



矩阵论

(第2版)

方保镕 周继东 李医民 编著

清华大学出版社

矩阵论 (第2版)

方保镕 周继东 李医民 编著

清华大学出版社
北京

内 容 简 介

本书比较全面、系统地介绍了矩阵的基本理论、方法及其应用。全书分上、下两篇,上篇为基础篇,下篇为应用篇,共8章,分别介绍了矩阵的几何理论(包括线性空间与线性算子,内积空间与等积变换), λ 矩阵与若尔当标准形,矩阵的分解,赋范线性空间与矩阵范数,矩阵微积分及其应用,广义逆矩阵及其应用,几类特殊矩阵与特殊积(如非负矩阵与正矩阵、循环矩阵与素矩阵、随机矩阵和双随机矩阵、单调矩阵、M矩阵与H矩阵、T矩阵与汉克尔矩阵以及克罗内克积、阿达马积与反积等),前7章每章均配有一定数量的习题。附录中还给出了15套模拟自测试题。所有习题和自测题(约1300题)的详细解答,即将由清华大学出版社另行出版。

本书可作为理工科大学各专业研究生的学位课程教材,也可作为理工科和师范类院校高年级本科生的选修课教材,并可供有关专业的教师和工程技术人员参考。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121993

图书在版编目(CIP)数据

矩阵论/方保镕,周继东,李医民编著. --2版. --北京:清华大学出版社,2013
ISBN 978-7-302-33269-5

I. ①矩… II. ①方… ②周… ③李… III. ①矩阵论 IV. ①O151.21

中国版本图书馆CIP数据核字(2013)第165785号

责任编辑:石磊 赵从棉

封面设计:傅瑞学

责任校对:赵丽敏

责任印制:刘海龙

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦A座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:北京密云胶印厂

经 销:全国新华书店

开 本:185mm×260mm 印 张:25.75 字 数:623千字

版 次:2004年11月第1版 2013年12月第2版 印 次:2013年12月第1次印刷

印 数:1~2500

定 价:45.00元

产品编号:048940-01

随着科学技术的迅速发展,古典的线性代数知识已不能满足现代科技的需要,矩阵的理论和方法业已成为现代科技领域必不可少的工具.诸如数值分析、优化理论、微分方程、概率统计、控制论、力学、电子学、网络等学科领域都与矩阵理论有着密切的联系,甚至在经济管理、金融、保险、社会科学等领域,矩阵理论和方法也有着十分重要的应用.可以毫不夸张地说,矩阵理论的发展极大地推动和丰富了其他众多学科的发展.工程中许多新的理论、方法和技术的诞生与发展就是矩阵理论的创造性应用与推广的结果.当今电子计算机及计算技术的迅速发展更为矩阵理论的应用开辟了更广阔的前景.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.从20世纪80年代,全国的工科院校已普遍把“矩阵论”作为研究生的必修课.为此,1989年我们根据国家教委制定的工科研究生学习“矩阵论”课程的基本要求编写了教材讲义,并于1993年和2004年分别由河海大学出版社和清华大学出版社先后正式出版,在部分高校讲授过多年.为使本书适应时代发展的要求,这次改版又对本书进行了充实更新,并对内容作了精心的处理.

本书内容分上、下两篇,上篇为基础篇,下篇为应用篇,共8章,比较全面、系统地介绍了矩阵的基本理论、方法及其应用.第1章介绍矩阵的几何理论,这部分内容既是线性代数知识的推广和深化,又是矩阵理论的基础,熟练掌握和深刻理解它们对后面内容的学习乃至将来正确处理实际问题有很大的作用.第2章至第4章主要介绍 λ 矩阵与若尔当标准形、矩阵的分解、赋范线性空间与矩阵范数.这些内容是矩阵理论研究、矩阵计算及应用中不可缺少的工具和手段.以上4章内容均为1991年国家教育委员会工科研究生数学课程教学指导小组对“矩阵论”课程所制定的基本要求,故本书把它们放入上篇作为基础篇,约为2~3学分(讲授36~54学时).考虑到矩阵理论的完整性、系统性,又能反映其应用性,同时也为满足某些专业多学时教学的需要,本书的下篇为应用篇,安排有:第5章介绍矩阵微积分及其应用;第6章介绍广义逆矩阵及其应用;第7章介绍几类特殊矩阵与特殊积(诸如非负矩阵与正矩阵、素矩阵与循环矩阵、随机矩阵和双随机矩阵、单调矩阵、M矩阵与H矩阵、T矩阵与汉克尔矩阵,矩阵的克罗内克积、阿达马积与反(Fan)积);第8章专门介绍了矩阵在其他方面的一些应用.本书前7章每章均配有数量一定的习题.附录中还给出了15套模拟自测试题.所有习题和自测题(约1300题)的详细解答,即将由清华大学出版社出版.目录中带*号的内容可用于选学或自学.

本书在编写过程中,力求做到:

1. 理论严谨,重点突出,既重视几何理论,又兼顾应用背景或具体应用;

2. 结构合理,既有系统性,适合全面阅读(多学时),又具有可分性,便于选读(少学时);
3. 取材丰富,涵盖多种矩阵理论与运算法则;
4. 深入浅出,文字流畅。

阅读本书只需具备高等数学和线性代数的基本知识.

作者诚挚地感谢王能超教授,他仔细审阅了全部书稿,并提出了不少有益的建议.

本书可作为理工科大学各专业研究生的学位课程教材,也可作为理工科和师范类院校高年级本科生的选修课教材,并可供有关专业的教师和工程技术人员参考.

由于编著者水平有限,书中如有不妥乃至谬误之处,期望读者批评指正.

编著者

2013年5月

上篇 基础篇

第 1 章 矩阵的几何理论	3
引言 矩阵是什么	3
1.1 线性空间上的线性算子与矩阵	3
1.1.1 线性空间	3
习题 1(1)	18
1.1.2 线性算子及其矩阵	23
习题 1(2)	54
1.2 内积空间上的等积变换	62
1.2.1 内积空间	63
习题 1(3)	73
1.2.2 等积变换及其矩阵	77
习题 1(4)	96
* 1.3 埃尔米特变换及其矩阵	99
1.3.1 对称变换与埃尔米特变换	100
1.3.2 埃尔米特正定、半正定矩阵	102
1.3.3 矩阵不等式	105
1.3.4 埃尔米特矩阵特征值的性质	107
* 1.3.5 一般的复正定矩阵	109
习题 1(5)	110
第 2 章 λ 矩阵与若尔当标准形	113
引言 什么是矩阵标准形	113
2.1 λ 矩阵	113
2.1.1 λ 矩阵的概念	113
2.1.2 λ 矩阵在相抵下的标准形	116
2.1.3 不变因子与初等因子	118
2.2 若尔当标准形	129
2.2.1 数字矩阵化为相似的若尔当标准形	129
* 2.2.2 若尔当标准形的其他求法	140
习题 2	147

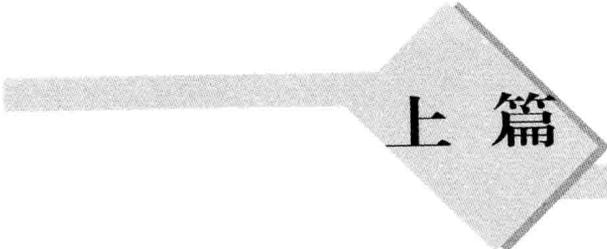
第3章 矩阵的分解	154
引言 矩阵分解的意义	154
3.1 矩阵的三角分解	154
3.1.1 消元过程的矩阵描述	154
3.1.2 矩阵的三角分解	157
3.1.3 常用的三角分解公式	162
3.2 矩阵的QR(正交三角)分解	167
3.2.1 QR分解的概念	167
3.2.2 QR分解的实际求法	170
3.3 矩阵的最大秩分解	176
3.4 矩阵的奇异值分解和极分解	180
3.5 矩阵的谱分解	184
3.5.1 正规矩阵	184
3.5.2 正规矩阵的谱分解	186
3.5.3 单纯矩阵的谱分解	189
习题3	192
第4章 赋范线性空间与矩阵范数	198
引言 范数是什么	198
4.1 赋范线性空间	198
4.1.1 向量的范数	198
4.1.2 向量范数的性质	204
习题4(1)	206
4.2 矩阵的范数	208
4.2.1 矩阵范数的定义与性质	208
4.2.2 算子范数	210
4.2.3 谱范数的性质和谱半径	215
习题4(2)	217
4.3 摄动分析与矩阵的条件数	220
4.3.1 病态方程组与病态矩阵	220
4.3.2 矩阵的条件数	221
*4.3.3 矩阵特征值的摄动分析	224
习题4(3)	228

下篇 应用篇

第5章 矩阵微积分及其应用	233
引言 讨论矩阵微积分的必要性	233
5.1 向量序列和矩阵序列的极限	233

5.1.1	向量序列的极限	233
5.1.2	矩阵序列的极限	235
5.2	矩阵级数与矩阵函数	238
5.2.1	矩阵级数	238
5.2.2	矩阵函数	245
5.3	函数矩阵的微分和积分	254
5.3.1	函数矩阵对实变量的导数	254
5.3.2	函数矩阵特殊的导数	258
5.3.3	矩阵的全微分	262
5.3.4	函数矩阵的积分	264
* 5.4	矩阵微分方程	265
5.4.1	常系数齐次线性微分方程组的解	266
5.4.2	常系数非齐次线性微分方程组的解	270
5.4.3	n 阶常系数微分方程的解	274
	习题 5	277
第 6 章	广义逆矩阵及其应用	286
引言	什么是广义逆矩阵	286
6.1	矩阵的几种广义逆	286
6.1.1	广义逆矩阵的基本概念	286
6.1.2	减号逆 \mathbf{A}^-	287
6.1.3	自反减号逆 \mathbf{A}_r^-	290
6.1.4	最小范数广义逆 \mathbf{A}_m^-	295
6.1.5	最小二乘广义逆 \mathbf{A}_l^-	299
6.1.6	加号逆 \mathbf{A}^+	300
6.2	广义逆在解线性方程组中的应用	306
6.2.1	线性方程组求解问题的提法	306
6.2.2	相容方程组的通解与 \mathbf{A}^-	307
6.2.3	相容方程组的极小范数解与 \mathbf{A}_m^-	309
6.2.4	矛盾方程组的最小二乘解与 \mathbf{A}_l^-	312
6.2.5	线性方程组的极小最小二乘解与 \mathbf{A}^+	317
	习题 6	318
第 7 章	几类特殊矩阵与特殊积	323
引言	什么是特殊矩阵与特殊积	323
7.1	非负矩阵	323
7.1.1	非负矩阵与正矩阵	323
7.1.2	不可约非负矩阵	329
7.1.3	素矩阵与循环矩阵	335

7.2	随机矩阵与双随机矩阵	336
7.3	单调矩阵	340
7.4	M 矩阵与 H 矩阵	341
7.4.1	M 矩阵	342
7.4.2	H 矩阵	346
7.5	T 矩阵与汉克尔矩阵	347
	习题 7(1)	349
7.6	克罗内克积	350
7.6.1	克罗内克积的概念	350
7.6.2	克罗内克积的性质	351
7.7	阿达马积	357
7.8	反积及非负矩阵的阿达马积	359
7.9	克罗内克积应用举例	359
7.9.1	矩阵的拉直	359
7.9.2	线性矩阵方程的解	361
	习题 7(2)	362
第 8 章 矩阵在数学内外的应用		363
	引言	363
8.1	矩阵在数学内部的应用	363
8.1.1	矩阵在代数中的应用	363
8.1.2	矩阵在几何中的应用	366
8.1.3	矩阵在图论中的应用	368
8.2	矩阵在数学之外的应用	372
8.2.1	矩阵在信息编码中的应用	372
8.2.2	矩阵在经济模型中的应用	374
8.2.3	矩阵在生物种群生长繁殖问题中的应用	376
8.2.4	矩阵在控制论中的应用	377
附录 模拟考试自测试题(共 15 套)		384
参考文献		401



上篇

基础篇

第 1 章 矩阵的几何理论

第 2 章 λ 矩阵与若尔当标准形

第 3 章 矩阵的分解

第 4 章 赋范线性空间与矩阵范数

引言 矩阵是什么

简单来说,矩阵就是“由数字纵横排列的一个数学符号”.但是,要想深刻了解矩阵的数学涵义,先要研究矩阵的几何理论.实际上,矩阵是“线性系统中线性算子在基(偶)下的一种数量表示”.因此,对抽象的线性算子的研究,就变成对具体的矩阵的研究.在这里,用得比较多的两个概念便是线性空间与线性算子.线性空间是对集合的元素在线性运算方面所表现的共性加以概括而形成的数学概念,而线性算子则是用来研究线性空间之间关系的主要工具.本章 1.1 节阐述线性空间上的线性算子及其矩阵;1.2 节介绍内积空间上常用的线性算子(正交变换与酉变换)及其矩阵;1.3 节扼要介绍埃尔米特变换及其矩阵.所有论述是在假定读者已经具备线性代数初步知识的基础上进行的,这里所讨论的内容既是已有线性代数知识的深化,也是本书的基础.

1.1 线性空间上的线性算子与矩阵

1.1.1 线性空间

1. 数环与数域

每一个数学概念都有其适用范围,线性空间的概念与在什么范围内取数有直接的关系,为了准确地叙述和理解线性空间这个数学概念,首先引入数域的概念.

定义 1.1.1 设 Z 为非空数集且其中任何两个相同或互异的数之和、差与积仍属于 Z (即数集关于加、减、乘法运算封闭),则称 Z 是一个数环.

只含一个 0 的数集 $Z = \{0\}$ 显然是个数环.

根据数环的定义有:

- (1) 任何数环 Z 必含有 0. 因为若 $a \in Z$, 则 $a - a = 0 \in Z$;
- (2) 若 $a \in Z$, 则 $-a \in Z$. 因为 $0 - a = -a \in Z$.

由此可知, $Z = \{0\}$ 是最小的数环.

定义 1.1.2 如果 P 是至少含有两个互异数的数环,并且其中任何两个数 a 与 b 之商 ($b \neq 0$) 仍属于 P (换言之,数集关于四则运算都封闭),则说 P 是一个数域.

根据数域的定义有:

- (1) 任何数域 P 中必含有 0 与 1, 因为 P 中至少有一个数 $a \neq 0$, 而 $a/a=1 \in P$.
- (2) 若 $a \neq 0$, 则 $1/a = a^{-1} \in P$.

全体整数(包括 0)组成一个数环. 全体有理数组成一个数域, 并且是最小的数域, 因为数中至少含有 0 与 1, 由 0 与 1 通过和、差、积运算形成整数环, 再加上商运算即形成有理数域, 记为 \mathbb{Q} .

全体实数组成一个数域, 叫做实数域, 记为 \mathbb{R} .

全体复数组成一个数域, 叫做复数域, 记为 \mathbb{C} .

读者可以验证, 形如 $a+b\sqrt{2}$ (其中 a, b 为有理数) 的数的全体也构成一个数域, 并且它包含了有理数域.

2. 线性空间

线性空间是线性代数中 n 维向量空间概念的抽象和推广. 为了便于理解这个抽象概念, 我们先回顾 n 维向量空间中的向量在加法及数与向量的乘法方面的运算性质, 然后再把具有同样运算性质的一切集合, 抽象概括为线性空间.

在 n 维向量空间

$$K^n = \{\alpha = (a_1, a_2, \dots, a_n) \mid a_i \in \mathbb{R} \text{ 或 } a_i \in \mathbb{C}, i = 1, 2, \dots, n\}$$

中, 向量 α 是有序数组, 且对向量的加法及数与向量乘法是封闭的(指运算结果都仍是 K^n 中的向量), 且满足如下 8 条性质(设 α, β, γ 都是 n 维向量, λ, μ 是常数):

- (1) $\alpha + \beta = \beta + \alpha$ (加法交换律);
- (2) $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$ (加法结合律);
- (3) $\alpha + \mathbf{0} = \alpha$ (存在零向量 $\mathbf{0}$);
- (4) $\alpha + (-\alpha) = \mathbf{0}$ (存在负向量 $-\alpha$);
- (5) $\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$ (数因子分配律);
- (6) $(\lambda + \mu)\alpha = \lambda\alpha + \mu\alpha$ (分配律);
- (7) $\lambda(\mu\alpha) = (\lambda\mu)\alpha$ (数因子结合律);
- (8) $1\alpha = \alpha$.

值得指出的是, 要研究的集合已远远超出了 n 维向量空间 K^n 的范围, 元素不一定是有序数组, 但集合中元素的加法及数与元素的乘法运算, 却具有 K^n 中相应的性质. 我们先看如下几个熟悉的例子.

例 1.1.1 以实数为系数, 次数不超过 n 的一元多项式的全体(包括 0), 记作

$$P[x]_n = \{a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 \mid a_n, a_{n-1}, \dots, a_1, a_0 \in \mathbb{R}\}.$$

按多项式相加及乘常数的规则, 则 $P[x]_n$ 对这两种运算是封闭的, 因为若 $f(x) \in P[x]_n, g(x) \in P[x]_n$, 则 $f(x) + g(x) \in P[x]_n$; 若 $k \in \mathbb{R}$, 则 $kf(x) \in P[x]_n$, 且易验证对 $P[x]_n$ 的这两种运算, 也有如 K^n 中所述的 8 条性质.

例 1.1.2 常系数二阶齐次线性微分方程

$$y'' - 3y' + 2y = 0$$

的解的集合

$$Y = \{ae^{2x} + be^x \mid a, b \in \mathbb{R}\},$$

对于函数的加法及数与函数乘法两种运算也是封闭的, 因为若 $y_1 = a_1 e^{2x} + b_1 e^x \in Y$, $y_2 = a_2 e^{2x} + b_2 e^x \in Y$, 则 $y_1 + y_2 = (a_1 + a_2) e^{2x} + (b_1 + b_2) e^x \in Y$; 当 $k \in \mathbb{R}$ 时, 则 $ky_1 = ka_1 e^{2x} + kb_1 e^x \in Y$, 且满足如 K^n 中所述的 8 条性质.

例 1.1.3 在所有 n 阶实矩阵的集合 $\mathbb{R}^{n \times n}$ 中, 如果 $A, B \in \mathbb{R}^{n \times n}$, 则 $A + B \in \mathbb{R}^{n \times n}$; 如果 $k \in \mathbb{R}$, 则 $kA \in \mathbb{R}^{n \times n}$. 即集合对于这两种运算是封闭的, 且也都满足如 K^n 中所述的 8 条性质.

此外, 在数学、力学及其他学科中, 还有如例 1.1.1~例 1.1.3 的大量这样的集合. 因此, 有必要不考虑集合的具体内容的涵义来研究这类集合的公共性质, 并把这类集合概括成一个数学名词, 于是就有如下的线性空间的概念.

定义 1.1.3 设 V 是一个非空集合, P 是一个数域. 如果 V 满足如下两个条件:

1. 在 V 中定义一个封闭的加法运算, 即当 $x, y \in V$ 时, 有唯一的和 $x + y \in V$, 并且加法运算满足 4 条性质:

$$(1) x + y = y + x \quad (\text{交换律});$$

$$(2) x + (y + z) = (x + y) + z \quad (\text{结合律});$$

$$(3) \text{存在零元素 } 0 \in V, \text{ 对于 } V \text{ 中任何一个元素 } x \text{ 都有 } x + 0 = x;$$

(4) 存在负元素, 即对任一元素 $x \in V$, 存在有一元素 $y \in V$, 使 $x + y = 0$, 且称 y 为 x 的负元素, 记为 $-x$, 于是有 $x + (-x) = 0$.

2. 在 V 中定义一个封闭的数乘运算(数与元素的乘法), 即当 $x \in V, \lambda \in P$ 时, 有唯一的 $\lambda x \in V$, 且数乘运算满足 4 条性质:

$$(5) (\lambda + \mu)x = \lambda x + \mu x \quad (\text{分配律});$$

$$(6) \lambda(x + y) = \lambda x + \lambda y \quad (\text{数因子分配律});$$

$$(7) \lambda(\mu x) = (\lambda\mu)x \quad (\text{结合律});$$

$$(8) 1x = x.$$

其中 x, y, z 表示 V 中的任意元素; λ, μ 是数域 P 中任意数; 1 是数域 P 中的单位数.

这时, 我们说 V 是数域 P 上的线性空间. 不管 V 的元素如何, 当 P 为实数域 \mathbb{R} 时, 则称 V 为实线性空间; 当 P 为复数域 \mathbb{C} 时, 就称 V 为复线性空间.

通常我们把 V 中满足 8 条性质且为封闭的加法及数乘两种运算, 统称线性运算. 简言之, 凡定义了线性运算的集合, 就称线性空间. 因此, 线性运算是线性空间的本质, 它反映了集合中元素之间的某种代数结构. 当仅研究集合的代数结构时, 便抽象出线性空间的概念.

下面列举一些线性空间的例子.

在 K^n 中, 所有实 n 维向量的集合 \mathbb{R}^n 是实数域 \mathbb{R} 上的线性空间; 所有复 n 维向量的集合 \mathbb{C}^n 是复数域 \mathbb{C} 上的线性空间. 作为特例, 几何空间全体向量组成的集合 \mathbb{R}^2 或 \mathbb{R}^3 是实数域 \mathbb{R} 上的线性空间.

例 1.1.1~例 1.1.3 中的集合, 在其各自的加法及数乘运算的定义下, 都构成实数域 \mathbb{R} 上的线性空间. 我们称例 1.1.3 所给的线性空间 $\mathbb{R}^{n \times n}$ 为矩阵空间.

此外, 检验集合是否构成线性空间, 逐条检验运算是否为线性运算是至关重要的. 例如, 次数等于 $n (n \geq 1)$ 的多项式的集合, 关于通常的多项式加法与数乘运算是不能构成线性空间的. 因为两个 n 次多项式的和可能不是 n 次多项式, 如当 $n > 1$ 时, $f(x) = x^n + x$, $g(x) = -x^n + 1$, 则 $f(x) + g(x) = x + 1$ 就不属于原来的集合, 亦即对加法运算不封闭, 故

不是线性空间. 还要注意, 检验线性运算不能只检验对运算的封闭性, 特别是当定义的加法及数乘运算不是通常意义下的运算时, 则应仔细检验其余 8 条性质.

下面再举一个不是线性空间的例子.

例 1.1.4 平面上全体向量组成的集合, 对于通常意义下的向量加法和如下定义的数乘

$$k \cdot \alpha = \mathbf{0},$$

虽然对两种运算都封闭, 但因 $1 \cdot \alpha = \mathbf{0}$, 不满足运算性质(8), 即定义的运算不是线性运算, 所以不是线性空间.

一般来说, 同一个集合, 若定义两种不同的线性运算, 就构成不同的线性空间; 若定义的运算不是线性运算, 也就不能构成线性空间. 所以, 线性空间的概念是集合与运算二者的结合. 为了对线性空间的理解更具有一般性, 请看下面的线性空间所表现的代数结构.

例 1.1.5 设 \mathbb{R}^+ 为所有正实数组成的数集, 其加法及数乘运算定义为(奇怪的加法与数乘)

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

证明 \mathbb{R}^+ 是 \mathbb{R} 上的线性空间.

证明 实际上要验证 10 条:

对加法封闭: 设 $a, b \in \mathbb{R}^+$, 则有 $a \oplus b = ab \in \mathbb{R}^+$;

对数乘封闭: 设 $k \in \mathbb{R}, a \in \mathbb{R}^+$, 则有 $k \circ a = a^k \in \mathbb{R}^+$;

1. $a \oplus b = ab = ba = b \oplus a$;
2. $(a \oplus b) \oplus c = (ab) \oplus c = (ab)c = a(bc) = a \oplus (b \oplus c)$;
3. 1 是零元素, 因为 $a \oplus 1 = a \cdot 1 = a$;
4. a 的负元素是 $1/a$, 因为 $a \oplus 1/a = a \cdot 1/a = 1$;
5. $k \circ (a \oplus b) = k \circ (ab) = (ab)^k = a^k b^k = (k \circ a) \oplus (k \circ b)$;
6. $(\lambda + \mu) \circ a = a^{\lambda + \mu} = a^\lambda \oplus a^\mu = (\lambda \circ a) \oplus (\mu \circ a)$;
7. $\lambda \circ (\mu \circ a) = \lambda \circ a^\mu = (a^\mu)^\lambda = a^{\lambda \mu} = (\lambda \mu) \circ a$;
8. $1 \circ a = a^1 = a$.

因此, \mathbb{R}^+ 是实线性空间.

线性空间还是物理、力学中满足叠加原理的系统的数学模型, 请看下例.

例 1.1.6 考察一根梁因受荷载而产生变形的问題(见图 1.1). 设所考察的情况都在弹性范围之内(我们不妨设应变与应力总是成比例的, 而在实际应用时结论只适用于小应变的情况).

设支点没有位移, 则挠度曲线是在区间 $[-a, a]$ 上的连续函数, 且有

$$f(-a) = 0, f(0) = 0, f(a) = 0.$$

考察所有这样的挠度曲线(函数形式)的集合:

$$D = \{f \mid f \in C^2[-a, a], f(-a) = 0, f(0) = 0, f(a) = 0\},$$

这里 $C^2[-a, a]$ 表示所有在区间 $[-a, a]$ 上有二阶连续导数的函数的集合.

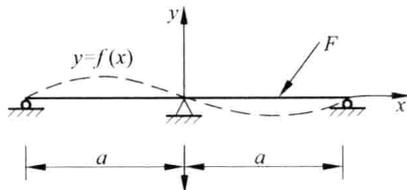


图 1.1 梁的挠度曲线

设荷载 F_1 产生挠度曲线 $y=f_1(x)$, F_2 产生挠度曲线 $y=f_2(x)$, 由叠加原理知: F_1 与 F_2 同时作用, 则产生的挠度曲线为 $y=f_1(x)+f_2(x)$, 这正是函数的加法.

这种加法显然是封闭的, 且满足线性空间定义前 4 条性质 (0 就是恒等于零的函数——对应于零荷载; 与 f 相反的元素就是 $-f$, 它对应于反方向的荷载).

如果荷载 F_1 是 F_2 的 k 倍, 即 $F_1=kF_2$, 则

$$f_1(x) = kf_2(x), \quad \text{即} \quad f_1 = kf_2.$$

这正是常数与函数的乘法, 它也是封闭的, 且满足线性空间定义后 4 条性质, 所以挠度曲线的集合 D 构成一个线性空间.

总之, 在不同的领域都可以举出许多线性空间的例子. 正因为线性空间是 n 维向量空间的抽象和推广, 所以为了几何直观, 有时我们又把线性空间叫做向量空间. 但这里的向量不一定是有序数组, 而是广义的向量, 可以是以数学对象 (如函数、矩阵等) 为向量, 也可以是以物理对象 (如力、速度等) 为向量.

3. 线性空间的基本性质

根据线性空间的定义, 可以推证线性空间的下述性质:

性质 1 线性空间的零元素是惟一的.

事实上, 如果 $\mathbf{0}_1, \mathbf{0}_2$ 是线性空间 V 中的两个零元素, 则根据定义有

$$\mathbf{0}_1 = \mathbf{0}_1 + \mathbf{0}_2 = \mathbf{0}_2 + \mathbf{0}_1 = \mathbf{0}_2.$$

性质 2 任一元素的负元素是惟一的.

事实上, 设 x_1, x_2 均为 $x \in V$ 的负元素, 则

$$x_1 = x_1 + (x + x_2) = (x_1 + x) + x_2 = x_2.$$

性质 3 设 $\lambda, 0, -1, 1 \in P, x, -x, \mathbf{0} \in V$, 则

$$(1) 0x = \mathbf{0};$$

$$(2) (-1)x = -x;$$

$$(3) \lambda \mathbf{0} = \mathbf{0};$$

$$(4) \text{若 } \lambda x = \mathbf{0}, \text{ 则 } \lambda = 0 \text{ 或 } x = \mathbf{0}.$$

事实上, 因为

$$x + 0x = 1x + 0x = (1+0)x = 1x = x,$$

所以

$$0x = \mathbf{0};$$

$$\text{因为} \quad x + (-1)x = 1x + (-1)x = [1+(-1)]x = 0x = \mathbf{0},$$

所以

$$(-1)x = -x;$$

$$\text{因为} \quad \lambda \mathbf{0} = \lambda[x + (-1)x] = \lambda x + (-\lambda)x = [\lambda + (-\lambda)]x = 0x = \mathbf{0},$$

故

$$\lambda \mathbf{0} = \mathbf{0};$$

最后, 若 $\lambda \neq 0$, 且 $x \neq \mathbf{0}$, 则

$$x = 1x = \left(\frac{1}{\lambda}\lambda\right)x = \frac{1}{\lambda}(\lambda x) = \frac{1}{\lambda}\mathbf{0} = \mathbf{0},$$

这与 $x \neq \mathbf{0}$ 矛盾, 故 $\lambda \neq 0$ 与 $x \neq \mathbf{0}$ 不能同时成立.

定义 1.1.4 只含一个元素的线性空间叫做零空间, 显然, 这个元素便是零元素.

4. 基、维数与坐标

由上面线性空间的定义容易知道,有限个向量组成的集合,总不能满足加法及数乘运算的封闭性,所以除只由一个零向量构成的零空间 $\{0\}$ 外,一般线性空间都有无穷多个向量.于是提出两个问题:

(1) 在无穷多个向量中能否找到有限个具有代表性的向量,使得线性空间中任一个向量都可以用这有限个向量来表示?

(2) 线性空间的向量是抽象的,如何把它与具体的数组向量 (a_1, a_2, \dots, a_n) 联系起来,使线性空间中抽象的线性运算转化为数组向量的线性运算?

为了圆满地解答这两个问题,首先需要定义线性空间中向量组的线性相关性等基本概念.

如果 $x_1, x_2, \dots, x_r (r \geq 1)$ 为线性空间 V 中一组向量, k_1, k_2, \dots, k_r 是数域 P 中的数,那么向量

$$x = k_1 x_1 + k_2 x_2 + \dots + k_r x_r \quad (1.1.1)$$

称为向量 x_1, x_2, \dots, x_r 的一个**线性组合**,有时也说向量 x 可用向量组 x_1, x_2, \dots, x_r **线性表示**.

例 1.1.7 对于例 1.1.2 中的线性空间 Y ,我们知道它的向量是微分方程 $y'' - 3y' + 2y = 0$ 的解 $y = c_1 e^x + c_2 e^{2x}$,其中 c_1 与 c_2 是独立的两个任意常数.这表明 y 是它的两个特解向量 e^x 与 e^{2x} 的线性组合.

如果式(1.1.1)中的 k_1, k_2, \dots, k_r 不全为零,且使

$$k_1 x_1 + k_2 x_2 + \dots + k_r x_r = 0, \quad (1.1.2)$$

则称向量组 x_1, x_2, \dots, x_r **线性相关**,否则就称其为**线性无关**.换句话说,如果等式(1.1.2)只有在 $k_1 = k_2 = \dots = k_r = 0$ 时才成立,则称 x_1, x_2, \dots, x_r **线性无关**.

显然,如果 x_1, x_2, \dots, x_r 中有一个为零元,则这 r 个元素必然是线性相关的,例如 $x_1 = 0$,则可置 $k_1 \neq 0$,而令 $k_2 = k_3 = \dots = k_r = 0$,使式(1.1.2)成立.

例 1.1.8 在 \mathbb{R}^n 中,设有两个向量组

$$\begin{cases} \varepsilon_1 = (1, 0, \dots, 0), \\ \varepsilon_2 = (0, 1, \dots, 0), \\ \vdots \\ \varepsilon_n = (0, 0, \dots, 1), \end{cases} \quad \text{及} \quad \begin{cases} \varepsilon'_1 = (1, 1, \dots, 1, 1), \\ \varepsilon'_2 = (0, 1, \dots, 1, 1), \\ \vdots \\ \varepsilon'_n = (0, 0, \dots, 0, 1). \end{cases}$$

分别考察方程组

$$k_1 \varepsilon_1 + k_2 \varepsilon_2 + \dots + k_n \varepsilon_n = 0 \quad \text{及} \quad k_1 \varepsilon'_1 + k_2 \varepsilon'_2 + \dots + k_n \varepsilon'_n = 0.$$

由于它们对应的系数行列式

$$\begin{vmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{vmatrix} \neq 0 \quad \text{及} \quad \begin{vmatrix} 1 & 1 & \cdots & 1 & 1 \\ 0 & 1 & \cdots & 1 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{vmatrix} \neq 0,$$

从而两个线性方程组都只有零解 $k_1 = k_2 = \dots = k_n = 0$,故这两个向量组是线性无关的.

又如,例 1.1.1 中的线性空间 $P[x]_n$,取向量组 $1, x, x^2, \dots, x^n$,当对任意 x ,使 $k_1 \cdot 1 + k_2 x + k_3 x^2 + \dots + k_{n+1} x^n = 0$ 时, k_1, k_2, \dots, k_{n+1} 必须同时为零,因此, $1, x, x^2, \dots, x^n$ 线性无

关; 同理, 例 1.1.2 中的向量组 e^x, e^{2x} , 当 $c_1 e^x + c_2 e^{2x} = 0$ 时, c_1 与 c_2 必须同时为零, 所以 e^x 与 e^{2x} 也线性无关.

与 \mathbb{R}^n 类似, 在线性空间 V 中下列命题成立.

命题 1.1.1 当 $r \geq 2$ 时, V 中的向量组 x_1, x_2, \dots, x_r 线性相关的充要条件是其中至少有一个向量可由向量组中其余向量线性表示; 而线性无关的充要条件则是其中每一个向量都不能由向量组中其余向量来线性表示.

命题 1.1.2 若 V 中向量组的某一子向量组线性相关, 则该向量组也线性相关.

命题 1.1.3 若 V 中某向量组线性无关, 则其任一子向量组也线性无关.

定义 1.1.5 设 V 是数域 P 上的线性空间, $x_1, x_2, \dots, x_n (n \geq 1)$ 是属于 V 的任意 n 个向量, 如果它满足:

- (1) x_1, x_2, \dots, x_n 线性无关;
- (2) V 中任一向量 x 均可由 x_1, x_2, \dots, x_n 来线性表示;

则称 x_1, x_2, \dots, x_n 是 V 的一组基(或基底), 并称 x_1, x_2, \dots, x_n 为基向量.

线性空间 V 的基向量所含向量的个数 n , 称为线性空间 V 的维数, 记为 $\dim V = n$, 并称 V 为 n 维线性空间, 可简记为 V^n .

由定义 1.1.5 可见, 维数实际上就是 V 中线性无关向量组中向量的最大个数, 而基只不过是 V 中的最大线性无关组而已.

特别地, 零空间的维数是 0.

若在 V 中可以找到任意多个线性无关的向量, 则称 V 是无限维线性空间.

按照上述定义, 不难看出, 例 1.1.1 中的线性空间 $P[x]_n$ 是 $n+1$ 维的, 因为前面已验证向量组 $1, x, x^2, \dots, x^n$ 线性无关, 且 $P[x]_n$ 中任一向量都可由这组向量线性表示, 即

$$f(x) = a_0 \cdot 1 + a_1 x + a_2 x^2 + \dots + a_n x^n.$$

因此, $1, x, x^2, \dots, x^n$ 是 $P[x]_n$ 中的一组基. 同样, 例 1.1.2 中的线性空间 Y 是二维的, e^x, e^{2x} 是 Y 中的一组基; 而例 1.1.3 中的矩阵空间 $\mathbb{R}^{n \times n}$ 是 n^2 维的线性空间, 这是因为 $\mathbb{R}^{n \times n}$ 的任

一向量 $A = \sum_{i,j=1}^n k_{ij} E_{ij}$, 其中 E_{ij} 是第 i 行第 j 列的元素为 1、其余为 0 的 n 阶方阵, 且易验证 $E_{ij} (i, j = 1, 2, \dots, n)$ 线性无关, 故 $\dim \mathbb{R}^{n \times n} = n^2$, $E_{ij} (i, j = 1, 2, \dots, n)$ 就是 $\mathbb{R}^{n \times n}$ 的一组基.

读者不难验证, 齐次线性方程组 $Ax = 0$ 的基础解系就是其解空间的一组基.

注意, 一个线性空间的基不是惟一的. 例如, 在例 1.1.8 中已验证 \mathbb{R}^n 中向量组 $\epsilon_1 = (1, 0, \dots, 0), \epsilon_2 = (0, 1, \dots, 0), \dots, \epsilon_n = (0, 0, \dots, 1)$ 及 $\epsilon'_1 = (1, 1, \dots, 1, 1), \epsilon'_2 = (0, 1, \dots, 1, 1), \dots, \epsilon'_n = (0, 0, \dots, 0, 1)$ 都是线性无关的, 而且对任一向量 $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, 分别有

$$x = x_1 \epsilon_1 + x_2 \epsilon_2 + \dots + x_n \epsilon_n,$$

$$x = x_1 \epsilon'_1 + (x_2 - x_1) \epsilon'_2 + \dots + (x_n - x_{n-1}) \epsilon'_n,$$

所以它们都是 \mathbb{R}^n 的基, 并且称 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 是 \mathbb{R}^n 的自然基.

可以证明(见习题 1(1)第 18 题), 线性空间里不同基所含向量的个数是相等的, 即线性空间的维数是确定的, 不会因选取的基不同而改变.

另外, 还必须指出, 线性空间的基、维数与所考虑的数域有关.

例 1.1.9 如果把复数域 \mathbb{C} 看作是自身上的线性空间, 数 1 就是一组基, 那么它是一维的; 如果把 \mathbb{C} 看作是实数域 \mathbb{R} 上的线性空间, 数 $1, i$ 就是一组基, 那么它是二维的.

关于无限维线性空间的例子也是有的. 如所有实系数多项式的集合(幂级数为向量)

$$P[x]_{\infty} = \{a_0 + a_1x + a_2x^2 + \cdots + a_nx^n + \cdots \mid a_i \in \mathbb{R}\}$$

在通常的多项式加法及数乘多项式的运算下形成的实线性空间, 就是无限维的, 因为对任意整数 N , 都有 N 个线性无关的向量 $1, x, x^2, \cdots, x^{N-1}$.

又如在 $[a, b]$ 上所有实连续函数的集合 $C[a, b]$, 它也是无限维的线性空间, 因为找不到有限个连续函数作为它的基.

关于无限维线性空间的讨论, 已超出本门课程的范围, 它属于“泛函分析”课程研究的对象, 本书主要讨论有限维线性空间.

定理 1.1.1 设 x_1, x_2, \cdots, x_n 是 V^n 的一组基, 对于任何向量 $x \in V^n$, 则它可惟一地由 x_1, x_2, \cdots, x_n 线性表示.

证明 设 x 由 x_1, x_2, \cdots, x_n 线性表示的式子有两个, 即

$$x = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n,$$

$$x = \mu_1 x_1 + \mu_2 x_2 + \cdots + \mu_n x_n,$$

以上两式相减得

$$(\lambda_1 - \mu_1)x_1 + (\lambda_2 - \mu_2)x_2 + \cdots + (\lambda_n - \mu_n)x_n = \mathbf{0}.$$

由 x_1, x_2, \cdots, x_n 的线性无关性, 知 $\lambda_1 = \mu_1, \lambda_2 = \mu_2, \cdots, \lambda_n = \mu_n$.

证毕

这样, 线性空间 V^n 可表示为

$$V^n = \{x = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n \mid \lambda_i \in \mathbb{R}, i = 1, 2, \cdots, n\},$$

这就较清楚地显示出线性空间中向量的构造, 即 V^n 中的基 x_1, x_2, \cdots, x_n 的确起到了“代表”的作用, 从而回答了前面提到的第一个问题. 现在, 我们来考虑第二个问题.

反之, 任给一组有序数 $\lambda_1, \lambda_2, \cdots, \lambda_n$, 总有惟一的向量

$$x = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n \in V^n,$$

这样, V^n 中的向量 x 与有序数组 $(\lambda_1, \lambda_2, \cdots, \lambda_n)$ 之间存在着一种一一对应的关系. 因此可以用这组有序数来表示向量 x . 于是我们有以下定义.

定义 1.1.6 设 x_1, x_2, \cdots, x_n 是线性空间 V^n 的一组基, 对于任一向量 $x \in V^n$, 总有且仅有一组有序数 a_1, a_2, \cdots, a_n 使

$$x = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n.$$

a_1, a_2, \cdots, a_n 这组有序数就称为向量 x 在基 x_1, x_2, \cdots, x_n 下的坐标, 并记作

$$\mathbf{X} = (a_1, a_2, \cdots, a_n) \quad \text{或} \quad \mathbf{X} = (a_1, a_2, \cdots, a_n)^{\top}.$$

例 1.1.10 若 x_1, x_2, \cdots, x_n 为 V^n 的一组基, 则每个基向量恰好可表示为

$$x_i = 0x_1 + \cdots + 0x_{i-1} + 1x_i + 0x_{i+1} + \cdots + 0x_n,$$

即相应的坐标为

$$\mathbf{X}_i = (\underbrace{0, \cdots, 0}_{i-1 \text{ 个}}, 1, 0, \cdots, 0), \quad i = 1, 2, \cdots, n.$$

由此还可见, 假如在定义 1.1.3 中没有第(8)条 $1x = x$ 的运算性质, 我们就无法把 x_i 写成基 x_1, x_2, \cdots, x_n 的线性组合, 从而基、维数、坐标等概念也就无从谈起了.

需要指出, 同一向量 x 在不同的基(或称坐标系)下的坐标往往是不同的.

例 1.1.11 在线性空间 $P[x]_n$ 中, 多项式

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

在基 $1, x, x^2, \dots, x^n$ 下的坐标就是它的系数构成的行向量 (a_0, a_1, \dots, a_n) . 若在 V^n 中取另一组基

$$1, (x-a), (x-a)^2, \dots, (x-a)^n,$$

则多项式 $f(x)$ 按泰勒(Taylor)公式展开为

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

因此, $f(x)$ 在新的一组基下的坐标是

$$\left(f(a), f'(a), \frac{f''(a)}{2!}, \dots, \frac{f^{(n)}(a)}{n!} \right).$$

又如, \mathbb{R}^n 的任一向量 (x_1, x_2, \dots, x_n) 在自然基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 下的坐标是 (x_1, x_2, \dots, x_n) , 但在例 1.1.8 中第二组基 $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$ 下的坐标却是 $(x_1, x_2 - x_1, \dots, x_n - x_{n-1})$.

建立了坐标以后, 就把抽象的向量 x 与具体的数组向量 (x_1, x_2, \dots, x_n) 联系起来, 并且还可以把 V^n 中抽象的线性运算与数组向量的线性运算联系起来.

设 $x, y \in V^n$, 有 $x = x_1x_1 + x_2x_2 + \dots + x_nx_n$, $y = y_1x_1 + y_2x_2 + \dots + y_nx_n$, 于是

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

$$\lambda x = (\lambda x_1)x_1 + (\lambda x_2)x_2 + \dots + (\lambda x_n)x_n,$$

即 $x + y$ 的坐标是 $(x_1 + y_1, x_2 + y_2, \dots, x_n + y_n) = (x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n)$, λx 的坐标是 $(\lambda x_1, \lambda x_2, \dots, \lambda x_n) = \lambda(x_1, x_2, \dots, x_n)$.

在 V^n 中, 任意 n 个线性无关的向量都可取作它的基或坐标系, 可以说 V^n 的基或坐标系有无穷多组, 但对不同的基或坐标系, 同一个向量的坐标一般是不同的. 那么, 不同的基与不同的坐标之间有怎样的关系呢? 为此, 先介绍由一组基改变为另一组基时的过渡矩阵的概念.

设 e_1, e_2, \dots, e_n 及 e'_1, e'_2, \dots, e'_n 是 V^n 中的两组基, 且

$$\begin{cases} e'_1 = c_{11}e_1 + c_{21}e_2 + \dots + c_{n1}e_n, \\ e'_2 = c_{12}e_1 + c_{22}e_2 + \dots + c_{n2}e_n, \\ \vdots \\ e'_n = c_{1n}e_1 + c_{2n}e_2 + \dots + c_{nn}e_n, \end{cases}$$

或者写成矩阵形式

$$(e'_1, e'_2, \dots, e'_n) = (e_1, e_2, \dots, e_n)C, \quad (1.1.3)$$

其中矩阵

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

称为由基 e_1, e_2, \dots, e_n 变到基 e'_1, e'_2, \dots, e'_n 的过渡矩阵. 由于 e'_1, e'_2, \dots, e'_n 线性无关, 故过渡矩阵 C 可逆, 即 C^{-1} 存在.

例如, 对于例 1.1.8 中给出的两组基, 易知有

$$(e'_1, e'_2, \dots, e'_n) = (e_1, e_2, \dots, e_n) \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix},$$

则由基 $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_n$ 变到基 $\boldsymbol{\varepsilon}'_1, \boldsymbol{\varepsilon}'_2, \dots, \boldsymbol{\varepsilon}'_n$ 的过渡矩阵

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}.$$

上面由过渡矩阵给出了两组基之间的变换关系,下面再讨论向量的坐标变换问题.

设 $\boldsymbol{x} \in V^n$, 且 \boldsymbol{x} 在两组基下的坐标分别为 (x_1, x_2, \dots, x_n) 及 $(x'_1, x'_2, \dots, x'_n)$, 即

$$\boldsymbol{x} = x_1 \boldsymbol{e}_1 + x_2 \boldsymbol{e}_2 + \cdots + x_n \boldsymbol{e}_n = x'_1 \boldsymbol{e}'_1 + x'_2 \boldsymbol{e}'_2 + \cdots + x'_n \boldsymbol{e}'_n,$$

写成矩阵形式为

$$\boldsymbol{x} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = (\boldsymbol{e}'_1, \boldsymbol{e}'_2, \dots, \boldsymbol{e}'_n) \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix}. \quad (1.1.4)$$

将式(1.1.3)代入式(1.1.4),得

$$(\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n) \mathbf{C} \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix}.$$

由于 $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n$ 线性无关且为基,故有

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \mathbf{C} \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix}. \quad (1.1.5)$$

由于 \mathbf{C} 可逆,故

$$\begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix} = \mathbf{C}^{-1} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}. \quad (1.1.6)$$

式(1.1.5)或式(1.1.6)称为在基变换式(1.1.3)下向量的坐标变换公式.

例 1.1.12 在 \mathbb{R}^4 中,已知向量 \boldsymbol{x} 在基 $\boldsymbol{e}_1 = (1, 2, -1, 0)$, $\boldsymbol{e}_2 = (1, -1, 1, 1)$, $\boldsymbol{e}_3 = (-1, 2, 1, 1)$, $\boldsymbol{e}_4 = (-1, -1, 0, 1)$ 下的坐标为 (x_1, x_2, x_3, x_4) , 求当基改变为基 $\boldsymbol{e}'_1 = (2, 1, 0, 1)$, $\boldsymbol{e}'_2 = (0, 1, 2, 2)$, $\boldsymbol{e}'_3 = (-2, 1, 1, 2)$, $\boldsymbol{e}'_4 = (1, 3, 1, 2)$ 时, 向量 \boldsymbol{x} 在新基下的坐标.

解 引入自然基来表示基 $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4$, 即

$$(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4) = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3, \boldsymbol{\varepsilon}_4) \begin{pmatrix} 1 & 1 & -1 & -1 \\ 2 & -1 & 2 & -1 \\ -1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix} = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3, \boldsymbol{\varepsilon}_4) \mathbf{A}.$$

于是

$$(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3, \boldsymbol{\varepsilon}_4) = (\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4) \mathbf{A}^{-1}.$$

同样,基 e'_1, e'_2, e'_3, e'_4 也可用自然基表示,即

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

所以有

$$(e'_1, e'_2, e'_3, e'_4) = (e_1, e_2, e_3, e_4) \mathbf{A}^{-1} \mathbf{B}.$$

于是得过渡矩阵

$$\mathbf{C} = \mathbf{A}^{-1} \mathbf{B}.$$

再由坐标变换公式(1.1.6)可得

$$(x'_1, x'_2, x'_3, x'_4) = (x_1, x_2, x_3, x_4) (\mathbf{C}^{-1})^T.$$

由计算可得

$$(\mathbf{C}^{-1})^T = (\mathbf{C}^T)^{-1} = (\mathbf{B}^T (\mathbf{A}^T)^{-1})^{-1} = \mathbf{A}^T (\mathbf{B}^T)^{-1} = \begin{pmatrix} 0 & -1 & 0 & 1 \\ 1 & 1 & 0 & -1 \\ -1 & 0 & 0 & 1 \\ 1 & 0 & 1 & -1 \end{pmatrix},$$

因此有

$$(x'_1, x'_2, x'_3, x'_4) = (x_1, x_2, x_3, x_4) \begin{pmatrix} 0 & -1 & 0 & 1 \\ 1 & 1 & 0 & -1 \\ -1 & 0 & 0 & 1 \\ 1 & 0 & 1 & -1 \end{pmatrix},$$

即所求的坐标变换公式为

$$\begin{aligned} x'_1 &= x_2 - x_3 + x_4, & x'_2 &= -x_1 + x_2, \\ x'_3 &= x_4, & x'_4 &= x_1 - x_2 + x_3 - x_4. \end{aligned}$$

5. 线性子空间

1) 子空间的概念

大家知道,过坐标原点的一条直线或一个平面都是三维几何空间的子集,而且它们关于向量加法及数乘运算分别构成一个一维和二维的线性空间.这种现象,也可以推广到 n 维线性空间,即线性空间的子集,它在原线性空间的加法及数乘运算下,仍构成线性空间.这一特性,对深入认识线性空间的性质、线性空间的分解与构造都是必不可少的,为此,我们引入下面的定义.

定义 1.1.7 设 V_1 是数域 P 上线性空间 V 的一个子集,且这个子集对 V 已有的加法及数乘运算也构成线性空间,则称 V_1 为 V 的线性子空间,简称子空间,记为 $V_1 \subseteq V$,当 $V_1 \neq V$ 时,记为 $V_1 \subset V$.

例 1.1.13 设 $C[a, b]$ 是定义在 $[a, b]$ 上的连续函数的全体组成的集合,它是一个线性空间,而 V_1 是定义在同一区间上不超过 n 次的实系数多项式的全体,显然 V_1 是 $C[a, b]$ 的一个子空间.

又如, n 元齐次线性方程组的解空间就是 \mathbb{R}^n 的子空间.

需要指出,验证一个非空子集是否为线性空间,只需验证这个子集对原线性空间已有的加法及数乘两种运算的封闭性就行了,而其余 8 条运算性质它自然满足. 我们有如下的定理.

定理 1.1.2 设 V_1 是线性空间 V 的一个非空子集,则 V_1 是 V 的一个子空间的充分必要条件为

- (1) 如果 $x, y \in V_1$, 则 $x + y \in V_1$;
- (2) 如果 $x \in V_1, k \in P$, 则 $kx \in V_1$.

定理的结论是明显的. 因为根据上面的两条封闭性,容易推出存在负元素与存在零元素这两条性质,即如果 $x \in V_1$, 则 $-x = (-1)x \in V_1$, 且 $x + (-x) = \mathbf{0} \in V_1$; 再者,因为 V_1 的元素自然也是 V 的元素,从而 V_1 的元素承袭了它们在 V 中的结合律、交换律、分配律、……其余的性质. 这样,线性运算的所有性质就全部满足了.

例 1.1.14 设平面上不经过坐标原点的直线为

$$V_1 = \{(2k, k+1) \mid k \in \mathbb{R}\}.$$

试验证它不是二维几何空间的子空间.

事实上,设 $\alpha = (2k_1, k_1 + 1) \in V_1, \beta = (2k_2, k_2 + 1) \in V_1$, 而 $\alpha + \beta = (2(k_1 + k_2), (k_1 + k_2) + 2) \notin V_1$, 所以 V_1 不是二维几何空间的子空间.

容易看出,每个线性空间至少有两个子空间,一个是它自身,另一个是仅由零向量所构成的子集合,称后者为**零子空间**. 这两个子空间通常称为**平凡子空间**,而其他的子空间称为**非平凡子空间(或真子空间)**.

既然线性子空间本身也是一个线性空间,因此,前面引入的关于基、维数、坐标等概念当然也可以应用到线性子空间上. 因为在线性子空间中不可能比在整个空间中有更多数目的线性无关的向量,所以,任何一个线性子空间的维数不能超过整个空间的维数,即有

$$\dim V_1 \leq \dim V.$$

例 1.1.15 设 $A \in \mathbb{R}^{m \times n}$, 齐次线性方程组

$$Ax = \mathbf{0}$$

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,记为 $N(A)$ 或 $\ker(A)$. 因为解空间的基就是齐次线性方程组的基础解系,所以 $\dim(N(A)) = n - \text{rank}(A)$.

下面讨论线性子空间的生成问题.

设 x_1, x_2, \dots, x_m 是线性空间 V 中一组向量,不难看出,这组向量所有可能的线性组合的集合

$$V_1 = \{k_1 x_1 + k_2 x_2 + \dots + k_m x_m\}$$

是非空的,而且容易验证 V_1 对 V 的线性运算是封闭的,因而 V_1 是 V 的一个子空间,这个子空间称为由 x_1, x_2, \dots, x_m 生成的子空间,记为

$$\text{Span}(x_1, x_2, \dots, x_m) = \{k_1 x_1 + k_2 x_2 + \dots + k_m x_m\}. \quad (1.1.7)$$

在有限维线性空间 V 中,任何一个子空间都可以由式(1.1.7)得到. 事实上,设 V_1 是 V 的一个子空间, V_1 当然是有限维的,令 x_1, x_2, \dots, x_m 是 V_1 的一组基,于是

$$V_1 = \text{Span}(x_1, x_2, \dots, x_m)$$

就是 m 维子空间. 特别地,零子空间就是由零向量生成的子空间 $\text{Span}(\mathbf{0})$.

例 1.1.16 试求由向量

$$\alpha_1 = \begin{pmatrix} 1 \\ 3 \\ 2 \\ 1 \end{pmatrix}, \quad \alpha_2 = \begin{pmatrix} 4 \\ 9 \\ 5 \\ 4 \end{pmatrix}, \quad \alpha_3 = \begin{pmatrix} 3 \\ 7 \\ 4 \\ 3 \end{pmatrix}$$

所生成的 \mathbb{R}^4 的子空间的基和维数.

解 设

$$k_1 \alpha_1 + k_2 \alpha_2 + k_3 \alpha_3 = \mathbf{0},$$

即

$$\begin{cases} k_1 + 4k_2 + 3k_3 = 0, \\ 3k_1 + 9k_2 + 7k_3 = 0, \\ 2k_1 + 5k_2 + 4k_3 = 0, \\ k_1 + 4k_2 + 3k_3 = 0. \end{cases}$$

解此线性方程组,得

$$k_2 = 2k_1, \quad k_3 = -3k_1.$$

于是有

$$\alpha_1 + 2\alpha_2 - 3\alpha_3 = \mathbf{0},$$

故 $\alpha_1, \alpha_2, \alpha_3$ 线性相关,又显见 α_1 与 α_2 (或 α_2 与 α_3 , 或 α_3 与 α_1) 线性无关,因此所论子空间的维数是 2,即

$$\dim \text{Span}(\alpha_1, \alpha_2, \alpha_3) = 2,$$

基是由 α_1 与 α_2 (或 α_2 与 α_3 , 或 α_3 与 α_1) 所组成.

2) 子空间的交与和

前面我们讨论了由线性空间的向量生成子空间的方法,这里将要讨论子空间的两种重要的运算——交与和,可以看成是由子空间生成的子空间.

定义 1.1.8 设 V_1 和 V_2 是 n 维线性空间 V 的两个子空间,由同时属于这两个子空间中的向量构成的子集合,叫做 V_1 与 V_2 的交,记作 $V_1 \cap V_2$.

定理 1.1.3 设 V_1, V_2 是数域 P 上的线性空间 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 因每一个子空间 V_1, V_2 都含有零向量,所以它们的交也含有零向量,于是 $V_1 \cap V_2$ 是非空集合;另一方面,若 $x, y \in V_1 \cap V_2$,那么 $x, y \in V_1, x, y \in V_2$,因 V_1, V_2 都是子空间,故 $x+y \in V_1, x+y \in V_2$,即 $x+y \in V_1 \cap V_2$;又因 $kx \in V_1, kx \in V_2$,故 $kx \in V_1 \cap V_2$,从而由定理 1.1.2 知 $V_1 \cap V_2$ 是 V 的子空间. 证毕

定义 1.1.9 设 V_1, V_2 都是数域 P 上的线性空间 V 的子空间,且 $x \in V_1, y \in V_2$,由所有 $x+y$ 这样的向量构成的集合叫做 V_1 与 V_2 的和或和空间,记作 $V_1 + V_2$.

定理 1.1.4 如果 V_1, V_2 都是数域 P 上的线性空间 V 的子空间,则它们的和 $V_1 + V_2$ 也是 V 的子空间.

证明 显然 $V_1 + V_2$ 非空,又对任两个向量 $x_1, x_2 \in V_1$ 及 $y_1, y_2 \in V_2$,有

$$(x_1 + y_1) + (x_2 + y_2) = (x_1 + x_2) + (y_1 + y_2) \in V_1 + V_2.$$

又因 $k(x_1 + y_1) = kx_1 + ky_2 \in V_1 + V_2$,这就证明了 $V_1 + V_2$ 是 V 的子空间.

证毕

例 1.1.17 在三维几何空间 \mathbb{R}^3 中, 用 V_1 表示过坐标原点的直线, V_2 表示一个通过坐标原点而且与 V_1 垂直的平面, 那么 V_1 与 V_2 的交是 $\{\mathbf{0}\}$, 而 V_1 与 V_2 的和是整个三维几何空间.

注意, 这里子空间的交与集合的交的概念是一致的, 但子空间的和与集合的并的概念是不一致的. 例如在例 1.1.17 中, $V_1 \cap V_2 = \{\mathbf{0}\}$, 表示子空间的交与集合的交一致; 但 $V_1 + V_2 = \mathbb{R}^3$, 而 $V_1 \cup V_2 = \{\text{过原点的直线与垂直平面上的点}\}$.

子空间的运算“交”与“和”, 易证其交换律与结合律均成立, 但“交”与“和”构成的分配律一般不成立.

例 1.1.18 设在 \mathbb{R}^2 中, $\mathbf{x}_1, \mathbf{x}_2$ 是两个线性无关的向量, 令

$$Q = \text{Span}(\mathbf{x}_1), \quad S = \text{Span}(\mathbf{x}_2), \quad T = \text{Span}(\mathbf{x}_1 + \mathbf{x}_2),$$

则

$$Q \cap (S + T) = Q \cap \mathbb{R}^2 = Q,$$

$$(Q \cap S) + (Q \cap T) = \{\mathbf{0}\} + \{\mathbf{0}\} = \{\mathbf{0}\}.$$

由此可见

$$Q \cap (S + T) \neq (Q \cap S) + (Q \cap T).$$

为便于讨论子空间交与和的维数公式, 先给出基的扩充定理.

定理 1.1.5 设 V_1 是数域 P 上 n 维线性空间 V 的一个 m 维子空间, $\alpha_1, \alpha_2, \dots, \alpha_m$ 是 V_1 的一组基, 那么 $\alpha_1, \alpha_2, \dots, \alpha_m$ 必定可扩充为整个空间的基, 也就是说, 在 V 中必定可以找到 $n-m$ 个向量 $\alpha_{m+1}, \alpha_{m+2}, \dots, \alpha_n$, 使得 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 V 的一组基.

证明 对维数 $n-m$ 应用数学归纳法, 当 $n-m=0$ 时定理显然成立, 因为 $\alpha_1, \alpha_2, \dots, \alpha_m$ 已经是 V 的基, 现在假定 $n-m=k$ 时定理成立, 我们考虑 $n-m=k+1$ 时的情形.

既然 $\alpha_1, \alpha_2, \dots, \alpha_m$ 还不是 V 的一组基, 它又是线性无关, 那么在 V 中必定有一个向量 α_{m+1} 不能被 $\alpha_1, \alpha_2, \dots, \alpha_m$ 线性表示, 把 α_{m+1} 补充进去, $\alpha_1, \alpha_2, \dots, \alpha_m, \alpha_{m+1}$ 必定是线性无关的, 这时子空间 $\text{Span}(\alpha_1, \alpha_2, \dots, \alpha_m, \alpha_{m+1})$ 是 $m+1$ 维的. 因为

$$n - (m+1) = (n-m) - 1 = k+1 - 1 = k,$$

由归纳假设知 $\text{Span}(\alpha_1, \alpha_2, \dots, \alpha_m, \alpha_{m+1})$ 的基 $\alpha_1, \alpha_2, \dots, \alpha_m, \alpha_{m+1}$ 可以扩充为整个空间 V 的基. 证毕

定理 1.1.6 (维数公式) 设 V_1 与 V_2 是数域 P 上的线性空间 V 的两个子空间, 则

$$\dim V_1 + \dim V_2 = \dim(V_1 + V_2) + \dim(V_1 \cap V_2). \quad (1.1.8)$$

证明 设 $\dim V_1 = n_1, \dim V_2 = n_2, \dim(V_1 \cap V_2) = m$, 我们要证明 $\dim(V_1 + V_2) = n_1 + n_2 - m$.

取 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ 为 $V_1 \cap V_2$ 的基. 根据定理 1.1.5, 将它依次扩充为 V_1, V_2 的一组基

$$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m; \mathbf{y}_1, \dots, \mathbf{y}_{n_1-m},$$

$$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m; \mathbf{z}_1, \dots, \mathbf{z}_{n_2-m},$$

即

$$V_1 = \text{Span}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m, \mathbf{y}_1, \dots, \mathbf{y}_{n_1-m}),$$

$$V_2 = \text{Span}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m, \mathbf{z}_1, \dots, \mathbf{z}_{n_2-m}).$$

所以

$$V_1 + V_2 = \text{Span}(\mathbf{x}_1, \dots, \mathbf{x}_m, \mathbf{y}_1, \dots, \mathbf{y}_{n_1-m}, \mathbf{z}_1, \dots, \mathbf{z}_{n_2-m}).$$

设

$$k_1 x_1 + \cdots + k_m x_m + p_1 y_1 + \cdots + p_{n_1-m} y_{n_1-m} + q_1 z_1 + \cdots + q_{n_2-m} z_{n_2-m} = \mathbf{0}.$$

令

$$\begin{aligned} x &= k_1 x_1 + \cdots + k_m x_m + p_1 y_1 + \cdots + p_{n_1-m} y_{n_1-m} \\ &= -q_1 z_1 - \cdots - q_{n_2-m} z_{n_2-m}, \end{aligned}$$

由上面的第一个等式知 $x \in V_1$, 由第二个等式知 $x \in V_2$, 于是 $x \in V_1 \cap V_2$, 故可令

$$x = l_1 x_1 + \cdots + l_m x_m,$$

则

$$l_1 x_1 + \cdots + l_m x_m = -q_1 z_1 - \cdots - q_{n_2-m} z_{n_2-m},$$

即

$$l_1 x_1 + \cdots + l_m x_m + q_1 z_1 + \cdots + q_{n_2-m} z_{n_2-m} = \mathbf{0}.$$

由于 $x_1, x_2, \cdots, x_m, z_1, \cdots, z_{n_2-m}$ 线性无关, 所以

$$l_1 = \cdots = l_m = q_1 = \cdots = q_{n_2-m} = 0,$$

因而 $x = \mathbf{0}$, 从而有

$$k_1 x_1 + \cdots + k_m x_m + p_1 y_1 + \cdots + p_{n_1-m} y_{n_1-m} = \mathbf{0}.$$

由于 $x_1, x_2, \cdots, x_m, y_1, \cdots, y_{n_1-m}$ 线性无关, 又得

$$k_1 = \cdots = k_m = p_1 = \cdots = p_{n_1-m} = 0.$$

这就证明了 $x_1, \cdots, x_m, y_1, \cdots, y_{n_1-m}, z_1, \cdots, z_{n_2-m}$ 线性无关, 因而它是 $V_1 + V_2$ 的一组基, $V_1 + V_2$ 的维数为 $n_1 + n_2 - m$, 于是维数公式成立. 证毕

从上述维数公式看出, 两个子空间和的维数往往要比维数的和来得小. 例如, 通过原点的两个不同平面的和是整个几何空间, 维数为 3, 而这两个平面维数之和等于 4. 由此也可推出这两个平面的交必然是一维的直线.

这里需要指出, 由子空间的和 $V_1 + V_2$ 的定义知道, 若 $x \in V_1, y \in V_2$, 则 $V_1 + V_2$ 中任一向量 z 可以表示为 $z = x + y$. 但是, 一般来说, 这种表示式并不是惟一的. 例如, 在 \mathbb{R}^3 中, 若 V_1 是 $x_1 = (1, 0, 0), x_2 = (0, 1, 0)$ 所生成的子空间, V_2 是 $y_1 = (1, 1, 0)$ 所生成的子空间, 则显然有 $z = (1, 2, 0) \in V_1 + V_2$. 但是, 一方面 $(1, 2, 0) = (1, 2, 0) + (0, 0, 0)$, 其中 $(1, 2, 0) \in V_1, (0, 0, 0) \in V_2$; 另一方面, $(1, 2, 0) = (0, 1, 0) + (1, 1, 0)$, 其中 $(0, 1, 0) \in V_1, (1, 1, 0) \in V_2$. 这就说明和空间中的向量 $z = (1, 2, 0)$ 的表示法不惟一. 针对这种现象, 我们在子空间的和中, 把能惟一表示成 $z = x + y$ 的那种和, 给予特别的定义.

定义 1.1.10 如果 $V_1 + V_2$ 中的任一向量只能惟一地表示为子空间 V_1 的一个向量与子空间 V_2 的一个向量的和, 则称 $V_1 + V_2$ 为直和(或直接和), 记为 $V_1 \oplus V_2$ 或 $V_1 \dot{+} V_2$.

定理 1.1.7 $V_1 + V_2$ 为直和的充要条件是: V_1 与 V_2 之交 $V_1 \cap V_2$ 为零子空间, 即

$$V_1 \cap V_2 = \{\mathbf{0}\}.$$

证明 充分性. 用反证法. 设 $V_1 \cap V_2$ 为零子空间, 若 $V_1 + V_2$ 的向量 z 不能惟一地表示为 V_1 与 V_2 的向量的和, 则必有 $x_1, x_2 \in V_1, y_1, y_2 \in V_2$, 且 $x_1 \neq x_2, y_1 \neq y_2$, 使

$$z = x_1 + y_1 \quad \text{及} \quad z = x_2 + y_2,$$

两式相减得

$$(x_1 - x_2) + (y_1 - y_2) = \mathbf{0},$$

因而有向量

$$w = (x_1 - x_2) = -(y_1 - y_2) \neq 0$$

既属于 V_1 又属于 V_2 , 即 $w \in V_1 \cap V_2$, 从而 $V_1 \cap V_2$ 不能为零子空间, 与假设矛盾.

必要性. 仍用反证法. 设 $V_1 + V_2$ 为直和, 若 $V_1 \cap V_2$ 不为零子空间, 则在 $V_1 \cap V_2$ 中至少有一向量 $x \neq 0$. 由于 $V_1 \cap V_2$ 是线性空间, 所以有 $-x \in V_1 \cap V_2$, 那么, 对 $V_1 + V_2$ 的零向量有 $0 = x + (-x)$, 其中 $x \in V_1, -x \in V_2$, 同时此零向量 0 又是 V_1 中的零向量与 V_2 中的零向量之和 $0 = 0 + 0$, 这就说明 $V_1 + V_2$ 的零向量表示法不惟一, 从而与 $V_1 + V_2$ 是直和的假设矛盾. 证毕

推论 1 $V_1 + V_2$ 是直和的充要条件为

$$\dim(V_1 + V_2) = \dim V_1 + \dim V_2. \quad (1.1.9)$$

证明 由公式(1.1.8)有

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

由定理 1.1.7 知, $V_1 + V_2$ 为直和的充要条件是 $V_1 \cap V_2 = \{0\}$, 这与 $\dim(V_1 \cap V_2) = 0$ 等价, 从而 $\dim(V_1 + V_2) = \dim V_1 + \dim V_2$. 证毕

推论 2 设 $V_1 + V_2$ 为直和, 若 x_1, \dots, x_{n_1} 是 V_1 的基, y_1, \dots, y_{n_2} 是 V_2 的基, 则 $x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}$ 为 $V_1 \oplus V_2$ 的基.

证明 由式(1.1.9)知, $\dim(V_1 \oplus V_2) = n_1 + n_2$, 而向量组

$$x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}$$

线性无关(事实上, 设其线性相关, 即有不全为 0 的常数 $k_1, \dots, k_{n_1}, l_1, \dots, l_{n_2}$, 使

$$k_1 x_1 + \dots + k_{n_1} x_{n_1} + l_1 y_1 + \dots + l_{n_2} y_{n_2} = 0,$$

则

$$x = k_1 x_1 + \dots + k_{n_1} x_{n_1} = -l_1 y_1 - \dots - l_{n_2} y_{n_2} \neq 0.$$

从而 $x \in V_1, x \in V_2$, 即 $x \in V_1 \cap V_2 \neq \{0\}$, 这与定理 1.1.7 矛盾). 这样, 线性无关向量组所含向量的个数正好等于 $V_1 \oplus V_2$ 的维数, 所以 $x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}$ 是 $V_1 \oplus V_2$ 的基. 证毕

定理 1.1.8 设 V_1 是 n 维线性空间 V 的一个子空间, 则一定存在 V 的一个子空间 V_2 , 使

$$V = V_1 \oplus V_2.$$

证明 设 x_1, \dots, x_m 是 V_1 的一组基, 由定理 1.1.5 知, 可以把它扩充为 V 的一组基 $x_1, \dots, x_m, x_{m+1}, x_{m+2}, \dots, x_n$, 令

$$V_2 = \text{Span}(x_{m+1}, x_{m+2}, \dots, x_n),$$

显然满足直和维数公式(1.1.9), 从而有 $V = V_1 \oplus V_2$, 即 V_2 为所求. 证毕

定理 1.1.8 表明线性空间 V 可作直和分解, 且易知这种直和分解不是惟一的.

关于子空间的和与交的概念以及有关的定理, 可以推广到多个子空间的情形.

习 题 1 (1)

1. 有没有一个向量的线性空间? 有没有两个向量的线性空间? 有没有 m 个向量的线性空间?
2. 在几何空间 \mathbb{R}^3 中, 按通常的向量加法和数乘的运算, 下列集合是否是实数域 \mathbb{R} 上的线性空间?
 - (1) 过原点的平面 H 上的所有向量集合;
 - (2) 位于第一象限, 以原点为始点的向量集合;

- (3) 位于第一、三象限,以原点为始点的向量集合;
 (4) x 轴上的向量集合;
 (5) 平面 H 上,不平行于某向量的向量集合.

3. 按通常的数的加法和数乘,下列数集是否构成有理数域 \mathbb{Q} 上的线性空间? 是否构成实数域 \mathbb{R} 上的线性空间?

(1) 整数集 \mathbb{Z} ; (2) 有理数集 \mathbb{Q} ; (3) 实数集 \mathbb{R} ; (4) 复数集 \mathbb{C} ; (5) 单个数的集合 $\{0\}$; (6) n 个数的集合 $N = \{a_1, a_2, \dots, a_n\}$.

4. 齐次线性方程组

$$\begin{cases} 2x_1 + x_2 + x_3 + x_4 = 0, \\ x_1 + 2x_2 + x_3 + 3x_4 = 0 \end{cases}$$

的全部解是否构成线性空间?

5. 在 n 维向量空间 P^n 中,下列 n 维向量的集合 V ,是否构成数域 P 上的线性空间?

(1) $V = \{(a, b, a, b, \dots, a, b) \mid a, b \in P\}$;

(2) $V = \{(a_1, a_2, \dots, a_n) \mid \sum_{i=1}^n a_i = 1\}$.

6. 检验以下集合对于矩阵的加法和实数与矩阵的乘法,是否构成实数域上的线性空间?

(1) 设 A 是 n 阶实数矩阵, A 的实系数多项式的全体;

(2) n 阶实对称(或上三角)矩阵的全体;

(3) 形如 $\begin{bmatrix} 0 & a \\ -a & b \end{bmatrix}$ 的二阶方阵的全体.

7. 设 $V = \{x \mid x = C_1 \sin t + C_2 \sin 2t + \dots + C_n \sin nt, C_i \in \mathbb{R}, 0 \leq t \leq 2\pi\}$. V 中元素对于通常三角多项式的加法和数乘运算,是否构成实数域上的线性空间? 若 V 是 \mathbb{R} 上的线性空间,证明 $\{\sin t, \sin 2t, \dots, \sin nt\}$ 是 V 的一组基,试提出确定 C_i 的方法.

8. 设 V 是有序实数对的集合: $V = \{(a, b) \mid a, b \in \mathbb{R}\}$,规定如下的加法与数乘运算为

(1) $(a, b) \oplus (c, d) = (a+c, b+d)$ 与 $k \circ (a, b) = (ka, b)$;

(2) $(a, b) \oplus (c, d) = (a, b)$ 与 $k \circ (a, b) = (ka, kb)$;

(3) $(a, b) \oplus (c, d) = (a+c, b+d)$ 与 $k \circ (a, b) = (k^2 a, k^2 b)$;

(4) $(a, b) \oplus (c, d) = (a+c, b+d+ac)$ 与 $k \circ (a, b) = \left(ka, kb + \frac{k(k-1)}{2} a^2\right)$.

那么 V 关于运算 \oplus, \circ 是否构成 \mathbb{R} 上的线性空间?

9. 证明: 线性空间定义中,加法的交换率 $\alpha + \beta = \beta + \alpha$ 不是独立的,即它可由其余 7 条性质推出.

10. 求线性方程组

$$\begin{cases} x_1 + x_2 - 3x_3 - x_4 = 0, \\ 3x_1 - x_2 - 3x_3 + 4x_4 = 0, \\ x_1 + 5x_2 - 9x_3 - 8x_4 = 0 \end{cases}$$

的解空间 V 的维数与基.

11. 证明: $x^2 + x, x^2 - x, x + 1$ 是线性空间 $P[x]_2$ 的基底,并求出 $2x^2 + 7x + 3$ 在此基下的坐标.

12. 在 \mathbb{R}^4 中有两组基

$$\mathbf{x}_1 = (1, 0, 0, 0), \quad \mathbf{x}_2 = (0, 1, 0, 0),$$

$$\mathbf{x}_3 = (0, 0, 1, 0), \quad \mathbf{x}_4 = (0, 0, 0, 1),$$

与

$$\mathbf{y}_1 = (2, 1, -1, 1), \quad \mathbf{y}_2 = (0, 3, 1, 0),$$

$$\mathbf{y}_3 = (5, 3, 2, 1), \quad \mathbf{y}_4 = (6, 6, 1, 3).$$

- (1) 求由基 x_1, x_2, x_3, x_4 到基 y_1, y_2, y_3, y_4 的过渡矩阵;
 (2) 求向量 $x = (\xi_1, \xi_2, \xi_3, \xi_4)$ 在基 y_1, y_2, y_3, y_4 下的坐标;
 (3) 求对两组基有相同坐标的非零向量.
13. 求下列线性空间的维数与一组基:

- (1) 习题 6(2) 中的空间;
 (2) 例 1.1.5 中的空间 \mathbb{R}^+ ;
 (3) 实数域 \mathbb{R} 上由矩阵 A 的实系数多项式的全体组成的空间, 其中

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega^2 \end{pmatrix}, \quad \omega = \frac{-1 + \sqrt{3}i}{2}, \quad \omega^2 = \bar{\omega}, \quad \omega^3 = 1.$$

14. 设线性空间 V^4 的基(Ⅰ) $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 和基(Ⅱ) $\beta_1, \beta_2, \beta_3, \beta_4$ 满足

$$\begin{cases} \alpha_1 + 2\alpha_2 = \beta_3, \\ \alpha_2 + 2\alpha_3 = \beta_1, \\ \beta_1 + 2\beta_2 = \alpha_3, \\ \beta_2 + 2\beta_3 = \alpha_4. \end{cases}$$

- (1) 求由基(Ⅰ)到基(Ⅱ)的过渡矩阵;
 (2) 求向量 $\alpha = 2\beta_1 - \beta_2 + \beta_3 + \beta_4$ 在基(Ⅰ)下的坐标;
 (3) 判断是否存在非零元素 $\alpha \in V^4$, 使得 α 在基(Ⅰ)和基(Ⅱ)下的坐标相同.
15. 在 $\mathbb{R}^{2 \times 2}$ 中, 求由基(Ⅰ):

$$A_1 = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 1 \\ 2 & 2 \end{bmatrix}, \quad A_3 = \begin{bmatrix} -2 & 1 \\ 1 & 2 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 1 & 3 \\ 1 & 2 \end{bmatrix}$$

到基(Ⅱ):

$$B_1 = \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \quad B_3 = \begin{bmatrix} -1 & 2 \\ 1 & 1 \end{bmatrix}, \quad B_4 = \begin{bmatrix} -1 & -1 \\ 0 & 1 \end{bmatrix}$$

的过渡矩阵.

16. 设 $P[x]_3$ 的两基为:

$$\begin{aligned} \text{(Ⅰ)} \quad & f_1(x) = 1, & f_2(x) &= 1+x, \\ & f_3(x) = 1+x+x^2, & f_4(x) &= 1+x+x^2+x^3; \\ \text{(Ⅱ)} \quad & g_1(x) = 1+x^2+x^3, & g_2(x) &= x+x^2+x^3, \\ & g_3(x) = 1+x+x^2, & g_4(x) &= 1+x+x^3. \end{aligned}$$

- (1) 求由基(Ⅰ)到基(Ⅱ)的过渡矩阵;
 (2) 求 $P[x]_3$ 中在基(Ⅰ)和基(Ⅱ)下有相同坐标的全体多项式.
17. 设非齐次线性方程组 $AX = b$ 有解, 其通解表达式为

$$\xi + k_1 \alpha_1 + k_2 \alpha_2 + \cdots + k_{n-r} \alpha_{n-r},$$

其中 k_1, k_2, \dots, k_{n-r} 为任意数. 试证: 向量组 $\xi, \xi + \alpha_1, \xi + \alpha_2, \dots, \xi + \alpha_{n-r}$ 是方程组 $AX = b$ 所有解的极大线性无关组, 但是 $AX = b$ 的所有解集合不构成线性空间.

18. 证明线性空间的替换定理: 设 $J = \{\alpha_1, \alpha_2, \dots, \alpha_t\}$ 与 $K = \{\beta_1, \beta_2, \dots, \beta_s\}$ 是 n 维线性空间 V 的两个向量组, 其中 J 线性无关. 如果每个 $\alpha_i \in J$ 都可由 K 线性表示, 则 $s \leq t$; 且可将 K 中的某 s 个向量换成 $\alpha_1, \alpha_2, \dots, \alpha_s$, 使得新的向量组生成的子空间与 K 生成的子空间相同.

19. 证明有限维线性空间的任意两个基所含向量的个数相同.
 20. 求出书中例 1.1.5 所述线性空间 \mathbb{R}^+ 的一个基与维数.
 21. 求下列线性空间的基及维数.
 (1) $V_1 = \{\alpha = (x_1, x_2, x_3) \mid 2x_1 - x_2 = 0\}$;

$$(2) V_2 = \{X | AX = XA, X \in \mathbb{R}^{3 \times 3}\}, A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix};$$

$$(3) V_3 = \{A_{n \times n} | A^T = -A\}.$$

$$22. \text{ 设 } V = \{\alpha = (x, y) | x, y \in \mathbb{C}\},$$

(1) V 在通常向量加法和数乘下, 在复数域上是多少维空间?

(2) V 在通常向量加法和数乘下, 在实数域上是多少维空间?

23. 设 $P[x]_{n+1}$ 是次数不大于 n 的实系数多项式空间,

$$W = \{f(x) | f(1) = 0, f(x) \in P[x]_{n+1}\},$$

证明 W 是一个线性空间, 并求一组基及维数.

24. 设 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 是线性空间中 4 个线性无关的向量, 求

$$W = L[\alpha_1 + \alpha_2, \alpha_2 + \alpha_3, \alpha_3 + \alpha_4, \alpha_4 + \alpha_1]$$

的基及维数.

$$25. \mathbb{R}^4 \text{ 中, 设 } \alpha_1 = (1, 2, 2, 3)^T, \alpha_2 = (1, 1, 2, 3)^T, \alpha_3 = (-1, 1, -4, -5)^T, \alpha_4 = (1, -3, 6, 7)^T,$$

(1) 设 $W = L[\alpha_1, \alpha_2, \alpha_3, \alpha_4]$, 求 W 的一组基及维数;

(2) $A = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$, 求 $N(A)$ 的基及维数; (其中 $N(A)$ 称为矩阵 A 的零空间, 即为齐次线性方程组 $AX=0$ 的解空间);

(3) 记 $R(A) = \{y | y = AX, \forall X \in \mathbb{R}^4\} = \{y = x_1\alpha_1 + x_2\alpha_2 + x_3\alpha_3 + x_4\alpha_4\} = L[\alpha_1, \alpha_2, \alpha_3, \alpha_4]$ (即 $R(A)$ 为 A 的像空间).

求 $R(A)$ 的基及维数.

26. 在 n 维向量空间 \mathbb{R}^n 中, 问分量满足下列条件的全体向量能否构成子空间?

(1) $x_1 + x_2 + \cdots + x_n = 0$; (2) $x_1 + x_2 + \cdots + x_n = 1$.

27. 假定 $\alpha_1, \alpha_2, \alpha_3$ 是 \mathbb{R}^3 的一组基, 试求由

$$\alpha'_1 = \alpha_1 - 2\alpha_2 + 3\alpha_3, \quad \alpha'_2 = 2\alpha_1 + 3\alpha_2 + 2\alpha_3, \quad \alpha'_3 = 4\alpha_1 + 13\alpha_2$$

生成的子空间 $\text{Span}(\alpha'_1, \alpha'_2, \alpha'_3)$ 的基.

28. 判断下列子集是否构成子空间?

(1) 给定矩阵 $P \in \mathbb{R}^{n \times n}, \mathbb{R}^{n \times n}$ 的子集

$$V_1 = \{A | AP = PA, A \in \mathbb{R}^{n \times n}\};$$

(2) $\mathbb{R}^{2 \times 2}$ 的子集

$$V_1 = \{A | \det A = 0, A \in \mathbb{R}^{2 \times 2}\};$$

$$V_2 = \{A | A^2 = A, A \in \mathbb{R}^{2 \times 2}\}.$$

29. 试证: 在 \mathbb{R}^4 中, 由 $(1, 1, 0, 0), (1, 0, 1, 1)$ 生成的子空间与由 $(2, -1, 3, 3), (0, 1, -1, -1)$ 生成的子空间相同.

30. 已知 $P = \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}$ 及 $\mathbb{R}^{2 \times 2}$ 的子空间 (见习题 28(1))

$$V_1 = \{A | AP = PA, A \in \mathbb{R}^{2 \times 2}\};$$

(1) 求 V_1 的基与维数;

(2) 写出 V_1 中的矩阵的一般形式.

31. 设 V_1, V_2 都是线性空间 V 的子空间, 且 $V_1 \subseteq V_2$. 证明: 如果 $\dim V_1 = \dim V_2$, 则 $V_1 = V_2$.

32. 求 \mathbb{R}^4 的子空间

$$V = \{(a_1, a_2, a_3, a_4) | a_1 - a_2 + a_3 - a_4 = 0\},$$

$$W = \{(a_1, a_2, a_3, a_4) | a_1 + a_2 + a_3 + a_4 = 0\}$$

的交 $V \cap W$ 的一组基.

33. 给定 $\mathbb{R}^{2 \times 2} = \{A = (a_{ij})_{2 \times 2} \mid a_{ij} \in \mathbb{R}\}$ (即数域 \mathbb{R} 上的二阶实方阵按通常矩阵的加法与数乘构成的线性空间) 的子集

$$V = \{A = (a_{ij})_{2 \times 2} \mid a_{11} + a_{22} = 0, a_{ij} \in \mathbb{R}\}.$$

(1) 证明: V 是 $\mathbb{R}^{2 \times 2}$ 的子空间.

(2) 求 V 的维数与基.

34. 判别下列集合是否构成子空间:

(1) $W_1 = \{\alpha = (x, y, z) \mid x^2 + y^2 + z^2 \leq 1, \mathbb{R}$ 为实数域};

(2) $W_2 = \{A \mid A^2 = I, A \in \mathbb{R}^{n \times n}\}$;

(3) \mathbb{R}^3 中, $W_3 = \left\{ \alpha = (x_1, x_2, x_3) \mid \int_0^t (x_1 \tau^2 + x_2 \tau + x_3) d\tau = 0 \right\}$;

(4) $W_4 = \left\{ A = (a_{ij})_{m \times n} \mid \sum_{i=1}^m \sum_{j=1}^n a_{ij} = 0, A \in \mathbb{R}^{m \times n} \right\}$.

35. 证明: 设 $U_1 = L[\alpha_1, \alpha_2, \dots, \alpha_r], U_2 = L[\beta_1, \beta_2, \dots, \beta_r]$, 则 $U_1 = U_2$ 的充分必要条件是生成元 $\alpha_1, \alpha_2, \dots, \alpha_r$ 与 $\beta_1, \beta_2, \dots, \beta_r$ 等价.

36. 设向量 $\alpha_1 = (1, 0, 2, 1)^T, \alpha_2 = (2, 0, 1, -1)^T, \alpha_3 = (1, 0, 1, 0)^T, \beta_1 = (1, 1, 0, 1)^T, \beta_2 = (4, 1, 3, 1)^T$, 令 $V_1 = L[\alpha_1, \alpha_2, \alpha_3], V_2 = L[\beta_1, \beta_2]$, 求 $V_1 + V_2$ 的基与维数.

37. 设 $W_1 = L[\alpha_1, \alpha_2], W_2 = L[\beta_1, \beta_2]$, 其中 $\alpha_1 = (1, 2, 1, 0)^T, \alpha_2 = (-1, 1, 1, 1)^T, \beta_1 = (2, -1, 0, 1)^T, \beta_2 = (1, -1, 3, 7)^T$, 求 $W_1 + W_2$ 与 $W_1 \cap W_2$ 的维数, 并求出 $W_1 \cap W_2$.

38. 设子空间 $V_1 = \{\alpha = (x_1, x_2, x_3, x_4)^T \mid x_1 + x_2 + x_3 + x_4 = 0\}, V_2 = \{\alpha = (x_1, x_2, x_3, x_4)^T \mid x_1 - x_2 + x_3 - x_4 = 0\}$, 求 $V_1 + V_2$ 的一组基及 $V_1 \cap V_2$ 的维数, 并求 $V_1 \cap V_2$ 的一组基.

39. 求由向量 x_i 生成的子空间与由向量 y_i 生成的子空间的交与和的基及维数:

$$(1) \begin{cases} x_1 = (1, 2, 1, 0), \\ x_2 = (-1, 1, 1, 1), \end{cases} \quad \begin{cases} y_1 = (2, -1, 0, 1), \\ y_2 = (1, -1, 3, 7); \end{cases}$$

$$(2) \begin{cases} x_1 = (1, 2, -1, -2), \\ x_2 = (3, 1, 1, 1), \\ x_3 = (-1, 0, 1, -1), \end{cases} \quad \begin{cases} y_1 = (2, 5, -6, -5), \\ y_2 = (-1, 2, -7, 3). \end{cases}$$

40. 设 e_1, e_2, \dots, e_n 与 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 是 n 维线性空间 V 的两组基. 证明:

(1) 在两组基下坐标完全相同的全体向量的集合 V_1 是 V 的子空间;

(2) 若空间 V 的每个向量在这两组基下的坐标完全相同, 则 $e_i = \varepsilon_i (i=1, 2, \dots, n)$.

41. 试证明所有二阶方阵的集合形成的实线性空间 $\mathbb{R}^{2 \times 2}$, 是所有二阶实对称矩阵的集合形成的子空间与所有二阶反对称矩阵的集合形成的子空间的直和.

42. 设 V_1, V_2 分别是齐次线性方程组

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

与

$$x_1 = x_2 = \dots = x_n$$

的解空间, 试证明 $P^n = V_1 \oplus V_2$.

43. 证明: 每个 n 维线性空间都可以表示成 n 个一维子空间的直和.

44. 设 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 是 \mathbb{R}^4 的一组基, $V_1 = L[2\alpha_1 + \alpha_2, \alpha_1], V_2 = L[\alpha_3 - \alpha_4, \alpha_1 + \alpha_4]$, 证明 $\mathbb{R}^4 = V_1 \oplus V_2$.

45. 设连续实函数线性空间 U 中, 函数 $\alpha_1 = 1, \alpha_2 = \cos x, \alpha_3 = \cos 2x, \alpha_4 = \cos 3x$, 子空间 $W_1 = L[\alpha_1, \alpha_2], W_2 = L[\alpha_3, \alpha_4]$, 证明:

(1) $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 线性无关;

(2) $U = W_1 \oplus W_2$.

46. 设 A 与 B 分别为 $m \times n, s \times n$ 矩阵, 齐次方程组 $AX = 0$ 和 $BX = 0$ 无公共解, 且 $\text{rank}(A) = r$, 方程组 $BX = 0$ 的解空间的维数为 r , 证明 $N(A) \oplus N(B) = \mathbb{R}^n$.

47. 设 $A, B, C, D \in \mathbb{R}^{n \times n}$ 两两可交换, 且 $AC + BD = I$ (I 为单位矩阵), 证明: $N(AB) = N(A) \oplus N(B)$.

1.1.2 线性算子及其矩阵

为了研究线性空间之间的关系,本节引进线性算子的概念,而抽象的线性算子又与具体的矩阵发生联系.下面在介绍线性算子基本概念的基础上,重点讨论它的特殊情形——线性变换.

1. 线性空间上的线性算子

定义 1.1.11 设 M 与 M' 为两个集合,对于每个 $x \in M$,如果根据某种法则 \mathcal{A} ,在 M' 中有确定的 x' 与之对应,那么称 \mathcal{A} 为由 M 到 M' 的一个映射,或称算子.记为 $\mathcal{A}: M \rightarrow M'$,或 $\mathcal{A}(x) = x'$.

此时, x' 叫做 x 在 \mathcal{A} 下的像, x 叫做 x' 的原像, M 是 \mathcal{A} 的定义域, x' 的全体构成 \mathcal{A} 的值域,记为 $\mathcal{A}(M)$.由定义可知 $\mathcal{A}(M) \subseteq M'$.

例如,在高等数学中的函数 $y = f(x)$.若定义域为 $[a, b]$,值域为 $[c, d]$,则对应规则 f 就是由区间 $[a, b]$ 到区间 $[c, d]$ 的一个映射或算子,换言之,算子(或映射)的概念是函数概念的推广.

在各类函数中,有一类最基本的函数——线性函数,例如, $y = f(x) = ax$ (a 为任意常数),它具有下列性质:

$$(1) \text{ 可加性: } f(x_1 + x_2) = a(x_1 + x_2) = ax_1 + ax_2 = f(x_1) + f(x_2);$$

$$(2) \text{ 齐次性: } f(kx) = a(kx) = k(ax) = kf(x).$$

下面我们把过原点的线性函数具备的这两条性质推广到自变量是向量的算子,就是线性算子或线性映射的概念.

定义 1.1.12 设 V 与 V' 为数域 P 上的两个线性空间, \mathcal{A} 是由 V 到 V' 的一个算子,且对于 V 的任何两个向量 $x_1, x_2 \in V$ 和任何数 $\lambda \in P$,有

$$\mathcal{A}(x_1 + x_2) = \mathcal{A}(x_1) + \mathcal{A}(x_2), \quad (1.1.10)$$

$$\mathcal{A}(\lambda x_1) = \lambda \mathcal{A}(x_1), \quad (1.1.11)$$

则称 \mathcal{A} 是由 V 到 V' 的线性算子(或线性映射).

例 1.1.19 设 V 是 n 维向量空间 \mathbb{R}^n , V' 是 m 维向量空间 \mathbb{R}^m , A 是 $m \times n$ 矩阵,将 A 左乘 $x \in \mathbb{R}^n$,得到的是 \mathbb{R}^m 中的一个向量 x' ,因而将矩阵 A 左乘 $x \in \mathbb{R}^n$,即可定义一个由 \mathbb{R}^n 到 \mathbb{R}^m 的算子 \mathcal{A} ,即

$$\mathcal{A}(x) = Ax, \quad x \in \mathbb{R}^n.$$

不难验证,如果 $y \in \mathbb{R}^n, k \in \mathbb{R}$,则有

$$\mathcal{A}(x + y) = A(x + y) = Ax + Ay = \mathcal{A}(x) + \mathcal{A}(y),$$

$$\mathcal{A}(kx) = A(kx) = kAx = k\mathcal{A}(x),$$

所以 \mathcal{A} 是一个线性算子.

例 1.1.20 图 1.2 表示一根简支梁.载荷 F 产生一挠度曲线 $y = f(x)$,载荷 Q 产生一挠度曲线 $y = q(x)$.如果载荷 F 与 Q 都作用到梁上,就产生挠度曲线

$$y = f(x) + q(x).$$

如果把载荷 F 乘 k 倍,即施以载荷 kF ,相应的挠度曲线为

$$y = kf(x),$$

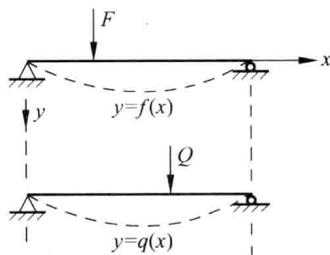


图 1.2 叠加原理

这种性质在力学上,称为“叠加原理”.

这一性质可以用算子(映射)的观点表示为:

如果

$$F \rightarrow y = f(x), \quad Q \rightarrow y = q(x),$$

则 (1) $F+Q \rightarrow y = f(x) + q(x)$,

(2) $kF \rightarrow y = kf(x)$.

这种性质,在数学里称为“线性”.因此,线性空间上的线性算子是研究物理、力学中满足叠加原理的系统的重要数学模型.

有时,线性算子的两个条件(可加性与齐次性)也可以写成另一形式.即 \mathcal{A} 是线性算子的充分必要条件是:对任何 $\mathbf{x}_1, \mathbf{x}_2 \in V$ 及 $\lambda_1, \lambda_2 \in P$, 有

$$\mathcal{A}(\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2) = \lambda_1 \mathcal{A}(\mathbf{x}_1) + \lambda_2 \mathcal{A}(\mathbf{x}_2). \quad (1.1.12)$$

事实上,由式(1.1.10)和式(1.1.11)有

$$\begin{aligned} \mathcal{A}(\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2) &= \mathcal{A}(\lambda_1 \mathbf{x}_1) + \mathcal{A}(\lambda_2 \mathbf{x}_2) \\ &= \lambda_1 \mathcal{A}(\mathbf{x}_1) + \lambda_2 \mathcal{A}(\mathbf{x}_2). \end{aligned}$$

反之,只要在式(1.1.12)中令 $\lambda_1 = \lambda_2 = 1$, 就得到式(1.1.10); 再令 $\lambda_2 = 0$, 便可得式(1.1.11).

因此,我们常常将式(1.1.12)作为线性算子的定义也是可行的.

例 1.1.21 将线性空间 V 中每一向量映射成线性空间 V' 中的零向量的算子 \mathcal{O} 叫做零算子,它是一个线性算子,因为

$$\mathcal{O}(\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2) = \mathbf{0} = \lambda_1 \mathbf{0} + \lambda_2 \mathbf{0} = \lambda_1 \mathcal{O}(\mathbf{x}_1) + \lambda_2 \mathcal{O}(\mathbf{x}_2),$$

即满足式(1.1.12), 所以是线性算子.

又如,任意给定数 $k \in P$, 数域 P 上线性空间 V 到自身的一个算子 \mathcal{K} 叫做 V 上的由数 k 决定的数乘算子,即 $\mathcal{K}(\mathbf{x}) = k\mathbf{x}, \forall \mathbf{x} \in V$. 读者不难验证 \mathcal{K} 也是一个线性算子.

例 1.1.22 在线性空间 $P[x]_n$ 中,微分算子是一个线性算子,这个算子通常用 \mathcal{D} 表示,即

$$\mathcal{D}(f(x)) = f'(x), \quad \forall f(x) \in P[x]_n, \quad x \in \mathbb{R}.$$

事实上,对任意的 $f(x), g(x) \in P[x]_n$, 及 $\lambda_1, \lambda_2 \in \mathbb{R}$, 有

$$\begin{aligned} \mathcal{D}(\lambda_1 f(x) + \lambda_2 g(x)) &= (\lambda_1 f(x) + \lambda_2 g(x))' = \lambda_1 f'(x) + \lambda_2 g'(x) \\ &= \lambda_1 \mathcal{D}(f(x)) + \lambda_2 \mathcal{D}(g(x)). \end{aligned}$$

例 1.1.23 在 $[a, b]$ 上一切实连续函数的线性空间 $C[a, b]$ 中,定义积分算子 \mathcal{A} , 即

$$\mathcal{A}(f(t)) = \int_a^x f(t) dt, \quad \forall f(t) \in C[a, b],$$

则 \mathcal{A} 是线性空间 $C[a, b]$ 上的线性算子.

事实上,有

$$\begin{aligned} \mathcal{A}(\lambda_1 f(t) + \lambda_2 g(t)) &= \int_a^x (\lambda_1 f(t) + \lambda_2 g(t)) dt \\ &= \lambda_1 \int_a^x f(t) dt + \lambda_2 \int_a^x g(t) dt \\ &= \lambda_1 \mathcal{A}(f(t)) + \lambda_2 \mathcal{A}(g(t)). \end{aligned}$$

从例 1.1.22 和例 1.1.23 可见,作为数学分析的两大运算——微分与积分,从算子(或映射)的角度来看都是线性算子,可见,线性算子是非常广泛的.

然而,非线性算子也是大量存在的.例如, $\mathcal{A}(\mathbf{A}) = \det \mathbf{A}, \mathbf{A} \in \mathbb{R}^{n \times n}$, 则 \mathcal{A} 是由 $\mathbb{R}^{n \times n}$ 到实数

集 \mathbb{R} 的一个算子(或映射),但不是线性算子.这是因为在一般情况下, $\det(\mathbf{A}+\mathbf{B})\neq\det\mathbf{A}+\det\mathbf{B}$,故不满足可加性, $\mathcal{A}(\mathbf{A}+\mathbf{B})\neq\mathcal{A}(\mathbf{A})+\mathcal{A}(\mathbf{B})$.又如,如果定义 $\mathcal{A}(\mathbf{x})=1$,那么 \mathcal{A} 是个算子(或映射),但不是线性算子,这是因为 $\mathcal{A}(\mathbf{x}_1+\mathbf{x}_2)=1$,而 $\mathcal{A}(\mathbf{x}_1)+\mathcal{A}(\mathbf{x}_2)=1+1=2$,故

$$\mathcal{A}(\mathbf{x}_1+\mathbf{x}_2)\neq\mathcal{A}(\mathbf{x}_1)+\mathcal{A}(\mathbf{x}_2).$$

不过,非线性算子不是我们这里研究的对象.

线性算子有以下性质:

(1) 线性算子 \mathcal{A} 把 V 中的零向量变为 V' 中的零向量;把向量 \mathbf{x} 的负向量 $-\mathbf{x}$ 变为 \mathbf{x} 的像 $\mathcal{A}(\mathbf{x})$ 的负向量 $-\mathcal{A}(\mathbf{x})$.

事实上,这是因为

$$\mathcal{A}(\mathbf{0}) = \mathcal{A}(0\mathbf{x}) = 0\mathcal{A}(\mathbf{x}) = 0\mathbf{x}' = \mathbf{0},$$

$$\mathcal{A}(-\mathbf{x}) = \mathcal{A}(-1\mathbf{x}) = -1\mathcal{A}(\mathbf{x}) = -\mathcal{A}(\mathbf{x}).$$

(2) 线性算子 \mathcal{A} 把线性相关的向量组仍变为线性相关的向量组,即若 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r$ 线性相关,则它们的像 $\mathcal{A}(\mathbf{x}_1), \mathcal{A}(\mathbf{x}_2), \dots, \mathcal{A}(\mathbf{x}_r)$ 也线性相关.

事实上,如果

$$k_1\mathbf{x}_1 + k_2\mathbf{x}_2 + \dots + k_r\mathbf{x}_r = \mathbf{0},$$

其中 $k_i (i=1, 2, \dots, r)$ 不全为零,用 \mathcal{A} 作用上式两端,则有

$$k_1(\mathcal{A}(\mathbf{x}_1)) + k_2(\mathcal{A}(\mathbf{x}_2)) + \dots + k_r(\mathcal{A}(\mathbf{x}_r)) = \mathcal{A}(\mathbf{0}) = \mathbf{0}.$$

但要注意,线性算子可能把线性无关的向量组变为线性相关的向量组,如零算子 \mathcal{O} 就是这样.

所以,如果 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r$ 为 V 中的一组线性无关的向量,为使它们的像 $\mathcal{A}(\mathbf{x}_1), \mathcal{A}(\mathbf{x}_2), \dots, \mathcal{A}(\mathbf{x}_r)$ 在 V' 中为线性无关的向量,必须对线性算子 \mathcal{A} 补充条件,这就是同构算子的概念.

2. 同构算子与线性空间同构

定义 1.1.13 设 \mathcal{A} 是由 V 到 V' 的线性算子,且是“一对一”的,即满足

(1) $\mathcal{A}(V)=V'$ (全映射);

(2) 若 $\mathbf{x}_1, \mathbf{x}_2 \in V$,当 $\mathbf{x}_1 \neq \mathbf{x}_2$ 时,有 $\mathcal{A}(\mathbf{x}_1) \neq \mathcal{A}(\mathbf{x}_2)$;换言之,由 $\mathcal{A}(\mathbf{x}_1) = \mathcal{A}(\mathbf{x}_2)$,就有 $\mathbf{x}_1 = \mathbf{x}_2$ (可逆映射);

那么称 \mathcal{A} 为 V 与 V' 间的一个同构算子.

若 V 与 V' 之间存在同构算子,则称 V 与 V' 是同构的线性空间,简称 V 与 V' 同构.

简单地说,一对一的线性算子称为同构算子.

例 1.1.24 不高于3次的实系数多项式(包括0)与以实数为分量的四维向量之间可以建立一一对应的关系

$$a_0 + a_1x + a_2x^2 + a_3x^3 \leftrightarrow (a_0, a_1, a_2, a_3).$$

容易验证这种对应关系就是一个一对一的线性算子,所以由不高于3次的实系数多项式组成的线性空间 $P[x]_3$ 与由实数域上四维向量全体组成的线性空间 \mathbb{R}^4 同构.

根据线性空间同构的定义,可推知同构线性空间的基本性质:

性质 1 传递性: 设 V_1, V_2, V_3 是数域 P 上的线性空间,如果 V_1 与 V_2 同构, V_2 与 V_3 同构,则 V_1 与 V_3 也同构.

事实上,设 $\mathbf{x}_1 \in V_1, \mathbf{x}_2 \in V_2, \mathbf{x}_3 \in V_3$,若 $\mathbf{x}_1 \leftrightarrow \mathbf{x}_2 = \mathcal{A}(\mathbf{x}_1), \mathbf{x}_2 \leftrightarrow \mathbf{x}_3 = \mathcal{B}(\mathbf{x}_2)$,则令 $\mathbf{x}_1 \leftrightarrow \mathbf{x}_3 = \sigma(\mathbf{x}_1)$,于是,当 $\mathbf{y}_1 \leftrightarrow \mathbf{y}_2 = \mathcal{A}(\mathbf{y}_1), \mathbf{y}_2 \leftrightarrow \mathbf{y}_3 = \mathcal{B}(\mathbf{y}_2)$ 时,有

$$\lambda x_1 + \mu y_1 \leftrightarrow \mathcal{A}(\lambda x_1 + \mu y_1) = \lambda x_2 + \mu y_2 \leftrightarrow \mathcal{B}(\lambda x_2 + \lambda y_2) = \lambda x_3 + \mu y_3,$$

因此必有

$$\lambda x_1 + \mu y_1 \leftrightarrow \lambda x_3 + \mu y_3 = \sigma(\lambda x_1 + \mu y_1) = \lambda \sigma(x_1) + \mu \sigma(y_1).$$

这就表明 σ 是一对一的线性算子, 故 V_1 与 V_3 是同构的.

性质 2 同构的线性空间中的零向量必定是互相对应的.

事实上, 若 $x_1 (\in V_1) \leftrightarrow x_2 (\in V_2)$, 则 $0x_1 \leftrightarrow 0x_2$, 即 $0_{V_1} \leftrightarrow 0_{V_2}$.

性质 3 同构的线性空间中的线性相关向量组对应于线性相关向量组, 线性无关向量组对应于线性无关向量组.

事实上, 如果设 \mathcal{A} 是由 V 到 V' 的同构算子, 那么只需证明: V 中向量组 x_1, x_2, \dots, x_m 线性相关当且仅当 $\mathcal{A}(x_1), \mathcal{A}(x_2), \dots, \mathcal{A}(x_m)$ 线性相关.

先证必要性. 由 x_1, x_2, \dots, x_m 线性相关, 则必有不全为零的数 k_1, k_2, \dots, k_m , 使 $k_1 x_1 + k_2 x_2 + \dots + k_m x_m = \mathbf{0}$, 于是得

$$k_1 \mathcal{A}(x_1) + k_2 \mathcal{A}(x_2) + \dots + k_m \mathcal{A}(x_m) = \mathbf{0},$$

所以 $\mathcal{A}(x_1), \mathcal{A}(x_2), \dots, \mathcal{A}(x_m)$ 也线性相关.

再证充分性. 由 $\mathcal{A}(x_1), \mathcal{A}(x_2), \dots, \mathcal{A}(x_m)$ 线性相关, 则必有不全为零的数 k_1, k_2, \dots, k_m , 使

$$k_1 \mathcal{A}(x_1) + k_2 \mathcal{A}(x_2) + \dots + k_m \mathcal{A}(x_m) = \mathbf{0},$$

即

$$\mathcal{A}(k_1 x_1 + k_2 x_2 + \dots + k_m x_m) = \mathbf{0}.$$

因为 \mathcal{A} 是一对一的线性算子, 只有 $\mathcal{A}(\mathbf{0}) = \mathbf{0}$, 故

$$k_1 x_1 + k_2 x_2 + \dots + k_m x_m = \mathbf{0},$$

即 x_1, x_2, \dots, x_m 线性相关, 从而得证.

更重要的, 有下面的定理.

定理 1.1.9 数域 P 上两个有限维线性空间同构的充要条件是两空间的维数相等.

证明 必要性. 因为线性空间的维数就是线性无关向量最多的个数, 所以由同构线性空间的性质 3 即可推知, 同构的线性空间有相同的维数.

再证充分性. 设 V_1^n 与 V_2^n 是数域 P 上两个 n 维的线性空间, 在 V_1^n 中取一组基 e_1, e_2, \dots, e_n , 在 V_2^n 中也取一组基 e'_1, e'_2, \dots, e'_n . 现在定义由 V_1^n 到 V_2^n 的算子如下:

$$\mathcal{A}(k_1 e_1 + k_2 e_2 + \dots + k_n e_n) = k_1 e'_1 + k_2 e'_2 + \dots + k_n e'_n.$$

显然, \mathcal{A} 是一对一的算子, 而且是线性算子, 即满足

$$\mathcal{A}(x + y) = \mathcal{A}(x) + \mathcal{A}(y),$$

$$\mathcal{A}(\lambda x) = \lambda \mathcal{A}(x),$$

故 V_1^n 与 V_2^n 同构. 证毕

推论 数域 P 上的任何 n 维线性空间 V^n 都与特殊的线性空间 $K^n = \{(a_1, a_2, \dots, a_n) \mid a_i \in P\}$ 同构.

因此, 引入同构的概念给研究抽象的线性空间 V^n 带来极大的便利, 即使 V^n 可以代表不同的线性空间, 其元素可能完全不同, 但利用同构关系, 都可能将 V^n 中的问题通过基转化到向量空间 K^n 中的问题加以研究.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}. \quad (1.1.15)$$

注意,基像 $\mathcal{A}(e_i)$ 的坐标恰是矩阵 \mathbf{A} 的第 i 列 ($i=1,2,\dots,n$),因而 \mathbf{A} 的行数等于 V^m 的维数,而 \mathbf{A} 的列数等于 V^n 的维数,因此 \mathbf{A} 通常记为 $\mathbf{A}_{m \times n}$. 由于表达式(1.1.13)是惟一的,所以此时对应的矩阵 $\mathbf{A}_{m \times n}$ 也惟一.

定义 1.1.16 式(1.1.14)中的矩阵 \mathbf{A} 称为线性算子 \mathcal{A} 在基偶 $\{e_1, e_2, \dots, e_n\}$ 与 $\{e'_1, e'_2, \dots, e'_m\}$ 下的矩阵表示.

反过来,给出了 $m \times n$ 矩阵(1.1.15),由式(1.1.13)或式(1.1.14)就能得到 n 个向量.不妨令 $y_i = a_{1i}e'_1 + a_{2i}e'_2 + \cdots + a_{mi}e'_m$ ($i=1,2,\dots,n$).下面证明,以这 n 个向量为一组基像的线性算子有且只有一个.

定理 1.1.11 若 e_1, e_2, \dots, e_n 是 n 维线性空间 V^n 的一组基,而 y_1, y_2, \dots, y_n 是 m 维线性空间 V^m 中任意 n 个向量,则存在惟一一个线性算子 \mathcal{A} ,把 e_1, e_2, \dots, e_n 分别映射为 y_1, y_2, \dots, y_n ,即

$$y_i = \mathcal{A}(e_i), \quad i = 1, 2, \dots, n. \quad (1.1.16)$$

证明 在 V^n 中任取向量 $x = \sum_{i=1}^n k_i e_i$, 定义算子 \mathcal{A} :

$$x = \sum_{i=1}^n k_i e_i \rightarrow \sum_{i=1}^n k_i y_i = y.$$

显然,这样定义的算子 \mathcal{A} 满足式(1.1.16),这是因为

$$\begin{aligned} e_i &= 0e_1 + \cdots + 0e_{i-1} + 1e_i + 0e_{i+1} + \cdots + 0e_n, \\ \mathcal{A}(e_i) &= 0y_1 + \cdots + 0y_{i-1} + 1y_i + 0y_{i+1} + \cdots + 0y_n = y_i, \end{aligned}$$

故有

$$\mathcal{A}(e_i) = y_i, \quad i = 1, 2, \dots, n.$$

还可证明, \mathcal{A} 是 V^n 到 V^m 的一个线性算子,事实上,

$$\begin{aligned} \mathcal{A}(\lambda x + \mu x') &= \mathcal{A}\left(\lambda \sum_{i=1}^n k_i e_i + \mu \sum_{i=1}^n k'_i e_i\right) = \mathcal{A}\left(\sum_{i=1}^n (\lambda k_i + \mu k'_i) e_i\right) \\ &= \sum_{i=1}^n (\lambda k_i + \mu k'_i) y_i = \sum_{i=1}^n \lambda k_i y_i + \sum_{i=1}^n \mu k'_i y_i \\ &= \lambda y + \mu y' = \lambda \mathcal{A}(x) + \mu \mathcal{A}(x'). \end{aligned}$$

再由定理 1.1.10 知,由 V^n 到 V^m 的线性算子是由基像 $\mathcal{A}(e_1), \mathcal{A}(e_2), \dots, \mathcal{A}(e_n)$ 惟一确定的.

证毕

4. 线性算子的运算

设 V_1, V_2, V_3 是数域 P 上的线性空间,把 V_1 到 V_2 的所有线性算子组成的集合记为 $\mathcal{D}(V_1, V_2)$. 类似地, $\mathcal{D}(V_2, V_3)$ 和 $\mathcal{D}(V_1, V_3)$ 分别表示 V_2 到 V_3 和 V_1 到 V_3 的所有线性算子组成的集合.

定义 1.1.17 设 $\mathcal{A}, \mathcal{B} \in \mathcal{D}(V_1, V_2)$, 如果有

$$(\mathcal{A} + \mathcal{B})(\mathbf{x}) = \mathcal{A}(\mathbf{x}) + \mathcal{B}(\mathbf{x}), \quad \forall \mathbf{x} \in V_1,$$

则称 $\mathcal{A} + \mathcal{B}$ 为 \mathcal{A} 与 \mathcal{B} 的和; 又设 $\mathcal{A} \in \mathcal{D}(V_1, V_2), \mathcal{B} \in \mathcal{D}(V_2, V_3)$, 如果有

$$(\mathcal{B}\mathcal{A})(\mathbf{x}) = \mathcal{B}(\mathcal{A}(\mathbf{x})), \quad \forall \mathbf{x} \in V_1, \quad (1.1.17)$$

则称 $\mathcal{B}\mathcal{A}$ 为 \mathcal{A} 与 \mathcal{B} 的乘积, 显然它是 V_1 到 V_3 的算子.

下面定理将证明 $\mathcal{A} + \mathcal{B}$ 与 $\mathcal{B}\mathcal{A}$ 均为线性算子.

定理 1.1.12 (1) 设 $\mathcal{A}, \mathcal{B} \in \mathcal{D}(V_1, V_2)$, 则 $\mathcal{A} + \mathcal{B} \in \mathcal{D}(V_1, V_2)$;

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

证明 (1) 对 $\forall \mathbf{x}, \mathbf{y} \in V_1, \forall k \in P$, 有

$$\begin{aligned} (\mathcal{A} + \mathcal{B})(\mathbf{x} + \mathbf{y}) &= \mathcal{A}(\mathbf{x} + \mathbf{y}) + \mathcal{B}(\mathbf{x} + \mathbf{y}) \\ &= \mathcal{A}(\mathbf{x}) + \mathcal{A}(\mathbf{y}) + \mathcal{B}(\mathbf{x}) + \mathcal{B}(\mathbf{y}) \\ &= (\mathcal{A}(\mathbf{x}) + \mathcal{B}(\mathbf{x})) + (\mathcal{A}(\mathbf{y}) + \mathcal{B}(\mathbf{y})) \\ &= (\mathcal{A} + \mathcal{B})(\mathbf{x}) + (\mathcal{A} + \mathcal{B})(\mathbf{y}), \\ (\mathcal{A} + \mathcal{B})(k\mathbf{x}) &= \mathcal{A}(k\mathbf{x}) + \mathcal{B}(k\mathbf{x}) \\ &= k\mathcal{A}(\mathbf{x}) + k\mathcal{B}(\mathbf{x}) = k(\mathcal{A}(\mathbf{x}) + \mathcal{B}(\mathbf{x})) \\ &= k(\mathcal{A} + \mathcal{B})(\mathbf{x}), \end{aligned}$$

因此, $\mathcal{A} + \mathcal{B}$ 是 V_1 到 V_2 的线性算子.

(2) 对 $\forall \mathbf{x}, \mathbf{y} \in V_1, \forall k \in P$, 有

$$\begin{aligned} (\mathcal{B}\mathcal{A})(\mathbf{x} + \mathbf{y}) &= \mathcal{B}(\mathcal{A}(\mathbf{x} + \mathbf{y})) = \mathcal{B}(\mathcal{A}(\mathbf{x}) + \mathcal{A}(\mathbf{y})) \\ &= \mathcal{B}(\mathcal{A}(\mathbf{x})) + \mathcal{B}(\mathcal{A}(\mathbf{y})) = \mathcal{B}\mathcal{A}(\mathbf{x}) + \mathcal{B}\mathcal{A}(\mathbf{y}), \\ (\mathcal{B}\mathcal{A})(k\mathbf{x}) &= \mathcal{B}(\mathcal{A}(k\mathbf{x})) = \mathcal{B}(k\mathcal{A}(\mathbf{x})) \\ &= k(\mathcal{B}(\mathcal{A}(\mathbf{x}))) = k(\mathcal{B}\mathcal{A})(\mathbf{x}), \end{aligned}$$

则 $\mathcal{B}\mathcal{A}$ 是 V_1 到 V_3 的线性算子. 证毕

容易验证 $\mathcal{A} + \mathcal{B}$ 满足交换律和结合律. 但是, $\mathcal{B}\mathcal{A}$ 通常不满足交换律, 即 $\mathcal{B}\mathcal{A} \neq \mathcal{A}\mathcal{B}$. 事实上, 当 $V_1 \neq V_3$ 时, $\mathcal{A}\mathcal{B}$ 未必有意义. 即使 $V_1 = V_2 = V_3$, $\mathcal{B}\mathcal{A}$ 与 $\mathcal{A}\mathcal{B}$ 都有意义, 但 $\mathcal{B}\mathcal{A}$ 与 $\mathcal{A}\mathcal{B}$ 也未必相等.

零算子 \mathcal{O} 具有性质

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

并且对每一个 $\mathcal{A} \in \mathcal{D}(V_1, V_2)$, 它的负算子 $-\mathcal{A} \in \mathcal{D}(V_1, V_2)$ 满足

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

设 $\mathcal{A}, \mathcal{B} \in \mathcal{D}(V_1, V_2)$, 线性算子的减法定义为

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

利用线性算子的乘法以及数乘算子 \mathcal{H} 可以定义线性算子的数量乘法. 设 $\mathcal{A} \in \mathcal{D}(V_1, V_2), k \in P$, 定义 k 与 \mathcal{A} 的数量乘积 $k\mathcal{A}$ 为

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

即

$$(k\mathcal{A})(\mathbf{x}) = \mathcal{H}(\mathcal{A}(\mathbf{x})) = k\mathcal{A}(\mathbf{x}), \quad \forall \mathbf{x} \in V_1.$$

由定理 1.1.12 知, $k\mathcal{A} \in \mathcal{D}(V_1, V_2)$.

对于线性算子,前面已经定义了加法、乘法和数量乘法 3 种运算. 如果由 V_1 到 V_2 的线性算子以及由 V_2 到 V_3 的线性算子在基偶给定之后,它们和矩阵便可建立一一对应的关系. 此时,以上的 3 种线性算子的运算与矩阵的相应的 3 种运算也一一对应,即

(1) 当 $\mathcal{A} \leftrightarrow \mathbf{A}, \mathcal{B} \leftrightarrow \mathbf{B}$ 时,有 $\mathcal{A} + \mathcal{B} \leftrightarrow \mathbf{A} + \mathbf{B}$;

(2) 当 $\mathcal{A} \leftrightarrow \mathbf{A}, \mathcal{B} \leftrightarrow \mathbf{B}$ 时,有 $\mathcal{B}\mathcal{A} \leftrightarrow \mathbf{B}\mathbf{A}$;

(3) 当 $\mathcal{A} \leftrightarrow \mathbf{A}, \forall k \in P$,有 $k\mathcal{A} \leftrightarrow k\mathbf{A}$.

进一步还可以看出,从 n 维线性空间 V^n 到 m 维线性空间 V^m 的所有线性算子,按线性算子的加法和数乘运算规则,形成数域 P 上的一个线性空间,而 $m \times n$ 矩阵的全体也构成数域 P 上的一个线性空间,因此这两个线性空间关于加法和数乘运算是同构的,其中零矩阵对应零算子. 从而对线性算子的研究常常转化为对矩阵的研究.

有时,我们常常利用线性算子的矩阵 \mathbf{A} ,来计算一个向量 \mathbf{x} 的像 $\mathcal{A}(\mathbf{x})$ 的坐标.

设 V^n 到 V^m 的线性算子 \mathcal{A} 在基偶 $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ 与 $\mathbf{e}'_1, \mathbf{e}'_2, \dots, \mathbf{e}'_m$ 下的矩阵为 $\mathbf{A} = (a_{ij})_{m \times n}$, 向量 $\mathbf{x} \in V^n$ 在基 $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ 下的坐标是 (x_1, x_2, \dots, x_n) , 则 $\mathcal{A}(\mathbf{x})$ 在基 $\mathbf{e}'_1, \mathbf{e}'_2, \dots, \mathbf{e}'_m$ 下的坐标 (y_1, y_2, \dots, y_m) , 可按公式

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \mathbf{A} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad (1.1.18)$$

来计算.

事实上,由假设有

$$\mathbf{x} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

则

$$\mathcal{A}(\mathbf{x}) = \mathcal{A}(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = (\mathbf{e}'_1, \mathbf{e}'_2, \dots, \mathbf{e}'_m) \mathbf{A} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

另一方面,有

$$\mathcal{A}(\mathbf{x}) = (\mathbf{e}'_1, \mathbf{e}'_2, \dots, \mathbf{e}'_m) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}.$$

由于 $\mathbf{e}'_1, \mathbf{e}'_2, \dots, \mathbf{e}'_m$ 线性无关(且是基),根据 $\mathcal{A}(\mathbf{x})$ 坐标的惟一性,可得

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \mathbf{A} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

若记 $\mathbf{X}=(x_1, x_2, \dots, x_n)^T, \mathbf{Y}=(y_1, y_2, \dots, y_m)^T$, 则式(1.1.18)可记为

$$\mathbf{Y} = \mathbf{A}\mathbf{X}. \quad (1.1.19)$$

在线性代数中, 矩阵是表示线性方程组的一种简便形式 $\mathbf{A}\mathbf{X}=\mathbf{b}$, 其中 $\mathbf{b}=(b_1, b_2, \dots, b_m)^T$ 其解(集)有非常直观的几何意义, 即从线性算子的角度来理解, 求它的解(集)可以看成是在 \mathbb{R}^n 到 \mathbb{R}^m 的线性算子 $\mathcal{A}: \mathbf{X} \mapsto \mathbf{A}\mathbf{X}$ 下, 向量 $\mathbf{b} \in \mathbb{R}^m$ 的原像。特别地, 齐次线性方程组 $\mathbf{A}\mathbf{X}=\mathbf{0}$ 的解(集)恰好是线性算子 $\mathcal{A}: \mathbf{X} \mapsto \mathbf{A}\mathbf{X}$ 为“零点”的原像。

5. 线性变换与方阵

线性变换的定义

在线性算子的定义中, 线性空间 V' 与 V 可以相同也可以不同. 在 $V'=V$ 的情况下, 由 V 到 V' 的线性算子可用一个特殊名称来命名, 由此给出下面的定义。

定义 1.1.18 由 V 到 V 的线性算子 \mathcal{A} 叫做 V 上的**线性变换**。

换言之, 线性变换是线性空间 V 到自身的线性算子, 它是线性算子中最简单、最基本的一种. 此时, 与之相应的矩阵是方阵。

定义 1.1.19 如果对于任何 $x \in V$, 恒有 $\mathcal{A}(x)=x$, 则称 \mathcal{A} 为**恒等变换**或**单位变换**, 记为 \mathcal{I} . 与之相应的矩阵是单位矩阵 \mathbf{I} 或 \mathbf{E} 。

对于线性空间 V 中的两个线性变换 \mathcal{A}, \mathcal{B} , 可定义它们的和 $\mathcal{A}+\mathcal{B}$ 如下:

$$(\mathcal{A}+\mathcal{B})\xi \equiv \mathcal{A}(\xi) + \mathcal{B}(\xi), \quad \xi \in V.$$

不难验证 $\mathcal{A}+\mathcal{B}$ 也是 V 中的线性变换. 对线性变换 \mathcal{A}, \mathcal{B} 的乘积 $\mathcal{A}\mathcal{B}$, 也可证明它是 V 中的线性变换, 请读者自己证明。

线性变换的加法和乘法满足如下的运算规律:

- (1) 加法满足交换律、结合律;
- (2) 乘法满足结合律;
- (3) 乘法对加法有左、右分配律。

对于加法, 零变换 \mathcal{O} 有着特殊的地位, 它与任意线性变换 \mathcal{A} 的和仍等于 \mathcal{A} : $\mathcal{A}+\mathcal{O}=\mathcal{A}$. 对于每个线性变换 \mathcal{A} , 其负变换 $-\mathcal{A}$ 也是线性的, 且

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

对于乘法, 单位变换 \mathcal{I} 有着特殊的地位, 它与任意线性变换 \mathcal{A} 的乘积仍等于 \mathcal{A} :

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

数域 P 中每个数 k 都决定一个数乘变换 \mathcal{K} , 利用线性变换的乘法, 可以定义数域 P 中的数与线性变换的数量乘法

$$k\mathcal{A} \equiv \mathcal{K}\mathcal{A}.$$

线性空间 V 上的线性变换 \mathcal{A} 称为可逆的, 如果有 V 上的线性变换 \mathcal{B} 存在, 使得

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

称 \mathcal{B} 为 \mathcal{A} 的**逆变换**, 记作 \mathcal{A}^{-1} 。

现在来证明: 如果线性变换 \mathcal{A} 是可逆的, 那么它的逆变换 \mathcal{A}^{-1} 也是线性变换. 事实上, 对任意向量 $\alpha, \beta \in V$, 数 $k \in P$, 有

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

$$\begin{cases} \mathcal{A}(\boldsymbol{\varepsilon}_1) = -\frac{5}{7}\boldsymbol{\varepsilon}_1 - \frac{4}{7}\boldsymbol{\varepsilon}_2 + \frac{27}{7}\boldsymbol{\varepsilon}_3, \\ \mathcal{A}(\boldsymbol{\varepsilon}_2) = \frac{20}{7}\boldsymbol{\varepsilon}_1 - \frac{5}{7}\boldsymbol{\varepsilon}_2 + \frac{18}{7}\boldsymbol{\varepsilon}_3, \\ \mathcal{A}(\boldsymbol{\varepsilon}_3) = -\frac{20}{7}\boldsymbol{\varepsilon}_1 - \frac{2}{7}\boldsymbol{\varepsilon}_2 + \frac{24}{7}\boldsymbol{\varepsilon}_3, \end{cases}$$

从而,有

$$\mathbf{B} = \begin{pmatrix} -\frac{5}{7} & \frac{20}{7} & -\frac{20}{7} \\ -\frac{4}{7} & -\frac{5}{7} & -\frac{2}{7} \\ \frac{27}{7} & \frac{18}{7} & \frac{24}{7} \end{pmatrix}.$$

在 1.1.1 节已经知道,同一向量在不同基下的坐标往往不同.类似地,例 1.1.25 和例 1.1.26 又表明同一个线性变换在不同基下的矩阵也不同.那么,同一个线性变换在不同基下的矩阵之间究竟有什么关系呢?这就是下面要在几何上来解释矩阵相似的问题.

6. 相似矩阵的几何解释

定理 1.1.13 假定线性空间 V^n 上的线性变换 \mathcal{A} ,对于基 $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n$ 下的矩阵为 \mathbf{A} ,而对于另一组基 $\boldsymbol{e}'_1, \boldsymbol{e}'_2, \dots, \boldsymbol{e}'_n$ 下的矩阵为 \mathbf{B} ,且由基 $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n$ 到基 $\boldsymbol{e}'_1, \boldsymbol{e}'_2, \dots, \boldsymbol{e}'_n$ 的过渡矩阵为 \mathbf{C} ,则有

$$\mathbf{B} = \mathbf{C}^{-1}\mathbf{A}\mathbf{C}. \quad (1.1.20)$$

证明 根据定理假设有

$$\begin{aligned} \mathcal{A}(\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n) &= (\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n)\mathbf{A}, \\ \mathcal{A}(\boldsymbol{e}'_1, \boldsymbol{e}'_2, \dots, \boldsymbol{e}'_n) &= (\boldsymbol{e}'_1, \boldsymbol{e}'_2, \dots, \boldsymbol{e}'_n)\mathbf{B}, \\ (\boldsymbol{e}'_1, \boldsymbol{e}'_2, \dots, \boldsymbol{e}'_n) &= (\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n)\mathbf{C}. \end{aligned}$$

故得

$$\begin{aligned} \mathcal{A}(\boldsymbol{e}'_1, \boldsymbol{e}'_2, \dots, \boldsymbol{e}'_n) &= \mathcal{A}(\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n)\mathbf{C} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n)\mathbf{A}\mathbf{C}, \\ \mathcal{A}(\boldsymbol{e}'_1, \boldsymbol{e}'_2, \dots, \boldsymbol{e}'_n) &= (\boldsymbol{e}'_1, \boldsymbol{e}'_2, \dots, \boldsymbol{e}'_n)\mathbf{B} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n)\mathbf{C}\mathbf{B}. \end{aligned}$$

因此

$$(\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n)\mathbf{A}\mathbf{C} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n)\mathbf{C}\mathbf{B},$$

即

$$(\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n)(\mathbf{A}\mathbf{C} - \mathbf{C}\mathbf{B}) = \mathbf{0}.$$

由于 $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n$ 线性无关,所以矩阵 $\mathbf{A}\mathbf{C} - \mathbf{C}\mathbf{B}$ 的每一列的元素都为零,故有

$$\mathbf{A}\mathbf{C} - \mathbf{C}\mathbf{B} = \mathbf{0} \text{ (零矩阵)}.$$

又因过渡矩阵 \mathbf{C} 非奇异,从而有

$$\mathbf{B} = \mathbf{C}^{-1}\mathbf{A}\mathbf{C}.$$

证毕

式(1.1.20)给出的两个 \mathbf{A} 与 \mathbf{B} 之间的关系,在矩阵论中将起极其重要的作用,对此我们引入下面的定义.

定义 1.1.20 如果 A 与 B 是数域 P 上的两个 n 阶矩阵,且可找到 P 上的 n 阶非奇异矩阵 C ,使得 $B=C^{-1}AC$,则称 A 与 B 相似,记为 $A\sim B$.

按此定义,我们可以说,线性变换在不同基下的矩阵是相似的;反之,如果两个矩阵相似,那么它们可以看成同一个线性变换在两组不同基下的矩阵.

运用上述相似原理,当已知线性变换在某一组基下的矩阵,而要写出它在另一组基下的矩阵时,可以使问题大为简化.例如,在例 1.1.26 中,由(1)已求出线性变换 \mathcal{A} 在基 e_1, e_2, e_3 下的矩阵

$$A = \begin{pmatrix} 2 & 3 & 5 \\ -1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}.$$

如果要写出 \mathcal{A} 在自然基 $\varepsilon_1, \varepsilon_2, \varepsilon_3$ 下的矩阵,则只要算出由基 e_1, e_2, e_3 到基 $\varepsilon_1, \varepsilon_2, \varepsilon_3$ 的过渡矩阵 C .为此,容易看出

$$(e_1, e_2, e_3) = (\varepsilon_1, \varepsilon_2, \varepsilon_3) \begin{pmatrix} -1 & 0 & 3 \\ 0 & 1 & -1 \\ 2 & 1 & 0 \end{pmatrix},$$

所以

$$(\varepsilon_1, \varepsilon_2, \varepsilon_3) = (e_1, e_2, e_3) \begin{pmatrix} -1 & 0 & 3 \\ 0 & 1 & -1 \\ 2 & 1 & 0 \end{pmatrix}^{-1},$$

故

$$C = \begin{pmatrix} -1 & 0 & 3 \\ 0 & 1 & -1 \\ 2 & 1 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} -\frac{1}{7} & -\frac{3}{7} & \frac{3}{7} \\ \frac{2}{7} & \frac{6}{7} & \frac{1}{7} \\ \frac{2}{7} & -\frac{1}{7} & \frac{1}{7} \end{pmatrix}.$$

从而 \mathcal{A} 在自然基下的矩阵为

$$\begin{aligned} B = C^{-1}AC &= \begin{pmatrix} -1 & 0 & 3 \\ 0 & 1 & -1 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 3 & 5 \\ -1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} -\frac{1}{7} & -\frac{3}{7} & \frac{3}{7} \\ \frac{2}{7} & \frac{6}{7} & \frac{1}{7} \\ \frac{2}{7} & -\frac{1}{7} & \frac{1}{7} \end{pmatrix} \\ &= \begin{pmatrix} -\frac{5}{7} & \frac{20}{7} & -\frac{20}{7} \\ -\frac{4}{7} & -\frac{5}{7} & -\frac{2}{7} \\ \frac{27}{7} & \frac{18}{7} & \frac{24}{7} \end{pmatrix}. \end{aligned}$$

该结果与例 1.1.26 完全一致.但这里利用相似原理的方法显然比直接用定义的方法更简单、直观.

例 1.1.27 设线性变换 \mathcal{A} 在基 e_1, e_2, e_3, e_4 下的矩阵为

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 2 & 8 \\ 5 & 7 & 0 & 1 \\ 3 & 0 & 1 & 3 \\ 1 & 1 & 2 & 2 \end{pmatrix},$$

试求 \mathcal{A} 在基 $e'_1 = e_1, e'_2 = e_1 + e_2, e'_3 = e_1 + e_2 + e_3, e'_4 = e_1 + e_2 + e_3 + e_4$ 下的矩阵.

解 方法一: 由矩阵 \mathbf{A} 的定义知

$$\begin{cases} \mathcal{A}(e_1) = e_1 + 5e_2 + 3e_3 + e_4, \\ \mathcal{A}(e_2) = 3e_1 + 7e_2 + e_4, \\ \mathcal{A}(e_3) = 2e_1 + e_3 + 2e_4, \\ \mathcal{A}(e_4) = 8e_1 + e_2 + 3e_3 + 2e_4, \end{cases}$$

于是

$$\mathcal{A}(e'_1) = \mathcal{A}(e_1) = e_1 + 5e_2 + 3e_3 + e_4 = -4e'_1 + 2e'_2 + 2e'_3 + e'_4,$$

$$\mathcal{A}(e'_2) = \mathcal{A}(e_1 + e_2) = 4e_1 + 12e_2 + 3e_3 + 2e_4$$

$$= -8e'_1 + 9e'_2 + e'_3 + 2e'_4,$$

$$\mathcal{A}(e'_3) = \mathcal{A}(e_1 + e_2 + e_3) = 6e_1 + 12e_2 + 4e_3 + 4e_4$$

$$= -6e'_1 + 8e'_2 + 4e'_4,$$

$$\mathcal{A}(e'_4) = \mathcal{A}(e_1 + e_2 + e_3 + e_4) = 14e_1 + 13e_2 + 7e_3 + 6e_4$$

$$= e'_1 + 6e'_2 + e'_3 + 6e'_4,$$

所以在新基 e'_1, e'_2, e'_3, e'_4 下的矩阵为

$$\mathbf{B} = \begin{pmatrix} -4 & -8 & -6 & 1 \\ 2 & 9 & 8 & 6 \\ 2 & 1 & 0 & 1 \\ 1 & 2 & 4 & 6 \end{pmatrix}.$$

方法二: 可以利用相似原理求 \mathbf{B} . 由基 $\{e_i\}$ 到基 $\{e'_i\}$ 的过渡矩阵为

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

所以

$$\begin{aligned} \mathbf{B} = \mathbf{C}^{-1}\mathbf{A}\mathbf{C} &= \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 2 & 8 \\ 5 & 7 & 0 & 1 \\ 3 & 0 & 1 & 3 \\ 1 & 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -4 & -8 & -6 & 1 \\ 2 & 9 & 8 & 6 \\ 2 & 1 & 0 & 1 \\ 1 & 2 & 4 & 6 \end{pmatrix}. \end{aligned}$$

比相似更一般的概念是相抵.

定义 1.1.21 设 A 与 B 都是 $m \times n$ 阶矩阵, 如果存在非奇异的 m 阶方阵 D 和 n 阶方阵 C , 使

$$B = DAC \quad (1.1.21)$$

成立, 则称矩阵 A 与 B 是相抵的, 记为 $A \simeq B$.

相抵关系在几何上可解释为: 在两个不同维的线性空间 V^n 和 V^m 中, 同一个线性算子 \mathcal{A} 在不同的基偶下所对应的矩阵 A 与 B 之间的关系.

例如, 线性代数中对矩阵 A 进行初等行(列)变换, 就相当于在 A 的左(右)边乘上一个非奇异的初等运算矩阵, 那么变换后的矩阵和原矩阵 A 便是相抵关系.

特别地, 当 A, B 是方阵, 而且 $D = C^T$ 时, 又有如下定义.

定义 1.1.22 设 A 与 B 是两个 n 阶方阵, 如果存在非奇异的 n 阶方阵 C , 使得

$$B = C^T A C, \quad (1.1.22)$$

则称矩阵 A 与 B 是相合(或合同)的.

例如, 线性代数中用非退化的坐标线性变换化简一个二次型时, 它们所对应的矩阵就是这种相合的关系.

总之, 相抵、相似、相合反映了两矩阵之间的三种内在联系, 这三种关系是既有区别又有联系的, 相似与相合只不过是相抵的特殊情况, 而且相似与相合只有在 $C^T = C^{-1}$ 时(即 C 为正交阵)才一致.

如前所述, 在取定一组基之后, 就建立了由数域 P 上的 n 维线性空间 V 的线性变换到数域 P 上 n 阶矩阵之间的一个一一对应. 这种对应的重要性还表现在它们保持有相应的运算, 即有下面定理:

定理 1.1.14 设 $\xi_1, \xi_2, \dots, \xi_n$ 是数域 P 上 n 维线性空间 V 的一组基, 在这组基下, 每个线性变换对应一个 n 阶矩阵. 这个对应具有以下性质:

- (1) 线性变换的和对应于矩阵的和;
- (2) 线性变换的乘积对应于矩阵的乘积;
- (3) 线性变换与数的积对应于矩阵与数的积;
- (4) 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 以(2), (4)为例, 其余可类似地证明. 设 \mathcal{A}, \mathcal{B} 是两个线性变换, 它们在基 $\xi_1, \xi_2, \dots, \xi_n$ 下的矩阵分别为 A, B , 即

$$\begin{cases} \mathcal{A}(\xi_1, \xi_2, \dots, \xi_n) = (\xi_1, \xi_2, \dots, \xi_n)A, \\ \mathcal{B}(\xi_1, \xi_2, \dots, \xi_n) = (\xi_1, \xi_2, \dots, \xi_n)B. \end{cases}$$

证明(2). 由

$$\begin{aligned} (\mathcal{A}\mathcal{B})(\xi_1, \xi_2, \dots, \xi_n) &= \mathcal{A}[\mathcal{B}(\xi_1, \xi_2, \dots, \xi_n)] = \mathcal{A}[(\xi_1, \xi_2, \dots, \xi_n)B] \\ &= [\mathcal{A}(\xi_1, \xi_2, \dots, \xi_n)]B = (\xi_1, \xi_2, \dots, \xi_n)AB \end{aligned}$$

可知, 在基 $\xi_1, \xi_2, \dots, \xi_n$ 下, 线性变换 $\mathcal{A}\mathcal{B}$ 的矩阵为 AB .

再证明(4). 因为单位变换对应于单位矩阵, 故等式

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

与等式

$$AB = BA = I$$

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.

证毕

7. 矩阵的逆方阵和秩

现在详细考察 n 阶方阵

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}.$$

定义 1.1.23 如果 $\det(\mathbf{A}) \neq 0$ (即 $|\mathbf{A}| \neq 0$), 则称 n 阶方阵 \mathbf{A} 为非奇异的.

定义 1.1.24 如果存在 n 阶方阵 \mathbf{B} , 使得

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n \quad (1.1.23)$$

其中 \mathbf{I}_n 为 n 阶单位方阵, 则称 n 阶方阵 \mathbf{A} 为可逆的, 而 \mathbf{B} 称为可逆方阵 \mathbf{A} 的逆方阵.

引理 (Binet-Cauchy 公式) 设 \mathbf{A} 和 \mathbf{B} 各为 $n \times m$ 和 $m \times n$ 矩阵, 则有

$$\det(\mathbf{AB}) = \begin{cases} 0, & \text{当 } n > m \\ \det(\mathbf{A})\det(\mathbf{B}), & \text{当 } n = m \\ \sum_{1 \leq j_1 < \cdots < j_n \leq m} \det \mathbf{A} \begin{pmatrix} 1 & 2 & \cdots & n \\ j_1 & j_2 & \cdots & j_n \end{pmatrix} \det \mathbf{B} \begin{pmatrix} j_1 & j_2 & \cdots & j_n \\ 1 & 2 & \cdots & n \end{pmatrix}, & \text{当 } n < m \end{cases}.$$

证明从略(见文献[6]).

定理 1.1.15 n 阶方阵 \mathbf{A} 是非奇异的, 当且仅当 \mathbf{A} 是可逆的, 且当 \mathbf{A} 可逆时, 它的逆方阵惟一存在, 记作 \mathbf{A}^{-1} , 且有

$$\mathbf{A}^{-1} = (\det(\mathbf{A}))^{-1} \mathbf{A}^*, \quad (1.1.24)$$

其中 \mathbf{A}^* 是 \mathbf{A} 的伴随方阵.

先证必要性. 今设 \mathbf{A} 为非奇异方阵, 即 $\det(\mathbf{A}) \neq 0$, 由式(1.1.23), 令

$$\mathbf{B} = (\det(\mathbf{A}))^{-1} \mathbf{A}^*,$$

则有

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n.$$

这证明了 \mathbf{A} 为可逆方阵, 反之, 若存在方阵 \mathbf{B} 使得 $\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n$, 由引理(Binet-Cauchy 公式), 两边取行列式, 得

$$1 = \det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}),$$

因此 $\det(\mathbf{A}) \neq 0$. 这证明了 \mathbf{A} 为非奇异方阵.

证毕

上面得到一个可逆方阵 \mathbf{A} 的逆方阵 $\mathbf{B} = (\det(\mathbf{A}))^{-1} \mathbf{A}^*$. 下面证明可逆方阵的逆方阵惟一存在.

事实上, 如果方阵 \mathbf{A} 另有一个逆方阵 \mathbf{B}_1 , 则由定义有 $\mathbf{B}_1 \mathbf{A} = \mathbf{A} \mathbf{B}_1 = \mathbf{I}_n$, 所以

$$\mathbf{B}_1 = \mathbf{I}_n \mathbf{B}_1 = (\mathbf{BA}) \mathbf{B}_1 = \mathbf{B}(\mathbf{AB}_1) = \mathbf{B} \mathbf{I}_n = \mathbf{B},$$

这就证明了惟一性.

定理 1.1.15 证明了“非奇异”和“可逆”这两个概念是等价的. 关于逆方阵还有下列简单的性质:

$$(1) (\mathbf{A}^{-1})^{-1} = \mathbf{A};$$

$$(2) (\mathbf{AB})^{-1} = \mathbf{B}^{-1} \mathbf{A}^{-1};$$

$$(3) (c\mathbf{A})^{-1} = c^{-1}\mathbf{A}^{-1}, \forall c \in P, c \neq 0;$$

$$(4) (\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T;$$

$$(5) (\bar{\mathbf{A}})^{-1} = \bar{\mathbf{A}}^{-1}, \text{由(4)和(5), 有 } (\mathbf{A}^*)^{-1} = (\mathbf{A}^{-1})^*;$$

$$(6) \det(\mathbf{A}^{-1}) = \det(\mathbf{A})^{-1} \neq 0.$$

证 只要证(2)成立, 读者试自证其他的性质成立.

利用逆方阵的惟一性, 只要证明 $\mathbf{B}^{-1}\mathbf{A}^{-1}$ 是方阵 \mathbf{AB} 的逆方阵就行了. 令

$$(\mathbf{AB})(\mathbf{B}^{-1}\mathbf{A}^{-1}) = \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1} = \mathbf{AA}^{-1} = \mathbf{I}_n,$$

又

$$(\mathbf{B}^{-1}\mathbf{A}^{-1})(\mathbf{AB}) = \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I}_n,$$

故断言成立.

证毕

由性质(2), 用数学归纳法, 很易证明: 设 n 阶方阵 $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p$ 都是可逆方阵, 则 $\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_p$ 也是可逆方阵, 且其逆为

$$(\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_p)^{-1} = \mathbf{A}_p^{-1}\mathbf{A}_{p-1}^{-1}\cdots\mathbf{A}_2^{-1}\mathbf{A}_1^{-1}.$$

由此可见, 方阵的取逆运算和转置运算, 当作用在方阵的乘积上时, 具有类似的性质.

下面引进矩阵的另一个重要概念——秩.

定义 1.1.25 数域 P 上 $n \times m$ 非零矩阵 \mathbf{A} 的所有子式中必有一个阶数最大的非零子式, 其阶数称为矩阵 \mathbf{A} 的秩, 记作 $\text{秩}(\mathbf{A})$, 或 $\text{rank}(\mathbf{A})$. 零矩阵的秩定义为零.

由于矩阵 \mathbf{A} 的子式的阶数不超过 \mathbf{A} 的行数及列数, 所以有

定理 1.1.16 数域 P 上 $n \times m$ 矩阵 \mathbf{A} 的秩满足:

$$(1) 0 \leq \text{rank}(\mathbf{A}) \leq \min(n, m);$$

$$(2) \text{rank}(c\mathbf{A}) = \text{rank}(\mathbf{A}), \forall 0 \neq c \in \mathbb{F};$$

$$(3) \text{rank}(\mathbf{A}') = \text{rank}(\mathbf{A});$$

$$(4) \text{rank}(\bar{\mathbf{A}}) = \text{rank}(\mathbf{A}), \text{rank}(\mathbf{A}^*) = \text{rank}(\mathbf{A}).$$

定义 1.1.26 $n \times m$ 矩阵 \mathbf{A} 称为满秩的, 如果

$$\text{rank}(\mathbf{A}) = \min(n, m) \quad (1.1.25)$$

定理 1.1.17 数域 P 上 n 阶方阵 \mathbf{A} 非奇异当且仅当 \mathbf{A} 可逆, 当且仅当 \mathbf{A} 满秩, 即对 n 阶方阵, 非奇异、可逆、满秩是三个互相等价的概念.

证 由秩的定义可知, n 阶方阵 \mathbf{A} 的秩为 n (即满秩) 的必要且充分条件为 $\det(\mathbf{A}) \neq 0$, 即对方阵而言, “秩为 n ” 和 “行列式不等于零” 这两概念是等价的, 由定理 1.1.15, 就证明了定理. 证毕

由上可知, n 阶方阵 \mathbf{A} 的秩小于 n 的必要且充分条件为 $\det(\mathbf{A}) = 0$. 为了区别起见, 今后称适合条件 $\det(\mathbf{A}) \neq 0$ 的方阵 \mathbf{A} 为非退化的或非奇异的, 称适合条件 $\det(\mathbf{A}) = 0$ 的方阵 \mathbf{A} 为退化的或奇异的.

下面给出秩的另一个等价定义.

定理 1.1.18 数域 P 上 $n \times m$ 矩阵 \mathbf{A} 的秩为 r 的必要且充分条件是: 在 \mathbf{A} 中存在一个 r 阶子式不等于零, 且在 $r < \min(n, m)$ 时, 矩阵 \mathbf{A} 的所有 $r+1$ 阶子式为零.

证 从秩的定义可知, 在矩阵 \mathbf{A} 的秩为 r 时, 必然存在一个 r 阶子式不为零, 且所有阶数为 $r+1$ 的子式都等于零, 这证明了必要性, 下面来证充分性, 即设在 \mathbf{A} 中存在一个 r 阶子式不等于零, 且在 $r < \min(n, m)$ 时, 矩阵 \mathbf{A} 的所有 $r+1$ 阶子式为零, 我们来证矩阵 \mathbf{A} 的所

有 $r+k$ 阶子式也都为零, $2 \leq k \leq \min(n, m)$. 事实上, 设 A 的子式 M 的阶为 $r_1 > r+1$. 将行列式 M 按照前 $r+1$ 行作 Laplace 展开, 由线性代数 Laplace 定理可知, 它是 $\binom{r_1}{r+1}$ 项的和, 每一项是 M 的一个 $r+1$ 阶子式和数域 P 中一数的乘积. 因为行列式 M 的子式也是矩阵 A 的子式, 所以行列式 M 的 Laplace 展开式中每一项都是零, 因此 $M=0$. 证毕

这个定理有一个简单的推论: 设数域 P 上矩阵 A 的秩为 r , 则在矩阵 A 中必分别存在一个 $1, 2, \dots, r$ 阶不等于零的子式. 因此, 在具体求矩阵的秩时, 可以用这样一种技巧: 即如果矩阵 A 有一个 s 阶子式不为零, 则矩阵 A 的秩必大于或等于 s . 所以, 依次地去寻找 $1, 2, \dots$ 阶不为零的子式. 如果在找到一个 r 阶不为零的子式后, 找不到 $r+1$ 阶不为零的子式, 换句话说, 所有 $r+1$ 阶子式全为零, 这时矩阵 A 的秩就是 r . 因此, 从原则上讲, 是从低阶往高阶寻找不为零的子式的.

例 1.1.28 矩阵

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

有一个不为零的 1 阶子式, 而所有 2 阶子式都为零, 所以 $\text{rank}(A)=1$.

例 1.1.29 矩阵

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

有一个 1, 2, 3 阶各不为零的子式, 所以 A 的秩为 3.

定理 1.1.19 设 A, B, C 分别为数域 P 上 $n \times m, p \times q, n \times q$ 矩阵, 则有

$$\text{rank} \begin{bmatrix} A & C \\ \mathbf{0} & B \end{bmatrix} \geq \text{rank} \begin{bmatrix} A & \mathbf{0} \\ \mathbf{0} & B \end{bmatrix} = \text{rank}(A) + \text{rank}(B). \quad (1.1.26)$$

证 设 $r = \text{rank}(A), s = \text{rank}(B)$. 于是在 A 中存在 r 阶子矩阵 A_0 , 在 B 中存在 s 阶子矩阵 B_0 , 使得 $\det(A_0) \neq 0, \det(B_0) \neq 0$. 从而在矩阵 $\begin{bmatrix} A & \mathbf{0} \\ \mathbf{0} & B \end{bmatrix}$ 中存在 $r+s$ 阶子矩阵

$\begin{bmatrix} A_0 & \mathbf{0} \\ \mathbf{0} & B_0 \end{bmatrix}$ 非奇异. 由线性代数 Laplace 展开定理可知在矩阵 $\begin{bmatrix} A & C \\ \mathbf{0} & B \end{bmatrix}$ 中存在 $r+s$ 阶子矩阵

$\begin{bmatrix} A_0 & C_0 \\ \mathbf{0} & B_0 \end{bmatrix}$ 非奇异. 这证明了

$$\text{rank} \begin{bmatrix} A & C \\ \mathbf{0} & B \end{bmatrix} \geq \text{rank}(A) + \text{rank}(B), \text{rank} \begin{bmatrix} A & \mathbf{0} \\ \mathbf{0} & B \end{bmatrix} \geq \text{rank}(A) + \text{rank}(B).$$

所以问题化为证

$$\text{rank} \begin{bmatrix} A & \mathbf{0} \\ \mathbf{0} & B \end{bmatrix} \leq \text{rank}(A) + \text{rank}(B).$$

由定理 1.1.18, 考虑矩阵 $\begin{bmatrix} A & \mathbf{0} \\ \mathbf{0} & B \end{bmatrix}$ 的任意 $r+s+1$ 阶子方阵, 它必形如 $X = \begin{bmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & B_1 \end{bmatrix}$, 其中

\mathbf{A}_1 为 $n_1 \times m_1$ 矩阵, \mathbf{B}_1 为 $p_1 \times q_1$ 矩阵, 而

$$n_1 + p_1 = m_1 + q_1 = r + s + 1.$$

下面证明 $\det(\mathbf{X}) = 0$. 用反证法. 设若不然, 即 $\det(\mathbf{X}) \neq 0$, 由 Laplace 展开式可证 $n_1 = m_1$, $p_1 = q_1$, 且 $\det(\mathbf{X}) = \det(\mathbf{A}_1)\det(\mathbf{B}_1)$. 但是由 $n_1 + p_1 = r + s + 1$ 可知 $n_1 \geq r + 1$, 或者 $p_1 \geq s + 1$. 当 $n_1 \geq r + 1$, 由 $\text{rank}(\mathbf{A}) = r$ 可知 $\det(\mathbf{A}_1) = 0$; 当 $p_1 \geq s + 1$, 由 $\text{rank}(\mathbf{B}) = s$ 可知 $\det(\mathbf{B}_1) = 0$. 总之, 证明了 $\det(\mathbf{X}) = 0$, 这导出矛盾. 证毕

利用 Binet-Cauchy 公式, 可以讨论矩阵乘积的秩.

定理 1.1.20 设 \mathbf{A} 为 $n \times m$ 矩阵, \mathbf{B} 为 $m \times p$ 矩阵, 则有

$$\text{rank}(\mathbf{AB}) \leq \min(\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})). \quad (1.1.27)$$

证 记 $\text{rank}(\mathbf{AB}) = r$, 任取 \mathbf{AB} 的 r 阶子式, 由 Binet-Cauchy 公式可知

$$\det(\mathbf{AB}) \begin{pmatrix} i_1 & i_2 & \cdots & i_r \\ j_1 & j_2 & \cdots & j_r \end{pmatrix}$$

等于

$$\sum_{1 \leq k_1 \leq k_2 \leq \cdots \leq k_r \leq m} \det \mathbf{A} \begin{pmatrix} i_1 & i_2 & \cdots & i_r \\ k_1 & k_2 & \cdots & k_r \end{pmatrix} \det \mathbf{B} \begin{pmatrix} k_1 & k_2 & \cdots & k_r \\ i_1 & i_2 & \cdots & i_r \end{pmatrix}.$$

今 $\text{rank}(\mathbf{AB}) = r$, 所以存在一个 r 阶子式不等于零, 设为 $\det(\mathbf{AB}) \begin{pmatrix} i_1 & i_2 & \cdots & i_r \\ j_1 & j_2 & \cdots & j_r \end{pmatrix} \neq 0$, 因此存在 $1 \leq k_1 \leq k_2 \leq \cdots \leq k_r \leq m$, 使得

$$\det \mathbf{A} \begin{pmatrix} i_1 & i_2 & \cdots & i_r \\ k_1 & k_2 & \cdots & k_r \end{pmatrix} \det \mathbf{B} \begin{pmatrix} k_1 & k_2 & \cdots & k_r \\ i_1 & i_2 & \cdots & i_r \end{pmatrix} \neq 0.$$

于是 $\text{rank}(\mathbf{A}) \geq r$, $\text{rank}(\mathbf{B}) \geq r$. 这证明了定理. 证毕

非奇异方阵有下面的重要性质.

定理 1.1.21 设 \mathbf{A} 为 $n \times m$ 矩阵, \mathbf{P} 为 n 阶非奇异方阵, \mathbf{Q} 为 m 阶非奇异方阵, 则有

$$\text{rank}(\mathbf{PA}) = \text{rank}(\mathbf{AQ}) = \text{rank}(\mathbf{A}). \quad (1.1.28)$$

换句话说, 将非奇异方阵左乘或右乘在矩阵上, 其秩保持不变.

证 由定理 1.1.20

$$\text{rank}(\mathbf{PA}) \leq \min(\text{rank}(\mathbf{P}), \text{rank}(\mathbf{A})) \leq \text{rank}(\mathbf{A}).$$

由此可知 $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{P}^{-1}(\mathbf{PA})) \leq \text{rank}(\mathbf{PA})$. 这证明了 $\text{rank}(\mathbf{PA}) = \text{rank}(\mathbf{A})$. 而

$$\text{rank}(\mathbf{AQ}) = \text{rank}(\mathbf{Q}'\mathbf{A}') = \text{rank}(\mathbf{A}') = \text{rank}(\mathbf{A}).$$

所以证明了定理. 证毕

定理 1.1.22 设 \mathbf{A} 为 $n \times m$ 矩阵, \mathbf{B} 为 $m \times p$ 矩阵, 则有

$$\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) \leq \text{rank}(\mathbf{AB}) + m. \quad (1.1.29)$$

证 由定理 1.1.19, 有

$$\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) = \text{rank} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \leq \text{rank} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{I}_m & \mathbf{B} \end{bmatrix}.$$

而

$$\begin{pmatrix} \mathbf{I}_n & -\mathbf{A} \\ \mathbf{0} & \mathbf{I}_m \end{pmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{I}_m & \mathbf{B} \end{bmatrix} \begin{pmatrix} \mathbf{I}_m & -\mathbf{B} \\ \mathbf{0} & \mathbf{I}_p \end{pmatrix} = \begin{bmatrix} \mathbf{0} & -\mathbf{AB} \\ \mathbf{I}_m & \mathbf{0} \end{bmatrix}.$$

由定理 1.1.21,

$$\text{rank} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{I}_m & \mathbf{B} \end{bmatrix} = \text{rank} \begin{bmatrix} \mathbf{0} & -\mathbf{AB} \\ \mathbf{I}_m & \mathbf{0} \end{bmatrix} = \text{rank} \begin{bmatrix} \mathbf{AB} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

由定理 1.1.19

$$\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) \leq \text{rank} \begin{bmatrix} \mathbf{AB} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} = \text{rank}(\mathbf{AB}) + \text{rank}(\mathbf{I}_m) = m + \text{rank}(\mathbf{AB}).$$

所以式(1.1.29)成立.

证毕

这个定理给出了一种重要的矩阵技巧,用来计算有关各种矩阵的秩的不等式.

8. 线性变换的特征值问题

线性变换的特征值和特征向量(简称特征值问题)是相当重要的概念,它们不仅对线性变换本身的研究具有重要的意义,而且在物理、力学和工程技术中研究振动波时具有实际的意义.

定义 1.1.27 设 \mathcal{A} 是数域 P 上线性空间 V^n 的一个线性变换,如果存在 $\lambda \in P$ 以及非零向量 $\mathbf{x} \in V^n$,使得

$$\mathcal{A}(\mathbf{x}) = \lambda \mathbf{x}, \quad (1.1.30)$$

则称 λ 为 \mathcal{A} 的特征值,并称 \mathbf{x} 为 \mathcal{A} 的属于(或对应于)特征值 λ 的特征向量. 求特征值和特征向量,统称为特征值问题.

如何求出 \mathcal{A} 的特征值和特征向量呢? 实际上,它们可以转化成线性代数中关于矩阵的特征值问题来解决.

设 $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ 是 V^n 的一组基,线性变换 \mathcal{A} 在该组基下的矩阵为 \mathbf{A} . 如果 λ 是 \mathcal{A} 的特征值, \mathbf{x} 是相应的特征向量,则

$$\mathbf{x} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad (1.1.31)$$

把式(1.1.31)代入式(1.1.30),并利用 \mathcal{A} 在基下的矩阵表示可得

$$(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) \mathbf{A} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) \lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

由于 $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ 线性无关,所以

$$\mathbf{A} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad (1.1.32)$$

这说明特征向量 \mathbf{x} 的坐标 $\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ 满足齐次线性方程组

$$(\lambda I - A)X = 0. \quad (1.1.33)$$

因为 $x \neq 0$ 所以 $X \neq 0$, 即齐次线性方程组 (1.1.33) 有非零解. 方程组 (1.1.33) 有非零解的充分必要条件是它的系数行列式为零, 即

$$|\lambda I - A| = 0. \quad (1.1.34)$$

因此 \mathcal{A} 的特征值 λ , 也是矩阵 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 (1.1.33) 有非零解 X , 从而非零向量

$$x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n$$

满足式 (1.1.30), 即 λ 是线性变换 \mathcal{A} 的特征值, x 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 1.1.30 设线性变换 \mathcal{A} 在基 e_1, e_2, e_3 下的矩阵是

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

求 \mathcal{A} 的特征值与特征向量.

解 矩阵 A 的特征多项式为

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

所以矩阵 A (即线性变换 \mathcal{A}) 的特征值是 $\lambda_{1,2} = -1$ (二重) 和 $\lambda_3 = 5$.

对应于特征值 $\lambda_{1,2} = -1$, 齐次线性方程组 $(\lambda_{1,2} I - A)X = 0$ 的基础解系为

$$X_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}.$$

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

$$x_1 = e_1 - e_3, \quad x_2 = e_2 - e_3.$$

对应于特征值 $\lambda_3 = 5$, 齐次线性方程组 $(\lambda_3 I - A)X = 0$ 的基础解系为 $X_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. 因此, 线

性变换 \mathcal{A} 属于特征值 5 的一个线性无关的特征向量是 $x_3 = e_1 + e_2 + e_3$.

例 1.1.31 在 $n+1$ 维线性空间 $P[x]_n$ 中, 求导变换 \mathcal{D}

$$\mathcal{D}(f(x)) = f'(x)$$

在基 $1, x, \frac{x^2}{2!}, \cdots, \frac{x^n}{n!}$ 下的矩阵是

$$D = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

D 的特征多项式为

$$|\lambda \mathbf{I} - \mathbf{D}| = \begin{vmatrix} \lambda & -1 & 0 & \cdots & 0 \\ 0 & \lambda & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & -1 \\ 0 & 0 & 0 & \cdots & \lambda \end{vmatrix} = \lambda^{n+1},$$

所以矩阵 \mathbf{D} 的特征值是 $\lambda=0$ ($n+1$ 重). 又由于相应的齐次线性方程组 $(\lambda \mathbf{I} - \mathbf{D})\mathbf{X} = \mathbf{0}$ 的基础解系为 $(1, 0, \dots, 0)^T$, 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.

在 n 维线性空间 V^n 中取定一组基之后, 线性变换的特征值就是它在这组基下的矩阵的特征值. 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.

下面讨论矩阵特征值问题的一些重要性质.

定理 1.1.23 相似矩阵有相同的特征多项式.

证明 设 $\mathbf{B} = \mathbf{C}^{-1}\mathbf{A}\mathbf{C}$, 则

$$\begin{aligned} |\lambda \mathbf{I} - \mathbf{B}| &= |\lambda \mathbf{I} - \mathbf{C}^{-1}\mathbf{A}\mathbf{C}| = |\mathbf{C}^{-1}(\lambda \mathbf{I} - \mathbf{A})\mathbf{C}| \\ &= |\mathbf{C}^{-1}| |\lambda \mathbf{I} - \mathbf{A}| |\mathbf{C}| = |\lambda \mathbf{I} - \mathbf{A}|. \end{aligned}$$

证毕

推论 相似矩阵有相同的特征值.

注意, 特征值完全由特征多项式确定, 所以相似矩阵 \mathbf{A} 和 \mathbf{B} 的特征值完全相同. 但是, 反之不真, 即有相同特征多项式的矩阵不一定相似, 例如

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

\mathbf{A} 与 \mathbf{I} 有相同的特征多项式 $(\lambda - 1)^2$ (亦即有相同的特征值), 可是对任何可逆矩阵 \mathbf{C} 都有 $\mathbf{C}^{-1}\mathbf{I}\mathbf{C} = \mathbf{I}$, 说明与单位矩阵 \mathbf{I} 相似的只有 \mathbf{I} 自己, 换言之, \mathbf{A} 与 \mathbf{I} 不相似.

定理 1.1.24 设 n 阶矩阵有特征值 λ , 对应的特征向量为 \mathbf{x} , 则

- (1) n 阶矩阵 $\mu\mathbf{A}$ 有特征值 $\mu\lambda$, 对应的特征向量仍为 \mathbf{x} (其中 μ 为任意常数);
- (2) 矩阵 \mathbf{A}^m 有特征值 λ^m , 对应的特征向量仍为 \mathbf{x} (m 为正整数);
- (3) 矩阵 \mathbf{A}^{-1} 有特征值 λ^{-1} ($\lambda \neq 0$), 对应的特征向量仍为 \mathbf{x} ;
- (4) 矩阵 \mathbf{A}^T 有特征值 λ (即转置矩阵有相同的特征值).

证明 (1) $(\mu\mathbf{A})\mathbf{x} = \mu(\mathbf{A}\mathbf{x}) = (\mu\lambda)\mathbf{x}$;

(2) 用数学归纳法证明:

$$\begin{aligned} \mathbf{A}^2\mathbf{x} &= \mathbf{A}(\mathbf{A}\mathbf{x}) = \mathbf{A}(\lambda\mathbf{x}) = \lambda(\mathbf{A}\mathbf{x}) \\ &= \lambda(\lambda\mathbf{x}) = (\lambda^2)\mathbf{x}. \end{aligned}$$

设 $\mathbf{A}^k\mathbf{x} = \lambda^k\mathbf{x}$, 则

$$\mathbf{A}^{k+1}\mathbf{x} = \mathbf{A}(\mathbf{A}^k\mathbf{x}) = \mathbf{A}(\lambda^k\mathbf{x}) = \lambda^k(\mathbf{A}\mathbf{x}) = \lambda^{k+1}\mathbf{x}.$$

$$(3) \mathbf{A}^{-1}\mathbf{x} = \mathbf{A}^{-1}\left(\frac{\lambda}{\lambda}\mathbf{x}\right) = \frac{1}{\lambda}\mathbf{A}^{-1}(\lambda\mathbf{x}) = \frac{1}{\lambda}\mathbf{I}\mathbf{x} = (\lambda^{-1})\mathbf{x};$$

(4) $|\lambda \mathbf{I} - \mathbf{A}| = |(\lambda \mathbf{I} - \mathbf{A})^T| = |\lambda \mathbf{I} - \mathbf{A}^T| = 0$, 矩阵 \mathbf{A}^T 与 \mathbf{A} 有相同的特征多项式, 即有相同的特征值.

证毕

定理 1.1.25 设 $A = (a_{ij}) \in \mathbb{R}^{n \times n}$, 则

$$|\lambda I - A| = \lambda^n + \sum_{k=1}^n (-1)^k b_k \lambda^{n-k}, \quad (1.1.35)$$

其中 $b_k (k=1, 2, \dots, n)$ 是 A 的所有 k 阶主子式之和, 特别地, 有

$$b_1 = a_{11} + a_{22} + \dots + a_{nn}, \quad b_n = |A|. \quad (1.1.36)$$

证明 记 $I = (e_1, e_2, \dots, e_n)$, $A = (\alpha_1, \alpha_2, \dots, \alpha_n)$, 其中 e_i 和 α_i 分别是 I 和 A 的第 i 列, 则

$$|\lambda I - A| = |(\lambda e_1 - \alpha_1, \lambda e_2 - \alpha_2, \dots, \lambda e_n - \alpha_n)|.$$

利用行列式的性质, 将上式右端拆成每列是 λe_i 或 $-\alpha_i$ 的行列式, 有

$$|\lambda I - A| = |(\lambda e_1, \lambda e_2 - \alpha_2, \dots, \lambda e_n - \alpha_n)| + \\ |(-\alpha_1, \lambda e_2 - \alpha_2, \dots, \lambda e_n - \alpha_n)|,$$

继续利用行列式的性质, 再把上面两个行列式继续拆成行列式之和, 合并后即有

$$|\lambda I - A| = \lambda^n |(e_1, e_2, \dots, e_n)| - \lambda^{n-1} \sum_{i=1}^n |(e_1, \dots, e_{i-1}, \alpha_i, \dots, e_{i+1}, \dots, e_n)| + \dots + \\ (-1)^k \lambda^{n-k} \sum_{\substack{1 \leq i_1 < \dots < i_k \leq n \\ 2 \leq k \leq n-1}} |(\dots, \alpha_{i_1}, \dots, \alpha_{i_k}, \dots)| + \dots + (-1)^n |A|,$$

式中行列式 $|(\dots, \alpha_{i_1}, \dots, \alpha_{i_k}, \dots)|$ 表示第 i_1 列, \dots , 第 i_k 列依次是 $\alpha_{i_1}, \dots, \alpha_{i_k}$, 而其余的列是单位矩阵 I 中相应的列. 例如, $n=5, k=3, i_1=1, i_2=2, i_3=4$, 则

$$|(\alpha_1, \alpha_2, e_3, \alpha_4, e_5)| = \begin{vmatrix} a_{11} & a_{12} & 0 & a_{14} & 0 \\ a_{21} & a_{22} & 0 & a_{24} & 0 \\ a_{31} & a_{32} & 1 & a_{34} & 0 \\ a_{41} & a_{42} & 0 & a_{44} & 0 \\ a_{51} & a_{52} & 0 & a_{54} & 1 \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{14} \\ a_{21} & a_{22} & a_{24} \\ a_{41} & a_{42} & a_{44} \end{vmatrix},$$

它是 A 的一个三阶主子式, 因此 b_3 是关于 $1 \leq i_1 < i_2 < i_3 \leq 5$ 求和, 即为 A 的所有三阶主子式之和. 证毕

由上述定理可知, 特征多项式可写为

$$f(\lambda) = |\lambda I - A| = \lambda^n + c_1 \lambda^{n-1} + \dots + c_{n-1} \lambda + c_n \\ = \lambda^n - (a_{11} + a_{22} + \dots + a_{nn}) \lambda^{n-1} + \dots + (-1)^n \det A. \quad (1.1.37)$$

根据代数方程的根与方程系数之间的关系, 即韦达 (Vieta) 定理, 便知特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 要满足下列 2 个条件:

(1) $\sum_{i=1}^n \lambda_i = -c_1 = \sum_{i=1}^n a_{ii}$, 式中 A 的主对角线上的元素之和 $\sum_{i=1}^n a_{ii}$ 叫做方阵 A 的迹, 记为 $\text{tr}A$, 即

$$\text{tr}A = \sum_{i=1}^n \lambda_i = \sum_{i=1}^n a_{ii}. \quad (1.1.38)$$

换言之, n 个特征值之和等于 A 的迹.

不难看出, 如果 n 阶矩阵 A 与 B 相似, 则有 $\text{tr}A = \text{tr}B$.

(2) $\prod_{i=1}^n \lambda_i = (-1)^n c_n = |A|$, 即 n 个特征值的积等于 A 的行列式.

例如,对于方阵 $A = \begin{bmatrix} 2 & 1 \\ 2 & 3 \end{bmatrix}$, 得

$$f(\lambda) = \begin{vmatrix} \lambda - 2 & -1 \\ -2 & \lambda - 3 \end{vmatrix} = \lambda^2 - 5\lambda + 4 = 0,$$

故知

$$c_1 = -5, \quad c_2 = 4, \quad \lambda_1 = 1, \quad \lambda_2 = 4,$$

因而有

$$\lambda_1 + \lambda_2 = (-1)c_1 = 5, \quad \lambda_1 \lambda_2 = c_2 = 4.$$

例 1.1.32 设

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_n \\ 1 & 0 & 0 & \cdots & 0 & -a_{n-1} \\ 0 & 1 & 0 & \cdots & 0 & -a_{n-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -a_1 \end{pmatrix},$$

矩阵 A 在控制论中称为友矩阵或相伴矩阵, 求 A 的特征多项式.

解 记

$$D_i = \begin{vmatrix} \lambda & 0 & 0 & \cdots & a_i \\ -1 & \lambda & 0 & \cdots & a_{i-1} \\ 0 & -1 & \lambda & \cdots & a_{i-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & -1 & \lambda + a_1 \end{vmatrix}, \quad i \geq 1, \quad D_0 = 1,$$

对 D_i 按第 1 行展开, 有

$$D_i = \lambda D_{i-1} + a_i, \quad i \geq 1.$$

由上式逐次递推得

$$\begin{aligned} D_n &= |\lambda I - A| = \lambda D_{n-1} + a_n = \lambda(\lambda D_{n-2} + a_{n-1}) + a_n \\ &= \lambda^2(\lambda D_{n-3} + a_{n-2}) + a_{n-1}\lambda + a_n \\ &= \lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + \cdots + a_{n-1}\lambda + a_n. \end{aligned}$$

例 1.1.33 设 $x, y \in \mathbb{C}^n, A = xy^H$, 求 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 0, 而 A 的一阶主子式之和为 $\text{tr}(xy^H) = y^H x$, 由定理 1.1.25 得

$$|\lambda I - xy^H| = \lambda^n - y^H x \lambda^{n-1} = \lambda^{n-1}(\lambda - y^H x).$$

例 1.1.34 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

$$\begin{pmatrix} I_m & -A \\ 0 & I_n \end{pmatrix} \begin{bmatrix} AB & 0 \\ B & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ B & BA \end{bmatrix} \begin{pmatrix} I_m & -A \\ 0 & I_n \end{pmatrix}.$$

因为 $\begin{pmatrix} I_m & -A \\ 0 & I_n \end{pmatrix}$ 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 1.1.23 有

$$\det(\lambda I_{m+n} - \begin{bmatrix} AB & 0 \\ B & 0 \end{bmatrix}) = \det(\lambda I_{m+n} - \begin{bmatrix} 0 & 0 \\ B & BA \end{bmatrix}),$$

即

$$\det \begin{pmatrix} \lambda \mathbf{I}_m - \mathbf{AB} & \mathbf{0} \\ -\mathbf{B} & \lambda \mathbf{I}_n \end{pmatrix} = \det \begin{pmatrix} \lambda \mathbf{I}_m & \mathbf{0} \\ -\mathbf{B} & \lambda \mathbf{I}_n - \mathbf{BA} \end{pmatrix},$$

从而

$$\lambda^n |\lambda \mathbf{I}_m - \mathbf{AB}| = \lambda^m |\lambda \mathbf{I}_n - \mathbf{BA}|. \quad \text{证毕}$$

由例 1.1.34 可知, m 阶方阵 \mathbf{AB} 与 n 阶方阵 \mathbf{BA} 具有相同的非零特征值, 从而有 $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$. 特别地, 若 \mathbf{A}, \mathbf{B} 为同阶方阵 (即 $m=n$), 则 \mathbf{AB} 与 \mathbf{BA} 具有相同的特征值.

下面进一步讨论线性变换 \mathcal{A} (或矩阵 \mathbf{A}) 的特征值以及特征向量之间的关系.

由代数基本定理 (即: n 次代数方程在复数域内有且仅有 n 个根 (重根按重数计算)) 可知, n 阶方阵 \mathbf{A} 在复数域内恰有 n 个特征值. 不妨假设 $\lambda_1, \lambda_2, \dots, \lambda_r$ 是 \mathbf{A} 的相异特征值, 其重数分别为 m_1, m_2, \dots, m_r , 则称 m_i 为 λ_i 的代数重复度. 显然有 $\sum_{i=1}^r m_i = n$.

由例 1.1.30 可见, 通过求 $(\lambda \mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ 的基础解系的途径求出了 \mathbf{A} 属于 λ_i 的线性无关的特征向量, 从而也就求出了 \mathcal{A} 属于 λ_i 的线性无关的特征向量. 其实, 齐次线性方程组 $(\lambda_i \mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ 的解空间中任一个非零向量都是 \mathbf{A} 属于 λ_i 的特征向量. 因此, 有下面的定义.

矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$, λ_i 是 \mathbf{A} 的一个特征值, 记齐次线性方程组 $(\lambda_i \mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ 的解空间 (包括 $\mathbf{x} = \mathbf{0}$) 为

$$V_{\lambda_i} = \{\mathbf{x} \mid (\lambda_i \mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}, \mathbf{x} \in \mathbb{C}^n\}.$$

对线性空间 V^n 上线性变换 \mathcal{A} , 属于特征值 λ_i 的全部特征向量再添上零向量所组成的集合, 也记为

$$V_{\lambda_i} = \{\mathbf{x} \mid \mathcal{A}(\mathbf{x}) = \lambda_i \mathbf{x}, \mathbf{x} \in V^n\},$$

则 V_{λ_i} 是 \mathbb{C}^n (或 V^n) 的一个子空间, 称 V_{λ_i} 为矩阵 \mathbf{A} (或线性变换 \mathcal{A}) 的属于 λ_i 的特征子空间. 显然, $\dim(V_{\lambda_i})$ 就是属于 λ_i 的线性无关特征向量的最大数目, 称 $\dim(V_{\lambda_i})$ 为特征值 λ_i 的几何重复度.

关于 λ_i 的代数重复度与几何重复度的关系, 有如下的定理.

定理 1.1.26 矩阵 \mathbf{A} 的任一特征值的几何重复度不大于它的代数重复度.

证明 设 λ_0 为 \mathbf{A} 的一个特征值, 齐次线性方程组 $(\lambda_0 \mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ 的基础解系为 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$, 由第 1 章基的扩充定理 1.1.5 知, 可取 $\mathbf{x}_{k+1}, \mathbf{x}_{k+2}, \dots, \mathbf{x}_n$ 使得

$$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \mathbf{x}_{k+1}, \dots, \mathbf{x}_n$$

构成 \mathbb{C}^n 的一组基, 令矩阵

$$\mathbf{C} = [\mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{x}_{k+1}, \mathbf{x}_{k+2}, \dots, \mathbf{x}_n],$$

故

$$\mathbf{C}^{-1}\mathbf{C} = [\mathbf{C}^{-1}\mathbf{x}_1, \mathbf{C}^{-1}\mathbf{x}_2, \dots, \mathbf{C}^{-1}\mathbf{x}_n] = \mathbf{I}_n,$$

即

$$\mathbf{C}^{-1}\mathbf{x}_1 = (1, 0, \dots, 0)^T,$$

$$\mathbf{C}^{-1}\mathbf{x}_2 = (0, 1, 0, \dots, 0)^T,$$

$$\vdots$$

$$\mathbf{C}^{-1}\mathbf{x}_k = (0, \dots, 0, 1, \dots, 0)^T,$$

$$\vdots$$

$$\mathbf{C}^{-1}\mathbf{x}_n = (0, \dots, 0, 1)^T.$$

于是

$$\begin{aligned} C^{-1}AC &= [C^{-1}Ax_1, C^{-1}Ax_2, \dots, C^{-1}Ax_k, \dots, C^{-1}Ax_n] \\ &= [\lambda_0 C^{-1}x_1, \lambda_0 C^{-1}x_2, \dots, \lambda_0 C^{-1}x_k, C^{-1}Ax_{k+1}, \dots, C^{-1}Ax_n] \\ &= \begin{pmatrix} \lambda_0 & & & & * \\ & \ddots & & & \\ & & \lambda_0 & & \\ \mathbf{0} & & & & \mathbf{A}_0 \end{pmatrix} = \mathbf{B}, \end{aligned}$$

式中 $\mathbf{0}$ 表示 $(n-k) \times k$ 零矩阵, $*$ 表示 $k \times (n-k)$ 矩阵, \mathbf{A}_0 表示 $n-k$ 阶方阵. 因此矩阵 \mathbf{B} 与 \mathbf{A} 相似, 从而有相同的特征多项式

$$|\lambda \mathbf{I}_n - \mathbf{A}| = |\lambda \mathbf{I}_n - \mathbf{B}| = (\lambda - \lambda_0)^k |\lambda \mathbf{I}_{n-k} - \mathbf{A}_0|,$$

这说明 λ_0 的代数重复度至少为 k , 而 λ_0 的几何重复度仅为 k (因为假定基础解系仅含 k 个向量), 所以几何重复度不大于代数重复度. 证毕

推论 如果 λ_0 的代数重复度为 1, 则它的几何重复度 $\dim(V_{\lambda_0}) = 1$.

例如, 在例 1.1.30 中, 有两个特征子空间 $V_1 = \text{Span}(\mathbf{x}_1, \mathbf{x}_2)$, $V_2 = \text{Span}(\mathbf{x}_3)$, $\lambda = -1$ 的代数重复度与几何重复度相等, 都为 2, 而 $\lambda = 5$ 的代数重复度为 1, 几何重复度 $\dim(V_2) = 1$.

例 1.1.35 设

$$\mathbf{A} = \begin{pmatrix} 3 & 1 & 0 \\ -4 & -1 & 0 \\ 4 & -8 & -2 \end{pmatrix},$$

特征多项式为

$$f(\lambda) = (\lambda - 1)^2(\lambda + 2),$$

其特征值为 $\lambda_{1,2} = 1, \lambda_3 = -2$, 将其逐一代入齐次线性方程组(1.1.33)得

$$\mathbf{x}_1 = \begin{pmatrix} 3 \\ -6 \\ 20 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

这里 $\lambda = 1$ 的代数重复度为 2, 可是它的特征子空间 $V_1 = \text{Span}(\mathbf{x}_1)$ 的维数为 1, 即几何重复度是 1, 它小于代数重复度, 而 $\lambda = -2$ 是单根, 它的代数重复度与几何重复度都为 1.

现在的问题是: 在例 1.1.30 中 $\mathbf{x}_1 = (1, 0, -1)^T$, $\mathbf{x}_2 = (0, 1, -1)^T$ 是属于 $\lambda_{1,2} = -1$ 的线性无关的特征向量, 而 $\mathbf{x}_3 = (1, 1, 1)^T$ 是属于 $\lambda_3 = 5$ 的特征向量, 那么, 把属于这两个不同特征值的线性无关的特征向量合起来的向量组 $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ 也能线性无关吗? 同理, 在例 1.1.35 中 $\mathbf{x}_1 = (3, -6, 20)^T$, $\mathbf{x}_2 = (0, 0, 1)^T$ 线性无关吗? 回答是肯定的, 有下面的定理.

定理 1.1.27 设 $\lambda_1, \lambda_2, \dots, \lambda_r$ 是矩阵 \mathbf{A} 的互异特征值, $\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_{s_i}^{(i)}$ 是属于 λ_i 的线性无关的特征向量(一般取齐次线性方程组 $(\lambda_i \mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ 的基础解系), 则

$$\mathbf{x}_1^{(1)}, \dots, \mathbf{x}_{s_1}^{(1)}, \mathbf{x}_1^{(2)}, \dots, \mathbf{x}_{s_2}^{(2)}, \dots, \mathbf{x}_1^{(r)}, \dots, \mathbf{x}_{s_r}^{(r)}$$

也线性无关.

证明 考察

$$k_{11}\mathbf{x}_1^{(1)} + \dots + k_{1s_1}\mathbf{x}_{s_1}^{(1)} + k_{21}\mathbf{x}_1^{(2)} + \dots + k_{2s_2}\mathbf{x}_{s_2}^{(2)} + \dots + k_{r1}\mathbf{x}_1^{(r)} + \dots + k_{rs_r}\mathbf{x}_{s_r}^{(r)} = \mathbf{0},$$

(1.1.39)

其中 $k_{ij} \in P (i=1, 2, \dots, r; j=1, 2, \dots, s_i)$. 对式(1.1.39)左乘以 $\lambda_1 \mathbf{I} - \mathbf{A}$, 利用 $(\lambda_1 \mathbf{I} - \mathbf{A})$

$\mathbf{A}\mathbf{x}_j^{(1)} = \mathbf{0} (j=1, 2, \dots, s_1)$ 与 $\mathbf{A}\mathbf{x}_j^{(i)} = \lambda_i \mathbf{x}_j^{(i)} (i=2, 3, \dots, r; j=1, 2, \dots, s_i)$, 则有

$$(\lambda_1 - \lambda_2)k_{21}\mathbf{x}_1^{(2)} + \dots + (\lambda_1 - \lambda_r)k_{r1}\mathbf{x}_1^{(r)} = \mathbf{0}.$$

再依次左乘 $(\lambda_2 \mathbf{I} - \mathbf{A}), (\lambda_3 \mathbf{I} - \mathbf{A}), \dots, (\lambda_{r-1} \mathbf{I} - \mathbf{A})$ 最后得

$$(\lambda_1 - \lambda_r)(\lambda_2 - \lambda_r) \dots (\lambda_{r-1} - \lambda_r)(k_{r1}\mathbf{x}_1^{(r)} + \dots + k_{r_s_r}\mathbf{x}_{s_r}^{(r)}) = \mathbf{0}.$$

因为

$$\lambda_i \neq \lambda_r, \quad i = 1, 2, \dots, r-1,$$

故得

$$k_{r1}\mathbf{x}_1^{(r)} + k_{r2}\mathbf{x}_2^{(r)} + \dots + k_{r_s_r}\mathbf{x}_{s_r}^{(r)} = \mathbf{0}.$$

因 $\mathbf{x}_1^{(r)}, \mathbf{x}_2^{(r)}, \dots, \mathbf{x}_{s_r}^{(r)}$ 是线性无关的, 所以

$$k_{r1} = k_{r2} = \dots = k_{r_s_r} = 0.$$

重复以上相应步骤, 可证得所有 $k_{ij} = 0 (i=1, 2, \dots, r; j=1, 2, \dots, s_i)$. 证毕

在定理 1.1.26 中, 如果对每个特征值 λ_i , 只取它的一个特征向量 \mathbf{x}_i , 则得下述推论.

推论 1 设 $\lambda_1, \lambda_2, \dots, \lambda_r$ 是 \mathbf{A} 的互异特征值, 又设 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r$ 是属于 $\lambda_1, \lambda_2, \dots, \lambda_r$ 的特征向量, 则 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r$ 线性无关.

推论 1 表明, 属于不同特征值的特征向量是线性无关的. 由此得到下面推论.

推论 2 n 阶矩阵 \mathbf{A} 若有 n 个不同的特征值 (即全是单根), 则 \mathbf{A} 一定有 n 个线性无关的特征向量.

推论 3 n 阶矩阵 \mathbf{A} 的每一个特征值的代数重复度等于几何重复度, 则 \mathbf{A} 一定有 n 个线性无关的特征向量.

这样, 我们就较圆满地解释了为什么在例 1.1.30 中 \mathbf{A} 的特征方程有重根, 但仍能找到 3 个线性无关的特征向量, 而在例 1.1.35 中 \mathbf{A} 的特征方程也有重根, 却找不到 3 个线性无关的特征向量.

由于线性变换 \mathcal{A} 的特征值 λ_i 与特征向量 \mathbf{X}_i 是通过相应矩阵 \mathbf{A} 的特征值 λ_i 与特征向量 \mathbf{X}_i 来求得的, 所以对于线性变换 \mathcal{A} 也有类似的定理.

定理 1.1.28 线性变换 \mathcal{A} 的任一特征值的几何重复度不超过它的代数重复度.

定理 1.1.29 设 $\lambda_1, \lambda_2, \dots, \lambda_r$ 是线性变换 \mathcal{A} 的 r 个互异特征值, $\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_{s_i}^{(i)}$ 是属于特征值 λ_i 的 s_i 个线性无关特征向量 ($i=1, 2, \dots, r$), 则 $\mathbf{x}_1^{(1)}, \dots, \mathbf{x}_{s_1}^{(1)}, \mathbf{x}_1^{(2)}, \dots, \mathbf{x}_{s_2}^{(2)}, \dots, \mathbf{x}_1^{(r)}, \dots, \mathbf{x}_{s_r}^{(r)}$ 线性无关.

对角矩阵是矩阵中最简单的一种. 下面要研究对于数域 P 上 n 维线性空间 V^n 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 1.1.28 设 \mathcal{A} 是数域 P 上 n 维线性空间 V^n 的一个线性变换, 如果 V^n 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.

定理 1.1.30 数域 P 上 n 维线性空间 V^n 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 $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ 下的矩阵是对角矩阵 $\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, 记为 \mathbf{A} , 则

$$(\mathcal{A}(\mathbf{e}_1), \mathcal{A}(\mathbf{e}_2), \dots, \mathcal{A}(\mathbf{e}_n)) = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)\mathbf{A}. \quad (1.1.40)$$

所以有

$$\mathcal{A}(\mathbf{e}_i) = \lambda_i \mathbf{e}_i, \quad i = 1, 2, \dots, n,$$

于是 $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ 是 \mathcal{A} 的 n 个线性无关的特征向量.

反之,如果 \mathcal{A} 有 n 个线性无关的特征向量 e_1, e_2, \dots, e_n , 则取 e_1, e_2, \dots, e_n 为 V^n 的一组基, 且由式(1.1.40)知, \mathcal{A} 在这组基下的矩阵是对角矩阵 Λ . 证毕

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 阶可逆方阵 C , 使得

$$C^{-1}AC = \Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n).$$

定义 1.1.29 如果 n 阶方阵 A 与对角矩阵相似, 则称矩阵 A 是可对角化的.

定理 1.1.31 n 阶方阵 A 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.

证明 必要性. 设满秩矩阵 C , 满足

$$C^{-1}AC = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n), \quad (1.1.41)$$

其中 $\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ 表示对角元素分别是 $\lambda_1, \lambda_2, \dots, \lambda_n$ 的 n 阶对角矩阵, 把 C 按列分块, 有

$$C = (x_1, x_2, \dots, x_n). \quad (1.1.42)$$

将式(1.1.42)左乘式(1.1.41)得

$$A(x_1, x_2, \dots, x_n) = (x_1, x_2, \dots, x_n)\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n),$$

于是有

$$Ax_i = \lambda_i x_i, \quad i = 1, 2, \dots, n. \quad (1.1.43)$$

因为 C 满秩, 所以 x_1, x_2, \dots, x_n 线性无关, 从而由式(1.1.43)知 A 有 n 个线性无关的特征向量.

充分性: 设 A 有 n 个线性无关的特征向量 x_1, x_2, \dots, x_n , 即 $Ax_i = \lambda_i x_i (i=1, 2, \dots, n)$, 令

$$C = (x_1, x_2, \dots, x_n),$$

显然 C 是满秩的, 故

$$\begin{aligned} AC &= A(x_1, x_2, \dots, x_n) = (Ax_1, Ax_2, \dots, Ax_n) \\ &= (\lambda_1 x_1, \lambda_2 x_2, \dots, \lambda_n x_n) \\ &= (x_1, x_2, \dots, x_n)\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) \\ &= C\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n), \end{aligned}$$

即

$$C^{-1}AC = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n). \quad \text{证毕}$$

从上面定理的证明过程, 我们可以看到化矩阵 A 为对角矩阵 Λ 的方法可归结为求矩阵 A 的特征值和特征向量.

推论 若 $C^{-1}AC = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, 则 $\lambda_1, \lambda_2, \dots, \lambda_n$ 是 A 的 n 个特征值, C 的第 i 个列向量是 A 的属于特征值 λ_i 的特征向量.

需要指出, 这里 $\lambda_1, \lambda_2, \dots, \lambda_n$ 的排列顺序必须与 x_1, x_2, \dots, x_n 的排列顺序相对应, 否则 C 就不是原来的矩阵了.

例 1.1.36 在例 1.1.30 中已经计算出矩阵 A 的特征值是 -1 和 5 , 而对应的特征向量是

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

令

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

则有

$$\mathbf{C}^{-1}\mathbf{A}\mathbf{C} = \text{diag}(-1, -1, 5).$$

如果取

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

则

$$\mathbf{C}^{-1}\mathbf{A}\mathbf{C} = \text{diag}(-1, 5, -1).$$

可见使之对角化的矩阵 \mathbf{C} 不是惟一的.

定义 1.1.30 当 n 阶矩阵 \mathbf{A} 有 n 个线性无关的特征向量时, 则称矩阵 \mathbf{A} 有完备的特征向量系; 否则, 称 \mathbf{A} 为亏损矩阵.

由定理 1.1.27、定理 1.1.29、定理 1.1.30 以及定理 1.1.31 可以直接推得下面定理.

定理 1.1.32 如果数域 P 上 n 维线性空间 V^n 上的线性变换 \mathcal{A} (或 n 阶矩阵 \mathbf{A}) 有 n 个不同的特征值, 则线性变换 \mathcal{A} (或矩阵 \mathbf{A}) 是可对角化的.

定理 1.1.33 数域 P 上 n 维线性空间 V^n 上的线性变换 \mathcal{A} (或矩阵 \mathbf{A}) 可对角化的充分必要条件是 \mathcal{A} (或 \mathbf{A}) 的每一个特征值的几何重复度等于代数重复度.

证明 设 \mathcal{A} (或 \mathbf{A}) 的互异特征值为 $\lambda_1, \lambda_2, \dots, \lambda_r, \lambda_i$ 的代数重复度记为 m_i , 几何重复度记为 s_i ($i=1, 2, \dots, r$), 则

$$m_1 + m_2 + \dots + m_r = n.$$

先证充分性. 如果 $m_i = s_i$ ($i=1, 2, \dots, r$), 则由定理 1.1.29 (或定理 1.1.27) 和定理 1.1.32 知, \mathcal{A} (或 \mathbf{A}) 是可对角化的.

再证必要性. 如果 \mathcal{A} (或 \mathbf{A}) 是可对角化的, 则由定理 1.1.30 (或定理 1.1.31) 知

$$s_1 + s_2 + \dots + s_r = n.$$

由定理 1.1.28 (或定理 1.1.26) 有 $s_i \leq m_i$ ($i=1, 2, \dots, r$), 从而

$$n = s_1 + s_2 + \dots + s_r \leq m_1 + m_2 + \dots + m_r = n.$$

故得 $s_i = m_i$ ($i=1, 2, \dots, r$).

证毕

例如, 在例 1.1.30 中的三维线性空间上的线性变换 \mathcal{A} 有 3 个线性无关的特征向量, 因此 \mathcal{A} 可对角化. 对于例 1.1.31 中 $n+1$ 维线性空间 $P[x]_n$ 上的求导变换 \mathcal{D} , 其零特征值的几何重复度小于代数重复度, 因此 \mathcal{D} 是不可对角化的.

经常称代数重复度和几何重复度相等的矩阵为单纯矩阵, 因此, 矩阵 \mathbf{A} 与对角矩阵相似的充要条件又可表述为 \mathbf{A} 是单纯矩阵.

若矩阵 \mathbf{A} 的特征值全是单根, 则 \mathbf{A} 必与对角矩阵相似; 但反之不真.

例如,在例 1.1.30 中 \mathbf{A} 是一个单纯矩阵,所以可以与对角矩阵相似(即可对角化);而在例 1.1.35 中 \mathbf{A} 是一个亏损矩阵,所以不与对角矩阵相似(即不可对角化).

定理 1.1.34 设数域 P 上 n 维线性空间 V^n 上的线性变换 \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

$$V^n = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \dots \oplus V_{\lambda_r}. \quad (1.1.44)$$

证明 必要性: 设 \mathcal{A} 可对角化, 在 V_{λ_i} 中取一组基 $\mathbf{x}_{s_i}^{(i)}, \dots, \mathbf{x}_{s_i}^{(i)}$ ($i=1, 2, \dots, r$), 则由定理 1.1.29 知 $\mathbf{x}_1^{(1)}, \dots, \mathbf{x}_{s_1}^{(1)}, \dots, \mathbf{x}_1^{(r)}, \dots, \mathbf{x}_{s_r}^{(r)}$ 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 1.1.30 知, $s_1 + s_2 + \dots + s_r = n$, 从而 $\mathbf{x}_1^{(1)}, \dots, \mathbf{x}_{s_1}^{(1)}, \dots, \mathbf{x}_1^{(r)}, \dots, \mathbf{x}_{s_r}^{(r)}$ 是 V^n 的一组基, 因此有直和 $V^n = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \dots \oplus V_{\lambda_r}$.

充分性: 设 $V^n = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \dots \oplus V_{\lambda_r}$, 在 V_{λ_i} ($i=1, 2, \dots, r$) 中取一组基 $\mathbf{x}_{s_i}^{(i)}, \dots, \mathbf{x}_{s_i}^{(i)}$, 则 $\mathbf{x}_1^{(1)}, \dots, \mathbf{x}_{s_1}^{(1)}, \dots, \mathbf{x}_1^{(r)}, \dots, \mathbf{x}_{s_r}^{(r)}$ 是 $V_{\lambda_1} \oplus V_{\lambda_2} \oplus \dots \oplus V_{\lambda_r}$ 的一组基, 即 V^n 的一组基. 这说明 V^n 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. 证毕

9. 线性变换的不变子空间

研究线性变换的不变子空间, 就是研究线性变换与子空间的关系, 由此可以进一步简化线性变换的矩阵.

定义 1.1.31 设 V^n 是数域 P 上的 n 维线性空间, \mathcal{A} 是 V^n 上的线性变换, V_1 是 V^n 的子空间, 如果对于任何 $\mathbf{x} \in V_1$ 恒有 $\mathcal{A}(\mathbf{x}) \in V_1$ (或 $\mathcal{A}(V_1) \subseteq V_1$), 则说 V_1 是关于 \mathcal{A} 的不变子空间.

例如, 求导数运算 \mathcal{D} 是线性空间 $P[x]_n$ 的一个线性变换, $P[x]_{n-1}$ 为一切次数不超过 $n-1$ 的多项式的集合形成的线性空间, 则 $P[x]_{n-1}$ 是 \mathcal{D} 的不变子空间.

又如, 设 \mathcal{A} 是线性空间 V^n 上的一个线性变换, V^n 中所有向量的像构成的集合称为线性变换 \mathcal{A} 的值域, 记为 $R(\mathcal{A})$, 即

$$R(\mathcal{A}) = \{\mathbf{y} = \mathcal{A}(\mathbf{x}) \mid \mathbf{x} \in V^n\}.$$

所有被 \mathcal{A} 变成零向量的原像构成的集合称为 \mathcal{A} 的核, 记作 $N(\mathcal{A})$, 即

$$N(\mathcal{A}) = \{\mathbf{x} \in V^n \mid \mathcal{A}(\mathbf{x}) = \mathbf{0}\}.$$

不难知道, \mathcal{A} 的值域和核都是 \mathcal{A} 的不变子空间.

一般称 $R(\mathcal{A})$ 的维数是线性变换 \mathcal{A} 的秩, 称 $N(\mathcal{A})$ 的维数是 \mathcal{A} 的零度, 记作 $\text{null}(\mathcal{A})$. 所以, 有时又称 $R(\mathcal{A})$ 为 \mathcal{A} 的秩空间, $N(\mathcal{A})$ 为 \mathcal{A} 的核空间.

整个线性空间 V 和零子空间, 对于每个线性变换 \mathcal{A} 而言, 都是 \mathcal{A} 的不变子空间, 但我们感兴趣的是异于这两个空间的不变子空间.

下面, 我们给出两类重要的子空间, 它们与线性方程组的理论有密切的关系.

定义 1.1.32 以 \mathbb{C}^m 表示全体 m 维复向量在复数域 \mathbb{C} 上构成的线性空间, \mathbf{A} 为 $m \times n$ 复矩阵, 其列(向量)为 $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$. 显然, $\mathbf{a}_i \in \mathbb{C}^m$ ($i=1, 2, \dots, n$). 子空间 $\text{Span}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ 称为矩阵 \mathbf{A} 的列空间(值域), 记作 $R(\mathbf{A})$, 即

$$R(\mathbf{A}) = \text{Span}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n).$$

记

$$\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n), \mathbf{y} = (y_1, y_2, \dots, y_n)^T \in \mathbb{C}^n,$$

则 $R(\mathbf{A})$ 可表成

$$R(\mathbf{A}) = \{\mathbf{A}\mathbf{y} \mid \mathbf{y} \in \mathbb{C}^n\}.$$

因此,在 V^n 的基下, \mathcal{A} 的矩阵 \mathbf{A} 有类似式(1.1.46)的形式,其中

$$\mathbf{A}_i = \begin{pmatrix} a_{11}^{(i)} & a_{12}^{(i)} & \cdots & a_{1n_i}^{(i)} \\ a_{21}^{(i)} & a_{22}^{(i)} & \cdots & a_{2n_i}^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_i1}^{(i)} & a_{n_i2}^{(i)} & \cdots & a_{n_in_i}^{(i)} \end{pmatrix}, \quad i = 1, 2, \cdots, s. \quad \text{证毕}$$

反之,若线性变换 \mathcal{A} 在基(1.1.45)下的矩阵是准对角矩阵(1.1.46),则容易验证由基(1.1.45)生成的子空间 $V_i = \text{Span}(\mathbf{e}_1^{(i)}, \mathbf{e}_2^{(i)}, \cdots, \mathbf{e}_{n_i}^{(i)}) (i=1, 2, \cdots, s)$ 是 \mathcal{A} 的不变子空间.

特别地,若所有 V_i 都是一维子空间时,则 $n_i=1, s=n$, 矩阵 \mathbf{A} 简化为对角矩阵

$$\mathbf{A} = \text{diag}(a_1, a_2, \cdots, a_n) = \begin{pmatrix} a_1 & & & \\ & a_2 & & \\ & & \ddots & \\ & & & a_n \end{pmatrix}.$$

由此可知,线性变换 \mathcal{A} 的矩阵简化为一个准对角矩阵(或对角矩阵)与线性空间 V^n 可分解为若干个不变子空间的直和是相当的.

习 题 1 (2)

1. 试说明下列对应规则是 \mathbb{R}^2 的一个变换,并说明其几何意义,其中 $(x_1, x_2) \in \mathbb{R}^2$.

- (1) $\mathcal{A}[(x_1, x_2)] = (x_1, -x_2)$;
- (2) $\mathcal{A}[(x_1, x_2)] = (-x_1, x_2)$;
- (3) $\mathcal{A}[(x_1, x_2)] = (-x_1, -x_2)$;
- (4) $\mathcal{A}[(x_1, x_2)] = (x_1, 0)$;
- (5) $\mathcal{A}[(x_1, x_2)] = (0, x_2)$.

2. 判别下面所定义的变换,哪些是线性的,哪些不是.

- (1) 在线性空间 V 中, $\mathcal{A}(\alpha) = \alpha + \beta$, 其中 $\beta \in V$ 是一固定的向量;
- (2) 在线性空间 V 中, $\mathcal{A}(\alpha) = \beta$, 其中 $\beta \in V$ 是一固定的向量;
- (3) 在 \mathbb{R}^3 中, $\mathcal{A}[(x_1, x_2, x_3)] = (x_1^2, x_2 + x_3, x_3^2)$, 其中 $(x_1, x_2, x_3) \in \mathbb{R}^3$;
- (4) 在 \mathbb{R}^3 中, $\mathcal{A}[(x_1, x_2, x_3)] = (2x_1 - x_2, x_2 + x_3, x_1)$, 其中 $(x_1, x_2, x_3) \in \mathbb{R}^3$;
- (5) 在 $P[x]_n$ 中, $\mathcal{A}[f(x)] = f(x+1)$, 其中 $f(x) \in P[x]_n$;
- (6) 在 $P[x]_n$ 中, $\mathcal{A}[f(x)] = f(x_0)$, 其中 $x_0 \in \mathbb{R}$ 是一固定的实数;
- (7) 在 \mathbb{R}^3 中, $\mathcal{A}[(x_1, x_2, x_3)] = (\cos x_1, \sin x_2, 0)$, 其中 $(x_1, x_2, x_3) \in \mathbb{R}^3$.

3. 在 \mathbb{R}^2 中, 设 $\alpha = (x_1, x_2)$. 证明: $\mathcal{A}_1(\alpha) = (x_2, -x_1)$ 与 $\mathcal{A}_2(\alpha) = (x_1, -x_2)$ 是 \mathbb{R}^2 中的两个线性变换, 并求 $\mathcal{A}_1 + \mathcal{A}_2, \mathcal{A}_1\mathcal{A}_2$ 及 $\mathcal{A}_2\mathcal{A}_1$.

4. 设任一矩阵 $\mathbf{A} \in P^{n \times n}$, 又给定矩阵 $\mathbf{C} \in P^{n \times n}$, 定义变换 \mathcal{A} 如下:

$$\mathcal{A}(\mathbf{A}) = \mathbf{CA} - \mathbf{AC}.$$

证明: (1) \mathcal{A} 是 $P^{n \times n}$ 中的线性变换;

(2) 对任意 $\mathbf{A}, \mathbf{B} \in P^{n \times n}$ 有 $\mathcal{A}(\mathbf{AB}) = \mathcal{A}(\mathbf{A})\mathbf{B} + \mathbf{A}\mathcal{A}(\mathbf{B})$.

5. 设 $V = \left\{ \begin{bmatrix} a & a+b \\ c & c \end{bmatrix} \mid a, b, c \in \mathbb{R} \right\}$, 作出 V 到 \mathbb{R}^3 的同构映射.

6. 在 $P[x]_n$ 中, $\mathcal{A}_1[f(x)] = f'(x), \mathcal{A}_2[f(x)] = xf(x)$. 证明: $\mathcal{A}_1\mathcal{A}_2 - \mathcal{A}_2\mathcal{A}_1$ 是恒等变换(即单位变换).

7. 设 e_1, e_2 是线性空间 V^2 的基, \mathcal{A}_1 与 \mathcal{A}_2 是 V^2 上的线性变换: $\mathcal{A}_1(e_1) = e'_1, \mathcal{A}_1(e_2) = e'_2$, 且 $\mathcal{A}_2(e_1 + e_2) = e'_1 + e'_2, \mathcal{A}_2(e_1 - e_2) = e'_1 - e'_2$. 证明: $\mathcal{A}_1 = \mathcal{A}_2$.

8. 在 $P[x]_2$ 中, 试求在基 $\alpha_1 = 1, \alpha_2 = x, \alpha_3 = x^2$ 下的矩阵为

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

的线性变换 \mathcal{A} , 求 $g = 2 + 4x - 7x^2$ 的像.

9. 设 \mathcal{A} 是 n 维线性空间 V 的一个线性变换, 对某个 $\xi \in V$ 有 $\mathcal{A}^{k-1}(\xi) \neq 0, \mathcal{A}^k(\xi) = 0$, 试证: $\xi, \mathcal{A}(\xi), \mathcal{A}^2(\xi), \dots, \mathcal{A}^{k-1}(\xi)$ 线性无关.

10. 若 n 维线性空间中线性变换 \mathcal{A} 使得对于 V 中的任何向量 ξ 都有 $\mathcal{A}^{n-1}(\xi) \neq 0, \mathcal{A}^n(\xi) = 0$, 求 \mathcal{A} 在某一基下的矩阵表示.

11. 在 $P[x]_3$ 中, 已知线性变换

$$\mathcal{D}f(x) = \frac{df(x)}{dx}, \quad f(x) \in P[x]_3$$

求 \mathcal{D} 在下列基下的矩阵:

(1) $1, x-2, (x-2)^2, (x-2)^3$;

(2) $1, 1+x, 1+x+x^2, 1+x+x^2+x^3$.

12. 在 $\mathbb{R}^{2 \times 2}$ 空间中, 线性变换 \mathcal{A} :

$$\mathcal{A}(X) = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} X \begin{bmatrix} -4 & 0 \\ 1 & 4 \end{bmatrix}, \quad X \in \mathbb{R}^{2 \times 2}.$$

求 \mathcal{A} 在基 $\alpha_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \alpha_2 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \alpha_3 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \alpha_4 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ 下的矩阵表示.

13. 设 \mathbb{R}^3 中, 线性变换 \mathcal{A} 为: $\mathcal{A}(\alpha_i) = \beta_i, i=1, 2, 3$, 其中 $\alpha_1 = (1, 0, -1)^T, \alpha_2 = (2, 1, 1)^T, \alpha_3 = (1, 1, 1)^T$, 与 $\beta_1 = (0, 1, 1)^T, \beta_2 = (-1, 1, 0)^T, \beta_3 = (1, 2, 1)^T$, 求:

(1) \mathcal{A} 在基 $\alpha_1, \alpha_2, \alpha_3$ 下的矩阵;

(2) \mathcal{A} 在标准基 $\varepsilon_1 = (1, 0, 0)^T, \varepsilon_2 = (0, 1, 0)^T, \varepsilon_3 = (0, 0, 1)^T$ 下的矩阵.

14. 在 $P[x]_4$ 中定义线性变换 \mathcal{A} : $\forall f(x) \in P[x]_4, \mathcal{A}[f(x)] = f'(x) - f(x)$.

(1) 求 \mathcal{A} 在基 $1, x, x^2, x^3$ 下的矩阵;

(2) 求 \mathcal{A} 在基 $1, 1+x, x+x^2, x^2+x^3$ 下的矩阵.

15. 在 $\mathbb{R}^{2 \times 2}$ 中定义线性变换 \mathcal{A} : $\forall X \in \mathbb{R}^{2 \times 2}, \mathcal{A}(X) = AX - XA$, 其中矩阵 $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, 求 \mathcal{A} 在标准基

$E_{11}, E_{12}, E_{21}, E_{22}$ 下的矩阵.

16. 求下列线性变换对所指定基底的矩阵:

(1) \mathbb{R}^3 中将向量投影到 xOy 平面上的线性变换 \mathcal{A} , 对基底 i, j, k (直角坐标轴上的单位向量的矩阵);

(2) 上述线性变换对基底 $\alpha = i, \beta = j, \gamma = i + j + k$ 的矩阵;

(3) \mathbb{R}^3 中的线性变换 \mathcal{A} : $\mathcal{A}[(a_1, a_2, a_3)] = (2a_1 - a_2, a_2 + a_3, a_1)$, 对基底 $i = (1, 0, 0), j = (0, 1, 0), k = (0, 0, 1)$ 的矩阵;

(4) 由 6 个函数

$$x_1 = e^{at} \cos bt, \quad x_2 = e^{at} \sin bt, \quad x_3 = te^{at} \cos bt,$$

$$x_4 = te^{at} \sin bt, \quad x_5 = \frac{1}{2} t^2 e^{at} \cos bt, \quad x_6 = \frac{1}{2} t^2 \sin bt$$

的所有实系数线性组合构成实数域 \mathbb{R} 上的一个 6 维线性空间

$$V^6 = \text{Span}(x_1, x_2, x_3, x_4, x_5, x_6).$$

求微分变换 \mathcal{D} 在基 $x_1, x_2, x_3, x_4, x_5, x_6$ 下的矩阵.

17. 设三维线性空间 V^3 上的线性变换 \mathcal{A} 在基 e_1, e_2, e_3 下的矩阵为

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

- (1) 求 \mathcal{A} 在 e_3, e_2, e_1 下的矩阵;
- (2) 求 \mathcal{A} 在 e_1, ke_2, e_3 下的矩阵, $k \neq 0$;
- (3) 求 \mathcal{A} 在 $e_1 + e_2, e_2, e_3$ 下的矩阵.

18. 在 \mathbb{R}^3 中, 设 $\alpha = (x_1, x_2, x_3)$, 定义线性变换 \mathcal{A} 为

$$\mathcal{A}(\alpha) = (2x_1 - x_2, x_2 + x_3, x_1).$$

试求 \mathcal{A} 在自然基 $\varepsilon_1 = (1, 0, 0), \varepsilon_2 = (0, 1, 0), \varepsilon_3 = (0, 0, 1)$ 下的矩阵.

19. 设 $\mathbb{R}^3 \rightarrow \mathbb{R}^2$ 的线性算子 \mathcal{A} 定义为

$$\mathcal{A}[(a_1, a_2, a_3)] = (2a_1 + 3a_2 - a_3, a_1 + a_3).$$

试求 \mathcal{A} 在基偶 $\varepsilon_1 = (1, 0, 0), \varepsilon_2 = (0, 1, 0), \varepsilon_3 = (0, 0, 1)$ 及 $\varepsilon'_1 = (1, 0), \varepsilon'_2 = (0, 1)$ 下的矩阵.

20. 在 V^3 中, 设线性变换 \mathcal{A} 在基 $\eta_1 = (-1, 1, 1), \eta_2 = (1, 0, -1), \eta_3 = (0, 1, 1)$ 下的矩阵为

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

求 \mathcal{A} 在自然基 $\varepsilon_1, \varepsilon_2, \varepsilon_3$ 下的矩阵 B .

21. 在 V^3 中给定两组基

$$\begin{aligned} e_1 &= (1, 0, 1), & e_2 &= (2, 1, 0), & e_3 &= (1, 1, 1), \\ \eta_1 &= (1, 2, -1), & \eta_2 &= (2, 2, -1), & \eta_3 &= (2, -1, -1), \end{aligned}$$

定义线性变换 $\mathcal{A}(e_i) = \eta_i (i=1, 2, 3)$.

- (1) 写出由基 $\{e_i\}$ 到基 $\{\eta_i\} (i=1, 2, 3)$ 的过渡矩阵;
- (2) 写出 \mathcal{A} 在基 $\{e_i\}$ 下的矩阵;
- (3) 写出 \mathcal{A} 在基 $\{\eta_i\}$ 下的矩阵.

22. 在 $\mathbb{R}^{2 \times 2}$ 中定义下列线性变换:

$$\begin{aligned} \mathcal{A}_1(A) &= \begin{bmatrix} a & b \\ c & d \end{bmatrix} A, & \mathcal{A}_2(A) &= A \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \\ \mathcal{A}_3(A) &= \begin{bmatrix} a & b \\ c & d \end{bmatrix} A \begin{bmatrix} a & b \\ c & d \end{bmatrix}, & \mathcal{A}_4(A) &= \begin{bmatrix} a & 2b \\ 2c & a \end{bmatrix} A, \end{aligned}$$

其中 $A \in \mathbb{R}^{2 \times 2}, a, b, c, d$ 都是实数.

- (1) 求 $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$ 在基 E_{ij} (第 i 行、第 j 列元素为 1, 其余为 0 的二阶方阵) 下的矩阵;
- (2) 求 \mathcal{A}_4 在基

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

下的矩阵.

23. 设线性空间 V^3 的两组基为 (I) e_1, e_2, e_3 ; (II) $\varepsilon_1, \varepsilon_2, \varepsilon_3$. 由基 (I) 变到基 (II) 的过渡矩阵

$$C = \begin{pmatrix} 1 & 0 & 1 \\ 0 & -1 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \text{ 线性变换 } \mathcal{A} \text{ 满足}$$

$$\begin{cases} \mathcal{A}(e_1 + 2e_2 + 3e_3) = e_1 + e_2, \\ \mathcal{A}(2e_1 + e_2 + 2e_3) = e_2 + e_3, \\ \mathcal{A}(e_1 + 3e_2 + 4e_3) = e_1 + e_3. \end{cases}$$

- (1) 求 \mathcal{A} 在基(II)下的矩阵 A ;
 (2) 求 $\mathcal{A}(e_1)$ 在基(I)下的坐标.

24. 设 $P[x]_2$ 的两组基为

$$(I) f_1(x) = 1 + 2x^2, \quad f_2(x) = x + 2x^2, \quad f_3(x) = 1 + 2x + 5x^2;$$

$$(II) g_1(x) = 1 - x, \quad g_2(x) = 1 + x^2, \quad g_3(x) = x + 2x^2.$$

线性变换 \mathcal{A} 满足

$$\mathcal{A}[f_1(x)] = 2 + x^2, \quad \mathcal{A}[f_2(x)] = x, \quad \mathcal{A}[f_3(x)] = 1 + x + x^2.$$

- (1) 求 \mathcal{A} 在基(II)下的矩阵 A ;
 (2) 设 $f(x) = 1 + 2x + 3x^2$, 求 $\mathcal{A}[f(x)]$.

25. 设 \mathcal{A} 是数域 P 上的 n 维线性空间 V^n 的一个线性变换. 证明: 如果 \mathcal{A} 在任意一组基下的矩阵都相同, 则 \mathcal{A} 是数乘变换.

26. 在 \mathbb{R}^3 中, 线性变换 \mathcal{A} 在基 $e_1 = (1, 0, 1), e_2 = (1, 1, 0), e_3 = (0, 1, 1)$ 下的矩阵为

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

线性变换 \mathcal{B} 在基 $\eta_1 = (1, 0, 0), \eta_2 = (1, 2, 0), \eta_3 = (1, 2, 3)$ 下的矩阵为

$$B_1 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

求 $\mathcal{A} + \mathcal{B}, \mathcal{A}\mathcal{B}, \mathcal{B}\mathcal{A}, \mathcal{B}(\mathcal{A} + \mathcal{B})$ 在基 $\{e_i\} (i=1, 2, 3)$ 下的矩阵.

27. 设 \mathcal{A} 是线性空间 \mathbb{R}^3 上的线性变换, 它在 \mathbb{R}^3 中基 $\alpha_1, \alpha_2, \alpha_3$ 下的矩阵表示为

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 3 \\ 2 & 1 & 5 \end{bmatrix}.$$

- (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;
 (2) 求 \mathcal{A} 在基 $\alpha_1, \alpha_2, \alpha_3$ 下的核与值域.

28. 设 \mathbb{R}^3 中的线性变换 \mathcal{A} 为

$$\mathcal{A}[(x, y, z)^T] = (x + y - z, y + z, x + 2y)^T.$$

- (1) 求 \mathcal{A} 的值域 $\mathcal{A}(V)$ 的维数及一组基;
 (2) 求 \mathcal{A} 的核 $\mathcal{A}^{-1}(0)$ 的维数及一组基.

29. 设 V^3 中从基 $\alpha_1, \alpha_2, \alpha_3$ 到基 $\beta_1, \beta_2, \beta_3$ 的过渡矩阵

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

线性变换满足

$$\begin{cases} \mathcal{A}(\alpha_1 + 2\alpha_2 + 3\alpha_3) = \beta_1 + \beta_2, \\ \mathcal{A}(2\alpha_1 + \alpha_2 + 2\alpha_3) = \beta_2 + \beta_3, \\ \mathcal{A}(\alpha_1 + 3\alpha_2 + 4\alpha_3) = \beta_1 + \beta_3. \end{cases}$$

- (1) 求 \mathcal{A} 在 $\alpha_1, \alpha_2, \alpha_3$ 下的矩阵 A ;
 (2) 求 β_1 的像 $\mathcal{A}(\beta_1)$ 在基 $\alpha_1, \alpha_2, \alpha_3$ 下的坐标.

30. 设 \mathcal{A} 是 n 维空间 V^n 的线性变换, $\alpha \in V^n, \mathcal{A}^{n-1}(\alpha) \neq \mathbf{0}, \mathcal{A}^n(\alpha) = \mathbf{0}$.

- (1) 证明: $\alpha, \mathcal{A}(\alpha), \dots, \mathcal{A}^{n-1}(\alpha)$ 是 V^n 的基;
 (2) 求 \mathcal{A} 在基 $\alpha, \mathcal{A}(\alpha), \dots, \mathcal{A}^{n-1}(\alpha)$ 下的矩阵;

31. 设 \mathcal{A} 是 \mathbb{R}^n 的线性变换, $\mathbf{X} = (x_1, x_2, \dots, x_n)^\top \in \mathbb{R}^n$,

$$\mathcal{A}(\mathbf{X}) = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \mathbf{X}.$$

- (1) 证明: $\mathcal{A}^n = \mathbf{0}$;
 (2) 求 \mathcal{A} 的核 $N(\mathcal{A})$ 及像空间 $R(\mathcal{A})$ 的基和维数.

32. 设线性算子 $\mathcal{A}: \mathbb{R}^4 \rightarrow \mathbb{R}^3$, 有

$$\mathcal{A}[(x_1, x_2, x_3, x_4)^\top] = (x_1 - x_2 + x_3 + x_4, x_1 + 2x_2 - x_4, x_1 + x_2 + 3x_3 - x_4)^\top,$$

求 \mathcal{A} 的 $N(\mathcal{A})$ 及 $R(\mathcal{A})$.

33. 设 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是线性空间 V^n 的基, \mathcal{A} 是 V^n 的一个线性变换, 证明: \mathcal{A} 是可逆线性变换的充分必要条件是基像 $\mathcal{A}(\alpha_1), \mathcal{A}(\alpha_2), \dots, \mathcal{A}(\alpha_n)$ 线性无关.

34. 设 $\alpha_1, \alpha_2, \dots, \alpha_k$ 和 $\beta_1, \beta_2, \dots, \beta_k$ 是 n 维线性空间 V^n 的两个线性无关组, 证明一定存在 V^n 上的可逆线性变换 \mathcal{A} , 使 $\mathcal{A}(\alpha_i) = \beta_i, i=1, 2, \dots, k$.

35. 设 W_1, W_2 是线性空间 V^n 的两个子空间, $W_1 \cap W_2 = \{\mathbf{0}\}, \dim W_1 + \dim W_2 = n$, 证明存在线性变换 \mathcal{A} , 使得

$$R(\mathcal{A}) = W_1, \quad N(\mathcal{A}) = W_2.$$

36. 已知 $\mathbb{R}^{n \times n}$ 中线性变换 \mathcal{A} 为: $\forall \mathbf{X} \in \mathbb{R}^{n \times n}$,

$$\mathcal{A}(\mathbf{X}) = \mathbf{X} - \mathbf{X}^\top,$$

求 \mathcal{A} 的值域 $R(\mathcal{A})$ 的基及维数.

37. 设 $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$, 证明: $\text{adj}(\mathbf{AB}) = \text{adj}(\mathbf{B}) \cdot \text{adj}(\mathbf{A})$.

38. 证明: 对任意 n 阶矩阵 \mathbf{A} , 有 $\text{rank}(\mathbf{A}^n) = \text{rank}(\mathbf{A}^{n+1})$.

39. 设 ω 是 n 次本原单位根 (可设 $\omega = e^{2\pi i/n} = \cos \frac{2\pi}{n} + i \sin \frac{2\pi}{n}$), 试求傅里叶 (Fourier) 矩阵

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \cdots & \omega^{(n-1)(n-1)} \end{bmatrix}$$

的逆矩阵.

40. 设 n 阶矩阵 \mathbf{A} 可逆, \mathbf{x} 与 \mathbf{y} 是 n 维列向量. 如果 $(\mathbf{A} + \mathbf{xy}^H)^{-1}$ 可逆, 证明 Sherman-Morrison 公式:

$$(\mathbf{A} + \mathbf{xy}^H)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1} \mathbf{xy}^H \mathbf{A}^{-1}}{1 + \mathbf{y}^H \mathbf{A}^{-1} \mathbf{x}}.$$

41. 设矩阵 \mathbf{A} 与 \mathbf{B} 相似, \mathbf{C} 与 \mathbf{D} 相似, 试证明 $\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{bmatrix}$ 与 $\begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}$ 相似.

42. 证明: $\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$.

43. 设 $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$, 证明: 若 $\mathbf{AB} = \mathbf{0}$, 则 $\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) \leq n$.

44. 设 \mathbf{A}, \mathbf{B} 是使积 \mathbf{AB} 有定义的任意矩阵, 证明:

$$\text{rank}(\mathbf{AB}) \leq \min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\}.$$

45. 设 $A \in \mathbb{R}^{m \times n}$, 证明:

$$\text{rank}(A) = \text{rank}(AA^T) = \text{rank}(A^T A).$$

46. 证明: $A^T A X = 0$ 的充分必要条件是 $A X = 0$.

47. 设 $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times s}$, 证明:

(1) 若 $\text{rank}(A) = n$, 则 $\text{rank}(AB) = \text{rank}(B)$;

(2) 若 $\text{rank}(B) = n$, 则 $\text{rank}(AB) = \text{rank}(A)$.

48. 举出 2×2 矩阵 A 与 B 的例子, 使得

(1) $\text{rank}(A+B) < \text{rank}(A)$ (或 $\text{rank}(B)$);

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

(3) $\text{rank}(A+B) > \text{rank}(A)$ (或 $\text{rank}(B)$).

49. 设 $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times m}$, 证明: 当 $m > n$ 时, 方阵 $C = AB$ 为降秩的.

50. 设 $\beta_1, \beta_2, \dots, \beta_m$ 线性无关, 且

$$\begin{aligned} \xi_i &= a_{1i}\beta_1 + a_{2i}\beta_2 + \dots + a_{mi}\beta_m \\ &= (\beta_1, \beta_2, \dots, \beta_m) \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{bmatrix}, \quad i = 1, 2, \dots, s. \end{aligned}$$

试证: 向量 $\xi_1, \xi_2, \dots, \xi_s$ 的秩 = 矩阵 $(a_{ij})_{m \times s}$ 的秩.

51. 试证: (1) $\text{tr}(AB)^k = \text{tr}(BA)^k, k=1, 2, \dots$, 其中 $\text{tr}(A)$ 表示矩阵 A 的迹.

(2) $\text{tr}(A^k) = \sum_{i=1}^n \lambda_i^k, \lambda_i$ 是 A 的特征值.

(3) 若 $P^{-1}AP = B$, 则 $\text{tr}(A) = \text{tr}(B) = \sum_{i=1}^n \lambda_i$.

52. 设 A 与 B 分别为 $m \times n, n \times m$ 矩阵, 证明:

$$\lambda^m |\lambda I_n - BA| = \lambda^n |\lambda I_m - AB|,$$

其中 I_n, I_m 为单位矩阵.

53. 设 \mathcal{A} 是数域 \mathbb{C} 上线性空间 V^3 的线性变换, 已知 \mathcal{A} 在 V^3 的基 e_1, e_2, e_3 下的矩阵

$$A = \begin{bmatrix} 3 & 1 & 0 \\ -4 & -1 & 0 \\ 4 & -8 & -2 \end{bmatrix}.$$

求 \mathcal{A} 的特征值与特征向量.

54. 求复数域 \mathbb{C} 上的线性空间 V 上的线性变换 \mathcal{A} 的特征值与特征向量. 已知 \mathcal{A} 在一组基下的矩阵为

$$(1) \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix};$$

$$(2) \begin{bmatrix} 0 & 2 & 1 \\ -2 & 0 & 3 \\ -1 & -3 & 0 \end{bmatrix};$$

$$(3) \begin{bmatrix} 3 & 1 & 0 \\ -4 & -1 & 0 \\ 4 & -8 & -2 \end{bmatrix};$$

$$(4) \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

55. 在上题中, 哪些是非亏损矩阵(单纯矩阵)? 对于非亏损矩阵, 求出可逆矩阵 P , 使得 $P^{-1}AP$ 为对角阵 A .

56. 设 \mathcal{A} 是线性空间 V 上的线性变换. 证明: \mathcal{A} 可逆的充分必要条件是 \mathcal{A} 没有等于零的特征值.

57. 证明: 不论 \mathbf{A}, \mathbf{B} 为怎样的两个 n 阶方阵, \mathbf{AB} 与 \mathbf{BA} 有相同的特征多项式, 因此有相同的特征值和迹.

58. 设 $\mathbf{AB}=\mathbf{BA}$, 证明: \mathbf{A} 与 \mathbf{B} 有公共的特征向量.

59. 设 V 是复数域 \mathbb{C} 上的线性空间, \mathcal{A} 是 V 的一个线性变换. 已知 \mathcal{A} 在基 $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3$ 下的矩阵为

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & -2 \\ 2 & 5 & -4 \\ -2 & -4 & 5 \end{bmatrix}.$$

求 \mathcal{A} 的全部特征值与特征向量.

60. 设 \mathcal{A} 是数域上的线性空间 V 的线性变换, $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ 分别为 \mathcal{A} 的三个互不相同特征值 $\lambda_1, \lambda_2, \lambda_3$ 对应的特征向量.

(1) 证明: $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ 是线性无关的;

(2) 证明: $\mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3$ 不是 \mathbf{A} 的特征向量.

61. 已知三阶矩阵

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

试求 \mathbf{A} 的伴随矩阵 \mathbf{A}^* 的特征值与特征向量.

62. 设 n 阶矩阵

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & \cdots & 2 \\ 2 & 2 & \cdots & 2 \\ \vdots & \vdots & & \vdots \\ 2 & 2 & \cdots & 2 \end{bmatrix}.$$

求 \mathbf{A} 的特征值与特征向量.

63. 已知两个 n 维列向量 $\boldsymbol{\alpha} = (a_1, a_2, \dots, a_n)^\top, \boldsymbol{\beta} = (b_1, b_2, \dots, b_n)^\top$ 都是非零列向量, 且 $\boldsymbol{\alpha}^\top \boldsymbol{\beta} = 0$, 若 $\mathbf{A} = \boldsymbol{\alpha} \boldsymbol{\beta}^\top$, 求 \mathbf{A} 的特征值与特征向量.

64. 设矩阵

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & 2 \\ 2 & 2 & 3 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{P}^{-1} \mathbf{A}^* \mathbf{P} = \mathbf{B}.$$

求 $\mathbf{B} + 4\mathbf{I}$ 的特征值与特征向量, 其中 \mathbf{A}^* 为 \mathbf{A} 的伴随矩阵.

65. 设 n 阶方阵 $\mathbf{A} = (a_{ij})_{n \times n}$, 且 $\sum_{j=1}^n |a_{ij}| < 1, i=1, 2, \dots, n$, 证明: \mathbf{A} 的每一个特征值 λ 的绝对值 $|\lambda| < 1$.

66. 设三阶方阵

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 1 \\ x & 4 & y \\ -3 & -3 & 5 \end{bmatrix}$$

的二重特征值 $\lambda=2$ 对应有两个线性无关的特征向量.

(1) 求 x, y ;

(2) 求 \mathbf{P} , 使 $\mathbf{P}^{-1} \mathbf{A} \mathbf{P} = \boldsymbol{\Lambda}$.

67. 设 n 阶方阵 \mathbf{A} 有 n 个相异的特征值, 若有 n 阶矩阵 \mathbf{B} , 使 $\mathbf{AB}=\mathbf{BA}$, 则 \mathbf{B} 可对角化.

68. 设 a_1, a_2 是 $A_{n \times n}$ 的两个不同特征值, 且有

$$\text{rank}(a_1 \mathbf{I} - \mathbf{A}) + \text{rank}(a_2 \mathbf{I} - \mathbf{A}) = n.$$

证明: A 可以对角化.

69. 设 $A^2 = I$, 试证: A 的特征值只能为 ± 1 ; 若特征值都等于 1, 则 $A = I$.

70. 若 A, B 均为 n 阶方阵, 且有一个可逆, 证明: AB 与 BA 相似, 且有相同的特征多项式.

71. 证明: 幂等矩阵 A (即 $A^2 = A$) 与对角矩阵相似, 并且 A 相似于 $\text{diag}(I_r, 0)$, 其中 $\text{rank}(A) = r$.

72. 若方阵 $A \neq 0$, 但 $A^k = 0$ (k 为某一正整数), 则 A 不可能相似于对角矩阵.

73. 设 x 是 A 的特征向量, 证明: x 也是 $f(A)$ 的特征向量 (其中 $f(x)$ 为 x 的任一多项式). 对应于 x 的 A 的特征值和 $f(A)$ 的特征值有什么关系?

74. 设 \mathbb{R}^3 中有向量 $\alpha = (x_1, x_2, x_3)$, 线性变换定义为

$$\mathcal{A}(\alpha) = (-2x_2 - 2x_3, -2x_1 + 3x_2 - x_3, -2x_1 - x_2 + 3x_3).$$

求 \mathbb{R}^3 中的一组基, 使 \mathcal{A} 在该基下的矩阵为对角矩阵.

75. 在 $P[x]_2$ 中, 设 $f(x) = k_1 + k_2x + k_3x^2$, 线性变换 \mathcal{A} 为

$$\mathcal{A}[f(x)] = (k_2 + k_3) + (k_1 + k_3)x + (k_1 + k_2)x^2.$$

(1) 试写出 \mathcal{A} 在基 $1, x, x^2$ 下的矩阵 A ;

(2) 求 $P[x]_2$ 的一组基, 使 A 在该基下的矩阵为对角矩阵.

76. 已知 $B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, 线性空间 $V = \{A = (a_{ij})_{2 \times 2} \mid a_{11} + a_{22} = 0, a_{ij} \in \mathbb{R}\}$ 的变换 \mathcal{A} 为: $\mathcal{A}(A) = B^T A -$

$A^T B (A \in V)$.

(1) 验证: \mathcal{A} 是线性变换;

(2) 求 V 的一组基, 使 \mathcal{A} 在该基下的矩阵为对角矩阵.

77. 设线性空间 V^n 的线性变换 \mathcal{A}_1 与 \mathcal{A}_2 满足

$$\mathcal{A}_1 \mathcal{A}_2 = \mathcal{A}_1 + \mathcal{A}_2.$$

证明: (1) 1 不是 \mathcal{A}_1 的特征值;

(2) 若 \mathcal{A}_1 的 n 个特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 互不相同, 则存在 V^n 的一组基, 使 \mathcal{A}_1 与 \mathcal{A}_2 在该基下的矩阵都是对角矩阵.

78. 已知 n 阶矩阵 A 满足 $A^2 = kA$ ($k \neq 0$), 试证: A 相似于对角矩阵.

79. 设矩阵 A 和 B 相似, 其中

$$A = \begin{bmatrix} 2 & -1 & 0 \\ 0 & x & 0 \\ 1 & -1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 2 & 0 \\ -1 & y & 0 \\ -2 & -2 & 1 \end{bmatrix}.$$

(1) 求 x, y 的值;

(2) 证明: A 和 B 均可相似对角化.

80. 设 A 为一个 n 阶矩阵且满足 $A^2 - 5A + 6I = 0$, 证明: A 相似于一个对角矩阵.

81. 设矩阵 A 满足方程 $A^2 - A + 2I = 0$, 问 A 可以对角化吗? 为什么? 将本题一般化有什么结论.

82. 设 $A \in \mathbb{R}^{n \times n}$, $A^2 = A$, 且 $\text{rank}(A) = r$.

(1) 证明 A 可对角化;

(2) 证明有 $\text{tr}(A) = r$;

(3) 证明 \mathbb{R}^n 可分解为 A 的两个特征子空间的直和;

(4) 求 $|2I - A|$.

83. 设 $\mathcal{A}_1, \mathcal{A}_2$ 是数域 \mathbb{C} 上的线性空间 V^n 的线性变换, 且 $\mathcal{A}_1 \mathcal{A}_2 = \mathcal{A}_2 \mathcal{A}_1$, 证明: 如果 λ_0 是 \mathcal{A}_1 的特征值, 那么 V_{λ_0} 是 \mathcal{A}_2 的不变子空间.

84. 已知 e_1, e_2, e_3, e_4 是四维线性空间 V^4 的一组基, V^4 的线性变换 \mathcal{A} 在这组基下的矩阵为

$$A = \begin{bmatrix} 1 & 0 & 2 & 1 \\ -1 & 2 & 1 & 3 \\ 1 & 2 & 5 & 5 \\ 2 & -2 & 1 & -2 \end{bmatrix}.$$

- (1) 求 \mathcal{A} 在基 $e'_1 = e_1 - 2e_2 + e_4, e'_2 = 3e_2 - e_3 - e_4, e'_3 = e_3 + e_4, e'_4 = 2e_4$ 下的矩阵;
 (2) 求 \mathcal{A} 的值域与核;
 (3) 在 \mathcal{A} 的核中选一组基, 把它扩充成 V^4 的一组基, 并求 \mathcal{A} 在这组基下的矩阵;
 (4) 在 \mathcal{A} 的值域中选一组基, 把它扩充成 V^4 的一组基, 并求 \mathcal{A} 在这组基下的矩阵.
85. 设 \mathcal{A}, \mathcal{B} 是 \mathbb{R}^3 的两个线性变换, 满足

$$\mathcal{A}[(a_1, a_2, a_3)] = (a_1 + a_2 + a_3, 0, 0), \quad \mathcal{B}[(a_1, a_2, a_3)] = (a_2, a_3, a_1).$$

试证: $\mathcal{A} + \mathcal{B}$ 的像子空间(值域)是 \mathbb{R}^3 , 即 $(\mathcal{A} + \mathcal{B})(\mathbb{R}^3) = \mathbb{R}^3$.

86. 设 \mathcal{A} 是 \mathbb{R}^3 的线性变换, 定义为 $\mathcal{A}[(a_1, a_2, a_3)] = (0, a_1, a_2)$, 求 \mathcal{A}^2 的像子空间 $R(\mathcal{A}^2)$ (值域)与核子空间 $N(\mathcal{A}^2)$ 的基与维数.

87. (1) 证明: $\mathcal{A}[(x_1, x_2, \dots, x_n)] = (0, x_1, \dots, x_{n-1})$ 是线性空间 P^n 的线性变换, 且 $\mathcal{A}^n = \mathcal{O}$ (零变换);
 (2) 求 \mathcal{A} 的核 $\mathcal{A}^{-1}(\mathcal{O})$ 的维数及像集合(值域) $\mathcal{A}(V)$ 的维数.
88. 设 $\mathcal{A}^2 = \mathcal{A}, \mathcal{B}^2 = \mathcal{B}$, 证明:
 (1) \mathcal{A} 与 \mathcal{B} 有相同的值域 $\Leftrightarrow \mathcal{A}\mathcal{B} = \mathcal{B}, \mathcal{B}\mathcal{A} = \mathcal{A}$;
 (2) \mathcal{A} 与 \mathcal{B} 有相同的核 $\Leftrightarrow \mathcal{A}\mathcal{B} = \mathcal{A}, \mathcal{B}\mathcal{A} = \mathcal{B}$.
89. 设 n 维线性空间 V 的一个线性变换 \mathcal{A} 有 n 个不同的特征值, 证明 \mathcal{A} 共有 2^n 个不变子空间.
90. 设 V 是数域 P 上的 n 维线性空间, \mathcal{A} 是 V 上的线性变换, \mathcal{A} 在基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 下的矩阵是

$$A = \begin{bmatrix} 3 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 3 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 3 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 3 \end{bmatrix}.$$

- (1) 证明: 如果 α_n 属于 \mathcal{A} 的不变子空间 W , 那么 $W = V$;
 (2) 证明: α_1 属于 \mathcal{A} 的任意一个非零不变子空间;
 (3) 证明: V 不能分解成 \mathcal{A} 的两个非凡不变子空间的直和;
 (4) 求 \mathcal{A} 的所有不变子空间.

1.2 内积空间上的等积变换

在线性空间中, 向量之间的运算只有加法和数乘, 统称为线性运算. 但是, 如果以解析几何中三维几何空间 \mathbb{R}^3 作为线性空间的一个模型, 我们会发现, \mathbb{R}^3 中诸如向量的长度、两个向量的夹角等度量概念在线性空间的理论中还未得到反映, 而这些度量性质在很多实际问题(包括几何问题)中却是很关键的. 因此, 有必要在一般的线性空间中引进内积运算, 从而导出内积空间的概念.

本节重点讨论实数域上的内积空间(即欧氏空间)以及几种重要的线性变换及其矩阵, 其中有正交变换(包括初等旋转变换和镜像变换)与正交矩阵、对称变换与对称矩阵等. 同时, 对复数域上的内积空间(即酉空间)以及酉变换、埃尔米特(Hermit)变换、正交投影变换等也将作简单介绍.

1.2.1 内积空间

在解析几何中,向量的长度与夹角等度量性质都可通过数量积(又称点积)来表达.假设 α, β 是 \mathbb{R}^3 中过原点的两个向量,则它们的数量积为

$$(\alpha, \beta) = \alpha \cdot \beta = |\alpha| |\beta| \cos\theta = x_1x_2 + y_1y_2 + z_1z_2, \quad (1.2.1)$$

其中 θ 是 α 与 β 的夹角, $(x_1, y_1, z_1), (x_2, y_2, z_2)$ 分别是 α, β 的坐标.

容易看出,式(1.2.1)定义的数量积具有如下的代数性质:

- (1) 对称性 $(\alpha, \beta) = (\beta, \alpha)$;
- (2) 可加性 $(\alpha_1 + \alpha_2, \beta) = (\alpha_1, \beta) + (\alpha_2, \beta)$;
- (3) 齐次性 $(k\alpha, \beta) = k(\alpha, \beta)$, k 为任何实数;
- (4) 非负性 $(\alpha, \alpha) \geq 0$; 当且仅当 $\alpha = \mathbf{0}$ 时, $(\alpha, \alpha) = 0$.

有了数量积的概念,向量的长度和夹角就可表示为

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

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

可见数量积的概念蕴含着长度和夹角的概念.因此,为了给抽象的线性空间引进长度、夹角等度量,我们可先以数量积所具备的4条代数性质为依据,在抽象的线性空间中引入与数量积相类似的概念,这就是内积的概念,并把定义了内积的线性空间叫做内积空间.不过,关于内积的定义要区别一下实数域和复数域,从而引出欧氏空间与酉空间.

1. 内积与欧几里得空间

定义 1.2.1 设 V 是实数域 \mathbb{R} 上的线性空间,对于 V 中任意两个向量 x, y (y 可以等于 x),如能给定某种规则使 x 与 y 对应着一个实数,记为 (x, y) ,并且满足以下条件:

- (1) $(x, y) = (y, x)$;
- (2) $(x + y, z) = (x, z) + (y, z)$;
- (3) $(kx, y) = k(x, y)$, $\forall k \in \mathbb{R}$;
- (4) $(x, x) \geq 0$, 当且仅当 $x = \mathbf{0}$ 时, $(x, x) = 0$;

则称该实数 (x, y) 是向量 x 与 y 的内积.

如此定义了内积的实线性空间 V 叫做欧几里得(Euclid)空间,简称欧氏空间(或实内积空间).

显然,欧氏空间与实线性空间的差别在于欧氏空间比实线性空间多定义了内积,或者说欧氏空间是一个特殊的实线性空间.

例如,在 n 维向量空间 \mathbb{R}^n 中,对于任意两个向量 $x = (a_1, a_2, \dots, a_n)$, $y = (b_1, b_2, \dots, b_n)$,我们规定

$$(x, y) = a_1b_1 + a_2b_2 + \dots + a_nb_n = \sum_{k=1}^n a_k b_k. \quad (1.2.2)$$

不难验证,这样确定的实数满足内积的4个条件,所以式(1.2.2)是 \mathbb{R}^n 中的内积,且是 \mathbb{R}^n 中最常用的内积定义.从而 \mathbb{R}^n 是 n 维的欧氏空间.

例 1.2.1 闭区间 $[a, b]$ ($b > a$) 上的实连续函数的全体按通常意义的加法和数乘运算

构成无穷维线性空间 $C[a, b]$. 对于函数 $f(x), g(x) \in C[a, b]$, 定义内积

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

则 $C[a, b]$ 构成欧氏空间(无穷维的).

事实上, 对于连续函数 $f(x), g(x)$, $\int_a^b f(x)g(x)dx$ 是唯一确定的实数. 同时, 满足

$$(1) (f, g) = \int_a^b f(x)g(x)dx = \int_a^b g(x)f(x)dx = (g, f);$$

$$(2) (kf, g) = \int_a^b kf(x)g(x)dx = k \int_a^b f(x)g(x)dx = k(f, g);$$

$$(3) (f+g, h) = \int_a^b (f(x)+g(x))h(x)dx \\ = \int_a^b f(x)h(x)dx + \int_a^b g(x)h(x)dx = (f, h) + (g, h);$$

$$(4) (f, f) = \int_a^b [f(x)]^2 dx \geq 0, \text{ 且 } \int_a^b f^2(x)dx = 0 \text{ 的充要条件为 } f(x) \equiv 0. \text{ 因此 } C[a, b]$$

是欧氏空间.

从内积的定义, 我们容易得到它的下面各基本性质:

$$(1) (x, ky) = k(x, y);$$

$$(2) (x, y+z) = (x, y) + (x, z);$$

$$(3) (x, \mathbf{0}) = (\mathbf{0}, x) = 0;$$

$$(4) \left(\sum_{i=1}^n \lambda_i x_i, \sum_{j=1}^m \mu_j y_j \right) = \sum_{i=1}^n \sum_{j=1}^m \lambda_i \mu_j (x_i, y_j). \quad (1.2.3)$$

因为由定义 1.2.1 有

$$(x, ky) = (ky, x) = k(y, x) = k(x, y);$$

$$(x, y+z) = (y+z, x) = (y, x) + (z, x) = (x, y) + (x, z);$$

$$(x, \mathbf{0}) = (x, 0y) = 0(x, y) = 0;$$

故性质(1)~性质(3)成立, 再用数学归纳法不难证明性质(4)也成立.

1) 向量的长度与夹角

与三维几何空间 \mathbb{R}^3 类似, 欧氏空间中抽象向量的长度(模)、夹角等度量可用内积来表示.

定义 1.2.2 非负实数 $\sqrt{(x, x)}$ 叫做向量 x 的长度或模, 记为 $|x|$. 长度等于 1 的向量叫做单位向量. 零向量的长度为 0.

这样定义的长度 $|x|$ 具有通常的性质: $|kx| = |k||x|$, 而且 $|x| = 0$ 的充要条件是 $x = \mathbf{0}$.

事实上, $|kx| = \sqrt{(kx, kx)} = \sqrt{k^2(x, x)} = |k||x|$, 而且

$$|x| = 0 \Leftrightarrow (x, x) = 0 \Leftrightarrow x = \mathbf{0}.$$

任何一个非零向量总可以化成单位向量. 因为对于任何 $x \neq \mathbf{0}$, 总有

$$\left| \frac{x}{|x|} \right| = \frac{1}{|x|} |x| = 1,$$

也就是说, $\frac{x}{|x|}$ 是一个单位向量, 即用向量 x 的长度 $|x|$ 去除 x , 得到一个与 x 同方向的单位向量. 通常称此过程为把 x 单位化或规范化.

在解析几何中,两个非零向量 x, y 的夹角 $\langle x, y \rangle$ 的余弦,可通过数量积(即内积)来表示

$$\cos \langle x, y \rangle = \frac{(x, y)}{|x| |y|}. \quad (1.2.4)$$

为了在抽象的欧氏空间中也能利用式(1.2.4)引入向量夹角的概念,我们必须证明不等式

$$\left| \frac{(x, y)}{|x| |y|} \right| \leq 1,$$

即

$$|(x, y)| \leq |x| |y| \quad (1.2.5)$$

对任意两个向量 x, y 均成立,且式(1.2.5)当且仅当 x, y 线性相关时,等号才成立. 式(1.2.5)就是所谓柯西-施瓦茨(Cauchy-Schwarz)不等式,又称为柯西-布涅柯夫斯基(Буниаковский)不等式.

事实上, x 与 y 有一个为零向量时,结论是显然的. 现设 x, y 都是非零向量,令 $u = x + ty, t$ 为实数域中任一数,则

$$(u, u) = (x + ty, x + ty) = (y, y)t^2 + 2(x, y)t + (x, x) \geq 0.$$

回忆初等代数中的知识,即如果实系数二次三项式

$$at^2 + 2bt + c, \quad a > 0,$$

对于任意实数 t 它都取非负值,则其系数之间必有判别式

$$\Delta = b^2 - ac \leq 0.$$

取 $a = (y, y), b = (x, y), c = (x, x)$, 得

$$(x, y)^2 \leq (x, x)(y, y),$$

即

$$(x, y)^2 \leq |x|^2 |y|^2,$$

从而有

$$|(x, y)| \leq |x| |y|.$$

等号成立的充要条件是 $u = 0$, 即 x 与 y 线性相关.

定义 1.2.3 非零向量 x 与 y 的夹角 $\langle x, y \rangle$ 规定为

$$\langle x, y \rangle = \arccos \frac{(x, y)}{|x| |y|}, \quad 0 \leq \langle x, y \rangle \leq \pi. \quad (1.2.6)$$

这个定义在形式上与解析几何中夹角的定义是完全一致的.

柯西-施瓦茨不等式(1.2.5)的几何解释是 $\langle x, y \rangle$ 满足

$$|\cos \langle x, y \rangle| \leq 1.$$

作为它的应用,请看下面的例子.

例 1.2.2 对于欧氏空间 \mathbb{R}^n 规定内积 $(x, y) = \sum_{i=1}^n x_i y_i$ 后,式(1.2.5)成为

$$\left| \sum_{i=1}^n x_i y_i \right| \leq \sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}. \quad (1.2.7)$$

对于欧氏空间 $C[a, b]$, 式(1.2.5)成为

$$\left| \int_a^b f(x)g(x)dx \right| \leq \sqrt{\int_a^b f^2(x)dx} \sqrt{\int_a^b g^2(x)dx}. \quad (1.2.8)$$

不等式(1.2.7)和(1.2.8)都是数学史上著名的不等式,其中式(1.2.8)又叫做施瓦茨不等式.

例 1.2.3 由柯西-施瓦茨不等式(1.2.5)又可推出如下的三角不等式

$$|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|. \quad (1.2.9)$$

事实上,因为

$$|\mathbf{x} + \mathbf{y}|^2 = (\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) = (\mathbf{x}, \mathbf{x}) + 2(\mathbf{x}, \mathbf{y}) + (\mathbf{y}, \mathbf{y}),$$

利用式(1.2.5)便得

$$(\mathbf{x}, \mathbf{x}) + 2(\mathbf{x}, \mathbf{y}) + (\mathbf{y}, \mathbf{y}) \leq (\mathbf{x}, \mathbf{x}) + 2\sqrt{(\mathbf{x}, \mathbf{x})}\sqrt{(\mathbf{y}, \mathbf{y})} + (\mathbf{y}, \mathbf{y}),$$

于是有

$$|\mathbf{x} + \mathbf{y}|^2 \leq (|\mathbf{x}| + |\mathbf{y}|)^2,$$

两边开方,便得式(1.2.9).

不等式(1.2.9)之所以被称为三角不等式,用三维空间向量解释,它表明三角形两边长之和不小于第三边长的事实.

不等式(1.2.9)还可推广到多个向量的情形,即

$$|\mathbf{x} + \mathbf{y} + \cdots + \mathbf{z}| \leq |\mathbf{x}| + |\mathbf{y}| + \cdots + |\mathbf{z}|. \quad (1.2.10)$$

由式(1.2.9)又可派生出以下两个不等式:

$$|\mathbf{x} - \mathbf{y}| \geq |\mathbf{x}| - |\mathbf{y}|;$$

$$|\mathbf{x} - \mathbf{z}| \leq |\mathbf{x} - \mathbf{y}| + |\mathbf{y} - \mathbf{z}|.$$

以后我们把 $|\mathbf{x} - \mathbf{y}|$ 称为向量 \mathbf{x} 与 \mathbf{y} 之间的距离.

2) 度量矩阵

这里将研究欧氏空间中向量的内积如何通过坐标来计算的问题,从而引出度量矩阵的概念.

假定 $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$ 是 n 维欧氏空间 V^n 的基.对于 V^n 的任意两个向量

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n,$$

$$\mathbf{y} = y_1\mathbf{e}_1 + y_2\mathbf{e}_2 + \cdots + y_n\mathbf{e}_n,$$

由内积性质(1.2.3)可得

$$(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^n x_i \mathbf{e}_i, \sum_{j=1}^n y_j \mathbf{e}_j \right) = \sum_{i,j=1}^n x_i y_j (\mathbf{e}_i, \mathbf{e}_j).$$

令 $a_{ij} = (\mathbf{e}_i, \mathbf{e}_j)$ ($i, j = 1, 2, \cdots, n$),并构造矩阵和列向量

$$\mathbf{A} = (a_{ij}) = \begin{pmatrix} (\mathbf{e}_1, \mathbf{e}_1) & (\mathbf{e}_1, \mathbf{e}_2) & \cdots & (\mathbf{e}_1, \mathbf{e}_n) \\ (\mathbf{e}_2, \mathbf{e}_1) & (\mathbf{e}_2, \mathbf{e}_2) & \cdots & (\mathbf{e}_2, \mathbf{e}_n) \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{e}_n, \mathbf{e}_1) & (\mathbf{e}_n, \mathbf{e}_2) & \cdots & (\mathbf{e}_n, \mathbf{e}_n) \end{pmatrix}, \quad (1.2.11)$$

$$\mathbf{X} = (x_1, x_2, \cdots, x_n)^T, \quad \mathbf{Y} = (y_1, y_2, \cdots, y_n)^T,$$

于是内积 (\mathbf{x}, \mathbf{y}) 可表示成:

$$(\mathbf{x}, \mathbf{y}) = \mathbf{X}^T \mathbf{A} \mathbf{Y}, \quad (1.2.12)$$

我们把式(1.2.11)中的 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 叫做基 $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$ 的度量矩阵,又叫做格拉姆(Gram)矩阵.

式(1.2.12)表明,向量 \mathbf{x} 与 \mathbf{y} 的内积可由 \mathbf{x}, \mathbf{y} 在某基下的坐标和度量矩阵来表示,因此,度量矩阵完全确定了内积.向量的长度与夹角等这些度量又可用内积来刻画,所以这就是度量矩阵的名词的涵义.

度量矩阵具有如下性质:

(1) 度量矩阵是对称正定矩阵.

事实上, 因为 $(e_i, e_j) = (e_j, e_i) (i, j = 1, 2, \dots, n)$, 所以有 $a_{ij} = a_{ji} (i, j = 1, 2, \dots, n)$, 即 A 是对称矩阵; 又因为对任意向量 $x \neq 0$, 从而 $X \neq 0$. 由式(1.2.12)知 $(x, x) = X^T A X > 0$, 故 A 是正定矩阵.

正因为如此, 我们常常用任意正定矩阵作为度量矩阵来规定内积.

(2) 两组不同基的度量矩阵是不同的, 但它们是相合的.

事实上, 设 e_1, e_2, \dots, e_n 和 e'_1, e'_2, \dots, e'_n 是 V^n 的两组基, 这两组基的度量矩阵分别是 A 和 B , 且由 e_1, e_2, \dots, e_n 到 e'_1, e'_2, \dots, e'_n 的过渡矩阵为 C , 即

$$(e'_1, e'_2, \dots, e'_n) = (e_1, e_2, \dots, e_n)C,$$

则 V_n 中任意两个向量 x, y 的内积可分别表示为

$$(x, y) = X^T A Y, \quad (x, y) = X'^T B Y',$$

其中 X 和 X' 分别是 x 在基 e_1, e_2, \dots, e_n 和 e'_1, e'_2, \dots, e'_n 下的坐标列向量; Y 和 Y' 分别是 y 在基 e_1, e_2, \dots, e_n 和 e'_1, e'_2, \dots, e'_n 下的坐标列向量.

根据坐标变换公式(1.1.5), 有

$$X = C X', \quad Y = C Y',$$

于是

$$(x, y) = X^T A Y = X'^T C^T A C Y' = X'^T B Y'$$

对任意 X' 和 Y' 都成立, 故得

$$B = C^T A C,$$

这就是说, 不同基的度量矩阵是相合的.

3) 正交性

(1) 正交的概念

在解析几何中, 两个非零向量正交的充要条件是它们夹角的余弦为零, 亦即它们的数量积为零. 内积既然是数量积的推广, 而且欧氏空间中两向量夹角的余弦仍以内积来定义, 即

$\cos \langle x, y \rangle = \frac{(x, y)}{|x||y|}$, 因而在欧氏空间中也可以借助内积来定义正交的概念.

定义 1.2.4 设 x, y 为欧氏空间的两个向量, 如果 $(x, y) = 0$, 则说 x 与 y 正交, 记为 $x \perp y$.

由 $(x, y) = 0$ 可知 $(y, x) = 0$, 所以 x 与 y 正交, y 与 x 也正交. 零向量与任何向量(包括与它自己)正交. 如果 $x \neq 0$, 由于 $(x, x) > 0$, 所以非零向量不会与自身正交, 换言之, 与自身正交的向量只能是零向量.

定理 1.2.1 如果向量 x 与 y 正交, 则有

$$|x + y|^2 = |x|^2 + |y|^2. \quad (1.2.13)$$

证明 因为 $|x + y|^2 = (x + y, x + y) = (x, x) + 2(x, y) + (y, y)$, 而 $(x, y) = 0$, 所以

$$|x + y|^2 = (x, x) + (y, y) = |x|^2 + |y|^2. \quad \text{证毕}$$

对于二维空间, 以上定理即是通常的商高定理: 直角三角形两直角边平方和等于斜边平方.

例 1.2.4 在 \mathbb{R}^2 中, 现给出两种不同的内积定义:

$$(\mathbf{x}, \mathbf{y})_1 = \mathbf{x}^T \mathbf{y}, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^2,$$

$$(\mathbf{x}, \mathbf{y})_2 = \mathbf{x}^T \mathbf{A} \mathbf{y}, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^2,$$

其中

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

由于在同一个线性空间 \mathbb{R}^2 里有两种不同的内积定义(其中 $(\mathbf{x}, \mathbf{y})_1$ 是常用的), 因此由 \mathbb{R}^2 产生两个欧氏空间, 分别记为 \mathbb{R}_1^2 与 \mathbb{R}_2^2 .

问: 向量 $\mathbf{x}_0 = (1, 1)^T, \mathbf{y}_0 = (-1, 1)^T$ 在这两个欧氏空间中是否正交?

解 由于

$$(\mathbf{x}_0, \mathbf{y}_0)_1 = (1, 1)(-1, 1)^T = 0, .$$

$$(\mathbf{x}_0, \mathbf{y}_0)_2 = (1, 1) \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = 1,$$

故 $\mathbf{x}_0, \mathbf{y}_0$ 在 \mathbb{R}_1^2 中正交, 而在 \mathbb{R}_2^2 中不正交.

此例表明, 向量正交与否, 与该欧氏空间的内积如何定义有关.

下面再把正交的概念推广到多个向量的情形.

定义 1.2.5 欧氏空间中一组非零向量, 如果它们两两正交, 则称其为一个正交向量组.

若 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ 是正交向量组, 则有

$$|\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_m|^2 = |\mathbf{x}_1|^2 + |\mathbf{x}_2|^2 + \dots + |\mathbf{x}_m|^2. \quad (1.2.14)$$

定理 1.2.2 如果 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ 是一组两两正交的非零向量, 则它们必是线性无关的.

证明 假定它们之间有线性关系

$$k_1 \mathbf{x}_1 + k_2 \mathbf{x}_2 + \dots + k_m \mathbf{x}_m = \mathbf{0},$$

下面要证明一切 $k_i (i=1, 2, \dots, m)$ 都必须为零. 为此, 用 $\mathbf{x}_i (i=1, 2, \dots, m)$ 与上式两端作内积, 得

$$(\mathbf{x}_i, k_1 \mathbf{x}_1 + k_2 \mathbf{x}_2 + \dots + k_m \mathbf{x}_m) = 0,$$

即

$$k_1 (\mathbf{x}_i, \mathbf{x}_1) + k_2 (\mathbf{x}_i, \mathbf{x}_2) + \dots + k_m (\mathbf{x}_i, \mathbf{x}_m) = 0.$$

由于 $i \neq j$ 时, $(\mathbf{x}_i, \mathbf{x}_j) = 0$, 故最后得

$$k_i (\mathbf{x}_i, \mathbf{x}_i) = 0, \quad i = 1, 2, \dots, m.$$

但 \mathbf{x}_i 是非零向量, 有 $(\mathbf{x}_i, \mathbf{x}_i) > 0$, 故必有

$$k_i = 0, \quad i = 1, 2, \dots, m.$$

所以, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ 线性无关. 证毕

反之, 从一组线性无关的向量出发, 必可构造出一组相同个数等价的两两正交的向量, 并且还可使每个新向量的长度(模)等于 1(即单位向量). 这种做法叫做线性无关向量组的正交规范化, 常用的方法是如下的施密特(Schmidt)方法.

(2) 施密特正交化方法

设 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ 是一组线性无关的向量. 施密特正交规范化步骤是先把它们正交化, 具体步骤为:

第一步 取 $\mathbf{y}'_1 = \mathbf{x}_1$ 作为正交向量组中的第一个向量.

第二步 令 $\mathbf{y}'_2 = \mathbf{x}_2 + k_{21}\mathbf{y}'_1$, 由正交条件 $(\mathbf{y}'_2, \mathbf{y}'_1) = 0$ 来决定待定常数 k_{21} . 由

$$(\mathbf{x}_2 + k_{21}\mathbf{y}'_1, \mathbf{y}'_1) = (\mathbf{x}_2, \mathbf{y}'_1) + k_{21}(\mathbf{y}'_1, \mathbf{y}'_1) = 0,$$

得

$$k_{21} = -\frac{(\mathbf{x}_2, \mathbf{y}'_1)}{(\mathbf{y}'_1, \mathbf{y}'_1)},$$

这样就得到两个正交的向量 $\mathbf{y}'_1, \mathbf{y}'_2$.

第三步 又令 $\mathbf{y}'_3 = \mathbf{x}_3 + k_{31}\mathbf{y}'_2 + k_{32}\mathbf{y}'_1$, 再由正交条件 $(\mathbf{y}'_3, \mathbf{y}'_2) = 0$ 及 $(\mathbf{y}'_3, \mathbf{y}'_1) = 0$ 来决定 k_{31}, k_{32} :

$$k_{31} = -\frac{(\mathbf{x}_3, \mathbf{y}'_2)}{(\mathbf{y}'_2, \mathbf{y}'_2)}, \quad k_{32} = -\frac{(\mathbf{x}_3, \mathbf{y}'_1)}{(\mathbf{y}'_1, \mathbf{y}'_1)},$$

到此我们已做出 3 个两两正交的向量 $\mathbf{y}'_1, \mathbf{y}'_2, \mathbf{y}'_3$.

第四步 如此继续进行, 一般式是

$$\begin{cases} \mathbf{y}'_m = \mathbf{x}_m + k_{m1}\mathbf{y}'_{m-1} + k_{m2}\mathbf{y}'_{m-2} + \cdots + k_{m,m-1}\mathbf{y}'_1, \\ k_{mi} = -\frac{(\mathbf{x}_m, \mathbf{y}'_{m-i})}{(\mathbf{y}'_{m-i}, \mathbf{y}'_{m-i})}, \quad i = 1, 2, \cdots, m-1, \end{cases} \quad (1.2.15)$$

直到 $m=n$. 这样得到的一组向量 $\mathbf{y}'_1, \mathbf{y}'_2, \cdots, \mathbf{y}'_n$ 显然是两两正交的.

再单位化, 即以 $|\mathbf{y}'_i|$ 除 $\mathbf{y}'_i (i=1, 2, \cdots, n)$, 就得到所要求的正交规范化的向量组

$$\mathbf{y}_i = \frac{1}{|\mathbf{y}'_i|} \mathbf{y}'_i, \quad i = 1, 2, \cdots, n.$$

例 1.2.5 试把向量组 $\mathbf{x}_1 = (1, 1, 0, 0), \mathbf{x}_2 = (1, 0, 1, 0), \mathbf{x}_3 = (-1, 0, 0, 1), \mathbf{x}_4 = (1, -1, -1, 1)$ 化为正交规范化向量组.

解 先正交化, 为此利用式(1.2.15)可得

$$\begin{aligned} \mathbf{y}'_1 &= \mathbf{x}_1 = (1, 1, 0, 0), \\ \mathbf{y}'_2 &= \mathbf{x}_2 - \frac{(\mathbf{x}_2, \mathbf{y}'_1)}{(\mathbf{y}'_1, \mathbf{y}'_1)} \mathbf{y}'_1 = \left(\frac{1}{2}, -\frac{1}{2}, 1, 0\right), \\ \mathbf{y}'_3 &= \mathbf{x}_3 - \frac{(\mathbf{x}_3, \mathbf{y}'_2)}{(\mathbf{y}'_2, \mathbf{y}'_2)} \mathbf{y}'_2 - \frac{(\mathbf{x}_3, \mathbf{y}'_1)}{(\mathbf{y}'_1, \mathbf{y}'_1)} \mathbf{y}'_1 = \left(-\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 1\right), \\ \mathbf{y}'_4 &= \mathbf{x}_4 - \frac{(\mathbf{x}_4, \mathbf{y}'_3)}{(\mathbf{y}'_3, \mathbf{y}'_3)} \mathbf{y}'_3 - \frac{(\mathbf{x}_4, \mathbf{y}'_2)}{(\mathbf{y}'_2, \mathbf{y}'_2)} \mathbf{y}'_2 - \frac{(\mathbf{x}_4, \mathbf{y}'_1)}{(\mathbf{y}'_1, \mathbf{y}'_1)} \mathbf{y}'_1 \\ &= (1, -1, -1, 1). \end{aligned}$$

再将 $\mathbf{y}'_1, \mathbf{y}'_2, \mathbf{y}'_3, \mathbf{y}'_4$ 单位化, 便有

$$\begin{aligned} \mathbf{y}_1 &= \frac{1}{|\mathbf{y}'_1|} \mathbf{y}'_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0\right), \\ \mathbf{y}_2 &= \frac{1}{|\mathbf{y}'_2|} \mathbf{y}'_2 = \left(\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, 0\right), \\ \mathbf{y}_3 &= \frac{1}{|\mathbf{y}'_3|} \mathbf{y}'_3 = \left(-\frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{3}{\sqrt{12}}\right), \\ \mathbf{y}_4 &= \frac{1}{|\mathbf{y}'_4|} \mathbf{y}'_4 = \left(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right). \end{aligned}$$

例 1.2.6 试把线性无关的向量组 $1, x, x^2, x^3$ 变为在 $[-1, 1]$ 上的正交多项式组. 在此, 内积定义为

$$(f, g) = \int_{-1}^1 f(x)g(x)dx.$$

解 取 $y_1 = 1$, 利用式(1.2.15)有

$$y_2 = x - \frac{(x, y_1)}{(y_1, y_1)}y_1 = x - \int_{-1}^1 x dx / \int_{-1}^1 1 \cdot dx = x,$$

$$y_3 = x^2 - \frac{(x^2, y_2)}{(y_2, y_2)}y_2 - \frac{(x^2, y_1)}{(y_1, y_1)}y_1 = x^2 - \frac{1}{3},$$

$$y_4 = x^3 - \frac{(x^3, y_3)}{(y_3, y_3)}y_3 - \frac{(x^3, y_2)}{(y_2, y_2)}y_2 - \frac{(x^3, y_1)}{(y_1, y_1)}y_1 = x^3 - \frac{3x}{5}.$$

将它们通分整理后得到的一组多项式

$$1, x, \frac{1}{3}(3x^2 - 1), \frac{1}{5}(5x^3 - 3x)$$

便是在 $[-1, 1]$ 上的正交多项式组. 实际上, 它们是勒让德(Legendre)多项式 $P_n(x)$ 中的前 4 个(常数因子不同):

$$P_0(x) = 1, \quad P_1(x) = x, \quad P_2(x) = \frac{1}{2}(3x^2 - 1), \quad P_3(x) = \frac{1}{2}(5x^3 - 3x).$$

(3) 标准正交基

解析几何中, 直角坐标系的概念推广到 n 维欧氏空间, 即为标准正交基. n 维欧氏空间既然是线性空间, 自然有一组 n 个线性无关向量作为它的基, 再由上一段的讨论知, 利用施密特正交化方法, 又可将这组基改造为一组由正交规范化向量组成的基. 因此, 有下面的定义.

定义 1.2.6 在欧氏空间 V^n 中, 由 n 个向量构成的正交向量组称为 V^n 的**正交基**. 由单位向量构成的正交基叫做**标准正交基**.

综上所述, 得到如下的定理.

定理 1.2.3 任何 n 维欧氏空间都有正交基和标准正交基.

例如, 在欧氏空间 \mathbb{R}^4 中, 自然基 $\boldsymbol{\varepsilon}_1 = (1, 0, 0, 0)$, $\boldsymbol{\varepsilon}_2 = (0, 1, 0, 0)$, $\boldsymbol{\varepsilon}_3 = (0, 0, 1, 0)$, $\boldsymbol{\varepsilon}_4 = (0, 0, 0, 1)$ 显然是一组标准正交基; 此外, 例 1.2.5 中得到的 y_1, y_2, y_3, y_4 也是 \mathbb{R}^4 的一组标准正交基, 可见欧氏空间的标准正交基不是惟一的.

例 1.2.7 已知欧氏空间 V 的某组基 $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4$ 的度量矩阵为

$$\boldsymbol{A} = \begin{pmatrix} 2 & 1 & 2 & 1 \\ 1 & 1 & 0 & 1 \\ 2 & 0 & 5 & 1 \\ 1 & 1 & 1 & 3 \end{pmatrix},$$

试求 V 的一组标准正交基.

解 根据度量矩阵的定义(1.2.11), 可以知道所给基 $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4$ 之间的内积, 例如, $(\boldsymbol{e}_1, \boldsymbol{e}_1) = 2, (\boldsymbol{e}_2, \boldsymbol{e}_2) = 1, \dots, (\boldsymbol{e}_4, \boldsymbol{e}_4) = 3.$

$$\begin{pmatrix} (\boldsymbol{e}_1, \boldsymbol{e}_1) & (\boldsymbol{e}_1, \boldsymbol{e}_2) & (\boldsymbol{e}_1, \boldsymbol{e}_3) & (\boldsymbol{e}_1, \boldsymbol{e}_4) \\ (\boldsymbol{e}_2, \boldsymbol{e}_1) & (\boldsymbol{e}_2, \boldsymbol{e}_2) & (\boldsymbol{e}_2, \boldsymbol{e}_3) & (\boldsymbol{e}_2, \boldsymbol{e}_4) \\ (\boldsymbol{e}_3, \boldsymbol{e}_1) & (\boldsymbol{e}_3, \boldsymbol{e}_2) & (\boldsymbol{e}_3, \boldsymbol{e}_3) & (\boldsymbol{e}_3, \boldsymbol{e}_4) \\ (\boldsymbol{e}_4, \boldsymbol{e}_1) & (\boldsymbol{e}_4, \boldsymbol{e}_2) & (\boldsymbol{e}_4, \boldsymbol{e}_3) & (\boldsymbol{e}_4, \boldsymbol{e}_4) \end{pmatrix} = \begin{pmatrix} 2 & 1 & 2 & 1 \\ 1 & 1 & 0 & 1 \\ 2 & 0 & 5 & 1 \\ 1 & 1 & 1 & 3 \end{pmatrix}. \quad (1.2.16)$$

再由施密特方法并借助式(1.2.16),先构造出一组正交基:

$$\begin{aligned} y'_1 &= e_1, \\ y'_2 &= e_2 - \frac{(e_2, y'_1)}{(y'_1, y'_1)} y'_1 = e_2 - \frac{1}{2} e_1, \\ y'_3 &= e_3 - \frac{(e_3, y'_2)}{(y'_2, y'_2)} y'_2 - \frac{(e_3, y'_1)}{(y'_1, y'_1)} y'_1 = e_3 + 2e_2 - 2e_1, \\ y'_4 &= e_4 - \frac{(e_4, y'_3)}{(y'_3, y'_3)} y'_3 - \frac{(e_4, y'_2)}{(y'_2, y'_2)} y'_2 - \frac{(e_4, y'_1)}{(y'_1, y'_1)} y'_1 \\ &= e_4 - e_3 - 3e_2 + 2e_1. \end{aligned}$$

最后,单位化(同理需借助式(1.2.16))得

$$\begin{aligned} y_1 &= \frac{\sqrt{2}}{2} e_1, \\ y_2 &= \frac{-\sqrt{2}}{2} e_1 + \sqrt{2} e_2, \\ y_3 &= -2e_1 + 2e_2 + e_3, \\ y_4 &= 2e_1 - 3e_2 - e_3 + e_4. \end{aligned}$$

这样的 y_1, y_2, y_3, y_4 便是 V 中的标准正交基.

定理 1.2.4 一组基为标准正交基的充分必要条件是它的度量矩阵为单位矩阵.

证明 设 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 为标准正交基,则由定义 1.2.6 有

$$(\varepsilon_i, \varepsilon_j) = \delta_{ij} = \begin{cases} 1, & \text{当 } i = j, \\ 0, & \text{当 } i \neq j, \end{cases}$$

这里的 δ_{ij} 叫做克罗内克(Kronecker)符号,于是由式(1.2.11)知,它的度量矩阵是单位矩阵.反之,如果以单位矩阵为度量矩阵,则由矩阵相等可得 $(\varepsilon_i, \varepsilon_j) = \delta_{ij}$,即 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 为标准正交基. 证毕

例如,在例 1.2.7 中原给定的基 e_1, e_2, e_3, e_4 的度量矩阵 A 不是单位矩阵,所以该基不是标准正交基;而后来用施密特法构造出的 y_1, y_2, y_3, y_4 ,不难验证它的度量矩阵是单位矩阵,因而它是标准正交基.

标准正交基的存在给我们研究问题带来极大的方便.例如,当欧氏空间 V^n 取基为任何一组标准正交基 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 时,那么根据式(1.2.12)知,向量 x 与 y 的内积总有最简单的形式:

$$(x, y) = X^T Y = x_1 y_1 + x_2 y_2 + \dots + x_n y_n, \quad (1.2.17)$$

而且向量 x 的坐标 x_1, x_2, \dots, x_n 还可通过内积表达出来,即有

$$x_i = (\varepsilon_i, x), \quad i = 1, 2, \dots, n. \quad (1.2.18)$$

事实上,设 $x = x_1 \varepsilon_1 + x_2 \varepsilon_2 + \dots + x_n \varepsilon_n$,以 $\varepsilon_i (i = 1, 2, \dots, n)$ 与等式两端作内积,便得 $x_i = (\varepsilon_i, x)$.

因此,在研究欧氏空间时,我们总采用其标准正交基.例如,在三维几何空间 \mathbb{R}^3 中,采用笛卡儿坐标系 i, j, k 作为标准正交基,此时任意两向量 x, y 的内积(数量积)表示为

$$(x, y) = x_1 y_1 + x_2 y_2 + x_3 y_3,$$

就是这个道理.

2. 酉空间介绍

欧氏空间是针对实线性空间而言的,即在实线性空间上定义内积运算便构成欧氏空间,这里将要介绍的酉空间,实际上就是一个特殊的复线性空间.酉空间的理论与欧氏空间的理论很相近,有一套平行的理论.本段只简单列举出复线性空间中的内积的定义与欧氏空间中内积定义的主要区别以及酉空间的主要结论,而不给出详细的说明.

如果 \mathbb{P} 为复数域,则前面欧氏空间给出的内积定义的非负性条件(4)有时不能成立.如在三维复向量空间 \mathbb{C}^3 中仍定义内积为 $(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^3 x_i y_i$,取 $\boldsymbol{x} = (3, 4, 5i)$,则 $(\boldsymbol{x}, \boldsymbol{x}) = 3^2 + 4^2 + (5i)^2 = 0$,这就是说,一个非零向量的长度等于0;又若取 $\boldsymbol{x} = (3, 4, 6i)$,则 $(\boldsymbol{x}, \boldsymbol{x}) < 0$.为了避免这一麻烦,对于 n 维复向量空间 \mathbb{C}^n 中的内积一开始就规定 $(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^n x_i \bar{y}_i$ (其中 \bar{y}_i 是 y_i 的共轭复数),则 $(\boldsymbol{x}, \boldsymbol{x}) = \sum_{i=1}^n x_i \bar{x}_i = \sum_{i=1}^n |x_i|^2$ 必然是非负的,而且除 $\boldsymbol{x} = \mathbf{0}$ 外, $(\boldsymbol{x}, \boldsymbol{x})$ 总大于零.不过这样一来,定义1.2.1中的对称性条件(1): $(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{y}, \boldsymbol{x})$,就不成立了,这时可用 $(\boldsymbol{x}, \boldsymbol{y}) = \overline{(\boldsymbol{y}, \boldsymbol{x})}$ 来代替条件(1).

这里 $\overline{(\cdot, \cdot)}$ 表示数 (\cdot, \cdot) 的共轭数,可见欧氏空间内积定义1.2.1中的条件(1)与(4)在复数域上不能并存.为此,我们选择修改(1),保留(4).

定义 1.2.7 设 V 是复数域 \mathbb{C} 上的线性空间,对于 V 中任意两个向量 $\boldsymbol{x}, \boldsymbol{y}$ (\boldsymbol{y} 可以等于 \boldsymbol{x}),如能给定某种规则,使 $\boldsymbol{x}, \boldsymbol{y}$ 对应着一个复数 $(\boldsymbol{x}, \boldsymbol{y})$,它能满足以下条件:

- (1) $(\boldsymbol{x}, \boldsymbol{y}) = \overline{(\boldsymbol{y}, \boldsymbol{x})}$;
- (2) $(\boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z}) = (\boldsymbol{x}, \boldsymbol{z}) + (\boldsymbol{y}, \boldsymbol{z}), \quad \forall \boldsymbol{z} \in V$;
- (3) $(k\boldsymbol{x}, \boldsymbol{y}) = k(\boldsymbol{x}, \boldsymbol{y}), \quad \forall k \in \mathbb{C}$;
- (4) $(\boldsymbol{x}, \boldsymbol{x}) \geq 0$, 当且仅当 $\boldsymbol{x} = \mathbf{0}$ 时 $(\boldsymbol{x}, \boldsymbol{x}) = 0$;

则称该复数 $(\boldsymbol{x}, \boldsymbol{y})$ 是向量 \boldsymbol{x} 与 \boldsymbol{y} 的内积.

如此定义了内积的复数域 \mathbb{C} 上的线性空间叫做酉空间(或 U 空间,或复内积空间).

例 1.2.8 在复 n 维向量空间 \mathbb{C}^n 中,对于任意两个向量

$$\boldsymbol{x} = (x_1, x_2, \dots, x_n), \quad \boldsymbol{y} = (y_1, y_2, \dots, y_n),$$

定义其内积为

$$(\boldsymbol{x}, \boldsymbol{y}) = x_1 \bar{y}_1 + x_2 \bar{y}_2 + \dots + x_n \bar{y}_n = \boldsymbol{x} \boldsymbol{y}^H, \quad (1.2.19)$$

其中 \boldsymbol{y}^H 表示 \boldsymbol{y} 的转置共轭向量,即 $\boldsymbol{y}^H = \bar{\boldsymbol{y}}^T$.

如果 $\boldsymbol{x}, \boldsymbol{y}$ 为列向量,则有

$$(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{y}^H \boldsymbol{x}.$$

显然,式(1.2.19)定义的内积满足定义1.2.7中的4个条件,故 \mathbb{C}^n 就是一个酉空间,仍以 \mathbb{C}^n 表示.

由式(1.2.19)立即可得

$$(\boldsymbol{x}, \boldsymbol{x}) = x_1 \bar{x}_1 + x_2 \bar{x}_2 + \dots + x_n \bar{x}_n = \sum_{i=1}^n |x_i|^2, \quad (1.2.20)$$

$|\boldsymbol{x}| = \sqrt{(\boldsymbol{x}, \boldsymbol{x})}$ 叫做向量 \boldsymbol{x} 的模或长度.

由于欧氏空间和酉空间都是在线性空间中引入了内积(虽然有稍许不同),故通常将这两个空间统称为内积空间.

欧氏空间与酉空间的差别不仅在于所用的数域不同,而且还在于内积定义中条件(1)不同,因而不涉及这两方面的空间性质便是相同的.下面列举出酉空间性质的主要结果.

$$(1) (\mathbf{x}, k\mathbf{y}) = \bar{k}(\mathbf{x}, \mathbf{y}).$$

$$(2) (\mathbf{x}, \mathbf{0}) = (\mathbf{0}, \mathbf{x}) = 0.$$

$$(3) \left(\sum_{i=1}^n \lambda_i \mathbf{x}_i, \sum_{j=1}^m \mu_j \mathbf{y}_j \right) = \sum_{i=1}^n \sum_{j=1}^m \lambda_i \bar{\mu}_j (\mathbf{x}_i, \mathbf{y}_j).$$

(4) 柯西-施瓦茨不等式仍成立,即对于任意两向量 \mathbf{x}, \mathbf{y} 有

$$(\mathbf{x}, \mathbf{y})(\mathbf{y}, \mathbf{x}) \leq (\mathbf{x}, \mathbf{x})(\mathbf{y}, \mathbf{y}).$$

上式当且仅当 \mathbf{x} 与 \mathbf{y} 线性相关时,等号成立.

应用这个不等式可以定义夹角的概念.

(5) 两个非零向量 \mathbf{x} 与 \mathbf{y} 的夹角 $\langle \mathbf{x}, \mathbf{y} \rangle$ 的余弦定义为

$$\cos^2 \langle \mathbf{x}, \mathbf{y} \rangle = \frac{(\mathbf{x}, \mathbf{y})(\mathbf{y}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})(\mathbf{y}, \mathbf{y})}.$$

当 $(\mathbf{x}, \mathbf{y}) = 0$ (或 $\langle \mathbf{x}, \mathbf{y} \rangle = \frac{\pi}{2}$) 时,称 \mathbf{x} 与 \mathbf{y} 正交.

在 n 维酉空间中,同样可以定义基和标准正交基的概念.关于标准正交基还有下列性质.

(6) 任意线性无关的向量组可以用施密特正交化方法正交化.

(7) 任一非零酉空间都存在正交基和标准正交基.

习 题 1 (3)

1. 设 V^n 为实数域上的一个 n 维向量空间, $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ 为其中一组基,两个向量

$$\mathbf{u} = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + \dots + x_n \mathbf{e}_n,$$

$$\mathbf{v} = y_1 \mathbf{e}_1 + y_2 \mathbf{e}_2 + \dots + y_n \mathbf{e}_n.$$

定义实函数 (\mathbf{u}, \mathbf{v}) 为

$$(\mathbf{u}, \mathbf{v}) = x_1 y_1 + 2x_2 y_2 + \dots + nx_n y_n.$$

证明: V^n 在此规定下是一个欧氏空间.

2. 验证 $(\mathbf{u}, \mathbf{v}) = x_1 y_1 - x_1 y_2 - x_2 y_1 + 3x_2 y_2$ (其中 $\mathbf{u} = (x_1, x_2), \mathbf{v} = (y_1, y_2)$) 是 \mathbb{R}^2 中的内积.

3. 在 \mathbb{R}^n 中, 设 $\boldsymbol{\alpha} = (\xi_1, \xi_2, \dots, \xi_n), \boldsymbol{\beta} = (\eta_1, \eta_2, \dots, \eta_n)$, 分别定义实数 $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ 如下:

$$(1) (\boldsymbol{\alpha}, \boldsymbol{\beta}) = \left(\sum_{i=1}^n \xi_i^2 \eta_i^2 \right)^{1/2};$$

$$(2) (\boldsymbol{\alpha}, \boldsymbol{\beta}) = \left(\sum_{i=1}^n \xi_i \right) \left(\sum_{j=1}^n \eta_j \right);$$

$$(3) (\boldsymbol{\alpha}, \boldsymbol{\beta}) = \left(\sum_{i=1}^n i \xi_i \eta_i \right);$$

判断它们是否为 \mathbb{R}^n 中向量 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 的内积.

4. 设 $\boldsymbol{\alpha} = (\xi_1, \xi_2), \boldsymbol{\beta} = (\eta_1, \eta_2)$ 是二维实线性空间 \mathbb{R}^2 的任意两个向量, 问 \mathbb{R}^2 对以下定义的实函数 $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ 是否构成欧氏空间?

$$(1) (\boldsymbol{\alpha}, \boldsymbol{\beta}) = \xi_1 \eta_1 + \xi_2 \eta_2 + 1;$$

$$(2) (\boldsymbol{\alpha}, \boldsymbol{\beta}) = \xi_1 \eta_1 - \xi_2 \eta_2;$$

$$(3) (\alpha, \beta) = 3\xi_1 \eta_1 + 5\xi_2 \eta_2;$$

$$(4) (\alpha, \beta) = \xi_1 \eta_1 + (\xi_1 - \xi_2)(\eta_1 - \eta_2).$$

5. 设 V 是多项式所组成的线性空间, 其内积由

$$(f, g) = \int_0^1 f(t)g(t)dt$$

给出. 设 $f(t) = t + 2, g(t) = t^2 - 2t - 3$, 求 (f, g) .

6. 设欧氏空间 $P[x]_2$ 中的内积定义为

$$(f, g) = \int_{-1}^1 f(x)g(x)dx,$$

(1) 求基 $1, x, x^2$ 的度量矩阵;

(2) 用坐标与度量矩阵乘积形式计算 $f(x) = 1 - x + x^2$ 与 $g(x) = 1 - 4x - 5x^2$ 的内积.

7. 设 $A = (a_{ij})$ 是一个 n 阶正定矩阵, 而 $\alpha = (x_1, x_2, \dots, x_n), \beta = (y_1, y_2, \dots, y_n)$, 在 \mathbb{R}^n 中定义实函数

$$(\alpha, \beta) = \alpha A \beta^T.$$

(1) 证明: 在这个意义下, \mathbb{R}^n 构成一欧氏空间;

(2) 求向量 $\varepsilon_1 = (1, 0, \dots, 0), \varepsilon_2 = (0, 1, \dots, 0), \dots, \varepsilon_n = (0, 0, \dots, 1)$ 的度量矩阵;

(3) 具体写出这个空间中的柯西-布涅柯夫斯基不等式.

8. 证明: 对任意实数 a_1, a_2, \dots, a_n , 下列不等式成立:

$$\sum_{i=1}^n |a_i| \leq \sqrt{n \sum_{i=1}^n a_i^2}.$$

9. 设 x_1, x_2, \dots, x_m 是欧氏空间 V^n 的一组向量, 而

$$\Delta = \begin{pmatrix} (x_1, x_1) & (x_1, x_2) & \cdots & (x_1, x_m) \\ (x_2, x_1) & (x_2, x_2) & \cdots & (x_2, x_m) \\ \vdots & \vdots & \ddots & \vdots \\ (x_m, x_1) & (x_m, x_2) & \cdots & (x_m, x_m) \end{pmatrix}.$$

试证明: $\det \Delta \neq 0$ 的充分必要条件是 x_1, x_2, \dots, x_m 线性无关.

10. 证明: 在欧氏空间 V^n 中, 采用坐标与度量矩阵乘法形式计算两个向量的内积时, 其值与选取的基无关.

11. (1) \mathbb{R}^n 中, $\alpha = (a_1, a_2, \dots, a_n)^T, \beta = (b_1, b_2, \dots, b_n)^T$, 判别实数 $(\alpha, \beta) = \sum_{i=1}^n |a_i b_i|$ 是否为内积?

(2) 验证: 若 $(\alpha, \beta)_1$ 与 $(\alpha, \beta)_2$ 是欧氏空间 V 的两个不同的内积, 则 $(\alpha, \beta) = (\alpha, \beta)_1 + (\alpha, \beta)_2$ 也是 V 的一个内积. 试创造一种新办法再构造 V 的一种内积.

12. 对 $x = (x_1, x_2)^T, y = (y_1, y_2)^T$, 规定

$$(x, y) = ax_1 y_1 + bx_1 y_2 + bx_2 y_1 + cx_2 y_2.$$

证明: (x, y) 是 \mathbb{R}^2 的内积 $\Leftrightarrow a > 0, ac > b^2$.

13. 设 $V = \{a \cos t + b \sin t, \text{其中 } a, b \text{ 为任意实数}\}$ 是实二维线性空间. 对任意 $f, g \in V$, 定义

$$(f, g) = f(0)g(0) + f\left(\frac{\pi}{2}\right) \cdot g\left(\frac{\pi}{2}\right).$$

证明: (f, g) 是 V 上的内积, 并求 $h(t) = 3\cos(t+7) + 4\sin(t+9)$ 的长度.

14. 设 $a_i (1 \leq i \leq n)$ 是正实数, x_i, y_i 是任意实数, 证明或否定不等式

$$\left(\sum_{i=1}^n a_i x_i y_i\right)^2 \leq \left(\sum_{i=1}^n a_i x_i^2\right) \left(\sum_{i=1}^n a_i y_i^2\right).$$

15. (1) $\forall x, y \in \mathbb{C}^n$, 定义内积 $(x, y) = x^H y$, 证明:

$$y^H x + x^H y = \frac{1}{2} (|x+y|^2 - |x-y|^2);$$

(2) 已知 $A \in \mathbb{R}^{n \times n}, A$ 为正定矩阵, $\forall x, y \in \mathbb{R}^n$, 证明: $|(Ax, y)| \leq (Ax, x)^{\frac{1}{2}} (Ay, y)^{\frac{1}{2}}$.

16. 设 u, v 为欧氏空间中两个非零向量, 证明:

(1) $u = av$ ($a > 0$) 必要且只要 u, v 的夹角为 0;

(2) $u = av$ ($a < 0$) 必要且只要 u, v 的夹角为 π .

17. 在 \mathbb{R}^n 中, 求 α, β 之间的夹角 $\langle \alpha, \beta \rangle$ (内积按通常意义). 设

(1) $\alpha = (2, 1, 3, 2), \quad \beta = (1, 2, -2, 1);$

(2) $\alpha = (1, 2, 2, 3), \quad \beta = (3, 1, 5, 1);$

(3) $\alpha = (1, 1, 1, 2), \quad \beta = (3, 1, -1, 0).$

18. 在 \mathbb{R}^4 中, 求一单位向量与 $(1, 1, -1, 1), (1, -1, -1, 1)$ 及 $(2, 1, 1, 3)$ 均正交.

19. 把向量

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

正交化、单位化.

20. 在 $P[x]_3$ 中定义内积

$$(f(x), g(x)) = \int_{-1}^1 f(x)g(x)dx,$$

求 $P[x]_3$ 的一个标准正交基(由基 $1, x, x^2, x^3$ 出发作正交单位化).

21. 求齐次线性方程组

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

的解空间(作为 \mathbb{R}^5 的子空间)的一组标准正交基.

22. 在 \mathbb{R}^2 中按照某种内积方式(不一定是通常的)构成欧氏空间, 记作 V^2 . 已知 V^2 的两组基为

(I) $e_1 = (1, 1), \quad e_2 = (1, -1);$

(II) $e'_1 = (0, 2), \quad e'_2 = (6, 12).$

e_i 与 e'_j 的内积为

$$(e_1, e'_1) = 1, \quad (e_1, e'_2) = 15, \quad (e_2, e'_1) = -1, \quad (e_2, e'_2) = 3.$$

(1) 求基(I)的度量矩阵;

(2) 求基(II)的度量矩阵;

(3) 求出欧氏空间 V^2 的一组标准正交基.

23. 设 $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5$ 是 \mathbb{R}^5 的一个标准正交基, 子空间 $W = L[\alpha_1, \alpha_2, \alpha_3]$, 其中 $\alpha_1 = \varepsilon_1 + \varepsilon_5, \alpha_2 = \varepsilon_1 - \varepsilon_3 + \varepsilon_4, \alpha_3 = 2\varepsilon_1 + \varepsilon_2 + \varepsilon_3$, 求 W 的标准正交基.

24. 设 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 为欧氏空间 V^n 的标准正交基, $\forall \alpha \in V^n$, 证明:

$$|\alpha|^2 = \sum_{i=1}^n (\alpha, \varepsilon_i)^2.$$

25. 试把 $\alpha_1 = (1, 0, 1, 0)^T, \alpha_2 = (0, 1, 0, 2)^T$ 扩充为 \mathbb{R}^4 的一个正交基.

26. 在 $\mathbb{R}^{2 \times 2}$ 中, 定义 $(A, B) = \sum_{i=1}^2 \sum_{j=1}^2 a_{ij}b_{ij}$, 其中 $A = (a_{ij})_{2 \times 2}, B = (b_{ij})_{2 \times 2}$, 证明: $E_{11}, E_{12}, E_{21}, E_{22}$ 是

$\mathbb{R}^{2 \times 2}$ 的一个标准正交基.

把 $\alpha_1 = 1, \alpha_2 = x, \alpha_3 = x^2$ 化为标准正交基.

27. 设 $A = (a_{ij}), B = (b_{ij}) \in \mathbb{C}^{n \times n}$.

(1) 证明 $(A, B) = \text{tr}(AB^H)$ 是 $\mathbb{C}^{n \times n}$ 的一个内积;

(2) 按(1)的内积, 矩阵 A 的长度是多少? 哪些是单位向量?

(3) 证明或否定: 基本矩阵 $E_{ij}, (1 \leq i, j \leq n)$ 是 $\mathbb{C}^{n \times n}$ 的一组标准正交基;

(4) 求 $\mathbb{R}^{2 \times 2}$ 的一组由可逆矩阵构成的标准正交基.

28. 在多项式空间 $P[x]_3$ 中, 定义内积为

$$\forall f(x), g(x) \in P[x]_3, (f(x), g(x)) = \int_0^{+\infty} f(x)g(x)e^{-x} dx,$$

把 $\alpha_1 = 1, \alpha_2 = x, \alpha_3 = x^2$ 化为标准正交基.

29. 设 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 是欧氏空间 V^n 的标准正交基, 则 $\beta_1, \beta_2, \dots, \beta_n$ 两两正交的充要条件是

$$\sum_{k=1}^n (\beta_i, \varepsilon_k) (\beta_j, \varepsilon_k) = 0, \quad i \neq j.$$

30. 已知欧氏空间 V^2 的基 e_1, e_2 的度量矩阵为 $A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$, 利用合同变换方法求 V^2 的一组标准正交

基(用 e_1, e_2 表示).

31. 设 e_1, e_2, \dots, e_n 为 n 维欧氏空间 V^n 的一组基, 证明任意两个向量

$$u = (e_1, e_2, \dots, e_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad v = (e_1, e_2, \dots, e_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

的内积可由等式

$$(u, v) = x_1 y_1 + \dots + x_n y_n$$

表达的充要条件是: e_1, e_2, \dots, e_n 为标准正交基.

32. 设 $\varepsilon_1, \varepsilon_2, \varepsilon_3$ 是三维欧氏空间中一组标准正交基, 证明:

$$\alpha_1 = \frac{1}{3}(2\varepsilon_1 + 2\varepsilon_2 - \varepsilon_3), \quad \alpha_2 = \frac{1}{3}(2\varepsilon_1 - \varepsilon_2 + 2\varepsilon_3), \quad \alpha_3 = \frac{1}{3}(\varepsilon_1 - 2\varepsilon_2 - 2\varepsilon_3)$$

也是一组标准正交基.

33. 设欧氏空间 V^n 的一组基为 e_1, e_2, \dots, e_n , 它的度量矩阵为 A . 证明: 存在实对称正定矩阵 C , 使得由

$$(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) = (e_1, e_2, \dots, e_n)C$$

确定的基 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 为 V^n 的标准正交基.

34. 已知欧氏空间 V^3 的一组基 e_1, e_2, e_3 的度量矩阵为 $A = \begin{pmatrix} 2 & 1 & 2 \\ 1 & 1 & 1 \\ 2 & 1 & 5 \end{pmatrix}$, 试利用施密特正交化方法求

V^3 的一组标准正交基(用 e_1, e_2, e_3 表示).

35. 设 e_1, e_2, e_3, e_4, e_5 是欧氏空间 V^5 的一组标准正交基, $V_1 = \text{Span}(\alpha_1, \alpha_2, \alpha_3)$, 其中 $\alpha_1 = e_1 + e_3, \alpha_2 = e_1 - e_2 + e_4, \alpha_3 = 2e_1 + e_2 + e_3$, 求 V_1 的一组标准正交基.

36. 取酉空间 U 的一组基 e_1, e_2, \dots, e_n , 如果 U 的任意两个向量

$$\alpha = \sum_{i=1}^n x_i e_i, \quad \beta = \sum_{i=1}^n y_i e_i$$

的内积为

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

试证: e_1, e_2, \dots, e_n 是 U 的标准正交基.

37. 在欧氏空间 \mathbb{R}^4 中, 求三个向量 $\alpha_1 = (1, 0, 1, 1)^T, \alpha_2 = (2, 1, 0, -3)^T$ 和 $\alpha_3 = (1, -1, 1, -1)^T$ 所生成的子空间的一个标准正交基.

38. (1) 复数域 \mathbb{C} 是实数域 \mathbb{R} 上的二维线性空间, 是否存在 \mathbb{C} 上的一个内积, 使得 i 与 $1+i$ 成为 \mathbb{C} 的一组标准正交基, 为什么?

(2) 试构造实线性空间 \mathbb{R}^3 上的一个内积, 使得向量组

$$e_1, e_1 + e_2, e_1 + e_2 + e_3$$

是一组标准正交基, 问 e_2 与 e_3 的长度各是多少? 它们的夹角又是多少?

39. 试尽可能一般性地讨论习题 38 中的问题.

40. 设二维欧氏空间的一组基为 α_1, α_2 , 其度量矩阵为

$$A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix},$$

试求 V 的一组标准正交基以及它到 α_1, α_2 的过渡矩阵.

41. 设 A 是反对称实矩阵(即 $A^T = -A$), 证明:

(1) A 的特征值为 0 或纯虚数;

(2) 设 $\alpha + \beta i$ 是 A 的属于一个非零特征值的特征向量, 其中 α, β 均为实向量, 则 α 与 β 正交.

42. 设 $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$, 定义 \mathbb{R}^2 上的二元(向量)函数 $\langle x, y \rangle$ 如下:

$$\langle x, y \rangle = x^T A y,$$

此二元函数与普通内积的差别是什么? 以此二元函数为基础, 建立相应的长度、角度等概念, 研究其中的正交与平行的定理.

1.2.2 等积变换及其矩阵

由 V 到自身的一切线性变换中, 有一类具有特殊性质的线性变换. 例如, 解析几何中, 空间 \mathbb{R}^2 的所有向量均绕原点依逆(或顺)时针旋转 θ 角的变换 \mathcal{A} , 就是一个特殊的线性变换, 它能使向量的长度与夹角保持不变, 即对任何 $x, y \in \mathbb{R}^2$, 有

$$\begin{cases} |\mathcal{A}(x)| = |x|, & |\mathcal{A}(y)| = |y|, \\ \arccos \frac{(\mathcal{A}(x), \mathcal{A}(y))}{|\mathcal{A}(x)| |\mathcal{A}(y)|} = \arccos \frac{(x, y)}{|x| |y|}. \end{cases}$$

这里 x, y 为非零向量. 从上两式不难看出有 $(\mathcal{A}(x), \mathcal{A}(y)) = (x, y)$, 亦即旋转变换 \mathcal{A} 保持向量 x 与 y 的内积不变. 把这种保持向量内积不变的线性变换推广到一般的内积空间, 就是本节将要讨论的等积变换.

1. 正交变换与正交矩阵

1) 正交变换的定义及性质

定义 1.2.8 设 V 是一个欧氏空间, \mathcal{A} 是 V 上的线性变换, 如果对于任何向量 $x, y \in V$, 变换 \mathcal{A} 恒能使下式成立:

$$(\mathcal{A}(x), \mathcal{A}(y)) = (x, y), \quad (1.2.21)$$

则说 \mathcal{A} 是 V 上的正交变换.

例如, 恒等变换即是一个正交变换; 坐标平面上的旋转变换也是一个正交变换. 下面讨论正交变换的一些性质.

定理 1.2.5 设 \mathcal{A} 是欧氏空间 V 上的线性变换, 下面写出的任一条件都是使 \mathcal{A} 成为正交变换的充要条件:

(1) \mathcal{A} 使向量长度保持不变, 即: 对任何 $x \in V$, 有

$$(\mathcal{A}(x), \mathcal{A}(x)) = (x, x); \quad (1.2.22)$$

(2) 任一组标准正交基经 \mathcal{A} 变换后的像仍是一组标准正交基;

(3) \mathcal{A} 在任一组标准正交基下的矩阵 A 满足

$$A^T A = A A^T = I \quad \text{或} \quad A^{-1} = A^T. \quad (1.2.23)$$

证明 (1) 只要让 $y = x$ 便知必要性成立. 为证明充分性, 分别对向量 $x, y, x + y$ 使用式(1.2.22), 即有

$$\begin{aligned}(\mathcal{A}(\mathbf{x}), \mathcal{A}(\mathbf{x})) &= (\mathbf{x}, \mathbf{x}), \\(\mathcal{A}(\mathbf{y}), \mathcal{A}(\mathbf{y})) &= (\mathbf{y}, \mathbf{y}), \\(\mathcal{A}(\mathbf{x} + \mathbf{y}), \mathcal{A}(\mathbf{x} + \mathbf{y})) &= (\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}),\end{aligned}$$

而

$$\begin{aligned}(\mathcal{A}(\mathbf{x} + \mathbf{y}), \mathcal{A}(\mathbf{x} + \mathbf{y})) &= (\mathcal{A}(\mathbf{x}) + \mathcal{A}(\mathbf{y}), \mathcal{A}(\mathbf{x}) + \mathcal{A}(\mathbf{y})) \\&= (\mathcal{A}(\mathbf{x}), \mathcal{A}(\mathbf{x})) + 2(\mathcal{A}(\mathbf{x}), \mathcal{A}(\mathbf{y})) + (\mathcal{A}(\mathbf{y}), \mathcal{A}(\mathbf{y})) \\&= (\mathbf{x}, \mathbf{x}) + 2(\mathcal{A}(\mathbf{x}), \mathcal{A}(\mathbf{y})) + (\mathbf{y}, \mathbf{y}), \\(\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) &= (\mathbf{x}, \mathbf{x}) + 2(\mathbf{x}, \mathbf{y}) + (\mathbf{y}, \mathbf{y}),\end{aligned}$$

故有

$$(\mathcal{A}(\mathbf{x}), \mathcal{A}(\mathbf{y})) = (\mathbf{x}, \mathbf{y}).$$

(2) 由正交变换的定义可知, 若 $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_n$ 为 V 的标准正交基, 则有 $(\mathcal{A}(\boldsymbol{\varepsilon}_i), \mathcal{A}(\boldsymbol{\varepsilon}_j)) = (\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j) = \delta_{ij}$, 即基像仍是一组标准正交基. 反之, 若 $\mathbf{x} = \sum_{i=1}^n x_i \boldsymbol{\varepsilon}_i, \mathbf{y} = \sum_{j=1}^n y_j \boldsymbol{\varepsilon}_j$, 则可推出

$$\begin{aligned}(\mathcal{A}(\mathbf{x}), \mathcal{A}(\mathbf{y})) &= (\mathcal{A}(\sum_{i=1}^n x_i \boldsymbol{\varepsilon}_i), \mathcal{A}(\sum_{j=1}^n y_j \boldsymbol{\varepsilon}_j)) \\&= \sum_{i,j=1}^n x_i y_j \delta_{ij} = \sum_{i=1}^n x_i y_i = (\sum_{i=1}^n x_i \boldsymbol{\varepsilon}_i, \sum_{j=1}^n y_j \boldsymbol{\varepsilon}_j) \\&= (\mathbf{x}, \mathbf{y}),\end{aligned}$$

故 \mathcal{A} 是正交变换.

(3) 假设 \mathcal{A} 在标准正交基 $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_n$ 下的矩阵为 \mathbf{A} , 则有

$$\mathcal{A}(\boldsymbol{\varepsilon}_i) = a_{1i} \boldsymbol{\varepsilon}_1 + a_{2i} \boldsymbol{\varepsilon}_2 + \dots + a_{ni} \boldsymbol{\varepsilon}_n, \quad i = 1, 2, \dots, n,$$

即

$$\mathcal{A}(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_n) = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_n) \mathbf{A},$$

其中

$$\mathbf{A} = (a_{ij})_{n \times n} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

另一方面, 有

$$\begin{aligned}(\mathcal{A}(\boldsymbol{\varepsilon}_i), \mathcal{A}(\boldsymbol{\varepsilon}_j)) &= (\sum_{k=1}^n a_{ki} \boldsymbol{\varepsilon}_k, \sum_{l=1}^n a_{lj} \boldsymbol{\varepsilon}_l) \\&= \sum_{k,l=1}^n a_{ki} a_{lj} (\boldsymbol{\varepsilon}_k, \boldsymbol{\varepsilon}_l) = \sum_{k,l=1}^n a_{ki} a_{lj} \delta_{kl} \\&= \sum_{k=1}^n a_{ki} a_{kj}, \quad i, j = 1, 2, \dots, n.\end{aligned}$$

由上面(2)知,

$$\begin{aligned}\mathcal{A} \text{ 为正交变换} &\Leftrightarrow (\mathcal{A}(\boldsymbol{\varepsilon}_i), \mathcal{A}(\boldsymbol{\varepsilon}_j)) = \delta_{ij} \\&\Leftrightarrow \sum_{k=1}^n a_{ki} a_{kj} = \delta_{ij}, \quad i, j = 1, 2, \dots, n \\&\Leftrightarrow \mathbf{A}^T \mathbf{A} = \mathbf{I}.\end{aligned}$$

不妨在 $\mathbf{A}^T \mathbf{A} = \mathbf{I}$ 两端同时取行列式, 有 $|\mathbf{A}^T| |\mathbf{A}| = |\mathbf{A}|^2 = 1$, 从而得 $|\mathbf{A}| = \pm 1 \neq 0$, 故知矩阵 \mathbf{A} 必可逆. 若在 $\mathbf{A}^T \mathbf{A} = \mathbf{I}$ 两端右乘 \mathbf{A}^{-1} 得

$$\mathbf{A}^{-1} = \mathbf{A}^T,$$

将它两端左乘 \mathbf{A} , 又有

$$\mathbf{A} \mathbf{A}^T = \mathbf{I}.$$

这样便证得 \mathcal{A} 为正交变换的充分必要条件是满足

$$\mathbf{A}^T \mathbf{A} = \mathbf{A} \mathbf{A}^T = \mathbf{I} \quad \text{或} \quad \mathbf{A}^{-1} = \mathbf{A}^T. \quad \text{证毕}$$

2) 正交矩阵

定义 1.2.9 满足式(1.2.23)的任何实方阵 \mathbf{A} 叫做正交矩阵.

由于正交矩阵有 $\mathbf{A}^{-1} = \mathbf{A}^T$, 这就给求逆带来了方便. 我们常用 $\mathbf{A} \mathbf{A}^T = \mathbf{I}$ 或 $\mathbf{A}^T \mathbf{A} = \mathbf{I}$ 是否成立来判断一个矩阵是否为正交矩阵.

例 1.2.9 判断下列矩阵是否为正交矩阵:

(1) $\mathbf{A} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$, 其中 θ 是实数;

(2) $\mathbf{B} = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{6} & \frac{2}{3} \\ 0 & -\frac{2\sqrt{2}}{3} & \frac{1}{3} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{6} & \frac{2}{3} \end{pmatrix}$.

解 (1)

$$\mathbf{A} \mathbf{A}^T = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

所以 \mathbf{A} 是正交矩阵.

(2) $\mathbf{B} \mathbf{B}^T = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{6} & \frac{2}{3} \\ 0 & -\frac{2\sqrt{2}}{3} & \frac{1}{3} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{6} & \frac{2}{3} \end{pmatrix},$

$$\begin{pmatrix} \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{6} & -\frac{2\sqrt{2}}{3} & \frac{\sqrt{2}}{6} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

所以 \mathbf{B} 也是正交矩阵.

由定理 1.2.5 的条件(3), 我们可以建立正交变换与正交矩阵之间的一一对应关系.

定理 1.2.6 在欧氏空间中, 正交变换在标准正交基下的矩阵是正交矩阵; 反过来, 如果线性变换 \mathcal{A} 在标准正交基下的矩阵是正交矩阵, 则 \mathcal{A} 是正交变换.

例 1.2.10 试写出 \mathbb{R}^2 中向量绕原点按顺时针旋转 θ 角的旋转变换 \mathcal{A} 在基 i, j 下的矩阵(见图 1.3).

解 根据旋转变换保持向量长度不变的性质, 不难写出基像的表达式

$$\begin{cases} \mathcal{A}(i) = i' = \cos\theta i - \sin\theta j, \\ \mathcal{A}(j) = j' = \sin\theta i + \cos\theta j. \end{cases}$$

于是得到 \mathcal{A} 在基 i, j 下的矩阵为

$$A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}. \quad (1.2.24)$$

上题已验证它为正交矩阵.

必须注意, 正交变换与正交矩阵的对应关系是建立在标准正交基上的. 如果取的基不是标准正交基, 则正交变换所对应的矩阵就不一定是正交矩阵了, 这由下面的两个例子可看出来.

例 1.2.11 在 \mathbb{R}^2 中, 设 \mathcal{A} 是向量关于 x 轴的反射变换, 即把 $\alpha = xi + yj$ 反射成 $\alpha' = xi - yj$ (见图 1.4).

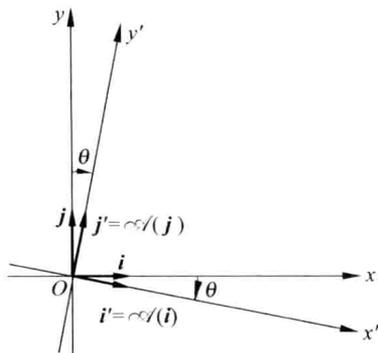


图 1.3

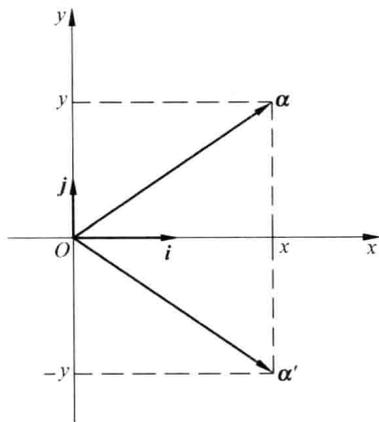


图 1.4

显然, 基像可表示为

$$\begin{cases} \mathcal{A}(i) = 1i + 0j, \\ \mathcal{A}(j) = 0i - 1j. \end{cases}$$

所以 \mathcal{A} 在 i, j 下的矩阵为

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

容易验证 $HH^T = I$, 故 H 也是正交矩阵.

但是, 如果采用 α, j 为基 (α 是以 i, j 为边的正方形的对角线向量, 即 $\alpha = i + j$), 此时, α, j 不再是标准正交基, 基像可表示为

$$\begin{cases} \mathcal{A}(\alpha) = i - j = 1\alpha - 2j, \\ \mathcal{A}(j) = -j = 0\alpha - 1j, \end{cases}$$

则反射变换 \mathcal{A} 在基 α, j 下的矩阵 $H = \begin{bmatrix} 1 & 0 \\ -2 & -1 \end{bmatrix}$ 不再是正交矩阵.

例 1.2.12 在 \mathbb{R}^2 中, 令向量 α 的意义同例 1.2.11 中的一样, 重新定义线性变换 \mathcal{A} : $\mathcal{A}(\alpha) = -\alpha$, $\mathcal{A}(j) = j$ (见图 1.5).

由于

$$\begin{cases} \mathcal{A}(\alpha) = -\alpha = -1\alpha + 0j, \\ \mathcal{A}(j) = j = 0\alpha + 1j, \end{cases}$$

易知 \mathcal{A} 关于基 α, j 的矩阵 $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ 是正交矩阵, 但 \mathcal{A} 却是非正交变换. 这是因为从初等几何学知道

$$\mathcal{A}(j + \alpha) = \mathcal{A}(j) + \mathcal{A}(\alpha) = j - \alpha = -i,$$

$$j + \alpha = f,$$

$$|j + \alpha|^2 = |f|^2 = |\alpha|^2 + |j|^2 - 2|\alpha||j|\cos\frac{3\pi}{4}$$

$$= 2 + 1 - 2\sqrt{2}\left(-\frac{1}{\sqrt{2}}\right) = 5,$$

$$|\mathcal{A}(j + \alpha)| = |-i| = 1,$$

所以 $|\mathcal{A}(f)| \neq |f|$, 因此 \mathcal{A} 不是正交变换.

例 1.2.11 与例 1.2.12 表明, 当基不是标准正交基时, 正交变换对应的矩阵可以不是正交矩阵, 而正交矩阵对应的线性变换也可以不是正交变换. 但是, 虽然基不是标准正交的, 而正交变换所对应的矩阵也有可能是正交的, 下面的例 1.2.13 就说明这个问题.

例 1.2.13 在 \mathbb{R}^2 中, 仍考察以 x 轴为对称轴的反射变换 \mathcal{A} .

令 $e_1 = \frac{1}{2}i + j, e_2 = -\frac{1}{2}i + j$, 则 e_1 与 e_2 关于 y 轴是对称的 (见图 1.6). 因 e_1, e_2 不正交, 故 e_1, e_2 不是 \mathbb{R}^2 的标准正交基, 但有

$$\begin{cases} \mathcal{A}(e_1) = \frac{1}{2}i - j = -e_2 = 0e_1 - 1e_2, \\ \mathcal{A}(e_2) = -\frac{1}{2}i - j = -e_1 = -1e_1 + 0e_2. \end{cases}$$

可见虽然 e_1, e_2 不是标准正交基, 但 \mathcal{A} 所对应的矩阵

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

却是正交矩阵.

关于正交矩阵还有如下的一些常用性质.

(1) 正交矩阵是非奇异的, 且 $\det A = 1$ 或 -1 (行列式等于 1 的正交矩阵叫正常的, 等于 -1 的叫非正常的).

事实上, 此性质的正确性由定理 1.2.5(3) 的证明过程可知.

(2) 正交矩阵的逆矩阵仍是正交矩阵.

事实上, 因为

$$(A^{-1})^T A^{-1} = (A^T)^{-1} A^{-1} = (AA^T)^{-1} = I,$$

所以 A^{-1} 是正交矩阵.

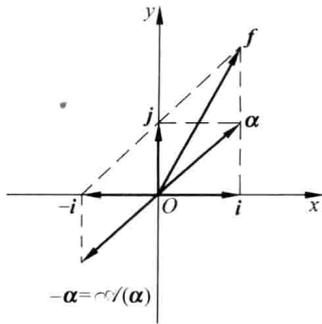


图 1.5

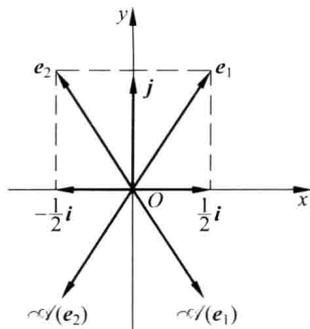


图 1.6

(3) 两个正交矩阵的乘积仍为正交矩阵.

事实上, 设 $\mathbf{A}_1, \mathbf{A}_2$ 都是正交矩阵, 则

$$(\mathbf{A}_1 \mathbf{A}_2)^T \mathbf{A}_1 \mathbf{A}_2 = \mathbf{A}_2^T \mathbf{A}_1^T \mathbf{A}_1 \mathbf{A}_2 = \mathbf{A}_2^T \mathbf{A}_2 = \mathbf{I}.$$

故结论成立.

(4) 实数域上方阵 \mathbf{A} 是正交矩阵的充分必要条件是: \mathbf{A} 的行(或列)向量组为标准正交向量组.

事实上, 不妨先对行进行证明. 设

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_n \end{pmatrix},$$

\mathbf{A} 的行向量组是 $\boldsymbol{\alpha}_1 = (a_{11}, \cdots, a_{1n}), \cdots, \boldsymbol{\alpha}_n = (a_{n1}, \cdots, a_{nn})$. 因为 \mathbf{A}^T 的 (i, j) 元^① = \mathbf{A} 的 (j, i) 元, 所以 $\mathbf{A}\mathbf{A}^T$ 的 (i, j) 元 = $\sum_{k=1}^n (\mathbf{A}$ 的 (i, k) 元 $\times \mathbf{A}^T$ 的 (k, j) 元) = $\sum_{k=1}^n (\mathbf{A}$ 的 (i, k) 元 $\times \mathbf{A}$ 的 (j, k) 元) = $\sum_{k=1}^n a_{ik} a_{jk} = (\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j)$, 其中 $i, j = 1, \cdots, n$. 即对于任意实方阵 \mathbf{A} 都有

$$\mathbf{A}\mathbf{A}^T \text{ 的 } (i, j) \text{ 元} = (\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j), \quad i, j = 1, \cdots, n. \quad (1.2.25)$$

如果 \mathbf{A} 是正交矩阵, 即有 $\mathbf{A}\mathbf{A}^T = \mathbf{I}$, 则

$$\mathbf{A}\mathbf{A}^T \text{ 的 } (i, j) \text{ 元} = \begin{cases} 1, & \text{当 } i = j \text{ 时,} \\ 0, & \text{当 } i \neq j \text{ 时.} \end{cases}$$

利用式(1.2.25)即得

$$(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j) = \begin{cases} 1, & \text{当 } i = j \text{ 时,} \\ 0, & \text{当 } i \neq j \text{ 时.} \end{cases}$$

这说明 \mathbf{A} 的行向量组 $\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_n$ 是标准正交向量组.

若 $\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_n$ 是标准正交向量组, 则

$$(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j) = \begin{cases} 1, & \text{当 } i = j \text{ 时,} \\ 0, & \text{当 } i \neq j \text{ 时.} \end{cases}$$

利用式(1.2.25)即得

$$\mathbf{A}\mathbf{A}^T \text{ 的 } (i, j) \text{ 元} = \begin{cases} 1, & \text{当 } i = j \text{ 时,} \\ 0, & \text{当 } i \neq j \text{ 时,} \end{cases}$$

所以

$$\mathbf{A}\mathbf{A}^T = \mathbf{I},$$

故 \mathbf{A} 为正交矩阵.

对列可类似证明.

这条性质告诉我们, 正交矩阵的任一行(或列)向量都是单位向量, 且任何两个不同的行(列)向量都是正交的, 即内积是零. 因此, 利用这一性质来构造一个正交矩阵是很方便的.

例 1.2.14 求正交矩阵 \mathbf{A} , 以 $\boldsymbol{\alpha}_1 = \left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right)$ 为其首行.

^① 矩阵的第 i 行、第 j 列元素, 简记为“ (i, j) 元”.

解 先求正交于 $\alpha_1 = \left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right)$ 的非零向量 $\alpha'_2 = (x, y, z)$, 则有

$$(\alpha_1, \alpha'_2) = \frac{x}{3} + \frac{2}{3}y + \frac{2}{3}z = 0 \quad \text{或} \quad x + 2y + 2z = 0,$$

显然 $\alpha'_2 = (0, 1, -1)$ 是其中一个解, 把 α'_2 单位化, 即得 A 的第二个行向量

$$\alpha_2 = \left(0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right).$$

再求同时正交于 α_1 与 α_2 的非零向量 $\alpha'_3 = (x, y, z)$, 则有

$$\begin{cases} (\alpha_1, \alpha'_3) = \frac{x}{3} + \frac{2}{3}y + \frac{2}{3}z = 0, \\ (\alpha_2, \alpha'_3) = \frac{y}{\sqrt{2}} - \frac{z}{\sqrt{2}} = 0, \end{cases}$$

即

$$\begin{cases} x + 2y + 2z = 0, \\ y - z = 0. \end{cases}$$

令 $z = -1$ 求得解 $\alpha'_3 = (4, -1, -1)$, 再单位化得 A 的第三个行向量

$$\alpha_3 = \left(\frac{4}{\sqrt{18}}, -\frac{1}{\sqrt{18}}, -\frac{1}{\sqrt{18}}\right),$$

从而得

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

注意, 从求解过程知道, 第一行是 $\left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right)$ 的正交矩阵不是惟一的. 另外, 如果只考虑行向量正交而不单位化, 则得到的矩阵 A 就不一定是正交矩阵.

由于标准正交基在欧氏空间中占有重要的地位, 因此, 我们不妨利用正交矩阵的性质(4)来讨论从一标准正交基改变到另一标准正交基的过渡矩阵的情况.

例 1.2.15 设 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 及 $\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_n$ 是欧氏空间 V^n 的两组标准正交基, 它们之间的过渡矩阵为 $A = (a_{ij})_{n \times n}$. 试证 A 为正交矩阵.

证明 由过渡矩阵的定义知

$$(\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_n) = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix},$$

即矩阵 A 的各列, 就是 $\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_n$ 在标准正交基 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 下的坐标, 即有

$$\varepsilon'_i = a_{1i}\varepsilon_1 + a_{2i}\varepsilon_2 + \cdots + a_{ni}\varepsilon_n, \quad i = 1, 2, \dots, n.$$

又因 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 与 $\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_n$ 都是标准正交基, 所以

$$(\varepsilon'_i, \varepsilon'_j) = (\varepsilon_i, \varepsilon_j) = \delta_{ij},$$

定义 1.2.10 R_{ij} 叫做初等旋转矩阵; 它所确定的线性变换叫做初等旋转变换(吉文斯(Givens)变换).

定理 1.2.7 设 R_{ij} 是初等旋转矩阵, 则有

(1) $\det R_{ij} = 1$;

(2) R_{ij} 对应的初等旋转变换是正交变换, R_{ij} 是正交矩阵.

证 (1) 由行列式展开定理知, $\det R_{ij} = c^2 + s^2 = 1$;

(2) 设 $\mathbf{x}, \mathbf{y} \in V, \mathbf{X} = (x_1, x_2, \dots, x_n)^T, \mathbf{Y} = (y_1, y_2, \dots, y_n)^T$ 分别是 \mathbf{x}, \mathbf{y} 在标准正交基下坐标列向量. 如果 $\mathbf{Y} = R_{ij}\mathbf{X}$, 由矩阵与列向量的乘法知, \mathbf{X} 的第 i 、第 j 个分量有了变化, 而其他分量都不变, 即有

$$\begin{cases} y_i = cx_i + sx_j \\ y_j = -sx_i + cx_j \\ y_k = x_k, \quad k \neq i, j. \end{cases} \quad (1.2.28)$$

不妨设 $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_n$ 是 V 的一组标准正交基, 于是有 $\mathbf{x} = x_1\boldsymbol{\varepsilon}_1 + x_2\boldsymbol{\varepsilon}_2 + \dots + x_n\boldsymbol{\varepsilon}_n, \mathbf{y} = y_1\boldsymbol{\varepsilon}_1 + y_2\boldsymbol{\varepsilon}_2 + \dots + y_n\boldsymbol{\varepsilon}_n$, 因此

$$(\mathbf{y}, \mathbf{y}) = \mathbf{Y}^T \mathbf{Y} = \sum_{m=1}^n y_m^2 = \sum_{m=1(m \neq i, j)}^n x_m^2 + y_i^2 + y_j^2 = \sum_{m=1}^n x_m^2 = (\mathbf{x}, \mathbf{x}),$$

从而有

$$|\mathbf{y}| = |\mathbf{x}|.$$

这说明向量 \mathbf{x} 的长度不变, 故 R_{ij} 对应的初等旋转变换是正交变换, R_{ij} 为正交矩阵. 证毕

初等旋转矩阵在简化矩阵方面有重要的应用. 例如, 对于非零 n 维列向量 ($n \times 1$ 矩阵) \mathbf{x} , 若以 R_{ij} 左乘, 即令 $\mathbf{y} = R_{ij}\mathbf{x}$, 根据式 (1.2.28) 易知, 只要按如下公式选取 R_{ij} 中的 c 和 s :

$$s = \frac{x_j}{\sqrt{x_i^2 + x_j^2}}, \quad c = \frac{x_i}{\sqrt{x_i^2 + x_j^2}}, \quad (1.2.29)$$

其中 x_i 和 x_j 不同时为零, 就可使 $y_i = \sqrt{x_i^2 + x_j^2} > 0, y_j = 0$. 换言之, 以 R_{ij} 左乘列向量, 可以使列向量得到简化, 即把第 j 个分量化为 0.

同理, 对于 n 维行向量 ($1 \times n$ 矩阵) \mathbf{x} , 则注意以 R_{ij} 右乘, 即令 $\mathbf{y} = \mathbf{x}R_{ij}$, 此时有

$$\begin{cases} y_i = cx_i - sx_j \\ y_j = sx_i + cx_j \\ y_k = x_k, \quad k \neq i, j, \end{cases} \quad (1.2.30)$$

我们可选取

$$s = \frac{-x_j}{\sqrt{x_i^2 + x_j^2}}, \quad c = \frac{x_i}{\sqrt{x_i^2 + x_j^2}}, \quad (1.2.31)$$

其中 x_i 与 x_j 不能同时为零, 就可使 $y_i = \sqrt{x_i^2 + x_j^2} > 0, y_j = 0$.

根据以上的这一性质, 进而可以推出:

对非零 n 维列(行)向量 \mathbf{x} 连续左(右)乘以 $R_{r,r+1}, R_{r,r+2}, \dots, R_{r,n}$, 可使 \mathbf{x} 的第 r 个分量变为一新的正数, 而第 $r+1$ 个到第 n 个分量全都被化为 0, 即分别有

$$\mathbf{y} = \mathbf{R}_m \cdots \mathbf{R}_{r,r+2} \mathbf{R}_{r,r+1} \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_{r-1} \\ y_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (1.2.32)$$

或

$$\mathbf{y} = \mathbf{x} \mathbf{R}_{r,r+1} \mathbf{R}_{r,r+2} \cdots \mathbf{R}_m = (x_1, \cdots, x_{r-1}, y_r, 0, \cdots, 0), \quad (1.2.33)$$

其中 $y_r = \sqrt{x_r^2 + \cdots + x_n^2} > 0$, 而 $\mathbf{R}_{r,r+1}, \cdots, \mathbf{R}_m$ 中的 c 和 s 按式(1.2.29)或式(1.2.31)选取.

例 1.2.16 已知向量 $\mathbf{x} = (1, 2, 3)^\top$, 试用初等旋转变换化 \mathbf{x} 使 \mathbf{x} 与 $\mathbf{e}_1 = (1, 0, 0)^\top$ 同方向.

解 利用式(1.2.29), 作

$$c = \cos \theta = \frac{x_1}{\sqrt{x_1^2 + x_2^2}} = \frac{1}{\sqrt{5}},$$

$$s = \sin \theta = \frac{x_2}{\sqrt{x_1^2 + x_2^2}} = \frac{2}{\sqrt{5}}.$$

于是得初等旋转变换的矩阵

$$\mathbf{R}_{12} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} & 0 \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

从而得

$$\mathbf{y} = \mathbf{R}_{12} \mathbf{x} = (\sqrt{5}, 0, 3)^\top.$$

同样再作

$$c = \frac{x_1}{\sqrt{x_1^2 + x_3^2}} = \sqrt{\frac{5}{14}}, \quad s = \frac{x_3}{\sqrt{x_1^2 + x_3^2}} = \frac{3}{\sqrt{14}},$$

由式(1.2.32)有

$$\hat{\mathbf{y}} = \mathbf{R}_{13} \mathbf{R}_{12} \mathbf{x} = \mathbf{R}_{13} \mathbf{y} = \begin{pmatrix} \sqrt{\frac{5}{14}} & 0 & \frac{3}{\sqrt{14}} \\ 0 & 1 & 0 \\ -\frac{3}{\sqrt{14}} & 0 & \sqrt{\frac{5}{14}} \end{pmatrix} \begin{pmatrix} \sqrt{5} \\ 0 \\ 3 \end{pmatrix}$$

$$= (\sqrt{14}, 0, 0)^\top = \sqrt{14} \mathbf{e}_1.$$

在第3章将会看到, 一个实可逆矩阵可以看成是由 n 个列向量(或行向量)组成, 因此它左连乘以一系列初等旋转矩阵, 就可使 \mathbf{A} 简化成结构较为简单的上三角矩阵.

2) 镜像变换

由例 1.2.11 已经知道, 平面 \mathbb{R}^2 中的反射变换 \mathcal{A} 是正交变换, 它在标准正交基 \mathbf{i}, \mathbf{j} 下的

矩阵 $H = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ 是正交矩阵, 而且 $\det H = -1$.

设 \mathbb{R}^2 中向量 ξ 关于 x 轴反射后变为 η , 则有

$$\eta = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \xi = H\xi.$$

我们把 $H = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ 称为 \mathbb{R}^2 中关于 x 轴的反射阵.

其实, 反射阵 H 可以用与 x 轴正交的单位向量 $\omega = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 来表示, 即

$$H = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - 2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} (0, 1) = I - 2\omega\omega^T.$$

现在把 \mathbb{R}^2 中关于 x 轴的反射变换, 先推广到关于与某单位坐标向量 ω 正交的轴 l 的反射变换(图 1.7), 在此基础上, 就很自然地引出欧氏空间 \mathbb{R}^n 的镜像变换的定义.

为此, 由几何直观从图 1.7 可得

$$\xi - \eta = 2\omega(\omega^T \xi),$$

其中 $\omega^T \xi$ 表示向量 ξ 在 ω 方向上的投影长度(即内积).

稍作整理, 可得

$$\eta = \xi - 2\omega(\omega^T \xi) = (I - 2\omega\omega^T)\xi. \quad (1.2.34)$$

式(1.2.34)表明, 由已知的坐标单位向量 ω 可以构造一个矩阵 $I - 2\omega\omega^T$, 以这矩阵左乘 ξ , 即得 ξ 的像 η . 这就意味着在 \mathbb{R}^2 中确定了一个线性变换.

下面我们来说明, 由坐标变换关系式(1.2.34)所确定的线性变换的确是关于 l 对称的反射变换.

事实上, 令 $H = I - 2\omega\omega^T$, 则 $\eta = H\xi$, 再按顺时针方向旋转坐标系, 使 ω 与 $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 重合, 亦即 l 轴与 x 轴重合.

$$\omega = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 \\ 1 \end{bmatrix} = e_2$$

的初等旋转阵为

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

且 $Q^{-1} = Q^T$, 于是 $Q\eta = QHQ^{-1}Q\xi$, 而

$$\begin{aligned} QHQ^{-1} &= Q(I - 2\omega\omega^T)Q^T = I - 2Q\omega\omega^TQ^T \\ &= I - 2(Q\omega)(Q\omega)^T \\ &= I - 2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} (0, 1) \\ &= I - 2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \end{aligned}$$

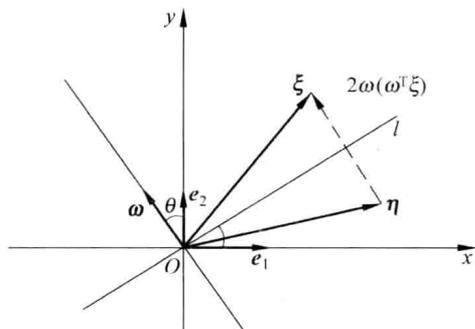


图 1.7

这就是关于 x 轴的反射阵,那么 $Q\eta$ 和 $Q\xi$ 就是关于 x 轴对称的.由此可知, η 和 ξ 是关于与 ω 垂直的轴 l 对称的.因此,由式(1.2.34)确定的变换是关于 l 轴的一种反射变换.由于

$$H = Q^{-1} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} Q,$$

根据正交矩阵的乘积仍是正交矩阵,正交矩阵的逆仍是正交矩阵的结果可知,矩阵 $H = I - 2\omega\omega^T$ 是正交矩阵,且有

$$\det H = \det(I - 2\omega\omega^T) = -1$$

定义 1.2.11 在欧氏空间 \mathbb{R}^n 中,设有线性变换将向量 ξ 映射成与单位向量 ω 正交的 $n-1$ 维子空间对称的像 η ,且有

$$\eta = (I - 2\omega\omega^T)\xi = H\xi,$$

则称这种线性变换为镜像变换,或豪斯霍尔德(Householder)变换,其中的矩阵

$$H = I - 2\omega\omega^T \quad (1.2.35)$$

称为初等反射矩阵,或豪斯霍尔德矩阵.

初等反射矩阵具有如下的性质.

定理 1.2.8 设 $H = I - 2\omega\omega^T$ 是 \mathbb{R}^n 中的初等反射矩阵,则

- (1) H 是对称的正交矩阵;
- (2) $\det H = \det(I - 2\omega\omega^T) = -1$.

证明 (1) 可以直接验证 $H^T = H$,且满足 $H^T H = I$,即

$$\begin{aligned} H^T H &= (I - 2\omega\omega^T)^T (I - 2\omega\omega^T) = (I - 2\omega\omega^T)^2 \\ &= I - 4\omega\omega^T + 4\omega(\omega^T \omega)\omega^T \\ &= I - 4\omega\omega^T + 4\omega\omega^T = I. \end{aligned}$$

所以 H 不仅对称,而且是正交矩阵.

(2) 如上一段所述,我们可以通过有限次的旋转使 ω 的第一个分量为正,其余分量全为 0(即 ω 与 $e_1 = (1, 0, \dots, 0)^T$ 同方向).因为 ω 是单位向量,所以通过有限次的旋转可使 ω 与 e_1 相同.令有限个这样的旋转矩阵的乘积为 Q ,则 $Q\omega = e_1$, Q 是正交矩阵.

另一方面,因为 $\eta = H\xi$,所以 $Q\eta = QHQ^{-1}Q\xi$,其中

$$\begin{aligned} QHQ^{-1} &= Q(I - 2\omega\omega^T)Q^{-1} = I - 2(Q\omega)(Q\omega)^T \\ &= I - 2 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} (1, 0, \dots, 0) = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}, \end{aligned} \quad (1.2.36)$$

从而

$$Q\eta = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} Q\xi.$$

由此可见,对于旋转后的向量 $Q\eta$ 与 $Q\xi$ 而言,仅第一个分量在此变换下改变了符号,也就是说 $Q\eta$ 可以看成是 $Q\xi$ 关于与 e_1 正交的 $n-1$ 维子空间的镜像.

由式(1.2.36)有

$$\mathbf{H} = \mathbf{Q}^{-1} \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} \mathbf{Q},$$

右端每一个矩阵都是正交矩阵,所以 \mathbf{H} 是正交矩阵,且

$$\det \mathbf{H} = \det(\mathbf{I} - 2\boldsymbol{\omega}\boldsymbol{\omega}^T) = -1. \quad \text{证毕}$$

定理 1.2.9 镜像变换可以使任何非零向量 $\boldsymbol{\xi}$ 变成与给定单位向量 $\boldsymbol{\zeta}$ 同方向的向量 $\boldsymbol{\eta}$ (图 1.8).

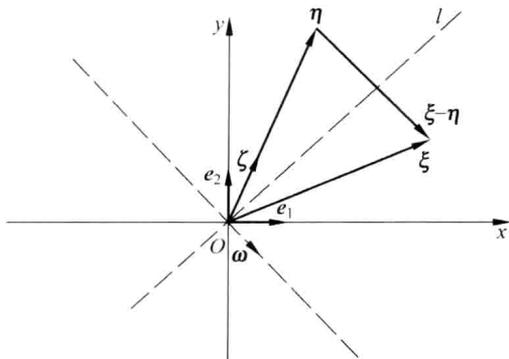


图 1.8

证明 显然此时 $\boldsymbol{\xi}$ 的像为

$$\boldsymbol{\eta} = |\boldsymbol{\xi}| \boldsymbol{\zeta}.$$

问题的关键是如何构造单位向量 $\boldsymbol{\omega}$, 作 $\mathbf{H} = \mathbf{I} - 2\boldsymbol{\omega}\boldsymbol{\omega}^T$, 使 $\boldsymbol{\eta} = \mathbf{H}\boldsymbol{\xi}$.

下面我们来导出满足这一要求的 $\boldsymbol{\omega}$ 的计算公式. 容易看出, $\boldsymbol{\xi}$ 在 $\boldsymbol{\omega}$ 方向的投影为

$$\begin{aligned} \frac{1}{2} |\boldsymbol{\xi} - \boldsymbol{\eta}| &= |\boldsymbol{\omega}| |\boldsymbol{\xi}| \cos(\boldsymbol{\omega}, \boldsymbol{\xi}) = \boldsymbol{\omega} \cdot \boldsymbol{\xi} \\ &= (\boldsymbol{\omega}, \boldsymbol{\xi}) = \boldsymbol{\omega}^T \boldsymbol{\xi}, \end{aligned}$$

所以向量 $\boldsymbol{\xi} - \boldsymbol{\eta} = 2(\boldsymbol{\omega}^T \boldsymbol{\xi})\boldsymbol{\omega}$, 从而得

$$\boldsymbol{\omega} = \frac{\boldsymbol{\xi} - \boldsymbol{\eta}}{2(\boldsymbol{\omega}^T \boldsymbol{\xi})} = \frac{\boldsymbol{\xi} - |\boldsymbol{\xi}| \boldsymbol{\zeta}}{|\boldsymbol{\xi} - |\boldsymbol{\xi}| \boldsymbol{\zeta}|}. \quad (1.2.37)$$

用这样的 $\boldsymbol{\omega}$ 作成的 \mathbf{H} 必有

$$\begin{aligned} \boldsymbol{\eta} &= \mathbf{H}\boldsymbol{\xi} \\ &= \left(\mathbf{I} - 2 \frac{(\boldsymbol{\xi} - |\boldsymbol{\xi}| \boldsymbol{\zeta})(\boldsymbol{\xi} - |\boldsymbol{\xi}| \boldsymbol{\zeta})^T}{|\boldsymbol{\xi} - |\boldsymbol{\xi}| \boldsymbol{\zeta}|^2} \right) \boldsymbol{\xi} \\ &= \boldsymbol{\xi} - 2(\boldsymbol{\xi} - |\boldsymbol{\xi}| \boldsymbol{\zeta}, \boldsymbol{\xi}) \frac{\boldsymbol{\xi} - |\boldsymbol{\xi}| \boldsymbol{\zeta}}{|\boldsymbol{\xi} - |\boldsymbol{\xi}| \boldsymbol{\zeta}|^2} \\ &= \boldsymbol{\xi} - (\boldsymbol{\xi} - |\boldsymbol{\xi}| \boldsymbol{\zeta}) = |\boldsymbol{\xi}| \boldsymbol{\zeta}. \end{aligned}$$

注意, 这里利用了等式

$$2(\boldsymbol{\xi} - |\boldsymbol{\xi}| \boldsymbol{\zeta}, \boldsymbol{\xi}) = |\boldsymbol{\xi} - |\boldsymbol{\xi}| \boldsymbol{\zeta}|^2. \quad \text{证毕}$$

则称这样的变换为初等酉变换.

由 $\mathbf{R}^H \mathbf{R} = \mathbf{I}$, 可知 \mathbf{R}_{ij} 是酉矩阵. 当 $\lambda = \mu = 0$ 时, \mathbf{R}_{ij} 是实的正交矩阵.

还有一种常用的酉变换叫做初等埃尔米特变换. 它的矩阵表示是

$$\mathbf{P} = \mathbf{I} - 2\omega\omega^H,$$

其中 ω 满足 $\omega^H \omega = 1$.

显然 \mathbf{P} 为酉矩阵.

4. 正交投影变换与正交投影矩阵

前面我们讨论了内积空间中的两类常用的正交变换(旋转与反射), 其实, 内积空间中还有一种叫做正交投影变换.

先考察 \mathbb{R}^3 中正交投影的含义, 在此基础上推广, 即得内积空间中正交投影变换的定义.

如图 1.9, 若将向量

$$\overrightarrow{OP} = xi + yj + zk$$

投影到坐标面 xOy 上, 得到另一向量

$$\overrightarrow{OP'} = x'i + y'j + z'k,$$

即有 $x' = x, y' = y, z' = 0$. 如把这个投影变换记为 \mathcal{A} , 显然它是线性变换. 因为容易验证: 若 $\mathcal{A}(\alpha) = \alpha', \mathcal{A}(\beta) = \beta'$, 则

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

$$\mathcal{A}(\lambda\alpha) = \lambda\mathcal{A}(\alpha) = \lambda\alpha'.$$

又因为

$$\begin{cases} \mathcal{A}(i) = i = 1i + 0j + 0k, \\ \mathcal{A}(j) = j = 0i + 1j + 0k, \\ \mathcal{A}(k) = \mathbf{0} = 0i + 0j + 0k, \end{cases}$$

或

$$\mathcal{A}(i, j, k) = (i, j, k) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

故此投影变换的矩阵为

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

由图 1.9 可见, 正交投影的概念实质上是反映了向量沿子空间 V_1 (z 轴) 到子空间 V_2 (xOy 平面) 的一种变换关系. 此时, V_1 与 V_2 是互相垂直的. 因此, 为了推广正交投影的含义, 必须先定义内积空间中子空间的正交关系.

1) 内积空间中子空间的正交

定义 1.2.14 设 V_1 与 V_2 是内积空间 V^n 的两个子空间, 如果对任意的 $x_1 \in V_1, x_2 \in V_2$, 都有

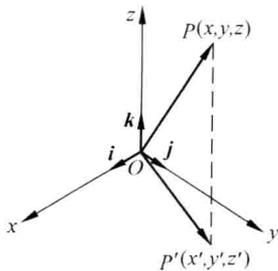


图 1.9

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

则 \mathcal{A} 在基 $\{\boldsymbol{\varepsilon}'_i\}$ 下的矩阵为 $\mathbf{B} = \mathbf{C}^{-1}\mathbf{A}\mathbf{C}$, 即

$$\mathbf{B} = \mathbf{C}^{-1} \text{diag}(1, 1, \dots, 1, 0, \dots, 0) \mathbf{C}. \quad (1.2.42)$$

↑
第 r 个

可以证明, 正交投影矩阵还有如下的性质.

定理 1.2.15 设 \mathbf{A} 是 n 阶实方阵, 则下面 3 个条件等价:

- (1) \mathbf{A} 是正交投影矩阵;
- (2) $\mathbf{A} = \mathbf{A}^T \mathbf{A}$;
- (3) $\mathbf{A}^2 = \mathbf{A} = \mathbf{A}^T$.

对其证明有兴趣的诸者, 可参阅参考文献[18].

应当注意到, 正交阵与正交投影阵是两个不相干的概念.

例 1.2.20 证明:

- (1) $\mathbf{A} = \begin{bmatrix} 1 & \\ & -1 \end{bmatrix}$ 是正交阵, 但不是正交投影阵;
- (2) $\mathbf{B} = \begin{bmatrix} 1 & \\ & 0 \end{bmatrix}$ 是正交投影阵, 但不是正交阵.

证明 (1) 因为 $\mathbf{A}^T \mathbf{A} = \mathbf{I}$, 所以 \mathbf{A} 是正交阵, 但 $\mathbf{A}^2 = \mathbf{I} \neq \mathbf{A}$, 所以 \mathbf{A} 不是正交投影阵.

(2) 因为 $\mathbf{B}^2 = \mathbf{B} = \mathbf{B}^T$, 所以 \mathbf{B} 是正交投影阵; 但 $|\mathbf{B}| = 0$, 而正交阵的行列式只能是 ± 1 , 所以 \mathbf{B} 不是正交阵. 证毕

最后, 我们还应注意到, 在正交投影的定义 1.2.16 中, 内积空间 V^n 作正交分解 $V^n = V_1 \oplus V_2$, 这一点尤其重要. 如果改写为直和 $V^n = V_1 \oplus V_2$, 此时沿 V_2 到 V_1 的投影 $\mathcal{A}(\mathbf{x}) = \mathbf{x}_1$ 就不再是正交投影, 前者为“直角”坐标系, 后者一般为“仿射”坐标系, 如图 1.10 所示.

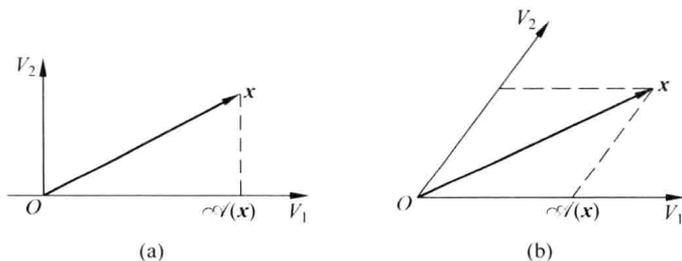


图 1.10

习 题 1 (4)

1. 设 $\boldsymbol{\beta}$ 是欧氏空间 V^n 中的单位向量, $\boldsymbol{\alpha} \in V^n$, 定义线性变换 \mathcal{A} 为: $\mathcal{A}(\boldsymbol{\alpha}) = \boldsymbol{\alpha} - 2(\boldsymbol{\beta}, \boldsymbol{\alpha})\boldsymbol{\beta}$, 证明 \mathcal{A} 是正交变换.

2. 证明: 欧氏空间中两个正交变换的乘积也是正交变换; 正交变换的逆变换也是正交变换.
3. 在欧氏空间 V^n 中, 取定标准正交基 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, 试证: 对任意正交矩阵 A 都惟一决定一个正交变换.

4. 试证线性变换 \mathcal{A} 为正交变换的充分必要条件是: 使任两点间的距离保持不变, 即

$$|\mathcal{A}(\alpha) - \mathcal{A}(\beta)| = |\alpha - \beta|.$$

5. 求证: 线性变换 \mathcal{A} 为正交变换的必要条件为: 保持任意两个向量的夹角不变. 说明条件不充分.
6. 设 u_1, u_2, \dots, u_m 与 v_1, v_2, \dots, v_m 为欧氏空间 V 的两组标准正交向量组. 证明: 必有正交变换把第一组变为第二组.
7. 证明: 第1题中的正交变换 \mathcal{A} 是第二类正交变换(即对应矩阵的行列式为 -1), 又称镜面反射.
8. (1) 设 α, β 是 n 维欧氏空间 V^n 中两个不同的单位向量, 证明: 存在镜面反射 \mathcal{A} , 使 $\mathcal{A}(\alpha) = \beta$.
- (2) 证明: n 维欧氏空间 V^n 中任一正交变换都可以表示成一系列镜像变换的乘积.
9. 如果 α, β 是欧氏空间 V 中长度相同的向量, 试证明必有正交变换 \mathcal{A} , 使 $\mathcal{A}(\alpha) = \beta$.
10. 设欧氏空间 V^n 的基 e_1, e_2, \dots, e_n 的度量矩阵为 G , 正交变换 \mathcal{A} 在该基下的矩阵为 A . 证明: $A^T G A = G$.

11. 试用初等旋转变换变化向量 $\alpha = (2, 3, 0, 5)^T$, 使所得向量与 $e_1 = (1, 0, 0, 0)^T$ 同方向.

12. 试用镜像变换(豪斯霍尔德变换)将向量 $\alpha = (2, 2, 0, 1)^T$ 变为与 $e_1 = (1, 0, 0, 0)^T$ 同方向的向量.

13. 已知 $u = \left(\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)^T$ 为单位向量, 作一初等反射矩阵 H , 并验证有 $H^{-1} = H$.

14. 设有线性变换定义为 $Hx = x - \lambda(x, \omega)\omega$, 其中 $x \in \mathbb{R}^n$, 而 ω 是欧氏长度为1的列向量, 问 λ 取何值时, H 是正交矩阵.

15. 求两个正交矩阵, 以 $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ 及 $\left(\frac{1}{6}, \frac{1}{6}, \frac{1}{2}, -\frac{5}{6}\right)$ 为前两行.

16. 证明: 上三角的正交矩阵必为对角矩阵, 且对角线上的元素为 $+1$ 或 -1 .

17. 设 A 为实对称阵, S 为实反对称矩阵, 有关系式 $AS = SA$, 而且 $A - S$ 非奇异, 求证: $(A + S)(A - S)^{-1}$ 是正交矩阵.

18. 设 A 与 B 都是正交矩阵, $|A| = -|B|$, 求证: $|A + B| = 0$.

19. 证明: 方阵 A 为正交矩阵必要且只要 A^T 为正交矩阵.

20. 证明: 正交矩阵的伴随矩阵也是正交矩阵.

21. 如果 A 是正交矩阵, 证明: 当 P 是正交矩阵时, $P^{-1}AP$ 也是正交的. 试举例说明 A 为正交矩阵时, 虽然 P 不是正交矩阵, 但 $P^{-1}AP$ 也可能还是正交的.

22. 设 A 是正交矩阵, 证明: $A^{-1}SA$ 是对称矩阵或是反对称矩阵, 由 S 是对称矩阵或是反对称矩阵而定.

23. 设 \mathcal{A} 为 n 维欧氏空间 \mathbb{R}^n 到其本身的一个变换, 其定义如下:

$$\mathcal{A}(\alpha) = \alpha - k(\alpha, \beta)\beta, \alpha \in \mathbb{R}^n,$$

其中 β 是 \mathbb{R}^n 中一个单位向量, 求 k 取何值时, \mathcal{A} 是正交变换.

24. 设 A 为 n 阶反对称矩阵(即 $A^T = -A$), 证明:

(1) A 的特征值为0或纯虚数, 且 $I - A$ 是可逆的;

(2) $S = (I + A)(I - A)^{-1}$ 是正交矩阵;

(3) 设 $\alpha + \beta i$ 是 A 的一个非零特征值的特征向量, 其中 α, β 均为实向量, 则 α 与 β 正交.

25. 设 $\alpha_1, \alpha_2, \dots, \alpha_n$ 与 $\beta_1, \beta_2, \dots, \beta_n$ 是 n 维欧氏空间两个线性无关的向量组, 证明存在正交变换 \mathcal{A} , 使 $\mathcal{A}(\alpha_i) = \beta_i, i = 1, 2, \dots, n$ 的充要条件是:

$$(\alpha_i, \alpha_j) = (\beta_i, \beta_j), \quad i, j = 1, 2, \dots, n.$$

26. (1) 设 A 是一个 n 阶实矩阵, 如果只要求 A 的列向量构成正交向量组, 问 $A^T A$ 有何特点? 此时, 能否找到方阵 B , 使得 $Q = AB$ 为正交阵;

(2) 如果 $|A| = -1$, 并且 A 的每一个元素都等于其本身的代数余子式乘以 -1 , 那么 A 是一个正交矩阵.

40. 设 A 为可逆的实对称矩阵, B 为实反对称矩阵; 且 $AB = BA$, 证明:

(1) $A - B$ 与 $A + B$ 均可逆;

(2) $(A + B)(A - B)^{-1}$ 与 $(A - B)^{-1}(A + B)$ 都是正交矩阵.

41. 设 n 维欧氏空间 V^n 的一组基 $\alpha_1, \alpha_2, \dots, \alpha_n$, 证明: 存在正定矩阵 C , 使得由

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

确定的向量组 $\beta_1, \beta_2, \dots, \beta_n$ 是 V^n 的一个标准正交基.

42. 已知

$$A = \begin{bmatrix} 3 & 0 & 8 \\ 3 & -1 & 6 \\ -2 & 0 & -5 \end{bmatrix},$$

试求 U 矩阵, 使得 $U^H A U$ 是上三角矩阵.

43. 证明: $(u, v)(v, u) \leq (u, u)(v, v)$, 其中 u, v 为西空间中任意向量.

44. 用向量 $\alpha_1 = (1, 0, 2, 1)^T, \alpha_2 = (2, 1, 2, 3)^T, \alpha_3 = (0, 1, -2, 1)^T$ 生成子空间 V , 求 V 的正交补 V^\perp 的基底及正交补空间 V^\perp .

45. 设 V_1, V_2 是欧氏空间 V 的两个子空间, 证明:

$$(V_1 + V_2)^\perp = V_1^\perp \cap V_2^\perp,$$

$$(V_1 \cap V_2)^\perp = V_1^\perp + V_2^\perp.$$

46. 求齐次线性方程组 $AX = 0$ 的解空间的 $N(A)$ 的一组标准正交基及 $N^\perp(A)$ 的一组标准正交基, 其中

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}.$$

47. 设 W 与 U 是 V^n 的两个子空间, 且 $V^n = W \oplus U$, 设 W 的基为 $\alpha_1, \alpha_2, \dots, \alpha_r, U$ 的基为 $\beta_1, \beta_2, \dots, \beta_{n-r}$, 则 U 是 W 的正交补的充要条件是 W 与 U 的两组基互相正交, 即 $(\alpha_i, \beta_j) = 0, i = 1, 2, \dots, r; j = 1, 2, \dots, n - r$.

48. 设 $P[x]_3$ 是次数不大于 3 的实域上多项式空间, 定义内积:

$$(f(x), g(x)) = \int_0^1 f(x)g(x)dx, \quad \forall f(x), g(x) \in P[x]_3,$$

设 W 是次数为零的多项式子空间, 求 W^\perp 的一个基.

49. 设实域的多项式空间 $P[x]_3$ 中, 两子空间 $V_1 = L[1, x], V_2 = L\left[x^2 - \frac{1}{3}, x^3 - \frac{3}{5}x\right]$, 在 $P[x]_3$ 中定义内积

$$(f(x), g(x)) = \int_{-1}^1 f(x)g(x)dx.$$

证明 V_1 与 V_2 互为正交补.

1.3 埃尔米特变换及其矩阵

在内积空间中, 除了 1.2 节已讨论了等积线性变换外, 还有一类十分重要的具有自伴性质(实数情况为对称)的线性变换(不一定等积), 它在力学、物理学、自动控制等工程技术中有着广泛的应用. 本节先介绍对称变换及对称矩阵, 再讨论更一般的埃尔米特变换与埃尔米特矩阵, 特别要讨论具有特殊性质的埃尔米特矩阵——正定(非负定)矩阵及其性质.

事实上,只要利用式(1.3.1)及式(1.1.18)即得.

(2) 若 A 不是对称矩阵, $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, 则有

$$(Ax, y) = (x, A^T y). \quad (1.3.4)$$

事实上, $(Ax, y) = (Ax)^T y = x^T A^T y = (x, A^T y)$.

2. 埃尔米特变换与埃尔米特矩阵

定义 1.3.2 如果酉空间 V 上的线性变换 \mathcal{A} 对于任何 $x, y \in V$ 恒有

$$(\mathcal{A}(x), y) = (x, \mathcal{A}(y)), \quad (1.3.5)$$

则说 \mathcal{A} 是埃尔米特变换.

若埃尔米特变换 \mathcal{A} 关于标准正交基的矩阵为 $A = (a_{ij})_{n \times n}$, 则对于任何 i, j 恒有

$$a_{ij} = \bar{a}_{ji} \quad \text{或} \quad A^H = A. \quad (1.3.6)$$

满足式(1.3.6)的矩阵 A 称为埃尔米特矩阵, 又称自伴矩阵. 若满足条件

$$A^H = -A,$$

则称 A 为反埃尔米特矩阵. 其中 A^H 表示矩阵 A 中的元素共轭后再转置.

引理 设 ξ_1 为 n 维单位列向量, 则存在 n 维列向量 ξ_2, \dots, ξ_n , 使得矩阵 $P = (\xi_1, \xi_2, \dots, \xi_n)$ 为酉矩阵.

证明 设 $\xi_1 = (a_1, a_2, \dots, a_n)^T$, 其中 $a_1 \neq 0$, 则向量组 ξ_1, e_2, \dots, e_n 线性无关(其中 e_j 为自然基向量). 用正交化方法可得标准正交向量组 $\xi_1, \xi_2, \dots, \xi_n$, 故 $P = (\xi_1, \xi_2, \dots, \xi_n)$ 为酉矩阵. 当 $a_1 = 0$ 时, 必有 $a_i \neq 0$ (i 为 $2 \sim n$ 间的某个整数), 我们用类似的方法可证明引理结论成立. 证毕

定理 1.3.2 (舒尔(Schur)定理) 任何 n 阶矩阵都酉相似于一个上三角阵, 即存在一个 n 阶酉矩阵 U 和一个上三角阵 R , 使得

$$U^H A U = T \quad \text{或} \quad A = U T U^H, \quad (1.3.7)$$

式中 T 的主对角元是 A 的特征值, 它们可以按所要求的次序排列.

证明 对矩阵阶数作归纳法. 对一阶矩阵结论是显然的. 假设阶数为 $n-1$ 时结论成立, 对 n 阶矩阵 A , 我们任取它的一个特征值 λ_1 , 相应的单位特征向量为 ξ_1 , 于是

$$A \xi_1 = \lambda_1 \xi_1, \quad \xi_1^H \xi_1 = 1. \quad (1.3.8)$$

由引理可知, 存在 n 维向量 ξ_2, \dots, ξ_n , 使矩阵 $P = (\xi_1, \xi_2, \dots, \xi_n)$ 为酉矩阵. 由式(1.3.8)得

$$\begin{aligned} P^H A P &= \begin{pmatrix} \xi_1^H \\ \xi_2^H \\ \vdots \\ \xi_n^H \end{pmatrix} A (\xi_1, \xi_2, \dots, \xi_n) \\ &= \begin{pmatrix} \lambda_1 & \xi_1^H A \xi_2 & \cdots & \xi_1^H A \xi_n \\ 0 & \xi_2^H A \xi_2 & \cdots & \xi_2^H A \xi_n \\ \vdots & \vdots & & \vdots \\ 0 & \xi_n^H A \xi_2 & \cdots & \xi_n^H A \xi_n \end{pmatrix} = \begin{pmatrix} \lambda_1 & \beta \\ \mathbf{0} & A_1 \end{pmatrix}, \end{aligned} \quad (1.3.9)$$

式中 β 为 $1 \times (n-1)$ 矩阵, $\mathbf{0}$ 为 $(n-1) \times 1$ 零矩阵, A_1 为 $n-1$ 阶矩阵. 根据归纳假设, 存在

$n-1$ 阶酉矩阵 U_1 和上三角阵 T_1 , 使得

$$U_1^H A_1 U_1 = T_1 \quad \text{或} \quad A_1 = U_1 T_1 U_1^H,$$

式中 T_1 对角线上的元素是 A_1 的特征值(当然也是 A 的特征值), 它们可按所要求的次序排列. 这样, 由式(1.3.9)有

$$\begin{aligned} P^H A P &= \begin{bmatrix} \lambda_1 & \beta \\ \mathbf{0} & U_1 T_1 U_1^H \end{bmatrix} \\ &= \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_1 \end{bmatrix} \begin{bmatrix} \lambda_1 & \beta U_1 \\ \mathbf{0} & T_1 \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_1^H \end{bmatrix}. \end{aligned} \quad (1.3.10)$$

令

$$U = P \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_1 \end{bmatrix}, \quad T = \begin{bmatrix} \lambda_1 & \beta U_1 \\ \mathbf{0} & T_1 \end{bmatrix},$$

则 U 为酉矩阵, T 为上三角阵. 由式(1.3.10)得

$$U^H A U = T.$$

可见矩阵阶数为 n 时结论也成立.

证毕

推论 若 A 为埃尔米特矩阵, 则 A 必酉相似于对角阵, 其对角元(A 的特征值)均为实数.

证明 利用定理 1.3.2, 再由 $A^H = A$ 可得

$$T^H = T.$$

这说明 T 是对角阵, 且对角元都是实数.

证毕

类似地, 若 A 是反埃尔米特矩阵, 则 A 必酉相似于对角阵, 其对角元(A 的特征值)均为零或纯虚数.

1.3.2 埃尔米特正定、半正定矩阵

这里先讨论埃尔米特正定与半正定矩阵的一些性质, 在此基础上下一段还将介绍矩阵不等式, 它在现代控制理论等学科中是很有用处的.

定义 1.3.3 设 A 为 n 阶埃尔米特矩阵, 如果对任意 n 维复向量 x 都有

$$x^H A x \geq 0, \quad (1.3.11)$$

则称 A 为非负定(半正定)矩阵, 记作 $A \geq 0$. 如果对任意 n 维非零复向量 x 都有

$$x^H A x > 0, \quad (1.3.12)$$

则称 A 为正定矩阵, 记作 $A > 0$.

值得提出的是, 矩阵 A 不是数, 本不应有大小, 这里引用不等式仅是一种记号而已.

按此定义, 显然有下列简单事实:

- (1) 单位矩阵 $I > 0$;
- (2) 若 $A > 0$, 数 $k > 0$, 则 $kA > 0$;
- (3) 若 $A \geq 0, B \geq 0$, 则 $A + B \geq 0$.

现在我们来讨论埃尔米特矩阵为正定(非负定)矩阵的条件. 在以下的讨论中, 矩阵 A 均指 n 阶埃尔米特矩阵.

定理 1.3.3 矩阵 A 为正定(非负定)矩阵的充分必要条件是 A 的所有特征值都是正数(非负数).

证明 设 $A > 0$ ($A \geq 0$), λ 是 A 的任一特征值, ξ 是对应的单位特征向量. 于是

$$\lambda = \xi^H A \xi > 0 \quad (\geq 0).$$

这就证明了条件的必要性.

再证条件的充分性. 由定理 1.3.2 的推论, 存在酉矩阵 V , 使得

$$A = V^H \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) V.$$

若 A 的特征值 λ_i ($i=1, \dots, n$) 都为正数(非负数), 则对任意 n 维非零向量 x , 都有

$$\begin{aligned} x^H A x &= (Vx)^H \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) (Vx) \\ &= y^H \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) y > 0 \quad (\geq 0), \end{aligned}$$

式中 $y = Vx \neq 0$. 这说明 $A > 0$ ($A \geq 0$). 证毕

推论 若 $A > 0$, 则 $\text{tr}(A) > \lambda_i$ ($i=1, \dots, n$). 若 $A \geq 0$, 则 $\text{tr}(A) \geq 0$ ($i=1, \dots, n$). 这里 λ_i 表示 A 的特征值, $\text{tr}(A)$ 表示 A 的对角元之和, 称为 A 的迹.

证明 记 $A = (a_{ij})_{n \times n}$, 其特征多项式为

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

易见

$$a_1 = -(a_{11} + a_{22} + \dots + a_{nn}) = -\text{tr}(A).$$

另一方面, 由多项式根与系数的关系有

$$a_1 = -(\lambda_1 + \lambda_2 + \dots + \lambda_n).$$

从而 $\text{tr}(A) = \lambda_1 + \lambda_2 + \dots + \lambda_n$. 由定理 1.3.3 立即得本推论.

定理 1.3.4 矩阵 A 为正定矩阵的充分必要条件是存在 n 阶非奇异矩阵 P , 使得 $A = P^H P$.

证明 条件的充分性是明显的, 下面证条件的必要性. 由定理 1.3.2 的推论知, 存在酉矩阵 V , 使得

$$A = V^H \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) V. \quad (1.3.13)$$

若 $A > 0$, 则由定理 1.3.3 可知, $\lambda_i > 0$ ($i=1, \dots, n$). 令

$$P = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n}) V,$$

则 P 非奇异, 且由式(1.3.13)有 $A = P^H P$. 证毕

若将条件中的“非奇异”去掉就得到 A 为非负定矩阵的充分必要条件.

定理 1.3.4' 矩阵 A 为非负定(半正定)矩阵的充分必要条件是存在 n 阶矩阵 P , 使得 $A = P^H P$.

推论 1 若 $A > 0$, 则 A 可逆且 $A^{-1} > 0$.

推论 2 若 $A > 0$, C 是任一 n 阶非奇异矩阵, 则 $C^H A C > 0$.

推论 3 若 $A \geq 0$, C 是任一 $n \times m$ 矩阵, 则 $C^H A C \geq 0$.

推论 1 的证明 当 $A > 0$ 时, 由定理 1.3.4 知存在非奇异矩阵 P , 使得 $A = P^H P$. 可见 A 可逆且有

$$A^{-1} = P^{-1} (P^H)^{-1} = P^{-1} (P^{-1})^H = Q^H Q,$$

其中 $Q = (P^{-1})^H$ 为非奇异矩阵. 由定理 1.3.4 可推得 $A^{-1} > 0$. 证毕

推论 2 和 **推论 3** 由读者自行证明.

定理 1.3.5 正定矩阵 A 的各阶顺序主子矩阵都是正定矩阵.

证明 设 A_k 是 A 的 k 阶顺序主子矩阵, \tilde{x} 是任意 k 维非零向量 ($1 \leq k \leq n$), 令 $x = \begin{bmatrix} \tilde{x} \\ \mathbf{0} \end{bmatrix}$, 其中 $\mathbf{0}$ 为 $n-k$ 维零向量. 将 A 作如下分块:

$$A = \begin{bmatrix} A_k & G \\ G^H & B \end{bmatrix},$$

于是

$$\begin{aligned} \tilde{x}^H A_k \tilde{x} &= (\tilde{x}^H, \mathbf{0}^H) \begin{bmatrix} A_k & G \\ G^H & B \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \mathbf{0} \end{bmatrix} \\ &= x^H A x > 0, \end{aligned}$$

这说明 A_k 是正定矩阵.

证毕

定理 1.3.6 矩阵 A 为正定矩阵的充分必要条件是 A 的各阶顺序主子式都为正数, 即

$$|A_k| > 0, \quad k = 1, \dots, n. \quad (1.3.14)$$

证明 先证条件的必要性. 当 $A > \mathbf{0}$ 时, A 的行列式 $\det A > 0$. 这是因为 $\det A$ 等于 A 的特征值的乘积. 由定理 1.3.5 知, A 的各阶顺序主子矩阵都是正定矩阵, 故它们的行列式均为正数, 即式(1.3.14)成立.

再证条件的充分性. 对矩阵阶数作归纳法. 阶为 1 时结论是显然的. 假设阶为 $n-1$ 时结论成立, 对 n 阶矩阵 A , 我们将它如下分块:

$$A = \begin{bmatrix} A_{n-1} & \alpha \\ \alpha^H & a_{nn} \end{bmatrix},$$

其中 A_{n-1} 为 A 的 $n-1$ 阶顺序主子矩阵. 于是

$$A = C^H \begin{bmatrix} A_{n-1} & \mathbf{0} \\ \mathbf{0}^H & b_{nn} \end{bmatrix} C, \quad (1.3.15)$$

式中 $b_{nn} = a_{nn} - \alpha^H A_{n-1}^{-1} \alpha$, 而

$$C = \begin{bmatrix} I_{n-1} & A_{n-1}^{-1} \alpha \\ \mathbf{0}^H & 1 \end{bmatrix},$$

根据归纳假设, 有 $A_{n-1} > \mathbf{0}$, 于是

$$b_{nn} = \frac{|A_n|}{|A_{n-1}|} > 0 \quad (|A_n| = |A|, |C| = 1),$$

从而

$$\begin{bmatrix} A_{n-1} & \mathbf{0} \\ \mathbf{0}^H & b_{nn} \end{bmatrix} > \mathbf{0}. \quad (1.3.16)$$

由推论 2, 从式(1.3.15)和式(1.3.16)可推出 $A > \mathbf{0}$, 这就证明了条件的充分性. 证毕

必须指出, 埃尔米特矩阵 A 的所有顺序主子式均非负, 并不能推出 A 是非负定矩阵. 例

如, 矩阵 $A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ 是埃尔米特矩阵, 且容易验证 A 的所有顺序主子式均非负,

但 A 不是非负定矩阵.

正定矩阵还有如下性质:

定理 1.3.7 设 A, B 都是 n 阶埃尔米特矩阵, 且 $B > \mathbf{0}$, 则存在非奇异矩阵 Q , 使得

$$Q^H B Q = I, \quad Q^H A Q = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n). \quad (1.3.17)$$

证明 由定理 1.3.4 知, 存在非奇异矩阵 P , 使得 $B = P^H P$, 由此得

$$(P^{-1})^H B P^{-1} = I. \quad (1.3.18)$$

又 $(P^{-1})^H A P^{-1}$ 亦为埃尔米特矩阵, 故有酉矩阵 U , 使

$$U^H (P^{-1})^H A P^{-1} U = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n). \quad (1.3.19)$$

令 $Q = P^{-1} U$, 则 Q 非奇异, 由式(1.3.18)和式(1.3.19)可知式(1.3.17)成立. 证毕

定理 1.3.8 n 阶埃尔米特矩阵 A 正定的充分必要条件是 A 的所有主子式全大于零.

证明 必要性: 对 A 的任一 k 阶主子式

$$|A^{(k)}| = \begin{vmatrix} a_{i_1 i_1} & a_{i_1 i_2} & \cdots & a_{i_1 i_k} \\ a_{i_2 i_1} & a_{i_2 i_2} & \cdots & a_{i_2 i_k} \\ \vdots & \vdots & & \vdots \\ a_{i_k i_1} & a_{i_k i_2} & \cdots & a_{i_k i_k} \end{vmatrix},$$

只要适当(若干次)对调 $|A^{(k)}|$ 的行和相应的列, 可使上述的 $|A^{(k)}|$ 成为一个 k 阶顺序主子式, 即存在可逆矩阵 P , 使 $P^H A P$ 的 k 阶顺序主子式为 $|A^{(k)}|$. 因为 $A > 0$, 由前面推论 2 知 $P^H A P > 0$, 从而由定理 1.3.6 有 $|A^{(k)}| > 0$.

充分性: 由定理 1.3.6 即得.

证毕

1.3.3 矩阵不等式

定义 1.3.4 设 A, B 都是 n 阶埃尔米特矩阵, 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 $B \leq A$).

显然, 当 $A \geq B$ 时, 对任意的 n 维列向量 x 都有

$$x^H A x \geq x^H B x.$$

反之亦然. 若 A, B 均为实对角阵:

$$A = \text{diag}(a_1, a_2, \dots, a_n),$$

$$B = \text{diag}(b_1, b_2, \dots, b_n),$$

则不难看出

$$A \geq B \Leftrightarrow a_i \geq b_i, \quad i = 1, \dots, n.$$

值得注意的是, 并非任意两个同阶的埃尔米特矩阵都能“比较大小”, 即并非 $A \geq B$ 或 $B \geq A$ 两者之中必有一成立(即使 A, B 都是正定矩阵). 例如, 对

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

来讲, $A \geq B$ 和 $B \geq A$ 均不成立.

下面讨论矩阵不等式的性质.

性质 1 由定义容易看出

- (1) 若 $A \geq B, B \geq C$, 则 $A \geq C$;
- (2) 若 $A \geq B, k$ 为正数, 则 $kA \geq kB$;
- (3) 若 $A_1 \geq B_1, A_2 \geq B_2$, 则 $A_1 + A_2 \geq B_1 + B_2$;
- (4) 若 $A \geq B, P$ 为 $n \times m$ 矩阵, 则 $P^H A P \geq P^H B P$.

性质 2 若 $A \geq 0$, 则 $A \leq \text{tr}(A)I$.

证明 由定理 1.3.2 推论, 存在酉矩阵 V , 使得

$$A = V^H \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) V.$$

再由定理 1.3.3 的推论, 有

$$\text{tr}(A) \geq \lambda_i, \quad i = 1, \dots, n.$$

因此

$$\begin{aligned} \text{tr}(A)I - A &= \text{tr}(A)I - V^H \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) V \\ &= V^H [\text{tr}(A)I - \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)] V \geq 0. \end{aligned}$$

证毕

性质 3 设 A, B 为同阶的正定矩阵, $A \geq B$, 则 $B^{-1} \geq A^{-1}$.

证明 由定理 1.3.7 可知, 存在非奇异矩阵 Q , 使得

$$Q^H B Q = I, \quad Q^H A Q = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) \equiv \Lambda,$$

因 $A \geq B$, 故 $\Lambda \geq I$. 从而

$$\Lambda^{-1} = \text{diag}(\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_n^{-1}) \leq I,$$

但

$$A^{-1} = Q \Lambda^{-1} Q^H, \quad B^{-1} = Q I Q^H,$$

故有 $B^{-1} \geq A^{-1}$.

证毕

从 $A \geq B > 0$ 能否推出 $A^2 \geq B^2$ 呢? 一般来讲是不能的. 例如, 令

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix},$$

则 $A > 0, B > 0$, 且 $A \geq B$. 但

$$A^2 - B^2 = \begin{bmatrix} 4 & 5 \\ 5 & 6 \end{bmatrix},$$

却不是非负定矩阵, 即 $A^2 \geq B^2$ 不成立. 然而, 我们有以下结果.

性质 4 若正定矩阵 A, B 可交换: $AB = BA$, 且 $A \geq B$, 则 $A^2 \geq B^2$.

证明 与性质 3 的证明相类似, 有

$$Q^H B Q = I, \quad Q^H A Q = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) \equiv \Lambda. \quad (1.3.20)$$

因 $A \geq B$, 故 $\Lambda \geq I$, 从而 $\lambda_i \geq 1 (i=1, \dots, n)$. 将式(1.3.20)改写成

$$B = (Q^H)^{-1} Q^{-1}, \quad A Q Q^H = (Q^H)^{-1} \Lambda Q^H, \quad (1.3.21)$$

由此得

$$A B^{-1} = (Q^H)^{-1} \Lambda Q^H,$$

这说明 $A B^{-1}$ 相似于 Λ , 因而其全部特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$. 由此可知 $(A B^{-1})^2$ 的全部特征值为 $\lambda_1^2, \lambda_2^2, \dots, \lambda_n^2$, 它们均不小于 1. 再利用 A, B 可交换这个条件, 有

$$(A B^{-1})^2 = A B^{-1} A B^{-1} = A^2 (B^{-1})^2,$$

且 $(A B^{-1})^2$ 为埃尔米特矩阵, 所以 $A^2 (B^{-1})^2 \geq I$. 由此推出

$$B^H (A^2 (B^{-1})^2) B \geq B^H B,$$

即 $A^2 \geq B^2$.

证毕

性质 5 矩阵型的施瓦兹(Schwartz)不等式

设 A, B 分别为 $n \times m$ 和 $m \times l$ 矩阵, 且 $A A^H$ 非奇异(这时 $n \leq m$, $\text{rank}(A) = n$), 则

$$B^H B \geq (A B)^H (A A^H)^{-1} (A B), \quad (1.3.22)$$

式中等号成立,当且仅当存在一个 $n \times l$ 矩阵 C ,使得 $B=A^H C$.

证明之前先看特殊情形: $n=l=1$. 这时 $A=(a_1, a_2, \dots, a_m)$, $B^H=(\bar{b}_1, \bar{b}_2, \dots, \bar{b}_m)$, 不等式(1.3.22)成为

$$\sum_{k=1}^m |b_k|^2 \geq \left(\sum_{k=1}^m a_k b_k \right)^H \left(\sum_{k=1}^m |a_k|^2 \right)^{-1} \left(\sum_{k=1}^m a_k b_k \right)$$

或

$$\left| \sum_{k=1}^m a_k b_k \right|^2 \leq \left(\sum_{k=1}^m |a_k|^2 \right) \left(\sum_{k=1}^m |b_k|^2 \right),$$

这就是熟知的不等式.

现在来证式(1.3.22),它可由下列诸式得出:

$$\begin{aligned} 0 &\leq (B - A^H(AA^H)^{-1}AB)^H (B - A^H(AA^H)^{-1}AB) \\ &= B^H B - 2B^H A^H(AA^H)^{-1}AB + B^H A^H(AA^H)^{-1}AA^H(AA^H)^{-1}AB \\ &= B^H B - B^H A^H(AA^H)^{-1}AB \\ &= B^H B - (AB)^H(AA^H)^{-1}(AB), \end{aligned}$$

这就证明了式(1.3.22). 又上式右端为零矩阵当且仅当 $B=A^H(AA^H)^{-1}AB$. 此条件满足时, 令 $C=(AA^H)^{-1}AB$, 则 $B=A^H C$. 反之, 如果 $B=A^H C$, 则

$$A^H(AA^H)^{-1}AB = A^H(AA^H)^{-1}AA^H C = A^H C = B,$$

这就证明了式(1.3.22)中等号成立当且仅当 $B=A^H C$.

证毕

1.3.4 埃尔米特矩阵特征值的性质

对于埃尔米特矩阵,它的特征值有许多重要的性质,这些性质对今后研究埃尔米特矩阵特征值的稳定性问题有着重要的作用.

定理 1.3.9 设 A 是 n 阶埃尔米特矩阵,则

$$\lambda_{\min}(A)I \leq A \leq \lambda_{\max}(A)I, \quad (1.3.23)$$

其中 $\lambda_{\min}(A)$ 和 $\lambda_{\max}(A)$ 分别表示 A 的最小和最大特征值.

证明 由定理 1.3.2 知,存在酉矩阵 U ,使得

$$A = U \text{diag}(\lambda_1, \dots, \lambda_n) U^H,$$

其中 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ 是矩阵 A 的特征值. 故 $\lambda_{\max}(A) = \lambda_1$, $\lambda_{\max}(A) - \lambda_i \geq 0 (i=1, 2, \dots, n)$, $\lambda_{\min}(A) = \lambda_n$, $\lambda_i - \lambda_{\min}(A) \geq 0 (i=1, 2, \dots, n)$. 再由定理 1.3.3 知

$$\lambda_{\max}(A)I - A = U[\lambda_{\max}(A)I - \text{diag}(\lambda_1, \dots, \lambda_n)]U^H \geq 0,$$

$$A - \lambda_{\min}(A)I = U[\text{diag}(\lambda_1, \dots, \lambda_n) - \lambda_{\min}(A)I]U^H \geq 0,$$

即 $\lambda_{\max}(A)I \geq A, A \geq \lambda_{\min}(A)I$.

证毕

下面介绍瑞利(Rayleigh)商的概念,再运用瑞利商研究埃尔米特矩阵的特征值.

定义 1.3.5 设 A 为 n 阶埃尔米特矩阵,对 $\forall x \in \mathbb{C}^n$ 且 $x \neq 0$,称

$$R(x) = \frac{x^H A x}{x^H x}, \quad x \neq 0 \quad (1.3.24)$$

为埃尔米特矩阵 A 的瑞利商.

显然,埃尔米特矩阵的瑞利商是实数,并且具有如下性质.

定理 1.3.10 设 A 是 n 阶埃尔米特矩阵,其特征值为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, 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) $\lambda_1 = \max_{x \neq 0} R(x), \lambda_n = \min_{x \neq 0} R(x)$. (1.3.25)

证明 (1) 由瑞利商的定义即得.

(2) 由定理 1.3.9 即得.

(3) 设 A 对应于特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 的标准正交特征向量为 x_1, x_2, \dots, x_n . 显然, $R(x_1) = \lambda_1, R(x_n) = \lambda_n$. 因此由(2)可知

$$\max_{x \neq 0} R(x) = \lambda_1, \quad \min_{x \neq 0} R(x) = \lambda_n. \quad \text{证毕}$$

推广定理 1.3.10 可得到如下定理.

定理 1.3.11 设 A 是 n 阶埃尔米特矩阵,其特征值为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, 相应的标准正交特征向量为 x_1, x_2, \dots, x_n . 记 $V_i^{(j)} = \text{Span}\{x_i, x_{i+1}, \dots, x_j\} (i \leq j)$, 则

$$\lambda_i = \max_{\substack{x \in V_i^{(j)} \\ x \neq 0}} R(x), \quad \lambda_j = \min_{\substack{x \in V_i^{(j)} \\ x \neq 0}} R(x). \quad (1.3.26)$$

证明 对 $\forall x \in V_i^{(j)}$ 且 $x \neq 0$, 有 $x = \alpha_i x_i + \alpha_{i+1} x_{i+1} + \dots + \alpha_j x_j$, 从而

$$\lambda_j \leq R(x) = \frac{|\alpha_i|^2 \lambda_i + |\alpha_{i+1}|^2 \lambda_{i+1} + \dots + |\alpha_j|^2 \lambda_j}{|\alpha_i|^2 + |\alpha_{i+1}|^2 + \dots + |\alpha_j|^2} \leq \lambda_i.$$

特别地,若 $x = x_i$, 则 $R(x) = \lambda_i$; 而 $x = x_j$ 时, 有 $R(x) = \lambda_j$, 因此

$$\lambda_i = \max_{\substack{x \in V_i^{(j)} \\ x \neq 0}} R(x), \quad \lambda_j = \min_{\substack{x \in V_i^{(j)} \\ x \neq 0}} R(x). \quad \text{证毕}$$

一般地,有如下著名的极大极小或极小极大定理.

定理 1.3.12 设 A 是 n 阶埃尔米特矩阵,其特征值为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, V_i 是 \mathbb{C}^n 中 i 维子空间, 则

$$\lambda_i = \max_{V_i} \min_{\substack{x \in V_i \\ x \neq 0}} R(x) \quad (1.3.27)$$

和

$$\lambda_i = \min_{V_{n-i+1}} \max_{\substack{x \in V_{n-i+1} \\ x \neq 0}} R(x), \quad (1.3.28)$$

其中 $i=1, 2, \dots, n$.

证明 这里只要证明式(1.3.27), 而式(1.3.28)的证明是类似的.

令 D_i 是以子空间 V_i 的一组标准正交基为列向量组成的 $n \times i$ 阶矩阵, 则

$$D_i^H D_i = I_i.$$

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 1.1.6 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

$$R(\tilde{x}) = \frac{|\alpha_i|^2 \lambda_i + |\alpha_{i+1}|^2 \lambda_{i+1} + \cdots + |\alpha_n|^2 \lambda_n}{|\alpha_i|^2 + |\alpha_{i+1}|^2 + \cdots + |\alpha_n|^2} \leq \lambda_i.$$

又因为 $\tilde{x} \in V_i$, 所以对任取的 i 维子空间 V_i 有

$$\min_{\substack{x \in V_i \\ x \neq 0}} R(x) \leq \lambda_i.$$

另一方面, 若取 $D_i = U_i$, 则 $\min_{\substack{x \in V_i \\ x \neq 0}} R(x) \geq \lambda_i$. 因此 $\lambda_i = \max_{V_i} \min_{\substack{x \in V_i \\ x \neq 0}} R(x)$. 证毕

最后, 我们应用极大极小定理 1.3.12 研究埃尔米特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 1.3.13 设 A, E 均为 n 阶埃尔米特矩阵, $B = A + E$, 且 A, B 和 E 的特征值分别为 $\lambda_1 \geq \cdots \geq \lambda_n, \mu_1 \geq \cdots \geq \mu_n$ 和 $\epsilon_1 \geq \cdots \geq \epsilon_n$, 则

$$\lambda_i + \epsilon_n \leq \mu_i \leq \lambda_i + \epsilon_1, \quad i = 1, 2, \cdots, n. \quad (1.3.29)$$

证明 设 x_1, \cdots, x_n 是 A 的对应于特征值 $\lambda_1, \cdots, \lambda_n$ 的标准正交特征向量, 令

$$V_{n-i+1} = \text{Span}\{x_i, x_{i+1}, \cdots, x_n\},$$

则由式(1.3.28)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 1.3.11 和定理 1.3.12 有

$$\mu_i \leq \lambda_i + \max_{\substack{x \in V_{n-i+1} \\ x \neq 0}} \frac{x^H E x}{x^H x} \leq \lambda_i + \epsilon_1, \quad i = 1, 2, \cdots, n. \quad (1.3.30)$$

另一方面, 令 $D = -E$, 记 D 的特征值为 $\delta_1 \geq \cdots \geq \delta_n$, 则 $A = B + D$. 由式(1.3.30)有

$$\lambda_i \leq \mu_i + \delta_1, \quad i = 1, 2, \cdots, n,$$

其中 $\delta_1 = -\epsilon_n$. 因此式(1.3.29)成立.

证毕

1.3.5 一般的复正定矩阵

定义 1.3.6 设 A 是 n 阶复矩阵, 如果对任意 n 维(复)列向量 x 都有

$$\text{Re}(x^H A x) \geq 0, \quad (1.3.31)$$

其中 $\text{Re}(z)$ 表示复数 z 的实部, 则称 A 为非负定(半正定)矩阵. 如果对任意 n 维非零(复)列向量 x , 式(1.3.31)中严格不等式成立, 则称 A 为正定矩阵.

容易看出, 当 A 为埃尔米特矩阵时, 式(1.3.31)成为 $x^H A x \geq 0$, 这时定义 1.3.6 和埃尔米特矩阵正定的定义是一致的. 因此, 定义 1.3.6 是埃尔米特矩阵正定的定义的拓广.

复正定矩阵有下列基本性质.

(1) 正定(非负定)矩阵 A 的所有特征值均有正实部(非负实部).

(2) 设 A 为 n 阶正定矩阵, P 为 n 阶非奇异矩阵, 则 $P^H A P$ 亦为正定矩阵.

(3) 设 A 为 n 阶非负定矩阵, C 为 $n \times m$ 矩阵, 则 $C^H A C$ 亦为非负定矩阵.

(4) 正定(非负定)矩阵 A 的所有主子矩阵均为正定(非负定)矩阵. 特别地, 正定矩阵 A 的对角元均有正(非负)实部.

下面, 我们来看复矩阵 A 为正定(非负定)矩阵的条件.

定理 1.3.14 n 阶复矩阵 A 为正定(非负定)矩阵的充分必要条件是 $A + A^H$ 为埃尔米特正定矩阵(非负定矩阵).

证明 由于

$$\mathbf{x}^H(\mathbf{A} + \mathbf{A}^H)\mathbf{x} = 2\operatorname{Re}(\mathbf{x}^H\mathbf{A}\mathbf{x}),$$

故定理成立. 证毕

定理 1.3.15 n 阶复矩阵 \mathbf{A} 为正定矩阵的充分必要条件是: 存在非奇异矩阵 \mathbf{P} , 使得

$$\mathbf{P}^H\mathbf{A}\mathbf{P} = \operatorname{diag}(1 + ia_1, 1 + ia_2, \dots, 1 + ia_n), \quad (1.3.32)$$

其中 $i = \sqrt{-1}$, $a_j (j=1, \dots, n)$ 均为实数.

证明 必要性: \mathbf{A} 可表示为

$$\mathbf{A} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^H) + \frac{1}{2}(\mathbf{A} - \mathbf{A}^H) = \mathbf{A}_H + \mathbf{A}_C, \quad (1.3.33)$$

其中 $\mathbf{A}_H = \frac{1}{2}(\mathbf{A} + \mathbf{A}^H)$ 为埃尔米特矩阵, $\mathbf{A}_C = \frac{1}{2}(\mathbf{A} - \mathbf{A}^H)$ 为反埃尔米特矩阵. 当 \mathbf{A} 为正定矩阵时, \mathbf{A}_H 为埃尔米特正定矩阵. 于是存在非奇异矩阵 \mathbf{P}_1 , 使得

$$\mathbf{P}_1^H\mathbf{A}_H\mathbf{P}_1 = \mathbf{I}, \quad \mathbf{P}_1^H\mathbf{A}_C\mathbf{P}_1 \equiv \mathbf{C}, \quad (1.3.34)$$

其中 \mathbf{C} 为反埃尔米特矩阵. 故存在酉矩阵 \mathbf{U} , 使得

$$\mathbf{U}^H\mathbf{C}\mathbf{U} = \operatorname{diag}(ia_1, ia_2, \dots, ia_n), \quad (1.3.35)$$

其中 $a_j (j=1, \dots, n)$ 均为实数. 以 \mathbf{U}^H 和 \mathbf{U} 分别左乘、右乘(1.3.34)两式的两端, 并记 $\mathbf{P}_1\mathbf{U}$ 为 \mathbf{P} , 则 \mathbf{P} 非奇异, 且有

$$\begin{aligned} \mathbf{P}^H\mathbf{A}_H\mathbf{P} &= \mathbf{I}, \\ \mathbf{P}^H\mathbf{A}_C\mathbf{P} &= \operatorname{diag}(ia_1, ia_2, \dots, ia_n), \end{aligned} \quad (1.3.36)$$

将式(1.3.36)中两式的两端分别相加, 即得式(1.3.32).

充分性: 对任意的 n 维非零列向量 \mathbf{x} , 有

$$\operatorname{Re}\{\mathbf{x}^H \operatorname{diag}(1 + ia_1, 1 + ia_2, \dots, 1 + ia_n)\mathbf{x}\} = \mathbf{x}^H\mathbf{x} > 0,$$

即 $\operatorname{diag}(1 + ia_1, 1 + ia_2, \dots, 1 + ia_n)$ 为正定矩阵. 若式(1.3.32)成立, 则由基本性质(2)可知, \mathbf{A} 为正定矩阵. 证毕

定义 1.3.7 设 \mathbf{A} 为 n 阶复矩阵. 如果 \mathbf{A} 的特征值均有负实部, 则称 \mathbf{A} 为稳定矩阵.

由基本性质(1)可知, 若 \mathbf{A} 为正定矩阵, 则 $-\mathbf{A}$ 为稳定矩阵. 反过来, 若 $-\mathbf{A}$ 为稳定矩阵, 则 \mathbf{A} 未必为正定矩阵.

习 题 1 (5)

1. 设 \mathbf{A}, \mathbf{B} 均为埃尔米特矩阵, 证明: \mathbf{AB} 为埃尔米特矩阵的充分必要条件是 $\mathbf{AB} = \mathbf{BA}$.
2. 证明: 任一复方阵都可以表示成埃尔米特矩阵与反埃尔米特矩阵之和.
3. 证明: 欧氏空间 V^n 的线性变换 \mathcal{A} 为反对称变换, 即

$$(\mathcal{A}(\boldsymbol{\alpha}), \boldsymbol{\beta}) = -(\boldsymbol{\alpha}, \mathcal{A}(\boldsymbol{\beta})), \quad \boldsymbol{\alpha}, \boldsymbol{\beta} \in V^n$$

的充分必要条件是 \mathcal{A} 在 V^n 的标准正交基下的矩阵 \mathbf{A} 为反对称矩阵(即 $\mathbf{A}^T = -\mathbf{A}$).

4. 证明:

- (1) 若 \mathbf{A} 是对称矩阵, 则 \mathbf{A}^{-1} 与 \mathbf{A}^n 也是对称矩阵.
- (2) 若 \mathbf{A} 是反对称矩阵, 则 \mathbf{A}^{-1} 也是反对称矩阵, 但不存在奇数阶可逆反对称矩阵.

5. 证明:

- (1) 与对称矩阵相合的矩阵仍为对称矩阵.

(2) 若方阵 A 与 B 相合, 则 $\text{rank}(A) = \text{rank}(B)$, 即在相合变换下秩是一个不变量.

6. 求证: 实反对称矩阵的特征值是零或纯虚数.

7. 设 A 是 n 阶实对称矩阵, 且 $A^2 = A$ (即 A 是幂等矩阵). 证明: 存在正交矩阵 Q , 使得

$$Q^{-1}AQ = \text{diag}(1, \dots, 1, 0, \dots, 0).$$

8. 证明: 酉空间 V^n 的埃尔米特变换 \mathcal{A} 在 V^n 的某组基下的矩阵为对角矩阵.

9. (1) 设 A 为半正定矩阵. 求证: $|A+I| \geq 1$, 且等号成立的充分必要条件是 $A=0$;

(2) 设 A 为半正定矩阵, B 是正定矩阵. 求证: $|A+B| \geq |B|$, 且等号成立的充分必要条件是 $A=0$.

10. 证明: 在 n 阶对称矩阵中, 正定矩阵只能与正定矩阵相似.

11. 求证: (1) 如果 A, B 都是 n 阶正定矩阵, 则 $A+B$ 也是正定矩阵.

(2) 设 A 为对称正定矩阵, M 为非奇异矩阵, 则与 A 相合的矩阵 M^TAM 以及 A^{-1} 均正定.

12. 求证: 对任何实矩阵 $A_{m \times n}$, 所得的矩阵 $B = A^T A$ 为半正定, 且当 A 的列向量组线性无关时, B 为正定矩阵.

13. 求证: (1) A 为实半正定矩阵的充分必要条件是 A 的主子式全大于等于零.

(2) 实对称矩阵 A 是负定的 (即对应的二次型 $x^T Ax < 0$) 充分必要条件是: A 的一切偶数阶主子式都大于零、一切奇数阶主子式都小于零.

14. 设 A 为实对称矩阵. 证明: 对任意奇数 m , 必有实对称矩阵 B , 使 $B^m = A$; 当 A 为半正定矩阵时, 对任意正整数 m , 有实对称矩阵 B , 使 $B^m = A$.

15. 设 A 是实对称正定矩阵, m 是任一正整数. 证明: 存在惟一的实对称正定矩阵 B , 使 $B^m = A$.

16. 证明: 每个非奇异实矩阵 A 必可表示为一个正定矩阵与一个正交矩阵的乘积, 即

$$A = B_1 Q_1 = Q_2 B_2,$$

其中 B_1, B_2 是正定矩阵, Q_1, Q_2 是正交矩阵, 且每种表示是惟一的.

17. 设 A, B 是两个 n 阶实对称矩阵, 且 B 为正定矩阵. 证明: 存在 n 阶实非奇异矩阵 P , 使 P^TAP 与 P^TBP 同时为对角矩阵.

18. 设 A 与 B 是两个实正定矩阵, 证明:

$$|A| + |B| \leq |A+B|.$$

19. 求正交矩阵 Q , 使 $Q^T A Q$ 为对角矩阵, 已知

$$(1) A = \begin{bmatrix} 2 & -2 & 0 \\ -2 & 1 & -2 \\ 0 & -2 & 0 \end{bmatrix}; \quad (2) A = \begin{bmatrix} 1 & 1 & 0 & -1 \\ 1 & 1 & -1 & 0 \\ 0 & -1 & 1 & 1 \\ -1 & 0 & 1 & 1 \end{bmatrix}.$$

20. 若 S, T 分别是实对称和反对称矩阵, 且 $\det(I - T - iS) \neq 0$, 试证: $(I + T + iS)(I - T - iS)^{-1}$ 是酉矩阵.

21. 设 A, B 均是实对称矩阵, 试证: A 与 B 正交相似的充要条件是 A 与 B 的特征值相同.

22. 设 $A^H = -A$, 试证: $U = (A + I)(A - I)^{-1}$ 是酉矩阵.

23. 设实对称矩阵

$$A = \begin{bmatrix} a & 1 & 1 \\ 1 & a & -1 \\ 1 & -1 & a \end{bmatrix}$$

求可逆矩阵 P , 使得 $P^{-1}AP$ 为对角阵, 并计算 $|A - 4I|$ 的值.

24. (1) 证明: 如果整数 a, b 均可以表示成四个整数的平方和, 那么 ab 也可以表示成四个整数的平方和;

(2) 将 1457 表示成四个整数的平方和.

25. 设 A 是埃尔米特矩阵, 如果对任意向量 x 均有 $x^H Ax = 0$, 则 $A = 0$.

26. 证明:

(1) 埃尔米特矩阵的特征值均为实数, 且属于不同特征值的特征向量彼此正交;

- (2) 反埃尔米特矩阵的特征值为零或纯虚数;
 (3) 酉矩阵特征值的模长为 1.

27. 设 n 阶酉矩阵 U 的特征值不等于 -1 , 试证: 矩阵 $I+U$ 满秩, 而且 $W=i(I-U)(I+U)^{-1}$ 是埃尔米特矩阵; 反之, 如果 W 是埃尔米特矩阵, 那么 $V=(I+iW)(I-iW)^{-1}$ 是酉矩阵.

28. 设 A 为正定埃尔米特矩阵, B 为反埃尔米特矩阵, 试证: AB 与 BA 的特征值实部为零.

29. 设 A, B 均是埃尔米特矩阵, 且 A 正定, 试证: AB 与 BA 的特征值都是实数.

30. 设 A 是半正定埃尔米特矩阵, $A \neq 0, B$ 是正定埃尔米特矩阵, 试证: $|A+B| > |B|$.

31. 设 A 是正定埃尔米特矩阵, B 是反埃尔米特矩阵, 试证: $A+B$ 为可逆矩阵.

32. 设 A, B 均是埃尔米特矩阵, 试证: A 与 B 酉相似的充要条件是 A 与 B 的特征值相同.

33. 设 A 是埃尔米特矩阵, 且 $A^2=I$, 则存在酉矩阵 U , 使得

$$U^H A U = \begin{bmatrix} I_r & 0 \\ 0 & -I_{n-r} \end{bmatrix}.$$

34. 设 A 是正定埃尔米特矩阵, 且 $A \in U^{n \times n}$ (酉矩阵空间), 则 $A=I$.

35. 试证: (1) 两个半正定埃尔米特矩阵之和是半正定的;

(2) 半正定埃尔米特矩阵与正定埃尔米特矩阵之和是正定的.

36. 设 $A^H=A$, 试证: 总存在 $t>0$, 使得 $A+tI$ 是正定埃尔米特矩阵, $A-tI$ 是负定埃尔米特矩阵.

37. 设 $A \in \mathbb{C}^{m \times n}$, 试证:

(1) $A^H A$ 和 AA^H 都是半正定的埃尔米特矩阵;

(2) $A^H A$ 和 AA^H 的非零特征值相同.

38. 设 $A \in \mathbb{C}^{n \times n}$, 那么 A 可以惟一的写成 $A=S+iT$, 其中 S, T 为埃尔米特矩阵, 且 A 可以惟一的写成 $A=B+C$, 其中 B 是埃尔米特矩阵, C 是反埃尔米特矩阵.

λ 矩阵与若尔当标准形

引言 什么是矩阵标准形

矩阵的标准形问题不仅在矩阵理论和矩阵计算中有着重要地位,而且在力学、控制理论等学科中也有着广泛的应用.通常所涉及的矩阵标准形有两种:

(1) 对角矩阵: $\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$, 其中主对角线上的元素均为常数,其余元素均为零;

(2) 若尔当标准形: $\begin{pmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_m \end{pmatrix}$, 其中

$$\mathbf{A}_i = \begin{pmatrix} \lambda_i & 1 & \cdots & 0 \\ & \lambda_i & \ddots & \vdots \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}, \quad i = 1, \cdots, m$$

叫做若尔当块.

由 1.1.2 节中第 8 段知,一个 n 阶矩阵 \mathbf{A} 如果具有 n 个线性无关的特征向量(称矩阵 \mathbf{A} 有完备的特征向量系,或称 \mathbf{A} 是非亏损矩阵),则 \mathbf{A} 必相似于对角矩阵,即为第 1 种标准形问题;如果 \mathbf{A} 的线性无关的特征向量的个数小于 n (称 \mathbf{A} 为亏损矩阵),则一定不能和对角矩阵相似.那么,后一情形 \mathbf{A} 能与何种具有“标准”形状的矩阵相似呢?这个问题就是本章要讨论的矩阵在相似下的若尔当标准形问题,即第 2 种标准形.我们将采用 λ 矩阵的理论导出矩阵的这种标准形,并利用矩阵的若尔当标准形研究线性变换的结构.

2.1 λ 矩阵

2.1.1 λ 矩阵的概念

在矩阵理论中,我们把矩阵定义为数的阵列,即它的元素是数域中的数.此类统称数字矩阵.现在我们把数字矩阵加以推广,引进 λ 矩阵.

定义 2.1.1 设 $a_{ij}(\lambda)$ ($i=1,2,\cdots,m; j=1,2,\cdots,n$) 为数域 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 λ 矩阵或多项式矩阵. 这样矩阵的全体记为 $P[\lambda]^{m \times n}$.

注意, 这里的 λ 是一个未定元, 而不是一个具体的数. 当 λ 取值为一个具体的数时(如 $\lambda=0$), 它就不是一个 λ 矩阵而是一个数字矩阵了.

显然数字矩阵和特征矩阵 $\lambda \mathbf{I} - \mathbf{A}$ 均为 λ 矩阵的特例.

如同数字矩阵一样, 对 λ 矩阵也可以进行加法和乘法运算.

由于 λ 矩阵的元素都是 λ 的多项式, 故 λ 矩阵的秩的定义以及初等变换都与数字矩阵的情况有些不同, 现叙述如下:

定义 2.1.2 若 λ 矩阵 $\mathbf{A}(\lambda)$ 至少有一个 r ($r \geq 1$) 阶子式不是零多项式, 而一切 $r+1$ 阶子式(如果有的话)都是零多项式, 则称 $\mathbf{A}(\lambda)$ 的秩为 r , 记为 $\text{rank} \mathbf{A}(\lambda)$.

零矩阵的秩定义为零. 若 n 阶 λ 矩阵的秩为 n , 则称 $\mathbf{A}(\lambda)$ 为满秩的或非奇异的.

例 2.1.1 n 阶数字矩阵 \mathbf{A} 的特征矩阵为 $\lambda \mathbf{I} - \mathbf{A}$, 显然 $|\lambda \mathbf{I} - \mathbf{A}| = \lambda^n + c_1 \lambda^{n-1} + \cdots + c_{n-1} \lambda + c_n$ (其中 $c_i \in \mathbf{P}, i=1, 2, \dots, n$) 是 λ 的多项式(不是零多项式), 因此 λ 矩阵 $\lambda \mathbf{I} - \mathbf{A}$ 的秩为 n , 且 $\lambda \mathbf{I} - \mathbf{A}$ 是满秩的(注意, 当 λ_0 为 \mathbf{A} 的特征值时, $\lambda_0 \mathbf{I} - \mathbf{A}$ 为一个数字矩阵, 此时其秩未必等于 n , 即数字矩阵 $\lambda_0 \mathbf{I} - \mathbf{A}$ 的秩小于 n).

定义 2.1.3 设 $\mathbf{A}(\lambda)$ 为一个 n 阶 λ 矩阵, 若存在 n 阶 λ 矩阵 $\mathbf{B}(\lambda)$ 使

$$\mathbf{A}(\lambda)\mathbf{B}(\lambda) = \mathbf{B}(\lambda)\mathbf{A}(\lambda) = \mathbf{I}, \quad (2.1.1)$$

则称 $\mathbf{A}(\lambda)$ 为可逆的, $\mathbf{B}(\lambda)$ 称为 $\mathbf{A}(\lambda)$ 的逆矩阵, 记为 $\mathbf{A}(\lambda)^{-1}$.

对数字矩阵, 满秩与可逆是两个等价的概念, 但对于 λ 矩阵, 这一结论不成立. 当然, 可逆的 λ 矩阵一定是满秩的, 但满秩的 λ 矩阵不一定可逆. 如例 2.1.1 中的特征矩阵 $\lambda \mathbf{I} - \mathbf{A}$ 是满秩的, 但不可逆. 事实上, 我们有如下定理.

定理 2.1.1 n 阶 λ 矩阵 $\mathbf{A}(\lambda)$ 可逆的充分必要条件是 $\mathbf{A}(\lambda)$ 的行列式为一个非零常数.

证明 若 λ 矩阵 $\mathbf{A}(\lambda)$ 可逆, 由定义, 存在 λ 矩阵 $\mathbf{B}(\lambda)$ 使式(2.1.1)成立, 两边取行列式有

$$|\mathbf{A}(\lambda)| |\mathbf{B}(\lambda)| = 1.$$

由于 $|\mathbf{A}(\lambda)|, |\mathbf{B}(\lambda)|$ 均为 λ 的多项式, 所以 $|\mathbf{A}(\lambda)|, |\mathbf{B}(\lambda)|$ 均为非零常数.

反之, 设 $|\mathbf{A}(\lambda)| = c \neq 0$, 则

$$\left(\frac{1}{c} \text{adj} \mathbf{A}(\lambda)\right) \cdot \mathbf{A}(\lambda) = \mathbf{A}(\lambda) \cdot \left(\frac{1}{c} \text{adj} \mathbf{A}(\lambda)\right) = \mathbf{I},$$

因而 $\mathbf{A}(\lambda)$ 是可逆的. 这里 $\text{adj} \mathbf{A}(\lambda)$ 是 $\mathbf{A}(\lambda)$ 的伴随矩阵.

证毕

定义 2.1.4 下列各种类型的变换, 叫做 λ 矩阵的初等变换:

- (1) 对调两行(列);
- (2) 用一个不为零的数 k 乘任一行(列)的所有元素;
- (3) 用 λ 的多项式 $\varphi(\lambda)$ 乘某行(列)的所有元素, 并加到另一行(列)的对应元素上去.

上述 3 种初等变换有相应的 3 种初等矩阵. 其实, 只要对单位矩阵 \mathbf{I} 施行上述 3 种初等

$$B(\lambda) = P_r P_{r-1} \cdots P_2 P_1 A(\lambda) Q_1 Q_2 \cdots Q_r.$$

令 $P(\lambda) = P_r P_{r-1} \cdots P_2 P_1$, $Q(\lambda) = Q_1 Q_2 \cdots Q_r$, 由于初等矩阵可逆, 所以 $P(\lambda)$, $Q(\lambda)$ 均可逆, 于是有

$$B(\lambda) = P(\lambda)A(\lambda)Q(\lambda). \quad \text{证毕}$$

由此定理可知, 若两个 λ 矩阵相抵, 则它们的秩必相等. 但此命题的逆命题却不成立, 这与数字矩阵不同, 具有相同秩的两个 λ 矩阵未必相抵. 例如

$$A(\lambda) = \begin{bmatrix} \lambda & 2 \\ 0 & \lambda \end{bmatrix}, \quad B(\lambda) = \begin{bmatrix} \lambda & -2 \\ \lambda & 2 \end{bmatrix}.$$

因为 $|A(\lambda)| = \lambda^2$, $|B(\lambda)| = 4\lambda$, 所以当 $\lambda \neq 0$ 时, $A(\lambda)$ 与 $B(\lambda)$ 的秩均为 2. 又由于初等变换是可逆的, 则由定理 2.1.2 的证明过程知, 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, 秩相等只是 λ 矩阵相抵的必要条件, 而并不是 λ 矩阵相抵的充分条件.

那么两个 λ 矩阵相抵的充分必要条件是什么呢? 这个问题将在 2.1.3 节中回答.

下面研究如何将 λ 矩阵化为标准形.

2.1.2 λ 矩阵在相抵下的标准形

这一段主要证明任何一个 n 阶 λ 矩阵经有限次初等变换, 总可以化为对角形, 即标准形.

引理 设 λ 矩阵的左上角元素 $a_{11}(\lambda) \neq 0$, $A(\lambda)$ 中至少有一个元素不能被它整除, 那么一定可找到与 $A(\lambda)$ 相抵的矩阵 $B(\lambda)$, 它的左上角元素也不为零, 但次数比 $a_{11}(\lambda)$ 的次数低.

证明 分三种情况讨论.

(1) 若在 $A(\lambda)$ 的第 1 列中有元素 $a_{i1}(\lambda)$ 不能被 $a_{11}(\lambda)$ 整除, 即

$$a_{i1}(\lambda) = a_{11}(\lambda)q(\lambda) + r(\lambda),$$

其中 $r(\lambda) \neq 0$, 且次数比 a_{11} 低.

此时, 对 $A(\lambda)$ 作两次初等行变换: 首先, 第 1 行乘以 $-q(\lambda)$ 加到第 i 行, 则第 i 行第 1 列的元素成为 $r(\lambda)$; 其次再对调第一行和第 i 行, 得到新的 λ 矩阵 $B(\lambda)$, 则 $B(\lambda)$ 左上角元素为 $r(\lambda)$, 故为引理所求矩阵.

(2) 在 $A(\lambda)$ 的第 1 行中有一元素 $a_{1i}(\lambda)$ 不能被 $a_{11}(\lambda)$ 整除, 这种情况与情况(1)类似.

(3) $A(\lambda)$ 中第 1 行与第 1 列的元素都可被 $a_{11}(\lambda)$ 整除, 但 $A(\lambda)$ 中的 $a_{ij}(\lambda)$ ($i > 1, j > 1$) 不能被 $a_{11}(\lambda)$ 整除.

此时, 我们设 $a_{11}(\lambda) = a_{11}(\lambda)\varphi(\lambda)$, 对 $A(\lambda)$ 作两次初等行变换: 首先, 第 1 行乘以 $-\varphi(\lambda)$ 加到第 i 行, 则第 i 行第 1 列的元素变为 0, 第 i 行第 j 列的元素变为 $a_{ij}(\lambda) - a_{1j}(\lambda)\varphi(\lambda)$; 其次把第 i 行的元素加到第 1 行, 则第 1 行第 1 列的元素仍为 $a_{11}(\lambda)$, 而第 1 行第 j 列的元素变为 $a_{1j}(\lambda) + [1 - \varphi(\lambda)]a_{1j}(\lambda)$, 它不能被 $a_{11}(\lambda)$ 所整除, 这就化为已经证明了的情况(2).

证毕

定理 2.1.3 设 $A(\lambda) \in P[\lambda]^{m \times n}$ 且 $\text{rank}(A(\lambda)) = r$, 则 $A(\lambda)$ 相抵于如下对角形, 称为 $A(\lambda)$ 的史密斯(Smith)标准形.

为史密斯标准形.

解 记 r_i 表示第 i 行, c_i 表示第 i 列, 计算过程如下:

$$\begin{aligned} \mathbf{A}(\lambda) &\stackrel{c_3+c_1}{\cong} \begin{pmatrix} -\lambda+1 & 2\lambda-1 & 1 \\ \lambda & \lambda^2 & 0 \\ \lambda^2+1 & \lambda^2+\lambda-1 & 1 \end{pmatrix} \stackrel{c_1+c_3}{\cong} \begin{pmatrix} 1 & 2\lambda-1 & -\lambda+1 \\ 0 & \lambda^2 & \lambda \\ 1 & \lambda^2+\lambda-1 & \lambda^2+1 \end{pmatrix} \\ &\stackrel{r_3-r_1}{\cong} \begin{pmatrix} 1 & 2\lambda-1 & -\lambda+1 \\ 0 & \lambda^2 & \lambda \\ 0 & \lambda^2-\lambda & \lambda^2+\lambda \end{pmatrix} \stackrel{\substack{c_2-(2\lambda-1)c_1 \\ c_3+(\lambda-1)c_1}}{\cong} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda^2 & \lambda \\ 0 & \lambda^2-\lambda & \lambda^2+\lambda \end{pmatrix} \\ &\stackrel{c_2+c_3}{\cong} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & \lambda^2 \\ 0 & \lambda^2+\lambda & \lambda^2-\lambda \end{pmatrix} \stackrel{c_3-\lambda c_2}{\cong} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & \lambda^2+\lambda & -\lambda^3-\lambda \end{pmatrix} \\ &\stackrel{\substack{(-1)c_3 \\ r_3-(\lambda+1)r_2}}{\cong} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda^3+\lambda \end{pmatrix}. \end{aligned}$$

此时, $d_1(\lambda)=1, d_2(\lambda)=\lambda, d_3(\lambda)=\lambda^3+\lambda$.

2.1.3 不变因子与初等因子

在 2.1.1 节中已经指出, 秩相等是两个 λ 矩阵相抵的必要条件, 此外还需具备什么性质才能使两个 λ 矩阵相抵呢? 又如何将矩阵化成若尔当标准形? 对这些问题的讨论, 不变因子和初等因子起着重要作用.

定义 2.1.6 λ 矩阵 $\mathbf{A}(\lambda)$ 最后化成的史密斯标准形(2.1.3), 其对角线的元素 $d_1(\lambda), d_2(\lambda), \dots, d_r(\lambda)$ 称为 $\mathbf{A}(\lambda)$ 的不变因子.

定义 2.1.7 设 $\mathbf{A}(\lambda)$ 的秩为 r , 对于正整数 $k(1 \leq k \leq r)$, $\mathbf{A}(\lambda)$ 中必有非零的 k 阶子式, 把 $\mathbf{A}(\lambda)$ 中全部 k 阶子式的最大公因式称为 $\mathbf{A}(\lambda)$ 的 k 阶行列式因子, 记为 $D_k(\lambda)$.

例 2.1.3 求

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

的各阶行列式因子.

解 由于两个一阶子式 $-\lambda+1$ 与 λ 的最大公因式为 1, 所以 $D_1(\lambda)=1$. 又考察两个二阶子式:

$$\begin{aligned} \begin{vmatrix} -\lambda+1 & \lambda^2 \\ \lambda & \lambda \end{vmatrix} &= \lambda(-\lambda^2-\lambda+1) = \varphi_1(\lambda), \\ \begin{vmatrix} -\lambda+1 & \lambda^2 \\ \lambda^2+1 & \lambda^2 \end{vmatrix} &= \lambda^3(-\lambda-1) = \varphi_2(\lambda), \end{aligned}$$

故 $\varphi_1(\lambda)$ 与 $\varphi_2(\lambda)$ 的最大公因式为 λ , 其余的二阶子式(还有 7 个)都包含因子 λ , 所以 $D_2(\lambda)=\lambda$.

于是有

$$\begin{cases} D_1(\lambda) \mid D_2(\lambda), D_2(\lambda) \mid D_3(\lambda), \dots, D_{r-1}(\lambda) \mid D_r(\lambda). \\ d_1(\lambda) = D_1(\lambda), d_2(\lambda) = D_2(\lambda)/D_1(\lambda), \dots, d_r(\lambda) = D_r(\lambda)/D_{r-1}(\lambda). \end{cases} \quad (2.1.5)$$

从而得到如下结论.

定理 2.1.5 λ 矩阵 $\mathbf{A}(\lambda)$ 的史密斯标准形是惟一的.

证明 因为 $\mathbf{A}(\lambda)$ 的各阶行列式因子是惟一的, 则由式(2.1.5)知 $\mathbf{A}(\lambda)$ 的不变因子也是惟一的. 因此 $\mathbf{A}(\lambda)$ 的史密斯标准形是惟一的. 证毕

应用 λ 矩阵的史密斯标准形, 可以证明如下定理.

定理 2.1.6 设 $\mathbf{A}(\lambda), \mathbf{B}(\lambda) \in P[\lambda]^{m \times n}$, 则 $\mathbf{A}(\lambda)$ 与 $\mathbf{B}(\lambda)$ 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 2.1.4 即得.

充分性 若 λ 矩阵 $\mathbf{A}(\lambda)$ 与 $\mathbf{B}(\lambda)$ 有相同的不变因子, 则 $\mathbf{A}(\lambda)$ 与 $\mathbf{B}(\lambda)$ 和同一史密斯标准形相抵, 因而 $\mathbf{A}(\lambda)$ 与 $\mathbf{B}(\lambda)$ 相抵. 证毕

一般来说, 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 2.1.4 求

$$\mathbf{A}(\lambda) = \begin{pmatrix} \lambda - a & -1 & \cdots & 0 \\ 0 & \lambda - a & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & \lambda - a \end{pmatrix}_{m \times m}$$

的行列式因子和不变因子.

解 由于 $\mathbf{A}(\lambda)$ 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 $D_{m-1}(\lambda) = 1$. 由式(2.1.5)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 $\mathbf{A}(\lambda)$ 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 $\mathbf{A}(\lambda)$ 的标准形为

$$\mathbf{A}(\lambda) \simeq \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & (\lambda - a)^m \end{pmatrix}_{m \times m}.$$

上述定理表明, λ 矩阵的不变因子是否相同是两个 λ 矩阵是否等价的本质特征. 其实,

不变因子还可以进一步分解,引出所谓初等因子的概念.

由于每个不变因子 $d_i(\lambda)$ 都是首项系数为 1 的 λ 的多项式,所以当 $\mathbf{A}(\lambda)$ 的秩为 r 时, r 个不恒等于零的 $d_i(\lambda)$ 在复数域内总可以分解为一次因式的幂积,即

$$\begin{cases} d_1(\lambda) = (\lambda - \lambda_1)^{e_{11}} (\lambda - \lambda_2)^{e_{12}} \cdots (\lambda - \lambda_s)^{e_{1s}}, \\ d_2(\lambda) = (\lambda - \lambda_1)^{e_{21}} (\lambda - \lambda_2)^{e_{22}} \cdots (\lambda - \lambda_s)^{e_{2s}}, \\ \vdots \\ d_r(\lambda) = (\lambda - \lambda_1)^{e_{r1}} (\lambda - \lambda_2)^{e_{r2}} \cdots (\lambda - \lambda_s)^{e_{rs}}, \end{cases} \quad (2.1.6)$$

其中 $\lambda_1, \lambda_2, \dots, \lambda_s$ 是 $d_r(\lambda)$ 中一切相异的根,且有些可能为复数.另外,由于 $d_i(\lambda)$ 可以被 $d_{i-1}(\lambda)$ 整除,所以有关系式

$$0 \leq e_{ij} \leq e_{2j} \leq \cdots \leq e_{rj}, \quad j = 1, 2, \dots, s,$$

也就是说, e_{ij} ($1 \leq i < r, 1 \leq j \leq s$) 可能为零,但 $e_{r1}, e_{r2}, \dots, e_{rs}$ 无一为零.

定义 2.1.8 当 $e_{ij} \neq 0$ 时,因子 $(\lambda - \lambda_j)^{e_{ij}}$ ($i = 1, 2, \dots, r; j = 1, 2, \dots, s$) 的全体叫做 $\mathbf{A}(\lambda)$ 的初等因子.

例如,若有某 λ 矩阵 $\mathbf{A}(\lambda)$ 通过一系列初等变换得到它的不变因子为

$$1, 1, (\lambda - 2)^5 (\lambda - 3)^3, (\lambda - 5)^5 (\lambda - 3)^4 (\lambda + 2),$$

则它的初等因子是

$$(\lambda - 2)^5, (\lambda - 3)^3, (\lambda - 5)^5, (\lambda - 3)^4, (\lambda + 2).$$

例 2.1.5 求矩阵 $\mathbf{A} = \begin{pmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ 1 & 0 & 2 \end{pmatrix}$ 的特征矩阵 $\lambda \mathbf{I} - \mathbf{A}$ 的不变因子和初等因子.

解 方法一: 用初等变换的方法将 $\lambda \mathbf{I} - \mathbf{A}$ 变为史密斯标准形:

$$\begin{aligned} \lambda \mathbf{I} - \mathbf{A} &= \begin{pmatrix} \lambda + 1 & -1 & 0 \\ 4 & \lambda - 3 & 0 \\ -1 & 0 & \lambda - 2 \end{pmatrix} \\ &\simeq \begin{pmatrix} -1 & & 0 & 0 \\ \lambda - 3 & (\lambda + 1)(\lambda - 3) + 4 & & 0 \\ 0 & & -1 & \lambda - 2 \end{pmatrix} \\ &\simeq \begin{pmatrix} 1 & 0 & 0 \\ 0 & (\lambda - 1)^2 & 0 \\ 0 & -1 & \lambda - 2 \end{pmatrix} \\ &\simeq \begin{pmatrix} 1 & 0 & 0 \\ 0 & (\lambda - 2)^2 & (\lambda - 2)(\lambda - 1)^2 \\ 0 & -1 & 0 \end{pmatrix} \\ &\simeq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (\lambda - 2)(\lambda - 1)^2 \end{pmatrix}, \end{aligned}$$

所以 $d_1(\lambda) = d_2(\lambda) = 1, d_3(\lambda) = (\lambda - 2)(\lambda - 1)^2$.

初等因子为 $\lambda - 2, (\lambda - 1)^2$.

方法二:用观察法先找出 $\lambda \mathbf{I} - \mathbf{A}$ 的各阶行列式因子 $D_i(\lambda) (i=1, 2, 3)$, 然后再用公式(2.1.5)求不变因子:

不难看出

$$D_3(\lambda) = \begin{vmatrix} \lambda+1 & -1 & 0 \\ 4 & \lambda-3 & 0 \\ -1 & 0 & \lambda-2 \end{vmatrix} = (\lambda-2)(\lambda-1)^2,$$

而 $D_1(\lambda) = D_2(\lambda) = 1$, 所以不变因子为

$$d_1(\lambda) = \frac{D_1(\lambda)}{D_0(\lambda)} = 1,$$

$$d_2(\lambda) = \frac{D_2(\lambda)}{D_1(\lambda)} = 1,$$

$$d_3(\lambda) = \frac{D_3(\lambda)}{D_2(\lambda)} = (\lambda-2)(\lambda-1)^2,$$

这里取 $D_0(\lambda) = 1$.

故初等因子为 $(\lambda-2), (\lambda-1)^2$.

例 2.1.6 求 n 阶若尔当块

$$\mathbf{J}_i = \begin{pmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_i & 1 \\ & & & & \lambda_i \end{pmatrix}_{n \times n}$$

的特征矩阵 $\lambda \mathbf{I} - \mathbf{J}_i$ 的不变因子和初等因子.

解 特征矩阵为

$$\lambda \mathbf{I} - \mathbf{J}_i = \begin{pmatrix} \lambda - \lambda_i & -1 & & & \\ & \lambda - \lambda_i & -1 & & \\ & & \ddots & \ddots & \\ & & & \lambda - \lambda_i & -1 \\ & & & & \lambda - \lambda_i \end{pmatrix}.$$

用观察法, 不难看出各阶行列式因子为

$$D_1(\lambda) = D_2(\lambda) = \cdots = D_{n-1}(\lambda) = 1,$$

而

$$|\lambda \mathbf{I} - \mathbf{J}_i| = \begin{vmatrix} \lambda - \lambda_i & -1 & & & \\ & \lambda - \lambda_i & \ddots & & \\ & & \ddots & -1 & \\ & & & \lambda - \lambda_i & \\ & & & & \lambda - \lambda_i \end{vmatrix} = (\lambda - \lambda_i)^n,$$

所以

$$D_n(\lambda) = (\lambda - \lambda_i)^n.$$

于是, $\lambda \mathbf{I} - \mathbf{J}_i$ 的不变因子为

$$d_1(\lambda) = d_2(\lambda) = \cdots = d_{n-1}(\lambda) = 1,$$

$$d_n(\lambda) = (\lambda - \lambda_i)^n,$$

因而初等因子只有一个: $(\lambda - \lambda_i)^n$.

由定义 2.1.8 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做成的初等因子中,次数最高的必在 $d_r(\lambda)$ 的分解中,次第二高的必在 $d_{r-1}(\lambda)$ 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 $5 \times 6 \lambda$ 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 $A(\lambda)$ 的史密斯标准形为

$$A(\lambda) \simeq \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda(\lambda - 1) & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda(\lambda - 1)^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda^2(\lambda - 1)^3(\lambda^2 + 1)^3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

由定理 2.1.6 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 2.1.7 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.

下面研究块对角矩阵初等因子的求法.

定理 2.1.8 设 λ 矩阵为块对角形

$$A(\lambda) = \begin{pmatrix} A_1(\lambda) & & & \\ & A_2(\lambda) & & \\ & & \ddots & \\ & & & A_m(\lambda) \end{pmatrix},$$

则 $A_i(\lambda) (i=1, 2, \dots, m)$ 的所有初等因子的集合是 $A(\lambda)$ 的全部初等因子.

证明 利用数学归纳法. 先证 $m=2$ 时成立.

$$J_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}_{m_i \times m_i}, \quad i = 1, 2, \dots, t,$$

试求若尔当标准形的特征矩阵的初等因子.

解 因为

$$\lambda I - J = \begin{pmatrix} \lambda I_1 - J_1 & & & \\ & \lambda I_2 - J_2 & & \\ & & \ddots & \\ & & & \lambda I_t - J_t \end{pmatrix},$$

其中 I_i 是 m_i 阶单位阵. 由例 2.1.6 知 $\lambda I_i - J_i$ 的初等因子是 $(\lambda - \lambda_i)^{m_i}$, 所以 $\lambda I - J$ 的初等因子为 $(\lambda - \lambda_1)^{m_1}, (\lambda - \lambda_2)^{m_2}, \dots, (\lambda - \lambda_t)^{m_t}$. 且有

$$\sum_{i=1}^t m_i = n.$$

由前面的讨论可知, 当 $d_i(\lambda)$ 已知时, $A(\lambda)$ 的初等因子可由式 (2.1.6) 和定义 2.1.8 来确定; 反之, 如果已知 $A(\lambda)$ 的秩 r 和全部的初等因子, 按式 (2.1.6) 排列, $d_i(\lambda)$ 也就被完全确定了, 因而 $A(\lambda)$ 的史密斯标准形

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

也就确定了.

例 2.1.8 已知 $A(\lambda)$ 的秩为 3, 初等因子为 $(\lambda - 1), (\lambda - 1)^2, (\lambda - 2), (\lambda - 5), (\lambda + 2)^2$, 试求 $A(\lambda)$ 的史密斯标准形.

解 将初等因子按式 (2.1.6) 排列如下:

$$\begin{cases} (\lambda - 1)^0 (\lambda - 2)^0 (\lambda - 5)^0 (\lambda + 2)^0, \\ (\lambda - 1) (\lambda - 2)^0 (\lambda - 5)^0 (\lambda + 2)^0, \\ (\lambda - 1)^2 (\lambda - 2) (\lambda - 5) (\lambda + 2)^2. \end{cases}$$

可见, 不变因子为 $d_1(\lambda) = 1, d_2(\lambda) = \lambda - 1, d_3(\lambda) = (\lambda - 1)^2 (\lambda - 2) (\lambda - 5) (\lambda + 2)^2$, 故 $A(\lambda)$ 的史密斯标准形为

$$\begin{pmatrix} 1 & & & \\ & \lambda - 1 & & \\ & & (\lambda - 1)^2 (\lambda - 2) (\lambda - 5) (\lambda + 2)^2 & \\ & & & \end{pmatrix}.$$

注意, 两个 λ 矩阵的初等因子相同, 它们可能不相抵. 例如

$$A(\lambda) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda - 4 & 0 & 0 \\ 0 & 0 & (\lambda - 4)^2 & 0 \end{pmatrix},$$

$$\mathbf{B}(\lambda) = \begin{bmatrix} \lambda-4 & 0 & 0 & 0 \\ 0 & (\lambda-4)^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

都是 3×4 的 λ 矩阵, 它们的初等因子都是 $\lambda-4, (\lambda-4)^2$, 但是它们的秩不等, 于是 $\mathbf{A}(\lambda)$ 与 $\mathbf{B}(\lambda)$ 并不相抵.

例 2.1.9 求下面矩阵 $\mathbf{A}(\lambda)$ 的初等因子、不变因子和史密斯标准形:

$$\mathbf{A}(\lambda) = \begin{bmatrix} 3\lambda+5 & (\lambda+2)^2 & 4\lambda+5 & (\lambda-1)^2 \\ \lambda+7 & (\lambda+2)^2 & \lambda+7 & 0 \\ \lambda-1 & 0 & 2\lambda-1 & (\lambda-1)^2 \\ 0 & 0 & (\lambda-2)(\lambda-5) & 0 \end{bmatrix}.$$

解 对 $\mathbf{A}(\lambda)$ 施行初等变换, 可得

$$\begin{aligned} \mathbf{A}(\lambda) &\stackrel{r_1-r_3}{\cong} \begin{bmatrix} 2\lambda+6 & (\lambda+2)^2 & 2\lambda+6 & 0 \\ \lambda+7 & (\lambda+2)^2 & \lambda+7 & 0 \\ \lambda-1 & 0 & 2\lambda-1 & (\lambda-1)^2 \\ 0 & 0 & (\lambda-2)(\lambda-5) & 0 \end{bmatrix} \\ &\stackrel{c_3-c_1}{\cong} \begin{bmatrix} 2\lambda+6 & (\lambda+2)^2 & 0 & 0 \\ \lambda+7 & (\lambda+2)^2 & 0 & 0 \\ \lambda-1 & 0 & \lambda & (\lambda-1)^2 \\ 0 & 0 & (\lambda-2)(\lambda-5) & 0 \end{bmatrix} \\ &\stackrel{r_1-r_2}{\cong} \begin{bmatrix} \lambda-1 & 0 & 0 & 0 \\ \lambda+7 & (\lambda+2)^2 & 0 & 0 \\ \lambda-1 & 0 & \lambda & (\lambda-1)^2 \\ 0 & 0 & (\lambda-2)(\lambda-5) & 0 \end{bmatrix} \\ &\stackrel{r_3-r_1}{\cong} \begin{bmatrix} \lambda-1 & 0 & 0 & 0 \\ \lambda+7 & (\lambda+2)^2 & 0 & 0 \\ 0 & 0 & \lambda & (\lambda-1)^2 \\ 0 & 0 & (\lambda-2)(\lambda-5) & 0 \end{bmatrix}. \end{aligned}$$

至此, $\mathbf{A}(\lambda)$ 已化为分块对角阵

$$\begin{bmatrix} \mathbf{A}_1(\lambda) & 0 \\ 0 & \mathbf{A}_2(\lambda) \end{bmatrix},$$

其中

$$\begin{aligned} \mathbf{A}_1(\lambda) &= \begin{pmatrix} \lambda-1 & 0 \\ \lambda+7 & (\lambda+2)^2 \end{pmatrix}, \\ \mathbf{A}_2(\lambda) &= \begin{pmatrix} \lambda & (\lambda-1)^2 \\ (\lambda-2)(\lambda-5) & 0 \end{pmatrix}. \end{aligned}$$

对于 $\mathbf{A}_1(\lambda)$, 有

$$D_1(\lambda) = 1, \quad D_2(\lambda) = (\lambda-1)(\lambda+2)^2,$$

所以

$$d_1(\lambda) = 1, \quad d_2(\lambda) = (\lambda - 1)(\lambda + 2)^2,$$

其初等因子为 $(\lambda - 1)$ 和 $(\lambda + 2)^2$; 而对于 $\mathbf{A}_2(\lambda)$, 有

$$D_1(\lambda) = 1, \quad D_2(\lambda) = (\lambda - 2)(\lambda - 5)(\lambda - 1)^2,$$

所以

$$d_1(\lambda) = 1, \quad d_2(\lambda) = (\lambda - 2)(\lambda - 5)(\lambda - 1)^2,$$

其初等因子为 $(\lambda - 2)$, $(\lambda - 5)$ 和 $(\lambda - 1)^2$.

由定理 2.1.8 可知, $\mathbf{A}(\lambda)$ 的初等因子为

$$\lambda - 1, (\lambda - 1)^2, \lambda - 2, \lambda - 5, (\lambda + 2)^2.$$

又 $\text{rank} \mathbf{A}(\lambda) = 4$, 故 $\mathbf{A}(\lambda)$ 的不变因子为

$$d_4(\lambda) = (\lambda - 2)(\lambda - 5)(\lambda - 1)^2(\lambda + 2)^2,$$

$$d_3(\lambda) = \lambda - 1,$$

$$d_2(\lambda) = d_1(\lambda) = 1,$$

因此, $\mathbf{A}(\lambda)$ 的史密斯标准形为

$$\begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & \lambda - 1 & & \\ & & & (\lambda - 2)(\lambda - 5)(\lambda - 1)^2(\lambda + 2)^2 & \end{pmatrix}.$$

例 2.1.10 在控制论中, 常见的矩阵是相伴矩阵(友矩阵):

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{pmatrix},$$

求它的特征矩阵的不变因子, 并将特征矩阵化为史密斯标准形.

解 设 \mathbf{A} 的特征矩阵为 $\mathbf{A}(\lambda)$, 即

$$\mathbf{A}(\lambda) = \lambda \mathbf{I} - \mathbf{A} = \begin{pmatrix} \lambda & -1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & -1 \\ a_n & a_{n-1} & a_{n-2} & \cdots & a_2 & \lambda + a_1 \end{pmatrix}.$$

将 $\mathbf{A}(\lambda)$ 的第 2, 3, \dots , n 列依次乘以 $\lambda, \lambda^2, \dots, \lambda^{n-1}$ 后都加到第 1 列上去, 得

$$\mathbf{A}(\lambda) \simeq \begin{pmatrix} 0 & -1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & -1 \\ f(\lambda) & a_{n-1} & a_{n-2} & \cdots & a_2 & \lambda + a_1 \end{pmatrix}.$$

把上式右端的矩阵记为 $\mathbf{B}(\lambda)$, 其中

$$f(\lambda) = \lambda^n + a_1 \lambda^{n-1} + \cdots + a_{n-2} \lambda^2 + a_{n-1} \lambda + a_n,$$

且

$$\det \mathbf{A}(\lambda) = \det \mathbf{B}(\lambda) = (-1)^{n+1} f(\lambda) (-1)^{n-1} = f(\lambda),$$

即 $D_n(\lambda) = f(\lambda)$. 由于将 $\mathbf{B}(\lambda)$ 的第 1 列第 n 行去掉后剩下的 $n-1$ 阶子式为 $(-1)^{n-1}$, 故

$$D_1(\lambda) = D_2(\lambda) = \cdots = D_{n-1}(\lambda) = 1,$$

从而

$$d_1(\lambda) = d_2(\lambda) = \cdots = d_{n-1}(\lambda) = 1,$$

$$d_n(\lambda) = f(\lambda),$$

所以 $\mathbf{A}(\lambda)$ 的史密斯标准形为

$$\begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & \ddots & & & \\ & & & 1 & & \\ & & & & & f(\lambda) \end{pmatrix}.$$

注 如果矩阵 \mathbf{A} 的特征矩阵的不变因子为

$$1, 1, \cdots, 1, f(\lambda),$$

则称 \mathbf{A} 为属于 $f(\lambda)$ 的相伴矩阵或友矩阵.

2.2 若尔当标准形

2.2.1 数字矩阵化为相似的若尔当标准形

虽然亏损矩阵不能相似于对角阵,但它能相似于一个形式上比对角阵稍复杂的若尔当标准形 \mathbf{J} . 由于若尔当标准形的独特结构揭示了两个矩阵相似的本质关系,故在数值计算和理论推导中经常采用. 利用它不仅容易求出矩阵 \mathbf{A} 的乘幂,还可以讨论矩阵函数和矩阵级数,求解矩阵微分方程. 因此,若尔当标准形的理论在数学、力学和计算方法中得到广泛的应用.

定义 2.2.1 形如

$$\mathbf{J}_i = \begin{pmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_i & 1 \\ & & & & \lambda_i \end{pmatrix}_{m_i \times m_i} \quad (2.2.1)$$

的方阵称为 m_i 阶若尔当块. 其中 λ_i 可以是实数,也可以是复数.

例如

$$\begin{bmatrix} 3 & 1 \\ & 3 \end{bmatrix}, \quad \begin{bmatrix} i & 1 & \\ & i & 1 \\ & & i \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix},$$

都是若尔当块. 特别地,一阶方阵是一阶若尔当块.

这是一个以常数矩阵为系数的 λ 的多项式,称为 $A(\lambda)$ 的多项式表示. A_L 称为 $A(\lambda)$ 的首项系数. 如果 $A(\lambda)$ 是 n 阶方阵, 并且 $A_L = \pm I_n$, 则称 $A(\lambda)$ 的首项系数为 1.

引理 设 $A(\lambda)$ 和 $B(\lambda)$ 分别为 L 次和 m 次的 n 阶 λ 矩阵, 即有

$$A(\lambda) = \sum_{i=0}^L A_i \lambda^i \quad \text{与} \quad B(\lambda) = \sum_{i=0}^m B_i \lambda^i,$$

且 $B(\lambda)$ 的首项系数 B_m 可逆, 即 $\det B_m \neq 0$, 则存在 λ 矩阵 $Q(\lambda), Q'(\lambda)$ 和 $R(\lambda), R'(\lambda)$, 使得

$$A(\lambda) = Q(\lambda)B(\lambda) + R(\lambda), \quad (2.2.4)$$

或

$$A(\lambda) = B(\lambda)Q'(\lambda) + R'(\lambda), \quad (2.2.5)$$

其中 $R(\lambda) \equiv 0, R'(\lambda) \equiv 0$ 或 $\deg R(\lambda) < m, \deg R'(\lambda) < m$.

证明 先证式(2.2.4). 若 $L < m$, 则可取 $Q(\lambda) = 0, R(\lambda) = A(\lambda)$ 便得所需的结果.

若 $L \geq m$, 用 $B_m^{-1} \lambda^{-m}$ 左乘 $B(\lambda)$, 得

$$\begin{aligned} B_m^{-1} \lambda^{-m} B(\lambda) &= I + \lambda^{-1} B_m^{-1} B_{m-1} + \lambda^{-2} B_m^{-1} B_{m-2} + \cdots + \\ &\quad \lambda^{-m+1} B_m^{-1} B_1 + \lambda^{-m} B_m^{-1} B_0, \end{aligned}$$

由上式可得

$$\lambda^L I = B_m^{-1} \lambda^{L-m} B(\lambda) - \lambda^{L-1} B_m^{-1} B_{m-1} - \lambda^{L-2} B_m^{-1} B_{m-2} - \cdots - \lambda^{L-m+1} B_m^{-1} B_1 - \lambda^{L-m} B_m^{-1} B_0,$$

于是

$$\begin{aligned} A(\lambda) &= A_L \lambda^L + A_{L-1} \lambda^{L-1} + \cdots + A_1 \lambda + A_0 \\ &= A_L (B_m^{-1} \lambda^{L-m} B(\lambda) - \lambda^{L-1} B_m^{-1} B_{m-1} - \lambda^{L-2} B_m^{-1} B_{m-2} - \cdots - \\ &\quad \lambda^{L-m+1} B_m^{-1} B_1 - \lambda^{L-m} B_m^{-1} B_0) + A_{L-1} \lambda^{L-1} + \cdots + A_1 \lambda + A_0 \\ &= A_L B_m^{-1} \lambda^{L-m} B(\lambda) + A^{(1)}(\lambda), \end{aligned} \quad (2.2.6)$$

其中 $A^{(1)}(\lambda)$ 的次数 $L_1 \leq L-1$, 且

$$A^{(1)}(\lambda) = A_{L_1}^{(1)} \lambda^{L_1} + \cdots + A_1^{(1)} \lambda + A_0^{(1)},$$

其中 $A_{L_1}^{(1)} \neq 0, L_1 < L$. 若 $L_1 \geq m$, 类似于上述步骤, 可得

$$A^{(2)}(\lambda) = A_{L_2}^{(2)} \lambda^{L_2} + \cdots + A_1^{(2)} \lambda + A_0^{(2)},$$

其中 $A_{L_2}^{(2)} \neq 0, L_2 < L_1$.

继续进行下去, 可以得到一组 λ 矩阵

$$A(\lambda), A^{(1)}(\lambda), A^{(2)}(\lambda), \cdots, A^{(r)}(\lambda),$$

它们的次数越来越低, 所以 $A^{(r)}(\lambda)$ 的次数 $L_r < m$, 而 $L_{r-1} \geq m$. 若记 $A(\lambda) = A^{(0)}(\lambda)$, 则根据式(2.2.6), 这组 λ 矩阵的每一个 $A^{(i-1)}(\lambda)$ 都有

$$A^{(i-1)}(\lambda) = A_{L_{i-1}}^{(i-1)} B_m^{-1} \lambda^{L_{i-1}-m} B(\lambda) + A^{(i)}(\lambda), \quad i = 1, 2, \cdots, r \quad (2.2.7)$$

最后得到

$$\begin{aligned} A(\lambda) &= (A_L B_m^{-1} \lambda^{L-m} + A_{L_1}^{(1)} B_m^{-1} \lambda^{L_1-m} + \cdots + A_{L_{r-1}}^{(r-1)} B_m^{-1} \lambda^{L_{r-1}-m}) B(\lambda) + A^{(r)}(\lambda) \\ &= Q(\lambda) B(\lambda) + R(\lambda), \end{aligned}$$

即证得式(2.2.4). 类似地可证式(2.2.5). 证毕

定理 2.2.1 矩阵 $A \sim B$ 的充要条件是它们相应的特征矩阵 $\lambda I - A \sim \lambda I - B$.

证明 必要性: 设 $A \sim B$, 则存在满秩矩阵 C , 使得

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

故

$$\lambda I - B = \lambda I - C^{-1}AC = C^{-1}(\lambda I - A)C,$$

即 $\lambda I - A$ 与 $\lambda I - B$ 相似,从而也相抵.

充分性: 设 $\lambda I - A \simeq \lambda I - B$, 则存在可逆 λ 矩阵 $U(\lambda), V(\lambda)$, 使得

$$\lambda I - A = U(\lambda)(\lambda I - B)V(\lambda)$$

或

$$U^{-1}(\lambda)(\lambda I - A) = (\lambda I - B)V(\lambda). \quad (2.2.8)$$

由本节引理中的式(2.2.5)和式(2.2.4)得

$$U(\lambda) = (\lambda I - A)Q(\lambda) + U_0, \quad (2.2.9)$$

$$V(\lambda) = R(\lambda)(\lambda I - A) + V_0. \quad (2.2.10)$$

由于 $\lambda I - A$ 是一次 λ 矩阵, 故上式中的 U_0 与 V_0 都为数字矩阵, 把式(2.2.9)代入式(2.2.8)得

$$U^{-1}(\lambda)(\lambda I - A) = (\lambda I - B)R(\lambda)(\lambda I - A) + (\lambda I - B)V_0$$

或

$$[U^{-1}(\lambda) - (\lambda I - B)R(\lambda)](\lambda I - A) = (\lambda I - B)V_0. \quad (2.2.11)$$

上式右端是一个次数为 1 的 λ 矩阵, 而左端 $\lambda I - A$ 也是一个次数为 1 的 λ 矩阵, 所以

$$U^{-1}(\lambda) - (\lambda I - B)R(\lambda) = C \quad (2.2.12)$$

必是一个数字矩阵, 因此式(2.2.11)可以写成

$$C(\lambda I - A) = (\lambda I - B)V_0. \quad (2.2.13)$$

现在证明 C 可逆, 且 $C = U_0^{-1}$.

由式(2.2.12)可得

$$U(\lambda)C = I - U(\lambda)(\lambda I - B)R(\lambda),$$

即

$$I = U(\lambda)C + U(\lambda)(\lambda I - B)R(\lambda).$$

由式(2.2.8)可把上式改为

$$I = U(\lambda)C + (\lambda I - A)V^{-1}(\lambda)R(\lambda).$$

把式(2.2.9)代入上式便得

$$\begin{aligned} I &= [(\lambda I - A)Q(\lambda) + U_0]C + (\lambda I - A)V^{-1}(\lambda)R(\lambda) \\ &= U_0C + (\lambda I - A)[Q(\lambda)C + V^{-1}(\lambda)R(\lambda)]. \end{aligned}$$

比较上式两边 λ 矩阵的次数知, 上式右边第二项必为零矩阵, 故

$$I = U_0C,$$

即 C 可逆, 且

$$C = U_0^{-1}, \quad (2.2.14)$$

把式(2.2.14)代入式(2.2.13)得

$$U_0^{-1}(\lambda I - A) = (\lambda I - B)V_0$$

或

$$\lambda I - A = U_0(\lambda I - B)V_0 = \lambda U_0V_0 - U_0BV_0.$$

比较上式两端得

$$U_0V_0 = I, \quad A = U_0BV_0,$$

故得

$$U_0 = V_0^{-1}, \quad A = V_0^{-1} B V_0,$$

即 $A \sim B$.

证毕

由定理 2.2.1 可见, $A \sim B$ 的问题既然转化为求它们的特征矩阵的相抵问题, 那么再由 2.1 节定理 2.1.6 知, 只要求出 $\lambda I - A$ 的初等因子和秩, 即可决定 B 的形式. 例如, 若取 B 为若尔当标准形, 就有下面的定理.

定理 2.2.2 每个 n 阶矩阵 A 都与一个若尔当标准形 J 相似, 且这个若尔当标准形在不计其中若尔当块的排列次序时, 完全由矩阵 A 惟一决定(即每个矩阵都有若尔当标准形).

证明 设 n 阶矩阵 A 的特征矩阵 $\lambda I - A$ 的初等因子为

$$(\lambda - \lambda_1)^{m_1}, (\lambda - \lambda_2)^{m_2}, \dots, (\lambda - \lambda_t)^{m_t}, \quad (2.2.15)$$

其中 $\lambda_1, \lambda_2, \dots, \lambda_t$ 中可能有相同的值, 指数 m_1, m_2, \dots, m_t 中也可能有相同的值, 但总有

$$\sum_{i=1}^t m_i = n.$$

另一方面, 由 2.1 节中的例 2.1.6 知, 形如式(2.2.1)的 m_i 阶若尔当块 J_i 的特征矩阵 $\lambda I - J_i$ 的初等因子为 $(\lambda - \lambda_i)^{m_i}$. 换言之, $\lambda I - A$ 的每个初等因子 $(\lambda - \lambda_i)^{m_i}$ 对应于一个若尔当块 J_i , 其阶数为 m_i . 那么, 这些若尔当块构成一个若尔当标准形 J (见式(2.2.2)), 而 J 的特征矩阵 $\lambda I - J$ 的初等因子是把各个若尔当块的初等因子合在一起构成的, 即 $\lambda I - J$ 的全部初等因子也为式(2.2.15), 亦即 $\lambda I - A$ 与 $\lambda I - J$ 有相同的初等因子, 且秩相等. 于是由定理 2.1.7 知

$$\lambda I - A \simeq \lambda I - J.$$

再由定理 2.2.1 知, 必有 $A \sim J$.

如果还有若尔当标准形 J' , 使得 $J' \sim A$, 则 J' 与 A 也有相同的初等因子. 因此, J' 与 J 在不计其若尔当块的排列次序时是相同的. 这样就证明了惟一性.

证毕

推论 矩阵 A 可对角化的充要条件是 A 的特征矩阵的初等因子全为一次式.

利用矩阵在相似变换下的若尔当标准形, 可得到线性变换的结构.

定理 2.2.3 设 \mathcal{A} 是复数域上 n 维线性空间 V 上的线性变换, 则在 V 中存在一组基使得 \mathcal{A} 在这组基下的矩阵是若尔当形矩阵.

证明 在 V 中任取一组基 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, 设线性变换 \mathcal{A} 在这组基下的矩阵是 A . 由定理 2.2.2 知, 存在可逆矩阵 P 使得 $P^{-1}AP = J$ 为若尔当形矩阵. 令

$$(\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_n) = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)P,$$

则线性变换 \mathcal{A} 在基 $\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_n$ 下的矩阵 $P^{-1}AP = J$ 为若尔当形矩阵.

证毕

定理 2.2.2 的证明过程实际上是给出了一个如何利用初等因子将 A 化为若尔当标准形的方法. 现将求解步骤归纳如下:

第一步 求出 n 阶方阵 $A = (a_{ij})$ 的特征矩阵 $\lambda I - A$ 的初等因子:

$$(\lambda - \lambda_1)^{m_1}, (\lambda - \lambda_2)^{m_2}, \dots, (\lambda - \lambda_t)^{m_t},$$

其中 $\lambda_1, \lambda_2, \dots, \lambda_t$ 可能有相同的, 指数 m_1, m_2, \dots, m_t 也可能有相同的, 且 $\sum_{i=1}^t m_i = n$.

第二步 写出每个初等因子 $(\lambda - \lambda_i)^{m_i}$ ($i=1, 2, \dots, t$) 对应的若尔当块

$$J_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}_{m_i \times m_i}, \quad i = 1, 2, \dots, t.$$

第三步 写出以这些若尔当块构成的若尔当标准形

$$J = \begin{pmatrix} J_1 & & \\ & J_2 & \\ & & \ddots \\ & & & J_t \end{pmatrix}.$$

例 2.2.1 求矩阵

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

的若尔当标准形.

解 在例 2.1.5 中已求出 $\lambda I - A$ 的初等因子为 $\lambda - 2, (\lambda - 1)^2$, 于是有

$$A \sim J = \begin{pmatrix} J_1 & \\ & J_2 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix},$$

其中 $J_1 = 2, J_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

例 2.2.2 求矩阵

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

的若尔当标准形.

解 因为

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

不难看出

$$D_4(\lambda) = \begin{vmatrix} \lambda - 1 & -2 & -3 & -4 \\ 0 & \lambda - 1 & -2 & -3 \\ 0 & 0 & \lambda - 1 & -2 \\ 0 & 0 & 0 & \lambda - 1 \end{vmatrix} = (\lambda - 1)^4,$$

而且可找到一个三阶子式

$$\begin{vmatrix} -2 & -3 & -4 \\ \lambda-1 & -2 & -3 \\ 0 & \lambda-1 & -2 \end{vmatrix} = -4\lambda(\lambda+1).$$

因为要求三阶行列式因子 $D_3(\lambda)$ 必须整除 $D_4(\lambda)$, 可见 $D_3(\lambda)=1$, 从而 $D_2(\lambda)=D_1(\lambda)=1$, 于是 $\lambda I - A$ 的不变因子为

$$d_1(\lambda) = d_2(\lambda) = d_3(\lambda) = 1, \quad d_4(\lambda) = (\lambda-1)^4,$$

即只有一个初等因子, 故

$$\mathbf{A} \sim \mathbf{J} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

例 2.2.3 求矩阵

$$\mathbf{A} = \begin{pmatrix} 2 & 0 & 2 & 1 \\ 6 & 1 & 4 & 4 \\ 10 & 0 & 0 & 4 \\ 7 & 0 & -7 & 2 \end{pmatrix}$$

的若尔当标准形.

解 矩阵 \mathbf{A} 的特征多项式为

$$\begin{aligned} D_4(\lambda) &= \det(\lambda I - \mathbf{A}) \\ &= \det \begin{pmatrix} \lambda-2 & 0 & -2 & -1 \\ -6 & \lambda-1 & -4 & -4 \\ -10 & 0 & \lambda & -4 \\ -7 & 0 & 7 & \lambda-2 \end{pmatrix} \\ &= (\lambda-2)(\lambda-1)^3. \end{aligned}$$

去掉 $\lambda I - \mathbf{A}$ 的第 4 行第 4 列后的余子式 M_{44} 为

$$\begin{aligned} M_{44} &= \det \begin{pmatrix} \lambda-2 & 0 & -2 \\ -6 & \lambda-1 & -4 \\ -10 & 0 & \lambda \end{pmatrix} \\ &= (\lambda-1)(\lambda^2 - 2\lambda - 20). \end{aligned}$$

同理可得

$$\begin{aligned} M_{44} &= \det \begin{pmatrix} -6 & -4 & -4 \\ -10 & \lambda & -4 \\ -7 & 7 & \lambda-2 \end{pmatrix} \\ &= -2(3\lambda^2 + 28\lambda - 40). \end{aligned}$$

因为 M_{44} 与 M_{12} 互素, 所以 $D_3(\lambda)=1$, 进而有

$$D_2(\lambda) = D_1(\lambda) = D_0(\lambda) = 1.$$

因此, $\lambda I - \mathbf{A}$ 的不变因子为

$$\begin{aligned} d_1(\lambda) &= d_2(\lambda) = d_3(\lambda) = 1, \\ d_4(\lambda) &= (\lambda-2)(\lambda-1)^3, \end{aligned}$$

初等因子为 $(\lambda-2)$ 和 $(\lambda-1)^3$,所以 \mathbf{A} 的若尔当标准形为

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

例 2.2.4 求矩阵

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

的若尔当标准形.

解 将 \mathbf{A} 写成分块形式

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \\ & \mathbf{A}_2 \end{pmatrix},$$

其中

$$\mathbf{A}_1 = \begin{bmatrix} 3 & 4 \\ -1 & -1 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}.$$

先分别求出子矩阵 $\mathbf{A}_1, \mathbf{A}_2$ 的若尔当标准形. 由

$$\lambda \mathbf{I} - \mathbf{A}_1 = \begin{bmatrix} \lambda - 3 & -4 \\ 1 & \lambda + 1 \end{bmatrix}$$

可知, \mathbf{A}_1 的不变因子为 $d_1(\lambda)=1, d_2(\lambda)=(\lambda-1)^2$,故初等因子为 $(\lambda-1)^2$. 因此

$$\mathbf{A}_1 \sim \mathbf{J}_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

再由

$$\lambda \mathbf{I} - \mathbf{A}_2 = \begin{bmatrix} \lambda - 2 & -1 \\ 1 & \lambda \end{bmatrix}$$

可知, \mathbf{A}_2 的不变因子为 $d_1(\lambda)=1, d_2(\lambda)=(\lambda-1)^2$,故初等因子为 $(\lambda-1)^2$. 因此

$$\mathbf{A}_2 \sim \mathbf{J}_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

于是得到 \mathbf{A} 的若尔当标准形为

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_1 & \\ & \mathbf{J}_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

综上所述,每个 n 阶矩阵 \mathbf{A} 总能与一个若尔当标准形相似,且不计若尔当块 \mathbf{J}_i 在对角线上的顺序, \mathbf{A} 的若尔当标准形 \mathbf{J} 是惟一的. 另外,不难看到,当 $m_i=1$ 时($i=1, 2, \dots, t$), \mathbf{J} 即化为对角形,此时 \mathbf{A} 必有 n 个线性无关的特征向量. 可见对角形只不过是若尔当标准形的特殊情况.

上面给出了矩阵 \mathbf{A} 的若尔当标准形的求法,但是还没有给出求所需要的非奇异矩阵 \mathbf{C}

的方法. 在一般情况下, 若只需化 A 为若尔当标准形, 就不必考虑求 C . 但若利用若尔当标准形求解微分方程组, 就免不了要求 C . 下面以三阶的形式来给出求矩阵 C 的方法.

若

$$C^{-1}AC = J = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 1 \\ 0 & 0 & \lambda_2 \end{pmatrix},$$

其中 $C = (x_1, x_2, x_3)$, 于是有

$$A(x_1, x_2, x_3) = (x_1, x_2, x_3) \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 1 \\ 0 & 0 & \lambda_2 \end{pmatrix},$$

即

$$(Ax_1, Ax_2, Ax_3) = (\lambda_1 x_1, \lambda_2 x_2, x_2 + \lambda_2 x_3),$$

由此可得

$$\begin{cases} (\lambda_1 I - A)x_1 = 0, \\ (\lambda_2 I - A)x_2 = 0, \\ (\lambda_2 I - A)x_3 = -x_2, \end{cases} \quad (2.2.16)$$

从而 x_1, x_2 依次是 A 的属于 λ_1, λ_2 的特征向量. 而 x_3 是式(2.2.16)中最后一个非齐次线性方程组的解向量, 求出这些解向量就得到我们所需要的矩阵 C .

又如, 如果

$$C^{-1}AC = J = \begin{pmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{pmatrix},$$

令 $C = (x_1, x_2, x_3)$, 则有

$$A(x_1, x_2, x_3) = (x_1, x_2, x_3) \begin{pmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{pmatrix},$$

由此可得

$$\begin{cases} (\lambda_1 I - A)x_1 = 0, \\ (\lambda_1 I - A)x_2 = -x_1, \\ (\lambda_1 I - A)x_3 = -x_2. \end{cases} \quad (2.2.17)$$

从而 x_1 是属于 λ_1 的特征向量. 而 x_2, x_3 是式(2.2.17)后面两个非齐次线性方程的解向量. 这样, 我们又得到所需要的 C .

应当注意, 任取上面线性方程组的解向量 y_1, y_2, y_3 , 当然不一定恰好是 x_1, x_2, x_3 . 但这没关系, 只要 y_1, y_2, y_3 线性无关, 它们就可以代替 x_1, x_2, x_3 , 线性方程组解的不惟一性, 正好说明所求 C 的不惟一性. 在一般情况下, 设 λ_1 是 A 的 k 重特征值, 则 x_1, x_2, \dots, x_k 可通过解下列各方程组

$$\begin{cases} (\lambda_1 I - A)x_1 = 0, \\ (\lambda_1 I - A)x_i = -x_{i-1}, \quad i = 2, 3, \dots, k \end{cases} \quad (2.2.18)$$

获得, 且这样得到的 x_1, x_2, \dots, x_k 线性无关(以上结论的证明过程冗长, 从略).

有时称 x_2, \dots, x_k 是 A 的属于 λ_1 的广义特征向量.

例 2.2.5 试分别计算例 2.2.1 和例 2.2.2 中所需要的非奇异矩阵 C .

解 因为 $\lambda_1=2, \lambda_2=1$ 分别是例 2.2.1 中的单特征值和二重特征值, 所以由式(2.2.16)有

$$\begin{cases} (2I - A)x_1 = 0, \\ (I - A)x_2 = 0, \\ (I - A)x_3 = -x_2. \end{cases}$$

可见 x_1 和 x_2 为对应于两个相异特征值 2 和 1 的特征向量, 且

$$x_1 = (0, 0, 1)^T, \quad x_2 = (1, 2, -1)^T,$$

而 x_3 是广义特征向量: $x_3 = (0, 1, -1)^T$. 故所求的矩阵 C 为

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

又 $\lambda_1=1$ 是例 2.2.2 的矩阵的 4 重特征值, 于是可利用式(2.2.18)求 C . 先解

$$(\lambda_1 - A)x_1 = 0 \quad \text{得} \quad x_1 = (8, 0, 0, 0)^T.$$

再解

$$(I - A)x_2 = -x_1 \quad \text{得} \quad x_2 = (4, 4, 0, 0)^T,$$

$$(I - A)x_3 = -x_2 \quad \text{得} \quad x_3 = (0, -1, 2, 0)^T,$$

$$(I - A)x_4 = -x_3 \quad \text{得} \quad x_4 = (0, 1, -2, 1)^T.$$

故得

$$C = \begin{pmatrix} 8 & 4 & 0 & 0 \\ 0 & 4 & -1 & 1 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

化矩阵为相似的若尔当标准形是非常重要的, 在很多问题中都涉及这种方法.

例 2.2.6 在解代数方程组中的应用. 设有方程组

$$Ax = b,$$

A 为 n 阶方阵, b 为列向量, 若有 C 存在, 使得

$$C^{-1}AC = J,$$

则做代换 $x=Cz$, 于是由 $Ax=b$ 得 $ACz=b$, 故有

$$C^{-1}ACz = C^{-1}b,$$

即有

$$Jz = C^{-1}b. \quad (2.2.19)$$

这时可按 J 的对角块 J_i 把原方程组分解成若干个小方程组来解. 例如, 已知

$$J = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad K = C^{-1}b = \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix},$$

则方程组 $Jz=k$ 可分解成两个独立的小方程组:

$$2z_1 = k_1; \quad \begin{cases} z_2 + z_3 = k_2, \\ z_3 = k_3. \end{cases}$$

显然,求解这样的两个独立的方程组是容易的.

例 2.2.7 解线性微分方程组

$$\begin{cases} \frac{dy_1}{dx} = -y_1 + y_2, \\ \frac{dy_2}{dx} = -4y_1 + 3y_2, \\ \frac{dy_3}{dx} = y_1 + 2y_3. \end{cases}$$

解 方程组右边的系数方阵为

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

令 $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$, 则原方程组改写为 $\frac{d\mathbf{y}}{dx} = \mathbf{A}\mathbf{y}$.

例 2.2.5 中已求出

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

使得

$$\mathbf{C}^{-1}\mathbf{A}\mathbf{C} = \mathbf{J} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

故做代换 $\mathbf{y} = \mathbf{C}\mathbf{z}$, $\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}$, 则由 $\frac{d\mathbf{y}}{dx} = \mathbf{A}\mathbf{y}$ 得

$$\frac{d\mathbf{C}\mathbf{z}}{dx} = \mathbf{A}\mathbf{C}\mathbf{z},$$

即

$$\frac{d\mathbf{z}}{dx} = \mathbf{J}\mathbf{z}.$$

或

$$\frac{d\mathbf{z}}{dx} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{z}.$$

上面方程的坐标写法是

$$\frac{dz_1}{dx} = 2z_1, \quad \frac{dz_2}{dx} = z_2 + z_3, \quad \frac{dz_3}{dx} = z_3,$$

显然解得

$$z_1 = k_1 e^{2x}, \quad z_3 = k_3 e^x.$$

再解第二个方程得

$$z_2 = e^x \left(\int k_3 e^x e^{-x} dx + k_2 \right) = e^x (k_3 x + k_2).$$

于是再由 $y=Cz$ 即得

$$\begin{aligned} y_1 &= e^x (k_3 x + k_2), \\ y_2 &= e^x (2k_3 x + 2k_2 + k_3), \\ y_3 &= k_1 e^{2x} - e^x (k_3 x + k_2 + k_3), \end{aligned}$$

式中 $k_i (i=1, 2, 3)$ 为任意常数.

* 2.2.2 若尔当标准形的其他求法

1. 凯莱-哈密顿定理与最小多项式

先给出 λ 矩阵的多项式写法. 例如, 二阶三次 λ 矩阵

$$A(\lambda) = \begin{pmatrix} \lambda^3 + \lambda + 1 & \lambda^2 - \lambda + 1 \\ \lambda - 1 & \lambda^3 + \lambda^2 + 2 \end{pmatrix},$$

可写为

$$\begin{aligned} A(\lambda) &= \begin{pmatrix} \lambda^3 + 0\lambda^2 + \lambda + 1 & 0\lambda^3 + \lambda^2 - \lambda + 1 \\ 0\lambda^3 + 0\lambda^2 + \lambda - 1 & \lambda^3 + \lambda^2 + 0\lambda + 2 \end{pmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \lambda^3 + \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \lambda^2 + \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} \lambda + \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}. \end{aligned}$$

一般地, 一个次数不超过 m 的 n 阶 λ 矩阵 $B(\lambda)$ 总可写为

$$B(\lambda) = B_0 \lambda^m + B_1 \lambda^{m-1} + \cdots + B_{m-1} \lambda + B_m,$$

其中 $B_i (i=0, 1, \cdots, m)$ 为 n 阶常数矩阵, 且称该式为 λ 矩阵 $B(\lambda)$ 的多项式写法.

同普通的关于某个变量的多项式一样. 形如

$$a_0 A^m + a_1 A^{m-1} + \cdots + a_{m-1} A + a_m I$$

的式子称为方阵 A 的多项式. 其中 a_i 为常数, m 为正整数.

例如, 设 λ 的多项式 $\varphi(\lambda)$ 为

$$\varphi(\lambda) = a_0 \lambda^m + a_1 \lambda^{m-1} + \cdots + a_{m-1} \lambda + a_m,$$

A 为 n 阶方阵, 则

$$\varphi(A) = a_0 A^m + a_1 A^{m-1} + \cdots + a_{m-1} A + a_m I$$

也是 n 阶方阵.

设 A 为任意 n 阶矩阵, 其特征多项式为

$$f(\lambda) = \det(\lambda I - A) = \lambda^n + a_1 \lambda^{n-1} + a_2 \lambda^{n-2} + \cdots + a_{n-1} \lambda + a_n.$$

矩阵 A 与其特征多项式之间有如下重要关系.

定理 2.2.4 (凯莱-哈密顿(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

$$(\lambda I - A)^* = C_1 \lambda^{n-1} + C_2 \lambda^{n-2} + \cdots + C_{n-1} \lambda + C_n,$$

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

$$\begin{aligned} & (\lambda I - A)(C_1\lambda^{n-1} + C_2\lambda^{n-2} + \dots + C_{n-1}\lambda + C_n) \\ &= I\lambda^n + a_1I\lambda^{n-1} + \dots + a_{n-1}I\lambda + a_nI, \end{aligned}$$

比较两边 λ 的同次幂的系数矩阵, 得

$$\begin{aligned} C_1 &= I, \\ C_2 - AC_1 &= a_1I, \\ C_3 - AC_2 &= a_2I, \\ &\vdots \\ C_n - AC_{n-1} &= a_{n-1}I, \\ -AC_n &= a_nI. \end{aligned}$$

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 $f(A) = \mathbf{0}$.

证毕

例 2.2.8 设

$$A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & \omega & \sqrt{2}i \\ 0 & 0 & \omega^2 \end{pmatrix},$$

其中 $\omega = \frac{-1 + \sqrt{3}i}{2}$, 计算 A^{100} .

解 易见矩阵 A 的特征多项式为

$$(\lambda - 1)(\lambda - \omega)(\lambda - \omega^2) = \lambda^3 - 1,$$

由定理 2.2.4, 有 $A^3 - I = \mathbf{0}$, 即 $A^3 = I$, 所以

$$A^{100} = (A^3)^{33} \cdot A = A.$$

例 2.2.9 设

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

求证:

$$A^n = A^{n-2} + A^2 - I, \quad n \geq 3.$$

证明 易知, 矩阵 A 的特征多项式为

$$f(\lambda) = (\lambda^2 - 1)(\lambda - 1).$$

令

$$\begin{aligned} g(\lambda) &= \lambda^n - \lambda^{n-2} - \lambda^2 + 1 \\ &= \lambda^{n-2}(\lambda^2 - 1) - (\lambda^2 - 1) \\ &= (\lambda^2 - 1)(\lambda^{n-2} - 1) \\ &= (\lambda^2 - 1)(\lambda - 1)(\lambda^{n-3} + \lambda^{n-4} + \dots + \lambda + 1), \quad \lambda \geq 3, \end{aligned}$$

由定理 2.2.4 可知

$$g(\mathbf{A}) = \mathbf{0},$$

因此

$$\mathbf{A}^n = \mathbf{A}^{n-2} + \mathbf{A}^2 - \mathbf{I}, \quad n \geq 3.$$

定义 2.2.4 设 \mathbf{A} 为 n 阶矩阵, 如果存在多项式 $\varphi(\lambda)$ 使得 $\varphi(\mathbf{A}) = \mathbf{0}$, 则称 $\varphi(\lambda)$ 为 \mathbf{A} 的化零多项式.

对任意 n 阶矩阵 \mathbf{A} , $f(\lambda)$ 是 \mathbf{A} 的特征多项式, 由定理 2.2.4 知 $f(\lambda)$ 为 \mathbf{A} 的化零多项式. 如果 $g(\lambda)$ 是任意多项式, 则 $g(\lambda)f(\lambda)$ 也是 \mathbf{A} 的化零多项式. 因此, 任意 n 阶矩阵 \mathbf{A} 的化零多项式总存在, 并且 \mathbf{A} 的特征多项式不一定是 \mathbf{A} 的次数最低的化零多项式.

例如, 设

$$\mathbf{A} = \begin{bmatrix} 3 & -1 & 0 \\ 0 & 2 & 0 \\ 1 & -1 & 2 \end{bmatrix},$$

它的特征多项式为 $f(\lambda) = (\lambda - 2)^2(\lambda - 3)$, 故 $f(\mathbf{A}) = \mathbf{0}$, 设 $m(\lambda) = (\lambda - 2)(\lambda - 3)$, 容易验证 $m(\mathbf{A}) = \mathbf{0}$.

定义 2.2.5 n 阶矩阵 \mathbf{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 \mathbf{A} 的最小多项式, 记为 $m(\lambda)$.

由定理 2.2.4 知, 任意 n 阶矩阵 \mathbf{A} 的最小多项式存在且次数不会超过 n .

定理 2.2.5 多项式 $\varphi(\lambda)$ 是矩阵 \mathbf{A} 的化零多项式的充要条件是 \mathbf{A} 的最小多项式 $m(\lambda)$ 能整除 $\varphi(\lambda)$, 即 $m(\lambda) | \varphi(\lambda)$. 特别地, 有 $m(\lambda) | f(\lambda)$, 即 \mathbf{A} 的最小多项式为其特征多项式的因式.

证明 设 $\varphi(\lambda)$ 是 \mathbf{A} 的化零多项式, 其次数自然不比 $m(\lambda)$ 的次数低, 以 $m(\lambda)$ 除 $\varphi(\lambda)$, 得

$$\varphi(\lambda) = q(\lambda)m(\lambda) + r(\lambda),$$

这里, $r(\lambda) \equiv 0$ 或 $r(\lambda)$ 的次数低于 $m(\lambda)$ 的次数. 于是

$$\varphi(\mathbf{A}) = q(\mathbf{A})m(\mathbf{A}) + r(\mathbf{A}).$$

由 $\varphi(\mathbf{A}) = \mathbf{0}$ 和 $m(\mathbf{A}) = \mathbf{0}$ 得 $r(\mathbf{A}) = \mathbf{0}$. 若 $r(\mathbf{A}) \neq \mathbf{0}$, 则 $r(\lambda)$ 为 \mathbf{A} 的次数低于 $m(\lambda)$ 的化零多项式, 这与 $m(\lambda)$ 为最小多项式矛盾. 因此, $r(\lambda) \equiv 0$, 即 $m(\lambda) | \varphi(\lambda)$.

反之, 若 $m(\lambda) | \varphi(\lambda)$, 则 $\varphi(\lambda) = q(\lambda)m(\lambda)$, 于是

$$\varphi(\mathbf{A}) = q(\mathbf{A})m(\mathbf{A}) = \mathbf{0},$$

即 $\varphi(\lambda)$ 为 \mathbf{A} 的化零多项式. 证毕

定理 2.2.6 相似矩阵有相同的最小多项式.

证明 设 $\mathbf{A} \sim \mathbf{B}$, 则存在可逆矩阵 \mathbf{P} , 使得

$$\mathbf{A} = \mathbf{P}^{-1}\mathbf{B}\mathbf{P},$$

若 $f(\mathbf{B}) = \mathbf{0}$, 则有

$$f(\mathbf{A}) = f(\mathbf{P}^{-1}\mathbf{B}\mathbf{P}) = \mathbf{P}^{-1}f(\mathbf{B})\mathbf{P} = \mathbf{0}. \quad \text{证毕}$$

注 定理 2.2.6 的逆定理不成立, 即最小多项式相同的两矩阵不一定相似. 例如, 设

$$\mathbf{A} = \begin{bmatrix} 2 & & \\ & 3 & \\ & & 3 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 2 & & \\ & 2 & \\ & & 3 \end{bmatrix},$$

它们的特征多项式分别为 $(\lambda-2)(\lambda-3)^2$ 和 $(\lambda-2)^2(\lambda-3)$,由它们的特征多项式不同知它们不相似,但是它们有相同的最小多项式 $(\lambda-2)(\lambda-3)$.

由定理 2.2.5 可知 $m(\lambda) \mid f(\lambda)$,那么商 $\frac{f(\lambda)}{m(\lambda)}$ 等于什么呢? 下面的定理回答了这个问题.

定理 2.2.7 n 阶矩阵 \mathbf{A} 的最小多项式等于它的特征矩阵 $(\lambda\mathbf{I}-\mathbf{A})$ 中的第 n 个不变因子 $d_n(\lambda)$, 因而

$$\frac{f(\lambda)}{m(\lambda)} = \frac{\det(\lambda\mathbf{I}-\mathbf{A})}{m(\lambda)} = \frac{\det(\lambda\mathbf{I}-\mathbf{A})}{d_n(\lambda)} = D_{n-1}(\lambda).$$

证明参见文献[5](第 215 页).

推论 1 方阵 \mathbf{A} 的特征多项式的根必是其最小多项式的根.

证明 将 \mathbf{A} 的特征矩阵 $\lambda\mathbf{I}-\mathbf{A}$ 化为史密斯标准形

$$\lambda\mathbf{I}-\mathbf{A} \simeq \begin{pmatrix} d_1(\lambda) & & & \\ & d_2(\lambda) & & \\ & & \ddots & \\ & & & d_n(\lambda) \end{pmatrix},$$

于是

$$f(\lambda) = \det(\lambda\mathbf{I}-\mathbf{A}) = d_1(\lambda)d_2(\lambda)\cdots d_n(\lambda).$$

若 λ_i 为 $f(\lambda)$ 的一个根,则由 $f(\lambda_i)=0$ 可知, λ_i 必为某个 $d_j(\lambda)$ 的根. 又由 $d_j(\lambda) \mid d_n(\lambda)$ 可知, λ_i 为 $d_n(\lambda)$ 的根,即 λ_i 为 $m(\lambda)$ 的根.

由推论 1 知,若 $f(\lambda)$ 分解为不同的一次因式的幂积,如

$$f(\lambda) = (\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_t)^{m_t},$$

其中 $\sum_{i=1}^t m_i = n$, 且每个 $m_i > 0$, 当 $i \neq j$ 时, $\lambda_i \neq \lambda_j$, 因此

$$m(\lambda) = (\lambda-\lambda_1)^{l_1}(\lambda-\lambda_2)^{l_2}\cdots(\lambda-\lambda_t)^{l_t},$$

其中 $1 \leq l_i \leq m_i, i=1, 2, \dots, t$. 特别地, 当每个 $m_i=1$ 时, 有 $t=n$, 且每个 $l_i=1$, 因而这时 $m(\lambda)=f(\lambda)$. 这就得到下面的推论:

推论 2 若矩阵 \mathbf{A} 的特征值互异, 则它的最小多项式就是特征多项式.

例 2.2.10 求下列方阵的特征多项式与最小多项式:

$$(1) \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}; \quad (2) \begin{pmatrix} a & c_1 & & & \\ & a & c_2 & & \\ & & a & \ddots & \\ & & & \ddots & c_{n-1} \\ & & & & a \end{pmatrix} \quad (\text{其中 } c_1, c_2, \dots, c_{n-1} \text{ 全不为 } 0);$$

$$(3) \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_n \\ 1 & 0 & \cdots & 0 & -a_{n-1} \\ 0 & 1 & \cdots & 0 & -a_{n-2} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & -a_2 \\ 0 & 0 & \cdots & 1 & -a_1 \end{pmatrix}.$$

解 (1) 特征多项式为

$$f(\lambda) = \det \begin{pmatrix} \lambda - 3 & -1 & \\ & \lambda - 3 & \\ & & \lambda - 3 \end{pmatrix} = (\lambda - 3)^3,$$

最小多项式可能是

$$(\lambda - 3), \quad (\lambda - 3)^2, \quad (\lambda - 3)^3$$

中的一个. 通过计算可知

$$\begin{aligned} \mathbf{A} - 3\mathbf{I} &= \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \neq \mathbf{0}, \\ (\mathbf{A} - 3\mathbf{I})^2 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \mathbf{0}, \end{aligned}$$

所以, 最小多项式为

$$m(\lambda) = (\lambda - 3)^2.$$

或者用初等变换将 $\lambda\mathbf{I} - \mathbf{A}$ 化为史密斯标准形:

$$\begin{aligned} \lambda\mathbf{I} - \mathbf{A} &\rightarrow \begin{pmatrix} \lambda - 3 & -1 & 0 \\ 0 & \lambda - 3 & 0 \\ 0 & 0 & \lambda - 3 \end{pmatrix} \\ &\rightarrow \begin{pmatrix} -1 & \lambda - 3 & 0 \\ \lambda - 3 & 0 & 0 \\ 0 & 0 & \lambda - 3 \end{pmatrix} \\ &\rightarrow \begin{pmatrix} -1 & \lambda - 3 & 0 \\ 0 & (\lambda - 3)^2 & 0 \\ 0 & 0 & \lambda - 3 \end{pmatrix} \\ &\rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda - 3 & 0 \\ 0 & 0 & (\lambda - 3)^2 \end{pmatrix}, \end{aligned}$$

故最小多项式为

$$m(\lambda) = d_3(\lambda) = (\lambda - 3)^2.$$

(2) 由例 2.1.4 知, 特征矩阵 $\lambda\mathbf{I} - \mathbf{A}$ 的第 n 个不变因子为 $d_n(\lambda) = (\lambda - a)^n$, 故

$$m(\lambda) = (\lambda - a)^n.$$

(3) 由例 2.1.10 知, $d_n(\lambda) = f(\lambda)$, 故 $m(\lambda) = f(\lambda)$, 即友矩阵的最小多项式就是它的特征多项式.

定理 2.2.8 块对角矩阵 $\mathbf{A} = \text{diag}(\mathbf{A}_1, \dots, \mathbf{A}_s)$ 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 \mathbf{A}_i 的最小多项式为 $m_i(\lambda) (i=1, \dots, s)$, 由于对任意多项式 $\varphi(\lambda)$, 有

$$\varphi(\mathbf{A}) = \text{diag}(\varphi(\mathbf{A}_1), \dots, \varphi(\mathbf{A}_s)),$$

如果 $\varphi(\lambda)$ 为 \mathbf{A} 的化零多项式, 则 $\varphi(\lambda)$ 必为 $\mathbf{A}_i (i=1, \dots, s)$ 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 $\varphi(\mathbf{A}_i) = \mathbf{0} (i=1, \dots, s)$, 从而 $\varphi(\mathbf{A}) = \mathbf{0}$. 因此, \mathbf{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. 证毕

定理 2.2.9 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 则 \mathbf{A} 的最小多项式为 \mathbf{A} 的第 n 个不变因子 $d_n(\lambda)$.

证明 由定理 2.2.2 知 \mathbf{A} 相似于若尔当标准形 $\mathbf{J} = \text{diag}(\mathbf{J}_1, \dots, \mathbf{J}_s)$, 其中 \mathbf{J}_i 为若尔当块. 由定理 2.1.6 和定理 2.2.6 知 \mathbf{A} 与 \mathbf{J} 有相同的不变因子和最小多项式. 而由定理 2.2.8 知 \mathbf{J} 的最小多项式为 $\mathbf{J}_1, \dots, \mathbf{J}_s$ 的最小多项式的最小公倍式. 因为 \mathbf{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 \mathbf{J} 的第 n 个不变因子 $d_n(\lambda)$. 因此 \mathbf{A} 的最小多项式就是 \mathbf{A} 的第 n 个不变因子 $d_n(\lambda)$. 证毕

由定理 2.2.2 的推论和定理 2.2.9 可得如下定理.

定理 2.2.10 n 阶矩阵 \mathbf{A} 相似于对角矩阵的充分必要条件是 \mathbf{A} 的最小多项式 $m(\lambda)$ 没有重零点.

例 2.2.11 如果 n 阶矩阵 \mathbf{A} 满足 $\mathbf{A}^2 = \mathbf{A}$, 则称矩阵 \mathbf{A} 为幂等矩阵. 证明: 幂等矩阵 \mathbf{A} 一定相似于对角矩阵.

证明 记 $\varphi(\lambda) = \lambda^2 - \lambda$, 则 $\varphi(\lambda)$ 是 \mathbf{A} 的化零多项式. 由定理 2.2.5 知 \mathbf{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. 根据定理 2.2.10 知 \mathbf{A} 相似于对角矩阵. 证毕

2. 幂零矩阵的若尔当标准形

设 \mathbf{A} 为一个非零的 n 阶幂零矩阵, 即存在正整数 m 使 $\mathbf{A}^m = \mathbf{0}$, 但 $\mathbf{A}^{m-1} \neq \mathbf{0}$. 称 m 为 \mathbf{A} 的幂零指标. 显然, 这时 \mathbf{A} 的最小多项式为 λ^m . 由此我们可得如下引理.

引理 \mathbf{A} 为一个幂零矩阵 $\Leftrightarrow \mathbf{A}$ 的特征值全为零.

由引理我们知道, 若 \mathbf{A} 为一个 n 阶幂零矩阵, 则 \mathbf{A} 有如下形式的若尔当标准形:

$$\mathbf{N} = \begin{bmatrix} \mathbf{N}_1 & & & \\ & \mathbf{N}_2 & & \\ & & \ddots & \\ & & & \mathbf{N}_s \end{bmatrix}, \quad \mathbf{N}_i = \begin{bmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & \ddots & \ddots \\ & & & \ddots & 1 \\ & & & & 0 \end{bmatrix}_{n_i \times n_i}, \quad i = 1, 2, \dots, s \quad (2.2.20)$$

定理 2.2.11 设 n 阶幂零矩阵 \mathbf{A} 的若尔当标准形为 (2.2.20), 幂零指标为 m . 则

- (1) $m = \max\{n_i | 1 \leq i \leq s\}$;
- (2) \mathbf{A} 的零度 (即 \mathbf{A} 的零化空间的维数) 等于 \mathbf{N} 中若尔当块的个数 s ;
- (3) 记 \mathbf{N} 中 k 阶若尔当块的个数为 ℓ_k , \mathbf{A}^k 的零度为 $\eta_k, 1 \leq k \leq n$, 则

$$\ell_1 = 2\eta_1 - \eta_2 = 2s - \eta_2, \quad (2.2.21)$$

$$\ell_k = 2\eta_k - \eta_{k-1} - \eta_{k+1}, \quad 2 \leq k \leq m. \quad (2.2.22)$$

证明 (1) 由于 A 与 N 相似, 所以 $A^k = \mathbf{0} \Leftrightarrow N^k = \mathbf{0}, k \in \mathbb{Z}_+$. 因

$$N^k = \begin{bmatrix} N_1^k & & & \\ & N_2^k & & \\ & & \ddots & \\ & & & N_s^k \end{bmatrix},$$

且 $N_i^{n_i} = \mathbf{0}, N_i^{n_i-1} \neq \mathbf{0}$, 所以 N 的幂零指标为 $m \Leftrightarrow n_i \leq m, 1 \leq i \leq s$ 且存在 i , 使 $n_i = m$.

(2) 设 A 的零度为 η_1 , 则

$$\eta_1 = n - r(A) = n - r(N) = \sum_{i=1}^s n_i - \sum_{i=1}^s (n_i - 1) = s.$$

(3) 根据 A^k 的零度等于 N^k 的零度, 等于 N_i^k 的零度之和 ($i=1, 2, \dots, s$), 且

$$N_i^k \text{ 的零度} = \begin{cases} k, & k \leq n_i \\ n_i, & k > n_i, \end{cases} \quad (2.2.23)$$

由式(2.2.23)有

$$\eta_1 = A \text{ 的零度} = N \text{ 的零度} = \sum_{i=1}^s (N_i \text{ 的零度}) = \sum_{i=1}^s 1 = s = \sum_{k \geq 1} \ell_k \quad (2.2.24)$$

$$\begin{aligned} \eta_2 = A^2 \text{ 的零度} &= N^2 \text{ 的零度} = \sum_{i=1}^s (N_i^2 \text{ 的零度}) \\ &= \sum_{i: n_i < 2} (N_i^2 \text{ 的零度}) + \sum_{i: n_i \geq 2} (N_i^2 \text{ 的零度}) = \ell_1 + 2 \sum_{k \geq 2} \ell_k, \end{aligned} \quad (2.2.25)$$

.....

$$\begin{aligned} \eta_j = A^j \text{ 的零度} &= N^j \text{ 的零度} = \sum_{i=1}^s (N_i^j \text{ 的零度}) \\ &= \sum_{i: n_i < j} (N_i^j \text{ 的零度}) + \sum_{i: n_i \geq j} (N_i^j \text{ 的零度}) = \sum_{k < j} k \ell_k + j \sum_{k \geq j} \ell_k, \end{aligned} \quad (2.2.26)$$

.....

由式(2.2.24)及式(2.2.25)即可推出式(2.2.21). 而式(2.2.22)可由式(2.2.26)推出. 证毕

我们在求幂零矩阵的若尔当标准形时, 总是把阶数大的若尔当块放在前面(在这个意义下, 对应的若尔当标准形是惟一的.)

上述定理给了我们一个求幂零矩阵的若尔当标准形的方法.

例 2.2.12 试证下列矩阵为幂零矩阵, 并求其若尔当标准形:

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

解 直接计算便知 $A^2 = \mathbf{0}$. 因此 A 为幂零矩阵, 且幂零指数为 2.

注意到任何幂零矩阵的若尔当标准形除了次对角线上的元素可能为 1 或 0 外, 其他位置的元素均为零, 因此 A 的若尔当标准形为

$$\mathbf{N} = \begin{bmatrix} 0 & a_1 & & \\ & 0 & a_2 & \\ & & 0 & a_3 \\ & & & 0 \end{bmatrix},$$

其中 a_1, a_2, a_3 为 0 或 1. 由于 $\mathbf{A} \neq \mathbf{0}$, 至少有一个 $a_i \neq 0$, 因此 $a_1 = 1$, 又由于 \mathbf{A} 的幂零指标为 2, 因此没有大于 2 阶的若尔当块, 即 $a_2 = 0$. 至于 a_3 为 0 或 1, 只需看 \mathbf{A} 的秩便知, 直接观察便知 $r(\mathbf{A}) > 1$ (因为秩小于或等于 1 的矩阵的每一行必成比例. 实际上很容易求出 $r(\mathbf{A}) = 2$), 因此 $a_3 = 1$. 所以 \mathbf{A} 的若尔当标准形为

$$\mathbf{N} = \begin{bmatrix} 0 & 1 & & \\ & 0 & & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}.$$

习 题 2

1. 下列矩阵能否与对角矩阵相似? 若能与对角矩阵相似, 则求出可逆矩阵 \mathbf{P} , 使 $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ 为对角矩阵.

$$(1) \mathbf{A} = \begin{bmatrix} 3 & 4 \\ 5 & 2 \end{bmatrix}; \quad (2) \mathbf{A} = \begin{bmatrix} 5 & -3 & 2 \\ 6 & -4 & 4 \\ 4 & -4 & 5 \end{bmatrix}; \quad (3) \mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{bmatrix}.$$

2. 求下列矩阵的不变因子及初等因子.

$$(1) \mathbf{A} = \begin{bmatrix} -1 & & & \\ & -2 & & \\ & & 1 & \\ & & & 2 \end{bmatrix}; \quad (2) \mathbf{A} = \begin{bmatrix} a & -b_1 & & 0 \\ & a & \ddots & \\ & & \ddots & -b_{n-1} \\ & & & a \end{bmatrix} (b_i \neq 0);$$

$$(3) \mathbf{A} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ -2 & -2 & -1 \end{bmatrix}; \quad (4) \mathbf{A} = \begin{bmatrix} -1 & -2 & -1 & 0 \\ 2 & -1 & 0 & -1 \\ 0 & 0 & -1 & -2 \\ 0 & 0 & 2 & -1 \end{bmatrix}.$$

3. 判断下列两个 λ 矩阵是否相抵.

$$\mathbf{A}(\lambda) = \begin{bmatrix} 3\lambda+1 & \lambda & 4\lambda-1 \\ 1-\lambda^2 & \lambda-1 & \lambda-\lambda^2 \\ \lambda^2+\lambda+2 & \lambda & \lambda^2+2\lambda \end{bmatrix}, \quad \mathbf{B}(\lambda) = \begin{bmatrix} \lambda+1 & \lambda-2 & \lambda^2-2\lambda \\ 2\lambda & 2\lambda-3 & \lambda^2-2\lambda \\ -2 & 1 & 1 \end{bmatrix}.$$

4. 求下列 λ 矩阵的不变因子:

$$(1) \begin{bmatrix} \lambda & -1 & 0 & 0 \\ 0 & \lambda & -1 & 0 \\ 0 & 0 & \lambda & -1 \\ 5 & 4 & 3 & \lambda+2 \end{bmatrix}; \quad (2) \begin{bmatrix} \lambda+\alpha & \beta & 1 & 0 \\ -\beta & \lambda+\alpha & 0 & 1 \\ 0 & 0 & \lambda+\alpha & \beta \\ 0 & 0 & -\beta & \lambda+\alpha \end{bmatrix}$$

5. 求下列矩阵的初等因子组:

$$(1) \begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ -2 & -2 & -1 \end{bmatrix}; \quad (2) \begin{bmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}.$$

6. 已知

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

求 \mathbf{A}^{100} .

7. 求下列 λ 矩阵的不变因子:

$$(1) \begin{bmatrix} \lambda-2 & -1 & 0 \\ 0 & \lambda-2 & -1 \\ 0 & 0 & \lambda-2 \end{bmatrix};$$

$$(2) \begin{bmatrix} \lambda & -1 & 0 & 0 \\ 0 & \lambda & -1 & 0 \\ 0 & 0 & \lambda & -1 \\ 5 & 4 & 3 & \lambda+2 \end{bmatrix};$$

$$(3) \begin{bmatrix} 0 & 0 & 1 & \lambda+2 \\ 0 & 1 & \lambda+2 & 0 \\ 1 & \lambda+2 & 0 & 0 \\ \lambda+2 & 0 & 0 & 0 \end{bmatrix}.$$

8. 求下列 λ 矩阵的初等因子:

$$(1) \begin{bmatrix} \lambda^3+2 & \lambda^3+1 \\ 2\lambda^3-\lambda^2-\lambda+3 & 2\lambda^3-\lambda^2-\lambda+2 \end{bmatrix};$$

$$(2) \begin{bmatrix} \lambda^3-2\lambda^2+2\lambda-1 & \lambda^2-2\lambda+1 \\ 2\lambda^3-2\lambda^2+\lambda-1 & 2\lambda^2-2\lambda \end{bmatrix}.$$

9. 求下列 λ 矩阵的史密斯标准形.

$$(1) \mathbf{A}(\lambda) = \begin{pmatrix} 0 & \lambda(\lambda-1) & 0 \\ \lambda & 0 & \lambda+1 \\ 0 & 0 & -\lambda+2 \end{pmatrix};$$

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

10. 化下列矩阵为史密斯标准形.

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

$$(2) \begin{pmatrix} 0 & 0 & 0 & \lambda^2 \\ 0 & 0 & \lambda^2-\lambda & 0 \\ 0 & (\lambda-1)^2 & 0 & 0 \\ \lambda^2-\lambda & 0 & 0 & 0 \end{pmatrix};$$

$$(3) \begin{pmatrix} 3\lambda^2+2\lambda-3 & 2\lambda-1 & \lambda^2+2\lambda-3 \\ 4\lambda^2+3\lambda-5 & 3\lambda-2 & \lambda^2+3\lambda-4 \\ \lambda^2+\lambda-4 & \lambda-2 & \lambda-1 \end{pmatrix};$$

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

11. 求下列 λ 矩阵的史密斯标准形.

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

$$(2) \begin{pmatrix} \lambda^2-1 & 0 \\ 0 & (\lambda-1)^2 \end{pmatrix};$$

$$(3) \begin{pmatrix} 0 & 0 & 1 & \lambda+4 \\ 0 & 1 & \lambda+4 & 0 \\ 1 & \lambda+4 & 0 & 0 \\ \lambda+4 & 0 & 0 & 0 \end{pmatrix}.$$

12. 设

$$\mathbf{A}_1 = \begin{bmatrix} 3 & 2 & -5 \\ 2 & 6 & -10 \\ 1 & 2 & -3 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 6 & 20 & -34 \\ 6 & 32 & -51 \\ 4 & 20 & -32 \end{bmatrix},$$

分别求 $\lambda I - A_1$ 与 $\lambda I - A_2$ 的史密斯标准形以及 A_1 与 A_2 的不变因子、行列式因子.

13. 证明: 相抵的 λ 矩阵具有相同的秩与相同的各阶行列式因子.
14. 证明: 两个相抵的 λ 矩阵(方阵)的行列式只相差一个非零的常数因子.
15. 秩相等的两个 $m \times n$ 的 λ 矩阵是否一定相抵?
16. 证明: 一个 $n \times n$ 的 λ 矩阵可逆的充分必要条件是: 行列式 $|\mathbf{A}(\lambda)|$ 是一个非零的数.

$$17. \text{ 求 } \mathbf{A}(\lambda) = \begin{pmatrix} \lambda - a & c_1 & & & \\ & \lambda - a & c_2 & & \\ & & \lambda - a & \ddots & \\ & & & \ddots & c_{n-1} \\ & & & & \lambda - a \end{pmatrix} \text{ 的不变因子与初等因子, 其中 } a, c_1, \dots, c_{n-1} \text{ 都是常数,}$$

且 $c_i \neq 0 (i=1, 2, \dots, n-1)$.

18. 求证: 两个 $m \times n$ 的 λ 矩阵相抵当且仅当它们有相同的各阶行列式因子.
19. 求出下列矩阵的若尔当标准形.

$$(1) \mathbf{A} = \begin{pmatrix} 2 & -1 & -1 \\ 2 & -1 & -2 \\ -1 & 1 & 2 \end{pmatrix}; \quad (2) \mathbf{A} = \begin{pmatrix} 4 & 6 & 0 \\ -3 & -5 & 0 \\ -3 & -6 & 1 \end{pmatrix};$$

(3) 第2题中的各矩阵.

20. 求下列各矩阵的若尔当标准形.

$$(1) \begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ -2 & -1 & -1 \end{pmatrix}; \quad (2) \begin{pmatrix} 3 & 7 & -3 \\ -2 & -5 & 2 \\ -4 & -10 & 3 \end{pmatrix}; \quad (3) \begin{pmatrix} 3 & 1 & 0 & 0 \\ -4 & -1 & 0 & 0 \\ 7 & 1 & 2 & 1 \\ -7 & -6 & -1 & 0 \end{pmatrix}.$$

21. 试判断下面4个矩阵, 哪些是相似的, 为什么?

$$\mathbf{A} = \begin{bmatrix} -3 & 3 & -2 \\ -7 & 6 & -3 \\ 1 & -1 & 2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 1 & -1 \\ -4 & 4 & -2 \\ -2 & 1 & 1 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} 0 & -1 & -1 \\ -3 & -1 & -2 \\ 7 & 5 & 6 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

22. 设矩阵

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

求 \mathbf{A}^5 .

23. 设矩阵

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & -1 \\ 2 & -1 & -2 \\ -1 & 1 & 2 \end{bmatrix},$$

求 \mathbf{A} 的若尔当标准形 \mathbf{J} , 并求相似变换矩阵 \mathbf{P} , 使得 $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{J}$.

24. 证明: 若尔当块

$$\mathbf{J}(a) = \begin{bmatrix} a & 1 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{bmatrix}$$

相似于矩阵

$$\begin{bmatrix} a & \epsilon & 0 \\ 0 & a & \epsilon \\ 0 & 0 & a \end{bmatrix}$$

这里 $\epsilon \neq 0$ 为任意实数.

25. 已知 10 阶矩阵

$$\mathbf{A} = \begin{bmatrix} a & 1 & & & & & & & & \\ & a & 1 & & & & & & & \\ & & \ddots & \ddots & & & & & & \\ & & & \ddots & \ddots & & & & & \\ & & & & & & & & & 1 \\ & & & & & & & & & a \end{bmatrix}_{10 \times 10}, \quad \mathbf{B} = \begin{bmatrix} a & 1 & & & & & & & & \\ & a & 1 & & & & & & & \\ & & \ddots & \ddots & & & & & & \\ & & & \ddots & \ddots & & & & & \\ & & & & & & & & & 1 \\ \epsilon & & & & & & & & & a \end{bmatrix}_{10 \times 10}$$

其中 $\epsilon = 10^{-10}$, 证明 \mathbf{A} 不相似于 \mathbf{B} .

26. 已知 $\mathbf{A}^2 = \mathbf{A}$, 证明: \mathbf{A} 相似于矩阵

$$\begin{bmatrix} 1 & & & & & & & & & \\ & \ddots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & 0 & & & & & & \\ & & & & \ddots & & & & & \\ & & & & & 0 & & & & \\ & & & & & & \ddots & & & \\ & & & & & & & 0 & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & 0 \end{bmatrix}$$

27. 设矩阵 \mathbf{A} 的特征值互不相同, 且有 $\mathbf{AB} = \mathbf{BA}$, 证明: $\lambda \mathbf{I} - \mathbf{B}$ 的初等因子均为一次因式.

28. 设 $\mathbf{A} \neq \mathbf{0}, \mathbf{A}^k = \mathbf{0} (k \geq 2)$. 证明: \mathbf{A} 不能与对角矩阵相似.

29. 求下列矩阵的若尔当标准形及其相似变换矩阵 \mathbf{P} .

$$(1) \begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ -2 & -2 & 1 \end{bmatrix}; \quad (2) \begin{bmatrix} -1 & 1 & 1 \\ -5 & 21 & 17 \\ 6 & -26 & -21 \end{bmatrix};$$

$$(3) \begin{bmatrix} 4 & 5 & -2 \\ -2 & -2 & 1 \\ -1 & -1 & 1 \end{bmatrix}; \quad (4) \begin{bmatrix} 3 & 0 & 8 \\ 3 & -1 & 6 \\ -2 & 0 & -5 \end{bmatrix}.$$

30. 用求矩阵秩的方法求下面矩阵的若尔当标准形.

$$\mathbf{A} = \begin{bmatrix} 8 & -3 & 6 \\ 3 & -2 & 0 \\ -4 & 2 & -2 \end{bmatrix}.$$

31. 试写出若尔当标准形均为

$$\mathbf{J} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix} \text{ 的两个矩阵}$$

32. 已知 $\mathbf{A}^2 = \mathbf{I}$, 证明: \mathbf{A} 相似于矩阵

$$\begin{bmatrix} 1 & & & & & & & & & \\ & \ddots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & -1 & & & & & & \\ & & & & \ddots & & & & & \\ & & & & & -1 & & & & \\ & & & & & & \ddots & & & \\ & & & & & & & -1 & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & -1 \end{bmatrix}.$$

33. 已知 $A^k = I$ (k 为正整数), 证明: A 与对角矩阵相似.

34. 求证: n 阶非零矩阵 A 可对角化的充要条件是对于任意的常数 k 都有 $r(kI - A) = r(kI - A)^2$.

35. 已知矩阵

$$A = \begin{bmatrix} 3 & 0 & 0 \\ a & 3 & 0 \\ c & b & 2 \end{bmatrix},$$

(1) 求 A 的所有可能的若尔当标准形;

(2) 给出 A 可对角化条件.

36. 求可逆矩阵 P 及 J , 使 $P^{-1}AP = J$, 其中

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

37. 设 $W = L[e^x, xe^x, x^2e^x, e^{2x}]$ 为函数向量 $e^x, xe^x, x^2e^x, e^{2x}$ 生成的四维空间, \mathcal{D} 为导数变换,

(1) 求 \mathcal{D} 在基 $e^x, xe^x, x^2e^x, e^{2x}$ 下的矩阵;

(2) 找一组基, 使 \mathcal{D} 在此基下为若尔当标准形.

38. 已知

$$A = \begin{bmatrix} 2 & 0 & 0 \\ a & 2 & 0 \\ b & c & -1 \end{bmatrix},$$

(1) 求出 A 的可能若尔当标准形;

(2) 给出 A 可对角化的条件.

39. 在多项式空间 $P[x]_{n-1}$ 中, \mathcal{D} 是 $P[x]_{n-1}$ 的一个导数变换, 证明 \mathcal{D} 在任一基下的矩阵是不可对角化的.

40. A 为 n 阶方阵, 证明: A^T 与 A 有相同的若尔当标准形.

41. 设 A 是主对角元为 0 的 n 阶上三角矩阵, 证明存在正整数 k , 使 $A^k = \mathbf{0}$.

42. 设 A 是 n 阶不可逆矩阵, 但不是幂零矩阵, 证明存在可逆矩阵 P , 使得

$$P^{-1}AP = \begin{bmatrix} B & \mathbf{0} \\ \mathbf{0} & C \end{bmatrix},$$

其中 B 是可逆块矩阵, C 为幂零块矩阵 (即 $C^k = \mathbf{0}$).

43. 设 $A, B \in C^{n \times n}$, 且 $A^n = \mathbf{0}, B^n = \mathbf{0}$, 但 $A^{n-1} \neq \mathbf{0}, B^{n-1} \neq \mathbf{0}$, 证明 A 与 B 相似.

44. 设 $A(\lambda)$ 为 5 阶 λ 矩阵, 其秩为 4, 初等因子为 $\lambda, \lambda^2, \lambda - 1, \lambda - 1, \lambda + 1, (\lambda + 1)^3$. 试求 $A(\lambda)$ 的不变因子并写出其标准形.

45. 已知 7 阶 λ 矩阵 $A(\lambda)$ 的秩为 5, 初等因子是 $\lambda, \lambda, \lambda^3, \lambda - 2, (\lambda - 2)^4, (\lambda - 4)^4$. 求 $A(\lambda)$ 的各阶子式的最高公因子.

46. 试证 $\begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}$ 的若尔当标准形是 $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

47. 求可逆矩阵 P , 使 $P^{-1} \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix} P = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

48. 试证: 假如有正整数 m , 使 $A^m = I$, 则 A 与对角矩阵相似. 这样的矩阵 A 的特征值只能是哪些数?

49. 证明: 任意方阵可表示为两个对称方阵的乘积, 且其中一个是可逆的.

50. 设 A 的特征值各不相等, 且 $AB = BA$. 求证: $\lambda I - B$ 的初等因子全是一次的.

51. 应用矩阵的若尔当标准形求解线性微分方程组

$$\begin{cases} \frac{dx_1}{dt} = -x_1 + x_2, \\ \frac{dx_2}{dt} = -4x_1 + 3x_2, \\ \frac{dx_3}{dt} = -8x_1 + 8x_2 - x_3, \end{cases}$$

这里 x_1, x_2, x_3 都是 t 的未知函数.

52. 试证: (1) 不等于零的幂零矩阵一定不相似于对角矩阵; (2) 设 A 具有惟一特征值但 A 不是对角矩阵, 证明 A 一定不相似于对角矩阵.

53. 在复数域上求矩阵

$$A = \begin{bmatrix} -4 & 2 & 10 \\ -4 & 3 & 7 \\ -3 & 1 & 7 \end{bmatrix}$$

的若尔当标准形 J , 并求出可逆矩阵 P , 使得 $P^{-1}AP = J$.

54. 设矩阵

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

试计算 $2A^8 - 3A^5 + A^4 + A^2 - 4I$.

55. 设矩阵

$$A = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}$$

试计算 $(A^4 - 5A^3 + 6A^2 + 6A - 8I)^{-1}$.

56. 已知三阶矩阵 A 的三个特征值为 $1, -1, 2$, 试将 A^{2n} 表示为 A 的二次式.

57. 求 $g(A) = A^7 - A^5 - 19A^4 + 28A^3 + 6A - 4I$, 其中

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

58. 已知矩阵 $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, 证明 $n \geq 3$ 时, 恒有 $A^n = A^{n-2} + A^2 - I$, 并计算 A^{1000} .

59. 设 A 为三阶方阵, 其特征值 $1, -1, 2$, 试将 A^{2n} 表示为 A 的二次矩阵多项式, 即形式为 $A^{2n} = aA^2 + bA + cI$.

60. 设 A 为 n 阶可逆方阵, 试将 A^{-1} 表示为 A 的 $n-1$ 次矩阵多项式的形式, 若

$$A = \begin{bmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{bmatrix}, \text{ 求 } A^{-1}.$$

61. 举例说明, 即使两个 n 阶矩阵 A, B 有相同的特征多项式和相同的最小多项式, 但 A 与 B 不一定相似.

62. 求下列矩阵的最小多项式:

$$(1) \begin{bmatrix} 3 & 1 & -1 \\ 0 & 2 & 0 \\ 1 & 1 & 1 \end{bmatrix}; \quad (2) \begin{bmatrix} 4 & -2 & 2 \\ -5 & 7 & -5 \\ -6 & 7 & -4 \end{bmatrix};$$

(3) n 阶单位矩阵 I_n ; (4) n 阶方阵 A , 其元素均为 1;

$$(5) \mathbf{B} = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 \\ -a_1 & a_0 & -a_3 & a_2 \\ -a_2 & a_3 & a_0 & -a_1 \\ -a_3 & -a_2 & a_1 & a_0 \end{bmatrix}.$$

63. n 阶方阵 A, B 有相同的特征多项式, 且它们的最小多项式等于特征多项式, 证明 A 和 B 相似.

64. 求 A 的最小多项式, 其中 A 为

$$\mathbf{A} = \begin{bmatrix} 3 & -3 & 2 \\ -1 & 5 & -2 \\ -1 & 3 & 0 \end{bmatrix}.$$

65. 设 $A \sim J$, 求 A 的最小多项式, 其中

$$\mathbf{J}_{6 \times 6} = \begin{bmatrix} 5 & & & & & \\ & 5 & 1 & 0 & & \\ & & 5 & 1 & & \\ & & & 5 & & \\ & & & & 2 & 1 \\ & & & & & 2 \end{bmatrix} = \begin{bmatrix} \mathbf{J}_1 & & \\ & \mathbf{J}_2 & \\ & & \mathbf{J}_3 \end{bmatrix}.$$

66. (1) A 为 4 阶方阵, 其特征值为 3, 2, 2, 2, 求 A 的可能若尔当标准形;

(2) A 为 4 阶方阵, A 的最小多项式为 $m_A(\lambda) = (\lambda-1)(\lambda-2)^2$, 求 A 的可能若尔当标准形.

矩阵的分解

引言 矩阵分解的意义

矩阵分解对矩阵理论及近代计算数学的发展起了关键作用. 所谓矩阵分解, 就是将一个矩阵写成结构比较简单的或性质比较熟悉的另一些矩阵的乘积. 例如, 在第 1 章中, 我们已经看到如果 n 阶方阵 A 是非亏损的, 则 A 可化为相似的对角矩阵

$$C^{-1}AC = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n), \quad (3.0.1)$$

其中 $\lambda_i (i=1, 2, \dots, n)$ 是 A 的特征值, 而 C 是由相应的 n 个线性无关的特征向量为列组成的可逆矩阵. 式(3.0.1)的本质就是将矩阵 A 分解为三个矩阵的乘积

$$A = C\Lambda C^{-1}. \quad (3.0.2)$$

又如, 在第 2 章中看到任一个实(复)方阵都与若尔当标准形 J 相似, 就意味着 A 有分解 $A = CJC^{-1}$, 但是, 上面仅仅限于在“相似”的条件下来研究矩阵的分解, 在应用上有很大的局限性, 往往达不到简化计算和深入理论的目的. 因此本章将要介绍的矩阵分解的特点是不一定考虑矩阵的相似关系而得出的几种分解, 并给出具体的矩阵分解方法和计算公式. 如在 3.1 节, 讨论的是以初等变换为依据的 LU 三角分解法; 3.2 节介绍的是以吉文斯与豪斯霍尔德变换为依据的 QR 正交三角分解法; 最后还将介绍矩阵的最大秩(满秩)分解、奇异值分解和极分解、谱分解等. 所有这些分解在数值代数和最优化问题的解法中都扮演着十分重要的角色.

3.1 矩阵的三角分解

3.1.1 消元过程的矩阵描述

只要我们对大家熟知的高斯消去法的消元过程用矩阵的语言作出描述, 就不难建立矩阵的三角分解理论.

设 n 元线性方程组为

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2, \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n, \end{cases}$$

或

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \quad (3.1.1)$$

其中, $\mathbf{A} = (a_{ij})_{n \times n}$, $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$, $\mathbf{b} = (b_1, b_2, \dots, b_n)^T$.

高斯消去法的基本思想是利用矩阵的初等行变换化系数矩阵 \mathbf{A} 为上三角矩阵. 不妨假定化 \mathbf{A} 为上三角矩阵的过程未用到行交换, 即按自然顺序进行消元. 其步骤如下:

记

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

如果第 1 个主元素 $a_{11}^{(1)} \neq 0$, 则分别从第 2 行减去第 1 行乘以 $a_{21}^{(1)}/a_{11}^{(1)}$, 第 3 行减去第 1 行乘以 $a_{31}^{(1)}/a_{11}^{(1)}$, 如此进行, 即可将 $\mathbf{A}^{(1)}$ 第 1 列上从第 2 到第 n 个元素全化为零. 得

$$\mathbf{A}^{(2)} = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2}^{(2)} & \cdots & a_{nn}^{(2)} \end{pmatrix}.$$

若令 $l_{i1} = a_{i1}^{(1)}/a_{11}^{(1)}$, 则新元素 $a_{ij}^{(2)}$ 的计算公式应为

$$a_{ij}^{(2)} = a_{ij}^{(1)} - l_{i1}a_{1j}^{(1)}, \quad i, j = 2, 3, \dots, n.$$

以上一共进行了 $n-1$ 次的倍加初等变换, 由线性代数知, 这相当于给 $\mathbf{A}^{(1)}$ 左乘 $n-1$ 个“倍加阵”

$$\mathbf{A}^{(2)} = \mathbf{I}(1, n(-l_{n1})) \cdots \mathbf{I}(1, 2(-l_{21}))\mathbf{A}^{(1)}.$$

若记

$$\mathbf{L}^{(1)} = \mathbf{I}(1, n(-l_{n1})) \cdots \mathbf{I}(1, 2(-l_{21}))$$

$$= \begin{pmatrix} 1 & & & & \\ -l_{21} & 1 & & & \\ \vdots & & \ddots & & \\ -l_{n1} & & & 1 & \end{pmatrix} = \begin{pmatrix} 1 & & & & \\ -\frac{a_{21}^{(1)}}{a_{11}^{(1)}} & 1 & & & \\ \vdots & & \ddots & & \\ -\frac{a_{n1}^{(1)}}{a_{11}^{(1)}} & & & 1 & \end{pmatrix}, \quad (3.1.2)$$

于是有

$$\mathbf{L}^{(1)}\mathbf{A}^{(1)} = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2}^{(2)} & \cdots & a_{nn}^{(2)} \end{pmatrix} = \mathbf{A}^{(2)}. \quad (3.1.3)$$

由此可见, $\mathbf{A}^{(1)} = \mathbf{A}$ 的第 1 列除主元素 $a_{11}^{(1)}$ 外, 其余元素全被化为零, 因此特殊的下三角矩阵 (3.1.2) 称为消元矩阵.

类似地, 若主元素 $a_{22}^{(2)} \neq 0$, 同样可作消元矩阵

$$L^{(2)} = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & -l_{32} & 1 & & \\ & \vdots & & \ddots & \\ & -l_{n2} & & & 1 \end{pmatrix},$$

其中

$$l_{i2} = \frac{a_{i2}^{(2)}}{a_{22}^{(2)}}, \quad i = 3, 4, \dots, n,$$

则可将 $A^{(2)}$ 的第 2 列主对角线以下的元素全化为零, 即有

$$L^{(2)} A^{(2)} = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \cdots & a_{1n}^{(1)} \\ & a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2n}^{(2)} \\ & & a_{33}^{(3)} & \cdots & a_{3n}^{(3)} \\ & & \vdots & & \vdots \\ & & a_{n3}^{(3)} & \cdots & a_{nn}^{(3)} \end{pmatrix} = A^{(3)}. \quad (3.1.4)$$

如此继续, 经 $n-1$ 步后, 则可将 A 变为上三角矩阵

$$L^{(n-1)} \cdots L^{(2)} L^{(1)} A^{(1)} = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ & & \ddots & \vdots \\ & & & a_{nn}^{(n)} \end{pmatrix} = A^{(n)}. \quad (3.1.5)$$

这种对 A 的元素进行的消元过程便是有名的高斯消元过程, 而式(3.1.3)~式(3.1.5)是消元过程的矩阵描述. 显然, 高斯消元过程能够进行到底当且仅当每一步的主元素 $a_{11}^{(1)}$, $a_{22}^{(2)}$, \dots , $a_{n-1, n-1}^{(n-1)}$ 都不为零. 怎样判别 A 的前 $n-1$ 个主元素是否为零呢? 我们有下面的定理.

记 n 阶矩阵 A 的前 $n-1$ 个顺序主子式为

$$\Delta_1 = a_{11}, \Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \dots, \quad \Delta_{n-1} = \begin{vmatrix} a_{11} & \cdots & a_{1, n-1} \\ \vdots & & \vdots \\ a_{n-1, 1} & \cdots & a_{n-1, n-1} \end{vmatrix}.$$

定理 3.1.1 当 $\Delta_1, \Delta_2, \dots, \Delta_{n-1}$ 都不为零时, 则 $a_{kk}^{(k)} \neq 0 (k=1, 2, \dots, n-1)$.

证明 利用数学归纳法. 显然阶数为 1 时结论成立. 假设阶数为 $k-1$ 时成立, 求证对 k 阶矩阵结论也成立. 由归纳法假设知 $a_i^{(i)} \neq 0 (i=1, 2, \dots, k-1)$, 于是根据前面消元过程有

$$L^{(k-1)} \cdots L^{(2)} L^{(1)} A = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & \cdots & a_{1n}^{(1)} \\ & a_{22}^{(2)} & \cdots & \cdots & a_{2n}^{(2)} \\ & & \ddots & & \vdots \\ & & & a_{kk}^{(k)} & \cdots & a_{kn}^{(k)} \\ & & & \vdots & & \vdots \\ & & & a_{nk}^{(k)} & \cdots & a_{nn}^{(k)} \end{pmatrix} = A^{(k)}.$$

由于每一步倍加初等变换不改变矩阵 A 的行列式的值, 所以从上面 $A^{(k)}$ 的形状容易看出, 矩阵 A 的前 $n-1$ 个顺序主子式应满足下列关系:

$$\begin{aligned} \Delta_1 &= a_{11} = a_{11}^{(1)}, \\ \Delta_2 &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \begin{vmatrix} a_{11}^{(1)} & a_{12}^{(1)} \\ 0 & a_{22}^{(2)} \end{vmatrix} = a_{11}^{(1)} a_{22}^{(2)}, \\ &\vdots \\ \Delta_k &= \begin{vmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{vmatrix} = \begin{vmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1k}^{(1)} \\ & a_{22}^{(2)} & \cdots & a_{2k}^{(2)} \\ & & \ddots & \vdots \\ & & & a_{kk}^{(k)} \end{vmatrix} = a_{11}^{(1)} a_{22}^{(2)} \cdots a_{kk}^{(k)}. \end{aligned} \quad (3.1.6)$$

已设 $\Delta_k \neq 0$ 及 $a_{ii}^{(i)} \neq 0 (i=1, 2, \dots, k-1)$, 根据式(3.1.6)便有 $a_{kk}^{(k)} \neq 0$, 即对 k 阶矩阵结论也成立. 证毕

定理结论与条件交换时也成立. 即如果 $a_{kk}^{(k)} \neq 0 (k=1, 2, \dots, n-1)$, 亦可推出 $\Delta_k \neq 0 (k=1, 2, \dots, n-1)$.

但要注意, 这里必须强调只有在前 $n-1$ 个顺序主子式 $\Delta_k \neq 0 (k=1, 2, \dots, n-1)$ 时, 才能保证前 $n-1$ 个主元素 $a_{kk}^{(k)} \neq 0 (k=1, 2, \dots, n-1)$. 读者可能会猜想, 似乎从式(3.1.6)有 $\Delta_{n-1} = a_{11}^{(1)} a_{22}^{(2)} \cdots a_{n-1, n-1}^{(n-1)}$, 因此只要 $\Delta_{n-1} \neq 0$ 就可以推得前 $n-1$ 个主元素 $a_{kk}^{(k)} \neq 0 (k=1, 2, \dots, n-1)$. 这个猜想是不正确的, 因为我们在用归纳法证明的过程中, 是在假定前 $k-1$ 个主元素 $a_{11}^{(1)}, a_{22}^{(2)}, \dots, a_{k-1, k-1}^{(k-1)}$ 都不为零的前提下才推得式(3.1.6), 忽视了这一点就

不一定正确了. 例如, 有三阶矩阵 $\begin{pmatrix} 1 & 1 & 2 \\ 2 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$, 虽然 $\Delta_3 = \begin{vmatrix} 1 & 1 & 2 \\ 2 & 2 & 3 \\ 1 & 2 & 3 \end{vmatrix} = 1 \neq 0$, 可是在消元过程

中(不交换两行的情况)仍然有主元素 $a_{22}^{(2)} = 0$.

综上所述, 可得到下面的定理.

定理 3.1.2 高斯消元过程能够进行到底的充分必要条件是 \mathbf{A} 的前 $n-1$ 个顺序主子式都不为零, 即

$$\Delta_k \neq 0, \quad k = 1, 2, \dots, n-1. \quad (3.1.7)$$

3.1.2 矩阵的三角分解

如前所述, 当条件(3.1.7)满足时, 由式(3.1.5)并令 $\mathbf{U} = \mathbf{A}^{(n)}$, 有

$$\mathbf{L}^{(n-1)} \cdots \mathbf{L}^{(2)} \mathbf{L}^{(1)} \mathbf{A} = \mathbf{U}.$$

容易看出 $\det \mathbf{L}^{(k)} = 1 \neq 0 (k=1, 2, \dots, n-1)$, 所以 $n-1$ 个消元矩阵 $\mathbf{L}^{(k)}$ 可逆, 于是有

$$\mathbf{A} = (\mathbf{L}^{(1)})^{-1} (\mathbf{L}^{(2)})^{-1} \cdots (\mathbf{L}^{(n-1)})^{-1} \mathbf{U}.$$

按照逆矩阵的定义知

$$(\mathbf{L}^{(k)})^{-1} = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & L_{k+1,k} & 1 & \\ & & \vdots & & \ddots \\ & & L_{nk} & & & 1 \end{pmatrix}, \quad k = 1, 2, \dots, n-1,$$

它们是下三角矩阵,显然它们的连乘积仍然是下三角矩阵.令

$$L = (L^{(1)})^{-1}(L^{(2)})^{-1}\cdots(L^{(n-1)})^{-1} = \begin{pmatrix} 1 & & & & & \\ l_{21} & 1 & & & & \\ l_{31} & l_{32} & 1 & & & \\ l_{41} & l_{42} & l_{43} & \ddots & & \\ \vdots & \vdots & \vdots & \ddots & \ddots & 1 \\ l_{n1} & l_{n2} & l_{n3} & \cdots & l_{n,n-1} & 1 \end{pmatrix},$$

这是一个对角元素都是 1 的下三角矩阵,称为单位下三角矩阵,则得

$$A = LU. \quad (3.1.8)$$

这样, A 就分解成一个单位下三角矩阵 L 与一个上三角矩阵 U 的乘积. 这种分解, 矩阵 A 与上三角矩阵 U 显然不相似, 如果把式(3.1.8)写成 $A = LUI$, 则由第 1 章知 A 与 U 呈相抵关系, 即 $A \simeq U$.

一般地, 有如下的定义.

定义 3.1.1 如果方阵 A 可分解成一个下三角矩阵 L 和一个上三角矩阵 U 的乘积, 则称 A 可作三角分解或 LU 分解. 如果 L 是单位下三角矩阵, U 为上三角矩阵, 此时的三角分解称为杜利特(Doolittle)分解; 若 L 是下三角矩阵, 而 U 是单位上三角矩阵, 则称三角分解为克劳特(Crout)分解.

下面研究方阵三角分解的存在和惟一性问题.

首先指出, 一个方阵的三角分解不是惟一的, 从上面定义看到, 杜利特分解与克劳特分解就是两种不同的三角分解. 其实, 方阵的三角分解有无穷多, 这是因为如果设 D 是行列式不为零的任意对角矩阵, 则

$$A = LU = LDD^{-1}U = (LD)(D^{-1}U) = \tilde{L}\tilde{U},$$

其中 \tilde{L}, \tilde{U} 也分别是下、上三角矩阵, 从而 $A = \tilde{L}\tilde{U}$ 也是 A 的一个三角分解. 因 D 的任意性, 所以三角分解不惟一. 但是我们可以借助下面的基本定理来回答一个方阵满足什么条件存在三角分解? 怎样才是惟一?

定理 3.1.3(LDU 基本定理) 设 A 为 n 阶方阵, 则 A 可以惟一地分解为

$$A = LDU \quad (3.1.9)$$

的充分必要条件是 A 的前 $n-1$ 个顺序主子式 $\Delta_k \neq 0 (k=1, 2, \dots, n-1)$. 其中 L, U 分别是单位下、上三角矩阵, D 是对角矩阵

$$D = \text{diag}(d_1, d_2, \dots, d_n),$$

$$d_k = \frac{\Delta_k}{\Delta_{k-1}}, \quad k = 1, 2, \dots, n, \quad \Delta_0 = 1.$$

* 证明 充分性: 若 $\Delta_k \neq 0 (k=1, 2, \dots, n-1)$, 则由定理 3.1.2 知高斯消元过程得以完成, 即实现了一个杜利特分解 $A = L\tilde{U}$, 其中 L 为单位下三角矩阵, \tilde{U} 为上三角矩阵, 记

$$\tilde{U} = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ & u_{22} & \cdots & u_{2n} \\ & & \ddots & \vdots \\ & & & u_{nn} \end{pmatrix} = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ & & \ddots & \vdots \\ & & & a_{nn}^{(n)} \end{pmatrix} = A^{(n)},$$

由定理 3.1.1 知, $u_{ii} = a_{ii}^{(i)} \neq 0 (i=1, 2, \dots, n-1)$.

下面分两种情况讨论:

(1) 若 \mathbf{A} 非奇异, 由式(3.1.6)有 $\Delta_n = a_{11}^{(1)} a_{22}^{(2)} \cdots a_{nn}^{(n)} = |\mathbf{A}| \neq 0$, 所以 $a_{nn}^{(n)} = u_{nn} \neq 0$, 这时令 $\mathbf{D} = \text{diag}(a_{11}^{(1)}, a_{22}^{(2)}, \dots, a_{nn}^{(n)})$, 则

$$\mathbf{D}^{-1} = \text{diag}\left(\frac{1}{a_{11}^{(1)}}, \frac{1}{a_{22}^{(2)}}, \dots, \frac{1}{a_{nn}^{(n)}}\right).$$

于是有

$$\mathbf{A} = \mathbf{L}\tilde{\mathbf{U}} = \mathbf{L}\mathbf{D}(\mathbf{D}^{-1}\tilde{\mathbf{U}}) = \mathbf{L}\mathbf{D}\mathbf{U} \quad (3.1.10)$$

是 \mathbf{A} 的一个 \mathbf{LDU} 分解.

(2) 若 \mathbf{A} 奇异, 则 $a_{nn}^{(n)} = u_{nn} = 0$, 此时令 $\mathbf{D} = \text{diag}(a_{11}^{(1)}, \dots, a_{n-1, n-1}^{(n-1)}, 0)$, $\mathbf{D}_{n-1} = \text{diag}(a_{11}^{(1)}, \dots, a_{n-1, n-1}^{(n-1)})$, $\boldsymbol{\alpha} = (u_{1n}, \dots, u_{n-1, n})^T$, 则

$$\tilde{\mathbf{U}} \equiv \begin{bmatrix} \tilde{\mathbf{U}}_{n-1} & \boldsymbol{\alpha} \\ \mathbf{0}^T & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{D}_{n-1} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{D}_{n-1}^{-1} \tilde{\mathbf{U}}_{n-1} & \mathbf{D}_{n-1}^{-1} \boldsymbol{\alpha} \\ \mathbf{0}^T & 1 \end{bmatrix} = \mathbf{D}\mathbf{U},$$

因此不论哪种情况, 只要 $\Delta_k \neq 0 (k=1, 2, \dots, n-1)$ 时, 总存在一个 \mathbf{LDU} 分解式(3.1.9), 且结合式(3.1.6)容易知道, $d_k = a_{kk}^{(k)} = \frac{\Delta_k}{\Delta_{k-1}} (k=1, 2, \dots, n), \Delta_0 = 1$.

再证这个分解是惟一的, 仍分两种情况讨论.

(1) 当 \mathbf{A} 非奇异时, 有 $|\mathbf{A}| = |\mathbf{L}| |\mathbf{D}| |\mathbf{U}| \neq 0$, 所以 $\mathbf{L}, \mathbf{D}, \mathbf{U}$ 皆非奇异. 若还存在另一个 \mathbf{LDU} 分解 $\mathbf{A} = \mathbf{L}_1 \mathbf{D}_1 \mathbf{U}_1$, 这里 $\mathbf{L}_1, \mathbf{D}_1, \mathbf{U}_1$ 也非奇异, 于是有

$$\mathbf{L}\mathbf{D}\mathbf{U} = \mathbf{L}_1 \mathbf{D}_1 \mathbf{U}_1, \quad (3.1.11)$$

上式两端左乘以 \mathbf{L}_1^{-1} 以及右乘以 \mathbf{U}^{-1} 和 \mathbf{D}^{-1} , 得

$$\mathbf{L}_1^{-1} \mathbf{L} = \mathbf{D}_1 \mathbf{U}_1 \mathbf{U}^{-1} \mathbf{D}^{-1}, \quad (3.1.12)$$

但式(3.1.12)左端是单位下三角阵, 右端是单位上三角阵, 所以都应该是单位阵, 因此

$$\mathbf{L}_1^{-1} \mathbf{L} = \mathbf{I}, \quad \mathbf{D}_1 \mathbf{U}_1 \mathbf{U}^{-1} \mathbf{D}^{-1} = \mathbf{I},$$

即

$$\mathbf{L}_1 = \mathbf{L}, \quad \mathbf{U}_1 \mathbf{U}^{-1} = \mathbf{D}_1^{-1} \mathbf{D}.$$

由后一个等式类似地可得

$$\mathbf{U}_1 \mathbf{U}^{-1} = \mathbf{I}, \quad \mathbf{D}_1^{-1} \mathbf{D} = \mathbf{I},$$

即有

$$\mathbf{U}_1 = \mathbf{U}, \quad \mathbf{D}_1 = \mathbf{D}.$$

(2) 若 \mathbf{A} 奇异, 则式(3.1.11)可写成分块形式

$$\begin{bmatrix} \tilde{\mathbf{L}}_1 & \mathbf{0} \\ \boldsymbol{\beta}_1^T & 1 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{D}}_1 & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{U}}_1 & \boldsymbol{\alpha}_1 \\ \mathbf{0}^T & 1 \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{L}} & \mathbf{0} \\ \boldsymbol{\beta}^T & 1 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{D}} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{U}} & \boldsymbol{\alpha} \\ \mathbf{0}^T & 1 \end{bmatrix},$$

其中 $\tilde{\mathbf{L}}, \tilde{\mathbf{L}}_1$ 是 $n-1$ 阶单位下三角阵; $\tilde{\mathbf{U}}, \tilde{\mathbf{U}}_1$ 是 $n-1$ 阶单位上三角阵; $\tilde{\mathbf{D}}, \tilde{\mathbf{D}}_1$ 是 $n-1$ 阶对角阵; $\boldsymbol{\alpha}, \boldsymbol{\alpha}_1, \boldsymbol{\beta}, \boldsymbol{\beta}_1$ 是 $n-1$ 维列向量.

由此得出

$$\begin{bmatrix} \tilde{\mathbf{L}}_1 \tilde{\mathbf{D}}_1 \tilde{\mathbf{U}}_1 & \tilde{\mathbf{L}}_1 \tilde{\mathbf{D}}_1 \boldsymbol{\alpha}_1 \\ \boldsymbol{\beta}_1^T \tilde{\mathbf{D}}_1 \tilde{\mathbf{U}}_1 & \boldsymbol{\beta}_1^T \tilde{\mathbf{D}}_1 \boldsymbol{\alpha}_1 \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{L}} \tilde{\mathbf{D}} \tilde{\mathbf{U}} & \tilde{\mathbf{L}} \tilde{\mathbf{D}} \boldsymbol{\alpha} \\ \boldsymbol{\beta}^T \tilde{\mathbf{D}} \tilde{\mathbf{U}} & \boldsymbol{\beta}^T \tilde{\mathbf{D}} \boldsymbol{\alpha} \end{bmatrix},$$

其中 $\tilde{\mathbf{L}}_1, \tilde{\mathbf{D}}_1, \tilde{\mathbf{U}}_1$ 和 $\tilde{\mathbf{L}}, \tilde{\mathbf{D}}, \tilde{\mathbf{U}}$ 皆非奇异, 类似于前面的推理, 可得

$$\begin{aligned} \tilde{\mathbf{L}}_1 &= \tilde{\mathbf{L}}, & \tilde{\mathbf{D}}_1 &= \tilde{\mathbf{D}}, & \tilde{\mathbf{U}}_1 &= \tilde{\mathbf{U}}, \\ \boldsymbol{\alpha}_1 &= \boldsymbol{\alpha}, & \boldsymbol{\beta}_1^T &= \boldsymbol{\beta}^T. \end{aligned}$$

必要性: 假定 \mathbf{A} 有一个惟一的 \mathbf{LDU} 分解, 写成分块的形式便是

$$\begin{pmatrix} \mathbf{A}_{n-1} & \mathbf{x} \\ \mathbf{y}^T & a_{nn} \end{pmatrix} = \begin{pmatrix} \mathbf{L}_{n-1} & \mathbf{0} \\ \boldsymbol{\beta}^T & 1 \end{pmatrix} \begin{pmatrix} \mathbf{D}_{n-1} & \mathbf{0} \\ \mathbf{0} & d_n \end{pmatrix} \begin{pmatrix} \mathbf{U}_{n-1} & \boldsymbol{\alpha} \\ \mathbf{0}^T & 1 \end{pmatrix}, \quad (3.1.13)$$

其中 $\mathbf{L}_{n-1}, \mathbf{D}_{n-1}, \mathbf{U}_{n-1}, \mathbf{A}_{n-1}$ 分别是 $\mathbf{L}, \mathbf{D}, \mathbf{U}, \mathbf{A}$ 的 $n-1$ 阶顺序主子矩阵; $\mathbf{x}, \mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}$ 为 $n-1$ 维列向量. 由式(3.1.13)有下面的矩阵方程:

$$\mathbf{A}_{n-1} = \mathbf{L}_{n-1} \mathbf{D}_{n-1} \mathbf{U}_{n-1}, \quad (3.1.14)$$

$$\mathbf{y}^T = \boldsymbol{\beta}^T \mathbf{D}_{n-1} \mathbf{U}_{n-1}, \quad (3.1.15)$$

$$\mathbf{x} = \mathbf{L}_{n-1} \mathbf{D}_{n-1} \boldsymbol{\alpha}, \quad (3.1.16)$$

$$a_{nn} = \boldsymbol{\beta}^T \mathbf{D}_{n-1} \boldsymbol{\alpha} + d_n. \quad (3.1.17)$$

不然, 若 $\Delta_{n-1} = |\mathbf{A}_{n-1}| = 0$, 则由式(3.1.14)有

$$|\mathbf{A}_{n-1}| = |\mathbf{L}_{n-1}| |\mathbf{D}_{n-1}| |\mathbf{U}_{n-1}| = |\mathbf{D}_{n-1}| = 0.$$

于是有 $|\mathbf{L}_{n-1} \mathbf{D}_{n-1}| = |\mathbf{D}_{n-1}| = 0$, 即 $\mathbf{L}_{n-1} \mathbf{D}_{n-1}$ 奇异. 那么, 对于非齐次线性方程组(3.1.16)有无穷多非零解, 不妨设有 $\boldsymbol{\alpha}'$, 使 $\mathbf{L}_{n-1} \mathbf{D}_{n-1} \boldsymbol{\alpha}' = \mathbf{x}$, 而 $\boldsymbol{\alpha}' \neq \boldsymbol{\alpha}$. 同理, 因 $\mathbf{D}_{n-1} \mathbf{U}_{n-1}$ 奇异, 故 $(\mathbf{D}_{n-1} \mathbf{U}_{n-1})^T = \mathbf{U}_{n-1}^T \mathbf{D}_{n-1}^T$ 也奇异, 就有 $\boldsymbol{\beta}' \neq \boldsymbol{\beta}$, 使 $\mathbf{U}_{n-1}^T \mathbf{D}_{n-1}^T \boldsymbol{\beta}' = \mathbf{y}$, 或 $\boldsymbol{\beta}'^T \mathbf{D}_{n-1} \mathbf{U}_{n-1} = \mathbf{y}^T$. 取 $d'_n = a_{nn} - \boldsymbol{\beta}'^T \mathbf{D}_{n-1} \boldsymbol{\alpha}'$, 则有

$$\begin{pmatrix} \mathbf{A}_{n-1} & \mathbf{x} \\ \mathbf{y}^T & a_{nn} \end{pmatrix} = \begin{pmatrix} \mathbf{L}_{n-1} & \mathbf{0} \\ \boldsymbol{\beta}'^T & 1 \end{pmatrix} \begin{pmatrix} \mathbf{D}_{n-1} & \mathbf{0} \\ \mathbf{0} & d'_n \end{pmatrix} \begin{pmatrix} \mathbf{U}_{n-1} & \boldsymbol{\alpha}' \\ \mathbf{0}^T & 1 \end{pmatrix},$$

这与 \mathbf{A} 的 \mathbf{LDU} 分解的惟一性矛盾, 因此 $\Delta_{n-1} \neq 0$.

考察 $n-1$ 阶顺序主子矩阵 \mathbf{A}_{n-1} 由式(3.1.14)写成分块形式, 同样有 $\mathbf{A}_{n-2} = \mathbf{L}_{n-2} \mathbf{D}_{n-2} \mathbf{U}_{n-2}$. 由于 $|\mathbf{D}_{n-1}| \neq 0$, 所以 $|\mathbf{D}_{n-2}| \neq 0$, 可得

$$|\mathbf{A}_{n-2}| = |\mathbf{L}_{n-2}| |\mathbf{D}_{n-2}| |\mathbf{U}_{n-2}| = |\mathbf{D}_{n-2}| \neq 0,$$

从而 $\Delta_{n-2} \neq 0$. 以此类推可得 $\Delta_k \neq 0 (k=1, 2, \dots, n-1)$. 证毕

有了 \mathbf{LDU} 分解的存在和惟一性定理 3.1.3, 只要将对角矩阵 \mathbf{D} 与 \mathbf{L} 或 \mathbf{U} 进行不同的搭配, 我们就可以很方便地得出杜利特分解及克劳特分解的存在和惟一性定理.

推论 1 设 \mathbf{A} 是 n 阶方阵, 则 \mathbf{A} 可以惟一地进行杜利特分解的充分必要条件是 \mathbf{A} 的前 $n-1$ 个顺序主子式

$$\Delta_k = \begin{vmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{vmatrix} \neq 0, \quad k = 1, 2, \dots, n-1, \quad (3.1.18)$$

其中 \mathbf{L} 为单位下三角矩阵, $\tilde{\mathbf{U}}$ 是上三角矩阵, 即有

$$\mathbf{A} = \begin{pmatrix} 1 & & & & \\ l_{21} & 1 & & & \\ l_{31} & l_{32} & \ddots & & \\ \vdots & \vdots & \ddots & 1 & \\ l_{n1} & l_{n2} & \cdots & l_{n,n-1} & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ & u_{22} & \cdots & u_{2n} \\ & & \ddots & \vdots \\ & & & u_{nn} \end{pmatrix}, \quad (3.1.19)$$

并且若 \mathbf{A} 为奇异矩阵, 则 $u_{nn} = 0$; 若 \mathbf{A} 为非奇异矩阵, 则充要条件(3.1.18)可换为: \mathbf{A} 的各阶顺序主子式全不为零, 即:

$$\Delta_k \neq 0, \quad k = 1, 2, \dots, n. \quad (3.1.20)$$

推论 2 n 阶方阵 \mathbf{A} 可唯一地进行克劳特分解

$$\mathbf{A} = \tilde{\mathbf{L}}\mathbf{U} = \begin{pmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{pmatrix} \begin{pmatrix} 1 & u_{12} & \cdots & u_{1n} \\ & 1 & \cdots & u_{2n} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \quad (3.1.21)$$

的充要条件仍为式(3.1.18). 若 \mathbf{A} 为奇异矩阵, 则 $l_{nn}=0$; 若 \mathbf{A} 为非奇异矩阵, 则充要条件也可换为式(3.1.20).

由此可见, 在 \mathbf{A} 的三角分解中, 只要有一个三角矩阵是单位三角矩阵, 则分解总是唯一的; 否则, 若 \mathbf{L} 与 \mathbf{U} 两个都不是单位三角矩阵, 那么分解不惟一.

例 3.1.1 求矩阵

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

的 \mathbf{LDU} 分解.

解 因为 $\Delta_1=2, \Delta_2=5, \Delta_3=0$, 所以 \mathbf{A} 有惟一的 \mathbf{LDU} 分解. 下面我们仿照高斯消元过程的计算步骤来得到 \mathbf{A} 的 \mathbf{LDU} 分解. 由式(3.1.2)有消元阵

$$\mathbf{L}^{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \quad (\mathbf{L}^{(1)})^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix},$$

所以得

$$\mathbf{L}^{(1)}\mathbf{A} = \begin{pmatrix} 2 & -1 & 3 \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 5 & -1 \end{pmatrix} = \mathbf{A}^{(2)}.$$

再由 $\mathbf{A}^{(2)}$ 计算消元阵

$$\mathbf{L}^{(2)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix}, \quad (\mathbf{L}^{(2)})^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix},$$

得

$$\mathbf{L}^{(2)}\mathbf{A}^{(2)} = \begin{pmatrix} 2 & -1 & 3 \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & \frac{5}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & -\frac{1}{5} \\ 0 & 0 & 1 \end{pmatrix} = \mathbf{A}^{(3)},$$

即

$$\mathbf{L}^{(2)}\mathbf{L}^{(1)}\mathbf{A} = \mathbf{A}^{(3)}.$$

故

$$\begin{aligned} \mathbf{A} &= (\mathbf{L}^{(1)})^{-1} (\mathbf{L}^{(2)})^{-1} \mathbf{A}^{(3)} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & \frac{5}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & -\frac{1}{5} \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & \frac{5}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & -\frac{1}{5} \\ 0 & 0 & 1 \end{pmatrix} = \mathbf{LDU}. \end{aligned}$$

例 3.1.2 (三角分解未必存在) 可逆矩阵 $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ 不存在三角分解.

这里要指出, 矩阵 \mathbf{A} 的任何一种三角分解都需要假定 \mathbf{A} 的前 $n-1$ 个顺序主子式非零, 这等价于高斯消元过程中要求每一步的主元素不为零. 如果这个条件不满足, 可以考虑适当地交换 \mathbf{A} 的两行(不再是矩阵 \mathbf{A}), 直至满足三角分解条件为止. 这样处理对于解线性方程组来说, 相当于调整方程的顺序, 对解不产生影响.

3.1.3 常用的三角分解公式

在实际应用中, 如果矩阵 \mathbf{A} 的阶数 n 很高, 那么, 按例 3.1.1 的消元步骤来得出 \mathbf{A} 的三角分解是相当麻烦的. 下面我们将分别根据 \mathbf{A} 的不对称和对称的情况, 介绍两个常用的直接三角分解公式, 先给出 \mathbf{A} 在不对称情况下的克劳特分解计算公式.

1. 克劳特分解

设 \mathbf{A} 为 n 阶方阵(但不一定对称), 且有分解式

$$\mathbf{A} = \mathbf{LU},$$

即

$$\begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} l_{11} & & & & \\ \vdots & \ddots & & & \\ l_{i1} & \cdots & l_{ii} & & \\ \vdots & & & \ddots & \\ l_{n1} & \cdots & \cdots & \cdots & l_{nn} \end{pmatrix} \begin{pmatrix} 1 & u_{12} & \cdots & u_{1j} & \cdots & u_{1n} \\ & \ddots & & \vdots & & \vdots \\ & & 1 & u_{j-1,j} & \cdots & u_{j-1,n} \\ & & & \ddots & & \vdots \\ & & & & 1 & 1 \end{pmatrix}.$$

下面给出下三角矩阵 \mathbf{L} 的元素 l_{ij} 及单位上三角矩阵 \mathbf{U} 的元素 u_{ij} 的实际求法. 我们考察 \mathbf{A} 的第 i 行第 j 列的元素:

$$a_{ij} = (l_{i1}, \cdots, l_{ii}, 0, \cdots, 0) \begin{pmatrix} u_{1j} \\ \vdots \\ u_{j-1,j} \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

当 $i \geq j$ 时(表示下三角位置),有

$$a_{ij} = \sum_{k=1}^j l_{ik} u_{kj} = \sum_{k=1}^{j-1} l_{ik} u_{kj} + l_{ij},$$

得

$$l_{ij} = a_{ij} - \sum_{k=1}^{j-1} l_{ik} u_{kj}, \quad i = 1, \dots, n, \quad j = 1, \dots, i; \quad (3.1.22)$$

当 $i < j$ 时(表示上三角位置),有

$$a_{ij} = \sum_{k=1}^i l_{ik} u_{kj} = \sum_{k=1}^{i-1} l_{ik} u_{kj} + l_{ii} u_{ij},$$

得

$$u_{ij} = (a_{ij} - \sum_{k=1}^i l_{ik} u_{kj}) / l_{ii}, \quad i = 1, \dots, n-1, \quad j = i+1, \dots, n. \quad (3.1.23)$$

式(3.1.22)与式(3.1.23)称为克劳特分解公式.

从表面上看,以上分解公式是非线性的,似乎觉得用它们来求解 L 和 U 的元素有些困难,其实不然,只要我们注意“自上而下且先左后右”一行行地求解,便能顺利地求得 L 和 U 的全部元素.

矩阵 A 的 LU 分解计算步骤如下所示.

第 1 步: 计算 $l_{11} \quad u_{12} \quad u_{13} \quad \dots \quad u_{1n}$;

第 2 步: 计算 $l_{21} \quad l_{22} \quad u_{23} \quad \dots \quad u_{2n}$;

⋮

第 n 步: 计算 $l_{n1} \quad l_{n2} \quad l_{n3} \quad \dots \quad l_{nn}$.

这种“按行分解”的方式有利于大型稀疏矩阵的分块分解.实践证明此方法是有效的.

例 3.1.3 试将下列矩阵进行克劳特分解:

$$A = \begin{pmatrix} 4 & 8 & 4 \\ 2 & 7 & 2 \\ 1 & 2 & 3 \end{pmatrix}.$$

解 首先我们约定,当上限小于下限时,和号“ \sum ”取 0,根据式(3.1.22)和式(3.1.23)有

$$\begin{aligned} l_{11} &= a_{11} = 4, \\ u_{12} &= a_{12} / l_{11} = 2, \\ u_{13} &= a_{13} / l_{11} = 1, \\ l_{21} &= a_{21} = 2, \\ l_{22} &= a_{22} - l_{21} u_{12} = 3, \\ u_{23} &= (a_{23} - l_{21} u_{13}) / l_{22} = 0, \\ l_{31} &= a_{31} = 1, \\ l_{32} &= a_{32} - l_{31} u_{12} = 0, \\ l_{33} &= a_{33} - l_{31} u_{13} - l_{32} u_{23} = 2, \end{aligned}$$

则得

$$\mathbf{A} = \mathbf{LU} = \begin{pmatrix} 4 & & \\ 2 & 3 & \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ & 1 & 0 \\ & & 1 \end{pmatrix}.$$

类似地,可得到 n 阶方阵 \mathbf{A} 的杜利特分解的计算公式:

$$\begin{cases} u_{ij} = a_{ij} - \sum_{k=1}^{i-1} l_{ik}u_{kj}, & i = 1, \dots, n, \quad j = i, \dots, n, \\ l_{ij} = (a_{ij} - \sum_{k=1}^{j-1} l_{ik}u_{kj})/u_{jj}, & i = 2, \dots, n, \quad j = 1, \dots, i-1. \end{cases} \quad (3.1.24)$$

2. 乔累斯基(Cholesky)分解

如果 \mathbf{A} 是对称正定矩阵,则可以使三角分解的计算量大为减少,大约是前述的克劳特分解或杜利特分解工作量的一半.

定理 3.1.4 设 \mathbf{A} 为 n 阶对称正定矩阵,则存在一个实的非奇异下三角矩阵 \mathbf{L} , 使

$$\mathbf{A} = \mathbf{LL}^T. \quad (3.1.25)$$

如果限定 \mathbf{L} 的对角元素为正时,这种分解是惟一的.

证明 因为 \mathbf{A} 是对称正定的,所以由线性代数知,它的各阶顺序主子式 $\Delta_k > 0 (k=1, 2, \dots, n)$. 再由定理 3.1.3 知, \mathbf{A} 可惟一地分解为

$$\mathbf{A} = \tilde{\mathbf{L}}\mathbf{D}\tilde{\mathbf{U}}, \quad (3.1.26)$$

其中 $\tilde{\mathbf{L}}, \tilde{\mathbf{U}}$ 是单位下三角阵和单位上三角阵, \mathbf{D} 是对角阵, 记

$$\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_n).$$

又因为 $\mathbf{A}^T = (\tilde{\mathbf{L}}\mathbf{D}\tilde{\mathbf{U}})^T = \tilde{\mathbf{U}}^T\mathbf{D}\tilde{\mathbf{L}}^T$, 根据 $\mathbf{A} = \mathbf{A}^T$ 所以有

$$\tilde{\mathbf{L}}\mathbf{D}\tilde{\mathbf{U}} = \tilde{\mathbf{U}}^T\mathbf{D}\tilde{\mathbf{L}}^T.$$

由于式(3.1.26)的分解是惟一的,从而得

$$\tilde{\mathbf{L}} = \tilde{\mathbf{U}}^T, \tilde{\mathbf{U}} = \tilde{\mathbf{L}}^T,$$

于是由式(3.1.26)有

$$\mathbf{A} = \tilde{\mathbf{U}}^T\mathbf{D}\tilde{\mathbf{U}}. \quad (3.1.27)$$

因为 $|\tilde{\mathbf{U}}| = 1 \neq 0$, 故 $\tilde{\mathbf{U}}$ 可逆, 所以由式(3.1.27)有

$$\mathbf{D} = (\tilde{\mathbf{U}}^T)^{-1}\mathbf{A}\tilde{\mathbf{U}}^{-1} = (\tilde{\mathbf{U}}^{-1})^T\mathbf{A}\tilde{\mathbf{U}}^{-1}.$$

此式说明矩阵 \mathbf{A} 与 \mathbf{D} 呈相合关系, 而 \mathbf{A} 又为对称正定矩阵, 则 \mathbf{D} 也为对称正定矩阵(事实上, 设任取 $\mathbf{x} \neq \mathbf{0}$, 显然 $\tilde{\mathbf{U}}^{-1}\mathbf{x} \neq \mathbf{0}$, 二次型 $\mathbf{x}^T\mathbf{D}\mathbf{x} = \mathbf{x}^T(\tilde{\mathbf{U}}^{-1})^T\mathbf{A}\tilde{\mathbf{U}}^{-1}\mathbf{x} = (\tilde{\mathbf{U}}^{-1}\mathbf{x})^T\mathbf{A}(\tilde{\mathbf{U}}^{-1}\mathbf{x}) > 0$. 故 \mathbf{D} 为对称正定矩阵). 既然对角矩阵 \mathbf{D} 对称正定, 那么它的所有一阶主子式(对角元素) $d_i > 0 (i=1, 2, \dots, n)$, 所以 $\sqrt{d_i}$ 有意义. 令

$$\mathbf{D}^{\frac{1}{2}} = \text{diag}(\sqrt{d_1}, \sqrt{d_2}, \dots, \sqrt{d_n}),$$

则有惟一的表达式

$$\mathbf{A} = \tilde{\mathbf{L}}\mathbf{D}\tilde{\mathbf{U}} = \tilde{\mathbf{L}}\mathbf{D}^{\frac{1}{2}}\mathbf{D}^{\frac{1}{2}}\tilde{\mathbf{L}}^T = (\tilde{\mathbf{L}}\mathbf{D}^{\frac{1}{2}})(\tilde{\mathbf{L}}\mathbf{D}^{\frac{1}{2}})^T = \mathbf{LL}^T,$$

其中 $\mathbf{L} = \tilde{\mathbf{L}}\mathbf{D}^{\frac{1}{2}}$ 是对角线元素全为正数 $\sqrt{d_1}, \dots, \sqrt{d_n}$ 的下三角矩阵.

证毕

定义 3.1.2 式(3.1.25)称为是对称正定矩阵 \mathbf{A} 的乔累斯基分解,亦称为平方根分解. 下面我们给出下三角矩阵 \mathbf{L} 的实际求法.

$$\begin{pmatrix} a_{11} & & & & \\ \vdots & \ddots & & & \\ a_{i1} & \cdots & a_{ii} & & \\ \vdots & & \vdots & \ddots & \\ a_{n1} & \cdots & a_{ni} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} l_{11} & & & & \\ \vdots & \ddots & & & \\ l_{i1} & \cdots & l_{ii} & & \\ \vdots & & & \ddots & \\ l_{n1} & \cdots & l_{ni} & \cdots & l_{nn} \end{pmatrix} \begin{pmatrix} l_{11} & \cdots & l_{j1} & \cdots & l_{n1} \\ & \ddots & \vdots & & \vdots \\ & & l_{jj} & \cdots & l_{nj} \\ & & & \ddots & \vdots \\ & & & & l_{nn} \end{pmatrix},$$

由于 \mathbf{A} 对称,所以只要考虑下三角元素,即当 $i \geq j$ 时,有

$$a_{ij} = \sum_{k=1}^j l_{ik} l_{jk} = \sum_{k=1}^{j-1} l_{ik} l_{jk} + l_{ij} l_{jj},$$

即

$$l_{ij} = \left(a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk} \right) / l_{jj}, \quad i \geq j. \quad (3.1.28)$$

特别地,当 $i=j$ 时,有

$$l_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2}, \quad (3.1.29)$$

这里与克劳特三角分解不同的是矩阵 $\mathbf{U} = \mathbf{L}^T$, 在求得 \mathbf{L} 的第 r 行元素之后, \mathbf{L}^T 的第 r 列元素即已得出,所以其计算量将为克劳特分解的一半左右. 同时,由式(3.1.29)有

$$a_{ii} = \sum_{k=1}^i l_{ik}^2, \quad i = 1, 2, \dots, n.$$

所以

$$l_{ik}^2 \leq a_{ii} \leq \max_{1 \leq i \leq n} \{a_{ii}\},$$

于是

$$\max_{i,k} \{l_{ik}^2\} \leq \max_{1 \leq i \leq n} \{a_{ii}\}.$$

以上分析说明,分解过程中元素 l_{ik} 的数量级完全得到控制,从而计算过程是稳定的.

然而,这种 \mathbf{LL}^T 分解也还是有很大的缺陷. 因为它要完成 n 个开方运算,而绝大多数计算机上的开方运算是用子程序实现的(转化成对数计算),这样不但增加了运算量,而且还可能扩大误差,甚至有平方根号下出现负数的危险. 为了避免前述平方根运算,只要在上面式(3.1.25)中 \mathbf{L} 与 \mathbf{L}^T 之间插入一个特殊的对角矩阵 \mathbf{D} ,便能收到预定的效果.

定理 3.1.5 设 \mathbf{A} 为 n 阶对称正定矩阵,则 \mathbf{A} 可惟一地分解为

$$\mathbf{A} = \mathbf{LDL}^T, \quad (3.1.30)$$

其中 \mathbf{L} 为下三角矩阵; \mathbf{D} 为对角矩阵,且对角线元素是 \mathbf{L} 对角线元素的倒数. 即

于是有

$$A = LDL^T$$

证毕

下面我们给出下三角矩阵 L 的实际求法. 由 $A = LDL^T$, 即

$$\begin{pmatrix} a_{11} & & & & \\ \vdots & \ddots & & & \\ a_{i1} & \cdots & a_{ii} & & \\ \vdots & & \vdots & \ddots & \\ a_{n1} & \cdots & a_{ni} & \cdots & a_{nn} \end{pmatrix} \begin{matrix} \text{对称} \\ \\ \\ \\ \end{matrix} = \begin{pmatrix} l_{11} & & & & \\ \vdots & \ddots & & & \\ l_{i1} & \cdots & l_{ii} & & \\ \vdots & & \vdots & \ddots & \\ l_{n1} & \cdots & l_{ni} & \cdots & l_{nn} \end{pmatrix} \begin{pmatrix} 1 & \cdots & \frac{l_{j1}}{l_{11}} & \cdots & \frac{l_{n1}}{l_{11}} \\ & \ddots & \vdots & & \vdots \\ & & 1 & \cdots & \frac{l_{nj}}{l_{jj}} \\ & & & \ddots & \vdots \\ & & & & 1 \end{pmatrix},$$

比较两端, 当 $i \geq j$ 时(下三角元素), 有

$$a_{ij} = \sum_{k=1}^j l_{ik} (l_{jk} / l_{kk}) = \sum_{k=1}^{j-1} l_{ik} l_{jk} / l_{kk} + l_{ij},$$

故得

$$l_{ij} = a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk} / l_{kk}, \quad i = 1, \cdots, n, \quad j = 1, \cdots, i. \quad (3.1.31)$$

定义 3.1.3 式(3.1.30)称为是对称正定矩阵 A 的不带平方根乔累斯基分解, 亦称乔累斯基分解的变形.

从式(3.1.31)看出, 当 $i=j$ 时, 计算对角线元素

$$l_{ii} = a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2 / l_{kk}$$

就不再出现开方运算, 弥补了平方根分解(3.1.25)的缺陷. 因此, 它是求解对称正定线性方程组最常用的一个分解公式.

3.2 矩阵的 QR(正交三角)分解

在上节中, 已用初等变换所对应的初等矩阵(具体地说, 是倍加阵的连乘积)研究了矩阵的三角化问题, 相伴产生的是矩阵的 LU 三角分解. 这种三角分解对数值代数算法的发展起了重要的作用. 然而, 以初等变换为工具的 LU 分解方法并不能消除病态线性方程组不稳定问题, 而且有时候对于可逆矩阵也可能不存在三角分解, 因此需要寻找其他类似的矩阵分解. 20 世纪 60 年代以后, 人们又以正交(酉)变换为工具, 导出了矩阵的正交三角分解方法, 矩阵的正交三角分解即是一种对任何可逆矩阵均存在的理想分解, 从而对数值代数理论的近代发展作出了杰出贡献.

3.2.1 QR 分解的概念

定义 3.2.1 如果实(复)非奇异矩阵 A 能化成正交(酉)矩阵 Q 与实(复)非奇异上三角矩阵 R 的乘积, 即

$$A = QR, \quad (3.2.1)$$

则称式(3.2.1)是 A 的 QR 分解.

为简单起见,我们先不妨以实矩阵为对象研究 QR 分解的存在和惟一性以及它的构造方法,同时也给出复矩阵类似的结论.

定理 3.2.1 任何实的非奇异 n 阶矩阵 A 可以分解成正交矩阵 Q 和上三角矩阵 R 的乘积,且除去相差一个对角线元素之绝对值全等于 1 的对角矩阵因子 D 外,分解式(3.2.1)是惟一的.

证明 设 A 的各列向量依次为 $\alpha_1, \alpha_2, \dots, \alpha_n$, 由于 A 非奇异,所以 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关. 将它们按照施密特正交化法正交化,得到 n 个标准正交的向量 $\beta_1, \beta_2, \dots, \beta_n$, 且

$$\begin{cases} \beta_1 = b_{11}\alpha_1, \\ \beta_2 = b_{12}\alpha_1 + b_{22}\alpha_2, \\ \vdots \\ \beta_n = b_{1n}\alpha_1 + b_{2n}\alpha_2 + \dots + b_{nn}\alpha_n, \end{cases}$$

这里 b_{ij} 都是常数,且由变化过程知 $b_{ii} \neq 0 (i=1, 2, \dots, n)$. 写成矩阵形式有

$$(\beta_1, \beta_2, \dots, \beta_n) = (\alpha_1, \alpha_2, \dots, \alpha_n)B,$$

即

$$Q = AB,$$

其中

$$B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ & b_{22} & \cdots & b_{2n} \\ & & \ddots & \vdots \\ & & & b_{nn} \end{pmatrix}$$

是上三角矩阵($b_{ii} \neq 0 (i=1, 2, \dots, n)$). 显然, B 可逆,而且 $B^{-1} = R$ 也是上三角矩阵; 由于 Q 的各列标准正交,所以 Q 为正交矩阵. 从而有 $A = QR$.

为了证明惟一性,设 A 有两种形如(3.2.1)的分解式:

$$A = QR = Q_1 R_1, \quad (3.2.2)$$

其中 Q 和 Q_1 都是正交矩阵, R 和 R_1 都是非奇异上三角矩阵,由式(3.2.2)得

$$Q = Q_1 R_1 R^{-1} = Q_1 D,$$

式中 $D = R_1 R^{-1}$ 仍为实非奇异上三角矩阵,于是

$$I = Q^T Q = (Q_1 D)^T (Q_1 D) = D^T D. \quad (3.2.3)$$

设

$$D = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ & d_{22} & \cdots & d_{2n} \\ & & \ddots & \vdots \\ & & & d_{nn} \end{pmatrix},$$

代入式(3.2.3)并与单位矩阵相比较,得

$$d_{11}^2 = 1, \quad d_{12} = \cdots = d_{1n} = 0,$$

$$\begin{aligned} d_{22}^2 &= 1, & d_{23} &= \cdots = d_{2n} = 0, \\ &\vdots \\ d_{nn}^2 &= 1, \end{aligned}$$

从而有 $|d_{11}| = |d_{22}| = \cdots = |d_{nn}| = 1$, 即

$$\mathbf{D} = \begin{pmatrix} \pm 1 & 0 & \cdots & 0 \\ 0 & \pm 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \pm 1 \end{pmatrix},$$

这表明 \mathbf{D} 不仅是正交矩阵, 而且还是对角线元素的绝对值全为 1 的对角矩阵, 再由式 (3.2.2) 不难得

$$\mathbf{R}_1 = \mathbf{D}\mathbf{R}, \quad \mathbf{Q}_1 = \mathbf{Q}\mathbf{D}^{-1}. \quad \text{证毕}$$

显然, 当规定上三角阵 \mathbf{R} 和 \mathbf{R}_1 对角线上的元素为正实数时, 则 $\mathbf{D} = \mathbf{I}$, 从而 \mathbf{QR} 分解惟一.

例 3.2.1 设

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

写出 \mathbf{A} 的 \mathbf{QR} 分解.

解 令 $\boldsymbol{\alpha}_1 = (1, 1, 0)^T$, $\boldsymbol{\alpha}_2 = (1, -1, 0)^T$, $\boldsymbol{\alpha}_3 = (0, 1, 2)^T$, 由施密特方法得

$$\begin{aligned} \boldsymbol{\beta}'_1 &= \boldsymbol{\alpha}_1 = (1, 1, 0)^T, \\ \boldsymbol{\beta}'_2 &= \boldsymbol{\alpha}_2 - \frac{(\boldsymbol{\beta}'_1, \boldsymbol{\alpha}_2)}{(\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_1)} \boldsymbol{\beta}'_1 = \boldsymbol{\alpha}_2 = (1, -1, 0)^T, \\ \boldsymbol{\beta}'_3 &= \boldsymbol{\alpha}_3 - \frac{(\boldsymbol{\beta}'_1, \boldsymbol{\alpha}_3)}{(\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_1)} \boldsymbol{\beta}'_1 - \frac{(\boldsymbol{\beta}'_2, \boldsymbol{\alpha}_3)}{(\boldsymbol{\beta}'_2, \boldsymbol{\beta}'_2)} \boldsymbol{\beta}'_2 \\ &= \boldsymbol{\alpha}_3 - \frac{1}{2} \boldsymbol{\beta}'_1 + \frac{1}{2} \boldsymbol{\beta}'_2 = (0, 0, 2)^T. \end{aligned}$$

单位化后得

$$\begin{aligned} \boldsymbol{\beta}_1 &= \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0 \right)^T = \frac{1}{\sqrt{2}} \boldsymbol{\alpha}_1, \\ \boldsymbol{\beta}_2 &= \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0 \right)^T = \frac{1}{\sqrt{2}} \boldsymbol{\alpha}_2, \\ \boldsymbol{\beta}_3 &= (0, 0, 1)^T = -\frac{1}{4} \boldsymbol{\alpha}_1 + \frac{1}{4} \boldsymbol{\alpha}_2 + \frac{1}{2} \boldsymbol{\alpha}_3, \\ (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3) &= (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3) \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{4} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{4} \\ 0 & 0 & \frac{1}{2} \end{pmatrix}. \end{aligned}$$

所以

$$\begin{aligned}
 \mathbf{A} &= (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3) = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3) \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{4} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{4} \\ 0 & 0 & \frac{1}{2} \end{pmatrix}^{-1} \\
 &= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{4} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{4} \\ 0 & 0 & \frac{1}{2} \end{pmatrix}^{-1} \\
 &= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{2} & -\frac{1}{\sqrt{2}} \\ 0 & 0 & 2 \end{pmatrix} = \mathbf{QR}.
 \end{aligned}$$

定理 3.2.1 还可以推广到一般复矩阵的情况.

定理 3.2.2 设 \mathbf{A} 为 $m \times n$ 复矩阵 ($m \geq n$), 且 n 个列向量线性无关, 则 \mathbf{A} 有分解式

$$\mathbf{A} = \mathbf{UR}, \quad (3.2.4)$$

其中 \mathbf{U} 是 $m \times n$ 复矩阵, 且满足 $\mathbf{U}^H \mathbf{U} = \mathbf{I}$, \mathbf{R} 是 n 阶复非奇异上三角矩阵, 且除去相差一个对角元素的模全为 1 的对角矩阵因子外, 分解式 (3.2.4) 是惟一的.

实用上, 一般不用施密特正交化方法作 \mathbf{QR} 分解, 而是借助第 1 章介绍过的初等旋转变换或镜像变换对矩阵进行 \mathbf{QR} 分解.

3.2.2 QR 分解的实际求法

下面我们先考虑初等旋转变换的 \mathbf{QR} 分解, 然后再介绍镜像变换的 \mathbf{QR} 分解, 两种工具各有优缺点.

1. 吉文斯(Givens)方法

我们在 1.2 节已指出初等旋转矩阵的性质, 即用 \mathbf{R}_{ij} 左乘矩阵 \mathbf{A} 时, 仅影响 \mathbf{A} 的第 i 行和第 j 行, 且选适当的 \mathbf{R}_{ij} , 就可以消去 \mathbf{A} 的一个非零元素. 一般地说, 作一次旋转可以消去一个非零元素. 如果在作下一次旋转时不会影响前面已化为零的元素, 即不会重新又变成非零, 那么, 借助于初等旋转阵将 \mathbf{A} 约化成上三角阵就有希望. 事实上, 只要注意运算顺序, 完全能办到, 我们有

定理 3.2.3 任何实非奇异矩阵可通过左连乘初等旋转阵化为上三角阵.

证明 (1) 对实可逆矩阵 $\mathbf{A} = (a_{ij})$ 左乘以初等旋转阵 \mathbf{R}_{ij} 以后, 只改变 \mathbf{A} 的第 i 行和第 j 行元素.

设

$$\mathbf{A}' = \mathbf{R}_{ij} \mathbf{A},$$

则

$$a'_{ig} = ca_{ig} + sa_{jg}, \quad a'_{jg} = -sa_{ig} + ca_{jg}, \quad a'_{pg} = a_{pg}, \quad p \neq i, j; \quad g = 1, 2, \dots, n.$$

如果要使 A' 中第 g_0 列的第 j 个元素 $a'_{jg_0} = 0$, 那么, 只要 a_{ig_0} 和 a_{jg_0} 之一不等于零, 且取

$$s = \frac{a_{jg_0}}{\sqrt{a_{ig_0}^2 + a_{jg_0}^2}}, \quad c = \frac{a_{ig_0}}{\sqrt{a_{ig_0}^2 + a_{jg_0}^2}}$$

即可, 此时元素

$$a'_{ig_0} = \sqrt{a_{ig_0}^2 + a_{jg_0}^2} > 0,$$

也就是 A 的第 g_0 列的第 j 个元素化为零, 第 g_0 列的第 i 个元素变为正的, 而其他元素不变.

(2) 假设 $a_{11} \neq 0$, 取 $g_0 = 1$, 连续左乘 $R_{12}, R_{13}, \dots, R_{1n}$, 使得第 1 列除第 1 个元素为正外, 其他元素都被逐个地化为零, 即 A 化为

$$\begin{aligned} A^{(1)} &= R_{1n} R_{1, n-1} \cdots R_{12} A \\ &= \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & a_{22}^{(1)} & \cdots & a_{2n}^{(1)} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2}^{(1)} & \cdots & a_{nn}^{(1)} \end{pmatrix}, \end{aligned}$$

且 $a_{11}^{(1)} > 0$.

如果 $a_{11} = 0$, 则 A 从左乘以 R_{1i_0} 开始, 这里 i_0 是 $a_{i1} \neq 0$ 的最小足标, 由于 A 是可逆的, 故这样的 i_0 是可以找到的. 此时取 $g_0 = 1$, 矩阵 $R_{1i_0} A$ 的第 1 列第 1 个元素就为正了. 又由于 A 可逆, 右下角 $n-1$ 阶子式非零, 所以在 $a_{22}^{(1)}, a_{32}^{(1)}, \dots, a_{n2}^{(1)}$ 中至少有一个元素不等于零. 此时同样可认为 $a_{22}^{(1)} \neq 0$, 取 $g_0 = 2$, 连续左乘以 $R_{23}, R_{24}, \dots, R_{2n}$, 使 $A^{(1)}$ 化为

$$\begin{aligned} A^{(2)} &= R_{2n} R_{2, n-1} \cdots R_{23} A^{(1)} \\ &= \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2n}^{(2)} \\ 0 & 0 & a_{33}^{(2)} & \cdots & a_{3n}^{(2)} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & a_{n3}^{(2)} & \cdots & a_{nn}^{(2)} \end{pmatrix}, \end{aligned}$$

且第一行元素不变, $a_{22}^{(2)} > 0$.

要注意的是, 对 $A^{(1)}$ 所作的这些旋转并不影响 $A^{(1)}$ 中第一列上已得到的零元素, 这一点是很重要的.

继续进行下去, 最后 A 就化为上三角阵

$$\begin{aligned} A^{(n-1)} &= R_{n-1, n} \cdots R_{12} A & (3.2.5) \\ &= \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn}^{(n-1)} \end{pmatrix}, \end{aligned}$$

其中除了最后的 $a_{nn}^{(n-1)}$ 外, 所有的对角线元素都是正的. 显然, $a_{nn}^{(n-1)}$ 的符号与矩阵 A 的行列式的符号一致.

证毕

从上面定理的证明过程可以看出,利用初等旋转矩阵将 A 上三角化的过程实际上存在着 A 的一个 QR 分解. 因为由式(3.2.5)有 $A = (R_{n-1,n} \cdots R_{12})^{-1} A^{(n-1)}$, 令

$$R = A^{(n-1)},$$

$$Q = (R_{n-1,n} \cdots R_{12})^{-1},$$

由于每一个 R_{ij} 都是正交阵,所以它们连乘积的逆 Q 也是正交阵,从而得

$$A = QR.$$

例 3.2.2 用初等旋转矩阵求下列矩阵 A 的 QR 分解.

$$A = \begin{pmatrix} 12 & -20 & 41 \\ 9 & -15 & -63 \\ 20 & 50 & 35 \end{pmatrix}.$$

解 第一步:以 R_{12} 左乘 A , 消去第 2 行第 1 列处的元素.

$$a'_{11} = 15, \quad c = \frac{12}{15} = \frac{4}{5}, \quad s = \frac{9}{15} = \frac{3}{5},$$

故

$$R_{12} = \begin{pmatrix} \frac{4}{5} & \frac{3}{5} & 0 \\ -\frac{3}{5} & \frac{4}{5} & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$R_{12}A = \begin{pmatrix} 15 & -25 & -5 \\ 0 & 0 & -75 \\ 20 & 50 & 35 \end{pmatrix},$$

同理,有

$$a'_{21} = (15^2 + 20^2)^{\frac{1}{2}} = 25,$$

$$s = \frac{20}{25} = \frac{4}{5}, \quad c = \frac{15}{25} = \frac{3}{5},$$

$$R_{13} = \begin{pmatrix} \frac{3}{5} & 0 & \frac{4}{5} \\ 0 & 1 & 0 \\ -\frac{4}{5} & 0 & \frac{3}{5} \end{pmatrix},$$

$$\begin{pmatrix} 25 & 25 & 25 \\ 0 & 0 & -75 \\ 0 & 50 & 25 \end{pmatrix} = A^{(1)} = R_{13}R_{12}A.$$

第二步:根据 $A^{(1)}$ 中的元素算出 $a'_{22} = 50, c = 0, s = 1$. 由此确定

$$R_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix},$$

$$\mathbf{A}^{(2)} = \mathbf{R}_{23}\mathbf{A}^{(1)} = \begin{pmatrix} 25 & 25 & 25 \\ 0 & 50 & 25 \\ 0 & 0 & 75 \end{pmatrix},$$

即

$$\mathbf{R}_{23}\mathbf{R}_{13}\mathbf{R}_{12} \begin{pmatrix} 12 & -20 & 41 \\ 9 & -15 & -63 \\ 20 & 50 & 35 \end{pmatrix} = \begin{pmatrix} 25 & 25 & 25 \\ 0 & 50 & 25 \\ 0 & 0 & 75 \end{pmatrix},$$

即

$$\begin{aligned} \mathbf{Q} &= (\mathbf{R}_{23}\mathbf{R}_{13}\mathbf{R}_{12})^{-1} = \begin{pmatrix} \frac{12}{25} & \frac{9}{25} & \frac{20}{25} \\ -\frac{16}{25} & -\frac{12}{25} & \frac{15}{25} \\ \frac{15}{25} & -\frac{20}{25} & 0 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \frac{12}{25} & \frac{9}{25} & \frac{20}{25} \\ -\frac{16}{25} & -\frac{12}{25} & \frac{15}{25} \\ \frac{15}{25} & -\frac{20}{25} & 0 \end{pmatrix}^T = \begin{pmatrix} \frac{12}{25} & -\frac{16}{25} & \frac{15}{25} \\ \frac{9}{25} & -\frac{12}{25} & -\frac{20}{25} \\ \frac{20}{25} & \frac{15}{25} & 0 \end{pmatrix}, \end{aligned}$$

故

$$\mathbf{A} = \begin{pmatrix} \frac{12}{25} & -\frac{16}{25} & \frac{15}{25} \\ \frac{9}{25} & -\frac{12}{25} & -\frac{20}{25} \\ \frac{20}{25} & \frac{15}{25} & 0 \end{pmatrix} \begin{pmatrix} 25 & 25 & 25 \\ 0 & 50 & 25 \\ 0 & 0 & 75 \end{pmatrix} = \mathbf{QR}.$$

从上面论述可见,吉文斯方法需要作 $\frac{n(n-1)}{2}$ 个初等旋转矩阵的连乘积,当 n 较大时,计算工作量较大,因此,常利用镜像变换来进行QR分解.

2. 豪斯霍尔德(Housholder)方法

定理 3.2.4 任何实的 n 阶矩阵 \mathbf{A} 可用初等反射矩阵 $\mathbf{H} = \mathbf{I} - 2\boldsymbol{\omega}\boldsymbol{\omega}^T$ 化为上三角矩阵.

证明 设矩阵 \mathbf{A} 按列分块为

$$\mathbf{A} = (\boldsymbol{\alpha}_1^{(1)}, \boldsymbol{\alpha}_2^{(1)}, \dots, \boldsymbol{\alpha}_n^{(1)}),$$

其中 $\boldsymbol{\alpha}_i^{(1)}$ 是 \mathbf{A} 的第 i 个列向量.因为 $\boldsymbol{\alpha}_1^{(1)}$ 不会是零向量,于是可按式(1.2.37)构造单位向量 $\boldsymbol{\omega}^{(1)}$,使 $\boldsymbol{\alpha}_1^{(1)}$ 与单位向量 $\mathbf{e}_1 = (1, 0, \dots, 0)^T$ 同方向.从而必存在初等反射矩阵 $\mathbf{H}^{(1)}$,使

$$\mathbf{H}^{(1)}\mathbf{A} = \begin{pmatrix} |\boldsymbol{\alpha}_1^{(1)}| & * \\ 0 & \\ \vdots & \mathbf{A}_{n-1} \\ 0 & \end{pmatrix},$$

然后对矩阵 \mathbf{A}_{n-1} 再用 $n-1$ 阶的初等反射矩阵 $\hat{\mathbf{H}}^{(2)}$, 使

$$\hat{\mathbf{H}}^{(2)} \mathbf{A}_{n-1} = \begin{pmatrix} | a_{22}^{(2)} | & & * \\ 0 & & \\ \vdots & & \mathbf{A}_{n-2} \\ 0 & & \end{pmatrix},$$

因此有

$$\mathbf{H}^{(2)} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \hat{\mathbf{H}}^{(2)} & \\ 0 & & & \end{pmatrix},$$

使

$$\mathbf{H}^{(2)} \mathbf{H}^{(1)} \mathbf{A} = \begin{pmatrix} | a_{11}^{(1)} | & & & * \\ 0 & | a_{22}^{(2)} | & & * \\ 0 & 0 & & \\ \vdots & \vdots & & \mathbf{A}_{n-2} \\ 0 & 0 & & \end{pmatrix}.$$

显然, 左乘 $\mathbf{H}^{(2)}$ 不改变 $\mathbf{H}^{(1)} \mathbf{A}$ 的第 1 行和第 1 列的元素(这一点很重要).

下面我们证明 $\mathbf{H}^{(2)}$ 也是初等反射阵.

若 $\hat{\mathbf{H}}^{(2)} = \mathbf{I}_{n-1} - 2\hat{\boldsymbol{\omega}}\hat{\boldsymbol{\omega}}^T$, 这里 $\hat{\boldsymbol{\omega}}$ 是 $n-1$ 维向量空间 \mathbb{R}^{n-1} 中的单位向量, 记

$$\boldsymbol{\omega} = \begin{pmatrix} 0 \\ \hat{\boldsymbol{\omega}} \end{pmatrix},$$

它是 \mathbb{R}^n 中的单位向量, 易知

$$\mathbf{H}^{(2)} = \mathbf{I} - 2\boldsymbol{\omega}\boldsymbol{\omega}^T,$$

因此 $\mathbf{H}^{(2)}$ 也是一个初等反射阵.

这样继续进行, 经过 $n-1$ 次左乘初等反射矩阵, \mathbf{A} 便可化为一个上三角矩阵

$$\mathbf{H}^{(n-1)} \cdots \mathbf{H}^{(1)} \mathbf{A} = \mathbf{A}^{(n)}. \quad (3.2.6)$$

由于每一个子矩阵 $\mathbf{A}_{n-1}, \mathbf{A}_{n-2}, \cdots, \mathbf{A}_2$ 的第 1 列向量都不为零向量(否则与 \mathbf{A} 是非奇异的假定不合), 所以上述过程一定能够进行到底. 证毕

从式(3.2.6)有 $\mathbf{A} = (\mathbf{H}^{(n-1)} \cdots \mathbf{H}^{(1)})^{-1} \mathbf{A}^{(n)}$. 令

$$\mathbf{R} = \mathbf{A}^{(n)},$$

$$\mathbf{Q} = (\mathbf{H}^{(n-1)} \cdots \mathbf{H}^{(1)})^{-1}.$$

因为每一个 $\mathbf{H}^{(i)}$ ($i=1, 2, \cdots, n-1$) 是正交阵, 所以它们连乘积的逆也是正交阵, 从而有

$$\mathbf{A} = \mathbf{QR}.$$

从上面讨论可见, 豪斯霍尔德方法只需左乘 $n-1$ 个初等反射阵, 计算量大约是吉文斯方法的一半, 所以这是它的优势, 但对稀疏矩阵, 用吉文斯方法作 \mathbf{QR} 分解仍有方便之处.

例 3.2.3 试用豪斯霍尔德方法求矩阵

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

的 QR 分解.

解 因为 $\alpha_1^{(1)} = (2, 1, 2)^T \neq \mathbf{0}$, 作单位向量

$$\begin{aligned} \omega^{(1)} &= \frac{\alpha_1^{(1)} - |\alpha_1^{(1)}| e_1}{|\alpha_1^{(1)} - |\alpha_1^{(1)}| e_1|} = \frac{(2, 1, 2)^T - 3e_1}{|(2, 1, 2)^T - 3e_1|} \\ &= \left(-\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}} \right)^T, \end{aligned}$$

于是

$$\mathbf{H}^{(1)} = \mathbf{I}_3 - 2\omega^{(1)}\omega^{(1)T} = \begin{pmatrix} 2 & & \\ 3 & \frac{1}{3} & \\ & \frac{2}{3} & \frac{2}{3} \\ & \frac{2}{3} & -\frac{2}{3} \\ & & \frac{1}{3} \end{pmatrix},$$

从而得

$$\mathbf{H}^{(1)}\mathbf{A} = \begin{pmatrix} 3 & \frac{8}{3} & \frac{8}{3} \\ 0 & \frac{4}{3} & \frac{1}{3} \\ 0 & -\frac{1}{3} & -\frac{4}{3} \end{pmatrix} = \begin{pmatrix} 3 & \frac{8}{3} & \frac{8}{3} \\ & \mathbf{A}_2 & \\ & & \end{pmatrix}.$$

再对 \mathbf{A}_2 作如上的计算. 为此作

$$\begin{aligned} \hat{\omega}^{(2)} &= \frac{\left(\frac{4}{3}, -\frac{1}{3}\right)^T - \left|\left(\frac{4}{3}, -\frac{1}{3}\right)^T\right| e_1}{\left|\left(\frac{4}{3}, -\frac{1}{3}\right)^T - \left|\left(\frac{4}{3}, -\frac{1}{3}\right)^T\right| e_1\right|} \\ &= \left(\frac{4 - \sqrt{17}}{\sqrt{34 - 8\sqrt{17}}}, -\frac{1}{\sqrt{34 - 8\sqrt{17}}} \right)^T. \end{aligned}$$

于是有

$$\hat{\mathbf{H}}^{(2)} = \mathbf{I}_2 - 2\hat{\omega}^{(2)}\hat{\omega}^{(2)T} = \begin{pmatrix} 1 - \frac{2(4 - \sqrt{17})^2}{34 - 8\sqrt{17}} & \frac{2(4 - \sqrt{17})}{34 - 8\sqrt{17}} \\ \frac{2(4 - \sqrt{17})}{34 - 8\sqrt{17}} & 1 - \frac{2}{34 - 8\sqrt{17}} \end{pmatrix}.$$

再作

$$\mathbf{H}^{(2)} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{H}}^{(2)} \end{pmatrix} = \begin{pmatrix} 1 & & 0 \\ 0 & 1 - \frac{2(4 - \sqrt{17})^2}{34 - 8\sqrt{17}} & \frac{2(4 - \sqrt{17})}{34 - 8\sqrt{17}} \\ 0 & \frac{2(4 - \sqrt{17})}{34 - 8\sqrt{17}} & 1 - \frac{2}{34 - 8\sqrt{17}} \end{pmatrix},$$

从而有

$$\mathbf{H}^{(2)}\mathbf{H}^{(1)}\mathbf{A} = \begin{pmatrix} 3 & \frac{8}{3} & \frac{8}{3} \\ 0 & \frac{\sqrt{17}}{3} & \frac{8\sqrt{17}}{51} \\ 0 & 0 & \frac{5\sqrt{17}}{17} \end{pmatrix} = \mathbf{R},$$

故 \mathbf{A} 的 QR 分解为

$$\mathbf{A} = (\mathbf{H}^{(2)}\mathbf{H}^{(1)})^{-1}\mathbf{R} = \begin{pmatrix} \frac{2}{3} & \frac{2\sqrt{17}}{51} & -\frac{3\sqrt{17}}{17} \\ \frac{1}{3} & \frac{10\sqrt{17}}{51} & \frac{2\sqrt{17}}{17} \\ \frac{2}{3} & -\frac{7\sqrt{17}}{51} & \frac{2\sqrt{17}}{17} \end{pmatrix} \begin{pmatrix} 3 & \frac{8}{3} & \frac{8}{3} \\ 0 & \frac{\sqrt{17}}{3} & \frac{8\sqrt{17}}{51} \\ 0 & 0 & \frac{5\sqrt{17}}{17} \end{pmatrix}.$$

值得指出的是,上述的吉文斯方法和豪斯霍尔德方法也可以推广到复矩阵的情况,具体论述可参阅文献[2].

3.3 矩阵的最大秩分解

以上两节主要是介绍了 n 阶方阵的几种分解,从本节开始,将介绍几种常用的长方形的分解.在这一节给出长方形 \mathbf{A} 分解为两个与 \mathbf{A} 同秩的矩阵因子乘积的具体方法,并讨论不同分解之间的关系.它们在第 6 章广义逆矩阵的讨论中,是十分重要的.

首先,我们要弄清楚什么是矩阵的最大秩?

定义 3.3.1 设 $m \times n$ 矩阵

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

如果当 $m \leq n$ 时,存在有 $\text{rank}\mathbf{A} = m$; 或者当 $m \geq n$ 时,存在有 $\text{rank}\mathbf{A} = n$, 则称这两种长方形为最大秩长方形(满秩长方形),前者又称行最大秩矩阵(行满秩矩阵或矮矩阵),后者又称为列最大秩矩阵(列满秩矩阵或高矩阵).

显然,最大秩长方形具有如下性质

$$\text{rank}(\mathbf{A}\mathbf{A}^T) = m, \quad \mathbf{A} = (a_{ij})_{m \times n}, m \leq n \quad (3.3.1)$$

或

$$\text{rank}(\mathbf{A}^T\mathbf{A}) = n, \quad \mathbf{A} = (a_{ij})_{m \times n}, m \geq n. \quad (3.3.2)$$

下面我们给出式(3.3.1)的证明:

事实上,只要证明 $\mathbf{A}\mathbf{A}^T\mathbf{X} = \mathbf{0}$ 只有零解即可. 若 $\mathbf{A}\mathbf{A}^T\mathbf{X} = \mathbf{0}$, 则 $\mathbf{X}^T\mathbf{A}\mathbf{A}^T\mathbf{X} = \mathbf{0}$, 即 $(\mathbf{A}^T\mathbf{X})^T\mathbf{A}^T\mathbf{X} = \mathbf{0}$, 因此 $\mathbf{A}^T\mathbf{X} = \mathbf{0}$. 由于 \mathbf{A}^T 为 $n \times m$ 矩阵,且 $\text{rank}\mathbf{A}^T = \text{rank}\mathbf{A} = m$, 因此 $\mathbf{X} = \mathbf{0}$.

同理可以证得式(3.3.2).

定义 3.3.2 设 \mathbf{A} 为 $m \times n$ 且秩为 $r > 0$ 的复矩阵,且记为 $\mathbf{A} \in \mathbb{C}_r^{m \times n}$, 如果存在矩阵

$B \in \mathbb{C}_r^{m \times r}$ 和 $C \in \mathbb{C}_r^{r \times n}$, 使

$$A = BC, \quad (3.3.3)$$

则称式(3.3.3)的分解为矩阵 A 的最大秩分解(满秩分解).

显然, 当 A 是列最大秩(列满秩)或行最大秩(行满秩)矩阵时, A 的最大秩分解的两个因子中, 一个因子是单位矩阵, 另一个因子是 A 本身, 称这种最大秩分解为平凡分解.

定理 3.3.1 设 $A \in \mathbb{C}_r^{m \times n}$, 则一定存在 $B \in \mathbb{C}_r^{m \times r}$ 和 $C \in \mathbb{C}_r^{r \times n}$ 使得

$$A = BC.$$

证明 将 A 进行初等行变换, 化为行标准形 \tilde{A}_r (即对 A 进行有限次初等行变换得到的阶梯形矩阵):

$$\tilde{A}_r = \left(\begin{array}{cccccccccccccccc} 0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots & * \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & * \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & & & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & * & \cdots & * \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{array} \right) \left. \begin{array}{l} \\ \\ \\ \\ \\ \\ \end{array} \right\} \begin{array}{l} r \text{ 个非零行} \\ \\ \\ \\ m-r \text{ 个零行} \end{array},$$

其中 $*$ 表示不一定为 0 元素, 在 \tilde{A}_r 中第 k_j 列的元素除第 j 个元素为 1 外, 其余元素均为 0 ($1 \leq j \leq r$).

显然, \tilde{A}_r 的第 k_1, k_2, \dots, k_r 列是线性无关的, 则 A 的第 k_1, k_2, \dots, k_r 列也是线性无关的. 因此, 保留 A 的第 k_1, k_2, \dots, k_r 列而把其他各列都除去, 所得的矩阵记为 B , 则 B 是具有最大秩的 $m \times r$ 阶矩阵.

\tilde{A}_r 的第 j 列可以表示成

$$\tilde{\alpha}_j = \left(\begin{array}{c} c_{j1} \\ \vdots \\ c_{jp} \\ 0 \\ \vdots \\ 0 \end{array} \right) \left. \begin{array}{l} \\ \\ \end{array} \right\} \begin{array}{l} p \\ m-p \end{array},$$

式中 $k_p < j < k_{p+1}$ ($1 \leq p \leq r$), 因此它可用下列线性组合表示:

$$\tilde{\alpha}_j = c_{j1} \tilde{\alpha}_{k_1} + \cdots + c_{jp} \tilde{\alpha}_{k_p},$$

由初等行变换的性质可知, A 的第 j 列也可以用对应的线性组合表示, 即 $\alpha_j = c_{j1} \alpha_{k_1} + \cdots + c_{jp} \alpha_{k_p}$. 其次, 设除了位于 \tilde{A}_r 下侧 $m-r$ 个 0 行以外的 $r \times n$ 阶矩阵为 C , 显然 C 是具有最大秩的 $r \times n$ 阶矩阵, 当 $k_p < j < k_{p+1}$ 时, 则有

$$C \text{ 的第 } j \text{ 列} = \left(\begin{array}{c} c_{j1} \\ \vdots \\ c_{jp} \\ 0 \\ \vdots \\ 0 \end{array} \right) \left. \begin{array}{l} \\ \\ \end{array} \right\} \begin{array}{l} p \\ r-p \end{array},$$

而且, 因为 $B = (\alpha_{k_1}, \alpha_{k_2}, \dots, \alpha_{k_r})$, 则

$$A \text{ 的第 } j \text{ 列} = B \times (C \text{ 的第 } j \text{ 列}),$$

式中的 j 取任何一列也成立. 最后得

$$A = BC.$$

证毕

定理的证明过程给出了求 B, C 的方法, 下面举例说明之.

例 3.3.1 求矩阵

$$A = \begin{pmatrix} 1 & 4 & -1 & 5 & 6 \\ 2 & 0 & 0 & 4 & 6 \\ -1 & 2 & -4 & -4 & -19 \\ 1 & -2 & -1 & -1 & -6 \end{pmatrix}$$

为最大秩分解.

解 将 A 只进行初等行变换, 化为行标准形

$$A \simeq \begin{pmatrix} 1 & 0 & 0 & 2 & 3 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

于是取 A 的前三列组成矩阵

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

再取 A 的行标准形的前三个非零行, 组成矩阵

$$C = \begin{pmatrix} 1 & 0 & 0 & 2 & 3 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 5 \end{pmatrix},$$

容易验证 $A = BC$.

对于矩阵的性质, 一般来说, 对于“行”具有的性质, “列”也具有. 例如, 在例 3.3.1 中, 将 A 进行初等列变换化为列标准形, 得

$$A \simeq \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ -\frac{5}{7} & \frac{15}{14} & \frac{3}{7} & 0 & 0 \end{pmatrix},$$

于是, 将 A 的列标准形的前三列取作因子

$$\bar{B} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\frac{5}{7} & \frac{15}{14} & \frac{3}{7} \end{pmatrix}.$$

再将 A 的前三行取作因子

$$\tilde{C} = \begin{pmatrix} 1 & 4 & -1 & 5 & 6 \\ 2 & 0 & 0 & 4 & 6 \\ -1 & 2 & -4 & -4 & -19 \end{pmatrix},$$

容易验证 $A = \tilde{B}\tilde{C}$, 且

$$\text{rank}\tilde{B} = \text{rank}\tilde{C} = \text{rank}A = 3.$$

由此可见矩阵 A 的最大秩分解不是惟一的, 但最大秩分解之间, 有如下关系.

定理 3.3.2 设 $A \in \mathbb{C}_r^{m \times n}$, 且

$$A = BC = \tilde{B}\tilde{C}$$

均为 A 的最大秩分解, 则

(1) 存在矩阵 $Q \in \mathbb{C}_r^{r \times r}$, 使得

$$B = \tilde{B}Q, \quad C = Q^{-1}\tilde{C}. \quad (3.3.4)$$

(2) $C^H(CC^H)^{-1}(B^HB)^{-1}B^H = \tilde{C}^H(\tilde{C}\tilde{C}^H)^{-1}(\tilde{B}^H\tilde{B})^{-1}\tilde{B}^H$. (3.3.5)

证明 (1) 由 $BC = \tilde{B}\tilde{C}$, 有

$$BCC^H = \tilde{B}\tilde{C}C^H. \quad (3.3.6)$$

又由式(3.3.1)知 $\text{rank}C = \text{rank}(CC^H) = r$, $CC^H \in \mathbb{C}^{r \times r}$, 所以矩阵 CC^H 可逆. 在式(3.3.6)两端同时右乘 $(CC^H)^{-1}$ 得

$$B = \tilde{B}\tilde{C}C^H(CC^H)^{-1} = \tilde{B}Q_1, \quad (3.3.7)$$

其中 $Q_1 = \tilde{C}C^H(CC^H)^{-1}$.

同理可得

$$C = (B^HB)^{-1}B^H\tilde{B}\tilde{C} = Q_2\tilde{C}, \quad (3.3.8)$$

其中 $Q_2 = (B^HB)^{-1}B^H\tilde{B}$.

将式(3.3.7)和式(3.3.8)代入 $BC = \tilde{B}\tilde{C}$ 得

$$\tilde{B}\tilde{C} = \tilde{B}Q_1Q_2\tilde{C},$$

上式两端左乘 \tilde{B}^H 、右乘 \tilde{C}^H 得

$$\tilde{B}^H\tilde{B}\tilde{C}\tilde{C}^H = \tilde{B}^H\tilde{B}Q_1Q_2\tilde{C}\tilde{C}^H.$$

又由于 $\tilde{B}^H\tilde{B}, \tilde{C}\tilde{C}^H$ 均可逆, 上式两端分别左乘 $(\tilde{B}^H\tilde{B})^{-1}$, 右乘 $(\tilde{C}\tilde{C}^H)^{-1}$ 得

$$I_r = Q_1Q_2.$$

显然 Q_1, Q_2 均为 r 阶方阵. 若记 $Q_1 = Q$, 则 $Q_2 = Q^{-1}$, 即式(3.3.4)成立.

(2) 由式(3.3.4)有

$$\begin{aligned} & C^H(CC^H)^{-1}(B^HB)^{-1}B^H \\ &= (Q^{-1}\tilde{C})^H[Q^{-1}\tilde{C}(Q^{-1}\tilde{C})^H]^{-1}[(\tilde{B}Q)^H \times (\tilde{B}Q)]^{-1}(\tilde{B}Q)^H \\ &= \tilde{C}^H(Q^{-1})^H[Q^{-1}(\tilde{C}\tilde{C}^H)(Q^{-1})^H]^{-1} \times [Q^H(\tilde{B}^H\tilde{B})Q]^{-1}Q^H\tilde{B}^H \\ &= \tilde{C}^H(Q^{-1})^H[(Q^{-1})^H]^{-1}(\tilde{C}\tilde{C}^H)^{-1} \times QQ^{-1}(\tilde{B}^H\tilde{B})^{-1}(Q^H)^{-1}Q^H\tilde{B}^H \\ &= \tilde{C}^H(\tilde{C}\tilde{C}^H)^{-1}(\tilde{B}^H\tilde{B})^{-1}\tilde{B}^H, \end{aligned}$$

即式(3.3.5)成立. 证毕

式(3.3.5)表明, 矩阵 A 的最大秩分解虽不惟一, 但由最大秩分解所作出这种形式的乘积

$$C^H(CC^H)^{-1}(B^HB)^{-1}B^H$$

是相同的. 这个乘积表达式正是第6章中广义逆矩阵中的伪逆矩阵 A^+ (又称为穆尔-彭诺斯逆).

3.4 矩阵的奇异值分解和极分解

矩阵的奇异值分解在矩阵理论中的重要性是不言而喻的,例如古典控制中的频率法,正是由于有了矩阵奇异值分解的帮助而得到了新的发展.这里,只给出奇异值的性质以及矩阵按奇异值的分解.作为预备知识,我们先看下面的命题.

命题 3.4.1 设 $A \in \mathbb{C}^{m \times n}$, 则有

- (1) $A^H A$ 与 AA^H 的特征值均为非负实数;
- (2) $A^H A$ 与 AA^H 的非零特征值相同.

证明 (1) 设 $\mathbf{0} \neq \mathbf{x} \in \mathbb{C}^n$ 为矩阵 $A^H A$ 的特征值 λ 所对应的特征向量. 由于 $(A^H A)^H = A^H A$, 所以 $A^H A$ 是埃尔米特矩阵. 又因为 $\mathbf{x}^H A^H A \mathbf{x} = (A \mathbf{x})^H (A \mathbf{x}) \geq 0$, 故 $A^H A$ 是半正定的, 从而推得 $\lambda \geq 0$.

相仿可知 AA^H 的特征值也是非负实数.

- (2) 设 $A^H A$ 的特征值依大小顺序编号为

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0,$$

而 AA^H 的特征值也依大小顺序编号为

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_s > \mu_{s+1} = \mu_{s+2} = \cdots = \lambda_m = 0.$$

设 $\mathbf{0} \neq \mathbf{x}_i \in \mathbb{C}^n (i=1, 2, \cdots, r)$ 为 $A^H A$ 的非零特征值 $\lambda_i (i=1, 2, \cdots, r)$ 所对应的特征向量, 则由

$$A^H A \mathbf{x}_i = \lambda_i \mathbf{x}_i, \quad i = 1, 2, \cdots, r,$$

有

$$(AA^H) \mathbf{x}_i = \lambda_i \mathbf{x}_i, \quad i = 1, 2, \cdots, r,$$

且 $\mathbf{x}_i \neq \mathbf{0}$, 于是 λ_i 也是 AA^H 的非零特征值. 同理可证 AA^H 的非零特征值 μ_i 也是 $A^H A$ 的非零特征值. 如果还能证明 $A^H A$ 与 AA^H 的非零特征值的代数重复度亦相同, 则 $A^H A$ 与 AA^H 的非零特征值就全同了. 为此, 设 $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p$ 为 $A^H A$ 对应于特征值 $\lambda \neq 0$ 的线性无关的特征向量, 由于 $A^H A$ 属单纯矩阵, 故 p 即为 λ 的代数重复度. 显然, $A \mathbf{y}_i (i=1, 2, \cdots, p)$ 是 AA^H 对应于 $\lambda \neq 0$ 的特征向量. 为了证明这些特征向量 $A \mathbf{y}_i$ 线性无关, 令

$$k_1 A \mathbf{y}_1 + k_2 A \mathbf{y}_2 + \cdots + k_p A \mathbf{y}_p = \mathbf{0},$$

即

$$A(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p) \mathbf{k} = \mathbf{0},$$

其中 $\mathbf{k} = (k_1, k_2, \cdots, k_p)^T$. 于是

$$A^H A(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p) \mathbf{k} = \mathbf{0},$$

即

$$\lambda(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p) \mathbf{k} = \mathbf{0}.$$

已知 $\lambda \neq 0$, 故

$$(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p) \mathbf{k} = \mathbf{0},$$

已知 $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p$ 线性无关, 故 $\mathbf{k} = \mathbf{0}$, 即 $A \mathbf{y}_1, A \mathbf{y}_2, \cdots, A \mathbf{y}_p$ 线性无关, 因而 λ 也是 AA^H 的 p 重非零特征值.

定义 3.4.1 设 $A \in \mathbb{C}_r^{m \times n}$, $A^H A$ 的特征值为

$$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_r > \lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0,$$

则称 $\sigma_i = \sqrt{\lambda_i} (i=1, 2, \dots, r)$ 为矩阵 \mathbf{A} 的正奇异值, 简称奇异值.

由此定义和命题可知, \mathbf{A} 与 \mathbf{A}^H 有相同的奇异值.

例 3.4.1 设

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

求 \mathbf{A} 的奇异值.

解 由于

$$\mathbf{A}\mathbf{A}^H = \begin{pmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

显然, $\mathbf{A}\mathbf{A}^H$ 的正特征值为 5, 故 \mathbf{A} 的奇异值为 $\sqrt{5}$.

定义 3.4.2 设 $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{m \times n}$, 如果存在 m 阶酉矩阵 \mathbf{U} 和 n 阶酉矩阵 \mathbf{V} , 使得

$$\mathbf{B} = \mathbf{U}\mathbf{A}\mathbf{V}, \quad (3.4.1)$$

则称 \mathbf{A} 与 \mathbf{B} 酉等价或酉相抵.

定理 3.4.1 若 \mathbf{A} 与 \mathbf{B} 酉等价, 则 \mathbf{A} 与 \mathbf{B} 有相同的奇异值.

证明 因为 $\mathbf{B} = \mathbf{U}\mathbf{A}\mathbf{V}$, 所以有

$$\mathbf{B}^H\mathbf{B} = \mathbf{V}^H\mathbf{A}^H\mathbf{U}^H\mathbf{U}\mathbf{A}\mathbf{V} = \mathbf{V}^H\mathbf{A}^H\mathbf{A}\mathbf{V},$$

即 $\mathbf{A}^H\mathbf{A}$ 与 $\mathbf{B}^H\mathbf{B}$ 酉相似, 故它们有相同的特征值, 根据定义 3.4.1 知, \mathbf{A} 与 \mathbf{B} 有相同的奇异值. 证毕

定理 3.4.2 设 $\mathbf{A} \in \mathbb{C}_r^{m \times n}$, 则存在 m 阶酉矩阵 \mathbf{U} 和 n 阶酉矩阵 \mathbf{V} , 使得

$$\mathbf{U}^H\mathbf{A}\mathbf{V} = \begin{bmatrix} \mathbf{\Delta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (3.4.2)$$

或

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \mathbf{\Delta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}^H, \quad (3.4.3)$$

其中 $\mathbf{\Delta} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$, λ_i 为 $\mathbf{A}\mathbf{A}^H$ 的非零特征值, 且 $\sigma_i = \sqrt{\lambda_i} (i=1, 2, \dots, r)$, 而 σ_i 是 \mathbf{A} 的全部奇异值.

证明 由于 $\mathbf{A}\mathbf{A}^H$ 总是半正定的, 故存在酉阵 \mathbf{U} , 使得

$$\mathbf{U}^H\mathbf{A}\mathbf{A}^H\mathbf{U} = \begin{bmatrix} \mathbf{\Delta}\mathbf{\Delta}^H & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2, 0, \dots, 0).$$

记 $\mathbf{U} = (\mathbf{x}_1, \dots, \mathbf{x}_r, \mathbf{x}_{r+1}, \dots, \mathbf{x}_m) = (\mathbf{U}_1, \mathbf{U}_2)$, 其中 $\mathbf{U}_1 = (\mathbf{x}_1, \dots, \mathbf{x}_r)$, $\mathbf{U}_2 = (\mathbf{x}_{r+1}, \dots, \mathbf{x}_m)$, 代入上式有

$$(\mathbf{U}_1, \mathbf{U}_2)^H \mathbf{A}\mathbf{A}^H (\mathbf{U}_1, \mathbf{U}_2) = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2, 0, \dots, 0).$$

比较上式两端, 得

$$\mathbf{U}_1^H \mathbf{A}\mathbf{A}^H \mathbf{U}_1 = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2) = \mathbf{\Delta}^2 = \mathbf{\Delta}\mathbf{\Delta}^H, \quad (3.4.4)$$

$$\mathbf{U}_2^H \mathbf{A}\mathbf{A}^H \mathbf{U}_2 = \mathbf{0}. \quad (3.4.5)$$

由式(3.4.5)有

$$(\mathbf{A}^H \mathbf{U}_2)^H (\mathbf{A}^H \mathbf{U}_2) = \mathbf{0},$$

故

$$\mathbf{A}^H \mathbf{U}_2 = \mathbf{0} \quad \text{或} \quad \mathbf{U}_2^H \mathbf{A} = \mathbf{0}. \quad (3.4.6)$$

令

$$\mathbf{V}_1 = \mathbf{A}^H \mathbf{U}_1 (\Delta^{-1})^H, \quad (3.4.7)$$

则有

$$\mathbf{V}_1^H \mathbf{V}_1 = \Delta^{-1} \mathbf{U}_1^H \mathbf{A} \mathbf{A}^H \mathbf{U}_1 (\Delta^{-1})^H,$$

把式(3.4.4)代入上式,得

$$\mathbf{V}_1^H \mathbf{V}_1 = \Delta^{-1} \Delta \Delta^H (\Delta^{-1})^H = \mathbf{I}_r,$$

即 $\mathbf{V}_1 \in \mathbf{U}^{n \times r}$ (表示 $n \times r$ 酉矩阵集合). 再令 $\mathbf{V}_2 \in \mathbf{U}^{n \times (n-r)}$, 使

$$\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2) \in \mathbf{U}^{n \times n},$$

则有

$$\mathbf{V}_1^H \mathbf{V}_2 = \Delta^{-1} \mathbf{U}_1^H \mathbf{A} \mathbf{V}_2 = \mathbf{0},$$

从而有

$$\mathbf{U}_1^H \mathbf{A} \mathbf{V}_2 = \mathbf{0}, \quad (3.4.8)$$

故有

$$\mathbf{U}^H \mathbf{A} \mathbf{V} = \begin{bmatrix} \mathbf{U}_1^H \\ \mathbf{U}_2^H \end{bmatrix} \mathbf{A} (\mathbf{V}_1, \mathbf{V}_2) = \begin{bmatrix} \mathbf{U}_1^H \mathbf{A} \mathbf{V}_1 & \mathbf{U}_1^H \mathbf{A} \mathbf{V}_2 \\ \mathbf{U}_2^H \mathbf{A} \mathbf{V}_1 & \mathbf{U}_2^H \mathbf{A} \mathbf{V}_2 \end{bmatrix}.$$

由式(3.4.6)、式(3.4.7)和式(3.4.8)得

$$\mathbf{U}^H \mathbf{A} \mathbf{V} = \begin{bmatrix} \mathbf{U}_1^H \mathbf{A} \mathbf{A}^H \mathbf{U}_1 (\Delta^{-1})^H & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

再由式(3.4.4)便得

$$\mathbf{U}^H \mathbf{A} \mathbf{V} = \begin{bmatrix} \Delta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \quad \text{证毕}$$

\mathbf{A} 的式(3.4.3)形式的分解称为 \mathbf{A} 的奇异值分解, 实际上, 它表明 \mathbf{A} 与一个长方对角矩阵酉等价.

例 3.4.2 求例 3.4.1 中矩阵 \mathbf{A} 的奇异值分解.

解 由例 3.4.1 已知 \mathbf{A} 的奇异值 $\sigma_1 = \sqrt{5}$, 则 $\Delta = (\sqrt{5})$, 且由

$$\mathbf{A} \mathbf{A}^H = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

知 $\mathbf{A} \mathbf{A}^H$ 的特征值 $\lambda_1 = 5, \lambda_2 = \lambda_3 = 0$, 对应的特征向量可分别取为

$$\mathbf{x}_1 = (1, 0, 0)^T, \quad \mathbf{x}_2 = (0, 1, 0)^T, \quad \mathbf{x}_3 = (0, 0, 1)^T,$$

则 $\mathbf{U} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = \mathbf{I}_3$, 其中 $\mathbf{U}_1 = (\mathbf{x}_1), \mathbf{U}_2 = (\mathbf{x}_2, \mathbf{x}_3)$, 而

$$\mathbf{V}_1 = \mathbf{A}^H \mathbf{U}_1 (\Delta^{-1})^H = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \frac{1}{\sqrt{5}} = \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}} \right)^T,$$

故可取

$$\mathbf{V}_2 = \begin{bmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix},$$

于是得

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{5} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}.$$

例 3.4.3 求以下矩阵的奇异值分解

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}.$$

解 在实际计算中,应该求 $\mathbf{A}^H\mathbf{A}$ 或 $\mathbf{A}\mathbf{A}^H$ 中阶数较小的方阵的特征值. 因为

$$\mathbf{A}\mathbf{A}^H = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

为 2×2 方阵, 而 $\mathbf{A}^H\mathbf{A}$ 为 3×3 方阵, 故求前者的特征值方便些.

$$|\lambda\mathbf{I} - \mathbf{A}\mathbf{A}^H| = \begin{vmatrix} \lambda - 2 & 1 \\ 1 & \lambda - 2 \end{vmatrix} = (\lambda - 1)(\lambda - 3),$$

因此, $\mathbf{A}\mathbf{A}^H$ 的特征值为 $\lambda_1 = 1, \lambda_2 = 3$, 它们相应的单位特征向量分别为

$$\boldsymbol{\beta}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\beta}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

则 $\mathbf{A}^H\mathbf{A}$ 的三个特征值分别为 $\lambda_1 = 1, \lambda_2 = 3$ (非零特征值与前相同), 而 $\lambda_3 = 0$, 且属于它们的特征向量分别为

$$\boldsymbol{\alpha}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{\alpha}_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \quad \boldsymbol{\alpha}_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}.$$

注意, $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$ 可由埃尔米特 (Hermite) 矩阵 $\mathbf{A}^H\mathbf{A}$ 属于不同特征值的特征向量彼此正交得到. 令

$$\mathbf{V} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3), \quad \mathbf{U} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2), \quad \mathbf{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3} & 0 \end{bmatrix},$$

则

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^H = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \end{bmatrix}.$$

例 3.4.4 设可逆矩阵 \mathbf{A} 的奇异值分解为 $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^H$, 则其逆的奇异值分解为 $\mathbf{A}^{-1} = \mathbf{V}\mathbf{D}^{-1}\mathbf{U}^H$. 因此, 若 \mathbf{A} 的奇异值为 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$, 则 \mathbf{A}^{-1} 的奇异值为 $1/\sigma_n \geq 1/\sigma_{n-1} \geq \cdots \geq 1/\sigma_1 > 0$.

设 $\mathbf{A} = \mathbf{U}_1\mathbf{D}\mathbf{V}^H$ 是 \mathbf{A} 的奇异值分解, 令

$$\mathbf{P} = \mathbf{U}_1\mathbf{D}\mathbf{U}_1^H, \quad \mathbf{U} = \mathbf{U}_1\mathbf{V}^H,$$

即可得到矩阵的另一种有趣分解——极分解.

定理 3.4.3 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 则存在酉矩阵 \mathbf{U} 和惟一的半正定矩阵 \mathbf{P} , 使得

$$\mathbf{A} = \mathbf{P}\mathbf{U}, \quad (3.4.9)$$

上式称为矩阵 \mathbf{A} 的极分解. 矩阵 \mathbf{P} 与 \mathbf{U} 分别称为 \mathbf{A} 的埃尔米特因子和酉因子.

本节最后, 我们解释奇异值这个词的含义, 它有许多良好的性质(读者自行验证):

命题 3.4.2 (奇异值与特征值) 设 λ 是 n 阶矩阵 \mathbf{A} 的一个特征值, 又将 \mathbf{A} 的最大奇异值与最小奇异值分别记为 $\sigma_{\max}(\mathbf{A})$ 与 $\sigma_{\min}(\mathbf{A})$, 则 $\sigma_{\max}(\mathbf{A}) \geq |\lambda| \geq \sigma_{\min}(\mathbf{A})$. 换言之, 矩阵的最大奇异值与最小奇异值是其特征值的模的上下界.

例 3.4.5 设 $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, 则 \mathbf{A} 的特征值为 1, 但奇异值为 $\sqrt{(3 \pm \sqrt{5})/2}$.

命题 3.4.3 (奇异值与矩阵的迹) 设 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 则 $\text{tr}(\mathbf{A}^H\mathbf{A}) = \sum_{i=1}^r \sigma_i^2$. (请参照前面的舒赫不等式)

命题 3.4.4 (奇异值与奇异矩阵) 矩阵 \mathbf{A} 列满秩 $\Leftrightarrow \mathbf{A}$ 的奇异值均非 0. 特别地, 方阵 \mathbf{A} 非奇异 $\Leftrightarrow \mathbf{A}$ 的奇异值均非 0.

例 3.4.6 设 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 则 $\mathbf{A} = \mathbf{0} \Leftrightarrow \mathbf{A}^H\mathbf{A} = \mathbf{0}$. 换言之, 方阵 $\mathbf{A} = \mathbf{0} \Leftrightarrow$ 它的奇异值均为 0.

矩阵的奇异值较之其特征值的一个优点是非零奇异值的个数恰好是该矩阵的秩, 而矩阵的非零特征值的个数一般比其秩小(比如幂零矩阵无非零特征值). 因此常常利用此性质来计算矩阵的秩.

3.5 矩阵的谱分解

由于所有分解的目的不外乎简化计算或深化理论, 因此都要涉及一些特殊性质的“好矩阵”的分解. 这里讲的好矩阵是指可以(酉)对角化的矩阵, 对于这类“好矩阵”, 其分解方法比前几节更加直观和简单. 本节先介绍正规矩阵(可以酉对角化)的谱分解, 再介绍单纯矩阵(可相似对角化)的谱分解.

3.5.1 正规矩阵

定义 3.5.1 设 \mathbf{A} 是复数域上的方阵, 如果有

$$\mathbf{A}\mathbf{A}^H = \mathbf{A}^H\mathbf{A}, \quad (3.5.1)$$

则称 \mathbf{A} 为正规矩阵.

如果 \mathbf{A} 是实数域上的 n 阶方阵, 且有

$$\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A}, \quad (3.5.2)$$

则说 \mathbf{A} 为实正规矩阵.

我们容易验证,对称矩阵、反对称矩阵($\mathbf{A} = -\mathbf{A}^T$)、正交矩阵都是实正规矩阵;而酉矩阵、埃尔米特矩阵、反埃尔米特矩阵(即 $\mathbf{A} = -\mathbf{A}^H$)均属于复正规矩阵.

例如,若 \mathbf{A} 是埃尔米特矩阵(含对称矩阵),由于 $\mathbf{A}^H = \mathbf{A}$,则 $\mathbf{A}^H \mathbf{A} = \mathbf{A}^2 = \mathbf{A} \mathbf{A}^H$,故埃尔米特矩阵是正规矩阵.

又若 \mathbf{U} 是酉矩阵(含正交矩阵),即有 $\mathbf{U}^H \mathbf{U} = \mathbf{I} = \mathbf{U} \mathbf{U}^H$,所以 \mathbf{U} 是正规矩阵.

除了以上几种矩阵外,还有其他的正规矩阵.读者可以验证,矩阵 $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ 是正规矩

阵. 对角矩阵 $\mathbf{A} = \text{diag}(2+i, 1-i)$ 以及 $\mathbf{B} = \begin{bmatrix} 1 & 1-2i \\ 2+i & 1 \end{bmatrix}$ 都是正规矩阵.

由定理 1.3.2 的推论告诉我们,埃尔米特矩阵一定酉相似于一个对角矩阵.现在我们把埃尔米特矩阵的这种性质推广到更一般的情况,即关于正规矩阵有如下的定理.

定理 3.5.1 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$,则 \mathbf{A} 酉相似于对角矩阵的充分必要条件是 \mathbf{A} 为正规矩阵.

证 充分性:若 \mathbf{A} 是正规阵,即 $\mathbf{A}^H \mathbf{A} = \mathbf{A} \mathbf{A}^H$.根据舒赫定理,存在酉矩阵 \mathbf{U} 使得 $\mathbf{U}^H \mathbf{A} \mathbf{U} = \mathbf{T}$ 为上三角矩阵,且其对角元素是 \mathbf{A} 的特征值.于是有

$$\begin{aligned} \mathbf{T} \mathbf{T}^H &= (\mathbf{U}^H \mathbf{A} \mathbf{U}) (\mathbf{U}^H \mathbf{A} \mathbf{U})^H = \mathbf{U}^H \mathbf{A} \mathbf{U} \mathbf{U}^H \mathbf{A}^H \mathbf{U} \\ &= \mathbf{U}^H \mathbf{A} \mathbf{A}^H \mathbf{U} = \mathbf{U}^H \mathbf{A}^H \mathbf{A} \mathbf{U} \\ &= \mathbf{U}^H \mathbf{A}^H \mathbf{U} \cdot \mathbf{U}^H \mathbf{A} \mathbf{U} = \mathbf{T}^H \mathbf{T}. \end{aligned}$$

记

$$\mathbf{T} = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ & t_{22} & \cdots & t_{2n} \\ & & \ddots & \vdots \\ & & & t_{nn} \end{pmatrix},$$

由 $(\mathbf{T} \mathbf{T}^H)_{ii} = \sum_{j=i}^n |t_{ij}|^2$, $(\mathbf{T}^H \mathbf{T})_{ii} = \sum_{j=1}^i |t_{ji}|^2$, 因

$$(\mathbf{T} \mathbf{T}^H)_{ii} = (\mathbf{T}^H \mathbf{T})_{ii}, \quad i = 1, 2, \dots, n,$$

故有 $t_{ij} = 0 (i \neq j)$, 这表明 \mathbf{T} 是一个对角阵. 记

$$\mathbf{\Lambda} = \text{diag}(t_{11}, t_{22}, \dots, t_{nn}),$$

则

$$\mathbf{U}^H \mathbf{A} \mathbf{U} = \mathbf{\Lambda}.$$

这样就得到 \mathbf{A} 酉相似于对角阵 $\mathbf{\Lambda}$,且其对角元素是 \mathbf{A} 的特征值.

必要性: 设 \mathbf{A} 酉相似于对角阵 $\mathbf{\Lambda}$, 即

$$\mathbf{U}^H \mathbf{A} \mathbf{U} = \mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n),$$

则

$$\mathbf{A}^H = \text{diag}(\bar{\lambda}_1, \bar{\lambda}_2, \dots, \bar{\lambda}_n)$$

也是对角阵,因此 $\mathbf{\Lambda}$ 和 $\mathbf{\Lambda}^H$ 相乘是可以交换的, 即有

$$\mathbf{\Lambda} \mathbf{\Lambda}^H = \mathbf{\Lambda}^H \mathbf{\Lambda}.$$

但是 $\mathbf{A}\mathbf{A}^H = \mathbf{U}^H\mathbf{A}\mathbf{A}^H\mathbf{U}$, $\mathbf{A}^H\mathbf{A} = \mathbf{U}^H\mathbf{A}^H\mathbf{A}\mathbf{U}$, 故有

$$\mathbf{U}^H\mathbf{A}\mathbf{A}^H\mathbf{U} = \mathbf{U}^H\mathbf{A}^H\mathbf{A}\mathbf{U},$$

即

$$\mathbf{A}\mathbf{A}^H = \mathbf{A}^H\mathbf{A},$$

这说明 \mathbf{A} 是一个正规矩阵. 证毕

由此可见, 线性代数中介绍的实对称矩阵可正交相似于一个对角阵的问题, 只不过是本定理充分条件的特例. 除实对称矩阵是实正规矩阵外, 还有实反对称阵、正交阵、埃尔米特阵、反埃尔米特阵、酉矩阵等也都是正规阵, 所以定理 3.5.1 解决了一大类矩阵的酉对角化问题. 这是化二次型为标准形的理论基础.

由定理 3.5.1 可以得到下面的推论.

推论 设 \mathbf{A} 为正规矩阵, 若 \mathbf{A} 又为三角矩阵, 则 \mathbf{A} 为对角矩阵.

例 3.5.1 矩阵 $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ 可以相似对角化, 但它不能酉对角化, 所以它不是正规矩阵.

事实上, 此时 \mathbf{A} 的两个特征值为 $\lambda_1 = 0, \lambda_2 = 1$, 对应的两个非零特征向量取为 $\mathbf{X}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \mathbf{X}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, 可以验证它们线性无关, 所以有可逆矩阵 $\mathbf{C} = \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix}$, 使 $\mathbf{C}^{-1}\mathbf{A}\mathbf{C} = \begin{bmatrix} 0 & \\ & 1 \end{bmatrix}$, 即矩阵 \mathbf{A} 可以相似对角化. 但由于 $\mathbf{A}\mathbf{A}^T \neq \mathbf{A}^T\mathbf{A}$, 从而 \mathbf{A} 不是正规矩阵.

注意: 该矩阵是一个幂等矩阵.

定理 3.5.2 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 则 \mathbf{A} 为正规矩阵 $\Leftrightarrow \mathbf{A}$ 有 n 个两两正交的单位特征向量.

推论 正规矩阵属于不同特征值的特征向量是相互正交的.

正规矩阵有许多良好的数字特性, 比如下面的定理 3.5.3.

定理 3.5.3 设 $\mathbf{A} = (a_{ij})_{n \times n}$ 是复矩阵, $\lambda_1, \lambda_2, \dots, \lambda_n$ 为 \mathbf{A} 的 n 个特征值, 则

$$(1) \text{ (舒赫不等式) } \sum_{i=1}^n |\lambda_i|^2 \leq \sum_{i,j=1}^n |a_{ij}|^2;$$

$$(2) \mathbf{A} \text{ 为正规矩阵 } \Leftrightarrow \sum_{i=1}^n |\lambda_i|^2 = \sum_{i,j=1}^n |a_{ij}|^2.$$

证明 此处仅给出证明思路, 细节读者自行完成. 由第 1 章舒赫酉三角化定理 1.3.2 可知 $\mathbf{U}^H\mathbf{A}\mathbf{U} = \mathbf{B}$ 是上三角矩阵; 利用下述等式

$$\text{tr}(\mathbf{A}\mathbf{A}^H) = \sum_{i,j=1}^n |a_{ij}|^2$$

并比较 $\text{tr}(\mathbf{A}\mathbf{A}^H)$ 与 $\text{tr}(\mathbf{B}\mathbf{B}^H)$.

例 3.5.2 设 \mathbf{A} 为正规矩阵且幂零, 则 $\mathbf{A} = \mathbf{0}$.

事实上, 由于此时 \mathbf{A} 的特征值为 0, 再由定理 3.5.3 中(2), 即知 $\mathbf{A} = \mathbf{0}$.

3.5.2 正规矩阵的谱分解

设 \mathbf{A} 为正规矩阵, 则由定理 3.5.1 知, 存在酉矩阵 \mathbf{U} 使得 $\mathbf{U}^H\mathbf{A}\mathbf{U} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, 因而

$$\mathbf{A} = \mathbf{U} \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) \mathbf{U}^H.$$

令 $\mathbf{U} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_n)$, 则

$$\begin{aligned} \mathbf{A} &= (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_n) \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_1^H \\ \boldsymbol{\alpha}_2^H \\ \vdots \\ \boldsymbol{\alpha}_n^H \end{bmatrix} \\ &= \lambda_1 \boldsymbol{\alpha}_1 \boldsymbol{\alpha}_1^H + \lambda_2 \boldsymbol{\alpha}_2 \boldsymbol{\alpha}_2^H + \dots + \lambda_n \boldsymbol{\alpha}_n \boldsymbol{\alpha}_n^H. \end{aligned} \quad (3.5.3)$$

由于 $\lambda_1, \lambda_2, \dots, \lambda_n$ 为 \mathbf{A} 的特征值, $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_n$ 为对应的两两正交的单位特征向量, 故式(3.5.3)称为正规矩阵 \mathbf{A} 的谱分解或特征(值)分解. 若把式(3.5.3)中系数相同的放在一起(0特征值对应的项去掉), 然后把系数提出来, 则式(3.5.3)就变成

$$\mathbf{A} = \lambda_1 \mathbf{P}_1 + \lambda_2 \mathbf{P}_2 + \dots + \lambda_s \mathbf{P}_s, \quad (3.5.4)$$

其中 $\lambda_1, \lambda_2, \dots, \lambda_s$ 为 \mathbf{A} 的互不相同的非零特征值. 由于

$$\begin{aligned} (\boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^H)^H &= \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^H, \quad 1 \leq i \leq n, \\ (\boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^H)(\boldsymbol{\alpha}_j \boldsymbol{\alpha}_j^H) &= \mathbf{0}, \quad 1 \leq i \neq j \leq n, \\ (\boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^H)^2 &= \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^H, \quad 1 \leq i \leq n, \end{aligned}$$

所以

$$\mathbf{P}_i^H = \mathbf{P}_i, \quad \mathbf{P}_i^2 = \mathbf{P}_i, \quad \mathbf{P}_i \mathbf{P}_j = \mathbf{0}, \quad 1 \leq i \neq j \leq s \quad (3.5.5)$$

从定理 1.2.15 幂等矩阵与投影变换的对应关系可知, \mathbf{P}_i 是某正交投影变换(在某基下)的矩阵, 故常称为正交投影矩阵.

例 3.5.3 (谱分解的几何意义) 如果二阶实正规矩阵 \mathbf{A} 有两个相同的特征值 λ , 则 $\mathbf{A} = \lambda \mathbf{I}$ 就是它的谱分解; 如果有两个不同的特征值 λ_1 与 λ_2 , 则其谱分解为 $\mathbf{A} = \lambda_1 \mathbf{P}_1 + \lambda_2 \mathbf{P}_2$. 因此对任意 $\boldsymbol{\alpha} \in \mathbb{R}^2$, 有

$$\mathbf{A}\boldsymbol{\alpha} = \lambda_1 \mathbf{P}_1 \boldsymbol{\alpha} + \lambda_2 \mathbf{P}_2 \boldsymbol{\alpha}, \quad (3.5.6)$$

计算内积可得 $(\mathbf{P}_1 \boldsymbol{\alpha}, \mathbf{P}_2 \boldsymbol{\alpha}) = (\mathbf{P}_2 \boldsymbol{\alpha})^H \mathbf{P}_1 \boldsymbol{\alpha} = \boldsymbol{\alpha}^H \mathbf{P}_2^H \mathbf{P}_1 \boldsymbol{\alpha} = 0$, 所以 $\lambda_1 \mathbf{P}_1 \boldsymbol{\alpha}$ 与 $\lambda_2 \mathbf{P}_2 \boldsymbol{\alpha}$ 是正交的向量, 故式(3.5.6)将 $\mathbf{A}\boldsymbol{\alpha}$ 分解成了两个正交向量的和. 因此, 二维正规矩阵的谱分解实际上是平面的正交投影变换的推广. 对任意 n 阶正规矩阵的谱分解, 式(3.5.4)有类似的解释.

例 3.5.4 设 $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, 求 \mathbf{A} 的谱分解.

$$\text{解} \quad |\lambda \mathbf{I} - \mathbf{A}| = \begin{vmatrix} \lambda & -1 \\ 1 & \lambda \end{vmatrix} = \lambda^2 + 1 = 0,$$

故 \mathbf{A} 的特征值为一对共轭复数 $\lambda_1 = i, \lambda_2 = -i$, 它们所对应的单位正交特征向量取为 $\boldsymbol{\alpha}_1 =$

$$\begin{bmatrix} \frac{i}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \text{ 和 } \boldsymbol{\alpha}_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} \end{bmatrix}, \text{ 故 } \mathbf{A} \text{ 的谱分解为}$$

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \lambda_1 \boldsymbol{\alpha}_1 \boldsymbol{\alpha}_1^H + \lambda_2 \boldsymbol{\alpha}_2 \boldsymbol{\alpha}_2^H = i \begin{bmatrix} \frac{1}{2} & -\frac{i}{2} \\ \frac{i}{2} & \frac{1}{2} \end{bmatrix} - i \begin{bmatrix} \frac{1}{2} & \frac{i}{2} \\ -\frac{i}{2} & \frac{1}{2} \end{bmatrix}.$$

令

$$\mathbf{P}_1 = \begin{bmatrix} \frac{1}{2} & -\frac{i}{2} \\ \frac{i}{2} & \frac{1}{2} \end{bmatrix}, \quad \mathbf{P}_2 = \begin{bmatrix} \frac{1}{2} & \frac{i}{2} \\ -\frac{i}{2} & \frac{1}{2} \end{bmatrix},$$

可以验证有

$$\mathbf{P}_i^H = \mathbf{P}_i, \quad \mathbf{P}_i^2 = \mathbf{P}_i, \quad \mathbf{P}_i \mathbf{P}_j = \mathbf{0}, \quad 1 \leq i \neq j \leq 2,$$

对于任意 $\boldsymbol{\alpha} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ 有

$$\mathbf{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda_1 \mathbf{P}_1 \boldsymbol{\alpha} + \lambda_2 \mathbf{P}_2 \boldsymbol{\alpha} = \frac{i}{2} \begin{bmatrix} x_1 - ix_2 \\ ix_1 + x_2 \end{bmatrix} - \frac{i}{2} \begin{bmatrix} x_1 + ix_2 \\ -ix_1 + x_2 \end{bmatrix},$$

容易看出,上式右端的两个复向量 $\lambda_1 \mathbf{P}_1 \boldsymbol{\alpha}$ 与 $\lambda_2 \mathbf{P}_2 \boldsymbol{\alpha}$ 是正交的.

例 3.5.5 设 $\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$, 求 $\mathbf{A}^T \mathbf{A}$ 与 $\mathbf{A} \mathbf{A}^T$ 的谱分解.

解 直接计算可知

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad |\lambda \mathbf{I} - \mathbf{A}^T \mathbf{A}| = \lambda(\lambda - 1)(\lambda - 3),$$

只需计算非零特征值的特征向量,得属于特征值 1 与 3 的特征向量(必定正交)分别为 $(1, 0, -1)^T$ 与 $(1, 2, 1)^T$, 因此 $\mathbf{A}^T \mathbf{A}$ 的谱分解为

$$\mathbf{A}^T \mathbf{A} = 1 \times \left[\frac{1}{2} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix} \right] + 3 \times \left[\frac{1}{6} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \right].$$

类似地可得 $\mathbf{A} \mathbf{A}^T$ 的谱分解为

$$\mathbf{A} \mathbf{A}^T = 1 \times \left[\frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \right] + 3 \times \left[\frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right]. \quad (3.5.7)$$

例 3.5.6 如果 \mathbf{A} 是可逆埃尔米特矩阵,则可以利用 \mathbf{A} 的谱分解来求其逆矩阵.

设 \mathbf{A} 的谱分解为

$$\mathbf{A} = \sum_{i=1}^n \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^H,$$

则(证明留给读者)

$$\mathbf{A}^{-1} = \sum_{i=1}^n \frac{1}{\lambda_i} \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^H, \quad (3.5.8)$$

比如例 3.5.5 中的矩阵 $\mathbf{A} \mathbf{A}^T$ 是可逆对称的,故由其谱分解式(3.5.7)及式(3.5.8)可知其逆矩阵为

$$(\mathbf{A} \mathbf{A}^T)^{-1} = 1 \times \left[\frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \right] + \frac{1}{3} \times \left[\frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right] = \frac{1}{3} \times \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

3.5.3 单纯矩阵的谱分解

我们已知道, n 阶方阵当代数重复度与几何重复度相等时, 称之为单纯矩阵, 这样的矩阵可对角化, 但不一定可以酉对角化(即不一定是正规矩阵). 不过, 单纯矩阵也可以类似于正规矩阵定义 \mathbf{A} 的谱分解. 不妨设 $\lambda_1, \lambda_2, \dots, \lambda_n$ 是 \mathbf{A} 的 n 个特征值; $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ 是 \mathbf{A} 的 n 个线性无关的特征向量, 且有

$$\mathbf{A}\mathbf{x}_i = \lambda_i\mathbf{x}_i, \quad i = 1, 2, \dots, n$$

令

$$\mathbf{P} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n), \quad (3.5.9)$$

$$\mathbf{A} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}, \quad (3.5.10)$$

则

$$\mathbf{A} = \mathbf{P}\mathbf{A}\mathbf{P}^{-1}. \quad (3.5.11)$$

把式(3.5.11)两端取转置得

$$\mathbf{A}^T = (\mathbf{P}^T)^{-1}\mathbf{A}\mathbf{P}^T. \quad (3.5.12)$$

这表明 \mathbf{A}^T 也与对角矩阵相似. 因此, 设 $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ 是 \mathbf{A}^T 的 n 个线性无关的特征向量, 即

$$\mathbf{A}^T\mathbf{y}_i = \lambda_i\mathbf{y}_i, \quad i = 1, 2, \dots, n, \quad (3.5.13)$$

把上式两端取转置得

$$\mathbf{y}_i^T\mathbf{A} = \lambda_i\mathbf{y}_i^T, \quad i = 1, 2, \dots, n, \quad (3.5.14)$$

根据式(3.5.14), 我们称 \mathbf{y}_i^T 是 \mathbf{A} 的左特征向量, 称 \mathbf{x}_i 是 \mathbf{A} 的右特征向量.

由式(3.5.12)知

$$(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n) = (\mathbf{P}^T)^{-1} = (\mathbf{P}^{-1})^T, \quad (3.5.15)$$

把式(3.5.15)两端转置得

$$\mathbf{P}^{-1} = \begin{pmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_n^T \end{pmatrix}, \quad (3.5.16)$$

代入 $\mathbf{P}\mathbf{P}^{-1} = \mathbf{P}^{-1}\mathbf{P} = \mathbf{I}$ 得

$$(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \begin{pmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_n^T \end{pmatrix} = \begin{pmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_n^T \end{pmatrix} (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \mathbf{I},$$

此即

$$\mathbf{x}_1\mathbf{y}_1^T + \mathbf{x}_2\mathbf{y}_2^T + \dots + \mathbf{x}_n\mathbf{y}_n^T = \mathbf{I}. \quad (3.5.17)$$

比较两端即有

$$\mathbf{y}_i^T\mathbf{x}_j = \delta_{ij}, \quad i, j = 1, 2, \dots, n, \quad (3.5.18)$$

式(3.5.18)表明,矩阵 \mathbf{A} 的左特征向量 \mathbf{y}_i^T 与右特征向量 \mathbf{x}_i 正交($i \neq j$). 把式(3.5.9)与式(3.5.16)代入式(3.5.11)得

$$\begin{aligned} \mathbf{A} &= (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_n^T \end{pmatrix} \\ &= \lambda_1 \mathbf{x}_1 \mathbf{y}_1^T + \lambda_2 \mathbf{x}_2 \mathbf{y}_2^T + \dots + \lambda_n \mathbf{x}_n \mathbf{y}_n^T. \end{aligned}$$

令

$$\mathbf{G}_i = \mathbf{x}_i \mathbf{y}_i^T, \quad (3.5.19)$$

则得

$$\mathbf{A} = \sum_{i=1}^n \lambda_i \mathbf{G}_i, \quad (3.5.20)$$

式(3.5.20)称为单纯矩阵 \mathbf{A} 的谱分解. 即 \mathbf{A} 分解成 n 个矩阵 \mathbf{G}_i 之和的形式, 其线性组合系数是 \mathbf{A} 的谱(所有的特征值).

定理 3.5.4 设 \mathbf{A} 是 n 阶单纯矩阵, $\lambda_1, \lambda_2, \dots, \lambda_r$ 是 \mathbf{A} 的 r 个相异的特征值, 则 \mathbf{A} 可以进行满足下列性质的谱分解:

$$(1) \mathbf{A} = \sum_{j=1}^r \lambda_j \mathbf{E}_j; \quad (3.5.21)$$

$$(2) \mathbf{E}_j^2 = \mathbf{E}_j \quad (j = 1, 2, \dots, r); \quad (3.5.22)$$

$$(3) \mathbf{E}_i \mathbf{E}_j = \mathbf{0} \quad (i \neq j, i, j = 1, 2, \dots, r); \quad (3.5.23)$$

$$(4) \sum_{j=1}^r \mathbf{E}_j = \mathbf{I}. \quad (3.5.24)$$

证明 设对应于 λ_j 的线性无关的右特征向量为 $\mathbf{x}_1^j, \mathbf{x}_2^j, \dots, \mathbf{x}_{s_j}^j$, 其中 s_j 为对应 λ_j 的代数重复度或几何重复度, 而左特征向量为 $(\mathbf{y}_1^j)^T, \dots, (\mathbf{y}_{s_j}^j)^T$, 代入式(3.5.20)有

$$\mathbf{A} = \sum_{j=1}^r \lambda_j \left(\sum_{k=1}^{s_j} \mathbf{x}_k^j (\mathbf{y}_k^j)^T \right) = \sum_{j=1}^r \lambda_j \mathbf{E}_j,$$

即式(3.5.21)得证. 其中

$$\mathbf{E}_j = \sum_{k=1}^{s_j} \mathbf{x}_k^j (\mathbf{y}_k^j)^T = (\mathbf{x}_1^j, \mathbf{x}_2^j, \dots, \mathbf{x}_{s_j}^j) \begin{pmatrix} (\mathbf{y}_1^j)^T \\ \vdots \\ (\mathbf{y}_{s_j}^j)^T \end{pmatrix}.$$

根据式(3.5.17)得

$$\sum_{j=1}^r \mathbf{E}_j = \mathbf{I},$$

即式(3.5.24)得证.

再由式(3.5.18)得

$$\mathbf{E}_j^2 = \mathbf{E}_j,$$

$$\mathbf{E}_i \mathbf{E}_j = \mathbf{0},$$

即式(3.5.22)、式(3.5.23)得证.

证毕

这些矩阵 \mathbf{E}_j 称为矩阵 \mathbf{A} 的(谱分解的)成分矩阵或幂等矩阵. 注意, 与正规矩阵相比,

一般矩阵的谱分解中的成分矩阵不一定是 Hermite 矩阵. 因此 $Ax = \sum_{j=1}^r \lambda_j E_j x$ 中的诸向量 $E_j x$ 未必是正交的.

例 3.5.7 求单纯矩阵 $A = \begin{bmatrix} -1 & 3 & -1 \\ -3 & 5 & -1 \\ -3 & 3 & 1 \end{bmatrix}$ 的谱分解.

解

$$|\lambda I - A| = \begin{vmatrix} \lambda + 1 & -3 & 1 \\ 3 & \lambda - 5 & 1 \\ 3 & -3 & \lambda - 1 \end{vmatrix} = (\lambda - 1)(\lambda - 2)^2,$$

所以 A 有特征值 $\lambda_1 = 1, \lambda_2 = 2$ (二重), 对应于 λ_1, λ_2 的线性无关的特征向量分别为

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix}.$$

令 $P = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$, 则可求出

$$P^{-1} = \begin{bmatrix} 3 & -3 & 1 \\ -3 & 4 & -1 \\ -1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1^T \\ \mathbf{y}_2^T \\ \mathbf{y}_3^T \end{bmatrix},$$

因此

$$\begin{aligned} E_1 &= \mathbf{x}_1 \mathbf{y}_1^T = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} (3, -3, 1) = \begin{bmatrix} 3 & -3 & 1 \\ 3 & -3 & 1 \\ 3 & -3 & 1 \end{bmatrix}, \\ E_2 &= \mathbf{x}_2 \mathbf{y}_2^T = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} (-3, 4, -1) + \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix} (-1, 1, 0) \\ &= \begin{bmatrix} -3 & 4 & -1 \\ -3 & 4 & -1 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ -3 & 3 & 0 \end{bmatrix} = \begin{bmatrix} -2 & 3 & -1 \\ -3 & 4 & -1 \\ -3 & 3 & 0 \end{bmatrix}, \end{aligned}$$

则 A 的谱分解为 $A = E_1 + 2E_2$, 注意, 矩阵 E_1 与 E_2 的第一列不正交, 故向量 $E_1 x$ 与 $E_2 x$ 不正交.

推论 设 $A = \sum_{j=1}^r \lambda_j E_j$ 是单纯矩阵 A 的谱分解, 则

$$A^m = \sum_{j=1}^r \lambda_j^m E_j, \quad (3.5.25)$$

从而对任意多项式 $f(x)$, 有 $f(A) = \sum_{j=1}^r f(\lambda_j) E_j$.

例 3.5.8 设 A 如例 3.5.7, 求 $e^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n$.

解 利用定理 3.5.4 的(2), (3), (4), 有

$$A^n = (E_1 + 2E_2)^n = E_1 + 2^n E_2, \quad \forall n \geq 0,$$

从而

$$e^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n = eE_1 + e^2 E_2.$$

将在第5章把上述推论推广到一般函数,从而利用谱分解给出了一个求矩阵函数的简便方法.

习 题 3

1. 判定矩阵 $C = \begin{pmatrix} 3 & 2 & -1 \\ -1 & 0 & 0 \\ -1 & 3 & 0 \end{pmatrix}$ 和 $B = \begin{pmatrix} 0 & 2 & -1 \\ -1 & 4 & -1 \\ 1 & 3 & -5 \end{pmatrix}$ 能否进行 LU 分解,为什么?若能分解,试分解

之.

2. 对下列矩阵进行杜利特分解:

(1) $A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{pmatrix}$; (2) $B = \begin{pmatrix} 12 & -3 & 3 \\ -18 & 3 & -1 \\ 1 & 1 & 1 \end{pmatrix}$.

3. 求矩阵 $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix}$ 的 LU 分解.

4. 求矩阵

$$A = \begin{pmatrix} 5 & 2 & -4 & 0 \\ 2 & 1 & -2 & 1 \\ -4 & -2 & 5 & 0 \\ 0 & 1 & 0 & 2 \end{pmatrix}$$

的杜利特分解与克劳特分解.

5. 求对称正定矩阵

$$A = \begin{pmatrix} 5 & 2 & -4 \\ 2 & 1 & -2 \\ -4 & -2 & 5 \end{pmatrix}$$

的不带平方根的乔累斯基分解.

6. 对下面矩阵进行 LDU 分解:

(1) $A = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$; (2) $A = \begin{pmatrix} 1 & 2 & 3 & -1 \\ 2 & -1 & 9 & -7 \\ -3 & 4 & -3 & 19 \\ 4 & -1 & 6 & -21 \end{pmatrix}$,

其中 L 为单位下三角矩阵, D 为对角矩阵, U 为单位上三角矩阵.

7. (1) 对 $A = \begin{pmatrix} 16 & 4 & 8 \\ 4 & 5 & -4 \\ 8 & -4 & 22 \end{pmatrix}$ 进行 LDL^T 分解和乔累斯基分解 GG^T , 其中 L 为单位下三角矩阵, D 为

对角矩阵, G 为下三角矩阵;

(2) 对 $A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{pmatrix}$ 进行克劳特分解和不带平方根的乔累斯基分解.

8. 分解 4 阶希尔伯特矩阵 $H_4 = (h_{ij}) = (i+j-1)^{-1}$ 为

(1) LDL^T 形式, 其中 L 为单位下三角矩阵, D 为对角矩阵;

(2) LL^T 的形式(楚列斯基分解).

9. 设 A 是秩为 r 的 n 阶矩阵, 且 $\det A_k \neq 0 (k=1, 2, \dots, r)$, 证明: A 可进行三角分解 $A=LU$, 且可使得 L 或 U 为可逆矩阵.

10. 将正定矩阵

$$A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 5 & -1 \\ -1 & -1 & 5 \end{bmatrix}$$

做楚列斯基(Cholesky)分解 $A=U^T U$.

11. 求下列矩阵的正交三角(UR)分解表达式

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

12. 求矩阵

$$A = \begin{bmatrix} 1 & \frac{1}{2} & 5 \\ 1 & -\frac{1}{2} & 2 \\ -1 & \frac{1}{2} & -2 \\ 1 & -\frac{3}{2} & 0 \end{bmatrix}$$

的 QR 分解.

13. 设 A 是 n 阶可逆实矩阵, 则 A 可表成一个正交矩阵 Q 与正定矩阵 S 的乘积, 即 $A=QS$.

14. A 为实 n 阶矩阵, 存在一个正交矩阵 Q , 使 A 正交相似于分块上三角矩阵(拟上三角矩阵), 即

$$Q^T A Q = Q^{-1} A Q = R,$$

其中 R 为分块上三角矩阵, 其对角元上对角块矩阵为 1 阶或 2 阶块方阵. 每一个 1 阶块是 A 的实特征值, 而每一个 2 阶实矩阵的两个特征值是 A 的一对共轭的特征值.

15. 用吉文斯变换求矩阵

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

的 QR 分解.

16. 用豪斯霍尔德变换求矩阵

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

的 QR 分解.

17. 已知矩阵

$$A = \begin{bmatrix} 1 & -2 & 6 & 0 \\ 2i & -4i & 0 & 0 \\ -2i & 4i & 0 & 15 \\ 0 & 0 & 3 & 0 \end{bmatrix},$$

分别用豪斯霍尔德矩阵和吉文斯矩阵(又称平面旋转矩阵)求 A 的 QR 分解.

18. 用施密特正交化方法求矩阵

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

的 QR 分解.

19. 用豪斯霍尔德变换使矩阵

$$A = \begin{pmatrix} 0 & 12 & 16 \\ 12 & 288 & 309 \\ 16 & 309 & 312 \end{pmatrix},$$

正交相似于三对角矩阵.

20. 试用三种方法(豪斯霍尔德、吉文斯、施密特)求矩阵

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

的 QR 分解.

21. 已知实对称矩阵

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

试用豪斯霍尔德矩阵和吉文斯矩阵使 A 正交相似于实对称三对角矩阵.

22. 对下列矩阵做满秩分解:

$$(1) A = \begin{bmatrix} 1 & 1 & 2 & 2 \\ -1 & 1 & 2 & -4 \\ 1 & 4 & 8 & 2 \end{bmatrix}; \quad (2) A = \begin{bmatrix} 1 & -1 & 2 & 3 \\ -1 & 0 & -1 & 0 \\ 3 & 2 & -1 & -6 \\ 0 & -1 & 1 & 3 \end{bmatrix}.$$

23. 对 A 做满秩分解

$$A = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 1 & 2 & -1 & 1 \\ 2 & 2 & -2 & -1 \\ -2 & -4 & 2 & -2 \end{bmatrix}.$$

24. 求矩阵 A 的满秩分解表达式:

$$(1) A = \begin{bmatrix} 2 & 1 & -2 & 3 & 1 \\ 2 & 5 & -1 & 4 & 1 \\ 1 & 3 & -1 & 2 & 1 \end{bmatrix}; \quad (2) A = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 2 & 3 & 1 & 3 & 1 \end{bmatrix};$$

$$(3) A = \begin{bmatrix} 1 & 2 & 1 & 0 & 1 & 2 \\ 1 & 2 & 2 & 1 & 3 & 3 \\ 2 & 4 & 3 & 1 & 4 & 5 \\ 4 & 8 & 6 & 2 & 8 & 10 \end{bmatrix}; \quad (4) A = \begin{bmatrix} 1 & 2 & 0 & 1 & 1 & 10 \\ 3 & 6 & 1 & 4 & 2 & 36 \\ 2 & 4 & 0 & 2 & 2 & 27 \\ 6 & 12 & 1 & 7 & 5 & 73 \end{bmatrix}.$$

25. 设矩阵 A 的满秩分解为 $A=BC$, 证明:

$$CX = 0 \Leftrightarrow AX = 0.$$

26. 设 A 的任两个分解为 $A=BC=B_1C_1$, 则它们之间的关系为

$$B = B_1M, \quad C = M^{-1}C_1.$$

其中 M 是 $r \times r$ 阶可逆矩阵.

27. 设 $A \in \mathbb{C}^{m \times n}$, A 的分块为

$$A = \begin{bmatrix} X & Y \\ Z & W \end{bmatrix},$$

其中 $X \in \mathbb{C}^{r \times r}$, $\text{rank}(X) = \text{rank}(A) = r$, 证明 A 有如下形式的满秩分解:

$$A = \begin{bmatrix} X \\ Z \end{bmatrix} [I_r \mid X^{-1}Y] = \begin{bmatrix} I_r \\ ZX^{-1} \end{bmatrix} [X, Y].$$

28. 设 $A^2 = A$ 且 A 满足满秩分解 $A=BC$, 证明 $CB=I$.

29. 设 $A \in \mathbb{R}^{m \times n}$, $\text{rank}(A) = r$, 则存在 $m \times r$ 矩阵 W 和 $r \times n$ 矩阵 R , 使

$$A_{m \times n} = W_{m \times r} R_{r \times n},$$

其中 R 为行满秩矩阵, W 是 r 个标准正交列向量组成的列满秩矩阵, 称为正交满秩分解.

30. 设 $A \in \mathbb{R}^{n \times n}$ 的满秩分解为 $A=BC$, 则齐次方程组 $AX=0$ 的解与齐次方程组 $CX=0$ 同解.

31. 设 A, B, C 分别为 $m \times n, n \times k, k \times p$ 阶矩阵, 证明:

$$\text{rank}(ABC) \geq \text{rank}(AB) + \text{rank}(BC) - \text{rank}(B).$$

32. $A \in \mathbb{R}^{n \times m}, B \in \mathbb{R}^{m \times n}$, 若 $BA=I$, 则称 B 为 A 的左逆矩阵(见第6章定义). 证明: A 有左逆矩阵的充分必要条件 A 为列满秩矩阵.

33. 设矩阵 $F \in \mathbb{C}_r^{m \times r}$ (列满秩), $G \in \mathbb{C}_r^{r \times n}$ (行满秩), 证明:

$$\text{rank}(FG) = r.$$

34. 求矩阵 $A = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ 的奇异值分解.

35. 求 $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$ 的奇异值分解.

36. 设 $A = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 2 & 0 \end{bmatrix}$, 求 A 的奇异值分解.

37. 设 A 为埃尔米特正定矩阵, A 的奇异值就是 A 的特征值, 试证明之.

38. A 为 n 阶可逆矩阵, 则 A 的行列式的绝对值是 A 的奇异值之积.

39. 已知

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

求 A 的奇异分解表达式.

40. 已知 $A \in \mathbb{C}^{m \times n}$ (秩为 $r > 0$) 的奇异值分解表达式为

$$A = U \begin{bmatrix} \Delta & 0 \\ 0 & 0 \end{bmatrix} V^H,$$

试求矩阵 $B = \begin{bmatrix} A \\ A \end{bmatrix}$ 的奇异值分解表达式.

41. 设 $A=UDV^H$ 为矩阵 A 的一个奇异值分解.

(1) 证明: U 的列向量为 AA^H 的特征向量, 称其为矩阵 A 的左奇异向量;

(2) 证明: V 的列向量为 AA^H 的特征向量, 称其为矩阵 A 的右特征向量;

(3) 举反例说明依据(1)和(2)中确定的酉矩阵 U 和 V 不一定是 A 的奇异值分解.

42. 设 $A, B \in \mathbb{C}^{m \times n}$, 如果存在 m 阶酉矩阵 U 和 n 阶酉矩阵 V , 使得 $B = UAV$, 那么称 A 与 B 酉相抵, 证明:

(1) 酉相抵是一种等价关系;

(2) 若 A 与 B 酉相抵, 则 A 与 B 有相同的奇异值.

43. 设 $A \in \mathbb{C}^{m \times n}$, U 和 V 分别为 m, n 阶酉矩阵, 试证: UA 和 AV 的奇异值与 A 的奇异值相同.

44. 设 $A \in \mathbb{C}^{m \times n}$,

(1) 证明: $A^H A$ 与 AA^H 的非零特征值相同;

(2) 设 $A^H A$ 的非零特征值 $\lambda_1, \lambda_2, \dots, \lambda_r$ 对应的正交的特征向量为 $\alpha_1, \alpha_2, \dots, \alpha_r$, 则 AA^H 的特征值 $\lambda_1, \lambda_2, \dots, \lambda_r$ 对应的特征向量为 $A\alpha_1, A\alpha_2, \dots, A\alpha_r$, 且它们也是正交向量组.

45. $A \in \mathbb{C}^{m \times n}$, U, V 分别为 m, n 阶酉矩阵, 证明: UA, AV 的奇异值与 A 的奇异值相同.

46. $A \in \mathbb{C}^{m \times n}$, U, V 分别为 m, n 阶酉矩阵, 若 $B = UAV$, 称 A 与 B 酉等价, 证明: B 与 A 的奇异值相同.

47. 设 $A \in \mathbb{C}^{n \times n}$, 用 A 的奇异值分解证明 A 的极分解:

$$A = GU = UH,$$

其中 G, H 为半正定埃尔米特矩阵, U 为酉矩阵.

48. 设矩阵

$$(1) A = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 3 \end{bmatrix}; \quad (2) B = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix},$$

验证 A 与 B 是正规矩阵, 并求 A 与 B 的谱分解表达式.

49. 已知矩阵

$$A = \begin{bmatrix} 0 & 2 & 4 \\ \frac{1}{2} & 0 & 2 \\ \frac{1}{4} & \frac{1}{2} & 0 \end{bmatrix},$$

验证 A 是单纯矩阵, 并求 A 的谱分解表达式.

50. 试求一酉矩阵 P , 使 $P^{-1}AP$ 为对角矩阵.

$$(1) A = \begin{bmatrix} -1 & i & 0 \\ -i & 0 & -i \\ 0 & i & -1 \end{bmatrix}; \quad (2) A = \begin{bmatrix} 0 & i & 1 \\ -i & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

这两个矩阵是正规矩阵吗?

51. 证明: 若 A 是正规矩阵, 则 A 的奇异值就是 A 的特征值的模.

52. 证明: 两个正规矩阵相似(酉等价)的充分必要条件是它们的特征多项式相同.

53. 设 A 为正规矩阵, 则 A 的谱分解式有

$$A = \lambda_1 U_1 U_1^H + \lambda_2 U_2 U_2^H + \dots + \lambda_n U_n U_n^H,$$

其中 U_1, U_2, \dots, U_n 是 A 的 n 个特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 对应的标准正交的特征向量.

54. $A, B \in \mathbb{C}^{m \times n}$, 且都是埃尔米特矩阵, A 是半正定矩阵, 证明: $\text{tr}(AB) \geq u_n(\text{tr}(A))$, 其中 u_n 为 B 的最小特征值, $\text{tr}(AB)$ 表示 AB 主对角元素之和.

55. 设矩阵 $A = (a_{ij})_{n \times n}$ 与 $B = (b_{ij})_{n \times n}$ 酉相似, 证明

$$(1) \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 = \sum_{i=1}^n \sum_{j=1}^n |b_{ij}|^2;$$

(2) 设 $\lambda_1, \lambda_2, \dots, \lambda_n$ 为 A 的 n 个特征值, 则

$$\sum_{i=1}^n |\lambda_i|^2 \leq \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2;$$

(3) 若 A 为正规矩阵, 则

$$\sum_{i=1}^n |\lambda_i|^2 = \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2.$$

56. (1) 设 A 为酉矩阵且是埃尔米特矩阵, 则 A 的特征值为 1 或 -1;

(2) 若 A 是正规矩阵, 且 A 的特征值 $|\lambda| = 1$, 则 A 是酉矩阵.

57. A 为 n 阶正规矩阵, $\lambda_i (i=1, 2, \dots, n)$ 是 A 的特征值, 证明 $A^H A$ 与 AA^H 的特征值为 $|\lambda_i|^2, i=1, 2, \dots, n$.

58. 设 A 为正规矩阵, 证明:

(1) 若对于正数 m , 有 $A^m = \mathbf{0}$, 则 $A = \mathbf{0}$;

(2) 若 $A^2 = A$, 则 $A^H = A$;

(3) 若 $A^3 = A^2$, 则 $A^2 = A$.

59. $A = (a_{ij})_{n \times n}, \lambda_1, \lambda_2, \dots, \lambda_n$ 为 A 的特征值, 若 $\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 = \sum_{i=1}^n |\lambda_i|^2$, 证明 A 是正规矩阵.

60. 若 A 为实矩阵, 且 $A^T A = AA^T$, 则 A 必是对称矩阵.

61. A 是正规矩阵, 证明:

(1) A 的特征向量也是 A^H 的特征向量;

(2) $\forall X \in \mathbb{C}^n, AX$ 与 $A^H X$ 的长度相等.

62. n 阶方阵 A 正规的充分必要条件是它与一个具有互异的特征值且与 A 有相同的特征向量的矩阵

B 可交换 (即 $AB = BA$).

63. 设 $A, B \in \mathbb{C}^{n \times n}, A, B$ 为正规矩阵, 且 $AB = BA$, 则存在 n 阶酉阵 U , 使得

$$U^H A U = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}, \quad U^H B U = \begin{bmatrix} u_1 & & \\ & \ddots & \\ & & u_n \end{bmatrix}.$$

64. 若两个正规矩阵可交换, 证明它们的乘积也是正规矩阵.

引言 范数是什么

我们在第1章的内积空间中已经看到,距离概念可以由长度导出,而长度可以由内积导出.但研究 $m \times n$ 矩阵函数构成的线性空间是无限维的,当然没有内积的概念,因此我们将在本章研究比用内积导出的长度更为广泛的概念——范数,以使范数能够应用在更广的范围.那么,什么是矩阵的范数呢?我们先看下面的例子:

设 x 是复数,则当 $|x| < 1$ 时有

$$(1-x)^{-1} = 1+x+x^2+\cdots+x^m+\cdots,$$

何时上式对于矩阵也成立,即设 \mathbf{A} 是 n 阶矩阵,要使公式

$$(\mathbf{I}-\mathbf{A})^{-1} = \mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^m + \cdots$$

成立,相当于 $|x| < 1$ 的条件是什么?在下章将容易知道一个充分条件是 $\rho(\mathbf{A}) < 1$,但计算矩阵的谱半径 $\rho(\mathbf{A})$ 又与矩阵的特征值有关,这是不容易计算的.我们能否改进这个条件?答案是肯定的,只需将 $\rho(\mathbf{A}) < 1$ 换成 $\|\mathbf{A}\| < 1$,即 \mathbf{A} 的范数小于1,任何一种范数即可!因此,矩阵的范数可以看作是实数的绝对值或者复数的模的推广,也是向量长度的推广,是一种衡量矩阵(包括向量)大小的尺度.

在计算数学中,特别是在数值代数中,研究数值方法的收敛性、稳定性及误差分析等问题时,范数理论都扮演着十分重要的角色.因此,本章首先在线性空间中定义向量的范数,引出赋范线性空间的概念;然后进一步讨论矩阵的范数及其性质,以及与范数有关的矩阵谱半径、条件数等.

4.1 赋范线性空间

4.1.1 向量的范数

向量的范数是用来刻画向量大小的一种度量.在第1章的内积空间中,由于用内积定义了向量的长度 $|x| = \sqrt{(x,x)}$,用长度来表示 x 的大小可带来许多方便.把这种长度的概念进一步推广,这就是所谓范数的概念,先看下面的例子.

例 4.1.1 设有平面向量 $x = ai + bj$,记长度为 $\|x\| = \sqrt{a^2 + b^2}$,那么 $\|x\|$ 具有下面3条性质:

- (1) 若 $\mathbf{x} \neq \mathbf{0}$, 则 $\|\mathbf{x}\| > 0$; 当且仅当 $\mathbf{x} = \mathbf{0}$ 时, 有 $\|\mathbf{x}\| = 0$.
 (2) $\|k\mathbf{x}\| = |k| \|\mathbf{x}\|$, k 为任意实数.
 (3) 对于任意平面向量 \mathbf{x} 和 \mathbf{y} , 有三角不等式

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|.$$

对于一般的线性空间, 引入满足上述 3 条性质的纯量(或函数), 即可用它来描述向量的大小, 称之为范数.

定义 4.1.1 如果 V 是数域 P 上的线性空间, 且对于 V 的任一向量 \mathbf{x} , 对应着一个实值函数 $\|\mathbf{x}\|$, 它满足以下 3 个条件.

- (1) 非负性: 当 $\mathbf{x} \neq \mathbf{0}$ 时, $\|\mathbf{x}\| > 0$; 当且仅当 $\mathbf{x} = \mathbf{0}$ 时, $\|\mathbf{x}\| = 0$;
 (2) 齐次性: $\|k\mathbf{x}\| = |k| \|\mathbf{x}\|$, $k \in P$;
 (3) 三角不等式: $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$, $\mathbf{x}, \mathbf{y} \in V$;

则称 $\|\mathbf{x}\|$ 为 V 上向量 \mathbf{x} 的范数(norm).

例 4.1.2 设 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$, 它的长度 $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$ 就是一种范数.

为了说明这里的 $\|\mathbf{x}\|$ 是范数, 只需验证它满足 3 个条件就行了.

(1) 当 $\mathbf{x} \neq \mathbf{0}$ 时, 至少有一分量不为 0, 所以 $\|\mathbf{x}\| > 0$; 当 $\mathbf{x} = \mathbf{0}$ 时, $\|\mathbf{x}\| = \sqrt{0^2 + \dots + 0^2} = 0$.

(2) 对任意数 $k \in \mathbb{R}$, 因为 $k\mathbf{x} = (kx_1, kx_2, \dots, kx_n)$, 所以 $\|k\mathbf{x}\| = \sqrt{(kx_1)^2 + \dots + (kx_n)^2} = |k| \sqrt{x_1^2 + \dots + x_n^2} = |k| \|\mathbf{x}\|$.

(3) 对任意两个向量 $\mathbf{x} = (x_1, \dots, x_n)$, $\mathbf{y} = (y_1, \dots, y_n)$, 有

$$\begin{aligned} \mathbf{x} + \mathbf{y} &= (x_1 + y_1, \dots, x_n + y_n), \\ \|\mathbf{x} + \mathbf{y}\|^2 &= (x_1 + y_1)^2 + \dots + (x_n + y_n)^2 \\ &= x_1^2 + \dots + x_n^2 + 2(x_1 y_1 + \dots + x_n y_n) + y_1^2 + \dots + y_n^2. \end{aligned}$$

根据 \mathbb{R}^n 中内积的定义有

$$(\mathbf{x}, \mathbf{y}) = x_1 y_1 + x_2 y_2 + \dots + x_n y_n.$$

再利用第 1 章柯西-施瓦茨不等式有

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^2 &= \|\mathbf{x}\|^2 + 2(\mathbf{x}, \mathbf{y}) + \|\mathbf{y}\|^2 \\ &\leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 \\ &= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2, \end{aligned}$$

从而得

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|.$$

这就证明了 $\|\mathbf{x}\| = \sqrt{x_1^2 + \dots + x_n^2}$ 是 \mathbb{R}^n 上的一种范数. 这种范数称为 2-范数或欧氏范数, 记为

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + \dots + x_n^2} = (\mathbf{x}^T \mathbf{x})^{\frac{1}{2}}. \quad (4.1.1)$$

同理, 对于复向量 $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{C}^n$, 2-范数的形式为

$$\|\mathbf{x}\|_2 = \sqrt{|x_1|^2 + \dots + |x_n|^2} = (\mathbf{x}^H \mathbf{x})^{\frac{1}{2}} \quad (4.1.1)'$$

其实, n 维向量空间中的范数并不是惟一的.

例 4.1.3 验证 $\|x\| = \max_i |x_i|$ 是 \mathbb{R}^n 上的一种范数, 这里 $x = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$.

事实上, 当 $x \neq 0$ 时, $\|x\| = \max_i |x_i| > 0$; 当 $x = 0$ 时, $\|x\| = 0$. 又对任意 $k \in \mathbb{R}$, 有

$$\|kx\| = \max_i |kx_i| = |k| \max_i |x_i| = |k| \|x\|,$$

对 \mathbb{R}^n 中任意两个向量 $x = (x_1, x_2, \dots, x_n)^T, y = (y_1, y_2, \dots, y_n)^T$, 有

$$\begin{aligned} \|x+y\| &= \max_i |x_i + y_i| \leq \max_i |x_i| + \max_i |y_i| \\ &= \|x\| + \|y\|, \end{aligned}$$

因此 $\|x\| = \max_i |x_i|$ 确实是 \mathbb{R}^n 上的一种范数.

我们称例 4.1.3 中的范数为 ∞ -范数, 记为

$$\|x\|_\infty = \max_i |x_i|. \quad (4.1.2)$$

例 4.1.4 验证 $\|x\| = \sum_{i=1}^n |x_i|$ 也是 \mathbb{R}^n 上的一种范数, 其中 $x = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$.

事实上, 当 $x \neq 0$ 时, 显然有 $\|x\| = \sum_{i=1}^n |x_i| > 0$; 当 $x = 0$ 时, $\|x\| = 0 + \dots + 0 = 0$.

又对于任意 $k \in \mathbb{R}$, 有

$$\|kx\| = \sum_{i=1}^n |kx_i| = |k| \sum_{i=1}^n |x_i| = |k| \|x\|.$$

对任意两个向量 $x, y \in \mathbb{R}^n$, 有

$$\begin{aligned} \|x+y\| &= \sum_{i=1}^n |x_i + y_i| \leq \sum_{i=1}^n (|x_i| + |y_i|) \\ &= \sum_{i=1}^n |x_i| + \sum_{i=1}^n |y_i| = \|x\| + \|y\|, \end{aligned}$$

因此知 $\|x\| = \sum_{i=1}^n |x_i|$ 是 \mathbb{R}^n 上的一种范数, 称它为 1-范数, 记为

$$\|x\|_1 = \sum_{i=1}^n |x_i|. \quad (4.1.3)$$

由例 4.1.2~例 4.1.4 可知, 在一个线性空间中, 可以定义多种向量范数, 实际上可以定义无限多种范数. 例如对于不小于 1 的任意实数 p 及 $x = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$, 可以证明实值函数

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}, \quad 1 \leq p < +\infty \quad (4.1.4)$$

都是 \mathbb{R}^n 中的范数.

为此, 先引入以下两个引理.

引理 4.1.1 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

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}. \quad (4.1.5)$$

* 证明 若 $a=0$ 或 $b=0$, 则式(4.1.5)显然成立. 下面考虑 a, b 均为正数的情况.

对 $x > 0, 0 < \mu < 1$, 记 $f(x) = x^\mu - \mu x$. 容易验证 $f(x)$ 在 $x=1$ 处达到最大值 $1-\mu$, 从而 $f(x) \leq 1-\mu$, 即

$$x^\mu \leq 1 - \mu + \mu x.$$

对任意正实数 c, d , 在上式中令 $x = \frac{c}{d}, \mu = \frac{1}{p}, 1 - \mu = \frac{1}{q}$, 则 $c^{\frac{1}{p}} d^{\frac{1}{q}} \leq \frac{c}{p} + \frac{d}{q}$. 由此再令 $a = c^{\frac{1}{p}}, b = d^{\frac{1}{q}}$, 即得式(4.1.5).

引理 4.1.2 设 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T, \mathbf{y} = (y_1, y_2, \dots, y_n)^T \in \mathbb{R}^n$, 则

$$\sum_{i=1}^n |x_i y_i| \leq \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^n |y_i|^q \right)^{\frac{1}{q}}, \quad (4.1.6)$$

其中实数 $p > 1, q > 1$ 且 $\frac{1}{p} + \frac{1}{q} = 1$. 此时, 式(4.1.6)称**赫尔德(Hölder)不等式**.

***证明** 如果 $\mathbf{x} = \mathbf{0}$ 或 $\mathbf{y} = \mathbf{0}$, 则式(4.1.6)显然成立. 下面设 $\mathbf{x} \neq \mathbf{0}, \mathbf{y} \neq \mathbf{0}$. 令

$$a = \frac{|x_i|}{\left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}}, \quad b = \frac{|y_i|}{\left(\sum_{i=1}^n |y_i|^q \right)^{\frac{1}{q}}},$$

则由式(4.1.5)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

$$\frac{\sum_{i=1}^n |x_i y_i|}{\left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^n |y_i|^q \right)^{\frac{1}{q}}} \leq \frac{\sum_{i=1}^n |x_i|^p}{p \left(\sum_{i=1}^n |x_i|^p \right)} + \frac{\sum_{i=1}^n |y_i|^q}{q \left(\sum_{i=1}^n |y_i|^q \right)} = \frac{1}{p} + \frac{1}{q} = 1.$$

由上式即得式(4.1.6).

证毕

定理 4.1.1 对任意向量 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n, 1 \leq p < +\infty$, 由式(4.1.4)定义的 $\|\mathbf{x}\|_p$ 是 \mathbb{R}^n 上的向量范数.

***证明** (1) 当 $\mathbf{x} \neq \mathbf{0}$ 时, \mathbf{x} 至少有一个分量不为零, 故 $\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} > 0$;

(2) 对 $\forall k \in \mathbb{R}, \forall \mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$, 则

$$\begin{aligned} \|k\mathbf{x}\|_p &= \|(kx_1, kx_2, \dots, kx_n)\|_p \\ &= \left(\sum_{i=1}^n |kx_i|^p \right)^{\frac{1}{p}} = |k| \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \\ &= |k| \|\mathbf{x}\|_p; \end{aligned}$$

(3) 对 $\forall \mathbf{x} = (x_1, x_2, \dots, x_n)^T, \mathbf{y} = (y_1, y_2, \dots, y_n)^T \in \mathbb{R}^n$, 下面设 $p > 1$, 记 $q = \frac{p}{p-1}$, 则

$q > 1$, 且 $\frac{1}{p} + \frac{1}{q} = 1$. 从而由式(4.1.6)有

$$\begin{aligned} \sum_{i=1}^n |x_i + y_i|^p &= \sum_{i=1}^n |x_i + y_i| |x_i + y_i|^{p-1} \\ &\leq \sum_{i=1}^n |x_i| |x_i + y_i|^{p-1} + \sum_{i=1}^n |y_i| |x_i + y_i|^{p-1} \end{aligned}$$

$$\begin{aligned}
&\leq \left(\sum_{i=1}^n |x_i|^\rho \right)^{\frac{1}{\rho}} \left(\sum_{i=1}^n |x_i + y_i|^{(\rho-1)q} \right)^{\frac{1}{q}} + \\
&\quad \left(\sum_{i=1}^n |y_i|^\rho \right)^{\frac{1}{\rho}} \left(\sum_{i=1}^n |x_i + y_i|^{(\rho-1)q} \right)^{\frac{1}{q}} \\
&= \left[\left(\sum_{i=1}^n |x_i|^\rho \right)^{\frac{1}{\rho}} + \left(\sum_{i=1}^n |y_i|^\rho \right)^{\frac{1}{\rho}} \right] \left(\sum_{i=1}^n |x_i + y_i|^{(\rho-1)q} \right)^{\frac{1}{q}} \\
&= \left[\left(\sum_{i=1}^n |x_i|^\rho \right)^{\frac{1}{\rho}} + \left(\sum_{i=1}^n |y_i|^\rho \right)^{\frac{1}{\rho}} \right] \left(\sum_{i=1}^n |x_i + y_i|^\rho \right)^{\frac{1}{q}},
\end{aligned}$$

因此

$$\left(\sum_{i=1}^n |x_i + y_i|^\rho \right)^{\frac{1}{\rho}} \leq \left(\sum_{i=1}^n |x_i|^\rho \right)^{\frac{1}{\rho}} + \left(\sum_{i=1}^n |y_i|^\rho \right)^{\frac{1}{\rho}},$$

即

$$\|\mathbf{x} + \mathbf{y}\|_\rho \leq \|\mathbf{x}\|_\rho + \|\mathbf{y}\|_\rho. \quad \text{证毕}$$

因此 $\|\mathbf{x}\|_\rho$ 是 \mathbb{R}^n 中的一种范数. 我们称式(4.1.4)为向量的 p -范数或称 l_p 范数.

其实,我们可以把3种常用范数 $\|\mathbf{x}\|_1$, $\|\mathbf{x}\|_2$, $\|\mathbf{x}\|_\infty$ 统一在 p -范数之中. 因为在式(4.1.4)中,当 $p=1$ 时,显然 $\|\mathbf{x}\|_p = \|\mathbf{x}\|_1$; 当 $p=2$ 时,显然 $\|\mathbf{x}\|_p = \|\mathbf{x}\|_2$; p 趋向于 ∞ 时,有

$$\lim_{p \rightarrow \infty} \|\mathbf{x}\|_p = \|\mathbf{x}\|_\infty. \quad (4.1.7)$$

对于式(4.1.7)证明如下:

事实上,将 $|x_1|, |x_2|, \dots, |x_n|$ 中最大者记为 $|x_{i_0}|$ ($\neq 0$). 由于 $\|\mathbf{x}\|_\infty = \max_i |x_i| = |x_{i_0}|$, 而

$$\begin{aligned}
\|\mathbf{x}\|_p &= \left(\sum_{i=1}^n |x_{i_0}|^\rho \frac{|x_i|^\rho}{|x_{i_0}|^\rho} \right)^{\frac{1}{\rho}} \\
&= |x_{i_0}| \left(\sum_{i=1}^n \frac{|x_i|^\rho}{|x_{i_0}|^\rho} \right)^{\frac{1}{\rho}},
\end{aligned}$$

而

$$|x_{i_0}|^\rho \leq \sum_{i=1}^n |x_i|^\rho \leq n |x_{i_0}|^\rho,$$

两边同时除以 $|x_{i_0}|^\rho$ 并开 p 次方得

$$1 \leq \left(\sum_{i=1}^n \frac{|x_i|^\rho}{|x_{i_0}|^\rho} \right)^{\frac{1}{\rho}} \leq n^{\frac{1}{\rho}}.$$

由于 $\lim_{p \rightarrow \infty} n^{\frac{1}{p}} = 1$, 从而有

$$\lim_{p \rightarrow \infty} \left(\sum_{i=1}^n \frac{|x_i|^\rho}{|x_{i_0}|^\rho} \right)^{\frac{1}{\rho}} = 1,$$

故

$$\lim_{p \rightarrow \infty} \|\mathbf{x}\|_p = |x_{i_0}| = \|\mathbf{x}\|_\infty.$$

这样, \mathbb{R}^n 中的范数 $\|\mathbf{x}\|_1, \|\mathbf{x}\|_2, \|\mathbf{x}\|_\infty$ 可依次写为 $1, 2, \infty$ 范数, 而且不难看出, 式(4.1.2)、式(4.1.3)和式(4.1.4)对复向量空间 \mathbb{C}^n 中的向量仍然成立(注意, $|x_i|$ 应理解为复数的模).

例 4.1.5 计算 \mathbb{C}^4 的向量 $\mathbf{x}=(3i, 0, -4i, -12)^T$ 的 $1, 2, \infty$ 范数, 这里 $i=\sqrt{-1}$.

解

$$\begin{aligned}\|\mathbf{x}\|_1 &= \sum_{i=1}^4 |x_i| = |3i| + |-4i| + |-12| = 19, \\ \|\mathbf{x}\|_2 &= \sqrt{\mathbf{x}^H \mathbf{x}} = \sqrt{(3i)(-3i) + (-4i)(4i) + (-12)^2} = 13, \\ \|\mathbf{x}\|_\infty &= \max(|x_1|, |x_2|, |x_3|, |x_4|) \\ &= \max(3, 0, 4, 12) = 12.\end{aligned}$$

由此可见, 在同一个线性空间中, 不同定义的范数其大小可能不同.

例 4.1.6 在 \mathbb{R}^n (或 \mathbb{C}^n)中, 若 $\mathbf{x}=(x_1, x_2, \dots, x_n)^T$, 如果仍按式(4.1.4)的规律, 但取 $0 < p < 1$ 来定义某个实值函数:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}, \quad 0 < p < 1,$$

试验证它不是 \mathbb{R}^n (或 \mathbb{C}^n)中的范数.

解 显然它不满足定义 4.1.1 中的条件(3), 所以它不是 \mathbb{R}^n (或 \mathbb{C}^n)上的范数. 例如, 在 \mathbb{R}^2 中, 取 $\mathbf{x}=(1, 0)^T, \mathbf{y}=(0, 1)^T$, 则 $\|\mathbf{x}+\mathbf{y}\|_{\frac{1}{2}}=4, \|\mathbf{x}\|_{\frac{1}{2}}=1, \|\mathbf{y}\|_{\frac{1}{2}}=1$, 所以

$$\|\mathbf{x}+\mathbf{y}\|_{\frac{1}{2}} \leq \|\mathbf{x}\|_{\frac{1}{2}} + \|\mathbf{y}\|_{\frac{1}{2}}$$

不成立, 那么 $\|\mathbf{x}\|_{\frac{1}{2}}$ 就不是 \mathbb{R}^2 中的向量范数.

下面的例子, 给出由已知的某种范数去构造出新的向量范数的一种方法.

例 4.1.7 设 $\|\cdot\|_a$ 是 \mathbb{C}^m 上的一种向量范数, 给定矩阵 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 且 \mathbf{A} 的 n 个列向量线性无关, 对任意 $\mathbf{x}=(x_1, x_2, \dots, x_n)^T \in \mathbb{C}^n$, 规定

$$\|\mathbf{x}\|_\beta = \|\mathbf{Ax}\|_a,$$

则 $\|\mathbf{x}\|_\beta$ 是 \mathbb{C}^n 中的向量范数.

证明 (1) 设 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是矩阵 \mathbf{A} 的 n 个线性无关的列向量, 从而对任何 $\mathbf{x}=(x_1, x_2, \dots, x_n)^T \neq \mathbf{0}$, 有

$$\begin{aligned}\mathbf{Ax} &= (\alpha_1, \alpha_2, \dots, \alpha_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\ &= x_1 \alpha_1 + x_2 \alpha_2 + \dots + x_n \alpha_n \neq \mathbf{0}.\end{aligned}$$

由于 $\|\cdot\|_a$ 是 \mathbb{C}^m 上的向量范数, 故 $\|\mathbf{Ax}\|_a > 0$, 即

$$\|\mathbf{x}\|_\beta = \|\mathbf{Ax}\|_a > 0;$$

(2) 对 $\forall k \in \mathbb{C}, \forall \mathbf{x} \in \mathbb{C}^n$, 有

$$\begin{aligned}\|k\mathbf{x}\|_\beta &= \|\mathbf{A}(k\mathbf{x})\|_a = \|k\mathbf{Ax}\|_a = |k| \|\mathbf{Ax}\|_a \\ &= |k| \|\mathbf{x}\|_\beta;\end{aligned}$$

(3) 对 $\forall \mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, 有

$$\begin{aligned}\|\mathbf{x}+\mathbf{y}\|_\beta &= \|\mathbf{A}(\mathbf{x}+\mathbf{y})\|_a = \|\mathbf{Ax}+\mathbf{Ay}\|_a \\ &\leq \|\mathbf{Ax}\|_a + \|\mathbf{Ay}\|_a = \|\mathbf{x}\|_\beta + \|\mathbf{y}\|_\beta,\end{aligned}$$

故 $\|\mathbf{x}\|_\beta$ 是 \mathbb{C}^n 中的向量范数.

上述构造新的向量范数的方法也可推广到一般的线性空间中, 即可定义出抽象向量的范数.

例 4.1.8 设 V 是数域 P 上的 n 维线性空间, e_1, e_2, \dots, e_n 是 V 的一组基, 则对 $\forall \alpha \in V$, 有惟一表示式

$$\alpha = x_1 e_1 + x_2 e_2 + \dots + x_n e_n.$$

记 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in P^n$, 又设 $\|\cdot\|$ 是 P^n 上的向量范数, 规定

$$\|\alpha\|_\mu = \|\mathbf{x}\|. \quad (4.1.8)$$

试证 $\|\alpha\|_\mu$ 是 V 中向量的一种范数.

证明 对 $\forall \alpha \in V$, 如果 $\alpha \neq \mathbf{0}$, 则其坐标向量 $\mathbf{x} \neq \mathbf{0}$, 从而 $\|\alpha\|_\mu = \|\mathbf{x}\| > 0$; 如果 $\alpha = \mathbf{0}$, 则其坐标向量 $\mathbf{x} = \mathbf{0}$, 于是 $\|\alpha\|_\mu = \|\mathbf{x}\| = 0$.

对 $\forall k \in P, k\alpha = \sum_{i=1}^n kx_i e_i$, 即 $k\alpha$ 的坐标向量为 $k\mathbf{x} = (kx_1, kx_2, \dots, kx_n)^T$, 故

$$\|k\alpha\|_\mu = \|k\mathbf{x}\| = |k| \|\mathbf{x}\| = |k| \|\alpha\|_\mu.$$

对 $\forall \beta \in V$, 则 $\beta = \sum_{i=1}^n y_i e_i$, $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$, $\alpha + \beta$ 的坐标向量为 $\mathbf{x} + \mathbf{y}$, 于是

$$\|\alpha + \beta\|_\mu = \|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| = \|\alpha\|_\mu + \|\beta\|_\mu,$$

因此 $\|\alpha\|_\mu$ 是 V 上的向量范数. 证毕

上例中最常用的是取 \mathbf{x} 的 2-范数来定义抽象向量 α 的范数:

$$\|\alpha\|_\mu = \|\alpha\|_E = \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}}. \quad (4.1.9)$$

由此可见, 在一般的线性空间 V^n 中, 可以有无穷多种范数.

定义 4.1.2 定义了向量范数 $\|\cdot\|$ 的线性空间 V^n , 称为赋范线性空间. 其中 $\|\cdot\|$ 表示泛指的任何一种范数.

4.1.2 向量范数的性质

前面已经看到, 在同一个线性空间内向量的范数可以有无穷多种(只要 p 在 1 与 ∞ 之间取值). 但我们自然要问: 这些范数之间有什么重要关系呢? 下面的定理告诉我们, 有限维线性空间上的不同范数是等价的.

定义 4.1.3 设 $\|\mathbf{x}\|_a, \|\mathbf{x}\|_b$ 是 n 维线性空间 V^n 上定义的任意两种范数, 若存在两个与 \mathbf{x} 无关的正常数 c_1, c_2 , 使得

$$c_1 \|\mathbf{x}\|_b \leq \|\mathbf{x}\|_a \leq c_2 \|\mathbf{x}\|_b, \quad \forall \mathbf{x} \in V^n, \quad (4.1.10)$$

则称 $\|\mathbf{x}\|_a$ 与 $\|\mathbf{x}\|_b$ 是等价的.

下面的命题可以帮助我们更好地理解范数等价, 其证明可以直接从式(4.1.10)得到.

命题 设 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \dots$ 是线性空间 V 中的向量序列, \mathbf{x}^* 是 V 中某给定向量. 设 $\|\cdot\|_a$ 与 $\|\cdot\|_b$ 是 V 的两个向量范数, 则 $\|\cdot\|_a$ 与 $\|\cdot\|_b$ 等价的充分必要条件是

$$\lim_{n \rightarrow \infty} \|\mathbf{x}_n - \mathbf{x}^*\|_a = 0 \Leftrightarrow \lim_{n \rightarrow \infty} \|\mathbf{x}_n - \mathbf{x}^*\|_b = 0, \quad (4.1.11)$$

此时称序列 $\{\mathbf{x}_n\}$ 按范数收敛于 \mathbf{x}^* , 换句话说, 两个范数等价 \Leftrightarrow 它们具有相同的敛散性(即有相同的极限).

定理 4.1.2 有限维线性空间上的不同范数是等价的.

* **证明** 设 $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_n$ 是 V^n 的一个基, 于是 V^n 中任意向量 \mathbf{x} 可以表示为

$$\mathbf{x} = x_1 \boldsymbol{\varepsilon}_1 + x_2 \boldsymbol{\varepsilon}_2 + \dots + x_n \boldsymbol{\varepsilon}_n.$$

定义

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2},$$

显然它是一种范数. 若存在正常数 c'_1, c'_2 和 c''_1, c''_2 , 使 $c'_1 \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_a \leq c'_2 \|\mathbf{x}\|_2$ 和 $c''_1 \|\mathbf{x}\|_b \leq \|\mathbf{x}\|_2 \leq c''_2 \|\mathbf{x}\|_b$ 成立, 则显然有 $c'_1 c''_1 \|\mathbf{x}\|_b \leq \|\mathbf{x}\|_a \leq c'_2 c''_2 \|\mathbf{x}\|_b$, 令 $c_1 = c'_1 c''_1, c_2 = c'_2 c''_2$, 便得不等式(4.1.10). 因此只要对 $b=2$ 证明式(4.1.10)成立就行了.

考察

$$\|\mathbf{x}\|_a = \|x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + \cdots + x_n \mathbf{e}_n\|_a,$$

它可以看作是 n 个坐标分量 (x_1, x_2, \cdots, x_n) 的函数, 记

$$\|\mathbf{x}\|_a = \varphi(x_1, x_2, \cdots, x_n),$$

可以证明 $\varphi(x_1, x_2, \cdots, x_n)$ 是坐标分量 x_1, x_2, \cdots, x_n 的连续函数.

事实上, 设另一个向量为

$$\mathbf{x}' = x'_1 \mathbf{e}_1 + x'_2 \mathbf{e}_2 + \cdots + x'_n \mathbf{e}_n,$$

$$\|\mathbf{x}'\|_a = \varphi(x'_1, x'_2, \cdots, x'_n),$$

则

$$\begin{aligned} & |\varphi(x'_1, x'_2, \cdots, x'_n) - \varphi(x_1, x_2, \cdots, x_n)| \\ &= |\|\mathbf{x}'\|_a - \|\mathbf{x}\|_a| \\ &\leq \|\mathbf{x}' - \mathbf{x}\|_a \\ &= \|(x'_1 - x_1)\mathbf{e}_1 + (x'_2 - x_2)\mathbf{e}_2 + \cdots + (x'_n - x_n)\mathbf{e}_n\|_a \\ &\leq |x'_1 - x_1| \|\mathbf{e}_1\|_a + |x'_2 - x_2| \|\mathbf{e}_2\|_a + \cdots + \\ &\quad |x'_n - x_n| \|\mathbf{e}_n\|_a. \end{aligned}$$

由于 $\|\mathbf{e}_i\|_a (i=1, 2, \cdots, n)$ 是常数, 因此当 x'_i 与 x_i 充分接近时, $\varphi(x'_1, x'_2, \cdots, x'_n)$ 就充分接近 $\varphi(x_1, x_2, \cdots, x_n)$, 这就证明了 $\varphi(x_1, x_2, \cdots, x_n)$ 是连续函数.

根据连续函数的性质可知, 在有界闭集

$$x_1^2 + x_2^2 + \cdots + x_n^2 = 1 \quad (4.1.12)$$

上(即 n 维欧氏空间 \mathbb{R}^n 的单位球)函数 $\varphi(x_1, x_2, \cdots, x_n)$ 可达到最大值 M 和最小值 m , 因为在式(4.1.12)中的 x_i 不能全为零, 因此 $m > 0$, 记

$$d = \|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2},$$

则向量

$$\mathbf{y} = \frac{x_1}{d} \mathbf{e}_1 + \frac{x_2}{d} \mathbf{e}_2 + \cdots + \frac{x_n}{d} \mathbf{e}_n$$

的分量满足

$$\left(\frac{x_1}{d}\right)^2 + \left(\frac{x_2}{d}\right)^2 + \cdots + \left(\frac{x_n}{d}\right)^2 = 1.$$

因此, 向量 \mathbf{y} 在单位球上, 从而有

$$0 < m \leq \|\mathbf{y}\|_a = \varphi\left(\frac{x_1}{d}, \frac{x_2}{d}, \cdots, \frac{x_n}{d}\right) \leq M.$$

但 $\mathbf{y} = \frac{1}{d} \cdot \mathbf{x}$, 故

$$md \leq \|\mathbf{x}\|_a \leq Md,$$

即

$$m \| \boldsymbol{x} \|_2 \leq \| \boldsymbol{x} \|_a \leq M \| \boldsymbol{x} \|_2. \quad \text{证毕}$$

注意, 定理 4.1.2 不能推广到无穷维空间, 在无限维的线性空间中, 两个向量范数是可以不等价的, 见下例.

例 4.1.9 设 $V = C[0, 1]$ 是闭区间 $[0, 1]$ 上全体实连续函数组成的无限维线性空间, 则

$$\| f \|_\infty = \max_{0 \leq x \leq 1} | f(x) | \quad (4.1.13)$$

与

$$\| f \|_1 = \int_0^1 | f(x) | dx \quad (4.1.14)$$

均是 V 中的范数. 它们等价吗? 考虑 V 中如下的函数列 f_n , 其中 $f_1 = 1$, 而对于每个 $n \geq 2$, 有

$$f_n = \begin{cases} 2nx, & 0 \leq x \leq \frac{1}{2n}, \\ -2nx + 2, & \frac{1}{2n} < x \leq \frac{1}{n}, \\ 0, & x > \frac{1}{n}, \end{cases}$$

画出这些函数的草图易知 $\lim_{n \rightarrow \infty} \| f_n \|_1 = 0$, 但 $\lim_{n \rightarrow \infty} \| f_n \|_\infty = 1$, 因此由前述命题知这两个范数不等价.

由于等价的范数导致相同的收敛性, 而函数的微积分学均由极限定义, 因此等价的范数将导致相同的微积分学.

如上所述, 正因为有了各种范数的等价性, 才保证了在各种范数下考虑向量序列收敛的一致性. 因此, 我们常常根据不同的要求选择一种方便的范数来研究收敛性问题.

习 题 4 (1)

- 求向量 $\boldsymbol{\alpha} = (1, 1, \dots, 1)$ 的 l_1, l_2 及 l_∞ 范数.
- 设 $\boldsymbol{\alpha} = (1, -2, 3)^\top, \boldsymbol{\beta} = (0, 2, 3)^\top$, 计算 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 的 3 种常用范数.
- 设 $\| \boldsymbol{\alpha} \|_a$ 与 $\| \boldsymbol{\alpha} \|_b$ 是 C^n 上的两种范数, k_1, k_2 是正的常数. 证明: 下列函数
 - $\max(\| \boldsymbol{\alpha} \|_a, \| \boldsymbol{\alpha} \|_b)$;
 - $k_1 \| \boldsymbol{\alpha} \|_a + k_2 \| \boldsymbol{\alpha} \|_b$
 是 C^n 上的范数.
- 设 $\| \cdot \|$ 是西空间 C^n 的向量范数, 证明向量范数的下列基本性质:
 - 零向量的范数为零;
 - 当 \boldsymbol{x} 是非零向量时, $\left\| \frac{\boldsymbol{x}}{\| \boldsymbol{x} \|} \right\| = 1$;
 - $\| -\boldsymbol{x} \| = \| \boldsymbol{x} \|$;
 - $|\| \boldsymbol{x} \| - \| \boldsymbol{y} \| | \leq \| \boldsymbol{x} - \boldsymbol{y} \|$;
 - $\| \boldsymbol{x} + \boldsymbol{y} \| \geq \| \boldsymbol{x} \| - \| \boldsymbol{y} \|$.
- 任取 $\boldsymbol{X} = (4i, -3i, 12, 0)^\top \in C^4$, 其中 $i^2 = -1$, 试计算 $\| \boldsymbol{X} \|_1, \| \boldsymbol{X} \|_2, \| \boldsymbol{X} \|_\infty$.

6. $\boldsymbol{A} = \begin{bmatrix} -1 & 0 & 2 & i \\ 3+i & 5 & 1+i & 0 \\ 2 & i & 2 & -4 \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} -1 \\ 2 \\ 0 \\ -i \end{bmatrix}, i = \sqrt{-1}$

 计算 $\| \boldsymbol{Ax} \|_1, \| \boldsymbol{Ax} \|_2, \| \boldsymbol{Ax} \|_\infty$.

7. 设 a_1, a_2, \dots, a_n 都是正实数, 向量 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$, 证明由 $\|\mathbf{x}\| = \left(\sum_{i=1}^n a_i x_i^2\right)^{\frac{1}{2}}$ 定义的非负实数是 \mathbb{R}^n 空间的一个向量范数.

8. 证明: 若 $\boldsymbol{\alpha}$ 是二维向量, 即 $\boldsymbol{\alpha} = (x_1, x_2)^T$, 则 $\|\boldsymbol{\alpha}\| = \max(|x_1|, |x_2|, \frac{2}{3}(|x_1| + |x_2|))$ 是向量范数. 再画出 $\|\boldsymbol{\alpha}\| \leq 1$ 的图形.

9. 求证: 对 $p = \frac{1}{2}$, 不能得到在 $\|\boldsymbol{\alpha}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}$ 意义下的范数.

10. 画出曲线 $x_1^{\frac{2}{3}} + x_2^{\frac{2}{3}} = 1$ 的草图. 若记二维实向量为 $\boldsymbol{\alpha} = (x_1, x_2)^T$, 实函数 $\|\boldsymbol{\alpha}\| = (|x_1|^{\frac{2}{3}} + |x_2|^{\frac{2}{3}})^{\frac{3}{2}}$ 是否为向量范数?

11. 区间 $[a, b]$ 上全体实值连续函数的集合, 按照通常的函数加法和数乘运算, 构成 \mathbb{R} 上的线性空间, 记作 $C[a, b]$. 对于 $f(t) \in C[a, b]$, 分别定义实数:

$$(1) \|f(t)\|_1 = \int_a^b |f(t)| dt;$$

$$(2) \|f(t)\|_\infty = \max_{t \in [a, b]} |f(t)|.$$

验证 $\|f(t)\|_1$ 与 $\|f(t)\|_\infty$ 都是 $C[a, b]$ 中的向量范数.

12. 设 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 是对称正定矩阵, 对于 \mathbb{R}^n 中的列向量 $\boldsymbol{\alpha}$, 定义实数 $\|\boldsymbol{\alpha}\| = \sqrt{\boldsymbol{\alpha}^T \mathbf{A} \boldsymbol{\alpha}}$, 验证 $\|\boldsymbol{\alpha}\|_{\mathbf{A}}$ 是 \mathbb{R}^n 中的向量范数.

13. 在 \mathbb{R}^2 中, 将向量 $\boldsymbol{\alpha} = (\xi_1, \xi_2)$ 表示成平面上直角坐标系中的点 (ξ_1, ξ_2) , 分别画出下列不等式决定的 $\boldsymbol{\alpha}$ 全体所对应的几何图形:

$$\|\boldsymbol{\alpha}\|_1 \leq 1, \quad \|\boldsymbol{\alpha}\|_2 \leq 1, \quad \|\boldsymbol{\alpha}\|_\infty \leq 1.$$

14. 设 $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n \times n}$, 且 \mathbf{B} 可逆, 对于 \mathbb{C}^n 中的列向量 $\boldsymbol{\alpha}$, 定义实数 $\|\boldsymbol{\alpha}\| = \|\mathbf{A}\boldsymbol{\alpha}\|_1 + 3\|\mathbf{B}\boldsymbol{\alpha}\|_2$, 验证 $\|\boldsymbol{\alpha}\|$ 是 \mathbb{C}^n 中的向量范数.

15. 设 $\mathbf{X} = (x_1, x_2, \dots, x_n)^T \in \mathbb{C}^n$, 定义

$$\|\mathbf{X}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}, \quad 0 < p < 1,$$

问 $\|\mathbf{X}\|_p$ 是否为 \mathbb{C}^n 上的向量范数? 如果是, 请予以证明; 如果不是, 请举反例说明.

16. 设 $\mathbf{A} \in \mathbb{C}^{m \times n}$, $\|\cdot\|_a$ 是 \mathbb{C}^n 上的一种向量范数, 对于任意的 $\mathbf{X} \in \mathbb{C}^n$, 定义

$$\|\mathbf{X}\|_b = \|\mathbf{A}\mathbf{X}\|_a.$$

证明: $\|\mathbf{X}\|_b$ 是 \mathbb{C}^n 上的向量范数.

17. 设 \mathbf{A} 为 n 阶正定埃尔米特矩阵, 对任意 $\mathbf{X} \in \mathbb{C}^n$, 定义

$$\|\mathbf{X}\|_{\mathbf{A}} = \sqrt{\mathbf{X}^H \mathbf{A} \mathbf{X}},$$

试证: $\|\mathbf{X}\|_{\mathbf{A}}$ 是一种向量范数.

18. 设 V 是复数域 \mathbb{C} 上的 n 维线性空间, $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_n$ 是 V 的一组基, 任取 $\boldsymbol{\alpha} \in V$, $\mathbf{X} = (x_1, x_2, \dots, x_n)^T \in \mathbb{C}^n$ 为 $\boldsymbol{\alpha}$ 在基 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_n$ 下的坐标, 已知 $\|\cdot\|$ 是 \mathbb{C}^n 上的一种向量范数, 定义

$$\|\boldsymbol{\alpha}\|_v = \|\mathbf{X}\|,$$

试证: $\|\boldsymbol{\alpha}\|_v$ 是 V 上的向量范数.

19. 设区间 $[a, b]$ 上全体实值连续函数的集合, 在通常函数的加法与数乘运算构成一个 \mathbb{R} 上的线性空间, 记为 $C[a, b]$, $\forall f(x) \in C[a, b]$, 定义实数:

$$\|f(t)\| = \int_a^b |f(t)| dt,$$

证明 $\|f(t)\|$ 是 $C[a, b]$ 上的向量范数.

20. 设 $\boldsymbol{\alpha} = (x_1, x_2, \dots, x_n)^T$, 则范数 $\|\boldsymbol{\alpha}\|$ 是变元 x_1, x_2, \dots, x_n 的一致连续函数, 即对任意给定的正数 ε , 必存在正数 δ , 对于 $\boldsymbol{\beta} = (y_1, y_2, \dots, y_n)^T$, 当 $\max_i |y_i| < \delta$ 时, 有 $\|\boldsymbol{\alpha} + \boldsymbol{\beta}\| - \|\boldsymbol{\alpha}\| < \varepsilon$, 试证明之.

21. 证明: 若 $\alpha \in \mathbb{C}^n$, 则向量范数 $\|\alpha\|_1$, $\|\alpha\|_2$ 及 $\|\alpha\|_\infty$ 两两等价, 且有

$$(1) \|\alpha\|_\infty \leq \|\alpha\|_1 \leq n \|\alpha\|_\infty;$$

$$(2) \|\alpha\|_\infty \leq \|\alpha\|_2 \leq \sqrt{n} \|\alpha\|_\infty;$$

$$(3) \frac{1}{n} \|\alpha\|_1 \leq \|\alpha\|_2 \leq \sqrt{n} \|\alpha\|_1;$$

$$(4) \|\alpha\|_2 \leq \|\alpha\|_1 \leq \sqrt{n} \|\alpha\|_2.$$

22. 已知对称正定矩阵 $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, 直接验证: \mathbb{R}^2 中的向量范数 $\|\alpha\|_A$ 与 $\|\alpha\|_2$ 等价, 且有

$$\|\alpha\|_2 \leq \|\alpha\|_A \leq \sqrt{3} \|\alpha\|_2,$$

其中

$$\|\alpha\|_A = \sqrt{\alpha^T A \alpha}.$$

23. 证明: 在 \mathbb{R}^n 中当且仅当 α, β 线性相关而且 $\alpha^T \beta \geq 0$ 时, 才有 $(\alpha, \beta) = \|\alpha\|_2 \cdot \|\beta\|_2$.

24. 设数域 \mathbb{R} 上的多项式空间 $P[x]_{n-1}$ 的两组基为

$$(I) f_1(x) = 1, f_2(x) = x, \dots, f_n(x) = x^{n-1};$$

$$(II) (g_1, g_2, \dots, g_n) = (f_1, f_2, \dots, f_n)C.$$

其中 C 为可逆矩阵, $f(x) \in P[x]_{n-1}$ 在基 (I) 与基 (II) 下的坐标分别为

$$\alpha = (\xi_1, \xi_2, \dots, \xi_n)^T, \quad \beta = (\eta_1, \eta_2, \dots, \eta_n)^T.$$

(1) 证明: 对任意的 $f(x) \in P[x]_{n-1}$, $\|\alpha\|_2 = \|\beta\|_2$ 的充分必要条件是 $C^T C = I$;

(2) 取 $n=3$, 找出使 (1) 成立的基 (II), 而且 $g_i(x) \neq \pm f_j(x)$ ($i, j=1, 2, 3$).

25. 设 $f(x) \in C[a, b]$, 定义

$$\|f(x)\|_p = \left(\int_a^b |f(x)|^p dx \right)^{1/p}.$$

试证: $\|f(x)\|_p$ 是 $C[a, b]$ 上的向量范数.

26. 设 $\|\alpha\|$ 是 \mathbb{C}^n 上的向量范数, $A \in \mathbb{C}^{n \times n}$, 则 $\|A\alpha\|$ 也是 \mathbb{C}^n 上的范数的充分必要条件是 A 为可逆矩阵.

27. 设 $\alpha \in \mathbb{C}^n$, 试证:

$$\frac{1}{n} \|\alpha\|_1 \leq \|\alpha\|_\infty \leq \|\alpha\|_2 \leq \|\alpha\|_1.$$

28. 设 a_1, a_2, \dots, a_n 是正实数, $\alpha = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$, 试证: $\|\alpha\| = \left(\sum_{i=1}^n a_i x_i^2 \right)^{1/2}$ 是向量范数.

29. 设 A 是正定埃尔米特矩阵, 试证: 若 $\alpha \in \mathbb{C}^n$, 则 $\|\alpha\| = (\alpha^H A \alpha)^{1/2}$ 是 α 的向量范数, 称此范数为椭圆范数.

4.2 矩阵的范数

本节将进一步把范数的概念推广到 $m \times n$ 矩阵上. 一个 $m \times n$ 矩阵当然可以看作是 $m \times n$ 维的向量, 因此可以按向量定义范数的办法来定义矩阵的范数. 但是, 由于在线性空间中只需考虑加法运算与数乘运算, 而现在矩阵空间 $\mathbb{C}^{m \times n}$ 中的矩阵, 还必须考虑矩阵与向量以及矩阵与矩阵之间的乘法运算, 因此在定义矩阵范数时, 必须多一条反映矩阵乘法的公理.

4.2.1 矩阵范数的定义与性质

定义 4.2.1 设 $A \in \mathbb{C}^{m \times n}$, 按某一法则在 $\mathbb{C}^{m \times n}$ 上规定 A 的一个实值函数, 记作 $\|A\|$, 它满足下面 4 个条件:

- (1) 非负性: 如果 $\mathbf{A} \neq \mathbf{0}$, 则 $\|\mathbf{A}\| > 0$; 如果 $\mathbf{A} = \mathbf{0}$, 则 $\|\mathbf{A}\| = 0$.
- (2) 齐次性: 对任意的 $k \in \mathbb{C}$, $\|k\mathbf{A}\| = |k| \|\mathbf{A}\|$.
- (3) 三角不等式: 对任意 $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{m \times n}$, $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$.
- (4) 次乘性: 当矩阵乘积 \mathbf{AB} 有意义时, 若有

$$\|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|,$$

则称 $\|\mathbf{A}\|$ 为矩阵范数.

注 1 (次乘性的合理性) 如果将矩阵范数定义中的次乘性的不等式反向, 即 $\|\mathbf{AB}\| \geq \|\mathbf{A}\| \|\mathbf{B}\|$, 则幂零矩阵 ($\mathbf{A}^2 = \mathbf{0}$) 的矩阵范数将是 0, 与非负性不符.

注 2 (次乘性的意义) 设 $\|\mathbf{A}\| < 1$, 则次乘性保证了 $\|\mathbf{A}^k\| \rightarrow 0 (k \rightarrow \infty)$. 因此矩阵范数的次乘性实际上保证了矩阵幂级数的敛散性的“合理性”.

如前所述, 我们若把 $m \times n$ 矩阵 \mathbf{A} 看成是一个 $m \times n$ 维的向量, 那么很自然地就可以仿照前面的式 (4.1.3), 式 (4.1.2) 和式 (4.1.1)' 来得出矩阵的几种范数. 为简单起见, 下面给出方阵的几个例子.

例 4.2.1 设 $\mathbf{A} = (a_{ij}) \in \mathbb{C}^{n \times n}$, 试验证下面规定的实值函数

$$\|\mathbf{A}\|_{m_1} = \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|, \quad (4.2.1)$$

$$\|\mathbf{A}\|_{m_\infty} = n \cdot \max_{i,j} |a_{ij}|, \quad (4.2.2)$$

$$\|\mathbf{A}\|_{m_2} = \left(\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \right)^{\frac{1}{2}}, \quad (4.2.3)$$

都是矩阵 \mathbf{A} 的范数.

事实上, 由于矩阵范数定义中的前 3 条与向量范数定义的 3 条类似, 故它们满足矩阵范数定义的前 3 条是显然的, 现只证它们满足矩阵范数定义的第 (4) 条次乘性即可.

$$\begin{aligned} \|\mathbf{AB}\|_{m_1} &= \sum_{i=1}^n \sum_{j=1}^n \left| \sum_{k=1}^n a_{ik} b_{kj} \right| \leq \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n |a_{ik}| |b_{kj}| \\ &\leq \left(\sum_{i=1}^n \sum_{k=1}^n |a_{ik}| \right) \left(\sum_{k=1}^n \sum_{j=1}^n |b_{kj}| \right) = \|\mathbf{A}\|_{m_1} \|\mathbf{B}\|_{m_1}, \end{aligned}$$

$$\begin{aligned} \|\mathbf{AB}\|_{m_\infty} &= n \cdot \max_{i,j} \left| \sum_{k=1}^n a_{ik} b_{kj} \right| \\ &\leq n \cdot \max_{i,j} \sum_{k=1}^n |a_{ik}| |b_{kj}| \\ &\leq n \cdot \max_{i,j} (n \cdot \max_k |a_{ik}| |b_{kj}|) \\ &\leq (n \cdot \max_{i,k} |a_{ik}|) (n \cdot \max_{k,j} |b_{kj}|) \\ &= \|\mathbf{A}\|_{m_\infty} \|\mathbf{B}\|_{m_\infty}, \end{aligned}$$

$$\begin{aligned} \|\mathbf{AB}\|_{m_2} &= \left(\sum_{i=1}^n \sum_{j=1}^n \left| \sum_{k=1}^n a_{ik} b_{kj} \right|^2 \right)^{\frac{1}{2}} \\ &\leq \left(\sum_{i=1}^n \sum_{j=1}^n \left(\sum_{k=1}^n |a_{ik}| |b_{kj}| \right)^2 \right)^{\frac{1}{2}} \end{aligned}$$

$$\begin{aligned} &\leq \left(\sum_{i=1}^n \sum_{j=1}^n \left(\sum_{k=1}^n |a_{ik}|^2 \right) \left(\sum_{k=1}^n |b_{kj}|^2 \right) \right)^{\frac{1}{2}} \\ &= \left(\sum_{i=1}^n \sum_{k=1}^n |a_{ik}|^2 \right)^{\frac{1}{2}} \left(\sum_{k=1}^n \sum_{j=1}^n |b_{kj}|^2 \right)^{\frac{1}{2}} \\ &= \| \mathbf{A} \|_{m_2} \| \mathbf{B} \|_{m_2}. \end{aligned}$$

式(4.2.3)实际上可视为 $n \times n$ 维向量 \mathbf{A} 的 2-范数,我们称这种矩阵范数为 \mathbf{A} 的弗罗贝尼乌斯(Frobenius)范数,或简称为 \mathbf{F} -范数,记为 $\| \mathbf{A} \|_{\mathbf{F}}$.

与向量范数类似,矩阵范数也具有相应性质,我们将不加证明地给出如下定理.

定理 4.2.1 设 $\mathbf{A} \in \mathbb{C}^{m \times n}$, $\| \mathbf{A} \|$ 是 $\mathbb{C}^{m \times n}$ 上的矩阵范数,则 $\mathbb{C}^{m \times n}$ 上的任意两个矩阵范数等价.

这个定理的证明与上节定理 4.1.2 的证明类似.

如前所述,矩阵范数也是多种多样的,而矩阵范数与向量范数又是有差异的,可是,在矩阵范数的定义 4.2.1 中,虽然考虑到了矩阵的乘法性质,但还没有将矩阵和线性算子(或线性变换)联系起来.矩阵的“真正范数”应能体现矩阵的这两层含义,或者说矩阵自身的范数应考虑到矩阵的乘法以及线性算子(变换)作用下的复合效果,这样定义出来的矩阵范数,才是最有用的或说是最合理的.这就是下面将要介绍的算子范数.

4.2.2 算子范数

在第 1 章中,我们知道线性算子 \mathcal{A} 作用在向量 $\boldsymbol{\alpha}$ 上,在某基偶下有如下的对应关系:

$$\mathcal{A}(\boldsymbol{\alpha}) \leftrightarrow \mathbf{A}\mathbf{x}, \quad \mathbf{A} \in \mathbb{C}^{m \times n}, \quad \mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{C}^n,$$

若将坐标向量 \mathbf{x} 视作矩阵,根据矩阵范数定义 4.2.1 的次乘性(4),应有

$$\| \mathbf{A}\mathbf{x} \| \leq \| \mathbf{A} \| \| \mathbf{x} \|, \quad (4.2.4)$$

因此矩阵 \mathbf{A} 的范数应满足不等式

$$\| \mathbf{A} \| \geq \frac{\| \mathbf{A}\mathbf{x} \|}{\| \mathbf{x} \|}. \quad (4.2.5)$$

遗憾的是,对于任意的矩阵 \mathbf{A} ,式(4.2.5)的右端可能不是一个常数,而和向量 \mathbf{x} 有关,但可以肯定的是,矩阵 \mathbf{A} 的范数绝对不能取 $\frac{\| \mathbf{A}\mathbf{x} \|}{\| \mathbf{x} \|}$,因为如果 \mathbf{A} 是非零不可逆矩阵,则存在非零向量 \mathbf{x} 使得 $\mathbf{A}\mathbf{x} = \mathbf{0}$,于是有 $\| \mathbf{A} \| = 0$,这与矩阵的非负性矛盾.因此应该取式(4.2.5)右端的最大值或者上确界(最小上界),即定义

$$\| \mathbf{A} \| = \sup_{\| \mathbf{x} \| \neq 0} \frac{\| \mathbf{A}\mathbf{x} \|}{\| \mathbf{x} \|}, \quad (4.2.6)$$

当 $\| \mathbf{x} \| = 1$ 时,有

$$\| \mathbf{A} \| = \sup_{\| \mathbf{x} \| = 1} \| \mathbf{A}\mathbf{x} \| = \max_{\| \mathbf{x} \| = 1} \| \mathbf{A}\mathbf{x} \|. \quad (4.2.7)$$

因向量赋范线性空间的单位闭球或单位球面($\| \mathbf{x} \| = 1$)皆为有界闭集,而 $\| \mathbf{A}\mathbf{x} \|$ 为 \mathbf{x} 的连续函数,故在单位球或单位球面上取得最大值,所以公式(4.2.7)中的“sup”可以换为“max”.

在不等式(4.2.4)中,同时出现了矩阵范数和向量范数,尽管它们各自的范数可能有不同的取法,但都应当保持这个不等式成立,这就是矩阵范数与向量范数相容(协调)的概念.

定义 4.2.2 设 $A \in \mathbb{C}^{m \times n}$, $x \in \mathbb{C}^n$, 如果取定的向量范数 $\|x\|$ 和矩阵范数 $\|A\|$ 满足不等式(4.2.4)

$$\|Ax\| \leq \|A\| \cdot \|x\|,$$

则称矩阵范数 $\|A\|$ 与向量范数 $\|x\|$ 是相容的.

式(4.2.6)和式(4.2.7)实际上是一个从向量范数出发构造与之相容的矩阵范数的方法, 将其写成下面的定理.

定理 4.2.2 设 $A \in \mathbb{C}^{m \times n}$, $x = (x_1, x_2, \dots, x_n)^T \in \mathbb{C}^n$, 且在 \mathbb{C}^n 中已规定了向量的范数(即 \mathbb{C}^n 是 n 维赋范线性空间). 定义

$$\|A\| = \sup_{\|x\| \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\|=1} \|Ax\|, \quad (4.2.8)$$

则上式定义了一个与向量范数 $\|\cdot\|$ 相容的矩阵范数, 称为由向量范数 $\|\cdot\|$ 诱导的矩阵范数或算子范数.

证明 为简单起见, 下面只要证明由 $\|A\| = \max_{\|x\|=1} \|Ax\|$ 定义的矩阵范数同时满足定义 4.2.1 中的 4 个条件和相容性条件(4.2.4)即可.

下面先验证相容性条件, 设 $y \neq 0$ 为任意一个向量, 则 $x = \frac{1}{\|y\|}y$ 满足条件 $\|x\| = 1$, 于是有

$$\begin{aligned} \|Ay\| &= \|A(\|y\|x)\| = \|y\| \|Ax\| \leq \|y\| \|A\| \\ &= \|A\| \|y\|. \end{aligned}$$

再验证 $\|A\|$ 确实是矩阵范数.

(1) 若 $A \neq 0$, 则一定可以找到 $\|x\| = 1$ 的向量 x , 使得 $Ax \neq 0$, 从而 $\|Ax\| \neq 0$, 所以 $\|A\| = \max_{\|x\|=1} \|Ax\| > 0$; 若 $A = 0$, 则一定有 $\|A\| = \max_{\|x\|=1} \|0x\| = 0$.

(2) 由式(4.2.8), 对 $\forall k \in \mathbb{C}$ 有

$$\|kA\| = \max_{\|x\|=1} \|kAx\| = |k| \max_{\|x\|=1} \|Ax\| = |k| \|A\|.$$

(3) 对于矩阵 $A+B$, 可以找到向量 x_0 , 使得

$$\|A+B\| = \|(A+B)x_0\|, \quad \|x_0\| = 1.$$

于是

$$\begin{aligned} \|A+B\| &= \|(A+B)x_0\| = \|Ax_0 + Bx_0\| \\ &\leq \|Ax_0\| + \|Bx_0\| \\ &\leq \|A\| \|x_0\| + \|B\| \|x_0\| = \|A\| + \|B\|. \end{aligned}$$

(4) 对于矩阵 AB , 可找到向量 x_0 , 使得

$$\|x_0\| = 1, \quad \|ABx_0\| = \|AB\|,$$

于是

$$\begin{aligned} \|AB\| &= \|ABx_0\| = \|A(Bx_0)\| \leq \|A\| \|Bx_0\| \\ &\leq \|A\| \|B\| \|x_0\| = \|A\| \|B\|, \end{aligned}$$

这就证明了式(4.2.8)定义的实值函数确实是一种矩阵范数.

证毕

显然, n 阶单位矩阵 I 的从属于任何向量范数的算子范数 $\|I\| = \max_{\|x\|=1} \|Ix\| = 1$; 而

对于 \mathbf{I} 的非算子范数, 如 $\|\mathbf{I}\|_{m_1} = n$, $\|\mathbf{I}\|_F = \sqrt{n}$, 它们都大于 1. 由于 $\mathbf{x} = \mathbf{I}\mathbf{x}$, 所以 $\|\mathbf{x}\| \leq \|\mathbf{I}\| \|\mathbf{x}\|$, 当取 $\|\mathbf{x}\| = 1$ 时, 有 $\|\mathbf{I}\| \geq 1$. 这说明单位矩阵的算子范数是所有与 $\|\mathbf{x}\|$ 相容的范数 $\|\mathbf{I}\|$ 中值最小的一个.

直接从式(4.2.8)来求矩阵的算子范数的值显然是很不方便的. 为此, 下面介绍几种常用的算子范数的求法. 当在式(4.2.8)中取向量 \mathbf{x} 的范数依次为 $\|\mathbf{x}\|_1$, $\|\mathbf{x}\|_2$, $\|\mathbf{x}\|_\infty$ 时, 希望由它们诱导的 3 种算子范数 $\|\mathbf{A}\|_1$, $\|\mathbf{A}\|_2$, $\|\mathbf{A}\|_\infty$ 的值可以通过矩阵 \mathbf{A} 的元素及 $\mathbf{A}^H\mathbf{A}$ 的特征值具体地表示出来, 有下面的定理.

定理 4.2.3 设 $\mathbf{A} = (a_{ij}) \in \mathbb{C}^{m \times n}$, $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{C}^n$, 则从属于向量 \mathbf{x} 的 3 种范数 $\|\mathbf{x}\|_1$, $\|\mathbf{x}\|_2$, $\|\mathbf{x}\|_\infty$ 的算子范数依次是

$$(1) \|\mathbf{A}\|_1 = \max_j \sum_{i=1}^m |a_{ij}| \quad (\text{称为列范数}); \quad (4.2.9)$$

$$(2) \|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^H\mathbf{A})} \quad (\text{称为谱范数}), \quad (4.2.10)$$

其中 $\lambda_{\max}(\mathbf{A}^H\mathbf{A})$ 是矩阵 $\mathbf{A}^H\mathbf{A}$ 特征值绝对值的最大值;

$$(3) \|\mathbf{A}\|_\infty = \max_i \sum_{j=1}^n |a_{ij}| \quad (\text{称为行范数}). \quad (4.2.11)$$

证明 (1) 对于任何非零向量 \mathbf{x} , 设 $\|\mathbf{x}\|_1 = 1$, 则有

$$\begin{aligned} \|\mathbf{Ax}\|_1 &= \sum_{i=1}^m \left| \sum_{j=1}^n a_{ij}x_j \right| \\ &\leq \sum_{i=1}^m \sum_{j=1}^n |a_{ij}| |x_j| \\ &= \sum_{j=1}^n \sum_{i=1}^m |a_{ij}| |x_j| \\ &= \sum_{j=1}^n \left(\sum_{i=1}^m |a_{ij}| \right) |x_j| \\ &\leq \max_j \sum_{i=1}^m |a_{ij}| \sum_{j=1}^n |x_j| \\ &= \max_j \sum_{i=1}^m |a_{ij}|, \end{aligned}$$

所以

$$\|\mathbf{Ax}\|_1 \leq \max_j \sum_{i=1}^m |a_{ij}|.$$

设在 $j = j_0$ 时, $\sum_{i=1}^m |a_{ij}|$ 达到最大值, 即

$$\sum_{i=1}^m |a_{ij_0}| = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|.$$

取向量

$$\mathbf{x}_0 = (0, \dots, 0, 1, 0, \dots, 0)^T,$$

其中第 j_0 个分量为 1. 显然 $\|\mathbf{x}_0\|_1 = 1$, 而且

$$\|\mathbf{Ax}_0\|_1 = \sum_{i=1}^m \left| \sum_{j=1}^n a_{ij}x_j \right| = \sum_{i=1}^m |a_{ij_0}| = \max_j \sum_{i=1}^m |a_{ij}|,$$

于是

$$\| \mathbf{A} \|_1 = \max_{\| \mathbf{x} \|_1 = 1} \| \mathbf{A} \mathbf{x} \|_1 = \max_j \sum_{i=1}^m | a_{ij} |.$$

(2) 因为 $\| \mathbf{A} \|_2 = \max_{\| \mathbf{x} \|_2 = 1} \| \mathbf{A} \mathbf{x} \|_2$, 但是

$$\| \mathbf{A} \mathbf{x} \|_2^2 = (\mathbf{A} \mathbf{x}, \mathbf{A} \mathbf{x}) = (\mathbf{x}, \mathbf{A}^H \mathbf{A} \mathbf{x}).$$

显然, 矩阵 $\mathbf{A}^H \mathbf{A}$ 是埃尔米特矩阵且非负, 从而它的特征值也都是非负实数.

设 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ 为 $\mathbf{A}^H \mathbf{A}$ 的特征值, 而 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ 是对应于这些特征值的一组标准正交特征向量, 任何一个范数为 1 的向量 \mathbf{x} 可表示为

$$\mathbf{x} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 + \dots + a_n \mathbf{x}_n,$$

则

$$(\mathbf{x}, \mathbf{x}) = | a_1 |^2 + | a_2 |^2 + \dots + | a_n |^2 = 1.$$

又

$$\begin{aligned} \| \mathbf{A} \mathbf{x} \|_2^2 &= (\mathbf{x}, \mathbf{A}^H \mathbf{A} \mathbf{x}) \\ &= (a_1 \mathbf{x}_1 + \dots + a_n \mathbf{x}_n, \lambda_1 a_1 \mathbf{x}_1 + \dots + \lambda_n a_n \mathbf{x}_n) \\ &= \lambda_1 | a_1 |^2 + \lambda_2 | a_2 |^2 + \dots + \lambda_n | a_n |^2 \\ &\leq \lambda_1 (| a_1 |^2 + | a_2 |^2 + \dots + | a_n |^2) \\ &= \lambda_1 = \lambda_{\max}(\mathbf{A}^H \mathbf{A}), \end{aligned}$$

而对于向量 $\mathbf{x} = \mathbf{x}_1$, 我们有

$$\begin{aligned} \| \mathbf{A} \mathbf{x}_1 \|_2^2 &= (\mathbf{x}_1, \mathbf{A}^H \mathbf{A} \mathbf{x}_1) = (\mathbf{x}_1, \lambda_1 \mathbf{x}_1) = \bar{\lambda}_1 (\mathbf{x}_1, \mathbf{x}_1) \\ &= \lambda_1 = \lambda_{\max}(\mathbf{A}^H \mathbf{A}), \end{aligned}$$

所以

$$\| \mathbf{A} \|_2 = \max_{\| \mathbf{x} \|_2 = 1} \| \mathbf{A} \mathbf{x} \|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^H \mathbf{A})}.$$

(3) 设 $\| \mathbf{x} \|_\infty = 1$, 则

$$\begin{aligned} \| \mathbf{A} \mathbf{x} \|_\infty &= \max_i \left| \sum_{j=1}^n a_{ij} x_j \right| \\ &\leq \max_i \sum_{j=1}^n | a_{ij} | | x_j | \\ &\leq \max_i \sum_{j=1}^n | a_{ij} |. \end{aligned}$$

所以

$$\max_{\| \mathbf{x} \|_\infty = 1} \| \mathbf{A} \mathbf{x} \|_\infty \leq \max_i \sum_{j=1}^n | a_{ij} |.$$

设 $\sum_{j=1}^n | a_{ij} |$ 在 $i = i_0$ 时达到最大值, 取下列向量

$$\mathbf{x}_0 = (x_1, x_2, \dots, x_n)^T,$$

其中

$$x_j = \begin{cases} \frac{| a_{i_0 j} |}{a_{i_0 j}}, & \text{当 } a_{i_0 j} \neq 0; \\ 1, & \text{当 } a_{i_0 j} = 0. \end{cases}$$

易知

$$\|\mathbf{x}_0\|_\infty = \max_j |x_j| = 1,$$

且当 $i=i_0$ 时,

$$\left| \sum_{j=1}^n a_{ij} x_j \right| = \max_i \sum_{j=1}^n |a_{ij}|,$$

从而

$$\|\mathbf{Ax}_0\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|,$$

于是

$$\|\mathbf{A}\|_\infty = \max_{\|\mathbf{x}\|_\infty=1} \|\mathbf{Ax}\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|. \quad \text{证毕}$$

比较习题 4(2)中 14 题和本定理中的式(4.2.9)和式(4.2.10)可知, $\|\mathbf{A}\|_{m_1}$ 和 $\|\mathbf{A}\|_1$ 均与 $\|\mathbf{x}\|_1$ 相容, $\|\mathbf{A}\|_F$ 和 $\|\mathbf{A}\|_2$ 均与 $\|\mathbf{x}\|_2$ 相容, 只不过是算子范数 $\|\mathbf{A}\|_1$, $\|\mathbf{A}\|_2$ 分别是与 $\|\mathbf{x}\|_1$, $\|\mathbf{x}\|_2$ 相容的范数中值最小的一个.

例 4.2.2 设 $\mathbf{A} = \mathbf{I}$, 则 $\|\mathbf{I}\|_{m_1} = \sum_{i=1}^n \sum_{j=1}^n |a_{ij}| = n \neq \|\mathbf{I}\|_1 = 1$; $\|\mathbf{I}\|_F = \left(\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \right)^{\frac{1}{2}} = \sqrt{n} \neq \|\mathbf{I}\|_2 = 1$; $\|\mathbf{I}\|_{m_\infty} = n \cdot \max_{i,j} |a_{ij}| = n \neq \|\mathbf{I}\|_\infty = 1$. 可见 $\|\mathbf{A}\|_{m_1}$, $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|_{m_\infty}$ 都不是算子范数.

例 4.2.3 设

$$\mathbf{A} = \begin{bmatrix} 1 & -2 \\ -3 & 4 \end{bmatrix},$$

试计算 $\|\mathbf{A}\|_1$, $\|\mathbf{A}\|_\infty$, $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|_2$.

解 $\|\mathbf{A}\|_1 = 6$, $\|\mathbf{A}\|_\infty = 7$, $\|\mathbf{A}\|_F = \sqrt{30} \approx 5.477$,

$$\|\mathbf{A}\|_2 = \sqrt{15 + \sqrt{221}} \approx 5.46.$$

由此例可见, \mathbf{A} 的各种范数的大小可能是不等的.

值得提出的是, 尽管 $\|\mathbf{A}\|_F$ 不是算子范数, 但它有优点, 所以它成为人们常用的矩阵范数之一, 请看下面的定理.

定理 4.2.4 设 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 而 $\mathbf{U} \in \mathbb{C}^{m \times m}$ 与 $\mathbf{V} \in \mathbb{C}^{n \times n}$ 都是酉矩阵, 证明

$$\|\mathbf{UA}\|_F = \|\mathbf{A}\|_F = \|\mathbf{AV}\|_F, \quad (4.2.12)$$

即给 \mathbf{A} 左乘或右乘以酉矩阵(正交矩阵)后, $\|\cdot\|_F$ 的值不变.

证明 若记 \mathbf{A} 的第 i 列为 $\boldsymbol{\alpha}_i (i=1, 2, \dots, n)$, 则有

$$\begin{aligned} \|\mathbf{UA}\|_F^2 &= \|\mathbf{U}(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_n)\|_F^2 \\ &= \|(\mathbf{U}\boldsymbol{\alpha}_1, \mathbf{U}\boldsymbol{\alpha}_2, \dots, \mathbf{U}\boldsymbol{\alpha}_n)\|_F^2 \\ &= \sum_{i=1}^n \|\mathbf{U}\boldsymbol{\alpha}_i\|_2^2 = \sum_{i=1}^n \|\boldsymbol{\alpha}_i\|_2^2 = \|\mathbf{A}\|_F^2, \end{aligned}$$

即

$$\|\mathbf{UA}\|_F = \|\mathbf{A}\|_F,$$

又

$$\begin{aligned}\|AV\|_F &= \|(AV)^H\|_F = \|V^H A^H\|_F \\ &= \|A^H\|_F = \|A\|_F.\end{aligned}$$

证毕

推论 与 A 酉(或正交)相似的矩阵的 F -范数是相同的, 即若 $B=U^H A U$, 则 $\|B\|_F = \|A\|_F$, 其中 $A \in \mathbb{C}^{m \times m}$, U 为酉矩阵.

4.2.3 谱范数的性质和谱半径

我们知道, 矩阵的算子范数 $\|A\|_2$ 称为 A 的谱范数, 它的值是通过矩阵 $A^H A$ 的最大特征值来计算的, 尽管求特征值比较麻烦, 但这种范数有非常好的性质, 所以在矩阵分析和系统理论中常常使用. 下面专门讨论谱范数的性质.

定理 4.2.5 设 $A \in \mathbb{C}^{m \times n}$, 则

- (1) $\|A\|_2 = \max_{\|x\|_2 = \|y\|_2 = 1} |y^H A x|, x \in \mathbb{C}^n, y \in \mathbb{C}^m$;
- (2) $\|A^H\|_2 = \|A\|_2$;
- (3) $\|A^H A\|_2 = \|A\|_2^2$.

证明 (1) 对满足 $\|x\|_2 = \|y\|_2 = 1$ 的 x 与 y 有

$$|y^H A x| \leq \|y\|_2 \|A x\|_2 \leq \|A\|_2.$$

又设有 $\|x\|_2 = 1$ 的 x , 并使 $\|A x\|_2 = \|A\|_2 \neq 0$, 若令 $y = \frac{A x}{\|A x\|_2}$, 就有

$$|y^H A x| = \frac{\|A x\|_2^2}{\|A x\|_2} = \|A x\|_2 = \|A\|_2,$$

从而 $\max_{\|x\|_2 = \|y\|_2 = 1} |y^H A x| = \|A\|_2$.

- (2) $\|A\|_2 = \max_{\|x\|_2 = \|y\|_2 = 1} |y^H A x|$
 $= \max_{\|x\|_2 = \|y\|_2 = 1} |x^H A^H y| = \|A^H\|_2$.

- (3) 由 $\|A^H A\|_2 \leq \|A^H\|_2 \|A\|_2, \|A^H\|_2 = \|A\|_2$, 可知

$$\|A^H A\|_2 \leq \|A\|_2^2. \quad (4.2.13)$$

令 $\|x\|_2 = 1$, 并使 $\|A x\|_2 = \|A\|_2$, 于是

$$\begin{aligned}\|A^H A\|_2 &\geq \max_{\|x\|_2 = 1} |x^H A^H A x| \\ &= \max_{\|x\|_2 = 1} \|A x\|_2^2 = \|A\|_2^2.\end{aligned} \quad (4.2.14)$$

由式(4.2.13)和式(4.2.14)即知(3)成立.

定理 4.2.6 设 $A \in \mathbb{C}^{m \times n}, U \in \mathbb{C}^{m \times m}, V \in \mathbb{C}^{n \times n}$ 且 $U^H U = I_m, V^H V = I_n$, 则

$$\|U A V\|_2 = \|A\|_2. \quad (4.2.15)$$

证明 令 $v = V^H x, u = U y$, 则

$$\begin{aligned}\|x\|_2 = 1 &\text{ 当且仅当 } \|v\|_2 = 1; \\ \|y\|_2 = 1 &\text{ 当且仅当 } \|u\|_2 = 1;\end{aligned}$$

于是

$$\begin{aligned}\|A\|_2 &= \max_{\|x\|_2 = \|y\|_2 = 1} |y^H Ax| \\ &= \max_{\|v\|_2 = \|u\|_2 = 1} |u^H UAVv| \\ &= \|UAV\|_2.\end{aligned}$$

证毕

对任何一种算子范数,还有如下的性质.

定理 4.2.7 设 $A \in \mathbb{C}^{n \times n}$, 若 $\|A\| < 1$, 则 $I - A$ 为非奇异矩阵, 且

$$\|(I - A)^{-1}\| \leq (1 - \|A\|)^{-1}. \quad (4.2.16)$$

证明 设 x 为任一非零向量, 则

$$\begin{aligned}\|(I - A)x\| &= \|x - Ax\| \\ &\geq \|x\| - \|Ax\| \\ &\geq \|x\| - \|A\| \|x\| \\ &= (1 - \|A\|) \|x\| > 0.\end{aligned}$$

于是, 若 $x \neq 0$, 则 $(I - A)x \neq 0$, 从而方程

$$(I - A)x = 0$$

无非零解, 故矩阵 $I - A$ 非奇异.

因为 $I - A$ 非奇异, 故有

$$(I - A)(I - A)^{-1} = I,$$

于是

$$\begin{aligned}(I - A)^{-1} &= [(I - A) + A](I - A)^{-1} \\ &= (I - A)(I - A)^{-1} + A(I - A)^{-1} \\ &= I + A(I - A)^{-1},\end{aligned}$$

从而

$$\begin{aligned}\|(I - A)^{-1}\| &\leq \|I\| + \|A\| \|(I - A)^{-1}\| \\ &= 1 + \|A\| \|(I - A)^{-1}\|,\end{aligned}$$

即

$$\|(I - A)^{-1}\| \leq \frac{1}{1 - \|A\|}. \quad \text{证毕}$$

下面我们引进一个数值代数中讨论收敛性时经常遇到的概念——谱半径.

定义 4.2.3 设 $A \in \mathbb{C}^{n \times n}$, $\lambda_1, \lambda_2, \dots, \lambda_n$ 为 A 的特征值, 我们称

$$\rho(A) = \max_i |\lambda_i| \quad (4.2.17)$$

为 A 的谱半径.

谱半径在几何上可以解释为: 以原点为圆心、能包含 A 的全部特征值的圆的半径中最小的一个.

定理 4.2.8 (特征值上界定理) 对任意矩阵 $A \in \mathbb{C}^{n \times n}$, 总有

$$\rho(A) \leq \|A\|, \quad (4.2.18)$$

即 A 的谱半径 $\rho(A)$ 不会超过 A 的任何一种范数.

证明 设 λ 是 A 的任一特征值, x 为相应的特征向量, 则有 $Ax = \lambda x$, 再由相容性条件, 有

$$|\lambda| \|x\| = \|\lambda x\| \leq \|A\| \|x\|,$$

即有 $|\lambda| \leq \|A\|$, 故得

$$\rho(\mathbf{A}) \leq \|\mathbf{A}\|. \quad \text{证毕}$$

特别地,如果 \mathbf{A} 为正规矩阵(包含实对称矩阵),则有下列的结果.

定理 4.2.9 如果 $\mathbf{A} \in \mathbb{C}^{n \times n}$,且 \mathbf{A} 为正规矩阵,则

$$\rho(\mathbf{A}) = \|\mathbf{A}\|_2. \quad (4.2.19)$$

证明 由于 \mathbf{A} 是正规矩阵,所以存在有 $\mathbf{U}, \mathbf{U}^H \mathbf{U} = \mathbf{I}$,使

$$\mathbf{U}^H \mathbf{A} \mathbf{U} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = \mathbf{\Lambda},$$

于是由定理 4.2.6 知

$$\begin{aligned} \|\mathbf{A}\|_2 &= \|\mathbf{U}^H \mathbf{A} \mathbf{U}\|_2 = \|\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)\|_2 \\ &= \sqrt{(\mathbf{\Lambda}^H \mathbf{\Lambda}) \text{ 的特征值绝对值的最大值}} = \sqrt{\max_i |\lambda_i|} \\ &= \sqrt{\max_i |\lambda_i|^2} = \rho(\mathbf{A}). \end{aligned} \quad \text{证毕}$$

定理 4.2.10 对任意非奇异矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$, \mathbf{A} 的谱范数为

$$\|\mathbf{A}\|_2 = \sqrt{\rho(\mathbf{A}^H \mathbf{A})} = \sqrt{\rho(\mathbf{A} \mathbf{A}^H)}. \quad (4.2.20)$$

证明 $\|\mathbf{A}\|_2 = \sqrt{(\mathbf{A}^H \mathbf{A}) \text{ 的最大特征值}}$

$$= \sqrt{\max_i |\lambda_i(\mathbf{A}^H \mathbf{A})|} = \sqrt{\rho(\mathbf{A}^H \mathbf{A})}.$$

又因为 $\mathbf{A} \mathbf{A}^H = \mathbf{A}(\mathbf{A}^H \mathbf{A})\mathbf{A}^{-1}$, 即 $\mathbf{A} \mathbf{A}^H \sim \mathbf{A}^H \mathbf{A}$, 所以它们有相同的特征值,从而有

$$\|\mathbf{A}\|_2 = \sqrt{\rho(\mathbf{A}^H \mathbf{A})} = \sqrt{\rho(\mathbf{A} \mathbf{A}^H)}. \quad \text{证毕}$$

习 题 4 (2)

1. 求矩阵 $\mathbf{A} = \begin{bmatrix} -1 & 2 & 1 \\ 1 & 0 & i \end{bmatrix}$ 和 $\mathbf{B} = \begin{bmatrix} -i & 2 & 3 \\ 1 & 0 & i \end{bmatrix}$ 的范数 $\|\cdot\|_1, \|\cdot\|_\infty$ 及 $\|\cdot\|_2$.

2. 设 $\mathbf{A} = \begin{pmatrix} 2 & 1 & 0 \\ -1 & 2 & 3 \\ 0 & -2 & 1 \end{pmatrix}$, 计算 $\|\mathbf{A}\|_1, \|\mathbf{A}\|_2, \|\mathbf{A}\|_\infty$ 及 $\|\mathbf{A}\|_F$.

3. 设 $\mathbf{P} \in \mathbb{R}^{n \times n}$ 可逆, 又已知 $\|\boldsymbol{\alpha}\| = \|\mathbf{P}^{-1} \boldsymbol{\alpha}\|_1$ 是 \mathbb{R}^n 中的向量范数, 试求 \mathbf{A} 从属于该向量范数 $\|\boldsymbol{\alpha}\|$ 的矩阵算子范数 $\|\mathbf{A}\|$ (提示: 将 \mathbf{A} 这样的算子范数 $\|\mathbf{A}\|$ 用另一个矩阵的 1-范数来表示即可).

4. 设 $\mathbf{A} \in \mathbb{R}^{2 \times 2}, \boldsymbol{\alpha} = (x_1, x_2)^T$, 常数 $h > 0$. 证明: 实数 $\max\left(|x_1|, \frac{|x_2 - x_1|}{h}\right)$ 是向量 $\boldsymbol{\alpha}$ 的范数, 并求从属于该向量范数的矩阵算子范数 $\|\mathbf{A}\|_h$ (提示: 将 \mathbf{A} 这样的算子范数 $\|\mathbf{A}\|_h$ 用另一个矩阵的 ∞ -范数来表示即可).

5. 设 $\mathbf{A} = (a_{ij})_{n \times n}$, 举例说明 $\max_{1 \leq i, j \leq n} |a_{ij}|$ 不是矩阵 \mathbf{A} 的范数.

6. 设 $\mathbf{A} = (a_{ij})_{m \times n}$, 分别定义实数:

$$(1) \|\mathbf{A}\| = \sqrt{mn} \cdot \max_{i,j} |a_{ij}|;$$

$$(2) \|\mathbf{A}\| = \max\{m, n\} \cdot \max_{i,j} |a_{ij}|;$$

验证它们都是 $\mathbb{C}^{m \times n}$ 中的矩阵范数.

7. 设 $\mathbf{P} \in \mathbb{C}^{n \times n}$ 可逆, 已知 $\mathbb{C}^{n \times n}$ 中有矩阵范数 $\|\cdot\|_M$, 对于 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 定义实数 $\|\mathbf{A}\| = \|\mathbf{P}^{-1} \mathbf{A} \mathbf{P}\|_M$, 验证 $\|\mathbf{A}\|$ 是 $\mathbb{C}^{n \times n}$ 中的一种矩阵范数.

8. 给定 $\mathbb{C}^{n \times n}$ 中的两种矩阵范数 $\|\cdot\|_M$ 与 $\|\cdot\|_S$, 对于 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 分别定义实数:

$$(1) \|A\| = \max\{\|A\|_M, \|A\|_S\};$$

$$(2) \|A\| = \|A\|_M + 2\|A\|_S;$$

验证它们都是 $\mathbb{C}^{n \times n}$ 中的矩阵范数.

9. 证明:

$$\frac{1}{\sqrt{n}} \|A\|_F \leq \|A\|_2 \leq \|A\|_F.$$

10. 对所有的非奇异矩阵 A 和 B , 在算子范数意义下证明:

$$(1) \|A^{-1}\| \geq \frac{1}{\|A\|};$$

$$(2) \|A^{-1} - B^{-1}\| \leq \|A^{-1}\| \|B^{-1}\| \|A - B\|.$$

$$11. \text{证明: 若 } \|A\| < 1, \text{ 则 } \|I - (I - A)^{-1}\| \leq \frac{\|A\|}{1 - \|A\|}.$$

12. 设 λ 为矩阵 $A \in \mathbb{C}^{n \times n}$ 的特征值, 证明:

$$|\lambda| \leq \sqrt[m]{\|A^m\|}.$$

13. 设矩阵 A 非奇异, λ 是它的任一个特征值, 证明:

$$|\lambda| \geq \frac{1}{\|A^{-1}\|}.$$

14. 试证:

(1) $\mathbb{C}^{n \times n}$ 上的矩阵 m_1 范数与 \mathbb{C}^n 上向量的 1 范数是相容的;

(2) $\mathbb{C}^{n \times n}$ 上的矩阵 m_∞ 范数与 \mathbb{C}^n 上向量的 ∞ 范数是相容的.

15. 设 $A \in \mathbb{C}^{m \times n}$, $\alpha \in \mathbb{C}^n$, 证明矩阵范数 $\|A\|_{m_1}$ 与向量的 p -范数 $\|\alpha\|_p$ ($1 \leq p < \infty$) 相容.

16. 设 $A = (a_{ij})_{m \times n}$, 列向量 $\alpha \in \mathbb{C}^n$, 证明矩阵范数 $\|A\| = \sqrt{mn} \cdot \max_{i,j} |a_{ij}|$ 与向量的 2-范数相容.

17. 设 $A = (a_{ij})_{m \times n}$, 列向量 $\alpha \in \mathbb{C}^n$. 证明: 矩阵范数

$$\|A\| = \max\{m, n\} \max_{i,j} |a_{ij}|$$

与向量 2-范数和 ∞ -范数都相容.

18. 证明

$$(1) \|A\|_F = \left(\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2} \text{ 与向量范数 } \|x\|_2 \text{ 相容};$$

$$(2) \|A\|_m = \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \text{ 与向量范数 } \|x\|_1 \text{ 相容}.$$

19. 设 $A = (a_{ij})_{n \times n}$ 及 \mathbb{C}^n 的列向量 $x = (x_1, x_2, \dots, x_n)^T$, 证明: 矩阵范数 $\|A\| = n \cdot \max |a_{ij}|$ 与向量 2 范数 $\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{1/2}$ 相容.

20. 设 $\|A\|_a$ 是 $\mathbb{C}^{m \times n}$ 中的一个已知矩阵范数, 给定 \mathbb{C}^n 中一个向量 $\alpha \neq 0$, 定义: $\forall x \in \mathbb{C}^m, \|x\|_u = \|x\alpha^H\|_a$, 证明 $\|x\|_u$ 是 \mathbb{C}^m 中的向量范数, 且是与 $\|A\|_a$ 相容的范数.

21. 试证: 对于 $\mathbb{C}^{m \times n}$ 中任意给定的矩阵范数 $\|\cdot\|_M$, 一定有多多个与之相容的向量范数.

22. 举例说明: $\mathbb{C}^{m \times n}$ ($n > 1$) 中矩阵的 1-范数与 \mathbb{C}^n 中向量的 ∞ -范数不相容.

23. 设 $\|\cdot\|$ 是 $\mathbb{C}^{n \times n}$ 上的矩阵范数, 证明:

$$(1) \|I\| \geq 1;$$

(2) 设 A 为 n 阶可逆矩阵, λ 是 A 的任意特征值, 那么

$$\|A^{-1}\|^{-1} \leq |\lambda| \leq \|A\|.$$

24. 已知 $u \in \mathbb{R}^n$ ($n > 1$) 为一个单位列向量, 令 $A = I - uu^T$, 试证:

$$(1) \|A\|_2 = 1;$$

(2) 对任意的 $X \in \mathbb{R}^n$, 如果有 $AX \neq X$, 那么

$$\|AX\|_2 < \|X\|_2.$$

25. 设 $A, B \in \mathbb{C}^{n \times n}$, 证明:

$$\|AB\|_F \leq \min\{\|A\|_2 \|B\|_F, \|A\|_F \|B\|_2\}.$$

26. 设 $A \in \mathbb{C}^{n \times n}$ 的特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 证明:

$$\|A\|_F \geq (|\lambda_1|^2 + |\lambda_2|^2 + \dots + |\lambda_n|^2)^{\frac{1}{2}}.$$

27. 证明: 若 $A \in \mathbb{C}^{n \times n}$, 则有 $\|A\|_2 \leq \|A\|_F$.

28. 举例说明矩阵的谱半径不是矩阵范数.

29. 设 $A, B \in \mathbb{C}^{n \times n}$ 都是对称矩阵, 求证:

$$\rho(A+B) \leq \rho(A) + \rho(B).$$

30. 证明下面的矩阵 A 的谱半径 $\rho(A) \leq 1$,

$$A = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{6} & \frac{1}{6} & \frac{3}{6} & \frac{1}{6} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{3}{7} \end{pmatrix}.$$

31. 设 $\|\cdot\|_o$ 是 $\mathbb{C}^{n \times n}$ 上的矩阵范数, P 是可逆矩阵, 对任意的矩阵 $A \in \mathbb{C}^{n \times n}$, 记 $\|A\|_\beta = \|P^{-1}AP\|_o$, 证明 $\|\cdot\|_\beta$ 也是 $\mathbb{C}^{n \times n}$ 上的矩阵范数.

32. 设 $A = (a_{ij}) \in \mathbb{C}^{n \times n}$, 试证: $\|A\| = [\text{tr}(A^H A)]^{1/2}$ 是矩阵范数.

33. 设 $A \in \mathbb{C}^{n \times n}$, 且其奇异值为 $\sigma_1, \sigma_2, \dots, \sigma_r$, 试证

$$\|A\|_2 = \max_i \sigma_i.$$

34. 设 $A \in \mathbb{C}^{n \times n}$, 证明:

$$(1) \|A\|_2 = \|A^H\|_2 = \|A^T\|_2 = \|\bar{A}\|_2;$$

$$(2) \|A^H A\|_2 = \|A A^H\|_2 = \|A\|_2^2;$$

$$(3) \|A\|_2 = \max_{\|x\|_2 = \|y\|_2 = 1} |y^H A x|;$$

$$(4) \|A\|_2^2 \leq \|A\|_1 \|A\|_\infty.$$

35. 设 $A \in \mathbb{C}^{n \times n}$, 且 $A^H A = I_{n \times n}$, 试证:

$$\|A\|_2 = 1, \quad \|A\|_F = \sqrt{n}.$$

36. 设 $\|\cdot\|$ 是 $\mathbb{C}^{n \times n}$ 上的矩阵范数, 证明: 对任意的 $A \in \mathbb{C}^{n \times n}$ 都有 $\rho(A) = \lim_{k \rightarrow \infty} \|A^k\|^{1/k}$.

37. 设 A 为埃尔米特矩阵, 证明

$$\|A\|_2 = \max_i |\lambda_i(A)|.$$

38. 证明:

(1) 酉矩阵 U 的谱范数等于 1;

(2) 设 $A \in \mathbb{C}^{n \times n}$, U, V 为 n 阶酉阵, 则

$$\|UA\|_2 = \|AV\|_2 = \|UAV\|_2 = \|A\|_2.$$

39. 设 $A \in \mathbb{C}^{n \times n}$, 且 A 可逆, 若 λ 是 A 的任一特征值, 证明:

$$\frac{1}{\|A^{-1}\|_2} \leq |\lambda| \leq \|A\|_2.$$

40. 设 $\|A\|_\rho$ 是由向量 $\|x\|_\rho$ 诱导的矩阵范数, A 为可逆矩阵, 证明:

$$(1) \|A^{-1}\|_\rho \geq \frac{1}{\|A\|_\rho};$$

$$(2) \frac{1}{\|A^{-1}\|_\rho} = \min_{x \neq 0} \frac{\|Ax\|_\rho}{\|x\|_\rho}.$$

41. A 是 n 阶正规矩阵, 证明

$$\|A\|_2 = \rho(A) \quad (\rho(A) \text{ 是 } A \text{ 的谱半径}).$$

42. 设 $A = (a_{ij}) \in \mathbb{R}^{n \times n}$, 且 $a_{ij} \geq 0$, 每一行元素之和为常数 α , 证明: A 的谱半径 $\rho(A) = \|A\|_\infty$.

43. (1) 设 $A \in \mathbb{C}^{n \times n}$, A 的特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 证明:

$$\|A\|_F \geq \left(\sum_{i=1}^n |\lambda_i|^2 \right)^{1/2};$$

(2) 设 $A \in \mathbb{C}^{m \times n}$, $\text{rank}(A) = r$, 且 A 的非零奇异值为 $\sigma_1, \sigma_2, \dots, \sigma_r$, 证明:

$$\|A\|_F = (\sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2)^{1/2}.$$

44. (1) 设 $A \in \mathbb{C}^{m \times n}$, $B \in \mathbb{C}^{n \times s}$, 记 $\sigma_1(M)$ 为矩阵 M 的最大奇异值, 证明:

$$\sigma_1(AB) \leq \sigma_1(A)\sigma_1(B);$$

(2) 若 $A, B \in \mathbb{C}^{m \times n}$, 则

$$\sigma_1(A+B) \leq \sigma_1(A) + \sigma_1(B).$$

45. 设 σ_1 和 σ_n 是矩阵 A 的最大奇异值和最小奇异值, 证明: $\sigma_1 = \|A\|_2$; 当 A 是非奇异矩阵时, $\|A^{-1}\|_2 = \frac{1}{\sigma_n}$.

46. 设 $A \in \mathbb{C}^{n \times n}$, $\|A\| < 1$, 证明:

(1) $I+A$ 可逆, 且

$$\frac{\|I\|}{1 + \|A\|} \leq \|(I+A)^{-1}\| \leq \frac{\|I\|}{1 - \|A\|};$$

(2) $\|I - (I+A)^{-1}\| \leq \frac{\|A\|}{1 - \|A\|}$.

47. 设 $X \in \mathbb{R}^n$, A 是 n 阶实对称正定矩阵, 定义向量范数

$$\|X\|_A = \sqrt{X^T A X},$$

又设 $f(t)$ 是 m 次实系数多项式, 证明

$$\|f(A)X\|_A \leq \max\{|f(\lambda_i)|\} \|X\|_A,$$

式中 λ_i 是 A 的特征值, $i=1, 2, \dots, n$.

4.3 摄动分析与矩阵的条件数

在数值计算中, 通常存在两类误差影响计算结果的精度, 即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. 下面我们将分别研究在线性方程组求解和矩阵特征值求解过程中, 因原始数据的摄动而引起问题的解有多大的变化, 即研究问题解的稳定性.

4.3.1 病态方程组与病态矩阵

在求解线性方程组问题中, 假定系数矩阵和自由项的元素有摄动, 该摄动有时会对解的精度产生巨大的影响. 不妨先看一个简单的例子.

考察一个二元线性方程组

$$\begin{bmatrix} 1 & 0.99 \\ 0.99 & 0.98 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad (4.3.1)$$

可以验证, 该方程组的精确解是 $x_1 = 100, x_2 = -100$.

如果系数矩阵有摄动 $\delta A = \begin{bmatrix} 0 & 0 \\ 0 & 0.01 \end{bmatrix}$, 并且右端项也有摄动 $\delta b = \begin{bmatrix} 0 \\ 0.001 \end{bmatrix}$, 则摄动后的线性方程组为

$$\begin{bmatrix} 1 & 0.99 \\ 0.99 & 0.99 \end{bmatrix} \begin{bmatrix} x_1 + \delta x_1 \\ x_2 + \delta x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1.001 \end{bmatrix}.$$

可以验证, 这个方程组的精确解变为 $x_1 + \delta x_1 = -0.1, x_2 + \delta x_2 = \frac{10}{9}$.

可见, 系数矩阵和右端项的微小摄动引起了解的巨大变化. 这种现象通常叫做“病态”. 所以, 我们有如下的定义.

定义 4.3.1 如果系数矩阵 A 或常数项 b 的微小变化, 引起方程组 $Ax = b$ 解的巨大变化, 则称方程组为病态方程组, 其系数矩阵 A 就叫做对应于解方程组(或求逆)的病态矩阵; 反之, 方程组就称为良态方程组, A 称良态矩阵.

应该指出, 谈到“病态矩阵”概念时, 必须明确它是对什么而言的. 因为对于解方程组(或求逆)来说是病态矩阵, 对于求特征值来说并不一定是病态的, 反之亦然. 所以, 我们不能笼统地说某个矩阵是“病态”的, 这里所说的“病态”就是相对于解方程组而言的.

还应指出, “病态”是系数矩阵本身的特性, 与所用的计算工具和计算方法无关. 但是, 实际计算中“病态”的程度却是通过所用的计算工具等表现出来的, 例如计算机的字长愈长, “病态”现象在程度上就会相对地减轻.

了解“病态”的概念以后, 我们希望能给出衡量一个矩阵是否“病态”的标准. 一个矩阵的“病态”程度, 大体上可以用“条件数”来衡量.

4.3.2 矩阵的条件数

设有方程组

$$Ax = b. \quad (4.3.2)$$

我们先来弄清楚系数矩阵和自由项有一个微小的变化时, 方程组的解是怎样变化的, 这个问题也叫做摄动分析.

设 A 是精确的, b 有误差 δb , 解为 $x + \delta x$, 则

$$\begin{aligned} A(x + \delta x) &= b + \delta b, \\ \delta x &= A^{-1} \delta b, \\ \|\delta x\| &\leq \|A^{-1}\| \|\delta b\|, \end{aligned} \quad (4.3.3)$$

这里对范数不加下标, 以表示任取一种范数.

由式(4.3.2), 有

$$\|b\| \leq \|A\| \|x\|,$$

即

$$\frac{1}{\|x\|} \leq \frac{\|A\|}{\|b\|}, \quad (4.3.4)$$

设 $b \neq 0$, 于是由式(4.3.3)及式(4.3.4)得下面定理.

定理 4.3.1 设 A 是非奇异矩阵, $Ax = b \neq 0$, 且

$$A(x + \delta x) = b + \delta b,$$

则

$$\frac{\|\delta x\|}{\|x\|} \leq \|A^{-1}\| \|A\| \frac{\|\delta b\|}{\|b\|}, \quad (4.3.5)$$

上式给出了解的相对误差的上界,常数项 b 的相对误差在解中可能放大 $\|A^{-1}\| \|A\|$ 倍.

现设 b 是精确的, A 有微小误差(摄动) δA , 解为 $x + \delta x$, 则

$$\begin{aligned} (A + \delta A)(x + \delta x) &= b, \\ (A + \delta A)\delta x &= -(\delta A)x, \end{aligned} \quad (4.3.6)$$

如果 δA 不受限制的话, $A + \delta A$ 可能奇异, 而

$$(A + \delta A) = A(I + A^{-1}\delta A),$$

由上节定理 4.2.7 知, 当 $\|A^{-1}\delta A\| < 1$ 时, $(I + A^{-1}\delta A)^{-1}$ 存在, 由式(4.3.6)有

$$\delta x = -(I + A^{-1}\delta A)^{-1}A^{-1}(\delta A)x,$$

因此

$$\|\delta x\| \leq \frac{\|A^{-1}\| \|\delta A\| \|x\|}{1 - \|A^{-1}\delta A\|}.$$

如果矩阵 A 的微小变化 δA 满足条件

$$\|A^{-1}\| \|\delta A\| < 1, \quad (4.3.7)$$

则有

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{\|A^{-1}\| \|A\| \frac{\|\delta A\|}{\|A\|}}{1 - \|A^{-1}\| \|A\| \frac{\|\delta A\|}{\|A\|}}. \quad (4.3.8)$$

这一不等式说明了解的相对变化与系数矩阵的相对变化之间的关系, 从其中可以看出, 只要满足条件式(4.3.7), 对于系数矩阵的同样相对变化来说, $\|A^{-1}\| \|A\|$ 愈大, 解的相对变化也愈大; $\|A^{-1}\| \|A\|$ 愈小, 解的相对变化就愈小. 综合式(4.3.5)与式(4.3.8)可以看出, $\|A^{-1}\| \|A\|$ 在某种程度上刻画了方程组的解对于原始数据变化的灵敏度, 也就是刻画了方程组“病态”的程度, 于是引进下述定义.

定义 4.3.2 设 A 为非奇异矩阵, 称数 $\text{cond}(A) = \|A^{-1}\|_p \|A\|_p$ ($p=1, 2$ 或 ∞) 为矩阵 A 的条件数.

由此看出矩阵的条件数与范数有关, 它刻画了方程组解的相对误差可能的放大率, 若 $\text{cond}(A) \gg 1$, 则方程组(4.3.2)是“病态”的(即 A 是“病态”矩阵, 或者说 A 是坏条件的); 若 $\text{cond}(A)$ 相对地小, 则方程组(4.3.2)是“良态”的(或者说 A 是好条件的). 究竟条件数多大矩阵才算病态, 一般来讲是没有具体标准的, 也只是相对而言.

通常使用的条件数, 有

$$(1) \text{cond}(A)_\infty = \|A^{-1}\|_\infty \|A\|_\infty;$$

(2) A 的谱条件数

$$\begin{aligned} \text{cond}(A)_2 &= \|A\|_2 \|A^{-1}\|_2 \\ &= \sqrt{\frac{\lambda_{\max}(A^H A)}{\lambda_{\min}(A^H A)}}. \end{aligned}$$

显然, 当 A 是实对称矩阵时,

$$\text{cond}(A)_2 = \frac{|\lambda_1|}{|\lambda_n|}, \quad (4.3.9)$$

其中 λ_1 与 λ_n 分别为矩阵 \mathbf{A} 的按模最大和最小特征值.

条件数的性质:

(1) 对任何非奇异矩阵 \mathbf{A} , 都有 $\text{cond}(\mathbf{A})_p \geq 1$. 事实上,

$$\text{cond}(\mathbf{A})_p = \|\mathbf{A}^{-1}\|_p \|\mathbf{A}\|_p \geq \|\mathbf{A}^{-1}\mathbf{A}\|_p = 1.$$

(2) 设 \mathbf{A} 为非奇异矩阵, $k \neq 0$ (常数), 则

$$\text{cond}(k\mathbf{A})_p = \text{cond}(\mathbf{A})_p.$$

(3) 如果 \mathbf{A} 为正交矩阵, 则 $\text{cond}(\mathbf{A})_2 = 1$; 如果 \mathbf{A} 为非奇异矩阵, \mathbf{R} 为正交矩阵, 则

$$\begin{aligned} \text{cond}(\mathbf{R}\mathbf{A})_F &= \text{cond}(\mathbf{A}\mathbf{R})_F \\ &= \text{cond}(\mathbf{A})_F. \end{aligned}$$

例 4.3.1 对希尔伯特(Hilbert)矩阵

$$\mathbf{H}_n = \begin{pmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & \cdots & \frac{1}{2n-1} \end{pmatrix},$$

求 \mathbf{H}_3 的条件数.

解

$$\mathbf{H}_3 = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix}, \quad \mathbf{H}_3^{-1} = \begin{pmatrix} 9 & -36 & 30 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{pmatrix},$$

$$\|\mathbf{H}_3\|_\infty = \frac{11}{6}, \quad \|\mathbf{H}_3^{-1}\|_\infty = 408,$$

所以 $\text{cond}(\mathbf{H}_3)_\infty = 748$.

同样可计算 $\text{cond}(\mathbf{H}_6)_\infty = 2.9 \times 10^6$, 对于一般矩阵 \mathbf{H}_n , 当 n 愈大时, 病态愈严重.

例 4.3.2 设

$$\mathbf{A} = \begin{bmatrix} 1 & 10^4 \\ 1 & 1 \end{bmatrix},$$

计算 $\text{cond}(\mathbf{A})_\infty$.

解

$$\mathbf{A}^{-1} = \frac{1}{10^4 - 1} \begin{bmatrix} -1 & 10^4 \\ 1 & -1 \end{bmatrix},$$

$$\text{cond}(\mathbf{A})_\infty = \frac{(1 + 10^4)^2}{10^4 - 1} \approx 10^4.$$

可见, 当矩阵 \mathbf{A} 的元素大小不均匀时, 往往造成很高的条件数, 在这种情况下, 如果要解方程组, 则可以对 \mathbf{A} 的行引进适当的比例因子, 以减少条件数.

例 4.3.3 设有方程组

$$\begin{bmatrix} 1 & 10^4 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 10^4 \\ 2 \end{bmatrix},$$

在 \mathbf{A} 的第 1 行引进比例因子, 如用 $s_1 = \max_{1 \leq j \leq 2} |a_{1j}| = 10^4$ 除第 1 个方程, 得 $\mathbf{A}'\mathbf{x} = \mathbf{b}'$, 即

$$\begin{bmatrix} 10^{-4} & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix},$$

$$(\mathbf{A}')^{-1} = \frac{1}{1-10^{-4}} \begin{bmatrix} -1 & 1 \\ 1 & -10^{-4} \end{bmatrix},$$

于是

$$\text{cond}(\mathbf{A}')_{\infty} = \frac{4}{1-10^{-4}} \approx 4.$$

4.3.3 矩阵特征值的摄动分析

前面已从线性方程组求解问题的摄动分析, 引出了矩阵的条件数的概念. 现在再讨论矩阵元素的摄动对矩阵特征值的影响, 我们会发现这种影响依然可以用“条件数”来说明.

下面先给出特征值的圆盘定理.

一般来说, 精确计算矩阵的特征值往往是很困难的(因为 5 次及 5 次以上的代数方程无公式解), 因此更为合理的办法是对特征值范围作出估计. 复数域上 n 阶矩阵的特征值可以用复平面上的点来表示, 所以对这些点的位置的估计就是特征值的估计.

定义 4.3.3 设 $\mathbf{A} = (a_{ij})$ 为任一 n 阶复数矩阵, 复平面上的 n 个圆盘

$$G_i(\mathbf{A}): |z - a_{ii}| \leq R_i, \quad i = 1, 2, \dots, n,$$

这里以 $R_i = \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|$ 为半径的圆(即圆盘的边界), 称为矩阵 \mathbf{A} 的 Gerschgorin 圆, 简称盖尔圆.

定理 4.3.2(盖尔定理又称圆盘定理) 设 $\mathbf{A} = (a_{ij}) \in \mathbb{C}^{n \times n}$, 则

(1) \mathbf{A} 的特征值都在 n 个圆盘 $G_i(\mathbf{A})$ 的并集内(换句话说, \mathbf{A} 的每个特征值都落在 \mathbf{A} 的某个圆盘之内), 即

$$\lambda(\mathbf{A}) \subseteq \bigcup_{i=1}^n G_i(\mathbf{A});$$

(2) 矩阵 \mathbf{A} 的任一个由 m 个圆盘组成的连通区域中, 有且只有 \mathbf{A} 的 m 个特征值(当 \mathbf{A} 的主对角线上有相同元素时, 则按重复次数计算, 有相同特征值时也需按重复次数计算).

证明 (1) 设 λ 是 \mathbf{A} 的任一个特征值, 相应的特征向量为 $\mathbf{x} = (x_1, \dots, x_n)^T$. 令 $|x_{i_0}| = \max_i |x_i|$, 则 $|x_{i_0}| > 0$. 由 $\mathbf{Ax} = \lambda\mathbf{x}$ 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

$$|\lambda - a_{i_0 i_0}| = \sum_{\substack{j=1 \\ j \neq i_0}}^n |a_{i_0 j}| \frac{|x_j|}{|x_{i_0}|} \leq \sum_{\substack{j=1 \\ j \neq i_0}}^n |a_{i_0 j}|,$$

即

$$\lambda \in G_{i_0}(\mathbf{A}), \quad \text{故} \quad \lambda(\mathbf{A}) \subseteq \bigcup_{i=1}^n G_i(\mathbf{A}).$$

(2) 记 $\mathbf{D} = \text{diag}(a_{11}, \dots, a_{nn})$, $\mathbf{B} = \mathbf{A} - \mathbf{D}$, 令

$$\mathbf{A}(t) = \mathbf{D} + t\mathbf{B}, \quad 0 \leq t \leq 1.$$

不失一般性, 假定 n 个圆盘中前面 m 个圆盘 $G_1(\mathbf{A}), \dots, G_m(\mathbf{A})$ 构成一连通区域 $G' =$

$\bigcup_{i=1}^m G_i(\mathbf{A})$, 并与其余的 $n-m$ 个圆盘分离, 即 G' 与 $G'' = \bigcup_{i=m+1}^n G_i(\mathbf{A})$ 不相交. 记

$$G_i(\mathbf{A}(t)) = \left\{ |z - a_{ii}| \leq t \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \right\}, \quad i = 1, 2, \dots, n,$$

$$G'(t) = \bigcup_{i=1}^m G_i(\mathbf{A}(t)), \quad G''(t) = \bigcup_{i=m+1}^n G_i(\mathbf{A}(t)),$$

则

$$G_i(\mathbf{A}(t)) \subseteq G_i(\mathbf{A}), \quad i = 1, 2, \dots, n,$$

$$G'(t) \subseteq G', \quad G''(t) \subseteq G'',$$

且对所有 $t \in [0, 1]$, $G'(t)$ 与 $G''(t)$ 都不相交. 特别地, $G'(0)$ 恰好包含 $\mathbf{A}(0)$ 的 m 个特征值 $a_{11}, a_{22}, \dots, a_{mm}$. 由(1)知, 对所有的 $t \in [0, 1]$, $\mathbf{A}(t)$ 的特征值都包含在 $G'(t) \cup G''(t)$ 中.

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 $\lambda(\mathbf{A})$ 是矩阵元素的连续函数. 此时由于 $G'(t)$ 与 $G''(t)$ 不相交, 则当 t 增加时, $\mathbf{A}(t)$ 的特征值不能从 $G'(t)$ 跳跃到 $G''(t)$. 由以上讨论可知 $G'(0)$ 恰好包含 $\mathbf{A}(0)$ 的 m 个特征值, 所以对所有的 $t \in [0, 1]$, $G'(t)$ 恰好包含 $\mathbf{A}(t)$ 的 m 个特征值. 因此, $G'(1)$ (即 G') 恰好包含 $\mathbf{A}(1)$ (即 \mathbf{A}) 的 m 个特征值. 证毕

例 4.3.4 估计下面矩阵的特征值范围:

$$\mathbf{A} = \begin{pmatrix} 1 & 0.1 & 0.2 & 0.3 \\ 0.5 & 3 & 0.1 & 0.2 \\ 1 & 0.3 & -1 & 0.5 \\ 0.2 & -0.3 & -0.1 & -4 \end{pmatrix}.$$

解 盖尔定理(1)所指的 4 个圆盘为

$$|z - 1| \leq 0.1 + 0.2 + 0.3 = 0.6,$$

$$|z - 3| \leq 0.5 + 0.1 + 0.2 = 0.8,$$

$$|z + 1| \leq 1 + 0.3 + 0.5 = 1.8,$$

$$|z + 4| \leq 0.2 + 0.3 + 0.1 = 0.6.$$

画在复平面上如图 4.1 所示.

从图 4.1 可以看到, 第 1、第 3 两个圆盘是相交的(含有两个特征值), 而另外的两个都是孤立的(即除自身外不与其他 3 个圆盘的任何一个相交, 各含有一个特征值). 两个相交的圆盘的并集构成一个连通区域.

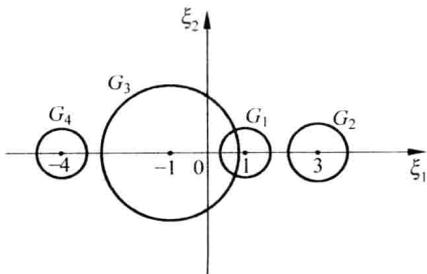


图 4.1

值得指出,由两个或两个以上的盖尔圆构成的连通部分,可能在其中的一个盖尔圆中有两个或两个以上的特征值,而在另外的一个或几个盖尔圆中没有特征值(见例 4.3.5).

例 4.3.5 讨论矩阵 $A = \begin{bmatrix} 1 & -0.8 \\ 0.5 & 0 \end{bmatrix}$ 的特征值分布状况.

解 特征方程为

$$(\lambda - 1)\lambda + 0.4 = 0,$$

故 A 的两个特征值为 $\lambda_{1,2} = \frac{1}{2}(1 \pm i\sqrt{0.6})$. 又 A 的两个圆盘为

$$|z - 1| \leq 0.8, \quad |z| \leq 0.5.$$

但因 $|\lambda_1| = |\lambda_2| = \sqrt{0.4} = 0.632456 > 0.5$, 因此这两个特征值全部落在圆盘 $|z - 1| \leq 0.8$ 内,而在圆盘 $|z| \leq 0.5$ 之外!(见图 4.2)

从上述例子,我们不难得到如下推论.

推论 若 n 阶实矩阵的每个圆盘与其余圆盘分离,则 A 的特征值均为实数. 为了得到有效的估计,往往希望每个圆盘只包含 A 的一个特征值(必然是实特征值),这就需要将每个圆盘的半径缩小. 通常有两种做法,其一是对 A^T 使用圆盘定理;其二是利用相似矩阵的特征值相同(即相似变换不改变特征值)这一性质,对矩阵 A 施行相似变换后,再应用圆盘定理. 确切地说,先构造矩阵 $B = DAD^{-1}$,

其中 $D = \text{diag}(d_1, d_2, \dots, d_n)$, d_j 的选取原则是:欲使第 j 个圆盘缩小,可取 $0 < d_j < 1$,而其余的 d_k 取为 1,此时 B 的其余圆盘相对放大;反之,欲使第 j 个圆盘放大,可取 $d_j > 1$,而其余的 d_k 取为 1,此时 B 的其余圆盘相对缩小. 于是,这样得到的相似矩阵 $B = DAD^{-1} =$

$\left(a_{ij} \frac{d_i}{d_j}\right)_{n \times n}$ 与 $A = (a_{ij})_{n \times n}$ 有相同的对角元.

例 4.3.6 隔离矩阵 $A = \begin{bmatrix} 20 & 5 & 0.8 \\ 4 & 10 & 1 \\ 1 & 2 & 10i \end{bmatrix}$ 的特征值.

解 A 的 3 个盖尔圆为

$$G_1: |z - 20| \leq 5.8,$$

$$G_2: |z - 10| \leq 5,$$

$$G_3: |z - 10i| \leq 3.$$

G_1 与 G_2 相交;而 G_3 孤立,其中恰好有 A 的一个特征值,记作 λ_3 (见图 4.3).

选取

$$D = \text{diag}(1, 1, 2),$$

则

$$B = DAD^{-1} = \begin{bmatrix} 20 & 5 & 0.4 \\ 4 & 10 & 0.5 \\ 2 & 4 & 10i \end{bmatrix}$$

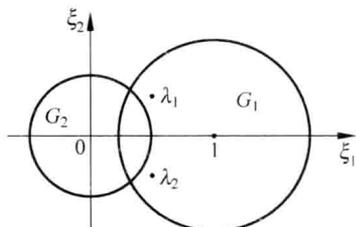


图 4.2

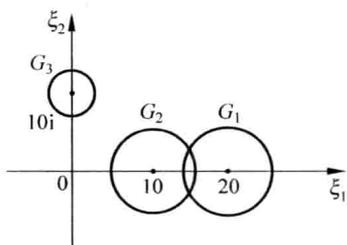


图 4.3

的3个盖尔圆为

$$G'_1: |z - 20| \leq 5.4,$$

$$G'_2: |z - 10| \leq 4.5,$$

$$G'_3: |z - 10i| \leq 6.$$

易见,这是3个孤立的盖尔圆,每个盖尔圆中恰好有 B 的(也是 A 的)一个特征值(见图 4.4). 注意 G'_3 中的特征值就是 G_3 中的特征值 λ_3 , 所以 A 的3个特征值分别位于 G'_1, G'_2 及 G'_3 之中,按上面推理,它们都是实数.

例 4.3.6 表明,对于矩阵 A ,适当选取正数 d_1, d_2, \dots, d_n ,可以获得只含 A 的一个特征值的孤立盖尔圆. 这样能够得到 A 的更精确的特征值的包含区域. 从几何上看,矩阵 A 的特征值全部位于以原点为圆心,谱半径 $\rho(A)$ 为半径的圆盘内.

必须注意,上述两种方法均有局限性,如当 A 对角线上元素均相同时会失效;又当 A 对角线上的元素为实数,但有非实的特征值时也失效.

定理 4.3.3 设 $A, E \in \mathbb{C}^{n \times n}$, 并且存在可逆矩阵 P 使得

$$P^{-1}AP = \text{diag}(\lambda_1, \dots, \lambda_n), \quad (4.3.10)$$

$A+E$ 有特征值 μ_1, \dots, μ_n , 则对任一 μ_j , 存在 λ_i 使得

$$|\lambda_i - \mu_j| \leq \|P^{-1}EP\|_{\infty}. \quad (4.3.11)$$

此外,如果 λ_i 是 m 重特征值,且圆盘

$$G_i = \{z \mid |z - \lambda_i| \leq \|P^{-1}EP\|_{\infty}\} \quad (4.3.12)$$

与圆盘

$$G_k = \{z \mid |z - \lambda_k| \leq \|P^{-1}EP\|_{\infty}\}, \quad \lambda_k \neq \lambda_i \quad (4.3.13)$$

不相交,则 G_i 恰好包含 $A+E$ 的 m 个特征值.

证明 令 $B = P^{-1}(A+E)P$, 则 B 的特征值为 μ_1, \dots, μ_n . 记 $P^{-1}EP = F = (f_{ij}) \in \mathbb{C}^{n \times n}$, 则 B 的对角元是 $\lambda_k + f_{kk}$. 由盖尔定理知存在 i 使得

$$|\lambda_i + f_{ii} - \mu_j| \leq \sum_{\substack{k=1 \\ k \neq i}}^n |f_{ik}|.$$

因此

$$|\lambda_i - \mu_j| \leq \sum_{k=1}^n |f_{ik}| \leq \|P^{-1}EP\|_{\infty}.$$

再从盖尔定理 4.3.2 的结论(2)可直接推得定理 4.3.3 的第二部分结论. 证毕

定理 4.3.4 设 $A, E \in \mathbb{C}^{n \times n}$, 并存在可逆矩阵 P 使得

$$P^{-1}AP = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n), \quad (4.3.14)$$

则对 $A+E$ 的任一特征值 μ , 有

$$\min_i |\lambda_i - \mu| \leq \|P^{-1}EP\|, \quad (4.3.15)$$

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 $B = P^{-1}(A+E)P = \Lambda + F$, 其中 $F = P^{-1}EP = (f_{ij}) \in \mathbb{C}^{n \times n}$. 对 $A+E$ 的任一特

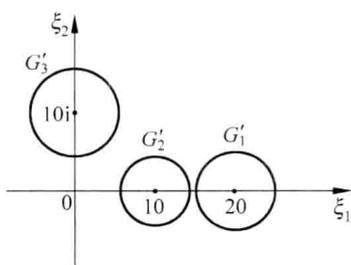


图 4.4

征值 μ , 若 $\mathbf{A} - \mu\mathbf{I}$ 奇异, 则存在某个 i , 使得 $\mu = \lambda_i$. 于是定理得证.

下面假定 $\mathbf{A} - \mu\mathbf{I}$ 非奇异, 因为 $\mathbf{A} - \mu\mathbf{I} + \mathbf{F}$ 奇异, 所以存在非零向量 $\mathbf{y} \in \mathbb{C}^{n \times n}$, 使得

$$(\mathbf{A} - \mu\mathbf{I})\mathbf{y} = -\mathbf{F}\mathbf{y},$$

则

$$\mathbf{y} = (\mathbf{A} - \mu\mathbf{I})^{-1}\mathbf{F}\mathbf{y},$$

从而

$$\begin{aligned} \|\mathbf{y}\| &\leq \|(\mathbf{A} - \mu\mathbf{I})^{-1}\| \|\mathbf{F}\| \|\mathbf{y}\| \\ &= (\min_i |\lambda_i - \mu|)^{-1} \|\mathbf{F}\| \|\mathbf{y}\|, \end{aligned}$$

因此

$$\min_i |\lambda_i - \mu| \leq \|\mathbf{F}\| = \|\mathbf{P}^{-1}\mathbf{E}\mathbf{P}\|. \quad \text{证毕}$$

由式(4.3.15)可得

$$\min_i |\lambda_i - \mu| \leq \|\mathbf{P}^{-1}\| \|\mathbf{P}\| \|\mathbf{E}\|, \quad (4.3.16)$$

上式表明: $\|\mathbf{P}^{-1}\| \|\mathbf{P}\|$ 反映了矩阵 \mathbf{A} 的特征值对摄动 \mathbf{E} 的“灵敏度”.

定义 4.3.4 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 并且存在可逆矩阵 \mathbf{P} 使得 $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \text{diag}(\lambda_1, \dots, \lambda_n)$, 则称 $\|\mathbf{P}^{-1}\| \|\mathbf{P}\|$ 为矩阵 \mathbf{A} 关于特征值问题的“条件数”, 简称特征条件数, 记为 $\zeta(\mathbf{P})$.

若 $\zeta(\mathbf{P}) = \|\mathbf{P}^{-1}\| \|\mathbf{P}\|$ 不是很大, 则 \mathbf{A} 的特征值问题是良态的.

如果矩阵 \mathbf{A} 是埃尔米特矩阵, 并且其摄动矩阵 \mathbf{E} 也是埃尔米特矩阵, 则有如下结论:

定理 4.3.5 设 \mathbf{A}, \mathbf{E} 均为 n 阶埃尔米特矩阵, $\mathbf{B} = \mathbf{A} + \mathbf{E}$, 且 \mathbf{A}, \mathbf{B} 和 \mathbf{E} 的特征值分别为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n, \mu_1 \geq \mu_2 \geq \dots \geq \mu_n$ 和 $\varepsilon_1 \geq \varepsilon_2 \geq \dots \geq \varepsilon_n$, 则

$$|\lambda_i - \mu_i| \leq \|\mathbf{E}\|_2, \quad i = 1, 2, \dots, n. \quad (4.3.17)$$

证明 因为 \mathbf{E} 是埃尔米特矩阵, 则

$$\|\mathbf{E}\|_2 = \max\{|\varepsilon_1|, |\varepsilon_n|\}.$$

于是, 由定理 1.3.13 即得式(4.3.17).

证毕

由定理 4.3.5 可知, 埃尔米特矩阵 \mathbf{A} 的特征条件数 $\zeta(\mathbf{P}) = \|\mathbf{P}^{-1}\| \|\mathbf{P}\| = 1$, 这是比较小的. 所以这种矩阵的特征值问题是良态的.

习 题 4 (3)

1. 求矩阵 $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 10^{-10} \end{bmatrix}$ 的条件数 $\text{cond}(\mathbf{A})_\infty$.
2. 在 $p=1, \infty$ 的范数意义下, 求矩阵 $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 1.001 & 2.001 \end{bmatrix}$ 的条件数.
3. 设 $\mathbf{A} = \begin{bmatrix} 2\lambda & \lambda \\ 1 & 1 \end{bmatrix}$, 证明: 当 $\lambda = \pm \frac{2}{3}$ 时, $\text{cond}(\mathbf{A})_\infty$ 有最小值.
4. 设 n 阶矩阵 ($n=101$)

$$\mathbf{A} = \begin{pmatrix} 1 & & & \\ & 10^{-1} & & \\ & & \ddots & \\ & & & 10^{-1} \end{pmatrix},$$

求 $\det(\mathbf{A})$ 和 $\text{cond}(\mathbf{A})_\infty$. 结果说明了什么?

5. 方程组

$$\begin{pmatrix} 6 & 13 & -17 \\ 13 & 29 & -38 \\ -17 & -38 & 50 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$$

是否病态?

6. 4阶希尔伯特矩阵 \mathbf{H}_4 的元素为 $h_{ij} = (i+j-1)^{-1}$, 求 \mathbf{H}_4 在 1-范数和 ∞ -范数意义下的条件数及 $\det(\mathbf{H}_4)$.

7. 已知威尔逊(Wilson)矩阵

$$\mathbf{W} = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix},$$

求解线性方程组 $\mathbf{W}\mathbf{x} = \mathbf{b}$. 其中

(1) $\mathbf{b} = (32, 23, 33, 31)^\top$;

(2) $\mathbf{b} = (32.01, 22.99, 33.01, 30.99)^\top$.

并讨论右端项 \mathbf{b} 的摄动对解 \mathbf{x} 的影响.

8. (1) 求线性方程组 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 的准确解, 这里

$$\mathbf{A} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 0.3 \\ 1 \\ 1 \end{pmatrix};$$

(2) 当(1)中的 $\mathbf{b} = \begin{pmatrix} 0.3 \\ 1 \\ 1.01 \end{pmatrix}$ 时, 再解 $\mathbf{A}\mathbf{x} = \mathbf{b}$.

9. 已知线性方程组

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

的真解为 $(9, -36, 30)^\top$, 若方程组变为

$$\begin{pmatrix} 1 & 0.50 & 0.33 \\ 0.50 & 0.33 & 0.25 \\ 0.33 & 0.25 & 0.20 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1.0 \\ 0.0 \\ 0.0 \end{pmatrix},$$

用高斯消去法求解(取二位有效数字运算), 并把两者结果进行比较, 能得出什么结论?

10. 设 $\mathbf{A} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 25 \\ 12 \\ 77 \\ 60 \\ 57 \\ 60 \\ 319 \\ 420 \end{pmatrix}$, 方程组 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 的真解为 $\mathbf{x} = (1, 1, 1, 1)^\top$.

- (1) 用三位有效数字(切断)运算,求 $Ax=b$ 的解.
 (2) 用三位有效数字(舍入)运算,求 $Ax=b$ 的解; 再与(1)比较.
 (3) 用五位有效数字(舍入)运算,求解 $Ax=b$,再把结果与(1)、(2)比较.
11. 利用盖尔定理估计下面矩阵 A 的特征值的分布范围:

$$A = \begin{bmatrix} 1 & -0.5 & -0.5 & 0 \\ -0.5 & 1.5 & i & 0 \\ 0 & -0.5i & 5 & 0.5i \\ -1 & 0 & 0 & 5i \end{bmatrix}.$$

12. 用相似方法,求 A 的孤立圆盘,其中

$$A = \begin{bmatrix} 0.9 & 0.01 & 0.12 \\ 0.01 & 0.8 & 0.13 \\ 0.01 & 0.02 & 0.4 \end{bmatrix}.$$

13. 证明 A 至少有两个实特征值,其中

$$A = \begin{bmatrix} 9 & 1 & -2 & 1 \\ 0 & 8 & 1 & 1 \\ -1 & 0 & 4 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

14. 在盖尔定理中,如果一个连通部分是由两个外切圆构成的,证明: 每个圆上不可能都有两个特征值.

15. 求 A 的圆盘,其中

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

16. 求 A 的圆盘并估计其特征值,其中

$$A = \begin{bmatrix} 10 & 1 & 2 & 3 \\ 5 & 30 & 1 & 2 \\ 10 & 3 & -10 & 5 \\ 2 & 3 & 1 & -10 \end{bmatrix}.$$

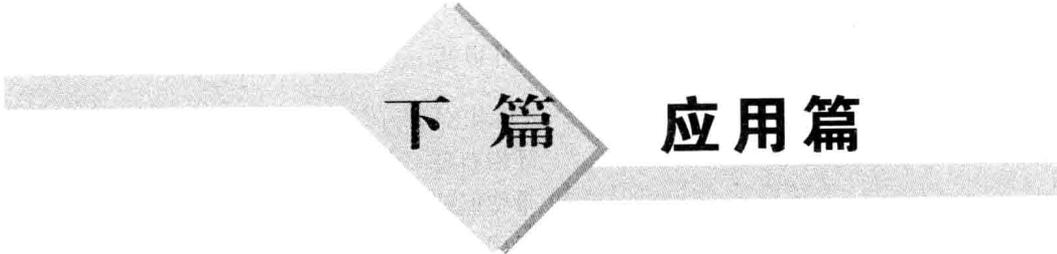
17. 设 n 阶矩阵 A 满足对角强优条件(亦称“严格对角占优”条件)

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|, \quad i = 1, 2, \dots, n,$$

证明: $|A| \neq 0$.

18. 求 A 的圆盘,其中

$$A = \begin{bmatrix} 10 & -8 \\ 5 & 0 \end{bmatrix}.$$



下篇

应用篇

第 5 章 矩阵微积分及其应用

第 6 章 广义逆矩阵及其应用

第 7 章 几类特殊矩阵与特殊积

第 8 章 矩阵在数学内外的应用

附录 模拟考试自测试题(共 15 套)

第 5 章

矩阵微积分及其应用

引言 讨论矩阵微积分的必要性

在线性代数中,只讨论矩阵的加(减)法、乘法和求逆为核心的代数运算,而完全没有涉及类似于数学分析中的极限、级数、微积分等运算,可是,在研究运筹学以及线性系统的可控制等方面的问题时,这些运算又是十分必要的.比如在图像处理、模式识别或移动通信等领域,常需要利用特定的线性变换将高维向量压缩成低维向量或将低维向量还原为高维向量,并且使误差尽可能小,描述此类问题的数学模型是:相当于求以矩阵 U 为自变量的函数

$$J(U) = \|U\alpha - \beta\|, \quad \text{其中 } U \in \mathbb{R}^{m \times n}, \alpha \in \mathbb{R}^n, \beta \in \mathbb{R}^m$$

在约束条件 $U^T U = I$ 或 $U U^T = I$ 下的最小值点(矩阵).解决此类优化问题的一个可行办法是求矩阵函数 $J(U)$ 关于未知矩阵 U 的导数,这就需要研究矩阵的微积分.在第 4 章建立了范数概念之后,就解决了两个矩阵之间的距离问题,一旦有了距离概念,就能够和高等数学几乎完全平行地讨论矩阵序列的极限和矩阵函数的连续性,进而讨论矩阵函数的微分学和积分学理论.以上这些,就是本章所要介绍的矩阵微积分内容.

5.1 向量序列和矩阵序列的极限

矩阵微积分理论的建立,同大学里所学的数学分析课程一样,也是以极限理论为基础的.因此,下面我们先讨论向量和矩阵序列的极限运算.

5.1.1 向量序列的极限

定义 5.1.1(按范数收敛) 设 $x^{(k)}, x \in \mathbb{C}^n (k=1, 2, \dots)$, 若

$$\|x^{(k)} - x\| \rightarrow 0, \quad k \rightarrow +\infty, \quad (5.1.1)$$

则称向量序列 $\{x^{(k)}\}$ 收敛于向量 x , 或说向量 x 是向量序列 $\{x^{(k)}\}$ 当 $k \rightarrow +\infty$ 时的极限, 可记为

$$\lim_{k \rightarrow +\infty} x^{(k)} = x \quad (5.1.2)$$

或

$$x^{(k)} \rightarrow x, \quad k \rightarrow +\infty. \quad (5.1.3)$$

由向量范数之间的等价关系可知,在某一向量范数意义下收敛,在其他向量范数意义下也一定收敛.

例 5.1.1 设 $\mathbf{x} = (1, 1, \dots, 1)^T \in \mathbb{R}^n$, 且

$$\mathbf{x}^{(k)} = \left(1 + \frac{1}{2^k}, 1 + \frac{1}{3^k}, \dots, 1 + \frac{1}{(n+1)^k} \right)^T,$$

试证 $\lim_{k \rightarrow +\infty} \mathbf{x}^{(k)} = \mathbf{x}$.

证明 由于 $\lim_{k \rightarrow +\infty} \|\mathbf{x}^{(k)} - \mathbf{x}\|_{\infty} = \lim_{k \rightarrow +\infty} \max_{2 \leq i \leq n+1} \frac{1}{i^k} = \lim_{k \rightarrow +\infty} \frac{1}{2^k} = 0$, 所以 $\lim_{k \rightarrow +\infty} \mathbf{x}^{(k)} = \mathbf{x}$.

证毕

例 5.1.2 设

$$\mathbf{x}^{(k)} = \sum_{m=0}^k \frac{t^m}{m!}, \quad t \in [0, 1], \quad k = 0, 1, 2, \dots$$

是线性空间 $P[t]_k$ (多项式集合) 的向量序列, 又

$$\|\mathbf{x}^{(k)}\|_{\infty} = \max_{t \in [0, 1]} |\mathbf{x}^{(k)}|.$$

显然, 当 $n > k$ 时

$$\|\mathbf{x}^{(k)} - \mathbf{x}^{(n)}\|_{\infty} \leq \frac{1}{(k+1)!} + \frac{1}{(k+2)!} + \dots + \frac{1}{n!},$$

只要 k, n 充分大, 上面不等式右端可以任意小, 由微积分理论知

$$\lim_{k \rightarrow +\infty} \mathbf{x}^{(k)} = e^t.$$

虽然向量序列 $\{\mathbf{x}^{(k)}\}$ 的极限存在, 但 e^t 并不是 $P[t]_k$ 中的向量, 由此, 在赋范线性空间中, 柯西收敛原理只是向量序列收敛的必要条件, 即若 $\{\mathbf{x}^{(k)}\}$ 收敛, 则对任意 $\varepsilon > 0$, 存在正整数 N , 使当 $k, n > N$ 时, 恒有

$$\|\mathbf{x}^{(k)} - \mathbf{x}^{(n)}\| < \varepsilon$$

成立, 但其逆不真 (因为极限可能不属于赋范线性空间).

定义 5.1.2 若赋范线性空间中任一收敛向量序列的极限均属于此赋范线性空间, 则称此赋范线性空间为完备的赋范线性空间, 或称为巴拿赫 (Banach) 空间. 在巴拿赫空间中, 柯西收敛原理成立.

如果赋范线性空间是 \mathbb{C}^n , 则向量序列在范数意义下的收敛定义 5.1.1 与下面的各坐标分量序列的同时收敛等价.

定理 5.1.1 设 $\mathbf{x}^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})^T$, $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{C}^n$, 则向量序列 $\{\mathbf{x}^{(k)}\}$ 收敛于 \mathbf{x} 的充要条件为: 每一个坐标分量序列 $\{x_i^{(k)}\}$ 收敛于 x_i , 即

$$\lim_{k \rightarrow +\infty} x_i^{(k)} = x_i, \quad i = 1, 2, \dots, n. \quad (\text{按坐标收敛}) \quad (5.1.4)$$

证明 必要性: 设 $\|\mathbf{x}\| = \|\mathbf{x}\|_{\infty} = \max_i |x_i|$, 则

$$\|\mathbf{x}^{(k)} - \mathbf{x}\|_{\infty} \rightarrow 0, \quad k \rightarrow +\infty.$$

由于 $\max_i |x_i^{(k)} - x_i| \geq |x_i^{(k)} - x_i|$, 所以有

$$|x_i^{(k)} - x_i| \rightarrow 0,$$

即

$$\lim_{k \rightarrow +\infty} x_i^{(k)} = x_i, \quad i = 1, 2, \dots, n.$$

充分性: 如果

$$\lim_{k \rightarrow +\infty} x_i^{(k)} = x_i, \quad i = 1, 2, \dots, n,$$

则

$$\lim_{k \rightarrow +\infty} \max_i |x_i^{(k)} - x_i| = 0,$$

即

$$\|\mathbf{x}^{(k)} - \mathbf{x}\|_{\infty} \rightarrow 0, \quad k \rightarrow +\infty.$$

再由范数的等价性知, 上述结论对 \mathbb{C}^n 中任意向量范数均成立.

证毕

例 5.1.3 试考察下列两向量序列的收敛性.

$$(1) \mathbf{x}^{(k)} = \left(\frac{1}{2^k}, \frac{\sin k}{k} \right) (k = 1, 2, 3, \dots).$$

$$(2) \mathbf{y}^{(k)} = \left(\sum_{i=1}^k \frac{1}{2^i}, \sum_{i=1}^k \frac{1}{i} \right) (k = 1, 2, 3, \dots).$$

解 (1) 因为当 $k \rightarrow +\infty$ 时, $\frac{1}{2^k} \rightarrow 0$, $\frac{\sin k}{k} \rightarrow 0$, 所以

$$\lim_{k \rightarrow +\infty} \mathbf{x}^{(k)} = \left(\lim_{k \rightarrow +\infty} \frac{1}{2^k}, \lim_{k \rightarrow +\infty} \frac{\sin k}{k} \right) = (0, 0) = \mathbf{0},$$

即收敛.

(2) 因为

$$\sum_{i=1}^k \frac{1}{2^i} = \frac{1}{2} \frac{1 - \left(\frac{1}{2}\right)^{k+1}}{1 - \frac{1}{2}} \rightarrow 1, \quad k \rightarrow +\infty,$$

而 $\sum_{i=1}^k \frac{1}{i}$ 当 $k \rightarrow +\infty$ 时为调和级数, 发散, 因而 $\{\mathbf{y}^{(k)}\}$ 发散.

从定理 5.1.1 和例 5.1.3 可见, 赋范线性空间 \mathbb{C}^n 中的向量序列 $\{\mathbf{x}^{(k)}\}$ 若收敛, 则极限 \mathbf{x} 仍属于 \mathbb{C}^n , 所以 \mathbb{C}^n 总是完备的赋范线性空间.

5.1.2 矩阵序列的极限

由于 n 阶矩阵可以看作是一个 $n \times n$ 维向量, 其收敛性可以和 \mathbb{C}^n 中的向量一样考虑. 所以, 我们可以按照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 5.1.3 设有矩阵序列 $\{\mathbf{A}^{(k)}\}$, 其中 $\mathbf{A}^{(k)} = (a_{ij}^{(k)}) \in \mathbb{C}^{n \times n}$, 且当 $k \rightarrow +\infty$ 时, $a_{ij}^{(k)} \rightarrow a_{ij}$, 则称 $\{\mathbf{A}^{(k)}\}$ 收敛, 并把矩阵 $\mathbf{A} = (a_{ij})$ 叫做 $\{\mathbf{A}^{(k)}\}$ 的极限, 或称 $\{\mathbf{A}^{(k)}\}$ 收敛于 \mathbf{A} , 记为

$$\lim_{k \rightarrow +\infty} \mathbf{A}^{(k)} = \mathbf{A} \quad \text{或} \quad \mathbf{A}^{(k)} \rightarrow \mathbf{A}. \quad (5.1.5)$$

不收敛的矩阵序列称为发散的.

下面我们来证明, 定义 5.1.3 和在范数意义下的收敛

$$\|\mathbf{A}^{(k)} - \mathbf{A}\| \rightarrow 0, \quad k \rightarrow +\infty \quad (5.1.6)$$

(其中 $\|\cdot\|$ 为任一矩阵范数) 是等价的.

事实上, 取矩阵范数 (见第 4 章 m_1 范数)

$$\|\mathbf{A}\| = \sum_{i,j=1}^n |a_{ij}|.$$

设 $\lim_{k \rightarrow +\infty} \mathbf{A}^{(k)} = \mathbf{A}$, 即对每个 i, j 有

$$\lim_{k \rightarrow +\infty} a_{ij}^{(k)} = a_{ij}, \quad i, j = 1, 2, \dots, n,$$

即

$$\lim_{k \rightarrow +\infty} |a_{ij}^{(k)} - a_{ij}| = 0, \quad i, j = 1, 2, \dots, n,$$

故有

$$\lim_{k \rightarrow +\infty} \|\mathbf{A}^{(k)} - \mathbf{A}\| = \lim_{k \rightarrow +\infty} \sum_{i,j=1}^n |a_{ij}^{(k)} - a_{ij}| = 0.$$

反之, 设 $\lim_{k \rightarrow +\infty} \|\mathbf{A}^{(k)} - \mathbf{A}\| = 0$, 即

$$\lim_{k \rightarrow +\infty} \sum_{i,j=1}^n |a_{ij}^{(k)} - a_{ij}| = 0.$$

于是对 $i, j = 1, 2, \dots, n$, 有

$$\lim_{k \rightarrow +\infty} |a_{ij}^{(k)} - a_{ij}| = 0.$$

即

$$\lim_{k \rightarrow +\infty} \mathbf{A}^{(k)} = \mathbf{A}.$$

再由矩阵范数的等价性, 有

$$c_1 \|\mathbf{A}^{(k)} - \mathbf{A}\| \leq \|\mathbf{A}^{(k)} - \mathbf{A}\|_a \leq c_2 \|\mathbf{A}^{(k)} - \mathbf{A}\|,$$

$$c'_1 \|\mathbf{A}^{(k)} - \mathbf{A}\|_a \leq \|\mathbf{A}^{(k)} - \mathbf{A}\| \leq c'_2 \|\mathbf{A}^{(k)} - \mathbf{A}\|_a,$$

其中 $\|\cdot\|_a$ 为任一矩阵范数. 从而知

$$\lim_{k \rightarrow +\infty} \|\mathbf{A}^{(k)} - \mathbf{A}\| = 0 \Leftrightarrow \lim_{k \rightarrow +\infty} \|\mathbf{A}^{(k)} - \mathbf{A}\|_a = 0.$$

关于矩阵序列的极限运算有下述性质.

性质 1 设 $\lim_{k \rightarrow +\infty} \mathbf{A}^{(k)} = \mathbf{A}$, $\lim_{k \rightarrow +\infty} \mathbf{B}^{(k)} = \mathbf{B}$, 则

$$\lim_{k \rightarrow +\infty} (a\mathbf{A}^{(k)} + b\mathbf{B}^{(k)}) = a\mathbf{A} + b\mathbf{B}, \quad a, b \in \mathbb{C};$$

性质 2 设 $\lim_{k \rightarrow +\infty} \mathbf{A}^{(k)} = \mathbf{A}$, $\lim_{k \rightarrow +\infty} \mathbf{B}^{(k)} = \mathbf{B}$, $\mathbf{A}^{(k)}, \mathbf{B}^{(k)} \in \mathbb{C}^{n \times n}$, 则

$$\lim_{k \rightarrow +\infty} \mathbf{A}^{(k)} \mathbf{B}^{(k)} = \mathbf{A} \mathbf{B};$$

性质 3 设 $\lim_{k \rightarrow +\infty} \mathbf{A}^{(k)} = \mathbf{A}$, 且 $\mathbf{A}^{(k)} (k=1, 2, \dots), \mathbf{A}$ 均可逆, 则 $\{(\mathbf{A}^{(k)})^{-1}\}$ 也收敛, 且

$$\lim_{k \rightarrow +\infty} (\mathbf{A}^{(k)})^{-1} = \mathbf{A}^{-1}.$$

性质 1 和性质 2 利用定义容易验证. 下面证明性质 3.

事实上, 记 $\text{adj}\mathbf{A}$ 为 \mathbf{A} 的伴随矩阵, 则

$$(\mathbf{A}^{(k)})^{-1} = \frac{\text{adj}\mathbf{A}^{(k)}}{\det\mathbf{A}^{(k)}},$$

其中

$$\text{adj}\mathbf{A}^{(k)} = \begin{pmatrix} A_{11}^{(k)} & A_{21}^{(k)} & \cdots & A_{n1}^{(k)} \\ A_{12}^{(k)} & A_{22}^{(k)} & \cdots & A_{n2}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n}^{(k)} & A_{2n}^{(k)} & \cdots & A_{nn}^{(k)} \end{pmatrix}.$$

因为 $A_{ij}^{(k)}$ 是 $|\mathbf{A}^{(k)}|$ 中元素 $a_{ij}^{(k)}$ 的代数余子式, 所以 $A_{ij}^{(k)}$ 是 $\mathbf{A}^{(k)}$ 的元素的 $(n-1)$ 次多项式, 由多项式函数的连续性知

$$\lim_{k \rightarrow +\infty} A_{ij}^{(k)} = A_{ij},$$

其中 A_{ij} 是 $|\mathbf{A}|$ 中元素 a_{ij} 的代数余子式, 故有

$$\lim_{k \rightarrow +\infty} \text{adj} \mathbf{A}^{(k)} = \text{adj} \mathbf{A}.$$

同理, $\det \mathbf{A}^{(k)}$ 是 $\mathbf{A}^{(k)}$ 的元素的 n 次多项式, 所以

$$\lim_{k \rightarrow +\infty} \det \mathbf{A}^{(k)} = \det \mathbf{A} \neq 0,$$

从而有

$$\lim_{k \rightarrow +\infty} (\mathbf{A}^{(k)})^{-1} = \lim_{k \rightarrow +\infty} \frac{\text{adj} \mathbf{A}^{(k)}}{\det \mathbf{A}^{(k)}} = \frac{\text{adj} \mathbf{A}}{\det \mathbf{A}} = \mathbf{A}^{-1}.$$

下面研究由矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 的幂组成的矩阵序列

$$\mathbf{A}, \mathbf{A}^2, \dots, \mathbf{A}^k, \dots \quad (5.1.7)$$

关于矩阵序列 (5.1.7) 有下述两个收敛定理:

定理 5.1.2 设有矩阵序列 $\{\mathbf{A}^{(k)}\}$: $\mathbf{A}, \mathbf{A}^2, \dots, \mathbf{A}^k, \dots$, 则 $\lim_{k \rightarrow +\infty} \mathbf{A}^k = \mathbf{0}$ 的充分必要条件是矩阵 \mathbf{A} 的所有特征值的模都小于 1, 即 \mathbf{A} 的谱半径小于 1.

$$\rho(\mathbf{A}) < 1. \quad (5.1.8)$$

证明 必要性: 设 $\lim_{k \rightarrow +\infty} \mathbf{A}^k = \mathbf{0}$, 由于 $\mathbf{A}^k = \mathbf{T} \mathbf{J}^k \mathbf{T}^{-1}$, 故有

$$\lim_{k \rightarrow +\infty} \mathbf{J}^k = \mathbf{0}. \quad (5.1.9)$$

显然, 式 (5.1.9) 成立的充要条件为

$$\lim_{k \rightarrow +\infty} \mathbf{J}_i^k = \mathbf{0}, \quad i = 1, 2, \dots, r, \quad (5.1.10)$$

其中

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_1 & & & \\ & \mathbf{J}_2 & & \\ & & \ddots & \\ & & & \mathbf{J}_r \end{bmatrix}_{n \times n}, \quad \mathbf{J}_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & 1 & \\ & & \lambda_i & \ddots \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix}_{n_i \times n_i},$$

\mathbf{J}_i 为若尔当块, $\lambda_i (i=1, 2, \dots, r)$ 可能有相同的, 初等因子的指数 $n_i (i=1, 2, \dots, r)$ 也可能有相同的, 且 $\sum_{i=1}^r n_i = n$.

又因为 k 充分大 ($k \geq n_i - 1$) 时,

$$\mathbf{J}_i^k = \begin{bmatrix} \lambda_i^k & C_k^1 \lambda_i^{k-1} & \dots & C_k^{n_i-1} \lambda_i^{k-n_i+1} \\ & \lambda_i^k & \dots & \vdots \\ & & \ddots & \vdots \\ & & & \lambda_i^k \end{bmatrix}_{n_i \times n_i},$$

所以, 式 (5.1.10) 成立的充要条件为

$$\lim_{k \rightarrow +\infty} \lambda_i^k = 0, \quad i = 1, 2, \dots, r,$$

故 $|\lambda_i| < 1 (i=1, 2, \dots, r)$, 从而 $\rho(\mathbf{A}) < 1$.

充分性: 设 \mathbf{A} 的谱半径 $\rho(\mathbf{A}) < 1$, 则 $\lim_{k \rightarrow +\infty} \lambda_i^k = 0 (i=1, 2, \dots, r)$, 由此可知 $\lim_{k \rightarrow +\infty} \mathbf{J}_i^k = \mathbf{0}$, 从而有

$$\begin{aligned} \lim_{k \rightarrow +\infty} \mathbf{J}^k &= \mathbf{0}, \\ \lim_{k \rightarrow +\infty} \mathbf{A}^k &= \mathbf{0}. \end{aligned} \quad \text{证毕}$$

注 当 \mathbf{A}^k 收敛于零矩阵时, 此时称 \mathbf{A} 为收敛矩阵.

定理 5.1.3 若对于矩阵 \mathbf{A} 的某一范数有 $\|\mathbf{A}\| < 1$, 则

$$\lim_{k \rightarrow +\infty} \mathbf{A}^k = \mathbf{0}.$$

证明 因为 $\|\mathbf{A}^k\| \leq \|\mathbf{A}\|^k$, 所以当 $\|\mathbf{A}\| < 1$ 时, 有

$$\lim_{k \rightarrow +\infty} \|\mathbf{A}^k\| = 0,$$

即

$$\lim_{k \rightarrow +\infty} \|\mathbf{A}^k - \mathbf{0}\| = 0.$$

再由式(5.1.6)知

$$\lim_{k \rightarrow +\infty} \mathbf{A}^k = \mathbf{0}. \quad \text{证毕}$$

5.2 矩阵级数与矩阵函数

5.2.1 矩阵级数

在建立矩阵函数以及表示系统微分方程的解时, 常常用到矩阵级数. 特别是矩阵的幂级数在矩阵分析中占有重要的地位. 矩阵级数理论与数学分析中数项级数的相应的定义与性质完全类似, 现给出如下.

定义 5.2.1 设有矩阵序列

$$\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(k)}, \dots,$$

其中 $\mathbf{A}^{(k)} = (a_{ij}^{(k)}) \in \mathbb{C}^{n \times n}$, 称无穷和

$$\mathbf{A}^{(0)} + \mathbf{A}^{(1)} + \mathbf{A}^{(2)} + \dots + \mathbf{A}^{(k)} + \dots$$

为矩阵级数, 记为 $\sum_{k=0}^{+\infty} \mathbf{A}^{(k)}$, $\mathbf{A}^{(k)}$ 称为矩阵级数的一般项, 即有

$$\sum_{k=0}^{+\infty} \mathbf{A}^{(k)} = \mathbf{A}^{(0)} + \mathbf{A}^{(1)} + \mathbf{A}^{(2)} + \dots + \mathbf{A}^{(k)} + \dots. \quad (5.2.1)$$

定义 5.2.2 级数(5.2.1)前 $k+1$ 项的和

$$\mathbf{S}^{(k)} = \mathbf{A}^{(0)} + \mathbf{A}^{(1)} + \mathbf{A}^{(2)} + \dots + \mathbf{A}^{(k)}$$

称为级数(5.2.1)的部分和, 如果矩阵序列 $\{\mathbf{S}^{(k)}\}$ 收敛, 且有极限 \mathbf{S} , 即有

$$\lim_{k \rightarrow +\infty} \mathbf{S}^{(k)} = \mathbf{S},$$

则称矩阵数(5.2.1)收敛, \mathbf{S} 称为级数的和, 记作

$$\mathbf{S} = \sum_{k=0}^{+\infty} \mathbf{A}^{(k)}. \quad (5.2.2)$$

不收敛的矩阵级数称为是发散的.

显然,矩阵级数(5.2.2)收敛的充要条件是 n^2 个数项级数 $\sum_{k=0}^{+\infty} a_{ij}^{(k)}$ ($i, j = 1, 2, \dots, n$) 都收敛,且和为

$$S = \begin{pmatrix} \sum_{k=0}^{+\infty} a_{11}^{(k)} & \sum_{k=0}^{+\infty} a_{12}^{(k)} & \cdots & \sum_{k=0}^{+\infty} a_{1n}^{(k)} \\ \sum_{k=0}^{+\infty} a_{21}^{(k)} & \sum_{k=0}^{+\infty} a_{22}^{(k)} & \cdots & \sum_{k=0}^{+\infty} a_{2n}^{(k)} \\ \vdots & \vdots & & \vdots \\ \sum_{k=0}^{+\infty} a_{n1}^{(k)} & \sum_{k=0}^{+\infty} a_{n2}^{(k)} & \cdots & \sum_{k=0}^{+\infty} a_{nn}^{(k)} \end{pmatrix}.$$

由矩阵级数收敛的定义易知

(1) 若 $\sum_{k=0}^{+\infty} A^{(k)}$ 收敛,则 $\lim_{k \rightarrow +\infty} A^{(k)} = \mathbf{0}$;

(2) 若 $\sum_{k=0}^{+\infty} A^{(k)} = S, \sum_{k=0}^{+\infty} B^{(k)} = S'$,则

$$\sum_{k=0}^{+\infty} (A^{(k)} \pm B^{(k)}) = S \pm S';$$

(3) 若 $\sum_{k=0}^{+\infty} A^{(k)} = S$,则 $\sum_{k=0}^{+\infty} \mu A^{(k)} = \mu S, \mu \in \mathbb{C}$.

定义 5.2.3 设矩阵级数 $\sum_{k=0}^{+\infty} A^{(k)} = A^{(0)} + A^{(1)} + A^{(2)} + \cdots + A^{(k)} + \cdots$, 其中 $A^{(k)} = (a_{ij}^{(k)}) \in \mathbb{C}^{n \times n}$. 如果 n^2 个数项级数

$$a_{ij}^{(0)} + a_{ij}^{(1)} + a_{ij}^{(2)} + \cdots + a_{ij}^{(k)} + \cdots, \quad i, j = 1, 2, \dots, n \quad (5.2.3)$$

都绝对收敛,则称矩阵级数绝对收敛.

定理 5.2.1 矩阵级数 $\sum_{k=0}^{+\infty} A^{(k)}$ 绝对收敛的充分必要条件是 $\sum_{k=0}^{+\infty} \|A^{(k)}\| = \|A^0\| + \|A^{(1)}\| + \|A^{(2)}\| + \cdots + \|A^{(k)}\| + \cdots$ 收敛,其中 $\|A^{(k)}\|$ 为 $A^{(k)}$ 的任何一种范数.

证明 若 $\sum_{k=0}^{+\infty} A^{(k)}$ 绝对收敛,于是有一正数 M 存在,它与 i, j 无关,使得

$$\sum_{k=0}^{+\infty} |a_{ij}^{(k)}| < M, \quad k \geq 0, \quad i, j = 1, 2, \dots, n,$$

从而有

$$\sum_{k=0}^{+\infty} \|A^{(k)}\|_{m_1} = \sum_{k=0}^{+\infty} \left(\sum_{i=1}^n \sum_{j=1}^n |a_{ij}^{(k)}| \right) < n^2 M,$$

故 $\sum_{k=0}^{+\infty} \|A^{(k)}\|_{m_1}$ 为收敛级数,由矩阵范数的等价性, $\sum_{k=0}^{+\infty} \|A^{(k)}\|$ 也为收敛级数.

反之,如果 $\sum_{k=0}^{+\infty} \|A^{(k)}\|$ 收敛,则 $\sum_{k=0}^{+\infty} \|A^{(k)}\|_{m_1}$ 也收敛,由 $|a_{ij}^{(k)}| \leq \|A^{(k)}\|_{m_1}$ ($i, j = 1, 2, \dots, n$) 可知,如式(5.2.3)的 n^2 个数项级数中,每一个级数都是绝对收敛的,故根据定义 5.2.3 知级数(5.2.2)绝对收敛. 证毕

例 5.2.1 设 $A \in \mathbb{C}^{n \times n}$, 则矩阵幂级数

$$I + A + \frac{A^2}{2!} + \cdots + \frac{A^{(k)}}{k!} + \cdots \quad (5.2.4)$$

绝对收敛.

证明 因为 $\left\| \frac{A^k}{k!} \right\| \leq \frac{\|A\|^k}{k!}$, 且

$$\begin{aligned} & \|I\| + \|A\| + \frac{\|A\|^2}{2!} + \cdots + \frac{\|A\|^k}{k!} + \cdots \\ &= e^{\|A\|} - 1 + \|I\|, \end{aligned}$$

故级数(5.2.4)绝对收敛, 其和称为 A 的矩阵指数函数, 记作 e^A , 即

$$e^A = I + A + \frac{A^2}{2!} + \cdots + \frac{A^k}{k!} + \cdots \quad (5.2.5)$$

同理, 有

$$\sin A = A - \frac{A^3}{3!} + \cdots + (-1)^{k-1} \frac{A^{2k-1}}{(2k-1)!} + \cdots \quad (5.2.6)$$

$$\cos A = I - \frac{A^2}{2!} + \cdots + (-1)^k \frac{A^{2k}}{(2k)!} + \cdots \quad (5.2.7)$$

$e^A, \sin A, \cos A$ 是最简单的矩阵函数, 后面将作进一步地讨论.

定理 5.2.2 设两个矩阵级数

$$A^{(1)} + A^{(2)} + \cdots + A^{(k)} + \cdots, \quad A^{(k)} \in \mathbb{C}^{n \times n}, \quad (5.2.8)$$

$$B^{(1)} + B^{(2)} + \cdots + B^{(k)} + \cdots, \quad B^{(k)} \in \mathbb{C}^{n \times n} \quad (5.2.9)$$

都绝对收敛, 其和分别为 A, B , 则将它们按项相乘后作成的矩阵级数

$$\begin{aligned} & A^{(1)}B^{(1)} + (A^{(1)}B^{(2)} + A^{(2)}B^{(1)}) + \cdots + \\ & (A^{(1)}B^{(k)} + A^{(2)}B^{(k-1)} + \cdots + A^{(k)}B^{(1)}) + \cdots \end{aligned} \quad (5.2.10)$$

绝对收敛, 且具有和 AB .

证明 因为级数(5.2.8)与级数(5.2.9)绝对收敛, 所以级数 $\sum_{k=1}^{+\infty} \|A^{(k)}\|$ 与 $\sum_{k=1}^{+\infty} \|B^{(k)}\|$ 收敛. 这两个级数的乘积所成的级数

$$\begin{aligned} & (\|A^{(1)}\| \|B^{(1)}\| + (\|A^{(1)}\| \|B^{(2)}\| + \|A^{(2)}\| \|B^{(1)}\|)) + \cdots + \\ & (\|A^{(1)}\| \|B^{(k)}\| + \|A^{(2)}\| \|B^{(k-1)}\| + \cdots + \|A^{(k)}\| \|B^{(1)}\|) + \cdots \end{aligned} \quad (5.2.11)$$

也是收敛的, 而与矩阵级数(5.2.10)相应的范数级数为

$$\begin{aligned} & \|A^{(1)}B^{(1)}\| + \|A^{(1)}B^{(2)} + A^{(2)}B^{(1)}\| + \cdots + \\ & \|A^{(1)}B^{(k)} + A^{(2)}B^{(k-1)} + \cdots + A^{(k)}B^{(1)}\| + \cdots. \end{aligned} \quad (5.2.12)$$

由矩阵范数的性质知, 级数(5.2.12)的每一项不大于级数(5.2.11)的对应项, 从而矩阵级数(5.2.10)绝对收敛. 下面求式(5.2.10)的和.

设 $S_1^{(N)}, S_2^{(N)}$ 与 $S_3^{(N)}$ 分别表示矩阵级数(5.2.8), (5.2.9)及(5.2.10)的部分和, 以 $\tilde{S}_1^{(N)}$,

$\tilde{S}_2^{(N)}$ 与 $\tilde{S}_3^{(N)}$ 分别表示级数 $\sum_{k=1}^{+\infty} \|A^{(k)}\|$, $\sum_{k=1}^{+\infty} \|B^{(k)}\|$ 及级数(5.2.11)的部分和, 则

$$\begin{aligned} S_1^{(N)} S_2^{(N)} &= (A^{(1)}B^{(1)} + A^{(1)}B^{(2)} + \cdots + A^{(1)}B^{(N)}) + \\ & (A^{(2)}B^{(1)} + A^{(2)}B^{(2)} + \cdots + A^{(2)}B^{(N)}) + \cdots + \\ & (A^{(N)}B^{(1)} + A^{(N)}B^{(2)} + \cdots + A^{(N)}B^{(N)}), \end{aligned}$$

$$\begin{aligned} \mathbf{S}_3^{(N)} &= \mathbf{A}^{(1)}\mathbf{B}^{(1)} + (\mathbf{A}^{(1)}\mathbf{B}^{(2)} + \mathbf{A}^{(2)}\mathbf{B}^{(1)}) + \cdots + \\ &\quad (\mathbf{A}^{(1)}\mathbf{B}^{(N)} + \mathbf{A}^{(2)}\mathbf{B}^{(N-1)} + \cdots + \mathbf{A}^{(N)}\mathbf{B}^{(1)}) \\ &= (\mathbf{A}^{(1)}\mathbf{B}^{(1)} + \mathbf{A}^{(1)}\mathbf{B}^{(2)} + \cdots + \mathbf{A}^{(1)}\mathbf{B}^{(N)}) + \\ &\quad (\mathbf{A}^{(2)}\mathbf{B}^{(1)} + \mathbf{A}^{(2)}\mathbf{B}^{(2)} + \cdots + \mathbf{A}^{(2)}\mathbf{B}^{(N-1)}) + \cdots + \mathbf{A}^{(N)}\mathbf{B}^{(1)}. \end{aligned}$$

在上面两个等式中以 $\|\mathbf{A}^{(i)}\|$ 替换 $\mathbf{A}^{(i)}$, 以 $\|\mathbf{B}^{(i)}\|$ 替换 $\mathbf{B}^{(i)}$, 便得到相应的 $\tilde{\mathbf{S}}_1^{(N)}, \tilde{\mathbf{S}}_2^{(N)}$ 与 $\tilde{\mathbf{S}}_3^{(N)}$ 的表达式. 不难知道

$$\|\mathbf{S}_1^{(N)}\mathbf{S}_2^{(N)} - \mathbf{S}_3^{(N)}\| \leq \tilde{\mathbf{S}}_1^{(N)}\tilde{\mathbf{S}}_2^{(N)} - \tilde{\mathbf{S}}_3^{(N)}. \quad (5.2.13)$$

在式(5.2.13)中, 让 $N \rightarrow +\infty$, 则 $\tilde{\mathbf{S}}_1^{(N)}\tilde{\mathbf{S}}_2^{(N)} - \tilde{\mathbf{S}}_3^{(N)} \rightarrow 0$, 故得

$$\begin{aligned} \lim_{N \rightarrow +\infty} \mathbf{S}_1^{(N)}\mathbf{S}_2^{(N)} &= \lim_{N \rightarrow +\infty} \mathbf{S}_1^{(N)} \cdot \lim_{N \rightarrow +\infty} \mathbf{S}_2^{(N)} = \mathbf{A}\mathbf{B} \\ &= \lim_{N \rightarrow +\infty} \mathbf{S}_3^{(N)}. \end{aligned} \quad \text{证毕}$$

从绝对收敛的定义及数学分析中相应的结果, 我们立刻得到下面关于判别矩阵级数收敛性的一些性质.

性质 1 设矩阵级数 $\sum_{k=0}^{+\infty} \mathbf{A}^{(k)}$ 绝对收敛, 则

(1) 级数 $\sum_{k=0}^{+\infty} \mathbf{A}^{(k)}$ 收敛;

(2) 级数 $\sum_{k=0}^{+\infty} \mathbf{A}^{(k)}$ 在任意改变各项的次序后仍然收敛, 且其和不变.

这条性质的证明, 读者只要依照数项级数的狄利克雷(Dirichlet)定理的证明立刻可得.

性质 2 设 \mathbf{P}, \mathbf{Q} 为 n 阶非奇异矩阵, 若级数 $\sum_{k=0}^{+\infty} \mathbf{A}^{(k)}$ 收敛(或绝对收敛), 则矩阵级数

$\sum_{k=0}^{+\infty} \mathbf{P}\mathbf{A}^{(k)}\mathbf{Q}$ 也收敛(或绝对收敛).

证明 设 $\sum_{k=0}^{+\infty} \mathbf{A}^{(k)} = \mathbf{S}$, 部分和 $\sum_{k=0}^{+\infty} \mathbf{A}^{(k)} = \mathbf{S}_N$, 则

$$\lim_{N \rightarrow +\infty} \mathbf{S}_N = \mathbf{S}.$$

显然有

$$\begin{aligned} \lim_{N \rightarrow +\infty} \mathbf{P}\mathbf{S}_N\mathbf{Q} &= \mathbf{P}(\lim_{N \rightarrow +\infty} \mathbf{S}_N)\mathbf{Q} = \mathbf{P}\mathbf{S}\mathbf{Q} \\ &= \mathbf{P}\left(\sum_{k=0}^{+\infty} \mathbf{A}^{(k)}\right)\mathbf{Q}, \end{aligned}$$

即当 $\sum_{k=0}^{+\infty} \mathbf{A}^{(k)}$ 收敛时, $\sum_{k=0}^{+\infty} \mathbf{P}\mathbf{A}^{(k)}\mathbf{Q}$ 也收敛.

若 $\sum_{k=0}^{+\infty} \mathbf{A}^{(k)}$ 绝对收敛, 由定理 5.2.1 知, $\sum_{k=0}^{+\infty} \|\mathbf{A}^{(k)}\|$ 也是收敛的, 由矩阵范数的相容性, 有

$$\begin{aligned} \|\mathbf{P}\mathbf{A}^{(k)}\mathbf{Q}\| &\leq \|\mathbf{P}\| \|\mathbf{A}^{(k)}\| \|\mathbf{Q}\| = \|\mathbf{P}\| \|\mathbf{Q}\| \|\mathbf{A}^{(k)}\| \\ &= M \|\mathbf{A}^{(k)}\|. \end{aligned}$$

其中 $M = \|\mathbf{P}\| \|\mathbf{Q}\|$, 由此可知 $\sum_{k=0}^{+\infty} \|\mathbf{P}\mathbf{A}^{(k)}\mathbf{Q}\|$ 也收敛, 于是 $\sum_{k=0}^{+\infty} \mathbf{P}\mathbf{A}^{(k)}\mathbf{Q}$ 绝对收敛. 证毕

对于矩阵级数也有幂级数的概念.

定义 5.2.4 形如

$$c_0 \mathbf{I} + c_1 \mathbf{A} + c_2 \mathbf{A}^2 + \cdots + c_k \mathbf{A}^k + \cdots \quad (5.2.14)$$

的矩阵级数称为矩阵幂级数, 其中 $c_i \in \mathbb{C}$, $\mathbf{A} \in \mathbb{C}^{n \times n}$.

因为数项级数

$$\|c_0 \mathbf{I}\| + \|c_1 \mathbf{A}\| + \|c_2 \mathbf{A}^2\| + \cdots + \|c_k \mathbf{A}^k\| + \cdots$$

的每一项不大于数项级数

$$|c_0| \|\mathbf{I}\| + |c_1| \|\mathbf{A}\| + |c_2| \|\mathbf{A}\|^2 + \cdots + |c_k| \|\mathbf{A}\|^k + \cdots \quad (5.2.15)$$

的对应项, 所以, 若级数(5.2.15)收敛, 则矩阵幂级数(5.2.14)绝对收敛. 于是我们得到下面定理.

定理 5.2.3 若正项级数 $|c_0| \|\mathbf{I}\| + \sum_{k=1}^{+\infty} |c_k| \|\mathbf{A}\|^k$ 收敛, 则矩阵幂级数 $c_0 \mathbf{I} + c_1 \mathbf{A} + c_2 \mathbf{A}^2 + \cdots + c_k \mathbf{A}^k + \cdots$ 绝对收敛, 其中 $\|\mathbf{A}\|$ 为矩阵 \mathbf{A} 的某种范数.

推论 若矩阵 \mathbf{A} 的某一种范数 $\|\mathbf{A}\|$ 在幂级数

$$\sum_{k=0}^{+\infty} c_0 z^k = c_0 + c_1 z + c_2 z^2 + \cdots + c_k z^k + \cdots$$

的收敛圆内, 则矩阵幂级数 $\sum_{k=0}^{+\infty} c_k \mathbf{A}^k$ 绝对收敛.

例 5.2.2 设

$$\mathbf{A} = \begin{pmatrix} 0.2 & 0.5 & 0.1 \\ 0.1 & 0.5 & 0.3 \\ 0.2 & 0.4 & 0.2 \end{pmatrix},$$

试证: 矩阵幂级数

$$\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^k + \cdots \quad (5.2.16)$$

绝对收敛.

证明 因为幂级数 $1 + x + x^2 + \cdots + x^k + \cdots$ 的收敛半径为 1, 而 $\|\mathbf{A}\|_\infty = \max_i \left(\sum_{j=1}^n |a_{ij}| \right) = 0.9 < 1$. 故由上面推论知级数(5.2.16)绝对收敛. 证毕

注意, 定理 5.2.3 的推论中的范数并不要求 \mathbf{A} 的任何一种范数均在收敛圆内, 而仅要求 \mathbf{A} 的某一种范数在收敛圆内. 在例 5.2.2 中若取 $\|\mathbf{A}\|_1 = 1.4 > 1$ 就无法判定矩阵级数是否绝对收敛, 而取 $\|\mathbf{A}\|_\infty$ 就能判定它是绝对收敛的. 这正是定理 5.2.3 及其推论的局限之处.

定理 5.2.4 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 如果 \mathbf{A} 的谱半径 $\rho(\mathbf{A})$ 的值在纯量 z 的幂级数 $\sum_{k=0}^{+\infty} c_k z^k$ 的收敛圆内, 那么矩阵幂级数 $\sum_{k=0}^{+\infty} c_k \mathbf{A}^k$ 绝对收敛; 如果 \mathbf{A} 的特征值中有一个在幂级数 $\sum_{k=0}^{+\infty} c_k z^k$ 的收敛圆外, 则矩阵幂级数 $\sum_{k=0}^{+\infty} c_k \mathbf{A}^k$ 发散.

证明 设 J 为 A 的若尔当标准形, 则

$$A = TJT^{-1} = T \begin{pmatrix} J_1 & & \\ & J_2 & \\ & & \ddots \\ & & & J_r \end{pmatrix} T^{-1},$$

其中

$$J_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & 1 & \\ & & \lambda_i & \ddots \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{pmatrix}_{n_i \times n_i},$$

因此

$$A^k = TJ^k T^{-1} = T \begin{pmatrix} J_1^k & & \\ & J_2^k & \\ & & \ddots \\ & & & J_r^k \end{pmatrix} T^{-1}.$$

于是

$$\sum_{k=0}^{+\infty} c_k A^k = T \begin{pmatrix} \sum_{k=0}^{+\infty} c_k J_1^k & & \\ & \sum_{k=0}^{+\infty} c_k J_2^k & \\ & & \ddots \\ & & & \sum_{k=0}^{+\infty} c_k J_r^k \end{pmatrix} T^{-1},$$

而

$$\sum_{k=0}^{+\infty} c_k J_i^k = \begin{pmatrix} \sum_{k=0}^{+\infty} c_k \lambda_i^k & \sum_{k=0}^{+\infty} c_k C_k^1 \lambda_i^{k-1} & \sum_{k=0}^{+\infty} c_k C_k^2 \lambda_i^{k-2} & \cdots & \sum_{k=0}^{+\infty} c_k C_k^{n_i-1} \lambda_i^{k-n_i+1} \\ & \sum_{k=0}^{+\infty} c_k \lambda_i^k & \sum_{k=0}^{+\infty} c_k C_k^1 \lambda_i^{k-1} & \cdots & \sum_{k=0}^{+\infty} c_k C_k^{n_i-2} \lambda_i^{k-n_i+2} \\ & & \ddots & \cdots & \vdots \\ & & & \ddots & \sum_{k=0}^{+\infty} c_k \lambda_i^k \end{pmatrix}, \quad (5.2.17)$$

其中

$$C_k^i = \frac{k(k-1)\cdots(k-i+1)}{i!}, \quad \text{当 } i \geq 1,$$

$$C_k^0 = 0.$$

令 R 为幂级数 $\sum_{k=0}^{+\infty} c_k z^k$ 的收敛半径, 当 $|\lambda_i| > R$ 时, $\sum_{k=0}^{+\infty} c_k J_i^k$ 发散; 当 $|\lambda_i| < R$ 时, 级数

$$\sum_{k=0}^{+\infty} c_k \lambda_i^k, \sum_{k=0}^{+\infty} c_k k \lambda_i^{k-1}, \dots, \sum_{k=0}^{+\infty} c_k k(k-1)\cdots(k-n_i+1) \lambda_i^{k-n_i+1}$$

均绝对收敛. 因此, 级数

$$\sum_{k=0}^{+\infty} c_k \lambda_i^k, \sum_{k=0}^{+\infty} c_k C_k^1 \lambda_i^{k-1}, \dots, \sum_{k=0}^{+\infty} c_k C_k^{n_i-1} \lambda_i^{k-n_i+1}$$

也都绝对收敛. 其余同理. 因此, 式(5.2.17)右端的所有级数都绝对收敛, 故当所有的

$|\lambda_i| < R$ 时, $\sum_{k=0}^{+\infty} c_k J_i^k$ 绝对收敛, 从而级数 $\sum_{k=0}^{+\infty} c_k \mathbf{A}^k$ 绝对收敛. 证毕

定理 5.2.5 矩阵幂级数 $\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^k + \cdots$ 绝对收敛的充要条件是 \mathbf{A} 的谱半径 $\rho(\mathbf{A}) < 1$, 且该级数的和为 $(\mathbf{I} - \mathbf{A})^{-1}$.

证明 因为幂级数 $1 + x + x^2 + \cdots + x^k + \cdots$ 的收敛半径为 1, 所以当 \mathbf{A} 的谱半径 $\rho(\mathbf{A}) < 1$ 时, 由定理 5.2.4 知, 级数

$$\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^k + \cdots$$

绝对收敛.

反之, 若 $\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^k + \cdots$ 绝对收敛, 则 $\lim_{k \rightarrow +\infty} \mathbf{A}^k = \mathbf{0}$, 由定理 5.1.2 知, \mathbf{A} 的谱半径小于 1.

下面计算级数 $\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^k + \cdots$ 的和. 因为 $\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^k + \cdots$ 绝对收敛, 而 $\mathbf{I} - \mathbf{A}$ 本身作为矩阵级数也绝对收敛, 由定理 5.2.2 知, 它们的乘积

$$(\mathbf{I} - \mathbf{A})(\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^k + \cdots) = \mathbf{I},$$

由逆矩阵的惟一性, 有

$$(\mathbf{I} - \mathbf{A})^{-1} = \mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^k + \cdots. \quad \text{证毕}$$

例 5.2.3 求 $\sum_{k=0}^{+\infty} \begin{bmatrix} 0.1 & 0.7 \\ 0.3 & 0.6 \end{bmatrix}^k$ 的和.

解 设 $\mathbf{A} = \begin{bmatrix} 0.1 & 0.7 \\ 0.3 & 0.6 \end{bmatrix}$, 容易求得它的谱半径 $\rho(\mathbf{A}) = \frac{0.7 + \sqrt{1.09}}{2} < 1$, 从而

$$\sum_{k=0}^{+\infty} \begin{bmatrix} 0.1 & 0.7 \\ 0.3 & 0.6 \end{bmatrix}^k = (\mathbf{I} - \mathbf{A})^{-1} = \frac{1}{0.15} \begin{bmatrix} 0.4 & 0.7 \\ 0.3 & 0.9 \end{bmatrix}.$$

定理 5.2.6 设矩阵 \mathbf{A} 的某种范数 $\|\mathbf{A}\| < 1$, 则对任何非负整数 k , 有

$$\begin{aligned} & \|(\mathbf{I} - \mathbf{A})^{-1} - (\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^k)\| \\ & \leq \frac{\|\mathbf{A}\|^{k+1}}{1 - \|\mathbf{A}\|}. \end{aligned} \quad (5.2.18)$$

证明 因为对任何非负整数 k , 有

$$\begin{aligned} & \mathbf{I} - [(\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^k) + (\mathbf{A}^{k+1} + \mathbf{A}^{k+2} + \cdots + \mathbf{A}^{k+m})](\mathbf{I} - \mathbf{A}) \\ & = \mathbf{A}^{k+m+1}, \end{aligned}$$

所以有

$$\begin{aligned} & (\mathbf{I} - \mathbf{A})^{-1} - (\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^k) \\ & = \mathbf{A}^{k+1} + \mathbf{A}^{k+2} + \cdots + \mathbf{A}^{k+m} + \mathbf{A}^{k+m+1}(\mathbf{I} - \mathbf{A})^{-1}. \end{aligned}$$

记

$$\mathbf{B}_m = \mathbf{A}^{k+1} + \mathbf{A}^{k+2} + \cdots + \mathbf{A}^{k+m},$$

因为 $\|A\| < 1$, 所以必有 $A^m \rightarrow 0$ (定理 5.1.3), 故 $A^{k+m+1}(I-A)^{-1} \rightarrow 0$,

$$\lim_{m \rightarrow +\infty} B_m = (I-A)^{-1} - (I+A+A^2+\cdots+A^k).$$

由范数的连续性知, 有

$$\|(I-A)^{-1} - (I+A+A^2+\cdots+A^k)\| = \lim_{m \rightarrow +\infty} \|B_m\|.$$

但

$$\begin{aligned} \|B_m\| &\leq \|A^{k+1}\| + \|A^{k+2}\| + \cdots + \|A^{k+m}\| \\ &\leq \|A\|^{k+1} + \|A\|^{k+2} + \cdots + \|A\|^{k+m} \\ &= \frac{\|A\|^{k+1}(1-\|A\|^m)}{1-\|A\|}, \end{aligned}$$

故对一切正整数 m , 有

$$\|B_m\| \leq \frac{\|A\|^{k+1}}{1-\|A\|},$$

于是

$$\|(I-A)^{-1} - (I+A+A^2+\cdots+A^k)\| \leq \frac{\|A\|^{k+1}}{1-\|A\|}. \quad \text{证毕}$$

定理 5.2.6 表明, 以 $(I-A)^{-1}$ 为部分和 $I+A+A^2+\cdots+A^k$ 的近似时, 其误差如式(5.2.18)所示.

5.2.2 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均是 n 阶方阵. 本节我们将以定理 5.2.4 为依据, 先通过收敛的矩阵级数给出矩阵函数的定义, 然后再讨论矩阵函数值的计算等问题.

1. 矩阵函数 $e^A, \sin A, \cos A$ 的定义与性质

在函数论中, 我们熟知纯量 z 的级数

$$\begin{aligned} e^z &= 1 + \frac{z}{1!} + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots + \frac{z^k}{k!} + \cdots, \\ \sin z &= z - \frac{z^3}{3!} + \frac{z^5}{5!} - \cdots + (-1)^k \frac{z^{2k+1}}{(2k+1)!} + \cdots, \\ \cos z &= 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \cdots + (-1)^k \frac{z^{2k}}{(2k)!} + \cdots \end{aligned}$$

在整个复平面上都是收敛的, 于是, 根据定理 5.2.4 可知, 不论 $A \in \mathbb{C}^{n \times n}$ 是任何矩阵, 矩阵幂级数

$$\begin{aligned} I + \frac{A}{1!} + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots + \frac{A^k}{k!} + \cdots, \\ A - \frac{A^3}{3!} + \frac{A^5}{5!} - \cdots + (-1)^k \frac{A^{2k+1}}{(2k+1)!} + \cdots, \\ I - \frac{A^2}{2!} + \frac{A^4}{4!} - \cdots + (-1)^k \frac{A^{2k}}{(2k)!} + \cdots \end{aligned}$$

都是绝对收敛的, 因此它们都有和. 受数学分析的启发, 我们很自然地采用记号 $e^A, \sin A, \cos A$ 来依次表示它们的和, 即有

$$e^A = \sum_{k=0}^{+\infty} \frac{A^k}{k!}, \quad (5.2.19)$$

$$\sin A = \sum_{k=0}^{+\infty} (-1)^k \frac{A^{2k+1}}{(2k+1)!}, \quad (5.2.20)$$

$$\cos A = \sum_{k=0}^{+\infty} (-1)^k \frac{A^{2k}}{(2k)!}. \quad (5.2.21)$$

定义 5.2.5 设实函数 $y=f(x)$, $A, B \in \mathbb{C}^{n \times n}$, 称 $B=f(A)$ 为矩阵 A 的函数.

我们称式(5.2.19)为矩阵指数函数, 式(5.2.20)及式(5.2.21)为矩阵三角函数, 矩阵函数有如下的一些性质:

定理 5.2.7 如果 $AB=BA$, 则有

$$e^A \cdot e^B = e^B \cdot e^A = e^{A+B}.$$

证明 由于 $A+B=B+A$, 所以只需证明 $e^{A+B}=e^A \cdot e^B$.

根据矩阵级数的乘法, 我们有

$$\begin{aligned} e^A \cdot e^B &= \left(I + A + \frac{1}{2!}A^2 + \cdots \right) \left(I + B + \frac{1}{2!}B^2 + \cdots \right) \\ &= I + (A+B) + \frac{1}{2!}(A^2 + 2AB + B^2) + \frac{1}{3!}(A^3 + 3A^2B + 3AB^2 + B^3) + \cdots \\ &= I + (A+B) + \frac{1}{2!}(A+B)^2 + \frac{1}{3!}(A+B)^3 + \cdots \\ &= e^{A+B}. \end{aligned}$$

证毕

推论 1 对任意矩阵 $A \in \mathbb{C}^{n \times n}$, e^A 总是可逆的(非奇异的)且 $(e^A)^{-1} = e^{-A}$.

事实上, 因为 $e^A \cdot e^{-A} = e^{-A} \cdot e^A = I$, 故有 $(e^A)^{-1} = e^{-A}$.

推论 2 $(e^A)^m = e^{mA}$ (m 为整数).

值得指出的是, 如果矩阵 A, B 的乘法不满足交换律, 即有 $AB \neq BA$, 则 $e^A \cdot e^B = e^B \cdot e^A = e^{A+B}$ 不再成立. 例如, 令

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix},$$

则易知 $A^2=A, B^2=B$, 从而有

$$A = A^2 = A^3 = \cdots, \quad B = B^2 = B^3 = \cdots,$$

于是得

$$\begin{aligned} e^A &= I + (e-1)A = \begin{bmatrix} e & e-1 \\ 0 & 1 \end{bmatrix}, \\ e^B &= I + (e-1)B = \begin{bmatrix} e & 1-e \\ 0 & 1 \end{bmatrix}, \end{aligned}$$

从而

$$e^A \cdot e^B = \begin{bmatrix} e^2 & -(e-1)^2 \\ 0 & 1 \end{bmatrix}, \quad e^B \cdot e^A = \begin{bmatrix} e^2 & (e-1)^2 \\ 0 & 1 \end{bmatrix},$$

故 $e^A \cdot e^B \neq e^B \cdot e^A$.

又由 $A+B = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$, $(A+B)^2 = 2(A+B)$, 于是有 $(A+B)^k = 2^{k-1}(A+B)$ ($k=1, 2, \dots$).

由此容易推出

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

这就说明 $e^A \cdot e^B, e^B \cdot e^A$, 以及 e^{A+B} 两两互不相等.

读者还不难推得如下的一些结果

$$\begin{cases} e^{iA} = \cos A + i \sin A, & i = \sqrt{-1}, \\ \cos A = \frac{1}{2}(e^{iA} + e^{-iA}), \\ \sin A = \frac{1}{2i}(e^{iA} - e^{-iA}), \\ \cos(-A) = \cos A, \\ \sin(-A) = -\sin A. \end{cases} \quad (5.2.22)$$

如果 $AB=BA$, 则还有下面有趣的结果

$$\begin{cases} \cos(A+B) = \cos A \cos B - \sin A \sin B, \\ \cos 2A = \cos^2 A - \sin^2 A, \\ \sin(A+B) = \sin A \cos B + \cos A \sin B, \\ \sin 2A = 2 \sin A \cos A. \end{cases} \quad (5.2.23)$$

2. 矩阵函数值的求法

如同求通常的函数值一样, 当给定了自变量矩阵 A 时, 如何求矩阵函数 $f(A)$ 的值呢? 如果不研究切实可行的算法, 矩阵函数值的计算是相当复杂的. 例如, 简单的矩阵函数 A^{101} 就要计算 100 次矩阵的乘法; 若 A 为 5 阶方阵, 则要进行 22500 次加法和乘法运算. 研究如何方便地计算矩阵函数是非常有意义的. 为此, 我们将介绍两种常用的算法.

(1) 递推公式计算法

通过下面的例题来说明递推计算法.

例 5.2.4 已知 4 阶矩阵 A 的特征值为 $\pi, -\pi, 0, 0$, 求 $\sin A, \cos A, e^A$.

解 因为 A 的特征方程为

$$\det(\lambda I - A) = (\lambda - \pi)(\lambda + \pi)\lambda^2 = \lambda^4 - \pi^2 \lambda^2 = 0,$$

由凯莱-哈密顿定理(见定理 2.2.4)知, 有

$$A^4 = \pi^2 A^2.$$

从而得

$$\begin{aligned} \sin A &= A - \frac{1}{3!}A^3 + \frac{1}{5!}A^5 - \frac{1}{7!}A^7 + \frac{1}{9!}A^9 - \dots \\ &= A - \frac{1}{3!}A^3 + \frac{1}{5!}\pi^2 A^3 - \frac{1}{7!}\pi^4 A^3 + \frac{1}{9!}\pi^6 A^3 - \dots \\ &= A + \left(-\frac{1}{3!} + \frac{1}{5!}\pi^2 - \frac{1}{7!}\pi^4 + \frac{1}{9!}\pi^6 - \dots\right)A^3 \\ &= A + \frac{\sin \pi - \pi}{\pi^3}A^3 = A - \pi^{-2}A^3, \end{aligned}$$

$$\begin{aligned}
\cos A &= I - \frac{1}{2!}A^2 + \frac{1}{4!}A^4 - \frac{1}{6!}A^6 + \frac{1}{8!}A^8 - \cdots \\
&= I - \frac{1}{2!}A^2 + \frac{1}{4!}\pi^2 A^2 - \frac{1}{6!}\pi^4 A^2 + \frac{1}{8!}\pi^6 A^2 - \cdots \\
&= I + \left(-\frac{1}{2!} + \frac{1}{4!}\pi^2 - \frac{1}{6!}\pi^4 + \frac{1}{8!}\pi^6 - \cdots\right)A^2 \\
&= I + \frac{\cos \pi - 1}{\pi^2}A^2 = I - 2\pi^{-2}A^2, \\
e^A &= I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \frac{1}{4!}A^4 + \cdots \\
&= I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \frac{1}{4!}\pi^2 A^2 + \frac{1}{5!}\pi^2 A^3 + \frac{1}{6!}\pi^4 A^2 + \frac{1}{7!}\pi^4 A^3 + \cdots \\
&= I + A + \left(\frac{1}{2!} + \frac{1}{4!}\pi^2 + \frac{1}{6!}\pi^4 + \cdots\right)A^2 + \left(\frac{1}{3!} + \frac{1}{5!}\pi^2 + \frac{1}{7!}\pi^4 + \cdots\right)A^3 \\
&= I + A + \frac{p}{\pi^2}A^2 + \frac{e^\pi - \pi - 1 - p}{\pi^3}A^3,
\end{aligned}$$

这里

$$p = \frac{1}{2!}\pi^2 + \frac{1}{4!}\pi^4 + \frac{1}{6!}\pi^6 + \cdots = \frac{e^\pi + e^{-\pi}}{2} - 1 = \cosh \pi - 1.$$

(2) 相似标准形算法

先介绍一个有关的定理.

定理 5.2.8 设方阵 A 的 n 个特征值为 $\lambda_i (i=1, 2, \dots, n)$, $f(z)$ 为一多项式, 则方阵多项式 $f(A)$ 的 n 个特征值为 $f(\lambda_i) (i=1, 2, \dots, n)$.

证明 设 $\varphi(\lambda)$ 为 A 的特征多项式, 则有

$$\varphi(\lambda) = |\lambda I - A| = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n).$$

记 $\mu - f(z) \equiv F(z)$, μ 为一与 z 无关的参数, 即

$$\begin{aligned}
F(z) &\equiv \mu - f(z) = b_0 z^k + b_1 z^{k-1} + \cdots + b_k \\
&\equiv b_0 (z - \alpha_1)(z - \alpha_2) \cdots (z - \alpha_k),
\end{aligned}$$

于是有

$$F(A) = b_0 (A - \alpha_1 I)(A - \alpha_2 I) \cdots (A - \alpha_k I).$$

但是

$$\begin{aligned}
|A - \alpha_j I| &= (-1)^n |\alpha_j I - A| \\
&= (-1)^n \varphi(\alpha_j), \quad j = 1, 2, \dots, k,
\end{aligned}$$

所以

$$\begin{aligned}
|F(A)| &= |\mu I - f(A)| = b_0^n |A - \alpha_1 I| |A - \alpha_2 I| \cdots |A - \alpha_k I| \\
&= b_0^n \prod_{j=1}^k (-1)^n \varphi(\alpha_j) = b_0^n \prod_{j=1}^k (-1)^n (\alpha_j - \lambda_1) \cdots (\alpha_j - \lambda_n) \\
&= b_0^n \prod_{j=1}^k (\lambda_1 - \alpha_j)(\lambda_2 - \alpha_j) \cdots (\lambda_n - \alpha_j) \\
&= \prod_{i=1}^k F(\lambda_i) = \prod_{i=1}^k [\mu - f(\lambda_i)],
\end{aligned}$$

$$\text{即 } |F(\mathbf{A})| = |\mu \mathbf{I} - f(\mathbf{A})| = \prod_{i=1}^k [\mu - f(\lambda_i)],$$

这表示 $f(\mathbf{A})$ 的 n 个特征值为 $f(\lambda_i) (i=1, 2, \dots, n)$.

证毕

利用上面的定理和矩阵幂级数的概念, 我们又可以得到实际求矩阵函数值的方法.

(i) \mathbf{A} 为非亏损矩阵的情况.

假定 \mathbf{A} 与对角矩阵相似, 即存在非奇异矩阵 \mathbf{C} , 使

$$\mathbf{C}^{-1}\mathbf{A}\mathbf{C} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n),$$

于是有

$$\mathbf{C}^{-1}\mathbf{A}^m\mathbf{C} = \text{diag}(\lambda_1^m, \lambda_2^m, \dots, \lambda_n^m).$$

若 $f(z)$ 为一任意多项式, 则由定理 5.2.8 知 $f(\mathbf{A})$ 的特征值为 $f(\lambda_i) (i=1, 2, \dots, n)$. 从而有

$$\mathbf{C}^{-1}f(\mathbf{A})\mathbf{C} = \text{diag}(f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n)). \quad (5.2.24)$$

若 $f(z) = \sum_{k=0}^{+\infty} c_k z^k$ 在整个复平面上收敛, 则由定理 5.2.4 知矩阵幂级数 $f(\mathbf{A}) = \sum_{k=0}^{+\infty} c_k \mathbf{A}^k$ 是

绝对收敛的. 于是其 $N+1$ 项部分和记为 $f_N(\mathbf{A}) = \sum_{k=0}^N c_k \mathbf{A}^k \xrightarrow{N \rightarrow +\infty} f(\mathbf{A})$. 再以 $f_N(\mathbf{A})$ 代替式(5.2.24)中的 $f(\mathbf{A})$ 便得

$$\mathbf{C}^{-1}f_N(\mathbf{A})\mathbf{C} = \text{diag}(f_N(\lambda_1), f_N(\lambda_2), \dots, f_N(\lambda_n)).$$

再由 $\lim_{N \rightarrow +\infty} f_N(\mathbf{A}) = f(\mathbf{A})$ 及 $\lim_{N \rightarrow +\infty} f_N(\lambda_i) = f(\lambda_i)$, 便得

$$\mathbf{C}^{-1}f(\mathbf{A})\mathbf{C} = \text{diag}(f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n)),$$

即

$$f(\mathbf{A}) = \mathbf{C} \cdot \text{diag}(f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n)) \cdot \mathbf{C}^{-1}. \quad (5.2.25)$$

在式(5.2.25)中以 $e^{\mathbf{A}}, \sin \mathbf{A}, \cos \mathbf{A}$ 代替 $f(\mathbf{A})$, 就得到计算矩阵函数 $e^{\mathbf{A}}, \sin \mathbf{A}, \cos \mathbf{A}$ 值的公式

$$\begin{cases} e^{\mathbf{A}} = \mathbf{C} \cdot \text{diag}(e^{\lambda_1}, e^{\lambda_2}, \dots, e^{\lambda_n}) \cdot \mathbf{C}^{-1}, \\ \sin \mathbf{A} = \mathbf{C} \cdot \text{diag}(\sin \lambda_1, \sin \lambda_2, \dots, \sin \lambda_n) \cdot \mathbf{C}^{-1}, \\ \cos \mathbf{A} = \mathbf{C} \cdot \text{diag}(\cos \lambda_1, \cos \lambda_2, \dots, \cos \lambda_n) \cdot \mathbf{C}^{-1}. \end{cases} \quad (5.2.26)$$

例 5.2.5 设

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 2 & 3 \end{bmatrix},$$

求 \mathbf{A}^{101} .

解 因为 \mathbf{A} 的特征多项式为 $\det(\lambda \mathbf{I} - \mathbf{A}) = (\lambda - 1)(\lambda - 4)$, 所以两个特征值为 1 和 4, 它们对应的特征向量为 $(-1, 1)^T$ 和 $(1, 2)^T$.

从而化 \mathbf{A} 为对角阵的非奇异矩阵为

$$\mathbf{C} = \begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{C}^{-1} = \begin{bmatrix} -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

应用式(5.2.25)可得

$$\begin{aligned} \mathbf{A}^{101} &= \mathbf{C} \begin{pmatrix} \lambda_1^{101} & \\ & \lambda_2^{101} \end{pmatrix} \mathbf{C}^{-1} \\ &= \frac{1}{3} \begin{pmatrix} 4^{101} + 2 & 4^{101} - 1 \\ 2(4^{101} - 1) & 2 \times 4^{101} + 1 \end{pmatrix}. \end{aligned}$$

例 5.2.6 设

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix},$$

计算 $\mathbf{A}^4 + 2\mathbf{A}^3 + \mathbf{A}$ 及 $e^{\mathbf{A}}$.

解 因为特征多项式为

$$\begin{aligned} \det(\lambda \mathbf{I} - \mathbf{A}) &= \begin{vmatrix} \lambda - 3 & -2 \\ -1 & \lambda - 2 \end{vmatrix} = \lambda^2 - 5\lambda + 4 \\ &= (\lambda - 1)(\lambda - 4), \end{aligned}$$

可见 1 与 4 都是单特征值,且它们对应的特征向量依次为 $(-1, 1)^T$ 与 $(2, 1)^T$. 于是化 \mathbf{A} 为对角阵的非奇异矩阵是

$$\mathbf{C} = \begin{bmatrix} -1 & 2 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{C}^{-1} = \begin{pmatrix} -\frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

令式(5.2.25)中 $f(\mathbf{A}) = \mathbf{A}^4 + 2\mathbf{A}^3 + \mathbf{A}$, 得

$$\begin{aligned} \mathbf{A}^4 + 2\mathbf{A}^3 + \mathbf{A} &= \mathbf{C} \cdot \text{diag}(1^4 + 2 \times 1^3 + 1, 4^4 + 2 \times 4^3 + 4) \cdot \mathbf{C}^{-1} \\ &= \begin{bmatrix} -1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 4 & \\ & 388 \end{bmatrix} \begin{pmatrix} -\frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix} \\ &= \begin{bmatrix} 260 & 256 \\ 128 & 132 \end{bmatrix}. \end{aligned}$$

再令式(5.2.25)中 $f(\mathbf{A}) = e^{\mathbf{A}}$, 得

$$\begin{aligned} e^{\mathbf{A}} &= \mathbf{C} \cdot \text{diag}(e, e^4) \cdot \mathbf{C}^{-1} \\ &= \begin{bmatrix} -1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} e & \\ & e^4 \end{bmatrix} \begin{pmatrix} -\frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{3}(2e^4 + e) & \frac{2}{3}(e^4 - e) \\ \frac{1}{3}(e^4 - e) & \frac{1}{3}(2e + e^4) \end{pmatrix}. \end{aligned}$$

例 5.2.7 设矩阵

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

试求 $e^{\mathbf{A}}$, $\sin \mathbf{A}$, $\cos \mathbf{A}$.

解 因为 \mathbf{A} 的特征多项式是

$$\det(\lambda \mathbf{I} - \mathbf{A}) = (\lambda + 2)(\lambda - 1)^2,$$

可见 -2 是单特征值, 1 是二重特征值. 不难求得属于 -2 的特征向量是 $(-1, 1, 1)^T$, 而属于 1 的特征向量是 $(-2, 1, 0)^T$ 和 $(0, 0, 1)^T$. 于是化 \mathbf{A} 为对角阵的非奇异矩阵

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

应用公式(5.2.25)直接可得

$$\begin{aligned} e^{\mathbf{A}} &= \mathbf{C} \cdot \text{diag}(e^{-2}, e, e) \cdot \mathbf{C}^{-1} \\ &= \begin{pmatrix} 2e - e^{-2} & 2e - 2e^{-2} & 0 \\ e^{-2} - e & 2e^{-2} - e & 0 \\ e^{-2} - e & 2e^{-2} - 2e & e \end{pmatrix}, \\ \sin \mathbf{A} &= \begin{pmatrix} 2\sin 1 + \sin 2 & 2\sin 1 + 2\sin 2 & 0 \\ -\sin 2 - \sin 1 & -2\sin 2 - \sin 1 & 0 \\ -\sin 2 - \sin 1 & -2\sin 2 - 2\sin 1 & \sin 1 \end{pmatrix}, \\ \cos \mathbf{A} &= \begin{pmatrix} 2\cos 1 - \cos 2 & 2\cos 1 - 2\cos 2 & 0 \\ \cos 2 - \cos 1 & 2\cos 2 - \cos 1 & 0 \\ \cos 2 - \cos 1 & 2\cos 2 - 2\cos 1 & \cos 1 \end{pmatrix}. \end{aligned}$$

在这里必须指出, 假定方阵 \mathbf{A} 可化为对角矩阵这个条件较强, 一般不易满足, 例如

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

这是一个亏损矩阵, 它不能与对角矩阵相似, 因此不能用前面的方法求 $e^{\mathbf{A}}$ 等. 下面我们给出另一种情况的算法.

(ii) \mathbf{A} 为亏损矩阵的情况.

定义 5.2.6 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 显然存在 \mathbf{A} 的若尔当标准形分解 $\mathbf{A} = \mathbf{T} \mathbf{J} \mathbf{T}^{-1}$, 如果函数 $f(z)$ 在各个特征值 $\lambda_i (i=1, 2, \dots, r)$ 处具有直到 $n_i - 1$ 阶导数 ($n_i (i=1, 2, \dots, r)$ 为初等因子的指数), 则定义矩阵函数为

$$f(\mathbf{A}) = \mathbf{T} \cdot f(\mathbf{J}) \cdot \mathbf{T}^{-1}, \quad (5.2.27)$$

其中

$$\begin{aligned} f(\mathbf{J}) &= \text{diag}(f(\mathbf{J}_1(\lambda_1)), f(\mathbf{J}_2(\lambda_2)), \dots, f(\mathbf{J}_r(\lambda_r))), \\ f(\mathbf{J}_i(\lambda_i)) &= \begin{pmatrix} f(\lambda_i) & f'(\lambda_i) & \frac{1}{2!} f''(\lambda_i) & \cdots & \frac{1}{(n_i - 1)!} f^{(n_i - 1)}(\lambda_i) \\ & f(\lambda_i) & f'(\lambda_i) & \cdots & \frac{1}{(n_i - 2)!} f^{(n_i - 2)}(\lambda_i) \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & f'(\lambda_i) \\ & & & & f(\lambda_i) \end{pmatrix}_{n_i \times n_i}. \end{aligned} \quad (5.2.28)$$

① 事实上, 我们将函数 $f(z)$ 展开为幂级数, 再把约当块 $\mathbf{J}_i(\lambda_i)$ 代入就可证明这里的 $f(\mathbf{J}_i(\lambda_i))$

我们称 $f(\mathbf{A})$ 的表示式(5.2.28)为 $f(\mathbf{A})$ 的若尔当标准形.

特别,若 \mathbf{A} 为单纯矩阵,则 \mathbf{J} 为对角矩阵,显然此时 $f(\mathbf{J})$ 也是对角矩阵,且与前面用幂级数方法定义 $f(\mathbf{A})$ 是一致的.

例 5.2.8 设

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

求 $e^{\mathbf{A}t}$ (其中 t 为参数).

解 因为

$$|\lambda \mathbf{I} - \mathbf{A}| = \begin{vmatrix} \lambda - 3 & -1 & 1 \\ -1 & \lambda - 2 & 1 \\ -2 & -1 & \lambda \end{vmatrix} = (\lambda - 1)(\lambda - 2)^2,$$

显然 $n_1 = 1, n_2 = 2$, 于是 \mathbf{A} 的若尔当标准形为

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

由于对应 $\lambda_1 = 1$ 的特征向量为 $(0, 1, 1)^T$, 而对应 $\lambda_2 = 2$ 的线性无关的特征向量只有一个 $(1, 0, 1)^T$, 为此再求出它的一个广义特征向量 $(1, 1, 1)^T$. 于是非奇异矩阵 \mathbf{T} 取为

$$\mathbf{T} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \mathbf{T}^{-1} = \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}.$$

容易看出

$$f(\mathbf{J}_1(\lambda_1)) = f(\lambda_1) = e^t,$$

$$f(\mathbf{J}_2(\lambda_2)) = \begin{pmatrix} e^{2t} & te^{2t} \\ 0 & e^{2t} \end{pmatrix},$$

所以

$$\begin{aligned} e^{\mathbf{A}t} &= \mathbf{T} \cdot \text{diag}(f(\mathbf{J}_1(\lambda_1)), f(\mathbf{J}_2(\lambda_2))) \cdot \mathbf{T}^{-1} \\ &= \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} e^t & 0 & 0 \\ 0 & e^{2t} & te^{2t} \\ 0 & 0 & e^{2t} \end{pmatrix} \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} (t+1)e^{2t} & te^{2t} & -te^{2t} \\ -e^t + e^{2t} & e^{2t} & e^t + e^{2t} \\ -e^t + (t+1)e^{2t} & te^{2t} & e^t - te^{2t} \end{pmatrix}. \end{aligned}$$

利用矩阵函数的若尔当标准形表示法,还可以求出一些特殊矩阵函数的值.

例 5.2.9 (1) 已知

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

求 $\frac{1}{\mathbf{A}}$;

(2) 已知

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

求 \mathbf{A} 的平方根 $\sqrt{\mathbf{A}}$.

解 (1) $\det(\lambda\mathbf{I} - \mathbf{A}) = (\lambda + 1)^2(\lambda - 5)$, 对应 $\lambda_1 = -1$ 的线性无关的特征向量是 $(1, 0, -1)^\top$ 和 $(0, 1, -1)^\top$, 而对应 $\lambda_2 = 5$ 的特征向量是 $(1, 1, 1)^\top$, 所以 \mathbf{A} 为单纯矩阵, 可以化为对角矩阵, 于是得

$$\begin{aligned} f(\mathbf{A}) &= \frac{1}{\mathbf{A}} = \mathbf{T} \cdot \text{diag}\left(-1, -1, \frac{1}{5}\right) \mathbf{T}^{-1} \\ &= \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} -1 & & \\ & -1 & \\ & & \frac{1}{5} \end{pmatrix} \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \\ &= \begin{pmatrix} -\frac{3}{5} & \frac{2}{5} & \frac{2}{5} \\ \frac{2}{5} & -\frac{3}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{2}{5} & -\frac{3}{5} \end{pmatrix}. \end{aligned}$$

(2) 显然 $\det(\lambda\mathbf{I} - \mathbf{A}) = (\lambda - 1)^4$, 它没有 4 个线性无关的特征向量, 是一个亏损矩阵, 因此只能化为若尔当标准形

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \begin{pmatrix} 1 & 1 & & \\ & 1 & 1 & \\ & & 1 & 1 \\ & & & 1 \end{pmatrix},$$

其中

$$\mathbf{T} = \begin{pmatrix} 8 & 4 & 0 & 0 \\ 0 & 4 & -1 & 1 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{T}^{-1} = \begin{pmatrix} \frac{1}{8} & -\frac{1}{8} & -\frac{1}{16} & 0 \\ 0 & \frac{1}{4} & \frac{1}{8} & 0 \\ 0 & 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

令 $f(z) = \sqrt{z}$, 直接利用式(5.2.28)有

$$f(\mathbf{J}(1)) = \begin{pmatrix} f(1) & f'(1) & \frac{1}{2!}f''(1) & \frac{1}{3!}f'''(1) \\ & f(1) & f'(1) & \frac{1}{2!}f''(1) \\ & & f(1) & f'(1) \\ & & & f(1) \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \frac{1}{2} & -\frac{1}{8} & \frac{1}{16} \\ 0 & 1 & \frac{1}{2} & -\frac{1}{8} \\ 0 & 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

所以

$$f(\mathbf{A}) = \sqrt{\mathbf{A}} = \mathbf{T} \cdot \text{diag}(f(\mathbf{J}(1)))\mathbf{T}^{-1}$$

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

注意,按式(5.2.27)定义的矩阵函数 $f(\mathbf{A})$ 与非奇异矩阵 \mathbf{T} 的选择无关.

5.3 函数矩阵的微分和积分

当矩阵的元素是函数时,对矩阵就可以引进微积分运算.例如在研究线性系统的可控制等方面的问题时,这种运算就是必要的了.这里研究的微积分法,从本质上说只不过是用矩阵重新描述普通数学分析中微积分的若干结果.

5.3.1 函数矩阵对实变量的导数

定义 5.3.1 若矩阵 $\mathbf{A} = (a_{ij})$ 的诸元素 a_{ij} 均是变量 t 的函数,即

$$\mathbf{A}(t) = \begin{pmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t) \\ a_{21}(t) & a_{22}(t) & \cdots & a_{2n}(t) \\ \vdots & \vdots & & \vdots \\ a_{m1}(t) & a_{m2}(t) & \cdots & a_{mn}(t) \end{pmatrix},$$

则称 $\mathbf{A}(t)$ 为函数矩阵.推而广之,变量 t 还可以是向量,也可以是矩阵.

对于函数矩阵我们可以像普通函数那样,引入极限、连续、微分和积分等概念.先看极限与连续.

如果所有的元素 $a_{ij}(t)$ 在 $t=t_0$ 时,存在极限,即有

$$\lim_{t \rightarrow t_0} a_{ij}(t) = a_{ij}, \quad a_{ij} \text{ 为一常数,}$$

则称矩阵 $\mathbf{A}(t)$ 有极限, 且极限值为 \mathbf{A} (常量矩阵), 即

$$\lim_{t \rightarrow t_0} \mathbf{A}(t) = \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

一个函数矩阵的极限, 具有通常函数极限的相似性质. 例如, 当 $t \rightarrow t_0$ 时, 函数矩阵 $\mathbf{A}(t)$ 和 $\mathbf{B}(t)$ 有极限 \mathbf{A} 和 \mathbf{B} , 则有

$$\lim_{t \rightarrow t_0} [\mathbf{A}(t) + \mathbf{B}(t)] = \mathbf{A} + \mathbf{B},$$

$$\lim_{t \rightarrow t_0} [\mathbf{A}(t)\mathbf{B}(t)] = \mathbf{A}\mathbf{B},$$

$$\lim_{t \rightarrow t_0} k\mathbf{A}(t) = k\mathbf{A},$$

其中 \mathbf{A}, \mathbf{B} 均为常量矩阵, k 为常数.

如果所有函数 $a_{ij}(t)$ 在某一点或某一区间上是连续的, 则称此函数矩阵在此点或在此区间上也是连续的.

对于多变量的函数矩阵, 也可以有与上述类似的规定, 这里就不一一重复了.

定义 5.3.2 设 $\mathbf{A}(t) = (a_{ij}(t))_{m \times n}$, 若 $a_{ij}(t) (i=1, 2, \dots, m; j=1, 2, \dots, n)$ 在 $t=t_0$ 处 (或 $[a, b]$ 上) 可导, 则称 $\mathbf{A}(t)$ 在点 $t=t_0$ 处 (或在 $[a, b]$ 上) 可导, 且记为

$$\begin{aligned} \mathbf{A}'(t_0) &= \left. \frac{d\mathbf{A}(t)}{dt} \right|_{t=t_0} = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{A}(t_0 + \Delta t) - \mathbf{A}(t_0)}{\Delta t} \\ &= \begin{pmatrix} a'_{11}(t_0) & a'_{12}(t_0) & \cdots & a'_{1n}(t_0) \\ a'_{21}(t_0) & a'_{22}(t_0) & \cdots & a'_{2n}(t_0) \\ \vdots & \vdots & & \vdots \\ a'_{m1}(t_0) & a'_{m2}(t_0) & \cdots & a'_{mn}(t_0) \end{pmatrix}_{m \times n}. \end{aligned}$$

不难证明函数矩阵的导数运算有下列性质:

(1) $\mathbf{A}(t)$ 为常数矩阵的充分必要条件是 $\mathbf{A}'(t) = \mathbf{0}$;

(2) 设 $\mathbf{A}(t) = (a_{ij}(t))_{m \times n}$ 与 $\mathbf{B}(t) = (b_{ij}(t))_{m \times n}$ 可导, 则

$$\frac{d}{dt} (\mathbf{A}(t) \pm \mathbf{B}(t)) = \mathbf{A}'(t) \pm \mathbf{B}'(t); \quad (5.3.1)$$

(3) 若 $k(t)$ 是可导的实函数, $\mathbf{A}(t)$ 可导, 则

$$\frac{d}{dt} (k(t)\mathbf{A}(t)) = k'(t)\mathbf{A}(t) + k(t)\mathbf{A}'(t); \quad (5.3.2)$$

(4) 设 $\mathbf{A}(t)$ 与 $\mathbf{B}(t)$ 都可导, 则

$$\frac{d}{dt} (\mathbf{A}(t)\mathbf{B}(t)) = \mathbf{A}'(t)\mathbf{B}(t) + \mathbf{A}(t)\mathbf{B}'(t); \quad (5.3.3)$$

(5) 若 $\mathbf{A}(t)$ 与 $\mathbf{A}^{-1}(t)$ 都有导数, 则

$$\frac{d\mathbf{A}^{-1}(t)}{dt} = -\mathbf{A}^{-1}(t)\mathbf{A}'(t)\mathbf{A}^{-1}(t); \quad (5.3.4)$$

(6) 设函数矩阵 $\mathbf{A}(t)$ 是 t 的函数, 而 $t=f(x)$ 是 x 的实值函数. 且 $\mathbf{A}(t)$ 与 $f(x)$ 均可导, 则有

$$\frac{d\mathbf{A}(t)}{dx} = \frac{d\mathbf{A}(t)}{dt} f'(x) = f'(x) \frac{d\mathbf{A}(t)}{dt}. \quad (5.3.5)$$

仅证明式(5.3.3)和式(5.3.4), 其余容易验证.

由导数定义, 有

$$\begin{aligned} \frac{d\mathbf{A}\mathbf{B}}{dt} &= \lim_{\Delta t \rightarrow 0} \frac{\mathbf{A}(t+\Delta t)\mathbf{B}(t+\Delta t) - \mathbf{A}(t)\mathbf{B}(t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\mathbf{A}(t+\Delta t)\mathbf{B}(t+\Delta t) - \mathbf{A}(t)\mathbf{B}(t+\Delta t) + \mathbf{A}(t)\mathbf{B}(t+\Delta t) - \mathbf{A}(t)\mathbf{B}(t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{[\mathbf{A}(t+\Delta t) - \mathbf{A}(t)]\mathbf{B}(t+\Delta t)}{\Delta t} + \lim_{\Delta t \rightarrow 0} \frac{\mathbf{A}(t)[\mathbf{B}(t+\Delta t) - \mathbf{B}(t)]}{\Delta t} \\ &= \frac{d\mathbf{A}}{dt}\mathbf{B} + \mathbf{A} \frac{d\mathbf{B}}{dt}. \end{aligned}$$

式(5.3.4)的证明, 只须注意到 $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$, 将它两端对 t 求导, 并应用公式(5.3.3)得

$$\frac{d\mathbf{A}}{dt}\mathbf{A}^{-1} + \mathbf{A} \frac{d\mathbf{A}^{-1}}{dt} = \mathbf{0} \quad (\text{零矩阵}).$$

将上式左端第一项移至等式右端, 然后等式两端左乘以 \mathbf{A}^{-1} 即得式(5.3.4).

注意, 式(5.3.3)的交换律一般不成立. 例如

$$\begin{aligned} \frac{d}{dt}\mathbf{A}^2(t) &= \frac{d}{dt}(\mathbf{A}(t)\mathbf{A}(t)) \\ &= \mathbf{A}'(t)\mathbf{A}(t) + \mathbf{A}(t)\mathbf{A}'(t) \\ &\neq 2\mathbf{A}(t)\mathbf{A}'(t). \end{aligned}$$

函数矩阵的导数本身也是一个函数矩阵, 还可以再进行导数运算, 故可以定义函数矩阵对实变量的高阶导数:

$$\frac{d^k \mathbf{A}(t)}{dt^k} = \frac{d}{dt} \left(\frac{d^{k-1} \mathbf{A}(t)}{dt^{k-1}} \right), \quad k = 1, 2, \dots, n.$$

例 5.3.1 不论 \mathbf{A} 是任何常量方阵, 总有

$$(1) \quad \frac{d}{dt} e^{\mathbf{A}t} = \mathbf{A}e^{\mathbf{A}t} = e^{\mathbf{A}t}\mathbf{A}; \quad (5.3.6)$$

$$(2) \quad \frac{d}{dt} \cos \mathbf{A}t = -\mathbf{A}(\sin \mathbf{A}t) = -(\sin \mathbf{A}t)\mathbf{A}; \quad (5.3.7)$$

$$(3) \quad \frac{d}{dt} \sin \mathbf{A}t = \mathbf{A}(\cos \mathbf{A}t) = (\cos \mathbf{A}t)\mathbf{A}. \quad (5.3.8)$$

证明 这里只证式(5.3.6), 而式(5.3.7), 式(5.3.8)的证明完全类似, 故不赘述. 为证式(5.3.6), 首先注意

$$(e^{\mathbf{A}t})_{ij} = \sum_{k=0}^{+\infty} \frac{1}{k!} t^k (\mathbf{A}^k)_{ij}.$$

上式右边是 t 的幂级数, 不管 t 值如何, 它总是收敛的. 因此, 逐项微分, 有

$$\frac{d}{dt} ((e^{\mathbf{A}t})_{ij}) = \sum_{k=1}^{+\infty} \frac{1}{(k-1)!} t^{k-1} (\mathbf{A}^k)_{ij}.$$

于是,有

$$\begin{aligned} \frac{d}{dt} e^{At} &= \sum_{k=1}^{+\infty} \frac{1}{(k-1)!} t^{k-1} A^k \\ &= \begin{cases} A \sum_{k=1}^{+\infty} \frac{1}{(k-1)!} t^{k-1} A^{k-1} = A e^{At}, \\ \left(\sum_{k=1}^{+\infty} \frac{1}{(k-1)!} t^{k-1} A^{k-1} \right) A = e^{At} A. \end{cases} \end{aligned}$$

证毕

例 5.3.2 设

$$A(t) = \begin{bmatrix} 1 & t^2 \\ t & 0 \end{bmatrix},$$

试计算 $\frac{d^2 A}{dt^2}$, $\frac{dA^{-1}}{dt}$.

解 显然有

$$\frac{dA}{dt} = \begin{bmatrix} 0 & 2t \\ 1 & 0 \end{bmatrix},$$

$$\frac{d^2 A}{dt^2} = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}.$$

要计算 $\frac{dA^{-1}}{dt}$, 必须先求 $A^{-1}(t)$:

$$\begin{aligned} A^{-1} &= \frac{1}{\begin{vmatrix} 1 & t^2 \\ t & 0 \end{vmatrix}} \begin{bmatrix} 0 & -t^2 \\ -t & 1 \end{bmatrix} = -\frac{1}{t^3} \begin{bmatrix} 0 & -t^2 \\ -t & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & \frac{1}{t} \\ \frac{1}{t^2} & -\frac{1}{t^3} \end{bmatrix}. \end{aligned}$$

所以

$$\begin{aligned} \frac{dA^{-1}}{dt} &= - \begin{bmatrix} 0 & \frac{1}{t} \\ \frac{1}{t^2} & -\frac{1}{t^3} \end{bmatrix} \begin{bmatrix} 0 & 2t \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & \frac{1}{t} \\ \frac{1}{t^2} & -\frac{1}{t^3} \end{bmatrix} \\ &= \begin{bmatrix} 0 & -\frac{1}{t^2} \\ -\frac{2}{t^3} & \frac{3}{t^4} \end{bmatrix}. \end{aligned}$$

例 5.3.3 设 $A = (a_{ij}(t))_{m \times n}$, 试证

$$\frac{d}{dt} \operatorname{tr} A = \operatorname{tr} \frac{dA}{dt}. \quad (5.3.9)$$

式(5.3.9)表明迹 tr 和 $\frac{d}{dt}$ 两种运算, 对方阵来说可以交换次序.

证明 因为 $\text{tr}\mathbf{A} = a_{11}(t) + a_{22}(t) + \cdots + a_{nn}(t)$, 所以

$$\frac{d}{dt}\text{tr}\mathbf{A} = \frac{d}{dt}a_{11} + \frac{d}{dt}a_{22} + \cdots + \frac{d}{dt}a_{nn},$$

另外, 由于

$$\frac{d\mathbf{A}}{dt} = \begin{pmatrix} \frac{da_{11}}{dt} & \cdots & \frac{da_{1n}}{dt} \\ \vdots & & \vdots \\ \frac{da_{n1}}{dt} & \cdots & \frac{da_{nn}}{dt} \end{pmatrix},$$

所以

$$\text{tr} \frac{d\mathbf{A}}{dt} = \frac{da_{11}}{dt} + \frac{da_{22}}{dt} + \cdots + \frac{da_{nn}}{dt} = \frac{d}{dt}\text{tr}\mathbf{A}. \quad \text{证毕}$$

5.3.2 函数矩阵特殊的导数

在自动控制的理论及一些实际问题中, 还经常遇到矩阵的特殊的导数问题. 常见的有数量函数对于向量、向量对于向量、矩阵对于向量以及矩阵对于矩阵的导数(微商)等, 现分述如下.

1. 数量函数对于向量的导数

在场论中, 我们对数量函数 $f(x, y, z)$ 定义梯度为

$$\text{grad}f = \nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right).$$

这可以理解为数量函数 $f(x, y, z)$ 对向量 (x, y, z) 的导数. 下面我们将这一概念推广到一般情形.

定义 5.3.3 设 $\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$, $f(\mathbf{x}) = f(x_1, x_2, \cdots, x_n)$ 是以向量 \mathbf{x} 为自变量的数量函数, 即为 n 元函数, 则规定数量函数 $f(\mathbf{x})$ 对于向量 \mathbf{x} 的导数为

$$\frac{df}{d\mathbf{x}} = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n} \right)^T. \quad (5.3.10)$$

显然, 若还有向量 \mathbf{x} 的数量函数

$$h(\mathbf{x}) = h(x_1, x_2, \cdots, x_n),$$

则下列导数法则成立:

$$\frac{d[f(\mathbf{x}) \pm h(\mathbf{x})]}{d\mathbf{x}} = \frac{df(\mathbf{x})}{d\mathbf{x}} \pm \frac{dh(\mathbf{x})}{d\mathbf{x}},$$

$$\frac{df(\mathbf{x})h(\mathbf{x})}{d\mathbf{x}} = \frac{df(\mathbf{x})}{d\mathbf{x}}h(\mathbf{x}) + f(\mathbf{x})\frac{dh(\mathbf{x})}{d\mathbf{x}}.$$

例 5.3.4 设 $\mathbf{A} = (a_{ij})_{n \times n}$ 为常量矩阵, $\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$. 证明数量函数 $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ 对于向量 \mathbf{x} 的导数为

$$\frac{df}{d\mathbf{x}} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x}. \quad (5.3.11)$$

证明 因为 $\mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i,j=1}^n a_{ij} x_i x_j$, 由式(5.3.10)得

$$\begin{aligned}
 \frac{df}{dx} &= \left(\frac{\partial}{\partial x_1} \sum_{i,j=1}^n a_{ij} x_i x_j, \dots, \frac{\partial}{\partial x_n} \sum_{i,j=1}^n a_{ij} x_i x_j \right)^T \\
 &= \begin{pmatrix} 2a_{11}x_1 + (a_{12} + a_{21})x_2 + \dots + (a_{1n} + a_{n1})x_n \\ \vdots \\ (a_{n1} + a_{n1})x_1 + (a_{n2} + a_{2n})x_2 + \dots + 2a_{nn}x_n \end{pmatrix} \\
 &= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n \end{pmatrix} + \begin{pmatrix} a_{11}x_1 + a_{21}x_2 + \dots + a_{n1}x_n \\ \vdots \\ a_{1n}x_1 + a_{2n}x_2 + \dots + a_{nn}x_n \end{pmatrix} \\
 &= \mathbf{Ax} + \mathbf{A}^T \mathbf{x} = (\mathbf{A} + \mathbf{A}^T) \mathbf{x}.
 \end{aligned}$$

证毕

作为特例:

(1) 当 \mathbf{A} 是实对称矩阵时, 二次型 $\mathbf{x}^T \mathbf{Ax}$ 对 \mathbf{x} 的导数为

$$\frac{d\mathbf{x}^T \mathbf{Ax}}{d\mathbf{x}} = 2\mathbf{Ax}; \quad (5.3.12)$$

(2) $\mathbf{A} = \mathbf{I}$ 时, 函数 $f(\mathbf{x}) = x_1^2 + x_2^2 + \dots + x_n^2$ 对 \mathbf{x} 的导数为

$$\frac{df(\mathbf{x})}{d\mathbf{x}} = 2\mathbf{x}. \quad (5.3.13)$$

例 5.3.5 令 $\mathbf{x} = (\xi_1(t), \xi_2(t), \dots, \xi_n(t))^T$, $f(\mathbf{x}) = f(\xi_1, \dots, \xi_n)$. 证明

$$\frac{df}{dt} = \left(\frac{df}{d\mathbf{x}} \right)^T \frac{d\mathbf{x}}{dt}. \quad (5.3.14)$$

证明 由偏导数的链式法则有

$$\begin{aligned}
 \frac{df}{dt} &= \frac{\partial f}{\partial \xi_1} \frac{d\xi_1}{dt} + \frac{\partial f}{\partial \xi_2} \frac{d\xi_2}{dt} + \dots + \frac{\partial f}{\partial \xi_n} \frac{d\xi_n}{dt} \\
 &= \left(\frac{\partial f}{\partial \xi_1}, \frac{\partial f}{\partial \xi_2}, \dots, \frac{\partial f}{\partial \xi_n} \right) \left(\frac{d\xi_1}{dt}, \frac{d\xi_2}{dt}, \dots, \frac{d\xi_n}{dt} \right)^T \\
 &= \left(\frac{df}{d\mathbf{x}} \right)^T \frac{d\mathbf{x}}{dt}.
 \end{aligned}$$

证毕

由于向量是特殊的矩阵, 因此, 我们可以定义数量函数对于矩阵的导数.

定义 5.3.4 设 $\mathbf{A} \in \mathbb{R}^{m \times n}$, $f(\mathbf{A})$ 为矩阵 \mathbf{A} 的数量函数, 即看成是 $m \times n$ 元函数, 则规定数量函数 $f(\mathbf{A})$ 对于矩阵 \mathbf{A} 的导数为

$$\frac{df}{d\mathbf{A}} = \left(\frac{\partial f}{\partial a_{ij}} \right)_{m \times n} = \begin{pmatrix} \frac{\partial f}{\partial a_{11}} & \dots & \frac{\partial f}{\partial a_{1n}} \\ \vdots & & \vdots \\ \frac{\partial f}{\partial a_{m1}} & \dots & \frac{\partial f}{\partial a_{mm}} \end{pmatrix}. \quad (5.3.15)$$

例 5.3.6 求二次型 $\mathbf{x}^T \mathbf{Ax}$ 对矩阵 \mathbf{A} 的导数, 其中 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 是实对称的.

解
$$\frac{d}{d\mathbf{A}} (\mathbf{x}^T \mathbf{Ax}) = \left(\frac{\partial}{\partial a_{ij}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \right)_{n \times n} = (x_i x_j)_{n \times n} = \mathbf{xx}^T.$$

例 5.3.7 设 $\mathbf{X} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$,

$$F(\mathbf{X}) = a^2 + b^2 + c^2 + d^2 - 2e + 15f,$$

求 $\frac{dF}{d\mathbf{X}}$.

解
$$\frac{dF}{d\mathbf{X}} = \begin{pmatrix} \frac{\partial F}{\partial a} & \frac{\partial F}{\partial b} & \frac{\partial F}{\partial c} \\ \frac{\partial F}{\partial d} & \frac{\partial F}{\partial e} & \frac{\partial F}{\partial f} \end{pmatrix} = \begin{bmatrix} 2a & 2b & 2c \\ 2d & -2 & 15 \end{bmatrix}.$$

例 5.3.8 设 $\mathbf{X} = \begin{bmatrix} x & y \\ u & v \end{bmatrix}$, 证明 $\left| \frac{d|\mathbf{X}|}{d\mathbf{X}} \right| = |\mathbf{X}|$.

证明 因为 $|\mathbf{X}| = xv - uy$,

$$\frac{d|\mathbf{X}|}{d\mathbf{X}} = \begin{pmatrix} \frac{\partial |\mathbf{X}|}{\partial x} & \frac{\partial |\mathbf{X}|}{\partial y} \\ \frac{\partial |\mathbf{X}|}{\partial u} & \frac{\partial |\mathbf{X}|}{\partial v} \end{pmatrix} = \begin{bmatrix} v & -u \\ -y & x \end{bmatrix},$$

所以

$$\left| \frac{d|\mathbf{X}|}{d\mathbf{X}} \right| = \begin{vmatrix} v & -u \\ -y & x \end{vmatrix} = xv - uy = |\mathbf{X}|.$$

证毕

例 5.3.9 设 $\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}$, $f(\mathbf{X}) = \text{tr}\mathbf{X} = x_{11} + x_{22}$, 求 $\frac{df}{d\mathbf{X}}$.

解
$$\frac{df}{d\mathbf{X}} = \frac{d}{d\mathbf{X}}(\text{tr}\mathbf{X}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}.$$

例 5.3.10 设 $\mathbf{X} = (x_{ij})_{n \times n}$, 求 $\frac{d}{d\mathbf{X}}(\text{tr}(\mathbf{X}\mathbf{X}^T))$ 及 $\frac{d}{d\mathbf{X}}(\text{tr}\mathbf{X}^2)$.

解 因为

$$\frac{\partial}{\partial x_{ij}}(\text{tr}(\mathbf{X}\mathbf{X}^T)) = \frac{\partial}{\partial x_{ij}} \left(\sum_{i=1}^n \sum_{j=1}^n x_{ij}^2 \right) = 2x_{ij},$$

所以

$$\frac{d}{d\mathbf{X}}(\text{tr}(\mathbf{X}\mathbf{X}^T)) = 2\mathbf{X}.$$

又因为

$$\begin{aligned} \frac{\partial}{\partial x_{ij}}(\text{tr}\mathbf{X}^2) &= \text{tr} \left(\frac{\partial}{\partial x_{ij}}(\mathbf{X}^2) \right) = \text{tr} \left(2\mathbf{X} \frac{\partial \mathbf{X}}{\partial x_{ij}} \right) \\ &= 2\text{tr}(\mathbf{X}\mathbf{E}_{ij}) = 2x_{ji}, \end{aligned}$$

这里 \mathbf{E}_{ij} 表示第 i 行第 j 列的元素为 1 而其余元素为 0 的矩阵, 故

$$\frac{d}{d\mathbf{X}}(\text{tr}\mathbf{X}^2) = 2\mathbf{X}^T.$$

2. 矩阵对于矩阵的导数

由于向量是特殊的矩阵, 所以向量对于向量的导数、矩阵对于向量的导数等都可视作矩阵对于矩阵的导数的特殊情况来计算.

定义 5.3.5 设矩阵 \mathbf{F} 是以 $\mathbf{A} \in \mathbb{C}^{m \times n}$ 为自变量的 $p \times q$ 矩阵, 即

$$\mathbf{F}(\mathbf{A}) = \begin{pmatrix} f_{11}(\mathbf{A}) & f_{12}(\mathbf{A}) & \cdots & f_{1q}(\mathbf{A}) \\ f_{21}(\mathbf{A}) & f_{22}(\mathbf{A}) & \cdots & f_{2q}(\mathbf{A}) \\ \vdots & \vdots & & \vdots \\ f_{p1}(\mathbf{A}) & f_{p2}(\mathbf{A}) & \cdots & f_{pq}(\mathbf{A}) \end{pmatrix}_{p \times q},$$

其元素 $f_{ks}(\mathbf{A})$ 是以矩阵 $\mathbf{A} = (a_{ij})_{m \times n}$ 的元素为自变量的 mn 元函数, 则规定矩阵 $\mathbf{F}(\mathbf{A})$ 对于矩阵 \mathbf{A} 的导数为

$$\frac{d\mathbf{F}}{d\mathbf{A}} = \left(\frac{\partial \mathbf{F}}{\partial a_{ij}} \right)_{pm \times qn} = \begin{pmatrix} \frac{\partial \mathbf{F}}{\partial a_{11}} & \frac{\partial \mathbf{F}}{\partial a_{12}} & \cdots & \frac{\partial \mathbf{F}}{\partial a_{1n}} \\ \frac{\partial \mathbf{F}}{\partial a_{21}} & \frac{\partial \mathbf{F}}{\partial a_{22}} & \cdots & \frac{\partial \mathbf{F}}{\partial a_{2n}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial \mathbf{F}}{\partial a_{m1}} & \frac{\partial \mathbf{F}}{\partial a_{m2}} & \cdots & \frac{\partial \mathbf{F}}{\partial a_{mn}} \end{pmatrix}, \quad (5.3.16)$$

其中

$$\frac{\partial \mathbf{F}}{\partial a_{ij}} = \begin{pmatrix} \frac{\partial f_{11}}{\partial a_{ij}} & \frac{\partial f_{12}}{\partial a_{ij}} & \cdots & \frac{\partial f_{1q}}{\partial a_{ij}} \\ \frac{\partial f_{21}}{\partial a_{ij}} & \frac{\partial f_{22}}{\partial a_{ij}} & \cdots & \frac{\partial f_{2q}}{\partial a_{ij}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_{p1}}{\partial a_{ij}} & \frac{\partial f_{p2}}{\partial a_{ij}} & \cdots & \frac{\partial f_{pq}}{\partial a_{ij}} \end{pmatrix}, \quad \begin{matrix} i = 1, 2, \cdots, m, \\ j = 1, 2, \cdots, n. \end{matrix}$$

例 5.3.11 设 $\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$, 求向量 $\mathbf{x}^T = (x_1, x_2, \cdots, x_n)$ 对向量 \mathbf{x} 的导数.

解 由式(5.3.16)有

$$\frac{d\mathbf{x}^T}{d\mathbf{x}} = \begin{pmatrix} \frac{\partial \mathbf{x}^T}{\partial x_1} \\ \frac{\partial \mathbf{x}^T}{\partial x_2} \\ \vdots \\ \frac{\partial \mathbf{x}^T}{\partial x_n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \mathbf{I}.$$

例 5.3.12 设 $\mathbf{x} = (x_1, x_2, \cdots, x_n)$, $\mathbf{y} = (y_1, y_2, \cdots, y_n)^T = f(\mathbf{x})$, 其中

$$\begin{cases} y_1 = f_1(x_1, x_2, \cdots, x_n) = f_1(\mathbf{x}), \\ y_2 = f_2(x_1, x_2, \cdots, x_n) = f_2(\mathbf{x}), \\ \vdots \\ y_n = f_n(x_1, x_2, \cdots, x_n) = f_n(\mathbf{x}), \end{cases}$$

求 $\frac{d\mathbf{f}}{d\mathbf{x}}$.

解 由公式(5.3.16)有

$$\frac{df}{dx} = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}. \quad (5.3.17)$$

式(5.3.17)的行列式 $\det(f'(x))$ 称为雅可比(Jacobi)行列式.

例 5.3.13 求 $y^T A^T A y$ 对于 A 的导数, 其中 $A \in \mathbb{C}^{m \times n}$, $y \in \mathbb{C}^n$ 是常量列向量.

解

$$\begin{aligned} \frac{d(y^T A^T A y)}{dA} &= \left(\frac{\partial(y^T A^T A y)}{\partial a_{ij}} \right) \\ &= \left(\frac{\partial(y^T A^T)}{\partial a_{ij}} A y + y^T A^T \frac{\partial(A y)}{\partial a_{ij}} \right) \\ &= \left(y^T \frac{\partial A^T}{\partial a_{ij}} A y + y^T A^T \frac{\partial A}{\partial a_{ij}} y \right) \\ &= \left(\sum_{k=1}^n a_{ik} y_k y_j \right) + \left(\sum_{k=1}^n a_{ik} y_k y_j \right) \\ &= A y y^T + A y y^T = 2A y y^T. \end{aligned}$$

例 5.3.14 设 $f(x)$ 是向量 $x = (x_1, x_2, \dots, x_n)^T$ 的函数, 而 $x_i = x_i(u)$, $i = 1, 2, \dots, n$, $u = (u_1, u_2, \dots, u_n)^T$. 试证明

$$\begin{aligned} \frac{df}{du} &= \frac{dx^T}{du} \frac{df}{dx}. \\ \text{证明 } \frac{df}{du} &= \begin{pmatrix} \frac{\partial f}{\partial u_1} \\ \frac{\partial f}{\partial u_2} \\ \vdots \\ \frac{\partial f}{\partial u_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \frac{\partial x_1}{\partial u_1} + \frac{\partial f}{\partial x_2} \frac{\partial x_2}{\partial u_1} + \dots + \frac{\partial f}{\partial x_n} \frac{\partial x_n}{\partial u_1} \\ \frac{\partial f}{\partial x_1} \frac{\partial x_1}{\partial u_2} + \frac{\partial f}{\partial x_2} \frac{\partial x_2}{\partial u_2} + \dots + \frac{\partial f}{\partial x_n} \frac{\partial x_n}{\partial u_2} \\ \vdots \\ \frac{\partial f}{\partial x_1} \frac{\partial x_1}{\partial u_n} + \frac{\partial f}{\partial x_2} \frac{\partial x_2}{\partial u_n} + \dots + \frac{\partial f}{\partial x_n} \frac{\partial x_n}{\partial u_n} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_2}{\partial u_1} & \dots & \frac{\partial x_n}{\partial u_1} \\ \frac{\partial x_1}{\partial u_2} & \frac{\partial x_2}{\partial u_2} & \dots & \frac{\partial x_n}{\partial u_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial x_1}{\partial u_n} & \frac{\partial x_2}{\partial u_n} & \dots & \frac{\partial x_n}{\partial u_n} \end{pmatrix} \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} \\ &= \frac{dx^T}{du} \frac{df}{dx}. \end{aligned}$$

证毕

5.3.3 矩阵的全微分

定义 5.3.6 设矩阵 $F = (f_{ij})_{m \times n}$, 则规定矩阵 F 的全微分为

$$dF = (df_{ij})_{m \times n}.$$

例 5.3.15 设

$$\mathbf{F} = \begin{pmatrix} s+t & s^2-2t \\ 2s+t^3 & t^2 \end{pmatrix},$$

那么

$$d\mathbf{F} = \begin{pmatrix} ds+dt & 2sds-2dt \\ 2ds+3t^2dt & 2tdt \end{pmatrix}.$$

矩阵的全微分有如下的运算性质:

- (1) $d(\mathbf{F} \pm \mathbf{G}) = d\mathbf{F} \pm d\mathbf{G}$;
- (2) $d(k\mathbf{F}) = kd\mathbf{F}$;
- (3) 当 \mathbf{A} 是常量矩阵时, $d\mathbf{A} = \mathbf{0}$;
- (4) $d(\mathbf{X}^\top) = (d\mathbf{X})^\top$;
- (5) $d(\text{tr}\mathbf{X}) = \text{tr}(d\mathbf{X})$.

证明 (1), (2), (3) 是显然的. 下面证明(4). 设

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix},$$

那么

$$d(\mathbf{X}^\top) = \begin{pmatrix} dx_{11} & \cdots & dx_{m1} \\ \vdots & & \vdots \\ dx_{1n} & \cdots & dx_{mn} \end{pmatrix} = (d\mathbf{X})^\top.$$

再证(5). 性质(5)是说“d”与“tr”两种运算的次序可交换. 设

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nn} \end{pmatrix},$$

那么

$$\begin{aligned} \text{tr}\mathbf{X} &= x_{11} + x_{22} + \cdots + x_{nn}, \\ d(\text{tr}\mathbf{X}) &= dx_{11} + dx_{22} + \cdots + dx_{nn}, \\ d\mathbf{X} &= \begin{pmatrix} dx_{11} & \cdots & dx_{1n} \\ \vdots & & \vdots \\ dx_{n1} & \cdots & dx_{nn} \end{pmatrix}, \end{aligned}$$

$$\text{tr}(d\mathbf{X}) = dx_{11} + dx_{22} + \cdots + dx_{nn} = d(\text{tr}\mathbf{X}). \quad \text{证毕}$$

定理 5.3.1 设 $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top$, 矩阵 $\mathbf{F} = (f_{ij})_{s \times m}$, 其中 f_{ij} 都是 x_i 的实函数, 那么

$$d\mathbf{F} = \sum_{i=1}^n \frac{\partial \mathbf{F}}{\partial x_i} dx_i. \quad (5.3.18)$$

在证明定理之前, 首先, 我们指出矩阵 \mathbf{F} 中的元素 f_{ij} 都是 x_1, x_2, \dots, x_n 的 n 元函数; 其次, 注意到式(5.3.18)的右端表示矩阵 \mathbf{F} 对变量 x_i 的偏导数, 而左端是矩阵 \mathbf{F} 的全微分. 下面来证明这个定理.

证明

$$\begin{aligned}
 d\mathbf{F} &= \begin{pmatrix} df_{11} & \cdots & df_{1m} \\ \vdots & & \vdots \\ df_{s1} & \cdots & df_{sm} \end{pmatrix} \\
 &= \begin{pmatrix} \frac{\partial f_{11}}{\partial x_1} dx_1 & \cdots & \frac{\partial f_{1m}}{\partial x_1} dx_1 \\ \vdots & & \vdots \\ \frac{\partial f_{s1}}{\partial x_1} dx_1 & \cdots & \frac{\partial f_{sm}}{\partial x_1} dx_1 \end{pmatrix} + \cdots + \begin{pmatrix} \frac{\partial f_{11}}{\partial x_n} dx_n & \cdots & \frac{\partial f_{1m}}{\partial x_n} dx_n \\ \vdots & & \vdots \\ \frac{\partial f_{s1}}{\partial x_n} dx_n & \cdots & \frac{\partial f_{sm}}{\partial x_n} dx_n \end{pmatrix} \\
 &= \frac{\partial \mathbf{F}}{\partial x_1} dx_1 + \cdots + \frac{\partial \mathbf{F}}{\partial x_n} dx_n.
 \end{aligned}$$

证毕

定理 5.3.2 (1) 设 $\mathbf{A} = \mathbf{BC}$, 则 $d\mathbf{A} = (d\mathbf{B})\mathbf{C} + \mathbf{B}(d\mathbf{C})$;

(2) 设 $\mathbf{A} = \mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_r$, 则

$$d\mathbf{F} = (d\mathbf{A}_1)\mathbf{A}_2\cdots\mathbf{A}_r + \mathbf{A}_1(d\mathbf{A}_2)\mathbf{A}_3\cdots\mathbf{A}_r + \mathbf{A}_1\cdots\mathbf{A}_{r-1}(d\mathbf{A}_r).$$

证明 只要证明了(1), 那么(2)是显然的. 设

$$\mathbf{A} = (a_{ij})_{m \times n}, \quad \mathbf{B} = (b_{ij})_{m \times s}, \quad \mathbf{C} = (c_{ij})_{s \times n},$$

那么

$$\begin{aligned}
 da_{ij} &= [(db_{i1})c_{1j} + b_{i1}(dc_{1j})] + \cdots + [(db_{is})c_{sj} + b_{is}(dc_{sj})] \\
 &= [(db_{i1})c_{1j} + \cdots + (db_{is})c_{sj}] + [b_{i1}(dc_{1j}) + \cdots + b_{is}(dc_{sj})],
 \end{aligned}$$

所以

$$d\mathbf{F} = (d\mathbf{B})\mathbf{C} + \mathbf{B}(d\mathbf{C}).$$

证毕

推论 (1) $d(\boldsymbol{\alpha}^T \mathbf{x}) = \boldsymbol{\alpha}^T d\mathbf{x} = (d\mathbf{x})^T \boldsymbol{\alpha}$, 其中 $\boldsymbol{\alpha}$ 是 $n \times 1$ 常量矩阵, \mathbf{x} 是 n 维列向量.

(2) $d(\mathbf{A}\mathbf{x}) = \mathbf{A}d\mathbf{x}$.

(3) $d(\mathbf{x}^T \mathbf{A}\mathbf{x}) = \mathbf{x}^T (\mathbf{A}^T + \mathbf{A})d\mathbf{x}$, 其中 \mathbf{A} 是 $m \times n$ 常量矩阵, \mathbf{x} 是 n 维列向量.

证 (1) 由定理 5.3.2 知, $d(\boldsymbol{\alpha}^T \mathbf{x}) = (d\boldsymbol{\alpha}^T)\mathbf{x} + \boldsymbol{\alpha}^T(d\mathbf{x})$, 而 $d\boldsymbol{\alpha}^T = \mathbf{0}$, 即有 $d(\boldsymbol{\alpha}^T \mathbf{x}) = \boldsymbol{\alpha}^T d\mathbf{x}$.

又因为 $\boldsymbol{\alpha}^T d\mathbf{x}$ 是 1×1 矩阵, 所以 $\boldsymbol{\alpha}^T d\mathbf{x} = (\boldsymbol{\alpha}^T d\mathbf{x})^T = (d\mathbf{x})^T \boldsymbol{\alpha}$.

(2) 显然.

$$\begin{aligned}
 (3) \quad d(\mathbf{x}^T \mathbf{A}\mathbf{x}) &= (d\mathbf{x}^T)\mathbf{A}\mathbf{x} + \mathbf{x}^T(d\mathbf{A})\mathbf{x} + \mathbf{x}^T \mathbf{A}d\mathbf{x} \\
 &= (d\mathbf{x})^T \mathbf{A}\mathbf{x} + \mathbf{x}^T \mathbf{A}d\mathbf{x} = ((d\mathbf{x})^T \mathbf{A}\mathbf{x})^T + \mathbf{x}^T \mathbf{A}d\mathbf{x} \\
 &= \mathbf{x}^T \mathbf{A}^T d\mathbf{x} + \mathbf{x}^T \mathbf{A}d\mathbf{x} = \mathbf{x}^T (\mathbf{A}^T + \mathbf{A})d\mathbf{x}.
 \end{aligned}$$

证毕

5.3.4 函数矩阵的积分

定义 5.3.7 设函数矩阵

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

我们定义

$$\int \mathbf{A}(t) dt = \begin{pmatrix} \int a_{11}(t) dt & \int a_{12}(t) dt & \cdots & \int a_{1n}(t) dt \\ \vdots & \vdots & & \vdots \\ \int a_{n1}(t) dt & \int a_{n2}(t) dt & \cdots & \int a_{nn}(t) dt \end{pmatrix},$$

$$\int_a^b \mathbf{A}(t) dt = \begin{pmatrix} \int_a^b a_{11}(t) dt & \int_a^b a_{12}(t) dt & \cdots & \int_a^b a_{1n}(t) dt \\ \vdots & \vdots & & \vdots \\ \int_a^b a_{n1}(t) dt & \int_a^b a_{n2}(t) dt & \cdots & \int_a^b a_{nn}(t) dt \end{pmatrix},$$

这里显然假设积分 $\int a_{ij}(t) dt (i, j = 1, 2, \dots, n)$ 是存在的.

例 5.3.16 设

$$\mathbf{A}(t) = \begin{pmatrix} \frac{t^2}{EJ} & \frac{t}{EJ} \\ \frac{t}{EJ} & \frac{1}{EJ} \end{pmatrix},$$

其中 E, J 为力学常数, 则

$$\int_0^l \mathbf{A}(t) dt = \begin{pmatrix} \frac{l^3}{3EJ} & \frac{l^2}{2EJ} \\ \frac{l^2}{2EJ} & \frac{l}{EJ} \end{pmatrix}.$$

在微积分中, 大家熟知, 常数因子可以提到积分号外. 在矩阵积分中, 亦有此性质; 各元素与积分变量无关的矩阵可以提到积分号外, 但要注意不可改变左右次序, 即: 设矩阵 \mathbf{A} 的元素 a_{ij} 为区域 Ω 的函数, 而矩阵 \mathbf{B} 的元素 b_{ij} 与 Ω 无关, 则仿照上述积分的定义有

$$\int_{\Omega} \mathbf{A} \mathbf{B} d\Omega = \left(\int_{\Omega} \mathbf{A} d\Omega \right) \mathbf{B}, \quad (5.3.19)$$

$$\int_{\Omega} \mathbf{B} \mathbf{A} d\Omega = \mathbf{B} \int_{\Omega} \mathbf{A} d\Omega. \quad (5.3.20)$$

事实上, 式(5.3.19)的

$$\begin{aligned} \text{左端} \int_{\Omega} \mathbf{A} \mathbf{B} d\Omega &= \left(\int_{\Omega} \left(\sum_{k=1}^n a_{ik} b_{kj} \right) d\Omega \right)_{ij} \\ &= \left(\sum_{k=1}^n \left(\int_{\Omega} a_{ik} d\Omega \right) b_{kj} \right)_{ij} = \left(\int_{\Omega} \mathbf{A} d\Omega \right) \mathbf{B}. \end{aligned}$$

同理可证式(5.3.20).

证毕

5.4 矩阵微分方程

在线性控制系统中, 常常涉及求解线性微分方程组的问题. 利用矩阵函数表示线性微分方程组的解, 形式比较简单, 从而使得复杂的系统状态方程的求解问题得到简化.

$$\begin{aligned} \mathbf{x}'(0) &= \mathbf{A}\mathbf{x}(0), \\ \mathbf{x}''(0) &= \mathbf{A}^2\mathbf{x}(0), \\ \mathbf{x}'''(0) &= \mathbf{A}^3\mathbf{x}(0), \\ &\vdots \end{aligned}$$

所以

$$\begin{aligned} \mathbf{x} = \mathbf{x}(t) &= \mathbf{x}(0) + \mathbf{A}\mathbf{x}(0)t + \frac{1}{2!}\mathbf{A}^2\mathbf{x}(0)t^2 + \cdots \\ &= \left(\mathbf{I} + (\mathbf{A}t) + \frac{1}{2!}(\mathbf{A}t)^2 + \cdots \right) \mathbf{x}(0) \\ &= e^{\mathbf{A}t}\mathbf{x}(0). \end{aligned}$$

由此可见,微分方程组(5.4.1)在给定初始条件(5.4.2)下的解必定具有

$$\mathbf{x} = e^{\mathbf{A}t}\mathbf{x}(0)$$

的形式.下面我们来证明它确实是(5.4.1)的解.

事实上

$$\begin{aligned} \frac{d\mathbf{x}}{dt} &= \frac{d}{dt}(e^{\mathbf{A}t}\mathbf{x}(0)) = \left(\frac{d}{dt}e^{\mathbf{A}t} \right) \mathbf{x}(0) + e^{\mathbf{A}t} \frac{d\mathbf{x}(0)}{dt} \\ &= \mathbf{A}e^{\mathbf{A}t}\mathbf{x}(0) = \mathbf{A}\mathbf{x}. \end{aligned}$$

又当 $t=0$ 时, $\mathbf{x}(0) = e^0\mathbf{x}(0) = \mathbf{I}\mathbf{x}(0) = \mathbf{x}(0)$, 因此这个解是满足初始条件的. 这样, 我们实际上证明了下面的定理.

定理 5.4.1 一阶线性常系数微分方程组的定解问题

$$\begin{cases} \frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t), \\ \mathbf{x}(0) = (x_1(0), x_2(0), \cdots, x_n(0))^T, \end{cases} \quad (5.4.3)$$

有惟一解 $\mathbf{x} = e^{\mathbf{A}t}\mathbf{x}(0)$.

同理可证定解问题

$$\begin{cases} \frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t), \\ \mathbf{x}|_{t=t_0} = \mathbf{x}(t_0) \end{cases} \quad (5.4.4)$$

的惟一解是

$$\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0).$$

若未知函数 $\mathbf{X}(t)$ 不是列向量, 而是 $n \times m$ 矩阵

$$\mathbf{X}(t) = \begin{pmatrix} x_{11}(t) & x_{12}(t) & \cdots & x_{1m}(t) \\ x_{21}(t) & x_{22}(t) & \cdots & x_{2m}(t) \\ \vdots & \vdots & & \vdots \\ x_{n1}(t) & x_{n2}(t) & \cdots & x_{nm}(t) \end{pmatrix},$$

则方程

$$\frac{d\mathbf{X}}{dt} = \mathbf{A}\mathbf{X}(t)$$

就是 $n \times m$ 个未知函数的线性微分方程组. 下面将讨论该方程组满足初始条件

$$\mathbf{X}(t) \Big|_{t=t_0} = \mathbf{X}(t_0) = \begin{pmatrix} x_{11}(t_0) & x_{12}(t_0) & \cdots & x_{1m}(t_0) \\ x_{21}(t_0) & x_{22}(t_0) & \cdots & x_{2m}(t_0) \\ \vdots & \vdots & & \vdots \\ x_{n1}(t_0) & x_{n2}(t_0) & \cdots & x_{nm}(t_0) \end{pmatrix}$$

的定解问题.

定理 5.4.2 设定解问题为

$$\begin{cases} \frac{d\mathbf{X}}{dt} = \mathbf{A}\mathbf{X}(t), \\ \mathbf{X}(t) \Big|_{t=t_0} = \mathbf{X}(t_0), \end{cases} \quad (5.4.5)$$

其中 $\mathbf{X}(t)$ 是 t 的可微函数的 $n \times m$ 矩阵, $\mathbf{X}(t_0)$ 是 $n \times m$ 常数矩阵, \mathbf{A} 是给定的 n 阶常数方阵, 则

(1) 定解问题(5.4.5)的解为

$$\mathbf{X}(t) = e^{\mathbf{A}(t-t_0)} \mathbf{X}(t_0),$$

并且这个解是惟一的;

(2) 解 $\mathbf{X}(t)$ 的秩与 t 的取值无关.

证明 (1) 因为

$$\begin{aligned} \frac{d\mathbf{X}(t)}{dt} &= \frac{d}{dt} (e^{\mathbf{A}(t-t_0)} \mathbf{X}(t_0)) \\ &= \mathbf{A} e^{\mathbf{A}(t-t_0)} \mathbf{X}(t_0) \\ &= \mathbf{A}\mathbf{X}(t), \end{aligned}$$

且

$$\begin{aligned} \mathbf{X}(t) \Big|_{t=t_0} &= (e^{\mathbf{A}(t-t_0)} \mathbf{X}(t_0)) \Big|_{t=t_0} \\ &= \mathbf{I}\mathbf{X}(t_0) = \mathbf{X}(t_0), \end{aligned}$$

因此 $\mathbf{X}(t) = e^{\mathbf{A}(t-t_0)} \mathbf{X}(t_0)$ 是式(5.4.5)的解.

若 $\mathbf{V}(t)$ 也是式(5.4.5)的解, 令

$$\mathbf{Y}(t) = e^{-\mathbf{A}t} \mathbf{V}(t),$$

则

$$\begin{aligned} \frac{d\mathbf{Y}(t)}{dt} &= e^{-\mathbf{A}t} (-\mathbf{A})\mathbf{V}(t) + e^{-\mathbf{A}t} \frac{d\mathbf{V}(t)}{dt} \\ &= -e^{-\mathbf{A}t} \mathbf{A}\mathbf{V}(t) + e^{-\mathbf{A}t} \mathbf{A}\mathbf{V}(t) \\ &= \mathbf{0}, \end{aligned}$$

故对 $\forall t \in \mathbb{C}$, 有

$$\mathbf{Y}(t) = k,$$

从而

$$\mathbf{Y}(t_0) = k.$$

由初始条件知

$$\mathbf{Y}(t_0) = e^{-\mathbf{A}t_0} \mathbf{V}(t_0) = e^{-\mathbf{A}t_0} \mathbf{X}(t_0),$$

故

$$e^{-\mathbf{A}t_0} \mathbf{X}(t_0) = k.$$

从而

$$\mathbf{Y}(t) = e^{-At}\mathbf{V}(t) = k = e^{-At_0}\mathbf{X}(t_0),$$

因此

$$\mathbf{V}(t) = e^{A(t-t_0)}\mathbf{X}(t_0),$$

即式(5.4.5)的解是惟一的.

(2) 由于 $e^{A(t-t_0)}$ 对于任何的 t 都是可逆的矩阵, 因而 $\mathbf{X}(t)$ 的秩与 $\mathbf{X}(t_0)$ 的秩相同, 故 $\mathbf{X}(t)$ 与 $\mathbf{X}(t_0)$ 的秩与 t 的取值无关.

以上定理表明: 齐次线性方程组(5.4.3)~(5.4.5)的定解问题与普通齐次微分方程的定解问题具有相同的简单形式的解, 这样为解决复杂的定常系统状态方程求解问题带来了极大的方便. 但是, 要注意的是, 这里的解是用矩阵函数表示的, 而矩阵乘法不满足交换律, 所以初始矩阵 $\mathbf{X}(t_0)$ 必须右乘.

例 5.4.1 求定解问题

$$\begin{cases} \frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t), \\ \mathbf{x}(0) = (1, 1, 1)^T, \end{cases} \quad \mathbf{A} = \begin{pmatrix} 3 & -1 & 1 \\ 2 & 0 & -1 \\ 1 & -1 & 2 \end{pmatrix}$$

的解

$$\text{解} \quad |\lambda\mathbf{I} - \mathbf{A}| = \begin{vmatrix} \lambda - 3 & 1 & -1 \\ -2 & \lambda & 1 \\ -1 & 1 & \lambda - 2 \end{vmatrix} = \lambda(\lambda - 2)(\lambda - 3),$$

故 \mathbf{A} 有 3 个不同的特征值, 从而 \mathbf{A} 可与对角形矩阵相似. 与特征值 $\lambda_1 = 0, \lambda_2 = 2, \lambda_3 = 3$ 相应的 3 个线性无关的特征向量为

$$\mathbf{x}_1 = (1, 5, 2)^T, \quad \mathbf{x}_2 = (1, 1, 0)^T, \quad \mathbf{x}_3 = (2, 1, 1)^T.$$

故得

$$\mathbf{P} = \begin{pmatrix} 1 & 1 & 2 \\ 5 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix} \quad \text{及} \quad \mathbf{P}^{-1} = -\frac{1}{6} \begin{pmatrix} 1 & -1 & -1 \\ -3 & -3 & 9 \\ -2 & 2 & -4 \end{pmatrix},$$

所以, 由定理 5.4.1 可得所求的解为(见式(5.2.27)).

$$\begin{aligned} \mathbf{x} &= e^{At}\mathbf{x}(0) = \mathbf{P} \begin{pmatrix} 1 \\ e^{2t} \\ e^{3t} \end{pmatrix} \mathbf{P}^{-1}\mathbf{x}(0) \\ &= \begin{pmatrix} 1 & 1 & 2 \\ 5 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ e^{2t} \\ e^{3t} \end{pmatrix} \left(-\frac{1}{6}\right) \begin{pmatrix} 1 & -1 & -1 \\ -3 & -3 & 9 \\ -2 & 2 & -4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ &= -\frac{1}{6} \begin{pmatrix} -1 + 3e^{2t} - 8e^{3t} \\ -5 + 3e^{2t} - 4e^{3t} \\ -2 - 4e^{3t} \end{pmatrix}. \end{aligned}$$

例 5.4.2 求常系数齐次线性微分方程组

$$\begin{cases} \frac{dx_1(t)}{dt} = 2x_1 + 2x_2 - x_3, \\ \frac{dx_2(t)}{dt} = -x_1 - x_2 + x_3, \\ \frac{dx_3(t)}{dt} = -x_1 - 2x_2 + 2x_3, \end{cases}$$

在初始条件

$$\mathbf{x}(0) = \begin{pmatrix} x_1(0) \\ x_2(0) \\ x_3(0) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}$$

下的解.

解 系数矩阵为

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

从而定解问题的解为

$$\mathbf{x}(t) = e^{At} \mathbf{x}(0),$$

下面求 e^{At} .

由于 \mathbf{A} 的特征值 $\lambda=1$ 为 3 重根, 此时无 3 个线性无关的特征向量, 所以 \mathbf{A} 为亏损矩阵, 不能对角化. 但易求出 \mathbf{A} 的若尔当标准形及相似变换矩阵分别为

$$\mathbf{J} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix},$$

从而

$$e^{At} = \mathbf{P} e^{Jt} \mathbf{P}^{-1} = \mathbf{P} \begin{pmatrix} e^t & te^t & 0 \\ 0 & e^t & 0 \\ 0 & 0 & e^t \end{pmatrix} \mathbf{P}^{-1},$$

故定解问题的解为

$$\mathbf{y}(t) = e^{At} \mathbf{x}(0) = \begin{pmatrix} e^t \\ e^t \\ 3e^t \end{pmatrix}.$$

5.4.2 常系数非齐次线性微分方程组的解

考虑一阶常系数非齐次线性微分方程组的定解问题

$$\begin{cases} \frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} + \mathbf{f}(t), \\ \mathbf{x}|_{t=t_0} = \mathbf{x}(t_0). \end{cases} \quad (5.4.6)$$

这里 $f(t) = (f_1(t), f_2(t), \dots, f_n(t))^T$ 是已知向量函数, A 及 x 意义如前面所述. 改写方程为

$$\frac{dx}{dt} - Ax = f(t),$$

并以 e^{-At} 左乘方程两边, 即得

$$e^{-At} \left(\frac{dx}{dt} - Ax \right) = e^{-At} f(t),$$

即

$$\frac{d}{dt} (e^{-At} x) = e^{-At} f(t).$$

在 $[t_0, t]$ 上求积分, 可得

$$e^{-At} x - e^{-At_0} x(t_0) = \int_{t_0}^t e^{-A\tau} f(\tau) d\tau,$$

即

$$x = e^{A(t-t_0)} x(t_0) + \int_{t_0}^t e^{A(t-\tau)} f(\tau) d\tau.$$

它就是我们要考虑的定解问题的解.

例 5.4.3 求定解问题

$$\begin{cases} \frac{dx}{dt} = Ax + f(t), \\ x(0) = (1, 1, 1)^T \end{cases}$$

的解, 其中矩阵 A 与例 5.4.1 的 A 相同, $f(t) = (0, 0, e^{2t})^T$.

解 由前面的讨论可知, 此问题的解为

$$x = e^{At} x(0) + \int_0^t e^{A(t-\tau)} f(\tau) d\tau.$$

这里 $e^{At} x(0)$ 在例 5.4.1 已经求出, 故只须计算积分

$$I = \int_0^t e^{A(t-\tau)} f(\tau) d\tau.$$

由于 $(P$ 与 P^{-1} 见例 5.4.1)

$$\begin{aligned} e^{A(t-\tau)} f(\tau) &= P \begin{pmatrix} 1 & & \\ & e^{2(t-\tau)} & \\ & & e^{3(t-\tau)} \end{pmatrix} P^{-1} \begin{pmatrix} 0 \\ 0 \\ e^{2\tau} \end{pmatrix} \\ &= P \begin{pmatrix} 1 & & \\ & e^{2(t-\tau)} & \\ & & e^{3(t-\tau)} \end{pmatrix} \left(-\frac{1}{6} \right) \begin{pmatrix} -e^{2\tau} \\ 9e^{2\tau} \\ -4e^{2\tau} \end{pmatrix} \\ &= -\frac{1}{6} P \begin{pmatrix} -e^{2\tau} \\ 9e^{2\tau} \\ -4e^{3t-\tau} \end{pmatrix} \\ &= -\frac{1}{6} \begin{pmatrix} -e^{2\tau} + 9e^{2t} - 8e^{3t-\tau} \\ -5e^{2\tau} + 9e^{2t} - 4e^{3t-\tau} \\ -2e^{2\tau} - 4e^{3t-\tau} \end{pmatrix}, \end{aligned}$$

将这一结果对变量 τ 从 0 到 t 进行积分, 即得

$$\mathbf{u} = -\frac{1}{6} \begin{pmatrix} \frac{1}{2} + \left(9t + \frac{15}{2}\right)e^{2t} - 8e^{3t} \\ \frac{5}{2} + \left(9t + \frac{3}{2}\right)e^{2t} - 4e^{3t} \\ 1 + 3e^{2t} - 4e^{3t} \end{pmatrix},$$

因此 $\mathbf{x} = e^{At}\mathbf{x}(0) + \mathbf{u}$, 即

$$\mathbf{x} = -\frac{1}{6} \begin{pmatrix} -\frac{1}{2} + \left(9t + \frac{21}{2}\right)e^{2t} - 16e^{3t} \\ -\frac{5}{2} + \left(9t + \frac{9}{2}\right)e^{2t} - 8e^{3t} \\ -1 + 3e^{2t} - 8e^{3t} \end{pmatrix}.$$

例 5.4.4 求常系数非齐次线性方程组

$$\begin{cases} \frac{dy_1(t)}{dt} = 2y_1 - y_2 + y_3 + e^{2t}, \\ \frac{dy_2(t)}{dt} = 3y_2 - y_3, \\ \frac{dy_3(t)}{dt} = 2y_1 + y_2 + 3y_3 + te^{2t} \end{cases}$$

在初始条件

$$\mathbf{y}(0) = \begin{pmatrix} y_1(0) \\ y_2(0) \\ y_3(0) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

下的解.

解 将方程组写成向量方程

$$\frac{d\mathbf{y}(t)}{dt} = \mathbf{A}\mathbf{y}(t) + \mathbf{f}(t),$$

其中

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

$$\mathbf{y}(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{pmatrix},$$

$$\mathbf{f}(t) = \begin{pmatrix} e^{2t} \\ 0 \\ te^{2t} \end{pmatrix}.$$

由非齐次方程组的定解问题可求解公式得

$$\mathbf{y}(t) = e^{At}\mathbf{y}(0) + \int_0^t e^{A(t-\tau)}\mathbf{f}(\tau)d\tau.$$

由于 \mathbf{A} 的若尔当标准形 \mathbf{J} 及相似变换矩阵 \mathbf{P} 分别为

$$\mathbf{J} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix},$$

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

而且

$$\mathbf{P}^{-1} = \begin{pmatrix} -\frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix},$$

于是

$$\begin{aligned} e^{At} \mathbf{y}(0) &= \mathbf{P} e^{At} \mathbf{P}^{-1} \mathbf{y}(0) \\ &= \begin{pmatrix} -1 & 0 & 1 \\ 1 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} e^{2t} & t e^{2t} & 0 \\ 0 & e^{2t} & 0 \\ 0 & 0 & e^{4t} \end{pmatrix} \begin{pmatrix} -\frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} -2te^{2t} + e^{4t} \\ 2(1+t)e^{2t} - e^{4t} \\ 2te^{2t} + e^{4t} \end{pmatrix}, \end{aligned}$$

被积函数

$$\begin{aligned} e^{A(t-\tau)} \mathbf{f}(\tau) &= \mathbf{P} e^{J(t-\tau)} \mathbf{P}^{-1} \mathbf{f}(\tau) \\ &= \begin{pmatrix} -1 & 0 & 1 \\ 1 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} e^{2(t-\tau)} & (t-\tau)e^{2(t-\tau)} & 0 \\ 0 & e^{2(t-\tau)} & 0 \\ 0 & 0 & e^{4(t-\tau)} \end{pmatrix} \\ &\quad \times \begin{pmatrix} -\frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} e^{2\tau} \\ 0 \\ \tau e^{2\tau} \end{pmatrix} \\ &= \frac{e^{2t}}{2} \begin{pmatrix} 1 - 2t + e^{2(t-\tau)} + \tau + \tau e^{2(t-\tau)} \\ 1 + 2t - e^{2(t-\tau)} - \tau - \tau e^{2(t-\tau)} \\ -1 + 2t + e^{2(t-\tau)} - \tau + \tau e^{2(t-\tau)} \end{pmatrix}, \end{aligned}$$

代入公式, 定解问题的解为

$$\mathbf{y}(t) = e^{At} \mathbf{y}(0) + \int_0^t e^{A(t-\tau)} \mathbf{f}(\tau) d\tau$$

$$= e^{2t} \begin{pmatrix} \frac{11}{8}e^{2t} - \frac{3}{8} - \frac{7}{4}t - \frac{3}{4}t^2 \\ -\frac{11}{8}e^{2t} + \frac{19}{8} + \frac{11}{4}t + \frac{3}{4}t^2 \\ \frac{11}{8}e^{2t} - \frac{3}{8} + \frac{5}{4}t + \frac{3}{4}t^2 \end{pmatrix}.$$

5.4.3 n 阶常系数微分方程的解

先考虑 n 阶常系数齐次线性方程的定解问题

$$\begin{cases} y^{(n)} + a_1 y^{(n-1)} + a_2 y^{(n-2)} + \cdots + a_n y = 0, \\ y^{(i)}(t) \Big|_{t=0} = y_0^{(i)}, \quad i = 0, 1, \cdots, n-1, \end{cases} \quad (5.4.7)$$

其中 a_1, a_2, \cdots, a_n 为常数.

由于常系数线性微分方程组的矩阵形式解已经得到, 因此, 人们自然想把高阶微分方程化为一阶方程组来求解.

令

$$\begin{cases} x_1 = y, \\ x_2 = y' = x_1', \\ \vdots \\ x_n = y^{(n-1)} = x_{n-1}', \end{cases}$$

即

$$\begin{cases} x_1' = x_2, \\ x_2' = x_3, \\ \vdots \\ x_{n-1}' = x_n, \\ x_n' = -a_n x_1 - a_{n-1} x_2 - \cdots - a_1 x_n. \end{cases}$$

再令

$$\begin{aligned} \mathbf{x}(t) &= (x_1(t), x_2(t), \cdots, x_n(t))^T, \\ \mathbf{x}(0) &= (x_1(0), x_2(0), \cdots, x_n(0))^T = (y_0, y_0', \cdots, y_0^{(n-1)})^T, \end{aligned}$$

则定解问题(5.4.7)可写成

$$\begin{cases} \frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t), \\ \mathbf{x}(t) \Big|_{t=0} = \mathbf{x}(0), \end{cases} \quad (5.4.8)$$

其中

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{pmatrix}.$$

显然, A 为友矩阵.

由定理 5.4.1 得, 定解问题(5.4.8)的解为 $\mathbf{x}(t) = e^{At} \mathbf{x}(0)$. 由于定解问题(5.4.7)的解是式(5.4.8)的一个分量, 从而定解问题(5.4.7)的解为

$$\begin{aligned} \mathbf{y} &= (1, 0, 0, \dots, 0) \mathbf{x}(t) \\ &= (1, 0, 0, \dots, 0) e^{At} \mathbf{x}(0) \\ &= (1, 0, 0, \dots, 0) e^{At} \begin{pmatrix} y_0 \\ y_0' \\ \vdots \\ y_0^{(n-1)} \end{pmatrix}. \end{aligned}$$

对于 n 阶常系数非齐次线性方程的定解问题

$$\begin{cases} y^{(n)} + a_1 y^{(n-1)} + a_2 y^{(n-2)} + \dots + a_n y = f(t), \\ y^{(i)}(t) \Big|_{t=0} = y_0^{(i)}, \quad i = 0, 1, \dots, n-1, \end{cases} \quad (5.4.9)$$

可作类似讨论, 进而得到定解问题(5.4.9)的解是方程组

$$\begin{cases} \frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{b}f(t), \\ \mathbf{x}(t) \Big|_{t=0} = \mathbf{x}(0) \end{cases}$$

的解的第一个分量, 其中

$$\begin{aligned} \mathbf{x}(t) &= \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} = \begin{pmatrix} y \\ y' \\ \vdots \\ y^{(n-1)} \end{pmatrix}, \\ \mathbf{x}(0) &= \begin{pmatrix} x_1(0) \\ x_2(0) \\ \vdots \\ x_n(0) \end{pmatrix} = \begin{pmatrix} y_0 \\ y_0' \\ \vdots \\ y_0^{(n-1)} \end{pmatrix}, \\ \mathbf{A} &= \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \dots & -a_1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}. \end{aligned}$$

由定解问题(5.4.6)知(令 $\mathbf{f}(t) = \mathbf{b}f(t)$)

$$\mathbf{x}(t) = e^{At} \mathbf{x}(0) + \int_0^t e^{A(t-\tau)} \mathbf{b}f(\tau) d\tau,$$

从而定解问题(5.4.9)的解为

$$\mathbf{y}(t) = (1, 0, \dots, 0) \left(e^{At} \mathbf{x}(0) + \int_0^t e^{A(t-\tau)} \mathbf{b}f(\tau) d\tau \right).$$

由上述讨论可知, 不论是一阶常系数线性微分方程组, 还是 n 阶常系数线性微分方程, 求解定解问题的关键在于计算矩阵函数 e^{At} , 而计算矩阵函数 e^{At} 是前面 5.5.2 节的基本内容, 从而求解常系数一阶线性微分方程组及 n 阶线性微分方程的定解问题的解的问题已经解决.

例 5.4.5 设某一动态微分方程为

$$y''' + 7y'' + 14y' + 8y = 6f(t),$$

其中 y 为系统的输出函数, $f(t)$ 为系统的输入函数, 求 $y(t)$.

解 令

$$\begin{cases} x_1 = y, \\ x_2 = y', \\ x_3 = y'', \end{cases}$$

则

$$\begin{cases} x_1' = x_2, \\ x_2' = x_3, \\ x_3' = y''' = -8x_1 - 14x_2 - 7x_3 + 6f(t). \end{cases}$$

写成向量方程组为

$$\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{b}f(t),$$

其中

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -8 & -14 & -7 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 0 \\ 0 \\ 6 \end{pmatrix}.$$

由于 \mathbf{A} 为友矩阵, 故特征多项式为

$$\varphi(\lambda) = \lambda^3 + 7\lambda^2 + 14\lambda + 8 = (\lambda + 1)(\lambda + 2)(\lambda + 4).$$

故 \mathbf{A} 的特征值分别为

$$\lambda_1 = -1, \quad \lambda_2 = -2, \quad \lambda_3 = -4.$$

从而 \mathbf{A} 的若尔当标准形 \mathbf{J} 及相似变换矩阵 \mathbf{P} 分别为

$$\mathbf{J} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -4 \end{pmatrix},$$

$$\mathbf{P} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & -2 & -4 \\ 1 & 4 & 16 \end{pmatrix},$$

而且

$$\mathbf{P}^{-1} = \frac{1}{6} \begin{pmatrix} 16 & 12 & 2 \\ -12 & -15 & -3 \\ 2 & 3 & 1 \end{pmatrix}.$$

方法 1 计算 $e^{\mathbf{A}t}$ 及 $e^{\mathbf{A}(t-\tau)}$, 并代入非齐次线性方程组的求解公式

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{b}f(\tau)d\tau,$$

其中

$$\mathbf{x}(0) = \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix}$$

为任意给定的初始条件. 因而原方程组的解为

$$\begin{aligned} y(t) &= (1, 0, \dots, 0)x(t) \\ &= (1, 0, \dots, 0) \left(e^{At}x(0) + \int_0^t e^{A(t-\tau)}bf(\tau)d\tau \right). \end{aligned}$$

此种方法在计算 e^{At} 及 $e^{A(t-\tau)}$ 的过程中显得烦琐, 下面介绍另一种方法.

方法 2 令

$$x = Pz,$$

其中 $z = (z_1, z_2, z_3)^T$, 则

$$x' = Pz',$$

因而

$$Pz' = APz + bf(t),$$

两边左乘 P^{-1} , 得

$$\begin{aligned} z' &= P^{-1}APz + P^{-1}bf(t) \\ &= \begin{pmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -4 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} + \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix} f(t), \end{aligned}$$

由 $y = (1, 0, 0)x$, $x = Pz$ 得

$$\begin{aligned} y &= (1, 0, 0) \begin{pmatrix} 1 & 1 & 1 \\ -1 & -2 & -4 \\ 1 & 4 & 16 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} \\ &= z_1 + z_2 + z_3, \end{aligned}$$

由于

$$\begin{cases} z_1 = e^{-t} \left(k_1 + \int 2f(t)e^t dt \right), \\ z_2 = e^{-2t} \left(k_2 + \int -3f(t)e^{2t} dt \right), \\ z_3 = e^{-4t} \left(k_3 + \int f(t)e^{4t} dt \right), \end{cases}$$

从而原方程组的解为

$$y = k_1 e^{-t} + k_2 e^{-2t} + k_3 e^{-4t} + 2e^{-t} \int f(t)e^t dt - 3e^{-2t} \int f(t)e^{2t} dt + e^{-4t} \int f(t)e^{4t} dt,$$

其中, k_1, k_2, k_3 为 3 个独立的任意常数.

以上我们仅仅介绍了矩阵函数在一阶常系数齐次、非齐次线性微分方程组和 n 阶常系数线性微分方程中的应用, 这是较为简单的情形. 同样地, 在一阶变系数齐次及非齐次线性微分方程组中, 矩阵函数都有应用, 但已经远比上面所讲的困难多了, 在此不作更深入的讨论.

习 题 5

1. 若 $\lim_{k \rightarrow \infty} A^{(k)} = A$, 证明: $\lim_{k \rightarrow \infty} \|A^{(k)}\| = \|A\|$, 其中 $A^{(k)}, A \in \mathbb{C}^{m \times n}$, $\|\cdot\|$ 是 $\mathbb{C}^{m \times n}$ 中的任何一种矩阵范数.
2. 设 $A^{(k)}, B^{(k)} \in \mathbb{C}^{m \times n}$, $\alpha_k, \beta_k \in \mathbb{C}$, 且 $\lim_{k \rightarrow \infty} A^{(k)} = A$, $\lim_{k \rightarrow \infty} B^{(k)} = B$, $\lim_{k \rightarrow \infty} \alpha_k = \alpha$, $\lim_{k \rightarrow \infty} \beta_k = \beta$. 证明:

$$\lim_{k \rightarrow \infty} (\alpha_k A^{(k)} + \beta_k B^{(k)}) = \alpha A + \beta B.$$

3. 下列矩阵是否为收敛矩阵? 为什么?

$$(1) \mathbf{A} = \begin{pmatrix} 0.2 & 0.1 & 0.2 \\ 0.5 & 0.5 & 0.4 \\ 0.1 & 0.3 & 0.2 \end{pmatrix}; \quad (2) \mathbf{B} = \begin{pmatrix} \frac{1}{6} & -\frac{4}{3} \\ -\frac{1}{3} & \frac{1}{6} \end{pmatrix}.$$

4. 设 $\mathbf{A} = \begin{pmatrix} 0 & a & a \\ a & 0 & a \\ a & a & 0 \end{pmatrix}$, 讨论实数 a 取何值时, \mathbf{A} 为收敛矩阵?

5. 设矩阵级数 $\sum_{k=0}^{\infty} \mathbf{A}^{(k)}$ 收敛(或绝对收敛), 证明: $\sum_{k=0}^{\infty} \mathbf{P}\mathbf{A}^{(k)}\mathbf{Q}$ 也收敛(或绝对收敛), 并且有等式

$$\sum_{k=0}^{\infty} \mathbf{P}\mathbf{A}^{(k)}\mathbf{Q} = \mathbf{P} \left(\sum_{k=0}^{\infty} \mathbf{A}^{(k)} \right) \mathbf{Q}. \text{ 其中 } \mathbf{A}^{(k)} \in \mathbb{C}^{m \times n}, \mathbf{P} \in \mathbb{C}^{s \times m}, \mathbf{Q} \in \mathbb{C}^{n \times t}.$$

6. 讨论下列矩阵幂级数的敛散性.

$$(1) \sum_{k=1}^{\infty} \frac{1}{k^2} \begin{bmatrix} 1 & 7 \\ -1 & -3 \end{bmatrix}^k; \quad (2) \sum_{k=0}^{\infty} \frac{k}{6^k} \begin{bmatrix} 1 & -8 \\ -2 & 1 \end{bmatrix}^k.$$

7. 计算矩阵幂级数 $\sum_{k=0}^{\infty} \begin{bmatrix} 0.1 & 0.7 \\ 0.3 & 0.6 \end{bmatrix}^k$.

8. (1) 设矩阵序列 $\{\mathbf{A}^k\}$, 其中 $\mathbf{A}^k \in \mathbb{C}^{m \times n}$, 若 $\lim_{k \rightarrow \infty} \mathbf{A}^k = \mathbf{A}$, 证明 $\lim_{k \rightarrow \infty} \|\mathbf{A}^k\| = \|\mathbf{A}\|$;

(2) 举例说明, $\lim_{k \rightarrow \infty} \|\mathbf{A}^k\| = \|\mathbf{A}\|$, 不一定有

$$\lim_{k \rightarrow \infty} \mathbf{A}^k = \mathbf{A}.$$

9. 设

$$\mathbf{A} = \begin{bmatrix} 0 & a & a \\ a & 0 & a \\ a & a & 0 \end{bmatrix},$$

问 a 取何值时, 有 $\lim_{k \rightarrow \infty} \mathbf{A}^k = \mathbf{0}$.

10. 设矩阵

$$\mathbf{A} = \begin{bmatrix} \frac{1}{2} & 1 & 1 \\ 0 & \frac{1}{3} & 1 \\ 0 & 0 & \frac{1}{5} \end{bmatrix},$$

求 $\lim_{k \rightarrow \infty} \mathbf{A}^k$.

11. 求 $\lim_{k \rightarrow \infty} \left(\frac{\mathbf{A}}{\rho(\mathbf{A})} \right)^k$, 其中矩阵 \mathbf{A} 为 $\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$.

12. 已知 \mathbf{A} 是一个 5 阶方阵, 其特征多项式为

$$f(\lambda) = |\lambda \mathbf{I} - \mathbf{A}| = \left(\lambda - \frac{1}{3} \right) (\lambda - 1)^2 \left(\lambda - \frac{1}{4} \right)^2,$$

且 $\text{rank}(\mathbf{I} - \mathbf{A}) = 3$, 试证: 矩阵序列 $\{\mathbf{A}^k\}$ 收敛.

13. 判断矩阵幂级数

$$\sum_{k=0}^{\infty} \begin{bmatrix} \frac{1}{6} & -\frac{1}{3} \\ -\frac{4}{3} & \frac{1}{6} \end{bmatrix}^k$$

的敛散性.

14. 已知

$$\mathbf{A} = \begin{bmatrix} \frac{1}{5} & \frac{3}{5} \\ \frac{3}{5} & \frac{1}{5} \end{bmatrix},$$

(1) 求证: 矩阵幂级数 $\sum_{k=1}^{\infty} k^2 \mathbf{A}^k$ 收敛;

(2) 求矩阵幂级数 $\sum_{k=1}^{\infty} k^2 \mathbf{A}^k$ 的收敛和.

15. 已知 \mathbf{A} 为一个 n 阶矩阵, 且 $\rho(\mathbf{A}) < 1$, 求 $\sum_{k=0}^{\infty} k^2 \mathbf{A}^k$.

16. 求 Von Neumann 级数 $\sum_{k=0}^{\infty} \begin{bmatrix} 0.2 & 0.5 \\ 0.7 & 0.4 \end{bmatrix}^k$ 的和.

17. 试找一个收敛的二阶可逆矩阵序列, 但其极限矩阵不可逆.

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 $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$, $\mathbf{C} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$, $\lim_{k \rightarrow \infty} \mathbf{A}^k = \mathbf{M}$, 求 $\lim_{k \rightarrow \infty} \mathbf{C}^k$ 以及

$$\lim_{k \rightarrow \infty} \begin{bmatrix} 0.2 & 0 & 1 & 1 \\ 0.3 & 0.7 & 0 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^k.$$

20. 讨论矩阵幂级数

$$\sum_{k=1}^{\infty} \frac{1}{k^2} \begin{bmatrix} -2 & 1 & -1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}^k$$

的敛散性.

21. 已知矩阵 \mathbf{A} 的某种范数 $\|\mathbf{A}\| < 1$, 求 $\sum_{k=1}^{\infty} k \mathbf{A}^{k-1}$.

22. 设 $\mathbf{C}, \mathbf{D} \in \mathbb{R}^{n \times n}$, $\mathbf{A} = \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{D} & \mathbf{I}_n \end{bmatrix}$, $\lim_{k \rightarrow \infty} \mathbf{C}^k = \mathbf{S}$, 且 $\rho(\mathbf{C}) < 1$, 求 $\lim_{k \rightarrow \infty} \mathbf{A}^k$, 并求

$$\lim_{k \rightarrow \infty} \begin{bmatrix} 0.1 & 0.3 & 0 & 0 \\ 0 & 0.6 & 0 & 0 \\ \hline 1 & 0 & 1 & 0 \\ \hline 1 & 2 & 0 & 1 \end{bmatrix}^k.$$

23. 设 \mathbf{A} 为 5 阶方阵, 其特征多项式

$$f(\lambda) = |\lambda \mathbf{I} - \mathbf{A}| = (\lambda - 1)^2 (\lambda - 0.5)^2 (\lambda - 0.25),$$

且 $\text{rank}(\mathbf{I} - \mathbf{A}) = 3$, 证明 $\lim_{k \rightarrow \infty} \mathbf{A}^k$ 收敛.

24. (1) 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 若 $\|\mathbf{A}\| < 1$, 则 $\lim_{k \rightarrow \infty} \mathbf{A}^k = \mathbf{0}$.

(2) $\mathbf{A} \in \mathbb{C}^{n \times n}$, $\lim_{k \rightarrow \infty} \mathbf{A}^k = \mathbf{0} \Leftrightarrow \rho(\mathbf{A}) < 1$.

25. 设 $\mathbf{A}^k \in \mathbb{C}^{n \times n}$, $\lim_{k \rightarrow \infty} \mathbf{A}^k = \mathbf{A}$, 且 $(\mathbf{A}^k)^{-1}$ 和 \mathbf{A}^{-1} 均存在, 则 $\lim_{k \rightarrow \infty} [\mathbf{A}^k]^{-1} = \mathbf{A}^{-1}$, 试举一具体例子, 说明条件 $(\mathbf{A}^k)^{-1}$ 与 \mathbf{A} 都可逆是不可少的.

26. 讨论矩阵幂级数 $\sum_{k=0}^{\infty} \mathbf{A}^k$ 的敛散性, 其中

$$\mathbf{A} = \begin{bmatrix} 0.2 & 0.1 & 0.2 \\ 0.5 & 0.5 & 0.4 \\ 0.1 & 0.3 & 0.2 \end{bmatrix}.$$

27. 证明: $e^{A+2\pi i I} = e^A$ (其中 $i = \sqrt{-1}$), $\sin(\mathbf{A} + 2\pi \mathbf{I}) = \sin \mathbf{A}$.

28. 证明: 若 \mathbf{A} 为实反对称矩阵 ($\mathbf{A}^T = -\mathbf{A}$), 则 $e^{\mathbf{A}}$ 为正交矩阵.

29. 证明: 若 \mathbf{A} 是埃尔米特矩阵, 则 $e^{i\mathbf{A}}$ 是酉矩阵.

30. 设

$$\mathbf{A} = \begin{bmatrix} -2 & 1 & 0 \\ -4 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

(1) 试求 \mathbf{A} 的特征多项式;

(2) 利用凯莱-哈密顿定理计算 $e^{\mathbf{A}}$ 和 $\sin \mathbf{A}$.

31. 设 $\mathbf{A} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$, 求 $e^{\mathbf{A}}$, $e^{t\mathbf{A}}$ ($t \in \mathbb{R}$) 及 $\sin \mathbf{A}$.

32. 设 $f(z) = \ln z$, 求 $f(\mathbf{A})$. 这里 \mathbf{A} 为

$$(1) \mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}; \quad (2) \mathbf{A} = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

33. 对下列矩阵 \mathbf{A} , 求矩阵函数 $e^{t\mathbf{A}}$.

$$(1) \mathbf{A} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}; \quad (2) \mathbf{A} = \begin{bmatrix} 2 & -2 & 3 \\ 1 & 1 & 1 \\ 1 & 3 & -1 \end{bmatrix};$$

$$(3) \mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -8 & -12 & -6 \end{bmatrix}; \quad (4) \mathbf{A} = \begin{bmatrix} -2 & 1 & 3 \\ 0 & -3 & 0 \\ 0 & 2 & -2 \end{bmatrix}.$$

34. 求下列三类矩阵的矩阵函数 $\cos \mathbf{A}$, $\sin \mathbf{A}$, $e^{\mathbf{A}^2}$:

(1) 当 \mathbf{A} 为幂等矩阵 ($\mathbf{A}^2 = \mathbf{A}$) 时;

(2) 当 \mathbf{A} 为对合矩阵 ($\mathbf{A}^2 = \mathbf{I}$) 时;

(3) 当 \mathbf{A} 为幂零矩阵 ($\mathbf{A}^2 = \mathbf{0}$) 时.

35. 计算下列矩阵函数:

$$(1) \mathbf{A} = \begin{pmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{pmatrix}, \text{ 求 } \mathbf{A}^{1000};$$

$$(2) \mathbf{A} = \begin{bmatrix} 4 & 2 & -5 \\ 6 & 4 & -9 \\ 5 & 3 & -7 \end{bmatrix}, \text{ 求 } e^{\mathbf{A}};$$

$$(3) \mathbf{A} = \begin{bmatrix} 0 & -1 \\ 4 & 4 \end{bmatrix}, \text{ 求 } \arcsin \frac{\mathbf{A}}{4};$$

$$(4) \mathbf{A} = \begin{bmatrix} 16 & 8 \\ 8 & 4 \end{bmatrix}, \text{ 求 } (\mathbf{I} + \mathbf{A})^{-1} \text{ 及 } \mathbf{A}^{\frac{1}{2}}.$$

36. 求矩阵 A 的最小多项式, 已知

$$(1) A = \begin{bmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{bmatrix}; \quad (2) A = \begin{bmatrix} 3 & 1 & -1 \\ -2 & 0 & 2 \\ -1 & -1 & 3 \end{bmatrix};$$

$$(3) A = \text{diag} \left[\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}, [3], \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix} \right].$$

37. 设 3 阶矩阵 A 的特征多项式与最小多项式分别为 $f(\lambda) = \lambda^3 - 5\lambda^2$ 与 $m(\lambda) = \lambda^2 - 5\lambda$, 试用它们计算矩阵 A^4 的表达式.

38. 用若尔当标准形法求 e^{At} , 其中

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

$$39. A = \begin{bmatrix} 1 & 2 & -6 \\ 1 & 0 & -3 \\ 1 & 1 & -4 \end{bmatrix}, \text{求 } e^A, e^{At}, \sin A.$$

$$40. \text{设 } f(x) = x^{\frac{1}{2}}, A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \text{求 } f(A).$$

$$41. \text{设 } A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \text{求 } e^{At}.$$

$$42. A = \frac{\pi}{2} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \text{求 } \sin A.$$

$$43. \text{已知 } A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix},$$

(1) 求 P , 使 $P^{-1}AP = J$;

(2) 求 $e^{\frac{i\pi}{2}A}$ 的若尔当标准形.

44. 设 4 阶方阵 A 的若尔当标准形是由两个二阶非零若尔当块构成, 且 $A^2 = \mathbf{0}$, 求 $\sin A$.

45. A^H 为 n 阶矩阵, 证明:

(1) $(e^A)^H = e^{A^H}$;

(2) 若 A 为反对称矩阵 $A^T = -A$, 则 e^A 是正交阵;

(3) 若 $A^H = A$, 则 e^A 是酉阵.

46. 已知矩阵

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

求矩阵函数 $f(A)$ 的若尔当表示, 并计算 $e^A, e^{At}, \arctan \frac{A}{4}, \sin \frac{\pi}{2} A, \cos \pi A$.

47. 已知矩阵

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix},$$

求矩阵函数 $f(\mathbf{A})$ 的若尔当表示, 并计算 $e^{\mathbf{A}}, \sin \mathbf{A}, \cos \pi \mathbf{A}, \ln(\mathbf{I} + \mathbf{A})$.

48. 已知矩阵 \mathbf{A} , 求矩阵函数 $f(\mathbf{A})$ 的若尔当表示, 并计算 $e^{\mathbf{A}}, \sin \frac{\pi}{2} \mathbf{A}, \cos \pi \mathbf{A}, \ln(\mathbf{I} + \mathbf{A})$.

$$(1) \mathbf{A} = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}; \quad (2) \mathbf{A} = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

49. 已知

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

求 $e^{\mathbf{A}}, \sin \mathbf{A}, \cos \mathbf{A}$.

50. 已知

$$\mathbf{A} = \begin{bmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix},$$

(1) 求 $e^{\mathbf{A}}, \sin \mathbf{A}$;

(2) 分别求可逆矩阵 $\mathbf{Q}_1, \mathbf{Q}_2$, 使得 $\mathbf{Q}_1^{-1} e^{\mathbf{A}} \mathbf{Q}_1$ 与 $\mathbf{Q}_2^{-1} \sin \mathbf{A} \mathbf{Q}_2$ 均为若尔当标准形矩阵.

51. 已知矩阵 \mathbf{A} , 求矩阵函数 $f(\mathbf{A})$ 的多项式表示, 并计算 $e^{\mathbf{A}}$.

52. 已知矩阵 $\mathbf{A} = \begin{bmatrix} 2 & 1 & 0 \\ -1 & 0 & 0 \\ -2 & -1 & 2 \end{bmatrix}$, 求矩阵函数 $f(\mathbf{A})$ 的多项式表示, 并计算 $\cos \pi \mathbf{A}$.

53. 设 \mathbf{A} 为 n 阶矩阵, 试证:

(1) $e^{2\pi i \mathbf{I}} = \mathbf{I}, e^{2\pi i \mathbf{I} + \mathbf{A}} = e^{\mathbf{A}}$, 这里 $i^2 = -1$;

(2) $\sin 2\pi \mathbf{I} = \mathbf{0}, \cos 2\pi \mathbf{I} = \mathbf{I}$;

(3) $\sin(2\pi \mathbf{I} + \mathbf{A}) = \sin \mathbf{A}$;

(4) $|e^{\mathbf{A}}| = e^{\text{tr}(\mathbf{A})}$;

(5) $\|e^{\mathbf{A}}\| \leq e^{\|\mathbf{A}\|}$.

54. 求下列各矩阵函数的幂级数表示, 并求出相应幂级数的收敛范围.

(1) $e^{\mathbf{A}}$; (2) $\sin \frac{\pi}{2} \mathbf{A}$; (3) $(3\mathbf{I} - \mathbf{A})^{-1}$.

55. 计算矩阵幂级数 $\sum_{k=1}^{\infty} \frac{k^2}{2^k} \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}^k$ 之和.

56. 已知

$$\mathbf{A} = \begin{bmatrix} 0.9 & 1 \\ 0 & 0.8 \end{bmatrix},$$

求 $\mathbf{I} + 2\mathbf{A} + 3\mathbf{A}^2 + \cdots + k\mathbf{A}^{k-1} + \cdots$ 之和.

57. 已知 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 那么

(1) $\frac{d}{dt} e^{t\mathbf{A}} = \mathbf{A} e^{t\mathbf{A}} = e^{t\mathbf{A}} \mathbf{A}$;

(2) $\frac{d}{dt} \cos t\mathbf{A} = -\mathbf{A} [\sin(t\mathbf{A})] = -[\sin(t\mathbf{A})] \mathbf{A}$;

(3) $\frac{d}{dt} \sin(t\mathbf{A}) = \mathbf{A} [\cos(t\mathbf{A})] = [\cos(t\mathbf{A})] \mathbf{A}$.

58. 已知函数矩阵

$$\mathbf{A}(t) = \begin{bmatrix} 2t-1 & e^t \\ t & 0 \\ t-1 & 0 \end{bmatrix},$$

- (1) 证明: $\mathbf{A}(t)$ 在点 $t=0$ 处连续;
 (2) 证明: $\mathbf{A}(t)$ 在点 $t=1$ 处不连续.

59. 设

$$\mathbf{A}(t) = \begin{bmatrix} \cos t & \sin t \\ 2+t & 0 \end{bmatrix},$$

- (1) 求 $\frac{d\mathbf{A}(t)}{dt}$; (2) 求 $\int_0^\pi \mathbf{A}(t) dt$.

60. 已知函数矩阵

$$\mathbf{A}(x) = \begin{bmatrix} \sin x & \cos x & x \\ \frac{\sin x}{x} & e^x & x^2 \\ 1 & 0 & x^3 \end{bmatrix},$$

其中 $x \neq 0$. 试求 $\lim_{x \rightarrow 0} \mathbf{A}(x)$, $\frac{d\mathbf{A}(x)}{dx}$, $\frac{d^2\mathbf{A}(x)}{dx^2}$, $\left| \frac{d\mathbf{A}(x)}{dx} \right|$.

61. 已知函数矩阵

$$\mathbf{A}(x) = \begin{bmatrix} e^{2x} & xe^x & x^2 \\ e^{-x} & 2e^{2x} & 0 \\ 3x & 0 & 0 \end{bmatrix},$$

试求 $\int_0^1 \mathbf{A}(x) dx$ 和 $\frac{d}{dx} \left(\int_0^{x^2} \mathbf{A}(t) dt \right)$.

62. 设 $\mathbf{A}(t) = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}$, 求 $\frac{d}{dt} \mathbf{A}(t)$, $\frac{d}{dt} \mathbf{A}^{-1}(t)$, $\frac{d}{dt} |\mathbf{A}(t)|$, $\left| \frac{d}{dt} \mathbf{A}(t) \right|$.

63. 设 $\mathbf{A}(t)$ 是 n 阶可微矩阵, 说明关系式

$$\frac{d}{dt} [\mathbf{A}(t)]^m = m[\mathbf{A}(t)]^{m-1} \frac{d}{dt} \mathbf{A}(t)$$

一般不成立. 问此式在什么条件下能成立?

64. 已知

$$\sin \mathbf{A}t = \frac{1}{4} \begin{pmatrix} \sin 5t + 3\sin t & 2\sin 5t - 2\sin t & \sin 5t - \sin t \\ \sin 5t - \sin t & 2\sin 5t + 2\sin t & \sin 5t - \sin t \\ \sin 5t - \sin t & 2\sin 5t - 2\sin t & \sin 5t + 3\sin t \end{pmatrix},$$

求 \mathbf{A} .

65. 已知函数矩阵

$$e^{\mathbf{A}t} = \begin{bmatrix} 2e^{2t} - e^t & e^{2t} - e^t & e^t - e^{2t} \\ e^{2t} - e^t & 2e^{2t} - e^t & e^t - e^{2t} \\ 3e^{2t} - 3e^t & 3e^{2t} - 3e^t & 3e^t - 2e^{2t} \end{bmatrix},$$

求矩阵 \mathbf{A} .

66. 求 $\int_0^t \mathbf{A}(\tau) d\tau$, 其中

$$\mathbf{A}(t) = \begin{bmatrix} e^{2t} & te^t & 1+t \\ e^{-2t} & 2e^{2t} & \sin t \\ 3t & 0 & t \end{bmatrix}.$$

67. 若函数矩阵 $\mathbf{A}(t)$ 在 $[t_0, t]$ 上可积, 证明:

$$\left\| \int_{t_0}^t \mathbf{A}(t) dt \right\|_1 \leq \int_{t_0}^t \|\mathbf{A}(t)\|_1 dt.$$

68. 设

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{X}(0) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix},$$

(1) 求 $e^{\mathbf{A}t}$;

(2) 求解 $\begin{cases} \dot{\mathbf{X}}(t) = \mathbf{A}\mathbf{X}(t) \\ \mathbf{X}(0) = (1, 0, 0, -1)^T. \end{cases}$

69. 设 $\mathbf{X} = [x_{ij}]_{n \times n}$, 求 $\frac{d}{d\mathbf{X}} \text{tr}(\mathbf{X}), \frac{d}{d\mathbf{X}} \det(\mathbf{X})$.

70. 设 $f(\mathbf{A}) = \|\mathbf{A}\|_F^2 = \text{tr}(\mathbf{A}^T \mathbf{A})$, 其中 $\mathbf{A} \in \mathbb{R}^{m \times n}$ 是矩阵变量, 求 $\frac{df}{d\mathbf{A}}$.

71. 设 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T, \mathbf{A} = (a_{ij})_{n \times n}$ 是实对称矩阵, $\mathbf{y} = (y_1, y_2, \dots, y_n)^T, c$ 为常数, 试求 $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{y}^T \mathbf{x} + c$ 对于 \mathbf{x} 的导数.

72. 设 \mathbf{X} 为 $n \times m$ 矩阵, \mathbf{A} 和 \mathbf{B} 依次为 $n \times n$ 和 $m \times n$ 的常数矩阵. 证明:

$$(1) \frac{d}{d\mathbf{X}} (\text{tr}(\mathbf{B}\mathbf{X})) = \frac{d}{d\mathbf{X}} (\text{tr}(\mathbf{X}^T \mathbf{B}^T)) = \mathbf{B}^T;$$

$$(2) \frac{d}{d\mathbf{X}} (\text{tr}(\mathbf{X}^T \mathbf{A} \mathbf{X})) = (\mathbf{A} + \mathbf{A}^T) \mathbf{X}.$$

73. 设 \mathbf{x} 为 n 维列向量, \mathbf{u} 为 n 维常数列向量, \mathbf{A} 为 n 阶常数对称矩阵. 证明:

$$\frac{d}{d\mathbf{X}} (\mathbf{x} - \mathbf{u})^T \mathbf{A} (\mathbf{x} - \mathbf{u}) = 2\mathbf{A}(\mathbf{x} - \mathbf{u}).$$

74. 设 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 是矩阵变量, 且 $\det \mathbf{A} \neq 0$, 令 $f(\mathbf{A}) = \det \mathbf{A}$, 证明:

$$\frac{df}{d\mathbf{A}} = \det(\mathbf{A}^{-1})^T.$$

75. 设 $\mathbf{B} \in \mathbb{R}^{n \times n}$ 是给定矩阵, $\mathbf{A} \in \mathbb{R}^{n \times m}$ 是矩阵变量, $f(\mathbf{A}) = \text{tr}(\mathbf{A}^T \mathbf{B} \mathbf{A})$, 试求 $\frac{df}{d\mathbf{A}}$.

76. 设 $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{x} \in \mathbb{R}^n$ 是向量变量, $F(\mathbf{x}) = \mathbf{A}\mathbf{x}, G(\mathbf{x}) = \mathbf{x}^T \mathbf{A}^T$, 试求 $\frac{dF(\mathbf{x})}{d\mathbf{x}}, \frac{dF(\mathbf{x})}{d\mathbf{x}^T}, \frac{dG(\mathbf{x})}{d\mathbf{x}}$.

77. 设 $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{b} \in \mathbb{R}^m, \mathbf{x} \in \mathbb{R}^n$ 是向量变量, $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$, 试求 $\frac{df}{d\mathbf{x}}$.

78. 设 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T, f(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_n(\mathbf{x}))^T$, 其中

$$f_i(\mathbf{x}) = \sum_{j=1}^n a_{ij} x_j + \delta_i, \quad i = 1, 2, \dots, n,$$

求 $\frac{d}{d\mathbf{x}} f(\mathbf{x})$.

79. 设 $\mathbf{A}(t) = \begin{bmatrix} e^{2t} & te^t & t^2 \\ e^{-t} & 2e^{2t} & 0 \\ 3t & 0 & 0 \end{bmatrix}$, 求 $\int \mathbf{A}(t) dt, \int_0^1 \mathbf{A}(t) dt$ 与 $\frac{d}{dt} \int_0^t \mathbf{A}(x) dx$.

80. 求微分方程的解:

$$\begin{cases} \frac{d\mathbf{x}}{dt} = \begin{bmatrix} -1 & 2 \\ -2 & 1 \end{bmatrix} \mathbf{x}(t), \\ \mathbf{x}(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \end{cases}$$

81. 求非齐次微分方程的解:

$$\begin{cases} \frac{d\mathbf{x}}{dt} = \begin{bmatrix} 3 & 5 \\ -5 & 3 \end{bmatrix} \mathbf{x}(t) + \begin{pmatrix} e^{-t} \\ 0 \end{pmatrix}, \\ \mathbf{x}(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \end{cases}$$

82. 求微分方程组

$$\begin{cases} \frac{dx_1}{dt} = -2x_1 + x_2 + 1, \\ \frac{dx_2}{dt} = -4x_1 + 2x_2 + 2, \\ \frac{dx_3}{dt} = x_1 + x_3 + e^t - 1 \end{cases}$$

满足初始条件 $x_1(0)=1, x_2(0)=1, x_3(0)=-1$ 的解.

83. 求解三阶齐次线性微分方程

$$\begin{cases} y'''(t) - 5y'' + 7y' - 3y = 0, \\ y(0) = 1, \quad y'(0) = 0, \quad y''(0) = 0. \end{cases}$$

84. 设 $\mathbf{Z}(t)$ 是非齐次常系数微分方程组

$$\dot{\mathbf{X}}(t) = \mathbf{A}\mathbf{X}(t) + \mathbf{f}(t)$$

的一个解, 证明满足初始条件 $\mathbf{X}(t_0)$ 的解为

$$\mathbf{X}(t) = \mathbf{Z}(t) + e^{\mathbf{A}(t-t_0)} [\mathbf{X}(t_0) - \mathbf{Z}(t_0)].$$

85. 设 \mathbf{A} 为 n 阶常数矩阵, 证明: 线性非齐次微分方程

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{f}(t), \quad \mathbf{x}(t_0) = \mathbf{x}_0$$

的解为

$$\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)} \mathbf{x}_0 + \int_{t_0}^t e^{\mathbf{A}(t-\tau)} \mathbf{f}(\tau) d\tau.$$

86. 求解微分方程组

$$\begin{cases} \dot{\mathbf{X}}(t) = \begin{bmatrix} 3 & -1 & 1 \\ 2 & 0 & -1 \\ 1 & -1 & 2 \end{bmatrix} \mathbf{X}(t) + \begin{bmatrix} 0 \\ 0 \\ e^{2t} \end{bmatrix}, \\ \mathbf{X}(0) = (1, 1, 1)^T. \end{cases}$$

87. 设微分方程组 $\dot{\mathbf{X}}(t) = \mathbf{A}\mathbf{X}(t)$, 其中系数矩阵 \mathbf{A} 可对角化, 试推导该方程组的一般解为

$$\dot{\mathbf{X}}(t) = C_1 e^{\lambda_1 t} \boldsymbol{\alpha}_1 + C_2 e^{\lambda_2 t} \boldsymbol{\alpha}_2 + \cdots + C_n e^{\lambda_n t} \boldsymbol{\alpha}_n,$$

其中 $\lambda_1, \lambda_2, \dots, \lambda_n$ 是 \mathbf{A} 的 n 个特征值, $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_n$ 是对应于这些特征值的 n 个线性无关的特征向量, C_1, C_2, \dots, C_n 为任意常数.

88. $\mathbf{A} = \begin{bmatrix} 1 & -1 & 4 \\ 3 & 2 & -1 \\ 2 & 1 & -1 \end{bmatrix}$, 求 $\dot{\mathbf{X}}(t) = \mathbf{A}\mathbf{X}(t)$ 的一般解(通解).

89. 设二阶齐次微分方程组:

$$\begin{cases} \ddot{\mathbf{X}}(t) + \mathbf{A}^2 \mathbf{X}(t) = \mathbf{0}, \\ \mathbf{X}(0) = \mathbf{X}_0, \quad \dot{\mathbf{X}}(0) = \mathbf{X}_1. \end{cases}$$

证明: (1) $\sin \mathbf{A}t, \cos \mathbf{A}t$ 是该方程组的两个解;

(2) 若 \mathbf{A} 可逆, 则通解为

$$\mathbf{X} = (\sin \mathbf{A}t) c_1 + (\cos \mathbf{A}t) c_2$$

且 $\mathbf{X}(t) = (\sin \mathbf{A}t) \mathbf{A}^{-1} \mathbf{x}_1 + (\cos \mathbf{A}t) \mathbf{X}_0$ 是满足初始条件的解.

90. 设 $\ddot{\mathbf{X}} + \mathbf{A}^2 \mathbf{X}(t) = \mathbf{0}$, \mathbf{A} 是半正定的埃尔米特矩阵, 求通解.

引言 什么是广义逆矩阵

广义逆矩阵是通常逆矩阵的推广,这种推广的必要性,是线性方程组的求解问题的实际需要.设有线性方程组

$$Ax = b, \quad (6.0.1)$$

当 A 是 n 阶方阵,且 $\det A \neq 0$ 时,则方程组(6.0.1)的解存在且惟一,并可写成

$$x = A^{-1}b. \quad (6.0.2)$$

但是,在许多实际问题中所遇到的矩阵 A 往往是奇异方阵或是任意的 $m \times n$ 矩阵(一般 $m \neq n$),显然不存在通常的逆矩阵 A^{-1} .这就促使人们去想象能否推广逆矩阵的概念,引进某种具有普通逆矩阵类似性质的矩阵 G ,使得其解仍可以表示为类似于式(6.0.2)的紧凑形式呢?即有

$$x = Gb. \quad (6.0.3)$$

1920年穆尔(Moore)首先引进了广义逆矩阵这一概念,其后30年未能引起人们的重视,直到1955年,彭诺斯(Penrose)以更明确的形式(4个矩阵方程)给出了穆尔的广义逆矩阵的定义之后,广义逆矩阵的研究才进入了一个新的时期.由于广义逆矩阵在数理统计、系统理论、最优化理论、现代控制理论等许多领域中的重要应用已为人们所认识,因而大大推动了对广义逆矩阵的研究,使得这一学科得到迅速的发展,已成为矩阵论的一个重要分支.

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.

6.1 矩阵的几种广义逆

6.1.1 广义逆矩阵的基本概念

1955年,彭诺斯指出:对任意复数矩阵 $A_{m \times n}$,如果存在复矩阵 $G_{n \times m}$,满足

$$AGA = A, \quad (6.1.1)$$

$$GAG = G, \quad (6.1.2)$$

$$(GA)^H = GA, \quad (6.1.3)$$

$$(AG)^H = AG, \quad (6.1.4)$$

则称 G 为 A 的一个穆尔-彭诺斯广义逆,并把上面4个方程叫做穆尔-彭诺斯方程,简称 M-P 方程.

由于 M-P 的 4 个方程都各有一定的解释, 并且应用起来各有方便之处, 所以出于不同的目的, 常常考虑满足部分方程的 G , 叫做弱逆. 为引用的方便, 我们给出如下的广义逆矩阵的定义.

定义 6.1.1 设 $A \in \mathbb{C}^{m \times n}$, 若有某个 $G \in \mathbb{C}^{n \times m}$, 满足 M-P 方程(6.1.1)~(6.1.4)中的全部或一部分, 则称 G 为 A 的广义逆矩阵, 简称为广义逆.

譬如有某个 G , 只满足式(6.1.1), 则 G 为 A 的 $\{1\}$ 广义逆, 记为 $G \in A\{1\}$; 如果另一个 G' 满足式(6.1.1)和式(6.1.2), 则 G' 为 A 的 $\{1,2\}$ 广义逆, 记为 $G' \in A\{1,2\}$; 如果 $G \in A\{1,2,3,4\}$, 则 G 同时满足 4 个方程, 它就是穆尔-彭诺斯广义逆等. 总之, 按照定义 6.1.1 可推得, 满足 1 个、2 个、3 个、4 个穆尔-彭诺斯方程的广义逆矩阵共有 15 种, 即

$$C_1^1 + C_2^1 + C_3^1 + C_4^1 = 15.$$

但应用较多的是以下 5 种

$$A\{1\}, A\{1,2\}, A\{1,3\}, A\{1,4\}, A\{1,2,3,4\}.$$

下面将会看到, 只有 $A\{1,2,3,4\}$ 是唯一确定的, 其他各种广义逆矩阵都不惟一确定, 每一种广义逆矩阵又都包含着一类矩阵, 分述如下:

- (1) $A\{1\}$: 其中任意一个确定的广义逆, 称作减号逆, 或 g 逆, 记为 A^- ;
- (2) $A\{1,2\}$: 其中任意一个确定的广义逆, 称作自反减号逆, 记为 $A_{\bar{r}}^-$;
- (3) $A\{1,3\}$: 其中任意一个确定的广义逆, 称作最小范数广义逆, 记为 $A_{\bar{m}}^-$;
- (4) $A\{1,4\}$: 其中任意一个确定的广义逆, 称作最小二乘广义逆, 记为 $A_{\bar{l}}^-$;
- (5) $A\{1,2,3,4\}$: 惟一的一个, 称作加号逆, 或伪逆, 或穆尔-彭诺斯逆, 记为 A^+ .

为叙述简单起见, 下面我们仅限于对实矩阵进行讨论. 类似地, 对于复矩阵也有相应的结果.

6.1.2 减号逆 A^-

定义 6.1.2 设有 $m \times n$ 实矩阵 A ($m \leq n$, 当 $m > n$ 时, 可讨论 A^T). 若有一个 $n \times m$ 实矩阵(记为 A^-)存在, 使下式成立, 则称 A^- 为 A 的减号逆或 g 逆:

$$AA^-A = A. \quad (6.1.5)$$

当 A^{-1} 存在时, 显然 A^{-1} 满足上式, 可见减号逆 A^- 是普通逆矩阵 A^{-1} 的推广; 另外, 由 $AA^-A=A$ 得

$$(AA^-A)^T = A^T, \quad \text{即} \quad A^T(A^-)^T A^T = A^T.$$

可见, 当 A^- 为 A 的一个减号逆时, $(A^-)^T$ 就是 A^T 的一个减号逆.

例 6.1.1 设 $A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, $C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

易知

$$ABA = A, \quad ACA = A,$$

故 B 与 C 均为 A 的减号逆.

例 6.1.2 若 $A = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{m \times n}$, 则

$$A^- = \begin{bmatrix} I_r & * \\ * & * \end{bmatrix}_{n \times m},$$

其中 * 任意选取.

证明 因为对任意的 $\begin{bmatrix} I_r & * \\ * & * \end{bmatrix}_{n \times m}$, 都有

$$\begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{m \times n} \begin{bmatrix} I_r & * \\ * & * \end{bmatrix}_{n \times m} \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{m \times n} = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{m \times n},$$

所以

$$A^- = \begin{bmatrix} I_r & * \\ * & * \end{bmatrix}_{n \times m}.$$

反之,任意的 $G = \begin{bmatrix} G_1 & G_2 \\ G_3 & G_4 \end{bmatrix}_{n \times m}$, 若满足

$$\begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} G_1 & G_2 \\ G_3 & G_4 \end{bmatrix} \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

必须有 $G_1 = I_r$, 即 G 为 $\begin{bmatrix} I_r & * \\ * & * \end{bmatrix}$ 的形状.

证毕

例 6.1.2 表明,标准形 $\begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ 的减号逆存在,而且不是惟一的,填一些数到 * 位置,就是一个减号逆,填不同数,就得到不同减号逆.

下面我们讨论当 A 为非零矩阵时,如何用初等变换的方法来构造它的任意一个减号逆,即讨论 A^- 的存在性.

引理 设 $B_{m \times n} = P_{m \times m} A_{m \times n} Q_{n \times n}$, 其中 P, Q 都是满秩方阵,如果已知 B 的减号逆为 B^- , 则矩阵 A 的减号逆

$$A^- = QB^-P. \quad (6.1.6)$$

证明 因为已知 B^- 是 B 的减号逆, 所以有

$$BB^-B = B,$$

即

$$(PAQ)B^-(PAQ) = PAQ,$$

由于 P 与 Q 非奇异, 故有

$$A(QB^-P)A = A,$$

从而有

$$A^- = QB^-P. \quad \text{证毕}$$

这个引理说明,两个相抵的矩阵 A, B (即满足 $B=PAQ$), 如果其中一个的减号逆可求出来, 那么, 另一个的减号逆也可以求出来.

定理 6.1.1 (存在性) 任给 $m \times n$ 矩阵 A , 那么减号逆 A^- 一定存在, 但不惟一.

证明 分两种情况讨论, 如果 $\text{rank} A = 0$, 即 $A = \mathbf{0}_{m \times n}$, 这时对任意的 $X \in \mathbb{R}^{n \times m}$, 都有 $\mathbf{0}X\mathbf{0} = \mathbf{0}$, 所以任意 $n \times m$ 矩阵 X 都是零矩阵的减号逆.

再设 $\text{rank} \mathbf{A} = r < \min\{m, n\}$ ($r > 0$), 那么存在 m 阶满秩矩阵 \mathbf{P} 与 n 阶满秩矩阵 \mathbf{Q} , 使得

$$\mathbf{PAQ} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \mathbf{B} \in \mathbb{R}^{m \times n}.$$

由例 6.1.2 知, 存在

$$\mathbf{B} = \begin{bmatrix} \mathbf{I}_r & * \\ * & * \end{bmatrix} \quad (* \text{ 可任意选取}).$$

再由引理知, 存在

$$\mathbf{A}^- = \mathbf{Q} \begin{bmatrix} \mathbf{I}_r & * \\ * & * \end{bmatrix} \mathbf{P},$$

只要 \mathbf{A} 非满秩. 由于 $*$ 的任意性, 所以 \mathbf{A}^- 不惟一.

证毕

这个定理的证明过程, 实际上是给出了一种求矩阵减号逆 \mathbf{A}^- 的方法.

例 6.1.3 设 $\mathbf{A} = \begin{bmatrix} 1 & -1 & 2 \\ 2 & 2 & 3 \end{bmatrix}$, 求 \mathbf{A}^- .

解 为将 \mathbf{A} 通过初等行与列变换, 化为一个等价的标准形, 我们在 \mathbf{A} 的右边放上一个 \mathbf{I}_2 , 在 \mathbf{A} 的下方放上一个 \mathbf{I}_3 , 当 \mathbf{A} 变成 \mathbf{I}_r 时, 则 \mathbf{I}_2 就变成 \mathbf{P} , 而 \mathbf{I}_3 就变成 \mathbf{Q} .

$$\begin{aligned} \left(\begin{array}{ccc|cc} \mathbf{A} & \mathbf{I}_2 \\ \mathbf{I}_3 & \mathbf{0} \end{array} \right) &= \left(\begin{array}{ccc|cc} 1 & -1 & 2 & 1 & 0 \\ 2 & 2 & 3 & 0 & 1 \\ \hline 1 & 0 & 0 & & \\ 0 & 1 & 0 & & \\ 0 & 0 & 1 & & \end{array} \right) \xrightarrow[\substack{c_3 + (-2)c_1}]{c_1 + c_2} \left(\begin{array}{ccc|cc} 1 & 0 & 0 & 1 & 0 \\ 2 & 4 & -1 & 0 & 1 \\ \hline 1 & 1 & -2 & & \\ 0 & 1 & 0 & & \\ 0 & 0 & 1 & & \end{array} \right) \\ &\xrightarrow[\substack{c_2 + 4c_3}]{c_1 + 2c_3} \left(\begin{array}{ccc|cc} 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ \hline -3 & -7 & -2 & & \\ 0 & 1 & 0 & & \\ 2 & 4 & 1 & & \end{array} \right) \xrightarrow{c_2 \leftrightarrow c_3} \left(\begin{array}{ccc|cc} 1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \\ \hline -3 & -2 & -7 & & \\ 0 & 0 & 1 & & \\ 2 & 1 & 4 & & \end{array} \right) \\ &\xrightarrow{(-1)r_2} \left(\begin{array}{ccc|cc} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ \hline -3 & -2 & -7 & & \\ 0 & 0 & 1 & & \\ 2 & 1 & 4 & & \end{array} \right). \end{aligned}$$

因此有

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \mathbf{A} \begin{bmatrix} -3 & -2 & -7 \\ 0 & 0 & 1 \\ 2 & 1 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

即

$$\mathbf{PAQ} = [\mathbf{I}_2 \quad \mathbf{0}] = \mathbf{B},$$

其中

$$P = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad Q = \begin{bmatrix} -3 & -2 & -7 \\ 0 & 0 & 1 \\ 2 & 1 & 4 \end{bmatrix}.$$

但标准形 B 的减号逆为

$$B^- = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ * & * \end{pmatrix} \quad (* \text{ 为任意选取的实数}),$$

故得

$$A^- = Q \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ * & * \end{pmatrix} P \quad (* \text{ 为任意实数}).$$

如果取 $*$ 为 0, 则 $A^- = Q \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} P = \begin{pmatrix} -3 & 2 \\ 0 & 0 \\ 2 & -1 \end{pmatrix}$, 这不过是其中的一个减号逆.

设有 $A \in \mathbb{R}^{m \times n}$, 我们来看 $\text{rank} A$ 与 $\text{rank} A^-$ 之间的关系.

定理 6.1.2 $\text{rank} A^- \geq \text{rank} A$.

证明 因为 $AA^-A = A$, 即 $(AA^-)A = A$, 所以有

$$\text{rank}(AA^-) \geq \text{rank} A.$$

又因为 $\text{rank} A^- \geq \text{rank}(AA^-)$, 故

$$\text{rank} A^- \geq \text{rank}(AA^-) \geq \text{rank} A.$$

证毕

这个定理说明, A^- 的秩总不会小于 A 的秩, 这从例 6.1.2 也可看出.

6.1.3 自反减号逆 A_r^-

众所周知, 对于普通的逆矩阵 A^{-1} , 有 $(A^{-1})^{-1} = A$, 但这一事实对于减号逆 A^- 一般不成立. 例如, 由例 6.1.1 知

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad A^- = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

但

$$A^-AA^- = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \neq A^-,$$

即 $(A^-)^- \neq A$. 为了使 A 与 A^- 能互为减号逆, 我们不妨对前面式(6.1.5)定义的减号逆 A^- 加以限制, 使 A^- 具有这种“自反”的性质. 为此下面我们给出自反减号逆矩阵的定义.

定义 6.1.3 对于一个 $m \times n$ 实矩阵 A , 使

$$AGA = A \quad \text{及} \quad GAG = G$$

同时成立的 $n \times m$ 实矩阵 G , 称为是 A 的一个自反减号逆, 用 A_r^- 表示, 即有

$$AA_r^-A = A \quad \text{及} \quad A_r^-AA_r^- = A_r^-.$$

显然, A_r^- 是一种特殊的减号逆. 此时, 它满足自反性质 $(A_r^-)_r^- = A$, 以后在求减号逆时, 常常求的是这种具有自反性质的减号逆.

下面我们介绍构造自反减号逆的方法. 先引进所谓“矩阵的右逆、左逆”的概念.

1. 矩阵的右逆 A_R^{-1} 和左逆 A_L^{-1}

若 $G \in \mathbb{R}^{n \times m}$, $A \in \mathbb{R}^{m \times n}$ 使 $AG = I$, 或 $GA = I$, 则 G 必为 A 的减号逆(因为 $AGA = A$), 但这样的 G 并不一定是 A 的逆, 因此可以引入下面的定义:

定义 6.1.4 设 $A \in \mathbb{R}^{m \times n}$, 若有 $G \in \mathbb{R}^{n \times m}$, 使得

$$AG = I \quad \text{或} \quad GA = I, \quad (6.1.7)$$

则称 G 为 A 的右逆(或左逆), 记为 A_R^{-1} (或 A_L^{-1}), 即

$$AA_R^{-1} = I \quad \text{或} \quad A_L^{-1}A = I.$$

在一般情况下, $A_R^{-1} \neq A_L^{-1}$. 若 $A_R^{-1} = A_L^{-1}$, 则 A^{-1} 存在, 且 $A^{-1} = A_R^{-1} = A_L^{-1}$.

定理 6.1.3 (1) 设 A 是行最大秩(行满秩)的 $m \times n$ 实矩阵($m \leq n$), 则必存在 A 的右逆, 且

$$A_R^{-1} = A^T(AA^T)^{-1}; \quad (6.1.8)$$

(2) 设 A 是列最大秩(列满秩)的 $n \times m$ 实矩阵($m \geq n$), 则必存在 A 的左逆, 且

$$A_L^{-1} = (A^T A)^{-1} A^T. \quad (6.1.9)$$

证明 因为 A 是行(或列)满秩矩阵, 所以 AA^T (或 $A^T A$) 为满秩方阵, 故有

$$AA^T(AA^T)^{-1} = I_{m \times m} = (AA^T)^{-1}AA^T$$

或

$$(A^T A)^{-1}A^T A = I_{n \times n} = A^T A(A^T A)^{-1},$$

所以有

$$A_R^{-1} = A^T(AA^T)^{-1}$$

或

$$A_L^{-1} = (A^T A)^{-1}A^T. \quad \text{证毕}$$

这里要指出的是, 对于行(或列)满秩的 $m \times n$ 矩阵 A , A_R^{-1} 和 A_L^{-1} 是不可能同时存在的; 当且仅当 $m = n$ 且 A 为满秩矩阵时, A_R^{-1} 和 A_L^{-1} 才同时存在, 并且都等于逆矩阵 A^{-1} (从式(6.1.8)、式(6.1.9)可直接看出).

另外, 由式(6.1.8)(或式(6.1.9))所定义的右逆(或左逆)满足 M-P 方程(6.1.1)~(6.1.4), 即

$$(1) \quad AA_R^{-1}A = A \quad (AA_L^{-1}A = A);$$

$$(2) \quad A_R^{-1}AA_R^{-1} = A_R^{-1} \quad (A_L^{-1}AA_L^{-1} = A_L^{-1});$$

$$(3) \quad (A_R^{-1}A)^T = A_R^{-1}A \quad (A_L^{-1}A)^T = A_L^{-1}A;$$

$$(4) \quad (AA_R^{-1})^T = AA_R^{-1} \quad (AA_L^{-1})^T = AA_L^{-1}.$$

证明 以(1),(3)为例证明, 其余读者补证.

$$(1) \quad AA_R^{-1}A = A[A^T(AA^T)^{-1}]A = AA^T(AA^T)^{-1}A = A;$$

$$(3) \quad (A_R^{-1}A)^T = [A^T(AA^T)^{-1}A]^T = A^T[(AA^T)^{-1}]^T A \\ = A^T[(AA^T)^T]^{-1}A = A^T[AA^T]^{-1}A \\ = A_R^{-1}A.$$

换言之, 对于行(或列)满秩的矩阵 A , 按式(6.1.8)计算出的右逆(或按式(6.1.9)计算的左逆)不但是 A 的自反减号逆, 而且也是 A 的最小范数广义逆、最小二乘广义逆和加号逆.

2. A_r^- 的计算方法

显然,自反减号逆必为减号逆,但反之不真.下面介绍计算自反减号逆 A_r^- 的方法,以后我们将会看到,这里讲的这些方法可以类似地推广到求 A_m^- , A_l^- 和 A^+ 中去,具有相当的普遍性.

设 $A \in \mathbb{R}^{m \times n}$, 分两种情况讨论:

(1) 若 A 是行(或列)满秩矩阵,即 $\text{rank} A = m \leq n$ (或 $\text{rank} A = n \leq m$), 则

$$A_r^- = A_R^{-1} = A^T(AA^T)^{-1} \quad (\text{或 } A_r^- = A_L^{-1} = (A^T A)^{-1} A^T);$$

(2) 若 A 既不是行满秩也不是列满秩,即

$\text{rank} A = r < \min\{m, n\}$ ($r > 0$), 则 A 进行一系列的初等变换,可变成标准形

$$PAQ = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

其中 I_r 为 r 阶单位矩阵,即

$$A = P^{-1} \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} Q^{-1} = P^{-1} \begin{bmatrix} I_r \\ \mathbf{0} \end{bmatrix} (I_r, \mathbf{0}) Q^{-1}.$$

然后令

$$B = P^{-1} \begin{bmatrix} I_r \\ \mathbf{0} \end{bmatrix}, \quad C = (I_r, \mathbf{0}) Q^{-1},$$

从而有

$$A = BC^{\textcircled{1}}. \quad (6.1.10)$$

计算

$$B_L^{-1} = (B^T B)^{-1} B^T, \quad C_R^{-1} = C^T (C C^T)^{-1},$$

于是

$$A_r^- = C_R^{-1} B_L^{-1}. \quad (6.1.11)$$

下面我们证明由式(6.1.11)构造出来的 A_r^- 的确是一个自反减号逆,即它满足 M-P 方程中的第一和第二个.事实上

由于

$$\begin{aligned} AA_r^- A &= BCC_R^{-1} B_L^{-1} BC = BC = A, \\ A_r^- AA_r^- &= C_R^{-1} B_L^{-1} BCC_R^{-1} B_L^{-1} = C_R^{-1} B_L^{-1} = A_r^-, \end{aligned}$$

故 $A_r^- = C_R^{-1} B_L^{-1}$ 是 A 的自反减号逆.

例 6.1.4 设

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

试求 A 的自反减号逆.

① A 的这种分解称为“最大秩分解”,又称为满秩分解,其中 B 为列满秩矩阵,正好取 P^{-1} 的前 r 列; C 为行满秩矩阵,正好取 Q^{-1} 的前 r 行.详见第 3 章.

解 因为 $\text{rank} \mathbf{A} = 2$, 所以 \mathbf{A} 为行满秩矩阵, 属于第一种情况, 故

$$\begin{aligned} \mathbf{A}_r^- &= \mathbf{A}_R^{-1} = \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} \\ &= \begin{pmatrix} 1 & 0 \\ 2 & -1 \\ -1 & 2 \end{pmatrix} \left(\begin{bmatrix} 1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{pmatrix} 1 & 0 \\ 2 & -1 \\ -1 & 2 \end{pmatrix} \right)^{-1} \\ &= \begin{pmatrix} 1 & 0 \\ 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{bmatrix} 6 & -4 \\ -4 & 5 \end{bmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ 2 & -1 \\ -1 & 2 \end{pmatrix} \cdot \frac{1}{14} \begin{bmatrix} 5 & 4 \\ 4 & 6 \end{bmatrix} \\ &= \frac{1}{14} \begin{pmatrix} 5 & 4 \\ 6 & 2 \\ 3 & 8 \end{pmatrix}. \end{aligned}$$

例 6.1.5 设

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

试求 \mathbf{A} 的自反减号逆.

解 因为 $\text{rank} \mathbf{A} = 2$, 所以 \mathbf{A} 为列满秩矩阵, 也属于第一种情况, 故

$$\begin{aligned} \mathbf{A}_r^- &= \mathbf{A}_L^{-1} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \\ &= \left(\begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 1 \end{pmatrix} \right)^{-1} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix} \\ &= \frac{1}{11} \begin{bmatrix} -4 & 7 & 1 \\ 7 & -4 & 1 \end{bmatrix}. \end{aligned}$$

例 6.1.6 设

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 2 \\ 2 & 4 & 0 \end{pmatrix},$$

求 \mathbf{A} 的自反减号逆.

解 因为 $\text{rank} \mathbf{A} = 2 < 3$, 所以 \mathbf{A} 既非行满秩矩阵又非列满秩矩阵, 属于第二种情况. 为此, 先对 \mathbf{A} 施行一系列的初等变换:

$$\left(\begin{array}{ccc|ccc} \mathbf{A} & \mathbf{I}_3 & & & & \\ \mathbf{I}_3 & & * & & & \end{array} \right) = \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 1 & 0 \\ 2 & 4 & 0 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & & & \\ 0 & 1 & 0 & & * & \\ 0 & 0 & 1 & & & \end{array} \right) \xrightarrow{\substack{r_3 + (-2)r_1 \\ (\frac{1}{2})r_2}} \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & -2 & 0 & 1 \\ \hline 1 & 0 & 0 & & & \\ 0 & 1 & 0 & & * & \\ 0 & 0 & 1 & & & \end{array} \right)$$

$$\begin{aligned}
 & \xrightarrow{c_2 + (-2)c_1} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & -2 & 0 & 1 \\ \hline 1 & -2 & 0 & & & \\ 0 & 1 & 0 & & * & \\ 0 & 0 & 1 & & & \end{array} \right) \xrightarrow{c_2 \leftrightarrow c_3} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & -2 & 0 & 1 \\ \hline 1 & 0 & -2 & & & \\ 0 & 0 & 1 & & * & \\ 0 & 1 & 0 & & & \end{array} \right) \\
 & = \begin{bmatrix} \mathbf{A}_1 & \mathbf{P} \\ \mathbf{Q} & * \end{bmatrix},
 \end{aligned}$$

因此有

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

即

$$\mathbf{PAQ} = \mathbf{A}_1 = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (\text{标准形}),$$

其中

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 0 & 1 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

令

$$\begin{aligned}
 \mathbf{B} &= \mathbf{P}^{-1} \begin{bmatrix} \mathbf{I}_2 \\ \mathbf{0} \end{bmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \\
 &= \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 2 & 0 \end{pmatrix} \quad (\text{即为 } \mathbf{P}^{-1} \text{ 的前两列}),
 \end{aligned}$$

$$\begin{aligned}
 \mathbf{C} &= (\mathbf{I}_2 \quad \mathbf{0}) \mathbf{Q}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} 1 & 0 & -2 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\
 &= \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (\text{即为 } \mathbf{Q}^{-1} \text{ 的前两行}),
 \end{aligned}$$

于是

$$\mathbf{C}_R^{-1} = \mathbf{C}^T (\mathbf{C}\mathbf{C}^T)^{-1}$$

$$\begin{aligned}
&= \begin{pmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 1 \end{pmatrix} \left(\begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 1 \end{pmatrix} \right)^{-1} \\
&= \begin{pmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 1 \end{pmatrix} \cdot \frac{1}{5} \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix} = \frac{1}{5} \begin{pmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 5 \end{pmatrix}, \\
\mathbf{B}_L^{-1} &= (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \\
&= \left(\begin{bmatrix} 1 & 0 & 2 \\ 0 & 2 & 0 \end{bmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 2 & 0 \end{pmatrix} \right)^{-1} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 2 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 5 & 0 \\ 0 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 2 & 0 \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 4 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 2 & 0 \end{bmatrix} \\
&= \frac{1}{10} \begin{bmatrix} 2 & 0 & 4 \\ 0 & 5 & 0 \end{bmatrix}.
\end{aligned}$$

所以

$$\begin{aligned}
\mathbf{A}_r^- &= \mathbf{C}_R^{-1} \mathbf{B}_L^{-1} \\
&= \frac{1}{5} \begin{pmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 5 \end{pmatrix} \cdot \frac{1}{10} \begin{bmatrix} 2 & 0 & 4 \\ 0 & 5 & 0 \end{bmatrix} = \frac{1}{50} \begin{pmatrix} 2 & 0 & 4 \\ 4 & 0 & 8 \\ 0 & 25 & 0 \end{pmatrix}.
\end{aligned}$$

值得指出的是,由式(6.1.11)确定的自反广义逆 \mathbf{A}_r^- 并不惟一.这是因为用式(6.1.8)来计算右逆 \mathbf{C}_R^{-1} 和用式(6.1.9)来计算左逆 \mathbf{B}_L^{-1} (只不过其中的一种方法)并非惟一.下面不妨给出计算行(或列)满秩矩阵的右逆(或左逆)的一般表达式.

设 $\mathbf{A} \in \mathbb{R}^{m \times n}$, 且 $\text{rank} \mathbf{A} = m$, 则 \mathbf{A} 的右逆的一般表达式为

$$\mathbf{G} = \mathbf{V} \mathbf{A}^T (\mathbf{A} \mathbf{V} \mathbf{A}^T)^{-1}, \quad (6.1.12)$$

其中 \mathbf{V} 是使得等式 $\text{rank}(\mathbf{A} \mathbf{V} \mathbf{A}^T) = \text{rank} \mathbf{A}$ 成立的任意 n 阶方阵.

事实上,用 \mathbf{A} 左乘式(6.1.12)两端,得

$$\mathbf{A} \mathbf{G} = \mathbf{A} \mathbf{V} \mathbf{A}^T (\mathbf{A} \mathbf{V} \mathbf{A}^T)^{-1}.$$

由于 $\text{rank}(\mathbf{A} \mathbf{V} \mathbf{A}^T) = \text{rank} \mathbf{A} = m$, 所以 $\mathbf{A} \mathbf{V} \mathbf{A}^T$ 是最大秩方阵. 因此,有

$$\mathbf{A} \mathbf{G} = (\mathbf{A} \mathbf{V} \mathbf{A}^T) (\mathbf{A} \mathbf{V} \mathbf{A}^T)^{-1} = \mathbf{I},$$

即 $\mathbf{G} = \mathbf{V} \mathbf{A}^T (\mathbf{A} \mathbf{V} \mathbf{A}^T)^{-1}$ 是 \mathbf{A} 的右逆的一般表达式.

当取 $\mathbf{V} = \mathbf{I}_n$ 时,式(6.1.12)就变成了式(6.1.8). 所以由式(6.1.8)给出的 $\mathbf{A}_R^{-1} = \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1}$ 只是 \mathbf{A} 的所有右逆中的一个.

同理,可写出列最大秩矩阵 \mathbf{A} 的左逆的一般表达式

$$\mathbf{G} = (\mathbf{A}^T \mathbf{U} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{U}, \quad (6.1.13)$$

其中 \mathbf{U} 是使关系式 $\text{rank}(\mathbf{A}^T \mathbf{U} \mathbf{A}) = \text{rank} \mathbf{A}$ 成立的任意的 m 阶方阵.

6.1.4 最小范数广义逆 \mathbf{A}_m^-

定义 6.1.5 设 $\mathbf{A} \in \mathbb{R}^{m \times n}$ ($m \leq n$), 如果有一个 $n \times m$ 阶矩阵 \mathbf{G} , 满足

$$\mathbf{A} \mathbf{G} \mathbf{A} = \mathbf{A} \quad \text{及} \quad (\mathbf{G} \mathbf{A})^T = \mathbf{G} \mathbf{A}, \quad (6.1.14)$$

则称 G 为 A 的一个最小范数广义逆, 记为 A_m^- .

显然, 最小范数广义逆是用条件 $(GA)^T = GA$ 对减号逆 A^- 进行限制后所得出的一个特殊减号逆.

最小范数广义逆 A_m^- 通常有下列的计算方法:

(1) 当 A 为行(或列)满秩矩阵时, 则

$$A_m^- = A_R^- = A^T(AA^T)^{-1} \quad (\text{或 } A_m^- = A_L^- = (A^T A)^{-1} A^T).$$

(2) 当 $\text{rank}A = r < \min\{m, n\}$ 时, 将 A 满秩分解为 $A = BC$, 其中 B 为列满秩矩阵, C 为行满秩矩阵, 则 $A_m^- = C_R^{-1} B_L^-$.

事实上, 我们只要补充证明它满足 M-P 第三个方程:

$$\begin{aligned} (A_m^- A)^T &= (C_R^{-1} B_L^- BC)^T = [C^T (CC^T)^{-1} (B^T B)^{-1} B^T BC]^T = [C^T (CC^T)^{-1} C]^T \\ &= C^T (CC^T)^{-1} C = A_m^- A. \end{aligned}$$

所以它是最小范数广义逆.

在一般情况下, 用满秩分解来求 A_m^- 是很麻烦的(因为要求 P, Q, P^{-1}, Q^{-1} 计算量较大), 我们可以利用下面的方法.

(3) 对于 $A \in \mathbb{R}^{m \times n}$ (假定 $m \leq n$), 有

$$A_m^- = A^T (AA^T)^-. \quad (6.1.15)$$

证明 因 $(AA^T)^-$ 是一个减号逆, 故有

$$AA^T (AA^T)^- AA^T = AA^T.$$

设 $\text{rank}A = r$, 则按满秩分解有 $A = BC$, $\text{rank}B = \text{rank}C = r$, 以 $A = BC$ 代入上式, 使得

$$(BC)(BC)^T (AA^T)^- (BC)(BC)^T = (BC)(BC)^T,$$

即

$$BCC^T B^T (AA^T)^- BCC^T B^T = BCC^T B^T,$$

用 $B(B^T B)^{-1} (CC^T)^{-1} C$ 右乘上式两端, 得

$$BC(BC)^T (AA^T)^- B I_r I_r C = B I_r I_r C,$$

即

$$AA^T (AA^T)^- A = A,$$

故 $A^T (AA^T)^-$ 满足最小范数广义逆的第一个条件; 其次它也满足第二个条件, 因为有

$$(A^T (AA^T)^- A)^T = A^T (AA^T)^- A.$$

故 $A_m^- = A^T (AA^T)^-$ 为 A 的一个最小范数广义逆. 证毕

因为减号逆 $(AA^T)^-$ 不是惟一的, 所以最小范数广义逆 A_m^- 也不是惟一的.

最小范数广义逆具有下面的性质.

定理 6.1.4 条件(6.1.14)与下面关系式等价

$$GAA^T = A^T. \quad (6.1.16)$$

事实上, 对式(6.1.16)两端右乘 G^T , 得

$$GAA^T G^T = A^T G^T,$$

即

$$GA(GA)^T = (GA)^T.$$

对上式两端取转置, 得

$$(GA)(GA)^T = GA,$$

可见有

$$(\mathbf{GA})^T = \mathbf{GA},$$

代入式(6.1.16)得

$$(\mathbf{GA})^T \mathbf{A}^T = \mathbf{A}^T,$$

即有

$$\mathbf{AGA} = \mathbf{A}.$$

反之,我们可以由式(6.1.14)推出(6.1.16).这是因为 $(\mathbf{GA})^T = \mathbf{GA}$,将其置换式(6.1.14)中第一个条件左边的 \mathbf{GA} ,得

$$\mathbf{A}(\mathbf{GA})^T = \mathbf{A},$$

上式两端取转置,得

$$\mathbf{GAA}^T = \mathbf{A}^T.$$

例 6.1.7 设

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 2 & 0 & 2 \\ 2 & 4 & 6 \end{pmatrix},$$

求 \mathbf{A}_m^- .

解 因为 $\text{rank } \mathbf{A} = r = 2 < \min\{3, 4\}$,用满秩分解法来求 $\mathbf{A}_m^- = \mathbf{C}_R^{-1} \mathbf{B}_L^{-1}$.易求得

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ 0 & -2 & 1 & 0 \\ -2 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix},$$

使得

$$\mathbf{PAQ} = \mathbf{A}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

$$\text{其中 } \mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \text{再求 } \mathbf{P}^{-1} = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{pmatrix}, \mathbf{Q}^{-1} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{B} = \mathbf{P}^{-1} \begin{pmatrix} \mathbf{I}_2 \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 2 & 0 \\ 2 & 4 \end{pmatrix},$$

$$\mathbf{C} = (\mathbf{I}_2 \quad \mathbf{0}) \mathbf{Q}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$

$$\mathbf{C}_R^{-1} = \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1} = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix},$$

$$\mathbf{B}_L^{-1} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T = \frac{1}{10} \begin{bmatrix} 0 & 2 & 4 & 0 \\ 1 & -1 & -2 & 2 \end{bmatrix},$$

故

$$\mathbf{A}_m^- = \mathbf{C}_R^{-1}\mathbf{B}_L^{-1} = \frac{1}{30} \begin{pmatrix} -1 & 5 & 10 & -2 \\ 2 & -4 & -8 & 4 \\ 1 & 1 & 2 & 2 \end{pmatrix}.$$

例 6.1.8 对例 6.1.7 的 \mathbf{A} 用方法(3)求 \mathbf{A}_m^- .

解 $\mathbf{A}_m^- = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^-$.

令

$$\mathbf{B} = \mathbf{A}\mathbf{A}^T = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 2 & 0 & 2 \\ 2 & 4 & 6 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 & 2 \\ 2 & 0 & 0 & 4 \\ 3 & 1 & 2 & 6 \end{pmatrix}$$

$$= 2 \begin{pmatrix} 7 & 2 & 4 & 14 \\ 2 & 1 & 2 & 4 \\ 4 & 2 & 4 & 8 \\ 14 & 4 & 8 & 28 \end{pmatrix}.$$

先按式(6.1.6)求出 \mathbf{B} 的任一个减号逆(非自反). 为此, 对 \mathbf{B} 作一系列的初等变换, 易求得

$$\mathbf{P} = \frac{1}{2} \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & -\frac{2}{3} & 0 & 0 \\ 0 & -2 & 1 & 0 \\ -2 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} 0 & 1 & 0 & -2 \\ 1 & -2 & -2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

使得

$$\mathbf{P}\mathbf{B}\mathbf{Q} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

故

$$\mathbf{B}^- = \mathbf{Q} \begin{bmatrix} \mathbf{I}_2 & * \\ * & * \end{bmatrix} \mathbf{P} \quad (\text{为简单计, 取 } * \text{ 为零})$$

$$= \frac{1}{2} \begin{pmatrix} \frac{1}{3} & -\frac{2}{3} & 0 & 0 \\ -\frac{2}{3} & \frac{7}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

从而

$$\begin{aligned} \mathbf{A}_m^- &= \begin{pmatrix} 1 & 1 & 2 & 2 \\ 2 & 0 & 0 & 4 \\ 3 & 1 & 2 & 6 \end{pmatrix} \cdot \frac{1}{2} \begin{pmatrix} \frac{1}{3} & -\frac{2}{3} & 0 & 0 \\ -\frac{2}{3} & \frac{7}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} -\frac{1}{3} & \frac{5}{3} & 0 & 0 \\ \frac{2}{3} & -\frac{4}{3} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 \end{pmatrix}. \end{aligned}$$

从这里看出,用式(6.1.15)求 \mathbf{A}_m^- 避免了求 \mathbf{P}^{-1} 和 \mathbf{Q}^{-1} 以及左、右逆,因而减小了计算量. 尽管形式上与例 6.1.7 不一样,但它们均满足最小范数广义逆的定义.

6.1.5 最小二乘广义逆 \mathbf{A}_l^-

定义 6.1.6 设 $\mathbf{A} \in \mathbb{R}^{m \times n} (m \leq n)$, 若有一个 $n \times m$ 阶矩阵 \mathbf{G} 满足

$$\mathbf{AGA} = \mathbf{A} \quad \text{及} \quad (\mathbf{AG})^T = \mathbf{AG},$$

则称 \mathbf{G} 为 \mathbf{A} 的一个最小二乘广义逆, 记为 \mathbf{A}_l^- .

显然, 最小二乘广义逆也是减号逆的一种.

求 \mathbf{A}_l^- 有下列方法:

(1) 当 \mathbf{A} 行(或列)满秩时, 有

$$\mathbf{A}_l^- = \mathbf{A}_R^{-1} = \mathbf{A}^T(\mathbf{AA}^T)^{-1} \quad (\text{或 } \mathbf{A}_l^- = \mathbf{A}_L^{-1} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T).$$

(2) 当 $\text{rank} \mathbf{A} = r < \min\{m, n\}$, 将 \mathbf{A} 满秩分解成 $\mathbf{A} = \mathbf{BC}$, 其中 \mathbf{B} 列满秩, \mathbf{C} 行满秩, 则

$$\mathbf{A}_l^- = \mathbf{C}_R^{-1}\mathbf{B}_L^{-1}.$$

事实上, 我们只要补充证明它满足 M-P 第 4 个方程:

$$\begin{aligned} (\mathbf{AA}_l^-)^T &= [\mathbf{BCC}_R^{-1}\mathbf{B}_L^{-1}]^T = [\mathbf{BCC}^T(\mathbf{CC}^T)^{-1}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T]^T \\ &= [\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T]^T = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T = \mathbf{AA}_l^-. \end{aligned}$$

所以它构成最小二乘广义逆.

在一般情况下, 用满秩分解方法求 \mathbf{A}_l^- 比较麻烦, 此时可按下面方法来求.

(3) 对于 $\mathbf{A} \in \mathbb{R}^{m \times n} (m \leq n)$, 有

$$\mathbf{A}_l^- = (\mathbf{A}^T\mathbf{A})^- \mathbf{A}^T. \quad (6.1.17)$$

因 $(\mathbf{A}^T\mathbf{A})^-$ 是一个减号逆, 故有

$$\mathbf{A}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^- \mathbf{A}^T\mathbf{A} = \mathbf{A}^T\mathbf{A}.$$

设 $\text{rank} \mathbf{A} = r$, 则按秩分解有 $\mathbf{A} = \mathbf{BC}$, $\text{rank} \mathbf{B} = \text{rank} \mathbf{C} = r$, 以 $\mathbf{A} = \mathbf{BC}$ 代入上式, 使得

$$(\mathbf{BC})^T(\mathbf{BC})(\mathbf{A}^T\mathbf{A})^- (\mathbf{BC})^T\mathbf{BC} = (\mathbf{BC})^T\mathbf{BC},$$

即

$$\mathbf{C}^T\mathbf{B}^T\mathbf{BC}(\mathbf{A}^T\mathbf{A})^- \mathbf{C}^T\mathbf{B}^T\mathbf{BC} = \mathbf{C}^T\mathbf{B}^T\mathbf{BC}.$$

用 $\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}(\mathbf{CC}^T)^{-1}\mathbf{C}$ 左乘上式两端, 得

$$\mathbf{BI}_r\mathbf{I}_r\mathbf{C}(\mathbf{A}^T\mathbf{A})^- \mathbf{C}^T\mathbf{B}^T\mathbf{BC} = \mathbf{BI}_r\mathbf{I}_r\mathbf{C},$$

即

$$BC(A^T A)^{-1}(BC)^T BC = BC.$$

或写成

$$A(A^T A)^{-1} A^T A = A.$$

故 $(A^T A)^{-1} A^T$ 满足最小二乘广义逆的第一个条件; 其次它也满足第二个条件, 因为

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

故 $(A^T A)^{-1} A^T$ 为 A 的一个最小二乘广义逆.

证毕

因为减号逆 $(A^T A)^{-1}$ 不是惟一的, 故最小二乘广义逆也不是惟一的.

例 6.1.9 设

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

试求 A_1^- .

解 例 6.1.5 中的矩阵 A 即为此处的矩阵 A . 因为 $\text{rank} A = r = 2$, 故 A 为列满秩矩阵, 有

$$\begin{aligned} A_1^- &= A_1^{-1} = (A^T A)^{-1} A^T \\ &= \left(\begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 1 \end{pmatrix} \right)^{-1} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 6 & 5 \\ 5 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix} = \frac{1}{11} \begin{bmatrix} 6 & -5 \\ -5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix} \\ &= \frac{1}{11} \begin{bmatrix} -4 & 7 & 1 \\ 7 & -4 & 1 \end{bmatrix}. \end{aligned}$$

例 6.1.10 设

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 2 & 0 & 2 \\ 2 & 4 & 6 \end{pmatrix},$$

试求 A_1^- .

解 例 6.1.7 中的矩阵 A 即为此处的矩阵 A , 又由计算方法(2)知满秩分解所得的 A_1^- 也是 A_m^- (或 A_r^-), 故由前面的结果得

$$A_1^- = A_m^- = \frac{1}{30} \begin{pmatrix} -1 & 5 & 10 & -2 \\ 2 & -4 & -8 & 4 \\ 1 & 1 & 2 & 2 \end{pmatrix}.$$

6.1.6 加号逆 A^+

前面我们对减号逆 A^- 加以不同的限制, 得出减号逆的具有不同性质的减号逆, 如自反广义逆 A_r^- 、最小范数广义逆 A_m^- 、最小二乘广义逆 A_1^- 等. 其实, 还有一类更特殊也更为重要广义逆, 这就是将要介绍的加号逆 A^+ . 它的实质是在减号逆的条件 $AGA = A$ 的基础上用上述所有条件同时加以限制. 用这样的方式得出的 A^+ , 不仅在应用上特别重要, 而且有很多有趣的性质.

定义 6.1.7 设 $A \in \mathbb{R}^{m \times n}$, 若存在 $n \times m$ 阶矩阵 G , 同时满足

- (1) $AGA = A$;
- (2) $GAG = G$;
- (3) $(AG)^T = AG$;
- (4) $(GA)^T = GA$;

则称 G 为 A 的加号逆, 或伪逆, 或摩尔-彭诺斯逆, 记为 A^+ .

从定义可以看出, 加号逆必同时是减号逆、自反广义逆、最小范数广义逆和最小二乘广义逆. 在 4 个条件中, G 与 A 完全处于对称地位, 因此 A 也是 A^+ 的加号逆, 即有

$$(A^+)^+ = A.$$

另外可见, 加号逆很类似于通常的逆阵, 因为通常的逆 A^{-1} 也有下列 4 个类似的性质:

- (1) $AA^{-1}A = A$;
- (2) $A^{-1}AA^{-1} = A^{-1}$;
- (3) $AA^{-1} = I$;
- (4) $A^{-1}A = I$.

由定义 6.1.7 中的条件(3)和(4)还可看出, AA^+ 与 A^+A 都是对称矩阵.

下面来看一个例子.

例 6.1.11

- (1) 设 $0 \in \mathbb{R}^{m \times n}$, 则 $0^+ = 0 \in \mathbb{R}^{n \times m}$;
- (2) 设 $A = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ 是 n 阶方阵, 则 $A^+ = A$;
- (3) 设对角阵

$$A = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}, \quad \text{则} \quad A^+ = \begin{pmatrix} \lambda_1^+ & & \\ & \ddots & \\ & & \lambda_n^+ \end{pmatrix},$$

其中

$$\lambda_i^+ = \begin{cases} \frac{1}{\lambda_i}, & \text{当 } \lambda_i \neq 0 \text{ 时;} \\ 0, & \text{当 } \lambda_i = 0 \text{ 时.} \end{cases}$$

例如, $2^+ = \frac{1}{2}$, $(-\frac{1}{3})^+ = -3$, $0^+ = 0$ 等.

证明 (1) 显然, 只要证明(3), 那么(2)是(3)的直接推论. 下面证明(3). 不失一般性, 我们令

$$\lambda_1, \dots, \lambda_s \neq 0, \quad \lambda_{s+1} = \dots = \lambda_n = 0,$$

那么, 令

$$B = \begin{pmatrix} \lambda_1^{-1} & & & & \\ & \ddots & & & \\ & & \lambda_s^{-1} & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix}.$$

容易验证 B 满足定义 6.1.7 的 4 个条件,从而

$$A^+ = B.$$

证毕

定理 6.1.5 若 $A \in \mathbb{R}^{m \times n}$, 且 $A = BC$ 是最大秩分解, 则

$$X = C^T(CC^T)^{-1}(B^TB)^{-1}B^T \quad (6.1.18)$$

是 A 的加号逆.

证明

$$\begin{aligned} AXA &= BCC^T(CC^T)^{-1}(B^TB)^{-1}B^TBC \\ &= BC = A, \end{aligned}$$

$$\begin{aligned} XAX &= C^T(CC^T)^{-1}(B^TB)^{-1}B^TBCC^T(CC^T)^{-1}(B^TB)^{-1}B^T \\ &= C^T(CC^T)^{-1}(B^TB)^{-1}B^T = X, \end{aligned}$$

$$\begin{aligned} (AX)^T &= [BCC^T(CC^T)^{-1}(B^TB)^{-1}B^T]^T \\ &= [B(B^TB)^{-1}B^T]^T = B(B^TB)^{-1}B^T = AX, \end{aligned}$$

$$\begin{aligned} (XA)^T &= [C^T(CC^T)^{-1}(B^TB)^{-1}(B^TB)C]^T \\ &= [C^T(CC^T)^{-1}C]^T = C^T(CC^T)^{-1}C = XA, \end{aligned}$$

因此, 式(6.1.18)是加号逆.

证毕

推论 设 $A \in \mathbb{R}^{m \times n}$, $\text{rank } A = r$, 则

(1) 当 $r = n$ (即 A 列满秩) 时

$$A^+ = (A^T A)^{-1} A^T.$$

(2) 当 $r = m$ (即 A 行满秩) 时,

$$A^+ = A^T (AA^T)^{-1}.$$

这只要注意到 $A = AI_n = I_m A$ 即可.

定理 6.1.6 对于任意 $A \in \mathbb{R}^{m \times n}$, 其加号逆 A^+ 存在且惟一.

证明 令 $G = A_m^- AA_1^-$, 则 G 满足(由 A_m^-, A_1^- 性质)

$$(1) \quad AGA = AA_m^- AA_1^- A = AA_1^- A = A;$$

$$(2) \quad GAG = A_m^- AA_1^- AA_m^- AA_1^- = A_m^- AA_m^- AA_1^- = A_m^- AA_1^- = G;$$

$$(3) \quad (GA)^T = (A_m^- AA_1^- A)^T = (A_m^- A)^T = A_m^- A = GA;$$

$$(4) \quad (AG)^T = (AA_m^- AA_1^-)^T = (AA_1^-)^T = AA_1^- = AG.$$

显然, $A_m^- AA_1^-$ 就是 A 的加号逆. 下面证明惟一性.

设 X 与 Y 均是 A 的加号逆, 于是同时有

$$AXA = A, \quad AYA = A,$$

用 Y 右乘上面的第一式, 再利用 AY 和 AX 的对称性, 使得

$$AXAY = AY,$$

$$\begin{aligned} AY &= (AY)^T = (AXAY)^T = (AY)^T (AX)^T \\ &= AYAX = (AYA)X = AX, \end{aligned}$$

即

$$AY = AX.$$

类似地, 得

$$YA = XA,$$

用 Y 左乘等式 $AY = AX$, 并利用上式, 使得

$$YAY = YAX = XAX.$$

但是

$$YAY = Y, \quad XAX = X,$$

故最终得 $Y=X$, 这表明 A^+ 是惟一的.

证毕

推论 若 A 是 n 阶满秩方阵, 即 A^{-1} 存在, 则

$$A^+ = A^{-1} = A^-, \quad (6.1.19)$$

这是因为前面我们已直接验证 A^{-1} 满足定义 6.1.7 的 4 个条件, 再由 A^+ 的惟一性即知式(6.1.19)成立.

换句话说, 当 $|A| \neq 0$ 时, 这 3 种逆是统一的, 且是惟一的, 一般情况下, 当 A^{-1} 不存在时, A^+ 总是存在的, 而 A^- 存在但不惟一.

下面我们来证明 A^+ 的一些特殊性质.

定理 6.1.7 (1) $(A^T)^+ = (A^+)^T$.

(2) $A^+ = (A^T A)^+ A^T = A^T (A A^T)^+$.

(3) $(A^T A)^+ = A^+ (A^T)^+$.

(4) $\text{rank} A = \text{rank} A^+ = \text{rank} A^+ A = \text{rank} A A^+$.

证明 (1) 令 $X = (A^+)^T$, 下面证明 X 是 A^T 的加号逆.

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

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

$$(A^T X)^T = (A^T (A^+)^T)^T = A^+ A = (A^+ A)^T$$

$$= A^T (A^+)^T = A^T X.$$

类似地可证明 $(X A^T)^T = X A^T$.

(2) 设 $A = BC$, 则 $A^T A$ 的最大秩分解可写成

$$A^T A = C^T (B^T B C),$$

于是利用式(6.1.18), 有

$$\begin{aligned} (A^T A)^+ &= (B^T B C)^T (B^T B C C^T B^T B)^{-1} (C C^T)^{-1} C \\ &= C^T (B^T B) (B^T B)^{-1} (C C^T)^{-1} (B^T B)^{-1} (C C^T)^{-1} C \\ &= C^T (C C^T)^{-1} (B^T B)^{-1} (C C^T)^{-1} C. \end{aligned}$$

再利用式(6.1.18), 得

$$\begin{aligned} (A^T A)^+ A^T &= C^T (C C^T)^{-1} (B^T B)^{-1} (C C^T)^{-1} C (C^T B^T) \\ &= C^T (C C^T)^{-1} (B^T B)^{-1} (C C^T)^{-1} (C C^T) B^T \\ &= C^T (C C^T)^{-1} (B^T B)^{-1} B^T \\ &= A^+. \end{aligned}$$

同样可证 $A^T (A A^T)^+ = A^+$.

(3) $(A^T A)^+ = (A^T A)^+ A^T A (A^T A)^+ = A^+ [A (A^T A)^+]$

$$= A^+ (A^T)^+.$$

证毕

(4) 由 $A = A A^+ A, A^+ = A^+ A A^+$ 知

$$\begin{aligned} \text{rank} A &= \text{rank}(A A^+ A) \leq \text{rank}(A^+ A) \leq \text{rank} A^+ \\ &= \text{rank}(A^+ A A^+) \leq \text{rank}(A A^+) \leq \text{rank} A. \end{aligned}$$

注意,对于同阶可逆矩阵 \mathbf{A}, \mathbf{B} 有 $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$, 定理 6.1.6 之(3)表明对于特殊的矩阵 \mathbf{A} 和 \mathbf{A}^{\top} , 加号逆 $(\mathbf{A}^{\top}\mathbf{A})^{+}$ 有类似的性质. 但是一般来讲, 这个性质不成立, 即

$$(\mathbf{AB})^{+} \neq \mathbf{B}^{+}\mathbf{A}^{+}.$$

例如, 取

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

$$\text{则 } \mathbf{AB} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}.$$

不难验证

$$\mathbf{A}^{+} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{B}^{+} = \mathbf{B}^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix},$$

$$(\mathbf{AB})^{+} = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix},$$

但

$$\mathbf{B}^{+}\mathbf{A}^{+} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \neq (\mathbf{AB})^{+}.$$

此外, $(\mathbf{A}^2)^{+}$ 也未必等于 $(\mathbf{A}^{+})^2$. 例如, 设

$$\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix},$$

不难验证

$$\mathbf{A}^{+} = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}, \quad \text{显然 } \mathbf{A}^2 = \mathbf{A}, \quad \text{故 } (\mathbf{A}^2)^{+} = \mathbf{A}^{+} = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}.$$

但

$$(\mathbf{A}^{+})^2 = \frac{1}{4} \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix} \neq (\mathbf{A}^2)^{+}.$$

下面介绍 \mathbf{A}^{+} 的各种算法, 有些算法前面虽然已讲过, 但为了完整起见, 现综述如下:

- (1) 如果 \mathbf{A} 为满秩方阵, 则 $\mathbf{A}^{+} = \mathbf{A}^{-1}$;
- (2) 如果 $\mathbf{A} = \text{diag}(d_1, d_2, \dots, d_n)$, $d_i \in \mathbb{R}$ ($i=1, 2, \dots, n$), 则

$$\mathbf{A}^{+} = \text{diag}(d_1^{+}, d_2^{+}, \dots, d_n^{+}),$$

其中

$$d_i^{+} = \begin{cases} 0, & \text{当 } d_i = 0 \text{ 时,} \\ \frac{1}{d_i}, & \text{当 } d_i \neq 0 \text{ 时;} \end{cases}$$

- (3) 如果 \mathbf{A} 为行满秩矩阵, 则

$$\mathbf{A}^{+} = \mathbf{A}_{\mathbb{R}}^{-1} = \mathbf{A}^{\top}(\mathbf{A}\mathbf{A}^{\top})^{-1};$$

- (4) 如果 \mathbf{A} 为列满秩矩阵, 则

$$\mathbf{A}^{+} = \mathbf{A}_{\mathbb{L}}^{-1} = (\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top};$$

- (5) 如果 \mathbf{A} 为降秩的 $m \times n$ 矩阵, 可用满秩分解求 \mathbf{A}^{+} , 即将 \mathbf{A} 满秩分解成 $\mathbf{A} = \mathbf{BC}$, 其中 \mathbf{B} 列满秩, \mathbf{C} 行满秩, 且

$$\text{rank} \mathbf{B} = \text{rank} \mathbf{C} = r = \text{rank} \mathbf{A} < \min\{m, n\},$$

则有

$$\mathbf{A}^+ = \mathbf{C}_R^{-1} \mathbf{B}_L^{-1} = \mathbf{C}^+ \mathbf{B}^+, \quad (6.1.20)$$

这里

$$\mathbf{B}_L^{-1} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T, \quad \mathbf{C}_R^{-1} = \mathbf{C}^T (\mathbf{C} \mathbf{C}^T)^{-1}.$$

前面在介绍计算 \mathbf{A}_r^- , \mathbf{A}_m^- 和 \mathbf{A}_l^- 的方法时, 已验证了 $\mathbf{C}_R^{-1} \mathbf{B}_L^{-1}$ 满足 M-P 的 4 个方程, 所以它也必为加号逆, 即式 (6.1.20) 成立. 必须注意的是, 这里的 \mathbf{B}_L^{-1} 与 \mathbf{C}_R^{-1} 只能按式 (6.1.8) 和式 (6.1.9) 求出 (记作 $\mathbf{B}_L^{-1} = \mathbf{B}^+$, $\mathbf{C}_R^{-1} = \mathbf{C}^+$). 如果按左、右逆的通式写出别的形式作为 \mathbf{B}_L^{-1} 与 \mathbf{C}_R^{-1} , 就不能保证 \mathbf{A}^+ 是惟一的.

例 6.1.12 设

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

求其加号逆 \mathbf{A}^+ .

解 首先对 \mathbf{A} 进行满秩分解

$$\mathbf{A} \rightarrow \begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{2} & \frac{1}{2} \end{pmatrix},$$

所以 \mathbf{A} 的满秩分解为

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

由于

$$\mathbf{B}_L^{-1} = \mathbf{B}^{-1} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix},$$

而 $\mathbf{C}_R^{-1} = \mathbf{C}^T (\mathbf{C} \mathbf{C}^T)^{-1}$, 于是

$$\mathbf{C}_R^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{2}{3} & 0 & 0 \\ 0 & \frac{3}{4} & -\frac{1}{4} \\ 0 & -\frac{1}{4} & \frac{3}{4} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & 0 & 0 \\ 0 & \frac{3}{4} & -\frac{1}{4} \\ 0 & -\frac{1}{4} & \frac{3}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{4} \\ -\frac{1}{3} & \frac{1}{4} & \frac{1}{4} \end{pmatrix},$$

故

$$A^+ = C_R^{-1} B_L^{-1} = \begin{pmatrix} \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{4} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{6} & \frac{1}{12} & \frac{1}{6} \\ -\frac{1}{6} & \frac{5}{12} & -\frac{1}{6} \end{pmatrix}.$$

6.2 广义逆在解线性方程组中的应用

在这一节,我们将会看到广义逆理论能够把相容线性方程组的一般解、极小范数解以及矛盾方程组的最小二乘解、极小最小二乘解(最佳逼近解)全部概括和统一起来,从而,以线性代数古典理论所不曾有的姿态解决了一般线性方程组的求解问题.

6.2.1 线性方程组求解问题的提法

考虑非齐次线性方程组

$$Ax = b, \quad (6.2.1)$$

其中 $A \in \mathbb{C}^{m \times n}$, $b \in \mathbb{C}^m$ 给定,而 $x \in \mathbb{C}^n$ 为待定向量.

若 $\text{rank}(A \mid b) = \text{rank}A$, 则方程组(6.2.1)有解,或称方程组相容;否则,若 $\text{rank}(A \mid b) \neq \text{rank}A$, 则方程组(6.2.1)无解,或称方程组不相容或矛盾方程组.

关于线性方程组的求解问题,常见的有以下几种情形:

(1) 当方程组(6.2.1)在相容时,若系数矩阵 $A \in \mathbb{C}^{n \times n}$, 且非奇异(即 $\det A \neq 0$), 则有唯一的解

$$x = A^{-1}b. \quad (6.2.2)$$

但当 A 是奇异方阵或长方矩阵时,它的解不是惟一的. 此时 A^{-1} 不存在或无意义,那么我们自然会想到,这时是否也能用某个矩阵 G 把一般解(无穷多)表示成

$$x = Gb \quad (6.2.3)$$

的形式呢? 这个问题的回答是肯定的,我们将会发现 A 的减号逆 A^- 充当了这一角色.

(2) 如果方程组(6.2.1)相容,且其解有无穷多个,怎样求具有极小范数的解,即

$$\min_{Ax=b} \|x\|, \quad (6.2.4)$$

其中 $\|\cdot\|$ 是欧氏范数. 可以证明,满足该条件的解是惟一的,称之为极小范数解.

(3) 如果方程组(6.2.1)不相容,则不存在通常意义下的解,但在许多实际问题中,要求出这样的解

$$x = \min_{x \in \mathbb{C}^n} \|Ax - b\|, \quad (6.2.5)$$

其中 $\|\cdot\|$ 是欧氏范数. 我们称这个问题为求矛盾方程组的最小二乘问题,相应的 x 称为矛

盾方程组的最小二乘解.

(4) 一般来说, 矛盾方程组的最小二乘解是不惟一的. 但在最小二乘解的集合中, 具有极小范数的解

$$x = \min_{\min \|Ax-b\|} \|x\| \quad (6.2.6)$$

是惟一的, 称之为极小范数最小二乘解, 或最佳逼近解.

广义逆矩阵与线性方程组的求解有着极为密切的联系. 利用前一节的减号逆 A^- (特别有用的是自反减号逆 A_r^-)、最小范数广义逆 A_m^- 、最小二乘广义逆 A_l^- 以及加号逆 A^+ 可以给出上述诸问题的解.

6.2.2 相容方程组的通解与 A^-

对于一个 $m \times n$ 的相容的线性方程组 (6.2.1), 不论系数矩阵 A 是方阵还是长方矩阵, 是满秩的还是降秩的, 我们都会有一个标准的求解方法, 并且能把它的解表达成非常简洁的形式. 下面用定理形式给出.

定理 6.2.1 如果线性方程组 (6.2.1) 是相容的, A^- 是 A 的任一个减号逆, 则线性方程组 (6.2.1) 的一个特解可表示成

$$x = A^- b, \quad (6.2.7)$$

而通解可以表示成

$$x = A^- b + (I - A^- A)z, \quad (6.2.8)$$

其中 z 是与 x 同维的任意向量.

证明 因为 $Ax=b$ 相容, 所以必有一个 n 维向量 w , 使

$$Aw = b \quad (6.2.9)$$

成立. 又由于 A^- 是 A 的一个减号逆, 所以 $AA^-A=A$, 则有 $AA^-Aw=Aw$, 亦即 $AA^-b=b$, 由此得出

$$x = A^- b$$

是方程组 (6.2.1) 的一个特解.

其次, 在式 (6.2.8) 两端左乘 A , 则有

$$Ax = AA^-b + A(I - A^-A)z = AA^-b,$$

由于 $A(A^-b)=b$, 所以式 (6.2.8) 确定的 x 是方程组 (6.2.1) 的解. 而且当 \tilde{x} 为任意一个解时, 若令 $z = \tilde{x} - A^-b$, 则有

$$\begin{aligned} (I - A^-A)z &= (I - A^-A)(\tilde{x} - A^-b) \\ &= \tilde{x} - A^-b - A^-A\tilde{x} + A^-AA^-b \\ &= \tilde{x} - A^-b - A^-b + A^-b \\ &= \tilde{x} - A^-b, \end{aligned}$$

从而得

$$\tilde{x} = A^-b + (I - A^-A)z.$$

这表明由式 (6.2.8) 确定的解是方程组 (6.2.1) 的通解. 证毕

特别地, 当 $b=0$ 时, $Ax=b=0$ 即为齐次线性方程组, 而齐次线性方程组总是有解的, 因此, 有如下的结果.

推论 齐次线性方程组

$$Ax = 0$$

的通解为

$$x = (I - A^{-1}A)z. \quad (6.2.10)$$

注 由式(6.2.8)、式(6.2.10)可得相容方程组解的结构为:线性非齐次方程组的通解为它的一个特解加上对应的齐次线性方程组的通解.

例 6.2.1 求解

$$\begin{cases} x_1 + 2x_2 - x_3 = 1, \\ -x_2 + 2x_3 = 2. \end{cases} \quad (6.2.11)$$

解 将方程组写成矩阵形式

$$Ax = b,$$

其中

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

由于 $\text{rank}A = \text{rank}(A \mid b) = 2$, 所以方程组是相容的. 又因为自反减号逆也是一种减号逆, 且此处 A 为行满秩矩阵, 因此利用 6.1 节例 6.1.4 的结果, 有

$$A^{-} = A_r^{-} = A^T(AA^T)^{-1} = \frac{1}{14} \begin{bmatrix} 5 & 4 \\ 6 & 2 \\ 3 & 8 \end{bmatrix}.$$

利用公式(6.2.8)可立即求得方程组(6.2.11)的通解:

$$\begin{aligned} x &= A^{-}b + (I - A^{-}A)z \\ &= \frac{1}{14} \begin{bmatrix} 13 + 9z_1 - 6z_2 - 3z_3 \\ 10 - 6z_1 + 4z_2 + 2z_3 \\ 19 - 3z_1 + 2z_2 + z_3 \end{bmatrix}, \end{aligned}$$

即

$$\begin{cases} x_1 = \frac{1}{14}(13 + 9z_1 - 6z_2 - 3z_3), \\ x_2 = \frac{1}{14}(10 - 6z_1 + 4z_2 + 2z_3), \\ x_3 = \frac{1}{14}(19 - 3z_1 + 2z_2 + z_3), \end{cases}$$

其中

$$z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \text{ 为任意向量.}$$

例 6.2.2 求齐次线性方程组

$$\begin{cases} x_1 + 2x_2 - x_3 = 0, \\ -x_2 + 2x_3 = 0 \end{cases}$$

的通解.

解 由例 6.2.1 的结果及公式(6.2.10),得

$$\begin{aligned} x &= (I - A^- A)z \\ &= \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \frac{1}{14} \begin{pmatrix} 5 & 4 \\ 6 & 2 \\ 3 & 8 \end{pmatrix} \begin{pmatrix} 1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \right] \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} \\ &= \frac{1}{14} \begin{pmatrix} 9z_1 - 6z_2 - 3z_3 \\ -6z_1 + 4z_2 + 2z_3 \\ -3z_1 + 2z_2 + z_3 \end{pmatrix}, \end{aligned}$$

其中 $z = (z_1, z_2, z_3)^T$ 为 \mathbb{C}^3 中任意常数.

从上面两例子可以看出,用减号逆来表示相容方程组的通解 $x = A^- b + (I - A^- A)z$ 是很方便的,这是线性方程组理论的一个重大发展.但是,如何在无穷多个解向量中求出一个长度最短的解向量呢?这便是下面要研究的极小范数解.

6.2.3 相容方程组的极小范数解与 A_m^-

定义 6.2.1 对于相容的线性方程组 $Ax = b$,如果存在与 b 无关的 A 的某些特殊减号逆 G ,使得 Gb 和其他的解相比较,具有最小范数,即

$$\|Gb\|_2 \leq \|x\|_2, \quad (6.2.12)$$

其中 x 是 $Ax = b$ 的解. $\|\cdot\|_2$ 是欧几里得范数,则我们称 $x = Gb$ 为极小范数解,简记为 LN 解.

现在要问:相容方程组 $Ax = b$ 的极小范数解可以用什么样的广义逆来表示?极小范数解是否惟一?

定理 6.2.2 在相容线性方程组 $Ax = b$ 的一切解中具有极小范数解的充要条件是

$$x = A_m^- b, \quad (6.2.13)$$

其中 A_m^- 是 A 的最小范数广义逆.

证明 先证必要性.设 G 是 A 的减号逆,那么 $Ax = b$ 的一般解是 $Gb + (I - GA)z$, z 是任意向量.如果 Gb 具有最小范数,则对任意向量 z 及一切与 A 相容的向量 b ,有

$$\|Gb\|_2 \leq \|Gb + (I - GA)z\|_2.$$

或者等价地,对任意向量 z 及任意解向量 \bar{x} ,有

$$\|GA\bar{x}\|_2 \leq \|GA\bar{x} + (I - GA)z\|_2, \quad (6.2.14)$$

其中 $b = A\bar{x}$,不等式(6.2.14)意味着如下关系是成立的:

$$(GA\bar{x}, GA\bar{x})^{1/2} \leq (GA\bar{x} + (I - GA)z, GA\bar{x} + (I - GA)z)^{1/2},$$

即

$$(GA\bar{x}, GA\bar{x}) \leq (GA\bar{x} + (I - GA)z, GA\bar{x} + (I - GA)z),$$

或

$$\begin{aligned} (GA\bar{x}, GA\bar{x}) &\leq (GA\bar{x}, GA\bar{x}) + 2(GA\bar{x}, (I - GA)z) + \\ &\quad ((I - GA)z, (I - GA)z), \end{aligned}$$

即

$$0 \leq ((I - GA)z, (I - GA)z) + 2(\bar{x}, (GA)^T(I - GA)z).$$

上式右边第一项是向量 $(I - GA)z$ 的范数的平方,恒大于等于零;第二项是任意解向量 \tilde{x} 与向量 $(GA)^T(I - GA)z$ 的内积.由于 \tilde{x}, z 的任意性,显然上述不等式成立的充要条件是:

$$(GA)^T(I - GA) = \mathbf{0},$$

由此推出

$$(GA)^T = (GA)^T GA,$$

两边转置得

$$GA = (GA)^T GA.$$

可见有

$$(GA)^T = GA$$

(即满足定义 6.1.5 第二个条件),从而 $G = A_m^-$,说明极小范数解的形式是 $x = A_m^- b$,定理的必要性得证.

关于定理的充分性,只要将上面的过程倒推回去便可以完成.

证毕

定理 6.2.3 相容的线性方程组 $Ax = b$,具有惟一的极小范数解.

证明 设 G_1 和 G_2 是 A 的两个不同的最小范数广义逆,由等价的公式(6.1.16)应有

$$G_1 A A^T = A^T, \quad G_2 A A^T = A^T,$$

所以

$$G_1 A A^T = G_2 A A^T,$$

$$(G_1 - G_2) A A^T = \mathbf{0}.$$

上式两边同时右乘以 $G_1^T - G_2^T$,得

$$(G_1 - G_2) A A^T (G_1^T - G_2^T) = \mathbf{0}$$

或

$$[(G_1 - G_2)A][(G_1 - G_2)A]^T = \mathbf{0}.$$

上式成立仅当 $(G_1 - G_2)A = \mathbf{0}$ 才有可能.因此有

$$(G_1 - G_2)A\tilde{x} = \mathbf{0} (\tilde{x} \text{ 为任意解向量}).$$

又由于 $A\tilde{x} = b$,所以有 $G_1 b = G_2 b$.这说明,不同的最小范数广义逆 G_1 和 G_2 ,按 $x = Gb$ 求得的极小范数解却是惟一的.

证毕

注 按定义, $x = A_m^- b$ 是相容 $Ax = b$ 的极小范数解,虽然 A_m^- 不是惟一的,但极小范数解 $A_m^- b$ 是惟一的(见例 6.2.4).

例 6.2.3 求方程组 $AX = b$ 的极小范数解,其中

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

解 由例 6.1.4 知,此矩阵 A 是行满秩矩阵,因此 AA^T 是满秩方阵, AA^T 的减号逆 $(AA^T)^-$ 就是普通的逆矩阵 $(AA^T)^{-1}$.所以由式(6.1.15)可作出一个最小范数广义逆

$$A_m^- = A^T (AA^T)^- = A^T (AA^T)^{-1},$$

而这正是行满秩矩阵的右逆,在本章例 6.1.4 中我们已经求得这个 A 的右逆为

$$A_R = A_m^- = \frac{1}{14} \begin{bmatrix} 5 & 4 \\ 6 & 2 \\ 3 & 8 \end{bmatrix}.$$

根据公式(6.2.13),我们可求得方程组的极小范数解

$$\mathbf{x} = \mathbf{A}_m^- \mathbf{b} = \frac{1}{14} \begin{pmatrix} 13 \\ 10 \\ 19 \end{pmatrix}. \quad (6.2.15)$$

又在例 6.2.1 中已求得这个方程组的一般解为

$$\mathbf{x} = \frac{1}{14} \begin{pmatrix} 13 + 9z_1 - 6z_2 - 3z_3 \\ 10 - 6z_1 + 4z_2 + 2z_3 \\ 19 - 3z_1 + 2z_2 + z_3 \end{pmatrix},$$

如果令 $z_1 = z_2 = 0, z_3 = 1$, 代入上式可得一特解

$$\mathbf{x} = \frac{1}{14} \begin{pmatrix} 10 \\ 12 \\ 20 \end{pmatrix}. \quad (6.2.16)$$

分别对式(6.2.15)、式(6.2.16)的 \mathbf{x} 求其范数有

$$\begin{aligned} \|\mathbf{x}\|_2 &= \sqrt{\mathbf{x}^T \mathbf{x}} \\ &= \frac{1}{14} \sqrt{13^2 + 10^2 + 19^2} = \frac{1}{14} \sqrt{630}, \\ \|\mathbf{x}\|_2 &= \sqrt{\mathbf{x}^T \mathbf{x}} = \frac{1}{14} \sqrt{10^2 + 12^2 + 20^2} = \frac{1}{14} \sqrt{644}. \end{aligned}$$

显然式(6.2.15)中 \mathbf{x} 的范数比式(6.2.16)中 \mathbf{x} 的范数要小.

例 6.2.4 求方程组

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 1, \\ x_1 + x_3 = 0, \\ 2x_1 + 2x_3 = 0, \\ 2x_1 + 4x_2 + 6x_3 = 2 \end{cases}$$

的极小范数解与通解.

解 由所给方程组可知

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 2 & 0 & 2 \\ 2 & 4 & 6 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 2 \end{pmatrix}.$$

由于

$$\text{rank} \mathbf{A} = \text{rank}(\mathbf{A} \vdots \mathbf{b}) = 2,$$

所以,方程组相容.由例 6.1.7 的结果知

$$\mathbf{A}_m^- = \mathbf{C}_R^{-1} \mathbf{B}_L^{-1} = \frac{1}{30} \begin{pmatrix} -1 & 5 & 10 & -2 \\ 2 & -4 & -8 & 4 \\ 1 & 1 & 2 & 2 \end{pmatrix},$$

故得方程组的极小范数解为

$$\mathbf{x} = \mathbf{A}_m^- \mathbf{b} = \frac{1}{30} \begin{pmatrix} -1 & 5 & 10 & -2 \\ 2 & -4 & -8 & 4 \\ 1 & 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{6} \\ \frac{1}{3} \\ \frac{1}{6} \end{pmatrix}.$$

注 如果 \mathbf{A}_m^- 取本章例 6.1.8 的结果, 则得到的极小范数解为

$$\mathbf{x} = \mathbf{A}_m^- \mathbf{b} = \frac{1}{2} \begin{pmatrix} -\frac{1}{3} & \frac{5}{3} & 0 & 0 \\ \frac{2}{3} & -\frac{4}{3} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{6} \\ \frac{1}{3} \\ \frac{1}{6} \end{pmatrix}.$$

可见, 尽管 \mathbf{A}_m^- 的形式不同, 但极小范数解却是相同的.

而方程组的通解为

$$\begin{aligned} \mathbf{x} &= \mathbf{G}\mathbf{b} + (\mathbf{I} - \mathbf{G}\mathbf{A})\mathbf{z} \\ &= \begin{pmatrix} -\frac{1}{6} \\ \frac{1}{3} \\ \frac{1}{6} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \frac{1}{30} \begin{pmatrix} -1 & 5 & 10 & -2 \\ 2 & -4 & -8 & 4 \\ 1 & 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 2 & 0 & 2 \\ 2 & 4 & 6 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} \\ &= \frac{1}{3} \begin{pmatrix} -\frac{1}{2} + z_1 + z_2 - z_3 \\ 1 + z_1 + z_2 - z_3 \\ \frac{1}{2} - z_1 - z_2 + z_3 \end{pmatrix}. \end{aligned}$$

6.2.4 矛盾方程组的最小二乘解与 \mathbf{A}_1^-

线性方程组理论告诉我们: 不相容的线性方程组是没有解的, 但是, 有了广义逆矩阵这个工具, 我们可以研究这类方程组的最优近似解的问题.

定义 6.2.2 对于不相容的线性方程组 $\mathbf{A}\mathbf{x} = \mathbf{b}$, 如果有这样的解 $\hat{\mathbf{x}}$, 使它的误差向量的 2 范数为最小:

$$\|\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}\|_2 = \min_{\mathbf{x} \in \mathbb{C}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|, \quad (6.2.17)$$

即

$$\|\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}\|_2 \leq \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2,$$

则称 $\hat{\mathbf{x}}$ 是方程组 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 的**最小二乘解**. 这是因为和任何其他近似解 \mathbf{x} 相比较, $\hat{\mathbf{x}}$ 所导致的误差平方和 $\|\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}\|_2^2$ 是最小的.

注 最小二乘解并不是方程组 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 的解. 现在的问题是: 是否有这样的矩阵 \mathbf{G} , 对于任意的向量 \mathbf{b} , 都使 $\mathbf{x} = \mathbf{G}\mathbf{b}$ 为方程组 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 的最小二乘解? 下面的定理回答了这一问题.

定理 6.2.4 不相容方程组 $Ax=b$ 有最小二乘解的充要条件是

$$x = A_1^- b,$$

其中 A_1^- 是 A 的最小二乘广义逆.

证明 先证必要性. 设 G 是一矩阵 (不必是矩阵 A 的减号逆). 如果 Gb 是不相容方程组 $Ax=b$ 的最小二乘解, 于是有

$$\|AGb - b\|_2 \leq \|Ax - b\|_2.$$

上式右边可以改写成

$$\begin{aligned} \|Ax - b\|_2 &= \|AGb - b + Ax - AGb\|_2 \\ &= \|(AG - I)b + A(x - Gb)\|_2 \\ &= \|(AG - I)b + A\bar{x}\|_2, \end{aligned}$$

其中 $\bar{x} = x - Gb$. 因此上述不等式可改写为

$$\begin{aligned} \|AGb - b\|_2 &= \|(AG - I)b\|_2 \\ &\leq \|(AG - I)b + A\bar{x}\|_2. \end{aligned}$$

仿照定理 6.2.2 的证明过程可知, 上述不等式成立的充要条件是

$$((AG - I)b, A\bar{x}) = (b, (AG - I)^T A\bar{x}) = 0,$$

而上式等于零的充要条件又是 $(AG - I)^T A = 0$, 即

$$A^T AG = A^T. \quad (6.2.18)$$

式(6.2.18)两边同时右乘以 A , 得 $A^T AGA = A^T A$, 所以有 $AGA = A$; 另外, 在式(6.2.18)两边同时左乘以 G^T , 得

$$G^T A^T AG = G^T A^T$$

或

$$(AG)^T AG = (AG)^T,$$

两边取转置, 并比较等式两边可得

$$(AG)^T = AG.$$

由最小二乘广义逆的定义知 $G = A_1^-$. 这说明不相容方程组 $Ax=b$ 的最小二乘解的形式是 $x = A_1^- b$, 定理必要性得证.

关于定理充分性证明, 读者可以自己完成. 证毕

必须注意, 矛盾方程组 (不相容方程组) 的最小二乘解 \hat{x} 导致的误差平方和 (即在最小二乘意义下) $\|A\hat{x} - b\|_2^2$ 是惟一的, 但是, 最小二乘解可以不惟一. 为此, 有下面的定理:

定理 6.2.5 不相容方程组 $Ax=b$ 的最小二乘解的通式为

$$\hat{x} = A_1^- b + (I - A_1^- A)z, \quad (6.2.19)$$

其中 z 是任意列向量.

证明 先证式(6.2.19)中的 \hat{x} 确为最小二乘解. 因为 $A_1^- b$ 是 $Ax=b$ 的最小二乘解, 所以 $\|AA_1^- b - b\|$ 取最小值, 而

$$\begin{aligned} A[A_1^- b + (I - A_1^- A)z] &= AA_1^- b + (A - AA_1^- A)z \\ &= AA_1^- b, \end{aligned}$$

所以, $\hat{x} = A_1^- b + (I - A_1^- A)z$ 也为最小二乘解.

再证 $Ax=b$ 的任一个最小二乘解 x_0 必可表示成式(6.2.19)的形式. 事实上, 由于 $x_0, A_1^- b$ 都是最小二乘解, 故有

$$\|Ax_0 - b\|_2 = \|AA_1^-b - b\|_2 = \min.$$

由于 A_1^- 满足 M-P 第 1 和第 4 个方程, 故有

$$\begin{aligned} (AA_1^-)^T &= AA_1^-, & AA_1^-A &= A, \\ (AA_1^- - I)^T A &= 0. \end{aligned} \quad (6.2.20)$$

考虑误差向量范数平方, 有

$$\begin{aligned} \|Ax_0 - b\|_2^2 &= \|AA_1^-b - b + A(x_0 - A_1^-b)\|_2^2 \\ &= \|AA_1^-b - b\|_2^2 + 2(AA_1^-b - b)^T A(x_0 - A_1^-b) + \\ &\quad \|A(x_0 - A_1^-b)\|_2^2, \end{aligned}$$

即有

$$\begin{aligned} \|Ax_0 - b\|_2^2 - \|AA_1^-b - b\|_2^2 \\ = \|A(x_0 - A_1^-b)\|_2^2 + 2b^T(AA_1^- - I)^T A(x_0 - A_1^-b). \end{aligned}$$

将式(6.2.20)代入上式有

$$\|Ax_0 - b\|_2^2 - \|AA_1^-b - b\|_2^2 = \|A(x_0 - A_1^-b)\|_2^2.$$

又由 $\|Ax_0 - b\|_2 = \|AA_1^-b - b\|_2 = \min$ 知

$$\|A(x_0 - A_1^-b)\|_2^2 = 0.$$

于是

$$A(x_0 - A_1^-b) = 0,$$

这说明 $x_0 - A_1^-b$ 为齐次方程组 $Ax=0$ 的一个解, 再由齐次方程组的通解公式(6.2.10)知, 存在 z_0 使得

$$x_0 - A_1^-b = (I - A_1^-A)z_0,$$

即

$$x_0 = A_1^-b + (I - A_1^-A)z_0. \quad \text{证毕}$$

如定理 6.2.5 所述, 不相容方程组的最小二乘解不是惟一的, 而由前面 6.1 节知道 A 的最小二乘广义逆 A_1^- 也不是惟一的. 现在要找出计算最小二乘广义逆 A_1^- 的通式.

引理 设 $A \in \mathbb{R}^{m \times n}$, A_1^- 是 A 的某个最小二乘广义逆, 则另一个矩阵 $G \in \mathbb{R}^{n \times m}$ 也是 A 的最小二乘广义逆的充要条件是

$$AG = AA_1^-. \quad (6.2.21)$$

证明 充分性: 设 G 满足式(6.2.21), 在式(6.2.21)两端右乘 A , 由于 A_1^- 是某个已知的最小二乘广义逆, 于是有

$$AGA = AA_1^-A = A,$$

故 G 满足 M-P 第 1 个方程. 又因为

$$(AG)^T = (AA_1^-)^T = AA_1^- = AG,$$

故 G 满足 M-P 第 4 个方程.

必要性: 设 G 也是最小二乘广义逆, 则有

$$\begin{aligned} AA_1^- &= AGAA_1^- = (AG)^T(AA_1^-)^T = G^T A^T (A_1^-)^T A^T \\ &= G^T (AA_1^-A)^T = G^T A^T = (AG)^T = AG, \end{aligned}$$

即式(6.2.21)成立.

定理 6.2.6 设 $A \in \mathbb{R}^{m \times n}$, A_1^- 是某个最小二乘广义逆, 则 A 的任何最小二乘广义逆都可表示成

$$G = A_1^- + (I_n - A_1^-A)U, \quad (6.2.22)$$

其中 U 是任意的 $n \times m$ 矩阵.

证明 首先证明, 对于任何 $U \in \mathbb{R}^{n \times m}$, 式(6.2.22)所确定的 G 是 A 的最小二乘广义逆. 事实上,

$$\begin{aligned} AG &= AA_1^- + A(I_n - A_1^- A)U \\ &= AA_1^- + (A - AA_1^- A)U = AA_1^- + (A - A)U \\ &= AA_1^-. \end{aligned}$$

由引理知, G 为 A 的最小二乘广义逆.

再证明对任意的最小二乘广义逆 G , 必存在 $U \in \mathbb{R}^{n \times m}$, 使 G 具有式(6.2.22)的形式. 事实上, 取 $U = G - A_1^-$ 即可. 因为由引理有 $AG = AA_1^-$, 所以

$$\begin{aligned} &A_1^- + (I_n - A_1^- A)(G - A_1^-) \\ &= A_1^- + G - A_1^- - A_1^- AG + A_1^- AA_1^- \\ &= A_1^- + G - A_1^- - A_1^- AA_1^- + A_1^- AA_1^- \\ &= G. \end{aligned}$$

证毕

从上述定理可以看出, 最小二乘广义逆的通式(6.2.22)与最小二乘解 \hat{x} 的通式(6.2.19)形式上有类似之处, 重要的是先用 6.1 节讲过的方法(1)、(2)、(3)求出 A 的某一个 A_1^- , 然后再用通式(6.2.22)或式(6.2.19)求得其他的(不同的)最小二乘广义逆或最小二乘解.

例 6.2.5 求不相容方程组

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 1, \\ x_1 + x_3 = 0, \\ 2x_1 + 2x_3 = 1, \\ 2x_1 + 4x_2 + 6x_3 = 3 \end{cases}$$

两个不同的最小二乘解, 并比较它们的最小误差.

解 由所给方程组可知

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 2 & 0 & 2 \\ 2 & 4 & 6 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 3 \end{pmatrix}.$$

由于 $\text{rank}A = 2 \neq \text{rank}(A \vdots b) = 3$, 故此方程组为不相容方程组(即矛盾方程组).

例 6.1.7 中的矩阵 A 即为此处的矩阵 A . 由于用满秩分解法所得的 $C_R^{-1}B_L^{-1}$ 既是 A_m^- , 也是 A_1^-, A_r^- 和 A^+ , 它们又都属于 A^- , 所以由前面例 6.1.7 的结果知

$$A_1^- = \frac{1}{30} \begin{pmatrix} -1 & 5 & 10 & -2 \\ 2 & -4 & -8 & 4 \\ 1 & 1 & 2 & 2 \end{pmatrix},$$

于是其中的一个最小二乘解为

$$\hat{x}_1 = A_1^- b = \frac{1}{30} \begin{pmatrix} -1 & 5 & 10 & -2 \\ 2 & -4 & -8 & 4 \\ 1 & 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 3 \end{pmatrix} = \frac{1}{10} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

若取 $U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$, 按通式(6.2.22)又可得到另一个最小二乘广义逆

$$G = A_1^- + (I_3 - A_1^- A)U = \frac{1}{30} \begin{pmatrix} 9 & 5 & 10 & -2 \\ 12 & -4 & -8 & 4 \\ -9 & 1 & 2 & 2 \end{pmatrix}.$$

于是第二个最小二乘解为

$$\hat{x}_2 = Gb = \frac{1}{30} \begin{pmatrix} 9 & 5 & 10 & -2 \\ 12 & -4 & -8 & 4 \\ -9 & 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 3 \end{pmatrix} = \frac{1}{30} \begin{pmatrix} 13 \\ 16 \\ -1 \end{pmatrix}.$$

经计算,这两个最小二乘解的“最小误差平方和”分别是:

$$\|A\hat{x}_1 - b\|_2^2 = \frac{4}{10} = 0.4,$$

$$\|A\hat{x}_2 - b\|_2^2 = \frac{4}{10} = 0.4.$$

可见,尽管最小二乘解不同,但是它们的“最小误差平方和”却是相同的.

注 一般来说,不相容方程组的最小二乘解不是惟一的,但在系数矩阵 A 的列向量线性无关时(即 A 为列满秩矩阵),解是惟一的(证明见文献[3]).此时,必须取 $A_1^- = A_L^{-1} = (A^T A)^{-1} A^T$, 注意到 $A_L^{-1} A = I_n$, 故 $\hat{x} = A_L^{-1} b = (A^T A)^{-1} A^T b$.

例 6.2.6 求矛盾方程组

$$\begin{cases} x_1 + 2x_2 = 1, \\ 2x_1 + x_2 = 0, \\ x_1 + x_2 = 0 \end{cases}$$

的最小二乘解.

解 系数矩阵 A 和向量 b 为

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

A 为列满秩矩阵,在例 6.1.9 中已求得

$$\begin{aligned} A_L^{-1} &= (A^T A)^{-1} A^T \\ &= \frac{1}{11} \begin{pmatrix} -4 & 7 & 1 \\ 7 & -4 & 1 \end{pmatrix} = A_1^-, \end{aligned}$$

于是,最小二乘解为

$$\begin{aligned} \hat{x} &= A_L^{-1} b = A_1^- b \\ &= \frac{1}{11} \begin{pmatrix} -4 & 7 & 1 \\ 7 & -4 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{11} \begin{pmatrix} -4 \\ 7 \end{pmatrix}, \end{aligned}$$

即

$$\hat{x}_1 = -\frac{4}{11}, \quad \hat{x}_2 = \frac{7}{11}.$$

将 \hat{x} 代入误差平方的公式得

$$\|A\hat{x} - b\|_2^2 = \frac{1}{11}.$$

在最小二乘曲线拟合和多元线性回归分析中常常要计算矛盾方程组的最小二乘解. 广义逆矩阵的理论使得求矛盾方程组最小二乘解的方法简单化、标准化了. 整个求解的关键在于求出 A 的最小二乘广义逆 A^- , 而用不着先求误差平方和, 再利用极值条件, 最后求解一个新的方程组等一系列烦琐的步骤.

6.2.5 线性方程组的极小最小二乘解与 A^+

由于加号逆既是减号逆又是极小范数逆、最小二乘逆, 故对于方程组 $Ax = b$, 不论其是否有解, 均可用加号逆 A^+ 来讨论(设 z 是任意 n 维向量):

(1) 当 $Ax = b$ 相容时, $x = A^+b + (I - A^+A)z$ 是通解; $x = A^+b$ 是极小范数解.

(2) 当 $Ax = b$ 不相容时, $x = A^+b$ 是最小二乘解; $x = A^+b + (I - A^+A)z$ 是最小二乘解的通解.

(3) 在下面的定理中, 我们将要证明对于矛盾方程组 $Ax = b$ (即不相容), $x = A^+b$ 不但是最小二乘解, 而且是具有极小范数的最小二乘解, 或最佳逼近解, 简记为 LNLS 解.

定理 6.2.7 矛盾方程组 $Ax = b$ 的极小范数最小二乘解(即最佳逼近解)为

$$x = A^+b. \quad (6.2.23)$$

证明 由式(6.2.19)我们已经证明了矛盾方程组的最小二乘解的一般表达式为

$$Gb + (I - GA)z,$$

其中 z 是任意 n 维向量, G 是最小二乘广义逆, 且满足

$$AGA = A, \quad (AG)^T = AG.$$

进一步设 Gb 是极小范数最小二乘解, 则对任意向量 b 和 z 应成立不等式

$$\|Gb\|_2 \leq \|Gb + (I - GA)z\|_2.$$

模仿定理 6.2.2 的证明过程知, 上式成立的充要条件是

$$(Gb, (I - GA)z) = (b, G^T(I - GA)z) = 0,$$

而保证上式成立的充要条件是

$$G^T(I - GA) = 0 \quad \text{或} \quad G^T = G^TGA.$$

仿照式(6.1.14)与式(6.1.16)等价的证明方法, 可证 $G^T = G^TGA$ 与 $GAG = G, (GA)^T = GA$ 等价. 由此可见, $x = A^+b$ 是极小范数最小二乘解. 证毕

例 6.2.7 求方程组

$$\begin{cases} x_1 + 2x_2 = 1, \\ 2x_3 = 1, \\ 2x_1 + 4x_2 = 3 \end{cases}$$

的极小最小二乘解.

解 由方程组知

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 2 \\ 2 & 4 & 0 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}.$$

因为 $\text{rank}\mathbf{A}=2, \text{rank}(\mathbf{A} \vdots \mathbf{b})=3$, 故此方程组不相容. 根据前面的讨论知, 它的极小最小二乘解为 $\hat{\mathbf{x}}=\mathbf{A}^+\mathbf{b}$, 为此先求 \mathbf{A}^+ . 由例 6.1.6 的解法(即满秩分解法)知

$$\mathbf{A}^+ = \mathbf{C}_R^{-1}\mathbf{B}_L^{-1} = \frac{1}{50} \begin{pmatrix} 2 & 0 & 4 \\ 4 & 0 & 8 \\ 0 & 25 & 0 \end{pmatrix},$$

因此, 方程组的极小最小二乘解为

$$\mathbf{x} = \mathbf{A}^+\mathbf{b} = \frac{1}{50} \begin{pmatrix} 2 & 0 & 4 \\ 4 & 0 & 8 \\ 0 & 25 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} = \frac{1}{50} \begin{pmatrix} 14 \\ 28 \\ 25 \end{pmatrix}.$$

例 6.2.8 求矛盾方程组

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 1, \\ x_1 + x_3 = 0, \\ 2x_1 + 2x_3 = 1, \\ 2x_1 + 4x_2 + 6x_3 = 3 \end{cases}$$

的极小最小二乘解.

解 由例 6.1.10 知

$$\mathbf{A}^+ = \mathbf{A}_1^- = \frac{1}{30} \begin{pmatrix} -1 & 5 & 10 & -2 \\ 2 & -4 & -8 & 4 \\ 1 & 1 & 2 & 2 \end{pmatrix},$$

故方程组的极小最小二乘解为

$$\mathbf{x} = \mathbf{A}^+\mathbf{b} = \frac{1}{30} \begin{pmatrix} -1 & 5 & 10 & -2 \\ 2 & -4 & -8 & 4 \\ 1 & 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 3 \end{pmatrix} = \frac{1}{10} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

注意, 此时 $\|\mathbf{Ax}-\mathbf{b}\|_2 = \frac{2}{\sqrt{10}}, \|\mathbf{x}\|_2 = \frac{\sqrt{14}}{10}$ 均为最小. 说明所求的解是所有最小二乘解中 2 范数最小的.

通过本节的讨论使我们体会到: 如果能方便地求得系数矩阵 \mathbf{A} 的加号逆 \mathbf{A}^+ , 则用它来表示相容或不相容线性方程组的解, 是一种既简单又严谨的计算方法.

习 题 6

1. 求下列矩阵的减号逆 \mathbf{A}^- :

$$(1) \mathbf{A}_1 = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \\ 1 & 0 & 2 \end{pmatrix}; \quad (2) \mathbf{A}_2 = \begin{pmatrix} 2 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}.$$

2. 设 $A \in \mathbb{C}^{m \times n}$, $\text{rank} A = r$, 若有 m 阶可逆矩阵 P 和 n 阶置换矩阵 T , 使得

$$PAT = \begin{bmatrix} I_r & S \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad S \in \mathbb{C}^{(m-r) \times (n-r)},$$

证明: 对任一个 $L \in \mathbb{C}^{(n-r) \times (m-r)}$, 矩阵 $G = T \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & L \end{bmatrix} P$ 是 A 的一个减号逆; 若取 $L = \mathbf{0}$, 则相应的 G 是 A 的一个自反减号逆 A^- .

3. 设 A 是 $m \times n$ 零矩阵, 哪一类矩阵 G 是 A 的减号逆? 哪一类矩阵 G 是 A 的自反减号逆?

4. 设 $m \times n$ 矩阵 A 除第 i_0 行第 j_0 列的元素为 1 外, 其余元素均为 0, 哪一类矩阵 G 是 A 的减号逆? 哪一类矩阵是 A 的自反减号逆?

5. 设 B 是所有元素全为 1 的 n 阶方阵, 记 $A = (a-b)I + bB$, 证明: 若 $a + (n-1)b = 0$, 则 $G = (a-b)^{-1}I$ 是 A 的减号逆.

6. 已知

$$A = \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix},$$

证明: $G = -(a_1^2 + a_2^2 + a_3^2)^{-1}A$ 是 A 的减号逆.

7. 已知矩阵

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

求 A 的一个减号逆和自反减号逆.

8. 对第 1 题中的两个矩阵, 分别求出方程 $A_1 x = b_1$ 和 $A_2 x = b_2$ 的通解, 其中 $b_1 = (1, 0, 1, 1)^T$, $b_2 = (2, 1, 1)^T$.

9. 已知

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

(1) 求 A 的一个减号逆和自反减号逆;

(2) 求 A^+ .

10. 求矩阵

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

的摩尔-彭诺斯逆 A^+ .

11. 求 $A = \begin{bmatrix} 1 & 0 \\ 2 & -1 \\ -1 & 2 \end{bmatrix}$ 的左逆 A_L^{-1} .

12. 设 $A \in \mathbb{R}^{m \times n}$, 且 $A = PBQ$, 其中 P 为 $m \times k$ 列满秩矩阵, Q 为 $s \times n$ 行满秩矩阵, 证明 $\text{rank}(A) = \text{rank}(B)$.

13. 设 $A \in \mathbb{C}^{m \times n}$, $B \in \mathbb{C}^{n \times m}$, $BA = I_n$, $\text{rank}(AB) = n$, $m \geq n$.

(1) 求 AB 的全部特征值;

(2) 证明: $R(AB) = R(A)$, $N(AB) = N(B)$.

14. 设矩阵 $A \in \mathbb{C}^{m \times m}$, $C \in \mathbb{C}^{n \times n}$ 是可逆的, 证明:

(1) 若 $B \in \mathbb{C}^{m \times n}$ 是左可逆的, 则 ABC 是左可逆;

(2) 若 $B \in \mathbb{C}^{m \times n}$ 是右可逆的, 则 ABC 是右可逆.

15. $A \in \mathbb{R}^{m \times n}$ 是一个行满秩矩阵, 证明 A 有右逆为

$$G = VA^T (AVA^T)^{-1},$$

其中 V 是使 $\text{rank}(AVA^T) = \text{rank}(A)$ 成立的任一 n 阶方阵.

16. $A \in \mathbb{C}^{m \times n}$ 是左可逆的, 证明存在 $\epsilon > 0$, 使满足条件 $\|A - B\| < \epsilon$ 的每一个矩阵 $B \in \mathbb{C}^{m \times n}$ 都是左可逆的, 换言之, 单边可逆矩阵对小扰动是稳定的.

17. 设

$$(1) A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 1 & 2 & 2 \end{bmatrix}; (2) A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & 0 \\ 2 & 1 & -1 \end{bmatrix}, \text{求 } A^+.$$

18. 已知矩阵

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

分别求 A^+ , B^+ .

19. 设 $A \in \mathbb{C}^{m \times n}$, P 与 Q 分别为 m 阶与 n 阶酉矩阵, 试证

$$(PAQ)^+ = Q^+ A^+ P^+.$$

20. 证明下列等式

$$(1) (A^H A)^+ = A^+ (A^H)^+; (AA^H)^+ = (A^H)^+ A^+;$$

$$(2) (A^H A)^+ = A^+ (AA^H)^+ A = A^H (AA^H)^+ (A^H)^+;$$

$$(3) AA^+ = (AA^H)(AA^H)^+ = (AA^H)^+(AA^H);$$

$$(4) A^+ A = (A^H A)(A^H A)^+ = (A^H A)^+(A^H A);$$

(5) 如果 $A^H = A$, 那么

$$(A^2)^+ = (A^+)^2, \quad A^2 (A^2)^+ = (A^2)^+ A^2 = AA^+;$$

(6) 如果 $A^H = A$, 那么

$$AA^+ = A^+ A.$$

21. 设 A 是一个正规矩阵, 证明 $AA^+ = A^+ A$.

22. 设 $A \in \mathbb{R}^{m \times n}$, 且 A 的 n 个列是标准正交的, 证明

$$A^+ = A^T.$$

23. A 是幂等且是埃尔米特阵(即 A 是正交投影矩阵), 证明 $A^+ = A$.

24. $A \in \mathbb{C}^{m \times n}$, $\text{rank}(A) = 1$, 证明:

(1) 存在数 a_1, a_2, \dots, a_m 与 b_1, b_2, \dots, b_n , 使

$$A = (a_1, a_2, \dots, a_m)^T (b_1, b_2, \dots, b_n);$$

(2) $A^+ = \frac{1}{a} A^H$, 其中 $a = \sum_{i=1}^m \sum_{j=1}^n |a_i b_j|^2$.

25. $A \in \mathbb{R}^{m \times n}$, $\text{rank}(A) = r$, U, V 是正交矩阵, 若 $A = U \begin{bmatrix} A_1 & 0 \\ 0 & 0 \end{bmatrix} V$, 其中 A_1 为可逆矩阵, 则

$$A^+ = V^H \begin{bmatrix} A_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^H.$$

26. 证明: $A^+ AB = A^+ AC$ 的充分必要条件是 $AB = AC$.

27. $A \in \mathbb{R}^{m \times n}$, 且 $\text{rank}(A) = n$, A 的正交上三角分解为: $A = QR$, 其中 $Q_{m \times n}$ 的 n 个列标准正交, R 是正

对角线元的上三角矩阵,证明:

$$A^+ = R^{-1}Q^T.$$

28. 设 $A \in \mathbb{R}^{m \times n}$, $A^T A$ 的特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 对应的 n 个标准正交的特征向量 X_1, X_2, \dots, X_n 组成正交矩阵 $Q = (X_1, X_2, \dots, X_n)$, 则

$$A^+ = QA^+Q^T A^T,$$

其中 $A = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$.

29. 举例说明有关摩尔-彭诺斯逆的下列命题不真:

- (1) $(AB)^+ = B^+ A^+$;
- (2) $(A^k)^+ = (A^+)^k$, k 为正整数;
- (3) 若 λ 是 A 的特征值, 则 λ 是 A^+ 的特征值;
- (4) 若 P, Q 为可逆矩阵, 则 $(PAQ)^+ = Q^{-1}A^+P^{-1}$.

30. 求下列矩阵的极小范数广义逆 A_m^- :

$$(1) A_1 = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 4 \end{bmatrix}; (2) A_2 = \begin{bmatrix} 1 & 0 & 3 \\ 2 & 3 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

并分别求方程 $A_1 x = b_1$ 和 $A_2 x = b_2$ 的最小范数解, 其中 $b_1 = (1, -1)^T$, $b_2 = (3, 0, 1)^T$.

31. 已知

$$A = \begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{pmatrix}, \quad b = \begin{pmatrix} 4 \\ -2 \\ -2 \end{pmatrix},$$

- (1) 用广义逆矩阵方法判定线性方程组 $Ax = b$ 是否相容?
- (2) 指出 $Ax = b$ 的极小范数解或极小范数最小二乘解(指出解的类型).

32. 设矩阵 $G \in A\{1, 4\}$ (即 G 为最小二乘广义逆), 则 $x \in \mathbb{C}^n$ 是不相容线性方程组 $Ax = b$ 的最小二乘解的充分必要条件是: 对任何 $b \in \mathbb{C}^m$, x 是方程组 $Ax = AGb$ 的解. 试证明之.

33. 求下列矩阵的最小二乘广义逆 A_1^- :

$$(1) A_1 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 1 \end{pmatrix}; (2) A_2 = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 2 & 0 & 2 & 2 \\ 4 & 2 & 4 & 2 \end{pmatrix}.$$

并分别求不相容方程组 $A_1 x = b_1$ 和 $A_2 x = b_2$ 的最小二乘解, 其中 $b_1 = (1, 0, 0)^T$, $b_2 = (0, 1, 0, 1)^T$.

34. 用满秩分解法求下列矩阵的极小最小二乘广义逆 A^+ :

$$(1) A_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 2 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}; (2) A_2 = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 2 & 2 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

并分别求不相容方程组 $A_1 x = b_1$ 和 $A_2 x = b_2$ 的极小最小二乘解, 其中 $b_1 = (1, 2, 1, 2)^T$, $b_2 = (0, 1, 0, 0)^T$.

35. 证明: 线性方程组 $Ax = b$ 有解的充分必要条件是 $AA^+b = b$ 和 $\text{rank}A = \text{rank}(A \mid b)$.

36. 设 $A \in \mathbb{C}^{m \times n}$, P, Q 分别是 m 阶和 n 阶可逆矩阵.

- (1) 证明 $Q^{-1}AP^{-1} \in (PAQ)^-$;
- (2) 举例说明 $(PAQ)^+ = Q^{-1}A^+P^{-1}$ 不真.

37. 证明: 设 $A \in \mathbb{C}^{m \times n}$, $B \in \mathbb{C}^{p \times q}$, $C \in \mathbb{C}^{m \times q}$, 则矩阵方程 $AXB = C$ 相容的充分必要条件是: 对某个 A^- , B^- 有 $AA^-CB^-B = C$ 成立, 且方程的通解为

$$X = A^-CB^- + (Z - A^-AZBB^-),$$

其中 $\mathbf{Z} \in \mathbb{C}^{m \times p}$ 任意.

38. 设非齐次线性方程组 $\mathbf{A}\mathbf{X}=\mathbf{b}$ 有解, 证明: 此方程组的一般解为 $\mathbf{X}=\mathbf{A}^-\mathbf{b}$, 其中 \mathbf{A}^- 是 \mathbf{A} 的任意一个广义逆.

39. 已知线性方程组

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 1, \\ x_1 + x_3 = 0, \\ 2x_1 + 2x_3 = 1, \\ 2x_1 + 4x_2 + 6x_3 = 3. \end{cases}$$

(1) 证明此线性方程组无解 (或为不相容线性方程组);

(2) 求此方程组的极小最小二乘解 \mathbf{X} ;

(3) 求 $\|\mathbf{X}\|_2$, 并求 $\mathbf{b}=(1,0,1,3)^\top$ 到 $R(\mathbf{A})$ 的最短距离, 这里 \mathbf{A} 为此方程组的系数矩阵.

40. 求线性方程组

$$\begin{cases} x_1 + 2x_2 = 1, \\ 2x_3 = 1, \\ 2x_1 + 4x_2 = 3 \end{cases}$$

的极小最小二乘解.

41. 设矩阵 \mathbf{A} 及向量 $\boldsymbol{\alpha}$ 为

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 2 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \boldsymbol{\alpha} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

(1) 证明 $\overline{\boldsymbol{\alpha}} \in R(\mathbf{A})$;

(2) 在 $R(\mathbf{A})$ 中求一向量 \mathbf{Y}_0 , 使 \mathbf{Y}_0 与向量 $\boldsymbol{\alpha}$ 距离很近;

(3) 求 $\|\mathbf{Y}_0 - \boldsymbol{\alpha}\|_2$.

42. 设平面曲线的 4 个点坐标为 $(-1,3)^\top, (0,0)^\top, (1,2)^\top, (2,5)^\top$, 求一与这些点吻合的二次曲线.

43. 设 $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\text{rank}(\mathbf{A})=n$, $\mathbf{A}\mathbf{x}=\mathbf{b}$ 为不相容的线性方程组.

(1) 求 $\mathbf{A}\mathbf{x}=\mathbf{b}$ 的极小最小二乘解 \mathbf{x}_0 ;

(2) 利用矩阵的正交三角分解 $\mathbf{A}=\mathbf{Q}_{m \times n}\mathbf{R}_{n \times n}$, 证明极小最小二乘解为 $\mathbf{x}_0=\mathbf{R}^{-1}\mathbf{Q}^\top\mathbf{b}$.

44. 用奇异值分解证明:

(1) $R(\mathbf{A}\mathbf{A}^+)=R(\mathbf{A}\mathbf{A}^\text{H})=R(\mathbf{A})$;

(2) $R(\mathbf{A}^+\mathbf{A})=R(\mathbf{A}^\text{H}\mathbf{A})=R(\mathbf{A}^+)=R(\mathbf{A}^\text{H})$;

(3) $N(\mathbf{A}\mathbf{A}^+)=N(\mathbf{A}\mathbf{A}^\text{H})=N(\mathbf{A}^+)=N(\mathbf{A}^\text{H})$;

(4) $N(\mathbf{A}^+\mathbf{A})=N(\mathbf{A}^\text{H}\mathbf{A})=N(\mathbf{A})$.

几类特殊矩阵与特殊积

引言 什么是特殊矩阵与特殊积

本章综合介绍几类特殊的矩阵,以及几种矩阵的特殊积.它们在许多科学技术领域内,都有着不同程度的应用.这里所谓的特殊矩阵,是指它的形状或其所具有的性质有某些特性的矩阵.例如,大家熟知的对角矩阵、单位矩阵、三角矩阵、对称矩阵、埃尔米特矩阵、正交矩阵、酉矩阵、正规矩阵、正定矩阵、非负定矩阵、吉文斯(Givens)旋转矩阵以及豪斯霍尔德反射矩阵等,这些特殊矩阵前面已作过介绍,不在本章的研究范围内.在这里,主要讨论非负矩阵与正矩阵、素矩阵与循环矩阵、随机矩阵、单调矩阵、M矩阵与H矩阵,以及其他的诸如T矩阵(特普利茨(Toeplitz)矩阵)与汉克尔(Hankel)矩阵等一些重要的特殊矩阵.再说矩阵的特殊积,也不是前面谈的一般乘积 AB ,因为一般乘积要求 A 的列数等于 B 的行数,而这里引入的几种矩阵的特殊积,它们不受矩阵的行数和列数的限制,显然它们是前述矩阵乘积 AB 的推广.用这样的特殊积来表示矩阵方程的解,有时显得十分地简洁.

7.1 非负矩阵

在数理经济学、概率论、弹性系统微振动理论等许多领域里,常常出现“元素都是非负实数”的矩阵,这类矩阵在数学上把它归成一类,叫做“非负矩阵”,它的基本特征已被认为是矩阵论的经典性内容之一.为此,本节介绍非负矩阵的一些基本性质,包括著名的佩龙-弗罗贝尼乌斯(Perron Frobenius)定理,以及正矩阵、不可约非负矩阵、素矩阵等概念.

7.1.1 非负矩阵与正矩阵

定义 7.1.1 设 $A = (a_{ij}) \in \mathbb{R}^{m \times n}$, 如果

$$a_{ij} \geq 0, \quad i = 1, \dots, m; j = 1, \dots, n, \quad (7.1.1)$$

即 A 的所有元素是非负的,则称 A 为非负矩阵,记作 $A \geq 0$; 若式(7.1.1)中严格不等号成立,即 $a_{ij} > 0$ ($i = 1, \dots, m; j = 1, \dots, n$),则称 A 为正矩阵,记为 $A > 0$.

设 $A, B \in \mathbb{R}^{m \times n}$, 如果成立 $A - B \geq 0$, 则记作 $A \geq B$; 如果成立 $A - B > 0$, 则记作 $A > B$.

必须注意,非负矩阵和正矩阵的概念与第1章1.3节中非负定矩阵和正定矩阵的概念是不同的.

对于任意的 $\mathbf{A}=(a_{ij})\in\mathbb{C}^{m\times n}$, 引进记号

$$|\mathbf{A}|=(|a_{ij}|), \quad (7.1.2)$$

即表示以 a_{ij} 之模 $|a_{ij}|$ 为元素所得的非负矩阵; 特别地, 当 $\mathbf{x}=(x_1, \dots, x_n)^T \in \mathbb{C}^n$ 时, $|\mathbf{x}|=(|x_1|, \dots, |x_n|)^T$ 表示一个非负向量.

注意, 这里使用的记号 $|\mathbf{A}|$ 与 $|\mathbf{x}|$, 不要与前面讲的“方阵的行列式”和“向量的长度”概念混淆了.

由定义 7.1.1 可直接得到如下定理.

定理 7.1.1 设 $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D} \in \mathbb{C}^{m \times n}$, 则

- (1) $|\mathbf{A}| \geq 0$, 并且 $|\mathbf{A}| = 0$ 当且仅当 $\mathbf{A} = \mathbf{0}$;
- (2) 对任意复数 α , 有 $|\alpha\mathbf{A}| = |\alpha||\mathbf{A}|$;
- (3) $|\mathbf{A} + \mathbf{B}| \leq |\mathbf{A}| + |\mathbf{B}|$;
- (4) 若 $\mathbf{A} \geq 0, \mathbf{B} \geq 0, a, b$ 是非负实数, 则 $a\mathbf{A} + b\mathbf{B} \geq 0$;
- (5) 若 $\mathbf{A} \geq \mathbf{B}$, 且 $\mathbf{C} \geq \mathbf{D}$, 则 $\mathbf{A} + \mathbf{C} \geq \mathbf{B} + \mathbf{D}$;
- (6) 若 $\mathbf{A} \geq \mathbf{B}$, 且 $\mathbf{B} \geq \mathbf{C}$, 则 $\mathbf{A} \geq \mathbf{C}$.

一般由 $\mathbf{A} \geq 0$ 和 $\mathbf{A} \neq 0$, 不能导出 $\mathbf{A} > 0$.

定理 7.1.2 设 $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D} \in \mathbb{C}^{n \times n}, \mathbf{x} \in \mathbb{C}^n$, 则

- (1) $|\mathbf{Ax}| \leq |\mathbf{A}||\mathbf{x}|$;
- (2) $|\mathbf{AB}| \leq |\mathbf{A}||\mathbf{B}|$;
- (3) 对任意正整数 m , 有 $|\mathbf{A}^m| \leq |\mathbf{A}|^m$;
- (4) 若 $0 \leq \mathbf{A} \leq \mathbf{B}, 0 \leq \mathbf{C} \leq \mathbf{D}$, 则 $0 \leq \mathbf{AC} \leq \mathbf{BD}$;
- (5) 若 $0 \leq \mathbf{A} \leq \mathbf{B}$, 对任意正整数 m , 有 $0 \leq \mathbf{A}^m \leq \mathbf{B}^m$;
- (6) 若 $\mathbf{A} \geq 0$ ($\mathbf{A} > 0$), 对任意正整数 m , $\mathbf{A}^m \geq 0$ ($\mathbf{A}^m > 0$);
- (7) 若 $\mathbf{A} > 0, \mathbf{x} \geq 0$ 且 $\mathbf{x} \neq 0$, 则 $\mathbf{Ax} > 0$;
- (8) 若 $|\mathbf{A}| \leq \mathbf{B}$, 则 $\|\mathbf{A}\|_2 \leq \||\mathbf{A}|\|_2 \leq \|\mathbf{B}\|_2$.

证明 (1)~(7)显然成立, 下面证明(8).

因为 $\forall \mathbf{x} \in \mathbb{C}^n$, 都有

$$|\mathbf{Ax}| \leq |\mathbf{A}||\mathbf{x}| \leq \mathbf{B}|\mathbf{x}|,$$

则

$$\|\mathbf{Ax}\|_2 = \||\mathbf{Ax}|\|_2 \leq \||\mathbf{A}||\mathbf{x}|\|_2 \leq \|\mathbf{B}|\mathbf{x}|\|_2,$$

于是

$$\max_{\|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2 = \max_{\|\mathbf{x}\|_2=1} \||\mathbf{Ax}|\|_2 \leq \max_{\|\mathbf{x}\|_2=1} \||\mathbf{A}||\mathbf{x}|\|_2 \leq \max_{\|\mathbf{x}\|_2=1} \|\mathbf{B}|\mathbf{x}|\|_2.$$

由上式有

$$\|\mathbf{A}\|_2 \leq \||\mathbf{A}|\|_2 \leq \|\mathbf{B}\|_2.$$

证毕

定理 7.1.3(谱半径的单调性) 设 $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n \times n}$, 若 $|\mathbf{A}| \leq \mathbf{B}$, 则

$$\rho(\mathbf{A}) \leq \rho(|\mathbf{A}|) \leq \rho(\mathbf{B}). \quad (7.1.3)$$

证明 由定理 7.1.2 的(3)和(5)知, 对任意正整数 m , 有 $|\mathbf{A}^m| \leq |\mathbf{A}|^m \leq \mathbf{B}^m$, 由定理 7.1.2(8), 有

$$\|\mathbf{A}^m\|_2 \leq \||\mathbf{A}^m|\|_2 \leq \|\mathbf{B}^m\|_2,$$

从而

$$\|A^m\|_2^{\frac{1}{m}} \leq \| |A| \|^{\frac{1}{m}} \leq \|B^m\|_2^{\frac{1}{m}}.$$

由于 $(\rho(A))^m = \rho(A^m) \leq \|A^m\|_2$, 所以对 $m=1, 2, \dots$, 有 $\rho(A) \leq \|A^m\|_2^{\frac{1}{m}}$. 另一方面, 对任意 $\epsilon > 0$, 矩阵 $\tilde{A} = [\rho(A) + \epsilon]^{-1}A$ 的谱半径严格小于1, 由定理5.1.2知 $\lim_{m \rightarrow \infty} \tilde{A}^m = \mathbf{0}$, 于是当 $m \rightarrow \infty$ 时, $\|\tilde{A}^m\|_2 \rightarrow 0$. 因此, 存在正整数 k , 使得当 $m > k$ 时, $\|\tilde{A}^m\|_2 < 1$, 即对所有 $m > k$ 有 $\|A^m\|_2 \leq [\rho(A) + \epsilon]^m$ 或 $\|A^m\|_2^{\frac{1}{m}} \leq \rho(A) + \epsilon$, 故 $\lim_{m \rightarrow \infty} \|A^m\|_2^{\frac{1}{m}} = \rho(A)$.

同理有 $\lim_{m \rightarrow \infty} \| |A| \|^{\frac{1}{m}} = \rho(|A|)$ 和 $\lim_{m \rightarrow \infty} \|B^m\|_2^{\frac{1}{m}} = \rho(B)$, 即得

$$\rho(A) \leq \rho(|A|) \leq \rho(B). \quad \text{证毕}$$

由定理7.1.3立即得到如下推论.

推论1 设 $A, B \in \mathbb{R}^{n \times n}$, 若 $0 \leq A \leq B$, 则 $\rho(A) \leq \rho(B)$.

推论2 设 $A \in \mathbb{R}^{n \times n}$, 若 $A \geq 0$, $A^{(k)}$ 是 A 的任一主子矩阵, 则 $\rho(A^{(k)}) \leq \rho(A)$. 特别地, $\max_{1 \leq i \leq n} \{a_{ii}\} \leq \rho(A)$.

事实上, 对任意正整数 k ($1 \leq k \leq n$), 用 \tilde{A} 表示把 $A^{(k)}$ 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 $\rho(A^{(k)}) = \rho(\tilde{A})$ 并且 $0 \leq \tilde{A} \leq A$, 则由推论1知, $\rho(A^{(k)}) = \rho(\tilde{A}) \leq \rho(A)$.

佩龙(Perron)在1907年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7.1.4(佩龙定理) 设 $A \in \mathbb{R}^{n \times n}$, 且 $\rho(A)$ 为其谱半径, 若 $A > 0$ (正矩阵), 则

- (1) $\rho(A)$ 为 A 的正特征值, 其对应的一个特征向量 $y \in \mathbb{R}^n$ 必为正向量;
- (2) 对 A 的任何其他特征值 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, $x = (x_1, \dots, x_n)^T$ 是相应的特征向量, 则

$$Ax = \mu x, \quad \text{且} \quad |\mu| = \rho(A). \quad (7.1.4)$$

令 $y = (|x_1|, \dots, |x_n|)^T$, 下面证明 y 是 A 对应于特征值 $\rho(A)$ 的正特征向量.

因为 $Ax = \mu x$, 所以对于 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, 有

$$\rho(A) y \leq Ay,$$

即

$$(A - \rho(A)I)y \geq 0. \quad (7.1.5)$$

下面证明式(7.1.5)的等号成立. 用反证法, 设 $(A - \rho(A)I)y = z \neq \mathbf{0}$, 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

$$Az \geq \epsilon Ay. \quad (7.1.6)$$

因为 $Az = A(A - \rho(A)I)y$, 所以

$$A^2 y = Az + \rho(A)Ay \geq [\varepsilon + \rho(A)]Ay.$$

令 $[\varepsilon + \rho(A)]^{-1}A = B$, 则有 $B > 0$, $\rho(B) < 1$, 且

$$BAy \geq Ay.$$

由上式可逐步推得

$$B^k Ay \geq Ay, \quad k = 1, 2, \dots \quad (7.1.7)$$

又因为 $\rho(B) < 1$, 则由定理 5.1.2 知, 当 $k \rightarrow \infty$ 时, $B^k \rightarrow 0$. 对不等式(7.1.7)两端取极限, 即得 $Ay \leq 0$, 这与 $Ay > 0$ 矛盾, 因此有 $z = 0$. 于是证明了

$$Ay = \rho(A)y. \quad (7.1.8)$$

这表明 $\rho(A) = |\mu|$ 是 A 的特征值, 而 y 是 A 的正特征向量.

下面证明(2). 只要证明除 $\rho(A)$ 外, A 不可能还有其他特征值 λ 能满足 $|\lambda| = \rho(A)$.

不妨假设 λ 是 A 的特征值, 且满足 $|\lambda| = \rho(A)$, 相应的特征向量为 $u = (u_1, \dots, u_n)^T$, 则

$$Au = \lambda u. \quad (7.1.9)$$

令 $v = (|u_1|, \dots, |u_n|)^T$, 重复上面证明(1)的讨论可得

$$Av = \rho(A)v. \quad (7.1.10)$$

而由式(7.1.9)可得

$$\lambda u_i = \sum_{j=1}^n a_{ij} u_j, \quad j = 1, \dots, n,$$

从而

$$\rho(A) |u_i| = \left| \sum_{j=1}^n a_{ij} u_j \right|. \quad (7.1.11)$$

将式(7.1.10)代入式(7.1.11)得

$$\left| \sum_{j=1}^n a_{ij} u_j \right| = \sum_{j=1}^n a_{ij} |u_j|, \quad i = 1, \dots, n.$$

由于 $a_{ij} > 0$, 则上式表明所有的 u_j 有相同的辐角 φ , 即

$$u_j = |u_j| e^{i\varphi}, \quad \text{其中 } i = \sqrt{-1}, j = 1, \dots, n,$$

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 $e^{i\varphi}$, 故 u 也是 A 对应于特征值 $\rho(A)$ 的特征向量, 即

$$Au = \rho(A)u. \quad (7.1.12)$$

由式(7.1.9)和式(7.1.12)可得 $\lambda = \rho(A)$.

最后证明(3). 令 $B = \rho^{-1}(A)A = (b_{ij})$, 则 $B > 0$ 且 $\rho(B) = 1$. 欲证明(3), 只需证明 1 是 B 的单特征值即可, 或者说, 在 B 的若尔当标准形中对应于特征值 1 只有一个一阶若尔当块.

根据结论(1)知, 存在向量 $y = (y_1, \dots, y_n)^T > 0$, 使得

$$By = y, \quad (7.1.13)$$

从而对任何正整数 k 都有

$$B^k y = y. \quad (7.1.14)$$

令 $y_s = \max_i y_i > 0$, $y_t = \min_i y_i > 0$, 则由式(7.1.14)可得

$$y_s \geq y_i = \sum_{l=1}^n b_{il}^{(k)} y_l \geq b_{ij}^{(k)} y_j \geq b_{ij}^{(k)} y_t,$$

其中 $b_{ij}^{(k)}$ 表示 \mathbf{B}^k 的 (i, j) 位置上元素, 从而有 $b_{ij}^{(k)} \leq \frac{y_j}{y_i}$, 这表明对所有 $k > 1$, $b_{ij}^{(k)}$ 是有界的.

假若 \mathbf{B} 的若尔当标准形中有一个对应于特征值 1 的若尔当块的阶数大于 1, 不妨设其为 2, 则存在可逆矩阵 \mathbf{P} 使得

$$\mathbf{B} = \mathbf{P} \begin{pmatrix} 1 & 1 & & & \\ 0 & 1 & & & \\ & & \mathbf{J}_1(\lambda_1) & & \\ & & & \ddots & \\ & & & & \mathbf{J}_m(\lambda_m) \end{pmatrix} \mathbf{P}^{-1},$$

其中

$$\mathbf{J}_i(\lambda_i) = \begin{pmatrix} \lambda_i & 1 & & & \\ & \lambda_i & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{pmatrix},$$

并且 $|\lambda_i| < 1$ ($i=1, \dots, m$), 则对 $k \geq 1$ 有

$$\mathbf{B}^k = \mathbf{P} \begin{pmatrix} 1 & k & & & \\ & 1 & & & \\ & & \mathbf{J}_1^k(\lambda_1) & & \\ & & & \ddots & \\ & & & & \mathbf{J}_m^k(\lambda_m) \end{pmatrix} \mathbf{P}^{-1},$$

这与 $b_{ij}^{(k)}$ 有界相矛盾, 故 \mathbf{B} 的若尔当标准形中对应于特征值 1 的若尔当块是一阶的.

下面证明 \mathbf{B} 的若尔当标准形中对应于特征值 1 的一阶若尔当块只有一个. 设 \mathbf{B} 的若尔当标准形为

$$\mathbf{J} = \begin{pmatrix} \mathbf{I}_r & & & \\ & \mathbf{J}_1(\lambda_1) & & \\ & & \ddots & \\ & & & \mathbf{J}_l(\lambda_l) \end{pmatrix},$$

其中 \mathbf{I}_r 为 r 阶单位矩阵, 且 $|\lambda_i| < 1$ ($i=1, \dots, l$).

如果 $r > 1$, 令 $\mathbf{C} = \mathbf{J} - \mathbf{I}$, 则由例 1.1.15 知, $\dim(N(\mathbf{C})) = n - \text{rank}(\mathbf{C}) = r$, 由于 \mathbf{B} 与 \mathbf{J} 相似, 故 $\dim(N(\mathbf{B} - \mathbf{I})) = r$. 因为 $r > 1$, 所以除有向量 \mathbf{y} 满足式(7.1.13)外, 必然还有另一向量 $\mathbf{z} = (z_1, \dots, z_n)^T \in \mathbb{R}^n$ 满足

$$\mathbf{B}\mathbf{z} = \mathbf{z}, \quad (7.1.15)$$

并且 \mathbf{z} 与 \mathbf{y} 线性无关, 令

$$\tau = \max_i \left(\frac{z_i}{y_i} \right) = \frac{z_j}{y_j}, \quad (7.1.16)$$

则有 $\tau\mathbf{y} \geq \mathbf{z}$, 且不可能取等号, 于是

$$\mathbf{B}(\tau\mathbf{y} - \mathbf{z}) > 0.$$

利用式(7.1.13)和式(7.1.15), 上式可写成

$$\tau\mathbf{y} - \mathbf{z} > 0.$$

写出上式的第 j 个分量,则有

$$\tau > \frac{z_j}{y_j},$$

这与式(7.1.16)中 τ 的定义相矛盾,故 $r=1$.

证毕

推论 正矩阵 A 的“模等于 $\rho(A)$ ”的特征值是惟一的.

但是,以上结论对一般的非负矩阵未必成立.例如,4阶非负矩阵

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

容易验证 $\rho(A)=3$ 是 A 的特征值,与它对应的特征向量为 $x=(\alpha, \alpha, \beta, 0)^T$,其中 α, β 可取正数.但 $\rho(A)=3$ 并不是 A 的单特征值(实际为二重),而且没有对应于特征值 $\rho(A)=3$ 的正特征向量.同时,还可看出 A 还有异于 $\rho(A)$ 的特征值 $\lambda=-3$,使得 $|\lambda|=\rho(A)$,即 A 的“模等于 $\rho(A)$ ”的特征值并不惟一.

佩龙定理有许多重要应用,一个漂亮而有效的应用是,利用占优非负矩阵的谱半径和主对角元可以得到矩阵的特征值包含区域.

定理 7.1.5 设 $A=(a_{ij})_{n \times n}$, $B=(b_{ij})_{n \times n} \in \mathbb{R}^{n \times n}$ 为非负矩阵, $|a_{ij}| \leq b_{ij}$, $i, j=1, 2, \dots, n$, 则

$$\lambda(A) \subset \bigcup_{i=1}^n \{z \in \mathbb{C} \mid |z - a_{ii}| \leq \rho(B) - b_{ii}\} \quad (7.1.17)$$

证明 可以假定 $B > 0$,事实上,若 B 中有元素为零,考虑 $B_\epsilon = (b_{ij} + \epsilon)_{n \times n}$,其中 $\epsilon > 0$,则 $B_\epsilon > 0$,且 $\lim_{\epsilon \rightarrow 0} (\rho(B_\epsilon) - (b_{ii} + \epsilon)) = \rho(B) - b_{ii}$.由佩龙定理,存在正向量 $x = (x_1, x_2, \dots, x_n)^T$,

使得 $Bx = \rho(B)x$,因而 $\sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| x_j \leq \sum_{\substack{j=1 \\ j \neq i}}^n b_{ij} x_j = \rho(B)x_i - b_{ii}x_i, i = 1, 2, \dots, n$.于是

$\frac{1}{x_i} \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| x_j \leq \rho(B) - b_{ii}$.由4.3节圆盘定理的推论,便知式(7.1.17)成立. 证毕

对于正矩阵还有如下的性质,这个结果在数理经济学中有直接的应用.

定理 7.1.6 设 $A \in \mathbb{R}^{n \times n}$,如果 $A > 0$, x 是 A 的对应于特征值 $\rho(A)$ 的正特征向量,又 y 是 A^T 的对应于特征值 $\rho(A)$ 的任一正特征向量,则

$$\lim_{m \rightarrow \infty} [\rho(A)^{-1} A]^m = (y^T x)^{-1} x y^T. \quad (7.1.18)$$

证明 记 $B = \rho(A)^{-1} A$,则 $B > 0$.由定理7.1.4及其证明过程知, $\rho(B)=1$ 是 B 的单特征值,并且在 B 的若尔当标准形中对应于特征值 1 只有一个一阶若尔当块,因此 B 的标准形为

$$J = \begin{pmatrix} 1 & & & \\ & J_1(\lambda_1) & & \\ & & \ddots & \\ & & & J_l(\lambda_l) \end{pmatrix},$$

其中 λ_i 是 B 的特征值,且 $|\lambda_i| < 1$ ($i=1, \dots, l$).于是 $\lim_{m \rightarrow \infty} B^m$ 存在,记 $\lim_{m \rightarrow \infty} B^m = P$.由于

$$P = \lim_{m \rightarrow \infty} B^m = B \lim_{m \rightarrow \infty} B^{m-1} = BP,$$

例 7.1.1 设有矩阵

$$\mathbf{A} = \begin{pmatrix} b_1 & c_1 & & & \\ a_2 & b_2 & c_2 & & \\ & \ddots & \ddots & \ddots & \\ & & a_{n-1} & b_{n-1} & c_{n-1} \\ & & & a_n & b_n \end{pmatrix},$$

其中 a_i, b_i, c_i 都不为零, 即三对角矩阵,

$$\mathbf{B} = \begin{pmatrix} 4 & -1 & -1 & 0 \\ -1 & 4 & 0 & -1 \\ -1 & 0 & 4 & -1 \\ 0 & -1 & -1 & 4 \end{pmatrix}.$$

由于它们无论怎样行列重排, 都不能形成左下角为零矩阵而对角线是两个低阶的方阵 \mathbf{A}_{11} 和 \mathbf{A}_{22} , 所以 \mathbf{A} 和 \mathbf{B} 是不可约矩阵.

可约的概念来源于线性方程组的求解问题. 一个线性方程组的系数矩阵是可约的, 表明该方程组可通过适当调整方程和未知数的次序, 化为两个低阶的方程组来求解. 即如果线性方程组

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

的系数矩阵 \mathbf{A} 可约时, 则可找到置换矩阵 \mathbf{P} 使 \mathbf{A} 呈

$$\mathbf{PAP}^T = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{pmatrix}.$$

于是原方程组可化为

$$\mathbf{PAP}^T(\mathbf{P}\mathbf{x}) = \mathbf{P}\mathbf{b}.$$

依次记 $\mathbf{y} = \mathbf{P}\mathbf{x} = (\mathbf{y}_1^T, \mathbf{y}_2^T)^T$ 和 $\hat{\mathbf{b}} = \mathbf{P}\mathbf{b} = (\hat{\mathbf{b}}_1^T, \hat{\mathbf{b}}_2^T)^T$, 就有

$$\begin{cases} \mathbf{A}_{11}\mathbf{y}_1 + \mathbf{A}_{12}\mathbf{y}_2 = \hat{\mathbf{b}}_1, \\ \mathbf{A}_{22}\mathbf{y}_2 = \hat{\mathbf{b}}_2. \end{cases}$$

于是方程组化为两个独立的低阶方程组, 比直接解原方程组要方便、简单.

同样, \mathbf{A} 的特征多项式也化为两个低阶矩阵的特征多项式的乘积.

从定义 7.1.2 直接可得如下定理.

定理 7.1.7 设 $\mathbf{A} \in \mathbb{R}^{n \times n}$, 则

- (1) \mathbf{A} 为不可约矩阵的充分必要条件是 \mathbf{A}^T 为不可约矩阵;
- (2) 如果 \mathbf{A} 是不可约非负矩阵, \mathbf{B} 是 n 阶非负矩阵, 则 $\mathbf{A} + \mathbf{B}$ 是不可约非负矩阵.

对于一个给定的矩阵, 直接根据定义 7.1.2 判断是否可约, 绝非易事, 因为 n 阶矩阵共有 $n!$ 个置换矩阵, 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 7.1.8 $n(\geq 2)$ 阶非负矩阵 \mathbf{A} 不可约的充分必要条件是存在正整数 $s \leq n-1$, 使得

$$(\mathbf{I} + \mathbf{A})^s > \mathbf{0}.$$

证明 必要性. 这只需证明对任意向量 $\mathbf{y} \geq \mathbf{0} (\mathbf{y} \neq \mathbf{0})$ 都有不等式

$$(\mathbf{I} + \mathbf{A})^{n-1} \mathbf{y} > \mathbf{0}.$$

我们首先证明在条件 $\mathbf{y} \geq \mathbf{0}$ 与 $\mathbf{y} \neq \mathbf{0}$ 下, 向量 $\mathbf{z} = (\mathbf{I} + \mathbf{A})\mathbf{y}$ 中零坐标的个数小于向量 \mathbf{y} 中零坐

标的个数. 假若相反, 那么 y 与 z 有相同的零坐标个数 (因为 z 的零坐标个数不会多于 y 的零坐标个数). 所以, 不失一般性, 设

$$y = \begin{bmatrix} u \\ \mathbf{0} \end{bmatrix}, \quad z = \begin{bmatrix} v \\ \mathbf{0} \end{bmatrix}, \quad u, v > 0,$$

这里列向量 u, v 有相同维数. 又令

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

则有

$$\begin{bmatrix} u \\ \mathbf{0} \end{bmatrix} + \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{bmatrix} u \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} v \\ \mathbf{0} \end{bmatrix},$$

因此得 $A_{21}u = 0$. 又因 $u > 0$, 故有 $A_{21} = \mathbf{0}$, 这与 A 为不可约矩阵相矛盾, 所以 z 与 y 有相同的零坐标个数是不可能的. 从而证明了向量 z 的零坐标个数小于向量 y 的零坐标个数.

上述结果表明: 向量 $y (0 \leq y \neq \mathbf{0})$ 每用 $I + A$ 左乘一次, 其零坐标个数至少减少一个. 因此得

$$(I + A)^{n-1} y > 0.$$

充分性. 设有 $(I + A)^s > 0$, 如果 A 是可约的, 则存在置换矩阵 P , 使得

$$P(I + A)P^T = \begin{pmatrix} A_{11} + I^{(1)} & A_{12} \\ \mathbf{0} & A_{22} + I^{(2)} \end{pmatrix} = \begin{bmatrix} \tilde{A}_{11} & A_{12} \\ \mathbf{0} & \tilde{A}_{22} \end{bmatrix},$$

对任意正整数 k , 都有

$$P(I + A)^k P^T = \begin{bmatrix} \tilde{A}_{11} & A_{12} \\ \mathbf{0} & \tilde{A}_{22} \end{bmatrix}^k,$$

故对所有的上述 k 值, 有

$$(I + A)^k = P^T \begin{bmatrix} \tilde{A}_{11} & A_{12} \\ \mathbf{0} & \tilde{A}_{22} \end{bmatrix}^k P,$$

此等式表明无论正整数 k 为何值, $(I + A)^k$ 中永远有零元素, 因此 $(I + A)^s > 0$ 是不可能的. 这就证明 A 不可能是可约的. 证毕

例如, 非负矩阵

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

是不可约的. 因为 $s = 3 - 1 = 2$ 时, 即有

$$(I + A)^2 = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}^2 = \begin{pmatrix} 5 & 4 & 1 \\ 4 & 6 & 4 \\ 1 & 4 & 5 \end{pmatrix} > 0.$$

在前面, 我们证明了正矩阵的佩龙定理, 而正矩阵是不可约非负矩阵的一种特殊情形. 1912年, 弗罗贝尼乌斯(Frobenius)把上述定理推广到不可约非负矩阵上.

定理 7.1.9 (佩龙-弗罗贝尼乌斯定理) 设 $A \in \mathbb{R}^{n \times n}$ 是不可约非负矩阵, 则

(1) A 有一正实特征值恰等于它的谱半径 $\rho(A)$, 并且存在正向量 $x \in \mathbb{R}^n$, 使得 $Ax = \rho(A)x$;

(2) $\rho(A)$ 是 A 的单特征值;

(3) 当 A 的任意元素(一个或多个)增加时, $\rho(A)$ 增加.

* 证明 (1) 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

$$0 \leq A < B_{k+1} < B_k,$$

由定理 7.1.3 的推论 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

$$\rho(A) \leq \lambda. \quad (7.1.21)$$

因为 $B_k > 0$, 则由定理 7.1.4(佩龙定理)知, 存在向量 $y_k = (y_1^{(k)}, \dots, y_n^{(k)})^T > 0$, 使得

$$B_k y_k = \rho(B_k) y_k. \quad (7.1.22)$$

令

$$x_j^{(k)} = \left(\sum_{i=1}^n (y_i^{(k)})^2 \right)^{-\frac{1}{2}} \cdot y_j^{(k)}, \quad x_k = (x_1^{(k)}, \dots, x_n^{(k)})^T,$$

则 $x_k > 0$, 并且 $\|x_k\|_2 = 1$, 同时有

$$B_k x_k = \rho(B_k) x_k.$$

令 $S = \{x \geq 0 \mid \|x\|_2 = 1, x \in \mathbb{R}^n\}$, 则 S 是 \mathbb{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

$$\lim_{m \rightarrow \infty} x_{k_m} = x \in S.$$

因为

$$\begin{aligned} \lambda x &= \lim_{m \rightarrow \infty} \rho(B_{k_m}) \lim_{m \rightarrow \infty} x_{k_m} = \lim_{m \rightarrow \infty} (\rho(B_{k_m}) x_{k_m}) \\ &= \lim_{m \rightarrow \infty} (B_{k_m} x_{k_m}) = Ax, \end{aligned}$$

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 $\lambda \leq \rho(A)$, 再由式(7.1.21)得

$$\lambda = \rho(A), \quad (7.1.23)$$

也就是说 $\rho(A)$ 是 A 的特征值, 且有

$$Ax = \rho(A)x, \quad (7.1.24)$$

其中 $x \geq 0$ 且 $x \neq 0$.

下面证明 $\rho(A) > 0$ 和 $x > 0$. 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 $b_{ij} = \alpha a_{ij} \left(\sum_{j=1}^n a_{ij} \right)^{-1}$. 则 $0 \leq B \leq A$, 并且 $\sum_{j=1}^n b_{ij} = \alpha (i=1, \dots, n)$, 即 B 的每一行之和是常数 α , 故 $\|B\|_\infty = \alpha$. 令 $x = (1, \dots, 1)^T$, 则 $Bx = \alpha x = \|B\|_\infty x$, 这说明 x 是 B 对应于特征值 $\|B\|_\infty$ 的特征向量. 又因为对任意相容矩阵范数 $\|\cdot\|$ 有 $\rho(B) \leq \|B\|$, 故 $\rho(B) = \|B\|_\infty = \alpha$. 再由定理 7.1.3 推论 1 得 $\rho(B) \leq \rho(A)$, 则有

$$\rho(A) \geq \|B\|_\infty = \alpha > 0,$$

因此, $\rho(A)$ 是 A 的正特征值.

显然, $1+\rho(\mathbf{A})$ 是 $\mathbf{I}+\mathbf{A}$ 的特征值, 即存在非负向量 $\mathbf{x} \in \mathbb{R}^n$ 且 $\mathbf{x} \neq \mathbf{0}$, 使得 $(\mathbf{I}+\mathbf{A})\mathbf{x} = (1+\rho(\mathbf{A}))\mathbf{x}$, 从而有 $(\mathbf{I}+\mathbf{A})^{n-1}\mathbf{x} = (1+\rho(\mathbf{A}))^{n-1}\mathbf{x}$. 从定理 7.1.7 可知 $(\mathbf{I}+\mathbf{A})^{n-1} > \mathbf{0}$, 而由定理 7.1.2(7) 有 $(\mathbf{I}+\mathbf{A})^{n-1}\mathbf{x} > \mathbf{0}$. 因此 $\mathbf{x} = (1+\rho(\mathbf{A}))^{1-n}(\mathbf{I}+\mathbf{A})^{n-1}\mathbf{x} > \mathbf{0}$, 这就说明特征向量 \mathbf{x} 是正的. 故(1)得证.

为了证明(2), 采用反证法. 如果 $\rho(\mathbf{A})$ 是 \mathbf{A} 的重特征值, 则 $1+\rho(\mathbf{A}) = \rho(\mathbf{I}+\mathbf{A})$ 是 $\mathbf{I}+\mathbf{A}$ 的重特征值, 从而 $(1+\rho(\mathbf{A}))^{n-1} = (\rho(\mathbf{I}+\mathbf{A}))^{n-1} = \rho((\mathbf{I}+\mathbf{A})^{n-1})$ 是 $(\mathbf{I}+\mathbf{A})^{n-1}$ 的重特征值. 另一方面, 因为 $(\mathbf{I}+\mathbf{A})^{n-1} > \mathbf{0}$ (正矩阵), 由定理 7.1.4 知 $\rho((\mathbf{I}+\mathbf{A})^{n-1})$ 是 $(\mathbf{I}+\mathbf{A})^{n-1}$ 的单特征值. 这个矛盾说明 $\rho(\mathbf{A})$ 是 \mathbf{A} 的单特征值.

(3) 由定理 7.1.3 推论 1 即得.

证毕

例 7.1.2 对于不可约非负矩阵

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

其谱半径 $\rho(\mathbf{A}) = 1 + \sqrt{10}$ 就是它的一个正的单特征值, 而属于 $\rho(\mathbf{A})$ 的正特征向量是 $(2, \sqrt{10}, 2)^T$. 并且, “模等于 $\rho(\mathbf{A})$ ” 的特征值 $1 + \sqrt{10}$ 也只有一个.

值得提出的是, 对于一般不可约非负矩阵 \mathbf{A} , 佩龙-弗罗贝尼乌斯定理并不能保证 \mathbf{A} 的“模等于 $\rho(\mathbf{A})$ ”的特征值是惟一的. 现看一个简单例子.

例 7.1.3 设

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}_{n \times n},$$

不难验证 \mathbf{A} 是不可约非负矩阵, 它的几个特征值是

$$\lambda_j = e^{i\frac{2j\pi}{n}}, \quad j = 0, 1, \dots, n-1,$$

从而它的特征值的模都等于谱半径 $\rho(\mathbf{A}) = 1$.

注 n 阶非负不可约矩阵 \mathbf{A} 的“模等于 $\rho(\mathbf{A})$ ”的 m 个不同特征值可以表示为

$$\lambda_j = \rho(\mathbf{A}) e^{i\frac{2j\pi}{m}}, \quad j = 0, 1, \dots, m-1, \quad (7.1.25)$$

也就是说, 它们“均匀”地分布在以原点为圆心、 $\rho(\mathbf{A})$ 为半径的圆周上(证明从略).

例 7.1.4 设

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

则 \mathbf{A} 是不可约非负的, 它的特征值是

$$\lambda_{1,2} = \pm \sqrt{1+\sqrt{2}}, \quad \lambda_{3,4} = \pm i \sqrt{\sqrt{2}-1},$$

其特征值的分布情况, 如图 7.1 所示.

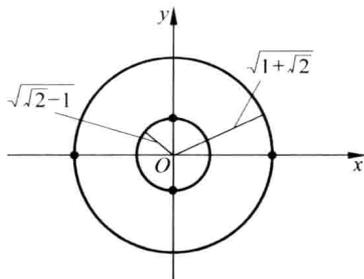


图 7.1 黑点表示特征值所在位置

对于不可约非负矩阵,定理 7.1.9 连同第 4 章的圆盘定理可以给出谱半径的估计界限,它在理论上尤其在矩阵迭代分析中有重要应用.

定理 7.1.10 设 $A = (a_{ij})_{n \times n}$ 为不可约非负矩阵,则或者

$$\sum_{j=1}^n a_{ij} = \rho(A), \quad i = 1, 2, \dots, n, \quad (7.1.26)$$

或者

$$\min_{1 \leq i \leq n} \sum_{j=1}^n a_{ij} < \rho(A) < \max_{1 \leq i \leq n} \sum_{j=1}^n a_{ij}. \quad (7.1.27)$$

证明 若 A 的每行元素之和均等于 μ , 令 $\xi = (1, 1, \dots, 1)^T$, 则有 $A\xi = \mu\xi$, 所以 μ 为 A 的一个特征值, 因而 $\mu \leq \rho(A)$. 另一方面, 由圆盘定理, 有某个 $i, 1 \leq i \leq n$, 使 $|\rho(A) - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij}$, 所以 $\rho(A) \leq a_{ii} + \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} = \mu$, 因此 $\mu = \rho(A)$, 这时式(7.1.26)成立. 现在假设 A 的各行之和不全相同, 则可以用减小(或增加) A 的某些正元素的方法得到一个不可约非负矩阵 $B = (b_{ij})_{n \times n}$ (或 $C = (c_{ij})_{n \times n}$), 使得

$$\sum_{j=1}^n b_{ij} = a = \min_{1 \leq k \leq n} \sum_{j=1}^n a_{kj}, \quad \sum_{j=1}^n c_{ij} = b = \max_{1 \leq k \leq n} \sum_{j=1}^n a_{kj}, \quad 1 \leq i \leq n,$$

因而 $\rho(B) = a, \rho(C) = b$. 因 $B \leq A \leq C$, 由定理 7.1.3 的推论 1 知 $\rho(B) \leq \rho(A) \leq \rho(C)$. 如果 $\rho(A) = \rho(C)$, 即 $\rho(A) = \max_{1 \leq k \leq n} \sum_{j=1}^n a_{kj} = b$, 因 A 为不可约非负矩阵, 所以由定理 7.1.9, A 有正的对应于特征值 $\rho(A)$ 的特征向量 $y = (y_1, y_2, \dots, y_n)^T$, 因此 $Ay = \rho(A)y = by$, 即

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

设 $y_{i_1} = y_{i_2} = \dots = y_{i_s} = \max\{y_1, y_2, \dots, y_n\}$, 则可设

$$\{y_1, y_2, \dots, y_n\} = \{y_{i_1}, y_{i_2}, \dots, y_{i_s}\} \cup \{y_{j_1}, y_{j_2}, \dots, y_{j_t}\}.$$

显然 $s+t=n$. 若 $s < n$, 则因

$$\sum_{j=1}^n a_{i_k, j}y_j = by_{i_k}, \quad k = 1, 2, \dots, s,$$

所以

$$by_{i_k} = \sum_{j=1}^n a_{i_k, j}y_j \leq \sum_{j=1}^n a_{i_k, j}y_{i_k} \leq by_{i_k}, \quad k = 1, 2, \dots, s,$$

因此

$$\sum_{j=1}^n a_{i_k, j}y_j = \sum_{j=1}^n a_{i_k, j}y_{i_k}, \quad k = 1, 2, \dots, s. \quad (7.1.28)$$

这说明

$$a_{i_k, j_1} = a_{i_k, j_2} = \dots = a_{i_k, j_t} = 0, \quad k = 1, 2, \dots, s.$$

否则的话, 式(7.1.28)等号不成立. 因此推出, A 的第 i_1, i_2, \dots, i_s 行中, 除第 i_1, i_2, \dots, i_s 列外, 其余各列元素均为零. 因此, A 为可约的, 与 A 为不可约非负矩阵矛盾, 所以 $s=n$, 这说明 $y = c(1, 1, \dots, 1)^T$ 对某个 $c \in \mathbb{R}, c \neq 0$ 成立.

因此

$$\sum_{j=1}^n a_{ij} = b, \quad 1 \leq i \leq n$$

这与 \mathbf{A} 的各行之和不全相同的假设相矛盾, 因此 $\rho(\mathbf{A}) \neq b$. 同样可证 $\rho(\mathbf{A}) \neq a$. 于是, $a < \rho(\mathbf{A}) < b$. 证毕

推论 \mathbf{A} 为不可约非负矩阵, 则对任意给定的正向量 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$, 或者有

$$\frac{1}{x_i} \sum_{j=1}^n a_{ij} x_j = \rho(\mathbf{A}), \quad i = 1, 2, \dots, n \quad (7.1.29)$$

或者有

$$\min_{1 \leq i \leq n} \left(\frac{1}{x_i} \sum_{j=1}^n a_{ij} x_j \right) < \rho(\mathbf{A}) < \max_{1 \leq i \leq n} \left(\frac{1}{x_i} \sum_{j=1}^n a_{ij} x_j \right). \quad (7.1.30)$$

证明 给定 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T > 0$, 令 $\mathbf{D} = \text{diag}(x_1, x_2, \dots, x_n)$, 将定理 7.1.10 应用于非负不可约矩阵 $\mathbf{B} = \mathbf{D}^{-1} \mathbf{A} \mathbf{D}$, 便得到推论结论. 证毕

7.1.3 素矩阵与循环矩阵

现转到非负矩阵进一步的分类问题上. 为此, 引进一类介于不可约非负矩阵与正矩阵之间的矩阵——素矩阵与循环矩阵的概念. 素矩阵有多种不同的定义方式, 这里采用按谱半径的重数来定义, 另外的方式作为性质.

定义 7.1.3 设 \mathbf{A} 是 n 阶非负矩阵, 且有 m 个特征值的模均等于谱半径 $\rho(\mathbf{A})$, 则当 $m=1$ 时, 就称方阵 \mathbf{A} 为素矩阵(或本原矩阵); 当 $m>1$ 时, 就称 \mathbf{A} 是循环矩阵(或非素矩阵). m 统称为 \mathbf{A} 的非素性指标.

例如, 前面例 7.1.2 中的 \mathbf{A} 是一个素矩阵, 而例 7.1.3 及例 7.1.4 中的 \mathbf{A} 分别是指标为 n 和 2 的循环矩阵.

又如, 正矩阵都是素矩阵, 但反之不真.

由定义 7.1.3 即得如下结论.

定理 7.1.11 设 \mathbf{A}, \mathbf{B} 均为 n 阶非负矩阵, 并且 \mathbf{A} 是素矩阵, 则

- (1) \mathbf{A}^T 也是素矩阵;
- (2) 对任一正整数 k , \mathbf{A}^k 也是素矩阵;
- (3) $\mathbf{A} + \mathbf{B}$ 也是素矩阵.

定理 7.1.12 非负矩阵 \mathbf{A} 是素矩阵(本原矩阵)的充分必要条件, 是存在某个正整数 k , 使得 $\mathbf{A}^k > \mathbf{0}$.

(证略.)

例如, 非负矩阵

$$\mathbf{A} = \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

都是素矩阵, 因为不难验证 $\mathbf{A}^2 > \mathbf{0}$, $\mathbf{B}^4 > \mathbf{0}$.

类似地可以证明, 对于素矩阵 \mathbf{A} , 佩龙定理以及定理 7.1.6 的结论仍然成立.

例如, 对于一个非负素矩阵 \mathbf{A} , 除特征值 $\rho(\mathbf{A})$ 外, 其余特征值的模都小于 $\rho(\mathbf{A})$. (正矩阵也有这个性质, 但对不可约非负矩阵这个结论不再成立, 见例 7.1.3)

本节最后要注意的是, 前面介绍的佩龙-弗罗贝尼乌斯定理不能照搬到可约非负矩阵上. 但是, 由于任一非负矩阵 $\mathbf{A} \geq \mathbf{0}$ 都可表示成不可约的正矩阵序列 $\{\mathbf{A}_m\}$ 的极限:

$$\mathbf{A} = \lim_{m \rightarrow \infty} \mathbf{A}_m, \quad \forall \mathbf{A}_m > 0, \quad (7.1.31)$$

所以不可约非负矩阵的某些性质,在较弱的形式下,对于可约非负矩阵亦能成立.下面我们把这种较弱的形式,不加证明地用定理表述如下:

定理 7.1.13 设 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 为非负矩阵,则有结论:

(1) $\rho(\mathbf{A})$ 是 \mathbf{A} 的特征值,且属于 $\rho(\mathbf{A})$ 的特征向量可取作非负的,即存在不为零的非负向量 \mathbf{x} ,使得 $\mathbf{Ax} = \rho(\mathbf{A})\mathbf{x}$ (注意,这里 $\rho(\mathbf{A})$ 和 \mathbf{x} 不一定是正的);

(2) \mathbf{A} 的特征值可分成若干组,每组中的特征值模都相等,而且“均匀”地分布在以原点为圆心的某一圆周上(注意,这里 \mathbf{A} 的所有特征值的模都不超过即小于等于 $\rho(\mathbf{A})$).

7.2 随机矩阵与双随机矩阵

这里介绍一类重要的非负矩阵——随机矩阵,并简要介绍随机矩阵一些性质以及它的应用背景.

定义 7.2.1 设 $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{n \times n}$ 是非负矩阵,如果 \mathbf{A} 的每一行上的元素之和都等于 1,即

$$\sum_{j=1}^n a_{ij} = 1, \quad i = 1, 2, \dots, n, \quad (7.2.1)$$

则称 \mathbf{A} 为随机矩阵;如果 \mathbf{A} 还满足

$$\sum_{i=1}^n a_{ij} = 1, \quad j = 1, 2, \dots, n, \quad (7.2.2)$$

则称 \mathbf{A} 为双随机矩阵.

\mathbf{A} 之所以称为随机矩阵,是因为 \mathbf{A} 的每一行可以看成有 n 个点的样本空间上的离散概念分布.这样的矩阵常常出现在城市间的人口流动模型、马尔可夫(Markov)链的研究及经济学和运筹学等领域的各种各样的数学模型问题中.

随机矩阵是一类特殊的非负矩阵,因此上节所述的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.

定理 7.2.1 设 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 是随机矩阵,则有

$$\rho(\mathbf{A}) = 1. \quad (7.2.3)$$

证明 因为 \mathbf{A} 是随机矩阵,所以 \mathbf{A} 的每一行元素之和为 1,则 $\|\mathbf{A}\|_{\infty} = 1$. 令 $\mathbf{x} = (1, \dots, 1)^T$,显然 $\mathbf{Ax} = \mathbf{x} = \|\mathbf{A}\|_{\infty} \mathbf{x}$,即 \mathbf{x} 是 \mathbf{A} 对应于特征值 $\|\mathbf{A}\|_{\infty}$ 的特征向量,而 $\rho(\mathbf{A}) \leq \|\mathbf{A}\|_{\infty}$,同时又有 $\|\mathbf{A}\|_{\infty} \leq \rho(\mathbf{A})$,故得 $\rho(\mathbf{A}) = \|\mathbf{A}\|_{\infty} = 1$. 证毕

从以上证明可知, n 阶随机矩阵 \mathbf{A} 有特征值 1,并且有相应的特征向量 $\mathbf{x} = (1, \dots, 1)^T$;反之,如果 n 阶非负矩阵 \mathbf{A} 有特征值 1 且对应于 1 的特征向量为 $\mathbf{x} = (1, \dots, 1)^T$,则 \mathbf{A} 是随机矩阵.于是,我们得到如下定理:

定理 7.2.2 n 阶非负矩阵 \mathbf{A} 是随机矩阵的充分必要条件是 $\mathbf{x} = (1, \dots, 1)^T \in \mathbb{R}^n$ 为 \mathbf{A} 对应于特征值 1 的特征向量,即 $\mathbf{Ax} = \mathbf{x}$.

容易验证,同阶随机矩阵之积仍是随机矩阵.

由定理 7.1.10 可知,随机矩阵 \mathbf{A} 的谱半径 $\rho(\mathbf{A}) = 1$.

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 7.2.3 设 n 阶非负矩阵 \mathbf{A} 的谱半径 $\rho(\mathbf{A}) > 0$,且有 $\mathbf{x} = (x_1, \dots, x_n)^T > 0$,则矩阵

\mathbf{A} 能相似于数 $\rho(\mathbf{A})$ 与某个随机矩阵 \mathbf{P} 的乘积, 即

$$\mathbf{A} = \mathbf{D}(\rho(\mathbf{A})\mathbf{P})\mathbf{D}^{-1}, \quad (7.2.4)$$

其中 $\mathbf{D} = \text{diag}(x_1, \dots, x_n)$. 即 $(\mathbf{D}^{-1}\mathbf{A}\mathbf{D})/\rho(\mathbf{A})$ 是随机矩阵.

证明 因为

$$\sum_{j=1}^n a_{ij}x_j = \rho(\mathbf{A})x_i, \quad i = 1, \dots, n, \quad (7.2.5)$$

引入对角矩阵

$$\mathbf{D} = \text{diag}(x_1, \dots, x_n)$$

及矩阵

$$\mathbf{P} = \frac{1}{\rho(\mathbf{A})}\mathbf{D}^{-1}\mathbf{A}\mathbf{D},$$

则

$$p_{ij} = \frac{1}{\rho(\mathbf{A})}x_i^{-1}a_{ij}x_j \geq 0, \quad i, j = 1, \dots, n,$$

而由式(7.2.5)可得

$$\sum_{j=1}^n p_{ij} = 1, \quad i = 1, \dots, n,$$

即 \mathbf{P} 是随机矩阵, 且有 $\mathbf{A} = \mathbf{D}(\rho(\mathbf{A})\mathbf{P})\mathbf{D}^{-1}$. 证毕

随机矩阵在随机过程中有着重要的应用. 设某个过程或系统可能出现 n 个随机事件 S_1, \dots, S_n , 且在时间序列 t_0, t_1, t_2, \dots 的每一瞬间, 这些事件有一个且只有一个能够出现. 如果在时刻 t_{k-1} ($k \geq 1$) 处于事件 S_i , 则下一时刻 t_k 将以概率 $p_{ij}(k)$ 转移到事件 S_j ($i, j = 1, \dots, n$; $k = 1, 2, \dots$). 若对所有 $k \geq 1$ 概率 $p_{ij}(k)$ 与 k 无关, 则称这个过程为纯马尔可夫链.

当给出了条件概率矩阵 $\mathbf{P} = (p_{ij})_{n \times n}$ 时, 显然满足

$$p_{ij} \geq 0, \quad \sum_{j=1}^n p_{ij} = 1, \quad i, j = 1, 2, \dots, n,$$

即 \mathbf{P} 是随机矩阵, 我们称之为该过程的转移矩阵.

在实际应用中常要考虑到随机矩阵 \mathbf{A} 的幂序列 $\{\mathbf{A}^m\}$ 的收敛性. 由于 \mathbf{A} 的谱半径 $\rho(\mathbf{A}) = 1$, 且 \mathbf{A} 的任一模等于 1 的特征值所对应的若尔当块都是一阶的(证略), 故有如下结果.

定理 7.2.4 设 \mathbf{A} 为不可约随机矩阵, 则极限 $\lim_{m \rightarrow \infty} \mathbf{A}^m$ 存在的充分必要条件是 \mathbf{A} 为本原矩阵.

下面简单考虑一下随机矩阵在齐次马尔可夫链中的应用.

设某个过程可能出现 n 个状态 S_1, S_2, \dots, S_n , 假如过程从状态 S_i 的概率 a_{ij} 只依赖于这两个状态, 则称该过程为有限马尔可夫过程. 令 $\mathbf{A} = (a_{ij})_{n \times n}$, 则 $a_{ij} \geq 0$, 且 $\sum_{i=1}^n a_{ij} = 1$, 即 \mathbf{A} 为随机矩阵, 称 \mathbf{A} 为该过程的转移矩阵.

用 $(\mathbf{A}, \mathbf{p}^{(0)})$ 表示某个有限齐次马尔可夫过程, 其中 $\mathbf{A} = (a_{ij})_{n \times n}$ 为转移矩阵, $\mathbf{p}^{(0)} = (p_1^{(0)}, p_2^{(0)}, \dots, p_n^{(0)})^T$ 为初始(概率)分布向量, $p_j^{(0)}$ 表示过程初始处在状态 S_j 的概率, $1 \leq j \leq n$. 设 $\mathbf{p}^{(k)} = (p_1^{(k)}, p_2^{(k)}, \dots, p_n^{(k)})^T$ 为第 k 个(概率)分布向量, 其中 $p_j^{(k)}$ 为过程在第 k 步

后处在状态 S_j 的概率, $1 \leq j \leq n$. 显然 $p_j^{(k)} \geq 0$, 且 $\sum_{j=1}^n p_j^{(k)} = 1$. 由假定, 有

$$p_j^{(k)} = \sum_{i=1}^n a_{ij} p_i^{(k-1)}, \quad 1 \leq j \leq n, \quad k = 1, 2, \dots$$

因此有

$$\mathbf{p}^{(k)} = \mathbf{A}^T \mathbf{p}^{(k-1)}, \quad k = 1, 2, \dots$$

因而

$$\mathbf{p}^{(k)} = (\mathbf{A}^T)^k \mathbf{p}^{(0)} = (\mathbf{A}^k)^T \mathbf{p}^{(0)}, \quad k = 1, 2, \dots \quad (7.2.6)$$

在马尔可夫链的研究中, 一个重要问题是讨论当 $k \rightarrow \infty$ 时, $\mathbf{p}^{(k)}$ 的变化趋势. 按式(7.2.6)这主要取决于 \mathbf{A}^k 的变化趋势.

例 7.2.1 设某少数民族现有 1800 人, 居住在 A, B, C 三个部落的人数分别为 200, 600, 1000, 假定每年每个部落的所有人分为相等的两半分别迁往其他两个部落, 试问一年、两年以及无限长久之后, 该民族在三个部落的人口分布情况.

解 设 $\mathbf{p}^{(k)} = (p_1^{(k)}, p_2^{(k)}, p_3^{(k)})^T$ 为 k 年后分布在三个部落的人口比率向量, 其中 $p_1^{(k)}$, $p_2^{(k)}$, $p_3^{(k)}$ 分别为 k 年后分布在 A, B, C 三个部落的人口比率. 设 $\mathbf{p}^{(0)} = (p_1^{(0)}, p_2^{(0)}, p_3^{(0)})^T$ 为初始人口分布比率向量. 由假定, 有

$$\mathbf{p}^{(0)} = (p_1^{(0)}, p_2^{(0)}, p_3^{(0)})^T = \left(\frac{1}{9}, \frac{1}{3}, \frac{5}{9}\right)^T,$$

过程的转移矩阵为

$$\mathbf{A} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix},$$

因此

$$\mathbf{p}^{(k)} = (\mathbf{A}^k)^T \mathbf{p}^{(0)} = \mathbf{A}^k \mathbf{p}^{(0)}, \quad k = 1, 2, \dots.$$

令 $k=1$, 则

$$\mathbf{p}^{(1)} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{9} \\ \frac{1}{3} \\ \frac{5}{9} \end{bmatrix} = \begin{bmatrix} \frac{4}{9} \\ \frac{1}{3} \\ \frac{2}{9} \end{bmatrix},$$

$k=2$, 则

$$\mathbf{p}^{(2)} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} \frac{4}{9} \\ \frac{1}{3} \\ \frac{2}{9} \end{bmatrix} = \begin{bmatrix} \frac{5}{18} \\ \frac{1}{3} \\ \frac{7}{18} \end{bmatrix},$$

故一年后分布在 A, B, C 三部落的人数分别为 800, 600, 400; 两年后分布在 A, B, C 三部落的人数分别为 500, 600, 700.

由于 $\mathbf{I} + \mathbf{A} > \mathbf{0}$, 故由定理 7.1.8, \mathbf{A} 为不可约非负矩阵, 计算得

$$\mathbf{A}^2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} > \mathbf{0},$$

所以由定理 7.1.12 知 \mathbf{A} 为本原矩阵(素矩阵). 再由定理 7.2.4 知 $\lim_{k \rightarrow \infty} \mathbf{A}^k$ 存在. 因 $|\lambda \mathbf{I} - \mathbf{A}| = (\lambda - 1) \left(\lambda + \frac{1}{2} \right)^2$, 所以 \mathbf{A} 有特征值 $\lambda_1 = 1, \lambda_2 = -\frac{1}{2}$. 分别解齐次方程组

$$(\lambda_1 \mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0} \quad \text{和} \quad (\lambda_2 \mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$$

得相互正交的特征向量为

$$\xi_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \xi_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad \xi_3 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}.$$

令 $\mathbf{P} = (\xi_1, \xi_2, \xi_3)$, 则

$$\mathbf{A} = \mathbf{P} \begin{bmatrix} 1 & & \\ & \frac{1}{2} & \\ & & -\frac{1}{2} \end{bmatrix} \mathbf{P}^{-1},$$

所以

$$\mathbf{A}^k = \mathbf{P} \begin{bmatrix} 1 & & \\ & \left(\frac{1}{2}\right)^k & \\ & & \left(-\frac{1}{2}\right)^k \end{bmatrix} \mathbf{P}^{-1}.$$

因此

$$\lim_{k \rightarrow \infty} \mathbf{A}^k = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 1 & & \\ & 0 & \\ & & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

于是

$$\begin{aligned} \lim_{k \rightarrow \infty} \mathbf{p}^{(k)} &= (\lim_{k \rightarrow \infty} (\mathbf{A}^k)^T) \mathbf{p}^{(0)} = (\lim_{k \rightarrow \infty} \mathbf{A}^k) \mathbf{p}^{(0)} \\ &= \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{9} \\ \frac{1}{3} \\ \frac{5}{9} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}, \end{aligned}$$

因此在无限年后分布在 A, B, C 的三个部落的人分别为 600, 600, 600.

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 7.2.5 设 $A \in \mathbb{R}^{n \times n}$ 是双随机矩阵,则

- (1) $\rho(A) = 1$, 且 $x = (1, \dots, 1)^T$ 是 A 与 A^T 对应于特征值 1 的特殊向量;
- (2) $\|A\|_2 \geq 1$.

7.3 单调矩阵

本节简要介绍一类矩阵 A , 其特点是它的逆矩阵 A^{-1} 是非负的矩阵——单调矩阵, 并说明它在求解线性方程组中的应用.

定义 7.3.1 设 $A \in \mathbb{R}^{n \times n}$, 如果它的逆矩阵 $A^{-1} \geq 0$, 则称 A 为单调矩阵.

例如, 矩阵

$$A = \begin{pmatrix} 1 & -\frac{1}{2} & \frac{1}{8} \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix}$$

的逆矩阵为

$$A^{-1} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{8} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix} \geq 0,$$

故 A 为单调矩阵.

下面给出 20 世纪 50 年代由卡拉茨(Collatz)提出的判别矩阵 A 为单调矩阵的一个充分必要条件, 它可作为单调矩阵概念的等价条件.

定理 7.3.1 设 $A \in \mathbb{R}^{n \times n}$, 则 A 为单调矩阵的充分必要条件是: 可从 $Ax \geq 0$ 推出 $x \geq 0$, 这里 x 是列向量.

证明 必要性: 如果 A 是单调矩阵, 则 $A^{-1} \geq 0$. 若 $Ax \geq 0$, 则必有 $A^{-1}(Ax) \geq 0$, 所以 $x \geq 0$.

充分性: 反之, 若可从 $Ax \geq 0$ 推出 $x \geq 0$, 则 A 非奇异. 事实上, 设 $Ax = 0$ 有解 \bar{x} , 即 $A\bar{x} = 0$, 于是 $A\bar{x} \geq 0$, 由假设知 $\bar{x} \geq 0$; 再由

$$A(-\bar{x}) = -A\bar{x} = 0,$$

又可推得 $-\bar{x} \geq 0$, 故只能 $\bar{x} = 0$, 从而 $Ax = 0$ 仅有零解, 故 A 非奇异, 即 A^{-1} 存在. 证毕

从单调矩阵的定义可推知: 若 A 是单调矩阵, 则 A 是非奇异的. 于是线性方程组 $Ax = b$ 有惟一解 $\bar{x} = (\bar{x}_1, \dots, \bar{x}_n)^T$, 并有如下结果.

定理 7.3.2 设 A 为单调矩阵, 若能找到向量 $x' = (x'_1, \dots, x'_n)^T$ 和 $x'' = (x''_1, \dots, x''_n)^T$ 分别使 $Ax' \leq b, Ax'' \geq b$, 则有估计式

$$x' \leq \bar{x} \leq x'' \quad (7.3.1)$$

或

$$x'_i \leq \bar{x}_i \leq x''_i, \quad i = 1, \dots, n. \quad (7.3.2)$$

证明 由于 $Ax' \leq b$, 因此, $Ax' - b \leq 0$, 又 A 是单调矩阵, 即有 $A^{-1} \geq 0$, 于是有

$$A^{-1}(Ax' - b) \leq 0,$$

即

$$x' - A^{-1}b \leq 0.$$

由 $\tilde{x} = A^{-1}b$, 便得 $x' \leq \tilde{x}$. 类似地有 $x'' \geq \tilde{x}$.

证毕

式(7.3.1)的意义在于: 当找到满足 $Ax' \leq b$ 的向量 x' , 便直接得到 $x' \leq \tilde{x}$, 即 x' 为解向量 \tilde{x} 的下界. 同样, 只要找到满足 $Ax'' \geq b$ 的 x'' , 便知 x'' 是 \tilde{x} 的上界.

作为式(7.3.1)的应用, 考虑线性方程组

$$\begin{cases} x_1 - \frac{1}{4}x_2 - \frac{1}{4}x_3 = 0.6, \\ -\frac{1}{4}x_1 + x_2 - \frac{1}{4}x_4 = 0.6, \\ -\frac{1}{4}x_1 + x_3 - \frac{1}{4}x_4 = 0.6, \\ -\frac{1}{4}x_2 - \frac{1}{4}x_3 + x_4 = 0.66, \end{cases} \quad (7.3.3)$$

系数矩阵 A 的逆矩阵是

$$A^{-1} = \begin{pmatrix} \frac{7}{6} & \frac{2}{6} & \frac{2}{6} & \frac{1}{6} \\ \frac{2}{6} & \frac{7}{6} & \frac{1}{6} & \frac{2}{6} \\ \frac{2}{6} & \frac{1}{6} & \frac{7}{6} & \frac{2}{6} \\ \frac{1}{6} & \frac{2}{6} & \frac{2}{6} & \frac{7}{6} \end{pmatrix} \geq 0.$$

若取 $x' = (1.2, 1.2, 1.2, 1.2)^T$, 容易计算出

$$Ax' = (0.6, 0.6, 0.6, 0.6)^T,$$

于是

$$Ax' \leq b = (0.6, 0.6, 0.6, 0.66)^T$$

成立, 故原方程组的解的下界为 $\tilde{x}_i \geq 1.2 (i=1, 2, 3, 4)$. 实际上, 该方程组的精确解为

$$\tilde{x} = (1.21, 1.22, 1.22, 1.27)^T.$$

7.4 M 矩阵与 H 矩阵

1937年, 奥斯乔斯基(Ostrowski)发现一类具有特殊构造的矩阵, 其非对角元素 ($i \neq j$) $a_{ij} \leq 0$, 即这种矩阵 A 都可以表示成 $A = sI - B$, 且 $s > 0, B \geq 0$, 故这种矩阵与非负矩阵有一定的联系, 称为闵科夫斯基(Minkovski)矩阵, 简称 M 矩阵. 随后, 数学家与经济学家将矩阵 A 推广到复矩阵, 在 M 矩阵的基础上又提出了 H 矩阵的概念. M 矩阵和 H 矩阵在偏微分方程的有限差分法、经济学中的投入产出法、运筹学中的线性余问题及概率统计的马尔可夫过程等很多领域都具有重要的应用. 本节先介绍 M 矩阵及其基本性质, 再简要介绍 H 矩阵的概念.

7.4.1 M 矩阵

定义 7.4.1 设 $A \in \mathbb{R}^{n \times n}$, 且可表示为

$$A = sI - B, \quad s > 0, \quad B \geq 0. \quad (7.4.1)$$

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.

例 7.4.1

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

是一个非奇异 M 矩阵, 因为 $A = 2I - B$, 其中 $s = 2 > 0$,

$$B = \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix} \geq 0,$$

且 $\rho(B) = 1, s > \rho(B)$.

从此例可看出, M 矩阵不一定是对称矩阵, 而且从后面的定理 7.4.2(4) 中可以发现, M 矩阵的对角元素 a_{ii} 总是正的. 为了进一步讨论 M 矩阵的性质, 引入所谓 Z 型矩阵的概念.

设 $A = (a_{ij})_{n \times n}$, 且

$$a_{ij} \leq 0, \quad i \neq j, i, j = 1, 2, \dots, n, \quad (7.4.2)$$

则称 A 为 Z 型矩阵. 全体 n 阶 Z 型矩阵的集合用记号 $Z^{n \times n}$ 表示. 显然, M 矩阵是 Z 型矩阵的特殊情况.

下面先给出非奇异 M 矩阵的一些特性.

定理 7.4.1 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, 且 $A^{-1} \geq D^{-1} \geq 0$;
- (2) D 的每个实特征值为正数;
- (3) $\det D \geq \det A > 0$.

证明 (1) 由假设有

$$A = sI - B, \quad B \geq 0, \quad s > \rho(B),$$

对任意给定的实数 $\omega \leq 0$, 考察矩阵

$$C = A - \omega I = (s - \omega)I - B.$$

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为正数. 由于 $D \in Z^{n \times n}$, 故存在足够小的正数 ϵ , 使得

$$P = I - \epsilon D \geq 0.$$

因为 $D \geq A$, 所以 $Q = I - \epsilon A \geq I - \epsilon D = P \geq 0$. 由定理 7.1.11 知 $\rho(Q)$ 为 Q 的非负特征值, 故有

$$\det[(1 - \rho(Q))I - \epsilon A] = \det(Q - \rho(Q)I) = 0.$$

由此可知 $\frac{1}{\varepsilon}(1-\rho(Q))$ 为 A 的实特征值. 由于上面已证非奇异 M 矩阵的特征值必为正数, 所以 $1-\rho(Q)>0$, 于是 $0\leq\rho(Q)<1$. 根据定理 5.2.5 得

$$(\varepsilon A)^{-1} = (I - Q)^{-1} = I + Q + Q^2 + \cdots \geq 0,$$

从而有 $A^{-1} \geq 0$. 又由定理 7.1.2(5) 有

$$0 \leq P^k \leq Q^k, \quad k = 1, 2, \cdots.$$

而由定理 7.1.3 推论 1 得 $\rho(P) \leq \rho(Q) < 1$, 于是有

$$(I - P)^{-1} = (\varepsilon D)^{-1} = I + P + P^2 + \cdots \leq (\varepsilon A)^{-1},$$

得 $A^{-1} \geq D^{-1} \geq 0$, 即(1)得证.

(2) 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.

(3) 由上面的分析, 只需证明: 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. 设 A_1, D_1 分别是 A 及 D 的前 $n-1$ 行、前 $n-1$ 列构成的矩阵, 则 A_1, D_1 都属于 $Z^{(n-1) \times (n-1)}$, 且 $A_1 \leq D_1$.

由于矩阵

$$\tilde{A} = \begin{pmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & a_{nn} \end{pmatrix} \in Z^{n \times n}$$

满足 $A \leq \tilde{A}$, 故由(2)知, \tilde{A} 的所有实特征值为正数. 因而 A_1 的所有实特征值亦为正数. 按归纳法假设, 即有 $\det(D_1) \geq \det(A_1) > 0$, 而由(1)知 $A^{-1} \geq D^{-1} \geq 0$, 于是

$$(A^{-1})_{n,n} \geq (D^{-1})_{n,n} \geq 0,$$

(这里 $(A^{-1})_{n,n}$ 表示 A^{-1} 的 (n, n) 元素) 即

$$(A^{-1})_{n,n} = \frac{\det(A_1)}{\det(A)} \geq \frac{\det(D_1)}{\det(D)} = (D^{-1})_{n,n} \geq 0,$$

因此 $|A| > 0, |D| > 0$, 并利用归纳假设得

$$\det(D) \geq \det(A) \det(D_1) / \det(A_1) \geq \det(A) > 0. \quad \text{证毕}$$

从上述定理的证明可见, 若 $A, D \in Z^{n \times n}$, 且 $A \leq D$, 则 A 为非奇异 M 矩阵蕴涵着 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, 若定理 7.4.1 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”, 则此定理的结论(1)~(3)仍然成立.

下面的定理是非奇异 M 矩阵的一个基本定理, 它提供了非奇异 M 矩阵的多种等价条件.

定理 7.4.2 设 $A \in Z^{n \times n}$, 则以下各命题彼此等价:

- (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) \mathbf{A} 的每一个实特征值为正数;

(7) 存在 \mathbf{A} 的一种分裂 $\mathbf{A} = \mathbf{P} - \mathbf{Q}$, 使得 $\mathbf{P}^{-1} \geq \mathbf{0}, \mathbf{Q} \geq \mathbf{0}$ 且 $\rho(\mathbf{P}^{-1}\mathbf{Q}) < 1$;

(8) \mathbf{A} 非奇异, 且 $\mathbf{A}^{-1} \geq \mathbf{0}$.

证明 (1) \Rightarrow (2). 由定理 7.4.1(3) 即得.

(2) \Rightarrow (3). 设 $\mathbf{A}^{(k)}$ 是 \mathbf{A} 的任一 k 阶主子矩阵, K 表示 $\mathbf{A}^{(k)}$ 在 \mathbf{A} 中的行(列)序数集, λ 是 $\mathbf{A}^{(k)}$ 的任一实特征值. 下面用反证法证明 $\lambda > 0$.

若不然, $\lambda \leq 0$, 定义矩阵 $\mathbf{B} = (b_{ij}) \in \mathbb{Z}^{n \times n}$ 如下:

$$b_{ij} = \begin{cases} a_{ii} - \lambda, & \text{当 } i = j \text{ 时;} \\ a_{ij}, & \text{当 } i \neq j \text{ 且 } i, j \in K \text{ 时;} \\ 0, & \text{当 } i \neq j \text{ 且 } i, j \notin K \text{ 时,} \end{cases}$$

则 $\mathbf{B} \geq \mathbf{A}$, 并且由(2)知矩阵 \mathbf{B} 非奇异. 另一方面, 记 $\mathbf{B}^{(k)} = \mathbf{A}^{(k)} - \lambda \mathbf{I}$, 因为 λ 是 $\mathbf{A}^{(k)}$ 的特征值, 则 $\det(\mathbf{B}^{(k)}) = 0$, 从而 $\det(\mathbf{B}) = \det(\mathbf{B}^{(k)}) \prod_{i \notin K} b_{ii} = 0$. 这与 \mathbf{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 \mathbf{A} 的任一主子矩阵的实特征值均为正数, 故 \mathbf{A} 的任一主子式均为正数.

(4) \Rightarrow (5). 显然成立.

(5) \Rightarrow (6). 由定理 1.1.25 得

$$\det(\mathbf{A} - \lambda \mathbf{I}) = (-\lambda)^n + b_1(-\lambda)^{n-1} + \cdots + b_n, \quad (7.4.3)$$

其中 b_k 是 \mathbf{A} 的所有 k 阶主子式之和. 由(5)知 $b_k > 0 (k=1, \dots, n)$, 因此式(7.4.2)不可能有非正的实根, 即 \mathbf{A} 的所有实特征值均为正数.

(6) \Rightarrow (1). 设 $\mathbf{A} = s\mathbf{I} - \mathbf{B}, s > 0$ 且 $\mathbf{B} \geq \mathbf{0}$, 则 $s - \rho(\mathbf{B})$ 为 \mathbf{A} 的实特征值, 由(6)知它是正数, 即 $s > \rho(\mathbf{B})$. 因此 \mathbf{A} 为非奇异 M 矩阵.

(1) \Rightarrow (7). 取 $\mathbf{P} = s\mathbf{I}, \mathbf{Q} = \mathbf{B}$, 并且 s, \mathbf{B} 满足

$$\mathbf{A} = s\mathbf{I} - \mathbf{B}, \quad s > \rho(\mathbf{B}), \quad \mathbf{B} \geq \mathbf{0},$$

则 $\mathbf{P}^{-1} \geq \mathbf{0}, \mathbf{Q} \geq \mathbf{0}$, 并且 $\rho(\mathbf{P}^{-1}\mathbf{Q}) = \rho\left(\frac{1}{s}\mathbf{B}\right) = \frac{1}{s}\rho(\mathbf{B}) < 1$.

(7) \Rightarrow (8). 由(7)得 $\mathbf{A} = \mathbf{P}(\mathbf{I} - \mathbf{C})$, 其中 $\mathbf{C} = \mathbf{P}^{-1}\mathbf{Q}$. 因为 $\rho(\mathbf{C}) < 1$, 则由定理 5.2.5 有

$$\mathbf{A}^{-1} = (\mathbf{I} - \mathbf{C})^{-1}\mathbf{P}^{-1} = (\mathbf{I} + \mathbf{C} + \mathbf{C}^2 + \cdots)\mathbf{P}^{-1},$$

所以从 $\mathbf{C} = \mathbf{P}^{-1}\mathbf{Q} \geq \mathbf{0}$, 得 $\mathbf{A}^{-1} \geq \mathbf{0}$.

(8) \Rightarrow (1). 记 $\mathbf{A}^{-1} = \mathbf{G} = (g_{ij})$, 由(8)知 $\mathbf{G} \geq \mathbf{0}$. 由于 $\mathbf{A}\mathbf{G} = \mathbf{I}$, 故 $\sum_{j=1}^n a_{ij}g_{ji} = 1 (i=1, 2, \dots, n)$.

又因为 $a_{ij} \leq 0, g_{ji} \geq 0 (i \neq j)$, 则

$$a_{ii}g_{ii} = 1 - \sum_{j \neq i} a_{ij}g_{ji} \geq 1, \quad i = 1, 2, \dots, n.$$

由 $g_{ii} \geq 0$ 及上式得

$$a_{ii} > 0, \quad g_{ii} > 0, \quad i = 1, 2, \dots, n.$$

令 $s \geq \max_{1 \leq i \leq n} |a_{ii}|$, 则 $\mathbf{B} = s\mathbf{I} - \mathbf{A} \geq \mathbf{0}$. 由定理 7.1.11 知, $\rho(\mathbf{B})$ 是 \mathbf{B} 的特征值, 并且有相应的非负特征向量 $\mathbf{x} \geq \mathbf{0}$. 于是从 $\mathbf{B}\mathbf{x} = \rho(\mathbf{B})\mathbf{x}$ 得

$$\mathbf{A}\mathbf{x} = (s - \rho(\mathbf{B}))\mathbf{x}.$$

由于 \mathbf{A} 可逆, 所以 $s \neq \rho(\mathbf{B})$. 从而

$$\mathbf{A}^{-1}\mathbf{x} = \frac{1}{s - \rho(\mathbf{B})}\mathbf{x}.$$

因为 $\mathbf{A}^{-1} \geq 0, \mathbf{x} \geq 0$ 且 $\mathbf{x} \neq 0$, 所以 $s > \rho(\mathbf{B})$. 故 \mathbf{A} 是非奇异 M 矩阵.

证毕

例 7.4.2 设

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

显然, $\mathbf{A} \in Z^{3 \times 3}$, 又因为 $\det \mathbf{A} \neq 0$, \mathbf{A} 是非奇异的, 而且

$$\mathbf{A}^{-1} = \frac{1}{4} \begin{pmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{pmatrix} > 0,$$

所以 \mathbf{A} 符合定理 7.4.2(8), 故 \mathbf{A} 是非奇异 M 矩阵.

同时, \mathbf{A} 可表示为 $\mathbf{A} = s\mathbf{I} - \mathbf{B}, s = 2$,

$$\mathbf{B} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

\mathbf{B} 的特征值是 $\sqrt{2}, 0, -\sqrt{2}$; $\rho(\mathbf{B}) = \sqrt{2} < 2 = s$. 则 \mathbf{A} 符合定义 7.4.1, 故是非奇异 M 矩阵.

由此可见, 非奇异 M 矩阵等价的定义各有所长, 可根据情况适当选择.

下面的例子可以看到, 非奇异 M 矩阵之和不一定是非奇异 M 矩阵.

例 7.4.3 设

$$\mathbf{A} = \begin{bmatrix} 0.5 & -1 \\ 0 & 0.5 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0.5 & 0 \\ -1 & 0.5 \end{bmatrix}.$$

容易验证 \mathbf{A} 和 \mathbf{B} 都是 M 矩阵, 但 $\mathbf{A} + \mathbf{B}$ 是奇异的.

定理 7.4.3 设 $\mathbf{A} \in Z^{n \times n}$ 是对称的, 则 \mathbf{A} 为非奇异 M 矩阵的充分必要条件是 \mathbf{A} 为正定矩阵.

证明 根据定理 7.4.2(4), $\mathbf{A} \in Z^{n \times n}$ 为非奇异 M 矩阵等价于 \mathbf{A} 的所有主子式为正数; 而在 \mathbf{A} 是实对称的条件下, 由定理 1.3.8, \mathbf{A} 的所有主子式为正数等价于 \mathbf{A} 是正定的.

证毕

定理 7.4.4 设 $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ 是非奇异 M 矩阵, 则 \mathbf{AB} 为非奇异 M 矩阵的充分必要条件是 $\mathbf{AB} \in Z^{n \times n}$.

证明很容易(应用定理 7.4.2(8)), 留作练习.

以上讨论了非奇异 M 矩阵的一些基本性质, 但一般的 M 矩阵与非奇异的 M 矩阵在应用中几乎同等重要. 但由于一般的 M 矩阵(尤其是奇异的 M 矩阵)研究的难度大, 故其理论比起非奇异 M 矩阵来要弱一些. 为此, 我们将不加证明地介绍一般 M 矩阵的一些特性.

定理 7.4.5 设 $\mathbf{A} \in Z^{n \times n}$, 则以下各命题等价.

- (1) \mathbf{A} 是 M 矩阵;
- (2) 对每个 $\epsilon > 0, \mathbf{A} + \epsilon\mathbf{I}$ 是非奇异 M 矩阵;
- (3) \mathbf{A} 的任意主子矩阵的每个实特征值非负;

- (4) \mathbf{A} 的所有主子式非负;
 (5) 对每个 $k=1, 2, \dots, n$, \mathbf{A} 的所有 k 阶主子式之和为非负实数;
 (6) \mathbf{A} 的每个实特征值非负.

对于 \mathbf{A} 是不可约的奇异 M 矩阵也有如下定理.

定理 7.4.6 设 \mathbf{A} 为不可约的奇异 M 矩阵, 则

- (1) $\text{rank} \mathbf{A} = n - 1$;
 (2) 存在正向量 $\mathbf{x} > 0$, 使得 $\mathbf{A}\mathbf{x} = \mathbf{0}$;
 (3) \mathbf{A} 的所有真主子矩阵为非奇异的 M 矩阵, 特别有 $a_{ii} > 0$ ($1 \leq i \leq n$);
 (4) 对任意 $\mathbf{x} \in \mathbb{R}^n$, 若 $\mathbf{A}\mathbf{x} \geq 0$, 则 $\mathbf{A}\mathbf{x} = \mathbf{0}$.

7.4.2 H 矩阵

下面将 n 阶方阵 \mathbf{A} 推广到复矩阵, 且利用 \mathbf{A} 中的元素取模构造出一个新的比较矩阵, 记为 $H(\mathbf{A})$, 如果 $H(\mathbf{A})$ 是非奇异的 M 矩阵, 则定义 \mathbf{A} 为 H 矩阵.

定义 7.4.2 设 $\mathbf{A} = (a_{ij}) \in \mathbb{C}^{n \times n}$, 并设

$$H(\mathbf{A}) = (m_{ij}) \in \mathbb{R}^{n \times n},$$

其中

$$m_{ij} = \begin{cases} |a_{ij}|, & j = i, \\ -|a_{ij}|, & j \neq i, \end{cases} \quad i, j = 1, \dots, n, \quad (7.4.4)$$

$H(\mathbf{A})$ 称为 \mathbf{A} 的比较矩阵.

定义 7.4.3 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 如果 \mathbf{A} 的比较矩阵 $H(\mathbf{A})$ 是非奇异的 M 矩阵, 则称 \mathbf{A} 为非奇异 H 矩阵, 简称 H 矩阵.

下面简要给出 H 矩阵的一些性质.

定理 7.4.7 设 $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n \times n}$, \mathbf{A} 是非奇异 M 矩阵, $H(\mathbf{B}) \geq \mathbf{A}$, 则

- (1) \mathbf{B} 是 H 矩阵;
 (2) \mathbf{B} 是非奇异的, 且 $\mathbf{A}^{-1} \geq |\mathbf{B}^{-1}| \geq 0$;
 (3) $|\det \mathbf{B}| \geq \det \mathbf{A} > 0$.

证明 这一定理是将 M 矩阵中定理 7.4.1 的基本结果推广至复矩阵 \mathbf{B} .

(1) 从定理 7.4.1(2) 知 $H(\mathbf{B})$ 的每个实特征值为正数, 再由定理 7.4.2 的(6) 知 $H(\mathbf{B})$ 是非奇异的 M 矩阵. 从而由定义 7.4.3 推知 \mathbf{B} 为 H 矩阵.

(2) 现在取对角酉矩阵

$$\mathbf{D} = \text{diag}(\bar{b}_{11}/|b_{11}|, \dots, \bar{b}_{nn}/|b_{nn}|),$$

则 \mathbf{DB} 的主对角元素是正数

$$|b_{11}|, \dots, |b_{nn}|,$$

从非奇异 M 矩阵的定义知, 可将 \mathbf{A} 表示为

$$\mathbf{A} = s\mathbf{I} - \mathbf{P}, \quad \mathbf{P} \geq 0, \quad s > \rho(\mathbf{P}),$$

并令

$$\mathbf{R} \equiv s\mathbf{I} - \mathbf{DB}.$$

由于

$$|\mathbf{R}| = \begin{vmatrix} |s - |b_{11}|| & |b_{12}| & \cdots & |b_{1n}| \\ |b_{21}| & |s - |b_{22}|| & \cdots & |b_{2n}| \\ \vdots & \vdots & & \vdots \\ |b_{n1}| & |b_{n2}| & \cdots & |s - |b_{nn}|| \end{vmatrix} \\ \leq \mathbf{P} + \mathbf{A} - \mathbf{H}(\mathbf{B}) \leq \mathbf{P}.$$

由定理 7.1.3 知

$$\rho(\mathbf{R}) \leq \rho(|\mathbf{R}|) \leq \rho(\mathbf{P}) < s,$$

因此 $\mathbf{DB} = s\mathbf{I} - \mathbf{R}$ 可逆, 故 \mathbf{B} 必非奇异, 而且

$$|\mathbf{B}^{-1}| = |(\mathbf{DB})^{-1}| = |s^{-1}(\mathbf{I} - s^{-1}\mathbf{R})^{-1}| \\ = \left| \sum_{k=0}^{\infty} s^{-k-1} \mathbf{R}^k \right| \leq \sum_{k=0}^{\infty} \frac{1}{s^{k+1}} \mathbf{P}^k = \mathbf{A}^{-1}.$$

(3) 仿照定理 7.4.1(3)的推导, 可得

$$\frac{\det(\mathbf{A}_{11})}{\det(\mathbf{A})} = (\mathbf{A}^{-1})_{nn} \geq |(\mathbf{B}^{-1})_{nn}| = \left| \frac{\det(\mathbf{B}_{11})}{\det(\mathbf{B})} \right|,$$

从而 $|\det(\mathbf{B})| \geq \det(\mathbf{A})$.

证毕

定理 7.4.8 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 则有如下明显性质:

- (1) $\mathbf{H}(\mathbf{A}) \in \mathbb{Z}^{n \times n}$;
- (2) $\mathbf{H}(\mathbf{A}) = \mathbf{A}$ 的充分必要条件是 $\mathbf{A} \in \mathbb{Z}^{n \times n}$;
- (3) $\mathbf{H}(\mathbf{A})$ 可表示为非负对角矩阵与具有零对角的非负矩阵之差:

$$\mathbf{H}(\mathbf{A}) = |\text{diag}(a_{11}, \cdots, a_{nn})| - [|\mathbf{A}| - |\text{diag}(a_{11}, \cdots, a_{nn})|], \quad (7.4.5)$$

这里 $|\mathbf{X}| \equiv [|x_{ij}|]$ 表示矩阵 $\mathbf{X} = (x_{ij}) \in \mathbb{C}^{n \times n}$ 的逐个元素取绝对值后的矩阵;

- (4) \mathbf{A} 为 M 矩阵的充分必要条件是 $\mathbf{H}(\mathbf{A}) = \mathbf{A}$, 且 \mathbf{A} 为 H 矩阵;
- (5) 若 \mathbf{A} 是 M 矩阵, 则式(7.4.5)成为

$$\mathbf{A} = \text{diag}(a_{11}, \cdots, a_{nn}) - [\text{diag}(a_{11}, \cdots, a_{nn}) - \mathbf{A}], \quad (7.4.6)$$

此表示式可供替代定义中的表示式 $\mathbf{A} = s\mathbf{I} - \mathbf{B}$.

证明留作练习.

7.5 T 矩阵与汉克尔矩阵

本节将主要介绍 T 矩阵(特普利茨矩阵)、汉克尔(Hankel)矩阵及其一些性质.

在数据处理、有限单元法、概率统计以及滤波理论等广泛的科学技术领域里, 常常遇到如下 n 阶矩阵

$$\mathbf{A} = \begin{pmatrix} a_0 & a_{-1} & a_{-2} & \cdots & a_{-n+1} \\ a_1 & a_0 & a_{-1} & \cdots & a_{-n+2} \\ a_2 & a_1 & a_0 & \cdots & a_{-n+3} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n-2} & a_{n-3} & a_{n-4} & \cdots & a_{-1} \\ a_{n-1} & a_{n-2} & a_{n-3} & \cdots & a_0 \end{pmatrix}, \quad (7.5.1)$$

其中位于任一条平行于主对角线的直线上的元素全相同. 这样的矩阵称为特普利茨矩阵, 简称 **T** 矩阵.

T 矩阵式(7.5.1)也可简记为

$$\mathbf{A} = (a_{i-j})_{i,j=1}^n.$$

20 世纪 60 年代以来, 有关 T 矩阵的快速算法已有相当发展. 60 年代中期, 特伦奇 (Trench) 提出了 T 矩阵的快速求逆算法. 几年后, 左哈 (Zohar) 进一步讨论了特伦奇的算法, 把对称正定条件减弱为强非奇异 (指所有顺序主子矩阵非奇异), 这样就把通常求逆的计算量 (或计算复杂性) 从 $O(n^3)$ 级减少为 $O(n^2)$ 级. 60 年代末还出现了以 T 矩阵为系数矩阵的线性方程组快速数值解法 (不通过求逆). 至于一般 T 矩阵特征值的快速算法还较少见.

T 矩阵的性质不易探讨, 因此人们很早就把兴趣集中到了与 T 矩阵联系密切的矩阵或特殊的 T 矩阵的研究上. 例如, 具有以下形式的 $n+1$ 阶矩阵:

$$\mathbf{H}_{n+1} = \begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_n \\ a_1 & a_2 & a_3 & \cdots & a_{n+1} \\ a_2 & a_3 & a_4 & \cdots & a_{n+2} \\ \vdots & \vdots & \vdots & & \vdots \\ a_n & a_{n+1} & a_{n+2} & \cdots & a_{2n} \end{pmatrix} = (a_{i+j})_{i,j=0}^n, \quad (7.5.2)$$

其中沿着所有平行于副对角线的直线上有相同的元素. 这样的矩阵称为汉克尔矩阵.

汉克尔矩阵在用最小二乘法求数据的多项式拟合曲线问题中, 有着广泛的应用.

设 $(x_i, y_i) (i=1, \dots, m)$ 是一组观测数据, 其中节点 x_i 互异, 寻找一个 n 次非零多项式

$$f(x) = \mu_0 + \mu_1 x + \cdots + \mu_n x^n, \quad n < m,$$

使得

$$S(\mu_0, \mu_1, \dots, \mu_n) = \sum_{i=1}^m [y_i - f(x_i)]^2 = \sum_{i=1}^m (y_i - \sum_{j=0}^n \mu_j x_i^j)^2$$

达到最小. 由高等数学多元函数求极值问题可知, $(\mu_0, \mu_1, \dots, \mu_n)$ 是极值点的必要条件为

$$\frac{\partial S}{\partial \mu_k} = -2 \sum_{i=1}^m x_i^k (y_i - \sum_{j=0}^n \mu_j x_i^j) = 0, \quad k = 0, 1, \dots, n. \quad (7.5.3)$$

记

$$a_k = \sum_{i=1}^m x_i^k, \quad \beta_k = \sum_{i=1}^m x_i^k y_i,$$

则由式(7.5.3)得

$$\sum_{j=0}^n a_{k+j} \mu_j = \beta_k, \quad k = 0, 1, \dots, n, \quad (7.5.4)$$

写成矩阵形式即为

$$\mathbf{H}_{n+1} \mathbf{u} = \mathbf{b}, \quad (7.5.5)$$

其中 \mathbf{H}_{n+1} 是形如式(7.5.2)的汉克尔矩阵, $\mathbf{u} = (\mu_0, \mu_1, \dots, \mu_n)^\top$, $\mathbf{b} = (\beta_0, \beta_1, \dots, \beta_n)^\top$. 可见问题转化为求解以汉克尔矩阵为系数矩阵的线性方程组问题.

汉克尔矩阵有如下的性质.

定理 7.5.1 汉克尔矩阵 \mathbf{H}_{n+1} 是非奇异的.

证明 利用反证法,若 $\det \mathbf{H}_{n+1} = 0$, 则式(7.5.4)的齐次方程组

$$\sum_{j=0}^n a_{k+j} \mu_j = 0, \quad k = 0, 1, \dots, n$$

有非零解. 将上面方程组中第 k 个方程乘以 μ_k , 然后对所有 k 求和, 使得

$$\begin{aligned} 0 &= \sum_{k=0}^n \mu_k \left(\sum_{j=0}^n a_{k+j} \mu_j \right) = \sum_{k=0}^n \mu_k \sum_{j=0}^n \mu_j \sum_{i=1}^m x_i^{k+j} \\ &= \sum_{i=1}^m \left(\sum_{k=0}^n \mu_k x_i^k \right) \left(\sum_{j=0}^n \mu_j x_i^j \right) \\ &= \sum_{i=1}^m \left(\sum_{j=0}^n \mu_j x_i^j \right)^2 = \sum_{i=1}^m f^2(x_i), \end{aligned}$$

据此就有

$$f(x_1) = f(x_2) = \dots = f(x_m) = 0.$$

又因 $m > n$, 故 $f(x) \equiv 0$, 与假设 $f(x) \neq 0$ 矛盾.

证毕

可以直接验证, T 矩阵与汉克尔矩阵是可以互相转化的. 事实上, 设 T 矩阵为 \mathbf{A} , 汉克尔矩阵为 \mathbf{H}_{n+1} , 则用矩阵

$$\mathbf{J} = \begin{bmatrix} & & & 1 \\ & & \cdot & \\ & & \cdot & \\ 1 & & & \end{bmatrix} \quad (7.5.6)$$

乘矩阵 \mathbf{H}_{n+1} , 其结果 \mathbf{JH}_{n+1} 或 $\mathbf{H}_{n+1}\mathbf{J}$ 都是 T 矩阵, 且有

$$(\mathbf{JH}_{n+1})^T = \mathbf{H}_{n+1}\mathbf{J}. \quad (7.5.7)$$

反之, 用 \mathbf{J} 乘 T 矩阵 \mathbf{A} , 则 \mathbf{JA} 或 \mathbf{AJ} 都是汉克尔矩阵.

从而, 可把对 T 矩阵的研究转化为对汉克尔矩阵的研究.

习 题 7 (1)

1. 设 n 阶矩阵 $\mathbf{A} = (a_{ij})$ 为非负矩阵, 且是非奇异的, 问逆矩阵 $\mathbf{A}^{-1} = \mathbf{B} = (b_{ij})$ 满足

$$b_{ij} \geq 0, \quad i \neq j, \quad \mathbf{B} \geq \mathbf{0}$$

的条件是什么?

2. 设 \mathbf{A} 为 n 阶实矩阵, \mathbf{x}, \mathbf{y} 是 n 维向量. 证明:

(1) 如果 $\mathbf{A} > \mathbf{0}$, $\mathbf{x} \geq \mathbf{0}$, 且 $\mathbf{x} \neq \mathbf{0}$, 则 $\mathbf{Ax} > \mathbf{0}$;

(2) 如果 $\mathbf{A} > \mathbf{0}$, $\mathbf{x} \geq \mathbf{y}$, 则 $\mathbf{Ax} \geq \mathbf{Ay}$;

(3) 如果对所有 $\mathbf{x} \geq \mathbf{0}$ 都有 $\mathbf{Ax} \geq \mathbf{0}$, 则 $\mathbf{A} \geq \mathbf{0}$.

3. 证明: 设 $\mathbf{A} \in \mathbb{R}^{n \times n}$, 若 $\mathbf{A} > \mathbf{0}$, 则 \mathbf{A} 是素矩阵, 但反之未必.

4. 设 \mathbf{A} 是 n 阶非负素矩阵, 证明 $\rho(\mathbf{A}) > 0$.

5. 设 \mathbf{A} 是 n 阶不可约非负对称矩阵, 证明: \mathbf{A} 是素矩阵的充分必要条件是 $\mathbf{A} + \rho(\mathbf{A})\mathbf{I}$ 非奇异.

6. 设 $\mathbf{A} = \begin{bmatrix} 7 & 2 & 2 \\ 2 & 1 & 1 \\ 4 & 2 & 2 \end{bmatrix}$, 求 $\rho(\mathbf{A})$ 和 $\lim_{n \rightarrow \infty} [\rho(\mathbf{A})^{-1} \mathbf{A}]^n$.

7. 设 A 是 n 阶不可约非负矩阵, 证明: 如果 $a_{ii} > 0 (i=1, 2, \dots, n)$, 则 $A^{n-1} > 0$.
8. 设矩阵 A 的逆矩阵 A^{-1} 是单调矩阵, 问 A 是怎样的矩阵?
9. 设 $A \in \mathbb{R}^{n \times n}$ 是正随机矩阵, 证明: A 的谱半径 $\rho(A) = 1$.
10. 设矩阵 A 和它的逆矩阵 A^{-1} 都是 M 矩阵, 问 A 是怎样的矩阵?
11. 矩阵 $A = (a_{ij})_{n \times n}$ 为 M 矩阵的充分必要条件是: $a_{ii} > 0, a_{ij} \leq 0 (i \neq j)$ 以及 $\rho(B) < 1$, 这个结论是否正确? 这里 $B = I - D^{-1}A$, 而 $D = \text{diag}(a_{11}, \dots, a_{nn})$.
12. 设矩阵 A 和 B 都是 M 矩阵, 问 $A+B$ 是否为 M 矩阵?
13. 设 A 是 n 阶非奇异 M 矩阵, x 是 n 维列向量, 证明: 若 $Ax \geq 0$, 则 $x \geq 0$.
14. 证明: 设 $A \in \mathbb{R}^{n \times n}$ 是对称矩阵, 则 A 为非奇异 M 矩阵的充分必要条件是 A 为正定矩阵.
15. 证明: 设 $A, B \in \mathbb{R}^{n \times n}$ 是 M 矩阵, 则 AB 为 M 矩阵的充分必要条件是 $AB \in \mathbb{R}^{n \times n}$. 特别地, 若 $A, B \in \mathbb{R}^{2 \times 2}$ 是 M 矩阵, 则 AB 为 M 矩阵.

7.6 克罗内克积

前面定义过两个矩阵 A 和 B 的乘积 AB , 它要求 A 的列数必须等于 B 的行数. 下面引进一种新的乘法运算, 它对矩阵的行数和列数没有任何要求.

7.6.1 克罗内克积的概念

定义 7.6.1 设 $A = (a_{ij}) \in \mathbb{C}^{m \times n}, B = (b_{ij}) \in \mathbb{C}^{p \times q}$, 则称如下的分块矩阵

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix} \in \mathbb{C}^{mp \times nq}$$

为 A 的克罗内克(Kronecker)积, 或称 A 与 B 的直积, 或张量积, 简记为 $A \otimes B = (a_{ij}B)_{mp \times nq}$. 即 $A \otimes B$ 是一个 $m \times n$ 块的分块矩阵, 最后是一个 $mp \times nq$ 矩阵.

例 7.6.1 设 $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, B = \begin{bmatrix} x \\ y \end{bmatrix}$, 那么

$$A \otimes B = \begin{bmatrix} aB & bB \\ cB & dB \end{bmatrix} = \begin{bmatrix} ax & bx \\ ay & by \\ cx & dx \\ cy & dy \end{bmatrix}_{4 \times 2},$$

$$B \otimes A = \begin{bmatrix} xA \\ yA \end{bmatrix} = \begin{bmatrix} xa & xb \\ xc & xd \\ ya & yb \\ yc & yd \end{bmatrix}_{4 \times 2} = \begin{bmatrix} ax & bx \\ cx & dx \\ ay & by \\ cy & dy \end{bmatrix}_{4 \times 2}.$$

由这个例子可以看出, $A \otimes B$ 与 $B \otimes A$ 一般不是同一矩阵, 即克罗内克积不满足交换律, 但它们的阶数是相同的. 例如:

$$A = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \quad B = (1 \quad -1),$$

$$\text{则 } A \otimes B = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 1 & -1 \end{bmatrix}, \quad B \otimes A = \begin{bmatrix} 1 & 0 & -1 & 0 \\ -1 & 1 & 1 & -1 \end{bmatrix},$$

所以

$$\mathbf{A} \otimes \mathbf{B} \neq \mathbf{B} \otimes \mathbf{A}.$$

例 7.6.2 设 $\mathbf{A} = \text{diag}(a_1, a_2, \dots, a_m)$, $\mathbf{B} = \text{diag}(b_1, b_2, \dots, b_n)$, \mathbf{A}, \mathbf{B} 都是对角矩阵, 则显然 $\mathbf{A} \otimes \mathbf{B} = \text{diag}(a_1 b_1, a_1 b_2, \dots, a_1 b_n, a_2 b_1, a_2 b_2, \dots, a_2 b_n, \dots, a_m b_1, a_m b_2, \dots, a_m b_n)$ 也是对角矩阵, 而当 \mathbf{A} 和 \mathbf{B} 都是单位矩阵时, 有

$$\mathbf{I}_n \otimes \mathbf{I}_m = \mathbf{I}_m \otimes \mathbf{I}_n = \mathbf{I}_{mn}.$$

不难验证, 当 \mathbf{A} 和 \mathbf{B} 都是上(下)三角矩阵时, $\mathbf{A} \otimes \mathbf{B}$ 也是上(下)三角矩阵.

7.6.2 克罗内克积的性质

不难验证, 矩阵的克罗内克积满足下列运算律:

- (1) $k(\mathbf{A} \otimes \mathbf{B}) = k\mathbf{A} \otimes \mathbf{B} = \mathbf{A} \otimes k\mathbf{B}, k \in \mathbb{C}$;
- (2) 分配律 $(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes \mathbf{C} + \mathbf{B} \otimes \mathbf{C}$;
- (3) 结合律 $(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C})$.

下面我们来研究克罗内克积的另一个重要性质, 这条性质对进一步研究克罗内克积有着重要的作用.

定理 7.6.1 设 $\mathbf{A} = (a_{ij})_{m \times n}$, $\mathbf{B} = (b_{ij})_{s \times r}$, $\mathbf{C} = (c_{ij})_{n \times p}$, $\mathbf{D} = (d_{ij})_{r \times t}$, 则

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}. \quad (7.6.1)$$

证明 因为

$$\begin{aligned} (\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) &= (a_{ij} \mathbf{B})(c_{ij} \mathbf{D}) \\ &= \left(\sum_{k=1}^n a_{ik} c_{kj} \mathbf{BD} \right) = (\mathbf{AC})_{ij} \mathbf{BD} \\ &= \mathbf{AC} \otimes \mathbf{BD}, \end{aligned}$$

式中 $(\mathbf{AC})_{ij}$ 是矩阵 \mathbf{AC} 中第 i 行第 j 列的元素.

证毕

推论 若 $\mathbf{A} = (a_{ij})_{m \times m}$, $\mathbf{B} = (b_{ij})_{n \times n}$, 则

$$\mathbf{A} \otimes \mathbf{B} = (\mathbf{A} \otimes \mathbf{I}_n)(\mathbf{I}_m \otimes \mathbf{B}) = (\mathbf{I}_m \otimes \mathbf{B})(\mathbf{A} \otimes \mathbf{I}_n).$$

定理 7.6.2 设 $\mathbf{A} = (a_{ij})_{m \times n}$, 则 $r(\mathbf{A}) \leq 1 \Leftrightarrow \mathbf{A}$ 可以表示成一个行向量和一个列向量的克罗内克积.

证明 不妨 $\mathbf{A} \neq \mathbf{0}$ (否则自然成立), 设 $r(\mathbf{A}) = 1$, 则 \mathbf{A} 的每一列成比例. 任取 \mathbf{A} 的一个非零列 $\mathbf{A}^{(j)}$, 并设 $\mathbf{A}^{(k)} = a_k \mathbf{A}^{(j)}, k = 1, 2, \dots, m$. 则

$$\mathbf{A} = [a_1 \mathbf{A}^{(j)}, a_2 \mathbf{A}^{(j)}, \dots, a_m \mathbf{A}^{(j)}] = [a_1, a_2, \dots, a_m] \otimes \mathbf{A}^{(j)}$$

为一个行向量和一个列向量的克罗内克积; 反之, 若 \mathbf{A} 是一个行向量和一个列向量的克罗内克积, 则显然 \mathbf{A} 的每一列成比例, 因此 $r(\mathbf{A}) \leq 1$.

证毕

例 7.6.3 可以利用定理 7.6.1 和定理 7.6.2 来计算两个秩为 1 的矩阵的乘积. 例如

$$\begin{bmatrix} 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -2 & 1 & 1 & -1 & 3 \\ 2 & -1 & -1 & 1 & -3 \\ -2 & 1 & 1 & -1 & 3 \\ -2 & 1 & 1 & -1 & 3 \end{bmatrix}$$

$$\begin{aligned}
&= \left[(1, 1, -1, -1) \otimes \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} \right] \otimes \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix} \otimes (-2, 1, 1, -1, 3) \\
&= (1, 1, -1, -1) \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} (-2, 1, 1, -1, 3) \\
&= (-2) \otimes \begin{pmatrix} -2 & 1 & 1 & -1 & 3 \\ 2 & -1 & -1 & 1 & -3 \\ -2 & 1 & 1 & -1 & 3 \\ 2 & -1 & -1 & 1 & -3 \end{pmatrix} = \begin{bmatrix} 4 & -2 & -2 & 2 & -6 \\ -4 & 2 & 2 & -2 & 6 \\ 4 & -2 & -2 & 2 & -6 \\ -4 & 2 & 2 & -2 & 6 \end{bmatrix}.
\end{aligned}$$

下面的例子可心解释克罗内克积 $A \otimes B$ 与 $B \otimes A$ 的几何意义.

例 7.6.4 设 $V = \mathbb{R}^m, W = \mathbb{R}^n$ 是两个(向量)线性空间, 则集合

$$S = \left\{ \sum_{i=1}^k \mathbf{v}_i \otimes \mathbf{w}_i \mid \mathbf{v}_i \in V, \mathbf{w}_i \in W, k \geq 0 \right\}$$

也构成一个线性空间, 称为 V 与 W 的张量积空间, 记为 $V \otimes W$. 显然 $V \otimes W$ 同构于 \mathbb{R}^{mn} , 并且, 如果 $\alpha_1, \alpha_2, \dots, \alpha_m$ 和 $\beta_1, \beta_2, \dots, \beta_n$ 分别为 V 和 W 的基, 则

$$\alpha_1 \otimes \beta_1, \alpha_1 \otimes \beta_2, \dots, \alpha_1 \otimes \beta_n, \alpha_2 \otimes \beta_1, \alpha_2 \otimes \beta_2, \dots, \alpha_2 \otimes \beta_n, \dots, \alpha_m \otimes \beta_1, \alpha_m \otimes \beta_2, \dots, \alpha_m \otimes \beta_n \quad (7.6.2)$$

是 $V \otimes W$ 的一组基. 又设 \mathcal{A} 是 V 中某个线性变换, \mathcal{B} 是 W 中某个线性变换, 则可以定义张量积空间 $V \otimes W$ 中的一个线性变换 \mathcal{T} : $\mathcal{T}(\mathbf{v} \otimes \mathbf{w}) = \mathcal{A}(\mathbf{v}) \otimes \mathcal{B}(\mathbf{w}), \mathbf{v} \in V, \mathbf{w} \in W$, 称 \mathcal{T} 为 \mathcal{A} 与 \mathcal{B} 的张量积, 记为 $\mathcal{A} \otimes \mathcal{B}$. 如果 \mathcal{A} 和 \mathcal{B} 在基 $\alpha_1, \alpha_2, \dots, \alpha_m$ 和 $\beta_1, \beta_2, \dots, \beta_n$ 下的矩阵为 A 和 B , 则 $\mathcal{A} \otimes \mathcal{B}$ 在基(7.6.2)下的矩阵为 $A \otimes B$, 而 $\mathcal{A} \otimes \mathcal{B}$ 在基

$$\alpha_1 \otimes \beta_1, \alpha_2 \otimes \beta_1, \dots, \alpha_m \otimes \beta_1, \alpha_1 \otimes \beta_2, \alpha_2 \otimes \beta_2, \dots, \alpha_m \otimes \beta_2, \dots, \alpha_1 \otimes \beta_n, \alpha_2 \otimes \beta_n, \dots, \alpha_m \otimes \beta_n \quad (7.6.3)$$

下的矩阵则为 $B \otimes A$.

定理 7.6.3 设 $A = (a_{ij})_{m \times n}, B = (b_{ij})_{p \times q}$, 则

$$(A \otimes B)^T = A^T \otimes B^T, \quad (7.6.4)$$

$$(A \otimes B)^H = A^H \otimes B^H. \quad (7.6.5)$$

证明 因为

$$\begin{aligned}
(A \otimes B)^T &= (a_{ij} B)^T = \begin{pmatrix} a_{11} B & \cdots & a_{1n} B \\ \vdots & & \vdots \\ a_{m1} B & \cdots & a_{mn} B \end{pmatrix}^T \\
&= \begin{pmatrix} a_{11} B^T & \cdots & a_{m1} B^T \\ \vdots & & \vdots \\ a_{1n} B^T & \cdots & a_{mn} B^T \end{pmatrix} = A^T \otimes B^T.
\end{aligned}$$

同理可证 $(A \otimes B)^H = A^H \otimes B^H$.

证毕

推论 若 A, B 是对称(埃尔米特或酉)矩阵, 则 $A \otimes B$ 也是对称(埃尔米特或酉)矩阵.

定理 7.6.5, 有

$$r(\mathbf{A} \otimes \mathbf{B}) = r(\mathbf{A})r(\mathbf{B}) = nq. \quad (7.6.9)$$

但由克罗内克积定义得

$$\mathbf{A} \otimes \mathbf{B} = [\mathbf{x}_1 \otimes \mathbf{y}_1, \dots, \mathbf{x}_1 \otimes \mathbf{y}_q, \dots, \mathbf{x}_m \otimes \mathbf{y}_1, \dots, \mathbf{x}_m \otimes \mathbf{y}_q]. \quad (7.6.10)$$

因此, 上两式说明 $\mathbf{A} \otimes \mathbf{B}$ 是列满秩的, 即 $\mathbf{x}_i \otimes \mathbf{y}_j$ ($i=1, \dots, m; j=1, \dots, q$) 是线性无关的.

反之, 设向量组 $\mathbf{x}_i \otimes \mathbf{y}_j$ ($i=1, \dots, m; j=1, \dots, q$) 是线性无关的, 则式(7.6.10)说明 $\mathbf{A} \otimes \mathbf{B}$ 是列满秩的, 即式(7.6.9)成立. 如果 $r(\mathbf{A}) < n$, 则必有 $r(\mathbf{B}) > q$, 这是不可能的, 因为 \mathbf{B} 的列数等于 q , 故 $r(\mathbf{A}) = n$. 同理, $r(\mathbf{B}) = q$, 即 \mathbf{A}, \mathbf{B} 都是列满秩的, 因此 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ 和 $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_q$ 都是线性无关的. 证毕

定理 7.6.8 设 \mathbf{A}, \mathbf{B} 分别为 m, p 阶方阵, 则有

$$|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^p |\mathbf{B}|^m. \quad (7.6.11)$$

证明 设 \mathbf{A}, \mathbf{B} 的若尔当标准型分别为 $\mathbf{J}_1, \mathbf{J}_2$, 则存在可逆矩阵 \mathbf{M}, \mathbf{P} , 使 $\mathbf{M}^{-1}\mathbf{A}\mathbf{M} = \mathbf{J}_1$, $\mathbf{P}^{-1}\mathbf{B}\mathbf{P} = \mathbf{J}_2$. 由定理 7.6.1 和定理 7.6.4, 有

$$(\mathbf{M} \otimes \mathbf{P})^{-1}(\mathbf{A} \otimes \mathbf{B})(\mathbf{M} \otimes \mathbf{P}) = (\mathbf{M}^{-1} \otimes \mathbf{P}^{-1})(\mathbf{A} \otimes \mathbf{B})(\mathbf{M} \otimes \mathbf{P}) = \mathbf{J}_1 \otimes \mathbf{J}_2,$$

因此, $|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{J}_1 \otimes \mathbf{J}_2|$. 但由于 $\mathbf{J}_1, \mathbf{J}_2$ 都是上三角矩阵, $\mathbf{J}_1 \otimes \mathbf{J}_2$ 也是上三角矩阵, 且对角线元素为

$$\lambda_1 \mu_1, \lambda_1 \mu_2, \dots, \lambda_1 \mu_p, \lambda_2 \mu_1, \lambda_2 \mu_2, \dots, \lambda_2 \mu_p, \dots, \lambda_m \mu_1, \lambda_m \mu_2, \dots, \lambda_m \mu_p,$$

其中 λ_i, μ_j 分别为 \mathbf{A}, \mathbf{B} 的特征值. 因此

$$|\mathbf{J}_1 \otimes \mathbf{J}_2| = \left(\prod_{i=1}^m \lambda_i \right)^p \left(\prod_{j=1}^p \mu_j \right)^m = |\mathbf{A}|^p |\mathbf{B}|^m. \quad \text{证毕}$$

我们定义克罗内克积的幂, 记 $\mathbf{A}^{[k]} = \underbrace{\mathbf{A} \otimes \mathbf{A} \otimes \dots \otimes \mathbf{A}}_k, k > 0$.

定理 7.6.9 设 $\mathbf{A} = (a_{ij})_{m \times p}, \mathbf{B} = (b_{ij})_{p \times n}$, 则有

$$(\mathbf{A}\mathbf{B})^{[k]} = \mathbf{A}^{[k]} \mathbf{B}^{[k]} \quad (7.6.12)$$

证明 利用归纳法. 当 $k=1$ 时显然成立. 设 $k-1$ 时成立, 则

$$\begin{aligned} (\mathbf{A}\mathbf{B})^{[k]} &= \mathbf{A}\mathbf{B} \otimes (\mathbf{A}\mathbf{B})^{[k-1]} = \mathbf{A}\mathbf{B} \otimes \mathbf{A}^{[k-1]} \mathbf{B}^{[k-1]} \\ &= (\mathbf{A} \otimes \mathbf{A}^{[k-1]}) (\mathbf{B} \otimes \mathbf{B}^{[k-1]}) = \mathbf{A}^{[k]} \mathbf{B}^{[k]}. \end{aligned} \quad \text{证毕}$$

下面讨论矩阵 \mathbf{A}, \mathbf{B} 的特征值与 $\mathbf{A} \otimes \mathbf{B}$ 的特征值的关系.

定理 7.6.10 设 $\lambda_1, \lambda_2, \dots, \lambda_m$ 是 $\mathbf{A}_{m \times m}$ 的 m 个特征值, $\mu_1, \mu_2, \dots, \mu_p$ 是 $\mathbf{B}_{p \times p}$ 的 p 个特征值, 那么 $\mathbf{A} \otimes \mathbf{B}$ 的 mp 个特征值为 $\lambda_i \mu_j$ ($i=1, 2, \dots, m; j=1, 2, \dots, p$).

证明 由 2.2 节知, \mathbf{A} 与 \mathbf{B} 一定与若尔当标准形相似, 即存在可逆矩阵 \mathbf{P} 与 \mathbf{Q} , 使得

$$\begin{aligned} \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{J}_1 &= \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ & & \lambda_m \end{pmatrix}, \\ \mathbf{Q}^{-1}\mathbf{B}\mathbf{Q} = \mathbf{J}_2 &= \begin{pmatrix} \mu_1 & & * \\ & \ddots & \\ & & \mu_p \end{pmatrix}, \end{aligned}$$

即有

$$A = P \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ & & \lambda_m \end{pmatrix} P^{-1},$$

$$B = Q \begin{pmatrix} \mu_1 & & * \\ & \ddots & \\ & & \mu_p \end{pmatrix} Q^{-1}.$$

从而由式(7.6.1)有

$$A \otimes B = (P \otimes Q) \left[\begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ & & \lambda_m \end{pmatrix} \otimes \begin{pmatrix} \mu_1 & & * \\ & \ddots & \\ & & \mu_p \end{pmatrix} \right] (P^{-1} \otimes Q^{-1})$$

$$= (P \otimes Q) \begin{pmatrix} \begin{pmatrix} \mu_1 & & * \\ & \ddots & \\ & & \mu_p \end{pmatrix} & & * \\ \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{pmatrix} \begin{pmatrix} \mu_1 & & * \\ & \ddots & \\ & & \mu_p \end{pmatrix} \end{pmatrix} (P \otimes Q)^{-1},$$

即有

$$A \otimes B \sim \begin{pmatrix} \lambda_1 \mu_1 & & & & * \\ & \ddots & & & \\ & & \lambda_1 \mu_p & & \\ & & & \ddots & \\ & & & & \lambda_m \mu_1 \\ & & & & & \ddots \\ & & & & & & \lambda_m \mu_p \end{pmatrix}.$$

从而 $A \otimes B$ 的 mp 个特征值为 $\lambda_i \mu_j$ ($i=1, \dots, m; j=1, \dots, p$).

证毕

由例 7.6.1 已看到, 克罗内克积的交换律不成立, 即 $A \otimes B$ 一般不等于 $B \otimes A$. 但是, 我们仍有下面的性质.

定理 7.6.11 设 A 为 m 阶矩阵, B 为 n 阶矩阵, 则有 $A \otimes B$ 相似于 $B \otimes A$.

证明 容易验证, 对矩阵 $A \otimes I_n$ 进行一系列“相合”变换^①, 可以变成 $I_n \otimes A$, 即存在一个 mn 阶置换矩阵(有限个初等矩阵的乘积) P , 使

$$P^T (A \otimes I_n) P = I_n \otimes A.$$

同理, 对矩阵 $I_m \otimes B$ 也有

$$P^T (I_m \otimes B) P = B \otimes I_m.$$

再由此种初等矩阵的性质知 $PP^T = I$, 有

$$\begin{aligned} P^T (A \otimes B) P &= P^T (A \otimes I_n) (I_m \otimes B) P \\ &= P^T (A \otimes I_n) P P^T (I_m \otimes B) P \\ &= (I_n \otimes A) (B \otimes I_m) \\ &= B \otimes A. \end{aligned}$$

证毕

① 对矩阵的行和相应的列进行相同的初等变换, 这里是指对调矩阵的第 i 行与第 j 行, 然后再对调第 i 列与第 j 列.

例 7.6.5 求矩阵

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 & 0 & 2 & 0 \\ 2 & 4 & 0 & 0 & 2 & 4 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 2 & 1 & 2 & 1 & 2 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 3 & 6 \end{bmatrix}$$

的特征值及相应特征向量.

解 令

$$\mathbf{B} = \begin{bmatrix} 2 & 0 & 2 \\ 1 & 1 & 1 \\ 0 & 0 & 3 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix},$$

则可看出 $\mathbf{A} = \mathbf{B} \otimes \mathbf{C}$, 容易求出 \mathbf{B} 的特征值为 1, 2, 3; \mathbf{C} 的特征值为 1, 2. 因此, 求得 \mathbf{A} 的 6 个特征值为 1, 2, 2, 3, 4, 6.

容易求出对应于 1, 2, 3 的特征向量分别为 $\mathbf{x}_1 = (0, 1, 0)^T$, $\mathbf{x}_2 = (1, 1, 0)^T$, $\mathbf{x}_3 = (4, 3, 2)^T$; \mathbf{C} 的对应于 1, 2 的特征向量分别为 $\mathbf{y}_1 = (1, -1)^T$, $\mathbf{y}_2 = (0, 1)^T$. 因此 \mathbf{A} 的特征向量分别为

$$\begin{aligned} \boldsymbol{\alpha}_1 &= \mathbf{x}_1 \otimes \mathbf{y}_1 = (0, 0, 1, -1, 0, 0)^T, & \boldsymbol{\alpha}_2 &= \mathbf{x}_1 \otimes \mathbf{y}_2 = (0, 0, 0, 1, 0, 0)^T, \\ \boldsymbol{\alpha}_3 &= \mathbf{x}_2 \otimes \mathbf{y}_1 = (1, -1, 1, -1, 0, 0)^T, & \boldsymbol{\alpha}_4 &= \mathbf{x}_3 \otimes \mathbf{y}_1 = (4, -4, 3, -3, 2, -2)^T, \\ \boldsymbol{\alpha}_5 &= \mathbf{x}_2 \otimes \mathbf{y}_2 = (0, 1, 0, 1, 0, 0)^T, & \boldsymbol{\alpha}_6 &= \mathbf{x}_3 \otimes \mathbf{y}_2 = (0, 4, 0, 3, 0, 2)^T. \end{aligned}$$

关于克罗内克积的多项式的特征值问题, 我们有下面的结论.

定理 7.6.12 设 $f(x, y) = \sum_{i,j=0}^p \alpha_{ij} x^i y^j$ 是变量 x, y 的复系数多项式, 对于 $\mathbf{A} \in \mathbb{C}^{m \times m}$, $\mathbf{B} \in \mathbb{C}^{n \times n}$ 定义 mn 阶矩阵:

$$f(\mathbf{A}, \mathbf{B}) = \sum_{i,j=0}^p \alpha_{ij} \mathbf{A}^i \otimes \mathbf{B}^j. \quad (7.6.13)$$

如果 \mathbf{A} 和 \mathbf{B} 的特征值分别是 $\lambda_1, \dots, \lambda_m$ 和 μ_1, \dots, μ_n , 它们对应的特征向量分别是 $\mathbf{x}_1, \dots, \mathbf{x}_m$ 和 $\mathbf{y}_1, \dots, \mathbf{y}_n$, 则矩阵 $f(\mathbf{A}, \mathbf{B})$ 的特征值是 $f(\lambda_r, \mu_s)$, 而对应 $f(\lambda_r, \mu_s)$ 的特征向量为 $\mathbf{x}_r \otimes \mathbf{y}_s$ ($r=1, \dots, m; s=1, \dots, n$).

证明 由

$$\mathbf{A} \mathbf{x}_r = \lambda_r \mathbf{x}_r, \quad \mathbf{B} \mathbf{y}_s = \mu_s \mathbf{y}_s,$$

有

$$\mathbf{A}^i \mathbf{x}_r = \lambda_r^i \mathbf{x}_r, \quad \mathbf{B}^j \mathbf{y}_s = \mu_s^j \mathbf{y}_s,$$

于是

$$\begin{aligned} f(\mathbf{A}, \mathbf{B}) \mathbf{x}_r \otimes \mathbf{y}_s &= \left(\sum_{i,j=0}^p \alpha_{ij} \mathbf{A}^i \otimes \mathbf{B}^j \right) (\mathbf{x}_r \otimes \mathbf{y}_s) \\ &= \sum_{i,j=0}^p (\alpha_{ij} \mathbf{A}^i \otimes \mathbf{B}^j) (\mathbf{x}_r \otimes \mathbf{y}_s) \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i,j=0}^p \alpha_{ij} (\mathbf{A}^i \mathbf{x}_r \otimes \mathbf{B}^j \mathbf{y}_s) \\
 &= \sum_{i,j=0}^p \alpha_{ij} \lambda_r^i \mu_s^j \mathbf{x}_r \otimes \mathbf{y}_s \\
 &= f(\lambda_r, \mu_s) \mathbf{x}_r \otimes \mathbf{y}_s.
 \end{aligned}$$

证毕

特别地,若取 $f(x, y) = xy$, 则有

$$f(\mathbf{A}, \mathbf{B}) = \mathbf{A} \otimes \mathbf{B}.$$

应用本定理,便有如下推论.

推论 1 $\mathbf{A} \otimes \mathbf{B}$ 的特征值为 mn 个数 $\lambda_r \mu_s$ ($r=1, \dots, m; s=1, \dots, n$), 且对应 $\lambda_r \mu_s$ 的特征向量为 $\mathbf{x}_r \otimes \mathbf{y}_s$.

若取 $f(x, y) = x + y$, 即 $f(x, y) = xy^0 + x^0 y$, 则

$$f(\mathbf{A}, \mathbf{B}) = \mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{B}.$$

推论 2 $\mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{B}$ 的特征值是 $\lambda_r + \mu_s$, 其对应的特征向量是 $\mathbf{x}_r \otimes \mathbf{y}_s$ ($r=1, \dots, m; s=1, \dots, n$).

矩阵 $\mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{B}$ 称为 \mathbf{A} 与 \mathbf{B} 的克罗内克和.

最后还要介绍一个在数理统计中很有用的矩阵.

定义 7.6.2 元素为 1 或 -1 的方阵 $\mathbf{H} \in \mathbb{R}^{n \times n}$, 若有

$$\mathbf{H}\mathbf{H}^T = n\mathbf{I}_n, \quad (7.6.14)$$

则称 \mathbf{H} 为 n 阶阿达马 (Hadamard) 矩阵.

定理 7.6.13 设 \mathbf{H}_m 与 \mathbf{H}_n 均为阿达马矩阵, 则矩阵 $\mathbf{H}_m \otimes \mathbf{H}_n$ 为 mn 阶的阿达马矩阵.

证明 因为

$$\begin{aligned}
 (\mathbf{H}_m \otimes \mathbf{H}_n)(\mathbf{H}_m \otimes \mathbf{H}_n)^T &= (\mathbf{H}_m \otimes \mathbf{H}_n)(\mathbf{H}_m^T \otimes \mathbf{H}_n^T) \\
 &= (\mathbf{H}_m \mathbf{H}_m^T) \otimes (\mathbf{H}_n \mathbf{H}_n^T) = (m\mathbf{I}_m) \otimes (n\mathbf{I}_n) \\
 &= mn\mathbf{I}_{mn},
 \end{aligned}$$

故按定义, $\mathbf{H}_m \otimes \mathbf{H}_n$ 是 mn 阶的阿达马矩阵.

证毕

本节讨论的克罗内克积, 尤其是阿达马矩阵在数理统计中应用很广.

7.7 阿达马积

阿达马乘法远比通常矩阵乘法简单, 但未被广泛地了解. 它出现在很多问题中, 诸如周期函数卷积的三角矩阵, 积分方程核的积, 偏微分方程中的弱极小原理, 概率论中的特征函数, 组合论中的结合方案研究, 算子理论中关于无限矩阵的阿达马积等.

定义 7.7.1 设 $\mathbf{A} = (a_{ij}), \mathbf{B} = (b_{ij}) \in \mathbb{C}^{m \times n}$. 用 $\mathbf{A} \circ \mathbf{B}$ 表示 \mathbf{A} 和 \mathbf{B} 的对应元素相乘而得到的 $m \times n$ 矩阵:

$$\mathbf{A} \circ \mathbf{B} = \begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1n}b_{1n} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2n}b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1}b_{m1} & a_{m2}b_{m2} & \cdots & a_{mn}b_{mn} \end{pmatrix}, \quad (7.7.1)$$

称为 A 和 B 的阿达马积,也称为舒尔积.

不巧,这种乘积的记号“ \circ ”和映射复合的记号相重,因此,在容易引起误会的场合使用这种记号应加以说明.

阿达马积的可相乘条件是只要两个矩阵有相同的行数和相同的列数.

显然,如此乘积与通常矩阵乘积不同,它是可交换的,即

$$A \circ B = B \circ A. \quad (7.7.2)$$

定理 7.7.1 设 $A, B, C \in \mathbb{C}^{m \times n}$, 则有

- (1) $A \circ (B + C) = A \circ B + A \circ C$;
- (2) $A \circ (B \circ C) = (A \circ B) \circ C$;
- (3) $(A \circ B)^T = A^T \circ B^T$;
- (4) $(A \circ B)^H = A^H \circ B^H$;
- (5) 如果 A 和 B 是自伴矩阵(即埃尔米特矩阵),那么 $A \circ B$ 也是自伴矩阵;
- (6) 如果 A 和 B 是斜自伴(即反埃尔米特)矩阵,那么 $A \circ B$ 是自伴矩阵;
- (7) 如果 A 是自伴矩阵, B 是斜自伴矩阵,那么 $A \circ B$ 是斜自伴矩阵.

这些基本性质可直接由阿达马积的定义推出,留作练习.

在应用中,阿达马积还有如下的一些性质(这里证明从略).

定理 7.7.2 设 $A, B \in \mathbb{C}^{m \times n}$, 则

- (1) $\text{rank}(A \circ B) \leq (\text{rank} A)(\text{rank} B)$;
- (2) 若 A, B 是半正定矩阵,则 $A \circ B$ 也是半正定矩阵;
- (3) 若 B 是正定矩阵, A 是半正定矩阵且没有零对角元素,则 $A \circ B$ 是正定矩阵;
- (4) 若 A 和 B 都是正定矩阵,则 $A \circ B$ 也是正定矩阵.

以上定理中的(2)~(4)表明:正定矩阵类和半正定矩阵类在阿达马积运算下是封闭的.这是一个定性的结果,还有各种定量的结果.

定理 7.7.3 设 $A, B \in \mathbb{C}^{n \times n}$ 是半正定矩阵,则成立

$$\lambda_{\min}(A \circ B) \geq \lambda_{\min}(A)\lambda_{\min}(B) \quad (7.7.4)$$

和

$$\lambda_{\max}(A \circ B) \leq \lambda_{\max}(A)\lambda_{\max}(B), \quad (7.7.5)$$

其中 $\lambda_{\min}(A)$ 和 $\lambda_{\max}(A)$ 分别表示 A 的最小特征值和最大特征值.

此定理是一个比较弱的定量估计.例如,在 A 正定, $B = A^{-1}$ 的情形,按照式(7.7.4),有

$$\lambda_{\min}(A \circ A^{-1}) \geq \lambda_{\min}(A)\lambda_{\min}(A^{-1}) = \lambda_{\min}(A)/\lambda_{\max}(A),$$

得到的是一个很粗糙的下界;而依下述定理就可得到好得多的有实用价值的下界.

定理 7.7.4 设 $A, B \in \mathbb{C}^{n \times n}$ 是半正定矩阵,则成立

$$\lambda_{\min}(A \circ B) \geq \lambda_{\min}(AB^T) \quad (7.7.6)$$

和

$$\lambda_{\min}(A \circ B) \geq \lambda_{\min}(AB). \quad (7.7.7)$$

例如,对于 A 正定, $B = A^{-1}$ 的情形,按照式(7.7.7)便有

$$\lambda_{\min}(A \circ A^{-1}) \geq 1.$$

7.8 反积及非负矩阵的阿达马积

定义 7.8.1 设 $A=(a_{ij}), B=(b_{ij}) \in \mathbb{C}^{m \times n}$. 令

$$c_{ij} = \begin{cases} a_{ii}b_{ii}, & j = i, \\ -a_{ij}b_{ij}, & j \neq i, \end{cases} \quad i = 1, \dots, m, \quad j = 1, \dots, n.$$

记 $A \star B = (c_{ij}) \in \mathbb{C}^{m \times n}$, 并称其为 A 和 B 的反积(Fan 积).

反积是阿达马积的一种变异.

关于反积以及非负矩阵的阿达马积有如下的基本性质.

定理 7.8.1 (1) 若 $A, B \in \mathbb{R}^{n \times n}$ 是 M 矩阵, 则 $A \star B$ 也是 M 矩阵;

(2) 若 $A, B \in \mathbb{C}^{n \times n}$ 是 H 矩阵, 则 $A \star B$ 也是 H 矩阵, $A \circ B$ 是非奇异的.

定理 7.8.2 设 $A, B \in \mathbb{R}^{n \times n}, A \geq 0, B \geq 0$, 则

(1) $A \circ B \geq 0$, 也就是说, 非负矩阵类在阿达马积下是封闭的;

(2) $\rho(A \circ B) \leq \rho(A)\rho(B)$.

定理 7.8.3 设 $A, B \in \mathbb{R}^{n \times n}$ 是 M 矩阵, 则 $A \circ B^{-1}$ 也是 M 矩阵.

以上 3 个定理的证明从略.

7.9 克罗内克积应用举例

利用矩阵克罗内克积的性质, 能够方便地研究一般线性矩阵方程

$$A_1 X B_1 + A_2 X B_2 + \dots + A_p X B_p = C \quad (7.9.1)$$

的相容性及其解法等问题. 这里 $A_i \in \mathbb{C}^{m \times m}, B_i \in \mathbb{C}^{n \times s}, C \in \mathbb{C}^{m \times s}$ 为已知矩阵, $X \in \mathbb{C}^{m \times n}$ 是未知矩阵.

对于矩阵方程(7.9.1), 可以将其转化为通常的线性方程组

$$Gx = c \quad (7.9.2)$$

来讨论. 其中系数矩阵 G 与 $A_i, B_i (i=1, 2, \dots, p)$ 有关, 向量 x 与矩阵 X 有关, 向量 c 与矩阵 C 有关. 为此, 先引入下面矩阵拉直的概念.

7.9.1 矩阵的拉直

定义 7.9.1 设 $A=(a_{ij})_{m \times n}$, 将 A 的各行依次按列纵排得到的 mn 维列向量, 这种运算称为 A 的拉直, 记为 \vec{A} , 即

$$\vec{A} = (a_{11}, a_{12}, \dots, a_{1n}, a_{21}, a_{22}, \dots, a_{2n}, \dots, a_{m1}, a_{m2}, \dots, a_{mn})^T. \quad (7.9.3)$$

从定义 7.9.1 可看出, \vec{A} 是 $mn \times 1$ 阶矩阵, 即为一个列向量. 这个列向量先把 A 的第 1 行按顺序写在前面, 依次再写第 2 行、第 3 行等, 最后写第 m 行.

例 7.9.1 设 $A = \begin{bmatrix} 1 & -1 \\ 3 & 1 \end{bmatrix}$, 则

$$\vec{A} = (1, -1, 3, 1)^T.$$

定理 7.9.1 拉直算子是线性的,即

$$\overrightarrow{\mathbf{A} + \mathbf{B}} = \overrightarrow{\mathbf{A}} + \overrightarrow{\mathbf{B}}, \quad \overrightarrow{k\mathbf{A}} = k\overrightarrow{\mathbf{A}}.$$

这些都是显然的.

定理 7.9.2 (1) $\mathbf{xy}^T = \mathbf{x} \otimes \mathbf{y}$, 其中 \mathbf{x}, \mathbf{y} 为 n 维列向量;

(2) $\mathbf{E}_{ij} = \mathbf{e}_i \mathbf{e}_j^T$, 其中 \mathbf{E}_{ij} 表示 (i, j) 元素为 1, 其余元素为 0 的 $m \times n$ 阶矩阵, \mathbf{e}_i 表示第 i 个元素为 1, 其余元素为 0 的列向量;

$$(3) \mathbf{A} \mathbf{e}_i = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{pmatrix};$$

$$(4) \mathbf{e}_j^T \mathbf{A} = (a_{j1}, a_{j2}, \dots, a_{jn});$$

$$(5) \overrightarrow{\mathbf{E}_{ij}} = \mathbf{e}_i \otimes \mathbf{e}_j.$$

这个定理的证明作为练习.

定理 7.9.3 设 $\mathbf{A} = (a_{ij})_{m \times n}$, $\mathbf{B} = (b_{ij})_{n \times p}$, $\mathbf{C} = (c_{ij})_{p \times q}$, 则

$$\overrightarrow{\mathbf{ABC}} = (\mathbf{A} \otimes \mathbf{C}^T) \overrightarrow{\mathbf{B}}. \quad (7.9.4)$$

证明 证明分两步, 先证

$$\overrightarrow{\mathbf{AE}_{ij}\mathbf{C}} = (\mathbf{A} \otimes \mathbf{C}^T) \overrightarrow{\mathbf{E}_{ij}}, \quad (7.9.5)$$

其中 \mathbf{E}_{ij} 为 $n \times p$ 阶矩阵.

事实上,

$$\begin{aligned} \overrightarrow{\mathbf{AE}_{ij}\mathbf{C}} &= \overrightarrow{\mathbf{Ae}_i \mathbf{e}_j^T \mathbf{C}} = \overrightarrow{\mathbf{Ae}_i (\mathbf{C}^T \mathbf{e}_j)^T} \\ &= \mathbf{Ae}_i \otimes \mathbf{C}^T \mathbf{e}_j. \end{aligned}$$

另一方面, 有

$$(\mathbf{A} \otimes \mathbf{C}^T) \overrightarrow{\mathbf{E}_{ij}} = (\mathbf{A} \otimes \mathbf{C}^T) (\mathbf{e}_i \otimes \mathbf{e}_j) = \mathbf{Ae}_i \otimes \mathbf{C}^T \mathbf{e}_j,$$

即证明了式(7.9.5). 下面再证明式(7.9.4), 由于

$$\mathbf{B} = (b_{ij})_{n \times p} = \sum_{i=1}^n \sum_{j=1}^p b_{ij} \mathbf{E}_{ij},$$

所以

$$\begin{aligned} \overrightarrow{\mathbf{ABC}} &= \overrightarrow{\mathbf{A} \left(\sum_{i=1}^n \sum_{j=1}^p b_{ij} \mathbf{E}_{ij} \right) \mathbf{C}} \\ &= \sum_{i=1}^n \sum_{j=1}^p b_{ij} \overrightarrow{\mathbf{AE}_{ij}\mathbf{C}} \\ &= \sum_{i=1}^n \sum_{j=1}^p b_{ij} (\mathbf{A} \otimes \mathbf{C}^T) \overrightarrow{\mathbf{E}_{ij}} \\ &= (\mathbf{A} \otimes \mathbf{C}^T) \sum_{i=1}^n \sum_{j=1}^p b_{ij} \overrightarrow{\mathbf{E}_{ij}} \\ &= (\mathbf{A} \otimes \mathbf{C}^T) \overrightarrow{\mathbf{B}}. \end{aligned}$$

证毕

推论 设 $\mathbf{A} = \mathbf{A}_{m \times m}$, $\mathbf{B} = \mathbf{B}_{n \times n}$, $\mathbf{X} = \mathbf{X}_{m \times n}$, 则

$$(1) \overrightarrow{\mathbf{AX}} = (\mathbf{A} \otimes \mathbf{I}_n) \overrightarrow{\mathbf{X}};$$

$$(2) \overrightarrow{XB} = (\mathbf{I}_m \otimes \mathbf{B}^T) \vec{X};$$

$$(3) \overrightarrow{AX + XB} = (\mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{B}^T) \vec{X}.$$

7.9.2 线性矩阵方程的解

定理 7.9.4 矩阵 $\mathbf{X} \in \mathbb{C}^{m \times n}$ 是矩阵方程(7.9.1)的解的充分必要条件是 $\mathbf{x} = \vec{X}$ 为通常的线性方程组

$$\mathbf{G}\mathbf{x} = \mathbf{c} \quad (7.9.6)$$

的解,其中 $\mathbf{G} = \sum_{i=1}^p \mathbf{A}_i \otimes \mathbf{B}_i^T, \mathbf{c} = \vec{C}$.

证明 对矩阵方程(7.9.1)两端拉直,有

$$\begin{aligned} \vec{C} &= \overrightarrow{\sum_{i=1}^p \mathbf{A}_i \mathbf{X} \mathbf{B}_i} = \sum_{i=1}^p \overrightarrow{\mathbf{A}_i \mathbf{X} \mathbf{B}_i} \\ &= \sum_{i=1}^p (\mathbf{A}_i \otimes \mathbf{B}_i^T) \vec{X} \\ &= \mathbf{G} \vec{X}, \end{aligned}$$

即

$$\mathbf{G}\mathbf{x} = \mathbf{c}.$$

故矩阵方程(7.9.1)的解与通常的线性方程组(7.9.6)的解相同.

证毕

这样,欲求矩阵方程(7.9.1)的解,只要将它转化为通常的线性方程组(7.9.6)求解就行了.

推论 1 矩阵方程(7.9.1)有解(相容)的充要条件是

$$\text{rank}(\mathbf{G} \mid \mathbf{c}) = \text{rank}(\mathbf{G}).$$

推论 2 矩阵方程(7.9.1)有惟一解的充要条件是 $\text{rank}(\mathbf{G} \mid \mathbf{c}) = \text{rank}(\mathbf{G}) = ms$.

下面我们来讨论矩阵方程(7.9.1)两个重要的特殊情况.

(1) 设 $\mathbf{A} \in \mathbb{C}^{m \times m}, \mathbf{B} \in \mathbb{C}^{n \times n}, \mathbf{C} \in \mathbb{C}^{m \times n}$, 方程

$$\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{B} = \mathbf{C}. \quad (7.9.7)$$

定理 7.9.5 矩阵方程(7.9.7)有惟一解 $\mathbf{X} \in \mathbb{C}^{m \times n}$ 的充要条件是 \mathbf{A} 和 $-\mathbf{B}$ 没有相同的特征值,即

$$\lambda_i + \mu_j \neq 0, \quad i = 1, \dots, m, \quad j = 1, \dots, n. \quad (7.9.8)$$

证明 将矩阵方程(7.9.7)两端拉直,并利用定理 7.9.3 推论(3)的结论知,方程(7.9.7)等价于

$$(\mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{B}^T) \vec{X} = \vec{C}. \quad (7.9.9)$$

再由定理 7.9.4 的推论 2 知,方程(7.9.7)有惟一解的充要条件是矩阵 $\mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{B}^T$ 是非奇异的,即矩阵 $\mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{B}^T$ 没有零特征值.

如果设 \mathbf{A} 的特征值为 $\lambda_1, \lambda_2, \dots, \lambda_m, \mathbf{B}$ (或 \mathbf{B}^T) 的特征值为 $\mu_1, \mu_2, \dots, \mu_n$, 则由定理 7.6.10 知,矩阵 $\mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{B}^T$ 的特征值为 $\lambda_i + \mu_j$ ($i = 1, \dots, m; j = 1, \dots, n$), 于是方程(7.9.7)有惟一解的充要条件是 $\lambda_i + \mu_j \neq 0$, 即 \mathbf{A} 与 $-\mathbf{B}$ 没有相同的特征值. 证毕

推论 设 $A \in \mathbb{C}^{m \times m}$, 则矩阵方程 $AX - XA = 0$ (即 $AX = XA$) 必有非零解 $X \in \mathbb{C}^{m \times m}$.

(2) 设 $A \in \mathbb{C}^{m \times m}$, $B \in \mathbb{C}^{n \times n}$, $C \in \mathbb{C}^{m \times n}$, 方程

$$X + AXB = C. \quad (7.9.10)$$

定理 7.9.6 矩阵方程(7.9.10)有惟一解 $X \in \mathbb{C}^{m \times n}$ 的充要条件是 $\lambda_i \mu_j \neq -1$ ($i=1, \dots, m$; $j=1, \dots, n$), λ_i 和 μ_j 分别为 A 与 B 的特征值.

证明 把方程(7.9.10)两端拉直, 有

$$\vec{C} = \overrightarrow{I_m X I_n + AXB} = (I_m \otimes I_n + A \otimes B^T) \vec{X},$$

于是方程(7.9.10)有惟一解的充要条件是矩阵 $I_m \otimes I_n + A \otimes B^T$ 的特征值全不为零, 由定理 7.6.10 知

$$1 + \lambda_i \mu_j \neq 0. \quad \text{证毕}$$

习 题 7 (2)

1. 设 $A = \text{diag}(1, 2, \dots, n)$, m 阶矩阵 B 的特征值是 $\lambda_1, \lambda_2, \dots, \lambda_m$ ($m > 1$), 试计算 $A \otimes B$ 的特征值.
2. 设 $A_{m \times m}$ 和 $B_{n \times n}$ 都是酉矩阵, 计算 $(A^H \otimes B)(A \otimes B^H)$.
3. 设 $A^2 = A, B^2 = B$, 证明: $(A \otimes B)^2 = A \otimes B$.
4. 设 A 和 B 都是(半)正定矩阵, 证明: $A \otimes B$ 也是(半)正定矩阵.
5. 用克罗内克积求矩阵 A 的特征值及相应的特征向量.

$$A = \begin{pmatrix} 2 & 0 & 0 & 0 & 2 & 0 \\ 2 & 4 & 0 & 0 & 2 & 4 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 2 & 1 & 2 & 1 & 2 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 3 & 6 \end{pmatrix}$$

6. 设 $A \in \mathbb{C}^{m \times m}$, $B \in \mathbb{C}^{n \times n}$, 它们的特征向量分别为 ξ 和 η , 证明: $\xi \otimes \eta$ 是 $A \otimes B$ 的特征向量.
7. 证明: 两个反埃尔米特矩阵的直积是埃尔米特矩阵.
8. 设 x 是 m 维列向量, y 是 n 维列向量, 且 $\|x\|_2 = \|y\|_2 = 1$, 试计算 $\|x \otimes y\|_2$.
9. 设 $x \in \mathbb{R}^n$ 是单位列向量, $A \in \mathbb{R}^{n \times n}$ 是正交矩阵, 计算 $\|A \otimes x\|_F$.
10. 设 A 和 B 分别是 m 阶和 n 阶酉矩阵, 试计算 $\|A \otimes B\|_2$.
11. 已知 n 阶矩阵 A 和 B 的特征值分别为 $\lambda_1, \lambda_2, \dots, \lambda_n$ 和 $\mu_1, \mu_2, \dots, \mu_n$. 问矩阵方程 $A^2 X + X B^2 - 2AXB = 0$ 有非零解的充分必要条件是什么?
12. 证明: $(A \otimes B)^+ = A^+ \otimes B^+$.

矩阵在数学内外的应用

引言

一位网友在自动化网论坛中写道：“请问大侠们，学好矩阵理论对研究生们真的很有用吗？”回答是肯定的。我们说，随着科学技术的发展，矩阵的应用无处不在，矩阵的渗透会越来越深入。可以说矩阵已经成为一门独立的理论和工具，在各个领域发挥着越来越重要的作用。本章将分别从数学内与外两个方面，列举若干矩阵的应用实例，目的是抛砖引玉，启发读者举一反三，鼓励大家在今后的学习和工作中，大胆推广矩阵的应用。

8.1 矩阵在数学内部的应用

8.1.1 矩阵在代数中的应用^[9]

1. 关于零点定理的证明

在代数中，零点定理对于代数几何有着特别重要的意义，它有多个证明方法，这里我们给出一个用到矩阵的证明。

引理 8.1.1 (诺特(Noether)规范化引理) 设 F 是一个无限域， f 是 $F[x_1, x_2, \dots, x_n]$ 中的一个次数为 d 的多项式， $n \geq 2, d \geq 1$ ，则存在 $\lambda_1, \lambda_2, \dots, \lambda_{n-1} \in F$ 使得在多项式

$$f(x_1 + \lambda_1 x_n, \dots, x_{n-1} + \lambda_{n-1} x_n, x_n) \quad (8.1.1)$$

中 x_n^d 的系数不为零。

证明 设 f_d 是 f 的 d 次齐次部分，则在多项式 (8.1.1) 中， x_n^d 的系数是 $f_d(\lambda_1, \dots, \lambda_{n-1}, 1)$ 。因为 $f_d(x_1, \dots, x_{n-1}, 1)$ 是 $F[x_1, \dots, x_{n-1}]$ 中的非零多项式并且 F 是无限的，存在 F^{n-1} 中的一点 $(\lambda_1, \dots, \lambda_{n-1})$ 使得 $f_d(\lambda_1, \dots, \lambda_{n-1}, 1) \neq 0$ ，这可以对变量的个数作归纳证明。

证毕

代数闭域一定是无限域。事实上，设 $K = \{a_0, a_1, \dots, a_m\}$ 是个有限域， $a_1 \neq 0$ ，则多项式

$$f(x) = a_1 + \prod_{j=0}^m (x - a_j) \text{ 在 } K \text{ 中没有根。}$$

定理 8.1.1 (Hilbert 零点定理) 设 F 是一个代数闭域， L 是 $F[x_1, x_2, \dots, x_n]$ 的一个真理想，则存在 $(a_1, a_2, \dots, a_n) \in F^n$ ，使得 $f(a_1, a_2, \dots, a_n) = 0$ 对所有 $f \in L$ 成立。

证明 假设 $L \neq \{0\}$, 否则结论平凡地成立. 对 n 作归纳, $n=1$ 的情形是显然的, 因为 $F[x]$ 的任何一个非零真理想 L 都是由一个非常量的多项式生成, 而这样一个生成元在 F 中有一个根 a , 因为 F 是代数闭的, 所以 $\forall f \in L, f(a)=0$ 成立.

现在设 $n \geq 2$, 并且假设定理对 $n-1$ 个变量的多项式环成立. 根据引理 8.1.1, 若有必要作变量替换并且把多项式乘以一个非零元, 我们可以假设 L 含有一个多项式 g , 具有形式

$$g = g_0 + g_1 x_n + \cdots + g_{k-1} x_n^{k-1} + x_n^k,$$

其中 $g_j \in F[x_1, x_2, \cdots, x_{n-1}]$, $j=0, 1, \cdots, k-1$, 用 L' 记 L 中不含有变量 x_n 的多项式的集合, 则 L' 是 $F[x_1, x_2, \cdots, x_{n-1}]$ 的一个真理想, 由归纳假设, 存在一点 $(a_1, a_2, \cdots, a_{n-1}) \in F^{n-1}$, 使得 L' 中的每个多项式在这点取值为 0, 我们断言

$$G = \{f(a_1, a_2, \cdots, a_{n-1}, x_n) \mid f \in L\}$$

是 $F[x_n]$ 的一个真理想. G 显然是 $F[x_n]$ 的理想, 假设 G 不是真理想, 则有 $f \in L$ 使得 $f(a_1, \cdots, a_{n-1}, x_n) = 1$ 将 f 写成

$$f = f_0 + f_1 x_n + \cdots + f_d x_n^d$$

其中所有 $f_i \in F[x_1, x_2, \cdots, x_{n-1}]$,

$$f_0(a_1, a_2, \cdots, a_{n-1}) = 1, f_1(a_1, a_2, \cdots, a_{n-1}) = \cdots = f_d(a_1, a_2, \cdots, a_{n-1}) = 0,$$

考虑 $k+d$ 阶矩阵

$$\mathbf{A}(x_1, x_2, \cdots, x_{n-1}) = \begin{pmatrix} f_0 & f_1 & \cdots & f_{k-1} & 0 & 0 & \cdots & 0 \\ 0 & f_0 & \cdots & f_{k-2} & f_{k-1} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & & & & \\ 0 & \cdots & 0 & f_0 & f_1 & \cdots & f_{d-1} & f_d \\ g_0 & g_1 & \cdots & g_{k-1} & 1 & 0 & \cdots & 0 \\ 0 & g_0 & \cdots & g_{k-2} & g_{k-1} & 1 & \cdots & 0 \\ & & \ddots & & & & \ddots & \\ 0 & \cdots & 0 & g_0 & g_1 & \cdots & g_{k-1} & 1 \end{pmatrix},$$

其中 \mathbf{A} 的对角线上有 k 个 f_0 , d 个 1. \mathbf{A} 的行列式

$$R(x_1, x_2, \cdots, x_{n-1}) = \det \mathbf{A}(x_1, x_2, \cdots, x_{n-1})$$

就是 f 和 g 关于 x_n 的结式. 熟知且易验证 R 是 f 和 g 的线性组合, 即存在 $u, w \in F[x_1, x_2, \cdots, x_{n-1}]$ 满足 $R = uf + wg$. 于是, $R \in L$, 从而 $R \in L'$. 因为 $\mathbf{A}(a_1, a_2, \cdots, a_{n-1})$ 是一个对角元素全为 1 的下三角矩阵, $R(a_1, a_2, \cdots, a_{n-1}) = 1$. 这与 $R \in L'$ 矛盾, 因为 L' 中的每个多项式在点 $(a_1, a_2, \cdots, a_{n-1})$ 取值都是 0. 所以 G 是 $F[x_n]$ 的一个真理想.

作为 $F[x_n]$ 的真理想, G 由某个多项式 $h(x_n)$ 生成, $\deg h \geq 1$ 或者 $h=0$. 因为 F 是代数闭的, 无论哪种情况都有 $a_n \in F$ 使得 $h(a_n) = 0$. 因此对所有 $f \in L, f(a_1, \cdots, a_{n-1}, a_n) = 0$.

证毕

2. 多项式根的界

我们想估计复系数多项式

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

的根的范围, 这里 $n \geq 2$. Carmichael 和 Mason 关于 $p(z)$ 的根 z 的界是

$$|z| \leq \left(1 + \sum_{i=0}^{n-1} |a_i|^2\right)^{1/2}. \quad (8.1.2)$$

现在可以利用伙伴矩阵的奇异值来改进这个界. $p(z)$ 的伙伴矩阵是

$$\mathbf{G}(p) = \begin{bmatrix} -a_{n-1} & -a_{n-2} & \cdots & -a_1 & -a_0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

设 z_1, z_2, \dots, z_n 是 $p(z)$ 的根, 且 $|z_1| \geq |z_2| \geq \cdots \geq |z_n|$, 则 z_1, z_2, \dots, z_n 是 $\mathbf{G}(p)$ 的特征值. 记 $\mathbf{G}(p)$ 的奇异值为 $s_1 \geq s_2 \geq \cdots \geq s_n$, 根据 Weyl 定理, 有

$$\prod_{i=1}^k |z_i| \leq \prod_{i=1}^k s_i, \quad k = 1, 2, \dots, n; \quad (8.1.3)$$

$$\prod_{i=k}^n |z_i| \geq \prod_{i=k}^n s_i, \quad k = 1, 2, \dots, n. \quad (8.1.4)$$

下面介绍 Roger A. Horn 求出伙伴矩阵 $\mathbf{G}(p)$ 的奇异值 $s_i (i=1, 2, \dots, n)$ 的方法.

记

$$\gamma = \sum_{i=0}^{n-1} |a_i|^2,$$

则

$$\mathbf{I} - \mathbf{G}(p)\mathbf{G}(p)^* = \begin{bmatrix} 1 - \gamma & a_{n-1} & a_{n-2} & \cdots & a_1 \\ \bar{a}_{n-1} & 0 & 0 & \cdots & 0 \\ \bar{a}_{n-2} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \bar{a}_1 & 0 & 0 & \cdots & 0 \end{bmatrix} \quad (8.1.5)$$

式(8.1.5)中的埃尔米特矩阵至多有两列线性无关, 所以它的秩至多为 2, 从而它有至少 $n-2$ 个零特征值. 于是 $\mathbf{G}(p)\mathbf{G}(p)^*$ 有至少 $n-2$ 个特征值等于 1. 设 $\mathbf{G}(p)\mathbf{G}(p)^*$ 的其余两个特征值为 λ 和 μ , 我们有

$$n-2 + \lambda + \mu = \operatorname{tr} \mathbf{G}(p)\mathbf{G}(p)^* = \|\mathbf{G}(p)\|_F^2 = n-1 + \gamma,$$

$$\lambda\mu \cdot \underbrace{1 \cdots 1}_{n-2} = \lambda\mu = \det \mathbf{G}(p)\mathbf{G}(p)^* = |\det \mathbf{G}(p)|^2 = |a_0|^2,$$

可见 λ 和 μ 是方程 $x^2 - (\gamma+1)x + |a_0|^2 = 0$ 的根. 这样我们就得到了以下引理.

引理 8.1.2 $\mathbf{G}(p)$ 的奇异值是

$$s_1 = \left\{ \frac{\gamma+1 + [(\gamma+1)^2 - 4|a_0|^2]^{1/2}}{2} \right\}^{1/2},$$

$$s_n = \left\{ \frac{\gamma+1 - [(\gamma+1)^2 - 4|a_0|^2]^{1/2}}{2} \right\}^{1/2},$$

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

将式(8.1.3)、式(8.1.4)和引理 8.1.2 合起来可以得到以下定理.

定理 8.1.2

$$\prod_{i=1}^k |z_i| \leq s_1, \quad k = 1, 2, \dots, n-1,$$

$$\prod_{i=k}^n |z_i| \geq s_n, \quad k = 2, 3, \dots, n.$$

显然,定理 8.1.2 的特殊情形 $|z_1| \leq s_1$ (从而 $|z_i| \leq s_i, i=1,2,\dots,n$) 优于式(8.1.2)的界.

8.1.2 矩阵在几何中的应用

1. 有限几何^[9]

定义 8.1.1 设 S 是一个有限集合, S 中的元素称为点, 设 S_1, S_2, \dots, S_m 为 S 的子集, 每个 S_i 含有至少两个点, 这些子集 S_i 称为线. 如果它们满足下面三条公理:

- (1) 任给两点, 存在一条线包含这两点;
- (2) 两条线相交于至多一个点;
- (3) 存在至少两条线;

则说这些点和线构成一个有限几何.

公理(1)和公理(2)合起来表明: 任给两点, 存在惟一的一条线包含它们.

定义 8.1.2 设给定有限几何 $G=(S; S_1, \dots, S_m)$, 其中 $S=\{p_1, p_2, \dots, p_n\}$, S_j 是线(含至少两个点), 我们把“包含点 p_i 的线的条数”称为点 p_i 的**次数** d_i , 而如下形成的 0-1 矩阵 $A=(a_{ij})_{n \times m}$, 其中

$$a_{ij} = \begin{cases} 1, & \text{若点 } p_i \text{ 在线 } S_j \text{ 上,} \\ 0, & \text{否则,} \end{cases}$$

称为 G 的**关联矩阵**.

显然, A 的行 i 对应点而列 j 对应线, 且有

$$d_i = \sum_{j=1}^m a_{ij} \geq 2, \quad k_j = \sum_{i=1}^n a_{ij}.$$

定理 8.1.3 每个有限几何的线的条数不少于点的个数.

证明 用 J 表示所有元素都是 1 的方阵, 其阶数由上下文确定. 下面要证明 $m \geq n$.

我们有

$$AA^T = \text{diag}(d_1 - 1, d_2 - 1, \dots, d_n - 1) + J.$$

由于 $d_i \geq 2, i=1, 2, \dots, n$, 所以 AA^T 等一个正定矩阵加一个半正定矩阵, 因而 AA^T 是正定矩阵, 故

$$n = \text{rank}(AA^T) = \text{rank}(A),$$

但 A 是 $n \times m$ 的, 故 $m \geq n$.

证毕

定理 8.1.4 若一个有限几何的线数等于点数, 则任意两条线都相交.

证明 沿用上面的记号, 现在 $m=n$, 记 $B=J-A^T$, 则

$$AB = AJ - AA^T = \begin{bmatrix} 0 & d_1 - 1 & d_1 - 1 & \cdots & d_1 - 1 \\ d_2 - 1 & 0 & d_2 - 1 & \cdots & d_2 - 1 \\ d_3 - 1 & d_3 - 1 & 0 & \cdots & d_3 - 1 \\ \vdots & \vdots & \vdots & & \vdots \\ d_n - 1 & d_n - 1 & d_n - 1 & \cdots & 0 \end{bmatrix},$$

$$\det(AB) = \prod_{i=1}^n (d_i - 1) \det(J - I) \neq 0,$$

所以 $\det \mathbf{B} \neq 0$, 从而 \mathbf{B} 有一条广义对角线上的每个元素都不为 0. 因为 \mathbf{B} 是 0-1 矩阵, 那条广义对角线上的元素全为 1, 于是 \mathbf{A} 有一条广义对角线上的元素全为 0. 可见存在置换矩阵 \mathbf{P} , 使得 \mathbf{PA} 的主对角元素全为 0. 因此将点重新命名就可以假定 \mathbf{A} 的主对角元全为 0. 此时, 点 p_i 不在线 S_i 上, 而线 S_i 上的每个点产生一条通过 p_i 的线, 所以 $k_i \leq d_i, i = 1, 2, \dots, n$. 但是, 由于

$$\sum_{i=1}^n k_i = \sum_{i,j=1}^n a_{ij} = \sum_{i=1}^n d_i,$$

得到 $k_i = d_i, i = 1, 2, \dots, n$.

我们来数不同的两个点组成的无序对的个数. 注意到每对不同的点都位于惟一的一条线上, 有

$$\begin{aligned} \binom{n}{2} &= \sum_{i=1}^n \binom{k_i}{2} = \sum_{i=1}^n \binom{d_i}{2}, \\ &= \text{Card}\{(i, j) \mid 1 \leq i < j \leq n, S_i \cap S_j \neq \emptyset\} \end{aligned}$$

其中 Card 表示集合的势, 这就证明任意两条线都必须相交.

证毕

2. 勾股定理证明中的矩阵方法^[13]

设一个直角三角形的斜边长为 c , 两个直角边长分别为 a, b . 假若选取三角形面积为 S , 两个锐角 α, β 及斜边 c 是待研究的变量, 则必有如下的函数关系:

$$f(S, c, \alpha, \beta) = 0. \quad (8.1.6)$$

下面我们应用 20 世纪初物理学中“量纲分析法”来建立 S 与 c, α, β 之间的一种数学模型. 许多物理量都是有量纲的, 它分为基本量纲和导出量纲. 基本量纲包括时间 $[t] = T$ 、质量 $[m] = M$ 和长度 $[l] = L$, 其他量都是导出量, 如, 加速度 $[a] = LT^{-2}$ 就是由基本量纲导出的. 等号两端各变量遵循量纲一致的原则, 将上述各量的量纲列成矩阵如下, 每列代表 1 个变量的量纲数据:

	c	α	β	S
L	1	0	0	2
M	0	0	0	0
T	0	0	0	0

这样, 就可以建立一个线性方程组:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}. \quad (8.1.7)$$

式(8.1.7)右端的向量可以看成是某线性变换作用在向量 $\begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \end{bmatrix}$ 上的像. 解之得 $x_{11} = 2$,

$x_{21} = 0, x_{31} = 0$, 故有关系式:

$$S = \lambda c^2 \quad (8.1.8)$$

由于式(8.1.6)中只有 3 个基本量纲, 所以对对应式(8.1.8)中的 λ 为惟一待确定的无量

纲量(常数). 式(8.1.8)说明在直角三角形中, 面积与斜边的平方成比例. 假若作斜边上的高, 将三角形分成两个相似的直角三角形, 面积分别为 S_1, S_2 , 则边长 a, b 成为两个小直角三角形的斜边, 三角形相似图如图 8.1 所示.

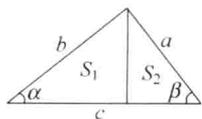


图 8.1 三角形相似图

根据上面得到的结论和相似原理, 同理有 $S_1 = \lambda b^2, S_2 = \lambda a^2$, 而 $S = S_1 + S_2$, 则

$$\lambda c^2 = \lambda b^2 + \lambda a^2,$$

勾股定理得证, 即

$$c^2 = a^2 + b^2. \quad (8.1.9)$$

8.1.3 矩阵在图论中的应用

“图论”是数学的一个分支, 它以图为研究对象, 图的基本元素是节点和边(也称线、弧、枝), 用节点表示所研究的对象, 用边表示研究对象之间的某种特定关系. 在方法上常常应用矩阵表示一个图的各种关系, 不仅是给出图的一种表示方法, 而且可以充分利用矩阵代数中的各种运算, 来研究图的结构特征(即性质), 且便于计算机处理.

1. 邻接矩阵的概念

图论中所研究的图就是节点和边的集合, 记作 $G = (V, E)$, 其中 V 表示非空的节点集合, E 表示边的集合.

在图论的研究中有如下常用的概念:

顶点数和边数 集合 V 的元素个数称为图 G 的节点数; 集合 E 的元素的个数称为图 G 的边数.

端点和关联边 若 $e = [v_i, v_j] \in E$, 则称点 v_i 和 v_j 是 e 的端点, 而称 e 是点 v_i 和 v_j 的关联边.

相邻点和相邻边 若 v_i 和 v_j 与同一条边相关联, 则 v_i 与 v_j 是相邻点; 若 e_i 与 e_j 有一个共同端点, 则 e_i 与 e_j 为相邻边.

环和多重边 若边的两个端点同属一个节点, 则称该边为环; 若两个端点之间多于一条边, 则称之为多重边.

多重图和简单图 含有多重边的图称为多重图; 无环、无多重边的图称为简单图.

次 以点 v 为端点的边数称为这个点的次, 记作 $d(v)$.

零图 一条边也没有的图.

子图 在研究和描述的性质和图的局部结构中, 子图的概念占有重要地位. 对于图 $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$, 若 $E_1 \subset E_2$, 则称 G_1 为 G_2 的子图.

连通图 如果图 G 的任意两点至少有一条通路连接起来, 则图 G 称为连通图, 否则称为不连通图.

树与图的生成树 若一个连通图中不存在任何回路, 则称之为树, 由树的定义直接得下列性质.

- (1) 树中任意两节点之间至多只有一条边;
- (2) 树中边数比节点数少 1;
- (3) 树任意去掉一条边, 就变为不连通图;
- (4) 树任意添加一条边, 就会构成一个回路.

不难看出,任意一个连通图或者就是一个树,或者去掉一些边后形成一个树.一个连通图去掉一些边后形成的树称为该连通图的生成树.一般来说,一个连通图的生成树可能不止一个.

图的矩阵 一个图由它的邻接性和关联性完全决定,这种信息可用矩阵表示,常用的有邻接矩阵和关联矩阵等.在无向图中前后相继连接的一串边的集合称为路.在有向图中,顺向的首尾相接的一串有向边的集合称为有向路.通常用顺次的节点或边来表示路或有向路,如图 8.2 中, $\{e_1, e_2, e_4\}$ 为一条路,该路也可用 $\{v_1, v_2, v_4, v_5\}$ 来表示.起点与终点为同一节点的路称为回路(或圈).

2. 无向图的邻接矩阵

定义 8.1.3 设图 G 的顶点集合为 $V(G) = \{v_1, v_2, \dots, v_p\}$, 边集为 $E(G) = \{e_1, e_2, \dots, e_q\}$, 则图 G 的邻接矩阵 $A(G) = (a_{ij})_{p \times p}$, 其中 a_{ij} 是使 v_i, v_j 连接的边的条数, 并且 $v_i \rightarrow v_i$ 是环时, $a_{ii} = 2$, 否则 $a_{ii} = 0$.

例 8.1.1 G 如图 8.3 所示.

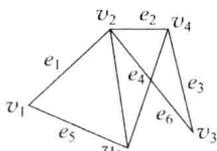


图 8.2

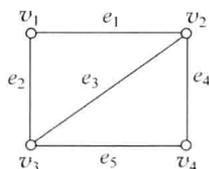


图 8.3

G 的邻接矩阵为

$$A(G) = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}.$$

例 8.1.2 G 如图 8.4 与图 8.5 所示.

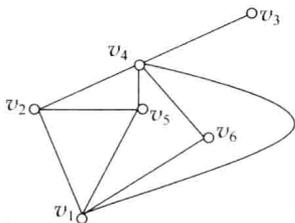


图 8.4

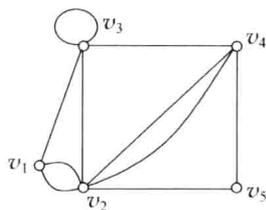


图 8.5

它们对应的邻接矩阵分别为

$$A(G) = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\text{和 } A(G) = \begin{bmatrix} 0 & 2 & 1 & 0 & 0 \\ 2 & 0 & 1 & 2 & 1 \\ 1 & 1 & 2 & 1 & 0 \\ 0 & 2 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

3. 有向图的邻接矩阵

定义 8.1.4 设 V_1, V_2, \dots, V_n 是有向图 D 的节点, 称矩阵 $A = (a_{ij})_{n \times n}$ 为 D 的邻接矩阵, 其中 a_{ij} 是以 v_i 为始点, V_j 为终点的边的条数 ($i, j = 1, 2, \dots, n$).

例 8.1.3 有向图 D 如图 8.6 所示.

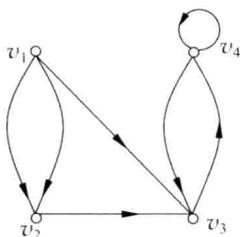


图 8.6

有向图 D 的邻接矩阵为

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

由定义知, 有向图 D 的邻接矩阵 A 具有以下性质:

- (1) 简单图的邻接矩阵是一个 0-1 的矩阵; 对角线为 0, 但不一定对称.
- (2) 矩阵的各行之和是相应顶点的出度, 各个列之和是相应顶点的入度. 所有元素相加之和与边数相等.

(3) 矩阵 A^n 的 (i, j) 位置元素为: 由 v_i 到 v_j 的长度等于 n 的通路数目, 而 (i, i) 位置的元素为: v_i 到自身的回路的数目.

4. 图论中重要定理的证明

定理 8.1.5 设图 G 的顶点集为 $V(G) = \{v_1, v_2, \dots, v_p\}$, 邻接矩阵为 A , 则 G 中从顶点 v_i 到顶点 v_j 长度为 k 的通路的条数为: $A^k = (a_{ij}^{(k)})_{p \times p}$ 中的 $a_{ij}^{(k)}$.

证明 对 k 用归纳法.

当 $k=1$ 时, 显然结论成立. 假设 k 时, 定理成立, 考虑 $k+1$ 的情形.

记 A^m 的 i 行 j 列元素为 $a_{ij}^{(m)}$, $m \geq 2$, 因为 $A^m A = A^{m+1}$, 所以

$$a_{ij}^{(m+1)} = a_{i1}^{(m)} a_{1j} + a_{i2}^{(m)} a_{2j} + \dots + a_{ip}^{(m)} a_{pj}. \quad (8.1.10)$$

而从 v_i 到 v_j 长为 $k+1$ 的通路无非是从 v_i 经 k 步到某顶点 v_m ($1 \leq m \leq p$), 再从 v_m 走一步到 v_j ; 由归纳假设从 v_i 到 v_m 长为 k 的通路共计 $a_{im}^{(k)}$ 条, 而从 v_m 到 v_j 长为 1 的通路为 a_{mj} 条, 所以长为 $k+1$ 的从 v_i 经 k 步到 v_m 并再经一步到 v_j 的通路共有 $a_{im}^{(k)} a_{mj}$ 条, 故从 v_i 经 $k+1$ 步到 v_j 的通路共有 $a_{ij}^{(k+1)} = \sum_{m=1}^p a_{im}^{(k)} a_{mj}$ 条. 证毕

例 8.1.4 如例 8.1.1 中

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

$a_{14}^{(2)} = 2$ 表示长为 2 的从 $v_1 \rightarrow v_4$ 的通路有 2 条, 即 $v_1 e_1 v_2 e_4 v_4$ 和 $v_1 e_2 v_3 e_5 v_4$.

$$A^3 = A^2 A = \begin{bmatrix} 2 & 1 & 1 & 2 \\ 1 & 3 & 2 & 1 \\ 1 & 2 & 3 & 1 \\ 2 & 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 5 & 5 & 2 \\ 5 & 4 & 5 & 5 \\ 5 & 5 & 4 & 5 \\ 2 & 5 & 5 & 2 \end{bmatrix}.$$

$a_{12}^{(3)} = 5$ 表示长为 3 的 $v_1 \rightarrow v_2$ 的通路有 5 条, 即:

$$v_1 e_1 v_2 e_1 v_1 e_1 v_2, \quad v_1 e_2 v_3 e_2 v_1 e_1 v_2, \quad v_1 e_1 v_2 e_3 v_3 e_3 v_2, \\ v_1 e_1 v_2 e_4 v_4 e_4 v_2, \quad v_1 e_2 v_3 e_5 v_4 e_4 v_2.$$

例 8.1.5 求出图 8.7 中有向图的邻接矩阵 A , 找出从 v_1 到 v_4 长度为 2 和 4 的路, 用计算 A^2, A^3 和 A^4 来验证该结论.

解 v_1 到 v_4 长度为 2 的路有 1 条: $v_1 v_2 v_4$, 长度为 4 的路有 3 条: $v_1 v_2 v_3 v_2 v_4, v_1 v_2 v_4 v_2 v_4$ 和 $v_1 v_4 v_3 v_2 v_4$.

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}; \quad A^2 = \begin{bmatrix} 0 & 1 & 2 & 1 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}; \\ A^3 = \begin{bmatrix} 0 & 3 & 2 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 2 & 1 & 0 \\ 0 & 2 & 2 & 1 \end{bmatrix}; \quad A^4 = \begin{bmatrix} 0 & 3 & 4 & 3 \\ 0 & 4 & 3 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 3 & 3 & 2 \end{bmatrix}.$$

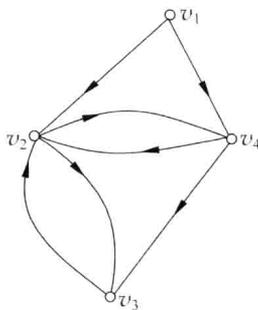


图 8.7

推论 设 G 的邻接矩阵为 $A, M = (m_{ij})_{p \times p} = \sum_{n=1}^k A^n$, 则 m_{ij} 为长度不超过 k 的从 v_i 到 v_j 通路的条数.

定理 8.1.6 (友谊定理) 如果在图 G 中任何两个不同的顶点都恰好有一个共同的邻居, 那么 G 有一个顶点与其他所有顶点都邻接.

这个结果用通俗的比方是说: 如果一群人中任何两个人都恰好有一个共同的朋友, 那么有一个人是每个人的朋友. “恰好有一个”就是有且只有一个.

证明 G 至少有 3 个顶点, 否则 G 只有一个顶点, 从而定理平凡地成立.

我们先要证明: 两个顶点 x, y 若不邻接则它们的度数相同. 事实上, 用 $N(x)$ 记与 x 邻接的顶点的集合. 定义映射 $\mathcal{A}: N(x) \rightarrow N(y)$, 对于 $Z \in N(x)$, $\mathcal{A}(x)$ 定义为 Z 与 y 共同的邻居. $\mathcal{A}(x) \neq x$, 因为 x 与 y 不邻接. 显然 \mathcal{A} 为双射, 这就证明了 x 和 y 的度数相同.

下面用反证法证明友谊定理. 假设定理结论不成立. 设有一个顶点的度数为 $k > 1$, 我们将证明所有顶点的度数都为 k . 设 Ω 是度数为 k 的顶点的集合, Γ 是度数不为 k 的顶点的集合, 假如 Γ 不为空集, 则由第一段的结论, Ω 中的每个顶点都与 Γ 中的每个顶点相邻接. 如果 Ω 或 Γ 是单点集, 那么那个顶点与 G 的所有其他顶点相邻接, 这与假设矛盾. 所以 Ω 和 Γ 都不是单点集, 但是在此情形下, Ω 中有两个顶点在 Γ 中有两个共同的邻居, 这与定理的假设矛盾. 因此, Γ 为空集, G 的所有顶点的度数都是 k , 即 G 是 k 正则的.

下一步需要证明 G 的顶点个数 $n = k(k-1) + 1$.

事实上, 我们可用两种方法来计算 G 的长度为 2 的路的条数 t . 一方面由假设 $t = \binom{n}{2}$;

另一方面对于每个顶点 v , 恰好有 $\binom{k}{2}$ 条长度为 2 的路以 v 为中间点, 因此 $t = n \cdot \binom{k}{2}$. 由

$$\binom{n}{2} = n \cdot \binom{k}{2} \text{ 解得 } n = k(k-1) + 1.$$

当 $k=1, 2$ 时, 给出 $n=1, 3$, 这两种情形定理显然成立. 下面假设 $k \geq 3$.

现在运用图 G 的邻接矩阵的特征值概念引出与 $k \geq 3$ 的矛盾. 设 \mathbf{A} 是 G 的邻接矩阵, 用 \mathbf{J} 记元素全为 1 的 n 阶矩阵, 则 $\text{tr} \mathbf{A} = 0$. 由定理的条件及 G 是 k 正则的事实得

$$\mathbf{A}^2 = (k-1)\mathbf{I} + \mathbf{J}.$$

因为 \mathbf{J} 的特征值是 n (1 重) 和 0 ($n-1$ 重), \mathbf{A}^2 的特征值是 $k-1+n=k^2$ (1 重) 和 $k-1$ ($n-1$ 重), 由谱映射定理, \mathbf{A} 的特征值是 k (1 重) 和 $\pm\sqrt{k-1}$, 设 \mathbf{A} 有 r 个特征值等于 $\sqrt{k-1}$, 有 s 个特征值等于 $-\sqrt{k-1}$, 则由 \mathbf{A} 的特征值之和等于它的迹得

$$k + r\sqrt{k-1} - s\sqrt{k-1} = 0,$$

所以, $r \neq s$ 且

$$\sqrt{k-1} = \frac{k}{s-r}.$$

可见 $\sqrt{k-1} = h$ 是有理数, 从而是整数 (因为 \sqrt{m} 是有理数 $\Rightarrow \sqrt{m}$ 必为整数). 从

$$h(s-r) = k = h^2 + 1$$

知 h 能整除 $h^2 + 1$, 于是 h 也能整除 $h^2 + 1 - h^2 = 1$, 这就得出 $h = 1$, 因而 $k = 2$. 这与假设 $k \geq 3$ 矛盾. 这就证明了假设定理的结论不成立是错的. 证毕

这个定理的证明取自文献[9].

8.2 矩阵在数学之外的应用

8.2.1 矩阵在信息编码中的应用

密码学在经济和军事方面,起着极其重要的作用,密码学中将信息代码称为密码,尚未换成密码的文字信息称为明文,由密码表示的信息称为密文. 所谓编码,是将明文加上密码加密成密文发送出去,而译码是将密文通过密码解密成明文. 由于两过程是相反的,因此矩阵的逆运算就发挥了较好的作用.

1929年,希尔(Hill)通过矩阵理论对传输信息进行加密处理,提出了在密码史上有重要地位的希尔加密算法. 下面我们介绍这种算法的基本思想.

第一步 加密

假定 26 个英文字母与数字之间有以下的一一对应关系:

$$\begin{array}{ccccccc} A & B & C & \cdots & X & Y & Z \\ \uparrow & \uparrow & \uparrow & & \uparrow & \uparrow & \uparrow \\ 1 & 2 & 3 & \cdots & 24 & 25 & 26 \end{array}$$

若要发出信息 action(行为;作用;诉讼),现需要利用矩阵乘法给出加密方法和加密后得到的密文. 一般是将单词中从左到右,每 3 个字母分为一组,并将对应的 3 个整数排成三维的列向量. 使用上述代码,则此信息的编码是: 1, 3, 20, 9, 15, 14. 可以写成两个向量:

$$\mathbf{b}_1 = \begin{bmatrix} 1 \\ 3 \\ 20 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} 9 \\ 15 \\ 14 \end{bmatrix},$$

或者写成一个矩阵(称为编码矩阵).

$$\mathbf{B} = \begin{bmatrix} 1 & 9 \\ 3 & 15 \\ 20 & 14 \end{bmatrix}.$$

现任选一个三阶的可逆矩阵,例如

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

于是将要发出的信息(编码矩阵)左乘 \mathbf{A} 对其加密,变成密文矩阵:

$$\mathbf{AB} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 2 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 9 \\ 3 & 15 \\ 20 & 14 \end{bmatrix} = \begin{bmatrix} 67 & 81 \\ 44 & 52 \\ 43 & 43 \end{bmatrix} = \mathbf{C}.$$

第二步 解密

在收到信息 $\begin{bmatrix} 67 & 81 \\ 44 & 52 \\ 43 & 43 \end{bmatrix}$ 后,可予以解密(当然,这里可逆矩阵 \mathbf{A} 是事先约定的,这个可逆

矩阵 \mathbf{A} 称为解密的钥匙或称为“密钥”).即用

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

从密码中恢复明码,得到明文矩阵:

$$\mathbf{A}^{-1}\mathbf{C} = \begin{bmatrix} 0 & 1 & -1 \\ 2 & -2 & -1 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 67 & 81 \\ 44 & 52 \\ 43 & 43 \end{bmatrix} = \begin{bmatrix} 1 & 9 \\ 3 & 15 \\ 20 & 14 \end{bmatrix}.$$

反过来查表,即可得到信息 action.

[注 1] 选择不同的可逆矩阵 \mathbf{A} (密钥),则可得到不同的密文.如:选择可逆矩阵

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 4 & 3 \end{bmatrix},$$

action 的编码矩阵是

$$\mathbf{B} = \begin{bmatrix} 1 & 9 \\ 3 & 15 \\ 20 & 14 \end{bmatrix},$$

则加密后的密文矩阵 \mathbf{C} 为

$$\mathbf{AB} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 4 & 3 \end{bmatrix} \begin{bmatrix} 1 & 9 \\ 3 & 15 \\ 20 & 14 \end{bmatrix} = \begin{bmatrix} 67 & 81 \\ 28 & 62 \\ 75 & 129 \end{bmatrix} = \mathbf{C}.$$

因为

$$\mathbf{A}^{-1} = \begin{bmatrix} 1 & 3 & -2 \\ -\frac{3}{2} & -3 & \frac{5}{2} \\ 1 & 1 & -1 \end{bmatrix},$$

所以明文矩阵为

$$\mathbf{A}^{-1}\mathbf{C} = \begin{bmatrix} 1 & 3 & -2 \\ -\frac{3}{2} & -3 & \frac{5}{2} \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 67 & 81 \\ 28 & 62 \\ 75 & 129 \end{bmatrix} = \begin{bmatrix} 1 & 9 \\ 3 & 15 \\ 20 & 14 \end{bmatrix}.$$

反过来查表即可得到信息 action.

[注2] 在假设中,也可将单词中从左到右,每4个字母分为一组,并将对应的4个整数排成四维向量,最后不足4个字母时用0补上.如信息 action,可以写成两个向量:

$$\mathbf{b}_1 = \begin{bmatrix} 1 \\ 3 \\ 20 \\ 9 \end{bmatrix}, \quad \mathbf{b}^2 = \begin{bmatrix} 15 \\ 14 \\ 0 \\ 0 \end{bmatrix},$$

编码矩阵

$$\mathbf{B} = \begin{bmatrix} 1 & 15 \\ 3 & 14 \\ 20 & 0 \\ 9 & 0 \end{bmatrix}.$$

设可逆矩阵(密钥)

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

于是加密后的密文矩阵为

$$\mathbf{AB} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 15 \\ 3 & 14 \\ 20 & 0 \\ 9 & 0 \end{bmatrix} = \begin{bmatrix} 103 & 43 \\ 70 & 14 \\ 38 & 0 \\ 9 & 0 \end{bmatrix} = \mathbf{C},$$

所以明文矩阵为

$$\mathbf{A}^{-1}\mathbf{C} = \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 103 & 43 \\ 70 & 14 \\ 38 & 0 \\ 9 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 15 \\ 3 & 14 \\ 20 & 0 \\ 9 & 0 \end{bmatrix}.$$

反过来查表,即可得到信息 action.

8.2.2 矩阵在经济模型中的应用^[13]

在经济社会中,商品交换是常事,小到二人之间的实物交换,大到产业之间的交易,都或多或少体现了商品交易中的商品价值.如何对已产出的产品进行合理的定价才能维持整个经济社会正常的运作?在这个复杂的经济系统中,矩阵可以将其转化为线性方程组,从而轻松地解决.

下面以 3 种产业之间商品交换为例,来说明商品交换的基础模型.当然,这种模型可以推广到 n 个产业的商品交换问题.

设 A, B, C 分别代表 3 种不同的产业,它们在不存在货币制度的前提下,使用实物交换制度.产业 A 将商品的 $\frac{1}{3}$ 留给自己, $\frac{1}{2}$ 给产业 B , $\frac{1}{6}$ 给产业 C ;产业 B 将商品的 $\frac{1}{3}$ 留给自己, $\frac{1}{3}$ 给产业 C ;产业 C 将商品的 $\frac{1}{3}$ 留给自己, $\frac{1}{6}$ 给产业 A , $\frac{1}{2}$ 给产业 B .应该如何给这 3 种产业进行定价,才能使这个经济体持续发展?

根据上述信息,可以构造一个有向图来表示实际实物交换的整个系统,如图 8.8 所示.

图中的箭头表示商品在某两个产业之间流出和流入的交换状况.

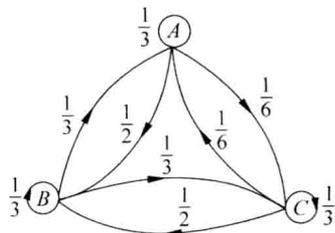


图 8.8 实物交换系统图

将 3 种产业 A, B, C 的产品分配情况列成表格,则会清晰地出现一个方阵 D :

	A	B	C
A	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
B	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{2}$
C	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$

 $\Rightarrow D = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \end{bmatrix}.$

若 3 种产业所生产出的价值分别为 x_1, x_2, x_3 ,根据上述信息,则可得下列线性方程组:

$$\begin{cases} \frac{1}{3}x_1 + \frac{1}{3}x_2 + \frac{1}{3}x_3 = x_1, \\ \frac{1}{2}x_1 + \frac{1}{3}x_2 + \frac{1}{2}x_3 = x_2, \\ \frac{1}{6}x_1 + \frac{1}{3}x_2 + \frac{1}{6}x_3 = x_3, \end{cases} \quad \text{即 } DX = IX.$$

移项后得下列齐次线性方程组:

$$\begin{cases} -\frac{2}{3}x_1 + \frac{1}{3}x_2 + \frac{1}{3}x_3 = 0, \\ \frac{1}{2}x_1 - \frac{2}{3}x_2 + \frac{1}{2}x_3 = 0, \\ \frac{1}{6}x_1 + \frac{1}{3}x_2 - \frac{5}{6}x_3 = 0, \end{cases} \quad \text{即 } (D - I)X = 0.$$

通过对系数矩阵 $D - I$ 的初等变换,得

$$\begin{bmatrix} -4 & 2 & 2 \\ 3 & -4 & 3 \\ 1 & 2 & -5 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & -5 \\ 3 & -4 & 3 \\ -2 & 1 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & -\frac{7}{5} \\ 0 & 1 & -\frac{9}{5} \\ 0 & 0 & 0 \end{bmatrix},$$

即有

$$\begin{cases} x_1 = \frac{7}{5}x_2, \\ x_2 = \frac{9}{5}x_3, \end{cases} \quad (x_3 \text{ 为自由未知量})$$

令 $x_3=5$, 得到一个基础解系 $(7, 9, 5)^\top$, 全部的解包括了所有 $(7, 9, 5)^\top$ 的倍数. 因此, 对 3 种产业 A, B, C 的总投入分配应该按照 $7:9:5$ 的比例取值, 才能使产业体持续、协调地发展.

8.2.3 矩阵在生物种群生长繁殖问题中的应用

利用矩阵方程、矩阵乘法、矩阵对角化和高次幂的知识, 能很好地预测、验证生物种群的发展情况.

假如动物种群中雌性最大生存年龄为 N 年. 将区间 $[0, N]$ 等分成 n 个年龄组, 第 i 组的年龄段为 $[\frac{i-1}{n}N, \frac{i}{n}N]$, 该年龄段的生育率(平均生育幼体的数目)为 a_i , 存活率为 b_i (指第 i 阶段存活到第 $i+1$ 阶段的数目与第 i 阶段动物总数之比). 若初始时刻动物种群年龄分布为 $\mathbf{X}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})^\top$, 取 $t_k = \frac{k}{n}N, k=1, 2, \dots$, 那么, 在 t_k 时刻动物种群年龄分布为 $\mathbf{X}^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})^\top$.

随着时间的变化, 种群各年龄段动物数目也会发生变化. 根据平衡关系, 时刻 t_k 第一个年龄组雌性动物数等于在时段 $[t_{k-1}, t_k]$ 内各年龄阶段中动物生育的幼体数目总和, 即

$$x_1^{(k)} = a_1 x_1^{(k-1)} + a_2 x_2^{(k-1)} + \dots + a_n x_n^{(k-1)}.$$

又由于

$$x_{i+1}^{(k)} = b_i x_i^{(k-1)}, \quad i = 1, 2, \dots, n-1,$$

则有

$$\begin{cases} x_1^{(k)} = a_1 x_1^{(k-1)} + a_2 x_2^{(k-1)} + \dots + a_n x_n^{(k-1)}, \\ x_2^{(k)} = b_1 x_1^{(k-1)}, \\ x_3^{(k)} = b_2 x_2^{(k-1)}, \\ \vdots \\ x_n^{(k)} = b_{n-1} x_{n-1}^{(k-1)}, \end{cases}$$

即

$$\mathbf{X}^{(k)} = \mathbf{A}\mathbf{X}^{(k-1)}, \quad k = 1, 2, \dots, n-1.$$

其中系数矩阵为

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 & \cdots & a_{n-1} & a_n \\ b_1 & 0 & \cdots & 0 & 0 \\ 0 & b_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & b_{n-1} & 0 \end{bmatrix},$$

于是

$$\mathbf{X}^{(k)} = \mathbf{A}^k \mathbf{X}^{(0)}, \quad k = 1, 2, \dots, n-1.$$

知道初始时刻, 动物种群的年龄分布, 由上式就能算出 t_k 时刻种群数目的分布 $\mathbf{X}^{(k)}$.

假如 $N=15$, 分成 3 个年龄组 $[0, 5]$, $[6, 10]$, $[11, 15]$, 统计得 3 个年龄组雌性动物的生育率分别为 0, 4, 3, 存活率为 0.5, 0.25, 0; 初始时刻 3 个年龄组雌性动物的数量为 1000, 1000, 2000, 那么 10 年后动物总数又如何?

$$\text{相当于 } \mathbf{X}^{(0)} = (1000, 1000, 2000)^T, \quad \mathbf{A} = \begin{bmatrix} 0 & 4 & 3 \\ 0.5 & 0 & 0 \\ 0 & 0.25 & 0 \end{bmatrix},$$

则

$$\mathbf{A}^2 = \begin{bmatrix} 2 & 0.75 & 0 \\ 0 & 2 & 1.50 \\ 0.125 & 0 & 0 \end{bmatrix}, \quad \mathbf{X}^{(2)} = \mathbf{A}^2 \mathbf{X}^{(0)} = (2750, 5000, 125)^T.$$

说明 10 年后, 3 个年龄段的动物总数分别发展为 2750, 5000, 125, 当 n 足够大时, 涉及矩阵高次幂 \mathbf{A}^n 的计算. 此时, 需要使用矩阵对角化的知识(见第 2 章), 再通过 $n \rightarrow \infty$ 时的极限知识(第 5 章)来研究动物总数的整个趋势, 从而科学地预测和分析动物数量动态的变化过程.

8.2.4 矩阵在控制论中的应用^{[10][17]}

在控制系统理论中, 常需要讨论如下定常线性系统:

$$\begin{cases} \frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \\ \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t), \end{cases} \quad (8.2.1)$$

其中 $\mathbf{A} \in \mathbb{C}^{n \times n}$, $\mathbf{B} \in \mathbb{C}^{n \times m}$, $\mathbf{C} \in \mathbb{C}^{s \times n}$, 三者均为常系数矩阵, $\mathbf{x}(t) = (x_1(t), x_2(t), \dots, x_n(t))^T$ 称为系统在时刻 t 的状态向量; $\mathbf{u}(t) = (u_1(t), u_2(t), \dots, u_m(t))^T$ 称为系统在时刻 t 的输入或控制向量; $\mathbf{y}(t) = (y_1(t), y_2(t), \dots, y_s(t))^T$ 称为系统在时刻 t 的输出或观测向量; \mathbf{A} 为系数矩阵; \mathbf{B} 为输入矩阵; \mathbf{C} 为输出矩阵.

上述定常线性系统(8.2.1)是控制论中最简单而又最基本的模型, 常简称为系统 $(\mathbf{A}, \mathbf{B}, \mathbf{C})$.

要控制一个定常系统(8.2.1), 需要了解系统的状态 $\mathbf{x}(t)$. 但在一般情况下, 不能直接测量到它, 必须通过得到的观测向量 $\mathbf{y}(t)$ 反过来判断 $\mathbf{x}(t)$. 能否通过观测向量 $\mathbf{y}(t)$ 确定系统的全部状态, 这便是所谓的可观测性问题, 掌握了系统的状态后, 能否控制它使其达到预期的目的, 便是所谓的可控制性问题. 除此以外, 还得分析控制系统是否能够正常工作的问题, 也就是稳定性和可检测性的问题. 下面我们将针对以上三个问题分别作简单介绍, 在讨论过程中将会使用到比较多的矩阵理论知识.

在这一节只讨论连续型的定常线性系统.

1. 系统的可观测性

定义 8.2.1 对于一个定常线性系统, 若在有限时间区间 $[0, t_1]$ 内能通过观测系统的输出 $\mathbf{y}(t)$ 而唯一地确定初始状态 $\mathbf{x}(0)$, 则称此系统是**完全能观测的**.

定理 8.2.1 系统 $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ 完全能观测 $\Leftrightarrow n$ 阶对称方阵

$$\mathbf{M}(0, t_1) = \int_0^{t_1} e^{\mathbf{A}^H t} \mathbf{C}^H \mathbf{C} e^{\mathbf{A} t} dt \quad (8.2.2)$$

为非奇异矩阵. 其中 $\mathbf{A}^H, \mathbf{C}^H$ 分别为 \mathbf{A}, \mathbf{C} 的共轭转置.

证明 充分性:

由前面 5.4.2 节知

$$\mathbf{x}(t) = e^{At} \mathbf{x}(0) + \int_0^t e^{A(t-s)} \mathbf{B} \mathbf{u}(s) ds,$$

所以

$$\mathbf{y}(t) = \mathbf{C} \mathbf{x}(t) = \mathbf{C} e^{At} \mathbf{x}(0) + \mathbf{C} \int_0^t e^{A(t-s)} \mathbf{B} \mathbf{u}(s) ds,$$

故

$$\mathbf{C} e^{At} \mathbf{x}(0) = \mathbf{y}(t) - \mathbf{C} \int_0^t e^{A(t-s)} \mathbf{B} \mathbf{u}(s) ds.$$

上式两边左乘 $e^{A^H t} \mathbf{C}^H$, 并从 0 到 t_1 积分, 得

$$\mathbf{M}(0, t_1) \mathbf{x}(0) = \int_0^{t_1} e^{A^H t} \mathbf{C}^H \left(\mathbf{y}(t) - \mathbf{C} \int_0^t e^{A(t-s)} \mathbf{B} \mathbf{u}(s) ds \right) dt.$$

由于 $\mathbf{M}(0, t_1)$ 非奇异, 所以

$$\mathbf{x}(0) = \mathbf{M}(0, t_1)^{-1} \int_0^{t_1} e^{A^H t} \mathbf{C}^H \left(\mathbf{y}(t) - \mathbf{C} \int_0^t e^{A(t-s)} \mathbf{B} \mathbf{u}(s) ds \right) dt.$$

这说明 $\mathbf{x}(0)$ 是唯一确定的, 因而系统是完全能观测的.

必要性: 设系统是完全能观测的, 则满足

$$\mathbf{y}(t) = \mathbf{C} e^{At} \mathbf{x}(0) + \mathbf{C} \int_0^t e^{A(t-s)} \mathbf{B} \mathbf{u}(s) ds. \quad (8.2.3)$$

的 $\mathbf{x}(0)$ 是唯一确定的. 若 $\mathbf{M}(0, t_1)$ 奇异, 则存在非零的 $\boldsymbol{\alpha} \in \mathbb{C}^n$ 满足

$$\mathbf{M}(0, t_1) \boldsymbol{\alpha} = \mathbf{0}$$

因而 $\boldsymbol{\alpha}^H \mathbf{M}(0, t_1) \boldsymbol{\alpha} = 0$, 于是

$$\int_0^{t_1} \boldsymbol{\alpha}^H e^{A^H t} \mathbf{C}^H \mathbf{C} e^{At} \boldsymbol{\alpha} dt = 0.$$

由此得到

$$\mathbf{y}(t) = \mathbf{C} e^{At} (\mathbf{x}(0) + \boldsymbol{\alpha}) + \mathbf{C} \int_0^t e^{A(t-s)} \mathbf{B} \mathbf{u}(s) ds.$$

这说明若 $\mathbf{x}(0)$ 满足式(8.2.2), 则 $\mathbf{x}(0) + \boldsymbol{\alpha}$ 也满足式(8.2.2). 因而由 $\boldsymbol{\alpha} \neq \mathbf{0}$ 知 $\mathbf{x}(0)$ 不是唯一确定的, 矛盾. 所以 $\mathbf{M}(0, t_1)$ 非奇异. 证毕

为了更实用, 我们给出如下判别条件(证明从略).

定理 8.2.2 系统 $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ 完全能观测 \Leftrightarrow 下列矩阵

$$\mathbf{M} = \begin{bmatrix} \mathbf{C} \\ \mathbf{C} \mathbf{A} \\ \vdots \\ \mathbf{C} \mathbf{A}^{n-1} \end{bmatrix}$$

的秩为 n , \mathbf{M} 称为能观测性矩阵.

例 8.2.1 设某系统的状态方程与输出方程为

$$\begin{cases} \frac{d\mathbf{x}(t)}{dt} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix} \mathbf{u}(t), \\ \mathbf{y}(t) = \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix} \mathbf{x}(t), \end{cases}$$

试判断该系统的能观测性.

解 能观测性矩阵为

$$\mathbf{M} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 2 & -2 \\ 0 & 2 \\ 0 & 4 \end{bmatrix}$$

所以 \mathbf{M} 的秩为 2, 说明系统是完全能观测的.

2. 系统的能控性

定义 8.2.2 对一个定常线性系统, 若在某个有限时间区间 $[0, t_1]$ 内存在着输入 $\mathbf{u}(t)$ ($0 \leq t \leq t_1$), 能使系统从初始状态 $\mathbf{x}(0)$ 转移到 $\mathbf{x}(t) = \mathbf{0}$, 则称此状态 $\mathbf{x}(0)$ 是能控的. 若对任意初始状态 $\mathbf{x}(0)$ 都是能控的, 则称此系统是完全能控的.

定理 8.2.3 系统 $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ 完全能控 $\Leftrightarrow n$ 阶对称方阵

$$\mathbf{W}(0, t_1) = \int_0^{t_1} e^{-\mathbf{A}t} \mathbf{B} \mathbf{B}^H e^{-\mathbf{A}^H t} dt \quad (8.2.4)$$

为非奇异矩阵, 其中 $\mathbf{A}^H, \mathbf{B}^H$ 分别为 \mathbf{A}, \mathbf{B} 的共轭转置.

证明 充分性: 设 $\mathbf{W}(0, t_1)$ 非奇异, 由第 5 章知

$$\mathbf{x}(t_1) = e^{\mathbf{A}t_1} \mathbf{x}(0) + \int_0^{t_1} e^{\mathbf{A}(t_1-t)} \mathbf{B} \mathbf{u}(t) dt. \quad (8.2.5)$$

令

$$\mathbf{u}(t) = -\mathbf{B}^H e^{-\mathbf{A}^H t} \mathbf{W}(0, t_1)^{-1} \mathbf{x}(0), \quad (8.2.6)$$

将 $\mathbf{u}(t)$ 代入式 (8.2.5) 得

$$\begin{aligned} \mathbf{x}(t_1) &= e^{\mathbf{A}t_1} \mathbf{x}(0) - \int_0^{t_1} e^{\mathbf{A}(t_1-t)} \mathbf{B} \mathbf{B}^H e^{-\mathbf{A}^H t} \mathbf{W}(0, t_1)^{-1} \mathbf{x}(0) dt \\ &= e^{\mathbf{A}t_1} \mathbf{x}(0) - e^{\mathbf{A}t_1} \left(\int_0^{t_1} e^{-\mathbf{A}t} \mathbf{B} \mathbf{B}^H e^{-\mathbf{A}^H t} dt \right) \mathbf{W}(0, t_1)^{-1} \mathbf{x}(0) \\ &= e^{\mathbf{A}t_1} \mathbf{x}(0) - e^{\mathbf{A}t_1} \mathbf{W}(0, t_1) \mathbf{W}(0, t_1)^{-1} \mathbf{x}(0) = \mathbf{0}. \end{aligned}$$

这说明在控制输入 $\mathbf{u}(t)$ 作用下, 能使系统 $\mathbf{x}(0)$ 转移到 $\mathbf{x}(t_1) = \mathbf{0}$. 由于 $\mathbf{x}(0)$ 的任意性, 因此系统是完全能控的.

必要性: 设有向量 $\boldsymbol{\alpha} = (a_1, a_2, \dots, a_n)^T$ 满足

$$\mathbf{W}(0, t_1) \boldsymbol{\alpha} = \mathbf{0} \quad (8.2.7)$$

则 $\boldsymbol{\alpha}^H \mathbf{W}(0, t_1) \boldsymbol{\alpha} = 0$. 因而

$$\int_0^{t_1} \boldsymbol{\alpha}^H (e^{-\mathbf{A}t} \mathbf{B} \mathbf{B}^H e^{-\mathbf{A}^H t}) \boldsymbol{\alpha} dt = 0,$$

因此推出

$$\boldsymbol{\alpha}^H e^{-\mathbf{A}t} \mathbf{B} = \mathbf{0}, \quad 0 \leq t \leq t_1. \quad (8.2.8)$$

设 $\boldsymbol{\alpha} \neq \mathbf{0}$, 由于系统是完全能控制的, 所以存在某个 $\mathbf{u}(t)$ 使其作用于系统上满足 $\mathbf{x}(t_1) = \mathbf{0}$, 故由式 (8.2.5), 有

$$e^{\mathbf{A}t_1} \mathbf{x}(0) + \int_0^{t_1} e^{\mathbf{A}(t_1-t)} \mathbf{B} \mathbf{u}(t) dt = \mathbf{0},$$

所以

$$\begin{aligned} \mathbf{x}(0) &= -e^{-A t_1} \int_0^{t_1} e^{A(t_1-t)} \mathbf{B} \mathbf{u}(t) dt \\ &= -\int_0^{t_1} e^{-A t} \mathbf{B} \mathbf{u}(t) dt. \end{aligned}$$

再由式(8.2.8)得

$$\boldsymbol{\alpha}^H \mathbf{x}(0) = -\int_0^{t_1} \boldsymbol{\alpha}^H e^{-A t} \mathbf{B} \mathbf{u}(t) dt.$$

由于 $\mathbf{x}(0)$ 是任意的, 所以可取 $\mathbf{x}(0) = \boldsymbol{\alpha}$, 于是

$$\boldsymbol{\alpha}^H \boldsymbol{\alpha} = 0.$$

这说明 $\boldsymbol{\alpha} = \mathbf{0}$, 与假设 $\boldsymbol{\alpha} \neq \mathbf{0}$ 矛盾. 这也就是说, 我们证明了如果式(8.2.7)成立, 则必有 $\boldsymbol{\alpha} = \mathbf{0}$. 因此, $\mathbf{W}(0, t_1)$ 是非奇异的. 证毕

系统的可控性有许多等价的描述, 一些常用的描述总结在下面的定理之中, 以便我们今后使用.

定理 8.2.4 设 $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times m}$, 则下面命题等价:

- (1) (\mathbf{A}, \mathbf{B}) 是可控的;
- (2) 对任意的 $t > 0$, 格拉姆(Gram)矩阵

$$\mathbf{W}(t) = \int_0^t e^{A \tau} \mathbf{B} \mathbf{B}^T e^{A^T \tau} d\tau \quad (8.2.9)$$

是正定的;

- (3) 对任意的 $t > 0$, 若 $\mathbf{v} \in \mathbb{R}^n$ 使得 $\mathbf{v}^T e^{A \tau} \mathbf{B} = \mathbf{0}$, 对一切的 $\tau \in [0, t]$ 成立, 则必有 $\mathbf{v} = \mathbf{0}$;
- (4) 可控性矩阵

$$\mathbf{G} = [\mathbf{B}, \mathbf{A}\mathbf{B}, \dots, \mathbf{A}^{n-1}\mathbf{B}] \quad (8.2.10)$$

是行满秩的;

- (5) 对一切的 $\lambda \in \lambda(\mathbf{A})$, 有 $\text{rank}[\mathbf{A} - \lambda \mathbf{I}, \mathbf{B}] = n$;
- (6) 若 $\mu \in \mathbb{R}$ 和 $\mathbf{v} \in \mathbb{R}^n$ 满足 $\mathbf{v}^T \mathbf{A} = \mu \mathbf{v}^T$ 和 $\mathbf{v}^T \mathbf{B} = \mathbf{0}$, 则必有 $\mathbf{v} = \mathbf{0}$.

这个定理的证明请参见: 徐树方著《控制论中的矩阵计算》p82-84.

有了这一结果, 我们立即可得到下面的一些应用起来十分方便的结果.

推论 设 $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times m}$, 则有

(1) 对任意的非奇异矩阵 $\mathbf{T} \in \mathbb{R}^{n \times n}$ 和非奇异矩阵 $\mathbf{G} \in \mathbb{R}^{m \times m}$, (\mathbf{A}, \mathbf{B}) 可控的充要条件是 $(\mathbf{T}^{-1}\mathbf{A}\mathbf{T}, \mathbf{T}^{-1}\mathbf{B}\mathbf{G})$ 是可控的;

(2) 若 \mathbf{A} 和 \mathbf{B} 有如下形状的分块

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix},$$

其中 $\mathbf{A}_{22} \in \mathbb{R}^{k \times k}$, 则 (\mathbf{A}, \mathbf{B}) 可控蕴涵着 $(\mathbf{A}_{22}, \mathbf{B}_2)$ 亦可控;

(3) 对任意的 $\mathbf{F} \in \mathbb{R}^{m \times n}$, (\mathbf{A}, \mathbf{B}) 可控的充要条件是 $(\mathbf{A} + \mathbf{B}\mathbf{F}, \mathbf{B})$ 是可控的.

此推论之(1)中的两个系统 (\mathbf{A}, \mathbf{B}) 和 $(\mathbf{T}^{-1}\mathbf{A}\mathbf{T}, \mathbf{T}^{-1}\mathbf{B}\mathbf{G})$ 常称为是等价的. 因此, 推论的(1)是说两个等价的系统有相同的可控性, 而(2)是说一个可控系统的子系统也是可控的, (3)表明反馈控制不改变系统的可控性.

例 8.2.2 已知

$$\mathbf{A} = \begin{bmatrix} 2 & -2 & 1 \\ 1 & 1 & 0 \\ -1 & 3 & -1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 2 & 1 \\ 1 & 0 \\ -1 & -1 \end{bmatrix},$$

试判断系统 $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ 是否完全可控?

解 由条件知,系统 $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ 具有三个状态变量、两个输入. 下面用定理 8.2.4 中的 (4) 来判断:

$$\begin{aligned} \mathbf{W} &= \begin{bmatrix} 2 & 1 & 1 & 1 & -2 & 0 \\ 1 & 0 & 3 & 1 & 4 & 2 \\ -1 & -1 & 2 & 0 & 6 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 1 & -5 & -1 & -10 & -4 \\ 1 & 0 & 3 & 1 & 4 & 2 \\ 0 & -1 & 5 & 1 & 10 & 4 \end{bmatrix} \\ &\rightarrow \begin{bmatrix} 0 & 1 & -5 & -1 & -6 & -4 \\ 1 & 0 & 3 & 1 & 4 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \end{aligned}$$

所以 \mathbf{W} 的秩为 2, 由此知系统 $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ 不是完全能控的.

3. 系统的可稳定性与可检测性

在控制系统(8.2.1)的分析与设计中,其稳定性是首先需要考虑的问题. 什么是控制系统(8.2.1)的稳定性? 回答这个问题,必须先从事力学中的自由系统

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \quad (8.2.11)$$

所谓的渐近稳定性谈起,然后再来定义控制系统(8.2.1)的稳定性.

如果对任意给定的 $\mathbf{x}_0 \in \mathbb{R}^n$, 该系统(8.2.11)所对应于 $\mathbf{x}(0) = \mathbf{x}_0$ 之解 $\mathbf{x}(t)$ 满足

$$\lim_{t \rightarrow \infty} \mathbf{x}(t) = \mathbf{0}, \quad (8.2.12)$$

则称自由系统(8.2.11)是渐近稳定的,简称稳定的.

这时,当自由系统(8.2.11)受到一个微小的扰动后,在足够长的时间内,系统能回到平衡状态.

下述定理给出了自由系统稳定性的一个判断方法.

定理 8.2.5 自由系统(8.2.11)渐近稳定的充要条件是 \mathbf{A} 的特征值的实部均为负数,即

$$\lambda(\mathbf{A}) \subset \{Z \in \mathbb{C} \mid \operatorname{Re} z < 0\}. \quad (8.2.13)$$

该定理的证明可以从 $\mathbf{x}(t) = e^{\mathbf{A}t} \mathbf{x}_0$ 和 \mathbf{A} 的若尔当分解立即得到,请读者作为练习自己补出.

设 $\mathbf{A} \in \mathbb{R}^{n \times n}$, 若 \mathbf{A} 满足条件(8.2.13), 则称 \mathbf{A} 是稳定矩阵.

可是,一般来讲,条件(8.2.13)并不是恒满足的. 若条件(8.2.13)不满足,则其对应的自由系统是不稳定的. 对于不稳定的系统而言,微小扰动就可能导导致系统的动态特性发生巨大的变化. 因此,控制论面临的另一个问题是,我们能否在控制系统(8.2.1)中,通过适当的选择控制向量 $\mathbf{u}(t)$, 使系统在“最佳状态”下运行?

最常见的一种控制是状态反馈控制,令

$$\mathbf{u}(t) = \mathbf{F}\mathbf{x}(t), \quad (8.2.14)$$

其中 $F \in \mathbb{R}^{m \times n}$, 叫做反馈矩阵. 将式(8.2.14)代入控制系统(8.2.1)的第一个方程, 得

$$\dot{x} = (A + BF)x, \quad (8.2.15)$$

其中 $A + BF$ 叫做闭环系统的状态矩阵.

只要反馈矩阵 F_0 选得适当, 闭环系统(8.2.15)就有可能是稳定的. 因此, 有如下定义:

定义 8.2.3 考察控制系统(8.2.1), 若存在 $F \in \mathbb{R}^{m \times n}$, 使得 $A + BF$ 是稳定矩阵(即闭环系统 $\dot{x} = (A + BF)x$ 是渐近稳定的), 则称控制系统(8.2.1)是可稳定的. 通常, 此时我们亦说矩阵对 (A, B) 是可稳定的. 如果 (A^T, C^T) 是可稳定的, 则称系统(8.2.1)是可检测的. 此时, 亦说矩阵对 (C, A) 是可检测的.

关于系统的可稳定性有下面常用的等价描述.

定理 8.2.6 (A, B) 可稳定的充要条件是: 若有 $\lambda \in \{z \in \mathbb{C} \mid \operatorname{Re} z \geq 0\}$ 和 $x \in \mathbb{C}^n$ 满足

$$x^T A = \lambda x^T \quad \text{和} \quad x^T B = 0, \quad (8.2.16)$$

则必有 $x = 0$.

证明 必要性: 设 (A, B) 可稳定, 则由定义知, 必存在 $F \in \mathbb{R}^{m \times n}$, 使得 $A + BF$ 是稳定矩阵, 也就是 $A + BF$ 的特征值之实部均为负数. 而式(8.2.16)成立蕴涵着

$$x^T (A + BF) = \lambda x^T$$

成立, 但 $\operatorname{Re} \lambda \geq 0$, 故 x 必然是零向量(若不然, 则 $A + BF$ 有正半平面的特征值, 与 $A + BF$ 稳定相矛盾.)

充分性: 设 A 的实舒赫(Schur)分解为

$$Q^T A Q = \begin{bmatrix} A_{11} & A_{12} \\ \mathbf{0} & A_{22} \end{bmatrix},$$

其中 $Q \in \mathbb{R}^{n \times n}$ 是正交阵, $A_{11} \in \mathbb{R}^{(n-k) \times (n-k)}$, $A_{22} \in \mathbb{R}^{k \times k}$, 并满足

$$\lambda(A_{11}) \subset \{z \in \mathbb{C} \mid \operatorname{Re} z < 0\},$$

$$\lambda(A_{22}) \subset \{z \in \mathbb{C} \mid \operatorname{Re} z \geq 0\},$$

将 $Q^T B$ 分块为

$$Q^T B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad B_z \in \mathbb{R}^{k \times m}.$$

下证充分性条件蕴涵着 (A_{22}, B_2) 是可控的. 事实上, 对任意的 $\lambda \in \lambda(A_{22})$ 和 $v \in \mathbb{C}^k$, 若有

$$v^T A_{22} = \lambda v^T \quad \text{和} \quad v^T B_2 = 0$$

成立, 则必有

$$x^T A = \lambda x^T \quad \text{和} \quad x^T B = 0$$

成立, 其中 $x^T = [0, v^T] Q^T$. 再注意 $\lambda \in \lambda(A_{22})$ 蕴涵着 $\operatorname{Re} \lambda \geq 0$. 故由充分性条件知, 必有 $x = 0$. 而 Q 正交, 故有 $v = 0$. 根据定理 8.2.4 知, (A_{22}, B_2) 是可控的. 再应用文献[17]中定理 7.2.2 可知, 存在 F_2 , 使得 $A_{22} + B_{22} F_2$ 的特征值均为 -1 . 这样, 令 $F = [0, F_2] Q^T$, 则矩阵

$$Q^T (A + BF) Q = Q^T A Q + Q^T B F Q = \begin{bmatrix} A_{11} & A_{12} + B_1 F_2 \\ \mathbf{0} & A_{22} + B_2 F_2 \end{bmatrix}$$

的特征值的实部均为负数, 即 (A, B) 是可稳定的.

证毕

利用定理 8.2.6, 立即得到下面几个十分有用的结果:

推论 1 (C, A) 是可检测的充要条件是: 若有 $\lambda \in \{z \in \mathbb{C} \mid \operatorname{Re} z \geq 0\}$ 和 $x \in \mathbb{C}^n$ 满足

$$Ax = \lambda x \quad \text{和} \quad Cx = 0, \quad (8.2.17)$$

则必有 $x=0$.

推论 2 设 $A \in \mathbb{C}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{l \times n}$, 则有

- (1) 若 (A, B) 可控, 则 (A, B) 是可稳定的;
- (2) 若 (C, A) 可观测, 则 (C, A) 是可检测的.

推论 3 设 $A \in \mathbb{C}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{l \times n}$, 则有

- (1) (A, B) 可稳定的充要条件是 (A, BB^T) 可稳定;
- (2) (C, A) 可检测的充要条件是 $(C^T C, A)$ 可测.

请读者作为练习, 给出这几个推论的证明.

模拟考试自测试题(共15套)

自测题一

一、(35分) 设 \mathbb{P} 为数域, $V=\mathbb{P}^{m \times n}$, $C=(c_{ij}) \in \mathbb{P}^{m \times n}$, $\mathcal{A}(X)=XC, \forall X \in V$.

(1) 证明: V 为数域 \mathbb{P} 上的线性空间;

(2) 求出 V 的一组基,及 $\dim V$,并验证之;

(3) 证明 \mathcal{A} 是 V 上的线性变换,并求 \mathcal{A} 在 V 上的上述基下的矩阵 A ;

(4) 当 $m=n=2, C=\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ 时,分别求 \mathcal{A} 和 A 的全部特征值和特征向量;

(5) 问 \mathcal{A} 和 A 可否对角化?若能,试求 V 的基使 \mathcal{A} 在此基下的矩阵为对角阵 Λ ,并求可逆矩阵 P ,使 $P^{-1}AP$ 为对角阵 Λ ;

(6) 给出从第(2)题中的基到第(5)题中的基变换公式,过渡矩阵和相应的坐标变换公式.

二、(15分) 设 $A=\begin{pmatrix} 3 & 1 & -1 \\ 1 & 2 & -1 \\ 2 & 1 & 0 \end{pmatrix}$.

(1) 求 A 的行列式因子、不变因子和初等因子;

(2) 求 A 的若尔当标准形;

(3) 求 $\sin A$.

三、(10分) 设 $A=\begin{pmatrix} \frac{1}{5} & \frac{1}{6} & \frac{1}{10} \\ \frac{1}{8} & 0 & \frac{1}{4} \\ \frac{1}{10} & \frac{1}{6} & \frac{1}{5} \end{pmatrix}$.

(1) 求证:矩阵序列 $\{A^k\}$ 收敛,并求当 $k \rightarrow \infty$ 时 $\{A^k\}$ 的极限;

(2) 证明:矩阵幂级数 $\sum_{k=0}^{\infty} A^k$ 绝对收敛,并写出其和矩阵的表达式.

四、(20分) 设 $A=\begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}, B=\begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix}$.

(1) 求 $\|A\|_1, \|A\|_{m_1}, \|A\|_{\infty}, \|A\|_{m_{\infty}}, \text{cond}(A)_{\infty}$ 以及 $\rho(A)$.

(2) 问 A 可否进行 LU 分解,为什么?

(3) 求 B^-, B^+, B_r^- .

五、(10分) 设线性空间 V^n 中, 从基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 到基 $\beta_1, \beta_2, \dots, \beta_n$ 的过渡矩阵为 C . 证明: 有非零向量 $\alpha \in V^n$, 使 α 在两组基下坐标相同的充要条件是 1 为矩阵 C 的特征值.

六、(10分) 设 A 是 n 阶实对称正定矩阵, 证明: 存在非奇异矩阵 P , 使 A 与单位矩阵相合, 即

$$P^T A P = I$$

自测题二

一、(8分) 设矩阵 $A = \lambda I$ ($\lambda \in \mathbb{R}$, I 单位矩阵), 按通常意义下的矩阵加法与数乘运算, 问下列集合 V 是否构成实数域 \mathbb{R} 上的线性空间? 若是, 试求出它的维数与一组基.

$$V = \{a_0 I + a_1 A + \dots + a_n A^n \mid a_i \in \mathbb{R}\}.$$

二、(12分)

(1) 在 $P[x]_2$ 中, 设 $\alpha \in P[x]_2$ 在基 $1, x, x^2$ 下的坐标为 $(1, 0, -1)$, 试写出 α 在另一组基 $1+x, x+x^2, x^2$ 下的坐标;

(2) 设 $\alpha = (a_1, a_2, a_3) \in \mathbb{R}^3$, 定义变换 \mathcal{A} 为

$$\mathcal{A}[(a_1, a_2, a_3)] = (a_1^2, a_1 + a_2, a_3),$$

问: \mathcal{A} 是否为线性变换, 为什么?

(3) 在 \mathbb{R}^2 中, 函数 $\langle \alpha, \beta \rangle = \langle (a_1, a_2), (b_1, b_2) \rangle = a_1 b_1 - a_2 b_2$ 是不是内积? 若不是, 给出反例.

三、(20分) 在 \mathbb{R}^3 中, 已知一组基 $\alpha_1 = (1, 1, 1)^T$, $\alpha_2 = (1, 2, 3)^T$, $\alpha_3 = (1, 0, 0)^T$.

(1) 试将 $\alpha_1, \alpha_2, \alpha_3$ 改造为正交基 e_1, e_2, e_3 ;

(2) 若有线性变换 \mathcal{A} , 它在基 $\alpha_1, \alpha_2, \alpha_3$ 下的矩阵为 $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, 求 \mathcal{A} 在上述正交基 e_1, e_2, e_3 下的

矩阵.

四、(15分)

(1) 设 A 是 n 阶实对称矩阵, 且 A 的特征值分别为 $1, \sqrt{2}, \dots, \sqrt{n}$. 试求 $\|A\|_F, \rho(A), \|A\|_2, \text{cond}(A)_2, \|(A^{-1})^m\|_2$;

(2) 举例说明 $\mathbb{C}^{n \times n}$ ($n > 1$) 中的矩阵范数 $\|A\|_1$ 与 \mathbb{C}^n 中的向量范数 $\|\alpha\|$ 不相容.

五、(15分) 设 $A = \begin{pmatrix} 2 & -1 & 2 \\ 5 & -3 & 3 \\ -1 & 0 & -2 \end{pmatrix}$.

(1) 求 A 的若尔当标准形;

(2) 求 $\cos A$ 的若尔当标准形.

六、(10分) 设矩阵 $A = \begin{pmatrix} 0.23 & 0.4 & 0.14 \\ 0.25 & 0 & 0.35 \\ 0.11 & 0.33 & 0.25 \end{pmatrix}$.

(1) 证明: 矩阵序列 $\{A^k\}$ 收敛, 并求其当 $k \rightarrow \infty$ 时的极限;

(2) 证明: 矩阵幂级数 $\sum_{k=0}^{\infty} A^k$ 绝对收敛, 并写出其和矩阵.

七、(10分) 已知 $A = \begin{pmatrix} -1 & 2 & 1 \\ -1 & 2 & 1 \\ 0 & 3 & 2 \end{pmatrix}$, $b = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$, 求方程组 $Ax = b$ 的极小范数最小二乘解.

八、证明题:(10分)

(1) 设 A 为实 n 阶非奇异矩阵,证明:如果 A 与 $-A$ 在实数域 \mathbb{R} 上相合,则 n 必为偶数.

(2) 设 n 阶方阵 $A = (a_{ij})_{n \times n}$, 且 $\sum_{j=1}^n |a_{ij}| < 1, i = 1, 2, \dots, n$, 证明: A 的每一个特征值 λ 的绝对值 $|\lambda| < 1$.

自测题三

一、(10分) 按通常意义下矩阵的加法和数乘运算,问下列集合 V 是否构成实数域 \mathbb{R} 上的线性空间?若是,试求它们的维数,并写出一组基.

(1) V 为所有实的 n 阶对称与反对称矩阵的全体;

(2) V 为由对角矩阵 $D = \text{diag}(d_1, \dots, d_n)$ 所生成的实系数 n 次多项式 $f(D)$ 的全体,其中 $d_i (i = 1, 2, \dots, n)$ 互异.

二、(20分) V 表示实数域上次数不超过 3 的多项式与零多项式构成的线性空间,对 $\forall f(x) = ax^2 + bx + c \in V$, 在 V 上定义线性变换:

$$\mathcal{A}[f(x)] = 3ax^2 + (2a + 2b + 3c)x + (a + b + 4c).$$

(1) 求: \mathcal{A} 在 V 的一组基 $x^2, x, 1$ 下的矩阵;

(2) 判断 \mathcal{A} 是否可以对角化,为什么?

(3) 在 V 中定义内积 $(f, g) = \int_0^1 f(t)g(t)dt$, 求基 $x^2, x, 1$ 的度量矩阵.

三、(15分) 设 $\alpha_1, \alpha_2, \alpha_3$ 是三维欧氏空间 V 的一组标准正交基,求 V 的一个正交变换 \mathcal{A} , 使得

$$\begin{cases} \mathcal{A}(\alpha_1) = \frac{2}{3}\alpha_1 + \frac{2}{3}\alpha_2 - \frac{1}{3}\alpha_3, \\ \mathcal{A}(\alpha_2) = \frac{2}{3}\alpha_1 - \frac{1}{3}\alpha_2 + \frac{2}{3}\alpha_3. \end{cases}$$

四、(15分) 设复数域 \mathbb{C} 上的线性空间 V^3 的一组基为 $\alpha_1, \alpha_2, \alpha_3$, 线性变换 \mathcal{A} 在该基下的矩阵为

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

求 V^3 的另一组基 $\beta_1, \beta_2, \beta_3$ (可用 $\alpha_1, \alpha_2, \alpha_3$ 表示), 使 \mathcal{A} 在该基下的矩阵为若尔当标准形.

五、(10分) 设 α 是给定的 n 维非零列向量, $\|A\|_F$ 是 $\mathbb{C}^{n \times n}$ 中矩阵的弗罗贝尼乌斯范数, 定义实值函数 $\|x\| = \|\alpha x^T\|_F$ (其中任意 $x \in \mathbb{C}^n$). 试证明 $\|x\|$ 是 \mathbb{C}^n 上的向量范数, 且矩阵的 F 范数与它相容.

六、(10分) 矩阵 $A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}$, 求 $\lim_{k \rightarrow \infty} \left(\frac{A}{\rho(A)} \right)^k$.

七、(10分) 设 $A \in \mathbb{C}^{n \times n}$, 证明: $\rho(A) < 1$ 的充分必要条件是存在某种矩阵范数 $\|\cdot\|$, 使得 $\|A\| < 1$.

八、(10分) 设 A, B 是两个 n 阶正交矩阵, 且 $|AB| = -1$, 试证明:

(1) $|A^T B| = |AB| = |A^T B^T| = -1$;

(2) $|A+B| = 0$.

自测题四

一、(15分) 矩阵空间 $\mathbb{R}^{2 \times 2}$ 的子空间 $V = \{X = (x_{ij})_{2 \times 2} \mid x_{11} + x_{12} + x_{21} = 0\}$ (按通常矩阵的加法和数量乘法), 在 V 中定义线性变换 $\mathcal{A}(X) = X + X^T$ (对任意 X 属于 V).

(1) 求 \mathcal{A} 在所给基 $\mathbf{E}_1 = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}$, $\mathbf{E}_2 = \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix}$, $\mathbf{E}_3 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ 下的矩阵;

(2) 在 V 中求一个基, 使 \mathcal{A} 在该基下的矩阵为对角形.

二、(10分) 设 $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ 是欧氏空间 V^3 中一组标准正交基,

$$\begin{cases} \mathbf{a}_1 = \frac{1}{3}(2\mathbf{e}_1 + 2\mathbf{e}_2 - \mathbf{e}_3), \\ \mathbf{a}_2 = \frac{1}{3}(2\mathbf{e}_1 - \mathbf{e}_2 + 2\mathbf{e}_3), \\ \mathbf{a}_3 = \frac{1}{3}(\mathbf{e}_1 - 2\mathbf{e}_2 - 2\mathbf{e}_3). \end{cases}$$

若有线性变换 \mathcal{A} 能将 $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ 变为 $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$, 试证明 \mathcal{A} 是正交变换.

三、(10分) 求解下列各题.

(1) 已知向量 $\mathbf{x} = (\xi_1, \xi_2, \dots, \xi_n)^T \in \mathbb{R}^n$, 求初等反射矩阵 \mathbf{H} , 使 $\mathbf{H}\mathbf{x} = (\xi_1, \eta_2, 0, \dots, 0)^T$;

(2) $\mathbf{A} = \begin{pmatrix} 1 & -8 \\ -2 & 1 \end{pmatrix}$, 且幂级数 $\sum_{k=0}^{\infty} \frac{k}{6^k} x^k$ 的收敛半径为 6, 则矩阵幂级数 $\sum_{k=0}^{\infty} \frac{k}{6^k} \mathbf{A}^k$ 收敛吗? 为什么?

四、(10分) 设

$$\mathbf{A} = \begin{pmatrix} \frac{1}{\sqrt{2}} & a & 0 \\ 0 & 0 & 1 \\ b & c & 0 \end{pmatrix},$$

问 a, b, c 为何值时, \mathbf{A} 为正交矩阵?

五、(20分) 设 $\mathbf{A} = \begin{pmatrix} 3 & 1 & -1 \\ 1 & 2 & -1 \\ 2 & 1 & 0 \end{pmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}$.

(1) 求 $\|\mathbf{A}\|_1, \|\mathbf{A}\|_{\infty}, \|\mathbf{A}_{m \times n}\|, \rho(\mathbf{A})$;

(2) 问 \mathbf{A} 可否进行 LU 分解, 为什么?

(3) 求 $\mathbf{A} \otimes \mathbf{B}$ 的秩, $(\mathbf{A} \otimes \mathbf{B})^2$ 的所有特征根;

(4) 求 $\mathbf{B}^-, \mathbf{B}^+, \mathbf{B}_r^-, \rho(\mathbf{A})$;

(5) 求 \mathbf{A} 的若尔当标准形;

(6) 求矩阵函数 $\frac{1}{\mathbf{A}}$ 的值.

六、(10分)

(1) \mathbf{A} 是 n 阶矩阵, 取 \mathbb{R}^n 的子空间 $V_{\lambda} = \{x | \mathbf{A}x = \lambda x\}$, 若 λ 不是 \mathbf{A} 的特征根, 则 $\dim V_{\lambda}$ 等于多少? 若 λ 是 \mathbf{A} 的特征根, 则 $\dim V_{\lambda}$ 等于多少?

(2) 设 \mathbf{A} 是 n 阶豪斯霍尔德矩阵, 求 \mathbf{A}^2 .

七、(10分) 设 $\mathbf{A} \in \mathbb{R}^{n \times m}$, $\mathbf{B} = \mathbf{A}^T \mathbf{A}$, 试证:

(1) \mathbf{B} 为半正定矩阵;

(2) 当 \mathbf{A} 的列向量组线性无关时, \mathbf{B} 为正定矩阵.

八、(15分) 证明题:

(1) 设矩阵 \mathbf{A} 非奇异, λ 是 \mathbf{A} 的任意一个特征值, 证明: $|\lambda| \geq \frac{1}{\|\mathbf{A}^{-1}\|}$;

(2) 设 \mathbf{A} 为 n 阶可逆矩阵, \mathbf{B} 为 n 阶矩阵, 若对某种矩阵范数有 $\|\mathbf{B}\| < \frac{1}{\|\mathbf{A}^{-1}\|}$, 则 $\mathbf{A} + \mathbf{B}$ 可逆.

自测题五

一、(10分) 设有矩阵空间 $\mathbb{R}^{2 \times 2}$ 的子空间

$$V_1 = \left\{ A = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \mid x_1 - x_2 + x_3 - x_4 = 0 \right\}, V_2 = L[B_1, B_2], \text{ 其中 } B_1 = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}, B_2 = \begin{bmatrix} 0 & -2 \\ 0 & 1 \end{bmatrix}.$$

- (1) 求 V_1 的一组基及维数;
 (2) 求 $V_1 + V_2$ 及 $V_1 \cap V_2$ 的维数.

二、(15分)

(1) 设 \mathbb{R}^3 中基为 $\mathbf{x}_1 = (1, 1, 1)^T, \mathbf{x}_2 = (0, 1, 1)^T, \mathbf{x}_3 = (0, 0, 1)^T$, 试写出某向量 $\mathbf{x} = (a_1, a_2, a_3)^T$ 在该基下的坐标表达式;

(2) 设线性空间 V 的线性变换 \mathcal{A} 有两个不同的特征值 λ_1 和 λ_2 , 记特征子空间 $V_{\lambda_i} = \{\mathbf{x} \mid \mathcal{A}(\mathbf{x}) = \lambda_i \mathbf{x}, \mathbf{x} \in V\} (i=1, 2)$, 试计算 $\dim(V_{\lambda_1} \cap V_{\lambda_2})$.

三、(10分) 设 $P[x]_3$ 中的多项式为 $f(x) = a_0 + a_1x + a_2x^2 + a_3x^3$, 线性变换 \mathcal{A} 定义为

$$\mathcal{A}[f(x)] = (a_0 - a_2) + (a_1 - a_3)x + (a_2 - a_0)x^2 + (a_3 - a_1)x^3.$$

求 $P[x]_3$ 的一组基, 使 \mathcal{A} 在该基下的矩阵为对角矩阵.

四、(12分) 求解下列各题.

- (1) 设 A 是可逆矩阵, 求 $\int_0^1 e^{At} dt$;
 (2) 在欧氏空间中, 求满足条件 $(\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j) = \delta_{ij}$ 的正交基 $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_n$ 的度量矩阵 A .

五、(18分)

(1) 已知 $\mathbf{x} = (-1, i, 0, 1), i = \sqrt{-1}$, 求

$$\|\mathbf{x}\|_1, \|\mathbf{x}\|_2, \|\mathbf{x}\|_\infty, \|\mathbf{x}^T \mathbf{x}\|_{m_1}, \|\mathbf{x}^T\|_F, \|\mathbf{x}^T \mathbf{x}\|_{m_\infty}, \|\mathbf{x}^T \mathbf{x}\|_\infty;$$

(2) 设 $A = \begin{pmatrix} -1 & 1 & 1 \\ -5 & 21 & 17 \\ 6 & -26 & -21 \end{pmatrix}$, 求 A 的若尔当标准形.

六、(15分) 设 $A = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 2 & 1 & 1 \end{pmatrix}, B = \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix}$.

- (1) 求 A 的最大秩分解;
 (2) 求 A^\dagger ;
 (3) 求方程组 $A\mathbf{x} = B$ 的解, 指出是哪种解.

七、(10分) 设 V^n 是实数域 \mathbb{R} 上的线性空间, $\mathbf{x} \in V^n$ 在基(I): $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_n$ 下的坐标为 $\boldsymbol{\alpha} = (\xi_1, \xi_2, \dots, \xi_n)^T$, 在基(II): $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_n$ 下的坐标为 $\boldsymbol{\beta} = (\eta_1, \eta_2, \dots, \eta_n)^T$, 且由基(I)变为基(II)的过渡矩阵为 C , $\|\cdot\|_2$ 表示 \mathbb{R}^2 中向量的2范数. 证明: $\|\boldsymbol{\alpha}\|_2 = \|\boldsymbol{\beta}\|_2$ 的充分必要条件是 C 为正交矩阵.

八、(10分) 证明: $A^+AB = A^+AC$ 的充分必要条件是 $AB = AC$.

自测题六

一、(12分) 设 $\mathbb{R}^{2 \times 2}$ 的矩阵子空间 $V = \{X = (x_{ij})_{2 \times 2} \mid x_{12} + x_{21} = 0\}$, 在 V 中定义线性变换 $\mathcal{A}(X) = B^T X - X^T B$, 其中 $B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ (对任意 X 属于 V).

- (1) 求 V 的维数, 并写出 V 的一组基;
 (2) 求 \mathcal{A} 在该基下的矩阵;
 (3) 在 V 中求一个基, 使 \mathcal{A} 在该基下的矩阵为对角形.

二、(10分) 设 $\mathbf{A} = \begin{pmatrix} 4 & -5 & 2 \\ 5 & -7 & 3 \\ 6 & -9 & 4 \end{pmatrix}$.

- (1) 求 \mathbf{A} 的不变因子和初等因子;
 (2) 求 \mathbf{A} 的若尔当标准形;
 (3) 问 \mathbf{A} 可否进行 LU 分解, 为什么?

三、(12分)

(1) 设 $\mathbf{A} = \begin{bmatrix} t^2 & 1 \\ t & 0 \end{bmatrix}$, 求 $\frac{d^2 \mathbf{A}}{dt^2}$, $\frac{d\mathbf{A}^{-1}}{dt}$, $\frac{d\mathbf{A}^2}{dt}$;

(2) 设 $\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix}$, $f(\mathbf{X}) = x_{11} + x_{12}^2 + x_{13}^3$, 求 $\frac{df}{d\mathbf{X}}$.

四、(9分)

(1) 设 $\mathbf{A} = \frac{1}{2}(\mathbf{B} + \mathbf{I})$, 试证: \mathbf{A} 为幂等矩阵 ($\mathbf{A}^2 = \mathbf{A}$) 的充分必要条件是 $\mathbf{B}^2 = \mathbf{I}$;

(2) 设 \mathbf{A}, \mathbf{B} 为 n 阶方阵, 且满足 $\mathbf{A}^2 = \mathbf{A}, \mathbf{B}^2 = \mathbf{B}$, 以及 $(\mathbf{A} + \mathbf{B})^2 = \mathbf{A} + \mathbf{B}$, 证明 $\mathbf{AB} = \mathbf{0}$.

五、求解下列各题(每题6分, 共24分).

(1) 设 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ ($m > 1$) 是 \mathbb{R}^n 中两两正交的单位向量, 记 $\mathbf{A} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)$, 求 \mathbf{A}^+ ;

(2) 设 $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, 讨论矩阵幂级数 $\sum_{k=1}^{\infty} \frac{(-1)^k}{k^2} \mathbf{A}^k$ 的敛散性;

(3) 设 $\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 2 & 5 \end{bmatrix}$, 求 $(2\mathbf{A}^4 - 12\mathbf{A}^3 + 19\mathbf{A}^2 - 29\mathbf{A} + 37\mathbf{I})^{-1}$;

(4) 设 $\mathbf{A} = \text{diag}(1, 2, \dots, n)$, m 阶矩阵 \mathbf{B} 的特征根是 $\lambda_1, \lambda_2, \dots, \lambda_m$ ($m > 1$), 求 $\mathbf{A} \otimes \mathbf{B}$ 的所有特征根.

六、(15分) 证明题:

(1) 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, $\mathbf{D} = \begin{pmatrix} \mathbf{A} \\ \mathbf{A} \end{pmatrix}$, 则 $\mathbf{D}^+ = \frac{1}{2}(\mathbf{A}^+ \ \vdots \ \mathbf{A}^+)$;

(2) 设 $\mathbf{A} = \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}$, 则 $\mathbf{X} = -(a_1^2 + a_2^2 + a_3^2)^{-1} \mathbf{A}$ 是 \mathbf{A} 的减号逆;

(3) 设 \mathbf{A} 为实反对称矩阵, 则 $e^{\mathbf{A}}$ 是正交矩阵.

七、(10分) 已知欧氏空间 V^n 的一组标准正交基 $\alpha_1, \alpha_2, \dots, \alpha_n$, 且向量 $\alpha_0 = \alpha_1 + 2\alpha_2 + \dots + n\alpha_n$, 定义变换 $\mathcal{A}(\alpha) = \alpha + k(\alpha, \alpha_0)\alpha_0$, $\alpha \in V^n$, 实数 $k \neq 0$.

证明: (1) \mathcal{A} 是线性变换;

(2) \mathcal{A} 是正交变换的充要条件是 $k = -\frac{2}{1^2 + 2^2 + \dots + n^2}$.

八、(8分) 设复矩阵 $\mathbf{A} = (a_{ij}) \in \mathbb{C}^{n \times n}$ 的特征值集合(即 \mathbf{A} 的谱)为 $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$, 试证明不等式:

$$\sum_{i=1}^n |\lambda_i|^2 \leq \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2.$$

上式当且仅当 \mathbf{A} 为正规矩阵时等号成立.

自测题七

一、(10分) 设 \mathbb{R}^4 的两个子空间为

$$V_1 = \{\alpha = (a_1, a_2, a_3, a_4) \mid a_1 + a_2 - a_4 = 0\},$$

$$V_2 = L(\beta_1, \beta_2), \quad \beta_1 = (0, 1, 1, 1), \quad \beta_2 = (1, 1, 1, 0).$$

(1) 求 $V_1 + V_2$ 的基与维数;

(2) 求 $V_1 \cap V_2$ 的基与维数.

二、(12分)

(1) 在 $P^{2 \times 2}$ 中定义线性变换

$$\mathcal{A}(X) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} X.$$

求 \mathcal{A} 在基 $E_{11}, E_{12}, E_{21}, E_{22}$ 下的矩阵;

(2) 设 A, B 为 n 阶方阵, $AB = A + B$, 证明 $AB = BA$.

三、(12分)

(1) 设 $A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 3 & 2 \end{pmatrix}$, 求 A 的若尔当标准形;

(2) 设 $A = (a_{ij}) \in \mathbb{R}^{n \times n}$, 求 $\frac{d}{dA} \text{tr} A$;

(3) 已知 $e^{At} = \begin{pmatrix} 2e^{2t} - e^t & e^{2t} - e^t & e^t - e^{2t} \\ e^{2t} - e^t & 2e^{2t} - e^t & e^t - e^{2t} \\ 3e^{2t} - 3e^t & 3e^{2t} - 3e^t & 3e^t - 2e^{2t} \end{pmatrix}$, 求 A . 四、(21分) 设 $A = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}$, $B =$

$\begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix}$. 计算:

(1) $\|A\|_1, \|A\|_\infty$; (2) $\rho(A), \text{cond}(B)_\infty$; (3) $\text{rank}(A \otimes B), A \otimes B$ 的特征值; (4) A^-, A^+ .

五、(15分) 已知 $A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \end{pmatrix}$, $b = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$.

(1) 求 A 的一个非平凡最大秩分解;

(2) 用广义逆矩阵法判别方程组 $Ax = b$ 是否相容?

(3) 求线性方程组 $Ax = b$ 的极小范数解.

六、(14分)

(1) 设可逆方阵 $P \in \mathbb{R}^{n \times n}$, 且知 $\|x\|_p = \|Px\|_2$ 是 \mathbb{R}^2 上的向量范数. 若 $\|A\|_p$ 表示 $\mathbb{R}^{n \times n}$ 上从属于向量范数 $\|x\|_p$ 的矩阵范数, 试导出 $\|A\|_p$ 与矩阵的2范数之间的关系式;

(2) 用Gerschgorin定理说明

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 4 & 0 & 1 \\ 1 & 0 & 6 & 2 \\ 0 & 1 & 1 & 8 \end{pmatrix}$$

至少有两个实特征值.

七、(16分) 证明:

(1) 正交变换的特征值等于 ± 1 ;

(2) 设 \mathcal{A} 为数域 P 上 n 维线性空间 V 的一个对合变换($\mathcal{A}^2 = \mathcal{I}$), 则 V 可以分解成特征子空间的直和, 试分解之.

自测题八

一、(20分) 实数域 \mathbb{R} 上次数不超过2的多项式与零多项式构成的线性空间 $P[t]_2$ 的基为:

$f_1(t) = 1 - t, f_2(t) = 1 + t^2, f_3(t) = t + 2t^2$, 定义线性变换

$$\mathcal{A}[f_1(t)] = 2 + t^2, \quad \mathcal{A}[f_2(t)] = t, \quad \mathcal{A}[f_3(t)] = 1 + t + t^2,$$

(1) 求 \mathcal{A} 在已知基下的矩阵 \mathbf{A} ;

(2) 设 $f(t) = 1 + 2t + 3t^2$, 求 $\mathcal{A}[f(t)]$;

(3) 在 $P[t]_2$ 中定义内积 $(f, g) = \int_0^1 f(t)g(t)dt$, 求基 $1, t, t^2$ 的度量矩阵.

二、(15分)

(1) 设 \mathbf{A} 为 n 阶方阵, \mathbf{A} 的 n 个特征值为 $2, 4, \dots, 2n$, 试求 $|\mathbf{A} - 3\mathbf{I}|$ 的值(\mathbf{I} 为 n 阶单位矩阵);

(2) 设

$$\mathbf{A} = \begin{pmatrix} 8 & -3 & 6 \\ 3 & -2 & 0 \\ -4 & 2 & -2 \end{pmatrix},$$

求 \mathbf{A} 的若尔当标准形.

三、(15分) 已知

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}.$$

(1) 求 $e^{\mathbf{A}t}$;

(2) 用矩阵函数方法求微分方程

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t)$$

满足初始条件 $\mathbf{x}(0) = (1, 0, 0, -1)^T$ 的解.

四、(15分) 设 $\mathbf{A} = \begin{pmatrix} 0 & -2 & -2 \\ -2 & 3 & -1 \\ -2 & -1 & 3 \end{pmatrix}, \mathbf{B} = \begin{bmatrix} 4 & 5 \\ 0 & 3 \end{bmatrix}$.

(1) 求 $\|\mathbf{A}\|_\infty, \rho(\mathbf{A})$;

(2) 求 $\mathbf{A} \otimes \mathbf{B}$ 的秩和所有特征根;

(3) 求 $\mathbf{B}_i^-, \mathbf{B}^+$.

五、(10分) 用吉文斯变换求 $\mathbf{A} = \begin{pmatrix} 3 & 5 & 5 \\ 0 & 3 & 4 \\ 4 & 0 & 5 \end{pmatrix}$ 的 \mathbf{QR} 分解.

六、(15分) 证明题:

(1) 设 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 为对称正定矩阵, 对于 \mathbb{R}^n 中的列向量 $\boldsymbol{\alpha}$, 定义实数 $\|\boldsymbol{\alpha}\|_{\mathbf{A}} = \sqrt{\boldsymbol{\alpha}^T \mathbf{A} \boldsymbol{\alpha}}$, 验证 $\|\boldsymbol{\alpha}\|_{\mathbf{A}}$ 是 \mathbb{R}^n 中

的向量范数:

(2) 设 $A \in \mathbb{R}^{m \times n}$, 证明 $A^T(AA^T)^{-1}$ 是 A 的右逆;

(3) 设 A 为正交矩阵, 若 $|A| = -1$, 证明 A 一定有特征根 -1 .

七、(10分) 利用特征多项式及凯莱-哈密顿定理, 证明: 任意可逆矩阵 A 的逆阵 A^{-1} 都可以表示为 A 的多项式.

自测题九

一、(5分) 设 V 是有序实数对组成的集合: $V = \{(a, b) | a, b \in \mathbb{R}\}$, 加法和数乘运算定义为

$$(a_1, b_1) \oplus (a_2, b_2) = (a_1, b_1), \quad k \circ (a, b) = (ka, kb), \quad k \in \mathbb{R}.$$

问: V 关于运算 \oplus, \circ 是否构成 \mathbb{R} 上的线性空间? 并说明理由.

二、(15分) 设 V 表示数域 P 上 2 阶矩阵全体组成的线性空间, 定义 V 的一个变换 \mathcal{A} 如下:

$$\mathcal{A}(X) = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} X, \quad X \in V.$$

(1) 证明: \mathcal{A} 是线性变换;

(2) 求 \mathcal{A} 在基 $E_{11}, E_{12}, E_{21}, E_{22}$ 下的矩阵;

(3) 求 \mathcal{A} 的核 $\ker \mathcal{A}$, 给出 $\ker \mathcal{A}$ 的维数及一组基.

三、(10分) 已知 $\mathbb{C}^{n \times n}$ 中的两种矩阵范数 $\|\cdot\|_a$ 与 $\|\cdot\|_b$, 对于 $A \in \mathbb{C}^{n \times n}$, 验证 $\|A\| = \|A\|_a + \|A^H\|_b$ 是 $\mathbb{C}^{n \times n}$ 中的矩阵范数.

四、(15分) 设 $A = \begin{pmatrix} 1 & 0 & -1 & 1 \\ 1 & -2 & -3 & -1 \\ -1 & 4 & 5 & 3 \end{pmatrix}$, $b = \begin{pmatrix} 3 \\ 3 \\ -3 \end{pmatrix}$. 求:

(1) A 的最大秩分解;

(2) A^+ ;

(3) 用广义逆矩阵方法判断方程组 $Ax = b$ 是否有解?

五、(15分) 设 $A = \begin{pmatrix} 3 & 1 & 0 \\ -4 & -1 & 0 \\ 4 & -8 & -2 \end{pmatrix}$, $B = \begin{bmatrix} 3 & 4 \\ 5 & 2 \end{bmatrix}$, 求

(1) $\text{rank}(A \otimes B)$, $A \otimes B$ 的所有特征值, 及 $|A \otimes B|$;

(2) $f(B) = B^2 - B + 3I$ 的特征值.

六、(10分) 用豪斯霍尔德变换求 $A = \begin{pmatrix} 1 & 1 & 3 \\ 2 & 0 & 2 \\ 2 & 1 & 1 \end{pmatrix}$ 的 QR 分解.

七、(15分)

(1) 设 W 为 n 维欧氏空间 V 的 $n-1$ 维子空间, 且 V 中非零向量 α 与 W 正交, 即 $(\alpha, W) = 0$. 若 \mathcal{A} 为 V 的一个线性变换, 对 $\forall \xi \in W$, 使 $\mathcal{A}(\xi) = \xi$ 而 $\mathcal{A}(\alpha) = -\alpha$, 证明 \mathcal{A} 为 V 的一个正交变换.

(2) 设 \mathcal{A} 是欧氏空间 V^n 中的正交变换, 它在标准正交基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 下的矩阵为 A , 且 A 的特征值 $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 都是实数. 证明: 该正交变换 \mathcal{A} 也是对称变换.

八、(15分) 证明题

(1) 设 n 阶方阵 A 的行列式为 d , 而 $I - A$ 的特征根的绝对值都小于 1. 证明: $0 < |d| < 2^n$, 其中 $|d|$ 表示 d 的绝对值;

(2) 设 A 是一个三阶正交方阵, 且 $|A| = 1$, 证明: 存在实数 t , $-1 \leq t \leq 3$, 使 $A^3 - tA^2 + tA - I = 0$.

自测题十

一、(12分) (1) 试求数域 P 上线性空间

$$V = \left\{ \mathbf{x} = (x_1, x_2, x_3, x_4)^T \mid \mathbf{Ax} = \mathbf{0}, \mathbf{A} = \begin{pmatrix} 3 & 2 & -5 & 4 \\ 3 & -1 & 3 & -3 \\ 3 & 5 & -13 & 11 \end{pmatrix} \right\}$$

的一组基及维数.

(2) 设 $P^{n \times n}$ 为数域 P 上全体 n 阶方阵所构成的空间, 记 V_1 为 P 上全体 n 阶对称方阵构成的子空间, V_2 为 P 上一切 n 阶反对称方阵构成的子空间. 试证: $P^{n \times n} = V_1 \oplus V_2$.

二、(14分) 设 $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3, \boldsymbol{\varepsilon}_4$ 是四维线性空间 V 的基, 线性变换 \mathcal{A} 在这组基下的矩阵为

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

求:

(1) \mathcal{A} 的核与维数;

(2) 在 \mathcal{A} 的核中选取基, 将它扩充为 V 的一组基, 并求 \mathcal{A} 在这组基下的矩阵.

三、(12分) 设 $\mathbf{A} \in \mathbb{R}^{m \times n}$, 且 $\text{rank} \mathbf{A} = n$, 行向量 $\boldsymbol{\alpha} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, $\boldsymbol{\beta} = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$, 在 \mathbb{R}^n 中定义实函数 $(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{\alpha} \mathbf{A}^T \mathbf{A} \boldsymbol{\beta}^T$.

(1) 证明: 在该实函数定义下 \mathbb{R}^n 成为欧氏空间;

(2) 求 \mathbb{R}^n 中自然基 $\boldsymbol{\varepsilon}_1 = (1, 0, \dots, 0), \dots, \boldsymbol{\varepsilon}_n = (0, \dots, 0, 1)$ 的度量矩阵.

四、(15分)

(1) 设 $\mathbf{B} = \lim_{n \rightarrow \infty} \mathbf{A}^n$, 证明 \mathbf{B} 是幂等阵;

(2) 设 $\mathbf{A} = \begin{pmatrix} 2 & -\frac{1}{2} \\ 2 & 0 \end{pmatrix}$, 求 $\sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{2^k}$;

(3) 设 $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$, 求 \mathbf{A}^+ .

五、(13分) 求线性方程组 $\begin{cases} 2x_1 + 4x_2 + x_3 + x_4 = 10, \\ x_1 + 2x_2 - x_3 + 2x_4 = 6, \\ -x_1 - 2x_2 - 2x_3 + x_4 = -7, \end{cases}$ 的全部最小二乘解和极小范数最小二乘解.

六、(12分)

(1) 设 $\boldsymbol{\alpha} = (1, 1, \dots, 1)^T \in \mathbb{C}^n, \forall \mathbf{x} \in \mathbb{C}^n$, 定义向量范数 $\|\mathbf{x}\| = \|\boldsymbol{\alpha} \mathbf{x}^T\|_F$, 试写出这种向量范数与向量 2 范数之间的关系;

(2) 证明: 设 $\mathbf{A} \in \mathbb{C}^{m \times m}, \mathbf{B} \in \mathbb{C}^{n \times n}$, 则 $\det(\mathbf{A} \otimes \mathbf{B}) = (\det \mathbf{A})^n (\det \mathbf{B})^m$.

七、(10分) 设线性空间 V^3 的线性变换 \mathcal{A} 在基 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$ 下的矩阵 $\mathbf{A} = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}$, 证明: $W =$

$L(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_3 - \boldsymbol{\alpha}_1)$ 是 \mathcal{A} 的不变子空间.

八、(12分) 设 \mathcal{A} 为线性空间 V 上的线性变换.

(1) 证明: $\dim(\mathcal{A}(V)) + \dim(\mathcal{A}^{-1}(\theta)) = \dim V$;

(2) 是否有 $\mathcal{A}(V) + \mathcal{A}^{-1}(\theta) = V$? 若有, 给出证明; 若没有, 给出反例.

自测题十一

一、(10分)

设 $\mathbb{R}^{2 \times 2}$ 的一组基为 $\mathbf{A}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \mathbf{A}_3 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \mathbf{A}_4 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, 另一组基为 $\mathbf{B}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{B}_2 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \mathbf{B}_3 = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}, \mathbf{B}_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$

(1) 求从前基到后基的过渡矩阵;

(2) 求从后基到前基的过渡矩阵.

二、填空(每题4分,共20分)

(1) 设 $\mathbf{A} = \begin{bmatrix} 2 & -3 & 8 & 2 \\ 2 & 12 & -2 & 12 \\ 1 & 3 & 1 & 4 \end{bmatrix}$, 则 \mathbf{A} 的值域 $R(\mathbf{A}) = \{ \mathbf{y} | \mathbf{y} = \mathbf{A}\mathbf{x}, \mathbf{x} \in \mathbb{R}^4 \}$ 的维数 $\dim R(\mathbf{A}) =$ _____.

(2) 设 $\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 2 & -1 & 0 & 1 \\ 1 & -1 & 3 & 7 \end{bmatrix}$, V_1, V_2 分别为齐次线性方程组 $\mathbf{A}\mathbf{x} = \mathbf{0}, \mathbf{B}\mathbf{x} = \mathbf{0}$ 的解空间, 则 $\dim(V_1 \cap V_2) =$ _____.

(3) 设 $\mathbf{A}_n = \begin{bmatrix} \frac{n+(-1)^n}{n} & \left(1-\frac{1}{n}\right)^{\frac{1}{n}} \\ \frac{n+1}{3n} & \left(\frac{2n+1}{2n-1}\right)^n \end{bmatrix}$, 则 $\lim_{n \rightarrow \infty} \mathbf{A}_n =$ _____.

(4) 设 $\mathbf{A} = \begin{bmatrix} 2 & -1 & 3 \\ 1 & 2 & 1 \\ 2 & 0 & 2 \end{bmatrix}$, 则 \mathbf{A} 的 LDU 分解为 $\mathbf{A} =$ _____.

(5) 设 $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ -2 & 5 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 2 & 4 \\ 2 & 0 \end{bmatrix}$, 则 $\mathbf{A} \otimes \mathbf{B} =$ _____.

三、(15分) 已知 $\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$.

(1) 求 $\lambda \mathbf{I} - \mathbf{A}$ 的全部行列式因子, 不变因子及史密斯标准形;

(2) 求 \mathbf{A} 的若尔当标准形 \mathbf{J} 及过渡矩阵 \mathbf{P} , 使 $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{J}$.

四、(10分)

设函数矩阵

$$\mathbf{A}(t) = \begin{bmatrix} \sin t & -\cos t \\ \cos t & \sin t \end{bmatrix},$$

求 $\int_0^t \mathbf{A}(t) dt$ 和 $\left(\int_0^t \mathbf{A}(t) dt \right)'$.

五、(15分)

(1) 设 $f(\mathbf{X}) = \|\mathbf{X}\|_F^2 = \text{tr}(\mathbf{X}^T \mathbf{X})$, 其中 $\mathbf{X} = (x_{ij})_{m \times n} \in \mathbb{R}^{m \times n}$ 是矩阵变量, 求 $\frac{df}{d\mathbf{x}}$;

(2) 设 $\mathbf{A} = (a_{ij})_{m \times n} \in \mathbb{R}^{m \times n}, \mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$, 是向量变量, $F(\mathbf{x}) = \mathbf{A}\mathbf{x}$, 求 $\frac{dF}{d\mathbf{x}^T}$.

六、(15分) 已知微分方程组

$$\begin{cases} \frac{dx}{dt} = Ax \\ x(0) = x_0 \end{cases}, \quad \text{其中 } A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & -1 \\ 0 & 1 & 1 \end{bmatrix}, \quad x_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

(1) 求矩阵 A 的若尔当标准形 J 和可逆矩阵 P , 使 $P^{-1}AP = J$;

(2) 求矩阵 A 的最小多项式 $m_A(\lambda)$;

(3) 计算矩阵函数 e^{At} ;

(4) 求该微分方程组的解.

七、(15分)

(1) 设线性空间 V^n 中, 从基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 到基 $\beta_1, \beta_2, \dots, \beta_n$ 的过渡矩阵为 P , 证明: 有非零向量 $\alpha \in V^n$, 使 α 在两组基下坐标相同的充要条件是 1 为矩阵 P 的特征值.

(2) 证明: 与任意 n 阶方阵可交换的矩阵必是纯量矩阵.

自测题十二

一、(5分) 判断下列变换 \mathcal{A} 哪些是线性变换.

(1) \mathbb{R}^3 中, $X = (x_1, x_2, x_3) \in \mathbb{R}^3$, $\mathcal{A}[(x_1, x_2, x_3)] = (x_1 + 1, x_2 + 2, x_3 + x_1)$;

(2) $\mathbb{R}^{n \times n}$ 中, $A \in \mathbb{R}^{n \times n}$, $X \in \mathbb{R}^{n \times n}$, $\mathcal{A}(X) = AX + A$;

(3) $\mathbb{R}^{n \times n}$ 中, $A, B \in \mathbb{R}^{n \times n}$, $\forall X \in \mathbb{R}^{n \times n}$, $\mathcal{A}(X) = AX + XB$;

(4) $\mathbb{R}^{2 \times 2}$ 中, $A \in \mathbb{R}^{2 \times 2}$, $\mathcal{A}(A) = A^*$;

其中 A^* 是二阶方阵 A 的伴随矩阵.

二、(5分) 设矩阵空间 $\mathbb{R}^{2 \times 2}$ 的子空间 $V = \{X = (x_{ij}) \mid x_{11} + x_{12} + x_{21} = 0\}$ 构成线性空间(按通常的矩阵加法和数量乘法), 试求 V 的维数, 并写出 V 的一组基.

三、(15分) 设矩阵 $A = \begin{bmatrix} -1 & 0 & 1 \\ 1 & 2 & 0 \\ -4 & 0 & 3 \end{bmatrix}$,

(1) 求 A 的初等因子组;

(2) 求 A 的若尔当标准形 J ;

(3) 求可逆矩阵 P , 使得 $P^{-1}AP = J$;

(4) 求 A^k .

四、(15分) 设微分方程组

$$\begin{cases} \frac{dx}{dt} = Ax \\ x(0) = x_0 \end{cases}, \quad \text{其中 } A = \begin{bmatrix} 3 & 1 & -1 \\ -2 & 0 & 2 \\ -1 & -1 & 3 \end{bmatrix}, \quad x_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

(1) 求 A 的最小多项式 $m_A(\lambda)$; (2) 求 e^{At} ; (3) 求该方程组的解.

五、(20分)

(1) 设 A 是 n 阶埃尔米特正交矩阵, 定义

$$\|x\|_A = (x^H Ax)^{\frac{1}{2}}, \quad \forall x \in \mathbb{C}^n,$$

试证上述函数是向量范数;

(2) 设 $A \in \mathbb{R}^{n \times n}$, $\|\cdot\|$ 是 $\mathbb{R}^{n \times n}$ 上的一个算子范数, 如果 $\|A\| < 1$, 证明 $I - A$ 可逆, 且有 $\|(I - A)^{-1}\| \leq$

$$\frac{1}{1 - \|A\|}.$$

六、(10分)

判别矩阵级数 $\sum_{k=0}^{\infty} \begin{bmatrix} \frac{1}{6} & -\frac{1}{3} \\ -\frac{4}{3} & \frac{1}{6} \end{bmatrix}^k$ 的收敛性,若收敛,求其和.

七、(15分) 设矛盾方程 $\mathbf{Ax}=\mathbf{b}$, 其中

$$\mathbf{A} = \begin{bmatrix} 2 & 4 & 1 & 1 \\ 1 & 2 & -1 & 2 \\ -1 & -2 & -2 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 0 \\ 3 \end{bmatrix}.$$

(1) 求 \mathbf{A} 的满秩分解 $\mathbf{A}=\mathbf{FG}$;

(2) 由上面分解来计算 \mathbf{A}^+ ;

(3) 写出该方程组最小二乘解表达式,并求出极小范数最小二乘解.

八、(10分)

设 \mathcal{A} 是 \mathbb{C}^n 的线性变换, $W \neq \{0\}$ 是 \mathbb{C}^n 的一个子空间, 且是 \mathcal{A} 的不变子空间, 证明 W 中必有 \mathcal{A} 的特征向量.

自测题十三

一、(5分) 在 \mathbb{R}^3 中线性变换 \mathcal{A} 将基

$$\alpha_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \quad \alpha_2 = \begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix}, \quad \alpha_3 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

变为基

$$\beta_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad \beta_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \quad \beta_3 = \begin{bmatrix} 0 \\ 3 \\ -2 \end{bmatrix}.$$

(1) 求 \mathcal{A} 在基 $\alpha_1, \alpha_2, \alpha_3$ 下的矩阵表示;

(2) 求向量 $\xi=(1,2,3)^T$ 及 $\mathcal{A}(\xi)$ 在基 $\alpha_1, \alpha_2, \alpha_3$ 下的坐标;

(3) 求向量 $\xi=(1,2,3)^T$ 及 $\mathcal{A}(\xi)$ 在基 $\beta_1, \beta_2, \beta_3$ 下的坐标.

二、(5分) 设实数域上的多项式

$$p_1(x)=x^3+2x^2+2x+3, \quad p_2(x)=x^3+x^2+2x+3$$

$$p_3(x)=-x^3+x^2-4x-5, \quad p_4(x)=x^3-3x^2+6x+7$$

求线性空间 $W=\text{span}(p_1, p_2, p_3, p_4)$ 的一组基和维数.

三、(15分) 计算:

(1) 已知 \mathbf{A} 可逆, 求 $\int_0^1 e^{\mathbf{A}t} dt$ (用矩阵 \mathbf{A} 或其逆矩阵表示);

(2) 设 $\alpha=(a_1, a_2, a_3, a_4)^T$ 是给定的常向量, $\mathbf{X}=(x_{ij})_{2 \times 4}$ 是矩阵变量, 求 $\frac{d(\mathbf{X}\alpha)^T}{d\mathbf{X}}$;

(3) 设 3 阶方阵 \mathbf{A} 的特征多项式为 $|\lambda\mathbf{I}-\mathbf{A}|=\lambda^2(\lambda-6)$, 且 \mathbf{A} 可对角化, 求 $\lim_{k \rightarrow \infty} \left(\frac{\mathbf{A}}{\rho(\mathbf{A})}\right)^k$.

四、(15分) 已知矛盾方程组 $\mathbf{AX}=\mathbf{b}$,

$$\begin{cases} x_1 + x_2 = 1, \\ x_1 + 2x_2 = 1, \\ 2x_1 + 3x_2 = 1. \end{cases}$$

- (1) 求 \mathbf{A} 的满秩分解 $\mathbf{A} = \mathbf{F}\mathbf{G}$;
 (2) 求 \mathbf{A} 的广义逆 \mathbf{A}^+ ;
 (3) 求该方程组的最小二乘解 \mathbf{X}_{LS} .

五、(15分) 利用 Gerschgorin 圆盘定理证明矩阵:

$$\mathbf{A} = \begin{bmatrix} 2 & \frac{1}{2} & \frac{1}{2^2} & \cdots & \frac{1}{2^{n-1}} \\ \frac{2}{3} & 4 & \frac{2}{3^2} & \cdots & \frac{2}{3^{n-1}} \\ \frac{3}{4} & \frac{3}{4^2} & 6 & \cdots & \frac{3}{4^{n-1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{n}{n+1} & \frac{n}{(n+1)^2} & \frac{n}{(n+1)^3} & \cdots & 2n \end{bmatrix}$$

- (1) 能与对角矩阵相似; (2) 特征值全为实数.

六、(15分)

(1) 设 $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \\ 0 & 2 \\ 1 & 0 \end{bmatrix}$, 求 \mathbf{A} 的奇异值分解;

(2) 求矩阵 $\mathbf{A} = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 1 & 2 & -1 & 1 \\ 2 & 2 & -2 & -1 \end{bmatrix}$ 的满秩分解.

七、(15分) 已知

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & -1 \\ 0 & 2 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

求 $e^{\mathbf{A}t}$, $\sin \mathbf{A}t$ 和 $\cos \mathbf{A}t$.

八、(15分) 设 \mathbf{A} 是可逆矩阵, $\frac{1}{\|\mathbf{A}^{-1}\|} = \alpha$, $\|\mathbf{B} - \mathbf{A}\| = \beta$ (这里矩阵范数都是算子范数), 如果 $\beta < \alpha$, 证明:

(1) \mathbf{B} 是可逆矩阵; (2) $\|\mathbf{B}^{-1}\| \leq \frac{1}{\alpha - \beta}$; (3) $\|\mathbf{B}^{-1} - \mathbf{A}^{-1}\| \leq \frac{\beta}{\alpha(\alpha - \beta)}$.

自测题十四

一、(20分) 填空题

(1) 已知矩阵 $\mathbf{A} = \begin{bmatrix} 1 & -1 & 1 \\ i & 2+i & 1+i \\ 2 & i & 1-i \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $i = \sqrt{-1}$, 则 $\|\mathbf{A}\|_1 =$ _____; $\|\mathbf{A}\|_\infty =$ _____;

$\|\mathbf{Ax}\|_1 =$ _____; $\|\mathbf{Ax}\|_\infty =$ _____.

(2) 已知 \mathcal{A} 是线性空间 V^3 上的线性变换, 且有

$$\mathcal{A}(\alpha_1) = \alpha_1 + 2\alpha_2 + \alpha_3, \quad \mathcal{A}(\alpha_2) = \alpha_2 - \alpha_3, \quad \mathcal{A}(\alpha_3) = \alpha_3$$

则变换 \mathcal{A} 在基 $\{\alpha_1, \alpha_2, \alpha_3\}$ 下的变换矩阵为 _____.

(3) 已知矩阵 A 的 LU 分解为 $\begin{bmatrix} 1 & & \\ 2 & 1 & \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 0 \\ & 3 & 6 \\ & & 1 \end{bmatrix}$, 则 A 的 LDU 分解为_____.

(4) 方阵 A 的谱半径 $\rho(A) < 1$ 时, 有 $\lim_{n \rightarrow \infty} A^n =$ _____.

(5) 设 $A_{3 \times 3}$ 的特征值为 $1, 2, 0$, 则矩阵 $\sin A$ 的特征值为_____.

(6) 已知正规矩阵 $A_{3 \times 3}$ 的特征值为 $\lambda_1 = -2, \lambda_2 = 1 + 2i, \lambda_3 = 1$, 则 A 的奇异值为_____.

(7) 如果 $A_{3 \times 3}$ 的最小多项式 $m_A(\lambda) = (\lambda - 2)(\lambda + 3)^2$, 则 A 的若尔当矩阵为_____.

(8) 设 $f(z) = \sqrt{z}, A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$, 则 $f(A) =$ _____.

二、(10分) 在欧氏空间 \mathbb{R}^3 中求一个正交变换 \mathcal{A} , 使得 $\mathcal{A}(\alpha_1) = e_3, \mathcal{A}(\alpha_2) = e_1$, 其中 $e_i (i=1, 2, 3)$ 为 \mathbb{R}^3 的基本单位向量(称自然基), 其中

$$\alpha_1 = \frac{1}{\sqrt{2}} (1, 0, 1)^T, \quad \alpha_2 = \frac{1}{\sqrt{2}} (1, 0, -1)^T.$$

三、(10分) 设 $\mathbb{R}^{2 \times 2}$ 中, $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, 定义变换

$$\mathcal{A}(X) = AX - XA, \quad \forall X \in \mathbb{R}^{2 \times 2}.$$

(1) 证明 \mathcal{A} 是 $\mathbb{R}^{2 \times 2}$ 的线性变换;

(2) 求 \mathcal{A} 在基 $E_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, E_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, E_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, E_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ 下的矩阵;

(3) 证明 0 是 \mathcal{A} 的特征值;

(4) 试讨论特征值为 0 的重数对于 a, b, c, d 的依赖关系.

四、(20分) (1) 证明过渡矩阵必是可逆矩阵;

(2) 若 n 阶方阵 A 既是正规矩阵, 又是幂零矩阵, 证明 $A = 0$.

五、(20分) 已知

$$A(t) = \begin{bmatrix} \sin t & -\cos t \\ e^t \cos t & e^t \sin t \end{bmatrix}$$

求 (1) $\lim_{t \rightarrow 0} A(t), \frac{dA(t)}{dt}, \frac{d^2 A(t)}{dt^2}, \frac{d|A(t)|}{dt}, \left| \frac{dA(t)}{dt} \right|$ 和 $\frac{dA^{-1}(t)}{dt}$.

(2) $\int A(t) dt, \int_0^1 A(t) dt$ 和 $\frac{d}{dx} \int_0^{x^2} A(t) dt$.

六、(10分) 讨论幂级数 $\sum_{k=1}^{\infty} \frac{1}{k} A^k$ 的敛散性, 其中 $A = \begin{bmatrix} -2 & 1 \\ -1 & 0 \end{bmatrix}$.

七、(10分) 设 $A_i \in \mathbb{R}^{m \times n}, b_i \in \mathbb{R}^m (i=1, 2, \dots, k), f(x) = \sum_{i=1}^k \|A_i x - b_i\|_2^2$.

(1) 求 $\frac{df}{dx}$ 和 $\frac{d}{dx^T} \left(\frac{df}{dx} \right)$;

(2) 如果 $x_0 \in \mathbb{R}^n$ 是 $f(x)$ 的极小值点, 证明 x_0 为下面方程组的解:

$$\left(\sum_{i=1}^k A_i^T A_i \right) x = \sum_{i=1}^k A_i^T b_i.$$

自测题十五

一、(10分) 填空

(1) \mathbb{R}^2 的两组基为 $\alpha_1 = (1, 2)^T, \alpha_2 = (1, 1)^T$ 与 $\beta_1 = (1, 0)^T, \beta_2 = (1, -1)^T$, 设 \mathcal{A} 在 α_1, α_2 下的矩阵

为 $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, 求 \mathcal{A} 在 β_1, β_2 下的矩阵为_____.

(2) 矩阵 $A = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 1 & -1 \\ 1 & 0 & 2 & 1 \end{bmatrix}$ 的满秩分解表达式 $A=BC$, 其中 $B=$ _____, $C=$ _____.

(3) 已知 $A = \begin{bmatrix} 2 & 2 & 0 \\ 8 & 2 & a \\ 0 & 0 & 6 \end{bmatrix}$ 为单纯矩阵, 求 $a=$ _____.

(4) 已知矩阵 $A = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$, $\mathbb{C}^{2 \times 2}$ 的子集 $V = \{X | AX=0, X \in \mathbb{C}^2\}$, 求 V 的一组基_____ 以及 $\dim V=$ _____.

二、(10分) (1) 设 $X \in \mathbb{R}^{2 \times 2}$, 定义变换 $\mathcal{A}(X) = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} X$, 试证 \mathcal{A} 是 $\mathbb{R}^{2 \times 2}$ 上的线性变换;

(2) 设 $A, B \in \mathbb{R}^{2 \times 2}$, 记 $\text{tr}(A) = a_{11} + a_{22}$, 问函数 $\langle A, B \rangle = \text{tr}(A+B)$ 是否为内积, 为什么?

(3) 设 $f(t), g(t) \in C[0, 1]$, 定义内积 $(f, g) = \int_0^1 f(t)g(t)dt$, 取 $f(t) = t, g(t) = e^t$, 试具体写出 Cauchy-Schwarz 不等式的表达式.

三、(15分) \mathbb{R}^n 中, 设 n 维非零列向量为 α 和 β , 矩阵 $A = \alpha\beta^T$, 又已知 q_1, q_2, \dots, q_{n-1} 是 $n-1$ 个线性无关向量组, 且 $\beta^T q_i = 0, i=1, 2, \dots, n-1$, 证明:

(1) α 是矩阵 A 的一个特征向量, 并求 α 对应的特征值;

(2) 证明 A 可对角化, 求矩阵 P , 使 $P^{-1}AP$ 为对角阵.

四、(15分) (1) 设 A 是 n 阶矩阵, 对任意 $0 \neq x \in \mathbb{R}^n$ 均有 $Ax \neq x$, 证明: $I-A$ 可逆并求其逆.

(2) 设 $A \in \mathbb{C}^{n \times n}$, 证明:

$$\|e^A\| \leq e^{\|A\|} \quad \text{及} \quad |e^A| = e^{\text{tr}(A)}.$$

五、(15分) 已知 3 阶矩阵

$$A = \begin{bmatrix} 1 & -1 & 1 \\ x & 4 & y \\ -3 & -3 & 5 \end{bmatrix}$$

的二重特征值 $\lambda=2$ 对应两个线性无关的特征向量.

(1) 求 x, y ;

(2) 求可逆矩阵 P , 使得 $P^{-1}AP$ 为对角矩阵;

(3) 求 A 的谱分解表达式.

六、(15分) (1) 设 A 和 B 是 n 阶正定埃尔米特矩阵, 则 $|\lambda B - A| = 0$ 的根全是正数, 试证明之.

(2) 已知矩阵

$$A = \begin{bmatrix} \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \\ \frac{1}{4} & \frac{2}{5} & \frac{1}{6} & \frac{1}{7} \\ \frac{1}{4} & \frac{1}{5} & \frac{3}{6} & \frac{1}{7} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{3}{7} \end{bmatrix},$$

试证: $\rho(A) \leq 1$.

七、(10分) 设

$$\mathbf{A} = \begin{bmatrix} 9 & 1 & -2 & 1 \\ 0 & 8 & 1 & 1 \\ -1 & 0 & 4 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

- (1) 写出 \mathbf{A} 的 4 个盖尔圆;
 (2) 应用盖尔圆定理证明矩阵 \mathbf{A} 至少有两个实特征值.

八、(10分) 假设三维线性空间 V 上的线性变换 \mathcal{A} 在 V 的基 $\alpha_1, \alpha_2, \alpha_3$ 下的矩阵为 $\mathbf{J} = \begin{bmatrix} 2 & a & c \\ 0 & 2 & b \\ 0 & 0 & -1 \end{bmatrix}$,

问: 当 a, b, c, d 满足什么条件时, 存在 V 的一组基, 使得 \mathcal{A} 的矩阵是

$$\mathbf{K} = \begin{bmatrix} 2 & 0 & 0 \\ 2 & d & 0 \\ 2 & 2 & 2 \end{bmatrix} ?$$

参 考 文 献

- [1] 蒋尔雄,高坤敏,吴景琨.线性代数[M].北京:人民教育出版社,1978.
- [2] 程云鹏.矩阵论[M].西安:西北工业大学出版社,1989.
- [3] 黄有度,等.矩阵论及其应用[M].合肥:中国科学技术大学出版社,1995.
- [4] 戴华.矩阵论[M].北京:科学出版社,2001.
- [5] 张远达.线性代数原理[M].上海:上海教育出版社,1980.
- [6] 许以超.线性代数与矩阵论[M].2版.北京:高等教育出版社,2008.
- [7] 张跃辉.矩阵理论与应用[M].北京:科学出版社,2011.
- [8] 张贤达.矩阵分析与应用[M].北京:清华大学出版社,2004.
- [9] 詹兴致.矩阵论[M].北京:高等教育出版社,2008.
- [10] 苏育才,等.矩阵理论[M].北京:科学出版社,2006.
- [11] [美]Horn R A,等.矩阵分析[M].杨奇,译.北京:机械工业出版社,2005.
- [12] Houck D J, Paul M E. On a theorem of de Bruijn and Erdős Linear Algebra Appl, 23(1979), 157-165.
- [13] 李明.线性代数中矩阵的应用研究[J].常州工学院学报,2011, 24(314).
- [14] 李剑,等.矩阵分析与应用习题解答[M].北京:清华大学出版社,2007.
- [15] Arnold V I. Mathematical Methods of Classical Mechanics[M]. New York: Springer,1978.
- [16] 须田信英,等.自动控制中的矩阵理论[M].曹长修,译.北京:科学出版社,1979.
- [17] 徐树方.控制论中的矩阵计算[M].北京:高等教育出版社,2011.
- [18] 钱吉林,等.矩阵及其广义逆[M].武汉:华中师范大学出版社,1998.
- [19] 陈景良,陈向晖.特殊矩阵[M].北京:清华大学出版社,2001.
- [20] 张凯院.矩阵论典型题解析及自测试题[M].西安:西北工业大学出版社,2001.
- [21] 林升旭.矩阵论学习辅导与典型题解析[M].武汉:华中科技大学出版社,2003.
- [22] 魏丰,等.矩阵分析指导[M].北京:北京理工大学出版社,2005.

清华大学出版社数字出版网站

WQBook  
www.wqbook.com

ISBN 978-7-302-33269-5



9 787302 332695 >

定价：45.00元