

北京邮电大学

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 _____

学 院 _____

专 业 _____

班 级 _____

学 号 _____

指导教师 _____

2023 年 12 月

矩阵函数的求法和矩阵分解方法研究

摘 要

本论文研究了矩阵函数的求法和矩阵分解的方法。本论文以矩阵理论的基础知识为铺垫，对于矩阵函数的求法，详细探讨了待定系数法、数项级数求和法、对角型法和 Jordan 标准型法这四种矩阵函数的求解方法；对于矩阵分解方法，详细探讨了矩阵的 LU 分解、QR 分解、满秩分解、奇异值分解四种分解方法，并给出了其实际运用，即矩阵分解求矩阵广义逆。本文展示了以上方法的步骤并以例题演示，总结了《矩阵理论与方法》学期课程内容。

关键词 矩阵理论 矩阵函数 矩阵分解

RESEARCH ON METHODS FOR MATRIX FUNCTION COMPUTATION AND MATRIX DECOMPOSITION

ABSTRACT

This paper explores methods for matrix function computation and matrix decomposition. Building upon foundational knowledge in matrix theory, the paper delves into four approaches for matrix function computation: the method of undetermined coefficients, series summation method, diagonalization method, and Jordan canonical form method. Additionally, it thoroughly investigates four matrix decomposition techniques: LU decomposition, QR decomposition, full rank decomposition, and singular value decomposition, illustrating their practical application in finding the generalized inverse of matrices through decomposition. The paper presents the step-by-step procedures for these methods and provides illustrative examples, summarizing the content covered in the course "Matrix Theory and Methods."

KEY WORDS Matrix Function Matrix Theory Matrix Decomposition

目 录

第一章 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	1
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	3
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	4
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
第二章 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 $Jordan$ 标准型的求解	12
2.1.3 欧式空间中线性变换的求法	16
2.2 向量范数与矩阵范数	22
2.2.1 向量范数介绍	22
2.2.2 矩阵范数介绍	23
2.2.3 矩阵可逆性条件、条件数和谱半径介绍	24
2.3 矩阵函数介绍	25
2.3.1 矩阵序列介绍	25
2.3.2 矩阵级数介绍	26
2.3.3 矩阵函数介绍	29
2.3.4 函数矩阵对矩阵的导数	30
第三章 矩阵函数的求法研究	36
3.1 待定系数法	36
3.1.1 待定系数法求矩阵函数的步骤推导	36

3.1.2 举例展示求法	36
3.2 数项级数求和法	38
3.2.1 数项级数求和法求矩阵函数的步骤推导	38
3.2.2 举例展示求法	38
3.3 对角型法	39
3.3.1 对角型法求矩阵函数的步骤推导	39
3.3.2 举例展示求法	40
3.4 Jordan 标准型法	41
3.4.1 Jordan 标准型法求矩阵函数的步骤推导	41
3.4.2 举例展示求法	42
第四章 矩阵分解方法研究	44
4.1 矩阵的 LU 分解	44
4.1.1 矩阵 LU 分解的步骤推导	44
4.1.2 举例展示求法	45
4.2 矩阵的 QR 分解	47
4.2.1 矩阵 QR 分解的步骤推导	47
4.2.2 举例展示求法	49
4.3 矩阵的满秩分解	50
4.3.1 矩阵满秩分解的步骤推导	50
4.3.2 举例展示求法	51
4.4 矩阵的奇异值分解	52
4.4.1 矩阵奇异值分解的步骤推导	52
4.4.2 举例展示求法	53
4.5 利用矩阵分解求矩阵广义逆	55
4.5.1 矩阵广义逆介绍	55
4.5.2 利用矩阵满秩分解求矩阵广义逆	56
4.5.3 利用矩阵奇异值分解求矩阵广义逆	56
4.5.4 举例展示求法	57
第五章 总结	59
参考文献	60

第一章 引言

1.1 背景介绍

1.1.1 矩阵理论与方法介绍

矩阵理论是线性代数的一个重要分支，为解决复杂的数学和工程问题提供了有力的工具。矩阵的运算包括乘法、逆矩阵、特征值和特征向量等，它们在解决线性方程组和理解线性变换方面起着关键作用。矩阵分解方法如 LU 分解、QR 分解等也是研究热点。

在矩阵理论的应用中，我们需要进行各种计算，如特征值与特征向量的计算、矩阵函数的计算、矩阵的微分与积分等。矩阵理论在科学、工程、统计学等领域广泛应用，并且对于解决复杂的线性代数问题具有特殊优势，如 Jordan 标准型。总之，矩阵理论为我们理解和应用线性代数提供了强大的工具和方法。

1.1.2 函数矩阵和矩阵函数介绍

函数矩阵是一类特殊的矩阵，其元素是函数而非常数。与之相对地，矩阵函数则是将矩阵作为自变量的函数。本文中的函数矩阵定义为：以变量 t 的函数 $a_{ij}(t)$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$) 为元素的矩阵 $A(t)$ 。

函数矩阵的微积分运算实质上是对矩阵内元素（函数）进行的微积分运算，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，其运算结果也仍为函数矩阵。

矩阵函数的概念与通常的函数概念类似，是以 n 阶矩阵为自变量和函数值（因变量）的一种函数。矩阵函数将会在 2.3 节给出严格的定义，它是以矩阵序列与矩阵级数为基础的对一元函数概念的推广。借助于 Hamilton - Cayley 定理，可以将矩阵函数的求值问题转化为矩阵多项式的计算问题；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。本文第三章将会对这一点进行详细地叙述。

函数矩阵与矩阵函数是两个不同的概念，但在一些情形下，矩阵函数在其定义域内的值可以看做函数矩阵。例如，矩阵指数函数可以看做相应变量的函数矩阵。

1.1.3 线性代数方程组求解介绍

线性代数方程组是由若干个线性方程组成的方程组，是科学和工程中常见的数学模型，许多应用问题对这类方程组进行求解。

对于一个线性代数方程组来说，解的情况通常有唯一解、无穷多解、无解。矩阵的分解如 LU 分解、QR 分解、满秩分解和奇异值分解为高效解决线性代数方程组提供了途径。因此，深入研究矩阵的分解方法对于提高线性代数方程组的求解效率具有重要意义。这一点将在本文第四章进行详细讲解。

1.2 问题介绍

很多实际问题，如两点边值问题¹的数值求解：

$$u''(x) + p(x)u'(x) + q(x)u(x) = f(x), \quad x \in I := (-1, 1)$$

往往可以转化为求解如下线性系统：

$$\mathbf{A}\bar{\mathbf{U}} = \bar{\mathbf{b}}$$

其中矩阵 \mathbf{A} 和向量 $\bar{\mathbf{b}}$ 是由下式得到的：

$$a_{ij} = \epsilon(\mathbf{D}^2)_{ij} + p(x_i)(\mathbf{D}^1)_{ij} + q(x_i)\delta_{ij}, \quad 1 \leq i, j \leq N-1$$

$$b_i = f(x_i) - [\epsilon(\mathbf{D}^2)_{i0} + p(x_i)(\mathbf{D}^1)_{i0}]c_+ - [\epsilon(\mathbf{D}^2)_{iN} + p(x_i)(\mathbf{D}^1)_{iN}]c_-, \quad 1 \leq i \leq N-1$$

更加一般地，我们考虑可以把很多问题的数值求解，转化为求解线性系统

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

其中 $\mathbf{A} = \mathbf{A}(f_1(\mathbf{D}))$, $\mathbf{b} = \mathbf{b}(f_2(\mathbf{D}))$ 。

\mathbf{D} 是一个 N 阶方阵； $f_1(\mathbf{D})$ 和 $f_2(\mathbf{D})$ 分别是关于矩阵 \mathbf{D} 的矩阵函数，矩阵 \mathbf{A} 和向量 \mathbf{b} 分别是关于 $f_1(\mathbf{D})$ 和 $f_2(\mathbf{D})$ 的函数矩阵；线性系统 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 可以利用矩阵分解来求解。因此，本文以下讨论的矩阵函数的求解与矩阵分解方法的研究，对于实际问题的数值求解是非常有意义的。

1.2.1 矩阵函数的求法问题介绍

如上例所示，矩阵函数的求法是矩阵理论中的一个关键问题。一般的函数求解时，可以直接应用微积分规则，但这并不适用于矩阵函数。

引入两道例题（第一次课 ppt 的例 1、3），初步演示求解矩阵函数的问题：

例题 1.1 $e^A = ?$, 其中 $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

解：

$$e^x = \sum_{k=0}^{+\infty} \frac{1}{k!} x^k = 1 + x + \frac{1}{2!} x^2 + \frac{1}{3!} x^3 + \dots$$

$$e^A = \sum_{k=0}^{+\infty} \frac{1}{k!} A^k = I + A + \frac{1}{2!} A^2 + \frac{1}{3!} A^3 + \dots = eA = \begin{pmatrix} e & 0 \\ 0 & e \end{pmatrix}$$

¹J. Shen, T. Tang, Spectral and High-order Methods with Applications, Science Press, 2006.

例题 1.2 $e^A = ?$, 其中 $A = \begin{pmatrix} 4 & 6 & 0 \\ -3 & -5 & 0 \\ -3 & -6 & 1 \end{pmatrix}$

解:

$$\begin{aligned} e^A &= \sum_{k=0}^{+\infty} \frac{1}{k!} A^k = I + A + \frac{1}{2!} A^2 + \frac{1}{3!} A^3 + \dots \\ &= P^{-1} I P + P^{-1} \Lambda P + \frac{1}{2!} P^{-1} \Lambda^2 P + \frac{1}{3!} P^{-1} \Lambda^3 P + \dots \\ &= P^{-1} \begin{pmatrix} e^{-2} & & \\ & e & \\ & & e \end{pmatrix} P \end{aligned}$$

在介绍矩阵函数的求法问题之前, 本文将引入一些预备知识, 即第二章对 Jordan 标准型、向量范数、矩阵范数、矩阵序列与矩阵级数的介绍等, 为后续矩阵函数的引入奠定基础。

铺垫完预备知识后, 第三章将详细介绍矩阵函数的求解方法, 包括待定系数法、数项级数求和法、对角型法和 Jordan 标准型法四种方法。

1.2.2 矩阵分解的方法问题介绍

同样如上例所示, 当求解一个线性系统时, 矩阵分解在其中起到了重要作用。矩阵分解指的是将一个矩阵拆分成由若干个基本矩阵相乘的形式, 比如三角矩阵、对称矩阵等, 有助于减少矩阵计算时的计算量。

第二章将介绍一些预备知识, 例如欧式空间中线性变换的求法, 矩阵可逆性、条件数与谱半径等。它们作为矩阵分解中常用的性质和指标, 与 LU 分解、QR 分解等方法直接相关。

铺垫完预备知识后, 第四章将详细介绍矩阵分解的方法, 包括 LU 分解、QR 分解、满秩分解和奇异值分解四种方法。

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

在《矩阵论》一书中介绍了求矩阵函数值的四种常见方法: 待定系数法、数项技术求和法、对角型法和 Jordan 标准型法。这些方法各有优劣, 取决于问题的特性和计算的复杂度。第三章将一一介绍这些方法的原理与推导过程, 并通过相关例子进行介绍。

1.3.2 矩阵分解方法研究现状

在《矩阵论》一书中介绍了四种矩阵分解的方法，分别是矩阵的 LU 分解、矩阵的 QR 分解、矩阵的满秩分解和矩阵的奇异值分解。每种分解方法都有其独特的适用场景和特点，如 LU 分解适用于求解线性方程组，而奇异值分解则常用于降维和数据压缩。第四章将一一介绍这些方法的原理与推导过程，并通过相关例子进行介绍。

1.4 本论文工作简述

1.4.1 本论文对上述问题研究简述

本论文深入研究了矩阵函数的求解与矩阵分解的方法。在引言中，我们概述了论文结构和研究问题；预备知识部分涵盖了欧式空间、向量与矩阵范数、矩阵函数的基础知识；第三章详细阐述了矩阵函数求解的四种方法；第四章详细阐述了矩阵分解的四大方法。

1.4.2 本论文创新点或特点简述

本论文的主要参考资料是《矩阵论》（张凯院、徐仲，西北工业大学出版社 2017 版）。在撰写过程中，还参考了《矩阵理论与方法》课程学习过程中的 ppt 与作业作为补充。

对于矩阵函数的求法和矩阵分解方法，本文将一一介绍这些方法的原理、步骤与推导过程，并通过相关例题进行演示，让概念易于理解。

1.4.3 本论文撰写结构简述

本文主要论述了矩阵函数的求法和矩阵分解的方法。

第一章的引言部分，主要介绍了本篇论文的基本结构和研究问题；

第二章的预备知识部分，总结了求解矩阵函数与研究矩阵分解所需的基础知识，内容包括欧式空间与线性变换、向量范数与矩阵范数、矩阵函数的介绍三个部分；

第三章归纳了矩阵函数的四种求解方法，包括待定系数法、数项级数求和法、对角型法和 Jordan 标准型法；

第四章归纳了矩阵分解问题的四种方法，分别是矩阵的 LU 分解、矩阵的 QR 分解、矩阵的满秩分解和矩阵的奇异值分解。

第二章 预备知识

2.1 欧式空间与线性变换

2.1.1 欧式空间与线性变换介绍

欧式空间是一种特殊的线性空间。它在线性空间的基础上引入了内积结构，使得空间中的向量具有明确的几何意义。相比之下，线性空间则更为抽象，侧重于向量之间的线性运算的代数性质。

先引入定义，介绍线性空间，欧式空间的严格定义将在 2.1.3 节中引入：

定义 2.1 设 V 是一个非空集合，它的元素用 $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ 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 $\boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z}) = (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z}$;

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

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

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 $\boldsymbol{x} + \boldsymbol{y} = \mathbf{0}$ ，则称 \boldsymbol{y} 为 \boldsymbol{x} 的负元素，记为 $-\boldsymbol{x}$ ，于是有 $\boldsymbol{x} + (-\boldsymbol{x}) = \mathbf{0}$ 。

(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：

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

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

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

4) $1\boldsymbol{x} = \boldsymbol{x}$ 。

则称 V 为数域 K 上的线性空间或向量空间。

特殊地，当 K 为实数域 \mathbf{R} 时，称 V 为实线性空间；当 K 为复数域 \mathbf{C} 时，称 V 为复线性空间。

线性代数中的线性相关，线性无关，基等概念同样可以推广到线性空间中，其相关结论推广后仍然成立，例如：部分相关则全体相关；全体无关则部分无关；包含零元素肯定相关；相关集中至少有一元素可由其它元素线性表示等。

这些概念可以引出线性空间的基与坐标，定义如下：

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

- (1) $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_r$ 线性无关;
- (2) V 中任一向量 \boldsymbol{x} 都是 $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_r$ 的线性组合.

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

需要指出, 一个线性空间的基不是唯一的, 但是维数是唯一确定的。

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, 可以从中引出线性变换的定义:

定义 2.3 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)。

定义 2.4 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。

由此可以知道线性变换的简单性质:

$$T\mathbf{0} = T(\mathbf{0}\boldsymbol{x}) = \mathbf{0}(T\boldsymbol{x}) = \mathbf{0}$$

$$T(-\boldsymbol{x}) = T((-1)\boldsymbol{x}) = (-1)T\boldsymbol{x} = -T\boldsymbol{x}$$

这就表明, 线性变换把线性空间的零向量变为零向量; 把向量 \boldsymbol{x} 的负向量 $-\boldsymbol{x}$ 变为 \boldsymbol{x} 的象 $T\boldsymbol{x}$ 的负向量 $-T\boldsymbol{x}$ 。又线性变换把线性相关的向量组仍变为线性相关的向量组, 即当

$$k_1\boldsymbol{x}_1 + k_2\boldsymbol{x}_2 + \dots + k_s\boldsymbol{x}_s = \mathbf{0}$$

其中 $k_i (i = 1, 2, \dots, s)$ 不全为零, 则有

$$k_1(T\boldsymbol{x}_1) + k_2(T\boldsymbol{x}_2) + \dots + k_s(T\boldsymbol{x}_s) = T\mathbf{0} = \mathbf{0}$$

再结合定义 2.4, 可得

$$\boldsymbol{x} = (E_1, \dots, E_n) \begin{pmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_n \end{pmatrix}$$

$$T(\mathbf{x}) = T\left((E_1, \dots, E_n) \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}\right) = (T(E_1), \dots, T(E_n)) \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}$$

实际上，微分和积分两个运算，从变换（或算子）的角度来看都是线性变换（或线性算子）。可以通过以下两个例题（《矩阵论》例 1.11, 1.12）来说明：

在线性空间 P_n 中，求微分是其一个线性变换，这里用 $D(\cdot)$ 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 $k, l \in \mathbf{R}$ ，有：

$$\begin{aligned} D(kf(t) + lg(t)) &= (kf(t) + lg(t))' \\ &= kf'(t) + lg'(t) = k(Df(t)) + l(Dg(t)) \end{aligned}$$

图 2-1 例题 2.1

定义在闭区间 $[a, b]$ 上的所有实连续函数的集合 $C(a, b)$ 构成 \mathbf{R} 上的一个线性空间，在 $C(a, b)$ 上定义变换 J ，即

$$J(f(t)) = \int_a^t f(u) du \quad (\forall f(t) \in C(a, b))$$

则 J 是 $C(a, b)$ 的一个线性变换

图 2-2 例题 2.2

整理成文字版本：

例题 2.1 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 $k, l \in \mathbf{R}$ ，有

$$\begin{aligned} D(kf(t) + lg(t)) &= (kf(t) + lg(t))' \\ &= kf'(t) + lg'(t) = k(Df(t)) + l(Dg(t)) \end{aligned}$$

例题 2.2 定义在闭区间 $[a, b]$ 上的所有实连续函数的集合 $C(a, b)$ 构成 \mathbf{R} 上的一个线性空间。在 $C(a, b)$ 上定义变换 J ，即

$$J(f(t)) = \int_a^t f(u) du \quad (\forall f(t) \in C(a, b))$$

则 J 是 $C(a, b)$ 的一个线性变换。

事实上, 因为有

$$\begin{aligned} J(kf(t) + lg(t)) &= \int_a^t (kf(u) + lg(u)) \, du \\ &= k \int_a^t f(u) \, du + l \int_a^t g(u) \, du = k(Jf(t)) + l(Jg(t)) \end{aligned}$$

得知了这些性质之后, 接下来将讨论线性变换的运算:

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

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

线性变换 T 的负变换 $-T$ 定义为

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

$T_1 + T_2$ 和 $-T$ 都是 V 的线性变换。

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

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

kT 也是线性变换。

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

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

T_1T_2 也是 V 的线性变换。

需要强调, 线性变换的乘法不满足交换律。

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

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

则称 S 是 T 的逆变换, 记为 $S = T^{-1}$. 且有

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

T^{-1} 是线性变换。

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

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

定义 T 的零次幂为

$$T^0 = T_e$$

于是可以建立线性变换的指数法则

$$T^{m+n} = T^mT^n, \quad (T^m)^n = T^{mn}$$

其中 $m, n \in \mathbf{N}_0$ 。当 T 是可逆变换时, 定义 T 的负整数次幂为

$$T^{-n} = (T^{-1})^n \quad (n \in \mathbf{N}_0)$$

这样就把指数法则式推广到负整数次幂的情形。

设 $f(t) = a_0t^m + a_1t^{m-1} + \dots + a_{m-1}t + a_m$ 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

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

也是 V 的一个线性变换, 称其为线性变换 T 的多项式。

得知了线性变换的多项式, 接下来引入一道例题 (第三次课 ppt-1), 说明其求解方法。

例题 2.3 在矩阵空间 $\mathbf{R}^{2 \times 2}$ 中, 给定矩阵 $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $T_1(\mathbf{X}) = \mathbf{A}\mathbf{X} \ (\forall \mathbf{X} \in \mathbf{R}^{2 \times 2})$,

求 $T(\mathbf{X}) = f(T_1)(\mathbf{X})$

解: 找到 $\mathbf{R}^{2 \times 2}$ 的一组基:

$$\mathbf{X}_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \mathbf{X}_2 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \mathbf{X}_3 = \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}, \mathbf{X}_4 = \begin{pmatrix} 0 & -1 \\ 0 & 1 \end{pmatrix}$$

其中

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)\boldsymbol{\alpha}, \quad \boldsymbol{\alpha} = \frac{1}{2}(\mathbf{x}_1 + \mathbf{x}_3, \mathbf{x}_2 + \mathbf{x}_4, -\mathbf{x}_1 + \mathbf{x}_3, -\mathbf{x}_2 + \mathbf{x}_4)^T$$

因此

$$T_1(\mathbf{X}) = T_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)\boldsymbol{\alpha} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)\text{diag}(1, 1, -1, -1)\boldsymbol{\alpha}$$

$$T_1^k(\mathbf{X}) = T_1^k(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)\boldsymbol{\alpha} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)\text{diag}(1, 1, (-1)^k, (-1)^k)\boldsymbol{\alpha}$$

$$f(T_1)(\mathbf{X}) = f(T_1)(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)\boldsymbol{\alpha} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)f(\text{diag}(1, 1, -1, -1))\boldsymbol{\alpha}$$

有限维线性空间的向量可以用坐标表示，线性变换可以用矩阵表示。线性空间的任一向量都可由基向量唯一线性表示，所以只要能够确定基向量的象，就能确定线性空间的任一向量。

接下来介绍线性变换的矩阵表示。

定义 2.5 设 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ 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

$$\begin{cases} T(\mathbf{x}_1) = a_{11}\mathbf{x}_1 + a_{21}\mathbf{x}_2 + \cdots + a_{n1}\mathbf{x}_n \\ T(\mathbf{x}_2) = a_{12}\mathbf{x}_1 + a_{22}\mathbf{x}_2 + \cdots + a_{n2}\mathbf{x}_n \\ \quad \quad \quad \dots\dots\dots \\ T(\mathbf{x}_n) = a_{1n}\mathbf{x}_1 + a_{2n}\mathbf{x}_2 + \cdots + a_{nn}\mathbf{x}_n \end{cases}$$

用矩阵表示即为：

$$T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = (T\mathbf{x}_1, T\mathbf{x}_2, \dots, T\mathbf{x}_n) = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)\mathbf{A}$$

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

举例说明求解线性变换的矩阵的方法：（《矩阵论》例 1.15）

整理成文字版本：

例题 2.4 在矩阵空间 $\mathbf{R}^{2 \times 2}$ 中，给定矩阵 $\mathbf{B} = \begin{bmatrix} 0 & 1 \\ 4 & 0 \end{bmatrix}$ ，线性变换为 $T(\mathbf{X}) = \mathbf{X}\mathbf{B} (\forall \mathbf{X} \in \mathbf{R}^{2 \times 2})$ ， $\mathbf{R}^{2 \times 2}$ 的基为 $\mathbf{E}_{11}, \mathbf{E}_{12}, \mathbf{E}_{21}, \mathbf{E}_{22}$ ，求 T 在这个基下的矩阵。

$$\begin{aligned}
T(\mathbf{E}_{11}) &= \mathbf{E}_{11}\mathbf{B} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \mathbf{E}_{12} & T(\mathbf{E}_{21}) &= \mathbf{E}_{21}\mathbf{B} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{E}_{22} \\
T(\mathbf{E}_{12}) &= \mathbf{E}_{12}\mathbf{B} = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} = 4\mathbf{E}_{11} & T(\mathbf{E}_{22}) &= \mathbf{E}_{22}\mathbf{B} = \begin{bmatrix} 0 & 0 \\ 4 & 0 \end{bmatrix} = 4\mathbf{E}_{21} + \mathbf{E}_{22} \\
\Rightarrow \mathbf{A}_1 &= \begin{bmatrix} 0 & 4 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 \end{bmatrix}
\end{aligned}$$

图 2-3 例题 2.4

解：计算基像组，有

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

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

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

$$T(\mathbf{E}_{22}) = \mathbf{E}_{22}\mathbf{B} = \begin{bmatrix} 0 & 0 \\ 4 & 0 \end{bmatrix} = 0\mathbf{E}_{11} + 0\mathbf{E}_{12} + 4\mathbf{E}_{21} + 1\mathbf{E}_{22}$$

故 T 在这个基下的矩阵为

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

下面的定理介绍了 $T_1 + T_2, kT_1, T_1T_2$ 在这个基下的矩阵。

定理 2.1 设 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ 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 \mathbf{A}, \mathbf{B} 。则有下列结论：

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

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

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

$$(4) T_1^{-1}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)\mathbf{A}^{-1};$$

为了由线性变换的矩阵 \mathbf{A} 计算向量 \mathbf{x} 的象 $T\mathbf{x}$ 的坐标，还可以引入如下定理：

定理 2.2 设线性变换 T 在线性空间 V^n 的基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$ 下的矩阵是 \boldsymbol{B} , 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

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

对于一个有限维线性空间 V , 取定一组基后, V 的任一线性变换都可以用矩阵来表示。为了研究线性变换性质, 希望这个矩阵越简单越好, 如对角矩阵。

现在讨论如何选择线性空间的基, 使线性变换在该基下的矩阵形状最简单的问题。引入定义, 论述线性变换的特征值和特征向量的概念:

定义 2.6 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

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

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。

定义 2.7 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

$$\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix} \quad \text{式 (2-1)}$$

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); 而相应于方程组 $\boldsymbol{A} \begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \\ \vdots \\ \boldsymbol{\xi}_n \end{bmatrix} = \lambda_0 \begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \\ \vdots \\ \boldsymbol{\xi}_n \end{bmatrix}$ 的非零解向量 $(\boldsymbol{\xi}_1,$

$\boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_n)^T$ 称为 \boldsymbol{A} 的属于特征值 λ_0 的特征向量。

接下来介绍求特征值与特征向量的一般步骤:

- (1) 在 V 中任取一组基 $\boldsymbol{E}_1, \boldsymbol{E}_2, \dots, \boldsymbol{E}_n$, 写出 T 在这组基下的矩阵 \boldsymbol{A}
- (2) 求 \boldsymbol{A} 的特征多项式 $|\lambda\boldsymbol{I} - \boldsymbol{A}|$ 在 K 上的全部根, 它们就是 T 的全部特征值
- (3) 把所求得特征值逐个代入方程组 $(\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{X} = 0$, 并求出它的一组基础解系。(它们就是属于这个特征值的全部线性无关的特征向量在基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$ 下的坐标)

2.1.2 Jordan 标准型的求解

在 2.1.1 中, 我们简单介绍了欧氏空间与线性变换的意义。而将一个矩阵转化为其 Jordan 标准型, 意义在于选择适当的基, 使得后续的计算过程变得简单。

Jordan 标准型是线性代数中的一种矩阵形式，通过以下步骤求解：首先，计算线性变换的特征值和特征向量。然后，构建相应的 **Jordan** 块，选择适当的特征向量组合，构成线性变换的 **Jordan** 标准型。

首先我们先引入最小多项式的定义：

定义 2.8 首项系数是 1（简称首 1），次数最小，且以矩阵 \mathbf{A} 为根的 λ 的多项式，称为 \mathbf{A} 的最小多项式，常用 $m(\lambda)$ 表示。

为了证明矩阵论中极为重要的 **Hamilton-Cayley** 定理，需要先介绍如下定理。

定理 2.3 任意 n 阶矩阵与三角矩阵相似。

接下来就可以引入 **Hamilton-Cayley** 定理：

定义 2.9 n 阶矩阵 \mathbf{A} 是其特征多项式的矩阵根（零点），即令

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

则有

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

要在与 \mathbf{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。接下来引入定义，介绍 **Jordan** 标准型：

定义 2.10 由

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

给出的矩阵 \mathbf{J} 称为矩阵 \mathbf{A} 的 **Jordan** 标准形， $\mathbf{J}_i(\lambda_i)$ 称为因式 $(\lambda - \lambda_i)^{m_i}$ 对应的 **Jordan** 块，

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

可得如下定理：

定理 2.4 设 \mathbf{A} 是 n 阶复矩阵，且其特征多项式的某种分解式是

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

则存在 n 阶复可逆矩阵 P , 使

$$P^{-1}AP = J$$

该定理改用线性变换来说, 就是: 设 T 是复数域 \mathbf{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 J 。

解释完定理, 接下来引入一道例题 (《矩阵论》例 1.26), 解释如何求解 Jordan 标准形:

例题 1.26

$$\lambda I - A = \begin{bmatrix} \lambda+1 & -1 & 0 \\ 4 & \lambda-3 & 0 \\ -1 & 0 & \lambda-2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (\lambda-2)(\lambda-1)^2 \end{bmatrix} \Rightarrow A \sim J = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

图 2-4 例题 2.5

例题 2.5 求矩阵 A 的 Jordan 标准形, 其中 $A = \begin{bmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}$

解: 求 $\lambda I - A$ 的初等因子组, 由于 $\lambda I - A =$

$$\begin{bmatrix} \lambda+1 & -1 & 0 \\ 4 & \lambda-3 & 0 \\ -1 & 0 & \lambda-2 \end{bmatrix} \rightarrow \begin{bmatrix} -1 & 0 & 0 \\ \lambda-3 & (\lambda+1)(\lambda-3)+4 & 0 \\ 0 & -1 & \lambda-2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & (\lambda-1)^2 & 0 \\ 0 & -1 & \lambda-2 \end{bmatrix} \rightarrow$$

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

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

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

可以总结: 在复数域 \mathbf{C} 上, 求矩阵 A 的 Jordan 标准型的步骤:

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

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

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

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

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

$$J_i(\lambda_i) = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \lambda_i & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix}_{m_i \times m_i} \quad (i = 1, 2, \dots, s)$$

第三步: 写出以这些 Jordan 块构成的 Jordan 标准形

$$J = \begin{bmatrix} J_1(\lambda_1) & & & \\ & J_2(\lambda_2) & & \\ & & \ddots & \\ & & & J_s(\lambda_s) \end{bmatrix}$$

接下来求使矩阵 A 相似于 Jordan 标准形时所用的可逆矩阵 P , P 的求法如下:

第四步: 假如

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

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

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

即

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

由此可得

$$\left. \begin{aligned} (\lambda_1 I - A)\mathbf{x}_1 &= \mathbf{0} \\ (\lambda_2 I - A)\mathbf{x}_2 &= \mathbf{0} \\ (\lambda_2 I - A)\mathbf{x}_3 &= -\mathbf{x}_2 \end{aligned} \right\} \quad \text{式 (2-2)}$$

从而 $\mathbf{x}_1, \mathbf{x}_2$ 依次是 A 的属于 λ_1, λ_2 的特征向量。 \mathbf{x}_3 是式 (2-2) 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 P 。

于是可以继续求例题 2.5 中的矩阵 P :

$$(2I - A)x_1 = 0, \quad (I - A)x_2 = 0, \quad (I - A)x_3 = -x_2$$

得特征向量 x_1, x_2 及广义特征向量 x_3 依次为

$$x_1 = (0, 0, 1)^T, \quad x_2 = (1, 2, -1)^T, \quad x_3 = (0, 1, -1)^T$$

故所求的矩阵 P 是

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

2.1.3 欧式空间中线性变换的求法

在解析几何中, 通常 \mathbf{R}^3 中的向量长度、夹角等度量性质, 都可通过向量的数量积来表达。为了给线性空间引进长度等概念, 可先引入内积概念, 进而引出欧氏空间的定义。

定义 2.11 设 V 是实数域 \mathbf{R} 上的线性空间, 对于 V 中任意两个向量 x 与 y , 按照某种规则定义一个实数, 用 (x, y) 来表示, 且它满足下述 4 个条件:

- (1) 交换律: $(x, y) = (y, x)$;
- (2) 分配律: $(x, y + z) = (x, y) + (x, z)$;
- (3) 齐次性: $(kx, y) = k(x, y) \quad (\forall k \in \mathbf{R})$;
- (4) 非负性: $(x, x) \geq 0$, 当且仅当 $x = 0$ 时, $(x, x) = 0$.

则称实数 (x, y) 为向量 x 与 y 的内积, 而称 V 为 *Euclid* 空间, 简称欧氏空间或实内积空间。

接下来以一道例题 (《矩阵论》例 1.36) 说明欧氏空间中关于线性变换的常见问题的求法:

例题 2.6 在欧氏空间 $\mathbf{R}^{2 \times 2}$ 中, 矩阵 A 与 B 的内积定义为 $(A, B) = \text{tr}(A^T B)$, 子空间

$$V = \left\{ X = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \mid x_3 - x_4 = 0 \right\}$$

V 中的线性变换为

$$T(X) = XB_0 \quad (\forall X \in V), \quad B_0 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

- (1) 求 V 的一个标准正交基;
- (2) 验证 T 是 V 中的对称变换;
- (3) 求 V 的一个标准正交基, 使 T 在该基下的矩阵为对角矩阵.

解:

(1) 设 $\mathbf{X} \in V$, 则

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$

故 V 的一个标准正交基为

$$\mathbf{X}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{X}_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{X}_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$

(2) 计算基象组:

$$T(\mathbf{X}_1) = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} = 1\mathbf{X}_1 + 2\mathbf{X}_2 + 0\mathbf{X}_3$$

$$T(\mathbf{X}_2) = \begin{bmatrix} 2 & 1 \\ 0 & 0 \end{bmatrix} = 2\mathbf{X}_1 + 1\mathbf{X}_2 + 0\mathbf{X}_3$$

$$T(\mathbf{X}_3) = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 \\ 3 & 3 \end{bmatrix} = 0\mathbf{X}_1 + 0\mathbf{X}_2 + 3\mathbf{X}_3$$

设 $T(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)\mathbf{A}$, 则

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

易见 \mathbf{A} 是对称矩阵, T 是对称变换。

(3) 求正交矩阵 \mathbf{Q} 使得 $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} = \mathbf{J}$: 根据线性代数课程中介绍的算法, 求得

$$\mathbf{J} = \begin{bmatrix} 3 & & \\ & 3 & \\ & & -1 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 1 & 0 & 0 \end{bmatrix}$$

令 $(Y_1, Y_2, Y_3) = (X_1, X_2, X_3)Q$, 求得标准正交基

$$Y_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, \quad Y_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad Y_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}$$

总结思路如下:

第一小问, 求非空集合 V 的一组标准正交基, 因此我们先求得一组基, 再正交化、单位化, 得到 V 的标准正交基。

第二小问, 我们需要判断相应矩阵是否为对称矩阵。可以利用第一小问求出的标准正交基, 求线性变换的对应矩阵, 具体方法可以参考 2.1.1 节内容。于是证明该线性变换是对称变换。

第三小问, 要求 V 的一个标准正交基, 使 T 在该基下的矩阵为对角矩阵。实际上就是求一组新的标准正交基 (Y_1, Y_2, Y_3) 使得等式 $T(Y_1, Y_2, Y_3) = (Y_1, Y_2, Y_3)J$ 成立。

再引入一道例题 (第八次课 ppt-1) 进一步说明:

1. ppt P8 例题

令 $x_{11} = -x_{12} - x_{21}$

$$X = \begin{pmatrix} -x_{11} - x_{12} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}, \quad Y = \begin{pmatrix} -y_{11} - y_{12} & y_{12} \\ y_{21} & y_{22} \end{pmatrix}$$

$$= x_{12} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} + x_{21} \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix} + x_{22} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\Rightarrow Y_1' = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad Y_2' = \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}, \quad Y_3' = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\Rightarrow e_1 = \frac{1}{\sqrt{2}} Y_1', \quad e_2 = \frac{1}{\sqrt{2}} Y_2', \quad e_3 = Y_3'$$

$$\Rightarrow X = \begin{pmatrix} k_1 & -k_1 \\ 0 & k_2 \end{pmatrix} = (e_1, e_2, e_3) \begin{pmatrix} k_1 & -k_1 \\ k_2 & k_2 \end{pmatrix}, \quad k_1 = -k_2, \quad k_2 = 0, \quad k_3 = 0$$

$$T(e_1, \dots, e_n) = (e_1, \dots, e_n) \begin{pmatrix} 2\sqrt{2} & 0 \\ \sqrt{2} & 0 \\ 0 & 0 \end{pmatrix}, \quad A_0 = \begin{pmatrix} 2\sqrt{2} & 0 \\ \sqrt{2} & 0 \\ 0 & 0 \end{pmatrix}$$

$$\lambda I - A_0 = \begin{pmatrix} \lambda - 2\sqrt{2} & 0 \\ -\sqrt{2} & \lambda \\ 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda - 2\sqrt{2} & 0 \\ 0 & 0 & \lambda(\lambda - 3) \end{pmatrix} \Rightarrow J = \begin{pmatrix} 3 & & \\ & 2 & \\ & & 1 \end{pmatrix}$$

$$\begin{aligned} (3I - A_0)x_1 &= 0 & \Rightarrow x_1 &= \begin{pmatrix} \sqrt{2} \\ 1 \\ 0 \end{pmatrix}, \quad x_2 = \begin{pmatrix} -1 \\ \sqrt{2} \\ 0 \end{pmatrix}, \quad x_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\ (-I - A_0)x_2 &= 0 \\ (3I - A_0)x_3 &= 0 \end{aligned}$$

$$\Rightarrow P = \begin{pmatrix} \sqrt{2} & -1 & 0 \\ 1 & \sqrt{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad P^{-1} = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\Rightarrow E_1 = \frac{\sqrt{2}}{2} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad E_2 = \sqrt{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad E_3 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\Rightarrow X = (E_1, \dots, E_n) P^{-1} \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} = (E_1, E_2, E_3) \begin{pmatrix} -\sqrt{2} \\ \sqrt{2} \\ -1 \end{pmatrix}$$

$$T^3(x) = \begin{pmatrix} 108 & -52 \\ -56 & -21 \end{pmatrix}, \quad T^k(x) = (E_1, E_2, E_3) \begin{pmatrix} 3^k & & \\ & 2^k & \\ & & 1^k \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} (x_1 e_1) \\ (x_2 e_2) \\ (x_3 e_3) \end{pmatrix}$$

图 2-5 例题 2.7

整理成文字版:

例题 2.7 设矩阵空间 $\mathbf{R}^{2 \times 2}$ 的子空间为

$$V = \{X = (x_{ij})_{2 \times 2} \mid x_{11} + x_{12} + x_{21} = 0, x_{ij} \in \mathbf{R}\}$$

V 中的线性变换为 $T(X) = X + 2X^T$ 。

$$\text{求 } (T^3)(\mathbf{X}), \quad \mathbf{X} = \begin{pmatrix} 4 & -4 \\ 0 & -3 \end{pmatrix} \in V$$

$$\text{求 } (T^k)(\mathbf{X}), \quad \forall \mathbf{X} \in V$$

解：令 $x_{11} = -x_{12} - x_{21}$

$$\mathbf{X} = x_{11} \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} + x_{21} \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix} + x_{21} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = x_1 \mathbf{X}_1 + x_2 \mathbf{X}_2 + x_3 \mathbf{X}_3$$

Schmidt 正交化得到一组标准正交基：

$$\mathbf{Y}'_1 = \mathbf{X}_1 = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$$

$$\mathbf{Y}'_2 = \mathbf{X}_2 - \frac{(\mathbf{X}_2, \mathbf{Y}'_1)}{(\mathbf{Y}'_1, \mathbf{Y}'_1)} \mathbf{Y}'_1 = \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} \\ 1 & 0 \end{pmatrix}$$

$$\mathbf{Y}'_3 = \mathbf{X}_3 - \frac{(\mathbf{X}_3, \mathbf{Y}'_2)}{(\mathbf{Y}'_2, \mathbf{Y}'_2)} \mathbf{Y}'_2 - \frac{(\mathbf{X}_3, \mathbf{Y}'_1)}{(\mathbf{X}_3, \mathbf{Y}'_1)} \mathbf{Y}'_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

单位化：

$$\mathbf{e}_1 = \frac{1}{|\mathbf{Y}'_1|} \mathbf{Y}'_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$$

$$\mathbf{e}_2 = \frac{1}{|\mathbf{Y}'_2|} \mathbf{Y}'_2 = \frac{\sqrt{2}}{\sqrt{3}} \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} \\ 1 & 0 \end{pmatrix}$$

$$\mathbf{e}_3 = \frac{1}{|\mathbf{Y}'_3|} \mathbf{Y}'_3 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

求对应的 k_1, k_2, k_3 ：

$$x = \begin{pmatrix} 4 & -4 \\ 0 & -3 \end{pmatrix} = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix}$$

可得

$$k_1 = (x, \mathbf{e}_1) = -4\sqrt{2}, \quad k_2 = (x, \mathbf{e}_2) = 0, \quad k_3 = (x, \mathbf{e}_3) = -3$$

求 $T\mathbf{e}_1$ 对应的 k_1, k_2, k_3 :

$$T\mathbf{e}_1 = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix}$$

可得:

$$k_1 = (T\mathbf{e}_1, \mathbf{e}_1) = 2, \quad k_2 = (T\mathbf{e}_1, \mathbf{e}_2) = \sqrt{3} \quad k_3 = (T\mathbf{e}_1, \mathbf{e}_3) = 0$$

相同的方法可求得 $T\mathbf{e}_2, T\mathbf{e}_3$ 对应的 k_1, k_2, k_3 分别为:

$$k_1 = (T\mathbf{e}_2, \mathbf{e}_1) = \sqrt{3}, \quad k_2 = (T\mathbf{e}_2, \mathbf{e}_2) = 0 \quad k_3 = (T\mathbf{e}_2, \mathbf{e}_3) = 0$$

$$k_1 = (T\mathbf{e}_3, \mathbf{e}_1) = 0, \quad k_2 = (T\mathbf{e}_3, \mathbf{e}_2) = 0 \quad k_3 = (T\mathbf{e}_3, \mathbf{e}_3) = 3$$

由

$$T(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)\mathbf{A}_0$$

可得

$$\mathbf{A} = \begin{pmatrix} 2 & \sqrt{3} & 0 \\ \sqrt{3} & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

$$\lambda\mathbf{I} - \mathbf{A}_0 = \begin{pmatrix} \lambda - 2 & -\sqrt{3} & 0 \\ -\sqrt{3} & \lambda & 0 \\ 0 & 0 & \lambda - 3 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & (\lambda - 3) & 0 \\ 0 & 0 & (\lambda + 1)(\lambda - 3) \end{pmatrix}$$

不变因子: $d_1(\lambda) = 1, \quad d_2(\lambda) = (\lambda - 3), \quad d_3(\lambda) = (\lambda + 1)(\lambda - 3)$

初等因子: $(\lambda - 3); \quad (\lambda + 1), (\lambda - 3)$

初等因子组: $(\lambda - 3), \quad (\lambda + 1), \quad (\lambda - 3)$

Jordan 块: $\mathbf{J}_1(\lambda_1) = (3), \quad \mathbf{J}_2(\lambda_2) = (-1), \quad \mathbf{J}_3(\lambda_3) = (3)$

$$\text{Jordan 标准型: } \mathbf{J}_1 = \begin{pmatrix} 3 & & \\ & -1 & \\ & & 3 \end{pmatrix}$$

因此

$$\mathbf{P} = (x_1, x_2, x_3), \quad \mathbf{P}\mathbf{J} = \mathbf{A}_0\mathbf{P}$$

$$(3x_1, -x_2, 3x_3) = (\mathbf{A}_0x_1, \mathbf{A}_0x_2, \mathbf{A}_0x_3)$$

可知

$$x_1 = \begin{pmatrix} \sqrt{3} \\ 1 \\ 0 \end{pmatrix}, \quad x_2 = \begin{pmatrix} -1 \\ \sqrt{3} \\ 0 \end{pmatrix}, \quad x_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\mathbf{P} = (x_1, x_2, x_3) = \begin{pmatrix} \sqrt{3} & -1 & 0 \\ 1 & \sqrt{3} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{P}^{-1} = \begin{pmatrix} \frac{\sqrt{3}}{4} & \frac{1}{4} & 0 \\ -\frac{1}{4} & \frac{\sqrt{3}}{4} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

接下来求新的一组基:

$$T(\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3) = (\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3)\mathbf{J}$$

可求得

$$\mathbf{E}_1 = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = \begin{pmatrix} \sqrt{3} \\ 1 \\ 0 \end{pmatrix} = \frac{2}{\sqrt{6}} \begin{pmatrix} -2 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\mathbf{E}_2 = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = \begin{pmatrix} -1 \\ \sqrt{3} \\ 0 \end{pmatrix} = \sqrt{2} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$$\mathbf{E}_3 = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \sqrt{2} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

通过坐标变换得到

$$x = (\mathbf{E}_1, \dots, \mathbf{E}_n)\mathbf{P}^{-1} \begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix} = (\mathbf{E}_1, \dots, \mathbf{E}_n) \begin{pmatrix} l_1 \\ \vdots \\ l_n \end{pmatrix}$$

$$x = \begin{pmatrix} 4 & -4 \\ 0 & -3 \end{pmatrix} = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) \begin{pmatrix} -4\sqrt{2} \\ 0 \\ -3 \end{pmatrix} = (\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3)\mathbf{P}^{-1} \begin{pmatrix} -4\sqrt{2} \\ 0 \\ -3 \end{pmatrix} = (\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3) \begin{pmatrix} -\sqrt{6} \\ \sqrt{2} \\ -3 \end{pmatrix}$$

综上所述

$$T(x) = (\mathbf{E}_1, \dots, \mathbf{E}_n) \mathbf{J} \begin{pmatrix} l_1 \\ \vdots \\ l_n \end{pmatrix} \Rightarrow (T^k)(x) = (\mathbf{E}_1, \dots, \mathbf{E}_n) \mathbf{J}^k \begin{pmatrix} l_1 \\ \vdots \\ l_n \end{pmatrix}$$

$$(T^3)(x) = (\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_n) \begin{pmatrix} 27 & & \\ & -1 & \\ & & 27 \end{pmatrix} \begin{pmatrix} -\sqrt{6} \\ \sqrt{2} \\ -3 \end{pmatrix} = \begin{pmatrix} 108 & -52 \\ -56 & -81 \end{pmatrix}$$

$$(T^k)(x) = (\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_n) \begin{pmatrix} 3^k & & \\ & (-1)^k & \\ & & 3^k \end{pmatrix} \begin{pmatrix} \frac{\sqrt{3}}{4} & \frac{1}{4} & 0 \\ -\frac{1}{4} & \frac{\sqrt{3}}{4} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} (x, \mathbf{e}_1) \\ (x, \mathbf{e}_2) \\ (x, \mathbf{e}_3) \end{pmatrix}$$

2.2 向量范数与矩阵范数

2.2.1 向量范数介绍

由上述可见，向量的长度可用来刻画收敛的性质。对于一般的线性空间，可以用范数的概念定义向量的长度，定义如下：

定义 2.12 如果 V 是数域 K 上的线性空间，对任意的 $\mathbf{x} \in V$ ，定义一个实值函数 $\|\mathbf{x}\|$ ，它满足以下三个条件：

- (1) 非负性：当 $\mathbf{x} \neq 0$ 时， $\|\mathbf{x}\| > 0$ ；当 $\mathbf{x} = 0$ 时， $\|\mathbf{x}\| = 0$ ；
- (2) 齐次性： $\|a\mathbf{x}\| = |a| \|\mathbf{x}\|$ ($a \in K, \mathbf{x} \in V$)；
- (3) 三角不等式： $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ ($\mathbf{x}, \mathbf{y} \in V$)

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

以下是常见的范数：

- (1) $\|\mathbf{x}\|_1 = \sum |\xi_i|$ 是一种向量范数，记为 1-范数；
- (2) $\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})}$ 是一种向量范数，记 2-范数；
- (3) $\|\mathbf{x}\| = \max_i |\xi_i|$ 是一种向量范数，记为 ∞ -范数；
- (4) $\|\mathbf{x}\|_p = (\sum_{i=1}^n |\xi_i|^p)^{\frac{1}{p}}$ ($1 \leq p < \infty$) 是一种向量范数，记为 p -范数或 l_p 范数。

我们在任意线性空间中，都可以找到向量 \boldsymbol{x} 相应的 p -范数，方法如下：

例题 2.8 给定线性空间 V^n 的基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$ ，设 $\boldsymbol{x} \in V^n$ 在该基下的坐标向量为 $\boldsymbol{\alpha} = (\xi_1, \xi_2, \dots, \xi_n)^T$ ，那么

$$\|\boldsymbol{x}\|_p = \|\boldsymbol{\alpha}\|_p \quad (1 \leq p < +\infty)$$

满足范数定义三个条件。因此，它是 V^n 上的范数，也称为 \boldsymbol{x} 的 p -范数。

接下来讨论线性空间 V^n 上的向量范数的等价性。

定理 2.5 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

$$c_1 \|\boldsymbol{x}\|_\beta \leq \|\boldsymbol{x}\|_\alpha \leq c_2 \|\boldsymbol{x}\|_\beta \quad (\forall \boldsymbol{x} \in V) \quad \text{式 (2-3)}$$

定理 2.6 满足式 (2-3) 两种范数是等价的。

也就是说，有限维线性空间上的不同范数是等价的。

2.2.2 矩阵范数介绍

矩阵空间 $\mathbf{C}^{m \times n}$ 是一个 mn 维的线性空间，将 $m \times n$ 矩阵 \boldsymbol{A} 看做线性空间 $\mathbf{C}^{m \times n}$ 中的“向量”，可以按照例题 2.8 的方式定义 \boldsymbol{A} 的范数，不同的是矩阵之间还有乘法运算。下面引出矩阵范数的定义：

定义 2.13 设 $\boldsymbol{A} \in \mathbf{C}^{m \times n}$ ，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： $\|\alpha \boldsymbol{A}\| = |\alpha| \|\boldsymbol{A}\|$ ($\alpha \in \mathbf{C}$)；

(3) 三角不等式： $\|\boldsymbol{A} + \boldsymbol{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\|$ ($\boldsymbol{A} \in \mathbf{C}^{m \times n}$)，

则称 $\|\boldsymbol{A}\|$ 为 \boldsymbol{A} 的广义矩阵范数。若对 $\mathbf{C}^{m \times n}, \mathbf{C}^{n \times l}$ 及 $\mathbf{C}^{m \times l}$ 上的同类广义矩阵范数 $\|\cdot\|$ ，还满足下面一个条件：

(4) 相容性：

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

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

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实现：

定义 2.14 对于 $\mathbf{C}^{m \times n}$ 上的矩阵范数 $\|\cdot\|_M$ 和 \mathbf{C}^m 与 \mathbf{C}^n 上的同类向量范数 $\|\cdot\|_V$ ，如果

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

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

下面有几种常用的矩阵范数。可以先引入定理，规定矩阵范数，使矩阵范数与已知的向量范数相容。

定理 2.7 已知 \mathbf{C}^m 和 \mathbf{C}^n 上的同类向量范数 $\|\cdot\|$ ，设 $\mathbf{A} \in \mathbf{C}^{m \times n}$ ，则函数

$$\|\mathbf{A}\| = \max_{\|\mathbf{x}\|=1} \|\mathbf{Ax}\| \quad \text{式 (2-4)}$$

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

称由式 (2-4) 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 $\mathbf{C}^{n \times n}$ 上的任何一种从属范数，有

$$\|\mathbf{I}\| = \max_{\|\mathbf{x}\|=1} \|\mathbf{Ix}\| = 1$$

我们可以知道，矩阵范数是与向量范数密切相关的。以下定理介绍了三种常见的矩阵范数：

定理 2.8 设 $\mathbf{A} = (a_{ij})_{m \times n} \in \mathbf{C}^{m \times n}$ ， $\mathbf{x} = (\xi_1, \xi_2, \dots, \xi_n)^T \in \mathbf{C}^n$ ，则从属于向量 \mathbf{x} 的三种范数 $\|\mathbf{x}\|_1$ ， $\|\mathbf{x}\|_2$ ， $\|\mathbf{x}\|_\infty$ 的矩阵范数计算公式依次为：

(1) 列和范数 $\|\mathbf{A}\|_1 = \max_j \sum_{i=1}^m |a_{ij}|$ ；

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

(3) 行和范数 $\|\mathbf{A}\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|$ 。

2.2.3 矩阵可逆性条件、条件数和谱半径介绍

首先介绍矩阵的可逆性条件，可以根据范数 $\|\mathbf{A}\|$ 的大小来判断 $\mathbf{I} - \mathbf{A}$ 是否为可逆矩阵。引入如下定理：

定理 2.9 设 $\mathbf{A} \in \mathbf{C}^{n \times n}$ ，且对 $\mathbf{C}^{n \times n}$ 上的某种矩阵范数 $\|\cdot\|$ ，有 $\|\mathbf{A}\| < 1$ ，则矩阵 $\mathbf{I} - \mathbf{A}$ 可逆，且有

$$\|(\mathbf{I} - \mathbf{A})^{-1}\| \leq \frac{\|\mathbf{I}\|}{1 - \|\mathbf{A}\|}$$

接下来介绍条件数的概念：

定理 2.10 设 $\mathbf{A} \in \mathbf{C}^{n \times n}$ 可逆， $\mathbf{B} \in \mathbf{C}^{n \times n}$ ，且对 $\mathbf{C}^{n \times n}$ 上的某种矩阵范数 $\|\cdot\|$ ，有 $\|\mathbf{A}^{-1}\mathbf{B}\| < 1$ ，则有以下结论：

(1) $\mathbf{A} + \mathbf{B}$ 可逆；

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

(3) $\frac{\|A^{-1} - (A + B)^{-1}\|}{\|A^{-1}\|} \leq \frac{\|A^{-1}B\|}{1 - \|A^{-1}B\|}$.

对于该定理, 若令 $\text{cond}(A) = \|A\| \|A^{-1}\|$, $d_A = \|\delta A\| \|A\|^{-1}$, 则当 $\|A^{-1}\| \|\delta A\| < 1$ 时, 由结论 (2) 与 (3) 可得

$$\|I - (I + A^{-1}\delta A)^{-1}\| \leq \frac{d_A \text{cond}(A)}{1 - d_A \text{cond}(A)}$$

$$\frac{\|A^{-1} - (A + \delta A)^{-1}\|}{\|A^{-1}\|} \leq \frac{d_A \text{cond}(A)}{1 - d_A \text{cond}(A)}$$

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

接下来介绍矩阵的谱半径及其性质。矩阵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中都有重要地位, 现引入其定义与相关定理:

定义 2.15 设 $A \in \mathbf{C}^{n \times n}$ 的 n 个特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 称

$$\rho(A) = \max_i |\lambda_i|$$

为 A 的谱半径。

定理 2.11 设 $A \in \mathbf{C}^{n \times n}$, 则对 $\mathbf{C}^{n \times n}$ 上任何一种矩阵范数 $\|\cdot\|$, 都有

$$\rho(A) \leq \|A\|$$

定理 2.12 设 $A \in \mathbf{C}^{n \times n}$, 对任意的正数 ε , 存在某种矩阵范数 $\|\cdot\|$, 使得

$$\|A\|_M \leq \rho(A) + \varepsilon$$

2.3 矩阵函数介绍

2.3.1 矩阵序列介绍

矩阵序列收敛与发散的定义如下:

定义 2.16 设有矩阵序列 $\{A^{(k)}\}$, 其中 $A^{(k)} = (a_{ij}^{(k)})_{m \times n} \in \mathbf{C}^{m \times n}$, 当 $a_{ij}^{(k)} \rightarrow a_{ij} (k \rightarrow \infty)$ 时, 称 $\{A^{(k)}\}$ 收敛, 或称矩阵 $A = (a_{ij})_{m \times n}$ 为 $\{A^{(k)}\}$ 的极限, 或称 $\{A^{(k)}\}$ 收敛于 A , 记为

$$\lim_{k \rightarrow \infty} A^{(k)} = A \quad \text{或} \quad A^{(k)} \rightarrow A$$

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。接下来介绍几条性质：

(1) 设 $\mathbf{A}^{(k)} \rightarrow \mathbf{A}_{m \times n}$, $\mathbf{B}^{(k)} \rightarrow \mathbf{B}_{m \times n}$, 则

$$\lim_{k \rightarrow \infty} (\alpha \mathbf{A}^{(k)} + \beta \mathbf{B}^{(k)}) = \alpha \mathbf{A} + \beta \mathbf{B} \quad (\forall \alpha, \beta \in \mathbf{C})$$

(2) 设 $\mathbf{A}^{(k)} \rightarrow \mathbf{A}_{m \times n}$, $\mathbf{B}^{(k)} \rightarrow \mathbf{B}_{n \times l}$, 则

$$\lim_{k \rightarrow \infty} \mathbf{A}^{(k)} \mathbf{B}^{(k)} = \mathbf{A} \mathbf{B}$$

(3) 设 $\mathbf{A}^{(k)}$ 与 \mathbf{A} 都是可逆矩阵，且 $\mathbf{A}^{(k)} \rightarrow \mathbf{A}$, 则

$$(\mathbf{A}^{(k)})^{-1} \rightarrow \mathbf{A}^{-1}$$

由矩阵收敛的定义，可得如下定理：

定理 2.13 设 $\mathbf{A}^{(k)} \in \mathbf{C}^{m \times n}$, 则

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

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

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

在给出收敛矩阵的定义之前，需要定义矩阵的有界性：

定义 2.17 矩阵序列 $\{\mathbf{A}^{(k)}\}$ 称为有界的，如果存在常数 $M > 0$, 使得对一切 k 都有

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

定义 2.18 设 \mathbf{A} 为方阵，且 $\mathbf{A}^k \rightarrow \mathbf{O} (k \rightarrow \infty)$, 则称 \mathbf{A} 为收敛矩阵。

于是可得如下定理：

定理 2.14 \mathbf{A} 为收敛矩阵的充要条件是 $\rho(\mathbf{A}) < 1$ 。

定理 2.15 \mathbf{A} 为收敛矩阵的充分条件是只要有一种矩阵范数 $\|\cdot\|$, 使得 $\|\mathbf{A}\| < 1$ 。

2.3.2 矩阵级数介绍

矩阵级数的定义如下：

定义 2.19 把定义 2.16 中的矩阵序列所形成的无穷和 $\mathbf{A}^{(0)} + \mathbf{A}^{(1)} + \mathbf{A}^{(2)} + \dots + \mathbf{A}^{(k)} + \dots$ 称为矩阵级数, 记为 $\sum_{k=0}^{\infty} \mathbf{A}^{(k)}$, 则有

$$\sum_{k=0}^{\infty} \mathbf{A}^{(k)} = \mathbf{A}^{(0)} + \mathbf{A}^{(1)} + \mathbf{A}^{(2)} + \dots + \mathbf{A}^{(k)} + \dots \quad \text{式 (2-5)}$$

矩阵级数的收敛与发散定义如下:

定义 2.20 记 $\mathbf{S}^{(N)} = \sum_{k=0}^N \mathbf{A}^{(k)}$, 称其为矩阵级数式 (2-5) 的部分和。如果矩阵序列 $\{\mathbf{S}^{(N)}\}$ 收敛, 且有极限 \mathbf{S} , 则有

$$\lim_{N \rightarrow \infty} \mathbf{S}^{(N)} = \mathbf{S}$$

那么就称矩阵级数式 (式 2-5) 收敛, 而且有和 \mathbf{S} , 记为

$$\mathbf{S} = \sum_{k=0}^{\infty} \mathbf{A}^{(k)}$$

不收敛的矩阵级数称为是发散的。

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

$$\sum_{k=0}^{\infty} a_{ij}^{(k)} = s_{ij} \quad (i = 1, 2, \dots, m; j = 1, 2, \dots, n) \quad \text{式 (2-6)}$$

矩阵级数式 (式 2-5) 绝对收敛与其元素形成的级数式 (式 2-6) 绝对收敛是等价的, 因此得到如下定义:

定义 2.21 如果式 (2-6) 中左端 mn 个数项级数都是绝对收敛的, 则称矩阵级数式 (2-5) 是绝对收敛的。

接下来讨论矩阵幂级数, 给出如下定理:

定理 2.16 设方阵 \mathbf{A} 对某一矩阵范数 $\|\cdot\|$ 有 $\|\mathbf{A}\| < 1$, 则对任何非负整数 N , 以 $(\mathbf{I} - \mathbf{A})^{-1}$ 为部分和 $\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \dots + \mathbf{A}^N$ 的近似矩阵时, 其误差为

$$\|(\mathbf{I} - \mathbf{A})^{-1} - (\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \dots + \mathbf{A}^N)\| \leq \frac{\|\mathbf{A}\|^{N+1}}{1 - \|\mathbf{A}\|}$$

定理 2.17 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 \mathbf{A} 满足 $\rho(\mathbf{A}) < r$, 则矩阵幂级数

$$\sum_{k=0}^{\infty} c_k \mathbf{A}^k \quad \text{式 (2-7)}$$

是绝对收敛的;

如果 $\rho(\mathbf{A}) > r$, 则矩阵幂级数式 (式 2-7) 是发散的。

例题 2.9

$$S^{(N)} = \sum_{k=1}^N A^{(k)} = \begin{bmatrix} \sum_{k=1}^N \frac{1}{2^k} & \sum_{k=1}^N \frac{\pi}{3 \times 4^k} \\ 0 & \sum_{k=1}^N \frac{1}{k(k+1)} \end{bmatrix} = \begin{bmatrix} 1 - \left(\frac{1}{2}\right)^N & \frac{\pi}{9} \left[1 - \left(\frac{1}{4}\right)^N\right] \\ 0 & \frac{N}{N+1} \end{bmatrix}$$

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

图 2-6 例题 2.9

引入一道例题（《矩阵论》例 3.2），演示如何判断矩阵级数的敛散性：
整理成文字版本：

例题 2.9 研究矩阵级数 $\sum_{k=1}^{\infty} A^{(k)}$ 的收敛性，其中

$$A^{(k)} = \begin{bmatrix} \frac{1}{2^k} & \frac{\pi}{3 \times 4^k} \\ 0 & \frac{1}{k(k+1)} \end{bmatrix} \quad (k = 1, 2, \dots)$$

解： 因为

$$S^{(N)} = \sum_{k=1}^N A^{(k)} = \begin{bmatrix} \sum_{k=1}^N \frac{1}{2^k} & \sum_{k=1}^N \frac{\pi}{3 \times 4^k} \\ 0 & \sum_{k=1}^N \frac{1}{k(k+1)} \end{bmatrix} = \begin{bmatrix} 1 - \left(\frac{1}{2}\right)^N & \frac{\pi}{9} \left[1 - \left(\frac{1}{4}\right)^N\right] \\ 0 & \frac{N}{N+1} \end{bmatrix}$$

所以

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

于是由定义 2.20 知，所给级数收敛，且其和就是这里的二阶矩阵 S 。

再引入一道例题（第十一次课 ppt-2），求矩阵幂级数的和：

例题 2.10 求矩阵幂级数 $\sum_{k=0}^{\infty} \frac{1}{2^k} \begin{bmatrix} 2 & -1/2 \\ 2 & 0 \end{bmatrix}^k$ 的和

解：

$$\begin{pmatrix} 2 & -1/2 \\ 2 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 2 & -1 \end{pmatrix} = PJP^{-1}$$

$$\begin{aligned} \mathbf{A}^k &= \mathbf{P} \mathbf{J}^k \mathbf{P}^{-1} = \begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 2 & -1 \end{pmatrix} \\ \sum_{k=0}^{+\infty} \frac{1}{2^k} \mathbf{A}^k &= \mathbf{P} \left(\sum_{k=0}^{+\infty} \frac{1}{2^k} \mathbf{J}^k \right) \mathbf{P}^{-1} = \mathbf{P} \begin{pmatrix} \sum_{k=0}^{+\infty} \frac{1}{2^k} & \sum_{k=0}^{+\infty} \frac{k}{2^k} \\ 0 & \sum_{k=0}^{+\infty} \frac{1}{2^k} \end{pmatrix} \mathbf{P}^{-1} \\ 2M &= 1 + \frac{1+1}{2^1} + \frac{2+1}{2^2} + \frac{3+1}{2^3} + \dots + \frac{k+1}{2^k} + \dots, \\ 2M - M &= 1 + \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^3} + \dots + \frac{1}{2^k} + \dots = \sum_{k=0}^{+\infty} \frac{1}{2^k} = 2 \\ \sum_{k=0}^{+\infty} \frac{1}{2^k} \mathbf{A}^k &= \mathbf{P} \begin{pmatrix} \sum_{k=0}^{+\infty} \frac{1}{2^k} & \sum_{k=0}^{+\infty} \frac{k}{2^k} \\ 0 & \sum_{k=0}^{+\infty} \frac{1}{2^k} \end{pmatrix} \mathbf{P}^{-1} = \mathbf{P} \begin{pmatrix} 2 & 2 \\ 0 & 2 \end{pmatrix} \mathbf{P}^{-1} \\ &= \begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 2 & 2 \\ 0 & 2 \end{pmatrix} \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} 4 & -1 \\ 4 & 0 \end{pmatrix} \end{aligned}$$

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。矩阵函数的定义如下：

定义 2.22 设一元函数 $f(z)$ 能够展开为 z 的幂级数

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

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

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

举例说明，式 (2-8) 为矩阵指数函数；式 (2-9) 和式 (2-10) 为矩阵三角函数。

$$\mathbf{e}^{\mathbf{A}} = \mathbf{I} + \mathbf{A} + \frac{1}{2!} \mathbf{A}^2 + \frac{1}{3!} \mathbf{A}^3 + \dots \quad \text{式 (2-8)}$$

$$\cos \mathbf{A} = \mathbf{I} - \frac{1}{2!} \mathbf{A}^2 + \frac{1}{4!} \mathbf{A}^4 - \dots \quad \text{式 (2-9)}$$

$$\sin \mathbf{A} = \mathbf{A} - \frac{1}{3!}\mathbf{A}^3 + \frac{1}{5!}\mathbf{A}^5 - \dots \quad \text{式 (2-10)}$$

由此可以推出：

$$\begin{aligned} e^{j\mathbf{A}} &= \cos \mathbf{A} + j \sin \mathbf{A} \quad (j = \sqrt{-1}) \\ \cos \mathbf{A} &= \frac{1}{2}(e^{j\mathbf{A}} + e^{-j\mathbf{A}}), \quad \sin \mathbf{A} = \frac{1}{2j}(e^{j\mathbf{A}} - e^{-j\mathbf{A}}) \\ \cos(-\mathbf{A}) &= \cos \mathbf{A}, \quad \sin(-\mathbf{A}) = -\sin \mathbf{A} \end{aligned}$$

引入如下定理：

定理 2.18 如果 $\mathbf{AB} = \mathbf{BA}$ ，则 $e^{\mathbf{A}}e^{\mathbf{B}} = e^{\mathbf{B}}e^{\mathbf{A}} = e^{\mathbf{A}+\mathbf{B}}$ 。

关于矩阵函数的求法，在第三章中会详细介绍。

2.3.4 函数矩阵对矩阵的导数

函数矩阵是一类特殊的矩阵，其元素是函数而非常数。与之相对地，矩阵函数则是将矩阵作为自变量的函数。本文中的函数矩阵定义为：以变量 t 的函数 $a_{ij}(t)$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$) 为元素的矩阵 $\mathbf{A}(t) = (a_{ij}(t))_{m \times n}$ 。

定义 2.23 如果函数矩阵 $\mathbf{A}(t) = (a_{ij}(t))_{n \times n}$ 的每一个元素 $a_{ij}(t)$ 是变量的可导函数，则称 $\mathbf{A}(t)$ 可导，其导数（微商）定义为

$$\mathbf{A}'(t) = \frac{d}{dt}\mathbf{A}(t) = \left(\frac{d}{dt}a_{ij}(t) \right)_{m \times n}$$

可以得到如下定理：

定理 2.19 设 $\mathbf{A}(t), \mathbf{B}(t)$ 是能够进行下面运算的两个可导的函数矩阵，则有

$$\begin{aligned} \frac{d}{dt}(\mathbf{A}(t) + \mathbf{B}(t)) &= \frac{d}{dt}\mathbf{A}(t) + \frac{d}{dt}\mathbf{B}(t) \\ \frac{d}{dt}(\mathbf{A}(t)\mathbf{B}(t)) &= \frac{d}{dt}\mathbf{A}(t) \cdot \mathbf{B}(t) + \mathbf{A}(t) \cdot \frac{d}{dt}\mathbf{B}(t) \\ \frac{d}{dt}(a\mathbf{A}(t)) &= \frac{da}{dt} \cdot \mathbf{A}(t) + a \frac{d}{dt}\mathbf{A}(t) \end{aligned}$$

这里， $a = a(t)$ 为 t 的可导函数。

定理 2.20 设 n 阶矩阵 \mathbf{A} 与 t 无关，则有

$$\begin{aligned} \frac{d}{dt}e^{t\mathbf{A}} &= \mathbf{A}e^{t\mathbf{A}} = e^{t\mathbf{A}}\mathbf{A} \\ \frac{d}{dt}\cos(t\mathbf{A}) &= -\mathbf{A}(\sin(t\mathbf{A})) = -(\sin(t\mathbf{A}))\mathbf{A} \end{aligned}$$

$$\frac{d}{dt} \sin(t\mathbf{A}) = \mathbf{A}(\cos(t\mathbf{A})) = (\cos(t\mathbf{A}))\mathbf{A}$$

接下来讨论函数矩阵的积分运算。给出积分定义：

定义 2.24 如果函数矩阵 $\mathbf{A}(t)$ 的每个元素 $a_{ij}(t)$ 都是区间 $[t_0, t_1]$ 上的可积函数，则定义 $\mathbf{A}(t)$ 在 $[t_0, t_1]$ 上的积分为

$$\int_{t_0}^{t_1} \mathbf{A}(t) dt = \left(\int_{t_0}^{t_1} a_{ij}(t) dt \right)_{m \times n}$$

接下来讨论其他微分概念：

1. 函数对矩阵的导数

定义 2.25 设 $\mathbf{X} = (\xi_{ij})_{n \times n}$, mn 元函数 $f(\mathbf{X}) = f(\xi_{11}, \xi_{12}, \dots, \xi_{1n}, \xi_{21}, \dots, \xi_{mn})$, 定义 $f(\mathbf{X})$ 对矩阵 \mathbf{X} 的导数为

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

2. 函数对矩阵的导数

定义 2.26 设 $\mathbf{X} = (\xi_{ij})_{m \times n}$, mn 元函数 $f_{ij}(\mathbf{X}) = f_{ij}(\xi_{11}, \xi_{12}, \dots, \xi_{1n}, \xi_{21}, \dots, \xi_{mn})$ ($i = 1, 2, \dots, r; j = 1, 2, \dots, s$)。定义函数矩阵

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

对矩阵 X 的导数为

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

其中

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

求导公式表（第 12 次课 ppt-其他微分概念）如下：

Identities: vector-by-vector $\frac{\partial y}{\partial \mathbf{x}}$

Condition	Expression	Numerator layout, i.e. by y and \mathbf{x}^\top	Denominator layout, i.e. by y^\top and \mathbf{x}
a is not a function of \mathbf{x}	$\frac{\partial a}{\partial \mathbf{x}} =$	$\mathbf{0}$	
	$\frac{\partial \mathbf{x}}{\partial \mathbf{x}} =$	\mathbf{I}	
\mathbf{A} is not a function of \mathbf{x}	$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} =$	\mathbf{A}	\mathbf{A}^\top
\mathbf{A} is not a function of \mathbf{x}	$\frac{\partial \mathbf{x}^\top \mathbf{A}}{\partial \mathbf{x}} =$	\mathbf{A}^\top	\mathbf{A}
a is not a function of \mathbf{x} , $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial a\mathbf{u}}{\partial \mathbf{x}} =$	$a \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	
$a = a(\mathbf{x}), \mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial a\mathbf{u}}{\partial \mathbf{x}} =$	$a \frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u} \frac{\partial a}{\partial \mathbf{x}}$	$a \frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial a}{\partial \mathbf{x}} \mathbf{u}^\top$
\mathbf{A} is not a function of \mathbf{x} , $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial \mathbf{A}\mathbf{u}}{\partial \mathbf{x}} =$	$\mathbf{A} \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \mathbf{A}^\top$
$\mathbf{u} = \mathbf{u}(\mathbf{x}), \mathbf{v} = \mathbf{v}(\mathbf{x})$	$\frac{\partial (\mathbf{u} + \mathbf{v})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$	
$\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}$
$\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial f(\mathbf{g}(\mathbf{u}))}{\partial \mathbf{x}} =$	$\frac{\partial f(\mathbf{g})}{\partial \mathbf{g}} \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial f(\mathbf{g})}{\partial \mathbf{g}}$

图 2-7 求导公式表 1

Identities: scalar-by-vector $\frac{\partial y}{\partial \mathbf{x}} = \nabla_{\mathbf{x}} y$

Condition	Expression	Numerator layout, i.e. by \mathbf{x}^\top ; result is row vector	Denominator layout, i.e. by \mathbf{x} ; result is column vector
a is not a function of \mathbf{x}	$\frac{\partial a}{\partial \mathbf{x}} =$	$\mathbf{0}^\top$ [4]	$\mathbf{0}$ [4]
a is not a function of \mathbf{x} , $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial a\mathbf{u}}{\partial \mathbf{x}} =$	$a \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	
$u = u(\mathbf{x}), v = v(\mathbf{x})$	$\frac{\partial (u + v)}{\partial \mathbf{x}} =$	$\frac{\partial u}{\partial \mathbf{x}} + \frac{\partial v}{\partial \mathbf{x}}$	
$u = u(\mathbf{x}), v = v(\mathbf{x})$	$\frac{\partial uv}{\partial \mathbf{x}} =$	$u \frac{\partial v}{\partial \mathbf{x}} + v \frac{\partial u}{\partial \mathbf{x}}$	
$u = u(\mathbf{x})$	$\frac{\partial g(u)}{\partial \mathbf{x}} =$	$\frac{\partial g(u)}{\partial u} \frac{\partial u}{\partial \mathbf{x}}$	
$u = u(\mathbf{x})$	$\frac{\partial f(g(u))}{\partial \mathbf{x}} =$	$\frac{\partial f(g)}{\partial g} \frac{\partial g(u)}{\partial u} \frac{\partial u}{\partial \mathbf{x}}$	
$\mathbf{u} = \mathbf{u}(\mathbf{x}), \mathbf{v} = \mathbf{v}(\mathbf{x})$	$\frac{\partial (\mathbf{u} \cdot \mathbf{v})}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}^\top \mathbf{v}}{\partial \mathbf{x}} =$	$\mathbf{u}^\top \frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^\top \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$ • assumes numerator layout of $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}, \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \mathbf{v} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \mathbf{u}$ • assumes denominator layout of $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}, \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$
$\mathbf{u} = \mathbf{u}(\mathbf{x}), \mathbf{v} = \mathbf{v}(\mathbf{x})$, \mathbf{A} is not a function of \mathbf{x}	$\frac{\partial (\mathbf{u} \cdot \mathbf{A}\mathbf{v})}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}^\top \mathbf{A}\mathbf{v}}{\partial \mathbf{x}} =$	$\mathbf{u}^\top \mathbf{A} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^\top \mathbf{A}^\top \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$ • assumes numerator layout of $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}, \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \mathbf{A}\mathbf{v} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \mathbf{A}^\top \mathbf{u}$ • assumes denominator layout of $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}, \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$
	$\frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}^\top} =$	\mathbf{H} , the Hessian matrix ^[5]	

图 2-8 求导公式表 2

a is not a function of \mathbf{x}	$\frac{\partial(\mathbf{a} \cdot \mathbf{x})}{\partial \mathbf{x}} = \frac{\partial(\mathbf{x} \cdot \mathbf{a})}{\partial \mathbf{x}} =$ $\frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}} =$	\mathbf{a}^\top	\mathbf{a}
\mathbf{A} is not a function of \mathbf{x} \mathbf{b} is not a function of \mathbf{x}	$\frac{\partial \mathbf{b}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{b}^\top \mathbf{A}$	$\mathbf{A}^\top \mathbf{b}$
\mathbf{A} is not a function of \mathbf{x}	$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{x}^\top (\mathbf{A} + \mathbf{A}^\top)$	$(\mathbf{A} + \mathbf{A}^\top) \mathbf{x}$
\mathbf{A} is not a function of \mathbf{x} \mathbf{A} is symmetric	$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} =$	$2\mathbf{x}^\top \mathbf{A}$	$2\mathbf{A} \mathbf{x}$
\mathbf{A} is not a function of \mathbf{x}	$\frac{\partial^2 \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}^2} =$	$\mathbf{A} + \mathbf{A}^\top$	
\mathbf{A} is not a function of \mathbf{x} \mathbf{A} is symmetric	$\frac{\partial^2 \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}^2} =$	$2\mathbf{A}$	
	$\frac{\partial(\mathbf{x} \cdot \mathbf{x})}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{x}} =$	$2\mathbf{x}^\top$	$2\mathbf{x}$
a is not a function of \mathbf{x} , $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial(\mathbf{a} \cdot \mathbf{u})}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^\top \mathbf{u}}{\partial \mathbf{x}} =$	$\mathbf{a}^\top \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$ • assumes numerator layout of $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \mathbf{a}$ • assumes denominator layout of $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$
\mathbf{a}, \mathbf{b} are not functions of \mathbf{x}	$\frac{\partial \mathbf{a}^\top \mathbf{x} \mathbf{x}^\top \mathbf{b}}{\partial \mathbf{x}} =$	$\mathbf{x}^\top (\mathbf{a} \mathbf{b}^\top + \mathbf{b} \mathbf{a}^\top)$	$(\mathbf{a} \mathbf{b}^\top + \mathbf{b} \mathbf{a}^\top) \mathbf{x}$
$\mathbf{A}, \mathbf{b}, \mathbf{C}, \mathbf{D}, \mathbf{e}$ are not functions of \mathbf{x}	$\frac{\partial (\mathbf{A} \mathbf{x} + \mathbf{b})^\top \mathbf{C} (\mathbf{D} \mathbf{x} + \mathbf{e})}{\partial \mathbf{x}} =$	$(\mathbf{D} \mathbf{x} + \mathbf{e})^\top \mathbf{C}^\top \mathbf{A} + (\mathbf{A} \mathbf{x} + \mathbf{b})^\top \mathbf{C} \mathbf{D}$	$\mathbf{D}^\top \mathbf{C}^\top (\mathbf{A} \mathbf{x} + \mathbf{b}) + \mathbf{A}^\top \mathbf{C} (\mathbf{D} \mathbf{x} + \mathbf{e})$
\mathbf{a} is not a function of \mathbf{x}	$\frac{\partial \ \mathbf{x} - \mathbf{a}\ }{\partial \mathbf{x}} =$	$\frac{(\mathbf{x} - \mathbf{a})^\top}{\ \mathbf{x} - \mathbf{a}\ }$	$\frac{\mathbf{x} - \mathbf{a}}{\ \mathbf{x} - \mathbf{a}\ }$

图 2-9 求导公式表 3

Identities: vector-by-scalar $\frac{\partial \mathbf{y}}{\partial x}$

Condition	Expression	Numerator layout, i.e. by \mathbf{y} , result is column vector	Denominator layout, i.e. by \mathbf{y}^\top , result is row vector
\mathbf{a} is not a function of x	$\frac{\partial \mathbf{a}}{\partial x} =$	$\mathbf{0}^{[4]}$	
a is not a function of x , $\mathbf{u} = \mathbf{u}(x)$	$\frac{\partial a \mathbf{u}}{\partial x} =$	$a \frac{\partial \mathbf{u}}{\partial x}$	
\mathbf{A} is not a function of x , $\mathbf{u} = \mathbf{u}(x)$	$\frac{\partial \mathbf{A} \mathbf{u}}{\partial x} =$	$\mathbf{A} \frac{\partial \mathbf{u}}{\partial x}$	$\frac{\partial \mathbf{u}}{\partial x} \mathbf{A}^\top$
$\mathbf{u} = \mathbf{u}(x)$	$\frac{\partial \mathbf{u}^\top}{\partial x} =$	$\left(\frac{\partial \mathbf{u}}{\partial x} \right)^\top$	
$\mathbf{u} = \mathbf{u}(x), \mathbf{v} = \mathbf{v}(x)$	$\frac{\partial (\mathbf{u} + \mathbf{v})}{\partial x} =$	$\frac{\partial \mathbf{u}}{\partial x} + \frac{\partial \mathbf{v}}{\partial x}$	
$\mathbf{u} = \mathbf{u}(x), \mathbf{v} = \mathbf{v}(x)$	$\frac{\partial (\mathbf{u}^\top \times \mathbf{v})}{\partial x} =$	$\left(\frac{\partial \mathbf{u}}{\partial x} \right)^\top \times \mathbf{v} + \mathbf{u}^\top \times \frac{\partial \mathbf{v}}{\partial x}$	$\frac{\partial \mathbf{u}}{\partial x} \times \mathbf{v} + \mathbf{u}^\top \times \left(\frac{\partial \mathbf{v}}{\partial x} \right)^\top$
$\mathbf{u} = \mathbf{u}(x)$	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial x} =$	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial x}$	$\frac{\partial \mathbf{u}}{\partial x} \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}$
Assumes consistent matrix layout; see below.			
$\mathbf{u} = \mathbf{u}(x)$	$\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial x} =$	$\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}} \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial x}$	$\frac{\partial \mathbf{u}}{\partial x} \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}$
Assumes consistent matrix layout; see below.			
$\mathbf{U} = \mathbf{U}(x), \mathbf{v} = \mathbf{v}(x)$	$\frac{\partial (\mathbf{U} \times \mathbf{v})}{\partial x} =$	$\frac{\partial \mathbf{U}}{\partial x} \times \mathbf{v} + \mathbf{U} \times \frac{\partial \mathbf{v}}{\partial x}$	$\mathbf{v}^\top \times \left(\frac{\partial \mathbf{U}}{\partial x} \right) + \frac{\partial \mathbf{v}}{\partial x} \times \mathbf{U}^\top$

图 2-10 求导公式表 4

Identities: scalar-by-matrix $\frac{\partial y}{\partial \mathbf{X}}$

Condition	Expression	Numerator layout, i.e. by \mathbf{X}^T	Denominator layout, i.e. by \mathbf{X}
a is not a function of \mathbf{X}	$\frac{\partial a}{\partial \mathbf{X}} = \mathbf{0}^{[6]}$	$\mathbf{0}^{[6]}$	$\mathbf{0}^{[6]}$
a is not a function of \mathbf{X} , $u = u(\mathbf{X})$	$\frac{\partial au}{\partial \mathbf{X}} = a \frac{\partial u}{\partial \mathbf{X}}$		
$u = u(\mathbf{X}), v = v(\mathbf{X})$	$\frac{\partial(u+v)}{\partial \mathbf{X}} = \frac{\partial u}{\partial \mathbf{X}} + \frac{\partial v}{\partial \mathbf{X}}$		
$u = u(\mathbf{X}), v = v(\mathbf{X})$	$\frac{\partial uv}{\partial \mathbf{X}} = u \frac{\partial v}{\partial \mathbf{X}} + v \frac{\partial u}{\partial \mathbf{X}}$		
$u = u(\mathbf{X})$	$\frac{\partial y(u)}{\partial \mathbf{X}} = \frac{\partial y(u)}{\partial u} \frac{\partial u}{\partial \mathbf{X}}$		
$u = u(\mathbf{X})$	$\frac{\partial f(g(u))}{\partial \mathbf{X}} = \frac{\partial f(g)}{\partial g} \frac{\partial g(u)}{\partial u} \frac{\partial u}{\partial \mathbf{X}}$		
$\mathbf{U} = \mathbf{U}(\mathbf{X})$	$\frac{\partial \mathbf{g}(\mathbf{U})}{\partial X_{ij}} =$	$\text{tr} \left(\frac{\partial \mathbf{g}(\mathbf{U})}{\partial \mathbf{U}} \frac{\partial \mathbf{U}}{\partial X_{ij}} \right)$	$\text{tr} \left(\left(\frac{\partial \mathbf{g}(\mathbf{U})}{\partial \mathbf{U}} \right)^T \frac{\partial \mathbf{U}}{\partial X_{ij}} \right)$
		Both forms assume numerator layout for $\frac{\partial \mathbf{U}}{\partial X_{ij}}$, i.e. mixed layout if denominator layout for \mathbf{X} is being used.	
a and b are not functions of \mathbf{X}	$\frac{\partial a^T \mathbf{X} b}{\partial \mathbf{X}} =$	ba^T	ab^T
a and b are not functions of \mathbf{X}	$\frac{\partial a^T \mathbf{X}^T b}{\partial \mathbf{X}} =$	ab^T	ba^T
a, b and \mathbf{C} are not functions of \mathbf{X}	$\frac{\partial (\mathbf{X}a + b)^T \mathbf{C} (\mathbf{X}a + b)}{\partial \mathbf{X}} =$	$((\mathbf{C} + \mathbf{C}^T)(\mathbf{X}a + b)a^T)^T$	$(\mathbf{C} + \mathbf{C}^T)(\mathbf{X}a + b)a^T$
a, b and \mathbf{C} are not functions of \mathbf{X}	$\frac{\partial (\mathbf{X}a)^T \mathbf{C} (\mathbf{X}b)}{\partial \mathbf{X}} =$	$(\mathbf{C}\mathbf{X}ba^T + \mathbf{C}^T \mathbf{X}ab^T)^T$	$\mathbf{C}\mathbf{X}ba^T + \mathbf{C}^T \mathbf{X}ab^T$

图 2-11 求导公式表 5

Identities: matrix-by-scalar $\frac{\partial \mathbf{Y}}{\partial x}$

Condition	Expression	Numerator layout, i.e. by \mathbf{Y}
$\mathbf{U} = \mathbf{U}(x)$	$\frac{\partial a \mathbf{U}}{\partial x} = a \frac{\partial \mathbf{U}}{\partial x}$	$a \frac{\partial \mathbf{U}}{\partial x}$
\mathbf{A}, \mathbf{B} are not functions of x , $\mathbf{U} = \mathbf{U}(x)$	$\frac{\partial \mathbf{A} \mathbf{U} \mathbf{B}}{\partial x} =$	$\mathbf{A} \frac{\partial \mathbf{U}}{\partial x} \mathbf{B}$
$\mathbf{U} = \mathbf{U}(x), \mathbf{V} = \mathbf{V}(x)$	$\frac{\partial (\mathbf{U} + \mathbf{V})}{\partial x} =$	$\frac{\partial \mathbf{U}}{\partial x} + \frac{\partial \mathbf{V}}{\partial x}$
$\mathbf{U} = \mathbf{U}(x), \mathbf{V} = \mathbf{V}(x)$	$\frac{\partial (\mathbf{U} \mathbf{V})}{\partial x} =$	$\mathbf{U} \frac{\partial \mathbf{V}}{\partial x} + \frac{\partial \mathbf{U}}{\partial x} \mathbf{V}$
$\mathbf{U} = \mathbf{U}(x), \mathbf{V} = \mathbf{V}(x)$	$\frac{\partial (\mathbf{U} \otimes \mathbf{V})}{\partial x} =$	$\mathbf{U} \otimes \frac{\partial \mathbf{V}}{\partial x} + \frac{\partial \mathbf{U}}{\partial x} \otimes \mathbf{V}$
$\mathbf{U} = \mathbf{U}(x), \mathbf{V} = \mathbf{V}(x)$	$\frac{\partial (\mathbf{U} \circ \mathbf{V})}{\partial x} =$	$\mathbf{U} \circ \frac{\partial \mathbf{V}}{\partial x} + \frac{\partial \mathbf{U}}{\partial x} \circ \mathbf{V}$
$\mathbf{U} = \mathbf{U}(x)$	$\frac{\partial \mathbf{U}^{-1}}{\partial x} =$	$-\mathbf{U}^{-1} \frac{\partial \mathbf{U}}{\partial x} \mathbf{U}^{-1}$
$\mathbf{U} = \mathbf{U}(x, y)$	$\frac{\partial^2 \mathbf{U}^{-1}}{\partial x \partial y} =$	$\mathbf{U}^{-1} \left(\frac{\partial \mathbf{U}}{\partial x} \mathbf{U}^{-1} \frac{\partial \mathbf{U}}{\partial y} - \frac{\partial^2 \mathbf{U}}{\partial x \partial y} + \frac{\partial \mathbf{U}}{\partial y} \mathbf{U}^{-1} \frac{\partial \mathbf{U}}{\partial x} \right) \mathbf{U}^{-1}$
any matrix function defined s its derivative, and $\mathbf{g}(\mathbf{X})$ is	$\frac{\partial \mathbf{g}(x\mathbf{A})}{\partial x} =$	$\mathbf{A} \mathbf{g}'(x\mathbf{A}) = \mathbf{g}'(x\mathbf{A}) \mathbf{A}$
\mathbf{A} is not a function of x	$\frac{\partial e^{x\mathbf{A}}}{\partial x} =$	$\mathbf{A} e^{x\mathbf{A}} = e^{x\mathbf{A}} \mathbf{A}$

图 2-12 求导公式表 6

第三章 矩阵函数的求法研究

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 \mathbf{A} 的特征多项式为 $\varphi(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A})$ 。如果首 1 多项式

$$\psi(\lambda) = \lambda^m + b_1\lambda^{m-1} + \cdots + b_{m-1}\lambda + b_m \quad (1 \leq m \leq n) \quad \text{式 (3-1)}$$

满足 $\psi(\mathbf{A}) = \mathbf{O}$ 且 $\psi(\lambda)$ 整除 $\varphi(\lambda)$ (矩阵 \mathbf{A} 的最小多项式与特征多项式均满足这些条件)。那么, $\psi(\lambda)$ 的零点都是 \mathbf{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, \dots, r_s ($r_1 + r_2 + \dots + r_s = m$), 则有

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

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

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

其中 $r(z)$ 是次数低于 m 的多项式, 于是可由

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

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

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

3.1.2 举例展示求法

整理成文字版:

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

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

例题3.5

$$\varphi(\lambda) = \det(\lambda I - A) = (\lambda - 2)^3 \Rightarrow m(\lambda) = (\lambda - 2)^3, \psi(\lambda) = (\lambda - 2)^2$$

$$e^A: f(\lambda) = e^\lambda, \text{ 设 } f(\lambda) = \psi(\lambda)g(\lambda) + (a + b\lambda)$$

$$\Rightarrow \begin{cases} f(2) = e^2 \\ f'(2) = e^2 \end{cases} \text{ 或 } \begin{cases} a + 2b = e^2 \\ b = e^2 \end{cases} \Rightarrow \begin{cases} a = -e^2, b = e^2 \\ r(\lambda) = e^2(\lambda - 1) \end{cases}$$

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

$$e^{tA}: f(\lambda) = e^{t\lambda}$$

$$\Rightarrow \begin{cases} f(2) = e^{2t} \\ f'(2) = te^{2t} \end{cases} \text{ 或 } \begin{cases} a + 2b = e^{2t} \\ b = te^{2t} \end{cases} \Rightarrow \begin{cases} a = (1 - 2t)e^{2t} \\ b = te^{2t} \end{cases}$$

$$\Rightarrow r(\lambda) = e^{2t}[(1 - 2t) + t\lambda]$$

$$\Rightarrow e^{tA} = f(A) = r(A) = e^{2t} \begin{bmatrix} 1 & 0 & 0 \\ t & 1 - t & t \\ t & -t & 1 + t \end{bmatrix}$$

图 3-1 例题 3.1

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

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

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

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

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

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

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

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

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 $\psi(\lambda)$ 形如式 (3-1), 且满足 $\psi(\mathbf{A}) = \mathbf{O}$, 即

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

或者

$$\mathbf{A}^m = k_0 \mathbf{I} + k_1 \mathbf{A} + \cdots + k_{m-1} \mathbf{A}^{m-1} \quad (k_i = -b_{m-i}) \quad \text{式 (3-2)}$$

由此可以求出

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

于是有

$$\begin{aligned} f(\mathbf{A}) &= \sum_{k=0}^{\infty} c_k \mathbf{A}^k = (c_0 \mathbf{I} + c_1 \mathbf{A} + \cdots + c_{m-1} \mathbf{A}^{m-1}) + \\ & c_m (k_0 \mathbf{I} + k_1 \mathbf{A} + \cdots + k_{m-1} \mathbf{A}^{m-1}) + \cdots + \\ & c_{m+l} (k_0^{(l)} \mathbf{I} + k_1^{(l)} \mathbf{A} + \cdots + k_{m-1}^{(l)} \mathbf{A}^{m-1}) + \cdots = \\ & \left(c_0 + \sum_{l=0}^{\infty} c_{m+l} k_0^{(l)} \right) \mathbf{I} + \left(c_1 + \sum_{l=0}^{\infty} c_{m+l} k_1^{(l)} \right) \mathbf{A} + \cdots + \\ & \left(c_{m-1} + \sum_{l=0}^{\infty} c_{m+l} k_{m-1}^{(l)} \right) \mathbf{A}^{m-1} \end{aligned} \quad \text{式 (3-3)}$$

这表明, 利用式 (3-3) 可以将一个矩阵幂级数的求和问题, 转化为 m 个数项级数的求和问题。当式 (3-2) 中只有少数几个系数不为零时, 式 (3-3) 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

例题 3.2 设 $\mathbf{A} = \begin{bmatrix} \pi & 0 & 0 & 0 \\ 0 & -\pi & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, 求 $\sin \mathbf{A}$

解: $\varphi(\lambda) = \det(\lambda \mathbf{I} - \mathbf{A}) = \lambda^4 - \pi^2 \lambda^2$ 。由于 $\varphi(\mathbf{A}) = \mathbf{O}$, 所以 $\mathbf{A}^4 = \pi^2 \mathbf{A}^2, \mathbf{A}^5 =$

$\pi^2 \mathbf{A}^3, \mathbf{A}^7 = \pi^4 \mathbf{A}^3, \dots$. 于是有

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

3.3 对角型法

3.3.1 对角型法求矩阵函数的步骤推导

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

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

则有 $\mathbf{A} = \mathbf{P} \mathbf{A} \mathbf{P}^{-1}, \mathbf{A}^2 = \mathbf{P} \mathbf{A}^2 \mathbf{P}^{-1}, \dots$, 于是可得

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

从而

$$\begin{aligned} f(\mathbf{A}) &= \sum_{k=0}^{\infty} c_k \mathbf{A}^k = \mathbf{P} \begin{bmatrix} \sum_{k=0}^{\infty} c_k \lambda_1^k & & \\ & \ddots & \\ & & \sum_{k=0}^{\infty} c_k \lambda_n^k \end{bmatrix} \mathbf{P}^{-1} \\ &= \mathbf{P} \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix} \mathbf{P}^{-1} \end{aligned} \quad \text{式 (3-4)}$$

这表明，当 \mathbf{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

用一道例题（《矩阵论》例 3.7）进行演示：

例题 3.3

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

$$\Rightarrow \lambda_1 = -2, \mathbf{p}_1 = (-1, 1, 1)^T$$

$$\lambda_2 = \lambda_3 = 1, \mathbf{p}_2 = (-2, 1, 0)^T, \mathbf{p}_3 = (0, 0, 1)^T$$

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

$$\Rightarrow e^{\mathbf{A}t} = \begin{bmatrix} 2e^{-\frac{t}{2}} & 2e^{-\frac{t}{2}} & 0 \\ \frac{t}{2}e^{-\frac{t}{2}} & \frac{t}{2}e^{-\frac{t}{2}} & 0 \\ \frac{t}{2}e^{-\frac{t}{2}} & \frac{t}{2}e^{-\frac{t}{2}} & e^t \end{bmatrix}$$

$$e^{t\mathbf{A}} = \begin{bmatrix} 2e^t - \frac{t}{2}e^t & 2e^t - \frac{t}{2}e^t & 0 \\ \frac{t}{2}e^t - e^t & \frac{t}{2}e^t - e^t & 0 \\ \frac{t}{2}e^t - e^t & \frac{t}{2}e^t - 2e^t & e^t \end{bmatrix}$$

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

图 3-2 例题 3.3

整理成文字版：

例题 3.3 设 $\mathbf{A} = \begin{bmatrix} 4 & 6 & 0 \\ -3 & -5 & 0 \\ -3 & -6 & 1 \end{bmatrix}$ ，分别求 $e^{\mathbf{A}}$, $e^{t\mathbf{A}} (t \in \mathbf{R})$ 及 $\cos \mathbf{A}$

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

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

则有

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

利用式 (3-4), 求得

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

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

$$\cos A = P \begin{bmatrix} \cos(-2) & & \\ & \cos 1 & \\ & & \cos 1 \end{bmatrix} P^{-1} = \begin{bmatrix} 2 \cos 1 - \cos 2 & 2 \cos 1 - 2 \cos 2 & 0 \\ \cos 2 - \cos 1 & 2 \cos 2 - \cos 1 & 0 \\ \cos 2 - \cos 1 & 2 \cos 2 - \cos 1 & \cos 1 \end{bmatrix}$$

3.4 Jordan 标准型法

3.4.1 Jordan 标准型法求矩阵函数的步骤推导

设 A 的 Jordan 标准形为 J , 则有可逆矩阵 P , 使得

$$P^{-1}AP = J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_s \end{bmatrix}$$

其中

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ & & & \lambda_i \end{bmatrix}_{m_i \times m_i}$$

可求得

$$\begin{aligned}
 f(\mathbf{J}_i) &= \sum_{k=0}^{\infty} c_k \mathbf{J}_i^k = \sum_{k=0}^{\infty} c_k \begin{bmatrix} \lambda_i^k & C_k^1 \lambda_i^{k-1} & \cdots & C_k^{m_i-1} \lambda_i^{k-m_i+1} \\ & \lambda_i^k & \ddots & \vdots \\ & & \ddots & C_k^1 \lambda_i^{k-1} \\ & & & \lambda_i^k \end{bmatrix} \\
 &= \begin{bmatrix} f(\lambda_i) & \frac{1}{1!} f'(\lambda_i) & \cdots & \frac{1}{(m_i-1)!} f^{(m_i-1)}(\lambda_i) \\ & f(\lambda_i) & \ddots & \vdots \\ & & \ddots & \frac{1}{1!} f'(\lambda_i) \\ & & & f(\lambda_i) \end{bmatrix} \quad \text{式 (3-5)}
 \end{aligned}$$

$$\begin{aligned}
 f(\mathbf{A}) &= \sum_{k=0}^{\infty} c_k \mathbf{A}^k = \sum_{k=0}^{\infty} c_k \mathbf{P} \mathbf{J}^k \mathbf{P}^{-1} = \mathbf{P} \left(\sum_{k=0}^{\infty} c_k \mathbf{J}^k \right) \mathbf{P}^{-1} = \\
 \mathbf{P} \begin{bmatrix} \sum_{k=0}^{\infty} c_k \mathbf{J}_1^k & & \\ & \ddots & \\ & & \sum_{k=0}^{\infty} c_k \mathbf{J}_s^k \end{bmatrix} \mathbf{P}^{-1} &= \mathbf{P} \begin{bmatrix} f(\mathbf{J}_1) & & \\ & \ddots & \\ & & f(\mathbf{J}_s) \end{bmatrix} \mathbf{P}^{-1} \quad \text{式 (3-6)}
 \end{aligned}$$

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

依旧使用例题 3.2 的题干，但是可以使用 Jordan 标准型法进行求解：

例题 3.4 设 $\mathbf{A} = \begin{bmatrix} \pi & 0 & 0 & 0 \\ 0 & -\pi & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ ，求 $\sin \mathbf{A}$

解： 矩阵 A 是一个 *Jordan* 标准形，它的三个 *Jordan* 块为

$$\mathbf{J}_1 = \pi, \quad \mathbf{J}_2 = -\pi, \quad \mathbf{J}_3 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

根据式 (3-5)，求得

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

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

再由式 (3-6)，可得 (取 $\mathbf{P} = \mathbf{I}$)

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

第四章 矩阵分解方法研究

4.1 矩阵的 LU 分解

4.1.1 矩阵 LU 分解的步骤推导

当条件式 $\Delta_r \neq 0$ ($r = 1, 2, \dots, n-1$) 满足时, 由 $\mathbf{A}^{(r-1)} = \mathbf{L}_r \mathbf{A}^{(r)}$ 有

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

容易求出

$$\mathbf{L} = \mathbf{L}_1 \mathbf{L}_2 \dots \mathbf{L}_{n-1} = \begin{bmatrix} 1 & & & & \\ c_{21} & 1 & & & \\ \vdots & \vdots & \ddots & & \\ c_{n-1,1} & c_{n-1,2} & \dots & 1 & \\ c_{n1} & c_{n2} & \dots & c_{n,n-1} & 1 \end{bmatrix} \quad \text{式 (4-1)}$$

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 $\mathbf{A}^{(n-1)} = \mathbf{U}$ (或 \mathbf{R}), 则得

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

这样 \mathbf{A} 就分解成一个单位下三角矩阵与一个上三角矩阵的乘积。

接下来给出 LU 分解和 LDU 分解的定义:

定义 4.1 如果方阵 \mathbf{A} 可分解成一个下三角矩阵 \mathbf{L} 和一个上三角矩阵 \mathbf{U} 的乘积, 则称 \mathbf{A} 可作三角分解或 LU (LR) 分解。如果方阵 \mathbf{A} 可分解成 $\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{U}$, 其中 \mathbf{L} 是单位下三角矩阵, \mathbf{D} 是对角矩阵, \mathbf{U} 是单位上三角矩阵, 则称 \mathbf{A} 可作 LDU 分解。

于是可以另外知道直接计算可逆矩阵 \mathbf{A} 的三角分解的方法, 给出 Doolittle 分解和 Crout 分解的定义:

定义 4.2 设矩阵 \mathbf{A} 有唯一的 LDU 分解, 若把 $\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{U}$ 中的 \mathbf{D} 与 \mathbf{U} 结合起来, 并且用 $\hat{\mathbf{U}}$ 来表示, 就得到唯一的分解为

$$\mathbf{A} = \mathbf{L}(\mathbf{D}\mathbf{U}) = \mathbf{L}\hat{\mathbf{U}}$$

称为 \mathbf{A} 的 Doolittle 分解; 若把 $\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{U}$ 中的 \mathbf{L} 与 \mathbf{D} 结合起来, 并且用 $\hat{\mathbf{L}}$ 来表示, 就得到唯一的分解为

$$\mathbf{A} = (\mathbf{L}\mathbf{D})\mathbf{U} = \hat{\mathbf{L}}\mathbf{U}$$

称为 \mathbf{A} 的 Crout 分解。

当 A 为实对称正定矩阵时, $\Delta_k > 0 (k = 1, 2, \dots, n)$. 于是 A 有唯一的 LDU 分解, 即 $A = LDU$, 其中 $D = \text{diag}(d_1, d_2, \dots, d_n)$, 且 $d_i > 0 (i = 1, 2, \dots, n)$. 令

$$\tilde{D} = \text{diag}(\sqrt{d_1}, \sqrt{d_2}, \dots, \sqrt{d_n})$$

则有 $A = L\tilde{D}^2U$. 由 $A^T = A$ 得到 $L\tilde{D}^2U = U^T\tilde{D}^2L^T$, 再由分解的唯一性有 $L = U^T, U = L^T$, 因而有

$$A = L\tilde{D}^2L^T = LDL^T$$

或者

$$A = L\tilde{D}^2L^T = (L\tilde{D})(L\tilde{D})^T = GG^T \quad \text{式 (4-2)}$$

这里 $G = L\tilde{D}$ 是下三角矩阵。

可引出 Cholesky 分解的定义如下:

定义 4.3 称式 (4-2) 为实对称正定矩阵的 Cholesky 分解 (平方根分解、对称三角分解)。

4.1.2 举例展示求法

以一道例题 (《矩阵论》例 4.1) 说明 LU 分解的方法:

2. 例题 4.1

$\Delta_1 = 2, \Delta_2 = 5 \Rightarrow$ LDU 分解

$$L_1 = \begin{bmatrix} 1 & & \\ \frac{1}{2} & 1 & \\ & & 1 \end{bmatrix}, L_1^T = \begin{bmatrix} 1 & & \\ & \frac{1}{2} & \\ & & 1 \end{bmatrix}$$

$$\Rightarrow L_1^{-1}A^{(0)} = \begin{bmatrix} 2 & -1 & 3 \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 5 & -1 \end{bmatrix} = A^{(1)}$$

$$L_2 = \begin{bmatrix} 1 & & \\ 0 & 1 & \\ 0 & 2 & 1 \end{bmatrix}, L_2^T = \begin{bmatrix} 1 & & \\ 0 & 1 & \\ 0 & -2 & 1 \end{bmatrix}$$

$$L_2^{-1}A^{(1)} = \begin{bmatrix} 2 & -1 & 3 \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{5}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & -\frac{1}{5} \\ 0 & 0 & 1 \end{bmatrix} = A^{(2)}$$

$$\Rightarrow L = L_1 L_2 = \begin{bmatrix} 1 & & \\ \frac{1}{2} & 1 & \\ 1 & 2 & 1 \end{bmatrix}$$

$$\Rightarrow A^{(0)}: A = L L_2 A^{(2)} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{5}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & -\frac{1}{5} \\ 0 & 0 & 1 \end{bmatrix}$$

图 4-1 例题 4.1

整理成文字版本:

例题 4.1 求矩阵 A 的 LDU 分解, 其中

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

解: 因为 $\Delta_1 = 2, \Delta_2 = 5$, 所以 A 有唯一的 LDU 分解。构造矩阵

$$L_1 = \begin{bmatrix} 1 & & \\ \frac{1}{2} & 1 & \\ 1 & 0 & 1 \end{bmatrix}, \quad L_1^{-1} = \begin{bmatrix} 1 & & \\ -\frac{1}{2} & 1 & \\ -1 & 0 & 1 \end{bmatrix}$$

计算, 得

$$L_1^{-1}A^{(0)} = \begin{bmatrix} 2 & -1 & 3 \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 5 & -1 \end{bmatrix} = A^{(1)}$$

对 $A^{(1)}$ 构造矩阵, 有

$$L_2 = \begin{bmatrix} 1 & & \\ 0 & 1 & \\ 0 & 2 & 1 \end{bmatrix}, \quad L_2^{-1} = \begin{bmatrix} 1 & & \\ 0 & 1 & \\ 0 & -2 & 1 \end{bmatrix}$$

计算, 得

$$L_2^{-1}A^{(1)} = \begin{bmatrix} 2 & -1 & 3 \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{5}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & -\frac{1}{5} \\ 0 & 0 & 1 \end{bmatrix} = A^{(2)}$$

由式 (4-1) 可求出

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

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

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

4.2 矩阵的 QR 分解

4.2.1 矩阵 QR 分解的步骤推导

先给出 QR 分解的定义与相关定理:

定义 4.4 如果实 (复) 可逆矩阵 \mathbf{A} 能够化成正交 (酉) 矩阵 \mathbf{Q} 与实 (复) 可逆上三角矩阵 \mathbf{R} 的乘积, 即

$$\mathbf{A} = \mathbf{Q}\mathbf{R} \quad \text{式 (4-3)}$$

则称式 (4-3) 为 \mathbf{A} 的 QR 分解。

定理 4.1 设 \mathbf{A} 是 n 阶实 (复) 可逆矩阵, 则存在正交 (酉) 矩阵 \mathbf{Q} 和实 (复) 可逆上三角矩阵 \mathbf{R} , 使 \mathbf{A} 有 QR 分解式 (4-3)。除去相差一个对角元素的绝对值 (模) 全等于 1 的对角矩阵因子外, 分解式 (4-3) 是唯一的。

证: 记矩阵 \mathbf{A} 的 n 个列向量依次为 $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$, 因为 \mathbf{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 $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$ 。

对 $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ 正交化, 可得

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

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

$$\begin{cases} \mathbf{a}_1 = \mathbf{b}_1 \\ \mathbf{a}_2 = k_{21}\mathbf{b}_1 + \mathbf{b}_2 \\ \dots\dots\dots \\ \mathbf{a}_n = k_{n1}\mathbf{b}_1 + k_{n2}\mathbf{b}_2 + \dots + k_{n,n-1}\mathbf{b}_{n-1} + \mathbf{b}_n \end{cases}$$

用矩阵形式表示为

$$(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)\mathbf{C}$$

其中

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

再对 $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$ 单位化, 可得

$$\mathbf{q}_i = \frac{1}{|\mathbf{b}_i|} \mathbf{b}_i \quad (i = 1, 2, \dots, n)$$

于是有

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

令

$$\mathbf{Q} = (\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n)$$

$$\mathbf{R} = \text{diag}(|\mathbf{b}_1|, |\mathbf{b}_2|, \dots, |\mathbf{b}_n|) \cdot \mathbf{C}$$

式 (4-4)

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

为了证明唯一性, 设 \mathbf{A} 有两个分解式 $\mathbf{A} = \mathbf{QR} = \mathbf{Q}_1\mathbf{R}_1$, 由此可得

$$\mathbf{Q} = \mathbf{Q}_1\mathbf{R}_1\mathbf{R}^{-1} = \mathbf{Q}_1\mathbf{D}$$

其中 $\mathbf{D} = \mathbf{R}_1\mathbf{R}^{-1}$ 仍为可逆上三角矩阵, 于是

$$\mathbf{I} = \mathbf{Q}^H\mathbf{Q} = (\mathbf{Q}_1\mathbf{D})^H(\mathbf{Q}_1\mathbf{D}) = \mathbf{D}^H\mathbf{D}$$

这表明 D 不仅为正交 (酉) 矩阵, 而且还是对角元素的绝对值 (模) 全为 1 的对角矩阵。从而 $R_1 = DR, Q_1 = QD^{-1}$ 。

证明结束。

该定理推广如下:

定理 4.2 设 A 是 $m \times n$ 实 (复) 矩阵, 且其 n 个列线性无关, 则 A 有分解

$$A = QR$$

其中 Q 是 $m \times n$ 实 (复) 矩阵, 且满足 $Q^T Q = I (Q^H Q = I)$, R 是 n 阶实 (复) 可逆上三角矩阵。该分解除去相差一个对角元素的绝对值 (模) 全等于 1 的对角矩阵因子外是唯一的。

4.2.2 举例展示求法

以一道例题 (《矩阵论》例 4.6) 演示矩阵的 QR 分解:

整理成文字版本:

例题 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 $\mathbf{a}_1 = (1, 2, 1)^T, \mathbf{a}_2 = (2, 1, 2)^T, \mathbf{a}_3 = (2, 2, 1)^T$, 正交化可得

$$\mathbf{b}_1 = \mathbf{a}_1 = (1, 2, 1)^T$$

$$\mathbf{b}_2 = \mathbf{a}_2 - \mathbf{b}_1 = (1, -1, 1)^T$$

$$\mathbf{b}_3 = \mathbf{a}_3 - \frac{1}{3}\mathbf{b}_2 - \frac{7}{6}\mathbf{b}_1 = \left(\frac{1}{2}, 0, -\frac{1}{2}\right)^T$$

根据式 (4-4) 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 $A = QR$ 。

4.3 矩阵的满秩分解

4.3.1 矩阵满秩分解的步骤推导

满秩分解指的是将非零矩阵分解为列满秩矩阵与行满秩矩阵的乘积，在广义逆矩阵的研究中意义重大。

先给出满秩分解的定义：

定义 4.5 设 $A \in C_r^{m \times n} (r > 0)$ ，如果存在矩阵 $F \in C_r^{m \times r}$ 和 $G \in C_r^{r \times n}$ ，使得

$$A = FG \quad \text{式 (4-5)}$$

则称式 (4-5) 为矩阵 A 的满秩分解。

当 A 是满秩 (列满秩或行满秩) 矩阵时， A 可分解为一个因子是单位矩阵，另一个因子是 A 本身，称此满秩分解为平凡分解。

矩阵满秩分解的步骤推导可由该定理及其证明给出：

定理 4.3 设 $A \in C_r^{m \times n} (r > 0)$ ，则 A 有满秩分解式 (4-5)。

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

$$A \xrightarrow{\text{行}} B = \begin{bmatrix} G \\ O \end{bmatrix}, \quad G \in C_r^{r \times n}$$

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

$$P^{-1} = \left[\begin{array}{c|c} F & S \end{array} \right] \quad (F \in C_r^{m \times r}, S \in C_{m-r}^{m \times (m-r)})$$

则有

$$A = P^{-1}B = \left[\begin{array}{c|c} f & S \end{array} \right] \begin{bmatrix} G \\ O \end{bmatrix} = FG$$

其中 F 是列满秩矩阵， G 是行满秩矩阵。证明结束。

需要指出，矩阵 A 的满秩分解式 (4-5) 不是唯一的。这是因为若取 D 是任一个 r 阶可逆矩阵，则式 (4-5) 可改写为

$$A = (FD)(D^{-1}G) = \tilde{F}\tilde{G}$$

这是 A 的另一个满秩分解。

按照定理 4.3，可以使用矩阵的初等行变换方法求矩阵的满秩分解。

4.3.2 举例展示求法

以课本例 4.10 为示例，演示矩阵的 QR 分解：

3. 例 4.10

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

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

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

图 4-2 例题 4.3

整理成文字版本：

例题 4.3 求矩阵 A 的满秩分解，其中

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

解： 根据定理 4.3 的证明过程提供的算法，需要求出阶梯形矩阵 B 及诸初等矩阵的乘积 P 。为此，对矩阵 $[A: I]$ 进行初等行变换，当 A 所在的位置成为阶梯形矩阵 B 时， I 所在的位置就是进行初等行变换对应的初等矩阵的乘积 P ，即

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

$$\xrightarrow{\text{行}} \left[\begin{array}{cccc|ccc} -1 & 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 3 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 1 \end{array} \right]$$

则有

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

可求得

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

于是有

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

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

矩阵的奇异值分解在最优化问题、特征值问题、最小二乘方问题、广义逆矩阵问题及统计学等方面都有重要应用。先给出奇异值的定义和一个定理：

定义 4.6 设 $\mathbf{A} \in \mathbf{C}_r^{m \times n} (r > 0)$, $\mathbf{A}^H \mathbf{A}$ 的特征值为

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0$$

则称 $\sigma_i = \sqrt{\lambda_i} (i = 1, 2, \dots, n)$ 为 \mathbf{A} 的奇异值；当 \mathbf{A} 为零矩阵时，它的奇异值都是 0。

定理 4.4 设 $\mathbf{A} \in \mathbf{C}_r^{m \times n} (r > 0)$ ，则存在 m 阶酉矩阵 \mathbf{U} 和 n 阶酉矩阵 \mathbf{V} ，使得

$$\mathbf{U}^H \mathbf{A} \mathbf{V} = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \quad \text{式 (4-6)}$$

其中 $\mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$ ，而 $\sigma_i (i = 1, 2, \dots, r)$ 为矩阵 \mathbf{A} 的全部非零奇异值。

接下来介绍矩阵奇异值分解的步骤：

1. 令 $\mathbf{B} = \mathbf{A}^H \mathbf{A}$ ，计算特征值特征向量 $\mathbf{B}\mathbf{x} = \lambda\mathbf{x}$ ，得到 λ_k, \mathbf{v}_k

$$\mathbf{B}\mathbf{v}_k = \lambda_k \mathbf{v}_k, \quad k = 1, 2, \dots, N$$

$V = (\nu_1, \dots, \nu_N)$ 是酉矩阵: $V^H V = I$

$$\begin{aligned} BV = V\Lambda &\Rightarrow A^H AV = V\Lambda \Rightarrow A^H AV_1 = V_1 \Sigma^2 \\ &\Rightarrow A^H AV_1 \Sigma^{-1} = V_1 \Sigma \end{aligned}$$

2. 令 $U_1 = AV_1 \Sigma^{-1}$, 有 $A^H U_1 = V_1 \Sigma$, 可知

$$U_1^H A = \Sigma V_1^H \Rightarrow A = U_1 \Sigma V_1^H$$

3. 将 U_1 扩充成酉矩阵 $U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}$

$$A = UDV^H$$

4.4.2 举例展示求法

以课本例 4.14 为示例, 演示矩阵的 QR 分解:

3. 例 4.14

$$B = A^T A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}, \lambda_1 = 3, \lambda_2 = 1, \lambda_3 = 0 \Rightarrow \xi_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \xi_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \xi_3 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

$$\Rightarrow \text{rank} A = 2, \Sigma = \begin{bmatrix} \sqrt{3} & & 0 \\ & 1 & \\ & & 0 \end{bmatrix}$$

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

$$U_1 = AV_1 \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & 0 \end{bmatrix}, U_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, U = [U_1; U_2] = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\Rightarrow A = U \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} V^T$$

图 4-3 例题 4.4

整理成文字版本:

例题 4.4 求矩阵 $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ 的奇异值分解。

解:

1.

$$A^H AV = V\Lambda \Rightarrow A^H AV_1 = V_1 \Sigma^2 \Rightarrow A^H AV_1 \Sigma^{-1} = V_1 \Sigma$$

即

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix} = \mathbf{B}, \quad |\lambda \mathbf{I} - \mathbf{B}| = \lambda(\lambda - 1)(\lambda - 3)$$

$$\lambda_1 = 3: \quad 3\mathbf{I} - \mathbf{B} = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & -1 \\ -1 & -1 & 1 \end{bmatrix}, \quad \xi_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

$$\lambda_2 = 1: \quad 1\mathbf{I} - \mathbf{B} = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & -1 \\ -1 & -1 & -1 \end{bmatrix}, \quad \xi_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

$$\lambda_3 = 0: \quad 0\mathbf{I} - \mathbf{B} = \begin{bmatrix} -1 & 0 & -1 \\ 0 & -1 & -1 \\ -1 & -1 & -2 \end{bmatrix}, \quad \xi_s = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

$$r_A = 2: \quad \Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{V} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \end{bmatrix}, \quad \mathbf{V}_1 = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{6}} & 0 \end{bmatrix}$$

2.

$$\mathbf{U}_1^H \mathbf{A} = \Sigma \mathbf{V}_1^H \implies \mathbf{A} = \mathbf{U}_1 \Sigma \mathbf{V}_1^H$$

即

$$\mathbf{U}_1 = \mathbf{A} \mathbf{V}_1 \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & 0 \end{bmatrix}$$

$$3. \text{ 取 } \mathbf{U}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \text{ 则 } \mathbf{U} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ 可得}$$

$$\mathbf{U}^T \mathbf{A} \mathbf{V} = \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \mathbf{D}$$

则 \mathbf{A} 的奇异值分解为

$$\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{V}^T = \mathbf{U} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{V}^T$$

4.5 利用矩阵分解求矩阵广义逆

4.5.1 矩阵广义逆介绍

逆矩阵的概念只是对可逆矩阵才有意义，但是在实际问题中，遇到的矩阵不一定是方阵，即便是方阵也不一定可逆。

广义逆矩阵就是将逆矩阵概念进一步推广得到的，若：

1. 该矩阵对于不可逆矩阵甚至长方矩阵都存在；
2. 它具有通常逆矩阵的一些性质；
3. 当矩阵可逆时，它还原到通常的逆矩阵。

称满足以上三个条件的矩阵为广义逆矩阵。

现介绍 Penrose 的广义逆矩阵的定义：

定义 4.7 设矩阵 $\mathbf{A} \in \mathbf{C}^{m \times n}$ ，若矩阵 $\mathbf{X} \in \mathbf{C}^{n \times m}$ 满足以下 4 个 Penrose 方程

$$(1) \mathbf{A} \mathbf{X} \mathbf{A} = \mathbf{A}$$

$$(2) \mathbf{X} \mathbf{A} \mathbf{X} = \mathbf{X}$$

$$(3) (\mathbf{A} \mathbf{X})^H = \mathbf{A} \mathbf{X}$$

$$(4) (\mathbf{X} \mathbf{A})^H = \mathbf{X} \mathbf{A}$$

则称 \mathbf{X} 为 \mathbf{A} 的 Moore-Penrose 逆，记为 \mathbf{A}^+ 。

4.5.2 利用矩阵满秩分解求矩阵广义逆

下面的定理说明任意矩阵的 Moore-Penrose 逆存在，并给出满秩分解的一般方法：

定理 4.5 对任意 $A \in \mathbf{C}^{m \times n}$, A^+ 存在并且唯一。

证明其存在性 (唯一性证明省略)：

设 $\text{rank} A = r$, 若 $r = 0$, 则 $A = O_{m \times n}$, 由定义知 $A^+ = O_{n \times m}$; 若 $r > 0$, 由定理 4.4 可知, A 可进行满秩分解：

$$A = FG \quad (F \in \mathbf{C}_r^{m \times r}, G \in \mathbf{C}_r^{r \times n})$$

令 $X = G^+F^+$, 则有

$$AXA = FG \cdot G^+F^+ \cdot FG = FG = A$$

$$XAX = G^+F^+ \cdot FG \cdot G^+F^+ = G^+F^+ = X$$

$$(AX)^H = (FG \cdot G^+F^+)^H = (F F^+)^H = F F^+ = F \cdot F G^+ \cdot F^+ = AX$$

$$(XA)^H = (G^+F^+ \cdot FG)^H = (G^+ G)^H = G^+ G = G^+ \cdot F^+ F \cdot G = XA$$

故

$$A^+ = G^+F^+ = G^H(F^H A G^H)^{-1} F^H \quad \text{式 (4-7)}$$

式 (4-7) 即为任意矩阵的 Moore-Penrose 逆的满秩分解方法。

4.5.3 利用矩阵奇异值分解求矩阵广义逆

下面的定理给出奇异值分解的一般方法：

定理 4.6 设 $A \in \mathbf{C}_r^{m \times n}$ 的不可逆值分解为

$$A = U \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix}_{m \times n} V^H$$

那么

$$A^+ = V \begin{bmatrix} \Sigma_r^{-1} & O \\ O & O \end{bmatrix}_{n \times m} U^H \quad \text{式 (4-8)}$$

这里 U, V, Σ_r 的意义同定理 4.4。

证：直接验证 $X = V \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix}_{n \times m} U^H$ 满足 4 个 Penrose 方程即可。

式 (4-8) 是任意矩阵的 Moore-Penrose 逆的奇异值分解方法。

4.5.4 举例展示求法

接下来以两道例题（第十五次课 ppt-2）展示如何利用矩阵满秩分解、奇异值分解求矩阵广义逆。

例题 4.5 设 $\mathbf{A} = \begin{bmatrix} -1 & 0 & 1 \\ 2 & 0 & -2 \end{bmatrix}$ ，求 \mathbf{A}^+ 。

解：（方法一）利用满秩分解公式可得

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

从而 \mathbf{A} 的伪逆矩阵是

$$\begin{aligned} \mathbf{A}^+ &= \mathbf{C}^H(\mathbf{C}\mathbf{C}^H)^{-1}(\mathbf{B}^H\mathbf{B})^{-1}\mathbf{B}^H \\ &= \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \left(\begin{bmatrix} 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right)^{-1} \left(\begin{bmatrix} -1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right)^{-1} \begin{bmatrix} -1 & 2 \end{bmatrix} \\ &= \frac{1}{10} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \begin{bmatrix} -1 & 2 \end{bmatrix} = \frac{1}{10} \begin{bmatrix} -1 & 2 \\ 0 & 0 \\ 1 & -2 \end{bmatrix} \end{aligned}$$

（方法二）先求 \mathbf{A} 的奇异值分解：

$$\mathbf{A}\mathbf{A}^H = \begin{bmatrix} 2 & -4 \\ -4 & 8 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 10 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}$$

令

$$\mathbf{V}_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & -2 \\ \sqrt{5} & -\sqrt{5} \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 \\ 2 & 0 & -2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ 2 \\ -\frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \sqrt{10} \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\mathbf{A}^+ = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{bmatrix} = \frac{1}{10} \begin{bmatrix} -1 & 2 \\ 0 & 0 \\ 1 & -2 \end{bmatrix}$$

例题 4.6 设 $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$, 求 A^+ 。

解: (方法一) 由满秩分解公式可得

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

于是其伪逆矩阵为

$$\begin{aligned} A^+ &= C^H(CC^H)^{-1}(B^HB)^{-1}B^H \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} (\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix})^{-1} (\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix})^{-1} \begin{bmatrix} 1 & 2 \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} \\ &= \frac{1}{10} \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \end{aligned}$$

(方法二) 先求 A 的奇异值分解:

$$AA^H = \begin{bmatrix} 2 & 4 \\ 4 & 8 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ 2 & -\frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 10 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix}$$

令:

$$V_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 2 \\ \sqrt{5} & \sqrt{5} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$A = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ 2 \\ \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \sqrt{10} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$A^+ = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 1 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$$

第五章 总结

参考文献

- [1] 张凯院, 徐仲等. 矩阵论 [M]. 西北工业大学出版社, 2017.