北京郵電大學

课 程 期 末 论 文

题　　目： 矩阵函数的求法和矩阵分解方法研究

姓　　名： 付容天

学　　院： 计算机学院（国家示范性软件学院）

专　　业： 计算机科学与技术

班　　级： 2020211314

学　　号： 2020211616

任课教师： 李昊辰

2022 年 1 月

矩阵函数的求法和矩阵分解方法研究

摘　要

本文对矩阵函数的求法和矩阵分解方法进行了研究，主要参考了西北工业大学出版社《矩阵论》一书，从矩阵理论的基础知识开始介绍，再给出一些在矩阵理论中常见的方法、定理和结论。在本文的主要部分介绍了矩阵函数的四种求解方法：待定系数法、数项级数求和法、对角型法以及 Jordan 标准型法；也介绍了矩阵的四种分解方法：矩阵的 LU 分解、矩阵的 QR 分解、矩阵的满秩分解以及矩阵的奇异值分解。本文梳理了矩阵理论与方法课程一学期的主要内容。

关键词：矩阵理论　矩阵函数　矩阵分解

Research on the Method of Matrix Function and Matrix Decomposition

Abstract

This paper studies the method of matrix function and matrix decomposition method, mainly referring to the book *Matrix Theory* by Northwestern Polytechnical University Press, starting from the basic knowledge of matrix theory, and then giving some common methods in matrix theory, theorems and conclusions. In the main part of this paper, four methods for solving matrix functions are introduced: undetermined coefficient method, numerical series summation method, diagonal method and Jordan standard method; four decomposition methods of matrix are also introduced: matrix LU Decomposition, QR Decomposition, Full Rank Decomposition and Singular Value Decomposition. This paper sorts out the main content of a semester of matrix theory and methods course.

KEY WORDS: Matrix Theory   Matrix Function   Matrix Decomposition

目录

1 引言

1.1 背景介绍

1.1.1 矩阵理论与方法介绍

矩阵理论是一门研究矩阵在数学上的应用的科目。矩阵理论本来是线性代数的一个小分支，但其后由于陆续在图论、代数、组合数学和统计上得到应用，渐渐发展为一门独立的学科。在矩阵理论的应用过程中，主要涉及到的有各种计算及其方法，其中常用的计算主要有：特征值与特征向量的计算、矩阵的最小多项式的计算、矩阵的 Schmidt 正交化、向量范数与矩阵范数的计算、矩阵谱半径的计算、矩阵函数的计算、矩阵的微分与积分、矩阵分解以及广义逆矩阵的计算。使用的计算方法主要来自线性代数的基础知识及其推广。

1.1.2 函数矩阵和矩阵函数介绍

一般来讲，函数矩阵指的是一个矩阵，其中的每个元素都是一个已经定义的函数，常用在线性系统的分析过程中；而矩阵函数则是指一个函数，其自变量与因变量都是一个确定大小的矩阵，实际上是一种映射法则，实现的结果是一种线性变换。

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线性方程组成的代数方程组，可能包含数个未知变量以及若干线性约束关系，求解线性代数方程组就是在给定这些约束关系的前提下，解出这些未知变量的具体数值或者相互之间的约束关系。线性代数方程组的解通常有以下三种情况：唯一解、无穷多解、无解。通过秩计算、初等变换、求逆矩阵等方法，可以确定待求的线性代数方程组的解属于上述三种情况中的哪一种情况。

1.2 问题介绍

1.2.1 矩阵函数的求法问题介绍

与高等数学引入函数概念的过程类似，此处引入矩阵函数之前，我们首先介绍矩阵序列和矩阵级数的概念，由此得到极限与收敛的概念，之后可以自然地引入矩阵函数的定义和概念。矩阵函数的求法从根本上来说是基于矩阵函数的幂级数展开，具体的方法主要有以下四种：待定系数法、数项级数求和法、对角型法以及 Jordan 标准型法。

1.2.2 矩阵分解的方法问题介绍

矩阵分解的一个重要的作用就是减小计算量，通过将复杂矩阵及逆行分解、进而得到某些确定结构下的矩阵，这种操作使得复杂矩阵以数个较为简单矩阵的乘积形式来呈现，从而在后续的计算中大大减少计算量，这是非常有意义的。矩阵分解从根本上来说是基于一些基本操作与已知结论，具体的分解方法主要有以下四种：矩阵的 LU 分解、矩阵的 QR 分解、矩阵的满秩分解、以及矩阵的奇异值分解，其中矩阵的奇异值分解还可以用来求解矩阵的广义逆矩阵，这在实际应用中是非常有意义的。

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、数项级数求和法、对角型法以及 Jordan 标准型法。这四种方法操作起来思路明确，理论难度小，但是有些时候可能会比较繁琐、且容易导致计算错误。在一些较新的研究中，有研究者提出了一些其他的矩阵函数值的求解方法，包括但不限于：利用留数定理计算矩阵函数值、利用矩阵的谱分解计算矩阵函数值等。

1.3.2 矩阵分解方法研究现状

在《矩阵论》一书中介绍了常见的矩阵分解的四种方法：矩阵的 LU 分解、矩阵的 QR 分解、矩阵的满秩分解以及矩阵的奇异值分解。在一些较新的研究中，有研究者提出了一些新的矩阵的分解方法，包括但不限于：矩阵的谱分解、利用牛顿法进行矩阵的分解等。

1.4 本论文工作简述

1.4.1 本论文对上述问题研究简述

本论文将从矩阵理论的基本概念与方法出发，简要介绍基本概念与方法之后，引入矩阵函数的概念，通过对矩阵函数的求法进行研究和总结，并进一步分析不同情况下不同方法之间的区别。之后本论文将总结矩阵分解的若干常见方法，并进一步分析不同情况下不同方法之间的区别。

1.4.2 本论文特点简述

本论文主要基于《矩阵论》(张凯院、徐仲，西北工业大学出版社 2017 版)，并结合《Matrix Computations》（Gene H. Golub, Charles F. Van Loan, The Johns Hopkins University Press, 4th Edition）以及北京邮电大学计算机学院（国家示范性软件学院）的课程《矩阵理论与方法》的课程 PPT（2021 年秋季）进行撰写。本论文将从基本概念出发逐步引出矩阵理论中的核心概念和关键方法，并给出重要定理的证明，最后总结出矩阵函数的常见求法和矩阵分解的常用方法。

1.4.3 本论文撰写结构简述

第一章引言部分主要介绍本论文的背景以及主要书写内容；第二章预备知识将给出矩阵理论的必备基础知识以及一般的分析思路和方法；第三章是矩阵函数的求法研究，将从矩阵函数的定义出发，研究并总结常见的四种矩阵函数的求法，并在每种求法的理论推导之后给出一个具体的例子加以说明；第四章是矩阵分解方法研究，将从矩阵分解的重要意义出发，说明矩阵分解的必要性和重要性，研究并总结常见的四种矩阵分解的方法，并在每种方法的理论推导之后给出一个具体的例子加以说明；第五章将总结本论文的整个书写过程以及对于矩阵理论与方法课程的感受；第六章将给出所有的参考文献。

2 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

欧氏空间（Euclidean space）是一类特殊的向量空间，对三维空间中的向量可以讨论长度、夹角等几何性质。之前的学习中所涉及到的空间都是欧氏空间，但实际上欧氏空间可以拓展到更一般的情况，也就是线性空间。对于给定的数域 K 和一个非空集合 V，用 x，y，z 表示 V 中的元素、用 l，m，n 表示 K 的元素，并定义此时的加法运算和数乘运算，如果以下性质均满足：

结合律 $x+(y+z)=(x+y)+z$

交换律 $x+y=y+x$

存在零元素0, 使 $x+0=x$

存在负元素, 即对任意向量 $x\in V$,存在向量 $y\in V$,使 $x+y=0$,称 y 为 x 的负元素,记为 $-x$

数因子分配律 $k(x+y)=kx+ky$

分配律 $(k+l)x=kx+lx$

结合律 $k(lx)=(kl)x$

$1x=x$

则称 V 为数域 K 上的线性空间或向量空间。

线性空间的意义在于给出了一个关于向量以及其他诸多数学概念的抽象的代数系统，使得许多问题均可以在线性空间的讨论下进行求解，这种一般性的方法作为一种处理问题的手段在数学的许多领域有着重要的应用，比如多项式空间和矩阵空间。视非空集合 V 中的元素为向量，则同 n 维线性空间的概念一样，此时也可以引出线性组合、线性相关与线性无关、基与维数、坐标、子空间与子空间的交与和、同构、商空间、线性变换等概念。下面将讨论线性变换的概念。

由于线性空间中的任意元素都可以看作是基底的线性组合，那么对于给定的映射法则 T，如果 T 满足：

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

则称 T 为 V 的一个线性变换或线性算子。这个式子的实质是映射法则 T 对向量的线性运算是封闭的。从已有结论出发，可以得到以下性质或定义：线性相关的向量组经线性变换仍为线性相关、V 中所有向量的象形成的集合称为 T 的值域 $R(T)$、V 中所有被 T 变为零向量的原象构成核空间 $N(T)$、核子空间、象子空间、T 的秩、T 的亏（或零度）。

线性变换也具有可运算性。线性变换的两个最基本的变换是单位变换和恒等变换，单位变换 T_e 满足

$T_e x = x$，恒等变换 T_0 满足 $T_0 x = 0$，接下来定义线性变换的几种运算如下：

$$
\begin{aligned}
\text{加法}\quad & (T_1 + T_2)x = T_1 x + T_2 x \\
\text{数乘}\quad & (kT)x = k(Tx) \\
\text{乘法}\quad & (T_1 T_2)x = T_1(T_2 x) \\
\text{逆变换}\quad & (ST)x = (TS)x = x \\
\text{多项式}\quad & f(T) = a_0 T^m + a_1 T^{m-1} + \cdots + a_{m-1}T + a_m T_e \\
& \quad (\text{其中} T^{m+n} = T^m T^n, (T^m)^n = T^{mn})
\end{aligned}
$$

通过坐标，可以将线性变换用矩阵表示出来，从而可将抽象的线性变换转化为具体的矩阵来处理。由于线性空间中的任意元素都可以视为基向量的线性组合，因此只要能够确定基向量在线性变换下的象，就能够确定线性空间中任意元素在线性变换下的象。对于基向量 $x_1, x_2, \cdots, x_n$，线性变换 T 有：

$$
\begin{cases}
Tx_1 = a_{11}x_1 + a_{21}x_2 + \cdots + a_{n1}x_n \\
Tx_2 = a_{12}x_1 + a_{22}x_2 + \cdots + a_{n2}x_n \\
\quad \cdots \\
Tx_n = a_{1n}x_1 + a_{2n}x_2 + \cdots + a_{nn}x_n
\end{cases}
$$

采用矩阵乘法形式，可将上述方程组表示为：

$$
T(x_1, x_2, \cdots, x_n) = (x_1, x_2, \cdots, x_n)A
$$

其中

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

这时就将矩阵 A 称为线性变换 T 在 V^n 的基 $(x_1, x_2, \cdots, x_n)$ 下的矩阵，简称 A 为 T 的矩阵。一些常见的线性变换的矩阵在形式上是简单的，如零变换 T_0 的矩阵是零矩阵 O、单位变换的矩阵是单位矩阵 I、数乘变换的矩阵是数量矩阵 mI 等。

2.1.2 Jordan 标准型的求解

将一个矩阵转化为其 Jordan 标准型，意义在于选择适当的线性空间的基或坐标系，从而优化矩阵的结构、使得后续的计算过程变得简单。在线性代数课程中已经学过特征值与特征向量的概念和求解方法，下面将以此为基础讨论线性变换对应矩阵的 Jordan 标准型的求解。

首先可以证明，线性变换的矩阵的特征多项式与基的选择无关，它直接被线性变换所决定，这保证了线性变换矩阵特征多项式的唯一性。定义最小多项式：首项系数是 1，次数最小，且以矩阵 A 为根的 λ 的多项式称为 A 的最小多项式。借助 Hamilton-Cayley 定理可以得到如下结论：A 的最小多项式是其特征多项式的因式，也就是：矩阵 A 的最小多项式 $m(\lambda)$ 可整除以 A 为根的任意首 1 多项式 $\psi(\lambda)$，且 $m(\lambda)$ 是唯一的。

Jordan 标准型从结构上来说是一个准对角矩阵:

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

其中

$$J_i(\lambda_i)=\begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \lambda_i & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix} \qquad (i=1,2,\cdots,s)$$

称为一个 Jordan 块。

为计算矩阵的 Jordan 标准型，需要先用初等变换将矩阵化为标准形，其中初等变换指的是要把矩阵的某一行（列）的乘以数域 K 上的 λ 多项式的结果加到另一行（列）对应元素上去，标准形是指一个对角矩阵（对角线上的元素是首 1 多项式且前面的元素可以整除后面的元素）。可以证明标准形对角线上的非零元素 $d_i(\lambda)$ 不随矩阵的初等变换而改变，因此称 $d_i(\lambda)$ 为矩阵的不变因子或不变因式。若将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，那么这样的不可约因式（连同它们的幂指数）称为矩阵的一个初等因子，初等因子的全体称为初等因子组。下面给出求解 Jordan 标准型的具体步骤:

第一步　求特征矩阵$\lambda I - A$的初等因子组

第二步　写出每个初等因子$(\lambda-\lambda_i)^{m_i}$对应的 Jordan 块:

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

第三步　写出所有 Jordan 块构成的 Jordan 标准型:

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

2.1.3　欧式空间中线性变换的求法

在实际应用中，常见的问题涉及到欧氏空间中线性变换的一些性质，下面以《矩阵论》书例 1.36 为基础说明常见问题的处理方法。

（1）要求非空集合 V 的一组标准正交基，思路是先求一组基，再进行正交化，如有需要再进行单位

化。对于本题:

$$X=\begin{bmatrix}x_1 & x_2\\ x_3 & x_4\end{bmatrix}=x_1\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}+x_2\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}+x_3\begin{bmatrix}0 & 0\\ 1 & 1\end{bmatrix}$$

所以 V 的一个标准正交基为

$$X_1=\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}\quad X_2=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\quad X_3=\frac{1}{\sqrt{2}}\begin{bmatrix}0 & 0\\ 1 & 1\end{bmatrix}$$

(2)对称变换的判定方法是看相应的矩阵是否是对称矩阵,本题(1)问中已经求得了一组标准正交基,现在的问题就是如何用这组标准正交基求出线性变换对应的矩阵。通过计算基向组就可以得到这个矩阵,也就是通过等式 $T(X_1,X_2,X_3)=(X_1,X_2,X_3)A$ 就可以计算得到线性变换对应的矩阵 A 为:

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

这样就可以证明,这个线性变换确实是对称变换。

(3)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。题目的要求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 $T(Y_1,Y_2,Y_3)=(Y_1,Y_2,Y_3)\Lambda$ 成立,其中 Λ 为对角矩阵。根据线性代数的知识可以知道,线性变换对应的矩阵 A 可以通过左乘和右乘某矩阵从而化为仅由特征值构成的对角矩阵,因此容易得到

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

下面的问题就在于求出新的标准正交基 (Y_1,Y_2,Y_3),下面先进行线性变换的推导,然后得到答案:

$$\begin{aligned}T(X_1,X_2,X_3)&=(X_1,X_2,X_3)A\\ T(X_1,X_2,X_3)&=(X_1,X_2,X_3)Q\Lambda Q^{-1}\\ T(X_1,X_2,X_3)Q&=(X_1,X_2,X_3)Q\Lambda\\ T(Y_1,Y_2,Y_3)&=(Y_1,Y_2,Y_3)\Lambda\end{aligned}$$

其中

$$(Y_1,Y_2,Y_3)=(X_1,X_2,X_3)Q$$

由此便可以求得符合题意的新的标准正交基 (Y_1,Y_2,Y_3):

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

2.2 向量范数与矩阵范数

2.2.1 向量范数介绍

在向量空间中，长度的度量是向量的模，对于一般的线性空间，起到长度度量的概念是范数概念，范数是比长度更为一般的概念。对于数域 K 上的线性空间 V，对任意的 $x \in V$，定义一个实值函数 $||x||$，如果该实值函数满足

$$
\begin{aligned}
\text{非负性} \quad & \text{当}x \neq 0\text{时}, ||x|| > 0; \text{当}x = 0\ , ||x|| = 0 \\
\text{齐次性} \quad & ||ax|| = |a|\ ||x||\ (a \in K, x \in V) \\
\text{三角不等式} \quad & ||x + y|| \leq ||x|| + ||y||\ (x, y \in V)
\end{aligned}
$$

则称 $||x||$ 为 V 上的向量 x 的范数，简称向量范数。下面给出一些常见的向量范数：

$$
\begin{aligned}
2 - \text{范数} \quad & ||x||_2 = \sqrt{|\xi_2|^2 + |\xi_2|^2 + \cdots + |\xi_n|^2} \\
\infty - \text{范数} \quad & ||x||_\infty = \max_i |\xi_i| \\
1 - \text{范数} \quad & ||x||_1 = \sum_{i=1}^{n} |\xi_i| \\
p - \text{范数} \quad & ||x||_p = (\sum_{i=1}^{n} |\xi_i|^p)^{1/p} \\
\text{椭圆范数} \quad & ||x||_A = (x^T A x)^{1/2}
\end{aligned}
$$

需要指出的是，向量范数通常和向量空间中基的选取有关，但是有定理保证线性空间上向量范数的等价性，也就是说，如果存在有限维线性空间 V 上的任意两种向量范数 $||x||_\alpha$ 和 $||x||_\beta$，且满足：

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

那么就称这两种向量范数是等价的。

2.2.2 矩阵范数介绍

在以矩阵为元素的线性空间中，起到长度度量作用的概念是范数概念，也就是矩阵范数。和向量范数类似，设 $A \in C^{n\times n}$，定义一个实值函数 $||A||$，对于下面的四条性质

$$
\begin{aligned}
\text{非负性} \quad & \text{当}A \neq 0\text{时}, ||A|| > 0; \text{当}A = 0\text{时}, ||A|| = 0 \\
\text{齐次性} \quad & ||\alpha A|| = |\alpha|\ ||A||\ (\alpha \in C) \\
\text{三角不等式} \quad & ||A + B|| \leq ||A|| + ||B||\ (B \in C^{n\times n}) \\
\text{相容性} \quad & ||AB|| \leq ||A||\ ||B||
\end{aligned}
$$

如果该实值函数满足前三条性质，则称 $||A||$ 为 A 的广义矩阵范数；如果该实值函数满足全部的四条性质，则称 $||A||$ 为 A 的矩阵范数。下面给出一些常见的矩阵范数：

$$
\begin{aligned}
\text{列和范数} \quad & ||A||_1 = \max_j |a_{ij}| \\
\text{谱范数} \quad & ||A||_2 = \sqrt{\lambda_1}, \quad \lambda_1\text{为}A^H A\text{的最大特征值} \\
\text{行和范数} \quad & ||A||_\infty = \max_i |a_{ij}|
\end{aligned}
$$

实际上，矩阵范数和向量范数是紧密相关的，有什么样的向量范数就有什么样的矩阵范数，由向量范数导出的矩阵范数简称为从属范数，上述三种常见矩阵范数都是对应的向量范数的从属范数。

2.2.3 矩阵可逆性条件、谱半径和条件数介绍

在有了矩阵范数的概念之后，下面介绍范数的应用场合和一些新的概念。

在判断矩阵的可逆性时,可以根据范数 $||A||$ 的大小来判断矩阵 $I-A$ 是否为可逆矩阵。对于 $A \in C^{n\times n}$ 以及相应的某种矩阵范数 $||\cdot||$，如果有 $||A|| < 1$，那么矩阵 $I-A$ 可逆，且有

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

这条定理给出了矩阵可逆性的判断方法。

谱半径在诸多理论中都有着重要的应用，称 $\rho(A) = \max_i |\lambda_i|$ 为矩阵 A 的谱半径，其中 λ_i 为矩阵 A 的特征值。就方阵而言，矩阵范数就是矩阵谱半径的上界，对于任意给定的矩阵都可以构造出一种矩阵范数使得该范数与谱半径充分接近，这在科学计算中应用广泛。

条件数是求矩阵逆的摄动的一个重要量。设 $\mathrm{cond}(A) = ||A||\,||A^{-1}||$，则当 $||A^{-1}\,||\delta A|| < 1$ 时，称 $\mathrm{cond}(A)$ 为矩阵 A 的条件数。一般来说，条件数越大，$(A+\delta A)^{-1}$ 和 A^{-1} 的相对误差就越大。

2.3 矩阵函数介绍

2.3.1 矩阵序列介绍

同数学分析一样，矩阵分析理论的建立也是基于极限理论的，因此此处先介绍矩阵序列的相关概念。按照正整数 k 的顺序，将 $C^{m\times n}$ 中的矩阵排成一列，$A_1, A_2, \cdots, A_k, \cdots$，称这列有序的矩阵为矩阵序列，称 A_k 为矩阵序列的一般项。下面给出矩阵序列收敛的定义：设有矩阵序列 $\{A^{(k)}\}$，其中 $A^{(k)} = (a_{ij}^{(k)})_{m\times n} \in C^{m\times n}$，当 $a_{ij}^{(k)} \to a_{ij}$ 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

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

且不收敛的矩阵序列称为发散。和数列收敛的性质类似，矩阵序列收敛也有一些有用的性质：

$$
\begin{aligned}
& \text{设}A^{(k)} \to A_{m\times n}\text{以及}B^{(k)} \to B_{m\times n}\text{则有} \\
& \text{性质 1} \quad \lim_{k\to\infty}(\alpha A^{(k)} + \beta B^{(k)}) = \alpha A + \beta B \\
& \text{性质 2} \quad \lim_{k\to\infty} A^{(k)} B^{(k)} = AB \\
& \text{性质 3} \quad (A^{(k)})^{-1} \to A^{-1}
\end{aligned}
$$

在判断矩阵序列的收敛性的时候，有一些常用的结论，比如：$A^{(k)} \to O$ 的充分必要条件是 $||A^{(k)}|| \to 0$、$A^{(k)} \to A$ 的充分必要条件是 $||A^{(k)} - A|| \to 0$，其中 $||\cdot||$ 为任意一种矩阵范数。在实际应用中，一种常见的矩阵序列是由方阵构成的矩阵序列，如果有 $A^{(k)} \to O$，则称 A 为收敛矩阵，且 A 为收敛矩阵的两个充分必要条件是 $\rho(A) < 1$ 或者 $||A|| < 1$，其中 $||\cdot||$ 为任意一种矩阵范数。

2.3.2 矩阵级数介绍

矩阵级数是建立矩阵函数理论的基础，称矩阵序列形成的无穷项和 $A^{(0)} + A^{(1)} + \cdots + A^{(k)} + \cdots$ 为矩阵级数，记为 $\sum_{k=0}^{\infty} A^{(k)}$，可称为矩阵级数式。记 $S^{(N)} = \sum_{k=0}^{N} A^{(k)}$ 为矩阵级数式的部分和，如果矩阵序列 $\{S^{(N)}\}$ 收敛，且有极限 S，则有 $\lim_{N\to\infty} S^{(N)} = S$，那么就称矩阵级数式收敛，且有和 S。不收敛的矩阵级数称为是发散的。最后给出矩阵级数绝对收敛的概念，矩阵级数绝对收敛指的是其中的每个数项级数都是绝对收敛的。下面给出矩阵级数收敛性的一些性质：

性质 1　若矩阵级数式绝对收敛, 则它也一定收敛, 且任意调换其项顺序所得级数仍收敛, 且其和不变

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

性质 3　如果$\displaystyle\sum_{k=0}^{\infty} A^{(k)}$是收敛 (或绝对收敛) 的, 那么$\displaystyle\sum_{k=0}^{\infty} PA^{(k)}Q$也是收敛 (或绝对收敛) 的

性质 4　设级数S_1和S_2绝对收敛, 其和为A和B, 则S_1和S_2按项相乘所得矩阵级数收敛, 且有和AB

对于矩阵级数而言，有一种常见的矩阵级数，即矩阵幂级数。对于方阵幂级数（Neumann 级数）而言，方阵 A 的幂级数 $\sum_{k=0}^{\infty} A^k = I + A + A^2 + \cdots + A^k + \cdots$ 收敛的充分必要条件是 A 为收敛矩阵，并且在收敛时，其和为 $(I - A)^{-1}$。对于矩阵幂级数 $\sum_{k=0}^{\infty} c_k A^k$ 而言，其对应的纯量幂级数 $f(z) = \sum_{k=0}^{\infty} c_k z^k$ 的收敛半径为 r，则如果方阵 A 满足 $\rho(A) < r$，则矩阵幂级数是绝对收敛的；如果 $\rho(A) > 1$，则矩阵幂级数是发散的。并且，如果纯量幂级数式在整个复平面上是收敛的，那么不论 A 是何种矩阵，矩阵幂级数式总是绝对收敛的。

2.3.3 矩阵函数介绍

矩阵函数是以 n 阶矩阵为自变量和因变量的一种函数。如果一元函数 $f(z)$ 能展开为 z 的幂级数 $f(z) = \sum_{k=0}^{\infty} c_k z^k \quad (|z| < r)$，其中 r 为收敛半径，则当 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

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

2.3.4 函数矩阵对矩阵的导数

在数学分析课程中，建立起函数概念之后便开始讨论函数的导数概念，在矩阵分析领域也是如此。函数矩阵是指以变量 t 的函数 $a_{ij}(t)$ 为元素的矩阵，在形式上可以认为是由函数作为元素所构成的一个矩阵。函数矩阵的微分和积分就是指作为其中每个元素的函数各自的微分和积分，运算所得的结果仍是一个函数矩阵。除了函数矩阵的导数以外，还有纯量对向量、向量对向量、矩阵对向量、矩阵对矩阵的导数问题。

函数对矩阵的导数：设 $X=(\xi_{ij})_{m\times n}$，mn 元函数 $f(X)=f(\xi_{11},\xi_{12},\cdots,\xi_{1n},\xi_{21},\cdots,\xi_{mn})$，则：

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

如上所示，函数对矩阵的导数所得到的结果是一个矩阵，该矩阵的行数和列数与矩阵 X 一样，其中的每个元素都是函数对矩阵 X 相应位置上的元素的偏导数。

函数矩阵对矩阵的导数：设 $X=(\xi_{ij})_{m\times n}$，mn 元函数 $f_{ij}(X)=f_{ij}(\xi_{11},\xi_{12},\cdots,\xi_{1n},\xi_{21},\cdots,\xi_{mn})$，其中 $i=1,2,\cdots,r;j=1,2,\cdots,s$，则：

$$F(X)=\begin{bmatrix} f_{11}(X) & \cdots & f_{1s}(X) \\ \vdots & & \vdots \\ f_{r1}(X) & \cdots & f_{rs}(X) \end{bmatrix} \qquad \frac{dF}{dX}=\begin{bmatrix} \frac{\partial F}{\partial \xi_{11}} & \frac{\partial F}{\partial \xi_{12}} & \cdots & \frac{\partial F}{\partial \xi_{1n}} \\ \frac{\partial F}{\partial \xi_{21}} & \frac{\partial F}{\partial \xi_{22}} & \cdots & \frac{\partial F}{\partial \xi_{2n}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial F}{\partial \xi_{m1}} & \frac{\partial F}{\partial \xi_{m2}} & \cdots & \frac{\partial F}{\partial \xi_{mn}} \end{bmatrix}$$

其中

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

可以认为函数矩阵对矩阵的导数是基于函数对矩阵的导数的：函数矩阵中的每个元素对矩阵进行导数，就是函数对矩阵的导数。

3　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

待定系数法是以 Hamilton-Cayley 定理为基础的一种求矩阵函数的方法。设 n 阶矩阵 A 的特征多项式为 $\phi(\lambda)=\det(\lambda I-A)$，且设首 1 多项式为 $\psi(\lambda)$，如果 $\psi(A)=O$，且 $\psi(\lambda)$ 整除 $\phi(\lambda)$，则根据 Hamilton-Cayley 定理知道 $\psi(\lambda)$ 的零点都是 A 的特征值，记 $\psi(\lambda)$ 的所有互异零点重数之和为 m，则

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

其中 $r(z)$ 为次数低于 m 的多项式，在确定出 $r(z)$ 之后，便知道 $f(A)=r(A)$。值得一提的是，求解 $r(z)$ 的时候可以回避掉 $g(z)$ 的计算，依据的原理是适当求导之后的 $\psi(z)g(z)$ 的值仍为零，具体的操作过程将在下面结合具体的例子加以说明。

3.1.2　举例展示求法

下面以《矩阵论》一书例 3.5 为基础说明待定系数法求矩阵函数的具体过程。

（1）容易得到最小多项式为 $m(\lambda)=(\lambda-2)^2$，故取 $\psi(\lambda)=(\lambda-2)^2$，此时最高次数为 2，故设 $r(\lambda)=a+b\lambda$，由于 2 为特征值，所以有下面的方程组：

$$\begin{cases}f(2)=e^2\\f'(2)=e^2\end{cases}$$

容易解得 $a=-e^2$，$b=e^2$，于是 $r(\lambda)=e^2(\lambda-1)$，故

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

（2）仍取 $\psi(\lambda)=(\lambda-2)^2$，此时最高次数为 2，故设 $r(\lambda)=a+b\lambda$，由于 2 为特征值，所以有下面的方程组：

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

容易解得 $a=(1-2t)e^{2t}$，$b=te^{2t}$，于是 $r(\lambda)=e^{2t}[(1-2t)+t\lambda]$，故

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

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

数项级数求和法是根据最小多项式导出的矩阵递推关系来求解求矩阵函数的方法。由于首 1 多项式 $\psi(\lambda)$ 满足 $\psi(A)=O$，也即

$$A^m+b_1A^{m-1}+\cdots+b_{m-1}A+b_mI=O$$

那么就可以得到

$$A^m=k_0I+k_1A+\cdots+k_{m-1}A^{m-1}\quad(k_i=-b_{m-i})$$

这就是关于 A^m 的递推关系，也就是说任意的 A^n 总能够通过序列的前 m 项来表示出来，这就将矩阵幂级数的求和问题转化为了 m 个矩阵的求和问题，即有

$$\begin{aligned}f(A)=\sum_{k=0}^{\infty}c_kA^k&=(c_0I+c_1A+\cdots+c_{m-1}A^{m-1})+c_m(k_0I+k_1A+\cdots+k_{m-1}A^{m-1})+\cdots\\&+c_{m+l}(k_0^{(l)}I+k_1^{(l)}A+\cdots+k_{m-1}^{(l)}A^{m-1})+\cdots\\&=(c_0+\sum_{l=0}^{\infty}c_{m+l}k_0^{(l)})I+(c_1+\sum_{l=0}^{\infty}c_{m+l}k_1^{(l)})A+\cdots+(c_{m-1}+\sum_{l=0}^{\infty}c_{m+l}k_{m-1}^{(l)})A^{m-1}\end{aligned}$$

3.2.2 举例展示求法

下面以《矩阵论》一书例 3.6 为基础说明数项级数求和法求矩阵函数的具体过程。

容易得到矩阵 A 的特征多项式为 $\phi(\lambda)=\det(\lambda I-A)=\lambda^4-\pi^2\lambda^2$，因此有 $\phi(A)=O$，也就有 $A^4=\pi^2A^2, A^5=\pi^2A^3, A^7=\pi^2A^5,\cdots$，于是

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

3.3 对角型法

3.3.1 对角型法求矩阵函数的步骤推导

对角型法就是线性代数课程中已经介绍过的求矩阵函数的方法。设 A 相似于对角矩阵 Λ，即有可逆矩阵 P，使得

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

则有矩阵函数 $\sum_{k=0}^{N} c_k A^k = \sum_{k=0}^{N} c_k P\Lambda^k P^{-1} = P(\sum_{k=0}^{N} c_k \Lambda^k)P^{-1}$，于是

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

3.3.2 举例展示求法

下面以《矩阵论》一书例 3.7 为基础说明对角型法求矩阵函数的具体过程。

矩阵 A 的特征多项式 $\phi(\lambda) = \det(\lambda I - A) = (\lambda+2)(\lambda-1)^2$，那么根据线性代数的知识可以构造矩阵

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

其中 p_1，p_2，p_3 为矩阵的三个特征向量，那么就有

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

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

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

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

Jordan 标准型法在形式上类似于对角型法。矩阵 A 的 Jordan 标准型为 J，则有可逆矩阵 P，使得

$$P^{-1}AP = J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_s \end{bmatrix} \qquad J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ & & & \lambda_i \end{bmatrix}_{m_i \times m_i}$$

那么矩阵函数就可以求解，形式为：

$$f(A) = P \begin{bmatrix} f(J_1) & & \\ & \ddots & \\ & & f(J_s) \end{bmatrix} P^{-1}$$

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

3.4.2 举例展示求法

下面以《矩阵论》一书例 3.6 为基础说明数项级数求和法求矩阵函数的具体过程。

矩阵 A 的三个 Jordan 块为

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

且根据求解过程可得

$$\sin J_1 = 0 \quad \sin J_2 = 0 \quad \sin J_3 = \begin{bmatrix} \sin 0 & \frac{1}{1!}\cos 1 \\ 0 & \sin 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

因此将此三个 Jordan 块进行组装便可以得到

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

4 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

矩阵的 LU 分解又称矩阵的三角分解，该分解方法是基于矩阵的 Gauss 消去法导出的。矩阵的 LU 分解得到的结果是方阵 A 被表示成一个下三角矩阵 L 和上三角矩阵 U 的乘积，也就是 $A = LU$。进一步，如果分解的结果可以是 $A = LDU$，其中 D 为对角矩阵，L 和 U 分别为单位下三角矩阵和单位上三角矩阵，那么则称这种分解为矩阵的 LDU 分解。Gauss 消去法是由一系列初等行变换组成的，对方阵 A 实行这些初等行变换等价于对方阵 A 左乘一系列的初等行变换矩阵，矩阵 LU 分解的关键一步就是找出这些初等行变换矩阵，也就是要找到

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

使得方阵 $A^{(n-1)}$ 为上三角矩阵, 记为 $U = A^{(n-1)}$, 又因为可以证明此时的矩阵 $L = L_1 L_2 \cdots L_{n-1}$ 是一个单位下三角矩阵, 所以这样就完成了矩阵的 LU 分解。对于一个方阵是否可做 LDU 分解, 可用顺序主子式来判断: 方阵 $A = (a_{ij})_{n\times n}$ 可做 LDU 分解当且仅当 A 的顺序主子式 $\Delta_k \neq 0$, 其中 $(k = 1, 2, \cdots, n-1)$。矩阵的 LU 分解和矩阵的 LDU 分解在具体操作上只有一点不同，这个不同将在后面结合具体的例子来说明。

当矩阵的前 $n-1$ 个顺序主子式均非零的条件不满足的时候，需要先对 A 左乘（或右乘）置换矩阵 P 之后再进行分解，得到 $PA = LDU$。

最后介绍两个概念：矩阵的 Doolittle 分解是指 $A = L(DU) = L\hat{U}$；而矩阵的 Crout 分解是指 $A = (LD)U = \hat{L}U$。

4.1.2 举例展示求法

下面以《矩阵论》一书例 4.1 为基础说明矩阵的 LU 分解和矩阵的 LDU 分解的具体过程。

因为顺序主子式 $\Delta_1 = 2 \neq 0$，$\Delta_2 = 5 \neq 0$，因此可以唯一地进行矩阵的 LDU 分解。构造矩阵

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

此时注意矩阵 L_1 结构上的特殊性保证了其在求逆的时候只需要将下三角部分取相反数即可，这减少了计算量。那么就有

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

此时再构造

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

然后计算得到

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

此时就得到了一个上三角矩阵 $A^{(2)}$，在矩阵的 LU 分解中，此时可以令 $L=L_1L_2$ 以及 $U=A^{(2)}$，分解完成，但是在矩阵的 LDU 分解中，还需要将 $A^{(2)}$ 进一步分解成一个对角矩阵和一个单位上三角矩阵（可以先设出单位上三角矩阵的待定形式之后进行反解），也就是

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

那么令 $L=L_1L_2$，则可以得到方阵 A 的 LDU 分解为

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

4.2 矩阵的 QR 分解

4.2.1 矩阵 QR 分解的步骤推导

矩阵的 QR 分解又称矩阵的正交三角分解，所使用的方法基于 Schmidt 正交化、或者 Givens 变换或者 Householder 变换。Schmidt 正交化的概念已经十分熟悉，下面介绍 Givens 变换和 Householder 变换的概念。Givens 变换指的是初等旋转变换，Givens 矩阵为

$$T_{ij}=\begin{bmatrix}1&&&&&&&&&&\\&\ddots&&&&&&&&&\\&&1&&&&&&&&\\&&&c&&&&-s&&&\\&&&&1&&&&&&\\&&&&&\ddots&&&&&\\&&&&&&1&&&&\\&&&-s&&&&c&&&\\&&&&&&&&1&&\\&&&&&&&&&\ddots&\\&&&&&&&&&&1\end{bmatrix}$$

其中两个含有 $-s$ 的行分别为第 i 行和第 j 行 $(i<j)$，并且 $c^2+s^2=1$。对一个向量进行 Givens 变换所得到的结果是改变这个向量的方向但是不改变这个向量的模（长度），而且对任何一个非零向量经过有限次的 Givens 变换总能时使其方向和 e 向量的方向相同，这一点由这个定理保证：设 $x=(\xi_1,\xi_2,\cdots,\xi_n)^T\neq 0$，则存在有限个 Givens 矩阵的乘积，记作 T，使得 $Tx=|x|e_1$，那么就有这样一个推理：设非零列向量 $x\in R^n$ 即单位列向量 $z\in R^n$，则存在有限个 Givens 矩阵的乘积，记作 T，使得 $Tx=|x|z$。

Householder 变换指的是对称变换，Householder 矩阵为：$H = I - 2uu^T$，对一个向量左乘 Householder 矩阵 H，实现的效果是将向量 x 映射为关于与单位向量 u 正交的 $n-1$ 维子空间对称的向量 y。Householder 矩阵又称为初等反射矩阵，它具有一些常见性质：

$$
\begin{aligned}
&(1)\ H^T = H \quad (\text{对称矩阵}) \\
&(2)\ H^T H = I \quad (\text{正交矩阵}) \\
&(3)\ H^2 = I \quad (\text{对合矩阵}) \\
&(4)\ H^{-1} = H \quad (\text{自逆矩阵}) \\
&(5)\ \det H = -1
\end{aligned}
$$

并且有这样一个定理：对于任何非零列向量 $x \in R^n (n > 1)$ 及单位列向量 $z \in R^n$，存在 Householder 矩阵 H，使得 $Hx = |x|z$，这和 Givens 矩阵的一个定理是类似的。而且，需要指出的是，初等旋转矩阵是两个初等反射矩阵的乘积，这也由相应的定理保证。

在介绍完 Givens 变换和 Householder 变换之后，下面来正式介绍矩阵的 QR 分解：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 $A = QR$，则称此式为 A 的 QR 分解。矩阵的 QR 分解有三个思路：一种方法是使用 Schmidt 正交化、一种方法是使用 Givens 变换、一种方法是使用 Householder 变换。

对于 Schmidt 正交化的思路，有如下定理：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，且除去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 $A = QR$ 是唯一的。下面将给出这个定理的证明。

记矩阵 A 的 n 个列向量依次为 $a_1, a_2, \cdots, a_n$，因为 A 可逆，故这 n 个列向量线性无关，可以进行正交化。在线性代数课程中已经对 Schmidt 正交化进行了足够详细的讨论，故本论文将在此处直接给出结果。对它们使用 Schmidt 正交化方法进行正交化，可以得到 n 个标准正交列向量 $q_1, q_2, \cdots, q_n$，且有

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

其中

$$
\begin{cases}
b_1 = a_1 \\
b_2 = a_2 - k_{21} b_1 \\
\quad \cdots \\
b_n = a_n - k_{n,n-1} b_{n-1} - \cdots - k_{n1} b_1
\end{cases}
\qquad \text{其中} k_{ij} = \frac{(a_i, b_j)}{(b_j, b_j)}
$$

且

$$
q_i = \frac{1}{|b_i|} b_i \qquad C = \begin{bmatrix} 1 & k_{21} & \cdots & k_{n1} \\ & 1 & \cdots & k_{n2} \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}
$$

之后令 $Q = (q_1, q_2, \cdots, q_n)$ 以及 $R = diag(|b_1|, |b_2|, \cdots, |b_n|) \cdot C$，就完成了矩阵的 QR 分解。

上面所述的定理是从 Schmidt 正交化的角度进行矩阵的 QR 分解，下面的定理将从 Givens 变换的角度将矩阵的 QR 分解进行推广：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 $Q^TQ=I(Q^HQ=I)$，R 是 n 阶实（复）可逆上三角矩阵，该分解除去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。下面将给出这个定理的证明。

第一步：由 $\det A\neq 0$ 知，A 的第一列 $b^{(1)}=(a_{11},a_{12},\cdots,a_{n1})^T\neq 0$，那么从前面关于 Givens 变换的叙述知道，存在有限个 Givens 矩阵的乘积，记作 T_1，使得 $T_1b^{(1)}=|b^{(1)}|e_1$，其中 $e_1\in R^n$，令 $a_{11}^{(1)}=|b^{(1)}|$ 则有

$$T_1A=\left[\begin{array}{c:ccc} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ \hdashline 0 & & & \\ \vdots & & A^{(1)} & \\ 0 & & & \end{array}\right]$$

第二步：由 $\det A^{(1)}\neq 0$ 知，$A^{(1)}$ 的第一列 $b^{(2)}=(a_{22}^{(1)},a_{32}^{(1)},\cdots,a_{n2}^{(1)})^T\neq 0$，那么从前面关于 Givens 变换的叙述知道，存在有限个 Givens 矩阵的乘积，记作 T_2，使得 $T_2b^{(2)}=|b^{(2)}|e_1$，其中 $e_1\in R^{n-1}$，令 $a_{22}^{(2)}=|b^{(2)}|$ 则有

$$T_1A=\left[\begin{array}{c:ccc} a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2n}^{(2)} \\ \hdashline 0 & & & \\ \vdots & & A^{(2)} & \\ 0 & & & \end{array}\right]$$

这样不断迭代下去，就能找到一系列的矩阵 T_i，最终得到

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

最后令

$$T=\begin{bmatrix} I_{n-2} & O \\ O & T_{n-1} \end{bmatrix}\cdots\begin{bmatrix} I_2 & O \\ O & T_3 \end{bmatrix}\begin{bmatrix} 1 & O \\ O & T_2 \end{bmatrix}T_1$$

这样，矩阵 T 就是有限个 Givens 矩阵的乘积且有

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

将上面最终得到的上三角矩阵记为 R，且令 $Q=T^{-1}$，这样就完成了矩阵的 QR 分解。

上述过程是从 Givens 变换的角度来进行矩阵的 QR 分解，下面将从 Householder 变换的角度来进行矩阵的 QR 分解，有如下定理：任何实可逆矩阵 A 可通过左连乘 Householder 矩阵化为可逆上三角矩阵。下面将给出这个定理的证明。

第一步：由 $\det A\neq 0$ 知，A 的第一列 $b^{(1)}=(a_{11},a_{12},\cdots,a_{n1})^T\neq 0$，那么从前面关于 Householder

变换的叙述知道，存在 Householder 矩阵 H_1，使得 $H_1 b^{(1)} = |b^{(1)}| e_1$，其中 $e_1 \in R^n$，令 $a_{11}^{(1)} = |b^{(1)}|$ 则有

$$H_1 A = \left[\begin{array}{c:ccc} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ \hdashline 0 & & & \\ \vdots & & A^{(1)} & \\ 0 & & & \end{array}\right]$$

第二步：由 $\det A^{(1)} \neq 0$ 知，$A^{(1)}$ 的第一列 $b^{(2)} = (a_{22}^{(1)}, a_{32}^{(1)}, \cdots, a_{n2}^{(1)})^T \neq 0$，那么从前面关于 Householder 变换的叙述知道，存在 Householder 矩阵 H_2，使得 $H_2 b^{(2)} = |b^{(2)}| e_1$，其中 $e_1 \in R^{n-1}$，令 $a_{22}^{(2)} = |b^{(2)}|$ 则有

$$H_1 A = \left[\begin{array}{c:ccc} a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2n}^{(2)} \\ \hdashline 0 & & & \\ \vdots & & A^{(2)} & \\ 0 & & & \end{array}\right]$$

这样不断迭代下去，就能找到一系列的矩阵 H_i，最终得到

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

最后令

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

并且注意到当 H_u 是 $n-l$ 阶 Householder 矩阵（即 $H_u = I_{n-l} - 2uu^T$）的时候，令 $v = (0, u)^T \in R^n$，则有 $v^T v = u^T u = 1$，且容易证明 $\begin{bmatrix} I_l & O \\ O & H_u \end{bmatrix} = I_n - 2vv^T$，这样，矩阵 S 就是有限个 Householder 矩阵的乘积且有

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

将上面最终得到的上三角矩阵记为 R，且令 $Q = S^{-1}$，这样就完成了矩阵的 QR 分解。

这样，我们给出了三种思路来进行矩阵的 QR 分解：一种方法是使用 Schmidt 正交化、一种方法是使用 Givens 变换、一种方法是使用 Householder 变换，这三种方法将在下面的例子中分别加以说明。

4.2.2 举例展示求法

下面以《矩阵论》一书例 4.6 为基础说明使用 Schmidt 正交化方法进行矩阵 QR 分解的具体过程。

可以提取出矩阵 A 的三个列向量分别为

$$a_1 = (1,2,1)^T \quad a_2 = (2,1,2)^T \quad a_3 = (2,2,1)^T$$

使用 Schmidt 正交化方法对这三个列向量进行正交化得到

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

对得到的三个正交的列向量进行单位化之后可以构造出矩阵 Q 为

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

这样就构造出了矩阵 Q，故现在的任务是构造出矩阵 R 就能完成矩阵的 QR 分解，根据步骤推导所述的内容，容易构造出

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

下面以《矩阵论》一书例 4.7 为基础说明使用 Givens 变换方法进行矩阵 QR 分解的具体过程。

对 A 的第一列 $b^{(1)} = (0\ 1\ 1)^T$ 构造 T_1，使得 $T_1 b^{(1)} = |b^{(1)}|e_1$ 有

$$\begin{aligned}
T_{12} &= \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad T_{12}b^{(1)} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \\
T_{13} &= \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}, \quad T_{13}(T_{12}b^{(1)}) = \begin{bmatrix} \sqrt{2} \\ 0 \\ 0 \end{bmatrix} \\
T_1 = T_{13}T_{12} &= \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -1 & 0 & 0 \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}, \quad T_1 A = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & -1 & -1 \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \\
\text{令} \quad A^{(1)} &= \begin{bmatrix} -1 & -1 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}
\end{aligned}$$

此后再对 $A^{(1)}$ 的第一列进行类似的操作，不断迭代，按照理论推导中叙述的思路，最后就能得到

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

这样就完成了矩阵 A 的 QR 分解。

下面以《矩阵论》一书例 4.8 为基础说明使用 Householder 变换方法进行矩阵 QR 分解的具体过程。

对 A 的第一列 $b^{(1)}=(3,6,6)^T$ 构造 T_1，使得 $T_1b^{(1)}=|b^{(1)}|e_1$ 有

$$b^{(1)}=\begin{bmatrix}3\\6\\6\end{bmatrix},\quad b^{(1)}-|b^{(1)}|e_1=6\begin{bmatrix}-1\\1\\1\end{bmatrix},\quad u=\frac{1}{\sqrt{3}}\begin{bmatrix}-1\\1\\1\end{bmatrix}H_1=I-2uu^T=\frac{1}{3}\begin{bmatrix}1&2&2\\2&1&-2\\2&-2&1\end{bmatrix},\quad H_1A=\begin{bmatrix}9&48&15\\0&9&-3\\0&-12&9\end{bmatrix},$$

此后再对 $A^{(1)}$ 的第一列进行类似的操作，不断迭代，按照理论推导中叙述的思路，最后就能得到

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

4.3 矩阵的满秩分解

4.3.1 矩阵满秩分解的步骤推导

矩阵的满秩分解是指将非零矩阵分解为列满秩矩阵与行满秩矩阵的乘积。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 $A=FG$，则称此式为矩阵 A 的满秩分解。并且当 A 是满秩（列满秩或行满秩）矩阵时，A 可以分解为一个因子是单位矩阵，另一个因子是 A 本身，此时称此满秩分解为平凡分解。

当 $\mathrm{rank}A=r$ 时，可以对 A 进行初等行变换将 A 化为阶梯形矩阵 B，即 $A\to B=\begin{bmatrix}G\\O\end{bmatrix}$，其中 $G\in C_r^{r\times n}$。于是存在有限个 m 阶初等矩阵的乘积 P 使得 $PA=B$，也就有

$$A=P^{-1}B=\left[\begin{array}{c:c}F&S\end{array}\right]\begin{bmatrix}G\\O\end{bmatrix}=FG\qquad(F\in C_r^{m\times r},S\in C_{m-r}^{m\times(m-r)},G\in C_r^{r\times n})$$

以上叙述给出了使用初等行变换方法进行矩阵满秩分解的思路，但是求解矩阵 P 及其逆矩阵的计算有时是十分麻烦的，下面给出使用 Hermite 标准型的思路来回避掉这些可能会比较复杂的运算。

设 $B\in C_r^{m\times n}(r>0)$，且满足

(1) B的前r行中每一行至少含一个非零元素, 且第一个非零元素为 1, 而后$m-r$行元素均为零

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

(3) B中的$j_1,j_2,\cdots,j_r$列为单位矩阵I_m的前r列

那么就称 B 为 Hermite 标准型。实际上，Hermite 标准型就是初等变换意义下的行最简形。另外，如果 B 满足

(1) B的后$m-r$行元素均为零

(2) B中的$j_1,j_2,\cdots,j_r$列为单位矩阵I_m的前r列

则称 B 为拟 Hermite 标准型。在矩阵的满秩分解过程中使用（拟）Hermite 标准型可以简化计算：满秩分解式 $A=FG$ 中可取 F 为 A 的 $j_1,j_2,\cdots,j_r$ 列构成的 $m\times r$ 矩阵，G 为 B 的前 r 行构成的 $r\times n$ 矩阵，这条性质由相应的定理保证。

上面实际上给出了两种进行矩阵的满秩分解的思路：一种是用初等行变换进行求解，另一种是用（拟）Hermite 标准型进行求解，下面将分别用例子演示这两种方法。

4.3.2 举例展示求法

下面以《矩阵论》一书例 4.10 为基础说明使用初等行变换方法来进行矩阵满秩分解的具体过程。

需要对矩阵 A 进行初等行变换使其成为阶梯形矩阵 B，并且需要得到初等矩阵的乘积 P，根据线性代数中的思路，先将矩阵 A 扩展成为矩阵 $\left[\begin{array}{c:c} A & I \end{array}\right]$ 之后再进行初等行变换便能得到 $\left[\begin{array}{c:c} B & P \end{array}\right]$，容易得到

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

于是有

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

下面以《矩阵论》一书例 4.11 为基础说明使用 Hermite 思路来进行矩阵满秩分解的具体过程。

对矩阵 A 进行初等行变换得到：

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

其中 B 是 Hermite 标准型，因为 B 的第一列和第三列构成了 I_3 的前两列，所以 F 为 A 的第一列和第三列构成的 3×2 矩阵，从而有

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

4.4 矩阵的奇异值分解

4.4.1 矩阵奇异值分解的步骤推导

矩阵的奇异值分解是一种基于矩阵 A^HA 的特征值的分解方法。在正式介绍矩阵的奇异值分解之前，首先需要下面三个易证结论：

$$\begin{aligned}&(1)\ 设 A\in C_r^{m\times n}(r>0),则 A^HA 是\ \text{Hermite}\ 矩阵,\ 且其特征值均是非负实数\\&(2)\ \text{rank}(A^HA)=\text{rank}A\\&(3)\ 设 A\in C^{m\times n},则 A=O 的充要条件是 A^HA=O\end{aligned}$$

下面先给出奇异值的概念：设 $A\in C_r^{m\times n}(r>0)$，$A^HA$ 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。矩阵的奇异值分解由这样一个定理保证：设 $A\in C_r^{m\times n}(r>0)$，则存在 m 阶酉矩阵 U 和 n 阶酉矩阵 V，使得 $U^HAV=\begin{bmatrix}\Sigma&O\\O&O\end{bmatrix}$，其中 $\Sigma=\text{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)$，而 $\sigma_i(i=1,2,\cdots,r)$ 为矩阵 A 的全部非零奇异值。

下面将给出这个定理的证明。

记 Hermite 矩阵 A^HA 的特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0$，则存在 n 阶酉矩阵 V，使得

$$V^H(A^HA)V = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = \begin{bmatrix} \Sigma^2 & O \\ O & O \end{bmatrix}$$

将矩阵 V 分块得到 $V = \left[\begin{array}{c:c} V_1 & V_2 \end{array}\right]$，其中 $V_1 \in C_r^{n\times r}, V_2 \in C_{n-r}^{n\times(n-r)}$，那么上式可以改写为

$$A^HAV = V\begin{bmatrix} \Sigma^2 & O \\ O & O \end{bmatrix}$$

那么有 $A^HAV_1 = \Sigma^2$ 以及 $A^HAV_2 = O$，则有 $(AV_1\Sigma^{-1})^H(AV_1\Sigma^{-1}) = I_r$ 以及 $(AV_2)^H(AV_2)$，此时可令 $U_1 = AV_1\Sigma^{-1}$, 则有 $U_1^HU_1 = I_r$, 即 U_1 的 r 个列向量是两两正交的单位向量, 记作 $U_1 = (u_1, u_2, \cdots, u_r)$。现在可将 $u_1, u_2, \cdots, u_r$ 扩充为 C^m 的标准正交基，记增添的向量为 $u_{r+1}, \cdots, u_m$，并构造矩阵 $U_2 = (u_{r+1}, \cdots, u_m)$，这样就得到了矩阵 $U = \left[\begin{array}{c:c} U_1 & U_2 \end{array}\right]$，于是可得

$$U^HAV = U^H\left[\begin{array}{c:c} AV_1 & AV_2 \end{array}\right] = \begin{bmatrix} U_1^H \\ U_2^H \end{bmatrix}\left[\begin{array}{c:c} U_1\Sigma & O \end{array}\right] = \begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}$$

4.4.2 举例展示求法

下面以《矩阵论》一书例 4.14 为基础说明矩阵奇异值分解的具体过程。

首先计算出矩阵 $B = A^TA$ 的特征值与特征向量：

$$\lambda_1 = 3 \quad \lambda_2 = 1 \quad \lambda_3 = 0 \xi_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \quad \xi_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \xi_3 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

于是可得 $\mathrm{rank}A = 2$ 以及 $\Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix}$，利用上述的特征向量可以构造出矩阵 V、利用步骤推导中的思路可以求得矩阵 U，如下：

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

然后可以使用步骤推导中的思路求得矩阵 U：

$$U_1 = AV_1\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & 0 \end{bmatrix} \quad U_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad U = \left[\begin{array}{c:c} U_1 & U_2 \end{array}\right] = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

因此矩阵 A 的奇异值分解为:

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

4.4.3 利用奇异值分解求矩阵广义逆

矩阵的奇异值分解的一个重要的应用就是求矩阵的广义逆矩阵。下面先介绍广义逆矩阵的概念：设矩阵 $A \in C^{m\times n}$，若矩阵 $X \in C^{n\times m}$ 对于以下四个 Penrose 方程:

$$\begin{aligned} &(1)\ AXA = A \\ &(2)\ XAX = X \\ &(3)\ (AX)^H = AX \\ &(4)\ (XA)^H = XA \end{aligned}$$

如果矩阵 X 满足以上四个方程中的一个或几个，则称其为矩阵 A 的广义逆；如果四个方程全部满足，则称其为矩阵 A 的 Moore-Penrose 逆，记为 A^+，且这种广义逆矩阵是唯一的，它具有以下性质:

$$\begin{aligned} &(1)\ \mathrm{rank}A^+ = \mathrm{rank}A \\ &(2)\ (A^+)^+ = A \\ &(3)\ (A^H)^+ = (A^+)^H, (A^T)^+ = (A^+)^T \\ &(4)\ (A^HA)^+ = A^+(A^H)^+, (AA^H)^+ = (A^H)^+A^+ \\ &(5)\ A^+ = (A^HA)^+A^H = A^H(AA^H)^+ \\ &(6)\ R(A^+) = R(A^H), N(A^+) = N(A^H) \end{aligned}$$

最后需要指出，当矩阵 A 为满秩方阵的时候，则有 $A^+ = A^{-1}$。

可以使用矩阵的奇异值分解来求矩阵的 Moore-Penrose 逆，这由下面的定理保证：设 $A \in C_r^{m\times n}$ 的奇异值分解为 $A = U \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix}_{m\times n} V^H$，则有 $A^+ = V \begin{bmatrix} \Sigma_r^{-1} & O \\ O & O \end{bmatrix}_{m\times n} U^H$。这条定理的证明是容易的，只需要说明此处构造出的矩阵 A^+ 满足上述四个 Penrose 方程即可。由于应用这条定理求矩阵的 Moore-Penrose 逆的操作是十分显然的，故此处不再单独举例展示该定理的具体使用步骤。

5 总结

经过了近两周的学习和写作，我终于完成了《矩阵函数的求法和矩阵分解方法研究》的论文。从看到题目、到复习书本知识和课堂 PPT、再到最后的动笔写作，这其中的每一步对我来说都是新的尝试与挑战，这是我在大学的第一篇真正有些长度和难度的文章。

在本文中，我先介绍了本论文的背景以及主要书写内容；然后叙述了矩阵理论的必备基础知识以及一般的分析思路和方法；然后就是矩阵函数的求法研究，从矩阵函数的定义出发，研究并总结常见的四种矩阵函数的求法，并在每种求法的理论推导之后给出一个具体的例子加以说明；之后讨论了矩阵的分解方法，从矩阵分解的重要意义出发，说明矩阵的必要性和重要性，研究并总结常见的四种矩阵分解的方法，并在每种方法的理论推导之后给出一个具体的例子加以说明。

在本文的书写过程中，一开始我想要使用 Word 进行写作，但是后来由于排版、公式编辑等诸多不方便的问题，我开始接触到 LaTex 排版系统，于是我决定使用它来进行论文的写作。这虽然增加了写作的成本，但是最终呈现出来的成果是让我颇为满意的：整篇文章的排版一致且优美，且公式部分由于全部用源代码写成，故风格一致且方便修改，可以说，使用 LaTex 增加的学习成本和时间成本和最终呈现出来的结果相比是十分值得的。

在复习这个学期所学的矩阵理论知识的时候，我也完整地梳理了这学期所学的知识并掌握了整个知识脉络，这对我书写本论文有着极大的帮助，使得整篇论文最后呈现为一个知识的整体而非知识的片段。我还时刻注意回顾线性代数课程中的所学知识，尽量将矩阵课程作为线性代数课程的更一般化、更深入的课程而不是一门独立的、毫无联系的课程，这有助于我理解矩阵理论中的诸多抽象概念，矩阵理论中的抽象概念往往可以在线性代数课程中找到对应的比较具体的概念。从我的理解来看，线性代数课程初步介绍了变换与转化，而矩阵理论则更为深入地讲解了抽象空间中变换与转化的一般法则，只有学习完矩阵理论，才能完整地完成从具体空间到抽线空间的思维转变，我认为这将对我后续的学习起到十分重要的作用。

总的来说，这次论文的写作使我受益匪浅，不仅仅加深了我对矩阵理论课程的理解，而且还锻炼了诸多与论文写作相关的能力。短暂的矩阵理论与方法课程已经结束，本论文的写作也即将结束，最后我想感谢本门课程的李昊辰老师。李老师课前精心备课，在课上耐心讲授，并且不囿于课本知识，给我们介绍了许多不在课本上、但是十分有用的内容，这让我印象深刻。

除了《矩阵论》课本与习题和课堂 PPT 以外，我在必要的时候还参考了《Matrix Computations》一书，该书在某些方面给出了更为深刻的论述，增加了我对本门课程中一些概念的理解。此外，我还参考了一些网络资料。所有的参考文献将在本论文的最后列出。

6 参考文献

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

[2] 张凯院, 徐仲. 矩阵论导教-导学-导考 [M]. 西安: 西北工业大学出版社,2014.

[3]Gene H. Golub, Charles F. Van Loan, *Matrix Computations*[M], Baltimore: The Johns Hopkins University Press, 2013.

[4]Wikipedia. 矩陣理論 [DB/OL].https://zh.m.wikipedia.org/wiki/矩陣理論.

[5] 百度百科. 欧几里得空间 [DB/OL].https://baike.baidu.com/item/欧几里得空间.

[6] 刘芝秀, 黄小杰, 金本清, 谢杰华. 留数定理在计算矩阵函数值中的应用 [J]. 江西理工大学学报,2013.05.001.

[7] 陈文冰. 几类矩阵分解及其应用 [D]. 天津：天津工业大学，2018.

[8] 赖淑珍. 非负矩阵分解若干算法研究与应用 [D]. 成都：电子科技大学，2014.