北京郵電大學

期 末 课 程 论 文

题目：矩阵函数的求法和矩阵分解方法研究

姓　　名　李永琰

学　　院　计算机学院

专　　业　计算机类

班　　级　**2021211314**

学　　号　**2021211353**

班内序号　**14**

任课教师　李昊辰

2022 年 12 月

摘要

矩阵的运算是数值分析领域的重要问题. 将矩阵分解为简单矩阵的组合可以在理论和实际应用上简化矩阵的运算. 在现代多种领域中, 都能看到矩阵的出现. 本文基于矩阵论，介绍了若干重要的定理, 并主要分析了矩阵函数的求法和矩阵分解的方法.

关键词: 矩阵理论, 线性变换, 矩阵函数, 矩阵分解, 线性方程组

Abstract

The operation of matrices is an important problem in the field of numerical analysis. Decomposing matrices into combinations of simple matrices can be found in theory and practical applications to simplify matrix operations. In many modern fields, the emergence of matrices can be seen. This article is based on moments Matrix theory, introduces some important theorems, and mainly analyzes the method of matrix function calculation and matrix factorization.

Key Words: Matrix Theory, Linear Transformation, Matrix Function, Matrix Decomposition, Linear System of Equations

目录

摘要 I

Abstract II

第 1 章 引言 1

1.1 背景介绍 . . . 1

1.1.1 矩阵理论与方法介绍 . . . 1

1.1.2 函数矩阵和矩阵函数介绍 . . . 1

1.1.3 线性代数方程组求解介绍 . . . 2

1.2 问题介绍 . . . 2

1.2.1 矩阵函数的求法问题介绍 . . . 2

1.2.2 矩阵分解的方法问题介绍 . . . 2

1.3 上述问题国内外研究成果介绍 . . . 3

1.3.1 矩阵函数的求法研究现状 . . . 3

1.3.2 矩阵分解方法研究现状 . . . 3

1.4 本论文工作简述 . . . 4

1.4.1 本论文对上述问题研究简述 . . . 4

1.4.2 本论文创新点或特点简述 . . . 4

1.4.3 本论文撰写结构简述 . . . 4

第 2 章 预备知识 5

2.1 欧式空间与线性变换 . . . 5

2.1.1 欧式空间与线性变换介绍 . . . 5

2.1.2 若尔当标准型的求解 . . . 17

2.1.3 欧式空间中线性变换的求法 . . . 23

2.2 向量范数与矩阵范数 . . . 27

2.2.1 向量范数介绍 . . . 27

2.2.2 矩阵范数介绍 . . . 29

2.2.3 矩阵可逆性条件、谱半径和条件数介绍 . . . 31

2.3 矩阵函数介绍 . . . 34

2.3.1 矩阵序列介绍 . . . 34

2.3.2 矩阵级数介绍 . . . 36

2.3.3 矩阵函数介绍 . . . 38

2.3.4 函数矩阵对矩阵的导数 . . . 39

第 3 章 矩阵函数的求法研究 41
3.1 待定系数法 41
3.1.1 待定系数法求矩阵函数的步骤推导 41
3.1.2 举例展示求法 41
3.2 数项级数求和法 42
3.2.1 数项级数求和法求矩阵函数的步骤推导 42
3.2.2 举例展示求法 43
3.3 对角型法 43
3.3.1 对角型法求矩阵函数的步骤推导 43
3.3.2 举例展示求法 44
3.4 若尔当标准型法 45
3.4.1 若尔当标准型法求矩阵函数的步骤推导 45
3.4.2 举例展示求法 46

第 4 章 矩阵分解方法研究 48
4.1 矩阵的 LU 分解 48
4.1.1 矩阵 LU 分解的步骤推导 48
4.1.2 举例展示求法 50
4.2 矩阵的 QR 分解 51
4.2.1 矩阵 QR 分解的步骤推导 51
4.2.2 举例展示求法 55
4.3 矩阵的满秩分解 58
4.3.1 矩阵满秩分解的步骤推导 58
4.3.2 举例展示求法 59
4.4 矩阵的奇异值分解 60
4.4.1 矩阵奇异值分解的步骤推导 60
4.4.2 举例展示求法 61
4.4.3 利用奇异值分解求矩阵广义逆 62

第 5 章 总结 64

参考文献 65

致谢 66

第 1 章 引言

1.1 背景介绍

1.1.1 矩阵理论与方法介绍

矩阵理论是一门研究矩阵在数学上应用的科目, 其是学习数值分析、最优化理论、概率统计、运筹学、控制理论、力学、电学、信息科学、管理科学与工程的基础, 尤其是计算机的广泛应用, 为矩阵理论的应用开辟了广阔的前景.

根据世界数学发展史记载, 矩阵概念产生于 19 世纪 50 年代, 是为了解线性方程组的需要而产生的. 然而, 在公元前我国就已经有了矩阵的萌芽. 在我国的《九章算术》一书中已经有所描述, 只是没有将它作为一个独立的概念加以研究, 而仅用它解决实际问题, 所以没能形成独立的矩阵理论.

1850 年, 英国数学家西尔维斯特 (SylveSter, 1814–1897) 在研究方程的个数与未知量的个数不相同的线性方程组时, 由于无法使用行列式, 所以引入了矩阵的概念.

1855 年, 英国数学家凯莱 (Caylag, 1821–1895) 在研究线性变换下的不变量时, 为了简洁、方便, 引入了矩阵的概念. 1858 年, 凯莱在《矩阵论的研究报告》中, 定义了两个矩阵相等、相加以及数与矩阵的数乘等运算和算律, 同时, 定义了零矩阵、单位阵等特殊矩阵, 更重要的是在该文中他给出了矩阵相乘、矩阵可逆等概念, 以及利用伴随阵求逆阵的方法, 证明了有关的算律, 如矩阵乘法有结合律, 没有交换律, 两个非零阵乘积可以为零矩阵等结论, 定义了转置阵、对称阵、反对称阵等概念.

1878 年, 德国数学家弗罗伯纽斯 (Frobeniws, 1849–1917) 在他的论文中引入了 λ 矩阵的行列式因子、不变因子和初等因子等概念, 证明了两个 λ 矩阵等价当且仅当它们有相同的不变因子和初等因子, 同时给出了正交矩阵的定义,1879 年, 他又在自己的论文中引进矩阵秩的概念.

矩阵的理论发展非常迅速, 到 19 世纪末, 矩阵理论体系已基本形成. 到 20 世纪, 矩阵理论得到了进一步的发展. 目前, 它已经发展成为在物理、控制论、机器人学、生物学、经济学等学科有大量应用的数学分支.

1.1.2 函数矩阵和矩阵函数介绍

矩阵函数是以矩阵为自变量且取值为矩阵的一类函数, 它是对一元函数概念的推广. 起先, 矩阵函数是一个收敛的矩阵幂级数的和来定义. 之后, 根据计算矩阵函数值的 Jordan 标准形方法, 对矩阵函数的概念进行了拓宽. 因此矩阵函数的基础是矩阵序列与矩阵级数.

函数矩阵是以函数为元素的矩阵. 函数矩阵的导数与积分是将通常函数的导数与积分等概念形式上推广到矩阵的情形. 当一个矩阵的元素都是变量 t 的函数时, 可以建立矩阵对变量 t 的导数与积分概念; 当一个多元函数的自变量都是矩阵 $\boldsymbol{X}$ 的元素时, 可以

建立起多元函数对矩阵 $\boldsymbol{X}$ 的导数概念; 当一个矩阵的元素都是矩阵 $\boldsymbol{X}$ 的元素的多元函数时, 可以建立矩阵对矩阵 $\boldsymbol{X}$ 的导数概念.

函数矩阵和矩阵函数是两个不同的概念. 但是, 在一些情形下, 矩阵函数在其定义内的值可以看做函数矩阵. 比如, 矩阵 $\boldsymbol{A}$ 的指数还能输 $e^{\boldsymbol{A}t}$ 可以看作变量 t 的函数矩阵. 借助于矩阵的指数函数, 可以给出某些线性微分方程组和线性矩阵方程一般解的解析表达式.

1.1.3 线性代数方程组求解介绍

线性方程组的求解问题在科学技术与经济管理领域有着广泛的应用. 线性规划问题, 某些工程问题, 经济问题等都可以转化成线性方程组求解问题.

一般 n 元线性方程组是指形如

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

的方程组, 其中 $x_1, x_2, \cdots, x_n$ 为未知量, $a_{ij}(i = 1, 2, \cdots, m; j = 1, 2, \cdots, n)$ 为方程组的系数, m 为方程个数, $b_j(j = 1, 2, \cdots, m)$ 为方程组的常数项. 若常数项全为零, 则称其为 n 元齐次线性方程组; 若常数项不全为零, 则称其为 n 元非齐次先线性方程组.

易知齐次线性方程组一定有解, 即零解必为它的一个解, 所以求解齐次线性方程组实际上是探究其是否有非零解. 而非齐次线性方程组的求解则更加复杂, 因为一个线性方程组的解可能是无解, 唯一解, 无穷多个解, 所以在解非齐次线性方程组时需先对其解进行判定再探究其解的结构.

1.2 问题介绍

1.2.1 矩阵函数的求法问题介绍

矩阵函数理论时矩阵理论的一个重要组成部分. 矩阵函数把对矩阵的研究带入分析领域. 同时也解决了数学领域及工程技术等其它领域的计算难题.

矩阵函数定义的引出把矩阵理论延伸到分析, 从而使矩阵的研究又提升到一个新的层次, 增加了新的手段, 同时也使矩阵理论在数学、物理等许多领域有了新的应用.

然而, 矩阵函数的形式复杂多样, 而且计算方法和理论也不尽相同, 因此需要根据实际问题选取不同的解法, 以达到事半功倍的效果.

1.2.2 矩阵分解的方法问题介绍

矩阵分解对矩阵理论及近代计算数学的发展起了关键的作用. 寻求矩阵在各种意义下的分解形式, 是对矩阵有关的数值计算和理论都有着极为重要的意义. 因为这些分解

式的特殊形式, 一是能明显的反映出原矩阵的某些特征; 二是分解的方法与过程提供了某些有效的数值计算方法和理论分析依据. 这些分解在数值代数和最优化问题的解决中都有着十分重要的角色以及在其它领域方面起着必不可少的作用.

1.3 上述问题国内外研究成果介绍

1.3.1 矩阵函数的求法研究现状

待定系数法是以 Hamilton-Cayley 定理为基础的一种求矩阵函数的方法. 借助于 Hamilton-Cayley 定理, 可以将矩阵函数的求值问题 (即矩阵幂级数的求和问题) 转化为矩阵多项式的计算问题.

数项级数求和法是根据最小多项式导出的矩阵递推关系来求解矩阵函数的方法.

对角形法是线性代数课程中已经介绍过的求矩阵函数的方法. 其是矩阵函数中较简单的一种, 这取决于可对角化矩阵的优良性质.

Jordan 标准型法在形式上类似于对角形法. 借助于矩阵的 Jordan 标准形理论, 可以将矩阵函数的求值问题转化为矩阵的乘法运算问题.

1.3.2 矩阵分解方法研究现状

对于一般的 n 阶方阵而言, 前 $n-1$ 个顺序主子式不等于零是可进行三角分解的充要条件. 在此条件下, n 阶方阵的 LDU 分解、Crout 分解以及 Doolittle 分解都是唯一的.

矩阵的 LDU 分解可以通过它的 Crout 分解或者 Doolittle 分解构造出来, 矩阵的 Doolittle 分解也可以通过它的 Crout 分解构造出来, 反之亦然.

借助于矩阵的三角分解, 可以将一般方阵的求逆计算转换为上三角矩阵和下三角矩阵的求逆计算, 也可以将一般线性代数方程组的求解问题转化为两个三角方程组的求解问题, 也就是求解线性代数方程组的追赶法.

任何矩阵都可以进行 QR 分解. 方阵可以分解为正交矩阵与上三角矩阵的乘积; 当 $m > n$ 时, $m \times n$ 矩阵可以分解为列向量组标准正交的矩阵与上三角矩阵的乘积; 当 $m < n$ 时, $m \times n$ 矩阵可以分解为行向量组标准正交的矩阵与上三角矩阵的乘积.

借助于矩阵的 QR 分解, 可以将一般方阵的求逆计算转化为三角矩阵的求逆计算, 也可以将一般线性代数方程组的求解问题转化为三角方程组的求解问题.

任何矩阵都可以进行满秩分解和奇异值分解. 相对而言, 求矩阵的满秩分解要比求矩阵的奇异值分解容易一些, 特别是采用 Hermite 标准形方法求矩阵的满秩分解更为简单.

矩阵的奇异值概念是对矩阵特征值概念的推广, 矩阵的奇异值分解是对矩阵的正交相似对角化问题的推广. 矩阵的奇异值可以给出该矩阵的值域和零空间的基, 也可以给

出以该矩阵为系数矩阵的线性代数方程组的最小二乘解和具有最小 2-范数的最小二乘解.

1.4 本论文工作简述

1.4.1 本论文对上述问题研究简述

本论文总结了函数矩阵的各种求法以及矩阵分解的各类方法, 并举例具体说明每种方法的优点与其局限性, 针对不同的矩阵来选择不同的方法, 以此提高效率、优化解答、过程, 更好的应用于实际当中.

1.4.2 本论文创新点或特点简述

本论文在分析研究学者理论的基础上, 从系统化、全局化的角度入手结合了多个角度，全面的对矩阵函数的求解和矩阵的分解进行了分析.

经过查阅大量资料, 对于该课题的研究大多数并没有给出理论依据, 仅提供了方法, 本论文更加注重方法的理论建设, 知其然并知其所以然, 更好地能够得到应用于实践中的结果.

1.4.3 本论文撰写结构简述

本论文分为四部分:

第一部分主要介绍研究上述问题的基石. 主要包括线性空间与线性变换、向量范数与矩阵范数以及矩阵函数的介绍;

第二部分是对矩阵函数求法展开的研究. 探讨了待定系数法、数项级数求和法、对角型法和若尔当标准型法求解矩阵函数的步骤推导, 并具体展示了具体求解过程;

第三部分是对矩阵分解方法展开的研究. 探讨了 LU 分解、QR 分解、满秩分解和奇异值分解的步骤推导, 并具体展示了具体求解过程;

第四部分是对本论文工作的总结.

第 2 章　预备知识

本章是正式开始研究两个大问题之前的准备工作, 主要介绍了欧氏空间、线性变换、向量范数、矩阵范数以及矩阵函数, 这些都是之后研究问题的基石.

2.1　欧式空间与线性变换

本节研究欧氏空间及线性变换的性质, 特别是几种重要的线性变换、对若尔当标准形的求解和线性变换的求法.

2.1.1　欧式空间与线性变换介绍

线性空间

线性空间是线性代数中最基本的概念之一, 也是学习现代矩阵论的重要基础. 线性空间的概念, 是某类事物从量的方面的一个抽象. 我们不考虑集合的对象, 抽去它们的具体内容的含义, 来研究这类集合的公共性质, 并引进一个概括性名词. 于是就有如下的线性空间的概念.

定义 2.1　设 V 是一个非空集合, 他的元素用 $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ 等表示, 并称之为向量; K 是一个数域, 它的元素用 k, l, m 等表示. 如果 V 满足条件:

(1) 在 V 中定义一个加法运算, 即当 $\boldsymbol{x}, \boldsymbol{y} \in V$ 时, 有唯一的和 $\boldsymbol{x} + \boldsymbol{y} \in V$, 且加法运算满足以下性质:

1) 结合律 $\boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{x}) = (\boldsymbol{x} + \boldsymbol{y}) = z$;

2) 交换律 $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$;

3) 存在零元素 $\boldsymbol{0}$, 使 $\boldsymbol{x} + \boldsymbol{0} = \boldsymbol{x}$;

4) 存在负元素, 即对任一向量 $\boldsymbol{x} \in V$, 存在向量 $\boldsymbol{y} \in V$, 使 $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{0}$, 则称 $\boldsymbol{y}$ 为 $\boldsymbol{x}$ 的负元素, 记为 $-\boldsymbol{x}$, 于是有 $\boldsymbol{x} + (-\boldsymbol{x}) = \boldsymbol{0}$.

(2) 在 V 中定义数乘（数与向量的乘法）运算, 即当 $\boldsymbol{x} \in V, k \in K$ 时, 有唯一的乘积 $k\boldsymbol{x} \in V$, 且数乘运算满足以下性质:

5) 数因子分配律 $k(\boldsymbol{x} + \boldsymbol{y} = k\boldsymbol{x} + k\boldsymbol{y})$;

6) 分配律 $(k + l)\boldsymbol{x} = k\boldsymbol{x} + l\boldsymbol{x}$;

7) 结合律 $k(l\boldsymbol{x}) = (kl)\boldsymbol{x}$;

8) $1 \cdot \boldsymbol{x} = \boldsymbol{x}$

则称 V 为数域 K 上的线性空间或向量空间.

V 中所定义的加法及数乘运算统称为 V 的线性运算. 在不致产生混淆时, 将数域 K 上的线性空间简称为线性空间. 数 k 与向量 $\boldsymbol{x}$ 的乘积 $k\boldsymbol{x}$ 也可写成 $\boldsymbol{x}k$.

需要指出, 不管 V 的元素如何, 当 K 为实数域 $\mathbb{R}$ 时, 则称 V 为实线性空间; 当 K 为复数域 $\mathbb{C}$ 时, 就称 V 为复线性空间.

例 2.1　设 $\mathbb{R}^+$ 为所有正实数组成的数集, 其加法与数乘运算分别定义为

$$m \oplus n = mn, \quad k \circ m = m^k$$

证明 $\mathbb{R}^+$ 是 $\mathbb{R}$ 是上的线性空间.

证　设 $m, n \in \mathbb{R}^+$,　$k \in R$，则有

$$m \oplus n = mn \in \mathbb{R}^+, \quad k \circ m = m^k \in \mathbb{R}^+$$

即 $\mathbb{R}^+$ 对所定义的加法运算“$\oplus$”与数乘运算“$\circ$”是封闭的，且有

(1) $(m \oplus n) \oplus p = (mn) \oplus p = mnp = m \oplus (np) = m \oplus (n \oplus p)$

(2) $m \oplus n = mn = nm = n \oplus m$

(3) 1 是零元素，因为 $m \oplus 1 = m \times 1 = m$

(4) m 的负元素是 $\dfrac{1}{m}$，因为 $m \oplus \dfrac{1}{m} = m \oplus \dfrac{1}{m} = 1$

(5) $k \circ (m \oplus n) = k \circ (mn) = (mn)^k = m^k n^k = (k \circ m) \oplus (k \circ n)$

(6) $(k + l) \circ m = m^{k+l} = m^k m^l = (k \circ m) \oplus (l \circ m)$

(7) $k \circ (l \circ m) = k \circ m^l = m^{lk} = m^{kl} = (kl) \circ m$

(8) $1 \circ m = m^1 = m$

成立, 故 $\mathbb{R}^+$ 是实线性空间.

定义 2.2　设 V 是数域 K 上的线性空间, $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r (r \geqslant 1)$ 是属于 V 的任意 r 个向量, 如果它满足

(1) $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r$ 线性无关;

(2) V 中任一向量 x 都是 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r$ 的线性组合.

则称 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r$ 为 V 的一个基或基底, 并称 $\boldsymbol{x}_i (i = 1, 2, \cdots, r)$ 为基向量.

且由上述定义可见, 线性空间的维数是其基中所含向量的个数.

同时, 一个线性空间的基不是唯一的. 例如, n 维向量组

$$\begin{cases} \boldsymbol{e}_1 = (1, 0, \cdots, 0) \\ \boldsymbol{e}_2 = (0, 1, \cdots, 0) \\ \cdots \\ \boldsymbol{e}_n = (0, 0, \cdots, 1) \end{cases} \quad \begin{cases} \boldsymbol{y}_1 = (1, 1, \cdots, 1, 1) \\ \boldsymbol{y}_2 = (0, 1, \cdots, 1, 1) \\ \cdots \\ \boldsymbol{y}_n = (0, 0, \cdots, 0, 1) \end{cases} \tag{2.1.1}$$

都实线性空间 $\boldsymbol{R}^n$ 的基. 这是因为

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

从而它们各自都线性无关, 而且对于任一向量 $\boldsymbol{x} = (\xi_1, \xi_2, \cdots, \xi_n) \in R^n$, 分别有

$$\boldsymbol{x} = \xi_1\boldsymbol{e}_1 + \xi_2\boldsymbol{e}_2 + \cdots + \xi_n\boldsymbol{e}_n$$
$$\boldsymbol{x} = \xi_1\boldsymbol{y}_1 + (\xi_2 - \xi_1)\boldsymbol{y}_2 + \cdots + (\xi_n - \xi_{n-1})\boldsymbol{y}_n$$

定义 2.3　称线性空间 V^n 的一个基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r$ 为 V^n 的一个坐标系. 设向量 $\boldsymbol{x} \in V^n$, 它在该基下的线性表示式为

$$\boldsymbol{x} = \xi_1\boldsymbol{x}_1 + \xi_2\boldsymbol{x}_2 + \cdots + \xi_n\boldsymbol{x}_n$$

则称 $\xi_1, \xi_2, \cdots, \xi_n$ 为 $\boldsymbol{x}$ 在该坐标系中的坐标或分量, 记为

$$(\xi_1, \xi_2, \cdots, \xi_n)^T$$

需要指出, 在不同的坐标系 (或基) 中, 同一向量的坐标一般是不同的.

欧氏空间

欧式空间是一种特殊的线性空间, 定义如下.

定义 2.4　设 V 是实数域 $\mathbb{R}$ 上的线性空间, 对于 V 中任意两个向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$, 按照某种规则定义一个实数, 用 $(\boldsymbol{x}, \boldsymbol{y})$ 来表示, 且它满足下述 4 个条件:

(1) 交换律: $(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{y}, \boldsymbol{x})$;

(2) 分配律: $(\boldsymbol{x}, \boldsymbol{y} + z) = (\boldsymbol{x}, \boldsymbol{y}) + (\boldsymbol{x}, z)$;

(3) 齐次性: $(k\boldsymbol{x}, \boldsymbol{y}) = k(\boldsymbol{x}, \boldsymbol{y})(\forall k \in \mathbb{R})$;

(4) 非负性: $(\boldsymbol{x}, \boldsymbol{x}) \geqslant 0$, 当且仅当 $\boldsymbol{x} = \boldsymbol{0}$ 时, $(\boldsymbol{x}, \boldsymbol{x}) = 0$.

则称实数 $(\boldsymbol{x}, \boldsymbol{y})$ 为向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的内积, 而称 V 为 Euclid 空间, 简称欧氏空间或实内积空间.

显然, 欧氏空间是定义了内积的实线性空间. 因此, 又有内积空间之称, 可见, 欧氏空间是一个特殊的实线性空间.

因为向量的内积与向量的线性运算是彼此无关的运算, 所以不论内积如何规定, 都不会影响该实线性空间的维数. 欧氏空间的子空间显然也是欧氏空间.

线性变换

线性空间是某类客观事物从量的方面的一个抽象, 而线性变换则研究线性空间中元素之间的最基本联系.

定义 2.5　设 V 是数域 K 上的线性空间, T 是 V 到自身的一个映射, 使对任意向量 $\boldsymbol{x} \in V$, V 中都有唯一的向量 $\boldsymbol{y}$ 与之对应, 则称 T 是 V 的一个变换或算子, 记为 $T\boldsymbol{x} = \boldsymbol{y}$, 称 $\boldsymbol{y}$ 为 $\boldsymbol{x}$ 在 T 下的象, 而 $\boldsymbol{x}$ 是 $\boldsymbol{y}$ 的原象 (或象源).

例如, 平面上所有起点在原点的向量的集合, 形成实二维线性空间 R^2. 在 R^2 中绕原点的旋转就是 R^2 的一个变换.

定义 2.6　如果数域 K 上的线性空间 V 的一个变换 T 具有以下性质:

$$T(k\boldsymbol{x} + l\boldsymbol{y}) = k(T\boldsymbol{x}) + l(T\boldsymbol{y})$$

其中 $\boldsymbol{x}, \boldsymbol{y} \in V, k, l \in K$. 则称 T 为 V 的一个线性变换或线性算子.

例 2.2　把线性空间 R^2 的所有向量均绕原点依顺 (或逆) 时针方向旋转 θ 角的变换, 就是一个线性变换. 这时象 (η_1, η_2) 与原象 (ξ_1, ξ_2) 之间的关系为

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}$$

例 2.3　在线性空间 P_n 中, 求微分是其一个线性变换, 这里用 D 表示, 即

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

事实上, 对任意的 $f(t), g(t) \in P_n$ 及 $k, l \in \mathbb{R}$, 有

$$D(kf(t) + lg(t)) = (kf(t) + lg(t))' = kf'(t) + lg'(t) = k(Df(t)) + l(Dg(t))$$

不过需要注意, 线性变换可能把线性无关的向量组变为线性相关的向量组. 如零变换 T_0 就是这样.

定义 2.6　设 T 是线性空间 V 的线性变换, V 中所有向量的象形成的集合, 称为 T 的值域, 用 $R(T)$ 表示, 即

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

V 中所有被 T 变为零向量的原象构成的集合, 称为 T 的核, 用 $N(T)$ 表示, 即

$$N(T) = \{\boldsymbol{x} \mid T\boldsymbol{x} = \boldsymbol{0}, \boldsymbol{x} \in V\}$$

线性变换的运算

下面, 为了讨论线性变换的运算, 引入单位变换的零变换的概念.

把线性空间 V 的任一向量都变为其自身的变换是一个线性变换, 称为单位变换或恒等变换, 记为 T_e, 于是有

$$T_e\boldsymbol{x} = \boldsymbol{0} \quad (\forall \boldsymbol{x} \in V)$$

如果 T_1, T_2 是 V 的两个变换, 且对于任意向量 $\boldsymbol{x} \in V$, 都有 $T_1\boldsymbol{x} = T_2\boldsymbol{x}$, 那么就称 T_1 与 T_2 相等, 记为

$$T_1 = T_2$$

对于线性空间的线性变换, 下面定义它们的集中运算方式.

1. 加法

设 T_1, T_2 是线性空间 V 的两个线性变换, 定义它们的和 T_1+T_2 为

$$(T_1+T_2)\boldsymbol{x}=T_1\boldsymbol{x}+T_2\boldsymbol{x} \quad (\forall \boldsymbol{x}\in V)$$

下面证明, 线性变换 T_1 与 T_2 的和 T_1+T_2 仍是 V 的线性变换. 事实上, 对任意 $\boldsymbol{x},\boldsymbol{y}\in V, k,l\in K$, 由定义 2.6 有

$$\begin{aligned}(T_1+T_2)(k\boldsymbol{x}+l\boldsymbol{y}) =&T_1(k\boldsymbol{x}+l\boldsymbol{y})+T_2(k\boldsymbol{x}+l\boldsymbol{y})=\\ &k(T_1\boldsymbol{x})+l(T_1\boldsymbol{y})+k(T_2\boldsymbol{x})+l(T_2\boldsymbol{y})=\\ &k(T_1\boldsymbol{x}+T_2\boldsymbol{x})+l(T_1\boldsymbol{y}+T_2\boldsymbol{y})=\\ &k(T_1+T_2)\boldsymbol{x}+l(T_1+T_2)\boldsymbol{y}\end{aligned}$$

这就表明 T_1+T_2 是 V 的线性变换.

另外, 定义线性变换 T 的负变换 $-T$ 定义为

$$(-T)\boldsymbol{x}=-(T\boldsymbol{x}) \quad (\forall \boldsymbol{x}\in V)$$

因此, 线性变换的加法具有以下基本性质:

(1) $T_1+T_2=T_2+T_1$;

(2) $(T_1+T_2)+T_3=T_1+(T_2+T_3)$;

(3) $T+T_0=T$;

(4) $T+(-T)=T_0$.

2. 线性变换与数的乘法

设 $k\in K$, T 为线性空间 V 中的线性变换, 定义数 k 与 T 的乘积 kT 为

$$(kT)\boldsymbol{x}=k(T\boldsymbol{x}) \quad (\forall \boldsymbol{x}\in V)$$

因此, 线性变换的数乘具有以下基本性质:

(1) $k(T_1+T_2)=kT_1+k$;

(2) $(k+l)T=kT+lT$;

(3) $(kl)T=k(lT)$;

(4) $1\ T=T$.

3. 线性变换的乘法

设 T_1, T_2 是线性空间 V 的两个线性变换, 定义 T_1 与 T_2 的乘积 T_1T_2 为

$$(T_1T_2)\boldsymbol{x}=T_1(T_2)\boldsymbol{x} \quad (\forall \boldsymbol{x}\in V)$$

即 T_1T_2 是先施行 T_2, 然后施行 T_1 的变换. 且

$$\begin{aligned}(T_1T_2)(k\boldsymbol{x}+l\boldsymbol{y}) =&T_1(T_2(k\boldsymbol{x}+l\boldsymbol{y})) = \\ &k(T_1(T_2\boldsymbol{x}))+l(T_1(T_2\boldsymbol{y})) = \\ &k(T_1T_2)\boldsymbol{x}+l(T_1T_2)\boldsymbol{y}\end{aligned}$$

因此, 线性变换的乘积具有以下基本性质:

(1) $(T_1T_2)T_3 = T_1(T_2T_3)$;

(2) $T_eT = TT_e = T$;

(3) 一般地, $T_1T_2 \neq T_2T_1$;

(4) $T_1(T_2+T_3) = T_1T_2+T_1T_3$;

(5) $(T_1+T_2)T_3 = T_1T_3+T_2T_3$.

4. 逆变换

同逆矩阵的概念类似, 若 T 是 V 的线性变换, 且存在线性变换 S, 使得

$$(ST)\boldsymbol{x} = (TS)\boldsymbol{x} = \boldsymbol{x} \quad (\forall \boldsymbol{x} \in V)$$

则称 S 是 T 的逆变换, 记为 $S = T^{-1}$. 且有

$$T^{-1}T = TT^{-1} = T_e$$

5. 线性变换的多项式

设 n 是正整数, T 是线性空间 V 的线性变换. 定义 T 的 n 次幂为

$$T^n = T^{n-1}T \quad (n = 2, 3, \cdots)$$

定义 T 的零次幂为

$$T^0 = T_e$$

设 $f(T) = a_0t^m + a_1t^{m-1} + \cdots + a_{m-1}t + a_m$ 是纯量 t 的 m 次多项式, T 是 V 的一个线性变换, 则由线性变换的运算可知

$$f(T) = a_0T^m + a_1T^{m-1} + \cdots + a_{m-1}T + a_mT_e$$

也是 V 的一个线性变换或线性算子, 称其为线性变换 T 的多项式.

线性变换的矩阵

有限维线性空间的向量可以用坐标表示, 更进一步, 这里将通过坐标把线性变换用矩阵表示出来, 从而可以把比较抽象的线性变换转化为具体的矩阵来处理.

定义 2.7 设 $x_1, x_2, \cdots, x_n$ 为数域 P 上线性空间 V 的一组基, T 为 V 的线性变换. 基向量的象可以被基线性表出, 设

$$\begin{cases} T(x_1) = a_{11}\boldsymbol{x}_1 + a_{21}\boldsymbol{x}_2 + \cdots + a_{n1}\boldsymbol{x}_n \\ T(x_1) = a_{11}\boldsymbol{x}_1 + a_{21}\boldsymbol{x}_2 + \cdots + a_{n1}\boldsymbol{x}_n \\ \cdots \\ T(x_1) = a_{11}\boldsymbol{x}_1 + a_{21}\boldsymbol{x}_2 + \cdots + a_{n1}\boldsymbol{x}_n \end{cases}$$

用矩阵表示即为

$$T(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n) = (T\boldsymbol{x}_1, T\boldsymbol{x}_2, \cdots, T\boldsymbol{x}_n) = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)\boldsymbol{A}$$

其中, $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$, 矩阵 $\boldsymbol{A}$ 称为线性变换 T 在基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 下的矩阵.

需要注意

(1) 给定 V^n 的基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 和线性变换 T, 矩阵 A 是唯一的;

(2) 单位变换在任意一组基下的矩阵皆为单位矩阵; 零变换在任意一组基下的矩阵皆为零矩阵; 数乘变换在任意一组基下的矩阵皆为数量矩阵.

例 2.4 在矩阵空间 $R^{2\times2}$ 中, 给定矩阵

$$\boldsymbol{B} = \begin{bmatrix} 0 & 1 \\ 4 & 0 \end{bmatrix}$$

线性变换 $T(\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{B}(\forall \boldsymbol{X} \in R^{2\times2})$, $R^{2\times2}$ 的两个基为

(1): $\boldsymbol{E}_{11}, \boldsymbol{E}_{12}, \boldsymbol{E}_{21}, \boldsymbol{E}_{22}$;

(2): $\boldsymbol{B}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \boldsymbol{B}_2 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \boldsymbol{B}_3 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \boldsymbol{B}_4 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

分别求 T 在这两个基下的矩阵.

解 计算 (1) 的基象组, 有

$$T(\boldsymbol{E}_{11}) = \boldsymbol{E}_{11}\boldsymbol{B} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = 0\boldsymbol{E}_{11} + 1\boldsymbol{E}_{12} + 0\boldsymbol{E}_{21} + 0\boldsymbol{E}_{22}$$

$$T(\boldsymbol{E}_{12}) = \boldsymbol{E}_{12}\boldsymbol{B} = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} = 4\boldsymbol{E}_{11} + 0\boldsymbol{E}_{12} + 0\boldsymbol{E}_{21} + 0\boldsymbol{E}_{22}$$

$$T(\boldsymbol{E}_{21}) = \boldsymbol{E}_{21}\boldsymbol{B} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = 0\boldsymbol{E}_{11} + 0\boldsymbol{E}_{12} + 0\boldsymbol{E}_{21} + 1\boldsymbol{E}_{22}$$

$$T(\boldsymbol{E}_{22}) = \boldsymbol{E}_{22}\boldsymbol{B} = \begin{bmatrix} 0 & 0 \\ 4 & 0 \end{bmatrix} = 0\boldsymbol{E}_{11} + 0\boldsymbol{E}_{12} + 4\boldsymbol{E}_{21} + 0\boldsymbol{E}_{22}$$

故 T 在基 (1) 下的矩阵为

$$\boldsymbol{A}_1 = \begin{bmatrix} 0 & 4 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

计算 (2) 的基象组, 有

$$T(\boldsymbol{B}_1) = \boldsymbol{B}_1\boldsymbol{B} = \begin{bmatrix} 4 & 1 \\ 4 & 1 \end{bmatrix} = 1\boldsymbol{B}_1 + 3\boldsymbol{B}_2 - 3\boldsymbol{B}_3 + 3\boldsymbol{B}_4$$

$$T(\boldsymbol{B}_2) = \boldsymbol{B}_2\boldsymbol{B} = \begin{bmatrix} 4 & 1 \\ 0 & 1 \end{bmatrix} = 1\boldsymbol{B}_1 - 1\boldsymbol{B}_2 + 1\boldsymbol{B}_3 + 3\boldsymbol{B}_4$$

$$T(\boldsymbol{B}_3) = \boldsymbol{B}_3\boldsymbol{B} = \begin{bmatrix} 4 & 1 \\ 0 & 0 \end{bmatrix} = 0\boldsymbol{B}_1 + 0\boldsymbol{B}_2 + 1\boldsymbol{B}_3 + 3\boldsymbol{B}_4$$

$$T(\boldsymbol{B}_4) = \boldsymbol{B}_4\boldsymbol{B} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = 0\boldsymbol{B}_1 + 0\boldsymbol{B}_2 + 1\boldsymbol{B}_3 - 1\boldsymbol{B}_4$$

故 T 在基 (2) 下的矩阵为

$$\boldsymbol{A}_2 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 3 & -1 & 0 & 0 \\ -3 & 1 & 1 & 1 \\ 3 & 3 & 3 & -1 \end{bmatrix}$$

显然 $\boldsymbol{A}_1 \neq \boldsymbol{A}_2$.

定理 2.1　设 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 是数域 K 上的线性空间 V^n 的一个基, 线性变换 T_1, T_2 在该基下的矩阵依次是 $\boldsymbol{A}, \boldsymbol{B}$. 则有如下结论:

(1) $(T_1 + T_2)(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n) = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)(\boldsymbol{A} + \boldsymbol{B})$;

(2) $(kT_1)(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n) = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)(k\boldsymbol{A})$;

(3) $(T_1T_2)(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n) = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)\boldsymbol{A}\boldsymbol{B}$;

(4) $T_1^{-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n) = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)\boldsymbol{A}^{-1}$.

定理 2.2　设线性变换 T 在线性空间 V^n 的基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 下的矩阵是 $\boldsymbol{A}$, 向量 $\boldsymbol{x}$ 在该基下的坐标是 $\boldsymbol{\alpha}$, 则 $T\boldsymbol{x}$ 在该基下的坐标是

$$\boldsymbol{\beta} = \boldsymbol{A\alpha} \tag{2.1.2}$$

利用 (2.1.2) 式, 可以直接由线性变换的矩阵 $\boldsymbol{A}$, 来计算一个向量 $\boldsymbol{x}$ 的象 $T\boldsymbol{x}$ 的坐标.

为了利用矩阵研究线性变换, 有必要弄清楚线性变换的矩阵是怎么样随基的改变而改变的, 从而建立矩阵相似的概念.

定理 2.3　设线性空间 V^n 的线性变换为 T, T 在 V^n 的两个基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 和 $\boldsymbol{y}_1$, $\boldsymbol{y}_2, \cdots, \boldsymbol{y}_n$ 下的矩阵依次是 $\boldsymbol{A}$ 和 $\boldsymbol{B}$, 并且

$$(\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n) = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)\boldsymbol{C}$$

那么

$$\boldsymbol{B} = \boldsymbol{C}^{-1}\boldsymbol{AC} \tag{2.1.3}$$

式 (2.1.3) 给出的两个矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 之间的关系, 在矩阵论中将起极其重要的作用. 引入如下定义.

定义 2.8　设 $\boldsymbol{A}, \boldsymbol{B}$ 为数域 K 上的两个 n 阶矩阵, 如果存在 K 上的 n 阶可逆矩阵 $\boldsymbol{P}$, 使得 $\boldsymbol{B} = \boldsymbol{P}^{-1}\boldsymbol{AP}$, 则称 $\boldsymbol{A}$ 相似于 $\boldsymbol{B}$, 记为 $\boldsymbol{A} \sim \boldsymbol{B}$.

那么, 相似矩阵有如下基本性质:

反身性: $\boldsymbol{A} \sim \boldsymbol{A}$.

对称性: 如果 $\boldsymbol{A} \sim \boldsymbol{B}$, 那么 $\boldsymbol{B} \sim \boldsymbol{A}$.

传递性: 如果 $\boldsymbol{A} \sim \boldsymbol{B}, \boldsymbol{B} \sim \boldsymbol{C}$, 那么 $\boldsymbol{A} \sim \boldsymbol{C}$.

特征值与特征向量

现在讨论如何选择线性空间的基, 使线性变换的该基下的矩阵形状最简单的问题. 为此, 先论述线性变换的特征值和特征向量的概念. 它们对于线性变换的研究, 起着十分重要的作用.

定义 2.9　设 T 是数域 K 上的线性空间 V^n 的线性变换, 且对 K 中某一数 λ_0, 存在非零向量 $\boldsymbol{x} \in V^n$, 使得

$$T\boldsymbol{x} = \lambda_0\boldsymbol{x} \tag{2.1.4}$$

成立, 则称 λ_0 为 T 的特征值, $\boldsymbol{x}$ 为 T 的属于 λ_0 的特征向量.

式 (2.1.4) 表明, 在几何上, 特征向量 $\boldsymbol{x}$ 的方位, 经过线性变换后保持不变.

定义 2.10　设 $\boldsymbol{A}=(a_{ij})_{n\times n}$ 是数域 K 上的 n 阶矩阵, λ 是参数, $\boldsymbol{A}$ 的特征矩阵 $\lambda\boldsymbol{I}-\boldsymbol{A}$ 的行列式

$$\det(\lambda\boldsymbol{I}-\boldsymbol{A})=\begin{vmatrix}\lambda-a_{11} & -a_{12} & \cdots & -a_{1n}\\ -a_{21} & \lambda-a_{22} & \cdots & -a_{1n}\\ \vdots & \vdots & & \vdots\\ -a_{n1} & -a_{n2} & \cdots & \lambda-a_{nn}\end{vmatrix}$$

称为矩阵 $\boldsymbol{A}$ 的特征多项式, 它是 K 上的一个 n 次多项式, 记为 $\varphi(\lambda)$. $\varphi(\lambda)$ 的根 (或零点) λ_0 称为 $\boldsymbol{A}$ 的特征值 (根); 而相对应与方程组

$$(\lambda_0\boldsymbol{I}-\boldsymbol{A})\begin{bmatrix}\xi_1\\ \xi_2\\ \vdots\\ \xi_n\end{bmatrix}=\boldsymbol{0} \tag{2.1.5}$$

的非零解向量 $(\xi_1,\xi_2,\cdots,\xi_n)^T$ 称为 $\boldsymbol{A}$ 的属于特征值 λ_0 的特征向量.

那么, 计算特征值和特征向量的步骤如下:

(1) 取定数域 K 上的线性空间 V^n 的一个基, 写出线性变换 T 在该基下的矩阵 $\boldsymbol{A}$;

(2) 求出 $\boldsymbol{A}$ 的特征多项式 $\varphi(\lambda)$ 在数域 K 上的全部根, 他们就是 T 的全部特征值;

(3) 把求得的特征值逐个带入方程组 (2.1.5) 中, 解出矩阵 $\boldsymbol{A}$ 属于每个特征值的全部线性无关的特征向量;

(4) 以 $\boldsymbol{A}$ 的属于每个特征值的特征向量为 V^n 中取定基下的坐标, 即得 T 的相应特征向量.

由行列式的展开法则可得 n 阶矩阵 $\boldsymbol{A}=(a_{ij})_{n\times n}$ 的特征多项式

$$\begin{aligned}\varphi(\lambda)&=\det(\lambda\boldsymbol{I}-\boldsymbol{A})=\\ &\lambda^n-(a_{11}+a_{22}+\cdots+a_{nn})\lambda^{n-1}+\cdots+(-1)^n\det\boldsymbol{A}\end{aligned}$$

如果 $\boldsymbol{A}$ 有 n 个特征值 $\lambda_1,\lambda_2,\cdots,\lambda_n$, 则由上式可知

$$\sum_{i=1}^n\lambda_i=\sum_{i=1}^n a_{ii},\quad \lambda_1\lambda_2\cdots\lambda_n=\det\boldsymbol{A} \tag{2.1.6}$$

引入记号

$$\mathrm{tr}\boldsymbol{A}=\sum_{i=1}^n a_{ii}$$

称为矩阵 $\boldsymbol{A}$ 的迹或追迹. 式 (2.1.6) 表明, 矩阵 $\boldsymbol{A}$ 的所有特征值的和等于 $\boldsymbol{A}$ 的迹, 而 $\boldsymbol{A}$ 的全体特征值的乘积等于 $\det\boldsymbol{A}$.

定理 2.4(Hamilton-Cayley)　n 阶矩阵 $\boldsymbol{A}$ 是其特征多项式的矩阵根 (零点), 即令

$$\varphi(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n$$

则有

$$\varphi(\boldsymbol{A}) = \boldsymbol{A}^n + a_1\boldsymbol{A}^{n-1} + \cdots + a_{n-1}\boldsymbol{A} + a_n\boldsymbol{I} + \boldsymbol{O} \tag{2.1.7}$$

当 n 阶矩阵 $\boldsymbol{A}$ 可逆时, 它的特征多项式中的常数项 $a_n \neq 0$. 由 (2.1.7) 可得

$$\boldsymbol{A}^{-1} = -\frac{1}{a_n}(\boldsymbol{A}^{n-1} + a_1\boldsymbol{A}^{n-1} + \cdots + a_{n-2}\boldsymbol{A} + a_{n-1}\boldsymbol{I})$$

即 $\boldsymbol{A}$ 的逆矩阵能够由它的 $n-1$ 次矩阵多项式表示. 此外, 无论 $\boldsymbol{A}$ 是否可逆, 它的 n 次幂也能够由它的次数不超过 $n-1$ 的矩阵多项式表示. 根据后一结论, 可以简化矩阵多项式的计算问题.

例 2.5　计算矩阵多项式 $\boldsymbol{A}^{100} + 2\boldsymbol{A}^{50}$, 其中

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

解　令 $\psi(\lambda) = \lambda^{100} + 2\lambda^{50}$, 可求得 $\boldsymbol{A}$ 的特征多项式为

$$\varphi(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A}) = (\lambda - 1)^2(\lambda - 2)$$

用 $\varphi(\lambda)$ 除 $\phi(\lambda)$, 可得

$$\psi(\lambda) = \varphi(\lambda)q(\lambda) + b_0 + b_1\lambda + b_2\lambda^2$$

将 $\lambda = 1, 2$ 分别带入上式, 则有

$$\begin{cases} b_0 + b_1 + b_2 = 3b_0 + 2b_1 + 4b_2 = 2^{100} + 2^{51} \end{cases}$$

对 $\psi(\lambda)$ 关于 λ 求导, 得到

$$\psi'(\lambda) = [2(\lambda - 1)(\lambda - 2) + (\lambda - 1)^2]q(\lambda) + \varphi(\lambda)q'(\lambda) + b_1 + 2b_2\lambda$$

将 $\lambda = 1$ 带入上式, 可得

$$b_1 + 2b_2 = 200$$

从而求得

$$b_0 = 2^{100} + 2^[51] - 400, \quad b_1 = 606 - 2^{101} - 2^{52}, \quad b_2 = -203 + 2^{100} + 2^{51}$$

故

$$\boldsymbol{A}^{100} + 2\boldsymbol{A}^{50} = \psi(\boldsymbol{A}) = b_0\boldsymbol{I} + b_1\boldsymbol{A} + b_2\boldsymbol{A}^2$$

以矩阵 $\boldsymbol{A}$ 为根的多项式有时是很多的, 但是它们之间却有一定的关系. 引入如下定义.

定义 2.11　首项系数是 1(简称首 1), 次数最小, 且以矩阵 $\boldsymbol{A}$ 为根的 λ 的多项式, 称为 $\boldsymbol{A}$ 的最小多项式, 常用 $m(\lambda)$ 表示.

根据定理 2.4, 显然 $\boldsymbol{A}$ 的最小多项式 $m(\lambda)$ 的次数不大于它的特征多项式 $\varphi(\lambda)$ 的次数.

例 2.6　求矩阵 $\boldsymbol{A} = \begin{bmatrix} 3 & -3 & 2 \\ -1 & 5 & -2 \\ -1 & 3 & 0 \end{bmatrix}$ 的最小多项式.

解　设 $f(\lambda) = \lambda + k (k \in \mathbb{R})$, 由于 $f(\boldsymbol{A}) = \boldsymbol{A} + k\boldsymbol{I} \neq \boldsymbol{O}$, 所以任何一次多项式都不是 $\boldsymbol{A}$ 的最小多项式. 注意到 $\boldsymbol{A}$ 的特征多项式

$$\varphi(\lambda) = (\lambda - 2)^2(\lambda - 4)$$

且对于它的二次因式

$$\psi(\lambda) = (\lambda - 2)(\lambda - 4) = \lambda^2 - 6\lambda + 8$$

有

$$\psi(\boldsymbol{A}) = \boldsymbol{A}^2 - 6\boldsymbol{A} + 8\boldsymbol{I} + \boldsymbol{O}$$

于是由定义 2.11, 有 $m(\lambda) = \psi(\lambda)$.

这就是说, $\boldsymbol{A}$ 的最小多项式是其特征多项式的因式. 这个事实具有一般性, 引出如下定理。

定理 2.5　矩阵 $\boldsymbol{A}$ 的最小多项式 $m(\lambda)$ 可以整除以 $\boldsymbol{A}$ 为根的任意首 1 多项式 $\psi(\lambda)$, 且 $m(\lambda)$ 是唯一的.

证　假若 $m(\lambda)$ 不能整除 $\psi(\lambda)$, 则有

$$\psi(\lambda) = m(\lambda)q(\lambda) + r(\lambda)$$

其中 $r(\lambda)$ 的次数小于 $m(\lambda)$ 的次数. 于是由

$$\psi(\boldsymbol{A}) = m(\boldsymbol{A})q(\boldsymbol{A}) + r(\boldsymbol{A})$$

知 $r(\boldsymbol{A}) = \boldsymbol{O}$, 这就与 $m(\lambda)$ 是 $\boldsymbol{A}$ 的最小多项式相矛盾.

定理 2.6　矩阵 $\boldsymbol{A}$ 的最小多项式 $m(\lambda)$ 与其特征多项式 $\varphi(\lambda)$ 的零点相同 (不计重数).

证　由定理 2.4 知 $\varphi(\boldsymbol{A}) = \boldsymbol{O}$, 再由定理 2.5 知 $m(\lambda)$ 能够整除 $\varphi(\lambda)$, 所以 $m(\lambda)$ 的零点是 $\varphi(\lambda)$ 的零点.

定理 2.7　设 n 阶矩阵 $\boldsymbol{A}$ 的特征多项式为 $\varphi(\lambda)$, 特征矩阵 $\lambda\boldsymbol{I}-\boldsymbol{A}$ 的全体 $n-1$ 阶子式的最大公因式为 $d(\lambda)$, 则 $\boldsymbol{A}$ 的最小多项式为

$$m(\lambda)=\frac{\varphi(\lambda)}{d(\lambda)}$$

需要指出, 最小多项式相同是矩阵相似的必要条件, 并非充分条件.

对角矩阵

对角矩阵是较简单的矩阵之一. 无论是计算它的乘积、逆矩阵还是特征值等, 都甚为方便. 下面讨论哪些线性变换在适当的基下的矩阵是对角矩阵的问题.

定理 2.8　设 T 是线性空间 V^n 的线性变换, T 在某一基下的矩阵 $\boldsymbol{A}$ 可以为对角矩阵的充要条件是 T 有 n 个线性无关的特征向量.

证　设 T 在 V^n 的基 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 下的矩阵是对角矩阵

$$\boldsymbol{A}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$$

这就意味着有

$$T\boldsymbol{x}_i=\lambda_i\boldsymbol{x}_i\quad(i=1,2,\cdots,n)$$

因而 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 就是 T 的 n 个线性无关的特征向量.

反之, 如果 T 有 n 个线性无关的特征向量 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$, 即有

$$T\boldsymbol{x}_i=\lambda_i\boldsymbol{x}_i\quad(i=1,2,\cdots,n)$$

那么就取 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 为 V^n 的基, 于是在这个基下 T 的矩阵是对角矩阵.

定理 2.9　n 阶矩阵 $\boldsymbol{A}$ 与对角矩阵相似的充要条件是, $\boldsymbol{A}$ 有 n 个线性无关的特征向量, 或 $\boldsymbol{A}$ 有完备的特征向量系.

2.1.2　若尔当标准型的求解

前面指出, 一切 n 阶矩阵 $\boldsymbol{A}$ 可以分解成许多相似类. 现在需要在与 $\boldsymbol{A}$ 相似的全体矩阵中, 找出一个较简单的矩阵来作为这个相似类的标准形. 当然以对角矩阵作为标准形最好, 可惜不是每一个矩阵都能与对角矩阵相似.

定理 2.10　设 T 是复数域 $\mathbb{C}$ 上的线性空间 V^n 的线性变换, 任取 V^n 的一个基, T 在该基下的矩阵是 $\boldsymbol{A}$, T(或 $\boldsymbol{A}$) 的特征多项式可分解因式为

$$\varphi(\lambda) = (\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_s)^{m_s} \quad (m_1 + m_2 + \cdots + m_n = n) \tag{2.1.8}$$

则 V^n 可分解成不变子空间的直和

$$V^n = V_1 \oplus V_2 \oplus \cdots \oplus V_s$$

其中 $V_i = \{\boldsymbol{x} \mid (T - \lambda_i T_e)^{m_i}\boldsymbol{x} = \boldsymbol{0}, \boldsymbol{x} \in V^n\}$ 是线性变换 $(T - \lambda_i T_e)^{m_i}$ 的核子空间.

如果给每个子空间 V_i 选一适当的基, 每个子空间的基合并起来即为 V^n 的基, 且 T 在该基下的矩阵为以下形式的准对角矩阵

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{J}_1(\lambda_1) & & & \\ & \boldsymbol{J}_2(\lambda_2) & & \\ & & \ddots & \\ & & & \boldsymbol{J}_s(\lambda_s) \end{bmatrix} \tag{2.1.9}$$

其中

$$\boldsymbol{J}_i(\lambda_i) = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \lambda_i & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix}_{m_i \times m_i} \qquad (i = 1, 2, \cdots, s)$$

定义 2.12　由式 (2.1.9) 给出的矩阵 $\boldsymbol{J}$ 称为矩阵 $\boldsymbol{A}$ 的 Jordan 标准形, $\boldsymbol{J}_i(\lambda_i)$ 称为因式 $(\lambda - \lambda_i)^{m_i}$ 对应的 Jordan 块.

由相似矩阵的的定义可得到如下定理.

定理 2.11　设 $\boldsymbol{A}$ 是 n 阶复矩阵, 且其特征多项式的某种分解式是 (2.1.8), 则存在 n 阶复可逆矩阵 $\boldsymbol{P}$, 使

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J}$$

即, 设 T 是复数域 $\mathbb{C}$ 上线性空间 V^n 的线性变换, 在 V^n 中必存在一个基, 使 T 在该基下的矩阵是 Jordan 标准形 $\boldsymbol{J}$.

虽然上述定理已经肯定了一般矩阵的 Jordan 标准形是存在的, 但是仍旧无法准确地求出矩阵的 Jordan 标准形. 而讨论矩阵的 Jordan 标准形的求法, 涉及以下形式的多项式矩阵或 $\lambda-$ 矩阵

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

的理论, 其中 $aij(\lambda)(i,j=1,2,\cdots,n)$ 为数域 K 上的纯量 λ 的多项式. 如果 $\boldsymbol{A}=(a_{ij})_{n\times n}$ 是数域 K 上的 n 阶矩阵, 则 $\boldsymbol{A}$ 的特征矩阵

$$\lambda\boldsymbol{I}-\boldsymbol{A}=\begin{bmatrix}\lambda-a_{11} & -a_{12} & \cdots & -a_{1n}\\ -a_{21} & \lambda-a_{22} & \cdots & -a_{2n}\\ \vdots & \vdots & & \vdots\\ -a_{n1} & -a_{n2} & \cdots & \lambda-a_{nn}\end{bmatrix}$$

就是一个特殊的多项式矩阵.

多项式矩阵 $\boldsymbol{A}(\lambda)$ 的标准形, 是指使用矩阵的初等变换将 $\boldsymbol{A}(\lambda)$ 化为多项式矩阵, 有

$$\boldsymbol{A}(\lambda)\rightarrow\begin{bmatrix}d_1(\lambda) & & & & & & \\ & d_2(\lambda) & & & & & \\ & & \ddots & & & & \\ & & & d_s(\lambda) & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0\end{bmatrix}$$

其中

$$d_1(\lambda)\mid d_2(\lambda),d_2(\lambda)\mid d_3(\lambda),\cdots d_{s-1}(\lambda)\mid d_s(\lambda)\quad(s\geqslant n)$$

且 $d_i(\lambda)(i=1,2,\cdots,s)$ 是首 1 多项式.

例 2.7　用初等变化化多项式矩阵

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

为标准形.

解

$$\boldsymbol{A}(\lambda)\rightarrow\begin{bmatrix}-\lambda+1 & 2\lambda-1 & 1\\ \lambda & \lambda^2 & 0\\ \lambda^2+1 & \lambda^2+\lambda-1 & 1\end{bmatrix}\rightarrow$$

$$\begin{bmatrix}1 & 2\lambda-1 & -\lambda+1\\ 0 & \lambda^2 & \lambda\\ 1 & \lambda^2+\lambda-1 & \lambda^2+1\end{bmatrix}\rightarrow$$

$$\begin{bmatrix}1 & 2\lambda-1 & -\lambda+1\\0 & \lambda^2 & \lambda\\0 & \lambda^2-\lambda & \lambda^2+\lambda\end{bmatrix}\rightarrow$$

$$\begin{bmatrix}1 & 0 & 0\\0 & \lambda^2 & \lambda\\0 & \lambda^2-\lambda & \lambda^2+\lambda\end{bmatrix}\rightarrow\begin{bmatrix}1 & 0 & 0\\0 & \lambda & \lambda^2\\0 & \lambda^2+\lambda & \lambda^2-\lambda\end{bmatrix}\rightarrow$$

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

最后所得的矩阵是 $\boldsymbol{A}(\lambda)$ 的标准形, 此时 $d_1(\lambda)=1, d_2(\lambda)=\lambda, d_3(\lambda)=\lambda^3+\lambda$.

可以证明, 一个多项式矩阵 $\boldsymbol{A}(\lambda)$ 的标准形式的对角线上的非零元素 $d_i(\lambda)$ 不随矩阵的初等变换而改变. 因此, 通常称 $d_i(\lambda)(i=1,2,\cdots,s)$ 为 $\boldsymbol{A}(\lambda)$ 的不变因子或不变因式.

如果以 $D_i(\lambda)(i=1,2,\cdots,s)$ 表示 $\boldsymbol{A}(\lambda)$ 的一切 i 阶子式的最大 (高) 公因式 (常称之为 $\boldsymbol{A}(\lambda)$ 的 i 阶行列式因子, 由行列式性质知 $D_i(\lambda)$ 不随初等变换而改变), 则 $\boldsymbol{A}(\lambda)$ 的不变因子的计算公式为

$$d_i(\lambda)=\frac{D_i(\lambda)}{D_{i-1}(\lambda)},\quad D_0(\lambda)=1\quad(i=1,2,\cdots,s)$$

表明, $\boldsymbol{A}(\lambda)$ 的标准形式被 $D_i(\lambda)(i=1,2,\cdots,s)$ 唯一决定.

把 $\boldsymbol{A}(\lambda)$ 的每个次数大于零的不变因子 $d_i(\lambda)$ 分解为不可约因式的乘积, 这样的不可约因式 (连同它们的幂指数) 称为 $\boldsymbol{A}(\lambda)$ 的一个初等因子, 初等因子的全体称为 $\boldsymbol{A}(\lambda)$ 的初等因子组.

确定 $\boldsymbol{A}(\lambda)$ 的初等因子组的一个简便方法是: 用初等变换将 $\boldsymbol{A}(\lambda)$ 化为对角矩阵, 若记对角线上的非零多项式为 $f_i(\lambda)(i=1,2,\cdots,s)$, 那么诸次数大于零的 $f_i(\lambda)$ 的全体不可约因式, 就是 $\boldsymbol{A}(\lambda)$ 的初等因子组.

要注意的是, 初等因子组是随系数域不同而不同的. 因为有些不变因子在有理数域上不可约, 但在实数域 $\mathbb{R}$ 或复数域 $\mathbb{C}$ 上确是可约的.

那么, 可以得到, 在复数域 $\mathbb{C}$ 上, 求 n 阶矩阵 $\boldsymbol{A}$ 的 Jordan 标准形的步骤如下:

第一步: 求特征矩阵 $\lambda\boldsymbol{I}-\boldsymbol{A}$ 的初等因子组, 设为

$$(\lambda-\lambda_1)^{m_1},(\lambda-\lambda_2)^{m_2},\cdots,(\lambda-\lambda_s)^{m_s}$$

其中, $\lambda_1,\lambda_2,\cdots,\lambda_s$ 可能由相同的, 指数 $m_1,m_2,\cdots,m_s$ 也可能有相同的, 且

$$m_1+m_2+\cdots+m_s=n$$

第二步: 写出每个初等因子 $(\lambda - \lambda_i)^{m_i}(i = 1, 2, \cdots, s)$ 对应的 Jordan 块

$$\boldsymbol{J}_i(\lambda_i) = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \lambda_i & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix} \quad (i = 1, 2, \cdots, s)$$

第三步: 写出以这些 Jordan 块构成的 Jordan 标准形

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{J}_1(\lambda_1) & & & \\ & \boldsymbol{J}_2(\lambda_2)7 & & \\ & & \ddots & \\ & & & \boldsymbol{J}_s(\lambda_s) \end{bmatrix}$$

例 2.8　求矩阵 $\boldsymbol{A}$ 的 Jordan 标准形, 其中

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

解　求 $\lambda\boldsymbol{I} - \boldsymbol{A}$ 的初等因子组

$$\lambda\boldsymbol{I} - \boldsymbol{A} = \begin{bmatrix} \lambda+1 & -1 & 0 \\ 4 & \lambda-3 & 0 \\ -1 & 0 & \lambda-2 \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & (\lambda-1)^2 & 0 \\ 0 & -1 & \lambda-2 \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (\lambda-2)(\lambda-1)^2 \end{bmatrix}$$

因此, 所求的初等因子组为 $\lambda - 2, (\lambda - 1)^2$. 于是有

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

定理 2.12　每个 n 阶复矩阵 $\boldsymbol{A}$ 都与一个 Jordan 标准形相似, 这个 Jordan 标准形除去其中 Jordan 块的排列次序外, 是被 $\boldsymbol{A}$ 唯一确定的.

对于所需要的可逆矩阵 $\boldsymbol{P}$, 下面给出特殊情况下 $\boldsymbol{P}$ 的计算方法.

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ & \lambda_2 & 1 \\ & & \lambda_2 \end{bmatrix}$$

其中 $\boldsymbol{P} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$, 于是有

$$\boldsymbol{A}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)\begin{bmatrix} \lambda_1 & 0 & 0 \\ & \lambda_2 & 1 \\ & & \lambda_2 \end{bmatrix}$$

即

$$(\boldsymbol{A}\boldsymbol{x}_1, \boldsymbol{A}\boldsymbol{x}_2, \boldsymbol{A}\boldsymbol{x}_3) = (\lambda_1\boldsymbol{x}_1, \lambda_2\boldsymbol{x}_2, \boldsymbol{x}_2 + \lambda_2\boldsymbol{x}_3)$$

由此可得

$$\begin{cases} (\lambda_1\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x}_1 = \boldsymbol{0} \\ (\lambda_2\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x}_2 = \boldsymbol{0} \\ (\lambda_2\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x}_3 = -\boldsymbol{x} - 2 \end{cases} \tag{2.1.10}$$

从而 $\boldsymbol{x}_1, \boldsymbol{x}_2$ 依次是 $\boldsymbol{A}$ 的属于 λ_1, λ_2 的特征向量. $\boldsymbol{x}_3$ 是最后一个非齐次线性方程组的解向量. 求出这些解向量就得到了所需要的矩阵 $\boldsymbol{P}$.

在一般情况下, 如果 λ_1 是 $\boldsymbol{A}$ 的 k 重特征值, 则 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k$ 可由解下面各方程组

$$(\lambda_1\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x}_1 = \boldsymbol{0}(\lambda_1\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x}_i = -\boldsymbol{x}_{i-1} \quad (i = 2, 3, \cdots k)$$

而获得. 这样得到的 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k$ 线性无关, 于是

$$\boldsymbol{P} = (\boldsymbol{x}_2, \boldsymbol{x}_3, \cdots, \boldsymbol{x}_k, \cdots)$$

称 $\boldsymbol{x}_2, \boldsymbol{x}_3, \cdots, \boldsymbol{x}_k$ 为 $\boldsymbol{A}$ 属于 λ_1 的广义特征向量.

例 2.9　求例 2.8 中, 使矩阵 $\boldsymbol{A}$ 相似于 Jordan 标准形时所用的可逆矩阵 $\boldsymbol{P}$.

解　因为 $\lambda_1 = 2, \lambda_2 = 1$ 分别是其单特征值和二重特征值, 所以可用式 (2.1.10) 求 $\boldsymbol{P} = (p_{ij})_{3\times 3}$, 这里 $(p_{1i}, p_{2i}, p_{3i})^T = \boldsymbol{x}_i (i = 1, 2, 3)$. 解方程组

$$(2\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x}_1 = \boldsymbol{0}, \quad (\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x}_2 = \boldsymbol{0}, \quad (\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x}_3 = -\boldsymbol{x}_2$$

得特征向量 $\boldsymbol{x}_1, \boldsymbol{x}_2$ 及广义特征向量 $\boldsymbol{x}_3$ 依次为

$$\boldsymbol{x}_1 = (0, 0, 1)^T, \quad \boldsymbol{x}_2 = (1, 2, -1)^T, \quad \boldsymbol{x}_3 = (0, 1, -1)^T$$

故所求矩阵 $\boldsymbol{P}$ 为

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

2.1.3 欧式空间中线性变换的求法

本节讨论在欧氏空间中的线性变换, 特别是正交变换和对称变换, 它们与几何密切相关.

欧氏空间的性质

假定 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 是 n 维欧氏空间 V^n 的基, 对于 V^n 的任意两个向量

$$\boldsymbol{x} = \xi_1\boldsymbol{x}_1 + \xi_2\boldsymbol{x}_2 + \cdots + \xi_n\boldsymbol{x}_n, \boldsymbol{y} = \eta_1\boldsymbol{x}_1 + \eta_2\boldsymbol{x}_2 + \cdots + \eta_n\boldsymbol{x}_n$$

可得

$$(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i,j=1}^{n} \xi_i\eta_j(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sum_{i,j=1}^{n} a_{ij}\xi_i\eta_j$$

其中 $\boldsymbol{a}_{ij} = (\boldsymbol{x}_i, \boldsymbol{x}_j)(i, j = 1, 2, \cdots, n)$, 用矩阵乘法表示, 则有

$$(\boldsymbol{x}, \boldsymbol{y}) = (\xi_1, \xi_2, \cdots, \xi_n)\boldsymbol{A} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix}$$

这里

$$\boldsymbol{A} = (a_{ij})_{n\times n} = \begin{bmatrix} (\boldsymbol{x}_1, \boldsymbol{x}_1) & \cdots & (\boldsymbol{x}_1, \boldsymbol{x}_n) \\ (\boldsymbol{x}_2, \boldsymbol{x}_1) & \cdots & (\boldsymbol{x}_2, \boldsymbol{x}_n) \\ \vdots & & \vdots \\ (\boldsymbol{x}_n, \boldsymbol{x}_1) & \cdots & (\boldsymbol{x}_n, \boldsymbol{x}_n) \end{bmatrix}$$

可以看出, 只要知道其中任意两个基向量的内积, 也就知道了矩阵 $\boldsymbol{A}$, 从而也就知道了任意两个向量的内积. 因此, 称 $\boldsymbol{A}$ 为 V^n 对于基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 的度量矩阵, 且是对称矩阵、正定矩阵.

正交性

通常, 两个向量垂直的充分必要条件是它们夹角的余弦为零, 亦即它们的数量积为零. 在一般的欧氏空间中, 仍以内积定义二向量夹角的余弦.

定义 2.13　如果对于欧氏空间中的两个向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$, 有 $(\boldsymbol{x},\boldsymbol{y})=0$, 则称 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 正交或垂直, 记为 $\boldsymbol{x}\perp\boldsymbol{y}$.

定义 2.14　如果欧氏空间中一组非零向量两两正交, 则成为正交向量组.

定理 2.13　在欧氏空间中, 如果 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 是正交向量组, 则它们必线性无关.

定义 2.15　在欧氏空间 V^n 中, 由 n 个非零向量组成的正交向量组称为 V^n 的正交基; 由单位向量组成的正交基称为标准正交基或法正交基.

把一个正交基进行单位化, 就得到一个标准正交基.

一个基为标准正交基的充要条件是它的度量矩阵为单位矩阵. 事实上, 标准正交基 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 满足

$$(\boldsymbol{x}_i,\boldsymbol{x}_j)=\delta_{ij}=\begin{cases}1 & (i=j)\\ 0 & (i\neq j)\end{cases}$$

这里的 δ_{ij} 称为 Kronecker 记号. 于是其度量矩阵是单位矩阵. 反之, 如果单位矩阵为度量矩阵, 则由矩阵相等可得 $(\boldsymbol{x}_i,\boldsymbol{x}_j)=\delta_{ij}$, 即 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 为标准正交基.

定理 2.14　对于欧氏空间 V^n 的任一基 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$, 都可以找到一个标准正交基, 令其为 $\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_n$. 换言之, 任一非零欧氏空间都有正交基和标准正交基.

证　应用下面论述的关于向量组的 Schmidt 正交化方法, 给出定理的构造性证明. 为此取 $\boldsymbol{y}_1'=\boldsymbol{x}_1$, 作为所求正交基中的第一个向量. 再令

$$\boldsymbol{y}_2'=\boldsymbol{x}_2+k\boldsymbol{y}_1'$$

由正交条件 $(\boldsymbol{y}_2',\boldsymbol{y}_1')=0$ 来决定待定常数 k. 由

$$(\boldsymbol{x}_2+k\boldsymbol{y}_1',\boldsymbol{y}_1')=(\boldsymbol{x}_2,\boldsymbol{y}_1')+k(\boldsymbol{y}_1',\boldsymbol{y}_1')=0$$

得

$$k=-\frac{(\boldsymbol{x}_2,\boldsymbol{y}_1')}{(\boldsymbol{y}_1',\boldsymbol{y}_1')}$$

这样就得到两个正交的向量 $\boldsymbol{y}_1',\boldsymbol{y}_2'$, 且 $\boldsymbol{y}_2'\neq\mathbf{0}$. 又令

$$\boldsymbol{y}_3'=\boldsymbol{x}_3+k_2\boldsymbol{y}_2'+k_1\boldsymbol{y}_1'$$

再有正交条件 $(\boldsymbol{y}_3',\boldsymbol{y}_2')=0$ 及 $(\boldsymbol{y}_3',\boldsymbol{y}_1')=0$ 来决定出 k_1 和 k_2 为

$$k_2=-\frac{(\boldsymbol{x}_3,\boldsymbol{y}_2')}{(\boldsymbol{y}_2',\boldsymbol{y}_2')},\quad k_1=-\frac{(\boldsymbol{x}_3,\boldsymbol{y}_1')}{\boldsymbol{y}_1',\boldsymbol{y}_1'}$$

到此, 已经做出三个两两正交的向量 $\boldsymbol{y}_1', \boldsymbol{y}_2', \boldsymbol{y}_3'$, 且 $\boldsymbol{y}_3' \neq \mathbf{0}$. 继续这样进行下去, 设已做出 m 个两两正交且不为零的向量 $\boldsymbol{y}_1', \boldsymbol{y}_2', \cdots, \boldsymbol{y}_m'$, 为求出第 $m+1$ 个与之正交的向量, 令

$$\boldsymbol{y}_{m+1}' = \boldsymbol{x}_{m+1} + l_m\boldsymbol{y}_m' + l_{m-1}\boldsymbol{y}_{m-1}' + \cdots + l_2\boldsymbol{y}_2' + l_1\boldsymbol{y}_1'$$

使用 m 个正交条件

$$(\boldsymbol{y}_{m+1}', \boldsymbol{y}_i') = 0(i = 1, 2, \cdots, m)$$

来决定 $l_m, l_{m-1}, \cdots, l_2, l_1$. 根据 $\boldsymbol{y}_1', \boldsymbol{y}_2', \cdots, \boldsymbol{y}_m'$ 两两正交的假设, 可得

$$(\boldsymbol{x}_{m+1}, y_i') + l_i(\boldsymbol{y}_i', \boldsymbol{y}_i')$$

故

$$l_i = -\frac{(\boldsymbol{x}_{m+1}, \boldsymbol{y}_i')}{(\boldsymbol{y}_i', \boldsymbol{y}_i')}(i = 1, 2, \cdots, m)$$

于是 $\boldsymbol{y}_{m+1}'$ 就被确定出来了.

采用上述 Schmidt 正交化方法, 可由已知基构造出 n 个两两正交的线性无关的非零向量 $\boldsymbol{y}_1', \boldsymbol{y}_2', \cdots, \boldsymbol{y}_n'$, 从而形成 V^n 的一组正交基. 再以 $|\boldsymbol{y}_i'|$ 除 $\boldsymbol{y}_i'(i = 1, 2, \cdots, n)$, 就得到定理所要求的标准正交基.

$$\boldsymbol{y}_i = \frac{1}{|\boldsymbol{y}_i'|}\boldsymbol{y}_i'(i = 1, 2, \cdots, n)$$

上述是由基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 构造标准正交基 $\boldsymbol{y}_1', \boldsymbol{y}_2', \cdots, \boldsymbol{y}_n'$ 的过程, 有时也称为把基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 正交单位化或正交规范化.

例 2.10　试把向量组 $\boldsymbol{x}_1 = (1, 1, 0, 0)$,$\boldsymbol{x}_2 = (1, 0, 1, 0)$, $\boldsymbol{x}_3 = (-1, 0, 0, 1)$,$\boldsymbol{x}_4 = (1, -1, -1, 1)$ 正交单位化.

解　先把它们正交化，使用 $l_i = -\dfrac{(\boldsymbol{x}_{m+1}, \boldsymbol{y}_i')}{\boldsymbol{y}_i', \boldsymbol{y}_i'}$，可得

$$\begin{aligned}
\boldsymbol{y}_1' &= \boldsymbol{x}_1 = (1, 1, 0, 0) \\
\boldsymbol{y}_2' &= \boldsymbol{x}_2 - \frac{(\boldsymbol{x}_2, \boldsymbol{y}_1')}{(\boldsymbol{y}_1', \boldsymbol{y}_1')}y_1' = (\frac{1}{2}, -\frac{1}{2}, 1, 0) \\
\boldsymbol{y}_3' &= \boldsymbol{x}_3 - \frac{(\boldsymbol{x}_3, \boldsymbol{y}_2')}{(\boldsymbol{y}_2', \boldsymbol{y}_2')}y_2' - \frac{(\boldsymbol{x}_3, \boldsymbol{y}_1')}{(\boldsymbol{y}_1', \boldsymbol{y}_1')}y_1' = (-\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 1) \\
\boldsymbol{y}_4' &= \boldsymbol{x}_4 - \frac{(\boldsymbol{x}_4, \boldsymbol{y}_3')}{(\boldsymbol{y}_3', \boldsymbol{y}_3')}y_3' - \frac{(\boldsymbol{x}_4, \boldsymbol{y}_2')}{(\boldsymbol{y}_2', \boldsymbol{y}_2')}y_2' - \frac{(\boldsymbol{x}_4, \boldsymbol{y}_1')}{(\boldsymbol{y}_1', \boldsymbol{y}_1')}y_1' = (1, -1, -1, 1)
\end{aligned}$$

再单位化，则有

$$\boldsymbol{y}_1 = \frac{1}{|\boldsymbol{y}_1'|}\boldsymbol{y}_1' = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0)$$

$$
\begin{aligned}
y_2 &= \frac{1}{|y_2'|}y_2' = (\frac{1}{\sqrt{6}}, \frac{-1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, 0) \\
y_3 &= \frac{1}{|y_3'|}y_3' = (-\frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{3}{\sqrt{12}},) \\
y_4 &= \frac{1}{|y_4'|}y_4' = (\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2})
\end{aligned}
$$

正交变换与正交矩阵

由解析几何知, 在旋转变化之下, 向量的长度保持不变. 在线性空间中, 能保持向量长度不变的线性变换, 在实际中应用是很广泛的.

定义 2.16　设 $\boldsymbol{V}$ 为欧氏空间, $\boldsymbol{T}$ 是 V 的一个线性变换, 如果 $\boldsymbol{T}$ 保持 V 中任意向量 $\boldsymbol{x}$ 的长度不变, 则有

$$(\boldsymbol{T}\boldsymbol{x}, \boldsymbol{T}\boldsymbol{x}) = (\boldsymbol{x}, \boldsymbol{x})$$

那么称 $\boldsymbol{T}$ 是 V 的一个正交变换.

定理 2.15　线性变换 $\boldsymbol{T}$ 为正交变换的充要条件是, 对于欧氏空间 V 中任意向量 $\boldsymbol{x}, \boldsymbol{y}$, 都有 $(\boldsymbol{T}\boldsymbol{x}, \boldsymbol{T}\boldsymbol{y}) = (\boldsymbol{x}, \boldsymbol{y})$.

定义 2.17　如果实方阵 $\boldsymbol{Q}$ 满足 $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$, 则称 $\boldsymbol{Q}$ 为正交矩阵.

$\boldsymbol{Q}$ 是正交矩阵的充要条件是它的列向量是两两正交的单位向量, 此外, 正交矩阵还有如下性质.

(1) 正交矩阵都是可逆的.

(2) 正交矩阵的逆矩阵仍是正交矩阵.

(3) 两个正交矩阵的乘积仍未正交矩阵.

对称变换与对称矩阵

定义 2.18　设 $\boldsymbol{T}$ 是欧式空间 V 的一个线性变换, 且对 V 中任意两个向量 $\boldsymbol{x}, \boldsymbol{y}$, 都有

$$(\boldsymbol{T}\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}, \boldsymbol{T}\boldsymbol{y})$$

则称 $\boldsymbol{T}$ 为 V 中的一个对称变换.

定理 2.16　欧氏空间的线性变换是实对称变换的充要条件是, 它对于标准正交基的矩阵是实对称矩阵.

实对称矩阵有如下性质.

(1) 实对称矩阵的特征值都是实数.

(2) 实对称矩阵的不同特征值所对应的特征向量是正交的.

2.2 向量范数与矩阵范数

在计算数学中, 特别是在数值代数中, 研究数值方法的收敛性、稳定性及误差分析等问题时, 范数理论显得十分重要. 本节主要讨论 n 维向量空间 C^n 中的向量范数与矩阵空间 $C^{m\times n}$ 中的矩阵范数的理论及其性质.

2.2.1 向量范数介绍

向量范数的概念及 l_p 范数

把一个向量 (或线性空间的元素) 与一个非负实数相联系, 在许多场合下, 这个实数可以作为向量大小的一种度量. 向量范数就是这样的实数, 它们在研究数值方法的收敛性和误差分析等方面与有着重要的作用. 现定义如下.

定义 2.19　如果 V 是数域 K 上的线性空间, 对任意的 $\boldsymbol{x}\in V$, 定义一个实值函数 $\|\boldsymbol{x}\|$, 它满足以下三个条件:

(1) 非负性: 当 $\boldsymbol{x}\neq\boldsymbol{0}$ 时, $\|\boldsymbol{x}\|>0$; 当 $\boldsymbol{x}=\boldsymbol{0}$ 时, $\|\boldsymbol{x}\|=0$;

(2) 齐次性: $\|a\boldsymbol{x}\|=|a|\,\|\boldsymbol{x}\|(a\in K,\boldsymbol{x}\in V)$;

(3) 三角不等式: $\|\boldsymbol{x}+\boldsymbol{y}\|\leqslant\|\boldsymbol{x}\|+\|\boldsymbol{y}\|(\boldsymbol{x},\boldsymbol{y}\in V)$.

则称 $\|\boldsymbol{x}\|$ 为 V 上向量 $\boldsymbol{x}$ 的范数, 简称向量范数.

例 2.11　证明 $\|\boldsymbol{x}\|=\sqrt{|\xi_1|^2+|\xi_2|^2+\cdots+|\xi_n|^2}$ 是 C^n 上的一种范数, $\boldsymbol{x}=(\xi_1,\xi_2,\cdots,\xi_n)\in C^n$.

证　当 $\boldsymbol{x}\neq\boldsymbol{0}$ 时, 显然 $\|\boldsymbol{x}\|>0$; 当 $\boldsymbol{x}=\boldsymbol{0}$ 时, 有 $\|\boldsymbol{x}\|=0$.

对于任意的复数 a, 有

$$\|a\boldsymbol{x}\|=|a|\sqrt{|\xi_1|^2+|\xi_2|^2+\cdots+|\xi_n|^2}=|a|\,\|\boldsymbol{x}\|$$

对于任意两个向量 $\boldsymbol{x},\boldsymbol{y}\in C^n$, 有

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2=(\boldsymbol{x}+\boldsymbol{y},\boldsymbol{x}+\boldsymbol{y})=(\boldsymbol{x},\boldsymbol{x})+2\mathrm{Re}(\boldsymbol{x},\boldsymbol{y})+(\boldsymbol{y},\boldsymbol{y})$$

因此,

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2\leqslant\|\boldsymbol{x}\|^2+2\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|+\|\boldsymbol{y}\|^2=(\|\boldsymbol{x}\|+\|\boldsymbol{y}\|)^2$$

即 $\|\boldsymbol{x}+\boldsymbol{y}\|\leqslant\|\boldsymbol{x}\|+\|\boldsymbol{y}\|$.

称例 2.11 中的范数为向量的 2 -范数, 记为 $\|\boldsymbol{x}\|_2$, 即

$$\|\boldsymbol{x}\|_2=\sqrt{|\xi_1|^2+|\xi_2|^2+\cdots+|\xi_n|^2}\tag{2.2.1}$$

例 2.12　证明 $\|\boldsymbol{x}\|=\max\limits_i|\xi_i|$ 是 C^n 上的一种范数, 这里 $\boldsymbol{x}=(\xi_1,\xi_2,\cdots,\xi_n)\in C^n$.

证 当 $\boldsymbol{x} \neq \boldsymbol{0}$ 时, $\|\boldsymbol{x}\| = \max_i |\xi_i| > 0$; 当 $\boldsymbol{x} = \boldsymbol{0}$ 时, 显然有 $\|\boldsymbol{x}\| = 0$.

又对任意的 $a \in C$, 有

$$\|a\boldsymbol{x}\| = \max_i |a\xi_i| = |a| \max_i |\xi_i|$$

对 C^n 的任意两个向量 $\boldsymbol{x} = (\xi_1, \xi_2, \cdots, \xi_n), \boldsymbol{y} = (\eta_1, \eta_2, \cdots, \eta_n)$, 有

$$\|\boldsymbol{x} + \boldsymbol{y}\| = \max_i |\xi_i + \eta_i| \leqslant \max_i |\xi_i| + \max_i |\eta_i| = \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$$

因此, $\|\boldsymbol{x}\| = \max_i |\xi_i|$ 是 C^n 的一种范数.

称例 2.12 中的范数为向量的 ∞ -范数, 记为 $\|\boldsymbol{x}\|_\infty$, 即

$$\|\boldsymbol{x}\|_\infty = \max_i |\xi_i| \tag{2.2.2}$$

例 2.13 证明 $\|\boldsymbol{x}\| = \sum_{i=1}^{n} |\xi_i|$ 是 C^n 上的一种范数, 其中 $\boldsymbol{x} = (\xi_1, \xi_2, \cdots, \xi_n) \in C^n$.

证 当 $\boldsymbol{x} \neq \boldsymbol{0}$ 时, 显然 $\|\boldsymbol{x}\| = \sum_{i=0}^{n} |\xi_i| > 0$; 当 $\boldsymbol{x} = \boldsymbol{0}$ 时, 由于 $\boldsymbol{x}$ 的每一分量都是零, 故 $\|\boldsymbol{x}\| = 0$.

又对于任意 $a \in C^n$, 有

$$\|\boldsymbol{x}\| = \sum_{i=1}^{n} |a\xi_i| = |a| \sum_{i=1}^{n} |\xi_i| = |a| \ \|\boldsymbol{x}\|$$

对于任意两个向量 $\boldsymbol{x}, \boldsymbol{y} \in C^n$, 有

$$\begin{aligned} \|\boldsymbol{x} + \boldsymbol{y}\| = \sum_{i=1}^{n} |\xi_i + \eta_i| \leqslant \sum_{i=1}^{n} (|xi_i| + |\eta_i|) = \\ = \sum_{i=1}^{n} |\xi_i| + \sum_{i=1}^{n} |\eta_i| = \|\boldsymbol{x} + \boldsymbol{y}\| \end{aligned}$$

因此, $\|\boldsymbol{x}\| = \sum_{i=1}^{n} |\xi_i|$ 是 C^n 的一种范数.

称例 2.13 中的范数为向量的 1 -范数, 记为 $\|\boldsymbol{x}\|_1$, 即

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |\xi_i| \tag{2.2.3}$$

由例 2.11 ~ 例 2.13 可知, 在一个线性空间中, 可以定义多种向量范数, 实际上可以定义无限多种范数. 例如, 对于不小于 1 的任意实数 p 及 $\boldsymbol{x} = (\xi_1, \xi_2, \cdots, \xi_n) \in C^n$, 可以证明实值函数

$$(\sum_{i=1}^{n} |\xi_i|^p)^{1/p} \quad (1 \leqslant p < +\infty)$$

满足向量范数的三个条件. 称 $(\sum_{i=1}^{n} |\xi_i|^p)^{1/p}$ 为向量 $\boldsymbol{x}$ 的 p -范数或 l_p 范数, 记为 $\|\boldsymbol{x}\|_p$, 即

$$\|\boldsymbol{x}\|_p = (\sum i = 1^n |\xi_i|^p)^{1/p} \tag{2.2.4}$$

线性空间 V^n 上的向量范数的等价性

定理 2.17　设 $\|\boldsymbol{x}\|_\alpha$ 和 $\|\boldsymbol{x}\|_\beta$ 为有限维线性空间 V 上的任意两种向量范数 (它们不限于 p -范数), 则存在两个与向量 $\boldsymbol{x}$ 无关的正常数 c_1 和 c_2, 使满足

$$c_1\|\boldsymbol{x}\|_\beta \leqslant \|\boldsymbol{x}\|_\alpha \leqslant c_2\|\boldsymbol{x}\|_\beta \quad (\forall \boldsymbol{x} \in V)$$

且满足这一不等式的两种范数被称为是等价的.

对于 C^n 上向量 $\boldsymbol{x}$ 的 p -范数, 也满足不等式

$$1\|\boldsymbol{x}\|_\infty \leqslant \|\boldsymbol{x}\|_1 \leqslant n\|\boldsymbol{x}\|_\infty \qquad 1\|\boldsymbol{x}\|_\infty \leqslant \|\boldsymbol{x}\|_2 \leqslant \sqrt{n}\|\boldsymbol{x}\|_\infty$$

以上两式表明, 对某一向量 $\boldsymbol{x}$ 而言, 如果它的某一种范数小 (或大), 那么它的另两种范数也小 (或大).

2.2.2　矩阵范数介绍

矩阵空间 $C^{n\times n}$ 是一个 mn 维的线性空间, 将 $m\times n$ 矩阵 $\boldsymbol{A}$ 看做线性空间 $C^{m\times n}$ 中的 “向量”, 于是可以定义 $\boldsymbol{A}$ 的范数. 但是, 矩阵之间还有乘法运算, 它应该在定义矩阵范数时予以体现.

矩阵范数的定义与性质

定义 2.20　设 $\boldsymbol{A} \in C^{m\times n}$, 定义一个实值函数 $\|\boldsymbol{A}\|$, 它满足以下三个条件:

(1) 非负性: 当 $\boldsymbol{A} \neq \boldsymbol{O}$ 时, $\|\boldsymbol{A}\| > 0$; 当 $\boldsymbol{A} = \boldsymbol{O}$ 时, $\|\boldsymbol{A}\| = 0$;

(2) 齐次性: $\|\alpha\boldsymbol{A}\| = |\alpha|\ \|\boldsymbol{A}\|(\alpha \in C)$;

(3) 三角不等式: $\|\boldsymbol{A}+\boldsymbol{B}\| \leqslant \|\boldsymbol{A}\| + \|\boldsymbol{B}\|(\boldsymbol{B} \in C^{m\times n})$. 则称 $\|\boldsymbol{A}\|$ 为 $\boldsymbol{A}$ 的广义矩阵范数. 若对 $C^{m\times n}, C^{n\times l}$ 及 $C^{m\times l}$ 上的同类广义矩阵范数 $\|\cdot\|$, 还满足下面一个条件:

(4) 相容性:

$$\|\boldsymbol{AB}\| \leqslant \|\boldsymbol{A}\|\ \|\boldsymbol{B}\| \quad (\boldsymbol{B} \in C^{n\times l})$$

则称 $\|\boldsymbol{A}\|$ 为 $\boldsymbol{A}$ 的矩阵范数.

如同向量范数的情况一样, 矩阵范数也是多种多样的. 但是, 在数值方法中进行某种估计时, 遇到的多数情况时: 矩阵范数常与向量范数混合在一起使用, 而矩阵范数经常是作为两个线性空间上的线性映射 (变换) 出现的. 因此, 考虑一些矩阵范数时, 应该使它能与向量范数联系起来. 这可由矩阵范数与向量范数相容的概念来体现.

定义 2.21　对于 $bmC^{m\times n}$ 上的矩阵范数 $\|\cdot\|_M$ 和 C^m 与 C^n 上的同类向量范数 $\|\cdot\|_V$, 如果

$$\|\boldsymbol{Ax}\|_V \leqslant \|\boldsymbol{A}\|_M\|\boldsymbol{x}\|_V \quad (\forall \boldsymbol{A} \in C^{m\times n}, \forall \boldsymbol{x} \in C^n)$$

则称矩阵范数 $\|\cdot\|_M$ 与向量范数 $\|\cdot\|_V$ 是相容的.

例 2.14　设 $\boldsymbol{A} = (a_{ij})_{m\times n} \in C^{n\times n}$, 证明函数

$$\|\boldsymbol{A}\|_F = (\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2)^{1/2} = (\mathrm{tr}(\boldsymbol{A}^H\boldsymbol{A}))^{1/2} \tag{2.2.5}$$

是 $C^{n\times n}$ 上的矩阵范数, 且与向量范数 $\|\cdot\cdot\|_2$ 相容.

证　显然, $\|\boldsymbol{A}\|_F$ 具有非负性和齐次性. 设 $\boldsymbol{B} \in C^{m\times n}$, 且 $\boldsymbol{A}$ 的第 j 列分别为 $\boldsymbol{a}_j, b_j(j = 1, 2, \cdots, n)$, 则有

$$\begin{aligned}\|\boldsymbol{A}+\boldsymbol{B}\|_F^2 =& \|\boldsymbol{a}_1+\boldsymbol{b}_1\|_2^2 + \cdots + \|\boldsymbol{a}_n+\boldsymbol{b}_n\|_2^2 \leqslant \\ & (\|\boldsymbol{a}_1\|_2 + \|\boldsymbol{b}_1\|_2)^2 + \cdots + (\|\boldsymbol{a}_n\|_2 + \|\boldsymbol{b}_n\|_2)^2 \leqslant \\ & \|\boldsymbol{A}\|_F^2 + 2\|\boldsymbol{A}\|_F\|\boldsymbol{B}\|_F + \|\boldsymbol{B}\|_F^2 = \\ & (\|\boldsymbol{A}\|_F + \|\boldsymbol{B}\|_F)^2\end{aligned}$$

即三角不等式成立.

再设 $\boldsymbol{B} = (b_{ij})_{n\times l} \in C^{n\times l}$, 则 $\boldsymbol{AB} = (\sum\limits_{k=1}^{n} a_{ik}b_{kj})_{m\times l} \in C^{m\times l}$, 于是有

$$\|\boldsymbol{AB}\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{l}|\sum_{k=1}^{n} a_{ik}b_{kj}|^2 \leqslant \sum_{i=1}^{m}\sum_{j=1}^{l}(\sum_{k=1}^{n}|a_{ik}||b_{kj}|)^2$$

可得

$$\|\boldsymbol{AB}\|_F^2 \leqslant \|\boldsymbol{A}\|_F^2\|\boldsymbol{B}\|_F^2$$

即 $\|\boldsymbol{A}\|_F$ 是 $\boldsymbol{A}$ 的矩阵范数.

取 $\boldsymbol{B} = \boldsymbol{x} \in C^{n\times l}$, 则有

$$\|\boldsymbol{Ax}\|_2 = \|\boldsymbol{AB}\|_F \leqslant \|\boldsymbol{A}\|_F\|\boldsymbol{B}\|_F = \|\boldsymbol{A}\|_F\|\boldsymbol{x}\|_2$$

即矩阵范数 $\|\cdot\|_F$ 与向量范数 $\|\cdot\|_2$ 相容.

范数 (2.2.5) 又称为 Frobenuis 范数, 或简称为 F -范数.

几种常用的矩阵范数

现在给出一种规定矩阵范数的具体方法, 使矩阵范数与已知的向量范数相容.

定理 2.18　已知 C^m 和 C^n 上的同类向量范数 $\|\cdot\|$, 设 $\boldsymbol{A} \in C^{m\times n}$, 则函数

$$\|\boldsymbol{A}\| = \max_{\|\boldsymbol{x}\|=1}\|\boldsymbol{Ax}\| \tag{2.2.6}$$

是 $C^{m\times n}$ 上的矩阵范数, 且与已知的向量范数相容.

称由式 (2.2.6) 给出的矩阵范数为由向量范数导出的矩阵范数, 简称为从属范数. 对于 $C^{m\times n}$ 上的任何一种从属范数, 有

$$\|\boldsymbol{I}\| = \max_{\|\boldsymbol{x}\|=1}\|\boldsymbol{I}\boldsymbol{x}\| = 1$$

但对于一般的矩阵范数 (设该矩阵范数与某向量范数相容), 由于

$$\|\boldsymbol{x}\| = \|\boldsymbol{I}\boldsymbol{x}\| \leqslant \|\boldsymbol{I}\|\ \|\boldsymbol{x}\|$$

对任意的 $\boldsymbol{x} \in C^n$ 成立, 所以 $\|\boldsymbol{I}\| \geqslant 1$.

上面论述表明, 矩阵范数是与向量范数密切相关的, 有什么样的向量范数就有什么样的矩阵范数. 当在式 (2.2.6) 中取向量 $\boldsymbol{x}$ 的范数 $\|\boldsymbol{x}\|$ 依次为 $\|\boldsymbol{x}\|_1, \|\boldsymbol{x}\|_2, \|\boldsymbol{x}\|_\infty$ 时, 就得到三种常用的矩阵范数, 如下.

定理 2.19　设 $\boldsymbol{A} = (a_{ij})_{m\times n} \in C^{m\times n}, \boldsymbol{x} = (\xi_1, \xi_2, \cdots, \xi_n)^T \in C^n$, 则从属于向量 $\boldsymbol{x}$ 的三种范数 $\|\boldsymbol{x}\|_1, \|\boldsymbol{x}\|_2, \|\boldsymbol{x}\|_\infty$ 的矩阵范数计算公式依次为

(1) $\|\boldsymbol{A}\|_1 = \max\limits_{j} \sum\limits_{i=1}^{m} |a_{ij}|$;

(2) $\|\boldsymbol{A}\|_2 = \sqrt{\lambda_1}$, λ_1 为 $\boldsymbol{A}^H\boldsymbol{A}$ 的最大特征值;

(3) $\|\boldsymbol{A}\|_\infty = \max\limits_{i} \sum\limits_{j=1}^{n} |a_{ij}|$;

通常称 $\|\boldsymbol{A}\|_1, \|\boldsymbol{A}\|_2, \|\boldsymbol{A}\|_\infty$ 依次为列和范数、谱范数及行和范数.

2.2.3　矩阵可逆性条件、谱半径和条件数介绍

本节主要介绍矩阵范数的几点应用.

矩阵的可逆性条件

设 $\boldsymbol{A} \in C^{n\times n}$, 可以根据范数 $\|\boldsymbol{A}\|$ 的大小来判断 $\boldsymbol{I} - \boldsymbol{A}$ 是否为可逆矩阵.

定理 2.20　设 $\boldsymbol{A} \in C^{n\times n}$, 且对于 $C^{n\times n}$ 上的某种矩阵范数 $\|\cdot\|$, 有 $\|\boldsymbol{A}\| < 1$, 则矩阵 $\boldsymbol{I} - \boldsymbol{A}$ 可逆, 且有

$$\|(\boldsymbol{I} - \boldsymbol{A})^{-1}\| \leqslant \frac{\|\boldsymbol{I}\|}{1 - \|\boldsymbol{A}\|}$$

证　设矩阵范数 $\|\boldsymbol{A}\|$ 与向量范数 $\|\boldsymbol{x}\|_V$ 相容, 如果 $\det(\boldsymbol{I} - \boldsymbol{A}) = 0$, 则齐次线性方程组 $(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ 有非零解 $\boldsymbol{x}_0$, 即

$$(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x}_0 = \boldsymbol{0}$$

从而有

$$\|\boldsymbol{x}_0\|_V = \|\boldsymbol{A}\boldsymbol{x}_0\|_V \leqslant \|\boldsymbol{A}\|\ \|\boldsymbol{x}_0\|_V < \|\boldsymbol{x}_0\|_V$$

则是一个矛盾, 故 $\det(\boldsymbol{I} - \boldsymbol{A}) \neq 0$, 即 $\boldsymbol{I} - \boldsymbol{A}$ 可逆.

再由 $(\boldsymbol{I}-\boldsymbol{A})^{-1}(\boldsymbol{I}-\boldsymbol{A})=\boldsymbol{I}$ 可得

$$(\boldsymbol{I}-\boldsymbol{A})^{-1}=\boldsymbol{I}+(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{A}$$

利用范数的三角不等式与相容性可得

$$\|(\boldsymbol{I}-\boldsymbol{A}^{-1})\|\leqslant\|\boldsymbol{I}\|+\|(\boldsymbol{I}-\boldsymbol{A})^{-1}\|\,\|\boldsymbol{A}\|$$

解此不等式可得证.

现在考虑如下问题: 若矩阵 $\boldsymbol{A}$ 的范数 $\|\boldsymbol{A}\|$ 很小, 且由于 $\|\boldsymbol{A}\|$ 是它的元素的连续函数, 所以矩阵 $\boldsymbol{A}$ 接近于零矩阵 $\boldsymbol{O}$, 而 $\boldsymbol{I}-\boldsymbol{O}$ 的逆矩阵为 $\boldsymbol{I}$, 那么, $(\boldsymbol{I}-\boldsymbol{A})^{-1}$ 与单位矩阵 $\boldsymbol{I}$ 的逼近程度可由下面的定理给出.

定理 2.21　设 $\boldsymbol{A}\in C^{n\times n}$, 且对 $C^{n\times n}$ 上的某种矩阵范数 $\|\cdot\|$, 有 $\|\boldsymbol{A}\|<1$, 则

$$\|\boldsymbol{I}-(\boldsymbol{I}-\boldsymbol{A})^{-1}\|\leqslant\frac{\|\boldsymbol{A}\|}{1-\|\boldsymbol{A}\|}$$

证　因为 $\|\boldsymbol{A}\|<1$, 所以 $(\boldsymbol{I}-\boldsymbol{A})^{-1}$ 存在, 给 $(\boldsymbol{I}-\boldsymbol{A})-\boldsymbol{A}=-\boldsymbol{A}$ 右乘 $(\boldsymbol{I}-\boldsymbol{A})^{-1}$ 可得

$$\boldsymbol{I}-(\boldsymbol{I}-\boldsymbol{A})^{-1}=-\boldsymbol{A}(\boldsymbol{I}-\boldsymbol{A})^{-1}$$

利用范数的三角不等式与相容性可得

$$\|\boldsymbol{A}(\boldsymbol{I}-\boldsymbol{A})^{-1}\|\leqslant\|\boldsymbol{A}\|+\|\boldsymbol{A}\|\,\|\boldsymbol{A}(\boldsymbol{I}-\boldsymbol{A})^{-1}\|$$

即 $\|\boldsymbol{A}(\boldsymbol{I}-\boldsymbol{A})^{-1}\|\leqslant\dfrac{\|\boldsymbol{A}\|}{1-\|\boldsymbol{A}\|}$, 故

$$\|\boldsymbol{I}-(\boldsymbol{I}-\boldsymbol{A})^{-1}\|=\|-\boldsymbol{A}(\boldsymbol{I}-\boldsymbol{A})^{-1}\|\leqslant\frac{\|\boldsymbol{A}\|}{1-\|\boldsymbol{A}\|}$$

矩阵的谱半径及其性质

矩阵 $\boldsymbol{A}\in C^{n\times n}$ 的谱半径在特征值估计、广义逆矩阵、数值分析以及数值代数等理论的建树中, 都占有极其重要的地位.

定义 2.22　设 $\boldsymbol{A}\in C^{n\times n}$ 的 n 个特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_n$, 称

$$\rho(\boldsymbol{A})=\max_i|\lambda_i|$$

为 $\boldsymbol{A}$ 的谱半径.

定理 2.22　设 $\boldsymbol{A}\in C^{n\times n}$, 则对 $Cn\times n$ 上任何一种矩阵范数 $\|\cdot\|$, 都有

$$\rho(\boldsymbol{A})\leqslant\|\boldsymbol{A}\|$$

证 设 $\boldsymbol{A}$ 的属于特征值 λ 的特征向量为 $\boldsymbol{x}$, 取与矩阵范数 $\|\cdot\|$ 相容的向量范数 $\|\cdot\|_V$, 则由 $\boldsymbol{Ax} = \lambda\boldsymbol{x}$, 可得

$$|\lambda|\ \|\boldsymbol{x}\|_V = \|\lambda\boldsymbol{x}\|_V = \|\boldsymbol{Ax}\|_V \leqslant \|\boldsymbol{A}\|\ \|\boldsymbol{x}\|_V$$

因为 $\boldsymbol{x} \neq \boldsymbol{x}$, 所以 $|\lambda| \leqslant \|\boldsymbol{A}\|$, 从而 $\rho(\boldsymbol{A}) \leqslant \|\boldsymbol{A}\|$.

例 2.15 试用矩阵

$$\boldsymbol{A} = \begin{bmatrix} 1-j & 3 \\ 2 & 1+j \end{bmatrix} \quad (j = \sqrt{-1})$$

验证上述定义对三种常见范数的正确性.

解 因为 $\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = (\lambda-1)^2 - 5$, 所以 $\lambda_1(\boldsymbol{A}) = 1 + \sqrt{5}, \lambda_2(\boldsymbol{A}) = 1 - \sqrt{5}$, 从而

$$\rho(\boldsymbol{A}) = 1 + \sqrt{5}$$

又 $\|\boldsymbol{A}\|_1 = \|\boldsymbol{A}\|_\infty = 3 + \sqrt{2}$, 而

$$\boldsymbol{A}^H\boldsymbol{A} = \begin{bmatrix} 6 & 5+5j \\ 5-5j & 11 \end{bmatrix}, \quad \det(\lambda\boldsymbol{I} - \boldsymbol{A}^H\boldsymbol{A}) = \lambda^2 - 17\lambda + 16$$

由此得 $\lambda_1(\boldsymbol{A}^H\boldsymbol{A}) = 16, \lambda_2(\boldsymbol{A}^H\boldsymbol{A}) = 1$, 则有

$$\|\boldsymbol{A}\|_2 = \sqrt{\lambda_1(\boldsymbol{A}^H\boldsymbol{A})} = 4$$

可得

$$\rho(\boldsymbol{A}) < \|\boldsymbol{A}\|_1, \quad \rho(\boldsymbol{A}) < \|\boldsymbol{A}\|_2, \quad \rho(\boldsymbol{A}) < \|\boldsymbol{A}\|_\infty$$

例 2.16 设 $\boldsymbol{A} \in C^{n\times n}$, 则 $\rho(\boldsymbol{A}^k) = [\rho(\boldsymbol{A})]^k (k = 1, 2, \cdots)$.

证 设 $\boldsymbol{A}$ 的 n 个特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$, 可得, $\boldsymbol{A}^k$ 的 n 个特征值为 $\lambda_1^k, \lambda_2^k, \cdots, \lambda_n^k$, 则有

$$\rho(\boldsymbol{A}^k) = \max_i |\lambda_i^k| = (\max_i |\lambda_i|)^k = [\rho(\boldsymbol{A})]^k$$

定理 2.23 设 $\boldsymbol{A} \in C^{n\times n}$, 对任意的正数 ϵ, 存在某种矩阵范数 $\|\cdot\|_M$, 使得

$$\|\boldsymbol{A}\|_M \leqslant \rho(\boldsymbol{A}) + \epsilon$$

矩阵的条件数

设 $\boldsymbol{A} = (a_{ij})_{n\times n} \in C^{n\times n}$ 的元素 a_{ij} 带有误差 $\delta a_{ij} (i, j = 1, 2, \cdots, n)$, 则准确矩阵应为 $\boldsymbol{A} + \delta\boldsymbol{A}$, 其中 $\delta\boldsymbol{A} = (\delta a_{ij})$. 若 $\boldsymbol{A}$ 为可逆矩阵, 其逆矩阵 $\boldsymbol{A}^{-1}$ 与 $(\boldsymbol{A} + \delta\boldsymbol{A})^{-1}$ 的近似程度 (摄动) 如何?

定理 2.24　设 $\boldsymbol{A} \in C^{n\times n}$ 可逆, $\boldsymbol{B} \in C^{n\times n}$, 且对 $C^{n\times n}$ 上的某种矩阵范数 $\|\cdot\|$, 有 $\|\boldsymbol{A}^{-1}\| < 1$, 则有以下结论:

(1) $\boldsymbol{A} + \boldsymbol{B}$ 可逆;

(2) 记 $\boldsymbol{F} = \boldsymbol{I} - (\boldsymbol{I} + \boldsymbol{A}^{-1}\boldsymbol{B})^{-1}$, 则 $\|\boldsymbol{F}\| \leqslant \dfrac{\|\boldsymbol{A}^{-1}\boldsymbol{B}\|}{1 - \|\boldsymbol{A}^{-1}\boldsymbol{B}\|}$;

(3) $\dfrac{\|\boldsymbol{A}^{-1} - (\boldsymbol{A} + \boldsymbol{B})^{-1}\|}{\|\boldsymbol{A}^{-1}\|} \leqslant \dfrac{\|\boldsymbol{A}^{-1}\boldsymbol{B}\|}{1 - \|\boldsymbol{A}^{-1}\boldsymbol{B}\|}$.

在定理 2.24 中, 若令 $\operatorname{cond}(\boldsymbol{A}) = \|\boldsymbol{A}\|\,\|\boldsymbol{A}^{-1}\|, d_{\boldsymbol{A}} = \|\delta\boldsymbol{A}\|\,\|\boldsymbol{A}\|^{-1}$, 则当 $\|\boldsymbol{A}^{-1}\|\,\|\delta\boldsymbol{A}\| < 1$ 时, 由结论 (2) 与 (3) 可得

$$\|\boldsymbol{I} - (\boldsymbol{I} + \boldsymbol{A}^{-1}\delta\boldsymbol{A})^{-1}\| \leqslant \frac{d_{\boldsymbol{A}}\operatorname{cond}(\boldsymbol{A})}{1 - d_{\boldsymbol{A}}\operatorname{cond}(\boldsymbol{A})}\frac{\|\boldsymbol{A}^{-1} - (\boldsymbol{A} + \delta\boldsymbol{A})^{-1}\|}{\|\boldsymbol{A}^{-1}\|} \leqslant \frac{d_{\boldsymbol{A}}\operatorname{cond}(\boldsymbol{A})}{1 - d_{\boldsymbol{A}}\operatorname{cond}(\boldsymbol{A})}$$

称 $\operatorname{cond}(\boldsymbol{A})$ 为矩阵 $\boldsymbol{A}$ 的条件数, 它是一个求矩阵逆的摄动的一个重要量. 一般说来, 条件数愈大, $(\boldsymbol{A} + \delta\boldsymbol{A})^{-1}$ 与 $\boldsymbol{A}^{-1}$ 的相对误差就愈大.

2.3　矩阵函数介绍

本节将要介绍矩阵分析理论. 矩阵分析理论的建立, 同数学分析一样, 也是以极限理论为基础的, 其内容丰富, 是研究数值方法和其他数学分支以及许多工程问题的重要工具. 本节首先讨论矩阵序列的极限运算, 然后介绍矩阵序列和矩阵级数的收敛定理、矩阵幂级数和一些矩阵函数; 最后介绍矩阵的微分的概念及其性质.

2.3.1　矩阵序列介绍

定义 2.23　设有矩阵序列 $\{\boldsymbol{A}^{(k)}\}$, 其中 $\boldsymbol{A}^{(k)} = (a_{ij})_{m\times n} \in C^{m\times n}$, 当 $a_{ij}^{(k)} \to a_{ij} (k \to \infty)$ 时, 称 $\{\boldsymbol{A}^{(k)}\}$ 收敛, 或称矩阵 $\boldsymbol{A} = (a_{ij})_{m\times n}$ 为 $\{\boldsymbol{A}^{(k)}\}$ 的极限, 或称 $\{\boldsymbol{A}^{(k)}\}$ 收敛于 $\boldsymbol{A}$, 记为

$$\lim_{k\to\infty} \boldsymbol{A}^{(k)} = \boldsymbol{A} \qquad \boldsymbol{A}^{(k)} \to \boldsymbol{A}$$

不收敛的矩阵序列称为发散.

矩阵序列收敛的性质, 有许多与数列收敛的性质相类似.

性质 1　设 $\boldsymbol{A}^{(k)} \to \boldsymbol{A}_{m\times n}, \boldsymbol{B}^{(k)} \to \boldsymbol{B}_{m\times n}$, 则

$$\lim_{k\to\infty}(\alpha\boldsymbol{A}^{(k)} + \beta\boldsymbol{B}^{(k)}) = \alpha\boldsymbol{A} + \beta\boldsymbol{B}\ (\forall\alpha,\beta \in \mathbb{C})$$

性质 2　设 $\boldsymbol{A}^{(k)} \to \boldsymbol{A}_{m\times n}, \boldsymbol{B}^{(k)} \to \boldsymbol{B}_{n\times l}$, 则

$$\lim_{k\to\infty} \boldsymbol{A}^{(k)}\boldsymbol{B}^{(k)} = \boldsymbol{A}\boldsymbol{B}$$

性质 3　设 $\boldsymbol{A}^{(k)}$ 与 $\boldsymbol{A}$ 都是可逆矩阵, $\boldsymbol{A}^{(k)} \to \boldsymbol{A}$, 则

$$(\boldsymbol{A}^{(k)})^{-1} \to \boldsymbol{A}^{-1}$$

定理 2.25　设 $\boldsymbol{A}^{(k)} \in \boldsymbol{C}^{m\times n}$, 则

(1) $\boldsymbol{A}^{(k)} \to \boldsymbol{O}$ 的充要条件是 $\|\boldsymbol{A}^{(k)}\| \to 0$;

(2) $\boldsymbol{A}^{(k)} \to \boldsymbol{A}$ 的充要条件是 $\|\boldsymbol{A}^{(k)} - \boldsymbol{A}\| \to 0$.

这里, $\|\cdot\|$ 是 $\boldsymbol{C}^{m\times n}$ 上的任何一种矩阵范数.

证　(1) 由于 $\boldsymbol{C}^{m\times n}$ 上的矩阵范数等价, 所以只要对矩阵范数 $\|\cdot\|_{m_\infty}$ 证明结论成立即可. 已知 $\boldsymbol{A}^{(k)} \to \boldsymbol{O}$, 由定义可得

$$a_{ij}^{(k)} \to 0 \quad (i = 1, 2, \cdots, m; j = 1, 2, \cdots, n)$$

也就是 $\max\limits_{i,j}|a_{ij}^{(k)}| \to 0$, 即

$$\|\boldsymbol{A}^{(k)}\|_{m_\infty} = n \cdot \max_{i,j}|a_{ij}^{(k)}| \to 0$$

上述推导步步可逆, 所以结论 (1) 成立.

(2) 由于 $\boldsymbol{A}^{(k)} \to \boldsymbol{A}$ 等价于 $(\boldsymbol{A}^{(k)} - \boldsymbol{A}) \to \boldsymbol{O}$, 所以利用结论 (1) 即可得结论 (2).

定义 2.24　矩阵序列 $\{\boldsymbol{A}^{(k)}\}$ 称为有界的, 如果存在常数 $M > 0$, 使得对一切 k 都有

$$|a_{ij}^{(k)}| < M \ (i = 1, 2, \cdots, m; j = 1, 2, \cdots, n)$$

有界的矩阵序列 $\{\boldsymbol{A}^{(k)}\}$, 有收敛的子序列 $\{\boldsymbol{A}^{(k_s)}\}$.

在矩阵序列中, 最常见的是由一个方阵的幂构成的序列. 关于这样的矩阵序列, 有以下的概念和收敛定理.

定义 2.25　设 $\boldsymbol{A}$ 为方阵, 且 $\boldsymbol{A}^{(k)} \to \boldsymbol{O}(k \to \infty)$, 则称 $\boldsymbol{A}$ 为收敛矩阵.

定理 2.26　$\boldsymbol{A}$ 为收敛矩阵的充要条件是 $\rho(\boldsymbol{A}) < 1$.

证　充分性. 已知 $\rho(\boldsymbol{A}) < 1$, 对于 $\epsilon = \dfrac{1}{2}[1 - \rho(\boldsymbol{A})] > 0$, 存在矩阵范数 $\|\cdot\|_M$, 使得

$$\|\boldsymbol{A}\|_M \leqslant \rho(\boldsymbol{A}) + \epsilon = \frac{1}{2}[1 + \rho(\boldsymbol{A})] < 1$$

于是有 $\|\boldsymbol{A}^k\|_M \leqslant \|\boldsymbol{A}\|_M^k \to 0$, 可得 $\boldsymbol{A}^k \to \boldsymbol{O}$.

必要性. 已知 $\boldsymbol{A}^k \to \boldsymbol{O}$, 设 λ 是 $\boldsymbol{A}$ 的任一特征值, 对应的特征向量为 $\boldsymbol{x}$, 则有 $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}(\boldsymbol{x} \neq \boldsymbol{0})$. 因为

$$\lambda^k \boldsymbol{x} = \boldsymbol{A}^k \boldsymbol{x} \to \boldsymbol{0}$$

所以 $\lambda^k \to 0$, 从而 $|\lambda| < 1$, 故 $\rho(\boldsymbol{A}) < 1$.

定理 2.27　$\boldsymbol{A}$ 为收敛矩阵的充分条件是只要有一种矩阵范数 $\|\cdot\|$, 使得 $\|\boldsymbol{A}\| < 1$.

例 2.17 判断 $\boldsymbol{A}=\begin{bmatrix}0.1 & 0.3\\ 0.7 & 0.6\end{bmatrix}$ 是否为收敛矩阵.

解 因为 $\|\boldsymbol{A}\|_1 = 0.9 < 1$, 所以 $\boldsymbol{A}$ 是收敛矩阵.

2.3.2 矩阵级数介绍

在建立矩阵分析的理论时, 特别看重讨论矩阵级数, 特别是矩阵的幂级数, 因为它是建立矩阵函数的理论基础. 在讨论矩阵级数时, 自然应该定义它的收敛、发散以及和的概念. 这些都与数项级数的相应定义与性质完全类似.

定义 2.26 把定义 2.23 中的矩阵序列所形成的无穷和 $\boldsymbol{A}^{(0)}+\boldsymbol{A}^{(1)}+\boldsymbol{A}^{(2)}+\cdots+\boldsymbol{A}^{(k)}+\cdots$ 称为矩阵级数, 记为 $\sum\limits_{k=0}^{\infty}\boldsymbol{A}^{(k)}$, 则有

$$\sum_{k=0}^{\infty}\boldsymbol{A}^{(k)}=\boldsymbol{A}^{(0)}+\boldsymbol{A}^{(1)}+\boldsymbol{A}^{(2)}+\cdots+\boldsymbol{A}^{(k)}+\cdots \tag{2.3.1}$$

定义 2.27 记 $\boldsymbol{S}^{(N)}=\sum\limits_{k=0}^{N}\boldsymbol{A}^{(k)}$, 称其为上述矩阵级数式的部分和. 如果矩阵序列 $\{\boldsymbol{S}^{(N)}\}$ 收敛, 且有极限 $\boldsymbol{S}$, 则有

$$\lim_{N\to\infty}\boldsymbol{S}^{(N)}=\boldsymbol{S} \tag{2.3.2}$$

那么就称上述矩阵级数式收敛, 而且有和 $\boldsymbol{S}$, 记为

$$\boldsymbol{S}=\sum_{k=0}^{\infty}\boldsymbol{A}^{(k)} \tag{2.3.3}$$

不收敛的矩阵级数称为是发散的.

若用 s_{ij} 表示 $\boldsymbol{S}$ 的第 i 行第 j 列的元素, 那么, 和 $\sum\limits_{k=0}^{\infty}\boldsymbol{A}^{(k)}=\boldsymbol{S}$ 的意义指的是

$$\sum_{k=0}^{\infty}a_{ij}^{(k)}=s_{ij}\ (i=1,2,\cdots,m;j=1,2,\cdots,n) \tag{2.3.4}$$

例 2.18 研究矩阵级数 $\sum\limits_{k=1}^{\infty}\boldsymbol{A}^{(k)}$ 的收敛性, 其中

$$\boldsymbol{A}^{(k)}=\begin{bmatrix}\dfrac{1}{2^k} & \dfrac{\pi}{3\times 4^k}\\ 0 & \dfrac{1}{k(k+1)}\end{bmatrix}$$

解 因为

$$S^{(N)} = \sum_{k=1}^{N} A^{(k)} = \begin{bmatrix} 1-(\frac{1}{2})^N & \frac{\pi}{9}[1-(\frac{1}{4})^N] \\ 0 & \frac{N}{N+1} \end{bmatrix}$$

所以

$$S = \lim_{N\to\infty} S^{(N)} = \begin{bmatrix} 1 & \frac{\pi}{9} \\ 0 & 1 \end{bmatrix}$$

定义 2.28 如果式 (2.3.4) 中左端 mn 个数项级数都是绝对收敛的, 则称矩阵级数式 (2.3.1) 是绝对收敛的.

从绝对收敛的定义及数学分析中的相应结果, 立刻得到下面关于判别矩阵级数收敛性的一些法则.

性质 1 若矩阵级数式 (2.3.1) 是绝对收敛的, 则它也一定收敛, 并且任意调换其项的顺序所得的级数还是收敛的, 且其和不变.

性质 2 矩阵级数 $\sum\limits_{k=0}^{\infty} A^{(k)}$ 为绝对收敛的充要条件是正项级数 $\sum\limits_{k=0}^{\infty} \|A^{(k)}\|$ 收敛.

性质 3 如果 $\sum\limits_{k=0}^{\infty} A^{(k)}$ 是收敛 (或绝对收敛) 的, 那么 $\sum\limits_{k=0}^{N} PA^{(k)}Q$ 也是收敛 (或绝对收敛) 的, 并且有

$$\sum_{k=0}^{N} PA^{(k)}Q = P(\sum_{k=0}^{\infty} A^{(k)})Q$$

性质 4 设 $C^{m\times n}$ 中的两个矩阵级数

$$S_1 : A^{(1)} + A^{(2)} + \cdots + A^{(k)} + \cdots$$

$$S_2 : B^{(1)} + B^{(2)} + \cdots + B^{(k)} + \cdots$$

都绝对收敛, 其和分别为 A 与 B, 则级数 S_1 与 S_2 按项相乘所得的矩阵级数

$$S_3 := \sum_{k=1}^{\infty} (\sum_{i=1}^{k} A^{(i)} B^{(k+1-i)})$$

绝对收敛, 且有和 AB.

下面讨论矩阵幂级数, 首先从一个比较简单的方阵幂级数谈起.

定理 2.28 方阵 A 的幂级数 (Neumann 级数)

$$\sum_{k=0}^{\infty} A^k = I + A + A^2 + \cdots + A^k + \cdots$$

收敛的充要条件是 A 为收敛矩阵, 并且咋收敛时, 其和为 $(I - A)^{-1}$.

定理 2.29 设方阵 $\boldsymbol{A}$ 对某一矩阵范数 $\|\cdot\|$ 有 $\|\boldsymbol{A}\| < 1$, 则对任何非负整数 N, 以 $(\boldsymbol{I}-\boldsymbol{A})^{-1}$ 为部分和 $\boldsymbol{I}+\boldsymbol{A}+\boldsymbol{A}^2+\cdots+\boldsymbol{A}^N$ 的近似矩阵时, 其误差为

$$\|(\boldsymbol{I}-\boldsymbol{A})^{-1}-(\boldsymbol{I}+\boldsymbol{A}+\boldsymbol{A}^2+\cdots+\boldsymbol{A}^N)\| \leqslant \frac{\|\boldsymbol{A}\|^{N+1}}{1-\|\boldsymbol{A}\|}$$

定理 2.30 设幂级数

$$f(z)=\sum_{k=0}^{\infty} c_k z^k$$

的收敛半径为 r, 如果方阵 $\boldsymbol{A}$ 满足 $\rho(\boldsymbol{A}) < r$, 则矩阵幂级数

$$\sum_{k=0}^{\infty} c_k \boldsymbol{A}^k$$

是绝对收敛的; 如果 $\rho(\boldsymbol{A}) > r$, 则矩阵幂级数是发散的.

证 (1) 当 $\rho(\boldsymbol{A}) < r$ 时, 选定正数 ϵ, 使满足 $\rho(\boldsymbol{A})+\epsilon < r$. 则存在矩阵范数 $\|\cdot\|$, 使得 $\|\boldsymbol{A}\| \leqslant \rho(\boldsymbol{A})+\epsilon$. 从而有

$$\|c_k\boldsymbol{A}^k\| \leqslant |c_k|\ \|\boldsymbol{A}\|^k \leqslant |c_k|(\rho(\boldsymbol{A})+\epsilon)^k$$

因为 $\rho(\boldsymbol{A})+\epsilon < r$, 所以 $\sum\limits_{k=0}^{\infty} c_k(\rho(\boldsymbol{A})+\epsilon)^k$ 绝对收敛, 从而 $\sum\limits_{k=0}^{\infty}\|c_k\boldsymbol{A}^k\|$ 收敛. 从而 $\sum\limits_{k=0}^{\infty} c_k\boldsymbol{A}^k$ 绝对收敛.

(2) 当 $\rho(\boldsymbol{A}) > r$ 时, 设 $\boldsymbol{A}$ 的 n 个特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_n$, 则有某个 λ_l 满足 $|\lambda_l| > r$. 则存在可逆矩阵 $\boldsymbol{P}$, 使得

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\begin{bmatrix}\lambda_1 & * & \cdots & * \\ & \lambda_2 & \ddots & \vdots \\ & & \ddots & * \\ & & & \lambda_n\end{bmatrix}$$

而 $\sum\limits_{k=0}^{\infty} c_k\boldsymbol{B}^k$ 的对角线元素为 $\sum\limits_{k=0}^{\infty} c_k\lambda_i^k (i=1,2,\cdots,n)$. 因为 $|\lambda_l| > r$, 所以 $\sum c_k\lambda_l^k$ 发散, 从而 $\sum\limits_{k=0}^{\infty} c_k\boldsymbol{B}^k$ 发散. 从而 $\sum\limits_{k=0}^{\infty} c_k\boldsymbol{A}^k=\sum\limits_{k=0}^{\infty} c_k\boldsymbol{P}\boldsymbol{B}^k\boldsymbol{P}^{-1}$ 也发散.

2.3.3 矩阵函数介绍

矩阵函数的概念与通常的函数概念一样, 它是以 n 阶矩阵为自变量和函数值 (因变量) 的一种函数. 本节给出矩阵函数的定义及其性质.

定义 2.29　设一元函数 $f(z)$ 能够展开为 z 的幂级数

$$f(z) = \sum_{k=0}^{\infty} c_k z^k \ (|z| < r)$$

其中 $r > 0$ 表示该幂级数的收敛半径. 当 n 阶矩阵 $\boldsymbol{A}$ 的谱半径 $\rho(\boldsymbol{A}) < r$ 时, 把收敛的矩阵幂级数 $\sum\limits_{k=0}^{\infty} c_k \boldsymbol{A}^k$ 的和称为矩阵函数, 记为 $f(\boldsymbol{A})$, 即

$$f(\boldsymbol{A}) = \sum_{k=0}^{\infty} c_k \boldsymbol{A}^k$$

定理 2.31　如果 $\boldsymbol{AB} = \boldsymbol{BA}$, 则 $e^{\boldsymbol{A}}e^{\boldsymbol{B}} = e^{\boldsymbol{B}}e^{\boldsymbol{A}} = e^{\boldsymbol{A}+\boldsymbol{B}}$.

证　因为矩阵加法满足交换律, 所以只需证明 $e^{\boldsymbol{A}}e^{\boldsymbol{B}} = e^{\boldsymbol{A}+\boldsymbol{B}}$ 就行了.

$$\begin{aligned}
e^{\boldsymbol{A}}e^{\boldsymbol{B}} =&(\boldsymbol{I} + \boldsymbol{A} + \frac{1}{2!}\boldsymbol{A}^2 + \cdots)(\boldsymbol{I} + \boldsymbol{B} + \frac{1}{2!}\boldsymbol{B}^2 + \cdots) = \\
&\boldsymbol{I} + (\boldsymbol{A} + \boldsymbol{B}) + \frac{1}{2!}(\boldsymbol{A}^2 + \boldsymbol{AB} + \boldsymbol{BA} + \boldsymbol{B}^2) + \\
&\frac{1}{3!}(\boldsymbol{A}^3 + 3\boldsymbol{A}^2\boldsymbol{B} + 3\boldsymbol{AB}^2 + \boldsymbol{B}^3) + \cdots = \\
&\boldsymbol{I} + (\boldsymbol{A} + \boldsymbol{B}) + \frac{1}{2!}(\boldsymbol{A} + \boldsymbol{B})^2 + \frac{1}{3!}(\boldsymbol{A} + \boldsymbol{B})^3 + \cdots = e^{\boldsymbol{A}+\boldsymbol{B}}
\end{aligned}$$

例 2.19　设函数 $f(z) = \dfrac{1}{1-z}(|z| < 1)$, 求矩阵函数 $f(\boldsymbol{A})$.

解　因为

$$f(z) = \frac{1}{1-z} = \sum_{k=0}^{\infty} z^k \ (|z| < 1)$$

当方阵 $\boldsymbol{A}$ 的谱半径 $\rho(\boldsymbol{A}) < 1$ 时, 有

$$f(\boldsymbol{A}) = \sum_{k=0}^{\infty} \boldsymbol{A}^k$$

从而 $f(\boldsymbol{A}) = (\boldsymbol{I} - \boldsymbol{A})^{-1}$.

2.3.4　函数矩阵对矩阵的导数

定义 2.30　设 $\boldsymbol{X} = (\xi_{ij})_{m\times n}$, mn 元函数 $f(\boldsymbol{X}) = f(\xi_{11}, \xi_{12}, \cdots, \xi_{1n}, \xi_{21}, \cdots, \xi_{mn})$, 定义 $f(\boldsymbol{X})$ 对矩阵 $\boldsymbol{X}$ 的导数为

$$\frac{df}{d\boldsymbol{X}} = (\frac{\partial f}{\partial \xi_{ij}})_{m\times n} = \begin{bmatrix} \dfrac{\partial f}{\partial \xi_{11}} & \cdots & \dfrac{\partial f}{\partial \xi_{1n}} \\ \vdots & & \vdots \\ \dfrac{\partial f}{\partial \xi_{m1}} & \cdots & \dfrac{\partial f}{\partial \xi_{mn}} \end{bmatrix}$$

例 2.20　设 $\boldsymbol{x} = (\xi_1, \xi_2, \cdots, \xi_n)^T$, n 元函数 $f(\boldsymbol{x}) = f(\xi_1, \xi_2, \cdots, \xi_n)$, 求 $\frac{df}{d\boldsymbol{x}}$ 与 $\frac{df}{d\boldsymbol{x}^T}$.

解　根据定义, 有

$$\frac{df}{d\boldsymbol{x}} = (\frac{\partial f}{\partial \xi_1}, \frac{\partial f}{\partial \xi_2}, \cdots, \frac{\partial f}{\partial \xi_n})^T, \quad \frac{df}{d\boldsymbol{x}^T} = (\frac{\partial f}{\partial \xi_1}, \frac{\partial f}{\partial \xi_2}, \cdots, \frac{\partial f}{\partial \xi_n})$$

定义 2.31　设 $\boldsymbol{X} = (\xi_{ij})_{m\times n}$, mn 元函数 $f_{ij}(\boldsymbol{X}) = f_{ij}(\xi_{11}, \xi_{12}, \cdots, \xi_{1n}, \xi_{21}, \cdots, \xi_{mn})$, 其中 i 和 j 有如下范围 $(i = 1, 2, \cdots, r; j = 1, 2, \cdots, s)$. 定义函数矩阵

$$\boldsymbol{F}(\boldsymbol{X}) = \begin{bmatrix} f_{11}(\boldsymbol{X}) & \cdots & f_{1s}(\boldsymbol{X}) \\ \vdots & & \vdots \\ f_{r1}(\boldsymbol{X}) & \cdots & f_{rs}(\boldsymbol{X}) \end{bmatrix}$$

对矩阵 $\boldsymbol{X}$ 的导数为

$$\frac{d\boldsymbol{F}}{d\boldsymbol{X}} = \begin{bmatrix} \frac{\partial \boldsymbol{F}}{\partial xi_{11}} & \frac{\partial \boldsymbol{F}}{\partial xi_{12}} & \cdots & \frac{\partial \boldsymbol{F}}{\partial xi_{1n}} \\ \frac{\partial \boldsymbol{F}}{\partial xi_{21}} & \frac{\partial \boldsymbol{F}}{\partial xi_{22}} & \cdots & \frac{\partial \boldsymbol{F}}{\partial xi_{2n}} \\ \vdots & \vdots & & \vdots \frac{\partial \boldsymbol{F}}{\partial xi_{m1}} & \frac{\partial \boldsymbol{F}}{\partial xi_{m2}} & \cdots & \frac{\partial \boldsymbol{F}}{\partial xi_{mn}} \end{bmatrix}$$

其中

$$\frac{\partial \boldsymbol{F}}{\partial \xi_{ij}} = \begin{bmatrix} \frac{\partial f_{11}}{\partial \xi_{ij}} & \frac{\partial f_{12}}{\partial \xi_{ij}} & \cdots & \frac{\partial f_{1s}}{\partial \xi_{ij}} \\ \frac{\partial f_{21}}{\partial \xi_{ij}} & \frac{\partial f_{22}}{\partial \xi_{ij}} & \cdots & \frac{\partial f_{2s}}{\partial \xi_{ij}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_{r1}}{\partial \xi_{ij}} & \frac{\partial f_{r2}}{\partial \xi_{ij}} & \cdots & \frac{\partial f_{rs}}{\partial \xi_{ij}} \end{bmatrix}$$

例 2.21　设 $\boldsymbol{x} = (\xi_1, \xi_2, \cdots, \xi_n)^T$, n 元函数 $f(\boldsymbol{x}) = f(\xi_1, \xi_2, \cdots, \xi_n)$, 求 $\frac{d}{d\boldsymbol{x}^T}(\frac{df}{d\boldsymbol{x}})$.

解　由例 2.20 知

$$\frac{df}{d\boldsymbol{x}} = (\frac{\partial f}{\partial \xi_1}, \frac{\partial f}{\partial \xi_2}, \cdots, \frac{\partial f}{\partial \xi_n})^T$$

可得

$$\frac{d}{d\boldsymbol{x}^T}(\frac{df}{d\boldsymbol{x}}) = \begin{bmatrix} \frac{\partial^2 f}{\partial \xi_1^2} & \frac{\partial^2 f}{\partial \xi_1 \partial \xi_2} & \cdots & \frac{\partial f}{\partial \xi_1 \partial \xi_n} \\ \frac{\partial^2 f}{\partial \xi_2 \partial \xi_1} & \frac{\partial^2 f}{\partial \xi_2^2} & \cdots & \frac{\partial f}{\partial \xi_2 \partial \xi_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial \xi_n \partial \xi_1} & \frac{\partial^2 f}{\partial \xi_n \partial \xi_2} & \cdots & \frac{\partial f}{\partial \xi_n^2} \end{bmatrix}$$

第 3 章　矩阵函数的求法研究

本章探讨第一个问题, 已知 $\boldsymbol{A}$, 怎样计算矩阵函数值得问题.

3.1　待定系数法

3.1.1　待定系数法求矩阵函数的步骤推导

设 n 阶矩阵 $\boldsymbol{A}$ 的特征多项式为 $\varphi(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A})$. 如果首 1 多项式

$$\psi(\lambda) = \lambda^m + b_1\lambda^{m-1} + \cdots + b_{m-1}\lambda + b_m \quad (1 \leqslant m \leqslant n)$$

满足 $\psi(\boldsymbol{A}) = \boldsymbol{O}$ 且 $\psi(\lambda)$ 整除 $\varphi(\lambda)$(矩阵 $\boldsymbol{A}$ 的最小多项式与特征多项式均满足这些条件). 那么, $\psi(\lambda)$ 的零点就是 $\boldsymbol{A}$ 的特征值. 记 $\psi(\lambda)$ 的互异零点为 $\lambda_1, \lambda_2, \cdots, \lambda_s$, 相应的重数为 $r_1, r_2, \cdots, r_s$, 且 $r_1 + r_2 + \cdots + r_s = m$, 则有

$$\psi^{(l)}(\lambda_i) = 0 \quad (l = 0, 1, \cdots, r_i - 1; i = 1, 2, \cdots, s)$$

这里, $\psi^{(l)}(\lambda)$ 表示 $\psi(\lambda)$ 的 l 阶导数 (下同). 设

$$f(z) = \sum_{k=0}^{\infty} c_k z^k = \psi(z)g(z) + r(z)$$

其中 $r(z)$ 是次数低于 m 的多项式, 于是可由

$$f^{(l)}(\lambda_i) = r^{l}(\lambda_i) \quad (l = 0, 1, \cdots, r_i - 1; i = 1, 2, \cdots, s)$$

确定出 $r(z)$. 利用 $\psi(\boldsymbol{A}) = \boldsymbol{O}$, 可得

$$f(\boldsymbol{A}) = \sum_{k=0}^{\infty} c_k \boldsymbol{A}^k = r(\boldsymbol{A})$$

3.1.2　举例展示求法

例 3.1　设 $\boldsymbol{A} = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & 3 \end{bmatrix}$, 求 $e^{\boldsymbol{A}}$ 与 $e^{t\boldsymbol{A}}(t \in \mathbb{R})$.

解　$\varphi(\lambda) - \det(\lambda \boldsymbol{I} - \boldsymbol{A}) = (\lambda - 2)^3$, 容易求得 $\boldsymbol{A}$ 的最小多项式 $m(\lambda) = (\lambda - 2)^2$, 取 $\psi(\lambda) = (\lambda - 2)^2$.

(1) 取 $f(\lambda) = e^{\lambda}$, 设 $f(\lambda) = \psi(\lambda)g(\lambda) + (a + b\lambda)$, 则有

$$\begin{cases} f(2) = e^2 \\ f'(2) = e^2 \end{cases} \Rightarrow \begin{cases} a + 2b = e^2 \\ b = e^2 \end{cases}$$

解此方程组可得 $a=-e^2, b=e^2$. 于是 $r(\lambda)=e^2(\lambda-1)$, 从而

$$e^{\boldsymbol{A}}=f(\boldsymbol{A})=r(\boldsymbol{A})=e^2(\boldsymbol{A}-\boldsymbol{I})=e^2\begin{bmatrix}1&0&0\\1&0&1\\1&-1&2\end{bmatrix}$$

(2) 取 $f(\lambda)=e^{t\lambda}$, 设 $f(\lambda)=\psi(\lambda)g(\lambda)+(a+b\lambda)$, 则有

$$\begin{cases}f(2)=e^{2t}\\f'(2)=te^{2t}\end{cases}\Rightarrow\begin{cases}a+2b=e^{2t}\\b=te^{2t}\end{cases}$$

解此方程组可得 $a=(1-2t)e^{2t}, b=te^{2t}$. 于是 $r(\lambda)=e^{2t}[(1-2t)+t\lambda]$, 从而

$$e^{t\boldsymbol{A}}=f(\boldsymbol{A})=r(\boldsymbol{A})=e^{2t}[(1-2t)\boldsymbol{I}+t\boldsymbol{A}]=e^{2t}\begin{bmatrix}1&0&0\\t&1-t&t\\t&-t&1+t\end{bmatrix}$$

3.2 数项级数求和法

3.2.1 数项级数求和法求矩阵函数的步骤推导

设首 1 多项式 $\psi(\lambda)=\lambda^m+b_1\lambda^{m-1}+\cdots+b_{m-1}\lambda+b_m \quad (1\leqslant m\leqslant n)$ 且满足 $\psi(\boldsymbol{A})=\boldsymbol{O}$, 即

$$\boldsymbol{A}^m+b_1\boldsymbol{A}^{m-1}+\cdots+b_{m-1}\boldsymbol{A}+b_m\boldsymbol{I}=\boldsymbol{O}$$

或者

$$\boldsymbol{A}^m=k_0\boldsymbol{I}+k_1\boldsymbol{A}+\cdots+k_{m-1}\boldsymbol{A}^{m-1}\quad(k_i=-b_{m-i}) \tag{3.2.1}$$

由此可以求出

$$\begin{cases}\boldsymbol{A}^{m+1} &= k_0^{(1)}\boldsymbol{I}+k_1^{(1)}\boldsymbol{A}+\cdots+k_{m-1}^{(1)}\boldsymbol{A}^{m-1}\\ &\cdots\\ \boldsymbol{A}^{m+l} &= k_0^{(l)}\boldsymbol{I}+k_1^{(l)}\boldsymbol{A}+\cdots+k_{m-1}^{(l)}\boldsymbol{A}^{m-1}\\ &\cdots\end{cases}$$

于是有

$$\begin{aligned}f(\boldsymbol{A})=\sum_{k=0}^{\infty}c_k\boldsymbol{A}^k&=(c_0\boldsymbol{I}+c_1\boldsymbol{A}+\cdots+c_{m-1}\boldsymbol{A}^{m-1})+\\&c_m(k_0\boldsymbol{I}+k_1\boldsymbol{A}+\cdots+k_{m-1}\boldsymbol{A}^{m-1})+\cdots+\\&c_{m+l}(k_0^{(l)}\boldsymbol{I}+k_1^{(l)}\boldsymbol{A}+\cdots+k_{m-1}^{l}\boldsymbol{A}^{m-1})+\cdots=\\&(c_0+\sum_{l=0}^{\infty}c_{m+l}k_0^{(l)})\boldsymbol{I}+(c_1+\sum_{l=0}^{\infty}c_{m+l}k_1^{(l)})\boldsymbol{A}+\cdots+\\&(c_{m-1}+\sum_{l=0}^{\infty}c_{m+l}k_{m-1}^{(l)})\boldsymbol{A}^{m-1}\end{aligned}$$

这表明, 利用上式可以将一个矩阵幂级数的求和问题, 转换为 m 个数项级数的求和问题. 当式 (3.2.1) 中只有少数几个系数不为零时, 上式中需要计算的数项级数也只有少数几个.

3.2.2 举例展示求法

例 3.2 设 $\boldsymbol{A}=\begin{bmatrix}\pi & 0 & 0 & 0\\ 0 & -\pi & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0\end{bmatrix}$, 求 $\sin\boldsymbol{A}$.

解 $\varphi(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})=\lambda^4-\pi^2\lambda^2$. 由于 $\varphi(\boldsymbol{A})=\boldsymbol{O}$, 所以 $\boldsymbol{A}^4=\pi^2\boldsymbol{A}^2, \boldsymbol{A}^5=\pi^2\boldsymbol{A}^3, \boldsymbol{A}^7=\pi^4\boldsymbol{A}^3, \cdots$. 于是有

$$
\begin{aligned}
\sin\boldsymbol{A}=&\boldsymbol{A}-\frac{1}{3!}\boldsymbol{A}^3+\frac{1}{5!}\boldsymbol{A}^5-\frac{1}{7!}\boldsymbol{A}^7+\frac{1}{9!}\boldsymbol{A}^9-\cdots=\\
&\boldsymbol{A}-\frac{1}{3!}\boldsymbol{A}^3+\frac{1}{5!}\pi^2\boldsymbol{A}^3-\frac{1}{7!}\pi^4\boldsymbol{A}^3+\frac{1}{9}\pi^6\boldsymbol{A}^3-\cdots=\\
&\boldsymbol{A}+(-\frac{1}{3!}+\frac{1}{5!}\pi^2-\frac{1}{7!}\pi^4+\frac{1}{9!}\pi^6-\cdots)\boldsymbol{A}^3=\\
&\boldsymbol{A}+\frac{\sin\pi-\pi}{\pi^3}\boldsymbol{A}^3=\boldsymbol{A}-\pi^{-2}\boldsymbol{A}^3=\begin{bmatrix}0&0&0&0\\0&0&0&0\\0&0&0&1\\0&0&0&0\end{bmatrix}
\end{aligned}
$$

3.3 对角型法

3.3.1 对角型法求矩阵函数的步骤推导

设 $\boldsymbol{A}$ 相似于对角矩阵 Λ, 即有可逆矩阵 $\boldsymbol{P}$, 使得

$$
\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}
$$

则有 $\boldsymbol{A}=\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}, \boldsymbol{A}^2=\boldsymbol{P}\boldsymbol{\Lambda}^2\boldsymbol{P}^{-1}, \cdots$, 于是可得

$$
\sum_{k=0}^{N}c_k\boldsymbol{A}^k=\sum_{k=0}^{N}c_k\boldsymbol{P}\boldsymbol{\Lambda}^k\boldsymbol{P}^{-1}=\boldsymbol{P}\cdot\sum_{k=0}^{N}c_k\boldsymbol{\Lambda}^k\boldsymbol{P}^{k-1}=\boldsymbol{P}\begin{bmatrix}\sum\limits_{k=0}^{N}c_k\lambda_1^k & & \\ & \ddots & \\ & & \sum\limits_{k=0}^{N}c_k\lambda_n^k\end{bmatrix}\boldsymbol{P}^{-1}
$$

从而

$$f(\boldsymbol{A})=\sum_{k=0}^{\infty}c_k\boldsymbol{A}^k=\boldsymbol{P}\begin{bmatrix}\sum_{k=0}^{N}c_k\lambda_1^k & & \\ & \ddots & \\ & & \sum_{k=0}^{N}c_k\lambda_n^k\end{bmatrix}\boldsymbol{P}^{-1}=\boldsymbol{P}\begin{bmatrix}f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n)\end{bmatrix}\boldsymbol{P}^{-1}$$

这表明, 当 $\boldsymbol{A}$ 与对角矩阵相似时, 可以将矩阵幂级数的求和问题转化为求相似变换矩阵的问题.

3.3.2 举例展示求法

例 3.3 设 $\boldsymbol{A}=\begin{bmatrix}4 & 6 & 0\\ -3 & -5 & 0\\ -3 & -6 & 0\end{bmatrix}$, 分别求 $e^{\boldsymbol{A}}, e^{t\boldsymbol{A}}(t\in\mathbb{R})$ 及 $\cos\boldsymbol{A}$.

解 $\varphi(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})=(\lambda+2)(\lambda-1)^2$. 对应 $\lambda_1=-1$ 的特征向量 $\boldsymbol{p}_1=(-1,1,1)^T$; 对应 $\lambda_2=\lambda_3=1$ 的两个线性无关的特征向量 $\boldsymbol{p}_2=(-2,1,0)^T, \boldsymbol{p}_3=(0,0,1)^T$. 构造矩阵

$$\boldsymbol{P}=(\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3)=\begin{bmatrix}-1 & -2 & 0\\ 1 & 1 & 0\\ 1 & 0 & 1\end{bmatrix}$$

则有

$$\boldsymbol{P}^{-1}=\begin{bmatrix}1 & 2 & 0\\ -1 & -1 & 0\\ -1 & -2 & 1\end{bmatrix},\ \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\begin{bmatrix}-2 & & \\ & 1 & \\ & & 1\end{bmatrix}$$

利用上述公式, 求得

$$e^{\boldsymbol{A}}=\boldsymbol{P}\begin{bmatrix}e^{-2} & & \\ & e & \\ & & e\end{bmatrix}\boldsymbol{P}^{-1}=\begin{bmatrix}2e-e^{-2} & 2e-2e^{-2} & 0\\ e^{-2}-e & 2e^{-2}-e & 0\\ e^{-2}-e & 2e^{-2}-2e & e\end{bmatrix}$$

$$e^{t\boldsymbol{A}}=\boldsymbol{P}\begin{bmatrix}e^{-2t} & & \\ & e^{t} & \\ & & e^{t}\end{bmatrix}\boldsymbol{P}^{-1}=\begin{bmatrix}2e^{t}-e^{-2t} & 2e^{t}-2e^{-2t} & 0\\ e^{-2t}-e^{t} & 2e^{-2t}-e^{t} & 0\\ e^{-2t}-e^{t} & 2e^{-2}-2e^{t} & e^{t}\end{bmatrix}$$

$$\cos\boldsymbol{A}=\boldsymbol{P}\begin{bmatrix}\cos(-2) & & \\ & \cos 1 & \\ & & \cos 1\end{bmatrix}\boldsymbol{P}^{-1}=$$

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

例 3.4 设 $\boldsymbol{A} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, 求 $e^{\boldsymbol{A}}, e^{t\boldsymbol{A}}(t \in \mathbb{R}), \sin \boldsymbol{A}$.

解 $\varphi(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A}) = (\lambda+1)(\lambda-1)(\lambda-2)$. 对应 $\lambda_1 = -1$ 的特征向量 $\boldsymbol{p}_1 = (1, -3, 3)^T$, 对应 $\lambda_2 = 1$ 的特征向量 $\boldsymbol{p}_2 = (-1, 1, 1)^T$, 对应 $\lambda_3 = -2$ 的特征向量 $\boldsymbol{p}_3 = (1, 3, 2)^T$. 构造矩阵

$$\boldsymbol{P} = (\boldsymbol{p}_1, \boldsymbol{p}_2, bmp_3) = \begin{bmatrix} 1 & -1 & 1 \\ -3 & 1 & 0 \\ 3 & 1 & 0 \end{bmatrix}$$

则有

$$\boldsymbol{P}^{-1} = \frac{1}{6}\begin{bmatrix} 0 & -1 & 1 \\ 0 & 3 & 3 \\ 6 & 4 & 2 \end{bmatrix}, \ \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} -1 & & \\ & 1 & \\ & & 2 \end{bmatrix}$$

求得

$$e^{\boldsymbol{A}} = \boldsymbol{P}\begin{bmatrix} e^{-1} & & \\ & e & \\ & & e^2 \end{bmatrix}\boldsymbol{P}^{-1} = \frac{1}{6}\begin{bmatrix} 6e^2 & 4e^2 - 3e - e^{-1} & 2e^2 - 3e + e^{-1} \\ 0 & 3e + 3e^{-1} & 3e - 3e^{-1} \\ 0 & 3e - 3e^{-1} & 3e + 3e^{-1} \end{bmatrix}$$

$$e^{t\boldsymbol{A}} = \boldsymbol{P}\begin{bmatrix} e^{-t} & & \\ & e^t & \\ & & e^{2t} \end{bmatrix}\boldsymbol{P}^{-1} = \frac{1}{6}\begin{bmatrix} 6e^{2t} & 4e^{2t} - 3e^t - e^{-t} & 2e^{2t} - 3e^t + e^{-t} \\ 0 & 3e^t + 3e^{-t} & 3e^t - 3e^{-t} \\ 0 & 3e^t - 3e^{-t} & 3e^t + 3e^{-t} \end{bmatrix}$$

$$\sin \boldsymbol{A} = \boldsymbol{P}\begin{bmatrix} \sin(-1) & & \\ & \sin 1 & \\ & & \sin 2 \end{bmatrix}\boldsymbol{P}^{-1} = \frac{1}{6}\begin{bmatrix} 6\sin 2 & 4\sin 2 - 2\sin 1 & 2\sin 2 - 4\sin 1 \\ 0 & 0 & 6\sin 1 \\ 0 & 6\sin 1 & 0 \end{bmatrix}$$

3.4 若尔当标准型法

3.4.1 若尔当标准型法求矩阵函数的步骤推导

设 $\boldsymbol{A}$ 的 Jordan 标准形为 $\boldsymbol{J}$, 则有可逆矩阵 $\boldsymbol{P}$, 使得

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J} = \begin{bmatrix} \boldsymbol{J}_1 & & \\ & \ddots & \\ & & \boldsymbol{J}_s \end{bmatrix}$$

其中

$$\boldsymbol{J}_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ & & & \lambda_i \end{bmatrix}_{m_i \times m_i}$$

可求得

$$f(\boldsymbol{J}_i) = \sum_{k=0}^{\infty} c_k \boldsymbol{J}_i^k = \sum_{k=0}^{\infty} c_k \begin{bmatrix} \lambda_i^k & C_k^1 \lambda_i^{k-1} & \cdots & C_k^{m_i-1} \lambda_i^{k-m_i+1} \\ & \lambda_i^k & \ddots & \vdots \\ & & \ddots & C_k^1 \lambda_i^{k-1} \\ & & & \lambda_i^k \end{bmatrix} =$$

$$\begin{bmatrix} f(\lambda_i) & \frac{1}{1!} f'(\lambda_i) & \cdots & \frac{1}{(m_i-1)!} f^{m_i-1}(\lambda_i) \\ & f(\lambda_i) & \ddots & \vdots \\ & & \ddots & \frac{1}{1!} f'(\lambda_i) \\ & & & f(\lambda_i) \end{bmatrix}$$

$$f(\boldsymbol{A}) = \sum_{k=0}^{\infty} c_k \boldsymbol{A}^k = \sum_{k=0}^{\infty} c_k \boldsymbol{P} \boldsymbol{J}^k \boldsymbol{P}^{-1} = \boldsymbol{P}(\sum_{k=0}^{\infty} c_k \boldsymbol{J}^k) \boldsymbol{P}^{-1} =$$

$$\boldsymbol{P} \begin{bmatrix} \sum_{k=0}^{\infty} c_k \boldsymbol{J}_1^k & & \\ & \ddots & \\ & & \sum_{k=0}^{\infty} c_k \boldsymbol{J}_s^k \end{bmatrix} \boldsymbol{P}^{-1} = \boldsymbol{P} \begin{bmatrix} f(\boldsymbol{J}_1) & & \\ & \ddots & \\ & & f(\boldsymbol{J}_s) \end{bmatrix} \boldsymbol{P}^{-1}$$

这表明, 矩阵幂级数的求和问题可以转化为求矩阵的 Jordan 标准形及相似变换矩阵的问题.

3.4.2 举例展示求法

例 3.5 设 $\boldsymbol{A} = \begin{bmatrix} \pi & 0 & 0 & 0 \\ 0 & -\pi & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, 求 $\sin \boldsymbol{A}$.

解 矩阵 $\boldsymbol{A}$ 是一个 Jordan 标准形, 它的三个 Jordan 块为

$$\boldsymbol{J} = \pi, \ \boldsymbol{J}_2 = -\pi, \ \boldsymbol{J}_3 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

可以求得

$$\sin \boldsymbol{J}_1 = \sin\pi = 0,\ \ \sin \boldsymbol{J}_2 = \sin(-\pi) = 0$$

$$\sin \boldsymbol{J}_3 = \begin{bmatrix} \sin 0 & \frac{1}{1!}\cos 0 \\ 0 & \sin 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

取 $\boldsymbol{P} = \boldsymbol{I}$, 可得

$$\sin \boldsymbol{A} = \begin{bmatrix} \sin \boldsymbol{J}_1 & & \\ & \sin \boldsymbol{J}_2 & \\ & & \sin \boldsymbol{J}_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

第 4 章　矩阵分解方法研究

本章首先讨论以 Gauss 消去法为基础导出的矩阵的三角 (或 LU) 分解, 然后论述 20 世纪 60 年代后根据 Givens 变换与 Householder 变换发展起来的矩阵的 QR 分解. 这些分解方法在计算数学领域扮演着十分重要的角色, 尤其是以 QR 分解所建立的 QR 方法, 已对数值线性代数理论的近代发展呢起了关键作用. 最后介绍在广义逆矩阵等理论中, 经常遇到的矩阵的满秩分解和奇异值分解, 它与 QR 方法都是求解各类最小二乘问题和最优化问题的重要数学工具.

4.1　矩阵的 LU 分解

4.1.1　矩阵 LU 分解的步骤推导

设 $\boldsymbol{A}^{(0)} = \boldsymbol{A}$, 其元素 $a_{ij}^{(0)} = a_{ij}(i, j = 1, 2, \cdots, n)$. 记 $\boldsymbol{A}$ 的 k 阶顺序主子式为 $\Delta_k(k = 1, 2, \cdots, n)$. 如果 $\Delta_1 = a_{11}^{(0)} \neq 0$, 令 $c_{i1} = \dfrac{a_{i1}^{(0)}}{a_{11}^{(0)}}(i = 2, 3, \cdots, n)$, 并构造 Frobenius 矩阵

$$\boldsymbol{L}_1 = \begin{bmatrix} 1 & & & \\ c_{21} & 1 & & \\ \vdots & & \ddots & \\ c_{n1} & & & 1 \end{bmatrix}, \ \boldsymbol{L}_1^{-1} = \begin{bmatrix} 1 & & & \\ -c_{21} & 1 & & \\ \vdots & & \ddots & \\ -c_{n1} & & & 1 \end{bmatrix}$$

计算

$$\boldsymbol{L}_1^{-1}\boldsymbol{A}^{(0)} = \begin{bmatrix} a_{11}^{(0)} & a_{12}^{(0)} & \cdots & a_{1n}^{(0)} \\ & a_{22}^{(1)} & \cdots & a_{2n}^{(1)} \\ & \vdots & & \vdots \\ & a_{n2}^{(1)} & \cdots & a_{nn}^{(1)} \end{bmatrix} = \boldsymbol{A}^{(1)}$$

由此可见, $\boldsymbol{A}^{(0)} = \boldsymbol{A}$ 的第一列除主元 $a_{11}^{(0)}$ 外, 其余元素全被化为零. 上式还可写为

$$\boldsymbol{A}^{(0)} = \boldsymbol{L}_1\boldsymbol{A}^{(1)}$$

因为倍加初等变换不改变矩阵行列式的值, 所以由 $\boldsymbol{A}^{(1)}$ 得 $\boldsymbol{A}$ 得二阶顺序主子式为

$$\delta_2 = a_{11}^{(0)}a_{22}^{(1)}$$

如果 $\delta_2 \neq 0$, 则 $a_{22}^{(1)} \neq 0$. 令 $c_{i2} = \dfrac{a_{i2}^{(1)}}{a_{22}^{(1)}}(i = 3, 4, \cdots, n)$, 并构造 Frobenius 矩阵

$$\boldsymbol{L}_2 = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & c_{32} & 1 & & \\ & \vdots & & \ddots & \\ & c_{n2} & & & 1 \end{bmatrix}, \ \boldsymbol{L}_2^{-1} = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & -c_{32} & 1 & & \\ & \vdots & & \ddots & \\ & -c_{n2} & & & 1 \end{bmatrix}$$

由此可见, $\boldsymbol{A}^{(2)}$ 的前两列中主元以下的元素全为零. 上式还可写为

$$\boldsymbol{A}^{(1)} = \boldsymbol{L}_2\boldsymbol{A}^{(2)}$$

因为倍加初等变换不改变矩阵行列式的值, 所以由 $\boldsymbol{A}^{(2)}$ 得 $\boldsymbol{A}$ 的三阶顺序主子式为

$$\Delta_3 = a_{11}^{(0)}a_{22}^{(1)}a_{33}^{(3)}$$

如果可以一直进行下去, 则在第 $n-1$ 步之后便有

$$\boldsymbol{A}^{(n-1)} = \begin{bmatrix} a_{11}^{(0)} & a_{12}^{(0)} & \cdots & a_{1,n-1}^{(0)} & a_{1n}^{(0)} \\ & a_{22}^{(1)} & \cdots & a_{2,n-1}^{(1)} & a_{2,n}^{(1)} \\ & & \ddots & \vdots & \vdots \\ & & & a_{n-1,n-1}^{(n-2)} & a_{n-1,n}^{(n-2)} \\ & & & & a_{nn}^{(n-1)} \end{bmatrix}$$

这种对 $\boldsymbol{A}$ 的元素进行的消元过程叫做 Gauss 消元过程. Gauss 消元过程能够进行到底的条件是当且仅当

$$\Delta_r \neq 0\ (r = 1, 2, \cdots, n-1)$$

当上述条件满足时, 有

$$\boldsymbol{A} = \boldsymbol{A}^{(0)} = \boldsymbol{L}_1\boldsymbol{A}^{(1)} = \boldsymbol{L}_1\boldsymbol{L}_2\boldsymbol{A}^{(2)} = \cdots = \boldsymbol{L}_1\boldsymbol{L}_2\cdots\boldsymbol{L}_{n-1}\boldsymbol{A}^{(n-1)}$$

容易求出

$$\boldsymbol{L} = \boldsymbol{L}_1\boldsymbol{L}_2\cdots\boldsymbol{L}_{n-1} = \begin{bmatrix} 1 & & & & \\ c_{21} & 1 & & & \\ \vdots & \vdots & \ddots & & \\ c_{n-1,1} & c_{n-1,2} & \cdots & 1 & \\ c_{n,1} & c_{n,2} & \cdots & c_{n,n-1} & 1 \end{bmatrix}$$

这是一个对角元素都是 1 的下三角矩阵, 称为单位下三角矩阵. 若令 $\boldsymbol{A}^{(n-1)} = \boldsymbol{U}(\boldsymbol{R})$, 则得

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$$

需要指出, 一个方阵得 LU 分解并不唯一. 但有如下阐述.

设 $\boldsymbol{A} = (a_{ij})_{n\times n}$ 是 n 阶矩阵, 则当且仅当 $\boldsymbol{A}$ 的顺序主子式 $\Delta_k \neq 0 (k = 1, 2, \cdots, n-1)$ 时, $\boldsymbol{A}$ 可唯一地分解为 $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{U}$, 其中 $\boldsymbol{L}$ 是单位下三角矩阵, $\boldsymbol{U}$ 是单位上三角矩阵, 且

$$\boldsymbol{D} = \mathrm{diag}(d_1, d_2, \cdots, d_n)$$

其中 $d_k = \dfrac{\Delta_k}{\Delta_{k-1}} (k = 1, 2, \cdots, n; \Delta_0 = 1)$.

4.1.2 举例展示求法

例 4.1 求矩阵 $\boldsymbol{A}$ 的 LDU 分解, 其中

$$\boldsymbol{A}\begin{bmatrix}2 & -1 & 3\\ 1 & 2 & 1\\ 2 & 4 & 2\end{bmatrix}$$

解 因为 $\Delta_1 = 2, \Delta_2 = 5$, 所以 $\boldsymbol{A}$ 有唯一的 LDU 分解. 构造矩阵

$$\boldsymbol{L}_1 = \begin{bmatrix}1 & & \\ \frac{1}{2} & 1 & \\ 1 & 0 & 1\end{bmatrix},\ \boldsymbol{L}_1^{-1} = \begin{bmatrix}1 & & \\ -\frac{1}{2} & 1 & \\ -1 & 0 & 1\end{bmatrix}$$

计算, 得

$$\boldsymbol{L}_1^{-1}\boldsymbol{A}^{(0)} = \begin{bmatrix}2 & -1 & 3\\ 0 & \frac{5}{2} & -\frac{1}{2}\\ 0 & 5 & -1\end{bmatrix} = \boldsymbol{A}^{(1)}$$

对 $\boldsymbol{A}^{(1)}$ 构造矩阵, 有

$$\boldsymbol{L}_2 = \begin{bmatrix}1 & & \\ 0 & 1 & \\ 0 & 2 & 1\end{bmatrix},\ \boldsymbol{L}_2^{-1} = \begin{bmatrix}1 & & \\ 0 & 1 & \\ 0 & -2 & 1\end{bmatrix}$$

计算, 得

$$\boldsymbol{L}_2^{-1}\boldsymbol{A}^{(1)} = \begin{bmatrix}2 & -1 & 3\\ 0 & \frac{5}{2} & -\frac{1}{2}\\ 0 & 0 & 0\end{bmatrix} = \begin{bmatrix}2 & 0 & 0\\ 0 & \frac{5}{2} & 0\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}1 & -\frac{1}{2} & \frac{3}{2}\\ 0 & 1 & -\frac{1}{5}\\ 0 & 0 & 1\end{bmatrix}$$

求出

$$\boldsymbol{L} = \boldsymbol{L}_1\boldsymbol{L}_2 = \begin{bmatrix}1 & & \\ \frac{1}{2} & 1 & \\ 1 & 2 & 1\end{bmatrix}$$

于是得 $\boldsymbol{A}^{(0)} = \boldsymbol{A}$ 的 LDU 分解为

$$\boldsymbol{A} = \boldsymbol{L}_1\boldsymbol{L}_2\boldsymbol{A}^{(2)} = \begin{bmatrix}1 & 0 & 0\\ \frac{1}{2} & 1 & 0\\ 1 & 2 & 1\end{bmatrix}\begin{bmatrix}2 & 0 & 0\\ 0 & \frac{5}{2} & 0\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}1 & -\frac{1}{2} & \frac{3}{2}\\ 0 & 1 & -\frac{1}{5}\\ 0 & 0 & 1\end{bmatrix}$$

4.2 矩阵的 QR 分解

Givens 变换与 Householder 变换

设实数 c 与 s 满足 $c^2 + s^2 = 1$, 称

$$\boldsymbol{T}_{ij} = \begin{bmatrix} \boldsymbol{I} & & & & \\ & c & & s & \\ & & \boldsymbol{I} & & \\ & -s & & c & \\ & & & & \boldsymbol{I} \end{bmatrix} (i < j)$$

为 Givens 矩阵 (初等旋转矩阵), 也可记作 $\boldsymbol{T}_{ij} = \boldsymbol{T}_{ij}(c, s)$. 由 Givens 矩阵确定的线性变换称为 Givens 变换 (初等旋转变换).

有如下性质:

(1) Givens 矩阵是正交矩阵, 且有

$$[\boldsymbol{T}_{ij}(c, s)]^{-1} = [\boldsymbol{T}_{ij}(c, s)]^T = \boldsymbol{T}_{ij}(c, -s)$$

$$\det[\boldsymbol{T}_{ij}(c, s)] = 1$$

(2) 设 $\boldsymbol{x} = (\xi_1, \xi_2, \cdots, \xi_n)^T, \boldsymbol{y} = \boldsymbol{T}_{ij}\boldsymbol{x} = (\eta_1, \eta_2, \cdots, \eta_n)^T$, 则有

$$\begin{cases} \eta_i = c\xi_i + s\xi_j \\ \eta_j = -s\xi_j + c\xi_j \\ \eta_k = \eta_k \ (k \neq i, j) \end{cases}$$

设单位列向量 $\boldsymbol{u} \in R^n$, 称

$$\boldsymbol{H} = \boldsymbol{I} - 2\boldsymbol{u}\boldsymbol{u}^T$$

为 Householder 矩阵 (初等反射矩阵). 由 Householder 矩阵确定的线性变换称为 Householder 变换 (初等反射变换).

Householder 矩阵具有下列性质:

(1) $\boldsymbol{H}^T = \boldsymbol{H}$(对称矩阵);

(2) $\boldsymbol{H}^T\boldsymbol{H} = \boldsymbol{I}$(正交矩阵);

(3) $\boldsymbol{H}^2 = \boldsymbol{I}$(对合矩阵);

(4) $\boldsymbol{H}^{-1} = \boldsymbol{H}$(自逆矩阵);

(5) $\det \boldsymbol{H} = -1$.

4.2.1 矩阵 QR 分解的步骤推导

如果实 (复) 可逆矩阵 $\boldsymbol{A}$ 能够化成正交 (酉) 矩阵 $\boldsymbol{Q}$ 和实 (复) 可逆上三角矩阵 $\boldsymbol{R}$ 的乘积, 即

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$$

则称其为 $\boldsymbol{A}$ 的 QR 分解.

Schmidt **正交化方法**

记矩阵 $\boldsymbol{A}$ 的 n 个列向量依次为 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n$. 因为 $\boldsymbol{A}$ 可逆, 所以这 n 个列向量线性无关. 将它们按 Schmidt 正交化方法正交化之, 可得到 n 个标准正交列向量 $\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_n$.

对 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n$ 正交化, 可得

$$\begin{cases} \boldsymbol{b}_1 = & \boldsymbol{a}_1 \\ \boldsymbol{b}_2 = & \boldsymbol{a}_2 - k_{21}\boldsymbol{b}_1 \\ \cdots \\ \boldsymbol{b}_n = & \boldsymbol{a}_n - k_{n,n-1}\boldsymbol{b}_{n-1} - \cdots - k_{n1}\boldsymbol{b}_1 \end{cases}$$

其中 $k_{ij} = \dfrac{(\boldsymbol{a}_i, \boldsymbol{b}_j)}{(\boldsymbol{b}_j, \boldsymbol{b}_j)}$ $(j < i)$, 将上式改写为

$$\begin{cases} \boldsymbol{a}_1 = & \boldsymbol{b}_1 \\ \boldsymbol{a}_2 = & k_{21}\boldsymbol{b}_1 + \boldsymbol{b}_2 \\ \cdots \\ \boldsymbol{a}_n = & k_{n1}\boldsymbol{b}_1 + k_{n2}\boldsymbol{b}_2 + \cdots + k_{n,n-1}\boldsymbol{n-1} + \boldsymbol{b}_n \end{cases}$$

用矩阵形式表示为

$$(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n) = (\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_n)\boldsymbol{C}$$

其中

$$\boldsymbol{C} = \begin{bmatrix} 1 & k_{21} & \cdots & k_{n1} \\ & 1 & \cdots & k_{n2} \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}$$

再对 $\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_n$ 单位化, 可得

$$\boldsymbol{q}_i = \frac{1}{|\boldsymbol{b}_i|}\boldsymbol{b}_i \ (i = 1, 2, \cdots, n)$$

于是有

$$\begin{aligned}(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n) =& (\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_n)\boldsymbol{C} = \\ & (\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_n)\begin{bmatrix} |\boldsymbol{b}_1| & & & \\ & |\boldsymbol{b}_2| & & \\ & & \ddots & \\ & & & |\boldsymbol{b}_n| \end{bmatrix}\end{aligned}$$

令

$$\begin{cases}\boldsymbol{Q}= & (\boldsymbol{q}_1,\boldsymbol{q}_2,\cdots,\boldsymbol{q}_n)\\ \boldsymbol{R}= & \mathrm{diag}(|\boldsymbol{b}_1|,|\boldsymbol{b}_2|,\cdots,|\boldsymbol{b}_n|)\cdot\boldsymbol{C}\end{cases}$$

则 $\boldsymbol{Q}$ 是正交 (酉) 矩阵, R 是可逆上三角矩阵, 且有 $\boldsymbol{A}=\boldsymbol{QR}$.

Givens 变换方法 任何 n 阶实可逆矩阵 $\boldsymbol{A}=(a_{ij})_{n\times n}$ 可通过左连乘初等旋转矩阵化为可逆上三角矩阵.

第 1 步: 由 $\det\boldsymbol{A}\neq 0$ 知, $\boldsymbol{A}$ 的第 1 列 $\boldsymbol{b}^{(1)}=(a_{11},a_{21},\cdots,a_{n1})^T\neq\boldsymbol{0}$. 则存在有限个 Givens 矩阵的乘积, 记作 $\boldsymbol{T}_1$, 使得

$$\boldsymbol{T}_1\boldsymbol{b}^{(1)}=|\boldsymbol{b}^{(1)}|\boldsymbol{e}_1\quad(\boldsymbol{e}_1\in R^n)$$

令 $a_{11}^{(1)}=|\boldsymbol{b}^{(1)}|$, 则有

$$\boldsymbol{T}_1\boldsymbol{A}=\left[\begin{array}{c|ccc} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)}\\ \hline 0 & & & \\ \vdots & & \boldsymbol{A}^{(1)} & \\ 0 & & & \end{array}\right]$$

第 2 步: 由 $\det\boldsymbol{A}^{(1)}\neq 0$ 知, $\boldsymbol{A}^{(1)}$ 的第 1 列 $\boldsymbol{b}^{(2)}=(a_{22}^{(1)},a_{32}^{(1)},\cdots,a_{n2}^{(1)})^T\neq\boldsymbol{0}$. 则存在有限个 Givens 矩阵的乘积, 记作 $\boldsymbol{T}_2$, 使得

$$\boldsymbol{T}_2\boldsymbol{b}^{(2)}=|\boldsymbol{b}^{(2)}|\boldsymbol{e}_1\quad(\boldsymbol{e}_1\in R^{n-1})$$

令 $a_{22}^{(2)}=|\boldsymbol{b}^{(2)}|$, 则有

$$\boldsymbol{T}_2\boldsymbol{A}^{(1)}=\left[\begin{array}{c|ccc} a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2n}^{(2)}\\ \hline 0 & & & \\ \vdots & & \boldsymbol{A}^{(2)} & \\ 0 & & & \end{array}\right]$$

第 $n-1$ 步: 由 $\det\boldsymbol{A}^{(n-2)}\neq 0$ 知, $\boldsymbol{A}^{(n-2)}$ 的第 1 列 $\boldsymbol{b}^{(n-1)}=(a_{n-1,n-1}^{(n-2)},a_{n,n-1}^{(n-2)})^T\neq\boldsymbol{0}$. 则存在 Givens 矩阵 $\boldsymbol{T}_{n-1}$, 使得

$$\boldsymbol{T}_{n-1}\boldsymbol{b}^{(n-1)}=|\boldsymbol{b}^{(n-1)}|\boldsymbol{e}_1\quad(\boldsymbol{e}_1\in\boldsymbol{R}^2)$$

令 $a_{n-1,n-1}^{(n-1)}=|\boldsymbol{b}^{(n-1)}|$, 则有

$$\boldsymbol{T}_{n-1}\boldsymbol{A}^{(n-2)}=\begin{bmatrix} a_{n-1,n-1}^{(n-1)} & a_{n-1,n}^{(n-1)}\\ 0 & a_{nn}^{(n-1)}\end{bmatrix}$$

最后, 令

$$\boldsymbol{T}=\begin{bmatrix}\boldsymbol{I}_{n-2} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{T}_{n-1}\end{bmatrix}\cdots\begin{bmatrix}\boldsymbol{I}_2 & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{T}_3\end{bmatrix}\begin{bmatrix}1 & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{T}_2\end{bmatrix}\boldsymbol{T}_1$$

则 $\boldsymbol{T}$ 是有限个 Givens 矩阵的乘积, 使得

$$\boldsymbol{TA} = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1,n-1}^{(1)} & a_{1n}^{(1)} \\ & a_{22}^{(2)} & \cdots & a_{2,n-1}^{(2)} & a_{2,n}^{(2)} \\ & & \ddots & \vdots & \vdots \\ & & & a_{n-1,n-1}^{(n-1)} & a_{n-1,n}^{(n-1)} \\ & & & & a_{nn}^{(n-1)} \end{bmatrix}$$

将 $\boldsymbol{TA}$ 记作 $\boldsymbol{R}$, 那么就有 $\boldsymbol{A} = \boldsymbol{QR}$, 其中 $\boldsymbol{Q} = \boldsymbol{T}^{-1}$. 因为 $\boldsymbol{T}$ 是有限个 Givens 矩阵的乘积, 而 Givens 矩阵都是正交矩阵, 所以 $\boldsymbol{T}$ 是正交矩阵, 于是 $\boldsymbol{Q} = \boldsymbol{T}^{-1} = \boldsymbol{T}$ 也是正交矩阵.

Householder **变换方法**　任何实可逆矩阵 $\boldsymbol{A} = (a_{ij})_{n\times n}$ 可通过左连乘 Householder 矩阵化为可逆上三角矩阵.

第 1 步: 由 $\det \boldsymbol{A} \neq 0$ 知, $\boldsymbol{A}$ 的第 1 列 $\boldsymbol{b}^{(1)} = (a_{11}, a_{21}, \cdots, a_{n1})^T \neq \boldsymbol{0}$. 则存在 Householder 矩阵 $\boldsymbol{H}_1$, 使得

$$\boldsymbol{H}_1\boldsymbol{b}^{(1)} = |\boldsymbol{b}^{(1)}|\boldsymbol{e}_1 \quad (\boldsymbol{e}_1 \in R^n)$$

令 $a_{11}^{(1)} = |\boldsymbol{b}^{(1)}|$, 则有

$$\boldsymbol{H}_1\boldsymbol{A} = \left[\begin{array}{c|ccc} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ \hline 0 & & & \\ \vdots & & \boldsymbol{A}^{(1)} & \\ 0 & & & \end{array}\right]$$

第 2 步: 由 $\det \boldsymbol{A}^{(1)} \neq 0$ 知, $\boldsymbol{A}^{(1)}$ 的第 1 列 $\boldsymbol{b}^{(2)} = (a_{22}^{(1)}, a_{32}^{(1)}, \cdots, a_{n2}^{(1)})^T \neq \boldsymbol{0}$. 则存在 Householder 矩阵 $\boldsymbol{H}_2$, 使得

$$\boldsymbol{H}_2\boldsymbol{b}^{(2)} = |\boldsymbol{b}^{(2)}|\boldsymbol{e}_1 \quad (\boldsymbol{e}_1 \in R^{n-1})$$

令 $a_{22}^{(2)} = |\boldsymbol{b}^{(2)}|$, 则有

$$\boldsymbol{H}_2\boldsymbol{A}^{(1)} = \left[\begin{array}{c|ccc} a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2n}^{(2)} \\ \hline 0 & & & \\ \vdots & & \boldsymbol{A}^{(2)} & \\ 0 & & & \end{array}\right]$$

第 $n-1$ 步: 由 $\det \boldsymbol{A}^{(n-2)} \neq 0$ 知, $\boldsymbol{A}^{(n-2)}$ 的第 1 列 $\boldsymbol{b}^{(n-1)} = (a_{n-1,n-1}^{(n-2)}, a_{n,n-1}^{(n-2)})^T \neq \boldsymbol{0}$. 则存在 Householder 矩阵 $\boldsymbol{H}_{n-1}$, 使得

$$\boldsymbol{H}_{n-1}\boldsymbol{b}^{(n-1)} = |\boldsymbol{b}^{(n-1)}|\boldsymbol{e}_1 \quad (\boldsymbol{e}_1 \in \boldsymbol{R}^2)$$

令 $a_{n-1,n-1}^{(n-1)} = |\boldsymbol{b}^{(n-1)}|$, 则有

$$\boldsymbol{H}_{n-1}\boldsymbol{A}^{(n-2)} = \begin{bmatrix} a_{n-1,n-1}^{(n-1)} & a_{n-1,n}^{(n-1)} \\ 0 & a_{nn}^{(n-1)} \end{bmatrix}$$

最后, 令

$$S = \begin{bmatrix} I_{n-2} & O \\ O & H_{n-1} \end{bmatrix} \cdots \begin{bmatrix} I_2 & O \\ O & H_3 \end{bmatrix} \begin{bmatrix} 1 & O \\ O & H_2 \end{bmatrix} H_1$$

并注意到, 若 H_u 是 $n-l$ 阶 Householder 矩阵, 即

$$H_u = I_{n-l} = 2uu^T \quad (u \in R^{n-l},\ u^T u = 1)$$

令 $v = \begin{bmatrix} 0 \\ u \end{bmatrix} \in R^n$, 则 $v^T v = u^T u = 1$, 且

$$\begin{bmatrix} I_l & O \\ O & H_u \end{bmatrix} = \begin{bmatrix} I_i & O \\ O & I_{n-l} \end{bmatrix} - 2 \begin{bmatrix} O & O \\ O & uu^T \end{bmatrix} = I_n - 2 \begin{bmatrix} 0 \\ u \end{bmatrix} \begin{bmatrix} 0^T u^T \end{bmatrix} = I_n - 2vv^T$$

是 n 阶 Householder 矩阵. 因此, S 是有限个 Householder 矩阵的乘积, 且使得

$$SA = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1,n-1}^{(1)} & a_{1n}^{(1)} \\ & a_{22}^{(2)} & \cdots & a_{2,n-1}^{(2)} & a_{2,n}^{(2)} \\ & & \ddots & \vdots & \vdots \\ & & & a_{n-1,n-1}^{(n-1)} & a_{n-1,n}^{(n-1)} \\ & & & & a_{nn}^{(n-1)} \end{bmatrix}$$

将 SA 记作 R, 那么就有 $A = QR$, 其中 $Q = S^{-1}$. 因为 S 是有限个 Householder 矩阵的乘积, 而 Householder 矩阵都是正交矩阵, 所以 S 是正交矩阵, 于是 $Q = S^{-1} = S^T$ 也是正交矩阵.

4.2.2 举例展示求法

例 4.2 用 Schmidt 正交化方法求矩阵 A 的 QR 分解, 其中

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

解 令 $a_1 = (1,2,1)^T, a_2 = (2,1,2)^T, a_3 = (2,2,1)^T$, 正交可得

$$\begin{aligned} b_1 &= a_1 = (1,2,1)^T \\ b_2 &= a_2 - b_1 = (1,-1,1)^T \\ b_3 &= a_3 - \frac{1}{3}b_2 - \frac{7}{6}b_3 = (\frac{1}{2}, 0, -\frac{1}{2})^T \end{aligned}$$

构造矩阵

$$Q = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

$$R = \begin{bmatrix} \sqrt{6} & & \\ & \sqrt{3} & \\ & & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 1 & \frac{7}{6} \\ & 1 & \frac{1}{3} \\ & & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{6} & \sqrt{6} & \frac{7}{\sqrt{6}} \\ & \sqrt{3} & \frac{1}{\sqrt{3}} \\ & & \frac{1}{\sqrt{2}} \end{bmatrix}$$

则有 $\boldsymbol{A} = \boldsymbol{QR}$.

例 4.3 用初等旋转变换求矩阵 $\boldsymbol{A}$ 的 QR 分解, 其中

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

解 第 1 步: 对 $\boldsymbol{A}$ 的第 1 列 $\boldsymbol{b}^{(1)} = (0,1,1)^T$ 构造 $\boldsymbol{T}_1$, 使 $\boldsymbol{T}_1\boldsymbol{b}^{(1)} = |\boldsymbol{b}^{(1)}|\boldsymbol{e}_1$.

$$\boldsymbol{T}_{12} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ \boldsymbol{T}_{12}\boldsymbol{b}^{(1)} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$$\boldsymbol{T}_{13} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}, \ \boldsymbol{T}_{13}(\boldsymbol{T}_{12}\boldsymbol{b}^{(1)}) = \begin{bmatrix} \sqrt{2} \\ 0 \\ 0 \end{bmatrix}$$

$$\boldsymbol{T}_1 = \boldsymbol{T}_{13}\boldsymbol{T}_{12} = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -1 & 0 & 0 \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}, \ \boldsymbol{T}_1\boldsymbol{A} = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & -1 & -1 \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

第 2 步: 对 $\boldsymbol{A}^{(1)} = \begin{bmatrix} -1 & -1 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$ 的第 1 列 $\boldsymbol{b}^{(2)} = \begin{bmatrix} -1 \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$ 构造 $\boldsymbol{T}_2$, 使 $\boldsymbol{T}_2\boldsymbol{b}^{(2)} = |\boldsymbol{b}^{(2)}|\boldsymbol{e}_1$.

$$\boldsymbol{T}_{12} = \begin{bmatrix} -\sqrt{\frac{2}{3}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & -\sqrt{\frac{2}{3}} \end{bmatrix}, \ \boldsymbol{T}_{12}\boldsymbol{b}^{(2)} = \begin{bmatrix} \sqrt{\frac{3}{2}} \\ 0 \end{bmatrix}$$

$$\boldsymbol{T}_2 = \boldsymbol{T}_{12}, \ \boldsymbol{T}_2\boldsymbol{A}^{(1)} = \begin{bmatrix} \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{2}{\sqrt{3}} \end{bmatrix}$$

最后, 令

$$\boldsymbol{T}=\begin{bmatrix}1 & \\ & \boldsymbol{T}_2\end{bmatrix}\boldsymbol{T}_1=\begin{bmatrix}0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}}\\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}}\end{bmatrix}$$

则有 $\boldsymbol{A}=\boldsymbol{QR}$, 其中

$$\boldsymbol{Q}=\boldsymbol{T}^T=\begin{bmatrix}0 & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{3}}\end{bmatrix},\ \boldsymbol{R}=\begin{bmatrix}\sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ & \frac{3}{\sqrt{6}} & \frac{1}{\sqrt{6}}\\ & & -\frac{2}{\sqrt{3}}\end{bmatrix}$$

例 4.4　用 Householder 变换求矩阵 $\boldsymbol{A}$ 的 QR 分解, 其中

$$\boldsymbol{A}=\begin{bmatrix}3 & 14 & 9\\ 6 & 43 & 3\\ 6 & 22 & 15\end{bmatrix}$$

解　对 $\boldsymbol{A}$ 的第 1 列, 构造 Householder 矩阵, 有

$$\boldsymbol{b}^{(1)}=\begin{bmatrix}3\\6\\6\end{bmatrix},\ \boldsymbol{b}^{1}-|\boldsymbol{b}^{(1)}|\boldsymbol{e}_1=6\begin{bmatrix}-1\\1\\1\end{bmatrix},\ \boldsymbol{u}=\frac{1}{\sqrt{3}}\begin{bmatrix}-1\\1\\1\end{bmatrix}$$

$$\boldsymbol{H}_1=\boldsymbol{I}-2\boldsymbol{u}\boldsymbol{u}^T=\frac{1}{3}\begin{bmatrix}1 & 2 & 2\\ 2 & 1 & -2\\ 2 & -2 & 1\end{bmatrix},\ \boldsymbol{H}_1\boldsymbol{A}=\begin{bmatrix}9 & 48 & 15\\ 0 & 9 & -3\\ 0 & -12 & 9\end{bmatrix}$$

对 $\boldsymbol{A}^{(1)}=\begin{bmatrix}9 & -3\\ -12 & 9\end{bmatrix}$ 的第 1 列, 构造 Householder 矩阵, 有

$$\boldsymbol{b}^{(2)}=\begin{bmatrix}9\\-12\end{bmatrix},\ \boldsymbol{b}^{(2)}-|\boldsymbol{b}^{(2)}|\boldsymbol{e}_1=6\begin{bmatrix}-1\\-2\end{bmatrix},\ \boldsymbol{u}=\frac{1}{\sqrt{5}}\begin{bmatrix}-1\\-2\end{bmatrix}$$

$$\boldsymbol{H}_2=\boldsymbol{I}-2\boldsymbol{u}\boldsymbol{u}^T=\frac{1}{5}\begin{bmatrix}3 & -4\\ -4 & -3\end{bmatrix},\ \boldsymbol{H}_2\boldsymbol{A}^{(1)}=\begin{bmatrix}15 & -9\\ 0 & -3\end{bmatrix}$$

最后, 令

$$\boldsymbol{S}=\begin{bmatrix}1 & \\ & \boldsymbol{H}_2\end{bmatrix}\boldsymbol{H}_1=\frac{1}{15}\begin{bmatrix}5 & 10 & 10\\ -2 & 11 & -10\\ -14 & 2 & 5\end{bmatrix}$$

则有

$$\boldsymbol{Q}=\boldsymbol{S}^T=\frac{1}{15}\begin{bmatrix}5 & -2 & -14\\ 10 & 11 & 2\\ 10 & -10 & 5\end{bmatrix},\ \boldsymbol{R}=\begin{bmatrix}9 & 48 & 15\\ & 15 & -9\\ & & -3\end{bmatrix},\ \boldsymbol{A}=\boldsymbol{QR}$$

4.3 矩阵的满秩分解

本节介绍将非零矩阵分解为列满秩矩阵与行满秩矩阵的乘积问题. 这种分解理论在广义逆矩阵的研究中有重要的作用.

4.3.1 矩阵满秩分解的步骤推导

基本原理

设 $\boldsymbol{A} \in C_r^{m\times n}(r>0)$, 如果存在矩阵 $\boldsymbol{F} \in C_r^{m\times r}$ 和 $\boldsymbol{G} \in C_r^{r\times n}$, 使得

$$\boldsymbol{A} = \boldsymbol{F}\boldsymbol{G}$$

则称上式为矩阵 $\boldsymbol{A}$ 的满秩分解.

$\mathrm{rank}\boldsymbol{A} = r$ 时, 根据矩阵的初等变换理论, 对 $\boldsymbol{A}$ 进行初等变换, 可将 $\boldsymbol{A}$ 化为阶梯形矩阵 $\boldsymbol{B}$, 即

$$\boldsymbol{A} \to \boldsymbol{B} = \begin{bmatrix} \boldsymbol{G} \\ \boldsymbol{O} \end{bmatrix}, \ \boldsymbol{G} \in C_r^{r\times n}$$

于是存在有限个 m 阶初等矩阵的乘积, 记作 P, 使得 $\boldsymbol{PA} = \boldsymbol{B}$, 或者 $\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{B}$. 将 $\boldsymbol{P}^{-1}$ 分块为

$$\boldsymbol{P}^{-1} = \begin{bmatrix} \boldsymbol{F} \vdots \boldsymbol{S} \end{bmatrix} \ (\boldsymbol{F} \in C_r^{m\times r}, \ \boldsymbol{S} \in C_{m-r}^{m\times(m-r)})$$

则有

$$\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{B} = \begin{bmatrix} \boldsymbol{F} \vdots \boldsymbol{S} \end{bmatrix} \begin{bmatrix} \boldsymbol{G} \\ \boldsymbol{O} \end{bmatrix} = \boldsymbol{F}\boldsymbol{G}$$

其中 $\boldsymbol{F}$ 是列满秩矩阵, $\boldsymbol{G}$ 是行满秩矩阵.

Hermite 标准形方法

设 $\boldsymbol{B} \in C_r^{m\times n}$, 且满足

(1) $\boldsymbol{B}$ 的前 r 行中每行至少含有一个非零元素, 且第一个非零元素是 1, 而后 $m-r$ 行元素均为零;

(2) 若 $\boldsymbol{B}$ 中第 i 行的第一个非零元素 1 所在第 j_i 列 $(i=1,2,\cdots,r)$, 则 $j_1 < j_2 < \cdots < j_r$;

(3) $\boldsymbol{B}$ 中的 $j_1, j_2, \cdots, j_r$ 列为单位矩阵 $\boldsymbol{I}_m$ 的前 r 列.

那么就称 $\boldsymbol{B}$ 为 Hermite 标准形.

设 $\boldsymbol{B} \in C_r^{m\times n}$, 且满足

(1) $\boldsymbol{B}$ 的后 $m-r$ 行元素均为零;

(2) $\boldsymbol{B}$ 中的 $j_1, j_2, \cdots, j_r$ 列为单位矩阵 $\boldsymbol{I}_m$ 的前 r 列.

那么就称 $\boldsymbol{B}$ 为拟 Hermite 标准形.

设 $\boldsymbol{A} \in C_r^{m\times n}(r>0)$ 的 (拟) Hermite 标准形为 $\boldsymbol{B}$, 那么, 在 $\boldsymbol{A}$ 的满秩分解式中, 可取 $\boldsymbol{F}$ 为 $\boldsymbol{A}$ 的 $j_1, j_2, \cdots, j_r$ 列构成的 $m\times r$ 矩阵, $\boldsymbol{G}$ 为 $\boldsymbol{B}$ 的前 r 行构成的 $r\times n$ 矩阵.

由 $\boldsymbol{A} \to \boldsymbol{B}$ 知, 存在 m 阶可逆矩阵 $\boldsymbol{P}$, 使得 $\boldsymbol{PA} = \boldsymbol{B}$, 或者 $\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{B}$. 则将 $\boldsymbol{P}$ 分块为

$$\boldsymbol{P}^{-1} = \left[\boldsymbol{F} \vdots \boldsymbol{S}\right] \ (\boldsymbol{F} \in C_r^{m\times r},\ \boldsymbol{S} \in C_{m-r}^{m\times(m-r)})$$

可得满秩分解 $\boldsymbol{A} = \boldsymbol{FG}$, 其中 $\boldsymbol{G}$ 为 $\boldsymbol{B}$ 的前 r 行构成的 $r\times n$ 矩阵.

下面确定列满秩矩阵 $\boldsymbol{F}$. 参照 $\boldsymbol{A}$ 的 (拟) Hermite 标准形 $\boldsymbol{B}$, 构造 $n\times r$ 矩阵, 有

$$\boldsymbol{P} = (\boldsymbol{e}_{j_1}, \boldsymbol{e}_{j_2}, \cdots, \boldsymbol{e}_{j_r})$$

其中 $\boldsymbol{e}_j$ 表示单位矩阵 $\boldsymbol{I}_n$ 的第 j 个列向量, 则有

$$\boldsymbol{GP}_1 = \boldsymbol{I}_r,\ \boldsymbol{AP}_1 = (\boldsymbol{FG})\boldsymbol{P}_1 = \boldsymbol{F}(\boldsymbol{GP}_1) = \boldsymbol{F}$$

即 $\boldsymbol{F}$ 为 $\boldsymbol{A}$ 的 $j_1, j_2, \cdots, j_r$ 列构成的矩阵.

4.3.2 举例展示求法

例 4.5　求矩阵 $\boldsymbol{A}$ 的满秩分解, 其中

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

解

$$[\boldsymbol{A} \vdots \boldsymbol{I}] = \left[\begin{array}{cccc:ccc} -1 & 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 2 & -1 & 1 & 0 & 1 & 0 \\ 2 & 2 & -2 & -1 & 0 & 0 & 1 \end{array}\right]$$

$$\to \left[\begin{array}{cccc:ccc} -1 & 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 3 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

则有

$$\boldsymbol{B} = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{P} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix}$$

可求得

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

于是有

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

例 4.6　求矩阵 $A = \begin{bmatrix} 0 & 0 & 1 \\ 2 & 1 & 1 \\ 2j & j & 0 \end{bmatrix} (j = \sqrt{-1})$ 的满秩分解.

解　对 A 进行初等行变换, 可得

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

其中 B 是 Hermite 标准形. 因为 B 的第 1 列和第 3 列构成 I_3 的前两列, 所以 F 为 A 的第 1 列和第 3 列构成的 3×2 矩阵, 从而有

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

4.4　矩阵的奇异值分解

4.4.1　矩阵奇异值分解的步骤推导

首先, 引入如下结论.

(1) 设 $A \in C_r^{m\times n}(r>0)$, 则 A^HA 是 Hermite 矩阵, 且其特征值均为非负实数;

(2) $\mathrm{rank}(A^HA) = \mathrm{rank}A$;

(3) 设 $A \in C_r^{m\times n}$, 则 $A = O$ 的充要条件是 $A^HA = O$.

设 $A \in C_r^{m\times n}(r>0)$, A^HA 的特征值为

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0$$

则称 $\sigma_i = \sqrt{\lambda_i}(i = 1,2,\cdots,n)$ 为 A 的奇异值; 当 A 为零矩阵时, 它的奇异值都是 0.

设 $A \in C_r^{m\times n}(r>0)$, 则存在 m 阶酉矩阵 U 和 n 阶酉矩阵 V, 使得

$$U^HAV = \begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}$$

其中 $\Sigma = \mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)$, 而 $\sigma_i(i = 1,2,\cdots,r)$ 为矩阵 A 的全部非零奇异值.

记 Hermite 矩阵 $\boldsymbol{A}^H\boldsymbol{A}$ 的特征值为

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0$$

则存在 n 阶酉矩阵 $\boldsymbol{V}$, 使得

$$\boldsymbol{V}^H(\boldsymbol{A}^H\boldsymbol{A})\boldsymbol{V} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}^2 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}$$

将 $\boldsymbol{V}$ 分块为

$$\boldsymbol{V} = \left[\boldsymbol{V}_1 \vdots \boldsymbol{V}_2\right], \ \boldsymbol{V}_1 \in C_r^{m\times r}, \ \boldsymbol{V}_2 \in C_{n-r}^{n\times(n-r)}$$

并得到

$$\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{V} = \boldsymbol{V}\begin{bmatrix} \boldsymbol{\Sigma}^2 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}$$

则有

$$\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{V}_1 = \boldsymbol{V}_1\boldsymbol{\Sigma}^2, \ \boldsymbol{A}^H\boldsymbol{A}\boldsymbol{V}_2 = \boldsymbol{O}$$

根据第一式可得 $\boldsymbol{V}_1^H\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{V}_1 = \boldsymbol{\Sigma}^2$, 或者

$$(\boldsymbol{A}\boldsymbol{V}_1\boldsymbol{\Sigma}^{-1})^H(\boldsymbol{A}\boldsymbol{V}_1\boldsymbol{\Sigma}^{-1}) = \boldsymbol{I}$$

根据第二式可得 $(\boldsymbol{A}\boldsymbol{V}_2)^H(\boldsymbol{A}\boldsymbol{V}_2) = \boldsymbol{O}$, 或者 $\boldsymbol{A}\boldsymbol{V}_2 = \boldsymbol{O}$.

令 $\boldsymbol{U}_1 = \boldsymbol{A}\boldsymbol{V}_1\boldsymbol{\Sigma}^{-1}$, 则 $\boldsymbol{U}_1^H\boldsymbol{U}_1 = \boldsymbol{I}_r$, 即 $\boldsymbol{U}_1$ 的 r 个列是两两正交的单位向量, 记作 $\boldsymbol{U}_1 = (\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_r)$. 并且可将其扩充为 C^m 的标准正交基, 记新添的向量为 $\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_m$, 并构造矩阵 $\boldsymbol{U}_2 = (\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_m)$, 则

$$\boldsymbol{U} = \left[\boldsymbol{U}_1 \vdots \boldsymbol{U}_2\right]$$

是 m 阶酉矩阵, 且有

$$\boldsymbol{U}_1^H\boldsymbol{U}_1 = \boldsymbol{I}_r, \ \boldsymbol{U}_2^H\boldsymbol{U}_1 = \boldsymbol{O}$$

于是可得

$$\boldsymbol{U}^H\boldsymbol{A}\boldsymbol{V} = \boldsymbol{U}^H\left[\boldsymbol{A}\boldsymbol{V}_1 \vdots \boldsymbol{A}\boldsymbol{V}_2\right] = \begin{bmatrix} \boldsymbol{U}_1^H \\ \boldsymbol{U}_2^H \end{bmatrix}\left[\boldsymbol{U}_1\boldsymbol{\Sigma} \vdots \boldsymbol{O}\right] = \begin{bmatrix} \boldsymbol{Sigma} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}$$

即得到矩阵 $\boldsymbol{A}$ 的奇异值分解

$$\boldsymbol{A} = \boldsymbol{U}\begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}\boldsymbol{V}^H$$

4.4.2 举例展示求法

例 4.7 求矩阵 $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ 的奇异值分解.

解 计算

$$\boldsymbol{B} = \boldsymbol{A}^T\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

求得 $\boldsymbol{B}$ 的特征值为 $\lambda_1 = 3, \lambda_2 = 1, \lambda_3 = 0$, 对应的特征向量依次是

$$\boldsymbol{\xi}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \quad \boldsymbol{\xi}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad \boldsymbol{\xi}_3 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

可得

$$\text{rank}\boldsymbol{A} = 2, \boldsymbol{\Sigma} = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix}$$

且正交矩阵为

$$\boldsymbol{V} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \end{bmatrix}$$

计算

$$\boldsymbol{U}_1 = \boldsymbol{A}\boldsymbol{V}_1\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & 0 \end{bmatrix}$$

构造

$$\boldsymbol{U}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{U} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

则 $\boldsymbol{A}$ 的奇异值分解为

$$\boldsymbol{A} = \boldsymbol{U} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{V}^T$$

4.4.3 利用奇异值分解求矩阵广义逆

定义 4.1 设矩阵 $\boldsymbol{A} \in \boldsymbol{C}^{m\times n}$, 若矩阵 $\boldsymbol{X} \in \boldsymbol{C}^{n\times m}$ 满足以下 4 个 Penrose 方程

$$(1)\boldsymbol{AXA} = \boldsymbol{A} \qquad (2)\boldsymbol{XAX} = \boldsymbol{X}$$
$$(3)(\boldsymbol{AX})^H = \boldsymbol{AX} \quad (4)(\boldsymbol{XA})^H = \boldsymbol{XA}$$

则称 $\boldsymbol{X}$ 为 $\boldsymbol{A}$ 的 Moore-Penrose **逆**, 记为 $\boldsymbol{A}^+$.

定理 4.1　对任意 $\boldsymbol{A} \in C^{m\times n}$, $\boldsymbol{A}^+$ 存在并且唯一.

定义 4.2　设矩阵 $\boldsymbol{A} \in C^{m\times n}$, 矩阵 $\boldsymbol{X} \in C^{n\times m}$.

(1) 若 $\boldsymbol{X}$ 满足 Penrose 方程中的第 (i) 个方程, 则称 $\boldsymbol{X}$ 为 $\boldsymbol{A}$ 的 $\{i\}$ -逆, 记作 $\boldsymbol{A}^{(i)}$, 全体 $\{i\}$ -逆的集合记作 $\boldsymbol{A}\{i\}$. 这种广义逆矩阵共有四类;

(2) 若 $\boldsymbol{X}$ 满足 Penrose 方程中的第 $(i),(j)$ 个方程 $(i \neq j)$, 则称 $\boldsymbol{X}$ 为 $\boldsymbol{A}$ 的 $\{i,j\}$ -逆, 记作 $\boldsymbol{A}^{(i,j)}$, 全体 $\{i,j\}$ -逆的集合记作 $\boldsymbol{A}\{i,j\}$. 这种广义逆矩阵共有 6 类;;

(3) 若 $\boldsymbol{X}$ 满足 Penrose 方程中的第 $(i),(j),(k)$ 个方程 (i,j,k 互异), 则称 $\boldsymbol{X}$ 为 $\boldsymbol{A}$ 的 $\{i,j,k\}$ -逆, 记作 $\boldsymbol{A}^{(i,j,k)}$, 全体 $\{i,j,k\}$ -逆的集合记作 $\boldsymbol{A}\{i,j,k\}$. 这种广义逆矩阵共有 4 类;

(4) 若 $\boldsymbol{X}$ 满足 Penrose 方程 (1) ~ (4), 则称 $\boldsymbol{X}$ 为 $\boldsymbol{A}$ 的 Moore-Penrose 逆 $\boldsymbol{A}^+$, 这种广义逆矩阵是唯一的.

设 $\boldsymbol{A} \in C_r^{m\times n}$, 其奇异值分解 $\boldsymbol{A} = \boldsymbol{UDV}^H$, 有 $\boldsymbol{A} = \boldsymbol{U}_r\Delta\boldsymbol{V}_r^H, \Delta = \mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)$, 则

$$\boldsymbol{A}^+ = \boldsymbol{V}_r\Delta^{-1}\boldsymbol{U}_r^h$$

例 4.8　设 $\boldsymbol{A} = \begin{bmatrix} -1 & 0 & 1 \\ 2 & 0 & -2 \end{bmatrix}$, 求 $\boldsymbol{A}^+$.

解　先求 $\boldsymbol{A}$ 的奇异值分解.

$$\boldsymbol{AA}^H = \begin{bmatrix} 2 & -4 \\ -4 & 8 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt5} & \frac{2}{\sqrt5} \\ -\frac{2}{\sqrt5} & \frac{1}{\sqrt5} \end{bmatrix}\begin{bmatrix} 10 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt5} & -\frac{2}{\sqrt5} \\ \frac{2}{\sqrt5} & \frac{1}{\sqrt5} \end{bmatrix}$$

令 $\boldsymbol{V}_1 = \frac{1}{\sqrt{10}}\begin{bmatrix} \frac{1}{\sqrt5} & -\frac{2}{\sqrt5} \end{bmatrix}\begin{bmatrix} -1 & 0 & 1 \\ 2 & 0 & -2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt2} & 0 & \frac{1}{\sqrt2} \end{bmatrix}$

$$\boldsymbol{A} = \begin{bmatrix} \frac{1}{\sqrt5} \\ -\frac{2}{\sqrt5} \end{bmatrix}\begin{bmatrix} \sqrt{10} \end{bmatrix}\begin{bmatrix} -\frac{1}{\sqrt2} & 0 & \frac{1}{\sqrt2} \end{bmatrix}$$

则

$$\boldsymbol{A}^+ = \begin{bmatrix} -\frac{1}{\sqrt2} \\ 0 \\ \frac{1}{\sqrt2} \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{10}} \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt5} & -\frac{2}{\sqrt5} \end{bmatrix} = \frac{1}{10}\begin{bmatrix} -1 & 2 \\ 0 & 0 \\ 1 & -2 \end{bmatrix}$$

第 5 章　总结

本论文是对本学期《矩阵理论与方法》课程的报告.

这门课是一门很有用的而学科, 它发展完善、理论严谨、方法独特, 对培养我的逻辑能力、推理能力具有重要的作用, 同时又能广泛应用于各种领域. 矩阵论是对线性代数的延申, 很有必要深入研究. 用矩阵的理论与方法来处理现代工程技术中的各种问题已经越来越普遍. 在工程技术中引进矩阵理论不仅使理论的表达极为简捷, 而且对理论的实质刻画也更为深刻, 这一点是不容置疑的. 计算机和计算方法的普及发展, 不仅为矩阵理论的应用开辟了广阔的前景, 也使工程技术的研究发生新的变化, 开拓了崭新的研究途径.

透过对本课程的学习, 我了解了线性空间、线性变换、范数理论、矩阵分析、矩阵分解、广义逆矩阵等. 通过本次课题的研究, 我对这学期所学知识重新回顾, 构建起了知识框架, 更加深入了解了矩阵函数的求解和矩阵的分解方法, 也算是给这门课程画上一个圆满的句号.

相信通过这门课学习到的理论, 一定能在未来对我起到极大的帮助!

参考文献

[1] 张凯院, 徐仲等. 矩阵论 [J]. 西北工业大学出版社, 2017.

[2] 王辉. 矩阵理论以及应用 [D]. [S.l.] : 淮北师范大学, 2011.

致谢

论文即将完成之际, 我借此机会向我的老师、同学们致以深切的感谢!

感谢李辰昊老师的教学与制定的论文题目, 让我能够深入了解矩阵及矩阵分析, 对其原理有了较多的认知, 在此向他表示最中心的感谢!

然后感谢我的同学们, 在我需要帮助的时候给了我极大的照顾. 在此, 我一并向他们表示感谢!

最后, 感谢我的家人们和学校, 给我提供衣食住行的条件!