r) \times r}$ 和 $M \in \mathbb{R}^{(n-r) \times (m-r)}$ 是任意的矩阵.

证明 直接验证(8.2.2)定义的 G 满足 $AGA = A$, 并且满足 $AGA = A$ 的 G 具有形式(8.2.2). \square

定理 8.2.1 说明矩阵 A 的广义逆矩阵 A^- 一定存在,但一般不惟一. A 惟一的充分必要条件是 $m = n = r$, 即 A 非奇异,此时 A^- 就是 A 的逆矩阵 A^{-1} .

由减号逆的构造不难得出如下性质.

定理 8.2.2 设 A 是 $m \times n$ 矩阵, P 和 Q 分别是 m 阶和 n 阶非奇异方阵,且 $B = PAQ$, A^- 是 A 的减号逆,则

- (1) $\text{rank}(A) \leq \text{rank}(A^-)$;
- (2) AA^- 和 A^-A 是幂等矩阵,并且 $\text{rank}(AA^-) = \text{rank}(A^-A) = \text{rank}(A)$;
- (3) $Q^{-1}A^-P^{-1} \in B\{1\}$.

证明 由(8.2.2)即得(1)和(2). 因为

$$BQ^{-1}A^-P^{-1}B = PAQQ^{-1}A^-P^{-1}PAQ = PAA^-AQ = PAQ = B$$

所以 $Q^{-1}A^-P^{-1} \in B\{1\}$. \square

如果线性方程组 $Ax=b$ 有解,则称该方程组是相容方程组;否则称为不相容或矛盾方程组.下面的定理说明,减号逆可以直接给出相容方程组 $Ax =$

b 的解.

定理 8.2.3 设 A 是 $m \times n$ 矩阵, 则 $G \in A\{1\}$ 的充分必要条件为 $x = Gb$ 是相容方程组 $Ax = b$ 的解.

证明 必要性. 若 $G \in A\{1\}$, 则 G 满足 $AGA = A$. 对任意的 $b \in R(A)$, 存在 $y \in \mathbb{R}^n$ 使得 $Ay = b$. 对这个 y , 由 $AGA = A$ 有 $AGAy = Ay$, 即 $AGb = b$. 这说明 $x = Gb$ 是 $Ax = b$ 的解.

充分性. 对任意的 $y \in \mathbb{R}^n$, 令 $b = Ay$, 则 $b \in R(A)$. 因为 $x = Gb$ 是 $Ax = b$ 的解, 则有 $AGb = b$, 代入 $b = Ay$ 得 $AGAy = Ay$. 由 $y \in \mathbb{R}^n$ 的任意性可得 $AGA = A$. \square

减号逆主要应用于研究线性方程组. 下面先给出 Penrose 定理, 它在许多问题的讨论中具有重要作用. 然后由 Penrose 定理导出线性方程组的可解性、通解表达式以及减号逆的通式.

定理 8.2.4 (Penrose 定理) 设 A, B, C 分别为 $m \times n, p \times q, m \times q$ 矩阵, 则矩阵方程

$$AXB = C \quad (8.2.3)$$

有解的充分必要条件是

$$AA^-CB^-B = C \quad (8.2.4)$$

并且在有解的情况下, 其通解为

$$X = A^-CB + Y - A^-AYBB^- \quad (8.2.5)$$

其中 $Y \in \mathbb{R}^{n \times p}$ 是任意的矩阵.

证明 必要性. 如果矩阵方程 (8.2.3) 有解, 设 X 为其任一解, 则

$$C = AXB = AA^-AXB B^-B = AA^-CB^-B$$

充分性. 如果 (8.2.4) 成立, 则 $X = A^-CB^-$ 是矩阵方程 (8.2.3) 的解, 即矩阵方程 (8.2.3) 有解.

下面证明 (8.2.5) 是矩阵方程 (8.2.3) 的通解.

显然, (8.2.5) 给出的 X 满足方程 $AXB = C$; 另一方面, 矩阵方程 (8.2.3) 的任一解 X 都可通过适当选取矩阵 Y 得到, 即 X 总可写成

$$X = A^-CB + X - A^-AXBB^-$$

它具有 (8.2.5) 的形式. \square

把定理 8.2.4 应用于线性方程组 $Ax = b$, 即得其可解性条件和通解表达式.

定理 8.2.5 设 $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$, 则线性方程组 $Ax = b$ 有解的充分必要条件是

$$AA^-b = b \quad (8.2.6)$$

这时, $Ax = b$ 的通解是

$$x = A^-b + (I - A^-A)y \quad (8.2.7)$$

其中 $y \in \mathbb{R}^n$ 是任意的.

应用定理 8.2.4 于矩阵方程

$$AGA = A \quad (8.2.8)$$

可得其通解为

$$G = A^-AA^- + Y - A^-AYAA^- \quad (8.2.9)$$

其中 Y 是任意的 $n \times m$ 矩阵, A^- 是 A 的任一广义逆矩阵. 令

$$Y = A^- + Z \quad (8.2.10)$$

其中 Z 是任意的 $n \times m$ 矩阵, 则

$$G = A^-AA^- + A^- + Z - A^-A(A^- + Z)AA^- = A^- + Z - A^-AZAA^-$$

定理 8.2.6 设 $A \in \mathbb{R}^{m \times n}$, 则 $A \{1\}$ 的通式为

$$A \{1\} = \{A^- + Z - A^-AZAA^- \mid Z \in \mathbb{R}^{n \times m}\} \quad (8.2.11)$$

§ 8.3 极小范数广义逆 A_m^- 与线性方程组的极小范数解

有关矩阵 A 的极小范数广义逆 A_m^- 的存在性, 有如下结果.

定理 8.3.1 设 $m \times n$ 矩阵 A 的奇异值分解为

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^T \quad (8.3.1)$$

其中 $U \in \mathbb{R}^{m \times m}$ 和 $V \in \mathbb{R}^{n \times n}$ 是正交矩阵, $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r) > 0$, $r = \text{rank}(A) \geq 1$. 则 $G \in A \{1, 4\}$ 的充分必要条件是

$$G = V \begin{bmatrix} \Sigma^{-1} & K \\ 0 & M \end{bmatrix} U^T \quad (8.3.2)$$

其中 $K \in \mathbb{R}^{r \times (m-r)}$, $M \in \mathbb{R}^{(n-r) \times (m-r)}$ 是任意的矩阵.

证明 容易验证 (8.3.2) 定义的 G 满足 $AGA = A$ 和 $(GA)^T = GA$, 并且满足 $AGA = A$ 和 $(GA)^T = GA$ 的 G 具有形式 (8.3.2). \square

定理 8.3.1 说明矩阵 A 的极小范数广义逆 A_m^- 存在但不惟一. A 的极小

范数广义逆的全体 $A\{1,4\}$ 由下列定理表征.

定理 8.3.2 设 $A \in \mathbb{R}^{m \times n}$, $G \in A\{1,4\}$ 的充分必要条件是 G 满足

$$GAA^T = A^T \quad (8.3.3)$$

证明 如果 $G \in A\{1,4\}$, $GAA^T = (GA)^T A^T = A^T G^T A^T = A^T$. 反过来, 如果 G 满足 (8.3.3), 则 $GAA^T G^T = A^T G^T$, 即 $GA(GA)^T = (GA)^T$. 因此 GA 是对称矩阵, 即

$$(GA)^T = GA \quad (8.3.4)$$

由 (8.3.3) 和 (8.3.4), 有

$$\begin{aligned} (AGA - A)^T (AGA - A) &= ((GA)^T A^T - A^T)(AGA - A) \\ &= (GAA^T - A^T)(AGA - A) = 0 \end{aligned}$$

则 $AGA = A$. 于是, $G \in A\{1,4\}$. \square

定理 8.3.3 设 $A \in \mathbb{R}^{m \times n}$, A_m^- 是 A 的任一极小范数广义逆, 则

$$A\{1,4\} = \{G \in \mathbb{R}^{n \times m} \mid GA = A_m^- A\} \quad (8.3.5)$$

证明 如果 G 满足

$$GA = A_m^- A \quad (8.3.6)$$

则

$$\begin{aligned} AGA &= AA_m^- A = A \\ (GA)^T &= (A_m^- A)^T = A_m^- A = GA \end{aligned}$$

所以 $G \in A\{1,4\}$. 另一方面, 对任 $G \in A\{1,4\}$, 则

$$\begin{aligned} A_m^- A &= A_m^- AGA = (A_m^- A)^T (GA)^T \\ &= A^T (A_m^-)^T A^T G^T = A^T G^T = (GA)^T = GA \end{aligned} \quad \square$$

应用定理 8.2.4 于矩阵方程 (8.3.6), 即得 $A\{1,4\}$ 的通式.

定理 8.3.4 设 $A \in \mathbb{R}^{m \times n}$, A^- 是 A 的任一广义逆矩阵, A_m^- 是 A 的任一极小范数广义逆, 则

$$A\{1,4\} = \{G \in \mathbb{R}^{n \times m} \mid G = A_m^- + Z(I - AA^-), Z \in \mathbb{R}^{n \times m}\} \quad (8.3.7)$$

证明 由定理 8.2.4 知, 矩阵方程 (8.3.6) 有解, 并且其通解为

$$G = A_m^- AA^- + Y - YAA^- \quad (8.3.8)$$

其中 $Y \in \mathbb{R}^{n \times m}$. 令 $Y = A_m^- + Z, Z \in \mathbb{R}^{n \times m}$, 则 $G = A_m^- + Z(I - AA^-)$. \square

在上一节我们已经证明了相容线性方程组 $Ax = b$ 的解可以用广义逆矩阵 A^- 表示为

$$x = A^- b \quad (8.3.9)$$

并且给出了通解表达式

$$x = A^- b + (I - A^- A)y \quad (8.3.10)$$

现在我们要在相容线性方程组 $Ax = b$ 的解集中求范数最小的解, 即求广义逆矩阵 G 使得

$$\|Gb\|_2 = \min_{Ax=b} \|x\|_2 \quad (8.3.11)$$

我们称具有这种性质的解为相容线性方程组 $Ax = b$ 的极小范数解. 下面将证明与这种极小范数解相对应的广义逆矩阵就是上面讨论的极小范数广义逆.

定理 8.3.5 设 A 是 $m \times n$ 矩阵, 则 $G \in A\{1, 4\}$ 的充分必要条件为 $x = Gb$ 是相容方程组 $Ax = b$ 的极小范数解.

证明 必要性. 若 $G \in A\{1\}$, 则相容线性方程组 $Ax = b$ 的通解为

$$x = Gb + (I - GA)y$$

如果 G 还满足 $(GA)^T = GA$, 则 $x = Gb$ 是 $Ax = b$ 的极小范数解.

事实上, 由于

$$\begin{aligned} \|x\|_2^2 &= \|Gb + (I - GA)y\|_2^2 = (Gb + (I - GA)y)^T (Gb + (I - GA)y) \\ &= \|Gb\|_2^2 + \|(I - GA)y\|_2^2 + (Gb)^T (I - GA)y + ((I - GA)y)^T Gb \end{aligned}$$

对任意的 $b \in R(A)$, 存在 $z \in \mathbb{R}^n$ 使得 $Az = b$, 则

$$\begin{aligned} (Gb)^T (I - GA)y &= (GAz)^T (I - GA)y = z^T (GA)^T (I - GA)y \\ &= z^T GA(I - GA)y = z^T (GA - GAGA)y = z^T (GA - GA)y = 0 \end{aligned}$$

同理可证 $((I - GA)y)^T Gb = 0$. 因此, 有

$$\|x\|_2^2 = \|Gb\|_2^2 + \|(I - GA)y\|_2^2 \geq \|Gb\|_2^2$$

这说明 $x = Gb$ 是 $Ax = b$ 的极小范数解.

充分性. 若 $x = Gb$ 是相容线性方程组 $Ax = b$ 的极小范数解, 则由定理 8.2.3 知 $G \in A\{1\}$, 从而 $Ax = b$ 的通解为 $x = Gb + (I - GA)y$. 因为 Gb 是 $Ax = b$ 的极小范数解, 则由 (8.3.11) 知, 对任意的 $b, y \in \mathbb{R}^n$, 有

$$\|Gb\|_2 \leq \|Gb + (I - GA)y\|_2 \quad (8.3.12)$$

令 $b = Az, z \in \mathbf{R}^n$, 则

$$\|GAz\|_2 \leq \|GAz + (I - GA)y\|_2 \quad (8.3.13)$$

欲使上述不等式恒成立, 其充分必要条件是

$$(GAz)^\top(I - GA)y = 0 \quad (8.3.14)$$

对任意的 $y, z \in \mathbf{R}^n$ 恒成立. 由 y, z 的任意性, 即得

$$(GA)^\top(I - GA) = 0$$

这表明 GA 是对称矩阵, 从而 $(GA)^\top = GA$. 因此, $G \in A\{1, 4\}$. \square

虽然极小范数广义逆不惟一, 但是相容线性方程组 $Ax = b$ 的极小范数解却是惟一的.

定理 8.3.6 相容线性方程组 $Ax = b$ 的极小范数解是惟一的.

证明 对 $b \in R(A)$, 存在 $z \in \mathbf{R}^n$ 使得 $Az = b$. 设 G_1, G_2 是矩阵 A 的两个不同的极小范数广义逆, 则

$$x_1 = G_1 b = G_1 Az$$

和

$$x_2 = G_2 b = G_2 Az$$

均为 $Ax = b$ 的极小范数解. 由定理 8.3.2, 有

$$(G_1 - G_2)AA^\top = 0$$

上式两边右乘 $(G_1 - G_2)^\top$, 得

$$[(G_1 - G_2)A][(G_1 - G_2)A]^\top = 0$$

则有 $(G_1 - G_2)A = 0$, 从而

$$x_1 - x_2 = (G_1 - G_2)Az = 0$$

因此, $x_1 = x_2$. \square

§ 8.4 最小二乘广义逆 A_l 与矛盾方程组的最小二乘解

类似于定理 8.3.1, 有关矩阵 A 的最小二乘广义逆 A_l 有如下结果.

定理 8.4.1 设 $m \times n$ 矩阵 A 的奇异值分解为

$$A = U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^\top \quad (8.4.1)$$

其中 $U \in \mathbf{R}^{m \times m}$ 和 $V \in \mathbf{R}^{n \times n}$ 是正交矩阵, $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r) > 0$, $r = \text{rank}(A) \geq 1$. 则 $G \in A\{1, 3\}$ 的充分必要条件是

$$G = V \begin{bmatrix} \Sigma^{-1} & 0 \\ L & M \end{bmatrix} U^T \quad (8.4.2)$$

其中 $L \in \mathbf{R}^{(n-r) \times r}$ 和 $M \in \mathbf{R}^{(n-r) \times (m-r)}$ 是任意的矩阵.

矩阵 A 的最小二乘广义逆的全体 $A\{1, 3\}$ 由下列定理表征.

定理 8.4.2 设 $A \in \mathbf{R}^{m \times n}$, $G \in A\{1, 3\}$ 的充分必要条件是 G 满足

$$A^T A G = A^T \quad (8.4.3)$$

证明与定理 8.3.2 类似, 作为练习留给读者.

类似于定理 8.3.3, 可得

定理 8.4.3 设 $A \in \mathbf{R}^{m \times n}$, A_l 是 A 的任一最小二乘广义逆, 则

$$A\{1, 3\} = \{G \in \mathbf{R}^{n \times m} \mid AG = AA_l^-\} \quad (8.4.4)$$

证明作为练习留给读者.

应用定理 8.2.4 于矩阵方程

$$AG = AA_l^- \quad (8.4.5)$$

可得 $A\{1, 3\}$ 的通式.

定理 8.4.4 设 $A \in \mathbf{R}^{m \times n}$, A^- 是 A 的任一广义逆矩阵, A_l^- 是 A 的任一最小二乘广义逆, 则

$$A\{1, 3\} = \{G \in \mathbf{R}^{n \times m} \mid G = A_l^- + (I - A^- A)Z, Z \in \mathbf{R}^{n \times m}\} \quad (8.4.6)$$

证明 应用定理 8.2.4 于矩阵方程(8.4.5), 得其通解为

$$G = A^- AA_l^- + Y - A^- AY \quad (8.4.7)$$

其中 $Y \in \mathbf{R}^{n \times m}$. 令 $Y = A_l^- + Z$, $Z \in \mathbf{R}^{n \times m}$, 则 $G = A_l^- + (I - A^- A)Z$. \square

如果线性方程组 $Ax = b$ 不相容, 则它没有通常意义下的解, 残量 $b - Ax$ 不等于零. 但这类方程组在许多实际问题中经常出现. 对于这类方程组, 我们可以求这样的解, 使它的残量范数为最小

$$\|Ax - b\|_2 = \min \quad (8.4.8)$$

问题(8.4.8)就是求 $b \in \mathbf{R}^m$ 在 $\text{span}(A)$ 上的最佳逼近, 通常称问题(8.4.8)为线性最小二乘问题. 满足(8.4.8)的 x 称为不相容线性方程组 $Ax = b$ 的最小二乘解.

下面的定理建立了 A 的最小二乘广义逆 A_l^- 与不相容线性方程组 $Ax = b$ 的最小二乘解之间的关系.

定理 8.4.5 设 A 是 $m \times n$ 矩阵, 则 $G \in A\{1, 3\}$ 的充分必要条件为 $x = Gb$ 是不相容线性方程组 $Ax = b$ 的最小二乘解.

证明 必要性. 若 $G \in A\{1, 3\}$, 则 $x = Gb$ 是不相容线性方程组 $Ax = b$ 的最小二乘解.

事实上, 对任意的 $x \in \mathbb{R}^n$, 有

$$\begin{aligned}\|Ax - b\|_2^2 &= \|(AGb - b) + A(x - Gb)\|_2^2 \\ &= \|AGb - b\|_2^2 + \|A(x - Gb)\|_2^2 + 2(A(x - Gb))^T \\ &\quad \cdot (AGb - b)\end{aligned}\quad (8.4.9)$$

由定理 8.4.2 可得

$$(A(x - Gb))^T (AGb - b) = (x - Gb)^T A^T (AG - I)b = 0 \quad (8.4.10)$$

因此, 有

$$\|Ax - b\|_2^2 = \|AGb - b\|_2^2 + \|A(x - Gb)\|_2^2 \geq \|AGb - b\|_2^2$$

这说明 $x = Gb$ 是 $Ax = b$ 的最小二乘解.

充分性. 若 $x = Gb$ 是不相容线性方程组 $Ax = b$ 的最小二乘解, 则对任意的 $x \in \mathbb{R}^n, b \in \mathbb{R}^m$, 都有

$$\|AGb - b\|_2^2 \leq \|Ax - b\|_2^2 \quad (8.4.11)$$

由(8.4.9)知, 不等式(8.4.11)恒成立的充分必要条件是对任意的 $x \in \mathbb{R}^n, b \in \mathbb{R}^m$, 等式(8.4.10)恒成立. 由 b 和 $x - Gb$ 的任意性, 可得 $A^T(AG - I) = 0$. 由定理 8.4.2 知, $G \in A\{1, 3\}$. \square

定理 8.4.6 不相容线性方程组 $Ax = b$ 的最小二乘解必为相容线性方程组

$$A^T A x = A^T b \quad (8.4.12)$$

的解; 反之亦然.

证明 由定理 8.4.2 知, $x = A_l^- b$ 是线性方程组(8.4.12)的解, 所以线性方程组(8.4.12)相容.

如果 y 为不相容方程组 $Ax = b$ 的最小二乘解, 则由定理 1.6.11(或例 1.6.6)知, y 是方程组(8.4.12)的解.

反过来, 若 y 是方程组(8.4.12)的解, 则

$$\|Ax - b\|_2^2 = \|A(x - y) + (Ay - b)\|_2^2$$

$$\begin{aligned}
&= \|A(x-y)\|_2^2 + \|Ay-b\|_2^2 + 2(x-y)^T A^T(Ay-b) \\
&= \|A(x-y)\|_2^2 + \|Ay-b\|_2^2 \geq \|Ay-b\|_2^2
\end{aligned}$$

上式说明 y 是不相容方程组 $Ax=b$ 的最小二乘解. \square

由定理 8.4.6 可知, 当 A 为列满秩时, 不相容方程组 $Ax=b$ 的最小二乘解是惟一的.

定理 8.4.7 x 是不相容线性方程组 $Ax=b$ 的最小二乘解当且仅当 x 是相容线性方程组

$$Ax = AA_l b \quad (8.4.13)$$

的解, 并且 $Ax=b$ 的最小二乘解的通式为

$$x = A_l^- b + (I - A_l^- A)y \quad (8.4.14)$$

其中 $y \in \mathbb{R}^n$ 是任意的.

证明 由定理 8.4.5 知, $x = A_l^- b$ 是不相容线性方程组 $Ax=b$ 的最小二乘解, 并且 $x = A_l^- b$ 满足 (8.4.13), 因此线性方程组 $Ax = AA_l^- b$ 相容. 设 x 为 $Ax=b$ 的任一最小二乘解, 则有 $\|Ax-b\|_2 = \|AA_l^- b-b\|_2 = \min$. 由于

$$\begin{aligned}
\|Ax-b\|_2^2 - \|AA_l^- b-b\|_2^2 &= \|A(x-A_l^- b)\|_2^2 \\
&\quad + 2(x-A_l^- b)^T A^T(AA_l^- - I)b
\end{aligned}$$

由定理 8.4.2 及 $\|Ax-b\|_2 = \|AA_l^- b-b\|_2$, 可得

$$\|A(x-A_l^- b)\|_2 = 0$$

则 $A(x-A_l^- b)=0$, 这说明 x 是线性方程组 $Ax=AA_l^- b$ 的解.

反过来, 如果 x 是相容线性方程组 $Ax=AA_l^- b$ 的解, 则由定理 8.4.2 有

$$A^T Ax = A^T AA_l^- b = A^T b$$

由定理 8.4.6 知 x 为 $Ax=b$ 的最小二乘解.

因为 $A_l^- b$ 是相容线性方程组 $Ax=AA_l^- b$ 的一个特解, 而齐次线性方程组 $Ax=0$ 的通解为 $(I-A_l^- A)y$, 因此 $x=A_l^- b+(I-A_l^- A)y$ 是 $Ax=b$ 的最小二乘解的通式. \square

§ 8.5 广义逆矩阵 A^+ 与线性方程组的极小最小二乘解

设 A 是 $m \times n$ 的矩阵, 其秩为 $r (r \geq 1)$. 关于矩阵 A 的 Moore-Penrose

广义逆 A^+ 的存在性与惟一性, 有如下结论.

定理 8.5.1 设 A 是任意的 $m \times n$ 矩阵, A^+ 存在并且惟一.

证明 设 A 的奇异值分解为

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^T \quad (8.5.1)$$

其中 $U \in \mathbb{R}^{m \times m}$ 和 $V \in \mathbb{R}^{n \times n}$ 是正交矩阵, $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r) > 0$. 令

$$A^+ = V \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^T \quad (8.5.2)$$

直接验证便知, (8.5.2) 式定义的 A^+ 满足定义 8.1.1 中的 4 个 Penrose 方程, 故 A^+ 存在.

再证惟一性. 设矩阵 G_1, G_2 都是 A 的 Moore-Penrose 广义逆, 则

$$\begin{aligned} G_1 &= G_1 A G_1 = G_1 (A G_1)^T = G_1 G_1^T A^T = G_1 G_1^T (A G_2 A)^T \\ &= G_1 G_1^T A^T (A G_2)^T = G_1 (A G_1)^T A G_2 = G_1 A G_1 A G_2 \\ &= G_1 A G_2 = G_1 A G_2 A G_2 = (G_1 A)^T (G_2 A)^T G_2 \\ &= (G_2 A G_1 A)^T G_2 = (G_2 A)^T G_2 = G_2 A G_2 = G_2 \end{aligned} \quad \square$$

定理 8.5.1 的证明同时也给出了计算 A^+ 的一个方法. 利用 A 的满秩分解, 可以给出 A^+ 的另一个表示式.

定理 8.5.2 设 A 是 $m \times n$ 矩阵, 其满秩分解为

$$A = BC \quad (8.5.3)$$

其中 B 是 $m \times r$ 矩阵, C 是 $r \times n$ 矩阵, $\text{rank}(B) = \text{rank}(C) = \text{rank}(A) = r$, 则

$$A^+ = C^T (C C^T)^{-1} (B^T B)^{-1} B^T \quad (8.5.4)$$

证明 直接验证 (8.5.4) 式定义的 A^+ 满足定义 8.1.1 中的四个 Penrose 方程. \square

Moore-Penrose 广义逆 A^+ 的基本性质可概述为如下定理.

定理 8.5.3 设 A 是 $m \times n$ 矩阵, 则

- (1) $(A^+)^+ = A$;
- (2) $(A^+)^T = (A^T)^+$;
- (3) $A^+ A A^T = A^T = A^T A A^+$;
- (4) $(A^T A)^+ = A^+ (A^T)^+ = A^+ (A^+)^T$;
- (5) $A^+ = (A^T A)^+ A^T = A^T (A A^T)^+$;

$$(6) A^+ = A_m^- A A_l^-;$$

$$(7) \text{rank}(A) = \text{rank}(A^+) = \text{rank}(AA^+) = \text{rank}(A^+A);$$

$$(8) \text{若 } \text{rank}(A) = n, \text{ 则 } A^+ = (A^T A)^{-1} A^T;$$

$$(9) \text{若 } \text{rank}(A) = m, \text{ 则 } A^+ = A^T (AA^T)^{-1};$$

$$(10) \text{若 } U, V \text{ 分别为 } m, n \text{ 阶正交矩阵, 则 } (UAV)^+ = V^T A^+ U^T;$$

$$(11) \text{若 } A = \begin{bmatrix} R & 0 \\ 0 & 0 \end{bmatrix}, \text{ 其中 } R \text{ 为 } r \text{ 阶非奇异矩阵, 则 } A^+ = \begin{bmatrix} R^{-1} & 0 \\ 0 & 0 \end{bmatrix}_{n \times m}$$

证明 这里仅给出(6)的证明,其余留作练习.

记 $G = A_m^- A A_l^-$, 由 A_m^- 和 A_l^- 的性质, 可得

$$AGA = AA_m^- AA_l^- A = AA_l^- A = A$$

$$GAG = A_m^- AA_l^- AA_m^- AA_l^- = A_m^- AA_m^- AA_l^- = A_m^- AA_l^- = G$$

$$(AG)^T = (AA_m^- AA_l^-)^T = (AA_l^-)^T = AA_l = AG$$

$$(GA)^T = (A_m^- AA_l A)^T = (A_m^- A)^T = A_m^- A = GA$$

由 Moore-Penrose 广义逆的惟一性即得 $A^+ = G = A_m^- A A_l^-$. \square

值得指出的是 A^+ 的许多性质, A^+ 并不具备.

(1) 对任意 $m \times n$ 矩阵 A 和 $n \times p$ 矩阵 B , 等式 $(AB)^+ = B^+ A^+$ 一般不成立.

事实上, 若取 $A = (1, 1), B = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$, 则

$$AB = (1, 0), (AB)^+ = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

而

$$A^+ = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}, B^+ = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, B^+ A^+ = \begin{bmatrix} 1 \\ 1/2 \end{bmatrix}$$

可见 $(AB)^+ \neq B^+ A^+$.

如果 A 是列满秩矩阵, B 是行满秩矩阵, 由定理 8.5.2 和定理 8.5.3(8),

(9) 知, 等式 $(AB)^+ = B^+ A^+$ 成立.

(2) 对任意 $m \times n$ 矩阵 $A, AA^+ \neq A^+ A$.

(3) 对任意 $m \times n$ 矩阵 A , 若 P, Q 分别为 m, n 阶非奇异矩阵, 则

$$(PAQ)^+ \neq Q^{-1}A^+P^{-1}$$

(4) 对任意 n 阶奇异矩阵 A 和正整数 k , $(A^k)^+ \neq (A^+)^k$.

事实上, 当 $k=2$ 时, 若取 $A = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 0 \end{bmatrix}$, 则 $A^2 = \begin{bmatrix} 1/2 & 1/2 \\ 0 & 0 \end{bmatrix}$, A^+
 $= \begin{bmatrix} 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 \end{bmatrix}$, $(A^+)^2 = \begin{bmatrix} 1/2 & 0 \\ 1/2 & 0 \end{bmatrix} \neq (A^2)^+$.

利用 Moore-Penrose 广义逆可以给出线性方程组 $Ax=b$ 的可解性条件和通解表达式.

定理 8.5.4 设 $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, 则线性方程组 $Ax=b$ 有解的充分必要条件是

$$AA^+b = b \quad (8.5.5)$$

这时, $Ax=b$ 的通解是

$$x = A^+b + (I - A^+A)y \quad (8.5.6)$$

其中 $y \in \mathbb{R}^n$ 是任意的.

证明 如果线性方程组 $Ax=b$ 有解, 则由定理 8.2.5 知 (8.5.5) 成立; 反过来, 如果 (8.5.5) 成立, 则 $x = A^+b$ 就是 $Ax=b$ 的一个解. 因此线性方程组 $Ax=b$ 有解.

类似于定理 8.2.4 的证明, 可证 $Ax=b$ 的通解由 (8.5.6) 给出. \square

定理 8.5.5 设 $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, 则不相容线性方程组 $Ax=b$ 的最小二乘解的通式为

$$x = A^+b + (I - A^+A)y \quad (8.5.7)$$

其中 $y \in \mathbb{R}^n$ 是任意的.

证明 由定理 8.4.6 知, 不相容线性方程组 $Ax=b$ 的最小二乘解与相容线性方程组 $A^T Ax = A^T b$ 的解一致. 由定理 8.5.4 知, $A^T Ax = A^T b$ 的通解为

$$x = (A^T A)^+ A^T b + [I - (A^T A)^+ (A^T A)]y$$

由定理 8.5.3(5) 即得 (8.5.7). \square

不相容线性方程组 $Ax=b$ 的最小二乘解一般是不惟一的. 设 x_0 是 $Ax=b$ 的一个最小二乘解, 如果对于任意的最小二乘解 x 都有

$$\|x_0\|_2 \leq \|x\|_2 \quad (8.5.8)$$

则称 x_0 为 $Ax=b$ 的极小最小二乘解.

因为(8.5.7)给出了不相容线性方程组 $Ax=b$ 的最小二乘通解,并且

$$\|A^+b + (I - A^+A)y\|_2^2 = \|A^+b\|_2^2 + \|(I - A^+A)y\|_2^2 \geq \|A^+b\|_2^2$$

且等号成立当且仅当 $(I - A^+A)y=0$, 所以 $Ax=b$ 的极小最小二乘解惟一, 且为 $x=A^+b$. 进一步, 有如下定理.

定理 8.5.6 设 A 是 $m \times n$ 矩阵, 则 G 是 Moore-Penrose 广义逆 A^+ 的充分必要条件为 $x=Gb$ 是不相容线性方程组 $Ax=b$ 的极小最小二乘解.

证明 由定理 8.4.7 知, 不相容方程组 $Ax=b$ 的最小二乘解与相容方程组

$$Ax = AA_l^-b \quad (8.5.9)$$

的解一致. 因此, $Ax=b$ 的极小最小二乘解就是方程组(8.5.9)的极小范数解, 并且是惟一的, 即有

$$x = A_m^-AA_l^-b \quad (8.5.10)$$

由定理 8.5.3(6), 可得 $G = A_m^-AA_l^- = A^+$.

注意到上述论证是可逆的, 从而得到结论. \square

习 题

1. 设 $A \in \mathbb{R}^{m \times n}$, $\text{rank}(A) = r$. 证明: 如果存在 m 阶正交矩阵 Q 使得

$$Q^T A = \begin{pmatrix} R & S \\ 0 & 0 \end{pmatrix}$$

其中 R 是 r 阶非奇异上三角矩阵, S 是 $r \times (n-r)$ 矩阵, 则

$$A^+ = \begin{pmatrix} R^{-1} & X \\ Y & 0 \end{pmatrix} Q^T$$

其中 Y 是满足 $SY=0$ 的 $(n-r) \times r$ 任意矩阵, X 是 $r \times (m-r)$ 任意矩阵.

2. 设 $A \in \mathbb{R}^{m \times n}$, 证明:

(1) $A^+A = I_n$ 的充分必要条件是 $\text{rank}(A) = n$;

(2) $AA^+ = I_m$ 的充分必要条件是 $\text{rank}(A) = m$.

3. 求矩阵 $A = \begin{pmatrix} 2 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$ 的减号逆 A^- , 并求线性方程组 $Ax=b$ 的通解, 其中 $b = (2, 1, 1)^T$.

4. 设 $A \in \mathbb{R}^{m \times n}$, 证明: $A|1| = \{G \in \mathbb{R}^{n \times m} | G = A^+ + V(I - AA^+) + (I - A^+A)W\}$, 其中 V 和 W 分别为 $n \times m$ 和 $m \times n$ 任意矩阵.

5. 设 $A \in \mathbb{R}^{m \times n}$, 证明:

(1) 若 $G_1, G_2 \in A|1, 4|$, 则 $(G_1 - G_2)A = 0$;

$$(2) A_m = A^+ (AA^+)^{-1}.$$

6. 求矩阵 $A = \begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & 4 \end{pmatrix}$ 的极小范数广义逆 A_m^- , 并求线性方程组 $Ax = b$ 的极小范

数解, 其中 $b = (1, -1)^T$.

7. 设 $A \in \mathbb{R}^{m \times n}$, 证明:

(1) 若 $G_1, G_2 \in A\{1, 3\}$, 则 $A(G_1 - G_2) = 0$;

$$(2) A_l^- = (A^T A)^{-1} A^T.$$

8. 求矩阵 $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 1 \end{pmatrix}$ 的最小二乘广义逆 A_l^- , 并求不相容线性方程组 $Ax = b$ 的最小

二乘解, 其中 $b = (1, 0, 0)^T$.

9. 求下列矩阵的 Moore-Penrose 广义逆 A^+

$$(1) A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix} \quad (2) A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 3 \\ -1 & 3 & 1 \end{pmatrix}$$

$$(3) A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 2 & 0 & 1 & 1 \end{pmatrix} \quad (4) A = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 2 & 0 & 2 & 2 \\ 4 & 2 & 4 & 2 \end{pmatrix}$$

10. 证明: 如果 A 是 n 阶实对称矩阵, 并且 $A^2 = A$, 则 $A^+ = A$.

11. 设 A 是 n 阶正规矩阵, 证明: $AA^+ = A^+A$, 并且对任一自然数 k 有 $(A^k)^+ = (A^+)^k$.

12. 设 A 是 n 阶实矩阵, 证明: $AA^+ = A^+A$ 的充分必要条件是 $N(A) = N(A^T)$.

13. 证明: 若 $AB = 0$, 则 $B^+A^+ = 0$.

14. 设 $A \in \mathbb{R}^{m \times n}$, 证明:

$$(1) (A^T A)^+ = A^+ (A^T)^+ = A^+ (A^+)^T;$$

$$(2) (A^T A)^+ A^T = A^T (AA^T)^+ = A^+.$$

15. 用广义逆矩阵判断下列线性方程组是否相容. 如果相容, 求其通解和极小范数解; 如果不相容, 求其最小二乘通解和极小最小二乘解.

$$(1) \begin{cases} x_1 + 2x_2 + 3x_3 - x_4 = 1, \\ 3x_1 + 2x_2 + x_3 - x_4 = 1, \\ 2x_1 + 3x_2 + x_3 + x_4 = 1; \end{cases} \quad (2) \begin{cases} x_1 + x_2 = 0, \\ x_1 + x_3 = 0, \\ -x_1 = 1, \\ x_1 + x_2 + x_3 = 2. \end{cases}$$

第九章 Kronecker 积与线性矩阵方程

本章讨论含有未知矩阵的线性矩阵代数方程. 首先介绍矩阵的 Kronecker 积及其有关性质. Kronecker 积不仅在矩阵方程的研究中起着重要作用, 而且在其他方面也有许多应用.

§ 9.1 矩阵的 Kronecker 积

现在我们引入一种新的矩阵乘法运算, 由于它在实、复运算上没有区别, 因此以复矩阵进行阐述.

定义 9.1.1 设 $A = (a_{ij}) \in C^{m \times n}$, $B = (b_{ij}) \in C^{p \times q}$, 则称如下分块矩阵

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix} \in C^{mp \times nq}$$

为 A 与 B 的 **Kronecker 积** (直积, 张量积), 简记为 $A \otimes B = (a_{ij}B)$.

显然 $A \otimes B$ 和 $B \otimes A$ 是同阶矩阵, 但一般说来 $A \otimes B \neq B \otimes A$, 即矩阵的 Kronecker 积不满足交换律. 例如

$$A = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \quad B = [1 \quad -1]$$

则

$$A \otimes B = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 1 & -1 \end{bmatrix}, \quad B \otimes A = \begin{bmatrix} 1 & 0 & -1 & 0 \\ -1 & 1 & 1 & -1 \end{bmatrix}$$

对单位矩阵, 有

$$I_m \otimes I_n = I_n \otimes I_m = I_{mn}$$

定理 9.1.1 矩阵的 Kronecker 积具有下列基本性质:

- (1) 对任意复数 k , $(kA) \otimes B = A \otimes (kB) = k(A \otimes B)$;
- (2) $(A \otimes B) \otimes C = A \otimes (B \otimes C)$;

$$(3) A \otimes (B + C) = A \otimes B + A \otimes C,$$

$$(B + C) \otimes A = B \otimes A + C \otimes A;$$

$$(4) (A \otimes B)^T = A^T \otimes B^T;$$

$$(5) (A \otimes B)^H = A^H \otimes B^H;$$

$$(6) \text{ 若 } A \in C^{k \times m}, B \in C^{p \times q}, C \in C^{m \times n}, D \in C^{q \times r}, \text{ 则 } (A \otimes B)(C \otimes D) = (AC) \otimes (BD);$$

$$(7) (A \otimes B)' = A' \otimes B';$$

(8) 如果 A 和 B 都是对角矩阵、上(下)三角矩阵、实对称矩阵、Hermite 矩阵、正交矩阵、酉矩阵, 则 $A \otimes B$ 也分别是这种类型的矩阵.

证明 (1)~(5)由定义 9.1.1 即可证明.

$$\begin{aligned} (A \otimes B)(C \otimes D) &= \begin{bmatrix} a_{11}B & \cdots & a_{1m}B \\ \vdots & & \vdots \\ a_{k1}B & \cdots & a_{km}B \end{bmatrix} \begin{bmatrix} c_{11}D & \cdots & c_{1n}D \\ \vdots & & \vdots \\ c_{m1}D & \cdots & c_{mn}D \end{bmatrix} \\ &= \left(\sum_{l=1}^m a_{il}c_{lj}BD \right) = \left(\left(\sum_{l=1}^m a_{il}c_{lj} \right) BD \right) \\ &= (AC) \otimes (BD) \end{aligned}$$

由(5)和(6),有

$$(A \otimes B)(A' \otimes B^+)(A \otimes B) = (AA' A) \otimes (BB^+ B) = A \otimes B,$$

$$(A^+ \otimes B^+)(A \otimes B)(A^+ \otimes B^+) = (A' AA^+) \otimes (B^+ BB') = A^+ \otimes B^+,$$

$$\begin{aligned} [(A \otimes B)(A' \otimes B^+)]^H &= [(AA') \otimes (BB^+)]^H = (AA')^H \otimes (BB^+)^H \\ &= (AA^+) \otimes (BB^+) = (A \otimes B)(A^+ \otimes B^+), \end{aligned}$$

$$\begin{aligned} [(A^+ \otimes B^+)(A \otimes B)]^H &= [(A^+ A) \otimes (B^+ B)]^H = (A^+ A)^H \otimes (B^+ B)^H \\ &= (A' A) \otimes (B^+ B) = (A^+ \otimes B^+)(A \otimes B). \end{aligned}$$

这说明 $(A \otimes B)^+ = A^+ \otimes B^+$.

由定义 9.1.1, (5)和(6)即可证明(8). \square

定理 9.1.2 设 $f(x, y) = \sum_{i,j=0}^K c_{ij}x^i y^j$ 是变量 x, y 的复系数二元多项式,

对 $A \in C^{m \times m}, B \in C^{n \times n}$, 定义 mn 阶矩阵 $f(A, B) = \sum_{i,j=0}^K c_{ij}A^i \otimes B^j$, 其中 $A^0 = I_m, B^0 = I_n$. 如果 A 和 B 的特征值分别为 $\lambda_1, \lambda_2, \cdots, \lambda_m$ 和 $\mu_1, \mu_2, \cdots, \mu_n$, 则 $f(A, B)$ 的特征值为 $f(\lambda_i, \mu_j) (i=1, 2, \cdots, m, j=1, 2, \cdots, n)$.

证明 由 Schur 定理知存在酉矩阵 P, Q 使得

$$P^H A P = \begin{bmatrix} \lambda_1 & & * \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_m \end{bmatrix} \equiv A_1, \quad Q^H B Q = \begin{bmatrix} \mu_1 & & * \\ & \mu_2 & \\ & & \ddots \\ & & & \mu_n \end{bmatrix} \equiv B_1$$

其中 A_1, B_1 均为上三角矩阵, 由定理 9.1.1(8) 知; $P \otimes Q$ 是酉矩阵, 并且 $A_1 \otimes B_1$ 是上三角矩阵, 则

$$\begin{aligned} (P \otimes Q)^{-1} f(A, B) (P \otimes Q) &= (P \otimes Q)^H f(A, B) (P \otimes Q) \\ &= \sum_{i,j=0}^K c_{ij} (P \otimes Q)^H (A^i \otimes B^j) (P \otimes Q) \\ &= \sum_{i,j=0}^K c_{ij} (P^H A^i P \otimes Q^H B^j Q) = \sum_{i,j=0}^K c_{ij} A_1^i \otimes B_1^j = f(A_1, B_1) \end{aligned}$$

也是上三角矩阵. 因为

$$A_1^i \otimes B_1^j = \begin{bmatrix} \lambda_1^i B_1^j & & * \\ & \ddots & \\ & & \lambda_m^i B_1^j \end{bmatrix}, \lambda_1^i B_1^j = \begin{bmatrix} \lambda_1^i \mu_1^j & & * \\ & \ddots & \\ & & \lambda_1^i \mu_n^j \end{bmatrix}$$

则 $f(A_1, B_1)$ 的对角元, 即 $f(A, B)$ 的特征值为 $f(\lambda_i, \mu_j) (i=1, 2, \dots, m, j=1, 2, \dots, n)$. \square

由定理 9.1.2 可得如下结论.

定理 9.1.3 设 A, B 分别为 $m \times m$ 和 $n \times n$ 矩阵, 并且其特征值分别为 $\lambda_1, \lambda_2, \dots, \lambda_m$ 和 $\mu_1, \mu_2, \dots, \mu_n$, 则

- (1) $A \otimes B$ 的 mn 个特征值为 $\lambda_i \mu_j (i=1, \dots, m; j=1, \dots, n)$;
- (2) $A \otimes I_n + I_m \otimes B$ 的 mn 个特征值为 $\lambda_i + \mu_j (i=1, \dots, m; j=1, \dots, n)$;
- (3) $\det(A \otimes B) = (\det(A))^n (\det(B))^m$;
- (4) $\text{tr}(A \otimes B) = \text{tr}(A) \text{tr}(B)$;
- (5) 若 A, B 均为非奇异矩阵, 则 $A \otimes B$ 是非奇异矩阵, 并且

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1} \quad (9.1.1)$$

证明 由定理 9.1.2 即得(1)和(2),并且

$$\begin{aligned} |A \otimes B| &= \prod_{i=1}^m \left(\prod_{j=1}^n \lambda_i \mu_j \right) = \prod_{i=1}^m \left(\lambda_i^n \prod_{j=1}^n \mu_j \right) \\ &= \left(\prod_{i=1}^m \lambda_i^n \right) \left(\prod_{j=1}^n \mu_j^m \right) = |A|^n |B|^m \\ \operatorname{tr}(A \otimes B) &= \sum_{i=1}^m \sum_{j=1}^n \lambda_i \mu_j = \left(\sum_{i=1}^m \lambda_i \right) \left(\sum_{j=1}^n \mu_j \right) = \operatorname{tr}(A) \operatorname{tr}(B) \end{aligned}$$

由(3)可知 $A \otimes B$ 非奇异,并且由定理 9.1.1(6)有

$$(A \otimes B)(A^{-1} \otimes B^{-1}) = (AA^{-1}) \otimes (BB^{-1}) = I_{mn}$$

故(9.1.1)成立. \square

定理 9.1.4 设 $A \in C^{m \times n}, B \in C^{p \times q}$, 则

$$\operatorname{rank}(A \otimes B) = \operatorname{rank}(A) \operatorname{rank}(B)$$

证明 由定理 2.1.10 知,存在非奇异矩阵 $P \in C^{m \times m}, Q \in C^{n \times n}, S \in C^{p \times p}, T \in C^{q \times q}$ 使得

$$PAQ = \begin{bmatrix} I_{r_A} & 0 \\ 0 & 0 \end{bmatrix} = A_1, \quad SBT = \begin{bmatrix} I_{r_B} & 0 \\ 0 & 0 \end{bmatrix} = B_1$$

其中 $r_A = \operatorname{rank}(A), r_B = \operatorname{rank}(B)$. 由定理 9.1.1 有

$$A \otimes B = (P^{-1}A_1Q^{-1}) \otimes (S^{-1}B_1T^{-1}) = (P^{-1} \otimes S^{-1})(A_1 \otimes B_1)(Q^{-1} \otimes T^{-1})$$

由定理 9.1.3 知, $P^{-1} \otimes S^{-1}, Q^{-1} \otimes T^{-1}$ 均为非奇异矩阵, 则

$$\operatorname{rank}(A \otimes B) = \operatorname{rank}(A_1 \otimes B_1)$$

而 $\operatorname{rank}(A_1 \otimes B_1) = r_A r_B = \operatorname{rank}(A) \operatorname{rank}(B)$, 于是 $\operatorname{rank}(A \otimes B) = \operatorname{rank}(A) \operatorname{rank}(B)$. \square

定理 9.1.5 设 A, B 分别为 $m \times m$ 和 $n \times n$ 矩阵, 则存在一个 mn 阶排列矩阵 P 使得

$$P^T(A \otimes B)P = B \otimes A$$

证明 容易验证, 对矩阵 $A \otimes I_n$, 存在一个 mn 阶排列矩阵 P 使得

$$P^T(A \otimes I_n)P = I_n \otimes A$$

并且对这个 mn 阶排列矩阵 P 有

$$P^T(I_m \otimes B)P = B \otimes I_m$$

因为排列矩阵是正交矩阵, 则

$$\begin{aligned} P^T(A \otimes B)P &= P^T(A \otimes I_n)(I_m \otimes B)P = (P^T(A \otimes I_n)P)(P^T(I_m \otimes B)P) \\ &= (I_n \otimes A)(B \otimes I_m) = B \otimes A \end{aligned} \quad \square$$

对 Kronecker 积也有幂的概念. 记

$$A^{[k]} = \underbrace{A \otimes A \otimes \cdots \otimes A}_k$$

关于 Kronecker 积的幂, 有如下结果.

定理 9.1.6 设 $A \in C^{m \times n}$, $B \in C^{p \times q}$, 则

$$(AB)^{[k]} = A^{[k]}B^{[k]}$$

证明 对 k 作数学归纳法. 当 $k=1$ 时, 结论显然成立. 设对 $k-1$ 结论成立, 则由定理 9.1.1 有

$$\begin{aligned} (AB)^{[k]} &= (AB) \otimes (AB)^{[k-1]} = (AB) \otimes (A^{[k-1]}B^{[k-1]}) \\ &= (A \otimes A^{[k-1]})(B \otimes B^{[k-1]}) = A^{[k]}B^{[k]} \end{aligned} \quad \square$$

§ 9.2 矩阵的拉直与线性矩阵方程

9.2.1 矩阵的拉直

定义 9.2.1 设 $A = (a_{ij}) \in C^{m \times n}$, 记 $a_i = (a_{1i}, a_{2i}, \cdots, a_{mi})^T (i=1, 2, \cdots, n)$. 令

$$\text{vec}(A) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

则称 $\text{vec}(A)$ 为矩阵 A 的列拉直(列展开).

类似地, 可以考虑矩阵的行拉直(行展开).

定理 9.2.1 设 $A \in C^{m \times n}$, $B \in C^{m \times p}$, $C \in C^{p \times q}$, 则

$$\text{vec}(ABC) = (C^T \otimes A)\text{vec}(B) \quad (9.2.1)$$

证明 记 $B = (b_1, \cdots, b_p)$, $b_i \in C^n (i=1, \cdots, p)$, $C = (c_1, \cdots, c_q)$, $c_j \in C^p (j=1, \cdots, q)$, 则

$$\text{vec}(\mathbf{ABC}) = \text{vec}(\mathbf{ABc}_1, \mathbf{ABc}_2, \cdots, \mathbf{ABc}_q) = \begin{bmatrix} \mathbf{ABc}_1 \\ \mathbf{ABc}_2 \\ \vdots \\ \mathbf{ABc}_q \end{bmatrix}$$

而

$$\begin{aligned} \mathbf{ABc}_i &= c_{1i}\mathbf{Ab}_1 + c_{2i}\mathbf{Ab}_2 + \cdots + c_{pi}\mathbf{Ab}_p \\ &= (c_{1i}\mathbf{A}, c_{2i}\mathbf{A}, \cdots, c_{pi}\mathbf{A})\text{vec}(\mathbf{B}) \end{aligned}$$

故

$$\text{vec}(\mathbf{ABC}) = \begin{bmatrix} c_{11}\mathbf{A} & c_{21}\mathbf{A} & \cdots & c_{p1}\mathbf{A} \\ c_{12}\mathbf{A} & c_{22}\mathbf{A} & \cdots & c_{p2}\mathbf{A} \\ \vdots & \vdots & & \vdots \\ c_{1q}\mathbf{A} & c_{2q}\mathbf{A} & \cdots & c_{pq}\mathbf{A} \end{bmatrix} \text{vec}(\mathbf{B}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B}) \quad \square$$

由定理 9.2.1 直接得如下结论.

推论 9.2.1 设 $\mathbf{A} \in \mathbb{C}^{m \times m}$, $\mathbf{B} \in \mathbb{C}^{n \times n}$, $\mathbf{X} \in \mathbb{C}^{m \times n}$, 则

- (1) $\text{vec}(\mathbf{AX}) = (\mathbf{I}_n \otimes \mathbf{A})\text{vec}(\mathbf{X})$;
- (2) $\text{vec}(\mathbf{XB}) = (\mathbf{B}^T \otimes \mathbf{I}_m)\text{vec}(\mathbf{X})$;
- (3) $\text{vec}(\mathbf{AX} + \mathbf{XB}) = (\mathbf{I}_n \otimes \mathbf{A} + \mathbf{B}^T \otimes \mathbf{I}_m)\text{vec}(\mathbf{X})$.

9.2.2 线性矩阵方程

一般的线性矩阵方程可表示为

$$\mathbf{A}_1\mathbf{XB}_1 + \mathbf{A}_2\mathbf{XB}_2 + \cdots + \mathbf{A}_p\mathbf{XB}_p = \mathbf{C} \quad (9.2.2)$$

其中 $\mathbf{A}_i \in \mathbb{C}^{m \times m}$, $\mathbf{B}_i \in \mathbb{C}^{n \times n}$ ($i = 1, 2, \cdots, p$), $\mathbf{C} \in \mathbb{C}^{m \times n}$ 是已知矩阵, 而 $\mathbf{X} \in \mathbb{C}^{m \times n}$ 是未知矩阵. 利用矩阵的 Kronecker 积和拉直, 可以给出线性矩阵方程 (9.2.2) 的可解性及其解法.

定理 9.2.2 矩阵 $\mathbf{X} \in \mathbb{C}^{m \times n}$ 是矩阵方程 (9.2.2) 的解的充分必要条件为 $x = \text{vec}(\mathbf{X})$ 是如下线性方程组的解

$$\mathbf{G}x = \text{vec}(\mathbf{C}) \quad (9.2.3)$$

其中 $\mathbf{G} = \sum_{i=1}^p \mathbf{B}_i^T \otimes \mathbf{A}_i$.

证明 对矩阵方程 (9.2.2) 两边拉直, 并利用定理 9.2.1, 有

$$\begin{aligned}\operatorname{vec}(\boldsymbol{C}) &= \operatorname{vec}\left(\sum_{i=1}^p \boldsymbol{A}_i \boldsymbol{X} \boldsymbol{B}_i\right) = \sum_{i=1}^p \operatorname{vec}(\boldsymbol{A}_i \boldsymbol{X} \boldsymbol{B}_i) \\ &= \sum_{i=1}^p (\boldsymbol{B}_i^{\mathrm{T}} \otimes \boldsymbol{A}_i) \operatorname{vec}(\boldsymbol{X}) = \boldsymbol{G} \operatorname{vec}(\boldsymbol{X})\end{aligned}$$

因此矩阵方程(9.2.2)的解与线性方程组(9.2.3)的解相同,故定理得证. \square

由定理 9.2.2 和线性方程组的可解性理论直接得如下推论.

推论 9.2.2 矩阵方程(9.2.2)有解的充分必要条件是 $\operatorname{rank}[\boldsymbol{G}, \operatorname{vec}(\boldsymbol{C})] = \operatorname{rank}(\boldsymbol{G})$; 矩阵方程(9.2.2)有惟一解的充分必要条件是 \boldsymbol{G} 非奇异.

如果矩阵方程(9.2.2)有解,对线性方程组(9.2.3)应用定理 8.2.5,可得解 $x = \operatorname{vec}(\boldsymbol{X})$ 的通式. 如果矩阵方程(9.2.2)无解,可考虑下列最小二乘问题的解

$$\|\boldsymbol{A}_1 \boldsymbol{X} \boldsymbol{B}_1 + \boldsymbol{A}_2 \boldsymbol{X} \boldsymbol{B}_2 + \cdots + \boldsymbol{A}_p \boldsymbol{X} \boldsymbol{B}_p - \boldsymbol{C}\|_F = \min \quad (9.2.4)$$

(9.2.4)可以写成

$$\|\boldsymbol{G} \operatorname{vec}(\boldsymbol{X}) - \operatorname{vec}(\boldsymbol{C})\|_2 = \min \quad (9.2.5)$$

对最小二乘问题(9.2.5)应用定理 8.5.5,可得最小二乘解 $x = \operatorname{vec}(\boldsymbol{X})$ 的通式.

同样地,可以考虑矩阵方程(9.2.2)的极小最小二乘解.

§ 9.3 矩阵方程 $\boldsymbol{AXB} = \boldsymbol{C}$ 与矩阵最佳逼近问题

9.3.1 矩阵方程 $\boldsymbol{AXB} = \boldsymbol{C}$

考虑矩阵方程

$$\boldsymbol{AXB} = \boldsymbol{C} \quad (9.3.1)$$

其中 $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ 分别为 $m \times n, p \times q, m \times q$ 已知矩阵, \boldsymbol{X} 为 $n \times p$ 未知矩阵.

定理 8.2.4 给出了矩阵方程(9.3.1)有解的条件及有解时通解的表达式. 利用 Moore-Penrose 广义逆,可以给出矩阵方程(9.3.1)有解的另一个条件.

利用定理 9.2.1 将矩阵方程(9.3.1)两端列展开,可得线性方程组

$$(\boldsymbol{B}^{\mathrm{T}} \otimes \boldsymbol{A}) \operatorname{vec}(\boldsymbol{X}) = \operatorname{vec}(\boldsymbol{C}) \quad (9.3.2)$$

定理 9.3.1 设 $\boldsymbol{A} \in \mathbb{C}^{m \times n}, \boldsymbol{B} \in \mathbb{C}^{p \times q}, \boldsymbol{C} \in \mathbb{C}^{m \times q}$, 矩阵方程(9.3.1)有解的充分必要条件是

$$\boldsymbol{A} \boldsymbol{A}^+ \boldsymbol{C} \boldsymbol{B}^+ \boldsymbol{B} = \boldsymbol{C} \quad (9.3.3)$$

并且在有解的情况下,其通解为

$$X = A^+ CB^+ + Y - A^+ AYBB^+ \quad (9.3.4)$$

其中 $Y \in C^{m \times p}$ 是任意的.

证明 由定理 9.2.2 知矩阵方程(9.3.1)有解的充分必要条件是线性方程组(9.3.2)有解,而由定理 8.5.4 可知线性方程组(9.3.2)有解的充分必要条件是

$$(B^T \otimes A)(B^T \otimes A)^+ \text{vec}(C) = \text{vec}(C)$$

由定理 9.1.1 和定理 9.2.1,上式化为

$$\begin{aligned} \text{vec}(C) &= (B^T \otimes A)(B^T \otimes A)^+ \text{vec}(C) = (B^T \otimes A)((B^+)^T \otimes A^+) \text{vec}(C) \\ &= [(B^+ B)^T \otimes (AA^+)] \text{vec}(C) = \text{vec}(AA^+ CB^+ B) \end{aligned}$$

故矩阵方程(9.3.1)有解的充分必要条件是(9.3.3)成立.

若线性方程组(9.3.2)有解,由定理 8.5.4 知其通解为

$$\begin{aligned} \text{vec}(X) &= (B^T \otimes A)^+ \text{vec}(C) + [I - (B^T \otimes A)^+ (B^T \otimes A)] \text{vec}(Y) \\ &= ((B^+)^T \otimes A^+) \text{vec}(C) + [I - ((B^+)^T \otimes A^+)(B^T \otimes A)] \text{vec}(Y) \\ &= \text{vec}[A^+ CB^+] + \text{vec}(Y) - [((B^+)^T B^T) \otimes (A^+ A)] \text{vec}(Y) \\ &= \text{vec}[A^+ CB^+] + \text{vec}(Y) - \text{vec}[A^+ AYBB^+] \end{aligned} \quad (9.3.5)$$

因此矩阵方程(9.3.1)的通解为

$$X = A^+ CB^+ + Y - A^+ AYBB^+ \quad \square$$

如果矩阵方程(9.3.1)无解,则可以考虑如下最小二乘问题的解

$$\|AXB - C\|_F = \min \quad (9.3.6)$$

(9.3.6)可以写成

$$\|(B^T \otimes A) \text{vec}(X) - \text{vec}(C)\|_2 = \min \quad (9.3.7)$$

对最小二乘问题(9.3.7)应用定理 8.5.5,并由(9.3.5)可得最小二乘问题(9.3.6)的通解为

$$X = A^+ CB^+ + Y - A^+ AYBB^+ \quad (9.3.8)$$

其中 $Y \in C^{m \times p}$ 是任意矩阵.

对最小二乘问题(9.3.7)应用定理 8.5.6,则得矩阵方程(9.3.1)的极小最小二乘解

$$X = A^+ CB^+ \quad (9.3.9)$$

9.3.2 带约束的矩阵最佳逼近问题

在对电学、光学、自动控制的线性系统或结构系统进行复原或修正时,根据对系统的某种矩阵 $\mathbf{X} \in \mathbb{C}^{n \times p}$ 的元素的观测和先验的统计信息,得到矩阵 \mathbf{X} 的初步估计 $\tilde{\mathbf{X}} = (\tilde{x}_{ij}) \in \mathbb{C}^{n \times p}$. 但是 $\tilde{\mathbf{X}}$ 未必满足理论上对矩阵 \mathbf{X} 的约束,例如,矩阵 \mathbf{X} 应满足矩阵方程(9.3.1)或最小二乘问题(9.3.6). 为此,需要考虑这样的问题:在所有满足(9.3.1)或(9.3.6)的矩阵中,选取一个与 $\tilde{\mathbf{X}}$ 最“接近”的矩阵 $\hat{\mathbf{X}}$, 作为矩阵 \mathbf{X} 的最佳估计. 这类问题称为带约束的矩阵最佳逼近问题.

记

$$S_X = \{\mathbf{X} \in \mathbb{C}^{n \times p} \mid \|\mathbf{AXB} - \mathbf{C}\|_F = \min\} \quad (9.3.10)$$

则带约束的矩阵最佳逼近问题可表述为:对 $\tilde{\mathbf{X}} = (\tilde{x}_{ij}) \in \mathbb{C}^{n \times p}$, 在 S_X 上求 $\tilde{\mathbf{X}}$ 的最佳逼近 $\hat{\mathbf{X}}$, 即

$$\|\tilde{\mathbf{X}} - \hat{\mathbf{X}}\|_F = \inf_{\mathbf{X} \in S_X} \|\tilde{\mathbf{X}} - \mathbf{X}\|_F \quad (9.3.11)$$

定理 9.3.2 设 $\mathbf{A} \in \mathbb{C}^{m \times n}$, $\mathbf{B} \in \mathbb{C}^{p \times q}$, $\mathbf{C} \in \mathbb{C}^{m \times q}$, 则 $\tilde{\mathbf{X}} = (\tilde{x}_{ij}) \in \mathbb{C}^{n \times p}$ 在 S_X 上存在唯一的最佳逼近 $\hat{\mathbf{X}}$, 并且

$$\hat{\mathbf{X}} = \mathbf{A}^+ \mathbf{CB}^+ + \tilde{\mathbf{X}} - \mathbf{A}^+ \mathbf{A} \tilde{\mathbf{X}} \mathbf{B} \mathbf{B}^+ \quad (9.3.12)$$

证明 由(9.3.8)知 S_X 非空, 并且容易看出 S_X 是内积空间 $\mathbb{C}^{n \times p}$ 中的一个闭凸集. 由定理 1.6.13 知 $\tilde{\mathbf{X}}$ 在 S_X 上存在唯一的最佳逼近 $\hat{\mathbf{X}}$.

下面证明 $\tilde{\mathbf{X}}$ 在 S_X 上的最佳逼近 $\hat{\mathbf{X}}$ 由(9.3.12)表示.

设 \mathbf{A}, \mathbf{B} 的奇异值分解为

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \boldsymbol{\Sigma}_A & 0 \\ 0 & 0 \end{bmatrix} \mathbf{V}^H, \mathbf{B} = \mathbf{S} \begin{bmatrix} \boldsymbol{\Sigma}_B & 0 \\ 0 & 0 \end{bmatrix} \mathbf{T}^H \quad (9.3.13)$$

其中 $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2] \in \mathbb{C}^{m \times m}$, $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2] \in \mathbb{C}^{n \times n}$, $\mathbf{S} = [\mathbf{S}_1, \mathbf{S}_2] \in \mathbb{C}^{p \times p}$, $\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2] \in \mathbb{C}^{q \times q}$ 均为酉矩阵, $\boldsymbol{\Sigma}_A = \text{diag}(\sigma_1, \dots, \sigma_{r_1})$, $\sigma_i > 0 (i = 1, \dots, r_1)$, $r_1 = \text{rank}(\mathbf{A})$, $\boldsymbol{\Sigma}_B = \text{diag}(\mu_1, \dots, \mu_{r_2})$, $\mu_j > 0 (j = 1, \dots, r_2)$, $r_2 = \text{rank}(\mathbf{B})$, $\mathbf{U}_1 \in \mathbb{C}^{m \times r_1}$, $\mathbf{V}_1 \in \mathbb{C}^{n \times r_1}$, $\mathbf{S}_1 \in \mathbb{C}^{p \times r_2}$, $\mathbf{T}_1 \in \mathbb{C}^{q \times r_2}$, 则由(8.5.2)得

$$\mathbf{A}^+ \mathbf{A} = \mathbf{V}_1 \mathbf{V}_1^H, \mathbf{B} \mathbf{B}^+ = \mathbf{S}_1 \mathbf{S}_1^H, \mathbf{V}_1 \mathbf{V}_1^H + \mathbf{V}_2 \mathbf{V}_2^H = \mathbf{I}, \mathbf{S}_1 \mathbf{S}_1^H + \mathbf{S}_2 \mathbf{S}_2^H = \mathbf{I} \quad (9.3.14)$$

记

$$U^H C T = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}, V^H \tilde{X} S = \begin{bmatrix} \tilde{X}_{11} & \tilde{X}_{12} \\ \tilde{X}_{21} & \tilde{X}_{22} \end{bmatrix}, V^H Y S = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \quad (9.3.15)$$

其中 $C_{ij} = U_i^H C T_j$, $\tilde{X}_{ij} = V_i^H \tilde{X} S_j$, $Y_{ij} = V_i^H Y S_j$ ($i, j=1, 2$), 则

$$A^+ C B^+ = V \begin{bmatrix} \Sigma_A^{-1} C_{11} \Sigma_B^{-1} & 0 \\ 0 & 0 \end{bmatrix} S^H \quad (9.3.16)$$

(9.3.8)可改写为

$$X = V \begin{bmatrix} \Sigma_A^{-1} C_{11} \Sigma_B^{-1} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} S^H \quad (9.3.17)$$

从而

$$\begin{aligned} \|\tilde{X} - X\|_F &= \left\| V \left[V^H \tilde{X} S - \begin{bmatrix} \Sigma_A^{-1} C_{11} \Sigma_B^{-1} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \right] S^H \right\|_F \\ &= \left\| \begin{bmatrix} \tilde{X}_{11} - \Sigma_A^{-1} C_{11} \Sigma_B^{-1} & \tilde{X}_{12} - Y_{12} \\ \tilde{X}_{21} - Y_{21} & \tilde{X}_{22} - Y_{22} \end{bmatrix} \right\|_F \end{aligned}$$

因此, 当且仅当 $Y_{12} = \tilde{X}_{12}$, $Y_{21} = \tilde{X}_{21}$, $Y_{22} = \tilde{X}_{22}$, $\|\tilde{X} - X\|_F$ 达到最小. 于是 \tilde{X} 在 S_X 上的最佳逼近 \hat{X} 为

$$\begin{aligned} \hat{X} &= V \begin{bmatrix} \Sigma_A^{-1} C_{11} \Sigma_B^{-1} & \tilde{X}_{12} \\ \tilde{X}_{21} & \tilde{X}_{22} \end{bmatrix} S^H \\ &= A^+ C B^+ + V \begin{bmatrix} 0 & V_1^H \tilde{X} S_2 \\ V_2^H \tilde{X} S_1 & V_2^H \tilde{X} S_2 \end{bmatrix} S^H \\ &= A^+ C B^+ + V_2 V_2^H \tilde{X} S_1 S_1^H + \tilde{X} S_2 S_2^H \\ &= A^+ C B^+ + (I - V_1 V_1^H) \tilde{X} S_1 S_1^H + \tilde{X} S_2 S_2^H \\ &= A^+ C B^+ + \tilde{X} - Y_1 V_1^H \tilde{X} S_1 S_1^H \\ &= A^+ C B^+ + \tilde{X} - A^+ A \tilde{X} B B^+ \quad \square \end{aligned}$$

如果 $\tilde{X}=0$, 则得矩阵方程(9.3.1)的极小最小二乘解(9.3.9). 如果 A 是 n 阶单位矩阵, $p=n$, $C=BA$, 其中 A 是 q 阶对角矩阵, 则问题(9.3.11)称为

谱约束下的矩阵最佳逼近问题,这是一类矩阵特征值反问题.

§ 9.4 矩阵方程 $AX=B$ 的 Hermite 解 与矩阵最佳逼近问题*

在结构系统修正等实际应用中,人们要求如下矩阵方程

$$AX=B \quad (9.4.1)$$

的 Hermite 解 $X \in C^{m \times n}$, 其中 $A, B \in C^{m \times n}$.

定理 9.4.1 设 $A, B \in C^{m \times n}$ 并且 A 的奇异值分解为

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^H \quad (9.4.2)$$

其中 $U = [U_1, U_2] \in C^{m \times m}$, $V = [V_1, V_2] \in C^{n \times n}$ 均为酉矩阵, $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$, $\sigma_i > 0 (i=1, \dots, r)$, $r = \text{rank}(A)$, $U_1 \in C^{m \times r}$, $V_1 \in C^{n \times r}$, 则矩阵方程(9.4.1)有 Hermite 解的充分必要条件是

$$AB^H = BA^H, AA^+B = B \quad (9.4.3)$$

并且在有解的情况下,其通解为

$$X = A^+B + (A^+B)^H - A^+AB^H(A^+)^H + V_2GV_2^H \quad (9.4.4)$$

其中 $G \in C^{(n-r) \times (n-r)}$ 是任意的 Hermite 矩阵.

证明 必要性. 如果矩阵方程(9.4.1)有 Hermite 解 X , 则由定理 9.3.1 知, A 和 B 满足 $AA^+B=B$, 并且由 $AX=B$ 得 $AXA^H=BA^H$. 因为 X 是 Hermite 矩阵, 所以 AXA^H 也是 Hermite 矩阵, 于是 $BA^H=AXA^H=(AXA^H)^H=AB^H$.

充分性. 记

$$U^HBV = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \quad V^HXV = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \quad (9.4.5)$$

其中 $B_{ij} = U_i^HBV_j$, $X_{ij} = V_i^HXV_j$ ($i, j=1, 2$), 由 A 的奇异值分解(9.4.2)知, 矩阵方程(9.4.1)等价于

$$\begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \quad (9.4.6)$$

因为 $AA^+B=B$, 则

$$U_2^H B = 0 \quad (9.4.7)$$

于是 $B_{21}=0, B_{22}=0$. 从而矩阵方程(9.4.6)等价于

$$\begin{cases} \Sigma X_{11} = B_{11} \\ \Sigma X_{12} = B_{12} \end{cases}$$

因此 $X_{11} = \Sigma^{-1} B_{11}, X_{12} = \Sigma^{-1} B_{12}$. 由 A 的奇异值分解和(9.4.7)得

$$U^H B A^H U = \begin{pmatrix} U_1^H B V_1 \Sigma & 0 \\ 0 & 0 \end{pmatrix}$$

因为 $AB^H = BA^H$, 所以 $U_1^H B V_1 \Sigma$, 进而 $X_{11} = \Sigma^{-1} B_{11} = \Sigma^{-1} (U_1^H B V_1 \Sigma) \Sigma^{-1}$ 都是 Hermite 矩阵. 于是, 当 $X_{21} = X_{12}^H, X_{22} = G$ 是任意 $n-r$ 阶 Hermite 矩阵时,

$$\begin{aligned} X &= V \begin{bmatrix} \Sigma^{-1} U_1^H B V_1 & \Sigma^{-1} U_1^H B V_2 \\ V_2^H B^H U_1 \Sigma^{-1} & G \end{bmatrix} V^H \\ &= V_1 \Sigma^{-1} U_1^H B + V_2 V_2^H B^H U_1 \Sigma^{-1} V_1^H + V_2 G V_2^H \\ &= A^+ B + (A^+ B)^H - A^+ A B^H (A^+)^H + V_2 G V_2^H \end{aligned} \quad (9.4.8)$$

是矩阵方程(9.4.1)的 Hermite 解. \square

如果矩阵方程(9.4.1)有 Hermite 解, 其解一般是不唯一的. 记矩阵方程(9.4.1)Hermite 解的全体为 H_X . 对 $\tilde{X} = (\tilde{x}_{ij}) \in C^{n \times n}$, 可以考虑在 H_X 上求 \tilde{X} 的最佳逼近 \hat{X} , 即

$$\|\tilde{X} - \hat{X}\|_F = \inf_{X \in H_X} \|\tilde{X} - X\|_F \quad (9.4.9)$$

引理 9.4.1 设 $A \in C^{m \times n}$, 则对任意 n 阶 Hermite 矩阵 H 都有

$$\left\| A - \frac{A + A^H}{2} \right\|_F \leq \|A - H\|_F \quad (9.4.10)$$

证明 因为对任意 n 阶 Hermite 矩阵 H

$$A - \frac{A + A^H}{2} = \frac{A - H}{2} + \frac{H - A^H}{2}$$

并且 $\|H - A^H\|_F = \|(H - A)^H\|_F = \|H - A\|_F$, 则

$$\left\| A - \frac{A + A^H}{2} \right\|_F = \left\| \frac{A - H}{2} + \frac{H - A^H}{2} \right\|_F$$

$$\leq \left\| \frac{\mathbf{A} - \mathbf{H}}{2} \right\|_F + \left\| \frac{\mathbf{H} - \mathbf{A}^H}{2} \right\|_F = \|\mathbf{A} - \mathbf{H}\|_F \quad \square$$

定理 9.4.2 设 $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{m \times n}$ 满足条件(9.4.3), 则 $\tilde{\mathbf{X}} = (\tilde{x}_{ij}) \in \mathbb{C}^{n \times n}$ 在 $H_{\mathbf{X}}$ 上存在惟一的最佳逼近 $\hat{\mathbf{X}}$, 并且

$$\begin{aligned} \hat{\mathbf{X}} = & \mathbf{A}^+ \mathbf{B} + (\mathbf{A}^+ \mathbf{B})^H - \mathbf{A}^+ \mathbf{A} \mathbf{B}^H (\mathbf{A}^+)^H \\ & + \frac{1}{2} (\mathbf{I} - \mathbf{A}^+ \mathbf{A}) (\tilde{\mathbf{X}} + \tilde{\mathbf{X}}^H) (\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \end{aligned} \quad (9.4.11)$$

证明 由定理 9.4.1 知 $H_{\mathbf{X}}$ 非空, 并且 $H_{\mathbf{X}}$ 是内积空间 $\mathbb{C}^{n \times n}$ 中的一个闭凸集. 由定理 1.6.13 知 $\tilde{\mathbf{X}}$ 在 $H_{\mathbf{X}}$ 上存在惟一的最佳逼近 $\hat{\mathbf{X}}$.

令

$$\mathbf{X}_0 = \mathbf{A}^+ \mathbf{B} + (\mathbf{A}^+ \mathbf{B})^H - \mathbf{A}^+ \mathbf{A} \mathbf{B}^H (\mathbf{A}^+)^H \quad (9.4.12)$$

由(9.4.8)有

$$\mathbf{X} = \mathbf{X}_0 + \mathbf{V}_2 \mathbf{G} \mathbf{V}_2^H \quad (9.4.13)$$

$$\tilde{\mathbf{X}} - \mathbf{X} = \tilde{\mathbf{X}} - \mathbf{X}_0 - \mathbf{V}_2 \mathbf{G} \mathbf{V}_2^H = \tilde{\mathbf{X}} - \mathbf{X}_0 - \mathbf{V} \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G} \end{bmatrix} \mathbf{V}^H$$

于是

$$\begin{aligned} \|\tilde{\mathbf{X}} - \mathbf{X}\|_F^2 &= \left\| \mathbf{V}^H (\tilde{\mathbf{X}} - \mathbf{X}_0) \mathbf{V} - \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G} \end{bmatrix} \right\|_F^2 \\ &= \left\| \begin{bmatrix} \mathbf{V}_1^H (\tilde{\mathbf{X}} - \mathbf{X}_0) \mathbf{V}_1 & \mathbf{V}_1^H (\tilde{\mathbf{X}} - \mathbf{X}_0) \mathbf{V}_2 \\ \mathbf{V}_2^H (\tilde{\mathbf{X}} - \mathbf{X}_0) \mathbf{V}_1 & \mathbf{V}_2^H (\tilde{\mathbf{X}} - \mathbf{X}_0) \mathbf{V}_2 - \mathbf{G} \end{bmatrix} \right\|_F^2 \\ &= \|\mathbf{V}_1^H (\tilde{\mathbf{X}} - \mathbf{X}_0) \mathbf{V}_1\|_F^2 + \|\mathbf{V}_1^H (\tilde{\mathbf{X}} - \mathbf{X}_0) \mathbf{V}_2\|_F^2 \\ &\quad + \|\mathbf{V}_2^H (\tilde{\mathbf{X}} - \mathbf{X}_0) \mathbf{V}_1\|_F^2 + \|\mathbf{V}_2^H (\tilde{\mathbf{X}} - \mathbf{X}_0) \mathbf{V}_2 - \mathbf{G}\|_F^2 \end{aligned}$$

由引理 9.4.1 知, 当且仅当 $\mathbf{G} = \mathbf{V}_2^H (\frac{\tilde{\mathbf{X}} + \tilde{\mathbf{X}}^H}{2} - \mathbf{X}_0) \mathbf{V}_2$ 上式达到最小. 因此 $\tilde{\mathbf{X}}$ 在 $H_{\mathbf{X}}$ 上的最佳逼近 $\hat{\mathbf{X}}$ 为

$$\hat{\mathbf{X}} = \mathbf{X}_0 + \mathbf{V}_2 \mathbf{V}_2^H (\frac{\tilde{\mathbf{X}} + \tilde{\mathbf{X}}^H}{2} - \mathbf{X}_0) \mathbf{V}_2 \mathbf{V}_2^H$$

因为 $\mathbf{V}_2^H \mathbf{X}_0 \mathbf{V}_2 = 0$, $\mathbf{V}_2 \mathbf{V}_2^H = \mathbf{I} - \mathbf{A}^+ \mathbf{A}$, 则由上式即得(9.4.11). \square

定理 9.4.1 和定理 9.4.2 可应用于研究如下—类 Hermite 矩阵特征值反

问题.

给定实数 λ_i , 向量 $q_i \in C^n (i=1, 2, \dots, m, m \leq n)$ 和矩阵 $\tilde{X} \in C^{n \times n}$, 求 n 阶 Hermite 矩阵 X 使得

$$Xq_i = \lambda_i q_i, \quad i = 1, 2, \dots, m \quad (9.4.14)$$

满足(9.4.14)的 Hermite 矩阵全体记为 H_X , 并在 H_X 上求 \tilde{X} 的最佳逼近 \hat{X} , 即在 H_X 上求 n 阶 Hermite 矩阵 \hat{X} 满足(9.4.9). 这类问题在结构动力模型修正中具有重要的应用.

记

$$Q = [q_1, q_2, \dots, q_m], \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m) \quad (9.4.15)$$

则(9.4.14)等价于

$$XQ = QA \quad (9.4.16)$$

H_X 就是矩阵方程(9.4.16)的全体 Hermite 解. 因为 Hermite 矩阵 X 满足(9.4.16)当且仅当 Hermite 矩阵 X 满足

$$Q^H X = \Lambda Q^H \quad (9.4.17)$$

则由定理 9.4.1 和定理 9.4.2 即得如下结论.

定理 9.4.3 设 $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_m) \in R^{m \times m}$, $Q = [q_1, \dots, q_m] \in C^{n \times m}$, $\tilde{X} \in C^{n \times n}$, 并且 Q 的奇异值分解为

$$Q = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^H \quad (9.4.18)$$

其中 $U = [U_1, U_2] \in C^{n \times n}$, $V = [V_1, V_2] \in C^{m \times m}$ 均为酉矩阵, $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$, $\sigma_i > 0 (i=1, \dots, r)$, $r = \text{rank}(Q)$, $U_1 \in C^{n \times r}$, $V_1 \in C^{m \times r}$, 则存在 n 阶 Hermite 矩阵 X 满足(9.4.16)的充分必要条件是

$$Q^H Q \Lambda = \Lambda Q^H Q, \quad Q \Lambda Q^+ Q = Q \Lambda \quad (9.4.19)$$

如果 Q 和 Λ 满足(9.4.19), 则(9.4.16)的 Hermite 解 X 可表示为

$$X = Q \Lambda Q^+ + (Q^+)^H \Lambda Q^H - (Q^+)^H Q^H Q \Lambda Q^+ + U_2 G U_2^H \quad (9.4.20)$$

其中 $G \in C^{(n-r) \times (n-r)}$ 是任意的 Hermite 矩阵, 并且 \tilde{X} 在 H_X 上存在惟一的最佳逼近 \hat{X}

$$\begin{aligned} \hat{X} = & Q \Lambda Q^+ + (Q^+)^H \Lambda Q^H - (Q^+)^H Q^H Q \Lambda Q^+ \\ & + \frac{1}{2} (I - Q Q^+) (\tilde{X} + \tilde{X}^H) (I - Q Q^+) \end{aligned}$$

§ 9.5 矩阵方程 $AX + XB = C$ 和 $X - AXB = C^*$ 9.5.1 矩阵方程 $AX + XB = C$

设 $A \in C^{m \times m}$, $B \in C^{n \times n}$, $C \in C^{m \times n}$, 对矩阵方程

$$AX + XB = C \quad (9.5.1)$$

利用推论 9.2.1 将(9.5.1)两边列展开, 可得等价的线性方程组

$$[I_n \otimes A + B^T \otimes I_m] \text{vec}(X) = \text{vec}(C) \quad (9.5.2)$$

由推论 9.2.2 知矩阵方程(9.5.1)有解的充分必要条件是

$$\text{rank}[I_n \otimes A + B^T \otimes I_m, \text{vec}(C)] = \text{rank}[I_n \otimes A + B^T \otimes I_m] \quad (9.5.3)$$

并且矩阵方程(9.5.1)有惟一解的充分必要条件是矩阵 $I_n \otimes A + B^T \otimes I_m$ 非奇异.

定理 9.5.1 设 $A \in C^{m \times m}$, $B \in C^{n \times n}$, A 和 B 的特征值分别为 $\lambda_1, \lambda_2, \dots, \lambda_m$ 和 $\mu_1, \mu_2, \dots, \mu_n$, 则矩阵方程(9.5.1)有惟一解的充分必要条件是

$$\lambda_i + \mu_j \neq 0, \quad i = 1, 2, \dots, m; j = 1, 2, \dots, n \quad (9.5.4)$$

即 A 和 $-B$ 没有共同的特征值.

证明 因为矩阵方程(9.5.1)等价于线性方程组(9.5.2), 则由定理 9.1.2 知矩阵 $I_n \otimes A + B^T \otimes I_m$ 的特征值是 $\lambda_i + \mu_j$ ($i = 1, 2, \dots, m, j = 1, 2, \dots, n$). 由于(9.5.1)有惟一解的充分必要条件是矩阵 $I_n \otimes A + B^T \otimes I_m$ 非奇异, 即其特征值均非零, 故定理得证. \square

定理 9.5.2 设 $A \in C^{m \times m}$, $B \in C^{n \times n}$, 如果 A 和 B 的特征值均具有负实部, 则矩阵方程(9.5.1)有惟一解, 并且解 $X \in C^{m \times n}$ 可以表成

$$X = - \int_0^{+\infty} e^{At} C e^{Bt} dt \quad (9.5.5)$$

证明 由假设条件知, A 和 $-B$ 没有共同的特征值, 则矩阵方程(9.5.1)有惟一解. 因为

$$\frac{d}{dt}(e^{At} C e^{Bt}) = A e^{At} C e^{Bt} + e^{At} C e^{Bt} B$$

两端对 t 自 0 到 $+\infty$ 积分, 得

$$(e^{At} C e^{Bt})|_0^{+\infty} = A \left(\int_0^{+\infty} e^{At} C e^{Bt} dt \right) + \left(\int_0^{+\infty} e^{At} C e^{Bt} dt \right) B$$

即

$$A \left(- \int_0^{+\infty} e^{At} C e^{Bt} dt \right) + \left(- \int_0^{+\infty} e^{At} C e^{Bt} dt \right) B = C$$

这说明(9.5.5)确定的 X 为矩阵方程 $AX + XB = C$ 的解. \square

当 $B = A^H \in C^{n \times n}$, $C = -W \in C^{n \times n}$ 时, 矩阵方程 $AX + XB = C$ 变为著名的 Lyapunov 矩阵方程

$$AX + XA^H = -W \quad (9.5.6)$$

定理 9.5.3 设 $A \in C^{n \times n}$, $W \in C^{n \times n}$, 如果 A 是稳定矩阵, 则矩阵方程 (9.5.6) 有惟一解, 并且解 $X \in C^{n \times n}$ 可以表示为

$$X = \int_0^{+\infty} e^{At} W e^{A^H t} dt \quad (9.5.7)$$

进一步, 若 W 是 Hermite (非负定, 正定) 矩阵, 则解 X 也是 Hermite (非负定, 正定) 矩阵.

证明 因为矩阵方程 (9.5.6) 是矩阵方程 (9.5.1) 的特殊情况, 由定理 9.5.2 知矩阵方程 (9.5.6) 有惟一解 $X \in C^{n \times n}$, 且由 (9.5.7) 表示.

如果 W 是 Hermite 矩阵, 则由 $(e^{At})^H = e^{A^H t}$ 及 (9.5.7) 可知, X 是 Hermite 矩阵.

如果 W 是 Hermite 非负定矩阵, 则对任意 $y \in C^n$, 有

$$y^H X y = \int_0^{+\infty} (e^{A^H t} y)^H W (e^{A^H t} y) dt \geq 0 \quad (9.5.8)$$

因此 X 是 Hermite 非负定矩阵.

若 W 是 Hermite 正定矩阵, 则对任意 $y \in C^n$, 显然有 $y^H X y \geq 0$, 且等号成立当且仅当 $e^{A^H t} y = 0$. 因为 $e^{A^H t}$ 非奇异, 故 $y^H X y \geq 0$ 中等号成立当且仅当 $y = 0$, 所以 X 正定. \square

9.5.2 矩阵方程 $X - AXB = C$

设 $A \in C^{m \times m}$, $B \in C^{n \times n}$, $C \in C^{m \times n}$, 对矩阵方程

$$X - AXB = C \quad (9.5.9)$$

利用定理 9.2.1 将 (9.5.9) 两边列展开, 可得等价的线性方程组

$$[I_m - B^T \otimes A] \text{vec}(X) = \text{vec}(C) \quad (9.5.10)$$

由推论 9.2.2 知矩阵方程 (9.5.9) 有解的充分必要条件是

$$\text{rank}[I_m - B^T \otimes A, \text{vec}(C)] = \text{rank}[I_m - B^T \otimes A] \quad (9.5.11)$$

并且矩阵方程 (9.5.9) 有惟一解的充分必要条件是矩阵 $I_m - B^T \otimes A$ 非奇

异.

定理 9.5.4 设 $A \in C^{m \times m}, B \in C^{n \times n}$, A 和 B 的特征值分别为 $\lambda_1, \lambda_2, \dots, \lambda_m$ 和 $\mu_1, \mu_2, \dots, \mu_n$, 则矩阵方程(9.5.9)有惟一解的充分必要条件是

$$\lambda_i \mu_j \neq 1, \quad i = 1, \dots, m; j = 1, \dots, n \quad (9.5.12)$$

证明和定理 9.5.1 的证明完全类似, 故略去.

定理 9.5.5 设 $A \in C^{m \times m}, B \in C^{n \times n}$, 且 $\rho(A)\rho(B) < 1$, 则矩阵方程(9.5.9)有惟一解

$$X = \sum_{j=0}^{\infty} A^j C B^j \quad (9.5.13)$$

证明 如果 $\rho(A)\rho(B) < 1$, 则由定理 9.5.4 知矩阵方程(9.5.9)有惟一解. 直接验证可知 $X = \sum_{j=0}^{\infty} A^j C B^j$ 是矩阵方程(9.5.9)的解. \square

当 $B = A^H \in C^{n \times n}, C \in C^{m \times n}$ 时, 矩阵方程 $X - AXB = C$ 变为著名的 Stein 矩阵方程

$$X - AXA^H = C \quad (9.5.14)$$

由定理 9.5.5 即得如下推论.

推论 9.5.1 设 $A, C \in C^{n \times n}$, 且 $\rho(A) < 1$, 则矩阵方程 $X - AXA^H = C$ 有惟一解 $X = \sum_{j=0}^{\infty} A^j C (A^H)^j$. 进一步, 若 C 是 Hermite (非负定, 正定) 矩阵, 则解 X 也是 Hermite (非负定, 正定) 矩阵.

习 题

1. 设 $A = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}, B = \begin{bmatrix} 0 & 2 & 1 \\ 1 & 0 & -1 \end{bmatrix}$, 计算 $A \otimes B$.
2. 证明: 如果 A 和 B 均为 Hermite 正定矩阵, 则 $A \otimes B$ 是 Hermite 正定矩阵.
3. 证明: 如果 A 和 B 均为反 Hermite 矩阵, 则 $A \otimes B$ 是 Hermite 矩阵.
4. 设 A 和 B 均为 n 阶矩阵, 证明: 如果 $A^2 = A, B^2 = B$, 则 $(A \otimes B)^2 = A \otimes B$.
5. 设 $X \in R^{2 \times 2}, A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \end{bmatrix}, C = \begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix}$.
 - (1) 判断矩阵方程 $AXB = C$ 是否有解;
 - (2) 如果 $AXB = C$ 有解, 求其通解; 如果 $AXB = C$ 无解, 求其极小最小二乘解.
6. 设 $X \in R^{2 \times 2}, A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}, C = \begin{bmatrix} -2 & 4 \\ -3 & 3 \end{bmatrix}$.
 - (1) 证明: 矩阵方程 $AX - XB = C$ 有惟一解;
 - (2) 求矩阵方程 $AX - XB = C$ 的解.
7. 设 A 是 n 阶矩阵, 利用矩阵 A 的 Jordan 标准形求矩阵方程 $AX = XA$ 的解 X .

8. 设 A 是 n 阶可对角化矩阵, $\lambda_1, \lambda_2, \dots, \lambda_r$ 为其互异特征值, 并且它们的代数重数分别为 m_1, m_2, \dots, m_r . 证明: 矩阵方程 $AX = XA$ 的解空间有维数 $m_1^2 + m_2^2 + \dots + m_r^2$.

9. 设 A 是 n 阶正规矩阵, 并且其所有特征值互异. 证明: 矩阵方程 $AX = XA$ 的任意解 X 为正规矩阵.

10. 设 $X \in \mathbb{R}^{2 \times 3}$, $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 2 & 1 \\ 0 & -1 & 2 \\ 0 & 0 & -1 \end{pmatrix}$, $C = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$.

(1) 证明: 矩阵方程 $X - AXB = C$ 有惟一解;

(2) 求矩阵方程 $X - AXB = C$ 的解.

第十章 非负矩阵*

元素都是非负实数的矩阵称为非负矩阵. 这类矩阵在数理经济学、运筹学、概率论、弹性系数的微振动理论等许多领域都有重要的应用. 本章介绍非负矩阵的一些基本性质, 包括著名的 Perron-Frobenius 定理, 以及与非负矩阵有密切联系而又有特别重要应用价值的 M 矩阵等有关主要结果.

§ 10.1 非负矩阵与正矩阵

定义 10.1.1 设 $A = (a_{ij}), B = (b_{ij}) \in \mathbf{R}^{m \times n}$, 如果对所有的 i, j 都有 $a_{ij} \geq b_{ij}$, 则记为 $A \geq B$. 如果对所有的 i, j 都有 $a_{ij} > b_{ij}$, 则记为 $A > B$. 特别地, 如果 $A \geq 0$, 则称 A 是非负矩阵. 如果 $A > 0$, 则称 A 是正矩阵.

必须指出, 非负矩阵和正矩阵的概念与第五章中非负定矩阵和正定矩阵的概念是不同的, 这里的记号“ $\geq (>)$ ”与第五章中的记号“ $\geq (>)$ ”所代表的意义也完全不同.

对任意矩阵 $A = (a_{ij}) \in \mathbf{C}^{m \times n}$, 本章中我们用 $|A|$ 表示 A 的元素取模之后所得的非负矩阵, 即 $|A| = (|a_{ij}|)$; 特别地, 当 $x = (x_1, \dots, x_n)^T \in \mathbf{C}^n$ 时, $|x| = (|x_1|, \dots, |x_n|)^T$. 由定义 10.1.1 直接得到如下结论.

定理 10.1.1 设 A, B, C, D 均为 $m \times n$ 矩阵, 则

- (1) $|A| \geq 0$, 并且 $|A| = 0$ 当且仅当 $A = 0$;
- (2) 对任意复数 a 有 $|aA| = |a||A|$;
- (3) $|A + B| \leq |A| + |B|$;
- (4) 如果 $A \geq 0, B \geq 0, a, b$ 是非负实数, 则 $aA + bB \geq 0$;
- (5) 如果 $A \geq B$ 且 $C \geq D$, 则 $A + C \geq B + D$;
- (6) 如果 $A \geq B$ 且 $B \geq C$, 则 $A \geq C$.

一般说来, 由 $A \geq 0$ 和 $A \neq 0$ 不能导出 $A > 0$.

定理 10.1.2 设 A, B, C, D 均为 n 阶矩阵, x 是 n 维向量, 则

- (1) $|Ax| \leq |A||x|$;
- (2) $|AB| \leq |A||B|$;
- (3) 对任意正整数 m , 有 $|A^m| \leq |A|^m$;
- (4) 如果 $0 \leq A \leq B, 0 \leq C \leq D$, 则 $0 \leq AC \leq BD$;
- (5) 如果 $0 \leq A \leq B$, 对任意正整数 m , 有 $0 \leq A^m \leq B^m$;

(6) 如果 $A \geq 0 (A > 0)$, 对任意正整数 $m, A^m \geq 0 (A^m > 0)$;

(7) 如果 $A > 0, x \geq 0$ 且 $x \neq 0$, 则 $Ax > 0$;

(8) 如果 $|A| \leq B$, 则 $\|A\|_2 \leq \| |A| \|_2 \leq \|B\|_2$.

证明 (1)~(7)显然成立. 下面证明(8).

因为对任意 $x \in C^n$, 都有

$$|Ax| \leq |A| |x| \leq B |x|$$

则

$$\|Ax\|_2 = \| |Ax| \|_2 \leq \| |A| |x| \|_2 \leq \|B|x|\|_2$$

于是

$$\begin{aligned} \max_{\|x\|_2=1} \|Ax\|_2 &= \max_{\|x\|_2=1} \| |Ax| \|_2 \leq \max_{\|x\|_2=1} \| |A| |x| \|_2 \\ &\leq \max_{\|x\|_2=1} \|B|x|\|_2 \end{aligned}$$

由上式有 $\|A\|_2 \leq \| |A| \|_2 \leq \|B\|_2$. \square

定理 10.1.3 设 A, B 均为 n 阶矩阵, 如果 $|A| \leq B$, 则 $\rho(A) \leq \rho(|A|) \leq \rho(B)$.

证明 由定理 10.1.2(3)和(5)知, 对任意正整数 m , 有 $|A^m| \leq |A|^m \leq B^m$. 由定理 10.1.2(8), 有

$$\|A^m\|_2 \leq \| |A|^m \|_2 \leq \|B^m\|_2$$

从而

$$\|A^m\|_2^{\frac{1}{m}} \leq \| |A|^m \|_2^{\frac{1}{m}} \leq \|B^m\|_2^{\frac{1}{m}}$$

在上式中令 $m \rightarrow \infty$ 并利用定理 6.3.4 即得 $\rho(A) \leq \rho(|A|) \leq \rho(B)$.

由定理 10.1.3 立即得到如下推论.

推论 10.1.1 设 $A, B \in R^{n \times n}$, 如果 $0 \leq A \leq B$, 则 $\rho(A) \leq \rho(B)$.

推论 10.1.2 设 $A \in R^{n \times n}$, 如果 $A \geq 0$, \tilde{A} 是 A 的任一主子矩阵, 则 $\rho(\tilde{A}) \leq \rho(A)$. 特别地, $\max_{1 \leq i \leq n} |a_{ii}| \leq \rho(A)$.

证明 对任意正整数 $r (1 \leq r \leq n)$, 设 \tilde{A} 是 A 的任一 r 阶主子矩阵. 用 \hat{A} 表示把 \tilde{A} 的所有元素放在 A 的原来位置而把 0 放在其余位置所得的 n 阶矩阵, 则 $\rho(\tilde{A}) = \rho(\hat{A})$ 并且 $0 \leq \hat{A} \leq A$. 则推论 10.1.1 知, $\rho(\tilde{A}) = \rho(\hat{A}) \leq \rho(A)$. \square

1907 年 Perron 建立了正矩阵的特征值与特征向量的重要性质, 这就是下面的定理.

定理 10.1.4(Perron 定理) 设 $A \in R^{n \times n}$, 如果 $A > 0$, 则

- (1) $\rho(A)$ 为 A 的正特征值, 并且存在正向量 $y \in \mathbf{R}^n$ 使得 $Ay = \rho(A)y$;
 (2) A 的任何一个其他特征值 λ , 都有 $|\lambda| < \rho(A)$;
 (3) $\rho(A)$ 是 A 的单特征值.

证明 首先证明(1). 设 μ 是 A 的按模最大的特征值, $x = (x_1, x_2, \dots, x_n)^T$ 是相应的特征向量, 则

$$Ax = \mu x \text{ 且 } |\mu| = \rho(A) \quad (10.1.1)$$

令 $y = (|x_1|, |x_2|, \dots, |x_n|)^T$, 下面证明 y 是 A 对应于特征值 $\rho(A)$ 的正特征向量.

由于 $Ax = \mu x$, 所以对于正整数 $i (1 \leq i \leq n)$, 有

$$\mu x_i = \sum_{j=1}^n a_{ij} x_j$$

从而

$$\rho(A) |x_i| = |\mu x_i| \leq \sum_{j=1}^n a_{ij} |x_j|$$

写成矩阵形式就是

$$\rho(A)y \leq Ay$$

即

$$(A - \rho(A)I)y \geq 0 \quad (10.1.2)$$

下面证明(10.1.2)的等号成立. 用反证法, 设 $(A - \rho(A)I)y = z \neq 0$, 因为 A 是正矩阵, 且 z 是非负的非零向量, 所以 $Az > 0$. 又显然有 $Ay > 0$, 则存在 $\epsilon > 0$ 使得

$$Az \geq \epsilon Ay \quad (10.1.3)$$

因为

$$Az = A(A - \rho(A)I)y$$

所以

$$A^2 y = Az + \rho(A)Ay \geq [\epsilon + \rho(A)]Ay$$

令 $[\epsilon + \rho(A)]^{-1}A = B$, 则有 $B > 0, \rho(B) < 1$ 且

$$BAy \geq Ay$$

由上式可逐步推得

$$B^k Ay \geq Ay, \quad k = 1, 2, \dots, \quad (10.1.4)$$

因为 $\rho(\mathbf{B}) < 1$, 则由定理 6.3.2 知当 $k \rightarrow \infty$ 时, $\mathbf{B}^k \rightarrow 0$. 对不等式 (10.1.4) 两边取极限即得 $\mathbf{A}y \leq 0$. 这与 $\mathbf{A}y > 0$ 相矛盾, 因此有 $z = 0$. 于是我们证明了

$$\mathbf{A}y = \rho(\mathbf{A})y \quad (10.1.5)$$

这表明 $\rho(\mathbf{A}) = |\mu| > 0$ 是 \mathbf{A} 的特征值, 而 y 是 \mathbf{A} 的正特征向量.

为了证明 (2), 只要证明除 $\rho(\mathbf{A})$ 外, \mathbf{A} 不可能还有其他特征值 λ 满足 $|\lambda| = \rho(\mathbf{A})$.

假设 λ 是 \mathbf{A} 的满足 $|\lambda| = \rho(\mathbf{A})$ 的特征值, 相应的特征向量为 $u = (u_1, \dots, u_n)^T$, 则

$$\mathbf{A}u = \lambda u \quad (10.1.6)$$

令 $v = (|u_1|, \dots, |u_n|)^T$, 重复证明 (1) 的讨论可得

$$\mathbf{A}v = \rho(\mathbf{A})v \quad (10.1.7)$$

而由 (10.1.6) 可得

$$\lambda u_i = \sum_{j=1}^n a_{ij} u_j, \quad j = 1, 2, \dots, n$$

从而

$$\rho(\mathbf{A}) |u_i| = \left| \sum_{j=1}^n a_{ij} u_j \right| \quad (10.1.8)$$

由 (10.1.7) 和 (10.1.8) 可得

$$\left| \sum_{j=1}^n a_{ij} u_j \right| = \sum_{j=1}^n a_{ij} |u_j|, \quad i = 1, 2, \dots, n$$

因为 $a_{ij} > 0$, 则上式表明所有的 u_j 有相同的幅角 φ , 即

$$u_j = |u_j| e^{i\varphi}, \quad i = \sqrt{-1}; \quad j = 1, 2, \dots, n$$

其中 φ 是不依赖于 j 的常数. 于是 $u = e^{i\varphi} v$, 这表明 u, v 只差一个非零常数因子, 故 u 也是 \mathbf{A} 对应于特征值 $\rho(\mathbf{A})$ 的特征向量, 即

$$\mathbf{A}u = \rho(\mathbf{A})u \quad (10.1.9)$$

由 (10.1.6) 和 (10.1.9) 可得 $\lambda = \rho(\mathbf{A})$.

最后证明 (3). 令 $\mathbf{B} = \rho^{-1}(\mathbf{A})\mathbf{A} = (b_{ij})$, 则 $\mathbf{B} > 0$ 且 $\rho(\mathbf{B}) = 1$. 要证明 (3), 只须证明 1 是 \mathbf{B} 的单特征值即可, 或者说, 在 \mathbf{B} 的 Jordan 标准形中对应于特征值 1 只有一个 1 阶 Jordan 块.

根据结论 (1) 知, 存在向量 $y = (y_1, y_2, \dots, y_n)^T > 0$ 使得

$$By = y \quad (10.1.10)$$

从而对任何正整数 k 都有

$$B^k y = y \quad (10.1.11)$$

令

$$y_s = \max_i y_i > 0, \quad y_t = \min_i y_i > 0$$

则由(10.1.11)可得

$$y_s \geq y_t = \sum_{i=1}^n b_{ti}^{(k)} y_i \geq b_{ts}^{(k)} y_s \geq b_{ts}^{(k)} y_t$$

其中 $b_{ij}^{(k)}$ 表示 B^k 的 (i, j) 位置上元素. 从而有 $b_{ij}^{(k)} \leq \frac{y_s}{y_t}$, 这表明对所有 $k > 1$, $b_{ij}^{(k)}$ 是有界的.

假若 B 的 Jordan 标准形中有一个对应于特征值 1 的 Jordan 块的阶数大于 1, 不妨设其为 2, 则存在可逆矩阵 P 使得

$$B = P \begin{bmatrix} 1 & 1 & & \\ & 1 & & \\ & & J_1(\lambda_1) & \\ & & & \ddots \\ & & & & J_m(\lambda_m) \end{bmatrix} P^{-1}$$

其中 $J_i(\lambda_i) = \begin{bmatrix} \lambda_i & 1 & & 0 \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{bmatrix}$, 并且 $|\lambda_i| < 1 (i=1, \dots, m)$. 则对 $k \geq 1$ 有

$$B^k = P \begin{bmatrix} 1 & k & & \\ & 1 & & \\ & & J_1^k(\lambda_1) & \\ & & & \ddots \\ & & & & J_m^k(\lambda_m) \end{bmatrix} P^{-1}$$

这与 $b_{ij}^{(k)}$ 有界相矛盾, 故 B 的 Jordan 标准形中对应于特征值 1 的 Jordan 块是 1 阶的.

下面证明 B 的 Jordan 标准形中对应于特征值 1 的 1 阶 Jordan 块只有一个. 设 B 的 Jordan 标准形为

$$J = \begin{pmatrix} I_r & & & \\ & J_1(\lambda_1) & & \\ & & \ddots & \\ & & & J_l(\lambda_l) \end{pmatrix}$$

其中 I_r 为 r 阶单位矩阵, 且 $|\lambda_i| < 1 (i = 1, 2, \dots, l)$.

如果 $r > 1$, 令 $C = J - I$, 则由例 1.4.2 知, $\dim(N(C)) = n - \text{rank}(C) = r$. 由于 B 与 J 相似, 故 $\dim(N(B - I)) = r$. 因为 $r > 1$, 所以除有向量 y 满足 (10.1.10) 外, 必然还有另一向量 $z = (z_1, z_2, \dots, z_n)^T \in \mathbf{R}^n$ 满足

$$Bz = z \quad (10.1.12)$$

并且 z 与 y 线性无关. 令

$$\tau = \max_i \left(\frac{z_i}{y_i} \right) = \frac{z_j}{y_j} \quad (10.1.13)$$

则有 $\tau y \geq z$, 且不可能取等号. 于是

$$B(\tau y - z) > 0$$

利用 (10.1.10) 与 (10.1.12), 上式可写成

$$\tau y - z > 0$$

写出上式的第 j 个分量, 则有

$$\tau > \frac{z_j}{y_j}$$

这与 (10.1.13) 中 τ 的定义相矛盾, 故 $r = 1$. \square

定理 10.1.5 设 $A \in \mathbf{R}^{n \times n}$, 如果 $A > 0$, x 是 A 对应于特征值 $\rho(A)$ 的正特征向量, y 是 A^T 对应于特征值 $\rho(A)$ 的正特征向量, 则

$$\lim_{k \rightarrow \infty} [(\rho(A))^{-1} A]^k = (y^T x)^{-1} x y^T \quad (10.1.14)$$

证明 记 $B = \rho^{-1}(A) A$, 则 $B > 0$, 并且由定理 10.1.4 及其证明知, $\rho(B) = 1$ 是 B 的单特征值, 并且在 B 的 Jordan 标准形中对应于特征值 1 只有一个 1 阶 Jordan 块. 因此, B 的 Jordan 标准形为

$$J = \begin{bmatrix} 1 & & & \\ & J_1(\lambda_1) & & \\ & & \ddots & \\ & & & J_l(\lambda_l) \end{bmatrix}$$

其中 λ_i 是 B 的特征值, 且 $|\lambda_i| < 1 (i = 1, 2, \dots, l)$. 于是 $\lim_{k \rightarrow \infty} B^k$ 存在, 记 $\lim_{k \rightarrow \infty} B^k = P$. 因为

$$P = \lim_{k \rightarrow \infty} B^k = B \lim_{k \rightarrow \infty} B^{k-1} = BP.$$

记 $P = [p_1, p_2, \dots, p_n], p_i \in \mathbb{R}^n (i = 1, 2, \dots, n)$, 则有

$$Bp_i = p_i, \quad i = 1, 2, \dots, n$$

上式说明 p_1, p_2, \dots, p_n 都是 B 对应于特征值 1 的特征向量 (如果 $p_i \neq 0$). 因为 B 的特征值 $\rho(B) = 1$ 是单特征值, 且 x 也是 B 对应于特征值 $\rho(B) = 1$ 的正特征向量, 所以 $p_i (i = 1, \dots, n)$ 都与 x 线性相关. 不妨记为 $p_i = q_i x (i = 1, \dots, n)$, 并记 $q = [q_1, \dots, q_n]^T$, 则

$$P = [p_1, \dots, p_n] = [q_1 x, \dots, q_n x] = xq^T$$

因为 y 是 A^T 对应于特征值 $\rho(A)$ 的正特征向量, 则 y 是 B^T 对应于特征值 1 的正特征向量. 于是 $(B^T)^k y = y$, 从而 $y^T = y^T P = y^T xq^T$. 显然 $y^T x \neq 0$, 则 $q^T = (y^T x)^{-1} y^T$, 从而有

$$\lim_{k \rightarrow \infty} [(\rho(A))^{-1} A]^k = P = xq^T = (y^T x)^{-1} xy^T \quad \square$$

定理 10.1.4 的结论对一般的非负矩阵不一定成立. 例如, 4 阶非负矩阵

$$A = \begin{bmatrix} 0 & 3 & 0 & 0 \\ 3 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

容易计算 $\rho(A) = 3$, 并且 A 对应于特征值 $\rho(A) = 3$ 的特征向量为 $x = (\alpha, \alpha, \beta, 0)^T$, 其中 α, β 可取正数. 而 $\rho(A) = 3$ 不是 A 的单特征值, A 没有对应于特征值 $\rho(A) = 3$ 的正特征向量, 并且 A 有异于 $\rho(A)$ 的特征值 $\lambda = -3$ 使得 $|\lambda| = \rho(A)$. 但 $\rho(A)$ 仍是 A 的特征值, 并且 A 有对应于 $\rho(A)$ 的非负特征向量. 一般地, 有如下结果.

定理 10.1.6 (广义 Perron 定理) 设 $A \in \mathbb{R}^{n \times n}$, 如果 $A \geq 0$, 则 $\rho(A)$ 为 A

的特征值,而且其相应的特征向量 $x \geq 0$.

证明 令 k 为任一正整数, $B_k = A + \frac{1}{k}E$, 其中 E 为所有元素均为 1 的 n 阶矩阵, 则对 $k=1, 2, \dots$, 有

$$0 \leq A < B_{k+1} < B_k$$

由推论 10.1.1 得

$$\rho(A) \leq \rho(B_{k+1}) \leq \rho(B_k)$$

数列 $\{\rho(B_k)\}$ 单调下降且有下界, 故它有极限. 令 $\lim_{k \rightarrow \infty} \rho(B_k) = \lambda$, 则

$$\rho(A) \leq \lambda \quad (10.1.15)$$

因为 $B_k > 0$, 则由定理 10.1.4 可知, 存在向量 $y_k = [y_1^{(k)}, y_2^{(k)}, \dots, y_n^{(k)}]^T > 0$ 使得

$$B_k y_k = \rho(B_k) y_k \quad (10.1.16)$$

令

$$x_k = \left(\sum_{i=1}^n (y_i^{(k)})^2 \right)^{-1}, \quad y_k = [x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}]^T$$

则 $x_k > 0$, 并且

$$B_k x_k = \rho(B_k) x_k \quad (10.1.17)$$

其中 $\|x_k\|_2 = 1$.

令 $S = \{x \geq 0 \mid \|x\|_2 = 1, x \in \mathbf{R}^n\}$, 则 S 是 \mathbf{R}^n 中的有界闭集. 因为 $\{x_k\} \in S$, 所以在 $\{x_k\}$ 中存在一个收敛的子序列 $\{x_{k_m}\}$, 即

$$\lim_{m \rightarrow \infty} x_{k_m} = x \in S \quad (10.1.18)$$

因为

$$\lambda x = \lim_{m \rightarrow \infty} \rho(B_{k_m}) \lim_{m \rightarrow \infty} x_{k_m} = \lim_{m \rightarrow \infty} (\rho(B_{k_m}) x_{k_m}) = \lim_{m \rightarrow \infty} (B_{k_m} x_{k_m}) = Ax \quad (10.1.19)$$

于是 $x \neq 0$ 且 $x \geq 0$ 是 A 对应于特征值 λ 的特征向量. 由于 $\lambda \leq \rho(A)$, 则由 (10.1.15) 得

$$\lambda = \rho(A) \quad \square$$

关于非负矩阵的谱半径, 有如下估计.

定理 10.1.7 设 $A \in \mathbf{R}^{n \times n}$, 如果 $A \geq 0$, 则

(1) 如果 A 的每一行元素之和是常数, 则 $\rho(A) = \|A\|_\infty$;

(2) 如果 A 的每一列元素之和是常数, 则 $\rho(A) = \|A\|_1$;

(3) $\min_{1 \leq i \leq n} \sum_{j=1}^n a_{ij} \leq \rho(A) \leq \max_{1 \leq i \leq n} \sum_{j=1}^n a_{ij}$;

(4) $\min_{1 \leq j \leq n} \sum_{i=1}^n a_{ij} \leq \rho(A) \leq \max_{1 \leq j \leq n} \sum_{i=1}^n a_{ij}$.

证明 由定理 6.2.3 知, 对任意相容矩阵范数 $\|\cdot\|$ 有 $\rho(A) \leq \|A\|$. 如果 A 的每一行元素之和是常数, 则 $x = [1, \dots, 1]^T$ 是 A 对应于特征值 $\|A\|_\infty$ 的特征向量, 故有 $\rho(A) = \|A\|_\infty$.

把同样的讨论应用于 A^T 即得(2).

设 $\alpha = \min_{1 \leq i \leq n} \sum_{j=1}^n a_{ij}$, 构造 n 阶实矩阵 $B = (b_{ij})$; 若 $\alpha = 0$, 令 $B = 0$; 若 $\alpha > 0$, 令 $b_{ij} = \alpha a_{ij} (\sum_{j=1}^n a_{ij})^{-1}$. 则 $0 \leq B \leq A$, 并且 $\sum_{j=1}^n b_{ij} = \alpha (i = 1, \dots, n)$. 由(1)得 $\rho(B) = \alpha$, 并且由推论 10.1.1 得 $\rho(B) \leq \rho(A)$, 从而 $\alpha \leq \rho(A) \leq \|A\|_\infty$, 即(3)成立. 对 A^T 应用(3)即得(4). \square

§ 10.2 素矩阵与不可约非负矩阵

10.2.1 素矩阵

现在将 Perron 定理推广到一类非负矩阵上, 为此, 引进素矩阵的概念.

定义 10.2.1 设 A 是 n 阶非负矩阵, 如果存在一个正整数 m , 使得 $A^m > 0$, 则称 A 为素矩阵或本原矩阵.

例如

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

就是素矩阵. 因为不难验证 $A^4 > 0$. 显然, 正矩阵都是素矩阵, 但反之未必. 由定义 10.2.1 即得如下结论.

定理 10.2.1 设 A, B 均为 n 阶非负矩阵, 并且 A 是素矩阵, 则

- (1) A^T 也是素矩阵;
- (2) 对任一正整数 k , A^k 也是素矩阵;
- (3) $A + B$ 也是素矩阵.

定理 10.2.2 设 A 为 n 阶非负素矩阵, 则

- (1) $\rho(A)$ 为 A 的正特征值, 并且存在正向量 $x \in \mathbf{R}^n$ 使得 $Ax = \rho(A)x$;
- (2) A 的任何一个其他特征值 λ , 都有 $|\lambda| < \rho(A)$;
- (3) $\rho(A)$ 是 A 的单特征值.

证明 因为 A 是非负素矩阵, 所以存在正整数 m 使得 $A^m > 0$. 由定理 10.1.4 知, 存在向量 $x > 0$ 使得

$$A^m x = \rho(A^m)x \quad (10.2.1)$$

令

$$\hat{x} = x + (\rho(A))^{-1}Ax + \cdots + (\rho(A))^{-(m-1)}A^{m-1}x \quad (10.2.2)$$

则 $\hat{x} > 0$. 因为对任意正整数 k 都有 $\rho(A^k) = (\rho(A))^k$, 则

$$\begin{aligned} A\hat{x} &= A\left(\sum_{i=0}^{m-1}(\rho(A))^{-i}A^i x\right) = \sum_{i=0}^{m-2}(\rho(A))^{-i}A^{i+1}x + (\rho(A))^{1-m}A^m x \\ &= \sum_{i=1}^{m-1}(\rho(A))^{-i+1}A^i x + (\rho(A))^{1-m}\rho(A^m)x \\ &= \sum_{i=1}^{m-1}(\rho(A))^{-i+1}A^i x + \rho(A)x = \rho(A)\hat{x} \end{aligned} \quad (10.2.3)$$

上式说明 $\rho(A)$ 为 A 的正特征值, 并且相应的特征向量 $\hat{x} > 0$. 因为 x 和 \hat{x} 都是 A^m 对应于单特征值 $\rho(A^m)$ 的正特征向量, 所以 x 和 \hat{x} 线性相关. 因此 x 也是 A 对应于特征值 $\rho(A)$ 的正特征向量.

下面证明 $\rho(A)$ 是 A 的单特征值. 显然, $\rho(A)$ 作为 A 的特征值的重数 (代数重数) 不超过 $\rho(A^m) = (\rho(A))^m$ 作为 A^m 的特征值的重数. 由于 $(\rho(A))^m$ 是 A^m 的单特征值, 故 $\rho(A)$ 也是 A 的单特征值.

最后证明 A 的任一其他特征值 λ 都满足 $|\lambda| < \rho(A)$. 采用反证法. 若 A 有一特征值 λ 满足 $|\lambda| = \rho(A)$, 则存在非零向量 y 使得 $Ay = \lambda y$, 从而有 $A^m y = \lambda^m y$. 于是 $|\lambda^m| = (\rho(A))^m = \rho(A^m)$. 因为对 $A^m > 0$ 的任何其他特征值 μ 都有 $|\mu| < \rho(A^m)$, 所以 $\lambda^m = \rho(A^m)$. 于是

$$A^m y = \rho(A^m)y$$

由于 $\rho(A^m)$ 是 A^m 的单特征值, 则 y 与 x 线性相关. 因此 y 也是 A 对应于特征值 $\rho(A)$ 的特征向量, 从而 $\lambda = \rho(A)$. \square

完全类似于定理 10.1.5, 可以证明如下定理.

定理 10.2.3 设 A 是 n 阶非负素矩阵, x 是 A 对应于特征值 $\rho(A)$ 的正特征向量, y 是 A^T 对应于特征值 $\rho(A)$ 的正特征向量, 则

$$\lim_{k \rightarrow \infty} [(\rho(A))^{-1}A]^k = (y^T x)^{-1}xy^T \quad (10.2.4)$$

10.2.2 不可约非负矩阵

下面将 Perron 定理推广到更一般的非负矩阵上, 首先介绍不可约矩阵的概念.

定义 10.2.2 设 $A \in \mathbf{R}^{n \times n}$, 如果存在 n 阶排列矩阵 P , 使得

$$PAP^T = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \quad (10.2.5)$$

其中 $A_{kk} \in \mathbf{R}^{k \times k}$ ($1 \leq k \leq n-1$), 则称 A 为可约(可分)矩阵; 否则称 A 为不可约矩阵.

由定义可知, 一阶方阵、正矩阵以及非负素矩阵都是不可约的.

从定义 10.2.2 直接可得如下结论.

定理 10.2.4 设 A 是 n 阶矩阵, 则

- (1) A 为不可约矩阵的充分必要条件是 A^T 为不可约矩阵;
- (2) 如果 A 是不可约非负矩阵, B 是 n 阶非负矩阵, 则 $A+B$ 是不可约非负矩阵.

对一个给定的矩阵, 直接根据定义 10.2.2 判断其是否可约, 绝非一件轻而易举的事. 因为 n 阶矩阵共有 $n!$ 个排列矩阵, 逐一去尝试是不可能的, 下面给出一个判断非负矩阵是否可约的办法.

定理 10.2.5 n 阶非负矩阵 A 不可约的充分必要条件是存在正整数 $s \leq n-1$, 使得

$$(I+A)^s > 0$$

证明 必要性. 只要证明对任意不等于零的非负向量 $x \geq 0$ 都有

$$(I+A)^{n-1}x > 0 \quad (10.2.6)$$

也就证明了必要性. 为此, 令

$$\begin{cases} x_0 = x \\ x_k = (I+A)x_{k-1}, \quad k = 1, 2, \dots, n-1 \end{cases} \quad (10.2.7)$$

设 m_k 为向量 x_k 中零元素的个数, 显然 $m_0 \leq n-1$. 如果能证明当 x_{k-1} ($1 \leq k \leq n-1$) 不是正向量时有

$$m_k < m_{k-1} \quad (10.2.8)$$

则(10.2.6)成立.

采用反证法证明(10.2.8). 因为 $x_k = x_{k-1} + Ax_{k-1}$ 且 $A \geq 0$, 所以 $m_k \leq$

m_{k-1} . 假设 x_{k-1} ($1 \leq k \leq n-1$) 不是正向量但 $m_k = m_{k-1}$, 则存在排列矩阵 P 使得

$$Px_{k-1} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \quad Px_k = \begin{pmatrix} \beta \\ 0 \end{pmatrix} \quad (10.2.9)$$

其中 $\alpha, \beta \in \mathbf{R}^{n-m_{k-1}}$, 并且 $\alpha > 0, \beta > 0$. 由 (10.2.7) 得

$$Px_k = Px_{k-1} + PAP^T(Px_{k-1}) \quad (10.2.10)$$

记

$$PAP^T = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

其中 $A_{11} \in \mathbf{R}^{(n-m_{k-1}) \times (n-m_{k-1})}$, 则 (10.2.10) 化为

$$\begin{pmatrix} \beta \\ 0 \end{pmatrix} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix} + \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \quad (10.2.11)$$

从上式不难导出 $A_{21} = 0$. 这与 A 是不可约矩阵矛盾, 从而证明了 (10.2.8).

充分性. 设存在正整数 $s \leq n-1$ 使得 $(I+A)^s > 0$, 如果 A 是可约的, 则存在排列矩阵 P 使得 (10.2.5) 成立. 从而有

$$P(I+A)P^T = \begin{pmatrix} A_{11} + I & A_{12} \\ 0 & A_{22} + I \end{pmatrix} = \begin{pmatrix} \tilde{A}_{11} & A_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix}$$

其中 $\tilde{A}_{ii} = A_{ii} + I$ ($i=1, 2$). 对任意正整数 k , 都有

$$P(I+A)^k P^T = \begin{pmatrix} \tilde{A}_{11} & A_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix}^k$$

$$(I+A)^k = P^T \begin{pmatrix} \tilde{A}_{11} & A_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix}^k P$$

上式表明无论正整数 k 为何值, $(I+A)^k$ 中永远有零元素, 这与 $(I+A)^s > 0$ 矛盾. 因此 A 是不可约的. \square

1912 年 Frobenius 把 Perron 定理推广到不可约非负矩阵上.

定理 10.2.6 (Perron-Frobenius 定理) 设 A 是 n 阶不可约非负矩阵, 则

- (1) $\rho(A)$ 为 A 的正特征值, 并且存在正向量 $x \in \mathbf{R}^n$ 使得 $Ax = \rho(A)x$;
- (2) $\rho(A)$ 是 A 的单特征值;

(3) 当 A 的任一元素增加时, $\rho(A)$ 增加.

证明 因为 $A \geq 0$, 则由定理 10.1.6 知 $\rho(A)$ 是 A 的特征值. 又因为 A 是不可约的, 则由定理 10.1.7 知 $\rho(A) > 0$. 因此 $\rho(A)$ 是 A 的正特征值. 由定理 10.1.6 知存在非负向量 $x \in \mathbb{R}^n$ 且 $x \neq 0$ 使得 $Ax = \rho(A)x$, 于是 $(I + A)^{n-1}x = (1 + \rho(A))^{n-1}x$. 从定理 10.2.5 可知 $(I + A)^{n-1} > 0$, 而由定理 10.1.2(7) 有 $(I + A)^{n-1}x > 0$. 因此 $x = (1 + \rho(A))^{1-n}(I + A)^{n-1}x > 0$. 故 (1) 得证.

为了证明 (2), 采用反证法. 如果 $\rho(A)$ 是 A 的重特征值, 则 $1 + \rho(A) = \rho(I + A)$ 是 $I + A$ 的重特征值, 从而 $(1 + \rho(A))^{n-1} = (\rho(I + A))^{n-1} = \rho((I + A)^{n-1})$ 是 $(I + A)^{n-1}$ 的重特征值. 另一方面, 因为 $(I + A)^{n-1} > 0$, 由定理 10.2.2 知 $\rho((I + A)^{n-1})$ 是 $(I + A)^{n-1}$ 的单特征值. 这个矛盾说明 $\rho(A)$ 是 A 的单特征值.

(3) 由推论 10.1.1 即得. \square

对一个非负素矩阵 A , 除特征值 $\rho(A)$ 外其余特征值的模都小于 $\rho(A)$. 对不可约非负矩阵这个结论不再成立. 例如

$$A = \begin{bmatrix} 0 & 4 \\ 4 & 0 \end{bmatrix}$$

是不可约非负矩阵, 但不是素矩阵. 直接计算得 A 的两个特征值为 $\lambda_{1,2} = \pm 4$. 故 $\rho(A) = 4$, 并且 A 有一个特征值 -4 的模等于其谱半径.

§ 10.3 随机矩阵

本节简要介绍一类重要的非负矩阵——随机矩阵, 给出随机矩阵的定义, 讨论随机矩阵的一些基本性质, 并说明随机矩阵在 Markov 链中的应用.

首先考虑随机矩阵的一个背景材料. 设某个过程或系统可能出现 n 个状态 s_1, \dots, s_n . 如果过程在时刻 $k-1$ ($k \geq 1$) 处于状态 s_i , 则下一时刻将以概率 $t_{ij}(k)$ 转移到状态 s_j . 转移概率 $t_{ij}(k)$ 对所有的 i, j, k 都是已知的. 一旦给定了系统的初始状态, 系统的概率特性也就完全确定了. 如果对所有 $k \geq 1$ 概率 $t_{ij}(k)$ 都与 k 无关, 则称这个过程为有限齐次 Markov 过程(链). 这时, n 阶矩阵 $T = (t_{ij})$ 满足 $t_{ij} \geq 0$ ($i, j = 1, \dots, n$) 且 $\sum_{j=1}^n t_{ij} = 1$ ($i = 1, \dots, n$), 我们称之为该过程的转移矩阵.

定义 10.3.1 设 $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ 是非负矩阵, 如果 A 满足

$$\sum_{j=1}^n a_{ij} = 1, \quad i = 1, 2, \dots, n \quad (10.3.1)$$

则称 A 为**随机矩阵**. 如果 A 还满足

$$\sum_{i=1}^n a_{ij} = 1, \quad j = 1, 2, \dots, n \quad (10.3.2)$$

则称 A 为**双随机矩阵**.

有限齐次 Markov 过程的转移矩阵是随机矩阵. 随机矩阵在有限 Markov 过程理论中起着基本的作用, 它也经常出现在数理经济学及运筹学的各种模型问题中.

随机矩阵是一类特殊的非负矩阵, 因此前面所述非负矩阵的各种概念和结果, 对随机矩阵也适用. 下面考虑随机矩阵的一些特殊性质.

容易验证: 同阶随机矩阵之积仍是随机矩阵, 并且由定理 10.1.7 知, 随机矩阵的谱半径等于 1.

从定义可知, n 阶随机矩阵 A 有特征值 1, 并且有相应的正特征向量 $e = (1, 1, \dots, 1)^T$. 反之, 如果 n 阶非负矩阵 A 有特征值 1 且对应于 1 的特征向量为 $e = (1, 1, \dots, 1)^T$, 则 A 是随机矩阵. 于是我们得到下述定理.

定理 10.3.1 n 阶非负矩阵 A 是随机矩阵的充分必要条件是 $e = (1, 1, \dots, 1)^T \in \mathbb{R}^n$ 为 A 对应于特征值 1 的特征向量, 即 $Ae = e$.

具有正谱半径与对应正特征向量的非负矩阵与随机矩阵之间存在着密切关系.

定理 10.3.2 设 n 阶非负矩阵 A 的谱半径 $\rho(A) > 0$, 且有 $x = (x_1, x_2, \dots, x_n)^T > 0$ 使得 $Ax = \rho(A)x$, 则存在对角元为正数的对角矩阵 D 使得 $\frac{1}{\rho(A)} D^{-1}AD$ 为随机矩阵.

证明 因为存在 $x = (x_1, x_2, \dots, x_n)^T > 0$ 使得 $Ax = \rho(A)x$, 令

$$D = \text{diag}(x_1, x_2, \dots, x_n), P = \frac{1}{\rho(A)} D^{-1}AD$$

则 P 是非负矩阵, 并且满足

$$Pe = PD^{-1}x = \frac{1}{\rho(A)} D^{-1}Ax = D^{-1}x = e$$

由定理 10.3.1 知 P 是随机矩阵. \square

定理 10.3.3 设 A 为 n 阶随机矩阵, 则 A 对应于特征值 1 的初等因子是线性的.

证明 采用反证法. 假若 A 对应于特征值 1 的初等因子中有一个是非线性的, 不妨设其为 $(\lambda - 1)^2$, 则存在可逆矩阵 P 使得

$$A = P \begin{bmatrix} I_k & & & \\ & 1 & 1 & \\ & & 1 & \\ & & & J_1(\lambda_1) \\ & & & & \ddots \\ & & & & & J_s(\lambda_s) \end{bmatrix} P^{-1}$$

其中 $J_i(\lambda_i)$ 是 Jordan 块, 并且 $|\lambda_i| \leq 1 (i=1, \dots, s)$. 从而对任何 $m \geq 1$ 有

$$A^m = P \begin{bmatrix} I_k & & & \\ & 1 & m & \\ & & 1 & \\ & & & J_1^m(\lambda_1) \\ & & & & \ddots \\ & & & & & J_s^m(\lambda_s) \end{bmatrix} P^{-1}$$

因为 A^m 仍是随机矩阵, 所以 A^m 有界, 而上式与 A^m 有界矛盾. 故 A 对应于特征值 1 的初等因子是线性的. \square

在实际应用中, 经常要考虑随机矩阵 A 的幂序列 $\{A^m\}$ 的收敛性.

定理 10.3.4 设 A 为 n 阶随机矩阵, 则 $\lim_{m \rightarrow \infty} A^m$ 存在的充分必要条件是 A 的不等于 1 的特征值之模均小于 1.

证明 由定理 10.3.3 知, 存在可逆矩阵 P 使得

$$A = P \begin{bmatrix} I_k & & & \\ & J_1(\lambda_1) & & \\ & & \ddots & \\ & & & J_s(\lambda_s) \end{bmatrix} P^{-1} \quad (10.3.3)$$

其中 $J_i(\lambda_i)$ 是 Jordan 块, 并且 $\lambda_i \neq 1, |\lambda_i| \leq 1 (i=1, \dots, s)$. 则对任何 $m \geq 1$ 有

$$A^m = P \begin{bmatrix} I_k & & & \\ & J_1^m(\lambda_1) & & \\ & & \ddots & \\ & & & J_s^m(\lambda_s) \end{bmatrix} P^{-1} \quad (10.3.4)$$

由上式可见, $\lim_{m \rightarrow \infty} A^m$ 存在的充分必要条件是 $|\lambda_i| < 1 (i = 1, \dots, s)$. \square

现在考虑随机矩阵在齐次 Markov 链中的应用. 用 $\mu = (T, \pi^0)$ 表示某个有限齐次 Markov 链, 它是以 $T \in \mathbf{R}^{n \times n}$ 为转移矩阵, $\pi^0 = (\pi_1^0, \dots, \pi_n^0)$ 为初始概率分布向量的 Markov 过程, π_i^0 表示过程在初始时刻处于状态 s_i 的概率 ($1 \leq i \leq n$). 一般地, 用 $\pi^k = (\pi_1^k, \dots, \pi_n^k)$ 表示第 k 个概率分布向量, 其中 π_i^k 表示过程在 k 步后处在状态 s_i 的概率 ($1 \leq i \leq n$). 则显然有

$$\pi_i^k \geq 0, \quad i = 1, \dots, n, \quad \sum_{i=1}^n \pi_i^k = 1, \quad k = 0, 1, 2, \dots$$

如果 $\sum_{i=1}^n \pi_i^k \neq 1$, 则称 π^k 为分布向量.

定理 10.3.5 设 $\mu = (T, \pi^0)$ 为有限齐次 Markov 链, 则对正整数 $k \geq 1$ 有

$$\pi^k = \pi^{k-1} T = \pi^0 T^k \quad (10.3.5)$$

证明 对任意正整数 $k \geq 1$, 由于过程在第 k 步处于状态 s_i 的概率完全由它前一步所处的状态 $s_j (j = 1, \dots, n)$ 所决定, 故过程从第 $k-1$ 步的状态 s_j 转移到第 k 步的状态 s_i 的概率为 $\pi_j^{k-1} t_{ji}$. 因此, 过程在第 k 步处于状态 s_i 的总概率为

$$\pi_i^k = \sum_{j=1}^n \pi_j^{k-1} t_{ji}$$

由此即得 (10.3.5). \square

研究 Markov 链, 一个重要的问题是讨论当 $k \rightarrow \infty$ 时 π^k 的变化趋势. 由定理 10.3.5 知这主要取决于当 $k \rightarrow \infty$ 时 T^k 的变化趋势.

例 10.3.1 考虑某一地区人口流动问题. 设该地区有专门人才 1800 人, 分布在 A, B, C 三个单位. 每年每个单位把所有人才各分 $\frac{1}{2}$ 与其他两个单位交流. 今年 A, B, C 三个单位各有人才分别为 200 人, 600 人, 1000 人, 问明年、后年的分布情况如何? 很多年后各单位的人才期望值是多少?

解 用 $\pi_a^k, \pi_b^k, \pi_c^k$ 分别表示第 k 年 A, B, C 三个单位的人才数, 即分布向量为

$$\pi^k = (\pi_a^k, \pi_b^k, \pi_c^k)$$

这是一个具有 3 个状态的齐次 Markov 链, 转移矩阵和初始分布向量分别为

$$T = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}, \quad \pi^0 = (200, 600, 1000)$$

则由定理 10.3.5 得

$$\pi^1 = (800, 600, 400), \quad \pi^2 = (500, 600, 700)$$

并且用归纳法可证明

$$T^m = \begin{pmatrix} t_m & t_{m+1} & t_{m+1} \\ t_{m+1} & t_m & t_{m+1} \\ t_{m+1} & t_{m+1} & t_m \end{pmatrix}, \quad t^m = \frac{1}{3} \left[1 + \frac{(-1)^m}{2^{m-1}} \right]$$

因此

$$\lim_{m \rightarrow \infty} T^m = \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

由此可见

$$\pi = \lim_{m \rightarrow \infty} \pi^m = (600, 600, 600)$$

它是这个 Markov 链惟一的稳态分布向量.

双随机矩阵是一类特殊的随机矩阵, 因而它具有随机矩阵的所有性质, 并且还有如下结果.

定理 10.3.6 设 A 是 n 阶双随机矩阵, 则 $\rho(A) = 1$ 是 A 的特征值, 并且 $e = (1, 1, \dots, 1)^T \in \mathbf{R}^n$ 为 A 与 A^T 对应于特征值 1 的正特征向量.

§ 10.4 M 矩阵

M 矩阵这个术语由 Ostrowski 在 1937 年首先提出, 随后数学家和经济学家进一步发展了 M 矩阵的理论及其应用. 现在 M 矩阵在数值分析、经济系统的投入产出分析、数理经济学中一般均衡的稳定性、运筹学中线性互补问题等领域都具有重要的应用. 本节介绍 M 矩阵的定义及其基本性质.

定义 10.4.1 设 $A \in \mathbf{R}^{n \times n}$, 并且可表示为

$$A = sI - B, \quad s > 0, \quad B \geq 0 \quad (10.4.1)$$

若 $s \geq \rho(B)$, 则称 A 为 M 矩阵; 若 $s > \rho(B)$, 则称 A 为非奇异 M 矩阵.

记

$$Z^{n \times n} = \{A = (a_{ij}) \in \mathbf{R}^{n \times n} \mid a_{ij} \leq 0 (i, j = 1, 2, \dots, n, i \neq j)\}.$$

我们首先给出非奇异 M 矩阵的一些特性.

定理 10.4.1 设 $A \in Z^{n \times n}$ 为非奇异 M 矩阵, 且 $D \in Z^{n \times n}$ 满足 $D \geq A$. 则

- (1) A^{-1} 与 D^{-1} 存在, $A^{-1} \geq D^{-1} \geq 0$;
- (2) D 的每一个实特征值为正数;
- (3) $\det(D) \geq \det(A) > 0$.

证明 因为 A 为非奇异 M 矩阵, 则有

$$A = sI - B, \quad s > \rho(B), \quad B \geq 0$$

对任意实数 $\omega \leq 0$, 考虑矩阵

$$C = A - \omega I = (s - \omega)I - B$$

由于 $s - \omega > \rho(B)$, 故 C 也是非奇异 M 矩阵. 这表明非奇异 M 矩阵的每一个实特征值必是正数. 由于 $D \in Z^{n \times n}$, 故存在足够小的正数 ϵ , 使得

$$P = I - \epsilon D \geq 0$$

由于 $D \geq A$, 则得

$$Q = I - \epsilon A \geq I - \epsilon D = P \geq 0$$

由定理 10.1.6 知 $\rho(Q)$ 为 Q 的非负特征值, 所以

$$\det[(1 - \rho(Q))I - \epsilon A] = \det(Q - \rho(Q)I) = 0$$

由此可知 $\frac{1}{\epsilon}(1 - \rho(Q))$ 为 A 的实特征值. 因为上面已证非奇异 M -矩阵的每一个特征值为正数, 所以 $1 - \rho(Q) > 0$. 于是 $0 \leq \rho(Q) < 1$. 由定理 6.3.9 得

$$(I - Q)^{-1} = (\epsilon A)^{-1} = I + Q + Q^2 + \dots \geq 0$$

故有 $A^{-1} \geq 0$. 又由定理 10.1.2(5)有

$$0 \leq P^k \leq Q^k, \quad k = 1, 2, \dots$$

而由推论 10.1.1 得 $\rho(P) \leq \rho(Q) < 1$, 于是有

$$(I - P)^{-1} = (\epsilon D)^{-1} = I + P + P^2 + \dots \leq (\epsilon A)^{-1}$$

从而得 $A^{-1} \geq D^{-1} \geq 0$, 即(1)得证.

任取 $\alpha \leq 0$, 则 $D - \alpha I \geq A$, 由(1)得 $D - \alpha I$ 非奇异, 因而 D 的所有实特征

值为正数,于是(2)得证.

最后证明(3).由上面的分析,只须证明:若 $A \in Z^{n \times n}$ 的所有实特征值为正数,且 $D \in Z^{n \times n}$ 满足 $D \geq A$,则(3)成立.

事实上,对矩阵的阶数 n 应用归纳法.当 $n=1$ 时,(3)显然成立.假设对 $k(1 \leq k < n)$ 阶矩阵(3)成立.现在考虑 n 阶矩阵的情形.设 A_1 与 D_1 分别是 A 与 D 的 $n-1$ 阶顺序主子矩阵,则 $A_1, D_1 \in Z^{(n-1) \times (n-1)}$,且 $A_1 \leq D_1$.因为矩阵

$$\tilde{A} = \begin{pmatrix} A_1 & 0 \\ 0 & a_{nn} \end{pmatrix} \in Z^{n \times n}$$

满足 $A \leq \tilde{A}$,则由(2)知 \tilde{A} 的所有实特征值为正数,从而 A_1 的所有实特征值亦为正数.根据归纳法假设,有 $\det(D_1) \geq \det(A_1) > 0$.又因为 $A^{-1} \geq D^{-1} \geq 0$,则

$$(A^{-1})_{nn} \geq (D^{-1})_{nn} \geq 0$$

其中 $(A^{-1})_{nn}$ 表示 A^{-1} 的 (n, n) 元素,即

$$(A^{-1})_{nn} = \frac{\det(A_1)}{\det(A)} \geq \frac{\det(D_1)}{\det(D)} = (D^{-1})_{nn} \geq 0$$

因此 $\det(A) > 0, \det(D) > 0$,并利用归纳假设得

$$\det(D) \geq \det(A) \det(D_1) / \det(A_1) \geq \det(A) > 0 \quad \square$$

注意,从定理 10.4.1 的证明可见,若 $A, D \in Z^{n \times n}$,且 $D \geq A$,则 A 为非奇异 M 矩阵蕴含着 D 也是非奇异 M 矩阵,且有 $\det(D) \geq \det(A) > 0$.这个结论在许多实际问题中十分有用.此外,如果定理 10.4.1 中的假设“非奇异 M 矩阵”改换成“ A 的每个实特征值都是正数”,则定理 10.4.1 的结论(1)~(3)仍然成立.

定理 10.4.2 设 $A = (a_{ij}) \in Z^{n \times n}$,则下列诸命题彼此等价:

- (1) A 为非奇异 M-矩阵;
- (2) 若 $B \in Z^{n \times n}$ 且 $B \geq A$,则 B 非奇异;
- (3) A 的任意主子矩阵的每一个实特征值为正数;
- (4) A 的所有主子式为正数;
- (5) 对每个 $k(1 \leq k \leq n)$, A 的所有 k 阶主子式之和为正数;
- (6) A 的每一个实特征值为正数;
- (7) 存在 A 的一种分裂 $A = P - Q$,使得 $P^{-1} \geq 0, Q \geq 0$ 且 $\rho(P^{-1}Q) < 1$;

(8) A 非奇异且 $A^{-1} \geq 0$.

证明 (1) \rightarrow (2) 定理 10.4.1(3) 即得.

(2) \Rightarrow (3) 设 A_k 是 A 的任一 k 阶主子矩阵, K 表示 A_k 在 A 中的行(列)序号集, λ 是 A_k 的任一实特征值. 下面用反证法证明 $\lambda > 0$.

假若 $\lambda \leq 0$, 定义矩阵 $B = (b_{ij}) \in \mathbb{R}^{n \times n}$ 如下:

$$b_{ij} = \begin{cases} a_{ij} - \lambda, & \text{当 } i = j \text{ 时} \\ a_{ij}, & \text{当 } i \neq j \text{ 且 } i, j \in K \text{ 时} \\ 0, & \text{当 } i \neq j \text{ 且 } i, j \notin K \text{ 时} \end{cases}$$

则 $B \geq A$, 并且由(2)知矩阵 B 非奇异. 另一方面, 记 $B_k = A_k - \lambda I$, 因为 λ 是 A_k 的特征值, 则 $\det(B_k) = 0$, 从而 $\det(B) = \det(B_k) \prod_{i \notin K} b_{ii} = 0$. 这与 B 非奇异矛盾, 故 $\lambda > 0$.

(3) \Rightarrow (4) 因为实方阵的复特征值成共轭对出现, 所以实方阵的所有非实特征值的乘积为正数. 由(3)知 A 的任一主子矩阵的实特征值均为正数, 故 A 的任一主子式均为正数.

(4) \Rightarrow (5) 显然成立.

(5) \Rightarrow (6) 由定理 2.4.1 得

$$\det(A - \lambda I) = (-\lambda)^n + b_1(-\lambda)^{n-1} + \cdots + b_n \quad (10.4.2)$$

其中 b_k 是 A 的所有 k 阶主子式之和. 由(5)知 $b_k > 0$ ($k = 1, \dots, n$), 因此(10.4.2)不可能有非正的实根, 即 A 的所有实特征值均为正数.

(6) \Rightarrow (1) 设 $A = sI - B$, $s > 0$ 且 $B \geq 0$, 则 $s - \rho(B)$ 为 A 的实特征值, 由(6)知它是正数, 即 $s > \rho(B)$. 因此 A 为非奇异 M 矩阵.

(1) \Rightarrow (7) 取 $P = sI$, $Q = B$, 并且 s, B 满足

$$A = sI - B, s > \rho(B), B \geq 0$$

则 $P^{-1} \geq 0, Q \geq 0$, 并且 $\rho(P^{-1}Q) = \rho(\frac{1}{s}B) = \frac{1}{s}\rho(B) < 1$.

(7) \Rightarrow (8) 由(7)得 $A = P(I - C)$, 其中 $C = P^{-1}Q$. 因为 $\rho(C) < 1$, 则由定理 6.3.9 有

$$A^{-1} = (I - C)^{-1}P^{-1} = (I + C + C^2 + \cdots)P^{-1}$$

故从 $C = P^{-1}Q \geq 0$, 得 $A^{-1} \geq 0$.

(8) \Rightarrow (1) 记 $A^{-1} = G = (g_{ij})$, 则由(8)知 $G \geq 0$. 从 $AG = I$ 得

$$\sum_{j=1}^n a_{ij}g_{ji} = 1, \quad i = 1, 2, \dots, n$$

因为 $a_{ij} \leq 0, g_{ji} \geq 0 (i \neq j)$, 则

$$a_{ii}g_{ii} = 1 - \sum_{j \neq i}^n a_{ij}g_{ji} \geq 1, \quad i = 1, 2, \dots, n$$

由 $g_{ii} \geq 0$ 及上式得

$$a_{ii} > 0, g_{ii} > 0, \quad i = 1, 2, \dots, n$$

令 $s \geq \max_{1 \leq i \leq n} \{a_{ii}\}$, 则 $B = sI - A \geq 0$. 由定理 10.1.6 知 $\rho(B)$ 是 B 的特征值, 并且有相应的非负特征向量 $x \geq 0$. 于是从 $Bx = \rho(B)x$ 得

$$Ax = (s - \rho(B))x$$

因为 A 可逆, 所以 $s \neq \rho(B)$. 从而

$$A^{-1}x = \frac{1}{s - \rho(B)}x$$

因为 $A^{-1} \geq 0, x \geq 0$ 且 $x \neq 0$, 所以 $s > \rho(B)$. 因此 A 为非奇异 M 矩阵. \square

以上讨论了非奇异 M 矩阵的一些基本性质, 但一般 M 矩阵与非奇异 M 矩阵在应用中几乎同等重要. 下面我们给出一般 M 矩阵的一些特性.

定理 10.4.3 设 $A = (a_{ij}) \in \mathbb{R}^{n \times n}$, 则下列诸命题彼此等价:

- (1) A 是 M 矩阵;
- (2) 对每个 $\epsilon > 0, A + \epsilon I$ 是非奇异 M 矩阵;
- (3) A 的任意主子矩阵的每一个实特征值非负;
- (4) A 的所有主子式非负;
- (5) 对每个 $k (1 \leq k \leq n)$, A 的所有 k 阶主子式之和为非负实数;
- (6) A 的每个实特征值非负.

证明 (1) \Rightarrow (2) 由 M 矩阵定义即得.

(2) \Rightarrow (3) 设 A_k 为 A 的任一 k 阶主子矩阵, λ 为 A_k 的任一实特征值. 若 λ 为负实数, 由 (2) 知 $B = A - \lambda I$ 为非奇异 M 矩阵. 设 B_k 为 B 的 k 阶主子矩阵, 其行、列序号与 A_k 在 A 中的行、列序号相同. 由定理 10.4.2(4) 知 $\det(B_k) > 0$. 因为 λ 为 A_k 的特征值, 所以 $\det(B_k) = \det(A_k - \lambda I) = 0$, 故得出矛盾. 即 (3) 成立.

(3) \Rightarrow (4) 因为 A 的主子式等于相应主子矩阵的所有特征值的乘积, 而非实特征值之积为正数, 实特征值之积非负. 因此 (4) 成立.

(4) \Rightarrow (5) 显然成立.

(5) \Rightarrow (6) 类似于定理 10.4.2 证明中 (5) \Rightarrow (6) 的证法. 因为

$$\det(A - \lambda I) = (-\lambda)^n + b_1(-\lambda)^{n-1} + \dots + b_n \quad (10.4.3)$$

其中 b_k 是 A 的所有 k 阶主子式之和. 由 (5) 知 $b_k \geq 0$ ($k = 1, \dots, n$), 则 (10.4.3) 不可能有负实根, 即 A 的所有实特征值均非负.

(6) \Rightarrow (1) 用类似于定理 10.4.2 中证明 (6) \Rightarrow (1) 的方法即得. \square

习 题

1. 设 A 为 n 阶实矩阵, x, y 是 n 维向量, 证明:

(1) 如果 $A > 0, x \geq 0$ 且 $x \neq 0$, 则 $Ax > 0$;

(2) 如果 $A > 0, x \geq y$, 则 $Ax \geq Ay$;

(3) 如果对所有 $x \geq 0$ 都有 $Ax \geq 0$, 则 $A \geq 0$.

2. 设 $A = \begin{bmatrix} 7 & 2 & 2 \\ 2 & 1 & 1 \\ 4 & 2 & 2 \end{bmatrix}$, 求 $\rho(A)$ 和 $\lim_{n \rightarrow \infty} [\rho(A)^{-1} A]^n$.

3. 设 $A \in \mathbb{R}^{n \times n}, \lambda > 0$ 为 A 的一个单特征值, 且 A 的其他 $n-1$ 个特征值的模都小于 λ . 证明: $\lim_{k \rightarrow \infty} [\rho(A)^{-1} A]^k$ 存在, 并且若 x 是 A 对应于 λ 的任一特征向量, y 是 A^T 对应于 λ 的任一特征向量, 则

$$\lim_{k \rightarrow \infty} [\lambda^{-1} A]^k = (y^T x)^{-1} x y^T.$$

4. 设 $A \in \mathbb{R}^{n \times n}$, 且 $A \geq 0$. 证明: 对任意正向量 $x \in \mathbb{R}^n$, 有

$$\min_{1 \leq i \leq n} \frac{1}{x_i} \sum_{j=1}^n a_{ij} x_j \leq \rho(A) \leq \max_{1 \leq i \leq n} \frac{1}{x_i} \sum_{j=1}^n a_{ij} x_j$$

$$\min_{1 \leq j \leq n} \sum_{i=1}^n \frac{a_{ij}}{x_j} \leq \rho(A) \leq \max_{1 \leq j \leq n} \sum_{i=1}^n \frac{a_{ij}}{x_j}$$

5. 设 $A \in \mathbb{R}^{n \times n}, x \in \mathbb{R}^n$ 且 $A \geq 0$ 和 $x > 0$. 证明: 如果存在 $\alpha \geq 0, \beta \geq 0$ 使得 $\alpha x \leq Ax \leq \beta x$, 则 $\alpha \leq \rho(A) \leq \beta$; 如果 $\alpha x < Ax$, 则 $\alpha < \rho(A)$; 如果 $Ax < \beta x$, 则 $\rho(A) < \beta$.

6. 设 $A \in \mathbb{R}^{n \times n}$, 且 $A \geq 0$. 证明: 如果 λ 为 A 的任一具有正特征向量的特征值, 则 $\lambda = \rho(A)$.

7. 设 $A \in \mathbb{R}^{n \times n}$, 且 $A \geq 0$. 证明: 如果 A 有正特征向量, 则

$$\rho(A) = \max_{x > 0} \min_{1 \leq i \leq n} \frac{1}{x_i} \sum_{j=1}^n a_{ij} x_j = \min_{x > 0} \max_{1 \leq i \leq n} \frac{1}{x_i} \sum_{j=1}^n a_{ij} x_j$$

8. 设 $A \in \mathbb{R}^{n \times n}$, 且 $A \geq 0$. 证明: 如果 A 有正特征向量 x , 则对 $i = 1, 2, \dots, n$ 和 $m = 1, 2, \dots$, 有

$$\sum_{j=1}^n a_{ij}^{(m)} \leq \frac{\max_{1 \leq k \leq n} x_k}{\min_{1 \leq k \leq n} x_k} \rho(A)^m \quad \text{和} \quad \frac{\min_{1 \leq k \leq n} x_k}{\max_{1 \leq k \leq n} x_k} \rho(A)^m \leq \sum_{j=1}^n a_{ij}^{(m)}$$

其中 $A^m = (a_{ij}^{(m)})$. 特别地, 若 $\rho(A) > 0$, 则对 $m = 1, 2, \dots, [\rho(A)^{-1} A]^m$ 的元素均有界.

9. 设 $A \in \mathbb{R}^{n \times n}$, 且 $A \geq 0$. 对 $m = 1, 2, \dots$, 记 $A^m = (a_{ij}^{(m)})$, 证明: 如果 A 有正特征向量 x , 则对 $i = 1, 2, \dots, n$,

$$(1) \text{ 对 } m = 1, 2, \dots, \text{ 有 } \frac{1}{n} \rho(A)^m \frac{\min_{1 \leq k \leq n} x_k}{\max_{1 \leq k \leq n} x_k} \leq \max_{1 \leq j \leq n} [a_{ij}^{(m)}]^2;$$

$$(2) \lim_{m \rightarrow \infty} \left(\sum_{j=1}^n a_{ij}^{(m)} \right)^{\frac{1}{m}} = \rho(A).$$

10. 设 A 是 n 阶非负素矩阵, 证明: $\rho(A) > 0$.

11. 设 A 是 n 阶不可约非负对称矩阵, 证明: A 是素矩阵的充分必要条件是 $A + \rho(A)I$ 非奇异.

12. 设 A 是 n 阶不可约非负矩阵, 证明: 当 $\lambda > \rho(A)$ 时, 有 $(\lambda I - A)^{-1} > 0$.

13. 设 A 是 n 阶不可约非负矩阵, 证明: $\lim_{m \rightarrow \infty} [\rho(A)^{-1} A]^m$ 存在的充分必要条件是 A 为素矩阵.

14. 设 A 是 n 阶不可约非负矩阵, 证明: 如果 A 的所有对角元均为正数, 则 $A^{-1} > 0$.

15. 证明: 如果双随机矩阵 A 是可约的, 则存在排列矩阵 P 使得 $PAP^T =$

$$\begin{bmatrix} 0 & A_{12} \\ A_{21} & 0 \end{bmatrix}.$$

参 考 文 献

- [1] 北京大学数学系几何与代数教研室代数小组. 高等代数. 第二版, 北京: 高等教育出版社, 1988
- [2] 蒋尔雄, 高坤敏, 吴景琨. 线性代数. 北京: 人民教育出版社, 1978
- [3] 吕炯兴. 矩阵论. 北京: 航空工业出版社, 1993
- [4] 孙继广. 矩阵扰动分析. 北京: 科学出版社, 1987
- [5] 陈公宁. 矩阵理论与应用. 北京: 高等教育出版社, 1990
- [6] 许以超. 线性代数与矩阵论. 北京: 高等教育出版社, 1992
- [7] 黄琳. 系统与控制理论中的线性代数. 北京: 科学出版社, 1990
- [8] 史荣昌. 矩阵分析. 北京: 北京理工大学出版社, 1996
- [9] 罗家洪. 矩阵分析引论. 广州: 华南理工大学出版社, 1992
- [10] 周树荃, 戴华. 代数特征值反问题. 郑州: 河南科学技术出版社, 1991
- [11] Dai Hua. On the symmetric solutions of linear matrix equations. *Linear Algebra and Its Applications*, 1990, 131: 1~7
- [12] Lancaster P, Tismenetsky M. *The Theory of Matrices with Applications*. 2nd Edition, Academic Press, 1985
- [13] Horn R A, Johnson C R. *Matrix Analysis*. Cambridge University Press, 1985.
- [14] Bochner S, Martin W T. *Several Complex Variables*. Princeton University Press, 1948
- [15] Fleming W. 多元函数(上册). 庄亚栋译. 北京: 人民教育出版社, 1981

参 考 文 献

- [1] 北京大学数学系几何与代数教研室代数小组. 高等代数. 第二版, 北京: 高等教育出版社, 1988
- [2] 蒋尔雄, 高坤敏, 吴景琨. 线性代数. 北京: 人民教育出版社, 1978
- [3] 吕炯兴. 矩阵论. 北京: 航空工业出版社, 1993
- [4] 孙继广. 矩阵扰动分析. 北京: 科学出版社, 1987
- [5] 陈公宁. 矩阵理论与应用. 北京: 高等教育出版社, 1990
- [6] 许以超. 线性代数与矩阵论. 北京: 高等教育出版社, 1992
- [7] 黄琳. 系统与控制理论中的线性代数. 北京: 科学出版社, 1990
- [8] 史荣昌. 矩阵分析. 北京: 北京理工大学出版社, 1996
- [9] 罗家洪. 矩阵分析引论. 广州: 华南理工大学出版社, 1992
- [10] 周树荃, 戴华. 代数特征值反问题. 郑州: 河南科学技术出版社, 1991
- [11] Dai Hua. On the symmetric solutions of linear matrix equations. *Linear Algebra and Its Applications*, 1990, 131: 1~7
- [12] Lancaster P, Tismenetsky M. *The Theory of Matrices with Applications*. 2nd Edition, Academic Press, 1985
- [13] Horn R A, Johnson C R. *Matrix Analysis*. Cambridge University Press, 1985.
- [14] Bochner S, Martin W T. *Several Complex Variables*. Princeton University Press, 1948
- [15] Fleming W. 多元函数(上册). 庄亚栋译. 北京: 人民教育出版社, 1981