

# 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 杨晨

学 院 计算机学院

专 业 计算机类

班 级 2021211321

学 号 2021212171

班内序号                     

任课教师 李昊辰

2022 年    12 月

# 矩阵函数的求法和矩阵分解方法研究

## 摘 要

本论文介绍了欧式空间中的线性变换，介绍了范数的概念，研究了矩阵函数的求解方法和矩阵分解的方法。

**关键词：**矩阵函数 矩阵分解 范数 线性变换

# **Research on the method of matrix function and matrix factorization method**

## **ABSTRACT**

This paper introduces linear transformations in Euclidean space, introduces the concept of norms, and studies the solution methods of matrix functions and matrix factorization. (本课程不考翻译，此处翻译地不好不扣分)

**KEY WORDS:** Matrix functions Matrix factorization Norm Linear transformation

# 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 .....	3
1.4 本论文工作简述 .....	3
1.4.1 本论文对上述问题研究简述 .....	3
1.4.2 本论文创新点或特点简述 .....	3
1.4.3 本论文撰写结构简述 .....	3
第二章 预备知识 .....	4
2.1 欧式空间与线性变换 .....	4
2.1.1 线性变换介绍 .....	4
2.1.2 若尔当标准型的求解 .....	10
2.1.3 欧式空间中线性变换的求法(可参考课本例 1.36 和 ppt) .....	13
2.2 向量范数与矩阵范数 .....	19
2.2.1 向量范数介绍 .....	19
2.2.2 矩阵范数介绍 .....	20
2.2.3 矩阵可逆性条件、谱半径和条件数介绍 .....	21
2.3 矩阵函数介绍 .....	22
2.3.1 矩阵序列介绍 .....	22
2.3.2 矩阵级数介绍 .....	23
2.3.3 矩阵函数介绍 (参考课本 3.3.1) .....	25
2.4 函数矩阵对矩阵的导数 .....	25
第三章 矩阵函数的求法研究 .....	33
3.1 待定系数法 .....	33

3.1.1 待定系数法求矩阵函数的步骤推导 .....	33
3.1.2 举例展示求法 .....	33
3.2 数项级数求和法 .....	34
3.2.1 数项级数求和法求矩阵函数的步骤推导 .....	34
3.2.2 举例展示求法 .....	34
3.3 对角型法 .....	35
3.3.1 对角型法求矩阵函数的步骤推导 .....	35
3.3.2 举例展示求法 .....	35
3.4 若尔当标准型法 .....	36
3.4.1 若尔当标准型法求矩阵函数的步骤推导 .....	36
3.4.2 举例展示求法 .....	37
第四章 矩阵分解方法研究 .....	38
4.1 矩阵的 LU 分解 .....	38
4.1.1 矩阵 LU 分解的步骤推导 .....	38
4.1.2 举例展示求法 .....	39
4.2 矩阵的 QR 分解 .....	39
4.2.1 矩阵 QR 分解的步骤推导 .....	39
4.2.2 举例展示求法 .....	41
4.3 矩阵的满秩分解 .....	41
4.3.1 矩阵满秩分解的步骤推导 .....	41
4.3.2 举例展示求法 .....	42
4.4 矩阵的奇异值分解 .....	42
4.4.1 矩阵奇异值分解的步骤推导 .....	42
4.4.2 举例展示求法 .....	43
4.4.3 利用奇异值分解求矩阵广义逆 .....	44
第五章 总结 .....	46
参考文献 .....	48

# 第一章 引言

## 1.1 背景介绍

### 1.1.1 矩阵理论与方法介绍

很多实际问题，如两点边值问题<sup>①</sup>

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

的数值求解，往往转化为求解如下线性系统

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

其中矩阵  $A$  和向量  $\bar{b}$  是由下式得到的

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

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

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

$$Ax = b \quad (2)$$

其中  $A = A(f_1(D))$ ,  $b = b(f_2(D))$

$D$  是一个  $N$  阶方阵， $f_1(D)$  和  $f_2(D)$  分别是关于矩阵  $D$  的矩阵函数，矩阵  $A$  和向量  $b$  分别是关于  $f_1(D)$  和  $f_2(D)$  的函数矩阵。

系统(2)可以利用矩阵分解来求解。

### 1.1.2 函数矩阵和矩阵函数介绍

矩阵函数形如  $y = f(A)$ ,  $A \in R^{n \times n}$

设幕级数  $\sum_{m=0}^{\infty} c_m z^m$  的收敛半径是  $R$ ，且在收敛域内  $f(z) = \sum_{m=0}^{\infty} c_m z^m$ ，当矩阵  $A$  的谱半径  $\rho(A) < R$ ，定义  $f(A) = \sum_{m=0}^{\infty} c_m A^m$ ，称  $f(A)$  为矩阵  $A$  的函数。

### 1.1.3 线性代数方程组求解介绍

考虑非齐次线性方程组

$$Ax = b$$

其中  $A \in C^{m \times n}$ ,  $b \in C^m$  给定，而  $x \in C^n$  为待定向量。如果存在向量  $x$  使方程组成立，则称方程组相容，否则称为不相容或矛盾方程组。

关于线性方程组的求解问题，常见的有以下几种情形。

- (1) 方程组相容的条件是什么？在相容时求出其通解（若解不唯一的话）。
- (2) 如果方程组相容，其解可能有无穷多个，求出具有极小范数的解，即

$$\min_{Ax=b} \|x\|$$

其中  $\|\cdot\|$  是欧氏范数。可以证明，满足该条件的解是唯一的，称之为极小范数解。

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

(3) 如果方程组不相容, 则不存在通常意义下的解. 但在许多实际问题中, 需要求出极值问题

$$\min_{x \in \mathbb{C}^n} \|Ax - b\|$$

的解  $x$ , 其中  $\|\cdot\|$  是欧氏范数. 称这个极值问题为求矛盾方程组的最小二乘问题, 相应的  $x$  称为矛盾方程组的最小二乘解.

(4) 一般说来, 矛盾方程组的最小二乘解是不唯一的. 但在最小二乘解的集合中, 具有极小范数的解

$$\min_{\|Ax-b\|} \|x\|$$

是唯一的, 称之为极小范数最小二乘解.

## 1.2 问题介绍

### 1.2.1 矩阵函数的求法问题介绍

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

解:

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

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

解:

$$\begin{aligned} e^A &= \sum_{k=0}^{+\infty} \frac{1}{k!} A^k = I + A + \frac{1}{2!} A^2 + \frac{1}{3!} A^3 + \dots \\ &= \lim_{N \rightarrow +\infty} \sum_{k=0}^N \frac{1}{k!} A^k \end{aligned}$$

因为矩阵  $A$  可以被分解, 即

$$\begin{aligned} A &= \begin{pmatrix} 4 & 6 & 0 \\ -3 & -5 & 0 \\ -3 & -6 & 1 \end{pmatrix} = P^{-1} \Lambda P \\ &= \begin{pmatrix} -1 & -2 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} -2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ -1 & -1 & 0 \\ -1 & -2 & 1 \end{pmatrix} \end{aligned}$$

所以

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

### 1.2.2 矩阵分解的方法问题介绍

对于可逆方阵 $A$ ，可以进行 **LU 分解**，在此基础上，可以将  $U$  进一步分解，这样得到的便是 **LDU 分解**。

若把  $A = LDU$  中的  $L$  与  $D$  结合起来，并且用  $\hat{L}$  来表示，就得到 **Crout 分解**

$$A = (LD)U = \hat{L}U$$

若矩阵  $A$  是对称正定矩阵，那么  $A$  还可以进行 **Cholesky 分解**

## 1.3 国内外研究现状

### 1.3.1 矩阵函数的求法研究现状

常见的有四种方法可以求矩阵函数的值

待定系数法，数项级数求和法，对角形法，若尔当标准型法

后文会一一介绍并给出实例

### 1.3.2 矩阵分解方法研究现状

常见的有四种矩阵分解方法

LU 分解，QR 分解，满秩分解，奇异值分解

后文会一一介绍并给出实例

## 1.4 本论文工作简述

### 1.4.1 本论文对上述问题研究简述

本文主要论述了矩阵函数的求法和矩阵分解的方法。在文章的第一部分，总结了矩阵函数所必须的基础知识，主要包括欧式空间中的线性变换，范数等。文章的第二部分，归纳了矩阵函数的四种计算方法，包括了数项级数求和法、利用相似对角化计算、利用 Jordan 标准型法进行计算、利用待定系数法求解四种计算方法。文章的第三部分，归纳了矩阵分解的若干计算方法，包括了 LU 分解，QR 分解，满秩分解，奇异值分解四种计算方法，并给出了奇异值分解的应用

### 1.4.2 本论文创新点或特点简述

本论文将定义定理与例题结合起来，使得较为抽象的概念能够便于理解和掌握

### 1.4.3 本论文撰写结构简述

每小节的开头会先给出定义，定理作为铺垫，在这些的基础之上会用例题加以解释和说明，便于理解



## 第二章 预备知识

### 2.1 欧式空间与线性变换

#### 2.1.1 线性变换介绍

**定义 1.1** 设 $V$ 是一个非空集合, 它的元素用 $x, y, z$ 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 $x, y \in V$ 时, 有唯一的和 $x + y \in V$ , 且加法运算满足以下性质:

1) 结合律 $x + (y + z) = (x + y) + z$ ;

2) 交换律 $x + y = y + x$ ;

3) 存在零元素 $0$ , 使 $x + 0 = x$ ;

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

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

5) 数因子分配律 $k(x + y) = kx + ky$ ;

6) 分配律 $(k + l)x = kx + lx$ ;

7) 结合律 $k(lx) = (kl)x$ ;

8)  $1x = x$ .

则称 $V$ 为数域 $K$ 上的线性空间或向量空间.

**定义 1.3** 设 $V$ 是数域 $K$ 上的线性空间,  $x_1, x_2, \dots, x_r (r \geq 1)$ 是属于 $V$ 的任意  $r$  个向量, 如果它满足

(1)  $x_1, x_2, \dots, x_r$  线性无关;

(2)  $V$ 中任一向量 $x$ 都是  $x_1, x_2, \dots, x_r$  的线性组合,

则称 $x_1, x_2, \dots, x_r$ 为  $V$  的一个基或基底, 并称 $x_i (i = 1, 2, \dots, r)$ 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

**定义 1.4** 称线性空间 $V^n$ 的一个基  $x_1, x_2, \dots, x_n$  为 $V^n$ 的一个坐标系. 设向量  $x \in V^n$ , 它在该基下的线性表示式为

$$x = \xi_1 x_1 + \xi_2 x_2 + \dots + \xi_n x_n$$

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

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

必须指出, 在不同的坐标系(或基)中, 同一向量的坐标一般是不同的。

**定义 1.10** 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  $x \in V$ ,  $V$ 中都有唯一的向量 $y$ 与之对应, 则称  $T$  是 $V$ 的一个变换或算子, 记为  $Tx = y$ , 称  $y$  为  $x$  在  $T$  下的象, 而  $x$  是  $y$  的原象(或象源).

**定义 1.11** 如果数域  $K$  上的线性空间  $V$  的一个变换  $T$  具有以下性质:

$$T(kx + ly) = k(Tx) + l(Ty)$$

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

根据线性变换的定义 1.11, 有

$$T0 = T(0x) = 0(Tx) = 0, \quad T(-x) = T((-1)x) = (-1)Tx = -Tx$$

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  $x$  的负向量  $-x$  变为  $x$  的象  $Tx$  的负向量  $-Tx$ . 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, 即若

$$k_1x_1 + k_2x_2 + \cdots + k_sx_s = 0$$

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

$$k_1(Tx_1) + k_2(Tx_2) + \cdots + k_s(Tx_s) = T0 = 0$$

有了上面的几条定义, 类似地, 我们可以得到下面的式子

$$\begin{aligned} \because x &= (E_1, \dots, E_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \\ \therefore T(x) &= T \left( (E_1, \dots, E_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right) = (T(E_1), \dots, T(E_n)) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \end{aligned}$$

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

$$T_e x = x \quad (\forall x \in V) \quad (1.2.2)$$

把线性空间  $V$  中的任一向量都变为零向量的变换也是一个线性变换, 称为零变换, 记为  $T_0$ , 于是有

$$T_0 x = 0 \quad (\forall x \in V)$$

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

$$T_1 = T_2$$

对于线性空间上的线性变换, 定义它们的几种运算如下

### 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

$$(T_1 + T_2)x = T_1x + T_2x \quad (\forall x \in V)$$

线性变换  $T$  的负变换  $-T$  定义为

$$(-T)x = -(Tx) \quad (\forall x \in V)$$

容易验证  $T_1 + T_2$  和负变换是线性变换。

### 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

$$(kT)x = k(Tx) \quad (\forall x \in V)$$

容易验证 $kT$ 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

$$(T_1T_2)x = T_1(T_2x) \quad (\forall x \in V)$$

线性变换的乘积中, 满足结合律, 左右分配律, 而交换律一般不成立。

容易验证 $T_1T_2$ 是线性变换。

### 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

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

则称 $S$ 是 $T$ 的逆变换, 记为 $S = T^{-1}$ , 且有

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

### 5. 线性变换的多项式

若 $T$ 是 $V$ 的线性变换, 且存在线性变换 $S$ , 使得

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

则称 $S$ 是 $T$ 的逆变换, 记为 $S = T^{-1}$ , 且有

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

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

$$T^{m+n} = T^mT^n, \quad (T^m)^n = T^{mn} \quad (1.2.8)$$

其中 $m, n \in N_0$ . 当 $T$ 是可逆变换时, 定义 $T$ 的负整数次幂为

$$T^{-n} = (T^{-1})^n \quad (n \in N_0)$$

这样就把指数法则式 (1.2.8) 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

$$f(T) = a_0T^m + a_1T^{m-1} + \dots + a_{m-1}T + a_mT_0$$

也是 $V$ 的一个线性变换, 称其为线性变换 $T$ 的多项式.

下面以一道例题来演示如何求线性变换的多项式

**例** 在矩阵空间 $R^{2 \times 2}$ 中, 给定矩阵

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad T_1(X) = AX, \quad (\forall X \in R^{2 \times 2})$$

求:  $T(X) = f(T_1)(X)$

解:

$$X_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, X_2 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, X_3 = \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}, X_4 = \begin{pmatrix} 0 & -1 \\ 0 & 1 \end{pmatrix}$$

易知 $X_1, X_2, X_3, X_4$ 线性无关

$\therefore$  对  $\forall X \in R^{2 \times 2}$

$$X = (X_1, X_2, X_3, X_4) \alpha$$

其中  $\alpha$  为  $X$  在这组基下的坐标.

$$\text{不妨设 } \alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)^\top \quad X = \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix}$$

$$\text{有 } \begin{cases} \alpha_1 - \alpha_3 = x_1 \\ \alpha_2 - \alpha_4 = x_2 \\ \alpha_1 + \alpha_3 = x_3 \\ \alpha_2 + \alpha_4 = x_4 \end{cases}$$

$$\text{即 } \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$$

$$\text{故 } \alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)^\top = \frac{1}{2} \begin{pmatrix} x_1 + x_3 \\ x_2 + x_4 \\ -x_1 + x_3 \\ -x_2 + x_4 \end{pmatrix}$$

由题  $T_1(X) = AX$

$$T_1 x_1 = AX_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = (x_1, x_2, x_3, x_4) \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$T_1 x_2 = Ax_2 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} = (x_1, x_2, x_3, x_4) \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

$$T_1 x_3 = Ax_3 = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix} = (x_1, x_2, x_3, x_4) \begin{pmatrix} 0 \\ 0 \\ -1 \\ 0 \end{pmatrix}$$

$$T_1 x_4 = Ax_4 = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix} = (x_1, x_2, x_3, x_4) \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \end{pmatrix}$$

$$\therefore T(x_1, x_2, x_3, x_4) = (x_1, x_2, x_3, x_4)A$$

$$\text{其中 } A = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$$

$$T_1(x) = T_1(x_1, x_2, x_3, x_4)\alpha = (x_1, x_2, x_3, x_4)A\alpha$$

$$\therefore T_1^k(x) = (x_1, x_2, x_3, x_4) \text{diag}(1, 1, (-1)^k, (-1)^k) \alpha$$

$$f(T_1)x = (x_1, x_2, x_3, x_4)f(\text{diag}(1, 1, -1, -1))\alpha$$

### 定义 1.14

对于形如

$$\left. \begin{aligned} T x_1 &= a_{11}x_1 + a_{21}x_2 + \cdots + a_{n1}x_n \\ T x_2 &= a_{12}x_1 + a_{22}x_2 + \cdots + a_{n2}x_n \\ &\dots\dots\dots \\ T x_n &= a_{1n}x_1 + a_{2n}x_2 + \cdots + a_{nn}x_n \end{aligned} \right\}$$

采用矩阵乘法形式, 上式可表示为

$$T(x_1, x_2, \dots, x_n) \stackrel{\text{def}}{=} (Tx_1, Tx_2, \dots, Tx_n) = (x_1, x_2, \dots, x_n)A$$

其中

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

矩阵  $A$  的第  $i$  列恰是  $Tx_i$  的坐标 ( $i = 1, 2, \dots, n$ ).

上式中的矩阵  $A$  称为  $T$  在  $V^n$  的基  $x_1, x_2, \dots, x_n$  下的矩阵, 简称  $A$  为  $T$  的矩阵.

有了线性变换在基下的矩阵, 可以得到下面这个定理

**定理 1.10** 设线性变换  $T$  在线性空间  $V^n$  的基  $x_1, x_2, \dots, x_n$  下的矩阵是  $A$ , 向量  $x$  在该基下的坐标是  $\alpha$ , 则  $Tx$  在该基下的坐标是

$$\beta = A\alpha$$

因此, 计算时可以先在一组简单的基下求坐标, 然后通过坐标变换得到在其他基下的坐标, 下面给出坐标变换的步骤.

### 坐标变换的求法

原基  $(e_1, e_2 \cdots e_n)$ , 坐标  $\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$ , 新基  $(E_1, E_2 \cdots E_n)$

$$E_1 = (e_1, e_2 \cdots e_n) \begin{pmatrix} c_{11} \\ \vdots \\ c_{n1} \end{pmatrix}$$

$$E_2 = (e_1, e_2 \cdots e_n) \begin{pmatrix} c_{12} \\ \vdots \\ c_{n2} \end{pmatrix}$$

$\vdots$

$$E_n = (e_1, e_2 \cdots e_n) \begin{pmatrix} c_{1n} \\ \vdots \\ c_{nn} \end{pmatrix}$$

$$\text{故 } (E_1, E_2 \cdots E_n) = (e_1, e_2 \cdots e_n)C$$

由线性代数的知识, 可以证明矩阵  $C$  一定可逆

$$\therefore (e_1, e_2 \cdots e_n) = (E_1, E_2 \cdots E_n)C^{-1}$$

$$\therefore \text{原向量 } x = (e_1, e_2 \cdots e_n) \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = (E_1, E_2 \cdots E_n)C^{-1} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

$$\therefore \text{新坐标 } C^{-1} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

得到新坐标后, 我们就可以求线性变换在新基下的矩阵  $A$

$$T(E_1, \dots, E_n) = (E_1, \dots, E_n)A$$

**定义 1.16** 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  $\lambda_0$ , 存在非零向量  $x \in V^n$ , 使得

$$Tx = \lambda_0 x$$

成立, 则称  $\lambda_0$  为  $T$  的特征值,  $x$  为  $T$  的属于  $\lambda_0$  的特征向量.

下面以一道例题来说明如何求 $T$ 的特征值和特征向量

**例** 设线性变换 $T$ 在基 $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ 下的矩阵是

$$A = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}, \text{ 求 } T \text{ 特征值与特征向量.}$$

解:  $A$  的特征多项式

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

故  $T$  的特征值为:  $\lambda_1 = -1$  (二重),  $\lambda_2 = 5$

把  $\lambda = -1$  代入齐次方程组  $(\lambda I - A)X = 0$ , 得

$$\begin{cases} -2x_1 - 2x_2 - 2x_3 = 0 \\ -2x_1 - 2x_2 - 2x_3 = 0 \\ -2x_1 - 2x_2 - 2x_3 = 0 \end{cases} \quad \text{即 } x_1 + x_2 + x_3 = 0$$

它的一个基础解系为:  $(1, 0, -1), (0, 1, -1)$

因此, 属于  $-1$  的两个线性无关的特征向量为

$$y_1 = x_1 - x_3, \quad y_2 = x_2 - x_3$$

而属于  $-1$  的全部特征向量为

$$k_1 y_1 + k_2 y_2, (k_1, k_2 \in K \text{ 不全为零})$$

把  $\lambda = 5$  代入齐次方程组  $(\lambda I - A)X = 0$ , 得

$$\begin{cases} 4x_1 - 2x_2 - 2x_3 = 0 \\ -2x_1 + 4x_2 - 2x_3 = 0 \\ -2x_1 - 2x_2 + 4x_3 = 0 \end{cases}$$

解得它的一个基础解系为:  $(1, 1, 1)$

因此, 属  $5$  的一个线性无关的特征向量为

$$y_3 = x_1 + x_2 + x_3$$

而属于  $5$  的全部特征向量为

$$k_3 y_3, (k_3 \in K, k_3 \neq 0)$$

而线性代数中有这样一个定理

**定理 1.25**  $n$  阶矩阵  $A$  与对角矩阵相似的充要条件是,  $A$  有  $n$  个线性无关的特征向量, 或  $A$  有完备的特征向量系<sup>②</sup>.

即存在可逆矩阵  $P$  和对角矩阵  $\Lambda$ , 使得  $A = P^{-1} \Lambda P$ , 其中  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ ,  $P^{-1} = (a_1, a_2, \dots, a_n)$ ,  $\lambda_i$  是特征向量  $a_i$  对应的特征值

类似地, 若  $T$  有  $n$  个线性无关的特征向量, 则  $A = P^{-1} \Lambda P$

此时,  $T(E_1, \dots, E_n) = (E_1, \dots, E_n) P^{-1} \Lambda P$

<sup>②</sup> 当  $n$  阶矩阵  $A$  有  $n$  个线性无关的特征向量时, 就称  $A$  有完备的特征向量系

### 2.1.2 若尔当标准型的求解

**定理 1.17** 任意  $n$  阶矩阵与三角矩阵相似.

即, 若  $T$  没有  $n$  个线性无关的特征向量, 则  $A = P^{-1}BP$ , 其中  $B$  是一个上三角矩阵

**定理 1.18** (Hamilton-Cayley)  $n$  阶矩阵  $A$  是其特征多项式的矩阵根 (零点), 即令

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

则有

$$\varphi(A) = A^n + a_1A^{n-1} + \cdots + a_{n-1}A + a_nI = O$$

通过  $\varphi(\lambda)$  很多时候可以简化计算

**定义 1.19** 首项系数是 1 (简称首 1), 次数最小, 且以矩阵  $A$  为根的  $\lambda$  的多项式, 称为  $A$  的最小多项式, 常用  $m(\lambda)$  表示.

显然,  $m(\lambda)$  的次数不大于特征多项式  $\varphi(\lambda)$  的次数

**定理 1.21** 设  $n$  阶矩阵  $A$  的特征多项式为  $\varphi(\lambda)$ , 特征矩阵  $\lambda I - A$  的全体  $n-1$  阶子式的最大公因式为  $d(\lambda)$ , 则  $A$  的最小多项式为

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

求最小多项式分为以下三步

$$1. \text{定义 } \lambda \text{ 矩阵, 令 } A(\lambda) = \lambda I - A = \begin{bmatrix} a_{11}(\lambda) & a_{12}(\lambda) & \cdots & a_{1n}(\lambda) \\ a_{21}(\lambda) & a_{22}(\lambda) & \cdots & a_{2n}(\lambda) \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}(\lambda) & a_{n2}(\lambda) & \cdots & a_{nn}(\lambda) \end{bmatrix}$$

2. 将  $\lambda$  矩阵  $A(\lambda)$  写成标准型, 并写出不变因子  $d_k(\lambda)$

多项式矩阵  $A(\lambda)$  的标准形, 是指使用矩阵的初等变换将  $A(\lambda)$  化为如下形式的多项式矩阵:

其中

$$A(\lambda) \rightarrow \begin{bmatrix} d_1(\lambda) & & & & & \\ & d_2(\lambda) & & & & \\ & & \ddots & & & \\ & & & d_s(\lambda) & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix} \quad (1.2.37)$$

$d_1(\lambda) | d_2(\lambda), d_2(\lambda) | d_3(\lambda), \dots, d_{s-1}(\lambda) | d_s(\lambda), s \leq n$ , 且  $d_i(\lambda) (i = 1, 2, \dots, s)$  是首 1 多项式 (前面的几个  $d_i(\lambda)$  可能是 1).

可以证明, 一个多项式矩阵  $A(\lambda)$  的标准形式 (1.2.37) 的对角线上的非零元素  $d_i(\lambda) (i = 1, 2, \dots, s)$  不随矩阵的初等变换而改变. 因此, 通常称  $d_i(\lambda) (i = 1, 2, \dots, s)$  为  $A(\lambda)$  的不变因子或不变因式.

3.最后一个不变因子就是A的最小多项式,  $m(\lambda) = d_s(\lambda)$

下面以一道例题来演示求最小不变因子

**例** 求矩阵

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

的最小多项式

解

求  $\lambda I - A$  的初等因子组. 由于

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

不变因子为  $d_1(\lambda) = 1, d_2(\lambda) = 1, d_3(\lambda) = (\lambda-2)(\lambda-1)^2$  最后一个不变因子就是  $A$  的最小多项式:  $m(\lambda) = d_3(\lambda) = (\lambda-2)(\lambda-1)^2$

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

下面给出 Jordan 标准型的定义

**定义 1.21** 由式  $J = \begin{bmatrix} J_1(\lambda_1) & & \\ & J_2(\lambda_2) & \\ & & \ddots \\ & & & J_s(\lambda_s) \end{bmatrix}$  给出的矩阵  $J$  称为矩阵  $A$  的

Jordan 标准形.

其中

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

$J_i(\lambda_i)$  称为因式  $(\lambda - \lambda_i)^{m_i}$  对应的 Jordan 块

**定理 1.29** 设  $A$  是  $n$  阶复矩阵, 且其特征多项式的某种分解式是

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

则存在  $n$  阶复可逆矩阵  $P$ , 使

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

有了上面的铺垫, 那么 Jordan 标准型中的  $J$  就可以求解了

下面以一道例题来演示求矩阵的 Jordan 标准型中的  $J$

**例** 求矩阵



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

的 *Jordan* 标准形.

解

求  $\lambda I - A$  的初等因子组. 由于

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

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

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

在求出了  $J$  后, 下一步我们需要考虑  $P$  该如何去求

对于  $P^{-1}AP = J$ , 其中  $P$  的求法如下

假如

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

其中  $P = (x_1, x_2, x_3)$ , 于是有

$$A(x_1, x_2, x_3) = (x_1, x_2, x_3) \begin{bmatrix} \lambda_1 & 0 & 0 \\ & \lambda_2 & 1 \\ & & \lambda_2 \end{bmatrix}$$

即

$$(Ax_1, Ax_2, Ax_3) = (\lambda_1 x_1, \lambda_2 x_2, x_2 + \lambda_2 x_3)$$

由此可得

$$\left. \begin{aligned} (\lambda_1 I - A)x_1 &= 0 \\ (\lambda_2 I - A)x_2 &= 0 \\ (\lambda_2 I - A)x_3 &= -x_2 \end{aligned} \right\} \quad (1.2.39)$$

从而  $x_1, x_2$  依次是  $A$  的属于  $\lambda_1, \lambda_2$  的特征向量.  $x_3$  是式(1.2.39) 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  $P$ .

又如

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & 1 & \\ & \lambda_1 & 1 \\ & & \lambda_1 \end{bmatrix}, \quad P = (x_1, x_2, x_3)$$

则有

$$A(x_1, x_2, x_3) = (x_1, x_2, x_3) \begin{bmatrix} \lambda_1 & 1 & \\ & \lambda_1 & 1 \\ & & \lambda_1 \end{bmatrix}$$

即

$$(Ax_1, Ax_2, Ax_3) = (\lambda_1 x_1, x_1 + \lambda_1 x_2, x_2 + \lambda_1 x_3)$$

由此可得

$$\begin{cases} (\lambda_1 I - A)x_1 = 0 \\ (\lambda_1 I - A)x_2 = -x_1 \\ (\lambda_1 I - A)x_3 = -x_2 \end{cases} \quad (1.2.40)$$

从而  $x_1$  是  $A$  的属于  $\lambda_1$  的特征向量  $x_2, x_3$  是式 (1.2.40) 后两个非齐次线性方程组的解向量. 这样, 又得到了所需要的矩阵  $P$ .

在一般情况下, 如果  $\lambda_1$  是  $A$  的  $k$  重特征值, 则  $x_1, x_2, \dots, x_k$  可由解下面各方程组

$$\begin{cases} (\lambda_1 I - A)x_1 = 0 \\ (\lambda_1 I - A)x_i = -x_{i-1} \quad (i = 2, 3, \dots, k) \end{cases}$$

而获得. 这样得到的  $x_1, x_2, \dots, x_k$  线性无关 (其证明过程冗长, 从略), 于是就得到  $P$ . 称  $x_2, x_3, \dots, x_k$  为  $A$  的属于  $\lambda_1$  的广义特征向量.

以上一道例题为例演示  $P$  的求法

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

设  $P = (x_1, x_2, x_3)$ ,  $A = PJP^{-1}$ , 有  $AP = PJ$

即  $A(x_1, x_2, x_3) = (x_1, x_2, x_3)J$

$$\begin{cases} Ax_1 = 2x_1 \\ Ax_2 = x_2 \\ Ax_3 = x_2 + x_3 \end{cases}$$

$$\text{由此, 可得} \begin{cases} (2I - A)x_1 = 0 \\ (I - A)x_2 = 0 \\ (I - A)x_3 = -x_2 \end{cases}$$

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

$$\text{故 } P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 2 & 1 \\ 1 & -1 & -1 \end{pmatrix}$$

$$\text{可求 } P^{-1} = \begin{pmatrix} -1 & 1 & 1 \\ 1 & 0 & 0 \\ -2 & 1 & 0 \end{pmatrix}$$

### 2.1.3 欧式空间中线性变换的求法

**定义 1.22** 设  $V$  是实数域  $R$  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:  $(kx, y) = k(x, y) (\forall k \in R)$  ;  
 (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**

**定义 1.29** 如果实方阵  $Q$  满足  $Q^T Q = I$ , 则称  $Q$  为正交矩阵.

**定理 1.37** 欧氏空间的线性变换是正交变换的**充要条件**是, 它对于标准正交基的矩阵是正交矩阵.

**定义 1.23** 在欧氏空间  $V$  中, 非负实数  $\sqrt{(x, x)}$  称为  $V$  中向量  $x$  的长度 (或模, 范数), 记为  $|x|$  (或  $\|x\|_2$ ), 即

$$|x| = \sqrt{(x, x)}$$

由此定义可知, 非零向量的长度是正数, 只有零向量的长度才是零. 这样定义的长度符合性质:

$$\begin{aligned} |kx| &= |k| |x| \\ |x+y| &\leq |x| + |y| \end{aligned}$$

**定义 1.25** 如果对于欧氏空间中的两个向量  $x$  与  $y$ , 有  $(x, y) = 0$ , 则称  $x$  与  $y$  正交或垂直, 记为  $x \perp y$ .

**定义 1.24** 非零向量  $x$  与  $y$  的夹角  $\langle x, y \rangle$  规定为

$$\langle x, y \rangle = \arccos \frac{(x, y)}{|x||y|} \quad (0 \leq \langle x, y \rangle \leq \pi)$$

**定义 1.26** 如果欧氏空间中一组非零向量两两正交, 则称为正交向量组

**定理 1.32** 在欧氏空间中, 如果  $x_1, x_2, \dots, x_m$  是正交向量组, 则它们必线性无关.

**定义 1.27** 在欧氏空间  $V^n$  中, 由  $n$  个非零向量组成的正交向量组称为  $V^n$  的正交基; 由单位向量组成的正交基称为**标准正交基**或**法正交基**.

把一个正交基进行单位化, 就得到一个标准正交基

**定理 1.32** 在欧氏空间中, 如果  $x_1, x_2, \dots, x_m$  是正交向量组, 则它们必线性无关.

**定理 1.33** 对于欧氏空间  $V^n$  的任一基  $x_1, x_2, \dots, x_n$ , 都可找到一个标准正交基  $y_1, y_2, \dots, y_n$ . 换言之, 任一非零欧氏空间都有正交基和标准正交基.

下面给出 Schmidt 正交化的方法(或过程)

为此取  $y_1' = x_1$ , 作为所求正交基中的第一个向量. 再令

$$y_2' = x_2 + ky_1'$$

由正交条件  $(y_2', y_1') = 0$  来决定待定常数  $k$ . 由

$$(x_2 + ky_1', y_1') = (x_2, y_1') + k(y_1', y_1') = 0$$

得

$$k = -\frac{(x_2, y_1')}{(y_1', y_1')}$$

这样就得到两个正交的向量  $y_1', y_2'$ , 且  $y_2' \neq 0$ . 又令

$$y_3' = x_3 + k_2 y_2' + k_1 y_1'$$

再由正交条件  $(y_3', y_2') = 0$  及  $(y_3', y_1') = 0$  来决定出  $k_1$  和  $k_2$  为

$$k_2 = -\frac{(x_3, y_2')}{(y_2', y_2')}, \quad k_1 = -\frac{(x_3, y_1')}{(y_1', y_1')}$$

到此, 已经做出三个两两正交的向量  $y_1', y_2', y_3'$ , 且  $y_3' \neq 0$ . 继续这样进行下去, 设已做出  $m$  个两两正交且不为零的向量  $y_1', y_2', \dots, y_m'$ , 为求出第  $m+1$  个与之正交的向量, 令

$$y_{m+1}' = x_{m+1} + l_m y_m' + l_{m-1} y_{m-1}' + \dots + l_2 y_2' + l_1 y_1'$$

使用  $m$  个正交条件

$$(y_{m+1}', y_i') = 0 \quad (i = 1, 2, \dots, m)$$

来决定  $l_m, l_{m-1}, \dots, l_2, l_1$ . 根据  $y_1', y_2', \dots, y_m'$  为两两正交的假定, 可得

$$(x_{m+1}, y_i') + l_i (y_i', y_i') = 0$$

故

$$l_i = -\frac{(x_{m+1}, y_i')}{(y_i', y_i')} \quad (i = 1, 2, \dots, m) \quad (1.1.38)$$

于是  $y_{m+1}'$  就被确定出来了.

采用上述 Schmidt 正交化方法, 可由已知基构造出  $n$  个两两正交的非零向量  $y_1', y_2', \dots, y_n'$ . 根据定理 1.32, 知  $y_1', y_2', \dots, y_n'$  线性无关, 从而它们形成  $V^n$  的一个正交基. 再以  $|y_i'|$  除  $y_i'$  ( $i = 1, 2, \dots, n$ ), 就得到定理所要求的标准正交基

$$y_i = \frac{1}{|y_i'|} y_i' \quad (i = 1, 2, \dots, n)$$

上述是由基  $x_1, x_2, \dots, x_n$  构造标准正交基  $y_1, y_2, \dots, y_n$  的过程, 有时也称为把基  $x_1, x_2, \dots, x_n$  正交单位化或正交规范化.

下面以一道例题来演示正交单位化的过程

**例 1.33** 试把向量组

$x_1 = (1, 1, 0, 0), x_2 = (1, 0, 1, 0), x_3 = (-1, 0, 0, 1), x_4 = (1, -1, -1, 1)$  正交单位化.

解 先把它们正交化. 使用式 (1.3.18), 可得

$$y_1' = x_1 = (1, 1, 0, 0)$$

$$y_2' = x_2 - \frac{(x_2, y_1')}{(y_1', y_1')} y_1' = \left(\frac{1}{2}, -\frac{1}{2}, 1, 0\right)$$

$$y_3' = x_3 - \frac{(x_3, y_1')}{(y_1', y_1')} y_1' - \frac{(x_3, y_2')}{(y_2', y_2')} y_2' = \left(-\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 1\right)$$

$$y_4' = x_4 - \frac{(x_4, y_1')}{(y_1', y_1')} y_1' - \frac{(x_4, y_2')}{(y_2', y_2')} y_2' - \frac{(x_4, y_3')}{(y_3', y_3')} y_3' = (1, -1, -1, 1)$$

再单位化, 则有

$$y_1 = \frac{1}{|y_1'|} y_1' = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0\right)$$

$$y_2 = \frac{1}{|y_2'|} y_2' = \left(\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, 0\right)$$

$$y_3 = \frac{1}{|y_3'|} y_3' = \left(-\frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{3}{\sqrt{12}}\right)$$

$$y_4 = \frac{1}{|y_4'|} y_4' = \left(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right)$$

至此, 欧式空间中线性变换的求法已经构建完成, 其步骤如下

设  $V$  是欧式空间,  $T$  是  $V$  上的一个线性变换, 求  $z = (T^k)(x)$ , 其中  $x \in V$

0、任意找一组基, 利用 Schmidt 正交化方法得到  $V$  的一组标准正交基  $e_1, \dots, e_n$ ,

$x = k_1 e_1 + \dots + k_n e_n$ , 其中  $k_i = (x, e_i)$

0.1、求  $T$  在基  $e_1, \dots, e_n$  下的矩阵  $A_0 \Rightarrow T(e_1, \dots, e_n) = (e_1, \dots, e_n)A_0$

0.2、其中  $A_0 = PJP^{-1}$ ,  $J$  是 Jordan 标准型  $\Rightarrow T(e_1, \dots, e_n) = (e_1, \dots, e_n)PJP^{-1}$

0.3、 $T(e_1, \dots, e_n)P = (e_1, \dots, e_n)PJ$

1、得到一组新的基  $(E_1, \dots, E_n) = (e_1, \dots, e_n)P$ ,

2、通过坐标变换得到  $x = (E_1, \dots, E_n)P^{-1} \begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix} = (E_1, \dots, E_n) \begin{pmatrix} l_1 \\ \vdots \\ l_n \end{pmatrix}$

3、 $T$  在新基下的矩阵:  $T(E_1, \dots, E_n) = (E_1, \dots, E_n)J$

4、 $T(x) = (E_1, \dots, E_n)J \begin{pmatrix} l_1 \\ \vdots \\ l_n \end{pmatrix} \Rightarrow (T^k)(x) = (E_1, \dots, E_n)J^k \begin{pmatrix} l_1 \\ \vdots \\ l_n \end{pmatrix}$

以一道例题来演示

**例** 设矩阵空间  $R^{2 \times 2}$  的子空间为

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

$V$  中的线性变换为  $T(X) = X + 2X^T$

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

(2) 求  $(T^k)(X), \forall X \in V$

令  $x_{11} = -x_{12} - x_{21}$

$$\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} = x_{11} \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} + x_{21} \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix} + x_{22} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = x_1 X_1 + x_2 X_2 + x_3 X_3$$

且  $X_1, X_2, X_3$  线性无关

利用 Schmidt 正交化, 有

$$Y'_1 = X_1 = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$$

$$Y'_2 = X_2 - \frac{(X_2, Y'_1)}{(Y'_1, Y'_1)} Y'_1 = \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} \\ 1 & 0 \end{pmatrix}$$

$$Y'_3 = X_3 - \frac{(X_3, Y'_2)}{(Y'_2, Y'_2)} Y'_2 - \frac{(X_3, Y'_1)}{(Y'_1, Y'_1)} Y'_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

$$e_1 = \frac{1}{|Y'_1|} Y'_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$$

$$e_2 = \frac{1}{|Y'_2|} Y'_2 = \frac{\sqrt{2}}{\sqrt{3}} \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} \\ 1 & 0 \end{pmatrix}$$

$$e_3 = \frac{1}{|Y'_3|} Y'_3 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\text{由 } x = \begin{pmatrix} 4 & -4 \\ 0 & -3 \end{pmatrix} = (e_1, e_2, e_3) \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix}$$

$$k_1 = (x, e_1) = -4\sqrt{2},$$

得  $k_2 = (x, e_2) = 0,$

$$k_3 = (x, e_3) = -3$$

$$Te_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} -3 & 1 \\ 2 & 0 \end{pmatrix},$$

$$Te_2 = \frac{\sqrt{2}}{\sqrt{3}} \begin{pmatrix} -\frac{3}{2} & -\frac{3}{2} \\ 0 & 0 \end{pmatrix},$$

$$Te_3 = \begin{pmatrix} 0 & 0 \\ 0 & 3 \end{pmatrix}$$

$$\text{由 } Te_1 = (e_1, e_2, e_3) \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix}$$

$$k_1 = (Te_1, e_1) = 2$$

得  $k_2 = (Te_1, e_2) = \sqrt{3}$

$$k_3 = (Te_1, e_3) = 0$$

同理可得  $Te_2Te_3$  对应的  $k$

$$\therefore T(e_1, e_2, e_3) = (e_1, e_2, e_3)A_0$$

$$A_0 = \begin{pmatrix} 2 & \sqrt{3} & 0 \\ \sqrt{3} & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

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

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

初等因子:  $\lambda - 3; \lambda + 1; \lambda - 3$

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

$Jordan$  块:  $J_1(\lambda_1) = 3, J_2(\lambda_2) = -1, J_3(\lambda_3) = 3$

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

由  $A_0 = PJP^{-1}$ , 有  $PJ = A_0J$

设  $P = (x_1, x_2, x_3)$ , 则有

$$(3x_1, -x_2, 3x_3) = (A_0x_1, A_0x_2, A_0x_3)$$

$$\begin{cases} 3x_1 = A_0x_1 \\ -x_2 = A_0x_2 \\ 3x_3 = A_0x_3 \end{cases} \Rightarrow \begin{cases} x_1 = (\sqrt{3}, 1, 0)^T \\ x_2 = (-1, \sqrt{3}, 0)^T \\ x_3 = (0, 0, 1)^T \end{cases}$$

$$\therefore P(x_1, x_2, x_3) = \begin{pmatrix} \sqrt{3} & -1 & 0 \\ 1 & \sqrt{3} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$P^{-1} = \begin{pmatrix} \frac{\sqrt{3}}{4} & \frac{1}{4} & 0 \\ -\frac{1}{4} & \frac{\sqrt{3}}{4} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\therefore T(e_1, e_2, e_3)P = (e_1, e_2, e_3)PJ$$

$$\therefore \text{新基}(E_1, E_2, E_3) = (e_1, e_2, e_3)P$$

得

$$E_1 = (e_1, e_2, e_3) \begin{pmatrix} \sqrt{3} \\ 1 \\ 0 \end{pmatrix} = \frac{2}{\sqrt{6}} \begin{pmatrix} -2 & 1 \\ 1 & 0 \end{pmatrix}$$

$$E_2 = (e_1, e_2, e_3) \begin{pmatrix} -1 \\ \sqrt{3} \\ 0 \end{pmatrix} = \sqrt{2} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$$E_3 = (e_1, e_2, e_3) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

$$T(E_1, E_2, E_3) = (E_1, E_2, E_3)J$$

$$\text{通过坐标变换, 有 } x = (E_1, E_2, E_3)P^{-1} \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} = (E_1, E_2, E_3) \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix}$$

$$\therefore x = \begin{pmatrix} 4 & -4 \\ 0 & -3 \end{pmatrix} = (e_1, e_2, e_3) \begin{pmatrix} -4\sqrt{2} \\ 0 \\ -3 \end{pmatrix}$$

$$\therefore (E_1, E_2, E_3)P^{-1} \begin{pmatrix} -4\sqrt{2} \\ 0 \\ -3 \end{pmatrix} = (E_1, E_2, E_3) \begin{pmatrix} -\sqrt{6} \\ \sqrt{2} \\ -3 \end{pmatrix}$$

$$\therefore T(x) = (E_1, E_2, E_3)J \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} \Rightarrow (T^k)(x) = (E_1, E_2, E_3)J^k \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix}$$

$$\therefore (T^3)(x) = (E_1, E_2, E_3) \begin{pmatrix} 27 & & \\ & -1 & \\ & & 27 \end{pmatrix} \begin{pmatrix} -\sqrt{6} \\ \sqrt{2} \\ -3 \end{pmatrix} = \begin{pmatrix} 108 & -52 \\ -56 & -81 \end{pmatrix}$$

$$(T^k)(x) = (E_1, E_2, E_3) \begin{pmatrix} 3^k & & \\ & (-1)^k & \\ & & 3^k \end{pmatrix} \begin{pmatrix} \frac{\sqrt{3}}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{\sqrt{3}}{4} & 0 \\ -\frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} (x, e_1) \\ (x, e_2) \\ (x, e_3) \end{pmatrix}$$

## 2.2 向量范数与矩阵范数

### 2.2.1 向量范数介绍

**定义 2.1** 如果  $V$  是数域  $K$  上的线性空间, 对任意的  $x \in V$ , 定义一个实值函数  $\|x\|$ , 它满足以下三个条件:

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

则称  $\|x\|$  为  $V$  上向量  $x$  的范数, 简称**向量范数**.

下面是几种常见的范数

$\|x\| = \sum_{i=1}^n |\xi_i|$  为向量的 1-范数, 记为  $\|x\|_1$ , 即

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



$\|x\| = \sqrt{(x, x)}$  为向量的 2-范数, 记为  $\|x\|_2$ , 即

$$\|x\|_2 = \sqrt{(x, x)}$$

$\|x\| = \max_i |\xi_i|$  为向量的  $\infty$ -范数, 记为  $\|x\|_\infty$ , 即

$$\|x\|_\infty = \max_i |\xi_i|$$

事实上, 以上几种范数都可以被归类为下面这个范数

线性空间  $C^n$ , 设  $x = (\xi_1, \xi_2, \dots, \xi_n)^T \in C^n$

$\|x\| = (\sum_{i=1}^n |\xi_i|^p)^{1/p}$  为向量  $x$  的  $p$ -范数或  $l_p$  范数, 记为  $\|x\|_p$ , 即

$$\|x\|_p = \left( \sum_{i=1}^n |\xi_i|^p \right)^{1/p}$$

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

**例 2.6** 给定线性空间  $V^n$  的基  $x_1, x_2, \dots, x_n$ , 设  $x \in V^n$  在该基下的坐标向量为  $\tilde{x} = (\xi_1, \xi_2, \dots, \xi_n)^T$ , 那么

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

满足范数定义三个条件. 因此, 它是  $V^n$  上的范数, 也称为  $x$  的  $p$ -范数.

**定理 2.1** 设  $\|x\|_\alpha$  和  $\|x\|_\beta$  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  $x$  无关的正常数  $c_1$  和  $c_2$ , 使满足

$$c_1 \|x\|_\beta \leq \|x\|_\alpha \leq c_2 \|x\|_\beta \quad (\forall x \in V)$$

**定义 2.2** 满足上面不等式的两种范数是等价的

因此, 有限维线性空间上的不同范数是等价的。

**定理 2.2**  $C^n$  中的向量序列

$$x^{(k)} = (\xi_1^{(k)}, \xi_2^{(k)}, \dots, \xi_n^{(k)}) \quad (k = 1, 2, 3, \dots)$$

收敛到向量  $x = (\xi_1, \xi_2, \dots, \xi_n)$  的充要条件是对任何一种向量范数  $\|\cdot\|$ , 数列  $\{x^{(k)} - x\}$  收敛于零.

这就是说, 尽管不同向量范数可能具有不同的大小, 但是在各种向量范数下考虑向量序列收敛性问题时, 各种范数表现出了一致性, 我们可以根据实际需要选择合适的向量范数。

### 2.2.2 矩阵范数介绍

**定义 2.3** 设  $A \in C^{m \times n}$ , 定义一个实值函数  $\|A\|$ , 它满足以下三个条件:

- (1) 非负性: 当  $A \neq O$  时,  $\|A\| > 0$ ; 当  $A = O$  时,  $\|A\| = 0$ ;
- (2) 齐次性:  $\|\alpha A\| = |\alpha| \|A\| (\alpha \in C)$ ;

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

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

(4) 相容性:

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

则称  $\|A\|$  为  $A$  的矩阵范数.

**定义 2.4** 对于  $C^{m \times n}$  上的矩阵范数  $\|\cdot\|_M$  和  $C^m$  与  $C^n$  上的同类向量范数  $\|\cdot\|_V$ , 如果

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

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

**定理 2.4** 已知  $C^m$  和  $C^n$  上的同类向量范数  $\|\cdot\|$ , 设  $A \in C^{m \times n}$ , 则函数

$$\|A\| = \max_{\|x\|=1} \|Ax\| \quad (2.2.5)$$

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

称由式 (2.2.5) 给出的矩阵范数为由向量范数导出的矩阵范数, 简称为从属范数. 对于  $C^{n \times n}$  上的任何一种从属范数, 有

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

下面是几种常用的矩阵范数

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

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

$$(3) \|A\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|.$$

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

### 2.2.3 矩阵可逆性条件、谱半径和条件数介绍

**定理 2.6** 设  $A \in C^{n \times n}$ , 且对  $C^{n \times n}$  上的某种矩阵范数  $\|\cdot\|$ , 有  $\|A\| < 1$ , 则矩阵  $I - A$  可逆, 且有

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

**定理 2.7** 设  $A \in C^{n \times n}$ , 且对  $C^{n \times n}$  上的某种矩阵范数  $\|\cdot\|$ , 有  $\|A\| < 1$ , 则

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

**定理 2.8** 设  $A \in C^{n \times n}$  可逆,  $B \in C^{n \times n}$ , 且对  $C^{n \times n}$  上的某种矩阵范数

$\|\cdot\|$  , 有  $\|A^{-1}B\| < 1$ , 则有以下结论:

(1)  $A + B$  可逆;

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

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

在定理 2.8 中, 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

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

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, 往往这样的矩阵难以求逆.

**定义 2.5** 设  $A \in \mathbb{C}^{n \times n}$  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 2.9** 设  $A \in \mathbb{C}^{n \times n}$  , 则对  $\mathbb{C}^{n \times n}$  上任何一种矩阵范数  $\|\cdot\|$  , 都有

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

**定理 2.10** 设  $A \in \mathbb{C}^{n \times n}$  , 对任意的正数  $\varepsilon$  , 存在某种矩阵范数  $\|\cdot\|_M$  , 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

首先给出矩阵序列收敛和发散的定義。

**定义 3.1** 设有矩阵序列  $\{A^{(k)}\}$  , 其中  $A^{(k)} = (a_{ij}^{(k)})_{m \times n} \in \mathbb{C}^{m \times n}$  , 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

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

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.

**性质 1** 设  $A^{(k)} \rightarrow A_{m \times n}$  ,  $B^{(k)} \rightarrow B_{m \times n}$  , 则

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

**性质 2** 设  $A^{(k)} \rightarrow A_{m \times n}$  ,  $B^{(k)} \rightarrow B_{n \times 1}$  , 则

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

**定理 3.1** 设  $A^{(k)} \in \mathbb{C}^{m \times n}$  , 则

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

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

**定义 3.2** 矩阵序列  $\{A^{(k)}\}$  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 3.3** 设  $A$  为方阵, 且  $A^k \rightarrow O (k \rightarrow \infty)$ , 则称  $A$  为收敛矩阵.

**定理 3.2**  $A$  为收敛矩阵的充要条件是  $\rho(A) < 1$

**定理 3.3**  $A$  为收敛矩阵的充分条件是只要有一种矩阵范数  $\|\cdot\|$ , 使得  $\|A\| < 1$  这个定理表明, 在判断矩阵是否是收敛矩阵时, 我们可以去找它有没有小于 1 的矩阵范数

当然, 因为通常情况下我们不一定能恰好找到满足条件的矩阵范数, 所以更多的时候, 我们还是用计算谱半径的方法来判断

### 2.3.2 矩阵级数介绍

**定义 3.4** 把定义 3.1 中的矩阵序列所形成的无穷和  $A^{(0)} + A^{(1)} + A^{(2)} + \dots + A^{(k)} + \dots$  称为矩阵级数, 记为  $\sum_{k=0}^{\infty} A^{(k)}$  , 则有

$$\sum_{k=0}^{\infty} A^{(k)} = A^{(0)} + A^{(1)} + A^{(2)} + \dots + A^{(k)} + \dots \quad (3.2.1)$$

**定义 3.5** 记  $S^{(N)} = \sum_{k=0}^N A^{(k)}$  , 称其为矩阵级数式 (3.2.1) 的部分和. 如果矩阵序列  $\{S^{(N)}\}$  收敛, 且有极限  $S$  , 则有

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

那么就称矩阵级数式 (3.2.1) 收敛, 而且有和  $S$  , 记为

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

不收敛的矩阵级数称为是发散的.

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

$$\sum_{k=0}^{\infty} a_{ij}^{(k)} = s_{ij} \quad (i = 1, 2, \dots, m; j = 1, 2, \dots, n) \quad (3.2.4)$$

**定义 3.6** 如果式 (3.2.4) 中左端  $m \times n$  个数项级数都是绝对收敛的, 则称矩阵级数式 (3.2.1) 是绝对收敛的.

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

**定理 3.4** 方阵  $A$  的幂级数 (Neumann 级数)

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

矩阵级数收敛的充要条件是  $A$  为收敛矩阵, 并且在收敛时, 其和为  $(I - A)^{-1}$ .

**定理 3.5** 设方阵  $A$  对某一矩阵范数  $\|\cdot\|$  有  $\|A\| < 1$ , 则对任何非负整数  $N$ , 以  $(I - A)^{-1}$  为部分和  $I + A + A^2 + \cdots + A^N$  的近似矩阵时, 其误差为

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

**定理 3.6** 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  $A$  满足  $\rho(A) < r$ , 则矩阵幂级数

$$\sum_{k=0}^{\infty} c_k A^k \quad (3.2.11)$$

是绝对收敛的; 如果  $\rho(A) > r$ , 则矩阵幂级数式 (3.2.11) 是发散的.

下面以一道例题来演示如何判断矩阵幂级数的敛散性

**例** 讨论下列幂级数的敛散性

$$\sum_{k=1}^{\infty} \frac{1}{k^2} \begin{bmatrix} 1 & 7 \\ -1 & -3 \end{bmatrix}^k;$$

$$\begin{pmatrix} 1 & 7 \\ -1 & -3 \end{pmatrix} = \begin{pmatrix} -2 + \sqrt{3}i & -2 - \sqrt{3}i \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -1 - \sqrt{3}i & 0 \\ 0 & -1 + \sqrt{3}i \end{pmatrix} \begin{pmatrix} -\frac{\sqrt{3}}{6}i & \frac{1}{2} - \frac{\sqrt{3}}{3}i \\ \frac{\sqrt{3}}{6}i & \frac{1}{2} + \frac{\sqrt{3}}{3}i \end{pmatrix}$$

$$\rho(A) = 2 > R = \frac{1}{l} = 1, l = \lim_{k \rightarrow +\infty} \frac{k^2}{(k+1)} = 1 \text{ 发散}$$

下面以一道例题来演示如何求矩阵幂级数的和

**例** 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

$$\begin{aligned}
\sum_{k=0}^{+\infty} \frac{1}{2^k} A^k &= P \begin{pmatrix} \sum_{k=0}^{+\infty} \frac{1}{2^k} & \sum_{k=0}^{+\infty} \frac{k}{2^k} \\ 0 & \sum_{k=0}^{+\infty} \frac{1}{2^k} \end{pmatrix} P^{-1} = P \begin{pmatrix} 2 & 2 \\ 0 & 2 \end{pmatrix} P^{-1} \\
\text{令 } \sum_{k=0}^{+\infty} \frac{k}{2^k} &= M = \frac{1}{2^1} + \frac{2}{2^2} + \frac{3}{2^3} + \cdots + \frac{k}{2^k} + \cdots \\
2M &= \frac{1}{2^{1-1}} + \frac{2}{2^{2-1}} + \frac{3}{2^{3-1}} + \cdots + \frac{k}{2^{k-1}} + \cdots \\
2M &= 1 + \frac{1+1}{2^1} + \frac{2+1}{2^2} + \frac{3+1}{2^3} + \cdots + \frac{k+1}{2^k} + \cdots \\
2M - M &= 1 + \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots + \frac{1}{2^k} + \cdots = \sum_{k=0}^{+\infty} \frac{1}{2^k} = 2 \\
\sum_{k=0}^{+\infty} \frac{1}{2^k} A^k &= P \begin{pmatrix} \sum_{k=0}^{+\infty} \frac{1}{2^k} & \sum_{k=0}^{+\infty} \frac{k}{2^k} \\ 0 & \sum_{k=0}^{+\infty} \frac{1}{2^k} \end{pmatrix} P^{-1} = P \begin{pmatrix} 2 & 2 \\ 0 & 2 \end{pmatrix} P^{-1} \\
&= \begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 2 & 2 \\ 0 & 2 \end{pmatrix} \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} 4 & -1 \\ 4 & 0 \end{pmatrix}
\end{aligned}$$

### 2.3.3 矩阵函数介绍

**定义 3.7** 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  $A$  的谱半径  $\rho(A) < r$  时, 把收敛的矩阵幂级数  $\sum_{k=0}^{\infty} c_k A^k$  的和称为矩阵函数, 记为  $f(A)$ , 即

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

下面是一组常见的矩阵函数等式

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

**定理 3.7** 如果  $AB = BA$ , 则  $e^A e^B = e^B e^A = e^{A+B}$ .

在本论文的第三章会给出矩阵函数的求法

## 2.4 函数矩阵对矩阵的导数

**定义 3.9** 如果函数矩阵  $A(t) = (a_{ij}(t))_{m \times n}$  的每一个元素  $a_{ij}(t)$  是变量  $t$  的可导函数, 则称  $A(t)$  可导, 其导数 (微商) 定义为

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

由这个定义, 我们可以得到下面这个定理

**定理 3.8** 设  $A(t), B(t)$  是能够进行下面运算的两个可导的函数矩阵, 则有

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

这里,  $a = a(t)$  为  $t$  的可导函数.

**定理 3.9** 设  $n$  阶矩阵  $A$  与  $t$  无关, 则有

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

**定义 3.10** 如果函数矩阵  $A(t)$  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  $A(t)$  在  $[t_0, t_1]$  上的积分为

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

有了上面函数矩阵的一些基本概念和性质, 我们接下来介绍对矩阵的导数

### 1. 函数对矩阵的导数

**定义 3.11** 设  $X = (\xi_{ij})_{m \times n}$ ,  $m, n$  元函数  $f(X) = f(\xi_{11}, \xi_{12}, \dots, \xi_{1n}, \xi_{21}, \dots, \xi_{mn})$ , 定义  $f(X)$  对矩阵  $X$  的导数为

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

以一道例题为例, 来展示其形状

**例** 设  $x = (\xi_1, \xi_2, \dots, \xi_n)^T, A = (a_{ij})_{n \times n}$ ,  $n$  元函数  $f(x) = x^T A x$ , 求  $\frac{df}{dx}$

解

$$f(x) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} \xi_i \xi_j = \xi_1 \sum_{j=1}^n a_{1j} \xi_j + \dots + \xi_k \sum_{j=1}^n a_{kj} \xi_j + \dots + \xi_n \sum_{j=1}^n a_{nj} \xi_j$$

且有

$$\begin{aligned}\frac{\partial f}{\partial \xi_k} &= \xi_1 a_{1k} + \dots + \xi_{k-1} a_{k-1,k} + \left( \sum_{j=1}^n a_{kj} \xi_j + \xi_k a_{kk} \right) + \\ &\quad \xi_{k+1} a_{k+1,k} + \dots + \xi_n a_{nk} = \sum_{j=1}^n a_{kj} \xi_j + \sum_{i=1}^n a_{ik} \xi_i\end{aligned}$$

所以

$$\frac{df}{dx} = \begin{bmatrix} \frac{\partial f}{\partial \xi_1} \\ \vdots \\ \frac{\partial f}{\partial \xi_n} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n a_{1j} \xi_j \\ \vdots \\ \sum_{j=1}^n a_{nj} \xi_j \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^n a_{i1} \xi_i \\ \vdots \\ \sum_{i=1}^n a_{in} \xi_i \end{bmatrix} = \mathbf{A}\mathbf{x} + \mathbf{A}^T\mathbf{x} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x}$$

特别地, 当  $\mathbf{A}$  为对称矩阵时, 有

$$\frac{df}{dx} = 2\mathbf{A}\mathbf{x}$$

## 2. 函数矩阵对矩阵的导数

**定义 3.12** 设  $\mathbf{X} = (\xi_{ij})_{m \times n}$ ,

$m \times n$  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  $\mathbf{X}$  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

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

以两道例题为例, 来展示其形状

例  $\mathbf{x} = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}, \mathbf{F}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_l(\mathbf{x})]$

$$\frac{d\mathbf{F}}{d\mathbf{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial \xi_1} & \cdots & \frac{\partial f_l}{\partial \xi_1} \\ \vdots & & \vdots \\ \frac{\partial f_1}{\partial \xi_n} & \cdots & \frac{\partial f_l}{\partial \xi_n} \end{bmatrix}$$

例  $\mathbf{A} = (a_{ij})_{n \times n}, \quad \mathbf{x} = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}$



$$\frac{d(Ax)}{dx^T} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = A$$

各种求导公式,

分子布局  $\frac{\partial y}{\partial x} = \frac{dy}{dx^T}$

$$\begin{aligned} \frac{\partial y}{\partial \mathbf{x}} &= \left[ \frac{\partial y}{\partial x_1} \frac{\partial y}{\partial x_2} \cdots \frac{\partial y}{\partial x_n} \right] \\ \frac{\partial \mathbf{y}}{\partial x} &= \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix} \\ \frac{\partial \mathbf{y}}{\partial \mathbf{x}} &= \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \\ \frac{\partial y}{\partial \mathbf{X}} &= \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{21}} & \cdots & \frac{\partial y}{\partial x_{p1}} \\ \frac{\partial y}{\partial x_{12}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{p2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{1q}} & \frac{\partial y}{\partial x_{2q}} & \cdots & \frac{\partial y}{\partial x_{pq}} \end{bmatrix} \end{aligned}$$

分母布局  $\frac{\partial y}{\partial x} = \frac{dx}{dy^T}$

$$\begin{aligned} \frac{\partial y}{\partial \mathbf{x}} &= \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix} \\ \frac{\partial \mathbf{y}}{\partial x} &= \begin{bmatrix} \frac{\partial y_1}{\partial x} & \frac{\partial y_2}{\partial x} & \cdots & \frac{\partial y_m}{\partial x} \\ \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \end{aligned}$$

$$\frac{\partial y}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1q}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2q}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{p1}} & \frac{\partial y}{\partial x_{p2}} & \cdots & \frac{\partial y}{\partial x_{pq}} \end{bmatrix}$$

下面是维基百科上的求导公示表

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 $\mathbf{x}$
$\mathbf{a}$ is not a function of $\mathbf{x}$	$\frac{\partial \mathbf{a}}{\partial \mathbf{x}} =$	$\mathbf{0}$	
	$\frac{\partial \mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{I}$	
$\mathbf{A}$ 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 $\mathbf{x}$ , $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial \mathbf{A}\mathbf{u}}{\partial \mathbf{x}} =$	$\mathbf{A} \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \mathbf{A}^\top$
$\mathbf{u} = \mathbf{u}(\mathbf{x})$ , $\mathbf{v} = \mathbf{v}(\mathbf{x})$	$\frac{\partial (\mathbf{u} + \mathbf{v})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$	
$\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}$
$\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}} \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}$

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 $\mathbf{x}$ , $u = u(\mathbf{x})$	$\frac{\partial au}{\partial \mathbf{x}} =$		$a \frac{\partial u}{\partial \mathbf{x}}$
$u = u(\mathbf{x}), v = v(\mathbf{x})$	$\frac{\partial(u+v)}{\partial \mathbf{x}} =$		$\frac{\partial u}{\partial \mathbf{x}} + \frac{\partial v}{\partial \mathbf{x}}$
$u = u(\mathbf{x}), v = v(\mathbf{x})$	$\frac{\partial uv}{\partial \mathbf{x}} =$		$u \frac{\partial v}{\partial \mathbf{x}} + v \frac{\partial u}{\partial \mathbf{x}}$
$u = u(\mathbf{x})$	$\frac{\partial g(u)}{\partial \mathbf{x}} =$		$\frac{\partial g(u)}{\partial u} \frac{\partial u}{\partial \mathbf{x}}$
$u = u(\mathbf{x})$	$\frac{\partial f(g(u))}{\partial \mathbf{x}} =$		$\frac{\partial f(g)}{\partial g} \frac{\partial g(u)}{\partial u} \frac{\partial u}{\partial \mathbf{x}}$
$\mathbf{u} = \mathbf{u}(\mathbf{x}), \mathbf{v} = \mathbf{v}(\mathbf{x})$	$\frac{\partial(\mathbf{u} \cdot \mathbf{v})}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}^\top \mathbf{v}}{\partial \mathbf{x}} =$	$\mathbf{u}^\top \frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^\top \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$ • 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 <sup>[5]</sup>

$\mathbf{a}$ 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	$\frac{\partial^2 \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}^2} =$	$2 \mathbf{A}$	
	$\frac{\partial(\mathbf{x} \cdot \mathbf{x})}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{x}} =$	$2 \mathbf{x}^\top$	$2 \mathbf{x}$
$\mathbf{a}$ 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 $\mathbf{x}$	$\frac{\partial \ \mathbf{x} - \mathbf{a}\ }{\partial \mathbf{x}} =$	$\frac{(\mathbf{x} - \mathbf{a})^\top}{\ \mathbf{x} - \mathbf{a}\ }$	$\frac{\mathbf{x} - \mathbf{a}}{\ \mathbf{x} - \mathbf{a}\ }$

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.	
$\mathbf{U} = \mathbf{U}(x), \mathbf{v} = \mathbf{v}(x)$	$\frac{\partial (\mathbf{U} \times \mathbf{v})}{\partial x} =$	$\frac{\partial \mathbf{U}}{\partial x} \times \mathbf{v} + \mathbf{U} \times \frac{\partial \mathbf{v}}{\partial x}$	$\mathbf{v}^\top \times \left( \frac{\partial \mathbf{U}}{\partial x} \right) + \frac{\partial \mathbf{v}}{\partial x} \times \mathbf{U}^\top$

Identities: scalar-by-matrix $\frac{\partial y}{\partial \mathbf{X}}$			
Condition	Expression	Numerator layout, i.e. by $\mathbf{x}^\top$	Denominator layout, i.e. by $\mathbf{X}$
$a$ is not a function of $\mathbf{X}$	$\frac{\partial a}{\partial \mathbf{X}} =$	$\mathbf{0}^\top$ [6]	$\mathbf{0}$ [6]
$a$ is not a function of $\mathbf{X}$ , $u = u(\mathbf{X})$	$\frac{\partial a u}{\partial \mathbf{X}} =$	$a \frac{\partial u}{\partial \mathbf{X}}$	
$u = u(\mathbf{X}), v = v(\mathbf{X})$	$\frac{\partial (u + v)}{\partial \mathbf{X}} =$	$\frac{\partial u}{\partial \mathbf{X}} + \frac{\partial v}{\partial \mathbf{X}}$	
$u = u(\mathbf{X}), v = v(\mathbf{X})$	$\frac{\partial u v}{\partial \mathbf{X}} =$	$u \frac{\partial v}{\partial \mathbf{X}} + v \frac{\partial u}{\partial \mathbf{X}}$	
$u = u(\mathbf{X})$	$\frac{\partial g(u)}{\partial \mathbf{X}} =$	$\frac{\partial g(u)}{\partial u} \frac{\partial u}{\partial \mathbf{X}}$	
$u = u(\mathbf{X})$	$\frac{\partial f(g(u))}{\partial \mathbf{X}} =$	$\frac{\partial f(g)}{\partial g} \frac{\partial g(u)}{\partial u} \frac{\partial u}{\partial \mathbf{X}}$	
$\mathbf{U} = \mathbf{U}(\mathbf{X})$	$^{[5]} \frac{\partial g(\mathbf{U})}{\partial \mathbf{X}_{ij}} =$	$\text{tr} \left( \frac{\partial g(\mathbf{U})}{\partial \mathbf{U}} \frac{\partial \mathbf{U}}{\partial \mathbf{X}_{ij}} \right)$	$\text{tr} \left( \left( \frac{\partial g(\mathbf{U})}{\partial \mathbf{U}} \right)^\top \frac{\partial \mathbf{U}}{\partial \mathbf{X}_{ij}} \right)$
		Both forms assume numerator layout for $\frac{\partial \mathbf{U}}{\partial \mathbf{X}_{ij}}$ , i.e. mixed layout if denominator layout for $\mathbf{X}$ is being used.	
$a$ and $b$ are not functions of $\mathbf{X}$	$\frac{\partial a^\top \mathbf{X} b}{\partial \mathbf{X}} =$	$b a^\top$	$a b^\top$
$a$ and $b$ are not functions of $\mathbf{X}$	$\frac{\partial a^\top \mathbf{X}^\top b}{\partial \mathbf{X}} =$	$a b^\top$	$b a^\top$
$a, b$ and $\mathbf{C}$ are not functions of $\mathbf{X}$	$\frac{\partial (\mathbf{X} a + b)^\top \mathbf{C} (\mathbf{X} a + b)}{\partial \mathbf{X}} =$	$((\mathbf{C} + \mathbf{C}^\top)(\mathbf{X} a + b) a^\top)^\top$	$(\mathbf{C} + \mathbf{C}^\top)(\mathbf{X} a + b) a^\top$
$a, b$ and $\mathbf{C}$ are not functions of $\mathbf{X}$	$\frac{\partial (\mathbf{X} a)^\top \mathbf{C} (\mathbf{X} b)}{\partial \mathbf{X}} =$	$(\mathbf{C} \mathbf{X} b a^\top + \mathbf{C}^\top \mathbf{X} a b^\top)^\top$	$\mathbf{C} \mathbf{X} b a^\top + \mathbf{C}^\top \mathbf{X} a b^\top$

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

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

有了上面这些公式表，我们在计算时可以直接查表来做

例  $\frac{d(Ax)}{dx^T} = \frac{\partial Ax}{\partial x} = \mathbf{A} \frac{\partial x}{\partial x} = \mathbf{A}$

例  $\frac{dx^T Ax}{dx} = \frac{\partial x^T Ax}{\partial x} = (\mathbf{A} + \mathbf{A}^T)x$

## 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  $A$  的特征多项式为  $\varphi(\lambda) = \det(\lambda I - A)$ . 如果首 1 多项式

$$\psi(\lambda) = \lambda^m + b_1\lambda^{m-1} + \cdots + b_{m-1}\lambda + b_m \quad (1 \leq m \leq n) \quad (3.3.9)$$

满足  $\psi(A) = O$  且  $\psi(\lambda)$  整除  $\varphi(\lambda)$  (矩阵  $A$  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  $A$  的特征值. 记  $\psi(\lambda)$  的互异零点为  $\lambda_1, \lambda_2, \dots, \lambda_s$ , 相应的重数为  $r_1, r_2, \dots, r_s$  ( $r_1 + r_2 + \cdots + r_s = m$ ), 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

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

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  $\psi(A) = O$ , 可得

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

#### 3.1.2 举例展示求法

例  $A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & 3 \end{bmatrix}$ , 求  $e^A, e^{tA} \quad (t \in R)$

解

$$\varphi(\lambda) = |\lambda I - A| = \begin{vmatrix} \lambda - 2 & 0 & 0 \\ -1 & \lambda - 1 & -1 \\ -1 & 1 & \lambda - 3 \end{vmatrix} = \begin{vmatrix} \lambda - 2 & 0 & 0 \\ 0 & \lambda - 2 & 0 \\ 0 & 0 & (\lambda - 2)^2 \end{vmatrix}$$

$\therefore$  最小多项式  $m(\lambda) = (\lambda - 2)^2$

取  $\psi(\lambda) = (\lambda - 2)^2$

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

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

$$\text{解得} \begin{cases} a = -e^2 \\ b = e^2 \end{cases}$$

$$\therefore r(\lambda) = e^2(\lambda - 1)$$

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

设  $f(\lambda) = e^{t\lambda} = \psi(\lambda)g(\lambda) + r(\lambda)$ , 其中  $r(\lambda) = a + b\lambda$

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

$$\text{解得} \begin{cases} a = (1 - 2t)e^{2t} \\ b = te^{2t} \end{cases}$$

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

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

## 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 (3.3.9), 且满足  $\psi(A) = O$ , 即

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

或者

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

由此, 低次幂表示高次幂可以求出

于是有

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

### 3.2.2 举例展示求法

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

解

$$\varphi(\lambda) = \det(\lambda I - A) = \lambda^4 - \pi^2 \lambda^2. \text{ 由于 } \varphi(A) = O,$$

$$\text{所以 } A^4 = \pi^2 A^2, A^5 = \pi^2 A^3, A^7 = \pi^4 A^3, \cdots.$$

于是有

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

### 3.3 对角型法

#### 3.3.1 对角型法求矩阵函数的步骤推导

设  $A$  相似于对角矩阵  $\Lambda$ ，即有可逆矩阵  $P$ ，使得

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

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

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

从而

$$\begin{aligned} f(A) &= \sum_{k=0}^{\infty} c_k A^k = P \begin{bmatrix} \sum_{k=0}^{\infty} c_k \lambda_1^k & & \\ & \ddots & \\ & & \sum_{k=0}^{\infty} c_k \lambda_n^k \end{bmatrix} P^{-1} \\ &= P \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix} P^{-1} \end{aligned} \quad (3.3.15)$$

#### 3.3.2 举例展示求法

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

解

$\varphi(\lambda) = \det(\lambda I - A) = (\lambda + 2)(\lambda - 1)^2$ 。对应  $\lambda_1 = -2$  的特征向量  $p_1 = (-1, 1, 1)^T$ ；对应  $\lambda_2 = \lambda_3 = 1$  的两个线性无关的特征向量  $p_2 = (-2, 1, 0)^T, p_3 = (0, 0, 1)^T$ 。构造矩阵

$$P = (p_1, p_2, p_3) = \begin{bmatrix} -1 & -2 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

则有

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



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

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

利用式(3.3.15), 求得

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

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

### 3.4 若尔当标准型法

#### 3.4.1 若尔当标准型法求矩阵函数的步骤推导

设  $A$  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

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

$$\begin{bmatrix} f(\lambda_i) & \frac{1}{1!} f'(\lambda_i) & \cdots & \frac{1}{(m_i-1)!} f^{(m_i-1)}(\lambda_i) \\ & f(\lambda_i) & \ddots & \vdots \\ & & \ddots & \frac{1}{1!} f'(\lambda_i) \\ & & & f(\lambda_i) \end{bmatrix} \quad (3.3.16)$$

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

$$P \begin{bmatrix} \sum_{k=0}^{\infty} c_k J_1^k & & \\ & \ddots & \\ & & \sum_{k=0}^{\infty} c_k J_s^k \end{bmatrix} P^{-1} = P \begin{bmatrix} f(J_1) & & \\ & \ddots & \\ & & f(J_s) \end{bmatrix} P^{-1} \quad (3.3.17)$$

### 3.4.2 举例展示求法

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

矩阵  $A$  是一个 Jordan 标准形, 它的三个 Jordan 块为

$$J_1 = \pi, \quad J_2 = -\pi, \quad J_3 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

由式(3.3.16)求得

$$\begin{aligned} \sin J_1 &= \sin \pi = 0 \\ \sin J_2 &= \sin(-\pi) = 0 \\ \sin J_3 &= \begin{bmatrix} \sin 0 & \frac{1}{1!} \cos 0 \\ 0 & \sin 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

根据式(3.3.17), 可得(取  $P = I$ )

$$\sin A = \begin{bmatrix} \sin J_1 & & \\ & \sin J_2 & \\ & & \sin J_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$f(A) = \sum_{k=0}^{\infty} c_k A^k = P \cdot \sum_{k=0}^{\infty} c_k J^k \cdot P^{-1} = P \cdot \text{diag}(f(J_1), \dots, f(J_s)) \cdot P^{-1}$$

## 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

矩阵  $A$  满秩时, 将初等行变换的过程写成下面这个式子, 其中矩阵  $L$  的主对角元素都是 1

$$A = A^{(0)} = L_1 A^{(1)} = L_1 L_2 A^{(2)} = \cdots = L_1 L_2 \cdots L_{n-1} A^{(n-1)}$$

容易求出

$$L = L_1 L_2 \cdots L_{n-1} = \begin{bmatrix} 1 & & & & \\ c_{21} & 1 & & & \\ \vdots & \vdots & \ddots & & \\ c_{n-1,1} & c_{n-1,2} & \cdots & 1 & \\ c_{n1} & c_{n2} & \cdots & c_{n,n-1} & 1 \end{bmatrix}$$

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

$$A = LU$$

这样  $A$  就分解成一个单位下三角矩阵与一个上三角矩阵的乘积, 一般地有如下的定义.

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

高斯消去法和 LU 分解法, 两种方法的时间复杂度相同, 工程中解线性方程组时可以采用 LU 分解法

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

$$A = L(DU) = L\hat{U}$$

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

$$A = (LD)U = \hat{L}U$$

称为  $A$  的 **Crout 分解**.

若矩阵  $A$  是对称正定矩阵, 那么  $A = LDU$  的分解中,  $U = L^T$ ,  $D = \tilde{D}^2$  那么

$$A = L\tilde{D}^2 L^T = (L\tilde{D})(L\tilde{D})^T = GG^T \quad (4.1.32)$$

这里  $G = L\tilde{D}$  是下三角矩阵.

定义 4.3 称式 (4.1,32) 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

例 求矩阵  $A$  的 LU 分解  $A = \begin{bmatrix} 2 & -1 & 3 \\ 1 & 2 & 1 \\ 2 & 4 & 3 \end{bmatrix}$

$$A \rightarrow A^{(1)} \rightarrow A^{(2)} = U$$

$$A = \begin{bmatrix} 2 & -1 & 3 \\ 1 & 2 & 1 \\ 2 & 4 & 3 \end{bmatrix} \quad A^{(1)} = \begin{bmatrix} 2 & -1 & 3 \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 5 & 0 \end{bmatrix} \quad A^{(2)} = \begin{bmatrix} 2 & -1 & 3 \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}$$

$$L_1 = \begin{bmatrix} 1 & & \\ -\frac{1}{2} & 1 & \\ -1 & & 1 \end{bmatrix} \quad L_2 = \begin{bmatrix} 1 & & \\ & 1 & \\ & -2 & 1 \end{bmatrix}$$

$$U = A^{(2)} = L_2 L_1 A$$

$$\text{令 } L = L_1^{-1} L_2^{-1}, \text{ 则 } A = LU = \begin{bmatrix} 1 & & \\ \frac{1}{2} & 1 & \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 3 \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}$$

例 求矩阵  $A$  的 LDU 分解:  $A = \begin{bmatrix} 2 & -1 & 3 \\ 1 & 2 & 1 \\ 2 & 4 & 2 \end{bmatrix}$

$$A \rightarrow A^{(1)} \rightarrow A^{(2)} = U$$

$$A = \begin{bmatrix} 2 & -1 & 3 \\ 1 & 2 & 1 \\ 2 & 4 & 2 \end{bmatrix} \quad A^{(1)} = \begin{bmatrix} 2 & -1 & 3 \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 5 & -1 \end{bmatrix} \quad A^{(2)} = \begin{bmatrix} 2 & -1 & 3 \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}$$

$$L_1 = \begin{bmatrix} 1 & & \\ -\frac{1}{2} & 1 & \\ -1 & & 1 \end{bmatrix} \quad L_2 = \begin{bmatrix} 1 & & \\ & 1 & \\ & -2 & 1 \end{bmatrix}$$

$$\text{令 } L = L_1^{-1} L_2^{-1}, \text{ 则 } A = LA^{(2)}$$

和 LU 分解类似, 不同的是最后要将 LU 分解中的矩阵  $U$  的主对角元素抽出来

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

## 4.2 矩阵的 QR 分解

### 4.2.1 矩阵 QR 分解的步骤推导

定义 4.6 如果实(复) 可逆矩阵  $A$  能够化成正交 (酉) 矩阵  $Q$  与实(复) 可逆上三角矩阵  $R$  的乘积, 即

$$A = QR \quad (4.2.7)$$

则称式 (4.2.7) 为  $A$  的 QR 分解.

**定理 4.6** 设  $A$  是  $n$  阶实(复) 可逆矩阵, 则存在正交(酉) 矩阵  $Q$  和实(复) 可逆上三角矩阵  $R$ , 使  $A$  有 QR 分解式 (4.2.7). 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 (4.2.7) 是唯一的.

证 记矩阵  $A$  的  $n$  个列向量依次为  $a_1, a_2, \dots, a_n$ . 因为  $A$  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  $q_1, q_2, \dots, q_n$ .

对  $a_1, a_2, \dots, a_n$  正交化, 可得

$$b_1 = a_1$$

$$b_2 = a_2 - k_{21}b_1$$

....

$$b_n = a_n - k_{n,n-1}b_{n-1} - \dots - k_{n1}b_1$$

其中  $k_{ij} = \frac{(a_i, b_j)}{(b_j, b_j)} \quad (j < i)$ . 将上式改写为

$$a_1 = b_1$$

$$a_2 = k_{21}b_1 + b_2$$

....

$$a_n = k_{n1}b_1 + k_{n2}b_2 + \dots + k_{n,n-1}b_{n-1} + b_n$$

用矩阵形式表示为

$$(a_1, a_2, \dots, a_n) = (b_1, b_2, \dots, b_n)C$$

其中

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

再对  $b_1, b_2, \dots, b_n$  单位化, 可得

$$q_i = \frac{1}{|b_i|} b_i \quad (i = 1, 2, \dots, n)$$

于是有

$$(a_1, a_2, \dots, a_n) = (b_1, b_2, \dots, b_n)C = (q_1, q_2, \dots, q_n) \begin{bmatrix} |b_1| & & & \\ & |b_2| & & \\ & & \ddots & \\ & & & |b_n| \end{bmatrix} C$$

令

$$Q = (q_1, q_2, \dots, q_n) \quad \left. \begin{aligned} R &= \text{diag}(|b_1|, |b_2|, \dots, |b_n|) \cdot C \end{aligned} \right\}$$

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

为了证明唯一性, 设  $A$  有两个分解式  $A = QR = Q_1R_1$ , 由此可得

$$Q = Q_1R_1R^{-1} = Q_1D$$

其中  $D = R_1R^{-1}$  仍为可逆上三角矩阵. 于是

$$I = Q^H Q = (Q_1D)^H (Q_1D) = D^H D$$

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  $R_1 = DR, Q_1 = QD^{-1}$ .

证毕

#### 4.2.2 举例展示求法

**例** 试用 Schmidt 正交化方法求矩阵  $A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix}$  的 QR 分解.

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

$$b_1 = a_1 = (1, 2, 1)^T$$

$$b_2 = a_2 - b_1 = (1, -1, 1)^T$$

$$b_3 = a_3 - \frac{1}{3}b_2 - \frac{7}{6}b_1 = \left(\frac{1}{2}, 0, -\frac{1}{2}\right)^T$$

$$(a_1 \ a_2 \ a_3) = (b_1 \ b_2 \ b_3) \begin{pmatrix} 1 & 1 & \frac{7}{6} \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{aligned} (a_1 \ a_2 \ a_3) &= \begin{pmatrix} b_1 & b_2 & b_3 \\ |b_1| & |b_2| & |b_3| \end{pmatrix} \begin{pmatrix} 1 & 1 & \frac{7}{6} \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 1 \end{pmatrix} \\ &= (q_1 \ q_2 \ q_3) \begin{pmatrix} |b_1| & & \\ & |b_2| & \\ & & |b_3| \end{pmatrix} \begin{pmatrix} 1 & 1 & \frac{7}{6} \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 1 \end{pmatrix} = QR \end{aligned}$$

$$Q = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} \end{bmatrix}, R = \begin{bmatrix} \sqrt{6} & & \\ & \sqrt{3} & \\ & & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 1 & \frac{7}{6} \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{6} & \sqrt{6} & \frac{7}{\sqrt{6}} \\ & \sqrt{3} & \frac{1}{\sqrt{3}} \\ & & \frac{1}{\sqrt{2}} \end{bmatrix}$$

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

**定义 4.8** 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

$$A = FG \quad (4.3.1)$$

则称式 (4.3.1) 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.

**定理 4.13** 设  $A \in C_r^{m \times n} (r > 0)$ , 则  $A$  有满秩分解式 (4.3.1),

证  $rank A = r$  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

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

则有

$$A = P^{-1}B = [F \quad S] \begin{bmatrix} G \\ O \end{bmatrix} = FG$$

其中  $F$  是列满秩矩阵,  $G$  是行满秩矩阵.

证毕

按照定理 4.13, 我们可以利用初等行变换的方法求矩阵满秩分解

### 4.3.2 举例展示求法

例  $A = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 1 & 2 & -1 & 1 \\ 2 & 2 & -2 & -1 \end{bmatrix}$ , 求  $A = FG$

$$\begin{aligned} A &\rightarrow A^{(1)} \rightarrow A^{(2)} \\ A^{(1)} &= \begin{bmatrix} -1 & 0 & 1 & 2 \\ 0 & 2 & 0 & 3 \\ 0 & 2 & 0 & 3 \end{bmatrix} \quad A^{(2)} = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ A^{(1)} &= L_1^{-1}A \quad A^{(2)} = L_2^{-1}A^{(1)} \quad A = L_1L_2A^{(2)} \\ L_1^{-1} &= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \quad L_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \\ A &= L_1L_2A^{(2)} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 & 2 \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 & 2 \\ 0 & 2 & 0 & 3 \end{bmatrix} = FG \end{aligned}$$

## 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

定义 4.11 设  $A \in C_r^{m \times n} (r > 0)$ ,  $A^H A$  的特征值为

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

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

定理 4.15 设  $A \in R^{n \times n}$  可逆, 则存在正交矩阵  $P$  和  $Q$ , 使得

$$P^T A Q = \text{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n)$$

其中  $\sigma_i > 0 (i = 1, 2, \cdots, n)$  .

**定理 4.16** 设  $A \in C_r^{m \times n} (r > 0)$  , 则存在  $m$  阶酉矩阵  $U$  和  $n$  阶酉矩阵  $V$  , 使得

$$U^H A V = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}$$

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

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^H$$

这个式子就是矩阵  $A$  的奇异值分解

求奇异值分解的步骤如下

1、令  $B = A^H A$  , 计算特征值特征向量  $Bx = \lambda x$  得到  $\lambda_k, x_k$

$$B \frac{x_k}{|x_k|} = \lambda_k \frac{x_k}{|x_k|}, k = 1, 2, \dots, N$$

(因为  $B$  是Hermite矩阵, 所以  $x_1, \dots, x_N$  两两正交)

记  $v_k = \frac{x_k}{|x_k|}$  , 则有

$$Bv_k = \lambda_k v_k, k = 1, 2, \dots, N$$

$V = (v_1, \dots, v_N)$  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  $A^H U_1 = V_1 \Sigma \Rightarrow U_1^H A = \Sigma V_1^H \Rightarrow A = U_1 \Sigma V_1^H$

3、将  $U_1$  扩充成酉矩阵  $U = [U_1: U_2]$

最后, 我们得到了  $A = UDV^H$

#### 4.4.2 举例展示求法

**例** 求矩阵  $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$  的奇异值分解.

解

$$\text{令 } B = A^T A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

$$|\lambda I - B| = \lambda(\lambda - 1)(\lambda - 3) = 0$$

$$\text{得 } \lambda_1 = 3, \lambda_2 = 1, \lambda_3 = 0$$

特征值 3, 1, 0 对应的特征向量依次为

$$\xi_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \xi_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \xi_3 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix},$$

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

$$\sigma_1 = \sqrt{3}, \sigma_2 = 1, \sigma_3 = 0$$



$$\Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix}$$

$\because \xi_1, \xi_2, \xi_3$  两两正交

$\therefore \xi_1, \xi_2, \xi_3$  线性无关

$$B(\xi_1, \xi_2, \xi_3) = (\xi_1, \xi_2, \xi_3) \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}$$

已经正交，再单位化，有

$$B\left(\frac{\xi_1}{|\xi_1|}, \frac{\xi_2}{|\xi_2|}, \frac{\xi_3}{|\xi_3|}\right) = \left(\frac{\xi_1}{|\xi_1|}, \frac{\xi_2}{|\xi_2|}, \frac{\xi_3}{|\xi_3|}\right) \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}$$

$$\text{记 } q_1 = \frac{\xi_1}{|\xi_1|} = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix}, \quad q_2 = \frac{\xi_2}{|\xi_2|} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}, \quad q_3 = \frac{\xi_3}{|\xi_3|} = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{bmatrix}$$

$$\text{记 } V = (q_1, q_2, q_3) = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \end{bmatrix}$$

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

$$\therefore BV = AV$$

$$\because A^H AV = V \Lambda, \Sigma = \begin{bmatrix} \sqrt{3} & \\ & 1 \end{bmatrix}$$

$\therefore A^H AV_1 = V_1 \Sigma^2$ , 其中  $V_1$  是  $V$  去掉冗余项

$$\text{令 } U_1 = AV_1 \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & 0 \end{bmatrix}$$

$$\text{将 } U_1 \text{ 扩充成酉矩阵 } U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\text{得到 } A \text{ 的奇异值分解为 } A = U \begin{bmatrix} \sqrt{3} & & \\ & 1 & \\ & & 0 \end{bmatrix} V^H$$

#### 4.4.3 利用奇异值分解求矩阵广义逆

**定义 6.1** 设矩阵  $A \in \mathbb{C}^{m \times n}$ , 若矩阵  $X \in \mathbb{C}^{n \times m}$  满足以下 4 个 Penrose 方程

$$(1) \quad AXA = A;$$

$$(2) \quad XAX = X$$

$$(3) \quad (AX)^H = AX;$$

$$(4) \quad (XA)^H = XA$$

则称  $X$  为  $A$  的 **Moore-Penrose 逆**, 记为  $A^+$  .

**定理 6.1** 对任意  $A \in \mathbb{C}^{m \times n}$ ,  $A^+$  存在并且唯一.

上面定理保证了广义逆的存在, 下面这个定理则是说明广义逆如何去求

**定理 6.2** 设  $A \in \mathbb{C}_r^{m \times n}$  的不可逆值分解为

$$A = U \begin{bmatrix} \Sigma_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{m \times n} V^H$$

那么

$$A^+ = V \begin{bmatrix} \Sigma_r^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{n \times m} U^H$$

**例** 设  $A = \begin{bmatrix} -1 & 0 & 1 \\ 2 & 0 & -2 \end{bmatrix}$  求  $A^+$

先求  $A$  的奇异值分解:

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

$$\text{令 } V_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 \\ 2 & 0 & -2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

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

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

## 第五章 总结

本论文分别研究了矩阵函数的求法和矩阵分解方法。（总结研究结果和学习心得）

在欧式空间中，快速求解线性变换，本质上是下面这四步

0、求一组标准正交基  $e_1, \dots, e_n$ ，并求向量  $x$  的坐标

1、由  $e_1, \dots, e_n$  生成新的基  $E_1, \dots, E_n$

2、通过坐标变换得到向量在基  $E_1, \dots, E_n$  下的坐标

3、求  $T$  在基  $E_1, \dots, E_n$  下的矩阵  $A \quad T(E_1, \dots, E_n) = (E_1, \dots, E_n)A$

而其中  $A$  直接就是若尔当标准型，则  $A = J$ ； $T(E_1, \dots, E_n) = (E_1, \dots, E_n)J$

在第二章的 2.1 节中，我们构建了欧式空间中线性变换求法，其步骤如下

设  $V$  是欧式空间， $T$  是  $V$  上的一个线性变换，求  $z = (T^k)(x)$ ，其中  $x \in V$

0、任意找一组基，利用 Schmidt 正交化方法得到  $V$  的一组标准正交基  $e_1, \dots, e_n$ ，  
 $x = k_1 e_1 + \dots + k_n e_n$ ，其中  $k_i = (x, e_i)$

0.1、求  $T$  在基  $e_1, \dots, e_n$  下的矩阵  $A_0 \Rightarrow T(e_1, \dots, e_n) = (e_1, \dots, e_n)A_0$

0.2、其中  $A_0 = PJP^{-1}$ ， $J$  是 Jordan 标准型  $\Rightarrow T(e_1, \dots, e_n) = (e_1, \dots, e_n)PJP^{-1}$

0.3、 $T(e_1, \dots, e_n)P = (e_1, \dots, e_n)PJ$

1、得到一组新的基  $(E_1, \dots, E_n) = (e_1, \dots, e_n)P$ ，

2、通过坐标变换得到  $x = (E_1, \dots, E_n)P^{-1} \begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix} = (E_1, \dots, E_n) \begin{pmatrix} l_1 \\ \vdots \\ l_n \end{pmatrix}$

3、 $T$  在新基下的矩阵： $T(E_1, \dots, E_n) = (E_1, \dots, E_n)J$

4、 $T(x) = (E_1, \dots, E_n)J \begin{pmatrix} l_1 \\ \vdots \\ l_n \end{pmatrix} \Rightarrow (T^k)(x) = (E_1, \dots, E_n)J^k \begin{pmatrix} l_1 \\ \vdots \\ l_n \end{pmatrix}$

范数是对向量长度的推广，是对向量长度的一种度量；矩阵范数是定义在矩阵空间上的要求更高的一种度量。

矩阵序列是对数列概念的推广，借助矩阵范数，我们可以将矩阵序列的收敛性问题转化为正向数列的收敛性问题

矩阵级数的对常数项级数概念的推广，借助矩阵范数，可以将矩阵级数的绝对收敛问题转为正项级数的收敛性问题

矩阵函数是以矩阵自身为自变量的一种函数，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

在计算矩阵导数的时候，我们可以分为分子布局 and 分母布局两种，根据表达式去查表，进而得到我们想要的结果

对一般的方阵而言，如果满秩，则可以进行 LU 分解，在此条件下， $n$  阶方阵的 LDU 分解，Crout 分解以及 Doolittle 分解也都是唯一的

而 Crout 分解和 Doolittle 分解可以由 LDU 分解构造出来

借助 LU 分解，可以将一般的方阵求逆计算转化为上三角矩阵和下三角矩阵的求逆

计算，这在解线性方程组的时候是很有用的

任何矩阵都可以进行 QR 分解，借助 QR 分解可以将一般方阵的求逆计算转为三角矩阵的求逆计算，也可以将一般的线性代数方程组问题转化为三角方程组的求解问题

矩阵的奇异值分解是对特征值概念的推广，利用奇异值分解可以给出线性代数方程组的最小二乘解（求广义逆矩阵）

而广义逆矩阵在矩阵可逆的情况下，是完全一样的。在编程的时候，利用广义逆矩阵，可以极大地提高程序的健壮性。

## 参考文献

- [1] 张凯院，徐仲等，矩阵论，西北工业大学出版社，2017 年 8 月.