

北京航空航天大学研究生教材

矩阵基本理论与应用

王 磊 编著

北京航空航天大学出版社

内 容 简 介

本书共分五章,较全面系统地介绍了矩阵的基本理论、方法和典型应用。第1、2章是线性代数的基础理论,主要介绍线性空间与内积空间、线性映射与线性变换、矩阵与特征值等基本概念和性质。第3章矩阵分解,主要介绍九种典型的矩阵分解,这些内容是矩阵理论研究、计算及其应用中不可缺少的工具和手段。第4章矩阵分析,介绍了向量范数与矩阵范数、矩阵幂级数、矩阵函数及其应用等。第5章广义逆矩阵,介绍常用的四种广义逆及其在解线性方程组等方面的应用。

本书为北京航空航天大学电子信息类学术博士研究生的通用教材,也可作为电子信息类学术硕士和工程硕士的教材。

矩阵基本理论与应用

王 磊 编著

责任编辑 王 实

*

北京航空航天大学出版社出版发行

北京市海淀区学院路37号(邮编100191) <http://www.buaapress.com.cn>

发行部电话:(010)82317024 传真:(010)82328026

读者信箱:emsbook@buaacm.com.cn 邮购电话:(010)82316936

印刷有限公司印装 各地书店经销

*

开本:710×1 000 1/16 印张: 字数: 千字

2020年8月第1版 2020年8月第1次印刷 印数: 册

ISBN 978-7-5124- - 定价: .00元

若本书有倒页、脱页、缺页等印装质量问题,请与本社发行部联系调换。联系电话:(010)82317024

前言

矩阵理论是一门理论上高度抽象、又极具应用价值的数学基础课程。它不仅是数学的一个重要分支,而且已成为现代各种科学技术领域,如自动化与人工智能、计算机科学、力学与机械工程、管理科学与工程等处理大量有限维空间形式与数量关系的基础工具。因此,学习和掌握矩阵的基本理论和方法,对工科研究生是必不可少的。

本书的主要内容作为北京航空航天大学工科研究生的基础理论核心课程“矩阵理论”已讲授多年。该课程在讲授过程中,无论授课对象如何,课程设置和讲授内容基本保持不变,因此具有极强的“鲁棒性”。编者于2017年起独立承担自动化科学与电气工程学院的“矩阵理论”课程的教学,经过六轮次的课程探索与实践,对课程内容作了精心安排,并加入了诸多面向信息类专业的应用案例,深受广大学生喜爱。在编写本书的过程中,编者力求使其内容既具有一定的理论体系和理论深度,又注重深入浅出、简洁易懂,对大量较为抽象的概念赋予简明的几何意义或具体实例,帮助学生理解并掌握矩阵相关的基本概念、理论和方法,实现培养学生的数学思想与思维、应用与创新能力的目的。

本书共分5章,比较全面、系统地介绍了矩阵的基本理论、方法和典型应用。第1章线性代数引论是本书的基础理论,主要介绍线性空间与内积空间。第2章线性映射与矩阵是本书的理论核心,阐释了矩阵是有限维线性空间的线性变换的一种表达形式,并赋予矩阵以几何直观,使得在线性代数课程中难以理解的诸多疑问,如相似矩阵具有相同的特征值、矩阵乘法往往不可以交换等豁然开朗。第3章矩阵分解是应用最广的矩阵理论和方法,主要介绍广泛应用的满秩分解、三角分解、QR分解、Schur分解、谱分解和奇异值分解;此外,还介绍了相抵分解、对角化分解和Jordan分解问题,这些内容是矩阵理论研究、矩阵计算及应用不可缺少的工具和方法。第4章矩阵分析是矩阵理论对数学分析课程的推广,介绍了

范数理论、矩阵级数理论、矩阵函数与函数矩阵,这些内容已广泛应用于诸多理论分析和工程问题。第5章广义逆矩阵是矩阵理论中比较现代的部分,主要介绍常见的4种广义逆矩阵及其在线性方程组中的应用,并分析了正交投影矩阵与广义逆矩阵之间的联系。

特别地,每章都单独用一节来展开矩阵理论研究与应用。比如,第1章讨论了多项式插值与多项式空间的基之间的关系;第2章研究了图的矩阵表示问题;第3章将诸多应用如线性方程组与线性常微分方程组求解、矩阵特征值与逆矩阵的计算、正交投影与图像压缩等融入各种矩阵分解中;第4章研究了主元分析法对数据降维的基本原理;第5章讨论了广义逆矩阵在区间线性规划中的应用问题。这些内容具有较强的针对性和应用性,为进一步理解矩阵的基本理论与方法,开展研究和解决工程问题提供了重要案例。

在本书编写过程中,考虑到授课的实际情况,将诸如投影与最佳逼近、Hermite 矩阵、矩阵的 Kronecker 积、线性矩阵方程等问题以片段或应用的形式分散在一些章节中,希望读者在阅读过程中注意总结。

编者特别感谢北京航空航天大学数学科学学院杨小远教授和王永革副教授、自动化科学与电气工程学院林岩教授和秦世引教授、兰州大学数学与统计学院孙春友教授、中国科学院数学与系统科学学院李婵颖研究员,他们给予编者很大的鼓励和支持,并仔细审阅了全部书稿,提出了很多宝贵意见和建议。

编者感谢北京化工大学经济管理学院张强副教授、北京航空航天大学自动化科学与电气工程学院铁林副教授和刘克新副教授,他们仔细审阅了全部书稿,并提出了很多有益的建议。感谢王宏伟、张书源、高燕姗、秦鸿宇、丁祥涛、梁锦威、田挚昆、段安娜等研究生协助编辑和校对书稿。

由于编者水平有限,书中难免存在谬误与不当之处,敬请批评、指正。来信请寄:北京市海淀区学院路37号北京航空航天大学自动化科学与电气工程学院 王磊。邮编:100191;电子邮箱:lwang@buaa.edu.cn。

编 者

2020年6月

主要符号表

0	数字零
$\mathbf{0}$	零向量
\mathbf{O}	零矩阵
\emptyset	空集
F	数域 F
\mathbb{R}	实数域
\mathbb{C}	复数域
\mathbb{Q}	有理数域
\mathbb{N}	自然数集
\mathbb{R}^+	正实数集
F^n	数域 F 上 n 维有序数组构成的线性空间
\mathbb{R}^n	实数域上 n 维有序数组构成的线性空间(n 维欧氏空间)
\mathbb{C}^n	复数域上 n 维有序数组构成的线性空间(n 维西空间)
F^n	数域 F 上 n 维有序数组构成的线性空间
$F^{m \times n}$	数域 F 上全体 $m \times n$ 阶矩阵构成的线性空间
$\mathbb{R}^{m \times n}$	全体 $m \times n$ 阶实矩阵构成的线性空间
$\mathbb{C}^{m \times n}$	全体 $m \times n$ 阶复矩阵构成的线性空间
$\mathbb{C}_r^{m \times n}$	全体秩为 r 的 $m \times n$ 阶复矩阵集合
$P_n(x)$	定义在数域 F 上的多项式空间
$C[a, b]$	区间 $[a, b]$ 上全体实变量连续函数构成的线性空间
$(V, \ \cdot\)$	赋范线性空间 V
$\dim(W)$	线性空间 W 的维数
$R(f)$	映射 f 的值域
$R(A)$	矩阵 A 的列空间
$N(A)$	矩阵 A 的零空间
$R(T)$	线性映射 T 的像空间
$N(T)$	线性映射 T 的核空间
$V \cap W$	子空间 V 与 W 的交空间
$V + W$	子空间 V 与 W 的和空间
$V \dot{+} W$	子空间 V 与 W 的直和
$V \oplus W$	子空间 V 与 W 的正交直和

W^\perp	子空间 W 的正交补空间
$E(\lambda)$	特征值 λ 的特征子空间
$L(V, W)$	从线性空间 V 到 W 的线性映射空间
$L(V)$	定义在线性空间 V 上的线性变换空间
i	虚数单位, $i = \sqrt{-1}$
\bar{z}	复数 z 的共轭
$ z $	复数 z 的模
$\deg(p(x))$	一元多项式 $p(x)$ 的次数
$f_A(\lambda)$	矩阵 A 的特征多项式
$m_A(\lambda)$	矩阵 A 的最小多项式
α	列向量 α
$\{x_k\}$	向量序列 $x_1, x_2, \dots, x_k, \dots$
θ	线性空间的零向量
e_i	第 i 个分量为 1、其余为 0 的列向量
(α, β)	向量 α 与 β 的内积
$\langle \alpha, \beta \rangle$	向量 α 与 β 的夹角
$\text{span}(\alpha_1, \dots, \alpha_n)$	由向量 $\alpha_1, \dots, \alpha_n$ 张成的线性子空间
$\ \alpha\ $	向量 α 的长度
$\ \alpha\ _\sigma$	向量 α 的 σ 范数
$\alpha \perp \beta$	向量 α 与 β 正交
$\text{Proj}_V b$	向量 b 在线性空间 V 上的(正交)投影
$A(\lambda)$	λ 的矩阵
$\{A_k\}$	矩阵序列 $A_1, A_2, \dots, A_k, \dots$
\bar{A}	矩阵 A 的所有元素取共轭
A^T	矩阵 A 的转置
A^H	矩阵 A 的共轭转置
$\text{rank}(A)$	矩阵 A 的秩
$\text{tr}(A)$	方阵 A 的迹
A^*	矩阵 A 的伴随矩阵
$\text{vec}(A)$	矩阵 A 的列展开
$\rho(A)$	矩阵 A 的谱半径
A^{-1}	矩阵 A 的逆
A^+	矩阵 A 的伪逆
$A^-, A^{(1)}$	矩阵 A 的减号逆
$A_l^-, A^{(1,3)}$	矩阵 A 的最小二乘广义逆

$\mathbf{A}_m^-, \mathbf{A}^{(1,4)}$	矩阵 \mathbf{A} 的极小范数广义逆
$\mathbf{A}^{(i,j)}$	矩阵 \mathbf{A} 的 $\{i, j\}$ 广义逆
$\mathbf{A} \otimes \mathbf{B}$	矩阵 \mathbf{A} 与 \mathbf{B} 的直积
$\text{diag}(\lambda_1, \cdots, \lambda_n)$	对角线元素为 $\lambda_1, \cdots, \lambda_n$ 的 n 阶对角矩阵
$\text{diag}(\mathbf{A}_1, \cdots, \mathbf{A}_s)$	对角线元素为矩阵 $\mathbf{A}_1, \cdots, \mathbf{A}_s$ 的对角块矩阵
\mathbf{I}_r, \mathbf{I}	r 阶单位矩阵, 单位矩阵
\mathbf{E}_{ij}	在 (i, j) 处为 1、其余元素为零的矩阵
$ \mathbf{A} $	矩阵 \mathbf{A} 的行列式
$\ \mathbf{A}\ _\sigma$	矩阵 \mathbf{A} 的 σ 范数
$\mathbf{A}(\lambda) \cong \mathbf{B}(\lambda)$	$\mathbf{A}(\lambda)$ 与 $\mathbf{B}(\lambda)$ 相抵
\forall	任意
\Leftrightarrow	当且仅当
$\stackrel{\text{def}}{=}$	定义为
\max, \min	最大值, 最小值
s. t.	受约束于
$\operatorname{argmax} f(\mathbf{x})$	当 $f(\mathbf{x})$ 取最大值时, \mathbf{x} 的取值



第 1 章 线性代数引论	1
1.1 线性空间	1
1.2 线性子空间	5
1.3 基与坐标	9
1.4 内积空间.....	15
1.5 直和与投影.....	23
1.6 应用：多项式插值	31
本章习题	34
第 2 章 线性映射与矩阵	38
2.1 预备知识.....	38
2.2 线性映射.....	40
2.3 矩阵与同构.....	46
2.4 特征值与特征向量.....	55
2.5 酉变换与酉矩阵.....	62
2.6 应用：图的矩阵表示	67
本章习题	71
第 3 章 矩阵分解	74
3.1 相抵分解.....	74
3.2 满秩分解.....	76
3.3 三角分解.....	79
3.4 QR 分解	87
3.5 Schur 分解	91
3.6 对角化分解.....	96
3.7 谱分解	104
3.8 Jordan 分解	114
3.9 奇异值分解	126
本章习题.....	135
第 4 章 矩阵分析	140
4.1 向量范数	140

4.2	矩阵范数	147
4.3	相容范数	150
4.4	矩阵扰动分析	156
4.5	特征值估计	160
4.6	矩阵级数	167
4.7	矩阵函数	175
4.8	函数矩阵	189
4.9	应用：主元分析法	197
	本章习题.....	200
第 5 章	广义逆矩阵	203
5.1	基本概念	203
5.2	矩阵方程 $AXB=D$	204
5.3	减号逆	212
5.4	极小范数广义逆	216
5.5	最小二乘广义逆	220
5.6	加号逆	223
5.7	应用：区间线性规划	228
	本章习题.....	230
	参考文献.....	233
	汉英名称索引.....	235

第 1 章

线性代数引论

本章主要概述线性空间和内积空间的基础概念和基本理论. 这些概念是对常见的几何空间概念的推广与抽象. 本章内容是学习本书的基础.

1.1 线性空间

我们最熟悉的空间概念,莫过于生活在其间的三维空间. 实际上,三维空间是由无穷个点组成的,且点与点之间存在着相对位置关系. 为精确地刻画点与点之间的相对位置,需取定空间中的一点 O 作为起点,空间中的任一点 A 为终点,则以点 O 为起点、点 A 为终点的有向线段 \overrightarrow{OA} 表示空间中的一个向量 α ,其中向量 α 的长度(或模)和方向由有向线段 \overrightarrow{OA} 的长度和方向确定. 基于此,三维空间中的任一点与上述有向线段定义的向量之间存在着——对应关系. 此时的三维空间就是一个三维向量空间.

为方便地表示三维空间中的任一向量,过点 O 作三条互相垂直的数轴,分别记为 Ox (x 轴)、 Oy (y 轴)、 Oz (z 轴),三条数轴均以 O 为原点且具有相同的长度单位. 由此构建了一个标准的空间直角坐标系,记为 $Oxyz$. 在确定坐标系 $Oxyz$ 后,空间中的任一点 A 均可由一个三元有序数组唯一表示. 由此,我们建立了三维空间的点、向量与三元有序数组之间的一一对应关系,用三元数组表示三维空间的向量. 特别地,在空间直角坐标系 $Oxyz$ 下,无论是一个三元数组,还是空间中某一点的坐标,都常用 $(\alpha_1, \alpha_2, \alpha_3)$ 来表示. 同理, n 维空间中的任一向量可用一个 n 元数组 $(\alpha_1, \alpha_2, \dots, \alpha_n)$ 来表示. 当然,当 $n > 3$ 时, n 维空间没有直观的几何意义.

定义 1.1.1(n 维向量) 若 n 维向量写成

$$[\alpha_1, \alpha_2, \dots, \alpha_n]$$

的形式,则称为 n 维行向量;若 n 维向量写成

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$$

的形式,则称为 n 维列向量. 这 n 个数称为该向量的 n 个分量,其中 α_i 称为第 i 个分量.

注 1: 当 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是复数时, n 维向量称为 n 维复向量; 当 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是实数时, n 维向量称为 n 维实向量. n 维空间中的同一向量既可以写成行向量,又可以写成列向量. 在实际运算时,通常会根据运算的需要把 n 维向量写成行向量或列向量的形式.

显然, n 维行向量是一个 $1 \times n$ 的矩阵, n 维列向量则是一个 $n \times 1$ 的矩阵. 因此,可以用矩阵的运算来定义向量的运算.

设 $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]^T$ 和 $\beta = [\beta_1, \beta_2, \dots, \beta_n]^T, k \in \mathbb{R}$, 则有

(1) 加法运算

$$\alpha + \beta = \begin{bmatrix} \alpha_1 + \beta_1 \\ \alpha_2 + \beta_2 \\ \vdots \\ \alpha_n + \beta_n \end{bmatrix} \quad (1.1.1)$$

(2) 数乘运算

$$k\alpha = \begin{bmatrix} k\alpha_1 \\ k\alpha_2 \\ \vdots \\ k\alpha_n \end{bmatrix} \quad (1.1.2)$$

基于向量加法和数乘运算,对向量空间进行如下定义:

定义 1.1.2(向量空间) 设 V 是 n 维实向量的非空集合,若 V 对向量的加法和数乘两种运算都封闭,即对于任意向量 $\alpha, \beta \in V$ 和 $k \in \mathbb{R}$,都有 $\alpha + \beta \in V$ (V 对向量的加法封闭)和 $k\alpha \in V$ (V 对向量的数乘封闭),则称集合 V 为向量空间.

例 1.1.1 定义集合 $V_1 = \{[0, a_2, \dots, a_n]^T \mid a_2, \dots, a_n \in \mathbb{R}\}$, 对集合 V_1 中的任意向量 $\alpha = [0, \alpha_2, \dots, \alpha_n]^T$ 和 $\beta = [0, \beta_2, \dots, \beta_n]^T$, 以及任意 $k \in \mathbb{R}$, 有

$$\alpha + \beta = \begin{bmatrix} 0 \\ \alpha_2 + \beta_2 \\ \vdots \\ \alpha_n + \beta_n \end{bmatrix} \in V_1, \quad k\alpha = \begin{bmatrix} 0 \\ k\alpha_2 \\ \vdots \\ k\alpha_n \end{bmatrix} \in V_1$$

所以, V_1 对向量的加法和数乘运算封闭. 由向量空间定义知, V_1 是向量空间.

例 1.1.2 集合 $V_2 = \{[1, a_2, \dots, a_n]^T \mid a_2, \dots, a_n \in \mathbb{R}\}$, 对集合 V_2 中的任意向

量 $\alpha = [1, \alpha_2, \dots, \alpha_n]^T$ 和 $\beta = [1, \beta_2, \dots, \beta_n]^T$, 有

$$\alpha + \beta = \begin{bmatrix} 2 \\ \alpha_2 + \beta_2 \\ \vdots \\ \alpha_n + \beta_n \end{bmatrix} \notin V_2$$

所以, 集合 V_2 对向量的加法运算不封闭. 由向量空间定义可知, V_2 不是向量空间.

由定义 1.1.2 可知, n 维向量空间是一个集合, 集合中的元素都是 n 维实向量, 其定义两种基本运算——加法运算和数乘运算都在实数域 \mathbb{R} 内满足指定的规则. 现将向量空间中的实数域扩展为一般的数域, 将向量集合扩展为一般的非空集合, 将向量加法和数乘运算扩展为一般的加法和数乘运算, 由此可定义更一般的“向量空间”, 即线性空间.

定义 1.1.3 (数域) 设 F 是非空数集, 若 F 中任意两个数的和、差、积、商(除数不为 0)仍在该数集, 即对四则运算封闭, 则称该数集 F 为一个数域.

常见的有理数集 \mathbb{Q} 、实数集 \mathbb{R} 和复数集 \mathbb{C} 都是数域, 分别称为有理数域、实数域和复数域. 但数域并不限于它们. 例如, 数集 $\mathbb{Q}(\sqrt{3}) = \{a + b\sqrt{3} \mid a, b \in \mathbb{Q}\}$ 也是一个数域.

定义 1.1.4 (线性空间) 设 V 是一个非空集合, F 是一个数域. 在 V 上定义了一种代数运算, 称为加法, 记为“+”, 使得对 V 中任意两个元素 α, β 都有 V 中唯一的元素 $\alpha + \beta$ 与之对应, 称为 α 与 β 的和; 定义了 F 中的数与 V 中元素的一种代数运算, 称为数乘, 记为“ \cdot ”, 使得对 V 中任意元素 α 和 F 中任意元素 k 有 V 中唯一的元素 $k \cdot \alpha$ (常简写为 $k\alpha$) 与之对应, 称为 k 与 α 的积. 如果加法和数乘满足如下性质:

- (1) 交换律: $\alpha + \beta = \beta + \alpha$;
- (2) 结合律: $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$;
- (3) 存在零元素: 存在元素 $\theta \in V$, 满足 $\forall \alpha \in V, \alpha + \theta = \alpha$;
- (4) 存在负元素: $\forall \alpha \in V$ 存在一个元素, 记为 $-\alpha$, 满足 $\alpha + (-\alpha) = \theta$;
- (5) 数乘的结合律: $k(l\alpha) = (kl)\alpha$;
- (6) 数乘对加法的分配律: $k(\alpha + \beta) = k\alpha + k\beta$;
- (7) 数乘对数的加法的分配律: $(k + l)\alpha = k\alpha + l\alpha$;
- (8) 数乘的初始条件: $1 \cdot \alpha = \alpha$.

其中, α, β, γ 是 V 的任一元素, k 和 l 是数域 F 的任意数, 则称 V 是数域 F 上的线性空间, 记为 $(V, +, \cdot)$, V 中的元素称为向量, F 中的元素称为标量. 特别地, 当 $F = \mathbb{R}$ 时, 称 V 为实线性空间; 当 $F = \mathbb{C}$ 时, 称 V 为复线性空间.

注 2: 满足加法性质 (1) ~ (4) 的集合 V , 称为加群, 记为 $(V, +)$. 我们熟知的有理数集 \mathbb{Q} 、实数集 \mathbb{R} 和复数集 \mathbb{C} 等均是加群, 其中的加法都是指数的加法, 但自然数集合 \mathbb{N} 不是加群(为什么?). 在高等数学中, 闭区间 $[a, b]$ 上的连续函数全体 $C[a, b]$ 在函数加法下(即 $(f + g)(x) = f(x) + g(x)$)也构成一个加群.

注 3: 定义 1.1.4 中的“加法”与“数乘”均不限于普通加法与普通乘法.

例 1.1.3 设 $V=\mathbb{R}^+$, $F=\mathbb{R}$, 定义 V 中加法运算为 $x\oplus y=xy$; 定义 V 中元素与 F 中数的数乘为 $k\cdot x=x^k$, 其中, $x, y\in V, k\in F$. 证明 (V, \oplus, \cdot) 是实线性空间.

解: 对任意 $x, y\in V$ 和 $k\in F$, 有 $x\oplus y\in V$ 和 $k\cdot x\in V$. 现根据定义 1.1.4 一一验证线性空间所须满足的 8 条性质.

- (1) 交换律: $x\oplus y=xy=yx=y\oplus x$;
- (2) 结合律: $(x\oplus y)\oplus z=xyz=x\oplus(y\oplus z)$;
- (3) 存在零元素: 存在元素 $1\in V$ 使得 $\forall x\in V, x\oplus 1=x$;
- (4) 存在负元素: $\forall x\in V, \exists x^{-1}\in V$ 使得 $x\oplus x^{-1}=1$;
- (5) 数乘的结合律: $\forall k, l\in F, k\cdot(l\cdot x)=(kl)\cdot x=x^{kl}$;
- (6) 数乘对加法的分配律: $k\cdot(x\oplus y)=(k\cdot x)\oplus(k\cdot y)=x^k y^k$;
- (7) 数乘对数的加法的分配律: $(k+l)x=(k\cdot x)\oplus(l\cdot x)=x^{k+l}$;
- (8) 数乘的初始条件: $1\cdot x=x$.

因此, (V, \oplus, \cdot) 是实线性空间.

例 1.1.4 常见的线性空间包括以下例子:

- (1) 向量空间 \mathbb{R}^n ($n\geq 1$) 是线性空间的首个例子.
- (2) 设 V 为 \mathbb{C} 上所有 $m\times n$ 矩阵构成的集合, 即

$$V=\{(a_{ij})_{m\times n} \mid a_{ij}\in\mathbb{C}\}$$

在矩阵加法和数乘运算下, 集合 V 构成 \mathbb{C} 上的线性空间, 记为 $\mathbb{C}^{m\times n}$. 类似的可定义 $\mathbb{R}^{m\times n}$.

- (3) 正弦函数集合

$$S_{[x]}=\{a\sin(x+b), a, b\in\mathbb{R}\}$$

在函数加法和乘法运算下, 集合 $S_{[x]}$ 构成 \mathbb{R} 上的线性空间.

- (4) 一元多项式集合

$$P_n(x)=\left\{\sum_{i=0}^n a_i x^i \mid a_i\in\mathbb{C}\right\}$$

在多项式加法和数乘运算下, 集合 $P_n(x)$ 构成 \mathbb{C} 上的线性空间, 称为多项式空间.

(5) 分别定义在区间 $[a, b]$ 上全体多项式集合 S_1 , 全体可微函数集合 S_2 , 全体连续函数集合 S_3 , 全体可积函数集合 S_4 , 全体实函数集合 S_5 , 则 $S_1\subset S_2\subset S_3\subset S_4\subset S_5$, 且这 5 个集合均为 \mathbb{R} 上的线性空间.

(6) 设 $A\in\mathbb{C}^{m\times n}$, $x\in\mathbb{C}^n$, 齐次线性方程组 $Ax=0$ 的解集构成 \mathbb{C} 上的线性空间, 而非齐次线性方程组 $Ax=b$ 的解集则不构成 \mathbb{C} 上的线性空间.

(7) 设 S 是所有双边序列集合, 其中双边序列 $\{x_k \mid x_k\in\mathbb{R}, k=0, \pm 1, \pm 2, \dots\}$ 表示工程中由测量或采样得到的某一离散时间信号 (序列常采用行向量表示). 若序列 $\{y_k\}$ 是集合 S 中的元素, 定义 $\{x_k\}$ 与 $\{y_k\}$ 的和为 $\{x_k+y_k\}$, 数乘 $c\{x_k\}$ 为 $\{cx_k\}$, 则 S 是 \mathbb{R} 上的线性空间, 并称为离散时间信号空间.

注 4: 设 V 是数域 F 上的线性空间, 有

- (1) 零向量是唯一的;
- (2) 任一向量的负向量是唯一的;
- (3) 对任意 $k \in F$ 和 $\alpha \in V, 0\alpha = \theta, (-1)\alpha = -\alpha, k\theta = \theta$;
- (4) 若 $k\alpha = \theta$, 则 $k=0$ 或 $\alpha = \theta$.

注 5: 线性空间中的两种运算(加法和数乘)合称为线性运算. 若线性空间 V 的向量 α 可由 V 的一组向量 $\alpha_1, \dots, \alpha_n$ 通过线性运算获得, 即存在 $k_i \in F, i=1, \dots, n$, 满足

$$\alpha = k_1\alpha_1 + \dots + k_n\alpha_n$$

则称向量 α 是向量组 $\alpha_1, \dots, \alpha_n$ 的一个线性组合, 或者说, α 可由向量组 $\alpha_1, \dots, \alpha_n$ 线性表示.

例 1.1.5 试判断矩阵 A^2 是否可以由 A 和 I 线性表示, 其中

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

解: 由于

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

因此, 矩阵 A^2 可以由 A 和 I 线性表示.

【思考】 对一般的 n 阶复方阵 A , 它的幂矩阵 A^n 是否可由 $I, A, A^2, \dots, A^{n-1}$ 线性表示?

1.2 线性子空间

在很多实际问题中, 我们关心的仅仅是 \mathbb{R}^n (或 \mathbb{C}^n) 空间中适当向量所构成的子集. 此时, 无需根据线性空间定义判定该子集是否为线性空间, 只需验证子集中的线性运算是否封闭即可, 这就是子空间的判别法.

定义 1.2.1 (子空间) 设集合 V 是数域 F 上的线性空间, W 是 V 的一个非空子集. 若 W 的向量关于 V 的加法和数乘运算也构成 F 上的线性空间, 则称 W 是 V 的子空间.

例 1.2.1 取定矩阵 $A \in \mathbb{C}^{m \times n}$, 集合 $W = \{x \in \mathbb{C}^n \mid Ax = 0\}$ 是 \mathbb{C} 上的线性空间. 注意到集合 W 是 n 维向量空间 \mathbb{C}^n 的子集, 故 W 是 \mathbb{C}^n 的子空间.

例 1.2.2 仅由线性空间 V 的零向量组成的集合是 V 的一个子空间, 称为零子空间, 记为 $\{\theta\}$ (注意与零元素 θ 的区别); 线性空间 V 本身也是 V 的一个子空间. 子空间 V 和 $\{\theta\}$ 称为 V 的平凡子空间.

例 1.2.3 向量空间 \mathbb{R}^2 不是 \mathbb{R}^3 的子空间, 因为 \mathbb{R}^2 不是 \mathbb{R}^3 的子集 (\mathbb{R}^3 的向量是三元有序数组, 而 \mathbb{R}^2 中的向量为二元有序数组). 集合 $W = \{[\alpha_1, \alpha_2, 0]^T, \alpha_1, \alpha_2 \in \mathbb{R}\}$ 是

\mathbb{R}^3 的一个子集, 尽管它很像 \mathbb{R}^2 (见图 1.2.1), 但可以证明 W 是 \mathbb{R}^3 的一个子空间.

例 1.2.4 \mathbb{R}^3 中不通过原点的平面不是 \mathbb{R}^3 的子空间, 这是因为此平面不包含 \mathbb{R}^3 的零向量; 类似地, \mathbb{R}^2 中不通过原点的直线 L (见图 1.2.2) 也不是 \mathbb{R}^2 的子空间.

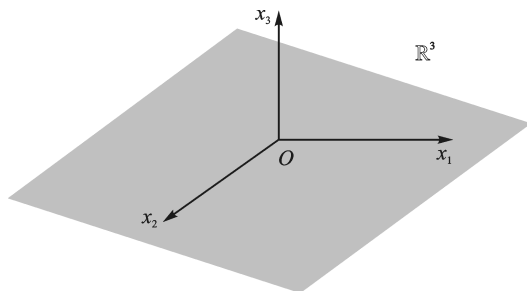


图 1.2.1 \mathbb{R}^3 的子空间: x_1x_2 平面

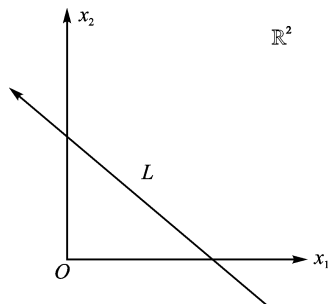


图 1.2.2 不是 \mathbb{R}^2 空间的直线 L

定理 1.2.1 (子空间判别法) 设集合 V 是数域 F 上的线性空间, W 是 V 的一个非空子集. 以下命题等价:

- (1) W 是 V 的子空间.
- (2) a. $\forall k \in F, \alpha \in W$, 有 $k\alpha \in W$;
b. $\forall \alpha, \beta \in W$, 有 $\alpha + \beta \in W$.
- (3) $\forall k, l \in F$ 和 $\alpha, \beta \in W$, 有 $k\alpha + l\beta \in W$.

例 1.2.1 (续) 证明集合 $W = \{x \in \mathbb{C}^n \mid Ax = 0\}$ 是 \mathbb{C} 上的线性空间.

解: 注意到集合 W 是 n 维向量空间 \mathbb{C}^n 的子集, 且对任意 $k, l \in F$ 和 $\alpha, \beta \in W$, 有

$$A(k\alpha + l\beta) = kA\alpha + lA\beta = 0$$

因此, $k\alpha + l\beta \in W$. 由定理 1.2.1 可知, W 是 V 的子空间.

注 1: 若 W 是 V 的子空间, 线性空间 V 的零向量也必在 W 中. 因此, 检验子空间最好的方法是首先观察零向量是否属于 W . 请参考例 1.2.4.

例 1.2.5 在数字信号处理中, 线性滤波器常用 n 阶线性差分方程描述

$$y_k = a_n u_{k+n} + a_{n-1} u_{k+n-1} + \cdots + a_1 u_{k+1} + a_0 u_k \quad (1.2.1)$$

式中: $k=0, \pm 1, \pm 2, \cdots$, $\{y_k\}$ 是输出信号, $\{u_k\}$ 是输入信号, $a_i \in \mathbb{R} (i=0, 1, \cdots, n)$. 如果 $\{y_k\}$ 是零序列, 则称差分方程 (1.2.1) 是齐次的.

考察齐次差分方程

$$u_{k+n} + a_{n-1} u_{k+n-1} + \cdots + a_1 u_{k+1} + a_0 u_k = 0 \quad (1.2.2)$$

定义方程 (1.2.2) 的解集为 \tilde{S} , 则 \tilde{S} 是非空集合 (零序列是方程 (1.2.2) 的一个解) 且它是信号空间 S (S 的定义见例 1.1.4) 的子集. 对任意输入信号 $\{u_k\}, \{v_k\} \in \tilde{S}$ 和 $\mu, \lambda \in \mathbb{R}$, $\{\mu u_k + \lambda v_k\}$ 也是差分方程 (1.2.2) 的解. 因此, 差分方程 (1.2.2) 的解集 \tilde{S} 是信号空间 S 的子空间.

例 1.2.6 设 W_1 和 W_2 是线性空间 V 的子空间, 定义 3 个集合

$$W_1 \cap W_2 = \{x \in V \mid x \in W_1 \text{ 且 } x \in W_2\}$$

$$W_1 \cup W_2 = \{x \in V \mid x \in W_1 \text{ 或 } x \in W_2\}$$

$$W_1 + W_2 = \{x_1 + x_2 \mid x_1 \in W_1, x_2 \in W_2\}$$

试判断它们是否是 V 的子空间.

解: 由于 W_1 和 W_2 是 V 的子空间, 则 $\theta \in W_1$ 且 $\theta \in W_2$, 故 $W_1 \cap W_2$ 是 V 的非空子集. 又知对任意 $k, l \in F$ 和 $\alpha, \beta \in W_1 \cap W_2$, 有

$$k\alpha + l\beta \in W_1 \text{ 且 } k\alpha + l\beta \in W_2$$

根据定理 1.2.1 可知, $W_1 \cap W_2$ 是 V 的子空间. 同理, $W_1 + W_2$ 也是 V 的子空间.

对于集合 $W_1 \cup W_2$ 而言, 若 $\alpha \in W_1, \beta \in W_2$, 则 $\alpha, \beta \in W_1 \cup W_2$. 而 $\alpha + \beta$ 却无法保证仍在集合 $W_1 \cup W_2$ 中. 因此, $W_1 \cup W_2$ 不是线性空间.

注 2: 关于例 1.2.6 的一个具体示例. 设 V 是 \mathbb{R} 上的三维向量空间, W_1 是 xOy 平面, W_2 是 Oz 轴, 则 $W_1 \cap W_2 = \{\theta\}$ 是零子空间, $W_1 + W_2 = V$ 是三维空间; 而 $W_1 \cup W_2$ 表示的是位于 xOy 平面和 Oz 轴所有向量的集合, 这显然不是一个线性空间.

定理 1.2.2 (交空间与和空间) 设 W_1 和 W_2 是数域 F 上线性空间 V 的子空间, 则集合 $W_1 \cap W_2 \stackrel{\text{def}}{=} \{\alpha \mid \alpha \in (W_1 \cap W_2)\}$ 和 $W_1 + W_2 \stackrel{\text{def}}{=} \{\alpha_1 + \alpha_2 \mid \alpha_1 \in W_1, \alpha_2 \in W_2\}$ 是 V 的子空间, 分别称为 W_1 与 W_2 的交(或交空间)与和(或和空间).

证明: 因为 $\theta \in W_1 \cap W_2$, 所以 $W_1 \cap W_2$ 非空. 对任意 $\alpha, \beta \in W_1 \cap W_2$, 有 $\alpha, \beta \in W_1$ 且 $\alpha, \beta \in W_2$.

又知 W_1 和 W_2 是 V 的子空间, 则对任意 $k \in F$ 有

$$\alpha + \beta \in W_1, \quad k\alpha \in W_1$$

$$\alpha + \beta \in W_2, \quad k\alpha \in W_2$$

从而有 $\alpha + \beta \in W_1 \cap W_2, k\alpha \in W_1 \cap W_2$. 因此, $W_1 \cap W_2$ 是 V 的子空间.

因为 $\theta \in W_1 + W_2$, 所以 $W_1 + W_2$ 非空. 对任意 $\alpha, \beta \in W_1 + W_2$, 必有

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

式中: $\alpha_1, \beta_1 \in W_1, \alpha_2, \beta_2 \in W_2$.

那么, 对任意 $k \in F$ 有

$$\alpha + \beta = (\alpha_1 + \beta_1) + (\alpha_2 + \beta_2) \in W_1 + W_2$$

$$k\alpha = k\alpha_1 + k\alpha_2 \in W_1 + W_2$$

因此, $W_1 + W_2$ 是 V 的子空间. 证毕.

注 3: $W_1 \cap W_2$ 是包含于 W_1 与 W_2 的最大子空间, $W_1 + W_2$ 是包含 W_1 与 W_2 的最小子空间. 同样地, 可定义多个子空间的交与和.

【思考】 如何衡量这里的“最大子空间”与“最小子空间”?

例 1.2.7 取 $V = \mathbb{R}^{n \times n}$, $W_1 = \{A \in V \mid A^T = A\}$, $W_2 = \{A \in V \mid A^T = -A\}$. 验证

W_1 和 W_2 是 V 的线性子空间, 且 $V=W_1+W_2$.

解: 对任意 $k, l \in \mathbb{R}$ 和 $\mathbf{A}, \mathbf{B} \in W_1$, 有 $(k\mathbf{A}+l\mathbf{B})^T = k\mathbf{A}^T + l\mathbf{B}^T$. 故 W_1 是 V 的线性子空间. 同理可证 W_2 是 V 的线性子空间.

又知对 V 的任一矩阵 \mathbf{A} 可分解为

$$\mathbf{A} = \frac{1}{2}(\mathbf{A}^T + \mathbf{A}) + \frac{1}{2}(-\mathbf{A}^T + \mathbf{A}) \quad (1.2.3)$$

式中: $\frac{1}{2}(\mathbf{A}^T + \mathbf{A}) \in W_1, \frac{1}{2}(-\mathbf{A}^T + \mathbf{A}) \in W_2$. 故 $V=W_1+W_2$.

例 1.2.7 表明任一实(复)方阵可分解为对称矩阵和反对称矩阵之和. 实际上, 该分解是唯一的(为什么?).

为刻画子空间, 令 $\alpha_1, \dots, \alpha_n$ 是线性空间 V 的一组向量, 并记

$$W = \{k_1\alpha_1 + \dots + k_n\alpha_n \mid k_i \in F, i=1, \dots, n\} \quad (1.2.4)$$

则有如下定理:

定理 1.2.3 若 $\alpha_1, \dots, \alpha_n$ 是线性空间 V 的一组向量, 由式(1.2.4)定义的集合 W 是 V 的一个线性子空间, 并称 W 是由向量组 $\alpha_1, \dots, \alpha_n$ 张成(或生成)的子空间, 记为 $\text{span}(\alpha_1, \dots, \alpha_n)$.

下面介绍两个与矩阵相关的重要子空间: 矩阵的零空间和列空间.

定义 1.2.2(矩阵零空间) 设矩阵 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 则矩阵 \mathbf{A} 的零空间(或核空间)定义为齐次线性方程组 $\mathbf{Ax}=\mathbf{0}$ 的解集, 记为

$$N(\mathbf{A}) = \{\mathbf{x} \in \mathbb{C}^n \mid \mathbf{Ax}=\mathbf{0}\}$$

定义 1.2.3(矩阵列空间) 设矩阵 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 则矩阵 \mathbf{A} 的列空间(或值空间)是由 \mathbf{A} 的列的所有线性组合组成的集合, 即

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

若将矩阵 \mathbf{A} 进行列分块, 并记为 $\mathbf{A}=[\alpha_1, \dots, \alpha_n]$, 则 \mathbf{A} 的列空间可表示为

$$R(\mathbf{A}) = \text{span}(\alpha_1, \dots, \alpha_n)$$

注 4: 矩阵 \mathbf{A} 的零空间是 \mathbb{C}^n 的子空间, 矩阵 \mathbf{A} 的列空间是 \mathbb{C}^m 的子空间. 矩阵 \mathbf{A} 的零空间是隐式定义的, 通常求解方程 $\mathbf{Ax}=\mathbf{0}$ 以获得它的显式表达. 与零空间不同, 矩阵的列空间是显式定义的.

例 1.2.8 求矩阵 $\mathbf{A} = \begin{bmatrix} 1 & 1 & 2 \\ -2 & 3 & 1 \end{bmatrix}$ 的零空间和列空间.

解: $R(\mathbf{A}) = \text{span}(\alpha_1, \alpha_2, \alpha_3)$, 其中 $\alpha_1 = [1, -2]^T, \alpha_2 = [1, 3]^T, \alpha_3 = [2, 1]^T$.

对矩阵 \mathbf{A} 进行初等变换得

$$\begin{bmatrix} 1 & 1 & 2 \\ -2 & 3 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

由此, $\mathbf{Ax}=\mathbf{0}$ 的一个基础解系为 $\beta = [1, 1, -1]^T$, 故 $N(\mathbf{A}) = \text{span}(\beta)$.

【思考】 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 其零空间和列空间有可能相同吗? 若这两个空间相同, 则

矩阵 A 具有何性质?

1.3 基与坐标

采用 $\text{span}(\alpha_1, \dots, \alpha_n)$ 描述线性(子)空间意味着已知该(子)空间的所有向量均可由向量组 $\alpha_1, \dots, \alpha_n$ 线性表示. 那么如何找到一组包含向量数目最小的向量组用于描述线性(子)空间? 这就涉及线性相关和线性无关这对概念了. 实际上, 线性空间中的线性相关和线性无关的定义与向量空间中的定义并无区别. 这里作简单回顾.

定义 1.3.1 (线性相关与线性无关) 设 V 是数域 F 上的线性空间, $\alpha_1, \dots, \alpha_n$ 是 V 中一组向量. 若向量方程

$$k_1 \alpha_1 + \dots + k_n \alpha_n = \theta, \quad k_1, \dots, k_n \in F$$

只有平凡解, 即 $k_1 = k_2 = \dots = k_n = 0$, 则称向量组 $\alpha_1, \dots, \alpha_n$ 线性无关; 否则称向量组 $\alpha_1, \dots, \alpha_n$ 线性相关.

从几何意义上看, 两个向量线性相关当且仅当它们落在通过原点的同一条直线上.

例 1.3.1 考察多项式空间 $P_1(x)$, 并令 $p_1(x) = 1, p_2(x) = x, p_3(x) = 4 - x$. 由于 $p_3 = 4p_1 - p_2$, 即 $4p_1 - p_2 - p_3 = 0$, 从而 p_1, p_2, p_3 是线性相关的.

例 1.3.2 设信号子空间 $\tilde{S} = \text{span}(\{x_k\}, \{y_k\}, \{z_k\})$, 其中 $x_k = (-1)^k, y_k = 2^k, z_k = 3^k$. 试判断信号 $\{x_k\}, \{y_k\}, \{z_k\}$ 是否线性无关.

解: 假设存在常数 l_1, l_2, l_3 对 $k = 0, \pm 1, \pm 2, \dots$ 均满足

$$l_1 x_k + l_2 y_k + l_3 z_k = 0$$

显然, 当且仅当 $l_1 = l_2 = l_3 = 0$ 时上式成立, 则信号 $\{x_k\}, \{y_k\}, \{z_k\}$ 线性无关.

实际上, 由于上式有 3 个未知数, 可建立如下方程组:

$$\begin{bmatrix} x_k & y_k & z_k \\ x_{k+1} & y_{k+1} & z_{k+1} \\ x_{k+2} & y_{k+2} & z_{k+2} \end{bmatrix} \begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix} = \mathbf{0}, \quad \forall k = 0, \pm 1, \pm 2, \dots$$

这个方程组的系数矩阵称为信号的 Casorati 矩阵, 其行列式称为 $\{x_k\}, \{y_k\}, \{z_k\}$ 的 Casorati 行列式. 显然, 如果对至少一个 k 值有 Casorati 行列式非零, 则说明这 3 个信号是线性无关的.

当 $x_k = (-1)^k, y_k = 2^k, z_k = 3^k$ 时, 取 $k = 0$ 时的 Casorati 矩阵为

$$C = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 2 & 3 \\ 1 & 4 & 9 \end{bmatrix}$$

该矩阵可逆. 因此, 信号 $\{x_k\}, \{y_k\}, \{z_k\}$ 线性无关.

注 1: 线性无关向量组的任一子集是线性无关的; 线性相关向量组的任一扩集仍

是线性相关的.

注 2: 单个零向量线性相关; 单个非零向量线性无关.

定理 1.3.1 设线性空间 V 的向量组 $\alpha_1, \dots, \alpha_n$ 线性无关, 而向量组 $\alpha_1, \dots, \alpha_n, \beta$ 线性相关 ($\beta \neq \theta$), 则 β 可由向量组 $\alpha_1, \dots, \alpha_n$ 线性表示且表示法唯一.

证明: 由向量组 $\alpha_1, \dots, \alpha_n, \beta$ 线性相关知, 必存在不全为零的数 k_1, \dots, k_n, k_{n+1} 使得

$$k_1 \alpha_1 + \dots + k_n \alpha_n + k_{n+1} \beta = \theta \quad (1.3.1)$$

若 $k_{n+1} = 0$, 则式 (1.3.1) 退化为

$$k_1 \alpha_1 + \dots + k_n \alpha_n = \theta \quad (1.3.2)$$

又知向量组 $\alpha_1, \dots, \alpha_n$ 线性无关, 则式 (1.3.2) 成立的充分必要条件是 $k_1 = \dots = k_n = 0$. 此时, 向量组 $\alpha_1, \dots, \alpha_n, \beta$ 线性无关, 这与已知条件矛盾. 故 $k_{n+1} \neq 0$, 即 β 可由向量组 $\alpha_1, \dots, \alpha_n$ 线性表示, 且有

$$\beta = -\frac{k_1}{k_{n+1}} \alpha_1 - \dots - \frac{k_n}{k_{n+1}} \alpha_n$$

下面证明唯一性. 假设向量 β 的表示法不唯一, 不妨设 β 有两种不同表示:

$$\beta = a_1 \alpha_1 + \dots + a_n \alpha_n = b_1 \alpha_1 + \dots + b_n \alpha_n \quad (1.3.3)$$

其中, a_i 和 b_i ($i=1, \dots, n$) 不同时为零.

对式 (1.3.3) 处理, 得

$$(a_1 - b_1) \alpha_1 + \dots + (a_n - b_n) \alpha_n = \theta$$

由于 $\alpha_1, \dots, \alpha_n$ 线性无关, 故 $a_i = b_i$ ($i=1, \dots, n$), 即 β 表示法唯一. 证毕.

在考察向量组时, 我们不仅关心向量之间是否线性无关, 而且关心这组向量中最多线性无关的向量个数. 为此, 引入定义:

定义 1.3.2 (极大线性无关组与秩) 设 $\alpha_1, \dots, \alpha_n$ 是线性空间 V 中的一组向量. 若 $\alpha_1, \dots, \alpha_n$ 中存在 r 个线性无关的向量 $\alpha_{i_1}, \dots, \alpha_{i_r}$ ($1 \leq i_j \leq n, j=1, \dots, r$), 并且 $\alpha_1, \dots, \alpha_n$ 中任一向量均可由向量组 $\alpha_{i_1}, \dots, \alpha_{i_r}$ 线性表示, 则称向量组 $\alpha_{i_1}, \dots, \alpha_{i_r}$ 为向量组 $\alpha_1, \dots, \alpha_n$ 的极大线性无关组, 数 r 称为向量组 $\alpha_1, \dots, \alpha_n$ 的秩, 记为 $\text{rank} [\alpha_1, \dots, \alpha_n] = r$.

注 3: 任一 $m \times n$ 阶矩阵均可看成包含 n 个 m 维列向量的向量组, 此时矩阵的 (列) 秩就是这一向量组的秩. 同理, 任一 $m \times n$ 阶矩阵也可看成包含 m 个 n 维行向量的向量组, 此时矩阵的 (行) 秩即是这一向量组的秩.

例 1.3.3 求向量组 $\alpha_1, \alpha_2, \alpha_3$ 的极大线性无关组, 其中

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

解: 令 $k_1 \alpha_1 + k_2 \alpha_2 = \theta$, 其中 $k_1, k_2 \in \mathbb{C}$. 将 α_1, α_2 代入向量方程得 $k_1 = 0, k_2 = 0$. 由此, α_1, α_2 线性无关. 又知 $\alpha_3 = i\alpha_1 + i\alpha_2$, 故 α_3 可由 α_1, α_2 线性表示. 因此, $\alpha_1, \alpha_2, \alpha_3$ 线性相关且 α_1, α_2 是向量组 $\alpha_1, \alpha_2, \alpha_3$ 的一个极大线性无关组.

注 4: 向量组 $\alpha_1, \dots, \alpha_n$ 的极大线性无关组一般不唯一.

【思考】 两组极大线性无关组之间有什么关系?

类比于向量组的极大线性无关组的定义,我们对线性空间引入基的概念.

定义 1.3.3(基) 设 V 是数域 F 上的线性空间, $\alpha_1, \dots, \alpha_n$ 是 V 中的一组向量. 若

(1) 向量组 $\alpha_1, \dots, \alpha_n$ 线性无关;

(2) V 中任一向量均可由 $\alpha_1, \dots, \alpha_n$ 线性表示,

则称 $\alpha_1, \dots, \alpha_n$ 是 V 的一个基底(或一组基).

例 1.3.4 单位矩阵 I 的 n 列可构成 \mathbb{R}^n (或 \mathbb{C}^n) 的一组基. 特别地, 令 e_1, \dots, e_n 是单位矩阵的 n 列, 则 e_1, \dots, e_n 称为 \mathbb{R}^n 的标准基.

例 1.3.5 已知 $A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, 求 $N(A)$ 的一组基.

解: 对矩阵 A 作初等变换, 得

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

因此, 方程组 $Ax = 0$ 的解为 $x_1 = -x_3, x_2 = 0$, 其基础解系取为 $[-1, 0, 1]^T$. 这意味着 $[-1, 0, 1]^T$ 是零空间 $N(A)$ 的一组基.

【思考】 明确线性空间 V 的一组基的重要意义是什么?

以 \mathbb{R}^2 平面为例, 我们之所以能够明确平面上每一点的原因在于平面上每一点与其坐标一一对应, 因此可采用点的坐标来表示点本身. 从线性空间的角度看, 坐标轴 xOy 表明 \mathbb{R}^2 空间通过选定标准基 e_1, e_2 来构建坐标系, 由此空间中的任一点都可由这组基唯一确定. 因此, 对于一般的线性空间, 明确一组基的重要原因在于强加一个“坐标系”. 在这个强加的坐标系下, 空间中的每个向量都可以唯一表示, 即向量的唯一表示定理.

定理 1.3.2(唯一表示定理) 设 x_1, \dots, x_n 是线性空间 V 的一组基, 则 V 中任一向量 x 都可由基 x_1, \dots, x_n 唯一表示.

定义 1.3.4(坐标) 设 x_1, \dots, x_n 是数域 F 上线性空间 V 的一组基, 对任意向量 $x \in V$, 令

$$x = \sum_{i=1}^n \alpha_i x_i = [x_1, \dots, x_n] \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$$

称有序数组 $[\alpha_1, \dots, \alpha_n]^T \in F^n$ 是 x 在基 x_1, \dots, x_n 下的坐标, 它由 x 与基 x_1, \dots, x_n 唯一确定.

例 1.3.6 证明 $E_{ij} \in \mathbb{R}^{2 \times 2} (i, j = 1, 2)$ 是 $\mathbb{R}^{2 \times 2}$ 的一组基, 并求矩阵 $A = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$ 在该组基下的坐标.

解: 考察方程

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

解得 $k_1 = k_2 = k_3 = k_4 = 0$. 从而 $\mathbf{E}_{11}, \mathbf{E}_{12}, \mathbf{E}_{21}, \mathbf{E}_{22}$ 线性无关.

对任意矩阵 $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{2 \times 2}$, 有

$$\mathbf{A} = a_{11} \mathbf{E}_{11} + a_{12} \mathbf{E}_{12} + a_{21} \mathbf{E}_{21} + a_{22} \mathbf{E}_{22}$$

根据基的定义知, $\mathbf{E}_{11}, \mathbf{E}_{12}, \mathbf{E}_{21}, \mathbf{E}_{22}$ 是 $\mathbb{R}^{2 \times 2}$ 的一组基.

又知 $\mathbf{A} = \mathbf{E}_{11} + 2\mathbf{E}_{12} + \mathbf{E}_{21} + \mathbf{E}_{22}$, 因此, 矩阵 \mathbf{A} 在 $\mathbf{E}_{11}, \mathbf{E}_{12}, \mathbf{E}_{21}, \mathbf{E}_{22}$ 下的坐标为 $[1, 2, 1, 1]^T$.

定义 1.3.5 (过渡矩阵) 设 x_1, \dots, x_n 和 y_1, \dots, y_n 是数域 F 上线性空间 V 的两组基, 令

$$y_i = a_{1i}x_1 + \dots + a_{ni}x_n = [x_1, \dots, x_n] \begin{bmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{bmatrix}, \quad i = 1, \dots, n$$

引入矩阵表示:

$$[y_1, \dots, y_n] = [x_1, \dots, x_n] \mathbf{A}$$

式中: $\mathbf{A} = (a_{ij})_{n \times n} \in F^{n \times n}$, 称 \mathbf{A} 是由基 x_1, \dots, x_n 到基 y_1, \dots, y_n 的过渡矩阵 (或变换矩阵).

命题 1.3.1 (过渡矩阵的性质) 设 V 是数域 F 上的线性空间, $\mathbf{A} \in F^{n \times n}$ 是由基 x_1, \dots, x_n 到基 y_1, \dots, y_n 的过渡矩阵, 则以下命题成立:

- (1) 过渡矩阵 \mathbf{A} 可逆;
- (2) 由基 y_1, \dots, y_n 到基 x_1, \dots, x_n 的过渡矩阵为 \mathbf{A}^{-1} ;
- (3) 任取 $x \in V$, 设 $x = \sum_{i=1}^n \alpha_i x_i = \sum_{i=1}^n \beta_i y_i$, 则

$$\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \quad \text{或} \quad \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \mathbf{A} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}$$

例 1.3.7 设 $f(x) \in P_n(x)$, 且 $f(x)$ 在基 $1, x, \dots, x^n$ 下的坐标为

$$\left[f(0), f'(0), \dots, \frac{f^{(n)}(0)}{n!} \right]^T$$

即 $f(x)$ 的表达式为

$$f(x) = f(0) + f'(0)x + \dots + \frac{f^{(n)}(0)}{n!}x^n$$

另一方面, 向量组 $1, x - x_0, \dots, (x - x_0)^n$ 也构成 $P_n(x)$ 的一组基, 则 $f(x)$ 在该组基下的坐标为

$$\left[f(x_0), f'(x_0), \dots, \frac{f^{(n)}(x_0)}{n!} \right]^T$$

设由基 $1, x, \dots, x^n$ 到基 $1, x - x_0, \dots, (x - x_0)^n$ 的过渡矩阵为 $A = (a_{ij})_{(n+1) \times (n+1)}$, 则有

$$[1, x - x_0, \dots, (x - x_0)^n] = [1, x, \dots, x^n] \begin{bmatrix} 1 & -x_0 & \cdots & (-x_0)^n \\ 0 & 1 & \cdots & n(-x_0)^{n-1} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

式中: 矩阵 A 的元素可根据二项式定理确定.

实际上, 例 1.3.7 说明了数学分析中泰勒公式和麦克劳林公式仅仅是由于选取的基不同而导致的公式形式不同, 本质上都是对同一空间中向量的表达.

定义 1.3.6 (维数) 在线性空间 V 中, 不同线性无关组中向量个数最大者叫作 V 的维数, 记为 $\dim V$. 当 $\dim V < \infty$, 称 V 为有限维空间, 否则称为无限维空间, 记为 $\dim V = \infty$.

例 1.3.8 离散时间信号空间 $S = \{\{x_k\} = [\dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots] \mid x_i \in \mathbb{R}\}$ 是 \mathbb{R} 上的无限维线性空间. 原因在于向量组 $\{1^k, 2^k, \dots\}$ 是线性无关向量组, 该集合包含无穷个线性无关的向量.

【思考】 由例 1.2.5 知, 齐次线性差分方程 (1.2.2) 的解集 \tilde{S} 是离散时间信号空间 S 的子空间. 请读者思考 \tilde{S} 的维数是多少.

定理 1.3.3 (维数与基的关系) 设 V 是有限维线性空间, 则 $\dim V = n$ 当且仅当 V 的任一组基有 n 个线性无关的向量.

证明: 必要性. 由 $\dim V = n$ 知, 线性空间 V 中必存在一组线性无关的向量 x_1, \dots, x_n . 假设对任意向量 $y \in V$, 有

$$k_1 x_1 + \dots + k_n x_n + k_{n+1} y = \theta$$

若 $k_{n+1} = 0$, 则必有 $k_1 = \dots = k_n = 0$. 由此, x_1, \dots, x_n, y 线性无关, 这表明线性空间 V 的维数大于 n , 故与 $\dim V = n$ 矛盾. 因此, $k_{n+1} \neq 0$ 且有

$$y = -\frac{k_1}{k_{n+1}} x_1 - \dots - \frac{k_n}{k_{n+1}} x_n$$

即: V 中任一向量均可由 x_1, \dots, x_n 线性表示, 所以 x_1, \dots, x_n 是 V 的一组基.

充分性. 设 x_1, \dots, x_n 是 V 的一组基, 则 $\dim V \geq n$. 现假设 $\dim V = m > n$, 则存在一组线性无关的向量 y_1, \dots, y_m , 它们可由基 x_1, \dots, x_n 表示, 即

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ \vdots \\ y_m = a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{cases}$$

由于 y_1, \dots, y_m 是线性无关的, 所以 $k_1 y_1 + \dots + k_m y_m = \theta$ 只有零解, 即

$$(k_1 a_{11} + \dots + k_m a_{m1}) x_1 + \dots + (k_1 a_{1n} + \dots + k_m a_{mn}) x_n = \theta$$

又知 x_1, \dots, x_n 线性无关, 所以

$$\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_m \end{bmatrix} = \mathbf{0} \quad (1.3.4)$$

注意到 $\text{rank}(\mathbf{A}) \leq n < m$, 故方程组 (1.3.4) 必有非零解, 这与方程 $k_1 \mathbf{y}_1 + \cdots + k_m \mathbf{y}_m = \mathbf{0}$ 只有零解矛盾. 因此, $\dim V = n$. 证毕.

例 1.3.9 求线性空间 \mathbb{C} 在实数域 \mathbb{R} 和复数域 \mathbb{C} 上的维数.

解: 复数域 \mathbb{C} 中任一复数均可由复数 1 和 i 线性表示. 现考察方程

$$k_1 \cdot 1 + k_2 \cdot i = 0$$

若 $k_1, k_2 \in \mathbb{R}$, 则上式成立当且仅当 $k_1 = k_2 = 0$, 这意味着 1 和 i 线性无关. 换言之, 向量组 1 和 i 构成 \mathbb{C} 的一组基. 此时, 实数域 \mathbb{R} 上的线性空间 \mathbb{C} 的维数为 2. 若 $k_1, k_2 \in \mathbb{C}$, 则有 $k_1 = i, k_2 = -1$ 时, $k_1 \cdot 1 + k_2 \cdot i = 0$. 这意味着 1 和 i 线性相关. 也就是说, 复数 1 和复数 i 分别构成 \mathbb{C} 的一组基. 此时, 复数域 \mathbb{C} 上的线性空间 \mathbb{C} 的维数为 1.

例 1.3.10 设矩阵 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 求 $N(\mathbf{A})$ 和 $R(\mathbf{A})$ 的维数.

解: $\dim N(\mathbf{A}) = n - \text{rank}(\mathbf{A}), \dim R(\mathbf{A}) = \text{rank}(\mathbf{A})$.

推论 1.3.1 (基扩充定理) n 维线性空间中任意 n 个线性无关的向量均为 V 的一个基底, 且任一线性无关向量组 $\mathbf{x}_1, \cdots, \mathbf{x}_r (1 \leq r < n)$ 可扩充为 V 的一个基底.

注 5: 基扩充定理是分析多个线性子空间与基或维数相关结果的重要理论基础. 特别注意的是, 在构造子空间的一组基时, 优先利用“加法”原则 (即可依据基的扩充定理增加线性无关的向量), 尽量避免“减法”原则 (即由空间的一组基通过删除某些向量后作为另一空间的一组基).

定理 1.3.4 (维数定理) 设 W_1 和 W_2 是线性空间 V 的两个子空间, 则

$$\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2)$$

证明: 设 W_1, W_2 和 $W_1 \cap W_2$ 的维数分别为 n_1, n_2 和 r . 取 $W_1 \cap W_2$ 的一组基为 $\mathbf{x}_1, \cdots, \mathbf{x}_r$ (优先利用“加法”原则), 并将这组向量分别扩充成为 W_1 和 W_2 的一组基, 记为

$$\mathbf{x}_1, \cdots, \mathbf{x}_r, \mathbf{y}_{r+1}, \cdots, \mathbf{y}_{n_1}; \quad \mathbf{x}_1, \cdots, \mathbf{x}_r, \mathbf{z}_{r+1}, \cdots, \mathbf{z}_{n_2}$$

于是

$$W_1 = \text{span}(\mathbf{x}_1, \cdots, \mathbf{x}_r, \mathbf{y}_{r+1}, \cdots, \mathbf{y}_{n_1})$$

$$W_2 = \text{span}(\mathbf{x}_1, \cdots, \mathbf{x}_r, \mathbf{z}_{r+1}, \cdots, \mathbf{z}_{n_2})$$

$$W_1 + W_2 = \text{span}(\mathbf{x}_1, \cdots, \mathbf{x}_r, \mathbf{y}_{r+1}, \cdots, \mathbf{y}_{n_1}, \mathbf{z}_{r+1}, \cdots, \mathbf{z}_{n_2})$$

现考察方程

$$\begin{aligned} \lambda_1 \mathbf{x}_1 + \cdots + \lambda_r \mathbf{x}_r + \mu_{r+1} \mathbf{y}_{r+1} + \cdots + \mu_{n_1} \mathbf{y}_{n_1} + \\ \xi_{r+1} \mathbf{z}_{r+1} + \cdots + \xi_{n_2} \mathbf{z}_{n_2} = \mathbf{0} \end{aligned} \quad (1.3.5)$$

式中: $\lambda_1, \cdots, \lambda_r, \mu_{r+1}, \cdots, \mu_{n_1}, \xi_{r+1}, \cdots, \xi_{n_2} \in F$ 为待定系数.

令 $w = \lambda_1 x_1 + \cdots + \lambda_r x_r + \mu_{r+1} y_{r+1} + \cdots + \mu_{n_1} y_{n_1}$, 则 $w \in W_1$. 又由式(1.3.5)知

$$w = -(\xi_{r+1} z_{r+1} + \cdots + \xi_{n_2} z_{n_2}) \in W_2$$

所以, $w \in W_1 \cap W_2$. 这表明 w 可由基 x_1, x_2, \cdots, x_r 线性表示.

由于向量 w 可由线性无关向量组 $x_1, \cdots, x_r, y_{r+1}, \cdots, y_{n_1}$ 线性表示, 故

$$\mu_{r+1} = \cdots = \mu_{n_1} = 0$$

进一步, 式(1.3.5)可改写为

$$\lambda_1 x_1 + \cdots + \lambda_r x_r + \xi_{r+1} z_{r+1} + \cdots + \xi_{n_2} z_{n_2} = \theta$$

又知 $x_1, \cdots, x_r, z_{r+1}, \cdots, z_{n_2}$ 线性无关, 故上式成立条件为

$$\lambda_1 = \cdots = \lambda_r = \xi_{r+1} = \cdots = \xi_{n_2} = 0$$

这表明方程(1.3.5)成立的条件为 $\lambda_1 = \cdots = \lambda_r = \mu_{r+1} = \cdots = \mu_{n_1} = \xi_{r+1} = \cdots = \xi_{n_2} = 0$, 即 $x_1, \cdots, x_r, y_{r+1}, \cdots, y_{n_1}, z_{r+1}, \cdots, z_{n_2}$ 线性无关. 注意到 $W_1 + W_2$ 的任一向量均可由向量组 $x_1, \cdots, x_r, y_{r+1}, \cdots, y_{n_1}, z_{r+1}, \cdots, z_{n_2}$ 线性表示, 故该向量组是 $W_1 + W_2$ 的一组基, 即

$$\dim(W_1 + W_2) = n_1 + n_2 - r$$

即维数定理成立. 证毕.

例 1.3.11 设 $W_1 = \text{span}(\alpha_1, \alpha_2, \alpha_3)$, $W_2 = \text{span}(\beta_1, \beta_2)$, 其中

$$\alpha_1 = [1, 2, 1, 0]^T, \quad \alpha_2 = [1, 1, 1, 0]^T, \quad \alpha_3 = [1, 0, 1, 1]^T$$

$$\beta_1 = [2, 1, 0, 0]^T, \quad \beta_2 = [0, 1, 0, 0]^T$$

求 $\dim(W_1 + W_2)$ 和 $\dim(W_1 \cap W_2)$.

解: 观察易知, $\alpha_1, \alpha_2, \alpha_3$ 线性无关, β_1, β_2 线性无关, 故有

$$\dim(W_1) = 3, \quad \dim(W_2) = 2$$

再对以下矩阵进行初等变换, 得

$$[\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2] = \begin{bmatrix} 1 & 1 & 1 & 2 & 0 \\ 2 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

于是,

$$\dim(W_1 + W_2) = \text{rank}(\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2) = 4$$

由维数定理可知, $\dim(W_1 \cap W_2) = 3 + 2 - 4 = 1$.

1.4 内积空间

我们已熟知二维空间和三维空间中的长度、距离和垂直等几何概念, 本节将在实(复)线性空间中引入类似的定义, 这些定义可为解决诸多实际问题提供有力的几何工具. 在实(复)线性空间中, 长度、距离和垂直这三个概念都是建立在内积基础上的.

定义 1.4.1(内积空间) 设 F 是实数域或复数域, V 是 F 上的线性空间, 若对 V 中任意两个向量 α 和 β 定义了一个数 $(\alpha, \beta) \in F$, 使得对任意向量 $x, y, z \in V$ 和 $k \in F$ 满足以下性质:

- (1) 共轭对称性: $(x, y) = \overline{(y, x)}$;
- (2) 可加性: $(x + y, z) = (x, z) + (y, z)$;
- (3) 齐次性: $(kx, y) = k(x, y)$;
- (4) 正定性: $(x, x) \geq 0$, 当且仅当 $x = \theta$ 时有 $(x, x) = 0$.

此时, V 称为一个内积空间, 数 (x, y) 称为 x 与 y 的内积. 有限维的实内积空间称为欧几里得空间(或欧氏空间). 有限维的复内积空间称为酉空间.

注 1: 欧氏(酉)空间的维数指它作为线性空间的维数; 欧氏(酉)空间的线性子空间仍是欧氏(酉)子空间.

注 2: 定义 1.4.1 中性质(3)齐次性仅对第一个向量成立, 对第二个向量则是“共轭齐次性”.

例 1.4.1 设 V 是数域 F 上的内积空间, $\forall x, y, z \in V$ 和 $k \in F$, 有

$$\begin{aligned}(x, y + z) &= \overline{(y + z, x)} \\ &= \overline{(y, x) + (z, x)} = (x, y) + (x, z) \\ (x, ky) &= \overline{k(y, x)} = \bar{k}(x, y)\end{aligned}$$

【思考】 $\forall y \in V, (\theta, y) = ?$

分析: $(\theta, y) = (0 \cdot \theta, y) = 0(\theta, y) = 0$.

例 1.4.2 在 \mathbb{R}^n 空间中, 取 $x = [x_1, \dots, x_n]^T, y = [y_1, \dots, y_n]^T$, 令

$$(x, y) = y^T x = x_1 y_1 + \dots + x_n y_n \quad (1.4.1)$$

易验证式(1.4.1)满足内积的四条性质, 故 \mathbb{R}^n 是欧氏空间, 仍记为 \mathbb{R}^n .

例 1.4.3 在 \mathbb{C}^n 空间中, 取 $x = [x_1, \dots, x_n]^T, y = [y_1, \dots, y_n]^T$, 令

$$(x, y) = y^H x = x_1 \bar{y}_1 + \dots + x_n \bar{y}_n \quad (1.4.2)$$

式中: $y^H = [\bar{y}_1, \dots, \bar{y}_n]$ 表示为向量 y 的共轭转置. 易验证式(1.4.2)满足内积四条性质, 故 \mathbb{C}^n 为酉空间, 仍记为 \mathbb{C}^n .

为讨论方便, 这里引入 Hermite 矩阵和正定矩阵的定义.

定义 1.4.2(Hermite 矩阵) 设矩阵 $A = (a_{ij}) \in \mathbb{C}^{n \times n}$, 若 $A^H = A$, 则称矩阵 A 是 Hermite 矩阵; 若 $A^H = -A$, 则称 A 是反 Hermite 矩阵, 其中 $A^H = (\bar{A})^T$.

定义 1.4.3(正定矩阵) 设 $x = [x_1, \dots, x_n]^T \in \mathbb{C}^n$ 是非零向量, $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ 是 Hermite 矩阵, 定义关于 $x_1, \dots, x_n, \bar{x}_1, \dots, \bar{x}_n$ 的复系数二次多项式

$$f(x) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} \bar{x}_i x_j$$

或等价写成

$$f(x) = x^H A x$$

称为复二次型, \mathbf{A} 称为复二次型 $f(\mathbf{x})$ 的矩阵. 若对任意向量 $\mathbf{x} \in \mathbb{C}^n$, 有 $f(\mathbf{x}) \geq 0$, 且有 $f(\mathbf{x}) = 0$ 当且仅当 $\mathbf{x} = \mathbf{0}$, 则称 $f(\mathbf{x})$ 是正定二次型, \mathbf{A} 是正定矩阵. 若 $\forall \mathbf{x} \in \mathbb{C}^n$, 有 $f(\mathbf{x}) \geq 0$, 则称 $f(\mathbf{x})$ 是半正定二次型, \mathbf{A} 是半正定矩阵. 类似地, 可定义“负定二次型”、“负定矩阵”、“半负定二次型”和“半负定矩阵”.

注 3: 与实对称矩阵相对应的复矩阵不是复对称矩阵, 而是复共轭对称矩阵, 习惯上称为 Hermite 矩阵. 实对称矩阵可以看成 Hermite 矩阵的特例, 也可以定义正定矩阵、半正定矩阵等.

例 1.4.4 设复二次型 $f(\mathbf{x}) = \mathbf{x}^H \mathbf{A} \mathbf{x}$, 其中, $\mathbf{x} = [x_1, x_2]^T \in \mathbb{C}^2$, 复二次型 $f(\mathbf{x})$ 的矩阵 \mathbf{A} 为

$$\mathbf{A} = \begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix}$$

则 $f(\mathbf{x}) = |x_1 + ix_2|^2 + |x_2|^2$. 显然, $f(\mathbf{x}) \geq 0$; 当且仅当 $\mathbf{x} = \mathbf{0}$ 时, $f(\mathbf{x}) = 0$. 因此, $f(\mathbf{x})$ 是正定二次型, \mathbf{A} 是正定矩阵.

若复二次型 $f(\mathbf{x})$ 的矩阵 \mathbf{A} 为

$$\mathbf{A} = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$$

则 $f(\mathbf{x}) = |x_1 + ix_2|^2$. 显然, $f(\mathbf{x}) \geq 0$; 当且仅当 $x_1 + ix_2 = 0$ 时, $f(\mathbf{x}) = 0$. 因此, $f(\mathbf{x})$ 是半正定二次型, \mathbf{A} 是半正定矩阵.

【思考】 设实二次型 $f(\mathbf{x}) = \mathbf{x}^H \mathbf{A} \mathbf{x}$, $\mathbf{x} = [x_1, \dots, x_n]^T \in \mathbb{R}^n$ 是未定元向量. 是否存在非实对称阵矩阵 \mathbf{A} 使得 $f(\mathbf{x})$ 是正定二次型?

例 1.4.5 在 \mathbb{C}^n 空间中, 取 $\mathbf{x} = [x_1, \dots, x_n]^T$, $\mathbf{y} = [y_1, \dots, y_n]^T$, 并设 Hermite 矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 是正定的. 令

$$(\mathbf{x}, \mathbf{y}) = \mathbf{y}^H \mathbf{A} \mathbf{x} \quad (1.4.3)$$

则易验证式(1.4.3)满足内积的四条性质, 故 \mathbb{C}^n 是酉空间.

注 4: 同一线性空间可定义不同的内积. 若无特殊说明, \mathbb{R}^n 空间的内积采用式(1.4.1), \mathbb{C}^n 空间的内积采用式(1.4.2).

【思考】 为什么在酉空间中采用共轭转置而不用普通的向量转置?

【分析】 例如, 若 $\mathbf{x} = [1, i]^T \in \mathbb{C}^2$, 则 $\mathbf{x}^T \mathbf{x} = -2 < 0$. 显然, $(\mathbf{x}, \mathbf{y}) = \mathbf{y}^T \mathbf{x}$ 不满足内积的正定性要求. 而采用共轭转置, 则保证了内积正定性要求.

例 1.4.6 设函数 $f, g \in C[a, b]$, 定义内积 $(f, g) = \int_a^b f(x)g(x)dx$, 则 $C[a, b]$ 为欧氏空间.

定义 1.4.4 (度量矩阵) 设 ζ_1, \dots, ζ_n 是内积空间 V 中的一组基, 称 n 阶矩阵

$$\mathbf{A} = ((\zeta_i, \zeta_j))_{n \times n} = \begin{bmatrix} (\zeta_1, \zeta_1) & (\zeta_1, \zeta_2) & \cdots & (\zeta_1, \zeta_n) \\ (\zeta_2, \zeta_1) & (\zeta_2, \zeta_2) & \cdots & (\zeta_2, \zeta_n) \\ \vdots & \vdots & \ddots & \vdots \\ (\zeta_n, \zeta_1) & (\zeta_n, \zeta_2) & \cdots & (\zeta_n, \zeta_n) \end{bmatrix}$$

为 V 关于基 ζ_1, \dots, ζ_n 的度量矩阵(或 Gram 矩阵), 常记为 $G(\zeta_1, \dots, \zeta_n)$.

注 5: 内积空间中内积(已确定基)与度量矩阵是一一对应的.

设 ζ_1, \dots, ζ_n 是内积空间 V 中的一组基, 对 V 中任意向量 x 和 y 有

$$\begin{aligned} x &= \sum_{i=1}^n \xi_i \zeta_i \\ y &= \sum_{i=1}^n \eta_i \zeta_i \\ (x, y) &= \sum_{i,j=1}^n \xi_i \bar{\eta}_j (\zeta_i, \zeta_j) \end{aligned}$$

式中: $\xi = [\xi_1, \dots, \xi_n]^T$ 和 $\eta = [\eta_1, \dots, \eta_n]^T$ 分别是向量 x 和 y 在基 ζ_1, \dots, ζ_n 下的坐标.

于是

$$(x, y) = \eta^H A \xi \quad (1.4.4)$$

式中: $A = ((\zeta_i, \zeta_j))_{n \times n}$ 为度量矩阵.

将式(1.4.4)代入 $(x, y) = \overline{(y, x)}$, 得

$$\eta^H A \xi = \eta^H A^H \xi \quad (1.4.5)$$

由于式(1.4.5)对任意向量 ξ 和 η 都成立, 因此 $A = A^H$, 即 A 是 Hermite 矩阵.

再由内积正定性知, $(x, x) = \xi^H A \xi \geq 0$, 当且仅当 $\xi = 0$ 时取等号. 因此, A 是正定矩阵.

命题 1.4.1(度量矩阵的性质) 设 $G(\varepsilon_1, \dots, \varepsilon_n)$ 和 $G(\zeta_1, \dots, \zeta_n)$ 均为内积空间 V 的度量矩阵, 则有

- (1) $G(\varepsilon_1, \dots, \varepsilon_n)$ 和 $G(\zeta_1, \dots, \zeta_n)$ 是正定 Hermite 矩阵;
- (2) $G(\varepsilon_1, \dots, \varepsilon_n)$ 和 $G(\zeta_1, \dots, \zeta_n)$ 合同(或相合), 即存在非奇异矩阵 P 使得

$$P^H G(\varepsilon_1, \dots, \varepsilon_n) P = G(\zeta_1, \dots, \zeta_n)$$

式中: P 是由基 $\varepsilon_1, \dots, \varepsilon_n$ 到基 ζ_1, \dots, ζ_n 的过渡矩阵.

证明: 性质(1)由上述分析整理可得; 这里仅证明性质(2).

设 V 中任意向量 x 和 y 分别表示为

$$\begin{aligned} x &= \sum_{i=1}^n \xi_i \varepsilon_i = \sum_{i=1}^n \xi'_i \zeta_i \\ y &= \sum_{i=1}^n \eta_i \varepsilon_i = \sum_{i=1}^n \eta'_i \zeta_i \end{aligned}$$

式中: $\xi = [\xi_1, \dots, \xi_n]^T$ 和 $\xi' = [\xi'_1, \dots, \xi'_n]^T$ 是向量 x 分别在基 $\varepsilon_1, \dots, \varepsilon_n$ 和基 ζ_1, \dots, ζ_n 下的坐标, $\eta = [\eta_1, \dots, \eta_n]^T$ 和 $\eta' = [\eta'_1, \dots, \eta'_n]^T$ 是向量 y 在基 $\varepsilon_1, \dots, \varepsilon_n$ 和基 ζ_1, \dots, ζ_n 下的坐标.

由此,

$$(x, y) = \eta^H A \xi = (\eta')^H A' \xi' \quad (1.4.6)$$

式中: $A=G(\varepsilon_1, \cdots, \varepsilon_n), A'=G(\zeta_1, \cdots, \zeta_n)$.

再设 P 是由基 $\varepsilon_1, \cdots, \varepsilon_n$ 到基 ζ_1, \cdots, ζ_n 的过渡矩阵, 则有 $\xi=P\xi'$ 和 $\eta=P\eta'$. 将其代入式(1.4.6)得

$$(\eta')^H P^H A P \xi' = (\eta')^H A' \xi'$$

由于度量矩阵是由基唯一确定的, 故 $P^H A P = A'$, 即度量矩阵 $G(\varepsilon_1, \cdots, \varepsilon_n)$ 和 $G(\zeta_1, \cdots, \zeta_n)$ 合同. 证毕.

定义 1.4.5(长度) 设 F 是实数域或复数域, V 是内积空间, 对任意向量 $x \in V$, 非负实数 $\sqrt{(x, x)}$ 称为向量 x 的长度(模), 记为 $\|x\|$. 长度为 1 的向量称为单位向量.

命题 1.4.2(长度性质) 设 V 是内积空间, 则 $\forall x, y \in V$ 和 $k \in F$ 有

- (1) 正定性: $\|x\| \geq 0, \|x\| = 0$ 当且仅当 $x = \theta$;
- (2) 齐次性: $\|kx\| = |k| \|x\|$;
- (3) 三角不等式: $\|x+y\| \leq \|x\| + \|y\|$;
- (4) 平行四边形法则: $\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$.

定理 1.4.1(Cauchy 不等式) 设 V 是数域 F 上的内积空间, 对任意 $x, y \in V$,

$$|(x, y)| \leq \|x\| \|y\|$$

式中: $|(x, y)| = \|x\| \|y\|$ 当且仅当 x, y 线性相关时成立.

证明: 当 $x = \theta$ 时, 该不等式显然成立. 当 $x \neq \theta$ 时, 对任意向量 $x, y \in V$ 和 $a \in F$, 有 $ax + y \in V$ 且 $(ax + y, ax + y) \geq 0$. 由此,

$$a^2 \|x\|^2 + 2a(x, y) + \|y\|^2 \geq 0$$

将 a 看作待定参数, 则上式为一元二次不等式. 其成立条件为

$$\Delta = [2(x, y)]^2 - 4\|x\|^2 \|y\|^2 \leq 0$$

整理可得 $|(x, y)| \leq \|x\| \|y\|$.

现考虑等号成立条件. 当 x, y 线性相关时, 存在数 $k \in F$ 使得 $x = ky$, 那么有 $|(x, y)| = \|x\| \|y\|$. 反之, 当 $|(x, y)| = \|x\| \|y\|$ 时, 若 $x = \theta$ 或 $y = \theta$ 时, x, y 线性相关; 若 x, y 均为非零向量, 且假设 x, y 线性无关, 则对任意 $a \in F$ 有 $ax + y \neq \theta$, 即

$$(ax + y, ax + y) > 0$$

取 $a = -\frac{(x, y)}{\|x\|^2}$, 并代入上式可得

$$|(x, y)|^2 < (x, x)(y, y) = \|x\|^2 \|y\|^2$$

与假设 $|(x, y)| = \|x\| \|y\|$ 矛盾, 故假设 x, y 线性无关不成立, 即 x, y 线性相关. 证毕.

推论 1.4.1(三角不等式) 设 V 是内积空间, 则 $\forall x, y \in V$, 有

$$\|x + y\| \leq \|x\| + \|y\|$$

注 6: Cauchy 不等式有着广泛的应用. 在根据内积空间定义不同内积可得到不同的 Cauchy 不等式.

(1) 对 \mathbb{R}^n 中任意两向量 $x = [x_1, \cdots, x_n]^T$ 和 $y = [y_1, \cdots, y_n]^T$, 有

$$\left| \sum_{i=1}^n x_i y_i \right| \leq \sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}$$

(2) 对 $C[a, b]$ 中任意两函数 $f(x)$ 和 $g(x)$, 有

$$\left| \int_a^b f(x) g(x) dx \right| \leq \sqrt{\int_a^b f^2(x) dx} \sqrt{\int_a^b g^2(x) dx}$$

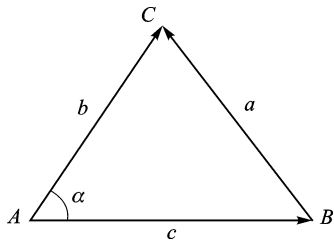


图 1.4.1 $\triangle ABC$ 示意图

在平面几何中, 利用余弦定理可建立夹角和边长间的关系. 若将 $\triangle ABC$ (见图 1.4.1) 的两边用 \mathbb{R}^2 空间的两个向量 \overrightarrow{AC} 和 \overrightarrow{AB} 表示, 则向量 \overrightarrow{AC} 和 \overrightarrow{AB} 间的夹角 α 满足

$$\cos \alpha = \frac{\|\overrightarrow{AC}\|^2 + \|\overrightarrow{AB}\|^2 - \|\overrightarrow{AC} - \overrightarrow{AB}\|^2}{2\|\overrightarrow{AC}\|\|\overrightarrow{AB}\|}$$

利用平行四边形公式整理得

$$\cos \alpha = \frac{(\overrightarrow{AC}, \overrightarrow{AB})}{\|\overrightarrow{AC}\|\|\overrightarrow{AB}\|}$$

上式是 \mathbb{R}^2 空间两向量的夹角. 基于此, 可将平面中向量的夹角定义扩展至一般的内积空间.

定义 1.4.6 (向量夹角) 设 V 是欧氏空间, 对 V 中任意向量 x 和 y , 定义向量 x 和 y 的夹角为

$$\alpha = \langle x, y \rangle = \arccos \frac{(x, y)}{\|x\|\|y\|} \in [0, \pi]$$

【思考】 若 V 是酉空间, 则向量夹角可采用定义 1.4.6 中的表达式吗?

正是内积定义的引入才有了向量长度和向量夹角的概念, 也带来了线性空间“图形化”的理解, 这使得线性空间的结构更加丰富多彩.

例 1.4.7 设 x_1, x_2 是 \mathbb{R}^2 的一组基, 则 \mathbb{R}^2 关于 x_1, x_2 的度量矩阵为

$$G(x_1, x_2) = \begin{bmatrix} (x_1, x_1) & (x_2, x_1) \\ (x_1, x_2) & (x_2, x_2) \end{bmatrix}$$

考察 $G(x_1, x_2)$ 的行列式 $|G(x_1, x_2)|$ 得

$$|G(x_1, x_2)| = \|x_1\|^2 \|x_2\|^2 - (x_1, x_2)^2$$

令向量 x_1 与 x_2 的夹角为 α , 则上式可改写为

$$|G(x_1, x_2)| = \|x_1\|^2 \|x_2\|^2 - \|x_1\|^2 \|x_2\|^2 \cos^2 \alpha = \|x_1\|^2 \|x_2\|^2 \sin^2 \alpha$$

上式表明, 行列式 $|G(x_1, x_2)|$ 恰好是以向量 x_1 与 x_2 为邻边的平行四边形的面积的平方.

实际上, 若设 x_1, \dots, x_k 是 n 维欧氏空间的一组线性无关向量, 并记 k 维子空间 $W = \text{span}(x_1, \dots, x_k)$ 关于基 x_1, \dots, x_k 的度量矩阵为 $G(x_1, \dots, x_k)$, 则该矩阵的行列式 $|G(x_1, \dots, x_k)|$ 的几何含义为 k 维超平行体体积的平方.

在二维或三维欧氏空间中,若两个非零向量夹角满足 $\cos \alpha = 0$ 即 $\alpha = \frac{\pi}{2}$, 则称两向量垂直(几何空间).“垂直”关系在几何空间尤为重要,在内积空间有如下定义.

定义 1.4.7(正交向量和正交向量组) 设 V 是内积空间,对 V 中向量 x 和 y ,若 $(x, y) = 0$, 则称向量 x 与 y 正交(或垂直),记为 $x \perp y$. 若一组非零向量两两正交,则称为正交向量组. 单位向量构成的正交向量组称为标准正交向量组.

注 7: 零向量 θ 与任何向量均正交. 正交向量组要求向量均为非零向量.

注 8: 向量 x 与 y 正交当且仅当 y 与 x 正交, 当且仅当 $\|x+y\|^2 = \|x\|^2 + \|y\|^2$, 这就是广为熟知的勾股定理. 该结果可推广至一般情形: 若 x_1, \dots, x_k 是 V 中正交向量组, 则

$$\|x_1 + \dots + x_k\|^2 = \|x_1\|^2 + \dots + \|x_k\|^2$$

例 1.4.8 求 \mathbb{R}^4 中的单位向量 x 使它与 $\alpha_1 = [1, 1, -1, 1]^T, \alpha_2 = [1, -1, -1, 1]^T, \alpha_3 = [2, 1, 1, 3]^T$ 均正交.

解: 设 $x = [x_1, x_2, x_3, x_4]^T$, 则有

$$\begin{cases} x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1 \\ x_1 + x_2 - x_3 + x_4 = 0 \\ x_1 - x_2 - x_3 + x_4 = 0 \\ 2x_1 + x_2 + x_3 + 3x_4 = 0 \end{cases}$$

$$\text{解得 } x = \left[\frac{2\sqrt{26}}{13}, 0, \frac{\sqrt{26}}{26}, -\frac{3\sqrt{26}}{26} \right]^T.$$

定理 1.4.2 正交向量组线性无关.

证明: 设向量组 x_1, x_2, \dots, x_s 是内积空间 V 的正交向量组, 现考察方程

$$k_1 x_1 + k_2 x_2 + \dots + k_s x_s = \theta \quad (1.4.7)$$

式中: $k_1, k_2, \dots, k_s \in F$ 为待定系数.

用 $x_i (i=1, 2, \dots, s)$ 与式(1.4.7)两端作内积得

$$(k_1 x_1 + k_2 x_2 + \dots + k_s x_s, x_i) = (\theta, x_i) = 0 \quad (1.4.8)$$

由于 x_1, x_2, \dots, x_s 是正交向量组, 所以当 $j \neq i$ 时, 有 $(x_j, x_i) = 0$; 当 $j = i$ 时, 有 $(x_j, x_i) \neq 0 (j, i=1, 2, \dots, s)$. 因此, 由式(1.4.7)可得

$$k_i (x_i, x_i) = 0$$

即 $k_i = 0 (i=1, 2, \dots, s)$. 因此, x_1, x_2, \dots, x_s 线性无关.

推论 1.4.2 在 n 维内积空间中, 正交向量组中的向量不会超过 n 个.

定义 1.4.8(标准正交基) 在 n 维内积空间中, 由 n 个向量组成的正交向量组称为正交基. 由单位向量组成的正交基称为标准正交基.

例 1.4.9 考察 $[-\pi, \pi]$ 上的三角函数组 $1, \cos t, \sin t, \dots, \cos kt, \sin kt, \dots$ 是否为正交向量组.

解: 事实上, 对任意 $m, n, l \in \mathbb{N}$ 且 $m \neq n$ 有

$$\int_{-\pi}^{\pi} \cos mt \cdot \cos nt dt = 0$$

$$\int_{-\pi}^{\pi} \sin mt \cdot \sin nt dt = 0$$

$$\int_{-\pi}^{\pi} \cos mt \cdot \sin lt dt = 0$$

因此,在区间 $[-\pi, \pi]$ 上的三角函数组 $1, \cos t, \sin t, \dots, \cos kt, \sin kt, \dots$,是正交向量组.我们将 $[-\pi, \pi]$ 上的光滑函数 $f(x)$ 展开为

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

式中:

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx dx, \quad n = 0, 1, \dots$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx dx, \quad n = 1, 2, \dots$$

上述展开称为傅里叶三角级数.由此可知, $1, \cos t, \sin t, \dots, \cos kt, \sin kt, \dots$ 是线性空间 $C[-\pi, \pi]$ 的一组正交基.

注 9: 将内积空间的一组正交基进行单位化处理即可得到标准正交基.线性空间 V 的向量组 ζ_1, \dots, ζ_n 是标准正交基,当且仅当 V 关于基 ζ_1, \dots, ζ_n 的度量矩阵为单位矩阵.在标准正交基下,内积的运算变得更加简单.

例 1.4.10 设 ζ_1, \dots, ζ_n 是线性空间 V 的标准正交基,并定义 $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]^T$ 和 $\beta = [\beta_1, \beta_2, \dots, \beta_n]^T$ 分别是 V 中向量 x 和 y 在基 ζ_1, \dots, ζ_n 下的坐标,则

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

定义 1.4.9(向量正交于集合) 设 V 是内积空间, W 是 V 的子集,若对任意向量 $y \in W$ 和 $x \in V$ 有 $x \perp y$,则称 x 正交(或垂直)于集合 W ,记为 $x \perp W$.

定义 1.4.10(两集合正交) 设 W_1 与 W_2 是内积空间 V 的两个子集,若对任意向量 $x \in W_1$ 和任意向量 $y \in W_2$ 有 $x \perp y$,则称 W_1 与 W_2 是正交的,记为 $W_1 \perp W_2$.

特别地,当上述定义中的子集是内积空间 V 的线性子空间,则有向量正交于子空间和两子空间正交的定义.

例 1.4.11 在直角坐标系 $Oxyz$ 中,定义过坐标原点的直线 x 和过坐标原点的平面 W_1 和 W_2 .若直线 x 在几何意义下垂直于平面 W_1 ,则有 $x \perp W_1$ (几何和内积意义下);若平面 W_1 与 W_2 在几何意义下相互垂直,则 W_1 不正交于 W_2 .

定义 1.4.11(子空间正交补) 设 W 是线性空间 V 的线性子空间,则集合 $W^\perp = \{x \in V | x \perp W\}$ 称为 W 的正交补.

定理 1.4.3 设 W 是线性空间 V 的子空间,则集合 W^\perp 是 V 的线性子空间.

证明: 对任意向量 $x, y \in W^\perp, z \in W$ 和 $\lambda, \mu \in F$,有

$$(\lambda x + \mu y, z) = \lambda(x, z) + \mu(y, z) = 0$$

故 $(\lambda x + \mu y) \in W^\perp$, 因此 W^\perp 是 V 的线性子空间.

【思考】 与线性子空间 W 正交的子空间是 W 的正交补空间吗?

分析: 在直角坐标系 $Oxyz$ 中, 假设 W 是位于 Ox 轴上所有向量的集合, V_1 是位于 Oy 轴上所有向量的集合, V_2 是 yOz 平面. 显然, $W \perp V_1$, 但 V_1 并不是 W 的正交补空间, V_2 才是 W 的正交补空间. 因此, 在说明子空间 V 是另一子空间 W 的正交补空间时不仅须说明子空间 V 和 W 是正交的, 还须说明所有与子空间 W 正交的向量都在子空间 V 中. 或者用如下定理表达.

定理 1.4.4 设 W 是线性空间 V 的线性子空间, 则 $V = W + W^\perp$.

证明: 设 $\dim(W) = m$, 并且 x_1, \dots, x_m 是 W 的一组标准正交基. 对任意 $x \in V$, 令

$$\alpha_1 = (x, x_1)x_1 + \dots + (x, x_m)x_m$$

$$\alpha_2 = x - \alpha_1$$

则 $\alpha_1 \in W$, 且

$$\begin{aligned} (\alpha_2, x_i) &= (x - \alpha_1, x_i) = (x, x_i) - \left(\sum_{j=1}^m (x, x_j)x_j, x_i \right) \\ &= (x, x_i) - (x, x_i)(x_i, x_i) = 0, i = 1, 2, \dots, m \end{aligned}$$

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

因为 x 是 V 中任意向量, 且 $x = \alpha_1 + \alpha_2$, 所以 $V = W + W^\perp$.

证毕.

注 10: 我们在说明线性空间 V 的两个子空间 W_1 和 W_2 相等时, 常采用以下两种方法:

(1) W_1 和 W_2 均是由同一向量组所张成的子空间, 或更确定地说明它们有相同的一组基;

(2) W_1 是 W_2 的子集(或 W_2 是 W_1 的子集)且 W_1 和 W_2 的维数相等.

1.5 直和与投影

由定理 1.4.4 知, 和空间可分解为子空间与它的正交补空间之和. 这一分解极具特殊性:

(1) 子空间与它的正交补空间正交(定义 1.4.11);

(2) 和空间中的任一向量唯一地分解为子空间的一个向量和它的正交补空间的一个向量.

特别是(2)中的向量唯一表示法是向量(正交)投影的理论基础. 本节将具体阐述这一问题.

定义 1.5.1(直和与正交直和) 设 W_1 与 W_2 是线性空间 V 的子空间, 若和空间 $W_1 + W_2$ 中任意向量均唯一地表示成 W_1 中的一个向量与 W_2 中的一个向量之和,

则称 $W_1 + W_2$ 是 W_1 与 W_2 的直和, 记为 $W_1 \dot{+} W_2$. 进一步, 若 $W_1 \perp W_2$, 则称直和 $W_1 \dot{+} W_2$ 是 W_1 与 W_2 的正交直和, 记为 $W_1 \oplus W_2$.

例 1.5.1 在直角坐标系 $Oxyz$ 中, 若 $W_1 = xOy$ 平面, $W_2 = yOz$ 平面, 试判断 $W_1 + W_2$ 是否为直和.

解: 设 y 轴上任一点坐标为 $[0, \beta, 0]^T$, 对于任意 $\mu \in \mathbb{R}$ 且 $\mu \neq 0$ 有

$$\begin{bmatrix} 0 \\ \beta \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \mu + \beta \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -\mu \\ 0 \end{bmatrix}$$

式中: $[0, \mu + \beta, 0]^T \in W_1$, $[0, -\mu, 0]^T \in W_2$. 这说明 y 轴上的点的表示方法不唯一. 故 $W_1 + W_2$ 不是直和.

若 $W_1 = x$ 轴, $W_2 = yOz$ 平面, 则对 $W_1 + W_2$ 空间任一点 $[\alpha, \beta, \gamma]^T$ 有

$$\begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} \alpha \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \beta \\ \gamma \end{bmatrix}$$

上式表示唯一, 故 $W_1 + W_2$ 是直和.

定理 1.5.1 (直和判定定理) 设 W_1 与 W_2 是线性空间 V 的两个子空间, 则以下命题等价:

- (1) $W_1 + W_2$ 是直和;
- (2) $W_1 + W_2$ 中零元素表示法唯一;
- (3) $W_1 \cap W_2 = \{\theta\}$;
- (4) $\dim(W_1 + W_2) = \dim W_1 + \dim W_2$.

证明: (1) \Rightarrow (2). 由于 $\theta \in W_1 + W_2$, 则根据定义 1.5.1 知, 表达式

$$\theta = \theta_1 + \theta_2, \quad \theta_1 \in W_1, \theta_2 \in W_2$$

唯一, 即 $W_1 + W_2$ 中零元素表示法唯一.

(2) \Rightarrow (3). 对任意向量 $\alpha \in W_1 \cap W_2$, 有

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

式中: $\alpha \in W_1, -\alpha \in W_2$.

又知 $\theta \in W_1 + W_2$ 且零元素表示法唯一, 故 $\alpha = \theta$. 因此, $W_1 \cap W_2 = \{\theta\}$.

(3) \Rightarrow (4). 若 $W_1 \cap W_2 = \{\theta\}$, 则根据维数定理得

$$\dim(W_1 + W_2) = \dim W_1 + \dim W_2$$

(4) \Rightarrow (1). 采用反证法证明. 假设存在向量 $\alpha \in W_1 + W_2$ 的表示法不唯一, 即

$$\alpha = \alpha_1 + \alpha_2 = \beta_1 + \beta_2$$

式中: $\alpha_1, \beta_1 \in W_1, \alpha_2, \beta_2 \in W_2$.

根据线性子空间定义知, $-\beta_1 \in W_1, -\alpha_2 \in W_2$. 于是

$$\alpha_1 - \beta_1 \in W_1, \quad -\alpha_2 + \beta_2 \in W_2$$

且有

$$\alpha_1 - \beta_1 = \alpha_1 + \alpha_2 - \beta_1 - \alpha_2 = \beta_2 - \alpha_2$$

于是

$$\alpha_1 - \beta_1 \in W_1 \cap W_2$$

$$\beta_2 - \alpha_2 \in W_1 \cap W_2$$

根据 $\dim(W_1 + W_2) = \dim W_1 + \dim W_2$ 得, $\dim(W_1 \cap W_2) = 0$, 即 $W_1 \cap W_2 = \{\theta\}$. 进而 $\alpha_1 = \beta_1$ 且 $\beta_2 = \alpha_2$. 这表明 α 表示法唯一, 即 $W_1 + W_2$ 是直和.

例 1.5.2 取 $V = \mathbb{R}^{n \times n}$, $W_1 = \{A \in V \mid A^T = A\}$, $W_2 = \{A \in V \mid A^T = -A\}$, 则 $V = W_1 + W_2$.

解: 由例 1.2.7 知, $V = W_1 + W_2$. 设 $B \in (W_1 \cap W_2)$, 则 $B = -B$, 即 $B = O$. 根据定理 1.5.1 知, $V = W_1 + W_2$.

例 1.5.3 定义 $\mathbb{R}^{2 \times 2}$ 的两个线性子空间分别为

$$W_1 = \left\{ \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \in \mathbb{R}^{2 \times 2} \mid \begin{cases} 2x_1 + 3x_2 - x_3 = 0 \\ x_1 + 2x_2 + 3x_3 - x_4 = 0 \end{cases} \right\}$$

$$W_2 = \text{span} \left(\begin{bmatrix} 2 & -1 \\ a+2 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 2 \\ 4 & a+8 \end{bmatrix} \right)$$

(1) 求 W_1 的一组基;

(2) 当 a 取何值时, $W_1 + W_2$ 是直和.

解: (1) 根据齐次线性方程组

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

得到方程组的基础解系为 $\alpha_1 = [1, -1, -1, 4]^T$ 和 $\alpha_2 = [-3, 2, 0, 1]^T$. 因此, W_1 的一组基为

$$\begin{bmatrix} 1 & -1 \\ -1 & -4 \end{bmatrix}, \begin{bmatrix} -3 & 2 \\ 0 & 1 \end{bmatrix}$$

(2) 若要求 $W_1 + W_2$ 是直和, 则须满足 $\dim(W_1 + W_2) = 4$. 又知 $\dim W_1 = 2$ 和 $\dim W_2 = 2$, 因此, 向量组

$$\begin{bmatrix} 1 & -1 \\ -1 & -4 \end{bmatrix}, \begin{bmatrix} -3 & 2 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 2 & -1 \\ a+2 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 2 \\ 4 & a+8 \end{bmatrix}$$

须线性无关, 即如下矩阵非奇异:

$$A = \begin{bmatrix} 1 & -3 & 2 & -1 \\ -1 & 2 & -1 & 2 \\ -1 & 0 & a+2 & 4 \\ -4 & 1 & 1 & a+8 \end{bmatrix}$$

经计算知, 当 $a \neq -1$ 且 $a \neq 7$ 时, $|A| \neq 0$, 即 $W_1 + W_2$ 是直和.

例 1.5.4 设 $A \in \mathbb{C}^{n \times n}$ 且满足 $A^2 = A$, 试证明 $\mathbb{C}^n = R(A) + R(I - A)$.

解: 对任意向量 $x \in \mathbb{C}^n$, 有

$$x = Ax + (I - A)x$$

式中: $Ax \in R(A)$, $(I - A)x \in R(I - A)$. 故 $\mathbb{C}^n = R(A) + R(I - A)$.

设 $y \in (R(A) \cap R(I - A))$, 则必存在向量 $x, z \in \mathbb{C}^n$ 满足

$$y = Ax = (I - A)z$$

对上式左右两端乘以 A 得

$$A^2 x = A(I - A)z$$

整理得 $y = 0$. 故 $R(A) \cap R(I - A) = \{0\}$. 因此, $R(A) + R(I - A)$ 是直和.

定理 1.5.2 若 W_1 与 W_2 是线性空间 V 的两个子空间, 且 $W_1 \perp W_2$, 则

$$W_1 + W_2 = W_1 \dot{+} W_2$$

证明: 因为 $W_1 \perp W_2$, 所以对任意向量 $\alpha \in W_1$ 和 $\beta \in W_2$, 有 $(\alpha, \beta) = 0$. 假设存在向量 $x \in W_1 \cap W_2$, 则必有 $(x, x) = 0$, 即 $x = \theta$. 这表明 $W_1 \cap W_2 = \{\theta\}$, 即 $W_1 + W_2 = W_1 \dot{+} W_2$.

定义 1.5.2 (直和分解与正交直和分解) 设 W_1 与 W_2 是线性空间 V 的两个子空间, 且 $V = W_1 \dot{+} W_2$, 则称表达式 $V = W_1 \dot{+} W_2$ 为直和分解. 若 $V = W_1 \oplus W_2$, 则称表达式 $V = W_1 \oplus W_2$ 为正交直和分解, 且有 $W_2 = W_1^\perp$.

注 1: 在直和分解中, 若子空间 W_1 给定, 则 W_2 不唯一; 在正交直和分解中, 若子空间 W_1 给定, 则 W_2 是唯一的且只能等于 W_1^\perp . 因此, 正交直和分解是唯一的.

例 1.5.5 在直角坐标系 $Oxyz$ 中, 假设 W_1 是位于 Ox 轴上所有向量的集合, W_2 是过坐标原点且不包括 Ox 轴的任一平面, 则 $\mathbb{R}^3 = W_1 \dot{+} W_2$ (直和分解不唯一). 倘若定义 $W_1^\perp = yOz$ 平面, 此时, $V = W_1 \oplus W_1^\perp$ (正交直和分解).

定理 1.5.3 设 $A = (a_{ij}) \in \mathbb{R}^{n \times n}$, 则 $N(A) \oplus R(A^T) = \mathbb{R}^n$ 且 $R(A^T) = [N(A)]^\perp$.

证明: 对任意向量 $y \in R(A^T)$, 存在向量 $z \in N(A)$ 使得 $y = A^T z$. 那么对任意向量 $x \in N(A)$ 和 $y \in R(A^T)$, 有

$$(x, y) = y^T x = z^T A x = 0$$

即 $N(A) \perp R(A^T)$.

另一方面, $N(A)$ 和 $R(A^T)$ 均是 \mathbb{R}^n 的子空间, 故 $N(A) + R(A^T)$ 也是 \mathbb{R}^n 的子空间. 又知

$$\dim [N(A) + R(A^T)] = \dim [N(A)] + \dim [R(A^T)]$$

$$\dim [R(A^T)] = \text{rank}(A)$$

故有

$$\dim [N(A) + R(A^T)] = \dim V$$

上式表明 $N(A) + R(A^T) = V$. 综上所述, $N(A) \oplus R(A^T) = \mathbb{R}^n$ 且 $R(A^T) = [N(A)]^\perp$.

注2:说明两子空间是正交直和常有如下思路:首先说明这两个子空间正交(由定理1.5.2知,两子空间正交意味着它们的和空间必为直和),然后再说明它们的和空间是 V 空间,或说明两个子空间的维数之和等于 V 空间的维数.请读者尝试采用不同方法证明本章的习题16和习题17,以加深对直和与正交直和的理解.

基于子空间的直和分解和正交直和分解,我们可以定义向量在子空间的投影和正交投影.

定义1.5.3(投影和正交投影) 设 W_1 和 W_2 是线性空间 V 的两个子空间且 $V=W_1+W_2$,对任意向量 $x \in V$ 均可唯一地分解成 $x=y+z$,其中 $y \in W_1, z \in W_2$,此时称向量 y 为向量 x 在 W_1 上的投影.特别地,若 $V=W_1 \oplus W_2$,则称向量 y 为向量 x 在 W_1 上的正交投影.

注3:之所以能利用(正交)直和分解定义向量(正交)投影的核心,原因就在于向量分解的唯一性.

例1.5.6 定义 \mathbb{R}^3 的线性子空间 $W=\text{span}(x_1, x_2)$,其中, $x_1=[2, 5, -1]^T$, $x_2=[-2, 1, 1]^T$.求向量 $y=[1, 2, 3]^T$ 在 W 和 W^\perp 的正交投影.

解:

方法1:由 $\dim W=2$ 知, $\dim W^\perp=1$,则 $W^\perp=\text{span}(x_3)$,其中 x_3 满足 $(x_1, x_3)=(x_2, x_3)=0$.经计算得 $x_3=[1, 0, 2]^T$.

又知,向量 y 在 W^\perp 上的正交投影可利用向量 y 与 x_3 的夹角 α 求得,即

$$\text{Proj}_{W^\perp} y = \|y\| \cos \alpha \frac{x_3}{\|x_3\|} = \frac{(y, x_3)}{\|x_3\|^2} x_3 = \left[\frac{7}{5}, 0, \frac{14}{5} \right]^T$$

利用直和分解的唯一性知,

$$\text{Proj}_W y = y - \text{Proj}_{W^\perp} y = \left[-\frac{2}{5}, 2, \frac{1}{5} \right]^T$$

方法2:向量 y 在向量 x_1 和 x_2 的正交投影分别为

$$\text{Proj}_{x_1} y = \frac{(y, x_1)}{\|x_1\|^2} x_1, \quad \text{Proj}_{x_2} y = \frac{(y, x_2)}{\|x_2\|^2} x_2$$

由于 x_1 和 x_2 正交(见图1.5.1),故

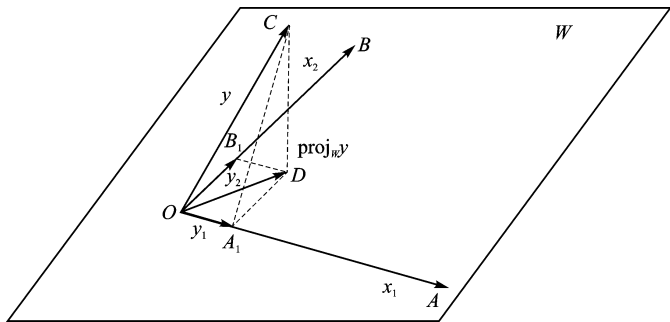


图 1.5.1

$$\text{Proj}_W \mathbf{y} = \frac{(\mathbf{y}, \mathbf{x}_1)}{\|\mathbf{x}_1\|^2} \mathbf{x}_1 + \frac{(\mathbf{y}, \mathbf{x}_2)}{\|\mathbf{x}_2\|^2} \mathbf{x}_2 = \left[-\frac{2}{5}, 2, \frac{1}{5} \right]^T$$

利用直和分解的唯一性求得

$$\text{Proj}_{W^\perp} \mathbf{y} = \mathbf{y} - \text{Proj}_W \mathbf{y} = \left[\frac{7}{5}, 0, \frac{14}{5} \right]^T$$

命题 1.5.1 若 W 是线性空间 V 的子空间, $\mathbf{x}_1, \dots, \mathbf{x}_n$ 是 W 的一组正交基. 对于 V 中任一向量 \mathbf{y} 均可唯一地表示为

$$\mathbf{y} = \text{Proj}_W \mathbf{y} + \text{Proj}_{W^\perp} \mathbf{y} \quad (1.5.1)$$

式中: $\text{Proj}_W \mathbf{y}$ 和 $\text{Proj}_{W^\perp} \mathbf{y}$ 分别为向量 \mathbf{y} 在空间 W 和补空间 W^\perp 上的正交投影, 且

$$\text{Proj}_W \mathbf{y} = \frac{(\mathbf{y}, \mathbf{x}_1)}{(\mathbf{x}_1, \mathbf{x}_1)} \mathbf{x}_1 + \dots + \frac{(\mathbf{y}, \mathbf{x}_n)}{(\mathbf{x}_n, \mathbf{x}_n)} \mathbf{x}_n \quad (1.5.2)$$

特别地, 若 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 是 W 的一组标准正交基, 则

$$\text{Proj}_W \mathbf{y} = (\mathbf{y}, \mathbf{x}_1) \mathbf{x}_1 + \dots + (\mathbf{y}, \mathbf{x}_n) \mathbf{x}_n \quad (1.5.3)$$

【应用】 现利用正交投影研究内积空间中的标准正交基的存在性问题. 已知 n 维内积空间的一组基, 能否利用这组基找到空间的一组标准正交基?

定理 1.5.4 有限维内积空间必存在标准正交基.

证明: 采用构造方法证明. 设 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 是内积空间 V 的一组基, 现依次构造正交向量组 $\mathbf{y}_1, \dots, \mathbf{y}_n$, 并将其单位化得 $\mathbf{z}_1, \dots, \mathbf{z}_n$ 使它们成为 V 的一组标准正交基.

为此, 定义

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{x}_1 \\ \mathbf{z}_1 &= \frac{\mathbf{y}_1}{\|\mathbf{y}_1\|} \end{aligned}$$

则 \mathbf{z}_1 是单位向量.

设向量 \mathbf{x}_2 在 \mathbf{z}_1 的正交投影为 $\text{Proj}_{\mathbf{z}_1} \mathbf{x}_2$, 则有

$$\text{Proj}_{\mathbf{z}_1} \mathbf{x}_2 = (\mathbf{x}_2, \mathbf{z}_1) \mathbf{z}_1$$

于是, 向量 $(\mathbf{x}_2 - \text{Proj}_{\mathbf{z}_1} \mathbf{x}_2)$ 必正交于向量 \mathbf{z}_1 . 为此, 定义

$$\begin{aligned} \mathbf{y}_2 &= \mathbf{x}_2 - \text{Proj}_{\mathbf{z}_1} \mathbf{x}_2 \\ \mathbf{z}_2 &= \frac{\mathbf{y}_2}{\|\mathbf{y}_2\|} \end{aligned}$$

则 $\mathbf{z}_1, \mathbf{z}_2$ 是单位正交向量组.

假设根据向量组 $\mathbf{x}_1, \dots, \mathbf{x}_{k-1}$ 已求得向量组 $\mathbf{z}_1, \dots, \mathbf{z}_{k-1}$ 是单位正交向量组. 现计算向量 \mathbf{z}_k 使得 $\mathbf{z}_1, \dots, \mathbf{z}_{k-1}, \mathbf{z}_k$ 是单位正交向量组. 设向量 \mathbf{x}_k 在 $\mathbf{z}_1, \dots, \mathbf{z}_{k-1}$ 的正交投影分别为 $\text{Proj}_{\mathbf{z}_1} \mathbf{x}_k, \dots, \text{Proj}_{\mathbf{z}_{k-1}} \mathbf{x}_k$, 则有

$$\text{Proj}_{\mathbf{z}_i} \mathbf{x}_k = (\mathbf{x}_k, \mathbf{z}_i) \mathbf{z}_i, \quad i = 1, \dots, k-1$$

由此可定义向量

$$\mathbf{y}_k = \mathbf{x}_k - \sum_{i=1}^{k-1} (\mathbf{x}_k, \mathbf{z}_i) \mathbf{z}_i$$

$$\mathbf{z}_k = \frac{\mathbf{y}_k}{\|\mathbf{y}_k\|}$$

则 $\mathbf{z}_1, \dots, \mathbf{z}_k$ 是单位正交向量组.

以此类推, 可构造单位正交向量组 $\mathbf{z}_1, \dots, \mathbf{z}_n$, 它们即是 V 的一组标准正交基. 证毕.

注 4: 上述证明方法提供了一种线性空间标准正交基的计算方法, 我们称为 Gram-Schmidt 正交化方法.

注 5: 设向量组 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 和 $\mathbf{z}_1, \dots, \mathbf{z}_n$ 均是 n 维内积空间 V 的一组基, 且由 Gram-Schmidt 正交化方法知, $\mathbf{y}_k = \mathbf{x}_k - \sum_{i=1}^{k-1} (\mathbf{x}_k, \mathbf{z}_i) \mathbf{z}_i, k=1, \dots, n$. 将其改写成矩阵形式

$$[\mathbf{x}_1, \dots, \mathbf{x}_n] = [\mathbf{z}_1, \dots, \mathbf{z}_n] \mathbf{A} \quad (1.5.4)$$

式中: \mathbf{A} 是由 $\mathbf{z}_1, \dots, \mathbf{z}_n$ 到基 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 的过渡矩阵, 定义为

$$\mathbf{A} = \begin{bmatrix} \|\mathbf{y}_1\| & (\mathbf{x}_2, \mathbf{z}_1) & \cdots & (\mathbf{x}_n, \mathbf{z}_1) \\ & \|\mathbf{y}_2\| & \cdots & (\mathbf{x}_n, \mathbf{z}_2) \\ & & \ddots & \vdots \\ 0 & & & \|\mathbf{y}_n\| \end{bmatrix}$$

例 1.5.7 已知 \mathbb{R}^4 中的一组基 $\mathbf{x}_1 = [1, 1, 0, 0]^T, \mathbf{x}_2 = [1, 0, 1, 0]^T, \mathbf{x}_3 = [-1, 0, 0, 1]^T, \mathbf{x}_4 = [1, -1, -1, 1]^T$, 求 \mathbb{R}^4 的一组标准正交基.

解: 令 $\mathbf{y}_1 = \mathbf{x}_1 = [1, 1, 0, 0]^T$, 则

$$\mathbf{z}_1 = \frac{1}{\|\mathbf{y}_1\|} \mathbf{y}_1 = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0 \right]^T$$

根据 Gram-Schmidt 正交化方法知

$$\mathbf{y}_2 = \mathbf{x}_2 - \frac{(\mathbf{x}_2, \mathbf{y}_1)}{(\mathbf{y}_1, \mathbf{y}_1)} \mathbf{y}_1 = \left[\frac{1}{2}, -\frac{1}{2}, 1, 0 \right]^T$$

$$\mathbf{y}_3 = \mathbf{x}_3 - \frac{(\mathbf{x}_3, \mathbf{y}_2)}{(\mathbf{y}_2, \mathbf{y}_2)} \mathbf{y}_2 - \frac{(\mathbf{x}_3, \mathbf{y}_1)}{(\mathbf{y}_1, \mathbf{y}_1)} \mathbf{y}_1 = \left[-\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 1 \right]^T$$

$$\mathbf{y}_4 = \mathbf{x}_4 - \frac{(\mathbf{x}_4, \mathbf{y}_1)}{(\mathbf{y}_1, \mathbf{y}_1)} \mathbf{y}_1 - \frac{(\mathbf{x}_4, \mathbf{y}_2)}{(\mathbf{y}_2, \mathbf{y}_2)} \mathbf{y}_2 - \frac{(\mathbf{x}_4, \mathbf{y}_3)}{(\mathbf{y}_3, \mathbf{y}_3)} \mathbf{y}_3 = [1, -1, -1, 1]^T$$

再依次进行单位化, 得

$$\mathbf{z}_2 = \left[\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, 0 \right]^T$$

$$\mathbf{z}_3 = \left[-\frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{3}{\sqrt{12}} \right]^T$$

$$\mathbf{z}_4 = \left[\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2} \right]^T$$

因此, z_1, z_2, z_3, z_4 是 \mathbb{R}^4 的一组标准正交基.

【应用】 正交直和分解的另一典型应用是求解最佳逼近问题.

定义 1.5.4(最佳逼近) 设 W 是线性空间 V 的非空子集, $\alpha \in V$ 为给定向量, 若存在 $x \in W$ 满足如下不等式:

$$\|\alpha - x\| \leq \|\alpha - y\|, \quad \forall y \in W$$

则称向量 x 是向量 α 在 W 的最佳逼近(最佳近似)向量.

定理 1.5.5(最佳逼近定理) 设 W 是有限维线性空间 V 的线性子空间, 则 V 中任一向量 x 在 W 上都有唯一的最佳逼近, 且 x 在 W 上的最佳逼近是 x 在 W 上的正交投影.

证明: 设 x 是 V 中任一向量, z 是 W 中任一向量, y 是 x 在 W 上的正交投影, 则

$$\|x - z\|^2 = \|(x - y) + (y - z)\|^2$$

注意到 $(y - z) \in W$ 且 $(x - y) \in W^\perp$, 故 $(x - y, y - z) = 0$, 即

$$\|x - z\|^2 = \|x - y\|^2 + \|y - z\|^2 \geq \|x - y\|^2$$

因此, y 是 x 在 W 上的最佳逼近.

现采用反证法证明最佳逼近的唯一性. 假设 $x^* \in W$ 也是 x 在 W 上的最佳逼近, 则 x^* 不是 x 在 W 上的正交投影(因为 x 在 W 上的正交投影是唯一的). 显然, W 中必存在一非零单位向量 α 使得 $(x - x^*, \alpha) \neq 0$. 不妨设 $(x - x^*, \alpha) = a$, 并定义 $\gamma = x^* + a\alpha$, 则 $\gamma \in W$ 且有

$$\|x - \gamma\|^2 = (x - x^* - a\alpha, x - x^* - a\alpha) = \|x - x^*\|^2 - |a|^2$$

又知 y 也是 x 在 W 上的最佳逼近, 且 $\forall z \in W$, 有 $\|x - z\|^2 \geq \|x - y\|^2$. 故当 $\gamma \in W$ 时,

$$\|x - \gamma\|^2 \geq \|x - y\|^2$$

显然, $\|x - x^*\|^2 \geq \|x - y\|^2 + |a|^2$. 注意到 $a \neq 0$, 故 $\|x - x^*\| > \|x - y\|$. 因此, x^* 不是 x 在 W 上的最佳逼近, 即 x 在 W 上的最佳逼近是唯一的. 证毕.

例 1.5.8(最小二乘问题) 在许多实际观测数据的处理问题中, 若已知变量 y 与变量 x_1, \dots, x_n 间呈线性关系, 即

$$y = \lambda_1 x_1 + \dots + \lambda_n x_n$$

但不知道系数 $\lambda_1, \dots, \lambda_n$. 为确定这些系数, 通常做 $m \geq n$ 次试验, 得到 m 组观测数据

$$(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}, y^{(k)}), \quad k = 1, \dots, m$$

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

$$\min_{c_k \in F} \sum_{k=1}^m \left| y^{(k)} - \sum_{i=1}^n c_i x_i^{(k)} \right|^2 = \sum_{k=1}^m \left| y^{(k)} - \sum_{i=1}^n \lambda_i x_i^{(k)} \right|^2$$

这就是最小二乘问题.

定义

$$a_i = \begin{bmatrix} x_i^{(1)} \\ x_i^{(2)} \\ \vdots \\ x_i^{(m)} \end{bmatrix} \in F^m, \quad b = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix} \in F^m$$

$$x = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \in F^n, \quad \lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} \in F^n$$

并记 $A = [a_1, a_2, \dots, a_n] \in F^{m \times n}$, 则最小二乘问题的矩阵形式为

$$\min_{x \in F^n} \|Ax - b\|^2 = \|A\lambda - b\|^2 \quad (1.5.5)$$

最小二乘问题也可以看作求解向量 b 在 $R(A)$ 空间上的最佳逼近问题. 后续章节将对这一问题从不同角度进行阐述.

【思考】 在最小二乘问题 1.5.8 中, 我们已知变量 y 与变量 x_1, \dots, x_n 间呈线性关系. 若已知变量 y 与变量 x_1, \dots, x_n 间呈非线性关系, 比如

$$y = \lambda_1 x_1 + \dots + \lambda_n x_n + \mu_1 x_1^2 + \dots + \mu_n x_n^2$$

此时, 能否利用式 (1.5.5) 求解待定系数 $\lambda_1, \dots, \lambda_n$ 和 μ_1, \dots, μ_n ?

1.6 应用: 多项式插值

在科学研究和工程中, 如汽车制造、机械工业等常常会遇到计算函数值等一类问题: 通过观测或实验得到一组数据, 确定了与自变量的某些点相应的函数值, 而需要获得未观察到的点的函数值. 当然, 这类问题的函数关系往往难以得到明显的解析表达式, 这需要我们构造一个适当的、比较简单的函数近似地代替寻求的函数. 这就是插值法, 用数学语言可表述如下:

设函数 $y = f(x)$ 定义在区间 $[a, b]$ 上, x_0, x_1, \dots, x_n 是 $[a, b]$ 上取定的 $(n+1)$ 个互异点, 且仅在这些点处的函数值 $y_i = f(x_i)$ 已知, 要构造一个函数 $g(x)$ 使得

$$g(x_i) = y_i, \quad i = 0, 1, \dots, n \quad (1.6.1)$$

且要求误差

$$r(x) = f(x) - g(x)$$

的绝对值在区间 $[a, b]$ 上比较小, 即函数 $g(x)$ 能较好地逼近 $f(x)$. 点 x_0, x_1, \dots, x_n 称为插值基点, 由插值基点确定的区间 (即以 $\min(x_0, x_1, \dots, x_n), \max(x_0, x_1, \dots, x_n)$ 为端点的区间) 称为插值区间, $f(x)$ 称为求插函数, $g(x)$ 称为插值函数, $r(x)$ 称为插值余项.

我们已知 $C[a, b]$ 是无限维线性空间, 要线性表示 $C[a, b]$ 中的一个连续函数

$f(x)$, 需给定该空间的一组基. 最常见的基莫过于 $1, x, x^2, \dots, x^n, \dots$. 对于足够光滑的函数 $f(x)$, 它总可以进行泰勒展开, 并可以用有限维多项式空间 $P_n(x)$ 中的一个多项式进行逼近. 基于此, 我们同样考虑用 $P_n(x)$ 的一个多项式化函数作为插值函数. 若选定 $P_n(x)$ 中的一组基为 $1, x, x^2, \dots, x^n$, 则

$$g(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n \quad (1.6.2)$$

将约束条件式(1.6.1)代入式(1.6.2), 得如下方程组:

$$\mathbf{D}^T \mathbf{a} = \mathbf{y} \quad (1.6.3)$$

式中: $\mathbf{a} = [a_0, a_1, \dots, a_n]^T$ 为待定向量, $\mathbf{y} = [y_0, y_1, \dots, y_n]^T$ 为给定向量, 系数矩阵 \mathbf{D} 定义为

$$\mathbf{D} = \begin{bmatrix} 1 & x_0 & \cdots & x_0^n \\ 1 & x_1 & \cdots & x_1^n \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \cdots & x_n^n \end{bmatrix}$$

在线性代数中, 形如 \mathbf{D} 的矩阵称为 Vardermode 矩阵, 它的行列式称为 Vardermode 行列式. 已知 x_0, x_1, \dots, x_n 是 $[a, b]$ 上取定的 $n+1$ 个互异点且

$$|\mathbf{D}^T| = |\mathbf{D}| = \prod_{0 \leq j < i \leq n} (x_i - x_j)$$

则 \mathbf{D} 是可逆矩阵. 相应地, 方程组(1.6.3)有唯一解 $\mathbf{a} = (\mathbf{D}^T)^{-1} \mathbf{y}$. 因此, 根据约束条件式(1.6.1), 可以唯一地确定插值函数 $g(x)$.

上述插值法看似简单, 但涉及矩阵求逆, 故计算比较烦琐. 为了避开矩阵求逆计算, 重新选取 $P_n(x)$ 空间中的一组基(本章的习题 10):

$$l_i(x) = \prod_{\substack{j=0, \\ j \neq i}}^n \frac{(x - x_j)}{(x_i - x_j)}, \quad i = 1, \dots, n \quad (1.6.4)$$

它们都是 n 次多项式, 称为 Lagrange 基本多项式. 显然, $l_i(x)$ 满足关系

$$l_i(x_k) = \begin{cases} 0, & k \neq i \\ 1, & k = i \end{cases}$$

令插值函数 $g(x)$ 满足如下表达:

$$g(x) = \sum_{i=0}^n b_i l_i(x) \quad (1.6.5)$$

并将约束条件式(1.6.1)代入式(1.6.5), 得

$$b_i = y_i, \quad i = 1, \dots, n$$

因此, 将

$$p_n(x) = \sum_{i=0}^n f(x_i) l_i(x)$$

称为 Lagrange 插值多项式. 显然, Lagrange 插值法比由式(1.6.2)确定的插值法简单得多.

采用多项式插值在低次时是比较简单方便的,在高次时则有数值不稳定的特点.因此,当插值基点较多时,自然设想将插值区间分段,然后采用低次插值并须保证在分点处的连续性,这就是局部化的分段插值.局部化的分段插值的一个严重缺点是插值函数在子区间的端点处不光滑,即导数不连续.对于一些实际问题,如汽车、飞机、船体外形的流线型设计,不但要求一阶导数连续,还要求二阶导数连续.此时就要采用样条插值.

样条本来是工程设计中用来描绘光滑曲线的简单工具,是一条富有弹性的细长木条.使用时,将它固定在一些给定点处,在其地方任其自然弯曲,并稍作调整,使样条具有满意的形状,然后沿样条画出曲线,称为样条曲线.样条函数就是对样条曲线进行数学模拟得到的.最常见的就是三次样条插值,即采用 $P_3(x)$ 空间的多项式函数作为插值函数.

设 $f(x)$ 是定义在区间 $[a, b]$ 上的二次连续可微函数,给定区间 $[a, b]$ 的一个划分(可由插值基点确定):

$$a = x_0 < x_1 < \cdots < x_n = b$$

则插值函数 $g(x)$ 在区间 $[a, b]$ 上可表示为

$$g(x) = \begin{cases} g_1(x), & x \in [x_0, x_1] \\ \vdots \\ g_i(x), & x \in [x_{i-1}, x_i] \\ \vdots \\ g_n(x), & x \in [x_{n-1}, x_n] \end{cases}$$

式中: $g_i(x) \in P_3(x)$ 且 $g(x)$ 满足式(1.6.1). 此时,插值函数 $g(x)$ 称为三次样条插值函数,简称三次样条.

由于三次样条 $g(x)$ 的二阶导数 $g''(x)$ 在每一个区间 $[x_{i-1}, x_i]$ ($i=0, 1, \cdots, n$) 上都是线性函数. 为讨论方便,我们采用 Lagrange 基本多项式定义 $g''(x)$, 得

$$g''_i(x) = m_{i-1} \frac{x_i - x}{h_i} + m_i \frac{x - x_{i-1}}{h_i}, \quad x \in [x_{i-1}, x_i] \quad (1.6.6)$$

式中: $h_i = x_i - x_{i-1}$, m_{i-1} 和 m_i 是待定系数.

根据要求, $g(x)$ 须满足以下条件: 对 $i=0, 1, \cdots, n$,

$$\begin{cases} g(x_i) = f(x_i) \\ g(x_i^-) = g(x_i^+) \\ g'(x_i^-) = g'(x_i^+) \\ g''(x_i^-) = g''(x_i^+) \end{cases} \quad (1.6.7)$$

由于 $g(x)$ 在每个区间上都是一个三次多项式,故须确定 4 个参数;由于一共有 n 个区间,故应确定 $4n$ 个参数. 由式(1.6.7)中的 4 个等式可依次确定 $n+1, n-1, n-1$ 和 $n-1$ 个方程,因此还须补充两个条件. 通常会根据具体问题,在区间的两个端点

处给出条件,称为边界条件.

例 1.6.1 观测得函数 $f(x)$ 在若干点的值见表 1.6.1. 求 $f(x)$ 的三次样条函数及 $f(3)$ 的近似值.

表 1.6.1 已有观测值

i	x_i	$f(x_i)$	$f'(x_i)$
1	0	0	8
2	2	16	—
3	4	36	—
4	6	54	—
5	10	82	7

解: 由式(1.6.7)列出方程组

$$\begin{bmatrix} 4 & 2 & 0 & 0 & 0 \\ 2 & 8 & 8 & 0 & 0 \\ 0 & 2 & 2 & 2 & 0 \\ 0 & 0 & 0 & 12 & 4 \\ 0 & 0 & 0 & 4 & 8 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ -1 \\ -2 \\ 0 \end{bmatrix}$$

解得 $m_1 = -1, m_2 = 2, m_3 = -1, m_4 = -1, m_5 = \frac{1}{2}$. 因此可得

$$g(x) = \begin{cases} 8 - \frac{1}{2}x^2 + \frac{1}{4}x^3, & x \in [0, 2] \\ 16 + 9(x-2) + (x-2)^2 - \frac{1}{4}(x-2)^3, & x \in [2, 4] \\ 36 + 10(x-4) - \frac{1}{2}(x-4)^2, & x \in [4, 6] \\ 54 + 8(x-6) - \frac{1}{2}(x-6)^2 + \frac{1}{16}(x-6)^3, & x \in [6, 10] \end{cases}$$

相应地, $g(3) = 25.75$.

对插值法感兴趣的读者,可阅读参考文献[15-16].

本章习题

1. 设 V 是数域 F 上的线性空间,证明:

- (1) 零向量是唯一的;
- (2) 任一向量的负向量是唯一的;
- (3) 对任意 $k \in F$ 和 $\alpha \in V, 0\alpha = \theta, (-1)\alpha = -\alpha, k\theta = \theta$;

(4) 若 $k\alpha = \theta$, 则 $k=0$ 或 $\alpha=\theta$.

2. 设 A 是 n 阶实方阵, 若在实数域上定义通常实矩阵的加法和数乘运算, 则下列集合不是实线性空间的是:

- a. $\{A \mid A \text{ 的行列式为 } 0\}$ b. $\{A \mid A \text{ 是对角元素为 } 0 \text{ 的矩阵}\}$
c. $\{A \mid A \text{ 是上三角矩阵}\}$ d. $\{A \mid A \text{ 是对称矩阵}\}$

3. 定义 \mathbb{R}^+ 上的加法与数乘运算如下:

$$\begin{aligned}\oplus: \alpha, \beta \in \mathbb{R}^+, \alpha \oplus \beta &= \alpha\beta \\ \circ: \alpha \in \mathbb{R}^+, k \in \mathbb{R}, k \circ \alpha &= \alpha^k\end{aligned}$$

试判断 $(\mathbb{R}^+, \oplus, \circ)$ 是否为线性空间.

4. 证明数集 $\mathbb{Q}(\sqrt{3})$ 是其自身上的线性空间.

5. 若 W 是有限维线性空间 V 的子空间, 且 $\dim W = \dim V$, 证明 $W = V$.

6. 令 H 是所有形如 $[a-3b, b-a, a, b]^T$ 的向量的集合, 其中, a 和 b 是任意实数, 证明 H 是 \mathbb{C}^4 的一个线性子空间.

7. 在 $\mathbb{R}^{2 \times 3}$ 中, 求子空间 $W = \left\{ \begin{bmatrix} x & y & 0 \\ z & 0 & t \end{bmatrix} \mid x+y+z=0, x, y, z, t \in \mathbb{R} \right\}$ 的维数和

一组基.

8. 设 $A \in \mathbb{C}^{n \times n}$, 记集合 $W = \{X \in \mathbb{C}^{n \times n} \mid AX = XA\}$.

(1) 证明 W 是 $\mathbb{C}^{n \times n}$ 的一个子空间;

(2) 当 $A = \text{diag}(1, 2, \dots, n)$ 时, 求 W 的维数和一组基.

9. 设 $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$, 求 $N(A) \cap R(A)$ 和 $N(A) + R(A)$.

10. 设 x_0, x_1, \dots, x_n 是 $[a, b]$ 上取定的 $n+1$ 个已知互异点, 定义

$$l_i(x) = \prod_{\substack{j=0, \\ j \neq i}}^n \frac{(x - x_j)}{(x_i - x_j)}, \quad i = 1, \dots, n$$

则 $l_i(x), i = 1, \dots, n$ 是 $P_n(x)$ 空间中的一组基.

11. 取 $P_2(t)$ 的一组基为 $1+t, t-2, t^2$, 求多项式 $p(t) = 2t^2 - t + 1$ 在该组基下的坐标.

12. 在 $\mathbb{R}^{2 \times 2}$ 中求向量 $A = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}$ 在基 $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ 下的坐标.

13. 设多项式空间 $P_2(x)$, 证明 $1, (x-1), (x-1)^2$ 是 $P_2(x)$ 的一组基; 并求从基 $1, x, x^2$ 到基 $1, (x-1), (x-1)^2$ 的过渡矩阵.

14. 设 $A = \begin{bmatrix} 0 & 22 & 2 \\ -1 & 45 & 3 \end{bmatrix}$, 求 $N(A)$ 和 $R(A)$ 的维数与基.

15. 设齐次线性方程组 $\mathbf{Ax}=\mathbf{0}$, 其中 $\mathbf{A} \in \mathbb{R}^{40 \times 42}$, $\mathbf{x} \in \mathbb{R}^{42}$. 若该方程组的基础解系由两个线性无关的解向量构成, 请判断非齐次线性方程组 $\mathbf{Ax}=\mathbf{b}$ 是否有解?

16. 证明命题 1.3.1.

17. 设 $P_2(x)$ 是 \mathbb{R} 上的线性空间, 定义函数

$$(f, g) = \int_{-1}^1 f(x)g(x)dx, \quad f(x), g(x) \in P_2(x)$$

(1) 证明 $P_2(x)$ 是欧氏空间;

(2) 求 $P_2(x)$ 关于基 $1, x, x^2$ 的度量矩阵;

(3) 试计算 $f(x)=1-x+x^2$ 和 $g(x)=1-4x-5x^2$ 的内积.

18. 设 t_0, \dots, t_n 是不同实数, $P_n(x)$ 是 \mathbb{R} 上的多项式空间. 定义

$$(f, g) = \sum_{i=0}^n f(t_i)g(t_i), \quad f(x), g(x) \in P_n(x)$$

(1) 证明 $P_n(x)$ 是欧氏空间;

(2) 设 $t_0=-2, t_1=-1, t_3=0, t_4=1, t_5=2$, 求 $\text{span}(1, x, x^2)$ 的一组标准正交基;

(3) 求 $f(x)=5-\frac{1}{2}x^4$ 到空间 $\text{span}(1, x, x^2)$ 的最佳逼近向量.

19. 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 是 Hermite 矩阵, 证明:

(1) \mathbf{A} 的所有特征值均为实数;

(2) 若 \mathbf{A} 是正定矩阵, 则 \mathbf{A} 的所有特征值均为正实数.

20. 设 $V = \{a \cos t + b \sin t \mid a, b \in \mathbb{R}\}$. 对任意 $f, g \in V$, 定义

$$(f, g) = f(0)g(0) + f\left(\frac{\pi}{2}\right)g\left(\frac{\pi}{2}\right)$$

(1) 证明 V 是二维实线性空间;

(2) 证明 (f, g) 是 V 上的内积;

(3) 求 $h(t)=3\cos(t+7)+4\sin(t+9)$ 的长度.

21. 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 且满足 $\mathbf{A}^2=\mathbf{A}$, 证明 $\mathbb{C}^n = N(\mathbf{A}) + N(\mathbf{A}-\mathbf{I})$.

22. 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 且满足 $\mathbf{A}^2=\mathbf{A}=\mathbf{A}^H$, 证明 $\mathbb{C}^n = N(\mathbf{A}) \oplus N(\mathbf{A}-\mathbf{I})$.

23. \mathbb{R}^n 中的点 \mathbf{y} 到 \mathbb{R}^n 的子空间 W 的距离定义为从 \mathbf{y} 到空间 W 中的点的最小距离. 试求 \mathbf{y} 到空间 $W = \{\mathbf{x}_1, \mathbf{x}_2\}$ 的距离, 其中

$$\mathbf{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \quad \mathbf{x}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

若选取 $\mathbf{x}_2 = [1, 2, 1]^T$, 则重新求该距离.

24. 设 $\zeta_1, \zeta_2, \zeta_3, \zeta_4, \zeta_5$ 是欧氏空间 \mathbb{R}^5 的一组标准正交基, 令 $\alpha_1 = \zeta_1 + \zeta_5, \alpha_2 = \zeta_1 - \zeta_2 + \zeta_4, \alpha_3 = 2\zeta_1 + \zeta_2 + \zeta_3$. 求 $W = \text{span}(\alpha_1, \alpha_2, \alpha_3)$ 的一组标准正交基.

25. 已知数据点 $(2,1), (5,2), (7,3), (8,3)$, 求最小二乘直线方程 $y = \mu_0 + \mu_1 t$.

26. 为测量飞机起飞性能, 飞机的水平位置从 $t=0$ 到 $t=10$ s 每秒测量一次, 具体位置(英尺)是: 0, 8.8, 29.9, 62.0, 104.7, 159.1, 222.0, 294.5, 380.4, 471.1, 571.7. 求出这组数据的最小二乘立方曲线

$$y = \mu_0 + \mu_1 t + \mu_2 t^2 + \mu_3 t^3$$

27. 请编写三次样条函数的程序.

28. 酒精测试仪通过测量驾驶员呼出的气体酒精含量来判断是否饮酒. 半导体传感器因响应速度快、驱动简单、成本低而在酒精测试仪中应用. 但它通常受环境温度影响较大. 这需要在实际应用中通过算法进行补偿. 已知环境温度为 $30\text{ }^{\circ}\text{C}$ 时的酒精浓度值为 0.211 mg/L . 在不同温度下的浓度值如表 1-1 所列, 试给出温度为 $30\text{ }^{\circ}\text{C}$ 、酒精浓度值为 0.211 mg/L 在 $20\sim 50\text{ }^{\circ}\text{C}$ 间的浓度值.

表 1-1 不同温度下的酒精浓度值

温度/ $^{\circ}\text{C}$	浓度/ $(\text{mg} \cdot \text{L}^{-1})$	温度/ $^{\circ}\text{C}$	浓度/ $(\text{mg} \cdot \text{L}^{-1})$
20	0.301	35	0.189
25	0.248	40	0.163
30	0.211	50	0.037

第 2 章

线性映射与矩阵

在线性代数中,我们引入矩阵简洁地表示线性方程组 $\mathbf{Ax}=\mathbf{b}$, $\mathbf{A}\in F^{m\times n}$. 对于二元或三元线性方程组,其解集有直观的几何意义,即若干直线或平面的交集;多元情况也有类似的意义. 当然,我们也可以从映射角度重新理解线性方程组 $\mathbf{Ax}=\mathbf{b}$: 将 $x\rightarrow\mathbf{Ax}$ 看成是一种从 F^n 到 F^m 的映射,向量 \mathbf{b} 是映射的像,向量 \mathbf{x} 是向量 \mathbf{b} 的原像. 由此,解方程 $\mathbf{Ax}=\mathbf{b}$ 就是求 F^n 中像为 \mathbf{b} 的原像 \mathbf{x} . 显然,矩阵 \mathbf{A} 与映射 $T:F^n\rightarrow F^m$ 之间有着特殊的天然联系. 本章的目的之一就是揭示二者之间的关系.

2.1 预备知识

定义 2.1.1(映射) 设 V 和 W 是两个非空集合,如果存在一个 V 到 W 的对应法则 f ,使得 V 中每一个元素 x 都有 W 中唯一的一个元素 y 与之对应,则称 f 是 V 到 W 的一个映射,记为 $y=f(x)$. 元素 $y\in W$ 称为元素 $x\in V$ 在映射 f 下的像,称 x 为 y 的原像. 集合 V 称为映射 f 的定义域. 当 V 中元素改变时, x 在映射 f 下的像的全体作为 W 的一个子集,称为映射 f 的值域,记为 $R(f)$.

我们通常用记号

$$f:V\rightarrow W$$

抽象地表示 f 是 V 到 W 的一个映射;而用记号

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

表示映射 f 所规定的元素之间的具体对应关系.

定义 2.1.2(单射、满射与双射) 设 V 和 W 是两个非空集合, f 是 V 到 W 的一个映射. 若对任意 $x_1, x_2\in V$, 当 $x_1\neq x_2$ 时有 $f(x_1)\neq f(x_2)$, 则称 f 是 V 到 W 的单映射(简称单射);若对任意 $y\in W$ 都有一个元素 $x\in V$ 使得 $f(x)=y$ (即 $R(f)=W$), 则称 f 是 V 到 W 的满映射(简称满射);若映射 f 既是单映射又是满映射, 则称 f 是 V 到 W 的一一映射或双映射(简称双射).

定义 2.1.3(映射相等) 设 f_1 是集合 V_1 到集合 W_1 的一个映射, f_2 是集合 V_2 到集合 W_2 的一个映射. 若 $V_1 = V_2, W_1 = W_2$, 并且对任意 $x \in V_1$ 有 $f_1(x) = f_2(x)$, 则称映射 f_1 和 f_2 相等, 记为 $f_1 = f_2$.

定义 2.1.4(映射乘积) 设 V_1, V_2 和 V_3 是三个非空集合, 并设 f_1 是 V_1 到 V_2 的一个映射, f_2 是 V_2 到 V_3 的一个映射. 由 f_1 和 f_2 确定的 V_1 到 V_3 的映射 $f_3: x \rightarrow f_2(f_1(x)), x \in V_1$, 称为映射 f_1 和 f_2 的乘积, 记为 $f_3 = f_2 \cdot f_1$, 或简写为 $f_3 = f_2 f_1$.

一般而言, 映射 f_1 和 f_2 的乘积不具有交换律, 即映射 f_1 和 f_2 的乘积与映射 f_2 和 f_1 的乘积不同.

定义 2.1.5(可逆映射) 设有映射 $f_1: V \rightarrow W$, 若存在映射 $f_2: W \rightarrow V$ 使得

$$f_2 \cdot f_1 = I_V, \quad f_1 \cdot f_2 = I_W$$

式中: $I_V: x \rightarrow x, x \in V$ 为 V 上的恒等映射, I_W 是 W 上的恒等映射. 我们称 f_2 为 f_1 的逆映射, 记为 f_1^{-1} . 若映射 f_1 有逆映射, 则称 f_1 为可逆映射.

定理 2.1.1 设映射 $f: V \rightarrow W$ 是可逆的, 则 f 的逆映射 f^{-1} 是唯一的.

定理 2.1.2 设映射 $f: V \rightarrow W$ 是可逆映射的充分必要条件是 f 是双射.

定义 2.1.6(变换) 设 V 是一个非空集合, V 到自身的映射称为 V 的变换; V 到自身的双射称为 V 的一一变换; 若 V 是有限集, V 的一一变换称为 V 的置换.

定义 2.1.7(一元多项式) 设 F 是数域, n 是自然数, λ 是一个文字(或符号), 形式表达式

$$g(\lambda) = a_n \lambda^n + a_{n-1} \lambda^{n-1} + \cdots + a_1 \lambda + a_0 \quad (2.1.1)$$

式中: $a_i \in F (i=0, 1, \cdots, n)$, 称为数域 F 上的一元多项式. 如果 $a_n \neq 0$, 则称 $a_n \lambda^n$ 为多项式的首项, n 称为 $g(\lambda)$ 的次数, 记为 $\deg(g(\lambda)) = n$, a_n 称为 $g(\lambda)$ 的首项系数. 若 $a_n = 1$, 则称 $g(\lambda)$ 为首 1 多项式.

对于多项式 $g(\lambda) = \sum_{i=0}^n a_{n-i} \lambda^{n-i}$ 和 $h(\lambda) = \sum_{i=0}^m b_{m-i} \lambda^{m-i} (n \geq m)$, 则 $g(\lambda)$ 与 $h(\lambda)$ 的加减、乘积分别定义为

$$g(\lambda) \pm h(\lambda) = \sum_{i=0}^n (a_{n-i} \pm b_{n-i}) \lambda^{n-i}$$

$$g(\lambda) h(\lambda) = \sum_{i=0}^{n+m} c_{n+m-i} \lambda^{n+m-i}$$

式中: $c_k = \sum_{i+j=k} a_i b_j, b_n = b_{n-1} = \cdots = b_{m+1} = 0$.

由上述定义易知, 两多项式经过加、减、乘运算后所得结果仍是数域 F 上的多项式, 且加法和乘法分别适合交换律、结合律, 乘法对加法的分配律和乘法消去律. 一般来说, 两多项式不能随便作除法, 但有如下定理.

定理 2.1.3 设 $g(\lambda)$ 和 $h(\lambda)$ 是数域 F 上的多项式且 $h(\lambda) \neq 0$, 则存在唯一的多项式 $p(\lambda)$ 和 $q(\lambda)$ 使得

$$g(\lambda) = p(\lambda)h(\lambda) + q(\lambda)$$

式中, $q(\lambda) = 0$ 或 $\deg(q(\lambda)) < \deg(h(\lambda))$.

定义 2.1.8(多项式整除) 设 $g(\lambda)$ 和 $h(\lambda)$ 是数域 F 上的多项式, 如果存在多项式 $p(\lambda)$ 使得 $g(\lambda) = p(\lambda)h(\lambda)$, 则称多项式 $h(\lambda)$ 整除 $g(\lambda)$, 记为 $h(\lambda) \mid g(\lambda)$, $h(\lambda)$ 是 $g(\lambda)$ 的因式, $g(\lambda)$ 是 $h(\lambda)$ 的倍式.

定义 2.1.9(公因式) 设 $g(\lambda)$ 和 $h(\lambda)$ 是数域 F 上的多项式, 若多项式 $p(\lambda)$ 既是 $g(\lambda)$ 的因式, 又是 $h(\lambda)$ 的因式, 则称 $p(\lambda)$ 是 $g(\lambda)$ 与 $h(\lambda)$ 的公因式. 若 $d(\lambda)$ 是 $g(\lambda)$ 与 $h(\lambda)$ 的公因式, 且 $g(\lambda)$ 与 $h(\lambda)$ 的任一公因式都是 $d(\lambda)$ 的因式, 则称 $d(\lambda)$ 是 $g(\lambda)$ 与 $h(\lambda)$ 的一个最大公因式.

定义 2.1.10(公倍式) 设 $g(\lambda)$ 和 $h(\lambda)$ 是数域 F 上的多项式, 若多项式 $p(\lambda)$ 既是 $g(\lambda)$ 的倍式, 又是 $h(\lambda)$ 的倍式, 则称 $p(\lambda)$ 是 $g(\lambda)$ 与 $h(\lambda)$ 的公倍式. 若 $d(\lambda)$ 是 $g(\lambda)$ 与 $h(\lambda)$ 的公倍式, 且 $g(\lambda)$ 与 $h(\lambda)$ 的任一公倍式都是 $d(\lambda)$ 的倍式, 则称 $d(\lambda)$ 是 $g(\lambda)$ 与 $h(\lambda)$ 的一个最小公倍式.

若无特殊说明, 最大公因式和最小公倍式都是指首 1 多项式.

定义 2.1.11(友矩阵) 设 $f(\lambda)$ 是数域 F 上的首 1 多项式, 其表达式为

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

定义 n 阶矩阵

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix} \quad \text{或} \quad A = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{bmatrix}$$

称为多项式 $f(\lambda)$ 的友矩阵(或伴侣矩阵).

定理 2.1.4 设多项式 $f(\lambda)$ 的友矩阵为 A , 则 A 的特征多项式为 $f(\lambda)$.

2.2 线性映射

定义 2.2.1(线性映射) 设 V 和 W 是数域 F 上的线性空间, 如果映射 $T: V \rightarrow W$ 满足下述性质:

- (1) 可加性: $\forall x, y \in V, T(x+y) = T(x) + T(y)$;
- (2) 齐次性: $\forall \lambda \in F, T(\lambda x) = \lambda T(x)$,

则称 T 为 V 到 W 的一个线性映射. 特别地, 当 $V=W$ 时, 映射 T 为 V 到自身的线性映射, 称 T 为 V 上的线性变换(或线性算子).

例 2.2.1 设 V 是数域 F 上的线性空间, 定义映射 $T: V \rightarrow V$, 分别满足

- (1) $T(x) = \theta, \forall x \in V$;
- (2) $T(x) = x, \forall x \in V$;

$$(3) T(\mathbf{x}) = -\mathbf{x}, \forall \mathbf{x} \in V,$$

则以上三个映射均为线性变换,分别称为零变换、恒等变换和负变换.

例 2.2.2 定义映射 $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2, \forall \mathbf{x} = [x_1, x_2]^T \in \mathbb{R}^2$, 它分别满足

$$(1) T(\mathbf{x}) = \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix} \mathbf{x}, k_1 \text{ 和 } k_2 \text{ 为正常数};$$

$$(2) T(\mathbf{x}) = [x_1, -x_2];$$

$$(3) T(\mathbf{x}) = \begin{bmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{bmatrix} \mathbf{x}, \varphi \text{ 为旋转角(逆时针取正)},$$

则以上三个映射均为线性变换,分别称为平面伸压变换、平面反射变换和平面旋转变换,它们是常见的图形变换.

例 2.2.3 在多项式空间 $P_n(x)$ 中,定义映射 $T: P_n \rightarrow P_n$ 满足

$$T(p(x)) = \frac{dp(x)}{dx}, \quad \forall p(x) \in P_n(x)$$

则映射 T 是 $P_n(x)$ 上的线性变换,称为微分变换(或微分算子).

例 2.2.4 在 $C[a, b]$ 空间中,定义映射 $T: C[a, b] \rightarrow C[a, b]$ 满足

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

则映射 T 是 $C[a, b]$ 上的线性变换,称为积分变换(或积分算子).

例 2.2.5 设 W 是线性空间 V 的非平凡子空间,定义映射 T 为

$$T(\mathbf{x}) = \text{Proj}_W \mathbf{x}, \quad \forall \mathbf{x} \in V$$

则映射 T 是 V 上的线性变换,称为正交投影变换.

解: 设向量组 $\alpha_1, \dots, \alpha_p$ 是 W 的一组标准正交基,则 V 中任一向量 \mathbf{x} 在 W 上的正交投影为

$$\text{Proj}_W \mathbf{x} = (\mathbf{x}, \alpha_1) \alpha_1 + \dots + (\mathbf{x}, \alpha_p) \alpha_p$$

因此,

$$T(\mathbf{x}) = \sum_{i=1}^p (\mathbf{x}, \alpha_i) \alpha_i$$

那么,对任意向量 $\mathbf{x}, \mathbf{y} \in V$ 和 $\lambda, \mu \in F$ 有

$$\begin{aligned} T(\lambda \mathbf{x} + \mu \mathbf{y}) &= \sum_{i=1}^p (\lambda \mathbf{x} + \mu \mathbf{y}, \alpha_i) \alpha_i \\ &= \lambda \sum_{i=1}^p (\mathbf{x}, \alpha_i) \alpha_i + \mu \sum_{i=1}^p (\mathbf{y}, \alpha_i) \alpha_i \\ &= \lambda T(\mathbf{x}) + \mu T(\mathbf{y}) \end{aligned}$$

综上所述,映射 T 是 V 上的线性变换.

例 2.2.6 设 V 是数域 F 上的线性空间, $\varepsilon_1, \dots, \varepsilon_n$ 和 η_1, \dots, η_n 是 V 的两组基,定义映射 $T: V \rightarrow V$ 为

$$T(x) = [\eta_1, \dots, \eta_n] [x_1, \dots, x_n]^T, \quad \forall x \in V$$

式中: $[x_1, \dots, x_n]^T$ 是向量 x 在基 $\varepsilon_1, \dots, \varepsilon_n$ 下的坐标. 根据定义 2.2.1 易证, 映射 T 是 V 上的线性变换.

例 2.2.7 设 $\alpha_1, \alpha_2, \alpha_3$ 是线性空间 V 的一组基, 定义映射 T 为

$$T(k_1\alpha_1 + k_2\alpha_2 + k_3\alpha_3) = (k_1 + k_2 + k_3)\alpha_1 + (k_2 + k_3)\alpha_2 + k_3\alpha_3$$

式中: $k_1, k_2, k_3 \in F$.

(1) 证明映射 T 是 $V \rightarrow V$ 的线性变换.

(2) 若 $\beta_0 = 2\alpha_1 + \alpha_2 + 3\alpha_3$, 求 $T(\beta_0)$.

解: (1) 首先证明映射 T 是 $V \rightarrow V$ 的线性变换. 对任意向量 $x, y \in V$, 令 $x = x_1\alpha_1 + x_2\alpha_2 + x_3\alpha_3, y = y_1\alpha_1 + y_2\alpha_2 + y_3\alpha_3$, 则有

$$\begin{aligned} T(x+y) &= T[(x_1+y_1)\alpha_1 + (x_2+y_2)\alpha_2 + (x_3+y_3)\alpha_3] \\ &= [(x_1+y_1) + (x_2+y_2) + (x_3+y_3)]\alpha_1 + \\ &\quad [(x_2+y_2) + (x_3+y_3)]\alpha_2 + (x_3+y_3)\alpha_3 \\ &= T(x) + T(y) \end{aligned}$$

对任意向量 $x \in V$ 和 $k \in F$, 有

$$\begin{aligned} T(kx) &= T(kx_1\alpha_1 + kx_2\alpha_2 + kx_3\alpha_3) \\ &= (kx_1 + kx_2 + kx_3)\alpha_1 + (kx_2 + kx_3)\alpha_2 + kx_3\alpha_3 \\ &= kT(x) \end{aligned}$$

因此, 映射 T 是 $V \rightarrow V$ 的线性变换.

(2) 求解 $T(\beta_0)$.

方法一: 令 $x_1=2, x_2=1, x_3=3$, 则

$$\begin{aligned} T(\beta_0) &= (2+1+3)\alpha_1 + (1+3)\alpha_2 + 3\alpha_3 \\ &= 6\alpha_1 + 4\alpha_2 + 3\alpha_3 \end{aligned}$$

方法二: 对任意 $x = [x_1, x_2, x_3]^T \in F^3$, 有

$$T(x_1\alpha_1 + x_2\alpha_2 + x_3\alpha_3) = [\alpha_1, \alpha_2, \alpha_3] \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} x$$

由于 β_0 在基 $\alpha_1, \alpha_2, \alpha_3$ 下的坐标为 $x = [2, 1, 3]^T$, 将其代入上式得

$$T(\beta_0) = [\alpha_1, \alpha_2, \alpha_3] \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = 6\alpha_1 + 4\alpha_2 + 3\alpha_3$$

例 2.2.8 考察在其自身的线性空间 $\mathbb{Q}(\sqrt{3})$, 定义映射 T 为

$$T(x + y\sqrt{3}) = x, \quad \forall x, y \in \mathbb{Q}$$

则映射 T 不是线性变换. 这是因为 $T(\sqrt{3} \cdot \sqrt{3}) = 3 \neq \sqrt{3}T(\sqrt{3})$.

定理 2.2.1 设 T 是数域 F 上线性空间 V 到 W 的线性映射, 若 $\alpha_1, \dots, \alpha_p$ 是 V

的一组向量, $k_1, \dots, k_p \in F$, 则

$$T(k_1\alpha_1 + \dots + k_p\alpha_p) = k_1T(\alpha_1) + \dots + k_pT(\alpha_p)$$

该定理可由定义 2.2.1 直接得到, 并有如下推论:

推论 2.2.1 设 T 是线性空间 V 到 W 的线性映射, 则

$$(1) T(\theta) = \theta', \theta \in V, \theta' \in W;$$

$$(2) T(-x) = -x, \forall x \in V;$$

(3) 若 $\alpha_1, \dots, \alpha_p$ 是 V 中一组线性相关向量, 则 $T(\alpha_1), \dots, T(\alpha_p)$ 是 W 中一组线性相关向量;

(4) 若 $T(\alpha_1), \dots, T(\alpha_p)$ 是 W 的一组线性无关向量, 则 $\alpha_1, \dots, \alpha_p$ 是 V 中一组线性无关向量.

注 1: 推论 2.2.1 性质(1)的几何意义是线性映射必须保持原点不动. 因此解析几何中常见的平移变换一般不是线性变换.

定理 2.2.2 设 T 是数域 F 上线性空间 V 到 W 的线性映射, 当且仅当 T 是单射时, V 中线性无关向量组的像是 W 中线性无关向量组.

证明: 充分性. 设 $\alpha_1, \dots, \alpha_p$ 是 V 中一组线性无关向量, 则对任一不全为零的数组 $k_1, \dots, k_p \in F$, 有 $k_1\alpha_1 + \dots + k_p\alpha_p \neq \theta$. 由于 T 是单射且 $T(\theta) = \theta$, 从而

$$T(k_1\alpha_1 + \dots + k_p\alpha_p) = k_1T(\alpha_1) + \dots + k_pT(\alpha_p) \neq \theta$$

上式说明 $T(\alpha_1), \dots, T(\alpha_p)$ 是 W 的一组线性无关向量.

必要性. 设 ζ_1, \dots, ζ_n 是 V 中一组基, 则 V 中任意向量 α 和 β 可表示为

$$\alpha = a_1\zeta_1 + \dots + a_n\zeta_n, \quad \beta = b_1\zeta_1 + \dots + b_n\zeta_n$$

式中, $a = [a_1, \dots, a_n]^T$ 和 $b = [b_1, \dots, b_n]^T$ 分别是向量 α 和 β 在基 ζ_1, \dots, ζ_n 下的坐标. 于是

$$T(\alpha) = a_1T(\zeta_1) + \dots + a_nT(\zeta_n)$$

$$T(\beta) = b_1T(\zeta_1) + \dots + b_nT(\zeta_n)$$

且有

$$T(\alpha) - T(\beta) = (a_1 - b_1)T(\zeta_1) + \dots + (a_n - b_n)T(\zeta_n)$$

当 $\alpha \neq \beta$ 时, 若有 $T(\alpha) = T(\beta)$, 则根据上式得

$$(a_1 - b_1)a_1T(\zeta_1) + \dots + (a_n - b_n)T(\zeta_n) = \theta$$

由于 $T(\zeta_1), \dots, T(\zeta_n)$ 线性无关, 故 $a_i = b_i, i=1, \dots, n$, 即 $a = b$. 由于 $\alpha \neq \beta$, 故其坐标 $a \neq b$. 因此, 当 $\alpha \neq \beta$ 时, $T(\alpha) \neq T(\beta)$. 这表明映射 T 是单射. 证毕.

推论 2.2.2 设线性空间 V 和 W 的维数相同, 且 T 是线性空间 V 到 W 的线性映射, 当且仅当 T 是单射时, V 中一组基的像是 W 中一组基. 此时映射 T 是双射.

若无特殊说明, 我们用 $\mathcal{L}(V, W)$ 表示线性空间 V 到 W 的所有线性映射的集合; 用 $\mathcal{L}(V, V)$ 表示定义在线性空间 V 的所有线性变换的集合, 简记为 $\mathcal{L}(V)$.

定义 2.2.2 (线性映射的加法运算) 设 $T_1, T_2 \in \mathcal{L}(V, W)$, 定义 T_1 与 T_2 的和为

$$(T_1 + T_2)(x) = T_1(x) + T_2(x), \quad \forall x \in V$$

式中: 等式右端的“+”表示线性空间 W 的加法运算.

定义 2.2.3(线性映射的数乘运算) 设 $T \in \mathcal{L}(V, W)$, $\lambda \in F$, 定义 λ 与 T 的数乘 $\lambda \cdot T$ 为

$$(\lambda T)(x) = \lambda \cdot T(x), \quad \forall x \in V$$

式中: 等式右端的“ \cdot ”表示线性空间 W 的数乘运算, 常省略.

定理 2.2.3 集合 $\mathcal{L}(V, W)$ 对定义 2.2.2 的加法和定义 2.2.3 的数乘构成数域 F 上的线性空间, 称为线性映射空间. 特别地, $\mathcal{L}(V)$ 称为线性变换空间.

证明: 对任意映射 $T_1, T_2 \in \mathcal{L}(V, W)$, 任意向量 $x, y \in V$ 和 $\lambda, \mu \in F$, 有

$$\begin{aligned} (T_1 + T_2)(\lambda x + \mu y) &= T_1(\lambda x + \mu y) + T_2(\lambda x + \mu y) \\ &= \lambda T_1(x) + \mu T_1(y) + \lambda T_2(x) + \mu T_2(y) \\ &= \lambda [T_1(x) + T_2(x)] + \mu [T_1(y) + T_2(y)] \\ &= \lambda (T_1 + T_2)(x) + \mu (T_1 + T_2)(y) \end{aligned}$$

因此, $T_1 + T_2$ 是 V 到 W 的线性映射. 同理可证, 对任意 $T \in \mathcal{L}(V, W)$ 和 $\lambda \in F$, λT 是 V 到 W 的线性映射.

我们还可验证线性映射的加法符合交换律和结合律, 零映射是零元, $(-T)$ 是映射 T 的负元; 并容易验证线性映射的数乘运算满足四条规则, 即对任意映射 $T_1, T_2, T \in \mathcal{L}(V, W)$ 和 $\lambda, \mu \in F$, 有: (1) $\lambda(T_1 + T_2) = \lambda T_1 + \lambda T_2$; (2) $(\lambda + \mu)T = \lambda T + \mu T$; (3) $\lambda(\mu T) = (\lambda\mu)T$; (4) $1 \cdot T = T$. 综上所述, $\mathcal{L}(V, W)$ 是数域 F 上的线性空间.

【思考】 线性空间 $\mathcal{L}(V, W)$ 的维数是多少?

此处请读者自行思考, 下一节将给出详细解答.

有两类重要的线性子空间伴随线性映射, 分别为线性映射的值空间和核空间.

定理 2.2.4(线性映射值空间和核空间) 设 $T \in \mathcal{L}(V, W)$, 定义

$$\begin{aligned} N(T) &= \{x \in V \mid T(x) = \theta\} \\ R(T) &= \{y \in W \mid y = T(x), \forall x \in V\} \end{aligned}$$

则 $N(T)$ 是 V 的子空间, $R(T)$ 是 W 的子空间. 我们称 $N(T)$ 是线性映射 T 的核空间(或零空间), $R(T)$ 是线性映射 T 的像空间(或值空间); 并称 $\dim N(T)$ 为线性映射 T 的零度(或亏), $\dim R(T)$ 为线性映射 T 的秩.

证明: (1) 当 $x = \theta$ 时, $Tx = \theta$. 故 $\theta \in R(T)$, 即 $R(T)$ 是 W 的非空子集.

对任意向量 $\alpha, \beta \in R(T)$ 和 $k \in F$, 有

$$\begin{aligned} T(\alpha) + T(\beta) &= T(\alpha + \beta) \in R(T) \\ kT(\alpha) &= T(k\alpha) \in R(T) \end{aligned}$$

由线性子空间定义可知, $R(T)$ 为 W 的子空间.

(2) 因为 $T(\theta) = \theta$, 故 $\theta \in N(T)$, 即 $N(T)$ 是 V 的非空子集. 对任意向量 $\alpha, \beta \in N(T)$ 和 $k \in F$, 有

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

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

因此, $N(T)$ 是 V 的子空间.

定理 2.2.5 (亏加秩定理) 设 $T \in \mathcal{L}(V, W)$, 则

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

即线性映射 T 的亏加秩等于其定义域 V 空间的维数.

证明: 设线性空间 V 的维数为 n , 其子空间 $N(T)$ 的维数为 m . 在 $N(T)$ 中取一组基 $\alpha_1, \dots, \alpha_m$, 并把它扩充为 V 的基 $\alpha_1, \dots, \alpha_m, \alpha_{m+1}, \dots, \alpha_n$, 则对任意向量 $x \in V$ 有

$$x = \sum_{i=1}^n k_i \alpha_i, \quad k_i \in F$$

从而

$$Tx = \sum_{i=1}^n k_i T(\alpha_i)$$

即

$$R(T) = \text{span}(T(\alpha_1), \dots, T(\alpha_m), T(\alpha_{m+1}), \dots, T(\alpha_n)) \quad (2.2.1)$$

注意到 $\alpha_1, \dots, \alpha_m \in N(T)$, 故 $T(\alpha_1) = \dots = T(\alpha_m) = \theta$. 由此, 式 (2.2.1) 可改写为

$$R(T) = \text{span}(T(\alpha_{m+1}), \dots, T(\alpha_n)) \quad (2.2.2)$$

现只需证明 $T(\alpha_{m+1}), \dots, T(\alpha_n)$ 线性无关, 即可说明 $T(\alpha_{m+1}), \dots, T(\alpha_n)$ 是 $R(T)$ 的一组基.

假设存在一组数 $l_i \in F, i = m+1, \dots, n$, 使得

$$\sum_{i=m+1}^n l_i T(\alpha_i) = \theta$$

则根据线性映射定义知

$$T\left(\sum_{i=m+1}^n l_i \alpha_i\right) = \theta$$

即

$$\sum_{i=m+1}^n l_i \alpha_i \in N(T)$$

又知向量组 $\alpha_1, \dots, \alpha_m$ 是 $N(T)$ 的一组基, 故向量 $\sum_{i=m+1}^n l_i \alpha_i$ 一定可由向量组 $\alpha_1, \dots, \alpha_m$ 线性表示, 即

$$\sum_{i=m+1}^n l_i \alpha_i = \sum_{i=1}^m l_i \alpha_i$$

整理得

$$\sum_{i=1}^n l_i \alpha_i = \theta$$

已知向量组 $\alpha_1, \dots, \alpha_m, \alpha_{m+1}, \dots, \alpha_n$ 线性无关, 故上式成立的充分必要条件为 $l_i = 0, i = 1, \dots, n$. 因此, 我们推断出 $T(\alpha_{m+1}), \dots, T(\alpha_n)$ 线性无关, 即 $\dim R(T) = n - m$. 证毕.

【思考】 $\forall A \in \mathbb{R}^{m \times n}$, 定义映射 $T: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{n \times m}$ 使得

$$T(A) = A^T$$

问: 映射 T 是线性映射吗? 请读者思考并说明原因.

2.3 矩阵与同构

现引入矩阵来研究线性映射和线性变换.

定义 2.3.1(矩阵) 设 V 和 W 均是数域 F 上的线性空间, 向量组 $\varepsilon_1, \dots, \varepsilon_n$ 和向量组 η_1, \dots, η_m 分别是 V 和 W 的一组基, 且 $T \in \mathcal{L}(V, W)$. 由此, $T(\varepsilon_i)$ 一定可以由基 η_1, \dots, η_m 线性表示. 不妨设为

$$\begin{cases} T(\varepsilon_1) = a_{11}\eta_1 + a_{21}\eta_2 + \dots + a_{m1}\eta_m \\ T(\varepsilon_2) = a_{12}\eta_1 + a_{22}\eta_2 + \dots + a_{m2}\eta_m \\ \vdots \\ T(\varepsilon_n) = a_{1n}\eta_1 + a_{2n}\eta_2 + \dots + a_{mn}\eta_m \end{cases}$$

上式可形式地记为

$$T(\varepsilon_1, \dots, \varepsilon_n) \stackrel{\text{def}}{=} [T(\varepsilon_1), \dots, T(\varepsilon_n)] = [\eta_1, \dots, \eta_m] A \quad (2.3.1)$$

其中

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \in F^{m \times n}$$

矩阵 A 称为线性映射 T 在 V 的基 $\varepsilon_1, \dots, \varepsilon_n$ 和 W 的基 η_1, \dots, η_m 下的矩阵. 特别地, 若 $V=W$, 式(2.3.1)改写为

$$T(\varepsilon_1, \dots, \varepsilon_n) \stackrel{\text{def}}{=} [T(\varepsilon_1), \dots, T(\varepsilon_n)] = [\varepsilon_1, \dots, \varepsilon_n] A \quad (2.3.2)$$

其中

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

矩阵 A 称为线性变换 T 在 V 的基 $\varepsilon_1, \dots, \varepsilon_n$ 下的矩阵.

例 2.3.1 定义平面旋转变换 $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$, 满足

$$T(\mathbf{x}) = \begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix} \mathbf{x}, \quad \forall \mathbf{x} = [x_1, x_2]^T \in \mathbb{R}^2$$

取定 \mathbb{R}^2 中的标准基 $\mathbf{e}_1, \mathbf{e}_2$, 则有

$$\begin{aligned} T(\mathbf{e}_1) &= \begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \cos \varphi \\ \sin \varphi \end{bmatrix} = [\mathbf{e}_1, \mathbf{e}_2] \begin{bmatrix} \cos \varphi \\ \sin \varphi \end{bmatrix} \\ T(\mathbf{e}_2) &= \begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\sin \varphi \\ \cos \varphi \end{bmatrix} = [\mathbf{e}_1, \mathbf{e}_2] \begin{bmatrix} -\sin \varphi \\ \cos \varphi \end{bmatrix} \end{aligned}$$

根据式(2.3.2)得

$$T(\mathbf{e}_1, \mathbf{e}_2) = [\mathbf{e}_1, \mathbf{e}_2] \begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix}$$

因此, 平面旋转变换 T 在标准基 $\mathbf{e}_1, \mathbf{e}_2$ 下的矩阵为

$$\mathbf{A} = \begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix}$$

例 2.3.2 在多项式空间 $P_n(x)$ 中, 定义微分变换 $T: P_n \rightarrow P_n$:

$$T(p(x)) = \frac{dp(x)}{dx}, \quad \forall p(x) \in P_n(x)$$

取定 $P_n(x)$ 的一组基 $1, x, x^2, \dots, x^n$, 则有

$$\begin{aligned} T(1) &= 0 = [1, x, x^2, \dots, x^n] [0, 0, \dots, 0]^T \\ T(x) &= 1 = [1, x, x^2, \dots, x^n] [1, 0, \dots, 0]^T \\ T(x^2) &= 2x = [1, x, x^2, \dots, x^n] [0, 2, \dots, 0]^T \\ &\vdots \\ T(x^n) &= nx^{n-1} = [1, x, x^2, \dots, x^n] [0, 0, \dots, n, 0]^T \end{aligned}$$

于是在微分变换 T 基 $1, x, x^2, \dots, x^n$ 下的矩阵为

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & n \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

例 2.3.3 设 V 是数域 F 上的线性空间, $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$ 和 $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n$ 是 V 的两组基, 定义线性变换 $T: V \rightarrow V$ 为

$$T(\mathbf{x}) = [\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n] [x_1, \dots, x_n]^T, \quad \forall \mathbf{x} \in V$$

式中: $[x_1, \dots, x_n]^T$ 是 \mathbf{x} 在基 $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$ 下的坐标. 求 T 在基 $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$ 下的矩阵.

解: 由 T 的定义式知

$$T(\boldsymbol{\varepsilon}_i) = [\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n] \mathbf{e}_i$$

于是

$$T(\varepsilon_1, \cdots, \varepsilon_n) = [\eta_1, \cdots, \eta_m] \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$

即 $T(\varepsilon_1, \cdots, \varepsilon_n) = [\eta_1, \cdots, \eta_m]$.

设矩阵 A 是基 $\varepsilon_1, \cdots, \varepsilon_n$ 到基 η_1, \cdots, η_m 的过渡矩阵, 则

$$[\eta_1, \cdots, \eta_m] = [\varepsilon_1, \cdots, \varepsilon_n]A$$

因此,

$$T(\varepsilon_1, \cdots, \varepsilon_n) = [\varepsilon_1, \cdots, \varepsilon_n]A$$

上式表明, T 在基 $\varepsilon_1, \cdots, \varepsilon_n$ 下的矩阵恰好为基 $\varepsilon_1, \cdots, \varepsilon_n$ 到基 η_1, \cdots, η_m 的过渡矩阵.

显然, 矩阵 A 由线性映射 T 唯一确定. 或者说, 矩阵是线性映射在某组基下的表达(或坐标化的线性变换). 反过来, 若给定 $m \times n$ 阶矩阵 A , 是否存在线性映射 $T \in \mathcal{L}(V, W)$ 使其在给定基下的矩阵恰为 A ? 若存在, 我们可以找到多少个线性映射?

定理 2.3.1 设 V 和 W 是数域 F 上的线性空间, 取定 $\varepsilon_1, \cdots, \varepsilon_n$ 和 η_1, \cdots, η_m 分别是 V 和 W 的一组基. 任取 $A = (a_{ij}) \in F^{m \times n}$, 则有且仅有一个线性映射 $T \in \mathcal{L}(V, W)$ 使其在 V 的基 $\varepsilon_1, \cdots, \varepsilon_n$ 和 W 的基 η_1, \cdots, η_m 下的矩阵恰为 A .

证明: V 中任意向量 x 均可由基 $\varepsilon_1, \cdots, \varepsilon_n$ 线性表示, 不妨设

$$x = \sum_{j=1}^n \alpha_j \varepsilon_j$$

式中: $\alpha = [\alpha_1, \cdots, \alpha_n]^T$ 是向量 x 在基 $\varepsilon_1, \cdots, \varepsilon_n$ 下的坐标.

定义映射 $T: V \rightarrow W$ 使得对 V 中任意向量 x 满足

$$T(x) = \sum_{j=1}^n \sum_{i=1}^m \alpha_j a_{ij} \eta_i$$

则对任意向量 $x, y \in V$ 和 $\lambda, \mu \in F$, 有

$$\begin{aligned} T(\lambda x + \mu y) &= \sum_{j=1}^n \sum_{i=1}^m (\lambda \alpha_j + \mu \beta_j) a_{ij} \eta_i \\ &= \lambda \sum_{j=1}^n \sum_{i=1}^m \alpha_j a_{ij} \eta_i + \mu \sum_{j=1}^n \sum_{i=1}^m \beta_j a_{ij} \eta_i \\ &= \lambda T(x) + \mu T(y) \end{aligned}$$

式中: $\beta = [\beta_1, \cdots, \beta_n]^T$ 是向量 y 在基 $\varepsilon_1, \cdots, \varepsilon_n$ 下的坐标.

因此, T 是线性映射. 现证明映射 T 在 V 的基 $\varepsilon_1, \cdots, \varepsilon_n$ 和 W 的基 η_1, \cdots, η_m 下的矩阵为 A . 由映射 T 定义式知, $\forall k=1, \cdots, n$, 则

$$T(\varepsilon_k) = \sum_{i=1}^m a_{ik} \eta_i = [\eta_1, \cdots, \eta_m] \begin{bmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{bmatrix}$$

于是

$$[T(\boldsymbol{\varepsilon}_1), \dots, T(\boldsymbol{\varepsilon}_n)] = [\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_m] \mathbf{A}$$

上式表明映射 T 在 V 的基 $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$ 和 W 的基 $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_m$ 下的矩阵为 \mathbf{A} .

为证明线性映射 T 的唯一性, 假设存在另一个线性映射 \tilde{T} , 且它在 V 的基 $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$ 和 W 的基 $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_m$ 下的矩阵也为 \mathbf{A} . 根据矩阵 \mathbf{A} 的定义, 有

$$\tilde{T}(\boldsymbol{\varepsilon}_k) = [\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_m] \begin{bmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{bmatrix}, \quad \forall k = 1, \dots, n$$

即 $T(\boldsymbol{\varepsilon}_k) = \tilde{T}(\boldsymbol{\varepsilon}_k), k = 1, \dots, n$. 因此, $T = \tilde{T}$ (本章的习题 5), 即线性映射 T 是唯一的. 证毕.

定理 2.3.1 表明线性映射 $T \in \mathcal{L}(V, W)$ 和 $\mathbf{A} = (a_{ij}) \in F^{m \times n}$ 存在着——对应的关系, 即存在着双射 $f: \mathcal{L}(V, W) \rightarrow F^{m \times n}$ 满足

$$f(T) = \mathbf{A}$$

实际上, 映射 f 不仅是双射, 还是线性映射, 具体原因如下:

设 $T_1, T_2 \in \mathcal{L}(V, W)$, 并令 T_1 和 T_2 在 V 的基 $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$ 和 W 的基 $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_m$ 下的矩阵分别为 \mathbf{A}_1 和 \mathbf{A}_2 , 则对任意 $\lambda, \mu \in F$, 有 $\lambda T_1 + \mu T_2$ 是线性映射, 且

$$\begin{aligned} (\lambda T_1 + \mu T_2)(\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n) &= [(\lambda T_1 + \mu T_2)(\boldsymbol{\varepsilon}_1), \dots, (\lambda T_1 + \mu T_2)(\boldsymbol{\varepsilon}_n)] \\ &= [\lambda T_1(\boldsymbol{\varepsilon}_1) + \mu T_2(\boldsymbol{\varepsilon}_1), \dots, \lambda T_1(\boldsymbol{\varepsilon}_n) + \mu T_2(\boldsymbol{\varepsilon}_n)] \\ &= \lambda [T_1(\boldsymbol{\varepsilon}_1), \dots, T_1(\boldsymbol{\varepsilon}_n)] + \mu [T_2(\boldsymbol{\varepsilon}_1), \dots, T_2(\boldsymbol{\varepsilon}_n)] \\ &= \lambda [\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_m] \mathbf{A}_1 + \mu [\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_m] \mathbf{A}_2 \\ &= [\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_m] (\lambda \mathbf{A}_1 + \mu \mathbf{A}_2) \end{aligned}$$

上式表明线性映射 $\lambda T_1 + \mu T_2$ 在 V 的基 $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$ 和 W 的基 $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_m$ 下的矩阵为 $\lambda \mathbf{A}_1 + \mu \mathbf{A}_2$, 即

$$f(\lambda T_1 + \mu T_2) = \lambda \mathbf{A}_1 + \mu \mathbf{A}_2$$

基于此, 我们引入同构映射定义.

定义 2.3.2 (同构映射) 设 V 和 W 是数域 F 上的线性空间, 若存在双射 $f: V \rightarrow W$ 满足

$$(1) f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y});$$

$$(2) f(\lambda \mathbf{x}) = \lambda f(\mathbf{x}).$$

式中: \mathbf{x} 和 \mathbf{y} 是 V 中任意向量, λ 是数域 F 的任意数, 则称 f 是 V 到 W 的同构映射, 并称线性空间 V 与 W 同构.

定理 2.3.2 设 V 和 W 是数域 F 上的线性空间, 它们的维数分别为 n 和 m , 则线性映射空间 $\mathcal{L}(V, W)$ 和矩阵空间 $F^{m \times n}$ 同构.

命题 2.3.1 (同构映射的性质) 设 V 和 W 是数域 F 上的线性空间, $T: V \rightarrow W$ 是同构映射, 则

$$(1) T(\boldsymbol{\theta}) = \boldsymbol{\theta}', \boldsymbol{\theta} \in V, \boldsymbol{\theta}' \in W;$$

(2) $T(-x) = -T(x), \forall x \in V$;

(3) $T(\sum \alpha_i x_i) = \sum \alpha_i T(x_i), \forall \alpha_i \in F \text{ 和 } x_i \in V$;

(4) V 的向量组 x_1, \dots, x_r 线性相关, 当且仅当其像 $T(x_1), \dots, T(x_r)$ 线性相关;

(5) 若 $\varepsilon_1, \dots, \varepsilon_n$ 是 V 的一组基, 则 $T(\varepsilon_1), \dots, T(\varepsilon_n)$ 是 W 的一组基;

(6) T 的逆映射 $T^{-1}: W \rightarrow V$ 存在且是同构映射.

证明: 由于同构映射 T 是线性映射, 故性质(1)~(4)由推论 2.2.1 可得. 现证明性质(5). 由定理 2.2.2 知, 若 $\varepsilon_1, \dots, \varepsilon_n$ 是 V 的一组基, 则向量组 $T(\varepsilon_1), \dots, T(\varepsilon_n)$ 必线性无关. 又知对任意向量 $y \in W$, 必存在 $x \in V$ 使得 $T(x) = y$, 其中 x 可由基 $\varepsilon_1, \dots, \varepsilon_n$ 线性表示为 $x = \sum_{j=1}^n \alpha_j \varepsilon_j$. 由此,

$$y = T\left(\sum_{j=1}^n \alpha_j \varepsilon_j\right) = \sum_{j=1}^n \alpha_j T(\varepsilon_j)$$

即 W 中任意向量 y 总可由向量组 $T(\varepsilon_1), \dots, T(\varepsilon_n)$ 线性表示. 于是, $T(\varepsilon_1), \dots, T(\varepsilon_n)$ 是 W 的一组基. 这同时表明线性空间 W 的维数与线性空间 V 的维数相同.

性质(6). 由定理 2.1.1 和定理 2.1.2 知, 同构映射存在唯一的逆映射 T^{-1} 且它也是双射. 于是, 对任意向量 $x, y \in V$ 和 $\lambda, \mu \in F$, 有

$$T^{-1}(T(x)) = x$$

$$T^{-1}(T(y)) = y$$

$$T^{-1}(T(\lambda x + \mu y)) = \lambda x + \mu y$$

进一步, 得

$$T^{-1}(\lambda T(x) + \mu T(y)) = \lambda T^{-1}(T(x)) + \mu T^{-1}(T(y))$$

即 T^{-1} 是线性映射. 因此, T^{-1} 是同构映射.

定理 2.3.3 线性空间同构当且仅当它们的维数相等.

证明: 必要性由命题 2.3.1 性质(5)证得. 这里只证明充分性. 设 V 和 W 均是数域 F 上的 n 维线性空间, 向量组 $\varepsilon_1, \dots, \varepsilon_n$ 和 η_1, \dots, η_n 分别是 V 和 W 的一组基.

定义映射 $T: V \rightarrow W$ 满足

$$T(x) = \sum_{i=1}^n \alpha_i \eta_i, \quad \forall x \in V$$

式中: $\alpha = [\alpha_1, \dots, \alpha_n]^T$ 是向量 x 在基 $\varepsilon_1, \dots, \varepsilon_n$ 下的坐标.

由此, 对任意向量 $x, y \in V$ 和 $\lambda, \mu \in F$, 有

$$T(\lambda x + \mu y) = \sum_{i=1}^n (\lambda \alpha_i + \mu \beta_i) \eta_i = \lambda T(x) + \mu T(y)$$

式中: $\beta = [\beta_1, \dots, \beta_n]^T$ 是向量 y 在基 $\varepsilon_1, \dots, \varepsilon_n$ 下的坐标. 上式表明 T 是线性映射.

若 $T(x) = T(y)$, 即

$$T(x) - T(y) = \sum_{i=1}^n (\alpha_i - \beta_i) \eta_i = \theta$$

则有 $\alpha = \beta$, 即 $x = y$. 于是当 $x \neq y$ 时, $T(x) \neq T(y)$, 这表明线性映射 T 是单射.

对任意向量 $z \in W$ 有

$$z = \sum_{i=1}^n \gamma_i \eta_i$$

式中: $\gamma = [\gamma_1, \dots, \gamma_n]^T$ 是向量 z 在基 η_1, \dots, η_n 下的坐标.

定义

$$\tilde{x} = \sum_{j=1}^n \gamma_j \varepsilon_j$$

则有 $\tilde{x} \in V$, 且 $T(\tilde{x}) = z$. 这表明线性映射 T 是满射.

综上所述, 映射 T 是同构映射, 即线性空间 V 和 W 是同构的. 证毕.

推论 2.3.1 任一实(复) n 维线性空间均与 \mathbb{R}^n (\mathbb{C}^n) 同构.

推论 2.3.2 设 V 和 W 是数域 F 上的线性空间, 它们维数分别为 n 和 m , 则

$$\dim(L(V, W)) = \dim(F^{m \times n}) = mn.$$

推论 2.3.3 设 V 是数域 \mathbb{R} (或 \mathbb{C}) 上的 n 维线性空间, 则线性变换空间 $L(V)$ 与 $\mathbb{R}^{n \times n}$ (或 $\mathbb{C}^{n \times n}$) 同构.

线性空间同构的概念比较抽象, 下面我们分别用定理并结合实例说明线性空间同构的意义.

定理 2.3.4 设映射 T 是 n 维线性空间 V 到 m 维线性空间 W 的线性映射, T 在 V 的基 $\varepsilon_1, \dots, \varepsilon_n$ 和 W 的基 η_1, \dots, η_m 下的矩阵为 A . 对任意向量 $x \in V$, 有

$$\beta = A\alpha$$

式中: $\beta \in F^m$ 和 $\alpha \in F^n$ 分别是像 $T(x)$ 和原像 x 在 V 的基 $\varepsilon_1, \dots, \varepsilon_n$ 和 W 的基 η_1, \dots, η_m 下的坐标.

证明: 对任意向量 $x \in V$, 定义

$$x = \sum_{j=1}^n \alpha_j \varepsilon_j$$

式中: $\alpha = [\alpha_1, \dots, \alpha_n]^T$ 是向量 x 在基 $\varepsilon_1, \dots, \varepsilon_n$ 下的坐标.

根据矩阵 A 的定义知

$$\begin{aligned} T(x) &= \sum_{j=1}^n \alpha_j T(\varepsilon_j) = [T(\varepsilon_1), \dots, T(\varepsilon_n)] \alpha \\ &= [\eta_1, \dots, \eta_m] A \alpha \end{aligned}$$

即向量 $A\alpha$ 是像 $T(x)$ 在基 η_1, \dots, η_m 下的坐标. 由于向量在同一组基下的坐标是唯一的, 故 $\beta = A\alpha$.

例 2.3.4 考察 \mathbb{R} 上的多项式空间, 定义微分映射 $T: P_3 \rightarrow P_2$:

$$T(p(x)) = \frac{dp(x)}{dx}, \quad \forall p(x) \in P_3(x)$$

取定 $P_3(x)$ 的一组基 $1, x, x^2, x^3$, 则 T 在 $P_3(x)$ 的一组基 $1, x, x^2, x^3$ 和 $P_2(x)$ 的一组基 $1, x, x^2$ 下的矩阵为

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}_{3 \times 4}$$

对于多项式空间 $P_3(x)$ 中任一多项式 $p(x) = a_0 + a_1x + a_2x^2 + a_3x^3, a_i \in \mathbb{R} (i=0, 1, 2, 3)$, 则有

$$T(p(x)) = a_1 + 2a_2x + 3a_3x^2 = [1, x, x^2] \begin{bmatrix} a_1 \\ 2a_2 \\ 3a_3 \end{bmatrix}$$

由于 $\dim(P_3(x)) = 4$, 故多项式空间 $P_3(x)$ 和 \mathbb{R}^4 同构. 不妨设 f 是 $P_3(x)$ 到 \mathbb{R}^4 的同构映射, 则

$$f(p(x)) = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} \stackrel{\text{def}}{=} \alpha$$

式中: $\alpha \in \mathbb{R}^4$ 是向量 $p(x)$ 在 $P_3(x)$ 的一组基 $1, x, x^2, x^3$ 下的坐标.

又知

$$A\alpha = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} a_1 \\ 2a_2 \\ 3a_3 \end{bmatrix} \stackrel{\text{def}}{=} \beta \in \mathbb{R}^3$$

式中: $\beta \in \mathbb{R}^3$ 是 $p(x)$ 的像 $T(p(x))$ 在 $P_2(x)$ 的一组基 $1, x, x^2$ 下的坐标.

上述过程如图 2.3.1 所示.

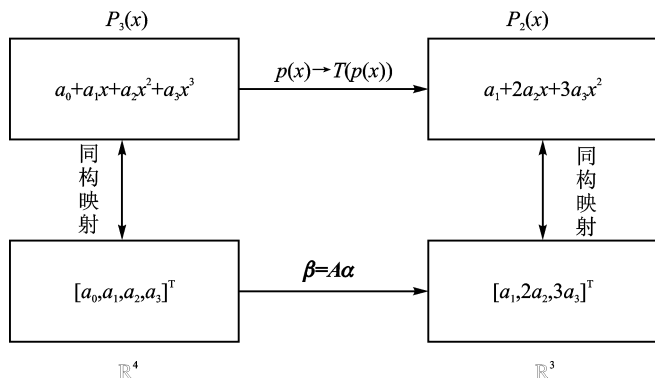


图 2.3.1 线性映射与矩阵的关系

由此, $m \times n$ 阶矩阵和 n 维向量相乘意味着某一 n 维线性空间的向量在进行线性变换; 反过来, 任一 n 维线性空间内的线性运算都可以等价于 \mathbb{R}^n (\mathbb{C}^n) 空间的向量的线性运算 (n 维线性空间与 \mathbb{R}^n 或 \mathbb{C}^n 同构); 任两线性空间定义的线性映射都可以等价于 $\mathbb{R}^{m \times n}$ ($\mathbb{C}^{m \times n}$) 中的矩阵与 \mathbb{R}^n (\mathbb{C}^n) 空间的向量作乘法运算 (线性映射空间与其对应的矩阵空间同构). 基于此, 我们可视同构空间为同一个线性空间.

我们已经了解到若取定线性空间中一组基, 则矩阵和线性映射是一一对应的. 一般而言, 同一线性映射在不同基下的矩阵是不同的. 下面回答同一线性映射在不同基下的矩阵之间的关系.

定理 2.3.5 设 V 和 W 是数域 F 上的 n 维和 m 维线性空间, $\varepsilon_1, \dots, \varepsilon_n$ 和 $\varepsilon'_1, \dots, \varepsilon'_n$ 是 V 的两组基, 由 $\varepsilon_1, \dots, \varepsilon_n$ 到 $\varepsilon'_1, \dots, \varepsilon'_n$ 的过渡矩阵为 Q ; η_1, \dots, η_m 和 η'_1, \dots, η'_m 是 W 的两组基, 由 η_1, \dots, η_m 到 η'_1, \dots, η'_m 的过渡矩阵为 P ; 设线性映射 $T \in \mathcal{L}(V, W)$ 在 V 的基 $\varepsilon_1, \dots, \varepsilon_n$ 和 W 的基 η_1, \dots, η_m 下的矩阵为 A , T 在 V 的基 $\varepsilon'_1, \dots, \varepsilon'_n$ 和 W 的基 η'_1, \dots, η'_m 下的矩阵为 B , 则

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

证明: 根据已知条件知

$$T(\varepsilon_1, \dots, \varepsilon_n) = [\eta_1, \dots, \eta_m]A$$

$$T(\varepsilon'_1, \dots, \varepsilon'_n) = [\eta'_1, \dots, \eta'_m]B$$

$$[\varepsilon'_1, \dots, \varepsilon'_n] = [\varepsilon_1, \dots, \varepsilon_n]Q$$

$$[\eta'_1, \dots, \eta'_m] = [\eta_1, \dots, \eta_m]P$$

又知 (见本章的习题 7)

$$T(\varepsilon'_1, \dots, \varepsilon'_n) = T(\varepsilon_1, \dots, \varepsilon_n)Q$$

则

$$T(\varepsilon'_1, \dots, \varepsilon'_n) = [\eta_1, \dots, \eta_m]AQ = [\eta'_1, \dots, \eta'_m]P^{-1}AQ$$

由于线性映射 T 的矩阵由基唯一确定, 所以

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

注 1: 矩阵 $A \in F^{m \times n}$ 经过有限次初等行变换或初等列变换变成矩阵 B , 则称矩阵 A 与 B 相抵 (或等价), 记为 $A \cong B$. 上式表明线性映射在不同基下的矩阵是相抵的; 反过来, 若两矩阵相抵, 则意味着它们是同一线性映射在不同基下的矩阵.

推论 2.3.4 设 V 是数域 F 上的 n 维线性空间, $\varepsilon_1, \dots, \varepsilon_n$ 和 $\varepsilon'_1, \dots, \varepsilon'_n$ 是 V 的两组基, 由 $\varepsilon_1, \dots, \varepsilon_n$ 到 $\varepsilon'_1, \dots, \varepsilon'_n$ 的过渡矩阵为 P , 线性变换 $T \in \mathcal{L}(V)$ 在基 $\varepsilon_1, \dots, \varepsilon_n$ 和基 $\varepsilon'_1, \dots, \varepsilon'_n$ 下的矩阵分别为 A 和 B , 则

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

注 2: 相似矩阵反映的是同一个线性变换. 故相似矩阵所具有的共同性质是线性变换所具有的.

例 2.3.5 设平面旋转变换 $T \in \mathcal{L}(\mathbb{R}^2)$, 其定义为

$$T(\mathbf{x}) = \begin{bmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{bmatrix} \mathbf{x}, \quad \forall \mathbf{x} \in \mathbb{R}^2$$

式中: φ 为旋转角. 经计算知,

$$T(\mathbf{e}_1, \mathbf{e}_2) = [\mathbf{e}_1, \mathbf{e}_2] \begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix}$$

故 T 在 \mathbb{R}^2 标准基 $\mathbf{e}_1, \mathbf{e}_2$ 下的矩阵为

$$\mathbf{A} = \begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix}$$

若选取向量 $\boldsymbol{\zeta}_1 = [\cos \varphi, \sin \varphi]^T$ 和 $\boldsymbol{\zeta}_2 = [-\sin \varphi, \cos \varphi]^T$ 为 \mathbb{R}^2 的一组基, 则有

$$T(\boldsymbol{\zeta}_1) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$T(\boldsymbol{\zeta}_2) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$T(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2) = I_2 = [\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2] \begin{bmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{bmatrix}$$

故 T 在 \mathbb{R}^2 基 $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2$ 下的矩阵为

$$\mathbf{B} = \begin{bmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{bmatrix}$$

定义从基 $\mathbf{e}_1, \mathbf{e}_2$ 到 \mathbb{R}^2 基 $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2$ 的过渡矩阵 \mathbf{P} 为

$$\mathbf{P} = \begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix}$$

则 $\mathbf{B} = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$ 成立.

我们引入两个重要的子空间: 线性映射的值空间和核空间. 在线性映射空间的同构空间——矩阵空间 $F^{m \times n}$ 中, 我们同样有矩阵的值空间(列空间)和核空间(零空间), 这里讨论它们之间的关系.

定理 2.3.6 设 V 和 W 是数域 F 上的 n 维和 m 维线性空间, 若 $T \in \mathcal{L}(V, W)$ 在 V 的基 $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$ 和 W 的基 $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_m$ 下的矩阵为 \mathbf{A} , 则

- (1) $\dim N(T) = \dim N(\mathbf{A})$;
- (2) $\dim R(T) = \dim R(\mathbf{A}) = \text{rank}(\mathbf{A})$;
- (3) $\dim N(\mathbf{A}) + \dim R(\mathbf{A}) = n$. (亏加秩)

证明: 任意向量 $\mathbf{x} \in V$ 均可由基 $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$ 线性表示, 不妨设为

$$\mathbf{x} = \sum_{j=1}^n \alpha_j \boldsymbol{\varepsilon}_j$$

式中: $\alpha = [\alpha_1, \dots, \alpha_n]^T$ 是向量 x 在基 $\varepsilon_1, \dots, \varepsilon_n$ 下的坐标.

注意到

$$T(x) = [T(\varepsilon_1), \dots, T(\varepsilon_n)] \alpha = [\varepsilon_1, \dots, \varepsilon_n] A \alpha$$

故 $T(x) = \theta$ 当且仅当 $A\alpha = 0$. 基于此, 定义映射 $f_1: N(T) \rightarrow N(A)$ 满足

$$f_1(x) = \alpha, \quad \forall x \in V$$

不难证明, f_1 是 $N(T)$ 到 $N(A)$ 的同构映射, 因此 $\dim N(T) = \dim N(A)$.

同理, 定义映射 $f_2: R(T) \rightarrow R(A)$ 满足

$$f_2(y) = A\alpha, \quad \forall y \in W$$

是 $R(T)$ 到 $R(A)$ 的同构映射, 即 $\dim R(T) = \dim R(A)$.

再根据性质(1)(2)以及定理 2.2.5 即得性质(3).

例 2.3.6 考察齐次线性差分方程

$$u_{k+n} + a_{n-1}u_{k+n-1} + \dots + a_1u_{k+1} + a_0u_k = 0 \quad (2.3.3)$$

$k=0, \pm 1, \pm 2, \dots$. 方程(2.3.3)的解集 \tilde{S} 是 S 的一个线性子空间. 若定义

$$x_k = \begin{bmatrix} u_k \\ u_{k+1} \\ \vdots \\ u_{k+n-1} \end{bmatrix} \in \mathbb{R}^n, \quad A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

则式(2.3.3)可改写为

$$x_{k+1} = Ax_k, \quad k=0, \pm 1, \pm 2, \dots \quad (2.3.4)$$

此时, 矩阵 A 恰好是多项式 $g(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0$ 的友矩阵.

由式(2.3.4)易验证, 当 x_0 给定, 序列 $\{u_k\}$ 唯一确定. 因此, 定义映射 $T: \tilde{S} \rightarrow \mathbb{R}^n$ 满足

$$T(\{u_k\}) = x_0$$

易证 T 是同构映射. 因此, $\dim(\tilde{S}) = \dim(\mathbb{R}^n) = n$, 即 \tilde{S} 是 S 的一个 n 维线性子空间.

2.4 特征值与特征向量

考察线性变换 $T: x \rightarrow Ax$, 其中 $A = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}$. 若原像 x 分别取为 $u = [-1, 1]^T$

和 $v = [2, 1]^T$, 则其像分别为

$$Au = \begin{bmatrix} -5 \\ -1 \end{bmatrix}, \quad Av = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

如图 2.4.1 所示,像 Au 与它的原像 u 存在一定的夹角,而像 Av 与它的原像 v 是线性相关的,即向量 v 在线性变换前后仅其长度发生了伸缩,方向保持不变.这表明存在着一些向量在线性变换前后方位保持不变(或变换前后向量始终在一条直线上),这就是我们常提到的特征向量.

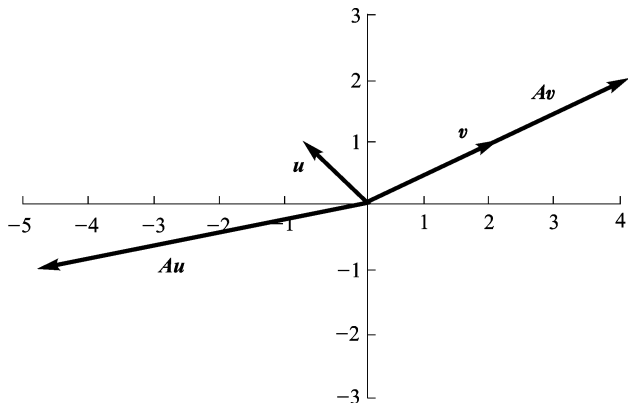


图 2.4.1 像与原像示意图

定义 2.4.1(线性变换的特征值和特征向量) 设线性变换 $T \in \mathcal{L}(V)$, 若存在 $\lambda_0 \in F$ 及 V 的非零向量 ξ 使得 $T\xi = \lambda_0 \xi$, 则称 λ_0 是 T 的一个特征值, 称 ξ 为 T 的属于特征值 λ_0 的一个特征向量.

例 2.4.1 设 V 是数域 F 上的 n 维线性空间, 定义恒等变换 $T_1 x = x, \forall x \in V$ 和零变换 $T_2(x) = \theta, \forall x \in V$, 则 V 的任意非零向量 ξ 都是 T_1 的属于特征值 $\lambda_0 = 1$ 的特征向量和 T_2 的属于特征值 $\lambda_0 = 0$ 的特征向量.

注 1: 从几何角度看, 特征向量在线性变换作用下保持共线, 即在同一直线上(有可能反向).

注 2: 设 ξ 是 T 的属于特征值 λ_0 的一个特征向量, 则 $k\xi$ 也是 T 的属于特征值 λ_0 的特征向量, 其中 $k \in F$ 且 $k \neq 0$.

注 3: 若 $\xi \in N(T)$ 且 $\xi \neq \theta$, 则 ξ 是属于 0 的特征向量.

注 4: 设 $T \in \mathcal{L}(V)$, ξ_1, \dots, ξ_n 是 V 的一组基, 且 $T\xi_i = \lambda_i \xi_i (i = 1, \dots, n)$, 则 T 在基 ξ_1, \dots, ξ_n 下的矩阵为对角阵.

考察一般情况: 取定线性空间 V 的一组基 ξ_1, \dots, ξ_n , 并有

$$T(\xi_1, \dots, \xi_n) = [\xi_1, \dots, \xi_n] A$$

式中: $A \in F^{n \times n}$ 是 T 在基 ξ_1, \dots, ξ_n 下的矩阵.

再设 λ 是 T 的一个特征值, ξ 为 T 的属于特征值 λ 的一个特征向量, 则

$$T\xi = \lambda\xi$$

且 ξ 可由基 ξ_1, \dots, ξ_n 线性表示, 即 $\xi = [\xi_1, \dots, \xi_n] \alpha, \alpha \in F^n$. 由此,

$$T\xi = T[\xi_1, \dots, \xi_n]\alpha = [\xi_1, \dots, \xi_n]A\alpha$$

$$\lambda\xi = \lambda[\xi_1, \dots, \xi_n]\alpha$$

由 $T\xi = \lambda\xi$ 得

$$\lambda[\xi_1, \dots, \xi_n]\alpha = [\xi_1, \dots, \xi_n]A\alpha$$

由向量坐标的唯一性知, $\lambda\alpha = A\alpha$. 这说明, 特征向量 ξ 在基 ξ_1, \dots, ξ_n 下的坐标满足齐次线性方程组

$$(\lambda I - A)\alpha = 0 \quad (2.4.1)$$

因为 $\xi \neq 0$, 所以 $\alpha \neq 0$, 即齐次线性方程组 (2.4.1) 有非零解. 方程组 (2.4.1) 有非零解的充分必要条件是它的系数矩阵行列式为零, 即

$$|\lambda I - A| = 0 \quad (2.4.2)$$

定义 2.4.2 (矩阵的特征值和特征向量) 设 $A \in F^{n \times n}$, λ 为一文字, 矩阵 $\lambda I - A$ 称为 A 的特征矩阵, 其行列式 $|\lambda I - A|$ 称为 A 的特征多项式, 方程 $|\lambda I - A| = 0$ 的根称为 A 的特征值 (或特征根). 方程 $(\lambda I - A)\alpha = 0$ 的非零解向量 α 称为属于特征值 λ 的特征向量.

注 5: λ 是线性变换 T 的特征值当且仅当 λ 是 A 的特征值; 向量 ξ 是线性变换 T 的特征向量当且仅当 α 是 A 的特征向量, 其中 $\xi = [\xi_1, \dots, \xi_n]\alpha$, α 是 ξ 在线性空间 V 的基 ξ_1, \dots, ξ_n 下的坐标.

例 2.4.2 设 $T \in \mathcal{L}(\mathbb{R}^3)$, 它在基 ξ_1, ξ_2, ξ_3 下的矩阵 A 为

$$\begin{bmatrix} 2 & -2 & 2 \\ -2 & -1 & 4 \\ 2 & 4 & -1 \end{bmatrix}$$

求 T 的全部特征值和特征向量.

解: 经计算得, A 的特征多项式为

$$f_A(\lambda) = |\lambda I - A| = (\lambda - 3)^2(\lambda + 6)$$

解得 $\lambda_1 = \lambda_2 = 3, \lambda_3 = -6$.

对于特征值 $\lambda_1 = \lambda_2 = 3$, 齐次方程组 $(3I - A)\alpha = 0$ 的基础解系为

$$\alpha_1 = [-2, 1, 0]^T, \quad \alpha_2 = [2, 0, 1]^T$$

因此, 线性变换 T 属于特征值 $\lambda_1 = \lambda_2 = 3$ 的有两个线性无关的特征向量, 分别为 $\xi_1 = -2\xi_1 + \xi_2$ 和 $\xi_2 = 2\xi_1 + \xi_3$.

同理, 有 $\xi_3 = \xi_1 + 2\xi_2 - 2\xi_3$. 因此, T 的全部特征向量进而表示为

$$\xi = k_1\xi_1 + k_2\xi_2 + k_3\xi_3$$

式中: $k_1, k_2, k_3 \in \mathbb{R}$ 且 k_1, k_2, k_3 不同时为零.

【思考】 矩阵 $A \in F^{n \times n}$ 有 n 个特征值吗?

例 2.4.3 考察矩阵 $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, 其特征多项式为 $f_A(\lambda) = \begin{vmatrix} \lambda & 1 \\ -1 & \lambda \end{vmatrix} = \lambda^2 + 1$.

若 $F = \mathbb{C}$, 则矩阵 A 有两个特征值, 分别为 $\pm i$; 若 $F = \mathbb{R}$, 则方程 $\lambda^2 + 1 = 0$ 无实根, 即

矩阵 A 在数域 \mathbb{R} 上无特征值. 因此, 矩阵的特征值依赖于 V 所在的数域 F .

定理 2.4.1 设 $\lambda_1, \dots, \lambda_n$ 是矩阵 $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ 的特征值, 则

$$\prod_{i=1}^n \lambda_i = |A|$$

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

式中: $\text{tr}(A)$ 称为矩阵 A 的迹.

例 2.4.4 设 λ 是可逆复方阵 A 的特征值, 试证明

- (1) λ^{-1} 是 A^{-1} 的特征值;
- (2) $\lambda^{-1} |A|$ 是 A^* 的特征值.

证明: (1) 当 A 可逆时, $\lambda \neq 0$. 令 x 是属于特征值 λ 的特征向量, 则有

$$Ax = \lambda x$$

对上式左右两端乘以 A^{-1} , 并整理得

$$A^{-1}x = \lambda^{-1}x$$

故 λ^{-1} 是 A^{-1} 的特征值.

(2) 根据 $AA^* = |A|I$ 知, $A^* = |A|A^{-1}$, 故 $\lambda^{-1}|A|$ 是 A^* 的特征值.

定义 2.4.3 (特征子空间) 设 λ 是矩阵 $A \in \mathbb{C}^{n \times n}$ 的一个特征值, 定义集合

$$E(\lambda) = \{x \in \mathbb{C}^n \mid Ax = \lambda x\}$$

则 $E(\lambda)$ 是 \mathbb{C}^n 的线性子空间, 称为属于特征值 λ 的特征子空间, $\dim(E(\lambda))$ 为特征值 λ 的几何重数.

注 6: 特征子空间 $E(\lambda)$ 也是齐次线性方程组 $(\lambda I - A)x = 0$ 的解空间, 还是矩阵 $(\lambda I - A)$ 的零空间.

注 7: 由一元 n 次多项式方程在复数域内有且仅有 n 个根知, n 阶矩阵 A 在复数域内必有 n 个特征值, 记为 $\lambda_1, \dots, \lambda_n$, 其中 λ_i 作为特征方程根的重数, 称为 λ_i 的代数重数.

例 2.4.5 求如下矩阵特征值的代数重数和几何重数:

$$J_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad J_2 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad J_3 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

解: 根据代数重数定义知, J_1, J_2 和 J_3 的特征值 1 的代数重数均为 3.

又知 $\text{rank}(I - J_1) = 0, \text{rank}(I - J_2) = 1, \text{rank}(I - J_3) = 2$, 故由亏加秩定理知, J_1, J_2 和 J_3 的特征值 1 的几何重数分别为 3, 2 和 1.

定理 2.4.2 复方阵的任一特征值的几何重数不超过它的代数重数.

证明: 设 λ_0 为 n 阶矩阵 $A \in \mathbb{C}^{n \times n}$ 的一个特征值, 其代数重数和几何重数分别为 m 和 k . 由此, 设 p_1, \dots, p_k 是特征子空间 $E(\lambda_0)$ 的一组基. 由基扩充定理知, 可将它扩充为 \mathbb{C}^n 的一组基, 记为 $p_1, \dots, p_k, p_{k+1}, \dots, p_n$.

定义 $P = [p_1, \dots, p_k, p_{k+1}, \dots, p_n] \in \mathbb{C}^{n \times n}$, 则 P 必可逆且有

$$P^{-1}AP = \begin{bmatrix} A_0 & B_1 \\ O & B_2 \end{bmatrix} \stackrel{\text{def}}{=} B$$

式中: A_0 为以 λ_0 为对角元素的 k 阶矩阵, B_2 为 $(n-k)$ 阶矩阵.

由于 A 与 B 相似, 故它们有相同的特征值. 注意到 B 的特征多项式为

$$|\lambda I_n - B| = (\lambda - \lambda_0)^k |\lambda I_{n-k} - B_2|$$

即 λ_0 在 B 中的代数重数至少为 k , 故它在 A 中的代数重数也至少为 k , 于是有 $m \geq k$. 证毕.

对于相似矩阵有如下结论.

命题 2.4.1 若 n 阶方阵 A 与 B 相似, 则

- (1) A 与 B 有相同的特征多项式与特征值;
- (2) A 与 B 有相同的秩与行列式;
- (3) A 与 B 有相同的迹.

注 8: 线性变换的矩阵的特征多项式与基的选取无关, 它直接由线性变换决定, 故可称之为线性变换的特征多项式.

关于矩阵的特征向量有如下结论.

定理 2.4.3 矩阵 A 的属于不同特征值的特征向量线性无关.

证明: 设 $\lambda_1, \dots, \lambda_r$ 是 n 阶矩阵 A 的 r ($2 \leq r \leq n$) 个互不相同的特征值, $\alpha_1, \dots, \alpha_r$ 是分别属于特征值 $\lambda_1, \dots, \lambda_r$ 的特征向量.

考察向量方程

$$k_1 \alpha_1 + \dots + k_r \alpha_r = 0 \quad (2.4.3)$$

式中: $k_1, \dots, k_r \in F$ 为待定系数.

对式 (2.4.3) 两端左乘矩阵 A, A^2, \dots, A^{n-1} , 得如下方程组:

$$\begin{aligned} k_1 \lambda_1 \alpha_1 + \dots + k_r \lambda_r \alpha_r &= 0 \\ k_1 \lambda_1^2 \alpha_1 + \dots + k_r \lambda_r^2 \alpha_r &= 0 \\ &\vdots \\ k_1 \lambda_1^{n-1} \alpha_1 + \dots + k_r \lambda_r^{n-1} \alpha_r &= 0 \end{aligned}$$

联合式 (2.4.3), 有

$$[k_1 \alpha_1, \dots, k_r \alpha_r] D^T = 0 \quad (2.4.4)$$

其中

$$D = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_r \\ \vdots & \vdots & & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \cdots & \lambda_r^{n-1} \end{bmatrix}$$

由于 $|D| = \prod_{1 \leq j < i \leq n} (\lambda_i - \lambda_j)$ 知, D 为可逆矩阵. 对式 (2.4.4) 两端右乘矩阵

$(D^T)^{-1}$ 得

$$[k_1 \alpha_1, \dots, k_r \alpha_r] = \mathbf{0}$$

即 $k_i \alpha_i = \mathbf{0} (i=1, \dots, n)$. 由于 α_i 为非零向量, 故 $k_i = 0 (i=1, \dots, n)$. 因此, $\alpha_1, \dots, \alpha_r$ 线性无关. 证毕.

注 9: 形如 D 的矩阵称为 Vardermonde 矩阵, 它的行列式称为 Vardermonde 行列式.

【应用】 在力学和工程技术中的很多问题都可归结为求解矩阵的最大特征值和相应的特征向量. 为讨论方便, 假设 $A \in \mathbb{R}^{n \times n}$ 的 n 个特征值满足 $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n| \geq 0$, 且属于它们的特征向量 x_1, \dots, x_n 线性无关. 则对任意向量 $x \in \mathbb{R}^n$, 有

$$x = a_1 x_1 + a_2 x_2 + \dots + a_n x_n$$

于是

$$Ax = a_1 \lambda_1 x_1 + a_2 \lambda_2 x_2 + \dots + a_n \lambda_n x_n$$

同理, 对任意 $k \in \mathbb{N}$ 有

$$A^k x = a_1 \lambda_1^k x_1 + a_2 \lambda_2^k x_2 + \dots + a_n \lambda_n^k x_n \quad (2.4.5)$$

对式 (2.4.5) 左右两端除以 λ_1^k 得,

$$\frac{1}{\lambda_1^k} A^k x = a_1 x_1 + a_2 \left(\frac{\lambda_2}{\lambda_1} \right)^k x_2 + \dots + a_n \left(\frac{\lambda_n}{\lambda_1} \right)^k x_n$$

当 k 趋于无穷大, 且当 $a_1 \neq 0$ 时,

$$\lim_{k \rightarrow \infty} \frac{1}{\lambda_1^k} A^k x = a_1 x_1$$

上式表明对于足够大的 k 值, 向量 $A^k x$ 与特征向量 x_1 在几何上几乎共线. 因此, 可以通过计算 $A^k x$ 估计特征向量 x_1 . 基于此, 我们可设计如下算法:

- (1) 选择初始向量 $x^{(0)}$ 且其最大分量为 1.
- (2) 对 $k=0, 1, \dots$,
 - i. 计算 $Ax^{(k)}$;
 - ii. 计算 $x^{(k+1)} = Ax^{(k)} / \mu_k$, 其中 μ_k 是 $Ax^{(k)}$ 模值最大的分量.
- (3) 序列 $\{\mu_k\}$ 收敛于特征值 λ_1 , 序列 $\{x^{(k)}\}$ 收敛于特征向量 x_1 .

上述方法就是近似计算实矩阵绝对值最大的特征值及其特征向量的方法, 我们常称之为幂法.

例 2.4.6 设 $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $x^{(0)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, 试利用幂法求 A 的绝对值最大的特征值

及其特征向量的近似值 (k 取到 5).

解: 由幂法算法知

$$\mathbf{A}\mathbf{x}^{(0)} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad \mu_0 = 3$$

$$\mathbf{x}^{(1)} = \begin{bmatrix} 0.333\ 3 \\ 1 \end{bmatrix}$$

$$\mathbf{A}\mathbf{x}^{(1)} = \begin{bmatrix} 2.333\ 3 \\ 5 \end{bmatrix}, \quad \mu_1 = 5$$

$$\mathbf{x}^{(1)} = \begin{bmatrix} 0.466\ 7 \\ 1 \end{bmatrix}$$

$$\mathbf{A}\mathbf{x}^{(2)} = \begin{bmatrix} 2.466\ 7 \\ 5.4 \end{bmatrix}, \quad \mu_2 = 5.4$$

$$\mathbf{x}^{(3)} = \begin{bmatrix} 0.456\ 8 \\ 1 \end{bmatrix}$$

$$\mathbf{A}\mathbf{x}^{(2)} = \begin{bmatrix} 2.456\ 8 \\ 5.370\ 4 \end{bmatrix}, \quad \mu_3 = 5.370\ 4$$

$$\mathbf{x}^{(4)} = \begin{bmatrix} 0.457\ 6 \\ 1 \end{bmatrix}$$

$$\mathbf{A}\mathbf{x}^{(4)} = \begin{bmatrix} 2.457\ 5 \\ 5.372\ 4 \end{bmatrix}, \quad \mu_4 = 5.372\ 4$$

$$\mathbf{x}^{(5)} = \begin{bmatrix} 0.457\ 4 \\ 1 \end{bmatrix}$$

$$\mathbf{A}\mathbf{x}^{(5)} = \begin{bmatrix} 2.457\ 4 \\ 5.372\ 3 \end{bmatrix}, \quad \mu_5 = 5.372\ 3$$

$$\mathbf{x}^{(6)} = \begin{bmatrix} 0.457\ 4 \\ 1 \end{bmatrix}$$

因此, $\mu_5 = 5.372\ 3$ 是矩阵 \mathbf{A} 的绝对值最大的特征值估计, 其特征向量估计为 $\mathbf{x}^{(6)}$. 这一结果与 MATLAB 计算结果一致.

幂法的优点是方法简单, 但其收敛速度与矩阵模值最大与次大特征值的比值有关, 因此对于某些矩阵其收敛速度较慢.

【思考】 若初试向量 $\mathbf{x}^{(0)}$ 与 \mathbf{x}_1 正交对结果会有影响吗?

【思考】 在知道特征值的一个较好的估计值后, 如何进一步提高精度?

我们常用反幂法进行求解. 设矩阵 \mathbf{A} 的特征值 λ_i 的估计值为 α_i , 则令

$$\mathbf{B} = (\mathbf{A} - \alpha_i \mathbf{I})^{-1}$$

根据例 2.4.4 知, \mathbf{B} 的特征值为

$$\frac{1}{\lambda_1 - \alpha_i}, \dots, \frac{1}{\lambda_n - \alpha_i}$$

且 \mathbf{A} 的属于特征值 $\lambda_1, \dots, \lambda_n$ 的特征向量恰是 \mathbf{B} 的属于上式这些特征值的特征向量

(本章的习题 15). 假设 α_i 更接近特征值 λ_i , 而不接近矩阵 \mathbf{A} 的其他特征值. 那么 $(\lambda_i - \alpha_i)^{-1}$ 就成为矩阵 \mathbf{B} 的模值最大的特征值. 此时利用幂法求解会快速地逼近 λ_i . 这就是反幂法的基本原理, 具体算法如下:

- (1) 选择特征值 λ 的估计值 α 和初始向量 $\mathbf{x}^{(0)}$ 且其最大分量为 1.
- (2) 对 $k=0, 1, \dots$,
 - i. 从 $(\mathbf{A} - \alpha \mathbf{I})\mathbf{y}^{(k)} = \mathbf{x}^{(k)}$ 解出 $\mathbf{y}^{(k)}$;
 - ii. 计算 $\mathbf{x}^{(k+1)} = \mathbf{y}^{(k)} / \mu_k$, 其中 $v_k = \alpha + \frac{1}{\mu_k}$, μ_k 是 $\mathbf{y}^{(k)}$ 模值最大的分量.
- (3) 序列 $\{v_k\}$ 收敛于特征值 λ , 序列 $\{\mathbf{x}^{(k)}\}$ 收敛于特征向量 \mathbf{x} .

对于多数简单情况, 幂法和反幂法是实用的, 另一个更全面、更广泛使用的迭代算法是 QR 算法, 详见 3.4 节内容.

【思考】 请阅读参考文献[19]“Eigenvectors from eigenvalues”, 试比较文献中所提及的计算矩阵特征向量的方法与定义法、幂法的优劣.

2.5 酉变换与酉矩阵

定义 2.5.1 (正交变换和酉变换) 若欧氏(酉)空间中的线性变换 T 保持向量的内积不变, 即对 V 的任意向量 \mathbf{x} 与 \mathbf{y} 有

$$(T(\mathbf{x}), T(\mathbf{y})) = (\mathbf{x}, \mathbf{y})$$

则称 T 为正交(酉)变换.

定义 2.5.2 (正交矩阵和酉矩阵) 若 n 阶实方阵 \mathbf{A} 满足 $\mathbf{A}^T \mathbf{A} = \mathbf{I}$ 或 $\mathbf{A} \mathbf{A}^T = \mathbf{I}$, 则称 \mathbf{A} 为正交矩阵; 若 n 阶复方阵 \mathbf{A} 满足 $\mathbf{A}^H \mathbf{A} = \mathbf{I}$ 或 $\mathbf{A} \mathbf{A}^H = \mathbf{I}$, 则称 \mathbf{A} 为酉矩阵.

定理 2.5.1 设 V 是 n 维欧氏(酉)空间, $T \in L(V)$, 则以下命题等价:

- (1) T 是正交(酉)变换;
- (2) T 保持长度不变, 即 $\|T(\mathbf{x})\| = \|\mathbf{x}\|$;
- (3) 若 ξ_1, \dots, ξ_n 是 V 中一组标准正交基, 则 $T(\xi_1), \dots, T(\xi_n)$ 也是 V 中一组标准正交基;

(4) T 在 V 的任一标准正交基下的矩阵 \mathbf{A} 为正交(酉)矩阵.

证明: (1) \Rightarrow (2). 由于 T 是正交(酉)变换, 故 $(T(\mathbf{x}), T(\mathbf{x})) = (\mathbf{x}, \mathbf{x})$, 两边开方得 $\|T\mathbf{x}\| = \|\mathbf{x}\|$, 即正交(酉)变换 T 保持长度不变;

(2) \Rightarrow (3). 设 ξ_1, \dots, ξ_n 是 V 中一组标准正交基, 则对 $i=1, \dots, n$, 有 $\|T(\xi_i)\| = \|\xi_i\| = 1$. 当 $i \neq j$ 时, 有

$$\begin{aligned} \|\xi_i + \xi_j\|^2 &= \|\xi_i\|^2 + \|\xi_j\|^2 \\ \|T(\xi_i + \xi_j)\|^2 &= (T(\xi_i + \xi_j), T(\xi_i + \xi_j)) \end{aligned}$$

$$= \|T(\xi_i)\|^2 + (T(\xi_i), T(\xi_j)) + (T(\xi_j), T(\xi_i)) + \|T(\xi_j)\|^2$$

又知 $\|T(\xi_i + \xi_j)\| = \|\xi_i + \xi_j\|$ 和 $\|T(\xi_i)\| = \|\xi_i\|$, 故

$$(T(\xi_i), T(\xi_j)) + (T(\xi_j), T(\xi_i)) = 0 \quad (2.5.1)$$

用 $i\xi_j$ 替换式(2.5.1)中的 ξ_j , 得

$$-(T(\xi_i), T(\xi_j)) + (T(\xi_j), T(\xi_i)) = 0 \quad (2.5.2)$$

联立式(2.5.1)和式(2.5.2), 得

$$(T(\xi_i), T(\xi_j)) = 0$$

所以, 向量组 $T(\xi_1), \dots, T(\xi_n)$ 是 V 的一组标准正交基.

(3) \Rightarrow (4). 设矩阵 $A = (a_{ij})_{n \times n}$ 是由标准正交基 ξ_1, \dots, ξ_n 到标准正交基 $T(\xi_1), \dots, T(\xi_n)$ 的过渡矩阵, 则对 $i=1, \dots, n$, 有

$$T(\xi_i) = [\xi_1, \dots, \xi_n] \begin{bmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{bmatrix}$$

于是,

$$(T(\xi_i), T(\xi_j)) = \sum_{k=1}^n \bar{a}_{ik} a_{kj} = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases} \quad (2.5.3)$$

所以, $A^H A = I$, 即 A 为正交(酉)矩阵.

(4) \Rightarrow (1). 设 $T(\xi_1, \dots, \xi_n) = [\xi_1, \dots, \xi_n] A$, 其中 A 为酉(正交)矩阵. 由式(2.5.3)知, $T(\xi_1), \dots, T(\xi_n)$ 是 V 的一组标准正交基.

对 V 中任意向量 x, y , 定义 $x = \sum_{i=1}^n a_i \xi_i, y = \sum_{i=1}^n b_i \xi_i$, 有

$$\begin{aligned} (T(x), T(y)) &= \left(\sum_{i=1}^n a_i \xi_i, \sum_{i=1}^n b_i \xi_i \right) \\ &= \sum_{i,j=1}^n a_i \bar{b}_j (T(\xi_i), T(\xi_j)) = \sum_{i=1}^n a_i \bar{b}_i = (x, y) \end{aligned}$$

即 T 是正交(酉)变换. 证毕.

【思考】 正交矩阵 A 的特征值一定是 ± 1 吗?

分析: 设 λ 是 A 的任一特征值, x 是属于 λ 的特征向量, 则有

$$Ax = \lambda x \quad (2.5.4)$$

对式(2.5.4)两端取转置, 得

$$x^T A^T = \lambda x^T \quad (2.5.5)$$

将式(2.5.4)和式(2.5.5)相乘得

$$x^T A^T A x = \lambda^2 x^T x \quad (2.5.6)$$

由 $A^T A = I$ 得,

$$(\lambda^2 - 1) x^T x = 0 \quad (2.5.7)$$

注意到 $\mathbf{x}^T \mathbf{x} \neq 0$, 故 $\lambda^2 = 1$, 即 $\lambda = \pm 1$.

上述分析是错误的. 这是在矩阵分析中极易犯的一个错误. 错误原因在于 $\mathbf{x}^T \mathbf{x} \neq 0$. 尽管 $\mathbf{x}^T \mathbf{x}$ 看起来是向量内积的形式, 但这是在欧氏空间定义的, 在西空间没有这种定义. 例如, 令 $\mathbf{x} = [1, i]^T$, 则 $\mathbf{x}^T \mathbf{x} = 0$. 在上述分析中, 若特征值 λ 不是实数, 则特征向量 \mathbf{x} 一定不是实向量. 在上例中, 我们利用 $\mathbf{x}^T \mathbf{x} \neq 0$ 这一结论实际上就已经认定了 \mathbf{x} 是实向量, 自然也就得出了 λ 是实特征值这一结论. 因此, 上面的证明过程应该改为:

对式(2.5.4)两端取共轭转置, 得

$$\mathbf{x}^H \mathbf{A}^H = \bar{\lambda} \mathbf{x}^H \quad (2.5.8)$$

将式(2.5.4)与式(2.5.8)相乘得

$$\mathbf{x}^H \mathbf{A}^H \mathbf{A} \mathbf{x} = |\lambda|^2 \mathbf{x}^H \mathbf{x}$$

由 $\mathbf{A}^T \mathbf{A} = \mathbf{I}$ 得,

$$(|\lambda|^2 - 1) \mathbf{x}^H \mathbf{x} = 0$$

注意到 $\mathbf{x}^H \mathbf{x} \neq 0$, 故 $|\lambda|^2 = 1$, 即 $|\lambda| = 1$.

例如, $\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ 是正交矩阵, 但其特征值为 $-\frac{1}{2} \pm i \frac{\sqrt{3}}{2}$ 和 1.

命题 2.5.1 正交(酉)矩阵 \mathbf{A} 满足如下性质:

- (1) 正交矩阵的行列式必为 ± 1 , 酉矩阵的行列式的模值为 1;
- (2) $\mathbf{A}^{-1} = \mathbf{A}^H$ 均为正交(酉)矩阵;
- (3) 正交(酉)矩阵的乘积仍为正交(酉)矩阵;
- (4) \mathbf{A} 的所有特征值的模值为 1.

定理 2.5.2 矩阵 \mathbf{A} 是 n 阶正交(酉)矩阵当且仅当矩阵 \mathbf{A} 的 n 个列(行)向量构成 n 维欧氏(酉)空间的一组标准正交基.

证明: 必要性. 设酉矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 对 \mathbf{A} 进行列分块, 并记为 $\mathbf{A} = [\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n]$, 其中 $\boldsymbol{\alpha}_i \in \mathbb{C}^n$. 由等式 $\mathbf{A}^H \mathbf{A} = \mathbf{I}$ 得

$$\begin{bmatrix} \boldsymbol{\alpha}_1^H \\ \vdots \\ \boldsymbol{\alpha}_n^H \end{bmatrix} [\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n] = \mathbf{I}$$

进而

$$\boldsymbol{\alpha}_i^H \boldsymbol{\alpha}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (2.5.9)$$

上式表明 $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n$ 是酉空间 \mathbb{C}^n 的一组单位正交向量组, 它们也构成了 n 维酉空间的一组标准正交基.

充分性. 设 $\mathbf{A} = [\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n]$, 其中 $\boldsymbol{\alpha}_i \in \mathbb{C}^n$. 若矩阵 \mathbf{A} 的这 n 个列向量构成 \mathbb{C}^n 的

一组标准正交基, 则式(2.5.9)成立, 进而 $\mathbf{A}^H \mathbf{A} = \mathbf{I}$. 证毕.

例 2.5.1 平面旋转变换 $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $\forall \mathbf{x} = [x_1, x_2]^T \in \mathbb{R}^2$, 满足 $T(\mathbf{x}) = \begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix} \mathbf{x}$, φ 为旋转角(逆时针取正). T 在标准基 $\mathbf{e}_1, \mathbf{e}_2$ 下的矩阵为

$$\mathbf{A} = \begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix}$$

由于 \mathbf{A} 是正交矩阵, 故 T 是正交变换. 在一般的 n 维欧氏空间中, 定义

$$\mathbf{T}(i, j) = (t_{kl}(i, j))_{n \times n} = \begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & \cos \varphi & 0 & \cdots & 0 & \sin \varphi & & \\ & & 0 & 1 & & & 0 & & \\ & & \vdots & & \ddots & & \vdots & & \\ & & 0 & & & 1 & 0 & & \\ & & -\sin \varphi & 0 & \cdots & 0 & \cos \varphi & & \\ & & & & & & & 1 & \\ & & & & & & & & \ddots & \\ & & & & & & & & & 1 \end{bmatrix}$$

式中: $t_{ii}(i, j) = t_{jj}(i, j) = \cos \varphi$, $t_{ij}(i, j) = \sin \varphi$, $t_{ji}(i, j) = -\sin \varphi$, 对于 $k \neq i, j$, $t_{kk}(i, j) = 1$, 且对于任意 $k \neq i, j$ 和 $l \neq i, j$, $t_{kl}(i, j) = 0$. 我们将矩阵 $\mathbf{T}(i, j)$ 称为 Givens 矩阵(或初等旋转矩阵).

命题 2.5.2 设 Givens 矩阵 $\mathbf{T}(i, j) \in \mathbb{R}^{n \times n}$, 则以下命题成立:

- (1) $\mathbf{T}(i, j)$ 是正交矩阵且 $(\mathbf{T}(i, j))^{-1} = (\mathbf{T}(i, j))^T$;
- (2) 设 $\mathbf{x} = [x_1, \dots, x_n]^T$, 若 $\mathbf{y} = \mathbf{T}(i, j)\mathbf{x} = [y_1, \dots, y_n]^T$, 则

$$y_k = x_k, k \neq i$$

$$y_i = \cos \varphi x_i + \sin \varphi x_j$$

$$y_j = -\sin \varphi x_i + \cos \varphi x_j$$

注 1: 命题 2.5.2 性质(2)表明若 $x_i^2 + x_j^2 \neq 0$, 定义

$$\cos \varphi = \frac{x_i}{\sqrt{x_i^2 + x_j^2}}, \quad \sin \varphi = \frac{x_j}{\sqrt{x_i^2 + x_j^2}}$$

则 $y_i = \sqrt{x_i^2 + x_j^2}$, $y_j = 0$. 此时, 若定义 $\mathbf{y} = \mathbf{T}(1, j)\mathbf{x}$, 则向量 \mathbf{y} 的第 1 个分量为 $\sqrt{x_1^2 + x_j^2}$, 第 j 个分量为 0. 进一步, 必存在着有限个 Givens 矩阵的乘积, 记为 \mathbf{T} 使得 $\mathbf{T}\mathbf{x} = |\mathbf{x}| \mathbf{e}_1$.

例 2.5.2 设 $\mathbf{x} = [0, 1, 1]^T$, 取 Givens 矩阵

$$T(1,2) = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad T(1,2)x = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

再取 Givens 矩阵

$$T(1,3) = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

定义 $T = T(1,3)T(1,2)$ 得

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

则有 $Tx = |x|e_1$.

另一类重要的酉矩阵是 Householder 矩阵, 其定义和性质如下.

定义 2.5.3 (Householder 矩阵) 设 $w \in \mathbb{C}^n$ 是单位向量, 定义矩阵

$$H(w) = I - 2ww^H$$

称为 Householder 矩阵 (或初等反射矩阵).

例 2.5.3 设 $w \in \mathbb{C}^n$ 是给定单位向量, 定义映射 $T: \mathbb{C}^n \rightarrow \mathbb{C}^n$ 使得对任意 $x \in \mathbb{C}^n$, $T(x) = y$, 其中, y 是向量 x 关于空间 W^\perp 的对称向量, $W = \text{span}(w)$. 如图 2.5.1 所示:

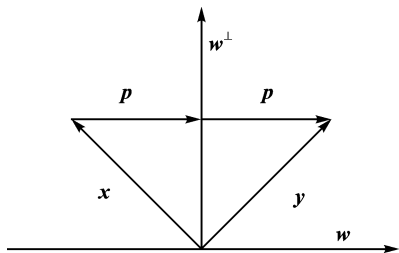


图 2.5.1 例 2.5.3 各向量关系示意图

$$x + 2p = y$$

$$x + p = \text{Proj}_{w^\perp} x = x - (x, w)w$$

由此, 解得

$$y = x - 2ww^H x = H(w)x \quad (2.5.10)$$

式中: $H(w)$ 是 Householder 矩阵.

根据式 (2.5.10) 知, T 是线性映射, 且在标准基下的矩阵为 $H(w)$. 因此, Householder 矩阵与 \mathbb{C}^n 中向量 x 相乘表示向量 x 沿向量 w 作为镜面的反射向量.

命题 2.5.3 Householder 矩阵 $H(w)$ 具有以下性质:

(1) $|H(w)| = -1$;

$$(2) (\mathbf{H}(\mathbf{w}))^H = \mathbf{H}(\mathbf{w}) = (\mathbf{H}(\mathbf{w}))^{-1};$$

(3) 设 $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ 且 $\mathbf{x} \neq \mathbf{y}$, 则存在单位向量 \mathbf{w} 使得 $\mathbf{H}(\mathbf{w})\mathbf{x} = \mathbf{y}$ 的充分必要条件是

$$\mathbf{x}^H \mathbf{x} = \mathbf{y}^H \mathbf{y}, \quad \mathbf{x}^H \mathbf{y} = \mathbf{y}^H \mathbf{x}$$

并且若上述条件成立, 则使 $\mathbf{H}(\mathbf{w})\mathbf{x} = \mathbf{y}$ 成立的单位向量 \mathbf{w} 可取为

$$\mathbf{w} = \frac{e^{i\theta}}{\|\mathbf{x} - \mathbf{y}\|} (\mathbf{x} - \mathbf{y})$$

其中 θ 为任一实数.

例 2.5.4 设 $\mathbf{x} = [0, 1, 1]^T$, 取 $\mathbf{y} = [\sqrt{2}, 0, 0]^T$, 并定义

$$\mathbf{e} = \frac{1}{\|\mathbf{x} - \mathbf{y}\|} (\mathbf{x} - \mathbf{y}) = \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

有 Householder 矩阵

$$\mathbf{H}(\mathbf{w}) = \mathbf{I} - 2\mathbf{w}\mathbf{w}^H = \begin{bmatrix} 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{\sqrt{2}}{2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

则 $\mathbf{H}(\mathbf{w})\mathbf{x} = \|\mathbf{x}\|\mathbf{e}_1$. 该结果与例 2.5.2 相同.

Givens 矩阵和 Householder 矩阵是矩阵计算的重要工具之一. 利用 Givens 变换和 Householder 变换是矩阵进行 QR 分解的重要技术手段.

【思考】 给定实方阵 \mathbf{A} , 是否有限个 Givens 矩阵或 Householder 矩阵的乘积, 记为 \mathbf{T} 使得 $\mathbf{T}\mathbf{A}$ 变成如下形式:

$$\begin{bmatrix} \lambda_1 & * & \cdots & * \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

其中, $*$ 为任意实数, $\lambda_1, \dots, \lambda_n$ 是矩阵 \mathbf{A} 的 n 个特征值.

2.6 应用: 图的矩阵表示

目前普遍认为, 图论是从 1736 年著名数学家欧拉研究哥尼斯堡七桥问题开始

的. 如图 2.6.1 所示, 哥尼斯堡被河流分割为四个地区(A、B、C 和 D), 它们通过河上的七座桥连接. 欧拉在哥尼斯堡发现, 当地市民正从事一项非常有趣的消遣活动: 在星期六做一次走过所有七座桥的散步, 每座桥只能经过一次而且起点与终点必须是同一地点. 这就是我们小时候玩的“一笔画游戏”. 欧拉把每个地区抽象为一个点, 称为图的“节点”, 将每座桥抽象为两个节点间的一条边. 基于这种简化, 欧拉解析证明了这种一次遍历七座桥的路径不存在. 欧拉对图的这种抽象表示至今仍在沿用.

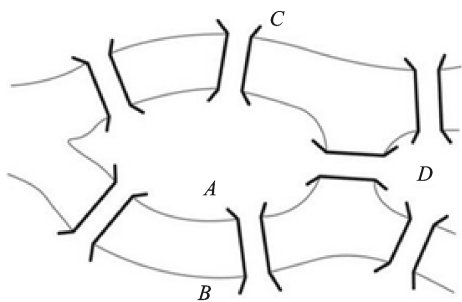


图 2.6.1 哥尼斯堡七桥问题

定义 2.6.1(图) 图 G 是有序三元组, 记作 $G = (V(G), E(G), \varphi_G)$, 其中, 非空集合 $V(G)$ 是 G 的节点集, 其元素称为节点(或结点、顶点), 集合 $E(G)$ 是 G 的边集, 其元素称为边, 而 $\varphi(G)$ 是集合 E 到集合 V 中元素有序对 $V \times V$ 的函数, 称为关联函数. 若 $V \times V$ 中元素全是无序对, 则图 G 称为无向图; 若 $V \times V$ 中元素全是有序对, 则图 G 称为有向图. 设边 $e \in E(G)$, 则存在 $x, y \in V(G)$ 和有序对

$(x, y) \in V \times V$ 使得 $\varphi_G(e) = (x, y)$, e 称为从 x 到 y 的有向边, x 称为边 e 的起点, y 称为边 e 的终点. 在无向图中, x 和 y 称为边 e 的端点. 去掉有向图 G 边上的方向得到的无向图称为 G 的基础图.

例 2.6.1 如图 2.6.2 所示, $G_1 = (V(G_1), E(G_1), \varphi_{G_1})$, 其中,

$$V(G_1) = \{v_1, v_2, v_3, v_4, v_5\}$$

$$E(G_1) = \{e_{12}, e_{23}, e_{34}, e_{45}, e_{15}\}$$

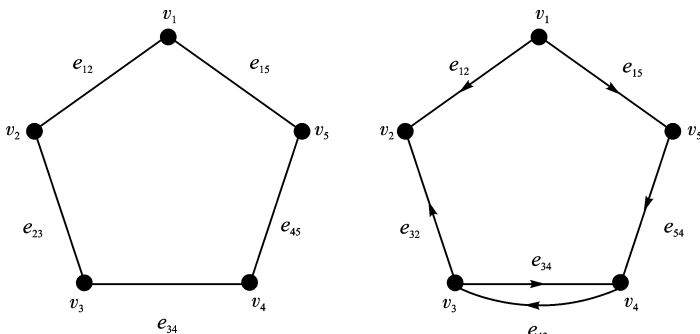


图 2.6.2 例 2.6.1 中的 G_1 (左) 和 G_2 (右)

而关联函数 φ_{G_1} 定义为

$$\varphi_{G_1}(e_{12}) = (v_1, v_2)$$

$$\varphi_{G_1}(e_{23}) = (v_2, v_3)$$

$$\varphi_{G_1}(e_{34}) = (v_3, v_4)$$

$$\varphi_{G_1}(e_{45}) = (v_4, v_5)$$

$$\varphi_{G_1}(e_{15}) = (v_1, v_5)$$

对于 $G_2 = (V(G_2), E(G_2), \varphi(G_2))$, 有

$$V(G_2) = \{v_1, v_2, v_3, v_4, v_5\}$$

$$E(G_2) = \{e_{12}, e_{32}, e_{34}, e_{43}, e_{54}, e_{15}\}$$

而关联函数 φ_{G_2} 定义为

$$\varphi_{G_2}(e_{12}) = (v_1, v_2)$$

$$\varphi_{G_2}(e_{32}) = (v_3, v_2)$$

$$\varphi_{G_2}(e_{34}) = (v_3, v_4)$$

$$\varphi_{G_2}(e_{43}) = (v_4, v_3)$$

$$\varphi_{G_2}(e_{54}) = (v_5, v_4)$$

$$\varphi_{G_2}(e_{15}) = (v_1, v_5)$$

定义 2.6.2(度) 设 $G = (V(G), E(G), \varphi_G)$ 是无向图, $v \in V(G)$ 的节点度定义为 G 中与 v 关联边的数目, 记为 $d_G(v)$.

定义 2.6.3(路) 设 u 和 v 是任意图 G 的节点, 图 G 的一条 $u-v$ 链是有限的节点和边交替序列 $u_0 e_1 u_1 e_2 \cdots u_{n-1} e_n u_n$ ($u = u_0, v = u_n$), 其中与边 e_i ($1 \leq i \leq n$) 相邻的两节点 u_{i-1} 和 u_i 正好是 e_i 的两个端点. 数 n (链中出现的边数) 称为链的长度. $u(u_0)$ 和 $v(u_n)$ 称为链的端点, 其余的节点称为链的内部点. 一条 $u-v$ 链, 当 $u \neq v$ 时, 称它为开的, 否则称为闭的. 边互不同的链称为迹, 内部点互不同的链称为路.

定义 2.6.4(连通图) 如果无向图 G 中每一对不同的节点 x 和 y 都有一条路, 则称 G 是连通图, 反之称为非连通图.

定义 2.6.5(连通分支) 设 $V_i, i=1, \cdots, m$ 是图 $G = (V(G), E(G), \varphi_G)$ 节点集 $V(G)$ 的子集, 满足: (1) $\bigcup_i V_i = V(G)$; (2) 对 $i \neq j$ 有, $V_i \cap V_j = \emptyset$. 若 $V_i (i=1, \cdots, m)$ 使得当且仅当两节点 v 和 u 属于同一子集 V_i 时, 节点 v 和 u 间存在一条路, 则 V_i 在 G 中导出的子图 G_i (以 V_i 为节点集, 以两两端点均在 V_i 中的全体边为边集合的 G 的子图) 称为 G 的连通分支, m 称为 G 的连通分支数.

由上述定义知, 图的数学抽象是三元组, 还知道图的形象直观的图形表示. 为便于计算, 我们介绍图的矩阵表示.

定义 2.6.6(邻接矩阵) 设图 $G = (V(G), E(G), \varphi_G)$, $V(G) = \{x_1, x_2, \cdots, x_n\}$. 若 a_{ij} 是图 G 中以 x_i 为起点且以 x_j 为终点的边的数目, 则 n 阶方阵 $A = (a_{ij})$ 称为 G 的邻接矩阵.

定义 2.6.7(关联矩阵) 设无向图 $G = (V(G), E(G), \varphi_G)$, $V(G) = \{x_1, x_2, \cdots, x_n\}$, $E = \{e_1, e_2, \cdots, e_m\}$. 若定义

$$m_{ij} = \begin{cases} 2, & e_j \text{ 关联于 } x_i, e_j \text{ 是自环} \\ 1, & e_j \text{ 关联于 } x_i, e_j \text{ 不是自环} \\ 1, & e_j \text{ 与 } x_i \text{ 不关联} \end{cases}$$

则称 $n \times m$ 阶矩阵 $M = (m_{ij})$ 为无向图 G 的关联矩阵.

定义 2.6.8(拉普拉斯矩阵) 设无向图 $G = (V(G), E(G), \varphi_G)$ 无自环, $V(G) = \{x_1, x_2, \dots, x_n\}$, 则称 n 阶方阵 $L = D - A$ 是图 G 的拉普拉斯矩阵, 其中, A 是图 G 的邻接矩阵, D 是 n 阶对角矩阵, 其对角线元素是对应节点的度.

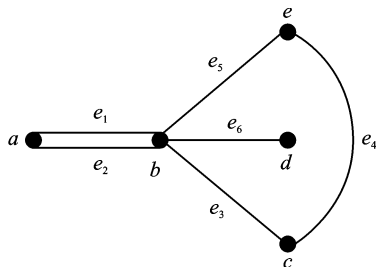


图 2.6.3 例 2.6.2 图 G

例 2.6.2 给出图 2.6.3 所示图 G 的关联矩阵、邻接矩阵和拉普拉斯矩阵.

解：图 G 的关联矩阵为

$$\begin{array}{c} e_1 \quad e_2 \quad e_3 \quad e_4 \quad e_5 \quad e_6 \\ \begin{matrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{matrix} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

图 G 的节点顺序为 a, b, c, d, e 的邻接矩阵和拉普拉斯矩阵是

$$A = \begin{bmatrix} 0 & 2 & 0 & 0 & 0 \\ 2 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}, \quad L = \begin{bmatrix} 2 & -2 & 0 & 0 & 0 \\ -2 & 5 & -1 & -1 & -1 \\ 0 & -1 & 2 & 0 & -1 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & -1 & 0 & 2 \end{bmatrix}$$

定理 2.6.1 设无向图 $G = (V(G), E(G), \varphi_G)$, $V(G) = \{x_1, x_2, \dots, x_n\}$, 且 $A = (a_{ij})_{n \times n}$ 为 G 的邻接矩阵, 则 A^k 中的 i 行 j 列元素 $a_{ij}^{(k)}$ 是图 G 中以 x_i 和 x_j 为端点且长度为 k 的链的数目.

定理 2.6.2 设图 G 的邻接矩阵为 $A \in \mathbb{R}^{n \times n}$, 作方阵 $R = A + A^2 + \dots + A^{n-1}$, 则图 G 连通的充分必要条件为 R 中的每个元素都不为零.

定理 2.6.3 设图 G 有 n 个节点和 k 个连通分支, 则 $\text{rank}(M) = n - k$, 其中 M 是 G 的关联矩阵.

定义 2.6.9(不可约矩阵) 如果存在 n 阶排列矩阵 P 使得

$$PAP^T = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

其中, $\mathbf{A}_{11} \in \mathbb{R}^{k \times k}$ ($1 \leq k \leq n-1$), 则称矩阵 \mathbf{A} 是可约矩阵; 否则称 \mathbf{A} 为不可约矩阵.

定理 2.6.4 无向图 G 是连通的当且仅当它的拉普拉斯矩阵 \mathbf{L} 不可约.

定理 2.6.5 无向图 G 的拉普拉斯矩阵 $\mathbf{L} = (l_{ij}) \in \mathbb{R}^{n \times n}$ 具有以下性质:

(1) \mathbf{L} 是半正定实对称矩阵.

(2) $\text{rank}(\mathbf{L}) = n - k$, 其中 k 为图的连通分支数; 特别地, 若 G 是连通图, 则 0 是矩阵 \mathbf{L} 的单根.

(3) 对任意向量 $\mathbf{x} = [x_1, \dots, x_n]^T \in \mathbb{R}^n$, 有

$$\mathbf{x}^T \mathbf{L} \mathbf{x} = -\frac{1}{2} \sum_{i,j=1}^n l_{ij} (x_i - x_j)^2$$

这里仅给出定理 2.6.5 性质(3)的证明. 将上式右端展开, 得

$$-\frac{1}{2} \sum_{i,j=1}^n l_{ij} (x_i - x_j)^2 = -\frac{1}{2} \sum_{i,j=1}^n l_{ij} (x_i^2 + x_j^2) + \sum_{i,j=1}^n l_{ij} x_i x_j$$

注意到

$$\begin{aligned} \sum_{i,j=1}^n l_{ij} x_i^2 &= \sum_{i=1}^n x_i^2 \left(\sum_{j=1}^n l_{ij} \right) = 0 \\ \sum_{i,j=1}^n l_{ij} x_j^2 &= \sum_{j=1}^n x_j^2 \left(\sum_{i=1}^n l_{ij} \right) = 0 \end{aligned}$$

故

$$-\frac{1}{2} \sum_{i,j=1}^n l_{ij} (x_i - x_j)^2 = \sum_{i,j=1}^n l_{ij} x_i x_j = \mathbf{x}^T \mathbf{L} \mathbf{x}$$

上述结论是拉普拉斯特征映射方法的核心理论基础, 它从图的角度去构建数据之间的关系, 是数据降维的常见方法. 有兴趣的读者可阅读参考文献[20].

本章习题

1. 设多项式 $f(\lambda)$ 的友矩阵为 \mathbf{A} , 证明 \mathbf{A} 的特征多项式为 $f(\lambda)$.

2. 拉普拉斯变换是求解和分析常系数线性微分方程的一种简便方法, 其定义如下: 设 $f(t)$ 为定义在 \mathbb{R}^+ 的实值函数, 且存在实数 τ 使得当 $s > \tau$ 时, 广义积分

$$F(s) = \int_0^{+\infty} f(t) e^{-st} dt$$

收敛, 则称上式右端的积分为函数 $f(t)$ 的拉普拉斯变换, $F(s)$ 称为 $f(t)$ 在拉普拉斯变换下的像函数, 常记为 $L(f(t))$, $f(t)$ 称为 $F(s)$ 的原函数. 试证明 $L: f(t) \rightarrow F(s)$ 是线性变换.

3. 设 V_1, V_2 和 V_3 均是数域 F 上的线性空间, 若 $T_1 \in \mathcal{L}(V_1, V_2)$, $T_2 \in \mathcal{L}(V_2, V_3)$, 证明 $T_2 T_1 \in \mathcal{L}(V_1, V_3)$.

4. 设 $T_1, T_2 \in \mathcal{L}(\mathbb{R}^{1 \times 2})$, 对任意 $[x_1, x_2] \in \mathbb{R}^{1 \times 2}$, 有 $T_1([x_1, x_2]) = [x_2, -x_1]$, $T_2([x_1, x_2]) = [x_1, -x_2]$, 试求 $T_1 + T_2, T_1 T_2$ 和 $T_2 T_1$.

5. 设 V 和 W 均是数域 F 上的线性空间, ζ_1, \dots, ζ_n 是 V 的一组基. 若 $T_1, T_2 \in \mathcal{L}(V, W)$, 且 $\forall k=1, \dots, n, T_1(\zeta_k) = T_2(\zeta_k)$. 证明 $T_1 = T_2$.

6. 复数域 \mathbb{C} 是 \mathbb{R} 上的线性空间, 证明 $\mathbb{C} \cong \mathbb{R}^2$.

7. 设线性映射 $T \in \mathcal{L}(V, W)$, 若 $\varepsilon_1, \dots, \varepsilon_n$ 和 $\varepsilon'_1, \dots, \varepsilon'_n$ 是 V 的两组基, 由 $\varepsilon_1, \dots, \varepsilon_n$ 到 $\varepsilon'_1, \dots, \varepsilon'_n$ 的过渡矩阵为 Q , 证明 $T(\varepsilon'_1, \dots, \varepsilon'_n) = T(\varepsilon_1, \dots, \varepsilon_n)Q$.

8. 设 $T_1 \in \mathcal{L}(V_1, V_2), T_2 \in \mathcal{L}(V_2, V_3), \varepsilon_1, \dots, \varepsilon_n, \eta_1, \dots, \eta_m$ 和 ζ_1, \dots, ζ_l 分别是 V_1, V_2 和 V_3 的基, T_1 在 V_1 的基 $\varepsilon_1, \dots, \varepsilon_n$ 和 V_2 的基 η_1, \dots, η_m 下的矩阵为 A , T_2 在 V_2 的基 η_1, \dots, η_m 和 V_3 的基 ζ_1, \dots, ζ_l 下的矩阵为 B , 求线性映射 $T_2 T_1$ 在 V_1 的基 $\varepsilon_1, \dots, \varepsilon_n$ 和 V_3 的基 ζ_1, \dots, ζ_l 下的矩阵. (理解矩阵相乘)

9. 设 V 是 n 维欧氏空间, $T \in \mathcal{L}(V)$, 若对任意向量 $x, y \in V$ 有

$$(T(x), y) = (x, T(y))$$

则称 T 是 V 上的对称变换. 证明 T 在 V 的一组标准正交基下的矩阵为对称矩阵.

10. 设 $T_1, T_2 \in \mathcal{L}(\mathbb{R}^n)$, 若对任意向量 $x, y \in \mathbb{R}^n$ 均有

$$(T_1(x), y) = (x, T_2(y))$$

则称 T_2 是 T_1 的一个伴随变换. 试证明线性变换 T_1 的伴随变换 T_2 存在且唯一.

11. 设 $T_1 \in \mathcal{L}(\mathbb{R}^2)$ 且满足

$$T_1: (x, y)^T \rightarrow (a_1x + b_1y, a_2x + b_2y)^T$$

求 T_1 和 T_1 的伴随变换 T_2 在标准基下的矩阵. (理解伴随变换可看作转置矩阵的几何意义.)

12. 证明 n 阶方阵 A 的属于特征值 λ 的全部特征向量再加上零向量构成 F^n 的一个线性子空间.

13. 设 $\lambda_1, \dots, \lambda_n$ 是 $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ 的特征值, 则

$$\prod_{i=1}^n \lambda_i = |A|$$

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

14. 设 $A \in \mathbb{C}^{m \times n}, B \in \mathbb{C}^{n \times m}$, 对任意 $\lambda \in \mathbb{C}$, 试证:

$$\lambda^n |\lambda I_m - AB| = \lambda^m |\lambda I_n - BA|$$

15. 设矩阵 $A \in \mathbb{C}^{n \times n}$ 有 s 个互异特征值, 分记为 $\lambda_1, \dots, \lambda_s$, 则以下两个集合

$$E(\lambda_1) + E(\lambda_2) + \dots + E(\lambda_s)$$

$$E(\lambda_1) \cap E(\lambda_2) \cap \dots \cap E(\lambda_s)$$

是 \mathbb{C}^n 的线性子空间吗? 若是, 求它们的维数.

16. 设 $T \in \mathcal{L}(V), W$ 是 V 的子空间. 若对任意 $\alpha \in W$, 都有 $T(\alpha) \in W$, 则称 W 是 T 的不变子空间. 证明 T 可对角化的充分必要条件是 V 可以分解为 T 的一维不变子空间的直和.

17. 设 λ 是 $A \in \mathbb{R}^{n \times n}$ 的特征值, $x \in \mathbb{R}^n$ 是属于 λ 的特征向量. 若 $\alpha \in \mathbb{R}$ 不是 A 的

特征值, 令 $B = (A - \alpha I)^{-1}$, 证明 $(\lambda - \alpha)^{-1}$ 是 B 的特征值, x 是其对应的特征向量.

18. 证明 Householder 矩阵 $H(w)$ 具有以下性质:

(1) $|H(w)| = -1$;

(2) $(H(xw))^H = H(w) = (H(w))^{-1}$;

(3) 设 $x, y \in \mathbb{C}^n$ 且 $x \neq y$, 则存在单位向量 w 使得 $H(w)x = y$ 的充分必要条件是

$$x^H x = y^H y, \quad x^H y = y^H x$$

并且若上述条件成立, 则使 $H(w)x = y$ 成立的单位向量 w 可取为

$$w = \frac{e^{i\theta}}{\|x - y\|} (x - y)$$

其中 θ 为任一实数.

19. 证明 Givens 矩阵可分解为两个 Householder 矩阵的乘积.

20. 近年来对众多实际网络的研究发现, 它们存在一个共同的特征, 称之为网络中的社团结构. 它是指网络中的节点可以分成组, 组内节点间的连接比较稠密, 组间节点的连接比较稀疏. 社团结构在实际系统中有着重要的意义. 在引文网中, 不同社团可能代表了不同的研究领域; 在万维网中, 不同社团可能表示了不同主题的主页; 在新陈代谢网络、神经网络中, 社团可能反映了功能单位; 在食物链网中, 社团可能反映了生态系统中的子系统. 在网络性质和功能的研究中, 社团结构也有显著的表现. 请读者思考利用拉普拉斯矩阵的特征值对图 2-1 中的网络进行社团划分(显然, 我们可划分为三个社团). 更多关于社团研究的论文请读者阅读参考文献[21-22]及其后引用文献.

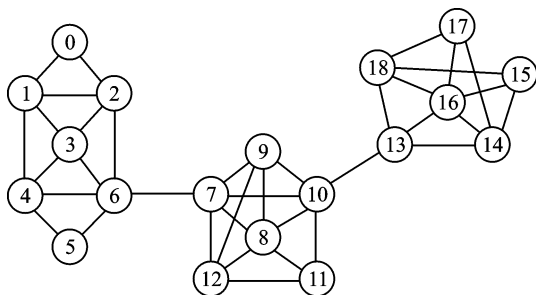


图 2-1 社团划分

第 3 章

矩阵分解

矩阵分解是将矩阵拆解为多个矩阵的乘积. 矩阵分解的一个常见类比是整数的因子分解, 例如将数字 12 分解为 2×6 或 3×4 或 $2 \times 2 \times 3$ 等形式. 类似于整数因子分解的多样性, 矩阵分解也有多种技术. 确定矩阵在不同情况下的分解形式, 对矩阵相关的理论分析和数值计算有着极为重要的意义. 因为矩阵的这些分解形式, 一般能明显地反映出原矩阵的某些特征(从线性变换的角度看, 分解形式反映了原线性变换的某些特征), 其涉及的分解方法、过程等可提供行之有效的理论分析或数值计算依据.

3.1 相抵分解

矩阵相抵是同一线性映射在不同基下的矩阵之间的关系, 这种关系也可以看作一种矩阵分解. 我们这里简单回顾一下.

引理 3.1.1 设矩阵 $A, B \in F^{m \times n}$, 则以下表述等价:

- (1) A 与 B 相抵;
- (2) 存在可逆矩阵 $P \in F^{m \times m}$ 和 $Q \in F^{n \times n}$ 使得 $A = PBQ$;
- (3) 矩阵 A 与 B 均可通过有限次初等行列变换得到同一个矩阵;
- (4) $\text{rank}(A) = \text{rank}(B)$.

注 1: 引理 3.1.1 性质(3)经初等行列变换所得矩阵的最简形式为 $\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$, 称为矩阵 A (或 B) 的相抵标准形, 其中 r 为矩阵 A (或 B) 的秩.

引理 3.1.2 设 $A \in F^{m \times n}$ 的秩为 r , 则存在可逆矩阵 $P \in F^{m \times m}$ 和 $Q \in F^{n \times n}$ 使得

$$A = P \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} Q \quad (3.1.1)$$

表达式(3.1.1)称为矩阵 A 的相抵分解式.

【应用】 矩阵相抵分解适用于任何矩阵,因此它是一种基本的矩阵分解.相抵分解常用于解决矩阵秩相关的命题.例如 $A \in \mathbb{C}^{m \times n}$,利用矩阵的相抵分解和矩阵乘积的转置公式容易证明矩阵的转置(或共轭转置)不改变矩阵的秩.

例 3.1.1 设 $A \in \mathbb{C}^{m \times n}$, $B \in \mathbb{C}^{n \times p}$,证明: $\text{rank}(A) + \text{rank}(B) - n \leq \text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$.

证明:首先证明第二个不等号成立.设 $A = P \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} Q$,并将矩阵 QB 按行分块,记为 $QB = \begin{bmatrix} C \\ D \end{bmatrix}$,则有

$$AB = P \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = P \begin{bmatrix} C \\ O \end{bmatrix} \quad (3.1.2)$$

式中: $P \in \mathbb{C}^{m \times m}$ 和 $Q \in \mathbb{C}^{n \times n}$ 是可逆矩阵, $C \in \mathbb{C}^{r \times p}$, $D \in \mathbb{C}^{(n-r) \times p}$, $r = \text{rank}(A)$.

由式(3.1.2)得:

$$\text{rank}(AB) = \text{rank} \left(\begin{bmatrix} C \\ O \end{bmatrix} \right) = \text{rank}(C)$$

又知 $\text{rank}(C) \leq r$ 且 $\text{rank}(C) \leq \text{rank}(QB) = \text{rank}(B)$.

因此, $\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$.

然后证明第一个不等号成立.由 $QB = \begin{bmatrix} C \\ D \end{bmatrix}$ 知, $\text{rank}(QB) \leq \text{rank}(D) + \text{rank}(C)$,故

$$\text{rank}(B) \leq \text{rank}(D) + \text{rank}(AB)$$

又知 $\text{rank}(D) \leq n - r$.因此, $\text{rank}(A) + \text{rank}(B) - n \leq \text{rank}(AB)$.证毕.

例 3.1.1 中的两个不等式是矩阵分析的重要不等式,在诸多的涉及矩阵乘积和矩阵秩的关系时往往需要根据这两个不等式进行放缩.请利用例 3.1.1 中的结论证明本章的习题 1.

在矩阵分解中,我们关心的一个基础问题是分解的唯一性问题.在随后矩阵分解的章节中,我们总会讨论这一问题.下面简单讨论相抵分解的唯一性问题.实际上,根据矩阵相抵的定义就能初步断定,由于初等行列变换次序不同可能会导致不同的相抵分解.例如,

$$A = P \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} Q = \left(P \begin{bmatrix} I_r & X \\ O & I \end{bmatrix} \right) \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \left(\begin{bmatrix} I_r & O \\ Y & I \end{bmatrix} Q \right) \quad (3.1.3)$$

式中: X 和 Y 是适当阶数的任意矩阵.由此,矩阵 A 的相抵分解是不唯一的.

【应用】 相抵分解的一个典型应用是求解矩阵的广义逆,请读者完成本章的习题 2.矩阵广义逆及相关知识将在后续章节详细展开,这里不再具体介绍.

相抵分解亦可用于线性方程组的求解问题.考察线性方程组 $Ax = b$ 的求解问题,其中 $A \in \mathbb{R}^{m \times n}$ 和 $b \in \mathbb{R}^m$ 已知, $x \in \mathbb{R}^n$ 是未知向量.

首先将矩阵 A 进行相抵分解,得

$$P \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} Qx = b \quad (3.1.4)$$

再令 $y = Qx$, 式(3.1.4)改写为

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix} y = P^{-1}b \quad (3.1.5)$$

由式(3.1.5)可解得 y , 再根据 $x = Q^{-1}y$ 求出 x , 其中 P^{-1} 和 Q^{-1} 可通过如下变换求得:

$$\begin{bmatrix} P^{-1} & O \\ O & I \end{bmatrix} \begin{bmatrix} A & I \\ I & * \end{bmatrix} \begin{bmatrix} Q^{-1} & O \\ O & I \end{bmatrix} = \begin{bmatrix} I_r & O & P^{-1} \\ O & O & * \\ Q^{-1} & & \end{bmatrix} \quad (3.1.6)$$

例 3.1.2 求解线性方程组 $Ax = b$, 其中

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

解: 由 A 表达式易知 $\text{rank}(A) = 2$, 并对如下分块矩阵进行初等变换:

$$\begin{bmatrix} A & I \\ I & * \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 2 & 0 & 1 \\ 1 & 0 & 0 & & \\ 0 & 1 & 0 & * & \\ 0 & 0 & 1 & & \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 1 \\ 1 & -1 & -1 & & \\ 0 & 0 & 1 & & * \\ 0 & 1 & 0 & & \end{bmatrix}$$

进而得

$$P^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \quad Q^{-1} = \begin{bmatrix} 1 & -1 & -1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad P^{-1}b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

由方程组(3.1.5)易得 $y = [1, 1, a]^T$, 进而解得

$$x = Q^{-1}y = \begin{bmatrix} -a \\ a \\ 1 \end{bmatrix}$$

上式即为线性方程组 $Ax = b$ 的解, 其中 a 为任意实数.

3.2 满秩分解

满秩分解是一种基本的矩阵分解, 其分解定理如下:

定理 3.2.1(满秩分解定理) 设 $A \in \mathbb{C}_r^{m \times n}$ ($r > 0$), 则存在列满秩矩阵 B 和行满秩矩阵 C 使得 $A = BC$.

证明:

方法一: 利用相抵分解证明. 由 $A = P \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} Q$ 和 $\begin{bmatrix} I_r & O \\ O & O \end{bmatrix} = \begin{bmatrix} I_r \\ O \end{bmatrix} [I_r, O]$ 知,

$$A = P \begin{bmatrix} I_r \\ O \end{bmatrix} [I_r, O] Q$$

令 $B = P \begin{bmatrix} I_r \\ O \end{bmatrix}$, $C = [I_r, O] Q$, 则易证 B 和 C 分别是列满秩矩阵和行满秩矩阵.

方法二: 利用线性空间性质证明. 将矩阵 A 进行列分块, 并令 $A = [a_1, \dots, a_n]$, 则有 $R(A) = \text{span}(a_1, \dots, a_n)$ 且 $\text{rank}(A) = \dim(R(A))$, 其中 $a_i \in \mathbb{C}^m, i = 1, \dots, n$.

任取 $R(A)$ 的一组基, 则 a_i 必可由基 b_1, \dots, b_r 线性表示, 即

$$a_i = [b_1, \dots, b_r] c_i$$

式中: $c_i \in \mathbb{C}^r$ 是 a_i 在基 b_1, \dots, b_r 下的坐标, $i = 1, \dots, n$.

进一步, 定义矩阵 $B = [b_1, \dots, b_r]$, $C = [c_1, \dots, c_n]$, 有

$$A = [a_1, \dots, a_n] = [b_1, \dots, b_r] [c_1, \dots, c_n] = BC$$

式中: $\text{rank}(B) = \dim(R(B)) = r$.

又知

$$\text{rank}(C) \geq \text{rank}(A) = r \text{ 且 } \text{rank}(C) \leq r$$

故 $\text{rank}(C) = r$. 综上所述, 矩阵 A 可分解为列满秩矩阵 B 和行满秩矩阵 C 的乘积.

注 1: 由于列空间 $R(A)$ 的基选取不同, 矩阵 A 的满秩分解也不唯一. 实际上, 上述证明也提供了一种满秩分解的计算方法.

例 3.2.1 求矩阵 $A = \begin{bmatrix} i & 1 & 1 \\ 1 & -i & 1 \end{bmatrix}$ 的满秩分解.

解: $a_1 = \begin{bmatrix} i \\ 1 \end{bmatrix}$ 和 $a_2 = \begin{bmatrix} 1 \\ -i \end{bmatrix}$ 线性相关, $a_1 = \begin{bmatrix} -i \\ 1 \end{bmatrix}$ 和 $a_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ 线性无关, 故可选取

a_1 和 a_3 构成列空间 $R(A)$ 的一组基.

定义

$$B = [a_1, a_3] = \begin{bmatrix} i & 1 \\ 1 & 1 \end{bmatrix}$$

则向量 a_1, a_2, a_3 在基 a_1, a_3 下的坐标 c_1, c_2, c_3 分别为

$$c_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad c_2 = \begin{bmatrix} -i \\ 0 \end{bmatrix}, \quad c_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

因此, 定义

$$C = [c_1, c_2, c_3] = \begin{bmatrix} 1 & -i & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

则 $A=BC$ 是矩阵 A 的满秩分解.

若选取

$$B=[a_2, a_3]=\begin{bmatrix} 1 & 1 \\ -i & 1 \end{bmatrix}$$

则矩阵 C 为如下形式:

$$C=\begin{bmatrix} i & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

注 2: 在挑选列空间 $R(A)$ 的基时,并不一定从 A 的列向量中选取. 比如在例 3.2.1 中,列满秩矩阵 B 也可选取为单位矩阵(单位矩阵的列向量不是矩阵 A 的任一系列向量),则 $A=I_2A$ 也是 A 的满秩分解. 之所以出现 $A=IA$ 形式的满秩分解,原因在于矩阵 A 本身是行满秩矩阵. 实际上,若 A 是列满秩矩阵,则 $A=AI$ 是 A 的满秩分解形式.

尽管满秩分解不唯一,但同一矩阵的不同满秩分解存在着一定关系.

定理 3.2.2 设 $A \in \mathbb{C}_r^{m \times n} (r > 0)$, $A=B_1C_1$ 和 $A=B_2C_2$ 是矩阵 A 的两种不同满秩分解,则存在可逆矩阵 $D \in \mathbb{C}^{r \times r}$ 使得 $B_1=B_2D$ 和 $C_1=D^{-1}C_2$.

证明: 将矩阵 B_1 和 B_2 进行列分块,并记为

$$B_1=[b_{11}, \dots, b_{1r}], \quad B_2=[b_{21}, \dots, b_{2r}]$$

由于 B_1 和 B_2 均为列满秩矩阵,则向量组 b_{11}, \dots, b_{1r} 和向量组 b_{21}, \dots, b_{2r} 均构成列空间 $R(A)$ 的一组基. 根据同一空间中不同基的关系,得

$$[b_{11}, \dots, b_{1r}] = [b_{21}, \dots, b_{2r}]D$$

式中: D 是从基 b_{21}, \dots, b_{2r} 到基 b_{11}, \dots, b_{1r} 的过渡矩阵.

因此,存在可逆矩阵 D 使得 $B_1=B_2D$. 再根据坐标与过渡矩阵的关系,易证 $C_1=D^{-1}C_2$. 证毕.

注 3: 定理 3.2.2 表明只要找到矩阵 A 的一个满秩分解表达式就可以构造无数个满秩分解.

【应用】 满秩分解是研究矩阵广义逆的一个强有力的工具. 请读者尝试完成本章的习题 2,关于矩阵广义逆内容将在第 5 章详细介绍.

例 3.2.2 求满足等式 $AB=I$ 或 $BA=I$ 的矩阵 B , 其中,矩阵 A 分别为

$$(1) A=\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}; (2) A=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}; (3) A=\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

解: 经计算,由(1)中的矩阵 A 知,不存在矩阵 B 使得 $AB=I$ 或 $BA=I$; 对于(2)中矩阵 A ,存在矩阵 B 使得 $AB=I$; 对于(3)中矩阵 A ,存在矩阵 B 使得 $BA=I$.

由例 3.2.2 知,尽管对于(2)、(3)中的矩阵 A 存在矩阵 B 使得 $AB=I$ 或 $BA=I$,但

矩阵 B 却不是相应矩阵 A 的逆矩阵(由于逆矩阵须同时满足 $AB=I$ 和 $BA=I$ 两个等式;对非方矩阵而言,最多只能满足这两个等式之一).

定理 3.2.3(右逆和左逆) 矩阵 $A \in \mathbb{C}_r^{m \times n}$ ($r > 0$) 有右逆(即存在矩阵 B 使得 $AB=I$) 的充分必要条件是 A 为行满秩矩阵;矩阵 A 有左逆(即存在矩阵 B 使得 $BA=I$) 的充分必要条件是 A 为列满秩矩阵.

具体证明留作课后习题. 定理 3.2.3 给出了矩阵左逆和右逆存在性的充分必要条件,但并未指出矩阵左逆或右逆的唯一性结果. 回顾例 3.2.2,对于(2)中的矩阵 A ,其右逆不唯一,可表示为

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

式中: $*$ 为任意复数.

实际上,只有当矩阵 A 为可逆矩阵时,其左逆和右逆才是唯一的;对于其他矩阵而言,其左逆或右逆或不存在,或存在但不唯一.

注 4: 设 $A \in \mathbb{C}_r^{r \times n}$, 则 AA^H 是 r 阶非奇异矩阵(利用本章习题 7 的结论). 根据 $AA^H(AA^H)^{-1}=I$, 得 $A^H(AA^H)^{-1}$ 是矩阵 A 的一个右逆. 同理, 当 $A \in \mathbb{C}_r^{m \times r}$ 时, $(A^HA)^{-1}A^H$ 是矩阵 A 的一个左逆.

3.3 三角分解

定义 3.3.1(三角矩阵) 设矩阵 $A = [a_{ij}] \in \mathbb{C}^{n \times n}$, 若 A 的对角线上(下)方的元素全为零, 即 $\forall i < j, a_{ij} = 0$ ($\forall i > j, a_{ij} = 0$), 则称矩阵 A 为下(上)三角矩阵. 通常将下三角矩阵和上三角矩阵统称为三角矩阵. 进一步, 将对角线元素全为正实数的三角矩阵称为正线三角矩阵, 将对角线元素全为 1 的三角矩阵称为单位三角矩阵.

三角矩阵是一类特殊矩阵, 其特殊之处在于任意两个下(上)三角矩阵的加法、乘积以及其逆矩阵(若存在)仍是下(上)三角矩阵. 特别地, 任意两个单位下(上)三角矩阵的乘积以及逆矩阵(若存在)仍是单位下(上)三角矩阵.

定理 3.3.1(LU 分解定理) 设 $A = [a_{ij}] \in \mathbb{C}^{n \times n}$ 是非奇异矩阵, 则存在唯一的单位下三角矩阵 L 和上三角矩阵 U 使得 $A=LU$ 成立的充分必要条件是 A 的所有顺序主子式均非零, 即

$$\Delta_i(A) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1i} \\ a_{21} & a_{22} & \cdots & a_{2i} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ii} \end{vmatrix} \neq 0, \quad i = 1, \cdots, n$$

证明: 必要性. 已知存在唯一的单位下三角矩阵 L 和上三角矩阵 U 使得 $A =$

LU , 需证明 A 的所有顺序主子式非零(可考虑将矩阵 A 分块, 然后通过分块矩阵找到矩阵 A 的顺序主子式矩阵).

将矩阵 A 、 L 和 U 作如下分块:

$$A = \begin{bmatrix} A_{i1} & A_{i2} \\ A_{i3} & A_{i4} \end{bmatrix} = \begin{bmatrix} L_{i1} & O \\ L_{i3} & L_{i4} \end{bmatrix} \begin{bmatrix} U_{i1} & U_{i2} \\ O & U_{i4} \end{bmatrix} \quad (3.3.1)$$

式中: A_{i1} 、 L_{i1} 和 U_{i1} 分别为矩阵 A 、 L 和 U 的 i 阶顺序主子式矩阵, $i=1, \dots, n$.

由式(3.3.1)知, $A_{i1} = L_{i1}U_{i1}$. 从而有

$$|A_{i1}| = |L_{i1}U_{i1}| = |U_{i1}| = u_{11} \cdot u_{22} \cdot \dots \cdot u_{ii}$$

式中: u_{ii} 是上三角矩阵 U 的第 i 个对角元素, $i=1, \dots, n$.

由矩阵 A 可逆推断出矩阵 U 可逆, 进而 $u_{11} \cdot u_{22} \cdot \dots \cdot u_{ii} \neq 0$. 因此, $u_{ii} \neq 0$, $\forall i=1, \dots, n$, 即 A 的所有顺序主子式非零.

【充分性分析: 矩阵分解的存在性证明一般通过找出分解式来说明. 因此, 这类问题通常采用构造法、归纳法等. 这里采用归纳法说明. **】**

充分性. 当矩阵 A 的阶数为 1 时, 结论显然成立. 假设对任意 $(n-1)$ 阶矩阵 A_{n-1} 有 LU 分解, 即

$$A_{n-1} = L_{n-1}U_{n-1}$$

式中: L_{n-1} 是 $(n-1)$ 阶单位下三角矩阵, U_{n-1} 是 $(n-1)$ 阶上三角矩阵.

现考察任意 n 阶矩阵 A . 为利用 $(n-1)$ 阶矩阵的 LU 分解式, 矩阵 A 可作如下分块处理

$$A = \begin{bmatrix} A_{n-1} & \beta \\ \alpha^T & a_{nn} \end{bmatrix}$$

式中: A_{n-1} 是矩阵 A 的 $n-1$ 阶顺序主子式矩阵, β 和 α 是 $n-1$ 维向量, a_{nn} 是标量.

若矩阵 A 可进行 LU 分解, 则须满足

$$\begin{bmatrix} A_{n-1} & \beta \\ \alpha^T & a_{nn} \end{bmatrix} = \begin{bmatrix} L_{n-1} & 0 \\ \alpha_L^T & 1 \end{bmatrix} \begin{bmatrix} U_{n-1} & \beta_U \\ 0 & u_{nn} \end{bmatrix}$$

式中: α_L^T 、 β_U 和 u_{nn} 待定且须满足以下条件:

$$\beta = L_{n-1}\beta_U$$

$$\alpha^T = \alpha_L^T U_{n-1}$$

$$a_{nn} = \alpha_L^T \beta_U + u_{nn}$$

由上式解得 $\beta_U = L_{n-1}^{-1}\beta$, $\alpha_L^T = \alpha^T U_{n-1}^{-1}$ 和 $u_{nn} = a_{nn} - \alpha^T A_{n-1}^{-1}$. 从而构造出 n 阶矩阵 A 的 LU 分解, 即分解存在性得证.

【注意: 该定理的充分性分析尚未结束. 上述证明仅给出存在性说明, 还须证明分解的唯一性. 矩阵分解的唯一性证明常用反证法. **】**

这里采用反证法证明. 假设上述分解是不唯一的, 即存在两种不同的 LU 分解,

并表示为

$$\mathbf{A} = \mathbf{L}_1 \mathbf{U}_1 = \mathbf{L}_2 \mathbf{U}_2$$

对上式等价变换得

$$\mathbf{L}_2^{-1} \mathbf{L}_1 = \mathbf{U}_2 \mathbf{U}_1^{-1} \quad (3.3.2)$$

其中,式(3.3.2)左边是单位下三角矩阵,式(2.3.2)右边是上三角矩阵.

为保证式(3.3.2)成立,有

$$\mathbf{L}_2^{-1} \mathbf{L}_1 = \mathbf{U}_2 \mathbf{U}_1^{-1} = \mathbf{I}$$

由此, $\mathbf{L}_1 = \mathbf{L}_2$, $\mathbf{U}_1 = \mathbf{U}_2$, 分解唯一性得证. 证毕.

例 3.3.1 求矩阵 $\mathbf{A} = \begin{bmatrix} 1 & 2 & -1 \\ 3 & 1 & 0 \\ -1 & -1 & -2 \end{bmatrix}$ 的 LU 分解.

分析: 可利用初等变换求 LU 分解, 即

$$[\mathbf{A}, \mathbf{I}] \rightarrow [\mathbf{U}, \mathbf{L}^{-1}]$$

解: 对 $[\mathbf{A}, \mathbf{I}]$ 进行初等变换得

$$\left[\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 3 & 1 & 0 & 0 & 1 & 0 \\ -1 & -1 & -2 & 0 & 0 & 1 \end{array} \right] \rightarrow \left[\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & -5 & 3 & -3 & 1 & 0 \\ 0 & 0 & -\frac{12}{5} & \frac{2}{5} & \frac{1}{5} & 1 \end{array} \right]$$

于是

$$\mathbf{U} = \begin{bmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \\ 0 & 0 & -\frac{12}{5} \end{bmatrix}, \quad \mathbf{L}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ \frac{2}{5} & \frac{1}{5} & 1 \end{bmatrix}$$

利用 \mathbf{L}^{-1} 求出单位下三角矩阵 \mathbf{L} , 得

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -1 & -\frac{1}{5} & 1 \end{bmatrix}$$

注 1: 非奇异上三角矩阵 \mathbf{U} 可进一步分解为对角矩阵和单位上三角矩阵的乘积, 即

$$\begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1,n-1} & u_{1n} \\ 0 & u_{22} & \cdots & u_{2,n-1} & u_{2n} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & u_{n-1,n-1} & u_{n-1,n} \\ 0 & 0 & \cdots & 0 & u_{nn} \end{bmatrix} =$$

$$\begin{bmatrix} u_{11} & & & & \\ & u_{22} & & & \\ & & \ddots & & \\ & & & u_{n-1,n-1} & \\ & & & & u_{nn} \end{bmatrix} \begin{bmatrix} 1 & \frac{u_{12}}{u_{11}} & \frac{u_{12}}{u_{11}} & \cdots & \frac{u_{1n}}{u_{11}} \\ & 1 & \frac{u_{23}}{u_{22}} & \cdots & \frac{u_{2n}}{u_{22}} \\ & & \ddots & \ddots & \vdots \\ & & & 1 & \frac{u_{n-1,n}}{u_{n-1,n-1}} \\ & & & & 1 \end{bmatrix} \quad (3.3.3)$$

因此,总结如下定理:

定理 3.3.2 (LDU 分解定理) 设 $\mathbf{A} = [a_{ij}] \in \mathbb{C}^{n \times n}$ 是非奇异矩阵,则存在唯一的单位下三角矩阵 \mathbf{L} , 对角矩阵 \mathbf{D} 和单位上三角矩阵 \mathbf{U} 使得 $\mathbf{A} = \mathbf{LDU}$ 成立的充分必要条件是 \mathbf{A} 的所有顺序主子式均非零,即

$$\Delta_i(\mathbf{A}) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1i} \\ a_{21} & a_{22} & \cdots & a_{2i} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ii} \end{vmatrix} \neq 0, \quad i = 1, \cdots, n$$

分解式 $\mathbf{A} = \mathbf{LDU}$ 称为矩阵 \mathbf{A} 的 LDU 分解.

注 2: 定理 3.3.2 中对角矩阵 $\mathbf{D} = \text{diag}(d_1, \cdots, d_n)$ 可由矩阵 \mathbf{A} 的顺序主子式求得

$$d_1 = a_{11} \\ d_i = \frac{\Delta_i(\mathbf{A})}{\Delta_{i-1}(\mathbf{A})}, \quad i = 2, \cdots, n$$

注 3: 若定义 $\tilde{\mathbf{L}} = \mathbf{LD}$, 矩阵 \mathbf{A} 的分解为 $\mathbf{A} = \tilde{\mathbf{L}}\mathbf{U}$, 其中, $\tilde{\mathbf{L}}$ 是下三角矩阵, \mathbf{U} 是单位上三角矩阵. 这是定理 3.3.1 的另一种表达, 常称为 Crout 分解. 定理 3.3.1 常称为 Doolittle 分解.

注 4: 在定理 3.3.1 和定理 3.3.2 中, 矩阵 \mathbf{A} 的非奇异性仅作为相应定理的已知条件, 并非分解存在性的充分必要条件. 换言之, 若矩阵 \mathbf{A} 是奇异矩阵且可作 LU 分解, 但其分解是不唯一的 (见例 3.3.2); 若矩阵 \mathbf{A} 是非奇异矩阵, 其 LU 分解可能不存在 (见例 3.3.3).

例 3.3.2 求矩阵 \mathbf{A}_1 和 \mathbf{A}_2 的 LU 分解, 其中

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$

解: \mathbf{A}_1 和 \mathbf{A}_2 可分解为如下表达式:

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ x & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & y \end{bmatrix}$$

$$\mathbf{A}_2 = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ * & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & z \end{bmatrix}$$

式中: $x+y=2$. 故矩阵 \mathbf{A}_1 的 LU 分解不唯一; 注意到 $*$ 不存在, 故矩阵 \mathbf{A}_2 不能进行 LU 分解.

例 3.3.3 求矩阵 $\mathbf{A} = \begin{bmatrix} 0 & 4 & 6 \\ 0 & -3 & -5 \\ 1 & -3 & -6 \end{bmatrix}$ 的 LU 分解.

解: 若对矩阵 \mathbf{A} 进行 LU 分解, 则

$$\begin{bmatrix} 0 & 4 & 6 \\ 0 & -3 & -5 \\ 1 & -3 & -6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} 0 & 4 & 6 \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$$

式中: l_{31} 不存在, 故矩阵 \mathbf{A} 不能作 LU 分解.

推论 3.3.1 (Cholesky 分解) 若 n 阶实对称矩阵 \mathbf{A} 是正定的, 则存在唯一的正线上三角矩阵 \mathbf{R} 使得 $\mathbf{A} = \mathbf{R}^T \mathbf{R}$.

分析: 矩阵 \mathbf{A} 是非奇异矩阵且其所有顺序主子式均大于零, 因此可利用 LDU 分解证明 (分析发现 LU 分解并不适合). 由定理 3.3.2 知, $\mathbf{A} = \mathbf{L} \mathbf{D} \mathbf{U}$. 利用 \mathbf{A} 为实对称矩阵得, $\mathbf{A} = \mathbf{A}^T = \mathbf{U}^T \mathbf{D} \mathbf{L}^T$. 根据 LDU 分解的唯一性推导出

$$\mathbf{L} = \mathbf{U}^T, \quad \mathbf{U} = \mathbf{L}^T$$

因此, $\mathbf{A} = \mathbf{U}^T \mathbf{D} \mathbf{U}$. 注意到对角矩阵 \mathbf{D} 中元素均大于零, 故可将 \mathbf{D} 分解为两个正定对角矩阵的乘积:

$$\mathbf{D} = \mathbf{D}^{\frac{1}{2}} \mathbf{D}^{\frac{1}{2}}, \quad \mathbf{D}^{\frac{1}{2}} = \text{diag}(\sqrt{d_1}, \dots, \sqrt{d_n})$$

显然, 推论 3.3.1 中的 \mathbf{R} 可定义为 $\mathbf{R} = \mathbf{D}^{\frac{1}{2}} \mathbf{U}$. Cholesky 分解的唯一性可通过反证法证明, 这里不再赘述.

例 3.3.4 求正定矩阵 $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ 的 Cholesky 分解.

解: 方法一: 将矩阵 \mathbf{A} 进行 LDU 分解, 得

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & \frac{3}{2} \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix}$$

令

$$\mathbf{R} = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \frac{\sqrt{6}}{2} \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & \frac{\sqrt{2}}{2} \\ 0 & \frac{\sqrt{6}}{2} \end{bmatrix}$$

则有 $\mathbf{A} = \mathbf{R}^T \mathbf{R}$.

方法二：设上三角矩阵 $\mathbf{R} = [r_{ij}]_{2 \times 2}$ ，直接比较 $\mathbf{A} = \mathbf{R}^T \mathbf{R}$ 的两端可知

$$a_{11} = r_{11}^2, \quad a_{12} = r_{11}r_{12}, \quad a_{22} = a_{12}^2 + a_{22}^2$$

解得

$$r_{11} = \sqrt{2}, \quad r_{12} = \frac{\sqrt{2}}{2}, \quad r_{22} = \frac{\sqrt{6}}{2}$$

例 3.3.4 求解方法二称为 Cholesky 算法. 请读者自行设计一般 n 阶正定实对称矩阵的 Cholesky 算法.

【应用】 Cholesky 分解的应用之一可参考本章习题 12. 下面讨论 LU 分解在线性方程组求解方面的应用. 设线性方程组为 $\mathbf{Ax} = \mathbf{b}$. 若系数矩阵 \mathbf{A} 有 LU 分解, 则

$$\mathbf{Ax} = (\mathbf{LU})\mathbf{x} = \mathbf{L}(\mathbf{Ux}) = \mathbf{b}$$

由此, 方程组 $\mathbf{Ax} = \mathbf{b}$ 等价转化为系数矩阵均是三角矩阵的两个线性方程组

$$\begin{cases} \mathbf{Ly} = \mathbf{b} \\ \mathbf{Ux} = \mathbf{y} \end{cases} \quad (3.3.4)$$

显然, 方程组 (3.3.4) 的求解不需要对一般矩阵 \mathbf{A} 进行求逆计算, 仅需利用迭代法即可简单快速求解, 这是计算机对大型线性方程组求解时常采用的方法. 这是因为 n 阶稠密矩阵的 LU 分解大约需要 $\frac{2}{3}n^3$ 次浮点运算, 而求 \mathbf{A}^{-1} 约需要 $2n^3$ 次浮点运算. 若 \mathbf{A} 是稀疏矩阵, 则 \mathbf{L} 和 \mathbf{U} 可能也是稀疏的, 但 \mathbf{A}^{-1} 很能是稠密; 这时用 LU 分解解方程组 $\mathbf{Ax} = \mathbf{b}$ 很可能比用 \mathbf{A}^{-1} 快很多. 更详细的内容请读者了解 Gauss 消元法, 这里不再赘述.

例 3.3.5 求解线性方程组:

$$\begin{bmatrix} 1 & -1 & 5 \\ 1 & 2 & -1 \\ 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 5 \end{bmatrix}$$

解: 将系数矩阵进行 LU 分解得

$$\begin{bmatrix} 1 & -1 & 5 \\ 1 & 2 & -1 \\ 2 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & \frac{5}{3} & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 5 \\ 0 & 3 & -6 \\ 0 & 0 & 4 \end{bmatrix}$$

故原方程组可化为以下两个方程组:

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & \frac{5}{3} & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 5 \end{bmatrix}, \quad \begin{bmatrix} 1 & -1 & 5 \\ 0 & 3 & -6 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

依次求解, 得

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 3 \\ -3 \\ 4 \end{bmatrix}, \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

【思考】若将例 3.3.5 中的系数矩阵 A 替换为如下矩阵:

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

则根据例 3.3.3 知,此时无法利用 LU 分解求解线性方程组 $Ax=b$. 但实际上,由于系数矩阵 A 是非奇异的,该方程组一定存在唯一解 $x=A^{-1}b$. 面临矩阵 A 无法进行 LU 分解的困难,应如何处理这一问题呢?

分析:在求解方程组的过程中,调整方程的顺序并不会改变未知变量和方程的解.因此,解决方案之一是对系数矩阵 A 进行初等行变换,使得变换后的矩阵所有顺序主子式非零.这样就可以对系数矩阵 A 进行 LU 分解了.现对例 3.3.3 中矩阵 A 进行初等行变换(将第 1 和第 3 行互换)得到

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

显然,矩阵 \tilde{A} 的所有顺序主子式均非零,可进行 LU 分解.那么,所有的非奇异矩阵都可以通过初等行变换实现变换后矩阵的所有顺序主子式非零吗?为解决这一问题,引入排列矩阵定义.

定义 3.3.2(排列矩阵) 设 e_i 是 n 阶单位矩阵 I 的第 i 个列向量,则矩阵 $P=[e_{i_1}, \dots, e_{i_n}]$ 称为一个 n 阶排列矩阵(或置换矩阵),其中, i_1, \dots, i_n 是 $1, \dots, n$ 的一个排列.

命题 3.3.1 若 P 是排列矩阵,则 P^T 和 P^{-1} 也是排列矩阵,且 $P^T=P^{-1}$.

注 5:将矩阵 A 的行按照 i_1, \dots, i_n 的次序重排,即排列矩阵 $[e_{i_1}, \dots, e_{i_n}]$ 左乘矩阵 A ;将 A 的列按照 i_1, \dots, i_n 的次序重排,即排列矩阵 $[e_{i_1}, \dots, e_{i_n}]$ 右乘矩阵 A .

引理 3.3.1 设 $A=[a_{ij}] \in \mathbb{C}^{n \times n}$ 是非奇异矩阵,则存在排列矩阵 P 使得 PA 的所有顺序主子式均非零.

证明:若 $a_{11}=0$,则必定存在 $a_{1j} \neq 0, j \in \{2, \dots, n\}$ (若不存在,则矩阵 A 奇异). 定义

$$P_1 = [e_j, e_2, \dots, e_{j-1}, e_1, e_{j+1}, \dots, e_n]$$

则矩阵 $P_1 A$ 的 1 阶顺序主子式不为零.若 $a_{11} \neq 0$,则定义 $P_1 = I$.

假设矩阵 $P_1 A$ 的 i 阶顺序主子式 $|D_i| \neq 0$, 其中, $i=1, \dots, i^*$, 而 $|D_{i^*+1}|=0$ (若矩阵 $P_1 A$ 不存在满足条件的 i^* , 意味着 $P_1 A$ 的所有顺序主子式均非零, 结论自然成立). 现证明存在排列矩阵 P_{i^*} 使得 $\forall i=1, \dots, i^*+1$, 矩阵 $P_{i^*} P_1 A$ 的 i 阶顺序主子

式均不为零.

由于矩阵 $P_1 A$ 的 i^* 阶顺序主子式 $|D_{i^*}| \neq 0$, 故 $P_1 A$ 的 i^* 阶顺序主子式矩阵 $D_{i^*} \in \mathbb{C}^{i^* \times i^*}$ 是非奇异矩阵, 即矩阵 D_{i^*} 的 i^* 个列向量线性无关. 当考察 $P_1 A$ 的 (i^*+1) 阶顺序主子式矩阵 D_{i^*+1} 时, 该矩阵前 i^* 行向量必线性无关 (由于该矩阵的前 i^* 列向量线性无关; 当这 i^* 列向量构成的向量组再增加一行时, 由新向量组构成的矩阵要么秩保持不变, 要么秩增加 1; 注意到新向量组构成的矩阵是 $i^*(i^*+1)$ 阶矩阵, 其秩不会超过 i^* , 故该矩阵的秩只能是 i^*).

注意到 $P_1 A$ 是非奇异矩阵 (A 是 n 阶非奇异矩阵), 故 $P_1 A$ 的前 (i^*+1) 列向量线性无关, 即由 $P_1 A$ 前 (i^*+1) 列向量构成的矩阵 A_{i^*+1} 的秩为 (i^*+1) . 又知矩阵 A_{i^*+1} 的前 i^* 行向量线性无关, 则必可从矩阵 A_{i^*+1} 余下的 $(n-i^*)$ 行向量找到与前 i^* 行向量均线性无关的向量 (若不存在, 则 A_{i^*+1} 的秩为 i^*). 不妨设 A_{i^*+1} 第 k 行向量与前 i^* 行向量均线性无关 ($i^*+1 \leq k \leq n$), 则定义

$$P_{i^*} = [e_1, \dots, e_{i^*}, e_k, e_{i^*+2}, \dots, e_{k-1}, e_{i^*+1}, e_{k+1}, \dots, e_n]$$

有 $\forall i=1, \dots, i^*+1$, 矩阵 $P_{i^*} P_1 A$ 的 i 阶顺序主子式 $|D_i| \neq 0$.

同理, 存在若干个排序矩阵 $P_n \cdots P_{i^*} P_1 A$ 使得 $P_n \cdots P_{i^*} P_1 A$ 的顺序主子式均非零. 定义 $P = P_n \cdots P_{i^*} P_1$, 则 P 仍是一个排列矩阵. 证毕.

引理 3.3.1 从理论上解决了非奇异矩阵的 LU 分解问题.

例 3.3.6 求解线性方程组

$$\begin{bmatrix} 0 & 4 & 6 \\ 0 & -3 & -5 \\ 1 & -3 & -6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ -2 \end{bmatrix}$$

解: 由于系数矩阵 A 的 1 阶和 2 阶顺序主子式均等于零, 故定义排列矩阵

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

则有

$$PA = \begin{bmatrix} 1 & -3 & -6 \\ 0 & -3 & -5 \\ 0 & 4 & 6 \end{bmatrix}$$

对 PA 进行 LU 分解得

$$\begin{bmatrix} 1 & -3 & -6 \\ 0 & -3 & -5 \\ 0 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{4}{3} & 1 \end{bmatrix} \begin{bmatrix} 1 & -3 & -6 \\ 0 & -3 & -5 \\ 0 & 0 & -\frac{2}{3} \end{bmatrix}$$

然后仿照例 3.3.5 步骤进行求解, 得

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

3.4 QR 分解

定义 3.4.1 (QR 分解) 若复方阵 A 可分解为 $A=QR$, 其中, Q 为酉矩阵, R 为上三角矩阵, 则称矩阵 A 可作 QR 分解 (或酉三角分解). 若分解式 $A=QR$ 中, 矩阵 A 是实方阵, Q 为正交矩阵, R 为上三角矩阵, 此时称分解式 $A=QR$ 为正交三角分解.

定理 3.4.1 若实方阵 A 满秩, 则存在正交矩阵 Q 及正线上三角矩阵 R 满足 $A=QR$ 且分解唯一.

证明: 将实方阵 A 按列分块, 并记为 $A=[a_1, \cdots, a_n]$, 则向量组 a_1, \cdots, a_n 是 \mathbb{R}^n 空间的一组基. 由 Gram-Schmidt 正交化方法知, 根据基 a_1, \cdots, a_n 可构造出 \mathbb{R}^n 的一组标准正交基 z_1, \cdots, z_n , 且这两组基有如下关系:

$$[a_1, \cdots, a_n] = [z_1, \cdots, z_n]R \quad (3.4.1)$$

式中: $y_k = x_k - \sum_{i=1}^{k-1} (x_k, z_i) z_i, k=1, \cdots, n$, 矩阵 R 定义为

$$R = \begin{bmatrix} \|y_1\| & (a_2, z_1) & \cdots & (a_n, z_1) \\ & \|y_2\| & \cdots & (a_n, z_2) \\ & & \ddots & \vdots \\ \mathbf{O} & & & \|y_n\| \end{bmatrix}$$

进一步将式(3.4.1)写成矩阵形式, 即 $A=QR$, 其中, $Q=[z_1, \cdots, z_n]$, 显然, Q 是正交矩阵, R 是正线上三角矩阵.

现采用反证法证明 QR 分解的唯一性. 假设存在两种不同的 QR 分解, 即

$$A = Q_1 R_1 = Q_2 R_2$$

令 $\tilde{Q} = Q_2^T Q_1, \tilde{R} = R_2 R_1^{-1}$, 则有 $\tilde{Q} = \tilde{R}$, 其中, \tilde{Q} 和 \tilde{R} 既为正交矩阵, 又为正线上三角矩阵. 另记

$$\tilde{R} = \begin{bmatrix} \tilde{r}_{11} & \tilde{r}_{12} & \cdots & \tilde{r}_{1n} \\ & \tilde{r}_{22} & \cdots & \tilde{r}_{2n} \\ & & \ddots & \vdots \\ \mathbf{O} & & & \tilde{r}_{nn} \end{bmatrix}$$

根据 $\tilde{R}^T \tilde{R} = I$, 得

$$\tilde{r}_{ii}^2 = 1, \quad i = 1, \cdots, n \quad (3.4.2)$$

$$\tilde{r}_{ij}=0, \quad j \neq i, i, j=1, \cdots, n$$

由于 $\tilde{\mathbf{R}}$ 的所有对角元素是正数, 解得 $\tilde{r}_{ii}=1, i=1, \cdots, n$. 进而 $\tilde{\mathbf{R}}=\mathbf{I}$, 即 $\mathbf{R}_1=\mathbf{R}_2$. 同时推导出 $\mathbf{Q}_1=\mathbf{Q}_2$. 因此, QR 分解是唯一的. 证毕.

注 1: 若一实方阵既是正交矩阵又是正线上三角矩阵, 则该矩阵一定是单位矩阵.

注 2: 若不要求上三角矩阵 \mathbf{R} 的对角元素全为正实数, 则方程 (3.4.2) 的解不唯一, 进而导致矩阵 \mathbf{A} 的 QR 分解不唯一. 例如, 考察 2 阶单位矩阵 \mathbf{I}_2 的 QR 分解问题. 显然,

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \stackrel{\text{def}}{=} \mathbf{Q}\mathbf{R}$$

是 \mathbf{A} 的 QR 分解,

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \stackrel{\text{def}}{=} \mathbf{Q}_1\mathbf{R}_1$$

也是 \mathbf{A} 的 QR 分解.

类似地, 对于可逆复方阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 我们有如下定理.

定理 3.4.2 设复方阵 \mathbf{A} 可逆, 则存在酉矩阵 \mathbf{U} 及正线上三角矩阵 \mathbf{R} 满足 $\mathbf{A} = \mathbf{U}\mathbf{R}$ 且分解唯一.

定理 3.4.2 的证明可参考定理 3.4.1. 这里不再具体展开.

【思考】 非方矩阵是否可作 QR 分解?

分析: 考察 \mathbf{A} 和 \mathbf{B} 两个矩阵

$$\mathbf{A} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \\ 4 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 3 & 3 \\ 0 & 0 \\ 4 & 4 \end{bmatrix}$$

根据 Gram-Schmidt 正交化方法, 矩阵 \mathbf{A} 和 \mathbf{B} 分别有如下分解:

$$\mathbf{A} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \\ 4 & 0 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & 0 \\ 0 & 1 \\ \frac{4}{5} & 0 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 3 & 3 \\ 0 & 0 \\ 4 & 4 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} \\ 0 \\ \frac{4}{5} \end{bmatrix} \begin{bmatrix} 5 & 5 \end{bmatrix}$$

由于矩阵 \mathbf{A} 是列满秩, 故其分解式为列正交规范矩阵 (满足 $\mathbf{Q}_1^H \mathbf{Q}_1 = \mathbf{I}_n$ 的 $m \times n$ 矩阵 \mathbf{Q}_1 称为列正交规范矩阵, \mathbf{Q}_1^H 称为行正交规范矩阵) 和正线上三角矩阵的乘积

(注意,非列满秩矩阵 B 则没有类似结论);因此,推广至一般的列满秩矩阵有如下分解:

$$A = \tilde{U}R \quad (3.4.3)$$

式中: $A \in \mathbb{C}_n^{m \times n}$, $\tilde{U} \in \mathbb{C}^{m \times n}$ 的列向量是列空间 $R(A)$ 的一组标准正交基, $R \in \mathbb{C}^{n \times n}$ 是正线上三角矩阵.

若将矩阵 A 和 B 分解为酉矩阵与类上三角矩阵(类上三角矩阵的阶数与矩阵 A 和 B 相同,一般为非方矩阵)的乘积,则可将上述分解中的正交规范矩阵的列向量补齐为 \mathbb{C}^m 空间中的一组标准正交基即可.此时,矩阵 A 和 B 可分解为

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 1 \\ 4 & 0 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & 0 & -\frac{4}{5} \\ 0 & 1 & 0 \\ \frac{4}{5} & 0 & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 3 & 3 \\ 0 & 0 \\ 4 & 4 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & 0 & -\frac{4}{5} \\ 0 & 1 & 0 \\ \frac{4}{5} & 0 & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 5 & 5 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

显然,对于行数不小于列数的矩阵 A ,总可以分解为酉矩阵 U 和类三角矩阵 R ,其中 $R = \begin{bmatrix} R_1 \\ O \end{bmatrix}_{m \times n}$, R_1 为上三角矩阵.进一步,若 A 为列满秩矩阵, R_1 为正线上三角矩阵;否则为奇异上三角矩阵.因此,有如下推论.

推论 3.4.1 矩阵 $A \in \mathbb{C}_n^{m \times n}$ 可分解为 $A = UR$, 其中, U 是 m 阶酉矩阵, $R = \begin{bmatrix} R_1 \\ O \end{bmatrix}_{m \times n}$, R_1 为正线上三角矩阵, $n \leq m$.

【应用】 QR 分解在矩阵计算中占据重要的地位.利用 QR 分解可解决诸多应用所涉及的最小二乘、特征值等问题.这里以线性方程组的求解应用和特征值的计算为例进行简要说明.

(1) 应用于线性方程组 $Ax = b$. 若对系数矩阵 A 进行 QR 分解,则原线性方程组可等价如下两个方程组:

$$Qy = b \quad (3.4.4)$$

$$Rx = y \quad (3.4.5)$$

式中: 方程(3.4.4)的解为 $y = Q^H b$ (Q 为酉矩阵,其逆矩阵只需共轭转置即可求得);方程(3.4.5)中的 R 为上三角矩阵,可先求解方程组(3.4.5)中最后一个方程,然后自此向前逐步求解每一个方程.

(2) 求矩阵的特征值、特征向量的一种有效方法: QR 算法(20 世纪在科学和工

程上有最大贡献与影响的十大算法之一,有兴趣的读者可阅读参考文献[23])是计算小规模矩阵所有特征值的稳定有效方法.

QR 算法的核心思想为: 给定可逆矩阵 $A \in \mathbb{R}^{n \times n}$, 定义 $A_1 = A$. 对 A_1 进行 QR 分解 $A_1 = Q_1 R_1$, 其中, 矩阵 Q_1 为正交矩阵, 矩阵 R_1 为上三角矩阵. 由此, 定义 A_2 为

$$A_2 = R_1 Q_1 = Q_1^T A_1 Q_1 \quad (3.4.6)$$

则矩阵 A_1 与 A_2 (酉) 相似. 于是, 矩阵 A_1 与 A_2 具有相同的特征值.

对矩阵 A_2 再作 QR 分解 $A_2 = Q_2 R_2$, 可定义矩阵 $A_3 = R_2 Q_2$.

重复上述过程, 可定义矩阵序列 $\{A_k, k=1, 2, \dots\}$. QR 算法的收敛性分析比较复杂, 这里不作详细介绍. 一个简单的结论是: 若非奇异矩阵的各特征值具有不同的模值, 则 QR 算法定义的矩阵序列收敛于上三角矩阵. 在有等模特征值 (包括重实特征值, 单对共轭复特征值或多对共轭复特征值等) 情况下, QR 算法的收敛性比较复杂, 可参考例 3.4.1.

例 3.4.1 利用 QR 算法求矩阵的特征值, 其中

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

解: (1) 令 $A = A_1$, 并对矩阵 A_1 进行 QR 分解: $A_1 = Q_1 R_1$, 其中,

$$Q_1 = \begin{bmatrix} -0.894\ 4 & 0.408\ 2 & 0.182\ 6 \\ -0.447\ 2 & -0.816\ 5 & -0.356\ 1 \\ 0 & -0.408\ 2 & 0.912\ 9 \end{bmatrix}$$

$$R_1 = \begin{bmatrix} -2.236\ 1 & -2.236\ 1 & -0.447\ 2 \\ 0 & -2.449\ 5 & -2.449\ 5 \\ 0 & 0 & 3.286\ 3 \end{bmatrix}$$

构造矩阵 $A_2 = R_1 Q_1$, 并对矩阵 A_2 进行 QR 分解: $A_2 = Q_2 R_2$, 其中,

$$A_2 = \begin{bmatrix} 3.000\ 0 & 1.095\ 4 & 0 \\ 1.095\ 4 & 3.000\ 0 & -1.341\ 6 \\ 0 & -1.341\ 6 & 3.000\ 0 \end{bmatrix}$$

$$Q_2 = \begin{bmatrix} -0.961\ 1 & 0.223\ 5 & -0.129\ 3 \\ -0.258\ 2 & -0.836\ 2 & 0.483\ 9 \\ 0 & 0.500\ 8 & 0.865\ 5 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} -2.898\ 3 & -1.512\ 3 & 0.412\ 4 \\ 0 & -3.189\ 1 & 2.408\ 8 \\ 0 & 0 & 1.081\ 9 \end{bmatrix}$$

同理, 可得

$$A_{10} = \begin{bmatrix} 4.728\ 5 & 0.078\ 1 & 0 \\ 0.078\ 1 & 3.003\ 5 & -0.002\ 0 \\ 0 & -0.002\ 0 & 1.268\ 0 \end{bmatrix}$$

$$\mathbf{A}_{20} = \begin{bmatrix} 4.732\ 1 & 0.000\ 8 & 0 \\ 0.000\ 8 & 3.000\ 0 & 0 \\ 0 & 0 & 1.267\ 9 \end{bmatrix}$$

读者可与矩阵 \mathbf{A} 的精确特征值作对比,其中的精确特征值分别为: $\lambda_1 = 3 + \sqrt{3} = 4.732\ 1, \lambda_2 = 3, \lambda_3 = 3 - \sqrt{3} = 1.267\ 9$.

(2) 矩阵 \mathbf{B} 的计算结果如下:

$$\mathbf{B}_2 = \begin{bmatrix} 2.333\ 3 & 0.356\ 3 & 5.455\ 4 \\ -0.712\ 7 & 1.238\ 1 & 0.583\ 2 \\ -0.218\ 2 & -0.233\ 3 & 0.428\ 6 \end{bmatrix}$$

$$\mathbf{B}_{10} = \begin{bmatrix} 2.049\ 6 & 1.046\ 4 & 3.778\ 1 \\ -0.022\ 6 & 1.521\ 9 & 3.978\ 4 \\ -0.015\ 1 & -0.319\ 1 & 0.428\ 6 \end{bmatrix}$$

$$\mathbf{B}_{20} = \begin{bmatrix} 1.999\ 2 & -0.818\ 7 & 3.655\ 0 \\ -0.002\ 2 & -0.332\ 6 & 3.723\ 3 \\ -0.000\ 7 & -0.745\ 4 & 2.333\ 3 \end{bmatrix}$$

$$\mathbf{B}_{100} = \begin{bmatrix} 2.000\ 0 & -0.816\ 5 & 3.651\ 5 \\ 0 & -0.333\ 3 & 3.726\ 8 \\ 0 & -0.745\ 4 & 2.333\ 3 \end{bmatrix}$$

从计算结果看,矩阵序列 $\{\mathbf{B}_k, k=1, 2, \dots\}$ 未收敛于上三角矩阵,计算其右下角 2 阶矩阵的特征值为 $1 \pm i$; 矩阵 \mathbf{B} 另一个特征值为 2. 矩阵 \mathbf{B} 的精确特征值为: $\lambda_1 = 2, \lambda_{2,3} = 1 \pm i$.

实际使用 QR 算法之前,一般会对矩阵进行处理,然后再进行 QR 算法,这样会大大节省计算量.

【思考】 能否利用 Givens 矩阵或 Householder 将满秩矩阵作 QR 分解,请读者尝试设计相应的分解方法?

3.5 Schur 分解

矩阵的 Schur 分解(Schur 引理)在理论分析中具有重要地位,是诸多重要定理证明的出发点. Schur 分解适用于所有的复方矩阵,相关结论如下.

定理 3.5.1(Schur 引理) 任意 n 阶复方矩阵 \mathbf{A} 相似于上三角矩阵 $\mathbf{\Lambda}$, 即存在可逆矩阵 \mathbf{P} 使得 $\mathbf{\Lambda} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ 为上三角矩阵,其中上三角矩阵 $\mathbf{\Lambda}$ 的对角元素是矩阵 \mathbf{A} 的特征值.

证明: 采用数学归纳法证明存在 m 个可逆矩阵 $\mathbf{P}_1, \dots, \mathbf{P}_m$ 使得

$$\mathbf{P}_m^{-1} \cdot \dots \cdot \mathbf{P}_1^{-1} \mathbf{A} \mathbf{P}_1 \cdot \dots \cdot \mathbf{P}_m = \begin{bmatrix} \mathbf{T}_m & \mathbf{C}_m \\ \mathbf{O} & \mathbf{A}_m \end{bmatrix}$$

式中, T_m 是以 $\lambda_1, \dots, \lambda_m$ 为对角线的上三角矩阵; A_m 是 $(n-m)$ 阶方阵, 其特征值为 $\lambda_{m+1}, \dots, \lambda_n; m=1, \dots, n-1$.

当 $m=1$ 时, 令 ζ_1 是属于特征值 λ_1 的单位特征向量. 利用基扩充定理将其扩充为 \mathbb{C}^n 空间的一组标准正交基 ζ_1, g_2, \dots, g_n , 并记 $P_1 = [\zeta_1, g_2, \dots, g_n]$. 于是有

$$P_1^{-1}AP_1 = \begin{bmatrix} \lambda_1 & C_1 \\ \mathbf{0} & A_1 \end{bmatrix}$$

由于矩阵 A 与 $P_1^{-1}AP_1$ 相似, 故矩阵 $A_1 \in \mathbb{C}^{(n-1) \times (n-1)}$ 的特征值为 $\lambda_2, \dots, \lambda_n$.

假设当 $m=k \in \{1, \dots, n-2\}$ 时存在可逆矩阵 P_1, \dots, P_k 使得

$$P_k^{-1} \cdot \dots \cdot P_1^{-1}AP_1 \cdot \dots \cdot P_k = \begin{bmatrix} T_k & C_k \\ \mathbf{0} & A_k \end{bmatrix}$$

当 $m=k+1$ 时, 仿照 P_1 构造方法知, 存在 $n-k$ 阶矩阵 V_{k+1} 使得

$$V_{k+1}^{-1}A_kV_{k+1} = \begin{bmatrix} \lambda_{k+1} & C_{k+1} \\ \mathbf{0} & A_{k+1} \end{bmatrix}$$

定义 $P_{k+1} = \begin{bmatrix} I_k & \mathbf{0} \\ \mathbf{0} & V_{k+1} \end{bmatrix}$, 则有

$$P_{k+1}^{-1}P_k^{-1} \cdot \dots \cdot P_1^{-1}AP_1 \cdot \dots \cdot P_kP_{k+1} = \begin{bmatrix} T_{k+1} & C_{k+1} \\ \mathbf{0} & A_{k+1} \end{bmatrix}$$

因此, 存在 $n-1$ 个可逆矩阵 P_1, \dots, P_{n-1} 使得

$$P_{n-1}^{-1} \cdot \dots \cdot P_1^{-1}AP_1 \cdot \dots \cdot P_{n-1} = \begin{bmatrix} T_{n-1} & C_{n-1} \\ \mathbf{0} & \lambda_n \end{bmatrix}$$

记 $P = P_1 \cdot \dots \cdot P_{n-1}$, 则上式右端为上三角矩阵. 证毕.

定理 3.5.2 (Schur 引理) 任意复方矩阵 A 酉相似于上三角矩阵 Λ , 即存在一酉矩阵 U 使得 $\Lambda = U^H A U$ 为上三角矩阵.

证明: 由定理 3.5.1 知, 存在可逆矩阵 P 使得

$$P^{-1}AP = \Lambda \quad (3.5.1)$$

式中: Λ 为上三角矩阵.

再对可逆矩阵 P 作 QR 分解, 得

$$P = UR \quad (3.5.2)$$

式中: U 是酉矩阵, R 为正线上三角矩阵.

将式(3.5.2)代入式(3.5.1), 得

$$R^{-1}U^H A U R = \Lambda \quad (3.5.3)$$

整理得到

$$U^H A U = R \Lambda R^{-1} \quad (3.5.4)$$

由于 R, R^{-1} 和 Λ 均为上三角矩阵, 故 $R \Lambda R^{-1}$ 仍是上三角矩阵. 证毕.

注 1: 与定理 3.5.1 相比, 定理 3.5.2 进一步说明了 Schur 分解中的满秩矩阵 P

可以是酉矩阵. 由于酉矩阵的逆矩阵是其共轭转置, 因此不需要复杂计算, 这为简化分析提供了重要的技术支持.

【思考】 是否存在正交矩阵 Q 使得实方阵 A 具有如下 Schur 分解:

$$Q^T A Q = \Lambda = \begin{bmatrix} \lambda_1 & & * \\ & \ddots & \\ \mathbf{O} & & \lambda_n \end{bmatrix}$$

分析: 考察矩阵 $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ 的 Schur 分解问题. 首先计算矩阵 A 的特征值 (因为上三角矩阵的对角线为 A 的特征值). 由特征多项式 $|\lambda I - A| = \lambda^2 + 1 = 0$ 得, $\lambda = \pm i$ 是矩阵 A 的两个特征值. 由此, A 必定为复矩阵而非实矩阵.

现假设存在正交矩阵 Q 使得 $Q^T A Q = \Lambda$, 则 $Q^T A Q$ 必为实矩阵, 而矩阵 A 是非实矩阵, 显然假设不成立, 即不一定存在正交矩阵 Q 使得 $Q^T A Q$ 为上三角矩阵.

出现这一问题的根源在于实矩阵 A 的特征值不一定全是实数, 可能有成对共轭复根. 因此, 在处理实矩阵特别是关系到矩阵特征值问题时, 不要輕易地将复矩阵中的结论直接引用.

定理 3.5.3 (实方阵 Schur 引理) 设 $A \in \mathbb{R}^{n \times n}$ 的特征值均为实数, 则存在正交矩阵 Q 使得

$$Q^T A Q = Q^{-1} A Q = \begin{bmatrix} \lambda_1 & & * \\ & \ddots & \\ \mathbf{O} & & \lambda_n \end{bmatrix}$$

下面利用 Schur 引理研究矩阵多项式.

定义 3.5.1 (矩阵多项式) 设 $A \in \mathbb{C}^{n \times n}$, $\varphi(\lambda) = a_n \lambda^n + a_{n-1} \lambda^{n-1} + \cdots + a_1 \lambda + a_0$, $a_i \in \mathbb{C} (i=0, 1, \cdots, n)$ 是数域 \mathbb{C} 上的多项式, 则

$$\varphi(A) = a_n A^n + a_{n-1} A^{n-1} + \cdots + a_1 A + a_0 I$$

称为矩阵多项式.

推论 3.5.1 设矩阵 $A \in \mathbb{C}^{n \times n}$ 的 n 个特征值为 $\lambda_1, \cdots, \lambda_n$, $\varphi(\lambda)$ 为任一多项式, 则矩阵多项式 $\varphi(A)$ 的 n 个特征值为 $\varphi(\lambda_1), \cdots, \varphi(\lambda_n)$.

证明: 根据定理 3.5.2 知, 存在一酉矩阵 U 使得 $U^H A U = \Lambda$, 其中 Λ 为上三角矩阵, 其对角线元素为矩阵 A 的特征值, 记为 $\lambda_1, \cdots, \lambda_n$. 注意到

$$\varphi(A) = \varphi(U \Lambda U^H) = U \varphi(\Lambda) U^H$$

式中: 上三角矩阵 $\varphi(\Lambda)$ 的对角线元素为 $\varphi(\lambda_1), \cdots, \varphi(\lambda_n)$. 由于 $\varphi(A)$ 酉相似于上三角矩阵 $\varphi(\Lambda)$, 故矩阵多项式 $\varphi(A)$ 的 n 个特征值为 $\varphi(\lambda_1), \cdots, \varphi(\lambda_n)$. 证毕.

注 2: n 阶矩阵 A 的属于特征值 λ_i 的特征向量 α_i 也是 $\varphi(A)$ 的属于特征值 $\varphi(\lambda_i)$ 的特征向量.

推论 3.5.2 设矩阵 $A \in \mathbb{C}^{n \times n}$ 的 n 个特征值为 $\lambda_1, \cdots, \lambda_n$, 则矩阵 A^m 的 n 个特征

值为 $\lambda_1^m, \dots, \lambda_n^m, m \in \mathbb{N}$.

【思考】 设 $f_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0$ 是矩阵 A 的特征多项式, 求矩阵多项式 $f_A(A)$.

分析: 以考察矩阵 $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ 为例. 矩阵 A 的特征多项式为 $f_A(\lambda) = \lambda^2 + 1$.

计算矩阵多项式 $f_A(A)$ 得, $f_A(A) = A^2 + I = O$.

实际上, 这一结论对一般的复方矩阵同样成立. 这就是著名的 Hamilton-Cayley 定理.

定理 3.5.4 (Hamilton-Cayley 定理) 设矩阵 $A \in \mathbb{C}^{n \times n}$ 的特征多项式为 $f_A(\lambda) = |\lambda I - A|$, 则 $f_A(A) = O$.

证明: 设矩阵 A 的特征多项式为 $f_A(\lambda) = (\lambda - \lambda_1) \cdot \dots \cdot (\lambda - \lambda_n)$ (注意方阵 A 的特征多项式的定义方式). 由定理 3.5.2 知, 存在酉矩阵 U 使得 $U^H A U = \Lambda$, 其中, Λ 为上三角矩阵, 其对角元素为 $\lambda_1, \dots, \lambda_n$. 进而, 矩阵多项式 $f_A(A)$ 可写成

$$f_A(A) = P f(\Lambda) P^{-1} = P (\Lambda - \lambda_1 I) \cdot \dots \cdot (\Lambda - \lambda_n I) P^{-1}$$

经计算知 $(\Lambda - \lambda_1 I) \cdot \dots \cdot (\Lambda - \lambda_n I) = O$, 故 $f_A(A) = O$.

注 3: 由于 $L(V) \cong \mathbb{C}^{n \times n}$, 故对线性变换 T 有平行的结果: $\forall T \in L(V)$ 且 $f_A(\lambda)$ 为 T 的特征多项式, 则 $f_A(T)$ 为零变换.

例 3.5.1 已知 $A = \begin{bmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{bmatrix}$, 计算矩阵多项式 $A^4 - 2A^2 + I$.

解: 矩阵 A 的特征多项式为

$$f_A(\lambda) = \begin{vmatrix} \lambda - 2 & 2 & -2 \\ 2 & \lambda + 1 & -4 \\ -2 & -4 & \lambda + 1 \end{vmatrix} = (\lambda - 1)^3$$

令 $g(\lambda) = \lambda^4 - 2\lambda^2 + 1$, 由多项式除法可得:

$$g(\lambda) = f_A(\lambda)(\lambda + 3) + 4(\lambda - 1)^2$$

由定理 3.5.4 知, $g(A) = 4(A - I)^2$. 将矩阵 A 的表达式代入上式得 $g(A) = O$.

【应用】 Hamilton-Cayley 定理的一个有趣且重要的应用是计算矩阵的逆矩阵.

推论 3.5.3 设复方阵 A 可逆, 其特征多项式为 $f_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0$, 则矩阵 A 的逆矩阵计算公式为

$$A^{-1} = -\frac{1}{a_0} (A^{n-1} + a_{n-1}A^{n-2} + \dots + a_1 I)$$

该结论只需将矩阵多项式方程 $f_A(A) = O$ 两端同时左乘(或右乘) A^{-1} 即可. 该推论表明任一 n 阶矩阵的逆矩阵(若存在)可由一个不超过 $n-1$ 次矩阵多项式表示.

例 3.5.2 求矩阵 $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ 的逆.

解: 矩阵 A 的特征多项式为 $f_A(\lambda) = \lambda^3 + 3\lambda^2 + 3\lambda + 1$. 由推论 3.5.3 知,

$$A^{-1} = A^2 - 3A + 3I = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

定义 3.5.2 (零化多项式) 给定复方阵 $A \in \mathbb{C}^{n \times n}$, 若存在多项式 $g(\lambda)$ 使得 $g(A) = O$, 则称 $g(\lambda)$ 为 A 的零化多项式.

定义 3.5.3 (最小多项式) 设复方阵 A 的零化多项式中最低次数的首 1 多项式称为 A 的最小多项式, 记为 $m_A(\lambda)$.

【思考】 特征多项式 $f_A(\lambda)$ 是矩阵 A 的零化多项式, 问 $f_A(\lambda)$ 是 A 的最小多项式吗?

分析: 以 $A = I_2$ 为例, 计算其最小多项式为 $m_A(\lambda) = \lambda - 1$, 而 $f_A(\lambda) = (\lambda - 1)^2$. 因此, 特征多项式并不一定是矩阵的最小多项式, 但两者有一定的关系, 详见如下定理.

定理 3.5.5 (最小多项式的性质) 设矩阵 $A \in \mathbb{C}^{n \times n}$, 则

(1) 矩阵 A 的最小多项式 $m_A(\lambda)$ 是唯一的, 且可整除 A 的任一零化多项式. 特别地, 有 $m_A(\lambda) \mid f_A(\lambda)$.

(2) 矩阵 A 的特征多项式 $f_A(\lambda)$ 与最小多项式 $m_A(\lambda)$ 具有相同的根 (不计重数).

证明: (1) 根据定理 2.1.3 知, 对任意多项式 $g(\lambda)$ 和 $h(\lambda)$, 必存在多项式 $q(\lambda)$ 以及多项式 $r(\lambda)$ 使得

$$g(\lambda) = h(\lambda)q(\lambda) + r(\lambda) \quad (3.5.5)$$

式中: $r(\lambda) = 0$ 或 $\deg[r(\lambda)] < \deg[h(\lambda)]$.

若 $g(\lambda)$ 和 $h(\lambda)$ 分别定义为矩阵的特征多项式 $f_A(\lambda)$ 和最小多项式 $m_A(\lambda)$, 则代入式 (3.5.5) 得 $r(A) = O$, 即 $m_A(\lambda) \mid f_A(\lambda)$.

再证唯一性. 假设 $m(\lambda)$ 也是 A 的最小多项式且 $m(\lambda) \neq m_A(\lambda)$, 则 $m_A(\lambda) \mid m(\lambda)$ 和 $m(\lambda) \mid m_A(\lambda)$ 均成立. 由于 $m(\lambda)$ 和 $m_A(\lambda)$ 次数相同且都是首 1 多项式, 故 $m(\lambda) = m_A(\lambda)$.

(2) 由 $m_A(\lambda) \mid f_A(\lambda)$ 知, 满足 $m_A(\lambda) = 0$ 的根 λ_i 一定是 $f_A(\lambda) = 0$ 的根. 故只需证明 $f_A(\lambda) = 0$ 的根是 $m_A(\lambda) = 0$ 的根即可.

设 λ_i 是矩阵 A 的特征值, x_i 是属于 λ_i 的特征向量. 根据推论 3.5.1 知, $m_A(\lambda_i)$ 是矩阵多项式 $m_A(A)$ 的特征值, 且 x_i 是属于 $m_A(A)$ 的特征向量, 即

$$m_A(A)x_i = m_A(\lambda_i)x_i$$

由于 $m_A(A)$ 是矩阵 A 的零化多项式, 故 $m_A(A) = O$, 进而有 $m_A(\lambda_i)x_i = 0$. 又知 x_i 是非零向量, 故 $m_A(\lambda_i) = 0$.

综上知, 矩阵 \mathbf{A} 的特征多项式 $f_{\mathbf{A}}(\lambda)$ 与最小多项式 $m_{\mathbf{A}}(\lambda)$ 有相同的根. 证毕.

例 3.5.3 求矩阵 $\mathbf{A} = \begin{bmatrix} 2 & -2 & 2 \\ -2 & -1 & 4 \\ 2 & 4 & -1 \end{bmatrix}$ 的最小多项式.

解: 矩阵 \mathbf{A} 的特征多项式为 $f_{\mathbf{A}}(\lambda) = (\lambda - 3)^2(\lambda + 6)$. 根据定理 3.5.5 知, 矩阵 \mathbf{A} 的最小多项式只能是 $(\lambda - 3)(\lambda + 6)$ 或 $(\lambda - 3)^2(\lambda + 6)$.

通过计算知, $(\mathbf{A} - 3\mathbf{I})(\mathbf{A} + 6\mathbf{I}) = \mathbf{O}$. 因此, $m_{\mathbf{A}}(\lambda) = (\lambda - 3)(\lambda + 6)$.

再次考察例 3.5.1, 由 \mathbf{A} 的特征多项式 $f_{\mathbf{A}}(\lambda) = (\lambda - 1)^3$ 知, \mathbf{A} 的最小多项式只能是 $(\lambda - 1)$, $(\lambda - 1)^2$ 或 $(\lambda - 1)^3$. 显然, $\mathbf{A} \neq \mathbf{I}$; 又知 $(\mathbf{A} - \mathbf{I})^2 = \mathbf{O}$, 则

$$m_{\mathbf{A}}(\lambda) = (\lambda - 1)^2$$

令 $g(\lambda) = \lambda^4 - 2\lambda^2 + 1$, 由多项式除法可得 $g(\lambda) = m_{\mathbf{A}}(\lambda)(\lambda - 1)^2$. 因此, $g(\mathbf{A}) = \mathbf{O}$. 与例 3.5.1 相比, 采用最小多项式计算矩阵多项式会更简单些.

同理, 利用最小多项式也可计算可逆矩阵的逆矩阵. 换言之, 推论 3.5.3 可改写为任一 n 阶可逆矩阵的逆矩阵可由一个不超过其最小多项式次数减 1 的矩阵多项式表示. 这一结论对矩阵的幂矩阵同样成立, 这里以例 3.5.4 进行说明.

例 3.5.4 设 $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 2 & 3 \end{bmatrix}$, 求 \mathbf{A}^{100} .

解: 由 $f_{\mathbf{A}}(\lambda) = (\lambda - 1)(\lambda - 4)$ 知, $\mathbf{A}^2 = 5\mathbf{A} - 4\mathbf{I}$. 利用该式进行迭代, 得

$$\mathbf{A}^3 = 5(5\mathbf{A} - 4\mathbf{I}) - 4\mathbf{A} = 21\mathbf{A} - 20\mathbf{I}$$

$$\mathbf{A}^4 = 21(5\mathbf{A} - 4\mathbf{I}) - 20\mathbf{A} = 85\mathbf{A} - 84\mathbf{I}$$

\vdots

由归纳得

$$\mathbf{A}^m = \left(\frac{4^m - 1}{3} \right) \mathbf{A} + \left(\frac{4 - 4^m}{3} \right) \mathbf{I}, \quad m = 0, 1, \dots$$

由此,

$$\mathbf{A}^{100} = \begin{bmatrix} \frac{2 + 4^{100}}{3} & \frac{-1 + 4^{100}}{3} \\ \frac{-2 + 2 \times 4^{100}}{3} & \frac{1 + 2 \times 4^{100}}{3} \end{bmatrix}$$

实际上, 无论是可逆矩阵的逆矩阵还是幂矩阵, 都可以看作矩阵函数的特例, 我们将在第 4 章的 4.7 节矩阵函数中详细展开.

3.6 对角化分解

定义 3.6.1 (单纯矩阵) 若 n 阶复方阵 \mathbf{A} 相似于对角矩阵, 则矩阵 \mathbf{A} 称为可对角化矩阵 (或单纯矩阵).

定义 3.6.1 隐含着单纯矩阵的对角化分解. 对角化分解实际上是 Schur 分解的一种特殊表达. 单纯矩阵的对角化分解是常见的矩阵分解, 其涉及的对角矩阵是一类特殊方阵: 对角矩阵的和、积、逆 (若存在) 仍是对角矩阵, 其对角元素就是它的特征值.

【思考】若线性变换 T 在某基下的矩阵为对角矩阵, 这里的某基如何确定?

下面给出复方阵为单纯矩阵的若干充分必要条件.

定理 3.6.1 设矩阵 $A \in \mathbb{C}^{n \times n}$ 的全部互异特征根为 $\lambda_1, \dots, \lambda_m, (m \leq n)$, 则以下表达等价:

- (1) A 是单纯矩阵;
- (2) A 有 n 个线性无关的特征向量;
- (3) 特征值 $\lambda_i (i=1, \dots, m)$ 的代数重数等于其几何重数;

$$(4) \sum_{i=1}^m \dim E(\lambda_i) = n;$$

- (5) 最小多项式 $m_A(\lambda)$ 无重根.

证明: 证明过程为 $(1) \Leftrightarrow (2); (1) \Rightarrow (3) \Rightarrow (4), (4) \Rightarrow (1); (1) \Rightarrow (5), (4) \Leftarrow (5).$

$(1) \Rightarrow (2)$. 假设存在可逆矩阵 P 使得:

$$P^{-1}AP = \text{diag}(\mu_1, \dots, \mu_n) \quad (3.6.1)$$

式中: μ_1, \dots, μ_n 是矩阵 A 的 n 个特征值.

将矩阵 P 进行列分块, 并记为 $P = [p_1, p_2, \dots, p_n]$. 式(3.6.1)可改写为

$$A[p_1, p_2, \dots, p_n] = [p_1, p_2, \dots, p_n] \text{diag}(\mu_1, \dots, \mu_n) \quad (3.6.2)$$

即 $Ap_i = \mu_i p_i (i=1, \dots, n)$, 这表明 p_i 是矩阵 A 的属于特征值 μ_i 的特征向量. 注意 P 是可逆矩阵, 故矩阵 A 有 n 个线性无关的特征向量.

$(1) \Leftarrow (2)$. 设 p_1, p_2, \dots, p_n 是矩阵 A 的 n 个线性无关向量, 且有 $Ap_i = \mu_i p_i (i=1, \dots, n)$. 定义 $P = [p_1, p_2, \dots, p_n]$, 则式(3.6.2)成立. 根据定义 3.6.1 知, A 是单纯矩阵.

$(1) \Rightarrow (3) \Rightarrow (4)$. 设矩阵 A 的特征多项式为

$$f_A(\lambda) = (\lambda - \lambda_1)^{d_1} (\lambda - \lambda_2)^{d_2} \cdots (\lambda - \lambda_m)^{d_m}$$

且存在可逆矩阵 P 使得

$$P^{-1}AP = A = \text{diag}(\underbrace{\lambda_1, \dots, \lambda_1}_{d_1}, \dots, \underbrace{\lambda_m, \dots, \lambda_m}_{d_m})$$

式中: $d_1 + d_2 + \dots + d_m = n$, d_i 是特征值 λ_i 的代数重数.

又知

$$\begin{aligned} \lambda_i I - A &= P(\lambda_i I - A)P^{-1} \\ &= P \text{diag}(\lambda_i - \lambda_1, \dots, \lambda_i - \lambda_{i-1}, \underbrace{0, \dots, 0}_{d_i}, \lambda_i - \lambda_{i+1}, \dots, \lambda_i - \lambda_m)P^{-1} \end{aligned}$$

从而

$$\operatorname{rank}(\lambda_i \mathbf{I} - \mathbf{A}) = n - d_i$$

这表明特征值 λ_i 的特征子空间 $E(\lambda_i)$ 的维数为

$$\dim E(\lambda_i) = n - \operatorname{rank}(\lambda_i \mathbf{I} - \mathbf{A}) = d_i, \quad i = 1, \cdots, m$$

因此, 特征值 λ_i 的代数重数等于其几何重数. 注意到代数重数之和为 n , 故几何重数之和也为 n .

(4) \Rightarrow (1). 由于 m 个特征子空间的维数之和为 n , 故可从这 m 个特征子空间找出 n 个线性无关的特征向量, 故 \mathbf{A} 是单纯矩阵.

(1) \Rightarrow (5). 考察矩阵多项式

$$g(\mathbf{A}) = (\mathbf{A} - \lambda_1 \mathbf{I}) \cdot \cdots \cdot (\mathbf{A} - \lambda_m \mathbf{I}) \quad (3.6.3)$$

由于 \mathbf{A} 是单纯矩阵, 故存在可逆矩阵 \mathbf{P} 使得 $\mathbf{P}^{-1} \mathbf{A} \mathbf{P} = \mathbf{\Lambda}$, 其中, $\mathbf{\Lambda}$ 是对角矩阵. 将其代入式 (3.6.3), 得

$$g(\mathbf{A}) = \mathbf{P}(\mathbf{\Lambda} - \lambda_1 \mathbf{I}) \cdot \cdots \cdot (\mathbf{\Lambda} - \lambda_m \mathbf{I}) \mathbf{P}^{-1} = \mathbf{O}$$

显然, $g(\lambda)$ 是矩阵 \mathbf{A} 的一个零化多项式. 根据定理 3.5.5 知, $g(\lambda)$ 必是矩阵 \mathbf{A} 的最小多项式, 故 \mathbf{A} 的最小多项式 $m_{\mathbf{A}}(\lambda)$ 无重根.

(4) \Leftarrow (5). 设矩阵 \mathbf{A} 的最小多项式为 $m_{\mathbf{A}}(\lambda) = (\lambda - \lambda_1) \cdot \cdots \cdot (\lambda - \lambda_m)$, 则

$$m_{\mathbf{A}}(\mathbf{A}) = (\mathbf{A} - \lambda_1 \mathbf{I}) \cdot \cdots \cdot (\mathbf{A} - \lambda_m \mathbf{I}) = \mathbf{O} \quad (3.6.4)$$

由等式 (3.6.4) 知 (见本章的习题 1)

$$\operatorname{rank}(\mathbf{A} - \lambda_1 \mathbf{I}) + \cdots + \operatorname{rank}(\mathbf{A} - \lambda_m \mathbf{I}) \leqslant (m-1)n \quad (3.6.5)$$

根据不等式 (3.6.5) 进一步得

$$\dim E(\lambda_1) + \cdots + \dim E(\lambda_m) \geqslant n \quad (3.6.6)$$

由于 m 个特征子空间的和空间是 \mathbb{C}^n 空间的线性子空间, 因此不等式 (3.6.6) 只能取等号, 即 $\sum_{i=1}^m \dim E(\lambda_i) = n$. 证毕.

例 3.6.1 设线性变换 $T \in \mathcal{L}(\mathbb{R}^3)$ 在 \mathbb{R}^3 空间的一组基 $\zeta_1, \zeta_2, \zeta_3$ 下的矩阵为 \mathbf{A} , 即 $T[\zeta_1, \zeta_2, \zeta_3] = [\zeta_1, \zeta_2, \zeta_3] \mathbf{A}$, 其中

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 0 & 0 \\ -2 & 0 & 4 \end{bmatrix}$$

问: (1) 线性变换 T 可否对角化; (2) 若 T 可对角化, 试求满秩矩阵 \mathbf{P} 使 $\mathbf{P}^{-1} \mathbf{A} \mathbf{P}$ 为对角矩阵.

解: (1) 方法一: 由矩阵 \mathbf{A} 的特征多项式 $f_{\mathbf{A}}(\lambda) = \lambda^2(\lambda - 5)$ 得, $\lambda_1 = \lambda_2 = 0, \lambda_3 = 5$. 又知 $\mathbf{A}(\mathbf{A} - 5\mathbf{I}) = \mathbf{O}$, 所以矩阵 \mathbf{A} 的最小多项式为 $m_{\mathbf{A}}(\lambda) = \lambda(\lambda - 5)$. 注意到 $m_{\mathbf{A}}(\lambda)$ 无重根, 故 \mathbf{A} 可对角化, 即线性变换 T 可对角化.

方法二: 矩阵 \mathbf{A} 的特征多项式为 $f(\lambda) = \lambda^2(\lambda - 5)$, 得到特征值 $\lambda_1 = \lambda_2 = 0, \lambda_3 = 5$. 由于

$$n - \operatorname{rank}(\mathbf{O} \mathbf{I} - \mathbf{A}) = 3 - 1 = 2$$

$$n - \text{rank}(5\mathbf{I} - \mathbf{A}) = 3 - 2 = 1$$

故 $\sum_{i=1}^3 \dim E(\lambda_i) = 3$, 故线性变换 T 可对角化.

(2) 对于 $\lambda_1 = \lambda_2 = 0$, 解方程 $(0\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ 得到两个线性无关的特征向量, 分别为 $\mathbf{x}_1 = [2, 0, 1]^T$, $\mathbf{x}_2 = [0, 1, 0]^T$;

对于 $\lambda_3 = 5$, 解方程 $(5\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ 得到一个特征向量为 $\mathbf{x}_3 = [1, 0, -2]^T$.

令 $\mathbf{P} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$, 故 \mathbf{P} 为可逆矩阵, 且满足 $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \text{diag}(0, 0, 5)$.

例 3.6.2 若矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 且 $\mathbf{A}^2 = \mathbf{A}$, 试判断 \mathbf{A} 是否可对角化.

解: 由 $\mathbf{A}^2 = \mathbf{A}$ 知, $g(\lambda) = \lambda^2 - \lambda$ 是 \mathbf{A} 的一个零化多项式 (注意 $g(\lambda)$ 不一定是 \mathbf{A} 的最小多项式). 由定理 3.5.5 知, \mathbf{A} 的最小多项式 $m_{\mathbf{A}}(\lambda)$ 只能是 $m_{\mathbf{A}}(\lambda) = \lambda$ 或 $m_{\mathbf{A}}(\lambda) = \lambda - 1$ 或 $m_{\mathbf{A}}(\lambda) = \lambda^2 - \lambda$. 无论 $m_{\mathbf{A}}(\lambda)$ 是哪一种表达式, 它均无重根. 因此, 矩阵 \mathbf{A} 是单纯矩阵.

推论 3.6.1 若复方阵 \mathbf{A} 的零化多项式 $g(\lambda)$ 无重根, 则矩阵 \mathbf{A} 是单纯矩阵.

推论 3.6.2 若 n 阶复方阵 \mathbf{A} 恰好有 n 个互异特征值, 则它必可对角化, 反之则不然.

注 1: 上述两个推论仅是复方阵 \mathbf{A} 为单纯矩阵的充分而非必要条件.

【应用】 单纯矩阵的对角化分解常用于矩阵的高次幂计算和线性常微分方程求解等问题. 这里通过两个例题分别说明.

例 3.6.3 设 $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 2 & 3 \end{bmatrix}$, 求 \mathbf{A}^{100} .

解: 矩阵 \mathbf{A} 的特征多项式 $f_{\mathbf{A}}(\lambda) = (\lambda - 1)(\lambda - 4)$, 则它的特征值为 $\lambda_1 = 1, \lambda_2 = 4$, 对应的特征向量分别为 $\mathbf{p}_1 = [1, -1]^T$, $\mathbf{p}_2 = [1, 2]^T$.

令 $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2] = \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}$, 有 $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \text{diag}(1, 4)$.

所以

$$\mathbf{A}^{100} = \mathbf{P} \begin{bmatrix} 1 & 0 \\ 0 & 4^{100} \end{bmatrix} \mathbf{P}^{-1} = \begin{bmatrix} \frac{2+4^{100}}{3} & \frac{-1+4^{100}}{3} \\ \frac{-2+2 \times 4^{100}}{3} & \frac{1+2 \times 4^{100}}{3} \end{bmatrix}$$

例 3.6.4 求解常系数线性常微分方程组

$$\begin{cases} \dot{x}_1 = x_1 - 3x_2 + 3x_3 \\ \dot{x}_2 = 3x_1 - 5x_2 + 3x_3 \\ \dot{x}_3 = 6x_1 - 6x_2 + 4x_3 \end{cases}$$

解: 记

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{bmatrix}$$

则原微分方程组可写成矩阵形式

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t)$$

计算矩阵 \mathbf{A} 的特征多项式为 $f_{\mathbf{A}}(\lambda) = (\lambda - 4)(\lambda + 2)^2$, 解得 $\lambda_1 = 4, \lambda_2 = \lambda_3 = -2$; 分别计算对应的特征向量得

$$\mathbf{p}_1 = [1, 1, 2]^T, \quad \mathbf{p}_2 = [1, 1, 0]^T, \quad \mathbf{p}_3 = [0, 1, 1]^T$$

因此, 矩阵 \mathbf{A} 是单纯矩阵.

$$\text{令 } \mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3] = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 2 & 0 & 1 \end{bmatrix}, \text{ 有}$$

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

定义 $\mathbf{y}(t) = \mathbf{P}^{-1}\mathbf{x}(t)$, 则常微分方程组 $\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t)$ 可等价变换为

$$\dot{\mathbf{y}}(t) = \begin{bmatrix} 4 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} \mathbf{y}(t)$$

定义 $\mathbf{y}(t) = [y_1(t), y_2(t), y_3(t)]^T$, 则上式可进一步写成

$$\begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

解得

$$\begin{cases} y_1 = c_1 e^{4t} \\ y_2 = c_2 e^{-2t} \\ y_3 = c_3 e^{-2t} \end{cases}$$

再由 $\mathbf{y}(t) = \mathbf{P}^{-1}\mathbf{x}(t)$ 解得

$$\begin{cases} x_1(t) = c_1 e^{4t} + c_2 e^{-2t} \\ x_2(t) = c_1 e^{4t} + (c_2 + c_3) e^{-2t} \\ x_3(t) = 2c_1 e^{4t} + c_3 e^{-2t} \end{cases}$$

式中: c_1, c_2 和 c_3 是任意常数.

上例通过定义线性变换 $\mathbf{y}(t) = \mathbf{P}^{-1}\mathbf{x}(t)$ 求解线性常微分方程. 实际上, 我们还可以利用其他的线性变换求解线性常微分方程, 最典型的莫过于拉普拉斯变换. 有兴趣的读者可尝试利用拉普拉斯变换求解上例.

【思考】 若定义 3.6.1 中的相似变换矩阵是酉矩阵, 则此时的单纯矩阵会表现出何种性质?

分析: 假设存在酉矩阵 \mathbf{U} 使得 $\mathbf{U}^H \mathbf{A} \mathbf{U} = \text{diag}(\lambda_1, \dots, \lambda_n)$, 并令 $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n]$,

则有

$$\mathbf{A}[\mathbf{u}_1, \cdots, \mathbf{u}_n] = [\mathbf{u}_1, \cdots, \mathbf{u}_n] \text{diag}(\lambda_1, \cdots, \lambda_n)$$

整理得 $\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i$ ($i=1, \cdots, n$), 即 \mathbf{u}_i 是属于特征值 λ_i 的特征向量. 注意到特征向量 \mathbf{u}_i 和 \mathbf{u}_j ($i \neq j$) 两两正交, 这表明矩阵 \mathbf{A} 有 n 个正交的特征向量.

进一步, 考察特征子空间 $E(\lambda_i)$ 和 $E(\lambda_j)$, 其中 $\lambda_i \neq \lambda_j$. 为讨论方便, 不妨设

$$E(\lambda_i) = \text{span}(\mathbf{u}_{i_1}, \cdots, \mathbf{u}_{i_{d_i}}), \quad E(\lambda_j) = \text{span}(\mathbf{u}_{j_1}, \cdots, \mathbf{u}_{j_{d_j}})$$

我们容易验证和空间 $E(\lambda_i) + E(\lambda_j)$ 是正交直和. 这表明属于不同特征值的特征向量必正交.

现考察一类特殊的矩阵: Hermite 矩阵(当然, 实对称矩阵也可看作 Hermite 矩阵的一种). 设复方阵 \mathbf{A} 是 Hermite 矩阵, 利用 Schur 引理可得

$$\mathbf{U}^H \mathbf{A} \mathbf{U} = \mathbf{\Lambda} \quad (3.6.7)$$

式中: \mathbf{U} 是酉矩阵, $\mathbf{\Lambda}$ 是上三角矩阵.

对式(3.6.7)左右两端进行共轭转置, 得

$$\mathbf{U}^H \mathbf{A} \mathbf{U} = \bar{\mathbf{\Lambda}} \quad (3.6.8)$$

由此, $\mathbf{\Lambda} = \bar{\mathbf{\Lambda}}$. 故 $\mathbf{\Lambda}$ 只能是对角矩阵, 且对角线元素等于其共轭. 因此, 有如下结论:

推论 3.6.3 设矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 则 \mathbf{A} 是 Hermite 矩阵当且仅当 \mathbf{A} 的所有特征值 $\lambda_1, \cdots, \lambda_n$ 为实数, 且存在酉矩阵 $\mathbf{U} \in \mathbb{C}^{n \times n}$ 使得 $\mathbf{U}^H \mathbf{A} \mathbf{U} = \text{diag}(\lambda_1, \cdots, \lambda_n)$.

推论 3.6.4 设矩阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$, 则 \mathbf{A} 是实对称矩阵当且仅当 \mathbf{A} 的所有特征值 $\lambda_1, \cdots, \lambda_n$ 为实数, 且存在正交矩阵 $\mathbf{Q} \in \mathbb{R}^{n \times n}$ 使得 $\mathbf{Q}^T \mathbf{A} \mathbf{Q} = \text{diag}(\lambda_1, \cdots, \lambda_n)$.

例 3.6.5 求正交矩阵 \mathbf{Q} 使得 $\mathbf{Q}^T \mathbf{B} \mathbf{Q}$ 为对角矩阵, 其中

$$\mathbf{B} = \begin{bmatrix} -1 & -3 & 3 & -3 \\ -3 & -1 & -3 & 3 \\ 3 & -3 & -1 & -3 \\ -3 & 3 & -3 & -1 \end{bmatrix}$$

解: 由 $|\lambda \mathbf{I} - \mathbf{B}| = (\lambda - 8)(\lambda + 4)^3$ 解得 $\lambda_1 = 8, \lambda_2 = \lambda_3 = \lambda_4 = -4$.

对于 $\lambda_1 = 8$, 解得属于它的特征向量为 $\boldsymbol{\alpha}_1 = [-1, 1, -1, 1]^T$.

对于 $\lambda_2 = \lambda_3 = \lambda_4 = -4$, 解得属于它的特征向量分别为

$$\boldsymbol{\alpha}_2 = [1, 1, 0, 0]^T, \boldsymbol{\alpha}_3 = [1, 0, -1, 0]^T, \boldsymbol{\alpha}_4 = [1, 0, 0, 1]^T$$

令 $V_1 = \text{span}(\boldsymbol{\alpha}_1)$, $V_2 = \text{span}(\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4)$, 则 $V_1 \perp V_2$. 故只需对 $\boldsymbol{\alpha}_1$ 作单位化处理, 对向量组 $\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4$ 作 Gram-Schmidt 正交化处理, 分别得到

$$\boldsymbol{\eta}_1 = \left[-\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, \frac{1}{2} \right]^T, \quad \boldsymbol{\eta}_2 = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0 \right]^T$$

$$\boldsymbol{\eta}_3 = \left[-\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, 0 \right]^T, \quad \boldsymbol{\eta}_4 = \left[\frac{1}{2\sqrt{3}}, -\frac{1}{2\sqrt{3}}, \frac{1}{2\sqrt{3}}, \frac{3}{2\sqrt{3}} \right]^T$$

令 $\mathbf{Q} = [\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \boldsymbol{\eta}_4]$, 则有 $\mathbf{Q}^T \mathbf{B} \mathbf{Q} = \text{diag}(8, -4, -4, -4)$.

注 2: 求 Hermite 矩阵 \mathbf{A} 酉相似于对角矩阵的步骤如下:

- (1) 求出 \mathbf{A} 的全部相异特征值及重数;
- (2) 对于每个特征值 λ , 求方程 $(\lambda \mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ 的一个基础解系, 并将其单位正交化处理;
- (3) 由标准正交特征向量生成酉矩阵 \mathbf{Q} , 则 $\mathbf{Q}^T \mathbf{A} \mathbf{Q}$ 是对角矩阵.

【思考】除了 Hermite 矩阵外, 还有哪些矩阵可以酉相似对角化?

例如, $\mathbf{A} = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}$, 其特征值为 $\lambda_{1,2} = \pm i$, 对应特征向量分别为

$$\mathbf{x}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}^T, \quad \mathbf{x}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}^T$$

定义 $\mathbf{U} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$, 有 $\mathbf{U}^H \mathbf{A} \mathbf{U} = \text{diag}(i, -i)$. 显然, 矩阵 \mathbf{A} 可酉相似对

角化.

定义 3.6.2 (正规矩阵) 设矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 若 $\mathbf{A}^H \mathbf{A} = \mathbf{A} \mathbf{A}^H$, 则称 \mathbf{A} 为正规矩阵 (或规范矩阵).

实对称矩阵、实反对称矩阵、Hermite 阵、反 Hermite 阵、正交矩阵、酉矩阵等都是正规矩阵. 再如, $\mathbf{B} = \mathbf{U}^H \text{diag}(3+5i, 2-i) \mathbf{U}$ 也是正规矩阵, 其中 \mathbf{U} 为任意二阶酉矩阵.

定理 3.6.2 复方阵 \mathbf{A} 是正规矩阵当且仅当 \mathbf{A} 酉相似于对角矩阵, 即 $\mathbf{A}^H \mathbf{A} = \mathbf{A} \mathbf{A}^H$ 当且仅当存在酉矩阵 \mathbf{U} 使得 $\mathbf{U}^H \mathbf{A} \mathbf{U} = \text{diag}(\lambda_1, \dots, \lambda_n)$.

证明: 必要性. 由 Schur 引理知, 对任意复方阵 \mathbf{A} , 存在酉矩阵 \mathbf{U} 使得

$$\mathbf{U}^H \mathbf{A} \mathbf{U} = \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & a_{12} & \cdots & a_{1n} \\ 0 & \lambda_2 & \cdots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

将上式代入等式 $\mathbf{A}^H \mathbf{A} = \mathbf{A} \mathbf{A}^H$, 得 $\mathbf{\Lambda}^H \mathbf{\Lambda} = \mathbf{\Lambda} \mathbf{\Lambda}^H$, 即

$$\begin{bmatrix} \bar{\lambda}_1 & 0 & \cdots & 0 \\ \bar{a}_{12} & \bar{\lambda}_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \bar{a}_{1n} & \bar{a}_{2n} & \cdots & \bar{\lambda}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & a_{12} & \cdots & a_{1n} \\ 0 & \lambda_2 & \cdots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 & a_{12} & \cdots & a_{1n} \\ 0 & \lambda_2 & \cdots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \begin{bmatrix} \bar{\lambda}_1 & 0 & \cdots & 0 \\ \bar{a}_{12} & \bar{\lambda}_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \bar{a}_{1n} & \bar{a}_{2n} & \cdots & \bar{\lambda}_n \end{bmatrix}$$

根据上式左右两端矩阵对角线元素相等,得

$$\begin{aligned} |\lambda_1|^2 &= |\lambda_1|^2 + |a_{12}|^2 + \cdots + |a_{1n}|^2 \\ |\lambda_2|^2 + |a_{12}|^2 &= |\lambda_2|^2 + |a_{23}|^2 + \cdots + |a_{2n}|^2 \\ &\vdots \\ |\lambda_n|^2 + |a_{1n}|^2 + |a_{2n}|^2 + \cdots + |a_{n-1,n}|^2 &= |\lambda_n|^2 \end{aligned}$$

依次求得 $a_{12} = \cdots = a_{1n} = 0, a_{23} = \cdots = a_{2n} = 0, \cdots, a_{n-1,n} = 0$. 由此,上三角矩阵 \mathbf{A} 退化为对角矩阵,即 \mathbf{A} 酉相似于对角矩阵 $\mathbf{\Lambda}$.

充分性. 已知存在酉矩阵 \mathbf{U} 使得矩阵 \mathbf{A} 酉相似于对角矩阵 $\mathbf{\Lambda}$, 即 $\mathbf{A} = \mathbf{U}^H \mathbf{\Lambda} \mathbf{U}$. 所以, $\mathbf{A}^H \mathbf{A} = \mathbf{U} \mathbf{\Lambda}^H \mathbf{\Lambda} \mathbf{U}^H, \mathbf{A} \mathbf{A}^H = \mathbf{U} \mathbf{\Lambda} \mathbf{\Lambda}^H \mathbf{U}^H$. 由于 $\mathbf{\Lambda}$ 是对角矩阵,故 $\mathbf{\Lambda}^H \mathbf{\Lambda} = \mathbf{\Lambda} \mathbf{\Lambda}^H$. 因此, $\mathbf{A}^H \mathbf{A} = \mathbf{A} \mathbf{A}^H$. 证毕.

推论 3.6.5 复方阵 \mathbf{A} 是正规矩阵当且仅当 \mathbf{A} 有 n 个特征向量构成 \mathbb{C}^n 空间的一组标准正交基,且属于 \mathbf{A} 的不同特征值的特征向量正交.

证明: 必要性. 若 \mathbf{A} 是正规矩阵,则根据定理 3.6.2 知,矩阵 \mathbf{A} 可酉相似对角化,即一定存在 n 个线性无关的特征向量构成 \mathbb{C}^n 的一组标准正交基. 显然,属于不同特征值的特征子空间正交,即矩阵 \mathbf{A} 的不同特征值的特征向量正交.

充分性. 矩阵 \mathbf{A} 有 n 个标准正交特征向量也就是 \mathbf{A} 可酉对角化. 根据定理 3.6.2, \mathbf{A} 是正规矩阵,由此得证.

例 3.6.6 设 $\mathbf{A} = \begin{bmatrix} 0 & i & -1 \\ -i & 0 & i \\ -1 & -i & 0 \end{bmatrix}$, 验证 \mathbf{A} 是正规矩阵,并求酉矩阵 \mathbf{U} 使得

$\mathbf{U}^H \mathbf{A} \mathbf{U}$ 为对角矩阵.

解: 由于

$$\mathbf{A} \mathbf{A}^H = \begin{bmatrix} 2 & i & -1 \\ -i & 2 & i \\ -1 & -i & 2 \end{bmatrix} = \mathbf{A}^H \mathbf{A}$$

故 \mathbf{A} 是正规矩阵.

由矩阵 \mathbf{A} 的特征多项式 $f_A(\lambda) = (\lambda - 2)(\lambda + 1)^2$ 知, \mathbf{A} 的特征值为 $\lambda_1 = 2, \lambda_2 = \lambda_3 = -1$.

对于 $\lambda_1 = 2$, 计算其特征向量为 $\mathbf{x}_1 = [1, -i, -1]^T$;

对于 $\lambda_2 = \lambda_3 = -1$, 计算其特征向量为 $\mathbf{x}_2 = [1, 0, 1]^T, \mathbf{x}_3 = [1, i, 0]^T$;

对 \mathbf{x}_1 作单位化处理, 对 $\mathbf{x}_2, \mathbf{x}_3$ 作 Gram-Schmit 正交化处理, 得

$$\boldsymbol{\alpha}_1 = [1, -i, -1]^T, \quad \boldsymbol{\alpha}_2 = \left[\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}} \right]^T, \quad \boldsymbol{\alpha}_3 = \left[-\frac{1}{\sqrt{2}}i, -\sqrt{2}, \frac{1}{\sqrt{2}}i \right]^T$$

令 $\mathbf{U} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3]$, 则 $\mathbf{U}^H \mathbf{A} \mathbf{U} = \text{diag}(2, -1, -1)$.

推论 3.6.6 实方阵 \mathbf{A} 是正交矩阵当且仅当 \mathbf{A} 的所有特征值的模值为 1, 且存在酉矩阵 \mathbf{U} 使得 $\mathbf{U}^H \mathbf{A} \mathbf{U} = \text{diag}(\lambda_1, \dots, \lambda_n)$, 其中 $\lambda_1, \dots, \lambda_n$ 是 \mathbf{A} 的 n 个特征值.

证明: 必要性. 设 λ 是矩阵 \mathbf{A} 的特征值, \mathbf{x} 是属于特征值 λ 的特征向量, 则有 $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. 进一步,

$$(\mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{x}) = \mathbf{x}^H \mathbf{A}^H \mathbf{A} \mathbf{x} = |\lambda|^2 \mathbf{x}^H \mathbf{x}$$

又知 \mathbf{A} 是正交矩阵, 故 $\mathbf{A}^H \mathbf{A} = \mathbf{I}$. 上式可进一步改写为

$$\mathbf{x}^H \mathbf{x} = |\lambda|^2 \mathbf{x}^H \mathbf{x}$$

注意到特征向量 \mathbf{x} 为非零向量, 故 $\mathbf{x}^H \mathbf{x} \neq 0$ 因此, $|\lambda|^2 = 1$, 即 $|\lambda| = 1$. 由于正交矩阵必为正规矩阵, 故矩阵 \mathbf{A} 可酉相似对角化.

充分性. 由 $\mathbf{A} = \mathbf{U}^H \text{diag}(\lambda_1, \dots, \lambda_n) \mathbf{U}$, 得

$$\mathbf{A} \mathbf{A}^H = \mathbf{U}^H \begin{bmatrix} \lambda_1 & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & \lambda_n \end{bmatrix} \begin{bmatrix} \bar{\lambda}_1 & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & \bar{\lambda}_n \end{bmatrix} \mathbf{U} = \mathbf{I}$$

因此, 矩阵 \mathbf{A} 是正交矩阵.

注 3: 正交矩阵酉相似对角化时的变换矩阵是酉矩阵, 不一定是正交矩阵.

推论 3.6.7 设矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 则 \mathbf{A} 是酉矩阵当且仅当 \mathbf{A} 的所有特征值的模值为 1, 且存在酉矩阵 \mathbf{U} 使得 $\mathbf{U}^H \mathbf{A} \mathbf{U} = \text{diag}(\lambda_1, \dots, \lambda_n)$, 其中 $\lambda_1, \dots, \lambda_n$ 是 \mathbf{A} 的 n 个特征值.

请读者注意推论 3.6.6 和推论 3.6.7 的区别.

3.7 谱分解

由定理 3.6.2 可知, 复方阵 \mathbf{A} 是正规矩阵当且仅当 \mathbf{A} 酉相似于对角矩阵, 即存在酉矩阵 \mathbf{U} 使得

$$\mathbf{U}^H \mathbf{A} \mathbf{U} = \text{diag}(\underbrace{\lambda_1, \dots, \lambda_1}_{d_1}, \dots, \underbrace{\lambda_m, \dots, \lambda_m}_{d_m}) \quad (3.7.1)$$

式中: $\lambda_1, \dots, \lambda_m$ 是矩阵 \mathbf{A} 的 m 个互异特征值, 其代数重数分别为 d_1, \dots, d_m 且有 $d_1 + \dots + d_m = n$.

为进一步研究分解式 (3.7.1), 定义 $\mathbf{u}_{j1}, \dots, \mathbf{u}_{jd_j}$ 是属于特征值 λ_j 的 d_j 个单位正交的特征向量, $j = 1, \dots, m$. 则式 (3.7.1) 中的酉矩阵可定义为

$$\mathbf{U} = [\mathbf{u}_{11}, \dots, \mathbf{u}_{1d_1}, \dots, \mathbf{u}_{m1}, \dots, \mathbf{u}_{md_m}] \in \mathbb{C}^{n \times n}$$

于是,式(3.7.1)可改写为

$$\begin{aligned}
 \mathbf{A} &= [\mathbf{u}_{11}, \cdots, \mathbf{u}_{md_m}] \begin{bmatrix} \lambda_1 & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & \lambda_m \end{bmatrix} \begin{bmatrix} \mathbf{u}_{11}^H \\ \vdots \\ \mathbf{u}_{md_m}^H \end{bmatrix} \\
 &= \lambda_1 \mathbf{u}_{11} \mathbf{u}_{11}^H + \cdots + \lambda_m \mathbf{u}_{md_m} \mathbf{u}_{md_m}^H \\
 &= \lambda_1 \left(\sum_{i=1}^{d_1} \mathbf{u}_{1i} \mathbf{u}_{1i}^H \right) + \cdots + \lambda_m \left(\sum_{i=1}^{d_m} \mathbf{u}_{mi} \mathbf{u}_{mi}^H \right) \quad (3.7.2)
 \end{aligned}$$

定义 $\mathbf{E}_j = \sum_{i=1}^{d_j} \mathbf{u}_{ji} \mathbf{u}_{ji}^H, j=1, \cdots, n$, 有

$$\mathbf{A} = \lambda_1 \mathbf{E}_1 + \cdots + \lambda_m \mathbf{E}_m \quad (3.7.3)$$

分解式(3.7.3)是酉矩阵 Schur 分解式(3.7.1)的另一等价形式,我们常称式(3.7.3)为酉矩阵 \mathbf{A} 的谱分解式,并作如下定义.

定义 3.7.1(正规矩阵谱分解) 设 $\lambda_1, \cdots, \lambda_m$ 是正规矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 的 m 个互异特征值,其代数重数分别为 d_1, \cdots, d_m 且 $d_1 + \cdots + d_m = n$. 矩阵 \mathbf{A} 的谱分解式为

$$\mathbf{A} = \sum_{j=1}^m \lambda_j \mathbf{E}_j$$

式中: $\mathbf{E}_j = \sum_{i=1}^{d_j} \mathbf{u}_{ji} \mathbf{u}_{ji}^H, j=1, \cdots, m$, 称为矩阵 \mathbf{A} 的谱阵, $\mathbf{u}_{j1}, \cdots, \mathbf{u}_{jd_j}$ 是属于特征值 λ_j 的 d_j 个单位正交的特征向量.

例 3.7.1 求正规矩阵 $\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & -1 & 1 \\ 1 & -1 & 0 & 1 \\ -1 & 1 & 1 & 0 \end{bmatrix}$ 的谱分解式.

解: 计算矩阵 \mathbf{A} 的特征值与特征向量, 分别为: $\lambda_1 = \lambda_2 = \lambda_3 = 1, \boldsymbol{\beta}_1 = [1, 1, 0, 0]^T, \boldsymbol{\beta}_2 = [1, 0, 1, 0]^T, \boldsymbol{\beta}_3 = [-1, 0, 0, 1]^T; \lambda_4 = -3, \boldsymbol{\beta}_4 = [1, -1, -1, 1]^T$.

将 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ 和 $\boldsymbol{\beta}_3$ 单位正交化, 并将 $\boldsymbol{\beta}_4$ 单位化, 得

$$\begin{aligned}
 \boldsymbol{\alpha}_1 &= \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0 \right]^T \\
 \boldsymbol{\alpha}_2 &= \left[\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, 0 \right]^T \\
 \boldsymbol{\alpha}_3 &= \left[-\frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{3}{\sqrt{12}} \right]^T \\
 \boldsymbol{\alpha}_4 &= \left[\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2} \right]^T
 \end{aligned}$$

定义 $\mathbf{E}_1 = \boldsymbol{\alpha}_1 \boldsymbol{\alpha}_1^H + \boldsymbol{\alpha}_2 \boldsymbol{\alpha}_2^H + \boldsymbol{\alpha}_3 \boldsymbol{\alpha}_3^H, \mathbf{E}_2 = \boldsymbol{\alpha}_4 \boldsymbol{\alpha}_4^H$, 则 $\mathbf{A} = \mathbf{E}_1 - 3\mathbf{E}_2$ 是 \mathbf{A} 的谱分解式,

其中

$$\mathbf{E}_1 = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} & -\frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & \frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{3}{4} \end{bmatrix}, \quad \mathbf{E}_2 = \begin{bmatrix} \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

定理 3.7.1 (正规矩阵谱阵的性质) 设正规矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 有谱分解式 $\mathbf{A} =$

$\sum_{j=1}^m \lambda_j \mathbf{E}_j$, 其中, $\lambda_1, \dots, \lambda_m$ 是 \mathbf{A} 的 m 个互异特征值, $\mathbf{E}_1, \dots, \mathbf{E}_m$ 是 \mathbf{A} 的 m 个谱阵, 则

对任意 $i, j = 1, \dots, m$ 且 $i \neq j$, 有性质

(1) $\mathbf{E}_j = \mathbf{E}_j^H = (\mathbf{E}_j)^2$;

(2) $\mathbf{E}_i \mathbf{E}_j = \mathbf{O}$;

(3) $\mathbf{E}_i \mathbf{A} = \mathbf{A} \mathbf{E}_i = \lambda_i \mathbf{E}_i$;

(4) $\sum_{k=1}^m \mathbf{E}_k = \mathbf{I}$;

(5) 谱阵集合 $\{\mathbf{E}_1, \dots, \mathbf{E}_m\}$ 唯一.

证明: 由正规矩阵的谱分解知, \mathbf{A} 的谱阵为

$$\mathbf{E}_j = \sum_{i=1}^{d_j} \mathbf{u}_{ji} \mathbf{u}_{ji}^H, \quad j = 1, \dots, m \quad (3.7.4)$$

式中: $\mathbf{u}_{j1}, \dots, \mathbf{u}_{jd_j}$ 是属于特征值 λ_j 的 d_j 个单位正交的特征向量.

性质(1)、(2). 由式(3.7.4)容易证明对任意 $i, j = 1, \dots, m$ 且 $i \neq j$

$$\begin{aligned} \mathbf{E}_j &= \mathbf{E}_j^H = (\mathbf{E}_j)^2 \\ \mathbf{E}_i \mathbf{E}_j &= \mathbf{O} \end{aligned}$$

性质(3). 又知

$$\begin{aligned} \mathbf{E}_i \mathbf{A} &= \mathbf{E}_i \left(\sum_{j=1}^m \lambda_j \mathbf{E}_j \right) = \lambda_i (\mathbf{E}_j)^2 = \lambda_i \mathbf{E}_i \\ \mathbf{A} \mathbf{E}_i &= \left(\sum_{j=1}^m \lambda_j \mathbf{E}_j \right) \mathbf{E}_i = \lambda_i (\mathbf{E}_j)^2 = \lambda_i \mathbf{E}_i \end{aligned}$$

故 $\mathbf{E}_i \mathbf{A} = \mathbf{A} \mathbf{E}_i = \lambda_i \mathbf{E}_i$.

性质(4). 令 $\mathbf{U} = [\mathbf{u}_{11}, \dots, \mathbf{u}_{1d_1}, \dots, \mathbf{u}_{m1}, \dots, \mathbf{u}_{md_m}] \in \mathbb{C}^{n \times n}$, 并由 $\mathbf{U} \mathbf{U}^H = \mathbf{I}$ 得

$$\sum_{j=1}^m \sum_{i=1}^{d_j} \mathbf{u}_{ji} \mathbf{u}_{ji}^H = \mathbf{I}$$

即 $\sum_{j=1}^m \mathbf{E}_j = \mathbf{I}$.

采用反证法证明性质(5). 假设存在两个不同的谱阵集合满足谱分解式, 两谱阵集合分别记为 $\{E_1, \dots, E_m\}$ 和 $\{G_1, \dots, G_m\}$, 故 $A = \sum_{i=1}^m \lambda_i E_i = \sum_{i=1}^m \lambda_i G_i$.

由性质(3)知, $AE_i = \lambda_i E_i$ 和 $AG_j = \lambda_j G_j$. 分别利用这两个等式得

$$AE_i G_j = \lambda_i E_i G_j$$

$$AE_i G_j = E_i AG_j = \lambda_j E_i G_j$$

即 $\lambda_i E_i G_j = \lambda_j E_i G_j$. 特别地, 当 $i \neq j$ 时, $\lambda_i \neq \lambda_j$. 故 $E_i G_j = O, i=1, \dots, m$.

又知, 对任意 $i=1, \dots, m$,

$$E_i = E_i I = E_i \left(\sum_{j=1}^m G_j \right) = E_i G_i = \left(\sum_{j=1}^m E_j \right) G_i = G_i$$

因此, 正规矩阵 A 的谱阵集合 $\{E_1, \dots, E_m\}$ 唯一.

定理 3.7.2 设 n 阶复方矩阵 A 有 m 个互异特征值 $\lambda_1, \dots, \lambda_m$, 则 A 为正规矩阵当且仅当存在 m 个 n 阶矩阵 E_1, \dots, E_m 使得对任意 $i, j=1, \dots, m$ 且 $i \neq j$, 有性质

$$(1) A = \sum_{k=1}^m \lambda_k E_k;$$

$$(2) E_j = E_j^H = (E_j)^2;$$

$$(3) E_i E_j = O;$$

$$(4) E_i A = A E_i = \lambda_i E_i;$$

$$(5) \sum_{k=1}^m E_k = I;$$

$$(6) \text{谱阵集合 } \{E_1, \dots, E_m\} \text{ 唯一.}$$

从上述分解看, 谱分解中的谱阵是一类特殊矩阵. 为此, 引入幂等矩阵定义.

定义 3.7.2 (幂等矩阵) 设 $E \in \mathbb{C}^{n \times n}$, 若 $E^2 = E$, 则称 E 为幂等矩阵 (或投影矩阵). Hermite 幂等矩阵称为正交投影矩阵.

定理 3.7.3 (幂等矩阵性质) 若 $E \in \mathbb{C}_r^{n \times n}$ 是幂等矩阵, 则

$$(1) E \text{ 为单纯矩阵且 Jordan 标准形为 } \begin{bmatrix} I_r & O \\ O & O \end{bmatrix};$$

$$(2) \operatorname{tr}(E) = r;$$

$$(3) Ex = x \Leftrightarrow x \in R(E), \text{ 其中 } x \in \mathbb{C}^n.$$

证明: 性质(1). 由 $E^2 = E$ 可知, $\varphi(\lambda) = \lambda(\lambda - 1)$ 是矩阵 E 的零化多项式, 故 E 为单纯矩阵, 且它的特征值只能为 0 或 1. 又知 $\operatorname{rank}(E) = r$, 故矩阵 E 的特征值 1 有 r 重根. 综上知, E 的 Jordan 标准形为 $\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$.

性质(2). 由性质(1)知, 存在可逆矩阵 P 使得

$$P^{-1}AP = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$$

又知 $\text{tr}(\mathbf{E}) = \text{tr}(\mathbf{P}^{-1}\mathbf{A}\mathbf{P})$, 故 $\text{tr}(\mathbf{E}) = r$.

性质(3). 由 $\mathbf{E}\mathbf{x} = \mathbf{x}$ 知, $\mathbf{x} \in R(\mathbf{E})$; 反之, 若 $\mathbf{x} \in R(\mathbf{E})$, 则存在向量 $\mathbf{y} \in \mathbb{C}^n$ 满足 $\mathbf{x} = \mathbf{E}\mathbf{y}$, 故 $\mathbf{E}\mathbf{x} = \mathbf{E}(\mathbf{E}\mathbf{y}) = \mathbf{E}^2\mathbf{y} = \mathbf{E}\mathbf{y} = \mathbf{x}$.

注 1: 定理 3.7.3 性质(3)的几何解释为: 向量 $\mathbf{x} \in R(\mathbf{E})$ 当且仅当向量 \mathbf{x} 在线性空间 $R(\mathbf{E})$ 的投影恰为它本身.

例 3.7.2 求向量 $\mathbf{b} \in \mathbb{C}^n$ 在 $V_m = \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_m)$ 上的正交投影, 其中向量组 $\mathbf{a}_1, \dots, \mathbf{a}_m$ 是 \mathbb{C}^n 空间的 m 个线性无关向量, $m \leq n$.

解: 定义矩阵 $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_m] \in \mathbb{C}^{n \times m}$, 则矩阵 \mathbf{A} 列满秩. 由此, $\mathbf{A}^H\mathbf{A}$ 为可逆矩阵, 并可定义矩阵 $\mathbf{P} = \mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H$. 显然, $\mathbf{P}^H = \mathbf{P}$ 且

$$\mathbf{P}^2 = \mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H = \mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H = \mathbf{P}$$

因此, \mathbf{P} 是正交投影矩阵.

考察向量 $\mathbf{P}\mathbf{b}$ 和空间 V_m , 容易证得以下结论:

- (1) $R(\mathbf{P}) = V_m$;
- (2) $N(\mathbf{P}) = R(\mathbf{P})^\perp$;
- (3) $\mathbf{P}\mathbf{b} \in R(\mathbf{P})$;
- (4) $\mathbf{b} = \mathbf{P}\mathbf{b} + (\mathbf{I} - \mathbf{P})\mathbf{b}$;
- (5) $(\mathbf{I} - \mathbf{P})\mathbf{b} \in N(\mathbf{P})$.

由此, 我们可推断出 $\mathbf{P}\mathbf{b}$ 是向量 \mathbf{b} 在 V_m 上的投影.

【思考】若例 3.7.2 中的向量组 $\mathbf{a}_1, \dots, \mathbf{a}_m$ 线性相关, 应如何求解?

注 2: 求取向量的(正交)投影可先找到(正交)投影矩阵, 然后用(正交)投影矩阵乘以向量即可得到该向量的(正交)投影. 倘若(正交)投影矩阵不易获得, 可优先考虑向量在其(正交)补空间的(正交)投影矩阵, 然后再利用(正交)投影求出向量的(正交)投影.

例 3.7.3 求向量 $\mathbf{b} \in \mathbb{C}^n$ 在 $W = \{\mathbf{x} \in \mathbb{C}^n \mid \mathbf{A}\mathbf{x} = \mathbf{0}, \mathbf{A} \in \mathbb{C}^{m \times n}\}$ 上的正交投影.

解: 已知向量 \mathbf{b} 在 $R(\mathbf{A}^H)$ 的正交投影(利用例 3.7.2 结果)为

$$\text{Proj}_{R(\mathbf{A}^H)} \mathbf{b} = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^{-1}\mathbf{A}\mathbf{b}$$

又知 $N(\mathbf{A}) \oplus R(\mathbf{A}^H) = \mathbb{C}^n$, 则向量 $\mathbf{b} \in \mathbb{C}^n$ 在 $N(\mathbf{A})$ 的正交投影为

$$\text{Proj}_{N(\mathbf{A})} \mathbf{b} = \mathbf{b} - \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^{-1}\mathbf{A}\mathbf{b}$$

【应用】诸多工程应用的最优问题求解都可以归结为提取某个希望的信号, 同时抑制其他所有干扰或噪声. 投影是解决这类问题的重要数学工具. 下面以最小二乘问题为例进行简单说明.

例 3.7.4(最小二乘问题) 考察线性方程组 $\mathbf{A}\mathbf{x} = \mathbf{b}$, 其中, $\mathbf{A} \in \mathbb{C}_n^{m \times n}$ 和 $\mathbf{b} \in \mathbb{C}^m$ 给定, $\mathbf{x} \in \mathbb{C}^n$ 待定. 求向量 \mathbf{x} 使得 $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$ 最小.

解: 设向量 \mathbf{b} 在 $R(\mathbf{A})$ 空间的正交投影为 $\text{Proj}_{R(\mathbf{A})} \mathbf{b}$, 则由图 3.7.1 所示, 向量 \mathbf{b} 可分解为 $\mathbf{b} = \text{Proj}_{R(\mathbf{A})} \mathbf{b} + (\mathbf{b} - \text{Proj}_{R(\mathbf{A})} \mathbf{b})$, 且有

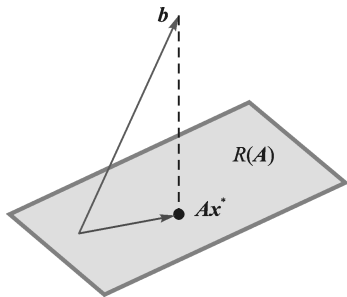


图 3.7.1 最小二乘问题示意图

$$\|Ax - b\| = \|(Ax - \text{Proj}_{R(A)} b) - (b - \text{Proj}_{R(A)} b)\| \quad (3.7.5)$$

式中: $b - \text{Proj}_{R(A)} b$ 为定向量.

对式(3.7.5)两边平方,得

$$\begin{aligned} \|Ax - b\|^2 &= \|Ax - \text{Proj}_{R(A)} b\|^2 + \|b - \text{Proj}_{R(A)} b\|^2 + \\ &\quad [Ax - \text{Proj}_{R(A)} b]^H [b - \text{Proj}_{R(A)} b] + \\ &\quad [b - \text{Proj}_{R(A)} b]^H [Ax - \text{Proj}_{R(A)} b] \end{aligned} \quad (3.7.6)$$

注意到向量 $[Ax - \text{Proj}_{R(A)} b]$ 与向量 $[b - \text{Proj}_{R(A)} b]$ 正交(请代入 $\text{Proj}_{R(A)} b$ 表达式自行证明), 式(3.7.6)可改写为

$$\|Ax - b\|^2 = \|Ax - \text{Proj}_{R(A)} b\|^2 + \|b - \text{Proj}_{R(A)} b\|^2 \geq \|b - \text{Proj}_{R(A)} b\|^2$$

上式等号成立的充分必要条件为 $Ax = \text{Proj}_{R(A)} b$.

又知 $\text{Proj}_{R(A)} b = A(A^H A)^{-1} A^H b$. 因此, 求 $\|Ax - b\|$ 最小可等价转化为求解线性方程组

$$Ax = A(A^H A)^{-1} A^H b. \quad (3.7.7)$$

方程组(3.7.7)左右两端同时乘以 A^H , 得

$$A^H A x = A^H b \quad (3.7.8)$$

由于矩阵 A 列满秩, 方程组(3.7.8)有唯一解 $x = (A^H A)^{-1} A^H b$. 因此, 方程组(3.7.7)的解为

$$x = (A^H A)^{-1} A^H b$$

仿照西矩阵的谱分解方法, 一般的单纯矩阵也有类似的结果.

定义 3.7.3(单纯矩阵谱分解) 设 $\lambda_1, \dots, \lambda_m$ 是单纯矩阵 $A \in \mathbb{C}^{n \times n}$ 的 m 个互异特征值, 其代数重数分别为 d_1, \dots, d_m , 则矩阵 A 的谱分解式定义为

$$A = \sum_{j=1}^m \lambda_j E_j$$

式中: $E_j = \sum_{i=1}^{d_j} \alpha_{ji} \beta_{ji}^H, j=1, \dots, m$, 称为 A 的谱阵, $\alpha_{j1}, \dots, \alpha_{jd_j}$ 是属于特征值 λ_j 的 d_j 个线性无关的单位特征向量, 行向量 $\beta_{jk}^H, k=1, \dots, d_j, j=1, \dots, m$, 是矩阵 $[\alpha_{11}, \dots,$

$\alpha_{1d_1}, \dots, \alpha_{m1}, \dots, \alpha_{md_m}]^{-1}$ 的第 $(\sum_{i=1}^{j-1} d_i + k)$ 行 (令 $d_0 = 0$).

注 3: 在定义 3.7.3 中, 所有列向量 α_{ji} 构成的 n 阶方阵恰好构成可逆变换矩阵 P , 而由所有行向量 β_{ji}^H 构成的 n 阶方阵恰好是 P^{-1} .

例 3.7.5 求矩阵 $A = \begin{bmatrix} 4 & 6 & 0 \\ -3 & -5 & 0 \\ -3 & -6 & 1 \end{bmatrix}$ 的谱分解.

解: 计算矩阵 A 的特征值分别为 $\lambda_1 = \lambda_2 = 1, \lambda_3 = -2$; 特征向量分别为

$$\alpha_1 = [2, -1, 0]^T, \quad \alpha_2 = [0, 0, 1]^T, \quad \alpha_3 = [-1, 1, 1]^T$$

于是

$$P = \begin{bmatrix} 2 & 0 & -1 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad P^{-1} = \begin{bmatrix} 1 & 1 & 0 \\ -1 & -2 & 1 \\ 1 & 2 & 0 \end{bmatrix}$$

根据定义 3.7.3 可知, $\beta_1 = [1, 1, 0]^T, \beta_2 = [-1, -2, 1]^T, \beta_3 = [1, 2, 0]^T$.

定义

$$E_1 = \alpha_1 \beta_1^H + \alpha_2 \beta_2^H = \begin{bmatrix} 2 & 2 & 0 \\ -1 & -1 & 0 \\ -1 & -2 & 1 \end{bmatrix}$$

$$E_2 = \alpha_3 \beta_3^H = \begin{bmatrix} -1 & -2 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 0 \end{bmatrix}$$

因此, $A = E_1 - 2E_2$ 是矩阵 A 的谱分解.

定理 3.7.4 设 n 阶复方阵 A 有 m 个互异特征值 $\lambda_1, \dots, \lambda_m$, 则 A 是单纯矩阵当且仅当存在 m 个 n 阶矩阵 E_1, \dots, E_m 使得对任意 $i, j = 1, \dots, m$ 且 $i \neq j$, 有性质

$$(1) A = \sum_{k=1}^m \lambda_k E_k;$$

$$(2) E_j = (E_j)^2;$$

$$(3) E_i E_j = O;$$

$$(4) E_i A = A E_i = \lambda_i E_i;$$

$$(5) \sum_{k=1}^m E_k = I;$$

$$(6) \text{谱阵集合 } \{E_1, \dots, E_m\} \text{ 唯一.}$$

证明: 必要性. 由定义 3.7.3 易证性质(1), 且有

$$[\alpha_{11}, \dots, \alpha_{md_m}] \begin{bmatrix} \beta_{11}^H \\ \vdots \\ \beta_{md_m}^H \end{bmatrix} = I \quad (3.7.9)$$

$$\begin{bmatrix} \beta_{11}^H \\ \vdots \\ \beta_{md_m}^H \end{bmatrix} [\alpha_{11}, \cdots, \alpha_{md_m}] = I \quad (3.7.10)$$

由式(3.7.9)知,

$$\sum_{j=1}^m \sum_{i=1}^{d_j} \alpha_{ji} \beta_{ji}^H = I$$

即性质(5)成立.

定义

$$X_j = [\alpha_{j1}, \cdots, \alpha_{jd_j}] \in \mathbb{C}^{n \times d_j}, \quad Y_j = \begin{bmatrix} \beta_{j1}^H \\ \vdots \\ \beta_{jd_j}^H \end{bmatrix} \in \mathbb{C}^{d_j \times n}$$

则 $E_j = X_j Y_j$ 且式(3.7.10)可改写为

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix} [X_1, \cdots, X_m] = I \quad (3.7.11)$$

由式(3.7.11)知, 当 $i, j = 1, \cdots, m$ 且 $i \neq j$ 时, $Y_j X_j = I_{d_j}$, $Y_i X_j = O$. 于是, 性质(2)和性质(3)成立. 进而,

$$\begin{aligned} E_i A &= E_i \left(\sum_{j=1}^m \lambda_j E_j \right) = \lambda_i E_i \\ A E_i &= \left(\sum_{j=1}^m \lambda_j E_j \right) E_i = \lambda_i E_i \end{aligned}$$

即性质(4)成立.

性质(6)可参考定理 3.7.2 中证明, 这里不再赘述.

充分性. 设 $\text{rank}(E_j) = d_j, j = 1, \cdots, m$, 则有

$$\sum_{j=1}^m d_j = \sum_{j=1}^m \text{tr}(E_j) = \text{tr}\left(\sum_{j=1}^m E_j\right) = \text{tr}(I_n) = n$$

由 $\dim(R(E_j)) = d_j$ 得, 可取列空间 $R(E_j)$ 的一组基 $\alpha_{j1}, \cdots, \alpha_{jd_j}$, 并定义

$$\begin{aligned} X_j &= [\alpha_{j1}, \cdots, \alpha_{jd_j}] \in \mathbb{C}^{n \times d_j} \\ X &= [X_1, \cdots, X_m] \in \mathbb{C}^{n \times n} \end{aligned}$$

由满秩分解知, $E_j = X_j Y_j$, 其中, $Y_j \in \mathbb{C}^{d_j \times n}, j = 1, \cdots, m$. 由此, 定义

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix} \in \mathbb{C}^{n \times n}$$

则有 $XY = \sum_{j=1}^m X_j Y_j = \sum_{j=1}^m E_j = I_n$. 故 X 是可逆矩阵.

再由 $YX = I_n$ 得, 当 $i, j = 1, \dots, m$ 且 $i \neq j$ 时, $Y_j X_j = I_{d_j}, Y_i X_j = O$. 由此, 计算 $E_j X_i$ 得

$$E_j X_i = X_j Y_j X_i = \begin{cases} X_j, & i = j \\ O, & i \neq j \end{cases}$$

考察矩阵 AX :

$$\begin{aligned} AX &= \left(\sum_{j=1}^m \lambda_j E_j \right) [X_1, \dots, X_m] = \left[\sum_{j=1}^m \lambda_j E_j X_1, \dots, \sum_{j=1}^m \lambda_j E_j X_m \right] \\ &= [\lambda_1 X_1, \dots, \lambda_m X_m] = \text{diag}(\lambda_1, \dots, \lambda_m) [X_1, \dots, X_m] \end{aligned}$$

即 $AX = \text{diag}(\lambda_1, \dots, \lambda_m)X$. 因此, A 是单纯矩阵. 证毕.

注 4: 同正规矩阵的谱阵相比, 单纯矩阵的谱阵是幂等矩阵, 但不是正交投影矩阵.

例 3.7.6 设 $A = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 0 & 0 \\ -2 & 0 & 4 \end{bmatrix}$, 求 $\sum_{i=1}^{100} A^i$.

解: A 的特征多项式为 $f_A(\lambda) = \lambda^2(\lambda - 5)$, 则 $\lambda_1 = \lambda_2 = 0$ 和 $\lambda_3 = 5$. 它们对应的单位特征向量分别为

$$\alpha_1 = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \sqrt{5} \\ 0 \end{bmatrix}, \quad \alpha_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \alpha_3 = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \sqrt{5} \\ -\frac{2}{\sqrt{5}} \end{bmatrix}$$

故 A 的谱分解为 $A = \lambda_1 E_1 + \lambda_2 E_2 = 5E_2$, 其中

$$E_2 = \alpha_3 \alpha_3^H = \begin{bmatrix} \frac{1}{5} & 0 & -\frac{2}{5} \\ 0 & 0 & 0 \\ -\frac{2}{5} & 0 & \frac{4}{5} \end{bmatrix}$$

由此,

$$\sum_{i=1}^{100} A^i = \left(\sum_{i=1}^{100} \lambda_2^i \right) E_2 = \frac{(5^{101} - 5)}{4} \begin{bmatrix} \frac{1}{5} & 0 & -\frac{2}{5} \\ 0 & 0 & 0 \\ -\frac{2}{5} & 0 & \frac{4}{5} \end{bmatrix}$$

推论 3.7.1 设单纯矩阵 $A \in \mathbb{C}^{n \times n}$ 的谱分解为 $A = \sum_{j=1}^m \lambda_j E_j$, $f(\lambda)$ 为数域 \mathbb{C} 上的任一多项式, 则

$$f(\mathbf{A}) = \sum_{j=1}^m f(\lambda_j) \mathbf{E}_j$$

式中, $\lambda_1, \dots, \lambda_m$ 为 \mathbf{A} 的 m 个互异特征值, $\mathbf{E}_j (j=1, \dots, m)$ 是矩阵 \mathbf{A} 的谱阵.

证明: 首先用数学归纳法证明对 $k=0, 1, 2, \dots$, 有

$$\mathbf{A}^k = \lambda_1^k \mathbf{E}_1 + \lambda_2^k \mathbf{E}_2 + \dots + \lambda_m^k \mathbf{E}_m$$

当 $k=0$ 和 $k=1$ 时, 上式显然成立. 假设当 $k=p$ 时, 有

$$\mathbf{A}^p = \lambda_1^p \mathbf{E}_1 + \lambda_2^p \mathbf{E}_2 + \dots + \lambda_m^p \mathbf{E}_m$$

现考察 $k=p+1$. 此时,

$$\begin{aligned} \mathbf{A}^{p+1} &= (\lambda_1^p \mathbf{E}_1 + \lambda_2^p \mathbf{E}_2 + \dots + \lambda_m^p \mathbf{E}_m) \mathbf{A} \\ &= \lambda_1^p \mathbf{E}_1 \mathbf{A} + \lambda_2^p \mathbf{E}_2 \mathbf{A} + \dots + \lambda_m^p \mathbf{E}_m \mathbf{A} \\ &= \lambda_1^{p+1} \mathbf{E}_1 + \lambda_2^{p+1} \mathbf{E}_2 + \dots + \lambda_m^{p+1} \mathbf{E}_m \end{aligned}$$

于是

$$\mathbf{A}^k = \lambda_1^k \mathbf{E}_1 + \lambda_2^k \mathbf{E}_2 + \dots + \lambda_m^k \mathbf{E}_m, \quad k=0, 1, 2, \dots \quad (3.7.12)$$

设 $f(\lambda) = a_n \lambda^n + a_{n-1} \lambda^{n-1} + \dots + a_1 \lambda + a_0$, 则

$$f(\mathbf{A}) = a_n \mathbf{A}^n + a_{n-1} \mathbf{A}^{n-1} + \dots + a_1 \mathbf{A} + a_0 \mathbf{I}$$

将式(3.7.12)代入上式得

$$\begin{aligned} f(\mathbf{A}) &= (a_n \lambda_1^n + a_{n-1} \lambda_1^{n-1} + \dots + a_1 \lambda_1 + a_0) \mathbf{E}_1 + \dots + \\ &\quad (a_n \lambda_m^n + a_{n-1} \lambda_m^{n-1} + \dots + a_1 \lambda_m + a_0) \mathbf{E}_m \end{aligned}$$

注意到 $f(\lambda_i) = a_n \lambda_i^n + a_{n-1} \lambda_i^{n-1} + \dots + a_1 \lambda_i + a_0$, 故 $f(\mathbf{A}) = \sum_{j=1}^m f(\lambda_j) \mathbf{E}_j$. 证毕.

推论 3.7.2 设单纯矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 的谱分解为 $\mathbf{A} = \sum_{j=1}^m \lambda_j \mathbf{E}_j$, 则

$$\mathbf{E}_i = \frac{1}{\prod_{l=1, l \neq i}^m (\lambda_i - \lambda_l)} \prod_{l=1, l \neq i}^m (\mathbf{A} - \lambda_l \mathbf{I}), \quad i=1, \dots, m$$

证明: 令 $f_i(\lambda) = \prod_{l=1, l \neq i}^m (\lambda - \lambda_l)$, 则由推论 3.7.1 知

$$f_i(\mathbf{A}) = f_i(\lambda_1) \mathbf{E}_1 + \dots + f_i(\lambda_m) \mathbf{E}_m$$

式中:

$$f_i(\lambda_j) = \begin{cases} 0, & j \neq i \\ f_i(\lambda_i), & j = i \end{cases}$$

因此, $f_i(\mathbf{A}) = f_i(\lambda_i) \mathbf{E}_i$, 即 $\mathbf{E}_i = f_i(\mathbf{A}) / f_i(\lambda_i)$. 证毕.

例 3.7.7 求矩阵 $\mathbf{A} = \begin{bmatrix} 4 & 6 & 0 \\ -3 & -5 & 0 \\ -3 & -6 & 1 \end{bmatrix}$ 的谱分解.

解: 矩阵 \mathbf{A} 的特征值为 $\lambda_1 = \lambda_2 = 1, \lambda_3 = -2$. 通过判断矩阵 $\mathbf{I} - \mathbf{A}$ 知,

$$\mathbf{I} - \mathbf{A} = \begin{bmatrix} -3 & -6 & 0 \\ 3 & 6 & 0 \\ 3 & 6 & 0 \end{bmatrix}$$

$\text{rank}(\mathbf{I} - \mathbf{A}) = 1$, 故 $\dim(\mathbf{E}(1)) = 3 - 1 = 2$. 显然, 矩阵 \mathbf{A} 是单纯矩阵.

定义 $f_1(\lambda) = \lambda + 2, f_2(\lambda) = \lambda - 1$, 则

$$\mathbf{E}_1 = \frac{1}{f_1(\lambda_1)} f_1(\mathbf{A}) = \begin{bmatrix} 2 & 2 & 0 \\ -1 & -1 & 0 \\ -1 & -2 & 1 \end{bmatrix}$$

$$\mathbf{E}_2 = \frac{1}{f_2(\lambda_2)} f_2(\mathbf{A}) = \begin{bmatrix} -1 & -2 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 0 \end{bmatrix}$$

因此, $\mathbf{A} = \mathbf{E}_1 - 2\mathbf{E}_2$.

通过例 3.7.5 和例 3.7.7 比较知, 利用推论 3.7.2 计算单纯矩阵的谱分解比利用定义法计算简单得多. 因此, 推论 3.7.2 为我们提供了一种计算单纯矩阵谱分解的简单方法.

3.8 Jordan 分解

由 Schur 引理知, 任一复方阵相似于上三角矩阵. 显然, 最简单的上三角矩阵就是对角矩阵, 但复方阵相似于对角矩阵的充分必要条件表明并非所有的复方阵都相似于对角矩阵. 此时, 同一线性变换下的矩阵的最简形式会是什么? 为研究这一问题, 我们引入 λ 矩阵的概念和理论.

定义 3.8.1 (λ 矩阵) 以 λ 多项式为元素的矩阵称为 λ 矩阵, 记为 $\mathbf{A}(\lambda)$, 即 $\mathbf{A}(\lambda) = [a_{ij}(\lambda)]_{m \times n}, a_{ij}(\lambda) \in P_n(\lambda)$.

例 3.8.1 判断 $\mathbf{A}(\lambda)$ 和 $\mathbf{B}(\lambda)$ 是否为 λ 矩阵, 其中

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

解: $\mathbf{A}(\lambda)$ 是 λ 矩阵; 由于 λ^{-2} 不是 λ 多项式, 故 $\mathbf{B}(\lambda)$ 不是 λ 矩阵.

注 1: 数字矩阵是特殊的 λ 矩阵; 复方阵 \mathbf{A} 的特征矩阵 $\lambda \mathbf{I} - \mathbf{A}$ 是 λ 矩阵.

注 2: λ 矩阵和数字矩阵一样有加、减、乘等运算且有相同的运算规律. 我们同样可定义正方 λ 矩阵的行列式、子式及 λ 矩阵的秩等.

定义 3.8.2 (λ 矩阵的秩) λ 矩阵 $\mathbf{A}(\lambda)$ 中非零子式的最高阶数 r 定义为 $\mathbf{A}(\lambda)$ 的秩, 记为 $\text{rank}(\mathbf{A}(\lambda)) = r$.

例 3.8.2 求 $A(\lambda) = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda+1 \end{bmatrix}$ 的行列式和秩.

解: $|A(\lambda)| = \lambda(\lambda+1)$, 故 $\text{rank}(A(\lambda)) = 2$ (注意: 在 λ 矩阵理论中, “当 $\lambda = 0$ 或 -1 时, λ 矩阵的秩为 1, 其余情况矩阵的秩为 2” 这种说法是错误的).

例 3.8.3 设 $A \in \mathbb{C}^{n \times n}$, 则 $f_A(\lambda) = |\lambda I - A|$ 是关于 λ 的一元 n 次多项式. 因此, A 的特征矩阵 $\lambda I - A$ 的秩为 n , 即 $\lambda I - A$ 总是满秩的.

定义 3.8.3 (λ 矩阵的逆矩阵) 设 $A(\lambda)$ 是 n 阶 λ 方阵, 若存在 n 阶 λ 方阵 $B(\lambda)$ 满足 $A(\lambda)B(\lambda) = B(\lambda)A(\lambda) = I$, 则称 λ 矩阵 $A(\lambda)$ 是可逆的, 并称 $B(\lambda)$ 为 $A(\lambda)$ 的逆矩阵, 记作 $A(\lambda)^{-1}$.

【思考】 在数字矩阵中, 满秩方阵和可逆矩阵是等价命题. 这一结论适用于 λ 矩阵吗?

由例 3.8.2 知, $A(\lambda) = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda+1 \end{bmatrix}$ 满秩. 若 $A(\lambda)$ 可逆, 则其逆矩阵应为

$$B(\lambda) = \begin{bmatrix} \frac{1}{\lambda+1} & 0 \\ 0 & \frac{1}{\lambda} \end{bmatrix}$$

但 $B(\lambda)$ 不是 λ 矩阵. 因此, λ 矩阵可逆不仅要求矩阵满秩, 还须增加条件.

定理 3.8.1 n 阶 λ 方阵 $A(\lambda)$ 可逆的充分必要条件是它的行列式 $|A(\lambda)|$ 为非零常数.

证明: 必要性. 若 λ 方阵 $A(\lambda)$ 可逆, 则存在 λ 方阵 $B(\lambda)$ 满足 $A(\lambda)B(\lambda) = I$. 对等式两端取行列式得

$$|A(\lambda)| |B(\lambda)| = 1$$

因为 $|A(\lambda)|$ 和 $|B(\lambda)|$ 均是 λ 的多项式, 所以 $|A(\lambda)|$ 和 $|B(\lambda)|$ 只能是零次多项式, 即行列式 $|A(\lambda)|$ 为非零常数.

充分性. 设 $|A(\lambda)| = \tau \neq 0$, $(A(\lambda))^*$ 是 $A(\lambda)$ 的伴随矩阵, 则

$$A(\lambda)(A(\lambda))^* = (A(\lambda))^*A(\lambda) = \tau I_n$$

注意到 $\frac{1}{\tau}(A(\lambda))^*$ 也是 n 阶 λ 矩阵, 则根据上式知, $\frac{1}{\tau}(A(\lambda))^*$ 是 $A(\lambda)$ 的逆矩阵.

证毕.

同数字矩阵一样, 我们引入 λ 矩阵的初等变换.

定义 3.8.4 (初等变换) 下列三种变换称为 λ 矩阵的初等变换:

- (1) λ 矩阵的两行(列)互换位置;
- (2) λ 矩阵的某一行(列)乘以非零常数 k ;
- (3) λ 矩阵的某一行(列)的 $\varphi(\lambda)$ 倍加到另一行(列), 其中 $\varphi(\lambda) \in P_n(\lambda)$.

对单位矩阵施行上述三种初等变换得到相应的三种 λ 矩阵的初等矩阵, 分别记

为 $P(i, j), P(i(k))$ 和 $P(i, j(\varphi))$, 并表示为

$$P(i, j) = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & 0 & \cdots & 1 & \cdots \\ & & & & 1 & & \\ & & \vdots & & \ddots & \vdots & \\ & & & & & 1 & \\ & 1 & \cdots & & 0 & \cdots & \\ & & & & & 1 & \\ & & & & & & \ddots \\ & & & & & & & 1 \end{bmatrix} \begin{matrix} \\ \\ \\ \text{第 } i \text{ 行} \\ \\ \\ \text{第 } j \text{ 行} \\ \\ \end{matrix}$$

$$P(i(c)) = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & c & \cdots & \cdots & \cdots \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \\ & & & & & & & 1 \end{bmatrix} \begin{matrix} \\ \\ \\ \text{第 } i \text{ 行} \\ \\ \\ \end{matrix}$$

$$P(i, j(\varphi)) = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & 1 & \cdots & \varphi(\lambda) & \cdots \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \\ & & & & & & & 1 \\ & & & & & & & & \ddots \\ & & & & & & & & & 1 \end{bmatrix} \begin{matrix} \\ \\ \\ \text{第 } i \text{ 行} \\ \\ \\ \text{第 } j \text{ 行} \\ \\ \end{matrix}$$

与数字矩阵一样,对 λ 矩阵作一次初等行变换意味着左乘相应的初等矩阵,对 λ 矩阵作一次初等列变换则意味着右乘相应的初等矩阵. 由于三种初等矩阵的行列式均为非零常数,故初等矩阵都是可逆的且对 λ 矩阵作初等变换不改变它的秩.

定义 3.8.5(λ 矩阵相抵) 若 λ 矩阵 $A(\lambda)$ 经过有限次初等变换变为 $B(\lambda)$, 则称 $A(\lambda)$ 与 $B(\lambda)$ 相抵, 记为 $A(\lambda) \cong B(\lambda)$.

注 3: λ 矩阵相抵则其秩相同,反之则不然,这与数字矩阵是有区别的.

例 3.8.4 考察 $A(\lambda)$ 和 $B(\lambda)$ 是否相抵,其中

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

解: $|A(\lambda)| = \lambda^2$, $|B(\lambda)| = 2\lambda$, 故两 λ 矩阵的秩均为 2. 由于初等变换是可逆的, 则两个相抵的 λ 方阵的行列式只能相差一非零常数, 故 $A(\lambda)$ 和 $B(\lambda)$ 不相抵.

定义 3.8.6 (行列式因子) 设 λ 矩阵 $A(\lambda)$ 的秩为 r , 对于正整数 $1 \leq k \leq r$, $A(\lambda)$ 的全部 k 阶子式的首 1 最大公因式称为 k 阶行列式因子, 记为 $D_k(\lambda)$.

例 3.8.5 求 $A(\lambda) = \begin{bmatrix} 1-\lambda & \lambda^2 & \lambda \\ \lambda & \lambda & -\lambda \\ 1+\lambda^2 & \lambda^2 & -\lambda^2 \end{bmatrix}$ 各阶行列式因子.

解: $A(\lambda)$ 的一阶子式为: $1-\lambda, \lambda^2, \lambda, \lambda, \lambda, -\lambda, 1+\lambda^2, \lambda^2, -\lambda^2$. 一阶因子的首 1 最大公因式为 $D_1(\lambda) = 1$.

$A(\lambda)$ 的 2 阶子式共 9 个, 分别为: $\lambda(1-\lambda-\lambda^2), -\lambda, -\lambda(\lambda^2+\lambda), -\lambda(\lambda^2+\lambda^3), -\lambda(\lambda+1), -\lambda(\lambda^3-\lambda^2), -\lambda, \lambda, 0$. 2 阶因子的首 1 最大公因式 $D_2(\lambda) = \lambda$.

$A(\lambda)$ 的 3 阶子式共 1 个, 即 $D_3(\lambda) = |A(\lambda)| = \lambda^3 + \lambda^2$.

定理 3.8.2 相抵的 λ 矩阵具有相同的秩和相同的各阶行列式因子.

分析: 只需证明 λ 矩阵经过 1 次初等变换, 秩和行列式因子不变.

设 λ 矩阵 $A(\lambda)$ 经过一次初等行变换变为 $B(\lambda)$, $f(\lambda)$ 与 $g(\lambda)$ 分别是 $A(\lambda)$ 与 $B(\lambda)$ 的 k 阶行列式因子. 我们分三种情况讨论以证明 $f(\lambda) = g(\lambda)$.

(1) $P(i, j)A(\lambda) = B(\lambda)$: $B(\lambda)$ 的每个 k 阶子式或等于 $A(\lambda)$ 的某个 k 阶子式, 或与 $A(\lambda)$ 的某一个 k 阶子式反号. 故 $f(\lambda)$ 是 $B(\lambda)$ 的 k 阶子式的首 1 公因式, 从而 $f(\lambda) | g(\lambda)$.

(2) $P(i(c))A(\lambda) = B(\lambda)$: $B(\lambda)$ 的每个 k 阶子式或等于 $A(\lambda)$ 的某个 k 阶子式, 或是 $A(\lambda)$ 的某一个 k 阶子式的 c 倍. 故 $f(\lambda)$ 是 $B(\lambda)$ 的 k 阶子式的首 1 公因式, 从而 $f(\lambda) | g(\lambda)$.

(3) $P(i+j(\varphi))A(\lambda) = B(\lambda)$: $B(\lambda)$ 中所有同时包含 i 行与 j 行的 k 阶子式和所有不包含 i 行的 k 阶子式都等于 $A(\lambda)$ 中对应的 k 阶子式; $B(\lambda)$ 中所有包含 i 行但不包含 j 行的 k 阶子式, 等于 $A(\lambda)$ 的一个 k 阶子式与另一个 k 阶子式的 $\mp \varphi(\lambda)$ 倍的和; 此时, $f(\lambda)$ 是 $B(\lambda)$ 的 k 阶子式的公因式, 从而 $f(\lambda) | g(\lambda)$.

如果 $A(\lambda)$ 经过一次初等列变换变成 $B(\lambda)$, 同理有 $f(\lambda) | g(\lambda)$. 注意到初等变换的可逆性, $B(\lambda)$ 也可以经过一次初等变换变成 $A(\lambda)$. 由上述讨论, 同样有 $g(\lambda) | f(\lambda)$. 因此, $f(\lambda) = g(\lambda)$.

定理 3.8.3 (Smith 标准形) 设 λ 矩阵 $A(\lambda)$ 的秩为 r , 则

$$\mathbf{A}(\lambda) \cong \begin{bmatrix} d_1(\lambda) & & & & & \\ & \ddots & & & & \\ & & d_r(\lambda) & & & \\ & & & 0 & & \\ & \mathbf{O} & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

式中: $d_i(\lambda)$ 是首 1 多项式, 且 $d_i(\lambda) \mid d_{i+1}(\lambda)$, 称此标准形为 $\mathbf{A}(\lambda)$ 的 Smith 标准形.

例 3.8.6 设 $\mathbf{A}(\lambda) = \begin{bmatrix} 1-\lambda & \lambda^2 & \lambda \\ \lambda & \lambda & -\lambda \\ 1+\lambda^2 & \lambda^2 & -\lambda^2 \end{bmatrix}$, 求其 Smith 标准形.

$$\begin{aligned} \text{解: } \mathbf{A}(\lambda) &\xrightarrow{\text{列①+列③}} \begin{bmatrix} 1 & \lambda^2 & \lambda \\ 0 & \lambda & -\lambda \\ 1 & \lambda^2 & -\lambda^2 \end{bmatrix} \xrightarrow{\text{列②+列①}(-\lambda^2)} \begin{bmatrix} 1 & 0 & \lambda \\ 0 & \lambda & -\lambda \\ 1 & 0 & -\lambda^2 \end{bmatrix} \\ &\xrightarrow{\text{列③+列①}(-\lambda)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 1 & 0 & -\lambda^2-\lambda \end{bmatrix} \xrightarrow{\text{行③+行①}(-1)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & -\lambda^2-\lambda \end{bmatrix} \\ &\xrightarrow{\text{列③+列②}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & -\lambda^2-\lambda \end{bmatrix} \xrightarrow{\text{列③}(-1)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda^2+\lambda \end{bmatrix} \end{aligned}$$

根据定理 3.8.2 和定理 3.8.3 易得如下关系:

$$\begin{aligned} D_1(\lambda) &= d_1(\lambda) \\ D_2(\lambda) &= d_1(\lambda) d_2(\lambda) \\ &\vdots \\ D_r(\lambda) &= d_1(\lambda) d_2(\lambda) \cdots d_r(\lambda) \end{aligned}$$

或者可以写成:

$$\begin{aligned} d_1(\lambda) &= D_1(\lambda) \\ d_2(\lambda) &= \frac{D_2(\lambda)}{D_1(\lambda)} \\ &\vdots \\ d_r(\lambda) &= \frac{D_r(\lambda)}{D_{r-1}(\lambda)} \end{aligned}$$

因此, 有以下推论.

推论 3.8.1 λ 矩阵的 Smith 标准形是唯一的.

定义 3.8.7 (不变因子) 在 λ 矩阵 $\mathbf{A}(\lambda)$ 的 Smith 标准形中, $d_1(\lambda), \dots, d_r(\lambda)$ 由 $\mathbf{A}(\lambda)$ 唯一确定的, 称为 $\mathbf{A}(\lambda)$ 的不变因子.

例 3.8.7 求下列 λ 矩阵的不变因子, 其中

$$\mathbf{A}(\lambda) = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}, \quad \mathbf{B}(\lambda) = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$$

解: 对于 $\mathbf{A}(\lambda)$, 有 $D_1(\lambda) = \lambda, D_2(\lambda) = \lambda^2$. 故 $d_1(\lambda) = D_1(\lambda) = \lambda, d_2(\lambda) = D_2(\lambda)/D_1(\lambda) = \lambda$. 对于 $\mathbf{B}(\lambda)$, 有 $D_1(\lambda) = 1, D_2(\lambda) = \lambda^2$. 故 $d_1(\lambda) = D_1(\lambda) = 1, d_2(\lambda) = D_2(\lambda)/D_1(\lambda) = \lambda^2$.

由于 $\mathbf{A}(\lambda)$ 和 $\mathbf{B}(\lambda)$ 的行列式因子不完全相同, 故这两个 λ 矩阵不相抵. 当然, 我们也可直接通过初等变换知, $\mathbf{A}(\lambda)$ 的 Smith 标准形就是它本身, 而 $\mathbf{B}(\lambda)$ 的 Smith 标准形为 $\text{diag}(1, \lambda^2)$.

推论 3.8.2 λ 矩阵 $\mathbf{A}(\lambda)$ 与 $\mathbf{B}(\lambda)$ 相抵当且仅当它们有完全一致的不变因子.

注 4: 初等变换不改变 λ 矩阵的不变因子. 例如, 再考察例 3.8.6:

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

其中, $\mathbf{A}(\lambda)$ 的不变因子可由其 Smith 标准形求得, 即 $d_1(\lambda) = 1, d_2(\lambda) = \lambda$ 和 $d_3(\lambda) = \lambda(\lambda+1)$. 当然, $\mathbf{A}(\lambda)$ 的不变因子也可由行列式因子求得, 这两种方法求得的结果一致.

注意到复数域 \mathbb{C} 上的一元多项式可分解成一次因子的幂的乘积形式. 设 $\mathbf{A}(\lambda)$ 的不变因子 $d_1(\lambda), \dots, d_r(\lambda)$ 可分解为

$$\begin{cases} d_1(\lambda) = (\lambda - \lambda_1)^{e_{11}} (\lambda - \lambda_2)^{e_{12}} \cdots (\lambda - \lambda_s)^{e_{1s}} \\ d_2(\lambda) = (\lambda - \lambda_1)^{e_{21}} (\lambda - \lambda_2)^{e_{22}} \cdots (\lambda - \lambda_s)^{e_{2s}} \\ \vdots \\ d_r(\lambda) = (\lambda - \lambda_1)^{e_{r1}} (\lambda - \lambda_2)^{e_{r2}} \cdots (\lambda - \lambda_s)^{e_{rs}} \end{cases} \quad (3.8.1)$$

式中: $\lambda_1, \dots, \lambda_s$ 互异, $e_{ij} (i=1, \dots, r; j=1, \dots, s)$ 为非负整数.

由 $d_i(\lambda) \mid d_{i+1}(\lambda)$ 知,

$$\begin{cases} 0 \leq e_{11} \leq e_{21} \leq \cdots \leq e_{r1} \\ 0 \leq e_{12} \leq e_{22} \leq \cdots \leq e_{r2} \\ \vdots \\ 0 \leq e_{1s} \leq e_{2s} \leq \cdots \leq e_{rs} \end{cases}$$

即序列 $e_{1j}, e_{2j}, \dots, e_{rj}$ 是非严格递增的. 由此引出初等因子的概念.

定义 3.8.8 (初等因子) 设 λ 矩阵 $\mathbf{A}(\lambda)$ 的不变因子为 $d_1(\lambda), \dots, d_r(\lambda)$, 且有分解式 (3.8.1), 则所有幂指数大于零的因子 $(\lambda - \lambda_j)^{e_{ij}}, i=1, \dots, r, j=1, \dots, s$, 统称为 λ 矩阵 $\mathbf{A}(\lambda)$ 的初等因子(组).

例 3.8.8 设 $\mathbf{A}(\lambda) = \text{diag}(\lambda(\lambda+1), \lambda^2, (\lambda+1)^2, 0)$. 求 $\mathbf{A}(\lambda)$ 的初等因子、不变因子和 Smith 标准形.

解: $A(\lambda)$ 的行列式因子为 $D_1(\lambda)=1, D_2(\lambda)=\lambda(\lambda+1), D_3(\lambda)=\lambda^3(\lambda+1)^3$, 所以 $A(\lambda)$ 的不变因子 $d_3(\lambda)=\lambda^2(\lambda+1)^2, d_2(\lambda)=\lambda(\lambda+1), d_1(\lambda)=1$. 因此, 写出其 Smith 标准形为

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda(\lambda+1) & 0 & 0 \\ 0 & 0 & \lambda^2(\lambda+1)^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

根据 Smith 标准形得, $A(\lambda)$ 的初等因子为: $\lambda, \lambda+1, \lambda^2, (\lambda+1)^2$.

注 5: 若 $A(\lambda) \cong B(\lambda)$, 则 $A(\lambda)$ 与 $B(\lambda)$ 有完全一致的行列式因子, 进而有相同的不变因子. 所以, $A(\lambda)$ 与 $B(\lambda)$ 有完全一致的初等因子. 这表明初等变换不改变 $A(\lambda)$ 的初等因子.

【思考】 若 $A(\lambda)$ 与 $B(\lambda)$ 有完全一致的初等因子, 则 $A(\lambda) \cong B(\lambda)$ 成立吗?

例 3.8.9 考察 $A(\lambda)$ 与 $B(\lambda)$ 是否相抵, 其中

$$A(\lambda) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & (\lambda-2)(\lambda-3) & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad B(\lambda) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (\lambda-2)(\lambda-3) \end{bmatrix}$$

解: $A(\lambda)$ 与 $B(\lambda)$ 均是 Smith 标准形, 但两者不同, 故 $A(\lambda)$ 与 $B(\lambda)$ 不相抵.

注意到例 3.8.9 中 $A(\lambda)$ 与 $B(\lambda)$ 有完全相同的初等因子: $(\lambda-2), (\lambda-3)$. 因此, $A(\lambda)$ 与 $B(\lambda)$ 有完全一致的初等因子, 并不意味着这两个矩阵相抵.

定理 3.8.4 λ 矩阵 $A(\lambda) \cong B(\lambda)$ 当且仅当它们有完全一致的初等因子, 且 $\text{rank}(A(\lambda)) = \text{rank}(B(\lambda))$.

分析: 由 Smith 标准形的唯一性可知, $A(\lambda)$ 与 $B(\lambda)$ 相抵, 故 $A(\lambda)$ 与 $B(\lambda)$ 具有完全一致的不变因子和初等因子. 另一方面, 记 $(\lambda - \lambda_j)^{e_{ij}} (i=1, \dots, r, j=1, \dots, s)$ 是 $A(\lambda)$ 与 $B(\lambda)$ 的初等因子, 则幂底因子为 $(\lambda - \lambda_j)$ 的最高幂指数必是不变因子 $d_r(\lambda)$ 的因子. 注意到 $A(\lambda)$ 与 $B(\lambda)$ 的秩相同, 且初等因子相同, 因此, 它们的不变因子 $d_r(\lambda)$ 也必相同. 同理, 可依次推断出 $A(\lambda)$ 与 $B(\lambda)$ 具有相同的不变因子 $d_{r-1}(\lambda), \dots, d_1(\lambda)$. 因此, $A(\lambda)$ 与 $B(\lambda)$ 具有相同的 Smith 标准形, 即 $A(\lambda) \cong B(\lambda)$.

定理 3.8.5 设 λ 矩阵 $A(\lambda)$ 为对角块矩阵, 即

$$A(\lambda) = \text{diag}(A_1(\lambda), \dots, A_s(\lambda))$$

则 $A_1(\lambda), \dots, A_s(\lambda)$ 初等因子的全体就是 $A(\lambda)$ 的全部初等因子, 其中 $A_i(\lambda), i=1, \dots, s$ 是适当阶数的 λ 矩阵.

由此, 在求取 λ 矩阵的 Smith 标准形、不变因子或初等因子时可先将 λ 矩阵作初等变换, 使得变换后的矩阵为对角(块)矩阵, 然后利用定理 3.8.5 可方便地求出 λ 矩阵初等因子, 进而求出 Smith 标准形或不变因子.

再次考察例 3.8.8. 由于 $A(\lambda)$ 是对角块 λ 矩阵, 故其初等因子为: $\lambda, \lambda+1, \lambda^2,$

$(\lambda+1)^2$. 基于此, 写成其不变因子为: $d_3(\lambda) = \lambda^2(\lambda+1)^2$, $d_2(\lambda) = \lambda(\lambda+1)$, $d_1(\lambda) = 1$. 进而得出其 Smith 标准形.

例 3.8.10 求特征矩阵 $\lambda I - A$ 的 Smith 标准形, 其中

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

解: 记 $A = \text{diag}(A_1, A_2)$, 其中,

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

又知

$$\lambda I - A_1 = \begin{bmatrix} \lambda & -1 \\ 1 & \lambda \end{bmatrix} \cong \begin{bmatrix} 1 & 0 \\ 0 & \lambda^2 + 1 \end{bmatrix}$$

故方阵 A_1 的初等因子是: $\lambda+i, \lambda-i$.

另有

$$\lambda I - A_2 = \begin{bmatrix} \lambda & -1 & -1 \\ -1 & \lambda & -1 \\ -1 & -1 & \lambda \end{bmatrix} \cong \begin{bmatrix} 1 & 0 & 0 \\ 0 & (\lambda+1) & 0 \\ 0 & 0 & (\lambda-2)(\lambda+1) \end{bmatrix}$$

因此, 方阵 A 的初等因子是: $\lambda+i, \lambda-i, \lambda+1, \lambda+1, \lambda-2$. 矩阵 $\lambda I - A$ 的 Smith 标准形是

$$\begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & \lambda+1 & & \\ & & & & (\lambda+i)(\lambda-i)(\lambda+1)(\lambda-2) & \end{bmatrix}$$

例 3.8.11 求 $A(\lambda) = \begin{bmatrix} \lambda-a & -1 & & & \\ & \lambda-a & -1 & & \\ & & \lambda-a & \ddots & \\ & & & \ddots & -1 \\ & & & & \lambda-a \end{bmatrix}_{n \times n}$ 的初等因子、不变

因子和 Smith 标准形.

解: 考察 $A(\lambda)$ 的 $(n-1)$ 阶子式

$$\begin{vmatrix} -1 & & & & \\ \lambda - a & -1 & & & 0 \\ & & \ddots & \ddots & \\ 0 & & & \lambda - a & -1 \end{vmatrix} = (-1)^{n-1}$$

故 $A(\lambda)$ 的 $(n-1)$ 阶子式的不变因子为 $d_{n-1}=1$.

又知 $|A| = (\lambda - a)^n$, 故 $A(\lambda)$ 的不变因子为 $d_1 = d_2 = \cdots = d_{n-1} = 1, d_n = (\lambda - a)^n$. 相应的, $A(\lambda)$ 的初等因子为 $(\lambda - a)^n$, 其 Smith 标准形为

$$A(\lambda) \cong \begin{bmatrix} 1 & & & \\ & 1 & & \mathbf{O} \\ & & \ddots & \\ \mathbf{O} & & & (\lambda - a)^n \end{bmatrix}$$

定理 3.8.6 复方阵 A 和 B 相似当且仅当它们的特征矩阵相抵.

证明: 必要性. 若 A 与 B 相似, 则存在可逆矩阵 P 使得 $P^{-1}AP = B$. 从而

$$P^{-1}(\lambda I - A)P = \lambda I - P^{-1}AP = \lambda I - B$$

故 $\lambda I - A$ 与 $\lambda I - B$ 相抵.

充分性. 由于 $\lambda I - A$ 与 $\lambda I - B$ 相抵, 故存在可逆 λ 矩阵 $P(\lambda)$ 和 $Q(\lambda)$ 使得

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

于是, 有

$$(P(\lambda))^{-1}(\lambda I - A) = (\lambda I - B)Q(\lambda) \quad (3.8.2)$$

根据定理 2.1.3 知, 存在可逆 λ 矩阵 $P_1(\lambda)$ 、 $Q_1(\lambda)$ 和数字矩阵 P_0 、 Q_0 满足

$$P(\lambda) = (\lambda I - A)P_1(\lambda) + P_0 \quad (3.8.3)$$

$$Q(\lambda) = Q_1(\lambda)(\lambda I - A) + Q_0 \quad (3.8.4)$$

将式(3.8.4)代入式(3.8.2), 并整理得

$$[P^{-1}(\lambda) - (\lambda I - B)Q_1(\lambda)](\lambda I - A) = (\lambda I - B)Q_0 \quad (3.8.5)$$

比较上式两端得 $T = P^{-1}(\lambda) - (\lambda I - B)Q_1(\lambda)$ 必为数字矩阵.

又知

$$I = P(\lambda)T + P(\lambda)(\lambda I - B)Q_1(\lambda)$$

故将式(3.8.3)代入上式得

$$\begin{aligned} I &= [(\lambda I - A)P_1(\lambda) + P_0]T + (\lambda I - A)(Q(\lambda))^{-1}Q_1(\lambda) \\ &= (\lambda I - A)[P_1(\lambda)T + (Q(\lambda))^{-1}Q_1(\lambda)] + P_0T \end{aligned}$$

比较上式两端得

$$\begin{aligned} P_1(\lambda)T + (Q(\lambda))^{-1}Q_1(\lambda) &= \mathbf{O} \\ P_0T &= I \end{aligned}$$

这表明 T 是可逆的. 此时, 由式(3.8.5)得

$$(\lambda I - A) = P_0(\lambda I - B)Q_0$$

比较上式两端得

$$P_0 Q_0 = I, \quad A = P_0 B Q_0$$

因此, $A = Q_0^{-1} B Q_0$, 即 A 和 B 相似. 证毕.

推论 3.8.3 复方阵 A 是单纯矩阵的充分必要条件是它的特征矩阵 $\lambda I - A$ 的初等因子为一次的.

推论 3.8.4 复方阵 A 是单纯矩阵的充分必要条件是它的特征矩阵 $\lambda I - A$ 的不变因子无重根.

例 3.8.12 判断矩阵 A 是否为单纯矩阵, 其中

$$A = \begin{bmatrix} a & 1 & & & \\ & a & 1 & & \\ & & a & \ddots & \\ & O & & \ddots & 1 \\ & & & & a \end{bmatrix}_{n \times n}$$

解: 由例 3.8.11 知, 特征矩阵 $\lambda I - A$ 的初等因子为 $(\lambda - a)^n$. 故当 $n \geq 1$ 时, 其初等因子不是一次, 即矩阵 A 不是单纯矩阵.

定义 3.8.9 (Jordan 块) 设 $A = [a_{ij}] \in \mathbb{C}^{n \times n}$, 其特征矩阵 $\lambda I - A$ 的初等因子为 $(\lambda - \lambda_1)^{n_1}, (\lambda - \lambda_2)^{n_2}, \dots, (\lambda - \lambda_s)^{n_s}$. 对 $(\lambda - \lambda_i)^{n_i}$ 作 n_i 阶矩阵

$$J_i = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & O & & \lambda_i & 1 \\ & & & & \lambda_i \end{bmatrix}_{n_i \times n_i}$$

则称矩阵 $J_i (i=1, \dots, s)$ 为 A 的 Jordan 块.

例 3.8.13 判断下列矩阵是否为 Jordan 块, 其中

$$A_1 = \begin{bmatrix} i+1 & 1 & 0 \\ 0 & i+1 & 1 \\ 0 & 0 & i+1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

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

解: 根据 Jordan 块定义知, 只有 A_1 是 Jordan 块.

例 3.8.14 求 Jordan 块 $J_i = \begin{bmatrix} \lambda_i & 1 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & O & & & \lambda_i \end{bmatrix}_{n_i \times n_i}$ 的最小多项式.

解: 由矩阵 J_i 的特征多项式为 $f_{J_i}(\lambda) = (\lambda - \lambda_i)^{n_i}$ 知, 其最小多项式可能为 $(\lambda - \lambda_i), \dots, (\lambda - \lambda_i)^{n_i}$. 经计算知, 对 $j = 1, \dots, n_i - 1, (J_i - \lambda_i I)^j \neq O$. 因此, 矩阵 J_i 的最小多项式为 $m_{J_i}(\lambda) = (\lambda - \lambda_i)^{n_i}$.

注 6: 任一 Jordan 块的最小多项式等于它的特征多项式, 也是 Jordan 块所对应的初等因子.

定义 3.8.10 (Jordan 标准形) 设 $A = [a_{ij}] \in \mathbb{C}^{n \times n}$, 其特征矩阵 $\lambda I - A$ 的初等因子为 $(\lambda - \lambda_1)^{n_1}, \dots, (\lambda - \lambda_s)^{n_s}$, 其对应的 Jordan 块分别记为 J_1, \dots, J_s , 则由 s 个 Jordan 块作 n 阶对角块矩阵 $J = \text{diag}(J_1, \dots, J_s)$ 称为 A 的 Jordan 标准形 (或 Jordan 法式).

例 3.8.15 设 $(\lambda I - A) \cong \text{diag}(\lambda(\lambda + 1), \lambda^2, (\lambda + 1)^2, 1, 1, 1)$, 求 A 的 Jordan 标准形.

解: $(\lambda I - A)$ 的初等因子为: $\lambda, (\lambda + 1), \lambda^2, (\lambda + 1)^2$. 依次写出对应的 Jordan 块为

$$J_1 = [0], \quad J_2 = [-1], \quad J_3 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad J_4 = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$$

因此, A 的 Jordan 标准形为

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

定理 3.8.7 (Jordan 标准形定理) 设矩阵 J 是复方阵 A 的 Jordan 标准形, 则矩阵 A 与矩阵 J 相似.

证明: 特征矩阵 $\lambda I - A$ 的初等因子为 $(\lambda - \lambda_1)^{n_1}, \dots, (\lambda - \lambda_s)^{n_s}$, 其对应的 Jordan 块分别为 J_1, \dots, J_s , 则 $J = \text{diag}(J_1, \dots, J_s)$ 是矩阵 A 的 Jordan 标准形. 显然, λ 矩阵 $\lambda I - J$ 与 $\lambda I - A$ 的初等因子完全相同. 因此, $\lambda I - A$ 与 $\lambda I - J$ 相抵, 即 A 与 J 相似.

例 3.8.16 求矩阵 A 的 Jordan 标准形 J , 并求 P 使得 $P^{-1}AP = J$, 其中

$$A = \begin{bmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{bmatrix}$$

解: 对矩阵 A 作初等变换得

$$\lambda I - A \cong \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda - 1 & 0 \\ 0 & 0 & (\lambda - 1)^2 \end{bmatrix}$$

则特征矩阵 $\lambda I - A$ 的初等因子为 $(\lambda - 1), (\lambda - 1)^2$. 由此, A 的 Jordan 标准形为

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

设 $P = [p_1, p_2, p_3]$, 则 $A[p_1, p_2, p_3] = [p_1, p_2, p_3]J$. 整理得

$$\begin{cases} Ap_1 = p_1 \\ Ap_2 = p_2 \\ Ap_3 = p_2 + p_3 \end{cases}$$

由 $Ap_i = p_i$ 解得两个线性无关的向量为 $p_1 = [3, 0, 1]^T$ 和 $p_2 = [0, 3, 1]^T$.

将 $p_2 = [0, 3, 1]^T$ 代入 $Ap_3 = p_2 + p_3$ 发现此方程无解(为什么?). 重新调整方程 $Ap_i = p_i$ 的解, 经观察得 $p_1 = [3, 0, 1]^T$ 和 $p_2 = [2, 1, 1]^T$. 此时, $p_3 = [-1, 0, 0]^T$.

因此,

$$P = \begin{bmatrix} 3 & 2 & -1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

例 3.8.17 求复方阵 A 的最小多项式.

解: 利用矩阵 A 的 Jordan 标准形 J 进行求解. 若标准形 J 仅包含一个 Jordan 块 J_1 时, 由例 3.8.14 知, $m_{J_1}(\lambda) = (\lambda - \lambda_1)^{k_1}$, 故 J_1 的最小多项式恰好等于它所对应的初等因子.

当 J 包含两个 Jordan 块 J_1 和 J_2 时, 下面分两种情况讨论:

(1) 若 $\lambda_1 = \lambda_2$, 此时 J_1 和 J_2 的最小多项式分别为

$$m_{J_1}(\lambda) = (\lambda - \lambda_1)^{k_1}, \quad m_{J_2}(\lambda) = (\lambda - \lambda_2)^{k_2}$$

经计算知, 矩阵 J 的最小多项式为 $m_{J_1}(\lambda)$ 和 $m_{J_2}(\lambda)$ 的最小公倍式.

(2) 若 $\lambda_i \neq \lambda_p$, 矩阵 J 的最小多项式仍为 $m_{J_i}(\lambda)$ 和 $m_{J_p}(\lambda)$ 的最小公倍式.

将上述结论推广至一般的 Jordan 标准形 J 时, J 的最小多项式等于特征矩阵 $\lambda I - A$ 的初等因子的最小公倍式. 又知所有初等因子的最高幂指数必为不变因子 $d_n(\lambda)$ 的因子, 因此, J 的最小多项式恰为不变因子 $d_n(\lambda)$.

定理 3.8.8 (Frobenius 定理) 设 $A \in \mathbb{C}^{n \times n}$, 其特征矩阵 $\lambda I - A$ 的 Smith 标准形为 $\text{diag}(d_1(\lambda), \dots, d_n(\lambda))$, 则 A 的最小多项式为 $m_A(\lambda) = d_n(\lambda)$.

例 3.8.18 求以下矩阵的最小多项式, 其中

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

解: 矩阵 $\lambda I - A_1$ 的初等因子为 $\lambda - 3, (\lambda - 3)^3$, 故 A_1 的最小多项式为 $(\lambda - 3)^3$.

矩阵 $\lambda I - A_2$ 的初等因子为 $(\lambda-2)^2, (\lambda-3)^2$, 故 A_2 的最小多项式为 $(\lambda-2)^2(\lambda-3)^2$.

3.9 奇异值分解

尽管 Jordan 分解在理论分析或数值计算时表现出诸多优点,但不可否认的是它也存在着一一定的局限性.归纳如下:

- (1) 矩阵 A 必须是方阵;
- (2) 尽管 Jordan 标准形是同一线性变换下最简单的上三角矩阵,但不一定是对角矩阵;
- (3) 相似变换矩阵是一般的可逆矩阵,非酉矩阵.

【思考】 是否存在一种在一定程度上能够克服上述分解局限性的矩阵分解?

分析: 针对任一矩阵 $A \in \mathbb{C}^{m \times n}$, 我们可构造两个 Hermite 矩阵 AA^H 与 $A^H A$, 则分别存在两个酉矩阵 $U \in \mathbb{C}^{m \times m}$ 与 $V \in \mathbb{C}^{n \times n}$ 使得

$$U^H AA^H U = \Sigma_1 \quad (3.9.1)$$

$$V^H A^H A V = \Sigma_2 \quad (3.9.2)$$

式中: Σ_1 和 Σ_2 分别是 m 阶和 n 阶对角矩阵, 其对角线元素分别是 Hermite 矩阵 AA^H 与 $A^H A$ 的特征值.

通过观察、比较和计算得, 矩阵 A 可作分解:

$$A = UDV^H \quad (3.9.3)$$

且须满足

$$DD^H = \Sigma_1, \quad D^H D = \Sigma_2 \quad (3.9.4)$$

式中: $D \in \mathbb{C}^{m \times n}$ 待定.

显然, 只要矩阵方程 (3.9.4) 有解 D , 则分解式 (3.9.3) 存在. 下面详细阐述这一分解的可行性.

引理 3.9.1 设 $A \in \mathbb{C}^{m \times n}$, 则 $\text{rank}(A^H A) = \text{rank}(AA^H) = \text{rank}(A)$.

证明: $\forall x \in N(A^H A)$, 有

$$x \in N(A^H A) \Leftrightarrow A^H A x = 0 \Leftrightarrow x^H A^H A x = 0 \Leftrightarrow A x = 0 \Leftrightarrow x \in N(A)$$

即 $N(A^H A) = N(A)$.

由定理 2.3.6 知

$$\begin{aligned} \text{rank}(A^H A) &= n - \dim N(A^H A) \\ &= n - \dim N(A) = \text{rank}(A) \end{aligned}$$

同理, $\text{rank}(A^H) = \text{rank}(AA^H)$. 又知 $\text{rank}(A) = \text{rank}(A^H)$, 故引理得证.

注 1: $N(A^H A) = N(A)$; $N(AA^H) = N(A^H)$.

引理 3.9.2 设 $A \in \mathbb{C}^{m \times n}$, 则 $A^H A$ 与 AA^H 都是半正定 Hermite 矩阵.

证明: $\forall x \in \mathbb{C}^n$ 且 $x \neq 0$, 有

$$\mathbf{x}^H(\mathbf{A}^H\mathbf{A})\mathbf{x} = (\mathbf{A}\mathbf{x})^H(\mathbf{A}\mathbf{x}) \geq 0$$

故 $\mathbf{A}^H\mathbf{A}$ 是半正定 Hermite 矩阵. 同理可证, $\mathbf{A}\mathbf{A}^H$ 是半正定 Hermite 矩阵.

注 2: 矩阵 \mathbf{A} 是半正定 Hermite 矩阵还有以下等价命题:

(1) 对于任意 n 阶可逆矩阵 \mathbf{P} , $\mathbf{P}^H\mathbf{A}\mathbf{P}$ 是半正定 Hermite 矩阵;

(2) Hermite 矩阵 \mathbf{A} 的 n 个特征值均为非负数;

(3) 存在 n 阶可逆矩阵 \mathbf{P} 使得 $\mathbf{P}^H\mathbf{A}\mathbf{P} = \begin{bmatrix} \mathbf{I}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}$, 其中 $r = \text{rank}(\mathbf{A})$;

(4) 存在秩为 $\text{rank}(\mathbf{A})$ 的矩阵 \mathbf{Q} 使得 $\mathbf{A} = \mathbf{Q}^H\mathbf{Q}$;

(5) 存在 n 阶 Hermite 矩阵 \mathbf{G} 使得 $\mathbf{A} = \mathbf{G}^2$;

(6) Hermite 矩阵 \mathbf{A} 的所有顺序主子式均非负.

引理 3.9.3 设 $\mathbf{A} \in \mathbb{C}_r^{m \times n}$, 则 $\mathbf{A}^H\mathbf{A}$ 与 $\mathbf{A}\mathbf{A}^H$ 的所有非零特征值完全相同且非零特征值的个数均为 r .

证明: 设 λ 是 $\mathbf{A}^H\mathbf{A}$ 的非零特征值, \mathbf{x} 是属于特征值 λ 的特征向量, 则

$$\mathbf{A}^H\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \quad (3.9.5)$$

且 $\mathbf{A}\mathbf{x} \neq \mathbf{0}$ (若 $\mathbf{A}\mathbf{x} = \mathbf{0}$, 则 $\mathbf{A}^H\mathbf{A}\mathbf{x} = \lambda\mathbf{x} = \mathbf{0}$, 进而 $\lambda = 0$). 等式 (3.9.5) 两端同时左乘矩阵 \mathbf{A} , 得

$$\mathbf{A}\mathbf{A}^H(\mathbf{A}\mathbf{x}) = \lambda(\mathbf{A}\mathbf{x})$$

上式表明 λ 也是 $\mathbf{A}\mathbf{A}^H$ 的特征值, 且 $\mathbf{A}\mathbf{x}$ 是属于 λ 的特征向量 (注意 $\mathbf{A}\mathbf{x} \neq \mathbf{0}$).

同理可证: 若 λ 是 $\mathbf{A}\mathbf{A}^H$ 的非零特征值, \mathbf{x} 是属于 λ 的特征向量, 则 λ 也是 $\mathbf{A}^H\mathbf{A}$ 的特征值, $\mathbf{A}^H\mathbf{x}$ 是属于 λ 的特征向量. 因此, $\mathbf{A}^H\mathbf{A}$ 与 $\mathbf{A}\mathbf{A}^H$ 具有相同的非零特征值 (不考虑代数重数).

现证明矩阵 $\mathbf{A}^H\mathbf{A}$ 与 $\mathbf{A}\mathbf{A}^H$ 的同一非零特征值的代数重数相同. 由引理 3.9.1 知, $\text{rank}(\mathbf{A}^H\mathbf{A}) = \text{rank}(\mathbf{A}\mathbf{A}^H) = r$. 又知矩阵 $\mathbf{A}^H\mathbf{A}$ 和 $\mathbf{A}\mathbf{A}^H$ 均为单纯矩阵, 故这两个矩阵均有 r 个非零特征值.

不妨设 $\lambda_1, \dots, \lambda_s$ 是矩阵 $\mathbf{A}^H\mathbf{A}$ 的 s 个非零且互异的特征值, 它们的代数重数分别为 d_1, \dots, d_s , 则有 $d_1 + \dots + d_s = r$, 且特征值 λ_i 有 d_i 个线性无关的特征向量. 令 $\mathbf{x}_{i1}, \dots, \mathbf{x}_{id_i}$ 是属于特征值 λ_i 的 d_i 个线性无关的特征向量, 则 $\mathbf{A}\mathbf{x}_{i1}, \dots, \mathbf{A}\mathbf{x}_{id_i}$ 是属于矩阵 $\mathbf{A}\mathbf{A}^H$ 特征值 λ_i 的特征向量. 令

$$k_{i1}\mathbf{A}\mathbf{x}_{i1} + \dots + k_{id_i}\mathbf{A}\mathbf{x}_{id_i} = \mathbf{0} \quad (3.9.6)$$

式中: k_{i1}, \dots, k_{id_i} 为待定复数, $i = 1, \dots, s$.

对式 (3.9.6) 两端左乘 \mathbf{A}^H , 整理得

$$\lambda_i(k_{i1}\mathbf{x}_{i1} + \dots + k_{id_i}\mathbf{x}_{id_i}) = \mathbf{0} \quad (3.9.7)$$

注意 $\lambda_i \neq 0$, 故 $k_{i1}\mathbf{x}_{i1} + \dots + k_{id_i}\mathbf{x}_{id_i} = \mathbf{0}$. 由于向量组 $\mathbf{x}_{i1}, \dots, \mathbf{x}_{id_i}$ 线性无关, 故 k_{i1}, \dots, k_{id_i} 均等于 0. 由此推导出向量组 $\mathbf{A}\mathbf{x}_{i1}, \dots, \mathbf{A}\mathbf{x}_{id_i}$ 线性无关. 这表明属于矩阵 $\mathbf{A}\mathbf{A}^H$ 特征值 λ_i 的特征向量至少有 d_i 个线性无关的特征向量. 对 i 由 1 到 s 加和得, 矩阵

$\mathbf{A}\mathbf{A}^H$ 的非零特征值数(计重数)必不小于 r . 已知 $\mathbf{A}\mathbf{A}^H$ 的非零特征值数等于 r , 故矩阵 $\mathbf{A}\mathbf{A}^H$ 特征值 λ_i 的特征向量有且仅有 d_i 个线性无关的特征向量. 换言之, 矩阵 $\mathbf{A}^H\mathbf{A}$ 和 $\mathbf{A}\mathbf{A}^H$ 的同一非零特征值 λ_i 的代数重数相同. 证毕.

例 3.9.1 设 $\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, 计算 $\mathbf{A}^H\mathbf{A}$ 和 $\mathbf{A}\mathbf{A}^H$ 的特征值.

解: $\mathbf{A}^H\mathbf{A} = \text{diag}(1, 1, 0)$, $\mathbf{A}\mathbf{A}^H = \mathbf{I}_2$. 显然, $\mathbf{A}^H\mathbf{A}$ 的特征值为 $\lambda_1=1, \lambda_2=1, \lambda_3=0$. $\mathbf{A}\mathbf{A}^H$ 的特征值 $\lambda_1=1, \lambda_2=1$.

注 3: 在计算 $\mathbf{A}^H\mathbf{A}$ 或 $\mathbf{A}\mathbf{A}^H$ 的特征值时, 可优先考虑计算阶数阶小的矩阵.

定义 3.9.1 (奇异值) 设 $\mathbf{A} \in \mathbb{C}_r^{m \times n}$, $\mathbf{A}^H\mathbf{A}$ 的特征值满足 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$, $\lambda_{r+1} = \lambda_{r+2} = \dots = \lambda_n = 0$, 称 $\sigma_i = \sqrt{\lambda_i}$ 为 \mathbf{A} 的奇异值. 特别地, 称 $\sigma_i, i=1, \dots, r$, 为 \mathbf{A} 的正奇异值.

例 3.9.2 计算 $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ i & 0 \end{bmatrix}$ 的正奇异值.

解: 因为

$$\mathbf{A}^H\mathbf{A} = \begin{bmatrix} 1 & 0 & -i \\ 2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ i & 0 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix}$$

计算 $\mathbf{A}^H\mathbf{A}$ 特征多项式为 $|\lambda\mathbf{I} - \mathbf{A}^H\mathbf{A}| = \lambda^2 - 7\lambda + 6$, 并求得它的特征值分别为 $\lambda_1=1, \lambda_2=6$. 由此得到 \mathbf{A} 的正奇异值为

$$\sigma_1 = \sqrt{\lambda_1} = 1, \quad \sigma_2 = \sqrt{\lambda_2} = \sqrt{6}$$

【思考】 复方阵的特征值与奇异值有何异同?

分析: 方阵的特征值和奇异值很容易被误认为完全相同或模值相同. 实际上, 这个结论并不正确. 这里以三个不同矩阵为例进行说明. 考察如下三个矩阵:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & k \\ 0 & 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$$

分别计算其特征值和奇异值为:

\mathbf{A} 的特征值与奇异值相同, 均为 $0, 2$;

\mathbf{B} 的特征值为 0 和 1 , 奇异值为 $\sigma_1 = \sqrt{k^2+1}, \sigma_2 = 0$;

\mathbf{C} 的特征值为 1 和 1 , 奇异值为 $\sigma_1 = \sqrt{\frac{1}{2}(k^2+2+k\sqrt{k^2+4})}, \sigma_2 = \frac{1}{\sigma_1}$.

从计算结果看出, 同一矩阵的特征值可能与其奇异值完全相同(矩阵 \mathbf{A}), 部分相同(矩阵 \mathbf{B}), 也可能完全不同(矩阵 \mathbf{C}); 对于特征值和奇异值不同的矩阵, 其模值也难有显而易见的结果. 常见的结果有:

命题 3.9.1 设 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$ 和 $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$ 分别是 n 阶正规矩

阵 \mathbf{A} 的奇异值和特征值, 则 $\sigma_i = |\lambda_i|, i=1, \dots, n$.

命题 3.9.2 设 σ_i 和 $\lambda_i (i=1, \dots, n)$ 分别是 n 阶复方阵 \mathbf{A} 的奇异值和特征值, 则

$$\sum_{i=1}^n |\lambda_i|^2 \leq \sum_{i=1}^n |\sigma_i|^2$$

上述两个命题留作课后习题, 这里就不展开具体证明了.

定理 3.9.1 (奇异值分解) 设 $\mathbf{A} \in \mathbb{C}_r^{m \times n}$, 则存在酉矩阵 $\mathbf{U} \in \mathbb{C}^{m \times m}$ 与酉矩阵 $\mathbf{V} \in \mathbb{C}^{n \times n}$ 使得

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \boldsymbol{\Sigma}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{V}^H$$

式中: $\boldsymbol{\Sigma}_r = \text{diag}(\sigma_1, \dots, \sigma_r), \sigma_i (i=1, \dots, r)$ 是矩阵 \mathbf{A} 的正奇异值.

证明: 由于 $\mathbf{A}\mathbf{A}^H$ 为 m 阶正规矩阵, 故存在酉矩阵 \mathbf{U} 使得

$$\mathbf{U}^H (\mathbf{A}\mathbf{A}^H) \mathbf{U} = \text{diag}(\sigma_1^2, \dots, \sigma_r^2, 0, \dots, 0)$$

式中: $\mathbf{U} = [\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_r, \boldsymbol{\alpha}_{r+1}, \dots, \boldsymbol{\alpha}_m], \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_r$ 分别属于特征值 $\sigma_1^2, \dots, \sigma_r^2$ 的单位正交特征向量, $\boldsymbol{\alpha}_{r+1}, \dots, \boldsymbol{\alpha}_m$ 则属于零特征值的 $(m-r)$ 个单位正交特征向量.

由引理 3.9.3 证明过程知, $\mathbf{A}^H \boldsymbol{\alpha}_1, \dots, \mathbf{A}^H \boldsymbol{\alpha}_r$ 分别是矩阵 $\mathbf{A}^H \mathbf{A}$ 属于特征值 $\sigma_1^2, \dots, \sigma_r^2$ 的特征向量. 又知对任意 $i, j=1, \dots, r$, 有

$$(\mathbf{A}^H \boldsymbol{\alpha}_i, \mathbf{A}^H \boldsymbol{\alpha}_j) = \boldsymbol{\alpha}_j^H \mathbf{A} \mathbf{A}^H \boldsymbol{\alpha}_i = \sigma_i^2 \boldsymbol{\alpha}_j^H \boldsymbol{\alpha}_i = \begin{cases} \sigma_i^2, & i=j \\ 0, & i \neq j \end{cases}$$

因此, $\frac{\mathbf{A}^H \boldsymbol{\alpha}_1}{\sigma_1}, \dots, \frac{\mathbf{A}^H \boldsymbol{\alpha}_r}{\sigma_r}$ 是单位正交向量组. 将向量组 $\frac{\mathbf{A}^H \boldsymbol{\alpha}_1}{\sigma_1}, \dots, \frac{\mathbf{A}^H \boldsymbol{\alpha}_r}{\sigma_r}$ 扩充为 \mathbb{C}^n 空间的一组标准正交基, 并记为

$$\frac{\mathbf{A}^H \boldsymbol{\alpha}_1}{\sigma_1}, \dots, \frac{\mathbf{A}^H \boldsymbol{\alpha}_r}{\sigma_r}, \boldsymbol{\beta}_{r+1}, \dots, \boldsymbol{\beta}_n$$

由此, 分别定义 $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$ 和 $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2]$, 其中

$$\begin{aligned} \mathbf{U}_1 &= [\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_r], \quad \mathbf{U}_2 = [\boldsymbol{\alpha}_{r+1}, \dots, \boldsymbol{\alpha}_m] \\ \mathbf{V}_1 &= \left[\frac{\mathbf{A}^H \boldsymbol{\alpha}_1}{\sigma_1}, \dots, \frac{\mathbf{A}^H \boldsymbol{\alpha}_r}{\sigma_r} \right], \quad \mathbf{V}_2 = [\boldsymbol{\beta}_{r+1}, \dots, \boldsymbol{\beta}_n] \end{aligned}$$

显然, $\mathbf{V}_1 = \mathbf{A}^H \mathbf{U}_1 \boldsymbol{\Sigma}_r^{-1}$. 于是,

$$\mathbf{U} \begin{bmatrix} \boldsymbol{\Sigma}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{V}^H = \mathbf{U}_1 \boldsymbol{\Sigma}_r \mathbf{V}_1^H = \mathbf{U}_1 \mathbf{U}_1^H \mathbf{A}$$

由于 $\mathbf{A}\mathbf{U}_2 = \mathbf{O}$, 故 $\mathbf{U}_2 \mathbf{U}_2^H \mathbf{A} = \mathbf{O}$. 因此, $\mathbf{U}_1 \mathbf{U}_1^H \mathbf{A} = (\mathbf{U}_1 \mathbf{U}_1^H + \mathbf{U}_2 \mathbf{U}_2^H) \mathbf{A} = \mathbf{A}$. 证毕.

注 4: 若 $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^H$ 是奇异值分解, 则酉矩阵 \mathbf{U} 的列向量为 $\mathbf{A}\mathbf{A}^H$ 的特征向量, 酉矩阵 \mathbf{V} 的列向量为 $\mathbf{A}^H \mathbf{A}$ 的特征向量.

根据定理 3.9.1 知, 矩阵的奇异值分解步骤可归结如下:

(1) 计算 $\mathbf{A}\mathbf{A}^H$ 的 m 个特征值, 确定对角矩阵 $\boldsymbol{\Sigma}_r$;

(2) 计算 $\mathbf{A}\mathbf{A}^H$ 的 m 个特征向量 $\alpha_1, \dots, \alpha_m$ (其中 $\alpha_1, \dots, \alpha_r$ 须是属于 $\mathbf{A}\mathbf{A}^H$ 非零特征值的特征向量), 将其标准化构成酉矩阵 \mathbf{U} ;

(3) 计算 $\mathbf{A}^H\alpha_1, \dots, \mathbf{A}^H\alpha_r$, 将其扩为 \mathbb{C}^n 的一组标准正交基, 构成酉矩阵 \mathbf{V} .

注 5: 读者也可先尝试构造酉矩阵 \mathbf{V} , 然后再依据 \mathbf{V} 构造酉矩阵 \mathbf{U} , 由此给出奇异值分解证明和相应的分解步骤. 下面用例 3.9.3 进行说明.

例 3.9.3 求矩阵 $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \\ 1 & 0 \end{bmatrix}$ 的奇异值分解.

解: 计算 $\mathbf{A}^H\mathbf{A}$ 得

$$\mathbf{A}^H\mathbf{A} = \begin{bmatrix} 5 & 2 \\ 2 & 5 \end{bmatrix}$$

其特征多项式为 $|\lambda\mathbf{I} - \mathbf{A}^H\mathbf{A}| = (\lambda - 7)(\lambda - 3)$. 显然, $\mathbf{A}^H\mathbf{A}$ 的特征值为 $\lambda_1 = 7, \lambda_2 = 3$; 其单位特征向量分别为

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

令

$$\mathbf{V} = \mathbf{V}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_r = \begin{bmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{3} \end{bmatrix}$$

定义

$$\mathbf{U}_1 = \mathbf{A}\mathbf{V}_1\boldsymbol{\Sigma}_r^{-1} = \begin{bmatrix} \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{14}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{14}} & \frac{1}{\sqrt{6}} \end{bmatrix} = [\mathbf{u}_1, \mathbf{u}_2]$$

并将 \mathbf{U}_1 中列向量 $\mathbf{u}_1, \mathbf{u}_2$ 扩充成 \mathbb{R}^3 中的标准正交基: 设 $\mathbf{u}_3 = [x_1, x_2, x_3]^T \in \mathbb{R}^3$, 由 $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ 是单位正交向量组, 得到如下方程组:

$$\begin{cases} 3x_1 + 2x_2 + x_3 = 0 \\ x_1 - 2x_2 + x_3 = 0 \\ |x_1|^2 + |x_2|^2 + |x_3|^2 = 1 \end{cases}$$

解得 $\mathbf{u}_3 = \frac{1}{\sqrt{21}} [2, -1, -4]^T$.

由此得到酉矩阵 \mathbf{U}

$$U = \begin{bmatrix} \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{21}} \\ \frac{2}{\sqrt{14}} & -\frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{21}} \\ \frac{1}{\sqrt{14}} & \frac{1}{\sqrt{6}} & -\frac{4}{\sqrt{21}} \end{bmatrix}$$

所以, \mathbf{A} 的奇异值分解为

$$\mathbf{A} = \begin{bmatrix} \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{21}} \\ \frac{2}{\sqrt{14}} & -\frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{21}} \\ \frac{1}{\sqrt{14}} & \frac{1}{\sqrt{6}} & -\frac{4}{\sqrt{21}} \end{bmatrix} \begin{bmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{3} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

例 3.9.4 求 $\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$ 的奇异值分解.

错误示范: 首先有

$$\mathbf{A}\mathbf{A}^H = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

计算其特征多项式为 $|\lambda \mathbf{I} - \mathbf{A}\mathbf{A}^H| = (\lambda - 3)(\lambda - 1)$. 由此得到 $\mathbf{A}\mathbf{A}^H$ 的特征值为 $\lambda_1 = 3$, $\lambda_2 = 1$. 分别求出这两个特征值对应的单位特征向量为

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

并定义酉矩阵 $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2]$, $\Sigma_r = \text{diag}(\sqrt{3}, 1)$.

因为 $\mathbf{A}^H\mathbf{A}$ 与 $\mathbf{A}\mathbf{A}^H$ 有相同的正特征值, 所以 $\mathbf{A}^H\mathbf{A}$ 的所有特征值是 $\lambda_1 = 3, \lambda_2 = 1, \lambda_3 = \lambda_4 = 0$. 分别求得各特征值对应的单位特征向量, 并定义酉矩阵 \mathbf{V} :

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4] = \begin{bmatrix} -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{3}} \\ 0 & 0 & 1 & 0 \\ \frac{2}{\sqrt{6}} & 0 & 0 & \frac{1}{\sqrt{3}} \end{bmatrix}$$

此时, $\mathbf{U} \begin{bmatrix} \Sigma_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{V}^H$ 是矩阵 \mathbf{A} 的奇异值分解吗? 经计算, 得

$$U \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix} V^H = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & -1 \end{bmatrix} \neq A$$

显然,这不是 A 的奇异值分解. 请大家思考上述分解出现的问题.

尽管 U 的列向量是矩阵 AA^H 的单位正交特征向量, V 的列向量是 $A^H A$ 的单位正交特征向量,但酉矩阵 U 和 V 的选择还是有一定的相关性: 或者先确定酉矩阵 U , 然后根据 $V_1 = A^H U_1 \Sigma_r^{-1}$ 确定酉矩阵 V 的前 r 列; 或者先确定酉矩阵 V , 然后根据 $U_1 = AV_1 \Sigma_r^{-1}$ 确定酉矩阵 U 的前 r 列. 单独计算酉矩阵 U 和 V 可能导致错误分解.

利用矩阵的奇异值分解可对矩阵作极分解. 具体定理如下:

定理 3.9.2(极分解) 任意复方阵 A 必有如下分解:

$$A = GW \quad (3.9.8)$$

式中: G 为半正定 Hermite 矩阵, W 为酉矩阵. 当矩阵 A 可逆时, G 是正定 Hermite 矩阵, 此时极分解式(3.9.8)唯一.

证明: (1) 对矩阵 A 应用奇异值分解, 得 $A = UDV^H$, 其中, U 和 V 均是酉矩阵, D 是半正定对角矩阵并可表示为

$$D = \begin{bmatrix} \sigma_1 & & O \\ & \ddots & \\ O & & \sigma_r \\ & & O & \\ & & & O \end{bmatrix}$$

式中: $\sigma_i, i=1, \dots, r$ 是矩阵 A 的正奇异值.

进一步, 由 $U^H U = I$ 得,

$$A = UDV^H = UDU^H UV^H$$

定义 $G = UDU^H, W = UV^H$, 则 G 是半正定 Hermite 矩阵, W 为酉矩阵.

(2) 当矩阵 A 是可逆矩阵时, 则 D 为正定对角矩阵, 其对角线元素恰为矩阵 A 的 n 个正奇异值. 此时极分解 $A = GW$ 中, G 是正定 Hermite 矩阵.

现证明极分解的唯一性. 假设矩阵 A 的极分解不唯一, 即

$$A = GW = \tilde{G}\tilde{W}$$

式中: G 和 \tilde{G} 是正定 Hermite 矩阵, W 和 \tilde{W} 是酉矩阵 \tilde{W} .

由于 \tilde{G} 是正定 Hermite 矩阵, 故存在正定对角阵 $\tilde{D} = \text{diag}(\mu_1, \dots, \mu_n)$ 和酉矩阵 \tilde{U} 满足 $\tilde{G} = \tilde{U}\tilde{D}\tilde{U}^H$, 其中正实数 μ_1, \dots, μ_n 是矩阵 \tilde{G} 的 n 个特征值.

定义 $\tilde{V} = \tilde{W}^H \tilde{U}$, 则矩阵 A 可作如下分解:

$$A = \tilde{U}\tilde{D}\tilde{V}^H \quad (3.9.9)$$

于是, $A = UDV^H$ 和 $A = \tilde{U}\tilde{D}\tilde{V}^H$ 是矩阵 A 的两种不同奇异值分解, 且集合 $\{\mu_1, \dots, \mu_n\}$ 和集合 $\{\sigma_1, \dots, \sigma_n\}$ 相等, 即 $\mu_{i_j} = \sigma_i$, 其中 $i, j=1, \dots, n, i_1, \dots, i_n$ 是 $1, \dots, n$ 的一个排列.

将矩阵 \mathbf{A} 的两种不同奇异值分解代入 $\mathbf{A}\mathbf{A}^H$ 得

$$\mathbf{U}\mathbf{D}^2\mathbf{U}^H = \tilde{\mathbf{U}}\tilde{\mathbf{D}}^2\tilde{\mathbf{U}}^H$$

整理上式得

$$\mathbf{D}^2\mathbf{U}^H\tilde{\mathbf{U}} = \mathbf{U}^H\tilde{\mathbf{U}}\tilde{\mathbf{D}}^2$$

令 $\mathbf{U}^H\tilde{\mathbf{U}} = [\tilde{u}_{ij}]_{n \times n}$, 则

$$\begin{bmatrix} \sigma_1^2 & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & \sigma_n^2 \end{bmatrix} \begin{bmatrix} \tilde{u}_{11} & \cdots & \tilde{u}_{1n} \\ \vdots & & \vdots \\ \tilde{u}_{n1} & \cdots & \tilde{u}_{nn} \end{bmatrix} = \begin{bmatrix} \tilde{u}_{11} & \cdots & \tilde{u}_{1n} \\ \vdots & & \vdots \\ \tilde{u}_{n1} & \cdots & \tilde{u}_{nn} \end{bmatrix} \begin{bmatrix} \mu_1^2 & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & \mu_n^2 \end{bmatrix}$$

上式展开整理得

$$\sigma_i^2 \tilde{u}_{ij} = \mu_j^2 \tilde{u}_{ij}, \quad i, j = 1, \cdots, n. \quad (3.9.10)$$

显然, 当 $\sigma_i \neq \mu_j$ 时, 式(3.9.10)成立当且仅当 $\tilde{u}_{ij} = 0$; 否则 $\sigma_i = \mu_j$. 故

$$\sigma_i \tilde{u}_{ij} = \mu_j \tilde{u}_{ij}, \quad i, j = 1, \cdots, n$$

将上式进一步改成矩阵形式为 $\mathbf{D}\mathbf{U}^H\tilde{\mathbf{U}} = \mathbf{U}^H\tilde{\mathbf{U}}\tilde{\mathbf{D}}$. 整理得 $\mathbf{G} = \tilde{\mathbf{G}}$, 并推导出 $\mathbf{W} = \tilde{\mathbf{W}}$. 唯一性得证.

注 6: 分解式(3.9.8)称为左极分解; 若 $\mathbf{A} = \mathbf{V}\mathbf{G}$, 其中 \mathbf{V} 是酉矩阵, \mathbf{G} 是半正定 Hermite 矩阵, 则称分解式 $\mathbf{A} = \mathbf{V}\mathbf{G}$ 为右极分解.

【应用】 奇异值分解在矩阵特征值、广义逆矩阵等矩阵分析和计算方面有着重要应用, 而且在图像处理、机器学习等领域有着广泛应用. 下面以奇异值分解在最小二乘问题和图像压缩的应用为例展开说明.

最小二乘问题 设 $\mathbf{A} \in \mathbb{C}^{m \times n}$, $\mathbf{b} \in \mathbb{C}^m$, 求 $\mathbf{x}^* \in \mathbb{C}^n$ 使得

$$\|\mathbf{A}\mathbf{x}^* - \mathbf{b}\| = \min_{\mathbf{x} \in \mathbb{C}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|$$

一般而言, $\mathbf{b} \notin R(\mathbf{A})$, 否则必存在 $\mathbf{x}^* \in \mathbb{C}^n$ 使得 $\mathbf{A}\mathbf{x}^* = \mathbf{b}$.

设矩阵 \mathbf{A} 的奇异值分解为 $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^H$, 则

$$\mathbf{A}\mathbf{x} - \mathbf{b} = \mathbf{U}\mathbf{D}\mathbf{V}^H\mathbf{x} - \mathbf{b} = \mathbf{U}(\mathbf{D}\mathbf{V}^H\mathbf{x} - \mathbf{U}^H\mathbf{b}) \quad (3.9.11)$$

令 $\mathbf{y} = \mathbf{V}^H\mathbf{x}$, $\mathbf{c} = \mathbf{U}^H\mathbf{b}$, 则式(3.9.11)可写为

$$\mathbf{A}\mathbf{x} - \mathbf{b} = \mathbf{U}(\mathbf{D}\mathbf{y} - \mathbf{c}) \quad (3.9.12)$$

由于 \mathbf{U} 是酉矩阵, 不改变向量的长度. 对式(3.9.12)两边取长度得

$$\|\mathbf{A}\mathbf{x} - \mathbf{b}\| = \|\mathbf{D}\mathbf{y} - \mathbf{c}\| \quad (3.9.13)$$

进一步, 定义

$$\mathbf{D} = \begin{bmatrix} \sigma_1 & & & & \mathbf{O} \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ \mathbf{O} & & & & \ddots \\ & & & & & 0 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}$$

则有 $D\mathbf{y} - \mathbf{c} = [\sigma_1 y_1 - c_1, \dots, \sigma_r y_r - c_r, -c_{r+1}, \dots, -c_m]^T$, 进而有

$$\|\mathbf{Ax} - \mathbf{b}\|^2 = \sum_{i=1}^r (\sigma_i y_i - c_i)^2 + \sum_{i=r+1}^m c_i^2 \quad (3.9.14)$$

显然, 当 $\sigma_1 y_1 - c_1 = \dots = \sigma_r y_r - c_r = 0$ 时, $\|\mathbf{Ax} - \mathbf{b}\|$ 达到最小. 由此可解出

$$y_i = \frac{c_i}{\sigma_i}, \quad i = 1, \dots, r \quad (3.9.15)$$

而 $y_i (i=r+1, \dots, n)$ 为任意复数. 再由 $\mathbf{y} = \mathbf{V}^H \mathbf{x}$ 可解得 $\mathbf{x} = \mathbf{Vy}$.

图像压缩 假设一副图像有 $m \times n$ 个像素, 若将 mn 个数据一起传送, 往往数据量太大. 现希望传送尽可能少的数据, 并且在接收端还能利用这些数据重构原图像. 为研究这一问题, 不妨用矩阵 $\mathbf{A} \in \mathbb{C}_r^{m \times n}$ 表示待传送的图像, 其中

$$\mathbf{A} = \begin{bmatrix} f(1,1) & f(1,2) & \cdots & f(1,n) \\ f(2,1) & f(2,2) & \cdots & f(2,n) \\ \vdots & \vdots & & \vdots \\ f(m,1) & f(m,2) & \cdots & f(m,n) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$f(x, y)$ 表示点 (x, y) 处图像的灰度或强度.

设矩阵 \mathbf{A} 的奇异值分解为

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & \mathbf{O} & & & \ddots \\ & & & & & 0 \end{bmatrix} \mathbf{V}^H$$

式中, $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_m]$, $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n]^H$, 且 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$.

由此,

$$\begin{aligned} \mathbf{A} &= [\mathbf{u}_1, \dots, \mathbf{u}_m] \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & \mathbf{O} & & & \ddots \\ & & & & & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^H \\ \vdots \\ \mathbf{v}_n^H \end{bmatrix} \\ &= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^H + \dots + \sigma_r \mathbf{u}_r \mathbf{v}_r^H \end{aligned}$$

若从中选择 $k (k < r)$ 个大奇异值以及这些奇异值对应的左右向量逼近原图像, 即

$$\hat{\mathbf{A}} = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^H \quad (3.9.16)$$

则可利用 $\hat{\mathbf{A}}$ “近似” \mathbf{A} 以实现图像的压缩. 当然, 若 k 值偏小, 则重构的图像质量可能不满意; 反之, 过大的 k 值又降低图像压缩和传送的效率. 因此, 需要根据事情选择合适的 k 值以兼顾图像传送效率和重构质量. 图 3.9.1 给出了利用奇异值分解的一个示例, 其中图(a)为原图, 图(b)为 $k=20$ 的图像重构, 图(c)为 $k=50$ 的图像重构, 图(d)为 $k=100$ 的图像重构.

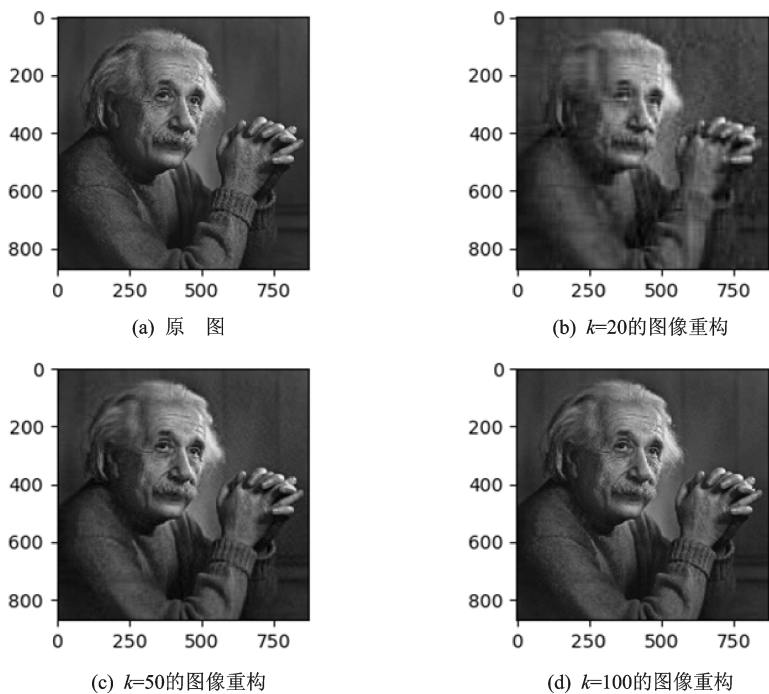


图 3.9.1 利用奇异值分解实例

【思考】 若已知矩阵 $B = \begin{bmatrix} 1 & \epsilon \\ -1 & \epsilon \end{bmatrix}$ 和 $C = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$, 如何求矩阵乘积 BC 的奇异值分解?

这是我们在应用奇异值分解时常会遇到的问题. 我们自然的想法是先求出矩阵乘积 BC , 然后再对 BC 进行奇异值分解. 在实际计算中, 若 $|\epsilon|$ 选取小于或大于计算截断误差时, 会发现 BC 的奇异值明显发生变化. 请读者思考并提出解决办法.

本章习题

1. 设 A_1, A_2, \dots, A_m 均为 n 阶复方阵, 且 $A_1 A_2 \cdots A_m = O, m \in \mathbb{N}$. 证明 $\text{rank}(A_1) + \text{rank}(A_2) + \cdots + \text{rank}(A_m) \leq (m-1)n$
2. 已知 $A \in \mathbb{C}^{m \times n}$, 求满足矩阵方程 $AXA = A$ 的所有 X , 并说明 X 唯一性条件.

(矩阵 A 的减号逆)

3. 设矩阵 $AB=BA=O$, 证明存在可逆矩阵 P, Q 使得

$$A = P \begin{bmatrix} I_m & O \\ O & O \end{bmatrix} Q, \quad B = Q^{-1} \begin{bmatrix} O & O \\ O & I_n \end{bmatrix} P^{-1}$$

4. 分别求满足条件(1)、(2)的矩阵 $A \in \mathbb{C}^{n \times n}$, 其中

(1) 幂等矩阵 $A^2=A$;

(2) 幂零矩阵 $A^2=O$.

5. 求下列矩阵的满秩分解:

$$(1) \begin{bmatrix} 0 & 1 & -2 & 4 & -i \\ 0 & -1 & 2 & -4 & i \\ 0 & -1 & 1+i & -3 & -i \end{bmatrix};$$

$$(2) \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 0 & 1 \end{bmatrix}.$$

6. 已知 $A \in \mathbb{C}^{m \times n}$, 利用满秩分解求满足矩阵方程 $AXA=A$ 的一类 X . (矩阵 A 的减号逆)

7. 设 $A \in \mathbb{C}_r^{m \times n}$, 利用满秩分解证明 $\text{rank}(A) = \text{rank}(AA^H) = \text{rank}(A^HA)$.

8. 证明定理 3.2.3.

9. 设 $A \in \mathbb{C}_{r_1}^{m \times n}$, $B \in \mathbb{C}_{r_2}^{n \times m}$, 则 AB 和 BA 具有完全相同的非零特征值.

10. 试编程(C语言或 MATLAB)矩阵的 LU 分解算法.

11. 利用 LU 分解求解方程组

$$\begin{cases} 3x_1 + 3x_2 + 4x_3 = 2 \\ x_1 + x_2 + 9x_3 = -7 \\ x_1 + 2x_2 - 6x_3 = 9 \end{cases}$$

试给出 n 阶正定实对称矩阵的 Cholesky 算法, 并用编程实现.

12. 平板稳态传热问题的解可近似用非齐次线性方程 $Ax=b$ 的解来逼近, 求该方程的解, 其中 $b = [5, 15, 0, 10, 0, 10, 20, 30]^T$, 矩阵 A 为

$$A = \begin{bmatrix} 4 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 4 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 4 & -1 & -1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 4 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 4 & -1 & -1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 4 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 4 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & 4 \end{bmatrix}$$

13. 设 A, B 均为 n 阶复方阵, 若存在复数 λ 和非零向量 $x \in \mathbb{C}^n$ 使得 $Ax = \lambda Bx$, 则称 λ 为广义特征值问题 $Ax = \lambda Bx$ 的特征值, 非零向量 x 称为对应于特征值 λ 特征向量. 在振动理论等应用问题中, A 和 B 均为 Hermite 矩阵, 且 B 是正定矩阵, 试求广义特征值问题 $Ax = \lambda Bx$ 的特征值与特征向量.

14. 试编程(C语言或 MATLAB)矩阵的 QR 分解算法.

15. 求 $A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix}$ 的 QR 分解.

16. 利用 QR 分解求解方程组

$$\begin{cases} 3x_1 + 3x_2 + 4x_3 = 2 \\ x_1 + x_2 + 9x_3 = -7 \\ x_1 + 2x_2 - 6x_3 = 9 \end{cases}$$

17. 利用 QR 分解解决优化问题: $\min_{b \in R(A)} \|Ax - b\|_2$, 其中, 矩阵 $A \in \mathbb{C}^{m \times n}$ ($n \leq m$) 和向量 $b \in \mathbb{C}^m$ 已知, 求向量 x .

18. 证明 QR 算法产生的矩阵序列 $\{A_k, k=1, 2, \dots\}$ 满足以下两个条件:

(1) $A_{k+1} = (Q_1 \cdot Q_2 \cdot \dots \cdot Q_k)^T A (Q_1 \cdot Q_2 \cdot \dots \cdot Q_k)$;

(2) $A^k = (Q_1 \cdot Q_2 \cdot \dots \cdot Q_k) (R_k \cdot R_{k-1} \cdot \dots \cdot R_1)$.

19. 证明秩为 1 的矩阵 A 的最小多项式是 $m_A(\lambda) = \lambda^2 - \text{tr}(A)\lambda$.

20. 求矩阵的最小多项式:

(1) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$; (2) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$; (3) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$.

21. 求题 20 中矩阵的逆矩阵.

22. 设 $A \in \mathbb{C}^{n \times n}$ 是首 1 多项式 $f(\lambda)$ 的友矩阵, 证明 $m_A(\lambda) = f(\lambda)$.

23. 已知矩阵 $A = \begin{bmatrix} 0 & 0 & 1 \\ x & 1 & y \\ 1 & 0 & 0 \end{bmatrix}$ 是单纯矩阵, 求 x 与 y .

24. 设 n 阶复方阵 A, B 是单纯矩阵, 且 $AB = BA$. 证明存在同一个 n 阶矩阵 P 使得 $P^{-1}AP$ 和 $P^{-1}BP$ 同时为对角阵.

25. 设复方阵 A 分别满足如下条件:

(1) $A^2 + A = 2I$;

(2) $A^k = I$;

(3) $A^k = O$ 且 $A \neq O$;

式中, k 为自然数. 试判断 A 是否为单纯矩阵.

26. 设 A 是 n 阶 Hermite 矩阵, 证明以下表达等价:

- (1) \mathbf{A} 是正定矩阵;
- (2) 对于任意 n 阶可逆矩阵 \mathbf{P} , 矩阵 $\mathbf{P}^H \mathbf{A} \mathbf{P}$ 正定;
- (3) \mathbf{A} 的 n 个特征值均为正数;
- (4) 存在 n 阶可逆矩阵 \mathbf{P} 使得 $\mathbf{P}^H \mathbf{A} \mathbf{P} = \mathbf{I}$;
- (5) 存在 n 阶可逆矩阵 \mathbf{Q} 使得 $\mathbf{A} = \mathbf{Q}^H \mathbf{Q}$;
- (6) 存在 n 阶可逆 Hermite 矩阵 \mathbf{G} 使得 $\mathbf{A} = \mathbf{G}^2$.

27. 设 $\mathbf{A} = [a_{ij}] \in \mathbb{C}^{n \times n}$, 证明 \mathbf{A} 是正规矩阵当且仅当 $\sum_{i,j=1}^n |a_{ij}|^2 = \sum_{i=1}^n |\lambda_i|^2$.

28. 设 $\mathbf{A} = [a_{ij}] \in \mathbb{C}^{n \times n}$, 证明 \mathbf{A} 是正规矩阵当且仅当 $\mathbf{A} + k\mathbf{I}$ 是正规矩阵, 其中 $k \in \mathbb{C}$.

29. 求 $\mathbf{A} = \begin{bmatrix} -2i & 4 & -2 \\ -4 & -2i & -2i \\ 2 & -2i & -5i \end{bmatrix}$ 的谱分解.

30. 求 $\mathbf{A} = \begin{bmatrix} 1 & 4 & 2 \\ 0 & -3 & 4 \\ 0 & 4 & 3 \end{bmatrix}$ 的谱分解并计算 \mathbf{A}^{100} .

31. 求矩阵 $\mathbf{A}(\lambda)$ 的行列式因子和 Smith 标准形 (利用两种不同方法求 Smith 标准形)

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

32. 求矩阵 $\mathbf{A}(\lambda)$ 的初等因子, 其中

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

33. 设 $\mathbf{A} \in \mathbb{C}^{7 \times 7}$, 且其特征矩阵 $\lambda\mathbf{I} - \mathbf{A} \cong \text{diag}((\lambda+1)^2, 1, (\lambda-2)^2, (\lambda-2)(\lambda+1), (\lambda-2), 1, 1)$, 求 $\lambda\mathbf{I} - \mathbf{A}$ 的初等因子、不变因子及 Smith 标准形、矩阵 \mathbf{A} 的最小多项式.

34. 实方矩阵 \mathbf{A} 和它的转置相似吗? 请说明理由.

35. 求 $\mathbf{A}^{100} - 2\mathbf{A}^{50}$, 其中

$$\mathbf{A} = \begin{bmatrix} 3 & 0 & 8 \\ 3 & 1 & 6 \\ -2 & 0 & 5 \end{bmatrix}$$

36. 求 $\mathbf{A} = \begin{bmatrix} -4 & 2 & 10 \\ -4 & 3 & 7 \\ -3 & 1 & 7 \end{bmatrix}$ 和 $\mathbf{B} = \begin{bmatrix} 2 & -1 & -1 \\ 2 & -1 & -2 \\ -1 & 1 & 2 \end{bmatrix}$ 的 Jordan 标准形及变换

矩阵.

37. 设 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ 和 $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$ 分别是 n 阶正规矩阵 \mathbf{A} 的奇异值和特征值, 则 $\sigma_i = |\lambda_i|, i=1, \cdots, n$.

38. 设 σ_i 和 $\lambda_i (i=1, \cdots, n)$ 分别是 n 阶复方阵 \mathbf{A} 的奇异值和特征值, 则

$$\sum_{i=1}^n |\lambda_i|^2 \leq \sum_{i=1}^n |\sigma_i|^2$$

39. 设 \mathbf{A} 是 n 阶实矩阵, 证明 $\text{rank}(\mathbf{A} - i\mathbf{I}) = \text{rank}(\mathbf{A} + i\mathbf{I})$.

40. 设 \mathbf{A} 是 n 阶矩阵, 证明

(1) 存在正整数 m 使得 $\text{rank}(\mathbf{A}^m) = \text{rank}(\mathbf{A}^{m+1})$;

(2) 对任意正整数 k , $\text{rank}(\mathbf{A}^m) = \text{rank}(\mathbf{A}^{m+j})$, 其中 m 由(1)确定.

41. 证明: 若 \mathbf{A} 与 \mathbf{B} 酉相抵, 则 \mathbf{A} 与 \mathbf{B} 有相同的奇异值.

42. 求下列矩阵的奇异值分解:

$$(1) \mathbf{A} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}; (2) \mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

43. 设矩阵 \mathbf{A} 是 Hermite 矩阵, 则以下等价命题:

(1) 矩阵 \mathbf{A} 是半正定的;

(2) 对于任意 n 阶可逆矩阵 \mathbf{P} , $\mathbf{P}^H \mathbf{A} \mathbf{P}$ 是半正定 Hermite 矩阵;

(3) Hermite 矩阵 \mathbf{A} 的 n 个特征值均为非负数;

(4) 存在 n 阶可逆矩阵 \mathbf{P} 使得 $\mathbf{P}^H \mathbf{A} \mathbf{P} = \begin{bmatrix} \mathbf{I}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}$, 其中 $r = \text{rank}(\mathbf{A})$;

(5) 存在秩为 $\text{rank}(\mathbf{A})$ 的矩阵 \mathbf{Q} 使得 $\mathbf{A} = \mathbf{Q}^H \mathbf{Q}$;

(6) 存在 n 阶 Hermite 矩阵 \mathbf{G} 使得 $\mathbf{A} = \mathbf{G}^2$;

(7) Hermite 矩阵 \mathbf{A} 的所有顺序主子式均非负.

第 4 章

矩阵分析

在初等代数中,我们定义了复数的模,并用模这个度量表示复数的大小(或表示几何意义下的长度),或依此考察复数间的距离.对于 n 维线性空间,我们在定义了内积这一概念后,向量就有了长度、夹角以及距离等度量概念,这是基于三维现实空间中相应概念的推广.本章将进一步把向量长度的概念推广至范数.

4.1 向量范数

定义 4.1.1(向量范数) 设 V 是数域 F 上的线性空间, $\|x\|$ 是以 $x \in V$ 为自变量的实值函数,若满足以下三条性质:

- (1) 正定性(或非负性): $\|x\| \geq 0$, $\|x\| = 0$ 当且仅当 $x = \theta$;
- (2) 齐次性: $\forall k \in F, x \in V, \|kx\| = |k| \|x\|$;
- (3) 三角不等式: $\forall x, y \in V, \|x + y\| \leq \|x\| + \|y\|$,

则称 $\|x\|$ 是向量 x 的范数, V 是数域 F 上的赋范线性空间,记为 $(V, \|\cdot\|)$.

注 1: 定义 4.1.1 中的三条性质称为范数的三条公理,即只要满足范数公理的实值函数均可定义为向量的范数.

注 2: 三角不等式的等价形式:

$$\forall x, y \in V, \quad |\|x\| - \|y\|| \leq \|x - y\|$$

例 4.1.1 $\forall x = [x_1, x_2, \dots, x_n]^T \in \mathbb{C}^n$, 定义

$$\|x\|_1 = \sum_{i=1}^n |x_i| \quad (4.1.1)$$

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i| \quad (4.1.2)$$

$$\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}} \quad (4.1.3)$$

则 $\|x\|_1$ 、 $\|x\|_\infty$ 和 $\|x\|_2$ 均是向量 x 的范数, 分别称为向量 x 的 1-范数、 ∞ -范数和 2-范数(或欧几里得范数).

证明: 1. 证明 $\|x\|_1$ 是向量 x 的范数.

(1) 正定性: $\|x\|_1 = \sum_{i=1}^n |x_i| \geq 0$, 且 $\|x\|_1 = 0$ 当且仅当 $x = \theta$;

(2) 齐次性: $\forall k \in \mathbb{C}, \|kx\|_1 = \sum_{i=1}^n |kx_i| = \sum_{i=1}^n |k| |x_i| = |k| \|x\|_1$;

(3) 三角不等式: $\forall x = [x_1, x_2, \dots, x_n]^T$ 和 $y = [y_1, y_2, \dots, y_n]^T \in \mathbb{C}^n$, 则

$$\|x + y\|_1 = \sum_{i=1}^n |x_i + y_i| \leq \sum_{i=1}^n |x_i| + \sum_{i=1}^n |y_i| = \|x\|_1 + \|y\|_1$$

因此, $\|x\|_1$ 是向量 x 的范数.

2. 证明 $\|x\|_\infty$ 是向量 x 的范数.

(1) 正定性: $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i| \geq 0$, 且 $\|x\|_\infty = 0$ 当且仅当 $x = \theta$;

(2) 齐次性: $\forall k \in \mathbb{C}, \|kx\|_\infty = \max_{1 \leq i \leq n} |kx_i| = |k| \max_{1 \leq i \leq n} |x_i| = |k| \|x\|_\infty$;

(3) 三角不等式: $\forall x = [x_1, x_2, \dots, x_n]^T$ 和 $y = [y_1, y_2, \dots, y_n]^T \in \mathbb{C}^n$, 则

$$\|x + y\|_\infty = \max_{1 \leq i \leq n} |x_i + y_i| \leq \max_{1 \leq i \leq n} |x_i| + \max_{1 \leq i \leq n} |y_i| = \|x\|_\infty + \|y\|_\infty$$

因此, $\|x\|_\infty$ 是向量 x 的范数.

3. 证明 $\|x\|_2$ 是向量 x 的范数.

(1) 正定性: $\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}} \geq 0$, 且 $\|x\|_2 = 0$ 当且仅当 $x = \theta$;

(2) 齐次性:

$$\forall k \in \mathbb{C}, \|kx\|_2 = \left(\sum_{i=1}^n |kx_i|^2 \right)^{\frac{1}{2}} = \left(\sum_{i=1}^n |k|^2 |x_i|^2 \right)^{\frac{1}{2}} = |k| \|x\|_2$$

(3) 三角不等式: $\forall x = [x_1, x_2, \dots, x_n]^T$ 和 $y = [y_1, y_2, \dots, y_n]^T \in \mathbb{C}^n$, 由 Cauchy 不等式得

$$\begin{aligned} \|x + y\|_2^2 &= \|x\|_2^2 + (x, y) + (y, x) + \|y\|_2^2 \\ &\leq \|x\|_2^2 + 2\|x\|_2\|y\|_2 + \|y\|_2^2 = (\|x\|_2 + \|y\|_2)^2 \end{aligned}$$

因此, $\|x\|_2$ 是向量 x 的范数.

例 4.1.1 表明同一线性空间可定义多个向量范数.

例 4.1.2 设向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 是 n 维线性空间 V 的一组基, V 中任意向量 α 在这组基下的坐标为 $x = [x_1, x_2, \dots, x_n]^T$. 由此, 可定义向量 α 的范数为

$$\|\alpha\| \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^n |x_i|^2} \quad (4.1.4)$$

证明: 由式(4.1.4)知,

(1) 正定性: $\|\alpha\| \geq 0$, 且 $\|\alpha\| = 0$ 当且仅当 $\alpha = \theta$;

(2) 齐次性: $\forall k \in F, \|k\alpha\| = \sqrt{\sum_{i=1}^n |kx_i|^2} = |k| \|\alpha\|$;

(3) 三角不等式: 设 V 中任意向量 α 和 β 在基 $\alpha_1, \dots, \alpha_n$ 下的坐标分别为 $x = [x_1, x_2, \dots, x_n]^T$ 和 $y = [y_1, y_2, \dots, y_n]^T$, 则

$$\begin{aligned}\|\alpha + \beta\| &= \sqrt{\sum_{i=1}^n |x_i + y_i|^2} \\ &\leq \sqrt{\sum_{i=1}^n |x_i|^2} + \sqrt{\sum_{i=1}^n |y_i|^2} = \|\alpha\| + \|\beta\|\end{aligned}$$

因此, 式(4.1.4)是向量 α 的范数.

【思考】 式(4.1.4)定义的范数是向量 2-范数吗?

分析: 尽管从形式上看式(4.1.4)定义的范数和向量 2-范数基本一致, 但这两种范数并不相同. 向量 2-范数是针对 \mathbb{C}^n 或 \mathbb{R}^n 空间中的向量而定义的范数, 或理解为向量 α 的 2-范数是基于向量 α 在 \mathbb{C}^n 或 \mathbb{R}^n 空间中的标准基下的坐标 x 定义的(从这个角度看, 向量的 2-范数式是式(4.1.4)的特例); 而式(4.1.4)则不限定线性空间, 只根据向量在一组特定基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 下的坐标 x 进行定义.

例 4.1.3 设 $1 \leq p \leq \infty$, 对任意向量 $x = [x_1, x_2, \dots, x_n]^T \in \mathbb{C}^n$, 定义

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \quad (4.1.5)$$

则 $\|x\|_p$ 是向量 x 的范数, 称为 p -范数.

证明: 根据向量范数的定义进行证明.

(1) 正定性显然成立;

(2) 齐次性: 对任意 $k \in \mathbb{C}$ 和 $x = [x_1, x_2, \dots, x_n]^T \in \mathbb{C}^n$, 有

$$\|kx\|_p = \left(\sum_{i=1}^n |kx_i|^p \right)^{\frac{1}{p}} = |k| \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} = |k| \|x\|_p$$

(3) 三角不等式: 当 $p=1$ 时, 三角不等式显然成立. 设 $p>1$, 并定义 $q = \frac{p}{p-1}$, 则 $q>1$ 且 $\frac{1}{p} + \frac{1}{q} = 1$.

再令 $x = [x_1, x_2, \dots, x_n]^T \in \mathbb{C}^n$ 和 $y = [y_1, y_2, \dots, y_n]^T \in \mathbb{C}^n$, 则

$$\begin{aligned}\sum_{i=1}^n |x_i + y_i|^p &= \sum_{i=1}^n |x_i + y_i| \cdot |x_i + y_i|^{p-1} \\ &\leq \sum_{i=1}^n |x_i| \cdot |x_i + y_i|^{p-1} + \sum_{i=1}^n |y_i| \cdot |x_i + y_i|^{p-1}\end{aligned}$$

利用 Holder 不等式, 上式可改写为

$$\sum_{i=1}^n |x_i + y_i|^p \leq \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^n |x_i + y_i|^{(p-1)q} \right)^{\frac{1}{q}} +$$

$$\begin{aligned}
 & \left(\sum_{i=1}^n |y_i|^p \right)^{\frac{1}{p}} \cdot \left(\sum_{i=1}^n |x_i + y_i|^{(p-1)q} \right)^{\frac{1}{q}} \\
 &= \left[\left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^n |y_i|^p \right)^{\frac{1}{p}} \right] \cdot \left(\sum_{i=1}^n |x_i + y_i|^p \right)^{\frac{1}{q}}
 \end{aligned}$$

因此,

$$\left(\sum_{i=1}^n |x_i + y_i|^p \right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^n |y_i|^p \right)^{\frac{1}{p}}$$

即 $\|\mathbf{x} + \mathbf{y}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p$. 综上知, $\|\mathbf{x}\|_p$ 是 \mathbf{x} 的范数.

注 3: 当 $p=1, 2$ 时, $\|\mathbf{x}\|_p$ 分别是例 4.1.1 中的 1-范数和 2-范数. 当 $p=\infty$ 时, $\lim_{p \rightarrow \infty} \|\mathbf{x}\|_p = \|\mathbf{x}\|_\infty$. 当 $0 < p < 1$ 时, 此时 $\|\mathbf{x}\|_p$ 不满足三角不等式.

注 4: 范数证明常涉及两个重要的不等式, 分别为

(1) Holder 不等式: 设 $p, q > 1$ 且 $\frac{1}{p} + \frac{1}{q} = 1$, 则对任意 $\mathbf{x} = [x_1, x_2, \dots, x_n]^T \in \mathbb{C}^n$ 和 $\mathbf{y} = [y_1, y_2, \dots, y_n]^T \in \mathbb{C}^n$, 有

$$\sum_{i=1}^n |x_i y_i| \leq \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^n |y_i|^q \right)^{\frac{1}{q}}$$

(2) Minkowski 不等式: $\forall \mathbf{x} = [x_1, x_2, \dots, x_n]^T \in \mathbb{C}^n, \mathbf{y} = [y_1, y_2, \dots, y_n]^T \in \mathbb{C}^n$ 以及 $p \geq 1$, 则

$$\left(\sum_{i=1}^n |x_i + y_i|^p \right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^n |y_i|^p \right)^{\frac{1}{p}}$$

【思考】 内积空间是赋范线性空间吗?

分析: 判断内积空间是否为赋范线性空间只需判断内积空间是否定义了符合范数三条公理的实值函数. 显然, 内积空间定义的向量长度符合这一要求. 因此, 内积空间是赋范线性空间. 实际上, 内积空间的长度定义不仅满足范数的三条公理, 还满足平行四边形法则; 而赋范线性空间并不一定满足这一法则.

例 4.1.4 考察空间 \mathbb{R}^2 , 若定义

$$\|\mathbf{x}\|_p = \sqrt[p]{|x_1|^p + |x_2|^p}$$

式中, $\mathbf{x} = [x_1, x_2]^T \in \mathbb{R}^2, p \geq 1$. 由例 4.1.3 知, 在赋范线性空间 $(\mathbb{R}^2, \|\cdot\|_p)$ 中, $\|\mathbf{x}\|_p$ 是一个范数. 当 $p=2$ 时, $\|\mathbf{x}\|_2$ 满足平行四边形法则, 由此可在 \mathbb{R}^2 中定义一个内积, 使得由它定义的长度正是 \mathbb{R}^2 中的 $\|\cdot\|_2$ (这是泛函中 Fréchet-Jordan-Von Neumann 定理). 当 $p \neq 2$ 时, 如果 \mathbb{R}^2 空间可定义内积, 使得对任意 $\mathbf{x} \in \mathbb{R}^2, (\mathbf{x}, \mathbf{x}) = \|\mathbf{x}\|_p^2$, 那么对 $\mathbf{x} = [x_1, x_2]^T$ 和 $\mathbf{y} = [y_1, y_2]^T$, 应有

$$\begin{aligned}
 \|\mathbf{x} + \mathbf{y}\|_p^2 + \|\mathbf{x} - \mathbf{y}\|_p^2 &= (|x_1 + y_1|^p + |x_2 + y_2|^p)^{\frac{2}{p}} + \\
 &\quad (|x_1 - y_1|^p + |x_2 - y_2|^p)^{\frac{2}{p}} \\
 &= 2(|x_1|^p + |x_2|^p)^{\frac{2}{p}} + 2(|y_1|^p + |y_2|^p)^{\frac{2}{p}}
 \end{aligned}$$

$$= 2(\|x\|_p^2 + \|y\|_p^2)$$

当 $x=e_1, y=e_2$ 时, 上式应有 $2^{\frac{2}{p}} + 2^{\frac{2}{p}} = 4$. 解得 $p=2$. 显然当 $p \neq 2$ 时, 上式显然不成立, 即当 $p \neq 2$ 时, 线性赋范空间 \mathbb{R}^2 中, 均没有内积.

例 4.1.5 设 $A \in \mathbb{C}^{n \times n}$ 是正定 Hermite 矩阵, 对任意向量 $x \in \mathbb{C}^n$, 定义 $\|x\|_A = \sqrt{x^H A x}$, 则 $\|x\|_A$ 是向量范数, 常称为加权范数或椭圆范数.

证明: 由 Cholesky 分解知, 存在可逆矩阵 W 使得 $A = W^H W$. 因此,

$$\|x\|_A = \sqrt{x^H A x} = \|Wx\|_2 \quad (4.1.6)$$

由式(4.1.6)容易证明正定性和齐次性. 下面证明三角不等式:

$$\|x+y\|_A = \|W(x+y)\|_2 \leq \|Wx\|_2 + \|Wy\|_2 = \|x\|_A + \|y\|_A$$

综上所述, $\|x\|_A$ 是向量范数.

注 5: 例 4.1.5 提供了一种借助已有向量范数构造新的向量范数的方法. 实际上, 我们可利用定义式(4.1.6)构造更一般的向量范数(见本章的习题 2).

【应用】 系统稳定性分析. 我们常采用二次型 Lyapunov 函数

$$V(x) = x^H P x = \|x\|_P^2 \quad (4.1.7)$$

研究线性系统乃至非线性系统的稳定性, 其中, P 是正定实对称矩阵.

【应用】 模式分类. 模式分类常通过已知类型属性的观测样本判断未知样本的类型属性, 其基本思想是根据未知样本的模式向量 x 与已知样本的模式向量 s_1, \dots, s_M 相似度来判断未知样本的类型属性. 最简单的方法就是用向量间的距离来度量模式向量的相似度, 比如经典的欧几里得距离, 就是采用向量的 2-范数:

$$D(x, s_i) = \|x - s_i\|_2 = \sqrt{(x - s_i)^H (x - s_i)} \quad (4.1.8)$$

【思考】 范数的本质是什么?

定理 4.1.1 线性空间 V 中任一范数 $\|x\|$ 都是其坐标的连续函数.

证明: 设 V 是 n 维线性空间, ζ_1, \dots, ζ_n 是 V 中一组基, 则对任意向量 x, y , 有

$$x = \xi_1 \zeta_1 + \dots + \xi_n \zeta_n = [\zeta_1, \dots, \zeta_n] \xi$$

$$y = \eta_1 \zeta_1 + \dots + \eta_n \zeta_n = [\zeta_1, \dots, \zeta_n] \eta$$

式中: $\xi = [\xi_1, \dots, \xi_n]^T$ 和 $\eta = [\eta_1, \dots, \eta_n]^T$ 分别是向量 x 和 y 在基 ζ_1, \dots, ζ_n 下的坐标.

由 $\|x\| = \|[\zeta_1, \dots, \zeta_n] \xi\|$ 知, 任一范数 $\|x\|$ 都是向量 x 坐标的函数, 故

$$\begin{aligned} |\|x\| - \|y\|| &\leq \|x - y\| = \left\| \sum_{i=1}^n (\xi_i - \eta_i) \zeta_i \right\| \\ &\leq \sum_{i=1}^n |(\xi_i - \eta_i)| \|\zeta_i\| \leq k \left(\sum_{i=1}^n |(\xi_i - \eta_i)|^2 \right)^{\frac{1}{2}} \end{aligned}$$

其中 $k = \left(\sum_{i=1}^n \|\zeta_i\|^2 \right)^{\frac{1}{2}}$ 是正常数.

当 $\xi_i \rightarrow \eta_i (i=1, \dots, n)$ 时, $|\|x\| - \|y\|| \rightarrow 0$, 即 $\|x\| \rightarrow \|y\|$. 因此, $\|x\|$ 是其坐标 ξ

的连续函数. 证毕.

例 4.1.6 画出 \mathbb{R}^2 空间不同范数下的“单位圆”.

解: 图 4.1.1 给出了 $\|x\|_1=1$, $\|x\|_2=1$, $\|x\|_5=1$ 和 $\|x\|_\infty=1$.

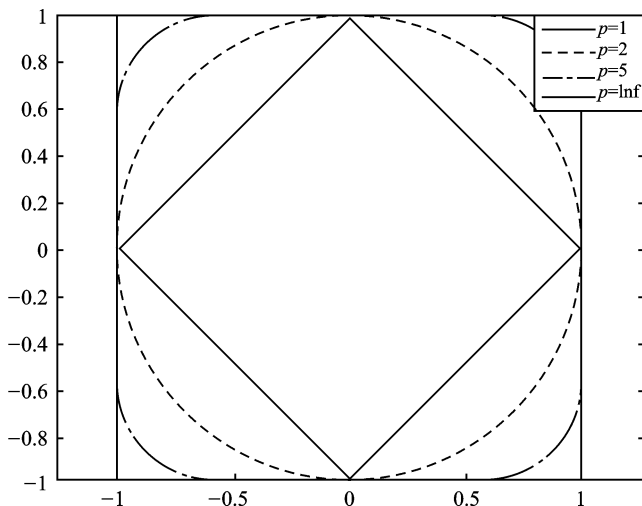


图 4.1.1 \mathbb{R}^2 空间的“单位圆”

从图 4.1.1 看出, 同一空间不同范数定义的单位圆是有区别的: 1-范数和 ∞ -范数的“单位圆”是正方形, 2-范数是单位圆. 尽管范数定义不同, 但其定义的“单位圆”均是有界闭集.

注 6: 有限维赋范线性空间中的“单位圆”具有重要意义. 在几何上, 它们相当于实数轴上的单位闭区间或其端点, 或三维空间中的单位球或球面. 因此, 它们都是有界闭集. 连续的函数在有界闭集上一定有最大值和最小值, 这是高等数学课程中的一个重要结论. 研究赋范线性空间上的连续函数或变换的一个重要技巧就是设法将函数的定义域限制或转移到“单位圆”上.

定义 4.1.2 (范数等价) 设 V 是数域 F 上的有限维线性空间, $\|x\|_\alpha$ 和 $\|x\|_\beta$ 是 V 中任意两个向量范数. 若存在正数 k_1 和 k_2 使得 $\forall x \in V$, 都有

$$k_1 \|x\|_\beta \leq \|x\|_\alpha \leq k_2 \|x\|_\beta$$

则称 $\|x\|_\alpha$ 与 $\|x\|_\beta$ 是等价的.

定理 4.1.2 有限维线性空间中的任意向量范数都是等价的.

证明: 设向量组 ζ_1, \dots, ζ_n 是线性空间 V 中一组基, 则对 V 中任意向量 x , 有

$$x = \xi_1 \zeta_1 + \dots + \xi_n \zeta_n = [\zeta_1, \dots, \zeta_n] \xi$$

式中: $\xi = [\xi_1, \dots, \xi_n]^T$ 是向量 x 在基 ζ_1, \dots, ζ_n 下的坐标.

令 $\|x\|_\alpha$ 与 $\|x\|_\beta$ 是 V 中两个不同范数. 当 $x = \theta$ 时, $k_1 \|x\|_\beta \leq \|x\|_\alpha \leq k_2 \|x\|_\beta$ 显然成立.

现考察 $x \neq \theta$ 情况. 由定理 4.1.1 知, $\|x\|_\alpha$ 与 $\|x\|_\beta$ 均是坐标 ξ 的连续函数. 定义

函数 $f(\xi)$ 和集合 S :

$$f(\xi) = \frac{\|x\|_a}{\|x\|_\beta}$$

$$S = \left\{ \eta \in F^n \mid \sum_{i=1}^n |\eta_i|^2 = 1 \right\}$$

则 $f(\xi)$ 也是 ξ 的连续函数, 集合 S 是 F^n 中的一个单位超球面 (或称为 F^n 中的一个 $n-1$ 维闭子流形) 且为有界闭集.

注意到

$$\begin{aligned} \frac{\|x\|_a}{\sum_{i=1}^n |\xi_i|^2} &= \left\| \frac{\xi_1 \zeta_1 + \cdots + \xi_n \zeta_n}{\sum_{i=1}^n |\xi_i|^2} \right\|_a = \|\zeta_1, \dots, \zeta_n\|_a' \\ \frac{\|x\|_\beta}{\sum_{i=1}^n |\xi_i|^2} &= \left\| \frac{\xi_1 \zeta_1 + \cdots + \xi_n \zeta_n}{\sum_{i=1}^n |\xi_i|^2} \right\|_\beta = \|\zeta_1, \dots, \zeta_n\|_\beta' \end{aligned}$$

式中:

$$\xi' = \left[\frac{\xi_1}{\left(\sum_{i=1}^n |\xi_i|^2\right)^{\frac{1}{2}}}, \dots, \frac{\xi_n}{\left(\sum_{i=1}^n |\xi_i|^2\right)^{\frac{1}{2}}} \right]^T \in S$$

此时, $f(\xi)$ 等价为

$$f(\xi') = \frac{\frac{\|x\|_a}{\sum_{i=1}^n |\xi_i|^2}}{\frac{\|x\|_\beta}{\sum_{i=1}^n |\xi_i|^2}}$$

由于连续函数 $f(\xi')$ 在有界闭集 S 上必有最大值 k_2 和最小值 k_1 , 则 $k_1 \|x\|_\beta \leq \|x\|_a \leq k_2 \|x\|_\beta$. 证毕.

命题 4.1.1 (范数等价的性质) 设 V 是数域 F 上的有限维线性空间, 向量范数 $\|x\|_a$ 与 $\|x\|_\beta$ 等价, 则

(1) 自反性: $1 \cdot \|x\|_a \leq \|x\|_a \leq 1 \cdot \|x\|_a$;

(2) 对称性: $\frac{1}{k_1} \|x\|_a \leq \|x\|_\beta \leq \frac{1}{k_2} \|x\|_a$;

(3) 传递性: 若 $\|x\|_\beta$ 与 $\|x\|_\gamma$ 等价, 则向量范数 $\|x\|_a$ 与 $\|x\|_\gamma$ 等价.

赋范线性空间中范数等价性结论是后续向量序列和矩阵序列的收敛性证明的重要基础理论.

4.2 矩阵范数

任一 $m \times n$ 阶的矩阵可看作 mn 维向量(矩阵空间 $\mathbb{C}^{m \times n}$ 与向量空间 \mathbb{C}^{mn} 同构), 故可将向量范数的定义和性质直接移植至矩阵.

定义 4.2.1(矩阵的向量范数) 对任意矩阵 $A \in \mathbb{C}^{m \times n}$, 定义 $\|A\|$ 是对应以 A 为自变量的实值函数, 且满足以下三条性质:

- (1) 正定性: $\|A\| \geq 0$, 当且仅当 $A = O$ 时有 $\|A\| = 0$;
- (2) 齐次性: $\forall k \in \mathbb{C}, \|kA\| = |k| \|A\|$;
- (3) 三角不等式: $\forall A$ 和 $B \in \mathbb{C}^{m \times n}$, 有 $\|A+B\| \leq \|A\| + \|B\|$;

则称 $\|A\|$ 是矩阵 A 的向量范数.

例 4.2.1 设 $A = (a_{ij}) \in \mathbb{C}^{m \times n}$, 则

$$\|A\|_{v1} = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|$$

$$\|A\|_{v\infty} = \max_{i,j} |a_{ij}|$$

$$\|A\|_{v2} = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{\frac{1}{2}}$$

$$\|A\|_{vp} = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^p \right)^{\frac{1}{p}}, \quad p \geq 1$$

均为矩阵 A 的向量范数.

类似于 4.1 节, 对于矩阵的向量范数, 我们同样有如下结论:

定理 4.2.1 矩阵 $A = (a_{ij}) \in \mathbb{C}^{m \times n}$ 的任一向量范数均是 A 元素的连续函数.

定理 4.2.2 线性空间 $\mathbb{C}^{m \times n}$ 的任意两个向量范数是等价的, 即对 $\|A\|_{v\alpha}$ 和 $\|A\|_{v\beta}$ 存在正数 k_1 及 k_2 使得对任意矩阵 $A \in \mathbb{C}^{m \times n}$, 有

$$k_1 \|A\|_{v\beta} \leq \|A\|_{v\alpha} \leq k_2 \|A\|_{v\beta}$$

尽管矩阵可视为拉直的向量, 但矩阵和向量还是有所不同, 典型的是矩阵有乘法运算. 因此, 在考虑范数时须兼顾矩阵的乘法运算.

定义 4.2.2(矩阵范数) 对任意矩阵 $A \in \mathbb{C}^{m \times n}$, 定义 $\|A\|$ 均是对应以 A 为自变量的实值函数, 且满足以下四条性质:

- (1) 正定性: $\|A\| \geq 0$, 当且仅当 $A = O$ 时有 $\|A\| = 0$;
- (2) 齐次性: $\forall k \in \mathbb{C}, \|kA\| = |k| \|A\|$;
- (3) 三角不等式: $\|A+B\| \leq \|A\| + \|B\|$;
- (4) 矩阵乘法相容性: $\|AB\| \leq \|A\| \|B\|$ (须矩阵乘积有意义),

则称 $\|A\|$ 是 A 的矩阵范数.

注 1: 若反向性质(4)中矩阵乘法相容性的不等式, 即 $\|AB\| \geq \|A\| \|B\|$, 此时幂

零矩阵(对于 n 阶复方阵 A , 存在正整数 k 使得 $A^k = O$, 则 A 称为幂零矩阵)的矩阵范数将是 0, 这与矩阵范数的正定性要求矛盾. 矩阵乘法相容性的不等式实际上保证了矩阵幂级数的收敛性(设 $\|A\| \leq 1$, 则根据矩阵乘法相容性得当 $k \rightarrow \infty$, $\|A^k\| \rightarrow 0$).

【思考】 设 $A = (a_{ij}) \in \mathbb{C}^{m \times n}$, $\|A\|_{v\infty} = \max_{i,j} |a_{ij}|$ 是矩阵范数吗?

分析: 考察 $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, $B = \begin{bmatrix} -1 & 0 \\ -1 & 1 \end{bmatrix}$, 则 $AB = \begin{bmatrix} -1 & 0 \\ -2 & 1 \end{bmatrix}$. 又知 $\|A\|_{v\infty} = 1$, $\|B\|_{v\infty} = 1$, $\|AB\|_{v\infty} = 2$, 则不等式 $\|AB\| \leq \|A\| \|B\|$ 不成立. 因此, $\|A\|_{v\infty}$ 不是矩阵范数.

例 4.2.2 设 $A = (a_{ij}) \in \mathbb{C}^{m \times n}$, 证明:

$$\|A\|_{v2} = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{\frac{1}{2}} = (\text{tr}(A^H A))^{\frac{1}{2}}$$

是 A 的矩阵范数.

证明: 设 $A = (a_{ij}) \in \mathbb{C}^{m \times n}$, $B = (b_{ij}) \in \mathbb{C}^{n \times p}$ 和 $D = (d_{ij}) \in \mathbb{C}^{m \times n}$, 则有 $AB = (c_{ij}) \in \mathbb{C}^{m \times p}$, 其中 $c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$.

(1) 正定性: $\|A\|_{v2} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} \geq 0$, 且 $\|A\|_{v2} = 0 \Leftrightarrow A = O$.

(2) 齐次性: 对任意 $k \in \mathbb{C}$, 有

$$\|kA\|_{v2} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |ka_{ij}|^2} = |k| \cdot \|A\|_{v2}$$

(3) 三角不等式:

$$\begin{aligned} \|A + D\|_{v2} &= \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij} + d_{ij}|^2} \\ &\leq \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} + \sqrt{\sum_{i=1}^m \sum_{j=1}^n |d_{ij}|^2} \\ &= \|A\|_{v2} + \|D\|_{v2} \end{aligned}$$

(4) 乘法相容性: $\|AB\|_{v2}^2 = \sum_{i=1}^m \sum_{j=1}^p |c_{ij}|^2$, 其中

$$\begin{aligned} |c_{ij}|^2 &= \left| \sum_{k=1}^n a_{ik} b_{kj} \right|^2 \leq \left(\sum_{k=1}^n |a_{ik} b_{kj}| \right)^2 \\ &\leq \left(\sum_{k=1}^n |a_{ik}|^2 \right) \left(\sum_{k=1}^n |b_{kj}|^2 \right) \end{aligned}$$

$$\|AB\|_{v2}^2 \leq \sum_{i,j} \left[\left(\sum_{k=1}^n |a_{ik}|^2 \right) \left(\sum_{k=1}^n |b_{kj}|^2 \right) \right] = \|A\|_{v2}^2 \|B\|_{v2}^2$$

因此, $\|A\|_{v2}$ 是 A 的矩阵范数.

注 2: $\|A\|_{v2}$ 范数称为 Frobenious 范数, 简称为 F-范数, 并常记为 $\|A\|_F$.

定理 4.2.3 设 $A = (a_{ij}) \in \mathbb{C}^{m \times n}$ 和 $x \in \mathbb{C}^n$, 则

(1) $\|UA\|_F = \|AV\|_F = \|UAV\|_F = \|A\|_F$, 其中 U 和 V 是酉矩阵;

(2) $\|A\|_F^2 = \sum_{i=1}^n \|\beta_i\|_2^2$, 其中矩阵 A 按列分块, 记为 $A = [\beta_1, \dots, \beta_n]$; 或 $\|A\|_F^2 =$

$\sum_{i=1}^m \|\alpha_i\|_2^2$, 其中矩阵 A 按行分块, 记为

$$A = \begin{bmatrix} \alpha_1^T \\ \vdots \\ \alpha_m^T \end{bmatrix}$$

(3) $\|Ax\|_2 \leq \|A\|_F \|x\|_2$.

证明: (1) 根据 F-范数定义知

$$\|UA\|_F = (\text{tr}((UA)^H(UA)))^{\frac{1}{2}} = (\text{tr}(A^H A))^{\frac{1}{2}} = \|A\|_F$$

$$\|AV\|_F = (\text{tr}((AV)(AV)^H))^{\frac{1}{2}} = (\text{tr}(A^H A))^{\frac{1}{2}} = \|A\|_F$$

$$\|UAV\|_F = (\text{tr}(U^H A^H A U))^{\frac{1}{2}} = (\text{tr}(A^H A))^{\frac{1}{2}} = \|A\|_F$$

(2) 根据向量 2-范数定义知,

$$\sum_{i=1}^n \|\beta_i\|_2^2 = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 = \sum_{i=1}^m \|\alpha_i\|_2^2$$

$$\text{因此, } \|A\|_F^2 = \sum_{i=1}^n \|\alpha_i\|_2^2 = \sum_{i=1}^m \|\beta_i\|_2^2.$$

(3) 对矩阵 A 行分块得

$$\|Ax\|_2 = \left\| \begin{bmatrix} \alpha_1^T x \\ \vdots \\ \alpha_m^T x \end{bmatrix} \right\|_2 = \left(\sum_{i=1}^m |\alpha_i^T x|^2 \right)^{\frac{1}{2}}$$

根据 Cauchy 不等式得, $|\alpha_i^T x| \leq \|\alpha_i\|_2 \|x\|_2$, 并将其代入上式得

$$\|Ax\|_2 \leq \left(\sum_{i=1}^m \|\alpha_i\|_2 \right) \|x\|_2 = \|A\|_F \|x\|_2$$

注 3: 定理 4.2.2 性质(1)称为 F-范数的酉不变性.

【思考】 如何根据已有的矩阵范数构造新的矩阵范数?

例 4.2.3 设 $A \in \mathbb{C}^{n \times n}$, $\|\cdot\|$ 是某一给定矩阵范数. 定义

$$\|A\|_m = \|P^{-1}AP\|$$

则 $\|\cdot\|_m$ 是矩阵范数, 其中 P 是 n 阶可逆矩阵.

证明: (1) 正定性: 由定义知, $\|A\|_m \geq 0$ 且等号成立的充分必要条件为 $A = O$.

(2) 齐次性: 对任意 $k \in \mathbb{C}$, $\|kA\|_m = |k| \|P^{-1}AP\| = |k| \|A\|_m$.

(3) 三角不等式: $\|A+B\|_m = \|P^{-1}(A+B)P\| \leq \|A\|_m + \|B\|_m$.

(4) 相容性: 对任意 A 和 $B \in \mathbb{C}^{n \times n}$,

$$\|AB\|_m = \|P^{-1}APP^{-1}BP\| \leq \|P^{-1}AP\| \|P^{-1}BP\| = \|A\|_m \|B\|_m$$

因此, $\|\cdot\|_m$ 是矩阵范数.

注 4: 单位矩阵的矩阵范数必不小于 1. 由矩阵乘法的相容性可知

$$\|I\| = \|I^2\| \leq \|I\|^2$$

请注意, 在矩阵的向量范数下, 单位矩阵的向量范数可以是任何正数.

4.3 相容范数

向量范数与矩阵范数在运算中经常同时出现, 这需建立它们之间的关系. 因此, 引入相容范数.

定义 4.3.1 (向量范数与矩阵范数相容) 若对任意 $A \in \mathbb{C}^{m \times n}$ 和 $x \in \mathbb{C}^n$, 向量范数 $\|x\|_v$ 与矩阵范数 $\|A\|_m$ 满足

$$\|Ax\|_v \leq \|A\|_m \|x\|_v$$

则称向量范数 $\|x\|_v$ 与矩阵范数 $\|A\|_m$ 相容.

【思考】 给定矩阵范数, 是否存在与之相容的向量范数? 反之, 给定向量范数, 是否存在与之相容的矩阵范数?

例 4.3.1 由定理 4.2.3 知, 对任意矩阵 $A \in \mathbb{C}^{m \times n}$ 和 $x \in \mathbb{C}^n$, $\|Ax\|_2 \leq \|A\|_F \|x\|_2$. 由此, 向量 2-范数和矩阵 F-范数是相容的.

定理 4.3.1 设 $\|A\|_m$ 是 $\mathbb{C}^{n \times n}$ 的一个矩阵范数, 则必存在 \mathbb{C}^n 上与之相容的向量范数.

证明: 取定非零向量 $\alpha = [\alpha_1, \dots, \alpha_n]^T \in \mathbb{C}^n$, 则对任意向量 $x = [x_1, \dots, x_n]^T \in \mathbb{C}^n$, 定义

$$\|x\|_v = \|x\alpha^T\|_m \quad (4.3.1)$$

则(1)正定性: $\|x\|_v \geq 0$ 成立; $\|x\|_v = 0$ 当且仅当 $x\alpha^T = O$, 即

$$\alpha_i x_j = 0, \quad i, j = 1, \dots, n$$

上式成立的充分必要条件为 $x = 0$.

(2) 齐次性: $\forall k \in \mathbb{C}, \|kx\|_v = |k| \|x\|_v$.

(3) 三角不等式: 设 $x = [x_1, \dots, x_n]^T \in \mathbb{C}^n$ 和 $y = [y_1, y_2, \dots, y_n]^T \in \mathbb{C}^n$, 则

$$\|x + y\|_v = \|x\alpha^T + y\alpha^T\|_m \leq \|x\alpha^T\|_m + \|y\alpha^T\|_m = \|x\|_v + \|y\|_v$$

(4) 相容性: $\forall A \in \mathbb{C}^{n \times n}, \|Ax\|_v = \|Ax\alpha^T\|_m \leq \|A\|_m \|x\alpha^T\|_m = \|A\|_m \|x\|_v$.

综上, 定义式(4.3.1)是一个向量范数, 且与给定的矩阵范数 $\|\cdot\|_m$ 相容.

定理 4.3.2 设 $\|x\|_v$ 是 \mathbb{C}^n 上的一个向量范数, 对任意矩阵 $A \in \mathbb{C}^{m \times n}$, 定义

$$\|A\| = \max_{\|x\|_v=1} \|Ax\|_v \quad (4.3.2)$$

则 $\|A\|$ 是一个与 $\|x\|_v$ 相容的矩阵范数, 称矩阵范数 (4.3.2) 是从属于向量范数 $\|\cdot\|_v$ 的算子范数或由向量范数 $\|\cdot\|_v$ 诱导的矩阵范数 (或简称诱导范数).

分析: 由于定义式 (4.3.2) 涉及极值问题, 故该定理首先须考虑最大值的存在性问题. 由于 $\|Ax\|_v$ 是向量 x 各分量的连续函数, 故它在有界闭集 $\|x\|_v = 1$ 可取到最大值. 因此, 定义式 (4.3.2) 有意义.

证明: (1) 正定性: $\|A\| \geq 0$ 显然成立, 且有 $A = O \Rightarrow \|A\| = 0$. $\|A\| = 0$ 意味着对满足 $\|x\|_v = 1$ 的任意向量 x 都有 $Ax = 0$. 此时, 齐次线性方程组 $Ax = 0$ 的解空间维数为 n . 根据定理 2.3.6 知, $\text{rank}(A) = 0$. 因此, $A = O$.

(2) 齐次性: 对任意 $k \in \mathbb{C}$, 有

$$\|kA\| = \max_{\|x\|_v=1} \|kAx\|_v = |k| \max_{\|x\|_v=1} \|Ax\|_v = |k| \|A\|.$$

(3) 三角不等式: 对任意矩阵 A 和 $B \in \mathbb{C}^{m \times n}$, 存在向量 $x_0 \in \mathbb{C}^n$ 且 $\|x_0\|_v = 1$ 满足

$$\begin{aligned} \|A+B\| &= \|(A+B)x_0\|_v \leq \|Ax_0\|_v + \|Bx_0\|_v \\ &\leq \max_{\|x\|_v=1} \|Ax\|_v + \max_{\|x\|_v=1} \|Bx\|_v = \|A\| + \|B\| \end{aligned}$$

(4) 矩阵乘法相容性: 对任意矩阵 $A \in \mathbb{C}^{p \times m}$ 和 $B \in \mathbb{C}^{m \times n}$, 必存在 $\|y_0\| = 1$ 的 $y_0 \in \mathbb{C}^n$ 满足

$$\|AB\| = \|(AB)y_0\|_v$$

若 $\|By_0\|_v = 0$, 有 $B = O$, 即 $\|AB\| = \|A\| \|B\| = 0$, 则性质 (4) 成立.

若 $\|By_0\|_v \neq 0$, 则有

$$\begin{aligned} \|AB\| &= \|(AB)y_0\|_v = \left\| A \left(\frac{1}{\|By_0\|_v} By_0 \right) \|By_0\|_v \right\|_v \\ &= \left\| A \left(\frac{1}{\|By_0\|_v} By_0 \right) \right\|_v \|By_0\|_v \\ &\leq \|A\| \|By_0\|_v \leq \|A\| \|B\| \end{aligned}$$

(5) 矩阵范数与向量范数的相容性: 若 $x = 0$, $\|Ax\|_v = \|A\| \|x\|_v = 0$. 若 $x \neq 0$, 有

$$\|Ax\|_v = \left\| A \frac{1}{\|x\|_v} x \right\|_v \|x\|_v \leq \|A\| \|x\|_v$$

因此, 矩阵范数 (4.3.2) 是从属于向量范数 $\|\cdot\|_v$ 的算子范数.

注 1: 我们常用到算子范数有如下等价定义式:

$$\|A\| = \max_{\|x\|_v \neq 0} \frac{\|Ax\|_v}{\|x\|_v} \quad (4.3.3)$$

常见的向量范数有 $\|x\|_1$ 、 $\|x\|_\infty$ 和 $\|x\|_2$, 则从属于它们的算子范数分别记为 $\|A\|_1$ 、 $\|A\|_\infty$ 和 $\|A\|_2$, 并分别称为列和范数、行和范数和谐范数.

定理 4.3.3 设 $A = (a_{ij}) \in \mathbb{C}^{m \times n}$ 和 $x \in \mathbb{C}^n$, $\sigma_{\max}(A)$ 是矩阵 A 的最大奇异值, 则从属于向量范数 $\|x\|_1$ 、 $\|x\|_\infty$ 和 $\|x\|_2$ 的算子范数分别为

$$\|\mathbf{A}\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}| \quad (\text{列和范数})$$

$$\|\mathbf{A}\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}| \quad (\text{行和范数})$$

$$\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^H \mathbf{A})} = \sigma_{\max}(\mathbf{A}) \quad (\text{谱范数})$$

分析: 定理 4.3.3 本质上是证明以下三个等式:

$$\max_{\|\mathbf{x}\|_1=1} \|\mathbf{A}\mathbf{x}\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}| \quad (4.3.4)$$

$$\max_{\|\mathbf{x}\|_\infty=1} \|\mathbf{A}\mathbf{x}\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}| \quad (4.3.5)$$

$$\max_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^H \mathbf{A})} \quad (4.3.6)$$

以证明式(4.3.4)为例, 证明过程应分为两个步骤: 首先证明 $\|\mathbf{A}\mathbf{x}\|_1 \leq \|\mathbf{A}\|_1$; 再证明存在 $\|\mathbf{x}_0\|_1=1$ 的向量 \mathbf{x}_0 满足 $\|\mathbf{A}\mathbf{x}_0\|_1 = \|\mathbf{A}\|_1$.

证明: (1) $\forall \mathbf{x} = [x_1, \dots, x_n]^T \in \mathbb{C}^n$ 且 $\sum_{k=1}^n |x_k| = 1$, 有

$$\begin{aligned} \|\mathbf{A}\mathbf{x}\|_1 &= \sum_{i=1}^m \left| \sum_{k=1}^n a_{ik} x_k \right| \leq \sum_{i=1}^m \sum_{k=1}^n (|a_{ik}| \cdot |x_k|) \\ &= \sum_{k=1}^n (|x_k| \sum_{i=1}^m |a_{ik}|) \leq \sum_{k=1}^n (|x_k| \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|) = \|\mathbf{A}\|_1 \end{aligned}$$

设矩阵 \mathbf{A} 的第 j^* 列对 $j=1, \dots, n$ 满足

$$\sum_{i=1}^m |a_{ij}| \leq \sum_{i=1}^m |a_{ij^*}|$$

则取 $\mathbf{x}_0 = \mathbf{e}_{j^*}$ 有

$$\|\mathbf{A}\mathbf{x}_0\|_1 = \sum_{i=1}^m |a_{ij^*}| = \|\mathbf{A}\|_1$$

上式表明 $\|\mathbf{A}\mathbf{x}\|_1$ 的最大值可以取到 $\|\mathbf{A}\|_1$. 因此, 式(4.3.4)成立.

(2) $\forall \mathbf{x} = [x_1, \dots, x_n]^T \in \mathbb{C}^n$ 且 $\max_{1 \leq k \leq n} |x_k| = 1$, 有

$$\begin{aligned} \|\mathbf{A}\mathbf{x}\|_\infty &= \max_{1 \leq i \leq m} \left| \sum_{k=1}^n a_{ik} x_k \right| \leq \max_{1 \leq i \leq m} \sum_{k=1}^n (|a_{ik}| |x_k|) \\ &\leq \max_{1 \leq i \leq m} \sum_{k=1}^n |a_{ik}| = \|\mathbf{A}\|_\infty \end{aligned}$$

设矩阵 \mathbf{A} 的第 i^* 行满足

$$\sum_{j=1}^n |a_{ij}| \leq \sum_{j=1}^n |a_{i^*j}|, \quad \forall i=1, \dots, m$$

定义

$$\mu_j = \begin{cases} \frac{|a_{i^*j}|}{a_{i^*j}}, & a_{i^*j} \neq 0 \\ 1, & a_{i^*j} = 0 \end{cases}, \quad j=1, \dots, n$$

并取 $\mathbf{x}_0 = [\mu_1, \dots, \mu_n]^T$, 则 $\|\mathbf{x}_0\|_\infty = 1$. 由此, 考察

$$\|\mathbf{Ax}\|_\infty = \max_{1 \leq i \leq n} \left| \sum_{j=1}^n a_{ij} \mu_j \right| = \sum_{j=1}^n |a_{i^*j}| = \|\mathbf{A}\|_\infty$$

上式表明 $\|\mathbf{Ax}\|_\infty$ 的最大值可以取到 $\|\mathbf{A}\|_\infty$. 因此, 式(4.3.5)成立.

(3) 设 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ 是 Hermite 矩阵 $\mathbf{A}^H \mathbf{A}$ 的 n 个非负特征值, $\mathbf{x}_1, \dots, \mathbf{x}_n$ 是属于特征值 $\lambda_1, \dots, \lambda_n$ 的单位正交特征向量, 则向量组 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 是 \mathbb{C}^n 空间的一组标准正交基, 且 $\forall \mathbf{x} \in \mathbb{C}^n$,

$$\mathbf{x} = k_1 \mathbf{x}_1 + \dots + k_n \mathbf{x}_n \quad (4.3.7)$$

式中: $[k_1, \dots, k_n]^T$ 是向量 \mathbf{x} 在基 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 下的坐标.

将式(4.3.7)代入 $\|\mathbf{x}\|_2$, 得

$$\|\mathbf{x}\|_2 = (\mathbf{x}^H \mathbf{x})^{\frac{1}{2}} = \left(\sum_{i=1}^n |k_i|^2 \right)^{\frac{1}{2}}$$

又知

$$\begin{aligned} \|\mathbf{Ax}\|_2^2 &= \mathbf{x}^H \mathbf{A}^H \mathbf{Ax} = \left(\sum_{i=1}^n k_i \mathbf{x}_i \right)^H \mathbf{A}^H \mathbf{A} \left(\sum_{i=1}^n k_i \mathbf{x}_i \right) \\ &= \sum_{i=1}^n |k_i|^2 \lambda_i^2 \leq \lambda_1^2 \sum_{i=1}^n |k_i|^2 = \lambda_1^2 \|\mathbf{x}\|_2^2 \end{aligned}$$

由此, $\|\mathbf{Ax}\|_2 \leq \lambda_1 \|\mathbf{x}\|_2$. 取 \mathbf{x}_1 是属于 λ_1 的单位特征向量, 则 $\|\mathbf{Ax}_1\|_2 = \lambda_1$, 即不等式 $\|\mathbf{Ax}\|_2 \leq \lambda_1 \|\mathbf{x}\|_2$ 可以取到等号. 因此, 式(4.3.6)成立. 证毕.

注 2: 由定理 4.2.3 知, $\|\mathbf{A}\|_F$ 与 $\|\mathbf{x}\|_2$ 是相容的, 而 $\|\mathbf{A}\|_2$ 作为从属于 $\|\mathbf{x}\|_2$ 的算子范数, 自然也是相容的, 所以相容于同一向量范数的矩阵范数有多个. 事实上, 我们总有 $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F$.

注 3: 若存在常数 M 使得对任意向量 $\mathbf{x} \in \mathbb{C}^n$, 有 $\|\mathbf{Ax}\|_v \leq M \|\mathbf{x}\|_v$, 则 $\|\mathbf{A}\|_v \leq M$, 即从属于范数 $\|\mathbf{x}\|_v$ 的算子范数 $\|\mathbf{A}\|_v$ 是使不等式 $\|\mathbf{Ax}\|_v \leq M \|\mathbf{x}\|_v$ 成立的最小常数.

例 4.3.2 设 $\mathbf{A} = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 2 & 3 \\ 1 & 2 & 0 \end{bmatrix}$, 分别计算 $\|\mathbf{A}\|_1$ 、 $\|\mathbf{A}\|_\infty$ 、 $\|\mathbf{A}\|_2$ 和 $\|\mathbf{A}\|_F$.

解: 由 $\|\mathbf{A}\|_1$ 、 $\|\mathbf{A}\|_\infty$ 和 $\|\mathbf{A}\|_F$ 定义计算得 $\|\mathbf{A}\|_1 = 5$, $\|\mathbf{A}\|_\infty = 5$, $\|\mathbf{A}\|_F = \sqrt{23}$; 对于 $\|\mathbf{A}\|_2$, 有

$$\mathbf{A}^H \mathbf{A} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 9 & 6 \\ 0 & 6 & 9 \end{bmatrix}$$

计算该矩阵特征值分别为 3、5 和 15. 因此, $\|\mathbf{A}\|_2 = \sqrt{15}$.

由例 4.3.2 看出, $\|\mathbf{A}\|_1$ 、 $\|\mathbf{A}\|_\infty$ 和 $\|\mathbf{A}\|_F$ 比较容易计算, 而计算 $\|\mathbf{A}\|_2$ 则比较复杂. 不可否认的是, $\|\mathbf{A}\|_2$ 对矩阵元素变化比较敏感. 因此, 在诸多不等式放缩时常采用谱范数.

注 4: 设 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ 是 Hermite 矩阵 \mathbf{A} 的 n 个特征值, 则由谱范数证明过程知

$$\lambda_n \leq R(\mathbf{x}) \leq \lambda_1 \quad (4.3.8)$$

其中 $R(\mathbf{x})$ 的定义为

$$R(\mathbf{x}) = \frac{\mathbf{x}^H \mathbf{A} \mathbf{x}}{\mathbf{x}^H \mathbf{x}}, \quad \forall \mathbf{x} \neq \mathbf{0}$$

称 $R(\mathbf{x})$ 为 Hermite 矩阵 \mathbf{A} 的 Rayleigh 商.

推论 4.3.1 设 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ 是 Hermite 矩阵 \mathbf{A} 的 n 个特征值, 则

$$\lambda_1 = \max_{\mathbf{x} \neq \mathbf{0}} R(\mathbf{x}), \quad \lambda_n = \min_{\mathbf{x} \neq \mathbf{0}} R(\mathbf{x})$$

基于 Rayleigh 商, 有如下著名的极大极小或极小极大定理.

定理 4.3.4 设 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ 是 Hermite 矩阵 \mathbf{A} 的 n 个特征值, V_i 是 \mathbb{C}^n 中 i 维子空间, 则 $\forall i=1, \cdots, n$,

$$\begin{aligned} \lambda_i &= \max_{\substack{\forall V_i \\ \mathbf{x} \in V_i \\ \mathbf{x} \neq \mathbf{0}}} \min R(\mathbf{x}) \\ \lambda_i &= \min_{\substack{\forall V_{n-i+1} \\ \mathbf{x} \in V_{n-i+1} \\ \mathbf{x} \neq \mathbf{0}}} \max R(\mathbf{x}) \end{aligned}$$

证明: 这里仅证明 $\lambda_i = \max_{\substack{\forall V_i \\ \mathbf{x} \in V_i \\ \mathbf{x} \neq \mathbf{0}}} \min R(\mathbf{x})$, 另一公式可采用类似方法证明.

设 V_i 是 \mathbb{C}^n 中任一 i 维子空间, 它的一组标准正交基为 $\mathbf{y}_1, \cdots, \mathbf{y}_i$. 令 $\mathbf{x}_1, \cdots, \mathbf{x}_n$ 是属于特征值 $\lambda_1, \cdots, \lambda_n$ 的单位正交特征向量, 并定义 $(n-i+1)$ 维子空间 W_{n-i+1} 为

$$W_{n-i+1} = \text{span}(\mathbf{x}_i, \mathbf{x}_{i+1}, \cdots, \mathbf{x}_n)$$

由于 $\dim(V_i) + \dim(W_{n-i+1}) = n+1 > n$, 则 V_i 和 W_{n-i+1} 的交空间必存在非零向量. 不妨假设 $\tilde{\mathbf{x}}_0 \in V_i \cap W_{n-i+1}$, 则 $\tilde{\mathbf{x}}_0$ 可表示为

$$\tilde{\mathbf{x}}_0 = k_i \mathbf{x}_i + k_{i+1} \mathbf{x}_{i+1} + \cdots + k_n \mathbf{x}_n$$

式中: $k_j \in \mathbb{C}, j=i, i+1, \cdots, n$.

相应地, 有

$$R(\tilde{\mathbf{x}}_0) = \frac{\sum_{j=i}^n |k_j|^2 \lambda_j}{\sum_{j=i}^n |k_j|^2} \leq \lambda_i$$

由此, $\min_{\substack{\mathbf{x} \in V_i \\ \mathbf{x} \neq \mathbf{0}}} R(\mathbf{x}) \leq \lambda_i$. 另一方面, 考虑 V_i 的一种特殊情况:

$$V_i = \text{span}(\mathbf{x}_1, \cdots, \mathbf{x}_i)$$

则对 V_i 中任意向量 x , 必存在一组常数 k_1, \dots, k_i 使得 $x = k_1 x_1 + \dots + k_i x_i$, 并有

$$R(x) = \frac{\sum_{j=1}^i |k_j|^2 \lambda_j}{\sum_{j=1}^i |k_j|^2}, \quad j = 1, \dots, i$$

显然, 对 V_i 中任意向量 x , 有 $R(x) \geq \lambda_i$. 因此, $\lambda_i = \max_{\substack{x \in V_i \\ x \neq 0}} \min R(x)$.

【应用】 利用极大极小定理可研究 Hermite 矩阵特征值的扰动问题(讨论 Hermite 矩阵元素发生变化时相应矩阵特征值的变化范围).

定理 4.3.5 设 A 和 E 均为 n 阶 Hermite 复方阵, 则对 $i = 1, \dots, n$, 有

$$\lambda_i(A) + \lambda_n(E) \leq \lambda_i(A + E) \leq \lambda_i(A) + \lambda_1(E)$$

式中: $\lambda_i(A)$ 是矩阵 A 的第 i 大特征值.

证明: 设 x_1, \dots, x_n 分别属于矩阵 A 的特征值 $\lambda_1, \dots, \lambda_n$ 的单位正交特征向量. 由向量组 x_1, \dots, x_n 中取前 i 个向量, 并定义 $V_{n-i+1} = \text{span}(x_i, \dots, x_n)$.

根据定理 4.3.4 知,

$$\begin{aligned} \lambda_i(A + E) &\leq \max_{\substack{x \in V_{n-i+1} \\ x \neq 0}} \frac{x^H(A + E)x}{x^H x} \\ &\leq \max_{\substack{x \in V_{n-i+1} \\ x \neq 0}} \frac{x^H A x}{x^H x} + \max_{\substack{x \in V_{n-i+1} \\ x \neq 0}} \frac{x^H E x}{x^H x} \end{aligned}$$

式中:

$$\begin{aligned} \max_{\substack{x \in V_{n-i+1} \\ x \neq 0}} \frac{x^H A x}{x^H x} &= \lambda_i(A) \\ \max_{\substack{x \in V_{n-i+1} \\ x \neq 0}} \frac{x^H E x}{x^H x} &\leq \lambda_1(E) \end{aligned}$$

因此, $\lambda_i(A + E) \leq \lambda_i(A) + \lambda_1(E)$. 另一方面, 将该不等式中的 A 用 $A + E$ 、 E 用 $-E$ 替换, 于是有

$$\lambda_i(A) \leq \lambda_i(A + E) + \lambda_1(-E)$$

式中: $\lambda_1(-E) = -\lambda_n(E)$. 此时有 $\lambda_i(A) + \lambda_n(E) \leq \lambda_i(A + E)$. 证毕.

如无特殊说明, 我们遇到矩阵和向量同时出现时, 总是假设矩阵范数和向量范数是相容的.

定理 4.3.6 设 $A \in \mathbb{C}^{n \times n}$, $\|A\|$ 是某一矩阵范数. 若 $\|A\| < 1$, 证明 $I - A$ 非奇异, 且

$$\|(I - A)^{-1}\| \leq \frac{\|I\|}{1 - \|A\|}$$

证明: 采用反证法. 假设矩阵 $I - A$ 奇异, 则齐次线性方程组 $(I - A)x = 0$ 有非

零解,即存在非零向量 \mathbf{x}_0 使得 $(\mathbf{I}-\mathbf{A})\mathbf{x}_0=\mathbf{0}$. 进一步整理得

$$\mathbf{A}\mathbf{x}_0=\mathbf{x}_0$$

对上式两端取与矩阵范数 $\|\cdot\|$ 相容的向量范数,得

$$\|\mathbf{x}_0\|=\|\mathbf{A}\mathbf{x}_0\|\leq\|\mathbf{A}\|\|\mathbf{x}_0\|$$

由于 $\mathbf{x}_0\neq\mathbf{0}$, 故 $\|\mathbf{x}_0\|\neq 0$. 由此, $\|\mathbf{A}\|\geq 1$, 这与 $\|\mathbf{A}\|<1$ 矛盾. 因此, 假设不成立, 即 $\mathbf{I}-\mathbf{A}$ 非奇异.

另一方面, 根据 $(\mathbf{I}-\mathbf{A})(\mathbf{I}-\mathbf{A})^{-1}=\mathbf{I}$ 得

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

对上式两端取矩阵范数, 得

$$\begin{aligned}\|(\mathbf{I}-\mathbf{A})^{-1}\|&=\|\mathbf{I}+\mathbf{A}(\mathbf{I}-\mathbf{A})^{-1}\|\\&\leq\|\mathbf{I}\|+\|\mathbf{A}\|\|(\mathbf{I}-\mathbf{A})^{-1}\|\end{aligned}$$

整理即得结论.

4.4 矩阵扰动分析

为解决科学与工程中的实际问题, 一般首先依据物理、电学、力学等规律建立数学模型, 然后基于数学模型给出求解模型的数值计算方法. 在模型求解过程中, 通常存在两类误差影响计算结果的精度: 数值计算方法引起的截断误差和计算环境引起的舍入误差. 为分析这些误差对模型解的影响, 人们将其归结为原始数据的扰动(或摄动)对解的影响.

例 4.4.1 考察线性方程组 $\mathbf{A}\mathbf{x}=\mathbf{b}$, 其中 $\mathbf{x}=[x_1, x_2]^T$,

$$\mathbf{A}=\begin{bmatrix} 1 & 0.99 \\ 0.99 & 0.98 \end{bmatrix}, \quad \mathbf{b}=\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

通过计算得, 该方程组的精确解为 $x_1=100, x_2=-100$.

若系数矩阵 \mathbf{A} 有一扰动 $\begin{bmatrix} 0 & 0 \\ 0 & 0.01 \end{bmatrix}$, 并且右端项 \mathbf{b} 也有一扰动 $\begin{bmatrix} 0 \\ 0.001 \end{bmatrix}$, 则扰动后的线性方程组为

$$\begin{bmatrix} 1 & 0.99 \\ 0.99 & 0.99 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1.001 \end{bmatrix}$$

通过计算得, 该方程组的精确解为 $x_1=-\frac{1}{10}, x_2=\frac{10}{9}$.

对比两组解向量知, 系数矩阵和右端项的微小扰动会引起解的强烈变化. 同时, 注意例 4.4.1 本身并没有截断误差和舍入误差, 因此原始数据的扰动对问题解的影响程度取决于问题本身的固有性质.

定义 4.4.1(病态问题) 病态问题是指输出结果相对于输入非常敏感的问题, 输入数据中哪怕是极少(或者极微妙)的扰动也会导致输出的较大改变. 相反的, 对于

输入不敏感的问题,我们称为良态问题.

矩阵扰动分析就是研究矩阵元素的变化对矩阵问题解的影响,它对矩阵理论和矩阵计算都具有重要意义.这里仅简要介绍矩阵 A 的逆矩阵,并对以 A 为系数矩阵的线性方程组的解的扰动进行分析.为讨论方便,假定 $\|\cdot\|$ 为 $\mathbb{C}^{n \times n}$ 上满足 $\|I\|=1$ 的相容矩阵范数.

设矩阵 $A \in \mathbb{C}^{n \times n}$ 并且 A 非奇异,经扰动变为 $A+E$,其中 $E \in \mathbb{C}^{n \times n}$ 称为扰动矩阵.我们需要解决:

(1) 什么条件下 $A+E$ 非奇异?

(2) 当 $A+E$ 非奇异时, $(A+E)^{-1}$ 与 A^{-1} 的近似程度.

定理 4.4.1 设 A 和 $E \in \mathbb{C}^{n \times n}$, $B=A+E$. 若 A 与 B 均非奇异,则

$$\frac{\|B^{-1} - A^{-1}\|}{\|A^{-1}\|} \leq \|A\| \|B^{-1}\| \frac{\|E\|}{\|A\|} \quad (4.4.1)$$

证明: 由

$$B^{-1} - A^{-1} = A^{-1}(A - B)B^{-1} = -A^{-1}EB^{-1} \quad (4.4.2)$$

对式(4.4.2)两端取矩阵范数,得

$$\|B^{-1} - A^{-1}\| \leq \|A^{-1}\| \|B^{-1}\| \|E\|$$

于是

$$\frac{\|B^{-1} - A^{-1}\|}{\|A^{-1}\|} \leq \|B^{-1}\| \|E\| = \|A\| \|B^{-1}\| \frac{\|E\|}{\|A\|}$$

定理 4.4.2 设 $A \in \mathbb{C}^{n \times n}$ 是非奇异矩阵, $E \in \mathbb{C}^{n \times n}$, 且满足条件

$$\|A^{-1}E\| < 1 \quad (4.4.3)$$

则 $A+E$ 非奇异, 并且有

$$\|(A+E)^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}E\|} \quad (4.4.4)$$

$$\frac{\|(A+E)^{-1} - A^{-1}\|}{\|A^{-1}\|} \leq \frac{\|A^{-1}E\|}{1 - \|A^{-1}E\|} \quad (4.4.5)$$

证明: 因为 $A+E = A(I + A^{-1}E)$, 其中 $\|A^{-1}E\| < 1$, 则由定理 4.3.6 知, $I + A^{-1}E$ 非奇异. 因此, $A+E$ 也是非奇异矩阵.

另由定理 4.3.6 知

$$\|(I + A^{-1}E)^{-1}\| \leq \frac{1}{1 - \|A^{-1}E\|}$$

则有

$$\begin{aligned} \|(A+E)^{-1}\| &= \|(I + A^{-1}E)^{-1}A^{-1}\| \\ &\leq \|(I + A^{-1}E)^{-1}\| \|A^{-1}\| \\ &\leq \frac{\|A^{-1}\|}{1 - \|A^{-1}E\|} \end{aligned}$$

又知

$$(\mathbf{A} + \mathbf{E})^{-1} - \mathbf{A}^{-1} = [(\mathbf{I} + \mathbf{A}^{-1}\mathbf{E})^{-1} - \mathbf{I}]\mathbf{A}^{-1}$$

则

$$\|(\mathbf{A} + \mathbf{E})^{-1} - \mathbf{A}^{-1}\| \leq \|(\mathbf{I} + \mathbf{A}^{-1}\mathbf{E})^{-1} - \mathbf{I}\| \|\mathbf{A}^{-1}\|$$

由定理 4.4.1 可得

$$\|(\mathbf{I} + \mathbf{A}^{-1}\mathbf{E})^{-1} - \mathbf{I}\| \leq \|(\mathbf{I} + \mathbf{A}^{-1}\mathbf{E})^{-1}\| \|\mathbf{A}^{-1}\mathbf{E}\| \leq \frac{\|\mathbf{A}^{-1}\mathbf{E}\|}{1 - \|\mathbf{A}^{-1}\mathbf{E}\|}$$

因此,

$$\|(\mathbf{A} + \mathbf{E})^{-1} - \mathbf{A}^{-1}\| \leq \frac{\|\mathbf{A}^{-1}\| \|\mathbf{A}^{-1}\mathbf{E}\|}{1 - \|\mathbf{A}^{-1}\mathbf{E}\|}$$

由于 $\|\mathbf{A}^{-1}\mathbf{E}\| \leq \|\mathbf{A}^{-1}\| \|\mathbf{E}\|$, 所以由定理 4.4.2 可得如下推论:

推论 4.4.1 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 是非奇异矩阵, $\mathbf{E} \in \mathbb{C}^{n \times n}$ 满足条件 $\|\mathbf{A}^{-1}\| \|\mathbf{E}\| < 1$, 则 $\mathbf{A} + \mathbf{E}$ 非奇异, 并且有

$$\frac{\|(\mathbf{A} + \mathbf{E})^{-1} - \mathbf{A}^{-1}\|}{\|\mathbf{A}^{-1}\|} \leq \frac{\kappa(\mathbf{A}) \frac{\|\mathbf{E}\|}{\|\mathbf{A}\|}}{1 - \kappa(\mathbf{A}) \frac{\|\mathbf{E}\|}{\|\mathbf{A}\|}} \quad (4.4.6)$$

其中, $\kappa(\mathbf{A}) = \|\mathbf{A}\| \|\mathbf{A}^{-1}\|$.

估计式 (4.4.6) 表明, $\kappa(\mathbf{A})$ 反映了 \mathbf{A}^{-1} 对于 \mathbf{A} 的扰动的敏感性. $\kappa(\mathbf{A})$ 愈大, $(\mathbf{A} + \mathbf{E})^{-1}$ 与 \mathbf{A}^{-1} 的相对误差就愈大. 在例 4.4.1 中, 当矩阵范数取 1-范数、2-范数或 ∞ -范数时, $\kappa(\mathbf{A})$ 均等于 39 601.

定义 4.4.2 (条件数) 设 n 阶矩阵 \mathbf{A} 非奇异, 称 $\kappa(\mathbf{A}) = \|\mathbf{A}\| \|\mathbf{A}^{-1}\|$ 为 \mathbf{A} 关于求逆的条件数, 也称为求解线性方程组 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 的条件数.

注 1: 条件数就是用来衡量输出相对于输入敏感度的指标. 良态问题和病态问题就是靠这个指标进行区分的: 如果 $\kappa(\mathbf{A})$ 很大, 则矩阵 \mathbf{A} 关于求逆是病态的.

现考察线性方程组

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad (4.4.7)$$

如果系数矩阵 \mathbf{A} 和右端项 \mathbf{b} 分别有扰动 \mathbf{E} 和 $\delta\mathbf{b}$, 则扰动后方程组为

$$(\mathbf{A} + \mathbf{E})(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b} + \delta\mathbf{b} \quad (4.4.8)$$

且有如下定理:

定理 4.4.3 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 是非奇异矩阵, \mathbf{b} 和 $\delta\mathbf{b} \in \mathbb{C}^n$, \mathbf{x} 是线性方程组 (4.4.7) 的解. 如果 $\mathbf{E} \in \mathbb{C}^{n \times n}$ 满足条件 $\|\mathbf{A}^{-1}\| \|\mathbf{E}\| < 1$, 则方程组 (4.4.8) 有唯一解 $\mathbf{x} + \delta\mathbf{x}$, 且有

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\kappa(\mathbf{A})}{1 - \kappa(\mathbf{A}) \frac{\|\mathbf{E}\|}{\|\mathbf{A}\|}} \left(\frac{\|\mathbf{E}\|}{\|\mathbf{A}\|} + \frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|} \right) \quad (4.4.9)$$

证明: 因为 $\|\mathbf{A}^{-1}\| \|\mathbf{E}\| < 1$, 由推论 4.4.1 知, 矩阵 $\mathbf{A} + \mathbf{E}$ 非奇异. 此时方程组 (4.4.8)

有唯一解 $x + \delta x$.

由式(4.4.7)和式(4.4.8)分别得, $x = A^{-1}b$ 和 $x + \delta x = (A + E)^{-1}(b + \delta b)$. 于是

$$\begin{aligned}\delta x &= (A + E)^{-1}(b + \delta b) - x \\ &= (A + E)^{-1}\delta b - x + (I + A^{-1}E)^{-1}A^{-1}b \\ &= (A + E)^{-1}\delta b - [I - (I + A^{-1}E)^{-1}]x \\ &= (A + E)^{-1}\delta b - (I + A^{-1}E)^{-1}A^{-1}Ex\end{aligned}$$

由 $\|A^{-1}E\| \leq \|A^{-1}\|\|E\|$ 、定理 4.3.6 和定理 4.4.2 可得

$$\begin{aligned}\|\delta x\| &\leq \frac{\|A^{-1}\|\|\delta b\|}{1 - \|A^{-1}E\|} + \frac{\|A^{-1}E\|\|x\|}{1 - \|A^{-1}E\|} \\ &\leq \frac{\|A^{-1}\|\|x\|}{1 - \|A^{-1}\|\|E\|} \left(\frac{\|\delta b\|}{\|x\|} + \|E\| \right)\end{aligned}$$

注意到 $\|b\| \leq \|A\|\|x\|$, 故由上式得

$$\begin{aligned}\frac{\|\delta x\|}{\|x\|} &\leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\|\|E\|} \left(\frac{\|\delta b\|}{\|x\|} + \|E\| \right) \\ &\leq \frac{\kappa(A)}{1 - \kappa(A) \frac{\|E\|}{\|A\|}} \left(\frac{\|\delta b\|}{\|b\|} + \frac{\|E\|}{\|A\|} \right)\end{aligned}$$

估计式(4.4.9)表明, $\kappa(A)$ 反映了线性方程组(4.4.7)的解 x 的相对误差对于 A 和 b 的相对误差的依赖程度. 由定理 4.4.3 知, 如果 $\kappa(A)$ 很大, 则线性方程组(4.4.7)是病态的. 在求矩阵的逆或求解线性方程组时, 可以通过变换降低矩阵的条件数, 即所谓预处理或预条件.

例 4.4.2 考察线性方程组

$$\begin{bmatrix} 1 & 10^5 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 10^5 \\ 2 \end{bmatrix}$$

的求解问题.

分析: 由系数矩阵 $A = \begin{bmatrix} 1 & 10^5 \\ 1 & 1 \end{bmatrix}$, 计算得

$$A^{-1} = \frac{1}{10^5 - 1} \begin{bmatrix} -1 & 10^5 \\ 1 & -1 \end{bmatrix}$$

取矩阵行和范数, 则

$$\kappa(A) = \frac{(1 + 10^5)^2}{10^5 - 1} \approx 10^5$$

因此矩阵 A 求逆和求解方程组 $Ax = b$ 都是病态的. 此时, 可用条件数 10^5 除方程组的第一个方程, 得

$$\begin{bmatrix} 10^{-5} & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

此时 $\kappa \approx 4$, 于是上面方程组是良态的.

注 2: 条件数 $\kappa(\mathbf{A})$ 大的两个常见原因: 矩阵的列之间的相关性过大 (此时矩阵为奇异或接近奇异的, 见例 4.4.1) 和矩阵的特征值差异大 (见例 4.4.2).

4.5 特征值估计

本节讨论矩阵元素的变化对矩阵特征值的影响. 由于一元多项式的根是其系数的连续函数, 且矩阵特征多项式的系数是矩阵元素的连续函数, 因此矩阵的特征值必然是矩阵元素的连续函数.

定义 4.5.1 (谱和谱半径) 给定复方阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 记

$$S_p(\mathbf{A}) = \{\lambda \mid \lambda \text{ 是 } \mathbf{A} \text{ 的特征值}\}$$

则称 $S_p(\mathbf{A})$ 是矩阵 \mathbf{A} 的谱, 称 \mathbf{A} 的特征值模的最大值为 \mathbf{A} 的谱半径, 记为 $\rho(\mathbf{A})$.

例 4.5.1 求 $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 2 & 2i \end{bmatrix}$ 的谱半径 $\rho(\mathbf{A})$.

解: 矩阵 \mathbf{A} 的特征值为 1 和 $2i$, 所以谱半径 $\rho(\mathbf{A}) = 2$.

【思考】 矩阵的谱半径可定义矩阵范数吗?

分析: 矩阵范数的正定性要求范数等于零的充分必要条件是矩阵本身为零矩阵. 考察矩阵 $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, 其谱半径为 0, 但 \mathbf{A} 为非零矩阵. 显然, 谱半径不可以定义为矩阵范数.

例 4.5.2 计算 $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}$ 的谱半径 $\rho(\mathbf{A})$, $\|\mathbf{A}\|_1$ 、 $\|\mathbf{A}\|_\infty$ 、 $\|\mathbf{A}\|_2$ 和 $\|\mathbf{A}\|_F$.

解: $\rho(\mathbf{A}) = 3.5616$, $\|\mathbf{A}\|_1 = 4$, $\|\mathbf{A}\|_\infty = 4$, $\|\mathbf{A}\|_2 = 3.5616$, $\|\mathbf{A}\|_F = 3.605$.

由例 4.5.2 计算结果知, 矩阵的范数均不小于其谱半径. 这是有理论依据的.

定理 4.5.1 复方阵的谱半径不大于它的任一矩阵范数.

证明: 设 λ 是复方阵 \mathbf{A} 的任一特征值, \mathbf{x} 是属于 λ 的特征向量, 则 $\mathbf{Ax} = \lambda\mathbf{x}$. 对任意矩阵范数 $\|\cdot\|$, 有

$$|\lambda| \|\mathbf{x}\| = \|\lambda\mathbf{x}\| = \|\mathbf{Ax}\| \leq \|\mathbf{A}\| \|\mathbf{x}\|$$

注意到 $\mathbf{x} \neq \mathbf{0}$, 则有 $|\lambda| \leq \|\mathbf{A}\|$, 即 $\rho(\mathbf{A}) \leq \|\mathbf{A}\|$. 证毕.

定理 4.5.2 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 任取正常数 ϵ , 则必存在某个矩阵范数 $\|\cdot\|$ 使得 $\|\mathbf{A}\| \leq \rho(\mathbf{A}) + \epsilon$.

证明: 由 Jordan 标准形定理知, 存在非奇异矩阵 $\mathbf{P} \in \mathbb{C}^{n \times n}$ 使得 $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{J}$, 其中, \mathbf{J} 是矩阵 \mathbf{A} 的 Jordan 标准形, 其对角线元素为 $\lambda_1, \lambda_2, \dots, \lambda_n$.

令 $\mathbf{A} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, 并定义 $\tilde{\mathbf{I}} = \mathbf{J} - \mathbf{A}$, 则有

$$\tilde{\mathbf{I}} = \begin{bmatrix} 0 & \delta_1 & & & \\ & 0 & \delta_2 & & \\ & & \ddots & \ddots & \\ & & & 0 & \delta_{n-1} \\ & & & & 0 \end{bmatrix}$$

式中: $\delta_i = 0$ 或 $1, i=1, 2, \dots, n-1$.

对任意正常数 ϵ , 令 $\mathbf{Q} = \text{diag}(1, \epsilon, \dots, \epsilon^{n-1})$, 则

$$(\mathbf{PQ})^{-1} \mathbf{A} (\mathbf{PQ}) = \mathbf{Q}^{-1} \mathbf{JQ} = \mathbf{A} + \epsilon \tilde{\mathbf{I}}$$

令 $\mathbf{S} = \mathbf{PQ}$, 则矩阵 \mathbf{S} 可逆且有

$$\|\mathbf{S}^{-1} \mathbf{AS}\|_1 = \|\mathbf{A} + \epsilon \tilde{\mathbf{I}}\|_1 \leq \rho(\mathbf{A}) + \epsilon$$

定义 $\|\mathbf{A}\|_m = \|\mathbf{S}^{-1} \mathbf{AS}\|_1$, 容易证明 $\|\mathbf{A}\|_m$ 是矩阵范数. 因此, 结论成立.

注 1: 定理 4.5.2 中构造的矩阵范数与给定的矩阵 \mathbf{A} 有关. 因此, 当用矩阵 \mathbf{A} 构造的矩阵范数来计算矩阵 \mathbf{B} ($\mathbf{B} \neq \mathbf{A}$) 矩阵范数时, 不等式 $\|\mathbf{B}\| \leq \rho(\mathbf{B}) + \epsilon$ 不一定成立.

定理 4.5.1 和定理 4.5.2 从不同角度提供了谱半径和矩阵范数之间的关系. 我们常用定理 4.5.1 来粗略估算矩阵的特征值, 通常选取易于计算的矩阵列和范数、行和范数或 F-范数, 此时可将矩阵的所有特征值限定在复平面的某一圆盘内. 不难发现, 这种估计方法简单但也比较粗糙. 为此, 引入盖尔圆盘定义和定理来改进特征值的估计方法.

定义 4.5.2 (盖尔圆盘) 设 $\mathbf{A} = (a_{ij}) \in \mathbb{C}^{n \times n}$, 令

$$\delta_i = \sum_{j=1, j \neq i}^n |a_{ij}|, \quad i=1, \dots, n$$

并定义

$$G_i = \{z \in \mathbb{C} \mid |z - a_{ii}| \leq \delta_i\}, \quad i=1, \dots, n$$

即 G_i 是复平面上以 a_{ii} 为圆心, δ_i 为半径的闭圆盘, 称为矩阵 \mathbf{A} 的一个盖尔圆盘.

例 4.5.3 计算矩阵 $\mathbf{A} = \begin{bmatrix} 1 & 0.02 & 0.11 \\ 0.01 & i & 0.14 \\ 0.02 & 0.01 & 0.5 \end{bmatrix}$ 的盖尔圆盘.

解: \mathbf{A} 的三个盖尔圆盘(见图 4.5.1)分别为

$$G_1 = \{z \in \mathbb{C} \mid |z - 1| \leq 0.13\}$$

$$G_2 = \{z \in \mathbb{C} \mid |z - i| \leq 0.15\}$$

$$G_3 = \{z \in \mathbb{C} \mid |z - 0.5| \leq 0.03\}$$

定理 4.5.3 (盖尔圆盘定理) 设 $\mathbf{A} = (a_{ij}) \in \mathbb{C}^{n \times n}$ 的 n 个盖尔圆盘为 G_1, \dots, G_n ,

则矩阵 \mathbf{A} 的任一特征值 $\lambda \in \bigcup_{i=1}^n G_i$.

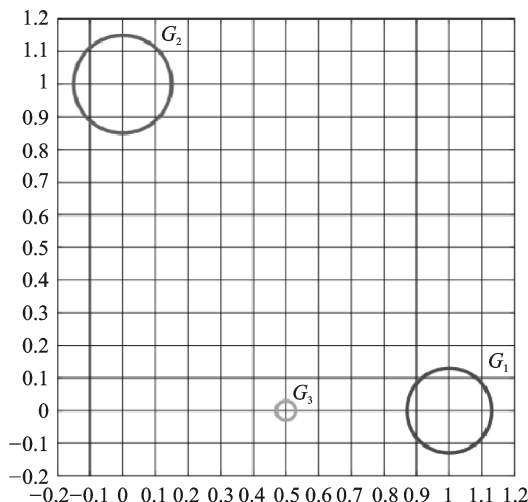


图 4.5.1 例 4.5.3 矩阵 A 的盖尔圆盘图示

证明: 设 λ 是矩阵 A 的任一特征值, $x = [x_1, \dots, x_n]^T$ 是属于特征值 λ 的特征向量, 则有

$$a_{i1}x_1 + \dots + a_{in}x_n = \lambda x_i, \quad i = 1, \dots, n \quad (4.5.1)$$

整理式(4.5.1)得

$$(\lambda - a_{ii})x_i = \sum_{k=1, k \neq i}^n a_{ik}x_k \quad (4.5.2)$$

定义

$$|x_\sigma| = \max_{1 \leq i \leq n} |x_i| > 0, \quad \sigma \in \{1, \dots, n\}$$

并考察式(4.5.2)的第 σ 个方程, 得

$$|\lambda - a_{\sigma\sigma}| |x_\sigma| = \left| \sum_{k=1, k \neq \sigma}^n a_{\sigma k} x_k \right|$$

注意到 $|x_\sigma| > 0$, 则上式可改写为

$$\begin{aligned} |\lambda - a_{\sigma\sigma}| &= \left| \sum_{k=1, k \neq \sigma}^n a_{\sigma k} \frac{x_k}{|x_\sigma|} \right| \\ &\leq \sum_{k=1, k \neq \sigma}^n \left(|a_{\sigma k}| \left| \frac{x_k}{|x_\sigma|} \right| \right) \leq \sum_{k=1, k \neq \sigma}^n |a_{\sigma k}| \end{aligned}$$

上式表明, λ 必在 A 的 G_σ 这一盖尔圆中. 因此, A 的任一特征值必在 $\bigcup_{i=1}^n G_i$ 内. 证毕.

例 4.5.4 试估计矩阵 $A = \begin{bmatrix} 2 & -1 & -2 & 0 \\ -1 & 3 & 2i & 0 \\ 0 & -i & 10 & i \\ -2 & 0 & 0 & 6i \end{bmatrix}$ 的特征值分布.

解: 由盖尔圆盘定理知, \mathbf{A} 的特征值 $\lambda \in \bigcup_{i=1}^4 G_i$, 其中 $G_1: |z-2| \leq 3$; $G_2: |z-3| \leq 3$; $G_3: |z-10| \leq 2$; $G_4: |z-6i| \leq 2$, 如图 4.5.2 所示.

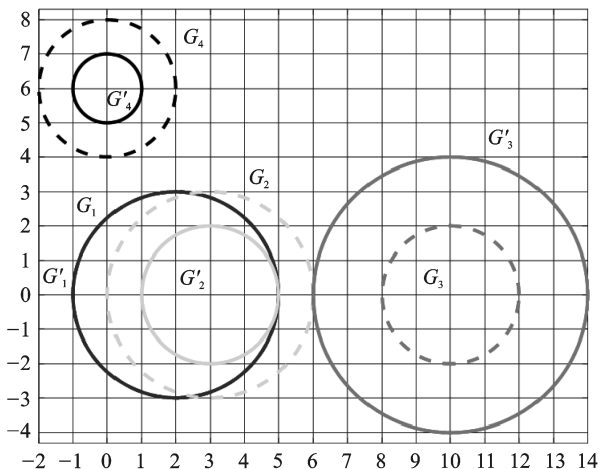


图 4.5.2 例 4.5.4 矩阵 \mathbf{A} 与 \mathbf{A}^T 的盖尔圆盘

由于 \mathbf{A}^T 与 \mathbf{A} 具有相同的特征值, 故我们还可以利用 \mathbf{A}^T 来估计 \mathbf{A} 的特征值. 计算 \mathbf{A}^T 的 4 个盖尔圆分别为: $G'_1: |z-2| \leq 3$; $G'_2: |z-3| \leq 2$; $G'_3: |z-10| \leq 4$; $G'_4: |z-6i| \leq 1$. 根据盖尔圆盘定理可知, 矩阵 \mathbf{A} 的 4 个特征值必在 $\bigcup_{i=1}^4 G'_i$ 中 (见图 4.5.2).

观察图 4.5.2 发现, \mathbf{A}^T 与 \mathbf{A} 的盖尔圆盘不尽相同, 因此矩阵 \mathbf{A} 的特征值满足 $\lambda \in (\bigcup_{i=1}^n G_i) \cap (\bigcup_{i=1}^n G'_i)$. 这表明, 我们可以结合矩阵的转置操作来提高特征值的估算精度.

注 2: 设 \mathbf{A}^T 的盖尔圆为 G'_1, \dots, G'_n , 则 G_i 与 G'_i 有相同的圆心. 因此, 矩阵 \mathbf{A} 的特征值必满足

$$\lambda \in (\bigcup_{i=1}^n G_i) \cap (\bigcup_{i=1}^n G'_i)$$

【思考】 n 阶复方阵 \mathbf{A} 的 n 个特征值是否恰好落入它的 n 个盖尔圆内?

例 4.5.5 考察矩阵 $\mathbf{A} = \begin{bmatrix} 0 & -0.4 \\ 0.9 & 1 \end{bmatrix}$, 其盖尔圆盘为

$$G_1: |z| \leq 0.4, \quad G_2: |z-1| \leq 0.9$$

经解析计算矩阵 \mathbf{A} 的特征值为 $\lambda_{1,2} = \frac{1 \pm i\sqrt{0.44}}{2}$. 由图 4.5.3 看出, 矩阵 \mathbf{A} 的所有特征值均在 G_2 内, 而 G_1 未包含任何特征值. 所以, 并非每个盖尔圆内都恰好有一个特征值.

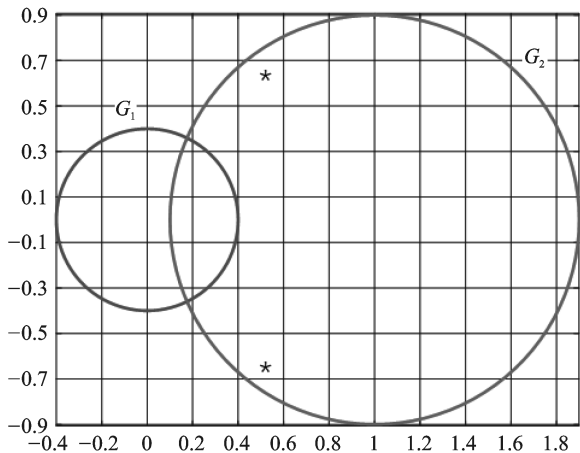


图 4.5.3 例 4.5.5 矩阵 A 的盖尔圆盘

定理 4.5.4 (盖尔圆盘定理续) 设 $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ 的盖尔圆盘为 G_1, \dots, G_n , 若其中的 k 个盖尔圆盘的并集形成一个连通的区域, 且该区域与其余 $n-k$ 个圆盘都不相交, 则此连通域内恰有 k 个特征值. 特别地, 孤立的盖尔圆内有且仅有一个特征值.

证明: 设 $D = \text{diag}(a_{11}, \dots, a_{nn}) \in \mathbb{C}^{n \times n}$ 和 $B = A - D$, 并令 $A(\epsilon) = D + \epsilon B$, 其中, $\epsilon \in [0, 1]$. 显然, $A(0) = D$, $A(1) = A$, 且 $A(\epsilon)$ 与 A 的盖尔圆有相同的圆心, 但前者的半径是后者的 ϵ 倍. 由于 $\epsilon \in [0, 1]$, 故 $A(\epsilon)$ 的任一盖尔圆盘都在 A 的相应盖尔圆内.

当 ϵ 从 0 连续地变为 1 时, 矩阵 $A(\epsilon)$ 的 n 个特征值将连续变化, 即特征值函数 (以 ϵ 为自变量) 在复平面上是 n 条连续的曲线. 每条曲线的起点分别为 A (或 D) 的对角元素, 即某一盖尔圆的圆心, 曲线终点为 A 的某一特征值. 由定理 4.5.3 知, 这 n 条连续曲线不能超出所有的盖尔圆. 因此, A 的 k 个盖尔圆所围的连通区域内有且只有 k 条曲线, 即有且只有 A 的 k 个特征值. 当 $k=1$ 时, 此连通域内有且仅有一个特征值. 证毕.

推论 4.5.1 设矩阵 $A \in \mathbb{C}^{n \times n}$ 的 n 个盖尔圆盘为 G_1, \dots, G_n , 若原点 $0 \notin \bigcup_{i=1}^n G_i$, 则矩阵 A 为非奇异矩阵.

推论 4.5.2 设 $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ 是对角占优矩阵, 即对 $i=1, \dots, n$, 有

$$|a_{ii}| > \sum_{i=1, j \neq i}^n |a_{ij}| \quad (\text{列对角占优})$$

或

$$|a_{ii}| > \sum_{j=1, j \neq i}^n |a_{ij}| \quad (\text{行对角占优})$$

则矩阵 A 非奇异.

推论 4.5.3 若复方阵 A 有 k 个孤立的盖尔圆, 则它至少有 k 个互异特征值. 特

别地,若矩阵 A 的所有盖尔圆两两互不相交,则 A 是单纯矩阵.

推论 4.5.4 若实方阵 A 有 k 个孤立的盖尔圆,则它至少有 k 个互异的实特征值.特别地,若矩阵 A 的所有盖尔圆两两互不相交,则它有 n 个互异的实特征值.

例 4.5.6 证明 n 阶矩阵 A 是单纯矩阵,其中

$$A = \begin{bmatrix} 2 & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \frac{1}{n} & 4 & \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & 2n \end{bmatrix}$$

证明: 矩阵 A 的 n 个盖尔圆为

$$\begin{aligned} G_1: |z-2| &\leq 1 - \frac{1}{n} \\ G_2: |z-4| &\leq 1 - \frac{1}{n} \\ &\vdots \\ G_n: |z-2n| &\leq 1 - \frac{1}{n} \end{aligned}$$

由于矩阵 A 的任意两个盖尔圆的圆心距都大于或等于 2,而它的任一盖尔圆的半径小于 1,故 A 的 n 个盖尔圆均是孤立的.由盖尔圆盘定理知,矩阵 A 有 n 个互异的实特征值.因此, A 是单纯矩阵.

例 4.5.7 证明矩阵 A 至少有两个实特征值,其中

$$A = \begin{bmatrix} 9 & 1 & -2 & 1 \\ 0 & 8 & 1 & 1 \\ -1 & 0 & 4 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

证明: 矩阵 A 的 4 个盖尔圆为 $G_1: |z-9| \leq 4; G_2: |z-8| \leq 2; G_3: |z-4| \leq 1$ 和 $G_4: |z-1| \leq 1$. 如图 4.5.4 所示, G_4 为孤立圆,表明矩阵 A 至少有一个实特征值; $G_1 \cup G_2 \cup G_3$ 中含有矩阵 A 的另外 3 个特征值,这表明矩阵 A 要么有 3 个实特征值,要么有一对共轭复根和 1 个实特征值.因此,矩阵 A 至少有两个实特征值.

注 3: 在使用盖尔圆估计矩阵 A 的特征值时,我们总希望获得更多的孤立圆,这时可采取如下方法:

取合适的非零实数 d_1, \cdots, d_n , 并令 $D = \text{diag}(d_1, \cdots, d_n)$, 则

$$B = DAD^{-1} = \left(a_{ij} \frac{d_i}{d_j} \right)_{n \times n}$$

显然,矩阵 A 与 B 相似,它们具有相同的特征值.我们可以依据矩阵 B 的盖尔圆

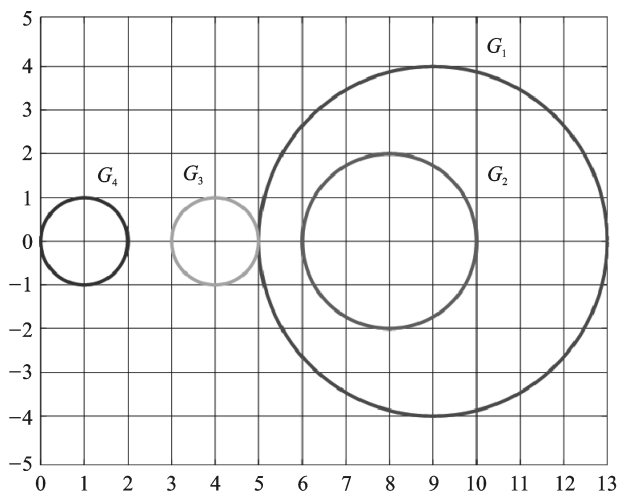


图 4.5.4 例 4.5.7 矩阵 A 的盖尔圆盘

来估计矩阵 A 的特征值. 通常 d_i 的选取办法如下:

(1) 若取 $d_i < 1$, 其余元素为 1, 则第 i 个盖尔圆 G_i 会缩小, 其余所有盖尔圆会放大;

(2) 若取 $d_i > 1$, 其余元素为 1, 则第 i 个盖尔圆 G_i 会放大, 而其余所有盖尔圆会缩小.

例 4.5.8 估计复方阵 $A = \begin{bmatrix} 9 & 1 & 1 \\ 1 & i & 1 \\ 1 & 1 & 3 \end{bmatrix}$ 的特征值分布范围.

解: 矩阵 A 的 3 个盖尔圆为 $G_1: |z-9| \leq 2, G_2: |z-i| \leq 2, G_3: |z-3| \leq 2$. 若选取

$$D = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

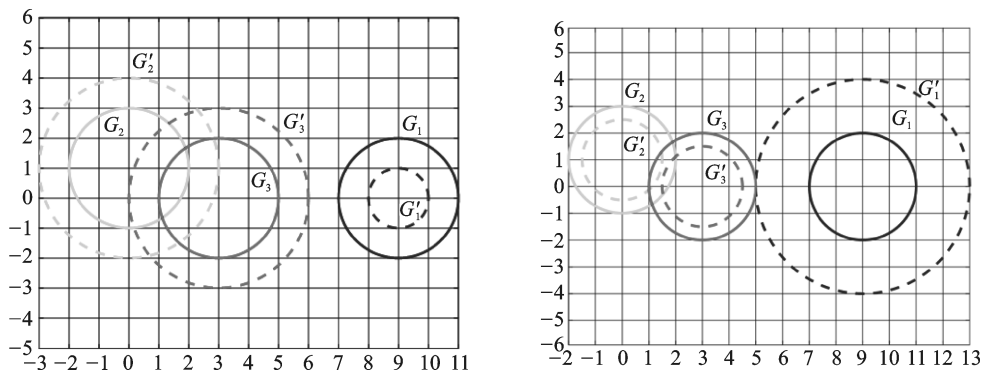
则有

$$B = DAD^{-1} = \begin{bmatrix} 9 & 0.5 & 0.5 \\ 2 & i & 1 \\ 2 & 1 & 3 \end{bmatrix}$$

此时, 矩阵 B 的 3 个盖尔圆盘分别为 $G'_1: |z-9| \leq 1, G'_2: |z-i| \leq 3, G'_3: |z-3| \leq 3$.

由图 4.5.5 左图中看出 G'_1 缩小, 而其余的盖尔圆盘 G'_2 和 G'_3 在放大.

若选取

图 4.5.5 例 4.5.8 矩阵 B 的盖尔圆盘

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

则有

$$B = DAD^{-1} = \begin{bmatrix} 9 & 2 & 2 \\ 0.5 & i & 1 \\ 0.5 & 1 & 3 \end{bmatrix}$$

此时, B 的 3 个盖尔圆为 $G'_1: |z-9| \leq 4, G'_2: |z-i| \leq 1.5, G'_3: |z-3| \leq 1.5$. 由图 4.5.5 右图知, 尽管 G'_1 放大, 但 G'_2 和 G'_3 在缩小, 且变成了两个孤立的盖尔圆. 此时说明 A 是单纯矩阵.

例 4.5.9 利用盖尔圆盘定理考察矩阵 $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ 的特征值分布.

解: 取 $D = \text{diag}(k, 1)$, 则有

$$B = DAD^{-1} = \begin{bmatrix} 0 & -k \\ \frac{1}{k} & 0 \end{bmatrix}$$

矩阵 B 的盖尔圆为 $G_1: |z| \leq k, G_2: |z| \leq \frac{1}{k}$. 当 $k > 1$ 时, $G_1 = G_1 \cup G_2$; 当 $k < 1$ 时, $G_2 = G_1 \cup G_2$. 无论何种情况, B 的盖尔圆都包含 A 的盖尔圆. 这说明对于某些特殊的矩阵, 利用对角阵 D 来改变矩阵的盖尔圆的这一方法会失效.

对于 Hermite 矩阵的特征值估计可结合定理 4.3.4 和定理 4.3.5 进行讨论.

4.6 矩阵级数

与数学分析一样, 矩阵分析理论的建立也是以极限理论为基础的, 其内容丰富,

是研究许多工程问题的重要工具. 本节首先讨论向量序列敛散性问题, 然后推广至矩阵序列的极限运算, 最后介绍矩阵级数.

定义 4.6.1(向量序列按范数收敛) 设 $(V, \|\cdot\|_a)$ 是 n 维赋范线性空间, $x_1, x_2, \dots, x_k, \dots$ 是 V 中一个向量序列, 记为 $\{x_k\}$. 若存在 V 的向量 x 满足

$$\lim_{k \rightarrow \infty} \|x_k - x\|_a = 0$$

则称向量序列 $\{x_k\}$ 按范数 $\|\cdot\|_a$ 收敛于 x , 记作

$$\lim_{k \rightarrow \infty} x_k = x \quad \text{或} \quad x_k \xrightarrow{\alpha} x$$

不收敛的向量序列称为发散的.

例 4.6.1 设 $V = \mathbb{R}$, $\|\cdot\|_a$ 取为绝对值, 则定义 4.6.1 中敛散性与高等数学中的相关概念一致.

定理 4.6.1 设 $(V, \|\cdot\|)$ 是 n 维赋范线性空间, $\{x_k\}$ 是 V 的一个向量序列. 若序列 $\{x_k\}$ 按某种范数收敛于 x , 则序列 $\{x_k\}$ 按任意范数收敛于 x , 即有限维空间中按范数收敛是等价的.

证明: 设 $\|\cdot\|_a$ 和 $\|\cdot\|_\beta$ 是 V 中任意两种范数, 则对 V 中任意向量 y 存在正常数 k 使得

$$\|y\|_\beta \leq k \|y\|_a$$

由此,

$$0 \leq \|x_k - x\|_\beta \leq k \|x_k - x\|_a$$

若 $\lim_{k \rightarrow \infty} \|x_k - x\|_a = 0$, 则由夹逼定理知

$$\lim_{k \rightarrow \infty} \|x_k - x\|_\beta = 0$$

即序列 $\{x_k\}$ 按 β 范数收敛于 x , 反之亦然.

定义 4.6.2(向量序列按坐标收敛) 设 $(V, \|\cdot\|_a)$ 是 n 维赋范线性空间, ζ_1, \dots, ζ_n 是 V 中一组基, $\{x_k\}$ 是 V 中一个向量序列, 并记向量序列 $\{x_k\}$ 中的任一向量 x_k 在基 ζ_1, \dots, ζ_n 下坐标为 $\xi_k = [\xi_1^{(k)}, \dots, \xi_n^{(k)}]^T \in F^n$. 若存在向量 $\xi = [\xi_1, \dots, \xi_n]^T \in F^n$ 满足

$$\lim_{k \rightarrow \infty} \xi_i^{(k)} = \xi_i, \quad i = 1, \dots, n$$

则称向量序列 $\{x_k\}$ 按坐标收敛于 x , 其中 ξ 是向量 x 在基 ζ_1, \dots, ζ_n 下坐标.

定理 4.6.2 设 $(V, \|\cdot\|)$ 是 n 维赋范线性空间, $\{x_k\}$ 是 V 中一个向量序列且 $x \in V$. 向量序列 $\{x_k\}$ 按范数收敛于向量 x 当且仅当它按坐标收敛于 x .

证明: 根据式(4.1.4)定义向量范数

$$\|x_k\| = \sqrt{\sum_{i=1}^n |\xi_i^{(k)}|^2}$$

则对任意 i 有

$$\lim_{k \rightarrow \infty} \|x_k - x\| = 0 \Leftrightarrow \lim_{k \rightarrow \infty} \|\xi_k - \xi\| = 0 \Leftrightarrow \lim_{k \rightarrow \infty} \xi_i^{(k)} = \xi_i$$

上式表明,序列 $\{\mathbf{x}_k\}$ 按范数收敛于 \mathbf{x} 当且仅当它按坐标收敛于 \mathbf{x} .证毕.

将 $m \times n$ 矩阵看作 mn 维向量即可定义矩阵序列的敛散性.

定义 4.6.3(矩阵序列按坐标收敛) 设矩阵序列 $\{\mathbf{A}_k\}$,其中矩阵 $\mathbf{A}_k = (a_{ij}^{(k)}) \in \mathbb{C}^{m \times n}$, $k=1,2,\dots$, 如果当 $k \rightarrow \infty$ 时,矩阵 \mathbf{A}_k 的每一个元素 $a_{ij}^{(k)}$ 都有极限 $a_{ij}^{(0)}$,即

$$\lim_{k \rightarrow \infty} a_{ij}^{(k)} = a_{ij}^{(0)}, \quad 1 \leq i \leq m, 1 \leq j \leq n$$

则称矩阵序列按元素收敛或按坐标收敛(或简称为矩阵序列 $\{\mathbf{A}_k\}$ 收敛), $\mathbf{A}_0 = (a_{ij}^{(0)})$ 称为序列的极限,记为 $\lim_{k \rightarrow \infty} \mathbf{A}_k = \mathbf{A}_0$.

例 4.6.2 设 $\mathbf{A}_k = \begin{bmatrix} \frac{2k^2+k+1}{k^2} & \frac{\sin k}{k} \\ e^{-k} \sin k & \left(1+\frac{1}{2k}\right)^k \end{bmatrix}$, 求极限 $\lim_{k \rightarrow \infty} \mathbf{A}_k$.

解: $\lim_{k \rightarrow \infty} a_{11}^{(k)} = 2, \lim_{k \rightarrow \infty} a_{12}^{(k)} = 0, \lim_{k \rightarrow \infty} a_{21}^{(k)} = 0, \lim_{k \rightarrow \infty} a_{22}^{(k)} = e^{\frac{1}{2}}$. 因此,

$$\lim_{k \rightarrow \infty} \mathbf{A}_k = \begin{bmatrix} 2 & 0 \\ 0 & e^{\frac{1}{2}} \end{bmatrix}$$

命题 4.6.1 设矩阵序列 $\{\mathbf{A}_k\}$ 和 $\{\mathbf{B}_k\}$ 分别收敛于矩阵 \mathbf{A} 和 \mathbf{B} ,则对任意 $c_1, c_2 \in \mathbb{C}$,有

- (1) $\lim_{k \rightarrow \infty} (c_1 \mathbf{A}_k + c_2 \mathbf{B}_k) = c_1 \mathbf{A} + c_2 \mathbf{B}$, 其中 $\mathbf{A}_k, \mathbf{B}_k \in \mathbb{C}^{m \times n}$;
- (2) $\lim_{k \rightarrow \infty} (\mathbf{A}_k \mathbf{B}_k) = \mathbf{A} \mathbf{B}$, 其中 $\mathbf{A}_k \in \mathbb{C}^{m \times n}, \mathbf{B}_k \in \mathbb{C}^{n \times p}$;
- (3) 若 \mathbf{A}_k 和 \mathbf{A} 为可逆矩阵,则 $\lim_{k \rightarrow \infty} \mathbf{A}_k^{-1} = \mathbf{A}^{-1}$.

推论 4.6.1 设 $\|\cdot\|$ 是 $\mathbb{C}^{m \times n}$ 上任一矩阵范数, $\mathbb{C}^{m \times n}$ 中矩阵序列 $\{\mathbf{A}_k\}$ 收敛于矩阵 \mathbf{A}_0 的充分必要条件是 $\lim_{k \rightarrow \infty} \|\mathbf{A}_k - \mathbf{A}_0\| = 0$.

推论 4.6.1 表明矩阵序列 $\{\mathbf{A}_k\}$ 按坐标收敛于 \mathbf{A}_0 当且仅当它按范数收敛于 \mathbf{A}_0 .

推论 4.6.2 复方阵 \mathbf{A} 的某一范数满足 $\|\mathbf{A}\| < 1$,则 $\lim_{k \rightarrow \infty} \mathbf{A}^k = \mathbf{O}$.

证明: 由不等式 $\|\mathbf{A}^k\| \leq \|\mathbf{A}\|^k$ 知,当 $k \rightarrow \infty$ 时, $\|\mathbf{A}^k\| \rightarrow 0$. 根据推论 4.6.1 得,
 $\lim_{k \rightarrow \infty} \mathbf{A}^k = \mathbf{O}$.

【思考】 若 $\lim_{k \rightarrow \infty} \mathbf{A}^k = \mathbf{O}$,则 $\|\mathbf{A}\| < 1$ 成立吗?

分析: 考察矩阵 $\mathbf{A} = \begin{bmatrix} 0.1 & 0 \\ 1 & 0.1 \end{bmatrix}$,显然, $\|\mathbf{A}\|_1 = 1.1 > 1$. 又知

$$\mathbf{A}^k = \begin{bmatrix} 0.1^k & 0 \\ k 0.1^{k-1} & 0.1^k \end{bmatrix}$$

当 $k \rightarrow \infty, \mathbf{A}^k \rightarrow \mathbf{O}$. 因此,上述结论不成立.

定理 4.6.3 设 $\mathbf{A} = (a_{ij}) \in \mathbb{C}^{n \times n}, \lim_{k \rightarrow \infty} \mathbf{A}^k = \mathbf{O}$ 的充分必要条件是 $\rho(\mathbf{A}) < 1$.

证明: 对任一 n 阶矩阵 \mathbf{A} ,必存在可逆矩阵 \mathbf{P} 使得 $\mathbf{P}^{-1} \mathbf{A} \mathbf{P} = \mathbf{J}$,其中 \mathbf{J} 为矩阵 \mathbf{A}

的 Jordan 标准形, 可表示为 $J = \text{diag}(J_1, \dots, J_s)$, $J_i (i=1, \dots, s)$ 为矩阵 A 的 Jordan 块, 可表示为

$$J_i = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_i & 1 \\ \mathbf{O} & & & & \lambda_i \end{bmatrix}_{n_i \times n_i}$$

此时, $A^k = PJ^kP^{-1}$, 式中, $J^k = \text{diag}(J_1^k, \dots, J_s^k)$, $J_i^k (i=1, \dots, s)$ 满足

$$J_i^k = \begin{bmatrix} \lambda_i^k & C_k^1 \lambda_i^{k-1} & C_k^2 \lambda_i^{k-2} & \cdots & C_k^{n_i-1} \lambda_i^{k-n_i+1} \\ & \lambda_i^k & C_k^1 \lambda_i^{k-1} & \cdots & C_k^{n_i-2} \lambda_i^{k-n_i+2} \\ & & \lambda_i^k & \ddots & \vdots \\ & & & \ddots & C_k^1 \lambda_i^{k-1} \\ \mathbf{O} & & & & \lambda_i^k \end{bmatrix}$$

当 $k \rightarrow \infty$ 时,

$$\lim_{k \rightarrow \infty} A^k = \mathbf{O} \Leftrightarrow \lim_{k \rightarrow \infty} J^k = \mathbf{O} \Leftrightarrow \lim_{k \rightarrow \infty} J_i^k = \mathbf{O}, \quad \forall i \Leftrightarrow \rho(A) < 1$$

例 4.6.3 设 $A = \begin{bmatrix} c & 2c \\ 3c & 2c \end{bmatrix}$, 其中 c 为实数. 求 $\lim_{k \rightarrow \infty} A^k = \mathbf{O}$ 的充分必要条件.

解: 由矩阵 A 的特征多项式 $f(\lambda) = (\lambda - 4c)(\lambda + c)$ 得, A 的特征值为 $\lambda_1 = 4c$, $\lambda_2 = -c$. 故 A 的谱半径 $\rho(A) = |4c|$. 为使 $\lim_{k \rightarrow \infty} A^k = \mathbf{O}$ 成立, 须满足 $\rho(A) = |4c| < 1$, 即 $-\frac{1}{4} < c < \frac{1}{4}$.

推论 4.6.3 若复方阵 A 的某一矩阵范数满足 $\|A\| < 1$, 则矩阵序列 $\{A^k\}$ 收敛于零矩阵.

定义 4.6.4 (矩阵级数) 设矩阵序列 $\{A_k \in \mathbb{C}^{m \times n}\}$, 称 $\sum_{k=1}^{\infty} A_k$ 为矩阵级数. 令 $S_N = \sum_{k=1}^N A_k$, 称 S_N 为矩阵级数的部分和. 若矩阵序列 $\{S_N\}$ 收敛且有极限 S , 即 $\lim_{N \rightarrow \infty} S_N = S$, 则称矩阵级数 $\sum_{k=1}^{\infty} A_k$ 收敛且有和 S . 不收敛的矩阵级数称为发散级数.

例 4.6.4 设 $A_k = \begin{bmatrix} \frac{1}{k(k+1)} & 0 \\ 0 & \frac{2k}{3^k} \end{bmatrix}, k=1, 2, \dots$. 求级数 $\sum_{k=1}^{\infty} A_k$.

解: 首先判断级数 $\sum_{k=1}^{\infty} A_k$ 的敛散性. 令 $a_{11}^{(k)} = \frac{1}{k(k+1)}, a_{22}^{(k)} = \frac{2k}{3^k}$, 则有

$$\sum_{k=1}^{\infty} a_{11}^{(k)} = \sum_{k=1}^{\infty} \left(\frac{1}{k} - \frac{1}{k+1} \right) = 1$$

$$\sum_{k=1}^{\infty} a_{22}^{(k)} = \sum_{k=1}^{\infty} \frac{2k}{3^k}$$

回顾当 $|x| < 1$ 时,

$$\sum_{k=1}^{\infty} kx^k = \frac{x}{(1-x)^2}$$

进而,

$$\sum_{k=1}^{\infty} \frac{2k}{3^k} = 2 \sum_{k=1}^{\infty} k \left(\frac{1}{3} \right)^k = \frac{3}{2}$$

因此, $\sum_{k=1}^{\infty} \mathbf{A}_k = \text{diag} \left(1, \frac{3}{2} \right)$.

命题 4.6.2 设 $\mathbf{A}_k \in \mathbb{C}^{m \times n}$, 则以下命题成立:

- (1) $\sum_{k=1}^{\infty} \mathbf{A}_k$ 收敛 $\Leftrightarrow mn$ 个数值级数 $\sum_{k=1}^{\infty} a_{ij}^{(k)}$ 收敛;
- (2) $\sum_{k=1}^{\infty} \mathbf{A}_k$ 发散 $\Leftrightarrow mn$ 个数值级数 $\sum_{k=1}^{\infty} a_{ij}^{(k)}$ 中至少有一个级数发散;
- (3) $\sum_{k=1}^{\infty} \mathbf{A}_k$ 收敛 $\Rightarrow \lim_{k \rightarrow \infty} \mathbf{A}_k = \mathbf{O}$.

定义 4.6.5(矩阵级数绝对收敛) 设 $\mathbf{A}_k \in \mathbb{C}^{m \times n}$, 若矩阵级数 $\sum_{k=1}^{\infty} \mathbf{A}_k$ 所对应的 mn 个数值级数 $\sum_{k=1}^{\infty} a_{ij}^{(k)}, 1 \leq j \leq n, 1 \leq i \leq m$ 均绝对收敛, 则称 $\sum_{k=1}^{\infty} \mathbf{A}_k$ 绝对收敛.

命题 4.6.3 设 $\mathbf{A}_k \in \mathbb{C}^{m \times n}$, 则以下命题成立:

- (1) 若 $\sum_{k=1}^{\infty} \mathbf{A}_k$ 绝对收敛, 则 $\sum_{k=1}^{\infty} \mathbf{A}_k$ 收敛. 但 $\sum_{k=1}^{\infty} \mathbf{A}_k$ 收敛并不蕴含 $\sum_{k=1}^{\infty} \mathbf{A}_k$ 绝对收敛.
- (2) 若 $\sum_{k=1}^{\infty} \mathbf{A}_k$ 绝对收敛于 \mathbf{S} , 对 $\sum_{k=1}^{\infty} \mathbf{A}_k$ 任意重组重排得矩阵级数 $\sum_{k=1}^{\infty} \mathbf{B}_k$, 则 $\sum_{k=1}^{\infty} \mathbf{B}_k$ 绝对收敛于 \mathbf{S} .

(3) 对任意常矩阵 $\mathbf{P} \in \mathbb{C}^{p \times m}$ 和 $\mathbf{Q} \in \mathbb{C}^{n \times q}$, 若矩阵级数 $\sum_{k=1}^{\infty} \mathbf{A}_k$ (绝对) 收敛, 则矩阵级数 $\sum_{k=1}^{\infty} \mathbf{P} \mathbf{A}_k \mathbf{Q}$ (绝对) 收敛; 反之, 则不然.

例 4.6.5 考察矩阵级数 $\sum_{k=1}^{\infty} \mathbf{A}_k$ 的敛散性, 其中

$$\mathbf{A}_k = \begin{bmatrix} (-1)^{k+1} & \frac{1}{k} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

解: $\sum_{k=1}^{\infty} (-1)^{k+1} \frac{1}{k}$ 是交错调和级数, 故该级数收敛但不绝对收敛. 因此, $\sum_{k=1}^{\infty} \mathbf{A}_k$ 收敛但不绝对收敛.

定理 4.6.4 设 $\mathbf{A}_k = (a_{ij}^{(k)}) \in \mathbb{C}^{m \times n}$, 矩阵级数 $\sum_{k=1}^{\infty} \mathbf{A}_k$ 绝对收敛当且仅当对任意矩阵范数, 数值级数 $\sum_{k=1}^{\infty} \|\mathbf{A}_k\|$ 收敛.

证明: 充分性. 由级数 $\sum_{k=1}^{\infty} \|\mathbf{A}_k\|_1$ 收敛知,

$$\lim_{N \rightarrow \infty} \sum_{k=1}^N \left(\max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}^{(k)}| \right) = 0$$

上式表明

$$\lim_{N \rightarrow \infty} \sum_{k=1}^N |a_{ij}^{(k)}| = 0, \quad 1 \leq j \leq n, 1 \leq i \leq m$$

即矩阵级数 $\sum_{k=1}^{\infty} \mathbf{A}_k$ 绝对收敛.

必要性. 若矩阵级数 $\sum_{k=1}^{\infty} \mathbf{A}_k$ 绝对收敛, 则容易证明 $\sum_{k=1}^{\infty} \|\mathbf{A}_k\|_1$ 收敛. 由范数等价性知, 对任意矩阵范数 $\|\cdot\|$ 存在正常数 k_1 满足 $\|\mathbf{A}_k\|_1 \geq k_1 \|\mathbf{A}_k\|$, 进而

$$0 \leq \sum_{k=1}^{\infty} \|\mathbf{A}_k\| \leq \frac{1}{k_1} \sum_{k=1}^{\infty} \|\mathbf{A}_k\|_1$$

由比较原理得到 $\sum_{k=1}^{\infty} \|\mathbf{A}_k\|$ 收敛. 证毕.

例 4.6.6 考察矩阵级数 $\sum_{k=1}^{\infty} \mathbf{A}_k$ 的敛散性, 其中

$$\mathbf{A}_k = \begin{bmatrix} \frac{1}{3^k} & 0 \\ 0 & \frac{1}{k(k+1)} \end{bmatrix}$$

解: 由 $\|\mathbf{A}_k\|_1 = \frac{1}{k(k+1)}$ 知, 数值级数 $\sum_{k=1}^{\infty} \|\mathbf{A}_k\|_1$ 收敛, 故矩阵级数 $\sum_{k=1}^{\infty} \mathbf{A}_k$ 绝对收敛.

倘若

$$\mathbf{A}_k = \begin{bmatrix} (-1)^{k-1} \frac{1}{k} & 0 \\ 0 & \frac{1}{k(k+1)} \end{bmatrix}$$

则 $\|\mathbf{A}_k\|_1 = \frac{1}{k}$. 由于调和级数发散, 故矩阵级数 $\sum_{k=1}^{\infty} \mathbf{A}_k$ 不绝对收敛. 此时无法利用定

理 4.6.4 判定矩阵级数 $\sum_{k=1}^{\infty} \mathbf{A}_k$ 是否收敛, 只能利用定义判断该矩阵范数是收敛但不是绝对收敛的.

定义 4.6.6 (矩阵幂级数) 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 定义矩阵级数 $\sum_{m=0}^{\infty} c_m \mathbf{A}^m$, 其中 $\mathbf{A}^0 = \mathbf{I}$, 则称该级数为矩阵幂级数.

对于复变量幂级数, 我们常用到 Abel 定理判定其敛散性. 这里简单回顾一下.

引理 4.6.1 设 $z \in \mathbb{C}$, 若幂级数 $\sum_{m=0}^{\infty} c_m z^m$ 在 $z = z_0$ 点收敛, 则对满足不等式 $|z| < |z_0|$ 的幂级数都绝对收敛. 反之, 若幂级数 $\sum_{m=0}^{\infty} c_m z^m$ 在 $z = z_0$ 点发散, 则对于 $|z| > |z_0|$ 的幂级数都发散.

若存在非负实数或无穷大数 r 满足 $|z| < r$ 时, 幂级数 $\sum_{m=0}^{\infty} c_m z^m$ 收敛; 而 $|z| > r$ 时, 幂级数 $\sum_{m=0}^{\infty} c_m z^m$ 都发散, 则称 r 为收敛半径.

定理 4.6.5 (Abel 型定理) 设复变量幂级数 $\sum_{m=0}^{\infty} c_m z^m$ 的收敛半径为 r , 矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 的谱半径为 $\rho(\mathbf{A})$, 则

(1) 当 $\rho(\mathbf{A}) < r$, $\sum_{m=0}^{\infty} c_m \mathbf{A}^m$ 绝对收敛;

(2) 当 $\rho(\mathbf{A}) > r$, $\sum_{m=0}^{\infty} c_m \mathbf{A}^m$ 发散.

证明: 当 $\rho(\mathbf{A}) < r$ 时, 必存在正常数 ϵ 使得 $\rho(\mathbf{A}) + \epsilon < r$. 又知对于给定的正常数 ϵ , 必存在某一矩阵范数 $\|\cdot\|$ 使得 $\|\mathbf{A}\| < \rho(\mathbf{A}) + \epsilon$, 即 $\|\mathbf{A}\| < r$.

由不等式

$$\|c_m \mathbf{A}^m\| \leq |c_m| \|\mathbf{A}\|^m < |c_m| r^m$$

和当 $|z| < r$ 时数值幂级数 $\sum_{m=0}^{\infty} c_m z^m$ 收敛知, $\sum_{m=0}^{\infty} \|c_m \mathbf{A}^m\|$ 收敛, 即 $\sum_{m=0}^{\infty} c_m \mathbf{A}^m$ 绝对收敛.

当 $\rho(\mathbf{A}) > r$ 时, 设 $|\lambda_0| = \rho(\mathbf{A})$, \mathbf{x}_0 为属于 λ_0 的单位特征向量, 则有 $\mathbf{A}\mathbf{x}_0 = \lambda_0 \mathbf{x}_0$. 假设矩阵级数 $\sum_{m=0}^{\infty} c_m \mathbf{A}^m$ 收敛, 则级数 $\mathbf{x}_0^H (\sum_{m=0}^{\infty} c_m \mathbf{A}^m) \mathbf{x}_0$ 收敛. 此时,

$$\begin{aligned} \mathbf{x}_0^H (\sum_{m=0}^{\infty} c_m \mathbf{A}^m) \mathbf{x}_0 &= \sum_{m=0}^{\infty} c_m \mathbf{x}_0^H \mathbf{A}^m \mathbf{x}_0 \\ &= \sum_{m=0}^{\infty} c_m \mathbf{x}_0^H \lambda_0^m \mathbf{x}_0 = \sum_{m=0}^{\infty} c_m \lambda_0^m \end{aligned}$$

【注意】 数值级数 $\sum_{m=0}^{\infty} c_m \lambda_0^m$ 发散, 故级数 $\mathbf{x}_0^H (\sum_{m=0}^{\infty} c_m \mathbf{A}^m) \mathbf{x}_0$ 发散, 这与假设矛盾.

因此,矩阵幂级数 $\sum_{m=0}^{\infty} c_m \mathbf{A}^m$ 发散.】

注 1: 当 $\rho(\mathbf{A})=r$ 时,矩阵幂级数 $\sum_{m=0}^{\infty} c_m \mathbf{A}^m$ 的敛散性无法由 Abel 型定理判断,只能根据矩阵级数的定义进行判断.

例 4.6.7 设 $\mathbf{A} = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$, 判断 $\sum_{k=1}^{\infty} \frac{1}{k} \mathbf{A}^k$ 的敛散性.

解: 幂级数 $\sum_{k=1}^{\infty} \frac{1}{k} z^k$ 的收敛半径为 1, 矩阵 \mathbf{A} 的谱半径也为 1. 因此无法利用 Abel 型定理判断该幂级数的敛散性. 现利用矩阵幂级数敛散性定义进行判定.

由于矩阵 \mathbf{A} 是 Jordan 标准型, 易求得

$$\mathbf{A}^k = \begin{bmatrix} (-1)^k & (-1)^{k+1} k \\ 0 & (-1)^k \end{bmatrix}$$

此时, 幂级数 $\sum_{k=1}^{\infty} \frac{1}{k} \mathbf{A}^k$ 写为如下表达式:

$$\sum_{k=1}^{\infty} \frac{1}{k} \mathbf{A}^k = \begin{bmatrix} \sum_{k=1}^{\infty} \frac{1}{k} (-1)^k & \sum_{k=1}^{\infty} (-1)^{k+1} \\ 0 & \sum_{k=1}^{\infty} \frac{1}{k} (-1)^k \end{bmatrix}$$

式中: $\sum_{k=1}^{\infty} \frac{1}{k} (-1)^k$ 收敛, 而 $\sum_{k=1}^{\infty} (-1)^{k+1}$ 发散, 故 $\sum_{k=1}^{\infty} \frac{1}{k} \mathbf{A}^k$ 发散.

注 2: 实际上, 矩阵级数 $\sum_{k=1}^{\infty} \frac{1}{k} \mathbf{A}^k$ 不是标准的矩阵幂级数 (矩阵幂级数序列编号从 0 开始), 但有限项的缺失或增加并不影响矩阵级数的敛散性. 因此, 我们仍可以借助矩阵幂级数的结论来判断非标准形的矩阵幂级数.

推论 4.6.4 若幂级数 $\sum_{m=0}^{\infty} c_m z^m$ 在整个复平面上都收敛, 则对任意复方阵 \mathbf{A} , 有矩阵幂级数 $\sum_{m=0}^{\infty} c_m \mathbf{A}^m$ 收敛.

例 4.6.8 已知 $e^z = \sum_{k=0}^{\infty} \frac{1}{k!} z^k$ 对任意复数 z 都收敛, 则对任意复方阵 \mathbf{A} , 矩阵幂级数 $\sum_{k=0}^{\infty} \frac{1}{k!} \mathbf{A}^k$ 收敛.

推论 4.6.5 (Neumann 级数) 矩阵幂级数 $\sum_{m=0}^{\infty} \mathbf{A}^m$ 收敛当且仅当 $\rho(\mathbf{A}) < 1$, 此时

$$\sum_{m=0}^{\infty} \mathbf{A}^m = (\mathbf{I} - \mathbf{A})^{-1}.$$

证明: 由于幂级数 $\sum_{m=0}^{\infty} z^m$ 的收敛半径为 1, 故当 $\rho(\mathbf{A}) < 1$ 时, $\sum_{m=0}^{\infty} \mathbf{A}^m$ 绝对收敛.

反之, 若幂级数 $\sum_{m=0}^{\infty} \mathbf{A}^m$ 收敛, 必有当 $k \rightarrow \infty$ 时, 矩阵序列 $\mathbf{A}^k \rightarrow \mathbf{O}$. 又知矩阵序列 $\mathbf{A}^k \rightarrow \mathbf{O}$

成立的充分必要条件为 $\rho(\mathbf{A}) < 1$. 故当矩阵幂级数 $\sum_{m=0}^{\infty} \mathbf{A}^m$ 收敛时, 必有 $\rho(\mathbf{A}) < 1$.

令 $\mathbf{S}_N = \sum_{k=0}^{N-1} \mathbf{A}^k$, 则 $\mathbf{S}_N(\mathbf{I} - \mathbf{A}) = \mathbf{I} - \mathbf{A}^N$. 又知当 $\rho(\mathbf{A}) < 1$ 时, 矩阵 $\mathbf{I} - \mathbf{A}$ 可逆, 则有

$$\mathbf{S}_N = (\mathbf{I} - \mathbf{A})^{-1} - \mathbf{A}^N (\mathbf{I} - \mathbf{A})^{-1}$$

注意到当 $N \rightarrow \infty$ 时, $\lim_{N \rightarrow \infty} \mathbf{A}^N = \mathbf{O}$. 因此, $\lim_{N \rightarrow \infty} \mathbf{S}_N = (\mathbf{I} - \mathbf{A})^{-1}$. 证毕.

注 3: 可逆矩阵 $(\mathbf{I} - \mathbf{A})^{-1}$ 与矩阵 \mathbf{A} 的乘积可交换, 即 $(\mathbf{I} - \mathbf{A})^{-1} \mathbf{A} = \mathbf{A} (\mathbf{I} - \mathbf{A})^{-1}$. 这是基于幂级数的定义推导的. 实际上, 所有可展开成幂级数的矩阵都具有这一性质. 例如, 当 $|z| < 1$ 时, 有 $\sum_{k=0}^{\infty} k z^k = \frac{z}{(1-z)^2}$. 相应地, 令 $\rho(\mathbf{A}) < 1$, 则有矩阵幂级数

$$\sum_{k=0}^{\infty} k \mathbf{A}^k = \mathbf{A} (\mathbf{I} - \mathbf{A})^{-2} = (\mathbf{I} - \mathbf{A})^{-2} \mathbf{A} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{A} (\mathbf{I} - \mathbf{A})^{-1}$$

例 4.6.9 判断幂级数 $\sum_{m=1}^{\infty} \mathbf{A}^m$ 的敛散性, 其中

$$\mathbf{A} = \begin{bmatrix} 0.2 & 0.5 & 0.2 \\ 0.1 & 0.5 & 0.3 \\ 0.1 & 0.4 & 0.2 \end{bmatrix}$$

解: 矩阵 \mathbf{A} 的列和范数为 $\|\mathbf{A}\|_1 = 0.9$, 故 $\rho(\mathbf{A}) < 1$. 此时, $\sum_{m=0}^{\infty} \mathbf{A}^m$ 收敛.

4.7 矩阵函数

实际上, 收敛的矩阵幂级数定义了一种特殊的映射, 我们将其称为矩阵函数. 它在控制理论、力学、信号处理等学科具有重要的应用.

定义 4.7.1 (矩阵函数) 设幂级数 $\sum_{m=0}^{\infty} c_m z^m$ 的收敛半径为 r , $z \in \mathbb{C}$. 当 $|z| < r$ 时, 幂级数收敛于函数 $f(z)$, 即

$$f(z) = \sum_{m=0}^{\infty} c_m z^m, \quad |z| < r$$

若复方阵 \mathbf{A} 满足 $\rho(\mathbf{A}) < r$, 称收敛的矩阵幂级数 $\sum_{m=0}^{\infty} c_m \mathbf{A}^m$ 为矩阵函数, 记为 $f(\mathbf{A})$.

常见的矩阵函数有

$$\begin{aligned} e^{\mathbf{A}} &= \sum_{m=0}^{\infty} \frac{1}{m!} \mathbf{A}^m, \quad \forall \mathbf{A} \in \mathbb{C}^{n \times n} \\ \sin \mathbf{A} &= \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m+1)!} \mathbf{A}^{2m+1}, \quad \forall \mathbf{A} \in \mathbb{C}^{n \times n} \\ \cos \mathbf{A} &= \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m)!} \mathbf{A}^{2m}, \quad \forall \mathbf{A} \in \mathbb{C}^{n \times n} \\ (\mathbf{I} - \mathbf{A})^{-1} &= \sum_{m=0}^{\infty} \mathbf{A}^m, \quad \forall \rho(\mathbf{A}) < 1 \\ \ln(\mathbf{I} + \mathbf{A}) &= \sum_{m=0}^{\infty} \frac{(-1)^m}{m+1} \mathbf{A}^{m+1}, \quad \forall \rho(\mathbf{A}) < 1 \end{aligned}$$

我们通常称 $e^{\mathbf{A}}$ 为矩阵指数函数, $\sin \mathbf{A}$ 为矩阵正弦函数, $\cos \mathbf{A}$ 为矩阵余弦函数.

命题 4.7.1 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 则以下结论成立:

(1) $\cos(-\mathbf{A}) = \cos \mathbf{A}, \sin(-\mathbf{A}) = -\sin \mathbf{A};$

(2) $e^{i\mathbf{A}} = \cos \mathbf{A} + i(\sin \mathbf{A});$

(3) $\cos \mathbf{A} = \frac{1}{2}(e^{i\mathbf{A}} + e^{-i\mathbf{A}});$

(4) $\sin \mathbf{A} = \frac{1}{2i}(e^{i\mathbf{A}} - e^{-i\mathbf{A}}).$

证明: 前两条性质可由矩阵函数定义直接证明, 具体如下:

$$\begin{aligned} \cos(-\mathbf{A}) &= \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m)!} (-\mathbf{A})^{2m} \\ &= \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m)!} \mathbf{A}^{2m} = \cos \mathbf{A} \\ \sin(-\mathbf{A}) &= \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m+1)!} (-\mathbf{A})^{2m+1} = -\sin \mathbf{A} \\ e^{i\mathbf{A}} &= \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m)!} \mathbf{A}^{2m} + i \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m+1)!} \mathbf{A}^{2m+1} \\ &= \cos \mathbf{A} + i(\sin \mathbf{A}) \end{aligned}$$

根据性质(1)、(2), 有

$$e^{i\mathbf{A}} = \cos \mathbf{A} + i(\sin \mathbf{A}), \quad e^{-i\mathbf{A}} = \cos \mathbf{A} - i(\sin \mathbf{A})$$

求解上式可得性质(3)、(4).

注 1: 尽管指数函数满足 $e^a e^b = e^{a+b} = e^b e^a$, 但该性质对矩阵指数函数一般不成立.

例 4.7.1 已知 $\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$ 和 $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, 分别计算 $e^{\mathbf{A}}, e^{\mathbf{B}}, e^{\mathbf{A}+\mathbf{B}}, e^{\mathbf{A}} e^{\mathbf{B}}$ 和 $e^{\mathbf{B}} e^{\mathbf{A}}$, 并比较 $e^{\mathbf{A}+\mathbf{B}}, e^{\mathbf{A}} e^{\mathbf{B}}$ 和 $e^{\mathbf{B}} e^{\mathbf{A}}$.

解: 矩阵 A 和 B 的最小多项式均为 $m_A(\lambda) = m_B(\lambda) = \lambda(\lambda - 1)$, 故

$$A(A - I) = O, \quad B(B - I) = O$$

即 $A^k = A, B^k = B, k = 1, 2, \dots$.

因此, 根据矩阵指数函数的定义得

$$\begin{aligned} e^A &= \sum_{m=0}^{\infty} \frac{1}{m!} A^m = I + \sum_{m=1}^{\infty} \frac{1}{m!} A \\ &= I + (e - 1)A = \begin{bmatrix} e & 1 - e \\ 0 & 1 \end{bmatrix} \end{aligned}$$

同理可得

$$e^B = I + (e - 1)B = \begin{bmatrix} e & e - 1 \\ 0 & 1 \end{bmatrix}$$

$$\text{由此, } e^A e^B = \begin{bmatrix} e^2 & (e-1)^2 \\ 0 & 1 \end{bmatrix}, e^B e^A = \begin{bmatrix} e^2 & -(e-1)^2 \\ 0 & 1 \end{bmatrix}.$$

又知 $A + B = \text{diag}(2, 0)$, 易求得 $\forall m \geq 1$,

$$(A + B)^m = \begin{bmatrix} 2^m & 0 \\ 0 & 0 \end{bmatrix}$$

故 $e^{A+B} = \text{diag}(e^2, 1)$. 显然, $e^{A+B}, e^A e^B$ 和 $e^B e^A$ 互不相等.

注 2: 矩阵函数的定义式提供了一种计算矩阵函数的方法. 在采用定义法计算矩阵函数时应巧妙地应用矩阵的最小多项式进行简化计算.

定理 4.7.1 设 $A, B \in \mathbb{C}^{n \times n}$, 若 $AB = BA$, 则 $e^A e^B = e^B e^A = e^{A+B}$.

证明: 因 $AB = BA$, 所以二项式定理成立, 故有

$$(A + B)^m = \sum_{k=0}^m C_m^k A^{m-k} B^k$$

所以

$$\begin{aligned} e^{A+B} &= \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{k=0}^m C_m^k A^{m-k} B^k \\ &= \sum_{m=0}^{\infty} \sum_{k=0}^m \frac{1}{k! (m-k)!} A^{m-k} B^k \\ &= \left(\sum_{m=0}^{\infty} \frac{1}{m!} A^m \right) \left(\sum_{k=0}^{\infty} \frac{1}{k!} B^k \right) = e^A e^B \end{aligned}$$

由定理 4.7.1 可得以下推论.

推论 4.7.1 设 $A \in \mathbb{C}^{n \times n}$, 则 $e^A e^{-A} = e^{-A} e^A = I$, 即 $(e^A)^{-1} = e^{-A}$.

推论 4.7.2 设 $A \in \mathbb{C}^{n \times n}$, 则 $\sin^2 A + \cos^2 A = I$.

注 3: 推论 4.7.1 表明无论矩阵 A 是否可逆, 矩阵指数函数 e^A 必可逆, 且其逆矩阵为 e^{-A} .

若矩阵函数 $f(A)$ 的自变量由矩阵 A 换成 At , 其中 t 为标量参数, 则有矩阵函数

表达式为

$$f(\mathbf{A}t) = \sum_{m=0}^{\infty} c_m (\mathbf{A}t)^m = \sum_{m=0}^{\infty} c_m t^m \mathbf{A}^m, \quad |t| \rho(\mathbf{A}) < r$$

我们称之为含参矩阵函数. 这类矩阵函数在线性常微分方程求解等应用中常会遇到.

例 4.7.2 设 $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, 求 $e^{\mathbf{A}t}$.

解: 由 $|\lambda \mathbf{I} - \mathbf{A}| = \lambda^2 + 1$ 知, $\mathbf{A}^2 + \mathbf{I} = \mathbf{O}$. 由此, 可得

$$\mathbf{A}^2 = -\mathbf{I}, \quad \mathbf{A}^3 = -\mathbf{A}, \quad \mathbf{A}^4 = \mathbf{I}, \quad \mathbf{A}^5 = \mathbf{A}, \dots$$

进而写出递归关系式 $\mathbf{A}^{2k} = (-1)^k \mathbf{I}, \mathbf{A}^{2k+1} = (-1)^k \mathbf{A}, k=1, 2, \dots$.

此时,

$$\begin{aligned} e^{\mathbf{A}t} &= \sum_{k=0}^{\infty} \frac{1}{k!} (\mathbf{A}t)^k \\ &= \left(1 - \frac{t^2}{2!} + \frac{t^4}{4!} - \dots\right) \mathbf{I} + \left(t - \frac{t^3}{3!} + \frac{t^5}{5!} - \dots\right) \mathbf{A} \\ &= (\cos t) \mathbf{I} + (\sin t) \mathbf{A} = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix} \end{aligned}$$

从上例可以看出, 尽管矩阵函数 $f(\mathbf{A})$ 是关于矩阵 \mathbf{A} 的幂级数, 但由于矩阵 \mathbf{A} 最小多项式 $m_{\mathbf{A}}(\lambda)$ 的存在, 矩阵函数 $f(\mathbf{A})$ 总可以表示成 $\mathbf{I}, \mathbf{A}, \mathbf{A}^2, \dots, \mathbf{A}^{l-1}$ 这 l 个矩阵的线性组合, 其中 $l = \deg(m_{\mathbf{A}}(\lambda))$. 比如在例 4.7.2 中, $e^{\mathbf{A}t} = (\cos t) \mathbf{I} + (\sin t) \mathbf{A}$. 当然, 若矩阵 \mathbf{A} 的最小多项式 $m_{\mathbf{A}}(\lambda)$ 未知, 则同样可利用矩阵 \mathbf{A} 特征多项式 $f(\lambda)$ 甚至它的任一零化多项式进行化简.

例 4.7.3 设 $\mathbf{A} \in \mathbb{C}^{4 \times 4}$, 其特征值分别为 $\pi, -\pi$ 和 0 (代数重数为 2), 求 $\sin \mathbf{A}$ 和 $\cos \mathbf{A}$.

解: 由矩阵 \mathbf{A} 的特征多项式 $|\lambda \mathbf{I} - \mathbf{A}| = \lambda^2 (\lambda - \pi) (\lambda + \pi) = \lambda^4 - \pi^2 \lambda^2$ 知, $\mathbf{A}^4 = \pi^2 \mathbf{A}^2$. 所以对 $m=1, 2, \dots$, 有 $\mathbf{A}^{2m+1} = \pi^{2(m-1)} \mathbf{A}^3$, 并将该式代入 $\sin \mathbf{A}$ 的定义式, 得

$$\begin{aligned} \sin \mathbf{A} &= \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m+1)!} \mathbf{A}^{2m+1} \\ &= \mathbf{A} + \sum_{m=1}^{\infty} \frac{(-1)^m \pi^{2(m-1)}}{(2m+1)!} \mathbf{A}^3 \\ &= \mathbf{A} - \frac{1}{\pi^2} \mathbf{A}^3 + \frac{1}{\pi^2} \sum_{m=0}^{\infty} \frac{(-1)^m \pi^{2(m+1)}}{(2m+1)!} \mathbf{A}^3 \\ &= \mathbf{A} + \frac{\sin \pi - \pi}{\pi^3} \mathbf{A}^3 = \mathbf{A} - \pi^{-2} \mathbf{A}^3 \end{aligned}$$

同理可求得 $\cos \mathbf{A} = \mathbf{I} - \frac{2}{\pi^2} \mathbf{A}^2$.

尽管定义法结合矩阵的零化多项式可简化矩阵函数的计算, 但该方法还是比较

烦琐. 计算矩阵函数有两种常见方法: 方法一是利用矩阵的相似变换, 方法二是利用谱上一致性. 这里首先介绍方法一.

定理 4.7.2 设复方阵 A 与 B 相似, 即存在可逆矩阵 P 使得 $P^{-1}AP = B$. 若 $f(A)$ 是矩阵函数, 则 $f(A) = Pf(B)P^{-1}$.

证明: 直接利用矩阵函数 $f(A)$ 定义进行证明. 根据矩阵函数 $f(A)$ 定义知

$$\begin{aligned} f(A) &= \sum_{m=0}^{\infty} c_m A^m = \sum_{m=0}^{\infty} c_m (PBP^{-1})^m \\ &= P \left(\sum_{m=0}^{\infty} c_m B^m \right) P^{-1} = Pf(B)P^{-1} \end{aligned}$$

基于定理 4.7.2, 我们对矩阵 A 分两种情况讨论: (1) A 为单纯矩阵; (2) A 不是单纯矩阵.

(1) 若 A 是单纯矩阵, 则存在可逆矩阵 P 使得 $P^{-1}AP = \text{diag}(\lambda_1, \dots, \lambda_n)$, 进而

$$f(A) = P \text{diag}(f(\lambda_1), \dots, f(\lambda_n)) P^{-1} \quad (4.7.1)$$

$$f(At) = P \text{diag}(f(\lambda_1 t), \dots, f(\lambda_n t)) P^{-1} \quad (4.7.2)$$

此时, 矩阵函数 $f(A)$ 和含参矩阵函数 $f(At)$ 仍是单纯矩阵. 特别地,

$$\begin{aligned} e^A &= P \text{diag}(e^{\lambda_1}, \dots, e^{\lambda_n}) P^{-1} \\ \sin A &= P \text{diag}(\sin \lambda_1, \dots, \sin \lambda_n) P^{-1} \end{aligned}$$

例 4.7.4 设矩阵 $A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$, 求矩阵函数 $\sin A$.

解: 由于 $\lambda_1 = 1, \lambda_2 = 2$, 故矩阵 A 是单纯矩阵. 分别计算变换矩阵及其逆矩阵为

$$P = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \quad P^{-1} = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$$

因此,

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

注 4: 利用单纯矩阵的谱分解式也可以方便地计算矩阵函数. 设单纯矩阵 A 有如下谱分解式:

$$A = \sum_{i=1}^k \lambda_i E_i \quad (4.7.3)$$

式中: $\lambda_1, \dots, \lambda_k$ 是矩阵 A 的 k 个互异特征值, E_1, \dots, E_k 是对应的谱阵.

将谱分解式(4.7.3)代入矩阵函数 $f(A)$ 的定义式, 得

$$f(A) = \sum_{m=0}^{\infty} c_m A^m = \sum_{i=1}^k f(\lambda_i) E_i \quad (4.7.4)$$

相比于式(4.7.1), 式(4.7.4)不需要求解变换矩阵, 也不需要三个矩阵相乘. 利

用谱分解再次考察例 4.7.4, 由推论 3.7.2 可计算出矩阵 \mathbf{A} 的谱阵, 分别为

$$\begin{aligned} \mathbf{E}_1 &= \frac{1}{1-2}(\mathbf{A} - 2\mathbf{I}) = \begin{bmatrix} 1 & -2 \\ 0 & 0 \end{bmatrix} \\ \mathbf{E}_2 &= \frac{1}{2-1}(\mathbf{A} - \mathbf{I}) = \begin{bmatrix} 0 & 2 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

因此,

$$\begin{aligned} \sin \mathbf{A} &= (\sin 1)\mathbf{E}_1 + (\sin 2)\mathbf{E}_2 \\ &= \begin{bmatrix} \sin 1 & 2\sin 2 - 2\sin 1 \\ 0 & \sin 2 \end{bmatrix} \end{aligned}$$

(2) 若矩阵 \mathbf{A} 为非单纯矩阵, 则存在可逆矩阵 \mathbf{P} 使得 $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{J}$, 其中 \mathbf{J} 为矩阵 \mathbf{A} 的 Jordan 标准形, 可表示为 $\mathbf{J} = \text{diag}(\mathbf{J}_1, \cdots, \mathbf{J}_s)$, $\mathbf{J}_i (i=1, \cdots, s)$ 为 Jordan 块, 可表示为

$$\mathbf{J}_i = \begin{bmatrix} \lambda_i & 1 & & & \mathbf{O} \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ \mathbf{O} & & & \lambda_i & 1 \\ & & & & \lambda_i \end{bmatrix}_{n_i \times n_i} \quad (4.7.5)$$

此时, $\mathbf{A}^k = \mathbf{P}\mathbf{J}^k\mathbf{P}^{-1}$, 式中, $\mathbf{J}^k = \text{diag}(\mathbf{J}_1^k, \cdots, \mathbf{J}_s^k)$, $\mathbf{J}_i^k (i=1, \cdots, s)$ 满足

$$\mathbf{J}_i^k = \begin{bmatrix} \lambda_i^k & C_k^1 \lambda_i^{k-1} & C_k^2 \lambda_i^{k-2} & \cdots & C_k^{n_i-1} \lambda_i^{k-n_i+1} \\ & \lambda_i^k & C_k^1 \lambda_i^{k-1} & \cdots & C_k^{n_i-2} \lambda_i^{k-n_i+2} \\ & & \ddots & \ddots & \vdots \\ \mathbf{O} & & & \lambda_i^k & C_k^1 \lambda_i^{k-1} \\ & & & & \lambda_i^k \end{bmatrix}$$

将上式代入矩阵函数 $f(\mathbf{A}) = \sum_{m=0}^{\infty} c_m \mathbf{A}^m$, 得

$$\begin{aligned} f(\mathbf{A}) &= \mathbf{P} \text{diag}(f(\mathbf{J}_1), \cdots, f(\mathbf{J}_s)) \mathbf{P}^{-1} \\ f(\mathbf{J}_i) &= \begin{bmatrix} \sum_{m=0}^{\infty} c_m \lambda_i^m & \sum_{m=0}^{\infty} c_m C_k^1 \lambda_i^{m-1} & \cdots & \sum_{m=0}^{\infty} c_m C_k^{n_i-1} \lambda_i^{m-n_i+1} \\ & \sum_{m=0}^{\infty} c_m \lambda_i^m & \ddots & \vdots \\ & & \ddots & \sum_{m=0}^{\infty} c_m \lambda_i^{m-1} \\ \mathbf{O} & & & \sum_{m=0}^{\infty} c_m \lambda_i^m \end{bmatrix} \end{aligned}$$

注意到当 $j < m$ 时, $\sum_{m=0}^{\infty} c_m C_m^j \lambda_i^{m-j} = 0$; 否则

$$\sum_{m=0}^{\infty} c_m C_m^j \lambda_i^{m-j} = \frac{1}{j!} f^{(j)}(\lambda_i)$$

因此,

$$f(\mathbf{A}) = \mathbf{P} \begin{bmatrix} f(J_1) & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & f(J_s) \end{bmatrix} \mathbf{P}^{-1} \quad (4.7.6)$$

$$f(\mathbf{J}_i) = \begin{bmatrix} f(\lambda_i) & f'(\lambda_i) & \cdots & \frac{1}{(n_i-1)!} f^{(n_i-1)}(\lambda_i) \\ & f(\lambda_i) & \ddots & \vdots \\ & & \ddots & f'(\lambda_i) \\ \mathbf{O} & & & f(\lambda_i) \end{bmatrix}_{n_i \times n_i} \quad (4.7.7)$$

式(4.7.6)和式(4.7.7)称为 *Sylvester* 公式.

例 4.7.5 设矩阵 $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -5 & 4 \end{bmatrix}$, 求矩阵函数 $e^{\mathbf{A}}$.

解: 由 $|\lambda \mathbf{I} - \mathbf{A}| = (\lambda - 1)^2(\lambda - 2)$ 知, 矩阵 \mathbf{A} 的特征值为 $\lambda_1 = \lambda_2 = 1, \lambda_3 = 2$. 又知 $\text{rank}(\mathbf{I} - \mathbf{A}) = 2$, 故 λ_1 的几何重数为 1. 因此, 矩阵 \mathbf{A} 不是单纯矩阵.

依次求得 \mathbf{A} 的 Jordan 标准形 \mathbf{J} 及变换矩阵 \mathbf{P} 为

$$\mathbf{J} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 1 & 1 \\ 4 & 1 & 2 \end{bmatrix}, \quad \mathbf{P}^{-1} = \begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -1 \\ -2 & 3 & -1 \end{bmatrix}$$

由 Sylvester 公式, 得

$$f(\mathbf{J}_1) = e^2, \quad f(\mathbf{J}_2) = \begin{bmatrix} e & e \\ 0 & e \end{bmatrix}$$

因此,

$$f(\mathbf{A}) = \mathbf{P} \begin{bmatrix} f(\mathbf{J}_1) & \mathbf{O} \\ \mathbf{O} & f(\mathbf{J}_2) \end{bmatrix} \mathbf{P}^{-1} = \begin{bmatrix} e^2 - 2e & -2e^2 + 5e & e^2 - 2e \\ 2e^2 - 4e & -4e^2 + 8e & 2e^2 - 3e \\ 4e^2 - 6e & -8e^2 + 11e & 4e^2 - 4e \end{bmatrix}$$

推论 4.7.3 设矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 的特征值为 $\lambda_1, \dots, \lambda_n$, $f(z) = \sum_{m=0}^{\infty} c_m z^m$ 的收敛半径为 r . 当 $\rho(\mathbf{A}) < r$ 时, 矩阵函数 $f(\mathbf{A})$ 的特征值为 $f(\lambda_1), \dots, f(\lambda_n)$.

推论 4.7.4 设矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 有分解式 $\mathbf{A} = \mathbf{PJP}^{-1}$, 其中 $\mathbf{J} = \text{diag}(\mathbf{J}_1, \dots, \mathbf{J}_s)$, $\mathbf{J}_i (i=1, \dots, s)$ 如式(4.7.5)所示, 幂级数 $f(z) = \sum_{m=0}^{\infty} c_m z^m$ 的收敛半径为 r . 当

$t\rho(\mathbf{A}) < r$ 时, 含参矩阵函数 $f(\mathbf{A}t) = \mathbf{P} \text{diag}(f(\mathbf{J}_1(\lambda_1 t)), \dots, f(\mathbf{J}_s(\lambda_s t))) \mathbf{P}^{-1}$, 其中

$$f(\mathbf{J}_i(\lambda_i t)) = \begin{bmatrix} f(\lambda_i t) & tf'(\lambda_i t) & \cdots & \frac{t^{n_i-1}}{(n_i-1)!} f^{(n_i-1)}(\lambda_i t) \\ & f(\lambda_i t) & \ddots & \vdots \\ & & \ddots & tf'(\lambda_i t) \\ \mathbf{O} & & & f(\lambda_i t) \end{bmatrix}_{n_i \times n_i} \quad (4.7.8)$$

例 4.7.6 设矩阵 $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$, 求矩阵函数 $\cos(\mathbf{A}t)$.

解: 已知 $f(\lambda_i t) = \cos 2t$, $tf'(\lambda_i t) = t \frac{d}{dx} \cos x|_{x=2t} = -t \sin 2t$, 故由推论 4.7.4 知

$$\cos(\mathbf{A}t) = \begin{bmatrix} \cos 2t & -t \sin 2t \\ 0 & \cos 2t \end{bmatrix}$$

尽管 Sylvester 公式非常漂亮, 但利用 Sylvester 公式计算矩阵函数有一定的计算难度, 主要表现在: (1) 相似变换矩阵 \mathbf{P} 及其逆 \mathbf{P}^{-1} 难求; (2) 3 个矩阵相乘 $\mathbf{P}f(\mathbf{J})\mathbf{P}^{-1}$ 计算烦琐. 为解决这一问题, 我们换个角度考虑矩阵函数的计算问题. 我们已知晓任一矩阵函数(须有定义)总可以表示为不超过 $(l-1)$ 次矩阵多项式. 该矩阵多项式由 l 个系数唯一确定, 若能确定这 l 个系数, 我们自然就可以确定出矩阵函数. 这就是矩阵函数的第二种计算方法.

定理 4.7.3 设矩阵 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 的最小多项式次数为 l , 幂级数 $f(z) = \sum_{m=0}^{\infty} c_m z^m$ 的收敛半径为 r . 若 $\rho(\mathbf{A}) < r$, 定义矩阵函数 $f(\mathbf{A})$, 则必存在唯一的 $(l-1)$ 次矩阵多项式 $p(\mathbf{A}) = \beta_0 \mathbf{I} + \beta_1 \mathbf{A} + \cdots + \beta_{l-1} \mathbf{A}^{l-1}$ 使得 $f(\mathbf{A}) = p(\mathbf{A})$.

证明: 对于多项式 $f(\lambda) = \lambda^m$ 和 $m_{\mathbf{A}}(\lambda)$, 必存在 $q_m(\lambda)$ 以及 $p_m(\lambda)$ 使得

$$\lambda^m = m_{\mathbf{A}}(\lambda) q_m(\lambda) + p_m(\lambda)$$

式中, $\deg(p_m(\lambda)) < l$.

因此,

$$\mathbf{A}^m = m_{\mathbf{A}}(\mathbf{A}) q_m(\mathbf{A}) + p_m(\mathbf{A}) = p_m(\mathbf{A})$$

将上式代入矩阵函数 $f(\mathbf{A})$ 的定义式, 得

$$f(\mathbf{A}) = \sum_{m=0}^{\infty} c_m \mathbf{A}^m = \sum_{m=0}^{\infty} c_m p_m(\mathbf{A}) \quad (4.7.9)$$

由于矩阵函数 $f(\mathbf{A})$ 绝对收敛, 所以对式(4.7.9)右端级数改变求和顺序仍会绝对收敛. 因此, 对式(4.7.9)右端矩阵多项式按 $\mathbf{I}, \mathbf{A}, \dots, \mathbf{A}^{l-1}$ 分别求和, 整理得

$$f(\mathbf{A}) = \beta_0 \mathbf{I} + \beta_1 \mathbf{A} + \cdots + \beta_{l-1} \mathbf{A}^{l-1} \quad (4.7.10)$$

式中,系数 $\beta_0, \beta_1, \dots, \beta_{l-1}$ 必存在. 这表明必存在 $(l-1)$ 次矩阵多项式 $p(\mathbf{A}) = \beta_0 \mathbf{I} + \beta_1 \mathbf{A} + \dots + \beta_{l-1} \mathbf{A}^{l-1}$ 使得 $f(\mathbf{A}) = p(\mathbf{A})$.

现证明矩阵多项式 $p(\mathbf{A})$ 的唯一性. 假设存在两个 $(l-1)$ 次矩阵多项式 $p(\mathbf{A})$ 和 $q(\mathbf{A})$ 满足 $f(\mathbf{A}) = p(\mathbf{A})$ 和 $f(\mathbf{A}) = q(\mathbf{A})$, 则 $p(\mathbf{A}) = q(\mathbf{A})$.

定义 $g(\lambda) = p(\lambda) - q(\lambda)$, 则 $g(\lambda)$ 是矩阵 \mathbf{A} 的零化多项式, 其次数为 $(l-1)$, 这显然与 \mathbf{A} 的最小多项式次数为 l 相矛盾. 因此假设不成立, 即 $p(\mathbf{A})$ 唯一.

定理 4.7.3 从理论上保证了 $(l-1)$ 次矩阵多项式 $p(\mathbf{A})$ 的存在性和唯一性. 为确定矩阵多项式 $p(\mathbf{A})$ 的系数, 须至少列出 l 个方程. 为此, 引入谱上给定和谱上一致概念.

定义 4.7.2 (谱上给定) 设 $\lambda_1, \dots, \lambda_s$ 是 n 阶复方阵 \mathbf{A} 的 s 个互异特征值,

$$m_{\mathbf{A}}(\lambda) = (\lambda - \lambda_1)^{n_1} (\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_s)^{n_s}$$

是 \mathbf{A} 的最小多项式, $\deg(m_{\mathbf{A}}(\lambda)) = l$. 若复函数 $f(z)$ 及其各阶导数 $f^{(j)}(z)$ 在 $z = \lambda_i$ 处的 n_i 个值 $f^{(j)}(\lambda_i)$ 均有界, $j = 0, 1, \dots, n_i - 1$, 则称 $f(z)$ 在矩阵 \mathbf{A} 的谱上给定 (或谱上有定义), 并称 $\lambda_1, \dots, \lambda_s$ 为谱点, $f^{(j)}(\lambda_i)$ 为 $f(z)$ 在矩阵 \mathbf{A} 上的谱值.

例 4.7.7 考察函数 $f(z)$ 在矩阵 \mathbf{A} 的谱上是否有定义, 其中

$$f(z) = \frac{1}{(z-3)(z-4)}, \quad \mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

解: 矩阵 \mathbf{A} 的最小多项式为 $m_{\mathbf{A}}(\lambda) = (\lambda - 2)(\lambda - 1)^2$, 其谱点分别为 $\lambda_1 = 2$, $\lambda_2 = 1$; 相应的谱值为

$$f(2) = \frac{1}{2}, \quad f(1) = \frac{1}{6}, \quad f'(1) = \frac{5}{36}$$

因此, $f(z)$ 在矩阵 \mathbf{A} 的谱上给定.

若将例 4.7.7 中矩阵 \mathbf{A} 替换为

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

则矩阵 \mathbf{A} 的最小多项式为 $m_{\mathbf{A}}(\lambda) = (\lambda - 1)(\lambda - 3)^2$, 其谱点分别为 $\lambda_1 = 1, \lambda_2 = 3$; 相应谱值为 $f(1) = \frac{1}{6}, f(3)$ 无界. 所以 $f(z)$ 在 \mathbf{A} 的谱上无定义.

定义 4.7.3 (谱上一致) 设复方阵 \mathbf{A} 的最小多项式为

$$m_{\mathbf{A}}(\lambda) = (\lambda - \lambda_1)^{n_1} (\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_s)^{n_s}, \quad \deg(m_{\mathbf{A}}(\lambda)) = l$$

若函数 $f(\lambda)$ 与 $p(\lambda)$ 在 \mathbf{A} 的谱上给定且满足

$$\begin{cases} f(\lambda_i) = p(\lambda_i) \\ f'(\lambda_i) = p'(\lambda_i) \\ \vdots \\ f^{(n_i-1)}(\lambda_i) = p^{(n_i-1)}(\lambda_i) \end{cases}, \quad i=1,2,\dots,s$$

则称函数 $f(\lambda)$ 与 $p(\lambda)$ 在矩阵 \mathbf{A} 的谱上一致。

例 4.7.8 考察函数 $f(z)$ 与 $p(z)$ 在矩阵 \mathbf{A} 的谱上一致条件, 其中

$$f(z) = e^z, \quad p(z) = \beta_0 + \beta_1 z, \quad \mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

解: 矩阵 \mathbf{A} 的最小多项式为 $m_A(\lambda) = (\lambda-1)(\lambda-2)$, 其谱点分别为 $\lambda_1=1, \lambda_2=2$; 相应谱值为 $f(1)=e, f(2)=e^2; p(1)=\beta_0+\beta_1, p(2)=\beta_0+2\beta_1$ 。

根据谱上一致条件, 得

$$\begin{cases} \beta_0 + \beta_1 = e \\ \beta_0 + 2\beta_1 = e^2 \end{cases}$$

解得 $\beta_0 = 2e - e^2, \beta_1 = e^2 - e$ 。此时,

$$f(\mathbf{A}) = \begin{bmatrix} e & 0 \\ 0 & e^2 \end{bmatrix} = p(\mathbf{A})$$

定理 4.7.4 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 幂级数 $f(z) = \sum_{m=0}^{\infty} c_m z^m$ 与多项式 $p(z) = \sum_{i=0}^k \beta_i z^i$ 在矩阵 \mathbf{A} 的谱上给定, 则 $f(\mathbf{A}) = p(\mathbf{A})$ 的充分必要条件是 $f(z)$ 与 $p(z)$ 在矩阵 \mathbf{A} 的谱上一致。

证明: 设矩阵 \mathbf{A} 有分解式 $\mathbf{A} = \mathbf{P}\mathbf{J}\mathbf{P}^{-1}$, 其中, $\mathbf{J} = \text{diag}(\mathbf{J}_1, \dots, \mathbf{J}_s)$, $\mathbf{J}_i (i=1, \dots, s)$ 为 Jordan 块, 如式(4.7.5)所示。于是

$$f(\mathbf{A}) = p(\mathbf{A}) \Leftrightarrow \text{diag}(f(\mathbf{J}_1), \dots, f(\mathbf{J}_s)) = \text{diag}(p(\mathbf{J}_1), \dots, p(\mathbf{J}_s))$$

即对 $i=1, \dots, s$, 有

$$f(\mathbf{J}_i) = p(\mathbf{J}_i) \quad (4.7.11)$$

由 Sylvester 公式知, 式(4.7.11)成立的充分必要条件为

$$\begin{bmatrix} f(\lambda_i) & \cdots & \frac{1}{(n_i-1)!} f^{(n_i-1)}(\lambda_i) \\ & \ddots & \vdots \\ \mathbf{O} & & f(\lambda_i) \end{bmatrix} = \begin{bmatrix} p(\lambda_i) & \cdots & \frac{1}{(n_i-1)!} p^{(n_i-1)}(\lambda_i) \\ & \ddots & \vdots \\ \mathbf{O} & & p(\lambda_i) \end{bmatrix}$$

进而, 对 $i=1, 2, \dots, s$, 有

$$\begin{cases} f(\lambda_i) = p(\lambda_i) \\ f'(\lambda_i) = p'(\lambda_i) \\ \vdots \\ f^{(n_i-1)}(\lambda_i) = p^{(n_i-1)}(\lambda_i) \end{cases}$$

上式表明 $f(z)$ 与 $p(z)$ 在矩阵 \mathbf{A} 是谱上一致的. 证毕.

根据定理 4.7.4 的谱上一致性条件, 可列出 l 个独立方程. 由此可求解出定理 4.7.3 中的 l 个系数 $\beta_0, \beta_1, \dots, \beta_{l-1}$. 这就是计算矩阵函数的第二种方法, 我们称之为谱上一致性方法.

例 4.7.9 设矩阵 $\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$, 求矩阵函数 $e^{\mathbf{A}}$.

解: 由矩阵 \mathbf{A} 的最小多项式 $m_{\mathbf{A}}(\lambda) = (\lambda-3)(\lambda-2)(\lambda-1)$ 定义多项式

$$p(\lambda) = a_2\lambda^2 + a_1\lambda + a_0$$

根据谱上一致性条件, 得

$$\begin{cases} f(1) = p(1) \\ f(2) = p(2) \\ f(3) = p(3) \end{cases} \Rightarrow \begin{cases} a_2 + a_1 + a_0 = e \\ 4a_2 + 2a_1 + a_0 = e^2 \\ 8a_2 + 3a_1 + a_0 = e^3 \end{cases}$$

将其代入 $p(\mathbf{A})$, 得

$$p(\mathbf{A}) = \begin{bmatrix} a_2 + a_1 + a_0 & 0 & 0 \\ 0 & 4a_2 + 2a_1 + a_0 & 0 \\ 0 & 0 & 8a_2 + 3a_1 + a_0 \end{bmatrix} = \begin{bmatrix} e & 0 & 0 \\ 0 & e^2 & 0 \\ 0 & 0 & e^3 \end{bmatrix}$$

因此, $f(\mathbf{A}) = \text{diag}(e, e^2, e^3)$.

例 4.7.10 设矩阵 $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -5 & 4 \end{bmatrix}$, 求矩阵函数 $e^{\mathbf{A}}$.

解: 由矩阵 \mathbf{A} 的最小多项式 $m_{\mathbf{A}}(\lambda) = (\lambda-2)(\lambda-1)^2$ 定义多项式

$$p(\lambda) = a_2\lambda^2 + a_1\lambda + a_0$$

根据谱上一致性条件, 得

$$\begin{cases} f(2) = p(2) \\ f(1) = p(1) \\ f'(1) = p'(1) \end{cases} \Rightarrow \begin{cases} 4a_2 + 2a_1 + a_0 = e^2 \\ a_2 + a_1 + a_0 = e \\ 2a_2 + a_1 = e \end{cases}$$

解得

$$\begin{cases} a_2 = e^2 - 2e \\ a_1 = -2e^2 + 5e \\ a_0 = e^2 - 2e \end{cases}$$

因此,

$$\begin{aligned} p(\mathbf{A}) &= a_2\mathbf{A}^2 + a_1\mathbf{A} + a_0\mathbf{I} \\ &= \begin{bmatrix} a_0 & a_1 & a_2 \\ 2a_2 & -5a_2 + a_0 & 4a_2 + a_1 \\ 8a_2 + 2a_1 & -18a_2 - 5a_1 & 11a_2 + 4a_1 + a_0 \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} e^2 - 2e & -2e^2 + 5e & e^2 - 2e \\ 2e^2 - 4e & -4e^2 + 8e & 2e^2 - 3e \\ 4e^2 - 6e & -8e^2 + 11e & 4e^2 - 4e \end{bmatrix}$$

注 5: 利用谱上一致性方法计算矩阵函数不仅适用于单纯矩阵,也适用于非单纯矩阵. 若需计算含参矩阵函数时,我们仍可以利用谱上一致性方法计算. 相关结论如下.

设 $f(\mathbf{A}t) = \sum_{m=0}^{\infty} c'_m \mathbf{A}^m$, 其中, $c'_m = c_m t^m$. 故含参矩阵函数 $f(\mathbf{A}t)$ 仍可以唯一地由 $l-1$ 次矩阵多项式 $p_t(\mathbf{A})$ 表示, 即

$$f(\mathbf{A}t) = \alpha_0(t)\mathbf{I} + \alpha_1(t)\mathbf{A} + \cdots + \alpha_{l-1}(t)\mathbf{A}^{l-1} = p_t(\mathbf{A})$$

式中, $\alpha_i(t) (i=0, 1, \cdots, l-1)$ 为待定含参数.

利用 Sylvester 公式, 可得如下方程组:

$$p_t^{(j)}(\lambda_i) = t^j f^{(j)}(\lambda_i t), \quad j=0, \cdots, n_i-1, i=1, \cdots, s \quad (4.7.12)$$

通过求解方程组 (4.7.12) 可确定待定的 l 个系数 $\alpha_i(t), i=0, 1, \cdots, l-1$.

例 4.7.11 设矩阵 $\mathbf{A} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$, 求矩阵函数 $e^{\mathbf{A}t}$.

解: 由矩阵 \mathbf{A} 的最小多项式 $m_{\mathbf{A}}(\lambda) = (\lambda-2)^3$ 定义 $f(\lambda t) = e^{\lambda t}$ 和多项式

$$p_t(\lambda) = \alpha_0(t) + \alpha_1(t)\lambda + \alpha_2(t)\lambda^2$$

依据式 (4.7.12) 建立如下方程组:

$$\begin{cases} \alpha_0(t) + 2\alpha_1(t) + 4\alpha_2(t) = e^{2t} \\ \alpha_1(t) + 4\alpha_2(t) = t e^{2t} \\ 2\alpha_2(t) = t^2 e^{2t} \end{cases}$$

解得 $\alpha_2(t) = \frac{t^2}{2} e^{2t}, \alpha_1(t) = e^{2t}(t-2t^2), \alpha_0(t) = e^{2t}(1-2t+2t^2)$. 故

$$\begin{aligned} e^{\mathbf{A}t} &= \alpha_0(t)\mathbf{I}_3 + \alpha_1(t)\mathbf{A} + \alpha_2(t)\mathbf{A}^2 \\ &= \begin{bmatrix} \alpha_0 + 2\alpha_1 + 4\alpha_2 & \alpha_1 + 4\alpha_2 & \alpha_2 \\ 0 & \alpha_0 + 2\alpha_1 + 4\alpha_2 & \alpha_1 + 4\alpha_2 \\ 0 & 0 & \alpha_0 + 2\alpha_1 + 4\alpha_2 \end{bmatrix} \\ &= \begin{bmatrix} e^{2t} & t e^{2t} & \frac{t^2}{2} e^{2t} \\ 0 & e^{2t} & t e^{2t} \\ 0 & 0 & e^{2t} \end{bmatrix} \end{aligned}$$

注 6: 利用谱上一致性方法计算矩阵函数时,可不选用矩阵的最小多项式,而选用矩阵的任一零化多项式. 此时须把选用的零化多项式当作矩阵的最小多项式,然后

仿照上面的方法求解即可.

再次考虑例 4.7.3. 由矩阵特征多项式 $g(\lambda) = \lambda^4 - \pi^2 \lambda^2$ 得, 令 $f(\lambda) = \sin \lambda$ 和 $p(\lambda) = \beta_0 + \beta_1 \lambda + \beta_2 \lambda^2 + \beta_3 \lambda^3$, 则有

$$\begin{cases} \beta_0 + \beta_1 \pi + \beta_2 \pi^2 + \beta_3 \pi^3 = 0 \\ \beta_0 - \beta_1 \pi + \beta_2 \pi^2 - \beta_3 \pi^3 = 0 \\ \beta_0 = \sin 0 \\ \beta_1 = \cos 0 \end{cases}$$

解得 $\beta_0 = 0, \beta_1 = 1, \beta_2 = 0, \beta_3 = -\frac{1}{\pi^2}$. 因此, $\sin \mathbf{A} = \mathbf{A} - \frac{1}{\pi^2} \mathbf{A}^3$.

例 4.7.12 设函数 $f(z) = \frac{1}{z}$, 矩阵 $\mathbf{A} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$, 求 $f(\mathbf{A})$.

分析: 因为 $f(z)$ 无法在 $z=0$ 处展开成幂级数形式, 所以无法根据定义 4.7.1 计算 $f(\mathbf{A})$. 换言之, 在定义 4.7.1 下, $f(\mathbf{A})$ 不能称为矩阵函数.

如果忽略定义 4.7.1 中对矩阵函数的收敛性要求, 分别采用方法一 Sylvester 公式和方法二谱上一致性方法进行计算.

方法一: Sylvester 公式. 由 Sylvester 公式知,

$$\begin{aligned} f(2) &= \frac{1}{2}, \quad f'(2) = -\frac{1}{z^2} \Big|_{z=2} = -\frac{1}{4} \\ \frac{1}{(3-1)!} f^2(2) &= \frac{1}{z^3} \Big|_{z=2} = \frac{1}{8} \end{aligned}$$

因此,

$$f(\mathbf{A}) = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & \frac{1}{8} \\ 0 & \frac{1}{2} & -\frac{1}{4} \\ 0 & 0 & \frac{1}{2} \end{bmatrix}$$

方法二: 谱上一致性方法. 由矩阵 \mathbf{A} 的最小多项式 $m_{\mathbf{A}}(\lambda) = (\lambda-2)^3$ 定义多项式

$$p(z) = \alpha_0 + \alpha_1 z + \alpha_2 z^2$$

则根据谱上一致性条件可得如下方程组:

$$\begin{cases} \alpha_0 + 2\alpha_1 + 4\alpha_2 = \frac{1}{2} \\ \alpha_1 + 4\alpha_2 = -\frac{1}{4} \\ 2\alpha_2 = \frac{1}{4} \end{cases}$$

解得 $\alpha_0 = \frac{3}{2}, \alpha_1 = -\frac{3}{4}$ 和 $\alpha_2 = \frac{1}{8}$. 故

$$f(\mathbf{A}) = \frac{3}{2}\mathbf{I} - \frac{3}{4}\mathbf{A} + \frac{1}{8}\mathbf{A}^2 = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & \frac{1}{8} \\ 0 & \frac{1}{2} & -\frac{1}{4} \\ 0 & 0 & \frac{1}{2} \end{bmatrix}$$

由上述计算结果知,两种方法都可以求解出相同的 $f(\mathbf{A})$,且求得的 $f(\mathbf{A})$ 就是矩阵 \mathbf{A} 的逆矩阵. 这表明例 4.7.12 定义的 $f(\mathbf{A})$ 也是有意义的,也可以称为矩阵函数. 同时也说明定义 4.7.1 是有局限性的. 该定义要求函数 $f(z)$ 能够在 $z=0$ 处展开为幂级数,即要求满足以下两个条件:

(1) $f^{(k)}(0)$ 存在, $k=0,1,\cdots$;

(2) $\lim_{k \rightarrow \infty} \frac{f^{(k+1)}(\xi)}{(k+1)!} z^{k+1} = 0$.

Sylvester 公式和谱上一致性方法所需条件为 $f(z)$ 在矩阵 \mathbf{A} 的谱上给定(尽管在定义或结论中仍要求矩阵的谱半径小于收敛半径,实际上我们可移除这一条件). 谱上给定条件比定义 4.7.1 的幂级数要求弱. 因此,可拓宽定义 4.7.1 矩阵函数的定义.

定义 4.7.4(矩阵函数) 设复方阵 \mathbf{A} 的最小多项式为

$$m_{\mathbf{A}}(\lambda) = (\lambda - \lambda_1)^{n_1} (\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_s)^{n_s}, \deg(m_{\mathbf{A}}(\lambda)) = l$$

若函数 $f(\lambda)$ 在矩阵 \mathbf{A} 的谱上给定,则矩阵函数 $f(\mathbf{A})$ 定义为 $f(\mathbf{A}) = p(\mathbf{A})$, 式中, $p(\mathbf{A}) = \beta_0 \mathbf{I} + \beta_1 \mathbf{A} + \cdots + \beta_{l-1} \mathbf{A}^{l-1}$, l 个系数 $\beta_0, \beta_1, \cdots, \beta_{l-1}$ 由以下方程组确定:

$$\begin{cases} f(\lambda_i) = p(\lambda_i) \\ f'(\lambda_i) = p'(\lambda_i) \\ \vdots \\ f^{(n_i-1)}(\lambda_i) = p^{(n_i-1)}(\lambda_i) \end{cases}, \quad i=1,2,\cdots,s$$

注 7: 矩阵函数也可以利用 Sylvester 公式定义,这里就不再赘述了. 特别地,当函数 $f(z)$ 能够展开为“ z ”的幂级数时,矩阵函数的定义 4.7.1 和定义 4.7.4 一致.

例 4.7.13 考察齐次线性差分方程

$$u_{k+3} + u_{k+2} - u_{k+1} - u_k = 0$$

当 $k < 0$ 时, $u_k = 0$ 且有 $u_1 = 0, u_2 = 1, u_3 = 0$. 求 $\{u_k\}$.

解: 定义

$$\mathbf{x}_k = \begin{bmatrix} u_k \\ u_{k+1} \\ u_{k+2} \end{bmatrix} \in \mathbb{R}^3, \quad \mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & -1 \end{bmatrix}$$

则有

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k, \quad \mathbf{x}_1 = [0, 1, 0]^T$$

对上述进行迭代得

$$\mathbf{x}_{k+1} = \mathbf{A}^k \mathbf{x}_1, \quad k=1, 2, \dots$$

由于 \mathbf{A} 是友矩阵, 可直接写出它的最小多项式为 $m_{\mathbf{A}}(\lambda) = (\lambda+1)^2(\lambda-1)$. 现利用谱上一致性求解矩阵函数 \mathbf{A}^k . 定义 $f(\lambda) = \lambda^k$ 和 $p(\lambda) = a_2\lambda^2 + a_1\lambda + a_0$, 则有如下方程组:

$$\begin{aligned} a_2 + a_1 + a_0 &= 1 \\ a_2 - a_1 + a_0 &= (-1)^k \\ -2a_2 + a_1 &= k(-1)^{k-1} \end{aligned}$$

$$\text{解得 } a_2 = \frac{1}{4} - \frac{1}{4}(1-2k)(-1)^k, \quad a_1 = \frac{1}{2} - \frac{1}{2}(-1)^k, \quad a_0 = \frac{1}{4} + \frac{1}{4}(3-2k)(-1)^k.$$

由此, 求得

$$\mathbf{A}^k = \begin{bmatrix} a_0 & a_2 + a_1 & 2a_2 \\ -2a_2 & 3a_2 + a_0 & a_1 \\ -a_2 - a_1 & -3a_2 + a_1 & -a_2 + a_1 + a_0 \end{bmatrix}$$

于是,

$$\mathbf{x}_k = \mathbf{A}^{k-1} \mathbf{x}_1 = \begin{bmatrix} a_2 + a_1 \\ 3a_2 + a_0 \\ -3a_2 + a_1 \end{bmatrix}$$

因此,

$$u_k = a_2 + a_1 = \frac{3}{4} - \frac{1}{4}(3-2k)(-1)^k$$

4.8 函数矩阵

定义 4.8.1 (函数矩阵) 以变量 t 的函数为元素的矩阵 $\mathbf{A}(t) = (a_{ij}(t))_{m \times n}$ 称为函数(值)矩阵; 若矩阵 $\mathbf{A}(t)$ 每个元素 $a_{ij}(t)$ 在 $[a, b]$ 上是连续、可微或可积时, 则称 $\mathbf{A}(t)$ 在 $[a, b]$ 上连续、可微或可积, 并定义

$$\begin{aligned} \mathbf{A}'(t) &= \frac{d}{dt} \mathbf{A}(t) = (a'_{ij}(t))_{m \times n} \\ \int_a^b \mathbf{A}(t) dt &= \left(\int_a^b a_{ij}(t) dt \right)_{m \times n} \end{aligned}$$

例 4.8.1 求函数矩阵 $\mathbf{A}(t) = \begin{bmatrix} \sin t & \cos t & t \\ e^t & e^{2t} & e^{3t} \\ 0 & 1 & t^2 \end{bmatrix}$ 的导数.

解:

$$\frac{d}{dt}\mathbf{A}(t) = \begin{bmatrix} \cos t & -\sin t & 1 \\ e^t & 2e^{2t} & 3e^{3t} \\ 0 & 0 & 2t \end{bmatrix}$$

命题 4.8.1 设 $\mathbf{A}(t)$ 和 $\mathbf{B}(t)$ 是适当阶的可微矩阵, 则

- (1) $\frac{d}{dt}(\mathbf{A}(t) + \mathbf{B}(t)) = \frac{d}{dt}\mathbf{A}(t) + \frac{d}{dt}\mathbf{B}(t);$
- (2) $\lambda(t)$ 为可微函数, $\frac{d}{dt}(\lambda(t)\mathbf{A}(t)) = \frac{d\lambda(t)}{dt}\mathbf{A}(t) + \lambda(t)\frac{d}{dt}\mathbf{A}(t);$
- (3) $\frac{d}{dt}(\mathbf{A}(t)\mathbf{B}(t)) = \frac{d\mathbf{A}(t)}{dt}\mathbf{B}(t) + \mathbf{A}(t)\frac{d\mathbf{B}(t)}{dt};$
- (4) 当 $u=f(t)$ 可微时, $\frac{d}{dt}(\mathbf{A}(u)) = f'(t)\frac{d}{dt}\mathbf{A}(t);$
- (5) 当 $\mathbf{A}(t)$ 是可逆矩阵时, 有 $\frac{d}{dt}(\mathbf{A}^{-1}(t)) = -\mathbf{A}^{-1}(t)\left(\frac{d}{dt}\mathbf{A}(t)\right)\mathbf{A}^{-1}(t).$

证明: (1) 令 $\mathbf{A}(t) = (a_{ij}(t))_{m \times n}$, $\mathbf{B}(t) = (b_{ij}(t))_{m \times n}$, 则

$$\begin{aligned} \frac{d}{dt}(\mathbf{A}(t) + \mathbf{B}(t)) &= \frac{d}{dt}[(a_{ij}(t))_{m \times n} + (b_{ij}(t))_{m \times n}] \\ &= (a'_{ij}(t))_{m \times n} + (b'_{ij}(t))_{m \times n} \\ &= \frac{d}{dt}\mathbf{A}(t) + \frac{d}{dt}\mathbf{B}(t) \end{aligned}$$

(2) 令 $\mathbf{A}(t) = (a_{ij}(t))_{m \times n}$, 则

$$\begin{aligned} \frac{d}{dt}(\lambda(t)\mathbf{A}(t)) &= \frac{d}{dt}(\lambda(t)a_{ij}(t))_{m \times n} \\ &= \frac{d\lambda(t)}{dt}(a_{ij}(t))_{m \times n} + \lambda(t)(a'_{ij}(t))_{m \times n} \\ &= \frac{d\lambda(t)}{dt}\mathbf{A}(t) + \lambda(t)\frac{d}{dt}\mathbf{A}(t) \end{aligned}$$

(3) 令 $\mathbf{A}(t) = (a_{ik}(t))_{m \times n}$, $\mathbf{B}(t) = (b_{kj}(t))_{n \times p}$, 则

$$\begin{aligned} \frac{d}{dt}(\mathbf{A}(t)\mathbf{B}(t)) &= \frac{d}{dt}\left(\sum_{k=1}^n a_{ik}(t)b_{kj}(t)\right)_{m \times p} \\ &= \left(\sum_{k=1}^n \left(\frac{d}{dt}a_{ik}(t)\right)b_{kj}(t) + \sum_{k=1}^n a_{ik}(t)\left(\frac{d}{dt}b_{kj}(t)\right)\right)_{m \times p} \\ &= \frac{d\mathbf{A}(t)}{dt}\mathbf{B}(t) + \mathbf{A}(t)\frac{d\mathbf{B}(t)}{dt} \end{aligned}$$

$$(4) \frac{d}{dt}(\mathbf{A}(u)) = (a'_{ij}(u))_{m \times n} = u'(a'(t))_{m \times n} = f'(t) \frac{d}{dt} \mathbf{A}(t).$$

(5) 等式 $\mathbf{A}(t)\mathbf{A}^{-1}(t) = \mathbf{I}$ ($\mathbf{A}(t)$ 为方阵) 两端对 t 求导, 得

$$\left(\frac{d}{dt}\mathbf{A}(t)\right)\mathbf{A}^{-1}(t) + \mathbf{A}(t)\left(\frac{d}{dt}\mathbf{A}^{-1}(t)\right) = \mathbf{O}$$

从而

$$\left(\frac{d}{dt}\mathbf{A}^{-1}(t)\right) = -\mathbf{A}^{-1}(t)\left(\frac{d}{dt}\mathbf{A}(t)\right)\mathbf{A}^{-1}(t)$$

命题 4.8.2 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 则有

$$(1) \frac{d}{dt} e^{\mathbf{A}t} = \mathbf{A}e^{\mathbf{A}t} = e^{\mathbf{A}t}\mathbf{A};$$

$$(2) \frac{d}{dt} \sin \mathbf{A}t = \mathbf{A} \cos \mathbf{A}t = (\cos \mathbf{A}t)\mathbf{A};$$

$$(3) \frac{d}{dt} \cos \mathbf{A}t = -\mathbf{A} \sin \mathbf{A}t = -(\sin \mathbf{A}t)\mathbf{A}.$$

推论 4.8.1 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 则有

$$(1) \int_{t_0}^t \mathbf{A} e^{\mathbf{A}s} ds = e^{\mathbf{A}t} - e^{\mathbf{A}t_0};$$

$$(2) \int_{t_0}^t \mathbf{A} \sin \mathbf{A}s ds = \cos \mathbf{A}t_0 - \cos \mathbf{A}t;$$

$$(3) \int_{t_0}^t \mathbf{A} \cos \mathbf{A}s ds = \sin \mathbf{A}t_0 - \sin \mathbf{A}t.$$

命题 4.8.3 设 $\mathbf{A}(t)$ 和 $\mathbf{B}(t)$ 是 $[a, b]$ 上适当阶的可积矩阵, $\lambda \in \mathbb{C}$, 则有

$$(1) \int_a^b (\mathbf{A}(t) + \mathbf{B}(t)) dt = \int_a^b \mathbf{A}(t) dt + \int_a^b \mathbf{B}(t) dt;$$

$$(2) \int_a^b \lambda \mathbf{A}(t) dt = \lambda \int_a^b \mathbf{A}(t) dt;$$

$$(3) \text{若 } \mathbf{A}(t) \text{ 在 } [a, b] \text{ 上连续, 则 } \forall t \in (a, b) \text{ 有 } \frac{d}{dt} \left(\int_a^t \mathbf{A}(\tau) d\tau \right) = \mathbf{A}(t);$$

$$(4) \text{若 } \mathbf{A}(t) \text{ 在 } [a, b] \text{ 上可微, 则有 } \int_a^b \mathbf{A}'(\tau) d\tau = \mathbf{A}(b) - \mathbf{A}(a).$$

定义 4.8.2 (矩阵对矩阵的导数) 设 $\mathbf{F}(\mathbf{X}) = (f_{ij}(\mathbf{X})) \in \mathbb{C}^{m \times n}$ 是 $\mathbf{X} \in \mathbb{C}^{p \times q}$ 的函数矩阵, $f_{ij}(\mathbf{X}) (i=1, \dots, m; j=1, \dots, n)$ 作为 $\mathbf{X} \in \mathbb{C}^{p \times q}$ 的多元函数是可微的. 令

$$\frac{d\mathbf{F}(\mathbf{X})}{d\mathbf{X}} = \left(\frac{\partial \mathbf{F}}{\partial x_{ij}} \right)_{m \times p \times nq} = \begin{bmatrix} \frac{\partial \mathbf{F}}{\partial x_{11}} & \frac{\partial \mathbf{F}}{\partial x_{12}} & \cdots & \frac{\partial \mathbf{F}}{\partial x_{1q}} \\ \frac{\partial \mathbf{F}}{\partial x_{21}} & \frac{\partial \mathbf{F}}{\partial x_{22}} & \cdots & \frac{\partial \mathbf{F}}{\partial x_{2q}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial \mathbf{F}}{\partial x_{p1}} & \frac{\partial \mathbf{F}}{\partial x_{p2}} & \cdots & \frac{\partial \mathbf{F}}{\partial x_{pq}} \end{bmatrix}$$

式中: $\frac{\partial \mathbf{F}}{\partial x_{ij}} = \left(\frac{\partial f_{kl}(\mathbf{X})}{\partial x_{ij}} \right) (k=1, \dots, m; l=1, \dots, n)$, 称 $\frac{d\mathbf{F}(\mathbf{X})}{d\mathbf{X}}$ 为 $\mathbf{F}(\mathbf{X})$ 对 \mathbf{X} 的导数.

注 1: 定义 4.8.2 既是对定义 4.8.1 中矩阵函数导数的扩展, 也隐含着向量对向量、向量对矩阵、矩阵对向量和矩阵对矩阵的导数定义.

例 4.8.2 设 $f(\mathbf{x}) = f(x_1, \dots, x_n)$ 是 n 元实可微函数, 则

$$\frac{df}{d\mathbf{x}} = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right]$$

式中: $\mathbf{x} = [x_1, \dots, x_n]$. 我们常称 $\frac{df}{d\mathbf{x}}$ 为 f 的梯度向量, 常记为 $\nabla(f)$.

例 4.8.3 设 $\mathbf{f}(\mathbf{x}) = [f_1, f_2, \dots, f_m]^T \in F^m$ 是 n 元可微向量函数, 即每个 n 元函数 $f_i(\mathbf{x}) = f_i(x_1, \dots, x_n)$, $1 \leq i \leq m$ 均为可微函数, 其中 $\mathbf{x} = [x_1, \dots, x_n]$ 为行向量.

$$\frac{d\mathbf{f}}{d\mathbf{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \dots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

矩阵 $\frac{d\mathbf{f}}{d\mathbf{x}}$ 常称为向量函数 \mathbf{f} 的 Jacobian 矩阵.

例 4.8.4 设常矩阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 和未定元向量 $\mathbf{x} \in \mathbb{R}^n$, 求 $\frac{d\mathbf{A}\mathbf{x}}{d\mathbf{x}^T}$.

解: 令 $\mathbf{y} = \mathbf{A}\mathbf{x} = [y_1, \dots, y_n]^T$, $\mathbf{x} = [x_1, \dots, x_n]^T$, 其中 $y_i = \sum_{j=1}^n a_{ij}x_j$, 则

$$\frac{d\mathbf{y}}{d\mathbf{x}^T} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \dots & \frac{\partial y_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial y_n}{\partial x_1} & \dots & \frac{\partial y_n}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} = \mathbf{A}$$

例 4.8.5 设常矩阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$, \mathbf{x} 和 $\mathbf{b} \in \mathbb{R}^n$, 求 $\frac{d}{d\mathbf{x}}(\mathbf{b}^T \mathbf{x})$ 和 $\frac{d}{d\mathbf{x}}(\mathbf{x}^T \mathbf{A} \mathbf{x})$.

解: 设 $\mathbf{x} = [x_1, \dots, x_n]^T$, $\mathbf{b} = [b_1, \dots, b_n]^T$, $\mathbf{A} = (a_{ij})$, 则

$$\begin{aligned} \mathbf{b}^T \mathbf{x} &= \sum_{i=1}^n b_i x_i \\ \mathbf{x}^T \mathbf{A} \mathbf{x} &= \sum_{i,j=1}^n a_{ij} x_i x_j \end{aligned}$$

于是,

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}(\mathbf{b}^{\mathrm{T}}\mathbf{x}) = \begin{bmatrix} \frac{\partial(\mathbf{b}^{\mathrm{T}}\mathbf{x})}{\partial x_1} \\ \frac{\partial(\mathbf{b}^{\mathrm{T}}\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial(\mathbf{b}^{\mathrm{T}}\mathbf{x})}{\partial x_n} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \mathbf{b}$$

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}) = \begin{bmatrix} \frac{\partial(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x})}{\partial x_1} \\ \frac{\partial(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x})}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n a_{i1}x_i + \sum_{j=1}^n a_{1j}x_j \\ \sum_{i=1}^n a_{i2}x_i + \sum_{j=1}^n a_{2j}x_j \\ \vdots \\ \sum_{i=1}^n a_{in}x_i + \sum_{j=1}^n a_{nj}x_j \end{bmatrix} = (\mathbf{A}^{\mathrm{T}} + \mathbf{A})\mathbf{x}$$

【应用】 求解最小二乘问题:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 = \|\mathbf{A}\boldsymbol{\lambda} - \mathbf{b}\|_2^2 \quad (4.8.1)$$

式中: $\mathbf{A} \in \mathbb{R}^{m \times n}$ 和 $\mathbf{b} \in \mathbb{R}^m$ 给定, $\mathbf{x} \in \mathbb{R}^n$ 为待定向量.

引理 4.8.1 (无约束极值的必要条件) 设 $\mathbf{X} \in \mathbb{R}^{m \times n}$, $f(\mathbf{X})$ 是关于矩阵变量 \mathbf{X} 的二阶连续可导函数. 若 \mathbf{X}_0 是 $f(\mathbf{X})$ 的一个极值点, 则

$$\left. \frac{\mathrm{d}f(\mathbf{X})}{\mathrm{d}\mathbf{X}} \right|_{\mathbf{x}=\mathbf{x}_0} = \mathbf{O}$$

若 $f(\mathbf{X})$ 还是关于矩阵变量 \mathbf{X} 的凸函数, 则 $f(\mathbf{X})$ 的极值点必是最值点.

由引理 4.8.1 知, 最小二乘问题 (4.8.1) 的极小值点须满足

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}(\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2) = 0$$

整理得

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}(\mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} - 2\mathbf{b}^{\mathrm{T}}\mathbf{A}\mathbf{x} + \mathbf{b}^{\mathrm{T}}\mathbf{b}) = 0 \quad (4.8.2)$$

式中:

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}(\mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x}) = 2\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x}$$

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}(\mathbf{b}^{\mathrm{T}}\mathbf{A}\mathbf{x}) = \mathbf{A}^{\mathrm{T}}\mathbf{b}$$

故式 (4.8.2) 等价于

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{A}^{\mathrm{T}}\mathbf{b} \quad (4.8.3)$$

因此,最小二乘问题(4.8.1)的解必满足方程(4.8.3).

【思考】 某些时候,求解最小二乘问题的方程(4.8.3)是病态的(常见情况是 \mathbf{A} 为列满秩,因计算 $(\mathbf{A}^T \mathbf{A})^{-1}$ 导致方程(4.8.3)的解有较大误差),此时如何获得最小二乘问题更可靠的解?

提示:若 \mathbf{A} 为列满秩,可考虑对矩阵 \mathbf{A} 进行 QR 分解.

【应用】 考察一阶线性常系数微分方程组的求解问题:

$$\begin{cases} \frac{dx_1(t)}{dt} = a_{11}x_1(t) + \cdots + a_{1n}x_n(t) + f_1(t) \\ \vdots \\ \frac{dx_n(t)}{dt} = a_{n1}x_1(t) + \cdots + a_{nn}x_n(t) + f_n(t) \end{cases} \quad (4.8.4)$$

方程(4.8.4)满足初始条件 $x_i(t_0) = c_i, i = 1, \cdots, n$. 由微分方程的理论知,上述方程的解存在且唯一.

方程(4.8.4)可改写为

$$\begin{cases} \frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{f}(t) \\ \mathbf{x}(t_0) = \mathbf{c} \end{cases} \quad (4.8.5)$$

式中: $\mathbf{x}(t) = [x_1(t), \cdots, x_n(t)]^T, \mathbf{f}(t) = [f_1(t), \cdots, f_n(t)]^T, \mathbf{c} = [c_1, \cdots, c_n]^T$ 和 $\mathbf{A} = (a_{ij})_{n \times n}$.

对式(4.8.5)积分得

$$\int_{t_0}^t \frac{d(e^{-\mathbf{A}t}\mathbf{x}(t))}{dt} dt = \int_{t_0}^t e^{-\mathbf{A}t}\mathbf{f}(t) dt \quad (4.8.6)$$

即

$$e^{-\mathbf{A}t}\mathbf{x}(t) - e^{-\mathbf{A}t_0}\mathbf{x}(t_0) = \int_{t_0}^t e^{-\mathbf{A}t}\mathbf{f}(t) dt$$

由此,方程(4.8.5)的解为

$$\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{c} + e^{\mathbf{A}t} \int_{t_0}^t e^{-\mathbf{A}t}\mathbf{f}(t) dt \quad (4.8.7)$$

或

$$\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{c} + \int_{t_0}^t e^{\mathbf{A}(t-\tau)}\mathbf{f}(\tau) d\tau \quad (4.8.8)$$

特别地,当 $\mathbf{f}(t) \equiv \mathbf{0}$ 时,即齐次线性方程组的解为

$$\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{c} \quad (4.8.9)$$

例 4.8.6 已知方程(4.8.5)中的参数分别为

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & 3 \end{bmatrix}, \quad \mathbf{f}(t) = [1, -t, t]^T, \quad \mathbf{c} = [1, 0, -1]^T$$

求该方程组的解.

$$\text{解: 由 } \lambda \mathbf{I} - \mathbf{A} \cong \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda - 2 & 0 \\ 0 & 0 & (\lambda - 2)^2 \end{bmatrix} \text{ 知, } m_{\mathbf{A}}(\lambda) = (\lambda - 2)^2.$$

令 $p_t(\lambda) = \alpha_0(t) + \alpha_1(t)\lambda$, 则有 $\alpha_0(t) + 2\alpha_1(t) = e^{2t}$, $\alpha_1(t) = te^{2t}$, 并解得 $\alpha_0(t) = (1 - 2t)e^{2t}$, $\alpha_1(t) = te^{2t}$. 因此,

$$e^{\mathbf{A}t} = \alpha_0(t)\mathbf{I} + \alpha_1(t)\mathbf{A} = e^{2t} \begin{bmatrix} 1 & 0 & 0 \\ t & 1 - t & t \\ t & -t & 1 + t \end{bmatrix}$$

此时, 方程的解为

$$\mathbf{x}(t) = e^{2t} \begin{bmatrix} \frac{3}{2} - \frac{1}{2}2e^{-2t} \\ \frac{1}{2}(t^2 + t - 2) + \left(-\frac{t^2}{2} + \frac{3t}{2} + 1\right)e^{-2t} \\ \left(2t^2 + t + \frac{1}{2}\right)e^{-2t} - \frac{3}{2} \end{bmatrix}$$

再考察 n 阶常系数微分方程的求解问题:

$$y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_0y = u(t) \quad (4.8.10)$$

式中, a_0, \cdots, a_{n-1} 为常数, $u(t)$ 为已知函数. 若 $u(t) \neq 0$ 时, 称方程 (4.8.10) 为非齐次微分方程, 否则称为齐次微分方程.

定义

$$\begin{cases} x_1 = y \\ x_2 = \dot{y} \\ \vdots \\ x_n = y^{(n-1)} \end{cases} \quad (4.8.11)$$

由此, 得到如下方程:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{f}(\mathbf{x}) \quad (4.8.12)$$

式中:

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \in \mathbb{R}^n$$

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & 0 \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix}$$

$$\mathbf{f}(t) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ u(t) \end{bmatrix}$$

式(4.8.12)在现代控制理论中称为状态方程, $\mathbf{x}(t)$ 称为状态向量, \mathbf{A} 为系统矩阵. 此时, 可将 n 阶常系数微分方程(4.8.10)转化为状态方程(4.8.12)进行求解.

例 4.8.7 求解非齐次微分方程组

$$y^{(3)} - 6\ddot{y} + 11\dot{y} - 6y = \sin t$$

式中: $\ddot{y}(0) = \dot{y}(0) = y(0) = 0$.

解: 令 $y = x_1, \dot{y} = x_2, \ddot{y} = x_3$, 则原方程可改写为 $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{f}(\mathbf{x})$, 其中 $\mathbf{x}(t) = [x_1(t), x_2(t), x_3(t)]^T$ 且 $\mathbf{x}(0) = [0, 0, 0]^T$,

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 6 & -11 & 6 \end{bmatrix}, \quad \mathbf{f}(\mathbf{x}) = \begin{bmatrix} 0 \\ 0 \\ \sin t \end{bmatrix}$$

由此,

$$\mathbf{x}(t) = e^{\mathbf{A}t} \mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-\tau)} \mathbf{f}(\tau) d\tau = \int_0^t e^{\mathbf{A}(t-\tau)} \begin{bmatrix} 0 \\ 0 \\ \sin \tau \end{bmatrix} d\tau$$

计算矩阵函数 $e^{\mathbf{A}t}$ 得

$$e^{\mathbf{A}t} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{bmatrix} \begin{bmatrix} e^t & 0 & 0 \\ 0 & e^{2t} & 0 \\ 0 & 0 & e^{3t} \end{bmatrix} \begin{bmatrix} 3 & -2.5 & 0.5 \\ -3 & 4 & -1 \\ 1 & -1.5 & 0.5 \end{bmatrix}$$

因此,

$$\mathbf{x}(t) = \frac{1}{20} \begin{bmatrix} 5e^t - 4e^{2t} + e^{3t} - 2\cos t \\ 5e^t - 8e^{2t} + 3e^{3t} + 2\sin t \\ 5e^t - 16e^{2t} + 9e^{3t} + 2\cos t \end{bmatrix}$$

$$\text{即 } y(t) = x_1(t) = \frac{1}{20} (5e^t - 4e^{2t} + e^{3t} - 2\cos t).$$

【应用】 系统的可控性是指对任意给定的系统初始状态, 均存在一个合适控制输入使得系统状态在有限时间转移至状态零点. 它表征了系统输入对系统状态的有效控制能力, 是现代控制理论的基础性概念. 例 4.8.8 即为线性定常系统(即采用线性常微分方程描述的系统)可控的充分必要条件.

例 4.8.8 n 阶 Hermite 矩阵 $\mathbf{W}(0, t_1)$ 是正定的, 当且仅当矩阵 $\text{rank}(\mathbf{C}) = n$, 其中

$$\mathbf{W}(0, t_1) = \int_0^{t_1} e^{-\mathbf{A}t} \mathbf{B} \mathbf{B}^T e^{-\mathbf{A}^T t} dt$$

$$C = [B, AB, \dots, A^{n-1}B]$$

$A \in \mathbb{R}^{n \times n}$ 是线性定常系统的系统矩阵, $B \in \mathbb{R}^{n \times p}$ 是线性定常系统的输入矩阵, t_1 是给定的正常数.

证明: 充分性. 采用反证法证明. 假设矩阵 $W(0, t_1)$ 不是正定的. 由 $W(0, t_1)$ 表达式知, 对任意非零向量 $x \in \mathbb{R}^n$, 有

$$x^T W(0, t_1) x = \int_0^{t_1} (B^T e^{-A^T t} x, B^T e^{-A^T t} x) dt \geq 0$$

因此, 必存在一非零向量 α 满足 $\alpha^T W(0, t_1) \alpha = 0$, 于是

$$\int_0^{t_1} (B^T e^{-A^T t} \alpha, B^T e^{-A^T t} \alpha) dt = 0$$

即对 $t \in [0, t_1]$,

$$\alpha^T e^{-At} B = 0$$

对上式求导直至 $n-1$ 次后, 再令 $t=0$ 得(此处也可以利用 4.7.4 矩阵函数定义进行讨论)

$$\alpha^T A^i B = 0, \quad i = 0, 1, \dots, n-1$$

进一步整理得 $\alpha^T C = 0$, 即 $C^T \alpha = 0$. 这意味着齐次线性方程组 $C^T \alpha = 0$ 有非零解, 由亏加秩定理知, $\text{rank}(C) < n$. 这与已知矛盾, 故矩阵 $W(0, t_1)$ 是正定的.

必要性. 假设 $\text{rank}(C) < n$, 则必存在一非零向量 α 满足 $\alpha^T C = 0$. 于是有

$$\alpha^T A^i B = 0, \quad i = 0, 1, \dots, n-1$$

由定义 4.7.4 知,

$$e^{-At} = \beta_0(t)I + \beta_1(t)A + \dots + \beta_{n-1}(t)A^{n-1}$$

式中: $\beta_i(t) (i=0, 1, \dots, n-1)$ 由矩阵 A 唯一确定.

此时,

$$\alpha^T e^{-At} B = \beta_0(t) \alpha^T B + \dots + \beta_{n-1}(t) A \alpha^{T^{n-1}} B = 0$$

即

$$\int_0^{t_1} (B^T e^{-A^T t} \alpha, B^T e^{-A^T t} \alpha) dt = 0$$

这表明矩阵 $W(0, t_1)$ 不是正定的, 与已知矛盾. 故 $\text{rank}(C) = n$. 证毕.

4.9 应用: 主元分析法

算法的复杂度和数据维数往往有着密切关系, 甚至与数据维数呈指数关联. 实际中, 处理成千上万甚至几十万维的数据情况也并不罕见, 此时的消耗可能难以接受, 需要对数据进行降维. 数据降维直观的好处是降低维数便于计算和可视化, 更深层次的意义在于有效信息的提取、综合及无用信息的摒弃.

主元分析法是一种常用的数据降维方法, 其目的是在“信息损失”较小的前提下, 将高维数据转换到低维. 基于此, 我们给出这一问题的数学描述.

假设一组数据有 n 表示样本数, 即 x_1, \dots, x_n , 其中, $x_i \in \mathbb{R}^p$ 表示每个样本向量 (或随机变量). 设降维后的样本向量为 $y_i \in \mathbb{R}^q$ ($q < p$, 否则失去降维的意义), 则线性空间 $\text{span}(y_1, \dots, y_n)$ 应与线性空间 $\text{span}(x_1, \dots, x_n)$ 同构. 由此, 存在同构映射 f 使得原像 x_i 与像 y_i 满足

$$y_i = f(x_i)$$

式中: f 为待定映射. 由于 f 是线性映射, 故上述可等价表示为

$$y_i = Wx_i = \begin{bmatrix} w_1^T x_i \\ \vdots \\ w_p^T x_i \end{bmatrix}$$

式中: 矩阵 $W \in \mathbb{R}^{q \times p}$, $w_i \in \mathbb{R}^p$ 为待定向量.

考察表达式 $w_k^T x_i$. 在欧氏空间中,

$$w_k^T x_i = (w_k, x_i)$$

若选取 $\|w_k\| = 1$, 则上式表示向量 x_i 在向量 w_k 的正交投影的长度. 由此, 我们选定 w_i ($i=1, \dots, q$) 长度为 1, 则 y_i 的每一个分量即是原像 x_i 在矩阵 W 行向量的正交投影. 为了保证像 y_i 尽量继承原始变量的特征, 应选择 w_i ($i=1, \dots, q$) 是正交向量组. 所以, 矩阵 W 是行正交规范矩阵, 即向量组 $w_i, i=1, \dots, q$ 是单位正交向量组.

为刻画“信息损失”较小, 我们考察二维平面的一个例子. 在图 4.9.1 中, 点 x_1, x_2, x_3, x_4 表示样本向量; 若将样本向量降为一维, 则待定向量 (组) w 是二维平面的一条过原点的直线. 当直线 w 绕原点旋转时, 样本向量 x_1, x_2, x_3, x_4 在直线 w 上的正交投影也会随之变化. 当 w 取为直线 w_1 时, 样本向量 x_1, x_2, x_3, x_4 在直线 w 上的正交投影重叠为一点, 此时降维后的样本没有意义了 (只剩下一个样本向量, 而其他样本信息全部丢失). 所以, 我们应反其道而行之, 要保证信息损失最小就要使得它在直线 w 上的正交投影尽可能地“散开”. 显然, 当 w 取为直线 w_2 时, 样本向量 x_1, x_2, x_3, x_4 在直线 w 上的正交投影“散开”最大.

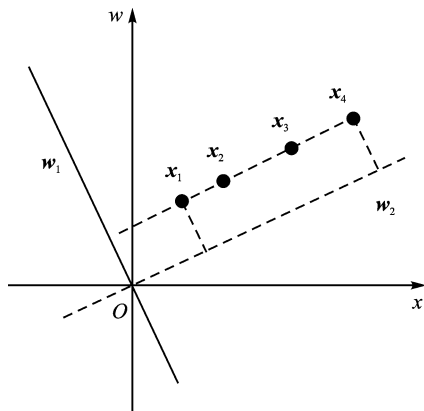


图 4.9.1 二维平面信息损失示例

在数学上,可以用方差来表述样本分散的程度.于是,优化目标可表达为

$$\hat{\mathbf{w}} = \operatorname{argmax}_{\mathbf{w}} \frac{1}{n} \sum_{i=1}^n (\mathbf{w}^T \mathbf{x}_i - \mu)^2 \quad (4.9.1)$$

式中: μ 是样本投用的均值,其定义式为

$$\mu = \frac{1}{n} \sum_{i=1}^n \mathbf{w}^T \mathbf{x}_i$$

注意到

$$\begin{aligned} \sigma^2 &= \frac{1}{n} \sum_{i=1}^n (\mathbf{w}^T \mathbf{x}_i - \mu)^2 \\ &= \frac{1}{n} \sum_{i=1}^n [\mathbf{w}^T (\mathbf{x}_i - \bar{\mathbf{x}})]^2 \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{w}^T (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{w} \\ &= \mathbf{w}^T \left[\frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^T \right] \mathbf{w} \end{aligned}$$

并定义矩阵

$$\mathbf{x} = [\mathbf{x}_1 - \bar{\mathbf{x}}, \mathbf{x}_2 - \bar{\mathbf{x}}, \dots, \mathbf{x}_n - \bar{\mathbf{x}}] \in \mathbb{R}^{p \times n}$$

则方差 σ^2 可进一步改写为

$$\sigma^2 = \mathbf{w}^T \left[\frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^T \right] \mathbf{w} = \mathbf{w}^T \mathbf{C} \mathbf{w} \quad (4.9.2)$$

式中: $\mathbf{C} = \frac{1}{n} \mathbf{x} \mathbf{x}^T \in \mathbb{R}^{p \times p}$ 是随机变量 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 的协方差矩阵.

由引理 3.9.2 知, \mathbf{C} 是半正定实对称矩阵,故

$$\sigma^2 = R(\mathbf{w}) \quad (4.9.3)$$

式中: $R(\mathbf{w})$ 是协方差矩阵 \mathbf{C} 的 Rayleigh 商.

方法一. 为讨论方便, 设 $\lambda_1 > \lambda_2 > \dots > \lambda_m$ 是矩阵 \mathbf{C} 的 m 个互异特征值, 其代数重数分别为 d_1, d_2, \dots, d_l 且有 $d_1 + d_2 + \dots + d_l = p$. 根据定理 4.3.4 知,

$$R(\mathbf{w}) \leq \lambda_1$$

上式等号成立的条件为向量 \mathbf{w} 取为属于特征值 λ_1 的 d_1 个单位正交特征向量组中的一个向量. 由此, 我们找到了 d_1 个 \mathbf{w} 向量. 那么余下的 $q - d_1$ 个 \mathbf{w} 向量应如何确定呢?

应在特征值为 $\lambda_2, \dots, \lambda_m$ 的特征子空间的和空间来寻找(为什么? 请读者思考), 即

$$\hat{\mathbf{w}} = \arg \max_{\mathbf{w} \in \left(\bigoplus_{i=2, \dots, m} E(\lambda_i) \right)} R(\mathbf{w}) \quad (4.9.4)$$

根据定理 4.3.4 知,

$$\max_{\substack{\oplus E(\lambda_i) \\ w \in (i=2, \dots, m)}} R(w) \leq \lambda_2$$

上式等号成立的条件为向量 w 取为属于特征值 λ_2 的 d_2 个单位正交特征向量组中的一个向量. 由此, 我们找到了 d_2 个 w 向量. 依此类推, 我们可以找到 q 个满足要求的向量 w . 这 q 个向量 w 恰好是协方差矩阵 C 的属于前 q 最大个特征值(计重数)的单位正交特征向量即可得到.

方法二. 由 Lagrange 乘子法构造目标函数:

$$L(w, \lambda) = w^T C w + \lambda(1 - w^T w) \quad (4.9.5)$$

对 $L(w, \mu)$ 求偏导数, 得

$$\frac{\partial L(w, \mu)}{\partial w} = 2Cw - 2\lambda w$$

令上式为零, 得以下极值条件

$$Cw = \lambda w$$

上式表明, λ 是协方差矩阵 C 的特征值, w 是属于特征值 λ 的单位特征向量.

求解目标函数 $Cw = \lambda w$, 得

$$\sigma^2 = \lambda \quad (4.9.6)$$

为保证方差最大, 只要我们对矩阵 C 的特征值进行排序, 然后选取属于前 q 最大特征值(计重数)的单位正交特征向量即可得到 W . 式(4.9.6)表明投影后的方程就是协方差矩阵的特征值.

MINST 数据库(见 <http://yann.lecun.com/exdb/mnist/>)是由 Yann 提供的手写数字数据库文件, 训练数据集包含 60 000 个样本, 测试数据集包含 10 000 样本. 数据集中的每张图片由 28×28 个像素点构成, 每个像素点用一个灰度值表示. 为说明主元分析性对降维的影响, 选取 28×28 像素进行列展开, 得到样本列向量 $x_i (i = 1, \dots, 60\,000)$. 选取 $q = 87$ (此时, 前 87 个特征值之和与 C 的所有特征值之和的比值超过 0.9). 利用 K-近邻算法将训练好后的模型在测试集上识别, 准确率为 97.28%, 识别所有测试样本所需时间为 85.34 s; 若直接用 K-近邻算法进行训练, 得出训练好后的模型在测试集上的识别准确率为 96.88%, 识别所有测试样所需时间为 688.90 s. 这表明主元分析法对原始图像的降维处理不仅能提高识别速度, 还能提高识别效率, 减少过拟合现象.

本章习题

1. 证明 $\lim_{p \rightarrow \infty} \|x\|_p = \|x\|_\infty$.

2. 设 $\|x\|_a$ 是 \mathbb{C}^m 上的向量范数, 给定矩阵 $A \in \mathbb{C}^{m \times n}$, 定义

$$\|x\|_\beta = \|Ax\|_a, \quad x \in \mathbb{C}^n$$

问 $\|\cdot\|_\beta$ 是 \mathbb{C}^n 上的向量范数吗? 若是, 请给出证明; 若不是, 请补充对矩阵 A 的

要求使其成为 \mathbb{C}^n 上的向量范数,并证明之.

3. 设 $\mathbf{x} \in \mathbb{C}^n$, 证明

$$\|\mathbf{x}\|_{\infty} \leq \|\mathbf{x}\|_1 \leq n \|\mathbf{x}\|_{\infty}$$

$$\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{n} \|\mathbf{x}\|_2$$

$$\|\mathbf{x}\|_{\infty} \leq \|\mathbf{x}\|_2 \leq \sqrt{n} \|\mathbf{x}\|_{\infty}$$

4. 设 $\mathbf{A} = (a_{ij}) \in \mathbb{C}^{n \times n}$, $\|\mathbf{A}\|_{v1} = \sum_{i,j=1}^n |a_{ij}|$ 是矩阵范数吗? 请说明理由.

5. 设 $\mathbf{A} = (a_{ij}) \in \mathbb{C}^{n \times n}$, $\|\mathbf{A}\|_{m\infty} = n \max_{i,j} |a_{ij}|$ 是矩阵范数吗? 请说明理由.

6. 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, 证明 $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F$.

7. 试证明对于任何矩阵 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 有

$$(1) \|\mathbf{A}^H\|_1 = \|\mathbf{A}^T\|_1 = \|\mathbf{A}\|_{\infty};$$

$$(2) \|\mathbf{A}^H\|_2 = \|\mathbf{A}^T\|_2 = \|\mathbf{A}\|_2;$$

$$(3) \|\mathbf{A}^H \mathbf{A}\|_2 = \|\mathbf{A}\|_2^2;$$

$$(4) \|\mathbf{A}\|_2^2 \leq \|\mathbf{A}\|_1 \|\mathbf{A}\|_{\infty}.$$

8. 若 \mathbf{L}_1 和 \mathbf{L}_2 是两个无向无权图的拉普拉斯矩阵, 且 0 均是这两个矩阵的单特征值, 证明:

$$\lambda_{n-1}(\mathbf{L}_1 + \mathbf{L}_2) \geq \lambda_{n-1}(\mathbf{L}_1) + \lambda_{n-1}(\mathbf{L}_2)$$

式中: $\lambda_{n-1}(\mathbf{L}_1)$ 是矩阵 \mathbf{L}_1 的次小特征值.

9. 设 \mathbf{A} 和 $\mathbf{B} \in \mathbb{C}^{n \times n}$ 且 \mathbf{A} 可逆, 若 $\|\mathbf{A}^{-1}\mathbf{B}\| < 1$, 证明:

(1) $(\mathbf{A} + \mathbf{B})$ 可逆;

$$(2) \|\mathbf{I} - (\mathbf{I} + \mathbf{A}^{-1}\mathbf{B})^{-1}\| \leq \frac{\|\mathbf{A}^{-1}\mathbf{B}\|}{1 - \|\mathbf{A}^{-1}\mathbf{B}\|}, \text{ 其中 } \|\mathbf{I}\| = 1.$$

10. 当矩阵范数取定为谱范数时, 求线性方程组 $\mathbf{Ax} = \mathbf{b}$ 的条件数.

11. 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, $\|\mathbf{A}\|$ 是矩阵范数. 若 $\|\mathbf{A}\| < 1$, 试利用盖尔圆盘定理证明 $\mathbf{I} - \mathbf{A}$ 非奇异.

$$12. \text{ 设 } \mathbf{A} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{6} & \frac{1}{6} & \frac{3}{6} & \frac{1}{6} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{3}{7} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{6} & \frac{1}{6} & \frac{3}{6} & \frac{1}{6} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{4}{7} \end{bmatrix}, \text{ 证明 } \rho(\mathbf{A}) < 1,$$

$$\rho(\mathbf{B}) = 1.$$

$$13. \text{ 设 } \mathbf{A} = \begin{bmatrix} 20 & 3 & 2.8 \\ 4 & 10 & 2 \\ 0.5 & 1 & 10i \end{bmatrix}, \text{ 试用盖尔圆盘定理证明该矩阵是单纯矩阵.}$$

14. 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 且 $\rho(\mathbf{A}) < 1$, 试求 $\sum_{k=1}^{\infty} k \mathbf{A}^k$ 及 $\sum_{k=1}^{\infty} k^2 \mathbf{A}^k$.

15. 若 $\mathbf{A} = \begin{bmatrix} -2 & 1 & -1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$, 试讨论 $\sum_{k=1}^{\infty} k^2 \mathbf{A}^k$ 的敛散性.

16. 设 $\mathbf{A} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{2}{3} \end{bmatrix}$, 判断 $\sum_{k=1}^{\infty} \frac{1}{k} \mathbf{A}^k$ 的敛散性.

17. 设 $\mathbf{A} = \begin{bmatrix} -1 & 0 \\ 2 & -1 \end{bmatrix}$, 判断 $\sum_{k=1}^{\infty} \frac{1}{k^2} \mathbf{A}^k$ 的敛散性.

18. 证明 $|\mathbf{e}^{\mathbf{A}}| = \mathbf{e}^{\text{tr}(\mathbf{A})}$.

19. 试用定义 4.7.1 计算矩阵函数 $\arcsin \frac{\mathbf{A}}{4}$, 其中 $\mathbf{A} = \begin{bmatrix} 0 & -1 \\ 4 & 4 \end{bmatrix}$.

20. 试分别用 Sylvester 公式和谱上一致性对以下矩阵函数:

(1) $\mathbf{A} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, 求 $\mathbf{A}^{\frac{1}{2}}$ 和 $\ln \mathbf{A}$;

(2) $\mathbf{A} = \begin{bmatrix} 0 & -1 \\ 4 & 4 \end{bmatrix}$, 求 $\arcsin \frac{\mathbf{A}}{4}$.

21. 求 $\frac{d}{dt}(\mathbf{e}^{\mathbf{A}t} \mathbf{Q} \mathbf{e}^{\mathbf{A}^H t})$, 其中 \mathbf{A} 和 \mathbf{Q} 为常复方阵.

22. 设矩阵函数 $\mathbf{A}(t) = \begin{bmatrix} \sin t & \cos t & 0 \\ t & e^t & t^2 \\ 1 & 0 & t^3 \end{bmatrix}$, 求 $\frac{d\mathbf{A}(t)}{dt}$, $\frac{d|\mathbf{A}(t)|}{dt}$, $\left| \frac{d\mathbf{A}(t)}{dt} \right|$.

23. 设 $\mathbf{X} \in \mathbb{R}^{n \times n}$, 求 $\frac{d}{d\mathbf{X}} \text{tr}(\mathbf{X})$, $\frac{d}{d\mathbf{X}} \text{tr}(\mathbf{X}\mathbf{X}^T)$, $\frac{d}{d\mathbf{X}} \text{tr}(\mathbf{X}^T \mathbf{X})$, $\frac{d}{d\mathbf{X}} \text{tr}(\mathbf{A}\mathbf{X})$, $\frac{d}{d\mathbf{X}} \text{tr}(\mathbf{X}^T \mathbf{B}\mathbf{X})$.

24. $\begin{cases} \frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) \\ \mathbf{x}(0) = [0, 1]^T \end{cases}$, $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$, 求 $\mathbf{x}(t)$.

25. 在主元分析法中, 优化目标为 $\max_{\mathbf{w}} \frac{1}{n} \mathbf{w}^T \mathbf{x} \mathbf{x}^T \mathbf{w}$. 若对 \mathbf{x} 作奇异值分解, 则主元分析法结果会有何异同?

第 5 章

广义逆矩阵

在线性代数中,我们就清楚地知道只有行列式不为零的方阵才可逆.可逆矩阵的存在让我们在解决诸多矩阵问题时可类似地引入“除法”运算(当然,矩阵并不存在除法运算).例如,可逆方阵在求解线性方程组时就非常简单,只需在方程组两端同时左乘系数矩阵的逆(可理解为方程组两端“除以”系数矩阵)即可得到方程组的解.然而,对于一般的 $m \times n$ 阶系数矩阵 A ,就难以利用可逆方阵的“除法”方便地求解线性方程组.本章引入的广义逆矩阵概念就是扩展了可逆矩阵的概念.

广义逆矩阵是矩阵理论的一个重要分支,已广泛地应用于控制理论、系统识别、优化理论和密码学等领域.本章主要介绍广义逆矩阵的定义,然后依次介绍四种典型的广义逆矩阵(减号逆、极小范数广义逆、最小二乘广义逆和加号逆)及其在线性方程组、正交投影等方面的应用.

5.1 基本概念

定义 5.1.1(广义逆矩阵) 设 $A \in \mathbb{C}^{m \times n}$, 若 $X \in \mathbb{C}^{n \times m}$ 满足以下 4 个 Penrose 方程的全部或者部分,则称矩阵 X 是矩阵 A 的广义逆矩阵.

- (1) $AXA = A$;
- (2) $XAX = X$;
- (3) $(AX)^H = AX$;
- (4) $(XA)^H = XA$.

注 1: 满足定义 5.1.1 中一个、两个、三个或四个 Penrose 方程的广义逆矩阵共计 15 种.若矩阵 G 是满足第 i 个 Penrose 方程的广义逆矩阵,则记为

$$G = A^{(i)}, \quad i = 1, 2, 3, 4$$

若矩阵 G 是满足第 i 和第 j 个 Penrose 方程的广义逆矩阵,则记为

$$G = A^{(i,j)}, \quad i, j = 1, 2, 3, 4 \text{ 且 } i \neq j$$

若矩阵 G 是满足第 i 、第 j 个和第 k 个 Penrose 方程的广义逆矩阵,则记为

$$G = A^{(i,j,k)}, \quad i, j, k = 1, 2, 3, 4 \text{ 且 } i, j, k \text{ 互不相等}$$

若矩阵 G 满足全部 4 个 Penrose 方程,则记为

$$G = A^{(1,2,3,4)} \text{ 或 } A^+$$

并将其称为加号逆或伪逆,或 Moore-Penrose 广义逆.我们将在后续章节中说明,矩阵 A 的加号逆唯一,其余各类广义逆矩阵一般不唯一.

每一广义逆矩阵都包含一类矩阵,分别用集合记为 $A\{i\}$ 、 $A\{i, j\}$ 、 $A\{i, j, k\}$ 表示.故有 $A^{(i)} \in A\{i\}$, $A^{(i,j)} \in A\{i, j\}$, $A^{(i,j,k)} \in A\{i, j, k\}$.

常见的广义逆有 $A^{(1)}$ 、 $A^{(1,3)}$ 、 $A^{(1,4)}$ 和 A^+ , 其中 $A^{(1)}$ 称为减号逆,记为 A^- ; $A^{(1,3)}$ 称为最小二乘广义逆,记为 A_l^- ; $A^{(1,4)}$ 为极小范数广义逆,记为 A_m^- .

5.2 矩阵方程 $AXB=D$

讨论广义逆矩阵面临的基础问题是其存在性问题.对此,本节主要讨论一般的矩阵方程 $AXB=D$ 的求解问题.

已知矩阵 $A \in \mathbb{C}_{r_A}^{m \times n}$, $B \in \mathbb{C}_{r_B}^{p \times q}$, $D \in \mathbb{C}_{r_D}^{m \times q}$, 待定矩阵 $X \in \mathbb{C}_{r_X}^{n \times p}$, 定义矩阵方程

$$AXB = D \quad (5.2.1)$$

定义 5.2.1(相容方程) 若存在矩阵 $X \in \mathbb{C}_{r_X}^{n \times p}$ 使矩阵方程(5.2.1)成立,则称方程(5.2.1)相容,否则称为不相容方程或矛盾方程.

下面采用三类不同方法讨论方程(5.2.1)的求解问题.

方法一: 利用矩阵分解求解.

由于矩阵 A 、 B 和 D 是适当阶数的任意矩阵,所以在利用矩阵分解求解方程(5.2.1)时,应优先考虑相抵分解、满秩分解或奇异值分解.注意到矩阵 A 、 B 和 D 并无关联,因此在利用矩阵分解求解一般的矩阵方程(5.2.1)比较困难.为讨论方便,这里仅考察 $r_A = r_B$ 这一情况.

假设 $r_A = r_B = r$, 并对矩阵 A 和 B 进行相抵分解得

$$P_A A Q_A = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}, \quad P_B B Q_B = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \quad (5.2.2)$$

式中: P_A 、 Q_A 、 P_B 和 Q_B 是适当阶数的可逆矩阵.

将式(5.2.2)代入式(5.2.1),得

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix} Q_A^{-1} X P_B^{-1} \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} = P_A D Q_B \quad (5.2.3)$$

定义 $Y = Q_A^{-1} X P_B^{-1}$ 和 $J = P_A D Q_B$, 并对其分块得

$$Y = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}, \quad J = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix}$$

式中: Y_{11} 和 $J_{11} \in \mathbb{C}^{r \times r}$.

由此, 式(5.2.3)化简为

$$\begin{bmatrix} Y_{11} & O \\ O & O \end{bmatrix} = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix}$$

因此, 我们有如下定理:

定理 5.2.1 考察矩阵方程(5.2.1), 如果存在非奇异矩阵 P_i 和 $Q_i (i=A, B)$ 满足式(5.2.2), 则方程(5.2.1)相容的充分必要条件是

$$P_A D Q_B = \begin{bmatrix} J_{11} & O \\ O & O \end{bmatrix}$$

此时, 方程(5.2.1)的解为

$$X = Q_A \begin{bmatrix} J_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} P_B$$

式中: $J_{11} \in \mathbb{C}^{r \times r}$; Y_{12} 、 Y_{21} 和 Y_{22} 是适当阶数的任意矩阵.

注 1: 若 $r_A = r_B = r$, 由定理 5.2.1 知, 方程(5.2.1)相容的必要条件为 $r_D \leq r$.

利用定理 5.2.1 可求解 Penrose 方程(1), 并有如下推论.

推论 5.2.1 考察矩阵方程 $AXA = A$, 存在非奇异矩阵 P 和 Q 使得 $PAQ =$

$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$, 则 G 是该方程解的充分必要条件是

$$G = Q \begin{bmatrix} I_r & K \\ L & M \end{bmatrix} P$$

式中: K 、 L 和 M 是适当阶数的任意矩阵. 当 A 为可逆矩阵时, G 唯一.

注 2: 推论 5.2.1 表明矩阵方程 $AXA = A$ 的解一定存在.

例 5.2.1 求解矩阵方程 $AXA = A$, 其中 $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$.

解:

$$\begin{bmatrix} A & I \\ I & * \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & * & * \\ 0 & 1 & 0 & * & * \\ 0 & 0 & 1 & * & * \end{bmatrix} \xrightarrow{\text{初等变换}} \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & -1 & * & * \\ 0 & 1 & -1 & * & * \\ 0 & 0 & 1 & * & * \end{bmatrix}$$

其中 $*$ 为任意复数.

由此, 解得

$$P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

故方程 $\mathbf{A}\mathbf{X}\mathbf{A}=\mathbf{A}$ 的解可表示为

$$\mathbf{G}=\mathbf{Q}\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ a & b \end{bmatrix}\mathbf{P}=\begin{bmatrix} 1-a & -b \\ -a & 1-b \\ a & b \end{bmatrix}$$

式中: a, b 为任意数.

注 3: 推论 5.2.1 中可逆矩阵 \mathbf{P} 和 \mathbf{Q} 可通过初等变换求解, 即

$$\begin{bmatrix} \mathbf{P} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{I} \\ \mathbf{I} & * \end{bmatrix} \begin{bmatrix} \mathbf{Q} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{PAQ} & \mathbf{P} \\ \mathbf{Q} & * \end{bmatrix}$$

同样地, 利用奇异值分解也可以方便地求解 $r_{\mathbf{A}}=r_{\mathbf{B}}$ 时的矩阵方程 (5.2.1), 并有如下结论:

定理 5.2.2 考察矩阵方程 (5.2.1), 若矩阵 \mathbf{A} 和 \mathbf{B} 有奇异值分解

$$\mathbf{A}=\mathbf{U}_{\mathbf{A}}\begin{bmatrix} \boldsymbol{\Sigma}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}\mathbf{V}_{\mathbf{A}}^{\mathrm{H}}, \quad \mathbf{B}=\mathbf{U}_{\mathbf{B}}\begin{bmatrix} \boldsymbol{\Lambda}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}\mathbf{V}_{\mathbf{B}}^{\mathrm{H}}$$

则方程 (5.2.1) 相容的充分必要条件是

$$\mathbf{U}_{\mathbf{A}}^{\mathrm{H}}\mathbf{D}\mathbf{V}_{\mathbf{B}}=\begin{bmatrix} \mathbf{J}_{11} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}$$

此时, 方程 (5.2.1) 的解为

$$\mathbf{X}=\mathbf{V}_{\mathbf{A}}\begin{bmatrix} \boldsymbol{\Sigma}_r^{-1}\mathbf{J}_{11}\boldsymbol{\Lambda}_r^{-1} & \mathbf{Y}_{12} \\ \mathbf{Y}_{21} & \mathbf{Y}_{22} \end{bmatrix}\mathbf{U}_{\mathbf{B}}^{\mathrm{H}}$$

式中: $\boldsymbol{\Sigma}_r$ 和 $\boldsymbol{\Lambda}_r$ 均为 r 阶对角矩阵, 其对角线元素分别为矩阵 \mathbf{A} 和 \mathbf{B} 的正奇异值; $\mathbf{J}_{11} \in \mathbb{C}^{r \times r}$; $\mathbf{Y}_{12}, \mathbf{Y}_{21}$ 和 \mathbf{Y}_{22} 是适当阶数的任意矩阵.

推论 5.2.2 考察矩阵方程 $\mathbf{A}\mathbf{X}\mathbf{A}=\mathbf{A}$, 若矩阵 \mathbf{A} 有奇异值分解 $\mathbf{A}=\mathbf{U}\begin{bmatrix} \boldsymbol{\Sigma}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}\mathbf{V}^{\mathrm{H}}$, 则矩阵 \mathbf{G} 是该方程解的充分必要条件是

$$\mathbf{G}=\mathbf{V}\begin{bmatrix} \boldsymbol{\Sigma}_r^{-1} & \mathbf{K} \\ \mathbf{L} & \mathbf{M} \end{bmatrix}\mathbf{U}^{\mathrm{H}}$$

式中: $\boldsymbol{\Sigma}_r$ 是 r 阶对角矩阵, 其对角线元素是矩阵 \mathbf{A} 的 r 个正奇异值; $\mathbf{K}, \mathbf{L}, \mathbf{M}$ 是适当阶数的任意矩阵. 当 \mathbf{A} 为可逆矩阵时, \mathbf{G} 唯一.

具体证明留作课后习题, 这里不再赘述了.

再考察例 5.2.1. 将矩阵 \mathbf{A} 进行奇异值分解, 得

$$\mathbf{A}=\begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3} & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} \\ 0 & \frac{\sqrt{6}}{3} & -\frac{\sqrt{3}}{3} \end{bmatrix}^{\mathrm{H}}$$

则矩阵 A 的减号逆可写为

$$\begin{aligned}
 G &= \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} \\ 0 & \frac{\sqrt{6}}{3} & -\frac{\sqrt{3}}{3} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{2}(a-b) & \frac{\sqrt{6}}{2}(a+b) \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}^H \\
 &= \begin{bmatrix} -\frac{2}{3}+a & -\frac{1}{3}+b \\ -\frac{1}{3}+a & \frac{2}{3}+b \\ \frac{1}{3}-a & \frac{1}{3}-b \end{bmatrix}
 \end{aligned}$$

式中: a, b 为任意数.

方法二: 利用 Penrose 定理求解.

定理 5.2.3 (Penrose 定理) 矩阵方程 (5.2.1) 相容的充分必要条件为 $AA^{(1)}DB^{(1)}B=D$, 其中 $A^{(1)} \in A\{1\}$, $B^{(1)} \in B\{1\}$; 此时方程 (5.2.1) 的通解为

$$X = A^{(1)}DB^{(1)} + Y - A^{(1)}AYBB^{(1)}$$

式中: Y 为任一 $n \times p$ 矩阵.

证明: 若 $AA^{(1)}DB^{(1)}B=D$ 成立, 则 $X=A^{(1)}DB^{(1)}$ 是方程 (5.2.1) 的一个解, 故矩阵方程相容 (相容充分性). 若方程 (5.2.1) 相容, 不妨设 X 是其一个解. 由推论 5.2.1 知, 任一矩阵的减号逆必存在, 故存在 $A^{(1)}$ 和 $B^{(1)}$ 满足

$$AA^{(1)}A=A, \quad BB^{(1)}B=B \quad (5.2.4)$$

将式 (5.2.4) 代入方程 (5.2.1), 得

$$AA^{(1)}AXB^{(1)}B=D \quad (5.2.5)$$

再将方程 (5.2.1) 代入式 (5.2.5) 得, $AA^{(1)}DB^{(1)}B=D$. 必要性得证.

通解的充分性. 将解 $X=A^{(1)}DB^{(1)}+Y-A^{(1)}AYBB^{(1)}$ 代入方程 (5.2.1) 左端得

$$\begin{aligned}
 AXB &= AA^{(1)}DB^{(1)}B + AYB - AA^{(1)}AYBB^{(1)}B \\
 &= D + AYB - AYB = D
 \end{aligned}$$

故 $X=A^{(1)}DB^{(1)}+Y-A^{(1)}AYBB^{(1)}$ 是方程 (5.2.1) 的解.

通解的必要性. 设 X 是方程 (5.2.1) 的任一解, 则

$$X = A^{(1)}DB^{(1)} + X - A^{(1)}DB^{(1)} \quad (5.2.6)$$

将方程 (5.2.1) 代入式 (5.2.6) 得

$$X = A^{(1)}DB^{(1)} + X - A^{(1)}AXB^{(1)} \quad (5.2.7)$$

显然, 方程 (5.2.1) 的任一解均可写成定理 5.2.3 中的通解形式. 证毕.

【应用】 利用 Penrose 定理可求解非齐次线性方程组和矩阵的减号逆.

推论 5.2.3 非齐次线性方程组 $Ax=b$ 相容的充分必要条件是 $AA^{(1)}b=b$, 其通解为 $x=A^{(1)}b+(I-A^{(1)}A)y, \forall y \in \mathbb{C}^n$.

推论 5.2.4 齐次线性方程组 $Ax=0$ 的通解为 $x=(I-A^{(1)}A)y, \forall y \in \mathbb{C}^n$.

推论 5.2.5 设 $A \in \mathbb{C}^{m \times n}$, 则 $A\{1\}=A^{(1)}+Z-A^{(1)}AZAA^{(1)}$, 其中, Z 为任一 $n \times m$ 阶矩阵.

证明: 令式(5.2.1)中 $B=D=A$, 并由 Penrose 定理得 X 的通解为

$$X=A^{(1)}AA^{(1)}+Y-A^{(1)}AYAA^{(1)}, \quad \forall Y \in \mathbb{C}^{n \times m}$$

对任意 $Z \in \mathbb{C}^{n \times m}$, 令 $Y=A^{(1)}+Z$ 并代入上式得

$$A\{1\}=A^{(1)}+Z-A^{(1)}AZAA^{(1)}$$

特别注意的是, 结论中出现的 $A^{(1)}$ 是指矩阵 A 的任一确定减号逆, 该减号逆可通过推论 5.2.1 和推论 5.2.2 求解. 尽管选定的减号逆可能不同, 但并不影响结论.

例 5.2.2 求线性方程组 $\begin{cases} x_1+x_3=1 \\ x_2+x_3=2 \end{cases}$ 的通解.

解: 令 $A=\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, b=\begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

方法一: 由例 5.2.1 计算知,

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

又知

$$AA^{(1)}b=\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix}=\begin{bmatrix} 1 \\ 2 \end{bmatrix}=b$$

故方程组相容. 此时, 该方程组的通解为

$$x=\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}+\begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}=\begin{bmatrix} 1-c_3 \\ 2-c_3 \\ c_3 \end{bmatrix}$$

式中: c_1, c_2, c_3 为任意数.

方法二: 根据推论 5.2.2 计算出 A 的一个减号逆为

$$A^{(1)}=\begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

则有

$$\mathbf{A}\mathbf{A}^{(1)}\mathbf{b} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \mathbf{b}$$

故方程组相容. 此时, 该方程组的通解为

$$\tilde{\mathbf{x}} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{3}(k_1 + k_2 - k_3) \\ 1 + \frac{1}{3}(k_1 + k_2 - k_3) \\ 1 - \frac{1}{3}(k_1 + k_2 - k_3) \end{bmatrix}$$

式中: k_1, k_2, k_3 为任意数. 注意比较上述两种方法得到的通解是等价的.

方法三: 利用直积和矩阵拉直求解.

定义 5.2.2 (Kronecker 积) 设 $\mathbf{A} = (a_{ij}) \in \mathbb{C}^{m \times n}$, $\mathbf{B} = (b_{ij}) \in \mathbb{C}^{p \times q}$, 称如下分块矩阵:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix} \in \mathbb{C}^{mp \times nq}$$

为 \mathbf{A} 与 \mathbf{B} 的 Kronecker 积(或直积、张量积), 简记为 $\mathbf{A} \otimes \mathbf{B} = (a_{ij}\mathbf{B})$.

例 5.2.3 计算 $\mathbf{A} \otimes \mathbf{B}$ 和 $\mathbf{B} \otimes \mathbf{A}$, 其中

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

$$\text{解: } \mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} -1 & 1 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \quad \mathbf{B} \otimes \mathbf{A} = \begin{bmatrix} -1 & 1 & 1 & -1 \\ 0 & -1 & 0 & 1 \end{bmatrix}.$$

由上例可以看出, 尽管 $\mathbf{A} \otimes \mathbf{B}$ 与 $\mathbf{B} \otimes \mathbf{A}$ 是同阶矩阵, 但一般来说 $\mathbf{A} \otimes \mathbf{B} \neq \mathbf{B} \otimes \mathbf{A}$, 即 Kronecker 积不满足交换律.

命题 5.2.1 (直积的性质) 矩阵的 Kronecker 积具有以下性质:

- (1) 对任意复数 k , $(k\mathbf{A}) \otimes \mathbf{B} = \mathbf{A} \otimes (k\mathbf{B}) = k(\mathbf{A} \otimes \mathbf{B})$;
- (2) $(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C})$;
- (3) $\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{C}$;
- (4) $(\mathbf{A} \otimes \mathbf{B})^H = \mathbf{A}^H \otimes \mathbf{B}^H$;
- (5) 若矩阵 \mathbf{A} 和 \mathbf{C} , 矩阵 \mathbf{B} 和 \mathbf{D} 均可相乘, 则 $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$;
- (6) $\text{rank}(\mathbf{A} \otimes \mathbf{B}) = \text{rank}(\mathbf{A})\text{rank}(\mathbf{B})$;
- (7) $(\mathbf{A} \otimes \mathbf{B})^+ = \mathbf{A}^+ \otimes \mathbf{B}^+$.

定理 5.2.4 设 $f(x, y) = \sum_{i,j=0}^k c_{ij} x^i y^j$ 是变量 x, y 的复二元多项式, 对任意矩阵 $A \in \mathbb{C}^{m \times m}$ 和 $B \in \mathbb{C}^{n \times n}$, 定义矩阵

$$f(A, B) = \sum_{i,j=0}^k c_{ij} (A^i \otimes B^j) \in \mathbb{C}^{mn \times mn}$$

式中: $A^0 = I_m, B^0 = I_n$. 若 A 和 B 的特征值分别为 $\lambda_1, \dots, \lambda_m$ 和 μ_1, \dots, μ_n , 则 $f(A, B)$ 的特征值为 $f(\lambda_i, \mu_j), i=1, \dots, m, j=1, \dots, n$.

证明: 由 Schur 引理知, 存在酉矩阵 P 和 Q 使得 $P^H A P = A_1$ 和 $Q^H B Q = B_1$, 其中, A_1 和 B_1 分别是对角元素为 $\lambda_1, \dots, \lambda_m$ 和 μ_1, \dots, μ_n 的上三角矩阵. 于是,

$$\begin{aligned} (P \otimes Q)^H f(A, B) (P \otimes Q) &= \sum_{i,j=0}^k c_{ij} (P \otimes Q)^H (A^i \otimes B^j) (P \otimes Q) \\ &= \sum_{i,j=0}^k c_{ij} (P^H A^i P) \otimes (Q^H B^j Q) \\ &= \sum_{i,j=0}^k c_{ij} (A_1^i \otimes B_1^j) = f(A_1, B_1) \end{aligned}$$

式中: $f(A_1, B_1)$ 是上三角矩阵, 且与 $f(A, B)$ 有相同的特征值. 又因为

$$\begin{aligned} A_1^i \otimes B_1^j &= \begin{bmatrix} \lambda_1^i B_1^j & * & * \\ & \ddots & * \\ O & & \lambda_m^i B_1^j \end{bmatrix} \\ \lambda_p^i B_1^j &= \begin{bmatrix} \lambda_p^i \mu_1^j & * & * \\ & \ddots & * \\ O & & \lambda_p^i \mu_n^j \end{bmatrix} \end{aligned}$$

其中, $p=1, \dots, m$.

所以, $f(A, B)$ 的特征值为 $f(\lambda_i, \mu_j), i=1, \dots, m, j=1, \dots, n$. 证毕.

由定理 5.2.4 可得以下常用结论:

推论 5.2.6 设 $A \in \mathbb{C}^{m \times m}, B \in \mathbb{C}^{n \times n}$, 其特征值分别为 $\lambda_1, \dots, \lambda_m$ 和 μ_1, \dots, μ_n , 则有

- (1) $A \otimes B$ 的特征值为 $\lambda_i \mu_j, i=1, \dots, m, j=1, \dots, n$.
- (2) $A \otimes I_n + I_m \otimes B$ 的特征值为 $\lambda_i + \mu_j, i=1, \dots, m, j=1, \dots, n$.

定义 5.2.3 (列拉直) 设 $A = (a_{ij}) \in \mathbb{C}^{m \times n}$, 并记 $a_i = [a_{1i}, a_{2i}, \dots, a_{mi}]^T, i=1, 2, \dots, n$. 令

$$\text{vec}(A) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

则称 $\text{vec}(\mathbf{A})$ 为矩阵 \mathbf{A} 的列拉直(或列展开).

类似地,可定义矩阵的行拉直(行展开).

定理 5.2.5 设 $\mathbf{A} \in \mathbb{C}^{m \times n}$, $\mathbf{B} \in \mathbb{C}^{n \times p}$ 和 $\mathbf{C} \in \mathbb{C}^{p \times q}$, 则

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B})$$

证明: 记 $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_p]$, $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_q]$, 则

$$\text{vec}(\mathbf{ABC}) = \text{vec}(\mathbf{ABc}_1, \mathbf{ABc}_2, \dots, \mathbf{ABc}_q) = \begin{bmatrix} \mathbf{ABc}_1 \\ \mathbf{ABc}_2 \\ \vdots \\ \mathbf{ABc}_q \end{bmatrix}$$

而 $\mathbf{ABc}_i = c_{1i}\mathbf{Ab}_1 + c_{2i}\mathbf{Ab}_2 + \dots + c_{pi}\mathbf{Ab}_p = [c_{1i}\mathbf{A}, c_{2i}\mathbf{A}, \dots, c_{pi}\mathbf{A}] \text{vec}(\mathbf{B})$, 故

$$\begin{aligned} \text{vec}(\mathbf{ABC}) &= \begin{bmatrix} c_{11}\mathbf{A} & c_{21}\mathbf{A} & \cdots & c_{p1}\mathbf{A} \\ c_{12}\mathbf{A} & c_{22}\mathbf{A} & \cdots & c_{p2}\mathbf{A} \\ \vdots & \vdots & & \vdots \\ c_{1q}\mathbf{A} & c_{2q}\mathbf{A} & \cdots & c_{pq}\mathbf{A} \end{bmatrix} \text{vec}(\mathbf{B}) \\ &= (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B}) \end{aligned}$$

根据上述定义和结论,有如下定理:

定理 5.2.6 矩阵方程(5.2.1)相容的充分必要条件为

$$\text{rank}(\mathbf{B}^T \otimes \mathbf{A}, \text{vec}(\mathbf{D})) = \text{rank}(\mathbf{B}^T \otimes \mathbf{A})$$

证明: 对矩阵方程(5.2.1)等式两端拉直,并利用定理 5.2.5,有

$$(\mathbf{B}^T \otimes \mathbf{A}) \text{vec}(\mathbf{X}) = \text{vec}(\mathbf{D}) \quad (5.2.8)$$

定义 $\mathbf{x} = \text{vec}(\mathbf{X})$, $\mathbf{G} = (\mathbf{B}^T \otimes \mathbf{A})$, 则式(5.2.8)可简写为 $\mathbf{Gx} = \text{vec}(\mathbf{D})$. 显然,该方程组相容的充分必要条件为 $\text{rank}(\mathbf{G}, \text{vec}(\mathbf{D})) = \text{rank}(\mathbf{G})$. 证毕.

注 4: 定理 5.2.3 和定理 5.2.6 分别给出了矩阵方程(5.2.1)相容的充分必要条件. 显然,这两个条件应该是互为等价的. 为讨论方便,我们利用 \mathbf{A}^+ 进行讨论(关于加号逆的具体性质详见 5.6 节). 选取 \mathbf{A}^+ 作为 \mathbf{A} 的一个已知减号逆,由推论 5.2.3 知,线性方程组 $\mathbf{Ax} = \mathbf{b}$ 相容的充分必要条件为 $\mathbf{AA}^+\mathbf{b} = \mathbf{b}$, 其通解为

$$\mathbf{x} = \mathbf{A}^+\mathbf{b} + (\mathbf{I} - \mathbf{A}^+\mathbf{A})\mathbf{y}, \quad \forall \mathbf{y} \in \mathbb{C}^n$$

由此,线性方程组(5.2.8)相容的充分必要条件为

$$(\mathbf{B}^T \otimes \mathbf{A})(\mathbf{B}^T \otimes \mathbf{A})^+ \text{vec}(\mathbf{D}) = \text{vec}(\mathbf{D})$$

由命题 5.2.1 得

$$\text{vec}(\mathbf{D}) = ((\mathbf{B}^+ \mathbf{B})^T \otimes (\mathbf{AA}^+)) \text{vec}(\mathbf{D}) = \text{vec}(\mathbf{AA}^+ \mathbf{DB}^+ \mathbf{B})$$

即 $\mathbf{AA}^+ \mathbf{DB}^+ \mathbf{B} = \mathbf{D}$. 该条件与定理 5.2.3 相容的充分必要条件一致.

由线性方程组 $\mathbf{Ax} = \mathbf{b}$ 通解得到方程(5.2.8)的通解为

$$\begin{aligned} \text{vec}(\mathbf{X}) &= (\mathbf{B}^T \otimes \mathbf{A})^+ \text{vec}(\mathbf{D}) + (\mathbf{I} - (\mathbf{B}^T \otimes \mathbf{A})^+ (\mathbf{B}^T \otimes \mathbf{A})) \text{vec}(\mathbf{Y}) \\ &= ((\mathbf{B}^+)^T \otimes \mathbf{A}^+) \text{vec}(\mathbf{D}) + \text{vec}(\mathbf{Y}) - ((\mathbf{B}^+ \mathbf{B})^T \otimes (\mathbf{AA}^+)) \text{vec}(\mathbf{Y}) \end{aligned}$$

$$= \text{vec}(\mathbf{A}^+ \mathbf{D} \mathbf{B}^+) + \text{vec}(\mathbf{Y}) - \text{vec}(\mathbf{A} \mathbf{A}^+ \mathbf{D} \mathbf{B}^+ \mathbf{B})$$

因此,方程(5.2.8)的通解式为 $\mathbf{X} = \mathbf{A}^+ \mathbf{D} \mathbf{B}^+ + \mathbf{Y} - \mathbf{A} \mathbf{A}^+ \mathbf{D} \mathbf{B}^+ \mathbf{B}$, 该条件与定理 5.2.3 一致.

【应用】 Lyapunov 矩阵方程定义为

$$\mathbf{A} \mathbf{X} + \mathbf{X} \mathbf{A}^H = -\mathbf{Q} \quad (5.2.9)$$

方程(5.2.9)在控制理论、通信和动力系统中起着非常重要的作用. 我们常根据 Lyapunov 矩阵方程的解来检测系统的稳定性、可控性和可观测性等问题.

定理 5.2.7 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$, $\mathbf{Q} \in \mathbb{C}^{n \times n}$, 若矩阵 \mathbf{A} 的所有特征值均具有负实部, 则矩阵方程(5.2.9)有唯一解, 且解 $\mathbf{X} \in \mathbb{C}^{n \times n}$ 可表示为

$$\mathbf{X} = \int_0^{+\infty} e^{\mathbf{A}^H t} \mathbf{Q} e^{\mathbf{A} t} dt \quad (5.2.10)$$

进一步, 若 \mathbf{Q} 是(半正定、正定) Hermite 矩阵, 则解 \mathbf{X} 也是(半正定、正定) Hermite 矩阵.

证明: 对矩阵方程(5.2.9)两端进行列展开, 得

$$(\mathbf{I}_n \otimes \mathbf{A} + \mathbf{A}^H \otimes \mathbf{I}_n) \text{vec}(\mathbf{X}) = -\text{vec}(\mathbf{Q}) \quad (5.2.11)$$

由于矩阵 $(\mathbf{I}_n \otimes \mathbf{A} + \mathbf{A}^H \otimes \mathbf{I}_n)$ 的特征值为矩阵 \mathbf{A} 的任意两个特征值之和(推论 5.2.6), 故 $(\mathbf{I}_n \otimes \mathbf{A} + \mathbf{A}^H \otimes \mathbf{I}_n)$ 的特征值均具有负实部. 相应地, 矩阵 $(\mathbf{I}_n \otimes \mathbf{A} + \mathbf{A}^H \otimes \mathbf{I}_n)$ 必可逆, 方程(5.2.11)有唯一解.

又知

$$\frac{d}{dt} (e^{\mathbf{A}^H t} \mathbf{Q} e^{\mathbf{A} t}) = \mathbf{A} e^{\mathbf{A}^H t} \mathbf{Q} e^{\mathbf{A} t} + e^{\mathbf{A}^H t} \mathbf{Q} e^{\mathbf{A} t} \mathbf{A}^H$$

两端对 t 从 0 到 ∞ 积分, 得

$$(e^{\mathbf{A}^H t} \mathbf{Q} e^{\mathbf{A} t}) \Big|_0^{+\infty} = \mathbf{A} \int_0^{+\infty} e^{\mathbf{A}^H t} \mathbf{Q} e^{\mathbf{A} t} dt + \int_0^{+\infty} e^{\mathbf{A}^H t} \mathbf{Q} e^{\mathbf{A} t} \mathbf{A}^H dt$$

经整理可得到矩阵方程(5.2.9), 这表明式(5.2.10)是矩阵方程(5.2.9)的解.

若 \mathbf{Q} 是 Hermite 矩阵, 则由 $(e^{\mathbf{A}^H})^H = e^{\mathbf{A} t}$ 及式(5.2.10)知, 解 \mathbf{X} 也是 Hermite 矩阵. 若 \mathbf{Q} 是(非负定、正定) Hermite 矩阵且 $e^{\mathbf{A}^H t}$ 是可逆矩阵, 则 $e^{\mathbf{A}^H t} \mathbf{Q} e^{\mathbf{A} t}$ 是(非负定、正定) Hermite 矩阵. 因此, 解 \mathbf{X} 也是(半正定、正定) Hermite 矩阵. 证毕.

实际上, 利用直积和矩阵拉直方法也可讨论诸如 $\mathbf{A} \mathbf{X} + \mathbf{X} \mathbf{B} = \mathbf{C}$ 和 $\mathbf{X} - \mathbf{A} \mathbf{X} \mathbf{B} = \mathbf{C}$ 等矩阵方程的求解问题. 有兴趣的读者可自行求解, 这里就不再赘述.

5.3 减号逆

推论 5.2.1 和推论 5.2.2 给出了矩阵减号逆的存在性结论, 并指出任一矩阵的减号逆必存在. 推论 5.2.3 也提供了矩阵减号逆的通解. 此外, 还可以利用满秩分解给出减号逆的一个特解. 特别地, 若 $\mathbf{A} \in \mathbb{C}_n^{m \times n}$, 则 \mathbf{A} 的左逆 $(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \in \mathbf{A} \{1\}$; 若

$A \in \mathbb{C}_m^{\times n}$, 则 A 的右逆 $A^H(AA^H)^{-1} \in A\{1\}$.

例 5.3.1 求矩阵 $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ 的一个减号逆.

解: 由于 A 是行满秩矩阵, 则 $A^H(AA^H)^{-1} \in A\{1\}$.

$$A^H(AA^H)^{-1} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

尽管利用奇异值分解或满秩分解(如例 5.3.1)求解矩阵的减号逆表达比较简洁, 但利用相抵分解(如例 5.2.1)求解的计算量相对较小.

命题 5.3.1(减号逆的性质)

- (1) $\text{rank}(A) \leq \text{rank}(A^-)$;
- (2) AA^- 和 A^-A 均为幂等矩阵, 且 $\text{rank}(AA^-) = \text{rank}(A^-A) = \text{rank}(A)$;
- (3) 若 $B = P^{-1}AQ^{-1}$, 则 $QA^-P \in B\{1\}$.

【应用】 考察非齐次线性方程组

$$Ax = b \quad (5.3.1)$$

式中, $A \in \mathbb{C}^{m \times n}$ 和 $b \in \mathbb{C}^m$ 给定, $x \in \mathbb{C}^n$ 为待定向量.

首先讨论 $A^{(1)}$ 与方程组 $Ax = b$ 相容性的关系.

命题 5.3.2(线性方程组相容条件) 设 $A \in \mathbb{C}^{m \times n}$, $b \in \mathbb{C}^m$, 则以下表达等价:

- (1) $\text{rank}(A) = \text{rank}(A, b)$;
- (2) $b \in R(A)$;
- (3) $AA^{(1)}b = b$.

命题 5.3.2 中的 3 个命题均是方程(5.3.1)相容的充分必要条件, 其中, 条件(1)是众所周知的方程相容条件, 该条件亦可用条件(2)表示; 条件(3)是推论 5.2.1 给出的充分必要条件. 下面简单说明条件(3)与条件(2)等价. 由命题 5.3.1 知, $AA^{(1)}$ 是幂等矩阵. 根据定理 3.7.4 知, $b \in R(A)$ 当且仅当 $AA^{(1)}b = b$. 即条件(2)和式(3)等价.

定理 5.3.1(减号逆与方程解的关系) 设 $A \in \mathbb{C}^{m \times n}$, 则 $G \in A\{1\}$ 的充分必要条件为 $x = Gb$ 是相容方程组 $Ax = b$ 的解.

证明: 必要性. 若 $G \in A\{1\}$, 则方程组 $Ax = b$ 相容的充分必要条件为 $AGb = b$. 这表明 $x = Gb$ 必为相容方程组 $Ax = b$ 的解.

充分性. 由于方程组 $Ax = b$ 相容的, 故 $b \in R(A)$. 换言之, 必存在向量 $y \in \mathbb{C}^n$ 使得 $b = Ay$. 已知条件表明, 对任意向量 $b \in \mathbb{C}^m$, $x = Gb$ 是相容方程组 $Ax = b$ 的解. 因此, 对任意向量 $y \in \mathbb{C}^n$, $x = GAy$ 是相容方程组 $Ax = b$ 的解. 将其代入相容方程组 $Ax = b$, 得

$$AGA y = A y \quad (5.3.2)$$

注意: 式(5.3.2)对 \mathbb{C}^n 中任意向量 y 都成立. 因此, $AGA = A$, 即 $G \in A\{1\}$.

【思考】 定理 5.3.1 表明了具有 $x = Gb$ 形式的解与减号逆的关系. 那么, 相容方程组(5.3.1)的解是否一定可以表示为 $x = Gb$ 这种形式呢?

分析: 由推论 5.2.3 知, 相容方程组(5.3.1)的通解式为

$$x = A^{(1)}b + (I - A^{(1)}A)y, \quad \forall y \in \mathbb{C}^n \quad (5.3.3)$$

从式(5.3.3)看, 相容方程组(5.3.1)的解似乎并不一定可表示为 $x = Gb$ 这种形式, 因为式(5.3.3)还包含了 $(I - A^{(1)}A)y$. 对此, 定义 $G = A^{(1)} + Z - A^{(1)}AZAA^{(1)}$ 和 $x = Gb$, 则有

$$\begin{aligned} x &= A^{(1)}b + Zb - A^{(1)}AZAA^{(1)}b \\ &= A^{(1)}b + (I - A^{(1)}A)Zb \end{aligned} \quad (5.3.4)$$

其中, Z 为任一 $n \times m$ 矩阵.

显然, 若对任意给定向量 y 和 b , 方程 $Zb = y$ 都能找到解矩阵 Z . 因此, 式(5.3.4)等价于相容方程(5.3.1)的通解式(5.3.3); 此时相容方程(5.3.1)的解一定可以表示为 $x = Gb$ 的形式, 其中 $G \in A\{1\}$.

根据定理 5.2.3 知, 方程 $Zb = y$ 相容当且仅当 $y b^{(1)}b = y$. 又知 $b^{(1)}b = 1$ 当且仅当 $b \neq 0$. 故方程 $Zb = y$ 相容的充分必要条件为 $b \neq 0$, 并有如下推论:

推论 5.3.1 设 $A \in \mathbb{C}^{m \times n}$, 非零向量 $b \in \mathbb{C}^m$, 则相容方程组 $Ax = b$ 的解一定可以用 $x = Gb$ 表示, 其中 $G \in A\{1\}$.

例 5.3.2 考察 $x = A^{(1)}b$ 与相容方程组 $Ax = b$ 解的关系, 其中

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

解: 由于 $b \in R(A)$, 故方程组 $Ax = b$ 相容. 易求得该方程组的解为 $x = [1, 0, c]^T, \forall c \in \mathbb{C}$.

另由减号逆定义求得

$$A^{(1)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ a & d \end{bmatrix}, \quad \forall a, d \in \mathbb{C}$$

则有 $x = A^{(1)}b = [1, 0, a]^T$. 因此, 相容方程组 $Ax = b$ 的解一定可以表示为 $x = A^{(1)}b$.

若 $b = 0$, 则不论 $A^{(1)}$ 取何矩阵, 方程解为 $x = A^{(1)}b = 0$; 而齐次方程组的解为 $x = [0, 0, c]^T, \forall c \in \mathbb{C}$. 此时, 相容方程组 $Ax = b$ 不能用 $x = A^{(1)}b$ 表示.

定理 5.3.2 (相容线性方程组唯一解条件) 设 $A \in \mathbb{C}^{m \times n}, A^{(1)}A = I_n$ 当且仅当矩阵 A 是列满秩的. 此时, 相容方程组 $Ax = b$ 有唯一解.

证明: 充分性. $A^{(1)}A$ 是幂等矩阵, 其特征值只能是 0 或 1. 由 A 是列满秩矩阵知, $A^{(1)}A$ 是可逆矩阵. 故 $A^{(1)}A$ 无零特征值, 所有特征值均为 1. 由此, $A^{(1)}A$ 相似于

单位矩阵, 即 $A^{(1)}A = I_n$.

必要性. 由 $A^{(1)}A = I_n$ 知, 矩阵 A 有左逆 $A^{(1)}$. 根据定理 3.2.3, A 是列满秩矩阵. 证毕.

注 1: 当矩阵 A 是列满秩矩阵, 此时它的任一减号逆 $A^{(1)}$ 都是 A 的左逆. 同理, 当矩阵 A 是行满秩矩阵, 则它的任一减号逆 $A^{(1)}$ 都是 A 的右逆.

例 5.3.3 求向量 $b \in \mathbb{C}^m$ 在 $R(A)$ 上的正交投影, 其中 $A \in \mathbb{C}^{m \times n}$.

分析: 回顾例 3.7.2, 若 A 为列满秩矩阵, 则向量 b 在 $R(A)$ 上的正交投影为

$$\text{Proj}_{R(A)} b = A(A^H A)^{-1} A^H b$$

而本例并未限定 A 为列满秩矩阵, 故 $(A^H A)^{-1}$ 不一定存在, 进而无法定义正交投影矩阵为 $A(A^H A)^{-1} A^H$. 现将正交投影矩阵中的求逆运算改为减号逆, 并定义

$$P = A^H (AA^H)^- A \quad (5.3.5)$$

显然, 若式 (5.3.5) 中的 P 是正交投影矩阵, 则 Pb 必是向量 b 在 $R(A)$ 空间上的正交投影.

定理 5.3.3 对任意矩阵 $A \in \mathbb{C}^{m \times n}$, $A(A^H A)^- A^H$ 是正交投影矩阵.

证明: 设矩阵 A 有奇异值分解: $A = U \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix} V^H$, 则 $A^H A = V \begin{bmatrix} \Sigma_r^2 & O \\ O & O \end{bmatrix} V^H$, 其

中, U, V 分别为 m 阶和 n 阶的正交矩阵, Σ_r 是 r 阶正定对角矩阵.

再由推论 5.2.2 知, $A^H A$ 的任一减号逆矩阵必可以表示为

$$G = V \begin{bmatrix} (\Sigma_r^2)^{-1} & K \\ L & M \end{bmatrix} V^H$$

式中: K, L 和 M 是适当阶数的任意矩阵.

由此,

$$AGA^H = U \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} U^H$$

即 AGA^H 是 Hermite 幂等矩阵; 换言之, $A(A^H A)^- A^H$ 是正交投影矩阵.

注 2: 对一般的矩阵 A , $(AA^H)^- \neq (A^H)^- A^-$.

推论 5.3.2 对任意矩阵 $A \in \mathbb{C}^{m \times n}$, $A(A^H A)^- A^H A = A$.

证明: 将矩阵 A 进行列分块, 并记为 $A = [a_1, \dots, a_n]$, 则 $a_i \in R(A)$, $i = 1, \dots, m$. 由 $A(A^H A)^- A^H$ 是幂等矩阵知, $A(A^H A)^- A^H a_i = a_i$. 因此,

$$A(A^H A)^- A^H [a_1, \dots, a_n] = [a_1, \dots, a_n]$$

即 $A(A^H A)^- A^H A = A$.

例 5.3.4 (最小二乘问题) 考察线性方程组 (5.3.1), 其中, 矩阵 $A \in \mathbb{C}^{m \times n}$ 和向量 $b \in \mathbb{C}^m$ 给定, 向量 $x \in \mathbb{C}^n$ 待定. 求向量 x 使得 $\|Ax - b\|_2$ 最小.

解: 由例 3.7.4 知, 求向量 x 使得 $\|Ax - b\|_2$ 最小可等价转化为求解方程组

$$Ax = \text{Proj}_{R(A)} b \quad (5.3.6)$$

式中, \mathbf{b} 在 $R(\mathbf{A})$ 的正交投影为 $\text{Proj}_{R(\mathbf{A})} \mathbf{b} = \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{b}$.

因此, 式(5.3.6)可等价写为

$$\mathbf{A}\mathbf{x} = \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{b} \quad (5.3.7)$$

显然, $\mathbf{x} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{b}$ 是方程(5.3.7)的一个解, 故该方程是相容的(也可以利用相容充分必要条件验证). 其通解为

$$\mathbf{x} = \mathbf{A}^{-} \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{b} + (\mathbf{I} - \mathbf{A}^{-} \mathbf{A}) \mathbf{y}, \quad \forall \mathbf{y} \in \mathbb{C}^n \quad (5.3.8)$$

回顾例 3.7.4 得到的使 $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$ 最小的等价方程组为

$$\mathbf{A}^H \mathbf{A} \mathbf{x} = \mathbf{A}^H \mathbf{b} \quad (5.3.9)$$

实际上, 方程组(5.3.7)、(5.3.9)均相容且为同解方程组. 这是一个非常有趣的结果. 尽管线性方程组 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 不一定相容, 但方程两端同时左乘 \mathbf{A}^H , 得新方程(5.3.9)必相容, 且该方程的任一解 \mathbf{x} 左乘以矩阵 \mathbf{A} 得到的向量恰为向量 \mathbf{b} 在 $R(\mathbf{A})$ 空间的正交投影.

5.4 极小范数广义逆

$\mathbf{A}^{(1,4)}$ 是矩阵 \mathbf{A} 的极小范数广义逆, 其存在性有如下结论:

定理 5.4.1 设矩阵 $\mathbf{A} \in \mathbb{C}^{m \times n}$ 的奇异值分解为

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \boldsymbol{\Sigma}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{V}^H$$

则 $\mathbf{G} \in \mathbf{A}\{1,4\}$ 的充分必要条件是

$$\mathbf{G} = \mathbf{V} \begin{bmatrix} \boldsymbol{\Sigma}_r^{-1} & \mathbf{K} \\ \mathbf{O} & \mathbf{M} \end{bmatrix} \mathbf{U}^H$$

式中: \mathbf{K} 和 \mathbf{M} 为适当阶数的任意矩阵.

定理 5.4.1 利用奇异值分解给出了矩阵 \mathbf{A} 的极小范数广义逆的矩阵通式(请思考能否利用相抵分解给出极小范数广义逆的通式). 该定理表明矩阵的极小范数广义逆一定存在, 且通常不唯一.

注 1: 若 $\mathbf{A} \in \mathbb{C}_n^{m \times n}$, 则 $(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \in \mathbf{A}\{1,4\}$; 若 $\mathbf{A} \in \mathbb{C}_m^{m \times n}$, 则 $\mathbf{A}^H (\mathbf{A} \mathbf{A}^H)^{-1} \in \mathbf{A}\{1,4\}$.

【思考】 能否利用 Penrose 定理求解矩阵的极小范数广义逆?

分析: $\mathbf{A}^{(1,4)}$ 同时满足 $\mathbf{A} \mathbf{A}^{(1,4)} \mathbf{A} = \mathbf{A}$ 和 $(\mathbf{A}^{(1,4)} \mathbf{A})^H = \mathbf{A}^{(1,4)} \mathbf{A}$ 两个方程, 其中第一个方程可由 Penrose 定理求解, 而另一方程则难以直接利用 Penrose 定理. 为解决这一问题, 引入下面的定理:

定理 5.4.2 设 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 则 $\mathbf{A}\{1,4\} = \{\mathbf{X} \in \mathbb{C}^{n \times m} \mid \mathbf{X} \mathbf{A} = \mathbf{A}_m^{-} \mathbf{A}\}$.

证明: 设矩阵 \mathbf{X} 满足 $\mathbf{X} \mathbf{A} = \mathbf{A}_m^{-} \mathbf{A}$, 则有 $\mathbf{A} \mathbf{X} \mathbf{A} = \mathbf{A} \mathbf{A}_m^{-} \mathbf{A} = \mathbf{A}$.

又知 $(\mathbf{X} \mathbf{A})^H = (\mathbf{A}_m^{-} \mathbf{A})^H = \mathbf{A}_m^{-} \mathbf{A} = \mathbf{X} \mathbf{A}$, 从而得到 $\mathbf{X} \in \mathbf{A}\{1,4\}$. 另一方面, 若 $\mathbf{X} \in$

$A \{1, 4\}$, 则有

$$A_m^- A = A_m^- A X A = (A_m^- A)^H (X A)^H = A^H (A_m^-)^H A^H X^H = (X A)^H = X A$$

综上所述, $A \{1, 4\} = \{X \in \mathbb{C}^{n \times m} \mid X A = A_m^- A\}$.

定理 5.4.3 设 $A_m^- \in A \{1, 4\}$, 则 $A \{1, 4\} = \{A_m^- + Z(I - A A_m^-) \mid Z \in \mathbb{C}^{n \times m}\}$.

证明: 由 Penrose 定理得方程 $X A = A_m^- A$ 的通解为

$$X = A_m^- A A_m^- + Y - Y A A_m^-, \quad \forall Y \in \mathbb{C}^{n \times m}$$

设 Z 是 $\mathbb{C}^{n \times m}$ 中任意矩阵, 将 $Y = A_m^- + Z$ 代入上式得

$$X = A_m^- + Z(I - A A_m^-)$$

因此, $A \{1, 4\} = \{A_m^- + Z(I - A A_m^-) \mid Z \in \mathbb{C}^{n \times m}\}$.

【应用】 矩阵的极小范数广义逆与相容线性方程组的解有密切关系. 考察相容线性方程组

$$A x = b \quad (5.4.1)$$

式中: $A \in \mathbb{C}^{m \times n}$ 和 $b \in \mathbb{C}^m$ 给定, $x \in \mathbb{C}^n$ 为待定向量.

当相容方程组(5.4.1)有多个解时, 往往需从中挑选一个合适的解. 此时, 一般挑选所有解中范数最小的解(或其中之一), 即定义如下优化问题:

$$\min_{\substack{A x = b \\ A A^{(1)} b = b}} \|x\|_2 \quad (5.4.2)$$

我们将满足优化问题(5.4.2)的解称为相容线性方程组(5.4.1)的极小范数解.

已知相容方程(5.4.1)的通解为 $x = A^{(1)} b + (I - A^{(1)} A) y$, $\forall y \in \mathbb{C}^n$, 则优化问题(5.4.2)可等价求矩阵 A 的某一(类)特定减号逆 G 使得 $\|G b + (I - G A) y\|_2$ 最小.

为此, 定义矩阵 G 是矩阵 A 的某一待减号逆, 则方程(5.4.1)的通解可表示为

$$x = G b + (I - G A) y, \quad \forall y \in \mathbb{C}^n \quad (5.4.3)$$

将式(5.4.3)代入 $\|x\|_2^2$, 得

$$\begin{aligned} \|x\|_2^2 &= \|G b\|_2^2 + \|(I - G A) y\|_2^2 + \\ &\quad (G b)^H (I - G A) y + ((I - G A) y)^H G b \end{aligned} \quad (5.4.4)$$

注意: 等式(5.4.4)右端第一项和第二项均为非负实数, 而由于向量 y 的任意性导致无法判断第三项和第四项之和的正负. 对此, 考察条件

$$(G b)^H (I - G A) = 0 \quad (5.4.5)$$

由于方程组(5.4.1)是相容的, 则对任意 $b \in R(A)$, 必存在 $z \in \mathbb{C}^n$ 使得 $A z = b$. 将其代入式(5.4.5), 得

$$G A z = A^H G^H G A z \quad (5.4.6)$$

式中: 向量 z 未知.

不难得到式(5.4.6)成立的一个充分条件为

$$G A = A^H G^H G A \quad (5.4.7)$$

由式(5.4.7)易得 $G A = (G A)^H$, 即 $G \in A \{1, 4\}$ 当且仅当式(5.4.7)成立. 因此,

我们选用 $G \in A\{1, 4\}$ 表示相容方程组 (5.4.1) 的通解, 则有

$$\|x\|_2^2 = \|Gb\|_2^2 + \|(I - GA)y\|_2^2 \quad (5.4.8)$$

显然, $\|x\|_2 \geq \|Gb\|_2$, 等号成立条件为当且仅当 $y = 0$. 这表明 $x = Gb$ 是相容方程组 (5.4.1) 的一个极小范数解, 并有如下定理:

定理 5.4.4 矩阵 $G \in A\{1, 4\}$ 的充分必要条件为 $x = Gb$ 是相容方程组 (5.4.1) 的极小范数解.

证明: 必要性证明可根据上述分析整理. 这里仅给出充分性证明.

若 $x = Gb$ 是相容方程组 (5.4.1) 的极小范数解, 则由定理 5.3.1 知, $G \in A\{1\}$. 又知 Gb 是相容方程组 (5.4.1) 的极小范数解, 则对任意向量 b 和 y , 有

$$\|Gb\|_2 \leq \|Gb + (I - GA)y\|_2 \quad (5.4.9)$$

令 $b = Az, z \in \mathbb{C}^n$, 则式 (5.4.9) 改写为

$$\|GAz\|_2 \leq \|GAz + (I - GA)y\|_2 \quad (5.4.10)$$

将等式 (5.4.10) 两端平方后, 整理得

$$\|(I - GA)y\|_2^2 + (GAz)^H(I - GA)y + ((I - GA)y)^H GAz \geq 0 \quad (5.4.11)$$

欲使不等式 (5.4.11) 对任意向量 z 和 y 恒成立, 有如下充分必要条件:

$$(GA)^H(I - GA) = O \quad (5.4.12)$$

由式 (5.4.12) 容易证得 $G \in A\{1, 4\}$. 证毕.

尽管定理 5.4.4 解决了相容方程组 (5.4.1) 的极小范数解的存在性问题. 但也很容易引发疑问: 选取不同的极小范数广义逆 G 所计算的 $\|Gb\|_2$ 结果相同吗? 若不相同, 是否还须从集合 $A\{1, 4\}$ 再选取某一特定 $G \in A\{1, 4\}$ 使得 $\|Gb\|_2$ 最小. 下面推论给出解答.

推论 5.4.1 相容方程 (5.4.1) 的极小范数解是唯一的.

证明: 设 G_1b 和 G_2b 是方程 (5.4.1) 的极小范数解, 则 $G_1, G_2 \in A\{1, 4\}$. 又知方程 (5.4.1) 是相容的, 则对任意向量 $b \in R(A)$, 必存在向量 $z \in \mathbb{C}^n$ 使得 $Az = b$. 此时有

$$G_1b = G_1Az, \quad G_2b = G_2Az$$

根据定理 5.4.2 知, $G_1A = G_2A$. 因此, $G_1b = G_2b$, 即相容方程 (5.4.1) 的极小范数解是唯一的. 证毕.

例 5.4.1 求相容方程组 $Ax = b$ 的极小范数解, 其中

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

解: 方法一: 由于矩阵 A 行满秩, 故 $A^H(AA^H)^{-1} \in A\{1, 4\}$.

$$A^H(AA^H)^{-1} = \begin{bmatrix} \frac{1}{9} & \frac{2}{9} \\ \frac{4}{9} & -\frac{1}{9} \\ -\frac{1}{9} & \frac{5}{18} \end{bmatrix}$$

此时, $\mathbf{x} = \mathbf{Gb} = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{6} \right]^T$ 是方程组的极小范数解.

方法二: 对矩阵 \mathbf{A} 进行奇异值分解, 得

$$\mathbf{A} = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ 2 & -1 \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{5}}{3} & 0 & -\frac{2}{3} \\ \frac{2\sqrt{5}}{15} & \frac{2\sqrt{5}}{5} & \frac{1}{3} \\ \frac{4\sqrt{5}}{15} & -\frac{\sqrt{5}}{5} & \frac{2}{3} \end{bmatrix}^H$$

于是, \mathbf{A} 的一个极小范数解为

$$\mathbf{G} = \begin{bmatrix} \frac{\sqrt{5}}{3} & 0 & -\frac{2}{3} \\ \frac{2\sqrt{5}}{15} & \frac{2\sqrt{5}}{5} & \frac{1}{3} \\ \frac{4\sqrt{5}}{15} & -\frac{\sqrt{5}}{5} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix}^H = \begin{bmatrix} \frac{1}{9} & \frac{2}{9} \\ \frac{4}{9} & -\frac{1}{9} \\ -\frac{1}{9} & \frac{5}{18} \end{bmatrix}$$

此时, $\mathbf{x} = \mathbf{Gb} = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{6} \right]^T$ 是方程组的极小范数解.

例 5.4.2 求向量 $\mathbf{b} \in \mathbb{C}^m$ 在 $R(\mathbf{A}^H)$ 上的正交投影, 其中 $\mathbf{A} \in \mathbb{C}^{m \times n}$.

解: 由定理 5.3.3 定义正交投影矩阵 $\mathbf{P} = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^- \mathbf{A}$. 则向量 $\mathbf{b} \in \mathbb{C}^m$ 在 $R(\mathbf{A}^H)$ 上的正交投影为

$$\text{Proj}_{R(\mathbf{A}^H)} \mathbf{b} = \mathbf{Pb} = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^- \mathbf{Ab} \quad (5.4.13)$$

现对矩阵 \mathbf{A} 进行奇异值分解, 即 $\mathbf{A} = \mathbf{U} \begin{bmatrix} \boldsymbol{\Sigma}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{V}^H$. 则 $\mathbf{A}\mathbf{A}^H$ 的任一极小范数广义逆可表示为

$$(\mathbf{A}\mathbf{A}^H)^- = \mathbf{U} \begin{bmatrix} \boldsymbol{\Sigma}_r^{-2} & \mathbf{K} \\ \mathbf{O} & \mathbf{M} \end{bmatrix} \mathbf{U}^H$$

式中: \mathbf{K} 和 \mathbf{M} 为适当阶数的任意矩阵.

由此, 正交投影矩阵 \mathbf{P} 为

$$\mathbf{P} = \mathbf{V} \begin{bmatrix} \mathbf{I}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{V}^H$$

又知 \mathbf{A} 的任一极小范数广义逆可表示为

$$\mathbf{A}_m^- = \mathbf{V} \begin{bmatrix} \boldsymbol{\Sigma}_r^{-1} & \mathbf{L} \\ \mathbf{O} & \mathbf{N} \end{bmatrix} \mathbf{U}^H$$

进而

$$\mathbf{A}_m^- \mathbf{A} = \mathbf{V} \begin{bmatrix} \mathbf{I}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{V}^H$$

式中: \mathbf{L} 和 \mathbf{N} 为适当阶数的任意矩阵.

因此, $\mathbf{P} = \mathbf{A}_m^- \mathbf{A}$. 也就是说, 向量 $\mathbf{b} \in \mathbb{C}^m$ 在 $R(\mathbf{A}^H)$ 上的正交投影可以用 \mathbf{A} 的极小范数广义逆表示, 即

$$\text{Proj}_{R(\mathbf{A})} \mathbf{b} = \mathbf{A}_m^- \mathbf{A} \mathbf{b}$$

5.5 最小二乘广义逆

$\mathbf{A}^{(1,3)}$ 是矩阵 \mathbf{A} 的最小二乘广义逆, 其存在性有如下结论:

定理 5.5.1 设 $\mathbf{A} \in \mathbb{C}_r^{m \times n}$ 的奇异值分解为

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \boldsymbol{\Sigma}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{V}^H$$

则 $\mathbf{G} \in \mathbf{A}\{1,3\}$ 的充分必要条件是

$$\mathbf{G} = \mathbf{V} \begin{bmatrix} \boldsymbol{\Sigma}_r^{-1} & \mathbf{O} \\ \mathbf{L} & \mathbf{M} \end{bmatrix} \mathbf{U}^H$$

式中: \mathbf{L} 和 \mathbf{M} 为适当阶数的任意矩阵.

定理 5.5.1 利用奇异值分解给出了矩阵 \mathbf{A} 的最小二乘广义逆的矩阵通式. 该定理表明矩阵的最小二乘广义逆一定存在, 且通常不唯一.

注 1: 若 $\mathbf{A} \in \mathbb{C}_n^{m \times n}$, 则 $(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \in \mathbf{A}\{1,3\}$; 若 $\mathbf{A} \in \mathbb{C}_m^{m \times n}$, 则 $\mathbf{A}^H (\mathbf{A} \mathbf{A}^H)^{-1} \in \mathbf{A}\{1,3\}$.

在利用 Penrose 定理求解矩阵最小二乘广义逆通式时, 我们遇到了与求解极小范数广义逆类似的问题. 对此, 引入以下定理:

定理 5.5.2 设 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 则 $\mathbf{A}\{1,3\} = \{\mathbf{X} \in \mathbb{C}^{n \times m} | \mathbf{A}\mathbf{X} = \mathbf{A}\mathbf{A}_l^-\}$.

证明: 假设矩阵 \mathbf{X} 满足 $\mathbf{A}\mathbf{X} = \mathbf{A}\mathbf{A}_l^-$, 则有 $\mathbf{A}\mathbf{X}\mathbf{A} = \mathbf{A}\mathbf{A}_l^- \mathbf{A} = \mathbf{A}$. 又知 $(\mathbf{A}\mathbf{X})^H = (\mathbf{A}\mathbf{A}_l^-)^H = \mathbf{A}\mathbf{A}_l^- = \mathbf{A}\mathbf{X}$, 从而 $\mathbf{X} \in \mathbf{A}\{1,3\}$.

另一方面,

$$\mathbf{A}\mathbf{A}_l^- = \mathbf{A}\mathbf{X}\mathbf{A}\mathbf{A}_l^- = (\mathbf{A}\mathbf{X})^H (\mathbf{A}\mathbf{A}_l^-)^H = \mathbf{X}^H \mathbf{A}^H (\mathbf{A}_l^-)^H \mathbf{A}^H = (\mathbf{A}\mathbf{X})^H = \mathbf{A}\mathbf{X}$$

综上所述, $\mathbf{A}\{1,3\} = \{\mathbf{X} \in \mathbb{C}^{n \times m} | \mathbf{A}\mathbf{X} = \mathbf{A}\mathbf{A}_l^-\}$.

定理 5.5.3 设 $\mathbf{A}_l^- \in \mathbf{A}\{1,3\}$, 则 $\mathbf{A}\{1,3\} = \{\mathbf{A}_l^- + (\mathbf{I} - \mathbf{A}_l^- \mathbf{A})\mathbf{Z} | \mathbf{Z} \in \mathbb{C}^{n \times m}\}$.

证明: 由 Penrose 定理得方程 $\mathbf{A}\mathbf{X} = \mathbf{A}\mathbf{A}_l^-$ 的通解为

$$\mathbf{X} = \mathbf{A}_l^- \mathbf{A} \mathbf{A}_l^- + \mathbf{Y} - \mathbf{A}_l^- \mathbf{A} \mathbf{Y}, \quad \forall \mathbf{Y} \in \mathbb{C}^{n \times m}$$

令 \mathbf{Z} 是 $\mathbb{C}^{n \times m}$ 中任意矩阵, 将 $\mathbf{Y} = \mathbf{A}_l^- + \mathbf{Z}$ 代入上式得

$$\mathbf{X} = \mathbf{A}_l^- + (\mathbf{I} - \mathbf{A}_l^- \mathbf{A})\mathbf{Z}$$

因此, $\mathbf{A}\{1,3\} = \{\mathbf{A}_l^- + (\mathbf{I} - \mathbf{A}_l^- \mathbf{A})\mathbf{Z} | \mathbf{Z} \in \mathbb{C}^{n \times m}\}$.

【应用】 矩阵的最小二乘广义逆与最小二乘问题有着密切关系. 考察线性方程组

$$Ax = b \quad (5.5.1)$$

式中: $A \in \mathbb{C}^{m \times n}$ 和 $b \in \mathbb{C}^m$ 给定, $x \in \mathbb{C}^n$ 为待定向量.

若方程组(5.5.1)不相容,则它没有通常意义的解.对此,我们退而求其次,求得残量范数(向量2范数)最小的解,即求向量 x 使得 $\|Ax - b\|_2$ 最小.通常将这一优化问题称为线性最小二乘问题(例1.5.8和例3.7.4已分别讨论),向量 x 为不相容方程组(5.5.1)的最小二乘解.

从几何角度看,求解不相容方程组(5.5.1)的最小二乘解的核心问题是求解向量 b 在 $R(A)$ 的正交投影.由例5.3.4知,向量 b 在 $R(A)$ 的正交投影可由矩阵 A 的减号逆表示,即 $\text{Proj}_{R(A)} b = A(A^H A)^- A^H b$.这里采用矩阵的最小二乘广义逆来表示正交投影矩阵.

定理 5.5.4 对任意矩阵 $A \in \mathbb{C}^{m \times n}$, AA^- 是正交投影矩阵.

证明: 由于 A^- 是 A 的一个减号逆且 AA^- 是幂等矩阵,故 AA^- 是幂等矩阵.又知 $AA^- = (AA^-)^H$,故 AA^- 是正交投影矩阵.

推论 5.5.1 向量 b 在 $R(A)$ 的正交投影为 AA^-b ,其中 $A \in \mathbb{C}^{m \times n}$ 和 $b \in \mathbb{C}^m$.

推论 5.5.2 给定 $A \in \mathbb{C}^{m \times n}$ 和 $b \in \mathbb{C}^m$, $A(A^H A)^- A^H b = AA^-b$.

定理 5.5.5 矩阵 $G \in A\{1,3\}$ 的充分必要条件为 $x = Gb$ 是不相容方程(5.5.1)的最小二乘解.

证明: 必要性. 设 $G \in A\{1,3\}$, 则 $\forall x \in \mathbb{C}^n$, 有

$$\begin{aligned} \|Ax - b\|_2^2 &= \| (AGb - b) + A(x - Gb) \|_2^2 \\ &= \|Ax - AGb\|_2^2 + \|b - AGb\|_2^2 + \\ &\quad (Ax - AGb)^H (b - AGb) + \\ &\quad (b - AGb)^H (Ax - AGb) \end{aligned} \quad (5.5.2)$$

由于 $G \in A\{1,3\}$, 则 $A^H = (AGA)^H = A^H (AG)^H = A^H GA$. 由此,

$$(Ax - AGb)^H (b - AGb) = (x - Gb)^H A^H (I - AG)b = 0$$

对上式左右两端进行共轭转置得

$$(b - AGb)^H (Ax - AGb) = 0$$

因此,

$$\|Ax - b\|_2^2 = \|Ax - AGb\|_2^2 + \|b - AGb\|_2^2 \geq \|b - AGb\|_2^2$$

这表明 $x = Gb$ 是不相容方程组 $Ax = b$ 的最小二乘解.

充分性. 若 $x = Gb$ 是不相容方程组 $Ax = b$ 的最小二乘解, 则对任意向量 $x \in \mathbb{C}^n$ 和 $b \in \mathbb{C}^m$, 有

$$\|Ax - b\|_2 \geq \|AGb - b\|_2$$

对上述不等式两端平方,并由式(5.5.2)知,上述不等式恒成立的充分必要条件

是对任意向量 $x \in \mathbb{C}^n$ 和 $b \in \mathbb{C}^m$, $(Ax - AGb)^H(b - AGb) = 0$. 由向量 b 和向量 $x - Gb$ 的任意性, 可得 $A^H = A^H GA$, 进而 $G \in A\{1, 3\}$. 证毕.

推论 5.5.3 向量 x 是不相容方程 (5.5.1) 的最小二乘解当且仅当 x 是相容方程组 $Ax = AA_l^-b$ 的解, 且 $Ax = b$ 的最小二乘解的通式为

$$x = A_l^-b + (I - A_l^-A)y, \quad \forall y \in \mathbb{C}^n$$

注 2: 方程组 $Ax = AA_l^-b$, $Ax = A(A^HA)^-A^Hb$ 和 $A^HAx = A^Hb$ 均是相容同解方程组.

例 5.5.1 求方程组 $Ax = b$ 的最小二乘解通式, 其中

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

解: 矩阵 A 的奇异值分解为

$$A = \begin{bmatrix} -\frac{\sqrt{5}}{5} & \frac{2\sqrt{5}}{5} \\ -\frac{2\sqrt{5}}{5} & \frac{\sqrt{5}}{5} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{\sqrt{5}}{5} & -\frac{2\sqrt{5}}{5} & 0 \\ -\frac{2\sqrt{5}}{5} & \frac{\sqrt{5}}{5} & 0 \\ 0 & 0 & 1 \end{bmatrix}^H$$

则取 A 的最小二乘广义逆为

$$A_l^- = \begin{bmatrix} -\frac{\sqrt{5}}{5} & -\frac{2\sqrt{5}}{5} & 0 \\ -\frac{2\sqrt{5}}{5} & \frac{\sqrt{5}}{5} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{5} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{\sqrt{5}}{5} & \frac{2\sqrt{5}}{5} \\ -\frac{2\sqrt{5}}{5} & \frac{\sqrt{5}}{5} \end{bmatrix}^H = \begin{bmatrix} \frac{1}{25} & \frac{2}{25} \\ \frac{2}{25} & \frac{4}{25} \\ 0 & 0 \end{bmatrix}$$

由此, 方程组的最小二乘通解为

$$x = A_l^-b + (I - A_l^-A)y = \begin{bmatrix} \frac{3}{25} \\ \frac{6}{25} \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{4}{5} & -\frac{2}{5} & 0 \\ -\frac{2}{5} & \frac{1}{5} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

式中: y_1, y_2 和 y_3 为任意数.

定理 5.5.6 不相容方程组 $Ax = b$ 具有唯一的最小二乘解当且仅当 A 是列满秩矩阵.

证明: 由定理 5.3.1 知, A 是列满秩矩阵当且仅当相容方程组 $Ax = AA_l^-b$ 有唯一解. 再由推论 5.5.3 知, A 是列满秩矩阵当且仅当不相容方程组 $Ax = b$ 具有唯一的最小二乘解.

例 5.5.2 求方程组 $Ax = b$ 最小二乘解的通式, 其中

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

解: 由 $\text{rank}(\mathbf{A}) = 2, \text{rank}(\mathbf{A} | \mathbf{b}) = 3$ 知, 方程组不相容. 又因为 \mathbf{A} 是列满秩矩阵, 故 $\mathbf{Ax} = \mathbf{b}$ 具有唯一的最小二乘解. 因此,

$$(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H = \begin{bmatrix} -\frac{4}{11} & \frac{7}{11} & \frac{1}{11} \\ \frac{7}{11} & -\frac{4}{11} & \frac{1}{11} \end{bmatrix} \in \mathbf{A}\{1, 3\}$$

方程组的最小二乘解为

$$\mathbf{x} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{b} = \begin{bmatrix} -\frac{4}{11} \\ \frac{7}{11} \\ 0 \end{bmatrix}$$

5.6 加号逆

$\mathbf{A}^{(1,2,3,4)}$ 是矩阵 \mathbf{A} 的最小二乘广义逆, 其存在性有如下结论:

定理 5.6.1 (加号逆存在性定理) 设矩阵 $\mathbf{A} \in \mathbb{C}_r^{m \times n}$ 的奇异值分解为

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \boldsymbol{\Sigma}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{V}^H$$

则

$$\mathbf{G} = \mathbf{V} \begin{bmatrix} \boldsymbol{\Sigma}_r^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{U}^H$$

是 \mathbf{A} 的一个加号逆.

证明: 将矩阵 \mathbf{A} 和 \mathbf{G} 依次代入 4 个 Penrose 方程, 得

$$\mathbf{AGA} = \mathbf{U} \begin{bmatrix} \boldsymbol{\Sigma}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{V}^H = \mathbf{A}$$

$$\mathbf{GAG} = \mathbf{V} \begin{bmatrix} \boldsymbol{\Sigma}_r^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{U}^H = \mathbf{G}$$

$$(\mathbf{AG})^H = \mathbf{U} \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{U}^H = \mathbf{AG}$$

$$(\mathbf{GA})^H = \mathbf{V} \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{V}^H = \mathbf{GA}$$

因此, 矩阵 \mathbf{G} 是矩阵 \mathbf{A} 的一个加号逆.

定理 5.6.1 利用奇异值分解给出了矩阵 \mathbf{A} 的最小二乘广义逆的一个特解. 该定理表明矩阵的加号逆一定存在, 这也表明矩阵的任一广义逆矩阵都存在.

定理 5.6.2 (加号逆唯一性定理) 任给矩阵 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 其加号逆 \mathbf{A}^+ 唯一.

证明: \mathbf{A}^+ 存在性由定理 5.6.1 给出. 这里假设矩阵 \mathbf{A} 的加号逆不唯一, 即矩阵 \mathbf{X} 和 \mathbf{Y} 均满足 4 个 Penrose 方程, 则

$$\begin{aligned}\mathbf{X} &= \mathbf{XAX} = \mathbf{XAYAX} = (\mathbf{XA})^H (\mathbf{YA})^H \mathbf{X} = \mathbf{A}^H \mathbf{X}^H \mathbf{A}^H \mathbf{Y}^H \mathbf{X} = (\mathbf{AXA})^H \mathbf{Y}^H \mathbf{X} \\ &= \mathbf{A}^H \mathbf{Y}^H \mathbf{X} = \mathbf{YAX} = \mathbf{YAYAX} = \mathbf{Y}(\mathbf{AY})^H (\mathbf{AX})^H = \mathbf{YY}^H \mathbf{A}^H \mathbf{X}^H \mathbf{A}^H \\ &= \mathbf{YY}^H (\mathbf{AXA})^H = \mathbf{YY}^H \mathbf{A}^H = \mathbf{YAY} = \mathbf{Y}\end{aligned}$$

下面分别利用奇异值分解和满秩分解来计算矩阵的加号逆.

方法一. 利用奇异值分解计算. 定理 5.6.1 已提供了基于奇异值分解的加号逆计算方法. 现对这一结果作进一步改进, 并有如下定理:

定理 5.6.3 设矩阵 $\mathbf{A} = (a_{ij}) \in \mathbb{C}_r^{m \times n}$, 则

$$\mathbf{A}^+ = \mathbf{V}_1 (\boldsymbol{\Sigma}_r^2)^{-1} \mathbf{V}_1^H \mathbf{A}^H$$

其中, $\boldsymbol{\Sigma}_r = \text{diag}(\sigma_1, \dots, \sigma_r)$, $\sigma_1, \dots, \sigma_r$ 是矩阵 \mathbf{A} 的 r 个正奇异值, $\mathbf{V}_1 \in \mathbb{C}^{n \times r}$ 是由 $\mathbf{A}^H \mathbf{A}$ 的属于特征值 $\sigma_1^2, \dots, \sigma_r^2$ 的 r 个单位正交特征向量构成.

证明: 设 $\mathbf{V}_2 \in \mathbb{C}^{n \times (n-r)}$ 是由 $\mathbf{A}^H \mathbf{A}$ 的属于特征值 0 的 $(n-r)$ 个单位正交特征向量构成, 并定义 $\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2)$, 则矩阵 $\mathbf{A} \in \mathbb{C}_r^{m \times n}$ 的奇异值分解为

$$\mathbf{A} = [\mathbf{U}_1, \mathbf{U}_2] \begin{bmatrix} \boldsymbol{\Sigma}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^H \\ \mathbf{V}_2^H \end{bmatrix} \quad (5.6.1)$$

式中, $\mathbf{U}_1 = \mathbf{A} \mathbf{V}_1 \boldsymbol{\Sigma}_r^{-1}$, $\mathbf{U}_2 \in \mathbb{C}^{m \times (m-r)}$ 是由 $\mathbf{A} \mathbf{A}^H$ 的属于特征值 0 的 $(m-r)$ 个单位正交特征向量组成.

根据定理 5.6.1 知,

$$\mathbf{A}^+ = [\mathbf{V}_1, \mathbf{V}_2] \begin{bmatrix} \boldsymbol{\Sigma}_r^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{U}_1^H \\ \mathbf{U}_2^H \end{bmatrix} = \mathbf{V}_1 \boldsymbol{\Sigma}_r^{-1} \mathbf{U}_1^H = \mathbf{V}_1 (\boldsymbol{\Sigma}_r^2)^{-1} \mathbf{V}_1^H \mathbf{A}^H$$

例 5.6.1 已知矩阵 $\mathbf{A} = \begin{bmatrix} -1 & 0 & 1 \\ 2 & 0 & -2 \end{bmatrix}$, 求 \mathbf{A}^+ .

解: 矩阵 \mathbf{A} 的奇异值分解为

$$\mathbf{A} = \begin{bmatrix} -\frac{\sqrt{5}}{5} & \frac{2\sqrt{5}}{5} \\ \frac{2\sqrt{5}}{5} & \frac{\sqrt{5}}{5} \end{bmatrix} \begin{bmatrix} \sqrt{10} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \\ -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \end{bmatrix}^H$$

则矩阵 \mathbf{A} 的加号逆为

$$\mathbf{A}^+ = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \\ -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{10}}{10} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{\sqrt{5}}{5} & \frac{2\sqrt{5}}{5} \\ \frac{2\sqrt{5}}{5} & \frac{\sqrt{5}}{5} \end{bmatrix}^H = \frac{1}{10} \begin{bmatrix} -1 & 2 \\ 0 & 0 \\ 1 & -2 \end{bmatrix}$$

考察一种特殊情况: $\mathbf{A} \in \mathbb{C}_1^{m \times n}$, 则矩阵 \mathbf{A} 有且仅有一个正奇异值, 不妨设 $\sigma_1 > 0$, 则 σ_1^2 是矩阵 $\mathbf{A}^H \mathbf{A}$ 唯一的非零特征值, 且 $(\Sigma_r^2)^{-1} = \sigma_1^{-2}$.

另一方面, 根据特征值与矩阵迹的关系, 得

$$\sigma_1^2 = \text{tr}(\mathbf{A}^H \mathbf{A}) = \sum_{i,j} |a_{ij}|^2$$

将上式代入 $\mathbf{A}^+ = \mathbf{V}_1 (\Sigma_r^2)^{-1} \mathbf{V}_1^H \mathbf{A}^H$, 得

$$\mathbf{A}^+ = \frac{1}{\sum_{i,j} |a_{ij}|^2} \mathbf{V}_1 \mathbf{V}_1^H \mathbf{A}^H$$

由于 $\mathbf{A}^H \mathbf{A} \mathbf{V}_2 = \mathbf{O}$, 故 $\mathbf{A} \mathbf{V}_2 = \mathbf{O}$. 又知 $\mathbf{V}_1 \mathbf{V}_1^H + \mathbf{V}_2 \mathbf{V}_2^H = \mathbf{I}_n$, 则

$$\mathbf{A}^H = (\mathbf{V}_1 \mathbf{V}_1^H + \mathbf{V}_2 \mathbf{V}_2^H) \mathbf{A}^H = \mathbf{V}_1 \mathbf{V}_1^H \mathbf{A}^H$$

因此, $\mathbf{A}^+ = \frac{1}{\sum_{i,j} |a_{ij}|^2} \mathbf{A}^H$, 并有如下推论:

推论 5.6.1 (秩 1 公式) 设 $\mathbf{A} = (a_{ij}) \in \mathbb{C}_1^{m \times n}$, $\mathbf{A}^+ = \frac{1}{\sum_{i,j} |a_{ij}|^2} \mathbf{A}^H$.

由推论 5.6.1 知, 例 5.6.1 中的加号逆 \mathbf{A}^+ 可直接写出来:

$$\mathbf{A}^+ = \frac{1}{10} \begin{bmatrix} -1 & 2 \\ 0 & 0 \\ 1 & -2 \end{bmatrix}$$

方法二. 利用满秩分解计算.

定理 5.6.4 设矩阵 $\mathbf{A} \in \mathbb{C}_r^{m \times n}$ 有满秩分解 $\mathbf{A} = \mathbf{B}\mathbf{C}$, 其中, $\mathbf{B} \in \mathbb{C}_r^{m \times r}$, $\mathbf{C} \in \mathbb{C}_r^{r \times n}$, 则 $\mathbf{A}^+ = \mathbf{C}^H (\mathbf{C}\mathbf{C}^H)^{-1} (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H = \mathbf{C}^H (\mathbf{B}^H \mathbf{A} \mathbf{C}^H)^{-1} \mathbf{B}^H$.

证明: 将 $\mathbf{A} = \mathbf{B}\mathbf{C}$ 和 $\mathbf{G} = \mathbf{C}^H (\mathbf{C}\mathbf{C}^H)^{-1} (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H$ 依次代入 4 个 Penrose 方程, 得

$$\begin{aligned} \mathbf{A}\mathbf{G}\mathbf{A} &= \mathbf{B}\mathbf{C}\mathbf{A} = \mathbf{A} \\ \mathbf{G}\mathbf{A}\mathbf{G} &= \mathbf{C}^H (\mathbf{C}\mathbf{C}^H)^{-1} (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H = \mathbf{G} \\ (\mathbf{A}\mathbf{G})^H &= \mathbf{B} (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H = \mathbf{A}\mathbf{G} \\ (\mathbf{G}\mathbf{A})^H &= \mathbf{C}^H (\mathbf{C}\mathbf{C}^H)^{-1} \mathbf{C} = \mathbf{G}\mathbf{A} \end{aligned}$$

因此, 矩阵 \mathbf{G} 是矩阵 \mathbf{A} 的加号逆.

推论 5.6.2 若 $\mathbf{A} \in \mathbb{C}_n^{m \times n}$, $\mathbf{A}^+ = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$; 若 $\mathbf{A} \in \mathbb{C}_m^{m \times n}$, $\mathbf{A}^+ = \mathbf{A}^H (\mathbf{A} \mathbf{A}^H)^{-1}$.

例 5.6.2 求 $A = \begin{bmatrix} 1 & 2 & 1 & 2 \\ 2 & 4 & 2 & 4 \\ 1 & 0 & 0 & 2 \end{bmatrix}$ 的伪逆.

解: 对矩阵 A 进行满秩分解为

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & \frac{1}{2} & 0 \end{bmatrix}$$

由此, 矩阵 A 的加号逆为

$$A^+ = C^H (B^H A C^H)^{-1} B^H = \begin{bmatrix} 0 & 0 & \frac{1}{5} \\ \frac{2}{25} & \frac{4}{25} & -\frac{2}{5} \\ \frac{1}{25} & \frac{2}{25} & -\frac{1}{5} \\ 0 & 0 & \frac{2}{5} \end{bmatrix}$$

命题 5.6.1 (加号逆的性质) 设 $A \in \mathbb{C}^{m \times n}$, 则

- (1) $(A^+)^+ = A$;
- (2) $(A^H)^+ = (A^+)^H$;
- (3) $(A^T)^+ = (A^+)^T$;
- (4) $\text{rank}(A^+) = \text{rank}(A)$;
- (5) $(A^H A)^+ = A^+ (A^H)^+$, $(A A^H)^+ = (A^H)^+ A^+$;
- (6) $A^+ = (A^H A)^+ A^H = A^H (A A^H)^+$.

证明: 性质(1)~(4)和(6)均可利用加号逆的定义进行证明, 这里仅以性质(5)为例进行说明.

由于

$$\begin{aligned} A^H A (A^+ (A^H)^+) A^H A &= A^H (A A^+)^H (A^+)^H A^H A \\ &= (A A^+ A)^H (A A^+)^H A = A^H A A^+ A = A^H A \\ (A^+ (A^H)^+) A^H A (A^+ (A^H)^+) &= A^+ (A A^+)^H (A A^+)^H (A^+)^H \\ &= A^+ A A^+ (A^+ A A^+)^H = A^+ (A^+)^H = A^+ (A^H)^+ \end{aligned}$$

故 $A^+ (A^H)^+ \in (A^H A)^{(1,2)}$.

又知

$$\begin{aligned} (A^H A A^+ (A^H)^+)^H &= A^+ (A^H)^+ A^H A = A^+ (A A^+)^H A = A^+ A A^+ A = A^+ A \\ A^H A A^+ (A^H)^+ &= A^H (A A^+)^H (A^H)^+ = (A^+ A A^+ A)^H = A^+ A \\ \text{故 } (A^H A A^+ (A^H)^+)^H &= A^H A A^+ (A^H)^+, \text{ 即 } A^+ (A^H)^+ \in (A^H A)^{\{3\}}. \\ (A^+ (A^H)^+ A^H A)^H &= A^H A A^+ (A^+)^H = A^H (A A^+)^H (A^+)^H \end{aligned}$$

$$\begin{aligned}
 &= (\mathbf{A}^+ \mathbf{A} \mathbf{A}^+ \mathbf{A})^H = \mathbf{A}^+ \mathbf{A} \mathbf{A}^+ (\mathbf{A}^H)^+ \mathbf{A}^H \mathbf{A} \\
 &= \mathbf{A}^+ \mathbf{A} \mathbf{A}^+ \mathbf{A} = \mathbf{A}^+ \mathbf{A}
 \end{aligned}$$

故 $(\mathbf{A}^+ (\mathbf{A}^H)^+ \mathbf{A}^H \mathbf{A})^H = \mathbf{A}^+ (\mathbf{A}^H)^+ \mathbf{A}^H \mathbf{A}$, 即 $\mathbf{A}^+ (\mathbf{A}^H)^+ \in (\mathbf{A}^H \mathbf{A})^{\{4\}}$.

综上所述, $\mathbf{A}^+ (\mathbf{A}^H)^+ = (\mathbf{A}^H \mathbf{A})^+$. 同理可证 $(\mathbf{A} \mathbf{A}^H)^+ = (\mathbf{A}^H)^+ \mathbf{A}^+$.

注 1: 对于一般的矩阵 \mathbf{A} 和 \mathbf{B} , $(\mathbf{A}\mathbf{B})^+ \neq \mathbf{B}^+ \mathbf{A}^+$, $\mathbf{A}^+ \mathbf{A} \neq \mathbf{A} \mathbf{A}^+ \neq \mathbf{I}$.

【应用】 加号逆与线性方程组的关系. 考察线性方程组

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad (5.6.2)$$

式中: $\mathbf{A} \in \mathbb{C}^{m \times n}$ 和 $\mathbf{b} \in \mathbb{C}^m$ 给定, $\mathbf{x} \in \mathbb{C}^n$ 为待定向量.

由前面章节内容, 总结以下结论:

推论 5.6.3 线性方程(5.6.2)相容的充分必要条件为 $\mathbf{A}\mathbf{A}^+ \mathbf{b} = \mathbf{b}$, 此时通解为

$$\mathbf{x} = \mathbf{A}^+ \mathbf{b} + (\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \mathbf{y}, \quad \forall \mathbf{y} \in \mathbb{C}^n \quad (5.6.3)$$

若方程(5.6.2)不相容, 则式(5.6.3)为其最小二乘解通式.

一般而言, 方程(5.6.2)的最小二乘解并不唯一, 我们通常把范数最小的一个解称为方程(5.6.2)的极小(范数)最小二乘解(或最佳逼近解).

定理 5.6.5 向量 $\mathbf{x} = \mathbf{A}^+ \mathbf{b}$ 既是相容方程(5.6.2)的唯一极小范数解, 也是不相容方程(5.6.2)的唯一最佳逼近解.

证明: $\mathbf{A}^+ \in \mathbf{A}\{1, 4\}$, 故 $\mathbf{x} = \mathbf{A}^+ \mathbf{b}$ 是相容方程(5.6.2)的唯一极小范数解.

由推论 5.5.3 知, 不相容方程(5.6.2)的最小二乘解当且仅当是相容方程 $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{A}^+ \mathbf{b}$ 的解. 又知相容方程 $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{A}^+ \mathbf{b}$ 的极小范数解唯一且可表示为

$$\mathbf{x} = \mathbf{A}^+ \mathbf{A} \mathbf{A}^+ \mathbf{b} = \mathbf{A}^+ \mathbf{b}$$

因此, $\mathbf{x} = \mathbf{A}^+ \mathbf{b}$ 是不相容方程组(5.6.2)的唯一最佳逼近解. 证毕.

例 5.6.3 求线性方程(5.6.2)的极小范数解或最佳逼近解, 其中

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 4 \\ -2 \\ -2 \end{bmatrix}$$

解: \mathbf{A} 的满秩分解为

$$\mathbf{A} = \mathbf{F}\mathbf{G} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

于是

$$\mathbf{A}^+ = \mathbf{G}^H (\mathbf{G}\mathbf{G}^H)^{-1} (\mathbf{F}^H \mathbf{F})^{-1} \mathbf{F}^H = \frac{1}{18} \begin{bmatrix} 5 & 2 & -1 \\ 1 & 1 & 1 \\ -4 & -1 & 2 \\ 6 & 3 & 0 \end{bmatrix}$$

由于 $\mathbf{A}\mathbf{A}^+ \mathbf{b} = [3, 0, -3]^T \neq \mathbf{b}$, 所以方程 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 不相容, 它的最佳逼近解为

$$\mathbf{x}_0 = \mathbf{A}^+ \mathbf{b} = [1, 0, -1, 1]^T$$

例 5.6.4 求向量 $b \in \mathbb{C}^n$ 在 $R(A)$ 和 $N(A)$ 上的正交投影, 其中 $A \in \mathbb{C}^{n \times n}$.

解: 由于 AA^+ 是正交投影矩阵, 且 $AA^+b \in R(A)$, 故 $\text{Proj}_{R(A)} b = AA^+b$. 又知 $N(A) \oplus R(A^H) = \mathbb{C}^n$ 是正交直和分解, 则 $b \in \mathbb{C}^n$ 在 $N(A)$ 上的正交投影为

$$\text{Proj}_{N(A)} b = (I - A^H(A^H)^+)b = (I - A^+A)b$$

5.7 应用: 区间线性规划

线性规划模型诞生于 20 世纪 40 年代, 最早时被称为生产组织与计划中的数学方法. 60 年代初, 康托洛维奇和库伯曼斯应用线性规划模型提出了资源最优分配理论, 并凭此荣获诺贝尔经济学奖; 自此, 线性规划模型受到诸多科学和工程领域的广泛关注.

区间线性规划模型(带双边约束的线性规划模型)作为线性规划模型的一种特殊形式, 更适用于解决部分拥有特殊性质的规划问题. 例如在交通规划问题中: 若给定了一个道路集, 我们要限制各条道路上的车总量不能过多, 否则会造成该集中道路的拥堵, 同时也不能让这个道路集上的车过少, 这样才能为其他道路集疏解压力. 我们可以定义约束条件中过多部分为资源分配问题, 即要求所有个体所占用的资源和要小于总资源数; 定义约束条件中过少部分为成本收益问题, 即要求每个个体的贡献和要大于某一成本. 既要考虑资源分配, 又需考虑成本收益的问题, 我们都可以称之为带双边约束的规划问题, 并可以通过区间线性规划模型进行求解.

定义 5.7.1 (区间线性规划) 设 $A = (a_{ij}) \in \mathbb{R}^{m \times n}$, $a = [a_1, \dots, a_m]^T$ 和 $b = [b_1, \dots, b_m]^T \in \mathbb{R}^m$, c 和 $x \in \mathbb{R}^n$, 则线性规划问题

$$\max c^T x \quad (5.7.1)$$

$$\text{s. t. } a_i \leq \sum_{j=1}^n a_{ij} x_j \leq b_i, \quad i = 1, \dots, m$$

称为区间线性规划(或带双边约束的线性规划).

若用 $a \leq b$ 表示 $a_i \leq b_i, i = 1, \dots, m$, 则该线性规划问题可简记为

$$\max c^T x \quad (5.7.2)$$

$$\text{s. t. } a \leq Ax \leq b$$

若向量 x 满足约束条件 $a \leq Ax \leq b$, 则称 x 为规划问题(5.7.2)的一个可行解. 记

$$S = \{x \in \mathbb{R}^n \mid a \leq Ax \leq b\}$$

称 S 为规划问题(5.7.2)的可行解集. 若 S 非空, 则称该规划问题是可行的. 若

$$\max\{c^T x, x \in S\} < \infty$$

则称该规划问题是有界的; 若存在 $x_0 \in S$ 使得

$$c^T x_0 = \max\{c^T x, x \in S\}$$

则称 x_0 为区间线性规划问题(5.7.2)的最优解.

下面两个定理通过广义逆矩阵刻画了规划问题(5.7.2)的有界性及最优解集.

定理 5.7.1 区间线性规划问题(5.7.2)有界,当且仅当 $\mathbf{c} \in R(\mathbf{A}^T)$, 或等价地, $\mathbf{c}^T \mathbf{A}^- \mathbf{A} = \mathbf{c}^T$.

证明: 设 \mathbf{x} 为任一可行解, $\mathbf{u} \in N(\mathbf{A})$, 则容易验证 $\mathbf{x} + \mathbf{u} \in S$. 这表明

$$S + N(\mathbf{A}) = S$$

进一步, 有

$$\max\{\mathbf{c}^T \mathbf{x}, \mathbf{x} \in S\} = \max\{\mathbf{c}^T \mathbf{x}, \mathbf{x} \in [S + N(\mathbf{A})]\}$$

又知

$$\mathbf{c}^T \mathbf{x} = \mathbf{c}^T \mathbf{A}^+ \mathbf{A} \mathbf{x} + \mathbf{c}^T (\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \mathbf{x}$$

则

$$\max\{\mathbf{c}^T \mathbf{x}, \mathbf{x} \in [S + N(\mathbf{A})]\} = l_1 + l_2$$

式中:

$$l_1 = \max\{\mathbf{c}^T \mathbf{A}^+ \mathbf{A} \mathbf{x}, \mathbf{x} \in S\}$$

$$l_2 = \max\{\mathbf{c}^T (\mathbf{I} - \mathbf{A}^+ \mathbf{A}) \mathbf{x}, \mathbf{x} \in N(\mathbf{A})\}$$

显然, l_1 是有限的. 另一方面, l_2 有限 $\Leftrightarrow \forall \mathbf{x} \in N(\mathbf{A}), \mathbf{c}^T \mathbf{x} = 0 \Leftrightarrow \mathbf{c} \in N(\mathbf{A})^\perp$. 由定理 1.5.3 知, $N(\mathbf{A})^\perp = R(\mathbf{A}^T)$. 故区间线性规划问题(5.7.2)有界, 当且仅当 $\mathbf{c} \in R(\mathbf{A}^T)$. 根据命题 5.3.2 知, $\mathbf{c} \in R(\mathbf{A}^T)$ 等价于 $\mathbf{c}^T \mathbf{A}^- \mathbf{A} = \mathbf{c}^T$. 证毕.

定理 5.7.2 假设区间线性规划问题(5.7.2)是可行有解的, $\text{rank}(\mathbf{A}) = m$, 则该问题的最优解集为

$$\mathbf{x} = \mathbf{A}^- \mathbf{u} + (\mathbf{I} - \mathbf{A}^- \mathbf{A}) \mathbf{v} \quad (5.7.3)$$

式中, 向量 $\mathbf{v} \in \mathbb{R}^n$ 任意, $\mathbf{u} = [u_1, \dots, u_m]^T \in \mathbb{R}^m$ 满足

$$u_i = \begin{cases} a_i, & \text{若 } (\mathbf{c}^T \mathbf{A}^-)_i < 0 \\ b_i, & \text{若 } (\mathbf{c}^T \mathbf{A}^-)_i > 0 \\ \theta_i a_i + (1 - \theta_i) b_i, & \text{若 } (\mathbf{c}^T \mathbf{A}^-)_i = 0 \end{cases} \quad (5.7.4)$$

$(\mathbf{c}^T \mathbf{A}^-)_i$ 表示行向量 $\mathbf{c}^T \mathbf{A}^-$ 的第 i 个分量, $0 \leq \theta_i \leq 1$.

证明: 由于 $\mathbf{A}^- \mathbf{A}$ 是幂等矩阵, 故 $R(\mathbf{A}^- \mathbf{A}) + R(\mathbf{I} - \mathbf{A}^- \mathbf{A}) = \mathbb{R}^n$. 由此, \mathbb{R}^n 的任一向量 \mathbf{x} 都可以表示为

$$\mathbf{x} = \mathbf{A}^- \mathbf{A} \mathbf{w} + (\mathbf{I} - \mathbf{A}^- \mathbf{A}) \mathbf{v}$$

式中: $\mathbf{w} \in \mathbb{R}^n, \mathbf{v} \in \mathbb{R}^n$.

根据该规划问题的有界性, 利用定理 5.7.1 知, $\mathbf{c} \in R(\mathbf{A}^T)$, 于是

$$\mathbf{c}^T (\mathbf{I} - \mathbf{A}^- \mathbf{A}) \mathbf{v} = 0$$

进一步,

$$\mathbf{c}^T \mathbf{x} = \mathbf{c}^T [\mathbf{A}^- \mathbf{A} \mathbf{w} + (\mathbf{I} - \mathbf{A}^- \mathbf{A}) \mathbf{v}] = \mathbf{c}^T \mathbf{A}^- \mathbf{A} \mathbf{w}$$

由于 $\text{rank}(\mathbf{A}) = m$, 故 $\mathbf{A} \mathbf{A}^- = \mathbf{I}_m$. 此时,

$$\mathbf{A} \mathbf{A}^- \mathbf{u} = \mathbf{u}$$

上式表明若向量 \mathbf{u} 已知, 方程组 $\mathbf{A} \mathbf{w} = \mathbf{u}$ 是相容方程组.

定义 $u = Aw$, 则 $c^T x = c^T A^{-1} u$. 此时易知, u 选取如式(5.7.4)所示, 则 $c^T x$ 取最大值; 相应地, 对于给定如式(5.7.4)所示的向量 u , 必可以找到向量 w . 这表明 $c^T x$ 的最大值可以取到. 证毕.

关于广义逆在一般规划问题中应用得更广泛的结果, 请读者参阅文献[24]及其后引用文献.

例 5.7.1 为缓解交通压力, 某市政府对 A、B、C 3 个区进行交通管控. 已知 A 区有 1 条高速路, 2 条普通公路; B 区有 2 条高速路, 4 条普通公路; C 区有 3 条高速路, 5 条普通公路. 假设 3 个区高速路上车辆数量相同, 3 个区普通公路上数量也相同. 现分别要求 A、B、C 的车辆承载量分别为 $[20, 30]$, $[30, 40]$, $[40, 80]$, 求此市所能容纳的最大车辆数.

解: 上述问题可转化为区间线性规划问题(5.7.1), 其中 $a = [20, 30, 40]^T$, $b = [30, 40, 80]^T$, $c = [6, 11]^T$,

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 5 \end{bmatrix}$$

由此求出矩阵 A 的一个减号逆 A^- 为

$$A^- = \begin{bmatrix} -5 & 0 & 2 \\ 3 & 0 & -1 \end{bmatrix}$$

经计算

$$c^T A^- A = c^T = [6, 11]$$

故该区间线性规划问题有界. 又知 $c^T A^- = [3, 0, 1]$, 由定理 5.7.2 得

$$u = \begin{bmatrix} 30 \\ 30\theta_2 + 40(1 - \theta_2) \\ 80 \end{bmatrix}$$

由式(5.7.3)可得该问题的最优解为 $x_0 = [10, 10]^T$, 此时该市所能容纳的最大车辆数为 $c^T x_0 = 170$.

本章习题

1. 矩阵左逆和右逆是广义逆矩阵吗? 请说明理由.
2. 证明推论 5.2.2.
3. 试用满秩分解给出矩阵方程 $AXA = A$ 的一个特解.
4. 证明命题 5.2.1.

5. 设 $X \in \mathbb{C}^{n \times n}$, $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \end{bmatrix}$, $D = \begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix}$. 试采

用不同方法判断矩阵方程 $AXB = D$ 是否相容; 若相容, 求其通解.

6. 求解矩阵方程 $\mathbf{XA}=\mathbf{B}$, 其中

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

7. 试用两种不同方法求矩阵 $\mathbf{A} = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$ 的一个减号逆.

8. 求 $\mathbf{Ax}=\mathbf{b}$ 的通解, 其中

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}, \quad \mathbf{b} = [2, 1, 1]^T.$$

9. 证明命题 5.3.1.

10. 若矩阵 $\mathbf{A} \in \mathbb{C}_m^{m \times n}$, 证明 $\mathbf{AA}^- = \mathbf{I}_m$.

11. 证明方程组 (5.3.7) 与方程组 (5.3.9) 是相容的, 且是同解方程组.

12. 设 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 若 \mathbf{A} 行满秩, 证明其极小范数广义逆唯一.

13. 试用不同方法求矩阵 $\mathbf{A} = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$ 的一个极小范数广义逆.

14. 试利用满秩分解给出矩阵的一个极小范数广义逆.

15. 求方程组 $\mathbf{Ax}=\mathbf{b}$ 的极小范数解, 其中

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}, \quad \mathbf{b} = [2, 1, 1]^T$$

16. 设有一组实验数据 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, 希望由实验数据拟合曲线使其误差

$$E(a, b) = \sum_{i=1}^n (y_i - ax_i - b)^2$$

最小, 假定拟合曲线为 $y=ax+b$. 试求 a 和 b .

17. 证明 $\mathbf{Ax}=\mathbf{AA}_l^- \mathbf{b}$ 和 $\mathbf{A}^H \mathbf{Ax}=\mathbf{A}^H \mathbf{b}$ 是相容的, 且为同解方程组.

18. 证明 $\mathbf{AA}_l^- = \mathbf{A}(\mathbf{A}^H \mathbf{A})^- \mathbf{A}^H$.

19. 设 $\mathbf{A} \in \mathbb{C}^{m \times n}$, 证明 $\mathbf{A}^+ = \mathbf{A}^{(1,4)} \mathbf{AA}^{(1,3)}$.

20. 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 是正规矩阵, 证明 $\mathbf{A}^+ \mathbf{A} = \mathbf{AA}^+$.

21. 若 $\mathbf{AB}=\mathbf{O}$, 证明 $\mathbf{B}^+ \mathbf{A}^+ = \mathbf{O}$.

22. 某养鸡场现有 1 000 只小鸡, 用黄豆和玉米混合饲料喂养. 每只鸡每天吃 1~1.3 kg 饲料. 从营养角度看, 每只鸡每天需要 0.004~0.006 kg 的钙, 并至少需要 0.21~0.23 kg 的蛋白质. 已知黄豆的蛋白质含量为 48%~52%, 钙含量为 0.5%~

0.8%, 其价格为 0.38~0.42 元/kg; 玉米的蛋白质含量为 8.5%~11.5%, 钙含量为 0.3%, 其价格为 0.2 元/kg. 问每天如何配料最节省?

23. 解非线性方程组常采用 Newton 迭代法, 其中所需的 Jacobi 矩阵要满足可逆条件. 若 Jacobi 矩阵不可逆, 能否用广义逆矩阵来代替它的普通逆? 请大家思考并尝试解决.

参考文献

- [1] Horn R A, Johnson C R. Matrix Analysis. Cambridge: Cambridge University Press, 2012.
- [2] 张绍飞, 赵迪. 矩阵论教程. 2 版. 北京: 机械工业出版社, 2012.
- [3] 戴华. 矩阵论. 北京: 科学出版社, 2001.
- [4] 张跃辉. 矩阵理论与应用. 北京: 科学出版社, 2011.
- [5] Lay D C, Lay S R, McDonald J J. 线性代数及其应用. 刘深泉, 张万芹, 陈玉珍, 等译. 北京: 机械工业出版社, 2018.
- [6] 张贤达. 矩阵分析与应用. 2 版. 北京: 清华大学出版社, 2013.
- [7] 程云鹏, 张凯院, 徐仲. 矩阵论. 西安: 西北工业大学出版社, 1999.
- [8] 陈祖明, 周家胜. 矩阵论引论. 北京: 北京航空航天大学出版社, 1998.
- [9] 庞晶, 周凤玲, 张余. 矩阵论. 北京: 化学工业出版社, 2013.
- [10] 陈公宁. 矩阵理论与应用. 北京: 高等教育出版社, 1990.
- [11] 徐树方. 矩阵计算的理论与方法. 北京: 北京大学出版社, 1995.
- [12] 史荣昌, 魏丰. 矩阵分析. 3 版. 北京: 北京理工大学出版社, 2010.
- [13] 李乔, 张晓东. 矩阵论十讲. 合肥: 中国科学技术大学出版社, 2015.
- [14] 胡寿松. 自动控制原理. 6 版. 北京: 科学出版社, 2013.
- [15] 冯康, 等. 数值计算方法. 北京: 国防工业出版社, 1978.
- [16] 林成森. 数值计算方法. 2 版. 北京: 科学出版社, 2005.
- [17] Bapat R B. 图与矩阵. 吴少川, 译. 哈尔滨: 哈尔滨工业大学出版社, 2014.
- [18] 殷剑宏, 吴开亚. 图论及其算法. 合肥: 中国科学技术大学出版社, 2003.
- [19] Denton Peter B, Parke Stephen J, Tao Terence, et al. Eigenvectors from Eigenvalues: a survey of a basic identity in linear algebra, arXiv: 1908. 03795v3. (2020-3-4) [2020-3-20]. <https://arxiv.org/abs/1908.03795?context=math.RA>.

- [20] Belkin M, Niyogi P. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, 2003, 15(6): 1373-1396.
- [21] Girvan M, Newman M E J. Community structure in social and biological networks//*Proceedings of the National Academy of Sciences of the United States of America*, 2001, 99(12): 7821-7826.
- [22] Newman M E J. Fast algorithm for detecting community structure in networks. *Physical Review E*, 2004, 69(6): 066-133.
- [23] Gene G, Frank U. The QR algorithm: 50 years later its genesis by John Francis and Vera Kublanovskaya and subsequent developments. *IMA Journal of Numerical Analysis*, 2009, 29: 467-485.
- [24] 王松桂, 杨振海. 广义逆矩阵及其应用. 北京: 北京工业大学出版社, 2006.
- [25] Nashed M Z. Generalized inverses to and applications. New York: Academic Press, 1976.
- [26] Noble B, Daniel J W. Applied linear algebra. 3rd ed. Englewood Cliffs: Prentice-Hall, 1988.

汉英名称索引

A

按范数收敛 normal convergence, 168
按坐标收敛 pointwise convergence, 168

B

半负定 negative semidefinite, 17
伴随矩阵 adjugate, 115
半正定 positive semidefinite, 17
标量 scalar, 3
标准正交基 orthonormal basis, 21
标准正交组 orthonormal set, 21
病态问题 ill-conditioned problem, 156
不变因子 invariant factor, 118
不变子空间 invariant subspace, 72
不可约矩阵 irreducible matrix, 70

C

Casorati 行列式 Casorati determinant, 9
Cauchy 不等式 Cauchy inequality, 19
差分方程 difference equation, 6
Cholesky 分解 Cholesky factorization, 83
Cholesky 算法 Cholesky algorithm, 84
初等变换 elementary operation, 115
初等矩阵 elementary matrix, 115
初等因子 elementary factor, 119

D

代数重数 algebraic multiplicity, 58
单纯矩阵 simple matrix, 96
单射 injection, 38
单位向量 unit vector, 19
度量矩阵 Gram matrix, 17
对称变换 symmetric transformation, 72
对角化 diagonalization, 96
对角占优矩阵 diagonally dominant matrix, 164
多项式插值 polynomial interpolation, 31
二次型 quadratic form, 17

F

F -范数 F -norm, 149
反对称矩阵 skew-symmetric matrix, 102
范数 norm, 140
负定 negative definite, 17
赋范线性空间 normed linear space, 140
傅里叶三角级数 Fourier series, 22
复数域 field of complex numbers, 3

G

盖尔圆 Gerschgorin circle, 161
盖尔圆盘 Gerschgorin disc, 161
Givens 矩阵 Givens matrix, 65

公倍式 common multiple, 40
共轭对称性 conjugate symmetry, 16
共轭转置 conjugate transpose, 16
公因式 common factor, 40
Gram-Schmidt 正交化方法 Gram-Schmidt-orthogonalization process, 29
关联矩阵 incidence matrix, 69
广义逆矩阵 generalized inverse matrix, 203
过渡矩阵 transfer matrix, 12

H

Hamilton-Cayley 定理 Hamilton-Cayley theorem, 94
函数空间 function space, 44
行和范数 row sum matrix norm, 151
行列式 determinant, 20
和空间 sum space, 7
合同(相合) congruent, 18
Hermite 矩阵 Hermite matrix, 16
Holder 不等式 Holder inequality, 143
Householder 变换 Householder transformation, 67
Householder 矩阵 Householder matrix, 66

J

Jacobian 矩阵 Jacobian matrix, 192
基 base, 11
迹 trace, 58
基础解系 fundamental set of solution, 8
极大线性无关组 maximal linearly independent set, 10
极分解 polar factorization, 132
几何重数 geometric multiplicity, 58
极小范数解 minimum norm solution, 217
极小范数逆 minimum norm inverse, 204
加号逆 Moore-Penrose inverse, 204
加权范数 weighted norm, 144
加群 additive group, 3
交空间 intersection space, 7

镜面反射 mirror reflection, 66
Jordan 标准形 Jordan normal form, 124
Jordan 块 Jordan block, 123
矩阵范数 matrix norm, 147
矩阵分解 matrix decomposition, 74
矩阵函数 matrix function, 175
矩阵级数 matrix series, 170
矩阵余弦函数 matrix cosine function, 176
矩阵正弦函数 matrix sine function, 176
矩阵指数函数 matrix exponential function, 176

K

可加性 additivity, 16
可逆矩阵 invertible matrix, 42
可逆映射 invertible mapping, 39
Kronecker 积 Kronecker product, 209

L

LDU 分解 LDU factorization, 82
LU 分解 LU factorization, 79
拉普拉斯矩阵 Laplacian matrix, 70
拉普拉斯特征映射 Laplacian Eigenmaps, 71
连通分支 connected component, 69
连通图 connected graph, 69
良态问题 well-conditioned problem, 157
列和范数 column sum matrix norm, 151
列空间 column space, 8
列展开 column stacking, 211
邻接矩阵 adjacency matrix, 69
零度 nullity, 44
零化多项式 annihilator polynomial, 95
零空间 null space, 8
Lyapunov 方程 Lyapunov equation, 144

M

满射 surjection, 38
满秩分解 full-rank decomposition, 76
矛盾方程 inconsistent equations, 204
幂等矩阵 idempotent matrix, 107

幂零矩阵 nilpotent matrix, 136
 模 modulus, 19
 模式分类 pattern classification, 144
 Moore-Penrose 广义逆 Moore-Penrose inverse, 204

N

内积 inner product, 16
 内积空间 inner product space, 16
 能控性 controllability, 196
 Neumann 级数 Neumann series, 174

O

欧几里得范数 Euclidean norm, 141
 欧氏空间 Euclidean space, 16

P

p -范数 p -norm, 142
 排列矩阵 permutation matrix, 85
 Penrose 定理 Penrose theorem, 207
 Penrose 方程 Penrose equation, 203
 平行四边形法则 parallelogram law, 19
 谱 spectrum, 160
 谱半径 spectral radius, 160
 谱点 spectral point, 183
 谱范数 spectral norm, 152
 谱分解 spectral decomposition, 105
 谱阵 spectral matrix, 105
 谱值 spectral value, 183

Q

QR 分解 QR decomposition, 87
 QR 算法 QR algorithm, 90
 齐次性 homogeneity, 16
 奇异值 singular value, 128
 奇异值分解 singular value decomposition, 129
 区间线性规划 interval linear program, 228

R

Rayleigh 商 Rayleigh quotient, 154

S

三次样条 cubic spline, 33
 三角不等式 triangle inequality, 19
 三角分解 triangular factorization, 79
 Schur 分解 Schur decomposition, 91
 上三角矩阵 upper triangular matrix, 79
 实数域 field of real numbers, 3
 数乘 scalar multiplication, 3
 数域 number field, 3
 双射 bijection, 38
 Smith 标准形 Smith normal form, 117
 算子范数 operator norm, 151
 Sylvester 公式 Sylvester reduction, 181

T

特征多项式 characteristic polynomial, 57
 特征矩阵 characteristic matrix, 57
 特征向量 eigenvector, 56
 特征值 eigenvalue, 56
 特征子空间 eigenspace, 58
 条件数 condition number, 158
 同构 isomorphism, 49
 投影 projection, 27
 投影矩阵 projection matrix, 107
 图 graph, 68
 图像压缩 image compression, 134

V

Vandermonde 行列式 Vandermonde determinant, 60
 Vandermonde 矩阵 Vandermonde matrix, 60

W

微分方程 differential equation, 100
 伪逆 pseudo inverse, 204
 维数 dimension, 13
 维数定理 dimension theorem, 14
 ∞ -范数 ∞ -norm, 140

无限维 infinite dimensional,13

X

系统矩阵 syatem matrix,196

系统稳定性 system stability,144

下三角矩阵 lower triangular matrix,79

线性变换 linear transformation,40

线性定常系统 linear time-invariant system,196

线性矩阵方程 linear matrix equation,204

线性空间 linear space,3

线性滤波器 linear filter,6

线性算子 linear operator,40

线性无关 linearly independent,9

线性相关 linearly dependent,9

线性映射 linear mapping,40

线性组合 linear combination,5

像 image,38

相抵 equivalent,53

相抵标准形 equivalent standard form,74

向量 vector,3

向量范数 vector norm,140

向量夹角 vector angle,20

向量空间 vector space,2

相容方程 system of consistent equations,204

相似 similar,53

相似变换 similarity transformation,53

协方差矩阵 covariance matrix,199

Y

$\{1\}$ -广义逆 $\{1\}$ -generalized inverse,204

$\{1,3\}$ -逆 $\{1,3\}$ -inverse,204

$\{1,4\}$ -逆 $\{1,4\}$ -inverse,204

一元多项式 unary polynomial,39

映射 mapping,38

诱导范数 induced norm,151

友矩阵 companion matrix,40

酉矩阵 unitary matrix,62

酉空间 unitary space (U-space),16

右逆 right inverse,79

原像 preimage,38

Z

正定二次型 positive definite quadratic form,17

正定矩阵 positive definite matrix,16

正规矩阵 normal matrix,102

正交 orthogonal,21

正交变换 orthogonal transformation,62

正交补 orthogonal complement,22

正交矩阵 orthogonal matrix,62

正交三角分解 QR factorization,87

正交投影矩阵 orthogonal projection matrix,107

正交直和分解 orthogonal direct sum decomposition,26

正交组 orthogonal set,21

秩 rank,10

直和 direct sum,23

置换矩阵 permutation matrix,85

主元分析法 principal component analysis,197

转置 transpose,16

状态方程 state equation,196

子空间 subspace,5

最佳逼近解 best approximate solution,227

最小二乘广义逆 generalized inverse of least square,204

最小二乘解 least square solution,221

坐标 coordinate,11

左逆 left inverse,79