北京郵電大學

期末课程论文

题目: 矩阵函数的求法和矩阵分解方法研究

姓　　名: 胡宇杭

学　　院: 计算机学院 (国家示范性软件学院)

专　　业: 计算机类

班　　级: 2022211320

学　　号: 2022212408

指导教师: 李昊辰

2024 年 1 月 7 日

矩阵函数的求法和矩阵分解方法研究

摘要

本论文的主要内容为基于现有的矩阵函数求法和矩阵分解方法，对相关技术进行了全面汇总，并提出了一套综合性的解决方案。重点探讨了不同矩阵函数求解和分解方法的理论基础及应用实践，同时对比了它们的优势和局限性。此外，本文还深入分析了这些方法在计算机程序上的实现和优化过程，展示了如何利用这些方法解决实际问题。这些研究成果为矩阵理论及其在科学计算中的应用提供了宝贵的参考。

关键词　矩阵函数求法　矩阵分解方法　程序实现

Research on Matrix Function Calculation and Matrix Decomposition Methods

Abstract

The main content of this paper is a comprehensive summary of existing matrix function calculation and matrix decomposition techniques, proposing an integrated solution. It focuses on the theoretical foundations and practical applications of different matrix function solutions and decomposition methods, while comparing their advantages and limitations. Additionally, the paper delves into the implementation and optimization of these methods in computer programs, demonstrating how to use them to solve real-world problems. These research findings offer valuable references for matrix theory and its application in scientific computing.

KEYWORDS Matrix Function Calculation Matrix Decomposition Methods Program implementation

目录

第 1 章 引言

1.1 背景介绍

1.1.1 矩阵理论与方法介绍

作为数学的一个重要分支，矩阵理论具有极为丰富的内容. 其由矩阵运算、特征值和特征向量、线性方程组、矩阵分解、矩阵微积分等方面组成.

矩阵理论的历史可以追溯到 18 世纪和 19 世纪的数学发展，但它的真正崭露头角是在 20 世纪早期。矩阵在矩阵代数、线性代数和线性空间中的研究变得非常重要。著名的数学家如 Sylvester、Cayley 和 Frobenius 等都为矩阵理论的发展作出了重要贡献。矩阵的应用范围迅速扩展，尤其是在工程、物理学、计算机科学和统计学等领域。

作为一种基本工具，矩阵理论在数学学科以及其他科学技术领域都有非常广泛的应用，包括但不限于:

1. **线性代数和线性方程组求解:** 矩阵用于表示和求解线性方程组，如高斯消元法和矩阵的逆运算等;
2. **统计学:** 协方差矩阵和相关矩阵用于描述和分析随机变量之间的关系;
3. **计算机图形学:** 矩阵用于实现图像变换、旋转和投影，以及三维图形的转换;
4. **机器学习和数据分析:** 矩阵用于数据降维、特征选择和各种机器学习算法的实现。

1.1.2 函数矩阵和矩阵函数介绍

函数矩阵: 函数矩阵是一个矩阵，其元素不是常数，而是函数。通常，一个函数矩阵会包含一个矩阵中的每个元素都是一个函数，这些函数可能依赖于一个或多个变量。函数矩阵可以表示为:

$$A(t) = \begin{bmatrix} a_{11}(t) & a_{12}(t) & \dots & a_{1n}(t) \\ a_{21}(t) & a_{22}(t) & \dots & a_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}(t) & a_{m2}(t) & \dots & a_{mn}(t) \end{bmatrix}$$

如果其中每一个元素都是变量 t 的可导函数，则称 $A(t)$ 可导，其导数定义为

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

函数矩阵也有与普通函数相似的性质:

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

这里，$a=a(t)$ 为 t 的可导函数.

总结一下，函数矩阵是一个矩阵，其元素是函数，而矩阵函数是一个函数，其输入和输出都是矩阵。这两个概念都在数学和工程领域中有广泛的应用，用于描述和处理复杂的多维数据和变量.

矩阵函数:

矩阵函数是以矩阵为自变量且取值为矩阵的一类函数，它是对一元函数概念的推广. 起先，矩阵函数是由一个收敛的矩阵幂级数的和来定义，之后根据计算矩阵函数值的 Jordan 标准形方法又对矩阵函数的概念进行了拓宽.

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 $\sum_{k=0}^{\infty}c_kA^k$ 的和称为矩阵函数，记为 $f(A)$，即

$$f(A)=\sum_{k=0}^{\infty}c_kA^k \tag{3.3.2}$$

矩阵函数在多个领域中具有广泛的应用，包括但不限于数值分析、微分方程求解、量子力学、机器学习和数据分析、信号处理等方面.

总之，矩阵函数是数学和工程领域中的一个关键工具，用于处理和分析多维数据和系统. 它们在各种应用中发挥着重要作用，从数值计算到科学研究和工程设计等领域都有广泛的应用. 理解矩阵函数有助于解决复杂的多维问题，并优化系统的设计和性能.

1.1.3 线性代数方程组求解

线性代数方程组求解的历史可以追溯到古希腊时期，但真正的突破发生在 17 世纪. 其中最早的方法之一是高斯消元法，由卡尔 · 弗里德里希 · 高斯于 1799 年提出。该方

法通过行变换将线性方程组转化为上三角形式，从而容易求解. 后来，利普希茨和克拉默分别独立提出了克拉默法则和利普希茨条件，这些方法也用于线性方程组的求解.

随着计算机科学的发展，各种数值方法和算法被开发出来，用于求解大规模和高维度的线性方程组。这些方法包括迭代法（如雅可比迭代、高斯-赛德尔迭代和共轭梯度法）、直接法（如 LU 分解和 QR 分解）以及特征值分解等.

线性代数方程组求解的方法可以大致分为以下几类:

1. **直接法:** 这些方法直接找到方程组的解，无需迭代。常见的直接法包括:
   - **高斯消元法:** 通过行变换将方程组转化为上三角形式，然后回代求解;
   - **LU 分解:** 将系数矩阵分解为下三角矩阵和上三角矩阵的乘积，然后通过前代和回代求解;
   - **QR 分解:** 将系数矩阵分解为正交矩阵和上三角矩阵的乘积，然后通过回代求解.
2. **迭代法:** 这些方法通过迭代逐步逼近方程组的解。常见的迭代法包括:
   - **雅可比迭代法:** 通过逐个更新解的各个分量来逼近精确解;
   - **高斯-赛德尔迭代法:** 与雅可比迭代法类似，但每次更新使用最新计算的分量;
   - **共轭梯度法:** 用于解决对称正定系数矩阵的线性方程组，通常用于优化问题.
3. **特征值分解:** 特征值和特征向量的方法可以用于求解对称或正定系数矩阵的线性方程组. 这包括使用 Cholesky 分解和对称幂法等方法;
4. **数值软件:** 许多数学库和软件包（如 NumPy、SciPy、MATLAB 等）提供了现成的线性代数方程组求解工具，使求解更加方便和高效.

线性代数方程组求解是数学和工程领域的一个重要问题，有着悠久的历史和多种方法. 选择合适的方法通常取决于问题的特点，例如系数矩阵的性质、方程组的规模和精度要求. 不同的方法都有各自的优点和局限性，熟练地选择和应用这些方法可以帮助有效地解决线性方程组问题.

1.2 问题介绍

1.2.1 矩阵函数的求法问题介绍

求解矩阵函数的方法取决于函数的类型和矩阵的特性. 常见的矩阵函数包括指数函数、对数函数、幂函数等.

求解矩阵函数主要的问题包括:

1. **定义和存在性：** 不是所有的数学函数都可以直接应用于矩阵. 首先需要确定函数对于给定的矩阵是否有意义，以及如何适当地定义这个函数;
2. **计算方法：** 矩阵函数的计算通常比简单的标量函数复杂得多. 需要选择合适的计算方法，如泰勒级数展开、特征分解、Jordan 形式或者积分表示等，这取决于矩阵的性质和函数的类型;
3. **数值稳定性：** 在计算矩阵函数时，保持数值稳定性是一个重要的问题. 某些方法可能在数值上不稳定，尤其是当矩阵的条件数较大时.

下面是几种常见的矩阵函数求解方法：

1. **泰勒级数：** 对于一些函数，比如指数函数和对数函数，可以通过将函数展开为泰勒级数来计算其在矩阵上的值;
2. **特征值分解：** 如果矩阵可对角化，即 $A = PDP^{-1}$，其中 D 是对角矩阵，那么 $f(A)$ 可以通过先对 D 中的每个特征值应用函数 f，然后重组矩阵来求得;
3. **Jordan 形式：** 对于不能对角化的矩阵，可以使用 Jordan 分解方法。这种方法涉及将矩阵转换为 Jordan 标准形式，然后在这种形式上应用函数;
4. **积分表示法：** 对于某些函数，可以通过积分的方式来定义矩阵函数.

1.2.2 矩阵分解的方法问题介绍

矩阵分解，也称为矩阵因式分解，是将一个矩阵分解为几个具有特定特性的矩阵乘积的过程。这种分解在数值分析、信号处理、统计学、机器学习等领域中有广泛的应用，因为它有助于简化复杂的矩阵运算，揭示数据的潜在结构，以及提高算法的效率和稳定性.

矩阵分解的主要问题和挑战通常包括：

1. **计算复杂度：** 矩阵分解通常涉及大量的计算，尤其是当处理大规模数据时. 因此，计算效率和优化是核心考虑因素;
2. **数值稳定性：** 在进行矩阵分解时，数值稳定性是一个重要问题. 某些分解方法可能在数值上不稳定，尤其是在处理条件数不佳的矩阵时;
3. **适用性和限制：** 不同的矩阵分解方法适用于不同类型的矩阵和问题. 例如，特征分解只适用于方阵，而奇异值分解（SVD）适用于任何形状的矩阵;

下面是几种常见的矩阵分解方法：

1. **LU 分解：** 将矩阵分解为一个下三角矩阵（L）和一个上三角矩阵（U）。这种分解常用于解线性方程组、计算行列式和矩阵的逆;

2. **QR 分解：**将矩阵分解为一个正交矩阵（Q）和一个上三角矩阵（R）。QR 分解在求解线性最小二乘问题和计算特征值等问题中非常有用;
3. **奇异值分解（SVD）：**将矩阵分解为三个矩阵的乘积，形式为 $U\Sigma V^H$. SVD 在信号处理、统计学和机器学习中非常有用，特别是在主成分分析（PCA）和低秩近似问题中;
4. **特征分解：**也称为谱分解，是将矩阵分解为其特征向量和特征值的乘积. 特征分解在理解线性变换的本质方面很有帮助，但只适用于方阵.

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

矩阵函数求解的主要聚焦方向包括数值方法改进、并行计算和分布式计算、高性能计算、自动化工具和软件的开发等方向. 具体包括：

1. **矩阵指数函数：**矩阵指数函数是一种常见的矩阵函数，它经常出现在线性常微分方程、控制理论和量子力学等领域中. 目前有多种数值方法用于计算矩阵指数函数，包括 Pade 逼近、泰勒级数展开、Krylov 子空间方法等。
2. **矩阵对数函数：**矩阵对数函数在矩阵计算中也是重要的. 它的计算通常涉及到特征值分解、Schur 分解或极化恒等式等技术;
3. **矩阵函数的插值方法：**近年来，研究人员提出了基于插值的方法，用于计算矩阵函数的近似值. 这些方法利用一系列已知的矩阵值来逼近矩阵函数，然后进行插值以获得更精确的近似;
4. **高性能计算和并行计算：**随着计算机性能的提高，矩阵函数的求解也得到了显著的改进。并行计算和分布式计算技术被广泛应用于加速矩阵函数的计算.

1.3.2 矩阵分解方法研究现状

矩阵分解的主要聚焦方向包括提高经典分解方法的效率和稳健性，探索高阶矩阵和张量分解方法，以及将深度学习技术与矩阵分解方法相结合等，具体如下：

1. **奇异值分解（SVD）：**奇异值分解是一种经典的矩阵分解方法，用于将矩阵分解为三个矩阵的乘积，通常应用于数据降维、主成分分析（PCA）和推荐系统等领域. 研究人员一直在寻找更高效和稀疏的 SVD 算法，以处理大规模数据;
2. **非负矩阵分解（NMF）：**非负矩阵分解是用于数据聚类和特征提取的方法，它约

束了分解出的矩阵因子为非负值; 近年来，研究者提出了各种改进的 NMF 算法，包括多任务 NMF、稀疏 NMF 等;

3. **Tucker 分解和高阶矩阵分解：** 对于高阶张量和多维数据，Tucker 分解和高阶矩阵分解方法变得越来越重要。这些方法有助于提取数据中的高级结构和模式;
4. **稀疏矩阵分解：** 稀疏矩阵分解方法用于处理大规模和稀疏数据。研究者关注如何设计更加高效的稀疏分解算法;
5. **深度学习方法：** 深度学习方法如神经网络和自动编码器也可以看作是矩阵分解方法的一种，研究人员不断探索如何将深度学习与传统的矩阵分解方法相结合，以获得更好的性能.

1.4 本论文工作简介

1.4.1 本论文对上述问题的研究简述

本文首先对矩阵理论与方法进行了概述，强调了矩阵在数学和工程领域中的重要性. 随后，我们提出了线性代数方程组的求解作为研究的第一个基础，并强调了其在解决实际问题中的关键作用. 接着，我们介绍了矩阵函数、函数矩阵和矩阵分解，这些概念为进一步的矩阵分析奠定了基础. 并在各部分给出了详细的代码实现.

1.4.2 本论文创新点或特点简述

本论文由浅入深、由面及点地提供了一套详尽的矩阵函数的求解方法和矩阵分解方法的推导过程和使用方法，同时提供了部分代码实现，契合了当今时代主题.

1.4.3 本论文撰写结构简述

本文围绕矩阵理论中的矩阵函数和矩阵分解的求解为核心展开:

1. **第一章：** 详细介绍了欧氏空间、线性变换、向量范数、矩阵范数、矩阵函数等概念，为下一步对各种求解方法提供理论基础;
2. **第二章：** 从待定系数法、数项级数求和法、对角型法、若尔当标准型法开始介绍了现有的矩阵函数的求法及其推导过程和使用示例;
3. **第三章：** 深入探讨矩阵分解的方法，提供了详细地推导过程和代码实现.

第 2 章 预备知识

2.1 欧氏空间与线性变换

2.1.1 欧氏空间与线性变换介绍

线性空间是线性代数最基本的概念之一，也是学习现代矩阵论的重要基础，现对线性空间做如下定义：

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

1. 在 V 中定义一个加法运算，即当 $\boldsymbol{x}, \boldsymbol{y} \in V$ 时，有唯一的和 $\boldsymbol{x}+\boldsymbol{y} \in V$，且加法运算满足以下性质：
   - 结合律 $\boldsymbol{x}+(\boldsymbol{y}+\boldsymbol{z})=(\boldsymbol{x}+\boldsymbol{y})+\boldsymbol{z}$;
   - 交换律 $\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{y}+\boldsymbol{x}$;
   - 存在 **零元素** 0，使 $\boldsymbol{x}+0=\boldsymbol{x}$;
   - 存在 **复元素**，即对任一向量 $\boldsymbol{x} \in V$，存在向量 $\boldsymbol{y} \in V$，使 $\boldsymbol{x}+\boldsymbol{y}=0$，则称 $\boldsymbol{y}$ 为 $\boldsymbol{x}$ 的负元素，记为 $-\boldsymbol{x}$.
2. 在 V 中定义数乘运算，即当 $\boldsymbol{x} \in V$，$k \in K$ 时，有唯一的乘积 $k\boldsymbol{x} \in V$，且数乘运算满足以下性质：
   - 数因子分配律 $k(\boldsymbol{x}+\boldsymbol{y})=k\boldsymbol{x}+k\boldsymbol{y}$;
   - 分配律 $(k+l)\boldsymbol{x}=k\boldsymbol{x}+l\boldsymbol{x}$;
   - 结合律 $k(l\boldsymbol{x})=(kl)\boldsymbol{x}$;
   - $1\ \boldsymbol{x}=\boldsymbol{x}$.

则称 V 为数域 K 上的 **线性空间**或 **向量空间**

有了线性空间的定义后，我们可以继续定义变换。

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

线性变换描述了将一个向量空间映射到另一个向量空间的线性操作，接下来我们通过前两个定义对其进行描述.

线性变换： 如果数域 K 上的线性空间 V 的一个变换 T 具有以下性质：

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

其中 $\boldsymbol{x}, \boldsymbol{y}\in V$，$k, l\in K$. 则称 T 为 V 的一个 **线性变换**或 **线性算子**，也就是变换 T 对向量的线性运算是封闭的.

即使在引入线性变换后，在线性空间中向量的基本运算也仅是线性运算，不涉及长度、夹角等度量概念，故我们需要通过内积（类似于数量积）对这些量进行描述。因此我们可以在原本的线性空间上定义内积，即 **欧氏空间**.

欧氏空间： 设 V 是实数域 R 上的线性空间，对于 V 中任意两个向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$，按照某种规律定义一个实数，用 $(\boldsymbol{x}, \boldsymbol{y})$ 来表示，且他满足下述 4 个条件：

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

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

2.1.2 若尔当标准形的求解

在介绍 **Jordan** 标准形之前，我们需要了解以下概念：

虽然我们在 2.1.1 中定义了向量内积的概念，但却没有定义向量的坐标，因此，我们在此定义线性空间的基（类似于坐标系）：

线性空间的基： 设 V 是数域 K 上的线性空间，$\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r \ (r\geq 1)$ 是属于 V 的任意 r 个向量，如果它满足：

1. $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r$ 线性无关;
2. V 中任一向量 $\boldsymbol{x}$ 都是 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r$ 的线性组合，或者说都可以被其线性表出.

则称 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r$ 为 V 的一个 **基**或 **基底**，并称 $\boldsymbol{x}_i \ (i=1, 2, \ldots, r)$ 为 **基向量**.

在有了 V^n 的基 $x_1, x_2, \ldots, x_n$ 后，我们可以任意向量 $\boldsymbol{x}\in V^n$ 表示为如下形式：

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

则向量 $\boldsymbol{x}$ 的坐标即为 $\xi_1,\ \xi_2,\ \ldots,\ \xi_n$，记为：

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

子空间：设 V_1 是数域 K 上的线性空间 V 的一个非空子集合，且对 V 已有的线性运算满足以下条件：

1. 如果 $\boldsymbol{x},\ \boldsymbol{y} \in V_1$，则 $\boldsymbol{x}+\boldsymbol{y} \in V_1$;
2. 如果 $\boldsymbol{x} \in V_1,\ k \in K$, 则 $k\boldsymbol{x} \in V_1$.

也就是运算的封闭性，则称 V_1 为 V 的 **线性子空间**或 **子空间**.

相似的，如果 T 是线性空间 V 的线性变换，V_1 是 V 的子空间，对于任一 $\boldsymbol{x} \in V_1$，都有 $T\boldsymbol{x} \in V_1$，则称 V_1 是 T 的 **不变子空间**.

核空间：设 $A \in R^{m\times n}$，称集合 $\{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x}=0\}$ 为 $\boldsymbol{A}$ 的 **核空间**，记为 $N(\boldsymbol{A})$，即：

$$N(\boldsymbol{A}) = \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x}=0\}$$

核子空间：对于线性空间 V 的线性变换 T，如果 $N(T)$ 是 V 的线性子空间，则称为 T 的 **核子空间**

我们知道，一切 n 阶矩阵 $\boldsymbol{A}$ 可以分成许多相似类，今要在与 $\boldsymbol{A}$ 相似的全体矩阵中，找出一个较简单的矩阵来作为这个相似类的标准形. 当然以对角矩阵作为标准形最好，可惜不是每一个矩阵都能与对角矩阵相似. 为了解决标准形问题，我们引入 **Jordan** 标准形的概念：

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

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

则 V^n 可分解成不变子空间的直和:

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

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

为每个子空间 V_i 选一适当的基，则每个子空间的基合并起来即为 V^n 的基，且 T

在该基下的矩阵为准对角矩阵:

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

其中

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

则称矩阵 $\boldsymbol{J}$ 为矩阵 $\boldsymbol{A}$ 的 **Jordan** 标准形, $\boldsymbol{J}_i(\lambda_i)$ 称为因式 $(\lambda-\lambda_i)^{m_i}$ 对应的 **Jordan 块**.

在了解了 **Jordan 标准型**的概念后，我们可以总结出在复数域 $\mathbf{C}$ 上，求 n 阶矩阵 $\boldsymbol{A}_{n\times n}$ 的 Jordan 标准形:

1. 求特征矩阵的 $\lambda\boldsymbol{I}-\boldsymbol{A}$ 的初等因子组 $(\lambda-\lambda_i)^{m_i}\quad(i=1,\ 2,\ \ldots,s,\ m_1+m_2+\cdots+m_s=n)$
   (a) 定义 λ 矩阵，令 $\boldsymbol{A}(\lambda)=\lambda\boldsymbol{I}-\boldsymbol{A}$
   (b) 求 $\boldsymbol{A}(\lambda)$ 的不变因子，即将 $\boldsymbol{A}(\lambda)$ 转化为标准型后其对角线上的非零元素 $d_i(\lambda)\ \ (i=1,\ 2,\ \ldots,\ s)$
   (c) 求 $\boldsymbol{A}(\lambda)$ 的初等因子，即将其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，这些不可约因式即为初等因子
   (d) $\boldsymbol{A}(\lambda)$ 的所有初等因子，称为 $\boldsymbol{A}(\lambda)$ 的初等因子组
2. 写出每个初等因子 $(\lambda-\lambda_i)^{m_i}\ \ (i=1,\ 2,\ \ldots,\ s)$ 对应的 Jordan 块
3. 写出以这些 Jordan 块构成的 Jordan 标准型

现以矩阵论第五版 1.2 节课后习题 19 (1) 作为样例对求解过程进行演示:

题目.

求矩阵 $\begin{bmatrix}3 & 7 & -3\\ -2 & -5 & 2\\ -4 & -10 & 3\end{bmatrix}$ 的 Jordan 标准形

解答.

1. 求特征矩阵的 $\lambda\boldsymbol{I}-\boldsymbol{A}$ 的初等因子组 $(\lambda-\lambda_i)^{m_i}$

   (a) 定义 λ 矩阵

$$\lambda\boldsymbol{I}-\boldsymbol{A}=\begin{bmatrix}\lambda-3 & -7 & 3\\ 2 & \lambda+5 & -2\\ 4 & 10 & \lambda-3\end{bmatrix}$$

   (b) 求 $\boldsymbol{A}(\lambda)$ 的不变因子

$$\lambda\boldsymbol{I}-\boldsymbol{A}=\begin{bmatrix}\lambda-3 & -7 & 3\\ 2 & \lambda+5 & -2\\ 4 & 10 & \lambda-3\end{bmatrix}\to\begin{bmatrix}\lambda & 2 & 3\\ 0 & \lambda-1 & -2\\ \lambda+1 & 3\lambda+1 & \lambda-3\end{bmatrix}\to\begin{bmatrix}\lambda & 2 & 3\\ 0 & \lambda-1 & -2\\ 1 & 2 & \lambda\end{bmatrix}$$

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

   故 $\boldsymbol{A}(\lambda)$ 的不变因子为 $d_1(\lambda)=1,\ d_2(\lambda)=\lambda-1,\ d_3(\lambda)=1+\lambda^2=(\lambda-i)(\lambda+i)$

   (c) 求 $\boldsymbol{A}(\lambda)$ 的初等因子、初等因子组

   易知，初等因子组为 $\lambda-1,\ \lambda+i,\ \lambda-i$

2. 写出每个初等因子对应的 Jordan 块

$$\boldsymbol{J}_1(\lambda_1)=\begin{bmatrix}1\end{bmatrix},\ \boldsymbol{J}_2(\lambda_2)=\begin{bmatrix}i\end{bmatrix},\ \boldsymbol{J}_3(\lambda_3)=\begin{bmatrix}-i\end{bmatrix}$$

3. 写出其构成的 Jordan 标准型

   于是有

$$\boldsymbol{A}\sim\boldsymbol{J}=\begin{bmatrix}\boldsymbol{J}_1(\lambda_1) & & \\ & \boldsymbol{J}_2(\lambda_2) & \\ & & \boldsymbol{J}_3(\lambda_3)\end{bmatrix}=\begin{bmatrix}1 & & \\ & i & \\ & & -i\end{bmatrix}$$

2.1.3 欧氏空间中线性变换的求法

在求解欧氏空间中的线性变换之前，我们需要先知道一些基础概念及其求法

正交向量组： 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**，易知这组向量之间线性无关.

标准正交基： 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**. 对于 V^n 中任一基 $x_1, x_2, \ldots, x_n$，都可找到一个标准正交基 $y_1, y_2, \ldots, y_n$. 换言之，任一非零欧氏空间都有正交基和标准正交基，且可以将任一基转化为标准正交基.

Schimidt 正交化: 令 $y_1' = x_1$, 则有 $y_{m+1}' = x_{m+1} + l_m y_m' + l_{m-1} y_{m-1}' + \cdots + l_2 y_2' + y_1'$, 其中 $l_i = -\frac{(x_{m+1},\ y_i')}{(y_i' \cdot y_i')}$ $(i = 1, 2, \ldots, m)$. 我们可以通过递推的方式求得正交基 $(y_1', y_2', \ldots, y_n')$, 然后以 $|y_i'|$ 除 y_i' $(i = 1, 2, \ldots, n)$, 得到标准正交基 $(y_1, y_2, \ldots, y_n)$, 其中 $y_i = \frac{y_i'}{|y_i'|}$.

在了解了上述内容后，我们可以得出在欧氏空间中线性变换的系统性求解方法:

设 V 是欧氏空间，T 是 V 上的一个线性变换，求解 $\boldsymbol{z} = (\boldsymbol{T}^k)(\boldsymbol{x}),\ \boldsymbol{x} \in V$ 的方法如下:

1. 任意找一组基，利用 **Schimidt** 正交化方法得到 V 的一组标准正交基 $e_1, \ldots, e_n$, $\boldsymbol{x} = k_1 e_1 + k_2 e_2 + \cdots + k_n e_n$，其中 $k_i = (\boldsymbol{x}, e_i)$
   (a) 求 T 在基 $e_1, \ldots, e_n$ 下的矩阵 $\boldsymbol{A}_0$，即 $\boldsymbol{T}(e_1, \ldots, e_n) = (e_1, \ldots, e_n)\boldsymbol{A}_0$
   (b) 求解 Jordan 标准形和 $\boldsymbol{P}$, $\boldsymbol{P}^{-1}$. 由于 $\boldsymbol{A}_0 = \boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1}$，有 $\boldsymbol{T}(e_1, \ldots, e_n) = (e_1, \ldots, e_n)\boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1}$
2. 变换得到 $\boldsymbol{T}(e_1, \ldots, e_n)\boldsymbol{P} = (e_1, \ldots, e_n)\boldsymbol{P}\boldsymbol{J}$，可以得到一组新的基 $(E_1, \ldots, E_n) = (e_1, \ldots, e_n)\boldsymbol{P}$, 则 T 在新基下的矩阵的矩阵为 $\boldsymbol{T}(E_1, \ldots, E_n) = (E_1, \ldots, E_n)\boldsymbol{J}$
3. 通过坐标变换得到 $\boldsymbol{x} = (E_1, \ldots, E_n)\boldsymbol{P}^{-1}\begin{bmatrix} k_1 \\ \vdots \\ k_n \end{bmatrix} = (E_1, \ldots, E_n)\begin{bmatrix} l_1 \\ \vdots \\ l_n \end{bmatrix}$
4. 则变换 T 可以表示为 $\boldsymbol{T}(x) = (E_1, \ldots, E_n)\boldsymbol{J}\begin{bmatrix} l_1 \\ \vdots \\ l_n \end{bmatrix}$，此时 $(\boldsymbol{T}^k)(x) = (E_1, \ldots, E_n)\boldsymbol{J}^k\begin{bmatrix} l_1 \\ \vdots \\ l_n \end{bmatrix}$

接下来，我们通过例题来演示上述步骤:

题目.

设矩阵空间 $R^{2\times2}$ 的子空间为 $V = \{X = (x_{ij})_{2\times2} \mid x_{11} + x_{12} + x_{21} = 0,\ \ x_{ij} \in R\}$,

V 上的线性变换 $\boldsymbol{T}(X) = X + 2X^T$，求 $(\boldsymbol{T}^k)(X)$，$\forall X \in V$.

解答.

1. 任意找一组基，利用 **Schimidt** 正交化方法得到 V 的一组标准正交基 e_1，$\ldots, e_n$，
   $\boldsymbol{x} = k_1e_1 + k_2e_2 + \cdots + k_ne_n$，其中 $k_i = (\boldsymbol{x}, e_i)$
   令 $x_{11} = -x_{12} - x_{21}$，则

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

定义 V 的内积为 $(X, Y) = tr(XY_T) = (x_{12}+x_{21})(y_{12}+y_{21}+x_{12}y_{12}+x_{21}y_{21}+x_{22}y_{22})$

任意找一组基

$$\begin{aligned} X &= \begin{bmatrix} -x_{12} - x_{21} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} = x_{12}\begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} + x_{21}\begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix} + x_{22}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\ &= x_{12}X_1 + x_{21}X_2 + x_{22}X_3 \end{aligned}$$

下面利用 **Schimidt** 正交化求解标准正交基

$$Y_1' = X_1 = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}$$

$$Y_2' = X_2 - \frac{(X_2, Y_1')}{Y_1', Y_1'}Y_1' = \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} \\ 1 & 0 \end{bmatrix}$$

$$Y_3' = X_3 - \frac{(X_3, Y_2')}{Y_2', Y_2'}Y_2' - \frac{(X_3, Y_1')}{Y_1', Y_1'}Y_1' = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} - \frac{0}{\frac{3}{2}}\begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} \\ 1 & 0 \end{bmatrix} - \frac{0}{2}\begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

得到 V 的一组正交基 Y_1', Y_2', Y_3'

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

故标准正交基 e_1，e_2，e_3 为

$$e_1 = \frac{1}{|Y_1'|}Y_1' = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}, e_2 = \frac{1}{|Y_2'|}Y_2' = \frac{\sqrt{2}}{\sqrt{3}}\begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} \\ 1 & 0 \end{bmatrix}, e_3 = \frac{1}{|Y_3'|}Y_3' = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\boldsymbol{x} = \begin{bmatrix} 4 & -4 \\ 0 & -3 \end{bmatrix} = \begin{bmatrix} e_1, e_2, e_3 \end{bmatrix}\begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix}, k_1 = (\boldsymbol{x}, e_1) = -4\sqrt{2}, k_2 = (\boldsymbol{x}, e_2) = 0, k_3 = (\boldsymbol{x}, e_3) = -3$$

(a) 求 T 在基 e_1, …, e_n 下的矩阵 $\boldsymbol{A}_0$

$$Te_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} -3 & 1 \\ 2 & 0 \end{bmatrix}, Te_2 = \frac{\sqrt{2}}{\sqrt{3}}\begin{bmatrix} -\frac{3}{2} & -\frac{3}{2} \\ 0 & 0 \end{bmatrix}, Te_3 = \begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix}$$

$$Te_1 = \begin{bmatrix} e_1, e_2, e_3 \end{bmatrix} \begin{bmatrix} 2 \\ \sqrt{3} \\ 0 \end{bmatrix}, Te_2 = \begin{bmatrix} e_1, e_2, e_3 \end{bmatrix} \begin{bmatrix} \sqrt{3} \\ 0 \\ 0 \end{bmatrix}$$

$$Te_3 = \begin{bmatrix} e_1, e_2, e_3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} \quad k_{ij} = (Te_i, e_j)$$

故有

$$T(e_1, \ldots, e_n) = (e_1, \ldots, e_n) \begin{bmatrix} 2 & \sqrt{3} & 0 \\ \sqrt{3} & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix} = (e_1, \ldots, e_n) A_0$$

(b) 求解 Jordan 标准形和 $\boldsymbol{P}, \boldsymbol{P}^{-1}$

$$\lambda I - A_0 = \begin{bmatrix} \lambda - 2 & -\sqrt{3} & 0 \\ -\sqrt{3} & \lambda & 0 \\ 0 & 0 & \lambda - 3 \end{bmatrix} \to \begin{bmatrix} -\sqrt{3} & \lambda - 2 & 0 \\ \lambda & -\sqrt{3} & 0 \\ 0 & 0 & \lambda - 3 \end{bmatrix}$$

$$\to \begin{bmatrix} -\sqrt{3} & 0 & 0 \\ \lambda & \frac{\lambda-2}{\sqrt{3}}\lambda - \sqrt{3} & 0 \\ 0 & 0 & \lambda - 3 \end{bmatrix} \to \begin{bmatrix} -\sqrt{3} & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}}(\lambda+1)(\lambda-3) & 0 \\ 0 & 0 & \lambda - 3 \end{bmatrix}$$

$$\to \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & \lambda - 3 \\ 0 & (\lambda+1)(\lambda-3) & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda - 3 & 0 \\ 0 & 0 & (\lambda+1)(\lambda-3) \end{bmatrix}$$

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

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 标准形为

$$J=\begin{bmatrix}3 & & \\ & -1 & \\ & & 3\end{bmatrix}$$

$$P=(x_1,\ x_2,\ x_3),PJ=A_0P\Rightarrow(3x_1,\ -x_2,\ 3x_3)=(A_0x_1,\ A_0x_2,\ A_0x_3)$$

$$(3I-A_0)x_1=\begin{bmatrix}1 & -\sqrt{3} & \\ -\sqrt{3} & 3 & \\ & & 0\end{bmatrix}x_1=0$$

$$(-I-A_0)x_2=\begin{bmatrix}-3 & -\sqrt{3} & \\ -\sqrt{3} & -1 & \\ & & 4\end{bmatrix}x_2=0$$

$$(3I-A_0)x_3=\begin{bmatrix}1 & -\sqrt{3} & \\ -\sqrt{3} & 3 & \\ & & 0\end{bmatrix}x_3=0$$

$$\Rightarrow x_1=(\sqrt{3},\ 1,\ 0)^T,\ x_2=(-1,\ \sqrt{3},\ 0)^T,\ x_3=(0,\ 0,\ 1)^T$$

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

2. 得到一组新的基 $(E_1,\ \ldots,\ E_n)=(e_1,\ \ldots,\ e_n)\boldsymbol{P}$

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

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

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

3. 通过坐标变换得到 $\boldsymbol{x}=(E_1,\ \ldots,\ E_n)\boldsymbol{P}^{-1}\begin{bmatrix}k_1\\ \vdots\\ k_n\end{bmatrix}=(E_1,\ \ldots,\ E_n)\begin{bmatrix}l_1\\ \vdots\\ l_n\end{bmatrix}$

$$x=\begin{bmatrix}4 & -4\\ 0 & -3\end{bmatrix}=(e_1,e_2,e_3)\begin{bmatrix}-4\sqrt{2}\\ 0\\ -3\end{bmatrix}=(E_1,E_2,E_3)P^{-1}\begin{bmatrix}-4\sqrt{2}\\ 0\\ -3\end{bmatrix}=(E_1,E_2,E_3)\begin{bmatrix}-\sqrt{6}\\ \sqrt{2}\\ -3\end{bmatrix}$$

4. 则变换 T 可以表示为 $\boldsymbol{T}(x)=(E_1,\ \ldots,\ E_n)\boldsymbol{J}\begin{bmatrix}l_1\\ \vdots\\ l_n\end{bmatrix}$

$$(T^k)(x)=(E_1,E_2,E_3)\begin{bmatrix}3^k & & \\ & (-1)^k & \\ & & 3^k\end{bmatrix}\begin{bmatrix}\frac{\sqrt{3}}{4} & \frac{1}{4} & 0\\ -\frac{1}{4} & \frac{\sqrt{3}}{4} & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}(x_1,e_1)\\ (x_2,e_2)\\ (x_3,e_3)\end{bmatrix}$$

2.2 向量范数与矩阵范数

2.2.1 向量范数介绍

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

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

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

下面介绍几种特殊的范数：

- **1-范数**在 n 维酉空间 C^n 上，复向量 $\boldsymbol{x}=(\xi_1,\ \xi_2,\ \ldots,\ \xi_n)\in C^n$，有范数 $\|\boldsymbol{x}\|=\sum_{i=1}^{n}|\xi_i|$，称为 **1-范数**，记作 $\|\boldsymbol{x}\|_1$;
- **2-范数**：在 n 维酉空间 C^n 或欧氏空间 R^n 上，复向量 $\boldsymbol{x}=(\xi_1,\ \xi_2,\ \ldots,\ \xi_n)$ 的长度 $\|\boldsymbol{x}\|=\sqrt{|\xi_1|^2+|\xi_2|^2+\cdots+|\xi_n|^2}$ 是 C^n 或 R^n 上的一种范数，称为 **2-范数**，记作 $\|\boldsymbol{x}\|_2$;

- **∞-范数**：在 n 维酉空间 C^n 上，复向量 $\boldsymbol{x} = (\xi_1, \xi_2, \ldots, \xi_n) \in C^n$，实值函数 $\|\boldsymbol{x}\| = \max\limits_i |\xi_i|$ 是一种范数，称为 **∞-范数**，记作 $\|\boldsymbol{x}\|_\infty$;
- **p-范数**：对于不小于 1 的任意实数 p 及 $\boldsymbol{x} = (\xi_1, \xi_2, \ldots, \xi_n) \in C^n$，实值函数 $\|\boldsymbol{x}\| = (\sum_{i=1}^n |\xi_i|^p)^{1/p} \quad (1 \le p < +\infty)$ 是一种范数，称为 **p-范数**或 l_p **范数**，记作 $\|\boldsymbol{x}\|_p$.

不难发现，令 **p-范数**的 $p = 1$，可以得到 $\|\boldsymbol{x}\|_1$；令 $p = 2$，可得 $\|\boldsymbol{x}\|_2$；还有 $\|\boldsymbol{x}\|_\infty = \lim_{p\to\infty} \|\boldsymbol{x}\|_p$.

2.2.2 矩阵范数介绍

矩阵范数：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 $A \ne O$ 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：$\|A + B\| \le \|A\| + \|B\| \; (B \in C^{m\times n})$;
4. 相容性：$\|AB\| \le \|A\| \, \|B\| \; (B \in C^{m\times l})$

如果满足前三条性质，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 $\|\cdot\|$，还满足第四条性质，则称 $\|A\|$ 为 A 的 **矩阵范数**.

多数情况下，矩阵范数常与向量范数混合在一起使用，而矩阵经常是作为两个线性空间上的线性映射 (变换) 出现的. 因此，考虑一些矩阵范数时，应该使它能与向量范数联系起来，即 **相容**.

相容：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

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

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

下面介绍几种常用的矩阵范数：

- **Frobenius 范数**：$C^{m\times n}$ 上的函数 $\|A\|_F = (\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2)^{1/2} = (tr(A^H A))^{1/2}$ 是矩阵范数，称为 **F-范数**，记作 $\|A\|_{m_2}$;
- **从属范数**：设 $A \in C^{m\times n}$，则函数 $\|A\| = \max\limits_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|$ 是矩阵范数，称为 **由向量范数导出的矩阵范数**，简称为 **从属范数**;

设 $A=(a_{ij})_{m\times n}\in C^{m\times n}$，$\boldsymbol{x}=(\xi_1,\ \xi_2,\ \ldots,\ \xi_n)^T\in C^n$，则从属于向量 $\boldsymbol{x}$ 的三种范数 $\|\boldsymbol{x}\|_1$，$\|\boldsymbol{x}\|_2$，$\|\boldsymbol{x}\|_\infty$ 的矩阵范数分别为：

1. **列和范数**：$\|A\|_1=\max\limits_j\sum_{i=1}^m|a_{ij}|$;
2. **谱范数**：$\|A\|_2=\sqrt{\lambda_1}$，λ_1 为 A^HA 的最大特征值;
3. **行和范数**：$|A\|_\infty=\max\limits_i\sum_{j=1}^n|a_{ij}|$.

2.2.3 矩阵可逆性条件、条件数和谱半径介绍

矩阵的可逆性条件：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

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

假如 $A=(a_{ij})_{n\times n}\in C^{n\times n}$ 的元素 a_{ij} 带有误差 δa_{ij} $(i,\ j=1,\ 2,\ \ldots,\ n)$，则其逆矩阵的近似程度可有下述摄动定理描述：

摄动定理：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 $\|A^{-1}B\|<1$，有以下结论：

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

在上述定理中,若令 $cond(A)=\|A\|\,\|A^{-1}\|$,$d_A=\|\delta A\|\,\|A\|^{-1}$,则当 $\|A^{-1}\|\,\|\delta A\|<1$ 时，由结论 2, 3 可得：

$$\|I-(I+A^{-1}\delta A)^{-1}\|\le\frac{d_Acond(A)}{1-d_Acond(A)}$$
$$\frac{\|A^{-1}-(A+\delta A)^{-1}\|}{\|A^{-1}\|}\le\frac{d_Acond(A)}{1-d_Acond(A)}$$

称 $cond(A)$ 为矩阵 A 的 **条件数**.

谱半径：设 $A\in C^{n\times n}$ 的 n 个特征值为 λ_1，λ_2，$\ldots$，λ_n，称

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

为 A 的 **谱半径**.

谱半径有如下的性质：

1. 设 $A \in C^{n\times n}$，且对 $C^{n\times n}$ 上任何一种矩阵范数 $\|\cdot\|$，都有

$$\rho(A) \le \|A\|$$

2. 设 $A \in C^{n\times n}$，对于任意的正数 ε，存在某种矩阵范数 $\|\cdot\|_M$，使得

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

2.3 矩阵函数介绍

矩阵函数是以矩阵为自变量且取值为矩阵的一类函数，它是对一元函数概念的推广. 起先，矩阵函数是由一个收敛的矩阵幂级数的和来定义，之后根据计算矩阵函数值的 Jordan 标准形方法又对矩阵函数的概念进行了拓宽. 因此，矩阵函数的基础是矩阵序列与矩阵级数.

2.3.1 矩阵序列介绍

矩阵序列：指无穷多个依次排列的同阶矩阵，记为 $\{A^{(k)}\}$

矩阵序列的敛散性：设有矩阵序列 $\{A^{(k)}\}$, 其中 $A^{(k)} = (a_{ij}^{(k)})_{m\times n}$, 当 $a_{ij}^{(k)} \to a_{ij} (k \to \infty)$ 时，称 $\{A^{(k)}\}$ 收敛，或称矩阵 $A = (a_{ij}^{(k)})_{m\times n}$ 为 $\{A^{(k)}\}$ 的极限，或称 $\{A^{(k)}\}$ 收敛于 A，记为

$$\lim_{k\to\infty} A^{(k)} = A \quad or \quad A^{(k)} \to A$$

不收敛的矩阵序列称为**发散**.

矩阵序列的有界性：如果存在常数 $M > 0$，使得对一切 k 都有

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

则称矩阵序列 $\{A^{(k)}\}$ 是 **有界**的.

收敛的矩阵序列有许多类似于数列收敛的性质：

1. 设 $A^{(k)} \to A_{m\times n}$，$B^{(k)} \to B_{m\times n}$，则

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

2. 设 $A^{(k)} \to A_{m\times n}$， $B^{(k)} \to B_{n\times l}$，则

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

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

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

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

矩阵序列收敛的充要条件：设 $A^{(k)} \in C^{m\times n}$，则

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

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

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

收敛矩阵：设 A 为方阵，且 $A^k \to O \ \ (k \to \infty)$，则称 A 为 **收敛矩阵**.

矩阵收敛的条件：

- 充要条件：$\rho(A) < 1$;
- 充分条件：存在一种矩阵范数 $\|\cdot\|$，使得 $\|A\| < 1$.

2.3.2 矩阵级数介绍

矩阵级数：对于矩阵序列 $\{A^{(k)}\}$，对其求和所形成的无穷和 $A^{(0)}+A^{(1)}+A^{(2)}+\cdots+A^{(k)}+\dots$ 称为 **矩阵级数**，记为 $\sum_{k=0}^{\infty} A^{(k)}$，则有

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

矩阵级数的敛散性：记 $S^{(N)} = \sum_{k=0}^{N} A^{(k)}$，称其为矩阵级数式 $\sum_{k=0}^{\infty} A^{(k)}$ 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 $\sum_{k=0}^{\infty} A^{(k)}$ **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.

也就是说，如果 $\sum_{k=0}^{\infty} a_{ij}^{(k)} = s_{ij} \ \ (i=1,\ 2,\ \dots,\ m;\ j=1,\ 2,\ \dots,\ n)$，则矩阵级数收敛.

绝对收敛：如果对于 $\sum_{k=0}^{\infty} a_{ij}^{(k)}$ 中的每一个数项级数，其都是绝对收敛的，则称矩阵级数式是 **绝对收敛**的.

收敛的矩阵级数有以下性质：

1. 若矩阵级数式 $\sum_{k=0}^{\infty}$ 是绝对收敛的，则它一定收敛，并且任意调换其项的顺序得到的级数还是收敛的，且和不变;
2. 矩阵级数 $\sum_{k=0}^{\infty}$ 为绝对收敛的充要条件是正项级数 $\sum_{k=0}^{\infty} \|A^{(k)}\|$ 收敛;
3. 如果 $\sum_{k=0}^{\infty} A^{(k)}$ 是收敛的（或绝对收敛）的，那么 $\sum_{k=0}^{\infty} PA^{(k)}Q$ 也是收敛（或绝对收敛）的，并且有

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

4. 设 $C^{n\times n}$ 中的两个矩阵级数

$$\begin{aligned} S_1: &\quad A^{(1)} + A^{(2)} + \cdots + A^{(k)} + \dots \\ S_2: &\quad B^{(1)} + B^{(2)} + \cdots + B^{(k)} + \dots \end{aligned}$$

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

$$\begin{aligned} S_3: \quad & A^{(1)}B^{(1)} + (A^{(1)}B^{(2)} + A^{(2)}B^{(1)}) + (A^{(1)}B^{(3)} + A^{(2)}B^{(2)+A^{(3)}B^{(1)}}) + \dots \\ & + (A^{(1)}B^{(k)} + A^{(2)}B^{(k-1)+\cdots+A^{(k)}B^{(1)}}) + \cdots = \sum_{k=1}^{\infty}(\sum_{i=1}^{k} A^{(i)}B^{(k+1-i)}) \end{aligned}$$

绝对收敛，且有和 AB

矩阵级数中，矩阵的幂级数是建立矩阵函数的理论基础，占有重要地位，因此我们单独讨论矩阵的幂级数.

矩阵的幂级数收敛的充要条件：方阵 A 的 **幂级数 (Neumann 级数)**

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

收敛的充要条件为 A 是收敛矩阵，并且在其收敛时有 $\sum_{k=0}^{\infty} A^k = (I-A)^{-1}$.

有了上述性质后，我们引出如下定理：

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

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

矩阵幂级数式的敛散性：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

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

时绝对收敛的；如果 $\rho(A) > r$，则上式是发散的.

2.3.3 矩阵函数介绍

矩阵函数：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

$$f(A) = \sum_{k=0}^{\infty} c_k A^k \tag{3.3.2}$$

代入规则：若 $f(z) = g(z)$，则 $f(A) = g(A)$

2.3.4 函数矩阵对矩阵的导数

函数矩阵的导数：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

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

有上述导数定义可以推出如下性质：

1. 设 $A(t)$, $B(t)$ 时能够进行下面运算的两个可导的函数矩阵，则有

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

这里，$a = a(t)$ 为 t 的可导函数

2. 设 n 阶矩阵 A 与 t 无关，则有

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

在定义函数矩阵对矩阵的导数之前，我们先给出函数对矩阵的导数的定义：

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

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

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

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

与函数对矩阵的导数类似，有函数矩阵对矩阵 X 的导数为

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

其中 $\frac{\partial F}{\partial \xi_{ij}}$ 为 f_{ab} $(a = 1, 2, \ldots, r; \ b = 1, 2, \ldots, s)$ 对 ξ_{ij} 的偏导数，即

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

第 3 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（首项系数为 1）多项式

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

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 λ_1, λ_2, …, λ_s，相应的重数为 r_1, r_2, …, r_s $(r_1 + r_2 + \cdots + r_s = m)$，则有

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

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

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

确定出 $r(z)$. 利用 $\psi(A) = O$，可得

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

按上述方法，在实际使用待定系数法求解函数矩阵时，我们通常遵循以下步骤:

1. 求出满足 $\psi(A) = O$, $\psi(A)|\varphi(A)$ 的首 1 多项式 (通常是最小多项式或特征多项式)
2. 构造多项式 $r(\lambda) = b_0 + b_1\lambda + \cdots + b_{m-1}\lambda^{m-1}$($m$ 为 $\psi(\lambda)$ 的最高次数)
3. 由方程 $f^{(l)}(\lambda_i) = r^{(l)}(\lambda_i)$ $(l = 0,\ 1,\ \ldots,\ r_i - 1)$($r_i$ 为 λ_i 对应的重数) 求出每一项的系数
4. 计算 $f(A) = r(A) = b_0 I + b_1 A + \cdots + b_{m-1}A^{m-1}$

下面一章节将通过《矩阵论第五版》习题 3.3.5 演示该过程.

3.1.2 举例展示求法

题目.

设 $A=\begin{bmatrix}2&1&0\\0&0&1\\0&1&0\end{bmatrix}$，求 e^A, $e^{tA}(t\in R)$, $sinA$

解答.

1. 求出满足 $\psi(A)=O$, $\psi(A)|\varphi(A)$ 的首 1 多项式 (通常是最小多项式或特征多项式)

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

$\varphi(\lambda)=det(\lambda I-A)=(\lambda-1)(\lambda+1)(\lambda-2)$，易知没有最小多项式，取 $\psi(A)=(\lambda-1)(\lambda+1)(\lambda-2)$.

2. 构造多项式 $r(\lambda)=b_0+b_1\lambda+\cdots+b_{m-1}\lambda^{m-1}$($m$ 为 $\psi(\lambda)$ 的最高次数)
取 $f(\lambda)=e^\lambda$，$r(\lambda)=a+b\lambda+c\lambda^2$，设 $f(\lambda)=\psi(\lambda)g(\lambda)+(a+b\lambda+c\lambda^2)$

3. 由方程 $f^{(l)}(\lambda_i)=r^{(l)}(\lambda_i)\quad(l=0,\ 1,\ \ldots,\ r_i-1)$($r_i$ 为 λ_i 对应的重数) 求出每一项的系数

$$\begin{cases}f(2)=e^2\\f'(2)=e^2\\f''(2)=e^2\end{cases}\quad or\quad\begin{cases}a+2b+4c=e^2\\b+4c=e^2\\c=e^2\end{cases}$$

解此方程组可得 $a=3e^2$, $b=-3e^2$, $c=e^2$. 于是 $r(\lambda)=e^2(\lambda^2-3\lambda+3)$

4. 计算 $f(A)=r(A)=b_0I+b_1A+\cdots+b_{m-1}A^{m-1}$

$$e^A=f(A)=r(A)=e^2(A^2-3A+3)=e^2\begin{bmatrix}-1&2&1\\0&1&0\\0&0&1\end{bmatrix}$$

同理，我们重复上述步骤 $3\sim4$，计算 e^{tA} 和 $sinA$

- 求解 e^{tA}：

– 由方程 $f^{(l)}(\lambda_i) = r^{(l)}(\lambda_i)$ $(l = 0,\ 1,\ \dots,\ r_i - 1)$($r_i$ 为 λ_i 对应的重数) 求出每一项的系数

$$\begin{cases} f(2) = e^{2t} \\ f'(2) = te^{2t} \\ f''(2) = (t^2+1)e^{2t} \end{cases} \quad or \quad \begin{cases} a + 2b + 4c = e^{2t} \\ b + 4c = te^{2t} \\ c = (t^2+1)e^{2t} \end{cases}$$

解此方程组可得 $a = (4t^2 - 2t + 5)e^{2t},\ b = (-4t^2 + t - 4)e^{2t},\ c = (t^2 + 1)e^{2t}$. 于是 $r(\lambda) = e^{2t}[(4t^2 - 2t + 5) + (-4t^2 + t - 4)\lambda + (t^2 + 1)\lambda^2]$

– 计算 $f(A) = r(A) = b_0 I + b_1 A + \cdots + b_{m-1}A^{m-1}$

$$\begin{aligned} e^{tA} = f(A) = r(A) &= e^{2t}[(4t^2 - 2t + 5)I + (-4t^2 + t - 4)A + (t^2 + 1)A^2] \\ &= e^{2t}\begin{bmatrix} 1 & -2t^2 + t - 2 & t^2 + 1 \\ 0 & 5t^2 - 2t + 6 & -4t^2 + t - 4 \\ 0 & -4t^2 + t - 4 & 5t^2 - 2t + 6 \end{bmatrix} \end{aligned}$$

- 求解 $sinA$:
  – 由方程 $f^{(l)}(\lambda_i) = r^{(l)}(\lambda_i)$ $(l = 0,\ 1,\ \dots,\ r_i - 1)$($r_i$ 为 λ_i 对应的重数) 求出每一项的系数

$$\begin{cases} f(2) = sin2 \\ f'(2) = cos2 \\ f''(2) = -sin2 \end{cases} \quad or \quad \begin{cases} a + 2b + 4c = sin2 \\ b + 4c = cos2 \\ c = -sin2 \end{cases}$$

解此方程组可得 $a = -3sin2 - 2cos2,\ b = 4sin2 + cos2,\ c = -sin2$. 于是 $r(\lambda) = -3sin2 - 2cos2 + (4sin2 + cos2)\lambda + (-sin2)\lambda^2$

  – 计算 $f(A) = r(A) = b_0 I + b_1 A + \cdots + b_{m-1}A^{m-1}$

$$\begin{aligned} sinA = f(A) = r(A) &= (-3sin2 - 2cos2)I + (4sin2 + cos2)A + (-sin2)A^2 \\ &= \begin{bmatrix} sin2 & sin2 + cos2 & -sin2 \\ 0 & 2sin2 - 2cos2 & 4sin2 + cos2 \\ 0 & 4sin2 + cos2 & -4sin2 - 2cos2 \end{bmatrix} \end{aligned}$$

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 $\psi(\lambda)$

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

满足 $\psi(A)=O$，有

$$A^m=k_0I+k_1+\cdots+k_{m-1}A^{m-1}\quad(k_i=-b_{m-i})$$

由此可以求出

$$\begin{cases}A^{m+1}=k_0A+k_1A^2+\cdots+k_{m-1}A^m=k_0k_{m-1}I+(k_0+k_1k_{m-1})A+\cdots+(k_{m-2}+k_{m-1}^2)A^{m-1}\\ \qquad\quad =k_0^{(1)}I+k_1^{(1)}A+\cdots+k_{m-1}^{(1)}A^{m-1}\\ \qquad\quad \cdots\cdots\\ A^{m+l}=k_0I+k_1A+\cdots+k_{m+l-1}A^{m+l-1}=k_0^{(l)}I+k_1^{(l)}A+\cdots+k_{m-1}^{(l)}A^{m-1}\\ \qquad\quad \cdots\cdots\end{cases}$$

其中，$k_0^{(l)}=k_0^{(l-1)}k_{m-1}^{(l-1)}$， $k_i^{(l)}=k_{i-1}^{(l-1)}+k_i^{(l-1)}k_{m-1}^{(l-1)}(i\geq 1)$，于是有

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

这表明，矩阵幂级数的求和问题可以转化为 m 个数项级数求和问题. 当 $\psi(A)$ 中的非零项很少时，这种求法的性能十分优越.

按上述方法，在实际使用待定系数法求解函数矩阵时，我们通常遵循以下步骤:

1. 求出满足 $\psi(A)=O$ 的首 1 多项式
2. 通过上述递推关系求出 A^m，A^{m+1}，A^{m+l}，...
3. 将原矩阵函数展开并化简为上述数项级数求和的形式，计算数项级数并求和得到答案

3.2.2 举例展示求法

题目.

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

解答.

1. 求出满足 $\psi(A)=O$ 的首 1 多项式

$$\lambda I-A=\begin{bmatrix}\lambda-\pi & 0 & 0 & 0\\ 0 & \lambda+\pi & 0 & 0\\ 0 & 0 & \lambda & -1\\ 0 & 0 & 0 & \lambda\end{bmatrix}\rightarrow\begin{bmatrix}\lambda-\pi & 0 & 0 & 0\\ 0 & \lambda+\pi & 0 & 0\\ 0 & 0 & \lambda & 0\\ 0 & 0 & 0 & \lambda\end{bmatrix}$$

   $\varphi(\lambda)=det(\lambda I-A)=\lambda^4-\pi^2\lambda^2$. 由于 $\varphi(A)=O$，并且注意到该式中非零项极少，可以使用数项级数求和法

2. 通过上述递推关系求出 A^m, A^{m+1}, A^{m+l}, $\ldots$
   $A^4=\pi^2A^2$, $A^5=\pi^2A^3$, $A^7=\pi^4A^3$, $\ldots$

3. 将原矩阵函数展开并化简为上述数项级数求和的形式，计算数项级数并求和得到答案

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

3.3 对角型法

3.3.1 对角型法求矩阵函数的步骤推导

设 A 相似于对角矩阵 Λ，即有可逆矩阵 P，使得

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

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

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

也就是，如果我们能求得 A 与对角矩阵相似，就可以将矩阵幂级数求和问题转化为求相似变换矩阵的问题

按上述方法，在实际使用对角型法求解函数矩阵时，我们通常遵循以下步骤：

1. 求出特征多项式 $\varphi(\lambda)$
2. 求可逆矩阵 P，使得 $P^{-1}AP = diag(\lambda_1,\ \lambda_2,\ \dots,\ \lambda_n)$
3. 计算 $f(A) = P \cdot diag(f(\lambda_1,\ \lambda_2,\ \dots, \lambda_n)) \cdot P^{-1}$

下面一章节将通过《矩阵论第五版》习题 3.3.5 演示该过程.

3.3.2 举例展示求法

题目.

设 $A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$，求 e^A，$e^{tA}(t \in R)$，$sinA$

解答.

1. 求出特征多项式 $\varphi(\lambda)$

   我们在 **3.1.2** 中已经求过了该矩阵的特征多项式 $\varphi(\lambda)=(\lambda+1)(\lambda-1)(\lambda-2)$，这里不再赘述.

2. 求可逆矩阵 P，使得 $P^{-1}AP=diag(\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_n)$

   特征值 $\lambda_1=-1$ 对应的特征向量为 $p_1=(1,\ -3,\ 3)^T$，特征值 $\lambda_2=1$ 对应的特征向量为 $p_2=(-1,\ 1,\ 1)^T$，特征值 $\lambda_3=2$ 对应的特征向量为 $p_3=(1,\ 0,\ 0)^T$. 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

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

3. 计算 $f(A)=P\cdot diag(f(\lambda_1,\ \lambda_2,\ \ldots,\lambda_n))\cdot P^{-1}$

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

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

$$sinA=P\begin{bmatrix}sin(-1) & & \\ & sin1 & \\ & & sin2\end{bmatrix}P^{-1}=\frac{1}{6}\begin{bmatrix}sin2 & 4sin2-2sin1 & 2sin2-4sin1\\ 0 & 0 & 6sin1\\ 0 & 6sin1 & 0\end{bmatrix}$$

3.4 若尔当标准型法

3.4.1 若尔当标准型法求矩阵函数的步骤推导

设 A 的 Jordan 标准形为 J，则有可逆矩阵 P，使得

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

其中

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

可求得

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

则有

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

按上述方法，在实际使用 Jordan 标准形法求解函数矩阵时，我们通常遵循以下步骤:

1. 求可逆矩阵 P，使得 $P^{-1}AP = diag(J_1,\ J_2,\ \ldots,\ J_s)$，其中 J_i 是 m_i 阶的 Jordan 块
2. 对于 $i = 1,\ 2,\ \ldots,\ s$，计算 $f^{(l)}(\lambda_i)(l = 0,\ 1,\ \ldots,\ m_i - 1)$，并构造 m_i 阶矩阵 $f(J_i) = f(\lambda_i)I_{m_i} + f'(\lambda_i)I_{m_i}^{(1)} + \cdots + \frac{f^{(m_i-1)}(\lambda_i)}{(m_i-1)!}I_{m_i}^{(m_i-1)}$
3. 计算 $f(A) = P \cdot diag(f(J_1),\ f(J_2),\ \ldots,\ f(J_s)) \cdot P^{-1}$

下面一章节将通过《矩阵论第五版》习题 3.3.6 演示该过程.

3.4.2 举例展示求法

由于之前已经给过 Jordan 标准形和 P 矩阵的详细求解方法，这里不再赘述.

题目.

设 $f(z) = lnz$，求 $f(A)$，这里 A 为:

$$(1)\quad A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix};\qquad (2)\quad A = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

解答.

1. 求可逆矩阵 P，使得 $P^{-1}AP = diag(J_1,\ J_2,\ \ldots,\ J_s)$，其中 J_i 是 m_i 阶的 Jordan 块

   对 A 求得

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

2. 对于 $i = 1,\ 2,\ \ldots,\ s$，计算 $f^{(l)}(\lambda_i)(l = 0,\ 1,\ \ldots,\ m_i - 1)$，并构造 m_i 阶矩阵 $f(J_i) = f(\lambda_i)I_{m_i} + f'(\lambda_i)I_{m_i}^{(1)} + \cdots + \frac{f^{(m_i-1)}(\lambda_i)}{(m_i-1)!}I_{m_i}^{(m_i-1)}$

$$f(J_1) = f(J) = \begin{bmatrix} ln1 & \frac{1}{1!}\frac{1}{1} & \frac{1}{2!}\frac{-1}{1^2} & \frac{1}{3!}\frac{2}{1^3} \\ & ln1 & \frac{1}{1!}\frac{1}{1} & \frac{1}{2!}\frac{-1}{1^2} \\ & & ln1 & \frac{1}{1!}\frac{1}{1} \\ & & & ln1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & -\frac{1}{2} & \frac{1}{3} \\ & 0 & 1 & -\frac{1}{2} \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}$$

3. 计算 $f(A) = P \cdot diag(f(J_1),\ f(J_2),\ \ldots,\ f(J_s)) \cdot P^{-1}$

$$lnA = f(A) = Pf(J)P^{-1} = \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ -\frac{1}{2} & 1 & 0 & \\ \frac{1}{3} & -\frac{1}{2} & 1 & 0 \end{bmatrix}$$

同理，对于 (2) 我们采取同样的步骤：

1. 求可逆矩阵 P，使得 $P^{-1}AP = diag(J_1,\ J_2,\ \ldots,\ J_s)$，其中 J_i 是 m_i 阶的 Jordan 块

   观察矩阵形式，发现其已经是一个 Jordan 标准形，且有

$$A = \begin{bmatrix} J_1 & \\ & J_2 \end{bmatrix},\quad J_1 = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix},\quad J_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

   故可取 $P = P^{-1} = I$.

2. 对于 $i = 1,\ 2,\ \ldots,\ s$，计算 $f^{(l)}(\lambda_i)(l = 0,\ 1,\ \ldots,\ m_i - 1)$，并构造 m_i 阶矩阵 $f(J_i) = f(\lambda_i)I_{m_i} + f'(\lambda_i)I_{m_i}^{(1)} + \cdots + \frac{f^{(m_i-1)}(\lambda_i)}{(m_i-1)!}I_{m_i}^{(m_i-1)}$

$$f(J_1) = \begin{bmatrix} ln2 & \frac{1}{2} \\ 0 & ln2 \end{bmatrix},\quad lnJ_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

3. 计算 $f(A) = P \cdot diag(f(J_1),\ f(J_2),\ \ldots,\ f(J_s)) \cdot P^{-1}$

$$lnA = f(A) = Pf(J)P^{-1} = f(J) = \begin{bmatrix} f(J_1) & \\ & f(J_2) \end{bmatrix} = \begin{bmatrix} ln2 & \frac{1}{2} & 0 & 0 \\ & ln2 & 0 & 0 \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}$$

第 4 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

矩阵的 LU 分解与 *Gauss* 消元过程密切相关，因此我们先推导 *Gauss* 消元的过程：

Gauss 消元过程：

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

$$L_1=\begin{bmatrix}1&&&\\c_{21}&1&&\\\vdots&&\ddots&\\c_{n1}&&&1\end{bmatrix},\quad L_1^{-1}\begin{bmatrix}1&&&\\-c_{21}&1&&\\\vdots&&\ddots&\\-c_{n1}&&&1\end{bmatrix}$$

计算

$$L_1^{-1}A^{(0)}=\begin{bmatrix}a_{11}^{(0)}&a_{12}^{(0)}&\cdots&a_{1n}^{(0)}\\&a_{22}^{(1)}&\cdots&a_{n}^{(1)}\\&\vdots&&\vdots\\&a_{n2}^{(1)}&\cdots&a_{nn}^{(1)}\end{bmatrix}=A^{(1)}$$

由此可见，$A^0=A$ 的第一列除主元 $a_{11}^{(0)}$ 外，其余元素全被化为零. 故有

$$A^{(0)}=L_1A^{(1)}$$

因为倍加初等变换不改变矩阵的行列式的值，所以由 $A^{(0)}$ 得 A 的二阶顺序主子式为

$$\Delta_2=a_{11}^{(0)}a_{22}^{(11)}$$

若 $\Delta_2\neq 0$，则 $a_{22}^{(1)}\neq 0$. 令 $c_{i2}=\frac{a_{i2}^{(1)}}{a_{22}^{(1)}}(i=3,\ 4,\ \ldots,\ n)$，并构造矩阵

$$L_2=\begin{bmatrix}1&&&&\\&1&&&\\&c_{32}&1&&\\&\vdots&&\ddots&\\&c_{n2}&&&1\end{bmatrix},\quad L_2^{-1}=\begin{bmatrix}1&&&&\\&1&&&\\&-c_{32}&1&&\\&\vdots&&\ddots&\\&-c_{n2}&&&1\end{bmatrix}$$

计算，得到

$$L_2^{-1}A^{(1)} = \begin{bmatrix} a_{11}^{(0)} & a_{12}^{(0)} & a_{13}^{(0)} & \cdots & a_{1n}^{(0)} \\ & a_{22}^{(1)} & a_{23}^{(1)} & \cdots & a_{n}^{(1)} \\ & & a_{33}^{(2)} & \cdots & a_{3n}^{(2)} \\ & & \vdots & & \vdots \\ & & a_{n3}^{(2)} & \cdots & a_{nn}^{(2)} \end{bmatrix} = A^{(2)}$$

同理，$A^{(2)}$ 的前两列中主元以下的元素全为零，有

$$A^{(1)} = L_2 A^{(2)}$$

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

$$\Delta_3 = a_{11}^{(0)} a_{22}^{(1)} a_{33}^{(2)}$$

如此迭代至第 $r-1$ 步，得到

$$A^{(r-1)} = \begin{bmatrix} a_{11}^{(0)} & \cdots & a_{1,r-1}^{(0)} & a_{1r}^{(0)} & \cdots & a_{1n}^{(0)} \\ & \ddots & \vdots & \vdots & & \vdots \\ & & a_{r-1,r-1}^{(r-2)} & a_{r-1,r}^{(r-2)} & \cdots & a_{r-1,n}^{(r-2)} \\ & & & a_{rr}^{(r-1)} & \cdots & a_{rn}^{(r-1)} \\ & & & \vdots & & \vdots \\ & & & a_{nr}^{(r-1)} & \cdots & a_{nn}^{(r-1)} \end{bmatrix}$$

若 $\Delta_r \neq 0$，则 $a_{rr}^{(r-1)} \neq 0$. 令 $C_{ir} = \frac{a_{ir}^{(r-1)}}{a_{rr}^{(r-1)}} (i = r+1,\ r+2,\ \dots,\ n)$，并构造 Frobenius 矩阵

$$L_r = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & c_{r+1,r} & 1 & & \\ & & \vdots & & \ddots & \\ & & c_{nr} & & & 1 \end{bmatrix}, \quad L_r^{-1} = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & -c_{r+1,r} & 1 & & \\ & & \vdots & & \ddots & \\ & & -c_{nr} & & & 1 \end{bmatrix}$$

计算可得

$$L_r^{-1}A^{(r-1)} = \begin{bmatrix} a_{11}^{(0)} & \cdots & a_{1r}^{(0)} & a_{1,r+1}^{(0)} & \cdots & a_{1n}^{(0)} \\ & \ddots & \vdots & \vdots & & \vdots \\ & & a_{rr}^{(r-1)} & a_{r,r+1}^{(r-1)} & \cdots & a_{rn}^{(r-1)} \\ & & & a_{r+1,r+1}^{(r)} & \cdots & a_{r+1,n}^{(r)} \\ & & & \vdots & & \vdots \\ & & & a_{n,r+1}^{(r)} & \cdots & a_{nn}^{(r)} \end{bmatrix}$$

同理，有

$$A^{(r-1)} = L_r A^{(r)}$$

故 A 的 $r+1$ 阶顺序主子式为

$$\Delta_{r+1} = a_{11}^{(0)} a_{22}^{(1)} \dots a_{rr}^{(r-1)} a_{r+1,r+1}^{(r)}$$

迭代至第 $n-1$ 步，有

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

以上便是 **Gauss 消元**的全过程. 根据前面所述，由于消元过程中未使用行、列的交换，因此其进行到底的条件是

$a_{11}^{(0)}$，$a_{22}^{(1)}$，…，$a_{n-1,n-1}^{(n-2)}$ 都不为零，即

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

从上述推导过程中不难发现，当 $\Delta_r \neq 0$ 时，有

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

令 $L = L_1L_2\dots L_{n-1}$，有

$$L = L_1L_2\dots L_{n-1} = \begin{bmatrix} 1 & & & & \\ c_{21} & 1 & & & \\ \vdots & \vdots & \ddots & & \\ c_{n-1,1} & c_{n-1,2} & \cdots & 1 & \\ c_{n1} & c_{n2} & \cdots & c_{n,n-1} & 1 \end{bmatrix}$$

L 是一个对角元素都为 1 的下三角矩阵,称为 **单位下三角矩阵**. 而由之前的推导, $A^{(n-1)}$ 是一个上三角矩阵，令 $U = A^{(n-1)}$，有

$$A = LU$$

如此，矩阵 A 就被分解成一个单位下三角矩阵与一个上三角矩阵的乘积. 现在，我们对矩阵的 **LU 分解**做如下定义:

LU 分解: 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 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**.

对于矩阵的 LU 分解有如下定理:

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

   $$D = diag(d_1,\ d_2,\ \dots, d_n)$$

   其中 $d_k = \frac{\Delta_k}{\Delta_{k-1}}(k = 1,\ 2,\ \dots,\ n;\ \Delta_0 = 1)$.
2. n 阶可逆矩阵 A 有三角分解 $A = LU$ 的充要条件是 A 的顺序主子式 $\Delta_k \neq 0$ $(k = 1,\ 2,\ \dots,\ n—1)$.
3. 设 A 是 n 阶可逆矩阵，则存在置换矩阵 P 使 PA 的 n 个顺序主子式非零，且有

   $$PA = L\hat{U} = LDU$$

由上述过程，我们可以得出矩阵的 LDU 分解方法:

1. 判断矩阵 A 的 k 阶顺序主子式都满足 $\Delta_k \neq 0$，如果不，可以参考定理 3 构造 $PAx = Pb$ 并求其解
2. 根据 Gauss 消元过程递推地计算 L_k，$A^{(k)}$

3. 计算 $L = L_1 L_2 \dots L_{n-1}$

4. 根据 $A = L_1 L_2 \dots L_{n-1} A^{(n-1)} = LA^{(n-1)}$ 得到矩阵的 LDU 分解

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

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

下面给出 Crout 分解的步骤推导. 设

$$\hat{L} = \begin{bmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix}, \quad U = \begin{bmatrix} 1 & u_{12} & \cdots & u_{1n} \\ & 1 & \cdots & u_{2n} \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}$$

根据 $A = \hat{L}U$，可以得到

$$a_{i1} = l_{i1} \quad or \quad l_{i1} = a_{i1} \quad (i = 1,\ 2,\ \dots,\ n) \tag{4.1.1}$$

$$a_{1j} = l_{11} u_{1j} \quad or \quad u_{1j} = \frac{a_{1j}}{l_{11}} \quad (j = 2,\ 3,\ \dots,\ n) \tag{4.1.2}$$

对于 $k = 2,\ 3,\ \dots,\ n$，当 $i \geq k$ 时，有

$$\because a_{ik} = l_{i1} u_{1k} + \cdots + l_{i,k-1} u_{k-1,k} + l_{ik}$$

$$\therefore l_{ik} = a_{ik} - (l_{i1} u_{1k} + \cdots + l_{i,k-1} u_{k-1,k}) \tag{4.1.3}$$

而当 $j > k$ 时，有

$$\because a_{kj} = l_{k1} u_{1j} + \cdots + l_{k,k-1} u_{k-1,j} + l_{kk} u_{kj}$$

$$\therefore u_{kj} = \frac{1}{l_{kk}} [a_{kj} - (l_{k1} u_{1j} + \cdots + l_{k,k-1} u_{k-1,j})] \tag{4.1.4}$$

在实际计算时，顺序如下

1. 利用式 (4.1.1) 计算 $\hat{L}$ 的第一列

   利用式 (4.1.2) 计算 U 的第一行

2. 利用式 (4.1.3) 计算 $\hat{L}$ 的第二列

   利用式 (4.1.4) 计算 U 的第二行

3. $\cdots\cdots$

4. 利用式 (4.1.3) 计算 $\hat{L}$ 的第 $n-1$ 列

   利用式 (4.1.4) 计算 U 的第 $n-1$ 行

5. 利用式 (4.1.3) 计算 $\hat{L}$ 的第 n 列

Doolittle 分解：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

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

与上述过程类似，我们可以得到 Doolittle 分解的计算公式为

$$\begin{cases} u_{ik} = a_{ik} - \sum_{r=1}^{i-1} l_{ir}u_{rk} \quad (k = i,\ i+1,\ \ldots, n) \\ l_{ki} = \frac{1}{u_{ii}}(a_{ki} - \sum_{r=1}^{i-1} l_{kr}u_{ri}) \quad (k = i+1,\ \ldots,\ n) \end{cases}$$

实对称矩阵的 Cholesky 分解：当 A 为实对称正定矩阵时，$\Delta_k > 0 \quad (k = 1,\ 2,\ \ldots,\ n)$. 于是 A 有唯一的 LDU 分解 $A = LDU$，其中 $D = diag(d_1,\ d_2,\ \ldots,\ n)$，且 $d_i > 0 \ \ (i = 1,\ 2,\ \ldots,\ n)$. 令

$$\widetilde{D} = diag(\sqrt{d_1},\ \sqrt{d_2},\ \ldots,\ \sqrt{d_n})$$

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

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

或者

$$A = L\widetilde{D}^2L^T = (L\widetilde{D})(L\widetilde{D})^T = GG^T$$

其中，$G = L\widetilde{D}$ 是下三角矩阵.

令 $G = (g_{ij})$，则由 $A = GG^T$ 两端相应元素相等可得

$$a_{ij} = g_{i1}g_{j1} + g_{i2}g_{j2} + \cdots + g_{ij}g_{ij} \quad (i > j), \quad a_{ii} = g_{i1}^2 + g_{i2}^2 + \cdots + g_{ii}^2$$

从而得到计算 g_{ij} 的递推关系式为

$$g_{ij} = \begin{cases} (a_{ii} - \sum_{k=1}^{i-1} g_{ik}^2)^{1/2} & (\mathrm{i = j}) \\ \frac{1}{g_{jj}}(a_{ij} - \sum_{k=1}^{j-1} g_{ik}g_{jk}) & (\mathrm{i > j}) \\ 0 & (\mathrm{i < j}) \end{cases}$$

实际求解过程是按照行优先原则对 g_{ij} 进行求解.

4.1.2 举例展示求法

题目.

求矩阵 A 的 LDU 分解和 Doolittle 分解，其中

$$A=\begin{bmatrix}5 & 2 & -4 & 0\\ 2 & 1 & -2 & 1\\ -4 & -2 & 5 & 0\\ 0 & 1 & 0 & 2\end{bmatrix}$$

解答.

Gauss 消元法:

1. 判断矩阵 A 的 k 阶顺序主子式都满足 $\Delta_k \neq 0$，如果不，可以参考定理 3 构造 $PAx = Pb$ 并求其解

   因为 $\Delta_1 = 5$，$\Delta_2 = 1$，$\Delta_3 = 1$，所以 A 有唯一的 LDU 分解.

2. 根据 Gauss 消元过程递推地计算 L_k，$A^{(k)}$

   对 A 构造矩阵

$$L_1=\begin{bmatrix}1 & & & \\ \frac{2}{5} & 1 & & \\ -\frac{4}{5} & 0 & 1 & \\ 0 & 0 & 0 & 1\end{bmatrix},\quad L_1^{-1}=\begin{bmatrix}1 & & & \\ -\frac{3}{5} & 1 & & \\ \frac{4}{5} & 0 & 1 & \\ 0 & 0 & 0 & 1\end{bmatrix}$$

   计算，得

$$L_1^{-1}A=\begin{bmatrix}5 & 2 & -4 & 0\\ 0 & \frac{1}{5} & -\frac{2}{5} & 1\\ 0 & -\frac{2}{5} & \frac{4}{5} & 0\\ 0 & 1 & 0 & 2\end{bmatrix}=A^{(1)}$$

   对 $A^{(1)}$ 构造矩阵

$$L_2=\begin{bmatrix}1 & & & \\ 0 & 1 & & \\ 0 & -2 & 1 & \\ 0 & 5 & 0 & 1\end{bmatrix},\quad L_2^{-1}=\begin{bmatrix}1 & & & \\ 0 & 1 & & \\ 0 & 2 & 1 & \\ 0 & -5 & 0 & 1\end{bmatrix}$$

计算，得

$$L_2^{-1}A^{(1)} = \begin{bmatrix} 5 & 2 & -4 & 0 \\ 0 & \frac{1}{5} & -\frac{2}{5} & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & -3 \end{bmatrix} = A^{(2)}$$

对 $A^{(2)}$ 构造矩阵

$$L_3 = \begin{bmatrix} 1 & & & \\ 0 & 1 & & \\ 0 & 0 & 1 & \\ 0 & 0 & 2 & 1 \end{bmatrix}, \quad L_3^{-1} = \begin{bmatrix} 1 & & & \\ 0 & 1 & & \\ 0 & 0 & 1 & \\ 0 & 0 & -2 & 1 \end{bmatrix}$$

计算，得

$$L_3^{-1}A^{(2)} = \begin{bmatrix} 5 & 2 & -4 & 0 \\ & \frac{1}{5} & -\frac{2}{5} & 1 \\ & & 1 & 2 \\ & & & -7 \end{bmatrix} = A^{(3)}$$

3. 计算 $L = L_1L_2 \dots L_{n-1}$

   令 $L = L_1L_2L_3$，有

$$L = L_1L_2L_3 = \begin{bmatrix} 1 & & & \\ \frac{2}{5} & 1 & & \\ -\frac{4}{5} & -2 & 1 & \\ 0 & 5 & 2 & 1 \end{bmatrix}$$

4. 根据 $A = L_1L_2 \dots L_{n-1}A^{(n-1)} = LA^{(n-1)}$ 得到矩阵的 LDU 分解

   于是，A 的 LDU 分解为

$$A = L_1L_2L_3A^{(3)} = \begin{bmatrix} 1 & & & \\ \frac{2}{5} & 1 & & \\ -\frac{4}{5} & -2 & 1 & \\ 0 & 5 & 2 & 1 \end{bmatrix} \begin{bmatrix} 5 & & & \\ & \frac{1}{5} & & \\ & & 1 & \\ & & & 7 \end{bmatrix} \begin{bmatrix} 1 & \frac{2}{5} & -\frac{4}{5} & 0 \\ & 1 & -2 & 5 \\ & & 1 & 2 \\ & & & 1 \end{bmatrix}$$

Doolittle 分解:

这里我们使用 Crout 分解来求 Doolittle 分解.

1. 利用式 (4.1.1) 计算 $\hat{L}$ 的第一列

利用式 (4.1.2) 计算 U 的第一行

$$l_{11} = a_{11} = 5, \quad l_{21} = a_{21} = 2, \quad l_{31} = a_{31} = -4, \quad l_{41} = a_{41} = 0$$
$$u_{12} = \frac{a_{12}}{l_{11}} = \frac{2}{5}, \quad u_{13} = \frac{a_{13}}{l_{11}} = -\frac{4}{5}, \quad u_{14} = \frac{a_{14}}{l_{11}} = 0$$

2. 利用式 (4.1.3) 计算 $\hat{L}$ 的第二列

利用式 (4.1.4) 计算 U 的第二行

$$l_{22} = a_{22} - l_{21}u_{12} = \frac{1}{5}, \quad l_{32} = a_{32} - l_{31}u_{12} = -\frac{2}{5}, \quad l_{42} = a_{42} - l_{41}u_{12} = 1$$
$$u_{23} = \frac{1}{l_{22}}[a_{23} - l_{21}u_{13}] = -2, \quad u_{24} = \frac{1}{l_{24}}[a_{24} - l_{21}u_{14}] = 5$$

3. 利用式 (4.1.3) 计算 $\hat{L}$ 的第三列

利用式 (4.1.4) 计算 U 的第三行

$$l_{33} = a_{33} - (l_{31}u_{13} + l_{32}u_{23}) = 1, \quad l_{43} = a_{43} - (l_{41}u_{13} + l_{42}u_{23}) = 2$$
$$u_{34} = \frac{1}{l_{33}}[a_{34} - (l_{31}u_{14} + l_{32}u_{24})] = 2$$

4. 利用式 (4.1.3) 计算 $\hat{L}$ 的第四列

$$l_{44} = a_{44} - (l_{41}u_{14} + l_{42}u_{24} + l_{43}u_{34}) = 7$$

求得

$$\widetilde{L} = \begin{bmatrix} 5 & & & \\ 2 & \frac{1}{5} & & \\ -4 & -\frac{2}{5} & 1 & \\ 0 & 1 & 2 & 7 \end{bmatrix}, \quad U = \begin{bmatrix} 1 & \frac{2}{5} & -\frac{4}{5} & 0 \\ & 1 & -2 & 5 \\ & & 1 & 2 \\ & & & 1 \end{bmatrix}$$

故 Doolittle 分解为

$$A = \widetilde{L}U = LDU = L\widetilde{U} = \begin{bmatrix} 1 & & & \\ \frac{2}{5} & 1 & & \\ -\frac{4}{5} & -2 & 1 & \\ 0 & 5 & 2 & 1 \end{bmatrix} \begin{bmatrix} 5 & 2 & -4 & 0 \\ & \frac{1}{5} & -\frac{2}{5} & 1 \\ & & 1 & 2 \\ & & & -7 \end{bmatrix}$$

题目.

求对称正定矩阵 A 的 Cholesky 分解，其中

$$A = \begin{bmatrix} 5 & 2 & -4 \\ 2 & 1 & -2 \\ -4 & -2 & 5 \end{bmatrix}$$

解答.

Cholesky 分解:

容易验证 A 式对称正定矩阵.

$$g_{11} = \sqrt{a_{11}} = \sqrt{5}$$
$$g_{21} = \frac{a_{21}}{g_{11}} = \frac{2}{\sqrt{5}}, \quad g_{22} = (a_{22} - g_{21}^2)^{1/2} = \frac{1}{\sqrt{5}}$$
$$g_{31} = \frac{a_{31}}{g_{11}} = -\frac{4}{\sqrt{5}}, \quad g_{32} = \frac{a_{32} - g_{31}g_{21}}{g_{22}} = -\frac{2}{\sqrt{5}}, \quad g_{33} = (a_{33} - g_{31}^2 - g_{32}^2)^{1/2} = 1$$

从而，得

$$A = \begin{bmatrix} \sqrt{5} & & \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \\ -\frac{4}{\sqrt{5}} & -\frac{2}{\sqrt{5}} & 1 \end{bmatrix} \begin{bmatrix} \sqrt{5} & \frac{2}{\sqrt{5}} & -\frac{4}{\sqrt{5}} \\ & \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ & & 1 \end{bmatrix}$$

4.1.3 代码实现

LU 分解

```
import numpy as np

def lu_decomposition(A):
    n =  len(A) # 获取矩阵的维数，即行数（假设矩阵是方阵）

    # 初始化下三角矩阵L和上三角矩阵U为零矩阵
    L = np.zeros((n, n))
    U = np.zeros((n, n))

    # 遍历矩阵的每一行
    for i in  range(n):
        # 计算上三角矩阵U的元素
        for k in  range(i, n):
             sum = 0  # 初始化求和变量，用于计算U矩阵的元素
            for j in  range(i):
                 sum += (L[i][j] * U[j][k])
            U[i][k] = A[i][k] -  sum  # 更新U矩阵的元素

        # 计算下三角矩阵L的元素
        for k in  range(i, n):
            if i == k:
```

```
                L[i][i] = 1 # 对角线上的元素设置为1
            else:
                 sum = 0 # 初始化求和变量，用于计算L矩阵的元素
                for j in  range(i):
                    sum += (L[k][j] * U[j][i])
                L[k][i] = (A[k][i] -  sum) / U[i][i] # 更新L矩阵的元素

    return L, U
```

4.2 矩阵的 QR 分解

4.2.1 矩阵 QR 分解的步骤推导

Givens 矩阵: 设实数 c 和 s 满足 $c^2+s^2=1$，对应的 *Givens* 矩阵为 $T_{ij}(c,\ s)(i\neq j)$，则

1. $T_{ij}(c,\ s)$ 是正交矩阵;
2. $[T_{ij}(c,\ s)]^{-1}=T_{ij}(c,\ -s)$;
3. $det[T_{ij}(c,\ s)]=1$;
4. 设实向量 **textitx** $=(\xi_1,\ \xi_2,\ \ldots,\ \xi_n)^T$，$y=T_{ij}(c,\ s)\boldsymbol{x}=(\eta_1,\ \eta_2,\ \ldots,\ eta_n)^T$，则

$$\eta_i=c\xi_i+s\xi_j,\quad \eta_j=-s\xi_i+c\xi_j$$
$$\eta_k=\xi_k\quad (k\neq i,\ j;\ k=1,\ 2,\ \ldots,\ n)$$

5. 设 $\boldsymbol{x}=(\xi_1,\ \xi_2,\ \ldots,\ \xi_n)^T\neq 0$，则存在有限个 Givens 矩阵的乘积，记作 T，使得 $T\boldsymbol{x}=|\boldsymbol{x}|e_1$

Householder 矩阵: 设 $u\in R^n$ 是单位列向量，对应 Householder 矩阵为 $H_u=I_n-2uu^T$，则

1. H_u 是对称 $(H^T=H)$、正交 $(H^TH=I)$、自逆 $(H^2=I)$、对合矩阵 $H^{-1}=H$
2. $detH=-1$
3. 任意给定非零列向量 $\boldsymbol{z}\in R^n(n>1)$ 及单位列向量 $\boldsymbol{z}\in R^n$，则存在 Householder 矩阵 H，使得 $H\boldsymbol{x}=|\boldsymbol{x}|\boldsymbol{z}$
4. 一个 Givens 矩阵可以表示为两个 Householder 矩阵的乘积

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 $A=QR$ 为 A 的 **QR 分解**.

Schimidt 正交化求 QR 分解:

记矩阵 A 的 n 个列向量依次为 a_1, a_2, $\ldots$, 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, $\ldots$, q_n.

对 a_1, a_2, $\ldots$, a_n 正交化，可得

$$\begin{cases} b_1 = a_1 \\ b_2 = a_2 - k_{21}b_1 \\ \quad \cdots\cdots \\ b_n = a_n - k_{n,n-1}b_{n-1} - \cdots - k_{n1}b_1 \end{cases}$$

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

$$\begin{cases} a_1 = b_1 \\ a_2 = k_{21}b_1 + b_2 \\ \quad \cdots\cdots \\ a_n = k_{n1}b_1 + k_{n2}b_2 + \cdots + kn,n-1b_{n-1} + b_n \end{cases}$$

用矩阵形式表示为

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

其中

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

对 b_1, b_2, $\ldots$, b_n 单位化，可以得到

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

于是有

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

令
$$\begin{cases} Q=(q_1,\ q_2,\ \dots,\ q_n) \\ R=diag(|b_1|,\ |b_2|,\ \dots,\ |b_n|)\cdot C \end{cases}$$
则 Q 是正交（酉）矩阵，R 是可逆上三角矩阵，且有 $A=QR.$

总结求解流程如下：

1. 对 A 的列向量组 $a_1,\ a_2,\ \dots, a_n$ 正交化得正交向量组 $b_1,\ b_2,\ \dots,\ b_n$
2. 构造正交矩阵 $Q=(q_1,\ q_2,\ \dots,\ q_n)$，其中 $q_j=\frac{b_j}{|b_j|}$
3. 构造上三角矩阵 $R=diag(|b_1|,\ |b_2|,\ \dots,\ |b_n|)\cdot C$，那么 $A=QR$

Givens 变换求 QR 分解：

由 $detA\neq 0$ 知，A 的第 1 列 $b^{(1)}=(a_{11},\ a_{21},\ \dots,\ a_{n1})^T\neq 0$. 存在有限个 Gievns 矩阵的乘积，记作 T_1，使得
$$T_1b^{(1)}=|b^{(1)}|e_1\quad(e_1\in R^n)$$
令 $a_{11}^{(1)}=|b^{(1)}|$，则有
$$T_1A=\begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & & & \\ \vdots & & A^{(1)} & \\ 0 & & & \end{bmatrix}$$
由 $detA^{(1)}\neq 0$ 知，$A^{(1)}$ 的第 1 列 $b^{(2)}=(a_{22}^{(1)},\ a_{23}^{(1)},\ \dots,\ a_{n2}^{(1)})$. 存在有限个 Givens 矩阵的乘积，记作 T_2，使得
$$T_2b^{(2)}=|b^{(2)}|e_1\quad(e_1\in R^{(n-1)})$$
令 $a_{22}^{(2)}=|b^{(2)}|$，则有
$$T_2A^{(1)}=\begin{bmatrix} a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{24}^{(2)} \\ 0 & & & \\ \vdots & & A^{(2)} & \\ 0 & & & \end{bmatrix}$$
以此类推，最后有
$$T_{n-1}b^{(n-1)}=|b^{(n-1)}|e_1\quad(e_1\in R^2)$$
令 $a_{n-1,n-1}^{n-1}=|b^{(n-1)}|$，则有
$$T_{n-1}A^{(n-2)}=\begin{bmatrix} a_{n-1,n-1}^{(n-1)} & a_{n-1,n}^{(n-1)} \\ 0 & a_{nn}^{(n-1)} \end{bmatrix}$$

最后，令

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

则 T 是有限个 Givens 矩阵的乘积，使得

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

记为 R，有 $A=QR$，其中 $Q=T^{-1}$. 因为 T 是有限个 Givens 矩阵的乘积，而 Givens 矩阵都是正交矩阵，所以 T 是正交矩阵，于是 $Q=T^{-1}=T^T$ 也是正交矩阵.

总结求解流程如下：

1. 对 A 的第 1 列 $b^{(1)}$ 构造有限个 Givens 矩阵的乘积 T_1，使得

$$T_1b^{(1)}=|b^{(1)}|e_1 \quad (e_1\in R^n)T_1A=\begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & & & \\ \vdots & & A^{(1)} & \\ 0 & & & \end{bmatrix}$$

2. 对 $A^{(1)}$ 的第 1 列 $b^{(2)}$ 构造 T_2
3. ······
4. 对 $A^{(n-2)}$ 的第 1 列 $b^{(n-1)}$ 构造 Givens 矩阵 T_{n-1}
5. 构造上三角矩阵 R，计算正交矩阵 Q，那么 $A=QR$.

Householder 变换求 QR 分解：

Householder 变化求解过程与 Givens 类似，这里不再赘述重复部分.

最后得出

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

并注意到，若 H_n 是 $n-l$ 阶 Householder 矩阵，即

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

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

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

是 n 阶 Householder 矩阵. 因此，S 是有限个 Householder 矩阵的乘积，且使得 SA 等于上述上三角矩阵，记为 R，则 $Q=S^{-1}=S^T$，有 $A=QR$.

4.2.2 举例展示求法

题目.

用 Schmidt 正交化方法求矩阵

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

的 QR 分解

解答.

1. 对 A 的列向量组 a_1，a_2，$\ldots,a_n$ 正交化得正交向量组 b_1，b_2，…，b_n
   令 $a_1=(0,\ 1,\ 1)^T$，$a_2=(1,\ 1,\ 0)^T$，$a_3=(1,\ 0,\ 1)^T$，正交化可得

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

2. 构造正交矩阵 $Q=(q_1,\ q_2,\ \ldots,\ q_n)$，其中 $q_j=\frac{b_j}{|b_j|}$
   根据给定方法构造矩阵

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

3. 构造上三角矩阵 $R=diag(|b_1|,\ |b_2|,\ \ldots,\ |b_n|)\cdot C$，那么 $A=QR$

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

则有 $A=QR$

题目.

用 Givens 变换求矩阵

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

的 QR 分解

解答.

1. 对 A 的第 1 列 $b^{(1)}$ 构造有限个 Givens 矩阵的乘积 T_1

   对 A 的第 1 列 $b^{(1)}=(2,\ 0,\ 2)^T$，构造 T_1 使得 $T_1b^{(1)}=|b^{(1)}|e_1$

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

$$T_1=T_{13},\quad T_1A=\begin{bmatrix}2\sqrt{2}&\frac{3}{\sqrt{2}}&\frac{3}{\sqrt{2}}\\0&2&2\\0&-\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{bmatrix}$$

2. 对 $A^{(1)}$ 的第 1 列 $b^{(2)}$ 构造 T_2

   对 $A^{(1)}=\begin{bmatrix}2&2\\-\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{bmatrix}$ 的第 1 列 $b^{(2)}=\begin{bmatrix}2\\-\frac{1}{\sqrt{2}}\end{bmatrix}$，构造 T_2，使得 $T_2b^{(2)}=|b^{(2)}|e_1$

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

$$T_2=T_{12},\quad T_2A^{(1)}=\begin{bmatrix}\frac{3}{\sqrt{2}}&\frac{7}{3\sqrt{2}}\\0&\frac{4}{3}\end{bmatrix}$$

3. 构造上三角矩阵 R，计算正交矩阵 Q，那么 $A=QR$.

   最后，令 $T=\begin{bmatrix}1&\\&T_2\end{bmatrix}T_1$，则有

$$Q=T^T=\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{3\sqrt{2}}&-\frac{2}{3}\\0&\frac{4}{3\sqrt{2}}&\frac{1}{3}\\\frac{1}{\sqrt{2}}&-\frac{1}{3\sqrt{2}}&\frac{2}{3}\end{bmatrix},\quad R=\begin{bmatrix}2\sqrt{2}&\frac{3}{\sqrt{2}}&\frac{3}{\sqrt{2}}\\&\frac{3}{\sqrt{2}}&\frac{7}{3\sqrt{2}}\\&&\frac{4}{3}\end{bmatrix}$$

   则有 $A=QR$

4.2.3 代码实现

QR 分解

```python
import numpy as np

# 对矩阵A进行Gram-Schmidt正交化处理。
def gram_schmidt(A):
    n, m = A.shape
    Q = np.zeros((n, m))

    for j in  range(m):
        q = A[:, j]
        # 从Q中已计算的列中减去在这些列上的投影，确保正交性
        for i in  range(j):
            q = q - np.dot(Q[:, i], A[:, j]) * Q[:, i]
        q = q / np.linalg.norm(q)  # 归一化向量
        Q[:, j] = q

    return Q

# 对矩阵A进行QR分解。
def qr_decomposition(A):

    Q = gram_schmidt(A)
    R = np.dot(Q.T, A)
    return Q, R
```

4.3 矩阵的满秩分解

4.3.1 矩阵满秩分解的步骤推导

满秩分解： 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

$$A = FG$$

则称其为矩阵 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**.

设 $A\in C_R^{m\times n}(r>0)$. $rankA=r$ 时，根据矩阵的初等变换理论，对 A 进行初等行变换，可将 A 化为阶梯形矩阵 B，即

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

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

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

则有

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

其中 F 是列满秩矩阵，G 是行满秩矩阵.

故可以使用矩阵的初等行变换方法求矩阵的满秩分解：

1. $\left[\begin{array}{c:c} A & I \end{array}\right] \xrightarrow{\text{行}} \left[\begin{array}{c:c} B & P \end{array}\right]$，其中 B 为阶梯形矩阵
2. 计算 P^{-1}（或者 P^{-1} 的前 r 列）
3. 取 F 为 P^{-1} 的前 r 列构成的列满秩矩阵，G 为 B 的前 r 行构成的行满秩矩阵，那么 $A=FG$

除此之外，我们还可以通过 **Hermite 标准形** 求解满秩分解.

Hermite 标准形： 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，且第一个非零元素是 1，而后 $m-1$ 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_i 列 $(i=1,\ 2,\ \ldots,\ r)$，则 $j_1<j_2<\cdots<j_r$;
3. B 中的 $j_1,\ j_2,\ \ldots,\ j_r$ 列为单位矩阵 I_m 的前 r 列;

那么就称 B 为 **Hermite 标准形**，即为初等变换意义下的行最简形.

拟 Hermite 标准形： 设 $B\in C_r^{m\times n}$ $(r>0)$，且满足：

1. B 的后 $m-r$ 行元素均为零;
2. B 中的 $j_1,\ j_2,\ \ldots,\ j_r$ 列为单位矩阵 I_m 的前 r 列;

那么就称 B 为 **拟 Hermite 标准形**.

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

为 B 的前 r 行构成的 $r\times n$ 矩阵.

由 $A \xrightarrow{\text{行}} B$ 知，存在 m 阶可逆矩阵 P，使得 $PA=B$，或者 $A=P^{-1}B$，可将 P^{-1} 分块为

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

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

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

$$P_1=(e_{j_1},\ \dots,\ e_{j_r})$$

其中 e_j 表示单位矩阵 I_n 的第 j 个列向量，则有

$$GP_1=I_r,\quad AP_1=(FG)P_1=F(GP_1)=F$$

即 F 为 A 的 j_1，j_2，$\dots$，j_r 列构成的矩阵.

具体步骤如下所示：

1. $A \xrightarrow{\text{行}} B$，其中 B 为 Hermite 标准形矩阵，且 B 的 j_1，j_2，$\dots$，j_r 列为单位矩阵 I_m 的前 r 列;
2. 取 F 为 A 的 j_1，j_2，$\dots$，j_r 列构成的列满秩矩阵，G 为 B 的前 r 行构成的行满秩矩阵，那么 $A=FG$.

4.3.2 举例展示求法

题目.

求下列各矩阵的满秩分解：

$$(1)\quad \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 1 & -1 \\ 1 & 0 & 2 & 1 \end{bmatrix};\quad (2)\quad \begin{bmatrix} 1 & -1 & 1 & 1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \end{bmatrix}$$

解答.

逆矩阵方法:

1. $\left[\begin{array}{c:c} A & I \end{array}\right] \xrightarrow{\text{行}} \left[\begin{array}{c:c} B & P \end{array}\right]$，其中 B 为阶梯形矩阵

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

则有

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

2. 计算 P^{-1}（或者 P^{-1} 的前 r 列）

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

3. 取 F 为 P^{-1} 的前 r 列构成的列满秩矩阵，G 为 B 的前 r 行构成的行满秩矩阵，那么 $A=FG$

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

故有

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

Hermite 标准形方法:

1. $A\xrightarrow{\text{行}}B$，其中 B 为 Hermite 标准形矩阵，且 B 的 j_1，j_2，$\ldots$，j_r 列为单位矩阵 I_m 的前 r 列;

$$A\xrightarrow{\text{行}}\begin{bmatrix}1&0&0&0\\0&1&-1&-1\\0&0&0&0\\0&0&0&0\end{bmatrix}=B$$

2. 取 F 为 A 的 j_1，j_2，$\ldots$，j_r 列构成的列满秩矩阵，G 为 B 的前 r 行构成的行满秩矩阵，那么 $A=FG$.

$rankB=2$ 且 B 中的第 1 列和第 2 列为单位矩阵的前两列，故

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

4.3.3 代码实现

满秩分解

```python
import numpy as np

def full_rank_decomposition(A):
    # 计算矩阵A的秩
    rank = np.linalg.matrix_rank(A)

    U, S, VT = svd_decomposition(A)

    # 构造满秩矩阵F和G
    F = U[:, :rank] @ np.diag(S[:rank])
    G = Vt[:rank, :]

    return F, G
```

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

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

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

$$U^HAV = \begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}$$

其中 $\Sigma = diag(\sigma_1,\ \sigma_2,\ \ldots,\ \sigma_r)$，而 $\sigma_i(i=1,\ 2,\ \ldots,\ r)$ 为矩阵 A 的全部非零奇异值.

记 Hermite 矩阵 A^HA 的特征值为

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

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 V 分块为

$$V = [V_1|V_2], \quad V_1 \in C_r^{n\times r},\ V_2 \in C_{n-r}^{n\times(n-r)}$$

这样，可以将上式改写为

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

则有

$$A^HAV_1 = V_1\Sigma^2, \quad A^HAV_2 = O$$

故有

$$\begin{aligned} V_1^HA^HAV_1 = \Sigma^2 \quad &or \quad (AV_1\Sigma^{-1})(AV_1\Sigma^{-1}) = I_r \\ (AV_2)^H(AV_2) = O \quad &or \quad AV_2 = O \end{aligned}$$

令 $U_1 = AV_1\Sigma^{-1}$，则 $U_1^HU_1 = I_r$，即 U_1 的 r 个列是两两正交的单位向量，记作 $U_1 = (u_1,\ u_2,\ \dots,\ u_r)$. 将其扩充为 C^m 的标准正交基，记增添的向量为 $u_{r+1},\ \dots,\ u_m$，并构造矩阵 $U_2 = (u_{r+1},\ \dots,\ u_m)$，则

$$U = \left[\begin{array}{c:c} U_1 & U_2 \end{array}\right] = (u_1,\ u_2,\ \dots,\ u_r,\ u_{r+1},\ \dots,\ u_m)$$

是 m 阶酉矩阵，且有

$$U_1^HU_1 = I_r, \quad U_2^HU_1 = O$$

于是可得

$$U^HAV = U^H\left[\begin{array}{c:c} AV_1 & AV_2 \end{array}\right] = \begin{bmatrix} U_1^H \\ U_2^H \end{bmatrix}\left[\begin{array}{c:c} U_1\Sigma & O \end{array}\right] = \begin{bmatrix} U_1^HU_1\Sigma & O \\ U_2^HU_1\Sigma & O \end{bmatrix} = \begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}$$

故有

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

称上式为矩阵 A 的 **奇异值分解**

总结求解过程如下:

1. 求酉矩阵 $V_{n\times n}$，使得

$$V^H(A^HA)V = diag(\lambda_1,\ \dots,\ \lambda_r,\ 0,\ \dots,0)$$

2. 计算 $U_1 = AV_1\Sigma^{-1}$，其中 V_1 为 V 的前 r 列构成的矩阵
3. 扩充 U_1 的 r 个列向量为 C^m 的标准正交基，并记由增加的 $m-r$ 个列向量构成的矩阵为 U_2，那么 $U = \left[\begin{array}{c:c} U_1 & U_2 \end{array}\right]$ 是酉矩阵
4. 写出 A 的奇异值分解 $A = U\begin{bmatrix}\Sigma & O\\ O & O\end{bmatrix}V^H$

4.4.2 举例展示求法

题目.

求 $A = \begin{bmatrix}1 & 0\\ 0 & 1\\ 1 & 1\end{bmatrix}$ 的奇异值分解.

解答.

1. 求酉矩阵 $V_{n\times n}$，使得 $V^H(A^HA)V = diag(\lambda_1,\ \dots,\ \lambda_r,\ 0,\ \dots,0)$

   计算

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

   求得 B 的特征值为 $\lambda_1 = 3,\ \lambda_2 = 1$，对应的特征向量依次为

$$\xi_1 = \begin{bmatrix}1\\ 1\end{bmatrix},\quad \xi_2 = \begin{bmatrix}-1\\ 1\end{bmatrix}$$

   于是可得

$$rankA = 2,\quad \Sigma = \begin{bmatrix}\sqrt{3} & 0\\ 0 & 1\end{bmatrix},\quad V = \begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}$$

2. 计算 $U_1 = AV_1\Sigma^{-1}$，其中 V_1 为 V 的前 r 列构成的矩阵

   有 $V_1 = V$，计算

$$U_1 = AV_1\Sigma^{-1} = \begin{bmatrix}\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}}\\ \frac{2}{\sqrt{6}} & 0\end{bmatrix}$$

3. 扩充 U_1 的 r 个列向量为 C^m 的标准正交基，并记由增加的 $m-r$ 个列向量构成的矩阵为 U_2，那么 $U=\left[\begin{array}{c:c} U_1 & U_2 \end{array}\right]$ 是酉矩阵

取 $U_2=\begin{bmatrix} -\frac{1}{\sqrt{3}} & \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}$ 构造正交矩阵

$$U=\left[\begin{array}{c:c} U_1 & U_2 \end{array}\right]=\begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix}$$

4. 写出 A 的奇异值分解 $A=U\begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}V^H$

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

4.4.3 代码实现

SVD 分解

```python
import numpy as np

def svd_decomposition(A):
    # 计算 A^T * A 和 A * A^T
    ATA = np.dot(A.T, A)
    AAT = np.dot(A, A.T)

    # 计算特征值和特征向量
    eig_vals_ATA, eig_vecs_ATA = np.linalg.eig(ATA)
    eig_vals_AAT, eig_vecs_AAT = np.linalg.eig(AAT)

    # 对特征值进行排序并计算奇异值
    singular_values = np.sqrt(np.sort(eig_vals_ATA)[::-1])

    # 对应的特征向量是奇异向量
    U = eig_vecs_AAT
    VT = eig_vecs_ATA.T

```

```
    # 构造对角矩阵 Sigma
    Sigma = np.zeros(A.shape)
    np.fill_diagonal(Sigma, singular_values)

    return U, Sigma, VT
```

4.5 利用矩阵分解求矩阵广义逆

4.5.1 矩阵广义逆介绍

逆矩阵的概念只是对可逆矩阵才有意义. 但是在实际问题中，遇到的矩阵不一定是方阵，即便是方阵也不一定可逆，这就需要考虑，可否将逆矩阵概念进一步推广，为此，引进下列条件:

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

称满足以上三个条件的矩阵为 **广义逆矩阵**.

Penrose 的广义逆矩阵：设矩阵 $A \in C^{m\times n}$，若矩阵 $X \in C^{n\times m}$ 满足以下 4 个 Penrose 方程 则称 X 为 A 的 **Moore-Penrose 逆**，记为 A^+.

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

易知一下特例:

1. 若 A 是可逆矩阵，则 $A^+ = A^{-1}$;
2. 若 $A = O_{m\times n}$，则 $A^+ = O_{n\times m}$;
3. 若 $A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$，则 $A^+ = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$

矩阵 Moore-Penrose 逆的存在性:

1. 对于任意 $A \in C^{m\times n}$，A^+ 存在并且唯一;
2. 设 $A \in C_r^{m\times n}$ 的不可逆值分解为

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

那么

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

4.5.2 利用矩阵满秩分解求矩阵广义逆

通过证明上述存在性定理一，我们可以完成利用矩阵满秩分解求矩阵广义逆的推导.

设 $rankA = r$. 若 $r = 0$，则 $A = O_{m\times n}$，则 $A^+ = O_{n\times m}$；若 $r > 0$，由矩阵满秩分解的存在性定理（设 $A \in C_r^{m\times n}(r > 0)$，则 A 有满秩分解式），A 可进行满秩分解：

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

令 $X = G^+F^+$，则有

$$\begin{aligned} AXA &= FG\cdot G^+F^+\cdot FG = FG = A \\ XAX &= G^+F^+\cdot FG\cdot G^+F^+ = G^+F^+ = X \\ (AX)^H &= (FG\cdot G^+F^+)^H = (FF^+)^H = FF^+ = F\cdot GG^+\cdot F^+ = AX \\ (XA)^H &= (G^+F^+\cdot FG)^H = (G^+G)^H = G^+G = G^+\cdot F^+F\cdot G = XA \end{aligned}$$

故

$$A^+ = G^+F^+ = G^H(F^HAG^H)^{-1}F^H$$

由上式我们可以总结出求矩阵的 Moore-Penrose 逆的满秩分解方法：

1. 求 A 的满秩分解 $A = FG$
2. 计算 $F^+ = (F^HF)^{-1}F^H$ 和 $G^+ = G^H(GG^H)^{-1}$
3. 计算 $A^+ = G^+F^+$

4.5.3 利用矩阵奇异值分解求矩阵广义逆

由上述存在性定理二，我们可以得出利用矩阵奇异值分解求矩阵广义逆的推导：

令 $\Sigma = \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix}$，则 $\Sigma^+ = \begin{bmatrix} \Sigma_r^{-1} & O \\ O & O \end{bmatrix}$

$$\begin{aligned} A^+ &= A^H(AA^H)^+ = (U\Sigma V^H)^H(U\Sigma V^H V\Sigma^H U^H) \\ &= V\Sigma^H U^H[(U\Sigma\Sigma^H U^H)^+ \\ &= V[(U\Sigma)^H)(U\Sigma(U\Sigma)^H)^+] \\ &= V(U\Sigma)^+ = V\Sigma^+ U^H \end{aligned}$$

由此可以总结出求矩阵的 Moore-Penrose 逆的奇异值分解方法：

1. 求 A 的奇异值分解 $A = U \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix}_{m\times n} V^H$
2. 计算 $A^+ = V \begin{bmatrix} \Sigma_r^{-1} & O \\ O & O \end{bmatrix}_{n\times m} U^H$

矩阵的 Moore-Penrose 逆是一种广义逆矩阵，它满足 4 个 Penrose 方程. 下面介绍满足一个或几个 Penrose 方程的广义逆矩阵：

设矩阵 $A \in C^{m\times n}$，矩阵 $X \in C^{n\times m}$

1. 若 X 满足 Penrose 方程中的第 (i) 个方程，则称 X 为 A 的 $\{i\}$-逆，记作 $A^{(i)}$，全体 $\{i\}$-逆的集合记作 $A\{i\}$. 这种广义逆矩阵共有 4 类;
2. 若 X 满足 Penrose 方程中的第 (i), (j) 个方程 $(i \neq j)$，称 X 为 A 的 $\{i,j\}$-逆，记作 $A^{(i,j)}$，全体 $\{i,j\}$-逆的集合记作 $A\{i,j\}$. 这种广义逆矩阵共有 6 类;
3. 若 X 满足 Penrose 方程中的第 (i), (j) (k) 个方程 (i, j, k **互异**)，称 X 为 A 的 $\{i,j,k\}$-逆，记作 $A^{(i,j,k)}$，全体 $\{i,j,k\}$-逆的集合记作 $A\{i,j,k\}$. 这种广义逆矩阵共有 4 类;

其中，应用较为广泛的广义逆矩阵有以下 5 种：

$$A\{1\}, \quad A\{1,2\}, \quad A\{1,3\}, \quad A\{1,4\}, \quad A^+$$

由于任意矩阵 $A \in C_r^{m\times n}(r > 0)$ 都可通过初等行变换化为 (拟) Hermite 标准形 B，即存在有限个初等矩阵的乘积，记作 Q，使得 $QA = B$. 根据矩阵 B，构造置换矩阵 (交换单位矩阵的列向量构成的矩阵) P，使得

$$QAP = \begin{bmatrix} I_r & K \\ O & O \end{bmatrix}$$

其中 K 是 $r \times (n-r)$ 子矩阵.

利用矩阵的 (拟) Hermite 标准形，容易求得矩阵的 $\{1\}$-逆和 $\{1,2\}$-逆.

由上式可知

$$QAP = \begin{bmatrix} I_r & K \\ O & O \end{bmatrix}, \quad A = Q^{-1}\begin{bmatrix} I_r & K \\ O & O \end{bmatrix}P^{-1}$$

则有

$$X = P\begin{bmatrix} I_r & O \\ O & L \end{bmatrix}Q, \quad X = P\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}Q$$

分别为 A 的 $\{1\}$-逆和 $\{1,2\}$-逆.

总结方法如下：

1. $\left[\begin{array}{c:c} A & I \end{array}\right] \xrightarrow{\text{行}} \left[\begin{array}{c:c} B & Q \end{array}\right]$，其中 B 为 Hermite 标准形矩阵
2. 构造置换矩阵 P，使得 $BP = \begin{bmatrix} I_r & K \\ O & O \end{bmatrix}_{m\times n}$
3. 计算 $X = P\begin{bmatrix} I_r & O \\ O & L \end{bmatrix}_{n\times m}Q$ 或 $X_0 = P\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}_{n\times m}Q$ 那么 $X \in A\{1\}$，$X_0 \in A\{1,2\}$.

4.5.4 举例展示求法

题目.

设 $A = \begin{bmatrix} 2 & 1 & 0 & 2 \\ 0 & 0 & 1 & 2 \\ 2 & 1 & 1 & 4 \end{bmatrix}$，分别求 $A^{(1)}, A^{(1,2)}$ 及 A^{+}.

解答.

1. $\left[\begin{array}{c:c} A & I \end{array}\right] \xrightarrow{\text{行}} \left[\begin{array}{c:c} B & Q \end{array}\right]$

$$\left[\begin{array}{c:c} A & I \end{array}\right] \xrightarrow{\text{行}} \begin{array}{cccc:ccc} 2 & 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & -1 \end{array}$$

故 A 的拟 Hermite 标准形为

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

2. 构造置换矩阵 P，使得 $BP=\begin{bmatrix}I_r&K\\O&O\end{bmatrix}_{m\times n}$

   构造可逆矩阵

$$Q=\begin{bmatrix}1&0&0\\0&1&0\\-1&-1&1\end{bmatrix},\quad P=(e_2,\ e_3,\ e_1,\ e_4)=\begin{bmatrix}0&0&1&0\\1&0&0&0\\0&1&0&0\\0&0&0&1\end{bmatrix}$$

3. 计算 $X=P\begin{bmatrix}I_r&O\\O&L\end{bmatrix}_{n\times m}Q$ 或 $X_0=P\begin{bmatrix}I_r&O\\O&O\end{bmatrix}_{n\times m}Q$

$$A^{(1)}=P\left[\begin{array}{cc:c}1&0&0\\0&1&0\\\hdashline 0&0&a\\0&0&b\end{array}\right]Q=\begin{bmatrix}-a&-a&a\\1&0&0\\0&1&0\\-b&-b&b\end{bmatrix},\quad A^{(1,2)}=\begin{bmatrix}0&0&0\\1&0&0\\0&1&0\\0&0&0\end{bmatrix}$$

下面求 A^+

1. 求 A 的满秩分解 $A=FG$

   因为拟 Hermite 标准形 B 的第 2 列和第 3 列构成 I_3 的前两列，所以 F 为 A 的第 2 列和第 3 列构成的 3×2 矩阵，从而有

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

2. 计算 $F^+ = (F^HF)^{-1}F^H$ 和 $G^+ = G^H(GG^H)^{-1}$

$$F^+ = (F^HF)^{-1}F^H = \frac{1}{3}\begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix}$$

$$G^+ = G^H(GG^H)^{-1} = \frac{1}{29}\begin{bmatrix} 10 & -8 \\ 5 & -4 \\ -4 & 9 \\ 2 & 10 \end{bmatrix}$$

3. 计算 $A^+ = G^+F^+$

$$A^+ = G^+F^+ = \frac{1}{87}\begin{bmatrix} 28 & -26 & 2 \\ 14 & -13 & 1 \\ -17 & 22 & 5 \\ -6 & 18 & 12 \end{bmatrix}$$