

# 矩阵函数的求法和矩阵分解方法研究

谢牧航 2022211363

## 摘 要

本论文综合研究了矩阵函数的求解方法和矩阵分解技术，涵盖了从传统理论到现代计算工具的广泛内容。首先，对经典的矩阵函数求解方法如待定系数法、数项级数求和法、对角型法以及若尔当标准型法进行了详细的阐述和深入的探讨。特别地，本文引入若尔当标准型法，扩展了求解矩阵函数的适用范围，并增强了理论在实际应用中的有效性。

在矩阵分解方面，论文不仅详尽讨论了 LU、QR、满秩和奇异值分解等经典分解方法，还包括了 PLU 分解等较新的技术。通过数学推导和实例演示，论文明确了每种方法背后的理论基础，证明了这些技术在实际问题中的应用价值和效能。

论文的一个显著特点是结合了理论分析和计算实现。它不仅提供了矩阵分解方法的理论框架，还提供了相应的 Python 实现代码，使得理论知识更加易于理解和应用。这些编程示例旨在帮助读者更好地掌握矩阵理论，并为其学习和实际应用提供支持。

此外，本论文还探讨了奇异值分解和广义逆在最小二乘法和最优化问题中的应用，突出了这些先进数学工具在数据分析和统计建模中的重要性。深入分析它们如何帮助解决现实世界问题，展示了矩阵理论与实践结合的研究视角。

综上所述，本论文总结了本学期矩阵理论课程的学习成果，并对矩阵理论的研究和实际应用提出了笔者的理解和方法。

**关键词** 矩阵理论 矩阵函数 矩阵分解 广义逆

# Research on the Methods of Matrix Function Calculation and Matrix Decomposition

Xie Muhang 2022211363

## ABSTRACT

This thesis comprehensively studies the methods for solving matrix functions and matrix decomposition techniques, covering a wide range from traditional theory to modern computational tools. Initially, classical methods for solving matrix functions such as the method of undetermined coefficients, power series summation, diagonalization, and Jordan canonical form are elaborately discussed and deeply explored. Notably, the introduction of the Jordan canonical form method has extended the applicability of solving matrix functions and enhanced the practical effectiveness of the theory.

In terms of matrix decomposition, the paper not only thoroughly discusses classic decomposition methods such as LU, QR, full rank, and singular value decomposition but also includes newer techniques like PLU decomposition. Through mathematical derivation and illustrative examples, the paper clarifies the theoretical basis behind each method, demonstrating their value and efficacy in solving practical problems.

A significant feature of the thesis is the integration of theoretical analysis with computational implementation. It provides not only the theoretical framework for matrix decomposition methods but also their corresponding Python code implementations, making the theoretical knowledge more understandable and applicable. These programming examples are intended to help readers grasp matrix theory better and provide support for their learning and practical application.

Additionally, the thesis explores the application of singular value decomposition and generalized inverses in least squares methods, highlighting the importance of these advanced mathematical tools in data analysis and statistical modeling. By analyzing how they contribute to solving real-world problems, the thesis presents a research perspective that combines matrix theory with practice.

In summary, this thesis summarizes the learning outcomes of the matrix theory course for this semester and proposes new insights and methods for the study and practical application of matrix theory.

**KEY WORDS** Matrix Theory; Matrix Functions; Matrix Decomposition; Generalized Inverse

# 目录

<b>第一章 引言 .....</b>	<b>5</b>
<b>1.1 背景介绍 .....</b>	<b>5</b>
1.1.1 矩阵理论与方法介绍 .....	5
1.1.2 函数矩阵和矩阵函数介绍 .....	5
1.1.3 线性代数方程组求解介绍 .....	6
<b>1.2 问题介绍 .....</b>	<b>6</b>
1.2.1 矩阵函数的求法问题介绍 .....	6
1.2.2 矩阵分解的方法问题介绍 .....	7
<b>1.3 上述问题国内外研究成果介绍 .....</b>	<b>7</b>
1.3.1 矩阵函数的求法研究现状 .....	7
1.3.2 矩阵分解方法研究现状 .....	8
<b>1.4 本论文工作简述 .....</b>	<b>8</b>
1.4.1 本论文对上述问题研究简述 .....	8
1.4.2 本论文创新点或特点简述 .....	9
1.4.3 本论文撰写结构简述 .....	9
<b>第二章 预备知识 .....</b>	<b>10</b>
<b>2.1 欧式空间与线性变换 .....</b>	<b>10</b>
2.1.1 欧式空间与线性变换介绍 .....	10
2.1.2 若尔当标准型的求解 .....	12
2.1.3 欧式空间中线性变换的求法 .....	13
<b>2.2 向量范数与矩阵范数 .....</b>	<b>19</b>
2.2.1 向量范数介绍 .....	19
2.2.2 矩阵范数介绍 .....	20
2.2.3 矩阵可逆性条件、条件数和谱半径介绍 .....	21
<b>2.3 矩阵函数介绍 .....</b>	<b>22</b>
2.3.1 矩阵序列介绍 .....	22
2.3.2 矩阵级数介绍 .....	23
2.3.3 矩阵函数介绍 .....	24
2.3.4 函数矩阵对矩阵的导数与积分 .....	24
<b>第三章 矩阵函数的求法研究 .....</b>	<b>26</b>
<b>3.1 待定系数法 .....</b>	<b>26</b>
3.1.1 待定系数法求矩阵函数的步骤推导 .....	26

3.1.2 举例展示求法 .....	26
<b>3.2 数项级数求和法 .....</b>	<b>27</b>
3.2.1 数项级数求和法求矩阵函数的步骤推导 .....	27
3.2.2 举例展示求法 .....	28
<b>3.3 对角型法 .....</b>	<b>29</b>
3.3.1 对角型法求矩阵函数的步骤推导 .....	29
3.3.2 举例展示求法 .....	29
<b>3.4 若尔当标准型法 .....</b>	<b>30</b>
3.4.1 若尔当标准型法求矩阵函数的步骤推导 .....	30
3.4.2 举例展示求法 .....	31
<b>第四章 矩阵分解方法研究 .....</b>	<b>33</b>
<b>4.1 矩阵的 LU 分解 .....</b>	<b>33</b>
4.1.1 矩阵 LU 分解的步骤推导 .....	33
4.1.2 举例展示求法 .....	35
<b>4.2 矩阵的 QR 分解 .....</b>	<b>37</b>
4.2.1 矩阵 QR 分解的步骤推导 .....	37
4.2.2 举例展示求法 .....	42
<b>4.3 矩阵的满秩分解 .....</b>	<b>45</b>
4.3.1 矩阵满秩分解的步骤推导 .....	45
4.3.2 举例展示求法 .....	46
<b>4.4 矩阵的奇异值分解 .....</b>	<b>46</b>
4.4.1 矩阵奇异值分解的步骤推导 .....	46
4.4.2 举例展示求法 .....	50
<b>4.5 利用矩阵分解求矩阵广义逆 .....</b>	<b>52</b>
4.5.1 矩阵广义逆介绍 .....	52
4.5.2 利用矩阵满秩分解求矩阵广义逆 .....	52
4.5.3 利用矩阵奇异值分解求矩阵广义逆 .....	53
4.5.4 举例展示求法 .....	53
<b>第五章 总结 .....</b>	<b>55</b>
<b>参考文献 .....</b>	<b>56</b>

# 第一章 引言

## 1.1 背景介绍

### 1.1.1 矩阵理论与方法介绍

矩阵理论是数学的一个重要分支，它在科学和工程的许多领域中都有着广泛的应用。矩阵，简而言之，是一个由数字、符号或表达式排列成的矩形阵列。这些元素被称为矩阵的“条目”或“元素”，可以用于表示和解决复杂的数学问题，如线性方程组的求解、向量空间的表示、线性变换等。

矩阵的基本运算包括加法、乘法、转置和求逆等。矩阵加法和乘法遵循特定的规则，这些规则是线性代数的核心。例如，两个矩阵的加法只有在它们具有相同的维度时才有定义。矩阵乘法则更为复杂，涉及行列间的乘积和求和。

矩阵可以按照不同的特性进行分类，如方阵（行数和列数相等的矩阵）、对称矩阵（转置与原矩阵相等的矩阵）、零矩阵（所有元素都为零的矩阵）等。不同类型的矩阵有着不同的性质和应用。

矩阵的重要性质包括秩、行列式和迹。秩描述了矩阵中线性独立行或列的最大数量，是解决线性方程组的关键。行列式与矩阵是否可逆密切相关，而迹（矩阵对角线上元素的和）在许多数学和物理问题中非常有用。

矩阵函数是将常规函数（如指数函数、对数函数）扩展到矩阵上的结果。矩阵分解，如 LU 分解、QR 分解和奇异值分解，是将矩阵分解为几个特定形式的乘积，这些分解在数值分析、数据压缩和信号处理等领域有着广泛应用。

矩阵理论在现代科学和工程学中扮演着关键角色。例如，在计算机视觉、量子物理、经济模型分析以及机器学习等领域，矩阵被用来表示和处理复杂的数据结构。

通过深入研究矩阵理论与方法，我们不仅能够更好地理解多维数据的结构和特性，还可以开发出更有效的计算算法，以解决实际问题。

### 1.1.2 函数矩阵和矩阵函数介绍

函数矩阵是一种特殊类型的矩阵，其元素由函数而非常规数字组成。这些函数通常是同一个变量的函数，例如时间或空间变量。函数矩阵在处理动态系统、控制理论和工程问题中特别有用。例如，在控制理论中，系统的动态可以通过一个以时间为变量的函数矩阵来描述。

与函数矩阵不同，矩阵函数是将普通的数学函数（如指数、对数、三角函数等）应用于矩阵。这意味着输入是一个矩阵，而输出也是一个矩阵。矩阵函数的概念对于理解和处理线性代数中的复杂问题至关重要。

矩阵函数的计算通常涉及到矩阵的特征值和特征向量。一个常用的方法是通过矩阵对角化或 Jordan 标准形来简化函数的计算。例如，如果矩阵可对角化，那么计算矩阵的指数函数可以通过计算其特征值的指数来实现。

矩阵函数在许多领域都有广泛应用。例如，在量子力学中，矩阵指数函数被用于描述系统的时间演化。在统计学中，矩阵对数函数可以用来计算多变量正态分布的概率密度函数。

理解函数矩阵和矩阵函数的概念及其在实际问题中的应用，对于深入探索线性代数以及其在各个科学和工程领域中的应用至关重要。这不仅涉及到数学理论，还包括计算方法和实际应用，形成理论与实践的完美结合。

### 1.1.3 线性代数方程组求解介绍

线性代数方程组是由多个线性方程构成的集合，通常用来描述一系列线性关系。这些方程组在数学、物理学、工程学和计算科学等领域都有广泛应用。

一个线性方程组可以表示为  $Ax = b$  的形式，其中  $A$  是一个已知的矩阵， $x$  是一个未知向量， $b$  是一个已知向量。解这个方程组意味着找到能满足所有方程的  $x$  的值。

线性方程组的求解方法主要分为两类：直接方法和迭代方法。

直接方法如：高斯消元法，通过行操作将矩阵转换为行阶梯形式，然后进行回代求解；LU 分解，将矩阵分解为一个下三角矩阵和一个上三角矩阵的乘积，便于求解。

迭代方法如：雅可比方法（Jacobi method），适用于对角占优矩阵，通过逐次迭代改进解的估计值；高斯-塞德尔方法（Gauss-Seidel method），类似于雅可比方法，但每次迭代时使用已更新的值。

线性方程组可能有唯一解、无解或无限多解。这取决于矩阵  $A$  的性质和方程组的构成。例如，如果  $A$  是非奇异（可逆）的，那么方程组有唯一解。

线性方程组在实际应用中非常常见。例如，在工程学中，它们用于建模物理系统；在经济学中，用于描述经济模型；在计算科学中，用于算法设计和数据分析。

在实际计算中，选择合适的方法求解线性方程组很重要，这不仅关乎计算的精确性，还涉及计算过程的数值稳定性和效率。

## 1.2 问题介绍

### 1.2.1 矩阵函数的求法问题介绍

矩阵函数的求法是线性代数和矩阵理论中的一个关键课题。它涉及将传统的数学函数（如指数、对数、幂函数等）应用于矩阵。这一领域面临的主要挑战在于矩阵并非单一数值，因此传统的函数求法需要相应地进行调整和扩展。

1. 待定系数法：基于 Hamilton-Cayley 定理，对函数改变形态后，对于多个特征值得到不同的式子，通过待定系数法求解。

2. 数项级数求和法：根据最小多项式，导出矩阵递推关系，从而转换为数项级数求和的问题。
3. 对角型法：如果矩阵可以对角化，即可以表示为一组特征值和对应特征向量的形式，那么矩阵函数可以通过对这些特征值应用函数并重新组合特征向量来求得。
4. 若尔当标准型法：对于不能对角化的矩阵，Jordan 标准型法提供了一种求解矩阵函数的方法。它涉及将矩阵转化为 Jordan 形式，然后在这种形式下计算函数。

以上是几种主要的矩阵函数求法。

除此之外，还有一些在有理数域的函数求值方法也可以类比到矩阵函数的求解，如帕德近似、积分表示法等。

### 1.2.2 矩阵分解的方法问题介绍

矩阵分解是数学和计算数学中的一个重要话题，它涉及将一个矩阵分解为多个具有特定性质的矩阵的乘积。这些方法在解决线性方程组、计算矩阵的逆、降维和数据分析等方面发挥着关键作用。以下是几种常用的矩阵分解方法及其应用：

1. LU 分解：在线性代数与数值分析中，LU 分解是矩阵分解的一种，将一个矩阵分解为一个下三角矩阵和一个上三角矩阵的乘积，有时需要再乘上一个置换矩阵。LU 分解可以被视为高斯消元法的矩阵形式。在数值计算上，LU 分解经常被用来解线性方程组、且在求逆矩阵和计算行列式中都是一个关键的步骤。
2. QR 分解：QR 分解法是一种将矩阵分解的方式。这种方式，把矩阵分解成一个正交矩阵与一个上三角矩阵的积。QR 分解经常用来解线性最小二乘法问题。QR 分解也是特定特征值算法即 QR 算法的基础。
3. 满秩分解：如果矩阵是满秩的，它可以分解为一个列满秩矩阵和行满秩矩阵的乘积。这种分解有助于理解矩阵的结构和解决相关数学问题。
4. 奇异值分解 (SVD)：将矩阵分解为三个矩阵的乘积：一个正交矩阵、一个对角矩阵（包含奇异值）和另一个正交矩阵的转置。SVD 在信号处理、统计学和机器学习中极为重要。

这些矩阵分解方法不仅有助于理解矩阵的内部结构，还在计算上提供了便利。例如，它们可以用于：

- 求解线性方程组：通过简化计算过程，加快求解速度。
- 计算矩阵逆和广义逆：特别是在矩阵不是方阵或不满足秩时，分解方法可以用来求解广义逆。
- 降维和数据压缩：如在 SVD 中，通过提取数据的主要成分，用于图像压缩和降噪。
- 数据分析和机器学习：分解方法可用于特征提取和数据的结构化表示。

## 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

矩阵函数的研究是线性代数和计算数学领域内的一个活跃领域。随着计算技术的发展和应用领域的扩展，对矩阵函数求法的研究也在不断进步。

早期的研究集中在理解和定义矩阵函数的基本性质，如矩阵指数函数、对数函数和幂函数等。这包括对矩阵函数的泰勒级数展开、特征值分析和对角化等方面的研究。

近年来，随着计算机科学的发展，对矩阵函数的数值计算方法进行了深入研究。这涉及开发更高效、更稳定的算法，以在保持计算准确性的同时减少计算成本。例如，Krylov 子空间方法在近似矩阵指数函数和其他相关函数方面表现出色。

矩阵函数的求法在许多应用领域中变得越来越重要。在量子物理、控制理论、图像处理和机器学习等领域，矩阵函数的有效计算对于理解和解决复杂的实际问题至关重要。

矩阵函数的求法研究不仅限于单一国家或机构，而是一个国际合作的成果。许多国家的研究者在这一领域做出了贡献，发表了大量关于矩阵函数理论和算法的论文。

当前的研究挑战包括开发适用于大规模矩阵的算法，处理非常规矩阵（如稀疏矩阵、非对称矩阵等）的函数计算，以及提高算法的并行化和可伸缩性。

### 1.3.2 矩阵分解方法研究现状

矩阵分解方法是数学、计算机科学以及工程领域的关键技术，近年来在理论和应用方面都取得了显著进展。

LU 分解、QR 分解、满秩分解和奇异值分解（SVD）等传统矩阵分解方法在理论上已相对成熟。这些方法的研究主要集中在提高计算效率、增强数值稳定性以及适应特殊类型矩阵（如稀疏矩阵、大规模矩阵）的需求。

随着现代计算环境的发展，特别是高性能计算和云计算的普及，矩阵分解方法的研究开始着重考虑算法的并行化和分布式计算。这涉及到算法在多核处理器和分布式计算环境中的优化。

除了传统方法，研究者还在探索新型的矩阵分解技术，如张量分解和非负矩阵分解（NMF）。这些新方法在处理更复杂的数据结构（如多维数据）和特定应用（如图像分析、文本挖掘）中显示出优势。

矩阵分解技术在多个应用领域都显示出其强大的能力，包括机器学习、数据挖掘、生物信息学和网络分析等。在这些领域，矩阵分解不仅用于数据降维和特征提取，还在模式识别和预测模型中扮演着关键角色。

矩阵分解方法的研究正变得越来越跨学科，涉及数学、计算机科学、统计学和各类应用科学。这种跨学科的合作推动了新算法的发展，同时也促进了理论与实际应用的紧密结合。

## 1.4 本论文工作简述

### 1.4.1 本论文对上述问题研究简述



本论文主要研究矩阵函数的求法和矩阵分解的方法。主要依据书籍《矩阵论》，探讨了矩阵函数的求法，如待定系数法、数项级数求和法、对角型法、若尔当标准型法等；研究了矩阵分解的方法，如 LU 分解、QR 分解、满秩分解和奇异值分解等。同时，继续探讨了矩阵的广义逆以及广义逆在求解线性系统中的应用。

#### 1.4.2 本论文创新点或特点简述

本论文在矩阵函数求解与矩阵分解方法研究方面，采用了创新性的研究路径与多元化的技术手段。不仅深入探讨了《矩阵论》中的传统方法，如相似对角化法，而且引入了若尔当标准型法，这一点使得研究内容适用于更广泛的线性变换问题，提高了理论方法的适用范围和实用价值。

在矩阵分解方面，论文不局限于传统的 LU 分解、QR 分解、满秩分解和奇异值分解等方法，还对 PLU 分解等先进分解技术进行了细致的研究和讨论。这些内容不但理论价值显著，也为工程实践提供了更多的选择和便利。

更重要的是，论文结合了当代计算技术的发展，特别是 Python 编程语言的广泛应用，提供了一系列的代码实现示例。这不仅降低了矩阵理论到实际应用的门槛，还为读者提供了直观、高效的学习工具，增强了理论知识的可操作性和实践性。

此外，论文介绍了奇异值分解和广义逆的在最小二乘法和最优化问题中的应用，展示了矩阵分解理论在统计学和数据科学领域中的强大功能和广泛用途。通过对这些高级话题的讨论，论文不仅扩展了矩阵理论的研究深度，也为跨学科领域提供了桥梁，使数学理论更加生动和实用。

#### 1.4.3 本论文撰写结构简述

本论文共分为五章，每章都围绕矩阵函数的求法和矩阵分解方法进行详细探讨。第一章为引言，介绍了矩阵论的基本概念及其在现代科学研究中的重要性，以及论文的研究目的和意义。第二章为预备知识，介绍了后面论述部分所必需的前置知识，如线性空间、线性变换等。

第三章对矩阵函数的求法进行了系统的阐述，包括待定系数法、数项级数求和法、对角型法和若尔当标准型法等，旨在为读者提供一系列求解矩阵函数的有效工具。

第四章深入研究了矩阵分解的理论和方法，重点讨论了 LU 分解、QR 分解、满秩分解以及奇异值分解。每一种分解方法都通过步骤推导和实际例子进行了详细说明，以帮助读者理解其数学原理和应用场景。第四章则专注于矩阵广义逆的求解技巧，展示了如何利用矩阵分解方法来求解广义逆，提供了一种在解决线性代数问题时的有效手段。

第五章是对整篇论文的总结，对矩阵函数的求法和矩阵分解方法进行了回顾和总结，并对未来的研究方向进行了展望。

此外，每章节末尾都提供了丰富的例题和习题，旨在加深读者对理论的理解和实践应用能力。本论文不仅提供了矩阵论的深入分析，同时也为数学爱好者和研究人员提供了一个实用的数学工具箱。

## 第二章 预备知识

### 2.1 欧式空间与线性变换

#### 2.1.1 欧式空间与线性变换介绍

##### 2.1.1.1 线性空间

要介绍欧式空间，我们首先需要介绍线性空间。

设  $V$  是一个非空集合，它的元素用  $x, y, z$  等表示，并称之为向量； $K$  是一个数域，它的元素用  $k, l, m$  等表示，如果  $v$  满足下列条件：

1. 在  $V$  中定义一个加法运算，即当  $x, y \in V$  时，有唯一的和  $x + y \in V$  且加法运算满足下列性质

- 结合律  $x + (y + z) = (x + y) + z$
- 交换律  $x + y = y + x$
- 存在零元素  $0$ ，使得  $x + 0 = x$
- 存在负元素，即对任一向量  $x \in V$ ，存在向量  $y \in V$ ，使  $x + y = 0$ ，则称  $y$  为  $x$  的负元素，记为  $-x$ ，于是有

$$x + (-x) = 0$$

2. 在  $V$  中定义数乘（数与向量的乘法）运算，即当  $x \in V$  时，有唯一的  $kx \in V$ ，且数乘运算满足下列性质：

- 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**。

在线性代数中， $R(A)$  和  $N(A)$  分别代表矩阵  $A$  的像空间（range）和零空间（null space）。

- $R(A)$  或  $\text{Im}(A)$  (Image of  $A$ )：矩阵  $A$  的像空间或列空间是由  $A$  的所有列向量的线性组合构成的空间。换句话说，它是  $A$  作用于  $\mathbb{R}^n$  下的所有可能结果的集合，记为  $\{Ax \mid x \in \mathbb{R}^n\}$ 。
- $N(A)$  或  $\text{Ker}(A)$  (Kernel of  $A$ )：矩阵  $A$  的零空间或核是所有映射到零向量的原像集合，即  $\{x \mid Ax = 0\}$ 。

##### 2.1.1.2 欧式空间

欧式空间是一种特殊的线性空间，其定义如下。

设  $V$  是实数域  $\mathbb{R}$  上的线性空间，对于  $V$  中任意两个向量  $x$  与  $y$ ，若函数  $\langle x, y \rangle$  满足以下四个条件，则  $\langle x, y \rangle$  称为内积，且  $V$  称为内积空间：

- 交换律： $\langle x, y \rangle = \langle y, x \rangle$

- 分配律:  $\langle x, y+z \rangle = \langle x, y \rangle + \langle x, z \rangle$
- 齐次律:  $\langle kx, y \rangle = k\langle x, y \rangle$  (对于所有  $k \in \mathbb{R}$ )
- 非负性:  $\langle x, x \rangle \geq 0$ , 且当且仅当  $x = 0$  时,  $\langle x, x \rangle = 0$

则函数  $\langle x, y \rangle$  称为  $x$  与  $y$  的内积, 而  $V$  称为 **Euclid 空间**, 简称**欧氏空间**或**实内积空间**。

欧式空间是定义了内积的实线性空间。注意到, 因为向量的内积与向量的线性运算是彼此无关的运算, 所以不论内积如何规定, 都不会影响该实线性空间的维数。欧氏空间的子空间仍然是欧氏空间。

欧氏空间是矩阵理论中的一个重要概念, 它为描述具有内积和度量的几何结构提供了基础。在欧氏空间中, 我们考虑的是具有欧氏内积的实数向量空间。这个内积定义了向量之间的夹角和长度, 是一种通过点积来度量向量之间相对位置的方式。

欧氏空间的关键性质是满足线性空间和度量空间的要求, 其中线性空间允许向量的线性组合, 而度量空间则提供了测量向量之间距离的工具。在欧氏空间中, 内积的引入使得我们能够定义向量的模长、夹角以及与内积相关的投影等概念, 这为几何和线性代数的研究提供了坚实的数学基础。

通过矩阵的表示, 欧氏空间中的向量和操作可以被转化为线性代数的形式, 使得对向量的运算更加方便。欧氏空间的理论不仅在纯数学领域中有着广泛的应用, 同时也渗透到物理学、工程学以及计算机科学等多个领域, 为解决实际问题提供了强大的数学工具。

如果数域扩展到复数域, 会有对应的线性空间称为酉空间, 在此不做介绍。

### 2.1.1.3 线性变换

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  $x \in V$ ,  $V$  中都有唯一的向量  $y$  与之对应, 则称  $T$  是  $V$  的一个**变换**或**算子**。记为

$$Tx = y$$

称  $y$  为  $x$  在  $T$  下的**象**, 而  $x$  是  $y$  的**原象** (或**象源**)。

如果数域  $K$  上的线性空间  $V$  的一个变换  $T$  具有下列性质:

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

其中,  $k, l \in K$ ,  $x, y \in V$ , 则称  $T$  为  $V$  的一个**线性变换**或**线性算子**。

常见的线性变换有单位变换、零变换、投影变换、对称变换、反对称变换、旋转变换、剪切变换、伸缩变换、剪伸变换、对合变换、置换变换、微分变换、积分变换等。

设  $x_1, x_2, \dots, x_n$  是数域  $K$  上的线性空间  $V^n$  的一个基, 线性变换  $T_1, T_2$  在该基下依次有  $n$  阶矩阵分别为  $A, B$ , 则有结论:

- $(T_1 + T_2)(x_1, x_2, \dots, x_n) = (x_1, x_2, \dots, x_n)(A + B)$
- $(kT_1)(x_1, x_2, \dots, x_n) = (x_1, x_2, \dots, x_n)(kA)$
- $(T_1 T_2)(x_1, x_2, \dots, x_n) = (x_1, x_2, \dots, x_n)(AB)$

$$\bullet T_1^{-1}(x_1, x_2, \dots, x_n) = (x_1, x_2, \dots, x_n)(A^{-1})$$

线性变换可以用矩阵表示。

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

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

设线性变换  $T$  在线性空间  $V^n$  的基  $x_1, x_2, \dots, x_n$  下的矩阵为  $A = (a_{ij})$ , 向量  $x$  在该基下的坐标为  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$ , 则  $T(x)$  在该基下的坐标为  $\beta = (\beta_1, \beta_2, \dots, \beta_n)^T$  可按公式

$$\beta = A\alpha$$

来计算。

### 2.1.2 若尔当标准型的求解

若尔当标准型是线性代数中矩阵理论的一个重要概念, 它描述了方阵的一种特殊标准形式。对于任意一个方阵, 若尔当标准型的存在性是矩阵理论中的基本定理之一。

若尔当标准型的特点在于, 矩阵可以分解为若干个若尔当块的直和形式。每个若尔当块都是一个上三角矩阵, 对角线上有相同的特征值, 并且除了对角线上的元素为 1, 其他元素均为 0。若尔当标准型的存在性保证了任意一个方阵都可以通过相似变换化为这种形式。

这个标准型的重要性在于简化矩阵的分析和计算。通过将矩阵转化为若尔当标准型, 我们可以更容易地理解矩阵的特征和性质, 进而解决与矩阵相关的问题。若尔当标准型也为矩阵的对角化提供了一种特殊的方式, 使得对角化过程更加直观和可操作。

总体而言, 若尔当标准型是矩阵理论中的一种标准形式, 它的存在性和具体形式为解决线性代数和矩阵分析中的问题提供了强有力的工具。

Jordan 标准型是一个准对角矩阵, 其形式大约如下:

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

设  $A$  是  $n$  阶复矩阵, 且其特征多项式的某种分解式  $\varphi(\lambda) = \prod_{i=1}^k (\lambda - \lambda_i)^{m_i}$  存在, 则存在  $n$  阶复矩阵  $P$ , 使

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

其中

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

为  $(\lambda - \lambda_i)^{m_i}$  对应的 Jordan 块。

改用线性变换来说明, 设  $T$  是  $V^n$  上的线性变换, 则存在  $V^n$  的基  $x_1, x_2, \dots, x_n$ , 使得  $T$  在该基下的矩阵为若尔当标准型。这说明了一般矩阵的若尔当标准型的存在性。但是仍旧无法准确求出矩阵的若尔当标准型。下面讨论求解若尔当标准型的方法。

下面给出求出每个若尔当块的方法。

为了求解矩阵的若尔当标准型, 我们需要先用初等变换<sup>1</sup>将矩阵  $(\lambda I - A)$  化为标准型。

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

其中  $d_i(\lambda)$  是  $\lambda$  的  $m_i$  次多项式, 且  $d_1(\lambda) \mid d_2(\lambda) \mid \dots \mid d_s(\lambda)$ 。且  $d_{i(\lambda)} (i = 1, 2, \dots, s)$  是首 1 多项式 (意思是前面的几个  $d_i(\lambda)$  可能是 1)。

可以证明, 一个多项式矩阵的标准形式的对角线上的非零元素  $d_i(\lambda)$  是唯一的, 不随矩阵的初等变换而改变。因此通常称  $d_i(\lambda)$  为矩阵  $A$  的第  $i$  个**不变因子**或**不变因式**。

把  $A(\lambda)$  的每个次数大于零的不变因子  $d_i(\lambda)$  分解为不可约因式的乘积, 这样的不可约因式 (连同他们的幂指数) 称为  $A(\lambda)$  的一个**初等因子**, 初等因子的全体称为  $A(\lambda)$  的**初等因子组**。每个初等因子  $(\lambda - \lambda_i)^{m_i}$  对应一个若尔当块  $J_i(\lambda_i)$ 。

### 2.1.3 欧式空间中线性变换的求法

欧式空间中线性变换的求法指的是求解线性变换的矩阵表示的过程。具体形式可能如“设  $V$  是欧氏空间,  $T$  是  $V$  上的一个线性变换, 求  $z = (T^k)(x)$ , 其中  $x \in V$ ”。

其实我们发现这里有个难点是  $T^k$  的求法, 假如在某个基下,  $T$  对应的矩阵比较特殊, 这样对  $T^k$  的计算就更加容易。这里我们有两种选择, 一是**使该矩阵进行相似对角化**, 二是**使该矩阵进行若尔当标准型化**。而我们要求的就是对应的基。

下面我们通过例 2 来演示这种方法。

#### 2.1.3.1 相似对角化

我们首先的通过例 1 来说明相似对角化方法的求法。

**例 1** 在矩阵空间  $\mathbb{R}^{2 \times 2}$  中, 设矩阵  $A$  与  $B$  的内积定义为  $(A, B) = \text{tr}(A^T B)$ , 于是

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

$V$  中的线性变换  $T$

<sup>1</sup>例如行初等变换, 是要把用数乘矩阵的某一行加到另一行对应的元素上, 这里我们一般改成数域  $K$  上的  $\lambda$  多项式。

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

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

解:

- (1) 这部分涉及到施密特正交化过程。

设  $X \in V$ , 则  $x_3 = x_4$ , 所以

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

发现已满足正交性, 只需单位化即可

$$e_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, e_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$

- (2) 对称变换的判断为看对应的矩阵是否为对称矩阵。我们要利用 (1) 中求得的标准正交基求出线性变换对应的矩阵。

由  $T(e_1, e_2, e_3) = (e_1, e_2, e_3)A$

$$T(e_1) = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} = 1e_1 + 2e_2 + 0e_3$$

$$T(e_2) = \begin{bmatrix} 2 & 1 \\ 0 & 0 \end{bmatrix} = 2e_1 + 1e_2 + 0e_3$$

$$T(e_3) = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 \\ 3 & 3 \end{bmatrix} = 0e_1 + 0e_2 + 3e_3$$

则

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

易见  $A = A^{-1}$ , 所以  $T$  是对称变换。

- (3) 对  $A$  进行相似对角化, 则  $A = P\Lambda P^{-1}$ , 其中  $\Lambda$  为  $A$  特征值组成的对角矩阵,  $P$  为  $A$  的特征向量组成的矩阵。相似对角化涉及到线性代数的知识, 这里不做介绍。

$$\Lambda = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad P = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \end{bmatrix}$$

由  $T(e_1, e_2, e_3) = (e_1, e_2, e_3)A$

$$\text{得 } T(e_1, e_2, e_3) = (e_1, e_2, e_3)P\Lambda P^{-1}$$

$$\text{令 } (E_1, E_2, E_3) = (e_1, e_2, e_3)P, \text{ 则 } T(E_1, E_2, E_3) = (E_1, E_2, E_3)\Lambda$$

$$E_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$

$$E_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

$$E_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}$$

由此求出符合题意的新标准正交基  $(E_1, E_2, E_3)$ 。

### 2.1.3.2 若尔当标准型化

事实上并不是所有的矩阵都能进行相似对角化, 只有满秩矩阵才能进行相似对角化。而若尔当标准型化是对于任意矩阵都能进行的一种标准化方法。下面介绍这种方法。

1. 任意选一组基, 利用施密特正交化方法得到  $V$  的一组标准正交基

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

- 求  $T$  在基  $e_1, \dots, e_n$  下的矩阵  $A_0 \Rightarrow T(e_1, \dots, e_n) = (e_1, \dots, e_n)A_0$ 。
- 其中  $A_0 = PJP^{-1} \Rightarrow T(e_1, \dots, e_n) = (e_1, \dots, e_n)PJP^{-1}$ , 其中  $J$  是若尔当标准型。
- $T(e_1, \dots, e_n)P = (e_1, \dots, e_n)PJ$ 。
- 

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

$$(E_1, \dots, E_n) = (e_1, \dots, e_n)P$$

3. 通过坐标变换得到  $x = (E_1, \dots, E_n)P^{-1} = (E_1, \dots, E_n)$

$$x = (E_1, \dots, E_n)P^{-1} \begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix} = (E_1, \dots, E_n) \begin{pmatrix} l_1 \\ \vdots \\ l_n \end{pmatrix}$$

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

5. 于是

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

下面我们通过例 2 来演示这种方法。

**例 2** 设定阶矩阵组  $R^{2 \times 2}$  的子空间为

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

对于  $V$  中的线性变换  $T(X) = X + 2X^T$ , 求:

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

$$(T^k)(X), \forall X \in V$$

**解：**首先找到  $V$  的一组标准正交基。令  $x_{11} = -x_{12} - x_{21}$ ，找一组基

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

下面将  $X_1, X_2, X_3$  进行施密特正交化过程

$$\begin{aligned} Y'_1 &= X_1 = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} \\ Y'_2 &= X_2 - \frac{X_2, Y'_1}{Y'_1, Y'_1} Y'_1 \\ &= \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} \\ 1 & 0 \end{bmatrix}, \\ Y'_3 &= X_3 - \frac{X_3, Y'_1}{Y'_1, Y'_1} Y'_1 - \frac{X_3, Y'_2}{Y'_2, Y'_2} Y'_2 \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

再进行单位化

$$\begin{aligned} e_1 &= \frac{Y'_1}{\|Y'_1\|} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} \\ e_2 &= \frac{Y'_2}{\|Y'_2\|} = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 & -1 \\ 2 & 0 \end{bmatrix} \\ e_3 &= \frac{Y'_3}{\|Y'_3\|} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

接下来求  $T$  在  $e_1, e_2, e_3$  下的矩阵  $A_0$

$$\begin{aligned} T(e_1) &= e_1 + 2e_1^T = \frac{1}{\sqrt{2}} \begin{bmatrix} -3 & 1 \\ 2 & 0 \end{bmatrix} \\ T(e_2) &= e_2 + 2e_2^T = \frac{1}{\sqrt{6}} \begin{bmatrix} -3 & 3 \\ 0 & 0 \end{bmatrix} \\ T(e_3) &= e_3 + 2e_3^T = \begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix} \\ T e_1 &= [e_1 \ e_2 \ e_3] \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} \end{aligned}$$



$$\begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} 2 \\ \sqrt{3} \\ 0 \end{bmatrix}$$

同理, 对于  $Te_2$

$$\begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} \sqrt{3} \\ 0 \\ 0 \end{bmatrix}$$

对于  $Te_3$

$$\begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$$

所以

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

下面求  $A_0$  的 Jordan 标准型

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

经过初等变换

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

求得不变因子:  $1, \lambda - 3, (\lambda + 1)(\lambda - 3)$  求得初等因子组:  $(\lambda - 3), (\lambda + 1), (\lambda - 3)$

所以 Jordan 标准型为

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

接下来求  $P = [x_1 \ x_2 \ x_3]$

$$A_0 = PJP^{-1}$$

即

$$PJ = A_0P \Rightarrow [3x_1 \ -x_2 \ 3x_3] = [A_0x_1 \ A_0x_2 \ A_0x_3]$$

解得

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

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

现在  $T(e_1, e_2, e_3) = (e_1, e_2, e_3)J$  转化为了

$$T(e_1, e_2, e_3) = (e_1, e_2, e_3)PJP^{-1} \Rightarrow T(e_1, e_2, e_3)P = (e_1, e_2, e_3)PJ$$

令  $(E_1, E_2, E_3) = (e_1, e_2, e_3)P$ , 则  $T(E_1, E_2, E_3) = (E_1, E_2, E_3)J$

$$E_1 = \frac{2}{\sqrt{6}} \begin{bmatrix} -2 & 1 \\ 1 & 0 \end{bmatrix}$$

$$E_2 = \sqrt{2} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

$$E_3 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

(1)

$$X = \begin{bmatrix} 4 & -4 \\ 0 & -3 \end{bmatrix} = [E_1 \ E_2 \ E_3] \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix}$$

解得

$$\begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} -\sqrt{6} \\ \sqrt{2} \\ -3 \end{bmatrix}$$

所以

$$(T^3)(X) = (E_1, E_2, E_3)J^3 \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} 108 & -52 \\ -56 & -81 \end{bmatrix}$$

(2)

$$(T^k)(X) = (E_1, E_2, E_3)J^k \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = (E_1, E_2, E_3) \begin{bmatrix} 3^k k_1 \\ (-1)^k k_2 \\ 3^k k_3 \end{bmatrix}$$

其中

$$\begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} (x, E_1) \\ (x, E_2) \\ (x, E_3) \end{bmatrix}$$

总之, 相似对角化和若尔当标准型化为线性变换提供了一种特殊的矩阵表示形式, 使得线性变换的性质更加明确, 计算更加方便。这种关系在矩阵理论和线性代数的研究中发挥着重要的作用。

## 2.2 向量范数与矩阵范数

### 2.2.1 向量范数介绍

范数是一种类似于“长度”概念的函数。例如在向量空间中, 一种常见的范数为向量的, 其在一定意义上是对向量长度的一种度量。对于更一般的概念, 范数是一个定义在向量空间或者矩阵空间上的函数, 其满足**非负性**、**齐次性**和**三角不等式**。范数的概念可以推广到更一般的空间, 例如函数空间、矩阵空间等。

对于向量范数, 现定义如下。

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

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

则  $\|x\|$  为  $V$  上向量  $x$  的范数, 简称**向量范数**。

对于向量范数, 还有等价的概念定义如下。

满足  $c_1 \|x\|_1 \leq \|x\|_2 \leq c_2 \|x\|_1$  的两个向量范数  $\|x\|_1$  和  $\|x\|_2$  称为**等价的**。

$C^n$  中的向量序列

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

收敛到向量  $x = (\xi_1, \xi_2, \dots, \xi_n)$ , 当且仅当对任何一个向量范数  $\|\cdot\|$ , 数列

$$\|x^{(k)} - x\|$$

的极限为零。

下面给出几种常见的向量范数。

$$1\text{-范数} \quad \|x\|_1 = \sum_{i=1}^n |\xi_i|$$

$$2\text{-范数} \quad \|x\|_2 = \sqrt{|\xi_1|^2 + |\xi_2|^2 + \cdots + |\xi_n|^2}$$

$$\infty\text{-范数} \quad \|x\|_\infty = \max_i |\xi_i|$$

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

$$\text{加权范数} \quad \|x\|_A = (x^T A x)^{1/2}$$

有限性线性空间上的不同范数是等价的。

### 2.2.2 矩阵范数介绍

类似的，矩阵的线性空间对应的是矩阵范数。定义如下：

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

- (1) 非负性：当  $A \neq O$  时， $\|A\| > 0$ ；当  $A = O$  时， $\|A\| = 0$ ；
- (2) 齐次性： $\|aA\| = |a| \|A\|$  ( $a \in C$ )；
- (3) 三角不等式： $\|A+B\| \leq \|A\| + \|B\|$  ( $B \in C^{m \times n}$ )。

则称  $\|\cdot\|$  为  $A$  的**广义矩阵范数**。

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

- (4) 相容性： $\|AB\| \leq \|A\| \|B\|$  ( $B \in C^{n \times l}$ )。

则称  $\|\cdot\|$  为  $C^{m \times n}$  上的**矩阵范数**。

下面给出几种常见的矩阵范数。

$$\text{列和范数} \quad \|A\|_1 = \max_j |a_{ij}|$$

$$\text{谱范数} \quad \|A\|_2 = \sqrt{\lambda_1}, \quad \lambda_1 \text{ 为 } A^H A \text{ 的最大特征值}$$

$$\text{行和范数} \quad \|A\|_\infty = \max_i |a_{ij}|$$

如同向量范数，矩阵范数也是多种多样的。

但是，在数值方法中进行某种估计时，遇到的多数情况是：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。这可由矩阵范数与向量范数相容的概念来实现，下面引入这个概念。

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  $\|\cdot\|_V$ ，如果满足

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

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

### 2.2.3 矩阵可逆性条件、条件数和谱半径介绍

前两节已经初步指出范数的应用场合，本节再列举几点范数的应用。

#### 2.2.3.1 矩阵的可逆性条件

设  $A$  是  $C^{n \times n}$  上的矩阵， $\|\cdot\|$  是  $C^{n \times n}$  上的矩阵范数，若有  $\|A\| < 1$ ，则矩阵  $I - A$  可逆，且有

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

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

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

下面介绍条件数的定义，其与近似逆矩阵的误差有重要联系。

#### 2.2.3.2 条件数

设  $A$  是  $C^{n \times n}$  上的矩阵，每个元素  $a_{ij}$  带有误差  $\delta a_{ij}$ ，则准确矩阵为  $A + \delta A$ ，其中  $\delta A = (\delta a_{ij})^2$ 。若  $A$  可逆，其逆矩阵  $A^{-1}$  和  $(A + \delta A)^{-1}$  的近似程度（摄动）如何呢？关于这个问题，有如下的摄动定理。

设  $A \in C^{n \times n}$  可逆， $B \in C^{n \times n}$ ，在某矩阵范数  $\|\cdot\|$  下，若  $\|A^{-1}B\| < 1$ ，则有以下结论：

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

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

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

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

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

#### 2.2.3.3 矩阵的谱半径及其性质

矩阵的谱半径在特征值估计、广义逆矩阵、数值分析以及数值代数等理论的建树中，都占有及其重要的地位。现论述如下：

设矩阵  $A \in C^{n \times n}$  的  $n$  个特征值是  $\lambda_1, \lambda_2, \dots, \lambda_n$ ，称

---

<sup>2</sup>又称  $\delta a_{ij}$  为  $A$  的元素的扰动， $\delta A$  为  $A$  的扰动矩阵。

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

为  $A$  的谱半径。

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

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

成立。这说明谱半径是矩阵所有范数的一种下确界。

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

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

成立。

谱半径和矩阵序列是否收敛有紧密的关系。这个我们会在后面的章节中讲到。

## 2.3 矩阵函数介绍

### 2.3.1 矩阵序列介绍

矩阵序列可类比数列, 其概念如下。

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

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

不收敛的矩阵序列称为**发散**。

下面给出一些矩阵序列中的定义。

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

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

成立。

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

要判断  $A$  是否为收敛矩阵, 可以通过下面的定理来判断。

1. 设  $A^{(k)} \in \mathbb{C}^{m \times n}$ , 则

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

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

2.  $A$  为收敛矩阵的充要条件是  $\rho(A) < 1$ 。这里,  $\rho(A)$  是  $A$  的谱半径。

3.  $A$  为收敛矩阵的充分条件是只要有一种矩阵范数使得  $\|A\| < 1$ 。

如例 3 所示:

**例 3** 判断矩阵  $A = \begin{bmatrix} 0.1 & 0.3 \\ 0.7 & 0.6 \end{bmatrix}$  是否为收敛矩阵。

**解:** 矩阵  $A$  的 1-范数为  $\|A\|_1 = 0.9$ , 因此  $A$  是收敛矩阵。

### 2.3.2 矩阵级数介绍

矩阵级数是矩阵函数理论的基础。其概念如下。

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

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

记  $S^{(N)} = \sum_{k=0}^N A^{(k)}$ ，称其为级数和的部分和。如果矩阵序列  $\{S^{(N)}\}$  收敛，且有极限  $S$ ，则有

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

那么就称矩阵级数收敛，而且有和为  $S$ ，记为

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

不收敛的矩阵级数称为发散的。

对于  $\sum_{k=0}^{\infty} a_{ij}^{(k)} = s_{ij}$ ，如果式左端  $mn$  个数项级数都是绝对收敛的，则称矩阵级数式是**绝对收敛**的。

矩阵级数式有一些类似于数项级数的性质，例如：

1. 若矩阵级数式是绝对收敛的，则它也一定收敛，并且任意调换其项顺序所得的级数还是收敛，且其和不变。
2. 矩阵级数  $\sum_{k=0}^{\infty} A^{(k)}$  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（绝对）收敛的，那么  $\sum_{k=0}^{\infty} PA^{(k)}Q$  也是（绝对）收敛的，且有  $\sum_{k=0}^{\infty} PA^{(k)}Q = P(\sum_{k=0}^{\infty} A^{(k)})Q$ 。
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  $AB$ 。

可以发现很多矩阵级数的性质与数项级数的性质相似。

类似数列幂级数，还有矩阵幂级数，如 Neumann 级数： $\sum_{k=0}^{\infty} A^k$ ，这是一种方阵幂级数。下面给出 Neumann 级数的性质。

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

设方阵  $A$  对某一种矩阵范数  $\|\cdot\|$  有  $\|A\| < 1$ ，则对任何非负整数  $N$ ，以  $(I - A)^{-1}$  为部分和  $I + A + A^2 + \dots + A^N$  的近似矩阵时，存在

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

成立，由此可以得到部分和的误差估计式。

现在研究矩阵幂级数  $\sum_{k=0}^{\infty} c_k A^k$  与对应的纯量幂级数  $\sum_{k=0}^{\infty} c_k z^k$  的关系。

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  $A$  的谱半径  $\rho(A) < r$ , 则相应矩阵幂级数

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

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

### 2.3.3 矩阵函数介绍

矩阵函数是矩阵序列和矩阵级数的推广, 其概念如下。

设一元函数  $f(x)$  能展开为  $x$  的幂级数

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

其中  $r > 0$  表示幂级数收敛的收敛半径。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

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

在数学分析中, 指数函数具有运算规则  $e^x e^y = e^y e^x = e^{x+y}$ , 但需注意在矩阵分析中, 只有当  $AB = BA$  时, 才有  $e^A e^B = e^B e^A = e^{A+B}$  成立。由此可推论出, 对于任意的  $n$  阶矩阵  $A$ , 有  $e^A e^{-A} = e^{-A} e^A = I$  和  $(e^A)^{-1} = e^{-A}$  成立。对于整数  $m$ , 还有  $(e^A)^m = e^{mA}$  成立。

当  $AB = BA$ , 类似的三角函数结论

$$\begin{cases} \cos(A+B) = \cos A \cos B - \sin A \sin B \\ \cos 2A = \cos^2 A - \sin^2 A \\ \sin(A+B) = \sin A \cos B + \cos A \sin B \\ \sin 2A = 2 \sin A \cos A \end{cases}$$

也成立。

### 2.3.4 函数矩阵对矩阵的导数与积分

函数矩阵对矩阵的导数的概念如下。

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, 并称

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

为矩阵  $A(t)$  的导数。

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

$$\frac{d}{dt}(A(t) + B(t)) = \frac{dA(t)}{dt} + \frac{dB(t)}{dt}$$



$$\frac{d}{dt}(A(t)B(t)) = \frac{dA(t)}{dt} \cdot B(t) + A(t) \cdot \frac{dB(t)}{dt}$$

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

成立。这里,  $a = a(t)$  是  $t$  的可导函数。

设  $n$  阶矩阵  $A$  与  $t$  无关, 则有

$$\frac{d}{dt}e^{tA} = Ae^{tA} = e^{tA}A$$

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

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

成立。

类似的, 还有函数矩阵对矩阵的积分的概念如下。

如果函数矩阵  $A(t)$  的每一个元素  $a_{ij}(t)$  都是区间  $[t_0, t_1]$  上的可积函数, 则称

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

为  $A(t)$  在区间  $[t_0, t_1]$  上的积分。

## 第三章 矩阵函数的求法研究

### 3.1 待定系数法

#### 3.1.1 待定系数法求矩阵函数的步骤推导

待定系数法以 Hamilton-Cayley 定理为基础来进行矩阵函数的求解。其步骤如下。

设  $n$  阶矩阵  $A$  的特征多项式  $\varphi(\lambda) = \det(\lambda I - A)$ 。如果首 1 多项式

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

满足  $\psi(\lambda) = O$  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  $\psi(\lambda)$  的互异零点分别为  $\lambda_1, \lambda_2, \dots, \lambda_s$  相应的重数分别为  $r_1, r_2, \dots, r_s$  ( $r_1 + r_2 + \cdots + r_s = m$ ), 则有

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

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

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

其中  $r(x)$  是次数低于  $m$  的多项式, 于是可由

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

确定  $r(x)$ 。利用  $\psi(A) = O$ , 可得

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

#### 3.1.2 举例展示求法

**例 4** 设矩阵  $A$  为

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

求  $e^A$  与  $e^{At}$  ( $t \in \mathbb{R}$ )。

**解:**

特征多项式  $\varphi(\lambda) = \det(\lambda I - A) = (\lambda - 2)^3$ ,  $A$  的最小多项式  $m(\lambda) = (\lambda - 2)^2$ , 故  $\varphi(\lambda)$  与  $(\lambda - 2)^2$  相同。

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

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

即

$$\begin{cases} a + 2b = e^2 \\ b = e^2 \end{cases}$$

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

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

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

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

即

$$\begin{cases} a + 2b = e^{2t} \\ b = te^{2t} \end{cases}$$

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

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

## 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

数项级数求和法利用级数逼近函数的思想来求矩阵函数。其步骤如下。

设首 1 多项式  $\psi(\lambda)$  形如式

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

且满足  $\psi(\lambda) = O$ , 即

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

或者

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

由此可得到递推

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

于是有

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

这表明, 利用上式可以将一个无限级数缩减为有限级数的形式, 从而转化为一个  $m$  个数项级数的求和问题。

### 3.2.2 举例展示求法

例 5 设  $A =$

$$\begin{bmatrix} \pi & 0 & 0 & 0 \\ 0 & -\pi & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

求  $\sin A$ .

$$\varphi(\lambda) = \det(\lambda I - A) = \lambda^4 - \pi^2 \lambda^2$$

由于  $\varphi(A) = O$ , 所以  $A^4 = \pi^2 A^2, A^5 = \pi^2 A^3, A^7 = \pi^4 A^3, \dots$  于是有

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

### 3.3 对角型法

#### 3.3.1 对角型法求矩阵函数的步骤推导

对角型法可以把矩阵函数的求解问题转化为对角矩阵函数的求解问题从而简化计算。其步骤如下。

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

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

则有  $A = P\Lambda P^{-1}$ ,  $A^2 = P\Lambda^2 P^{-1}$ , ..., 于是有

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

从而

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

这表明，当  $A$  相似于对角矩阵时，可以将矩阵函数的求和问题转化为函数值的求和问题。

#### 3.3.2 举例展示求法

**例 6** 设矩阵  $A$  为

$$A = \begin{bmatrix} 4 & 6 & 0 \\ -3 & -5 & 0 \\ -3 & -6 & 1 \end{bmatrix}$$

求  $e^A$ ,  $e^{At}$  ( $t \in \mathbb{R}$ ) 及  $\cos A$ 。

**解：**特征多项式  $\varphi(\lambda) = \det(\lambda I - A) = (\lambda + 2)(\lambda - 1)^2$ 。对应  $\lambda_1 = -2$  的特征向量  $p_1 = (-1, 1, 1)^T$ ，对应  $\lambda_2 = \lambda_3 = 1$  的两个线性无关的特征向量分别是  $p_2 = (-2, 1, 0)^T$ ,  $p_3 = (0, 0, 1)^T$ 。构造矩阵  $P$

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

则有

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

和

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

求得

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

同理

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

以及

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

### 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

类似求线性变换, 对于无法相似于对角矩阵的矩阵, 可以依然可以利用若尔当标准型来求矩阵函数。其步骤如下。

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

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

其中

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

可求得

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

故有

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

这表明，矩阵幂级数的求和问题可以转化为求解每个 Jordan 块的函数值然后相加得到最终结果。

### 3.4.2 举例展示求法

**例 7** 设  $f(x) = \ln x$ ，求  $f(A)$ ，这里  $A$  分别为

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

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

**解:**

(1)

$$\det(\lambda I - A) = (\lambda - 1)^4$$

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

$$\ln A = C \cdot \ln J \cdot C^{-1} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 & 0 \\ \frac{1}{3} & -\frac{1}{2} & 1 & 0 \end{bmatrix}$$

(2)  $J_1 = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ ,  $J_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ , 于是有

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

因此

$$\ln A = \begin{bmatrix} \ln J_1 & 0 \\ 0 & \ln J_2 \end{bmatrix} = \begin{bmatrix} \ln 2 & \frac{1}{2} & 0 & 0 \\ 0 & \ln 2 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$



## 第四章 矩阵分解方法研究

### 4.1 矩阵的 LU 分解

#### 4.1.1 矩阵 LU 分解的步骤推导

三角矩阵的计算, 例如求行列式、逆矩阵等较普通矩阵来说都较为简单, 所以我们就希望研究如何将普通矩阵分解为几个三角矩阵的乘积, 矩阵分解应运而生。

首先使用矩阵理论描写线性代数中学习的高斯消元过程。

设  $A^{(0)} = A$ , 其元素为  $a_{ij}^{(0)} (i, j = 1, 2, \dots, n)$ 。记  $A$  的  $k$  阶阶梯形式子式为  $\Delta_k (k = 1, 2, \dots, n)$ 。如果  $\Delta_1 = a_{11}^{(0)} \neq 0$ , 则  $c_{i1} = \frac{a_{i1}^{(0)}}{a_{11}^{(0)}} (i = 2, 3, \dots, n)$ , 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

$$L_1^{-1} A^{(0)} = \begin{bmatrix} a_{11}^{(0)} & a_{12}^{(0)} & \dots & a_{1n}^{(0)} \\ 0 & a_{22}^{(1)} & \dots & a_{2n}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2}^{(1)} & \dots & a_{nn}^{(1)} \end{bmatrix} = A^{(1)}$$

实际上  $L^{-1}$  矩阵就是体现了高斯消元过程中的行的加减过程, 而  $A^{(1)}$  矩阵就是经过变换后的矩阵。由此可见,  $A^{(1)}$  矩阵的第一列元素除  $a_{11}^{(0)}$  外都为 0。上式还可写为

$$A^{(1)} = L_1 A^{(0)}$$

将这种步骤进行  $n-1$  次, 即可得到一个上三角矩阵  $A^{(n-1)}$ , 这就是高斯消元过程。高斯消元过程能够进行到底的条件是当且仅当  $\Delta_r \neq 0 (r = 1, 2, \dots, n-1)$ , 即  $A$  的所有顺序主子式都不为 0。

我们发现  $A^{(n-1)}$  已满足上三角矩阵的要求, 我们将其认为是  $U$  矩阵, 接下来我们求解对应的  $L$  矩阵。由求解过程中我们知道

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

容易求出

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

这是一个对角元素均为 1 的下三角矩阵, 称为**单位下三角矩阵**, 于是得到

$$A = LU$$

这就是矩阵  $A$  的 **LU 分解** (或**三角分解**)。但是一个方阵的 LU 分解并不唯一。为了唯一分解, 在实际分解中, 我们通常使用 LDU 分解。我们可以提出一个对角矩阵  $D$ , 使得  $U$  矩阵也变成单位上三角矩阵。定理如下。

设  $A = (a_{ij})_{n \times n}$  是  $n$  阶方阵, 则当且仅当  $A$  的顺序主子式  $\Delta_k \neq 0$  ( $k = 1, 2, \dots, n-1$ ) 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

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

其中  $d_k = \frac{\Delta_k}{\Delta_{k-1}}$  ( $k = 1, 2, \dots, n; \Delta_0 = 1$ )。

设矩阵  $A$  有唯一的 LDU 分解。若把  $A = LDU$  中的  $D$  和  $U$  结合起来, 并且用  $\hat{U}$  表示, 就得到唯一的分解为

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

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

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

称为  $A$  的 **Crout 分解**。

除 LDU 分解外, 还有其他三角分解, 如 Cholesky 分解, 定义如下。

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

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

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

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

或者

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

这里  $G = L\tilde{D}$  是下三角矩阵, 称为实对称正定矩阵的 **Cholesky 分解** (平方根分解或对称三角分解)。

方阵  $A$  的 PLU 分解是是将它分解成一个置换矩阵  $P$ 、一个下三角矩阵  $L$  与上三角矩阵  $U$  的乘积, 即

$$A = PLU$$

所有的方阵都可以写成 PLU 分解的形式, 由于左乘置换矩阵  $P^{-1}$  是在交换行的顺序, 所以由  $P^{\{-1\}}A = LU$  推得适当的交换  $A$  的行的顺序, 即可将  $A$  做 LU 分解。PLU 分解有很高的数值

稳定性, 在许多数学计算编程软件中都有涉及。本文会在下方例题中稍作展示。如果同时取  $P^{-1}$  作为  $P$  的逆, 那么就得到了  $A$  的 LU 分解, 同时再可得到 LDU 分解。

#### 4.1.2 举例展示求法

**例 8** 求  $A$  的 LDU 分解, 其中

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

**解:**

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

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

计算, 得

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

对  $A^{(1)}$  构造矩阵, 得

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

计算, 得

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

可求出

$$L = L_1 L_2 = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

于是根据  $A^{(0)} = A$  的 LDU 分解为

$$A = L_1 L_2 A^{(2)} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{5}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & -\frac{1}{5} \\ 0 & 0 & 1 \end{bmatrix}$$

**例 9** 利用编程实现例 8 中  $A$  的 PLU 分解。

**解:**

使用 Python 语言 scipy 库中的 linalg 模块中的 lu 函数, 可以求得  $A$  的 PLU 分解。

```
import numpy as np
from scipy.linalg import lu

A = np.array([[2, -1, 3], [1, 2, 1], [2, 4, 2]])

P, L, U = lu(A)

print("P:\n", P)
print("L:\n", L)
print("U:\n", U)
```

```
[1]: import numpy as np
      from scipy.linalg import lu

      A = np.array([[2, -1, 3], [1, 2, 1], [2, 4, 2]])

      P, L, U = lu(A)

      print("P:\n", P)
      print("L:\n", L)
      print("U:\n", U)

P:
[[1.  0.  0.]
 [0.  0.  1.]
 [0.  1.  0.]]
L:
[[1.  0.  0.]
 [1.  1.  0.]
 [0.5 0.5 1.]]
U:
[[ 2. -1.  3.]
 [ 0.  5. -1.]
 [ 0.  0.  0.]]
```

图 1 Python PLU 分解结果

**例 10** 求矩阵  $A$  的 Cholesky 分解，其中

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

**解：** 矩阵分解式是对称正定矩阵。由构造递推式可以得到如下

$$\begin{aligned} g_{11} &= \sqrt{a_{11}} = \sqrt{5} \\ g_{21} &= \frac{a_{21}}{g_{11}} = \frac{-2}{\sqrt{5}}, \quad g_{22} = \sqrt{a_{22} - g_{21}^2} = \sqrt{\frac{11}{5}} \\ g_{31} &= \frac{a_{31}}{g_{11}} = 0, \quad g_{32} = \frac{a_{32} - g_{31}g_{21}}{g_{22}} = -\frac{\sqrt{5}}{\sqrt{11}} \\ g_{33} &= \sqrt{a_{33} - g_{31}^2 - g_{32}^2} = \sqrt{1 - \frac{5}{11}} = \sqrt{\frac{6}{11}} \end{aligned}$$

从而

$$A = \begin{bmatrix} \sqrt{5} & 0 & 0 \\ -\frac{2}{\sqrt{5}} & \sqrt{\frac{11}{5}} & 0 \\ 0 & -\frac{\sqrt{5}}{\sqrt{11}} & \sqrt{\frac{6}{11}} \end{bmatrix} \begin{bmatrix} \sqrt{5} - \frac{2}{\sqrt{5}} & 0 \\ 0 & \sqrt{\frac{11}{5}} - \frac{\sqrt{5}}{\sqrt{11}} \\ 0 & 0 & \sqrt{\frac{6}{11}} \end{bmatrix}$$

## 4.2 矩阵的 QR 分解

### 4.2.1 矩阵 QR 分解的步骤推导

QR 分解把矩阵分解成一个半正交矩阵与一个上三角矩阵的积。QR 分解经常用来解线性最小二乘法问题。QR 分解也是特定特征值算法即 QR 算法的基础。

QR 分解可以使用施密特正交化过程来计算。也可以使用 Givens 变换或者 Householder 变换来计算。

首先介绍 Givens 矩阵和 Householder 矩阵。

#### 4.2.1.1 Givens 矩阵

Givens 变换是初等旋转变换，其矩阵形式为

$$T_{ij} = \begin{bmatrix} I & & \\ & c & s \\ & -s & c \\ & & I \end{bmatrix} \quad (i < j)$$

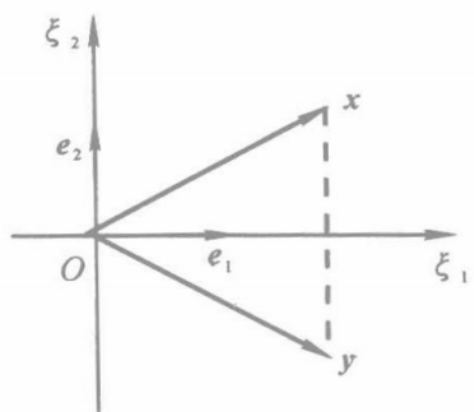
特别地，当  $i = 1$ ，即  $j = i + 1$ ，或  $j = n$  时，上式便为特为

$$T_{1,j} = \begin{bmatrix} c & s & & \\ -s & c & & \\ & & I & \\ & & & I \end{bmatrix}, \quad T_{ij} = \begin{bmatrix} I & & & \\ & c & s & \\ & -s & c & \\ & & & I \end{bmatrix}, \quad T_{in} = \begin{bmatrix} I & & & \\ & c & s & \\ & & I & \\ & -s & c & \end{bmatrix}$$

容易验证，当  $c^2 + s^2 = 1$  时，存在角度  $\theta$ ，使得  $c = \cos \theta$ ， $s = \sin \theta$ 。这对应将向量  $x = (\xi_1, \xi_2)^T$  旋转  $\theta$  角，得到向量  $y = (\eta_1, \eta_2)^T$  的变换。

#### 4.2.1.2 Householder 矩阵

在平面  $\mathbb{R}^2$  中，将向量  $x$  映射为关于得到  $e_1$ （或者关于  $e_2$  轴正交的直线）的向量  $y$  的变换，称为求  $x$  关于  $e_1$  轴的镜像（反射）（见图 2）变换。

图 2 关于  $e_1$  轴的镜像变换

设  $x = (\xi_1, \xi_2)^T$ ，则有

$$y = \begin{bmatrix} -\xi_1 \\ -\xi_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = (I - 2e_2e_2^T)x = Hx$$

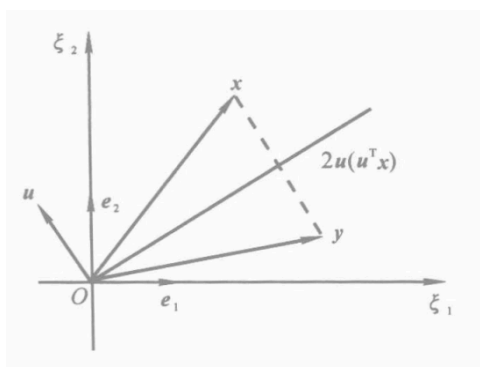
其中  $e_2 = (0, 1)^T$ ， $H$  是正交线性变换，且  $\det H = -1$ 。

将向量  $x$  映射为关于与单位向量  $u$  正交的直线对称的向量  $y$  的变换（见图 3）可描述如下：

$$x - y = 2u(u^T x)$$

$$y = x - 2u(u^T x) = (I - 2uu^T)x = Hx$$

容易验证， $H$  是正交矩阵，且由上知  $\det H = -1$ 。

图 3 关于与单位向量  $u$  正交的直线的对称变换

一般地，在  $\mathbb{R}^n$  中，将向量  $x$  映射为关于与向量  $u$  正交的  $n-1$  维子空间对称的向量  $y$  的镜像变换定义如下。

设单位列向量  $u \in \mathbb{R}^n$ ，称

$$H = I - 2uu^T$$

为 **Householder 矩阵（初等反射矩阵）**，由 Householder 变换定义的线性变换为 **Householder 变换（初等反射变换）**。

Householder 矩阵具有下列性质

1.  $H^T = H$ （对称矩阵）；

2.  $H^T H = I$  (正交矩阵);
3.  $H^2 = I$  (对合矩阵);
4.  $H^{-1} = H$  (自逆矩阵);
5.  $\det H = -1$  (行列式为  $-1$ )。

成立。

#### 4.2.1.3 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

$$A = QR$$

则称上式为  $A$  的 QR 分解。

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  $R$ , 使得  $A$  的 QR 分解式唯一。除去相差一个对角元素的绝对值(模)全部为 1 的对角矩阵因子外, 分解式是唯一的。

设矩阵  $A$  是  $m \times n$  实(复)矩阵, 且矩阵  $n$  列向量独立, 则  $A$  有分解

$$A = QR$$

其中  $Q$  是  $m \times n$  实(复)矩阵, 满足  $Q^T Q = I$  ( $Q^H Q = I$ ),  $R$  是  $n$  阶实(复)可逆上三角矩阵。该分解有去掉一个对角元素绝对值(模)全部为 1 的对角矩阵因子外, 分解式是唯一的。

于是我们可以使用施密特正交化方法来求解 QR 分解。

定理中矩阵  $A$  的  $n$  个列向量为  $a_1, a_2, \dots, a_n$ 。采用  $A$  可逆, 所以这  $n$  个列向量线性无关。将它们用 Schmidt 正交化方法生成正交化之, 可得到  $n$  个标准正交列向量  $q_1, q_2, \dots, q_n$ 。

对  $a_1, a_2, \dots, a_n$  正交化, 可得

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

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

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

用矩阵形式表示为

其中

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

用对  $b_1, b_2, \dots, b_n$  单位化, 可得

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

于是有

$$(a_1, a_2, \dots, a_n) = (b_1, b_2, \dots, b_n)C = \begin{bmatrix} b_1 & & & \\ & b_2 & & \\ & & \ddots & \\ & & & b_n \end{bmatrix} \begin{bmatrix} 1 & & & \\ k_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ k_{n1} & k_{n2} & \cdots & 1 \end{bmatrix}$$

$$(q_1, q_2, \dots, q_n) = \frac{1}{\|b_1\|} \frac{1}{\|b_2\|} \cdots \frac{1}{\|b_n\|} C$$

故

$$Q = (q_1, q_2, \dots, q_n) \quad R = \text{diag}(\|b_1\|, \|b_2\|, \dots, \|b_n\|) \cdot C$$

从而可见  $A$  的 QR 分解  $A = QR$  是唯一的, 因为

$$Q = Q_1 R_1 R_1^{-1} = Q_1 D$$

其中  $D = R R^{-1}$  为对角线上全为单位矩阵。于是

$$I = Q^H Q = (Q_1 D)^H (Q_1 D) = D^H D$$

这表明  $D$  不改变正交(酉)矩阵, 而上三角矩阵就是对角矩阵(模)全为 1 的对角矩阵乘以  $R_1$ ,  $Q_1 = Q D^{-1}$ 。

除此之外, 可以直接用 Givens 旋转和 Householder 反射, 把正交矩阵  $Q$  具体构造出来。

任何  $n$  阶实可逆矩阵  $A = (a_{ij})_{n \times n}$  可通过左连乘若干个 Givens 矩阵化为可逆上三角矩阵。

由  $\det A \neq 0$  知,  $A$  的第一列  $b^{(1)} = (a_{11}, a_{12}, \dots, a_{n1})^T \neq 0$ , 存在有限个 Givens 矩阵的乘积, 记作  $T_1$ , 使得

$$T_1 b^{(1)} = |b^{(1)}| e_1 \quad e_1 \in R^n$$

令  $a_{11}^{(1)} = |b^{(1)}|$ , 则有

$$T_1 A = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & & & \\ \vdots & & A^{(1)} & \\ 0 & & & \end{bmatrix}$$

由  $\det A^{(1)} \neq 0$  知,  $A^{(1)}$  的第 1 列  $b^{(2)} = (a_{22}^{(1)}, a_{32}^{(1)}, \dots, a_{n2}^{(1)})^T \neq 0$ , 存在有限个 Givens 矩阵的乘积, 记作  $T_2$ , 使得



$$T_2 b^{(2)} = |b^{(2)}| e_1 \quad e_1 \in R^{n-1}$$

令  $a_{22}^{(2)} = |b^{(2)}|$ , 则有

$$T_1 A = \begin{bmatrix} a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2n}^{(2)} \\ 0 & & & \\ \vdots & & A^{(2)} & \\ 0 & & & \end{bmatrix}$$

重复此步骤

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

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, 使得

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

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。

类似的, 任何  $n$  阶实可逆矩阵  $A = (a_{ij})_{n \times n}$  可通过左连乘若干个 Householder 矩阵化为可逆上三角矩阵。

由  $\det A \neq 0$  知,  $A$  的第一列  $b^{(1)} = (a_{11}, a_{12}, \dots, a_{n1})^H \neq 0$ , 存在 Householder 矩阵, 记作  $H_1$ , 使得

$$H_1 b^{(1)} = |b^{(1)}| e_1 \quad e_1 \in R^n$$

令  $a_{11}^{(1)} = |b^{(1)}|$ , 则有

$$H_1 A = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & & & \\ \vdots & & A^{(1)} & \\ 0 & & & \end{bmatrix}$$

由  $\det A^{(1)} \neq 0$  知,  $A^{(1)}$  的第 1 列  $b^{(2)} = (a_{22}^{(1)}, a_{32}^{(1)}, \dots, a_{n2}^{(1)})^H \neq 0$ , 存在 Householder 矩阵, 记作  $H_2$ , 使得

$$H_2 b^{(2)} = |b^{(2)}| e_1 \quad e_1 \in R^{n-1}$$

令  $a_{22}^{(2)} = |b^{(2)}|$ , 则有

$$H_1 A = \begin{bmatrix} a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2n}^{(2)} \\ 0 & & & \\ \vdots & & A^{(2)} & \\ 0 & & & \end{bmatrix}$$

重复此步骤

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

最后令

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

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

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

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

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

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

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

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。

#### 4.2.2 举例展示求法

**例 11** 用 Schmidt 正交化方法来求矩阵  $A$  的 QR 分解, 其中

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

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

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

$$b_2 = a_2 - \frac{1}{6}b_1 = (1, -1, 1)^T$$

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

构造矩阵

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

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

则有  $A = QR$ 。

**例 12** 用 Givens 旋转来求矩阵  $A$  的 QR 分解, 其中

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

**解:** 第 1 步: 对  $A$  的第 1 列  $b^{(1)} = (0, 1, 1)^T$  旋转  $T_1$ , 使得  $T_1 b^{(1)} = b_1^{(1)} e_1$ 。

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

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

$$T_1 = T_{13} T_{12} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ -1 & 0 & 0 \\ 0 & -1 & 1 \end{bmatrix}, \quad T_1 A = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & -1 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

第 2 步: 对  $A^{(1)}$  的第 1 列  $b^{(2)} = (-1, -1)^T$  旋转  $T_2$ , 使得  $T_2 b^{(2)} = |b_1^{(2)}| e_1$ 。

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

$$T_2 = T_{12}, \quad T_2 A^{(1)} = \begin{bmatrix} \frac{3}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{2}{\sqrt{3}} \end{bmatrix}$$

最后, 令

$$T = \begin{bmatrix} 1 & & \\ & T_2 & \\ & & \end{bmatrix}, \quad T_1 = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \end{bmatrix}$$

则有  $A = QR$ , 其中

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

**例 13** 用 Householder 反射来求矩阵  $A$  的 QR 分解, 其中

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

**解:** 第 1 步: 对  $A$  的第 1 列  $b^{(1)} = (3, 6, 6)^T$ , 构造 Householder 反射, 有

$$b^{(1)} = \begin{bmatrix} 3 \\ 6 \\ 6 \end{bmatrix}, \quad b^{(1)} - \|b^{(1)}\| e_1 = \begin{bmatrix} -6 \\ 6 \\ 6 \end{bmatrix}, \quad u = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

$$H_1 = I - 2uu^T = \frac{1}{3} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{bmatrix}, \quad H_1 A = \begin{bmatrix} 9 & 48 & 15 \\ 0 & 9 & -3 \\ 0 & -12 & 9 \end{bmatrix}$$

对  $A^{(1)}$  的第 1 列  $b^{(2)} = (9, -12)^T$ , 构造 Householder 反射, 有

$$b^{(2)} = \begin{bmatrix} 9 \\ -12 \end{bmatrix}, \quad b^{(2)} - \|b^{(2)}\| e_1 = \begin{bmatrix} -6 \\ -12 \end{bmatrix}, \quad u = \frac{1}{\sqrt{5}} \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

$$H_2 = I - 2uu^T = \frac{1}{5} \begin{bmatrix} 1 & 3 & -4 \\ 3 & -4 & -3 \end{bmatrix}, \quad H_2 A^{(1)} = \begin{bmatrix} 15 & -9 \\ 0 & -3 \end{bmatrix}$$

最后, 令

$$S = \begin{bmatrix} 1 & & \\ & H_2 & \\ & & \end{bmatrix}, \quad H_1 = \frac{1}{15} \begin{bmatrix} 5 & 10 & 10 \\ -2 & 11 & -10 \\ -14 & 2 & 5 \end{bmatrix}$$

则有

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

**例 14** 利用编程实现对例 12 中的矩阵  $A$  的 QR 分解。

**解:** 编写 Python 代码如下:

```
import numpy as np
from scipy.linalg import qr
```

```
A = np.array([[1, 2, 2], [2, 1, 2], [1, 2, 1]])
```

```
Q, R = qr(A)
```

```
Q, R
```

```
import numpy as np
from scipy.linalg import qr

# 定义矩阵A
A = np.array([[1, 2, 2], [2, 1, 2], [1, 2, 1]])

# 进行QR分解
Q, R = qr(A)

Q, R

(array([[ -4.08248290e-01,  5.77350269e-01, -7.07106781e-01],
        [ -8.16496581e-01, -5.77350269e-01, -3.33066907e-16],
        [ -4.08248290e-01,  5.77350269e-01,  7.07106781e-01]]),
 array([[-2.44948974, -2.44948974, -2.85773803],
        [ 0.          ,  1.73205081,  0.57735027],
        [ 0.          ,  0.          , -0.70710678]]))
```

图 4 Python QR 分解结果

## 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问题。

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

$$A = FG$$

则根据上式分解称为  $A$  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**。

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

$$A \rightarrow B = \begin{bmatrix} G \\ O \end{bmatrix}, \quad G \in \mathbb{C}_r^{r \times n}$$

于是存在着一个  $m$  阶初等矩阵的逆矩阵，记作  $P$ ，使得  $PA = B$ ，或者  $A = P^{\{-1\}}B$ 。将  $P^{\{-1\}}$  分块为

$$P^{-1} = [F \mid S] \quad (F \in \mathbb{C}_r^{m \times r}, S \in \mathbb{C}_{m-r}^{m \times (m-r)})$$

则有

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

其中  $F$  是列满秩矩阵,  $G$  是行满秩矩阵。

注意  $A$  的满秩分解式不是唯一的。这是因为若取  $D$  是任一  $r$  阶可逆矩阵, 则式可改写为

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

这是  $A$  的另一种满秩分解。

#### 4.3.2 举例展示求法

**例 15** 求解矩阵  $A$  的满秩分解, 其中

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

**解:** 将矩阵  $A$  与单位阵  $I$  扩充为增广矩阵  $[A \mid I]$  进行初等行变换, 将  $A$  所在的位置变换为阶梯形矩阵  $B$ , 且

$$[A \mid I] = \left[ \begin{array}{cccc|cccc} -1 & 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 2 & -1 & 1 & 0 & 1 & 0 \\ 2 & 2 & -2 & -1 & 0 & 0 & 1 \end{array} \right] \xrightarrow{\text{初等行变换}} \left[ \begin{array}{cccc|cccc} -1 & 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 3 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 1 \end{array} \right]$$

由此得到

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

因此,  $P^{-1}$  为

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

于是有

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

### 4.4 矩阵的奇异值分解

#### 4.4.1 矩阵奇异值分解的步骤推导

矩阵的奇异值分解在最优化问题、特征值问题、最小二乘法问题、广义逆矩阵问题和统计学等各方面都有广泛的应用。

奇异值分解 (SVD) 在最小二乘法和最优化中的应用是多方面的, 主要体现在以下几个方面:

1. 求解过定或欠定系统: 对于线性方程组  $Ax = b$ , 当方程组是过定的, 即方程的数量多于未知数 (方程过于充满), 直接求解可能不符合或不存在。此时, 我们可以通过最小化残差的平方和  $\|Ax - b\|^2$  来寻找最佳拟合。使用 SVD, 将矩阵  $A$  分解为  $U\Sigma V^T$ , 其中  $U$  和  $V$  是正交矩阵,  $\Sigma$  是对角矩阵。通过 SVD, 我们可以与相对应的方程组的解  $x = V\Sigma^{-1}U^Tb$ , 这里

的  $\Sigma^{-1}$  实际上是  $\Sigma$  中非零元素的倒数, 并且已经转置, 这样的解是最接近原方程组的最小二乘解。

2. 数据拟合: 在数据分析中, 经常需要对缺失数据进行拟合, 找到最佳的数据模型。通过 SVD, 我们可以在数据的高维度空间中, 寻找到最小的子空间。缺失的数据可以在这个空间中得到最佳估计。
3. 计算稳定性: 直接使用矩阵的转置  $A^T A$  来求解可能会导致数值不稳定的问题, 尤其是当  $A^T$  接近奇异时。而 SVD 由于其优越的数值稳定性, 可以提供更稳定的解。
4. 处理具有特殊结构的矩阵: 当矩阵  $A$  具有特殊的数学结构时, 例如是一个稀疏矩阵或者一个大矩阵。SVD 允许我们通过构造对角矩阵  $A^+$  来解决这个问题, 使得能够通过大型稀疏矩阵或者大型密集矩阵的奇异值分解来找到最优解。通过  $A^+$  可以找到最小的奇异值的逆, 从而获得。
5. 降维和特征提取: 在机器学习中, SVD 可以用来降低数据的维度, 提取出最重要的特征。这是通过保留最高的几个奇异值及其对应的向量来实现的, 这可以显著减少在大型或者复杂的数据集上的运算量, 同时仍然保持数据的主要特征。

关于这个最优化的问题, 我们也会在后面的例题中涉及。

首先介绍特殊情况, 在实对称正交矩阵上的正交对角分解。

若  $A$  是  $n$  阶对称矩阵, 则存在正交矩阵  $Q$ , 使得

$$Q^T A Q = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

其中  $\lambda_i (i = 1, 2, \dots, n)$  为矩阵  $A$  的特征值, 而  $Q$  的  $n$  个列向量组成  $A$  的一个完备的标准正交特征向量系。

对于实的非对称矩阵  $A$ , 不再有上式的分解, 但存在两个正交矩阵  $P$  和  $Q$ , 使得  $P^T A Q$  是对角矩阵, 即有下面的正交对角分解定理。

设  $A$  是  $\mathbb{R}^{n \times n}$  阶矩阵, 则存在正交矩阵  $P$  和  $Q$ , 使得

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

其中  $\sigma_i > 0 (i = 1, 2, \dots, n)$ 。

为  $A$  可逆, 所以  $A^T A$  为对称矩阵正定矩阵。于是存在正交矩阵  $Q$ , 使得

$$Q^T (A^T A) Q = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

其中  $\lambda_i > 0 (i = 1, 2, \dots, n)$  为  $A^T A$  的特征值。令

$$\sigma_i = \sqrt{\lambda_i}, \quad \Lambda = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$$

则有  $Q^T (A^T A) Q = \Lambda^2$ , 或者  $(A Q \Lambda^{-1})^T A Q = A$ 。

再令  $P = A Q \Lambda^{-1}$ , 则有

$$P^T P = (A Q \Lambda^{-1})^T (A Q \Lambda^{-1}) = I$$

即  $P$  为正交矩阵, 且满足

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

故

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

称上式为矩阵  $A$  的正交对角分解。

然而, 对于一般的矩阵  $A$ , 我们应该如何分解呢? 我们发现,  $A^H A$  满足上面的条件, 所以我们可以对  $A^H A$  进行上面的分解, 然后再返回来一一对比得到  $A$  的分解。

为了介绍矩阵的奇异值与奇异值分解, 需要下面的结论:

1. 设  $A \in \mathbb{C}^{m \times n} (r > 0)$ , 则  $A^H A$  是 Hermitian 矩阵, 且其特征值都是非负实数;
2.  $\text{rank}(A^H A) = \text{rank} A$ ;
3. 设  $A \in \mathbb{C}^{m \times n}$ , 则  $A=0$  的充要条件是  $A^H A = 0$ 。

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

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

则称  $\sigma_i = \sqrt{\lambda_i} (i = 1, 2, \dots, n)$  为  $A$  的奇异值; 当  $\lambda_i$  是  $A$  的正特征值, 它的奇异值都是 0。因此, 矩阵  $A$  的奇异值由  $A$  的正定矩阵  $A^H A$  的特征值的非负平方根给出。 $A$  的非零奇异值的个数等于  $A$  的秩。

设  $A \in \mathbb{C}^{m \times n} (r > 0)$ , 则存在  $m$  阶正交矩阵  $U$  和  $n$  阶正交矩阵  $V$ , 使得

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

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

记 Hermitian 矩阵  $A^H A$  的特征值为

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

可知存在  $n$  阶正交矩阵  $V$ , 使得

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

将  $V$  分块为

$$V = [V_1 \mid V_2], \quad V_1 \in \mathbb{C}_r^{n \times r}, \quad V_2 \in \mathbb{C}_{n-r}^{n \times (n-r)}$$

并改写式为

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

则有

$$A^H A V_1 = V_1 \Sigma^2, \quad A^H A V_2 = O$$



由第一式可得  $V_1^H A^H A V_1 = \Sigma^2$ , 或者

$$(A V_1 \Sigma^{-1})^H (A V_1 \Sigma^{-1}) = I_r,$$

由第二式可得  $(A V_2)^H (A V_2) = O$ , 或者  $A V_2 = O$ 。

令  $U_1 = A V_1 \Sigma^{-1}$ , 则  $U_1^H U_1 = I_r$ , 即  $U_1$  的  $r$  列是两两正交的单位向量, 记作  $U_1 = (u_1, u_2, \dots, u_r)$ 。可将  $u_1, u_2, \dots, u_r$  扩充为  $\mathbb{C}^m$  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

$$U = [U_1 \mid U_2] = (u_1, u_2, \dots, u_r, u_{r+1}, \dots, u_m)$$

是  $m$  阶酉矩阵, 且有

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

于是可得

$$U^H A V = U^H [A V_1 \mid A V_2] = \begin{bmatrix} U_1^H \\ U_2^H \end{bmatrix} [U_1 \Sigma \mid O] = \begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}$$

于是可得

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

称上式为矩阵  $A$  的奇异值分解。

$A$  的奇异值由  $A$  唯一确定, 但是酉矩阵  $U$  和  $V$  不唯一。所以,  $A$  的奇异值分解不是唯一的。

奇异值分解 (SVD) 在数值线性代数中是一个强大的工具, 特别是在解决过定或欠定的线性系统中, 这在最小二乘问题中很常见。以下是在最小二乘背景下应用 SVD 的步骤:

1. 问题表述: 假设我们想要解决系统  $Ax = b$ , 其中  $A$  是一个  $m \times n$  矩阵,  $m \geq n$ ,  $b$  是一个  $m$  维向量。这个系统可能没有精确解, 因为它是过定的, 所以我们寻求一个  $x$ , 它可以最小化残差  $\|Ax - b\|^2$ 。
2. 执行 SVD: 计算  $A$  的 SVD, 将  $A$  分解为三个矩阵:

$$A = U \Sigma V^T$$

其中  $U$  是一个  $m \times m$  正交矩阵,  $\Sigma$  是一个  $m \times n$  对角矩阵, 对角线上是非负实数 (奇异值),  $V$  是一个  $n \times n$  正交矩阵。

3. 识别非零奇异值: 在  $\Sigma$  中, 奇异值通常按降序排列。如果  $A$  是满秩的, 所有奇异值都是非零的。但是, 如果  $A$  是秩亏的, 一些奇异值可能为零或接近零。
4. 计算伪逆: 使用 SVD 中的奇异值和正交矩阵计算  $A$  的伪逆  $A^+$ :

$$A^+ = V \Sigma^+ U^T$$

其中  $\Sigma^+$  是  $\Sigma$  的伪逆, 通过取每个非零奇异值的倒数并转置矩阵来形成。

5. 求解  $x$ : 使用伪逆来计算最小二乘解:

$$x = A^+b$$

这个解最小化了最小二乘准则  $\|Ax - b\|^2$ 。

6. 解的分析: 分析解  $x$  的稳定性和可靠性。如果奇异值非常小, 它们可能放大数据中的噪声, 导致不稳定的解。

7. 验证: 通过将  $x$  代回原始方程  $Ax$  并与  $b$  比较来验证解。差异  $\|Ax - b\|$  应该是最小的。

#### 4.4.2 举例展示求法

**例 16** 求矩阵  $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$  的奇异值分解。

**解:** 计算

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

矩阵  $B$  的特征值分别为  $\lambda_1 = 3, \lambda_2 = 1, \lambda_3 = 0$ , 对应的特征向量分别是

$$\vec{\xi}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \quad \vec{\xi}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad \vec{\xi}_3 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

于是可得

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

计算正交矩阵  $V$  如下

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

计算

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

构造

$$U_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad U = [U_1 \mid U_2] = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

则  $A$  的奇异值分解为

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

**例 17** 利用编程实现对例 16 中的矩阵  $A$  的奇异值分解。

**解：**编写 Python 代码如下：

```
from scipy.linalg import svd
```

```
A = np.array([[1, 0, 1], [0, 1, 1], [0, 0, 0]])
```

```
U, S, Vh = svd(A)
```

```
U, S, Vh
```

```
from scipy.linalg import svd

# 定义矩阵A
A = np.array([[1, 0, 1], [0, 1, 1], [0, 0, 0]])

# 进行奇异值分解
U, S, Vh = svd(A)

U, S, Vh

(array([[ 0.70710678, -0.70710678,  0.          ],
        [ 0.70710678,  0.70710678,  0.          ],
        [ 0.          ,  0.          ,  1.          ]]),
 array([ 1.73205081,  1.          , -0.          ]),
 array([[ 4.08248290e-01,  4.08248290e-01,  8.16496581e-01],
        [-7.07106781e-01,  7.07106781e-01,  2.22044605e-16],
        [ 5.77350269e-01,  5.77350269e-01, -5.77350269e-01]]))
```

图 5 Python QR 分解结果

**例 18** 给出应用奇异值分解式求解齐次线性方程组  $Ax = 0$  的方法。

把问题转化为求向量  $x$  的最优值使得  $\|Ax\|_2$  的值最小。我们已经知道了  $x = 0$  是该方程的一个特解，为了避免  $x = 0$  这种情况，我们增加一个约束，比如  $\|x\|_2 = 1$ ，这样，问题就变为（带约束的优化问题）

$$\min \|Ax\|, \quad \|x\| = 1$$

$$\min \|Ax\| = \min \|UDV^T x\| = \min \|DV^T x\| \quad \text{且} \quad \|x\| = \|V^T x\|$$

定义  $y = V^T x$  则问题变为： $\min \|Dy\|, \quad \|y\| = 1$ （因为  $V$  为正交矩阵）

由于  $D$  是一个对角矩阵，对角元素按降序排列，因此最优解在  $y = (0, 0, \dots, 1)$  时取得，因为  $x = Vy$ ，所以最优解就是  $V$  的最小奇异值对应的列向量，即  $x = v_n$ 。

## 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,甚至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**。

设矩阵  $A \in \mathbb{C}^{m \times n}$ , 若矩阵  $X \in \mathbb{C}^{m \times n}$  满足下面四个方程

$$\begin{cases} AXA = A \\ XAX = X \\ (AX)^H = AX \\ (XA)^H = XA \end{cases}$$

中的一个或几个,则称  $X$  为  $A$  的**广义逆**;若四个方程都满足,则称  $X$  为  $A$  的**Moore-Penrose 逆**,记为  $A^+$ 。矩阵  $A$  的广义逆  $A^+$  存在且唯一。若矩阵  $A$  可逆(满秩),则  $A^+ = A^{-1}$ 。 $A^+$  满足如下性质:

1.  $\text{rank } A = \text{rank } A^+$ ;
2.  $(A^+)^+ = A$ ;
3.  $(A^H)^+ = (A^+)^H$ ,  $(A^T)^+ = (A^+)^T$ ;
4.  $(A^H A)^+ = A^+ (A^H)^+$ ,  $(A A^H)^+ = (A^H)^+ A^+$ ;
5.  $A^+ = (A^H A)^+ A^H = A^H (A A^H)^+$ ;
6.  $R(A^+) = R(A^H)$ ,  $N(A^+) = N(A^H)$ ;
7. 若  $A \in \mathbb{C}_n^{m \times n}$ , 则  $A^+ = (A^H A)^{-1} A^H$ ;
8. 若  $A \in \mathbb{C}_m^{m \times n}$ , 则  $A^+ = A^H (A A^H)^{-1}$ 。

矩阵  $A$  广义逆  $A^+$  的等价定义为:  $AA^+ = P_{R(A)}$ ,  $A^+A = P_{R(A^+)}$ 。

### 4.5.2 利用矩阵满秩分解求矩阵广义逆

在 4.3 节中介绍了矩阵  $A \in \mathbb{C}_r^{m \times n}$  的满秩分解的概念,并给出了两个等价矩阵进行满秩分解的方法。设  $A \in \mathbb{C}_r^{m \times n} (r > 0)$  的满秩分解为

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

那么,可以按照例 6.1 和定理 6.1 的方法计算矩阵的 Moore-Penrose 逆,即

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

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

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

可以证明构造出的  $A^+$  满足上述的四个式子。

#### 4.5.3 利用矩阵奇异值分解求矩阵广义逆

设  $A \in \mathbb{C}_r^{m \times n} (r > 0)$  的奇异值分解为  $A = U D V^H$ , 有  $A = U_r \Delta V_r^H$ , 其中  $U_r$  是  $U$  的前  $r$  列,  $V_r$  是  $V$  的前  $r$  列,  $\Delta = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$ 。则

$$A^+ = V_r \Delta^{-1} U_r^H = V \begin{bmatrix} \Sigma^{-1} & O \\ O & O \end{bmatrix} U^H$$

可以证明构造出的  $A^+$  满足上述的四个式子。

#### 4.5.4 举例展示求法

**例 19** 设

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

求  $A^+$ 。

**解：**（方法一）利用满秩分解得

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

所以

$$\begin{aligned} A^+ &= G^+ F^+ = G^H (G G^H)^{-1} (F F^H)^{-1} F^H \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \left( \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)^{-1} \left( \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 2 \end{bmatrix} \\ &= \frac{1}{10} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} \\ &= \frac{1}{10} \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \end{aligned}$$

（方法二）利用奇异值分解得

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

令

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

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

所以

$$A^+ = V_1 \Delta^{-1} U_1^H = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$$

(方法三) 利用 Python 代码如下:

```
from scipy.linalg import pinv
```

```
A = np.array([[1, 1], [2, 2]])
```

```
A_pinv = pinv(A)
```

```
A_pinv
```

```
from scipy.linalg import pinv

# 定义矩阵A
A = np.array([[1, 1], [2, 2]])

# 计算广义逆 (Moore-Penrose pseudo-inverse)
A_pinv = pinv(A)

A_pinv

array([[0.1, 0.2],
       [0.1, 0.2]])
```

图 6 Python 求广义逆结果

## 第五章 总结

本论文全面系统地研究了矩阵函数的求解方法和矩阵的分解技术。首先, 论文通过对待定系数法、数项级数求和法、对角型法以及若尔当标准型法等求解矩阵函数的传统方法进行了深入分析和总结。特别是, 引入了若尔当标准型法, 这一创新点显著扩展了矩阵函数求解方法的适用范围, 提升了理论的实际应用价值。

在矩阵分解的研究部分, 论文不仅详细探讨了 LU 分解、QR 分解、满秩分解、奇异值分解等经典方法, 还对 PLU 分解等分解方法进行了综述。通过丰富的数学推导与实例, 清晰地展示了每种分解方法的理论基础和计算步骤, 强调了它们在解决实际问题中的实用性和有效性。

论文的另一个亮点是将理论知识与当代计算工具相结合, 提供了矩阵分解方法的 Python 代码实现。这些编程示例不仅有助于读者更好地理解和掌握矩阵理论, 而且为矩阵理论的学习和应用提供了便捷的工具, 增强了论文的实用性和互动性。

同时, 论文也关注了如奇异值分解和广义逆在最小二乘法最优化问题中的应用, 深入分析了它们在统计建模和数据分析中的关键作用。通过对这些高级数学工具在解决实际问题中的应用进行阐述, 论文不仅拓宽了矩阵理论的应用领域, 还为读者提供了一种理论与实践相结合的研究视角。

总体来说, 本论文对本学期矩阵理论课程的学习和应用进行了总结和反思, 对矩阵理论的研究和应用提供了更多思路和方法, 具有一定的理论和实践价值。

## 参考文献

- [1] 张凯院, 徐仲等. 矩阵论[M]. 西北工业大学出版社. 2017 年 8 月.
- [2] 矩阵论每周作业 (见文末)