PROBLEM SOLUTIONS FOR MATRIX ANALYSIS AND APPLICATIONS

矩阵分析与应用习题解答

李剑　李细林　苏泳涛　丁子哲　编
Li Jian　Li Xilin　Su Yongtao　Ding Zizhe

张贤达　审校
Zhang Xianda

清华大学出版社

ISBN 978-7-302-14553-0

9 787302 145530 >

定价：29.00元

内容简介

本书是研究生教材《矩阵分析与应用》的配套参考书，由矩阵与线性方程组、特殊矩阵、矩阵的变换与分解、梯度分析与最优化、奇异值分析、总体最小二乘方法、特征分析、子空间分析与跟踪、投影分析共9章组成。每章均包含两部分内容：第一部分总结复习该章所涉及的主要理论知识，第二部分为习题的详细解答。所选习题分为基础题型、综合题型、应用题型。这些习题可以帮助读者巩固加深对基础概念的理解，提高综合运用知识的技能和解决实际应用问题的能力。

本书可供电子、通信、自动化、计算机等学科的研究生学习辅导之用，也可供相关专业和领域的教师和科研人员参考。

图书在版编目（CIP）数据

矩阵分析与应用习题解答/李剑等编. —北京：清华大学出版社，2007.5
ISBN 978-7-302-14553-0

Ⅰ. 矩…　Ⅱ. 李…　Ⅲ. 矩阵分析—高等学校—解题　Ⅳ. O151.21-44

中国版本图书馆 CIP 数据核字(2007)第 007519 号

责任编辑：王一玲
责任校对：白　蕾
责任印制：何　芊

出版发行：清华大学出版社
http://www.tup.com.cn
c-service@tup.tsinghua.edu.cn
社 总 机：010-62770175
投稿咨询：010-62772015
地　址：北京清华大学学研大厦 A 座
邮　编：100084
邮购热线：010-62786544
客户服务：010-62776969
印 装 者：北京鑫海金澳胶印有限公司
经　销：全国新华书店
开　本：185×260　**印　张**：19　**字　数**：436千字
版　次：2007年5月第1版　**印　次**：2007年5月第1次印刷
印　数：1～3000
定　价：29.00元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题，请与清华大学出版社出版部联系调换。联系电话：010-62770177 转 3103　　产品编号：023060-01

前　　言

作为数学的一个分支，矩阵理论是众多理工学科的重要数学工具，它极大地推动了信号与信息处理、通信、控制、航天等领域的发展。很多高校都开设有矩阵的相关课程。而要学好它，做习题是一个必不可少的环节。目前，国内外比较全面的矩阵理论习题解答书籍很少。鉴于此，作者编著了本书。

本书由矩阵与线性方程组、特殊矩阵、矩阵的变换与分解、梯度分析与最优化、奇异值分析、总体最小二乘方法、特征分析、子空间分析与跟踪、投影分析共9章组成，所包含的320道习题全部选自张贤达教授编著的《矩阵分析与应用》(清华大学出版社，2004)。每章又分为两个部分：前半章为总结与复习主要理论知识，后半章为习题的详细解答。

华罗庚先生在其《高等数学引论》的序言中精辟地论述到：习题的目的首先是熟悉和巩固学习了的东西；其二是启发大家灵活运用，独立思考；其三是融会贯通。本书的习题也充分体现了这三点，这些习题大致可以分为三类：基础题型用以巩固所学的知识，加深对基本概念的理解；综合题型训练读者灵活运用知识的能力，有一定的难度；应用题型大多选自信息科学方面的国际著名杂志论文，是对某个实际问题的数学抽象，有很强的开放性和可推广性，因此有利于培养读者运用所学理论知识解决实际应用问题的能力，并体会前沿研究中的矩阵应用，但读者在解答这类习题时并不需要专业背景知识。部分综合型和应用型习题在清华大学矩阵分析课程中被选为作业题，收到了良好的效果。

全书共9章，其中第1、2章由苏泳涛编写，第1、3、6章由丁子哲编写，第5、7章由李细林编写，第4、8、9 章由李剑编写，张贤达教授负责全书的审校。全书使用LaTex排版，符号和编排与《矩阵分析与应用》保持一致。

编写本书的主旨，当然不是“越俎代庖”，不可将书中的解答看作是“标准答案”。它们应作为读者在经过自己的独立思考而后对照的“参考答案”。若再进而起到抛砖引玉的作用，启发出更漂亮的解答，将是作者莫大的欣慰。

在本书编写过程中，作者得到了课题组高秋彬、朱峰、张玲、王煜航、胡亚峰、施维、张莉、武露、谢德光、李培胜、常冬霞、张道明的帮助，在此向他们表示衷心的感谢。

囿于作者水平有限，错误之处在所难免，恳请读者批评指正。

作者

2006年10月于清华园

目　录

第 1 章　矩阵与线性方程组

矩阵是描述和求解线性方程组最基本和最有用的工具。本章涉及向量和矩阵的基本概念，归纳了向量和矩阵的基本运算。

1.1　主要理论与方法

1.1.1　矩阵的基本运算

一、矩阵与向量

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}\right\} \tag{1.1}$$

它使用m个方程描述n个未知量之间的线性关系。这一线性方程组很容易用矩阵——向量形式简记为

$$\boldsymbol{Ax} = \boldsymbol{b} \tag{1.2}$$

式中

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{1.3}$$

称为$m \times n$矩阵，是一个按照长方阵列排列的复数或实数集合；而

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \tag{1.4}$$

分别为$n \times 1$向量和$m \times 1$向量，是按照列方式排列的复数或实数集合，统称列向量。类似地，按照行方式排列的复数或实数集合称为行向量，例如

$$\boldsymbol{a} = [a_1, a_2, \cdots, a_n] \tag{1.5}$$

是$1 \times n$向量。

二、矩阵的基本运算

1. 共轭转置：若$\boldsymbol{A}=[a_{ij}]$是一个$m\times n$矩阵，则$\boldsymbol{A}$的转置记作$\boldsymbol{A}^{\mathrm{T}}$，是一个$n\times m$矩阵，定义为$[\boldsymbol{A}^{\mathrm{T}}]_{ij}=a_{ji}$；矩阵$\boldsymbol{A}$的复数共轭$\boldsymbol{A}^*$ 定义为$[\boldsymbol{A}^*]_{ij}=a_{ji}^*$；复共轭转置记作$\boldsymbol{A}^{\mathrm{H}}$，定义为

$$\boldsymbol{A}^{\mathrm{H}}=\begin{bmatrix} a_{11}^* & a_{21}^* & \cdots & a_{m1}^* \\ a_{12}^* & a_{22}^* & \cdots & a_{m2}^* \\ \vdots & \vdots & & \vdots \\ a_{1n}^* & a_{2n}^* & \cdots & a_{mn}^* \end{bmatrix} \tag{1.6}$$

共轭转置又叫Hermitian伴随、Hermitian转置或Hermitian共轭。满足$\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$的正方复矩阵称为Hermitian矩阵或共轭对称矩阵。

2. 矩阵求和：两个$m\times n$矩阵$\boldsymbol{A}=[a_{ij}]$和$\boldsymbol{B}=[b_{ij}]$之和记作$\boldsymbol{A}+\boldsymbol{B}$，定义为$[\boldsymbol{A}+\boldsymbol{B}]_{ij}=a_{ij}+b_{ij}$。

3. 标量与矩阵相乘：令$\boldsymbol{A}=[a_{ij}]$是一个$m\times n$矩阵，且α是一个标量。乘积$\alpha\boldsymbol{A}$是一个$m\times n$矩阵，定义为$[\alpha\boldsymbol{A}]_{ij}=\alpha a_{ij}$。

4. 矩阵与向量相乘：$m\times n$矩阵$\boldsymbol{A}=[a_{ij}]$与$r\times 1$向量$\boldsymbol{x}=[x_1,x_2,\cdots,x_r]^{\mathrm{T}}$的乘积$\boldsymbol{A}\boldsymbol{x}$只有当$n=r$时才存在，它是一个$m\times 1$向量，定义为

$$[\boldsymbol{A}\boldsymbol{x}]_i=\sum_{j=1}^{n}a_{ij}x_j,\quad i=1,2,\cdots,m$$

5. 矩阵与矩阵相乘：$m\times n$矩阵$\boldsymbol{A}=[a_{ij}]$与$r\times s$矩阵$\boldsymbol{B}=[b_{ij}]$的乘积$\boldsymbol{AB}$只有当$n=r$时才存在，它是一个$m\times s$矩阵，定义为

$$[\boldsymbol{AB}]_{ij}=\sum_{k=1}^{n}a_{ik}b_{kj},\quad i=1,2,\cdots,m;\ j=1,2,\cdots,s$$

根据定义，容易验证矩阵的加法服从下面的运算规则。

- 加法交换律(commutative law of addition)：$\boldsymbol{A}+\boldsymbol{B}=\boldsymbol{B}+\boldsymbol{A}$
- 加法结合律(associative law of addition)：$(\boldsymbol{A}+\boldsymbol{B})+\boldsymbol{C}=\boldsymbol{A}+(\boldsymbol{B}+\boldsymbol{C})$

定理 1.1 矩阵的乘积服从下面的运算法则。

(1) 乘法结合律(associative law of multiplication)：若$\boldsymbol{A}\in C^{m\times n},\boldsymbol{B}\in C^{n\times p},\boldsymbol{C}\in C^{p\times q}$，则$\boldsymbol{A}(\boldsymbol{BC})=(\boldsymbol{AB})\boldsymbol{C}$。

(2) 乘法左分配律(left distributive law of multiplication)：若$\boldsymbol{A}$和$\boldsymbol{B}$是两个$m\times n$ 矩阵，且$\boldsymbol{C}$是一个$n\times p$矩阵，则$(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C}=\boldsymbol{AC}+\boldsymbol{BC}$。

(3) 乘法右分配律(right distributive law of multiplication)：若$\boldsymbol{A}$是一个$m\times n$ 矩阵，并且$\boldsymbol{B}$和$\boldsymbol{C}$是两个$n\times p$矩阵，则$\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C})=\boldsymbol{AB}+\boldsymbol{AC}$。

(4) 若α是一个标量，并且$\boldsymbol{A}$和$\boldsymbol{B}$是两个$m\times n$矩阵，则$\alpha(\boldsymbol{A}+\boldsymbol{B})=\alpha\boldsymbol{A}+\alpha\boldsymbol{B}$。

6. 逆矩阵：令$\boldsymbol{A}$是一个$n\times n$矩阵，若可以找到一个$n\times n$矩阵$\boldsymbol{A}^{-1}$满足$\boldsymbol{A}\boldsymbol{A}^{-1}=\boldsymbol{A}^{-1}\boldsymbol{A}=\boldsymbol{I}$，称矩阵$\boldsymbol{A}$可逆，并称$\boldsymbol{A}^{-1}$是矩阵$\boldsymbol{A}$的逆矩阵。

下面是共轭、转置、共轭转置和逆矩阵的性质。

(1) 矩阵的共轭、转置和共轭转置满足分配律：

$$(\boldsymbol{A}+\boldsymbol{B})^* = \boldsymbol{A}^* + \boldsymbol{B}^*$$
$$(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}} + \boldsymbol{B}^{\mathrm{T}}$$
$$(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{H}} = \boldsymbol{A}^{\mathrm{H}} + \boldsymbol{B}^{\mathrm{H}}$$

(2) 矩阵乘积的转置、共轭转置和逆矩阵满足关系式

$$(\boldsymbol{AB})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$$
$$(\boldsymbol{AB})^{\mathrm{H}} = \boldsymbol{B}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}$$
$$(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1} \quad (\boldsymbol{A},\boldsymbol{B}\ \text{为可逆的正方矩阵})$$

(3) 共轭、转置和共轭转置等符号均可与求逆符号交换，即有

$$(\boldsymbol{A}^*)^{-1} = (\boldsymbol{A}^{-1})^*, \quad (\boldsymbol{A}^{\mathrm{T}})^{-1} = (\boldsymbol{A}^{-1})^{\mathrm{T}}, \quad (\boldsymbol{A}^{\mathrm{H}})^{-1} = (\boldsymbol{A}^{-1})^{\mathrm{H}}$$

因此，常常分别采用紧凑的数学符号$\boldsymbol{A}^{-*}, \boldsymbol{A}^{-\mathrm{T}}$和$\boldsymbol{A}^{-\mathrm{H}}$。

(4) 对于任意矩阵$\boldsymbol{A}$，矩阵$\boldsymbol{B}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$都是Hermitian 矩阵。若$\boldsymbol{A}$可逆，则对于Hermitian矩阵$\boldsymbol{B}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$，有$\boldsymbol{A}^{-\mathrm{H}}\boldsymbol{B}\boldsymbol{A}^{-1}=\boldsymbol{A}^{-\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{A}^{-1}=\boldsymbol{I}$。

7. 幂等矩阵：矩阵$\boldsymbol{A}_{n\times n}$称为幂等矩阵(idempotent matrix)，若$\boldsymbol{A}^2=\boldsymbol{AA}=\boldsymbol{A}$。

8. 对合矩阵：矩阵$\boldsymbol{A}_{n\times n}$称为对合矩阵(involutory matrix)，若$\boldsymbol{A}^2=\boldsymbol{AA}=\boldsymbol{I}$。

三、向量的线性无关性与非奇异矩阵

1. 向量组线性相关/无关：一组m维向量$\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_n\}$称为线性无关，若方程

$$c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \cdots + c_n\boldsymbol{u}_n = \boldsymbol{0}$$

只有零解$c_1=c_2=\cdots=c_n=0$。若能够找到一组不全部为零的系数$c_1,c_2,\cdots,c_n$使得上述方程成立，则称m维向量组$\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_n\}$线性相关。

2. 奇异/非奇异矩阵：一个$n\times n$矩阵$\boldsymbol{A}$是非奇异的，当且仅当矩阵方程$\boldsymbol{Ax}=\boldsymbol{0}$只有零解$\boldsymbol{x}=\boldsymbol{0}$。若$\boldsymbol{A}$不是非奇异的，则称$\boldsymbol{A}$ 奇异。

四、初等行变换与阶梯型矩阵

1. 矩阵的初等行变换：令矩阵$\boldsymbol{A}\in C^{m\times n}$的$m$个行向量分别为$\boldsymbol{r}_1,\boldsymbol{r}_2,\cdots,\boldsymbol{r}_m$。下列运算称为矩阵$\boldsymbol{A}$的初等行运算(elementary row operation)或初等行变换：

(1) 互换矩阵的任意两行，如$\boldsymbol{r}_p\leftrightarrow\boldsymbol{r}_q$，称为 I 型初等行变换。

(2) 一行元素同乘一个非零常数α，如$\alpha\boldsymbol{r}_p\to\boldsymbol{r}_p$，称为 II 型初等行变换。

(3) 将第p行元素同乘一个非零常数β后，加给第q行，即$\beta\boldsymbol{r}_p+\boldsymbol{r}_q\to\boldsymbol{r}_q$，称为III型初等行变换。

2. 阶梯型矩阵：一个$m \times n$矩阵称为阶梯型(echelon form)矩阵，若

(1) 全部由零组成的所有行都位于矩阵的底部。

(2) 每一个非零行的首项元素总是出现在上一个非零行的首项元素的右边。

(3) 首项元素下面的同列元素全部为零。

1.1.2 向量空间、内积空间与线性映射

一、向量空间

以向量为元素的集合V称为向量空间，若加法运算定义为两个向量之间的加法，乘法运算定义为向量与标量域S中的标量之间的乘法，并且对于向量集合V中的向量$\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}$和标量域$S$中的标量$a_1, a_2$，以下两个闭合性和关于加法及乘法的八个公理(axiom) [也称公设(postulate)或定律(law)]满足：

闭合性(closure properties)

(c1) 若$\boldsymbol{x} \in V$和$\boldsymbol{y} \in V$，则$\boldsymbol{x} + \boldsymbol{y} \in V$，即$V$在加法下是闭合的，简称加法的闭合性(closure for addition)；

(c2) 若a_1是一个标量，$\boldsymbol{y} \in V$，则$a_1\boldsymbol{y} \in V$，即V在标量乘法下是闭合的，简称标量乘法的闭合性(closure for scalar multiplication)。

加法的公理

(a1) $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$, $\forall \boldsymbol{x}, \boldsymbol{y} \in V$，称为加法的交换律(commutative law for addition)；

(a2) $\boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{w}) = (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{w}$, $\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w} \in V$，称为加法的结合律(associative law for addition)；

(a3) 在V中存在一个零向量$\mathbf{0}$，使得对于任意向量$\boldsymbol{y} \in V$，恒有$\boldsymbol{y} + \mathbf{0} = \boldsymbol{y}$ (零向量的存在性)；

(a4) 给定一个向量$\boldsymbol{y} \in V$，存在另一个向量$-\boldsymbol{y} \in V$使得$\boldsymbol{y} + (-\boldsymbol{y}) = (-\boldsymbol{y}) + \boldsymbol{y} = \mathbf{0}$ (负向量的存在性)。

标量乘法的公理

(s1) $a(b\boldsymbol{y}) = (ab)\boldsymbol{y}$对所有向量$\boldsymbol{y}$和所有标量$a, b$成立，称为标量乘法的结合律(associative law for scalar multiplication)；

(s2) $a(\boldsymbol{x} + \boldsymbol{y}) = a\boldsymbol{x} + a\boldsymbol{y}$对所有向量$\boldsymbol{x}, \boldsymbol{y} \in V$和标量$a$ 成立，称为标量乘法的分配律(distributive law for scalar multiplication)；

(s3) $(a + b)\boldsymbol{y} = a\boldsymbol{y} + b\boldsymbol{y}$对所有向量$\boldsymbol{y}$和所有标量$a, b$成立(标量乘法的分配律)；

(s4) $1\boldsymbol{y} = \boldsymbol{y}$ 对所有$\boldsymbol{y} \in V$成立，称为标量乘法的单位律(unity law for scalar multiplication)。

二、实内积空间

实内积空间(real inner product space)是满足下列条件的实向量空间E：对E中每一对向量$\boldsymbol{x}, \boldsymbol{y}$，存在向量$\boldsymbol{x}$和$\boldsymbol{y}$的内积$\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ 服从以下公理：

(1) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle > 0, \ \forall \ \boldsymbol{x} \neq \mathbf{0}$，称为内积的严格正性(strict positivity)或称内积是正定的(positive definite)，并且$\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0 \Leftrightarrow \boldsymbol{x} = \mathbf{0}$；

(2) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$，称为内积的对称性(symmetry)；

(3) $\langle \boldsymbol{x}, \boldsymbol{y}+\boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle, \ \forall \ \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$；

(4) $\langle \alpha\boldsymbol{x}, \boldsymbol{y} \rangle = \alpha\langle \boldsymbol{x}, \boldsymbol{y} \rangle$对所有实向量$\boldsymbol{x}, \boldsymbol{y}$及所有实标量$\alpha$成立。

三、复内积空间

复内积空间(complex inner product space)是满足下列条件的复向量空间C：对C中每一对向量$\boldsymbol{x}, \boldsymbol{y}$，存在复向量$\boldsymbol{x}$和$\boldsymbol{y}$之间的内积$\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ 服从以下公理：

(1) $\boldsymbol{x} \neq \mathbf{0} \Rightarrow \langle \boldsymbol{x}, \boldsymbol{x} \rangle > 0$，称为内积的严格正性或称内积是正定的；

(2) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle^* = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$，称为内积的共轭对称性(conjugate symmetry)或Hermitian性；

(3) $\langle \boldsymbol{x}, \boldsymbol{y}+\boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle$，对所有向量$\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$成立；

(4) $\langle c\boldsymbol{x}, \boldsymbol{y} \rangle = c^*\langle \boldsymbol{x}, \boldsymbol{y} \rangle$对所有复向量$\boldsymbol{x}, \boldsymbol{y}$ 及所有复标量c成立。

四、线性映射

令V和W分别是R^m和R^n 的子空间，并且$T: V \mapsto W$是一映射。称T为线性映射或线性变换，若对于$\boldsymbol{v} \in V, \boldsymbol{w} \in W$和所有标量$c$，映射$T$满足线性关系式

$$T(\boldsymbol{v}+\boldsymbol{w}) = T(\boldsymbol{v}) + T(\boldsymbol{w}) \tag{1.7}$$

和

$$T(c\boldsymbol{v}) = cT(\boldsymbol{v}) \tag{1.8}$$

1.1.3 随机向量

一、随机向量的统计描述

1. 均值向量：考查$m\times 1$随机向量$\boldsymbol{x}(\xi) = [x_1(\xi), x_2(\xi), \cdots, x_m(\xi)]^{\mathrm{T}}$。令随机变量$x_i(\xi)$的均值$\mathrm{E}\{x_i(\xi)\} = \mu_i$，则随机向量的数学期望称为均值向量，记作$\boldsymbol{\mu}_x$，定义为

$$\boldsymbol{\mu}_x = \mathrm{E}\{\boldsymbol{x}(\xi)\} = \begin{bmatrix} \mathrm{E}\{x_1(\xi)\} \\ \mathrm{E}\{x_2(\xi)\} \\ \vdots \\ \mathrm{E}\{x_m(\xi)\} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{bmatrix} \tag{1.9}$$

式(1.9)表明，均值向量的元素是随机向量各个元素的均值。

2. 自相关矩阵：随机向量的自相关矩阵定义为

$$\boldsymbol{R}_x \overset{\mathrm{def}}{=} \mathrm{E}\{\boldsymbol{x}(\xi)\boldsymbol{x}^{\mathrm{H}}(\xi)\} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mm} \end{bmatrix} \tag{1.10}$$

式中，$r_{ii}, i=1,2,\cdots,m$表示随机变量$x_i(\xi)$的自相关函数，定义为

$$r_{ii} \stackrel{\text{def}}{=} \mathrm{E}\{|x_i(\xi)|^2\}, \quad i=1,2,\cdots,m \tag{1.11}$$

而r_{ij}表示随机变量$x_i(\xi)$和$x_j(\xi)$之间的互相关函数，定义为

$$r_{ij} \stackrel{\text{def}}{=} \mathrm{E}\{x_i(\xi)x_j^*(\xi)\}, \quad i,j=1,2,\cdots,m,\ i\neq j \tag{1.12}$$

显然，自相关矩阵是共轭对称的，即为Hermitian矩阵。

3. 自协方差矩阵：随机向量$\boldsymbol{x}(\xi)$的自协方差矩阵定义为

$$\boldsymbol{C}_x \stackrel{\text{def}}{=} \mathrm{E}\{[\boldsymbol{x}(\xi)-\boldsymbol{\mu}_x][\boldsymbol{x}(\xi)-\boldsymbol{\mu}_x]^{\mathrm{H}}\} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & c_{22} & \cdots & c_{2m} \\ \vdots & \vdots & & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mm} \end{bmatrix} \tag{1.13}$$

式中，主对角线的元素

$$c_{ii} \stackrel{\text{def}}{=} \mathrm{E}\{|x_i(\xi)-\mu_i|^2\}, \quad i=1,2,\cdots,m \tag{1.14}$$

表示随机变量$x_i(\xi)$的方差σ_i^2，即$c_{ii}=\sigma_i^2$，而非主对角线元素

$$c_{ij} \stackrel{\text{def}}{=} \mathrm{E}\{[x_i(\xi)-\mu_i][x_j(\xi)-\mu_j]^*\} = \mathrm{E}\{x_i(\xi)x_j^*(\xi)\} - \mu_i\mu_j^* = c_{ji}^* \tag{1.15}$$

表示随机变量$x_i(\xi)$和$x_j(\xi)$之间的协方差。自协方差矩阵也是Hermitian矩阵。

随机向量间的互相关矩阵与互协方差矩阵很容易由自相关矩阵与自协方差矩阵推广得到。

4. 统计不相关：两个随机向量$\boldsymbol{x}(\xi)$与$\boldsymbol{y}(\xi)$统计不相关，若它们的互协方差矩阵等于零矩阵，即$\boldsymbol{C}_{xy}=\boldsymbol{O}$。

5. 正交：两个随机向量$\boldsymbol{x}(\xi)$和$\boldsymbol{y}(\xi)$称为正交，若它们的互相关矩阵为零矩阵，即

$$\boldsymbol{R}_{xy} = \mathrm{E}\{\boldsymbol{x}(\xi)\boldsymbol{y}^{\mathrm{H}}(\xi)\} = \boldsymbol{O} \tag{1.16}$$

二、正态随机向量

若随机向量$\boldsymbol{x}(\xi) = [x_1(\xi), x_2(\xi), \cdots, x_m(\xi)]^{\mathrm{T}}$的各分量为联合正态分布的随机变量，则称$\boldsymbol{x}(\xi)$为正态随机向量。

1.1.4 内积与范数

一、向量的内积与范数

1. 常数向量的内积与范数

(1) 内积：两个$m\times 1$维常数向量$\boldsymbol{x} = [x_1, x_2, \cdots, x_m]^{\mathrm{T}}$ 和$\boldsymbol{y} = [y_1, y_2, \cdots, y_m]^{\mathrm{T}}$的内积(或叫点积)定义为

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \boldsymbol{x}^{\mathrm{H}}\boldsymbol{y} = \sum_{i=1}^{m} x_i^* y_i \tag{1.17}$$

(2) 范数：

(a) l_1范数

$$\|\boldsymbol{x}\|_1 \stackrel{\text{def}}{=} \sum_{i=1}^{m} |x_i| = |x_1| + |x_2| + \cdots + |x_m| \tag{1.18}$$

上述范数有时也叫和范数或1范数。

(b) l_2范数

$$\|\boldsymbol{x}\|_2 = (|x_1|^2 + |x_2|^2 + \cdots + |x_m|^2)^{1/2} \tag{1.19}$$

这一范数常称Euclidean范数，有时也称Frobenius范数。

(c) l_∞范数

$$\|\boldsymbol{x}\|_\infty = \max(|x_1|, |x_2|, \cdots, |x_m|) \tag{1.20}$$

也称无穷范数或极大范数。

(d) l_p范数

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{m} |x_i|^p\right)^{1/p}, \qquad p \geqslant 1 \tag{1.21}$$

l_p范数也叫Hölder范数[20]。

2. 随机向量的内积与范数

(1) 内积：若$\boldsymbol{x}(\xi)$和$\boldsymbol{y}(\xi)$ 分别是样本变量ξ的随机向量，则它们的内积定义为

$$\langle \boldsymbol{x}(\xi), \boldsymbol{y}(\xi) \rangle \stackrel{\text{def}}{=} \mathrm{E}\{\boldsymbol{x}^{\mathrm{H}}(\xi)\boldsymbol{y}(\xi)\} \tag{1.22}$$

其中，样本变量ξ可以是时间t、圆频率f、角频率ω和空间变量s等。

(2) 范数：随机向量$\boldsymbol{x}(\xi)$的范数定义为

$$\|\boldsymbol{x}(\xi)\|^2 \stackrel{\text{def}}{=} \mathrm{E}\{\boldsymbol{x}^{\mathrm{H}}(\xi)\boldsymbol{x}(\xi)\} \tag{1.23}$$

二、矩阵的范数

1. Frobenius范数

$$\|\boldsymbol{A}\|_{\mathrm{F}} \stackrel{\text{def}}{=} \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2\right)^{1/2} \tag{1.24}$$

这一定义可以视为向量的Euclidean范数对按照矩阵各行排列“长向量”

$$\boldsymbol{x} = [a_{11}, \cdots, a_{1n}, a_{21}, \cdots, a_{2n}, \cdots, a_{m1}, \cdots, a_{mn}]^{\mathrm{T}}$$

的推广。矩阵的Frobenius范数也被称为Euclidean范数、Schur范数、Hilbert-Schmidt范数或者l_2范数。

2. l_p范数

$$\|\boldsymbol{A}\|_p \stackrel{\text{def}}{=} \max_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p} \tag{1.25}$$

式中，$\|\boldsymbol{x}\|_p$是向量$\boldsymbol{x}$的l_p范数，由式(1.21)定义。l_p范数也称为Minkowski p 范数，或者简称p范数。

3. 行和范数(row-sum norm)

$$\|\boldsymbol{A}\|_{\text{row}} = \max_{1\leqslant i\leqslant m}\left\{\sum_{j=1}^{n}|a_{ij}|\right\} \tag{1.26}$$

4. 列和范数(column-sum norm)

$$\|\boldsymbol{A}\|_{\text{col}} = \max_{1\leqslant j\leqslant n}\left\{\sum_{i=1}^{m}|a_{ij}|\right\} \tag{1.27}$$

5. 谱范数(spectrum norm)

$$\|\boldsymbol{A}\|_{\text{spec}} = \sigma_{\max} = \sqrt{\lambda_{\max}} \tag{1.28}$$

1.1.5 基与Gram-Schmidt正交化

一、向量子空间的基

生成子空间W的线性无关的向量$\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_d\}$称为子空间$W$的基向量(basis vectors) 或简称为基。生成子空间W的基向量的个数称为子空间W 的维数，即有

$$d = \dim(\text{Span}\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_d\}) \tag{1.29}$$

二、Gram-Schmidt正交化

令$\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$ 是p维向量子空间W的任意一组基(即线性无关的向量)。于是，子空间W的标准正交基$\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_n\}$可以通过Gram-Schmidt正交化构造如下：

$$\begin{aligned}
&\boldsymbol{p}_1 = \boldsymbol{x}_1, \qquad \boldsymbol{u}_1 = \frac{\boldsymbol{p}_1}{\|\boldsymbol{p}_1\|} = \frac{\boldsymbol{x}_1}{\|\boldsymbol{x}_1\|}\\
&\boldsymbol{p}_k = \boldsymbol{x}_k - \sum_{i=1}^{k-1}(\boldsymbol{u}_i^{\text{H}}\boldsymbol{x}_k)\boldsymbol{u}_i, \qquad \boldsymbol{u}_k = \frac{\boldsymbol{p}_k}{\|\boldsymbol{p}_k\|}
\end{aligned} \tag{1.30}$$

式中，$2\leqslant k\leqslant n$。

1.1.6 矩阵的标量函数

一、矩阵的二次型

任意一个正方矩阵$\boldsymbol{A}$的二次型$\boldsymbol{x}^{\text{H}}\boldsymbol{A}\boldsymbol{x}$是一实标量。以实矩阵为例，考查二次型

$$\begin{aligned}
\boldsymbol{x}^{\text{T}}\boldsymbol{A}\boldsymbol{x} &= [x_1,x_2,x_3]\begin{bmatrix}1 & 4 & 2\\ -1 & 7 & 5\\ -1 & 6 & 3\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix}\\
&= x_1^2 - x_2x_1 - x_3x_1 + 4x_1x_2 + 7x_2^2 + 6x_3x_2 + 2x_1x_3 + 5x_2x_3 + 3x_3^2\\
&= x_1^2 + 7x_2^2 + 3x_3^2 + 3x_1x_2 + x_1x_3 + 11x_2x_3
\end{aligned}$$

这是变元x的二次型函数，故称$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$为矩阵$\boldsymbol{A}$的二次型。

一个复共轭对称矩阵$\boldsymbol{A}$称为

正定矩阵，若二次型$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x} > 0, \quad \forall \boldsymbol{x} \neq \boldsymbol{0}$;

半正定矩阵，若二次型$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x} \geqslant 0, \quad \forall \boldsymbol{x} \neq \boldsymbol{0}$ (也称非负定的);

负定矩阵，若二次型$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x} < 0, \quad \forall \boldsymbol{x} \neq \boldsymbol{0}$;

半负定矩阵，若二次型$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x} \leqslant 0, \quad \forall \boldsymbol{x} \neq \boldsymbol{0}$ (也称非正定的);

不定矩阵，若二次型$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}$既可能取正值，也可能取负值。

二、矩阵的迹

1. 定义：$n \times n$矩阵$\boldsymbol{A}$的对角元素之和称为$\boldsymbol{A}$的迹(trace)，记作$\mathrm{tr}(\boldsymbol{A})$，即

$$\mathrm{tr}(\boldsymbol{A}) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii} \tag{1.31}$$

2. 性质：

(1) 关于迹的等式 [27]

(a) 若$\boldsymbol{A}$和$\boldsymbol{B}$ 均为$n \times n$矩阵，则$\mathrm{tr}(\boldsymbol{A} \pm \boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) \pm \mathrm{tr}(\boldsymbol{B})$。

(b) 若c是一个复或者实的常数，则$\mathrm{tr}(c\boldsymbol{A}) = c\,\mathrm{tr}(\boldsymbol{A})$。

(c) 若$\boldsymbol{A}$和$\boldsymbol{B}$ 均为$n \times n$矩阵，并且c_1和c_2为常数，则$\mathrm{tr}(c_1\boldsymbol{A} \pm c_2\boldsymbol{B}) = c_1\mathrm{tr}(\boldsymbol{A}) \pm c_2\mathrm{tr}(\boldsymbol{B})$。

(d) 矩阵$\boldsymbol{A}$的转置、复数共轭和复共轭转置的迹分别为

$$\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}) = \mathrm{tr}(\boldsymbol{A})$$
$$\mathrm{tr}(\boldsymbol{A}^{*}) = [\mathrm{tr}(\boldsymbol{A})]^{*}$$
$$\mathrm{tr}(\boldsymbol{A}^{\mathrm{H}}) = [\mathrm{tr}(\boldsymbol{A})]^{*}$$

(e) 迹是相似不变量：若$\boldsymbol{A}$为$m \times n$ 矩阵，且$\boldsymbol{B}$为$n \times m$矩阵，则

$$\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{B}\boldsymbol{A})$$

(f) 若矩阵$\boldsymbol{A}$和$\boldsymbol{B}$均为$m \times m$矩阵，并且$\boldsymbol{B}$非奇异，则

$$\mathrm{tr}(\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}^{-1}) = \mathrm{tr}(\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A})$$

(g) 若$\boldsymbol{A}$是一个$m \times n$矩阵，则$\mathrm{tr}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) = 0 \Leftrightarrow \boldsymbol{A} = \boldsymbol{O}_{m \times n}$ (零矩阵)。

(h) $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x} = \mathrm{tr}(\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}})$和$\boldsymbol{y}^{\mathrm{H}}\boldsymbol{x} = \mathrm{tr}(\boldsymbol{x}\boldsymbol{y}^{\mathrm{H}})$。

(i) 分块矩阵的迹满足

$$\mathrm{tr}\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{D})$$

式中，$\boldsymbol{A} \in C^{m \times m}, \boldsymbol{B} \in C^{m \times n}, \boldsymbol{C} \in C^{n \times m}, \boldsymbol{D} \in C^{n \times n}$。

(j) 矩阵$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 和$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$的迹相等，且有

$$\mathrm{tr}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}) = \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|^2 \tag{1.32}$$

(k) 迹等于特征值之和，即

$$\mathrm{tr}(\boldsymbol{A}) = \lambda_1 + \lambda_2 + \cdots + \lambda_n \tag{1.33}$$

(l) 对于任何正整数k，有

$$\mathrm{tr}(\boldsymbol{A}^k) = \sum_{i=1}^{n} \lambda_i^k \tag{1.34}$$

式右的和称为$\boldsymbol{A}$的诸特征值的k次矩。

(2) 关于迹的不等式 [27]

(a) 对一个复矩阵$\boldsymbol{A} \in C^{m\times n}$，有$\mathrm{tr}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}) \geqslant 0$。

(b) 若$\boldsymbol{A}, \boldsymbol{B}$均为$m \times n$矩阵，则

$$\mathrm{tr}[(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B})^2] \leqslant \mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})\mathrm{tr}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}) \quad (\text{Cauchy-Schwartz不等式})$$

$$\mathrm{tr}[(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B})^2] \leqslant \mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})$$

$$\mathrm{tr}[(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B})^2] \leqslant \mathrm{tr}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}})$$

(c) Schur不等式：$\mathrm{tr}(\boldsymbol{A}^2) \leqslant \mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$。

(d) $\mathrm{tr}[(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}}] \leqslant 2[\mathrm{tr}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}) + \mathrm{tr}(\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}})]$。

(e) 若$\boldsymbol{A}$和$\boldsymbol{B}$为$m \times m$对称矩阵，则$\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) \leqslant \frac{1}{2}\mathrm{tr}(\boldsymbol{A}^2 + \boldsymbol{B}^2)$。

三、行列式

1. 定义：一个$n \times n$正方矩阵$\boldsymbol{A}$的行列式记作$\det(\boldsymbol{A})$或$|\boldsymbol{A}|$

$$\det(\boldsymbol{A}) = |\boldsymbol{A}| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \tag{1.35}$$

若$\boldsymbol{A} = \{a\} \in C^{1\times 1}$，则它的行列式由$\det(\boldsymbol{A}) = a$给出。

2. 性质：

(1) 关于行列式的等式关系 [27]

(a) 如果矩阵的两行(或列)互换位置，则行列式数值保持不变，但符号改变。

(b) 若矩阵的某行(或列)是其他行(或列)的线性组合，则$\det(\boldsymbol{A}) = 0$。特别地，若某行(或列)与另一行(或列) 成正比或相等，或者某行(或列)的元素均等于零，则$\det(\boldsymbol{A}) = 0$。

(c) 任何一个正方矩阵$\boldsymbol{A}$和它的转置矩阵$\boldsymbol{A}^{\mathrm{T}}$具有相同的行列式，即

$$\det(\boldsymbol{A}) = \det(\boldsymbol{A}^{\mathrm{T}}) \tag{1.36}$$

但$\det(\boldsymbol{A}^{\mathrm{H}}) = [\det(\boldsymbol{A}^{\mathrm{T}})]^*$。

(d) 单位矩阵的行列式等于1，即$\det(\boldsymbol{I}) = 1$。

(e) 一个Hermitian矩阵的行列式为实数，因为

$$\det(\boldsymbol{A}) = \det(\boldsymbol{A}^{\mathrm{H}}) = \det(\boldsymbol{A}^{\mathrm{T}}) \Rightarrow \det(\boldsymbol{A}) = \det(\boldsymbol{A}^*) = [\det(\boldsymbol{A})]^* \tag{1.37}$$

(f) 两个矩阵乘积的行列式等于它们的行列式的乘积，即

$$\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B}), \qquad \boldsymbol{A},\boldsymbol{B} \in C^{n\times n} \tag{1.38}$$

(g) 对于一个三角(上三角或下三角)矩阵$\boldsymbol{A}$，其行列式等于三角矩阵主对角线所有元素的乘积，即

$$\det(\boldsymbol{A}) = \prod_{i=1}^{n} a_{ii}$$

一个对角矩阵$\boldsymbol{A} = \mathrm{diag}(a_{11}, a_{22}, \cdots, a_{nn})$的行列式也等于其对角元素的乘积。

(h) 给定一个任意的常数(可以是复数) c，则

$$\det(c\boldsymbol{A}) = c^n \det(\boldsymbol{A}) \tag{1.39}$$

(i) 若$\boldsymbol{A}$非奇异，则$\det(\boldsymbol{A}^{-1}) = [\det(\boldsymbol{A})]^{-1}$。

(j) 对于矩阵$\boldsymbol{A}_{m\times m}, \boldsymbol{B}_{m\times n}, \boldsymbol{C}_{n\times m}, \boldsymbol{D}_{n\times n}$，分块矩阵的行列式满足

$$\boldsymbol{A}\text{非奇异} \iff \det\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \det(\boldsymbol{A})\det(\boldsymbol{D} - \boldsymbol{CA}^{-1}\boldsymbol{B}) \tag{1.40}$$

或

$$\boldsymbol{D}\text{非奇异} \iff \det\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \det(\boldsymbol{D})\det(\boldsymbol{A} - \boldsymbol{BD}^{-1}\boldsymbol{C}) \tag{1.41}$$

(2) 关于行列式的不等式关系[27]

(a) Cauchy-Schwartz不等式：若$\boldsymbol{A},\boldsymbol{B}$都是$m\times n$矩阵，则

$$|\det(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{B})|^2 \leqslant \det(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})\det(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})$$

(b) Hadamard不等式：对于$m\times m$矩阵$\boldsymbol{A}$，有

$$\det(\boldsymbol{A}) \leqslant \prod_{i=1}^{m}\left(\sum_{j=1}^{m}|a_{ij}|^2\right)^{1/2}$$

(c) Fischer不等式：若$\boldsymbol{A}_{m\times m}, \boldsymbol{B}_{m\times n}, \boldsymbol{C}_{n\times n}$，则

$$\det\left(\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^{\mathrm{H}} & \boldsymbol{C} \end{bmatrix}\right) \leqslant \det(\boldsymbol{A})\det(\boldsymbol{C})$$

(d) Minkowski不等式：若$\boldsymbol{A}_{m\times m} \neq \boldsymbol{O}_{m\times m}, \boldsymbol{B}_{m\times m} \neq \boldsymbol{O}_{m\times m}$ 半正定，则

$$\sqrt[m]{\det(\boldsymbol{A}+\boldsymbol{B})} \geqslant \sqrt[m]{\det(\boldsymbol{A})} + \sqrt[m]{\det(\boldsymbol{B})}$$

(e) 正定矩阵$\boldsymbol{A}$的行列式大于0，即$\det(\boldsymbol{A})>0$。

(f) 半正定矩阵$\boldsymbol{A}$的行列式大于或者等于0，即$\det(\boldsymbol{A})\geqslant 0$。

(g) 若$m\times m$矩阵$\boldsymbol{A}$半正定，则

$$[\det(\boldsymbol{A})]^{1/m}\leqslant \frac{1}{m}\det(\boldsymbol{A})$$

(h) 若矩阵$\boldsymbol{A}_{m\times m},\boldsymbol{B}_{m\times m}$均半正定，则

$$\det(\boldsymbol{A}+\boldsymbol{B})\geqslant\det(\boldsymbol{A})+\det(\boldsymbol{B})$$

(i) 若$\boldsymbol{A}_{m\times m}$正定，$\boldsymbol{B}_{m\times m}$半正定，则

$$\det(\boldsymbol{A}+\boldsymbol{B})\geqslant\det(\boldsymbol{A})$$

四、矩阵的秩

1. 定义：矩阵$\boldsymbol{A}_{m\times n}$ 的秩定义为该矩阵中线性无关的行和列的数目。

2. 性质：

(1) 秩的性质

(a) 秩是一个正整数。

(b) 秩等于或小于矩阵的行数或列数。

(c) 当$n\times n$矩阵$\boldsymbol{A}$的秩等于n时，则$\boldsymbol{A}$是非奇异矩阵，或称$\boldsymbol{A}$ 满秩(full rank)。

(d) 如果$\mathrm{rank}(\boldsymbol{A}_{m\times n})<\min\{m,n\}$，则称$\boldsymbol{A}$是秩亏缺的(rank deficient)。一个秩亏缺的正方矩阵称为奇异矩阵。

(e) 若$\mathrm{rank}(\boldsymbol{A}_{m\times n})=m\,(<n)$，则称矩阵$\boldsymbol{A}$具有满行秩(full row rank)。

(f) 若$\mathrm{rank}(\boldsymbol{A}_{m\times n})=n\,(<m)$，则称矩阵$\boldsymbol{A}$具有满列秩(full column rank)。

(g) 任何矩阵$\boldsymbol{A}$左乘满列秩矩阵或者右乘满行秩矩阵后，矩阵$\boldsymbol{A}$的秩保持不变。

(h) 当矩阵的秩$\mathrm{rank}(\boldsymbol{A}_{m\times n})=r\neq 0$时，至少存在一个$r\times r$子矩阵$\boldsymbol{X}_{r\times r}$满秩或非奇异。即是说，矩阵$\boldsymbol{A}_{m\times n}$可以分块为

$$\boldsymbol{A}_{m\times n}=\begin{bmatrix}\boldsymbol{X}_{r\times r} & \boldsymbol{Y}_{r\times(n-r)}\\ \boldsymbol{Z}_{(m-r)\times r} & \boldsymbol{W}_{(m-r)\times(n-r)}\end{bmatrix}$$

式中，$\boldsymbol{X}_{r\times r}$非奇异。

(2) 关于秩的等式

(a) 若$\boldsymbol{A}\in C^{m\times n}$，则$\mathrm{rank}(\boldsymbol{A}^{\mathrm{H}})=\mathrm{rank}(\boldsymbol{A}^{\mathrm{T}})=\mathrm{rank}(\boldsymbol{A}^{*})=\mathrm{rank}(\boldsymbol{A})$。

(b) 若$\boldsymbol{A}\in C^{m\times n}$和$c\neq 0$，则$\mathrm{rank}(c\boldsymbol{A})=\mathrm{rank}(\boldsymbol{A})$。

(c) 若$\boldsymbol{A}\in C^{m\times m}$和$\boldsymbol{C}\in C^{n\times n}$均非奇异，则对于任一矩阵$\boldsymbol{B}\in C^{m\times n}$有$\mathrm{rank}(\boldsymbol{AB})=\mathrm{rank}(\boldsymbol{B})=\mathrm{rank}(\boldsymbol{BC})=\mathrm{rank}(\boldsymbol{ABC})$。即是说，矩阵$\boldsymbol{B}$左乘与(或)右乘一个非奇异矩阵后，$\boldsymbol{B}$的秩保持不变。

(d) 若$\boldsymbol{A} \in C^{m\times n}$且$\boldsymbol{B} \in C^{m\times n}$，则$\text{rank}(\boldsymbol{A}) = \text{rank}(\boldsymbol{B})$，当且仅当存在非奇异矩阵$\boldsymbol{X} \in C^{m\times m}$和$\boldsymbol{Y} \in C^{n\times n}$ 使得$\boldsymbol{B} = \boldsymbol{XAY}$。

(e) 若$\boldsymbol{A} \in C^{m\times n}$，则

$$\text{rank}(\boldsymbol{AA}^{\text{T}}) = \text{rank}(\boldsymbol{A}^{\text{T}}\boldsymbol{A}) = \text{rank}(\boldsymbol{A})$$
$$\text{rank}(\boldsymbol{AA}^{\text{H}}) = \text{rank}(\boldsymbol{A}^{\text{H}}\boldsymbol{A}) = \text{rank}(\boldsymbol{A})$$

(f) 若$\boldsymbol{A} \in C^{m\times m}$，则

$$\text{rank}(\boldsymbol{A}) = m \iff \det(\boldsymbol{A}) \neq 0 \iff \boldsymbol{A}\text{非奇异}$$

(g) 若$m \times m$矩阵$\boldsymbol{A}$非奇异，且$\boldsymbol{B} \in C^{m\times n}, \boldsymbol{C} \in C^{n\times m}, \boldsymbol{D} \in C^{n\times n}$，则

$$\text{rank}\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = m \iff \boldsymbol{D} = \boldsymbol{CA}^{-1}\boldsymbol{B}$$

(3) 关于秩的不等式

(a) 对于任意$m \times n$矩阵$\boldsymbol{A}$均有$\text{rank}(\boldsymbol{A}) \leqslant \min\{m, n\}$。

(b) 若$\boldsymbol{A}, \boldsymbol{B} \in C^{m\times n}$，则$\text{rank}(\boldsymbol{A} + \boldsymbol{B}) \leqslant \text{rank}(\boldsymbol{A}) + \text{rank}(\boldsymbol{B})$。

(c) 若$\boldsymbol{A} \in C^{m\times k}$和$\boldsymbol{B} \in C^{k\times n}$，则

$$\text{rank}(\boldsymbol{A}) + \text{rank}(\boldsymbol{B}) - k \leqslant \text{rank}(\boldsymbol{AB}) \leqslant \min\{\text{rank}(\boldsymbol{A}), \text{rank}(\boldsymbol{B})\}$$

(d) 如果从任意矩阵中删去某些行与(或)某些列，则所得子矩阵的秩不可能大于原矩阵的秩。

1.1.7 逆矩阵

一、逆矩阵的性质

$n \times n$ 矩阵 $\boldsymbol{A}$ 的逆矩阵 $\boldsymbol{A}^{-1}$ 具有以下性质：

(1) $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{AA}^{-1} = \boldsymbol{I}$。

(2) $\boldsymbol{A}^{-1}$ 是唯一的。

(3) 逆矩阵的行列式等于原矩阵行列式的倒数，即 $|\boldsymbol{A}^{-1}| = \dfrac{1}{|\boldsymbol{A}|}$。

(4) 逆矩阵是非奇异的。

(5) $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$。

(6) 复共轭转置矩阵 $\boldsymbol{A}^{\text{H}}$ 的逆矩阵等于逆矩阵 $\boldsymbol{A}^{-1}$ 的复共轭转置，即 $(\boldsymbol{A}^{\text{H}})^{-1} = (\boldsymbol{A}^{-1})^{\text{H}}$。逆矩阵的复共轭转置常采用符号 $\boldsymbol{A}^{-\text{H}} = (\boldsymbol{A}^{-1})^{\text{H}}$ 简记之。

(7) 若 $\boldsymbol{A}^{\text{H}} = \boldsymbol{A}$, 则 $(\boldsymbol{A}^{-1})^{\text{H}} = \boldsymbol{A}^{-1}$。

(8) $(\boldsymbol{A}^*)^{-1} = (\boldsymbol{A}^{-1})^*$。

(9) 如果 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 都是可逆的，则

$$(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1} \tag{1.42}$$

更一般地，有

$$(\boldsymbol{ABC})^{-1}=\boldsymbol{C}^{-1}\boldsymbol{B}^{-1}\boldsymbol{A}^{-1} \tag{1.43}$$

(10) 若 $\boldsymbol{A}=\mathrm{diag}(a_1,a_2,\cdots,a_m)$ 为对角矩阵，则其逆矩阵

$$\boldsymbol{A}^{-1}=\mathrm{diag}(a_1^{-1},a_2^{-1},\cdots,a_m^{-1})$$

(11) 若 $\boldsymbol{A}$ 非奇异，则

$$\boldsymbol{A}\text{ 为正交矩阵} \iff \boldsymbol{A}^{-1}=\boldsymbol{A}^{\mathrm{T}}$$
$$\boldsymbol{A}\text{ 为酉矩阵} \iff \boldsymbol{A}^{-1}=\boldsymbol{A}^{\mathrm{H}}$$

二、矩阵求逆引理

1. Sherman-Morrison 公式： 令 $\boldsymbol{A}$ 是一个 $n\times n$ 的可逆矩阵，并且 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 是两个 $n\times 1$ 向量，使得 $(\boldsymbol{A}+\boldsymbol{x}\boldsymbol{y}^{\mathrm{H}})$ 可逆，则

$$(\boldsymbol{A}+\boldsymbol{x}\boldsymbol{y}^{\mathrm{H}})^{-1}=\boldsymbol{A}^{-1}-\frac{\boldsymbol{A}^{-1}\boldsymbol{x}\boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}^{-1}}{1+\boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}^{-1}\boldsymbol{x}} \tag{1.44}$$

2. Woodbury 公式：

$$\begin{aligned}(\boldsymbol{A}+\boldsymbol{UBV})^{-1}&=\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{UB}(\boldsymbol{B}+\boldsymbol{BVA}^{-1}\boldsymbol{UB})^{-1}\boldsymbol{BVA}^{-1}\\&=\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{I}+\boldsymbol{BVA}^{-1}\boldsymbol{U})^{-1}\boldsymbol{BVA}^{-1}\end{aligned} \tag{1.45}$$

或者

$$(\boldsymbol{A}-\boldsymbol{UV})^{-1}=\boldsymbol{A}^{-1}+\boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{I}-\boldsymbol{VA}^{-1}\boldsymbol{U})^{-1}\boldsymbol{VA}^{-1} \tag{1.46}$$

3. Duncan-Guttman 公式：

$$(\boldsymbol{A}-\boldsymbol{UD}^{-1}\boldsymbol{V})^{-1}=\boldsymbol{A}^{-1}+\boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{D}-\boldsymbol{VA}^{-1}\boldsymbol{U})^{-1}\boldsymbol{VA}^{-1} \tag{1.47}$$

分块矩阵的几种求逆公式：

(1) 矩阵 $\boldsymbol{A}$ 可逆时，为

$$\begin{bmatrix}\boldsymbol{A} & \boldsymbol{U}\\ \boldsymbol{V} & \boldsymbol{D}\end{bmatrix}^{-1}=\begin{bmatrix}\boldsymbol{A}^{-1}+\boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{D}-\boldsymbol{VA}^{-1}\boldsymbol{U})^{-1}\boldsymbol{VA}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{D}-\boldsymbol{VA}^{-1}\boldsymbol{U})^{-1}\\ -(\boldsymbol{D}-\boldsymbol{VA}^{-1}\boldsymbol{U})^{-1}\boldsymbol{VA}^{-1} & (\boldsymbol{D}-\boldsymbol{VA}^{-1}\boldsymbol{U})^{-1}\end{bmatrix}$$

(2) 矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{D}$ 可逆时，为

$$\begin{bmatrix}\boldsymbol{A} & \boldsymbol{U}\\ \boldsymbol{V} & \boldsymbol{D}\end{bmatrix}^{-1}=\begin{bmatrix}(\boldsymbol{A}-\boldsymbol{UD}^{-1}\boldsymbol{V})^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{D}-\boldsymbol{VA}^{-1}\boldsymbol{U})^{-1}\\ -\boldsymbol{D}^{-1}\boldsymbol{V}(\boldsymbol{A}-\boldsymbol{UD}^{-1}\boldsymbol{V})^{-1} & (\boldsymbol{D}-\boldsymbol{VA}^{-1}\boldsymbol{U})^{-1}\end{bmatrix}$$

(3) 矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{D}$ 可逆时，为

$$\begin{bmatrix}\boldsymbol{A} & \boldsymbol{U}\\ \boldsymbol{V} & \boldsymbol{D}\end{bmatrix}^{-1}=\begin{bmatrix}(\boldsymbol{A}-\boldsymbol{UD}^{-1}\boldsymbol{V})^{-1} & -(\boldsymbol{A}-\boldsymbol{UD}^{-1}\boldsymbol{V})^{-1}\boldsymbol{UD}^{-1}\\ -(\boldsymbol{D}-\boldsymbol{VA}^{-1}\boldsymbol{U})^{-1}\boldsymbol{VA}^{-1} & (\boldsymbol{D}-\boldsymbol{VA}^{-1}\boldsymbol{U})^{-1}\end{bmatrix} \tag{1.48}$$

或者

$$\begin{bmatrix}\boldsymbol{A} & \boldsymbol{U}\\ \boldsymbol{V} & \boldsymbol{D}\end{bmatrix}^{-1}=\begin{bmatrix}(\boldsymbol{A}-\boldsymbol{UD}^{-1}\boldsymbol{V})^{-1} & -(\boldsymbol{V}-\boldsymbol{DU}^{-1}\boldsymbol{A})^{-1}\\ (\boldsymbol{U}-\boldsymbol{AV}^{-1}\boldsymbol{D})^{-1} & (\boldsymbol{D}-\boldsymbol{VA}^{-1}\boldsymbol{U})^{-1}\end{bmatrix} \tag{1.49}$$

1.1.8 广义逆矩阵

一、左逆矩阵与右逆矩阵

满足 $\boldsymbol{LA}=\boldsymbol{I}$，但不满足 $\boldsymbol{AL}=\boldsymbol{I}$ 的矩阵 $\boldsymbol{L}$ 称为矩阵 $\boldsymbol{A}$ 的左逆矩阵 (left inverse)。类似地，满足 $\boldsymbol{AR}=\boldsymbol{I}$，但不满足 $\boldsymbol{RA}=\boldsymbol{I}$ 的矩阵称为矩阵 $\boldsymbol{A}$ 的右逆矩阵 (right inverse)。仅当 $m\geqslant n$ 时，矩阵 $\boldsymbol{A}\in C^{m\times n}$ 可能有左逆矩阵。反之，仅当 $m\leqslant n$ 时，矩阵 $\boldsymbol{A}\in C^{m\times n}$ 可能有右逆矩阵。

当 $m>n$ 并且 $\boldsymbol{A}$ 具有满列秩 (rank $\boldsymbol{A}=n$) 时，有左伪逆矩阵：

$$\boldsymbol{L}=(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{H}} \tag{1.50}$$

当 $m<n$ 并且 $\boldsymbol{A}$ 具有满行秩 (rank $\boldsymbol{A}=m$) 时，有右伪逆矩阵：

$$\boldsymbol{R}=\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{-1} \tag{1.51}$$

二、广义逆矩阵的定义与性质

定义 1.1 令 $\boldsymbol{A}$ 是一个 $m\times n$ 矩阵，具有任意秩。矩阵 $\boldsymbol{A}$ 的广义逆矩阵是一个 $n\times m$ 矩阵 $\boldsymbol{G}$，并使得当 $\boldsymbol{Ax}=\boldsymbol{y}$ 为一致方程时，$\boldsymbol{x}=\boldsymbol{Gy}$ 是线性方程 $\boldsymbol{Ax}=\boldsymbol{y}$ 的解。

定理 1.2 一致方程 $\boldsymbol{Ax}=\boldsymbol{y}$ 对 $\boldsymbol{y}\neq\boldsymbol{0}$ 有解 $\boldsymbol{x}=\boldsymbol{Gy}$，当且仅当 $\boldsymbol{AGA}=\boldsymbol{A}$。

命题 1.1 方程 $\boldsymbol{Ax}=\boldsymbol{0}$ 的解与矩阵 $\boldsymbol{A}$ 的任意行正交，并且线性无关。

定义 1.2 $m\times n$ 矩阵 $\boldsymbol{A}$ 的广义逆矩阵是一个满足

$$\boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{A}=\boldsymbol{A}$$

的 $n\times m$ 矩阵 $\boldsymbol{A}^{-}$。

定义 1.3 $m\times n$ 矩阵 $\boldsymbol{A}$ 的广义逆矩阵是满足下列条件之一的 $n\times m$ 矩阵 $\boldsymbol{A}^{-}$：
(1) $\boldsymbol{A}^{-}\boldsymbol{A}$ 为幂等矩阵，且 $\mathrm{rank}(\boldsymbol{A}^{-}\boldsymbol{A})=\mathrm{rank}(\boldsymbol{A})$；
(2) $\boldsymbol{A}\boldsymbol{A}^{-}$ 是幂等矩阵，且 $\mathrm{rank}(\boldsymbol{A}\boldsymbol{A}^{-})=\mathrm{rank}(\boldsymbol{A})$。

广义逆矩阵有以下几个重要的性质：
(1) $\boldsymbol{A}^{-}$ 存在 $\Leftrightarrow \boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{A}=\boldsymbol{A}$。
(2) $\boldsymbol{A}^{-}$ 存在 $\Leftrightarrow \boldsymbol{H}=\boldsymbol{A}^{-}\boldsymbol{A}$ 为幂等矩阵 (即$\boldsymbol{H}^2=\boldsymbol{H}$) 和 $\mathrm{rank}(\boldsymbol{H})=\mathrm{rank}(\boldsymbol{A})$。
(3) $\boldsymbol{A}^{-}$ 存在 $\Leftrightarrow \boldsymbol{F}=\boldsymbol{A}\boldsymbol{A}^{-}$ 为幂等矩阵 (即$\boldsymbol{F}^2=\boldsymbol{F}$) 和 $\mathrm{rank}(\boldsymbol{F})=\mathrm{rank}(\boldsymbol{A})$。

三、广义逆矩阵的计算

秩为 r 的矩阵 $\boldsymbol{A}_{m\times n}$ 的满秩分解算法：

算法 1.1 (矩阵的满秩分解算法)

步骤1 利用行初等变换将矩阵 $\boldsymbol{A}$ 化为阶梯型，即

$$\boldsymbol{A} \xrightarrow{\boldsymbol{E}_1} [\ \ ] \xrightarrow{\boldsymbol{E}_2} \cdots \xrightarrow{\boldsymbol{E}_k} \begin{bmatrix} \boldsymbol{G}_{r\times n} \\ \boldsymbol{O}_{(m-r)\times n} \end{bmatrix}$$

步骤 2 对单位矩阵执行逆行初等变换，得到逆矩阵

$$\boldsymbol{I} \xrightarrow{\boldsymbol{E}_k^{-1}} [\ \ ] \xrightarrow{\boldsymbol{E}_{k-1}^{-1}} \cdots \xrightarrow{\boldsymbol{E}_1^{-1}} [\ \ ] = \boldsymbol{P}^{-1}$$

步骤 3 利用逆矩阵 $\boldsymbol{P}^{-1}$ 的前 r 列构造矩阵 $\boldsymbol{F}$。

步骤 4 书写满秩分解结果 $\boldsymbol{A} = \boldsymbol{F}\boldsymbol{G}$。

算法 1.2 (广义逆矩阵的计算)

步骤 1 计算矩阵 $\boldsymbol{A}_{m\times n}$ 的满秩分解 $\boldsymbol{A} = \boldsymbol{F}\boldsymbol{G}$。

步骤 2 求广义逆矩阵 $\boldsymbol{A}^- = \boldsymbol{G}^{\mathrm{T}}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{G}^{\mathrm{T}})^{-1}\boldsymbol{F}^{\mathrm{T}}$。

四、一致方程的最小范数解

令 $n\times m$ 矩阵 $\boldsymbol{A}^-$ 是 $m\times n$ 矩阵 $\boldsymbol{A}$ 的任意一个广义逆矩阵，则

(1) 齐次方程 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ 的一个通解为 $\boldsymbol{x} = (\boldsymbol{I} - \boldsymbol{A}^-\boldsymbol{A})\boldsymbol{z}$，其中，$\boldsymbol{z}$ 是 $n\times 1$ 任意向量。

(2) 非齐次方程 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ 为一致方程的充分必要条件是

$$\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{y} = \boldsymbol{y}. \tag{1.52}$$

(3) 非齐次方程 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ 的一个通解为

$$\boldsymbol{x} = \boldsymbol{A}^-\boldsymbol{y} + (\boldsymbol{I} - \boldsymbol{A}^-\boldsymbol{A})\boldsymbol{z} \tag{1.53}$$

式中，$\boldsymbol{z}$ 为 $n\times 1$ 任意向量。

矩阵 $\boldsymbol{A}_{m\times n}$ 的伴随矩阵用符号 $\boldsymbol{A}^{\#}_{n\times m}$ 表示，定义为将 m 阶向量空间的内积等价变换为 n 阶向量的内积的一个映射，即有

$$\langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{y}\rangle_m = \langle \boldsymbol{x}, \boldsymbol{A}^{\#}\boldsymbol{y}\rangle_n \tag{1.54}$$

定理 1.3 $\boldsymbol{G}\boldsymbol{y}$ 是一致方程 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ 的最小范数解，当且仅当

$$\boldsymbol{A}\boldsymbol{G}\boldsymbol{A} = \boldsymbol{A}, \qquad (\boldsymbol{G}\boldsymbol{A})^{\#} = \boldsymbol{G}\boldsymbol{A} \tag{1.55}$$

当 $\boldsymbol{A}_{m\times n}$ 具有满行秩 m 时，线性方程 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ 的最小范数解为 $\boldsymbol{x}^{\mathrm{o}} = \boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{-1}\boldsymbol{y}$。

五、非一致方程的最小二乘解

令 $\boldsymbol{G}$ 是某个矩阵，则 $\hat{\boldsymbol{x}} = \boldsymbol{G}\boldsymbol{y}$ 是非一致方程 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ 的最小二乘解，当且仅当

$$\boldsymbol{A}^{\#}\boldsymbol{A}\boldsymbol{G} = \boldsymbol{A}^{\#} \tag{1.56}$$

或者等价为

$$AGA = A,\quad (AG)^{\#} = AG \tag{1.57}$$

非一致方程的最小二乘解有可能不是唯一的，但是不同的最小二乘解得到的 Ax 和 $Ax - y$ 是唯一的。非一致方程 $Ax = y$ 的最小二乘解的通解形式为

$$\hat{x} = Gy + (I - GA)z,\quad z \text{ 任意} \tag{1.58}$$

当非一致方程 $Ax = y$ 的矩阵 A 具有满列秩时，$x^{\circ} = (A^{\mathrm{H}}A)^{-1}A^{\mathrm{H}}y$ 是一个最小二乘解。

1.1.9 Moore-Penrose 逆矩阵

一、Moore-Penrose 逆矩阵的定义

令 A 是任意 $m \times n$ 矩阵，称矩阵 G 是 A 的广义逆矩阵，若 G 满足以下四个条件(常称Moore-Penrose条件)：

(1) $AGA = A$;

(2) $GAG = G$;

(3) AG 为 Hermitian 矩阵，即 $(AG)^{\mathrm{H}} = AG$;

(4) GA 为 Hermitian 矩阵，即 $(GA)^{\mathrm{H}} = GA$。

只满足条件 (1) 和 (2) 的矩阵 $G = A^{\dagger}$ 称为 A 的自反广义逆矩阵 (reflexive generalized inverse)。注意，对于只满足某些条件的广义逆矩阵，它的秩与原矩阵的秩不一定相等。另外，特别需要注意的是，与 $(AB)^{-1} = B^{-1}A^{-1}$ 不同，一般情况下，矩阵乘积 AB 的 Moore-Penrose 逆矩阵 $(AB)^{\dagger} \neq B^{\dagger}A^{\dagger}$。

二、Moore-Penrose 逆矩阵的计算

1. 方程求解法

算法 1.3

步骤1 计算矩阵 $B = AA^{\mathrm{H}}$。

步骤2 求解矩阵方程 $B^2X^{\mathrm{H}} = B$ 得到矩阵 X^{H}。

步骤3 计算 B 的 Moore-Penrose 逆矩阵 $B^{\dagger} = (AA^{\mathrm{H}})^{\dagger} = XBX^{\mathrm{H}}$。

步骤4 计算矩阵 A 的 Moore-Penrose 逆矩阵 $A^{\dagger} = A^{\mathrm{H}}(AA^{\mathrm{H}})^{\dagger} = A^{\mathrm{H}}B^{\dagger}$。

算法 1.4

步骤1 计算矩阵 $B = A^{\mathrm{H}}A$。

步骤2 求解矩阵方程 $B^2X^{\mathrm{H}} = B$ 得到矩阵 X^{H}。

步骤3 计算 B 的 Moore-Penrose 逆矩阵 $B^{\dagger} = (A^{\mathrm{H}}A)^{\dagger} = XBX^{\mathrm{H}}$。

步骤4 计算矩阵 A 的 Moore-Penrose 逆矩阵 $A^{\dagger} = (A^{\mathrm{H}}A)^{\dagger}A^{\mathrm{H}} = B^{\dagger}A^{\mathrm{H}}$。

若矩阵 $\boldsymbol{A}_{m\times n}$ 的列数大于行数，则矩阵乘积 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的维数比 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的维数小，故选择算法 1.3 可花费较少的计算量。反之，若 $\boldsymbol{A}$ 的行数大于列数，则选择算法 1.4。

2. KL 分解法

若 $\boldsymbol{A}=\boldsymbol{K}\boldsymbol{L}$ 是矩阵 $\boldsymbol{A}_{m\times n}$ 的满秩分解，则

$$\boldsymbol{G}=\boldsymbol{L}^{\mathrm{H}}(\boldsymbol{K}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{L}^{\mathrm{H}})^{-1}\boldsymbol{K}^{\mathrm{H}} \tag{1.59}$$

是 $\boldsymbol{A}_{m\times n}$ 的 Moore-Penrose 逆矩阵。

3. 递推法

算法 1.5 (求 Moore-Penrose 逆矩阵的列递推算法)

初始值 $\boldsymbol{A}_1^{\dagger}=\boldsymbol{a}_1^{\dagger}=(\boldsymbol{a}_1^{\mathrm{H}}\boldsymbol{a}_1)^{-1}\boldsymbol{a}_1^{\mathrm{H}}$。

递推　令 $k=2,3,\cdots,n$，进行以下计算：

$$\begin{aligned}
\boldsymbol{d}_k&=\boldsymbol{A}_{k-1}^{\dagger}\boldsymbol{a}_k\\
\boldsymbol{b}_k&=\begin{cases}(1+\boldsymbol{d}_k^{\mathrm{H}}\boldsymbol{d}_k)^{-1}\boldsymbol{d}_k^{\mathrm{H}}\boldsymbol{A}_{k-1}^{\dagger}, & \boldsymbol{a}_k-\boldsymbol{A}_{k-1}\boldsymbol{d}_k=\boldsymbol{0}\\ (\boldsymbol{a}_k-\boldsymbol{A}_{k-1}\boldsymbol{d}_k)^{\dagger}, & \boldsymbol{a}_k-\boldsymbol{A}_{k-1}\boldsymbol{d}_k\neq\boldsymbol{0}\end{cases}\\
\boldsymbol{A}_k^{\dagger}&=\begin{bmatrix}\boldsymbol{A}_{k-1}^{\dagger}-\boldsymbol{d}_k\boldsymbol{b}_k\\ \boldsymbol{b}_k\end{bmatrix}
\end{aligned}$$

4. 迹方法

已知矩阵 $\boldsymbol{A}_{m\times n}$ 的秩为 r。

算法 1.6 (求 Moore-Penrose 逆矩阵的迹方法) [13]

步骤1　计算 $\boldsymbol{B}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$。

步骤2　令 $\boldsymbol{C}_1=\boldsymbol{I}$。

步骤3　计算

$$\boldsymbol{C}_{i+1}=\frac{1}{i}\mathrm{tr}(\boldsymbol{C}_i\boldsymbol{B})\boldsymbol{I}-\boldsymbol{C}_i\boldsymbol{B},\quad i=1,2,\cdots,r-1$$

步骤4　计算

$$\boldsymbol{A}^{\dagger}=\frac{r}{\mathrm{tr}(\boldsymbol{C}_r\boldsymbol{B})}\boldsymbol{C}_r\boldsymbol{A}^{\mathrm{T}}$$

1.1.10 Hadamard 积与 Kronecker 积

一、矩阵的直和

$m\times m$ 矩阵 $\boldsymbol{A}$ 与 $n\times n$ 矩阵 $\boldsymbol{B}$ 的直和记作 $\boldsymbol{A}\oplus\boldsymbol{B}$，它是一个 $(m+n)\times(m+n)$ 矩阵，定义为

$$\boldsymbol{A}\oplus\boldsymbol{B}=\begin{bmatrix}\boldsymbol{A} & \boldsymbol{O}_{m\times n}\\ \boldsymbol{O}_{n\times m} & \boldsymbol{B}\end{bmatrix} \tag{1.60}$$

二、Hadamard 积

$m\times n$ 矩阵 $\boldsymbol{A}=[a_{ij}]$ 与 $m\times n$ 矩阵 $\boldsymbol{B}=[b_{ij}]$ 的 Hadamard 积记作 $\boldsymbol{A}\odot\boldsymbol{B}$，它仍然是一个 $m\times n$ 矩阵，定义为

$$\boldsymbol{A}\odot\boldsymbol{B}=[a_{ij}b_{ij}] \tag{1.61}$$

定理 1.4 若 $m\times m$ 矩阵 $\boldsymbol{A},\boldsymbol{B}$ 是正定 (或半正定) 的，则它们的 Hadamard 积 $\boldsymbol{A}\odot\boldsymbol{B}$ 也是正定 (或半正定) 的。

定理 1.5 令 $\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}$ 为 $m\times n$ 矩阵，并且 $\boldsymbol{1}=[1,1,\cdots,1]^{\mathrm{T}}$ 为 $n\times 1$ 求和向量，$\boldsymbol{D}=\mathrm{diag}(d_1,d_2,\cdots,d_m)$，其中，$d_i=\sum\limits_{j=1}^{n}a_{ij}$，则

$$\mathrm{tr}\left(\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{B}\odot\boldsymbol{C})\right)=\mathrm{tr}\left((\boldsymbol{A}^{\mathrm{T}}\odot\boldsymbol{B}^{\mathrm{T}})\boldsymbol{C}\right) \tag{1.62}$$

和

$$\boldsymbol{1}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{B}\odot\boldsymbol{C})\boldsymbol{1}=\mathrm{tr}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{C}) \tag{1.63}$$

定理 1.6 令 $\boldsymbol{A},\boldsymbol{B}$ 为 $n\times n$ 正方矩阵，并且 $\boldsymbol{1}=[1,1,\cdots,1]^{\mathrm{T}}$ 为 $n\times 1$ 求和向量。假定 $\boldsymbol{M}$ 是一个 $n\times n$ 对角矩阵 $\boldsymbol{M}=\mathrm{diag}(\mu_1,\mu_2,\cdots,\mu_n)$，而 $\boldsymbol{m}=\boldsymbol{M}\boldsymbol{1}$ 为 $n\times 1$ 向量，则有

$$\mathrm{tr}(\boldsymbol{A}\boldsymbol{M}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{M})=\boldsymbol{m}^{\mathrm{T}}(\boldsymbol{A}\odot\boldsymbol{B})\boldsymbol{m} \tag{1.64}$$

$$\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}^{\mathrm{T}})=\boldsymbol{1}^{\mathrm{T}}(\boldsymbol{A}\odot\boldsymbol{B})\boldsymbol{1} \tag{1.65}$$

$$\boldsymbol{M}\boldsymbol{A}\odot\boldsymbol{B}^{\mathrm{T}}\boldsymbol{M}=\boldsymbol{M}(\boldsymbol{A}\odot\boldsymbol{B}^{\mathrm{T}})\boldsymbol{M} \tag{1.66}$$

三、矩阵化函数和向量化函数

矩阵的向量化算子 vec 与迹之间有以下关系：

$$\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B})=(\mathrm{vec}(\boldsymbol{A}))^{\mathrm{T}}\mathrm{vec}(\boldsymbol{B}) \tag{1.67}$$

$m\times n$ 矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的 Hadamard 积的向量化函数为

$$\mathrm{vec}(\boldsymbol{A}\odot\boldsymbol{B})=\mathrm{vec}(\boldsymbol{A})\odot\mathrm{vec}(\boldsymbol{B})$$

$$\mathrm{vec}(\boldsymbol{A}\odot\boldsymbol{B})=\mathrm{diag}(\mathrm{vec}(\boldsymbol{A}))\mathrm{vec}(\boldsymbol{B})=\mathrm{diag}(\mathrm{vec}(\boldsymbol{B}))\mathrm{vec}(\boldsymbol{A})$$

四、Kronecker 积

$m\times n$ 矩阵 $\boldsymbol{A}$ 和 $p\times q$ 矩阵 $\boldsymbol{B}$ 的右 Kronecker 积 $\boldsymbol{A}\otimes\boldsymbol{B}$ 定义为

$$[\boldsymbol{A}\otimes\boldsymbol{B}]_{\mathrm{right}}=[a_{ij}\boldsymbol{B}]=\begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2n}\boldsymbol{B} \\ \vdots & \vdots & & \vdots \\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix} \tag{1.68}$$

$m \times n$ 矩阵 $\boldsymbol{A}$ 和 $p \times q$ 矩阵$\boldsymbol{B}$ 的左 Kronecker 积 $\boldsymbol{A} \otimes \boldsymbol{B}$ 定义为

$$[\boldsymbol{A} \otimes \boldsymbol{B}]_{\text{left}} = [\boldsymbol{A}b_{ij}] = \begin{bmatrix} \boldsymbol{A}b_{11} & \boldsymbol{A}b_{12} & \cdots & \boldsymbol{A}b_{1q} \\ \boldsymbol{A}b_{21} & \boldsymbol{A}b_{22} & \cdots & \boldsymbol{A}b_{2q} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{A}b_{p1} & \boldsymbol{A}b_{p2} & \cdots & \boldsymbol{A}b_{pq} \end{bmatrix} \tag{1.69}$$

Kronecker 积具有以下性质：

(1) 若矩阵 $\boldsymbol{A}_{m\times n} = \boldsymbol{ab}^{\mathrm{T}}$，则

$$\text{vec}(\boldsymbol{ab}^{\mathrm{T}}) = \boldsymbol{b} \otimes \boldsymbol{a} \tag{1.70}$$

(2) 令 $\boldsymbol{A}_{m\times p}$，$\boldsymbol{B}_{p\times q}$，$\boldsymbol{C}_{q\times n}$，则

$$\text{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}^{\mathrm{T}} \otimes \boldsymbol{A})\text{vec}(\boldsymbol{B}) \tag{1.71}$$

(3) 若 $\boldsymbol{A}$ 为单位矩阵 $\boldsymbol{I}_m$，而 $\boldsymbol{B} \in R^{m\times q}, \boldsymbol{C} \in R^{q\times n}$，则

$$\text{vec}(\boldsymbol{BC}) = (\boldsymbol{C}^{\mathrm{T}} \otimes \boldsymbol{I}_m)\text{vec}(\boldsymbol{B}) = (\boldsymbol{C}^{\mathrm{T}} \otimes \boldsymbol{B})\text{vec}(\boldsymbol{I}_q) = (\boldsymbol{I}_n \otimes \boldsymbol{B})\text{vec}(\boldsymbol{C}) \tag{1.72}$$

(4) 若 $\boldsymbol{C} = \boldsymbol{d}$ 为 q 向量，则

$$\boldsymbol{ABd} = \text{vec}(\boldsymbol{ABd}) = (\boldsymbol{d}^{\mathrm{T}} \otimes \boldsymbol{A})\text{vec}(\boldsymbol{B}) = (\boldsymbol{A} \otimes \boldsymbol{d}^{\mathrm{T}})\text{vec}(\boldsymbol{B}^{\mathrm{T}}) \tag{1.73}$$

(5) 对于矩阵 $\boldsymbol{A}_{m\times n}, \boldsymbol{B}_{n\times k}, \boldsymbol{C}_{l\times p}, \boldsymbol{D}_{p\times q}$，有

$$\boldsymbol{AB} \otimes \boldsymbol{CD} = (\boldsymbol{A} \otimes \boldsymbol{C})(\boldsymbol{B} \otimes \boldsymbol{D}) \tag{1.74}$$

(6) 若矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 分别有广义逆矩阵 $\boldsymbol{A}^{\dagger}$ 和 $\boldsymbol{B}^{\dagger}$，则

$$(\boldsymbol{A} \otimes \boldsymbol{B})^{\dagger} = \boldsymbol{A}^{\dagger} \otimes \boldsymbol{B}^{\dagger} \tag{1.75}$$

(7) 对于矩阵 $\boldsymbol{A}_{m\times n}, \boldsymbol{B}_{p\times q}$，有

$$\text{rank}(\boldsymbol{A} \otimes \boldsymbol{B}) = \text{rank}(\boldsymbol{A})\text{rank}(\boldsymbol{B}) \tag{1.76}$$

(8) 若 $\boldsymbol{A}$ 是 $m \times m$ 矩阵，$\boldsymbol{B}$ 是 $n \times n$ 矩阵，则

$$\det(\boldsymbol{A} \otimes \boldsymbol{B}) = (\det(\boldsymbol{A}))^n (\det(\boldsymbol{B}))^m \tag{1.77}$$

(9) 若 $\boldsymbol{A}$ 是 $m \times m$ 矩阵，$\boldsymbol{B}$ 是 $n \times n$ 矩阵，则

$$\text{tr}(\boldsymbol{A} \otimes \boldsymbol{B}) = \text{tr}(\boldsymbol{A})\text{tr}(\boldsymbol{B}) \tag{1.78}$$

(10) 对于矩阵 $\boldsymbol{A}_{m\times n}, \boldsymbol{B}_{p\times q}$，有

$$\exp(\boldsymbol{A} \otimes \boldsymbol{B}) = \exp(\boldsymbol{A}) \otimes \exp(\boldsymbol{B}) \tag{1.79}$$

(11) 作为式 (1.74) 的特例，若 $\boldsymbol{B}=\boldsymbol{I}_p$ 和 $\boldsymbol{C}=\boldsymbol{I}_q$，则

$$\boldsymbol{A}\otimes\boldsymbol{D}=(\boldsymbol{A}\boldsymbol{I}_p)\otimes(\boldsymbol{I}_q\boldsymbol{D})=(\boldsymbol{A}\otimes\boldsymbol{I}_q)(\boldsymbol{I}_p\otimes\boldsymbol{D}) \tag{1.80}$$

如果 $\boldsymbol{G}$ 是 $t\times u$ 矩阵，而 $\boldsymbol{F}$ 是 $q\times u$ 矩阵，即 $\boldsymbol{G}$ 和 $\boldsymbol{F}$ 有相同的列数，则这两个矩阵的 Khatri-Rao 积记为 $\boldsymbol{F}*\boldsymbol{G}$，并定义为

$$\boldsymbol{F}*\boldsymbol{G}=[\boldsymbol{f}_1\otimes\boldsymbol{g}_1,\boldsymbol{f}_2\otimes\boldsymbol{g}_2,\cdots,\boldsymbol{f}_u\otimes\boldsymbol{g}_u] \tag{1.81}$$

下面是向量化函数、Kronecker 乘幂和 Khatri-Rao 积的性质：

(1) 矩阵之和的向量化

$$\mathrm{vec}(\boldsymbol{A}+\boldsymbol{B})=\mathrm{vec}(\boldsymbol{A})+\mathrm{vec}(\boldsymbol{B}) \tag{1.82}$$

(2) 转置矩阵的向量化

$$\mathrm{vec}(\boldsymbol{A}^{\mathrm{T}})=\boldsymbol{K}_{pq}\mathrm{vec}(\boldsymbol{A}) \tag{1.83}$$

(3) 两个矩阵 $\boldsymbol{A}_{m\times n}$、$\boldsymbol{B}_{n\times p}$ 的向量化

$$\begin{aligned}\mathrm{vec}(\boldsymbol{A}\boldsymbol{B})&=(\boldsymbol{I}_s\otimes\boldsymbol{A})\mathrm{vec}(\boldsymbol{B})=(\boldsymbol{B}^{\mathrm{T}}\otimes\boldsymbol{I}_p)\mathrm{vec}(\boldsymbol{A})\\&=(\boldsymbol{B}^{\mathrm{T}}\otimes\boldsymbol{A})\mathrm{vec}(\boldsymbol{I}_q)\end{aligned} \tag{1.84}$$

(4) 三个矩阵 $\boldsymbol{A}_{m\times n}$、$\boldsymbol{B}_{n\times p}$、$\boldsymbol{C}_{p\times q}$ 的向量化

$$\mathrm{vec}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C})=(\boldsymbol{C}^{\mathrm{T}}\otimes\boldsymbol{A})\mathrm{vec}(\boldsymbol{B}) \tag{1.85}$$

(5) 矩阵的 Kronecker 乘幂

$$\boldsymbol{A}^{[k+1]}=\boldsymbol{A}\otimes\boldsymbol{A}^{[k]} \tag{1.86}$$

(6) 矩阵乘积的 Kronecker 乘幂

$$(\boldsymbol{A}\boldsymbol{B})^{[k]}=\boldsymbol{A}^{[k]}\boldsymbol{B}^{[k]} \tag{1.87}$$

(7) 三个矩阵乘积的迹

$$\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C})=(\mathrm{vec}(\boldsymbol{A}))^{\mathrm{T}}(\boldsymbol{I}_p\otimes\boldsymbol{B})\mathrm{vec}(\boldsymbol{C}) \tag{1.88}$$

(8) 四个矩阵乘积的迹

$$\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}\boldsymbol{D})=(\mathrm{vec}(\boldsymbol{D}^{\mathrm{T}}))^{\mathrm{T}}(\boldsymbol{C}^{\mathrm{T}}\otimes\boldsymbol{A})\mathrm{vec}(\boldsymbol{B}) \tag{1.89}$$

$$=(\mathrm{vec}(\boldsymbol{D}))^{\mathrm{T}}(\boldsymbol{A}\otimes\boldsymbol{C}^{\mathrm{T}})\mathrm{vec}(\boldsymbol{B}^{\mathrm{T}}) \tag{1.90}$$

(9) 矩阵内积的迹等于两个矩阵的向量化函数的内积，即

$$\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{D})=(\mathrm{vec}(\boldsymbol{A}))^{\mathrm{T}}\mathrm{vec}(\boldsymbol{D}) \tag{1.91}$$

(10) Khatri-Rao 积的结合律

$$\boldsymbol{A} * (\boldsymbol{D} * \boldsymbol{F}) = (\boldsymbol{A} * \boldsymbol{D}) * \boldsymbol{F} \tag{1.92}$$

(11) Khatri-Rao 积 $\boldsymbol{A} * \boldsymbol{B}$ 与 $\boldsymbol{B} * \boldsymbol{A}$ 之间的关系

$$\boldsymbol{A} * \boldsymbol{B} = \boldsymbol{K}_{nn}(\boldsymbol{B} * \boldsymbol{A}) \tag{1.93}$$

(12) Kronecker 积与 Khatri-Rao 积的乘积

$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{F} * \boldsymbol{G}) = \boldsymbol{AF} * \boldsymbol{BG} \tag{1.94}$$

式 (1.83) 和式 (1.93) 中的矩阵 $\boldsymbol{K}_{pq}$ 为交换矩阵。

1.2 习题与解答

1.1 证明矩阵加法的结合律$(\boldsymbol{A}+\boldsymbol{B})+\boldsymbol{C}=\boldsymbol{A}+(\boldsymbol{B}+\boldsymbol{C})$和矩阵乘法的右分配律$(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C}=\boldsymbol{AC}+\boldsymbol{BC}$。

思路分析：按照矩阵加法和乘法的基本定义可以直接证明该问题。

证明：设$\boldsymbol{A}_{ij}$表示矩阵$\boldsymbol{A}$的第i行，第j列上的元素，则有

$$
\begin{aligned}
[(\boldsymbol{A}+\boldsymbol{B})+\boldsymbol{C}]_{ij} &= (\boldsymbol{A}_{ij}+\boldsymbol{B}_{ij})+\boldsymbol{C}_{ij}\\
&= \boldsymbol{A}_{ij}+(\boldsymbol{B}_{ij}+\boldsymbol{C}_{ij})\\
&= [\boldsymbol{A}+(\boldsymbol{B}+\boldsymbol{C})]_{ij}\\
[(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C}]_{ij} &= \sum_k(\boldsymbol{A}_{ik}+\boldsymbol{B}_{ik})\boldsymbol{C}_{kj}\\
&= \sum_k\boldsymbol{A}_{ik}\boldsymbol{C}_{kj}+\sum_k\boldsymbol{B}_{ik}\boldsymbol{C}_{kj}\\
&= (\boldsymbol{AC})_{ij}+(\boldsymbol{BC})_{ij}\\
&= (\boldsymbol{AC}+\boldsymbol{BC})_{ij}
\end{aligned}
$$

上述两式就证明了矩阵加法的结合律与矩阵乘法的右分配律。

1.2 令

$$
\boldsymbol{X}=\begin{bmatrix}6&0&0\\-3&4&0\\0&5&1\end{bmatrix}\quad\text{和}\quad\boldsymbol{Y}=\begin{bmatrix}1&2&3\\0&-2&1\\7&0&-1\end{bmatrix}
$$

求$\boldsymbol{X}^2$、$\boldsymbol{Y}^2$、$\boldsymbol{XY}$、$\boldsymbol{YX}$，并证明

$$
(\boldsymbol{X}-\boldsymbol{Y})^2=\boldsymbol{X}^2+\boldsymbol{Y}^2-\boldsymbol{XY}-\boldsymbol{YX}=\begin{bmatrix}52&-37&-19\\-26&37&1\\-64&54&20\end{bmatrix}
$$

思路分析：直接按照矩阵乘法的定义求解和证明即可。

解：由于求解过程比较简单，这里直接给出结果

$$
\boldsymbol{X}^2=\begin{bmatrix}36&0&0\\-30&16&0\\-15&25&1\end{bmatrix},\quad\boldsymbol{Y}^2=\begin{bmatrix}22&-2&2\\7&4&-3\\0&14&22\end{bmatrix}
$$

$$
\boldsymbol{XY}=\begin{bmatrix}6&12&18\\-3&-14&-5\\7&-10&4\end{bmatrix},\quad\boldsymbol{YX}=\begin{bmatrix}0&23&3\\6&-3&1\\42&-5&-1\end{bmatrix}
$$

证明：

$$
\begin{aligned}
(\boldsymbol{X}-\boldsymbol{Y})^2 &= (\boldsymbol{X}-\boldsymbol{Y})(\boldsymbol{X}-\boldsymbol{Y})\\
&= \boldsymbol{X}^2-\boldsymbol{XY}-\boldsymbol{YX}+\boldsymbol{Y}^2
\end{aligned}
$$

将前面得到的4个结果代入上式，即可得到

$$(\boldsymbol{X}-\boldsymbol{Y})^2=\begin{bmatrix} 52 & -37 & -19 \\ -26 & 37 & 1 \\ -64 & 54 & 20 \end{bmatrix}$$

1.3 假定$\boldsymbol{A}$和$\boldsymbol{B}$具有相同的维数，证明

$$(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}}=(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{B}^{\mathrm{T}})=\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}+\boldsymbol{A}\boldsymbol{B}^{\mathrm{T}}+\boldsymbol{B}\boldsymbol{A}^{\mathrm{T}}$$

思路分析：在前面两题的基础上，这里又引入了对矩阵转置定义的考察。转置是矩阵的一种基本运算，其定义可见教材P3。

证明：设$\boldsymbol{A}_{ij}$表示矩阵$\boldsymbol{A}$ 的第i行，第j列上的元素，则根据矩阵转置的定义，$[\boldsymbol{A}^{\mathrm{T}}]_{ij}=\boldsymbol{A}_{ji}$，从而

$$\left[(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}}\right]_{ij}=[\boldsymbol{A}+\boldsymbol{B}]_{ji}=\boldsymbol{A}_{ji}+\boldsymbol{B}_{ji}=[\boldsymbol{A}^{\mathrm{T}}]_{ij}+[\boldsymbol{B}^{\mathrm{T}}]_{ij}$$

这样就可以得到$(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}}=\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{B}^{\mathrm{T}}$。根据上面的结论，可以继续得到

$$(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}}=(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{B}^{\mathrm{T}})=\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}+\boldsymbol{A}\boldsymbol{B}^{\mathrm{T}}+\boldsymbol{B}\boldsymbol{A}^{\mathrm{T}}$$

1.4 令$\boldsymbol{A}=\begin{bmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix}$，计算$\boldsymbol{A}^2$、$\boldsymbol{A}^4$和$\boldsymbol{A}^5$。

解：根据矩阵乘法的定义，可以直接计算得到结果，详细的计算过程不再给出。

$$\boldsymbol{A}^2=\begin{bmatrix} 0.28 & 0.72 \\ 0.24 & 0.76 \end{bmatrix}$$

$$\boldsymbol{A}^4=\begin{bmatrix} 0.2512 & 0.7488 \\ 0.2496 & 0.7504 \end{bmatrix}$$

$$\boldsymbol{A}^5=\begin{bmatrix} 0.2502 & 0.7498 \\ 0.2499 & 0.7501 \end{bmatrix}$$

1.5 已知线性方程组

$$\begin{cases} 2y_1-y_2=x_1 \\ y_1+2y_2=x_2 \\ -2y_1+3y_2=x_3 \end{cases}, \qquad \begin{cases} 3z_1-z_2=y_1 \\ 5z_1+2z_2=y_2 \end{cases}$$

用z_1,z_2表示x_1,x_2,x_3。

思路分析：两个线性方程组实际上给出了两组线性变换，将二者串接在一起，即可得到z_1,z_2与x_1,x_2,x_3之间的线性关系。

解：根据第一个方程组可得

$$\begin{bmatrix} 2 & -1 \\ 1 & 2 \\ -2 & 3 \end{bmatrix}\boldsymbol{y}=\boldsymbol{x}$$

其中$\boldsymbol{y}=[y_1,y_2]^{\mathrm{T}}$，$\boldsymbol{x}=[x_1,x_2,x_3]^{\mathrm{T}}$。

同理，根据第二个方程组可得

$$\begin{bmatrix} 3 & -1 \\ 5 & 2 \end{bmatrix} \boldsymbol{z} = \boldsymbol{y}$$

其中，$\boldsymbol{z}=[z_1,z_2]^{\mathrm{T}}$。

将第二式代入到第一式中，得到

$$\begin{bmatrix} 2 & -1 \\ 1 & 2 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 5 & 2 \end{bmatrix} \boldsymbol{z} = \boldsymbol{x}$$

即

$$\begin{bmatrix} 1 & -4 \\ 13 & 3 \\ 9 & 8 \end{bmatrix} \boldsymbol{z} = \boldsymbol{x}$$

上式即给出了z_1,z_2与x_1,x_2,x_3之间的线性关系。

1.6 利用初等行变换，将矩阵

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 2 & 8 & 4 \\ 0 & 0 & 0 & 1 & 4 & 9 & 7 \\ 0 & 3 & -11 & -3 & -8 & -15 & -32 \\ 0 & -2 & -8 & 1 & 6 & 13 & 21 \end{bmatrix}$$

化简为简约阶梯型矩阵。

思路分析：矩阵的初等行运算可以解决矩阵方程求解、矩阵求逆和矩阵的基本空间的基向量构造等复杂问题，是一种重要的工具。而一个阶梯型矩阵称为简约阶梯型，要求其每一非零行的首项元素等于1，并且每一个首一元素也是它所在的列唯一的非零元素。教材P9中的算法1.1.1给出了将一个矩阵化为简约阶梯型的步骤，下面按照此步骤求解。

解：用$\boldsymbol{r}_i$表示上一步化简后矩阵的第i行。

(a) $\boldsymbol{r}_1 \leftrightarrow \boldsymbol{r}_4$，$\boldsymbol{r}_2 \leftrightarrow \boldsymbol{r}_3$得到

$$\boldsymbol{A} \sim \begin{bmatrix} 0 & -2 & -8 & 1 & 6 & 13 & 21 \\ 0 & 3 & -11 & -3 & -8 & -15 & -32 \\ 0 & 0 & 0 & 1 & 4 & 9 & 7 \\ 0 & 0 & 0 & 0 & 2 & 8 & 4 \end{bmatrix}$$

(b) $\frac{1}{2} \times \boldsymbol{r}_4 \to \boldsymbol{r}_4$，$\frac{1}{3} \times \boldsymbol{r}_2 \to \boldsymbol{r}_2$，$-\frac{1}{2} \times \boldsymbol{r}_1 \to \boldsymbol{r}_1$得到

$$\boldsymbol{A} \sim \begin{bmatrix} 0 & 1 & 4 & -\frac{1}{2} & -3 & -\frac{13}{2} & -\frac{21}{2} \\ 0 & 1 & -\frac{11}{3} & -1 & -\frac{8}{3} & -5 & -\frac{32}{3} \\ 0 & 0 & 0 & 1 & 4 & 9 & 7 \\ 0 & 0 & 0 & 0 & 1 & 4 & 2 \end{bmatrix}$$

(c) $\boldsymbol{r}_3-4\boldsymbol{r}_4\to\boldsymbol{r}_3$得到

$$A\sim\begin{bmatrix}0&1&4&-\frac{1}{2}&-3&-\frac{13}{2}&-\frac{21}{2}\\0&1&-\frac{11}{3}&-1&-\frac{8}{3}&-5&-\frac{32}{3}\\0&0&0&1&0&-7&-1\\0&0&0&0&1&4&2\end{bmatrix}$$

(d) $\boldsymbol{r}_2-\boldsymbol{r}_1\to\boldsymbol{r}_2$得到

$$A\sim\begin{bmatrix}0&1&4&-\frac{1}{2}&-3&-\frac{13}{2}&-\frac{21}{2}\\0&0&-\frac{23}{3}&-\frac{1}{2}&\frac{1}{3}&\frac{3}{2}&-\frac{1}{6}\\0&0&0&1&0&-7&-1\\0&0&0&0&1&4&2\end{bmatrix}$$

(e) $\boldsymbol{r}_1+\frac{1}{2}\boldsymbol{r}_3\to\boldsymbol{r}_1$，$\boldsymbol{r}_2+\frac{1}{2}\boldsymbol{r}_3\to\boldsymbol{r}_2$得到

$$A\sim\begin{bmatrix}0&1&4&0&-3&-10&-11\\0&0&-\frac{23}{3}&0&\frac{1}{3}&-2&-\frac{2}{3}\\0&0&0&1&0&-7&-1\\0&0&0&0&1&4&2\end{bmatrix}$$

(f) $\boldsymbol{r}_1+3\boldsymbol{r}_4\to\boldsymbol{r}_1$，$\boldsymbol{r}_2-\frac{1}{3}\boldsymbol{r}_4\to\boldsymbol{r}_2$得到

$$A\sim\begin{bmatrix}0&1&4&0&0&2&-5\\0&0&-\frac{23}{3}&0&0&-\frac{10}{3}&-\frac{4}{3}\\0&0&0&1&0&-7&-1\\0&0&0&0&1&4&2\end{bmatrix}$$

(g) $-\frac{3}{23}\boldsymbol{r}_2\to\boldsymbol{r}_2$得到

$$A\sim\begin{bmatrix}0&1&4&0&0&2&-5\\0&0&1&0&0&\frac{10}{23}&\frac{4}{23}\\0&0&0&1&0&-7&-1\\0&0&0&0&1&4&2\end{bmatrix}$$

(h) $\boldsymbol{r}_1-4\boldsymbol{r}_2\to\boldsymbol{r}_1$得到

$$A\sim\begin{bmatrix}0&1&0&0&0&\frac{6}{23}&-\frac{131}{23}\\0&0&1&0&0&\frac{10}{23}&\frac{4}{23}\\0&0&0&1&0&-7&-1\\0&0&0&0&1&4&2\end{bmatrix}$$

1.7 利用教材中的算法1.1.2求解线性方程组

$$\begin{aligned} 2x_1 - 4x_2 + 3x_3 - 4x_4 - 11x_5 &= 28 \\ -x_1 + 2x_2 - x_3 + 2x_4 + 5x_5 &= -13 \\ -3x_3 + 2x_4 + 5x_5 &= -10 \\ 3x_1 - 5x_2 + 10x_3 - 7x_4 + 12x_5 &= 31 \end{aligned}$$

思路分析：用教材P12中的算法1.1.2求解线性方程组，实质上是利用了简约阶梯型矩阵与原增广矩阵等价的性质。

解：根据原方程组构造增广矩阵

$$\begin{bmatrix} 2 & -4 & 3 & -4 & -11 & 28 \\ -1 & 2 & -1 & 2 & 5 & -13 \\ 0 & 0 & -3 & 2 & 5 & -10 \\ 3 & -5 & 10 & -7 & 12 & 31 \end{bmatrix}$$

依据算法1.1.1将上述矩阵化为简约阶梯型矩阵

(a) $\boldsymbol{r}_1 + 2\boldsymbol{r}_2 \to \boldsymbol{r}_1$，$\boldsymbol{r}_4 + 3\boldsymbol{r}_2 \to \boldsymbol{r}_4$

$$\begin{bmatrix} 0 & 0 & 1 & 0 & -1 & 2 \\ -1 & 2 & -1 & 2 & 5 & -13 \\ 0 & 0 & -3 & 2 & 5 & -10 \\ 0 & 1 & 7 & -1 & 27 & -8 \end{bmatrix}$$

(b) $\boldsymbol{r}_2 - 2\boldsymbol{r}_4 \to \boldsymbol{r}_2$

$$\begin{bmatrix} 0 & 0 & 1 & 0 & -1 & 2 \\ -1 & 0 & -15 & 4 & -49 & 3 \\ 0 & 0 & -3 & 2 & 5 & -10 \\ 0 & 1 & 7 & -1 & 27 & -8 \end{bmatrix}$$

(c) $\boldsymbol{r}_2 + 15\boldsymbol{r}_1 \to \boldsymbol{r}_2$，$\boldsymbol{r}_3 + 3\boldsymbol{r}_1 \to \boldsymbol{r}_3$，$\boldsymbol{r}_4 - 7\boldsymbol{r}_1 \to \boldsymbol{r}_4$

$$\begin{bmatrix} 0 & 0 & 1 & 0 & -1 & 2 \\ -1 & 0 & 0 & 4 & -64 & 33 \\ 0 & 0 & 0 & 2 & 2 & -4 \\ 0 & 1 & 0 & -1 & 34 & -22 \end{bmatrix}$$

(d) $\dfrac{1}{2} \times \boldsymbol{r}_3 \to \boldsymbol{r}_3$

$$\begin{bmatrix} 0 & 0 & 1 & 0 & -1 & 2 \\ -1 & 0 & 0 & 4 & -64 & 33 \\ 0 & 0 & 0 & 1 & 1 & -2 \\ 0 & 1 & 0 & -1 & 34 & -22 \end{bmatrix}$$

(e) $-\boldsymbol{r}_2 \to \boldsymbol{r}_2$，$\boldsymbol{r}_2 + 4\boldsymbol{r}_3 \to \boldsymbol{r}_2$，$\boldsymbol{r}_4 + \boldsymbol{r}_3 \to \boldsymbol{r}_4$

$$\begin{bmatrix} 0 & 0 & 1 & 0 & -1 & 2 \\ 1 & 0 & 0 & 0 & 68 & -41 \\ 0 & 0 & 0 & 1 & 1 & -2 \\ 0 & 1 & 0 & 0 & 35 & -24 \end{bmatrix}$$

到这里，增广矩阵只差互换行的位置就可化成简约阶梯型了，其实已经很容易看出方程组的解。即特解为[27, 11, 1 −1, −1]$^{\mathrm{T}}$，对应的齐次方程组的通解为k[68, 35, −1, 1, −1]。综上所述，原方程组的解可以表示为

$$\boldsymbol{x}=\begin{bmatrix}27\\11\\1\\-1\\-1\end{bmatrix}+k\begin{bmatrix}68\\35\\-1\\1\\-1\end{bmatrix}\quad\forall\ k\in R$$

1.8 假定

$$1^3+2^3+\cdots+n^3=a_1n+a_2n^2+a_3n^3+a_4n^4$$

试求常数a_1,a_2,a_3,a_4。（提示：分别令$n=1,2,3,4$，得到线性方程组。）

解法一：根据自然数的立方和公式

$$1^3+2^3+\cdots+n^3=\left[\frac{n(n+1)}{2}\right]^2$$

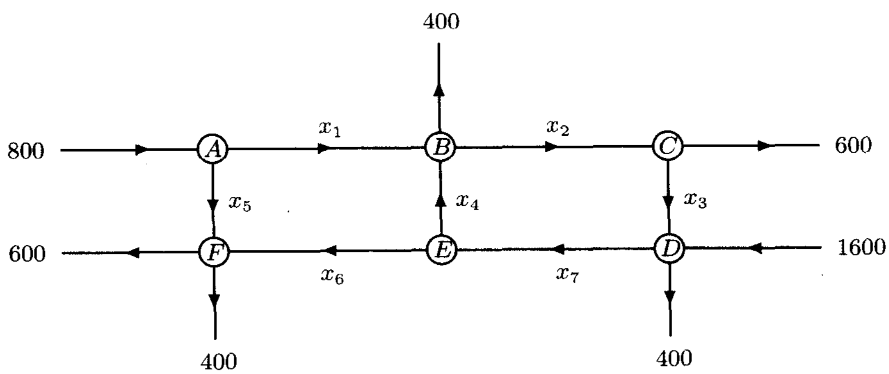
利用待定系数法不难得到 $a_1=0$，$a_2=\dfrac{1}{4}$，$a_3=\dfrac{1}{2}$，$a_4=\dfrac{1}{4}$。

解法二：另一个更为简单的方法是分别令$n=1,2,3,4$，并代入原式中，得到方程组

$$\begin{bmatrix}1&1&1&1\\2&4&8&16\\3&9&27&81\\4&16&64&256\end{bmatrix}\begin{bmatrix}a_1\\a_2\\a_3\\a_3\end{bmatrix}=\begin{bmatrix}1\\9\\36\\100\end{bmatrix}$$

由于系数矩阵可逆，所以该方程组有唯一解。直接计算即可得到与第一种方法相同的结果。值得说明的是，在第二种方法中我们令$n=1,2,3,4$，这样做实际上是为了计算的简便。当然也可以取其他值，并且都会得到相同的结果。

1.9 题1.9图画出了某城市6个交通枢纽的交通网络图[10]。其中，节点表示交通枢纽的编号，数字表示在交通高峰期每小时驶入和驶出某个交通枢纽的车辆数。



题1.9图 交通网络图

(1) 写出表示交通网络图各个交通枢纽的交通流量的线性方程组，并求解该方程组。

(2) 若$x_6=300$辆/小时，$x_7=1300$辆/小时，求交通流量$x_1\sim x_5$。

思路分析： 这实际上是一个求解线性方程组的问题。根据每个节点上的输入、输出的总和为零这一事实，可以列出6个方程，然后直接求解即可。

解： (1) 由题图所示，根据每个节点的输入输出关系列出方程，并组成方程组

$$
\begin{cases}
x_1+x_5=800\\
x_1-x_2+x_4=400\\
x_2-x_3=600\\
x_5+x_6=1000\\
x_4+x_6-x_7=0\\
x_3-x_7=-1200
\end{cases}
$$

经检验发现，系数矩阵的秩为5，小于未知数的个数。所以方程组存在无穷多组解。这里给出通解如下：

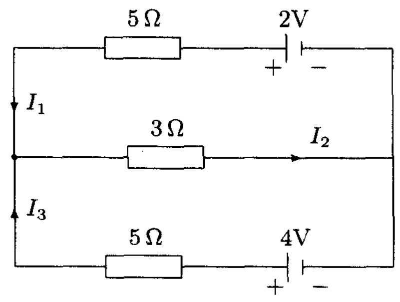
$$
\begin{bmatrix} x_1\\ x_2\\ x_3\\ x_4\\ x_5\\ x_6\\ x_7 \end{bmatrix}
=
\begin{bmatrix} 800\\ 400\\ -200\\ 0\\ 0\\ 1000\\ 1000 \end{bmatrix}
+k_1
\begin{bmatrix} -1\\ 0\\ 0\\ 1\\ 1\\ -1\\ 0 \end{bmatrix}
+k_2
\begin{bmatrix} 0\\ 1\\ 1\\ 1\\ 0\\ 0\\ 1 \end{bmatrix}
$$

其中，$k_1,k_2\in R$。

(2) 将x_6和x_7的值代入到上述通解中，可得到

$$
\begin{cases}
x_1=100\\
x_2=700\\
x_3=100\\
x_4=1000\\
x_5=700
\end{cases}
$$

1.10 题1.10图画出了一电路，求各个支路的电流。



题1.10图 电路图

思路分析：类似于前面的问题，仍然可以通过列出线性方程组来求解。与前面所不同的是，这里除了用到同一节点上电流输入、输出总和为零这一性质，还要用到一个环路上的电势和为零的性质。

解：由题图所示，将三条支路的电流作为三个未知量，根据基尔霍夫定律列出方程组如下：

$$\begin{cases} I_1 + I_3 = I_2 \\ 5I_1 + 3I_2 = 2 \\ 5I_3 + 3I_2 = 4 \end{cases}$$

解得

$$\begin{cases} I_1 = 0.0727 \text{ (A)} \\ I_2 = 0.5455 \text{ (A)} \\ I_3 = 0.4727 \text{ (A)} \end{cases}$$

1.11 证明：由所有2×2 实矩阵组成的集合是一向量空间，若向量加法定义为矩阵加法，向量与标量的乘法定义为矩阵与标量的乘法。（提示：使用向量空间的定义公式(1.2.1)。）

思路分析：根据教材P15页中给出的向量空间的定义，我们可以通过逐条验证的方法来证明题设中的集合在给定的加法和乘法下是一向量空间。由于需要验证的两个闭合性和关于加法及乘法的八个公理中大多比较简单，这里只选取其中的两个给出证明。

证明：(1) 只需选择2×2零矩阵（元素全部为零）作为该向量空间中的零向量，即可满足加法公理中的(a3)这一条。

(2) 对于给定的向量$\boldsymbol{X} \in V$（V为所有2×2实矩阵组成的向量空间），则负向量可以表示为

$$-\boldsymbol{X} = \begin{bmatrix} -x_{11} & -x_{12} \\ -x_{21} & -x_{22} \end{bmatrix}$$

其余各条的证明非常简单，这里不再详细给出。

1.12 令$F : R^3 \mapsto R^2$是一变换，定义为

$$F(\boldsymbol{x}) = \begin{bmatrix} 2x_1 - x_2 \\ x_2 + 5x_3 \end{bmatrix}, \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

试确定F是否为线性变换？

思路分析：确定F是否为线性变换，只需要根据线性变换的定义逐条加以验证即可。

证明：$\forall\ \boldsymbol{x}, \boldsymbol{y} \in R^3$，根据$F$的定义

$$\begin{aligned} F(\boldsymbol{x}+\boldsymbol{y}) &= \begin{bmatrix} 2(x_1+y_1) - (x_2+y_2) \\ (x_2+y_2) + 5(x_3+y_3) \end{bmatrix} \\ &= \begin{bmatrix} 2x_1 - x_2 \\ x_2 + 5x_3 \end{bmatrix} + \begin{bmatrix} 2y_1 - y_2 \\ y_2 + 5y_3 \end{bmatrix} \\ &= F(\boldsymbol{x}) + F(\boldsymbol{y}) \end{aligned}$$

另外，$\forall\ c \in R$，

$$F(c\boldsymbol{x}) = \begin{bmatrix} 2cx_1 - cx_2 \\ cx_2 + 5cx_3 \end{bmatrix} = cF(\boldsymbol{x})$$

所以，F满足线性变换的两条基本性质，它是一个线性变换。

1.13 令$\boldsymbol{A}$ 是一个3×4矩阵，证明$\boldsymbol{A}$的列线性相关。

思路分析：证明$\boldsymbol{A}$的各列线性相关等价于证明该矩阵非满列秩。另外，也可考虑$\boldsymbol{A}$的各列组成的向量组，利用向量组的线性相关性来证明。

证明：根据这两种思路，可以得到如下两种证明方法。

解法一：依题意，有$\text{rank}(\boldsymbol{A}) \leqslant \min\{3,4\} = 3$，所以$\boldsymbol{A}$一定不是列满秩矩阵，即其列线性相关。

解法二：由于R^n中的任意$n+1$个向量必定线性相关，当$n=3$时即为本题的情况，所以$\boldsymbol{A}$的列一定线性相关。

1.14 令$\boldsymbol{A}$ 是一个4×3矩阵，证明$\boldsymbol{A}$的行线性相关。

证明：完全同上，这里不再重复。

1.15 令V 是一n维子空间，且$\boldsymbol{b}$为任意$n \times 1$向量。证明：存在$n \times 1$向量$\boldsymbol{v}_0 \in V$，使得

$$\|\boldsymbol{b} - \boldsymbol{v}_0\| \leqslant \|\boldsymbol{b} - \boldsymbol{v}\|, \qquad \forall\ \boldsymbol{v} \in V$$

思路分析：该题目考察的是向量空间、基和向量之间的距离等一些概念。要证明的问题的物理意义可以表述为：若$\boldsymbol{b} \in V$，则$\boldsymbol{v}_0 = \boldsymbol{b}$；若$\boldsymbol{b} \notin V$，则一定可以在$V$内找到一个向量与$\boldsymbol{b}$的距离最近。

证明：设$\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_k$为$n$维子空间$V$的一组正交基（$k$为子空间的维数），将其扩充后，得到$\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_k,\boldsymbol{\beta}_{k+1},\cdots,\boldsymbol{\beta}_n$构成了$n$维线性空间$R^n$的一组正交基。若令矩阵

$$\boldsymbol{A}=[\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_k]$$
$$\boldsymbol{B}=[\boldsymbol{\beta}_{k+1},\cdots,\boldsymbol{\beta}_n]$$

显然，R^n中的任意向量$\boldsymbol{b}$可以唯一表示为

$$\boldsymbol{b}=[\boldsymbol{A},\boldsymbol{B}]\left[\boldsymbol{x}^{\mathrm{T}},\boldsymbol{y}^{\mathrm{T}}\right]^{\mathrm{T}}$$

另外，子空间V中的任意向量$\boldsymbol{v}$可以唯一表示为$\boldsymbol{v}=\boldsymbol{A}\boldsymbol{z}$。

事实上，只需要令$\boldsymbol{v}_0=\boldsymbol{A}\boldsymbol{x}$，则$\|\boldsymbol{b}-\boldsymbol{v}_0\|$必为最小。这是因为

$$\begin{aligned}\|\boldsymbol{b}-\boldsymbol{v}\|&=\|\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{z})+\boldsymbol{B}\boldsymbol{y}\|\\&=\sqrt{\|\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{z})\|^2+\|\boldsymbol{B}\boldsymbol{y}\|^2}\\&\geqslant\|\boldsymbol{B}\boldsymbol{y}\|\\&=\|\boldsymbol{b}-\boldsymbol{v}_0\|\end{aligned}$$

上式中的第二个等号的成立实际上是利用了构造的正交基的性质。

1.16 [11] 矩阵的秩在工程控制系统的设计起着重要的作用。一个离散时间的控制系统的状态空间模型包括了差分方程

$$\boldsymbol{x}_{k+1}=\boldsymbol{A}\boldsymbol{x}_k+\boldsymbol{B}\boldsymbol{u}_k,\qquad k=0,1,\cdots$$

式中，$\boldsymbol{A}\in R^{n\times n},\boldsymbol{B}\in R^{n\times m}$，并且$\boldsymbol{x}_k\in R^n$ 为描述系统在k时刻的状态的向量，简称状态向量；而$\boldsymbol{u}_k$为系统在k时刻的输入或控制向量。矩阵对$(\boldsymbol{A},\boldsymbol{B})$称为可控的，若

$$\operatorname{rank}\left([\boldsymbol{B},\boldsymbol{A}\boldsymbol{B},\boldsymbol{A}^2\boldsymbol{B},\cdots,\boldsymbol{A}^{n-1}\boldsymbol{B}]\right)=n$$

且$(\boldsymbol{A},\boldsymbol{B})$是可控的，则最多用$n$步即可将系统控制到任意一个指定的状态$\boldsymbol{x}$。试确定以下矩阵对是否可控：

(1) $\boldsymbol{A}=\begin{bmatrix}0.9&1&0\\0&-0.9&0\\0&0&0.5\end{bmatrix},\ \boldsymbol{B}=\begin{bmatrix}0\\1\\1\end{bmatrix}$

(2) $\boldsymbol{A}=\begin{bmatrix}0.8&-0.3&0\\0.2&0.5&1\\0&0&-0.5\end{bmatrix},\ \boldsymbol{B}=\begin{bmatrix}1\\1\\0\end{bmatrix}$

解：可以直接应用题目中给出的矩阵对可控的定义式来判断：

(1) 经计算后得到

$$[\boldsymbol{B},\boldsymbol{A}\boldsymbol{B},\boldsymbol{A}^2\boldsymbol{B}]=\begin{bmatrix}0&1&0\\1&-0.9&0.81\\1&0.5&0.25\end{bmatrix}$$

其秩为3，所以矩阵对$(\boldsymbol{A},\boldsymbol{B})$是可控的。

(2) 同理，此问中

$$[\boldsymbol{B},\boldsymbol{A}\boldsymbol{B},\boldsymbol{A}^2\boldsymbol{B}]=\begin{bmatrix}1&0.5&0.19\\1&0.7&0.45\\0&0&0\end{bmatrix}$$

其秩为2，所以矩阵对$(\boldsymbol{A},\boldsymbol{B})$是不可控的。

1.17 令 U 和 V 是 Euclidean n 空间 R^n 的两个子空间，并假定 V 是 U 的子集。证明：$\dim(V)\leqslant\dim(U)$。若$\dim(V)=\dim(U)$，证明U包含于V，因此$V=U$。

思路分析：此题考察的是向量组的线性相关性和子空间维数的有关性质，比较容易证明。第一问只需证明子空间U的基向量的个数小于V即可；第二问可以通过证明U的基向量可以由V的基向量线性表出来解决。

证明：设 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_p$ 和 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_q$ 分别是子空间 U 和 V 的基向量，这就意味着 $\dim(U)=p$，$\dim(V)=q$。由于V是U的子集，则向量组$\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_q$可以由向量组$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_p$线性表出。两个线性无关的向量组有这样的关系，若$\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_q$可以由$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_p$线性表出，则$q\leqslant p$ (教材中有相关说明)。因此，第一问得证。

当$p=q$时，定义矩阵$\boldsymbol{A}=[\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_p]$，$\boldsymbol{B}=[\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_q]$。根据题目中给出的条件，$V$是$U$的子集，则可以用线性变换来描述$\boldsymbol{A}$和$\boldsymbol{B}$的关系，即

$$\boldsymbol{B}=\boldsymbol{A}\boldsymbol{P}$$

其中，$\boldsymbol{P}$为$p\times p$阶方阵。因为$\boldsymbol{A}$和$\boldsymbol{B}$都是列满秩矩阵，所以$\boldsymbol{P}$为可逆矩阵。因此

$$\boldsymbol{A}=\boldsymbol{P}^{-1}\boldsymbol{B}$$

即向量组$\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_q$也可以由向量组$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_p$线性表出。这就表明，$U$的基向量也可以由$V$的基线性表示，即$U$包含于$V$。两个子空间相互包含，表明二者是相等的。

1.18 令正方矩阵$\boldsymbol{A}$和$\boldsymbol{B}$ 具有相同的维数，证明$\mathrm{tr}(\boldsymbol{A}\boldsymbol{B})=\mathrm{tr}(\boldsymbol{B}\boldsymbol{A})$。

思路分析：求迹是矩阵的一种重要运算，此题目实际上是要求证明迹的一条性质：迹是相似不变量。可根据迹的定义来证明。

证明：从方阵的迹的定义出发，有

$$\mathrm{tr}(\boldsymbol{A}\boldsymbol{B})=\sum_i\sum_j\boldsymbol{A}_{ij}\,\boldsymbol{B}_{ji}=\sum_j\sum_i\boldsymbol{B}_{ji}\,\boldsymbol{A}_{ij}=\mathrm{tr}(\boldsymbol{B}\boldsymbol{A})$$

到这里，问题已经证毕。实际上，并不要求矩阵$\boldsymbol{A}$和$\boldsymbol{B}$一定是具有相同维数的方阵，只要二者乘积是方阵就可以。

1.19 解释为什么

(1) $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\mathrm{tr}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})$。

(2) $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \mathrm{tr}(\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}})$。

解：

(1) 根据迹操作的定义，标量的迹就是其本身。

(2) 根据1.18题结论知，此等式必然成立。

1.20 令$\boldsymbol{A}$为$n \times n$矩阵，证明：

$$\frac{\partial|\boldsymbol{A} - \lambda\boldsymbol{I}|}{\partial\lambda} = -\sum_{k=1}^{n}|\boldsymbol{A}_k - \lambda\boldsymbol{I}|$$

式中，$\boldsymbol{A}_k$是$(n-1)\times(n-1)$维矩阵，它是从矩阵$\boldsymbol{A}$中删去第k行和第k列剩下的子矩阵。

思路分析：此题看似无从下手。容易想到的是按照一行或一列来展开行列式，然后用数学归纳法证明。可实际操作时却发现很难进行下去。如果利用行列式的另一种展开方法，则会得到简单的证明。

证明：首先，引入行列式的另一种求解方法，或称为展开形式：

对于$n \times n$矩阵$\boldsymbol{A} = [a_{ij}]$，其行列式可以表示为

$$|\boldsymbol{A}| = \sum_{p}\sigma(p)a_{1p_1}a_{2p_2}\cdots a_{np_n}$$

其中$p = (p_1, p_2, \cdots, p_n)$为$(1, 2, \cdots, n)$的一个重排列，也就是说$a_{1p_1}, a_{2p_2}, \cdots, a_{np_n}$相当于每一行每一列只取一个元素做乘积，所以求和号中应该共有$n!$项。而$\sigma(p)$ 则定义为

$$\sigma(p) = \begin{cases} +1 & p\text{可以经过偶数次互换元素恢复为自然排列顺序} \\ -1 & p\text{可以经过奇数次互换元素恢复为自然排列顺序} \end{cases}$$

详细内容请见文献[20]。有了这样的展开式，我们可以直接得到如下结论[20]：

如果$\boldsymbol{A}$矩阵的所有元素都是变量x的可微函数，则

$$\frac{\mathrm{d}|\boldsymbol{A}|}{\mathrm{d}x} = |\boldsymbol{D}_1| + |\boldsymbol{D}_2| + \cdots + |\boldsymbol{D}_n|$$

其中，$\boldsymbol{D}_i$表示除了i行上的元素被替换为$\mathrm{d}a_{ij}/\mathrm{d}t(j = 1, 2, \cdots, n)$外，其余行均与$\boldsymbol{A}$完全相同的矩阵。上述命题的证明并不复杂。

$$\begin{aligned}
\frac{\mathrm{d}|\boldsymbol{A}|}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\sum_{p}\sigma(p)a_{1p_1}a_{2p_2}\cdots a_{np_n} \\
&= \sum_{p}\sigma(p)\frac{\mathrm{d}(a_{1p_1}a_{2p_2})\cdots a_{np_n}}{\mathrm{d}t} \\
&= \sum_{p}\sigma(p)\left(a_{1p_1}^{'}a_{2p_2}^{'}\cdots a_{np_n} + a_{1p_1}a_{2p_2}^{'}\cdots a_{np_n} + \cdots + a_{1p_1}a_{2p_2}\cdots a_{np_n}^{'}\right) \\
&= \sum_{p}\sigma(p)a_{1p_1}^{'}a_{2p_2}\cdots a_{np_n} + \sum_{p}\sigma(p)a_{1p_1}a_{2p_2}^{'}\cdots a_{np_n} + \cdots \\
&\quad + \sum_{p}\sigma(p)a_{1p_1}a_{2p_2}\cdots a_{np_n}^{'} \\
&= |\boldsymbol{D}_1| + |\boldsymbol{D}_2| + \cdots + |\boldsymbol{D}_n|
\end{aligned}$$

本题的证明到了这里就非常简单了，只需要直接应用上面的命题即可。将上面的$\boldsymbol{A}$矩阵用$\boldsymbol{A}-\lambda\boldsymbol{I}$替换，得到

$$\boldsymbol{D}_i=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1,i-1} & a_{1,i} & a_{1,i+1} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2,i-1} & a_{2,i} & a_{2,i+1} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{i-1,1} & a_{i-1,2} & \cdots & a_{i-1,i-1} & a_{i-1,i} & a_{i-1,i+1} & \cdots & a_{i-1,n} \\ 0 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \\ a_{i+1,1} & a_{i+1,2} & \cdots & a_{i+1,i-1} & a_{i+1,i} & a_{i+1,i+1} & \cdots & a_{i+1,n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{n,i-1} & a_{n,i} & a_{n,i+1} & \cdots & a_{nn} \end{bmatrix}$$

从而，

$$|\boldsymbol{D}_i|=-|\boldsymbol{A}_i-\lambda\boldsymbol{I}|$$
$$\frac{\partial|\boldsymbol{A}-\lambda\boldsymbol{I}|}{\partial\lambda}=-\sum_{k=1}^{n}|\boldsymbol{A}_k-\lambda\boldsymbol{I}|$$

1.21 直线方程可以表示为$ax+by=-1$。证明一条通过点(x_1,y_1)和(x_2,y_2)的直线方程为

$$\begin{vmatrix} 1 & x & y \\ 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \end{vmatrix}=0$$

证明：通过(x_1,y_1)和(x_2,y_2)两点的直线方程可表示为

$$(y_2-y_1)x-(x_2-x_1)y=y_2x_1-x_2y_1$$

上式可等效写作

$$\begin{vmatrix} 1 & x & y \\ 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \end{vmatrix}=0$$

1.22 平面方程可以表示为$ax+by+cz=-1$。证明：通过三点(x_i,y_i,z_i), $i=1,2,3$的平面方程由

$$\begin{vmatrix} 1 & x & y & z \\ 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \end{vmatrix}=0$$

决定。

证明：将三点的坐标代入到平面方程中，得到方程组

$$\begin{cases} ax_1+by_1+cz_1=-1 \\ ax_2+by_2+cz_2=-1 \\ ax_3+by_3+cz_3=-1 \end{cases}$$

利用Gramer法则得到

$$a=\frac{\begin{vmatrix}-1 & y_1 & z_1\\ -1 & y_2 & z_2\\ -1 & y_3 & z_3\end{vmatrix}}{\begin{vmatrix}x_1 & y_1 & z_1\\ x_2 & y_2 & z_2\\ x_3 & y_3 & z_3\end{vmatrix}},\quad b=\frac{\begin{vmatrix}x_1 & -1 & z_1\\ x_2 & -1 & z_2\\ x_3 & -1 & z_3\end{vmatrix}}{\begin{vmatrix}x_1 & y_1 & z_1\\ x_2 & y_2 & z_2\\ x_3 & y_3 & z_3\end{vmatrix}},\quad c=\frac{\begin{vmatrix}x_1 & y_1 & -1\\ x_2 & y_2 & -1\\ x_3 & y_3 & -1\end{vmatrix}}{\begin{vmatrix}x_1 & y_1 & z_1\\ x_2 & y_2 & z_2\\ x_3 & y_3 & z_3\end{vmatrix}}$$

因此，待求平面方程可写作

$$\begin{vmatrix}-1 & y_1 & z_1\\ -1 & y_2 & z_2\\ -1 & y_3 & z_3\end{vmatrix}x+\begin{vmatrix}x_1 & -1 & z_1\\ x_2 & -1 & z_2\\ x_3 & -1 & z_3\end{vmatrix}y+\begin{vmatrix}x_1 & y_1 & -1\\ x_2 & y_2 & -1\\ x_3 & y_3 & -1\end{vmatrix}z=\begin{vmatrix}x_1 & y_1 & z_1\\ x_2 & y_2 & z_2\\ x_3 & y_3 & z_3\end{vmatrix}$$

化简后，即为

$$\begin{vmatrix}1 & x & y & z\\ 1 & x_1 & y_1 & z_1\\ 1 & x_2 & y_2 & z_2\\ 1 & x_3 & y_3 & z_3\end{vmatrix}=0$$

1.23 在不展开行列式的情况下，证明下列结果：

$$2\begin{vmatrix}a & b & c\\ d & e & f\\ x & y & z\end{vmatrix}=\begin{vmatrix}a+b & b+c & c+a\\ d+e & e+f & f+d\\ x+y & y+z & z+x\end{vmatrix}$$

和

$$2\begin{vmatrix}0 & a & b\\ a & 0 & c\\ b & c & 0\end{vmatrix}=\begin{vmatrix}b+a & c & c\\ b & a+c & b\\ a & a & c+b\end{vmatrix}$$

思路分析：这是一道常见的题目，求解时需要利用行列式的性质及一定的技巧。

证明：根据行列式的性质，可以直接得到

$$\begin{aligned}\begin{vmatrix}a+b & b+c & c+a\\ d+e & e+f & f+d\\ x+y & y+z & z+x\end{vmatrix}&=\begin{vmatrix}a & b+c & c+a\\ d & e+f & f+d\\ x & y+z & z+x\end{vmatrix}+\begin{vmatrix}b & b+c & c+a\\ e & e+f & f+d\\ y & y+z & z+x\end{vmatrix}\\ &=\begin{vmatrix}a & b & c+a\\ d & e & f+d\\ x & y & z+x\end{vmatrix}+\begin{vmatrix}b & c & c+a\\ e & f & f+d\\ y & z & z+x\end{vmatrix}\\ &=2\begin{vmatrix}a & b & c\\ d & e & f\\ x & y & z\end{vmatrix}\end{aligned}$$

对于第二个证明，只需要将第一问得到的结果进行变量替换，即令

$$\begin{aligned}&a=0,\quad b=u,\quad c=v\\ &d=u,\quad e=0,\quad f=w\\ &x=v,\quad y=w,\quad z=0\end{aligned}$$

则有

$$2\begin{vmatrix}0 & u & v\\ u & 0 & w\\ v & w & 0\end{vmatrix}=\begin{vmatrix}u & u+v & v\\ u & w & u+w\\ v+w & w & v\end{vmatrix}=\begin{vmatrix}v+u & w & w\\ v & u+w & v\\ u & u & w+v\end{vmatrix}$$

这实际上正是第二问要求证明的等式。

1.24 令$\boldsymbol{A}_{n\times n}$正定，并且$\boldsymbol{B}_{n\times n}$ 半正定，证明

$$\det(\boldsymbol{A}+\boldsymbol{B})\geqslant\det(\boldsymbol{A})$$

思路分析：本题及1.26题实际上是行列式的几个基本的不等式关系，见教材P58 ~ 59。证明思路相似，都要用到Minkowski不等式。因此，我们只给出本题目的证明，1.26题则不再重复证明。

证明：Minkowski不等式是这样表述的，若$\boldsymbol{A}_{n\times n}\neq\boldsymbol{O}_{n\times n}$，$\boldsymbol{B}_{n\times n}\neq\boldsymbol{O}_{n\times n}$半正定，则

$$\sqrt[n]{\det(\boldsymbol{A}+\boldsymbol{B})}\geqslant\sqrt[n]{\det(\boldsymbol{A})}+\sqrt[n]{\det(\boldsymbol{B})}$$

本题目中已知$\boldsymbol{A}$正定，$\boldsymbol{B}$半正定，自然可以套用上述结论，得到

$$\begin{aligned}\det(\boldsymbol{A}+\boldsymbol{B})&\geqslant\left(\sqrt[n]{\det(\boldsymbol{A})}+\sqrt[n]{\det(\boldsymbol{B})}\right)^n\\&\geqslant\left(\sqrt[n]{\det(\boldsymbol{A})}\right)^n\\&=\det(\boldsymbol{A})\end{aligned}$$

上式中第二个不等号处用到了半正定矩阵的行列式大于等于零的性质。

1.25 令$\boldsymbol{A}_{n\times n}$和$\boldsymbol{B}_{n\times n}$都是半正定矩阵，证明

$$\det(\boldsymbol{A}+\boldsymbol{B})\geqslant\det(\boldsymbol{A})+\det(\boldsymbol{B})$$

证明：思路同上。

1.26 令$\boldsymbol{A}_{12\times 12}$满足$\boldsymbol{A}^5=3\boldsymbol{A}$。试求$|\boldsymbol{A}|$的所有可能的数值。

解：依题意，可立即得到

$$\left|\boldsymbol{A}^5\right|=|3\boldsymbol{A}|$$

设$x=|\boldsymbol{A}|$，应用行列式的基本性质可得

$$x^5-3^{12}x=0$$

所以

$$x_1=0,\quad x_2=27\mathrm{e}^{\mathrm{j}\frac{\pi}{2}},\quad x_3=27\mathrm{e}^{\mathrm{j}\pi},\quad x_4=27\mathrm{e}^{\mathrm{j}\frac{3\pi}{2}},\quad x_5=27$$

1.27 已知$\boldsymbol{X}=[\boldsymbol{A},\boldsymbol{B}]$是一个分块矩阵，证明

$$|\boldsymbol{X}|^2=|\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}|=\begin{vmatrix}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}\\ \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A} & \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\end{vmatrix}$$

思路分析：此题考察的是矩阵转置的行列式与矩阵乘积的行列式的基本性质，证明较为简单。

证明：直接给出证明结果如下

$$|\boldsymbol{X}|^2=\left|\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\right|=\left|[\boldsymbol{A},\boldsymbol{B}]\begin{bmatrix}\boldsymbol{A}^{\mathrm{T}}\\ \boldsymbol{B}^{\mathrm{T}}\end{bmatrix}\right|=\left|\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\right|$$

同时

$$|\boldsymbol{X}|^2=\left|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right|=\left|\begin{bmatrix}\boldsymbol{A}^{\mathrm{T}}\\ \boldsymbol{B}^{\mathrm{T}}\end{bmatrix}[\boldsymbol{A},\boldsymbol{B}]\right|=\begin{vmatrix}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}\\ \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A} & \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\end{vmatrix}$$

1.28 [23, p.361] 已知$n\times n$矩阵$\boldsymbol{M}=\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}$。若矩阵$\boldsymbol{X}$的秩为$r_X$，且$\boldsymbol{y}$是一个正态分布的随机向量，其均值向量为$\boldsymbol{X}\boldsymbol{b}$，协方差矩阵为$\sigma^2\boldsymbol{I}$，即$\boldsymbol{y}\sim N(\boldsymbol{X}\boldsymbol{b},\sigma^2\boldsymbol{I})$，证明

(1) $\mathrm{E}\{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{y}\}=(n-r_X)\sigma^2$。

(2) $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{y}$和$\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}$是统计独立的随机变量。

(3) $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{y}/\sigma^2$服从自由度为$(n-r_X)$的$\mathcal{X}^2$分布，即$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{y}/\sigma^2\sim\mathcal{X}^2_{n-r_X}$。

(4) 当$\boldsymbol{X}\boldsymbol{b}=\boldsymbol{0}$时，$\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}/\sigma^2\sim\mathcal{X}^2_{r_X}$。

思路分析：这个问题较为复杂，除了涉及比较繁琐的推导外，还具有明确的物理意义，即子空间与投影的概念。下面在证明过程中逐一阐释。

证明：(1) 第一问比较简单，可以直接证明。其中需要使用迹函数的一些性质。由于已知“矩阵$\boldsymbol{X}$的秩为r_X”，且$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$可逆，所以可以推出$\boldsymbol{X}$的列数为$r_X$。进而，

$$\begin{aligned}\mathrm{E}\left\{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{y}\right\}&=\mathrm{tr}\left\{\mathrm{E}\left\{\boldsymbol{M}\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}\right\}\right\}\\&=\mathrm{tr}[\boldsymbol{M}(\sigma^2\boldsymbol{I}+\boldsymbol{X}\boldsymbol{b}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})]\\&=\mathrm{tr}[\sigma^2\boldsymbol{I}-\sigma^2\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}+\boldsymbol{M}\boldsymbol{X}\boldsymbol{b}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}]\\&=(n-r_X)\sigma^2\end{aligned}$$

第四个等号的成立用到了$\boldsymbol{M}\boldsymbol{X}=\boldsymbol{O}$这一事实。

(2) 题目中给出的矩阵$\boldsymbol{M}$称为正交投影矩阵，$\boldsymbol{M}\boldsymbol{y}$相当于向量$\boldsymbol{y}$到$\boldsymbol{X}$的列张成的子空间的正交补上的投影。而$\boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}$称为投影矩阵，若记为$\boldsymbol{N}$，则$\boldsymbol{M}+\boldsymbol{N}=\boldsymbol{I}$。以上内容详见教材最后一章。投影矩阵（包括正交投影矩阵）的重要性质为幂等性和对称性，即

$$\boldsymbol{M}^2=\boldsymbol{M},\qquad \boldsymbol{M}^{\mathrm{T}}=\boldsymbol{M}$$

设$\boldsymbol{u}=\boldsymbol{M}\boldsymbol{y}$，$\boldsymbol{v}=\boldsymbol{N}\boldsymbol{y}$，则可以得到$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{y}=\|\boldsymbol{u}\|^2$，$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{N}\boldsymbol{y}=\|\boldsymbol{v}\|^2$。这一问要求证明标量$\|\boldsymbol{u}\|^2$和$\|\boldsymbol{v}\|^2$独立，我们只需证明$\boldsymbol{u}$和$\boldsymbol{v}$独立即可。

观察$\boldsymbol{u}$和$\boldsymbol{v}$的结构可知，它们是正态分布的随机向量$\boldsymbol{y}$经过不同的线性变换得到的。易知，二者也都是正态分布的随机向量。这样，我们证明两个向量相互独立，只需证明它们不相关即可。求二者的互协方差矩阵，得到

$$\begin{aligned}\boldsymbol{C}_{\boldsymbol{u},\boldsymbol{v}}&=\mathrm{E}\left\{\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\right\}-\mathrm{E}\left\{\boldsymbol{u}\right\}\mathrm{E}\left\{\boldsymbol{v}^{\mathrm{T}}\right\}\\&=\boldsymbol{M}\mathrm{E}\left\{\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}\right\}\boldsymbol{N}^{\mathrm{T}}-\boldsymbol{M}\boldsymbol{X}\boldsymbol{b}(\boldsymbol{N}\boldsymbol{X}\boldsymbol{b})^{\mathrm{T}}\\&=\boldsymbol{O}\end{aligned}$$

上式中最后一个等式实际上用到了投影矩阵与正交投影矩阵正交的性质，即$\boldsymbol{M}\boldsymbol{N}=\boldsymbol{O}$以及$\boldsymbol{M}\boldsymbol{X}=\boldsymbol{O}$的事实。这样，我们就证明了向量$\boldsymbol{u}$和$\boldsymbol{v}$是独立的，从而证明了$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{y}$和$\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}$是独立的。

(3) 对正交投影矩阵$\boldsymbol{M}$做特征值分解，得到

$$\boldsymbol{M}=[\boldsymbol{U}_s,\boldsymbol{U}_n]\boldsymbol{\Sigma}\begin{bmatrix}\boldsymbol{U}_s^{\mathrm{T}}\\\boldsymbol{U}_n^{\mathrm{T}}\end{bmatrix}$$

其中，$\boldsymbol{\Sigma}=\mathrm{diag}(\sigma_1,\cdots,\sigma_{n-r_X},0,\cdots,0)$为对角阵，其前$n-r_X$个元素为1，后$r_X$个元素全部为零，这实际上是由投影矩阵和正交投影矩阵的幂等性质决定的。$\boldsymbol{U}_s$为对应于特征值为1的特征向量构成的矩阵，而$\boldsymbol{U}_n$则为对应于特征值零的特征向量构成的矩阵。此时，$\boldsymbol{M}$可重新写作$\boldsymbol{M}=\boldsymbol{U}_s\boldsymbol{U}_s^{\mathrm{T}}$。

接下来，定义$\boldsymbol{w}=\boldsymbol{U}_s^{\mathrm{T}}\boldsymbol{y}$，则易知

$$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{y}=\|\boldsymbol{w}\|^2$$

并且

$$\mathrm{E}\left\{\boldsymbol{w}\right\}=\boldsymbol{U}_s^{\mathrm{T}}\boldsymbol{X}\boldsymbol{b}$$

因为

$$\boldsymbol{M}\boldsymbol{X}=\boldsymbol{O}$$

所以

$$\boldsymbol{U}_s^{\mathrm{T}}\boldsymbol{X}=\boldsymbol{O},\quad \mathrm{E}\left\{\boldsymbol{w}\right\}=\boldsymbol{0}$$

另外，$\boldsymbol{w}$的自协方差矩阵可以表示为

$$\boldsymbol{C}_{\boldsymbol{w},\boldsymbol{w}}=\mathrm{E}\left\{\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}\right\}=\boldsymbol{U}_s^{\mathrm{T}}\mathrm{E}\left\{\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}\right\}\boldsymbol{U}_s=\sigma^2\boldsymbol{I}$$

综上可知，$\boldsymbol{w}\sim N(\boldsymbol{0},\sigma^2\boldsymbol{I})$。因此，$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{y}/\sigma^2$相当于$n-r_X$个独立的标准正态分布的随机变量的平方和，必然服从自由度为$n-r_X$的$\mathcal{X}^2_{n-r_X}$分布。

(4) 这一问的证明同上，不再重复。

1.29 令$\boldsymbol{A}^2=\boldsymbol{A}$，用下面两种方法证明$\text{rank}(\boldsymbol{I}-\boldsymbol{A})=n-\text{rank}(\boldsymbol{A})$：

(1) 利用矩阵$\boldsymbol{A}$的迹与秩相等的性质。

(2) 考虑线性方程组$(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{0}$的线性无关解。

思路分析：本题综合考察矩阵的秩、迹、矩阵与线性方程组的关系等概念。其求解思路在两个问题中已给出。

证明：(1) 由$\boldsymbol{A}$是幂等矩阵，我们可以得到

$$(\boldsymbol{I}-\boldsymbol{A})^2=\boldsymbol{I}-2\boldsymbol{A}+\boldsymbol{A}^2=\boldsymbol{I}-\boldsymbol{A}$$

即$\boldsymbol{I}-\boldsymbol{A}$也是幂等矩阵。由幂等矩阵的秩与迹相等的性质可以得到

$$\text{rank}(\boldsymbol{A})=\text{tr}(\boldsymbol{A}),\quad \text{rank}(\boldsymbol{I}-\boldsymbol{A})=\text{tr}(\boldsymbol{I}-\boldsymbol{A})$$

从而，有

$$\text{rank}(\boldsymbol{I}-\boldsymbol{A})=n-\text{tr}(\boldsymbol{A})=n-\text{rank}(\boldsymbol{A})$$

(2) 因为

$$\boldsymbol{A}(\boldsymbol{I}-\boldsymbol{A})=\boldsymbol{O}$$

所以

$$\text{Null}(\boldsymbol{A})\supseteq \text{range}(\boldsymbol{I}-\boldsymbol{A})$$

从而

$$n-\text{rank}(\boldsymbol{A})\geqslant \text{rank}(\boldsymbol{I}-\boldsymbol{A})$$

根据矩阵和的秩的不等式关系，即前面得到的

$$\text{rank}(\boldsymbol{A})+\text{rank}(\boldsymbol{I}-\boldsymbol{A})\geqslant n$$

从而使问题得证。

1.30 已知矩阵

$$\boldsymbol{A}=\begin{bmatrix}1&1&1&2\\-1&0&2&-3\\2&4&8&5\end{bmatrix}$$

求矩阵$\boldsymbol{A}$的秩和零维。 (提示：将矩阵$\boldsymbol{A}$化为阶梯型。)

思路分析：由于初等变换不会改变矩阵的秩，因而我们可以利用高斯消去法先把该矩阵化为阶梯型，从而求出其秩。在得到矩阵的秩后，其零空间的维数也就同时得到了。

解：使用高斯消去法对$\boldsymbol{A}$进行变换，用$\boldsymbol{a}_i$表示经过上一次变换后矩阵$\boldsymbol{A}$的第i行

(1) $\boldsymbol{a}_2+\boldsymbol{a}_1\to\boldsymbol{a}_2, \boldsymbol{a}_3-2\boldsymbol{a}_1\to\boldsymbol{a}_3$

$$\begin{bmatrix}1&1&1&2\\0&1&3&-1\\0&2&6&1\end{bmatrix}$$

(2) $\boldsymbol{a}_3 - 2\boldsymbol{a}_2 \to \boldsymbol{a}_3$

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

化简到这里，不难看出，$\text{rank}(\boldsymbol{A}) = 3$，并且$\dim[\text{Null}(\boldsymbol{A})] = 4 - 3 = 1$。

1.31 令$\boldsymbol{A}$是一个$m \times n$矩阵，证明$\text{rank}(\boldsymbol{A}) \leqslant m$和$\text{rank}(\boldsymbol{A}) \leqslant n$。

证明：根据教材P60中的定理1.62和定义1.6.5可知，$\text{rank}(\boldsymbol{A})$表示矩阵中线性无关的行（或列）的数目，必然小于或等于矩阵的行数和列数中最小的一个。自然就得到$\text{rank}(\boldsymbol{A}) \leqslant m$且$\text{rank}(\boldsymbol{A}) \leqslant n$。

1.32 考虑线性方程组

$$x_1 + 3x_2 - x_3 = a_1$$

$$x_1 + 2x_2 = a_2$$

$$3x_1 + 7x_2 - x_3 = a_3$$

(1) 确定线性方程组为一致方程的充分必要条件。

(2) 假定三种情况：

① $a_1 = 2, a_2 = 2, a_3 = 6$

② $a_1 = 1, a_2 = 0, a_3 = -2$

③ $a_1 = 0, a_2 = 1, a_3 = 2$

判断线性方程组是否为一致方程。若是一致方程，则给出相对应的解。

思路分析：教材P74页一致方程的定义1.8.2：线性方程$\boldsymbol{Ax} = \boldsymbol{y}$是一致方程，当且仅当方程组系数矩阵行之间存在的线性关系也存在于向量$\boldsymbol{y}$中。根据这一定义，我们可以方便求解本题。

解：(1) 经过观察可以发现，此方程组系数矩阵的第一行加上第二行的2倍正好等于第三行，则其满足一致方程的充要条件是$a_1 + 2a_2 = a_3$。

(2) 按照上面的充要条件关系，条件①③可构成一致方程。其中，第一组系数代入到方程后，得到的解为$[x_1, x_2, x_3]^{\mathrm{T}} = [0, 1, 1]^{\mathrm{T}} + k[2, -1, -1]^{\mathrm{T}}, \forall\, k \in R$；第三组系数代入到方程后，解为$[x_1, x_2, x_3]^{\mathrm{T}} = [3, -1, 0]^{\mathrm{T}} + k[2, -1, -1]^{\mathrm{T}}, \forall\, k \in R$。

1.33 令$\boldsymbol{A}$是一个3×4矩阵，其零维等于1。证明对3×1实向量$\boldsymbol{b}$的每一种选择，3×4线性方程$\boldsymbol{Ax} = \boldsymbol{b}$均是一致方程。

思路分析：上一个题目用到了一致方程的定义，本题目的证明可利用教材P74页的定理1.8.4：方程组是一致方程，当且仅当系数矩阵的秩等于增广矩阵的秩。

证明：依题意，3×4矩阵$\boldsymbol{A}$的零空间维数是1，则矩阵$\boldsymbol{A}$的秩为

$$\mathrm{rank}(\boldsymbol{A}) = 4 - \dim[\mathrm{Null}(\boldsymbol{A})] = 4 - 1 = 3$$

自然地，$\mathrm{rank}(\boldsymbol{A}) = \mathrm{rank}([\boldsymbol{A} \quad \boldsymbol{b}]) = 3$。

所以，对于任意的$\boldsymbol{b}$向量，$\boldsymbol{Ax} = \boldsymbol{b}$ 一定是一致方程。

1.34 设线性方程组

$$\begin{aligned} x_1 + a_1x_2 + a_1^2x_3 &= a_1^3 \\ x_1 + a_2x_2 + a_2^2x_3 &= a_2^3 \\ x_1 + a_3x_2 + a_3^2x_3 &= a_3^3 \\ x_1 + a_4x_2 + a_4^2x_3 &= a_4^3 \end{aligned}$$

(1) 证明：若a_1, a_2, a_3, a_4两两不相等，则该线性方程组无解。

(2) 设$a_1 = a_3 = c$和$a_2 = a_4 = -c$，其中，$c \neq 0$，且已知

$$\boldsymbol{\alpha}_1 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{\alpha}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

是线性方程组的两个解，写出此方程组的通解。

思路分析：线性方程组是否有解，与该方程组是否是一致方程等价。若要证明(1)中该方程组无解，我们只需证明该方程组不是一致方程即可。第(2)问中求通解比较简单，只需要利用给定的两个解求出未知量c，继而可得到通解。

证明：(1) 根据教材P74定理1.8.3可知，要证明方程无解，只需证明方程

$$\begin{bmatrix} 1 & a_1 & a_1^2 \\ 1 & a_2 & a_2^2 \\ 1 & a_3 & a_3^2 \\ 1 & a_4 & a_4^2 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} a_1^3 \\ a_2^3 \\ a_3^3 \\ a_4^3 \end{bmatrix}$$

是非一致方程，其中$\boldsymbol{x} = [x_1, x_2, x_3]^{\mathrm{T}}$。经过观察可以发现，该方程组的系数矩阵和相应的增广矩阵都是Vandermonde矩阵，其最为突出的性质就是当$a_1, a_2, \cdots, a_4$各异时，矩阵为满列秩矩阵（详细内容见教材第2章）。这就表明，系数矩阵的秩为3，而增广矩阵的秩为4。从而根据教材上的定理1.8.4，得到原方程组不是一致方程的结论，也就说明线性方程组无解。

(2) 若$a_1 = a_3 = c, \quad a_2 = a_4 = -c$，则上述方程可简写为

$$\begin{bmatrix} 1 & c & c^2 \\ 1 & -c & c^2 \\ 1 & c & c^2 \\ 1 & -c & c^2 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} c^3 \\ -c^3 \\ c^3 \\ -c^3 \end{bmatrix}$$

将$\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$两个解代入到上述方程中得到

$$c^2 = c^3 \quad \Rightarrow \quad c_1 = 0,\ c_2 = 1$$

因为已知要求$c \neq 0$，所以$c = 1$。再带回到方程组中，得到通解为

$$\boldsymbol{x} = \begin{bmatrix} -a \\ 1 \\ a \end{bmatrix}$$

其中，a为任意实数。

1.35 当α取何值时，线性方程组

$$\begin{aligned} (\alpha+3)x_1 + x_2 + 2x_3 &= \alpha \\ 3(\alpha+1)x_1 + \alpha x_2 + (\alpha+3)x_3 &= 3 \\ \alpha x_1 + (\alpha-1)x_2 + x_3 &= \alpha \end{aligned}$$

有唯一解、无解和无穷多解。当方程组有无穷多解时，求出它的通解。

思路分析：此题考察的仍然是有关一致方程和线性方程组求解方面的问题，按照前面几题的思路解答即可。

解：增广矩阵经过消去后可得

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

(1) 方程组有唯一解，要求系数矩阵可逆，即

$$[\alpha-(2-\alpha)(\alpha+1)]\left(1-\frac{\alpha}{3}\right) - 2\left[\alpha-1-\frac{\alpha(2-\alpha)}{3}\right] \neq 0$$

由此推出，$\alpha \neq 0$且$\alpha \neq 1$。

(2) 方程组无解，则该方程为非一致方程，要求

$$[\alpha-(2-\alpha)(\alpha+1)]\left(1-\frac{\alpha}{3}\right) - 2\left[\alpha-1-\frac{\alpha(2-\alpha)}{3}\right] = 0$$

且系数矩阵后两行所成的比例与向量$[0,3,\alpha]^{\mathrm{T}}$后两个元素所成的比例不等。从而得到$\alpha = 0$。

(3) 方程组有无穷多解，要求

$$[\alpha-(2-\alpha)(\alpha+1)]\left(1-\frac{\alpha}{3}\right) - 2\left[\alpha-1-\frac{\alpha(2-\alpha)}{3}\right] = 0$$

且方程为一致方程，从而$\alpha = 1$。此时，方程组的通解为$x_2 = 2x_3 - 3$，$x_1 = 1 - x_3$。

1.36 当α和β取何值时，线性方程组

$$x_1 + 3x_2 + 6x_3 + x_4 = 3$$
$$x_1 + x_2 + 2x_3 + 3x_4 = 1$$
$$x_1 - 5x_2 - 10x_3 + 12x_4 = \alpha$$
$$3x_1 - x_2 - \beta x_3 + 15x_4 = 3$$

有唯一解、无解和无穷多解。当方程组有无穷多解时，求出它的通解。

思路分析：思路同上题。

解：将方程组表示为$\boldsymbol{Ax} = \boldsymbol{b}$。对增广矩阵$[\boldsymbol{A}, \boldsymbol{b}]$做初等行变换后得到

$$\begin{bmatrix} 1 & 3 & 6 & 1 & 3 \\ 0 & -2 & -4 & 2 & -2 \\ 0 & 0 & 0 & 3 & \alpha+5 \\ 0 & 0 & -\beta+2 & 2 & 4 \end{bmatrix}$$

(1) 若$\beta \neq 2$，则$\mathrm{rank}(\boldsymbol{A}) = \mathrm{rank}([\boldsymbol{A}, \boldsymbol{b}]) = 4$ 上述方程组有唯一解。

(2) 若$\beta = 2$且$\alpha \neq 1$，则$\mathrm{rank}(\boldsymbol{A}) = 3, \mathrm{rank}([\boldsymbol{A}, \boldsymbol{b}]) = 4$ 上述方程组无解。

(3) 若$\beta = 2$且$\alpha = 1$，则$\mathrm{rank}(\boldsymbol{A}) = \mathrm{rank}([\boldsymbol{A}, \boldsymbol{b}]) = 3$ 上述方程有无穷多解。其特解为$\boldsymbol{x}_0 = [-8, 3, 0, 2]^{\mathrm{T}}$，导出组的基础解系只由一个向量构成，即$\boldsymbol{x}_1 = [0, -2, 1, 0]^{\mathrm{T}}$。所以，通解可表示为$\boldsymbol{x} = \boldsymbol{x}_0 + c\boldsymbol{x}_1$，其中$c$为任意实数。

1.37 [23] 已知矩阵方程$\boldsymbol{Ax} = \boldsymbol{b}$为

$$x_1 + 2x_2 + 3x_3 = 26$$
$$3x_1 + 7x_2 + 10x_3 = 87$$
$$2x_1 + 11x_2 + 7x_3 = 73$$

(1) 利用高斯消去法求解方程。

(2) 将矩阵$\boldsymbol{A}$的第j列用$\boldsymbol{b}$ 代替，并记所得矩阵为$\boldsymbol{A}_j$。证明(1)中求出的方程的解可以表示为

$$x_j = |\boldsymbol{A}_j| / |\boldsymbol{A}|, \quad j = 1, 2, 3$$

这一方法称为求解线性方程的Cramer法则。

(3) 证明对$\boldsymbol{A}_{n\times n}\boldsymbol{x}_{n\times 1} = \boldsymbol{b}_{n\times 1}$的一般情况，若$|\boldsymbol{A}| \neq 0$，则由Cramer法则求出的解$\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^{\mathrm{T}}$ 确实满足线性方程$\boldsymbol{Ax} = \boldsymbol{b}$。

思路分析：(1)的求解非常简单。(2)(3)涉及到Cramer法则，有必要对其做一回顾。由于Cramer法则在求解线性方程组时非常繁琐，因此在实际中很少使用，但其具有重要的理论意义。

解：(1) 对增广矩阵$[\boldsymbol{A}, \boldsymbol{b}]$进行初等行变换，得到

$$\begin{bmatrix} 1 & 2 & 3 & 26 \\ 0 & 1 & 1 & 9 \\ 0 & 0 & -6 & -42 \end{bmatrix}$$

解得

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 7 \end{bmatrix}$$

(2) 分别求各$\boldsymbol{A}_j$和$\boldsymbol{A}$的行列式，得$|\boldsymbol{A}_1| = -6$，$|\boldsymbol{A}_2| = -12$，$|\boldsymbol{A}_3| = -42$，$|\boldsymbol{A}| = -6$

$$\Rightarrow x_1 = 1, \quad x_2 = 2, \quad x_3 = 7$$

与上一问所得结果完全相同!

(3) 现在来证明Cramer法则。Cramer法则描述如下：如果线性方程组

$$\begin{aligned} a_{11}\,x_1 + a_{12}\,x_2 + \cdots + a_{1n}\,x_n &= b_1 \\ a_{21}\,x_1 + a_{22}\,x_2 + \cdots + a_{2n}\,x_n &= b_2 \\ \vdots \qquad\quad \vdots \qquad\quad & \vdots \\ a_{n1}\,x_1 + a_{n2}\,x_2 + \cdots + a_{nn}\,x_n &= b_n \end{aligned}$$

$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$系数行列式

$$|\boldsymbol{A}| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \neq 0$$

则上述线性方程组有唯一解

$$x_j = \frac{|\boldsymbol{A}_j|}{|\boldsymbol{A}|}, \quad j = 1, 2, \cdots, n$$

其中，$\boldsymbol{A}_j$是用$\boldsymbol{b}$向量替换$\boldsymbol{A}$的第j列得到的新的矩阵，即

$$\boldsymbol{A}_j = \begin{bmatrix} a_{11} & \cdots & a_{1j-1} & b_1 & a_{1j+1} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2j-1} & b_2 & a_{2j+1} & \cdots & a_{2n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nj-1} & b_n & a_{nj+1} & \cdots & a_{nn} \end{bmatrix}, \qquad j = 1, 2, \cdots, n$$

证明[49]：设$\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^{\mathrm{T}}$是方程组的解，则

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \qquad i = 1, 2, \cdots, n$$

设$\boldsymbol{A}_{ik}$表示矩阵$\boldsymbol{A}$第i行第k列元素的代数余子式，$\boldsymbol{A}_k$表示矩阵$\boldsymbol{A}$中第k列被向量$\boldsymbol{b}$替换后的结果

$$\sum_{i=1}^{n} \boldsymbol{A}_{ik} \sum_{j=1}^{n} a_{ij} x_j = \sum_{i=1}^{n} \boldsymbol{A}_{ik} b_i = |\boldsymbol{A}_k|$$

上式中第二个等号成立是由于行列式可以按列展开来求解。另外，

$$\sum_{i=1}^{n}\boldsymbol{A}_{ik}\sum_{j=1}^{n}a_{ij}x_j=\sum_{j=1}^{n}\sum_{i=1}^{n}a_{ij}\boldsymbol{A}_{ik}x_j$$
$$=\sum_{j=1,j\neq k}^{n}\sum_{i=1}^{n}a_{ij}\boldsymbol{A}_{ik}x_j+\sum_{i=1}^{n}a_{ik}\boldsymbol{A}_{ik}x_k$$

其中，

$$\sum_{j=1,j\neq k}^{n}\sum_{i=1}^{n}a_{ij}\boldsymbol{A}_{ik}=0$$

这是因为矩阵任一列的所有元素与另外任意一列相应元素的代数余子式的乘积之和为零。综合上两式，就有

$$|\boldsymbol{A}_k|=\sum_{i=1}^{n}a_{ij}\boldsymbol{A}_{ik}x_k=|\boldsymbol{A}|x_k$$

从而$x_k=|\boldsymbol{A}_k|/|\boldsymbol{A}|$。

1.38 假定$\boldsymbol{A}$和$\boldsymbol{B}$都是$n\times n$ 矩阵，并且$\boldsymbol{A}$非奇异。从线性方程组的角度证明：若$\boldsymbol{AB}=\boldsymbol{O}$ (零矩阵)，则$\boldsymbol{B}=\boldsymbol{O}$。

证明：将$\boldsymbol{B}$表示成按列分块形式，即$\boldsymbol{B}=[\boldsymbol{b}_1,\boldsymbol{b}_2,\cdots,\boldsymbol{b}_n]$，$\forall\ \boldsymbol{b}_i,i\in\{1,2,\cdots,n\}$，依题意有$\boldsymbol{A}\boldsymbol{b}_i=\mathbf{0}$。又因为已知$\boldsymbol{A}$可逆，所以$\boldsymbol{b}_i=\boldsymbol{A}^{-1}\mathbf{0}=\mathbf{0}$。因此，必然有$\boldsymbol{B}=\boldsymbol{O}$。

1.39 设向量组

$$\boldsymbol{a}_1=[1,1,1,3]^{\mathrm{T}},\qquad \boldsymbol{a}_2=[-1,-3,5,1]^{\mathrm{T}}$$
$$\boldsymbol{a}_3=[3,2,-1,p+2]^{\mathrm{T}},\qquad \boldsymbol{a}_4=[-2,-6,10,p]^{\mathrm{T}}$$

(1) p为何值时，此向量组线性无关？用$\boldsymbol{a}_1\sim\boldsymbol{a}_4$的线性组合表示$\boldsymbol{a}=[4,1,6,10]^{\mathrm{T}}$。

(2) p取何值时，该向量组线性相关？求出此时矩阵$[\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3,\boldsymbol{a}_4]$的秩和一个极大线性无关的向量组。

思路分析：向量组线性无关，等价于由这些向量构成的矩阵满秩。因此，这个问题可以转化为通过高斯消去法来检验这些向量构成的矩阵是否满秩，并在此基础上寻找极大线性无关向量组。

证明：(1)

$$\boldsymbol{A}=[\boldsymbol{a}_1\quad\boldsymbol{a}_2\quad\boldsymbol{a}_3\quad\boldsymbol{a}_4]=\begin{bmatrix}1&-1&3&-2\\1&-3&2&-6\\1&5&-1&10\\3&1&p+2&p\end{bmatrix}$$

对其进行初等行变换后得到

$$\begin{bmatrix} 1 & -1 & 3 & -2 \\ 0 & -2 & -1 & -4 \\ 0 & 0 & -7 & 0 \\ 0 & 0 & p-9 & p-2 \end{bmatrix}$$

若$p \neq 2$，则上述矩阵满秩，此时向量组$\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \boldsymbol{a}_4$线性无关。用$\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \boldsymbol{a}_4$的线性组合表示$\boldsymbol{a}$，就相当于求$\boldsymbol{x}$,使得$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{a}$.利用高斯消去法对增广矩阵进行行变换得到

$$[\boldsymbol{A}, \boldsymbol{a}] = \begin{bmatrix} 1 & -1 & 3 & -2 & 4 \\ 0 & -2 & -1 & -4 & -3 \\ 0 & 0 & -7 & 0 & -7 \\ 0 & 0 & p-9 & p-2 & -8 \end{bmatrix}$$

当$p = 2$时，显然方程组无解；当$p \neq 2$时，$\boldsymbol{x} = [2, (3p-4)/(p-2), 1, (1-p)/(p-2)]^{\mathrm{T}}$。

(2) $p = 2$时，向量组线性相关，$\mathrm{rank}(\boldsymbol{A}) = 3$。从(1)的行变换结果可以看出：$\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$就构成了一个极大线性无关向量组。

1.40 已知向量组

$$\boldsymbol{a}_1 = \begin{bmatrix} a \\ 2 \\ 10 \end{bmatrix}, \quad \boldsymbol{a}_2 = \begin{bmatrix} -2 \\ 1 \\ 5 \end{bmatrix}, \quad \boldsymbol{a}_3 = \begin{bmatrix} -1 \\ 1 \\ 4 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ b \\ c \end{bmatrix}$$

试分别求出满足以下条件的a, b, c值：

(1) $\boldsymbol{b}$可由$\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ 线性表示，且唯一。

(2) $\boldsymbol{b}$不能由$\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ 线性表示。

(3) $\boldsymbol{b}$可由$\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ 线性表示，但表示不唯一。并求出一般表达式。

解：令$\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3]$

(1) 该问题等价于求a, b, c为何值时，$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$有唯一解。

对增广矩阵进行初等行变换，得到

$$\begin{bmatrix} a & -2 & -1 & 1 \\ 2+a & -1 & 0 & b+1 \\ 4+a & 0 & 0 & c-3b+1 \end{bmatrix}$$

可见，当$a \neq -4$时，方程组有唯一解

$$\boldsymbol{x} = \left[\frac{c-3b+1}{4+a}, \frac{2c+ac-10b-4ab-2}{4+a}, \frac{-ac+5ab-4c+20b}{4+a}\right]^{\mathrm{T}}$$

(2) 当$a = -4$且$c \neq 3b-1$时，方程组为非一致方程，$\boldsymbol{b}$不能由$\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ 线性表示。

(3) 当$a = -4$且$c = 3b-1$时，可得无穷多解，其中特解为$\boldsymbol{x}_0 = [0, -b-1, 2b+1]^{\mathrm{T}}$，导出组的基础解系只由一个向量构成，即$\boldsymbol{x}_1 = [1, -2, 0]^{\mathrm{T}}$。所以，通解可表示成$\boldsymbol{x} = \boldsymbol{x}_0 + k\boldsymbol{x}_1, \forall\, k \in R$。

1.41 令矩阵$\boldsymbol{A}_{m\times n}$的秩$r=\operatorname{rank}(\boldsymbol{A})$。证明：若$\boldsymbol{A}$分块为

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}$$

式中，$r\times r$矩阵$\boldsymbol{A}_{11}$ 是一个秩为r的非奇异矩阵，则$\boldsymbol{A}_{22}=\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}$。（提示：注意$[\boldsymbol{A}_{21},\boldsymbol{A}_{22}]$的行是$[\boldsymbol{A}_{11},\boldsymbol{A}_{12}]$的$r$行的线性组合，即有$[\boldsymbol{A}_{21},\boldsymbol{A}_{22}]=\boldsymbol{F}[\boldsymbol{A}_{11},\boldsymbol{A}_{12}]$。类似地，矩阵$\boldsymbol{A}$ 的右边$n-r$列是左边r列的线性组合。）

思路分析： 本题目设计比较巧妙。任意非满秩矩阵，表示成分块矩阵形式后，看似各部分没有任何关系，但实际上却有如上所示的等式存在。证明较为容易，直接根据提示做即可。

证明： 由题意可知，矩阵$\boldsymbol{A}$的后$m-r$行可以表示成前r行的线性组合。即$\exists\boldsymbol{F}$，使得

$$\boldsymbol{A}_{21}=\boldsymbol{F}\boldsymbol{A}_{11},\qquad \boldsymbol{A}_{22}=\boldsymbol{F}\boldsymbol{A}_{12}$$

因为$\boldsymbol{A}_{11}$非奇异，所以

$$\boldsymbol{A}_{22}=\boldsymbol{F}\boldsymbol{A}_{12}=\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}$$

1.42 令两个向量相互正交，证明它们线性无关。

思路分析： 本题考察的是向量正交与线性无关之间的关系。向量正交的物理意义非常明确，$\boldsymbol{a}$与$\boldsymbol{b}$正交，即向量$\boldsymbol{a}$到向量$\boldsymbol{b}$的投影为零。这表明向量$\boldsymbol{a}$中不包含向量$\boldsymbol{b}$的任何成分；反之亦然。而线性相关性说的是某个向量能不能被另一些向量线性表出。显然，向量正交比向量线性无关的条件更强。下面来证明向量正交是向量线性无关的充分条件。

证明： 设$\boldsymbol{a},\boldsymbol{b}$为两个非零的正交向量，即$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}=0$。我们采用反证法。假设$\boldsymbol{a},\boldsymbol{b}$线性相关，则一定存在非零标量$k$，使得$\boldsymbol{a}=k\boldsymbol{b}$。此时，$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}=k\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}=k\|\boldsymbol{b}\|^2\neq 0$。因此，假设错误。两个向量相互正交，则它们一定线性无关。

1.43 矩阵$\boldsymbol{A}^2=\boldsymbol{A}$和$\boldsymbol{B}^2=\boldsymbol{B}$，并且$\boldsymbol{B}$的列是$\boldsymbol{A}$的列的线性组合。证明$\boldsymbol{A}\boldsymbol{B}=\boldsymbol{B}$。

证明： 因为$\boldsymbol{B}$的各列是$\boldsymbol{A}$的线性组合，所以必存在矩阵$\boldsymbol{Q}$，使得$\boldsymbol{B}=\boldsymbol{A}\boldsymbol{Q}$。在该式两边同时左乘$\boldsymbol{A}$，得到

$$\boldsymbol{A}\boldsymbol{B}=\boldsymbol{A}^2\boldsymbol{Q}=\boldsymbol{A}\boldsymbol{Q}=\boldsymbol{B}$$

上式利用了$\boldsymbol{A}^2=\boldsymbol{A}$的条件，证毕。

1.44 验证向量组

$$\boldsymbol{A}=\left\{\begin{bmatrix}1\\0\\1\\2\end{bmatrix},\begin{bmatrix}-1\\1\\1\\0\end{bmatrix},\begin{bmatrix}-1\\-2\\1\\0\end{bmatrix}\right\}$$

是一组正交向量。

解：对上述三个向量两两求内积，发现都得零，因此这是一组正交向量。

1.45 令$B=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3\}$是Euclidean空间$R^3$的一组正交基。给定向量$\boldsymbol{u}\in R^3$，试确定常数$a_1,a_2,a_3$使得

$$\boldsymbol{u}=a_1\boldsymbol{v}_1+a_2\boldsymbol{v}_2+a_3\boldsymbol{v}_3$$

思路分析：用一组基向量表示线性空间中的某个向量，实际上就是线性方程组的求解问题。

解法一：将该问题看作线性方程组的求解。设$\boldsymbol{a}=[a_1,a_2,a_3]^{\mathrm{T}}, \boldsymbol{V}=[\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3]$，则有$\boldsymbol{Va}=\boldsymbol{u}$。所以，$\boldsymbol{a}=\boldsymbol{V}^{-1}\boldsymbol{u}$，$\boldsymbol{V}$的非奇异性由向量组$\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3$的线性无关性来保证。

解法二：上述方法只利用了基向量的线性无关性，而并没有用到题目中所给出的更强的正交性。更为简单的解法是依次用$\boldsymbol{v}_1^{\mathrm{T}},\boldsymbol{v}_2^{\mathrm{T}},\boldsymbol{v}_3^{\mathrm{T}}$,左乘$\boldsymbol{u}$，再除以各自模的平方，则所得的结果就是$a_1,a_2,a_3$。以$a_1$的求解过程为例。$\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{u}=a_1\|\boldsymbol{v}_1\|^2+a_2\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_2+a_3\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_3=a_1\|\boldsymbol{v}_1\|^2$（利用了向量组的正交性），则$a_1=\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{u}/\|\boldsymbol{v}_1\|^2$。

1.46 使用Gram-Schmidt正交化方法构造子空间$V=\mathrm{Span}\{\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3\}$的正交基和标准正交基，其中

$$\boldsymbol{v}_1=\begin{bmatrix}0\\2\\1\\1\end{bmatrix},\quad \boldsymbol{v}_2=\begin{bmatrix}0\\3\\1\\1\end{bmatrix},\quad \boldsymbol{v}_3=\begin{bmatrix}1\\1\\1\\0\end{bmatrix}$$

思路分析：Gram-Schmidt正交化是矩阵分解、变换中的一种重要方法，应用范围非常广。教材1.5节给出了基本方法和数值稳定性更好的修正方法。直接套用，即可求解。

解：利用基本的Gram-Schmidt正交化方法，正交基的求解如下：

$$\boldsymbol{p}_1=\boldsymbol{v}_1=[0,2,1,1]^{\mathrm{T}}$$

$$\boldsymbol{p}_2=\boldsymbol{v}_2-\frac{\boldsymbol{p}_1^{\mathrm{T}}\boldsymbol{v}_2}{\boldsymbol{p}_1^{\mathrm{T}}\boldsymbol{p}_1}\boldsymbol{p}_1=\left[0,\frac{1}{3},-\frac{1}{3},-\frac{1}{3}\right]^{\mathrm{T}}$$

$$\boldsymbol{p}_3=\boldsymbol{v}_3-\frac{\boldsymbol{p}_1^{\mathrm{T}}\boldsymbol{v}_3}{\boldsymbol{p}_1^{\mathrm{T}}\boldsymbol{p}_1}\boldsymbol{p}_1-\frac{\boldsymbol{p}_2^{\mathrm{T}}\boldsymbol{v}_3}{\boldsymbol{p}_2^{\mathrm{T}}\boldsymbol{p}_2}\boldsymbol{p}_2=\left[1,0,\frac{1}{2},-\frac{1}{2}\right]^{\mathrm{T}}$$

标准正交基为

$$\boldsymbol{\omega}_1=\frac{\boldsymbol{p}_1}{\|\boldsymbol{p}_1\|}=\frac{1}{\sqrt{6}}[0,2,1,1]^{\mathrm{T}}$$

$$\boldsymbol{\omega}_2=\frac{\boldsymbol{p}_2}{\|\boldsymbol{p}_2\|}=\sqrt{3}\left[0,\frac{1}{3},-\frac{1}{3},-\frac{1}{3}\right]^{\mathrm{T}}$$

$$\boldsymbol{\omega}_3=\frac{\boldsymbol{p}_3}{\|\boldsymbol{p}_3\|}=\sqrt{\frac{2}{3}}\left[\frac{1}{2},0,\frac{1}{2},-\frac{1}{2}\right]^{\mathrm{T}}$$

1.47 已知3×5矩阵

$$\boldsymbol{A}=\begin{bmatrix}1&3&2&5&7\\2&1&0&6&1\\1&1&2&5&4\end{bmatrix}$$

用MATLAB函数orth(A)和null(A)分别求矩阵$\boldsymbol{A}$ 的列空间Span($\boldsymbol{A}$)的正交基和矩阵$\boldsymbol{A}$的零空间Null($\boldsymbol{A}$)。

解： 直接应用MATLAB中的函数$\boldsymbol{B}=\mathrm{orth}(\boldsymbol{A})$可得到$\boldsymbol{A}$的列空间的一组标准正交基。满足$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}=\boldsymbol{I}$,且$\boldsymbol{B}$的列数等于$\boldsymbol{A}$的秩。

$$\boldsymbol{B}=\mathrm{orth}(\boldsymbol{A})=\begin{bmatrix}-0.7233&0.5296&-0.4431\\-0.4318&-0.8476&-0.3083\\-0.5389&-0.0316&0.8418\end{bmatrix}$$

$\boldsymbol{C}=\mathrm{Null}(\boldsymbol{A})$得到$\boldsymbol{A}$的零空间的一组标准正交基

$$\boldsymbol{C}=\mathrm{null}(\boldsymbol{A})=\begin{bmatrix}-0.8758&-0.2921\\0.3299&-0.5715\\0.0288&-0.6529\\0.2736&0.1291\\-0.2199&0.3810\end{bmatrix}$$

1.48 令$\boldsymbol{x}$和$\boldsymbol{y}$是Euclidean n空间R^n的任意两个向量，请证明Cauchy-Schwartz不等式$|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}|\leqslant\|\boldsymbol{x}\|\|\boldsymbol{y}\|$。（提示：观察$\|\boldsymbol{x}-c\boldsymbol{y}\|^2\geqslant 0$对所有标量$c$成立。）

思路分析： Cauchy-Schwartz不等式证明的经典方法就是通过构造二次不等式，利用判别式小于等于零的性质加以证明。

证明： 构造关于实数c的函数$f(c)=\|\boldsymbol{x}-c\boldsymbol{y}\|^2=\|\boldsymbol{x}\|^2+c^2\|\boldsymbol{y}\|^2-2c\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$。显然，$f(c)\geqslant 0$。这表明以$c$为自变量，以$f(c)$为函数的曲线与横轴（$c$轴）没有交点或只有一个交点。因此要求判别式小于等于零，即

$$\left(-2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}\right)^2-4\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2\leqslant 0$$

从而有$|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}|\leqslant\|\boldsymbol{x}\|\|\boldsymbol{y}\|$,当且仅当$\boldsymbol{x}=c\boldsymbol{y}$时等号成立。

1.49 令$\boldsymbol{x}$和$\boldsymbol{y}$是Euclidean n 空间R^n的任意两个向量，证明三角不等式$\|\boldsymbol{x}+\boldsymbol{y}\|\leqslant\|\boldsymbol{x}\|+\|\boldsymbol{y}\|$。（提示：展开$\|\boldsymbol{x}+\boldsymbol{y}\|^2$，并利用Cauchy-Schwartz不等式。）

思路分析： 本题目的证明是上一题Cauchy-Schwartz不等式的直接应用。

证明：

$$\begin{aligned}\|\boldsymbol{x}+\boldsymbol{y}\|^2&=\|\boldsymbol{x}\|^2+\|\boldsymbol{y}\|^2+2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}\\&\leqslant\|\boldsymbol{x}\|^2+\|\boldsymbol{y}\|^2+2\|\boldsymbol{x}\|\|\boldsymbol{y}\|\\&=(\|\boldsymbol{x}\|+\|\boldsymbol{y}\|)^2\end{aligned}$$

所以

$$\|\boldsymbol{x}+\boldsymbol{y}\| \leqslant \|\boldsymbol{x}\|+\|\boldsymbol{y}\|$$

注意到，若$\boldsymbol{x}$和$\boldsymbol{y}$均为二维向量，这个不等式实际上就是众所周知的三角形两边之和大于第三边。

1.50 令$B=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n\}$是子空间$W$的标准正交基，且$\boldsymbol{u}$是子空间$W$内的向量。证明：若$\boldsymbol{u}=a_1\boldsymbol{v}_1+a_2\boldsymbol{v}_2+\cdots+a_n\boldsymbol{v}_n$，则

$$\|\boldsymbol{u}\|^2=|a_1|^2+|a_2|^2+\cdots+|a_n|^2$$

思路分析：此题考察的是正交基向量的基本概念。应用正交的性质，可直接证明。

证明：

$$\|\boldsymbol{u}\|^2=\left(\sum_{i=1}^{n}a_i\boldsymbol{v}_i\right)^{\mathrm{H}}\left(\sum_{j=1}^{n}a_j\boldsymbol{v}_j\right)=\sum_{i=1}^{n}|a_i|^2\|\boldsymbol{v}_i\|^2+\sum_{i\neq j,\ i,j=1}^{n}a_ia_j\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_j$$

因为$\boldsymbol{v}_i$是一组标准正交基，所以

$$\boldsymbol{v}_i^{\mathrm{H}}\boldsymbol{v}_j=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases}$$

代入到前面的结果中，直接得到

$$\|\boldsymbol{u}\|^2=\sum_{i=1}^{n}|a_i|^2$$

1.51 证明任意一组由R^3中的4个或更多个向量的集合不可能组成R^3的正交基。

证明：我们知道，R^n中任意$n+1$个向量必定线性相关，所以R^3中的4个或更多个向量必定是线性相关的，而线性相关的向量组不能构成一组正交基。

1.52 定义变换$H:R^2\mapsto R^2$为

$$H(\boldsymbol{x})=\begin{bmatrix}x_1+x_2-1\\ 3x_1\end{bmatrix},\qquad \boldsymbol{x}=\begin{bmatrix}x_1\\ x_2\end{bmatrix}$$

判断H是否为线性变换。

思路分析：判断一个变换是否为线性变换，只需按照线性变换的定义逐条验证即可。

解：$\forall$向量$\boldsymbol{x}=[x_1,x_2]^{\mathrm{T}}$，$\boldsymbol{y}=[y_1,y_2]^{\mathrm{T}}$

$$H(\boldsymbol{x}+\boldsymbol{y})=\begin{bmatrix}x_1+y_1+x_2+y_2-1\\ 3(x_1+y_1)\end{bmatrix}\neq H(\boldsymbol{x})+H(\boldsymbol{y})$$

$$H(k\boldsymbol{x})=\begin{bmatrix}kx_1+kx_2-1\\ 3kx_1\end{bmatrix}\neq kH(\boldsymbol{x}),\qquad \forall\ k\in R$$

因而，H不满足线性变换的定义，不是线性变换。

1.53 令

$$\boldsymbol{P}=\begin{bmatrix}0.5&0.2&0.3\\0.3&0.8&0.3\\0.2&0&0.4\end{bmatrix},\qquad \boldsymbol{x}_0=\begin{bmatrix}1\\0\\0\end{bmatrix}$$

假定一系统的状态向量可以用Markov链$\boldsymbol{x}_{k+1}=\boldsymbol{P}\boldsymbol{x}_k$, $k=0,1,\cdots$描述，试计算状态向量$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_{15}$，分析系统随时间的变化。

思路分析：本题结合了Markov链的背景。在一个Markov链中，任一时刻的状态向量完全由一步状态转移矩阵与初始状态向量确定。

解：依题意可得

$$\boldsymbol{x}_1=\boldsymbol{P}\boldsymbol{x}_0=\begin{bmatrix}0.5\\0.3\\0.2\end{bmatrix},\quad \boldsymbol{x}_2=\boldsymbol{P}^2\boldsymbol{x}_0=\begin{bmatrix}0.37\\0.45\\0.18\end{bmatrix},\quad\cdots,\quad \boldsymbol{x}_{15}=\boldsymbol{P}^{15}\boldsymbol{x}_0=\begin{bmatrix}0.3\\0.6\\0.1\end{bmatrix}$$

进一步地，根据$\boldsymbol{P}$矩阵的结构及Markov链的性质，可以判断该系统必将收敛到稳态。可以观察到，从$\boldsymbol{x}_{15}$向后，$\boldsymbol{x}_{16},\boldsymbol{x}_{17},\cdots$系统状态不再改变，都为$\boldsymbol{x}=[0.3,0.6,0.1]^{\mathrm{T}}$，并且，还可得到

$$\lim_{n\to\infty}\boldsymbol{P}^n=\begin{bmatrix}0.3&0.3&0.3\\0.6&0.6&0.6\\0.1&0.1&0.1\end{bmatrix}$$

1.54 已知矩阵函数

$$\boldsymbol{A}(x)=\begin{bmatrix}2x&-1&x&2\\4&x&1&-1\\3&2&x&5\\1&-2&3&x\end{bmatrix}$$

求$\dfrac{\mathrm{d}^3|\boldsymbol{A}(x)|}{\mathrm{d}x^3}$。（提示：按任意行或列展开行列式$|\boldsymbol{A}(x)|$，并且只需要关心$x^4$和$x^3$项。）

思路分析：该4阶行列式展开后，x的最高次幂为4。最终要求的是关于x的三阶导数，因此只需获得x^3和x^4的系数即可。

解：将上述行列式按照第一行展开，得到

$$\begin{aligned}|\boldsymbol{A}(x)|&=2x\begin{vmatrix}x&1&-1\\2&x&5\\-2&3&x\end{vmatrix}+\begin{vmatrix}4&1&-1\\3&x&5\\1&3&x\end{vmatrix}+x\begin{vmatrix}4&x&-1\\3&2&5\\1&-2&x\end{vmatrix}\\&\quad-2\begin{vmatrix}4&x&1\\3&2&x\\1&-2&3\end{vmatrix}\\&=a_1x^4+a_2x^3+a_3x^2+a_4x+a_5\end{aligned}$$

化简后，可得到$a_1=2,a_2=-3$，从而$\dfrac{\mathrm{d}^3|\boldsymbol{A}(x)|}{\mathrm{d}x^3}=48x-18$。

1.55 证明：若$\boldsymbol{A}_1$非奇异，则

$$\begin{vmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 \\ \boldsymbol{A}_3 & \boldsymbol{A}_4 \end{vmatrix} = |\boldsymbol{A}_1||\boldsymbol{A}_4 - \boldsymbol{A}_3\boldsymbol{A}_1^{-1}\boldsymbol{A}_2|$$

思路分析：要求证明分块矩阵行列式的等式关系，可通过将原矩阵表示成几个矩阵乘积的形式，从而将原行列式转化成几个简单行列式的乘积。

证明：因为

$$\begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 \\ \boldsymbol{A}_3 & \boldsymbol{A}_4 \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{O} \\ \boldsymbol{A}_3 & \boldsymbol{A}_4 - \boldsymbol{A}_3\boldsymbol{A}_1^{-1}\boldsymbol{A}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{A}_1^{-1}\boldsymbol{A}_2 \\ \boldsymbol{O} & \boldsymbol{I} \end{bmatrix}$$

所以

$$\begin{vmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 \\ \boldsymbol{A}_3 & \boldsymbol{A}_4 \end{vmatrix} = |\boldsymbol{A}_1||\boldsymbol{A}_4 - \boldsymbol{A}_3\boldsymbol{A}_1^{-1}\boldsymbol{A}_2||\boldsymbol{I}| = |\boldsymbol{A}_1||\boldsymbol{A}_4 - \boldsymbol{A}_3\boldsymbol{A}_1^{-1}\boldsymbol{A}_2|$$

1.56 求对称矩阵$\boldsymbol{A}$和非对称矩阵$\boldsymbol{B}$，使得下面各二次型可以写成$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$和$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x}$：

(1) $7x_1^2 + 14x_1x_2 + 5x_2^2$。

(2) $(x_1 - 2x_2)^2 + (3x_2 - x_3)^2 + (6x_1 - 4x_3)^2$。

(3) $2(x_1^2 + x_2^2 + x_3^2) - 2(x_1x_2 + x_1x_3 + x_2x_3)$。

(4) $a_1x_1^2 + a_2x_2^2 + a_3x_3^2 + b_1x_1x_2 + b_2x_1x_3 + b_3x_2x_3$。

思路分析：该题目实际上是希望读者注意到，任何一个二次型函数对应多个矩阵$\boldsymbol{A}$，使得$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$都可以表示原二次型。但其中，对称的矩阵$\boldsymbol{A}$却只有一个。

解：不加说明，我们直接给出答案。注意，$\boldsymbol{A}$矩阵是唯一的，而$\boldsymbol{B}$矩阵却不唯一。

(1)

$$\boldsymbol{A} = \begin{bmatrix} 7 & 7 \\ 7 & 5 \end{bmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} 7 & 4 \\ 10 & 5 \end{bmatrix}$$

(2)

$$\boldsymbol{A} = \begin{bmatrix} 37 & -2 & -24 \\ -2 & 13 & -3 \\ -24 & -3 & 17 \end{bmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} 37 & -1 & -24 \\ -3 & 13 & -2 \\ -24 & -4 & 17 \end{bmatrix}$$

(3)

$$\boldsymbol{A} = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} 2 & 0 & -1 \\ -2 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

(4)

$$\boldsymbol{A} = \begin{bmatrix} a_1 & \dfrac{b_1}{2} & \dfrac{b_2}{2} \\ \dfrac{b_1}{2} & a_2 & \dfrac{b_3}{2} \\ \dfrac{b_2}{2} & \dfrac{b_3}{2} & a_3 \end{bmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} a_1 & b_1 & b_2 \\ 0 & a_2 & b_3 \\ 0 & 0 & a_3 \end{bmatrix}$$

1.57 判断下列二次型的正定性：

(1) $f=-2x_1^2-8x_2^2-6x_3^2+2x_1x_2+2x_1x_3$。

(2) $f=x_1^2+4x_2^2+9x_3^2+15x_4^2-2x_1x_2+4x_1x_3+2x_1x_4-6x_2x_4-12x_3x_4$。

思路分析：判断二次型函数的正定性，可以直接考察函数本身，也可以转而考察相应的对称矩阵的正定性。这里我们采用后一种方法求解。

解：(1) 将代数式写成矩阵方程的形式

$$f=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=[x_1,x_2,x_3]\begin{bmatrix}-2&1&1\\1&-8&0\\1&0&-6\end{bmatrix}[x_1,x_2,x_3]^{\mathrm{T}}$$

经检验，发现$\boldsymbol{A}$矩阵的各阶顺序主子式负正相间，因而$\boldsymbol{A}$为负定矩阵。即该二次型为负定二次型。

(2) 方法同上

$$f=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=[x_1,x_2,x_3,x_4]\begin{bmatrix}1&-1&2&1\\-1&4&0&-3\\2&0&9&-6\\1&-3&-6&15\end{bmatrix}[x_1,x_2,x_3,x_4]^{\mathrm{T}}$$

经检验，发现$\boldsymbol{A}$ 矩阵的所有顺序主子式都大于零，所以$\boldsymbol{A}$为正定矩阵。即该二次型为正定二次型。

1.58 证明：

(1) 若$\boldsymbol{B}$为实的非奇异矩阵，则$\boldsymbol{A}=\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}$正定。

(2) 若$|\boldsymbol{C}|\neq 0$，则$\boldsymbol{A}=\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}}$正定。

思路分析：证明$\boldsymbol{A}$矩阵正定，可以将其转换为证明相应的二次型函数正定。

证明：(1) $\forall\ \boldsymbol{x}\in R^n$，易知

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{x}=\|\boldsymbol{B}^{\mathrm{T}}\boldsymbol{x}\|^2\geqslant 0$$

此外，由于$\boldsymbol{B}$非奇异，所以当且仅当$\boldsymbol{x}=\boldsymbol{0}$时，上式取等号。所以，$\boldsymbol{A}$为正定矩阵。

(2) 证明同上，不再重复。

1.59 设$n\times n$ $(n\geqslant 3)$矩阵

$$\boldsymbol{A}=\begin{bmatrix}1&a&\cdots&a\\a&1&\cdots&a\\\vdots&\vdots&&\vdots\\a&a&\cdots&1\end{bmatrix}$$

当a取何值时，矩阵$\boldsymbol{A}$的秩为$n-1$。

思路分析：此题考察的是对称矩阵特征值分解的灵活应用。从题目中可以看出，$\boldsymbol{A}$矩阵具有非常特殊的结构，即可以表示为$\boldsymbol{A}=a\boldsymbol{B}+\boldsymbol{I}$的形式，其中$\boldsymbol{B}$矩阵为除了对角线上元素为零，其余元素均为1的特殊矩阵。接下来，我们只需要考察$\boldsymbol{B}$矩阵的特征值分解情况即可。

解：对上述$\boldsymbol{B}$矩阵进行特征值分解可以发现，

$$\boldsymbol{B}=\boldsymbol{U\Sigma U}^{\mathrm{T}}$$

其中，$\boldsymbol{\Sigma}=\mathrm{diag}\{n-1,-1,\cdots,-1\}$。代入到$\boldsymbol{A}$的表达式中，有

$$\begin{aligned}\boldsymbol{A}&=a\boldsymbol{U\Sigma U}^{\mathrm{T}}+\boldsymbol{UU}^{\mathrm{T}}\\&=\boldsymbol{U}\begin{bmatrix}a(n-1)+1&&&\\&-a+1&&\\&&\ddots&\\&&&-a+1\end{bmatrix}\boldsymbol{U}^{\mathrm{T}}\end{aligned}$$

显然，当$a=-1/(n-1)$时，$\boldsymbol{A}$矩阵的秩为$n-1$。

1.60 已知$\boldsymbol{AB}=\boldsymbol{BA}=\boldsymbol{O}$（零矩阵）和$\mathrm{rank}(\boldsymbol{A}^2)=\mathrm{rank}(\boldsymbol{A})$，证明

(1) $\mathrm{rank}(\boldsymbol{A}+\boldsymbol{B})=\mathrm{rank}(\boldsymbol{A})+\mathrm{rank}(\boldsymbol{B})$。

(2) $\mathrm{rank}(\boldsymbol{A}^k+\boldsymbol{B}^k)=\mathrm{rank}(\boldsymbol{A}^k)+\mathrm{rank}(\boldsymbol{B}^k)$，其中，$k$是某个整数。

思路分析：此题证明比较困难。可以考虑先对矩阵$\boldsymbol{A}$进行Jordan标准型分解（见教材第4章），然后根据已给定的$\boldsymbol{A}$的秩的关系式来得到Jordan块的某行性质，进而考察$\boldsymbol{A}+\boldsymbol{B}$的特殊结构。

证明：(1) 首先，对矩阵$\boldsymbol{A}$进行Jordan标准型分解，即

$$\boldsymbol{A}=\boldsymbol{S}\begin{bmatrix}\boldsymbol{J}_{n_1}(\lambda_1)&&&0\\&\boldsymbol{J}_{n_2}(\lambda_2)&&\\&&\ddots&\\0&&&\boldsymbol{J}_{n_k}(\lambda_k)\end{bmatrix}\boldsymbol{S}^{-1}=\boldsymbol{S\Sigma S}^{-1}$$

其中，$\boldsymbol{S}$为可逆矩阵，

$$\boldsymbol{J}_k(\lambda_k)=\begin{bmatrix}\lambda_k&1&&&0\\&\lambda_k&1&&\\&&\ddots&\ddots&\\&&&\ddots&1\\0&&&&\lambda_k\end{bmatrix}$$

据此可以得到

$$\boldsymbol{A}^2=\boldsymbol{S}\begin{bmatrix}\boldsymbol{J}^2_{n_1}(\lambda_1)&&&0\\&\boldsymbol{J}^2_{n_2}(\lambda_2)&&\\&&\ddots&\\0&&&\boldsymbol{J}^2_{n_k}(\lambda_k)\end{bmatrix}\boldsymbol{S}^{-1}=\boldsymbol{S\Sigma S}^{-1}$$

从上面可以看出，经过Jordan标准型分解之后，$\boldsymbol{A}$和$\boldsymbol{A}^2$的秩实际上由$\boldsymbol{\Sigma}$和$\boldsymbol{\Sigma}^2$决定。因为已知$\text{rank}(\boldsymbol{A}^2) = \text{rank}(\boldsymbol{A})$，所以可以断定：矩阵$\boldsymbol{A}$中一定不存在下面形式的Jordan块

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

否则，$\boldsymbol{\Sigma}^2$的秩将小于$\boldsymbol{\Sigma}$的秩，从而$\boldsymbol{A}^2$的秩将小于$\boldsymbol{A}$的秩。也就是说，矩阵$\boldsymbol{A}$的0特征值对应的Jordan块一定是一阶的。

若定义$\boldsymbol{C} = \boldsymbol{S}^{-1}\boldsymbol{B}\boldsymbol{S}$，则易知，

$$\boldsymbol{AB} = \boldsymbol{O} \Leftrightarrow \boldsymbol{S\Sigma C} = \boldsymbol{O} \Leftrightarrow \boldsymbol{\Sigma C} = \boldsymbol{O}$$

若将$\boldsymbol{C}$按行分块，且与$\boldsymbol{\Sigma}$的各个Jordan块的维数相对应，即

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_1 \\ \boldsymbol{C}_2 \\ \vdots \\ \boldsymbol{C}_{n_k} \end{bmatrix}$$

由$\boldsymbol{\Sigma C} = \boldsymbol{O}$可知，$\boldsymbol{C}$矩阵与$\boldsymbol{\Sigma}$非零Jordan块相对应的行一定为零。

同理，由于$\boldsymbol{BA} = \boldsymbol{O}$，则$\boldsymbol{C}$矩阵与$\boldsymbol{\Sigma}$非零Jordan块相对应的列也一定为零。这样，有

$$\text{rank}(\boldsymbol{\Sigma} + \boldsymbol{C}) = \text{rank}(\boldsymbol{\Sigma}) + \text{rank}(\boldsymbol{C})$$

在$\boldsymbol{\Sigma} + \boldsymbol{C}$两端分别左乘$\boldsymbol{S}$,、右乘$\boldsymbol{S}^{-1}$，不会影响原矩阵的秩。最终得到

$$\text{rank}(\boldsymbol{A} + \boldsymbol{B}) = \text{rank}(\boldsymbol{\Sigma} + \boldsymbol{C}) = \text{rank}(\boldsymbol{\Sigma}) + \text{rank}(\boldsymbol{C}) = \text{rank}(\boldsymbol{A}) + \text{rank}(\boldsymbol{B})$$

(2) 因为已知$\text{rank}(\boldsymbol{A}^2) = \text{rank}(\boldsymbol{A})$，所以有

$$\text{rank}(\boldsymbol{A}^{2k}) = \text{rank}(\boldsymbol{A}^k)$$

成立。因为$\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{O}$，所以

$$\boldsymbol{A}^k\boldsymbol{B}^k = \boldsymbol{B}^k\boldsymbol{A}^k = \boldsymbol{O}$$

应用上一问结论，直接得到要证明的结果。

1.61 令$\boldsymbol{C}_{n\times n}$ 是一任意对称矩阵，证明：存在满足$\boldsymbol{AB} = \boldsymbol{O}$的两个唯一非负定矩阵$\boldsymbol{A}$和$\boldsymbol{B}$，使得$\boldsymbol{C} = \boldsymbol{A} - \boldsymbol{B}$。

思路分析： 利用对称矩阵的特征值分解，可以构造出满足条件的矩阵$\boldsymbol{A}$和$\boldsymbol{B}$，进而再证明二者是唯一的。

证明：将对称矩阵$\boldsymbol{C}$对角化，即进行特征值分解，得到

$$\boldsymbol{C} = \boldsymbol{U\Sigma U}^{\mathrm{T}}$$

其中，$\boldsymbol{\Sigma}$为对角阵，其对角线上的元素为$\boldsymbol{C}$矩阵的实特征值，$\boldsymbol{U}$为相应的特征向量构成的矩阵。

若将所有特征值重新编排，即正的特征值构成对角阵$\boldsymbol{\Sigma}_1$，负特征值构成$\boldsymbol{\Sigma}_2$，对应的特征向量构成的矩阵为$\boldsymbol{U}_1$和$\boldsymbol{U}_2$，则矩阵$\boldsymbol{C}$可以重新写作

$$\begin{aligned}\boldsymbol{C} &= [\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{U}_3]\begin{bmatrix} \boldsymbol{\Sigma}_1 & & \\ & \boldsymbol{\Sigma}_2 & \\ & & \boldsymbol{O} \end{bmatrix}\begin{bmatrix} \boldsymbol{U}_1^{\mathrm{T}} \\ \boldsymbol{U}_2^{\mathrm{T}} \\ \boldsymbol{U}_3^{\mathrm{T}} \end{bmatrix} \\ &= \boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{U}_1^{\mathrm{T}} + \boldsymbol{U}_2\boldsymbol{\Sigma}_2\boldsymbol{U}_2^{\mathrm{T}}\end{aligned}$$

其中，$\boldsymbol{U}_3$为对应于0特征值的特征向量构成的矩阵。

令$\boldsymbol{A} = \boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{U}_1^{\mathrm{T}}$，$\boldsymbol{B} = -\boldsymbol{U}_2\boldsymbol{\Sigma}_2\boldsymbol{U}_2^{\mathrm{T}}$，则容易验证，$\boldsymbol{A}$和$\boldsymbol{B}$都是非负定矩阵，且$\boldsymbol{C} = \boldsymbol{A} - \boldsymbol{B}$。另外，由于$\boldsymbol{U}_1, \boldsymbol{U}_2$是相互正交的，所以$\boldsymbol{AB} = \boldsymbol{O}$成立。

下面，我们证明$\boldsymbol{A}$和$\boldsymbol{B}$是唯一的。设二者的特征值分解为

$$\boldsymbol{A} = \boldsymbol{V}_1\boldsymbol{\Sigma}_a\boldsymbol{V}_1^{\mathrm{T}}, \quad \boldsymbol{B} = \boldsymbol{V}_2\boldsymbol{\Sigma}_b\boldsymbol{V}_2^{\mathrm{T}}$$

其中$\boldsymbol{V}_1$和$\boldsymbol{V}_2$分别为$n\times r(\boldsymbol{A})$和$n\times r(\boldsymbol{B})$维矩阵，$r(\boldsymbol{A})$和$r(\boldsymbol{B})$是两个矩阵的秩。因为$\boldsymbol{A}$和$\boldsymbol{B}$都是非负定矩阵，所以二者的特征值都大于等于零。由于$\boldsymbol{A}$和$\boldsymbol{B}$正交，所以$\boldsymbol{V}_1\boldsymbol{V}_2^{\mathrm{T}} = \boldsymbol{O}$。这样就可以得到

$$\mathrm{Null}(\boldsymbol{A}) = \mathrm{Null}(\boldsymbol{V}_1) \supseteq \mathrm{Col}(\boldsymbol{V}_2) = \mathrm{Col}(\boldsymbol{B})$$

从而有

$$r(\boldsymbol{A}) + r(\boldsymbol{B}) \leqslant n$$

因为有关系式$\boldsymbol{C} = \boldsymbol{A} - \boldsymbol{B}$，所以$\boldsymbol{C}$可以重新写作

$$\boldsymbol{C} = \boldsymbol{V}_1\boldsymbol{\Sigma}_a\boldsymbol{V}_1^{\mathrm{T}} + \boldsymbol{V}_2(-\boldsymbol{\Sigma}_b)\boldsymbol{V}_2^{\mathrm{T}} = [\boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{V}_3]\begin{bmatrix} \boldsymbol{\Sigma}_a & & \\ & -\boldsymbol{\Sigma}_b & \\ & & \boldsymbol{O} \end{bmatrix}\begin{bmatrix} \boldsymbol{V}_1^{\mathrm{T}} \\ \boldsymbol{V}_2^{\mathrm{T}} \\ \boldsymbol{V}_3^{\mathrm{T}} \end{bmatrix}$$

其中，$\boldsymbol{V}_3$由与$\boldsymbol{V}_1$和$\boldsymbol{V}_2$都正交的标准向量构成。这事实上给出了矩阵$\boldsymbol{C}$的特征值分解。我们知道，矩阵的特征值只有唯一一组，因而可以断定，$\boldsymbol{\Sigma}_a$为$\boldsymbol{\Sigma}_1$中的元素经过重排列后的矩阵，$\boldsymbol{\Sigma}_b$为$\boldsymbol{\Sigma}_2$中的元素经过重排列后的矩阵。而

$$\boldsymbol{V}_1 = \boldsymbol{U}_1\boldsymbol{Q}_1, \quad \boldsymbol{V}_2 = \boldsymbol{U}_2\boldsymbol{Q}_2$$

其中，$\boldsymbol{Q}_1$和$\boldsymbol{Q}_2$为某正交阵。虽然$\boldsymbol{V}_1$和$\boldsymbol{V}_2$并不是唯一，但不难证明满足条件的$\boldsymbol{A}$和$\boldsymbol{B}$是唯一的。

1.62 已知向量组$\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_p\}$中$\boldsymbol{x}_p\neq\boldsymbol{0}$。令$a_1,a_2,\cdots,a_{p-1}$为任意常数，并且

$$\boldsymbol{y}_i=\boldsymbol{x}_i+a_i\boldsymbol{x}_p,\quad i=1,2,\cdots,p-1$$

证明向量组$\{\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_{p-1}\}$线性无关的充分条件是向量组$\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_p\}$ 线性无关。

思路分析：证明向量组的线性相关性，可以直接利用定义证明，也可以将向量组看作矩阵，转而证明矩阵的满秩。下面，我们就从矩阵的角度出发（$\boldsymbol{X}$满列秩），给出此题的完整证明过程。

证明：设矩阵$\boldsymbol{X}$和$\boldsymbol{Y}$分别由两个向量组构成，即

$$\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_p],\qquad \boldsymbol{Y}=[\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_{p-1}]$$

由题意得，$\boldsymbol{Y}=\boldsymbol{X}\boldsymbol{A}$，其中

$$\boldsymbol{A}=\begin{bmatrix}1&&&\\&1&&\\&&\ddots&\\&&&1\\a_1&a_2&\cdots&a_{p-1}\end{bmatrix}$$

显然，$\text{rank}(\boldsymbol{A})=p-1$。接下来，根据矩阵乘积的秩的不等式，我们可以得到

$$\text{rank}(\boldsymbol{X})+\text{rank}(\boldsymbol{A})-p\leqslant\text{rank}(\boldsymbol{Y})\leqslant\min\{\text{rank}(\boldsymbol{X}),\text{rank}(\boldsymbol{A})\}$$

当向量组$\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_p\}$ 线性无关时，$\text{rank}(\boldsymbol{X})=p$，从而有

$$p-1\leqslant\text{rank}(\boldsymbol{Y})\leqslant p-1$$

即$\text{rank}(\boldsymbol{Y})=p-1$。所以，向量组$\{\boldsymbol{y}_1,\cdots,\boldsymbol{y}_{p-1}\}$是线性无关的。

1.63 证明：若$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{A}$，则$\boldsymbol{A}=\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}^2$。

证明：因为$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{A}$，且$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$为对称矩阵，所以$\boldsymbol{A}$为对称矩阵，即$\boldsymbol{A}=\boldsymbol{A}^{\mathrm{T}}$。在上式两边同时右乘$\boldsymbol{A}$ 得到$\boldsymbol{A}^2=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{A}$，证毕。

1.64 [23] 设

$$\boldsymbol{K}=\boldsymbol{K}^{\mathrm{T}}=\boldsymbol{K}^3,\quad \boldsymbol{K}\boldsymbol{1}=\boldsymbol{0},\qquad \boldsymbol{K}\begin{bmatrix}1\\2\\-3\end{bmatrix}=\begin{bmatrix}1\\2\\-3\end{bmatrix}$$

式中，$\boldsymbol{1}$是一个元素全部为1的向量。计算下列值，并且说明为什么在无需计算$\boldsymbol{K}$的情况下，可以得到下列值的原因：

(1) $\boldsymbol{K}$的阶数。

(2) $\boldsymbol{K}$的秩。

(3) $\boldsymbol{K}$的迹和行列式。

(4) $\boldsymbol{K}^{26}$的迹和行列式。

(5) 矩阵$6\boldsymbol{K}^{60}-7\boldsymbol{K}^{37}+3\boldsymbol{I}$的迹与行列式。

解: (1) 因为$\boldsymbol{K}$为对称矩阵，所以其行数等于列数，又因为$\boldsymbol{K}$可以左乘一个三维列向量，所以$\boldsymbol{K}$为3阶方阵。

(2) 对$\boldsymbol{K}$做特征值分解，得到

$$\boldsymbol{K}=\boldsymbol{U}\begin{bmatrix}\lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3\end{bmatrix}\boldsymbol{U}^{\mathrm{T}}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{T}}$$

因为$\boldsymbol{K}^3=\boldsymbol{K}$，所以可知3阶矩阵$\boldsymbol{K}$的特征值只能从$-1$，1，0中取。题目中已经给出$\lambda_1=1$和$\lambda_2=0$都是$\boldsymbol{K}$的特征值，但$\lambda_3$可以是$-1$，1，0中的任意一个。所以下面应该分情况进行讨论：

(a) 若$\lambda_3=-1$，则$\mathrm{rank}(\boldsymbol{K})=2$;

(b) 若$\lambda_3=1$，即1为二重特征值，则$\mathrm{rank}(\boldsymbol{K})=2$;

(c) 若$\lambda_3=0$，即0为二重特征值，则$\mathrm{rank}(\boldsymbol{K})=1$。

(3) $\mathrm{tr}(\boldsymbol{K})=\lambda_1+\lambda_2+\lambda_3$。当$\lambda_3$分别取$-1$，1，0时，对应的$\mathrm{tr}(\boldsymbol{K})$分别为0，2和1。$|\boldsymbol{K}|=\lambda_1\lambda_2\lambda_3=0$。

(4) $\mathrm{tr}(\boldsymbol{K}^{26})=\lambda_1^{26}+\lambda_2^{26}+\lambda_3^{26}$。当$\lambda_3$取$-1$，1，0时，对应的$\mathrm{tr}(\boldsymbol{K}^{26})$分别为2，2和1。$|\boldsymbol{K}^{26}|=(\lambda_1\lambda_2\lambda_3)^{26}=0$。

(5) $\mathrm{tr}(6\boldsymbol{K}^{60}-7\boldsymbol{K}^{37}+3\boldsymbol{I})=6\mathrm{tr}(\boldsymbol{K}^{60})-7\mathrm{tr}(\boldsymbol{K}^{37})+9$。$\lambda_3$取$-1$，1，0时，对应的结果分别为21，7，8。$|6\boldsymbol{K}^{60}-7\boldsymbol{K}^{37}+3\boldsymbol{I}|=|6\boldsymbol{\Sigma}^{60}-7\boldsymbol{\Sigma}^{37}+3\boldsymbol{I}|$，$\lambda_3$取$-1$，1，0时，对应的结果分别为96，12和18。

1.65 假定下面提到的每个逆矩阵都存在，证明以下结果：

(1) $(\boldsymbol{A}^{-1}+\boldsymbol{I})^{-1}=\boldsymbol{A}(\boldsymbol{A}+\boldsymbol{I})^{-1}$。

(2) $(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1}=\boldsymbol{A}(\boldsymbol{A}+\boldsymbol{B})^{-1}\boldsymbol{B}=\boldsymbol{B}(\boldsymbol{A}+\boldsymbol{B})^{-1}\boldsymbol{A}$。

(3) $(\boldsymbol{I}+\boldsymbol{A}\boldsymbol{B})^{-1}\boldsymbol{A}=\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{B}\boldsymbol{A})^{-1}$。

(4) $(\boldsymbol{A}+\boldsymbol{B})^{-1}=\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}$ 意味着 $\boldsymbol{A}+\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}=\boldsymbol{B}+\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}$。

(5) $\boldsymbol{A}-\boldsymbol{A}(\boldsymbol{A}+\boldsymbol{B})^{-1}\boldsymbol{A}=\boldsymbol{B}-\boldsymbol{B}(\boldsymbol{A}+\boldsymbol{B})^{-1}\boldsymbol{B}$。

思路分析: 本题的证明主要依据逆矩阵的若干性质，具体可参见本章的“主要方法与理论”或教材P67。

解: (1) 注意到 $\boldsymbol{A}\boldsymbol{A}^{-1}=\boldsymbol{A}^{-1}\boldsymbol{A}=\boldsymbol{I}$，所以

$$(\boldsymbol{A}^{-1}+\boldsymbol{I})^{-1}=(\boldsymbol{A}^{-1}+\boldsymbol{A}\boldsymbol{A}^{-1})^{-1}=[(\boldsymbol{I}+\boldsymbol{A})\boldsymbol{A}^{-1}]^{-1}=\boldsymbol{A}(\boldsymbol{A}+\boldsymbol{I})^{-1}$$

(2) 由于 $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$，$\boldsymbol{B}\boldsymbol{B}^{-1} = \boldsymbol{B}^{-1}\boldsymbol{B} = \boldsymbol{I}$，所以

$$\begin{aligned}(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1} &= (\boldsymbol{A}^{-1}+\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{B}^{-1})^{-1} = (\boldsymbol{I}+\boldsymbol{A}\boldsymbol{B}^{-1})^{-1}\boldsymbol{A}\\ &= (\boldsymbol{B}\boldsymbol{B}^{-1}+\boldsymbol{A}\boldsymbol{B}^{-1})^{-1}\boldsymbol{A} = \boldsymbol{B}(\boldsymbol{A}+\boldsymbol{B})^{-1}\boldsymbol{A}\end{aligned}$$

同理，由 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的对称性可得，$(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1} = \boldsymbol{A}(\boldsymbol{A}+\boldsymbol{B})^{-1}\boldsymbol{B}$。

(3) 本小题证明思路与上面类似：

$$\begin{aligned}(\boldsymbol{I}+\boldsymbol{A}\boldsymbol{B})^{-1}\boldsymbol{A} &= [\boldsymbol{A}^{-1}(\boldsymbol{I}+\boldsymbol{A}\boldsymbol{B})]^{-1} = (\boldsymbol{A}^{-1}+\boldsymbol{B})^{-1}\\ &= (\boldsymbol{A}^{-1}+\boldsymbol{B}\boldsymbol{A}\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}(\boldsymbol{I}+\boldsymbol{B}\boldsymbol{A})^{-1}\end{aligned}$$

(4) 一方面，

$$(\boldsymbol{A}+\boldsymbol{B})^{-1} = \boldsymbol{A}^{-1}(\boldsymbol{I}+\boldsymbol{B}\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}^{-1}\boldsymbol{B}^{-1}(\boldsymbol{B}^{-1}+\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1})^{-1}$$

另一方面，

$$(\boldsymbol{A}+\boldsymbol{B})^{-1} = (\boldsymbol{B}^{-1}\boldsymbol{A}+\boldsymbol{I})^{-1}\boldsymbol{B}^{-1} = (\boldsymbol{A}^{-1}+\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}$$

由于 $(\boldsymbol{A}+\boldsymbol{B})^{-1} = \boldsymbol{A}^{-1}\boldsymbol{B}^{-1}$，所以有

$$\boldsymbol{B}^{-1}+\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1} = \boldsymbol{A}^{-1}+\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}\boldsymbol{A}$$

两边同时左乘 $\boldsymbol{A}$ 和右乘 $\boldsymbol{B}$，得

$$\boldsymbol{A}+\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1} = \boldsymbol{B}+\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}$$

(5) 本题有两种证明方法。方法一：由 (2) 中结果知，

$$\begin{aligned}\boldsymbol{A}-\boldsymbol{A}(\boldsymbol{A}+\boldsymbol{B})^{-1}\boldsymbol{A} &= \boldsymbol{A}-\boldsymbol{A}(\boldsymbol{A}+\boldsymbol{B})^{-1}\boldsymbol{B}\boldsymbol{B}^{-1}\boldsymbol{A}\\ &= \boldsymbol{A}-\boldsymbol{B}(\boldsymbol{A}+\boldsymbol{B})^{-1}\boldsymbol{A}\boldsymbol{B}^{-1}\boldsymbol{A}\\ &= \boldsymbol{B}(\boldsymbol{A}+\boldsymbol{B})^{-1}[(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{B}^{-1}\boldsymbol{A}-\boldsymbol{A}\boldsymbol{B}^{-1}\boldsymbol{A}]\\ &= \boldsymbol{B}(\boldsymbol{A}+\boldsymbol{B})^{-1}\boldsymbol{A}\end{aligned}$$

另一方面，

$$\begin{aligned}\boldsymbol{B}-\boldsymbol{B}(\boldsymbol{A}+\boldsymbol{B})^{-1}\boldsymbol{B} &= \boldsymbol{B}(\boldsymbol{A}+\boldsymbol{B})^{-1}[(\boldsymbol{A}+\boldsymbol{B})-\boldsymbol{B}]\\ &= \boldsymbol{B}(\boldsymbol{A}+\boldsymbol{B})^{-1}\boldsymbol{A}\end{aligned}$$

等式得证。

方法二：根据推广的矩阵求逆公式，有

$$(\boldsymbol{A}+\boldsymbol{B})^{-1} = \boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}(\boldsymbol{I}+\boldsymbol{B}\boldsymbol{A}^{-1})^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}$$

$$(\boldsymbol{B}+\boldsymbol{A})^{-1} = \boldsymbol{B}^{-1}-\boldsymbol{B}^{-1}(\boldsymbol{I}+\boldsymbol{A}\boldsymbol{B}^{-1})^{-1}\boldsymbol{A}\boldsymbol{B}^{-1}$$

于是，

$$A(A+B)^{-1}A = A-(I+BA^{-1})^{-1}B = A-(B^{-1}+A^{-1})^{-1}$$
$$B(B+A)^{-1}B = B-(I+AB^{-1})^{-1}A = B-(A^{-1}+B^{-1})^{-1}$$

所以，

$$A-A(A+B)^{-1}A = B-B(A+B)^{-1}B = (A^{-1}+B^{-1})^{-1}$$

1.66 验证分块矩阵求逆公式：

(1) 矩阵 A 可逆时，为

$$\begin{bmatrix} A & U \\ V & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1}+A^{-1}U(D-VA^{-1}U)^{-1}VA^{-1} & -A^{-1}U(D-VA^{-1}U)^{-1} \\ -(D-VA^{-1}U)^{-1}VA^{-1} & (D-VA^{-1}U)^{-1} \end{bmatrix}$$

(2) 矩阵 A 和 D 可逆时，为

$$\begin{bmatrix} A & U \\ V & D \end{bmatrix}^{-1} = \begin{bmatrix} (A-UD^{-1}V)^{-1} & -A^{-1}U(D-VA^{-1}U)^{-1} \\ -D^{-1}V(A-UD^{-1}V)^{-1} & (D-VA^{-1}U)^{-1} \end{bmatrix}$$

(3) 矩阵 A 和 D 可逆时，为

$$\begin{bmatrix} A & U \\ V & D \end{bmatrix}^{-1} = \begin{bmatrix} (A-UD^{-1}V)^{-1} & -(A-UD^{-1}V)^{-1}UD^{-1} \\ -(D-VA^{-1}U)^{-1}VA^{-1} & (D-VA^{-1}U)^{-1} \end{bmatrix}$$

思路分析： 关于分块矩阵求逆公式的验证或证明主要依据逆矩阵的定义 $AA^{-1} = A^{-1}A = I$，思路比较简单，但具体计算过程可能比较繁琐。

解： (1) 因为矩阵 A 可逆，所以

$$\begin{bmatrix} A & U \\ V & D \end{bmatrix} \begin{bmatrix} A^{-1}+A^{-1}U(D-VA^{-1}U)^{-1}VA^{-1} & -A^{-1}U(D-VA^{-1}U)^{-1} \\ -(D-VA^{-1}U)^{-1}VA^{-1} & (D-VA^{-1}U)^{-1} \end{bmatrix} = \begin{bmatrix} W & X \\ Y & Z \end{bmatrix}$$

其中

$$\begin{aligned} W &= AA^{-1}+AA^{-1}U(D-VA^{-1}U)^{-1}VA^{-1}-U(D-VA^{-1}U)^{-1}VA^{-1} = I \\ X &= -AA^{-1}U(D-VA^{-1}U)^{-1}+U(D-VA^{-1}U)^{-1} = O \\ Y &= VA^{-1}+VA^{-1}U(D-VA^{-1}U)^{-1}VA^{-1}-D(D-VA^{-1}U)^{-1}VA^{-1} \\ &= VA^{-1}+(VA^{-1}U-D)(D-VA^{-1}U)^{-1}VA^{-1} = O \\ Z &= -VA^{-1}U(D-VA^{-1}U)^{-1}+D(D-VA^{-1}U)^{-1} \\ &= (D-VA^{-1}U)(D-VA^{-1}U)^{-1} = I \end{aligned}$$

另一方面，

$$\begin{bmatrix} A^{-1}+A^{-1}U(D-VA^{-1}U)^{-1}VA^{-1} & -A^{-1}U(D-VA^{-1}U)^{-1} \\ -(D-VA^{-1}U)^{-1}VA^{-1} & (D-VA^{-1}U)^{-1} \end{bmatrix} \begin{bmatrix} A & U \\ V & D \end{bmatrix} = \begin{bmatrix} M & N \\ P & Q \end{bmatrix}$$

其中

$$\begin{aligned}
M &= A^{-1}A + A^{-1}U(D - VA^{-1}U)^{-1}VA^{-1}A - A^{-1}U(D - VA^{-1}U)^{-1}V = I\\
N &= A^{-1}U + A^{-1}U(D - VA^{-1}U)^{-1}VA^{-1}U - A^{-1}U(D - VA^{-1}U)^{-1}D\\
&= A^{-1}U + A^{-1}U(D - VA^{-1}U)^{-1}(VA^{-1}U - D) = O\\
P &= -(D - VA^{-1}U)^{-1}VA^{-1}A + (D - VA^{-1}U)^{-1}V = O\\
Q &= -(D - VA^{-1}U)^{-1}VA^{-1}U + (D - VA^{-1}U)^{-1}D\\
&= (D - VA^{-1}U)^{-1}(D - VA^{-1}U) = I
\end{aligned}$$

(2) 因为矩阵 A 和D 可逆，所以

$$\begin{bmatrix} A & U \\ V & D \end{bmatrix}\begin{bmatrix} (A - UD^{-1}V)^{-1} & -A^{-1}U(D - VA^{-1}U)^{-1} \\ -D^{-1}V(A - UD^{-1}V)^{-1} & (D - VA^{-1}U)^{-1} \end{bmatrix} = \begin{bmatrix} W & X \\ Y & Z \end{bmatrix}$$

其中

$$\begin{aligned}
W &= A(A - UD^{-1}V)^{-1} - UD^{-1}V(A - UD^{-1}V)^{-1}\\
&= (A - UD^{-1}V)(A - UD^{-1}V)^{-1} = I\\
X &= -AA^{-1}U(D - VA^{-1}U)^{-1} + U(D - VA^{-1}U)^{-1} = O\\
Y &= V(A - UD^{-1}V)^{-1} - DD^{-1}V(A - UD^{-1}V)^{-1} = O\\
Z &= -VA^{-1}U(D - VA^{-1}U)^{-1} + D(D - VA^{-1}U)^{-1}\\
&= (D - VA^{-1}U)(D - VA^{-1}U)^{-1} = I
\end{aligned}$$

另一方面，

$$\begin{bmatrix} (A - UD^{-1}V)^{-1} & -A^{-1}U(D - VA^{-1}U)^{-1} \\ -D^{-1}V(A - UD^{-1}V)^{-1} & (D - VA^{-1}U)^{-1} \end{bmatrix}\begin{bmatrix} A & U \\ V & D \end{bmatrix} = \begin{bmatrix} M & N \\ P & Q \end{bmatrix}$$

其中

$$\begin{aligned}
M &= (A - UD^{-1}V)^{-1}A - A^{-1}U(D - VA^{-1}U)^{-1}V\\
&= [A^{-1} + A^{-1}U(D - VA^{-1}U)^{-1}VA^{-1}]A - A^{-1}U(D - VA^{-1}U)^{-1}V = I\\
N &= (A - UD^{-1}V)^{-1}U - A^{-1}U(D - VA^{-1}U)^{-1}D\\
&= [A^{-1} + A^{-1}U(D - VA^{-1}U)^{-1}VA^{-1}]U - A^{-1}U(D - VA^{-1}U)^{-1}D\\
&= A^{-1}U + A^{-1}U(D - VA^{-1}U)^{-1}(VA^{-1}U - D) = O\\
P &= -D^{-1}V(A - UD^{-1}V)^{-1}A + (D - VA^{-1}U)^{-1}V\\
&= -D^{-1}V - D^{-1}VA^{-1}U(D - VA^{-1}U)^{-1}V + (D - VA^{-1}U)^{-1}V\\
&= -D^{-1}V - D^{-1}(VA^{-1}U - D)(D - VA^{-1}U)^{-1}V = O
\end{aligned}$$

$$
\begin{aligned}
Q &= -D^{-1}V(A-UD^{-1}V)^{-1}U+(D-VA^{-1}U)^{-1}D \\
&= -D^{-1}VA^{-1}U-D^{-1}VA^{-1}U(D-VA^{-1}U)^{-1}VA^{-1}U+(D-VA^{-1}U)^{-1}D \\
&= -D^{-1}VA^{-1}U(D-VA^{-1}U)^{-1}D+(D-VA^{-1}U)^{-1}D=I
\end{aligned}
$$

这里利用了 Duncan-Guttman 矩阵求逆公式：

$$(A-UD^{-1}V)^{-1}=A^{-1}+A^{-1}U(D-VA^{-1}U)^{-1}VA^{-1}$$

(3) 因为矩阵 A 和 D 可逆，所以

$$\begin{bmatrix} A & U \\ V & D \end{bmatrix}\begin{bmatrix} (A-UD^{-1}V)^{-1} & -(A-UD^{-1}V)^{-1}UD^{-1} \\ -(D-VA^{-1}U)^{-1}VA^{-1} & (D-VA^{-1}U)^{-1} \end{bmatrix}=\begin{bmatrix} W & X \\ Y & Z \end{bmatrix}$$

其中

$$
\begin{aligned}
W &= A(A-UD^{-1}V)^{-1}-U(D-VA^{-1}U)^{-1}VA^{-1} \\
&= A[A^{-1}+A^{-1}U(D-VA^{-1}U)^{-1}VA^{-1}]-U(D-VA^{-1}U)^{-1}VA^{-1}=I \\
X &= -A(A-UD^{-1}V)^{-1}UD^{-1}+U(D-VA^{-1}U)^{-1} \\
&= -UD^{-1}-U(D-VA^{-1}U)^{-1}VA^{-1}UD^{-1}+U(D-VA^{-1}U)^{-1} \\
&= -UD^{-1}+U(D-VA^{-1}U)^{-1}(-VA^{-1}U+D)D^{-1}=O \\
Y &= V(A-UD^{-1}V)^{-1}-D(D-VA^{-1}U)^{-1}VA^{-1} \\
&= VA^{-1}+VA^{-1}U(D-VA^{-1}U)^{-1}VA^{-1}-D(D-VA^{-1}U)^{-1}VA^{-1}=O \\
Z &= -V(A-UD^{-1}V)^{-1}UD^{-1}+D(D-VA^{-1}U)^{-1} \\
&= -V(A-UD^{-1}V)^{-1}UD^{-1}+D[D^{-1}+D^{-1}V(A-UD^{-1}V)^{-1}UD^{-1}]=I
\end{aligned}
$$

另一方面，

$$\begin{bmatrix} (A-UD^{-1}V)^{-1} & -(A-UD^{-1}V)^{-1}UD^{-1} \\ -(D-VA^{-1}U)^{-1}VA^{-1} & (D-VA^{-1}U)^{-1} \end{bmatrix}\begin{bmatrix} A & U \\ V & D \end{bmatrix}=\begin{bmatrix} M & N \\ P & Q \end{bmatrix}$$

其中

$$
\begin{aligned}
M &= (A-UD^{-1}V)^{-1}A-(A-UD^{-1}V)^{-1}UD^{-1}V=I \\
N &= (A-UD^{-1}V)^{-1}U-(A-UD^{-1}V)^{-1}UD^{-1}D=O \\
P &= -(D-VA^{-1}U)^{-1}VA^{-1}A+(D-VA^{-1}U)^{-1}V=O \\
Q &= -(D-VA^{-1}U)^{-1}VA^{-1}U+(D-VA^{-1}U)^{-1}D=I
\end{aligned}
$$

1.67 用矩阵求逆公式 $A^{-1}=\dfrac{1}{\det(A)}\mathrm{adj}(A)$ 分别计算矩阵

$$A=\begin{bmatrix}1 & 3 & 2\\5 & 2 & 0\\2 & -1 & 1\end{bmatrix}$$
$$B=\begin{bmatrix}4 & 1 & 3\\0 & 2 & 0\\-4 & 1 & -4\end{bmatrix}$$

的逆矩阵。(注：adj(A) 表示矩阵A 的伴随矩阵。)

解： 对于矩阵 A，由于

$$\det(A)=-31$$
$$\mathrm{adj}(A)=\begin{bmatrix}2 & -5 & -9\\-5 & -3 & 7\\-4 & 10 & -13\end{bmatrix}^{\mathrm{T}}$$

所以

$$A^{-1}=\begin{bmatrix}-\frac{2}{31} & \frac{5}{31} & \frac{4}{31}\\ \frac{5}{31} & \frac{3}{31} & -\frac{10}{31}\\ \frac{9}{31} & -\frac{7}{31} & \frac{13}{31}\end{bmatrix}$$

对于矩阵 B，由于

$$\det(B)=-8$$
$$\mathrm{adj}(B)=\begin{bmatrix}-8 & 0 & 8\\7 & -4 & -8\\-6 & 0 & 8\end{bmatrix}^{\mathrm{T}}$$

所以

$$B^{-1}=\begin{bmatrix}1 & -\frac{7}{8} & \frac{3}{4}\\ 0 & \frac{1}{2} & 0\\ -1 & 1 & -1\end{bmatrix}$$

1.68 若 $Y=(AX+B)(CX+D)^{-1}$，试用 Y 表示 X。

解： 在 $Y=(AX+B)(CX+D)^{-1}$ 两边同时右乘 $(CX+D)$，得

$$Y(CX+D)=AX+B \quad\Longrightarrow\quad (A-YC)X=YD-B$$

由于 $(A-YC)$ 既可能是正方矩阵，也可能是长方矩阵；既可能是满秩的，也可能是秩亏缺的，故若记 $(A-YC)^{-}$ 表示 $(A-YC)$ 的广义逆矩阵，则有 $X=(A-YC)^{-}(YD-B)$。

1.69 当 $\boldsymbol{X}$ 是随机变量的矩阵时，不存在 $\boldsymbol{X}$ 的散布矩阵。与之不同，却存在 $\mathrm{vec}(\boldsymbol{X})$ 的散布矩阵，并且散布矩阵包含矩阵 $\boldsymbol{X}$ 的所有元素的方差和协方差。令 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 是矩阵 $\boldsymbol{X}$ 的列，它们的协方差矩阵为

$$\mathrm{cov}\left(\boldsymbol{x}_i, \boldsymbol{x}_j^{\mathrm{T}}\right) = \boldsymbol{M}_{ij}$$

于是，向量化函数 $\mathrm{vec}(\boldsymbol{X})$ 的方差-协方差矩阵 $\mathrm{var}(\mathrm{vec}(\boldsymbol{X}))$ 是一分块矩阵，其子矩阵为 $\boldsymbol{M}_{ij}$，即有

$$\mathrm{var}(\mathrm{vec}(\boldsymbol{X})) = \{\boldsymbol{M}_{ij}\}$$

对于 $\boldsymbol{M}_{ij} = m_{ij}\boldsymbol{V}$ 的特殊情况，若 $\boldsymbol{M} = \{m_{ij}\}$，证明

(1) $\mathrm{var}(\mathrm{vec}(\boldsymbol{X})) = \boldsymbol{M} \otimes \boldsymbol{V}$。

(2) $\mathrm{var}(\mathrm{vec}(\boldsymbol{TX})) = \boldsymbol{M} \otimes \boldsymbol{TVT}^{\mathrm{T}}$。

(3) $\mathrm{var}(\mathrm{vec}(\boldsymbol{X}^{\mathrm{T}})) = \boldsymbol{V} \otimes \boldsymbol{M}$。

思路分析： 有关 Kronecker 积的证明可按照定义和性质这两个思路来分析。

证明： 令矩阵 $\boldsymbol{X}$ 是 $p \times q$ 维矩阵，则

(1) 因为 $\mathrm{var}(\mathrm{vec}(\boldsymbol{X})) = \{\boldsymbol{M}_{ij}\} = \{m_{ij}\boldsymbol{V}\}$ 成立，且 $\boldsymbol{M} = \{m_{ij}\}$，所以根据 Kronecker 积的定义，有 $\mathrm{var}(\mathrm{vec}(\boldsymbol{X})) = \boldsymbol{M} \otimes \boldsymbol{V}$。

(2) 令 $\boldsymbol{Y} \stackrel{\mathrm{def}}{=} \boldsymbol{TX}$，则 $\boldsymbol{Y}$ 的任意两列的协方差矩阵可表示为

$$\mathrm{cov}(\boldsymbol{y}_i, \boldsymbol{y}_j^{\mathrm{T}}) = \mathrm{cov}(\boldsymbol{Tx}_i, (\boldsymbol{Tx}_j)^{\mathrm{T}}) = \boldsymbol{T}\mathrm{cov}(\boldsymbol{x}_i, \boldsymbol{x}_j^{\mathrm{T}})\boldsymbol{T}^{\mathrm{T}} = \boldsymbol{TM}_{ij}\boldsymbol{T}^{\mathrm{T}}$$

从而可得

$$\mathrm{var}(\mathrm{vec}(\boldsymbol{Y})) = \mathrm{var}(\mathrm{vec}(\boldsymbol{TX})) = \{\boldsymbol{TM}_{ij}\boldsymbol{T}^{\mathrm{T}}\} = \{m_{ij}\boldsymbol{TVT}^{\mathrm{T}}\} = \boldsymbol{M} \otimes \boldsymbol{TVT}^{\mathrm{T}}$$

(3) 由题意，矩阵 $\boldsymbol{X}$ 的任意两个元素 x_{ki} 和 x_{lj} 的协方差可表示为 $\mathrm{cov}(x_{ki}, x_{lj}) = m_{ij}v_{kl}$，其中 v_{kl} 是矩阵 $\boldsymbol{V}$ 的元素。矩阵 $\boldsymbol{X}^{\mathrm{T}}$ 的第 i 行第 k 列元素 $x'_{ik} = x_{ki}$，第 j 行第 l 列元素 $x'_{jl} = x_{lj}$，故 $\mathrm{cov}(x'_{ik}, x'_{jl}) = v_{kl}m_{ij}$，于是有 $\mathrm{var}(\mathrm{vec}(\boldsymbol{X}^{\mathrm{T}})) = \{v_{kl}\boldsymbol{M}\} = \boldsymbol{V} \otimes \boldsymbol{M}$。

1.70 证明：

(1) $\mathrm{vec}(\boldsymbol{xy}^{\mathrm{T}}) = \boldsymbol{y} \otimes \boldsymbol{x}$。

(2) $\mathrm{vec}(\boldsymbol{A} \otimes \boldsymbol{b}) = \mathrm{vec}(\boldsymbol{A}) \otimes \boldsymbol{b}$。

(3) $\mathrm{vec}(\boldsymbol{a}_{p\times 1}^{\mathrm{T}} \otimes \boldsymbol{B}_{m\times n}) = (\boldsymbol{I}_{pn} \otimes \boldsymbol{I}_m)(\boldsymbol{a} \otimes \mathrm{vec}(\boldsymbol{B}))$。

证明： (1) 设 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 分别是 $m \times 1$ 维和 $n \times 1$ 维列向量，则向量 $\mathrm{vec}(\boldsymbol{xy}^{\mathrm{T}})$ 中的第 $(i-1)m+j$ 个元素为 $y_i x_j$ $(1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m)$，因此 $\mathrm{vec}(\boldsymbol{xy}^{\mathrm{T}}) = \boldsymbol{y} \otimes \boldsymbol{x}$ 成立。

(2) 设 $\boldsymbol{A}$ 是 $m \times n$ 维矩阵，$\boldsymbol{b}$ 是 $p \times 1$ 维列向量。由于元素 $a_{ij}b_k$ $(1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n, 1 \leqslant k \leqslant p)$ 是向量 $\mathrm{vec}(\boldsymbol{A} \otimes \boldsymbol{b})$ 的第 $(j-1)mp+(i-1)p+k$ 个元素，也是向量 $\mathrm{vec}(\boldsymbol{A}) \otimes \boldsymbol{b}$ 的第 $(j-1)mp+(i-1)p+k$ 个元素，故有 $\mathrm{vec}(\boldsymbol{A} \otimes \boldsymbol{b}) = \mathrm{vec}(\boldsymbol{A}) \otimes \boldsymbol{b}$。

(3) 由于元素 $a_k b_{ij}$ $(1 \leqslant k \leqslant p, 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n)$ 是 $\mathrm{vec}(\boldsymbol{a}_{p\times 1}^{\mathrm{T}} \otimes \boldsymbol{B}_{m\times n})$ 的第 $(k-1)mn+(j-1)m+i$ 个元素，也是 $\boldsymbol{a} \otimes \mathrm{vec}(\boldsymbol{B})$ 的第 $(k-1)mn+(j-1)m+i$ 个元素，且 $\boldsymbol{I}_{pn} \otimes \boldsymbol{I}_m = \boldsymbol{I}_{pmn}$，故 $\mathrm{vec}(\boldsymbol{a}_{p\times 1}^{\mathrm{T}} \otimes \boldsymbol{B}_{m\times n}) = (\boldsymbol{I}_{pn} \otimes \boldsymbol{I}_m)(\boldsymbol{a} \otimes \mathrm{vec}(\boldsymbol{B}))$。

1.71 证明：

$$\mathrm{vec}(\boldsymbol{PQ}) = (\boldsymbol{Q}^{\mathrm{T}} \otimes \boldsymbol{P})\mathrm{vec}(\boldsymbol{I}) = (\boldsymbol{Q}^{\mathrm{T}} \otimes \boldsymbol{I})\mathrm{vec}(\boldsymbol{P}) = (\boldsymbol{I} \otimes \boldsymbol{P})\mathrm{vec}(\boldsymbol{Q})$$

证明： 令 $\boldsymbol{A}_{m\times p}, \boldsymbol{B}_{p\times q}, \boldsymbol{C}_{q\times n}$，首先证明 $\mathrm{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}^{\mathrm{T}} \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{B})$。

令 $\boldsymbol{B} = [\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_q]$，并且 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_q$ 是 $p \times 1$ 维基本向量，即 $\boldsymbol{B} = \sum\limits_{i=1}^{q} \boldsymbol{b}_i \boldsymbol{e}_i^{\mathrm{T}}$。于是由性质 $\mathrm{vec}(\boldsymbol{ab}^{\mathrm{T}}) = \boldsymbol{b} \otimes \boldsymbol{a}$ 知，

$$\begin{aligned}
\mathrm{vec}(\boldsymbol{ABC}) &= \mathrm{vec}\left(\sum_{i=1}^{q} \boldsymbol{A}\boldsymbol{b}_i \boldsymbol{e}_i^{\mathrm{T}} \boldsymbol{C}\right) = \sum_{i=1}^{q} \mathrm{vec}\left((\boldsymbol{A}\boldsymbol{b}_i)(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{e}_i)^{\mathrm{T}}\right) \\
&= \sum_{i=1}^{q}\left(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{e}_i \otimes \boldsymbol{A}\boldsymbol{b}_i\right) = (\boldsymbol{C}^{\mathrm{T}} \otimes \boldsymbol{A})\sum_{i=1}^{q}(\boldsymbol{e}_i \otimes \boldsymbol{b}_i) \\
&= (\boldsymbol{C}^{\mathrm{T}} \otimes \boldsymbol{A})\sum_{i=1}^{q} \mathrm{vec}(\boldsymbol{b}_i \boldsymbol{e}_i^{\mathrm{T}}) = (\boldsymbol{C}^{\mathrm{T}} \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{B})
\end{aligned}$$

令 $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ 分别为 $\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{I}$ 的各种组合，容易得到

$$\mathrm{vec}(\boldsymbol{PQ}) = (\boldsymbol{Q}^{\mathrm{T}} \otimes \boldsymbol{P})\mathrm{vec}(\boldsymbol{I}) = (\boldsymbol{Q}^{\mathrm{T}} \otimes \boldsymbol{I})\mathrm{vec}(\boldsymbol{P}) = (\boldsymbol{I} \otimes \boldsymbol{P})\mathrm{vec}(\boldsymbol{Q})$$

1.72 验证 $\mathrm{tr}(\boldsymbol{XA}) = (\mathrm{vec}(\boldsymbol{A}^{\mathrm{T}}))^{\mathrm{T}}\mathrm{vec}(\boldsymbol{X}) = (\mathrm{vec}(\boldsymbol{X}^{\mathrm{T}}))^{\mathrm{T}}\mathrm{vec}(\boldsymbol{A})$。求 $\mathrm{tr}(\boldsymbol{XA})$ 关于 $\mathrm{vec}(\boldsymbol{X})$ 的偏导，证明：

$$\frac{\partial}{\partial \boldsymbol{X}}\mathrm{tr}(\boldsymbol{XA}) = \boldsymbol{A}^{\mathrm{T}}$$

思路分析： 按照迹和向量化函数的定义求解，其中求偏导的时候会用到迹函数的梯度矩阵：

$$\frac{\partial \mathrm{tr}(\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}})}{\partial \boldsymbol{x}} = \frac{\partial \mathrm{tr}(\boldsymbol{y}\boldsymbol{x}^{\mathrm{T}})}{\partial \boldsymbol{x}} = \boldsymbol{y}$$

解： (1) 令

$$\boldsymbol{X}_{m\times n} = \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_m \end{bmatrix}, \quad \boldsymbol{A}_{n\times m} = [\boldsymbol{a}_1, \cdots, \boldsymbol{a}_m]$$

根据迹的定义，易知 $\mathrm{tr}(\boldsymbol{XA}) = \sum\limits_{i=1}^{m} \boldsymbol{x}_i \boldsymbol{a}_i$。而

$$\left(\mathrm{vec}(\boldsymbol{X}^{\mathrm{T}})\right)^{\mathrm{T}} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m], \quad \mathrm{vec}(\boldsymbol{A}) = \begin{bmatrix} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_m \end{bmatrix}$$

所以 $(\text{vec}(\boldsymbol{X}^{\mathrm{T}}))^{\mathrm{T}}\text{vec}(\boldsymbol{A}) = \sum_{i=1}^{m} \boldsymbol{x}_i \boldsymbol{a}_i$，即 $\text{tr}(\boldsymbol{XA}) = (\text{vec}(\boldsymbol{X}^{\mathrm{T}}))^{\mathrm{T}}\text{vec}(\boldsymbol{A})$。

又由 $\text{tr}(\boldsymbol{XA}) = \text{tr}(\boldsymbol{AX})$ 知，$\text{tr}(\boldsymbol{XA}) = (\text{vec}(\boldsymbol{A}^{\mathrm{T}}))^{\mathrm{T}}\text{vec}(\boldsymbol{X})$。

(2) 下面求 $\text{tr}(\boldsymbol{XA})$ 关于 $\text{vec}(\boldsymbol{X})$ 的偏导：

$$\frac{\partial}{\partial \text{vec}(\boldsymbol{X})}\text{tr}(\boldsymbol{XA}) = \frac{\partial}{\partial \text{vec}(\boldsymbol{X})}\left[\left(\text{vec}(\boldsymbol{A}^{\mathrm{T}})\right)^{\mathrm{T}} \text{vec}(\boldsymbol{X})\right] = \text{vec}(\boldsymbol{A}^{\mathrm{T}})$$

(3) 利用 (2) 中结果，可得

$$\frac{\partial}{\partial \boldsymbol{X}}\text{tr}(\boldsymbol{XA}) = \text{unvec}_{n,m}\left[\frac{\partial}{\partial \text{vec}(\boldsymbol{X})}\text{tr}(\boldsymbol{XA})\right] = \text{unvec}_{n,m}\left[\text{vec}(\boldsymbol{A}^{\mathrm{T}})\right] = \boldsymbol{A}^{\mathrm{T}}$$

1.73 设 $\boldsymbol{A}, \boldsymbol{G}$ 和 $\boldsymbol{H}$ 分别是 $m \times n, n \times m$ 和 $n \times p$ 矩阵。证明：若 $\text{rank}(\boldsymbol{A}) = \text{rank}(\boldsymbol{AH})$，则 $\boldsymbol{GAH} = \boldsymbol{H} \Longrightarrow \boldsymbol{G} = \boldsymbol{A}^{-}$。

思路分析： 广义逆矩阵的证明大多根据它的定义去寻找线索。本题即是从广义逆矩阵是一致方程的解这个定义（教材P75定义1.8.3）出发，根据教材的定理1.8.5来证明的。

证明： 由 $\text{rank}(\boldsymbol{A}) = \text{rank}(\boldsymbol{AH})$ 知，$\text{range}(\boldsymbol{A}) = \text{range}(\boldsymbol{AH})$。又

$$\boldsymbol{GAH} = \boldsymbol{H} \quad \Longrightarrow \quad \boldsymbol{AG}(\boldsymbol{AH}) = \boldsymbol{AH}$$

即是说对 $\forall\ \boldsymbol{y} \in \text{range}(\boldsymbol{AH}) = \text{range}(\boldsymbol{A})$，方程 $\boldsymbol{Ax} = \boldsymbol{y}$ 都有解 $\boldsymbol{x} = \boldsymbol{Gy}$。由定理 1.8.5 知，一致方程 $\boldsymbol{Ax} = \boldsymbol{y}$ 对 $\boldsymbol{y} \neq \boldsymbol{0}$ 有解 $\boldsymbol{x} = \boldsymbol{Gy}$ 的充分必要条件是 $\boldsymbol{AGA} = \boldsymbol{A}$，即 $\boldsymbol{G} = \boldsymbol{A}^{-}$ 是 $\boldsymbol{A}$ 的广义逆矩阵。

1.74 令 $\boldsymbol{I}$ 为 $n \times n$ 单位矩阵，$\boldsymbol{J}_{n\times n}$ 是全部元素等于 1 的矩阵。若 $a + (n-1)b = 0$，证明 $(a-b)^{-1}\boldsymbol{I}$ 是矩阵 $(a-b)\boldsymbol{I} + b\boldsymbol{J}$ 的满足定义 $\boldsymbol{AGA} = \boldsymbol{A}$ 的广义逆矩阵。

证明： 令 $\boldsymbol{A} = (a-b)\boldsymbol{I} + b\boldsymbol{J}$，$\boldsymbol{G} = (a-b)^{-1}\boldsymbol{I}$，则

$$\begin{aligned}\boldsymbol{AGA} &= [(a-b)\boldsymbol{I} + b\boldsymbol{J}]\,(a-b)^{-1}\boldsymbol{I}\,[(a-b)\boldsymbol{I} + b\boldsymbol{J}] \\ &= (a-b)^{-1}\left[(a-b)^2\boldsymbol{I} + 2(a-b)b\boldsymbol{J} + b^2\boldsymbol{J}^2\right]\end{aligned}$$

因为 $a + (n-1)b = 0$，所以 $a - b = -nb$。另外，由于 $\boldsymbol{J}$ 是全部元素为 1 的矩阵，因此 $\boldsymbol{J}^2 = n\boldsymbol{J}$。代入上式，可得

$$\boldsymbol{AGA} = -\frac{1}{nb}\left[n^2b^2\boldsymbol{I} - nb^2\boldsymbol{J}\right] = -nb\boldsymbol{I} + b\boldsymbol{J} = (a-b)\boldsymbol{I} + b\boldsymbol{J} = \boldsymbol{A}$$

故 $(a-b)^{-1}\boldsymbol{I}$ 是矩阵 $(a-b)\boldsymbol{I} + b\boldsymbol{J}$ 的满足定义 $\boldsymbol{AGA} = \boldsymbol{A}$ 的广义逆矩阵。

1.75 一个对角矩阵 $\boldsymbol{H}$ 称为 Hermitian 标准型，若它的对角线元素仅由 0 和 1 组成。对于任意一个正方矩阵 $\boldsymbol{A}$，总是存在非奇异矩阵 $\boldsymbol{C}$ 使得 $\boldsymbol{CA} = \boldsymbol{H}$ 为 Hermitian 标准型。证明 $\boldsymbol{C} = \boldsymbol{A}^{-}$ 是矩阵 $\boldsymbol{A}$ 的广义逆矩阵。

证明： 若对角矩阵 $\boldsymbol{H}$ 是 Hermitian 标准型，则 $\boldsymbol{H}$ 必然为幂等矩阵，即 $\boldsymbol{H}^2=\boldsymbol{H}$。由于 $\boldsymbol{H}^2=(\boldsymbol{CA})^2=\boldsymbol{H}=\boldsymbol{CA}$，所以 $\boldsymbol{CA}$ 也是幂等矩阵。又因为 $\boldsymbol{C}$ 非奇异，所以 $\mathrm{rank}(\boldsymbol{CA})=\mathrm{rank}(\boldsymbol{A})$。

根据广义逆矩阵的定义 1.8.5 知，$\boldsymbol{C}$ 是矩阵 $\boldsymbol{A}$ 的广义逆矩阵。

1.76 令 $\boldsymbol{A}_{m\times n}$ 和 $\boldsymbol{R}_{p\times n}$ 是两个复矩阵，并且 $\boldsymbol{N}_{n\times q}$ 是一满足 $\boldsymbol{RN}=\boldsymbol{O}$ 的任意矩阵，其秩为 $n-\mathrm{rank}(\boldsymbol{R})$，记

$$\boldsymbol{D}=\boldsymbol{N}(\boldsymbol{N}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{AN})^{-}\boldsymbol{N}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}$$

$$\boldsymbol{E}=\boldsymbol{A}^{\mathrm{H}}-\boldsymbol{A}^{\mathrm{H}}\boldsymbol{AD}$$

证明

(1) 方程

$$\begin{bmatrix}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} & \boldsymbol{R}^{\mathrm{H}}\\ \boldsymbol{R} & \boldsymbol{O}\end{bmatrix}\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{Y}\end{bmatrix}=\begin{bmatrix}\boldsymbol{A}^{\mathrm{H}}\\ \boldsymbol{O}\end{bmatrix} \tag{1.95}$$

是一致方程。

(2) 若

$$\begin{bmatrix}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} & \boldsymbol{R}^{\mathrm{H}}\\ \boldsymbol{R} & \boldsymbol{O}\end{bmatrix}^{-}=\begin{bmatrix}\boldsymbol{C}_1 & \boldsymbol{C}_2^{\mathrm{H}}\\ \boldsymbol{C}_2 & \boldsymbol{C}_3\end{bmatrix}$$

则

① $\boldsymbol{C}_2^{\mathrm{H}}$ 是 $\boldsymbol{R}$ 的广义逆矩阵。

② $\boldsymbol{RC}_1\boldsymbol{R}^{\mathrm{H}}=\boldsymbol{O}$。

③ $\boldsymbol{AC}_1\boldsymbol{A}^{\mathrm{H}}$ 为幂等矩阵，即 $(\boldsymbol{AC}_1\boldsymbol{A}^{\mathrm{H}})^2=\boldsymbol{AC}_1\boldsymbol{A}^{\mathrm{H}}$。

证明： (1) 设 $\boldsymbol{X}=\boldsymbol{NZ}$，将矩阵方程展开，得到

$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{ANZ}+\boldsymbol{R}^{\mathrm{H}}\boldsymbol{Y}=\boldsymbol{A}^{\mathrm{H}} \tag{1.96}$$

$$\boldsymbol{RNZ}+\boldsymbol{OY}=\boldsymbol{O} \tag{1.97}$$

若上述方程组是关于未知矩阵 $\boldsymbol{Z}$ 的一致方程，则其必然是关于 $\boldsymbol{X}$ 的一致方程。方程式 (1.97) 中，由于 $\boldsymbol{RN}=\boldsymbol{O}$，它对任意的 $\boldsymbol{Z}$ 和 $\boldsymbol{Y}$ 均成立，因而考察方程组的一致性等价于考察方程式 (1.96) 的一致性。将式 (1.96) 两边左乘 $\boldsymbol{N}^{\mathrm{H}}$，可得

$$\boldsymbol{N}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{ANZ}=\boldsymbol{N}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}} \tag{1.98}$$

由于 $\mathrm{rank}(\boldsymbol{N}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{AN})=\mathrm{rank}(\boldsymbol{N}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}})=\mathrm{rank}([\boldsymbol{N}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{AN},\boldsymbol{N}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}])$，由定理 1.8.3 知，矩阵方程式 (1.98) 为一致方程，所以矩阵方程式 (1.95) 是一致方程。

(2) 由于矩阵方程式 (1.95) 是一致方程，根据定理 1.8.5，其解为

$$\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{Y}\end{bmatrix}=\begin{bmatrix}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} & \boldsymbol{R}^{\mathrm{H}}\\ \boldsymbol{R} & \boldsymbol{O}\end{bmatrix}^{-}\begin{bmatrix}\boldsymbol{A}^{\mathrm{H}}\\ \boldsymbol{O}\end{bmatrix}$$

将其代入式 (1.95)，得

$$\begin{bmatrix} A^{\mathrm{H}}A & R^{\mathrm{H}} \\ R & O \end{bmatrix} \begin{bmatrix} C_1 & C_2^{\mathrm{H}} \\ C_2 & C_3 \end{bmatrix} \begin{bmatrix} A^{\mathrm{H}} \\ O \end{bmatrix} = \begin{bmatrix} A^{\mathrm{H}} \\ O \end{bmatrix}$$

展开得到

$$A^{\mathrm{H}}AC_1A^{\mathrm{H}} + R^{\mathrm{H}}C_2A^{\mathrm{H}} = A^{\mathrm{H}} \tag{1.99}$$

$$RC_1A^{\mathrm{H}} = O \tag{1.100}$$

由式 (1.95) 中证明易知，矩阵方程

$$\begin{bmatrix} A^{\mathrm{H}}A & R^{\mathrm{H}} \\ R & O \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} R^{\mathrm{H}} \\ O \end{bmatrix}$$

也是一致方程，因此

$$\begin{bmatrix} A^{\mathrm{H}}A & R^{\mathrm{H}} \\ R & O \end{bmatrix} \begin{bmatrix} C_1 & C_2^{\mathrm{H}} \\ C_2 & C_3 \end{bmatrix} \begin{bmatrix} R^{\mathrm{H}} \\ O \end{bmatrix} = \begin{bmatrix} R^{\mathrm{H}} \\ O \end{bmatrix}$$

展开得到

$$A^{\mathrm{H}}AC_1R^{\mathrm{H}} + R^{\mathrm{H}}C_2R^{\mathrm{H}} = R^{\mathrm{H}} \tag{1.101}$$

$$RC_1R^{\mathrm{H}} = O \tag{1.102}$$

又因为 $\begin{bmatrix} A^{\mathrm{H}}A & R^{\mathrm{H}} \\ R & O \end{bmatrix}$ 是 Hermitian 矩阵，故 $C_1 = C_1^{\mathrm{H}}$。由式 (1.100)、式 (1.101) 和式 (1.102) 可得

$$RC_1R^{\mathrm{H}} = O$$

$$R^{\mathrm{H}}C_2R^{\mathrm{H}} = R^{\mathrm{H}}$$

即 $RC_2^{\mathrm{H}}R = R$，故 C_2^{H} 是 R 的广义逆矩阵。将式 (1.99) 的两边左乘 AC_1，并利用式 (1.100)，可得 $(AC_1A^{\mathrm{H}})^2 = AC_1A^{\mathrm{H}}$，即 AC_1A^{H} 是幂等矩阵。

1.77 令

$$T = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}$$

证明 $-(a_1^2 + a_2^2 + a_3^2)^{-1}T$ 是 T 的广义逆矩阵 T^-。

证明： 由 $\boldsymbol{T}\boldsymbol{T}^{-}\boldsymbol{T}=\boldsymbol{T}$ 可知，本题等价证明 $\boldsymbol{T}^3=-(a_1^2+a_2^2+a_3^2)\boldsymbol{T}$。因为

$$\begin{aligned}
\boldsymbol{T}^3 &= \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}\begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}\begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & a_3^3+a_2^2a_3+a_1^2a_3 & -a_2^3-a_3^2a_2-a_1^2a_2 \\ -a_3^3-a_2^2a_3-a_1^2a_3 & 0 & a_1a_2^2+a_1a_3^2+a_1^3 \\ a_2^3+a_3^2a_2+a_1^2a_2 & -a_1a_2^2-a_1a_3^2-a_1^3 & 0 \end{bmatrix} \\
&= -(a_1^2+a_2^2+a_3^2)\begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \\
&= -(a_1^2+a_2^2+a_3^2)\boldsymbol{T}
\end{aligned}$$

故命题得证。

1.78　利用矩阵的满秩分解，求矩阵

$$\boldsymbol{A}=\begin{bmatrix} 1 & 2 & 4 & 3 \\ 3 & -1 & 2 & -2 \\ 5 & -4 & 0 & -7 \end{bmatrix},\qquad \boldsymbol{B}=\begin{bmatrix} 1 & 2 & -1 \\ 3 & 2 & 1 \\ -1 & -2 & -1 \\ 3 & 5 & 4 \end{bmatrix}$$

的广义逆矩阵 $\boldsymbol{A}^-$ 和 $\boldsymbol{B}^-$。

解： (1) 对矩阵 $\boldsymbol{A}$ 作行变换化为阶梯型的过程如下：

$$\begin{bmatrix} 1 & 2 & 4 & 3 \\ 3 & -1 & 2 & -2 \\ 5 & -4 & 0 & -7 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_2-3r_1}} \begin{bmatrix} 1 & 2 & 4 & 3 \\ 0 & -7 & -10 & -11 \\ 5 & -4 & 0 & -7 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_3-5r_1}}$$

$$\begin{bmatrix} 1 & 2 & 4 & 3 \\ 0 & -7 & -10 & -11 \\ 0 & -14 & -20 & -22 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_3-2r_2}} \begin{bmatrix} 1 & 2 & 4 & 3 \\ 0 & -7 & -10 & -11 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

对 $\boldsymbol{I}$ 作相应的初等行反变换如下：

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_3+2r_2}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_3+5r_1}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 5 & 2 & 1 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_2+3r_1}} \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 5 & 2 & 1 \end{bmatrix}$$

所以，$\boldsymbol{A}$ 的满秩分解 $\boldsymbol{A}=\boldsymbol{F}\boldsymbol{G}$ 中，

$$\boldsymbol{G}_{2\times 4}=\begin{bmatrix} 1 & 2 & 4 & 3 \\ 0 & -7 & -10 & -11 \end{bmatrix},\quad \boldsymbol{F}_{3\times 2}=\begin{bmatrix} 1 & 0 \\ 3 & 1 \\ 5 & 2 \end{bmatrix}$$

于是可得广义逆矩阵

$$\boldsymbol{A}^-=\boldsymbol{G}^{\mathrm{T}}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{G}^{\mathrm{T}})^{-1}\boldsymbol{F}^{\mathrm{T}}=\begin{bmatrix} 0.0687 & 0.0603 & 0.0518 \\ 0.0386 & 0.0019 & -0.0348 \\ 0.1337 & 0.0716 & 0.0094 \\ 0.0508 & -0.0056 & -0.0621 \end{bmatrix}$$

(2) 对 $\boldsymbol{B}$ 作初等行变换如下：

$$\begin{bmatrix} 1 & 2 & -1 \\ 3 & 2 & 1 \\ -1 & -2 & -1 \\ 3 & 5 & 4 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_4-r_2}} \begin{bmatrix} 1 & 2 & -1 \\ 3 & 2 & 1 \\ -1 & -2 & -1 \\ 0 & 3 & 3 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_2+3r_3}} \begin{bmatrix} 1 & 2 & -1 \\ 0 & -4 & -2 \\ -1 & -2 & -1 \\ 0 & 3 & 3 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_2\times\frac{1}{2}}}$$

$$\begin{bmatrix} 1 & 2 & -1 \\ 0 & -2 & -1 \\ -1 & -2 & -1 \\ 0 & 3 & 3 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_3-r_2}} \begin{bmatrix} 1 & 2 & -1 \\ 0 & -2 & -1 \\ -1 & 0 & 0 \\ 0 & 3 & 3 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_1+r_3}} \begin{bmatrix} 0 & 2 & -1 \\ 0 & -2 & -1 \\ -1 & 0 & 0 \\ 0 & 3 & 3 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_1-r_3}}$$

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & -2 & -1 \\ 0 & 2 & -1 \\ 0 & 3 & 3 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_3+r_2}} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -2 & -1 \\ 0 & 0 & -2 \\ 0 & 3 & 3 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_4+r_2+r_3}} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -2 & -1 \\ 0 & 0 & -2 \\ 0 & 1 & 0 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_2\leftrightarrow r_4}}$$

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \\ 0 & -2 & -1 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_3\times\frac{1}{2}}} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & -2 & -1 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_4+2r_2-r_3}} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

对 $\boldsymbol{I}$ 作相应的初等行反变换如下：

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_4-2r_2+r_3}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -2 & 1 & 1 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_3\times 2}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & -2 & 1 & 1 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_2\leftrightarrow r_4}}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -2 & 1 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_4-r_2-r_3}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -2 & 1 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 3 & -3 & -1 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_3-r_2}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -2 & 1 & 1 \\ 0 & 2 & 1 & -1 \\ 0 & 3 & -3 & -1 \end{bmatrix}$$

$$\xrightarrow{\boldsymbol{E}_{r_1\leftrightarrow r_3}} \begin{bmatrix} 0 & 2 & 1 & -1 \\ 0 & -2 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 3 & -3 & -1 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_1-r_3}} \begin{bmatrix} -1 & 2 & 1 & -1 \\ 0 & -2 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 3 & -3 & -1 \end{bmatrix}$$

$$\xrightarrow{\boldsymbol{E}_{r_3+r_2}} \begin{bmatrix} -1 & 2 & 1 & -1 \\ 0 & -2 & 1 & 1 \\ 1 & -2 & 1 & 1 \\ 0 & 3 & -3 & -1 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_2\times 2}} \begin{bmatrix} -1 & 2 & 1 & -1 \\ 0 & -4 & 2 & 2 \\ 1 & -2 & 1 & 1 \\ 0 & 3 & -3 & -1 \end{bmatrix}$$

$$\xrightarrow{\boldsymbol{E}_{r_2-3r_3}} \begin{bmatrix} -1 & 2 & 1 & -1 \\ -3 & 2 & -1 & -1 \\ 1 & -2 & 1 & 1 \\ 0 & 3 & -3 & -1 \end{bmatrix} \xrightarrow{\boldsymbol{E}_{r_4+r_2}} \begin{bmatrix} -1 & 2 & 1 & -1 \\ -3 & 2 & -1 & -1 \\ 1 & -2 & 1 & 1 \\ -3 & 5 & -4 & 2 \end{bmatrix}$$

故 $\boldsymbol{B}$ 的满秩分解 $\boldsymbol{B}=\boldsymbol{F}\boldsymbol{G}$ 中

$$\boldsymbol{G}_{3\times 3}=\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad \boldsymbol{F}_{4\times 3}=\begin{bmatrix} -1 & 2 & 1 \\ -3 & 2 & -1 \\ 1 & -2 & 1 \\ -3 & 5 & -4 \end{bmatrix}$$

于是可得广义逆矩阵

$$B^{-}=G^{\mathrm{T}}(F^{\mathrm{T}}BG^{\mathrm{T}})^{-1}F^{\mathrm{T}}=\begin{bmatrix}-0.1&0.5333&0.1&-0.1333\\0.3&-0.2667&-0.3&0.0667\\-0.3&-0.0667&0.3&0.2667\end{bmatrix}$$

1.79 设 $KA=O$ 和 $K^2=K$，且 K 非奇异，证明矩阵 A 的广义逆矩阵 $A^{-}=(A-K)^{-1}$。

证明： 令 $G=A-K$，首先证明矩阵 G 是可逆的。由已知条件知

$$KA-K^2=-K \implies K(A-K)=-K \implies KG=-K$$

若 G 不可逆，则 $\det(KG)=\det(K)\det(G)=0\neq\det(-K)$，故 G 必然是可逆的。同时，有 $KG^{-1}=-K$。

由 $G=A-K$ 知，$A=G+K$。所以，

$$\begin{aligned}A(A-K)^{-}A=AG^{-1}A&=(G+K)G^{-1}(G+K)=(I+KG^{-1})(G+K)\\&=(I-K)(G+K)=G+K-KG-K^2\\&=G+K=A\end{aligned}$$

故 $A^{-}=(A-K)^{-1}$ 是 A 的广义逆矩阵。

1.80 已知 $K^2=K$，且 Z^{-} 是矩阵 $Z=KAK$ 的广义逆矩阵。证明：$KZ^{-}K$ 也是 Z 的一个广义逆矩阵。

证明： 因为 Z^{-} 是 $Z=KAK$ 的广义逆矩阵，所以

$$\begin{aligned}&ZZ^{-}Z=Z \implies KAKZ^{-}KAK=Z \implies KAK^2Z^{-}K^2AK=Z\\&\implies KAK(KZ^{-}K)KAK=Z \implies Z(KZ^{-}K)Z=Z\end{aligned}$$

故 $KZ^{-}K$ 也是 Z 的一个广义逆矩阵。

1.81 设 G 是A 的一个广义逆矩阵。证明：它也是矩阵 AG 的一个广义逆矩阵，当且仅当 G^2 是 A 的一个广义逆矩阵。

证明： 首先证明充分性。因为 G^2 是 A 的一个广义逆矩阵，所以有 $AG^2A=A$。两边同时右乘 G，得到

$$AG^2AG=AG \implies (AG)G(AG)=AG$$

所以 G 是矩阵 AG 的一个广义逆矩阵。

现在证明必要性。因为 G 是 AG 的广义逆矩阵，所以有 $AGGAG=AG$。两边同时右乘 A，并利用 $AGA=A$，可得 $AG^2A=A$，即 G^2 是 A 的一个广义逆矩阵。命题得证。

1.82 满足 Moore-Penrose 逆矩阵两个条件

$$AGA = A, \quad GAG = G$$

的矩阵 G 称为矩阵 A 的自反广义逆矩阵 (reflexive generalized inverse)。证明：$\text{rank}(G) = \text{rank}(A)$ 对 $AGA = A$ 成立，当且仅当 G 是 A 的一个自反广义逆矩阵。

思路分析： 自反广义逆矩阵的证明也是按照定义来求解。

证明： (1) 充分性：由 G 是 A 的自反广义逆矩阵知，$AGA = A$，所以

$$\text{rank}(A) = \text{rank}(AGA) \leqslant \text{rank}(G)$$

而由 $GAG = G$，又有

$$\text{rank}(G) = \text{rank}(GAG) \leqslant \text{rank}(A)$$

所以 $\text{rank}(G) = \text{rank}(A)$，即充分性得证。

(2) 必要性：由于已知 $AGA = A$，所以欲证 G 是 A 的一个自反广义逆矩阵，仅需证明 $GAG = G$，即 A 是 G 的一个广义逆矩阵。因为

$$(AG)^2 = AGAG = AG$$

所以 AG 是幂等矩阵。又因为 $AGA = A$，即 G 是 A 的广义逆矩阵，所以

$$\text{rank}(AG) = \text{rank}(A) = \text{rank}(G)$$

根据广义逆矩阵的定义 1.8.5，A 是 G 的广义逆矩阵，即 $GAG = G$。

1.83 令

$$G = Q\begin{bmatrix} I_r & U \\ V & VU \end{bmatrix} P$$

其中，I_r 为单位矩阵，P 和 Q 非奇异，并且

$$PAQ = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}, \quad O \text{ 为零矩阵}$$

证明 G 是 A 的自反广义逆矩阵。

证明： 因为

$$\begin{aligned} GAG &= Q\begin{bmatrix} I_r & U \\ V & VU \end{bmatrix} PAQ \begin{bmatrix} I_r & U \\ V & VU \end{bmatrix} P \\ &= Q\begin{bmatrix} I_r & U \\ V & VU \end{bmatrix}\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}\begin{bmatrix} I_r & U \\ V & VU \end{bmatrix} P \\ &= Q\begin{bmatrix} I_r & U \\ V & VU \end{bmatrix} P \\ &= G \end{aligned}$$

又由于 $\boldsymbol{P},\boldsymbol{Q}$ 非奇异，所以

$$\operatorname{rank}(\boldsymbol{G})=\operatorname{rank}\left(\begin{bmatrix}\boldsymbol{I}_r & \boldsymbol{U}\\ \boldsymbol{V} & \boldsymbol{VU}\end{bmatrix}\right)=r=\operatorname{rank}(\boldsymbol{A})$$

由题 1.82 知，$\boldsymbol{G}$ 是 $\boldsymbol{A}$ 的自反广义逆矩阵。

1.84 证明

(1) 所有左和右逆矩阵都是自反广义逆矩阵，它们分别满足 Moore-Penrose 对称条件 $\boldsymbol{AGA}=\boldsymbol{A}$ 和 $\boldsymbol{GAG}=\boldsymbol{G}$之中的一个条件。

(2) 一个满行 (列) 秩矩阵 $\boldsymbol{A}$ 的所有广义逆矩阵 $\boldsymbol{A}^-$ 都是右 (左) 逆矩阵。

证明： (1) 令左伪逆矩阵 $\boldsymbol{L}$，右伪逆矩阵 $\boldsymbol{R}$。由于

$$\boldsymbol{LA}=\boldsymbol{I}\quad\Rightarrow\quad\boldsymbol{ALA}=\boldsymbol{A},\ \ \boldsymbol{LAL}=\boldsymbol{L}$$

$$\boldsymbol{AR}=\boldsymbol{I}\quad\Rightarrow\quad\boldsymbol{ARA}=\boldsymbol{A},\ \ \boldsymbol{RAR}=\boldsymbol{R}$$

从而命题得证。

(2) 因为 $\boldsymbol{A}$ 满行秩，所以 $\boldsymbol{AA}^{\mathrm{H}}$ 非奇异。于是，

$$\boldsymbol{AA}^-\boldsymbol{A}=\boldsymbol{A}\quad\Rightarrow\quad\boldsymbol{AA}^-\boldsymbol{AA}^{\mathrm{H}}=\boldsymbol{AA}^{\mathrm{H}}\quad\Rightarrow\quad\boldsymbol{AA}^-=\boldsymbol{I}$$

即 $\boldsymbol{A}^-$ 是右逆矩阵。同理，当 $\boldsymbol{A}$ 满列秩时，易证 $\boldsymbol{A}^-$ 是左逆矩阵。

1.85 求 3×1 向量 $\boldsymbol{a}=[1,5,7]^{\mathrm{T}}$ 的 Moore-Penrose 逆矩阵。

解： 由于 $\boldsymbol{a}$ 满列秩，故其 Moore-Penrose 逆矩阵即为 $\boldsymbol{a}$ 的左伪逆矩阵，即

$$\boldsymbol{a}^{\dagger}=(\boldsymbol{a}^{\mathrm{H}}\boldsymbol{a})^{-1}\boldsymbol{a}^{\mathrm{H}}=\left(\begin{bmatrix}1 & 5 & 7\end{bmatrix}\begin{bmatrix}1\\5\\7\end{bmatrix}\right)^{-1}\begin{bmatrix}1\\5\\7\end{bmatrix}=\begin{bmatrix}\frac{1}{75} & \frac{1}{15} & \frac{7}{75}\end{bmatrix}^{\mathrm{T}}$$

1.86 证明 $\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}$ 是 $\boldsymbol{AA}^{\mathrm{T}}$ 的 Moore-Penrose 逆矩阵。

思路分析： 证明某个矩阵是另一个矩阵的 Moore-Penrose 逆矩阵，只需验证该矩阵是否满足 Moore-Penrose 逆矩阵定义的四个条件即可。

证明： (1) $(\boldsymbol{AA}^{\mathrm{T}})[\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}](\boldsymbol{AA}^{\mathrm{T}})=\boldsymbol{AA}^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{AA}^{\mathrm{T}}=\boldsymbol{AA}^{\mathrm{T}}$。

(2) $[\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}](\boldsymbol{AA}^{\mathrm{T}})[\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}]=\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{AA}^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}$。

(3) $(\boldsymbol{AA}^{\mathrm{T}})[\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}]=\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$ 是 Hermitian 矩阵。

(4) $[\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}](\boldsymbol{AA}^{\mathrm{T}})=\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$ 是 Hermitian 矩阵。

故 $\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-2}\boldsymbol{A}^{\mathrm{T}}$ 是 $\boldsymbol{AA}^{\mathrm{T}}$ 的 Moore-Penrose 逆矩阵。

1.87 证明关于 Moore-Penrose 逆矩阵的下列定义条件 (1) 和条件 (2) 等价：

(1) $AGA = A,\quad GAG = G,\quad (AG)^{\#} = AG,\quad (GA)^{\#} = GA$。

(2) $A^{\#}AG = A^{\#},\quad G^{\#}GA = G^{\#}$。

证明： 首先证明从条件 (1) 可以导出条件 (2)：

$$A^{\#}AG = A^{\#}(AG)^{\#} = (AGA)^{\#} = A^{\#}$$
$$G^{\#}GA = G^{\#}(GA)^{\#} = (GAG)^{\#} = G^{\#}$$

然后证明从条件 (2) 可以导出条件 (1)：根据定理 1.9.3，满足条件 (1) 的矩阵 G 必然使得 Gy 是非一致方程 $Ax = y$ 的最小范数最小二乘解。又根据定理 1.8.8，非一致方程 $Ax = y$ 的最小二乘解的通解形式为 $Gy + (I - GA)z$，所以

$$\begin{aligned} & \| Gy \| \leqslant \| Gy + (I - GA)z \|, \quad \forall y, z \\ \Rightarrow\quad & \langle Gy, (I - GA)z\rangle = 0, \quad \forall y, z \\ \Rightarrow\quad & G^{\#}(I - GA) = O \\ \Rightarrow\quad & G^{\#} = G^{\#}GA \end{aligned}$$

由 G 和 A 的对称性，亦可得到 $A^{\#}AG = A^{\#}$。

综上所述，条件 (1) 和条件 (2) 等价。

1.88 设 A 是一对称矩阵，并且 M 是 A 的 Moore-Penrose 逆矩阵。证明：矩阵 M^2 是 A^2 的 Moore-Penrose 逆矩阵。

证明： 由于 M 是 A 的 Moore-Penrose 逆矩阵，所以它满足条件：

$$AMA = A,\quad MAM = M,\quad (AM)^{\#} = AM,\quad (MA)^{\#} = MA$$

(1)
$$\begin{aligned} A^2M^2A^2 &= AAMMAA = A(AM)^{\#}(MA)^{\#}A = AM^{\#}AAM^{\#}A \\ &= (AM^{\#}A)^2 = [(AMA)^{\#}]^2 = (A^{\#})^2 \\ &= A^2 \end{aligned}$$

(2)
$$\begin{aligned} M^2A^2M^2 &= MMAAMM = M(MA)^{\#}AMM = MA^{\#}M^{\#}AMM \\ &= MA^{\#}M^{\#}A^{\#}MM = MA^{\#}MM = MAMM \\ &= M^2 \end{aligned}$$

(3) 因为

$$\begin{aligned} (M^2A^2)^{\#} &= A^{\#}A^{\#}M^{\#}M^{\#} = A^{\#}(MA)^{\#}M^{\#} = A^{\#}MAM^{\#} \\ &= AMAM^{\#} = AM^{\#} = (MA)^{\#} = MA \\ M^2A^2 &= MMAA = M(MA)^{\#}A = MA^{\#}M^{\#}A \\ &= M(AMA)^{\#} = MA^{\#} = MA \end{aligned}$$

所以，$(\boldsymbol{M}^2\boldsymbol{A}^2)^{\#}=\boldsymbol{M}^2\boldsymbol{A}^2$。

(4) 因为

$$\begin{aligned}(\boldsymbol{A}^2\boldsymbol{M}^2)^{\#}&=\boldsymbol{M}^{\#}\boldsymbol{M}^{\#}\boldsymbol{A}^{\#}\boldsymbol{A}^{\#}=\boldsymbol{M}^{\#}(\boldsymbol{A}\boldsymbol{M})^{\#}\boldsymbol{A}^{\#}\\&=\boldsymbol{M}^{\#}\boldsymbol{A}\boldsymbol{M}\boldsymbol{A}^{\#}=\boldsymbol{M}^{\#}\boldsymbol{A}=(\boldsymbol{A}\boldsymbol{M})^{\#}=\boldsymbol{A}\boldsymbol{M}\\\boldsymbol{A}^2\boldsymbol{M}^2&=\boldsymbol{A}\boldsymbol{A}\boldsymbol{M}\boldsymbol{M}=\boldsymbol{A}(\boldsymbol{A}\boldsymbol{M})^{\#}\boldsymbol{M}=\boldsymbol{A}\boldsymbol{M}^{\#}\boldsymbol{A}^{\#}\boldsymbol{M}\\&=(\boldsymbol{A}\boldsymbol{M}\boldsymbol{A})^{\#}\boldsymbol{M}=\boldsymbol{A}\boldsymbol{M}\end{aligned}$$

所以，$(\boldsymbol{A}^2\boldsymbol{M}^2)^{\#}=\boldsymbol{A}^2\boldsymbol{M}^2$。

综上所述，矩阵 $\boldsymbol{M}^2$ 满足 $\boldsymbol{A}^2$ 的 Moore-Penrose 逆矩阵的四个条件，命题得证。

1.89 已知矩阵

$$\boldsymbol{A}=\begin{bmatrix}1&0&-1&1\\0&2&2&2\\-1&4&5&3\end{bmatrix}$$

用 KL 分解法求 Moore-Penrose 逆矩阵 $\boldsymbol{A}^{\dagger}$。

解： 首先进行 $\boldsymbol{A}$ 的满秩分解。利用行初等变换将 $\boldsymbol{A}$ 化为阶梯型，即

$$\boldsymbol{A}=\begin{bmatrix}1&0&-1&1\\0&2&2&2\\-1&4&5&3\end{bmatrix}\xrightarrow{\boldsymbol{E}_{r_3+r_1}}\begin{bmatrix}1&0&-1&1\\0&2&2&2\\0&4&4&4\end{bmatrix}\xrightarrow{\boldsymbol{E}_{r_3-2r_2}}\begin{bmatrix}1&0&-1&1\\0&2&2&2\\0&0&0&0\end{bmatrix}$$

所以

$$\boldsymbol{L}_{2\times 4}=\begin{bmatrix}1&0&-1&1\\0&2&2&2\end{bmatrix}$$

对单位矩阵执行逆行初等变换如下：

$$\boldsymbol{I}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\xrightarrow{\boldsymbol{E}_{r_3+2r_2}}\begin{bmatrix}1&0&0\\0&1&0\\0&2&1\end{bmatrix}\xrightarrow{\boldsymbol{E}_{r_3-r_1}}\begin{bmatrix}1&0&0\\0&1&0\\-1&2&1\end{bmatrix}$$

所以

$$\boldsymbol{K}_{3\times 2}=\begin{bmatrix}1&0\\0&1\\-1&2\end{bmatrix}$$

由此可得

$$\boldsymbol{A}^{\dagger}=\boldsymbol{L}^{\mathrm{H}}(\boldsymbol{K}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{L}^{\mathrm{T}})^{-1}\boldsymbol{K}^{\mathrm{H}}=\begin{bmatrix}0.2778&0.1111&-0.0556\\0.0556&0.0556&0.0556\\-0.2222&-0.0556&0.1111\\0.3333&0.1667&0\end{bmatrix}$$

1.90 分别利用递推法 (算法1.9.3) 和迹方法 (算法1.9.4) 求矩阵

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

的 Moore-Penrose 逆矩阵 $\boldsymbol{X}^{\dagger}$。

解： (1) 递推法。设 $\boldsymbol{x}_k$ 是矩阵 $\boldsymbol{X}$ 的第 k 列，$\boldsymbol{X}_k$ 是 $\boldsymbol{X}$ 的前 k 列组成的分块矩阵，则递推运算如下：令 $\boldsymbol{X}_2 = [\boldsymbol{X}_1, \boldsymbol{x}_2]$，其中

$$\boldsymbol{X}_1 = \boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 2 \end{bmatrix}, \quad \boldsymbol{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ -1 \end{bmatrix}$$

则有

$$\boldsymbol{X}_1^{\dagger} = \boldsymbol{x}_1^{\dagger} = (\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{x}_1)^{-1}\boldsymbol{x}_1^{\mathrm{H}} = \frac{1}{6}[1, 0, -1, 2], \quad \boldsymbol{d}_2 = \boldsymbol{X}_1^{\dagger}\boldsymbol{x}_2 = -\frac{1}{2},$$

$$\boldsymbol{b}_2 = (\boldsymbol{x}_2 - \boldsymbol{X}_1\boldsymbol{d}_2)^{\dagger} = \begin{bmatrix} \frac{1}{2} \\ 1 \\ \frac{1}{2} \\ 0 \end{bmatrix}^{\dagger} = \frac{2}{3}\begin{bmatrix} \frac{1}{2} & 1 & \frac{1}{2} & 0 \end{bmatrix}$$

计算

$$\boldsymbol{X}_2^{\dagger} = \begin{bmatrix} \boldsymbol{X}_1^{\dagger} - \boldsymbol{d}_2\boldsymbol{b}_2 \\ \boldsymbol{b}_2 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \end{bmatrix}, \quad \boldsymbol{d}_3 = \boldsymbol{X}_2^{\dagger}\boldsymbol{x}_3 = -\frac{1}{3}\begin{bmatrix} 1 \\ 3 \end{bmatrix},$$

$$\boldsymbol{b}_3 = (\boldsymbol{X}_3 - \boldsymbol{X}_2\boldsymbol{d}_3)^{\dagger} = \frac{1}{5}\begin{bmatrix} -1 & 0 & 1 & 1 \end{bmatrix}$$

最后，得到 $\boldsymbol{X}$ 的 Moore-Penrose 逆矩阵为

$$\boldsymbol{X}^{\dagger} = \begin{bmatrix} \boldsymbol{X}_2^{\dagger} - \boldsymbol{d}_3\boldsymbol{b}_3 \\ \boldsymbol{b}_3 \end{bmatrix} = \frac{1}{15}\begin{bmatrix} 4 & 5 & 1 & 6 \\ 2 & 10 & 8 & 3 \\ -3 & 0 & 3 & 3 \end{bmatrix}$$

(2) 迹方法。易知 $\mathrm{rank}(\boldsymbol{X}) = 3$，迹方法的计算过程如下：计算

$$\boldsymbol{B} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \begin{bmatrix} 6 & -3 & 1 \\ -3 & 3 & -2 \\ 1 & -2 & 10 \end{bmatrix}$$

令 $\boldsymbol{C} = \boldsymbol{I}_{3\times 3}$，计算

$$\boldsymbol{C}_2 = \mathrm{tr}(\boldsymbol{C}_1\boldsymbol{B})\boldsymbol{I} - \boldsymbol{C}_1\boldsymbol{B} = \begin{bmatrix} 13 & 3 & -1 \\ 3 & 16 & 2 \\ -1 & 2 & 9 \end{bmatrix}$$

$$\boldsymbol{C}_3 = \frac{1}{2}\mathrm{tr}(\boldsymbol{C}_2\boldsymbol{B})\boldsymbol{I} - \boldsymbol{C}_2\boldsymbol{B} = \begin{bmatrix} 26 & 28 & 3 \\ 28 & 59 & 9 \\ 3 & 9 & 9 \end{bmatrix}$$

于是矩阵 $\boldsymbol{X}$ 的 Moore-Penrose 逆矩阵为

$$\boldsymbol{X}^{\dagger} = \frac{3\boldsymbol{C}_3\boldsymbol{X}^{\mathrm{T}}}{\mathrm{tr}(\boldsymbol{C}_3\boldsymbol{B})} = \frac{1}{15}\begin{bmatrix} 4 & 5 & 1 & 6 \\ 2 & 10 & 8 & 3 \\ -3 & 0 & 3 & 3 \end{bmatrix}$$

1.91 验证

$$G=\frac{1}{12}\begin{bmatrix}5&-4&-1&3\\-2&4&-2&6\\-3&0&3&3\end{bmatrix}$$

确实是矩阵

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

的 Moore-Penrose 逆矩阵。

解： (1) 因为

$$AG=\frac{1}{12}\begin{bmatrix}1&0&-1\\-1&1&-1\\0&-1&2\\1&1&1\end{bmatrix}\begin{bmatrix}5&-4&-1&3\\-2&4&-2&6\\-3&0&3&3\end{bmatrix}=\frac{1}{3}\begin{bmatrix}2&-1&-1&0\\-1&2&-1&0\\-1&-1&2&0\\0&0&0&3\end{bmatrix}$$

所以 $(AG)^{\mathrm{H}}=AG$。

(2) 因为

$$GA=\frac{1}{12}\begin{bmatrix}5&-4&-1&3\\-2&4&-2&6\\-3&0&3&3\end{bmatrix}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$$

所以 $(GA^{\mathrm{H}})=GA$。

(3) 又因为

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

(4) $GAG=IG=G$

故 G 是 A 的 Moore-Penrose 逆矩阵。

1.92 证明：若 $Ax=b$ 为一致方程，则其通解为

$$x=A^{\dagger}b+(I-A^{\dagger}A)z$$

式中，$A^{\dagger}$ 是 A 的 Moore-Penrose 逆矩阵，并且 z 为任意向量。

证明： 由于 $Ax=b$ 是一致方程，由广义逆矩阵的定义 1.8.3 知，$x=A^{\dagger}b$ 是方程 $Ax=b$ 的解，即 $A^{\dagger}Ab=b$。于是

$$Ax=A[A^{\dagger}b+(I-A^{\dagger}A)z]=AA^{\dagger}b+(A-AA^{\dagger}A)z=b$$

即 $x=A^{\dagger}b+(I-A^{\dagger}A)z$ 是方程 $Ax=b$ 的通解。

1.93 令矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是使得矩阵乘积 $\boldsymbol{AB}$ 存在，并且 $\boldsymbol{B}_1=\boldsymbol{A}^\dagger\boldsymbol{AB}$ 和 $\boldsymbol{A}_1=\boldsymbol{AB}_1\boldsymbol{B}_1^\dagger$，证明

$$\boldsymbol{AB}=\boldsymbol{A}_1\boldsymbol{B}_1 \quad 和 \quad (\boldsymbol{AB})^\dagger=(\boldsymbol{A}_1\boldsymbol{B}_1)^\dagger=\boldsymbol{B}_1^\dagger\boldsymbol{A}_1^\dagger$$

上述结果是 Cline 建立的。

证明：将 $\boldsymbol{B}_1=\boldsymbol{A}^\dagger\boldsymbol{AB}$ 等式两边同时左乘 $\boldsymbol{A}$，得

$$\boldsymbol{AB}_1=\boldsymbol{AA}^\dagger\boldsymbol{AB}=\boldsymbol{AB}$$

将 $\boldsymbol{A}_1=\boldsymbol{AB}_1\boldsymbol{B}_1^\dagger$ 等式两边同时右乘 $\boldsymbol{B}_1$，得

$$\boldsymbol{A}_1\boldsymbol{B}_1=\boldsymbol{AB}_1\boldsymbol{B}_1^\dagger\boldsymbol{B}_1=\boldsymbol{AB}_1$$

所以 $\boldsymbol{A}_1\boldsymbol{B}_1=\boldsymbol{AB}=\boldsymbol{AB}_1$。

现在证明 $(\boldsymbol{AB})^\dagger=\boldsymbol{B}_1^\dagger\boldsymbol{A}_1^\dagger$。令 $\boldsymbol{Y}=\boldsymbol{AB}=\boldsymbol{A}_1\boldsymbol{B}_1$，$\boldsymbol{X}=\boldsymbol{B}_1^\dagger\boldsymbol{A}_1^\dagger$，于是证明 $(\boldsymbol{AB})^\dagger=\boldsymbol{B}_1^\dagger\boldsymbol{A}_1^\dagger$ 也即证明 $\boldsymbol{X}$ 满足 $\boldsymbol{Y}$ 的 Moore-Penrose 逆矩阵的四个条件：

(1) 因为 $\boldsymbol{YX}=\boldsymbol{A}_1\boldsymbol{B}_1\boldsymbol{B}_1^\dagger\boldsymbol{A}_1^\dagger=\boldsymbol{AB}_1\boldsymbol{B}_1^\dagger\boldsymbol{A}_1^\dagger=\boldsymbol{A}_1\boldsymbol{A}_1^\dagger$，所以 $\boldsymbol{YX}$ 是 Hermitian 矩阵。

(2) $\boldsymbol{YXY}=\boldsymbol{A}_1\boldsymbol{A}_1^\dagger\boldsymbol{A}_1\boldsymbol{B}_1=\boldsymbol{A}_1\boldsymbol{B}_1=\boldsymbol{Y}$。

(3) $\boldsymbol{XYX}=\boldsymbol{B}_1^\dagger\boldsymbol{A}_1^\dagger\boldsymbol{A}_1\boldsymbol{A}_1^\dagger=\boldsymbol{B}_1^\dagger\boldsymbol{A}_1^\dagger=\boldsymbol{X}$。

(4) 下面证明 $\boldsymbol{XY}$ 也是 Hermitian 矩阵。因为

$$\boldsymbol{A}^\dagger\boldsymbol{A}_1=\boldsymbol{A}^\dagger\boldsymbol{AB}_1\boldsymbol{B}_1^\dagger=\boldsymbol{A}^\dagger\boldsymbol{ABB}_1^\dagger=\boldsymbol{B}_1\boldsymbol{B}_1^\dagger$$

又因为

$$\boldsymbol{A}_1^\dagger\boldsymbol{A}_1=\boldsymbol{A}_1^\dagger\boldsymbol{AB}_1\boldsymbol{B}_1^\dagger=\boldsymbol{A}_1^\dagger\boldsymbol{A}_1\boldsymbol{B}_1\boldsymbol{B}_1^\dagger$$

且 $\boldsymbol{A}_1^\dagger\boldsymbol{A}_1$ 和 $\boldsymbol{B}_1\boldsymbol{B}_1^\dagger$ 都是 Hermitian 矩阵，所以

$$\boldsymbol{B}_1\boldsymbol{B}_1^\dagger\boldsymbol{A}_1^\dagger\boldsymbol{A}_1=(\boldsymbol{B}_1\boldsymbol{B}_1^\dagger)^{\mathrm{H}}(\boldsymbol{A}_1^\dagger\boldsymbol{A}_1)^{\mathrm{H}}=(\boldsymbol{A}_1^\dagger\boldsymbol{A}_1\boldsymbol{B}_1\boldsymbol{B}_1^\dagger)^{\mathrm{H}}=(\boldsymbol{A}_1^\dagger\boldsymbol{A}_1)^{\mathrm{H}}=\boldsymbol{A}_1^\dagger\boldsymbol{A}_1$$

将 $\boldsymbol{B}_1\boldsymbol{B}_1^\dagger$ 替换为 $\boldsymbol{A}^\dagger\boldsymbol{A}_1$，得

$$\boldsymbol{A}_1^\dagger\boldsymbol{A}_1=\boldsymbol{A}^\dagger\boldsymbol{A}_1\boldsymbol{A}_1^\dagger\boldsymbol{A}_1=\boldsymbol{A}^\dagger\boldsymbol{A}_1 \quad \Longrightarrow \quad \boldsymbol{A}_1^\dagger\boldsymbol{A}_1=\boldsymbol{B}_1\boldsymbol{B}_1^\dagger$$

于是 $\boldsymbol{XY}=\boldsymbol{B}_1^\dagger\boldsymbol{A}_1^\dagger\boldsymbol{A}_1\boldsymbol{B}_1=\boldsymbol{B}_1^\dagger\boldsymbol{B}_1\boldsymbol{B}_1^\dagger\boldsymbol{B}_1=\boldsymbol{B}_1^\dagger\boldsymbol{B}_1$ 是 Hermitian 矩阵。

综上所述，$\boldsymbol{AB}=\boldsymbol{A}_1\boldsymbol{B}_1$ 且 $(\boldsymbol{AB})^\dagger=(\boldsymbol{A}_1\boldsymbol{B}_1)^\dagger=\boldsymbol{B}_1^\dagger\boldsymbol{A}_1^\dagger$，命题得证。

1.94 令 U 是 P_3 空间的子空间，定义为

$$U=\{p(x)=a_0+a_1x+a_2x^2+a_3x^3:\ a_3=-2a_0+3a_1+a_2\}$$

证明 U 与 P_3 同构。

证明： 本题有两种证明方法。解法一：由于 P_3 空间的定义为：$P_3 = \{p(x) = a_0 + a_1x + a_2x^2 + a_3x^3\}$，所以，$P_3$ 与 U 具有相同的一组基 $1, x, x^2, x^3$。令 P_3 到 U 的映射为 T，因为

$$[a_0, a_1, a_2, a_3]\begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = [a_0, a_1, a_2, -2a_0 + 2a_1 + a_2]$$

即映射 T 的矩阵表示为

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

令向量 $\boldsymbol{\alpha} \overset{\text{def}}{=} [a_0, a_1, a_2, a_3], \boldsymbol{\beta} \overset{\text{def}}{=} [b_0, b_1, b_2, b_3]$，根据同构的定义：$T(\boldsymbol{\alpha}+\boldsymbol{\beta}) = T(\boldsymbol{\alpha})+T(\boldsymbol{\beta})$，显然矩阵 $\boldsymbol{A}$ 满足：$(\boldsymbol{\alpha}+\boldsymbol{\beta})\boldsymbol{A} = \boldsymbol{\alpha}\boldsymbol{A} + \boldsymbol{\beta}\boldsymbol{A}$，且 $(k\boldsymbol{\alpha})\boldsymbol{A} = k(\boldsymbol{\alpha}\boldsymbol{A})$。所以，$U$ 与 P 同构。

解法二：根据“数域 P 上两个有限维线性空间同构的充分必要条件是它们具有相同的维数”这一定理，显然 P 和 U 的维数都是 3，所以它们是同构的。

1.95 证明满足 $(\boldsymbol{I}-\boldsymbol{A})(\boldsymbol{I}+\boldsymbol{A}) = \boldsymbol{O}$ 的矩阵 $\boldsymbol{A}$ 为对合矩阵。

证明： 因为

$$(\boldsymbol{I}-\boldsymbol{A})(\boldsymbol{I}+\boldsymbol{A}) = \boldsymbol{O} \quad \Longrightarrow \quad \boldsymbol{I}-\boldsymbol{A}+\boldsymbol{A}-\boldsymbol{A}^2 = \boldsymbol{O} \quad \Longrightarrow \quad \boldsymbol{A}^2 = \boldsymbol{I}$$

所以 $\boldsymbol{A}$ 是对合矩阵。

1.96 设 $\boldsymbol{A}_{n\times n}$ 为对合矩阵，证明 $\boldsymbol{B} = \frac{1}{2}(\boldsymbol{I}+\boldsymbol{A})$ 为幂等矩阵。

证明： 由 $\boldsymbol{A}$ 是对合矩阵，知 $\boldsymbol{A}^2 = \boldsymbol{I}$。所以

$$\boldsymbol{B}^2 = \frac{1}{4}(\boldsymbol{I}+\boldsymbol{A})^2 = \frac{1}{4}(\boldsymbol{I}+2\boldsymbol{A}+\boldsymbol{A}^2) = \frac{1}{2}(\boldsymbol{I}+\boldsymbol{A}) = \boldsymbol{B}$$

所以 $\boldsymbol{B}$ 为幂等矩阵。

1.97 已知

$$P_{\text{ex}}(n) = \text{tr}\,(\boldsymbol{R}\boldsymbol{\Phi}(n-1))$$

式中，$\boldsymbol{R} = \text{E}\{\boldsymbol{x}(n)\boldsymbol{x}^{\text{H}}(n)\}$ 是 $M\times 1$ 随机数据向量 $\boldsymbol{x}(n)$ 的自相关矩阵，并且

$$\boldsymbol{\Phi}(n) \approx \lambda^2\boldsymbol{\Phi}(n-1) + \sigma^2\text{E}\left\{\hat{\boldsymbol{R}}^{-1}(n)\boldsymbol{x}(n)\boldsymbol{x}^{\text{H}}(n)\hat{\boldsymbol{R}}^{-1}(n)\right\}$$

其中，$0 < \lambda < 1$，并且

$$\hat{\boldsymbol{R}}(n) = \sum_{i=0}^{n}\lambda^{n-i}\boldsymbol{x}(i)\boldsymbol{x}^{\text{H}}(i)$$

是真实自相关矩阵 $\boldsymbol{R}$ 的样本估计。证明

$$P_{\text{ex}}(\infty) = \text{tr}\,(\boldsymbol{R\Phi}(\infty)) \approx \frac{1-\lambda}{1+\lambda}M\sigma^2$$

(提示：求逆矩阵的数学期望 $\text{E}\{\hat{\boldsymbol{R}}^{-1}\}$ 的近似。)

证明： 因为 $\hat{\boldsymbol{R}}(n-1) = \sum\limits_{i=1}^{n-1} \lambda^{n-i-1}\boldsymbol{x}(i-1)\boldsymbol{x}^{\text{H}}(i-1)$，所以

$$\hat{\boldsymbol{R}}(n) = \lambda\hat{\boldsymbol{R}}(n-1) + \boldsymbol{x}(n)\boldsymbol{x}^{\text{H}}(n)$$

令 $\hat{\boldsymbol{R}} = \hat{\boldsymbol{R}}(\infty)$，则当 $n \to \infty$ 时，$\hat{\boldsymbol{R}} \approx \lambda\hat{\boldsymbol{R}} + \boldsymbol{R}$，其中我们用真实的自相关矩阵 $\boldsymbol{R}$ 来近似瞬时值 $\boldsymbol{x}(\infty)\boldsymbol{x}^{\text{H}}(\infty)$，即

$$\hat{\boldsymbol{R}} \approx \frac{1}{1-\lambda}\boldsymbol{R} \quad 或 \quad \hat{\boldsymbol{R}}^{-1} \approx (1-\lambda)\boldsymbol{R}^{-1}$$

又因为

$$\begin{aligned}
\boldsymbol{\Phi}(\infty) &\approx \lambda^2\boldsymbol{\Phi}(\infty) + \sigma^2\text{E}\{\hat{\boldsymbol{R}}^{-1}\boldsymbol{x}(\infty)\boldsymbol{x}^{\text{H}}(\infty)\hat{\boldsymbol{R}}^{-1}\} \\
&\approx \lambda^2\boldsymbol{\Phi}(\infty) + \sigma^2(1-\lambda)^2\boldsymbol{R}^{-1}\text{E}\{\boldsymbol{x}(\infty)\boldsymbol{x}^{\text{H}}(\infty)\}\boldsymbol{R}^{-1} \\
&= \lambda^2\boldsymbol{\Phi}(\infty) + \sigma^2(1-\lambda)^2\boldsymbol{R}^{-1}
\end{aligned}$$

所以

$$\boldsymbol{\Phi}(\infty) \approx \frac{1-\lambda}{1+\lambda}\sigma^2\boldsymbol{R}^{-1}$$

于是有

$$P_{\text{ex}}(\infty) = \text{tr}(\boldsymbol{R\Phi}(\infty)) \approx \text{tr}\left\{\frac{1-\lambda}{1+\lambda}\sigma^2\boldsymbol{I}\right\} = \frac{1-\lambda}{1+\lambda}M\sigma^2$$

1.98 证明：对于任何 $m \times n$ 矩阵 $\boldsymbol{A}$，其向量化函数

$$\text{vec}(\boldsymbol{A}) = (\boldsymbol{I}_n \otimes \boldsymbol{A})\text{vec}(\boldsymbol{I}_n) = (\boldsymbol{A}^{\text{T}} \otimes \boldsymbol{I}_n)\text{vec}(\boldsymbol{I}_m)$$

证明： 由题 1.78 的证明，可知

$$\begin{aligned}
\text{vec}(\boldsymbol{A}) &= \text{vec}(\boldsymbol{IA}) = (\boldsymbol{A}^{\text{T}} \otimes \boldsymbol{I}_m)\text{vec}(\boldsymbol{I}_n) = (\boldsymbol{A}^{\text{T}} \otimes \boldsymbol{I}_n)\text{vec}(\boldsymbol{I}_m) \\
\text{vec}(\boldsymbol{A}) &= \text{vec}(\boldsymbol{AI}) = (\boldsymbol{I}_n \otimes \boldsymbol{A})\text{vec}(\boldsymbol{I}_n)
\end{aligned}$$

故命题得证。

1.99 证明 $\boldsymbol{A} \otimes \boldsymbol{B}$ 非奇异的充分必要条件是 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 分别非奇异，并证明 $(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$。

证明： 令 $\boldsymbol{A}_{m\times m}$ 和 $\boldsymbol{B}_{n\times n}$，则 $\boldsymbol{A}\otimes\boldsymbol{B}$ 是 $mn\times mn$ 维的矩阵。因为 $\text{rank}(\boldsymbol{A}\otimes\boldsymbol{B})=\text{rank}(\boldsymbol{A})\text{rank}(\boldsymbol{B})$，所以

$$\boldsymbol{A}\otimes\boldsymbol{B}\text{非奇异} \quad\Leftrightarrow\quad \text{rank}(\boldsymbol{A})\,\text{rank}(\boldsymbol{B})=mn \quad\Leftrightarrow\quad \boldsymbol{A},\boldsymbol{B}\text{非奇异}$$

故 $\boldsymbol{A}\otimes\boldsymbol{B}$ 非奇异的充分必要条件是 $\boldsymbol{A},\boldsymbol{B}$ 分别非奇异。

又因为

$$(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{A}^{-1}\otimes\boldsymbol{B}^{-1})=\boldsymbol{A}\boldsymbol{A}^{-1}\otimes\boldsymbol{B}\boldsymbol{B}^{-1}=\boldsymbol{I}\otimes\boldsymbol{I}=\boldsymbol{I}$$
$$(\boldsymbol{A}^{-1}\otimes\boldsymbol{B}^{-1})(\boldsymbol{A}\otimes\boldsymbol{B})=\boldsymbol{A}^{-1}\boldsymbol{A}\otimes\boldsymbol{B}^{-1}\boldsymbol{B}=\boldsymbol{I}\otimes\boldsymbol{I}=\boldsymbol{I}$$

故由逆矩阵的定义可知 $(\boldsymbol{A}\otimes\boldsymbol{B})=\boldsymbol{A}^{-1}\otimes\boldsymbol{B}^{-1}$。

1.100 证明 Kronecker 积的 Moore-Penrose 逆矩阵的关系 $(\boldsymbol{A}\otimes\boldsymbol{B})^{\dagger}=\boldsymbol{A}^{\dagger}\otimes\boldsymbol{B}^{\dagger}$。

证明： 按照 Moore-Penrose 逆矩阵定义的四个条件逐一证明 $\boldsymbol{A}^{\dagger}\otimes\boldsymbol{B}^{\dagger}$ 是 $\boldsymbol{A}\otimes\boldsymbol{B}$ 的 Moore-Penrose 逆矩阵：

(1)
$$\begin{aligned}(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{A}^{\dagger}\otimes\boldsymbol{B}^{\dagger})(\boldsymbol{A}\otimes\boldsymbol{B})&=(\boldsymbol{A}\boldsymbol{A}^{\dagger}\otimes\boldsymbol{B}\boldsymbol{B}^{\dagger})(\boldsymbol{A}\otimes\boldsymbol{B})\\&=\boldsymbol{A}\boldsymbol{A}^{\dagger}\boldsymbol{A}\otimes\boldsymbol{B}\boldsymbol{B}^{\dagger}\boldsymbol{B}\\&=\boldsymbol{A}\otimes\boldsymbol{B}\end{aligned}$$

(2)
$$\begin{aligned}(\boldsymbol{A}^{\dagger}\otimes\boldsymbol{B}^{\dagger})(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{A}^{\dagger}\otimes\boldsymbol{B}^{\dagger})&=(\boldsymbol{A}^{\dagger}\boldsymbol{A}\otimes\boldsymbol{B}^{\dagger}\boldsymbol{B})(\boldsymbol{A}^{\dagger}\otimes\boldsymbol{B}^{\dagger})\\&=\boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{A}^{\dagger}\otimes\boldsymbol{B}^{\dagger}\boldsymbol{B}\boldsymbol{B}^{\dagger}\\&=\boldsymbol{A}^{\dagger}\otimes\boldsymbol{B}^{\dagger}\end{aligned}$$

(3) 因为 $\boldsymbol{A}\boldsymbol{A}^{\dagger}$ 和 $\boldsymbol{B}\boldsymbol{B}^{\dagger}$ 都是 Hermitian 矩阵，所以

$$\begin{aligned}[(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{A}^{\dagger}\otimes\boldsymbol{B}^{\dagger})]^{\text{H}}&=(\boldsymbol{A}\boldsymbol{A}^{\dagger}\otimes\boldsymbol{B}\boldsymbol{B}^{\dagger})^{\text{H}}=(\boldsymbol{A}\boldsymbol{A}^{\dagger})^{\text{H}}\otimes(\boldsymbol{B}\boldsymbol{B}^{\dagger})^{\text{H}}\\&=\boldsymbol{A}\boldsymbol{A}^{\dagger}\otimes\boldsymbol{B}\boldsymbol{B}^{\dagger}=(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{A}^{\dagger}\otimes\boldsymbol{B}^{\dagger})\end{aligned}$$

(4)
$$\begin{aligned}[(\boldsymbol{A}^{\dagger}\otimes\boldsymbol{B}^{\dagger})(\boldsymbol{A}\otimes\boldsymbol{B})]^{\text{H}}&=(\boldsymbol{A}^{\dagger}\boldsymbol{A}\otimes\boldsymbol{B}^{\dagger}\boldsymbol{B})^{\text{H}}=(\boldsymbol{A}^{\dagger}\boldsymbol{A})^{\text{H}}\otimes(\boldsymbol{B}^{\dagger}\boldsymbol{B})^{\text{H}}\\&=\boldsymbol{A}^{\dagger}\boldsymbol{A}\otimes\boldsymbol{B}^{\dagger}\boldsymbol{B}=(\boldsymbol{A}^{\dagger}\otimes\boldsymbol{B}^{\dagger})(\boldsymbol{A}\otimes\boldsymbol{B})\end{aligned}$$

因此 $(\boldsymbol{A}\otimes\boldsymbol{B})^{\dagger}=\boldsymbol{A}^{\dagger}\otimes\boldsymbol{B}^{\dagger}$ 成立。

1.101 若 $\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}$ 为相同维数的正方矩阵，并且 $\boldsymbol{C}^{\text{T}}=\boldsymbol{C}$，证明

$$(\text{vec}(\boldsymbol{C}))^{\text{T}}(\boldsymbol{A}\otimes\boldsymbol{B})\text{vec}(\boldsymbol{C})=(\text{vec}(\boldsymbol{C}))^{\text{T}}(\boldsymbol{B}\otimes\boldsymbol{A})\text{vec}(\boldsymbol{C})$$

思路分析： 利用性质 $\text{tr}(\boldsymbol{A}^{\text{T}}\boldsymbol{B})=(\text{vec}(\boldsymbol{A}))^{\text{T}}\text{vec}(\boldsymbol{B})$ 和 $\text{vec}(\boldsymbol{ABC})=(\boldsymbol{C}^{\text{T}}\otimes\boldsymbol{A})\text{vec}(\boldsymbol{B})$。

证明： 因为

$$(\text{vec}(\boldsymbol{C}))^{\text{T}}(\boldsymbol{A}\otimes\boldsymbol{B})\text{vec}(\boldsymbol{C}) = (\text{vec}(\boldsymbol{C}))^{\text{T}}\text{vec}(\boldsymbol{B}\boldsymbol{C}\boldsymbol{A}^{\text{T}}) = \text{tr}(\boldsymbol{C}^{\text{T}}\boldsymbol{B}\boldsymbol{C}\boldsymbol{A}^{\text{T}}) \\ = \text{tr}(\boldsymbol{C}\boldsymbol{B}\boldsymbol{C}\boldsymbol{A}^{\text{T}})$$

且

$$(\text{vec}(\boldsymbol{C}))^{\text{T}}(\boldsymbol{B}\otimes\boldsymbol{A})\text{vec}(\boldsymbol{C}) = (\text{vec}(\boldsymbol{C}))^{\text{T}}\text{vec}(\boldsymbol{A}\boldsymbol{C}\boldsymbol{B}^{\text{T}}) = \text{tr}(\boldsymbol{C}^{\text{T}}\boldsymbol{A}\boldsymbol{C}\boldsymbol{B}^{\text{T}}) \\ = \text{tr}(\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}\boldsymbol{B}^{\text{T}}) = \text{tr}(\boldsymbol{B}\boldsymbol{C}^{\text{T}}\boldsymbol{A}^{\text{T}}\boldsymbol{C}^{\text{T}}) = \text{tr}(\boldsymbol{B}\boldsymbol{C}\boldsymbol{A}^{\text{T}}\boldsymbol{C}) \\ = \text{tr}(\boldsymbol{C}\boldsymbol{B}\boldsymbol{C}\boldsymbol{A}^{\text{T}})$$

所以 $(\text{vec}(\boldsymbol{C}))^{\text{T}}(\boldsymbol{A}\otimes\boldsymbol{B})\text{vec}(\boldsymbol{C}) = (\text{vec}(\boldsymbol{C}))^{\text{T}}(\boldsymbol{B}\otimes\boldsymbol{A})\text{vec}(\boldsymbol{C})$。

1.102 令 $\boldsymbol{A}$、$\boldsymbol{B}$、$\boldsymbol{C}$ 和 $\boldsymbol{D}$ 具有适当的维数，使得 $\boldsymbol{ABCD}$ 满足矩阵乘积定义。证明

$$\text{tr}(\boldsymbol{ABCD}) = (\text{vec}(\boldsymbol{D}^{\text{T}}))^{\text{T}}(\boldsymbol{C}^{\text{T}}\otimes\boldsymbol{A})\text{vec}(\boldsymbol{B}) = (\text{vec}(\boldsymbol{D}))^{\text{T}}(\boldsymbol{A}\otimes\boldsymbol{C}^{\text{T}})\text{vec}(\boldsymbol{B}^{\text{T}})$$

证明： 因为

$$(\text{vec}(\boldsymbol{D}^{\text{T}}))^{\text{T}}(\boldsymbol{C}^{\text{T}}\otimes\boldsymbol{A})\text{vec}(\boldsymbol{B}) = (\text{vec}(\boldsymbol{D}^{\text{T}}))^{\text{T}}\text{vec}(\boldsymbol{ABC}) \\ = \text{tr}(\boldsymbol{DABC}) = \text{tr}(\boldsymbol{ABCD})$$

且

$$(\text{vec}(\boldsymbol{D}))^{\text{T}}(\boldsymbol{A}\otimes\boldsymbol{C}^{\text{T}})\text{vec}(\boldsymbol{B}^{\text{T}}) = (\text{vec}(\boldsymbol{D}))^{\text{T}}\text{vec}(\boldsymbol{C}^{\text{T}}\boldsymbol{B}^{\text{T}}\boldsymbol{A}^{\text{T}}) \\ = \text{tr}(\boldsymbol{D}^{\text{T}}\boldsymbol{C}^{\text{T}}\boldsymbol{B}^{\text{T}}\boldsymbol{A}^{\text{T}}) = \text{tr}(\boldsymbol{ABCD})$$

故命题得证。

1.103 令 $\boldsymbol{A}$ 为 $m\times m$ 对称矩阵，$\boldsymbol{B}$ 为 $m\times n$ 矩阵，$\boldsymbol{C}=\boldsymbol{AB}$，且 $\boldsymbol{D}=\boldsymbol{I}_m-\boldsymbol{C}\boldsymbol{C}^{\dagger}$。证明

$$(\boldsymbol{AC})^{\dagger} = \boldsymbol{C}^{\dagger}\boldsymbol{A}^{\dagger}[\boldsymbol{I}_m-(\boldsymbol{D}\boldsymbol{A}^{\dagger})^{\dagger}\boldsymbol{D}\boldsymbol{A}^{\dagger}]$$

证明： 按照 Moore-Penrose 逆矩阵定义的四个条件逐一证明 $\boldsymbol{C}^{\dagger}\boldsymbol{A}^{\dagger}[\boldsymbol{I}_m-(\boldsymbol{D}\boldsymbol{A}^{\dagger})^{\dagger}\boldsymbol{D}\boldsymbol{A}^{\dagger}]$ 是 $\boldsymbol{AC}$ 的 Moore-Penrose 逆矩阵。

由 Moore-Penrose 逆的定义知，$(\boldsymbol{A}^{\dagger}\boldsymbol{A})^{\text{H}}=\boldsymbol{A}^{\dagger}\boldsymbol{A}$。因为

$$(\boldsymbol{A}^{\dagger}\boldsymbol{A})^{\text{H}} = \boldsymbol{A}^{\text{H}}(\boldsymbol{A}^{\dagger})^{\text{H}} = \boldsymbol{A}\boldsymbol{A}^{\dagger}$$

故 $\boldsymbol{A}^{\dagger}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{\dagger}$。又因为

$$\boldsymbol{DC} = (\boldsymbol{I}_m-\boldsymbol{C}\boldsymbol{C}^{\dagger})\boldsymbol{C} = \boldsymbol{C}-\boldsymbol{C}\boldsymbol{C}^{\dagger}\boldsymbol{C} = \boldsymbol{O}$$

所以

$$DA^{\dagger}DC = DAA^{\dagger}C = DAA^{\dagger}AB = DAB = DC = O$$

另外，还有关系式

$$C^{\dagger}A^{\dagger}AC = C^{\dagger}A^{\dagger}AAB = C^{\dagger}AA^{\dagger}AB = C^{\dagger}AB = C^{\dagger}C$$

(1)
$$\begin{aligned}
& ACC^{\dagger}A^{\dagger}[I_m - (DA^{\dagger})^{\dagger}DA^{\dagger}]AC \\
&= ACC^{\dagger}A^{\dagger}AC - ACC^{\dagger}A^{\dagger}(DA^{\dagger})^{\dagger}DA^{\dagger}AC \\
&= ACC^{\dagger}A^{\dagger}AC = ACC^{\dagger}C \\
&= AC
\end{aligned}$$

(2)
$$\begin{aligned}
& C^{\dagger}A^{\dagger}[I_m - (DA^{\dagger})^{\dagger}DA^{\dagger}]ACC^{\dagger}A^{\dagger}[I_m - (DA^{\dagger})^{\dagger}DA^{\dagger}] \\
&= [C^{\dagger}A^{\dagger}AC - C^{\dagger}A^{\dagger}(DA^{\dagger})^{\dagger}DA^{\dagger}AC]C^{\dagger}A^{\dagger}[I_m - (DA^{\dagger})^{\dagger}DA^{\dagger}] \\
&= C^{\dagger}CC^{\dagger}A^{\dagger}[I_m - (DA^{\dagger})^{\dagger}DA^{\dagger}] \\
&= C^{\dagger}A^{\dagger}[I_m - (DA^{\dagger})^{\dagger}DA^{\dagger}]
\end{aligned}$$

(3)
$$\begin{aligned}
& ACC^{\dagger}A^{\dagger}[I_m - (DA^{\dagger})^{\dagger}DA^{\dagger}] \\
&= A(I_m - D)A^{\dagger}[I_m - (DA^{\dagger})^{\dagger}DA^{\dagger}] \\
&= AA^{\dagger} - ADA^{\dagger} - AA^{\dagger}(DA^{\dagger})^{\dagger}DA^{\dagger} + ADA^{\dagger}(DA^{\dagger})^{\dagger}DA^{\dagger} \\
&= AA^{\dagger} - AA^{\dagger}(DA^{\dagger})^{\dagger}DA^{\dagger}
\end{aligned}$$

因为 $(AA^{\dagger})^{\mathrm{H}} = A^{\dagger}A = AA^{\dagger}$，且

$$\begin{aligned}
AA^{\dagger}(DA^{\dagger})^{\dagger}DA^{\dagger} &= AA^{\dagger}[(DA^{\dagger})^{\dagger}DA^{\dagger}]^{\mathrm{H}} = AA^{\dagger}A^{\dagger}D^{\mathrm{H}}[(DA^{\dagger})^{\dagger}]^{\mathrm{H}} \\
&= A^{\dagger}D^{\mathrm{H}}[(DA^{\dagger})^{\dagger}]^{\mathrm{H}} = (DA^{\dagger})^{\mathrm{H}}[(DA^{\dagger})^{\dagger}]^{\mathrm{H}} = [(DA^{\dagger})^{\dagger}DA^{\dagger}]^{\mathrm{H}} \\
&= (DA^{\dagger})^{\dagger}DA^{\dagger}
\end{aligned}$$

而

$$[AA^{\dagger}(DA^{\dagger})^{\dagger}DA^{\dagger}]^{\mathrm{H}} = (DA^{\dagger})^{\dagger}DA^{\dagger}AA^{\dagger} = (DA^{\dagger})^{\dagger}DA^{\dagger}$$

故 $[AA^{\dagger}(DA^{\dagger})^{\dagger}DA^{\dagger}]^{\mathrm{H}} = AA^{\dagger}(DA^{\dagger})^{\dagger}DA^{\dagger}$，即是说 $ACC^{\dagger}A^{\dagger}[I_m - (DA^{\dagger})^{\dagger}DA^{\dagger}]$ 是 Hermitian 矩阵。

(4)
$$\begin{aligned}
C^{\dagger}A^{\dagger}[I_m - (DA^{\dagger})^{\dagger}DA^{\dagger}]AC = C^{\dagger}A^{\dagger}AC = C^{\dagger}A^{\dagger}(DA^{\dagger})^{\dagger}DA^{\dagger}AC \\
= C^{\dagger}A^{\dagger}AC = C^{\dagger}C
\end{aligned}$$

因为 $(C^{\dagger}C)^{\mathrm{H}} = C^{\dagger}C$，即 $C^{\dagger}C$ 是 Hermitian 矩阵，所以 $C^{\dagger}A^{\dagger}[I_m - (DA^{\dagger})^{\dagger}DA^{\dagger}]AC$ 是 Hermitian 矩阵。

综上所述，命题得证。

第2章 特殊矩阵

本章首先复习一些比较常见的特殊矩阵的定义及其性质，然后结合习题深入探讨，灵活运用相关的性质进行推导和证明，总结一些解题方面的技巧。

2.1 主要理论与方法

2.1.1 对称矩阵、Hermitian 矩阵与循环矩阵

一、对称矩阵

1. 定义：满足条件 $\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}$ 的实正方矩阵。

2. 性质：若 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 都是对称矩阵，则 $\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}$，且 $\boldsymbol{A}^{-1},\boldsymbol{A}^{m}$ (m为正整数) 和 $\boldsymbol{A}+\boldsymbol{B}$ 仍是对称矩阵。

二、反对称矩阵

1. 定义：满足条件 $\boldsymbol{A}^{\mathrm{T}}=-\boldsymbol{A}$ 的正方矩阵。显然，为了满足反对称性，矩阵 $\boldsymbol{A}$ 的主对角线元素必定等于零。

2. 性质：

(1) 若 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 都是反对称矩阵，则 $\boldsymbol{A}^{-1},\boldsymbol{A}+\boldsymbol{B}$ 仍是反对称矩阵，而且 $\boldsymbol{A}^{k}$ 为对称矩阵 (k为偶数时) 或反对称矩阵 (k为奇数时)。

(2) 任意正方矩阵 $\boldsymbol{A}$ 都可以分解为一个对称矩阵 $\frac{1}{2}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})$ 和一个反对称矩阵 $\frac{1}{2}(\boldsymbol{A}-\boldsymbol{A}^{\mathrm{T}})$ 之和。

三、中心对称矩阵

1. 定义：相对于矩阵交叉对角线对称的矩阵，即其元素满足

$$p_{ij}=p_{n-j+1,n-i+1},\qquad \forall\, i,j=1,2,\cdots,n$$

2. 性质：

(1) 若 $\boldsymbol{A}$ 中心对称，则 $\boldsymbol{A}^{\mathrm{T}}$ 也中心对称。

(2) 若 $\boldsymbol{A}_i$ 中心对称，则 $\boldsymbol{A}=\sum\limits_{i=1}^{K}a_i\boldsymbol{A}_i$ 也中心对称。

(3) 若 $\boldsymbol{A}_{n\times n}$ 和 $\boldsymbol{B}_{n\times n}$ 中心对称，则 $\boldsymbol{AB}$ 也中心对称。

(4) 中心对称矩阵 $\boldsymbol{A}$ 的逆矩阵 $\boldsymbol{A}^{-1}$ 中心对称。

四、Hermitian 矩阵

1. 定义：一个正方矩阵 $\boldsymbol{A} = [a_{ij}] \in C^{n\times n}$ 称为 Hermitian 矩阵，若 $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{H}}$，其中，$\boldsymbol{A}^{\mathrm{H}} = (\boldsymbol{A}^*)^{\mathrm{T}} = [a_{ji}^*]$，即 Hermitian 矩阵是一种复共轭对称矩阵。

Hermitian矩阵又有以下几种特殊形式：

(1) 矩阵$\boldsymbol{A}$称为反Hermitian矩阵，若$\boldsymbol{A} = -\boldsymbol{A}^{\mathrm{H}}$。

(2) 中央复共轭对称矩阵$\boldsymbol{R}$是一个元素满足对称性$r_{ij} = r_{n-i+1,n-j+1}^*$的$n\times n$正方矩阵，也称中央Hermitian矩阵。

(3) 中央复共轭对称矩阵的一个特殊子类是双重对称的矩阵，它既是关于主对角线对称的Hermitian矩阵，又是关于交叉对角线对称的交叉对称矩阵，例如

$$\boldsymbol{R} = \begin{bmatrix} r_{11} & r_{21}^* & r_{31}^* & r_{41}^* \\ r_{21} & r_{22} & r_{32}^* & r_{31}^* \\ r_{31} & r_{32} & r_{22} & r_{21}^* \\ r_{41} & r_{31} & r_{21} & r_{11} \end{bmatrix}$$

在这种特殊情况下，$r_{ij} = r_{ji}^* = r_{n-j+1,n-i+1} = r_{n-i+1,n-j+1}^*$。

2. 性质：

(1) 对所有 $\boldsymbol{A} \in C^{n\times n}$，矩阵 $\boldsymbol{A} + \boldsymbol{A}^{\mathrm{H}}$，$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 和 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 均是 Hermitian 矩阵。

(2) 若 $\boldsymbol{A}$ 是 Hermitian 矩阵，则 $\boldsymbol{A}^k$ 对所有 $k = 1, 2, 3, \cdots$ 都是 Hermitian 矩阵。若 $\boldsymbol{A}$ 还是非奇异的，则 $\boldsymbol{A}^{-1}$ 是 Hermitian 矩阵。

(3) 若 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是 Hermitian 矩阵，则 $a\boldsymbol{A} + b\boldsymbol{B}$ 对所有实数 a 和 b 均是 Hermitian 矩阵。

(4) 对所有 $\boldsymbol{A} \in C^{n\times n}$，$\boldsymbol{A} - \boldsymbol{A}^{\mathrm{H}}$ 为反 Hermitian 矩阵。

(5) 若 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是反 Hermitian 矩阵，则 $\alpha\boldsymbol{A}+\beta\boldsymbol{B}$ 对所有实数 α 和 β 均是反 Hermitian 矩阵。

(6) 若 $\boldsymbol{A}$ 是 Hermitian 矩阵，则 $\mathrm{j}\,\boldsymbol{A}$ ($\mathrm{j} = \sqrt{-1}$) 是反 Hermitian 矩阵。

(7) 若 $\boldsymbol{A}$ 是反 Hermitian 矩阵，则 $\mathrm{j}\,\boldsymbol{A}$ 是 Hermitian 矩阵。

五、循环矩阵

循环矩阵 $\boldsymbol{C}$ 是仅由 n 个复值元素经过适当移位而组成的 $n\times n$ 正方矩阵。如果在范围 $1 \leqslant i, j \leqslant n$ 内矩阵元素具有关系：

$$\boldsymbol{C}_R(i,j) = \begin{cases} c_R(j-i), & j-i \geqslant 0 \\ c_R(n+j-i), & j-i < 0 \end{cases} \tag{2.1}$$

则称这样的循环矩阵为右循环矩阵，并记作 $\boldsymbol{C}_R$。

类似可定义左循环矩阵 $\boldsymbol{C}_L$，即

$$\boldsymbol{C}_L(i,j) = \begin{cases} c_L(n+1-i-j), & j+i \leqslant n+1 \\ c_L(2n+1-i-j), & j+i > n+1 \end{cases} \tag{2.2}$$

2.1.2 基本矩阵

一、基本向量

一个仅第 k 个元素为 1，而其他元素均为零的 $n\times 1$ 维向量称为单位向量或基本向量 (elementary vector)，记作 $\boldsymbol{e}_k$。

二、数量矩阵

对角元素均相同的对角矩阵叫做数量矩阵，记作

$$\boldsymbol{D}=\begin{bmatrix} d & & & 0\\ & d & & \\ & & \ddots & \\ 0 & & & d\end{bmatrix}=d\boldsymbol{I}$$

显然，$\boldsymbol{DA}=\boldsymbol{AD}=d\boldsymbol{A}$，其中矩阵 $\boldsymbol{A}$ 任意。

三、基本矩阵

对 $n\times n$ 单位矩阵进行初等行 (或列) 变换得到的矩阵称为基本矩阵[10],[14],[18],[27]。与矩阵的三种初等运算 (Ⅰ，Ⅱ，Ⅲ型初等行变换) 相对应，可以引出以下三类基本矩阵。

Ⅰ型基本矩阵 $\boldsymbol{E}_{(p,q)}$：互换单位矩阵 $\boldsymbol{I}_n$ 的第 p 行和第 q 行得到的矩阵，或互换单位矩阵的第 p 列和第 q 列得到的矩阵。

Ⅱ型基本矩阵 $\boldsymbol{E}_{\alpha(p)}$：用一个非零常数 α 乘以单位矩阵 $\boldsymbol{I}_n$ 的第 p 行 (或列) 得到的矩阵。

Ⅲ型基本矩阵 $\boldsymbol{E}_{(p)+\alpha(q)}$：用一个非零常数 α 乘以单位矩阵 $\boldsymbol{I}_n$ 的第 q 行 (或列)，再加到 $\boldsymbol{I}_n$ 的第 p 行 (或列) 得到的矩阵。

定理 2.1 对应于矩阵初等变换的三种基本矩阵的行列式分别为

(1) $\det\left(\boldsymbol{E}_{(p,q)}\right)=-1$。

(2) $\det\left(\boldsymbol{E}_{\alpha(p)}\right)=\alpha$。

(3) $\det\left(\boldsymbol{E}_{(p)+\alpha(q)}\right)=1$。

定理 2.2 $n\times n$ 矩阵 $\boldsymbol{A}$ 互换任意两行 (或列) 得到的矩阵 $\boldsymbol{A}_{(p),(q)}$ 的行列式与原矩阵 $\boldsymbol{A}$ 的行列式存在关系式：

$$\det\left(\boldsymbol{A}_{(p),(q)}\right)=-\det(\boldsymbol{A}),\quad \forall\ p\neq q$$

推论 2.1

$$\det(\boldsymbol{E}_p\boldsymbol{E}_{p-1}\cdots\boldsymbol{E}_1\boldsymbol{A})=\det(\boldsymbol{E}_p)\det(\boldsymbol{E}_{p-1})\cdots\det(\boldsymbol{E}_1)\det(\boldsymbol{A})$$
$$\det(\boldsymbol{A}\boldsymbol{E}_1\boldsymbol{E}_2\cdots\boldsymbol{E}_p)=\det(\boldsymbol{A})\det(\boldsymbol{E}_1)\det(\boldsymbol{E}_2)\cdots\det(\boldsymbol{E}_p)$$

推论 2.2 若 $\boldsymbol{A}_{n\times n}$ 为奇异矩阵，则 $\det(\boldsymbol{A})=0$。

2.1.3 置换矩阵、互换矩阵与选择矩阵

一、置换矩阵

一个正方矩阵称为置换矩阵，若它的每一行和每一列有一个且仅有一个非零元素 1。置换矩阵 $\boldsymbol{P}$ 有下列性质。

(1) $(\boldsymbol{P}_{m\times n})^{\mathrm{T}}=\boldsymbol{P}_{n\times m}$。

(2) $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}=\boldsymbol{P}\boldsymbol{P}^{\mathrm{T}}=\boldsymbol{I}$，这说明置换矩阵是正交矩阵。

(3) $\boldsymbol{P}^{\mathrm{T}}=\boldsymbol{P}^{-1}$。

(4) $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}$ 与 $\boldsymbol{A}$ 具有相同的对角线元素，但排列顺序可能不同。

二、交换矩阵

1. 定义：交换矩阵 $\boldsymbol{K}_{mn}$ 定义为满足 $\boldsymbol{K}_{mn}\mathrm{vec}(\boldsymbol{A}_{m\times n})=\mathrm{vec}(\boldsymbol{A}^{\mathrm{T}})$ 的特殊置换矩阵。

2. 性质：交换矩阵既是对合矩阵，也是对称矩阵，即

$$\boldsymbol{K}_{nn}^2=\boldsymbol{K}_{nn}\boldsymbol{K}_{nn}=\boldsymbol{I}_{nn} \tag{2.3}$$

$$\boldsymbol{K}_{nn}^{\mathrm{T}}=\boldsymbol{K}_{nn} \tag{2.4}$$

三、互换矩阵

1. 定义：互换矩阵 (exchange matrix) 常用符号 $\boldsymbol{J}$ 表示，定义为

$$\boldsymbol{J}=\begin{bmatrix} 0 & & & 1 \\ & & 1 & \\ & \cdot^{\cdot^{\cdot}} & & \\ 1 & & & 0 \end{bmatrix} \tag{2.5}$$

2. 性质：

(1) 互换矩阵既是对合矩阵，也是对称矩阵，即

$$\boldsymbol{J}^2=\boldsymbol{J}\boldsymbol{J}=\boldsymbol{I}\quad (\text{互换矩阵的对合性}) \tag{2.6}$$

$$\boldsymbol{J}^{\mathrm{T}}=\boldsymbol{J}\quad (\text{互换矩阵的对称性}) \tag{2.7}$$

(2) 若 $\boldsymbol{R}$ 是交叉对称矩阵，则 $\boldsymbol{R}^{\mathrm{T}}=\boldsymbol{J}\boldsymbol{R}\boldsymbol{J}$ 和 $\boldsymbol{R}=\boldsymbol{J}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{J}$。

(3) 中央对称矩阵 $\boldsymbol{R}=\boldsymbol{J}\boldsymbol{R}\boldsymbol{J}$，而中央复共轭对称矩阵 $\boldsymbol{R}=\boldsymbol{J}\boldsymbol{R}^*\boldsymbol{J}$。

四、移位矩阵

$n\times n$ 移位矩阵 (shift matrix) 定义为

$$\boldsymbol{P}=\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{2.8}$$

移位矩阵能使别的矩阵的首行或者最后一列移动位置。值得注意的是，移位矩阵既不具有对合性，也不是对称矩阵。

五、广义置换矩阵

一个正方矩阵称为广义置换矩阵 (generalized permutation matrix)，简称 g 矩阵，若其每行和每列有一个并且仅有一个非零元素。容易证明，一个正方矩阵是 g 矩阵，当且仅当它可以分解为一个置换矩阵和一个非奇异的对角矩阵之积，即有

$$\boldsymbol{G} = \boldsymbol{P}\boldsymbol{D} \tag{2.9}$$

式中，$\boldsymbol{D}$ 为非奇异的对角矩阵。

2.1.4 正交矩阵与酉矩阵

一、正交矩阵与酉矩阵

一实的正方矩阵 $\boldsymbol{Q} \in R^{n\times n}$ 称为正交矩阵，若

$$\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I} \tag{2.10}$$

一复值正方矩阵 $\boldsymbol{U} \in C^{n\times n}$ 称为酉矩阵，若

$$\boldsymbol{U}\boldsymbol{U}^{\mathrm{H}} = \boldsymbol{U}^{\mathrm{H}}\boldsymbol{U} = \boldsymbol{I} \tag{2.11}$$

实矩阵 $\boldsymbol{Q}_{m\times n}$ 称为半正交矩阵 (semi-orthogonal matrix)，若它只满足 $\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{I}_m$ 或者 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}_n$。类似地，复矩阵 $\boldsymbol{U}_{m\times n}$ 称为仿酉矩阵 (para-unitary matrix)，若它只满足 $\boldsymbol{U}\boldsymbol{U}^{\mathrm{H}} = \boldsymbol{I}_m$ 或者 $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U} = \boldsymbol{I}_n$。

酉矩阵的性质：

(1) $\boldsymbol{A}_{m\times m}$ 为酉矩阵 $\Leftrightarrow$ $\boldsymbol{A}$ 的列是标准正交的向量。

(2) $\boldsymbol{A}_{m\times m}$ 为酉矩阵 $\Leftrightarrow$ $\boldsymbol{A}$ 的行是标准正交的向量。

(3) $\boldsymbol{A}_{m\times m}$ 为酉矩阵 $\Leftrightarrow$ $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{I}_m$

$\Leftrightarrow$ $\boldsymbol{A}^{\mathrm{T}}$ 为酉矩阵

$\Leftrightarrow$ $\boldsymbol{A}^{\mathrm{H}}$ 为酉矩阵

$\Leftrightarrow$ $\boldsymbol{A}^{*}$ 为酉矩阵

$\Leftrightarrow$ $\boldsymbol{A}^{-1}$ 为酉矩阵

$\Leftrightarrow$ $\boldsymbol{A}^{i}$ 为酉矩阵，$i = 1, 2, \cdots$

(4) $\boldsymbol{A}_{m\times m}$ 为实矩阵：$\boldsymbol{A}$ 为酉矩阵 $\Leftrightarrow$ $\boldsymbol{A}$ 为正交矩阵。

(5) $\boldsymbol{A}_{m\times m}, \boldsymbol{B}_{m\times m}$ 为酉矩阵 $\Rightarrow$ $\boldsymbol{A}\boldsymbol{B}$ 为酉矩阵。

(6) 若 $\boldsymbol{A}_{m\times m}$ 为酉矩阵，则

① $\det(\boldsymbol{A})$ 的模等于 1。

② $\text{rank}(\boldsymbol{A}) = m$。

③ $\boldsymbol{A}$ 是正规矩阵，即 $\boldsymbol{A}\boldsymbol{A}^{\text{H}} = \boldsymbol{A}^{\text{H}}\boldsymbol{A}$。

④ λ为$\boldsymbol{A}$ 的特征值 $\Rightarrow$ $|\lambda| = 1$。

⑤ $\boldsymbol{B}_{m\times n}$ $\Rightarrow$ $\|\boldsymbol{A}\boldsymbol{B}\|_{\text{F}} = \|\boldsymbol{B}\|_{\text{F}}$。

⑥ $\boldsymbol{B}_{n\times m}$ $\Rightarrow$ $\|\boldsymbol{B}\boldsymbol{A}\|_{\text{F}} = \|\boldsymbol{B}\|_{\text{F}}$。

⑦ $\boldsymbol{x}_{m\times 1}$ $\Rightarrow$ $\|\boldsymbol{A}\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$。

(7) 若 $\boldsymbol{A}_{m\times m}, \boldsymbol{B}_{n\times n}$ 为酉矩阵，则

① $\boldsymbol{A} \oplus \boldsymbol{B}$ 为酉矩阵。

② $\boldsymbol{A} \otimes \boldsymbol{B}$ 为酉矩阵。

二、子酉变换

$m \times n$ 矩阵 $\boldsymbol{A}$ 称为部分等距变换 (partial isometry transformation) 或子酉变换，若

$$\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_n = \|\boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{A}\boldsymbol{x}_2\|_m, \quad \forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{C}(\boldsymbol{A}^{\#}) \tag{2.12}$$

或等价为

$$\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle_n = \langle \boldsymbol{A}\boldsymbol{x}_1, \boldsymbol{A}\boldsymbol{x}_2 \rangle_m, \quad \forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{C}(\boldsymbol{A}^{\#}) \tag{2.13}$$

式中，$\mathcal{C}(\boldsymbol{A}^{\#})$ 表示矩阵 $\boldsymbol{A}^{\#}$ 的列空间，而 $\boldsymbol{A}^{\#}$ 是矩阵$\boldsymbol{A}$ 的伴随矩阵，满足

$$\langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{y} \rangle_m = \langle \boldsymbol{x}, \boldsymbol{A}^{\#}\boldsymbol{y} \rangle_n \tag{2.14}$$

三、酉等价

一个满足 $\boldsymbol{B} = \boldsymbol{U}^{\text{H}}\boldsymbol{A}\boldsymbol{U}$ 的矩阵 $\boldsymbol{B} \in C^{n\times n}$ 被称为是与 $\boldsymbol{A} \in C^{n\times n}$ 酉等价的。如果 $\boldsymbol{U}$ 取实数 (因而是实正交的)，则称 $\boldsymbol{B}$ 是与 $\boldsymbol{A}$ 正交等价的。

若 $n \times n$ 矩阵 $\boldsymbol{A} = [a_{ij}]$ 和 $\boldsymbol{B} = [b_{ij}]$ 是酉等价的，则 $\text{tr}(\boldsymbol{A}^{\text{H}}\boldsymbol{A}) = \text{tr}(\boldsymbol{B}^{\text{H}}\boldsymbol{B})$。

四、正规矩阵

矩阵 $\boldsymbol{A} \in C^{n\times n}$ 称为正规矩阵 (normal matrix)，若 $\boldsymbol{A}^{\text{H}}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{\text{H}}$。

定理 2.3 矩阵 $\boldsymbol{A} \in C^{n\times n}$ 是正规矩阵，当且仅当存在 $n \times n$ 酉矩阵 $\boldsymbol{U}$，使得

$$\boldsymbol{U}^{\text{H}}\boldsymbol{A}\boldsymbol{U} = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \tag{2.15}$$

五、J 正交矩阵

一个对角元素只取 $+1$ 和 -1 两种值的 $N \times N$ 对角矩阵称为符号矩阵 (signature matrix)。

令 $\boldsymbol{J}$ 为 $N \times N$ 符号矩阵，满足

$$\boldsymbol{Q}\boldsymbol{J}\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{J} \tag{2.16}$$

的 $N \times N$ 矩阵称为 J 正交矩阵 (J-orthogonal matrix)，或称超正规矩阵 (hypernormal matrix)。

J 正交矩阵具有以下性质。

(1) J 正交矩阵 $\boldsymbol{Q}$ 非奇异，其行列式的绝对值等于 1。

(2) 任何一个 $N \times N$ 维 J 正交矩阵 $\boldsymbol{Q}$ 也可以等价定义为

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{J}\boldsymbol{Q} = \boldsymbol{J} \tag{2.17}$$

2.1.5 带型矩阵与三角矩阵

一、带型矩阵

满足条件 $a_{ij} = 0, |i-j| > k$ 的矩阵 $\boldsymbol{A} \in C^{m\times n}$ 称为带型矩阵 (banded matrix)。

二、三角矩阵

满足条件 $a_{ij} = 0, i > j$ 的正方矩阵 $\boldsymbol{U} = [u_{ij}]$ 称为上三角矩阵 (upper triangular matrix)，其一般形式为

$$\boldsymbol{U} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{bmatrix}$$

满足条件 $l_{ij} = 0, i < j$ 的正方矩阵 $\boldsymbol{L} = [l_{ij}]$ 称为下三角矩阵 (lower triangular matrix)，其一般形式为

$$\boldsymbol{L} = \begin{bmatrix} l_{11} & & & 0 \\ l_{12} & l_{22} & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix} \quad \rightarrow \quad |\boldsymbol{L}| = l_{11}l_{22}\cdots l_{nn}$$

上三角矩阵的性质如下：

(1) 上三角矩阵之积为上三角矩阵，即若 $\boldsymbol{U}_1, \boldsymbol{U}_2, \cdots, \boldsymbol{U}_k$ 各为上三角矩阵，则 $\boldsymbol{U} = \boldsymbol{U}_1\boldsymbol{U}_2\cdots\boldsymbol{U}_k$ 为上三角矩阵。

(2) 上三角矩阵 $\boldsymbol{U} = [u_{ij}]$ 的行列式等于对角线元素之积，即

$$\det(\boldsymbol{U}) = u_{11}u_{22}\cdots u_{nn} = \prod_{i=1}^{n} u_{ii}$$

(3) 上三角矩阵的逆矩阵为下三角矩阵。

(4) 上三角矩阵 $\boldsymbol{U}_{n\times n}$ 的 k 次幂 $\boldsymbol{U}^k$ 仍为上三角矩阵，并且其第 i 个对角线元素等于 u_{ii}^k。

(5) 上三角矩阵 $\boldsymbol{U}_{n\times n}=\boldsymbol{U}=[u_{ij}]$ 的特征值为 $u_{11},u_{22},\cdots,u_{nn}$。

(6) 正定 Hermitian 矩阵 $\boldsymbol{A}$ 可以分解为 $\boldsymbol{A}=\boldsymbol{T}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{T}$，其中，$\boldsymbol{T}$ 为单位上三角复矩阵，$\boldsymbol{D}$ 为实对角矩阵。

下三角矩阵的性质如下：

(1) 下三角矩阵之积为下三角矩阵，即若 $\boldsymbol{L}_1,\boldsymbol{L}_2,\cdots,\boldsymbol{L}_k$ 各为下三角矩阵，则 $\boldsymbol{L}=\boldsymbol{L}_1\boldsymbol{L}_2\cdots\boldsymbol{L}_k$ 为下三角矩阵。

(2) 下三角矩阵的行列式等于对角线元素之积，即

$$\det(\boldsymbol{L})=l_{11}l_{22}\cdots l_{nn}=\prod_{i=1}^{n}l_{ii}$$

(3) 下三角矩阵的逆矩阵为上三角矩阵。

(4) 下三角矩阵 $\boldsymbol{L}_{n\times n}$ 的 k 次幂 $\boldsymbol{L}^k$ 仍为下三角矩阵，且第 i 个对角线元素等于 l_{ii}^k。

(5) 下三角矩阵 $\boldsymbol{L}_{n\times n}$ 的特征值为 $l_{11},l_{22},\cdots,l_{nn}$。

(6) 一个正定矩阵 $\boldsymbol{A}_{n\times n}$ 能够分解为下三角矩阵 $\boldsymbol{L}_{n\times n}$ 与其转置之积，即

$$\boldsymbol{A}=\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}\quad(\boldsymbol{A}\ \text{正定，且}\ \boldsymbol{L}\ \text{下三角矩阵})$$

这一分解称为矩阵 $\boldsymbol{A}$ 的 Cholesky 分解。

三、分块三角矩阵的求逆公式

$$\begin{bmatrix}\boldsymbol{A} & \boldsymbol{O}\\ \boldsymbol{B} & \boldsymbol{C}\end{bmatrix}^{-1}=\begin{bmatrix}\boldsymbol{A}^{-1} & \boldsymbol{O}\\ -\boldsymbol{C}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} & \boldsymbol{C}^{-1}\end{bmatrix}\tag{2.18}$$

$$\begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{O}\end{bmatrix}^{-1}=\begin{bmatrix}\boldsymbol{O} & \boldsymbol{C}^{-1}\\ \boldsymbol{B}^{-1} & -\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{C}^{-1}\end{bmatrix}\tag{2.19}$$

$$\begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{O} & \boldsymbol{C}\end{bmatrix}^{-1}=\begin{bmatrix}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{C}^{-1}\\ \boldsymbol{O} & \boldsymbol{C}^{-1}\end{bmatrix}\tag{2.20}$$

假定对角线或者交叉对角线上的矩阵是可逆的。

2.1.6 中心化矩阵与对角加矩阵

一、求和向量与中心化矩阵

所有元素为1的向量称为求和向量 (summuing vector)，记为 $\mathbf{1}=[1,1,\cdots,1]^{\mathrm{T}}$。

定义矩阵

$$\boldsymbol{C}_n=\boldsymbol{I}_n-\bar{\boldsymbol{J}}_n=\boldsymbol{I}_n-\frac{1}{n}\boldsymbol{J}_n\tag{2.21}$$

这一矩阵在数理统计中称为中心化矩阵 (centering matrix)，其中 $\boldsymbol{J}=\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$，即所有元素为 1 的正方矩阵。

矩阵 $\boldsymbol{C}$ 对数据向量 $\boldsymbol{x}$ 的线性变换 $\boldsymbol{C}\boldsymbol{x}$ 是原数据向量的各个元素减去n 个数据的均值的结果，而数据的协方差可以用核矩阵为中心化矩阵的二次型 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{x}$ 表示。

二、对角加矩阵

令矩阵 $\boldsymbol{D}_{n\times n}$ 非奇异，$\boldsymbol{a}_{n\times 1}$ 和 $\boldsymbol{b}_{n\times 1}$ 为向量，并且 p 为标量。称

$$\boldsymbol{C}=\boldsymbol{D}+p\,\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}=\begin{bmatrix} d_{11}+p\,a_1b_1 & d_{12}+p\,a_1b_2 & \cdots & d_{1n}+p\,a_1b_n \\ d_{21}+p\,a_2b_1 & d_{22}+p\,a_2b_2 & \cdots & d_{2n}+p\,a_2b_n \\ \vdots & \vdots & & \vdots \\ d_{n1}+p\,a_nb_1 & d_{n2}+p\,a_nb_2 & \cdots & d_{nn}+p\,a_nb_n \end{bmatrix} \tag{2.22}$$

为对角加矩阵 (diagonal plus matrix)，式中，$p\neq -\left(\sum\limits_{i=1}^{n}\dfrac{a_ib_i}{d_{ii}}\right)^{-1}$。

2.1.7 相似矩阵与相合矩阵

一、相似矩阵

矩阵 $\boldsymbol{B}\in C^{n\times n}$ 称为矩阵 $\boldsymbol{A}\in C^{n\times n}$ 的相似矩阵，若存在一非奇异矩阵 $\boldsymbol{S}\in C^{n\times n}$ 使得 $\boldsymbol{B}=\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$。此时，线性变换 $\boldsymbol{A}\mapsto\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$ 称为矩阵 $\boldsymbol{A}$ 的相似变换。

令 $\boldsymbol{A},\boldsymbol{B}\in C^{n\times n}$。若 $\boldsymbol{B}$ 与 $\boldsymbol{A}$ 相似，则 $\det(\boldsymbol{B})=\det(\boldsymbol{A})$ 和 $\mathrm{tr}(\boldsymbol{B})=\mathrm{tr}(\boldsymbol{A})$，即相似矩阵的行列式相等，并具有相同的迹。

令 $\boldsymbol{A},\boldsymbol{B}\in C^{n\times n}$。若 $\boldsymbol{B}$ 与 $\boldsymbol{A}$ 相似，则 $\boldsymbol{B}$ 的特征多项式 $\det(\boldsymbol{B}-z\boldsymbol{I})$ 与 $\boldsymbol{A}$ 的特征多项式 $\det(\boldsymbol{A}-z\boldsymbol{I})$ 相同。

若 $\boldsymbol{A},\boldsymbol{B}\in C^{n\times n}$，并且 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 相似，则它们具有相同的特征值 (包括多重特征值在内)。

相似矩阵的重要性质。

(1) 相似矩阵 $\boldsymbol{B}\sim\boldsymbol{A}$ 具有相同的行列式，即 $|\boldsymbol{B}|=|\boldsymbol{A}|$。

(2) 若矩阵 $\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}=\boldsymbol{T}$ (上三角矩阵)，则 $\boldsymbol{T}$ 的对角元素给出矩阵 $\boldsymbol{A}$ 的特征值 λ_i。

(3) 两个相似矩阵具有完全相同的特征值。

(4) 若 $\boldsymbol{A}$ 的特征值各不相同，则一定可以找到一相似矩阵 $\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}=\boldsymbol{D}$ (对角矩阵)，其对角元素即是矩阵 $\boldsymbol{A}$ 的特征值。

(5) $n\times n$矩阵 $\boldsymbol{A}$ 与对角矩阵相似的充分必要条件是：矩阵 $\boldsymbol{A}$ 的 n 个特征向量线性无关。

(6) 相似矩阵 $\boldsymbol{B}=\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$ 意味着 $\boldsymbol{B}^2=\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}=\boldsymbol{S}^{-1}\boldsymbol{A}^2\boldsymbol{S}$，从而有 $\boldsymbol{B}^k=\boldsymbol{S}^{-1}\boldsymbol{A}^k\boldsymbol{S}$。就是说，若 $\boldsymbol{B}\sim\boldsymbol{A}$，则 $\boldsymbol{B}^k\sim\boldsymbol{A}^k$。这一性质称为相似矩阵的幂性质。

(7) 若矩阵 $\boldsymbol{B}=\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$ 和 $\boldsymbol{A}$ 均可逆，则 $\boldsymbol{B}^{-1}=\boldsymbol{S}^{-1}\boldsymbol{A}^{-1}\boldsymbol{S}$，即当两个矩阵相似时，它们的逆矩阵也相似。

在相似变换中最重要的是酉相似变换。如果矩阵 $\boldsymbol{A}$ 经过酉矩阵相似变换为 $\boldsymbol{B}$，就称 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是酉相似的。若 Hermitian 矩阵 $\boldsymbol{A}$ 经过酉矩阵 $\boldsymbol{U}^{-1}=\boldsymbol{U}^{\mathrm{H}}$ 相似变换为对角矩阵 $\boldsymbol{\Sigma}$，即有 $\boldsymbol{\Sigma}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}$，则 Hermitian 矩阵 $\boldsymbol{A}$ 与酉相似的对角矩阵 $\boldsymbol{\Sigma}$ 具有相同的特征值。这正是 Hermitian 矩阵$\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$ 的特征值分解 $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{H}}$ 的理论基础。

二、相合矩阵

令 $\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}\in C^{n\times n}$，并且 $\boldsymbol{C}$ 非奇异，则矩阵 $\boldsymbol{B}=\boldsymbol{C}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{C}$ 称为 $\boldsymbol{A}$ 的相合矩阵 (congruent matrix)，而线性变换 $\boldsymbol{A}\mapsto\boldsymbol{C}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{C}$ 称为相合变换。

“相合”具有以下两层涵义：

(1) 两个相合矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的二次型函数相吻合。考查二次型函数 $f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}$。若令 $\boldsymbol{x}=\boldsymbol{C}\boldsymbol{y}$，其中，$\boldsymbol{C}$ 为非奇异矩阵，则有

$$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}^{\mathrm{H}}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{C}\boldsymbol{y}=\boldsymbol{y}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{y} \tag{2.23}$$

式中，$\boldsymbol{B}=\boldsymbol{C}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{C}$。这表明，两个相合矩阵具有相同的二次型函数。

(2) Hermitian 矩阵 $\boldsymbol{A}$ 的规范相合矩阵 $\boldsymbol{D}$ 与酉相似对角化矩阵 (特征值矩阵) $\boldsymbol{\Sigma}$ 的元素具有以下关系：零元素的个数相同，对应的非零元素具有相同的符号。

2.1.8 Vandermonde 矩阵与 Fourier 矩阵

一、Vandermonde 矩阵

$n\times n$ 维 Vandermonde 矩阵是取以下特殊形式的矩阵：

$$\boldsymbol{A}=\begin{bmatrix}1 & x_1 & x_1^2 & \cdots & x_1^{n-1}\\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1}\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1}\end{bmatrix} \tag{2.24}$$

或

$$\boldsymbol{A}=\begin{bmatrix}1 & 1 & \cdots & 1\\ x_1 & x_2 & \cdots & x_n\\ x_1^2 & x_2^2 & \cdots & x_n^2\\ \vdots & \vdots & & \vdots\\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1}\end{bmatrix} \tag{2.25}$$

即矩阵每行 (或列) 的元素组成一个等比序列。

n 个参数 $x_1,x_2,\cdots,x_n$ 各异时，Vandermonde 矩阵非奇异。

二、Fourier 矩阵

Fourier 矩阵是 Vandermonde 矩阵的一种特例，定义为

$$\boldsymbol{F}=\frac{1}{\sqrt{N}}\begin{bmatrix}1 & 1 & 1 & \cdots & 1\\ 1 & w & w^2 & \cdots & w^{N-1}\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & w^{N-1} & w^{2(N-1)} & \cdots & w^{(N-1)(N-1)}\end{bmatrix},\quad w=\mathrm{e}^{-\mathrm{j}\,2\pi/N} \tag{2.26}$$

由此可导出对称形式的离散 Fourier 变换对

$$\hat{x}(k)=\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1}x(n)\mathrm{e}^{-\mathrm{j}\,2\pi nk/N},\quad k=0,1,\cdots,N-1 \tag{2.27}$$

$$x(n)=\frac{1}{\sqrt{N}}\sum_{k=0}^{N-1}\hat{x}(k)\mathrm{e}^{\mathrm{j}\,2\pi nk/N},\quad n=0,1,\cdots,N-1 \tag{2.28}$$

根据定义易知

$$\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F}=\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}=\boldsymbol{I} \tag{2.29}$$

Fourier 矩阵的性质：

(1) Fourier 矩阵为对称矩阵，即 $\boldsymbol{F}^{\mathrm{T}}=\boldsymbol{F}$。

(2) Fourier 矩阵的逆矩阵等于 Fourier 矩阵的共轭矩阵，即 $\boldsymbol{F}^{-1}=\boldsymbol{F}^{*}$。

(3) $\boldsymbol{F}^2=\boldsymbol{P}=[\boldsymbol{e}_1,\boldsymbol{e}_n,\boldsymbol{e}_{n-1},\cdots,\boldsymbol{e}_2]$ (置换矩阵)，其中，$\boldsymbol{e}_k$ 是标准向量 (仅第 k 个元素为 1，其他元素皆为 0 的向量)。

(4) $\sqrt{N}\boldsymbol{F}=\boldsymbol{C}+\mathrm{j}\,\boldsymbol{S}$，其中，矩阵 $\boldsymbol{C}$ 和 $\boldsymbol{S}$ 的第 i 行和第 j 列的元素分别为

$$C_{ij}=\cos\left(\frac{2\pi}{N}(i-1)(j-1)\right)$$

$$S_{ij}=\sin\left(\frac{2\pi}{N}(i-1)(j-1)\right)$$

式中，$i,j=1,2,\cdots,N$。

(5) $\boldsymbol{F}^4=\boldsymbol{I}$。

(6) $\boldsymbol{C}\boldsymbol{S}=(\boldsymbol{S}\boldsymbol{C})^{\mathrm{T}}$ 和 $\boldsymbol{C}^2+\boldsymbol{S}^2=\boldsymbol{I}$。

2.1.9 Hankel 矩阵

正方矩阵 $\boldsymbol{A}\in C^{(n+1)\times(n+1)}$ 称为 Hankel 矩阵，若

$$\boldsymbol{A}=\begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_n \\ a_1 & a_2 & a_3 & \cdots & a_{n+1} \\ a_2 & a_3 & a_4 & \cdots & a_{n+2} \\ \vdots & \vdots & \vdots & & \vdots \\ a_n & a_{n+1} & a_{n+2} & \cdots & a_{2n} \end{bmatrix} \tag{2.30}$$

Hankel 矩阵是一个交叉对角线上具有相同元素的矩阵。

2.1.10 Hadamard 矩阵

$\boldsymbol{H}_n\in R^{n\times n}$称为 Hadamard 矩阵，若它的所有元素取 +1 或者 −1，并且满足

$$\boldsymbol{H}_n\boldsymbol{H}_n^{\mathrm{T}}=\boldsymbol{H}_n^{\mathrm{T}}\boldsymbol{H}_n=n\boldsymbol{I}_n \tag{2.31}$$

定理 2.4 令 $n=2^k$，$k=1,2,\cdots$，则规范化的标准正交 Hadamard 矩阵具有通用构造公式：

$$\boldsymbol{H}_{2n}=\frac{1}{\sqrt{2}}\begin{bmatrix}\boldsymbol{H}_n & \boldsymbol{H}_n\\ \boldsymbol{H}_n & -\boldsymbol{H}_n\end{bmatrix},\qquad \boldsymbol{H}_2=\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} \tag{2.32}$$

也可以利用矩阵的 Kronecker 积写成 $\boldsymbol{H}_{2n}=\boldsymbol{H}_2\otimes\boldsymbol{H}_n$。

由于 Hadamard 矩阵是规范化的标准正交矩阵，并且元素只取 $+1$ 或 -1，故 Hadamard 矩阵是唯一只使用加法和减法的标准正交变换。Hadamard 矩阵可以用作移动通信中的编码，得到的码称为 Hadamard 码 (或称 Walsh-Hadamard 码)。另外，由于 Hadamard 矩阵的行向量之间的正交性，行向量可以用来仿真码分多址中各个用户的扩频波形向量。

2.2 习题与解答

2.1 证明多个正交矩阵的乘积仍然为正交矩阵。

证明：设 $\boldsymbol{A}_1,\cdots,\boldsymbol{A}_N$ 是 N 个 $n\times n$ 维的正交矩阵，即 $\boldsymbol{A}_i\boldsymbol{A}_i^{\mathrm{T}}=\boldsymbol{I}$，其乘积为 $\boldsymbol{B}=\boldsymbol{A}_1\boldsymbol{A}_2\cdots\boldsymbol{A}_N$，则

$$\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}=(\boldsymbol{A}_1\boldsymbol{A}_2\cdots\boldsymbol{A}_N)(\boldsymbol{A}_1\boldsymbol{A}_2\cdots\boldsymbol{A}_N)^{\mathrm{T}}=\boldsymbol{A}_1\cdots\boldsymbol{A}_{N-1}\boldsymbol{A}_N\boldsymbol{A}_N^{\mathrm{T}}\boldsymbol{A}_{N-1}^{\mathrm{T}}\cdots\boldsymbol{A}_1^{\mathrm{T}}=\boldsymbol{I}$$
$$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}=(\boldsymbol{A}_1\boldsymbol{A}_2\cdots\boldsymbol{A}_N)^{\mathrm{T}}(\boldsymbol{A}_1\boldsymbol{A}_2\cdots\boldsymbol{A}_N)=\boldsymbol{A}_N^{\mathrm{T}}\cdots\boldsymbol{A}_2^{\mathrm{T}}\boldsymbol{A}_1^{\mathrm{T}}\boldsymbol{A}_1\boldsymbol{A}_2\cdots\boldsymbol{A}_N=\boldsymbol{I}$$

故 $\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}=\boldsymbol{I}$，即 $\boldsymbol{B}$ 是正交矩阵，命题得证。

2.2 令 $\boldsymbol{A}$ 为实对称矩阵，$\boldsymbol{B}$ 为实反对称矩阵，且这两个矩阵是乘积可交换的，即 $\boldsymbol{AB}=\boldsymbol{BA}$。证明：若 $\boldsymbol{A}-\boldsymbol{B}$ 是非奇异的，则 $(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})^{-1}$ 是正交矩阵。

证明：由题意可知，$\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A},\boldsymbol{B}^{\mathrm{T}}=-\boldsymbol{B},\boldsymbol{AB}=\boldsymbol{BA}$，故 $(\boldsymbol{A}-\boldsymbol{B})(\boldsymbol{A}+\boldsymbol{B})=(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})=\boldsymbol{A}^2-\boldsymbol{B}^2$。因为

$$\begin{aligned}
&[(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})^{-1}]^{\mathrm{T}}[(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})^{-1}]\\
=&(\boldsymbol{A}-\boldsymbol{B})^{-\mathrm{T}}(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}}(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})^{-1}\\
=&(\boldsymbol{A}+\boldsymbol{B})^{-1}(\boldsymbol{A}-\boldsymbol{B})(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})^{-1}\\
=&(\boldsymbol{A}+\boldsymbol{B})^{-1}(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})^{-1}\\
=&\boldsymbol{I}
\end{aligned}$$

且

$$\begin{aligned}
&[(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})^{-1}][(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})^{-1}]^{\mathrm{T}}\\
=&(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})^{-1}(\boldsymbol{A}-\boldsymbol{B})^{-\mathrm{T}}(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}}\\
=&(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})^{-1}(\boldsymbol{A}+\boldsymbol{B})^{-1}(\boldsymbol{A}-\boldsymbol{B})\\
=&(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}+\boldsymbol{B})^{-1}(\boldsymbol{A}-\boldsymbol{B})^{-1}(\boldsymbol{A}-\boldsymbol{B})\\
=&\boldsymbol{I}
\end{aligned}$$

所以，$(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})^{-1}$ 是正交矩阵。

2.3 令 $\boldsymbol{E}_{\alpha(p)}\boldsymbol{A}$ 是使矩阵 $\boldsymbol{A}$ 第 p 行乘常数 α 的基本矩阵，且 $\boldsymbol{E}_{(p)+\alpha(q)}\boldsymbol{A}$ 第 q 行乘非零常数α 后，加到 $\boldsymbol{A}$ 的第 p 行的基本矩阵。证明基本矩阵的下列性质：

(1) $\det\left(\boldsymbol{E}_{\alpha(p)}\right)=\alpha$。

(2) $\det\left(\boldsymbol{E}_{(p)+\alpha(q)}\right)=1$。

证明：(1) 因为 $\boldsymbol{E}_{\alpha(p)}$ 是用常数 α 乘以单位矩阵 $\boldsymbol{I}_n$ 的第 p 行 (或第 p 列) 所得到的矩阵，即

$$\boldsymbol{E}_{\alpha(p)}=\begin{bmatrix}1 & \cdots & 0 & \cdots & 0\\ \vdots & \ddots & \vdots & & \vdots\\ 0 & \cdots & \alpha & \cdots & 0\\ \vdots & & \vdots & \ddots & \vdots\\ 0 & \cdots & 0 & \cdots & 1\end{bmatrix}$$

所以，$\det\left(\boldsymbol{E}_{\alpha(p)}\right)=\alpha(-1)^{p+p}\det\left(\boldsymbol{E}_p\right)=\alpha\det(\boldsymbol{I})=\alpha$，其中 $\boldsymbol{E}_p$ 是矩阵 $\boldsymbol{E}_{\alpha(p)}$ 去掉第 p 行第 p 列所得的矩阵。

(2) 不失一般性，假设 $p<q$，则

$$\boldsymbol{E}_{(p)+\alpha(q)}=\begin{array}{c}\begin{bmatrix}1 & \cdots & 0 & \cdots & 0 & \cdots & 0\\ \vdots & \ddots & \vdots & & \vdots & & \vdots\\ 0 & \cdots & 1 & \cdots & \alpha & \cdots & 0\\ \vdots & & \vdots & \ddots & \vdots & & \vdots\\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0\\ \vdots & & \vdots & & \vdots & \ddots & \vdots\\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1\end{bmatrix}\end{array}\begin{array}{l}\\ \\ \longleftarrow \text{第 } p \text{ 行}\\ \\ \longleftarrow \text{第 } q \text{ 行}\\ \\ \\ \end{array}$$

所以，$\det\left(\boldsymbol{E}_{(p)+\alpha(q)}\right)=1\cdot(-1)^{p+p}\det\left(\boldsymbol{E}_p\right)+\alpha(-1)^{p+q}\det\left(\boldsymbol{E}_{p,q}\right)$，其中 $\boldsymbol{E}_{p,q}$ 表示去掉第 p 行第 q 列后所得到的矩阵，它是对角线元素中存在零的上三角矩阵，因此 $\det\left(\boldsymbol{E}_{p,q}\right)=0$，从而有 $\det\left(\boldsymbol{E}_{(p)+\alpha(q)}\right)=\det\left(\boldsymbol{E}_p\right)=\det(\boldsymbol{I})=1$。

2.4 令 $\boldsymbol{A}_{n\times n}$ 是下三角矩阵。

(1) 求 $\boldsymbol{A}$ 可对角化的条件。

(2) 若 $a_{11}=a_{22}=\cdots=a_{nn}$，且至少有一个元素 $a_{ij}\neq 0\ (i>j)$，证明：矩阵$\boldsymbol{A}$ 不可对角化。

解：(1) 如果矩阵 $\boldsymbol{A}\in C^{n\times n}$ 与一个对角矩阵相似，则称 $\boldsymbol{A}$ 是可对角化的。$\boldsymbol{A}$ 与对角矩阵相似的充分必要条件是 $\boldsymbol{A}$ 的 n 个特征向量线性无关。由于特征多项式为

$$f(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})=(\lambda-a_{11})(\lambda-a_{22})\cdots(\lambda-a_{nn})$$

所以，当 $a_{ii}\neq a_{jj}\ (i\neq j,\ i,j=1,\cdots,n)$ 时，$\boldsymbol{A}$ 具有 n 个不同的特征值，其对应的 n 个特征向量必然线性无关。即 $a_{ii}\neq a_{jj}$ 时，$\boldsymbol{A}$ 一定是可对角化的。

(2) 假定

$$\boldsymbol{A}=\begin{bmatrix}a_{11} & & & \\ & a_{22} & & \\ & a_{i_0j_0} & \ddots & \\ & & & a_{nn}\end{bmatrix}$$

与对角矩阵

$$
\boldsymbol{B}=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}
$$

相似，则由于 $\boldsymbol{A}$ 的特征多项式 $f(\lambda)=(\lambda-a_{11})^n$，所以 $\lambda_1=\lambda_2=\cdots=\lambda_n=a_{11}$，即 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 具有相同的特征值。由此可知，$\boldsymbol{B}=a_{11}\boldsymbol{I}$ 是一个数量矩阵，而由于一个数量矩阵只能与它自身相似 (因为 $\boldsymbol{S}^{-1}\boldsymbol{B}\boldsymbol{S}=a_{11}\boldsymbol{S}^{-1}\boldsymbol{S}=a_{11}\boldsymbol{I}=\boldsymbol{B}$)，故 $\boldsymbol{A}$ 不可能与对角矩阵相似，即不可对角化。

2.5 令

$$
\boldsymbol{T}=\begin{bmatrix}t_{11} & t_{12} & \cdots & t_{1n}\\ 0 & t_{22} & \cdots & t_{2n}\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & t_{nn}\end{bmatrix}
$$

为上三角矩阵。证明：

(1) 若对某个 $1\leqslant i\leqslant n$ 有 $t_{ii}=0$，则 $\boldsymbol{T}$ 奇异。

(2) 若 $t_{ii}\neq 0, i=1,2,\cdots,n$，则 $\boldsymbol{T}$ 非奇异。

证明： 因为 $\det(\boldsymbol{T})=\prod\limits_{i=1}^{n}t_{ii}$，所以

$$
\boldsymbol{T}\text{ 非奇异} \quad\Leftrightarrow\quad \det(\boldsymbol{T})\neq 0 \quad\Leftrightarrow\quad t_{ii}\neq 0,\quad i=1,2,\cdots,n
$$

即只要存在一个 $t_{ii}=0$，$\boldsymbol{T}$ 都是奇异的，命题 (1) 和 (2) 得证。

2.6 设矩阵

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

求 c 值，使得 $\boldsymbol{B}=\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ 为对角矩阵。并求出矩阵 $\boldsymbol{P}$ 和 $\boldsymbol{B}$。

解： $\boldsymbol{A}$ 的特征多项式为

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

求解特征方程 $\det(\boldsymbol{A}-\lambda\boldsymbol{I})=0$ 得到矩阵 $\boldsymbol{A}$ 的 3 个特征值 $\lambda=1,-1,-1$，其中 -1 是二重根。将 $\lambda=1$ 和 $\lambda=-1$ 分别代入 $\boldsymbol{A}-\lambda\boldsymbol{I}$ 得

$$
\boldsymbol{A}-1\boldsymbol{I}=\begin{bmatrix}2 & 2 & -2\\ -c & -2 & c\\ 4 & 2 & -4\end{bmatrix},\quad \boldsymbol{A}+1\boldsymbol{I}=\begin{bmatrix}4 & 2 & -2\\ -c & 0 & c\\ 4 & 2 & -2\end{bmatrix}
$$

由 $\text{rank}(\boldsymbol{A}-1\boldsymbol{I})=2$ 以及 $\text{rank}(\boldsymbol{A}+1\boldsymbol{I})=1$ 易知，$c=0$。此时，

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

当 $\lambda = 1$ 时，由 $(\boldsymbol{A} - 1\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$，即

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

可得其解为 $x_1 = x_3$ 和 $x_2 = 0$，其中 x_3 任意。因此，与特征值 $\lambda = 1$ 对应的特征向量为

$$\boldsymbol{x} = \begin{bmatrix} a \\ 0 \\ a \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad a \neq 0$$

取 $a = 1$，得特征向量为 $\boldsymbol{x}_1 = [1, 0, 1]^{\mathrm{T}}$。

当 $\lambda = -1$ 时，由 $(\boldsymbol{A} + 1\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$，即

$$4x_1 + 2x_2 - 2x_3 = 0$$

可得其解为 $x_3 = 2x_1 + x_2$，其中 x_1 和 x_2 任意。因此，与特征值 $\lambda = -1$ 对应得特征向量为

$$\boldsymbol{x} = \begin{bmatrix} a \\ b \\ 2a + b \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

取 $a = 1, b = 0$ 和 $a = 0, b = 1$，得特征向量为 $\boldsymbol{x}_1 = [1, 0, 2]^{\mathrm{T}}$ 和 $\boldsymbol{x}_2 = [0, 1, 1]^{\mathrm{T}}$。

由于 $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ 线性无关，所以矩阵 $\boldsymbol{A}$ 一定与一个对角矩阵相似。因此，

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

$$\boldsymbol{B} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} 1 & & \\ & -1 & \\ & & -1 \end{bmatrix}$$

2.7 若 $\boldsymbol{A}$ 为幂等矩阵和对称矩阵，证明 $\boldsymbol{A}$ 是半正定的。

证明： 对称矩阵 $\boldsymbol{A}$ 是半正定的，若 $\boldsymbol{A}$ 的所有特征值都不小于零。假设对称矩阵 $\boldsymbol{A}$ 的特征值分解为 $\boldsymbol{A} = \boldsymbol{U\Sigma U}^{\mathrm{H}}$，其中 $\boldsymbol{\Sigma} = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$。由题意知，

$$\boldsymbol{A} = \boldsymbol{A}^2 = \boldsymbol{U\Sigma U}^{\mathrm{H}}\boldsymbol{U\Sigma U}^{\mathrm{H}} = \boldsymbol{U\Sigma}^2\boldsymbol{U}^{\mathrm{H}}$$

所以，$\boldsymbol{A}$ 的特征值为 $\lambda_i^2 \geqslant 0\ (i = 1, \cdots, n)$，即 $\boldsymbol{A}$ 是半正定的。

2.8 证明：若 $\boldsymbol{A}$ 为正定矩阵，并且 $F(z) = \sum\limits_{n=0}^{\infty} c_n z^n$，$c_n > 0$，则 $F(\boldsymbol{A})$ 正定。

证明： 由于矩阵 $\boldsymbol{A}$ 是 Hermitian 矩阵且正定，所以 $\boldsymbol{A}$ 的特征值分解为 $\boldsymbol{A} = \boldsymbol{U\Sigma U}^{\mathrm{H}}$，其中 $\boldsymbol{U}$ 是酉矩阵，$\boldsymbol{\Sigma} = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$，且 $\lambda_i > 0\ (i = 1, \cdots, n)$。于是 $F(\boldsymbol{A}) = \sum\limits_{n=0}^{\infty} c_n \boldsymbol{A}^n$ 的特征值 $\lambda_{F(\boldsymbol{A})_i} = \sum\limits_{n=0}^{\infty} c_n \lambda_i^n > 0$, $\forall i \in \{1, \cdots, n\}$，即 $F(\boldsymbol{A})$ 正定。

2.9 令矩阵 $\boldsymbol{A}$ 正定，并且其元素 $a_{ij} \geqslant 0$。证明

$$\boldsymbol{A}(n) = \begin{bmatrix} a_{11}^n & a_{12}^n \\ a_{21}^n & a_{22}^n \end{bmatrix}, \quad n > 0$$

正定。

证明： 利用 Hadamard 积的性质：若 $m \times m$ 矩阵 $\boldsymbol{A}, \boldsymbol{B}$ 是正定（或半正定）的，则它们的 Hadamard 积 $\boldsymbol{A} \odot \boldsymbol{B}$ 也是正定（或半正定）的。

当 $n = 1$ 时，$\boldsymbol{A}(1) = \boldsymbol{A}$ 正定。假设 $n-1$ 时，$\boldsymbol{A}(n-1)$正定，则 $\boldsymbol{A}(n) = \boldsymbol{A}(n-1) \odot \boldsymbol{A}$ 正定。由数学归纳法知，$\boldsymbol{A}(n)$ 正定。

2.10 证明：一个维数为奇数的反对称矩阵，其行列式必等于零。

证明： 设 $\boldsymbol{A}$ 是 $n \times n$ 维的反对称矩阵，其中 n 为奇数。因为

$$\det(\boldsymbol{A}) = \det(\boldsymbol{A}^{\mathrm{T}}) = \det(-\boldsymbol{A}) = (-1)^n \det(\boldsymbol{A}) = -\det(\boldsymbol{A})$$

所以，$\det(\boldsymbol{A}) = 0$。

2.11 令$\boldsymbol{x} = [1, 3, 5, 7, 9]^{\mathrm{T}}$，求矩阵$\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$，使得

(1) $1^2 + 3^2 + 5^2 + 7^2 + 9^2 = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$。

(2) $(1 + 3 + 5 + 7 + 9)^2/5 = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x}$。

(3) $(1-5)^2 + (3-5)^2 + (5-5)^2 + (7-5)^2 + (9-5)^2 = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{x}$。

解： 由 (1) 知，$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$，所以 $\boldsymbol{A} = \boldsymbol{I}$。

由 (2) 知，$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x} = \dfrac{1}{5}(\boldsymbol{1}^{\mathrm{T}}\boldsymbol{x})^2 = \dfrac{1}{5}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{1}\boldsymbol{1}^{\mathrm{T}}\boldsymbol{x}$，所以 $\boldsymbol{B} = \dfrac{1}{5}\boldsymbol{1}\boldsymbol{1}^{\mathrm{T}} = \dfrac{1}{5}\boldsymbol{J}_5$，其中 $\boldsymbol{J}_5$ 是 5×5 维所有元素都为 1 的矩阵。

由 (3) 知，$\boldsymbol{C}$ 为中心化矩阵，即 $\boldsymbol{C} = \boldsymbol{I}_5 - \bar{\boldsymbol{J}}_5 = \boldsymbol{I}_5 - \dfrac{1}{5}\boldsymbol{J}_5 = \boldsymbol{I} - \boldsymbol{B}$。

2.12 设

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \boldsymbol{P} \begin{bmatrix} \xi \\ \eta \\ \zeta \end{bmatrix}$$

为一正交变换，它将二次曲面方程

$$x^2 + ay^2 + z^2 + 2bxy + 2xz + 2yz = 4$$

化为椭圆柱面方程 $\eta^2 + 4\zeta^2 = 4$，求 a, b 的数值和正交矩阵 $\boldsymbol{P}$。

解： 将二次曲面方程用二次型表示为

$$\begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 1 & b & 1 \\ b & a & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 4$$

即

$$[\,\xi \quad \eta \quad \zeta\,]\boldsymbol{P}^{\mathrm{T}}\begin{bmatrix}1 & b & 1\\ b & a & 1\\ 1 & 1 & 1\end{bmatrix}\boldsymbol{P}\begin{bmatrix}\xi\\ \eta\\ \zeta\end{bmatrix}=4$$

又椭圆柱面方程的二次型表示为

$$[\,\xi \quad \eta \quad \zeta\,]\begin{bmatrix}0 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 4\end{bmatrix}\begin{bmatrix}\xi\\ \eta\\ \zeta\end{bmatrix}$$

所以

$$\underbrace{\boldsymbol{P}^{\mathrm{T}}\begin{bmatrix}1 & b & 1\\ b & a & 1\\ 1 & 1 & 1\end{bmatrix}\boldsymbol{P}}_{\boldsymbol{A}}=\underbrace{\begin{bmatrix}0 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 4\end{bmatrix}}_{\boldsymbol{\Sigma}}$$

由此可知，$\boldsymbol{A}=\boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{P}^{\mathrm{T}}$，即 $\boldsymbol{A}$ 的特征值为 0,1 和 4 。所以，

$$\det(\lambda\boldsymbol{I}-\boldsymbol{A})=\begin{vmatrix}\lambda-1 & -b & -1\\ -b & \lambda-a & -1\\ -1 & -1 & \lambda-1\end{vmatrix}=0$$

的根为 0,1 和 4 ，于是可得 $a=3, b=1$，以及

$$\boldsymbol{P}=\begin{bmatrix}\dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{3}} & \dfrac{1}{\sqrt{6}}\\ 0 & -\dfrac{1}{\sqrt{3}} & \dfrac{2}{\sqrt{6}}\\ -\dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{3}} & \dfrac{1}{\sqrt{6}}\end{bmatrix}$$

2.13 一个 n 阶 Helmert 矩阵 $\boldsymbol{H}_n$ 的第 1 行为 $n^{-1/2}\boldsymbol{1}_n^{\mathrm{T}}$，其他 $n-1$ 行具有分块形式 [23, p.71]

$$\frac{1}{\sqrt{\lambda_i}}\left[\boldsymbol{1}_i^{\mathrm{T}},-i,\boldsymbol{0}_{n-i-1}^{\mathrm{T}}\right],\quad \lambda_i=i(i+1),\ i=1,2,\cdots,n-1$$

式中，$\boldsymbol{1}_i^{\mathrm{T}}$ 和 $\boldsymbol{0}_i^{\mathrm{T}}$ 分别表示元素全部为 1 和 0 的 i 阶行向量。例如，

$$\boldsymbol{H}_4=\begin{bmatrix}1/\sqrt{4} & 1/\sqrt{4} & 1/\sqrt{4} & 1/\sqrt{4}\\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0\\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & 0\\ 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & -3\sqrt{12}\end{bmatrix}$$

将 n 阶 Helmert 矩阵分块为

$$\boldsymbol{H}=\begin{bmatrix}\boldsymbol{h}^{\mathrm{T}}\\ \boldsymbol{K}\end{bmatrix}$$

式中，$\boldsymbol{h}=n^{-1/2}\boldsymbol{1}$，而 $\boldsymbol{K}$ 表示 $\boldsymbol{H}$ 的最后 $n-1$ 行。

(1) 证明 $\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}=\boldsymbol{I}_n$。

(2) 对于 n 阶向量 $\boldsymbol{x}$，证明

$$n\bar{x}_n^2=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{h}\boldsymbol{h}^{\mathrm{T}}\boldsymbol{x}$$

式中，$\bar{x}_n = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$；并证明

$$S_n = \sum_{i=1}^{n}\left(x_i - \sum_{k=1}^{n} x_k/n\right)^2$$

可以表示为

$$S_n = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{x}$$

(3) 推导递推公式

$$S_n = S_{n-1} + (1-1/n)(\bar{x}_{n-1} - x_n)^2$$

式中，$\bar{x}_{n-1} = \frac{1}{n-1}\sum\limits_{i=1}^{n-1} x_i$。

证明：(1) 因为 $\boldsymbol{h}^{\mathrm{T}}\boldsymbol{h} = n^{-\frac{1}{2}}\boldsymbol{1}^{\mathrm{T}} \cdot n^{-\frac{1}{2}}\boldsymbol{1} = 1$，

$$\frac{1}{\sqrt{\lambda_i}}\left[\boldsymbol{1}_i^{\mathrm{T}}, -i, \boldsymbol{0}_{n-i-1}^{\mathrm{T}}\right] = \frac{1}{\sqrt{n\lambda_i}}\left[\boldsymbol{1}_i^{\mathrm{T}}, -i, \boldsymbol{0}_{n-i-1}^{\mathrm{T}}\right]\boldsymbol{1}_n = 0 \quad \Longrightarrow \quad \boldsymbol{K}\boldsymbol{h} = 0$$

又因为

$$\frac{1}{\lambda_i}\left[\boldsymbol{1}_i^{\mathrm{T}}, -i, \boldsymbol{0}_{n-i-1}^{\mathrm{T}}\right]\begin{bmatrix}\boldsymbol{1}_i\\ -i\\ \boldsymbol{0}_{n-i-1}\end{bmatrix} = \frac{1}{\lambda_i}(\boldsymbol{1}_i^{\mathrm{T}}\boldsymbol{1}_i + i^2) = 1$$

且

$$\frac{1}{\sqrt{\lambda_i\lambda_j}}\left[\boldsymbol{1}_i^{\mathrm{T}}, -i, \boldsymbol{0}_{n-i-1}^{\mathrm{T}}\right]\begin{bmatrix}\boldsymbol{1}_j\\ -j\\ \boldsymbol{0}_{n-j-1}\end{bmatrix} = \begin{cases}\dfrac{1}{\sqrt{\lambda_i\lambda_j}}(\boldsymbol{1}_i^{\mathrm{T}}\boldsymbol{1}_i - i) = 0, & \text{若 } i \leqslant j\\ \dfrac{1}{\sqrt{\lambda_i\lambda_j}}(\boldsymbol{1}_j^{\mathrm{T}}\boldsymbol{1}_j - j) = 0, & \text{若 } i > j\end{cases}$$

于是有 $\boldsymbol{K}\boldsymbol{K}^{\mathrm{T}} = \boldsymbol{I}$。因此，

$$\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}} = \begin{bmatrix}\boldsymbol{h}^{\mathrm{T}}\\ \boldsymbol{K}\end{bmatrix}\begin{bmatrix}\boldsymbol{h} & \boldsymbol{K}^{\mathrm{T}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{h}^{\mathrm{T}}\boldsymbol{h} & \boldsymbol{h}^{\mathrm{T}}\boldsymbol{K}^{\mathrm{T}}\\ \boldsymbol{K}\boldsymbol{h} & \boldsymbol{K}\boldsymbol{K}^{\mathrm{T}}\end{bmatrix} = \boldsymbol{I}$$

(2) 因为 $\boldsymbol{h}^{\mathrm{T}}\boldsymbol{x} = n^{-\frac{1}{2}}\boldsymbol{1}^{\mathrm{T}}\boldsymbol{x} = n^{-\frac{1}{2}}\sum\limits_{i=1}^{n} x_i$，所以 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{h}\boldsymbol{h}^{\mathrm{T}}\boldsymbol{x} = \left(\frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n} x_i\right)^2 = n\bar{x}_n^2$。又

$$S_n = \sum_{i=1}^{n}\left(x_i - \sum_{k=1}^{n} x_k/n\right)^2 = \sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{x}$$

其中 $\boldsymbol{C}$ 为中心化矩阵，即 $\boldsymbol{C} = \boldsymbol{I} - \frac{1}{n}\boldsymbol{1}_n\boldsymbol{1}_n^{\mathrm{T}} = \boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^{\mathrm{T}}$。由 (1) 中证明知，$\boldsymbol{H}$ 的行组成标准正交组，所以 $\boldsymbol{H}$ 是正交矩阵，即 $\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}} = \boldsymbol{H}^{\mathrm{T}}\boldsymbol{H} = \boldsymbol{I}$，而 $\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}} = \boldsymbol{h}\boldsymbol{h}^{\mathrm{T}} + \boldsymbol{K}^{\mathrm{T}}\boldsymbol{K} = \boldsymbol{I}$，所以 $\boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^{\mathrm{T}} = \boldsymbol{K}^{\mathrm{T}}\boldsymbol{K} = \boldsymbol{C}$。于是有，$S_n = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{x}$。

(3) 由 (2) 知，

$$S_n = \boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{C}\boldsymbol{x}_n = \boldsymbol{x}_n^{\mathrm{T}}(\boldsymbol{I} - \frac{1}{n}\boldsymbol{1}_n\boldsymbol{1}_n^{\mathrm{T}})\boldsymbol{x}_n = \boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{x}_n - \frac{1}{n}\boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{1}_n\boldsymbol{1}_n^{\mathrm{T}}\boldsymbol{x}_n$$

$$S_{n-1} = \boldsymbol{x}_{n-1}^{\mathrm{T}}\boldsymbol{x}_{n-1} - \frac{1}{n-1}\boldsymbol{x}_{n-1}^{\mathrm{T}}\boldsymbol{1}_{n-1}\boldsymbol{1}_{n-1}^{\mathrm{T}}\boldsymbol{x}_{n-1}$$

其中

$$\boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{x}_n = \boldsymbol{x}_{n-1}^{\mathrm{T}}\boldsymbol{x}_{n-1} + x_n^2,$$
$$\boldsymbol{1}_n^{\mathrm{T}}\boldsymbol{x}_n = \boldsymbol{1}_{n-1}^{\mathrm{T}}\boldsymbol{x}_{n-1} + x_n$$

所以

$$\begin{aligned}
S_n &= \boldsymbol{x}_{n-1}^{\mathrm{T}}\boldsymbol{x}_{n-1} + x_n^2 - \frac{1}{n}\left(x_n + \boldsymbol{1}_{n-1}^{\mathrm{T}}\boldsymbol{x}_{n-1}\right)^2 \\
&= \boldsymbol{x}_{n-1}^{\mathrm{T}}\boldsymbol{x}_{n-1} - \frac{1}{n-1}\boldsymbol{x}_{n-1}^{\mathrm{T}}\boldsymbol{1}_{n-1}\boldsymbol{1}_{n-1}^{\mathrm{T}}\boldsymbol{x}_{n-1} \\
&\quad + \frac{1}{n(n-1)}\left(\boldsymbol{1}_{n-1}^{\mathrm{T}}\boldsymbol{x}_{n-1}\right)^2 + \frac{n-1}{n}x_n^2 - \frac{2}{n}\boldsymbol{x}_n\boldsymbol{1}_{n-1}^{\mathrm{T}}\boldsymbol{x}_{n-1} \\
&= S_{n-1} + \frac{n}{n-1}\bar{x}_{n-1}^2 + \frac{n-1}{n}x_n^2 - \frac{2(n-1)}{n}x_n\bar{x}_{n-1}^2 \\
&= S_{n-1} + (1-\frac{1}{n})(\bar{x}_{n-1} - x_n)^2
\end{aligned}$$

命题得证。

2.14 令 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为对称矩阵，并且满足

$$|\boldsymbol{I} - \lambda\boldsymbol{A}|\,|\boldsymbol{I} - \mu\boldsymbol{B}| = |\boldsymbol{I} - \lambda\boldsymbol{A} - \mu\boldsymbol{B}|, \quad \forall\,\lambda, \mu$$

证明 $\boldsymbol{AB} = \boldsymbol{O}$ (零矩阵)。

证明： 令 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的特征值分解分别为 $\boldsymbol{A} = \boldsymbol{U\Sigma U}^{\mathrm{H}}$ 和 $\boldsymbol{B} = \boldsymbol{V\Lambda V}^{\mathrm{H}}$，则

$$|\boldsymbol{I} - \lambda\boldsymbol{A}||\boldsymbol{I} - \mu\boldsymbol{B}| = |\boldsymbol{I} - \lambda\boldsymbol{A} - \mu\boldsymbol{B}|$$
$$|\boldsymbol{I} - \lambda\boldsymbol{A}'||\boldsymbol{I} - \mu\boldsymbol{B}'| = |\boldsymbol{I} - \lambda\boldsymbol{A}' - \mu\boldsymbol{B}'|$$

其中，$\boldsymbol{A}' = \boldsymbol{\Sigma}$，$\boldsymbol{B}' = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{V\Lambda V}^{\mathrm{H}}\boldsymbol{U}$。因此，不失一般性，我们假定 $\boldsymbol{A}$ 为对角矩阵。

若 $\boldsymbol{A}$ 可逆，则有

$$|\boldsymbol{I} - \mu\boldsymbol{B}| = |\boldsymbol{I} - \mu(\boldsymbol{I} - \lambda\boldsymbol{A})^{-1}\boldsymbol{B}| \tag{2.33}$$

注意到式(2.33)中等式右边与 λ 无关，也就是说 λ 取任意值，等式均成立。为求矩阵 $\boldsymbol{B}$，令 $\lambda \to \infty$，即有

$$(\boldsymbol{I} - \lambda\boldsymbol{A})^{-1} = \lambda^{-1}(\lambda^{-1} - \boldsymbol{A})^{-1} \xrightarrow{\lambda\to\infty} \boldsymbol{O}$$

于是有 $|\boldsymbol{I} - \mu\boldsymbol{B}| = |\boldsymbol{I}| = 1$，由于 μ 任意选取时该等式都成立，因此 $\boldsymbol{B} = \boldsymbol{O}$，从而可得 $\boldsymbol{AB} = \boldsymbol{O}$。

若 $\boldsymbol{A}$ 不可逆，令 $\boldsymbol{A} = \mathrm{diag}\{\sigma_1, \cdots, \sigma_k, 0, \cdots, 0\}$，于是有

$$|\boldsymbol{I} - \mu\boldsymbol{B}| = |\boldsymbol{I} - \mu\tilde{\boldsymbol{B}}| \tag{2.34}$$

其中

$$\boldsymbol{B} \overset{\text{def}}{=} \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{bmatrix}$$

$$\tilde{\boldsymbol{B}} \overset{\text{def}}{=} \begin{bmatrix} \frac{1}{1-\lambda\sigma_1} & & & \\ & \ddots & & \\ & & \frac{1}{1-\lambda\sigma_k} & \\ & & & \boldsymbol{I} \end{bmatrix} \boldsymbol{B} \xrightarrow{\lambda\to\infty} \begin{bmatrix} \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{B}_{22} \end{bmatrix}$$

式(2.34)说明 $\boldsymbol{B}$ 与 $\tilde{\boldsymbol{B}}$ 的所有特征值相同，故有

$$\text{tr}(\boldsymbol{B}\boldsymbol{B}^{\text{H}}) = \text{tr}(\tilde{\boldsymbol{B}}\tilde{\boldsymbol{B}}^{\text{H}})$$

即矩阵 $\boldsymbol{B}$ 的所有元素模平方和等于矩阵 $\tilde{\boldsymbol{B}}$ 的所有元素模平方和，因此立即有

$$\boldsymbol{B} = \tilde{\boldsymbol{B}} = \begin{bmatrix} \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{B}_{22} \end{bmatrix}$$

从而有 $\boldsymbol{AB} = \boldsymbol{O}$。

综上所述，命题得证。

2.15 证明：若 $\boldsymbol{A}$ 为实反对称矩阵，则 $\boldsymbol{A}+\boldsymbol{I}$ 非奇异。

证明： 因为 $\boldsymbol{A}^{\text{T}} = -\boldsymbol{A}$，所以对于任意 $\boldsymbol{x} \in R^n$，有

$$\boldsymbol{x}^{\text{T}}\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{x}^{\text{T}}\boldsymbol{A}\boldsymbol{x})^{\text{T}} = \boldsymbol{x}^{\text{T}}\boldsymbol{A}^{\text{T}}\boldsymbol{x} = -\boldsymbol{x}^{\text{T}}\boldsymbol{A}\boldsymbol{x} \quad \Rightarrow \quad \boldsymbol{x}^{\text{T}}\boldsymbol{A}\boldsymbol{x} = 0$$

因此，如果 $\boldsymbol{y} \in R^n$ 满足 $(\boldsymbol{I}+\boldsymbol{A})\boldsymbol{y} = \boldsymbol{0}$，则

$$\boldsymbol{y}^{\text{T}}\boldsymbol{y} = \boldsymbol{y}^{\text{T}}\boldsymbol{y} + \boldsymbol{y}^{\text{T}}\boldsymbol{A}\boldsymbol{y} = \boldsymbol{y}^{\text{T}}(\boldsymbol{I}+\boldsymbol{A})\boldsymbol{y} = 0 \quad \Rightarrow \quad \boldsymbol{y} = \boldsymbol{0}$$

也就是说齐次方程组 $(\boldsymbol{I}+\boldsymbol{A})\boldsymbol{y} = \boldsymbol{0}$ 只有零解，于是 $\boldsymbol{I}+\boldsymbol{A}$ 是非奇异的。类似可证 $(\boldsymbol{I}-\boldsymbol{A})$ 也是非奇异的。

2.16 假定$\boldsymbol{A}$ 是一实反对称矩阵，证明 Cayley 变换 $\boldsymbol{T} = (\boldsymbol{I}-\boldsymbol{A})(\boldsymbol{I}+\boldsymbol{A})^{-1}$ 为正交矩阵。

证明：因为 $\boldsymbol{A}^{\mathrm{T}}=-\boldsymbol{A}$，且 $(\boldsymbol{I}+\boldsymbol{A})(\boldsymbol{I}-\boldsymbol{A})=(\boldsymbol{I}-\boldsymbol{A})(\boldsymbol{I}+\boldsymbol{A})$，所以

$$\begin{aligned}
\boldsymbol{T}^{\mathrm{T}}\boldsymbol{T}&=[(\boldsymbol{I}-\boldsymbol{A})(\boldsymbol{I}+\boldsymbol{A})^{-1}]^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{A})(\boldsymbol{I}+\boldsymbol{A})^{-1}\\
&=(\boldsymbol{I}+\boldsymbol{A}^{\mathrm{T}})^{-1}(\boldsymbol{I}-\boldsymbol{A}^{\mathrm{T}})(\boldsymbol{I}-\boldsymbol{A})(\boldsymbol{I}+\boldsymbol{A})^{-1}\\
&=(\boldsymbol{I}-\boldsymbol{A})^{-1}(\boldsymbol{I}+\boldsymbol{A})(\boldsymbol{I}-\boldsymbol{A})(\boldsymbol{I}+\boldsymbol{A})^{-1}\\
&=(\boldsymbol{I}-\boldsymbol{A})^{-1}(\boldsymbol{I}-\boldsymbol{A})(\boldsymbol{I}+\boldsymbol{A})(\boldsymbol{I}+\boldsymbol{A})^{-1}\\
&=\boldsymbol{I}\\
\boldsymbol{T}\boldsymbol{T}^{\mathrm{T}}&=(\boldsymbol{I}-\boldsymbol{A})(\boldsymbol{I}+\boldsymbol{A})^{-1}[(\boldsymbol{I}-\boldsymbol{A})(\boldsymbol{I}+\boldsymbol{A})^{-1}]^{\mathrm{T}}\\
&=(\boldsymbol{I}-\boldsymbol{A})(\boldsymbol{I}+\boldsymbol{A})^{-1}(\boldsymbol{I}-\boldsymbol{A})^{-1}(\boldsymbol{I}+\boldsymbol{A})\\
&=(\boldsymbol{I}-\boldsymbol{A})(\boldsymbol{I}-\boldsymbol{A})^{-1}(\boldsymbol{I}+\boldsymbol{A})^{-1}(\boldsymbol{I}+\boldsymbol{A})\\
&=\boldsymbol{I}
\end{aligned}$$

即 $\boldsymbol{T}^{\mathrm{T}}\boldsymbol{T}=\boldsymbol{T}\boldsymbol{T}^{\mathrm{T}}=\boldsymbol{I}$，故 $\boldsymbol{T}$ 是正交矩阵。

2.17 令 $\boldsymbol{A}$ 是正交矩阵，且 $\boldsymbol{A}+\boldsymbol{I}$ 非奇异。证明：矩阵 $\boldsymbol{A}$ 可表示为 Cayley 变换

$$\boldsymbol{A}=(\boldsymbol{I}-\boldsymbol{S})(\boldsymbol{I}+\boldsymbol{S})^{-1}$$

式中，$\boldsymbol{S}$ 为实反对称矩阵。

证明：由题 2.15 知，$\boldsymbol{S}$ 为实反对称矩阵时 $(\boldsymbol{I}+\boldsymbol{S})$ 非奇异即可逆，所以

$$\boldsymbol{A}=(\boldsymbol{I}-\boldsymbol{S})(\boldsymbol{I}+\boldsymbol{S})^{-1}\quad\Leftrightarrow\quad\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{S})=\boldsymbol{I}-\boldsymbol{S}\quad\Leftrightarrow\quad\boldsymbol{S}=(\boldsymbol{I}+\boldsymbol{A})^{-1}(\boldsymbol{I}-\boldsymbol{A})$$

即原命题等价于证明 $\boldsymbol{S}=(\boldsymbol{I}+\boldsymbol{A})^{-1}(\boldsymbol{I}-\boldsymbol{A})$ 是实反对称矩阵。因为

$$\begin{aligned}
\boldsymbol{S}^{\mathrm{T}}&=(\boldsymbol{I}-\boldsymbol{A}^{\mathrm{T}})(\boldsymbol{I}+\boldsymbol{A}^{\mathrm{T}})^{-1}=(\boldsymbol{I}-\boldsymbol{A}^{\mathrm{T}})\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{I}+\boldsymbol{A}^{\mathrm{T}})^{-1}\\
&=(\boldsymbol{A}-\boldsymbol{I})(\boldsymbol{A}^{-\mathrm{T}}+\boldsymbol{I})^{-1}=-(\boldsymbol{I}-\boldsymbol{A})(\boldsymbol{I}+\boldsymbol{A})^{-1}\\
&=-\boldsymbol{S}
\end{aligned}$$

所以，$\boldsymbol{S}$ 是反对称矩阵，命题得证。

2.18 令$\boldsymbol{P}$ 是一个$n\times n$置换矩阵，证明存在一个正整数k，使得 $\boldsymbol{P}^{k}=\boldsymbol{I}$。（提示：考虑矩阵序列$\boldsymbol{P},\boldsymbol{P}^{2},\boldsymbol{P}^{3},\cdots$。）

证明：设 $\boldsymbol{a}$ 是 $n\times 1$ 维向量。由于 $\boldsymbol{Pa}$ 相当于将向量 $\boldsymbol{a}$ 的元素位置进行交换，而 $\boldsymbol{a}$ 的元素位置的排列共有有限的 $n!$ 种可能，故在无穷反复的置换中必然出现重复的元素排列。设第 k_1 次置换与第 k_2 次置换重复 $(k_1>k_2)$，且 $k=k_1-k_2$，则有 $\boldsymbol{P}^{k_1-k_2}=\boldsymbol{P}^{k}=\boldsymbol{I}$。

2.19 假定 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$ 是两个 $n\times n$ 置换矩阵，证明 $\boldsymbol{PQ}$ 也是一个 $n\times n$ 置换矩阵。

证明： 设 $\boldsymbol{E}_i$ 表示将单位矩阵的任意两行互相交换后得到的矩阵，则 $n\times n$ 维矩阵 $\boldsymbol{P}$ 可表示为一系列的行交换操作，即 $\boldsymbol{P}=\boldsymbol{E}_n\boldsymbol{E}_{n-1}\cdots\boldsymbol{E}_1$。同理，$\boldsymbol{Q}=\boldsymbol{E}'_n\boldsymbol{E}'_{n-1}\cdots\boldsymbol{E}'_1$。故 $\boldsymbol{PQ}=\boldsymbol{E}_n\cdots\boldsymbol{E}_1\boldsymbol{E}'_n\cdots\boldsymbol{E}'_1$，即 $\boldsymbol{PQ}$ 由单位矩阵经过一系列行变换操作所得到，所以 $\boldsymbol{PQ}$ 也是一个 $n\times n$ 维的置换矩阵。

2.20 证明：对于每一个矩阵$\boldsymbol{A}$，都存在一个三角矩阵 $\boldsymbol{T}$，使得 $\boldsymbol{TA}$ 为酉矩阵。

思路分析： 本题首先利用 $(\boldsymbol{TA})(\boldsymbol{TA})^{\mathrm{H}}=\boldsymbol{I}$ 去寻找一个三角矩阵 $\boldsymbol{T}$，然后验证这样找到的矩阵 $\boldsymbol{T}$ 能使得 $(\boldsymbol{TA})^{\mathrm{H}}(\boldsymbol{TA})=\boldsymbol{I}$，即使得 $\boldsymbol{TA}$ 是酉矩阵。

证明： 令 $m\times n$ 维矩阵 $\boldsymbol{A}$ 的奇异值分解为 $\boldsymbol{A}=\boldsymbol{U\Sigma V}^{\mathrm{H}}$，其中 $\boldsymbol{U}$ 为 $m\times m$ 维矩阵，$\boldsymbol{V}$ 是 $n\times n$ 维矩阵，$\boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_1 & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}$，则

$$
\begin{aligned}
&(\boldsymbol{TA})(\boldsymbol{TA})^{\mathrm{H}}=\boldsymbol{TAA}^{\mathrm{H}}\boldsymbol{T}^{\mathrm{H}}=\boldsymbol{I}\\
\Rightarrow\ &\boldsymbol{TU\Sigma V}^{\mathrm{H}}\boldsymbol{V\Sigma}^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{T}^{\mathrm{H}}=(\boldsymbol{TU})\begin{bmatrix}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_1^{\mathrm{H}} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}(\boldsymbol{TU})^{\mathrm{H}}=\boldsymbol{I}\\
\Rightarrow\ &\boldsymbol{TU}=\begin{bmatrix}\boldsymbol{\Sigma}_1^{-1} & \boldsymbol{O}\end{bmatrix}\stackrel{\mathrm{def}}{=}\boldsymbol{B}
\end{aligned}
$$

令 $\boldsymbol{B}^{\mathrm{H}}$ 的 QR 分解为 $\boldsymbol{B}^{\mathrm{H}}=\boldsymbol{QR}$，其中 $\boldsymbol{Q}$ 为酉矩阵，$\boldsymbol{R}$ 为上三角矩阵，故有 $\boldsymbol{B}^{\mathrm{H}}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{T}^{\mathrm{H}}=\boldsymbol{QR}$。令 $\boldsymbol{U}^{\mathrm{H}}=\boldsymbol{Q},\boldsymbol{T}=\boldsymbol{R}^{\mathrm{H}}$，即可使 $(\boldsymbol{TA})(\boldsymbol{TA})^{\mathrm{H}}=\boldsymbol{I}$。

又因为

$$
\begin{aligned}
(\boldsymbol{TA})^{\mathrm{H}}(\boldsymbol{TA})&=\boldsymbol{V\Sigma}^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{TU\Sigma V}^{\mathrm{H}}=\boldsymbol{V\Sigma}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B\Sigma V}^{\mathrm{H}}\\
&=\boldsymbol{V}\begin{bmatrix}\boldsymbol{\Sigma}_1 & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_1^{-1}\\ \boldsymbol{O}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_1^{-1} & \boldsymbol{O}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_1 & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}\boldsymbol{V}^{\mathrm{H}}\\
&=\boldsymbol{V}\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{O}\end{bmatrix}\begin{bmatrix}\boldsymbol{I} & \boldsymbol{O}\end{bmatrix}\boldsymbol{V}^{\mathrm{H}}\\
&=\boldsymbol{I}
\end{aligned}
$$

故此时 $\boldsymbol{TA}$ 为酉矩阵，命题得证。

2.21 令 $\boldsymbol{A}$ 是一个给定的矩阵。证明：可以找到一个主对角线上的元素取 ±1 的对角矩阵 $\boldsymbol{J}$，使得 $\boldsymbol{JA}+\boldsymbol{I}$ 非奇异。

证明： 对于任意正方矩阵 $\boldsymbol{A}$，必然存在一个高斯消元矩阵 $\boldsymbol{J}$，其中 $\boldsymbol{J}$ 对角元素为 1，使得

$$
\boldsymbol{JA}=\begin{bmatrix}b_{11} & b_{12} & \cdots & b_{1n}\\ & b_{22} & \cdots & b_{2n}\\ & & \ddots & \vdots\\ & & & b_{nn}\end{bmatrix}=\boldsymbol{B}
$$

若矩阵 $\boldsymbol{A}$ 的秩为 $k\leqslant n$，则上三角矩阵自第 $k+1$ 行至 n 行全为零。

$$
\det(\boldsymbol{JA}+\boldsymbol{I})=(b_{11}+1)(b_{22}+1)\cdots(b_{nn}+1)
$$

高斯消元之后，若有某个对角元素 b_{ii} 为 -1，则将 $\boldsymbol{J}$ 的 i 行乘 -1，则 $\boldsymbol{B}$ 第 i 行也全乘了 -1，对角元素 b_{ii} 变为 1。所以一定存在一个主对角线上的元素取 ± 1 的对角矩阵 $\boldsymbol{J}$，使得 $\det(\boldsymbol{JA}+\boldsymbol{I})\neq 0$，即命题得证。

2.22 证明：若 $\boldsymbol{H}=\boldsymbol{A}+\mathrm{j}\boldsymbol{B}$ 为 Hermitian 矩阵，其中 $\boldsymbol{A},\boldsymbol{B}$ 为实矩阵，且 $\boldsymbol{A}$ 非奇异，则行列式

$$|\det(\boldsymbol{H})|^2=|\boldsymbol{A}|^2|\boldsymbol{I}+\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}|$$

证明： 由题意 $\boldsymbol{H}^{\mathrm{H}}=\boldsymbol{H}$ 知，$\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}$ 且 $\boldsymbol{B}^{\mathrm{H}}=-\boldsymbol{B}$，即 $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{T}}$ 且 $\boldsymbol{B}^{\mathrm{T}}=-\boldsymbol{B}$。因为

$$\det(\boldsymbol{H})=|\boldsymbol{A}+\mathrm{j}\boldsymbol{B}|=|\boldsymbol{A}||\boldsymbol{I}+\mathrm{j}\boldsymbol{A}^{-1}\boldsymbol{B}|$$
$$\det(\boldsymbol{H}^{\mathrm{T}})=|\boldsymbol{A}-\mathrm{j}\boldsymbol{B}|=|\boldsymbol{A}||\boldsymbol{I}-\mathrm{j}\boldsymbol{A}^{-1}\boldsymbol{B}|$$

而 $\det(\boldsymbol{H})=\det(\boldsymbol{H}^{\mathrm{T}})$，所以

$$|\det(\boldsymbol{H})|^2=|\boldsymbol{A}|^2|(\boldsymbol{I}+\mathrm{j}\boldsymbol{A}^{-1}\boldsymbol{B})(\boldsymbol{I}-\mathrm{j}\boldsymbol{A}^{-1}\boldsymbol{B})|=|\boldsymbol{A}|^2|\boldsymbol{I}+\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}|$$

2.23 令 $\boldsymbol{A},\boldsymbol{S}$ 为 $n\times n$ 矩阵，且 $\boldsymbol{S}$ 非奇异。

(1) 证明 $(\boldsymbol{S}^{-1}\boldsymbol{AS})^2=\boldsymbol{S}^{-1}\boldsymbol{A}^2\boldsymbol{S}$ 和 $(\boldsymbol{S}^{-1}\boldsymbol{AS})^3=\boldsymbol{S}^{-1}\boldsymbol{A}^3\boldsymbol{S}$。

(2) 利用数学归纳法证明 $(\boldsymbol{S}^{-1}\boldsymbol{AS})^k=\boldsymbol{S}^{-1}\boldsymbol{A}^k\boldsymbol{S}$，其中，$k$ 为正整数。

证明： (1) 由于 $\boldsymbol{S}$ 非奇异，因此 $\boldsymbol{S}^{-1}$ 存在，于是有

$$(\boldsymbol{S}^{-1}\boldsymbol{AS})^2=\boldsymbol{S}^{-1}\boldsymbol{ASS}^{-1}\boldsymbol{AS}=\boldsymbol{S}^{-1}\boldsymbol{A}^2\boldsymbol{S}$$
$$(\boldsymbol{S}^{-1}\boldsymbol{AS})^3=\boldsymbol{S}^{-1}\boldsymbol{AS}(\boldsymbol{S}^{-1}\boldsymbol{AS})^2=\boldsymbol{S}^{-1}\boldsymbol{ASS}^{-1}\boldsymbol{A}^2\boldsymbol{S}=\boldsymbol{S}^{-1}\boldsymbol{A}^3\boldsymbol{S}$$

(2) 假设 $(\boldsymbol{S}^{-1}\boldsymbol{AS})^{k-1}=\boldsymbol{S}^{-1}\boldsymbol{A}^{k-1}\boldsymbol{S}$，则

$$(\boldsymbol{S}^{-1}\boldsymbol{AS})^k=\boldsymbol{S}^{-1}\boldsymbol{AS}(\boldsymbol{S}^{-1}\boldsymbol{AS})^{k-1}=\boldsymbol{S}^{-1}\boldsymbol{ASS}^{-1}\boldsymbol{A}^{k-1}\boldsymbol{S}=\boldsymbol{S}^{-1}\boldsymbol{A}^k\boldsymbol{S}$$

综上所述，命题得证。

2.24 证明：若 $\boldsymbol{A}$ 是可对角化的，并且 $\boldsymbol{B}$ 与 $\boldsymbol{A}$ 相似，则 $\boldsymbol{B}$ 是可对角化的。(提示：假定 $\boldsymbol{S}^{-1}\boldsymbol{AS}=\boldsymbol{D}$和$\boldsymbol{W}^{-1}\boldsymbol{AW}=\boldsymbol{B}$。)

证明： 因为 $\boldsymbol{A}$ 是可对角化的，即 $\boldsymbol{A}=\boldsymbol{SDS}^{-1}$，其中 $\boldsymbol{D}$ 是对角矩阵。所以

$$\boldsymbol{B}=\boldsymbol{W}^{-1}\boldsymbol{AW}=\boldsymbol{W}^{-1}\boldsymbol{SDS}^{-1}\boldsymbol{W}=(\boldsymbol{S}^{-1}\boldsymbol{W})^{-1}\boldsymbol{D}(\boldsymbol{S}^{-1}\boldsymbol{W})$$

$\boldsymbol{B}$ 与对角矩阵相似，即 $\boldsymbol{B}$ 是可对角化的。

2.25 假定 $\boldsymbol{B}$与$\boldsymbol{A}$ 相似，证明：

(1) $\boldsymbol{B}+\alpha\boldsymbol{I}$ 与 $\boldsymbol{A}+\alpha\boldsymbol{I}$ 相似。

(2) $\boldsymbol{B}^{\mathrm{T}}$ 与 $\boldsymbol{A}^{\mathrm{T}}$ 相似。

(3) 若 $\boldsymbol{A},\boldsymbol{B}$ 非奇异，则 $\boldsymbol{B}^{-1}$ 与 $\boldsymbol{A}^{-1}$ 相似。

证明： 因为 $\boldsymbol{B}$ 与 $\boldsymbol{A}$ 相似，所以 $\boldsymbol{B}=\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$。

(1) 因为

$$\boldsymbol{B}+\alpha\boldsymbol{I}=\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}+\alpha\boldsymbol{S}^{-1}\boldsymbol{S}=\boldsymbol{S}^{-1}(\boldsymbol{A}+\alpha\boldsymbol{I})\boldsymbol{S}$$

所以，$\boldsymbol{B}+\alpha\boldsymbol{I}$ 与 $\boldsymbol{A}+\alpha\boldsymbol{I}$ 相似。

(2) 因为

$$\boldsymbol{B}^{\mathrm{T}}=(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S})^{\mathrm{T}}=\boldsymbol{S}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{S}^{-\mathrm{T}}=(\boldsymbol{S}^{-\mathrm{T}})^{-1}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{S}^{-\mathrm{T}})$$

所以，$\boldsymbol{B}^{\mathrm{T}}$ 与 $\boldsymbol{A}^{\mathrm{T}}$ 相似。

(3) 因为

$$\boldsymbol{B}^{-1}=(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S})^{-1}=\boldsymbol{S}^{-1}\boldsymbol{A}^{-1}\boldsymbol{S}$$

所以，$\boldsymbol{B}^{-1}$ 与 $\boldsymbol{A}^{-1}$ 相似。

2.26 假定 $\boldsymbol{A},\boldsymbol{B}$ 为 $n\times n$ 矩阵，并且 $\boldsymbol{B}$ 非奇异，证明 $\boldsymbol{A}\boldsymbol{B}$ 与 $\boldsymbol{B}\boldsymbol{A}$ 相似。

证明： 因为 $\boldsymbol{A}\boldsymbol{B}=\boldsymbol{B}^{-1}\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}=\boldsymbol{B}^{-1}(\boldsymbol{B}\boldsymbol{A})\boldsymbol{B}$，所以 $\boldsymbol{A}\boldsymbol{B}$ 与 $\boldsymbol{B}\boldsymbol{A}$ 相似。

2.27 证明：若$n\times n$矩阵$\boldsymbol{A}$与$n\times n$单位矩阵$\boldsymbol{I}$ 相似，则$\boldsymbol{A}=\boldsymbol{I}$。

证明： 因为 $\boldsymbol{A}$ 与 $\boldsymbol{I}$ 相似，所以 $\boldsymbol{A}=\boldsymbol{S}^{-1}\boldsymbol{I}\boldsymbol{S}=\boldsymbol{S}^{-1}\boldsymbol{S}=\boldsymbol{I}$。

2.28 令$\boldsymbol{A}$ 是一个$n\times n$实矩阵，证明$\boldsymbol{B}=(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})/2$为对称矩阵，而$\boldsymbol{C}=(\boldsymbol{A}-\boldsymbol{A}^{\mathrm{T}})/2$为反对称矩阵。

证明： 因为

$$\boldsymbol{B}^{\mathrm{T}}=\left[\frac{1}{2}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})\right]^{\mathrm{T}}=\frac{1}{2}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})=\boldsymbol{B}$$

所以，$\boldsymbol{B}$ 为对称矩阵。又因为

$$\boldsymbol{C}^{\mathrm{T}}=\left[\frac{1}{2}(\boldsymbol{A}-\boldsymbol{A}^{\mathrm{T}})\right]^{\mathrm{T}}=\frac{1}{2}(\boldsymbol{A}^{\mathrm{T}}-\boldsymbol{A})=-\boldsymbol{C}$$

所以，$\boldsymbol{C}$ 为反对称矩阵。

2.29 给定 $n+1$ 个不同的数 $x_0,x_1,\cdots,x_n$ 和任意 $n+1$ 个数的集合 $\{y_0,y_1,\cdots,y_n\}$，则存在一个唯一的多项式 $p(x)=a_0+a_1x+a_2x^2+\cdots+a_nx^n$，使得 $p(x_0)=y_0,p(x_1)=y_1,\cdots,p(x_n)=y_n$。求 $a_0,a_1,\cdots,a_n$ 的表达式。

解： 由题意可得如下方程组：

$$\begin{aligned} a_0+a_1x_0+a_2x_0^2+\cdots+a_nx_0^n&=y_0\\ a_0+a_1x_1+a_2x_1^2+\cdots+a_nx_1^n&=y_1\\ &\vdots\\ a_0+a_1x_n+a_2x_n^2+\cdots+a_nx_n^n&=y_n \end{aligned}$$

写成矩阵形式为

$$\boldsymbol{X}\boldsymbol{a}=\boldsymbol{y}$$

其中

$$\boldsymbol{X}=\begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n\\ 1 & x_1 & x_1^2 & \cdots & x_1^n\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{bmatrix}$$
$$\boldsymbol{a}=[a_0,a_1,\cdots,a_n]^{\mathrm{T}},\quad \boldsymbol{y}=[y_0,y_1,\cdots,y_n]^{\mathrm{T}}$$

由于 Vandermonde 矩阵 $\boldsymbol{X}$ 中 $x_0,x_1,\cdots,x_n$ 各不相同，且 $p(\boldsymbol{x})$ 唯一，故以上方程组有唯一解，即 Vandermonde 矩阵 $\boldsymbol{X}$ 满秩。所以，$\boldsymbol{a}=\boldsymbol{X}^{-1}\boldsymbol{y}$。

2.30 给定 $n+1$个 不同的数$\beta_0,\beta_1,\cdots,\beta_n$ 和任意 $n+1$ 个数的集合 $\{y_0,y_1,\cdots,y_n\}$，证明：存在唯一的一个多项式 $y(x)=a_0\mathrm{e}^{\beta_0x}+a_1\mathrm{e}^{\beta_1x}+\cdots+a_n\mathrm{e}^{\beta x}$ 满足约束条件 $y(0)=y_0,y'(0)=y_1,\cdots,y^{(n)}(0)=y_n$，其中，$y^{(k)}$ 表示 $\dfrac{\mathrm{d}^ky(x)}{\mathrm{d}x^k}$。

证明： 因为

$$y^{(k)}=a_0\beta_0^k\mathrm{e}^{\beta_0x}+a_1\beta_1^k\mathrm{e}^{\beta_1x}+\cdots+a_n\beta_n^k\mathrm{e}^{\beta_nx}$$

所以，由约束条件可得以下方程组：

$$\begin{aligned} a_0+a_1+\cdots+a_n&=y_0\\ a_0\beta_0+a_1\beta_1+\cdots+a_n\beta_n&=y_1\\ &\vdots\\ a_0\beta_0^n+a_1\beta_1^n+\cdots+a_n\beta_n^n&=y_n \end{aligned}$$

写成矩阵形式为

$$\boldsymbol{B}\boldsymbol{a}=\boldsymbol{y}$$

其中

$$
\boldsymbol{B}=\begin{bmatrix}1 & 1 & \cdots & 1\\ \beta_0 & \beta_1 & \cdots & \beta_n\\ \vdots & \vdots & \cdots & \vdots\\ \beta_0^n & \beta_1^n & \cdots & \beta_n^n\end{bmatrix}
$$

$$
\boldsymbol{a}=[a_0,a_1,\cdots,a_n]^{\mathrm{T}},\quad \boldsymbol{y}=[y_0,y_1,\cdots,y_n]^{\mathrm{T}}
$$

由于 Vandermonde 矩阵 $\boldsymbol{B}$ 中 $\beta_0,\cdots,\beta_n$ 各不相同，所以 $\mathrm{rank}(\boldsymbol{B})=n+1$，即 $\boldsymbol{B}$ 满秩。因此，矩阵方程 $\boldsymbol{Ba}=\boldsymbol{y}$ 有唯一解，即多项式 $y(x)$ 存在且唯一。

2.31 第 i 行和第 j 列元素为 $1/(i+j-1)$ 的 $n\times n$ 矩阵称为 Hilbert 矩阵。令 $\boldsymbol{A}$ 是一个 6×6 维 Hilbert 矩阵，并且

$$
\boldsymbol{b}=[1,2,1,1.414,1,2]^{\mathrm{T}},\qquad \boldsymbol{b}+\Delta\boldsymbol{b}=[1,2,1,1.4142,1,2]^{\mathrm{T}}
$$

试用 MATLAB 求解矩阵方程 $\boldsymbol{Ax}_1=\boldsymbol{b}$ 和 $\boldsymbol{Ax}_2=\boldsymbol{b}+\Delta\boldsymbol{b}$，并比较 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$。为什么尽管向量 $\boldsymbol{A}$ 的扰动很小，$\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 却相差很大？

解： 用 MATLAB 分别求解两个方程组，得到

$$
\boldsymbol{x}_1=10^6\times[-0.0065,0.1857,-1.2562,3.2714,-3.6163,1.4271]^{\mathrm{T}}
$$
$$
\boldsymbol{x}_2=10^6\times[-0.0065,0.1857,-1.2562,3.2721,-3.6171,1.4274]^{\mathrm{T}}
$$

此外，$\Delta\boldsymbol{x}=\boldsymbol{x}_1-\boldsymbol{x}_2=[1.5120,-42.3360,282.2400,-725.7600,793.8000,-310.4640]^{\mathrm{T}}$。由此可见，对于 $\boldsymbol{A}$ 的微小扰动，$\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 相差很大，其解释如下：

对于矩阵方程 $\boldsymbol{Ax}=\boldsymbol{b}$ 和 $\boldsymbol{A}(\boldsymbol{x}+\Delta\boldsymbol{x})=\boldsymbol{b}+\Delta\boldsymbol{b}$，有

$$
\Delta\boldsymbol{x}=\boldsymbol{A}^{-1}\Delta\boldsymbol{b}
$$

由矩阵范数的定义有

$$
\|\Delta\boldsymbol{x}\|\leqslant\|\boldsymbol{A}^{-1}\|\cdot\|\Delta\boldsymbol{b}\|,\quad \|\boldsymbol{b}\|\leqslant\|\boldsymbol{A}\|\cdot\|\boldsymbol{x}\|
$$

因此，

$$
\frac{\|\Delta\boldsymbol{x}\|}{\boldsymbol{x}}\leqslant(\|\boldsymbol{A}^{-1}\|\cdot\|\boldsymbol{A}\|)\frac{\|\Delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}
$$

我们把 $\|\boldsymbol{A}^{-1}\|\cdot\|\boldsymbol{A}\|$ 称为矩阵 $\boldsymbol{A}$ 的条件数，记为 $\mathrm{cond}(\boldsymbol{A})$，即上式可写为

$$
\frac{\|\Delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|}\leqslant\mathrm{cond}(\boldsymbol{A})\frac{\|\Delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}
$$

可以用 MATLAB 计算得到矩阵 $\boldsymbol{A}$ 的条件数为 1.4951×10^7，因此，$\boldsymbol{b}$ 相对微量的抖动就会造成解的相对比较大的变化。$\boldsymbol{b}$ 在实际中一般是观测值，所以对于条件数大的状态方程，在观测误差的扰动下，求解的状态值会有很大的误差。

2.32 矩阵 $\boldsymbol{A}=[a_{ij}], i,j=1,2,3,4$ 称为Lorentz矩阵，若变换 $\boldsymbol{x}=\boldsymbol{A}\boldsymbol{y}$ 使得二次型 $Q(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{x}$ 不变，即 $Q(\boldsymbol{x})=Q(\boldsymbol{y})$。证明两个Lorentz矩阵的乘积仍然为Lorentz矩阵。

证明： 根据题意知，$\boldsymbol{A}$ 是 Lorentz 矩阵，若 $Q(\boldsymbol{x})=Q(\boldsymbol{A}\boldsymbol{x})$，即

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{x}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{A}\boldsymbol{x} \quad\Longrightarrow\quad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{A}=\boldsymbol{\Lambda}$$

另，假设 $\boldsymbol{B}$ 也是 Lorentz 矩阵，即 $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{B}=\boldsymbol{\Lambda}$，所以

$$(\boldsymbol{A}\boldsymbol{B})^{\mathrm{T}}\boldsymbol{\Lambda}(\boldsymbol{A}\boldsymbol{B})=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{A}\boldsymbol{B}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{B}=\boldsymbol{\Lambda}$$

即两个 Lorentz 矩阵的乘积仍然是 Lorentz 矩阵。

2.33 一个 $n\times n$ 矩阵 $\boldsymbol{M}$ 称为 Markov 矩阵，若其元素满足条件 $m_{ij}\geqslant 0, \sum\limits_{i=1}^{n} m_{ij}=1, j=1,2,\cdots,n$。假定 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$ 均为 Markov 矩阵，证明：

(1) 对于常数 $0\leqslant\lambda\leqslant 1$，矩阵 $\lambda\boldsymbol{P}+(1-\lambda)\boldsymbol{Q}$ 是Markov矩阵。

(2) 矩阵乘积 $\boldsymbol{PQ}$ 也为 Markov 矩阵。

证明： (1) 设 $\boldsymbol{P}=[p_{ij}], \boldsymbol{Q}=[q_{ij}]$，则

$$\boldsymbol{M}=\lambda\boldsymbol{P}+(1-\lambda)\boldsymbol{Q} \quad\Longrightarrow\quad m_{ij}=\lambda p_{ij}+(1-\lambda)q_{ij}\geqslant 0$$

且

$$\sum_{i=1}^{n} m_{ij}=\lambda\sum_{i=1}^{n} p_{ij}+(1-\lambda)\sum_{i=1}^{n} q_{ij}=\lambda+(1-\lambda)=1$$

故，$\boldsymbol{M}=\lambda\boldsymbol{P}+(1-\lambda)\boldsymbol{Q}$ 是 Markov 矩阵。

(2) 令 $\boldsymbol{M}=\boldsymbol{PQ}$，则 $m_{ij}=\sum\limits_{k=1}^{n} p_{ik}q_{kj}$。由于$p_{ij}\geqslant 0, q_{ij}\geqslant 0, \forall i,j\in\{1,\cdots,n\}$，且

$$\sum_{i=1}^{n} m_{ij}=\sum_{i=1}^{n}\sum_{k=1}^{n} p_{ik}q_{kj}=\sum_{k=1}^{n}\sum_{i=1}^{n} p_{ik}q_{kj}=\sum_{k=1}^{n} q_{kj}=1$$

故，$\boldsymbol{PQ}$ 也是 Markov 矩阵。

2.34 一个 $n\times 1$ 向量 $\boldsymbol{x}$ 称为概率向量 (probability vector)，若其元素满足与概率公式类似的条件 $x_i\geqslant 0$ 和 $\sum\limits_{i=1}^{n} x_i=1$。证明：若 $\boldsymbol{x}$ 为概率向量，则矩阵 $\boldsymbol{M}$ 是 Markov 矩阵的必要条件是 $\boldsymbol{Mx}$ 为概率向量。

证明： 因为 $\boldsymbol{M}$ 是 Markov 矩阵，即 $\sum\limits_{i=1}^{n} m_{ij}=1$，所以

$$\sum_{i=1}^{n} y_i=\sum_{j=1}^{n}\sum_{i=1}^{n} m_{ij}x_j=\sum_{j=1}^{n} x_j=1$$

且 $y_i\geqslant 0, \forall i\in\{1,\cdots,n\}$，即 $\boldsymbol{y}=\boldsymbol{Mx}$ 是概率向量。

2.35 令 $y_i = y_i(x_1, x_2, \cdots, x_n), i = 1, 2, \cdots, n$ 是关于 $x_1, x_2, \cdots, x_n$ 的 n 个函数。矩阵 $\boldsymbol{J} = \boldsymbol{J}(\boldsymbol{y}, \boldsymbol{x}) = [\partial y_i / \partial x_j]$ 称为函数 $y_i(x_1, x_2, \cdots, x_n), i = 1, 2, \cdots, n$ 的 Jacobian 矩阵，其行列式称为 Jacobian 行列式。式中，$\boldsymbol{y} = [y_1, y_2, \cdots, y_n]^{\mathrm{T}}$，$\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^{\mathrm{T}}$。证明 $\boldsymbol{J}(\boldsymbol{z}, \boldsymbol{y})\boldsymbol{J}(\boldsymbol{y}, \boldsymbol{x}) = \boldsymbol{J}(\boldsymbol{z}, \boldsymbol{x})$。

证明： 因为

$$\boldsymbol{J}(\boldsymbol{z}, \boldsymbol{y}) = \left[\frac{\partial z_i}{\partial y_j}\right]_{ij}, \quad \boldsymbol{J}(\boldsymbol{y}, \boldsymbol{x}) = \left[\frac{\partial y_i}{\partial x_j}\right]_{ij}$$

所以，由复合函数求导知

$$\boldsymbol{J}(\boldsymbol{z}, \boldsymbol{y})\boldsymbol{J}(\boldsymbol{y}, \boldsymbol{x}) = \left[\sum_{k=1}^{n} \frac{\partial z_i}{\partial y_k}\frac{\partial y_k}{\partial x_j}\right]_{ij} = \left[\frac{\partial z_i}{\partial x_j}\right]_{ij} = \boldsymbol{J}(\boldsymbol{z}, \boldsymbol{x})$$

命题得证。

2.36 令 $p \times p$ 矩阵 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 分别为对称矩阵，并且 $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}$。证明：Jacobian 行列式 $|\boldsymbol{J}(\boldsymbol{Y}, \boldsymbol{X})|$ 等于矩阵 $\boldsymbol{A}$ 的行列式的 $p+1$ 次方，即 $|\boldsymbol{J}(\boldsymbol{Y}, \boldsymbol{X})| = |\boldsymbol{A}|^{p+1}$。

证明： 首先，按照教材定义的 II 型初等行(列)变换和 III 型初等行(列)变换，有以下性质：

(1) II 型初等行(列)变换：将矩阵的第 i 行(列)乘以一个常数 $a \neq 0$，该变换相当于将原矩阵左乘（右乘）一个对角矩阵 $\boldsymbol{D}$，其中 $\boldsymbol{D}$ 的第 i 行第 i 列对角元素为 a，其余为 1 。因此，$|\boldsymbol{D}| = a$，且 $\boldsymbol{D}^{-1}$ 具有与 $\boldsymbol{D}$ 相同的形式。

(2) III 型初等行(列)变换：将矩阵的第 j 行 (或第 i 列) 乘以 a $(a \neq 1)$ 并加到第 i 行 (或第 j 列) 上，这相当于左乘（右乘）一个矩阵 $\boldsymbol{E}$，其中，$\boldsymbol{E}$ 的主对角线元素为 1，第 i 行第 j 列为 a ，其余元素为 1。因此，$|\boldsymbol{E}| = 1$，且 $\boldsymbol{E}^{-1}$ 具有与 $\boldsymbol{E}$ 相同的形式。

对于一个矩阵 $\boldsymbol{A}$ 的初等变换等价于 $\boldsymbol{E}_1 \cdots \boldsymbol{E}_n \boldsymbol{A} \boldsymbol{E}_{n+1} \cdots \boldsymbol{E}_m$，其中，$\boldsymbol{E}_i$ 或者是 II 型初等行(列)变换，或者是 III 型初等行(列)变换。因此，若 $\boldsymbol{A}$ 是正方矩阵且非奇异，则有

$$\boldsymbol{A} = \boldsymbol{E}_n^{-1} \cdots \boldsymbol{E}_1^{-1} \boldsymbol{E}_m^{-1} \cdots \boldsymbol{E}_{n+1}^{-1} \stackrel{\text{def}}{=} \boldsymbol{F}_m \boldsymbol{F}_{m-1} \cdots \boldsymbol{F}_1$$

其中，$\boldsymbol{F}_i$ 或者是 II 型初等行(列)变换，或者是 III 型初等行(列)变换，于是

$$\boldsymbol{Y} = \boldsymbol{F}_m \cdots \boldsymbol{F}_1 \boldsymbol{X} \boldsymbol{F}_1^{\mathrm{T}} \cdots \boldsymbol{F}_m^{\mathrm{T}}$$

令 Jacobian 行列式 $D(\boldsymbol{X}; \boldsymbol{Y}) = |J(\boldsymbol{Y}; \boldsymbol{X})|$，下面的引理给出了 Jacobian 行列式的分解计算方法。

引理 2.1 令 $\boldsymbol{Y} = \boldsymbol{A}_n \boldsymbol{A}_{n-1} \cdots \boldsymbol{A}_1 \boldsymbol{X} \boldsymbol{A}_1^{\mathrm{T}} \cdots \boldsymbol{A}_{n-1}^{\mathrm{T}} \boldsymbol{A}_n^{\mathrm{T}}$，则

$$D(\boldsymbol{X}; \boldsymbol{Y}) = D(\boldsymbol{X}; \boldsymbol{Y}_1) D(\boldsymbol{Y}_1; \boldsymbol{Y}_2) \cdots D(\boldsymbol{Y}_{n-1}; \boldsymbol{Y})$$

其中，$\boldsymbol{Y}_i = \boldsymbol{A}_i \boldsymbol{Y}_{i-1} \boldsymbol{A}_i^{\mathrm{T}}$ $(i = 1, \cdots, n)$；$\boldsymbol{Y}_0 \stackrel{\text{def}}{=} \boldsymbol{X}, \boldsymbol{Y}_n \stackrel{\text{def}}{=} \boldsymbol{Y}$。

这个引理不难从 Jacobian 矩阵的性质：$D(\boldsymbol{X};\boldsymbol{Y}) = D(\boldsymbol{X};\boldsymbol{U})D(\boldsymbol{U};\boldsymbol{V})D(\boldsymbol{V};\boldsymbol{Y})$ 推广得到。由引理 2.1 知，

$$D(\boldsymbol{X};\boldsymbol{Y}) = D(\boldsymbol{Y}_0;\boldsymbol{Y}_1)D(\boldsymbol{Y}_1;\boldsymbol{Y}_2)\cdots D(\boldsymbol{Y}_{m-1};\boldsymbol{Y}_m)$$

其中 $\boldsymbol{Y}_i = \boldsymbol{F}_i\boldsymbol{Y}_{i-1}\boldsymbol{F}_i^{\mathrm{T}}\ (i=1,\cdots,m)$，$\boldsymbol{Y}_0 \stackrel{\mathrm{def}}{=} \boldsymbol{X}, \boldsymbol{Y}_m \stackrel{\mathrm{def}}{=} \boldsymbol{Y}$。

下面证明 $D(\boldsymbol{Y}_{i-1};\boldsymbol{Y}_i) = |\boldsymbol{F}_i|^{p+1}, (i=1,\cdots,m)$：

(1) 令 $\boldsymbol{G}$ 表示 $\boldsymbol{F}_i\ (i=1,\cdots,m)$ 中任意一个 II 型初等行(列)变换，则变换 $\boldsymbol{Y}_i = \boldsymbol{G}\boldsymbol{Y}_{i-1}\boldsymbol{G}^{\mathrm{T}}$ 意味着

$$y_{ii} = a^2x_{ii},\quad y_{ij} = ax_{ij}\ (i\neq j),\quad y_{jk} = x_{jk}\ (j,k\neq i)$$

因此，Jacobian 矩阵的对角元素有 $p-1$ 个元素为 a，1 个元素为 a^2，其余元素均为 1，故

$$D(\boldsymbol{Y}_{i-1};\boldsymbol{Y}_i) = a^{p+1} = |\boldsymbol{G}|^{p+1}$$

(2) 令 $\boldsymbol{H}$ 表示 $\boldsymbol{F}_i\ (i=1,\cdots,n)$ 中任意一个 III 型初等行(列)变换，则变换 $\boldsymbol{Y}_i = \boldsymbol{H}\boldsymbol{Y}_{i-1}\boldsymbol{H}^{\mathrm{T}}$ 意味着

$$\begin{aligned} y_{ii} &= x_{ii} + 2ax_{ij} + a^2x_{jj} \\ y_{ki} &= y_{ik} = x_{ik} + ax_{jk}\quad (k\neq i) \\ y_{jk} &= x_{jk}\quad (j,k\neq i) \end{aligned}$$

其 Jacobian 矩阵的主对角线元素为 1，主对角线下面的元素为 0，故

$$D(\boldsymbol{Y}_{i-1},\boldsymbol{Y}_i) = 1 = |\boldsymbol{H}|^{p+1}$$

由 (1) 和 (2) 知，

$$D(\boldsymbol{Y}_{i-1};\boldsymbol{Y}_i) = |\boldsymbol{F}_i|^{p+1},\quad (i=1,\cdots,m)$$

因此有

$$D(\boldsymbol{X};\boldsymbol{Y}) = |\boldsymbol{F}_m|^{p+1}\cdots|\boldsymbol{F}_1|^{p+1} = |\boldsymbol{F}_m\cdots\boldsymbol{F}_1|^{p+1} = |\boldsymbol{A}|^{p+1}$$

2.37 证明：若 $\boldsymbol{A}$ 为正规矩阵 (即满足 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$)，则 $\boldsymbol{A}-\lambda\boldsymbol{I}$ 为正规矩阵。

证明： 因为

$$\begin{aligned} (\boldsymbol{A}-\lambda\boldsymbol{I})(\boldsymbol{A}-\lambda\boldsymbol{I})^{\mathrm{H}} &= \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} - \lambda^*\boldsymbol{A} - \lambda\boldsymbol{A}^{\mathrm{H}} + |\lambda|^2\boldsymbol{I} \\ &= \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} - \lambda^*\boldsymbol{A} - \lambda\boldsymbol{A}^{\mathrm{H}} + |\lambda|^2\boldsymbol{I} \\ &= (\boldsymbol{A}-\lambda\boldsymbol{I})^{\mathrm{H}}(\boldsymbol{A}-\lambda\boldsymbol{I}) \end{aligned}$$

所以，$\boldsymbol{A}-\lambda\boldsymbol{I}$ 是正规矩阵。

2.38 若 $\boldsymbol{B}$ 为正规矩阵，并且存在一个角度 θ，使得 $\boldsymbol{A}\mathrm{e}^{\mathrm{j}\theta}+\boldsymbol{A}^{\mathrm{H}}\mathrm{e}^{-\mathrm{j}\theta}\geqslant 0$，其中，$\boldsymbol{A}^2=\boldsymbol{B}$。证明 $\boldsymbol{A}$ 是正规矩阵。

证明：令 $\boldsymbol{C}=\mathrm{e}^{\mathrm{j}\theta}\boldsymbol{A}$，它满足 $\boldsymbol{C}+\boldsymbol{C}^{\mathrm{H}}\geqslant 0$。若令 $\boldsymbol{C}=\boldsymbol{H}+\mathrm{j}\boldsymbol{J}$，其中 $\boldsymbol{H}=(\boldsymbol{C}+\boldsymbol{C}^{\mathrm{H}})/2$ 和 $\boldsymbol{J}=-\mathrm{j}(\boldsymbol{C}-\boldsymbol{C}^{\mathrm{H}})/2$ 分别为 $\boldsymbol{C}$ 的自伴随的实部和虚部，则

$$\boldsymbol{C}^2=\mathrm{e}^{2\mathrm{j}\theta}\boldsymbol{A}^2=(\boldsymbol{H}^2-\boldsymbol{J}^2)+\mathrm{j}(\boldsymbol{H}\boldsymbol{J}+\boldsymbol{J}\boldsymbol{H}) \tag{2.35}$$

由于 $\boldsymbol{B}=\boldsymbol{A}^2$ 是正规矩阵，所以 $\boldsymbol{C}^2$ 也是正规矩阵，且显然 $\boldsymbol{C}^2$ 与 $\boldsymbol{C}$ 是可交换的（两个矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是可交换的，若 $\boldsymbol{A}\boldsymbol{B}=\boldsymbol{B}\boldsymbol{A}$）。于是，由题 2.39 知，$(\boldsymbol{C}^2)^{\mathrm{H}}$ 与 $\boldsymbol{C}$ 是可交换的，即

$$(\boldsymbol{C}^2)^{\mathrm{H}}\boldsymbol{C}=\boldsymbol{C}(\boldsymbol{C}^2)^{\mathrm{H}} \quad\Longrightarrow\quad \boldsymbol{C}^{\mathrm{H}}\boldsymbol{C}^2=\boldsymbol{C}^2\boldsymbol{C}^{\mathrm{H}}$$

由此可见，$\boldsymbol{C}^{\mathrm{H}}$ 与 $\boldsymbol{C}^2$ 也是可交换的。所以

$$\begin{cases}\boldsymbol{C}^2(\boldsymbol{H}+\mathrm{j}\boldsymbol{J})=(\boldsymbol{H}+\mathrm{j}\boldsymbol{J})\boldsymbol{C}^2\\ \boldsymbol{C}^2(\boldsymbol{H}-\mathrm{j}\boldsymbol{J})=(\boldsymbol{H}-\mathrm{j}\boldsymbol{J})\boldsymbol{C}^2\end{cases} \quad\Longrightarrow\quad \begin{cases}\boldsymbol{C}^2\boldsymbol{H}=\boldsymbol{H}\boldsymbol{C}^2\\ \boldsymbol{C}^2\boldsymbol{J}=\boldsymbol{J}\boldsymbol{C}^2\end{cases}$$

即 $\boldsymbol{C}^2$ 分别与 $\boldsymbol{H}$ 和 $\boldsymbol{J}$ 都是可交换的。将式 (2.35) 两边左乘 $\boldsymbol{H}$，可得

$$\boldsymbol{H}\boldsymbol{C}^2=(\boldsymbol{H}^3-\boldsymbol{H}\boldsymbol{J}^2)+\mathrm{j}(\boldsymbol{H}^2\boldsymbol{J}+\boldsymbol{H}\boldsymbol{J}\boldsymbol{H}) \tag{2.36}$$

式 (2.36) 两边右乘 $\boldsymbol{H}$，可得

$$\boldsymbol{C}^2\boldsymbol{H}=(\boldsymbol{H}^3-\boldsymbol{J}^2\boldsymbol{H})+\mathrm{j}(\boldsymbol{H}\boldsymbol{J}\boldsymbol{H}+\boldsymbol{J}\boldsymbol{H}^2) \tag{2.37}$$

式 (2.36) 和式 (2.37) 相减，得

$$\boldsymbol{R}+\mathrm{j}\boldsymbol{S}=\boldsymbol{O}$$

其中 $\boldsymbol{R}=\boldsymbol{J}^2\boldsymbol{H}-\boldsymbol{H}\boldsymbol{J}^2, \boldsymbol{S}=\boldsymbol{H}^2\boldsymbol{J}-\boldsymbol{J}\boldsymbol{H}^2$。由自伴随性，可知

$$\begin{aligned}\boldsymbol{R}^{\mathrm{H}}-\mathrm{j}\boldsymbol{S}^{\mathrm{H}}=\boldsymbol{O} \quad &\Rightarrow \quad (\boldsymbol{H}\boldsymbol{J}^2-\boldsymbol{J}^2\boldsymbol{H})-\mathrm{j}(\boldsymbol{J}\boldsymbol{H}^2-\boldsymbol{H}^2\boldsymbol{J})=\boldsymbol{O}\\ &\Rightarrow \quad \boldsymbol{R}-\mathrm{j}\boldsymbol{S}=\boldsymbol{O}\end{aligned}$$

因此，$\boldsymbol{S}=\boldsymbol{O}$，即 $\boldsymbol{H}^2\boldsymbol{J}=\boldsymbol{J}\boldsymbol{H}^2$。又由于

$$\boldsymbol{C}+\boldsymbol{C}^{\mathrm{H}}\geqslant 0 \quad\Longrightarrow\quad \boldsymbol{H}\geqslant 0$$

所以，$\boldsymbol{H}\boldsymbol{J}=\boldsymbol{J}\boldsymbol{H}$。由题 2.40(1) 的结论可知，$\boldsymbol{C}$ 是正规矩阵，即 $\boldsymbol{A}$ 为正规矩阵。

2.39 满足条件 $\boldsymbol{A}\boldsymbol{B}=\boldsymbol{B}\boldsymbol{A}$ 的矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 称为可交换矩阵 (commute matrix)。证明：若 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 可交换，则 $\boldsymbol{A}^{\mathrm{H}}$ 和 $\boldsymbol{B}$ 可交换的条件是 $\boldsymbol{A}$ 为正规矩阵。

证明：由教材定理 2.4.4 知，若 $\boldsymbol{A}$ 是正规矩阵，则必然存在一个酉矩阵 $\boldsymbol{U}$，使得 $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\boldsymbol{\varLambda}$，其中 $\boldsymbol{\varLambda}$ 为对角矩阵，即 $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\varLambda}\boldsymbol{U}^{\mathrm{H}}$。由 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 可交换知，

$$\boldsymbol{A}\boldsymbol{B}=\boldsymbol{B}\boldsymbol{A} \quad\Leftrightarrow\quad \boldsymbol{U}\boldsymbol{\varLambda}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{B}=\boldsymbol{B}\boldsymbol{U}\boldsymbol{\varLambda}\boldsymbol{U}^{\mathrm{H}} \quad\Leftrightarrow\quad \boldsymbol{\varLambda}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{U}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{U}\boldsymbol{\varLambda}$$

记 $\boldsymbol{C}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{U}$，则有 $\boldsymbol{\Lambda}\boldsymbol{C}=\boldsymbol{C}\boldsymbol{\Lambda}$，即对 $\forall i,j,1\leqslant i,j\leqslant n$ 都有

$$\boldsymbol{\Lambda}_{ii}\boldsymbol{C}_{ij}=\boldsymbol{C}_{ij}\boldsymbol{\Lambda}_{jj}\quad\Rightarrow\quad\boldsymbol{\Lambda}_{ii}^{*}\boldsymbol{C}_{ij}=\boldsymbol{C}_{ij}\boldsymbol{\Lambda}_{jj}^{*}\quad\Rightarrow\quad\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{C}=\boldsymbol{C}\boldsymbol{\Lambda}^{\mathrm{H}}$$

所以

$$\begin{aligned}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{U}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{U}\boldsymbol{\Lambda}^{\mathrm{H}}\quad&\Rightarrow\quad\boldsymbol{U}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{B}=\boldsymbol{B}\boldsymbol{U}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}}\\&\Rightarrow\quad\boldsymbol{A}^{\mathrm{H}}\boldsymbol{B}=\boldsymbol{B}\boldsymbol{A}^{\mathrm{H}}\end{aligned}$$

命题得证。

2.40 令 $\boldsymbol{A}$ 为复矩阵，证明 $\boldsymbol{A}$ 为正规矩阵，当且仅当下列条件之一成立：

(1) $\boldsymbol{A}=\boldsymbol{B}+\mathrm{j}\boldsymbol{C}$，其中，$\boldsymbol{B}$ 和 $\boldsymbol{C}$ 为 Hermitian 矩阵，并且可交换。

(2) $\boldsymbol{A}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{U}$，其中，$\boldsymbol{U}$ 为酉矩阵，且 $\boldsymbol{D}$ 为对角矩阵。

(3) $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{H}$，其中，$\boldsymbol{U}$ 为酉矩阵，$\boldsymbol{H}$ 为 Hermitian 矩阵，并且 $\boldsymbol{U}$ 和 $\boldsymbol{H}$ 可交换。

证明：(1) 充分性：因为

$$\begin{aligned}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}&=(\boldsymbol{B}+\mathrm{j}\boldsymbol{C})(\boldsymbol{B}+\mathrm{j}\boldsymbol{C})^{\mathrm{H}}=(\boldsymbol{B}+\mathrm{j}\boldsymbol{C})(\boldsymbol{B}^{\mathrm{H}}-\mathrm{j}\boldsymbol{C}^{\mathrm{H}})\\&=\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}+\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}}+\mathrm{j}\boldsymbol{C}\boldsymbol{B}^{\mathrm{H}}-\mathrm{j}\boldsymbol{B}\boldsymbol{C}^{\mathrm{H}}\\&=\boldsymbol{B}^2+\boldsymbol{C}^2+\mathrm{j}\boldsymbol{C}\boldsymbol{B}-\mathrm{j}\boldsymbol{B}\boldsymbol{C}\\&=\boldsymbol{B}^2-\boldsymbol{C}^2\\\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}&=(\boldsymbol{B}+\mathrm{j}\boldsymbol{C})^{\mathrm{H}}(\boldsymbol{B}+\mathrm{j}\boldsymbol{C})=(\boldsymbol{B}^{\mathrm{H}}-\mathrm{j}\boldsymbol{C}^{\mathrm{H}})(\boldsymbol{B}+\mathrm{j}\boldsymbol{C})\\&=\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}+\boldsymbol{C}^{\mathrm{H}}\boldsymbol{C}-\mathrm{j}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{B}+\mathrm{j}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{C}\\&=\boldsymbol{B}^2+\boldsymbol{C}^2-\mathrm{j}\boldsymbol{C}\boldsymbol{B}+\mathrm{j}\boldsymbol{B}\boldsymbol{C}\\&=\boldsymbol{B}^2+\boldsymbol{C}^2\end{aligned}$$

所以，$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$，即 $\boldsymbol{A}$ 为正规矩阵。

必要性：由 $\boldsymbol{A}$ 是正规矩阵知，$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}$，其中 $\boldsymbol{U}$ 为酉矩阵，$\boldsymbol{\Lambda}$ 为对角矩阵。将 $\boldsymbol{A}$ 另写为

$$\boldsymbol{A}=\boldsymbol{U}\mathrm{real}(\boldsymbol{\Lambda})\boldsymbol{U}^{\mathrm{H}}+\boldsymbol{U}\mathrm{imag}(\boldsymbol{\Lambda})\boldsymbol{U}^{\mathrm{H}}$$

其中，$\mathrm{real}(\boldsymbol{\Lambda})$ 表示 $\boldsymbol{\Lambda}$ 的实部，$\mathrm{imag}(\boldsymbol{\Lambda})$ 表示 $\boldsymbol{\Lambda}$ 的虚部。令 $\boldsymbol{B}\overset{\mathrm{def}}{=}\boldsymbol{U}\mathrm{real}(\boldsymbol{\Lambda})\boldsymbol{U}^{\mathrm{H}}$，$\boldsymbol{C}\overset{\mathrm{def}}{=}\boldsymbol{U}\mathrm{imag}(\boldsymbol{\Lambda})\boldsymbol{U}^{\mathrm{H}}$，易知，$\boldsymbol{B}$ 和 $\boldsymbol{C}$ 都是 Hermitian 矩阵。又

$$\begin{aligned}\boldsymbol{B}\boldsymbol{C}&=\boldsymbol{U}\mathrm{real}(\boldsymbol{\Lambda})\boldsymbol{U}^{\mathrm{H}}\cdot\boldsymbol{U}\mathrm{imag}(\boldsymbol{\Lambda})\boldsymbol{U}^{\mathrm{H}}\\&=\boldsymbol{U}\mathrm{real}(\boldsymbol{\Lambda})\mathrm{imag}(\boldsymbol{\Lambda})\boldsymbol{U}^{\mathrm{H}}\\&=\boldsymbol{U}\mathrm{imag}(\boldsymbol{\Lambda})\mathrm{real}(\boldsymbol{\Lambda})\boldsymbol{U}^{\mathrm{H}}\\&=\boldsymbol{C}\boldsymbol{B}\end{aligned}$$

即 $\boldsymbol{B}$ 和 $\boldsymbol{C}$ 可交换，命题得证。

(2) 由题意，本题即证若 $\boldsymbol{A}$ 是正规矩阵，则它可用一个酉矩阵对角化。

必要性：由 Schur 分解知，对于一个复矩阵，我们总可以找到一个酉变换矩阵 $\boldsymbol{T}$，使 $\boldsymbol{A}$ 变为上三角矩阵，即

$$\boldsymbol{A}=\boldsymbol{T}\begin{bmatrix}\lambda_1 & b_{12} & \cdots & b_{1N}\\ & \lambda_2 & \cdots & b_{2N}\\ & & \ddots & \vdots\\ 0 & & & \lambda_N\end{bmatrix}\boldsymbol{T}^{-1}=\boldsymbol{T}\boldsymbol{D}\boldsymbol{T}^{-1}$$

由于 $\boldsymbol{T}$ 是酉矩阵，故

$$\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{T}\begin{bmatrix}\lambda_1^* & & & 0\\ b_{12}^* & \lambda_2^* & & \\ \vdots & \vdots & \ddots & \\ b_{1N}^* & b_{2N}^* & \cdots & \lambda_N^*\end{bmatrix}\boldsymbol{T}^{-1}=\boldsymbol{T}\boldsymbol{D}^{\mathrm{H}}\boldsymbol{T}^{-1}$$

因此，

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{T}\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}}\boldsymbol{T}^{-1}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{T}\boldsymbol{D}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{T}^{-1}$$

于是有 $\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}}=\boldsymbol{D}^{\mathrm{H}}\boldsymbol{D}$。由左右两边矩阵乘积的各元素相等，可得 $b_{ij}=0,1\leqslant i<j\leqslant N$，即上三角矩阵 $\boldsymbol{D}$ 实际上是一个对角矩阵。令 $\boldsymbol{U}=\boldsymbol{T}^{\mathrm{H}}$，则 $\boldsymbol{A}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{U}$，故必要性得证。

充分性：因为

$$\begin{aligned}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}&=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{U}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{D}^{\mathrm{H}}\boldsymbol{U}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}}\boldsymbol{U}\\&=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{D}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{U}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{D}^{\mathrm{H}}\boldsymbol{U}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{U}\\&=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\end{aligned}$$

所以，$\boldsymbol{A}$ 为正规矩阵。由此，命题得证。

(3) 充分性：因为

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=(\boldsymbol{U}\boldsymbol{H})(\boldsymbol{U}\boldsymbol{H})^{\mathrm{H}}=(\boldsymbol{H}\boldsymbol{U})(\boldsymbol{H}\boldsymbol{U})^{\mathrm{H}}=\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}=\boldsymbol{H}^2$$
$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{H}^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}\boldsymbol{H}=\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}=\boldsymbol{H}^2$$

所以，$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$，即 $\boldsymbol{A}$ 是正规矩阵。

必要性：本题必要性的证明等价为 $\boldsymbol{A}$ 是正规矩阵时，求一个酉矩阵 $\boldsymbol{U}$ ，使得 $\boldsymbol{H}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}$ 是 Hermitian 矩阵。记 $\boldsymbol{A}=\boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{\mathrm{H}}$，它满足正规矩阵的条件 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$，其中 $\boldsymbol{V}$ 是酉矩阵，$\boldsymbol{D}$ 是对角矩阵。矩阵 $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}$ 为 Hermitian 矩阵意味着

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{\mathrm{H}}=(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{\mathrm{H}})^{\mathrm{H}}$$

即

$$\boldsymbol{W}\boldsymbol{D}=(\boldsymbol{W}\boldsymbol{D})^{\mathrm{H}}$$

其中，矩阵 $\boldsymbol{W}=(\boldsymbol{UV})^{\mathrm{H}}\boldsymbol{V}$。显然，矩阵 $\boldsymbol{W}$ 也是一个酉矩阵。记 $\boldsymbol{D}=\mathrm{diag}([d_1,d_2,\cdots,d_n])$，则酉矩阵 $\boldsymbol{W}$ 可选择为

$$\boldsymbol{W}=\begin{bmatrix}\frac{d_1^*}{|d_1|} & & & \\ & \frac{d_2^*}{|d_2|} & & \\ & & \ddots & \\ & & & \frac{d_n^*}{|d_n|}\end{bmatrix}$$

由此可解得矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{H}$ 分别为

$$\boldsymbol{U}=\boldsymbol{V}(\boldsymbol{VW})^{\mathrm{H}},\qquad \boldsymbol{H}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}$$

现在我们来验证这样求得的矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{H}$ 是可交换的。记酉矩阵 $\boldsymbol{V}=[\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n]$，于是有

$$\boldsymbol{A}=\sum_{i=1}^{n}d_i\boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{H}}$$

$$\boldsymbol{V}=\sum_{i=1}^{n}=\frac{d_i}{|d_i|}\boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{H}}$$

$$\boldsymbol{H}=\sum_{i=1}^{n}|d_i|\boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{H}}$$

此时可计算得到

$$\boldsymbol{UH}=\boldsymbol{HU}=\sum_{i=1}^{n}d_i\boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{H}}=\boldsymbol{A}$$

即矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{H}$ 可交换，从而必要性得证。

2.41 令 $u_1,u_2,\cdots,u_n$ 是 n 阶线性微分方程

$$u^{(n)}+p_1(t)u^{(n-1)}+\cdots+p_n(t)u=0$$

的线性无关解。矩阵

$$\boldsymbol{W}=\begin{bmatrix}u_1 & u_2 & \cdots & u_n\\ u_1' & u_2' & \cdots & u_n'\\ \vdots & \vdots & & \vdots\\ u_1^{(n-1)} & u_2^{(n-1)} & \cdots & u_n^{(n-1)}\end{bmatrix}$$

称为 Wronskian 矩阵 [19]，其行列式 $W(u_1,u_2,\cdots,u_n)=\det(\boldsymbol{W})$ 称为 $u_1,u_2,\cdots,u_n$ 的 Wronskian 函数。证明：

$$W(t)=W(u_1,u_2,\cdots,u_n)=W(t_0)\exp\left(-\int_{t_0}^{t}p_1(s)\mathrm{d}s\right)$$

思路分析： 本题证明的关键是将微分方程转化为矩阵方程，而把 Wronskian 矩阵作为矩阵方程的解。为了证明本题，首先给出行列式求导和秩 1 更新的定理 [47]：

定理 2.5 (行列式的导数) 若 $\boldsymbol{A}_{n\times n}=[a_{ij}(t)]$ 是 t 的可微函数，则

$$\frac{\mathrm{d}\,(\det(\boldsymbol{A}))}{\mathrm{d}t}=\det(\boldsymbol{D}_1)+\det(\boldsymbol{D}_2)+\cdots+\det(\boldsymbol{D}_n)$$

其中，$\boldsymbol{D}_i$ 除第 i 行为相应的导数外，其余与 $\boldsymbol{A}$ 相同，即

$$[\boldsymbol{D}_i]_{k*}=\begin{cases}\boldsymbol{A}_{k*}, & 若 i\neq k\\ d\boldsymbol{A}_{k*}/\mathrm{d}t, & 若 i=k\end{cases}$$

定理 2.6 (行列式的秩 1 更新) 若 $\boldsymbol{A}_{n\times n}$ 是非奇异的，且若 $\boldsymbol{c}$ 和 $\boldsymbol{d}$ 都是 $n\times 1$ 维向量，则

$$\det(\boldsymbol{I}+\boldsymbol{c}\boldsymbol{d}^{\mathrm{T}})=1+\boldsymbol{d}^{\mathrm{T}}\boldsymbol{c}$$
$$\det(\boldsymbol{A}+\boldsymbol{c}\boldsymbol{d}^{\mathrm{T}})=\det(\boldsymbol{A})(1+\boldsymbol{d}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{c})$$

证明 [47]：将 n 阶线性微分方程 $u^{(n)}+p_1(t)u^{(n-1)}+\cdots+p_n(t)u=0$ 写成矩阵形式：

$$\begin{bmatrix}x_1'\\x_2'\\\vdots\\x_{n-1}'\\x_n'\end{bmatrix}=\begin{bmatrix}0&1&0&\cdots&0\\0&0&1&\cdots&0\\\vdots&\vdots&\vdots&\ddots&\vdots\\0&0&0&\cdots&1\\-p_n&-p_{n-1}&-p_{n-2}&\cdots&-p_1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_{n-1}\\x_n\end{bmatrix}$$

即

$$\boldsymbol{x}'=\boldsymbol{A}\boldsymbol{x}$$

其中

$$\boldsymbol{A}=\begin{bmatrix}0&1&0&\cdots&0\\0&0&1&\cdots&0\\\vdots&\vdots&\vdots&\ddots&\vdots\\0&0&0&\cdots&1\\-p_n&-p_{n-1}&-p_{n-2}&\cdots&-p_1\end{bmatrix},$$
$$\boldsymbol{x}=[x_1,x_2,\cdots,x_{n-1},x_n]^{\mathrm{T}}$$

而 $\boldsymbol{W}_i=[u_i,u_i',\cdots,u_i^{(n-1)}]^{\mathrm{T}}\ (i=1,\cdots,n)$ 是方程 $\boldsymbol{x}'=\boldsymbol{A}\boldsymbol{x}$ 的 n 个线性无关解。由此，Wronskian 矩阵可写为 $\boldsymbol{W}=[\boldsymbol{W}_1,\cdots,\boldsymbol{W}_n]$，于是有

$$\boldsymbol{W}'=\boldsymbol{A}\boldsymbol{W}$$

由于 Wronskian 函数 $W(u_1,u_2,\cdots,u_n)=\det(\boldsymbol{W})$ 是时间 t 的函数，所以利用行列式求导的定理求其关于时间的导数，得

$$\frac{\mathrm{d}W(t)}{\mathrm{d}t}=\frac{\mathrm{d}\,(\det(\boldsymbol{W}))}{\mathrm{d}t}=\sum_{i=1}^{n}\det(\boldsymbol{D}_i)$$

其中

$$\boldsymbol{D}_i = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ \vdots & \vdots & & \vdots \\ w'_{i1} & w'_{i2} & \cdots & w'_{in} \\ \vdots & \vdots & & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix} = \boldsymbol{W} + \boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{W}' - \boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{W}$$

其中 w_{ij} 为矩阵 $\boldsymbol{W}$ 第 i 行第 j 列的元素，$\boldsymbol{e}_i$ 为基本向量，其第 i 个元素为 1 ，其余为 0。将 $\boldsymbol{W}' = \boldsymbol{A}\boldsymbol{W}$ 代入上式，得

$$\begin{aligned} \boldsymbol{D}_i &= \boldsymbol{W} + \boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{W}' - \boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{W} \\ &= \boldsymbol{W} + \boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{W} - \boldsymbol{e}_i\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{W} \\ &= \left[\boldsymbol{I} + \boldsymbol{e}_i\left(\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{A} - \boldsymbol{e}_i\right)\right]\boldsymbol{W} \end{aligned}$$

利用行列式的秩 1 更新定理，得到

$$\det(\boldsymbol{D}_i) = \left(1 + \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{e}_i - \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{e}_i\right)\det(\boldsymbol{W}) = a_{ii}(t)\boldsymbol{W}(t)$$

故

$$\begin{aligned} \frac{\mathrm{d}W(t)}{\mathrm{d}t} &= \sum_{i=1}^{n}\det(\boldsymbol{D}_i) = \left(\sum_{i=1}^{n}a_{ii}(t)\right)\boldsymbol{W}(t) \\ &= \mathrm{tr}(\boldsymbol{A}(t))\boldsymbol{W}(t) \end{aligned}$$

换句话说，$\boldsymbol{W}(t)$满足一阶微分方程 $\boldsymbol{W}' = \tau\boldsymbol{W}$，其中 $\tau = \mathrm{tr}(\boldsymbol{A}(t))$，故其解为

$$\begin{aligned} W(t) &= W(t_0)\exp\left(\int_{t_0}^{t}\tau(s)\mathrm{d}s\right) \\ &= W(t_0)\exp\left(\int_{t_0}^{t}\mathrm{tr}(\boldsymbol{A}(s))\mathrm{d}s\right) \end{aligned}$$

因为 $\mathrm{tr}(\boldsymbol{A}(t)) = -p_1(t)$，所以

$$W(t) = W(t_0)\exp\left(-\int_{t_0}^{t}p_1(s)\mathrm{d}s\right)$$

命题得证。

2.42 证明：若$\boldsymbol{A}$是反对称矩阵，则 $\langle\boldsymbol{A}\boldsymbol{x},\boldsymbol{x}\rangle = 0$ 对任意向量$\boldsymbol{x} \in R^n$成立。

证明： 因为 $\boldsymbol{A} = -\boldsymbol{A}^{\mathrm{T}}$，所以

$$\langle\boldsymbol{A}\boldsymbol{x},\boldsymbol{x}\rangle = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} = -\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = -\langle\boldsymbol{x},\boldsymbol{A}\boldsymbol{x}\rangle = -\langle\boldsymbol{A}\boldsymbol{x},\boldsymbol{x}\rangle$$

即对任意向量 $\boldsymbol{x} \in R^n$ 都有 $\langle\boldsymbol{A}\boldsymbol{x},\boldsymbol{x}\rangle = 0$。

第3章　矩阵的变换与分解

矩阵的变换与分解在求解线性方程组、参数估计等问题中有广泛的应用。本章将对典型的矩阵变换与分解进行归类，给出其重要性质，并结合习题探讨如何对其加以灵活应用。

3.1　主要理论与方法

3.1.1　矩阵的变换

一、Householder变换

1. 定义：向量$\boldsymbol{x}$相对于向量$\boldsymbol{v}$的Householder变换记做

$$\boldsymbol{H}_{\boldsymbol{v}}\boldsymbol{x} = (\boldsymbol{P}_{\boldsymbol{v}}^{\perp} - \boldsymbol{P}_{\boldsymbol{v}})\boldsymbol{x} = \boldsymbol{P}_{\boldsymbol{v}}^{\perp}\boldsymbol{x} - \boldsymbol{P}_{\boldsymbol{v}}\boldsymbol{x} \tag{3.1}$$

其中，矩阵

$$\boldsymbol{H}_{\boldsymbol{v}} = \boldsymbol{P}_{\boldsymbol{v}}^{\perp} - \boldsymbol{P}_{\boldsymbol{v}} \tag{3.2}$$

称为向量$\boldsymbol{v}$的Householder (变换)矩阵。而

$$\boldsymbol{P}_{\boldsymbol{v}} = \boldsymbol{v}\langle\boldsymbol{v},\boldsymbol{v}\rangle^{-1}\boldsymbol{v}^{\mathrm{H}} \tag{3.3}$$

和

$$\boldsymbol{P}_{\boldsymbol{v}}^{\perp} = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{v}} = \boldsymbol{I} - \boldsymbol{v}\langle\boldsymbol{v},\boldsymbol{v}\rangle^{-1}\boldsymbol{v}^{\mathrm{H}} \tag{3.4}$$

分别为向量$\boldsymbol{v}$的投影矩阵和正交投影矩阵。事实上，$\boldsymbol{H}_{\boldsymbol{v}}\boldsymbol{x}$ 是向量$\boldsymbol{x}$关于与$\boldsymbol{v}$垂直的超平面$\boldsymbol{v}^{\perp}$ 的一个反射(镜像)。因此，Householder变换又叫镜像变换。

2. 性质：

(1) Householder矩阵是复共轭对称矩阵，即Hermitian矩阵。

(2) Householder矩阵为酉矩阵。

(3) Householder变换是对某个向量进行反射的保范数算子或保长度算子。

定理 3.1　(Householder变换的保范性)　给定任意三个向量$\boldsymbol{x},\boldsymbol{y}$和$\boldsymbol{v}$，则关系式

$$\langle\boldsymbol{H}_{\boldsymbol{v}}\boldsymbol{x},\boldsymbol{H}_{\boldsymbol{v}}\boldsymbol{y}\rangle = \langle\boldsymbol{x},\boldsymbol{y}\rangle \tag{3.5}$$

总是成立。

3. 变换算法：希望将$n\times 1$向量$\boldsymbol{x}$ 变换为一个新的(稀疏)向量$\|\boldsymbol{x}\|\boldsymbol{e}_k$，其中，$\boldsymbol{e}_k$仅有第$k$个元素为1，而其他元素皆为0。令实现这一变换的Householder矩阵为$\boldsymbol{H}_k$，则有

$$\boldsymbol{H}_k=\boldsymbol{I}-\boldsymbol{u}\boldsymbol{u}^{\mathrm{H}},\qquad \boldsymbol{u}=\frac{\boldsymbol{x}-\beta\boldsymbol{e}_k}{\sqrt{\bar{\beta}(\beta-x_k)}} \tag{3.6}$$

式中，$\boldsymbol{u}$是一个长度为2的向量，即$\|\boldsymbol{u}\|=\sqrt{2}$，并且

$$\bar{\beta}=-|x_k|\|\boldsymbol{x}\| \tag{3.7}$$

由于$\bar{\beta}x_k$必须是实数，故取

$$\beta=\pm\frac{x_k}{|x_k|}\|\boldsymbol{x}\|,\qquad x_k\neq 0 \tag{3.8}$$

当$x_k=0$时，直接取$\beta=\|\boldsymbol{x}\|$。由上式引出了两种类型的Householder变换。

1. 1型Householder变换

取$\beta=-\frac{x_k}{|x_k|}\|\boldsymbol{x}\|$，与之对应的Householder 矩阵为

$$\boldsymbol{H}_k^{(1)}=\boldsymbol{I}-\frac{1}{\|\boldsymbol{x}\|(\|\boldsymbol{x}\|+|x_k|)}(\boldsymbol{x}-\beta\boldsymbol{e}_k)(\boldsymbol{x}-\beta\boldsymbol{e}_k)^{\mathrm{H}} \tag{3.9}$$

这是通常采用的一种类型。Householder向量$\boldsymbol{u}$直接由式(3.6)计算，或者通过计算其元素得到，即

$$u_k^{(1)}=\frac{x_k}{|x_k|}\left(1+\frac{|x_k|}{\|\boldsymbol{x}\|}\right)^{1/2} \tag{3.10}$$

$$u_i^{(1)}=\frac{x_i}{|x_k|}\left(1+\frac{|x_k|}{\|\boldsymbol{x}\|}\right)^{-1/2},\quad i\neq k \tag{3.11}$$

2. 2型Householder变换

取$\beta=\frac{x_k}{|x_k|}\|\boldsymbol{x}\|$，Householder矩阵为

$$\boldsymbol{H}_k^{(2)}=\boldsymbol{I}-\frac{1+\frac{|x_k|}{\|\boldsymbol{x}\|}}{\|\boldsymbol{x}-x_k\boldsymbol{e}_k\|^2}(\boldsymbol{x}-\beta\boldsymbol{e}_k)(\boldsymbol{x}-\beta\boldsymbol{e}_k)^{\mathrm{H}} \tag{3.12}$$

此时，Householder向量$\boldsymbol{u}$由式(3.6) 计算，或者其元素直接取为

$$u_k^{(2)}=-\frac{x_k}{|x_k|}\frac{\|\boldsymbol{x}-x_k\boldsymbol{e}_k\|}{\|\boldsymbol{x}\|}\left(1+\frac{|x_k|}{\|\boldsymbol{x}\|}\right)^{-1/2} \tag{3.13}$$

$$u_i^{(2)}=\frac{x_i}{\|\boldsymbol{x}-x_k\boldsymbol{e}_k\|}\left(1+\frac{|x_k|}{\|\boldsymbol{x}\|}\right)^{1/2},\quad i\neq k \tag{3.14}$$

二、Givens旋转

对于一给定向量$\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$，则Givens变换可以使得其元素$x_i$和$x_j$ 通过角度$\theta=\arctan(s/c)$被旋转，而其他元素保持不变。Givens变换矩阵可以表示成单位矩阵的

秩2修正矩阵

$$
G(i,j,\theta)=\begin{array}{c}\begin{bmatrix}1&\cdots&0&\cdots&0&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&c&\cdots&s&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&-s&\cdots&c&\cdots&0\\ \vdots&&\vdots&&\vdots&&\vdots\\ 0&\cdots&0&\cdots&0&\cdots&1\end{bmatrix}\begin{array}{c}\\ \\ i\\ \\ j\\ \\ \\ \end{array}\\ \begin{array}{ccccccc}\quad&&i&&j&&\quad\end{array}\end{array}\tag{3.15}
$$

并满足

$$
c^2+s^2=1\tag{3.16}
$$

使用一连串Givens旋转，即可将矩阵$\boldsymbol{X}$主对角线以下的元素全部消去，得到上三角矩阵。

3.1.2 矩阵的分解

一、对角化分解

CS分解

定理 3.2 [24] (CS分解) 若$(k+j)\times(k+j)$矩阵

$$
\boldsymbol{Q}=\begin{bmatrix}\boldsymbol{Q}_{11}&\boldsymbol{Q}_{12}\\ \boldsymbol{Q}_{21}&\boldsymbol{Q}_{22}\end{bmatrix}
$$

是正交的，其中，$\boldsymbol{Q}_{11}$是$k\times k$矩阵，并且$k\geqslant j$；则存在正交矩阵$\boldsymbol{U}_1,\boldsymbol{V}_1\in R^{k\times k}$和正交矩阵$\boldsymbol{U}_2,\boldsymbol{V}_2\in R^{j\times j}$使得

$$
\begin{bmatrix}\boldsymbol{U}_1&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{U}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{Q}_{11}&\boldsymbol{Q}_{12}\\ \boldsymbol{Q}_{21}&\boldsymbol{Q}_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_1&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{V}_2\end{bmatrix}=\begin{bmatrix}\boldsymbol{I}_{k-j}&\boldsymbol{O}&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{C}&\boldsymbol{S}\\ \boldsymbol{O}&-\boldsymbol{S}&\boldsymbol{C}\end{bmatrix}\tag{3.17}
$$

其中

$$
\boldsymbol{C}=\mathrm{diag}(c_1,c_2,\cdots,c_j),\qquad c_i=\cos\theta_i\tag{3.18}
$$

$$
\boldsymbol{S}=\mathrm{diag}(s_1,s_2,\cdots,s_j),\qquad s_i=\sin\theta_i\tag{3.19}
$$

且$0\leqslant\theta_1\leqslant\theta_2\leqslant\cdots\leqslant\theta_j\leqslant\pi/2$。

二、三角化分解

1. Cholesky分解

设$\boldsymbol{A}=[a_{ij}]\in R^{n\times n}$是对称正定矩阵，$\boldsymbol{A}=\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}$称为矩阵$\boldsymbol{A}$的Cholesky分解，其中，$\boldsymbol{G}\in R^{n\times n}$是一个具有正的对角线元素的下三角矩阵，即

$$
\boldsymbol{G}=\begin{bmatrix}g_{11}&&&0\\ g_{21}&g_{22}&&\\ \vdots&\vdots&\ddots&\\ g_{n1}&g_{n2}&\cdots&g_{nn}\end{bmatrix}
$$

定理 3.3 (Cholesky分解) 如果$\boldsymbol{A} \in R^{n\times n}$是对称正定矩阵，则Cholesky分解$\boldsymbol{A} = \boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}$ 是唯一的。

2. LU分解

定理 3.4 (LU分解) 如果$\boldsymbol{A} \in R^{n\times n}$非奇异，并且其LU分解存在的话，则$\boldsymbol{A}$的LU分解是唯一的，且$\det(\boldsymbol{A}) = u_{11}u_{22}\cdots u_{nn}$。

算法 3.1.1 (矩阵的LU分解的初等行变换算法)

步骤1 利用初等行变换将矩阵$\boldsymbol{A}$ 化为阶梯型矩阵$\boldsymbol{U}$，即

$$\boldsymbol{A} \xrightarrow{\boldsymbol{E}_1} [\ \ ] \xrightarrow{\boldsymbol{E}_2} \cdots \xrightarrow{\boldsymbol{E}_k} [\ \ ] = \boldsymbol{U}$$

步骤 2 对单位矩阵执行与步骤1相应的初等行逆变换，得到单位下三角矩阵$\boldsymbol{L}$，即

$$\boldsymbol{I} \xrightarrow{\boldsymbol{E}_k^{-1}} [\ \ ] \xrightarrow{\boldsymbol{E}_{k-1}^{-1}} \cdots \xrightarrow{\boldsymbol{E}_1^{-1}} [\ \ ] = \boldsymbol{L}$$

输出 LU分解由$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$给出。

关于上述算法，有两点注意事项：

(1) 步骤2的初等行逆变换E_i^{-1}是步骤1对应的初等行变换E_i的逆变换。

(2)若在步骤1中定义$\alpha R_p + R_q$为将第p行元素乘常数α之后，加给第q行的初等行变换，则步骤2中相应初等行逆变换$(\alpha R_p + R_q)^{-1}$定义为$-\alpha R_p + R_q$。

3. QR分解

定理 3.5 (QR分解) 若$\boldsymbol{A} \in R^{m\times n}$，且$m \geqslant n$，则存在列正交的矩阵$\boldsymbol{Q} \in R^{m\times m}$和上三角矩阵$\boldsymbol{R} \in R^{m\times n}$使得$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$。当$m = n$时，$\boldsymbol{Q}$是正交矩阵。如果$\boldsymbol{A}$是非奇异的$n\times n$矩阵，则$\boldsymbol{R}$的所有对角线元素均为正，并且在这种情况下$\boldsymbol{Q}$和$\boldsymbol{R}$二者是唯一的。若$\boldsymbol{A}$是复矩阵，则$\boldsymbol{Q}$和$\boldsymbol{R}$取复值。

QR分解的算法有多种。教材中给出了基于修正Gram-Schmidt正交化，Householder变换和Givens旋转的QR分解算法。限于篇幅，这里不再列出。本章的习题中对这些算法都有涉及，读者可以在解题过程中加以体会。

三、三角—对角化分解

1. $\mathrm{LDM^T}$分解

定理 3.6 ($\mathrm{LDM^T}$分解) 若$\boldsymbol{A} \in R^{n\times n}$ 的所有子矩阵都是非奇异的，则存在两个唯一的单位下三角矩阵$\boldsymbol{L}$和$\boldsymbol{M}$以及一个唯一的对角矩阵$\boldsymbol{D} = \mathrm{diag}(d_1, d_2, \cdots, d_n)$ 使得$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{M}^{\mathrm{T}}$。

2. $\mathrm{LDL^T}$分解

定理 3.7 若$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{M}^{\mathrm{T}}$是非奇异的对称矩阵$\boldsymbol{A}$的$\mathrm{LDM^T}$分解，则$\boldsymbol{L} = \boldsymbol{M}$。

3. Schur分解

定理 3.8 (Schur分解) 若$\boldsymbol{A} \in C^{n\times n}$，则存在一个酉矩阵$\boldsymbol{U} \in C^{n\times n}$使得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{T} = \boldsymbol{D} + \boldsymbol{N} \tag{3.20}$$

其中，$D = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$；$\boldsymbol{N}$是严格上三角矩阵，即$n_{ij} = 0, \forall i \geqslant j$。

四、矩阵束的分解

定理 3.9 (广义Schur分解) 若$\boldsymbol{A},\boldsymbol{B}\in C^{n\times b}$，则存在酉矩阵$\boldsymbol{Q}$和$\boldsymbol{Z}$使得$\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{Z}=\boldsymbol{T}$和$\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{Z}=\boldsymbol{S}$均为上三角矩阵。如果对于某个$k$值，$t_{kk}$和$s_{kk}$二者都等于零，则$\lambda(\boldsymbol{A},\boldsymbol{B})$有无穷多个解。在其他情况下

$$\lambda(\boldsymbol{A},\boldsymbol{B})=\{t_{ii}/s_{ii}:\ s_{ii}\neq 0\}$$

定理 3.10 (广义实Schur分解) 若$\boldsymbol{A},\boldsymbol{B}\in R^{n\times n}$，则存在正交矩阵$\boldsymbol{Q}$和$\boldsymbol{Z}$使得$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Z}$和$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{Z}$均为上三角矩阵。

3.2 习题与解答

3.1 已知数据矩阵$\boldsymbol{X}=\begin{bmatrix}2 & 3\\ 1 & 2\\ 5 & 7\end{bmatrix}$。

(1) 试用Householder变换求$\boldsymbol{X}$的QR分解，并使用MATLAB函数[Q,R]=qr(X)校验计算的结果。

(2) 用Givens旋转求$\boldsymbol{X}$的QR分解。

思路分析： 利用4.1.3和4.7.4两节中给出的Householder变换和Givens旋转方法直接求解矩阵 $\boldsymbol{X}$的QR分解即可。

解：

(1) 根据矩阵$\boldsymbol{X}$的第一列，可直接求得第一次Householder变换向量和变换矩阵为

$$\boldsymbol{w}_1=[7.477,\ 1,\ 5]^{\mathrm{T}},\quad \boldsymbol{H}_1=\boldsymbol{I}-\frac{2\boldsymbol{w}_1\boldsymbol{w}_1^{\mathrm{T}}}{\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_1}=\begin{bmatrix}-0.3651 & -0.1826 & -0.9129\\ -0.1826 & 0.9756 & -0.1221\\ -0.9129 & -0.1221 & 0.3895\end{bmatrix}$$

经过第一次Householder变换后，矩阵$\boldsymbol{X}$变为

$$\boldsymbol{H}_1\boldsymbol{X}=\begin{bmatrix}-5.477 & -7.8507\\ 0 & 0.5488\\ 0 & -0.256\end{bmatrix}$$

接下来，令$\boldsymbol{x}=[0.5488, -0.256]^{\mathrm{T}}$，计算Householder向量和矩阵分别为

$$\boldsymbol{w}_2=[0,\ 1.154,\ -0.256]^{\mathrm{T}},\quad \boldsymbol{H}_2=\begin{bmatrix}1 & 0 & 0\\ 0 & -0.9062 & -0.4229\\ 0 & -0.4229 & 0.9062\end{bmatrix}$$

第二次变换后

$$\boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{X}=\begin{bmatrix}-5.4772 & -7.8507\\ 0 & -0.6056\\ 0 & 0\end{bmatrix}$$

从而，利用Householder变换得到的矩阵$\boldsymbol{X}$的QR分解中，

$$\boldsymbol{Q}=(\boldsymbol{H}_2\boldsymbol{H}_1)^{\mathrm{T}}=\begin{bmatrix}-0.3651 & -0.2206 & -0.9045\\ -0.1826 & -0.9357 & 0.3019\\ -0.9129 & 0.2754 & 0.3014\end{bmatrix},\quad \boldsymbol{R}=\begin{bmatrix}-5.4772 & -7.8507\\ 0 & -0.6056\\ 0 & 0\end{bmatrix}$$

经检验，利用MATLAB函数计算的结果与上面的结果完全一致。

(2) 为了消去$\boldsymbol{X}_{3,1}$，令$c=0.3714$，$s=-0.9285$，从而

$$\boldsymbol{G}_1=\begin{bmatrix}0.3714 & 0 & 0.9285\\ 0 & 1 & 0\\ -0.9285 & 0 & 0.3714\end{bmatrix},\quad \boldsymbol{G}_1\boldsymbol{X}=\begin{bmatrix}5.3853 & 7.6137\\ 1 & 2\\ 0 & -0.1857\end{bmatrix}$$

为了消去$\boldsymbol{X}_{2,1}$，令$c=0.9832$，$s=-0.1826$，从而

$$\boldsymbol{G}_2=\begin{bmatrix} 0.9832 & 0.1826 & 0 \\ -0.1826 & 0.9832 & 0 \\ 0 & 0 & 1 \end{bmatrix},\quad \boldsymbol{G}_2\boldsymbol{X}=\begin{bmatrix} 5.4774 & 7.8511 \\ 0 & 0.5761 \\ 0 & -0.1857 \end{bmatrix}$$

为了消去$\boldsymbol{X}_{3,2}$，令$c=0.9518$，$s=0.3068$，从而

$$\boldsymbol{G}_3=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.9518 & -0.3068 \\ 0 & 0.3068 & 0.9518 \end{bmatrix},\quad \boldsymbol{G}_3\boldsymbol{G}_2\boldsymbol{G}_1\boldsymbol{X}=\begin{bmatrix} 5.4774 & 7.8511 \\ 0 & 0.6053 \\ 0 & 0 \end{bmatrix}$$

综上，利用Givens旋转得到的QR分解中，

$$\boldsymbol{Q}=(\boldsymbol{G}_3\boldsymbol{G}_2\boldsymbol{G}_1)^{\mathrm{T}}=\begin{bmatrix} 0.3652 & 0.2203 & -0.9046 \\ 0.1826 & 0.9358 & 0.3016 \\ 0.9129 & -0.2753 & 0.3015 \end{bmatrix},\quad \boldsymbol{R}=\begin{bmatrix} 5.4774 & 7.8510 \\ 0 & 0.6053 \\ 0 & 0 \end{bmatrix}$$

与第一问所得结果对比后可以发现，除微小的计算误差外，二者完全相同。

3.2 已知线性方程组

$$\begin{aligned} x_1+2x_2&=-3 \\ x_1+3x_2&=10 \\ x_1-x_2&=6 \end{aligned}$$

(1) 利用Householder变换求解方程组。

(2) 利用Givens旋转求解方程组。

思路分析：利用上述两种变换求解方程组，可以先将增广矩阵通过一系列变换化为上三角矩阵，从而将原来复杂的方程组求解变得比较简单。两种变换的具体算法完全同上题。

证明：记方程组的增广矩阵为

$$\boldsymbol{A}=\begin{bmatrix} 1 & 2 & -3 \\ 1 & 3 & 10 \\ 1 & -1 & 6 \end{bmatrix}$$

(1) 根据$\boldsymbol{A}$矩阵的第一列，得到Householder变换向量和矩阵分别为

$$\boldsymbol{w}_1=[2.732,\ 1,\ 1],\quad \boldsymbol{H}_1=\begin{bmatrix} -0.5773 & -0.5774 & -0.5774 \\ -0.5774 & 0.7887 & -0.2113 \\ -0.5774 & -0.2113 & 0.7887 \end{bmatrix}$$

变换后，得到

$$\boldsymbol{H}_1\boldsymbol{A}=\begin{bmatrix} -1.7321 & -2.3094 & -7.5057 \\ 0 & 1.4226 & 8.3508 \\ 0 & -2.5774 & 4.3508 \end{bmatrix}$$

考察增广矩阵的第二列，得到

$$\boldsymbol{w}_2 = [0,\ 4.3665,\ -2.5774]^{\mathrm{T}},\quad \boldsymbol{H}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.4832 & 0.8755 \\ 0 & 0.8755 & 0.4832 \end{bmatrix}$$

变换后，得到

$$\boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{A} = \begin{bmatrix} -1.7321 & -2.3094 & -7.5057 \\ 0 & -2.9439 & -0.2262 \\ 0 & 0 & 9.4135 \end{bmatrix}$$

从而，方程组简化为

$$-1.7321x_1 - 2.3094x_2 = -7.5057$$
$$-2.9439x_2 = -0.2262$$

解得$x_2 = 0.0768$，$x_1 = 4.2309$。将解代入到原方程，得到拟合误差分别为$e_1 = 7.385$，$e_2 = 5.539$，$e_3 = 1.846$。

(2) 利用Givens变换求解。为了消去$\boldsymbol{A}_{3,1}$，令$c = 0.707$，$s = -0.707$，从而

$$\boldsymbol{G}_1 = \begin{bmatrix} 0.707 & 0 & 0.707 \\ 0 & 1 & 0 \\ -0.707 & 0 & 0.707 \end{bmatrix},\quad \boldsymbol{G}_1\boldsymbol{A} = \begin{bmatrix} 1.414 & 0.707 & 2.121 \\ 1 & 3 & 10 \\ 0 & -2.121 & 6.363 \end{bmatrix}$$

为了消去$\boldsymbol{A}_{2,1}$，令$c = 0.8164$，$s = -0.5774$，从而

$$\boldsymbol{G}_2 = \begin{bmatrix} 0.8164 & 0.5774 & 0 \\ -0.5774 & 0.8164 & 0 \\ 0 & 0 & 1 \end{bmatrix},\quad \boldsymbol{G}_2\boldsymbol{G}_1\boldsymbol{A} = \begin{bmatrix} 1.7318 & 2.3094 & 7.5056 \\ 0 & 2.0410 & 6.9393 \\ 0 & -2.121 & 6.3630 \end{bmatrix}$$

为了消去$\boldsymbol{A}_{3,2}$，令$c = 0.6934$，$s = 0.7206$，从而

$$\boldsymbol{G}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.6934 & -0.7206 \\ 0 & 0.7206 & 0.6934 \end{bmatrix},\quad \boldsymbol{G}_3\boldsymbol{G}_2\boldsymbol{G}_1\boldsymbol{A} = \begin{bmatrix} 1.7318 & 2.3094 & 7.5056 \\ 0 & 2.9436 & 0.2266 \\ 0 & 0 & 9.4126 \end{bmatrix}$$

最终，原方程组化简为

$$1.7318x_1 + 2.3094x_2 = 7.5056$$
$$2.9436x_2 = 0.2266$$

所得的结果同第一问。

3.3[8] 一个$m \times m$矩阵$\boldsymbol{A}$最多使用$m-1$次Householder变换即可三角化，变为

$$\boldsymbol{Q}\boldsymbol{A} = \boldsymbol{R}$$

式中，$\boldsymbol{R}$为上三角矩阵，并且

$$\boldsymbol{Q} = \boldsymbol{Q}_{m-1}\boldsymbol{Q}_{m-2}\cdots\boldsymbol{Q}_1$$

(1) 证明

$$\det(\boldsymbol{A})=(-1)^{m-1}\det(\boldsymbol{R})$$

(2) 利用Householder变换的保范性证明

$$|\det(\boldsymbol{A})|\leqslant\prod_{i=1}^{m}\|\boldsymbol{a}_i\|_2$$

式中，$\boldsymbol{a}_i$是矩阵$\boldsymbol{A}$的第i列。这一结果称为Hadanard定理。

思路分析：第一问的证明比较简单，直接利用$\boldsymbol{Q}$矩阵为酉矩阵的性质即可。第二问的证明中，要用到Householder变换的保范性，即变换前后的两个向量的范数相等。

证明：

(1) 首先来证明Householder变换矩阵的行列式值为-1。

任取非零向量$\boldsymbol{v}\in R^m$，构造与其对应的Householder变换矩阵

$$\boldsymbol{H}=\boldsymbol{I}-\frac{2\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}}{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}}$$

接下来，我们从$\boldsymbol{v}$出发，构造R^m中的一组正交基$\boldsymbol{v},\boldsymbol{u}_1,\cdots,\boldsymbol{u}_{m-1}$（要求$\boldsymbol{v}$是其中的一个基向量）。则不难得到，这组正交基中的任一向量都是$\boldsymbol{H}$的特征向量，因为

$$\boldsymbol{H}\boldsymbol{v}=-\boldsymbol{v},\qquad \boldsymbol{H}\boldsymbol{u}_i=\boldsymbol{u}_i$$

相应的特征值为-1（对应于特征向量$\boldsymbol{v}$）和$+1$（对应于其他特征向量$\boldsymbol{u}_i$）。因为对称矩阵的行列式为所有特征值的乘积，因此$\boldsymbol{H}$的行列式为-1。由于$\boldsymbol{v}$向量是任取的，所以以上的讨论具有普遍意义。从而

$$\det(\boldsymbol{A})=\left[\det(\boldsymbol{Q}_{m-1})\det(\boldsymbol{Q}_{m-2})\cdots\det(\boldsymbol{Q}_1)\right]^{-1}\det(\boldsymbol{R})=(-1)^{m-1}\det(\boldsymbol{R})$$

(2) 利用$\boldsymbol{R}$为上三角阵的事实及以上所得的结论，可知

$$|\det(\boldsymbol{A})|=|\det(\boldsymbol{R})|=\prod_{i=1}^{m}|\boldsymbol{R}_{ii}|\leqslant\prod_{i=1}^{m}\|\boldsymbol{r}_i\|=\prod_{i=1}^{m}\|\boldsymbol{a}_i\|$$

上式中，$\boldsymbol{r}_i$表示矩阵$\boldsymbol{R}$的第i列，$\boldsymbol{R}_{ii}$表示矩阵$\boldsymbol{R}$的第i个对角元素。不等号成立是因为一列中的任一元素的绝对值要小于该列的模。最后一个等号成立，是因为Householder变换具有保范性。

3.4 已知线性方程组

$$\begin{aligned}x_1-2x_2-x_3&=-1\\2x_1-4x_2+2x_3&=1\\2x_1+x_3&=1\\x_1+5x_2&=-2\end{aligned}$$

试用Gram-Schmidt正交化方法计算QR分解，并求解最小二乘问题。

思路分析：直接对增广矩阵的各列进行Gram-Schmidt正交化，所得到的正交的列向量组就构成了QR分解中的$\boldsymbol{Q}$矩阵。然后用$\boldsymbol{Q}^{\mathrm{T}}$矩阵左乘原增广矩阵，就得到了$\boldsymbol{R}$矩阵。从而，原来较复杂的方程组转化为一个非常简单的方程组，便于求解。

解：构造增广矩阵

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

将$\boldsymbol{A}$写成列分块的形式，即$\boldsymbol{A}=[\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3,\boldsymbol{a}_4]$，然后对这四个向量构成的向量组进行正交化。即

$$\begin{aligned}\boldsymbol{p}_1&=\boldsymbol{a}_1=[1,\ 2,\ 2,\ 1]^{\mathrm{T}}\\ \boldsymbol{u}_1&=\frac{\boldsymbol{p}_1}{\|\boldsymbol{p}_1\|}=[0.3162,\ 0.6325,\ 0.6325,\ 0.3162]^{\mathrm{T}}\end{aligned}$$

$$\begin{aligned}\boldsymbol{p}_2&=\boldsymbol{a}_2-(\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{a}_2)\boldsymbol{u}_1=[-1.5,\ -3,\ 1,\ 5.5]^{\mathrm{T}}\\ \boldsymbol{u}_2&=\frac{\boldsymbol{p}_2}{\|\boldsymbol{p}_2\|}=[-0.2301,\ -0.4602,\ 0.1534,\ 0.8437]^{\mathrm{T}}\end{aligned}$$

$$\begin{aligned}\boldsymbol{p}_3&=\boldsymbol{a}_3-(\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{a}_3)\boldsymbol{u}_1-(\boldsymbol{u}_2^{\mathrm{T}}\boldsymbol{a}_3)\boldsymbol{u}_2\\ &=[-1.6235,\ 0.7529,\ 0.0824,\ -0.0471]^{\mathrm{T}}\\ \boldsymbol{u}_3&=\frac{\boldsymbol{p}_3}{\|\boldsymbol{p}_3\|}=[-0.9059,\ 0.4201,\ 0.046,\ -0.0263]^{\mathrm{T}}\end{aligned}$$

$$\begin{aligned}\boldsymbol{p}_4&=\boldsymbol{a}_4-(\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{a}_4)\boldsymbol{u}_1-(\boldsymbol{u}_2^{\mathrm{T}}\boldsymbol{a}_4)\boldsymbol{u}_2-(\boldsymbol{u}_3^{\mathrm{T}}\boldsymbol{a}_4)\boldsymbol{u}_3\\ &=[-0.2154,\ -0.6103,\ 1.0051,\ -0.5744]^{\mathrm{T}}\\ \boldsymbol{u}_4&=\frac{\boldsymbol{p}_4}{\|\boldsymbol{p}_4\|}=[-0.1624,\ -0.4601,\ 0.7579,\ -0.4331]^{\mathrm{T}}\end{aligned}$$

从而，用Gram-Schmidt正交化计算的QR分解可以表示为

$$\boldsymbol{Q}=[\boldsymbol{u}_1,\boldsymbol{u}_2,\boldsymbol{u}_3,\boldsymbol{u}_4]=\begin{bmatrix}0.3162 & -0.2301 & -0.9059 & -0.1624\\ 0.6325 & -0.4602 & 0.4201 & -0.4601\\ 0.6325 & 0.1534 & 0.0460 & 0.7579\\ 0.3162 & 0.8437 & -0.0263 & -0.4331\end{bmatrix}.$$

$$\boldsymbol{R}=\begin{bmatrix}3.1623 & -1.5811 & 1.5811 & 0.3162\\ 0 & 6.5192 & -0.5369 & -1.7640\\ 0 & 0 & 1.7921 & 1.4245\\ 0 & 0 & 0 & 1.3263\end{bmatrix}$$

由此，得简化后的线性方程组

$$\begin{aligned}3.1623x_1-1.5811x_2+1.5811x_3&=0.3162\\ 6.5192x_2-0.5369x_3&=-1.7640\\ 1.7921x_3&=1.4245\end{aligned}$$

解得$x_1 = -0.4$，$x_2 = -0.2051$，$x_3 = 0.7949$。

3.5 计算下列矩阵的LDL$^{\mathrm{T}}$分解和Cholesky分解：

$$\boldsymbol{A} = \begin{bmatrix} 9 & 3 & -7 \\ 3 & 4 & 1 \\ -7 & 1 & 10 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 7 & 4 & -3 \\ 4 & 6 & 3 \\ -3 & 3 & 7 \end{bmatrix}$$

思路分析： 直接应用教材4.6节的算法，按照步骤求解即可。

解： 首先来求矩阵$\boldsymbol{A}$的LDL$^{\mathrm{T}}$分解。

(a) 用$\boldsymbol{a}_i$ 表示矩阵$\boldsymbol{A}$的第i行。对$\boldsymbol{A}$的各行进行初等变换。$\boldsymbol{a}_2 - \frac{1}{3}\boldsymbol{a}_1 \to \boldsymbol{a}_2, \boldsymbol{a}_3 + \frac{7}{9}\boldsymbol{a}_1 \to \boldsymbol{a}_3$，得到

$$\begin{bmatrix} 9 & 3 & -7 \\ 0 & 3 & \frac{10}{3} \\ 0 & \frac{10}{3} & \frac{41}{9} \end{bmatrix}$$

$\boldsymbol{a}_3 - \frac{10}{9}\boldsymbol{a}_2 \to \boldsymbol{a}_3$，得到

$$\begin{bmatrix} 9 & 3 & -7 \\ 0 & 3 & \frac{10}{3} \\ 0 & 0 & \frac{23}{27} \end{bmatrix}$$

(b) 设单位阵$\boldsymbol{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$，用$\boldsymbol{r}_i$表示$\boldsymbol{I}$矩阵的第$i$行。对$\boldsymbol{I}$的各行进行初等变换。

$\boldsymbol{r}_3 + \frac{10}{9}\boldsymbol{r}_2 \to \boldsymbol{r}_3$，得到

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{10}{9} & 1 \end{bmatrix}$$

$\boldsymbol{r}_2 + \frac{1}{3}\boldsymbol{r}_1 \to \boldsymbol{r}_2, \boldsymbol{r}_3 - \frac{7}{9}\boldsymbol{r}_1 \to \boldsymbol{r}_3$，得到

$$\begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{3} & 1 & 0 \\ -\frac{7}{9} & \frac{10}{9} & 1 \end{bmatrix}$$

这样便求得了$\boldsymbol{L}$和$\boldsymbol{U}$矩阵。

(c) $\boldsymbol{D}$矩阵是由$\boldsymbol{U}$矩阵的对角线上的元素组成的。到现在，已经得到了完整的LDL$^{\mathrm{T}}$分解，其中

$$\boldsymbol{L} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{3} & 1 & 0 \\ -\frac{7}{9} & \frac{10}{9} & 1 \end{bmatrix}, \quad \boldsymbol{D} = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & \frac{23}{27} \end{bmatrix}$$

下面求矩阵$\boldsymbol{A}$的Cholesky分解。依据按照教材P225给出的算法求解即可。设$\boldsymbol{A} = \boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}$，其中$\boldsymbol{G}$ 为下三角矩阵

(a) $g_{11}=\sqrt{A_{11}}=3, g_{21}=\dfrac{A_{21}}{g_{11}}=1, g_{31}=\dfrac{A_{31}}{g_{11}}=-\dfrac{7}{3}$

(b) $g_{22}=\sqrt{A_{22}-g_{21}^2}=\sqrt{3}, g_{32}=\dfrac{A_{32}-g_{31}g_{21}}{g_{22}}=\dfrac{10}{3\sqrt{3}}$

(c) $g_{33}=\sqrt{A_{33}-g_{31}^2-g_{32}^2}=\sqrt{\dfrac{23}{27}}$

这样，就可以得到矩阵$\boldsymbol{A}$的Cholesky分解的结果

$$\boldsymbol{G}=\begin{bmatrix} 3 & 0 & 0 \\ 1 & \sqrt{3} & 0 \\ -\frac{7}{3} & \frac{10}{3\sqrt{3}} & \sqrt{\frac{23}{27}} \end{bmatrix}$$

事实上，在求出了矩阵的LDL$^{\mathrm{T}}$ 分解后，可以直接利用下式求其Cholesky分解

$$\boldsymbol{G}=\boldsymbol{L}\times\sqrt{\boldsymbol{D}}$$

这样得到的结果与前面完全相同。

$\boldsymbol{B}$矩阵的LDL$^{\mathrm{T}}$ 分解过程同上，这里直接给出结果。$\boldsymbol{B}$矩阵的LU分解为

$$\boldsymbol{U}=\begin{bmatrix} 7 & 4 & -3 \\ 0 & \frac{26}{7} & \frac{33}{7} \\ 0 & 0 & -\frac{7}{26} \end{bmatrix},\qquad \boldsymbol{L}=\begin{bmatrix} 1 & 0 & 0 \\ \frac{4}{7} & 1 & 0 \\ -\frac{3}{7} & \frac{33}{26} & 1 \end{bmatrix}$$

$\boldsymbol{B}$的LDL$^{\mathrm{T}}$分解为

$$\boldsymbol{L}=\begin{bmatrix} 1 & 0 & 0 \\ \frac{4}{7} & 1 & 0 \\ -\frac{3}{7} & \frac{33}{26} & 1 \end{bmatrix},\qquad \boldsymbol{D}=\begin{bmatrix} 7 & & \\ & \frac{26}{7} & \\ & & -\frac{7}{26} \end{bmatrix}$$

由于矩阵$\boldsymbol{B}$不正定，所以无法进行Cholesky分解。

3.6 使用Gram-Schmidt 正交化方法和修正Gram-Schmidt正交化方法求数据矩阵

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

的QR分解，并比较两种方法的结果。

解: 将$\boldsymbol{A}$写成按列分块的形式，即$\boldsymbol{A}=[\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3]$。

首先使用Gram-Schmidt正交化方法进行计算。

$$\boldsymbol{p}_1=\boldsymbol{a}_1=[4,\ 2,\ 2,\ 1]^{\mathrm{T}}$$

$$\boldsymbol{u}_1=\frac{\boldsymbol{p}_1}{\|\boldsymbol{p}_1\|}=[0.8,\ 0.4,\ 0.4,\ 0.2]^{\mathrm{T}}$$

$$\boldsymbol{p}_2=\boldsymbol{a}_2-(\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{a}_2)\boldsymbol{u}_1=[0.4,\ -0.8,\ -0.8,\ 1.6]^{\mathrm{T}}$$
$$\boldsymbol{u}_2=\frac{\boldsymbol{p}_2}{\|\boldsymbol{p}_2\|}=[0.2,\ -0.4,\ -0.4,\ 0.8]^{\mathrm{T}}$$

$$\boldsymbol{p}_3=\boldsymbol{a}_3-(\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{a}_3)\boldsymbol{u}_1-(\boldsymbol{u}_2^{\mathrm{T}}\boldsymbol{a}_3)\boldsymbol{u}_2=[0,\ 1,\ -1,\ 0]^{\mathrm{T}}$$
$$\boldsymbol{u}_3=\frac{\boldsymbol{p}_3}{\|\boldsymbol{p}_3\|}=[0,\ 0.7071,\ -0.7071,\ 0]^{\mathrm{T}}$$

所以，用Gram-Schmidt正交化方法得到的矩阵$\boldsymbol{A}$的QR分解可以表示为

$$\boldsymbol{Q}=[\boldsymbol{u}_1,\boldsymbol{u}_2,\boldsymbol{u}_3]=\begin{bmatrix}0.8 & 0.2 & 0\\0.4 & -0.4 & 0.7071\\0.4 & -0.4 & -0.7071\\0.2 & 0.8 & 1\end{bmatrix}$$

$$\boldsymbol{R}=\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}=\begin{bmatrix}5 & 2 & 1\\0 & 2 & 1\\0 & 0 & 1.4142\end{bmatrix}$$

接下来，利用修正的Gram-Schmidt正交化方法计算QR分解。将$\boldsymbol{a}_1$标准化，并将其从$\boldsymbol{a}_2$，$\boldsymbol{a}_3$中减去

$$\boldsymbol{u}_1=\frac{\boldsymbol{a}_1}{\|\boldsymbol{a}_1\|}=[0.8,\ 0.4,\ 0.4,\ 0.2]^{\mathrm{T}}$$
$$\boldsymbol{a}_2^{(1)}=\boldsymbol{a}_2-(\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{a}_2)\boldsymbol{u}_1=[0.4,\ -0.8,\ -0.8,\ 1.6]^{\mathrm{T}}$$
$$\boldsymbol{a}_3^{(1)}=\boldsymbol{a}_3-(\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{a}_3)\boldsymbol{u}_1=[0.2,\ 0.6,\ -1.4,\ 0.8]$$

标准化$\boldsymbol{a}_2^{(1)}$，并将其从$\boldsymbol{a}_3^{(1)}$中减去，得到

$$\boldsymbol{u}_2=\frac{\boldsymbol{a}_2^{(1)}}{\left\|\boldsymbol{a}_2^{(1)}\right\|}=[0.2,\ -0.4,\ -0.4,\ 0.8]^{\mathrm{T}}$$

$$\boldsymbol{a}_3^{(2)}=\boldsymbol{a}_3^{(1)}-\left(\boldsymbol{u}_2^{\mathrm{T}}\boldsymbol{a}_3^{(1)}\right)\boldsymbol{u}_2=[0,\ 1,\ -1,\ 0]^{\mathrm{T}}$$

最后，将$\boldsymbol{a}_3^{(2)}$标准化，得到

$$\boldsymbol{u}_3=\frac{\boldsymbol{a}_3^{(2)}}{\left\|\boldsymbol{a}_3^{(2)}\right\|}=[0,\ 0.7071,\ -0.7071,\ 0]^{\mathrm{T}}$$

综上，利用修正的Gram-Schmidt正交化方法得到的QR分解可以表示如下：

$$\boldsymbol{Q}=[\boldsymbol{u}_1,\boldsymbol{u}_2,\boldsymbol{u}_3]=\begin{bmatrix}0.8 & 0.2 & 0\\0.4 & -0.4 & 0.7071\\0.4 & -0.4 & -0.7071\\0.2 & 0.8 & 1\end{bmatrix}$$

$$\boldsymbol{R}=\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}=\begin{bmatrix}5 & 2 & 1\\0 & 2 & 1\\0 & 0 & 1.4142\end{bmatrix}$$

可见，对于本例，用两种正交化方法得到的QR分解完全相同。

3.7 证明一个$n \times n$矩阵$\boldsymbol{A}$正定，当且仅当$\boldsymbol{A}$存在Cholesky分解，即对某个可逆的上三角矩阵$\boldsymbol{R}$，有$\boldsymbol{A} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$。

证明：首先证明充分性。因为$\boldsymbol{A} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$，所以$\boldsymbol{A}$为对称矩阵。任取$n$维向量$\boldsymbol{x}$，则有

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{x} = (\boldsymbol{R}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{R}\boldsymbol{x} \geqslant 0$$

由于$\boldsymbol{R}$是可逆矩阵，所以当且仅当$\boldsymbol{x} = 0$时，等号成立。这样就证明了$\boldsymbol{A}$为对称正定矩阵。

必要性由教材P225页给出，即任意对称正定矩阵都可以分解成$\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$，其中$\boldsymbol{R}$为上三角矩阵，教材还给出了具体的分解算法。

3.8 令Hermitian 矩阵$\boldsymbol{A}$非负定，证明：存在一个三角矩阵$\boldsymbol{T}$，使得$\boldsymbol{A} = \boldsymbol{T}\boldsymbol{T}^{\mathrm{H}}$。

思路分析：正定矩阵可以进行Cholesky分解，非负定矩阵并没有这样的性质。但我们仍然可以构造出下三角矩阵(对角元素不全为正)，使其与自身的乘积等于该非负定矩阵。证明过程需要利用矩阵特征值分解和QR分解的有关性质。

证明：如果能构造出$\boldsymbol{T}$矩阵来，则问题自然就解决了。下面，我们具体考虑如何构造。首先，对$\boldsymbol{A}$进行特征值分解，得到

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{T}}$$

若令$\boldsymbol{P} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{U}^{\mathrm{T}}$，则显然有$\boldsymbol{A} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}$。

接下来，对$\boldsymbol{P}$进行QR分解，设为$\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{R}$，其中$\boldsymbol{Q}$为正交阵，$\boldsymbol{R}$为上三角阵。将此式代回去可以发现

$$\boldsymbol{A} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{P} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$$

其中用到了$\boldsymbol{Q}$为正交阵的性质。这样，我们只需要令$\boldsymbol{T} = \boldsymbol{R}^{\mathrm{T}}$，即得到了满足题设要求的三角阵。

3.9 利用LU分解，求解线性方程组

$$\begin{aligned} x_1 + 3x_2 &= 1 \\ 2x_1 + x_2 &= 3 \\ 3x_1 + 4x_2 + x_3 &= 0 \end{aligned}$$

思路分析：对于形如$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$的线性方程组，首先对系数矩阵进行LU分解，得到$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$。用前向回代法求解下三角矩阵方程：$\boldsymbol{L}\boldsymbol{y} = \boldsymbol{b}$。最后用后向回代法求解上三角矩阵方程：$\boldsymbol{U}\boldsymbol{x} = \boldsymbol{y}$。

解：由方程组得到系数矩阵

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

对矩阵A进行LU分解如下

(a) $a_2-2a_1\to a_2, a_3-3a_1\to a_3$得到

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

$a_3-a_2\to a_3$得到

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

(b) 设单位阵$I=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$，对其进行相应的初等行逆变换。$r_3+r_2\to r_3$得到

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

$r_2+2r_1\to r_2, r_3+3r_1\to r_3$得到

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

从而，

$$A=\begin{bmatrix}1&3&0\\2&1&0\\3&4&1\end{bmatrix}=\begin{bmatrix}1&0&0\\2&1&0\\3&1&1\end{bmatrix}\begin{bmatrix}1&3&0\\0&-5&0\\0&0&1\end{bmatrix}$$

先求解方程组$Ly=[1,\ 3,\ 0]^{\mathrm{T}}$，得到$y=[1,\ 1,\ -4]^{\mathrm{T}}$。再求解方程组$Ux=y$，得到$x=\left[\frac{8}{5},-\frac{1}{5},-4\right]^{\mathrm{T}}$。

3.10 计算下列矩阵的满秩分解：

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

思路分析：利用教材P78页算法1.8.1可以直接计算任意给定矩阵的满秩分解。

解：考察矩阵A。不难看出，A实际上是满秩的。因此，最为简单的满秩分解之一就是写成单位矩阵与它本身的乘积。然而，为了方便读者熟悉算法1.8.1，我们还是给出详细的计算过程。设A的三行分别为a_1，a_2和a_3。对其进行初等行变换：

$\boldsymbol{a}_2-\boldsymbol{a}_1\rightarrow\boldsymbol{a}_2$，$\boldsymbol{a}_3-3\boldsymbol{a}_1\rightarrow\boldsymbol{a}_3$得到

$$\begin{bmatrix}1&2&3\\0&-2&-2\\0&-4&-7\end{bmatrix}$$

$\boldsymbol{a}_3-2\boldsymbol{a}_2\rightarrow\boldsymbol{a}_3$，得到

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

对单位阵做上述变换的逆变换，得到

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

从而，矩阵$\boldsymbol{A}$的满秩分解可表示为$\boldsymbol{A}=\boldsymbol{F}_1\boldsymbol{G}_1$，其中

$$\boldsymbol{F}_1=\begin{bmatrix}1&0&0\\1&1&0\\3&2&1\end{bmatrix},\quad \boldsymbol{G}_1=\begin{bmatrix}1&2&3\\0&-2&-2\\0&0&-3\end{bmatrix}$$

经检验，发现$\boldsymbol{F}_1$和$\boldsymbol{G}_1$的秩都为3，与$\boldsymbol{A}$相同。

$\boldsymbol{B}$的满秩分解计算方法完全同上，只是这里的$\boldsymbol{B}$矩阵不再满秩（秩为1）。若记$\boldsymbol{B}=\boldsymbol{F}_2\boldsymbol{G}_2$，则有

$$\boldsymbol{F}_2=[2,1]^{\mathrm{T}},\quad \boldsymbol{G}_2=[1,1]$$

3.11　证明：若$\boldsymbol{A}$正定，则$\boldsymbol{A}$存在LU分解$\boldsymbol{A}=\boldsymbol{LU}$，且$\boldsymbol{U}$的对角元素为正数。

思路分析： 对于正定矩阵，很容易想起Cholesky分解，分解结果与LU分解很相似。只不过前者分解得到下三角矩阵与它本身转置的乘积，而后者得到单位下三角矩阵与另外的上三角矩阵的乘积。从Cholesky分解出发，容易证明此题。

证明： 由于$\boldsymbol{A}$是正定矩阵，则通过Cholesky分解，我们可以得到$\boldsymbol{A}=\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}$，其中$\boldsymbol{G}$为对角线元素全为正数的下三角矩阵。定义

$$\boldsymbol{\Lambda}=\operatorname{diag}(g_{11},\cdots,g_{nn})$$

其中，g_{ii}为矩阵$\boldsymbol{G}$的对角元素，并且全部大于零。令

$$\boldsymbol{L}=\boldsymbol{G}\boldsymbol{\Lambda}^{-1},\quad \boldsymbol{U}=\boldsymbol{\Lambda}\boldsymbol{G}$$

显然，$\boldsymbol{L}$和$\boldsymbol{U}$分别为单位下三角矩阵和对角线元素全为正数的上三角矩阵。此题得证。

3.12　分别利用Householder变换和Givens旋转求解方程$\boldsymbol{Ax}=\boldsymbol{b}$，其中

$$\boldsymbol{A}=\begin{bmatrix}4&2\\2&0\\1&2\\1&2\end{bmatrix},\quad \boldsymbol{b}=\begin{bmatrix}1\\1\\-1\\1\end{bmatrix}$$

解：利用Householder变换和Givens旋转求解线性方程组的具体过程已经在本章的习题3.1中给出，这里不再重复。事实上，两者都是希望通过某种变换使得增广矩阵成为上三角矩阵，从而简化后续的计算。下面直接给出简化后的方程组

$$
\begin{aligned}
-4.6904x_1 - 2.5584x_2 &= -1.2792 \\
2.3355x_2 &= -0.5449
\end{aligned}
$$

解得$x_1 = 0.4$，$x_2 = -0.2333$。

3.13 利用MATLAB进行LU分解的函数$[\mathrm{L}, \mathrm{U}] = \mathrm{lu}(\mathbf{X})$计算下列矩阵的$\mathrm{LDL}^{\mathrm{T}}$ 分解：

$$
\boldsymbol{A} = \begin{bmatrix} 15 & 7 & 10 \\ 7 & 8 & 9 \\ 10 & 9 & 13 \end{bmatrix}
$$

证明：首先利用MATLAB函数计算矩阵的LU分解，得到

$$
\boldsymbol{L} = \begin{bmatrix} 1 & 0 & 0 \\ 0.4667 & 1 & 0 \\ 0.6667 & 0.9155 & 1 \end{bmatrix}, \quad \boldsymbol{U} = \begin{bmatrix} 15 & 7 & 10 \\ 0 & 4.7333 & 4.3333 \\ 0 & 0 & 2.3662 \end{bmatrix}
$$

进而，令

$$
\boldsymbol{D} = \begin{bmatrix} 15 & & \\ & 4.7333 & \\ & & 2.3662 \end{bmatrix}
$$

这样就得到了矩阵$\boldsymbol{A}$的$\mathrm{LDL}^{\mathrm{T}}$ 分解结果，即$\boldsymbol{A} = \boldsymbol{LDL}^{\mathrm{T}}$。

3.14 利用MATLAB函数$[\mathrm{Q}, \mathrm{R}] = \mathrm{qr}(\mathrm{A})$求矩阵

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

的QR分解。若$\boldsymbol{b} = [1, 2, 3, 4]^{\mathrm{T}}$，试求解方程$\boldsymbol{Ax} = \boldsymbol{b}$。

解：直接利用MATLAB函数，可得$\boldsymbol{A}$矩阵的QR分解为

$$
\boldsymbol{Q} = \begin{bmatrix} -0.2582 & -0.3545 & 0.8006 & -0.4082 \\ -0.5164 & -0.7089 & -0.4804 & 0 \\ -0.2582 & 0.4051 & -0.3203 & -0.8165 \\ -0.7746 & 0.4557 & 0.1601 & 0.4082 \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} -3.8730 & -2.0656 & -3.3566 \\ 0 & -1.3166 & 0.7089 \\ 0 & 0 & 0.4804 \\ 0 & 0 & 0 \end{bmatrix}
$$

用$\boldsymbol{Q}^{\mathrm{T}}$左乘$\boldsymbol{b}$得到，$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b} = [-5.1640,\ 1.2659,\ -0.4804,\ -1.2247]^{\mathrm{T}}$。从而，原方程组化简为

$$
\begin{aligned}
-3.873x_1 - 2.0656x_2 - 3.3566x_3 &= -5.164 \\
-1.3166x_2 + 0.7089x_3 &= 1.2659 \\
0.4804x_3 &= -0.4804
\end{aligned}
$$

解得$x_1 = 3$，$x_2 = -1.5$，$x_3 = -1$。

3.15 证明$M \times M$维Hermitian 正定矩阵$\boldsymbol{A} = [a_{ij}]$的Cholesky 分解$\boldsymbol{A} = \boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}$可以利用[7, p.441]

for $j = 1 : M$

$$g_{jj} = \left(a_{jj} - \sum_{k=1}^{j-1} |g_{jk}|^2\right)^{1/2}$$

for $i = j + 1 : M$

$$g_{ij} = \frac{1}{g_{jj}}\left(a_{ij} - \sum_{k=1}^{j-1} g_{ik}^* g_{jk}\right)$$

end i

end j

计算。

证明：设

$$\boldsymbol{G} = \begin{bmatrix} g_{11} & & & 0 \\ g_{21} & g_{22} & & \\ \vdots & \vdots & \ddots & \\ g_{n1} & g_{n2} & \cdots & g_{nn} \end{bmatrix}$$

比较$\boldsymbol{A} = \boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}$两边，可以得到

$$a_{ij} = \sum_{k=1}^{j} g_{jk} g_{ik}$$

经过简单变换，得到

$$g_{jj} g_{ij} = a_{ij} - \sum_{k=1}^{j-1} g_{jk} g_{ik} = v(i)$$

即知道了$\boldsymbol{G}$的前$j-1$列元素后，就可以计算得到$v(i)$。在上面的等式中，进而令$i = j$，可以得到

$$g_{ij} = \frac{v(i)}{g_{jj}} = \frac{v(i)}{\sqrt{v(j)}}$$

综上所述，就得到了题目中给出的迭代计算方法。事实上，仔细观察可以发现，此题目给出的算法就是教材4.6.1节中介绍的Gaxpy Cholesky算法。

3.16 求矩阵

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

的Jordan标准型。

解：按照求矩阵Jordan标准型的计算步骤求解即可。由于$\boldsymbol{A}$是上三角阵，所以不难得到其特征值为1和3，且前者的代数重数为2，后者为1。这样就可以确定Jordan标准型共有2个块，一个2阶，另一个1阶。

求解对应于特征值1的特征向量，得到$\boldsymbol{x}_1 = [1,0,0]^{\mathrm{T}}$，对应于特征值3的特征向量是$\boldsymbol{x}_3 = [-2,3,2]^{\mathrm{T}}$。求与特征值1对应的广义特征向量，即解方程$(\boldsymbol{A}-\boldsymbol{I})\boldsymbol{x} = \boldsymbol{x}_1$，得到$\boldsymbol{x}_2 = [0,\ -0.5,\ 0]^{\mathrm{T}}$。

综上所述，矩阵$\boldsymbol{A}$的Jordan标准型可以表示为

$$\boldsymbol{A} = \boldsymbol{P}\begin{bmatrix} 3 & & \\ & 1 & 1 \\ & & 1 \end{bmatrix}\boldsymbol{P}^{-1}$$

其中，

$$\boldsymbol{P} = \begin{bmatrix} -2 & 1 & 0 \\ 3 & 0 & 0.5 \\ 2 & 0 & 0 \end{bmatrix}$$

3.17 证明：若非奇异矩阵$\boldsymbol{A}$的$\mathrm{LDM}^{\mathrm{T}}$分解存在，则这一分解是唯一的。

思路分析： 若矩阵的LU分解存在，则这一分解是唯一的。从这里出发，很容易证明本题目。

证明： 设$\boldsymbol{A} = \boldsymbol{LDM}^{\mathrm{T}}$。若记$\boldsymbol{U} = \boldsymbol{DM}^{\mathrm{T}}$，则其为一上三角阵。由定理4.6.2可知，非奇异矩阵的LU分解一定是唯一的。这就表明，前述的$\boldsymbol{L}$和$\boldsymbol{U}$是唯一的。又因为$\boldsymbol{M}$是特殊的下三角阵，即对角线元素全部为1，则可知$\boldsymbol{D}$和$\boldsymbol{M}$也必然是唯一的。其中，$\boldsymbol{D}$是由$\boldsymbol{U}$的对角元素组成的对角阵，$\boldsymbol{M} = \boldsymbol{D}^{-1}\boldsymbol{U}$。

令$\boldsymbol{D}$矩阵为$\boldsymbol{U}$的对角线上的元素所组成的对角阵，则$\boldsymbol{M}$可唯一确定，即$\boldsymbol{M} = \boldsymbol{D}^{-1}\boldsymbol{U}$。这样就证明了$\boldsymbol{LDM}^{T}$分解的唯一性。

3.18 [4] 令$\boldsymbol{A} = \boldsymbol{I} + \boldsymbol{uu}^{\mathrm{T}}$，其中，$\boldsymbol{A} \in R^{n\times n}$和$\|\boldsymbol{u}\| = 1$。求矩阵$\boldsymbol{A}$的Cholesky因子的主对角线和次对角线的显式表达式。

思路分析： 由题意可知，$\boldsymbol{A}$为正定矩阵，并且其形式非常特殊，为单位阵的秩一修正。我们可以利用教材P225~226页给出的求矩阵Cholesky分解的步骤来得到$\boldsymbol{G}$矩阵前几行的解析表达形式，进而推测出分解结果的一般形式，最后用数学归纳法证明。

解：

$$\boldsymbol{A} = \boldsymbol{I} + \boldsymbol{uu}^{\mathrm{T}} = \begin{bmatrix} 1+u_1^2 & u_1u_2 & \cdots & u_1u_n \\ u_2u_1 & 1+u_2^2 & \cdots & u_2u_n \\ \vdots & & \ddots & \\ u_nu_1 & u_nu_2 & \cdots & 1+u_n^2 \end{bmatrix} = \boldsymbol{GG}^{\mathrm{T}}$$

其中，

$$\boldsymbol{G} = \begin{bmatrix} g_{1,1} & & & \\ g_{2,1} & g_{2,2} & & \\ \vdots & \vdots & \ddots & \\ g_{n,1} & g_{n,2} & \cdots & g_{n,n} \end{bmatrix}$$

根据对应元素分别相等的规律可得

$$g_{1,1}^2 = 1 + u_1^2 \Rightarrow g_{1,1} = \sqrt{1+u_1^2}$$

$$g_{2,1}g_{1,1} = u_1u_2 \Rightarrow g_{2,1} = \frac{u_1u_2}{\sqrt{1+u_1^2}}$$

同理，

$$g_{n,1}g_{1,1} = u_1u_n \Rightarrow g_{n,1} = \frac{u_1u_n}{\sqrt{1+u_1^2}}\,, \quad n > 1$$

$\boldsymbol{G}$矩阵的第二行元素可以表示为

$$g_{2,1}^2 + g_{2,2}^2 = u_2^2 + 1 \Rightarrow g_{2,2} = u_2^2 + 1 - \frac{u_1^2u_2^2}{1+u_1^2} = \sqrt{\frac{1+u_1^2+u_2^2}{1+u_1^2}}$$

$$g_{2,1}g_{n,1} + g_{2,2}g_{n,2} = u_2u_n \Rightarrow g_{n,2} = \frac{u_2u_n}{\sqrt{(1+u_1^2)(1+u_1^2+u_2^2)}}\,, \quad n > 2$$

由此可以发现规律：$\boldsymbol{G}$矩阵主对角线上的元素为

$$g_{i,i} = \sqrt{\frac{1+u_1^2+\cdots+u_i^2}{1+u_1^2+\cdots+u_{i-1}^2}}$$

次对角线上的元素为

$$g_{i+1,i} = \frac{u_iu_{i+1}}{\prod\limits_{k=1}^{i}(1+u_1^2+\cdots+u_k^2)^{\frac{1}{2}}}$$

利用数学归纳法不难证明上述结论，这里从略。

第4章 梯度分析与最优化

函数的极值问题研究具有非同寻常的实用价值。而通常一个实际问题中的函数变量不是一个简单的一元标量，而是一个向量，甚至一个矩阵，它们的梯度如何求，这是本章习题的重点，所以该章重要性毋庸置疑。本章习题具有一定难度，但是建议读者好好练习，熟练掌握向量、矩阵变元微分、求导的技能以及约束优化问题求解的通用方法。它将作为一门重要工具使您在科研工作中受益无穷。

4.1 主要理论与方法

4.1.1 梯度矩阵和Jacobian矩阵的定义

定义可微分实函数$\boldsymbol{F}(\boldsymbol{X}) \in R^{m\times p}$，其变量矩阵$\boldsymbol{X} \in R^{n\times q}$。则梯度矩阵表示为

$$\frac{\partial \boldsymbol{F}(\boldsymbol{X})}{\partial \boldsymbol{X}} = \begin{bmatrix} \partial f_{11}/\partial \boldsymbol{X} & \dots & \partial f_{1p}/\partial \boldsymbol{X} \\ \vdots & \ddots & \vdots \\ \partial f_{m1}/\partial \boldsymbol{X} & \dots & \partial f_{mp}/\partial \boldsymbol{X} \end{bmatrix} \tag{4.1}$$

Jacobian矩阵表示为

$$\mathrm{D}\boldsymbol{F}(\boldsymbol{X}) = \frac{\partial \mathrm{vec}\boldsymbol{F}(\boldsymbol{X})}{\partial \mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}} \tag{4.2}$$

注意 Jacobian矩阵是$mp \times nq$维的，而梯度矩阵是$mn \times pq$维。

4.1.2 实值函数相对于实向量的梯度

(1) 若$f(\boldsymbol{x}) = c$为常数，则梯度$\dfrac{\partial c}{\partial \boldsymbol{x}} = \boldsymbol{0}$。

(2) 线性法则：若$f(\boldsymbol{x})$和$g(\boldsymbol{x})$ 分别是向量$\boldsymbol{x}$的实值函数，c_1 和c_2为实常数，则

$$\frac{\partial[c_1 f(\boldsymbol{x}) + c_2 g(\boldsymbol{x})]}{\partial \boldsymbol{x}} = c_1 \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} + c_2 \frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}} \tag{4.3}$$

(3) 乘积法则：

① 若$f(\boldsymbol{x})$和$g(\boldsymbol{x})$ 都是向量$\boldsymbol{x}$的实值函数，则

$$\frac{\partial f(\boldsymbol{x})g(\boldsymbol{x})}{\partial \boldsymbol{x}} = g(\boldsymbol{x})\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} + f(\boldsymbol{x})\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}} \tag{4.4}$$

② 若$f(\boldsymbol{x}),g(\boldsymbol{x})$ 和$h(\boldsymbol{x})$都是向量$\boldsymbol{x}$ 的实值函数，则

$$\frac{\partial f(\boldsymbol{x})g(\boldsymbol{x})h(\boldsymbol{x})}{\partial \boldsymbol{x}}=g(\boldsymbol{x})h(\boldsymbol{x})\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}+f(\boldsymbol{x})h(\boldsymbol{x})\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}}+\quad f(\boldsymbol{x})g(\boldsymbol{x})\frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}} \tag{4.5}$$

(4) 商法则：若$g(\boldsymbol{x})\neq 0$，则

$$\frac{\partial f(\boldsymbol{x})/g(\boldsymbol{x})}{\partial \boldsymbol{x}}=\frac{1}{g^2(\boldsymbol{x})}\left[g(\boldsymbol{x})\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}-f(\boldsymbol{x})\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}}\right] \tag{4.6}$$

(5) 链式法则：若$\boldsymbol{y}(\boldsymbol{x})$ 是$\boldsymbol{x}$的向量值函数，则

$$\frac{\partial f(\boldsymbol{y}(\boldsymbol{x}))}{\partial \boldsymbol{x}}=\frac{\partial \boldsymbol{y}^{\mathrm{T}}(\boldsymbol{x})}{\partial \boldsymbol{x}}\frac{\partial f(\boldsymbol{y})}{\partial \boldsymbol{y}} \tag{4.7}$$

式中，$\dfrac{\partial \boldsymbol{y}^{\mathrm{T}}(\boldsymbol{x})}{\partial \boldsymbol{x}}$为$n\times n$矩阵。

(6) 若$n\times 1$向量$\boldsymbol{a}$与$\boldsymbol{x}$是无关的常数向量，则

$$\frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}}{\partial \boldsymbol{x}}=\boldsymbol{a},\qquad \frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{a}}{\partial \boldsymbol{x}}=\boldsymbol{a}$$

(7) 若$n\times 1$向量$\boldsymbol{a}$与$\boldsymbol{x}$是无关的常数向量，则

$$\frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{y}(\boldsymbol{x})}{\partial \boldsymbol{x}}=\frac{\partial \boldsymbol{y}^{\mathrm{T}}(\boldsymbol{x})}{\partial \boldsymbol{x}}\boldsymbol{a},\qquad \frac{\partial \boldsymbol{y}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{a}}{\partial \boldsymbol{x}}=\frac{\partial \boldsymbol{y}^{\mathrm{T}}(\boldsymbol{x})}{\partial \boldsymbol{x}}\boldsymbol{a}$$

(8) 若$\boldsymbol{A}$和$\boldsymbol{y}$均与向量$\boldsymbol{x}$无关，则

$$\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}}{\partial \boldsymbol{x}}=\boldsymbol{A}\boldsymbol{y},\qquad \frac{\partial \boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$$

(9) 令$\boldsymbol{A}$是一个与向量$\boldsymbol{x}$无关的矩阵，则

$$\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}}{\partial \boldsymbol{x}}=\boldsymbol{A},\qquad \frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}=(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})\boldsymbol{x}$$

特别地，若$\boldsymbol{A}$为对称矩阵，则有$\dfrac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}}=2\boldsymbol{A}\boldsymbol{x}$。

(10) 若$\boldsymbol{A}$是与向量$\boldsymbol{x}$ 无关，而$\boldsymbol{y}(\boldsymbol{x})$与向量$\boldsymbol{x}$的元素有关，则

$$\frac{\partial [\boldsymbol{y}(\boldsymbol{x})]^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}(\boldsymbol{x})}{\partial \boldsymbol{x}}=\frac{\partial [\boldsymbol{y}(\boldsymbol{x})]^{\mathrm{T}}}{\partial \boldsymbol{x}}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})\boldsymbol{y}(\boldsymbol{x})$$

(11) 若$\boldsymbol{A}$是一个与向量$\boldsymbol{x}$无关的矩阵，而$\boldsymbol{y}(\boldsymbol{x})$ 和$\boldsymbol{z}(\boldsymbol{x})$是与向量$\boldsymbol{x}$的元素有关的列向量，则

$$\frac{\partial [\boldsymbol{y}(\boldsymbol{x})]^{\mathrm{T}}\boldsymbol{A}\boldsymbol{z}(\boldsymbol{x})}{\partial \boldsymbol{x}}=\frac{\partial [\boldsymbol{y}(\boldsymbol{x})]^{\mathrm{T}}}{\partial \boldsymbol{x}}\boldsymbol{A}\boldsymbol{z}(\boldsymbol{x})+\frac{\partial [\boldsymbol{z}(\boldsymbol{x})]^{\mathrm{T}}}{\partial \boldsymbol{x}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}(\boldsymbol{x})$$

(12) 令$\boldsymbol{x}$为$n\times 1$向量，$\boldsymbol{a}$为$m\times 1$常数向量，$\boldsymbol{A}$和$\boldsymbol{B}$分别为$m\times n$和$m\times m$ 常数矩阵，且$\boldsymbol{B}$为对称矩阵，则

$$\frac{\partial (\boldsymbol{a}-\boldsymbol{A}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{B}(\boldsymbol{a}-\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}}=-2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}(\boldsymbol{a}-\boldsymbol{A}\boldsymbol{x})$$

4.1.3 实值函数的梯度矩阵

(1) 若$f(\boldsymbol{A})=c$为常数，其中，$\boldsymbol{A}$为$m\times n$矩阵，则梯度$\dfrac{\partial c}{\partial \boldsymbol{A}}=\boldsymbol{O}_{m\times n}$。

(2) 线性法则：若$f(\boldsymbol{A})$和$g(\boldsymbol{A})$ 分别是矩阵$\boldsymbol{A}$的实值函数，c_1和c_2为实常数，则

$$\frac{\partial[c_1 f(\boldsymbol{A})+c_2 g(\boldsymbol{A})]}{\partial \boldsymbol{A}}=c_1\frac{\partial f(\boldsymbol{A})}{\partial \boldsymbol{A}}+c_2\frac{\partial g(\boldsymbol{A})}{\partial \boldsymbol{A}} \tag{4.8}$$

(3) 乘积法则：若$f(\boldsymbol{A}),g(\boldsymbol{A})$和$h(\boldsymbol{A})$都是矩阵$\boldsymbol{A}$ 的实值函数，则

$$\frac{\partial f(\boldsymbol{A})g(\boldsymbol{A})}{\partial \boldsymbol{A}}=g(\boldsymbol{A})\frac{\partial f(\boldsymbol{A})}{\partial \boldsymbol{A}}+f(\boldsymbol{A})\frac{\partial g(\boldsymbol{A})}{\partial \boldsymbol{A}} \tag{4.9}$$

$$\frac{\partial f(\boldsymbol{A})g(\boldsymbol{A})h(\boldsymbol{A})}{\partial \boldsymbol{A}}=g(\boldsymbol{A})h(\boldsymbol{A})\frac{\partial f(\boldsymbol{A})}{\partial \boldsymbol{A}}+f(\boldsymbol{A})h(\boldsymbol{A})\frac{\partial g(\boldsymbol{A})}{\partial \boldsymbol{A}}+\quad f(\boldsymbol{A})g(\boldsymbol{A})\frac{\partial h(\boldsymbol{A})}{\partial \boldsymbol{A}} \tag{4.10}$$

(4) 商法则：若$g(\boldsymbol{A})\neq 0$，则

$$\frac{\partial f(\boldsymbol{A})/g(\boldsymbol{A})}{\partial \boldsymbol{A}}=\frac{1}{g^2(\boldsymbol{A})}\left[g(\boldsymbol{A})\frac{\partial f(\boldsymbol{A})}{\partial \boldsymbol{A}}-f(\boldsymbol{A})\frac{\partial g(\boldsymbol{A})}{\partial \boldsymbol{A}}\right] \tag{4.11}$$

(5) 链式法则：令$\boldsymbol{A}$为$m\times n$矩阵，且$y=f(\boldsymbol{A})$和$g(y)$分别是以矩阵$\boldsymbol{A}$和标量y为变元的实值函数，则

$$\frac{\partial g(f(\boldsymbol{A}))}{\partial \boldsymbol{A}}=\frac{\mathrm{d}g(y)}{\mathrm{d}y}\frac{\partial f(\boldsymbol{A})}{\partial \boldsymbol{A}} \tag{4.12}$$

(6) 若$\boldsymbol{A}\in R^{m\times n},\boldsymbol{x}\in R^{m\times 1},\boldsymbol{y}\in R^{n\times 1}$，则

$$\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}}{\partial \boldsymbol{A}}=\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}} \tag{4.13}$$

(7) 若$\boldsymbol{A}\in R^{n\times n}$非奇异，$\boldsymbol{x}\in R^{n\times 1},\boldsymbol{y}\in R^{n\times 1}$，则

$$\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{y}}{\partial \boldsymbol{A}}=-\boldsymbol{A}^{-\mathrm{T}}\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}^{-\mathrm{T}},\qquad \boldsymbol{A}^{-\mathrm{T}}=(\boldsymbol{A}^{-1})^{\mathrm{T}} \tag{4.14}$$

(8) 若$\boldsymbol{A}\in R^{m\times n},\boldsymbol{x},\boldsymbol{y}\in R^{n\times 1}$，则

$$\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}}{\partial \boldsymbol{A}}=\boldsymbol{A}(\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}+\boldsymbol{y}\boldsymbol{x}^{\mathrm{T}}) \tag{4.15}$$

(9) 若$\boldsymbol{A}\in R^{m\times n},\boldsymbol{x},\boldsymbol{y}\in R^{m\times 1}$，则

$$\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}}{\partial \boldsymbol{A}}=(\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}+\boldsymbol{y}\boldsymbol{x}^{\mathrm{T}})\boldsymbol{A} \tag{4.16}$$

(10) 指数函数的梯度

$$\frac{\partial \exp(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y})}{\partial \boldsymbol{A}}=\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}\exp(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}) \tag{4.17}$$

4.1.4 迹函数的梯度矩阵

(1) 单个矩阵的迹的梯度矩阵

① $\boldsymbol{W}$是$m \times m$ 矩阵时，有

$$\frac{\partial \operatorname{tr}(\boldsymbol{W})}{\partial \boldsymbol{W}} = \boldsymbol{I}_m \tag{4.18}$$

② $m \times m$矩阵$\boldsymbol{W}$可逆时，有

$$\frac{\partial \operatorname{tr}(\boldsymbol{W}^{-1})}{\partial \boldsymbol{W}} = -(\boldsymbol{W}^{-2})^{\mathrm{T}} \tag{4.19}$$

③ 对于两个向量的外积，有

$$\frac{\partial \operatorname{tr}(\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}})}{\partial \boldsymbol{x}} = \frac{\partial \operatorname{tr}(\boldsymbol{y}\boldsymbol{x}^{\mathrm{T}})}{\partial \boldsymbol{x}} = \boldsymbol{y} \tag{4.20}$$

(2) 两个矩阵乘积的迹的梯度

① 若$\boldsymbol{W} \in R^{m\times n}, \boldsymbol{A} \in R^{n\times m}$，则

$$\frac{\partial \operatorname{tr}(\boldsymbol{W}\boldsymbol{A})}{\partial \boldsymbol{W}} = \frac{\partial \operatorname{tr}(\boldsymbol{A}\boldsymbol{W})}{\partial \boldsymbol{W}} = \boldsymbol{A}^{\mathrm{T}} \tag{4.21}$$

特别地，有[3]

$$\frac{\partial \operatorname{tr}(\boldsymbol{W}\boldsymbol{A})}{\partial \boldsymbol{W}} = \frac{\partial \operatorname{tr}(\boldsymbol{A}\boldsymbol{W})}{\partial \boldsymbol{W}} = \boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}} - \operatorname{diag}(\boldsymbol{A}), \quad \boldsymbol{W} \text{ 为对称矩阵} \tag{4.22}$$

注释： 式(4.21)和式(4.22)似乎矛盾，实际上，对称矩阵的元素之间并不独立。按定义计算即可以求得式(4.22)。

② 当$\boldsymbol{W} \in R^{m\times n}, \boldsymbol{A} \in R^{m\times n}$时，有

$$\frac{\partial \operatorname{tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{A})}{\partial \boldsymbol{W}} = \frac{\partial \operatorname{tr}(\boldsymbol{A}\boldsymbol{W}^{\mathrm{T}})}{\partial \boldsymbol{W}} = \boldsymbol{A} \tag{4.23}$$

③ 对于$\boldsymbol{W} \in R^{m\times n}$，有

$$\frac{\partial \operatorname{tr}(\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}})}{\partial \boldsymbol{W}} = \frac{\partial \operatorname{tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W})}{\partial \boldsymbol{W}} = 2\boldsymbol{W} \tag{4.24}$$

④ 令$\boldsymbol{W} \in R^{m\times m}$，则

$$\frac{\partial \operatorname{tr}(\boldsymbol{W}^2)}{\partial \boldsymbol{W}} = \frac{\partial \operatorname{tr}(\boldsymbol{W}\boldsymbol{W})}{\partial \boldsymbol{W}} = 2\boldsymbol{W}^{\mathrm{T}} \tag{4.25}$$

⑤ 若$\boldsymbol{W}, \boldsymbol{A} \in R^{m\times m}$，并且$\boldsymbol{W}$非奇异，则

$$\frac{\partial \operatorname{tr}(\boldsymbol{A}\boldsymbol{W}^{-1})}{\partial \boldsymbol{W}} = -(\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1})^{\mathrm{T}} \tag{4.26}$$

(3) 三个矩阵乘积的迹的梯度

① 若$\boldsymbol{W} \in R^{m\times n}, \boldsymbol{A} \in R^{m\times m}$，则

$$\frac{\partial \operatorname{tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{W})}{\partial \boldsymbol{W}} = (\boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}})\boldsymbol{W} \tag{4.27}$$

特别地，当$\boldsymbol{A}$为对称矩阵时，有$\dfrac{\partial\,\mathrm{tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{W})}{\partial\boldsymbol{W}}=2\boldsymbol{A}\boldsymbol{W}$。

② 令$\boldsymbol{W}\in R^{m\times n},\boldsymbol{A}\in R^{n\times n}$，则

$$\frac{\partial\,\mathrm{tr}(\boldsymbol{W}\boldsymbol{A}\boldsymbol{W}^{\mathrm{T}})}{\partial\boldsymbol{W}}=\boldsymbol{W}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}) \tag{4.28}$$

特别地，当$\boldsymbol{A}$为对称矩阵时，$\dfrac{\partial\,\mathrm{tr}(\boldsymbol{W}\boldsymbol{A}\boldsymbol{W}^{\mathrm{T}})}{\partial\boldsymbol{W}}=2\boldsymbol{W}\boldsymbol{A}$。

③ 当$\boldsymbol{W},\boldsymbol{A},\boldsymbol{B}\in R^{m\times m}$，并且$\boldsymbol{W}$非奇异时，有

$$\frac{\partial\,\mathrm{tr}(\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B})}{\partial\boldsymbol{W}}=-(\boldsymbol{W}^{-1}\boldsymbol{B}\boldsymbol{A}\boldsymbol{W}^{-1})^{\mathrm{T}} \tag{4.29}$$

(4) 四个矩阵乘积的迹的梯度

① $\boldsymbol{W}\in R^{m\times n}$和$\boldsymbol{A}\in R^{p\times m}$时，有

$$\frac{\partial\,\mathrm{tr}(\boldsymbol{A}\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}})}{\partial\boldsymbol{W}}=2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{W} \tag{4.30}$$

② $\boldsymbol{W}\in R^{m\times n}$和$\boldsymbol{A}\in R^{p\times n}$时，有

$$\frac{\partial\,\mathrm{tr}(\boldsymbol{A}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}^{\mathrm{T}})}{\partial\boldsymbol{W}}=2\boldsymbol{W}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} \tag{4.31}$$

③ $\boldsymbol{W}\in R^{m\times n},\boldsymbol{A}\in R^{p\times m},\boldsymbol{B}\in R^{m\times p}$时，有

$$\frac{\partial\,\mathrm{tr}(\boldsymbol{A}\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{B})}{\partial\boldsymbol{W}}=(\boldsymbol{B}\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}})\boldsymbol{W} \tag{4.32}$$

④ $\boldsymbol{W}\in R^{m\times n},\boldsymbol{A}\in R^{p\times n},\boldsymbol{B}\in R^{n\times p}$时，有

$$\frac{\partial\,\mathrm{tr}(\boldsymbol{A}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{B})}{\partial\boldsymbol{W}}=\boldsymbol{W}(\boldsymbol{B}\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}) \tag{4.33}$$

⑤ $\boldsymbol{W}\in R^{m\times n},\boldsymbol{A}\in R^{n\times n},\boldsymbol{B}\in R^{m\times m}$时，有

$$\frac{\partial\,\mathrm{tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{B})}{\partial\boldsymbol{W}}=\boldsymbol{B}\boldsymbol{W}\boldsymbol{A}+\boldsymbol{B}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}^{\mathrm{T}} \tag{4.34}$$

⑥ $\boldsymbol{W}\in R^{m\times n}$和$\boldsymbol{A},\boldsymbol{B}\in R^{n\times m}$时，有

$$\frac{\partial\,\mathrm{tr}(\boldsymbol{W}\boldsymbol{A}\boldsymbol{W}\boldsymbol{B})}{\partial\boldsymbol{W}}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}} \tag{4.35}$$

4.1.5 行列式的梯度矩阵

(1) 单个非奇异矩阵的行列式的梯度

$$\begin{aligned}\frac{\partial|\boldsymbol{W}|}{\partial\boldsymbol{W}}&=|\boldsymbol{W}|(\boldsymbol{W}^{-1})^{\mathrm{T}}=(\boldsymbol{W}^{\#})^{\mathrm{T}}\\\frac{\partial|\boldsymbol{W}^{-1}|}{\partial\boldsymbol{W}}&=-|\boldsymbol{W}|^{-1}(\boldsymbol{W}^{-1})^{\mathrm{T}}\end{aligned}\tag{4.36}$$

式中，$\boldsymbol{W}^{\#}$是矩阵$\boldsymbol{A}$的伴随矩阵。

(2) 行列式对数的梯度

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{W}}\log|\boldsymbol{W}|&=\frac{1}{|\boldsymbol{W}|}\frac{\partial|\boldsymbol{W}|}{\partial\boldsymbol{W}},\quad \boldsymbol{W}\text{ 非奇异}\\&=(\boldsymbol{W}^{-1})^{\mathrm{T}},\quad \boldsymbol{W}\text{ 的元素相互独立}\\&=2\boldsymbol{W}^{-1}-\mathrm{diag}(\boldsymbol{W}^{-1}),\quad \text{若 }\boldsymbol{W}\text{ 为对称矩阵}\end{aligned}\tag{4.37}$$

(3) 两个矩阵乘积的行列式的梯度

$$\begin{aligned}\frac{\partial|\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}|}{\partial\boldsymbol{W}}&=2|\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}|(\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W},\quad \mathrm{rank}(\boldsymbol{W}_{m\times n})=m\\\frac{\partial|\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}|}{\partial\boldsymbol{W}}&=2|\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}|\boldsymbol{W}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W})^{-1},\quad \mathrm{rank}(\boldsymbol{W}_{m\times n})=n\\\frac{\partial|\boldsymbol{W}^{2}|}{\partial\boldsymbol{W}}&=2|\boldsymbol{W}|^{2}(\boldsymbol{W}^{-1})^{\mathrm{T}},\quad \mathrm{rank}(\boldsymbol{W}_{m\times m})=m\end{aligned}\tag{4.38}$$

(4) 三个矩阵乘积的行列式的梯度

$$\begin{aligned}\frac{\partial|\boldsymbol{A}\boldsymbol{W}\boldsymbol{B}|}{\partial\boldsymbol{W}}&=|\boldsymbol{A}\boldsymbol{W}\boldsymbol{B}|\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}})^{-1}\boldsymbol{B}^{\mathrm{T}}\\\frac{\partial|\boldsymbol{W}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{W}|}{\partial\boldsymbol{W}}&=2\boldsymbol{A}\boldsymbol{W}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{W})^{-1},\quad |\boldsymbol{W}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{W}|>0\\\frac{\partial|\boldsymbol{W}\boldsymbol{A}\boldsymbol{W}^{\mathrm{T}}|}{\partial\boldsymbol{W}}&=[(\boldsymbol{W}\boldsymbol{A}\boldsymbol{W}^{\mathrm{T}})^{-1}]^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A})\\&=2(\boldsymbol{W}\boldsymbol{A}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}\boldsymbol{A},\quad \boldsymbol{A}\text{ 为对称矩阵}\end{aligned}\tag{4.39}$$

4.1.6 矩阵微分

(1) 常数矩阵的微分矩阵为零矩阵，即

$$\mathrm{d}\boldsymbol{A}=\boldsymbol{O}\tag{4.40}$$

(2) 常数α与矩阵函数$\boldsymbol{U}$的乘积的微分矩阵，即

$$\mathrm{d}(\alpha\boldsymbol{U})=\alpha\mathrm{d}\boldsymbol{U}\tag{4.41}$$

(3) 矩阵转置的微分矩阵等于原矩阵的微分矩阵的转置，即

$$\mathrm{d}(\boldsymbol{U}^{\mathrm{T}})=(\mathrm{d}\boldsymbol{U})^{\mathrm{T}}\tag{4.42}$$

(4) 两个矩阵函数的和(差)的微分矩阵为

$$\mathrm{d}(\boldsymbol{U} \pm \boldsymbol{V}) = \mathrm{d}\boldsymbol{U} \pm \mathrm{d}\boldsymbol{V} \tag{4.43}$$

(5) 常数矩阵与矩阵函数乘积的微分矩阵为

$$\mathrm{d}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}) = \boldsymbol{A}(\mathrm{d}\boldsymbol{X})\boldsymbol{B} \tag{4.44}$$

(6) 矩阵函数乘积的微分矩阵为

$$\mathrm{d}(\boldsymbol{U}\boldsymbol{V}) = (\mathrm{d}\boldsymbol{U})\boldsymbol{V} + \boldsymbol{U}(\mathrm{d}\boldsymbol{V}) \tag{4.45}$$

$$\mathrm{d}(\boldsymbol{U}\boldsymbol{V}\boldsymbol{W}) = (\mathrm{d}\boldsymbol{U})\boldsymbol{V}\boldsymbol{W} + \boldsymbol{U}(\mathrm{d}\boldsymbol{V})\boldsymbol{W} + \boldsymbol{U}\boldsymbol{V}(\mathrm{d}\boldsymbol{W}) \tag{4.46}$$

特别地，若$\boldsymbol{A}$为常数矩阵，则

$$\mathrm{d}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}) = (\mathrm{d}\boldsymbol{X})\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}} + \boldsymbol{X}\boldsymbol{A}(\mathrm{d}\boldsymbol{X})^{\mathrm{T}} \tag{4.47}$$

$$\mathrm{d}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}) = (\mathrm{d}\boldsymbol{X})^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{X} \tag{4.48}$$

(7) 矩阵函数的Kronecker积的微分矩阵为

$$\mathrm{d}(\boldsymbol{U} \otimes \boldsymbol{V}) = (\mathrm{d}\boldsymbol{U}) \otimes \boldsymbol{V} + \boldsymbol{U} \otimes \mathrm{d}\boldsymbol{V} \tag{4.49}$$

(8) 矩阵函数的Hadamard积的微分矩阵，为

$$\mathrm{d}(\boldsymbol{U} \odot \boldsymbol{V}) = (\mathrm{d}\boldsymbol{U}) \odot \boldsymbol{V} + \boldsymbol{U} \odot \mathrm{d}\boldsymbol{V} \tag{4.50}$$

(9) 向量化函数vec($\boldsymbol{U}$)的微分矩阵等于$\boldsymbol{U}$的微分矩阵的向量化函数，即

$$\mathrm{d}(\mathrm{vec}(\boldsymbol{U})) = \mathrm{vec}(\mathrm{d}\boldsymbol{U}) \tag{4.51}$$

(10) 行列式的微分为

$$\mathrm{d}|\boldsymbol{X}| = |\boldsymbol{X}|\mathrm{tr}(\boldsymbol{X}^{-1}\mathrm{d}\boldsymbol{X}) \tag{4.52}$$

(11) 矩阵$\boldsymbol{U}$的迹的微分d(tr$\boldsymbol{U}$)等于微分矩阵d$\boldsymbol{U}$的迹tr(d$\boldsymbol{U}$)，即有

$$\mathrm{d}(\mathrm{tr}(\boldsymbol{U})) = \mathrm{tr}(\mathrm{d}\boldsymbol{U}) \tag{4.53}$$

$$\mathrm{d}(\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})) = 2\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}) \tag{4.54}$$

(12) 逆矩阵的微分矩阵为

$$\mathrm{d}(\boldsymbol{X}^{-1}) = -\boldsymbol{X}^{-1}(\mathrm{d}\boldsymbol{X})\boldsymbol{X}^{-1} \tag{4.55}$$

(13) Moore-Penrose逆矩阵的微分矩阵为

$$\begin{aligned}\mathrm{d}(\boldsymbol{X}^{\dagger}) = &-\boldsymbol{X}^{\dagger}(\mathrm{d}\boldsymbol{X})\boldsymbol{X}^{\dagger} + \boldsymbol{X}^{\dagger}(\boldsymbol{X}^{\dagger})^{\mathrm{T}}(\mathrm{d}\boldsymbol{X}^{\mathrm{T}})(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{\dagger}) + \\ &(\boldsymbol{I} - \boldsymbol{X}^{\dagger}\boldsymbol{X})(\mathrm{d}\boldsymbol{X}^{\mathrm{T}})(\boldsymbol{X}^{\dagger})^{\mathrm{T}}\boldsymbol{X}^{\dagger}\end{aligned} \tag{4.56}$$

$$\mathrm{d}(\boldsymbol{X}^{\dagger}\boldsymbol{X}) = \boldsymbol{X}^{\dagger}(\mathrm{d}\boldsymbol{X})(\boldsymbol{I} - \boldsymbol{X}^{\dagger}\boldsymbol{X}) + \left(\boldsymbol{X}^{\dagger}(\mathrm{d}\boldsymbol{X})(\boldsymbol{I} - \boldsymbol{X}^{\dagger}\boldsymbol{X})\right)^{\mathrm{T}} \tag{4.57}$$

$$\mathrm{d}(\boldsymbol{X}\boldsymbol{X}^{\dagger}) = (\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{\dagger})(\mathrm{d}\boldsymbol{X})\boldsymbol{X}^{\dagger} + \left((\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{\dagger})(\mathrm{d}\boldsymbol{X})\boldsymbol{X}^{\dagger}\right)^{\mathrm{T}} \tag{4.58}$$

(14) 矩阵对数的微分矩阵为

$$\mathrm{d}\log \boldsymbol{X} = \boldsymbol{X}^{-1}\mathrm{d}\boldsymbol{X} \tag{4.59}$$

$$\mathrm{d}\log(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}) = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}[(\mathrm{d}\boldsymbol{X})^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{X}] \tag{4.60}$$

微分矩阵和梯度是两个不同的概念，一般来讲，我们可以通过求解函数的微分矩阵来得到对应的梯度矩阵。表4.1.1和表4.1.2分别给出了迹函数和行列式函数的微分矩阵与梯度矩阵的对应关系。

表4.1.1 几种迹函数的微分矩阵与梯度矩阵的对应关系[1]

迹函数$f(\boldsymbol{X})$	微分矩阵$\mathrm{d}f(\boldsymbol{X})$	梯度矩阵$\partial f(\boldsymbol{X})/\partial \boldsymbol{X}$
$\mathrm{tr}(\boldsymbol{X})$	$\mathrm{tr}(\boldsymbol{I}\mathrm{d}\boldsymbol{X})$	$\boldsymbol{I}$
$\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})$	$2\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X})$	$2\boldsymbol{X}$
$\mathrm{tr}(\boldsymbol{A}\boldsymbol{X})$	$\mathrm{tr}(\boldsymbol{A}\mathrm{d}\boldsymbol{X})$	$\boldsymbol{A}^{\mathrm{T}}$
$\mathrm{tr}(\boldsymbol{X}^2)$	$2\mathrm{tr}(\boldsymbol{X}\mathrm{d}\boldsymbol{X})$	$2\boldsymbol{X}^{\mathrm{T}}$
$\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})$	$\mathrm{tr}\left(\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})\mathrm{d}\boldsymbol{X}\right)$	$(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})\boldsymbol{X}$
$\mathrm{tr}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})$	$\mathrm{tr}\left((\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right)$	$\boldsymbol{X}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})$
$\mathrm{tr}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X})$	$\mathrm{tr}\left((\boldsymbol{A}\boldsymbol{X}+\boldsymbol{X}\boldsymbol{A})\mathrm{d}\boldsymbol{X}\right)$	$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}$
$\mathrm{tr}(\boldsymbol{A}\boldsymbol{X}^{-1})$	$-\mathrm{tr}\left(\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}^{-1}\mathrm{d}\boldsymbol{X}\right)$	$-(\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}^{-1})^{\mathrm{T}}$
$\mathrm{tr}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}\boldsymbol{B})$	$\mathrm{tr}((\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}+\boldsymbol{B}\boldsymbol{X}\boldsymbol{A})\mathrm{d}\boldsymbol{X})$	$(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}+\boldsymbol{B}\boldsymbol{X}\boldsymbol{A})^{\mathrm{T}}$
$\mathrm{tr}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{B})$	$\mathrm{tr}((\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{B}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}})\mathrm{d}\boldsymbol{X})$	$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{B}\boldsymbol{X}\boldsymbol{A}$

表4.1.2 几种行列式函数的微分矩阵与梯度矩阵的对应关系[1]

行列式$f(\boldsymbol{X})$	微分矩阵$\mathrm{d}f(\boldsymbol{X})$	梯度矩阵$\partial f(\boldsymbol{X})/\partial \boldsymbol{X}$
$\lvert\boldsymbol{X}\rvert$	$\lvert\boldsymbol{X}\rvert\,\mathrm{tr}(\boldsymbol{X}^{-1}\mathrm{d}\boldsymbol{X})$	$\lvert\boldsymbol{X}\rvert\boldsymbol{X}^{-\mathrm{T}}$
$\lvert\boldsymbol{X}^2\rvert$	$2\lvert\boldsymbol{X}\rvert^2\mathrm{tr}\left(\boldsymbol{X}^{-1}\mathrm{d}\boldsymbol{X}\right)$	$2\lvert\boldsymbol{X}\rvert^2\boldsymbol{X}^{-\mathrm{T}}$
$\lvert\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rvert$	$2\lvert\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rvert\,\mathrm{tr}\left(\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})^{-1}\mathrm{d}\boldsymbol{X}\right)$	$2\lvert\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rvert(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}$
$\lvert\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\rvert$	$2\lvert\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\rvert\,\mathrm{tr}\left((\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right)$	$2\lvert\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\rvert\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}$
$\lvert\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}\rvert$	$\lvert\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}\rvert\,\mathrm{tr}\left(\boldsymbol{B}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B})^{-1}\boldsymbol{A}\mathrm{d}\boldsymbol{X}\right)$	$\lvert\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}\rvert\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}})^{-1}\boldsymbol{B}^{\mathrm{T}}$

4.1.7 标量函数的共轭梯度

(1) 若$f(\boldsymbol{x}) = c$为常数，则共轭梯度$\dfrac{\partial c}{\partial \boldsymbol{x}^*} = \boldsymbol{0}$。

(2) 线性法则：若$f(\boldsymbol{x})$和$g(\boldsymbol{x})$分别是向量$\boldsymbol{x}$的实值函数，c_1和c_2为复常数，则

$$\frac{\partial[c_1 f(\boldsymbol{x}) + c_2 g(\boldsymbol{x})]}{\partial \boldsymbol{x}^*} = c_1\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^*} + c_2\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}^*} \tag{4.61}$$

(3) 乘积法则：

① 若$f(\boldsymbol{x})$和$g(\boldsymbol{x})$ 都是向量$\boldsymbol{x}$的实值函数，则

$$\frac{\partial f(\boldsymbol{x})g(\boldsymbol{x})}{\partial \boldsymbol{x}^*} = g(\boldsymbol{x})\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^*} + f(\boldsymbol{x})\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}^*} \tag{4.62}$$

② 若$f(\boldsymbol{x}), g(\boldsymbol{x})$和$h(\boldsymbol{x})$都是向量$\boldsymbol{x}$ 的实值函数，则

$$\frac{\partial f(\boldsymbol{x})g(\boldsymbol{x})h(\boldsymbol{x})}{\partial \boldsymbol{x}^*} = g(\boldsymbol{x})h(\boldsymbol{x})\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^*} + f(\boldsymbol{x})h(\boldsymbol{x})\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}^*} + \quad f(\boldsymbol{x})g(\boldsymbol{x})\frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}^*} \tag{4.63}$$

(4) 商法则：若$g(\boldsymbol{x}) \neq 0$，则

$$\frac{\partial f(\boldsymbol{x})/g(\boldsymbol{x})}{\partial \boldsymbol{x}^*} = \frac{1}{g^2(\boldsymbol{x})}\left[g(\boldsymbol{x})\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^*} - f(\boldsymbol{x})\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}^*}\right] \tag{4.64}$$

(5) 链式法则：若$\boldsymbol{y}(\boldsymbol{x})$ 是$\boldsymbol{x}$的复向量值函数，则

$$\frac{\partial f(\boldsymbol{y}(\boldsymbol{x}))}{\partial \boldsymbol{x}^*} = \frac{\partial\,[\boldsymbol{y}(\boldsymbol{x})]^{\mathrm{T}}}{\partial \boldsymbol{x}^*}\frac{\partial f(\boldsymbol{y})}{\partial\,\boldsymbol{y}} \tag{4.65}$$

式中，$\dfrac{\partial[\boldsymbol{y}(\boldsymbol{x})]^{\mathrm{T}}}{\partial \boldsymbol{x}^*}$为$n \times n$矩阵。

(6) 若$n \times 1$向量为$\boldsymbol{a}$与$\boldsymbol{x}$无关的常数向量，则

$$\frac{\partial \boldsymbol{a}^{\mathrm{H}}\boldsymbol{x}}{\partial \boldsymbol{x}^*} = \boldsymbol{0}, \qquad \frac{\partial \boldsymbol{x}^{\mathrm{H}}\boldsymbol{a}}{\partial \boldsymbol{x}^*} = \boldsymbol{a} \tag{4.66}$$

(7) 令$\boldsymbol{A}$是一个与向量$\boldsymbol{x}$ 无关的矩阵，则

$$\begin{aligned} \frac{\partial \boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} &= \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}^*, \qquad & \frac{\partial \boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}^*} &= \boldsymbol{A}\boldsymbol{x} \\ \frac{\partial \boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{y}}{\partial \boldsymbol{A}} &= \boldsymbol{x}^*\boldsymbol{y}^{\mathrm{T}}, \qquad & \frac{\partial \boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{A}} &= \boldsymbol{x}^*\boldsymbol{x}^{\mathrm{T}} \end{aligned} \tag{4.67}$$

4.1.8 迹函数的共轭梯度

$$\begin{gathered} \frac{\partial \mathrm{tr}(\boldsymbol{y}\boldsymbol{x}^{\mathrm{H}})}{\partial \boldsymbol{x}^*} = \frac{\partial \mathrm{tr}(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y})}{\partial \boldsymbol{x}^*} = \boldsymbol{y} \\ \frac{\partial \mathrm{tr}(\boldsymbol{B}\boldsymbol{A}^{\mathrm{H}})}{\partial \boldsymbol{A}^*} = \frac{\partial \mathrm{tr}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{B})}{\partial \boldsymbol{A}^*} = \boldsymbol{B} \\ \frac{\partial \mathrm{tr}(\boldsymbol{A}^{\mathrm{H}})}{\partial \boldsymbol{A}^*} = \boldsymbol{I}, \quad \frac{\partial \mathrm{tr}(\boldsymbol{A})}{\partial \boldsymbol{A}^*} = \boldsymbol{O} \\ \frac{\partial \mathrm{tr}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{W}\boldsymbol{A})}{\partial \boldsymbol{A}^*} = \boldsymbol{W}\boldsymbol{A}, \quad \frac{\partial \mathrm{tr}(\boldsymbol{A}\boldsymbol{W}\boldsymbol{A}^{\mathrm{H}})}{\partial \boldsymbol{A}^*} = \boldsymbol{A}\boldsymbol{W} \end{gathered} \tag{4.68}$$

4.2 习题与解答

4.1 令$\boldsymbol{y}$是一实值随机向量，由$\boldsymbol{y}=\alpha\boldsymbol{x}+\boldsymbol{e}$ 描述。其中，$\boldsymbol{x}$是一个确定向量；α为常数；$\boldsymbol{e}$为一零均值的随机向量，其协方差矩阵为$\boldsymbol{R}_e$。现在希望利用数据向量$\boldsymbol{y}$，确定α的最优线性无偏估计。为此，令$\hat{\alpha}=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{y}$，求最优滤波器$\boldsymbol{w}$。

解：

$$\hat{\alpha}=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{y}=\hat{\alpha}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{w}^{\mathrm{T}}\boldsymbol{e}$$

显然若要$\hat{\alpha}$为无偏估计，则必须约束$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}=1$。线性估计的方差可以写为

$$\sigma_{\hat{\alpha}}^2=\mathrm{E}\left[\boldsymbol{w}^{\mathrm{T}}\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}\boldsymbol{w}\right]=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_e\boldsymbol{w}$$

欲最小化估计的方差，则以上问题等价于以下最优化问题：

$$\min\sigma_{\hat{\alpha}}^2=\min\boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_e\boldsymbol{w}$$

$$s.t.\quad \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}=1$$

构造Lagrange函数如下：

$$J(\boldsymbol{w},\lambda)=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_e\boldsymbol{w}+2\lambda(1-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x})$$

该函数分别关于$\boldsymbol{w}$和λ求导得到以下两方程：

$$\begin{aligned}\boldsymbol{R}_e\boldsymbol{w}&=\lambda\boldsymbol{x}\\ \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}&=1\end{aligned}\tag{4.69}$$

由协方差矩阵$\boldsymbol{R}_e$ 是半正定性有以下两种情况：

(1) 若$\boldsymbol{R}_e$满秩，则由式(4.69)有

$$\boldsymbol{w}=\lambda\boldsymbol{R}_e^{-1}\boldsymbol{x}$$

代入约束条件有

$$\lambda=\frac{1}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{R}_e^{-1}\boldsymbol{x}}$$

于是最优滤波器为

$$\boldsymbol{w}=\frac{\boldsymbol{R}_e^{-1}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{R}_e^{-1}\boldsymbol{x}}$$

最优估计方差为

$$\sigma_{\hat{\alpha}}^2=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_e\boldsymbol{w}=\frac{1}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{R}_e^{-1}\boldsymbol{x}}$$

(2) 若$\boldsymbol{R}_e$非满秩，则必有一个零特征值对应的特征向量，将这个向量放大一个合适比例必能使得约束条件成立。即，当$\boldsymbol{R}_e$奇异时，最优滤波器为噪声方差零特征值对应的特征向量，最优线性估计的方差为零。

$\boldsymbol{R}_e$非满秩就可以使得估计准确无误，似乎有些困惑。举一简单例子来说明问题。

$$\begin{bmatrix} y_1 \\ \vdots \\ y_{n-1} \\ y_n \end{bmatrix} = \alpha \begin{bmatrix} x_1 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_{n-1} \\ 0 \end{bmatrix}$$

其中$e_1, \cdots, e_{n-1}$是服从独立的标准正态分布。

由于$y_n = \alpha x_n$，即y_n是不含噪声的α的线性函数，则$\boldsymbol{w}$可以取为$\boldsymbol{w} = \left[0, \cdots, 0, \dfrac{1}{x_n}\right]$即可以准确无误估计出$\alpha$。同时我们可以验证$\boldsymbol{w}$是$\boldsymbol{R}_e$零特征值对应的特征向量。不难有

$$\boldsymbol{R}_e = \mathrm{E}\left[\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}\right] = \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & 0 \end{bmatrix}$$

很显然，上述$\boldsymbol{w}$是$\boldsymbol{R}_e$零特征值对应的特征向量。所以$\boldsymbol{R}_e$非满秩就可以使得估计准确无误也不足为奇。

4.2 令$f(t)$为一已知函数。考虑二次型函数的最小化：

$$\min Q(x) = \min \int_0^1 [f(t) - x_0 - x_1 t - \cdots - x_n t^n]^2 \mathrm{d}t$$

判断与线性方程组对应的矩阵是否病态?

思路分析： 这是一个用多项式拟合曲线的问题，多项式的系数是我们要求的。

解： 最小化$Q(x)$的必要条件是使得

$$\frac{\partial Q(x)}{\partial x_k} = 0, \quad k = 0, 1, \cdots, n \tag{4.70}$$

由

$$\frac{\partial Q(x)}{\partial x_k} = \int_0^1 [f(t) - x_0 - x_1 t - \cdots - x_n t^n] t^k \mathrm{d}t = 0$$

有

$$\int_0^1 f(t) t^k \mathrm{d}t = \sum_{j=0}^{n} x_j \int_0^1 t^{k+j} \mathrm{d}t = \sum_{j=0}^{n} x_j \frac{1}{k+j+1}$$

记

$$b_k = \int_0^1 f(t) t^k \mathrm{d}t$$

则式(4.70)可以表示成下面方程组

$$\sum_{j=0}^{n} x_j \frac{1}{k+j+1} = b_k, \quad k = 0, 1, \cdots, n$$

写成矩阵形式为

$$\boldsymbol{H}_n \boldsymbol{x} = \boldsymbol{b} \tag{4.71}$$

其中

$$\boldsymbol{H}_n = \left(\frac{1}{k+j+1}\right)_{k,j=0}^{n} = \begin{bmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n+1} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n+2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n+1} & \frac{1}{n+2} & \cdots & \frac{1}{2n+1} \end{bmatrix}$$

$$\boldsymbol{b} = [b_0, b_1 \cdots, b_n]^{\mathrm{T}}$$

$\boldsymbol{H}_n$称为Hilbert**矩阵**。它是严重病态的，条件数随n增加而迅速增大。可以通过Matlab数值计算当$n=2$时，Cond($\boldsymbol{H}_n$)=524，当$n=2$时，Cond($\boldsymbol{H}_n$)=1.495×10^7。条件数态大，方程(4.71)对舍入误差非常敏感，以至于数值解和精确解差别很大，使得曲线拟合失败。

4.3 考虑方程$\boldsymbol{y}=\boldsymbol{A\theta}+\boldsymbol{e}$，其中，$\boldsymbol{e}$ 为误差向量。定义加权误差平方和

$$E_w \stackrel{\mathrm{def}}{=} \boldsymbol{e}^{\mathrm{H}}\boldsymbol{W}\boldsymbol{e}$$

其中，$\boldsymbol{W}$为一对称正定矩阵，它对误差起加权作用。

(1) 求使E_w最小化的参数向量$\boldsymbol{\theta}$的解。这一解称为$\boldsymbol{\theta}$的加权最小二乘估计。

(2) 证明加权最小二乘准则相当于使误差或数据向量进行预白化。

解：(1)

$$E_w = \boldsymbol{e}^{\mathrm{H}}\boldsymbol{W}\boldsymbol{e} = (\boldsymbol{y}-\boldsymbol{A\theta})^{\mathrm{H}}\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{A\theta})$$

参考教材求共轭梯度公式(5.3.36)有

$$\frac{\partial E_w}{\partial \boldsymbol{\theta}^*} = \boldsymbol{A}^{\mathrm{H}}\boldsymbol{W}\boldsymbol{A}\boldsymbol{\theta} - \boldsymbol{A}^{\mathrm{H}}\boldsymbol{W}\boldsymbol{y}$$

令共轭梯度为零求得$\boldsymbol{\theta}$的加权最小二乘估计为

$$\hat{\theta}_{\mathrm{LS}} = (\boldsymbol{A}^{\mathrm{H}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{W}\boldsymbol{y}$$

(2) 假设误差向量的方差矩阵为$\sigma^2\boldsymbol{V}$，由$\boldsymbol{V}$的正定性可将其进行Cholesky分解为$\boldsymbol{V}=\boldsymbol{P}\boldsymbol{P}^{\mathrm{T}}$。令$\boldsymbol{\epsilon}=\boldsymbol{P}^{-1}\boldsymbol{e}$，$\boldsymbol{z}=\boldsymbol{P}^{-1}\boldsymbol{y}$有

$$\boldsymbol{z} = \boldsymbol{P}^{-1}\boldsymbol{A\theta} + \boldsymbol{\epsilon} = \boldsymbol{B\theta} + \boldsymbol{\epsilon} \tag{4.72}$$

由于

$$\mathrm{var}(\boldsymbol{\epsilon}) = \mathrm{var}(\boldsymbol{P}^{-1}\boldsymbol{e}) = \boldsymbol{P}^{-1}\mathrm{var}(\boldsymbol{e})\boldsymbol{P}^{-T} = \boldsymbol{P}^{-1}\sigma^2\boldsymbol{P}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}^{-T} = \sigma^2\boldsymbol{I}$$

即$\boldsymbol{\epsilon}$是白噪声。则式(4.72)的最小二乘估计为

$$\begin{aligned}\hat{\theta}_{\mathrm{LS}} &= (\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{z} = (\boldsymbol{A}^{\mathrm{H}}\boldsymbol{P}^{-\mathrm{T}}\boldsymbol{P}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{P}^{-\mathrm{T}}\boldsymbol{P}^{-1}\boldsymbol{z}\\ &= (\boldsymbol{A}^{\mathrm{H}}\boldsymbol{V}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{V}^{-1}\boldsymbol{z}\end{aligned}$$

则加权最小二乘$\boldsymbol{W} = \boldsymbol{V}^{-1}$，并且等价于把误差向量$\boldsymbol{e}$白化。

4.4 考虑本书题4.1中$\boldsymbol{y}$为复向量的情况。若令代价函数为

$$f(\boldsymbol{w}) = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{R}_e\boldsymbol{w}$$

并且给滤波器加约束条件$\mathrm{Re}(\boldsymbol{w}^{\mathrm{H}}\boldsymbol{x}) = b$，其中，$b$为一常数。试求最优滤波器$\boldsymbol{w}$。

思路分析： 本题涉及到实标量函数关于复向量的求导，可以参考教材P286~289，其解题思路完全同题4.1。

解： 约束条件$\mathrm{Re}(\boldsymbol{w}^{\mathrm{H}}\boldsymbol{x}) = \boldsymbol{b}$ 可以表示转换为

$$\boldsymbol{w}^{\mathrm{H}}\boldsymbol{x} + \boldsymbol{x}^{\mathrm{H}}\boldsymbol{w} = 2b$$

构造Lagrange函数如下：

$$J(\boldsymbol{w}, \lambda) = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{R}_e\boldsymbol{w} + \lambda(2b - \boldsymbol{w}^{\mathrm{H}}\boldsymbol{x} - \boldsymbol{x}^{\mathrm{H}}\boldsymbol{w})$$

该函数分别关于$\boldsymbol{w}$和λ求导，由协方差矩阵都是实矩阵的性质有以下两方程：

$$\begin{aligned}\boldsymbol{R}_e\boldsymbol{w} &= \lambda\boldsymbol{x}\\ \boldsymbol{w}^{\mathrm{H}}\boldsymbol{x} + \boldsymbol{x}^{\mathrm{H}}\boldsymbol{w} &= 2b\end{aligned}$$

同题4.1有以下两种情况：

(1) 当$\boldsymbol{R}_e$满秩时，

$$\boldsymbol{w} = \lambda\boldsymbol{R}_e^{-1}\boldsymbol{x}$$

代入约束条件并利用$\boldsymbol{R}_e^{-1} = (\boldsymbol{R}_e^{-1})^{\mathrm{H}}$有

$$\lambda\boldsymbol{x}^{\mathrm{H}}\boldsymbol{R}_e^{-1}\boldsymbol{x} + \lambda\boldsymbol{x}^{\mathrm{H}}\boldsymbol{R}_e^{-1}\boldsymbol{x} = 2b$$

则

$$\lambda = \frac{1}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{R}_e^{-1}\boldsymbol{x}}$$

于是最优滤波器为

$$\boldsymbol{w} = \frac{\boldsymbol{R}_e^{-1}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{R}_e^{-1}\boldsymbol{x}}$$

最优估计方差为

$$\sigma_{\hat{\alpha}}^2 = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{R}_e\boldsymbol{w} = \frac{1}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{R}_e^{-1}\boldsymbol{x}}$$

(2) 当$\boldsymbol{R}_c$奇异时，结论也同题4.1，这里不再赘述。

4.5 解释下列有约束最优化问题是否有解：

(1) $\min(x_1+x_2)$，约束条件为$x_1^2+x_2^2=2, 0\leqslant x_1\leqslant 1, 0\leqslant x_2\leqslant 1$；

(2) $\min(x_1+x_2)$，约束条件为$x_1^2+x_2^2\leqslant 1, x_1+x_2=4$；

(3) $\min(x_1x_2)$，约束条件为$x_1+x_2=3$。

解法一： 上面三个小题都可以通过做图很直观地找到答案，这里从略。

解法二：

(1) 如果$0\leqslant x_1<1$, $\|x_2\|>1$，则约束条件$0\leqslant x_2\leqslant 1$ 不可能满足。所以，当且仅当$x_1=1$ 满足约束条件，且$x_1=1, x_2=1$ 为问题的最优解。

(2) 从$x_1^2+x_2^2=1$ 有$\|x_1\|\leqslant 1, \|x_2\|\leqslant 1$。因此$-2\leqslant x_1+x_2\leqslant 2$，然而与$x_1+x_2=4$矛盾，所以该问题无解。

(3) Kuhn-Tucker 点：

$$\nabla f(x^*)=\sum_i \nabla g_i(x^*)\lambda_i+\sum_j \nabla h_j(x^*)\mu_j$$

$$\lambda_i g_i(x^*)=0, \quad \lambda_i\geqslant 0$$

针对本题，以上两个方程可以写为

$$x_2=\mu$$

$$x_1=\mu$$

从而有$x_1=x_2=\mu=\dfrac{3}{2}$。这是该问题唯一的KKT点。

令$L(x_1,x_2,\mu)=x_1x_2-\mu(x_1+x_2-3)$ 有

$$\nabla_{\mathbf{x}}^2 L(x_1,x_2,\mu)=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$$

$$\mathrm{d}^{\mathrm{T}}\nabla_{\mathbf{x}}^2 L(x_1,x_2,\mu)\mathrm{d}<0, \quad \forall \mathrm{d}\in T(x_1,x_2,\mu)$$

所以$\left(\dfrac{3}{2},\dfrac{3}{2}\right)$ 是唯一的KKT点,但不是该问题的最优解。所以该问题无最优解。

4.6 考虑约束优化问题

$$\min(x-1)(y+1), \qquad \text{约束条件为 } x-y=0$$

利用Lagrange乘数法证明极小点为$(1,1)$，且Lagrange乘数$\lambda=1$。若Lagrange函数取

$$\psi(x,y)=(x-1)(y+1)-1(x-y)$$

证明$\phi(x,y)$在$(0,0)$有一个鞍点，即点$(0,0)$不能使$\psi(x,y)$ 极小化。

证明：令

$$L(x,y,\mu)=(x-1)(y+1)-\mu(x-y)$$

KKT 点满足

$$\frac{\partial L(x,y,\mu)}{\partial x}=(y+1)-\mu=0$$
$$\frac{\partial L(x,y,\mu)}{\partial y}=(x-1)+\mu=0$$

以上方程的解为$(x,y)=(0,0)$。且

$$\nabla_{\mathbf{x}}^2 L(x,y,\mu)=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$$

可以得到

$$\mathrm{d}^{\mathrm{T}}\nabla_{\mathbf{x}}^2 L(x,y,\mu)\mathrm{d}>0,\quad \forall 0\neq \mathrm{d}\in T(x,y,\mu)$$

所以$(x,y)=(0,0)$ 是局部最小点。

4.7 求解约束优化问题$\min J(x,y,z)=x^2+y^2+z^2$，约束条件为$3x+4y-z=25$。

解：KKT 满足下列方程组：

$$2x=3\mu$$
$$2y=4\mu$$
$$2z=-\mu$$
$$3x+4y-z=25$$

该方程组的解为$(x,y,z)=\left(\frac{75}{26},\frac{50}{13},-\frac{25}{26}\right)$。因为$f(x,y,z)$ 是凸函数，约束$h(x)$ 是线性函数，所以KKT点是全局最优点。因此$(x,y,z)=\left(\frac{75}{26},\frac{50}{13},-\frac{25}{26}\right)$ 是全局极小点。

4.8 证明

$$\mathrm{d}(\boldsymbol{UVW})=(\mathrm{d}\boldsymbol{U})\boldsymbol{VW}+\boldsymbol{U}(\mathrm{d}\boldsymbol{V})\boldsymbol{W}+\boldsymbol{UV}(\mathrm{d}\boldsymbol{W})$$

思路分析：同微积分中推导多个函数乘积微分公式的思路一样，我们先得到两个矩阵乘积的微分公式：$\mathrm{d}(\boldsymbol{UV})$。

证明：令

$$\boldsymbol{P}=\boldsymbol{UV}$$

则

$$p_{ij}=\sum_k u_{ik}v_{kj}$$

由标量微分法则有

$$\mathrm{d}p_{ij} = \sum_k (\mathrm{d}u_{ik})\, v_{kj} + \sum_k u_{ik}(\mathrm{d}v_{kj})$$

所以

$$\mathrm{d}\boldsymbol{P} = (\mathrm{d}\boldsymbol{U})\boldsymbol{V} + \boldsymbol{U}(\mathrm{d}\boldsymbol{V})$$

则有

$$\begin{aligned}\mathrm{d}(\boldsymbol{UVW}) &= (\mathrm{d}(\boldsymbol{UV}))\boldsymbol{W} + (\boldsymbol{UV})\mathrm{d}\boldsymbol{W}\\ &= [(\mathrm{d}\boldsymbol{U})\boldsymbol{V} + \boldsymbol{U}(\mathrm{d}\boldsymbol{V})]\,\boldsymbol{W} + (\boldsymbol{UV})\mathrm{d}\boldsymbol{W}\\ &= (\mathrm{d}\boldsymbol{U})\boldsymbol{VW} + \boldsymbol{U}(\mathrm{d}\boldsymbol{V})\boldsymbol{W} + \boldsymbol{UV}(\mathrm{d}\boldsymbol{W})\end{aligned}$$

依此类推，有更一般的矩阵乘积微分结论

$$\mathrm{d}(\boldsymbol{U}_1\boldsymbol{U}_2\cdots\boldsymbol{U}_n) = (\mathrm{d}\boldsymbol{U}_1)\boldsymbol{U}_2\cdots\boldsymbol{U}_n + \boldsymbol{U}_1(\mathrm{d}\boldsymbol{U}_2)\boldsymbol{U}_3\cdots\boldsymbol{U}_n + \boldsymbol{U}_1\cdots\boldsymbol{U}_{n-1}\mathrm{d}\boldsymbol{U}_n$$

4.9 证明

$$\mathrm{d}[\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})] = 2\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X})$$

证明：

$$\begin{aligned}\mathrm{d}[\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})] &= \mathrm{tr}[\mathrm{d}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})] && \text{迹函数和微分的线性性}\\ &= \mathrm{tr}[(\mathrm{d}(\boldsymbol{X}^{\mathrm{T}}))\boldsymbol{X} + \boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}] && \text{乘积矩阵的微分性质}\\ &= \mathrm{tr}[(\mathrm{d}\boldsymbol{X})^{\mathrm{T}}\boldsymbol{X}] + \mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}) && \text{微分和转置的线性性}\\ &= \mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}) + \mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}) && \mathrm{tr}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A}^{\mathrm{T}})\\ &= 2\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X})\end{aligned}$$

4.10 求矩阵函数$\boldsymbol{AXB}$和$\boldsymbol{AX}^{-1}\boldsymbol{B}$的Jacobian矩阵。

解： (1)

$$\mathrm{d}(\boldsymbol{AXB}) = \boldsymbol{A}\mathrm{d}(\boldsymbol{X})\boldsymbol{B}$$

由教材中式(1.10.25)有

$$\mathrm{d}(\mathrm{vec}(\boldsymbol{AXB})) = (\boldsymbol{B}^{\mathrm{T}} \otimes \boldsymbol{A})\mathrm{d}(\mathrm{vec}\boldsymbol{X})$$

由Jacobian矩阵的定义有

$$D(AXB) = (B^{\mathrm{T}} \otimes A)$$

(2)

$$\begin{aligned} \mathrm{d}(AX^{-1}B) &= A\mathrm{d}(X^{-1})B \\ &= -AX^{-1}\mathrm{d}(X)X^{-1}B \end{aligned}$$

由教材中式(1.10.25)有

$$\mathrm{d}(\mathrm{vec}(AX^{-1}B)) = -[(X^{-1}B)^{\mathrm{T}} \otimes (AX^{-1})]\mathrm{d}(\mathrm{vec}X)$$

由Jacobian矩阵的定义有

$$D(AX^{-1}B) = -(X^{-1}B)^{\mathrm{T}} \otimes (AX^{-1})$$

4.11 求矩阵函数$X^{\mathrm{T}}AX$, XAX^{T}, XAX和$X^{\mathrm{T}}AX^{\mathrm{T}}$的Jacobian矩阵。

解: (1) 令X，A分别是$n \times m$，$n \times n$ 维矩阵，则

$$\mathrm{d}(X^{\mathrm{T}}AX) = \mathrm{d}(X^{\mathrm{T}})AX + X^{\mathrm{T}}A\mathrm{d}X$$

由教材中式(1.10.57)有

$$\begin{aligned} \mathrm{d}(\mathrm{vec}(X^{\mathrm{T}}AX)) &= [(AX)^{\mathrm{T}} \otimes I_m]\mathrm{d}(\mathrm{vec}X^{\mathrm{T}}) + [I_m \otimes (X^{\mathrm{T}}A)]\mathrm{d}(\mathrm{vec}X) \\ &= [(AX)^{\mathrm{T}} \otimes I_m]K_{nm}\mathrm{d}(\mathrm{vec}X) + [I_m \otimes (X^{\mathrm{T}}A)]\mathrm{d}(\mathrm{vec}X) \\ &= [((AX)^{\mathrm{T}} \otimes I_m)K_{nm} + I_m \otimes (X^{\mathrm{T}}A)]\mathrm{d}(\mathrm{vec}X) \end{aligned}$$

其中K_{nm}定义参考教材中式(1.10.20)。

由Jacobian矩阵的定义有

$$D(X^{\mathrm{T}}AX) = ((AX)^{\mathrm{T}} \otimes I_m)K_{nm} + I_m \otimes (X^{\mathrm{T}}A)$$

(2) 令X，A分别是$m \times n$，$n \times n$ 维矩阵，则

$$\mathrm{d}(XAX^{\mathrm{T}}) = \mathrm{d}(X)AX^{\mathrm{T}} + XA\mathrm{d}(X^{\mathrm{T}})$$

由教材中式(1.10.57)有

$$\begin{aligned} \mathrm{d}(\mathrm{vec}(XAX^{\mathrm{T}})) &= [(XA^{\mathrm{T}}) \otimes I_m]\mathrm{d}(\mathrm{vec}X) + [I_m \otimes (XA)]\mathrm{d}(\mathrm{vec}X^{\mathrm{T}}) \\ &= [(XA^{\mathrm{T}}) \otimes I_m]\mathrm{d}(\mathrm{vec}X) + [I_m \otimes (XA)]K_{mn}\mathrm{d}(\mathrm{vec}X) \\ &= [(XA^{\mathrm{T}}) \otimes I_m + (I_m \otimes (XA))K_{mn}]\mathrm{d}(\mathrm{vec}X) \end{aligned}$$

由Jacobian矩阵的定义有

$$\boldsymbol{D}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}) = (\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}) \otimes \boldsymbol{I}_m + (\boldsymbol{I}_m \otimes (\boldsymbol{X}\boldsymbol{A}))\boldsymbol{K}_{mn}$$

(3) 令$\boldsymbol{X}$，$\boldsymbol{A}$分别是$n \times m$，$m \times n$ 维矩阵，则

$$\mathrm{d}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}) = \mathrm{d}(\boldsymbol{X})\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{A}\mathrm{d}(\boldsymbol{X})$$

由教材中式(1.10.57)有

$$\begin{aligned}\mathrm{d}(\mathrm{vec}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X})) &= [(\boldsymbol{A}\boldsymbol{X})^{\mathrm{T}} \otimes \boldsymbol{I}_n]\mathrm{d}(\mathrm{vec}\boldsymbol{X}) + [\boldsymbol{I}_m \otimes (\boldsymbol{X}\boldsymbol{A})]\mathrm{d}(\mathrm{vec}\boldsymbol{X}) \\ &= [(\boldsymbol{A}\boldsymbol{X})^{\mathrm{T}} \otimes \boldsymbol{I}_n + \boldsymbol{I}_m \otimes (\boldsymbol{X}\boldsymbol{A})]\mathrm{d}(\mathrm{vec}\boldsymbol{X})\end{aligned}$$

由Jacobian矩阵的定义有

$$\boldsymbol{D}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}) = (\boldsymbol{A}\boldsymbol{X})^{\mathrm{T}} \otimes \boldsymbol{I}_n + \boldsymbol{I}_m \otimes (\boldsymbol{X}\boldsymbol{A})$$

(4) 令$\boldsymbol{X}$，$\boldsymbol{A}$都是$m \times n$ 维矩阵，则

$$\mathrm{d}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}) = \mathrm{d}(\boldsymbol{X}^{\mathrm{T}})\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\mathrm{d}(\boldsymbol{X}^{\mathrm{T}})$$

由教材中式(1.10.57)有

$$\begin{aligned}\mathrm{d}(\mathrm{vec}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})) &= [(\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}) \otimes \boldsymbol{I}_n]\mathrm{d}(\mathrm{vec}\boldsymbol{X}^{\mathrm{T}}) + [\boldsymbol{I}_m \otimes (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})]\mathrm{d}(\mathrm{vec}\boldsymbol{X}^{\mathrm{T}}) \\ &= [(\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}) \otimes \boldsymbol{I}_n + \boldsymbol{I}_m \otimes (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})]\boldsymbol{K}_{mn}\mathrm{d}(\mathrm{vec}\boldsymbol{X})\end{aligned}$$

由Jacobian矩阵的定义有

$$\boldsymbol{D}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}) = [(\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}) \otimes \boldsymbol{I}_n + \boldsymbol{I}_m \otimes (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})]\boldsymbol{K}_{mn}$$

4.12 求行列式对数$\log|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}|, \log|\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}|$，$\log|\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}|$ 和$\log|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}|$ 的Jacobian矩阵与梯度矩阵。

解： (1)

$$\begin{aligned}\mathrm{d}(\log|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}|) &= |\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}|^{-1}|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}|\mathrm{tr}[(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}\mathrm{d}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})] \\ &= \mathrm{tr}[(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}((\mathrm{d}\boldsymbol{X}^{\mathrm{T}})\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{X})] \\ &= \mathrm{tr}[(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{X}] + \mathrm{tr}[\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}\mathrm{d}\boldsymbol{X}^{\mathrm{T}}]\end{aligned}$$

则其梯度矩阵为

$$\frac{\partial \log|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}|}{\partial X} = ((\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})^{\mathrm{T}} + \boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}$$

Jacobian矩阵可以写为

$$\boldsymbol{D}(\log|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}|) = (\mathrm{vec}((\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})^{\mathrm{T}})^{\mathrm{T}} + (\mathrm{vec}(\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}))^{\mathrm{T}}$$

(2)

$$\begin{aligned}\mathrm{d}(\log|\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}|) &= |\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}|^{-1}|\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}|\mathrm{tr}[(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1}\mathrm{d}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})]\\ &= \mathrm{tr}[(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1}((\mathrm{d}\boldsymbol{X})\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}} + \boldsymbol{X}\boldsymbol{A}\mathrm{d}\boldsymbol{X}^{\mathrm{T}})]\\ &= \mathrm{tr}[\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1}\mathrm{d}\boldsymbol{X}] + \mathrm{tr}[(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}\boldsymbol{A}\mathrm{d}\boldsymbol{X}^{\mathrm{T}}]\end{aligned}$$

则其梯度矩阵为

$$\frac{\partial\log|\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}|}{\partial X} = (\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1})^{\mathrm{T}} + (\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}\boldsymbol{A}$$

Jacobian矩阵可以写为

$$\boldsymbol{D}(\log|\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}|) = [\mathrm{vec}(\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1})^{\mathrm{T}}]^{\mathrm{T}} + [\mathrm{vec}((\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}\boldsymbol{A})]^{\mathrm{T}}$$

(3)

$$\begin{aligned}\mathrm{d}(\log|\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}|) &= |\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}|^{-1}|\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}|\mathrm{tr}[(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X})^{-1}\mathrm{d}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X})]\\ &= \mathrm{tr}[(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X})^{-1}((\mathrm{d}\boldsymbol{X})\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}\boldsymbol{A}\mathrm{d}\boldsymbol{X})]\\ &= \mathrm{tr}[\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X})^{-1}\mathrm{d}\boldsymbol{X}] + \mathrm{tr}[(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{A}\mathrm{d}\boldsymbol{X}]\end{aligned}$$

则其梯度矩阵为

$$\frac{\partial\log|\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}|}{\partial X} = (\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X})^{-1})^{\mathrm{T}} + [(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{A}]^{\mathrm{T}}$$

Jacobian矩阵可以写为

$$\boldsymbol{D}(\log|\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}|) = [\mathrm{vec}(\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X})^{-1})^{\mathrm{T}}]^{\mathrm{T}} + [\mathrm{vec}((\boldsymbol{X}\boldsymbol{A}\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{A})^{\mathrm{T}}]^{\mathrm{T}}$$

(4)

$$\begin{aligned}\mathrm{d}(\log|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}|) &= |\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}|^{-1}|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}|\mathrm{tr}[(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1}\mathrm{d}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})]\\ &= \mathrm{tr}[(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1}((\mathrm{d}\boldsymbol{X}^{\mathrm{T}})\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{X}^{\mathrm{T}})]\\ &= \mathrm{tr}[\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1}\mathrm{d}\boldsymbol{X}^{\mathrm{T}}] + \mathrm{tr}[(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{X}^{\mathrm{T}}]\end{aligned}$$

则其梯度矩阵为

$$\frac{\partial\log|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}|}{\partial X} = \boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1} + (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}$$

Jacobian矩阵可以写为

$$\boldsymbol{D}(\log|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}|) = [\mathrm{vec}(\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1})]^{\mathrm{T}} + [\mathrm{vec}((\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})]^{\mathrm{T}}$$

4.13 证明：若$\boldsymbol{F}$是矩阵函数，并且可二次微分，则

$$\mathrm{d}^2\log|\boldsymbol{F}| = -\mathrm{tr}(\boldsymbol{F}^{-1}\mathrm{d}\boldsymbol{F})^2 + \mathrm{tr}(\boldsymbol{F}^{-1}\mathrm{d}^2\boldsymbol{F})$$

思路分析：先求其一次微分，再求二次微分。

证明：

$$\mathrm{d}\log|\boldsymbol{F}| = |\boldsymbol{F}|^{-1}|\boldsymbol{F}|\mathrm{tr}(\boldsymbol{F}^{-1}\mathrm{d}\boldsymbol{F}) = \mathrm{tr}(\boldsymbol{F}^{-1}\mathrm{d}\boldsymbol{F})$$

则

$$\begin{aligned}\mathrm{d}^2\log|\boldsymbol{F}| &= \mathrm{d}[\mathrm{tr}(\boldsymbol{F}^{-1}\mathrm{d}\boldsymbol{F})]\\ &= \mathrm{tr}\left(-\boldsymbol{F}^{-1}(\mathrm{d}\boldsymbol{F})\boldsymbol{F}^{-1}\mathrm{d}\boldsymbol{F} + \boldsymbol{F}^{-1}\mathrm{d}^2\boldsymbol{F}\right)\\ &= -\mathrm{tr}(\boldsymbol{F}^{-1}\mathrm{d}\boldsymbol{F})^2 + \mathrm{tr}(\boldsymbol{F}^{-1}\mathrm{d}^2\boldsymbol{F})\end{aligned}$$

注释：若$\boldsymbol{F}$是矩阵，则$\mathrm{d}^2\log|\boldsymbol{F}| = -\mathrm{tr}(\boldsymbol{F}^{-1}\mathrm{d}\boldsymbol{F})^2$，$\mathrm{d}\boldsymbol{F}$当常数对待。

4.14 假定$\boldsymbol{x}$是N维数据或文本向量，现在希望寻找一个$n\times N$线性变换矩阵$\boldsymbol{W}$对$\boldsymbol{x}$进行数据压缩：$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x}$，使得$n \ll N$。定义目标函数$J_W = \mathrm{tr}[(\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}\boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}}]$，其中，$\boldsymbol{S}_{xw}$和$\boldsymbol{S}_{xb}$ 分别是原数据向量$\boldsymbol{x}$的类内和类间散布矩阵。线性变换矩阵的优化准则是使目标函数J_W最大化。设$\boldsymbol{S}_{xw}^{-1}\boldsymbol{S}_{xb}$的特征值为$\lambda_1,\lambda_2,\cdots,\lambda_N$ $(\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_N)$，并且$\boldsymbol{u}_i$是与特征值$\lambda_i$对应的特征向量。证明$\boldsymbol{W} = [\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_N]$，并且$\max J_W = \sum_{i=1}^{n}\lambda_i$。(提示：使用矩阵微分求梯度矩阵$\partial J_W/\partial\boldsymbol{W}$)

证法一：

$$\begin{aligned}\mathrm{d}J_W &= \mathrm{d}\left(\mathrm{tr}[(\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}\boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}}]\right)\\ &= -\mathrm{tr}\left((\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\mathrm{d}(\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})(\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}\boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}}\right)\\ &\quad + \mathrm{tr}\left((\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}(\mathrm{d}\boldsymbol{W})\boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}} + (\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}\boldsymbol{S}_{xb}\mathrm{d}\,\boldsymbol{W}^{\mathrm{T}}\right)\\ &= -\mathrm{tr}\left(\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}\boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\mathrm{d}\boldsymbol{W}\right)\\ &\quad - \mathrm{tr}\left((\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}\boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}\boldsymbol{S}_{xw}\mathrm{d}\boldsymbol{W}^{\mathrm{T}}\right)\\ &\quad + \mathrm{tr}\left(\boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\mathrm{d}\boldsymbol{W}\right) + \mathrm{tr}\left((\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}\boldsymbol{S}_{xb}\mathrm{d}\,\boldsymbol{W}^{\mathrm{T}}\right)\end{aligned}$$

由$\boldsymbol{S}_{xw}$，$\boldsymbol{S}_{xb}$是对称矩阵有

$$\frac{\partial J_W}{\partial\boldsymbol{W}} = 2(\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\left(\boldsymbol{W}\boldsymbol{S}_{xb} - \boldsymbol{W}\boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}\boldsymbol{S}_{xw}\right)$$

令

$$\frac{\partial J_W}{\partial \boldsymbol{W}} = 0$$

有

$$\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}\boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}} = \boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}}$$

所以有

$$\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}\boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}} = \boldsymbol{S}_{xw}^{-1}\boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}}$$

由于$\boldsymbol{W}$满行秩，等式左右两边同乘$(\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}$有

$$(\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}\boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}} = (\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}\boldsymbol{S}_{xw}^{-1}\boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}}$$

所以最大化J_W问题转换为

$$\max_{W} \operatorname{tr}\left((\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}\boldsymbol{S}_{xw}^{-1}\boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}}\right)$$

不难验证上述最优化问题也等价于

$$\begin{gathered}\max_{W} \operatorname{tr}\left((\boldsymbol{W}\boldsymbol{S}_{xw}^{-1}\boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}}\right)\\ \boldsymbol{W}\boldsymbol{W}^{\mathrm{T}} = \boldsymbol{I}\end{gathered}$$

显然上述最优化问题的解在$\boldsymbol{W}$取$\boldsymbol{S}_{xw}^{-1}\boldsymbol{S}_{xb}$最大$n$个特征值对应的特征向量时$J_W$取到最大值。并且

$$\max J_W = \sum_{i=1}^{n}\lambda_i$$

同理，$\boldsymbol{W}$取$\boldsymbol{S}_{xw}^{-1}\boldsymbol{S}_{xb}$最小$n$个特征值对应的特征向量时$J_W$取到最小值。

证法二： 由于$\boldsymbol{S}_{xw}$具有对称正定性，可对其进行Cholesky分解：

$$\boldsymbol{G}_{xw} = \boldsymbol{S}_{xw}^{1/2}(\boldsymbol{S}_{xw}^{1/2})^{\mathrm{T}}$$

令

$$\boldsymbol{V} = \boldsymbol{W}\boldsymbol{S}_{xw}^{1/2}$$

则

$$\begin{aligned}\mathrm{d}J_W &= \operatorname{tr}[(\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}\boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}}]\\ &= \operatorname{tr}[(\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}})^{-1}\boldsymbol{V}\boldsymbol{S}_{xw}^{-1/2}\boldsymbol{S}_{xb}(\boldsymbol{S}_{xw}^{-1/2})^{\mathrm{T}}\boldsymbol{V}^{\mathrm{T}}]\end{aligned}$$

优化J_W等价于优化J_V：

$$J_V = \max_{V} \operatorname{tr}[(\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}})^{-1}\boldsymbol{V}\boldsymbol{S}_{xw}^{-1/2}\boldsymbol{S}_{xb}(\boldsymbol{S}_{xw}^{-1/2})^{\mathrm{T}}\boldsymbol{V}^{\mathrm{T}}]$$

该优化问题又等价于

$$\max_{V} \operatorname{tr}[\boldsymbol{V}\boldsymbol{S}_{xw}^{-1/2}\boldsymbol{S}_{xb}(\boldsymbol{S}_{xw}^{-1/2})^{\mathrm{T}}\boldsymbol{V}^{\mathrm{T}}]$$
$$\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{I}$$

显然J_W在$\boldsymbol{V}$取$\boldsymbol{S}_{xw}^{-1/2}\boldsymbol{S}_{xb}(\boldsymbol{S}_{xw}^{-1/2})^{\mathrm{T}}$的最大$n$个特征值对应的特征向量时达到最大值。并且

$$\max J_W = \sum_{i=1}^{n} \lambda_i$$

实际上这些特征值也是$\boldsymbol{S}_{xw}^{-1}\boldsymbol{S}_{xb}$的特征值。这是因为

$$\boldsymbol{S}_{xw}^{-1/2}\boldsymbol{S}_{xb}(\boldsymbol{S}_{xw}^{-1/2})^{\mathrm{T}}\boldsymbol{x} = \lambda\boldsymbol{x}$$
$$(\boldsymbol{S}_{xw}^{-1/2})^{\mathrm{T}}\boldsymbol{S}_{xw}^{-1/2}\boldsymbol{S}_{xb}(\boldsymbol{S}_{xw}^{-1/2})^{\mathrm{T}}\boldsymbol{x} = \lambda(\boldsymbol{S}_{xw}^{-1/2})^{\mathrm{T}}\boldsymbol{x}$$

注释： 实际上，以上优化问题是Grassmann 流形上的优化问题。J_W的最大值并不一定只是在$\boldsymbol{W}$取$\boldsymbol{S}_{xw}^{-1}\boldsymbol{S}_{xb}$ n个主特征向量(principle components)时取到。而只要是n个主特征值张成的空间(也称主空间principle subspace)即可。证明如下。

任意取可逆矩阵$\boldsymbol{A}^{n\times n}$。令$\boldsymbol{W}_o$ 由$\boldsymbol{S}_{xw}^{-1}\boldsymbol{S}_{xb}$ n个主特征向量组成。取$\boldsymbol{W} = \boldsymbol{A}\boldsymbol{W}_o$，则

$$\begin{aligned}
J_W &= \operatorname{tr}[(\boldsymbol{W}\boldsymbol{S}_{xw}\boldsymbol{W}^{\mathrm{T}})^{-1}\boldsymbol{W}\boldsymbol{S}_{xb}\boldsymbol{W}^{\mathrm{T}}] \\
&= \operatorname{tr}[(\boldsymbol{A}\boldsymbol{W}_o\boldsymbol{S}_{xw}(\boldsymbol{A}\boldsymbol{W}_o)^{\mathrm{T}})^{-1}\boldsymbol{A}\boldsymbol{W}_o\boldsymbol{S}_{xb}(\boldsymbol{A}\boldsymbol{W}_o)^{\mathrm{T}}] \\
&= \operatorname{tr}[\boldsymbol{A}^{-1}(\boldsymbol{W}_o\boldsymbol{S}_{xw}\boldsymbol{W}_o^{\mathrm{T}})^{-1}\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{W}_o\boldsymbol{S}_{xb}\boldsymbol{W}_o^{\mathrm{T}}\boldsymbol{A}] \\
&= J_{W_o}
\end{aligned}$$

4.15 证明[1]

$$\mathrm{d}(\boldsymbol{F}^{\dagger}\boldsymbol{F}) = \boldsymbol{F}^{\dagger}(\mathrm{d}\boldsymbol{F})(\boldsymbol{I}-\boldsymbol{F}^{\dagger}\boldsymbol{F}) + [\boldsymbol{F}^{\dagger}(\mathrm{d}\boldsymbol{F})(\boldsymbol{I}-\boldsymbol{F}^{\dagger}\boldsymbol{F})]^{\mathrm{T}}$$
$$\mathrm{d}(\boldsymbol{F}\boldsymbol{F}^{\dagger}) = (\boldsymbol{I}-\boldsymbol{F}\boldsymbol{F}^{\dagger})(\mathrm{d}\boldsymbol{F})\boldsymbol{F}^{\dagger} + [(\boldsymbol{I}-\boldsymbol{F}\boldsymbol{F}^{\dagger})(\mathrm{d}\boldsymbol{F})\boldsymbol{F}^{\dagger}]^{\mathrm{T}}$$

式中，$\boldsymbol{F}$是实矩阵，$\boldsymbol{F}^{\dagger}$ 是$\boldsymbol{F}$的Moore-Penrose逆矩阵。

证明： (1) 由Moore-Penrose逆矩阵的定义有$\boldsymbol{F}^{\dagger}\boldsymbol{F}$是对称矩阵，由Moore-Penrose逆矩阵的性质有$\boldsymbol{F}^{\dagger}\boldsymbol{F}$是幂等矩阵(参考教材P89)。

$$\begin{aligned}
\mathrm{d}(\boldsymbol{F}^{\dagger}\boldsymbol{F}) &= \mathrm{d}(\boldsymbol{F}^{\dagger}\boldsymbol{F}\boldsymbol{F}^{\dagger}\boldsymbol{F}) \\
&= (\mathrm{d}\boldsymbol{F}^{\dagger}\boldsymbol{F})\boldsymbol{F}^{\dagger}\boldsymbol{F} + \boldsymbol{F}^{\dagger}\boldsymbol{F}\mathrm{d}(\boldsymbol{F}^{\dagger}\boldsymbol{F}) \\
&= \boldsymbol{F}^{\dagger}\boldsymbol{F}\mathrm{d}(\boldsymbol{F}^{\dagger}\boldsymbol{F}) + (\boldsymbol{F}^{\dagger}\boldsymbol{F}\mathrm{d}(\boldsymbol{F}^{\dagger}\boldsymbol{F}))^{\mathrm{T}}
\end{aligned} \tag{4.73}$$

由Moore-Penrose逆矩阵的定义有

$$\mathrm{d}\boldsymbol{F} = \mathrm{d}\boldsymbol{F}\boldsymbol{F}^{\dagger}\boldsymbol{F} = (\mathrm{d}\boldsymbol{F})\boldsymbol{F}^{\dagger}\boldsymbol{F} + \boldsymbol{F}(\mathrm{d}\boldsymbol{F}^{\dagger}\boldsymbol{F})$$

将上式重新整理为

$$\boldsymbol{F}(\mathrm{d}\boldsymbol{F}^{\dagger}\boldsymbol{F}) = (\mathrm{d}\boldsymbol{F})(\boldsymbol{I} - \boldsymbol{F}^{\dagger}\boldsymbol{F}) \tag{4.74}$$

将式(4.74)代入式(4.73)有

$$\mathrm{d}(\boldsymbol{F}^{\dagger}\boldsymbol{F}) = \boldsymbol{F}^{\dagger}(\mathrm{d}\boldsymbol{F})(\boldsymbol{I} - \boldsymbol{F}^{\dagger}\boldsymbol{F}) + [\boldsymbol{F}^{\dagger}(\mathrm{d}\boldsymbol{F})(\boldsymbol{I} - \boldsymbol{F}^{\dagger}\boldsymbol{F})]^{\mathrm{T}}$$

(2) 证明方法完全同(1)，留给读者练习。

4.16 已知线性方程

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \vdots & \vdots & & \vdots \\ \lambda_1^{n-1} & \lambda_2 & \cdots & \lambda_n^{n-1} \end{bmatrix} \begin{bmatrix} \mathrm{d}\lambda_1 \\ \mathrm{d}\lambda_2 \\ \vdots \\ \mathrm{d}\lambda_n \end{bmatrix} = \begin{bmatrix} \mathrm{tr}(\mathrm{d}\boldsymbol{X}) \\ \mathrm{tr}(\boldsymbol{X}_0\mathrm{d}\boldsymbol{X}) \\ \vdots \\ \mathrm{tr}(\boldsymbol{X}_0^{n-1}\mathrm{d}\boldsymbol{X}) \end{bmatrix}$$

求$\mathrm{d}\lambda_i$。

解：记

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \vdots & \vdots & & \vdots \\ \lambda_1^{n-1} & \lambda_2 & \cdots & \lambda_n^{n-1} \end{bmatrix}^{-1} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{bmatrix}$$

则

$$\mathrm{d}\lambda_i = \mathrm{tr}\left(\sum_{j=1}^{n} v_{ij}\boldsymbol{X}_0^{j-1}\mathrm{d}\boldsymbol{X}\right)$$

注释：本题解方程解答到此就算结束了。但细心的读者会问这个线性方程为什么成立，这个方程是如何来的？

实际上我们知道

$$\mathrm{tr}\boldsymbol{X}^k = \sum_{i=0}^{n} \lambda_i^k$$

其中λ_i是$\boldsymbol{X}^{n\times n}$的特征值。则有

$$\mathrm{dtr}\boldsymbol{X}^k = k\sum_{i=0}^{n} \lambda_i^{k-1}\mathrm{d}\lambda_i = k\mathrm{tr}(\boldsymbol{X}^{k-1}\mathrm{d}\boldsymbol{X})$$

取$k = 1, 2, \cdots, n$，则可以得到以上线性方程。

本题结论是一个很有用的定理，它阐述了矩阵内元素的扰动对其特征值的影响。

4.17 计算标量函数$f(\boldsymbol{x}) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$和$f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$的Hessian矩阵。

解：由Hessian矩阵的定义(参考教材P282)有

(1)

$$\boldsymbol{H}f(\boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}}\left(\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}\right) = \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{x}^{\mathrm{T}}} = \boldsymbol{O}$$

(2)

$$\boldsymbol{H}f(\boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}}\left(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}\right) = \boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}}$$

4.18 证明标量函数$f(\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^{\mathrm{T}})$的Hessian矩阵为

$$\boldsymbol{H}[f(\boldsymbol{X})] = \boldsymbol{B}^{\mathrm{T}} \otimes \boldsymbol{A} + \boldsymbol{B} \otimes \boldsymbol{A}^{\mathrm{T}}$$

证明： 由Kronecker积的性质有(参考教材P13式(1.10.62))

$$\mathrm{tr}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^{\mathrm{T}}) = (\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}(\boldsymbol{B}^{\mathrm{T}} \otimes \boldsymbol{A})\mathrm{vec}\boldsymbol{X}$$

由Hessian 矩阵的定义(教材P282)有

$$\begin{aligned}
\boldsymbol{H}[f(\boldsymbol{X})] &= \frac{\partial}{\partial(\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}}\left(\frac{\partial f(\boldsymbol{X})}{\partial(\mathrm{vec}\boldsymbol{X})}\right) \\
&= \frac{\partial}{\partial(\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}}\left((\boldsymbol{B}^{\mathrm{T}} \otimes \boldsymbol{A})\mathrm{vec}\boldsymbol{X} + (\boldsymbol{B}^{\mathrm{T}} \otimes \boldsymbol{A})^{\mathrm{T}}\mathrm{vec}\boldsymbol{X}\right) \\
&= \boldsymbol{B}^{\mathrm{T}} \otimes \boldsymbol{A} + \boldsymbol{B} \otimes \boldsymbol{A}^{\mathrm{T}}
\end{aligned}$$

4.19 求行列式对数$\log|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}|, \log|\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}|$和$\log|\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}|$ 的Hessian矩阵。

思路分析： 求Hessian矩阵得先求Jacobian矩阵。我们直接利用题4.12的结论。

解： (1)

$$\frac{\partial f(\boldsymbol{X})}{\partial(\mathrm{vec}\boldsymbol{X})} = (\boldsymbol{D}f(\boldsymbol{X}))^{\mathrm{T}}$$

所以

$$\frac{\partial\left(\log|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}|\right)}{\partial(\mathrm{vec}\boldsymbol{X})} = \mathrm{vec}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})^{-1}) + \mathrm{vec}(\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1})$$

$$\begin{aligned}
&\mathrm{d}\left(\frac{\partial\left(\log|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}|\right)}{\partial(\mathrm{vec}\boldsymbol{X})}\right) \\
=&\mathrm{vec}\left(\boldsymbol{A}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})^{-1}\right) \\
&+ \mathrm{vec}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})^{-1}\left(\mathrm{d}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right)(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})^{-1}\right) \\
&+ \mathrm{vec}\left(\boldsymbol{A}\mathrm{d}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}\right) \\
&+ \mathrm{vec}\left(\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}\left(\mathrm{d}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{X}\right)(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}\right) \\
=&\left(\boldsymbol{I} \otimes (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{A}\right)\mathrm{vec}(\mathrm{d}\boldsymbol{X}) + \left(\boldsymbol{I} \otimes (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})^{-1})^{2}\right)\mathrm{vec}(\mathrm{d}\boldsymbol{X}^{\mathrm{T}}) \\
&+ \left(\boldsymbol{I} \otimes (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\right)\mathrm{vec}(\mathrm{d}\boldsymbol{X}) \\
&+ \left(\boldsymbol{I} \otimes (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}\boldsymbol{A}\right)\mathrm{vec}(\mathrm{d}\boldsymbol{X}) + \left(\boldsymbol{I} \otimes (\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1})^{2}\right)\mathrm{vec}(\mathrm{d}\boldsymbol{X}^{\mathrm{T}}) \\
&+ \left(\boldsymbol{I} \otimes (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\right)\mathrm{vec}(\mathrm{d}\boldsymbol{X})
\end{aligned}$$

则Hessian矩阵可以表示为

$$
\begin{aligned}
\boldsymbol{H}\left(\log|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}|\right) =&\frac{\partial}{\partial(\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}}\left(\frac{\partial\left(\log|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}|\right)}{\partial(\mathrm{vec}\boldsymbol{X})}\right)\\
=&\boldsymbol{I}\otimes(\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}+\boldsymbol{I}\otimes(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})^{-1})^2\\
&+\boldsymbol{I}\otimes\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}+\boldsymbol{I}\otimes\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})^{-1}\\
&+\boldsymbol{I}\otimes(\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1})^2\\
&+\boldsymbol{I}\otimes\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})^{-1}
\end{aligned}
$$

(2) 因为

$$
\frac{\partial\left(\log|\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}|\right)}{\partial(\mathrm{vec}\boldsymbol{X})}=\mathrm{vec}\left((\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}\right)+\mathrm{vec}\left((\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}\boldsymbol{A}\right)
$$

$$
\begin{aligned}
&\mathrm{d}\left(\frac{\partial\left(\log|\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}|\right)}{\partial(\mathrm{vec}\boldsymbol{X})}\right)\\
=&\mathrm{vec}\left((\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}((\mathrm{d}\boldsymbol{X})\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}+\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}^{\mathrm{T}})(\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}\right)\\
&+\mathrm{vec}\left((\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}(\mathrm{d}\boldsymbol{X})\boldsymbol{A}^{\mathrm{T}}\right)+\mathrm{vec}\left(\boldsymbol{A}(\mathrm{d}\boldsymbol{X})(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}\right)\\
&+\mathrm{vec}\left(\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}((\mathrm{d}\boldsymbol{X}^{\mathrm{T}})\boldsymbol{A}\boldsymbol{X}+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{X})(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}\right)\\
=&\left(\boldsymbol{I}\otimes\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\right)\mathrm{vec}\,(\mathrm{d}\boldsymbol{X})\\
&+\left(\boldsymbol{I}\otimes((\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}})^2\right)\mathrm{vec}\left(\mathrm{d}\boldsymbol{X}^{\mathrm{T}}\right)\\
&+\left(\boldsymbol{I}\otimes\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\right)\mathrm{vec}\,(\mathrm{d}\boldsymbol{X})+\left(\boldsymbol{I}\otimes(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}\boldsymbol{A}\right)\mathrm{vec}\,(\mathrm{d}\boldsymbol{X})\\
&+\left(\boldsymbol{I}\otimes(\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1})^2\right)\mathrm{vec}\left(\mathrm{d}\boldsymbol{X}^{\mathrm{T}}\right)\\
&+\left(\boldsymbol{I}\otimes(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\right)\mathrm{vec}\,(\mathrm{d}\boldsymbol{X})
\end{aligned}
$$

则Hessian矩阵可以写为

$$
\begin{aligned}
\boldsymbol{H}\left(\log|\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}|\right) =&\frac{\partial}{\partial(\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}}\left(\frac{\partial\left(\log|\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}|\right)}{\partial(\mathrm{vec}\boldsymbol{X})}\right)\\
=&\left(\boldsymbol{I}\otimes(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}\boldsymbol{A}\right)\\
&+\left(\boldsymbol{I}\otimes((\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}\boldsymbol{A}^{\mathrm{T}})^2\right)\\
&+\left(\boldsymbol{I}\otimes(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{A}\right)+\left(\boldsymbol{I}\otimes\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})^{-1}\right)\\
&+\left(\boldsymbol{I}\otimes(\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1})^2\right)\\
&+\left(\boldsymbol{I}\otimes\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})^{-1}\right)
\end{aligned}
$$

(3) 因为

$$\frac{\partial\left(\log|\boldsymbol{XAX}|\right)}{\partial(\mathrm{vec}\boldsymbol{X})}=\mathrm{vec}\left((\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\right)+\mathrm{vec}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\right)$$

则

$$\begin{aligned}
&\mathrm{d}\left(\frac{\partial\left(\log|\boldsymbol{XAX}|\right)}{\partial(\mathrm{vec}\boldsymbol{X})}\right)\\
=&\mathrm{vec}\left((\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}((\mathrm{d}\boldsymbol{X}^{\mathrm{T}})\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}(\mathrm{d}\boldsymbol{X}^{\mathrm{T}}))(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\right)\\
&+\mathrm{vec}\left((\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}(\mathrm{d}\boldsymbol{X}^{\mathrm{T}})\boldsymbol{A}^{\mathrm{T}}\right)+\mathrm{vec}\left(\boldsymbol{A}^{\mathrm{T}}(\mathrm{d}\boldsymbol{X}^{\mathrm{T}})(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\right)\\
&+\mathrm{vec}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}((\mathrm{d}\boldsymbol{X}^{\mathrm{T}})\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}(\mathrm{d}\boldsymbol{X}^{\mathrm{T}}))(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\right)\\
=&\left(\boldsymbol{I}\otimes\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\right)\mathrm{vec}\left(\mathrm{d}\boldsymbol{X}^{\mathrm{T}}\right)\\
&+\left(\boldsymbol{I}\otimes(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\right)\mathrm{vec}\left(\mathrm{d}\boldsymbol{X}^{\mathrm{T}}\right)\\
&+\left(\boldsymbol{I}\otimes\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\right)\mathrm{vec}\left(\mathrm{d}\boldsymbol{X}^{\mathrm{T}}\right)\\
&+\left(\boldsymbol{I}\otimes(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{A}^{\mathrm{T}}\right)\mathrm{vec}\left(\mathrm{d}\boldsymbol{X}^{\mathrm{T}}\right)\\
&+\left(\boldsymbol{I}\otimes(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1})^{2}\right)\mathrm{vec}\left(\mathrm{d}\boldsymbol{X}^{\mathrm{T}}\right)\\
&+\left(\boldsymbol{I}\otimes((\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{2}\right)\mathrm{vec}\left(\mathrm{d}\boldsymbol{X}^{\mathrm{T}}\right)
\end{aligned}$$

则有

$$\begin{aligned}
\boldsymbol{H}\left(\log|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}|\right)=&\frac{\partial}{\partial(\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}}\left(\frac{\partial\left(\log|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}|\right)}{\partial(\mathrm{vec}\boldsymbol{X})}\right)\\
=&\left(\boldsymbol{I}\otimes\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\right)\\
&+\left(\boldsymbol{I}\otimes(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\right)\\
&+\left(\boldsymbol{I}\otimes\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\right)+\left(\boldsymbol{I}\otimes(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{A}^{\mathrm{T}}\right)\\
&+\left(\boldsymbol{I}\otimes(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1})^{2}\right)\\
&+\left(\boldsymbol{I}\otimes((\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{2}\right)
\end{aligned}$$

4.20 求标量函数$f(\boldsymbol{X})=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{a}$的Hessian矩阵。

解:

$$\begin{aligned}
\mathrm{d}f(\boldsymbol{X})=\mathrm{d}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{a})&=\mathrm{d}\left(\mathrm{tr}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{a})\right)\\
&=\mathrm{tr}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right)+\mathrm{tr}\left(\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\mathrm{d}\boldsymbol{X}^{\mathrm{T}}\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial f(\boldsymbol{X})}{\partial(\mathrm{vec}\boldsymbol{X})}&=\mathrm{vec}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}\right)^{\mathrm{T}}+\mathrm{vec}\left(\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\right)\\
&=2\,\mathrm{vec}\left(\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\right)\\
&=2\left(\boldsymbol{I}\otimes\boldsymbol{A}\right)\mathrm{vec}\boldsymbol{X}
\end{aligned}$$

则Hessian矩阵为

$$\begin{aligned}\boldsymbol{H}[f(\boldsymbol{X})] &= \frac{\partial}{\partial(\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}}\left(\frac{\partial f(\boldsymbol{X})}{\partial(\mathrm{vec}\boldsymbol{X})}\right)\\ &= \frac{\partial}{\partial(\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}}\left[2\ (\boldsymbol{I}\otimes\boldsymbol{A})\,\mathrm{vec}\boldsymbol{X}\right]\\ &= 2\ (\boldsymbol{I}\otimes\boldsymbol{A})\end{aligned}$$

4.21 令观测数据向量由线性回归模型

$$\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\epsilon},\qquad \mathrm{E}\{\boldsymbol{\epsilon}\}=\boldsymbol{0},\quad \mathrm{E}\{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}}\}=\sigma^2\boldsymbol{I}$$

产生。现在希望设计一个滤波器矩阵$\boldsymbol{A}$，其输出向量$\boldsymbol{e}=\boldsymbol{A}\boldsymbol{y}$满足$\mathrm{E}\{\boldsymbol{e}-\boldsymbol{\epsilon}\}=\boldsymbol{0}$，并且可以使得$\mathrm{E}\{(\boldsymbol{e}-\boldsymbol{\epsilon})^{\mathrm{T}}(\boldsymbol{e}-\boldsymbol{\epsilon})\}$最小化。证明这个最优化问题等效为

$$\min\left[\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})-2\mathrm{tr}(\boldsymbol{A})\right]$$

约束条件为$\boldsymbol{A}\boldsymbol{X}=\boldsymbol{O}$，其中，$\boldsymbol{O}$为零矩阵。

证明： 由题意显然有

$$\begin{aligned}\mathrm{E}\{\boldsymbol{e}-\boldsymbol{\epsilon}\} &= \mathrm{E}\{\boldsymbol{A}\boldsymbol{y}-\boldsymbol{\epsilon}\}\\ &= \boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta}+\mathrm{E}\{\boldsymbol{A}\boldsymbol{\epsilon}-\boldsymbol{\epsilon}\}\\ &= \boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{0}\end{aligned}$$

一般来说，模型系数与数据无关，则即有

$$\boldsymbol{A}\boldsymbol{X}=\boldsymbol{O}$$

$$\begin{aligned}\mathrm{E}\{(\boldsymbol{e}-\boldsymbol{\epsilon})^{\mathrm{T}}(\boldsymbol{e}-\boldsymbol{\epsilon})\} &= \mathrm{E}\{(\boldsymbol{A}\boldsymbol{\epsilon}-\boldsymbol{\epsilon})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{\epsilon}-\boldsymbol{\epsilon})\}\\ &= \mathrm{E}\{(\boldsymbol{A}-\boldsymbol{I})^{\mathrm{T}}\sigma^2\boldsymbol{I}(\boldsymbol{A}-\boldsymbol{I})\}\\ &= \sigma^2[\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})-2\mathrm{tr}(\boldsymbol{A})]\end{aligned}$$

则命题得证。

4.22 证明最优化问题

$$\min\left[\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})-2\mathrm{tr}(\boldsymbol{A})\right],\qquad \text{约束条件为}\quad \boldsymbol{A}\boldsymbol{X}=\boldsymbol{O}\quad(\text{零矩阵})$$

的解矩阵为$\hat{\boldsymbol{A}}=\boldsymbol{I}-\boldsymbol{X}\boldsymbol{X}^{\dagger}$。

思路分析：我们接触得多的约束型一般是标量和向量的形式。如何将矩阵型的约束转化为标量型是问题的一个关键。实际上，约束条件等价于

$$\forall\,\boldsymbol{L},\qquad 2\,\mathrm{tr}(\boldsymbol{L}\boldsymbol{A}\boldsymbol{X})=0$$

系数2是为了下面证明的方便。

证法一： 构造Lagrange函数如下：

$$J(\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) - 2\,\mathrm{tr}(\boldsymbol{A}) + 2\,\mathrm{tr}(\boldsymbol{L}\boldsymbol{A}\boldsymbol{X})$$

$$\frac{\partial J(\boldsymbol{X})}{\partial \boldsymbol{X}} = 2\,\boldsymbol{A} - 2\,\boldsymbol{I} + 2\,\boldsymbol{L}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}$$

令$\partial J(\boldsymbol{X})/\partial \boldsymbol{X} = 0$于是有

$$\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{L}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}} \tag{4.75}$$

代入约束条件有

$$\boldsymbol{X} = \boldsymbol{L}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \tag{4.76}$$

显然方程有解且不唯一，但至少有一个解可以表示为

$$\boldsymbol{L} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{\dagger}\boldsymbol{X}^{\mathrm{T}}$$

将上式代入式(4.75)有

$$\begin{aligned}\boldsymbol{A} &= \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{\dagger}\boldsymbol{X}^{\mathrm{T}} \\ &= \boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{\dagger} \qquad\qquad \text{(根据M-P逆的性质，教材P87)}\end{aligned}$$

方程式(4.76)解$\boldsymbol{L}$不唯一，但是解矩阵$\boldsymbol{A}$是唯一的。

取方程式(4.76)的两个不同的解$\boldsymbol{L}_1$，$\boldsymbol{L}_2$，对应不同的解矩阵$\boldsymbol{A}_1$，$\boldsymbol{A}_2$

$$\boldsymbol{O} = \boldsymbol{X} - \boldsymbol{X} = \boldsymbol{L}_1^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} - \boldsymbol{L}_2^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = (\boldsymbol{L}_1^{\mathrm{T}} - \boldsymbol{L}_2^{\mathrm{T}})\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$$

则

$$(\boldsymbol{L}_1^{\mathrm{T}} - \boldsymbol{L}_2^{\mathrm{T}})\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{L}_1 - \boldsymbol{L}_2) = \boldsymbol{O}$$

所以

$$\boldsymbol{O} = (\boldsymbol{L}_1^{\mathrm{T}} - \boldsymbol{L}_2^{\mathrm{T}})\boldsymbol{X}^{\mathrm{T}} = \boldsymbol{I} - \boldsymbol{L}_1^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}} - \boldsymbol{I} - \boldsymbol{L}_2^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}} = \boldsymbol{A}_1 - \boldsymbol{A}_2$$

即证明了解矩阵$\boldsymbol{A}$是唯一的。

证法二： 方程式(4.76)以上步骤同证法一。方程式(4.76)的解可以表示为

$$\boldsymbol{L} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{\dagger}\boldsymbol{X}^{\mathrm{T}} + [\boldsymbol{I} - (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{\dagger}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}]\boldsymbol{Y} \tag{4.77}$$

将式(4.77)代入式(4.75)有

$$\begin{aligned}\boldsymbol{A} &= \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{\dagger}\boldsymbol{X}^{\mathrm{T}} + \boldsymbol{Y}\left(\boldsymbol{X}^{\mathrm{T}} - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{\dagger}\boldsymbol{X}^{\mathrm{T}}\right) \\ &= \boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{\dagger} + \boldsymbol{Y}\left(\boldsymbol{X}^{\mathrm{T}} - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\dagger}\right) \\ &= \boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{\dagger} + \boldsymbol{Y}\left(\boldsymbol{X}^{\mathrm{T}} - \boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}})^{\dagger}\boldsymbol{X}^{\mathrm{T}}\right) \\ &= \boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{\dagger}\end{aligned}$$

4.23 证明约束最优化问题

$$\min\ \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}, \qquad \text{约束条件为} \quad \boldsymbol{C}\boldsymbol{x}=\boldsymbol{b}$$

具有唯一解$\boldsymbol{x}^*=\boldsymbol{C}^{\dagger}\boldsymbol{b}$。

证明： 构造Lagrange优化函数如下：

$$J(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{l}^{\mathrm{T}}(\boldsymbol{b}-\boldsymbol{C}\boldsymbol{x})$$

分别关于$\boldsymbol{x}$和$\boldsymbol{l}$求偏导有

$$\boldsymbol{C}^{\mathrm{T}}\boldsymbol{l}=\boldsymbol{x} \tag{4.78}$$

$$\boldsymbol{C}\boldsymbol{x}=\boldsymbol{b} \tag{4.79}$$

将式(4.78)代入式(4.79)有

$$\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{l}=\boldsymbol{b} \tag{4.80}$$

则$\boldsymbol{l}$可以表示为

$$\boldsymbol{l}=(\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}})^{\dagger}\boldsymbol{b}+\left[\boldsymbol{I}-(\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}})^{\dagger}\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}}\right]\boldsymbol{y} \tag{4.81}$$

将式(4.81)代入式(4.78)有

$$\boldsymbol{x}^*=\boldsymbol{C}^{\mathrm{T}}\left\{(\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}})^{\dagger}\boldsymbol{b}+\left[\boldsymbol{I}-(\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}})^{\dagger}\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}}\right]\boldsymbol{y}\right\}$$

根据教材P90的结论

$$\boldsymbol{C}^{\dagger}=\boldsymbol{C}^{\mathrm{T}}(\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}})^{\dagger} \qquad (\boldsymbol{C}^{\mathrm{T}})^{\dagger}=(\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}})^{\dagger}\boldsymbol{C}$$

则有

$$\boldsymbol{x}^*=\boldsymbol{C}^{\dagger}\boldsymbol{b}+\left[\boldsymbol{C}^{\mathrm{T}}-\boldsymbol{C}^{\mathrm{T}}(\boldsymbol{C}^{\mathrm{T}})^{\dagger}\boldsymbol{C}^{\mathrm{T}}\right]\boldsymbol{y}=\boldsymbol{C}^{\dagger}\boldsymbol{b}$$

注释： 本题由式(4.80)到式(4.81)容易犯如下两个错误:

$$\boldsymbol{l}=(\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}})^{\dagger}\boldsymbol{b} \qquad \text{或}\boldsymbol{l}=(\boldsymbol{C}\boldsymbol{C})^{-1}\boldsymbol{b}$$

并且从这两个错误却都可以得到正确的结论。

4.24 若约束最优化问题为$\min \mathrm{tr}(\boldsymbol{A}\boldsymbol{V}\boldsymbol{A}^{\mathrm{T}})$，约束条件是$\boldsymbol{A}\boldsymbol{X}=\boldsymbol{W}$，证明

$$\boldsymbol{A}=\boldsymbol{W}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}_0^{\dagger}\boldsymbol{X})^{\dagger}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}_0^{\dagger}+\boldsymbol{Q}(\boldsymbol{I}-\boldsymbol{V}_0\boldsymbol{V}_0^{\dagger})$$

其中，$\boldsymbol{V}_0=\boldsymbol{V}+\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$，$\boldsymbol{Q}$是一任意矩阵。

证明： 构造拉格朗日目标函数如下：

$$f(\boldsymbol{A})=\mathrm{tr}(\boldsymbol{A}\boldsymbol{V}\boldsymbol{A}^{\mathrm{T}})+2\boldsymbol{L}^{\mathrm{T}}(\boldsymbol{W}-\boldsymbol{A}\boldsymbol{X})$$

目标函数分别关于$\boldsymbol{A}$和$\boldsymbol{L}$求偏导得

$$\boldsymbol{AV} = \boldsymbol{LX}^{\mathrm{T}}$$

$$\boldsymbol{AX} = \boldsymbol{W}$$

由以上两个一阶条件有矩阵方程

$$\begin{bmatrix} \boldsymbol{V}^{\mathrm{T}} & \boldsymbol{X} \\ \boldsymbol{X}^{\mathrm{T}} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}^{\mathrm{T}} \\ \boldsymbol{L}^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{W}^{\mathrm{T}} \end{bmatrix}$$

这是一个典型的矩阵方程，其解可以参考文献[1]定理3.23和定理3.24。这里不再论述，直接给出结论：

$$\boldsymbol{A} = \boldsymbol{W}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}_0^{\dagger}\boldsymbol{X})^{\dagger}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}_0^{\dagger} + \boldsymbol{Q}(\boldsymbol{I} - \boldsymbol{V}_0\boldsymbol{V}_0^{\dagger})$$

其中，$\boldsymbol{V}_0 = \boldsymbol{V} + \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$，$\boldsymbol{Q}$是一任意矩阵。

4.25 约束最优化问题为min tr$(\boldsymbol{AS}^{\mathrm{T}})$，约束条件为$\boldsymbol{AX} = \boldsymbol{O}$，$\boldsymbol{AA}^{\mathrm{T}} = \boldsymbol{I}$。证明：最优化问题的解为[1, pp.302~303]

$$\boldsymbol{A} = (\boldsymbol{SMS}^{\mathrm{T}})^{-1/2}\boldsymbol{SM}, \qquad \boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{\dagger}$$

证明： 定义Lagrange目标函数如下：

$$f(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{AS}^{\mathrm{T}}) - \mathrm{tr}(\boldsymbol{L}_1^{\mathrm{T}}\boldsymbol{AX}) - \frac{1}{2}\mathrm{tr}\boldsymbol{L}_2(\boldsymbol{AA}^{\mathrm{T}} - \boldsymbol{I})$$

其中$\boldsymbol{L}_1$和$\boldsymbol{L}_2$是拉格朗日乘子矩阵，$\boldsymbol{L}_2$选择为对称矩阵。将目标函数关于$\boldsymbol{A}$求微分有

$$\begin{aligned} \mathrm{d}f &= \mathrm{tr}(\mathrm{d}\boldsymbol{A})\boldsymbol{S}^{\mathrm{T}} - \mathrm{tr}\boldsymbol{L}_1^{\mathrm{T}}(\mathrm{d}\boldsymbol{A})\boldsymbol{X} - \frac{1}{2}\mathrm{tr}\boldsymbol{L}_2(\mathrm{d}\boldsymbol{A})\boldsymbol{A}^{\mathrm{T}} - \frac{1}{2}\mathrm{tr}\boldsymbol{L}_2\boldsymbol{A}(\mathrm{d}\boldsymbol{A})^{\mathrm{T}} \\ &= \mathrm{tr}\boldsymbol{S}^{\mathrm{T}}\mathrm{d}\boldsymbol{A} - \mathrm{tr}\boldsymbol{X}\boldsymbol{L}_1^{\mathrm{T}}\mathrm{d}\boldsymbol{A} - \mathrm{tr}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{L}_2\mathrm{d}\boldsymbol{A} \end{aligned}$$

则可以得到一阶条件

$$\boldsymbol{S}^{\mathrm{T}} = \boldsymbol{X}\boldsymbol{L}_1^{\mathrm{T}} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{L}_2 \tag{4.82}$$

$$\boldsymbol{AX} = 0 \tag{4.83}$$

$$\boldsymbol{AA}^{\mathrm{T}} = \boldsymbol{I} \tag{4.84}$$

式(4.82)两边左乘$\boldsymbol{X}\boldsymbol{X}^{\dagger}$有

$$\begin{aligned} \boldsymbol{X}\boldsymbol{X}^{\dagger}\boldsymbol{S}^{\mathrm{T}} &= \boldsymbol{X}\boldsymbol{X}^{\dagger}\boldsymbol{X}\boldsymbol{L}_1^{\mathrm{T}} + \boldsymbol{X}\boldsymbol{X}^{\dagger}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{L}_2 \\ &= \boldsymbol{X}\boldsymbol{L}_1^{\mathrm{T}} + (\boldsymbol{X}^{\dagger})^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{L}_2 \\ &= \boldsymbol{X}\boldsymbol{L}_1^{\mathrm{T}} \end{aligned}$$

将上式代入式(4.82)有

$$\boldsymbol{MS}^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{L}_2 \tag{4.85}$$

其中$\boldsymbol{M}=\boldsymbol{I}-\boldsymbol{X}\boldsymbol{X}^{\dagger}$。式(4.82)两边左乘$\boldsymbol{A}$有

$$\boldsymbol{A}\boldsymbol{S}^{\mathrm{T}}=\boldsymbol{L}_2=\boldsymbol{L}_2^{\mathrm{T}}=\boldsymbol{S}\boldsymbol{A}^{\mathrm{T}}$$

式(4.85)两边左乘$\boldsymbol{S}$并联合上式得到

$$\boldsymbol{S}\boldsymbol{M}\boldsymbol{S}^{\mathrm{T}}=\boldsymbol{L}_2^2$$

即

$$\boldsymbol{L}_2=(\boldsymbol{S}\boldsymbol{M}\boldsymbol{S}^{\mathrm{T}})^{1/2}$$

将上式代入式(4.85)得到

$$\boldsymbol{A}=(\boldsymbol{S}\boldsymbol{M}\boldsymbol{S}^{\mathrm{T}})^{-1/2}\boldsymbol{S}\boldsymbol{M}$$

4.26 [1, p.367] 令$\boldsymbol{S}$和$\boldsymbol{\Phi}$是两个已知的$p\times p$正定矩阵，并且$\boldsymbol{\Phi}$为对角矩阵。令f是一实值函数，由

$$f(\boldsymbol{A})=\log|\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{\Phi}|+\mathrm{tr}\left((\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{\Phi})^{-1}\boldsymbol{S}\right)$$

定义，其中$\boldsymbol{A}\in R^{p\times m}$，$1\leqslant m\leqslant p$。证明：

(1) 当$\boldsymbol{A}=\boldsymbol{\Phi}^{1/2}\boldsymbol{T}(\boldsymbol{\Lambda}-\boldsymbol{I}_m)^{1/2}$时，实值函数$f$达到最小。式中，$\boldsymbol{\Lambda}$是一个$m\times m$对角矩阵，它包含了矩阵$\boldsymbol{\Phi}^{-1/2}\boldsymbol{S}\boldsymbol{\Phi}^{-1/2}$的$m$个最大特征值，而矩阵$\boldsymbol{T}$ 是一个$p\times m$矩阵，由矩阵$\boldsymbol{\Phi}^{-1/2}\boldsymbol{S}\boldsymbol{\Phi}^{-1/2}$的$m$个主特征向量组成。

(2) 实值函数f的最小值为

$$m+\log|\boldsymbol{S}|+\sum_{i=m+1}^{p}(\lambda_i-\log\lambda_i-1)$$

式中，$\lambda_{m+1},\lambda_{m+2},\cdots,\lambda_p$表示矩阵$\boldsymbol{\Phi}^{-1/2}\boldsymbol{S}\boldsymbol{\Phi}^{-1/2}$的$p-m$个最小的特征值。

证明： 定义

$$\boldsymbol{\Omega}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{\Phi},\qquad \boldsymbol{C}=\boldsymbol{\Omega}^{-1}-\boldsymbol{\Omega}^{-1}\boldsymbol{S}\boldsymbol{\Omega}^{-1}$$

则有$f(\boldsymbol{A})=\log|\boldsymbol{\Omega}|+\mathrm{tr}\left(\boldsymbol{\Omega}^{-1}\boldsymbol{S}\right)$，所以

$$\begin{aligned}\mathrm{d}f&=\mathrm{tr}\boldsymbol{\Omega}^{-1}\mathrm{d}\boldsymbol{\Omega}-\mathrm{tr}\boldsymbol{\Omega}^{-1}(\mathrm{d}\boldsymbol{\Omega})\boldsymbol{\Omega}^{-1}\boldsymbol{S}=\mathrm{tr}\boldsymbol{C}\mathrm{d}\boldsymbol{\Omega}\\&=\mathrm{tr}(\boldsymbol{C}(\mathrm{d}\boldsymbol{A})\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}(\mathrm{d}\boldsymbol{A})^{\mathrm{T}})=2\mathrm{tr}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}\mathrm{d}\boldsymbol{A}\end{aligned}$$

令$\mathrm{d}f=0$ 有

$$\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{C}\boldsymbol{A}=0$$

等价于

$$\boldsymbol{A}=\boldsymbol{S}\boldsymbol{\Omega}^{-1}\boldsymbol{A}\tag{4.86}$$

式(4.86)两边同时右乘$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\boldsymbol{A}$ 有

$$\begin{aligned}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\boldsymbol{A} &= \boldsymbol{S}\boldsymbol{\Omega}^{-1}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\boldsymbol{A} = \boldsymbol{S}\boldsymbol{\Omega}^{-1}(\boldsymbol{\Omega}-\boldsymbol{\Phi})\boldsymbol{\Phi}^{-1}\boldsymbol{A}\\ &= \boldsymbol{S}\boldsymbol{\Phi}^{-1}\boldsymbol{A} - \boldsymbol{S}\boldsymbol{\Omega}^{-1}\boldsymbol{A} = \boldsymbol{S}\boldsymbol{\Phi}^{-1}\boldsymbol{A} - \boldsymbol{A}\end{aligned}$$

所以有

$$\boldsymbol{S}\boldsymbol{\Phi}^{-1}\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{I}_m + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\boldsymbol{A}) \tag{4.87}$$

假设rank$(\boldsymbol{A}) = m' \leqslant m$，选择$\boldsymbol{Q}$为一个$m \times m'$维的准正交矩阵(即$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}_{m'}$)，使得

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{M} \tag{4.88}$$

其中$\boldsymbol{M}$是一个对角元素为$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\boldsymbol{A}$的非零特征值的对角矩阵。所以式(4.87)可以表示为

$$\boldsymbol{S}\boldsymbol{\Phi}^{-1}\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{A}\boldsymbol{Q}(\boldsymbol{I}_m + \boldsymbol{M})$$

从而有

$$(\boldsymbol{\Phi}^{-1/2}\boldsymbol{S}\boldsymbol{\Phi}^{-1/2})\tilde{\boldsymbol{T}} = \tilde{\boldsymbol{T}}(\boldsymbol{I}_m + \boldsymbol{M}) \tag{4.89}$$

其中$\tilde{\boldsymbol{T}} = \boldsymbol{\Phi}^{-1/2}\boldsymbol{A}\boldsymbol{Q}\boldsymbol{M}^{-1/2}$是一个$p \times m'$维的准正交矩阵。将$\boldsymbol{\Omega}$重新写为

$$\boldsymbol{\Omega} = \boldsymbol{\Phi}^{1/2}(\boldsymbol{I} + \boldsymbol{\Phi}^{-1/2}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1/2})\boldsymbol{\Phi}^{1/2}$$

则$\boldsymbol{\Omega}$行列式可以表示为

$$|\boldsymbol{\Omega}| = |\boldsymbol{\Omega}||\boldsymbol{I} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\boldsymbol{A}|$$

$\boldsymbol{\Omega}$逆矩阵可以表示为

$$\boldsymbol{\Omega}^{-1} = \boldsymbol{\Phi}^{-1} - \boldsymbol{\Phi}^{-1}\boldsymbol{A}(\boldsymbol{I} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}$$

利用式(4.87)有

$$\begin{aligned}\boldsymbol{\Omega}^{-1}\boldsymbol{S} &= \boldsymbol{\Phi}^{-1}\boldsymbol{S} - \boldsymbol{\Phi}^{-1}\boldsymbol{A}(\boldsymbol{I} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\boldsymbol{A})^{-1}(\boldsymbol{I} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\boldsymbol{A})\boldsymbol{A}^{\mathrm{T}}\\ &= \boldsymbol{\Phi}^{-1}\boldsymbol{S} - \boldsymbol{\Phi}^{-1}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\end{aligned}$$

由式(4.86)有

$$\begin{aligned}f(\boldsymbol{A}) &= \log|\boldsymbol{\Omega}| + \mathrm{tr}\boldsymbol{\Omega}^{-1}\boldsymbol{S}\\ &= \log|\boldsymbol{\Phi}| + \log|\boldsymbol{I} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\boldsymbol{A}| + \mathrm{tr}\boldsymbol{\Phi}^{-1}\boldsymbol{S} - \mathrm{tr}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\boldsymbol{A}\\ &= p + \log|\boldsymbol{S}| + \left[\mathrm{tr}(\boldsymbol{\Phi}^{-1/2}\boldsymbol{S}\boldsymbol{\Phi}^{-1/2}) - \log|\boldsymbol{\Phi}^{-1/2}\boldsymbol{S}\boldsymbol{\Phi}^{-1/2}| - p\right]\\ &\quad - \left[\mathrm{tr}(\boldsymbol{I} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\boldsymbol{A}) - \log|\boldsymbol{I} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\boldsymbol{A}| - m\right]\\ &= p + \log|\boldsymbol{S}| + \sum_{i=1}^{p}(\lambda_i - \log\lambda_i - 1) - \sum_{j=1}^{m}(\nu_j - \log\nu_j - 1)\end{aligned}$$

其中$\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p$是$\boldsymbol{\Phi}^{-1/2}\boldsymbol{S}\boldsymbol{\Phi}^{-1/2}$的特征值，$\nu_1 \geqslant \nu_2 \geqslant \ldots \geqslant \nu_m$是$\boldsymbol{I}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\boldsymbol{A}$的特征值。由式(4.88)和式(4.89)我们知道$\nu_1,\nu_2,\ldots,\nu_{m'}$也是$\boldsymbol{\Phi}^{-1/2}\boldsymbol{S}\boldsymbol{\Phi}^{-1/2}$的特征值，其余特征值$\nu_{m'+1},\ldots,\nu_m$为1。若要最小化$f$，选择$\nu_1,\nu_2,\ldots,\nu_{m'}$尽可能大，即可选择$\nu_1=\lambda_1,\ldots,\nu_{m'}=\lambda_{m'},\nu_{m'+1}=1,\ldots,\nu_m=1$。于是$f$可以简化为

$$f(\boldsymbol{A})=p+\log|\boldsymbol{S}|+\sum_{i=m'+1}^{p}(\lambda_i-\log\lambda_i-1)$$

可以选择$m'=m$则可以得到f的最小值

$$f(\boldsymbol{A})=p+\log|\boldsymbol{S}|+\sum_{i=m+1}^{p}(\lambda_i-\log\lambda_i-1)$$

当$m'=m$时，$\boldsymbol{Q}$是一个正交矩阵，$\boldsymbol{T}=\tilde{\boldsymbol{T}}=\boldsymbol{\Phi}^{-1/2}\boldsymbol{A}\boldsymbol{Q}\boldsymbol{M}^{-1/2}$，$\boldsymbol{\Lambda}=\boldsymbol{I}+\boldsymbol{M}$，所以有

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{\Phi}^{1/2}\boldsymbol{T}(\boldsymbol{\Lambda}-\boldsymbol{I})\boldsymbol{T}^{\mathrm{T}}\boldsymbol{\Phi}^{1/2}$$

则$\boldsymbol{A}$可以选择为

$$\boldsymbol{A}=\boldsymbol{\Phi}^{1/2}\boldsymbol{T}(\boldsymbol{\Lambda}-\boldsymbol{I})^{1/2}$$

4.27 令$p>1$，$a_i\geqslant 0, i=1,2,\cdots,n$，证明：对每一组满足$\sum\limits_{i=1}^{n}x_i^q=1$ $(q=p/(p-1))$的非负实数$x_1,x_2,\cdots,x_n$，不等式

$$\sum_{i=1}^{n}a_ix_i\leqslant\left(\sum_{i=1}^{n}a_i^p\right)^{1/p}$$

成立。这一不等式称为$\left(\sum\limits_{i=1}^{n}a_i^p\right)^{1/p}$的表示定理[1, p.218]，并且等号成立，当且仅当$a_1=a_2=\cdots=a_n=0$或者$x_i^q=a_i^p\left(\sum\limits_{k=1}^{n}a_k^p\right)^{-1}$，$i=1,2,\cdots,n$。

证明： 考虑优化问题

$$\max\ \sum_{i=1}^{n}a_ix_i$$

$$s.t.\quad \sum_{i=1}^{n}x_i^q=1$$

构造拉格朗日函数

$$\psi(x)=\sum_{i=1}^{n}a_ix_i-\lambda q^{-1}\left(\sum_{i=1}^{n}x_i^q-1\right)$$

对其进行微分，可得

$$\mathrm{d}\psi(x)=\sum_{i=1}^{n}(a_i-\lambda x_i^{q-1})\mathrm{d}x_i$$

由上式可得一阶条件：

$$\lambda x_i^{q-1} = a_i, \qquad i = 1, \cdots, n$$
$$\sum_{i=1}^{n} x_i^q = 1$$

求解 x_i 和 λ 可得

$$x_i = \left(a_i^p / \sum_i a_i^p\right)^{1/q}, \qquad i = 1, \cdots, n$$
$$\lambda = \left(\sum_i a_i^p\right)^{1/p}$$

由于$q > 1$, $\psi(x)$是凹函数，所以$\sum a_i x_i$的最大值可以表示为

$$\sum_i a_i \left(a_i^p / \sum_i a_i^p\right)^{1/q} = \sum_i a_i^p / \left(\sum_i a_i^p\right)^{1/q} = \left(\sum_i a_i^p\right)^{1/p}$$

4.28 考虑M个实谐波信号的Pisarenko谐波分解的下列推广[2]。令噪声子空间的维数大于1，于是张成噪声子空间的矩阵$\boldsymbol{V}_n$ 的每一个列向量的元素都满足

$$\sum_{k=0}^{2M} v_k \mathrm{e}^{\mathrm{j}\,\omega_i k} = \sum_{k=0}^{2M} v_k \mathrm{e}^{-\mathrm{j}\,\omega_i k} = 0, \quad 1 \leqslant i \leqslant M$$

令

$$\bar{\boldsymbol{p}} = \boldsymbol{V}_n \boldsymbol{\alpha} = [\bar{p}_0, \bar{p}_1, \cdots, \bar{p}_{2M}]^{\mathrm{T}}$$

表示$\boldsymbol{V}_n$ 的列向量的非退化线性组合。所谓非退化，乃是指由向量$\bar{\boldsymbol{p}}$ 的元素构造的多项式$p(z)$ 至少具有$2M$ 阶，即$p(z) = \bar{p}_0 + \bar{p}_1 z + \cdots + \bar{p}_{2M} z^{2M}$，$\bar{p}_{2M} \neq 0$。于是，这一多项式也满足上面的式子。这意味着，所有谐波频率均可由多项式$p(z)$位于单位圆上的$2M$个根求出。现在希望选择系数向量$\boldsymbol{\alpha}$满足条件：$p_0 = 1$和$\sum_{k=1}^{K} p_k^2 = \min$。

(1) 令$\boldsymbol{v}^{\mathrm{T}}$是矩阵$\boldsymbol{V}_n$的第一行，而$\boldsymbol{V}$是由$\boldsymbol{V}_n$的其他所有行组成的矩阵。若$\boldsymbol{p}$是由$\bar{\boldsymbol{p}}$除第一个元素以外的其他元素组成的向量，试证明

$$\boldsymbol{\alpha} = \arg\ \min \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{V}^{\mathrm{T}} \boldsymbol{V} \boldsymbol{\alpha}$$

约束条件为$\boldsymbol{v}^{\mathrm{T}} \boldsymbol{\alpha} = 1$。

(2) 利用Làgrange乘子法证明约束优化问题的解为

$$\boldsymbol{\alpha} = \frac{(\boldsymbol{V}^{\mathrm{T}} \boldsymbol{V})^{-1} \boldsymbol{v}}{\boldsymbol{v}^{\mathrm{T}} (\boldsymbol{V}^{\mathrm{T}} \boldsymbol{V})^{-1} \boldsymbol{v}}, \qquad \boldsymbol{p} = \frac{\boldsymbol{V} (\boldsymbol{V}^{\mathrm{T}} \boldsymbol{V})^{-1} \boldsymbol{v}}{\boldsymbol{v}^{\mathrm{T}} (\boldsymbol{V}^{\mathrm{T}} \boldsymbol{V})^{-1} \boldsymbol{v}}$$

解：(1) 由题意知，张成噪声子空间的矩阵$\boldsymbol{V}_n$分块为

$$\boldsymbol{V}_n = \begin{bmatrix} \boldsymbol{v}^{\mathrm{T}} \\ \boldsymbol{V} \end{bmatrix}$$

于是

$$\bar{\boldsymbol{p}} = \boldsymbol{V}_n\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{v}^{\mathrm{T}} \\ \boldsymbol{V} \end{bmatrix}\boldsymbol{\alpha} = \begin{bmatrix} \bar{p}_0 \\ \bar{p}_1 \\ \vdots \\ \bar{p}_{2M} \end{bmatrix}$$

即有

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{\alpha} = \bar{p}_0$$
$$\boldsymbol{V}\boldsymbol{\alpha} = \begin{bmatrix} \bar{p}_1 \\ \vdots \\ \bar{p}_{2M} \end{bmatrix}$$

由于$\boldsymbol{p}$是由$\bar{\boldsymbol{p}}$除第一个元素以外的其他元素组成的向量.故

$$\boldsymbol{p} \overset{\text{def}}{=} [p_1, \cdots, p_{2M}]^{\mathrm{T}} = [\bar{p}_1, \cdots, \bar{p}_{2M}]^{\mathrm{T}}$$

进而有

$$\boldsymbol{V}\boldsymbol{\alpha} = \boldsymbol{p} = [p_1, \cdots, p_{2M}]^{\mathrm{T}}$$

由于希望选择$\boldsymbol{\alpha}$使得$\bar{\boldsymbol{p}} = \boldsymbol{V}_n\boldsymbol{\alpha}$的元素满足$\bar{p}_0 = 1$和$\sum\limits_{k=1}^{2M}\bar{p}_k^2 = \min$，而$\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\alpha} = \sum\limits_{k=1}^{2M}\bar{p}_k^2$，所以$\boldsymbol{\alpha}$是下列优化问题的解：

$$\boldsymbol{\alpha} = \arg\min\left(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\alpha}\right) = \min$$

约束条件为

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{\alpha} = 1 \tag{4.90}$$

(2) 利用Lagrange乘子法,构造目标函数

$$J(\boldsymbol{\alpha}, \lambda) = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\alpha} + \lambda(1 - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{\alpha})$$

由$\dfrac{\partial J(\boldsymbol{\alpha}, \lambda)}{\partial \boldsymbol{\alpha}} = 0$得$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\alpha} - \lambda\boldsymbol{v} = 0$，即有

$$\boldsymbol{\alpha} = \lambda(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})^{-1}\boldsymbol{v} \tag{4.91}$$

将式(4.91)代入约束条件式(4.90)，可求得

$$\lambda\boldsymbol{v}^{\mathrm{T}}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})^{-1}\boldsymbol{v} = 1$$

或

$$\lambda = \frac{1}{\boldsymbol{v}^{\mathrm{T}}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})^{-1}\boldsymbol{v}}$$

从而有结论

$$\boldsymbol{\alpha} = \frac{(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})^{-1}\boldsymbol{v}}{\boldsymbol{v}^{\mathrm{T}}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})^{-1}\boldsymbol{v}}$$

$$\boldsymbol{p} = \frac{\boldsymbol{V}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})^{-1}\boldsymbol{v}}{\boldsymbol{v}^{\mathrm{T}}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})^{-1}\boldsymbol{v}}$$

4.29 设矩阵$\boldsymbol{X}$的秩$\mathrm{rank}(\boldsymbol{X}) \leqslant r$，证明：对所有满足$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}_r$的半正交矩阵$\boldsymbol{A}$和所有矩阵$\boldsymbol{Z} \in R^{n\times r}$，恒有

$$\min \mathrm{tr}\left((\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{A}^{\mathrm{T}})(\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}\right) = 0$$

证法一： 构造拉格朗日函数

$$f(\boldsymbol{A}, \boldsymbol{Z}, \boldsymbol{L}) = \mathrm{tr}\left((\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{A}^{\mathrm{T}})(\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}\right) + \mathrm{tr}\left(\boldsymbol{L}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} - \boldsymbol{I}_r)\right)$$

目标函数关于$\boldsymbol{Z}$求偏导不难有

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{A}$$

将其代入目标函数有

$$\begin{aligned}\mathrm{tr}\left((\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{A}^{\mathrm{T}})(\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}\right) &= \mathrm{tr}\left(\boldsymbol{X}(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})\boldsymbol{X}^{\mathrm{T}}\right)\\ &= \mathrm{tr}\left(\boldsymbol{X}(\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}})\boldsymbol{X}^{\mathrm{T}}\right)\\ &= \mathrm{tr}\left(\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)\end{aligned}$$

取$\boldsymbol{X}^{\mathrm{T}}$的$\mathrm{rank}(\boldsymbol{X})$个线性无关的特征向量，并将其扩展为$r$个线性无关的向量。将这$r$个向量进行Schmidt正交化。不难验证这$r$个向量秩为$r$且满足半正交矩阵条件。选定$\boldsymbol{A}$为这$r$个向量组成的矩阵，则有

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \in \mathrm{Range}(\boldsymbol{X}^{\mathrm{T}}) \subset \mathrm{Range}(\boldsymbol{A})$$

于是必有

$$\mathrm{tr}\left(\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right) = 0$$

又由于

$$\mathrm{tr}\left((\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{A}^{\mathrm{T}})(\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}\right) \geqslant 0$$

则有

$$\min \mathrm{tr}\left((\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{A}^{\mathrm{T}})(\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}\right) = 0$$

证法二：

证法一是求解带约束问题的一般方法。实际上本题利用奇异值分解可以更简单得到证明。$\boldsymbol{X}$的奇异值分解表示如下：

$$\boldsymbol{X}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}=\begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_r\end{bmatrix}\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_r\end{bmatrix}^{\mathrm{T}}$$

令

$$\boldsymbol{Z}=\begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_r\end{bmatrix}\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r\end{bmatrix}$$

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_r\end{bmatrix}$$

以上选定的$\boldsymbol{A},\boldsymbol{Z}$正好使得

$$\operatorname{tr}\left((\boldsymbol{X}-\boldsymbol{Z}\boldsymbol{A}^{\mathrm{T}})(\boldsymbol{X}-\boldsymbol{Z}\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}\right)=0$$

第 5 章　奇异值分析

本章对矩阵的奇异值分析展开了专题讨论。首先分析了单个矩阵的(普通) 奇异值分解(OSVD)、奇异值的性质以及奇异值分解的数值计算；然后，以两个或者多个矩阵为对象，介绍了各种推广的奇异值分解：

(1) 两个矩阵乘积$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}$的奇异值分解，简称乘积奇异值分解(PSVD)；

(2) 两个矩阵组成的矩阵对$(\boldsymbol{A},\boldsymbol{B})$的广义奇异值分解(GSVD)；

(3) 矩阵三元组$(\boldsymbol{A},\boldsymbol{B},\boldsymbol{C})$的约束奇异值分解(RSVD)；

(4) 矩阵相对于某种分块结构的结构奇异值。

普通奇异值分解、乘积奇异值分解和广义奇异值分解都是约束奇异值分解的特例。

5.1　主要理论与方法

5.1.1　矩阵的奇异值分解

关于矩阵的奇异值分解，我们有下面的定理。

定理 5.1　(矩阵的奇异值分解)　令$\boldsymbol{A} \in R^{m\times n}$(或$C^{m\times n}$)，则存在正交(或酉)矩阵$\boldsymbol{U} \in R^{m\times m}$(或$C^{m\times m}$)和$\boldsymbol{V} \in R^{n\times n}$(或$C^{n\times n}$)使得

$$\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}} \ (\text{或}\boldsymbol{U\Sigma V}^{\mathrm{H}}) \tag{5.1}$$

式中

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix} \tag{5.2}$$

且$\boldsymbol{\Sigma}_1 = \mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)$，其对角元素按照顺序

$$\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0, \qquad r = \mathrm{rank}(\boldsymbol{A}) \tag{5.3}$$

排列。

记$\boldsymbol{U} = [\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_m]$，$\boldsymbol{V} = [\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n]$，则关于矩阵的奇异值分解，有下面汇总的一些常用概念和结论。

(1) $\boldsymbol{V}$的列向量$\boldsymbol{v}_i$ 称为矩阵$\boldsymbol{A}$的右奇异向量，即

$$\boldsymbol{A}\boldsymbol{v}_i = \begin{cases} \sigma_i\boldsymbol{u}_i, & i = 1,2,\cdots,r \\ 0, & i = r+1,r+2,\cdots,n \end{cases} \tag{5.4}$$

类似地，$\boldsymbol{U}$的列向量$\boldsymbol{u}_i$ 称为矩阵$\boldsymbol{A}$的左奇异向量，即

$$\boldsymbol{u}_i^{\mathrm{H}}\boldsymbol{A} = \begin{cases} \sigma_i \boldsymbol{v}_i^{\mathrm{H}}, & i = 1, 2, \cdots, r \\ 0, & i = r+1, r+2, \cdots, n \end{cases} \tag{5.5}$$

(2) 矩阵$\boldsymbol{A}$的奇异值分解式(5.1) 可以改写成向量表达形式：

$$\boldsymbol{A} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{H}} \tag{5.6}$$

这种表达有时称为$\boldsymbol{A}$的并向量(奇异值) 分解。

(3) 当矩阵$\boldsymbol{A}$的秩$r = \mathrm{rank}(\boldsymbol{A}) < \min\{m, n\}$时，由于奇异值$\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_h = 0$，$h = \min\{m, n\}$，故奇异值分解式(5.1)可以简化为

$$\boldsymbol{A} = \boldsymbol{U}_r \boldsymbol{\Sigma}_r \boldsymbol{V}_r^{\mathrm{H}} \tag{5.7}$$

式中

$$\boldsymbol{U}_r = [\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_r], \quad \boldsymbol{V}_r = [\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_r], \quad \boldsymbol{\Sigma}_r = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$$

式(5.7)称为矩阵$\boldsymbol{A}$的截尾奇异值分解或薄奇异值分解。

(4) 如果矩阵$\boldsymbol{A}_{m\times n}$具有秩$r$，则$m \times m$酉矩阵$\boldsymbol{U}$的前$r$列组成矩阵$\boldsymbol{A}$的列空间的标准正交基；$n \times n$酉矩阵$\boldsymbol{V}$的前$r$列组成矩阵$\boldsymbol{A}$的行空间(或$\boldsymbol{A}^{\mathrm{H}}$ 的列空间)的标准正交基；$\boldsymbol{V}$的后$n-r$列组成矩阵$\boldsymbol{A}$的零空间的标准正交基；$\boldsymbol{U}$的后$m-r$列组成矩阵$\boldsymbol{A}^{\mathrm{H}}$的零空间的标准正交基。

(5) 矩阵$\boldsymbol{A}$的谱范数或2-范数等于$\boldsymbol{A}$的最大奇异值，即

$$\|\boldsymbol{A}\|_{\mathrm{spec}} = \sigma_1 \tag{5.8}$$

矩阵$\boldsymbol{A}$的Frobenius范数等于该矩阵所有非零奇异值平方和的正平方根，即

$$\|\boldsymbol{A}\|_{\mathrm{F}} = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2} \tag{5.9}$$

(6) 对于一个$m \times n$矩阵$\boldsymbol{A}$，其条件数可以利用奇异值定义为

$$\mathrm{cond}(\boldsymbol{A}) = \sigma_1/\sigma_p, \qquad p = \min\{m, n\} \tag{5.10}$$

(7) 矩阵$\boldsymbol{A}_{m\times n}$的非零奇异值是$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$或者$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$的非零特征值的正平方根。

(8) 若$\boldsymbol{A} = \boldsymbol{U}\begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}\boldsymbol{V}^{\mathrm{H}}$是$m \times n$矩阵$\boldsymbol{A}$的奇异值分解，则$\boldsymbol{A}$的Moore-Penrose逆矩阵为

$$\boldsymbol{A}^{\dagger} = \boldsymbol{V}\begin{bmatrix} \boldsymbol{\Sigma}_1^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}\boldsymbol{U}^{\mathrm{H}}$$

(9) $m \times n$矩阵$\boldsymbol{A}$的共轭转置$\boldsymbol{A}^{\mathrm{H}}$的奇异值分解为

$$\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{V}\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{H}} \tag{5.11}$$

即矩阵$\boldsymbol{A}$和$\boldsymbol{A}^{\mathrm{H}}$具有完全相同的奇异值。

(10) $\boldsymbol{P}$和$\boldsymbol{Q}$分别为$m\times m$和$n\times n$酉矩阵时，$\boldsymbol{PAQ}^{\mathrm{H}}$ 的奇异值分解由

$$\boldsymbol{PAQ}^{\mathrm{H}}=\tilde{\boldsymbol{U}}\boldsymbol{\Sigma}\tilde{\boldsymbol{V}}^{\mathrm{H}} \tag{5.12}$$

给出，其中$\tilde{\boldsymbol{U}}=\boldsymbol{PU}$，$\tilde{\boldsymbol{V}}=\boldsymbol{QV}$。就是说，矩阵$\boldsymbol{PAQ}^{\mathrm{H}}$与$\boldsymbol{A}$具有相同的奇异值，即奇异值具有酉不变性，但奇异向量不同。

(11) $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$, $\boldsymbol{AA}^{\mathrm{H}}$的奇异值分解分别为

$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{V}\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}},\qquad \boldsymbol{AA}^{\mathrm{H}}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{H}} \tag{5.13}$$

其中

$$\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{\Sigma}=\mathrm{diag}(\sigma_1^2,\sigma_2^2,\cdots,\sigma_r^2,\overbrace{0,\cdots,0}^{n-r个}) \tag{5.14}$$

$$\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathrm{T}}=\mathrm{diag}(\sigma_1^2,\sigma_2^2,\cdots,\sigma_r^2,\overbrace{0,\cdots,0}^{m-r个}) \tag{5.15}$$

注：$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$和$\boldsymbol{AA}^{\mathrm{H}}$均为Hermitian矩阵。Hermitian 矩阵的奇异值分解与特征值分解是一致的。

(12) 在相差最小Frobenius范数的意义下，矩阵$\boldsymbol{A}$的最佳k秩近似为$\boldsymbol{A}_k=\sum_{i=1}^{k}\sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{H}}$。

(13) 若$\boldsymbol{A}$为方阵，则$\boldsymbol{A}$的行列式的绝对值等于$\boldsymbol{A}$的所有奇异值的乘积。若至少有一个奇异值为零，则$\boldsymbol{A}$是奇异的（行列式为零）。

5.1.2 矩阵奇异值分解的推广

下面的定理深刻地阐明了奇异值分解的含义。

定理 5.2 [4] 令$\boldsymbol{A}\in C^{m\times n}$ $(m>n)$的奇异值为

$$\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_n\geqslant 0$$

则

$$\sigma_k=\min_{\boldsymbol{E}\in C^{m\times n}}\{\|\boldsymbol{E}\|_2:\mathrm{rank}(\boldsymbol{A}+\boldsymbol{E})\leqslant k-1\},\qquad k=1,2,\cdots,n \tag{5.16}$$

并且存在一满足$\|\boldsymbol{E}_k\|_2=\sigma_k$的误差矩阵$\boldsymbol{E}$使得

$$\mathrm{rank}(\boldsymbol{A}+\boldsymbol{E}_k)=k-1,\qquad k=1,2,\cdots,n$$

其中$\|\cdot\|_2$表示矩阵的2-范数（或谱范数）。这一结果可以推广到约束奇异值分解的情况。我们有下面的定义。

定义 5.1 (约束奇异值)[28] 令$\boldsymbol{A}\in C^{m\times n},\boldsymbol{B}\in C^{m\times p},\boldsymbol{D}\in C^{p\times q},\boldsymbol{C}\in C^{q\times n}$，矩阵三元组$(\boldsymbol{A},\boldsymbol{B},\boldsymbol{C})$的约束奇异值定义为

$$\sigma_k(\boldsymbol{A},\boldsymbol{B},\boldsymbol{C})=\min_{\boldsymbol{D}\in C^{p\times q}}\{\|\boldsymbol{D}\|_2|\mathrm{rank}(\boldsymbol{A}+\boldsymbol{BDC})\leqslant k-1\},\quad k=1,2,\cdots,n \tag{5.17}$$

约束奇异值分解的特殊情况包括[15]：

(1) 矩阵三元组$(\boldsymbol{A}, \boldsymbol{I}_m, \boldsymbol{I}_n)$的约束奇异值分解就是矩阵$\boldsymbol{A}$的普通奇异值分解；

(2) 矩阵三元组$(\boldsymbol{I}_m, \boldsymbol{B}, \boldsymbol{C})$的约束奇异值分解就是$(\boldsymbol{B}^{\mathrm{H}}, \boldsymbol{C})$的乘积奇异值分解；

(3) 矩阵三元组$(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{I}_n)$的约束奇异值分解就是矩阵束$(\boldsymbol{A}, \boldsymbol{B})$的广义奇异值分解；

(4) 矩阵三元组$(\boldsymbol{A}, \boldsymbol{I}_m, \boldsymbol{C})$的约束奇异值分解就是矩阵束$(\boldsymbol{A}, \boldsymbol{C})$的广义奇异值分解。

关于乘积奇异值分解，有下面的定理。

定理 5.3 (乘积奇异值分解) [16] 令$\boldsymbol{B}^{\mathrm{T}} \in C^{m\times p}$，$\boldsymbol{C} \in C^{p\times n}$，则存在酉矩阵$\boldsymbol{U} \in C^{m\times m}, \boldsymbol{V} \in C^{n\times n}$和非奇异矩阵$\boldsymbol{Q} \in C^{p\times p}$使得

$$\boldsymbol{U}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{I} & & \\ & \boldsymbol{O}_B & \\ & & \boldsymbol{\Sigma}_B \end{bmatrix}, \qquad \boldsymbol{Q}^{-1}\boldsymbol{C}\boldsymbol{V}^{\mathrm{H}} = \begin{bmatrix} \boldsymbol{O}_C & & \\ & \boldsymbol{I} & \\ & & \boldsymbol{\Sigma}_C \end{bmatrix} \tag{5.18}$$

式中

$$\boldsymbol{\Sigma}_B = \mathrm{diag}(s_1, s_2, \cdots, s_r), \qquad 1 > s_1 \geqslant \cdots \geqslant s_r > 0$$
$$\boldsymbol{\Sigma}_C = \mathrm{diag}(t_1, t_2, \cdots, t_r), \qquad 1 > t_1 \geqslant \cdots \geqslant t_r > 0$$

且

$$s_i^2 + t_i^2 = 1, \qquad i = 1, 2, \cdots, r$$

关于广义奇异值分解，有下面的定理。

定理 5.4 (广义奇异值分解1) [40] 若$\boldsymbol{A} \in C^{m\times n}, m \geqslant n$和$\boldsymbol{B} \in C^{p\times n}$，则存在酉矩阵$\boldsymbol{U} \in C^{m\times m}$和$\boldsymbol{V} \in C^{p\times p}$以及非奇异矩阵$\boldsymbol{Q} \in C^{n\times n}$，使得

$$\boldsymbol{U}\boldsymbol{A}\boldsymbol{Q} = \begin{array}{c} \begin{array}{cc} k & n-k \end{array} \\ [\boldsymbol{\Sigma}_A \quad \boldsymbol{O}] \end{array}, \qquad \boldsymbol{\Sigma}_A = \begin{bmatrix} \boldsymbol{I}_r & & \\ & \boldsymbol{S}_A & \\ & & \boldsymbol{O}_A \end{bmatrix} \tag{5.19}$$

$$\boldsymbol{V}\boldsymbol{B}\boldsymbol{Q} = \begin{array}{c} \begin{array}{cc} k & n-k \end{array} \\ [\boldsymbol{\Sigma}_B \quad \boldsymbol{O}] \end{array}, \qquad \boldsymbol{\Sigma}_B = \begin{bmatrix} \boldsymbol{O}_B & & \\ & \boldsymbol{S}_B & \\ & & \boldsymbol{I}_{k-r-s} \end{bmatrix} \tag{5.20}$$

式中

$$\boldsymbol{S}_A = \mathrm{diag}(\alpha_{r+1}, \alpha_{r+2}, \cdots, \alpha_{r+s}), \quad \boldsymbol{S}_B = \mathrm{diag}(\beta_{r+1}, \beta_{r+2}, \cdots, \beta_{r+s}) \tag{5.21}$$

$$\left.\begin{aligned} 1 > \alpha_{r+1} \geqslant \cdots \geqslant \alpha_{r+s} > 0 \\ 0 < \beta_{r+1} \leqslant \cdots \leqslant \beta_{r+s} < 1 \\ \alpha_i^2 + \beta_i^2 = 1, \quad i = r+1, r+2, \cdots, r+s \end{aligned}\right\} \tag{5.22}$$

整数k, r和s分别为

$$k = \mathrm{rank}\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{bmatrix}, \qquad r = \mathrm{rank}\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{bmatrix} - \mathrm{rank}(\boldsymbol{B})$$

和

$$s = \mathrm{rank}(\boldsymbol{A}) + \mathrm{rank}(\boldsymbol{B}) - \mathrm{rank}\begin{bmatrix}\boldsymbol{A}\\ \boldsymbol{B}\end{bmatrix}$$

关于约束奇异值分解，有下面的定理。

定理 5.5 (约束奇异值分解) [28] 令$\boldsymbol{A} \in C^{m\times n}, \boldsymbol{B} \in C^{m\times p}$，$\boldsymbol{C} \in C^{q\times n}$，则存在非奇异矩阵$\boldsymbol{P} \in C^{m\times m}$和$\boldsymbol{Q} \in C^{n\times n}$以及酉矩阵$\boldsymbol{U} \in C^{p\times p}$和$\boldsymbol{V} \in C^{q\times q}$使得

$$\boldsymbol{PAQ} = \begin{array}{cc} \begin{array}{cc} n-t_1 & t_1 \end{array} & \\ \begin{bmatrix} \boldsymbol{\Sigma}_A & \\ & \boldsymbol{O}_A^{(2)} \end{bmatrix} & \begin{array}{c} m-t_1 \\ t_1 \end{array} \end{array}, \quad \boldsymbol{PBU} = \begin{bmatrix} \boldsymbol{\Sigma}_B \\ \boldsymbol{O}_B^{(2)} \end{bmatrix} \begin{array}{c} m \\ m-t_2 \end{array}, \quad \boldsymbol{VCQ} = \begin{array}{c} \begin{array}{cc} n-t_1 & t_1 \end{array} \\ [\boldsymbol{\Sigma}_C,\ \boldsymbol{O}_C^{(2)}] \end{array}$$

成立，其中

$$\boldsymbol{\Sigma}_A = \begin{array}{cc} \begin{array}{ccccc} j & k & l & r & s_1 \end{array} & \\ \begin{bmatrix} \boldsymbol{I}_j & & & & \\ & \boldsymbol{I}_k & & & \\ & & \boldsymbol{I}_l & & \\ & & & \boldsymbol{S}_A & \\ & & & & \boldsymbol{O}_A^{(1)} \end{bmatrix} & \begin{array}{c} j \\ k \\ l \\ r \\ s_2 \end{array} \end{array} \tag{5.23}$$

$$\boldsymbol{\Sigma}_B = \begin{array}{cc} \begin{array}{cccc} j & p-j-r-s_2 & r & s_2 \end{array} & \\ \begin{bmatrix} \boldsymbol{I}_j & & & \\ & \boldsymbol{O}_B^{(1)} & & \\ & & \boldsymbol{S}_B & \\ & & & \boldsymbol{I}_{s_2} \end{bmatrix} & \begin{array}{c} j \\ k+l \\ r \\ s_2 \end{array} \end{array} \tag{5.24}$$

$$\boldsymbol{\Sigma}_C = \begin{array}{cc} \begin{array}{cccc} j+k & l & r & s_1 \end{array} & \\ \begin{bmatrix} \boldsymbol{O}_C^{(1)} & & & \\ & \boldsymbol{I}_l & & \\ & & \boldsymbol{S}_C & \\ & & & \boldsymbol{I}_{s_1} \end{bmatrix} & \begin{array}{c} q-l-r-s_1 \\ l \\ r \\ s_1 \end{array} \end{array} \tag{5.25}$$

式中，$\boldsymbol{S}_A = \mathrm{diag}(\alpha_i), \boldsymbol{S}_B = \mathrm{diag}(\beta_i), \boldsymbol{S}_C = \mathrm{diag}(\gamma_i)$，并且

$$\alpha_i^2 + \beta_i^2 + \gamma_i^2 = 1, \quad i = s+1, s+2, \cdots, s+r \tag{5.26}$$

其中

$$1 > \alpha_i \geqslant \alpha_{i+1} > 0, \quad 0 < \beta_i \leqslant \beta_{i+1} < 1, \quad 1 > \gamma_i \geqslant \gamma_{i+1} > 0 \tag{5.27}$$

和

$$\frac{\alpha_i}{\beta_i\gamma_i} \geqslant \frac{\alpha_{i+1}}{\beta_{i+1}\gamma_{i+1}}, \quad i = s+1, s+2, \cdots, s+r-1 \tag{5.28}$$

其中，$s = j+k+l$。

除了这些推广的奇异值分解外，还有Takagi早在1925年提出的对称复矩阵的奇异值分解[26](现在习惯上称之为Takagi奇异值分解)和Doyle于1982年提出的结构奇异值的概念[25]。

5.2 习题与解答

5.1 已知矩阵

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

通过计算AA^{T}和$A^{\mathrm{T}}A$的特征值和特征向量，求矩阵A的奇异值分解。

解：矩阵AA^{T}为

$$AA^{\mathrm{T}}=\begin{bmatrix}2&2&0\\2&2&0\\0&0&0\end{bmatrix}$$

其特征值为0, 0, 4，相应的特征向量为

$$\begin{bmatrix}-\frac{\sqrt{2}}{2}\\\frac{\sqrt{2}}{2}\\0\end{bmatrix},\quad\begin{bmatrix}0\\0\\1\end{bmatrix},\quad\begin{bmatrix}\frac{\sqrt{2}}{2}\\\frac{\sqrt{2}}{2}\\0\end{bmatrix}$$

矩阵$A^{\mathrm{T}}A$为

$$A^{\mathrm{T}}A=\begin{bmatrix}2&2\\2&2\end{bmatrix}$$

其特征值为0, 4，相应的特征向量为

$$\begin{bmatrix}-\frac{\sqrt{2}}{2}\\\frac{\sqrt{2}}{2}\end{bmatrix},\quad\begin{bmatrix}\frac{\sqrt{2}}{2}\\\frac{\sqrt{2}}{2}\end{bmatrix}$$

由奇异值分解的性质知，矩阵A的奇异值分解为

$$A=\begin{bmatrix}\frac{\sqrt{2}}{2}&\frac{\sqrt{2}}{2}&0\\\frac{\sqrt{2}}{2}&-\frac{\sqrt{2}}{2}&0\\0&0&1\end{bmatrix}\begin{bmatrix}2&0\\0&0\\0&0\end{bmatrix}\begin{bmatrix}\frac{\sqrt{2}}{2}&\frac{\sqrt{2}}{2}\\\frac{\sqrt{2}}{2}&-\frac{\sqrt{2}}{2}\end{bmatrix}^{\mathrm{T}}$$

5.2 计算矩阵

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

的奇异值分解。

解：可通过类似于题5.1的方法求得A的奇异值分解，结果为

$$A=\begin{bmatrix}-\frac{1}{\sqrt{19}}&\sqrt{\frac{18}{19}}&0\\-\frac{3}{\sqrt{19}}&-\frac{1}{\sqrt{38}}&\frac{\sqrt{2}}{2}\\\frac{3}{\sqrt{19}}&\frac{1}{\sqrt{38}}&\frac{\sqrt{2}}{2}\end{bmatrix}\begin{bmatrix}\sqrt{38}&0\\0&0\\0&0\end{bmatrix}\begin{bmatrix}-\frac{\sqrt{2}}{2}&\frac{\sqrt{2}}{2}\\\frac{\sqrt{2}}{2}&\frac{\sqrt{2}}{2}\end{bmatrix}^{\mathrm{T}}$$

5.3 计算矩阵

$$A=\begin{bmatrix}3 & 4 & 5\\ 2 & 1 & 7\end{bmatrix}$$

的奇异值分解。

解：可通过类似于题5.1的方法求得$\boldsymbol{A}$的奇异值分解。在MATLAB或SCILAB数值计算软件中，可通过命令"[u s v] = svd(A)"求得矩阵$\boldsymbol{A}$的奇异值分解，计算结果为

$$\boldsymbol{A}=\begin{bmatrix}-0.6912 & -0.7226\\ -0.7226 & 0.6912\end{bmatrix}\begin{bmatrix}9.8511 & 0 & 0\\ 0 & 2.6373 & 0\end{bmatrix}\begin{bmatrix}-0.3572 & -0.2978 & -0.8853\\ -0.3540 & -0.8339 & 0.4234\\ -0.8643 & 0.4647 & 0.1925\end{bmatrix}^{\mathrm{T}}$$

5.4 已知矩阵

$$\boldsymbol{A}=\begin{bmatrix}-149 & -50 & -154\\ 537 & 180 & 546\\ -27 & 9 & -25\end{bmatrix}$$

求$\boldsymbol{A}$的奇异值以及与最小奇异值σ_1相对应的左、右奇异向量。

解：计算$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$得

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=\begin{bmatrix}48417 & -173097 & 7423\\ -173097 & 618885 & -26529\\ 7423 & -26529 & 1435\end{bmatrix}$$

计算矩阵$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$的特征值得到$\boldsymbol{A}$的奇异值为$\sqrt{6.6844\times 10^5}$, $\sqrt{297.2675}$和$\sqrt{2.9439}$。最小奇异值为$\sigma_3=\sqrt{2.9439}=1.7158$，对应的左和右奇异向量分别是矩阵$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$和$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$对应特征值$\sigma_3^2=2.9439$的右特征向量，计算得

$$\boldsymbol{u}_3=\begin{bmatrix}0.9631\\ 0.2690\\ -0.0099\end{bmatrix},\quad \boldsymbol{v}_3=\begin{bmatrix}0.6938\\ 0.0973\\ -0.7136\end{bmatrix}$$

5.5 令$\boldsymbol{A}=\boldsymbol{x}\boldsymbol{p}^{\mathrm{H}}+\boldsymbol{y}\boldsymbol{q}^{\mathrm{H}}$，其中，$\boldsymbol{x}\perp\boldsymbol{y}$和$\boldsymbol{p}\perp\boldsymbol{q}$。求矩阵$\boldsymbol{A}$的Frobenius范数$\|\boldsymbol{A}\|_{\mathrm{F}}$。(提示：计算$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$，并求$\boldsymbol{A}$的奇异值。)

解法一：由于$\boldsymbol{x}\perp\boldsymbol{y}$和$\boldsymbol{p}\perp\boldsymbol{q}$，容易计算得$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$为

$$\begin{aligned}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} &= (\boldsymbol{p}\boldsymbol{x}^{\mathrm{H}}+\boldsymbol{q}\boldsymbol{y}^{\mathrm{H}})(\boldsymbol{x}\boldsymbol{p}^{\mathrm{H}}+\boldsymbol{y}\boldsymbol{q}^{\mathrm{H}})\\ &= \boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}\boldsymbol{p}\boldsymbol{p}^{\mathrm{H}}+\boldsymbol{y}^{\mathrm{H}}\boldsymbol{y}\boldsymbol{q}\boldsymbol{q}^{\mathrm{H}}\end{aligned}$$

考虑到向量$\boldsymbol{p}$与$\boldsymbol{q}$正交，由$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$的表达式知，$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$有特征值和特征向量：

(1) 特征值$\boldsymbol{p}^{\mathrm{H}}\boldsymbol{p}\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}$，对应的特征向量为$\boldsymbol{p}$；

(2) 特征值$\boldsymbol{q}^{\mathrm{H}}\boldsymbol{q}\boldsymbol{y}^{\mathrm{H}}\boldsymbol{y}$，对应的特征向量为$\boldsymbol{q}$；

(3) 特征值0，对应的特征向量为与$\boldsymbol{p}$和$\boldsymbol{q}$正交的向量。

于是由奇异值与矩阵的范数之间的关系，得$\boldsymbol{A}$的Frobenius范数为

$$\|\boldsymbol{A}\|_{\mathrm{F}}=\sqrt{\boldsymbol{p}^{\mathrm{H}}\boldsymbol{p}\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}+\boldsymbol{q}^{\mathrm{H}}\boldsymbol{q}\boldsymbol{y}^{\mathrm{H}}\boldsymbol{y}}$$

解法二： 我们也可以直接计算矩阵$\boldsymbol{A}$的Frobenius范数如下：

$$
\begin{aligned}
\|\boldsymbol{A}\|_{\mathrm{F}} &= \sqrt{\operatorname{tr}\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\right)} \\
&= \sqrt{\operatorname{tr}\left(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}\boldsymbol{p}\boldsymbol{p}^{\mathrm{H}} + \boldsymbol{y}^{\mathrm{H}}\boldsymbol{y}\boldsymbol{q}\boldsymbol{q}^{\mathrm{H}}\right)} \\
&= \sqrt{\operatorname{tr}\left(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}\boldsymbol{p}\boldsymbol{p}^{\mathrm{H}}\right) + \operatorname{tr}\left(\boldsymbol{y}^{\mathrm{H}}\boldsymbol{y}\boldsymbol{q}\boldsymbol{q}^{\mathrm{H}}\right)} \\
&= \sqrt{\boldsymbol{p}^{\mathrm{H}}\boldsymbol{p}\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x} + \boldsymbol{q}^{\mathrm{H}}\boldsymbol{q}\boldsymbol{y}^{\mathrm{H}}\boldsymbol{y}}
\end{aligned}
$$

5.6 已知$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}$是矩阵$\boldsymbol{A}$的奇异值分解，矩阵$\boldsymbol{A}^{\mathrm{H}}$的奇异值与$\boldsymbol{A}$的奇异值有何关系？

解： 由$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}$可得$\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{V}\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{H}}$，故矩阵$\boldsymbol{A}^{\mathrm{H}}$与矩阵$\boldsymbol{A}$有相同的非零奇异值。

5.7 证明：若$\boldsymbol{A}$为正方矩阵，则$|\det(\boldsymbol{A})|$等于$\boldsymbol{A}$的奇异值之积。

证明： 令$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}$为矩阵$\boldsymbol{A}$的奇异值分解，在等式两边取行列式值并取绝对值得$|\det(\boldsymbol{A})| = \sigma_1\sigma_2\cdots\sigma_n$。

5.8 假定$\boldsymbol{A}$ 为可逆矩阵，求$\boldsymbol{A}^{-1}$的奇异值分解。

解： 令$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}$为矩阵$\boldsymbol{A}$的奇异值分解，由于$\boldsymbol{A}$ 为可逆矩阵，在等式两边同时取逆得$\boldsymbol{A}^{-1} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{H}}$，此即$\boldsymbol{A}^{-1}$的奇异值分解。

5.9 证明：若$\boldsymbol{A}$为$n \times n$正定矩阵，则$\boldsymbol{A}$的奇异值与$\boldsymbol{A}$的特征值相同。

证明： 由于$\boldsymbol{A}$为$n \times n$正定矩阵，故它存在特征值分解

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{H}}$$

其中$\boldsymbol{U}$为酉阵，$\boldsymbol{\Sigma}$为对角线元素都为正的对角阵。显然，$\boldsymbol{A}$的特征值分解也是$\boldsymbol{A}$的奇异值分解，故$\boldsymbol{A}$的奇异值与$\boldsymbol{A}$的特征值相同。

5.10 令$\boldsymbol{A}$ 为$m \times n$矩阵，且$\boldsymbol{P}$为$m \times m$正交矩阵。证明$\boldsymbol{P}\boldsymbol{A}$与$\boldsymbol{A}$的奇异值相同。矩阵$\boldsymbol{P}\boldsymbol{A}$与$\boldsymbol{A}$ 的左、右奇异向量有何关系？

证明： 设$\boldsymbol{A}$的奇异值分解为

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$$

显然，矩阵$\boldsymbol{P}\boldsymbol{A}$的奇异值分解为

$$\boldsymbol{P}\boldsymbol{A} = \boldsymbol{P}\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$$

观察知$\boldsymbol{P}\boldsymbol{A}$与$\boldsymbol{A}$的奇异值相同，且$\boldsymbol{P}\boldsymbol{A}$与$\boldsymbol{A}$的右奇异向量也相同。$\boldsymbol{P}\boldsymbol{A}$的左奇异向量为正交矩阵$\boldsymbol{P}\boldsymbol{U}$的列向量，而$\boldsymbol{A}$的左奇异向量为正交矩阵$\boldsymbol{U}$的列向量。

5.11 令$\boldsymbol{A}$ 是一个$m\times n$矩阵，并且$\lambda_1,\lambda_2,\cdots,\lambda_n$ 是矩阵$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$的特征值，相对应的特征向量为$\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_n$。证明$\boldsymbol{A}$的奇异值$\sigma_i$等于范数$\|\boldsymbol{A}\boldsymbol{u}_i\|$，即

$$\sigma_i=\|\boldsymbol{A}\boldsymbol{u}_i\|,\qquad i=1,2,\cdots,n$$

证明：由于$\lambda_i=\sigma_i^2$，我们有

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u}_i=\sigma_i^2\boldsymbol{u}_i$$

在上式两边同时左乘$\boldsymbol{u}_i^{\mathrm{T}}$得

$$(\boldsymbol{A}\boldsymbol{u}_i)^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u}_i=\sigma_i^2\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{u}_i$$

假设特征向量$\boldsymbol{u}_i$为单位长度向量，即得

$$\sigma_i=\|\boldsymbol{A}\boldsymbol{u}_i\|$$

5.12 令$\lambda_1,\lambda_2,\cdots,\lambda_n$和$\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_n$分别是矩阵$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$的特征值和特征向量。假定矩阵$\boldsymbol{A}$有$r$个非零的奇异值，证明$\{\boldsymbol{A}\boldsymbol{u}_1,\boldsymbol{A}\boldsymbol{u}_2,\cdots,\boldsymbol{A}\boldsymbol{u}_r\}$是列空间$\mathrm{Col}(\boldsymbol{A})$的一组正交基，并且$\mathrm{rank}(\boldsymbol{A})=r$。

证明：易知$\boldsymbol{A}$的秩为r，且$\boldsymbol{u}_i$为$\boldsymbol{A}$的右奇异向量，设$\boldsymbol{A}$有薄奇异值分解

$$\boldsymbol{A}=\boldsymbol{V}_r\boldsymbol{\Sigma}_r\boldsymbol{U}_r^{\mathrm{T}}$$

其中$\boldsymbol{U}_r=[\boldsymbol{u}_1,\cdots,\boldsymbol{u}_r]$，$\boldsymbol{V}_r=[\boldsymbol{v}_1,\cdots,\boldsymbol{v}_r]$，$\boldsymbol{\Sigma}_r=[\sqrt{\lambda_1},\cdots,\sqrt{\lambda_r}]$。

$\{\boldsymbol{A}\boldsymbol{u}_1,\boldsymbol{A}\boldsymbol{u}_2,\cdots,\boldsymbol{A}\boldsymbol{u}_r\}$的正交性可由Hermitian矩阵的特征值分解的性质得到。事实上，$(\boldsymbol{A}\boldsymbol{u}_i)^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u}_j=\lambda_i\lambda_j(\boldsymbol{u}_i)^{\mathrm{T}}\boldsymbol{u}_j=\lambda_i^2\delta_{i,j}$。

由于$\boldsymbol{A}\boldsymbol{u}_i=\sqrt{\lambda_i}\boldsymbol{v}_i$，有

$$\mathrm{Span}\{\boldsymbol{A}\boldsymbol{u}_1,\cdots,\boldsymbol{A}\boldsymbol{u}_r\}=\mathrm{Span}\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_r\}$$

故只需证明$\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_r\}$张成矩阵$\boldsymbol{A}$的列空间即可，这可验证如下：

$$\begin{aligned}\mathrm{Col}(\boldsymbol{A}) &= \{\boldsymbol{y}:\quad \boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}\}\\ &= \{\boldsymbol{y}:\quad \boldsymbol{y}=\boldsymbol{V}_r\boldsymbol{\Sigma}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{x}\}\\ &= \{\boldsymbol{y}:\quad \boldsymbol{y}=\sum_{i=1}^{r}\sqrt{\lambda_i}\boldsymbol{v}_i\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{x}\}\\ &= \{\boldsymbol{y}:\quad \boldsymbol{y}=\sum_{i=1}^{r}\left(\sqrt{\lambda_i}\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{x}\right)\boldsymbol{v}_i\}\end{aligned}$$

5.13 令$\boldsymbol{B},\boldsymbol{C}\in R^{m\times n}$，求复矩阵$\boldsymbol{A}=\boldsymbol{B}+\mathrm{j}\boldsymbol{C}$与实分块矩阵$\begin{bmatrix}\boldsymbol{B} & -\boldsymbol{C}\\ \boldsymbol{C} & \boldsymbol{B}\end{bmatrix}$的奇异值和奇异向量之间的关系。

解：设$\boldsymbol{A}$的奇异值分解为

$$\boldsymbol{A}=\boldsymbol{U\Sigma V}^{\mathrm{H}}$$

令酉阵$\boldsymbol{U}$的实部和虚部为$\boldsymbol{U}_R$和$\boldsymbol{U}_I$，即

$$\boldsymbol{U}=\boldsymbol{U}_R+\mathrm{j}\boldsymbol{U}_I$$

同样，可将酉阵$\boldsymbol{V}$表示成

$$\boldsymbol{V}=\boldsymbol{V}_R+\mathrm{j}\boldsymbol{V}_I$$

这样得到等式

$$\boldsymbol{B}+\mathrm{j}\boldsymbol{C}=(\boldsymbol{U}_R+\mathrm{j}\boldsymbol{U}_I)\boldsymbol{\Sigma}(\boldsymbol{V}_R^{\mathrm{T}}-\mathrm{j}\boldsymbol{V}_I^{\mathrm{T}})$$

对比上式两边的实部和虚部，分别令其相等，得等式

$$\boldsymbol{B}=\boldsymbol{U}_R\boldsymbol{\Sigma V}_R^{\mathrm{T}}+\boldsymbol{U}_I\boldsymbol{\Sigma V}_I^{\mathrm{T}}$$

$$\boldsymbol{C}=-\boldsymbol{U}_R\boldsymbol{\Sigma V}_I^{\mathrm{T}}+\boldsymbol{U}_I\boldsymbol{\Sigma V}_R^{\mathrm{T}}$$

由此，可将分块矩阵$\begin{bmatrix}\boldsymbol{B} & -\boldsymbol{C}\\ \boldsymbol{C} & \boldsymbol{B}\end{bmatrix}$表示成

$$\begin{bmatrix}\boldsymbol{B} & -\boldsymbol{C}\\ \boldsymbol{C} & \boldsymbol{B}\end{bmatrix}=\begin{bmatrix}\boldsymbol{U}_R & -\boldsymbol{U}_I\\ \boldsymbol{U}_I & \boldsymbol{U}_R\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{\Sigma}\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_R & -\boldsymbol{V}_I\\ \boldsymbol{V}_I & \boldsymbol{V}_R\end{bmatrix}^{\mathrm{T}}$$

另一方面，考虑到$\boldsymbol{U}$为酉阵，故

$$(\boldsymbol{U}_R+\mathrm{j}\boldsymbol{U}_I)(\boldsymbol{U}_R^{\mathrm{T}}-\mathrm{j}\boldsymbol{U}_I^{\mathrm{T}})=\boldsymbol{I}$$

比较等式两边的实部和虚部，令其分别相等，得等式

$$\boldsymbol{U}_R\boldsymbol{U}_R^{\mathrm{T}}+\boldsymbol{U}_I\boldsymbol{U}_I^{\mathrm{T}}=\boldsymbol{I}$$

$$-\boldsymbol{U}_R\boldsymbol{U}_I^{\mathrm{T}}+\boldsymbol{U}_I\boldsymbol{U}_R^{\mathrm{T}}=\boldsymbol{O}$$

由此可验证分块矩阵$\begin{bmatrix}\boldsymbol{U}_R & -\boldsymbol{U}_I\\ \boldsymbol{U}_I & \boldsymbol{U}_R\end{bmatrix}$为正交矩阵，即

$$\begin{bmatrix}\boldsymbol{U}_R & -\boldsymbol{U}_I\\ \boldsymbol{U}_I & \boldsymbol{U}_R\end{bmatrix}\begin{bmatrix}\boldsymbol{U}_R & -\boldsymbol{U}_I\\ \boldsymbol{U}_I & \boldsymbol{U}_R\end{bmatrix}^{\mathrm{T}}=\begin{bmatrix}\boldsymbol{U}_R\boldsymbol{U}_R^{\mathrm{T}}+\boldsymbol{U}_I\boldsymbol{U}_I^{\mathrm{T}} & \boldsymbol{U}_R\boldsymbol{U}_I^{\mathrm{T}}-\boldsymbol{U}_I\boldsymbol{U}_R^{\mathrm{T}}\\ \boldsymbol{U}_I\boldsymbol{U}_R^{\mathrm{T}}-\boldsymbol{U}_R\boldsymbol{U}_I^{\mathrm{T}} & \boldsymbol{U}_I\boldsymbol{U}_I^{\mathrm{T}}+\boldsymbol{U}_R\boldsymbol{U}_R^{\mathrm{T}}\end{bmatrix}=\begin{bmatrix}\boldsymbol{I} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{I}\end{bmatrix}$$

同样可验证分块矩阵$\begin{bmatrix}\boldsymbol{V}_R & -\boldsymbol{V}_I\\ \boldsymbol{V}_I & \boldsymbol{V}_R\end{bmatrix}$也为正交矩阵。故分解式

$$\begin{bmatrix}\boldsymbol{B} & -\boldsymbol{C}\\ \boldsymbol{C} & \boldsymbol{B}\end{bmatrix}=\begin{bmatrix}\boldsymbol{U}_R & -\boldsymbol{U}_I\\ \boldsymbol{U}_I & \boldsymbol{U}_R\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{\Sigma}\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_R & -\boldsymbol{V}_I\\ \boldsymbol{V}_I & \boldsymbol{V}_R\end{bmatrix}^{\mathrm{T}}$$

即为实分块矩阵$\begin{bmatrix}\boldsymbol{B} & -\boldsymbol{C}\\ \boldsymbol{C} & \boldsymbol{B}\end{bmatrix}$的奇异值分解，由此不难得到复矩阵$\boldsymbol{A}=\boldsymbol{B}+\mathrm{j}\boldsymbol{C}$与实分块矩阵$\begin{bmatrix}\boldsymbol{B} & -\boldsymbol{C}\\ \boldsymbol{C} & \boldsymbol{B}\end{bmatrix}$的奇异值和奇异向量之间的关系。

5.14 用矩阵$\boldsymbol{A}\in R^{m\times n}$ $(m\geqslant n)$的奇异向量表示$\begin{bmatrix}\boldsymbol{O} & \boldsymbol{A}^{\mathrm{T}}\\ \boldsymbol{A} & \boldsymbol{O}\end{bmatrix}$的特征向量。

解：令$\boldsymbol{A}$的奇异值分解为$\boldsymbol{A}=\boldsymbol{U\Sigma V}^{\mathrm{T}}$，即$\boldsymbol{A}=\sum_{i=1}^{n}\sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}}$，这里$\sigma_i$为对角矩阵$\boldsymbol{\Sigma}$的第$i$个对角线元素，$\boldsymbol{u}_i$和$\boldsymbol{v}_i$分别为$\boldsymbol{U}$和$\boldsymbol{V}$的第$i$个列向量。这时，对$1\leqslant i\leqslant n$，容易验证：

$$\begin{bmatrix}\boldsymbol{O} & \boldsymbol{A}^{\mathrm{T}}\\ \boldsymbol{A} & \boldsymbol{O}\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_i\\ \boldsymbol{u}_i\end{bmatrix}=\sigma_i\begin{bmatrix}\boldsymbol{v}_i\\ \boldsymbol{u}_i\end{bmatrix}$$

和

$$\begin{bmatrix}\boldsymbol{O} & \boldsymbol{A}^{\mathrm{T}}\\ \boldsymbol{A} & \boldsymbol{O}\end{bmatrix}\begin{bmatrix}-\boldsymbol{v}_i\\ \boldsymbol{u}_i\end{bmatrix}=-\sigma_i\begin{bmatrix}-\boldsymbol{v}_i\\ \boldsymbol{u}_i\end{bmatrix}$$

即$\begin{bmatrix}\boldsymbol{O} & \boldsymbol{A}^{\mathrm{T}}\\ \boldsymbol{A} & \boldsymbol{O}\end{bmatrix}$有$2n$个特征向量

$$\begin{bmatrix}\boldsymbol{v}_1\\ \boldsymbol{u}_1\end{bmatrix},\cdots,\begin{bmatrix}\boldsymbol{v}_n\\ \boldsymbol{u}_n\end{bmatrix},\begin{bmatrix}-\boldsymbol{v}_1\\ \boldsymbol{u}_1\end{bmatrix},\cdots,\begin{bmatrix}-\boldsymbol{v}_n\\ \boldsymbol{u}_n\end{bmatrix}$$

考虑到$\begin{bmatrix}\boldsymbol{O} & \boldsymbol{A}^{\mathrm{T}}\\ \boldsymbol{A} & \boldsymbol{O}\end{bmatrix}$的秩最多是$2n$，还剩下$m-n$个零特征值对应的特征向量，它们与以上列出的$2n$个特征向量正交，但具体形式并不唯一，例如，可取为

$$\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{u}_j\end{bmatrix},\quad n+1\leqslant j\leqslant m$$

注释：有时，我们通过计算增广矩阵$\begin{bmatrix}\boldsymbol{O} & \boldsymbol{A}^{\mathrm{T}}\\ \boldsymbol{A} & \boldsymbol{O}\end{bmatrix}$的特征值分解来计算矩阵$\boldsymbol{A}$的奇异值分解。

5.15 利用MATLAB函数[U, S, V] = svd(**X**)求解方程$\boldsymbol{Ax}=\boldsymbol{b}$，其中

$$\boldsymbol{A}=\begin{bmatrix}1 & 1 & 1\\ 3 & 1 & 3\\ 1 & 0 & 1\\ 2 & 2 & 1\end{bmatrix},\qquad \boldsymbol{b}=\begin{bmatrix}1\\ 4\\ 3\\ 2\end{bmatrix}$$

解：设$\boldsymbol{A}$的奇异值分解为$\boldsymbol{A}=\boldsymbol{USV}^{\mathrm{T}}$，则可将原方程写为

$$\boldsymbol{USV}^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{b}$$

令向量$\boldsymbol{c}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{b}$，向量$\boldsymbol{y}=\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}$，则可进一步将原方程简化为

$$\boldsymbol{Sy}=\boldsymbol{c}$$

通常这是一个矛盾方程，只能得到它的一个近似解

$$\boldsymbol{y}^*=[\boldsymbol{S}(1:r,1:r)]^{-1}\boldsymbol{c}(1:r)$$

这里假设只有前r个奇异值大于零，采用MATLAB的记法，$\boldsymbol{S}(1:r,1:r)$表示取矩阵$\boldsymbol{S}$的前r行和前r列个元素，$\boldsymbol{c}(1:r)$表示取向量$\boldsymbol{c}$的前r个元素。由此解得$\boldsymbol{x}$的一个近似解得

$$\boldsymbol{x}^*=\boldsymbol{Vy}^*$$

以下是求解方程的MATLAB代码：

```
% 系数A和b
A = [1 1 1 ; 3 1 3 ; 1 0 1 ; 2 2 1];
b = [1; 4; 3; 2];

% 方程求解
[U S V] = svd(A);
c = U' * b;
d = c(1:3);
x = V * ...
    S(1:3, 1:3)^(-1) * d
```

比较与用矩阵伪逆“ x = pinv(A) * b ”求得的结果，发现该方法求得的结果与用矩阵伪逆求得的结果是相同的。

5.16　假定计算机仿真的观测数据由

$$x(n) = \sqrt{20}\sin(2\pi 0.2n) + \sqrt{2}\sin(2\pi 0.215n) + w(n)$$

产生，其中，$w(n)$是一高斯白噪声，其均值为0，方差为1，并取$n = 1, 2, \cdots, 128$。试针对10次独立的仿真实验数据，分别确定自相关矩阵

$$\boldsymbol{R} = \begin{bmatrix} r(0) & r(-1) & \cdots & r(-2p) \\ r(1) & r(0) & \cdots & r(-2p+1) \\ \vdots & \vdots & & \vdots \\ r(M) & r(M-1) & \cdots & r(M-2p) \end{bmatrix}$$

的有效秩。式中，$r(k) = \dfrac{1}{128}\sum_{i=1}^{128-k} x(i)x(i+k)$表示观测信号的样本自相关函数(未知的观测数据皆令其等于0)，并取$M = 50$，$p = 10$。

解：我们采用教材P355的归一化奇异值方法来确定自相关矩阵的秩。即秩rank确定为

$$\text{rank} = \max_i \frac{\sigma_i}{\sigma_1} \geqslant \epsilon$$

这里取$\epsilon = 0.05$。MATLAB程序代码设计如下：

```
T = 128;
M = 50;
p = 10;
n = 1 : T;
x = sqrt(20)*sin(2*pi*0.2*n) ...
    + sqrt(2)*sin(2*pi*0.215*n) ...
    + randn(1, T);
r = xcorr(x)/T;
R = zeros(M+1, 2*p+1);
for m = 0 : M
    R(m+1, :) = r(T+m:-1:T+m-2*p);
end

[u s v] = svd(R);
d = diag(s);
d = d / d(1);
e = 0.05;
for rank = ...
    1 : 2*p+1
    if d(rank) / d(1) < e
        break;
    else
        continue;
    end
end
rank = rank - 1
```

程序独立的运行10次，相关矩阵的有效秩都正确的确定为4。图5.1画出了一次运行的奇异值分布的情况。

注释：实际中，我们常用有偏的样本自相关估计$\hat{r}(k)$来进行运算，默认的情况下，MATLAB的xcorr函数返回的是有偏的自相关函数估计，因此，我们无需进一步处理自相关函数。

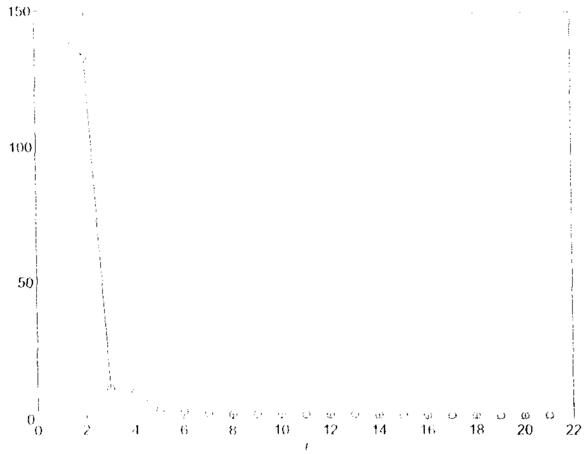


图 5.1 程序一次运行的奇异值σ_j的分布情况

5.17 [4] 使用奇异值分解证明：若$\boldsymbol{A} \in R^{m\times n}$ $(m \geqslant n)$，则存在$\boldsymbol{Q} \in R^{m\times n}$和$\boldsymbol{P} \in R^{n\times n}$，使得$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{P}$，其中$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}_n$，并且$\boldsymbol{P}$是对称的和非负定的。这一分解有时称为极分解(polar decomposition)，因为它与复数分解$z = |z|\mathrm{e}^{\mathrm{j}\arg(z)}$类似。

证明： 设$\boldsymbol{A}$的奇异值分解为$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}$，可将这一分解重写成$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{V}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{\Sigma}\boldsymbol{V}$，令$\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{V}$，$\boldsymbol{P} = \boldsymbol{V}^{\mathrm{H}}\boldsymbol{\Sigma}\boldsymbol{V}$，即得所需的极分解形式。

第6章 总体最小二乘方法

求解$\boldsymbol{Ax} = \boldsymbol{b}$的线性方程组，最小二乘方法只考虑到了数据向量$\boldsymbol{b}$中存在的误差。当系数矩阵$\boldsymbol{A}$也存在误差时，总体最小二乘方法是更好的选择。本章重点给出了总体最小二乘及其推广——约束总体最小二乘和结构总体最小二乘的重要性质。并希望通过习题，使读者对总体最小二乘的基本概念有更加深入的理解。

6.1 主要理论与方法

6.1.1 最小二乘

一、参数的唯一可辨识性

给定$m \times 1$数据向量$\boldsymbol{b}$和$m \times n$数据矩阵$\boldsymbol{A}$，我们希望求解矩阵方程$\boldsymbol{Ax} = \boldsymbol{b}$。如果无噪声存在，则适定方程组(矩阵$\boldsymbol{A}$正定)是一致的，方程存在唯一的解$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$。然而，矩阵方程超定(独立的方程个数大于独立的未知数个数)时，$\boldsymbol{Ax} = \boldsymbol{b}$将不再是一致方程。此时，我们会很自然地想到选择这样一种求解的准则：使误差的平方和

$$\phi = \|\Delta \boldsymbol{b}\|^2 = (\Delta \boldsymbol{b})^{\mathrm{T}} \Delta \boldsymbol{b} = (\boldsymbol{b} - \boldsymbol{Ax})^{\mathrm{T}} (\boldsymbol{b} - \boldsymbol{Ax}) \tag{6.1}$$

最小。这样得到的解$\boldsymbol{x}$称为最小二乘解。

当$m \times n$矩阵$\boldsymbol{A}$具有不同的秩时，上述方程的解有两种不同的情况。

情况1 $\mathrm{rank}(\boldsymbol{A}) = n$

由于$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$非奇异，所以方程有唯一的解：

$$x = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} \tag{6.2}$$

情况2 $\mathrm{rank}(\boldsymbol{A}) < n$

在这种情况下，由$\boldsymbol{x}$的不同解均得到相同的$\boldsymbol{Ax}$值。显而易见，虽然数据向量$\boldsymbol{b}$可以提供有关$\boldsymbol{Ax}$的某些信息，但是却无法区别对应于相同$\boldsymbol{Ax}$值的各个不同的未知参数向量$\boldsymbol{x}$。因此，称这样的参数向量是不可辨识的。更一般地，如果某参数的不同值给出在抽样空间上的相同分布，则称该参数是不可辨识的[21]。

二、Gauss-Markov定理

定理 6.1 (Gauss-Markov定理) 考虑线性方程组

$$\boldsymbol{A\theta} = \boldsymbol{b} + \boldsymbol{e} \tag{6.3}$$

式中，$m \times n$矩阵$\boldsymbol{A}$和$n \times 1$向量$\boldsymbol{\theta}$分别为常数矩阵和常数向量；$\boldsymbol{b}$为$m \times 1$向量，它存在随机误差向量$\boldsymbol{e} = [e_1, e_2, \cdots, e_m]^{\mathrm{T}}$。误差向量的均值向量和协方差矩阵分别为

$$\mathrm{E}\{\boldsymbol{e}\} = \boldsymbol{0}, \qquad \mathrm{cov}(\boldsymbol{e}) = \mathrm{E}\{\boldsymbol{e}\boldsymbol{e}^{\mathrm{H}}\} = \sigma^2 \boldsymbol{I}$$

$n \times 1$参数向量$\boldsymbol{\theta}$的最优无偏解$\hat{\boldsymbol{\theta}}$存在，当且仅当$\mathrm{rank}(\boldsymbol{A}) = n$。此时，最优无偏解由

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{b} \tag{6.4}$$

给出，其方差

$$\mathrm{var}(\hat{\boldsymbol{\theta}}) \leqslant \mathrm{var}(\tilde{\boldsymbol{\theta}}) \tag{6.5}$$

式中，$\tilde{\boldsymbol{\theta}}$是矩阵方程$\boldsymbol{A}\boldsymbol{\theta} = \boldsymbol{b} + \boldsymbol{e}$的任何一个其他解。

6.1.2 总体最小二乘

一、问题描述

考虑的是矩阵方程

$$(\boldsymbol{A} + \boldsymbol{E})\boldsymbol{x} = \boldsymbol{b} + \boldsymbol{e} \tag{6.6}$$

的求解，其中$\boldsymbol{E}$和$\boldsymbol{e}$分别为系数矩阵$\boldsymbol{A}$和数据向量$\boldsymbol{b}$的扰动向量。总体最小二乘方法可以表示为约束最优化问题：

$$\min_{\boldsymbol{D}, \boldsymbol{x}} \|\boldsymbol{D}\|_{\mathrm{F}}^2 \tag{6.7}$$

约束条件为

$$(\boldsymbol{b} + \boldsymbol{e}) \in \mathrm{Range}(\boldsymbol{A} + \boldsymbol{E}) \tag{6.8}$$

式中，$\|\boldsymbol{D}\|_{\mathrm{F}}$是矩阵$\boldsymbol{D}$的Frobenius范数，即

$$\|\boldsymbol{D}\|_{\mathrm{F}} = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij}^2 \right)^{1/2} = \sqrt{\mathrm{tr}(\boldsymbol{D}^{\mathrm{H}}\boldsymbol{D})} \tag{6.9}$$

且约束条件$(\boldsymbol{b} + \boldsymbol{e}) \in \mathrm{Range}(\boldsymbol{A} + \boldsymbol{E})$的涵义为：若$(\boldsymbol{b} + \boldsymbol{e}) \in C^{m \times 1}$，则一定可以找到某个$\boldsymbol{x} \in C^{n \times 1}$使得$\boldsymbol{b} + \boldsymbol{e} = (\boldsymbol{A} + \boldsymbol{E})\boldsymbol{x}$。

二、问题求解

当$m < n + 1$时，方程式(6.7)是欠定的，存在无穷多个解$\boldsymbol{x}$，而TLS方法可给出最小范数解。

在超定方程$(m > n)$的总体最小二乘解中，有两种可能的情况。

情况1　矩阵$\boldsymbol{B}$的奇异值σ_n明显比σ_{n+1}大，即最小的奇异值只有一个。此时，总体最小二乘问题可以归结为：求一具有最小范数平方的扰动矩阵$\boldsymbol{D} \in C^{m \times (n+1)}$使得$\boldsymbol{B} + \boldsymbol{D}$是非满秩的(如果满秩，则只有平凡解$\boldsymbol{z} = \boldsymbol{0}$)。

令$m \times (n + 1)$增广矩阵$\boldsymbol{B}$的奇异值分解为

$$\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}} \tag{6.10}$$

并且其奇异值按照顺序$\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_{n+1}$排列，与这些奇异值对应的右奇异向量为$\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{n+1}$。于是，根据上面的分析，总体最小二乘解为$\boldsymbol{z} = \boldsymbol{v}_{n+1}$。也就是说，原矩阵方程式$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 的最小二乘解由下式给出：

$$\boldsymbol{x}_{\mathrm{TLS}} = \frac{1}{v(1,n+1)} \begin{bmatrix} v(2,n+1) \\ \vdots \\ v(n+1,n+1) \end{bmatrix} \tag{6.11}$$

其中，$v(i,n+1)$是$\boldsymbol{V}$的第$n+1$列的第i个元素。

情况2　矩阵$\boldsymbol{B}$的最小奇异值多重(最后面若干个奇异值重复或非常接近)。

不妨令

$$\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_p > \sigma_{p+1} \approx \cdots \approx \sigma_{n+1} \tag{6.12}$$

且$\boldsymbol{v}_i$是子空间

$$S = \mathrm{Span}\{\boldsymbol{v}_{p+1}, \boldsymbol{v}_{p+2}, \cdots, \boldsymbol{v}_{n+1}\}$$

中的任一列向量，则上述任一右奇异向量$\boldsymbol{v}_i$ 都给出一组总体最小二乘解

$$\boldsymbol{x} = \boldsymbol{y}_i/\alpha_i\,, \qquad i = p+1, p+2, \cdots, n+1$$

其中，α_i是向量$\boldsymbol{v}_i$ 的第一个元素，而其他的元素组成向量$\boldsymbol{y}_i$，也即$\boldsymbol{v}_i = \begin{bmatrix} \alpha_i \\ \boldsymbol{y}_i \end{bmatrix}$。因此，会有$n+1-p$个总体最小二乘解。然而，可以找出在某种意义下唯一的总体最小二乘解。可能的唯一解有两种。

(1) 最小范数解：解向量由n个参数组成。

算法 6.1　(最小范数解的TLS算法)

步骤1　计算增广矩阵$\boldsymbol{B}$的SVD：

$$\boldsymbol{B} = \boldsymbol{U} \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_{n+1} \end{bmatrix} \boldsymbol{V}^{\mathrm{H}}$$

并存储矩阵$\boldsymbol{V}$。

步骤2　决定主奇异值的个数p，即利用$\sigma_p > \sigma_{n+1} + \epsilon \geqslant \sigma_{p+1} \geqslant \cdots \geqslant \sigma_{n+1}$确定$p$，其中，$\epsilon$是某个很小的正数。

步骤3　令$\boldsymbol{V}_1 = [\boldsymbol{v}_{p+1}, \boldsymbol{v}_{p+2}, \cdots, \boldsymbol{v}_{n+1}]$是$\boldsymbol{V}$ 的列分块形式，并计算Householder变换矩阵$\boldsymbol{Q}$使得

$$\boldsymbol{V}_1\boldsymbol{Q} = \left[\begin{array}{c:c} \alpha & 0 \cdots 0 \\ \hdashline \boldsymbol{y} & \times \end{array}\right]$$

其中，α是一个标量，$\times$代表其数值在下一步不起作用的块。

步骤4 若$\alpha \neq 0$，则$\boldsymbol{x}_{\mathrm{TLS}} = \boldsymbol{y}/\alpha$；若$\alpha = 0$，则对原设定的$p$无TLS解，应减小$p$，即使用$p \leftarrow p-1$，并重复以上步骤，直至求出唯一的TLS解。

(2) 最优最小二乘近似解：解向量仅包含p个参数。

算法 6.2 (SVD-TLS算法)

步骤1 计算增广矩阵$\boldsymbol{B}$ 的SVD，并存储右奇异矩阵$\boldsymbol{V}$。

步骤2 确定$\boldsymbol{B}$的有效秩p。

步骤3 计算$(p+1)\times(p+1)$矩阵$\boldsymbol{S}^{(p)}$。

步骤4 求$\boldsymbol{S}^{(p)}$的逆矩阵$\boldsymbol{S}^{-(p)}$，求最优最小二乘近似解。

6.1.3 约束总体最小二乘

定理 6.2 令

$$\boldsymbol{W}_x = \sum_{i=1}^{L} x_i \boldsymbol{G}_i - \boldsymbol{G}_{L+1} \tag{6.13}$$

则约束总体最小二乘的解向量就是满足下列函数极小化的变量$\boldsymbol{x}$:

$$\min_x \begin{bmatrix} \boldsymbol{x} \\ -1 \end{bmatrix}^{\mathrm{H}} \boldsymbol{C}^{\mathrm{H}} (\boldsymbol{W}_x^{\dagger})^{\mathrm{H}} \boldsymbol{W}_x^{\dagger} C \begin{bmatrix} \boldsymbol{x} \\ -1 \end{bmatrix} \tag{6.14}$$

式中，$\boldsymbol{W}_x^{\dagger} = (\boldsymbol{W}_x^{\mathrm{H}} \boldsymbol{W}_x)^{-1} \boldsymbol{W}_x^{\mathrm{H}}$ 是$\boldsymbol{W}_x$的伪逆矩阵。

6.1.4 结构总体最小二乘

定理 6.3 [22] 考虑STLS 问题

$$\min_{\boldsymbol{b}\in R^m,\, \boldsymbol{y}\in R^q} \sum_{i=1}^{m} w_i (a_i - b_i)^2, \qquad 服从约束条件 \begin{cases} \boldsymbol{B}(b)\boldsymbol{y} = \boldsymbol{0} \\ \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} = 1 \end{cases} \tag{6.15}$$

其中，$a_i, i = 1,2,\cdots,m$是数据向量$\boldsymbol{a} \in R^m$的元素；$\boldsymbol{B}(b) = \boldsymbol{B}_0 + \boldsymbol{B}_1 b_1 + \cdots + \boldsymbol{B}_m b_m$，$\boldsymbol{B}_i \in R^{p\times q}(i = 1,2,\cdots,m)$为固定的给定矩阵。STLS问题的求解过程如下。

(1) 求对应于最小标量τ的三元组$(\boldsymbol{u},\tau,\boldsymbol{v})$，其中$\boldsymbol{u} \in R^p, \boldsymbol{v} \in R^q, \tau \in R$，使得

$$\boldsymbol{A}\boldsymbol{v} = \tau \boldsymbol{D}_v \boldsymbol{u}, \qquad \boldsymbol{u}^{\mathrm{T}} \boldsymbol{D}_v \boldsymbol{u} = 1 \tag{6.16}$$

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} = \tau \boldsymbol{D}_u \boldsymbol{v}, \qquad \boldsymbol{v}^{\mathrm{T}} \boldsymbol{D}_u \boldsymbol{v} = 1 \tag{6.17}$$

式中，$\boldsymbol{A} = \boldsymbol{B}_0 + \sum_{i=1}^{m} a_i \boldsymbol{B}_i$；而$\boldsymbol{D}_u$定义为

$$\sum_{i=1}^{m} \boldsymbol{B}_i^{\mathrm{T}} (\boldsymbol{u}^{\mathrm{T}} \boldsymbol{B}_i \boldsymbol{v}) \boldsymbol{u} = \boldsymbol{D}_u \boldsymbol{v}$$

它是一个对称的正定或非负定矩阵，其元素是$\boldsymbol{u}$ 的元素的二次型。类似地，$\boldsymbol{D}_v$ 定义为

$$\sum_{i=1}^{m} \boldsymbol{B}_i^{\mathrm{T}} (\boldsymbol{u}^{\mathrm{T}} \boldsymbol{B}_i \boldsymbol{v}) \boldsymbol{v} = \boldsymbol{D}_v \boldsymbol{u}$$

它是一个对称的正定或非负定矩阵，其元素是$\boldsymbol{v}$ 的元素的二次型。

(2) 利用$\boldsymbol{y}=\boldsymbol{v}/\|\boldsymbol{v}\|$构造向量$\boldsymbol{y}$。

(3) 计算$\boldsymbol{b}$的元素

$$b_k = a_k - \tau \boldsymbol{u}^{\mathrm{T}} \boldsymbol{B}_k \boldsymbol{v}, \qquad k = 1, 2, \cdots, m \tag{6.18}$$

6.2 习题与解答

6.1 考虑线性方程$\boldsymbol{A\theta}+\boldsymbol{\epsilon}=\boldsymbol{x}$，其中，$\boldsymbol{\epsilon}$为加性有色噪声向量，满足条件$\mathrm{E}\{\boldsymbol{\epsilon}\}=\boldsymbol{0}$和$\mathrm{E}\{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}}\}=\boldsymbol{R}$。令$\boldsymbol{R}$已知，并使用加权误差函数$Q(\boldsymbol{\theta})=\boldsymbol{\epsilon}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\epsilon}$作为求参数向量$\boldsymbol{\theta}$最优估计$\hat{\boldsymbol{\theta}}_{\mathrm{WLS}}$的代价函数。这种方法称为加权最小二乘方法。证明

$$\hat{\boldsymbol{\theta}}_{\mathrm{WLS}}=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{x}$$

其中，加权矩阵$\boldsymbol{W}$的最优选择为$\boldsymbol{W}_{\mathrm{opt}}=\boldsymbol{R}^{-1}$。

思路分析： 这是一个利用加权最小二乘进行参数估计的问题。实际中经常会将问题建模成此类模型。求证的思路就是令代价函数关于权矩阵的梯度为零，从而得到权矩阵的最优值。

证明： 首先，将$\boldsymbol{\epsilon}=\boldsymbol{x}-\boldsymbol{A\theta}$代入到$Q(\boldsymbol{\theta})$的表达式中，得到

$$\begin{aligned} Q(\boldsymbol{\theta}) &= (\boldsymbol{x}-\boldsymbol{A\theta})^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{x}-\boldsymbol{A\theta}) \\ &= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{x}-\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{x}-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A\theta}+\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A\theta} \end{aligned}$$

求$Q(\boldsymbol{\theta})$关于$\boldsymbol{\theta}$的偏导数，并令其为零，得到

$$2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{x}=2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A\theta}$$

这里用到了$\boldsymbol{W}$为对称阵的条件。从而得到最优的参数估计

$$\hat{\boldsymbol{\theta}}_{\mathrm{WLS}}=\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{x}$$

从$\hat{\boldsymbol{\theta}}_{\mathrm{WLS}}$表达式出发，可以得到

$$\hat{\boldsymbol{\theta}}_{\mathrm{WLS}}=\boldsymbol{\theta}+\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\epsilon}$$

因此，估计误差的协方差矩阵可以表示为

$$\begin{aligned} \boldsymbol{K} &= \mathrm{E}\left\{\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{\mathrm{WLS}}\right)\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{\mathrm{WLS}}\right)^{\mathrm{T}}\right\} \\ &= \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{R}\boldsymbol{W}\boldsymbol{A}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}\right)^{-1} \end{aligned}$$

为了求得最优的加权矩阵，我们需要使估计误差的均方误差最小，即最小化$\boldsymbol{K}$的迹。令

$$\frac{\partial\,\mathrm{tr}(\boldsymbol{K})}{\partial\boldsymbol{W}}=\boldsymbol{O}$$

并利用教材第5章有关矩阵微分的知识，可以得到

$$\boldsymbol{A}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}\right)^{-2}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{R}\boldsymbol{W}\boldsymbol{A}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{R}\boldsymbol{W}\boldsymbol{A}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}\right)^{-2}\boldsymbol{A}^{\mathrm{T}}$$

不难看出，当$\boldsymbol{W}=\boldsymbol{R}^{-1}$时上式成立。

6.2 [48] 给定$\boldsymbol{A}\in R^{m\times n},\boldsymbol{x}\in R^n,\boldsymbol{b}\in R^m,\boldsymbol{C}\in R^{p\times n},\boldsymbol{d}\in R^p$，并且$\tau$是一个大于零的数。现在希望求解带有二次约束的最小二乘问题：

$$\min\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2,\quad \boldsymbol{x}\in S(\tau) \tag{6.19}$$

其中，$S(\tau)$是一个向量集合，定义为

$$S(\tau)=\{\boldsymbol{x}:\|\boldsymbol{C}\boldsymbol{x}-\boldsymbol{d}\|_2\leqslant\tau\} \tag{6.20}$$

(1) 证明：若$\|(\boldsymbol{I}-\boldsymbol{C}\boldsymbol{C}^{\dagger})\boldsymbol{d}\|_2>\tau$，则式(6.19)和式(6.20) 所表述的二次约束最小二乘问题无解。

(2) 二次约束最小二乘问题存在显式解，当且仅当存在$\boldsymbol{z}\in R^n$使得

$$\|\boldsymbol{C}[\boldsymbol{A}^{\dagger}\boldsymbol{b}+(\boldsymbol{I}-\boldsymbol{A}^{\dagger}\boldsymbol{A})\boldsymbol{z}]-\boldsymbol{d}\|_2\leqslant\tau$$

成立，并且对应的显式解由$\boldsymbol{x}=\boldsymbol{A}^{\dagger}+(\boldsymbol{I}-\boldsymbol{A}^{\dagger}\boldsymbol{A})\boldsymbol{z}$给出。 (提示：无约束最小二乘问题$\min\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2$的通解为$\boldsymbol{x}=\boldsymbol{A}^{\dagger}\boldsymbol{b}+\mathrm{Null}(\boldsymbol{A})=\boldsymbol{A}^{\dagger}\boldsymbol{b}+\mathrm{Range}(\boldsymbol{I}-\boldsymbol{A}^{\dagger}\boldsymbol{A})$。)

思路分析：这是一个带有不等式约束条件的最小二乘问题。若要求该问题有解，则必须满足的一个基本条件就是由约束条件确定的可行集不能为空，第一问我们从这一思路出发即可得到答案。第二问实际上就是直接应用第一问的结论，给出带约束的最小二乘问题的通解。

证明：(1) 观察约束条件。首先，我们来解这样的最小二乘问题

$$\min\|\boldsymbol{C}\boldsymbol{x}-\boldsymbol{d}\|_2$$

易知，结果为$\boldsymbol{x}=\left(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\right)^{\dagger}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{d}=\boldsymbol{C}^{\dagger}\boldsymbol{d}$。相应的，

$$\min\|\boldsymbol{C}\boldsymbol{x}-\boldsymbol{d}\|_2=\left\|\left(\boldsymbol{C}\boldsymbol{C}^{\dagger}-\boldsymbol{I}\right)\boldsymbol{d}\right\|_2$$

显然，当上式大于τ时，$S(\tau)$为空集，该约束最小二乘问题自然无解。

(2) 事实上，对于原无约束最小二乘问题，其通解可以表示为

$$\boldsymbol{x}=\boldsymbol{A}^{\dagger}\boldsymbol{b}+\mathrm{Null}(\boldsymbol{A})$$

其中，$\mathrm{Null}(\boldsymbol{A})$又可以等价表示为

$$\mathrm{Null}(\boldsymbol{A})=\{\boldsymbol{x}|\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}\}=\left\{\boldsymbol{y}|\boldsymbol{y}=\left(\boldsymbol{I}-\boldsymbol{A}^{\dagger}\boldsymbol{A}\right)\boldsymbol{z},\forall\boldsymbol{z}\in R^n\right\}$$

因此，只要存在满足约束条件

$$\|\boldsymbol{C}[\boldsymbol{A}^{\dagger}\boldsymbol{b}+(\boldsymbol{I}-\boldsymbol{A}^{\dagger}\boldsymbol{A})\boldsymbol{z}]-\boldsymbol{d}\|_2\leqslant\tau$$

的z，上述带约束的最小二乘问题就一定有解，其解可以表示成

$$x = A^{\dagger}b + (I - A^{\dagger}A)z$$

6.3 [4] 求解线性方程$Ax = b$的总体最小二乘问题也可以表示为

$$\min_{b+e\in\text{Range}(A+E)} \|D[E, e]T\|_{\text{F}}, \quad E \in R^{m\times n},\ e \in R^m$$

式中，$D = \text{diag}(d_1, d_2, \cdots, d_m)$和$T = \text{diag}(t_1, t_2, \cdots, t_{n+1})$非奇异。

(1) 证明：若$\text{rank}(A) < n$，则上述总体最小二乘问题有一个解，当且仅当$b \in \text{Range}(A)$。

(2) 证明：若$\text{rank}(A) = n$，$A^{\text{T}}D^2b = 0$，$|t_{n+1}|\|Db\|_2 \geqslant \sigma_n(DAT_1)$，$T_1 = \text{diag}(t_1, t_2, \cdots, t_n)$，则总体最小二乘问题无解。其中，$\sigma_n(C)$表示矩阵$C$的第$n$个奇异值。

思路分析： 此题要求证明：只有在满足在一定条件的前提下，总体最小二乘问题才有解。这也就告诉我们，总体最小二乘问题并不是总有解的。

证明： (1) 充分性：若$b \in \text{Range}(A)$，则只须令$E = O, e = 0$,即可使得

$$\min_{b+e\in\text{Range}(A+E)} \|D[E, e]T\|_{\text{F}} = 0$$

TLS问题有解。

必要性：根据教材P409式（7.2.7）所述，上述TLS问题的目标函数可以等价写作

$$\min \|D[A, b]Tz\|_2^2, \qquad \text{subject to} \quad z^{\text{T}}z = 1$$

其中，$z = T^{-1}[x, -1]^{\text{T}}$。

该问题类似于Rayleigh商问题，其解向量为矩阵$(D[A, b]T)^{\text{T}}(D[A, b]T)$ 的最小特征值对应的特征向量。

因为$\text{rank}(A) < n$，且D和T非奇异，所以$\text{rank}(D[A, b]T) = \text{rank}([A, b]) < n + 1$。因此，$(D[A, b]T)^{\text{T}}(D[A, b]T)$的最小特征值为零。根据上面的分析，该TLS问题的解满足

$$(D[A, b]T)^{\text{T}}(D[A, b]T)z = 0$$

也即

$$(D[A, b]T)z = D[A, b][x^{\text{T}}, -1]^{\text{T}} = 0 \Rightarrow [A, b][x^{\text{T}}, -1]^{\text{T}} = 0$$

所以就得到了$b \in \text{Range}(A)$。

(2) 按照上一问的思路，计算

$$(D[A, b]T)^{\text{T}}(D[A, b]T) = \begin{bmatrix} T_1A^{\text{T}}D^2AT_1 & 0 \\ 0 & t_{n+1}^2 b^{\text{T}}D^2b \end{bmatrix}$$

从上式可以看出，矩阵$(D[A, b]T)^{\text{T}}(D[A, b]T)$的特征值就是$T_1A^{\text{T}}D^2AT_1$的特征值再加上$|t_{n+1}|^2\|Db\|_2^2$。

由于题目中已知$|t_{n+1}|\|\boldsymbol{Db}\|_2 \geqslant \sigma_n(\boldsymbol{DAT}_1)$ 所以$(\boldsymbol{D}[\boldsymbol{A},\boldsymbol{b}]\boldsymbol{T})^{\mathrm{T}}(\boldsymbol{D}[\boldsymbol{A},\boldsymbol{b}]\boldsymbol{T})$ 的最小特征值应为$\sigma_n^2(\boldsymbol{DAT}_1)$ 。

TLS问题的解应该满足

$$\left(\begin{bmatrix} \boldsymbol{T}_1\boldsymbol{A}^{\mathrm{T}}\boldsymbol{D}^2\boldsymbol{AT}_1 & 0 \\ 0 & t_{n+1}^2\boldsymbol{b}^{\mathrm{T}}\boldsymbol{D}^2\boldsymbol{b} \end{bmatrix} - \sigma_n^2(\boldsymbol{DAT}_1)\boldsymbol{I}\right)\begin{bmatrix} \boldsymbol{T}_1^{-1}\boldsymbol{x} \\ -t_{n+1}^{-1} \end{bmatrix} = \boldsymbol{0}$$

显然，由于$|t_{n+1}|\|\boldsymbol{Db}\|_2 \neq \sigma_n(\boldsymbol{DAT}_1)$ ，而t_{n+1}^{-1} 不可能为零，所以上式无解。

6.4 考虑上题所述的总体最小二乘问题。证明：若$\boldsymbol{C} = \boldsymbol{D}[\boldsymbol{A},\boldsymbol{b}]\boldsymbol{T} = [\boldsymbol{A}_1,\boldsymbol{d}]$，并且$\sigma_n(\boldsymbol{C}) > \sigma_{n+1}(\boldsymbol{C})$，则总体最小二乘解满足$(\boldsymbol{A}_1^{\mathrm{T}}\boldsymbol{A}_1 - \sigma_{n+1}^2(\boldsymbol{C})\boldsymbol{I})\boldsymbol{T}_1^{-1}\boldsymbol{x} = t_{n+1}^{-1}\boldsymbol{A}_1^{\mathrm{T}}\boldsymbol{d}$。

思路分析：直接应用上题的结论，不难给出证明。

证明：类似于题6.4，根据本题所给的条件$\boldsymbol{D}[\boldsymbol{A},\boldsymbol{b}]\boldsymbol{T} = [\boldsymbol{A}_1,\boldsymbol{d}]$和$\sigma_n(\boldsymbol{C}) > \sigma_{n+1}(\boldsymbol{C})$可知，TLS问题的解一定满足

$$\left(\begin{bmatrix} \boldsymbol{A}_1^{\mathrm{T}}\boldsymbol{A}_1 & \boldsymbol{A}_1^{\mathrm{T}}\boldsymbol{d} \\ \boldsymbol{d}^{\mathrm{T}}\boldsymbol{A}_1 & \boldsymbol{d}^{\mathrm{T}}\boldsymbol{d} \end{bmatrix} - \sigma_{n+1}^2(\boldsymbol{C})\boldsymbol{I}\right)\boldsymbol{T}^{-1}\begin{bmatrix} \boldsymbol{x} \\ -1 \end{bmatrix} = 0$$

上述方程的前n行构成的方程组为

$$(\boldsymbol{A}_1^{\mathrm{T}}\boldsymbol{A}_1 - \sigma_{n+1}^2(\boldsymbol{C})\boldsymbol{I})\boldsymbol{T}_1^{-1}\boldsymbol{x} = t_{n+1}^{-1}\boldsymbol{A}_1^{\mathrm{T}}\boldsymbol{d}$$

6.5 已知数据点(1,3), (3,1), (5,7), (4,6), (7,4)，分别求总体最小二乘和一般最小二乘的拟合直线，并分析它们的距离平方和。

思路分析：按照教材7.3节给出的方法求解即可。通过此例，可以对总体最小二乘和一般最小二乘做一比较。

解：根据教材P423的算法7.3.1，计算数据均值向量

$$\bar{\boldsymbol{x}} = \frac{1}{5}\left\{\begin{bmatrix} 1 \\ 3 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \end{bmatrix} + \begin{bmatrix} 5 \\ 7 \end{bmatrix} + \begin{bmatrix} 4 \\ 6 \end{bmatrix} + \begin{bmatrix} 7 \\ 4 \end{bmatrix}\right\} = \begin{bmatrix} 4 \\ 4.2 \end{bmatrix}$$

构造矩阵

$$\boldsymbol{M} = \begin{bmatrix} -3 & -1.2 \\ -1 & -3.2 \\ 1 & 2.8 \\ 0 & 1.8 \\ 3 & -0.2 \end{bmatrix}$$

对$\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}$进行特征值分解

$$\begin{aligned} \boldsymbol{M}^{\mathrm{T}}\boldsymbol{M} &= \begin{bmatrix} 20 & 9 \\ 9 & 22.8 \end{bmatrix} \\ &= \begin{bmatrix} 0.6505 & -0.7595 \\ 0.7595 & 0.6505 \end{bmatrix}\begin{bmatrix} 30.5082 & 0 \\ 0 & 12.2918 \end{bmatrix}\begin{bmatrix} 0.6505 & -0.7595 \\ 0.7595 & 0.6505 \end{bmatrix}^{\mathrm{T}} \end{aligned}$$

这样便可以确定直线方程为

$$-0.7595(x-4)+0.6505(y-4.2)=0$$

计算距离平方和为

$$\boldsymbol{D}_{\text{TLS}}=\left\|\begin{bmatrix}-3 & -1.2\\ -1 & -3.2\\ 1 & 2.8\\ 0 & 1.8\\ 3 & -0.2\end{bmatrix}\begin{bmatrix}-0.7595\\ 0.6505\end{bmatrix}\right\|_2^2=12.2918$$

一般最小二乘拟合的代价函数可以表示为

$$\begin{aligned}D_{\text{LS}}^{(1)} &= \sum_{i=1}^{5}[m(x_i-\bar{x})+(y_i-\bar{y})]^2\\ &= (-3m-1.2)^2+(-m-3.2)^2+(m+2.8)^2+1.8^2+(3m-0.2)^2\end{aligned}$$

$$\frac{\partial D_{\text{LS}}^{(1)}}{\partial m}=40m+18=0 \Rightarrow m=-0.45$$

则直线方程为

$$-0.45(x-4)+(y-4.2)=0$$

距离平方和为

$$d_{\text{LS}}^{(1)}=\frac{D_{\text{LS}}^{(1)}}{1+m^2}=\frac{18.75}{1+0.45^2}=15.5925$$

$$\begin{aligned}D_{\text{LS}}^{(2)} &= \sum_{i=1}^{5}[(x_i-\bar{x})+m(y_i-\bar{y})]^2\\ &= (-3-1.2m)^2+(-1-3.2m)^2+(1+2.8m)^2+(1.8m)^2+(m-0.2m)^2\end{aligned}$$

$$\frac{\partial D_{\text{LS}}^{(2)}}{\partial m}=45.6m+18=0 \Rightarrow m=-0.3947$$

则直线方程为

$$(x-4)-0.3947(y-4.2)=0$$

距离平方和为

$$d_{\text{LS}}^{(2)}=\frac{D_{\text{LS}}^{(2)}}{1+m^2}=\frac{16.4474}{1+0.3947^2}=14.2305$$

第7章 特征分析

矩阵的特征分析包含了丰富多彩的内容。不仅标准的特征值分解有许多有趣的性质和广泛的应用，而且它还有各类既有趣又极为重要的推广：广义特征值分解、Rayleigh商、广义Rayleigh商、二次特征值和多个矩阵的联合对角化，它们之间有以下关系。

(1) 对称的正定矩阵束$(\boldsymbol{A},\boldsymbol{B})$的广义特征值分解等价为$\boldsymbol{B}^{-1}\boldsymbol{A}$的特征值分解；

(2) 满足对称矩阵$\boldsymbol{A}$的Rayleigh商$\lambda=\dfrac{\boldsymbol{u}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{u}}{\boldsymbol{u}^{\mathrm{H}}\boldsymbol{u}}$的极小值和极大值的解$(\lambda,\boldsymbol{u})$分别是与$\boldsymbol{A}$的最小和最大的特征值对应的特征对；

(3) 满足对称矩阵束$(\boldsymbol{A},\boldsymbol{B})$的广义Rayleigh商$\lambda=\dfrac{\boldsymbol{u}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{u}}{\boldsymbol{u}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{u}}$的极小值和极大值的解$(\lambda,\boldsymbol{u})$分别是与矩阵束$(\boldsymbol{A},\boldsymbol{B})$的最小和最大的广义特征值对应的广义特征对；

(4) 二次特征值问题通过线性化，可以转换为标准的特征值问题求解；

(5) 两个对称矩阵的联合对角化等同于这两个矩阵组成的矩阵束的广义特征值分解。

7.1 主要理论与方法

7.1.1 特征值分解

特征值的基本问题可以陈述为：给定一个$n\times n$维矩阵$\boldsymbol{A}$，确定标量λ的值，使得线性代数方程

$$\boldsymbol{A}\boldsymbol{u}=\lambda\boldsymbol{u},\qquad \boldsymbol{u}\neq\boldsymbol{0}\tag{7.1}$$

具有$n\times 1$非零解$\boldsymbol{u}$。这样的标量λ称为矩阵$\boldsymbol{A}$的特征值(eigenvalue)，向量$\boldsymbol{u}$称为与λ对应的特征向量(eigenvector)。

矩阵$\boldsymbol{A}\in C^{n\times n}$的所有特征值$\lambda\in C$的集合称为矩阵$\boldsymbol{A}$的谱，记作$\lambda(\boldsymbol{A})$。矩阵$\boldsymbol{A}$的谱半径是非负实数，定义为

$$\rho(\boldsymbol{A})=\max|\lambda|:\lambda\in\lambda(\boldsymbol{A})\tag{7.2}$$

称多项式$p(x)=\det(\boldsymbol{A}-x\boldsymbol{I})$为矩阵$\boldsymbol{A}$的特征多项式。显然，特征值$\lambda$是特征多项式$\det(\boldsymbol{A}-x\boldsymbol{I})$的根。值得注意的是，即使矩阵$\boldsymbol{A}$是实的，特征方程的根也可能是复的，而且根的多重数可以是任意的，甚至可以是n重根。若λ是特征多项式$\det(\boldsymbol{A}-x\boldsymbol{I})$的$\mu$重根，则称$\lambda$具有代数多重度$\mu$。若与$\lambda$对应的线性无关的特征向量的个数为$\gamma$，则称特征值$\lambda$具有几何多重度$\gamma$。一个特征值的几何多重度小于等于它的代数多重度。几何多重度等于代数多重度的特征值称为半单特征值。几何多重度小于代数多重度的特征值称为亏损特征值。

若矩阵$\boldsymbol{A}_{n\times n}$是一个一般的复矩阵，并且$\lambda$是其特征值，则满足

$$(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{v}=\boldsymbol{0} \quad 或 \quad \boldsymbol{A}\boldsymbol{v}=\lambda\boldsymbol{v} \tag{7.3}$$

的非零向量$\boldsymbol{v}$称为$\boldsymbol{A}$与特征值λ对应的右特征向量，而满足

$$\boldsymbol{u}^{\mathrm{H}}(\boldsymbol{A}-\lambda\boldsymbol{I})=\boldsymbol{0}^{\mathrm{T}} \quad 或 \quad \boldsymbol{u}^{\mathrm{H}}\boldsymbol{A}=\lambda\boldsymbol{u}^{\mathrm{H}} \tag{7.4}$$

的非零向量$\boldsymbol{u}$称为$\boldsymbol{A}$与特征值λ对应的左特征向量。通常，我们只讨论矩阵的右特征向量。

任意一个矩阵$\boldsymbol{A}$的单个特征值λ 的条件数定义为

$$\mathrm{cond}(\lambda)=\frac{1}{\cos\theta(\boldsymbol{u},\boldsymbol{v})} \tag{7.5}$$

式中，$\boldsymbol{u},\boldsymbol{v}$分别是与特征值$\lambda$对应的左和右特征向量。

特征值问题的求解由以下两步组成：

(1) 求出所有使矩阵$\boldsymbol{A}-\lambda\boldsymbol{I}$奇异的标量$\lambda$ (特征值)；

(2) 给出一个使矩阵$\boldsymbol{A}-\lambda\boldsymbol{I}$奇异的特征值$\lambda$，求出所有满足$(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{x}=\boldsymbol{0}$的非零向量$\boldsymbol{x}$，它就是与$\lambda$ 对应的特征向量。

矩阵的特征值分解有众多的重要性质，这里列出一些常用的性质，更详细的讨论参见教材或习题。

(1) 矩阵$\boldsymbol{A}$奇异，当且仅当至少有一个特征值$\lambda=0$。

(2) 矩阵$\boldsymbol{A}$和$\boldsymbol{A}^{\mathrm{T}}$具有相同的特征值。

(3) 对一个$n\times n$矩阵$\boldsymbol{A}$：

① 若λ是$\boldsymbol{A}$的特征值，则λ 也是$\boldsymbol{A}^{\mathrm{T}}$的特征值。

② 若λ是$\boldsymbol{A}$的特征值，则λ^*是$\boldsymbol{A}^{\mathrm{H}}$的特征值。

③ 若λ是$\boldsymbol{A}$的特征值，则$\lambda+\sigma^2$是$\boldsymbol{A}+\sigma^2\boldsymbol{I}$的特征值。

④ 若λ是矩阵$\boldsymbol{A}$的特征值，则$1/\lambda$是逆矩阵$\boldsymbol{A}^{-1}$ 的特征值。

(4) 与不同特征值$\lambda_1,\lambda_2,\cdots,\lambda_n$对应的非零特征向量$\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_n$ 线性无关。

(5) 矩阵$\boldsymbol{A}$的迹等于其所有特征值之和，行列式$\det(\boldsymbol{A})$等于$\boldsymbol{A}$所有特征值的乘积。

(6) 特征值与秩的关系：

① 若$n\times n$矩阵$\boldsymbol{A}$有r个非零特征值，则$\mathrm{rank}(\boldsymbol{A})\geqslant r$。

② 若0是$n\times n$矩阵$\boldsymbol{A}$的无多重的特征值，则$\mathrm{rank}(\boldsymbol{A})=n-1$。

③ 若$\mathrm{rank}(\boldsymbol{A}-\lambda\boldsymbol{I})\leqslant n-1$，则$\lambda$是矩阵$\boldsymbol{A}$的特征值。

(7) 若矩阵$\boldsymbol{A}$的特征值为λ，则矩阵多项式$f(\boldsymbol{A})=\boldsymbol{A}^n+c_1\boldsymbol{A}^{n-1}+\cdots+c_{n-1}\boldsymbol{A}+c_n\boldsymbol{I}$的特征值为

$$f(\lambda)=\lambda^n+c_1\lambda^{n-1}+\cdots+c_{n-1}\lambda+c_n \tag{7.6}$$

(8) 若$(\lambda,\boldsymbol{u})$是矩阵$\boldsymbol{A}$的特征对，则$(\mathrm{e}^{\lambda},\boldsymbol{u})$是矩阵指数函数$\mathrm{e}^{\boldsymbol{A}}$的特征对。

(9) 若$\lambda(A)$和$\lambda(B)$分别是矩阵$\boldsymbol{A}$和$\boldsymbol{B}$的特征值，而$\boldsymbol{u}(A)$和$\boldsymbol{u}(B)$分别是与特征值$\lambda(A)$和$\lambda(B)$对应的特征向量，则

① $\lambda(A)\lambda(B)$和$\boldsymbol{u}(A)\otimes\boldsymbol{u}(B)$是矩阵Kronecker积$\boldsymbol{A}\otimes\boldsymbol{B}$的一对特征值和特征向量。

② $\lambda(A)$和$\lambda(B)$分别是矩阵直和$\boldsymbol{A}\oplus\boldsymbol{B}$的特征值，与它们对应的特征向量分别为

$$\begin{bmatrix}\boldsymbol{u}(A)\\ \boldsymbol{0}\end{bmatrix} \quad 和 \quad \begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{u}(B)\end{bmatrix}$$

(10) 一个$n\times n$实矩阵$\boldsymbol{A}$是可对角化的，当且仅当$\boldsymbol{A}$ 具有n个线性无关的特征向量。

(11) 若$\boldsymbol{A}$的特征值不相同，则一定可以找到一个相似矩阵$\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}=\boldsymbol{D}$ (对角矩阵)，其对角元素即是矩阵$\boldsymbol{A}$的特征值。

7.1.2 矩阵函数

若

$$p(\boldsymbol{A})=p_n\boldsymbol{A}^n+p_{n-1}\boldsymbol{A}^{n-1}+\cdots+p_1\boldsymbol{A}+p_0\boldsymbol{I}=\boldsymbol{O}$$

则称

$$p(x)=p_nx^n+p_{n-1}x^{n-1}+\cdots+p_1x+p_0$$

是使矩阵$\boldsymbol{A}$零化的多项式，简称零化多项式(annihilating polynomial)。次数最小的零化多项式称为矩阵$\boldsymbol{A}$的最小多项式。

Cayley-Hamilton定理表明特征多项式$p(x)=\det(\boldsymbol{A}-x\boldsymbol{I})$是使矩阵$\boldsymbol{A}_{n\times n}$零化的多项式，即$p(\boldsymbol{A})=\boldsymbol{O}$。Cayley-Hamilton定理为我们计算矩阵幂、矩阵指数函数以及矩阵求逆等运算提供了非常有效的工具。

1. 利用Cayley-Hamilton定理计算逆矩阵和广义逆矩阵

若矩阵$\boldsymbol{A}_{n\times n}$非奇异，多项式$p(x)=\det(\boldsymbol{A}-x\boldsymbol{I})=p_nx^n+p_{n-1}x^{n-1}+\cdots+p_0$ 是矩阵$\boldsymbol{A}$的特征多项式，则

$$\boldsymbol{A}^{-1}=-\frac{1}{p_0}(p_n\boldsymbol{A}^{n-1}+p_{n-1}\boldsymbol{A}^{n-2}+\cdots+p_2\boldsymbol{A}+p_1\boldsymbol{I}) \tag{7.7}$$

Cayley-Hamilton定理还可用来计算矩阵的广义逆，我们有如下定理。

定理 7.1 令矩阵$\boldsymbol{A}$ 是任意一个$m\times n$矩阵，并且令

$$f(\lambda)=(-1)^m(a_0\lambda^m+a_1\lambda^{m-1}+\cdots+a_{m-1}\lambda+a_m),\quad a_0=1 \tag{7.8}$$

是矩阵乘积$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$的特征多项式$\det(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}-\lambda\boldsymbol{I})$。若$k$是满足$a_k\neq 0$的最大整数，则$\boldsymbol{A}$的广义逆矩阵由

$$\boldsymbol{A}^{\dagger}=-a_k^{-1}\boldsymbol{A}^{\mathrm{H}}\left[(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{k-1}+a_1(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{k-2}+\cdots+a_{k-2}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})+a_{k-1}\boldsymbol{I}\right] \tag{7.9}$$

确定。当$k=0$是使$a_k\neq 0$的最大整数时，广义逆矩阵$\boldsymbol{A}^{\dagger}=\boldsymbol{O}$。

2. 利用Cayley-Hamilton定理计算矩阵的幂

若x^K除$p(x)$的余式为$r(x)$，则$\boldsymbol{A}^K = r(\boldsymbol{A})$。设特征多项式$p(x) = \det(\boldsymbol{A}-x\boldsymbol{I}) = 0$的$n$个特征根为$\lambda_1, \lambda_2, \cdots, \lambda_n$，余式$r(x) = r_0 + r_1 x + \cdots + r_{n-1}x^{n-1}$的系数可从以下线性方程组解得。

$$\left.\begin{aligned} r_0 + \lambda_1 r_1 + \cdots + \lambda_1^{n-1} r_{n-1} &= \lambda_1^K \\ r_0 + \lambda_2 r_1 + \cdots + \lambda_2^{n-1} r_{n-1} &= \lambda_2^K \\ &\cdots \\ r_0 + \lambda_n r_1 + \cdots + \lambda_n^{n-1} r_{n-1} &= \lambda_n^K \end{aligned}\right\} \tag{7.10}$$

3. 利用Cayley-Hamilton定理计算矩阵指数函数

类似于标量指数函数

$$\mathrm{e}^{at} = 1 + at + \frac{1}{2!}a^2t^2 + \cdots + \frac{1}{k!}a^kt^k + \cdots$$

对一个已知矩阵$\boldsymbol{A}$，可以定义矩阵指数函数(matrix exponential function)

$$\mathrm{e}^{\boldsymbol{A}t} = \boldsymbol{I} + \boldsymbol{A}t + \frac{1}{2!}\boldsymbol{A}^2t^2 + \cdots + \frac{1}{k!}\boldsymbol{A}^kt^k + \cdots \tag{7.11}$$

令$\boldsymbol{A}$是一个$n \times n$常数矩阵，其特征多项式为

$$p(\lambda) = \det(\lambda\boldsymbol{I} - \boldsymbol{A}) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0$$

则矩阵指数函数$\mathrm{e}^{\boldsymbol{A}t}$由下式

$$\mathrm{e}^{\boldsymbol{A}t} = x_1(t)\boldsymbol{I} + x_2(t)\boldsymbol{A} + x_3(t)\boldsymbol{A}^2 + \cdots + x_n(t)\boldsymbol{A}^{n-1} \tag{7.12}$$

给出，式中，$x_k(t)$是n阶标量微分方程

$$x^{(n)}(t) + c_{n-1}x^{(n-1)}(t) + \cdots + c_1x'(t) + c_0x(t) = 0 \tag{7.13}$$

满足初始条件

$$\left.\begin{aligned} x_1(0) &= 1 \\ x_1'(0) &= 0 \\ &\cdots \\ x_1^{(n-1)}(0) &= 0 \end{aligned}\right\}, \quad \left.\begin{aligned} x_2(0) &= 0 \\ x_2'(0) &= 1 \\ &\cdots \\ x_2^{(n-1)}(0) &= 0 \end{aligned}\right\}, \quad \cdots, \quad \left.\begin{aligned} x_n(0) &= 0 \\ x_n'(0) &= 0 \\ &\cdots \\ x_n^{(n-1)}(0) &= 1 \end{aligned}\right\}$$

的解。

一般的，设一元（复变）函数$f(z)$在$z=0$的幂级数为$f(z) = \sum_{k=0}^{\infty} c_k z^k$（$|z| < r$），$\boldsymbol{A}$的谱半径$\rho(\boldsymbol{A}) < r$，则称$f(\boldsymbol{A}) = \sum_{k=0}^{\infty} c_k \boldsymbol{A}^k$为对应于$f(z)$的矩阵函数。这时，除了可以利用Cayley-Hamilton定理计算矩阵函数外，我们也可以利用矩阵的Jordan标准型来计算矩阵函数。我们有如下结论[4]。

设$\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X} = \mathrm{diag}(\boldsymbol{J}_1, \boldsymbol{J}_2, \cdots, \boldsymbol{J}_p)$为$\boldsymbol{A} \in C^{n\times n}$的Jordan标准型，其中

$$\boldsymbol{J}_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & & 1 \\ & & \ddots & \\ & & & \lambda_i \end{bmatrix}$$

为m_i阶Jordan块。如果$f(z)$在包含$\lambda(\boldsymbol{A})$的一个开集上解析，则

$$f(\boldsymbol{A}) = \boldsymbol{X}\mathrm{diag}(f(\boldsymbol{J_1}), f(\boldsymbol{J_2}), \cdots, f(\boldsymbol{J_p}))\boldsymbol{X}^{-1}$$

其中

$$f(\boldsymbol{J_i}) = \begin{bmatrix} f(\lambda_i) & f'(\lambda_i) & \cdots & \dfrac{f^{(m_i-1)}(\lambda_i)}{(m_i-1)!} \\ & f(\lambda_i) & \ddots & \\ & & \ddots & f'(\lambda_i) \\ & & & f(\lambda_i) \end{bmatrix}$$

特别地，当$\boldsymbol{A}$可对角化时，即$\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$，则有

$$f(\boldsymbol{A}) = \boldsymbol{X}\mathrm{diag}(f(\lambda_1), f(\lambda_2), \cdots, f(\lambda_n))\boldsymbol{X}^{-1}$$

一个$n\times n$实矩阵$\boldsymbol{A}$是可对角化的，当且仅当$\boldsymbol{A}$ 具有n个线性无关的特征向量。作为特例，若$n\times n$矩阵$\boldsymbol{A}$有n个不同的特征值，则$\boldsymbol{A}$是可对角化的。更一般地，即使矩阵$\boldsymbol{A}$具有多重根，它仍然有可能是可对角化的，因为$\boldsymbol{A}$的n个特征向量有可能是线性无关的。正规矩阵（包含Hermitian矩阵、斜Hermitian矩阵以及酉矩阵等）是一类可对角化的矩阵。

7.1.3 广义特征值分解

特征值分解的一种推广是广义特征值分解。令$\boldsymbol{A}$和$\boldsymbol{B}$是两个$n\times n$ 正方矩阵，它们组成一矩阵束(matrix pencil)或矩阵对(matrix pair)，记作$(\boldsymbol{A}, \boldsymbol{B})$。现在考虑广义特征值问题：求所有的标量$\lambda$使得方程

$$\boldsymbol{A}\boldsymbol{u} = \lambda\boldsymbol{B}\boldsymbol{u} \tag{7.14}$$

具有非零解$\boldsymbol{u} \neq \boldsymbol{0}$。我们称这样的标量$\lambda$和非零向量$\boldsymbol{u}$分别为矩阵束$(\boldsymbol{A}, \boldsymbol{B})$的广义特征值(generalized eigenvalue)和广义特征向量(generalized eigenvector)。

若将矩阵束的广义特征值记作$\lambda(\boldsymbol{A}, \boldsymbol{B})$，则广义特征值定义为

$$\lambda(\boldsymbol{A}, \boldsymbol{B}) = \{z \in C : \det(\boldsymbol{A} - z\boldsymbol{B}) = 0\} \tag{7.15}$$

正则矩阵束的特征值问题是广义特征值问题的一种特例。若$n\times n$矩阵$\boldsymbol{A}$和$\boldsymbol{B}$ 均为Hermitian矩阵，并且$\boldsymbol{B}$正定，则称$(\boldsymbol{A}, \boldsymbol{B})$为正则矩阵束。关于正则矩阵束的广义特征值分解，有如下定理。

定理 7.2 [44] 若$\lambda_1, \lambda_2, \cdots, \lambda_n$是正则矩阵束$(\boldsymbol{A}, \boldsymbol{B})_{n\times n}$的广义特征值，则

(1) 存在矩阵$\boldsymbol{X} \in C^{n\times n}$，使得

$$\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^{\mathrm{H}} = \boldsymbol{I}_n, \quad \boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{H}} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

或等价写作

$$\boldsymbol{X}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{X} = \boldsymbol{I}_n, \quad \boldsymbol{A}\boldsymbol{X} = \boldsymbol{B}\boldsymbol{X}\boldsymbol{\Lambda}$$

式中，$\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$。

(2) 所有广义特征值都是实的，即$\lambda_i \in R, i = 1, 2, \cdots, n$。

(3) 若记$\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n]$，则有

$$\boldsymbol{A}\boldsymbol{x}_i = \lambda_i \boldsymbol{B}\boldsymbol{x}_i, \quad i = 1, 2, \cdots, n$$

$$\boldsymbol{x}_i^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}_j = \delta_{ij}, \quad i, j = 1, 2, \cdots, n$$

式中，δ_{ij}为Kronecker δ函数。

下面是关于广义特征值问题$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{B}\boldsymbol{x}$的一些性质[46, pp.176~177]。

(1) 若矩阵$\boldsymbol{A}$和$\boldsymbol{B}$ 互换，则广义特征值将变为其倒数，但广义特征向量保持不变，即有

$$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{B}\boldsymbol{x} \ \Rightarrow \ \boldsymbol{B}\boldsymbol{x} = \frac{1}{\lambda}\boldsymbol{A}\boldsymbol{x}$$

(2) 若$\boldsymbol{B}$非奇异，则广义特征值分解简化为标准的特征值分解

$$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{B}\boldsymbol{x} \ \Rightarrow \ (\boldsymbol{B}^{-1}\boldsymbol{A})\boldsymbol{x} = \lambda\boldsymbol{x}$$

(3) 若$\boldsymbol{A}$和$\boldsymbol{B}$均为正定的Hermitian矩阵，则广义特征值必定是实的，并且与不同广义特征值对应的广义特征向量相对于正定矩阵$\boldsymbol{B}$正交，即

$$\boldsymbol{x}_i^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}_j = \boldsymbol{x}_i^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}_j = 0$$

(4) 若$\boldsymbol{A}$和$\boldsymbol{B}$均为实对称矩阵，并且$\boldsymbol{B}$正定，则广义特征值问题$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{B}\boldsymbol{x}$ 可以变换为标准的对称特征值问题

$$(\boldsymbol{L}^{-1}\boldsymbol{A}\boldsymbol{L}^{-\mathrm{T}})(\boldsymbol{L}^{\mathrm{T}}\boldsymbol{x}) = \lambda(\boldsymbol{L}^{\mathrm{T}}\boldsymbol{x})$$

式中，$\boldsymbol{L}$为下三角矩阵，它是Cholesky分解$\boldsymbol{B} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}$的因子。

(5) 若$\boldsymbol{A}$和$\boldsymbol{B}$均为实对称的正定矩阵，则广义特征值一定是正的。

(6) 如果$\boldsymbol{A}$奇异，则$\lambda = 0$必定是一个广义特征值。

(7) 若$\tilde{\boldsymbol{B}} = \boldsymbol{B} + (1/\alpha)\boldsymbol{A}$，其中，$\alpha$是任意一个不等于零的标量，则修正的广义特征值问题

$$\boldsymbol{A}\boldsymbol{x} = \tilde{\lambda}\tilde{\boldsymbol{B}}\boldsymbol{x}$$

的广义特征值$\tilde{\lambda}$与原广义特征值λ之间存在下列关系：

$$\frac{1}{\tilde{\lambda}} = \frac{1}{\lambda} + \frac{1}{\alpha}$$

7.1.4 Hermitian矩阵与Rayleigh商

Hermitian矩阵有很多很好的性质，且在实际应用中常常遇到。

Hermitian矩阵的一个重要性质是：任何Hermitian矩阵$\boldsymbol{A}$皆可对角化，且有相似范式

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{H}} \tag{7.16}$$

其中$\boldsymbol{U}$为酉矩阵，$\boldsymbol{\Sigma}$为对角线元素为实数的对角阵。即Hermitian矩阵的特征值是实的，且特征向量相互正交。

式(7.16)经常写成级数展开形式

$$\boldsymbol{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^{\mathrm{H}} \tag{7.17}$$

这一形式称为Hermitian矩阵的谱分解。

若Hermitian矩阵$\boldsymbol{A}$的特征值都大于零，则称该Hermitian矩阵正定；若Hermitian矩阵的特征值都大于等于零，则称该Hermitian矩阵半正定；若Hermitian矩阵的特征值都小于零，则称该Hermitian矩阵负定；若Hermitian矩阵的特征值都小于等于零，则称该Hermitian矩阵半负定；以上情况皆不是者，则称该Hermitian矩阵不定。

表7.1.1给出了实对称矩阵的分类及各类矩阵的充分必要条件。

表7.1.1 实对称矩阵的分类及各类矩阵的充分必要条件

名称	定义	充分必要条件
正定矩阵	特征值都大于零的实对称矩阵	所有主子式都大于零，即 $\det(\boldsymbol{A}_i) > 0,\ i = 1,2,\cdots,n$
半正定矩阵	特征值都不小于零的实对称矩阵	$\det(\boldsymbol{A}) = 0$ $\det(\boldsymbol{A}_i) \geqslant 0,\ i = 1,2,\cdots,n-1$
负定矩阵	特征值都小于零的实对称矩阵	$\det(\boldsymbol{A}_i)\begin{cases} < 0 & (i\text{为奇数}) \\ > 0 & (i\text{为偶数}) \end{cases}$ $i = 1,2,\cdots,n$
半负定矩阵	特征值都不大于零的实对称矩阵	$\det(\boldsymbol{A}) = 0$ $\det(\boldsymbol{A}_i)\begin{cases} \leqslant 0 & (i\text{为奇数}) \\ \geqslant 0 & (i\text{为偶数}) \end{cases}$ $i = 1,2,\cdots,n-1$
不定矩阵	特征值既有大于零，也有小于零的实对称矩阵	或有主子式 $\det(\boldsymbol{A}_{2k}) = 0$ 或有 $\det(\boldsymbol{A}_{2i+1}) > 0,\ \det(\boldsymbol{A}_{2j+1}) < 0,\ i \neq j$

正定矩阵具有以下性质：

(1) 正定矩阵$\boldsymbol{A}$的所有对角线元素都是正的。

(2) 正定矩阵的行列式大于零，即$\det(\boldsymbol{A}) > 0$，因而$\boldsymbol{A}$是非奇异的。

(3) 逆矩阵$\boldsymbol{A}^{-1}$是正定的。

(4) 正定矩阵$\boldsymbol{A}$的对角线元素满足不等式$a_{ii}a_{jj} > |a_{ij}|^2$，$i \neq j$。

(5) 正定矩阵$\boldsymbol{A}$具有最大绝对值的元素一定在对角线上。

(6) 正定矩阵删去第k行和第k列$(1 \leqslant k \leqslant n)$之后，仍然是正定的。

Hermitian矩阵是一类重要的正规矩阵。下面是Hermitian矩阵、斜Hermitian矩阵和酉矩阵等几种特殊的正规矩阵和一般的正规矩阵的特征值情况[48, p.183]。

$$
\text{正规矩阵}\begin{cases}\text{Hermitian矩阵} \Longleftrightarrow \text{特征值均为实值}\\ \text{斜Hermitian矩阵} \Longleftrightarrow \text{特征值均在虚轴上}\\ \text{酉矩阵} \Longleftrightarrow \text{特征值的复数模均等于1}\\ \text{其他正规矩阵} \Longleftrightarrow \text{特征值无上述要求}\end{cases}
$$

一个重要的Hermitian矩阵的二次型函数是Rayleigh商。Hermitian矩阵$\boldsymbol{A} \in C^{n\times n}$的Rayleigh商或Rayleigh-Ritz比$R(\boldsymbol{x})$是一个标量，定义为

$$
R(\boldsymbol{x}) = R(\boldsymbol{x}, \boldsymbol{A}) = \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} \tag{7.18}
$$

其中，$\boldsymbol{x}$是待选择的向量，其目的是使Rayleigh商最大化或者最小化。

Rayleigh商的重要性质有：

性质1 (齐次性) 若α和β为标量，则

$$
R(\alpha\boldsymbol{x}, \beta\boldsymbol{A}) = \beta R(\boldsymbol{x}, \boldsymbol{A}) \tag{7.19}
$$

性质2 (平移不变性)

$$
R(\boldsymbol{x}, \boldsymbol{A} - \alpha\boldsymbol{I}) = R(\boldsymbol{x}, \boldsymbol{A}) - \alpha \tag{7.20}
$$

性质3 (正交性)

$$
\boldsymbol{x} \perp (\boldsymbol{A} - R(\boldsymbol{x})\boldsymbol{I})\boldsymbol{x} \tag{7.21}
$$

性质4 (有界性) 当向量$\boldsymbol{x}$在所有非零向量的范围变化时，Rayleigh商$R(\boldsymbol{x})$落在一复平面的区域(称为矩阵$\boldsymbol{A}$的值域)内。这一区域是闭合的、有界的和凸的。若$\boldsymbol{A}$是Hermitian的，即满足$\boldsymbol{A} = \boldsymbol{A}^{\mathrm{H}}$，则这一区域是一个闭区间$[\lambda_1, \lambda_n]$。

性质5 (最小残差) 对于所有向量$\boldsymbol{x} \neq \boldsymbol{0}$和所有标量$\mu$，恒有

$$
\|[\boldsymbol{A} - R(\boldsymbol{x})\boldsymbol{I}]\boldsymbol{x}\| \leqslant \|[\boldsymbol{A} - \mu\boldsymbol{I}]\boldsymbol{x}\| \tag{7.22}
$$

Hermitian矩阵的Rayleigh商的有界性可以用下面的定理严格叙述。

定理 7.3 (Rayleigh-Ritz定理) 令$\boldsymbol{A} \in C^{n\times n}$是Hermitian的，并令$\boldsymbol{A}$的特征值按递增次序

$$
\lambda_{\min} = \lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_{n-1} \leqslant \lambda_n = \lambda_{\max} \tag{7.23}
$$

排列，则

$$\max_{\boldsymbol{x}\neq \boldsymbol{0}} \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} = \max_{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}=1} \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} = \lambda_{\max}, \quad 若\ \boldsymbol{A}\boldsymbol{x} = \lambda_{\max}\boldsymbol{x} \tag{7.24}$$

$$\min_{\boldsymbol{x}\neq \boldsymbol{0}} \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} = \min_{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}=1} \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} = \lambda_{\min}, \quad 若\ \boldsymbol{A}\boldsymbol{x} = \lambda_{\min}\boldsymbol{x} \tag{7.25}$$

更一般地，矩阵$\boldsymbol{A}$的所有特征向量和特征值分别称为Rayleigh商$R(\boldsymbol{x})$的临界点(critical point)和临界值(critical value)。

极小-极大原理或极大-极小原理也是一个关于Rayleigh商的重要定理。

定理 7.4 (Courant-Fischer定理) 令$n\times n$矩阵$\boldsymbol{A}$为Hermitian矩阵，其特征值为$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$，则$\lambda_k$ $(1\leqslant k\leqslant n)$服从关系式

$$\lambda_k = \min_{S,\dim(S)=n-k+1}\ \max_{\boldsymbol{x}\in S,\boldsymbol{x}\neq\boldsymbol{0}} \left\{\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\right\} \tag{7.26}$$

或

$$\lambda_k = \max_{S,\dim S=k}\ \min_{\boldsymbol{x}\in S,\boldsymbol{x}\neq\boldsymbol{0}} \left\{\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\right\} \tag{7.27}$$

Rayleigh商的推广形式称为广义Rayleigh商。令$\boldsymbol{A}$和$\boldsymbol{B}$均为$n\times n$维Hermitian矩阵，且$\boldsymbol{B}$是正定矩阵。矩阵束$(\boldsymbol{A},\boldsymbol{B})$的广义Rayleigh商或广义Rayleigh-Ritz比$R(\boldsymbol{x})$是一个标量(函数)，定义为

$$R(\boldsymbol{x}) = \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}} \tag{7.28}$$

其中，$\boldsymbol{x}$是待选择的向量，其目的是使广义Rayleigh 商最大化或者最小化。

广义Rayleigh商的分析可以转化为Rayleigh商的分析。定义一个新向量$\tilde{\boldsymbol{x}} = \boldsymbol{B}^{1/2}\boldsymbol{x}$，其中，$\boldsymbol{B}^{1/2}$表示正定矩阵$\boldsymbol{B}$的平方根。用$\boldsymbol{x} = \boldsymbol{B}^{-1/2}\tilde{\boldsymbol{x}}$代入广义Rayleigh商定义式(7.28)，则有

$$R(\tilde{\boldsymbol{x}}) = \frac{\tilde{\boldsymbol{x}}^{\mathrm{H}}\left(\boldsymbol{B}^{-1/2}\right)^{\mathrm{H}}\boldsymbol{A}\left(\boldsymbol{B}^{-1/2}\right)\tilde{\boldsymbol{x}}}{\tilde{\boldsymbol{x}}^{\mathrm{H}}\tilde{\boldsymbol{x}}} \tag{7.29}$$

这表明，矩阵束$(\boldsymbol{A},\boldsymbol{B})$的广义Rayleigh商等价为矩阵乘积$(\boldsymbol{B}^{-1/2})^{\mathrm{H}}\boldsymbol{A}(\boldsymbol{B}^{-1/2})$的Rayleigh商。

关于广义Rayleigh商，我们有结论：广义Rayleigh商取最大值和最小值的条件是

$$R(\boldsymbol{x}) = \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}} = \lambda_{\max}, \quad 若选择\ \boldsymbol{A}\boldsymbol{x} = \lambda_{\max}\boldsymbol{B}\boldsymbol{x} \tag{7.30}$$

$$R(\boldsymbol{x}) = \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}} = \lambda_{\min}, \quad 若选择\ \boldsymbol{A}\boldsymbol{x} = \lambda_{\min}\boldsymbol{B}\boldsymbol{x} \tag{7.31}$$

即是说，欲使广义Rayleigh商最大化，向量$\boldsymbol{x}$ 必须选取与矩阵束$(\boldsymbol{A},\boldsymbol{B})$最大广义特征值对应的特征向量；反之，需要使广义Rayleigh商最小化时，则应该取与矩阵束$(\boldsymbol{A},\boldsymbol{B})$ 最小广义特征值对应的特征向量作$\boldsymbol{x}$。

7.1.5 二次特征值问题

二次特征值问题与标准的特征值问题尤其是广义特征值问题，既存在密切的联系，又有着明显的不同，是非线性特征值问题的一个重要子类。二次特征值问题可叙述为[29]：求标量λ和非零向量$\boldsymbol{u},\boldsymbol{v}$ 满足方程

$$(\lambda^2\boldsymbol{M}+\lambda\boldsymbol{C}+\boldsymbol{K})\boldsymbol{u}=\boldsymbol{0},\qquad \boldsymbol{v}^{\mathrm{H}}(\lambda^2\boldsymbol{M}+\lambda\boldsymbol{C}+\boldsymbol{K})=\boldsymbol{0}^{\mathrm{T}} \tag{7.32}$$

式中，$\boldsymbol{M},\boldsymbol{C},\boldsymbol{K}$为$n\times n$复矩阵。满足上述方程的标量$\lambda$称为特征值，非零向量$\boldsymbol{u}$和$\boldsymbol{v}$分别称为与特征值$\lambda$对应的右和左特征向量。特征值和特征向量组成二次特征值问题的特征对。

令

$$\boldsymbol{Q}(\lambda)=\lambda^2\boldsymbol{M}+\lambda\boldsymbol{C}+\boldsymbol{K} \tag{7.33}$$

这是一个二次$n\times n$矩阵多项式。换言之，矩阵$\boldsymbol{Q}(\lambda)$的系数是λ的二次多项式。通常称$\boldsymbol{Q}(\lambda)$为λ矩阵[30]。于是，特征值λ是特征方程

$$\det\boldsymbol{Q}(z)=\det(z^2\boldsymbol{M}+z\boldsymbol{C}+\boldsymbol{K})=0 \tag{7.34}$$

的根，并称$\boldsymbol{Q}(\lambda)$为特征多项式。

根据矩阵$\boldsymbol{M},\boldsymbol{C},\boldsymbol{K}$的性质不同，二次特征值问题的特征值与特征向量的性质也有所不同，详见表7.1.2的汇总。

表7.1.2 二次特征值问题的特征值与特征向量的性质[45]

编号	矩阵性质	特征值性质	特征向量性质
1	$\boldsymbol{M}$非奇异	$2n$个有限大特征值	
2	$\boldsymbol{M}$奇异	有限大和无穷大特征值	
3	$\boldsymbol{M},\boldsymbol{C},\boldsymbol{K}$为实矩阵	实或共轭成对(λ,λ^*)	若$\boldsymbol{u}$是λ的右特征向量，则$\boldsymbol{u}^*$是λ^*的右特征向量
4	$\boldsymbol{M},\boldsymbol{C},\boldsymbol{K}$为Hermitian矩阵	实或共轭成对(λ,λ^*)	若$\boldsymbol{u}$是λ的右特征向量，则$\boldsymbol{u}^*$是λ^*的右特征向量
5	$\boldsymbol{M}$：Hermitian正定 $\boldsymbol{C},\boldsymbol{K}$：Hermitian半正定	$\mathrm{Re}(\lambda)\leqslant 0$	
6	$\boldsymbol{M},\boldsymbol{C}$对称正定 $\boldsymbol{K}$对称半正定 $\gamma(\boldsymbol{M},\boldsymbol{C},\boldsymbol{K})>0$	λ取正和负，n个大特征值与n个小特征值	与n个大特征值(或n个小特征值)对应的n个特征向量线性无关
7	$\boldsymbol{M},\boldsymbol{K}$：Hermitian矩阵 $\boldsymbol{M}$正定，$\boldsymbol{C}=-\boldsymbol{C}^{\mathrm{H}}$	特征值为纯虚数或共轭成对(λ,λ^*)	若$\boldsymbol{u}$是λ的右特征向量，则$\boldsymbol{u}$是$-\lambda^*$的左特征向量
8	$\boldsymbol{M},\boldsymbol{K}$实对称正定 $\boldsymbol{C}=-\boldsymbol{C}^{\mathrm{T}}$	特征值为纯虚数	

注：$\gamma(\boldsymbol{M},\boldsymbol{C},\boldsymbol{K})=\min\left\{(\boldsymbol{u}^{\mathrm{H}}\boldsymbol{C}\boldsymbol{u})^2-4(\boldsymbol{u}^{\mathrm{H}}\boldsymbol{M}\boldsymbol{u})(\boldsymbol{u}^{\mathrm{H}}\boldsymbol{K}\boldsymbol{u}):\|\boldsymbol{u}\|_2=1\right\}$

根据二次型函数的大小，二次特征值问题又可进一步分类如下[43]：

(1) 二次型函数满足$(\boldsymbol{u}^{\mathrm{H}}\boldsymbol{C}\boldsymbol{u})^2 < 4(\boldsymbol{u}^{\mathrm{H}}\boldsymbol{M}\boldsymbol{u})(\boldsymbol{u}^{\mathrm{H}}\boldsymbol{K}\boldsymbol{u})$的二次特征值问题称为椭圆二次特征值问题(elliptic QEP)。

(2) 二次型函数满足$(\boldsymbol{u}^{\mathrm{H}}\boldsymbol{C}\boldsymbol{u})^2 > 4(\boldsymbol{u}^{\mathrm{H}}\boldsymbol{M}\boldsymbol{u})(\boldsymbol{u}^{\mathrm{H}}\boldsymbol{K}\boldsymbol{u})$的二次特征值问题称为双曲线二次特征值问题(hyperbolic QEP)。

以下定理给出了椭圆二次特征值问题和双曲线二次特征值问题的判别方法。

定理 7.5 具有正定矩阵$\boldsymbol{M}$ 的二次特征值问题式(7.32)是椭圆二次特征值问题，当且仅当λ矩阵$\boldsymbol{Q}(\lambda)$对所有实特征值$\lambda \in R$正定。式(7.32)是双曲线二次特征值问题，当且仅当$\boldsymbol{Q}(\lambda)$对某些实特征值$\lambda \in R$ 负定。

二次特征值问题可通过分解法或线性化方法求解。

1. 分解法

定义λ矩阵$\boldsymbol{Q}(\lambda) = \lambda^2\boldsymbol{M} + \lambda\boldsymbol{C} + \boldsymbol{K}$，如果二次矩阵方程

$$\boldsymbol{Q}(S) = \boldsymbol{M}\boldsymbol{S}^2 + \boldsymbol{C}\boldsymbol{S} + \boldsymbol{K} = \boldsymbol{O} \tag{7.35}$$

存在一个解$\boldsymbol{S} \in C^{n\times n}$，则有

$$\boldsymbol{Q}(\lambda) = \boldsymbol{Q}(\lambda) - \boldsymbol{Q}(S) = (\lambda\boldsymbol{M} + \boldsymbol{M}\boldsymbol{S} + \boldsymbol{C})(\lambda\boldsymbol{I} - \boldsymbol{S}) \tag{7.36}$$

这样，求解特征值问题$\det[\boldsymbol{Q}(\lambda)] = 0$就转化为以下两个(一次)特征值子问题：

(1) 二次特征值是特征方程$\det(\lambda\boldsymbol{M} + \boldsymbol{M}\boldsymbol{S} + \boldsymbol{C}) = 0$的解，共得到$n$个解；

(2) 二次特征值是特征方程$\det(\lambda\boldsymbol{I} - \boldsymbol{S}) = 0$的$n$个解。

有趣的是，特征值问题$\det(\lambda\boldsymbol{M} + \boldsymbol{M}\boldsymbol{S} + \boldsymbol{C}) = 0$与广义特征值问题$(\boldsymbol{M}\boldsymbol{S} + \boldsymbol{C})\boldsymbol{u} = \lambda(-\boldsymbol{M})\boldsymbol{u}$ 等价。

2. 线性化方法

容易验证，若令$\dot{\boldsymbol{u}} = \lambda\boldsymbol{u}$，则二次特征值问题$(\lambda^2\boldsymbol{M} + \lambda\boldsymbol{C} + \boldsymbol{K})\boldsymbol{u} = \boldsymbol{0}$可以等价写作常见的状态空间方程形式：

$$L1: \quad \begin{bmatrix} \boldsymbol{C} & \boldsymbol{K} \\ -\boldsymbol{N} & \boldsymbol{O} \end{bmatrix}\begin{bmatrix} \dot{\boldsymbol{u}} \\ \boldsymbol{u} \end{bmatrix} + \lambda\begin{bmatrix} \boldsymbol{M} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{N} \end{bmatrix}\begin{bmatrix} \dot{\boldsymbol{u}} \\ \boldsymbol{u} \end{bmatrix} = \boldsymbol{0} \tag{7.37}$$

或者

$$L2: \quad \begin{bmatrix} \boldsymbol{O} & \boldsymbol{K} \\ -\boldsymbol{N} & \boldsymbol{O} \end{bmatrix}\begin{bmatrix} \dot{\boldsymbol{u}} \\ \boldsymbol{u} \end{bmatrix} + \lambda\begin{bmatrix} \boldsymbol{M} & \boldsymbol{C} \\ \boldsymbol{O} & \boldsymbol{N} \end{bmatrix}\begin{bmatrix} \dot{\boldsymbol{u}} \\ \boldsymbol{u} \end{bmatrix} = \boldsymbol{0} \tag{7.38}$$

式中，$\boldsymbol{N}$可以是任意$n \times n$非奇异矩阵。

选择友矩阵对$(\boldsymbol{A}, \boldsymbol{B})$ 如下[45],[42]：

(1) 第1友型(first companion form)

$$L1: \quad \boldsymbol{A} = \begin{bmatrix} -\boldsymbol{C} & -\boldsymbol{K} \\ \boldsymbol{N} & \boldsymbol{O} \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{M} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{N} \end{bmatrix} \tag{7.39}$$

(2) 第2友型(second companion form)

$$L2: \qquad A=\begin{bmatrix} O & -K \\ N & O \end{bmatrix}, \qquad B=\begin{bmatrix} M & C \\ O & N \end{bmatrix} \tag{7.40}$$

则可通过矩阵束$(\boldsymbol{A},\boldsymbol{B})$ 的广义特征值分解，获得二次特征值问题的解。

有必要指出，任何一个高次特征值问题

$$(\lambda^m \boldsymbol{A}_m+\lambda^{m-1}\boldsymbol{A}_{m-1}+\cdots+\lambda \boldsymbol{A}_1+\boldsymbol{A}_0)\boldsymbol{u}=\boldsymbol{0} \tag{7.41}$$

都可以线性化，例如[41]

$$\begin{bmatrix} -\boldsymbol{A}_0 & & & \\ & \boldsymbol{I} & & \\ & & \ddots & \\ & & & \boldsymbol{I} \end{bmatrix}\begin{bmatrix} \boldsymbol{u} \\ \lambda\boldsymbol{u} \\ \vdots \\ \lambda^{m-1}\boldsymbol{u} \end{bmatrix}=\lambda\begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \cdots & \boldsymbol{A}_m \\ \boldsymbol{I} & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{O} & \cdots & \boldsymbol{I} & \boldsymbol{O} \end{bmatrix}\begin{bmatrix} \boldsymbol{u} \\ \lambda\boldsymbol{u} \\ \vdots \\ \lambda^{m-1}\boldsymbol{u} \end{bmatrix} \tag{7.42}$$

就是说，m次特征值问题也可以线性化为标准的广义特征值问题$\boldsymbol{Ax}=\lambda\boldsymbol{Bx}$求解。

7.1.6 矩阵的近似联合对角化

矩阵的联合对角化在统计信号处理中有大量的应用实例。

我们已知的事实有：

(1) 任何一个$n\times n$维Hermitian矩阵$\boldsymbol{A}$可对角化。可通过对$\boldsymbol{A}$作特征值分解来实现。

(2) 对于两个$n\times n$维非负定Hermitian矩阵$\boldsymbol{A}$和$\boldsymbol{B}$，存在一个非奇异矩阵$\boldsymbol{W}$，使得$\boldsymbol{W}^{\mathrm{H}}\boldsymbol{AW}$和$\boldsymbol{W}^{\mathrm{H}}\boldsymbol{BW}$同时对角化。事实上，这等价于矩阵束$(\boldsymbol{A},\boldsymbol{B})$的广义特征值分解。

更一般的情况下，对于一组Hermitian矩阵$\mathcal{A}=\{\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_K\}$，并不一定存在一个可逆矩阵$\boldsymbol{W}$使得它们同时对角化。然而，实际问题中，我们常期望求得一可逆矩阵$\boldsymbol{W}$，使得一组Hermitian矩阵$\mathcal{A}=\{\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_K\}$尽可能好的对角化，即使得以下的表达式尽可能好的得到拟合

$$\boldsymbol{A}_k=\boldsymbol{W}\boldsymbol{\Lambda}_k\boldsymbol{W}^{\mathrm{H}}, \quad 1\leqslant k\leqslant K \tag{7.43}$$

或

$$\boldsymbol{W}^{\mathrm{H}}\boldsymbol{A}_k\boldsymbol{W}=\boldsymbol{\Lambda}_k, \quad 1\leqslant k\leqslant K \tag{7.44}$$

其中$\boldsymbol{\Lambda}_1,\boldsymbol{\Lambda}_2,\cdots,\boldsymbol{\Lambda}_K$是$K$个对角阵。

我们分两种情况讨论。

1. 正交近似联合对角化

若矩阵组$\mathcal{A}$中有一个正定矩阵，不妨设为$\boldsymbol{A}_1$，则我们可将$\mathcal{A}$中的矩阵"白化"为

$$\mathcal{A}'=\{\boldsymbol{I},\boldsymbol{A}_1^{-\frac{\mathrm{H}}{2}}\boldsymbol{A}_2\boldsymbol{A}_1^{-\frac{1}{2}},\cdots,\boldsymbol{A}_1^{-\frac{\mathrm{H}}{2}}\boldsymbol{A}_K\boldsymbol{A}_1^{-\frac{1}{2}}\}$$

其中$\boldsymbol{A}_1^{\frac{1}{2}}$表示正定矩阵$\boldsymbol{A}_1$的平方根。这时若能求得一酉矩阵$\boldsymbol{U}$ 使得矩阵组$\mathcal{A}'$近似联合对角化，则矩阵$\boldsymbol{W}=\boldsymbol{U}\boldsymbol{A}_1^{-\frac{1}{2}}$ 将使得矩阵组$\mathcal{A}$近似联合对角化。这样，矩阵的近似联合对角化问题简化为矩阵的正交近似联合对角化问题。

考虑加权系数相同的情况，且假定矩阵组$\mathcal{A}=\{\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_K\}$已经被预白化。此时，由于$\|\cdot\|_{\mathrm{F}}^2\geqslant 0$，故我们可取联合对角化准则

$$\min \quad J(\boldsymbol{U})=\sum_{k=1}^{K}\left\|\boldsymbol{A}_k-\boldsymbol{U}\boldsymbol{A}_k\boldsymbol{U}^{\mathrm{H}}\right\|_{\mathrm{F}}^2 \tag{7.45}$$

式中，$\boldsymbol{U}$为酉矩阵。由于$\boldsymbol{U}$为酉矩阵，与准则式(7.45)等价的是下式的准则

$$\min \quad J(\boldsymbol{U})=\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\left|[\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}_k\boldsymbol{U}]_{ij}\right|^2 \tag{7.46}$$

其中，$[\boldsymbol{X}]_{ij}$表示矩阵$\boldsymbol{X}$的第i行、第j列元素。

正交近似联合对角化的最小二乘算法和子空间算法基于准则式(7.45)。基于矩阵旋转的Givens旋转算法、Schur分解以及Jacobi算法等基于准则式(7.46)。

在正交近似联合对角化中，由于所选定的正定矩阵$\boldsymbol{A}_1$总是被精确对角化的，这导致对角化误差分配不均匀，为此，人们提出非正交近似联合对角化，即直接求得对角化矩阵$\boldsymbol{W}$。

2. 非正交近似联合对角化

传统的非正交近似联合对角化算法大都采取最小二乘准则，即优化目标函数

$$\min \quad J_{\mathrm{WLS1}}(\boldsymbol{W})=\sum_{k=1}^{K}\alpha_k\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\left|[\boldsymbol{W}^{\mathrm{H}}\boldsymbol{A}_k\boldsymbol{W}]_{ij}\right|^2,\quad \boldsymbol{W}\neq\boldsymbol{O} \tag{7.47}$$

或

$$\min \quad J_{\mathrm{WLS2}}(\boldsymbol{W},\boldsymbol{\Lambda}_1,\boldsymbol{\Lambda}_2,\cdots,\boldsymbol{\Lambda}_K)=\sum_{k=1}^{K}\alpha_k\|\boldsymbol{A}_k-\boldsymbol{W}\boldsymbol{\Lambda}_k\boldsymbol{W}^{\mathrm{H}}\|_{\mathrm{F}}^2 \tag{7.48}$$

其中，准则式(7.47)中的$\boldsymbol{W}$应避免平凡解零解$\boldsymbol{W}=\boldsymbol{O}$，$\alpha_k>0$为正的权重。Li和Zhang指出，最小二乘联合对角化准则无法排除奇异的或病态的对角化矩阵解（退化解），这严重限制了它们的性能。为此，Li和Zhang提出了下面的近似联合对角化准则[9]

$$\min \quad J(\boldsymbol{W})=\sum_{k=1}^{K}\alpha_k\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\left|[\boldsymbol{W}^{\mathrm{H}}\boldsymbol{A}_k\boldsymbol{W}]_{ij}\right|^2-\beta\log\mathrm{abs}\det(\boldsymbol{W}) \tag{7.49}$$

其中，β是正的权重，但它对对角化算法的性能没有影响（在相差一标量因子的差别下），为此我们常取$\beta=1$。

关于近似联合对角化准则式(7.49)，我们有下面的结论：

当且仅当存在一可逆矩阵$\boldsymbol{W}$同时精确对角化矩阵组$\mathcal{A}$中的所有矩阵时，对角化准则式(7.49)下无界。在近似联合对角化中，对角化准则式(7.49)的极小解的对角化误差的上界随目标函数的值指数下降。

特别地，对于正定矩阵组的近似联合对角化，我们可以利用Hadamard不等式来构造近似联合对角化准则。Hadamard不等式可叙述为：若矩阵$\boldsymbol{A}$正定，则有不等式

$$\det(\boldsymbol{A})\leqslant\det[\mathrm{diag}(\boldsymbol{A})]$$

当且仅当$\boldsymbol{A}$是对角矩阵，等式成立。由此，Pham构造出以下正定矩阵组的近似联合对角化准则[12]：

$$\min \quad J(\boldsymbol{W})=\frac{1}{2}\sum_{k=1}^{K}\alpha_k \log\frac{\det \operatorname{diag}(\boldsymbol{W}^{\mathrm{H}}\boldsymbol{A}_k\boldsymbol{W})}{\det(\boldsymbol{W}^{\mathrm{H}}\boldsymbol{A}_k\boldsymbol{W})} \tag{7.50}$$

非正交近似联合对角化算法大都较复杂，可参见具体的参考文献。

7.2 习题与解答

7.1 证明当$\boldsymbol{A}$为幂等矩阵时，矩阵$\boldsymbol{BA}$的特征值与$\boldsymbol{ABA}$的特征值相同。

证明：通过证明两个矩阵的特征多项式相同来证明两个矩阵的特征值相同。设$\boldsymbol{A}$为$n\times n$维的幂等矩阵，其秩为$r(r\leqslant n)$。由于$\boldsymbol{A}$为幂等矩阵，可设$\boldsymbol{A}$有相似形分解形式

$$\boldsymbol{A}=\boldsymbol{P}\begin{bmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{0}\end{bmatrix}\boldsymbol{P}^{-1} \tag{7.51}$$

其中$\boldsymbol{I}_r$是$r\times r$维的单位阵。则有

$$\begin{aligned}\boldsymbol{BA} &= \boldsymbol{BP}\begin{bmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{0}\end{bmatrix}\boldsymbol{P}^{-1}\\ &= \boldsymbol{PP}^{-1}\boldsymbol{BP}\begin{bmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{0}\end{bmatrix}\boldsymbol{P}^{-1}\\ &\sim \boldsymbol{P}^{-1}\boldsymbol{BP}\begin{bmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{0}\end{bmatrix}\\ &= \boldsymbol{C}\begin{bmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{0}\end{bmatrix}\end{aligned}$$

这里记$\boldsymbol{C}=\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P}$，$\sim$代表相似于。将$\boldsymbol{C}$分块为

$$\boldsymbol{C}=\begin{bmatrix}\boldsymbol{C}_1 & \boldsymbol{C}_2\\ \boldsymbol{C}_3 & \boldsymbol{C}_4\end{bmatrix}$$

其中$\boldsymbol{C}_1$为$r\times r$的方阵。利用相似矩阵具有相同的特征多项式的性质，可计算矩阵$\boldsymbol{BA}$的特征多项式为

$$\begin{aligned}\det(\lambda\boldsymbol{I}-\boldsymbol{BA}) &= \det\left(\begin{bmatrix}\lambda\boldsymbol{I}_r-\boldsymbol{C}_1 & \boldsymbol{O}\\ -\boldsymbol{C}_3 & \lambda\boldsymbol{I}_{n-r}\end{bmatrix}\right)\\ &= \lambda^{n-r}\det(\lambda\boldsymbol{I}_r-\boldsymbol{C}_1)\end{aligned} \tag{7.52}$$

同理，我们可以求得$\boldsymbol{ABA}$的一个相似形为

$$\begin{aligned}\boldsymbol{ABA} &= \boldsymbol{P}\begin{bmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{0}\end{bmatrix}\boldsymbol{P}^{-1}\boldsymbol{BP}\begin{bmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{0}\end{bmatrix}\boldsymbol{P}^{-1}\\ &\sim \begin{bmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{0}\end{bmatrix}\boldsymbol{P}^{-1}\boldsymbol{BP}\begin{bmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{0}\end{bmatrix}\\ &= \begin{bmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{0}\end{bmatrix}\boldsymbol{C}\begin{bmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{0}\end{bmatrix}\\ &= \begin{bmatrix}\boldsymbol{C}_1 & \\ & \boldsymbol{0}\end{bmatrix}\end{aligned}$$

由此得矩阵$\boldsymbol{ABA}$的特征多项式为

$$\begin{aligned}\det(\lambda\boldsymbol{I}-\boldsymbol{ABA}) &= \det\left(\begin{bmatrix}\lambda\boldsymbol{I}_r-\boldsymbol{C}_1 & \\ & \lambda\boldsymbol{I}_{n-r}\end{bmatrix}\right)\\ &= \lambda^{n-r}\det(\lambda\boldsymbol{I}_r-\boldsymbol{C}_1)\end{aligned} \tag{7.53}$$

比较式(7.52)和式(7.53)知$\boldsymbol{BA}$和$\boldsymbol{ABA}$有相同的特征多项式，因此，它们有相同的特征值(包括相同特征值的代数重度也相同)。

注释1：通常情况下，我们不能通过证明命题"若$\alpha \in \lambda(\boldsymbol{A})$，则$\alpha \in \lambda(\boldsymbol{B})$；且若$\beta \in \lambda(\boldsymbol{B})$，则$\beta \in \lambda(\boldsymbol{A})$"来证明$\lambda(\boldsymbol{A})$与$\lambda(\boldsymbol{B})$是完全相同的。这是因为$\lambda(\boldsymbol{A})$与$\lambda(\boldsymbol{B})$可能有重复的特征值。例如，若

$$\lambda(\boldsymbol{A}) = \{1, 1, 2\}$$

$$\lambda(\boldsymbol{B}) = \{1, 2, 2\}$$

则命题"若$\alpha \in \lambda(\boldsymbol{A})$，则$\alpha \in \lambda(\boldsymbol{B})$；且若$\beta \in \lambda(\boldsymbol{B})$，则$\beta \in \lambda(\boldsymbol{A})$"成立，但$\lambda(\boldsymbol{A})$与$\lambda(\boldsymbol{B})$却是不同的。

注意，我们常通过证明命题"若$a \in A$，则$a \in B$；且若$b \in B$，则$b \in A$"来证明两个集合A和B相等。在集合中，重复的元素被看成是一个元素，但特征值可能有重复的，且重复的次数也是我们关心的，因此，我们不能像证明两个集合相等一样证明两个矩阵的特征值相同。

注释2：幂等矩阵是可对角化的，且其特征值为0或1，因此我们可以得到对角化形式(7.51)。更一般的，设矩阵$\boldsymbol{A}$满足$\boldsymbol{A}^n = \boldsymbol{A}$，这里$n$是一正整数，则矩阵$\boldsymbol{A}$的一个零化多项式为$p(x) = x^n - x$。该零化多项式$p(x)$无重根，因此矩阵$\boldsymbol{A}$的最小多项式也必无重根，故矩阵$\boldsymbol{A}$可对角化。

7.2 设n阶矩阵$\boldsymbol{A}$的全部元素为1，求$\boldsymbol{A}$的n个特征值。

解：由于$\text{rank}(\boldsymbol{A}) = 1$，故$\boldsymbol{A}$有$n-1$个零特征值。唯一的一个非零特征值对应的特征向量为

$$\boldsymbol{p} = [\underbrace{1, \cdots, 1}_{n}]^{\mathrm{T}}$$

容易验证特征向量$\boldsymbol{p}$对应的特征值为$\frac{1}{n}$。综上可得，$\boldsymbol{A}$的n个特征值为

$$\frac{1}{n}, 0, \cdots, 0$$

7.3 设矩阵

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

(1) 已知$\boldsymbol{A}$的一个特征值为3，试求y值。

(2) 求矩阵$\boldsymbol{P}$，使$(\boldsymbol{AP})^{\mathrm{T}}\boldsymbol{AP}$为对角矩阵。

解：(1) 观察知$\boldsymbol{A}$为块对角化矩阵，容易计算$\lambda\boldsymbol{I}-\boldsymbol{A}$为

$$\begin{aligned}\lambda\boldsymbol{I}-\boldsymbol{A} &= \begin{vmatrix} \lambda & -1 & 0 & 0 \\ -1 & \lambda & 0 & 0 \\ 0 & 0 & \lambda-y & -1 \\ 0 & 0 & -1 & \lambda-2 \end{vmatrix} \\ &= (\lambda^2-1)[\lambda^2-(2+y)\lambda+2y-1]\end{aligned}$$

由于3是$\boldsymbol{A}$的一个特征值，故$|3\boldsymbol{I}-\boldsymbol{A}|=0$，由此容易解得$y=2$，由此矩阵$\boldsymbol{A}$为

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

(2) 显然这样的矩阵$\boldsymbol{P}$不是唯一的，我们求取这样的正交矩阵$\boldsymbol{P}$。观察知$\boldsymbol{A}$为块对角化的对称矩阵，矩阵$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$也是块对角化的，容易求得为

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 4 & 5 \end{bmatrix}$$

求得子块$\begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$的对角化矩阵为$\begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$。总的对角化矩阵$\boldsymbol{P}$为

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

7.4 令初始值$u(0)=2, v(0)=8$。利用特征值求解微分方程

$$u'(t)=3u(t)+v(t)$$

$$v'(t)=-2u(t)+v(t)$$

解：将微分方程写成矩阵形式

$$\begin{bmatrix} u'(t) \\ v'(t) \end{bmatrix}=\begin{bmatrix} 3 & 1 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} u(t) \\ v(t) \end{bmatrix}$$

$$\boldsymbol{x}'(t)=\boldsymbol{A}\boldsymbol{x}(t)$$

这里$\boldsymbol{x}(t)=\begin{bmatrix} u(t) \\ v(t) \end{bmatrix}$，$\boldsymbol{A}=\begin{bmatrix} 3 & 1 \\ -2 & 1 \end{bmatrix}$。易知$\boldsymbol{x}(t)$的解为

$$\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}(0)$$

其中$\boldsymbol{x}(0)=\begin{bmatrix}2\\8\end{bmatrix}$为初始向量。容易求得$\boldsymbol{A}$的特征值分解为

$$\boldsymbol{A}=\begin{bmatrix}\frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3}\\ -\frac{\sqrt{3}}{3}+\frac{\sqrt{3}}{3}\mathrm{j} & -\frac{\sqrt{3}}{3}-\frac{\sqrt{3}}{3}\mathrm{j}\end{bmatrix}\begin{bmatrix}2+\mathrm{j} & 0\\ 0 & 2-\mathrm{j}\end{bmatrix}\begin{bmatrix}\frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3}\\ -\frac{\sqrt{3}}{3}+\frac{\sqrt{3}}{3}\mathrm{j} & -\frac{\sqrt{3}}{3}-\frac{\sqrt{3}}{3}\mathrm{j}\end{bmatrix}^{-1}$$

这里$\mathrm{j}=\sqrt{-1}$。由此解得

$$\begin{aligned}\boldsymbol{x}(t) &= \begin{bmatrix}\frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3}\\ -\frac{\sqrt{3}}{3}+\frac{\sqrt{3}}{3}\mathrm{j} & -\frac{\sqrt{3}}{3}-\frac{\sqrt{3}}{3}\mathrm{j}\end{bmatrix}\begin{bmatrix}\mathrm{e}^{(2+\mathrm{j})t} & 0\\ 0 & \mathrm{e}^{(2-\mathrm{j})t}\end{bmatrix}\\ &\quad\times\begin{bmatrix}\frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3}\\ -\frac{\sqrt{3}}{3}+\frac{\sqrt{3}}{3}\mathrm{j} & -\frac{\sqrt{3}}{3}-\frac{\sqrt{3}}{3}\mathrm{j}\end{bmatrix}^{-1}\begin{bmatrix}2\\8\end{bmatrix}\\ &= \mathrm{e}^{2t}\begin{bmatrix}2\cos(t)+10\sin(t)\\ 8\cos(t)-12\sin(t)\end{bmatrix}\end{aligned}$$

7.5 令4×4维Hessenberg 矩阵

$$\boldsymbol{H}=\begin{bmatrix}a_1 & b_1 & c_1 & d_1\\ a_2 & b_2 & c_2 & d_2\\ 0 & b_3 & c_3 & d_3\\ 0 & 0 & c_4 & d_4\end{bmatrix}$$

证明以下两个结果：

(1) 若a_2,b_3,c_4均不等于零，并且$\boldsymbol{H}$的任意特征值λ为实数，则λ的几何多重度必定等于1。

(2) 若$\boldsymbol{H}$与对称矩阵$\boldsymbol{A}$相似，并且$\boldsymbol{A}$的某个特征值λ的代数多重度大于1，则a_2,b_3,c_4至少有一个等于零。

证明： (1) 由于a_2,b_3,c_4均不等于零，故对任意的实特征值λ，矩阵$\lambda\boldsymbol{I}-\boldsymbol{A}$的秩都是3，即仅有一个独立的特征向量，因此$\lambda$的几何多重度必定等于1。

(2) 若$\boldsymbol{H}$与对称矩阵$\boldsymbol{A}$相似，则$\boldsymbol{H}$必可对角化，且与$\boldsymbol{A}$有相同的特征值。又$\boldsymbol{A}$的某个特征值λ的代数多重度大于1，则$\boldsymbol{H}$的相应特征值λ的代数多重度也大于1。由于$\boldsymbol{H}$可对角化，$\boldsymbol{H}$的特征值λ的几何多重度必大于1，由第一问中的结论知，此时a_2,b_3,c_4中至少有一个等于零。

7.6 证明以下各题：

(1) 若$\boldsymbol{A}=\boldsymbol{A}^2=\boldsymbol{A}^{\mathrm{H}}\in C^{n\times n}$，并且$\mathrm{rank}(\boldsymbol{A})=r<n$，则存在$n\times n$酉矩阵$\boldsymbol{V}$，使得

$$\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}=\mathrm{diag}(\boldsymbol{I}_r,\boldsymbol{0})$$

(2) 若$\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}\in C^{n\times n}$，并且$\boldsymbol{A}^2=\boldsymbol{I}_n$，则存在酉矩阵$\boldsymbol{V}$，使得

$$\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}=\mathrm{diag}(\boldsymbol{I}_r,-\boldsymbol{I}_{n-r})$$

证明：(1) 考虑到$\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}$，故$\boldsymbol{A}$有分解形式

$$\boldsymbol{A}=\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}$$

这里$\boldsymbol{V}$为酉阵，$\boldsymbol{\Sigma}$为对角阵。由$\boldsymbol{A}^2=\boldsymbol{A}$可得

$$\boldsymbol{\Sigma}^2=\boldsymbol{\Sigma}$$

故对角阵$\boldsymbol{\Sigma}$的对角线上的元素满足关系式$\boldsymbol{\Sigma}_i^2=\boldsymbol{\Sigma}_i$，即$\boldsymbol{\Sigma}_i$为0或者1。总可以交换$\boldsymbol{V}$中列向量的顺序，使得所有的特征值1集中在一起，所有的特征值0也集中在一起。由于$\boldsymbol{A}$的秩为r，故经交换后，$\boldsymbol{\Sigma}$将有如下形式

$$\boldsymbol{\Sigma}=\mathrm{diag}(\boldsymbol{I}_r,\boldsymbol{0})$$

考虑到$\boldsymbol{V}^{-1}=\boldsymbol{V}^{\mathrm{H}}$，即得所需结论。

(2) 同(1)中的记号，由$\boldsymbol{A}^2=\boldsymbol{I}$可得$\boldsymbol{\Sigma}^2=\boldsymbol{I}$。即$\boldsymbol{\Sigma}_i=1$或者$-1$。同样，可交换矩阵$\boldsymbol{V}$中列向量的顺序，使得$\boldsymbol{\Sigma}$具有形式

$$\boldsymbol{\Sigma}=\mathrm{diag}(\boldsymbol{I}_r,-\boldsymbol{I}_{n-r})$$

这里r和$n-r$分别为$\boldsymbol{A}$的正和负惯性指数。考虑到$\boldsymbol{V}^{-1}=\boldsymbol{V}^{\mathrm{H}}$，即得所需结论。

7.7 令$\boldsymbol{H}=\boldsymbol{I}-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{H}}$为Householder变换矩阵。

(1) 证明：具有单位范数的向量$\boldsymbol{u}$是Householder 变换矩阵的特征向量，并求与之相对应的特征值。

(2) 若向量$\boldsymbol{w}$与$\boldsymbol{u}$ 正交，证明$\boldsymbol{w}$是矩阵$\boldsymbol{H}$的特征向量，并求与之对应的特征向量。

证明：(1) 容易计算得

$$\boldsymbol{H}\boldsymbol{u}=-\boldsymbol{u}$$

故$\boldsymbol{u}$为$\boldsymbol{H}$的特征值为-1的特征向量。

(2) 容易计算得

$$\boldsymbol{H}\boldsymbol{w}=\boldsymbol{w}$$

故$\boldsymbol{w}$是$\boldsymbol{H}$的特征值为1的特征向量。

7.8 设矩阵$\boldsymbol{A}$和$\boldsymbol{B}$相似，其中

$$\boldsymbol{A}=\begin{bmatrix}-2&0&0\\2&x&2\\3&1&1\end{bmatrix},\qquad \boldsymbol{B}=\begin{bmatrix}-1&0&0\\0&2&0\\0&0&y\end{bmatrix}$$

(1) 求x和y的值。

(2) 求可逆矩阵$\boldsymbol{P}$，使得$\boldsymbol{P}^{-1}\boldsymbol{AP}=\boldsymbol{B}$。

解：(1) 计算$\boldsymbol{A}$的特征多项式得

$$|\lambda\boldsymbol{I}-\boldsymbol{A}|=(\lambda+2)[\lambda^2-(x+1)\lambda+x-2]$$

由于矩阵$\boldsymbol{A}$和$\boldsymbol{B}$相似且$\boldsymbol{B}$为对角化形式，故矩阵$\boldsymbol{A}$有特征根$-1,2$和y。令$\lambda=-1$和$|\lambda\boldsymbol{I}-\boldsymbol{A}|=0$可解得$x=0$。因此有$|\lambda\boldsymbol{I}-\boldsymbol{A}|=(\lambda+2)(\lambda^2-\lambda-2)$，故得特征值$\lambda(\boldsymbol{A})=\{-2,2,-1\}$，进一步得$y=-2$。

(2) 通过求这些特征根对应的特征向量，不难求得矩阵$\boldsymbol{P}$为

$$\boldsymbol{P}=\begin{bmatrix}0 & 0 & \sqrt{\frac{1}{2}}\\ \sqrt{\frac{4}{5}} & \sqrt{\frac{1}{2}} & 0\\ -\sqrt{\frac{1}{5}} & \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}}\end{bmatrix}$$

这里以求取特征值-1对应的特征向量为例来展示计算过程。令$\boldsymbol{p}=[p_1,p_2,p_3]^{\mathrm{T}}$为待求的特征值$-1$对应的非零特征向量，则有

$$\begin{bmatrix}-2 & 0 & 0\\ 2 & 0 & 2\\ 3 & 1 & 1\end{bmatrix}\begin{bmatrix}p_1\\ p_2\\ p_3\end{bmatrix}=(-1)\begin{bmatrix}p_1\\ p_2\\ p_3\end{bmatrix}$$

化简得方程组

$$\begin{cases}-p_1=0\\ 2p_1+p_2+2p_3=0\\ 3p_1+p_2+2p_3=0\end{cases}$$

它的一个非零解为

$$\begin{cases}p_1=0\\ p_2=2\\ p_3=-1\end{cases}$$

通常情况下，我们约定$\boldsymbol{p}$为单位长度的向量，因此有$\boldsymbol{p}=\left[0,\sqrt{\frac{4}{5}},-\sqrt{\frac{1}{5}}\right]^{\mathrm{T}}$。其他的特征值对应的特征向量可以类似的求得，它们组成矩阵$\boldsymbol{P}$。类似的计算常在下面的习题中遇到，这里不再详细叙述其计算过程。

7.9 利用分块矩阵，可以将矩阵的特征值问题降维。令

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{B} & \boldsymbol{X}\\ \boldsymbol{O} & \boldsymbol{C}\end{bmatrix}\quad(\boldsymbol{O}\text{ 为零矩阵})$$

证明$\det(\boldsymbol{A}-\lambda\boldsymbol{I})=\det(\boldsymbol{B}-\lambda\boldsymbol{I})\det(\boldsymbol{C}-\lambda\boldsymbol{I})$。

证明：由行列式的性质即得。

7.10 令$\boldsymbol{u}$是矩阵$\boldsymbol{A}$与特征值λ对应的一个特征向量。

(1) 证明$\boldsymbol{u}$是矩阵$\boldsymbol{A}-3\boldsymbol{I}$的一个特征向量。

(2) 证明$\boldsymbol{u}$是矩阵$\boldsymbol{A}^2-4\boldsymbol{A}+3\boldsymbol{I}$的一个特征向量。

证明：(1) 容易计算得

$$(\boldsymbol{A}-3\boldsymbol{I})\boldsymbol{u}=(\lambda-3)\boldsymbol{u}$$

故$\boldsymbol{u}$是矩阵$(\boldsymbol{A}-3\boldsymbol{I})$的一个特征向量。

(2) 命题可通过类似的计算证明。

7.11 [18] 令

$$\boldsymbol{A}=\begin{bmatrix}-2 & 4 & 3\\ 0 & 0 & 0\\ -1 & 5 & 2\end{bmatrix},\qquad f(x)=x^{593}-2x^{15}$$

证明

$$\boldsymbol{A}^{593}-2\boldsymbol{A}^{15}=-\boldsymbol{A}=\begin{bmatrix}2 & -4 & -3\\ 0 & 0 & 0\\ 1 & -5 & -2\end{bmatrix}$$

解：计算$|x\boldsymbol{I}-\boldsymbol{A}|$，可得$\boldsymbol{A}$的一个化零多项式为

$$p(x)=|x\boldsymbol{I}-\boldsymbol{A}|=x(x-1)(x+1)$$

令$f(x)$除$p(x)$的余式为$r(x)=ax^2+bx+c$，即$f(x)=p(x)q(x)+r(x)$，分别代入$x=0,-1,1$可得关系式

$$\begin{cases}f(0)=r(0)\\ f(-1)=r(-1)\\ f(1)=r(1)\end{cases}$$

即

$$\begin{cases}c=0\\ a-b+c=1\\ a+b+c=-1\end{cases}$$

解得$a=c=0,b=-1$，即余式$r(x)=-x$，故有$\boldsymbol{A}^{593}-2\boldsymbol{A}^{15}=-\boldsymbol{A}$。

7.12 [18] 假定$a_0,a_1,a_2,\cdots$为正整数序列，并且满足递推关系$a_{k+1}=a_k+2a_{k-1}$，$\forall\, k\geqslant 1$。若$a_0=0$，$a_1=1$，求a_k值。（提示：建立向量$\begin{bmatrix}a_{k+1}\\ a_k\end{bmatrix}$与$\begin{bmatrix}a_1\\ a_0\end{bmatrix}$之间的关系，并运用Cayley-Hamilton定理。）

解：首先，我们写出向量$\begin{bmatrix}a_{k+1}\\ a_k\end{bmatrix}$与$\begin{bmatrix}a_k\\ a_{k-1}\end{bmatrix}$之间的关系式

$$\begin{bmatrix}a_{k+1}\\ a_k\end{bmatrix}=\begin{bmatrix}1 & 2\\ 1 & 0\end{bmatrix}\begin{bmatrix}a_k\\ a_{k-1}\end{bmatrix}$$

于是有

$$\begin{bmatrix}a_{k+1}\\ a_k\end{bmatrix}=\begin{bmatrix}1 & 2\\ 1 & 0\end{bmatrix}^k\begin{bmatrix}a_1\\ a_0\end{bmatrix}$$

令$A=\begin{bmatrix}1 & 2\\ 1 & 0\end{bmatrix}$，容易计算得$|\lambda I-A|=(\lambda-2)(\lambda+1)$，由Cayley-Hamilton定理，我们可设$A^k=c_0I+c_1A$，这里系数$c_0$，$c_1$满足

$$\begin{aligned}c_0+(-1)c_1&=(-1)^k\\ c_0+2c_1&=2^k\end{aligned}$$

由此可解得c_0，c_1为

$$c_0=\frac{2(-1)^k+2^k}{3}$$
$$c_1=\frac{2^k-(-1)^k}{3}$$

故

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

考虑到初始条件$a_0=0,a_1=1$，我们有

$$\begin{bmatrix}a_{k+1}\\ a_k\end{bmatrix}=\frac{1}{3}\begin{bmatrix}2^{k+1}+(-1)^k & 2^{k+1}-2(-1)^k\\ 2^k-(-1)^k & 2(-1)^k+2^k\end{bmatrix}\begin{bmatrix}1\\ 0\end{bmatrix}$$

故有

$$a_k=\frac{2^k}{3}-\frac{(-1)^k}{3}$$

注释1：我们也可以通过矩阵的Jordan标准型来计算矩阵函数。此例中，矩阵A可对角化为

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

故有

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

考虑到初始条件$a_0=0,a_1=1$，可求得a_k的表达式为

$$a_k=\begin{bmatrix}\sqrt{\frac{1}{5}} & \sqrt{\frac{1}{2}}\end{bmatrix}\begin{bmatrix}2^k & 0\\ 0 & (-1)^k\end{bmatrix}\begin{bmatrix}\sqrt{\frac{4}{5}} & -\sqrt{\frac{1}{2}}\\ \sqrt{\frac{1}{5}} & \sqrt{\frac{1}{2}}\end{bmatrix}^{-1}\begin{bmatrix}1\\ 0\end{bmatrix}$$

化简得

$$a_k=\frac{2^k}{3}-\frac{(-1)^k}{3}$$

由此我们得到同样的结果。在处理较简单的问题时，运用Cayley-Hamilton定理似乎显得更灵活简单些。但当矩阵函数形式较复杂时，通过Jordan标准型的计算或许更简洁。

注释2：此例中的正整数系列a_k也可以通过$\mathcal{Z}$变换的方式求得。这里a_k满足的差分方程为$z^2-z-2=0$，它的两个根为$z_1=-1, z_2=2$，因此可设a_k的通解表达式为

$$a_k=\alpha(-1)^k+\beta 2^k$$

其中的系数α，β可通过初始条件$a_0=0$，$a_1=1$确定。这导致同样的结果。

7.13 已知矩阵

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

求$\mathrm{e}^{\boldsymbol{A}t}$。

解：容易通过特征值分解求得$\boldsymbol{A}$的对角化形式为

$$\boldsymbol{A}=\begin{bmatrix}1&0&\dfrac{\sqrt{2}}{2}\\0&1&0\\0&0&\dfrac{\sqrt{2}}{2}\end{bmatrix}\begin{bmatrix}2&0&0\\0&2&0\\0&0&3\end{bmatrix}\begin{bmatrix}1&0&\dfrac{\sqrt{2}}{2}\\0&1&0\\0&0&\dfrac{\sqrt{2}}{2}\end{bmatrix}^{-1}$$

故有

$$\mathrm{e}^{\boldsymbol{A}t}=\begin{bmatrix}1&0&\dfrac{\sqrt{2}}{2}\\0&1&0\\0&0&\dfrac{\sqrt{2}}{2}\end{bmatrix}\begin{bmatrix}\mathrm{e}^{2t}&0&0\\0&\mathrm{e}^{2t}&0\\0&0&\mathrm{e}^{3t}\end{bmatrix}\begin{bmatrix}1&0&\dfrac{\sqrt{2}}{2}\\0&1&0\\0&0&\dfrac{\sqrt{2}}{2}\end{bmatrix}^{-1}$$

化简得

$$\mathrm{e}^{\boldsymbol{A}t}=\begin{bmatrix}\mathrm{e}^{2t}&0&\mathrm{e}^{3t}-\mathrm{e}^{2t}\\0&\mathrm{e}^{2t}&0\\0&0&\mathrm{e}^{3t}\end{bmatrix}$$

7.14 [18] 已知矩阵

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

求非奇异矩阵$\boldsymbol{S}$使相似矩阵$\boldsymbol{B}=\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$为对角矩阵。

解：显然，这样的$\boldsymbol{S}$是不唯一的，这里我们求取列向量为单位长度的变换矩阵。首先，求得$\boldsymbol{A}$的全部特征值和特征向量为

$$\lambda=\{1,2,3\}$$

和

$$\boldsymbol{p}=\left\{\begin{bmatrix}\dfrac{\sqrt{6}}{6}\\\dfrac{\sqrt{6}}{3}\\-\dfrac{\sqrt{6}}{6}\end{bmatrix},\begin{bmatrix}-\dfrac{\sqrt{3}}{3}\\\dfrac{\sqrt{3}}{3}\\-\dfrac{\sqrt{3}}{3}\end{bmatrix},\begin{bmatrix}\dfrac{\sqrt{2}}{2}\\0\\\dfrac{\sqrt{2}}{2}\end{bmatrix}\right\}$$

故取$\boldsymbol{S}$为

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

则$\boldsymbol{B}$将为对角阵

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

7.15 证明：满足$\boldsymbol{A}^2-\boldsymbol{A}=2\boldsymbol{I}$的矩阵$\boldsymbol{A}_{n\times n}$是可对角化的。

证明：由题设知$(\boldsymbol{A}-2\boldsymbol{I})(\boldsymbol{A}+\boldsymbol{I})=\boldsymbol{O}$，即$\boldsymbol{A}-2\boldsymbol{I}$和$\boldsymbol{A}+\boldsymbol{I}$的列分别是矩阵$\boldsymbol{A}$的特征值$-1$和2对应的特征向量。我们证明这样的独立的列共有$n$个，这样矩阵$\boldsymbol{A}$便可对角化。

首先，由$(\boldsymbol{A}-2\boldsymbol{I})(\boldsymbol{A}+\boldsymbol{I})=\boldsymbol{O}$知

$$\operatorname{rank}(\boldsymbol{A}-2\boldsymbol{I})+\operatorname{rank}(\boldsymbol{A}+\boldsymbol{I})\leqslant n$$

其次，我们有

$$\operatorname{rank}(\boldsymbol{A}-2\boldsymbol{I})+\operatorname{rank}(\boldsymbol{A}+\boldsymbol{I})\geqslant \operatorname{rank}[(\boldsymbol{A}-2\boldsymbol{I})-(\boldsymbol{A}+\boldsymbol{I})]=n$$

故必有$\operatorname{rank}(\boldsymbol{A}-2\boldsymbol{I})+\operatorname{rank}(\boldsymbol{A}+\boldsymbol{I})=n$。即矩阵$\boldsymbol{A}$有$n$个独立的特征向量，因此可对角化。

注释：事实上，从以下的两点观察也可知矩阵$\boldsymbol{A}$可对角化：

(1) 容易配方得到$(2\boldsymbol{A}-\boldsymbol{I})^2=9\boldsymbol{I}$，因此，$\boldsymbol{A}$可对角化。

(2) 由题设知$\boldsymbol{A}$的一个化零多项式为$p(x)=x^2-x-2$，该多项式无重根，因此$\boldsymbol{A}$的最小多项式也必无重根，故$\boldsymbol{A}$可对角化。

7.16 已知$\boldsymbol{u}=[1,1,-1]^{\mathrm{T}}$是矩阵

$$\boldsymbol{A}=\begin{bmatrix} 2 & -1 & 2 \\ 5 & a & 3 \\ -1 & b & -2 \end{bmatrix}$$

的一个特征向量。

(1) 求a,b和特征向量$\boldsymbol{u}$对应的特征值。

(2) 矩阵$\boldsymbol{A}$能否相似于对角矩阵？试说明理由。

解：(1) 令$\boldsymbol{A}\boldsymbol{u}=\lambda\boldsymbol{u}$，容易解得$\lambda=-1$和$a=-3,b=0$，矩阵$\boldsymbol{A}$为

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

(2) 令$|\lambda\boldsymbol{I}-\boldsymbol{A}|=0$，得方程$(\lambda+1)^3=0$，即$\boldsymbol{A}$的特征值$\lambda=-1$的代数多重度为3，显然，其几何多重度小于3，因此，$\boldsymbol{A}$不可对角化。

注释：如果使用数值计算软件，如MATLAB或SCILAB，计算$\boldsymbol{A}$的特征值分解，则由于计算误差的原因，会得到$\boldsymbol{A}$的3个不同特征值，即会错误地得到$\boldsymbol{A}$可对角化的结论。在MATLAB中可用命令“condeig(A)”来检查特征值的条件数，例如本题中的矩阵$\boldsymbol{A}$的3个特征值条件数约为5.8×10^9，这时，数值计算得到的结果就不一定可靠了。

7.17 证明：矩阵$\boldsymbol{A}$和$\boldsymbol{B}$的Kronecker积$\boldsymbol{A} \otimes \boldsymbol{B}$ 的非零特征值等于$\boldsymbol{A}$的特征值与$\boldsymbol{B}$的特征值的乘积，即$\lambda(\boldsymbol{A} \otimes \boldsymbol{B}) = \lambda(\boldsymbol{A})\lambda(\boldsymbol{B})$。

证明：令$\boldsymbol{A}$和$\boldsymbol{B}$的一个特征对分别为$\{\lambda_A, \boldsymbol{u}_A\}$和$\{\lambda_B, \boldsymbol{u}_B\}$，其中$\lambda_A$和$\lambda_B$均为非零的数。容易计算得

$$\begin{aligned}(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{u}_A \otimes \boldsymbol{u}_B) &= (\boldsymbol{A}\boldsymbol{u}_A) \otimes (\boldsymbol{B}\boldsymbol{u}_B) \\ &= \lambda_A \boldsymbol{u}_A \otimes \lambda_B \boldsymbol{u}_B \\ &= \lambda_A \lambda_B \boldsymbol{u}_A \otimes \boldsymbol{u}_B\end{aligned}$$

即$\boldsymbol{A}$与$\boldsymbol{B}$的非零特征值的乘积为$\boldsymbol{A} \otimes \boldsymbol{B}$ 的非零特征值。

7.18 令$\boldsymbol{A}$ 和$\boldsymbol{B}$均为Hermitian矩阵，并且λ_i 和μ_i分别是矩阵$\boldsymbol{A}$和$\boldsymbol{B}$ 的特征值。证明：若$c_1\boldsymbol{A} + c_2\boldsymbol{B}$具有特征值$c_1\lambda_i + c_2\mu_i$，其中$c_1, c_2$ 为任意标量，则$\boldsymbol{AB} = \boldsymbol{BA}$。

证明：由于$\boldsymbol{A}$和$\boldsymbol{B}$均为Hermitian矩阵，故存在酉阵$\boldsymbol{U}$和$\boldsymbol{V}$使得

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}_A\boldsymbol{U}^{\mathrm{H}}, \quad \boldsymbol{B} = \boldsymbol{V}\boldsymbol{\Sigma}_B\boldsymbol{V}^{\mathrm{H}}$$

其中对角阵$\boldsymbol{\Sigma}_A$和$\boldsymbol{\Sigma}_B$的对角线元素分别由$\boldsymbol{A}$和$\boldsymbol{B}$的特征值组成。

考虑到$c_1\boldsymbol{A} + c_2\boldsymbol{B}$也为Hermitian矩阵，且有特征值$c_1\lambda_i + c_2\mu_i$，故存在酉阵$\boldsymbol{W}$使得

$$c_1\boldsymbol{A} + c_2\boldsymbol{B} = \boldsymbol{W}(c_1\boldsymbol{\Sigma}_A + c_2\boldsymbol{\Sigma}_B)\boldsymbol{W}^{\mathrm{H}}$$

即有

$$c_1\boldsymbol{U}\boldsymbol{\Sigma}_A\boldsymbol{U}^{\mathrm{H}} + c_2\boldsymbol{V}\boldsymbol{\Sigma}_B\boldsymbol{V}^{\mathrm{H}} = c_1\boldsymbol{W}\boldsymbol{\Sigma}_A\boldsymbol{W}^{\mathrm{H}} + c_2\boldsymbol{W}\boldsymbol{\Sigma}_B\boldsymbol{W}^{\mathrm{H}}$$

由于c_1和c_2是任意的，因此必有

$$\begin{aligned}\boldsymbol{U}\boldsymbol{\Sigma}_A\boldsymbol{U}^{\mathrm{H}} &= \boldsymbol{W}\boldsymbol{\Sigma}_A\boldsymbol{W}^{\mathrm{H}} \\ \boldsymbol{V}\boldsymbol{\Sigma}_B\boldsymbol{V}^{\mathrm{H}} &= \boldsymbol{W}\boldsymbol{\Sigma}_B\boldsymbol{W}^{\mathrm{H}}\end{aligned}$$

故有$\boldsymbol{U} = \boldsymbol{V} = \boldsymbol{W}$，即$\boldsymbol{A}$和$\boldsymbol{B}$具有相同的对角化酉阵，易知此时$\boldsymbol{A}$和$\boldsymbol{B}$是可乘积交换的。

7.19 证明：若$\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$正定，则$|\boldsymbol{A}\lambda^2 + \boldsymbol{B}\lambda + \boldsymbol{C}| = 0$的根具有负的实部。

证明：令λ使得$|\boldsymbol{A}\lambda^2 + \boldsymbol{B}\lambda + \boldsymbol{C}| = 0$，则存在非零向量$\boldsymbol{p}$使得

$$(\boldsymbol{A}\lambda^2 + \boldsymbol{B}\lambda + \boldsymbol{C})\boldsymbol{p} = \boldsymbol{0}$$

将上式左乘$\boldsymbol{p}^{\mathrm{T}}$(复矩阵时左乘$\boldsymbol{p}^{\mathrm{H}}$)得

$$\boldsymbol{p}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}\lambda^2+\boldsymbol{p}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{p}\lambda+\boldsymbol{p}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{p}=0$$

由于$\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}$正定，故此关于$\lambda$的二次方程的系数都是正数，由根与系数的关系知，$\lambda$具有负的实部。

7.20 令$\boldsymbol{P},\boldsymbol{Q},\boldsymbol{R},\boldsymbol{X}$ 为2×2矩阵。证明：矩阵方程$\boldsymbol{P}\boldsymbol{X}^2+\boldsymbol{Q}\boldsymbol{X}+\boldsymbol{R}=\boldsymbol{O}$ (零矩阵)的解$\boldsymbol{X}$的每一个特征值都是$|\boldsymbol{P}\lambda^2+\boldsymbol{Q}\lambda+\boldsymbol{R}|=0$的根。

证明：令$\{\lambda,\boldsymbol{p}\}$是矩阵$\boldsymbol{X}$的一个特征对，即

$$\boldsymbol{X}\boldsymbol{p}=\lambda\boldsymbol{p}$$

由于$\boldsymbol{P}\boldsymbol{X}^2+\boldsymbol{Q}\boldsymbol{X}+\boldsymbol{R}=\boldsymbol{O}$，故有

$$(\boldsymbol{P}\boldsymbol{X}^2+\boldsymbol{Q}\boldsymbol{X}+\boldsymbol{R})\boldsymbol{p}=\boldsymbol{0}$$

即

$$(\boldsymbol{P}\lambda^2+\boldsymbol{Q}\lambda+\boldsymbol{R})\boldsymbol{p}=\boldsymbol{0}$$

由于$\boldsymbol{p}$为非零的特征向量，故$|\boldsymbol{P}\lambda^2+\boldsymbol{Q}\lambda+\boldsymbol{R}|=0$，即$\lambda$是$|\boldsymbol{P}\lambda^2+\boldsymbol{Q}\lambda+\boldsymbol{R}|=0$的根。

7.21 令

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

若定义矩阵束$(\boldsymbol{A},\boldsymbol{B})$的广义特征值$(\alpha,\beta)$是满足$\det(\beta\boldsymbol{A}-\alpha\boldsymbol{B})=0$的数值$\alpha$和$\beta$，试求$\alpha$和$\beta$。

解：容易计算得$\det(\beta\boldsymbol{A}-\alpha\boldsymbol{B})=-(\alpha^2+\beta^2)=0$，故$\alpha=\mathrm{j}\beta$，这里$\mathrm{j}=\sqrt{-1}$。

7.22 [39, p.286] 令矩阵束$(\boldsymbol{A},\boldsymbol{B})$的广义特征值$(\alpha,\beta)$如上题所定义。令$\lambda_i=(\alpha_i,\beta_i)$和$\lambda_j=(\alpha_j,\beta_j)$是矩阵束$(\boldsymbol{A},\boldsymbol{B})$的两个不同广义特征值，并且$\boldsymbol{u}_i$是与广义特征值$\lambda_i$对应的右广义特征向量，而$\boldsymbol{w}_j$是与$\lambda_j$对应的左广义特征向量，证明

$$\langle\boldsymbol{A}\boldsymbol{u}_i,\boldsymbol{w}_j\rangle=\langle\boldsymbol{B}\boldsymbol{u}_i,\boldsymbol{w}_j\rangle=0$$

证明：由广义特征值和特征向量的定义，容易写出等式

$$\begin{aligned}\beta_i\boldsymbol{A}\boldsymbol{u}_i&=\alpha_i\boldsymbol{B}\boldsymbol{u}_i\\ \beta_j\boldsymbol{w}_j^{\mathrm{T}}\boldsymbol{A}&=\alpha_j\boldsymbol{w}_j^{\mathrm{T}}\boldsymbol{B}\end{aligned}$$

将上两式分别左乘$\boldsymbol{w}_j^{\mathrm{T}}$和右乘$\boldsymbol{u}_i$可得到方程组

$$\begin{bmatrix}\beta_i&-\alpha_i\\\beta_j&-\alpha_j\end{bmatrix}\begin{bmatrix}\langle\boldsymbol{A}\boldsymbol{u}_i,\boldsymbol{w}_j\rangle\\\langle\boldsymbol{B}\boldsymbol{u}_i,\boldsymbol{w}_j\rangle\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$$

这一方程组有非零解的充分必要条件是$\alpha_i\beta_j-\alpha_j\beta_i=0$，即$\lambda_i=(\alpha_i,\beta_i)$和$\lambda_j=(\alpha_j,\beta_j)$是相同的广义特征值。故当$\lambda_i=(\alpha_i,\beta_i)$和$\lambda_j=(\alpha_j,\beta_j)$为不相同的特征值时，必有$\langle \boldsymbol{Au}_i,\boldsymbol{w}_j\rangle=\langle \boldsymbol{Bu}_i,\boldsymbol{w}_j\rangle=0$。

7.23 令矩阵$\boldsymbol{G}$是$\boldsymbol{A}$的广义逆矩阵，并且$\boldsymbol{A}$和$\boldsymbol{GA}$都是对称矩阵。证明$\boldsymbol{A}$的非零特征值的倒数是广义逆矩阵$\boldsymbol{G}$的一个特征值。

证明：由于$\boldsymbol{A}$为对称矩阵，故$\boldsymbol{A}$存在分解形式$\boldsymbol{A}=\boldsymbol{UDU}^{\mathrm{T}}$，这里$\boldsymbol{U}$为正交矩阵，$\boldsymbol{D}$为对角矩阵，其对角线上的元素为矩阵$\boldsymbol{A}$的特征值。由此得到$\boldsymbol{A}$的伪逆为$\boldsymbol{G}=\boldsymbol{UD}^{\dagger}\boldsymbol{U}^{\mathrm{T}}$，由于对角阵$\boldsymbol{D}$的伪逆仍为对角阵，且其对角线上的非零元素为原对角阵非零元素的倒数，故$\boldsymbol{A}$的非零特征值的倒数是其广义逆矩阵$\boldsymbol{G}$的特征值。

注释：通常情况下，$(\boldsymbol{AB})^{\dagger}\neq \boldsymbol{B}^{\dagger}\boldsymbol{A}^{\dagger}$。但当$\boldsymbol{A}$或$\boldsymbol{B}$中有一个为正交矩阵时，便可得到结论$(\boldsymbol{AB})^{\dagger}=\boldsymbol{B}^{\dagger}\boldsymbol{A}^{\dagger}$的。本题的证明用到这一结论，这个结论可以很容易地由教材P87的定理1.9.2来验证。

7.24 [31, p.226] 令$\boldsymbol{A}$是一个$n\times n$复矩阵，其特征值为$\lambda_1,\lambda_2,\cdots,\lambda_n$。证明$\boldsymbol{A}$为正规矩阵，当且仅当下列条件之一成立：

(1) $\boldsymbol{AA}^{\mathrm{H}}$的特征值为$|\lambda_1|^2,|\lambda_2|^2,\cdots,|\lambda_n|^2$。

(2) $\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}}$的特征值为$\lambda_1+\lambda_1^*,\lambda_2+\lambda_2^*,\cdots,\lambda_n+\lambda_n^*$。

证明：条件(1)和(2)的必要性很容易验证。事实上，若$\boldsymbol{A}$为正规矩阵，则存在酉矩阵$\boldsymbol{U}$使得$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{AU}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$。这时，可直接计算出$\boldsymbol{AA}^{\mathrm{H}}$的特征值为

$$|\lambda_1|^2,|\lambda_2|^2,\cdots,|\lambda_n|^2$$

$\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}}$的特征值为

$$\lambda_1+\lambda_1^*,\lambda_2+\lambda_2^*,\cdots,\lambda_n+\lambda_n^*$$

下面我们证明条件(1)和(2)的充分性。

(1) 设$\boldsymbol{A}$的Schur分解为$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{AU}=\boldsymbol{D}+\boldsymbol{N}$，其中$\boldsymbol{U}$为酉矩阵，$\boldsymbol{N}$为严格上三角阵，$\boldsymbol{D}$为对角阵，其对角线上的元素为$\boldsymbol{A}$的特征值，即$\boldsymbol{D}=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$。

容易计算得$\boldsymbol{AA}^{\mathrm{H}}=\boldsymbol{U}(\boldsymbol{D}+\boldsymbol{N})(\boldsymbol{D}+\boldsymbol{N})^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}}$，由于酉变换不改变矩阵的Frobenius范数，故可求得矩阵$\boldsymbol{AA}^{\mathrm{H}}$的范数为

$$\|\boldsymbol{AA}^{\mathrm{H}}\|_{\mathrm{F}}^2=\|(\boldsymbol{D}+\boldsymbol{N})(\boldsymbol{D}+\boldsymbol{N})^{\mathrm{H}}\|_{\mathrm{F}}^2$$

考虑到$\boldsymbol{D}$为对角阵，$\boldsymbol{N}$为严格上三角阵，我们可得到不等式

$$\begin{aligned}\|\boldsymbol{AA}^{\mathrm{H}}\|_{\mathrm{F}}^2 &= \|(\boldsymbol{D}+\boldsymbol{N})(\boldsymbol{D}+\boldsymbol{N})^{\mathrm{H}}\|_{\mathrm{F}}^2\\ &\geqslant \|\mathrm{diag}[(\boldsymbol{D}+\boldsymbol{N})(\boldsymbol{D}+\boldsymbol{N})^{\mathrm{H}}]\|_{\mathrm{F}}^2\\ &= \|\boldsymbol{DD}^{\mathrm{H}}+\mathrm{diag}(\boldsymbol{NN}^{\mathrm{H}})\|_{\mathrm{F}}^2\end{aligned}$$

另一方面，容易验证$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$为Hermitian矩阵，其范数的平方为其特征值的模的平方和，即

$$\|\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\|_{\mathrm{F}}^2=\sum_{i=1}^{n}|\lambda_i|^4=\|\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}}\|_{\mathrm{F}}^2$$

对比上两式的结论得

$$\|\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}}\|_{\mathrm{F}}^2\geqslant\|\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}}+\mathrm{diag}(\boldsymbol{N}\boldsymbol{N}^{\mathrm{H}})\|_{\mathrm{F}}^2$$

由于矩阵$\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}}$和$\mathrm{diag}(\boldsymbol{N}\boldsymbol{N}^{\mathrm{H}})$的对角线上的元素都是非负的，故该不等式取等号的条件是$\|\mathrm{diag}(\boldsymbol{N}\boldsymbol{N}^{\mathrm{H}})\|_{\mathrm{F}}^2=0$，即$\boldsymbol{N}=\boldsymbol{O}$。故存在酉阵$\boldsymbol{U}$使得$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}$为对角阵，此时容易验证$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$，因此，$\boldsymbol{A}$为正规矩阵。

(2) 同(1)，设$\boldsymbol{A}$有Schur分解$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\boldsymbol{D}+\boldsymbol{N}$。一方面，容易计算$\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}}$的范数为

$$\begin{aligned}\|\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}}\|_{\mathrm{F}}^2&=\|\boldsymbol{D}+\boldsymbol{D}^{\mathrm{H}}+\boldsymbol{N}+\boldsymbol{N}^{\mathrm{H}}\|_{\mathrm{F}}^2\\&=\|\boldsymbol{D}+\boldsymbol{D}^{\mathrm{H}}\|_{\mathrm{F}}^2+\|\boldsymbol{N}+\boldsymbol{N}^{\mathrm{H}}\|_{\mathrm{F}}^2\end{aligned}$$

另一方面，由于$\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}}$为Hermitian矩阵，其范数的平方为其所有特征值的模的平方和，即

$$\|\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}}\|_{\mathrm{F}}^2=\sum_{i=1}^{n}|\lambda_i+\lambda_i^*|^2=\|\boldsymbol{D}+\boldsymbol{D}^{\mathrm{H}}\|_{\mathrm{F}}^2$$

对比上两式的结论知$\|\boldsymbol{N}+\boldsymbol{N}^{\mathrm{H}}\|_{\mathrm{F}}^2=0$，由于$\boldsymbol{N}$为严格上三角阵，因此$\boldsymbol{N}=\boldsymbol{O}$。故存在酉阵$\boldsymbol{U}$使得$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}$为对角阵，此时容易验证$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$，因此，$\boldsymbol{A}$为正规矩阵。

7.25 利用特征方程证明：若λ是矩阵$\boldsymbol{A}$的实特征值，则$\boldsymbol{A}+\boldsymbol{A}^{-1}$的特征值的绝对值等于或大于2。

证明： 设λ是矩阵$\boldsymbol{A}$的特征值，则$\lambda+\lambda^{-1}$是矩阵$\boldsymbol{A}+\boldsymbol{A}^{-1}$的特征值，故

$$|\lambda+\lambda^{-1}|\geqslant2\sqrt{|\lambda\lambda^{-1}|}=2$$

7.26 令实矩阵$\boldsymbol{A}_{4\times4}$满足$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=2\boldsymbol{I}$，且$|\boldsymbol{A}|<0$，求伴随矩阵$\mathrm{adj}(\boldsymbol{A})$的一个特征值。

解： 由$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=2\boldsymbol{I}$得$\dfrac{\sqrt{2}}{2}\boldsymbol{A}$为正交矩阵，考虑到$|\boldsymbol{A}|<0$，故$\dfrac{\sqrt{2}}{2}\boldsymbol{A}$有特征值$-1$，即$\boldsymbol{A}$有特征值$-\sqrt{2}$。由$|\dfrac{\sqrt{2}}{2}\boldsymbol{A}|=-1$得$|\boldsymbol{A}|=-4$，由这些结论，易知伴随矩阵$\mathrm{adj}(\boldsymbol{A})=\det(\boldsymbol{A})\boldsymbol{A}^{-1}$的一个特征值为$-4\times(-\sqrt{2})^{-1}=2\sqrt{2}$。

7.27 设$\boldsymbol{A}_{4\times4}$满足条件$|3\boldsymbol{I}_4+\boldsymbol{A}|=0$, $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=2\boldsymbol{I}_4$和$|\boldsymbol{A}|<0$。求矩阵$\boldsymbol{A}$的伴随矩阵$\mathrm{adj}(\boldsymbol{A})=\det(\boldsymbol{A})\boldsymbol{A}^{-1}$的一个特征值。

解： 由$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=2\boldsymbol{I}_4$知$|\boldsymbol{A}|^2=16$，又$|\boldsymbol{A}|<0$，故$|\boldsymbol{A}|=-4$。由$|3\boldsymbol{I}_4+\boldsymbol{A}|=0$知$-3$为$\boldsymbol{A}$的一个特征值，故$-\dfrac{1}{3}$为$\boldsymbol{A}^{-1}$的一个特征值，因此伴随矩阵$\mathrm{adj}(\boldsymbol{A})=\det(\boldsymbol{A})\boldsymbol{A}^{-1}$的一个特征值为$(-4)\times\left(-\dfrac{1}{3}\right)=\dfrac{4}{3}$。

注释：与题7.26不同，本题中的矩阵$\frac{\sqrt{2}}{2}\boldsymbol{A}$不可能为正交矩阵。事实上，由于$-3$为$\boldsymbol{A}$的一个特征值，故$-\frac{3\sqrt{2}}{2}$为$\frac{\sqrt{2}}{2}\boldsymbol{A}$的一个特征值，因此，$\frac{\sqrt{2}}{2}\boldsymbol{A}$不可能为正交矩阵，而只可能为复的矩阵。

7.28 证明二次型$f = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ 在$\|\boldsymbol{x}\| = 1$ 时的最大值等于对称矩阵$\boldsymbol{A}$的最大特征值。(提示：将f化为标准二次型。)

证明：将$\boldsymbol{A}$分解为$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{T}}$，这里$\boldsymbol{U}$为正交矩阵，$\boldsymbol{D} = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$为对角阵。令$\boldsymbol{y} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}$，则可将$f$写成

$$f = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{y} = \sum_{i=1}^{n} \lambda_i y_i^2$$

由于有约束条件$\|\boldsymbol{x}\| = 1$，而正交变换不改变向量的长度，因此$\|\boldsymbol{y}\| = 1$。此时，f的最大值等于对称矩阵$\boldsymbol{A}$的最大特征值，当且仅当最大特征值λ_j对应的权重$y_j^2 = 1$时取到。

注释：将目标函数写成Rayleigh商的形式

$$f = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}$$

即可由Rayleigh商的性质得到需证的结论。

7.29 证明：若λ 是矩阵$\boldsymbol{AB}$ 的一个非零特征值，则它也是矩阵$\boldsymbol{BA}$ 的非零特征值($\boldsymbol{A}, \boldsymbol{B}$不一定为正方矩阵，但$\boldsymbol{AB}$ 和$\boldsymbol{BA}$ 分别是正方的)。

证明：设$\boldsymbol{A}$是$m \times n$维矩阵，$\boldsymbol{B}$是$n \times m$维矩阵，则容易验证矩阵恒等式

$$\begin{bmatrix} \boldsymbol{I}_m & -\boldsymbol{A} \\ \boldsymbol{0} & \boldsymbol{I}_n \end{bmatrix}\begin{bmatrix} \boldsymbol{AB} & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{BA} \end{bmatrix}\begin{bmatrix} \boldsymbol{I}_m & -\boldsymbol{A} \\ \boldsymbol{0} & \boldsymbol{I}_n \end{bmatrix}$$

由于矩阵$\begin{bmatrix} \boldsymbol{I}_m & -\boldsymbol{A} \\ \boldsymbol{0} & \boldsymbol{I}_n \end{bmatrix}$可逆，故矩阵$\begin{bmatrix} \boldsymbol{AB} & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{0} \end{bmatrix}$和$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{BA} \end{bmatrix}$相似，其特征多项式相等。分别计算其特征多项式，令其相等，得

$$\det\left(\begin{bmatrix} \lambda\boldsymbol{I}_m - \boldsymbol{AB} & \boldsymbol{0} \\ -\boldsymbol{B} & \lambda\boldsymbol{I}_n \end{bmatrix}\right) = \det\left(\begin{bmatrix} \lambda\boldsymbol{I}_m & \boldsymbol{0} \\ -\boldsymbol{B} & \lambda\boldsymbol{I}_n - \boldsymbol{BA} \end{bmatrix}\right)$$

即

$$\lambda^n \det(\lambda\boldsymbol{I}_m - \boldsymbol{AB}) = \lambda^m \det(\lambda\boldsymbol{I}_n - \boldsymbol{BA}) \tag{7.54}$$

故矩阵$\boldsymbol{AB}$和$\boldsymbol{BA}$有相同的非零特征值。

注释：令式(7.54)中的$\lambda = 1$，则得到矩阵恒等式

$$\det(\boldsymbol{I}_m - \boldsymbol{AB}) = \det(\boldsymbol{I}_n - \boldsymbol{BA})$$

这是一个有用的恒等式。特别地，令$\boldsymbol{A} = \boldsymbol{a}$，$\boldsymbol{B} = \boldsymbol{b}^{\mathrm{H}}$，有

$$\det(\boldsymbol{I} - \boldsymbol{a}\boldsymbol{b}^{\mathrm{H}}) = 1 - \boldsymbol{b}^{\mathrm{H}}\boldsymbol{a}$$

7.30 令$A_{n\times n}$为对称矩阵，其特征值为$\lambda_i, i=1,2,\cdots,n$。证明

$$\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^2=\sum_{k=1}^{n}\lambda_k^2$$

证法一：由于$A_{n\times n}$为对称矩阵，令其有分解形式

$$A_{n\times n}=UDU^{\mathrm{T}}$$

其中U为正交矩阵，D为对角矩阵，其对角线上的元素为矩阵A的特征值。由正交变换不改变矩阵的Frobenius范数的结论知$\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^2=\sum_{k=1}^{n}\lambda_k^2$。

证法二：我们也可以直接计算矩阵A的Frobenius范数如下

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^2 &= \mathrm{tr}(AA^{\mathrm{T}})\\ &= \mathrm{tr}(UD^2U^{\mathrm{T}})\\ &= \mathrm{tr}(D^2)\\ &= \sum_{k=1}^{n}\lambda_k^2\end{aligned}$$

此即我们欲证的结论。

7.31 设$A_{n\times n}$的全部特征值为$\lambda_1,\lambda_2,\cdots,\lambda_n$，且与$\lambda_i$ 对应的特征向量为u_i。试求

(1) $P^{-1}AP$的特征值与相对应的特征向量。

(2) $(P^{-1}AP)^{\mathrm{T}}$ 的特征值与相对应的特征向量。

解：(1) 设矩阵$D=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$，矩阵$U=[u_1,\cdots,u_n]$，则可将矩阵$A$写成$A=UDU^{-1}$。易知$P^{-1}AP=P^{-1}UD(P^{-1}U)^{-1}$，故矩阵$P^{-1}AP$的特征值仍为$\lambda_i, 1\leqslant i\leqslant n$，其第$i$个特征值对应的特征向量为矩阵$P^{-1}U$的第$i$个列向量。

(2) 通过同样的计算可得到矩阵$(P^{-1}AP)^{\mathrm{T}}$的特征值仍为$\lambda_i, 1\leqslant i\leqslant n$，第$i$个特征值对应的特征向量为矩阵$P^{\mathrm{T}}U^{-\mathrm{T}}$的第$i$个列向量。

7.32 设$n\times n$矩阵$A=\{a_{ij}\}$，其中

$$a_{ij}=\begin{cases}a, & i=j\\ b, & i\neq j\end{cases}$$

求A的特征值及特征向量。

解：将矩阵A写成

$$A=(a-b)I+b\mathbf{1}$$

这里$\boldsymbol{I}$为单位阵，$\mathbf{1}$为元素全为1的矩阵。由题7.2的结论知，矩阵$\boldsymbol{A}$的n个特征值为

$$\{(a-b)+\frac{b}{n}, a-b, \cdots, a-b\}$$

其中与特征值$(a-b)+\frac{b}{n}$对应的特征向量为

$$\boldsymbol{p}_1=[1,\cdots,1]^{\mathrm{T}}$$

与$\boldsymbol{p}_1$正交的向量为特征值$a-b$对应的特征向量。

7.33 令$p(z)=a_0+a_1z+\cdots+a_nz^n$ 为一多项式。证明：矩阵$\boldsymbol{A}$ 的特征向量一定是矩阵多项式$p(\boldsymbol{A})$的特征向量，但$p(\boldsymbol{A})$的特征向量不一定是$\boldsymbol{A}$ 的特征向量。

证明：设$\boldsymbol{p}$是矩阵$\boldsymbol{A}$的特征值λ对应的特征向量，容易验证$\boldsymbol{p}$是矩阵多项式$p(\boldsymbol{A})$的特征值$p(\lambda)$对应的特征向量。

通常，反过来的结论并不成立。我们举一个反例。令矩阵$\boldsymbol{A}$为

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

多项式$p(z)=z^2$。容易算得矩阵$p(\boldsymbol{A})$为

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

故任何一个2×1的向量都是$p(\boldsymbol{A})$的特征向量，但显然不一定为矩阵$\boldsymbol{A}$的特征向量。

7.34 令$\boldsymbol{A}$ 为实斜对称矩阵，即其元素$a_{ij}=-a_{ji}$。证明：

(1) $\boldsymbol{A}$的特征值为纯虚数或零。

(2) 若$\boldsymbol{u}+\mathrm{j}\,\boldsymbol{v}$是与特征值$\mathrm{j}\,\mu$ (其中，μ是非零的实数)对应的特征向量，并且$\boldsymbol{u}$和$\boldsymbol{v}$为实向量，则$\boldsymbol{u}$与$\boldsymbol{v}$正交。

证明：(1) 令$(\lambda,\boldsymbol{p})$为实斜对称矩阵$\boldsymbol{A}$的一个特征对，则有

$$\boldsymbol{A}\boldsymbol{p}=\lambda\boldsymbol{p} \tag{7.55}$$

对上式两边取复共轭转置得

$$\boldsymbol{p}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}=\lambda^*\boldsymbol{p}^{\mathrm{H}}$$

考虑到$\boldsymbol{A}$为实斜对称矩阵，则$\boldsymbol{A}^{\mathrm{H}}=-\boldsymbol{A}$，因此，上式化为

$$-\boldsymbol{p}^{\mathrm{H}}\boldsymbol{A}=\lambda^*\boldsymbol{p}^{\mathrm{H}} \tag{7.56}$$

对式(7.55)两边同时左乘$\boldsymbol{p}^{\mathrm{H}}$得

$$\boldsymbol{p}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{p}=\lambda\|\boldsymbol{p}\|^2 \tag{7.57}$$

对式(7.56)两边同时右乘$\boldsymbol{p}$得

$$-\boldsymbol{p}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{p}=\lambda^*\|\boldsymbol{p}\|^2 \tag{7.58}$$

等式(7.57)和式(7.58)两边同时相加得

$$(\lambda+\lambda^*)\|\boldsymbol{p}\|^2=0$$

由于$\boldsymbol{p}$为非零的特征向量，故有$\lambda+\lambda^*=0$，即$\boldsymbol{A}$的特征值为纯虚数或零。

(2) 由题设知

$$\boldsymbol{A}(\boldsymbol{u}+\mathrm{j}\boldsymbol{v})=\mathrm{j}\mu(\boldsymbol{u}+\mathrm{j}\boldsymbol{v})$$

展开上式，并比较等式两边的实部得

$$\boldsymbol{A}\boldsymbol{u}=-\mu\boldsymbol{v} \tag{7.59}$$

对式(7.59)两边左乘$\boldsymbol{u}^{\mathrm{T}}$得

$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u}=-\mu\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} \tag{7.60}$$

再对式(7.60)两边取转置，并代入$\boldsymbol{A}^{\mathrm{T}}=-\boldsymbol{A}$得

$$-\boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u}=-\mu\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u} \tag{7.61}$$

将等式(7.60)和式(7.61)两边同时相加，并考虑到$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}=\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}$，可得

$$-2\mu\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}=0$$

因此，若$\mu\neq 0$，则$\boldsymbol{u}$与$\boldsymbol{v}$正交。

7.35 令$\boldsymbol{A}$是一个正交矩阵，λ是$\boldsymbol{A}$的一个不等于±1，但其模为1的特征值，并且$\boldsymbol{u}+\mathrm{j}\,\boldsymbol{v}$是与该特征值对应的特征向量，其中，$\boldsymbol{u}$和$\boldsymbol{v}$为实向量。证明$\boldsymbol{u}$和$\boldsymbol{v}$正交。

证明：设矩阵$\boldsymbol{A}$的一个特征对为$(\lambda,\boldsymbol{u}+\mathrm{j}\boldsymbol{v})$，即

$$\boldsymbol{A}(\boldsymbol{u}+\mathrm{j}\boldsymbol{v})=\lambda(\boldsymbol{u}+\mathrm{j}\boldsymbol{v}) \tag{7.62}$$

这里$\boldsymbol{u}$和$\boldsymbol{v}$为实的向量。将上式两边转置得

$$(\boldsymbol{u}^{\mathrm{T}}+\mathrm{j}\boldsymbol{v}^{\mathrm{T}})\boldsymbol{A}^{\mathrm{T}}=\lambda(\boldsymbol{u}^{\mathrm{T}}+\mathrm{j}\boldsymbol{v}^{\mathrm{T}}) \tag{7.63}$$

将式(7.64)的两边分别同时左乘式(7.62)的两边，并考虑到$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{I}$得

$$(\lambda^2-1)(\boldsymbol{u}^{\mathrm{T}}+\mathrm{j}\boldsymbol{v}^{\mathrm{T}})(\boldsymbol{u}+\mathrm{j}\boldsymbol{v})=0 \tag{7.64}$$

若$\lambda\neq\pm1$，则必有

$$(\boldsymbol{u}^{\mathrm{T}}+\mathrm{j}\boldsymbol{v}^{\mathrm{T}})(\boldsymbol{u}+\mathrm{j}\boldsymbol{v})=0$$

展开上式，并令等式左边的实部和虚部分别等于零得

$$\|\boldsymbol{u}\| = \|\boldsymbol{v}\|, \quad \boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} = 0$$

即$\boldsymbol{u}$和$\boldsymbol{v}$正交，并等长。

注释：实斜对称矩阵和正交矩阵有着紧密的联系。设$\boldsymbol{A}$为实斜对称矩阵（它们构成一个李代数），则有

$$\mathrm{e}^{\boldsymbol{A}}(\mathrm{e}^{\boldsymbol{A}})^{\mathrm{T}} = \mathrm{e}^{\boldsymbol{A}}\mathrm{e}^{-\boldsymbol{A}} = \boldsymbol{I}$$

即$\mathrm{e}^{\boldsymbol{A}}$为正交矩阵。反之，设$\boldsymbol{U}$为行列式值为1的正交矩阵（它们构成一个李群），则有

$$\boldsymbol{O} = \log(\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}) = \log \boldsymbol{U} + (\log \boldsymbol{U})^{\mathrm{T}}$$

即$\log \boldsymbol{U}$为实斜对称矩阵。这种一一对应关系表现在特征值和特征向量上为:

(1) 若$(\lambda, \boldsymbol{p})$是实斜对称矩阵$\boldsymbol{A}$的一个特征对，则$(\mathrm{e}^{\lambda}, \boldsymbol{p})$是正交矩阵$\mathrm{e}^{\boldsymbol{A}}$的一个特征对。

(2) 若$(\lambda, \boldsymbol{p})$是正交矩阵$\boldsymbol{U}$的一个特征对，则$(\log \lambda, \boldsymbol{p})$是实斜对称矩阵$\log \boldsymbol{U}$的一个特征对。

这种映射关系将实斜对称矩阵的零或纯虚特征值映射为正交矩阵在单位圆上的特征值；反之，将正交矩阵在单位圆上的特征值映射为实斜对称矩阵的零或纯虚特征值，而特征向量保持不变。因此，题7.34第二问的命题和题7.35的命题是等价的。

7.36 一滤波器的抽头延迟线的输出由

$$y(k) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}(k)$$

给出，式中

$$\boldsymbol{a} = [a_0, a_1, \cdots, a_n]^{\mathrm{T}}$$
$$\boldsymbol{x}(k) = [x(k), x(k-1), \cdots, x(k-n)]^{\mathrm{T}}$$

令

$$\boldsymbol{R}_x \stackrel{\mathrm{def}}{=} \mathrm{E}\{\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\} = \boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q}^{\mathrm{T}}$$

式中

$$\boldsymbol{\Sigma} = \mathrm{diag}(\lambda_0, \lambda_1, \cdots, \lambda_n)$$

如果输出序列$\{y(k)\}$的均方值为

$$J_a = \frac{1}{2}\mathrm{E}\{y^2(k)\}$$

证明以下两个结论:

(1) 在条件$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a} = 1$ 的约束下，使J_a最小化等价于

$$J_w = \frac{1}{2}\sum_{i=0}^{n} w_i^2 \lambda_i$$

的最小化，其中

$$\boldsymbol{w}=[w_0,w_1,\cdots,w_n]^{\mathrm{T}},\quad \sum_{i=0}^{n}w_i^2=1$$

且$\boldsymbol{w}=\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{a}$。

(2) 若取$\boldsymbol{w}=[\pm1,0,\cdots,0]^{\mathrm{T}}$，则使$J_a$最小化的最优向量$\boldsymbol{a}=\pm\boldsymbol{a}_0$，其中，$\boldsymbol{a}_0$ 是矩阵$\boldsymbol{R}_x$相对于最小特征值λ_0 的特征向量。

证明：(1) 将J_a写成向量$\boldsymbol{a}$的二次型形式有

$$J_a=\frac{1}{2}\boldsymbol{a}^{\mathrm{T}}\mathrm{E}\{\boldsymbol{x}(k)\boldsymbol{x}^{\mathrm{T}}(k)\}\boldsymbol{a}=\frac{1}{2}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{R}_x\boldsymbol{a}$$

在上式中代入$\boldsymbol{w}=\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{a}$得

$$J_w=\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{R}_x\boldsymbol{Q}\boldsymbol{w}=\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{w}$$

即

$$J_w=\frac{1}{2}\sum_{i=0}^{n}w_i^2\lambda_i$$

故结论成立。

(2) 若使J_a取得最小值，也就是使得J_w取最小值，考虑到约束$\sum_{i=0}^{n}w_i^2=1$，J_w的最小值在$\boldsymbol{w}=[\pm1,0,\cdots,0]^{\mathrm{T}}$时取到。设$\boldsymbol{Q}=[\boldsymbol{a}_0,\boldsymbol{a}_1,\cdots,\boldsymbol{a}_n]$，则这时对应的向量$\boldsymbol{a}$为

$$\boldsymbol{a}=[\boldsymbol{a}_0,\boldsymbol{a}_1,\cdots,\boldsymbol{a}_n]\begin{bmatrix}\pm1\\0\\\vdots\\0\end{bmatrix}=\pm\boldsymbol{a}_0$$

即矩阵$\boldsymbol{R}_x$相对于最小特征值λ_0 的特征向量。

7.37 证明：一个$n\times n$ 实对称矩阵$\boldsymbol{A}$可以写作

$$\boldsymbol{A}=\sum_{i=1}^{n}\lambda_i\boldsymbol{Q}_i$$

式中，λ_i是$\boldsymbol{A}$的特征值；$\boldsymbol{Q}_i$为非负定矩阵，并且不仅满足正交条件

$$\boldsymbol{Q}_i\boldsymbol{Q}_j=\boldsymbol{O},\quad i\neq j$$

而且还是幂等矩阵，即$\boldsymbol{Q}_i^2=\boldsymbol{Q}_i$。矩阵$\boldsymbol{A}$的这一表示称为$\boldsymbol{A}$的谱分解[31, p.64]。

证明：由于$\boldsymbol{A}$为实对称矩阵，可设其有分解形式

$$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{T}}$$

其中$\boldsymbol{U}=[\boldsymbol{u}_1,\cdots,\boldsymbol{u}_n]$为正交矩阵，$\boldsymbol{D}=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$为对角矩阵，其对角上的元素为矩阵$\boldsymbol{A}$的特征值。令$\boldsymbol{Q}_i=\boldsymbol{u}_i\boldsymbol{u}_i^{\mathrm{T}}$，则可将矩阵$\boldsymbol{A}$写成

$$\boldsymbol{A}=\sum_{i=1}^{n}\lambda_i\boldsymbol{Q}_i$$

由于$\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{u}_j = 0, i \neq j$，易证$\boldsymbol{Q}_i\boldsymbol{Q}_j = \boldsymbol{O}, i \neq j$。$\boldsymbol{Q}_i$的非负定性也是显见的。故结论成立。

7.38 已知

$$\boldsymbol{A} = \begin{bmatrix} 3 & -1 & 0 \\ 0 & 5 & -2 \\ 0 & 0 & 9 \end{bmatrix}$$

求非奇异矩阵$\boldsymbol{S}$使得相似变换$\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S} = \boldsymbol{B}$为对角矩阵，并求对角矩阵$\boldsymbol{B}$。

解：参见题7.14的解法。这里

$$\boldsymbol{S} = \begin{bmatrix} 1 & -\dfrac{1}{\sqrt{5}} & \dfrac{1}{\sqrt{181}} \\ 0 & \dfrac{2}{\sqrt{5}} & -\dfrac{6}{\sqrt{181}} \\ 0 & 0 & \dfrac{12}{\sqrt{181}} \end{bmatrix}$$

$$\boldsymbol{B} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

7.39 已知矩阵

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

有三个线性无关的特征向量，求x和y应该满足的条件。

解：首先求得$\boldsymbol{A}$的特征多项式得

$$|\lambda \boldsymbol{I} - \boldsymbol{A}| = (\lambda - 1)^2(\lambda + 1)$$

即$\lambda = 1$是矩阵$\boldsymbol{A}$的重根，其代数重度为2，若$\boldsymbol{A}$有三个线性无关的特征向量，其几何重度也应为2，即矩阵

$$\boldsymbol{I} - \boldsymbol{A} = \begin{bmatrix} 1 & 0 & -1 \\ -x & 0 & -y \\ -1 & 0 & 1 \end{bmatrix}$$

的秩为1，易知其条件为$x + y = 0$。

7.40 设

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

试通过求矩阵$\boldsymbol{S}$使得$\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S} = \boldsymbol{D}$（对角矩阵），证明$\boldsymbol{A}$是可对角化的。

证明：可取矩阵$\boldsymbol{S}$为

$$\boldsymbol{S} = \begin{bmatrix} \dfrac{\sqrt{2}}{2} & \dfrac{\sqrt{2}}{2} \\ -\dfrac{\mathrm{j}\sqrt{2}}{2} & \dfrac{\mathrm{j}\sqrt{2}}{2} \end{bmatrix}$$

对角阵$\boldsymbol{D}$为

$$\boldsymbol{D}=\begin{bmatrix}1+\mathrm{j} & 0\\ 0 & 1-\mathrm{j}\end{bmatrix}$$

7.41 证明下列结论：

(1) 若$\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p$是矩阵$\boldsymbol{A}_{n\times n}$与同一特征值$\lambda$ 对应的特征向量，则$\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p$的任意一个线性组合也是$\boldsymbol{A}$的属于特征值$\lambda$的特征向量。

(2) 若向量$\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p$是矩阵$\boldsymbol{A}_{n\times n}$与不同特征值$\lambda_1,\lambda_2,\cdots,\lambda_p$对应的特征向量，则当$c_1,c_2,\cdots,c_p$中至少有两个不为零时，$c_1\boldsymbol{u}_1+c_2\boldsymbol{u}_2+\cdots+c_p\boldsymbol{u}_p$必定不是$\boldsymbol{A}$的特征向量。

证明： (1) 设$\boldsymbol{u}=\sum_{i=1}^{p}c_i\boldsymbol{u}_i$是向量$\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p$的一个线性组合，容易计算得

$$\boldsymbol{A}\boldsymbol{u}=\left(\lambda\sum_{i=1}^{p}c_i\right)\boldsymbol{u}$$

故$\boldsymbol{u}$仍是矩阵$\boldsymbol{A}$的一个特征向量，对应的特征值为$\lambda\sum_{i=1}^{p}c_i$。

(2) 首先，我们证明$\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p$是线性无关的。用反证法，设某个$\boldsymbol{u}_q$，$1\leqslant q\leqslant p$，可以表示成其他$\boldsymbol{u}_i$的线性组合，这里$1\leqslant i\leqslant p$且$i\neq q$，不妨记为

$$\boldsymbol{u}_q=\sum_{1\leqslant i\leqslant p,i\neq q}d_i\boldsymbol{u}_i$$

这里选择个数最少的一组$\boldsymbol{u}_i$，因此系数d_i是唯一确定的。在上式两边左乘矩阵$\boldsymbol{A}$得

$$\lambda_q\boldsymbol{u}_q=\sum_{1\leqslant i\leqslant p,i\neq q}\lambda_i d_i\boldsymbol{u}_i$$

由于$\boldsymbol{u}_q$分解的唯一性，上式只有在下述的两种情形下成立

(a) $\lambda_q=\lambda_i=0$，$1\leqslant i\leqslant p$且$i\neq q$；

(b) $\lambda_q\neq 0$但$\lambda_i=\lambda_q$，$1\leqslant i\leqslant p$且$i\neq q$。

这两种情况都导致λ_i，$1\leqslant i\leqslant q$中有相同的特征值，与题设矛盾。因此，特征向量$\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p$必线性无关。

现在，若设$c_1\boldsymbol{u}_1+c_2\boldsymbol{u}_2+\cdots+c_p\boldsymbol{u}_p$为矩阵$\boldsymbol{A}$的特征向量，不妨设对应的特征值为$\lambda$，则有

$$\boldsymbol{A}\sum_{i=1}^{p}c_i\boldsymbol{u}_i=\lambda\sum_{i=1}^{p}c_i\boldsymbol{u}_i$$

即

$$\sum_{i=1}^{p}c_i(\lambda_i-\lambda)\boldsymbol{u}_i=\boldsymbol{0}$$

此时，若$c_1,c_2,\cdots,c_p$中至少有两个不为零，上式要么导致特征向量$\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p$线性相关，要么导致$\lambda_i$，$1\leqslant i\leqslant p$，中有相同的特征值(与$\lambda$相同)，这都是与已知事实矛盾的，故结论成立。

7.42 设三阶实对称矩阵$\boldsymbol{A}$ 的特征值为$\lambda_1 = -1, \lambda_2 = 1, \lambda_3 = -2$，且与$\lambda_1$ 对应的特征向量为$\boldsymbol{u}_1 = [0,1,1]^{\mathrm{T}}$，求矩阵$\boldsymbol{A}$的表达式。

解：显然，矩阵$\boldsymbol{A}$的表达式不唯一。首先，将特征向量$\boldsymbol{u}_1$的长度归一化为1，即

$$\boldsymbol{u}_1 = \left[0, \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]^{\mathrm{T}}$$

再任取单位长度向量$\boldsymbol{u}_2$和$\boldsymbol{u}_3$，使得$\boldsymbol{u}_1$，$\boldsymbol{u}_2$和$\boldsymbol{u}_3$ 两两正交，那么矩阵

$$\boldsymbol{A} = -\boldsymbol{u}_1\boldsymbol{u}_1^{\mathrm{T}} + \boldsymbol{u}_2\boldsymbol{u}_2^{\mathrm{T}} - 2\boldsymbol{u}_3\boldsymbol{u}_3^{\mathrm{T}}$$

满足题设要求。

简单地取

$$\boldsymbol{u}_2 = [1,0,0]^{\mathrm{T}}$$

$$\boldsymbol{u}_3 = \left[0, \frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right]^{\mathrm{T}}$$

可以计算一个满足条件的矩阵$\boldsymbol{A}$为

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1.5 & 0.5 \\ 0 & 0.5 & -1.5 \end{bmatrix}$$

7.43 已知矩阵

$$\boldsymbol{A} = \begin{bmatrix} 1 & -1 & 1 \\ x & 4 & y \\ -3 & -3 & 5 \end{bmatrix}$$

有三个线性无关的特征向量，且$\lambda = 2$是$\boldsymbol{A}$的二重特征值。试求可逆矩阵$\boldsymbol{P}$，使得$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$为对角矩阵。

解：由于$\lambda = 2$是$\boldsymbol{A}$的二重特征值且$\boldsymbol{A}$有三个线性无关的特征向量，故$2\boldsymbol{I} - \boldsymbol{A}$的秩为1，即矩阵

$$2\boldsymbol{I} - \boldsymbol{A} = \begin{bmatrix} 1 & 1 & -1 \\ -x & -2 & -y \\ 3 & 3 & -3 \end{bmatrix}$$

的秩为1，观察知$x = 2, y = -2$。矩阵$\boldsymbol{A}$为

$$\boldsymbol{A} = \begin{bmatrix} 1 & -1 & 1 \\ 2 & 4 & -2 \\ -3 & -3 & 5 \end{bmatrix}$$

容易求得$\boldsymbol{A}$的另一个特征值为6，求得这些特征值对应特征向量，可得矩阵$\boldsymbol{P}$为

$$\boldsymbol{P} = \begin{bmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{1}{\sqrt{14}} \\ 0 & \frac{\sqrt{2}}{2} & -\frac{2}{\sqrt{14}} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & \frac{3}{\sqrt{14}} \end{bmatrix}$$

注意，由于特征根$\lambda=2$的几何重度为2，即使约定它的特征向量为单位长度的，这一对特征向量也是无法唯一确定的，这里的矩阵$\boldsymbol{P}$给出了一种可能的实现形式。

7.44 设向量$\boldsymbol{\alpha}=[\alpha_1,\alpha_2,\cdots,\alpha_n]^{\mathrm{T}}$和$\boldsymbol{\beta}=[\beta_1,\beta_2,\cdots,\beta_n]^{\mathrm{T}}$是两个正交的非零向量。若令$\boldsymbol{A}=\boldsymbol{\alpha}\boldsymbol{\beta}^{\mathrm{T}}$，试求

(1) $\boldsymbol{A}^2$;

(2) 矩阵$\boldsymbol{A}$的特征值和特征向量。

解： (1) 直接计算得$\boldsymbol{A}^2=\boldsymbol{O}$。

(2) 显然，任意一组与$\boldsymbol{\beta}$正交的$n-1$个线性无关的向量都是$\boldsymbol{A}$的特征向量，对应的特征值为0，除此之外，$\boldsymbol{A}$没有别的特征值了。事实上，$p(\lambda)=\lambda^2$便是$\boldsymbol{A}$的最小多项式了，由于$p(\lambda)$有重根0，故特征值0为矩阵$\boldsymbol{A}$的亏损特征值，其代数重度为n，几何重度为$n-1$。

注释： 由题7.29的结论知

$$\lambda|\lambda\boldsymbol{I}-\boldsymbol{\alpha}\boldsymbol{\beta}^{\mathrm{T}}|=\lambda^n|\lambda-\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\beta}|$$

即

$$|\lambda\boldsymbol{I}-\boldsymbol{\alpha}\boldsymbol{\beta}^{\mathrm{T}}|=\lambda^n$$

由此也可得到矩阵$\boldsymbol{A}$有n个0特征值。但$\boldsymbol{A}$只有$n-1$个线性独立的特征向量(由于$\boldsymbol{\alpha}\neq\boldsymbol{\beta}$, $\boldsymbol{\beta}$不可能为$\boldsymbol{A}$的特征向量，而所有与$\boldsymbol{\beta}$ 正交的向量都是矩阵$\boldsymbol{A}$的特征向量，故这样的线性独立的向量共有$n-1$个)，因此$\boldsymbol{A}$不可对角化。

7.45 已知矩阵

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

且$\boldsymbol{B}=(k\boldsymbol{I}+\boldsymbol{A})^2$，其中，$k$为实数。

(1) 求矩阵$\boldsymbol{B}$ 的对角化矩阵$\boldsymbol{\Lambda}$。

(2) 试问：k为何值时，矩阵$\boldsymbol{B}$是正定的？

解： (1) 注意到$\boldsymbol{A}$已为对称矩阵，计算得其特征值为$0,2,2$，故矩阵$\boldsymbol{B}$的对角化矩阵$\boldsymbol{\Lambda}$为

$$\boldsymbol{\Lambda}=\begin{bmatrix}k^2 & 0 & 0\\ 0 & (k+2)^2 & 0\\ 0 & 0 & (k+2)^2\end{bmatrix}$$

(2) 当$k\neq0$且$k\neq-2$时$\boldsymbol{B}$正定。

7.46 设

$$|\boldsymbol{A}|=\begin{vmatrix}a & -1 & c\\ 5 & b & 3\\ 1-c & 0 & -a\end{vmatrix}=-1$$

又$\boldsymbol{A}$的伴随矩阵$\mathrm{adj}(\boldsymbol{A})=|\boldsymbol{A}|\boldsymbol{A}^{-1}$有一个特征值为$\lambda$，与之对应的特征向量$\boldsymbol{u}=[-1,-1,1]^{\mathrm{T}}$。求$a,b,c$和$\lambda$的值。

解：特征向量$\boldsymbol{u}$也是矩阵$\boldsymbol{A}$的特征向量，即

$$\begin{bmatrix} a & -1 & c \\ 5 & b & 3 \\ 1-c & 0 & -a \end{bmatrix}\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}=k\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

从上式中容易得到$k=-1,a=c,b=-3$。

代入关系式$a=c,b=-3$，容易算得$|\boldsymbol{A}|=a-3$，由于$|\boldsymbol{A}|=-1$，故$a=c=2$。即矩阵$\boldsymbol{A}$为

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

$\boldsymbol{A}$的伴随矩阵$\mathrm{adj}(\boldsymbol{A})=|\boldsymbol{A}|\boldsymbol{A}^{-1}$的与特征向量$\boldsymbol{u}$对应的特征值为$-1\times\dfrac{1}{-1}=1$。

7.47 求矩阵

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

的正交基。

解：$\boldsymbol{A}$是一个满秩的方阵，它张成了整个$\boldsymbol{R}^4$空间，它的一组正交基为

$$\boldsymbol{Q}=\begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

注释：一般的，任意满秩的矩阵的正交基可以选择为单位矩阵。

7.48 求解广义特征值问题$\boldsymbol{Ax}=\lambda\boldsymbol{Bx}$时，矩阵$\boldsymbol{B}$必须是非奇异的。现在假定$\boldsymbol{B}$奇异，其广义逆矩阵为$\boldsymbol{B}^{\dagger}$。

(1) 令$(\lambda,\boldsymbol{x})$是矩阵$\boldsymbol{B}^{\dagger}\boldsymbol{A}$的一个特征对。证明：该特征对是矩阵束$(\boldsymbol{A},\boldsymbol{B})$ 的一个广义特征对，若$\boldsymbol{Ax}$是矩阵$\boldsymbol{BB}^{\dagger}$与特征值1对应的特征向量。

(2) 令$\lambda,\boldsymbol{x}$满足$\boldsymbol{Ax}=\lambda\boldsymbol{Bx}$。证明：若$\boldsymbol{x}$ 也是矩阵$\boldsymbol{B}^{\dagger}\boldsymbol{B}$与特征值1对应的特征向量，则$(\lambda,\boldsymbol{x})$是矩阵$\boldsymbol{B}^{\dagger}\boldsymbol{A}$的一个特征对。

证明：(1) 由题设知有等式

$$\boldsymbol{BB}^{\dagger}\boldsymbol{Ax}=\boldsymbol{Ax}$$

在上式右边代入条件$\boldsymbol{B}^{\dagger}\boldsymbol{Ax}=\lambda\boldsymbol{x}$即得

$$\lambda\boldsymbol{Bx}=\boldsymbol{Ax}$$

即$(\lambda, \boldsymbol{x})$是矩阵束$(\boldsymbol{A}, \boldsymbol{B})$ 的一个广义特征对。

(2) 由题设知条件$\boldsymbol{B}^{\dagger}\boldsymbol{B}\boldsymbol{x} = \boldsymbol{x}$，在等式$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{B}\boldsymbol{x}$两边同时左乘$\boldsymbol{B}^{\dagger}$即得

$$\boldsymbol{B}^{\dagger}\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$$

即$(\lambda, \boldsymbol{x})$是矩阵$\boldsymbol{B}^{\dagger}\boldsymbol{A}$的一个特征对。

7.49 令矩阵束$(\boldsymbol{A}, \boldsymbol{B})$与广义特征值$\lambda_1$对应的右特征向量为$\boldsymbol{u}_1$，左特征向量为$\boldsymbol{v}_1$，并且$\langle \boldsymbol{B}\boldsymbol{u}_1, \boldsymbol{B}\boldsymbol{v}_1\rangle = 1$。试证明：矩阵束$(\boldsymbol{A}, \boldsymbol{B})$和

$$\boldsymbol{A}_1 = \boldsymbol{A} - \sigma_1\boldsymbol{B}\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}, \qquad \boldsymbol{B}_1 = \boldsymbol{B} - \sigma_2\boldsymbol{A}\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}$$

具有一对相同的左和右特征向量。式中，假定移位因子σ_1和σ_2满足条件$1 - \sigma_1\sigma_2 \neq 0$。

证明： 我们证明$\boldsymbol{u}_1$和$\boldsymbol{v}_1$是矩阵束$(\boldsymbol{A}, \boldsymbol{B})$和$(\boldsymbol{A}_1, \boldsymbol{B}_1)$共同的右和左特征向量。先证明$\boldsymbol{u}_1$是矩阵束$(\boldsymbol{A}_1, \boldsymbol{B}_1)$的右特征向量。由题设知$\boldsymbol{A}\boldsymbol{u}_1 = \lambda_1\boldsymbol{B}\boldsymbol{u}_1$，因此有

$$\begin{aligned}\boldsymbol{A}_1\boldsymbol{u}_1 &= (\boldsymbol{A} - \sigma_1\boldsymbol{B}\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}})\boldsymbol{u}_1 \\ &= \lambda_1\boldsymbol{B}\boldsymbol{u}_1 - \sigma_1\boldsymbol{B}\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{u}_1 \\ &= (\lambda_1 - \sigma_1\boldsymbol{v}_1^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{u}_1)\boldsymbol{B}\boldsymbol{u}_1 \end{aligned} \tag{7.65}$$

$\boldsymbol{B}_1\boldsymbol{u}_1$可表示为

$$\begin{aligned}\boldsymbol{B}_1\boldsymbol{u}_1 &= (\boldsymbol{B} - \sigma_2\boldsymbol{A}\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}})\boldsymbol{u}_1 \\ &= \boldsymbol{B}\boldsymbol{u}_1 - \sigma_2\lambda_1\boldsymbol{B}\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{u}_1 \\ &= (1 - \sigma_2\lambda_1\boldsymbol{v}_1^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{u}_1)\boldsymbol{B}\boldsymbol{u}_1 \end{aligned} \tag{7.66}$$

比较式(7.65)和式(7.66)知，$\boldsymbol{A}_1\boldsymbol{u}_1$和$\boldsymbol{B}_1\boldsymbol{u}_1$共线，因此，$\boldsymbol{u}_1$是矩阵束$(\boldsymbol{A}_1, \boldsymbol{B}_1)$的一个右特征向量。

类似的计算表明$\boldsymbol{v}_1$是矩阵束$(\boldsymbol{A}_1, \boldsymbol{B}_1)$的一个左特征向量。因此，矩阵束$(\boldsymbol{A}, \boldsymbol{B})$和$(\boldsymbol{A}_1, \boldsymbol{B}_1)$有一对相同的右和左特征向量。

7.50 考虑无线通信中的一个码分多址(CDMA)系统，它共有K个用户[37]。其中，用户1为期望用户。一接收机接收所有K个用户发射的信号，其接收信号的向量形式由下式给出：

$$\boldsymbol{y}(n) = \sum_{k=1}^{K}\boldsymbol{y}_k(n) = \boldsymbol{h}_1 w_1(n) + \boldsymbol{H}\boldsymbol{w}(n) + \boldsymbol{v}(n)$$

式中，$w_1(n)$是期望用户发射的比特信号，它是我们希望检测的；$\boldsymbol{h}_1$是期望用户的等价特征波形向量，它是已知的；而$\boldsymbol{H}$和$\boldsymbol{w}(n)$ 分别为所有其他用户(简称干扰用户)的特征波形向量组成的矩阵和干扰比特向量。假定信道的加性噪声为高斯白噪声，各个噪声分量的均值都等于零，方差均为σ^2。

(1) 设计一最小方差接收机$\boldsymbol{f}$，使得接收机输出

$$\hat{w}_1(n) = \boldsymbol{f}^{\mathrm{T}}\boldsymbol{y}(n)$$

能够在满足约束条件

$$\boldsymbol{f}^{\mathrm{H}}\boldsymbol{h}_1 = 1$$

的同时，与$w_1(n)$之间的均方误差为最小。求最小方差接收机$\boldsymbol{f}$ 的表达式。

(2) 若期望用户的等价特征波形向量$\boldsymbol{h}_1$为

$$\boldsymbol{h}_1 = \boldsymbol{C}_1\boldsymbol{g}_1$$

式中

$$\boldsymbol{C}_1 = \begin{bmatrix} c_1(0) & & 0 \\ \vdots & \ddots & \\ c_1(P-1) & \cdots & c_1(0) \\ & \ddots & \vdots \\ 0 & & c_1(P-1) \end{bmatrix}, \quad \boldsymbol{g}_1 = \begin{bmatrix} g_1(0) \\ \vdots \\ g_1(L) \end{bmatrix}$$

这里，$c_1(0), c_1(1), \cdots, c_1(P-1)$是期望用户的扩频码，而$g_1(l)$代表第$l$条传输路径的参数。试设计一最小方差无畸变(MVDR) 波束形成器$\boldsymbol{g}$，并证明它恰好是矩阵束$(\boldsymbol{C}_1^{\mathrm{H}}\boldsymbol{R}_y^{-1}\boldsymbol{C}_1, \boldsymbol{C}_1^{\mathrm{H}}\boldsymbol{C}_1)$与最小广义特征值对应的广义特征向量。

解：(1) 考虑到约束，我们有Lagrange函数

$$J(\boldsymbol{f}) = \boldsymbol{f}^{\mathrm{H}}\boldsymbol{R}_y\boldsymbol{f} - \lambda(\boldsymbol{f}^{\mathrm{H}}\boldsymbol{h}_1 - 1)$$

其中$\boldsymbol{R}_y$为接收信号向量的自相关矩阵，λ为Lagrange乘子。将上式对$\boldsymbol{f}^*$求导得

$$\boldsymbol{R}_y\boldsymbol{f} - \lambda\boldsymbol{h}_1 = \boldsymbol{0}$$

由此解得

$$\boldsymbol{f} = \lambda\boldsymbol{R}_y^{-1}\boldsymbol{h}_1$$

其中Lagrange乘子λ可由约束$\boldsymbol{f}^{\mathrm{H}}\boldsymbol{h}_1 = 1$解得，为

$$\lambda = \frac{1}{\boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{R}_y^{-1}\boldsymbol{h}_1}$$

故解得$\boldsymbol{f}$为

$$\boldsymbol{f} = \frac{\boldsymbol{R}_y^{-1}\boldsymbol{h}_1}{\boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{R}_y^{-1}\boldsymbol{h}_1}$$

此时的接收机输出能量为

$$J_{\min}(\boldsymbol{f}) = \boldsymbol{f}^{\mathrm{H}}\boldsymbol{R}_y\boldsymbol{f} = \frac{1}{\boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{R}_y^{-1}\boldsymbol{h}_1} \tag{7.67}$$

(2) 考虑到多径时，$\boldsymbol{h}_1=\boldsymbol{C}_1\boldsymbol{g}_1$。我们在约束$\|\boldsymbol{h}_1\|=1$下最大化式(7.67)中得到的最小输出能量，即

$$\min \quad \frac{\boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{R}_y^{-1}\boldsymbol{h}_1}{\boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{h}_1}$$

也就是

$$\min \quad \frac{\boldsymbol{g}_1^{\mathrm{H}}\boldsymbol{C}_1^{\mathrm{H}}\boldsymbol{R}_y^{-1}\boldsymbol{C}_1\boldsymbol{g}_1}{\boldsymbol{g}_1^{\mathrm{H}}\boldsymbol{C}_1^{\mathrm{H}}\boldsymbol{C}_1\boldsymbol{g}_1}$$

故最佳的波束形成器$\boldsymbol{g}$恰好是矩阵束$(\boldsymbol{C}_1^{\mathrm{H}}\boldsymbol{R}_y^{-1}\boldsymbol{C}_1,\boldsymbol{C}_1^{\mathrm{H}}\boldsymbol{C}_1)$与最小广义特征值对应的广义特征向量。

7.51 假定

$$\begin{bmatrix}\boldsymbol{O} & \boldsymbol{A}\\ \boldsymbol{A}^{\mathrm{T}} & \boldsymbol{O}\end{bmatrix}\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}\end{bmatrix}=\lambda\begin{bmatrix}\boldsymbol{B}_1 & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{B}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}\end{bmatrix}$$

式中，$\boldsymbol{A}\in R^{m\times n},\boldsymbol{B}_1\in R^{m\times m}$，$\boldsymbol{B}_2\in R^{n\times n}$。假定矩阵$\boldsymbol{B}_1$和$\boldsymbol{B}_2$正定，并且分别具有Cholesky三角因子$\boldsymbol{G}_1$和$\boldsymbol{G}_2$。将上式描述的广义特征值问题与$\boldsymbol{G}_1^{-1}\boldsymbol{A}\boldsymbol{G}_2^{-\mathrm{T}}$联系起来。

解：这里，特征值λ是以下特征多项式的根

$$p(\lambda)=\begin{vmatrix}\lambda\boldsymbol{B}_1 & -\boldsymbol{A}\\ -\boldsymbol{A}^{\mathrm{T}} & \lambda\boldsymbol{B}_2\end{vmatrix}$$

由于$\boldsymbol{B}_1$可逆，可将$p(\lambda)$展开为

$$p(\lambda)=|\lambda\boldsymbol{B}_1||\lambda\boldsymbol{B}_2-\boldsymbol{A}^{\mathrm{T}}(\lambda\boldsymbol{B}_1)^{-1}\boldsymbol{A}|$$

$p(\lambda)$可化简为

$$p(\lambda)=\lambda^{m-n}|\boldsymbol{B}_1||\boldsymbol{B}_2||\lambda^2\boldsymbol{I}-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}_1^{-1}\boldsymbol{A}\boldsymbol{B}_2^{-1}|$$

在上式中代入$\boldsymbol{B}_1=\boldsymbol{G}_1\boldsymbol{G}_1^{\mathrm{T}}$和$\boldsymbol{B}_2=\boldsymbol{G}_2\boldsymbol{G}_2^{\mathrm{T}}$得

$$p(\lambda)=\lambda^{m-n}|\boldsymbol{B}_1||\boldsymbol{B}_2||\lambda^2\boldsymbol{I}-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{G}_1^{-\mathrm{T}}\boldsymbol{G}_1^{-1}\boldsymbol{A}\boldsymbol{G}_2^{-\mathrm{T}}\boldsymbol{G}_2^{-1}|$$

由题7.29的结论知：

$$|\lambda^2\boldsymbol{I}-(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{G}_1^{-\mathrm{T}}\boldsymbol{G}_1^{-1}\boldsymbol{A}\boldsymbol{G}_2^{-\mathrm{T}})\boldsymbol{G}_2^{-1}|=|\lambda^2\boldsymbol{I}-\boldsymbol{G}_2^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{G}_1^{-\mathrm{T}}\boldsymbol{G}_1^{-1}\boldsymbol{A}\boldsymbol{G}_2^{-\mathrm{T}})|$$

故有

$$p(\lambda)=\lambda^{m-n}|\boldsymbol{B}_1||\boldsymbol{B}_2||\lambda^2\boldsymbol{I}-\boldsymbol{G}_2^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{G}_1^{-\mathrm{T}}\boldsymbol{G}_1^{-1}\boldsymbol{A}\boldsymbol{G}_2^{-\mathrm{T}}|$$

更明确的，有

$$p(\lambda)=\lambda^{m-n}|\boldsymbol{B}_1||\boldsymbol{B}_2|\left|\lambda^2\boldsymbol{I}-\left(\boldsymbol{G}_1^{-1}\boldsymbol{A}\boldsymbol{G}_2^{-\mathrm{T}}\right)^{\mathrm{T}}\boldsymbol{G}_1^{-1}\boldsymbol{A}\boldsymbol{G}_2^{-\mathrm{T}}\right|$$

由此，我们把本题的广义特征值问题与矩阵$\boldsymbol{G}_1^{-1}\boldsymbol{A}\boldsymbol{G}_2^{-\mathrm{T}}$的特征值问题联系起来。

7.52 已知矩阵

$$
\begin{aligned}
\frac{\mathrm{d}x(t)}{\mathrm{d}t} &= \omega x(t) + 0y(t) + 0z(t) \\
\frac{\mathrm{d}y(t)}{\mathrm{d}t} &= 0x(t) + \omega y(t) + z(t) \\
\frac{\mathrm{d}z(t)}{\mathrm{d}t} &= 0x(t) + y(t) + \omega z(t)
\end{aligned}
$$

求三阶矩阵微分方程

$$\boldsymbol{\Phi}^{(n)}(t) + c_2\boldsymbol{\Phi}''(t) + c_1\boldsymbol{\Phi}'(t) + c_0\boldsymbol{\Phi}(t) = \boldsymbol{O}$$

满足初始条件

$$\boldsymbol{\Phi}(0) = \boldsymbol{I}, \quad \boldsymbol{\Phi}'(0) = \boldsymbol{A}, \quad \boldsymbol{\Phi}''(0) = \boldsymbol{A}^2$$

的解 $\boldsymbol{\Phi}(t)$。

解： 矩阵 $\boldsymbol{A}$ 为

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

容易计算得 $\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = (\lambda-\omega)(\lambda-\omega-1)(\lambda-\omega+1)$。故矩阵 $\boldsymbol{A}$ 的3个特征根为 ω，$\omega+1$ 和 $\omega-1$。容易计算得它们对应的特征向量为

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

因此矩阵 $\boldsymbol{A}$ 有相似标准型

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \omega & 0 & 0 \\ 0 & \omega+1 & 0 \\ 0 & 0 & \omega-1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix}^{-1}$$

由此计算 $\mathrm{e}^{\boldsymbol{A}t}$ 得

$$\mathrm{e}^{\boldsymbol{A}t} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mathrm{e}^{\omega t} & 0 & 0 \\ 0 & \mathrm{e}^{(\omega+1)t} & 0 \\ 0 & 0 & \mathrm{e}^{(\omega-1)t} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix}^{-1}$$

化简得

$$\mathrm{e}^{\boldsymbol{A}t} = \mathrm{e}^{\omega t} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \dfrac{\mathrm{e}^t + \mathrm{e}^{-t}}{2} & \dfrac{\mathrm{e}^t - \mathrm{e}^{-t}}{2} \\ 0 & \dfrac{\mathrm{e}^t - \mathrm{e}^{-t}}{2} & \dfrac{\mathrm{e}^t + \mathrm{e}^{-t}}{2} \end{bmatrix}$$

7.53 若 $\boldsymbol{M}, \boldsymbol{C}, \boldsymbol{K}$ 对称和正定，证明

$$|\lambda^2\boldsymbol{M} + \lambda\boldsymbol{C} + \boldsymbol{K}| = 0$$

的根有负的实部。

证明：参见题7.19的证明。

7.54 已知$m \times m$矩阵

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

证明

$$|\lambda \boldsymbol{I} - \boldsymbol{A}| = \prod_{i=1}^{m} \left(\lambda - 2 - 2\cos\frac{i\pi}{m+1} \right)$$

证明：令$\alpha = \lambda - 2$，我们可将矩阵$\lambda \boldsymbol{I} - \boldsymbol{A}$写成

$$\lambda \boldsymbol{I} - \boldsymbol{A} = \begin{bmatrix} \alpha & 1 & & & \\ 1 & \alpha & 1 & & \\ & & \ddots & & \\ & & 1 & \alpha & 1 \\ & & & 1 & \alpha \end{bmatrix}$$

令$D_m(\alpha) = \det(\lambda \boldsymbol{I} - \boldsymbol{A})$，我们求得$D_m(\alpha) = 0$的根即得$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$的根，从而得到$\det(\lambda \boldsymbol{I} - \boldsymbol{A})$的表达式。将$D_m(\alpha)$按第一行展开得

$$D_m(\alpha) = \alpha D_{m-1}(\alpha) - D_{m-2}(\alpha)$$

可将以上差分方程写成

$$D_m(\alpha) - r_1 D_{m-1}(\alpha) = r_2[D_{m-1}(\alpha) - r_1 D_{m-2}(\alpha)] \tag{7.68}$$

其中r_1，r_2是二次方程

$$x^2 - \alpha x + 1 = 0$$

的根，即$r_{1,2} = \dfrac{\alpha \pm \sqrt{\alpha^2 - 4}}{2}$。

$D_m(\alpha) = 0$的根必使得$r_1 \neq r_2$。事实上，$r_1 = r_2$导致$\alpha = \pm 2$。这里，我们排除$\alpha = \pm 2$是方程$D_m(\alpha) = 0$的根。若不然，则$r_1 = r_2 = \pm 1$。不妨设$\alpha = 2$，$r_1 = r_2 = 1$，则有

$$D_m(2) - D_{m-1}(2) = \cdots = D_3(2) - D_2(2) = 1$$

由于$D_2(2) = 3$，有$D_m(2) = m + 1 \neq 0$。同样的，若$\alpha = -2$，$r_1 = r_2 = -1$，则有

$$\begin{aligned} D_m(-2) + D_{m-1}(-2) &= -[D_{m-1}(-2) + D_{m-2}(-2)] \\ &= \cdots \\ &= (-1)^{m-3}[D_3(-2) + D_2(-2)] \\ &= (-1)^{m-2} \end{aligned}$$

由于$D_2(-2)=3$，所以可递推得$D_m(-2)=(-1)^m(m+1)\neq 0$。

由此知$D_m(\alpha)=0$的根必使得$r_1\neq r_2$。考虑到r_1和r_2的对称性，我们可由递推式(7.68)得到关系式

$$D_m(\alpha)-r_1D_{m-1}(\alpha)=r_2^m \tag{7.69}$$

$$D_m(\alpha)-r_2D_{m-1}(\alpha)=r_1^m \tag{7.70}$$

这里定义$D_1(\alpha)=\alpha$，$D_0(\alpha)=1$。由式(7.69)和式(7.70) 可解得$D_m(\alpha)$为

$$D_m(\alpha)=\frac{r_1^{m+1}-r_2^{m+1}}{r_1-r_2}$$

考虑到$r_1\neq r_2$，那么$D_m(\alpha)=0$的解将满足

$$r_1^{m+1}-r_2^{m+1}=0 \tag{7.71}$$

由于$r_1r_2=1$，故可在等式(7.71)的两边除以r_2得

$$\left(\frac{r_1}{r_2}\right)^{m+1}=1$$

由于$r_1\neq r_2$，故有

$$\frac{r_1}{r_2}=\exp\left(\frac{\mathrm{j}2\pi i}{m+1}\right),\quad i=1,2,\cdots,m$$

其中$\mathrm{j}=\sqrt{-1}$。不妨令$r_1=\dfrac{\alpha+\sqrt{\alpha^2-4}}{2}$，$r_2=\dfrac{\alpha-\sqrt{\alpha^2-4}}{2}$，则有

$$\frac{\alpha+\sqrt{\alpha^2-4}}{\alpha-\sqrt{\alpha^2-4}}=\exp\left(\frac{\mathrm{j}2\pi i}{m+1}\right),\quad i=1,2,\cdots,m \tag{7.72}$$

简化式(7.72)的左边得

$$\frac{(\alpha+\sqrt{\alpha^2-4})^2}{4}=\exp\left(\frac{\mathrm{j}2\pi i}{m+1}\right),\quad i=1,2,\cdots,m \tag{7.73}$$

对式(7.73)两边分别开根号得

$$\alpha+\sqrt{\alpha^2-4}=\pm 2\exp\left(\frac{\mathrm{j}\pi i}{m+1}\right),\quad i=1,2,\cdots,m \tag{7.74}$$

将等式(7.74)左边的α移到等式右边，并两边同时取平方，可解得α为

$$\alpha=\pm 2\cos\left(\frac{\pi i}{m+1}\right),\quad i=1,2,\cdots,m$$

去掉由开根号得到的重复的根，即得

$$\alpha=2\cos\left(\frac{\pi i}{m+1}\right),\quad i=1,2,\cdots,m$$

考虑到$\lambda = \alpha - 2$，即得$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$的$m$个根

$$\lambda_i = 2 + 2\cos\left(\frac{\pi i}{m+1}\right), \quad i = 1, 2, \cdots, m$$

由此得到$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \prod_{i=1}^{m}(\lambda - \lambda_i) = \prod_{i=1}^{m}\left(\lambda - 2 - 2\cos\frac{i\pi}{m+1}\right)$。

7.55 假定$n \times n$维Hermitian矩阵$\boldsymbol{A}$的特征值按照顺序$\lambda_1(\boldsymbol{A}) \geqslant \lambda_2(\boldsymbol{A}) \geqslant \cdots \geqslant \lambda_n(\boldsymbol{A})$排列。用Rayleigh商证明：

(1) $\lambda_1(\boldsymbol{A}+\boldsymbol{B}) \geqslant \lambda_1(\boldsymbol{A}) + \lambda_n(\boldsymbol{B})$。

(2) $\lambda_n(\boldsymbol{A}+\boldsymbol{B}) \geqslant \lambda_n(\boldsymbol{A}) + \lambda_n(\boldsymbol{B})$。

证明：由Rayleigh商的性质知，对称矩阵的最大(小)特征值为其Rayleigh商的最大(小)值。此题的结论可证明如下。

(1)

$$\begin{aligned}
\lambda_1(\boldsymbol{A}+\boldsymbol{B}) &= \max \frac{\boldsymbol{x}^{\mathrm{H}}(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} \\
&= \max\left(\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} + \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\right) \\
&\geqslant \max\left(\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} + \lambda_n(\boldsymbol{B})\right) \\
&= \max\left(\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\right) + \lambda_n(\boldsymbol{B}) \\
&= \lambda_1(\boldsymbol{A}) + \lambda_n(\boldsymbol{B})
\end{aligned}$$

(2)

$$\begin{aligned}
\lambda_n(\boldsymbol{A}+\boldsymbol{B}) &= \min \frac{\boldsymbol{x}^{\mathrm{H}}(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} \\
&= \min\left(\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} + \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\right) \\
&\geqslant \min \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} + \min \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} \\
&= \lambda_n(\boldsymbol{A}) + \lambda_n(\boldsymbol{B})
\end{aligned}$$

7.56 利用Rayleigh 商证明：对于任何一个$n \times n$对称矩阵$\boldsymbol{A}$和任何一个$n \times n$半正定矩阵$\boldsymbol{B}$，特征值服从不等式

$$\lambda_k(\boldsymbol{A}+\boldsymbol{B}) \geqslant \lambda_k(\boldsymbol{A}), \quad k = 1, 2, \cdots, n$$

若$\boldsymbol{B}$正定，则

$$\lambda_k(\boldsymbol{A}+\boldsymbol{B}) > \lambda_k(\boldsymbol{A}), \quad k = 1, 2, \cdots, n$$

证明：考虑到$\boldsymbol{B}$为半正定的，故$\lambda_n(\boldsymbol{B}) \geqslant 0$，对任意的$k = 1,2,\cdots,n$，由Courant-Fischer定理知

$$\begin{aligned}
\lambda_k(\boldsymbol{A}+\boldsymbol{B}) &= \min_{S,\dim(S)=n-k+1} \max_{\boldsymbol{x}\in S,\boldsymbol{x}\neq \boldsymbol{0}} \frac{\boldsymbol{x}^{\mathrm{H}}(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} \\
&= \min_{S,\dim(S)=n-k+1} \max_{\boldsymbol{x}\in S,\boldsymbol{x}\neq \boldsymbol{0}} \left(\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} + \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\right) \\
&\geqslant \min_{S,\dim(S)=n-k+1} \max_{\boldsymbol{x}\in S,\boldsymbol{x}\neq \boldsymbol{0}} \left(\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} + \lambda_n(\boldsymbol{B})\right) \\
&= \min_{S,\dim(S)=n-k+1} \max_{\boldsymbol{x}\in S,\boldsymbol{x}\neq \boldsymbol{0}} \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} + \lambda_n(\boldsymbol{B}) \\
&= \lambda_k(\boldsymbol{A}) + \lambda_n(\boldsymbol{B}) \\
&\geqslant \lambda_k(\boldsymbol{A})
\end{aligned}$$

此即所证结论。在上述的证明中，我们用到Rayleigh商的性质

$$\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} \geqslant \lambda_n(\boldsymbol{B})$$

当$\boldsymbol{B}$正定时，$\lambda_n(\boldsymbol{B}) > 0$，故$\lambda_k(\boldsymbol{A}+\boldsymbol{B}) \geqslant \lambda_k(\boldsymbol{A}) + \lambda_n(\boldsymbol{B}) > \lambda_k(\boldsymbol{A})$，即有结论$\lambda_k(\boldsymbol{A}+\boldsymbol{B}) > \lambda_k(\boldsymbol{A})$, $k = 1,2,\cdots,n$。

7.57 令$\boldsymbol{A}$和$\boldsymbol{B}$分别是$n \times n$ 对称矩阵，证明

$$\lambda_1(\boldsymbol{A}+\boldsymbol{B}) \leqslant \lambda_1(\boldsymbol{A}) + \lambda_1(\boldsymbol{B})$$

和

$$\lambda_n(\boldsymbol{A}+\boldsymbol{B}) \geqslant \lambda_n(\boldsymbol{A}) + \lambda_n(\boldsymbol{B})$$

证明：首先证明不等式$\lambda_1(\boldsymbol{A}+\boldsymbol{B}) \leqslant \lambda_1(\boldsymbol{A}) + \lambda_1(\boldsymbol{B})$，其中用到Rayleigh商的性质，即对称矩阵的最大(小)特征值为其Rayleigh商的最大(小)值。证明过程如下。

$$\begin{aligned}
\lambda_1(\boldsymbol{A}+\boldsymbol{B}) &= \max \frac{\boldsymbol{x}^{\mathrm{H}}(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} \\
&= \max \left(\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} + \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\right) \\
&\leqslant \max \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} + \max \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} \\
&= \lambda_1(\boldsymbol{A}) + \lambda_1(\boldsymbol{B})
\end{aligned}$$

可由同样的方法证明结论$\lambda_n(\boldsymbol{A}+\boldsymbol{B}) \geqslant \lambda_n(\boldsymbol{A}) + \lambda_n(\boldsymbol{B})$，参见题7.55第二问的证明。

7.58 令$\boldsymbol{A}$是一个$n \times n$矩阵(不一定对称)。证明对任意$n \times 1$向量$\boldsymbol{x}$，恒有不等式

$$(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})^2 \leqslant (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x})(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})$$

和

$$\frac{1}{2}\left|\frac{\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}\right| \leqslant \left(\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}\right)^{1/2}$$

证明：首先证明第一个不等式。注意到$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}$，利用Rayleigh商的性质，有不等式

$$\begin{aligned}\frac{(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})^2}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} &= \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}\\ &\leqslant \lambda_{\max}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A})\end{aligned}$$

注意到矩阵$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}$是半正定的，秩为1，其最大的特征值为$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}$，对应的特征向量为$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}$。矩阵$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}$剩余的$n-1$个特征值为零，对应的特征向量为与$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}$正交的向量。故$\lambda_{\max}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}$，即有不等式

$$\frac{(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})^2}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} \leqslant \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}$$

将上不等式左边和右边同乘因子$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$即得所需证明的不等式。

注意到$\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})\boldsymbol{x}=2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$，我们可将第一问证得的不等式重写为

$$\left[\frac{\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})\boldsymbol{x}}{2}\right]^2 \leqslant (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x})(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})$$

将上不等式两边同除因子$(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})^2$并开根号即可得第二问所需证明的不等式。

7.59 考虑对称矩阵序列

$$\boldsymbol{A}_r=[a_{ij}], \qquad i,j=1,2,\cdots,r$$

其中，$r=1,2,\cdots,n$。令$\lambda_i(\boldsymbol{A}_r)$, $i=1,2,\cdots,r$是矩阵$\boldsymbol{A}_r$ 的第i个特征值，并且

$$\lambda_1(\boldsymbol{A}_r)\geqslant\lambda_2(\boldsymbol{A}_r)\geqslant\cdots\geqslant\lambda_r(\boldsymbol{A}_r)$$

则

$$\lambda_{k+1}(\boldsymbol{A}_{i+1})\leqslant\lambda_k(\boldsymbol{A}_i)\leqslant\lambda_k(\boldsymbol{A}_{i+1})$$

这一结果称为Sturmian分离定理[31, p.117]。试使用Rayleigh商证明这一定理。

证明：注意到$\boldsymbol{A}_{i+1}$与$\boldsymbol{A}_i$之间的关系为

$$\boldsymbol{A}_{i+1}=\begin{bmatrix}\boldsymbol{A}_i & \boldsymbol{\alpha}\\ \boldsymbol{\alpha}^{\mathrm{T}} & a_{i+1,i+1}\end{bmatrix}$$

其中，$\boldsymbol{\alpha}=[a_{1,i+1},a_{2,i+1},\cdots,a_{i,i+1}]^{\mathrm{T}}$ 为一i维向量。记$\boldsymbol{x}_i$为一i维向量，那么由Courant-Fisher定理

$$\lambda_k=\min_{S,\dim(S)=n-k+1}\ \max_{\boldsymbol{x}\in S,\boldsymbol{x}\neq\boldsymbol{0}}\left\{\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\right\}$$

可得

$$\begin{aligned}
\lambda_k(\boldsymbol{A}_{i+1}) &= \min_{S,\dim(S)=i+1-k+1}\ \max_{\boldsymbol{x}_{i+1}\in S,\,\boldsymbol{x}_{i+1}\neq\boldsymbol{0}} \left\{\frac{\boldsymbol{x}_{i+1}^{\mathrm{H}}\boldsymbol{A}_{i+1}\boldsymbol{x}_{i+1}}{\boldsymbol{x}_{i+1}^{\mathrm{H}}\boldsymbol{x}_{i+1}}\right\} \\
&\geqslant \min_{S,\dim(S)=i+1-k+1}\ \max_{\boldsymbol{x}_{i+1}\in S,\,\boldsymbol{x}_{i+1}\neq\boldsymbol{0},\,\boldsymbol{x}_{i+1}=[\boldsymbol{x}_i,0]^{\mathrm{T}}} \left\{\frac{\boldsymbol{x}_{i+1}^{\mathrm{H}}\boldsymbol{A}_{i+1}\boldsymbol{x}_{i+1}}{\boldsymbol{x}_{i+1}^{\mathrm{H}}\boldsymbol{x}_{i+1}}\right\} \\
&= \min_{S,\dim(S)=i-k+1}\ \max_{\boldsymbol{x}_i\in S,\,\boldsymbol{x}_i\neq\boldsymbol{0}} \left\{\frac{\boldsymbol{x}_i^{\mathrm{H}}\boldsymbol{A}_i\boldsymbol{x}_i}{\boldsymbol{x}_i^{\mathrm{H}}\boldsymbol{x}_i}\right\} \\
&= \lambda_k(\boldsymbol{A}_i)
\end{aligned}$$

即$\lambda_k(\boldsymbol{A}_i)\leqslant\lambda_k(\boldsymbol{A}_{i+1})$。

类似地，由Courant-Fisher定理

$$\lambda_k=\max_{S,\dim S=k}\ \min_{\boldsymbol{x}\in S,\,\boldsymbol{x}\neq\boldsymbol{0}}\left\{\frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}\right\}$$

可得

$$\begin{aligned}
\lambda_{k+1}(\boldsymbol{A}_{i+1}) &= \max_{S,\dim(S)=k+1}\ \min_{\boldsymbol{x}_{i+1}\in S,\,\boldsymbol{x}_{i+1}\neq\boldsymbol{0}} \left\{\frac{\boldsymbol{x}_{i+1}^{\mathrm{H}}\boldsymbol{A}_{i+1}\boldsymbol{x}_{i+1}}{\boldsymbol{x}_{i+1}^{\mathrm{H}}\boldsymbol{x}_{i+1}}\right\} \\
&\leqslant \max_{S,\dim(S)=k+1}\ \min_{\boldsymbol{x}_{i+1}\in S,\,\boldsymbol{x}_{i+1}\neq\boldsymbol{0},\,\boldsymbol{x}_{i+1}=[\boldsymbol{x}_i,0]^{\mathrm{T}}} \left\{\frac{\boldsymbol{x}_{i+1}^{\mathrm{H}}\boldsymbol{A}_{i+1}\boldsymbol{x}_{i+1}}{\boldsymbol{x}_{i+1}^{\mathrm{H}}\boldsymbol{x}_{i+1}}\right\} \\
&= \max_{S,\dim(S)=k}\ \min_{\boldsymbol{x}_i\in S,\,\boldsymbol{x}_i\neq\boldsymbol{0}} \left\{\frac{\boldsymbol{x}_i^{\mathrm{H}}\boldsymbol{A}_i\boldsymbol{x}_i}{\boldsymbol{x}_i^{\mathrm{H}}\boldsymbol{x}_i}\right\} \\
&= \lambda_k(\boldsymbol{A}_i)
\end{aligned}$$

即$\lambda_k(\boldsymbol{A}_i)\geqslant\lambda_{k+1}(\boldsymbol{A}_{i+1})$。

综上证明可得

$$\lambda_{k+1}(\boldsymbol{A}_{i+1})\leqslant\lambda_k(\boldsymbol{A}_i)\leqslant\lambda_k(\boldsymbol{A}_{i+1})$$

7.60 证明教材中式(8.11.1) 是双曲线二次特征值问题，当且仅当$\boldsymbol{Q}(\lambda)$ 对某些实特征值$\lambda\in R$ 负定。

证明：$\boldsymbol{Q}(\lambda)$的表达式为

$$\boldsymbol{Q}(\lambda)=\lambda^2\boldsymbol{M}+\lambda\boldsymbol{C}+\boldsymbol{K}$$

首先证明充分性。设$\boldsymbol{Q}(\lambda)$对某些实特征值$\lambda\in R$ 负定，不妨设$\boldsymbol{Q}(\lambda_0)$负定，即二次型

$$\boldsymbol{p}^{\mathrm{T}}(\lambda_0^2\boldsymbol{M}+\lambda_0\boldsymbol{C}+\boldsymbol{K})\boldsymbol{p}<0$$

恒成立。由于$\boldsymbol{M}$正定，因此$\boldsymbol{p}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{p}>0$总成立，故关于$\lambda$的二次方程

$$\boldsymbol{p}^{\mathrm{T}}(\lambda^2\boldsymbol{M}+\lambda\boldsymbol{C}+\boldsymbol{K})\boldsymbol{p}=0$$

恒有实数解，因此其判别式总是大于零的，即恒有

$$(\boldsymbol{p}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{p})^2 > 4(\boldsymbol{p}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{p})(\boldsymbol{p}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{p})$$

故二次特征值问题是双曲线二次特征值问题。

问题必要性的证明可参考文献[17]。

7.61 对于两个相同维数的任意正方矩阵$\boldsymbol{A},\boldsymbol{B}$，证明

(1) $2(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}})-(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}}$半正定。

(2) $\mathrm{tr}[(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}}]\leqslant 2[\mathrm{tr}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})+\mathrm{tr}(\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}})]$。

(3) $\lambda[(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}}]\leqslant 2[\lambda(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})+\lambda(\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}})]$。

证明：(1) 容易验证等式

$$2(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}})-(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}}=(\boldsymbol{A}-\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})^{\mathrm{T}}$$

故矩阵$2(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}})-(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}}$半正定性。

(2) 由于半正定矩阵的对角线元素都是非负，故有不等式

$$\mathrm{tr}[2(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}})-(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}}]\geqslant 0$$

展开上述不等式即得所需证的不等式。

(3) 令矩阵$\mathcal{A}$和矩阵$\mathcal{B}$为

$$\mathcal{A}=(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}}$$

$$\mathcal{B}=2(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}})-(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}}$$

已经证明矩阵$\mathcal{B}$是半正定的，由题7.56的结论可得

$$\lambda(\mathcal{A}+\mathcal{B})\geqslant\lambda(\mathcal{A})$$

展开上述不等式即得所需证明的不等式。

7.62 令$\boldsymbol{A}$ 和$\boldsymbol{B}$是两个维数相同的半正定矩阵，证明[38],[36]

$$\sqrt{\mathrm{tr}(\boldsymbol{A}\boldsymbol{B})}\leqslant\frac{1}{2}[\mathrm{tr}(\boldsymbol{A})+\mathrm{tr}(\boldsymbol{B})]$$

等号成立，当且仅当$\boldsymbol{A}=\boldsymbol{B}$和$\mathrm{rank}(\boldsymbol{A})\leqslant 1$。

证明：设$\boldsymbol{A}$和$\boldsymbol{B}$均是$n\times n$维的半正定矩阵。由于$\boldsymbol{B}$半正定，设$\boldsymbol{B}$的特征值分解为

$$\boldsymbol{B}=\boldsymbol{U}\boldsymbol{\Sigma}_B\boldsymbol{U}^{\mathrm{H}}$$

其中$\boldsymbol{U}$为酉阵，$\boldsymbol{\Sigma}_B$为对角线元素非负的对角阵。由迹的性质有

$$\begin{aligned}\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) &= \mathrm{tr}(\boldsymbol{A}\boldsymbol{U}\boldsymbol{\Sigma}_B\boldsymbol{U}^{\mathrm{H}})\\ &= \mathrm{tr}(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}\boldsymbol{\Sigma}_B)\end{aligned}$$

记$\boldsymbol{A}_U=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}$，则有

$$\mathrm{tr}(\boldsymbol{AB})=\sum_{i=1}^{n}\boldsymbol{A}_{U,ii}\boldsymbol{\Sigma}_{B,i}$$

其中$\boldsymbol{A}_{U,ii}$表示$\boldsymbol{A}_U$的第i个对角线元素，$\boldsymbol{\Sigma}_{B,i}$为$\boldsymbol{\Sigma}_B$的第$i$个对角线元素。注意到$\boldsymbol{A}$是半正定的，则$\boldsymbol{A}_U$仍是半正定的，其对角线上的元素$\boldsymbol{A}_{U,ii}$都是非负的。

同样，由迹的性质有

$$\begin{aligned}\mathrm{tr}(\boldsymbol{A})&=\mathrm{tr}(\boldsymbol{A}_U)\\ \mathrm{tr}(\boldsymbol{B})&=\mathrm{tr}(\boldsymbol{\Sigma}_B)\end{aligned}$$

故所需证明的不等式可写为

$$\sqrt{\sum_{i=1}^{n}\boldsymbol{A}_{U,ii}\boldsymbol{\Sigma}_{B,i}}\leqslant\frac{1}{2}[\mathrm{tr}(\boldsymbol{A}_U)+\mathrm{tr}(\boldsymbol{\Sigma}_B)]$$

将上不等式两边同时取平方，得到等价的不等式

$$4\sum_{i=1}^{n}\boldsymbol{A}_{U,ii}\boldsymbol{\Sigma}_{B,i}\leqslant[\mathrm{tr}(\boldsymbol{A}_U)+\mathrm{tr}(\boldsymbol{\Sigma}_B)]^2$$

记

$$\boldsymbol{a}=[\boldsymbol{A}_{U,11},\cdots,\boldsymbol{A}_{U,nn}]^{\mathrm{T}}$$

$$\boldsymbol{b}=[\boldsymbol{\Sigma}_{B,1},\cdots,\boldsymbol{\Sigma}_{B,n}]^{\mathrm{T}}$$

和一个元素全为1的n维矢量

$$\mathbf{1}=[1,\cdots,1]^{\mathrm{T}}$$

那么所需证明的不等式可简写成

$$4\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}\leqslant[\boldsymbol{a}^{\mathrm{T}}\mathbf{1}+\mathbf{1}^{\mathrm{T}}\boldsymbol{b}]^2$$

由熟知的算术不等式，可得到

$$\begin{aligned}[\boldsymbol{a}^{\mathrm{T}}\mathbf{1}+\mathbf{1}^{\mathrm{T}}\boldsymbol{b}]^2&\geqslant[2\sqrt{\boldsymbol{a}^{\mathrm{T}}\mathbf{1}\mathbf{1}^{\mathrm{T}}\boldsymbol{b}}]^2\\ &=4\boldsymbol{a}^{\mathrm{T}}\mathbf{1}\mathbf{1}^{\mathrm{T}}\boldsymbol{b}\end{aligned}$$

由于矢量$\boldsymbol{a}$和$\boldsymbol{b}$中的元素都是非负的，易知$4\boldsymbol{a}^{\mathrm{T}}\mathbf{1}\mathbf{1}^{\mathrm{T}}\boldsymbol{b}\geqslant4\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}$，到此不等式证明完毕。

从以上的证明可以观察到，不等式取等号的条件是

$$\boldsymbol{a}^{\mathrm{T}}[\mathbf{1}\mathbf{1}^{\mathrm{T}}-\boldsymbol{I}]\boldsymbol{b}=0$$

和

$$\mathbf{1}^{\mathrm{T}}\boldsymbol{a}=\mathbf{1}^{\mathrm{T}}\boldsymbol{b}$$

这里$\boldsymbol{I}$为单位阵。第一个条件意味着：(1)要么矢量$\boldsymbol{a}$和矢量$\boldsymbol{b}$中各自至多有一个非零元素，且在矢量中的位置一样；(2)要么一个矢量中元素全为零，另一个可有非负元素。进一步考虑到取等号的条件(2)知，只能是矢量$\boldsymbol{a}$和矢量$\boldsymbol{b}$至多在相同的位置有一个非零元素，故$\mathrm{rank}(\boldsymbol{B}) \leqslant 1$且$\boldsymbol{a} = \boldsymbol{b}$。

不妨设$\boldsymbol{a}$的第一个元素允许为非零的(注意只能是正的)，剩下的元素都为0，即$\boldsymbol{A}_U$除了第一个对角线上的元素可能为正的，余下的对角线元素都为0。考虑到$\boldsymbol{A}_U$的半正定性，则对任意的矢量$\boldsymbol{x} = [x_1, \cdots, x_n]$，有不等式

$$a_1 x_1^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_{ij}x_i x_j \geqslant 0$$

其中，a_{ij}为$\boldsymbol{A}_U$的第i行、j列的元素。以上不等式只能在$a_{ij} = 0, \forall 1 \leqslant i \neq j \leqslant n$时才恒成立。因此，酉阵$\boldsymbol{U}$也同时使得$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}$对角化，且与$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{U}$有一样的对角线元素(即$\boldsymbol{a} = \boldsymbol{b}$)，故$\boldsymbol{A} = \boldsymbol{B}$，且$\mathrm{rank}(\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{B}) \leqslant 1$。至此证毕。

7.63 对于多项式$p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$，称$n \times n$ 矩阵

$$\boldsymbol{C}_p = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix}$$

为多项式$p(x)$的友矩阵。利用数学归纳法证明：对于$n \geqslant 2$，恒有

$$\det(\boldsymbol{C}_p - \lambda\boldsymbol{I}) = (-1)^n(a_0 + a_1\lambda + \cdots + a_{n-1}\lambda^{n-1} + \lambda^n) = (-1)^n p(\lambda)$$

证明：首先，容易验证$n = 2$时结论成立。当$n = 2$时有

$$\boldsymbol{C}_p = \begin{bmatrix} 0 & 1 \\ -a_0 & -a_1 \end{bmatrix}$$

直接计算$\det(\boldsymbol{C}_p - \lambda\boldsymbol{I})$得

$$\begin{aligned} \det(\boldsymbol{C}_p - \lambda\boldsymbol{I}) &= -\lambda(-a_1 - \lambda) - (-a_0) \\ &= (-1)^2(\lambda^2 + a_1\lambda + a_0) \end{aligned}$$

故此时结论成立。

设$n \geqslant 2$时结论成立，我们证明$n+1$时结论仍成立。$n+1$时的友矩阵为

$$\boldsymbol{C}_p = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_n \end{bmatrix}$$

将以上$(n+1)\times(n+1)$维矩阵的行列式按第一行展开为一系列$n\times n$维矩阵行列式的和，我们可以计算特征多项式$\det(\boldsymbol{C}_p-\lambda\boldsymbol{I})$如下

$$
\begin{aligned}
\det(\boldsymbol{C}_p-\lambda\boldsymbol{I}) &= \begin{vmatrix} -\lambda & 1 & 0 & \cdots & 0 \\ 0 & -\lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_n-\lambda \end{vmatrix} \\
&= -\lambda \begin{vmatrix} -\lambda & 1 & 0 & \cdots & 0 \\ 0 & -\lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_1 & -a_2 & -a_3 & \cdots & -a_n-\lambda \end{vmatrix} \\
&\quad - \begin{vmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & -\lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & 0 & \cdots & 1 \\ -a_0 & -a_2 & -a_3 & \cdots & -a_n-\lambda \end{vmatrix} \\
&= -\lambda(-1)^n(a_1+a_2\lambda+\cdots+a_n\lambda^{n-1}+\lambda^n)-(-1)^{n-1}(-a_0) \\
&= (-1)^{n+1}(a_0+a_1\lambda+a_2\lambda^2+\cdots+a_n\lambda^n+\lambda^{n+1}) \\
&= (-1)^{n+1}p(\lambda)
\end{aligned}
$$

故$n+1$时结论仍成立。由数学归纳法知，对于$n\geqslant 2$，恒有

$$\det(\boldsymbol{C}_p-\lambda\boldsymbol{I})=(-1)^n(a_0+a_1\lambda+\cdots+a_{n-1}\lambda^{n-1}+\lambda^n)=(-1)^np(\lambda)$$

7.64 令$p(x)=x^3+a_2x^2+a_1x+a_0$，并且λ是多项式$p(x)$的一个零点。

(1) 写出多项式$p(x)$的友矩阵$\boldsymbol{C}_p$。

(2) 解释为什么$\lambda^3=-a_2\lambda^2-a_1\lambda-a_0$，并证明$(1,\lambda,\lambda^2)$是多项式$p(x)$的友矩阵$\boldsymbol{C}_p$的特征向量。

解：(1) 参考题7.63，可写出多项式$p(x)$的友矩阵$\boldsymbol{C}_p$为

$$\boldsymbol{C}_p=\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_0 & -a_1 & -a_2 \end{bmatrix}$$

(2) 令$p(\lambda)=0$即得等式$\lambda^3=-a_2\lambda^2-a_1\lambda-a_0$，可直接验证$(1,\lambda,\lambda^2)$是多项式$p(x)$的友矩阵$\boldsymbol{C}_p$的特征向量，对应的特征值为$\lambda$。

7.65 [34] 令$\boldsymbol{A},\boldsymbol{B}$和$\boldsymbol{A}-\boldsymbol{B}$为半正定矩阵。证明：$\boldsymbol{B}^\dagger-\boldsymbol{A}^\dagger$ 是半正定的，当且仅当$\operatorname{rank}(\boldsymbol{A})=\operatorname{rank}(\boldsymbol{B})$。

证明： 首先验证一个广义逆的等式。令$\boldsymbol{P} \in C^{n\times r}$是列满秩矩阵，$\boldsymbol{Q} \in C^{r\times r}$为可逆矩阵，则矩阵$\boldsymbol{PQP}^{\mathrm{H}}$的伪逆$\boldsymbol{G} = (\boldsymbol{P}^{\dagger})^{\mathrm{H}}\boldsymbol{Q}^{-1}\boldsymbol{P}^{\dagger}$。事实上，由于$\boldsymbol{P}$是列满秩的，有$\boldsymbol{P}^{\dagger} = (\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{-1}\boldsymbol{P}^{\mathrm{H}}$，即$\boldsymbol{G} = \boldsymbol{P}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{-1}\boldsymbol{Q}^{-1}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{-1}\boldsymbol{P}^{\mathrm{H}}$，这里验证矩阵$\boldsymbol{G}$满足Moore-Penrose伪逆的4个条件，过程如下：

$$(\boldsymbol{PQP}^{\mathrm{H}})\boldsymbol{G}(\boldsymbol{PQP}^{\mathrm{H}}) = \boldsymbol{PQP}^{\mathrm{H}}\boldsymbol{P}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{-1}\boldsymbol{Q}^{-1}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{-1}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{PQP}^{\mathrm{H}} = \boldsymbol{PQP}^{\mathrm{H}}$$

$$\begin{aligned}\boldsymbol{G}(\boldsymbol{PQP}^{\mathrm{H}})\boldsymbol{G} &= \boldsymbol{P}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{-1}\boldsymbol{Q}^{-1}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{-1}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{PQP}^{\mathrm{H}}\boldsymbol{P}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{-1}\boldsymbol{Q}^{-1}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{-1}\boldsymbol{P}^{\mathrm{H}}\\ &= \boldsymbol{P}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{-1}\boldsymbol{Q}^{-1}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{-1}\boldsymbol{P}^{\mathrm{H}}\\ &= \boldsymbol{G}\end{aligned}$$

$$(\boldsymbol{PQP}^{\mathrm{H}})\boldsymbol{G} = \boldsymbol{PQP}^{\mathrm{H}}\boldsymbol{P}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{-1}\boldsymbol{Q}^{-1}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{-1}\boldsymbol{P}^{\mathrm{H}} = \boldsymbol{P}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{-1}\boldsymbol{P}^{\mathrm{H}}$$

$$\boldsymbol{G}(\boldsymbol{PQP}^{\mathrm{H}}) = \boldsymbol{P}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{-1}\boldsymbol{Q}^{-1}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{-1}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{PQP}^{\mathrm{H}} = \boldsymbol{P}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P})^{-1}\boldsymbol{P}^{\mathrm{H}}$$

显然，$(\boldsymbol{PQP}^{\mathrm{H}})\boldsymbol{G}$和$\boldsymbol{G}(\boldsymbol{PQP}^{\mathrm{H}})$都是Hermitian矩阵，因此，矩阵$\boldsymbol{G} = (\boldsymbol{P}^{\dagger})^{\mathrm{H}}\boldsymbol{Q}^{-1}\boldsymbol{P}^{\dagger}$为矩阵$\boldsymbol{PQP}^{\mathrm{H}}$的广义逆。

下面我们开始证明本题的结论。由于$\boldsymbol{A}$和$\boldsymbol{B}$均为半正定矩阵，故存在可逆矩阵$\boldsymbol{P}$，使得矩阵$\boldsymbol{A}$和$\boldsymbol{B}$可同时表示为

$$\boldsymbol{A} = \boldsymbol{P\Lambda}_A\boldsymbol{P}^{\mathrm{H}}$$

$$\boldsymbol{B} = \boldsymbol{P\Lambda}_B\boldsymbol{P}^{\mathrm{H}}$$

其中$\boldsymbol{\Lambda}_A$和$\boldsymbol{\Lambda}_B$为对角矩阵。由于$\boldsymbol{A}$和$\boldsymbol{B}$为半正定矩阵，它们的对角线元素都是非负的。另外，由于$\boldsymbol{A}-\boldsymbol{B}$为半正定矩阵，故$\boldsymbol{\Lambda}_A - \boldsymbol{\Lambda}_B$也为半正定的。因此，矩阵$\boldsymbol{A}$的秩不小于矩阵$\boldsymbol{B}$的秩，且相应的对角线元素不小于矩阵$\boldsymbol{B}$相应的对角线元素。

我们先证明必要条件。当$\mathrm{rank}(\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{B})$，则不妨设$\mathrm{rank}(\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{B}) = r$，这时$\boldsymbol{\Lambda}_A$和$\boldsymbol{\Lambda}_B$仅有$r$个非零元素，不妨设前$r$ 个元素是非零的，则可将矩阵$\boldsymbol{A}$和矩阵$\boldsymbol{B}$写成

$$\boldsymbol{A} = \boldsymbol{P}_r\mathrm{diag}(d_{A,1},\cdots,d_{A,r})\boldsymbol{P}_r^{\mathrm{H}}$$

$$\boldsymbol{B} = \boldsymbol{P}_r\mathrm{diag}(d_{B,1},\cdots,d_{B,r})\boldsymbol{P}_r^{\mathrm{H}}$$

其中，矩阵$\boldsymbol{P}_r$由矩阵$\boldsymbol{P}$的前r个列向量组成，$d_{A,1},\cdots,d_{A,r}$为对角矩阵$\boldsymbol{\Lambda}_A$的前r个非零对角元素，$d_{B,1},\cdots,d_{B,r}$为对角矩阵$\boldsymbol{\Lambda}_B$的前r个非零对角元素。根据以上的广义逆等式，我们可以求得矩阵$\boldsymbol{A}$和矩阵$\boldsymbol{B}$的伪逆为

$$\boldsymbol{A}^{\dagger} = (\boldsymbol{P}_r^{\dagger})^{\mathrm{H}}\mathrm{diag}\left(\frac{1}{d_{A,1}},\cdots,\frac{1}{d_{A,r}}\right)\boldsymbol{P}_r^{\dagger}$$

和

$$\boldsymbol{B}^{\dagger} = (\boldsymbol{P}_r^{\dagger})^{\mathrm{H}}\mathrm{diag}\left(\frac{1}{d_{B,1}},\cdots,\frac{1}{d_{B,r}}\right)\boldsymbol{P}_r^{\dagger}$$

由于$\boldsymbol{A}$的对角线元素不小于$\boldsymbol{B}$的对角线元素，因此，取倒数后，$\boldsymbol{A}^{\dagger}$的对角线元素不大于$\boldsymbol{B}^{\dagger}$的对角线元素，即$\boldsymbol{B}^{\dagger}-\boldsymbol{A}^{\dagger}$ 是半正定的。

充分性条件可由问题的对称性质得到。注意到若矩阵$\boldsymbol{A}$是半正定的，则$\boldsymbol{A}^{\dagger}$也是半正定的，且与矩阵$\boldsymbol{A}$有相同的秩。因此，若矩阵$\boldsymbol{B}^{\dagger}-\boldsymbol{A}^{\dagger}$半正定，则矩阵$\boldsymbol{B}$的秩不小于矩阵$\boldsymbol{A}$的秩，我们已知矩阵$\boldsymbol{A}$的秩不小于矩阵$\boldsymbol{B}$的秩，因此只能是矩阵$\boldsymbol{A}$和矩阵$\boldsymbol{B}$有相等的秩。

注释： 通常情况下，$(\boldsymbol{AB})^{\dagger}\neq\boldsymbol{B}^{\dagger}\boldsymbol{A}^{\dagger}$，因此，等式$(\boldsymbol{PQP}^{\mathrm{H}})^{\dagger}=(\boldsymbol{P}^{\dagger})^{\mathrm{H}}\boldsymbol{Q}^{-1}\boldsymbol{P}^{\dagger}$ 需要在一定的假设条件下成立（如本题证明中用到的条件）。

7.66　令$\boldsymbol{A}$ 是一个$n\times n$矩阵(不一定对称)。证明：对于每一个$n\times 1$向量$\boldsymbol{x}$，恒有

$$(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Ax})^2\leqslant(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{AA}^{\mathrm{T}}\boldsymbol{x})(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})$$

因此有

$$\frac{1}{2}\left|\frac{\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}\right|\leqslant\left(\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{AA}^{\mathrm{T}}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}\right)^{1/2}$$

证明： 参见题7.58。

7.67　证明：对于任何一个$n\times n$对称矩阵$\boldsymbol{A}$和任何一个半正定矩阵$\boldsymbol{B}$，恒有特征值不等式

$$\lambda_k(\boldsymbol{A}+\boldsymbol{B})\geqslant\lambda_k(\boldsymbol{A}),\qquad k=1,2,\cdots,n$$

并且若$\boldsymbol{B}$正定，则有

$$\lambda_k(\boldsymbol{A}+\boldsymbol{B})>\lambda_k(\boldsymbol{A}),\qquad k=1,2,\cdots,n$$

(提示：利用Rayleigh商的极大–极小原理。)

证明： 参见题7.56。

7.68 [1]　令$\boldsymbol{A}$ 是$n\times n$正定矩阵，其特征值$0<\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$。证明：矩阵

$$(\lambda_1+\lambda_n)\boldsymbol{I}_n-\boldsymbol{A}-(\lambda_1\lambda_n)\boldsymbol{A}^{-1}$$

半正定，并且其秩$\leqslant n-2$。（提示：利用$x^2-(a+b)x+ab\leqslant 0,\forall\, x\in[a,b]$。）

证明： 首先注意到矩阵$(\lambda_1+\lambda_n)\boldsymbol{I}_n-\boldsymbol{A}-(\lambda_1\lambda_n)\boldsymbol{A}^{-1}$仍为对称矩阵，我们可以通过检查它的特征值来证明它的半正定性和计算它的秩。令$\boldsymbol{A}$有分解形式

$$\boldsymbol{A}=\boldsymbol{UDU}^{\mathrm{T}}$$

其中$\boldsymbol{U}$为正交矩阵，$\boldsymbol{D}=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$为对角矩阵，则我们可将矩阵$(\lambda_1+\lambda_n)\boldsymbol{I}_n-\boldsymbol{A}-(\lambda_1\lambda_n)\boldsymbol{A}^{-1}$写成

$$\begin{aligned}(\lambda_1+\lambda_n)\boldsymbol{I}_n-\boldsymbol{A}-(\lambda_1\lambda_n)\boldsymbol{A}^{-1}&=(\lambda_1+\lambda_n)\boldsymbol{I}_n-\boldsymbol{UDU}^{\mathrm{T}}-\lambda_1\lambda_n\boldsymbol{UD}^{-1}\boldsymbol{U}^{\mathrm{T}}\\&=\boldsymbol{U}\left[(\lambda_1+\lambda_n)\boldsymbol{I}_n-\boldsymbol{D}-\lambda_1\lambda_n\boldsymbol{D}^{-1}\right]\boldsymbol{U}^{\mathrm{T}}\end{aligned}$$

容易写出对称矩阵$(\lambda_1+\lambda_n)\boldsymbol{I}_n-\boldsymbol{A}-(\lambda_1\lambda_n)\boldsymbol{A}^{-1}$的第$i$个特征值为

$$\lambda_1+\lambda_n-\lambda_i-\frac{\lambda_1\lambda_n}{\lambda_i}$$

或者进一步将其写成形式

$$-\frac{(\lambda_i-\lambda_1)(\lambda_i-\lambda_n)}{\lambda_i}$$

由于$0<\lambda_1\leqslant\lambda_i\leqslant\lambda_n$，故矩阵$(\lambda_1+\lambda_n)\boldsymbol{I}_n-\boldsymbol{A}-(\lambda_1\lambda_n)\boldsymbol{A}^{-1}$的每一个特征值都是非负的，且第1和第$n$个特征值为0，因此矩阵$(\lambda_1+\lambda_n)\boldsymbol{I}_n-\boldsymbol{A}-(\lambda_1\lambda_n)\boldsymbol{A}^{-1}$半正定，并且秩$\leqslant n-2$。

7.69 [1] 令$\boldsymbol{A}$ 是一个$n\times n$正定矩阵，其特征值$0<\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$。利用上一习题证明

$$1\leqslant(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{x})\leqslant\frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}$$

对所有满足$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}=1$的实向量$\boldsymbol{x}$成立。这一不等式称为Kantorovich不等式[35]。

证明：首先，我们证明不等式的左边。令$\boldsymbol{A}$有分解形式

$$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{T}}$$

其中$\boldsymbol{U}$为正交矩阵，$\boldsymbol{D}=\mathrm{diag}(\lambda_1,\cdots.\lambda_n)$为对角矩阵。令向量$\boldsymbol{y}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}$，则$\boldsymbol{y}$仍为单位长度向量。我们将表达式$(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{x})$写成$\boldsymbol{y}$的函数形式有

$$(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{x})=(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{y})(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{D}^{-1}\boldsymbol{y})$$

由Cauchy不等式得

$$\begin{aligned}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{x}) &= (\boldsymbol{y}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{y})(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{D}^{-1}\boldsymbol{y})\\ &= \left(\sum_{i=1}^{n}\lambda_i y_i^2)(\sum_{i=1}^{n}\lambda_i^{-1}y_i^2\right)\\ &\geqslant \left(\sum_{i=1}^{n}\lambda_i\lambda_i^{-1}y_iy_i\right)^2\\ &= \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}=1\end{aligned}$$

其中$\boldsymbol{y}=[y_1,\cdots,y_n]^{\mathrm{T}}$。由此证明了不等式的左边。

关于不等式右边的证明，需要用到题7.67的结论。由题7.67的结论知有不等式

$$\boldsymbol{x}^{\mathrm{T}}\left[(\lambda_1+\lambda_n)\boldsymbol{I}_n-\boldsymbol{A}-(\lambda_1\lambda_n)\boldsymbol{A}^{-1}\right]\boldsymbol{x}\geqslant 0$$

由题设知$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}=1$，可将上述不等式简化为

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}+\lambda_1\lambda_n\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{x}\leqslant\lambda_1+\lambda_n$$

再在上述不等式的左边运用算术不等式，可得不等式

$$2\sqrt{\lambda_1\lambda_n\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{x}}\leqslant\lambda_1+\lambda_n$$

将上述不等式两边同时取平方，即得所需证明的不等式的右边。

第8章　子空间分析与跟踪

子空间分析在现代信号处理中有着广泛的应用。本章习题内容主要包含子空间的一些基本性质，运用子空间的方法解决通信中的问题的应用。

8.1　主要理论与方法

定义 8.1　若$S=\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_m\}$是向量空间$V$的向量子集合，则$\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_m$的所有线性组合的集合$W$称为由$\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_m$张成的子空间，定义为

$$W=\mathrm{Span}\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_m\}=\{\boldsymbol{u}:\ \boldsymbol{u}=a_1\boldsymbol{u}_1+a_2\boldsymbol{u}_2+\cdots+a_m\boldsymbol{u}_m\} \tag{8.1}$$

定义 8.2　令W是一向量子空间。向量集合$\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p\}$称为$W$的一组基，若下列两个条件满足：

(1) 子空间W由向量$\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p$张成，即

$$W=\mathrm{Span}\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p\}$$

(2) 向量集合$B=\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p\}$是一线性无关的集合。

定理 8.1　令$\boldsymbol{P}$是到子空间S的正交投影算子，则对于复向量空间C^n内的任意向量$\boldsymbol{x}$，有

$$\min_{\boldsymbol{y}\in S}\|\boldsymbol{x}-\boldsymbol{y}\|_2=\|\boldsymbol{x}-\boldsymbol{P}\boldsymbol{x}\|_2 \tag{8.2}$$

定义 8.3 [4, p.76]　假定S_1和S_2是C^n的两个子空间，并且$\dim(S_1)=\dim(S_2)$，则这两个子空间之间的距离定义为

$$\mathrm{dist}(S_1,S_2)=\|\boldsymbol{P}_{S_1}-\boldsymbol{P}_{S_2}\|_{\mathrm{F}} \tag{8.3}$$

式中，$\boldsymbol{P}_{S_i}$是到子空间$S_i, i=1,2$ 的正交投影算子。

定义 8.4[4]　子空间H_1 与H_2之间的第i主角(principal angle) $\phi_i(H_1,H_2)$是介于0和$\pi/2$之间的角度，定义为

$$\phi_i(H_1,H_2)=\arccos\left(\max_{\boldsymbol{u}\in H_1}\max_{\boldsymbol{v}\in H_2}\boldsymbol{u}^{\mathrm{H}}\boldsymbol{v}\right)=\arccos\left(\boldsymbol{u}_i^{\mathrm{H}}\boldsymbol{v}_i\right) \tag{8.4}$$

约束条件为

$$\left.\begin{aligned} &\boldsymbol{u}^{\mathrm{H}}\boldsymbol{u}=\boldsymbol{v}^{\mathrm{H}}\boldsymbol{v}=1 \\ &\boldsymbol{u}^{\mathrm{H}}\boldsymbol{u}_j=0, \quad j=1,2,\cdots,i-1 \\ &\boldsymbol{v}^{\mathrm{H}}\boldsymbol{v}_j=0, \quad j=1,2,\cdots,i-1 \end{aligned}\right\} \tag{8.5}$$

式中，$\boldsymbol{u}_i$和$\boldsymbol{v}_i$是ϕ_i达到第i个最大值时的向量$\boldsymbol{u}$和$\boldsymbol{v}$。

定义 8.5[5] 子空间H_1 与H_2之间的最小角度$\phi(H_1,H_2)$是介于0和π/2之间的角度，其余弦定义为

$$\cos\phi(H_1,H_2)\stackrel{\text{def}}{=}\max\left\{\left|\boldsymbol{u}^{\mathrm{H}}\boldsymbol{v}\right|:\boldsymbol{u}\in H_1,\ \boldsymbol{v}\in H_2,\ \|\boldsymbol{u}\|=1,\ \|\boldsymbol{v}\|=1\right\} \tag{8.6}$$

定义 8.6 若$\boldsymbol{A}=[\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n]\in C^{m\times n}$为复矩阵，则其列向量的所有线性组合的集合构成一个子空间，称为矩阵$\boldsymbol{A}$ 的列空间(column space)或列张成(column span)，用符号Col($\boldsymbol{A}$)表示，即有

$$\mathrm{Col}(\boldsymbol{A})=\mathrm{Span}\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\} \tag{8.7}$$

$$=\left\{\boldsymbol{y}\in C^m:\ \boldsymbol{y}=\sum_{j=1}^{n}\alpha_j\boldsymbol{a}_j:\ \alpha_j\in C\right\} \tag{8.8}$$

类似地，矩阵$\boldsymbol{A}$的复共轭行向量$\boldsymbol{r}_1^*,\boldsymbol{r}_2^*,\cdots,\boldsymbol{r}_m^*\in C^n$的所有线性组合的集合称为矩阵$\boldsymbol{A}$ 的行空间(row spacc)或行张成(row span)，用符号Row($\boldsymbol{A}$)表示，即有

$$\mathrm{Row}(\boldsymbol{A})=\mathrm{Span}\{\boldsymbol{r}_1^*,\boldsymbol{r}_2^*,\cdots,\boldsymbol{r}_m^*\} \tag{8.9}$$

$$=\left\{\boldsymbol{y}\in C^n:\ \boldsymbol{y}=\sum_{i=1}^{m}\beta_i\boldsymbol{r}_i^*:\ \beta_i\in C\right\} \tag{8.10}$$

定义 8.7 若$\boldsymbol{A}$是一个$m\times n$复矩阵，则$\boldsymbol{A}$的值域(range)定义为

$$\mathrm{Range}(\boldsymbol{A})=\{\boldsymbol{y}\in C^m:\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y},\quad \boldsymbol{x}\in C^n\} \tag{8.11}$$

矩阵$\boldsymbol{A}$的零空间(null space)也称$\boldsymbol{A}$的核(kernel)，定义为满足齐次线性方程$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$的所有解向量的集合，即

$$\mathrm{Null}(\boldsymbol{A})=\mathrm{Ker}(\boldsymbol{A})=\{\boldsymbol{x}\in C^n:\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}\} \tag{8.12}$$

类似地，复矩阵$\boldsymbol{A}_{m\times n}$的共轭转置$\boldsymbol{A}^{\mathrm{H}}$的零空间定义为

$$\mathrm{Null}(\boldsymbol{A}^{\mathrm{H}})=\mathrm{Ker}(\boldsymbol{A}^{\mathrm{H}})=\{\boldsymbol{x}\in C^m:\boldsymbol{A}^{\mathrm{H}}\boldsymbol{x}=\boldsymbol{0}\} \tag{8.13}$$

零空间的维数称为$\boldsymbol{A}$的零化维(nullity)，即有

$$\mathrm{nullity}(\boldsymbol{A})=\dim[\mathrm{Null}\,(\boldsymbol{A})] \tag{8.14}$$

定理 8.2 若$\boldsymbol{A}$是$m \times n$ 复矩阵，则$\boldsymbol{A}$的行空间的正交补$(\text{Row}(\boldsymbol{A}))^{\perp}$是$\boldsymbol{A}$的零空间，并且$\boldsymbol{A}$的列空间的正交补$(\text{Col}(\boldsymbol{A}))^{\perp}$是$\boldsymbol{A}^{\text{H}}$ 的零空间，即有

$$(\text{Row}(\boldsymbol{A}))^{\perp} = \text{Null}(\boldsymbol{A}), \qquad (\text{Col}(\boldsymbol{A}))^{\perp} = \text{Null}(\boldsymbol{A}^{\text{H}}) \tag{8.15}$$

矩阵$\boldsymbol{A}$的向量子空间之间的关系如下：

(1) 矩阵$\boldsymbol{A}$的值域与列空间相等，即

$$\text{Range}(\boldsymbol{A}) = \text{Col}(\boldsymbol{A}) = \text{Span}\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n\}$$

(2) 矩阵$\boldsymbol{A}$的行空间与$\boldsymbol{A}^{\text{H}}$的列空间相等，即

$$\text{Row}(\boldsymbol{A}) = \text{Col}(\boldsymbol{A}^{\text{H}}) = \text{Range}(\boldsymbol{A}^{\text{H}})$$

(3) 矩阵$\boldsymbol{A}$的行空间的正交补等于$\boldsymbol{A}$的零空间，即

$$(\text{Row}(\boldsymbol{A}))^{\perp} = \text{Null}(\boldsymbol{A})$$

(4) 矩阵$\boldsymbol{A}$的列空间的正交补就是$\boldsymbol{A}^{\text{H}}$的零空间，即

$$(\text{Col}(\boldsymbol{A}))^{\perp} = \text{Null}(\boldsymbol{A}^{\text{H}})$$

定理 8.3 初等行变换不改变一个矩阵的行空间。

定理 8.4 初等列变换不改变一个矩阵的列空间。

定理 8.5 初等行变换不改变矩阵$\boldsymbol{A}$的零空间$\text{Null}(\boldsymbol{A})$。

定理 8.6 初等列变换不改变矩阵$\boldsymbol{A}^{\text{H}}$的零空间$\text{Null}(\boldsymbol{A}^{\text{H}})$。

定理 8.7 矩阵$\boldsymbol{A}_{m\times n}$的列空间与行空间的维数相等。这个共同的维数就是矩阵$\boldsymbol{A}$的秩$\text{rank}(\boldsymbol{A})$，它与零空间维数之间有下列关系：

$$\text{rank}(\boldsymbol{A}) + \dim[\text{Null}(\boldsymbol{A})] = n \tag{8.16}$$

下面两个定理表明用于构造列空间的基向量。

定理 8.8 若$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$是一个满列秩矩阵$\boldsymbol{A} \in R^{m\times n}$的QR分解，并且$\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n]$ 和$\boldsymbol{Q} = [\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_m]$ 是列分块的，则

$$\text{Span}\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k\} = \text{Span}\{\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_k\}, \qquad k = 1, 2, \cdots, n$$

特别地，若$\boldsymbol{Q} = [\boldsymbol{Q}_1, \boldsymbol{Q}_2]$，其中，$\boldsymbol{Q}_1$是$\boldsymbol{Q}$的前$n$列组成的分块，$\boldsymbol{Q}_2$是$\boldsymbol{Q}$的其他列组成的分块，则

$$\text{Range}(\boldsymbol{A}) = \text{Range}(\boldsymbol{Q}_1), \quad (\text{Range}(\boldsymbol{A}))^{\perp} = \text{Range}(\boldsymbol{Q}_2)$$

并且$\boldsymbol{A} = \boldsymbol{Q}_1\boldsymbol{R}_1$，$\boldsymbol{R}_1 = \boldsymbol{R}(1:n, 1:n)$，即$\boldsymbol{R}_1$是$\boldsymbol{R}$的左上方$n \times n$方块。

定理 8.9 令秩$\text{rank}(\boldsymbol{A})=r$的矩阵$\boldsymbol{A}_{m\times n}$具有以下奇异值分解：

$$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}$$

式中

$$\boldsymbol{U}=[\boldsymbol{U}_r,\tilde{\boldsymbol{U}}_r],\quad \boldsymbol{V}=[\boldsymbol{V}_r,\tilde{\boldsymbol{V}}_r],\quad \boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_r & \boldsymbol{O}_{r\times(n-r)}\\ \boldsymbol{O}_{(m-r)\times(n-r)} & \boldsymbol{O}_{(n-r)\times(n-r)}\end{bmatrix}$$

这里，$\boldsymbol{U}_r$和$\tilde{\boldsymbol{U}}_r$分别为$m\times r$和$m\times(m-r)$矩阵，$\boldsymbol{V}_r$和$\tilde{\boldsymbol{V}}_r$分别为$n\times r$和$n\times(n-r)$矩阵，并且$\boldsymbol{\Sigma}=\text{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)$。则有以下结论:

(1) 与非零奇异值对应的r个左奇异向量$\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_r$是列空间$\text{Col}(\boldsymbol{A})$的标准正交基，即有

$$\text{Col}(\boldsymbol{A})=\text{Span}\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_r\} \tag{8.17}$$

(2) 与零奇异值对应的$m-r$个左奇异向量$\boldsymbol{u}_{r+1},\boldsymbol{u}_{r+2},\cdots,\boldsymbol{u}_m$是零空间$\text{Null}(\boldsymbol{A}^{\mathrm{H}})$的标准正交基，即

$$\text{Null}(\boldsymbol{A}^{\mathrm{H}})=(\text{Col}(\boldsymbol{A}))^{\perp}=\text{Span}\{\boldsymbol{u}_{r+1},\boldsymbol{u}_{r+2},\cdots,\boldsymbol{u}_m\} \tag{8.18}$$

(3) 与非零奇异值对应的r个右奇异向量$\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_r$是行空间$\text{Row}(\boldsymbol{A})$的标准正交基，即

$$\text{Row}(\boldsymbol{A})=\text{Span}\{\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_r\} \tag{8.19}$$

(4) 与零奇异值对应的$n-r$个右奇异向量$\boldsymbol{v}_{r+1},\boldsymbol{v}_{r+2},\cdots,\boldsymbol{v}_n$是零空间$\text{Null}(\boldsymbol{A})$的标准正交基，即

$$\text{Null}(\boldsymbol{A})=(\text{Row}(\boldsymbol{A}))^{\perp}=\text{Span}\{\boldsymbol{v}_{r+1},\boldsymbol{v}_{r+2},\cdots,\boldsymbol{v}_n\} \tag{8.20}$$

8.2 习题与解答

8.1 令V是所有2×2 矩阵的向量空间，证明子空间

$$W=\left\{\boldsymbol{A}:\boldsymbol{A}=\begin{bmatrix}a & b\\ c & d\end{bmatrix},\quad ad=0, bc=0\right\}$$

不是V的子空间。

证明： 验证一个集合是否是某个空间的子空间，一个必要条件就是集合对该空间定义的乘加运算封闭。以下证明集合W 对空间V 的加法运算不封闭。

令

$$\boldsymbol{A}_1=\begin{bmatrix}a_1 & b_1\\ c_1 & d_1\end{bmatrix},\qquad \boldsymbol{A}_2=\begin{bmatrix}a_2 & b_2\\ c_2 & d_2\end{bmatrix}$$

满足 $a_1d_1=0$, $b_1c_1=0$, $a_2d_2=0$, $b_2c_2=0$，则

$$\boldsymbol{A}_1+\boldsymbol{A}_2=\begin{bmatrix}a_1+a_2 & b_1+b_2\\ c_1+c_2 & d_1+d_2\end{bmatrix}$$

由 $a_1d_1=0$, $a_2d_2=0$ 无法推得 $(a_1+a_2)(d_1+d_2)=0$，即无法保证 $\boldsymbol{A}_1+\boldsymbol{A}_2\in\boldsymbol{W}$。所以集合 W 不是V 的子空间。

8.2 令V是2×2 矩阵的向量空间，并且

$$W=\left\{\boldsymbol{A}:\boldsymbol{A}=\begin{bmatrix}a & 0\\ 0 & b\end{bmatrix},\quad a,b\text{为任意实数}\right\}$$

是V的一个子空间。若

$$\boldsymbol{B}_1=\begin{bmatrix}0 & 2\\ 1 & 0\end{bmatrix},\quad \boldsymbol{B}_2=\begin{bmatrix}0 & 3\\ 0 & 0\end{bmatrix},\quad \boldsymbol{B}_3=\begin{bmatrix}0 & 3\\ 2 & 0\end{bmatrix}$$

(1) 证明矩阵集合$\{\boldsymbol{B}_1,\boldsymbol{B}_2,\boldsymbol{B}_3\}$线性相关，并将$\boldsymbol{B}_3$表示为$\boldsymbol{B}_1$ 和$\boldsymbol{B}_2$ 的线性组合；

(2) 证明$\{\boldsymbol{B}_1,\boldsymbol{B}_2\}$ 是一个线性无关的矩阵集合。

解： (1) 由观察不难有

$$\boldsymbol{B}_3=2\,\boldsymbol{B}_1-\frac{1}{3}\boldsymbol{B}_2$$

即线性相关得证。

(2) 假设

$$\lambda_1\boldsymbol{B}_1+\lambda_2\boldsymbol{B}_2=\begin{bmatrix}0 & 2\lambda_1+3\lambda_2\\ \lambda_1 & 0\end{bmatrix}=\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}$$

显然，当且仅当 $\lambda_1=0$, $\lambda_2=0$，即$\{\boldsymbol{B}_1,\boldsymbol{B}_2\}$为线性无关的矩阵集合时，第二个等式才能成立。

8.3　令$\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p$是有限维的非零向量空间$V$的向量，并且$S=\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p\}$为一向量集合。判断下列结果的真与假：

(1) $\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p$的所有线性组合的集合为一向量空间；

(2) 若$\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_{p-1}\}$线性无关，则$S$也是线性无关的向量集合；

(3) 若向量集合S线性无关，则S是向量空间V的一组基；

(4) 若$V=\mathrm{Span}\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p\}$，则$S$ 的某个子集是V的一组基；

(5) 若$\dim(V)=p$和$V=\mathrm{Span}\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p\}$，则向量集合$S$不可能线性相关。

解：

(1) 显然满足组成向量空间的所有条件。

(2) 显然不成立。

(3) 结论不成立。当且仅当$\boldsymbol{S}$ 的维数同$\boldsymbol{V}$ 的维数时成立。

(4) 结论成立。

(5) 结论成立。

8.4　判断下列结果是否为真：

(1) 矩阵$\boldsymbol{A}$的行空间与$\boldsymbol{A}^{\mathrm{T}}$的列空间相同。

(2) 矩阵$\boldsymbol{A}$的行空间和列空间的维数相同，即使$\boldsymbol{A}$不是正方矩阵。

(3) 矩阵$\boldsymbol{A}$的行空间和零空间的维数之和等于$\boldsymbol{A}$的行数。

(4) 矩阵$\boldsymbol{A}^{\mathrm{T}}$的行空间与$\boldsymbol{A}$的列空间相同。

解：

(1) 结论显然成立。

(2) 结论成立。该结论的另一种表达是矩阵的行秩和列秩相等。

(3) 结论不成立。

(4) 结论显然成立。

8.5　令$\boldsymbol{A}$ 为$m\times n$矩阵，在子空间$\mathrm{Row}(\boldsymbol{A})$, $\mathrm{Col}(\boldsymbol{A})$, $\mathrm{Null}(\boldsymbol{A})$, $\mathrm{Row}(\boldsymbol{A}^{\mathrm{T}})$, $\mathrm{Col}(\boldsymbol{A}^{\mathrm{T}})$和$\mathrm{Null}(\boldsymbol{A}^{\mathrm{T}})$内，有几个不同的子空间？哪些位于$R^m$空间，哪些位于$R^n$空间？

解：

$$\mathrm{Row}(\boldsymbol{A})=\mathrm{Col}(\boldsymbol{A}^{\mathrm{T}})\in\mathcal{R}^n$$

$$\mathrm{Row}(\boldsymbol{A}^{\mathrm{T}})=\mathrm{Col}(\boldsymbol{A})\in\mathcal{R}^m$$

$$\mathrm{Null}(\boldsymbol{A})\in\mathcal{R}^n,\quad \mathrm{Null}(\boldsymbol{A}^{\mathrm{T}})\in\mathcal{R}^m$$

8.6　证明下列向量集合W 为向量子空间，或举反例说明它不是向量子空间：

(1) $W=\left\{\begin{bmatrix}a\\b\\c\\d\end{bmatrix}: \begin{matrix}2a+b=c\\ a+b+c=d\end{matrix}\right\}$; (2) $W=\left\{\begin{bmatrix}a-b\\3b\\3a-2b\\a\end{bmatrix}: a,b\text{为实数}\right\}$

(3) $W=\left\{\begin{bmatrix}2a+3b\\c+a-2b\\4c+a\\3c-a-b\end{bmatrix}: a,b,c\text{为实数}\right\}$

证明：显然，以上三个向量集合都是向量空间，验证从略。

8.7 已知

$$\boldsymbol{A}=\begin{bmatrix}8&2&9\\-3&-2&-4\\5&0&5\end{bmatrix},\qquad \boldsymbol{w}=\begin{bmatrix}2\\1\\-2\end{bmatrix}$$

判断$\boldsymbol{w}$是在列空间$\mathrm{Col}(\boldsymbol{A})$ 还是零空间$\mathrm{Null}(\boldsymbol{A})$?

解：不难验证 $\mathrm{rank}(\boldsymbol{A})=2$，增广矩阵 $\mathrm{rank}([\boldsymbol{A}|\boldsymbol{w}])=3$。所以$\boldsymbol{w}$不可能由矩阵$\boldsymbol{A}$的列向量线性表示，即 $\boldsymbol{w}$不在列空间$\mathrm{Col}(\boldsymbol{A})$。

又由于$\boldsymbol{Aw}=\boldsymbol{0}$，所以$\boldsymbol{w}$在零空间$\mathrm{Null}(\boldsymbol{A})$中。

注释：$\boldsymbol{w}$并非不在$\mathrm{Col}(\boldsymbol{A})$就在零空间$\mathrm{Null}(\boldsymbol{A})$中。对于一般的矩阵$\boldsymbol{A}\in\mathcal{R}^{m\times n}$，$\mathrm{Col}(\boldsymbol{A})\in\mathcal{R}^m$，而$\mathrm{Null}(\boldsymbol{A})\in\mathcal{R}^n$。

8.8 一个7×10矩阵能否有二维的零空间?

解：由教材中式(8.2.16)有，矩阵的秩和其零空间的维数之和等于10。若该矩阵的零空间的维数为2，则矩阵的秩等于8，显然不成立。则该矩阵零空间的维数不可能等于2。

8.9 试证明$\boldsymbol{v}$在矩阵$\boldsymbol{A}$的列空间$\mathrm{Col}(\boldsymbol{A})$内，若$\boldsymbol{Av}=\lambda\boldsymbol{v}$，且$\lambda\neq 0$。

解：由

$$\boldsymbol{v}=\frac{1}{\lambda}\boldsymbol{Av}$$

有，$\boldsymbol{v}$可以表示成$\boldsymbol{A}$的列向量的线性组合。即$\boldsymbol{v}$在矩阵$\boldsymbol{A}$的列空间 $\mathrm{Col}(\boldsymbol{A})$。

8.10 令$\boldsymbol{V}_1$和$\boldsymbol{V}_2$的列向量分别是C^n的同一子空间的正交基，证明对于C^n空间中任意n维向量$\boldsymbol{x}$，$\boldsymbol{V}_1\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{x}=\boldsymbol{V}_2\boldsymbol{V}_2^{\mathrm{H}}\boldsymbol{x}$成立。

证明：由于$\boldsymbol{V}_1$和$\boldsymbol{V}_2$的列向量是C^n的同一子空间的正交基，则必存在一个正交矩阵$\boldsymbol{P}\in C^n$使得

$$\boldsymbol{V}_1=\boldsymbol{V}_2\boldsymbol{P}$$

则$\forall\,\boldsymbol{x}$

$$\boldsymbol{V}_1\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{x}=\boldsymbol{V}_2\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{V}_2^{\mathrm{H}}\boldsymbol{x}=\boldsymbol{V}_2\boldsymbol{V}_2^{\mathrm{H}}\boldsymbol{x}$$

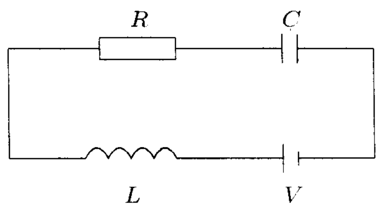
8.11 令V是一子空间，且S是V的生成元或张成集合。已知

$$S=\left\{\begin{bmatrix}1\\0\\-1\end{bmatrix},\begin{bmatrix}-1\\-2\\-3\end{bmatrix},\begin{bmatrix}1\\1\\2\end{bmatrix},\begin{bmatrix}2\\-1\\0\end{bmatrix}\right\}$$

求V的基，并计算$\dim(V)$。

解： 不难验证，$\dim(\boldsymbol{V})=3$，且任意三个列向量都可以做$\boldsymbol{V}$的基。

8.12 题8.12图中的电路由电阻R (Ω)、电感L (H)和电容C (F) 和初始电压源V 组成。令$b=R/(2L)$，并假定R,L,C 的值使得b 的数值也等于$1/\sqrt{LC}$ (例如，伏特计就是这种情况)。令$v(t)$ 是在时间t 测得的电容两端的瞬时电压，而H是将$v(t)$ 映射为$Lv''(t)+Rv'(t)+(1/C)v(t)$ 的线性变换的零空间。可以证明，v位于零空间H内，并且H由所有具有形式$v(t)=\mathrm{e}^{-bt}(c_1+c_2t)$的函数组成。求零空间$H$的一组基。



题8.12图 电路图

解： 解一元二次线性微分方程：

$$Lv''(t)+Rv'(t)+(1/C)v(t)=0$$

其通解为

$$v(t)=\mathrm{e}^{-bt}(c_1+c_2t)$$

其中$b=\dfrac{1}{\sqrt{LC}}$。可见H中的所有元素可以用$u_1(t)=\mathrm{e}^{-bt}$ 与$u_2(t)=t\mathrm{e}^{-bt}$ 线性表出。下面证明$u_1(t)$与$u_2(t)$线性无关。假设$u_1(t)$与$u_2(t)$线性相关，则存在不全为零的常数a_1,a_2,使得

$$a_1\mathrm{e}^{-bt}+a_2t\mathrm{e}^{-bt}=0$$

因为上式对任意t成立，故有

$$a_1\mathrm{e}^{-b(t+1)}+a_2(t+1)\mathrm{e}^{-b(t+1)}=0$$

写成矩阵形式如下：

$$\begin{bmatrix}\mathrm{e}^{-bt} & t\mathrm{e}^{-bt}\\ \mathrm{e}^{-b(t+1)} & (t+1)\mathrm{e}^{-b(t+1)}\end{bmatrix}\begin{bmatrix}a_1\\a_2\end{bmatrix}=0$$

不妨设$t=0$，则方程组变为

$$\begin{bmatrix} 1 & 0 \\ e^{-b} & e^{-b} \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}=0$$

由于

$$\begin{vmatrix} 1 & 0 \\ e^{-b} & e^{-b} \end{vmatrix} \neq 0$$

所以，方程组只有零解。因此假设不成立，即$u_1(t)$与$u_2(t)$线性无关。即有结论：$u_1(t)=e^{-bt}$与$u_2(t)=te^{-bt}$ 是H的一组基。

8.13 一质量为m的物体挂在一弹簧的末端。如果压紧该弹簧，然后再释放，这一质量-弹簧系统就会开始振荡。假定质量m与其静止位置的位移$y(t)$由函数

$$y(t)=c_1\cos(\omega t)+c_2\sin(\omega t)$$

描述，其中，ω是一个与质量m和弹簧有关的常数。固定ω，令c_1和c_2任意。

(1) 证明：描述质量-弹簧系统振荡的函数$y(t)$的集合为一向量空间V。

(2) 求向量空间V的一组基。

证明： 定义

$$\boldsymbol{V}=\{y(t)|y(t)=c_1\cos(\omega t)+c_2\sin(\omega t),\quad \forall c_1,c_2\in R\}$$

任取$y_1(t)$，$y_2(t)$如下：

$$y_1(t)=c_{11}\cos(\omega t)+c_{12}\sin(\omega t)$$
$$y_2(t)=c_{21}\cos(\omega t)+c_{22}\sin(\omega t)$$

则有

$$y_1(t)+y_2(t)=(c_{11}+c_{21})\cos(\omega t)+(c_{12}+c_{22})\sin(\omega t)\in \boldsymbol{V}$$

同时对于任意$c\in R$有

$$c\,y(t)=c\,c_1\cos(\omega t)+c\,c_2\sin(\omega t)\in \boldsymbol{V}$$

所以$\boldsymbol{V}$是一个线性空间。由于$\boldsymbol{V}$中的任一元素都可以由$\cos(\omega t)$，$\sin(\omega t)$线性表示。所以可以猜测$\{\cos(\omega t)，\sin(\omega t)\}$是线性空间$\boldsymbol{V}$的一组基。其线性无关性很容易验证，若$\cos(\omega t)$和$\sin(\omega t)$线性相关，则必存在一常数$c$，$\forall t$使得$\cos(\omega t)=c\sin(\omega t)$，则显然是不成立的，即$\cos(\omega t)$和$\sin(\omega t)$线性无关。所以$\{\cos(\omega t),\sin(\omega t)\}$是线性空间$\boldsymbol{V}$的一组基。

8.14 令

$$W=\left\{\begin{bmatrix} a \\ b \\ c \end{bmatrix}: a-3b-c=0\right\}$$

证明W是R^3的一个子空间。

解： 令

$$\boldsymbol{w}_1 = \begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix} \qquad \boldsymbol{w}_2 = \begin{bmatrix} a_2 \\ b_2 \\ c_2 \end{bmatrix}$$

$$\lambda_1 \boldsymbol{w}_1 + \lambda_2 \boldsymbol{w}_2 = \begin{bmatrix} \lambda_1 a_1 + \lambda_2 a_2 \\ \lambda_1 b_1 + \lambda_2 b_2 \\ \lambda_1 c_1 + \lambda_2 c_2 \end{bmatrix}$$

由于

$$(\lambda_1 a_1 + \lambda_2 a_2) - 3(\lambda_1 b_1 + \lambda_2 b_2) - (\lambda_1 c_1 + \lambda_2 c_2) = \lambda_1(a_1 - 3b_1 - c_1) + \lambda_2(a_2 - 3b_2 - c_2) = 0$$

所以$\lambda_1 \boldsymbol{w}_1 + \lambda_2 \boldsymbol{w}_2$在集合$\boldsymbol{W}$中。即$\boldsymbol{W}$是$R^3$的一个子空间。

8.15 令$\boldsymbol{A}$ 和$\boldsymbol{B}$是两个$m \times n$ 矩阵，并且$m \geqslant n$。证明

$$\min_{\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}=\boldsymbol{I}_n} \|\boldsymbol{A} - \boldsymbol{B}\boldsymbol{Q}\|_{\mathrm{F}}^2 = \sum_{i=1}^{n} [(\sigma_i^2(\boldsymbol{A}))^2 - 2\sigma_i(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}) + (\sigma_i(\boldsymbol{B}))^2]$$

式中，$\sigma_i(\boldsymbol{A})$是矩阵$\boldsymbol{A}$的第i个奇异值。

证明：

$$\begin{aligned} \|\boldsymbol{A} - \boldsymbol{B}\boldsymbol{Q}\|_{\mathrm{F}}^2 &= \mathrm{tr}\left((\boldsymbol{A} - \boldsymbol{B}\boldsymbol{Q})^{\mathrm{T}}(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{Q})\right) \\ &= \mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}) - 2\mathrm{tr}\left(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}\right) \\ &= \sum_{i=1}^{n}(\sigma_i^2(\boldsymbol{A}))^2 + \sum_{i=1}^{n}(\sigma_i^2(\boldsymbol{B}))^2 - 2\mathrm{tr}\left(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}\right) \end{aligned}$$

$$\min \|\boldsymbol{A} - \boldsymbol{B}\boldsymbol{Q}\|_{\mathrm{F}}^2 \quad \Longleftrightarrow \quad \max \mathrm{tr}\left(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}\right)$$

使$\mathrm{tr}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A})$ 最大的$\boldsymbol{Q}$ 可通过$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}$ 的奇异值分解得到。令

$$\boldsymbol{U}^{\mathrm{T}}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A})\boldsymbol{V} = \boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \cdots, \sigma_n)$$

定义一正交矩阵$\boldsymbol{W} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{U}$，则

$$\mathrm{tr}\left(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}\right) = \mathrm{tr}\left(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}\right) = \mathrm{tr}(\boldsymbol{W}\boldsymbol{\Sigma}) = \sum_{i=1}^{n} w_{ii}\sigma_i \leqslant \sum_{i=1}^{n}\sigma_i$$

等号当$\boldsymbol{W} = \boldsymbol{I}$ 时达到，此时$\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}$。所以有结论

$$\min_{\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}=\boldsymbol{I}_n} \|\boldsymbol{A} - \boldsymbol{B}\boldsymbol{Q}\|_{\mathrm{F}}^2 = \sum_{i=1}^{n} [(\sigma_i^2(\boldsymbol{A}))^2 - 2\sigma_i(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}) + (\sigma_i(\boldsymbol{B}))^2]$$

8.16 假定T是一个一对一线性变换，并且$T(\boldsymbol{u}) = T(\boldsymbol{v})$ 总是意味着$\boldsymbol{u} = \boldsymbol{v}$。证明：若像的集合$\{T(\boldsymbol{u}_1), T(\boldsymbol{u}_2), \cdots, T(\boldsymbol{u}_p)\}$线性相关，则向量集合$\{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_p\}$线性相关。

(注：该命题表明，一个一对一线性变换将线性无关的向量集合映射为线性无关的向量集合。)

解：若集合 $T(\boldsymbol{u}_1), T(\boldsymbol{u}_2), T(\boldsymbol{u}_p)$线性相关。不妨假设

$$\lambda_1 T(\boldsymbol{u}_1) + \lambda_2 T(\boldsymbol{u}_2) + \cdots + \lambda_p T(\boldsymbol{u}_p) = 0$$

由于 T 是线性变换。则有

$$T(\lambda_1 \boldsymbol{u}_1 + \lambda_2 \boldsymbol{u}_2 + \cdots + \lambda_p \boldsymbol{u}_p) = 0$$

由于 T 是一一变换，则有

$$\lambda_1 \boldsymbol{u}_1 + \lambda_2 \boldsymbol{u}_2 + \cdots + \lambda_p \boldsymbol{u}_p = 0$$

像线性相关,则原像也线性相关。即证明了其逆否命题：一一映射线性变换将线性无关的向量集合映射成线性无关的向量集。

8.17 已知矩阵

$$\boldsymbol{A} = \begin{bmatrix} 2 & 5 & -8 & 0 & 17 \\ 1 & 3 & -5 & 1 & 5 \\ -3 & -11 & 19 & -7 & -1 \\ 1 & 7 & -13 & 5 & -3 \end{bmatrix}$$

试求其列空间、行空间和零空间的基。

解：$\operatorname{rank}(A) = 3$

不难验证，行空间的基有124，134，234。

列空间的基有124，125，134，135，234，235，345。

8.18 已知

$$\boldsymbol{A} = \begin{bmatrix} -2 & 1 & -1 & 6 & -8 \\ 1 & -2 & -4 & 3 & -2 \\ 7 & -8 & -10 & -3 & 10 \\ 4 & -5 & -7 & 0 & 4 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 1 & -2 & -4 & 3 & -2 \\ 0 & 3 & 9 & -12 & 12 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

是两个行等价的矩阵，试求

(1) 矩阵$\boldsymbol{A}$的秩和零空间Null($\boldsymbol{A}$)的维数；

(2) 列空间Col($\boldsymbol{A}$)和行空间Row($\boldsymbol{A}$)的基；

(3) 如果希望求零空间Null($\boldsymbol{A}$)的基，下一步应该执行什么运算？

解：

(1) 因为$\boldsymbol{A}$与$\boldsymbol{B}$行等价，则由$\boldsymbol{B}$有$\operatorname{rank}(\boldsymbol{A}) = 2$，$\dim(\operatorname{Null}(\boldsymbol{A})) = 5 - 2 = 3$。

(2) $\boldsymbol{B}$的第一二两列组成Col($\boldsymbol{A}$)的基。$\boldsymbol{B}$的第一二两行组成Row($\boldsymbol{A}$)的基。

(3) 求Null($\boldsymbol{A}$)的基，等价于求解方程$\boldsymbol{Ax}=\boldsymbol{0}$。不难求得Null($\boldsymbol{A}$)的基为

$$\left\{\begin{bmatrix}-2\\-3\\1\\0\\0\end{bmatrix},\begin{bmatrix}5\\4\\0\\1\\0\end{bmatrix},\begin{bmatrix}-6\\-4\\0\\0\\1\end{bmatrix}\right\}$$

8.19 令$\boldsymbol{A}$ 是一个$n\times n$对称矩阵，证明：

(1) $(\mathrm{Col}\boldsymbol{A})^{\perp}=\mathrm{Null}(\boldsymbol{A})$。

(2) R^n内的每一个向量$\boldsymbol{x}$ 都可以写作$\boldsymbol{x}=\hat{\boldsymbol{x}}+\boldsymbol{z}$，式中，$\hat{\boldsymbol{x}}\in\mathrm{Col}(\boldsymbol{A})$，$\boldsymbol{z}\in\mathrm{Null}(\boldsymbol{A})$。

解：(1) $\forall\boldsymbol{z}\in(\mathrm{Col}\boldsymbol{A})^{\perp}$必满足

$$\boldsymbol{z}^{\mathrm{T}}\boldsymbol{Ax}=0,\qquad\forall\boldsymbol{x}\in\mathcal{R}^n$$

则必有$\boldsymbol{z}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{0}^{\mathrm{T}}$；又由于$\boldsymbol{A}$是对称矩阵，则有$\boldsymbol{Az}=\boldsymbol{0}$。即有$\boldsymbol{z}\in\mathrm{Null}(\boldsymbol{A})$，所以$(\mathrm{Col}\boldsymbol{A})^{\perp}\subset\mathrm{Null}(\boldsymbol{A})$。

$\forall\boldsymbol{z}\in\mathrm{Null}(\boldsymbol{A})$ 有$\boldsymbol{Az}=\boldsymbol{0}$。又由$\boldsymbol{A}$的对称性有$\boldsymbol{z}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{0}$，于是也有

$$\boldsymbol{z}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{0},\qquad\forall\boldsymbol{x}\in\mathcal{R}^n$$

即有，$\boldsymbol{z}\in(\mathrm{Col}\boldsymbol{A})^{\perp}$，所以$\mathrm{Null}(\boldsymbol{A})\subset(\mathrm{Col}\boldsymbol{A})^{\perp}$。

综上则有$\mathrm{Null}(\boldsymbol{A})=(\mathrm{Col}\boldsymbol{A})^{\perp}$。

(2) 由于

$$\mathrm{Col}\boldsymbol{A}\oplus(\mathrm{Col}\boldsymbol{A})^{\perp}=\mathcal{R}^n$$

所以，$\boldsymbol{x}$可以唯一表示为

$$\boldsymbol{x}=\hat{\boldsymbol{x}}+\boldsymbol{z}$$

其中$\hat{\boldsymbol{x}}\in\mathrm{Col}\boldsymbol{A}$，$\boldsymbol{z}\in(\mathrm{Col}\boldsymbol{A})^{\perp}$。又由(1)问的结论有$\boldsymbol{z}\in\mathrm{Null}(\boldsymbol{A})$。

8.20 考虑一码分多址(CDMA)系统，它共有K个用户。假定用户1为期望用户，其特征波形向量$\boldsymbol{s}_1$为已知，并满足单位能量条件$\langle\boldsymbol{s}_1,\boldsymbol{s}_1\rangle=\boldsymbol{s}_1^{\mathrm{T}}\boldsymbol{s}_1=1$。现有一接收机的观测数据向量为$\boldsymbol{y}(n)$，它包含了$K$个用户信号的线性混合。为了检测期望用户的信号，希望设计一多用户检测器$\boldsymbol{c}_1$，使检测器的输出能量最小化。若多用户检测器服从约束条件$\boldsymbol{c}_1=\boldsymbol{s}_1+\boldsymbol{U}_i\boldsymbol{w}$，其中，$\boldsymbol{U}_i$称为干扰子空间，意即它的列张成干扰子空间。

(1) 求线性检测器$\boldsymbol{c}_1$的LMS自适应算法。

(2) 求干扰子空间$\boldsymbol{U}_i$。

解：(1) 由题意，多用户检测器$\boldsymbol{c}_1$的设计实际是一个带约束条件的最小输出能量问题：

$$\boldsymbol{c}_1=\arg\min_{c_1}\mathrm{E}\{[\boldsymbol{c}_1^{\mathrm{T}}\boldsymbol{y}(n)]^2\}$$

约束条件为

$$c_1 = s_1 + U_i w$$

式中$\mathrm{E}\{[c_1^{\mathrm{T}} y(n)]^2\}$表示多用户检测器输出的能量。

由于相对于期望用户而言，s_1代表了信号子空间，U_i为干扰子空间，所以它们之间存在正交关系，即有$s_1^{\mathrm{T}} U_i = 0$。这表明,约束条件式表示的分解为正交分解。由梯度下降法，有

$$w(n+1) = w(n) - \mu \nabla_w \mathrm{E}\{[c_1^{\mathrm{T}}(n) y(n)]^2\}$$

式中$\nabla_w \mathrm{E}\{[c_1^{\mathrm{T}}(n) y(n)]^2\}$表示真实梯度，即

$$\begin{aligned}
\nabla_w \mathrm{E}\{[c_1^{\mathrm{T}}(n) y(n)]^2\} &= \frac{\partial}{\partial w^{\mathrm{T}}(n)} \mathrm{E}\{[c_1^{\mathrm{T}}(n) y(n)]^2\} \\
&= \mathrm{E}\left\{2 c_1^{\mathrm{T}}(n) y(n) \cdot \frac{\partial c_1^{\mathrm{T}}(n) y(n)}{w^{\mathrm{T}}(n)}\right\} \\
&= 2\mathrm{E}\{[s_1 + U_i w(n)]^{\mathrm{T}} y(n) \cdot U_i^{\mathrm{T}} y(n)\}
\end{aligned}$$

用瞬时值代替期望值，并令$z(n) = [s_1 + U_i w(n)]^{\mathrm{T}} y(n)$，即得瞬时梯度

$$\hat{\nabla}_w = 2[s_1 + U_i w(n)]^{\mathrm{T}} y(n) U_i^{\mathrm{T}} y(n) = 2 z(n) U_i^{\mathrm{T}} y(n)$$

于是，LMS自适应算法为

$$w(n+1) = w(n) - 2\mu z(n) U_i^{\mathrm{T}} y(n) \tag{8.21}$$

而多用户检测器c_1的自适应更新公式为

$$c_1(n+1) = s_1 + U_i w(n+1) \tag{8.22}$$

式(8.21)和式(8.22) 一起组成了多用户检测器的LMS自适应算法。

(2) 令$K \times K$矩阵

$$R_y \stackrel{\mathrm{def}}{=} \mathrm{E}\{y(n) y^{\mathrm{T}}(n)\}$$

表示观测数据向量$y(n)$的自相关矩阵，其特征值分解为

$$R_y = U \Sigma U^{\mathrm{T}}$$

则矩阵$U = [s_1, U_i]$的列张成数据空间，其中s_1张成信号子空间，而U_i张成干扰子空间。事实上，$P_{s_1} y(n)$表示观测数据向量$y(n)$在信号子空间的投影，因此，$y(n) - P_{s_1} y(n)$代表观测数据向量中的全部干扰部分。换言之，干扰向量可表示为

$$\begin{aligned}
i(n) &= y(n) - P_{s_1} y(n) \\
&= y(n) - s_1 \langle s_1, s_1 \rangle^{-1} s_1^{\mathrm{T}} y(n) \\
&= y(n) - s_1 s_1^{\mathrm{T}} y(n)
\end{aligned}$$

式中使用了单位能量条件$\langle \boldsymbol{s}_1, \boldsymbol{s}_1 \rangle = 1$。

注意到标量$\boldsymbol{s}_1^{\mathrm{T}}\boldsymbol{y}(n) = \boldsymbol{y}^{\mathrm{T}}(n)\boldsymbol{s}_1$，所以干扰向量可写为

$$\boldsymbol{i}(n) = \boldsymbol{y}(n) - (\boldsymbol{y}^{\mathrm{T}}(n)\boldsymbol{s}_1)\boldsymbol{s}_1$$

一旦干扰向量计算出后，即可计算其样本自相关矩阵

$$\boldsymbol{R}_i = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{i}(n)\boldsymbol{i}^{\mathrm{T}}(n)$$

注意$K \times K$维自相关矩阵$\boldsymbol{R}_y$是满秩的，即其秩为K，而$\boldsymbol{R}_i$的秩则为$K-1$，因为干扰信号部分被去除了。

第9章 投影分析

投影分为正交投影和斜投影两类。当两个子空间正交时，常采用正交投影进行参数的最优估计或者信号的最优滤波，因为观测数据向量在被投影的子空间上的分量可以被抽取，而且理论上还可以完全对消掉观测数据向量在另外一个子空间的分量。正交投影已成为最小二乘参数估计[4]、阵列信号处理[33]、自适应滤波[32]中的有力工具。然而，若两个子空间不正交，则正交投影不再有效，必须使用斜投影，才能达到抽取观测数据向量在一个子空间上的分量，而完全抑制它在另外一个子空间上的分量之目的。斜投影算子的一个重要应用是在CDMA多用户检测中的解相关检测（见习题9.17）。实际上还有很多算法从几何上看本质是一种斜投影，只是没有被发现罢了。

9.1 主要理论与方法

9.1.1 幂等矩阵的性质

(1) 幂等矩阵的特征值只取1和0两个数值。

(2) 所有的幂等矩阵(单位矩阵除外) $\boldsymbol{A}$ 都是奇异矩阵。

(3) 所有幂等矩阵的秩与迹相等，即$\text{rank}(\boldsymbol{A}) = \text{tr}(\boldsymbol{A})$。

(4) 若$\boldsymbol{A}$为幂等矩阵，则$\boldsymbol{A}^{\text{H}}$也为幂等矩阵，即有$\boldsymbol{A}^{\text{H}}\boldsymbol{A}^{\text{H}} = \boldsymbol{A}^{\text{H}}$。

(5) 若$\boldsymbol{A}$为幂等矩阵，则$\boldsymbol{I}_n - \boldsymbol{A}$也是幂等矩阵，且$\text{rank}(\boldsymbol{I}_n - \boldsymbol{A}) = n - \text{rank}(\boldsymbol{A})$。

(6) 所有对称的幂等矩阵(单位矩阵除外)都是半正定的。

(7) 令$n \times n$幂等矩阵$\boldsymbol{A}$的秩为r_A，则$\boldsymbol{A}$有r_A个特征值1和$n - r_A$个特征值0。

(8) 所有的幂等矩阵$\boldsymbol{A}$都是可对角化的：

$$\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{I}_{r_A} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix} \tag{9.1}$$

式中，$r_A = \text{rank}(\boldsymbol{A})$。

(9) 一个对称的幂等矩阵$\boldsymbol{A}$可以表示为$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^{\text{T}}$，其中，$\boldsymbol{L}$满足$\boldsymbol{L}^{\text{T}}\boldsymbol{L} = \boldsymbol{I}_{R_A}$。

定理 9.1 线性齐次算子$\boldsymbol{P}$ 是投影算子，当且仅当$\boldsymbol{P}$是幂等矩阵，即

$$\boldsymbol{P}^2 = \boldsymbol{P} \tag{9.2}$$

定理 9.2 若$\boldsymbol{F} \in \boldsymbol{C}^{n\times l}$，$\boldsymbol{G} \in \boldsymbol{C}^{n\times m}$，则$\text{Range}(\boldsymbol{F}) \perp \text{Range}(\boldsymbol{G})$ 当且仅当 $\boldsymbol{F}^{\text{H}}\boldsymbol{G} = \boldsymbol{0}$。

定理 9.3 $\text{Range}(\boldsymbol{A}) = \text{Range}(\boldsymbol{A}\boldsymbol{A}^{\text{H}})$，$\text{Null}(\boldsymbol{A}) = \text{Null}(\boldsymbol{A}^{\text{H}}\boldsymbol{A})$。

9.1.2 投影的性质

令$\boldsymbol{H}$是向量空间，而$\boldsymbol{M}$是$\boldsymbol{H}$的n维子空间。$\boldsymbol{P}_M$代表闭子空间$\boldsymbol{M}$的投影映射算子。则有

(1) $\boldsymbol{P}_M(\alpha\boldsymbol{x}+\beta\boldsymbol{y})=\alpha\boldsymbol{P}_M\boldsymbol{x}+\beta\boldsymbol{P}_M\boldsymbol{y},\quad \boldsymbol{x},\boldsymbol{y}\in H;\quad \alpha,\beta\in C$。

(2) $\|\boldsymbol{x}\|^2=\|\boldsymbol{P}_M\boldsymbol{x}\|^2+\|(\boldsymbol{I}-\boldsymbol{P}_M)\boldsymbol{x}\|^2$。

(3) 每一个$\boldsymbol{x}\in M$都具有以下的唯一表示：

$$\boldsymbol{x}=\boldsymbol{P}_M\boldsymbol{x}+(\boldsymbol{I}-\boldsymbol{P}_M)\boldsymbol{x}\quad (\text{正交分解}) \tag{9.3}$$

即$\boldsymbol{x}$可以唯一分解成M的元素与$M^{\perp}$的元素之和。

(4) $\boldsymbol{P}_M\boldsymbol{x}_n\to\boldsymbol{P}_M\boldsymbol{x}$，若$\|\boldsymbol{x}_n-\boldsymbol{x}\|\to 0$。

(5) $\boldsymbol{x}\in M$当且仅当$\boldsymbol{P}_M\boldsymbol{x}=\boldsymbol{x}$。

(6) $\boldsymbol{x}\in M^{\perp}$当且仅当$\boldsymbol{P}_M\boldsymbol{x}=\boldsymbol{0}$。

(7) $M_1\subseteq M_2$，当且仅当$\boldsymbol{P}_{M_1}\boldsymbol{P}_{M_2}\boldsymbol{x}=\boldsymbol{P}_{M_1}\boldsymbol{x}$对所有$\boldsymbol{x}\in H$恒成立。

定理 9.4 (空间的投影定理) 若数据向量$\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\cdots,\boldsymbol{\eta}_n$组成标准正交系，则随机变量$\boldsymbol{\xi}$的最佳均方估计由

$$\hat{\boldsymbol{\xi}}=\sum_{i=1}^{n}\langle\boldsymbol{\xi},\boldsymbol{\eta}_i\rangle\boldsymbol{\eta}_i \tag{9.4}$$

确定。

定理 9.5 若$M\times N$ $(M\geqslant N)$ 矩阵$\boldsymbol{A}$满列秩，则投影矩阵可以表示为

$$\boldsymbol{P}_{\boldsymbol{A}}=\boldsymbol{A}\langle\boldsymbol{A},\boldsymbol{A}\rangle^{-1}\boldsymbol{A}^{\mathrm{H}} \tag{9.5}$$

定理 9.6 令数据空间为$\{\boldsymbol{U}\}$，其相对应的投影矩阵是$\boldsymbol{P}_U$，正交投影矩阵为$\boldsymbol{P}_U^{\perp}$。现加入一个新的数据向量$\boldsymbol{u}$到$\{\boldsymbol{U}\}$的原向量组中。则$\{\boldsymbol{U},\boldsymbol{u}\}$空间的投影矩阵$\boldsymbol{P}_{U,u}$和正交投影矩阵$\boldsymbol{P}_{U,u}^{\perp}$可以表示为

$$\boldsymbol{P}_{U,u}=\boldsymbol{P}_U+\boldsymbol{P}_U^{\perp}\boldsymbol{u}\langle\boldsymbol{P}_U^{\perp}\boldsymbol{u},\boldsymbol{P}_U^{\perp}\boldsymbol{u}\rangle^{-1}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{P}_U^{\perp} \tag{9.6}$$

$$\boldsymbol{P}_{U,u}^{\perp}=\boldsymbol{P}_U^{\perp}-\boldsymbol{P}_U^{\perp}\boldsymbol{u}\langle\boldsymbol{P}_U^{\perp}\boldsymbol{u},\boldsymbol{P}_U^{\perp}\boldsymbol{u}\rangle^{-1}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{P}_U^{\perp} \tag{9.7}$$

实际上任何一个向量的投影和正交投影也有同样的更新公式

$$\boldsymbol{P}_{U,u}\boldsymbol{y}=\boldsymbol{P}_U\boldsymbol{y}+\boldsymbol{P}_U^{\perp}\boldsymbol{u}\langle\boldsymbol{P}_U^{\perp}\boldsymbol{u},\boldsymbol{P}_U^{\perp}\boldsymbol{u}\rangle^{-1}\langle\boldsymbol{u},\boldsymbol{P}_U^{\perp}\boldsymbol{y}\rangle \tag{9.8}$$

$$\boldsymbol{P}_{U,u}^{\perp}\boldsymbol{y}=\boldsymbol{P}_U^{\perp}\boldsymbol{y}-\boldsymbol{P}_U^{\perp}\boldsymbol{u}\langle\boldsymbol{P}_U^{\perp}\boldsymbol{u},\boldsymbol{P}_U^{\perp}\boldsymbol{u}\rangle^{-1}\langle\boldsymbol{u},\boldsymbol{P}_U^{\perp}\boldsymbol{y}\rangle \tag{9.9}$$

9.1.3 斜投影算子

考虑$n \times m$满列秩矩阵$\boldsymbol{H}$和$n \times k$满列秩矩阵$\boldsymbol{S}$组合成一个$n \times (m+k)$ 矩阵$[\boldsymbol{H}, \boldsymbol{S}]$，其中，$m+k<n$，使得矩阵$[\boldsymbol{H}, \boldsymbol{S}]$的列秩小于行数$n$，并且$\boldsymbol{H}$的列向量与$\boldsymbol{S}$的列向量线性无关。记$\boldsymbol{E}_{H|S}$是沿着值域空间Range($\boldsymbol{S}$)，到另一值域空间Range($\boldsymbol{H}$)的投影算子。$\boldsymbol{E}_{S|H}$是沿着值域空间Range($\boldsymbol{H}$)，到值域空间Range($\boldsymbol{S}$)的投影算子。它们可以分别表示为

$$\boldsymbol{E}_{H|S} = \boldsymbol{H}(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{P}_S^{\perp}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{P}_S^{\perp} \tag{9.10}$$

$$\boldsymbol{E}_{S|H} = \boldsymbol{S}(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{P}_H^{\perp}\boldsymbol{S})^{-1}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{P}_H^{\perp} \tag{9.11}$$

以上两个斜投影算子的性质总结如下：

(1) $\boldsymbol{E}_{H|S}$和$\boldsymbol{E}_{S|H}$均为幂等算子，即有

$$\boldsymbol{E}_{H|S}^2 = \boldsymbol{E}_{H|S}, \qquad \boldsymbol{E}_{S|H}^2 = \boldsymbol{E}_{S|H}$$

(2) $\boldsymbol{E}_{H|S}[\boldsymbol{H}, \boldsymbol{S}] = [\boldsymbol{H}, \boldsymbol{O}]$和$\boldsymbol{E}_{S|H}[\boldsymbol{H}, \boldsymbol{S}] = [\boldsymbol{O}, \boldsymbol{S}]$，或者等价为

$$\begin{aligned} \boldsymbol{E}_{H|S}\boldsymbol{H} = \boldsymbol{H}, &\qquad \boldsymbol{E}_{H|S}\boldsymbol{S} = \boldsymbol{O} \\ \boldsymbol{E}_{S|H}\boldsymbol{H} = \boldsymbol{O}, &\qquad \boldsymbol{E}_{S|H}\boldsymbol{S} = \boldsymbol{S} \end{aligned}$$

(3) 斜投影算子$\boldsymbol{E}_{H|S}$和$\boldsymbol{E}_{S|H}$的交叉项为零，即有

$$\boldsymbol{E}_{H|S}\boldsymbol{E}_{S|H} = \boldsymbol{O}, \qquad \boldsymbol{E}_{S|H}\boldsymbol{E}_{H|S} = \boldsymbol{O}$$

(4) 斜投影后，再正交投影，不会改变原斜投影：

$$\boldsymbol{E}_{H|S} = \boldsymbol{P}_H\boldsymbol{E}_{H|S}, \qquad \boldsymbol{E}_{S|H} = \boldsymbol{P}_S\boldsymbol{E}_{S|H}$$

(5) 令$\boldsymbol{B}^{\dagger} = (\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}}$ 表示矩阵$\boldsymbol{B}$的广义逆矩阵，则

$$\boldsymbol{H}^{\dagger}\boldsymbol{E}_{H|S} = (\boldsymbol{P}_S^{\perp}\boldsymbol{H})^{\dagger}, \qquad \boldsymbol{S}^{\dagger}\boldsymbol{E}_{S|H} = (\boldsymbol{P}_H^{\perp}\boldsymbol{S})^{\dagger}$$

(6) 斜投影矩阵与正交投影矩阵的关系：

$$\boldsymbol{P}_S^{\perp}\boldsymbol{E}_{H|S}\boldsymbol{P}_S^{\perp} = \boldsymbol{P}_S^{\perp}\boldsymbol{P}_{\boldsymbol{P}_S^{\perp}\boldsymbol{H}}\boldsymbol{P}_S^{\perp} = \boldsymbol{P}_{\boldsymbol{P}_S^{\perp}\boldsymbol{H}}$$

(7) 若子空间Range($\boldsymbol{H}$)与Range($\boldsymbol{S}$)正交，即Range($\boldsymbol{H}$) $\perp$ Range($\boldsymbol{S}$)，则

$$\boldsymbol{E}_{H|S} = \boldsymbol{P}_H, \qquad \boldsymbol{E}_{S|H} = \boldsymbol{P}_H^{\perp}$$

9.2 习题与解答

9.1 证明唯一的非奇异幂等矩阵为单位矩阵。

证明： 幂等矩阵的特征值只有0，1两种。非奇异的幂等矩阵的秩全为1，则该矩阵必为单位矩阵。

9.2 若$\boldsymbol{A}$为幂等矩阵，证明

(1) 矩阵$\boldsymbol{A}^k$具有与$\boldsymbol{A}$相同的特征值。

(2) $\boldsymbol{A}^k$与$\boldsymbol{A}$ 具有相同的秩。

解： 由于$\boldsymbol{A}$为幂等矩阵，则有$\boldsymbol{A}^k=\boldsymbol{A}$，则(1)(2)结论显然成立。

9.3 假定$\boldsymbol{A}$ 和$\boldsymbol{B}$ 为对称矩阵，并且$\boldsymbol{B}$ 正定。若$\boldsymbol{AB}$ 的所有特征值为1或者0，证明$\boldsymbol{AB}$是幂等矩阵。

证明： 由$\boldsymbol{B}$ 正定有，$\boldsymbol{AB}$相似于$\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{ABB}^{-\frac{1}{2}}$，则$\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{AB}^{\frac{1}{2}}$ 的所有特征值为1或者0。又因为$\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{AB}^{\frac{1}{2}}$是对称矩阵，所以其必是幂等矩阵。所以有

$$\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{AB}^{\frac{1}{2}}=\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{AB}^{\frac{1}{2}}\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{AB}^{\frac{1}{2}}=\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{ABAB}^{\frac{1}{2}}$$

上式左乘$\boldsymbol{B}^{-\frac{1}{2}}$，右乘$\boldsymbol{B}^{\frac{1}{2}}$则有$\boldsymbol{ABAB}=\boldsymbol{AB}$。

9.4 若$\boldsymbol{A},\boldsymbol{B}$为$n\times n$ 幂等矩阵，并且$\boldsymbol{AB}=\boldsymbol{BA}$，证明$\boldsymbol{AB}$也是幂等矩阵。

证明：

$$\boldsymbol{ABAB}=\boldsymbol{AABB}=\boldsymbol{AB}$$

9.5 令$\boldsymbol{X}$ 表示观测数据矩阵，现在用它估计向量$\boldsymbol{y}$。已知两个观测数据向量$\boldsymbol{x}_1=[2,1,2,3]^{\mathrm{T}},\boldsymbol{x}_2=[1,2,1,1]^{\mathrm{T}}$。若使用它们估计$\boldsymbol{y}=[1,2,3,2]^{\mathrm{T}}$，求估计的误差平方和。（提示：令最优滤波器为$\boldsymbol{w}_{\mathrm{opt}}$，则有观测方程$\boldsymbol{Xw}_{\mathrm{opt}}=\boldsymbol{y}$。）

解： $\boldsymbol{y}$到数据空间的正交投影为

$$\hat{\boldsymbol{y}}=\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}=\left[\frac{16}{9},\frac{94}{45},\frac{16}{9},\frac{34}{15}\right]^{\mathrm{T}}$$

$$\|\boldsymbol{y}-\hat{\boldsymbol{y}}\|^2=\frac{98}{45}$$

9.6 假定$\boldsymbol{V}_1,\boldsymbol{V}_2$分别由复向量$C^n$ 的子空间W的两组标准正交基组成，证明

$$\boldsymbol{V}_1\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{x}=\boldsymbol{V}_2\boldsymbol{V}_2^{\mathrm{H}}\boldsymbol{x}$$

对所有向量$\boldsymbol{x}$成立。

解法一： 记

$$C^n=W\oplus V$$

则$\forall\boldsymbol{x}$,$\boldsymbol{x}=\boldsymbol{x}_w+\boldsymbol{x}_v$。其中$\boldsymbol{x}_w\in W$,$\boldsymbol{x}_v\in V$。

由于

$$\boldsymbol{V}_1\boldsymbol{V}_1^{\mathrm{H}}=\boldsymbol{V}_1(\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{V}_1)^{-1}\boldsymbol{V}_1^{\mathrm{H}}\qquad \boldsymbol{V}_2\boldsymbol{V}_2^{\mathrm{H}}=\boldsymbol{V}_2(\boldsymbol{V}_2^{\mathrm{H}}\boldsymbol{V}_2)^{-1}\boldsymbol{V}_2^{\mathrm{H}}$$

所以有

$$\boldsymbol{V}_1\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{x}=\boldsymbol{V}_1\boldsymbol{V}_1^{\mathrm{H}}(\boldsymbol{x}_w+\boldsymbol{x}_v)=\boldsymbol{V}_1\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{x}_w=\boldsymbol{x}_w=\boldsymbol{V}_2\boldsymbol{V}_2^{\mathrm{H}}\boldsymbol{x}$$

解法二： 由题设有，$\boldsymbol{V}_2=\boldsymbol{V}_1\boldsymbol{P}$，其中$\boldsymbol{P}$是正交矩阵。于是有

$$\boldsymbol{V}_2\boldsymbol{V}_2^{\mathrm{H}}\boldsymbol{x}=\boldsymbol{V}_1\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{x}=\boldsymbol{V}_1\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{x}$$

实际上也可以从投影的角度来解释：

$$\boldsymbol{y}=\boldsymbol{V}_1\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{x}=\boldsymbol{V}_1\left(\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{V}_1\right)^{-1}\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{x}$$

即$\boldsymbol{y}$是向量$\boldsymbol{x}$到由$\boldsymbol{V}_1$张成的子空间的投影。同理$\boldsymbol{y}$也是向量$\boldsymbol{x}$到由$\boldsymbol{V}_2$张成的子空间的投影。由于$\boldsymbol{V}_1$，$\boldsymbol{V}_2$张成同一个空间，则投影都相同，命题得证。

9.7 证明：若投影算子$\boldsymbol{P}_1$和$\boldsymbol{P}_2$是可交换的，即$\boldsymbol{P}_1\boldsymbol{P}_2=\boldsymbol{P}_2\boldsymbol{P}_1$，则它们的乘积$\boldsymbol{P}=\boldsymbol{P}_1\boldsymbol{P}_2$是一投影算子；并求$\boldsymbol{P}$的值域Range($\boldsymbol{P}$)和零空间Null($\boldsymbol{P}$)。

解： 验证其是否是投影矩阵，只需验证其是否满足幂等性，即

$$\boldsymbol{P}\boldsymbol{P}=\boldsymbol{P}_1\boldsymbol{P}_2\boldsymbol{P}_1\boldsymbol{P}_2=\boldsymbol{P}_2\boldsymbol{P}_1\boldsymbol{P}_1\boldsymbol{P}_2=\boldsymbol{P}_2\boldsymbol{P}_1\boldsymbol{P}_2=\boldsymbol{P}_1\boldsymbol{P}_2\boldsymbol{P}_2=\boldsymbol{P}_1\boldsymbol{P}_2=\boldsymbol{P}$$

所以$\boldsymbol{P}$是投影矩阵。

由$\boldsymbol{P}$的构造，不难猜测$\boldsymbol{P}$的值域可以表示为

$$\mathrm{Range}(\boldsymbol{P})=\mathrm{Range}(\boldsymbol{P}_1)\cap\mathrm{Range}(\boldsymbol{P}_2)$$

证明如下：

显然Range($\boldsymbol{P}$) $\subset$ Range($\boldsymbol{P}_1$)。同时因为$\boldsymbol{P}=\boldsymbol{P}_2\boldsymbol{P}_1$，所以也有Range($\boldsymbol{P}$) $\subset$ Range($\boldsymbol{P}_2$)。所以

$$\mathrm{Range}(\boldsymbol{P})\subset\mathrm{Range}(\boldsymbol{P}_1)\cap\mathrm{Range}(\boldsymbol{P}_2)$$

$\forall\boldsymbol{x}\subset$ Range($\boldsymbol{P}_1$) $\cap$ Range($\boldsymbol{P}_2$)有

$$\boldsymbol{P}_1\boldsymbol{x}=\boldsymbol{x},\qquad \boldsymbol{P}_2\boldsymbol{x}=\boldsymbol{x}$$

$$P_1P_2x = P_1x = x$$

即$x \in \text{Range}(P)$ 于是有

$$\text{Range}(P) \supset \text{Range}(P_1) \cap \text{Range}(P_2)$$

命题得证。由于

$$\text{Range}(P) \oplus \text{Null}(P) = C^n$$

(注意这个结论只对P是投影算子成立) 所以有

$$\begin{aligned}\text{Null}(P) &= \overline{\text{Range}(P)} = \overline{\text{Range}(P_1) \cap \text{Range}(P_2)} \\ &= \overline{\text{Range}(P_1)} \cup \overline{\text{Range}(P_2)} = \text{Null}(P_1) \cup \text{Null}(P_2)\end{aligned}$$

9.8 已知矩阵

$$A = \begin{bmatrix} 6 & 2 \\ -7 & 6 \end{bmatrix}$$

和

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

分别求它们的Moore-Penrose逆矩阵$A^\dagger$，并解释为什么$AA^\dagger$和$A^\dagger A$分别是到矩阵A的列空间和行空间的正交投影。

解： 根据教材1.9节的知识，不难求得以上两个矩阵的Moore-Penrose逆矩阵分别为

$$\begin{bmatrix} \dfrac{3}{25} & \dfrac{-1}{25} \\ \dfrac{7}{50} & \dfrac{3}{25} \end{bmatrix} \qquad \begin{bmatrix} \dfrac{-1}{3} & \dfrac{2}{3} \\ \dfrac{1}{3} & \dfrac{1}{3} \\ \dfrac{2}{3} & \dfrac{-1}{3} \end{bmatrix}$$

由于

$$\text{Range}(A) = \text{Range}(AA^\dagger A) \subset \text{Range}(AA^\dagger) \subset \text{Range}(A)$$

所以

$$\text{Range}(AA^\dagger) = \text{Range}(A)$$

同时由于

$$AA^\dagger AA^\dagger = AA^\dagger$$

是幂等矩阵，所以有$AA^\dagger$ 是A列空间的投影。

同理证明

$$\text{Range}(A^{\text{H}}) = \text{Range}(A^{\text{H}}(A^\dagger)^{\text{H}}A^{\text{H}}) \subset \text{Range}(A^{\text{H}}(A^\dagger)^{\text{H}}) \subset \text{Range}(A^{\text{H}})$$

又因为

$$\text{Range}(\boldsymbol{A}^{\text{H}}(\boldsymbol{A}^{\dagger})^{\text{H}}) = \text{Range}(\boldsymbol{A}^{\dagger}\boldsymbol{A})$$

所以

$$\text{Range}(\boldsymbol{A}^{\dagger}\boldsymbol{A}) = \text{Range}(\boldsymbol{A}^{\text{H}})$$

同时由于

$$\boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{A}^{\dagger}\boldsymbol{A} = \boldsymbol{A}^{\dagger}\boldsymbol{A}$$

是幂等矩阵，所以有$\boldsymbol{A}^{\dagger}\boldsymbol{A}$ 是$\boldsymbol{A}^{\text{H}}$列空间的投影，即$\boldsymbol{A}$行空间的投影。

9.9 假定两个基向量

$$\boldsymbol{u}_1 = [-1, 2, -4, 3, 1]^{\text{T}}$$
$$\boldsymbol{u}_2 = [5, 6, 2, -2, -1]^{\text{T}}$$

生成向量空间$\text{Range}(\boldsymbol{U}) = \text{Span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$。试问向量

$$\boldsymbol{v} = [-31, -18, -34, 28, 11]^{\text{T}}$$

是否在向量空间$\text{Range}(\boldsymbol{U})$内，并加以证明。

解： 由于

$$\text{rank}\{\boldsymbol{u}_1, \boldsymbol{u}_2\} = 2 = \text{rank}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{v}\}$$

所以$\boldsymbol{v}$可以由$\boldsymbol{u}_1$，$\boldsymbol{u}_2$线性表示，实际上$\boldsymbol{v} = 6\boldsymbol{u}_1 - 5\boldsymbol{u}_2$。因此$\boldsymbol{v}$在$\text{Range}(\boldsymbol{U})$内。

9.10 证明下列关系为真：

$$\boldsymbol{X}_{1,k}(n) = \begin{bmatrix} \boldsymbol{0}_k^{\text{T}} \\ \boldsymbol{X}_{0,k-1}(n-1) \end{bmatrix}$$

和

$$\boldsymbol{P}_{1,k}(n) = \begin{bmatrix} 0 & \boldsymbol{0}_{k-1}^{\text{T}} \\ \boldsymbol{0}_{k-1} & \boldsymbol{P}_{0,k-1}(n-1) \end{bmatrix}$$

式中，$\boldsymbol{0}_k$为$k \times 1$维零向量。

证明： 由数据矩阵的定义,有

$$\begin{aligned}
\boldsymbol{X}_{1,k}(n) &= [z^{-1}\boldsymbol{x}(n), \cdots, z^{-k}\boldsymbol{x}(n)]^{\text{T}} \\
&= \begin{bmatrix} 0 & 0 & \cdots & 0 \\ x(1) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ x(n-1) & x(n-2) & \cdots & x(n-k) \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{0}_k^{\text{T}} \\ \boldsymbol{X}_{0,k-1}(n-1) \end{bmatrix}
\end{aligned}$$

将上面的结果代入投影矩阵的定义公式，得

$$
\begin{aligned}
\boldsymbol{P}_{1,k}(n) =& \boldsymbol{X}_{1,k}(n)\langle \boldsymbol{X}_{1,k}(n), \boldsymbol{X}_{1,k}(n)\rangle^{-1}\boldsymbol{X}_{1,k}^{\mathrm{T}}(n) \\
=& \begin{bmatrix} \mathbf{0}_k^{\mathrm{T}} \\ \boldsymbol{X}_{0,k-1}(n-1) \end{bmatrix} \left\langle \begin{bmatrix} \mathbf{0}_k^{\mathrm{T}} \\ \boldsymbol{X}_{0,k-1}(n-1) \end{bmatrix}, \begin{bmatrix} \mathbf{0}_k^{\mathrm{T}} \\ \boldsymbol{X}_{0,k-1}(n-1) \end{bmatrix} \right\rangle^{-1} \times \\
& \begin{bmatrix} \mathbf{0}_k^{\mathrm{T}} \\ \boldsymbol{X}_{0,k-1}(n-1) \end{bmatrix}^{\mathrm{T}} \\
=& \begin{bmatrix} \mathbf{0}_k^{\mathrm{T}} \\ \boldsymbol{X}_{0,k-1}(n-1) \end{bmatrix} \langle \boldsymbol{X}_{0,k-1}(n-1), \boldsymbol{X}_{0,k-1}(n-1)\rangle^{-1} \begin{bmatrix} \mathbf{0}_k^{\mathrm{T}} \\ \boldsymbol{X}_{0,k-1}(n-1) \end{bmatrix}^{\mathrm{T}} \\
=& \begin{bmatrix} 0 & \mathbf{0}_{k-1}^{\mathrm{T}} \\ \mathbf{0}_{k-1} & \boldsymbol{P}_{0,k-1}(n-1) \end{bmatrix}
\end{aligned}
$$

命题得证。

9.11 用逆矩阵

$$\langle \boldsymbol{X}_{1,p}(n-1), \boldsymbol{X}_{1,p}(n-1)\rangle^{-1}$$

表示逆矩阵

$$\langle \boldsymbol{X}_{1,p}(n), \boldsymbol{X}_{1,p}(n)\rangle^{-1}$$

解： 由教材中定义式(10.3.26)有

$$\boldsymbol{X}_{1,p}(n) = \begin{bmatrix} \boldsymbol{X}_{1,p}(n-1) \\ \boldsymbol{u} \end{bmatrix}$$

其中$\boldsymbol{u} = [x(n-1), x(n-2), \cdots, x(n-p)]$，则

$$
\begin{aligned}
\langle \boldsymbol{X}_{1,p}(n), \boldsymbol{X}_{1,p}(n)\rangle^{-1} &= \left(\boldsymbol{X}_{1,p}(n)^{\mathrm{H}}\boldsymbol{X}_{1,p}(n)\right)^{-1} \\
&= \left(\boldsymbol{X}_{1,p}(n-1)^{\mathrm{H}}\boldsymbol{X}_{1,p}(n-1) + \boldsymbol{u}^{\mathrm{H}}\boldsymbol{u}\right)^{-1}
\end{aligned}
$$

由矩阵求逆引理(教材P68:Sherman-Morrison 公式)有

$$\langle \boldsymbol{X}_{1,p}(n), \boldsymbol{X}_{1,p}(n)\rangle^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}^{\mathrm{H}}\boldsymbol{u}\boldsymbol{A}^{-1}}{1+\boldsymbol{u}\boldsymbol{A}^{-1}\boldsymbol{u}^{\mathrm{H}}}$$

上式中$\boldsymbol{A} = \langle \boldsymbol{X}_{1,p}(n-1), \boldsymbol{X}_{1,p}(n-1)\rangle$。

9.12 已知

$$\gamma_m(n-1) = \langle \boldsymbol{\pi}(n), \boldsymbol{P}_{1,m}^{\perp}(n)\boldsymbol{\pi}(n)\rangle$$

其中$\boldsymbol{\pi}(n) = [0, \cdots, 0, 1]^{\mathrm{T}}$的$n$维向量。证明

$$\gamma_m(n) = \langle \boldsymbol{\pi}(n), \boldsymbol{P}_{0,m-1}^{\perp}(n)\boldsymbol{\pi}(n)\rangle$$

证明：

$$
\begin{aligned}
\boldsymbol{P}_{0,m-1}^{\perp}(n) &= \boldsymbol{X}_{0,m-1}(n)\langle \boldsymbol{X}_{0,m-1}(n), \boldsymbol{X}_{0,m-1}(n)\rangle^{-1}\boldsymbol{X}_{0,m-1}^{\mathrm{T}}(n) \\
&= \boldsymbol{X}_{1,m}(n+1)\langle \boldsymbol{X}_{1,m}(n+1), \boldsymbol{X}_{1,m}(n+1)\rangle^{-1}\boldsymbol{X}_{1,m}^{\mathrm{T}}(n+1) \\
&= \boldsymbol{P}_{1,m}^{\perp}(n+1)
\end{aligned}
$$

则由$\gamma_m(n)$的定义有

$$
\gamma_m(n) = \langle \boldsymbol{\pi}(n), \boldsymbol{P}_{1,m}^{\perp}(n+1)\boldsymbol{\pi}(n)\rangle = \langle \boldsymbol{\pi}(n), \boldsymbol{P}_{0,m-1}^{\perp}(n)\boldsymbol{\pi}(n)\rangle
$$

9.13 给定一时间信号$\boldsymbol{v}(n) = [v(1), v(2), v(3), \cdots, v(n)]^{\mathrm{T}} = [4, 2, 4, \cdots]^{\mathrm{T}}$。计算

(1) 数据向量$\boldsymbol{v}(2)$和$\boldsymbol{v}(3)$。

(2) 向量$z^{-1}\boldsymbol{v}(2)$和$z^{-2}\boldsymbol{v}(2)$。

(3) 向量$z^{-1}\boldsymbol{v}(3)$和$z^{-2}\boldsymbol{v}(3)$。

(4) 若令$\boldsymbol{u}(n) = z^{-1}\boldsymbol{v}(n)$，计算投影矩阵$\boldsymbol{P}_u(2)$和$\boldsymbol{P}_u(3)$。

(5) 利用$\boldsymbol{u}(n)$求$\boldsymbol{v}(n)$的最小二乘预测。这一预测称为$\boldsymbol{v}(n)$的一步前向预测。

(6) 计算前向预测误差向量$\boldsymbol{e}_1^{\mathrm{f}}(2)$和$\boldsymbol{e}_1^{\mathrm{f}}(3)$。

解： 注意到$z^{-k}\boldsymbol{v}(n) = [0, \cdots, 0, v(1), \cdots, v(n-k)]^{\mathrm{T}}$，易得

(1) 数据向量

$$
\boldsymbol{v}(2) = \begin{bmatrix} v(1) \\ v(2) \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}
$$

$$
\boldsymbol{v}(3) = \begin{bmatrix} v(1) \\ v(2) \\ v(3) \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 4 \end{bmatrix}
$$

(2) 向量

$$
z^{-1}\boldsymbol{v}(2) = \begin{bmatrix} 0 \\ v(1) \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \end{bmatrix}
$$

$$
z^{-2}\boldsymbol{v}(2) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}
$$

(3) 向量

$$
z^{-1}\boldsymbol{v}(3) = \begin{bmatrix} 0 \\ v(1) \\ v(2) \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix}
$$

$$
z^{-2}\boldsymbol{v}(3) = \begin{bmatrix} 0 \\ 0 \\ v(1) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix}
$$

(4) 由于

$$\boldsymbol{u}(2) = z^{-1}\boldsymbol{v}(2) = \begin{bmatrix} 0 \\ v(1) \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \end{bmatrix}$$

故投影矩阵为

$$\begin{aligned}
\boldsymbol{P}_u(2) &= \boldsymbol{u}(2)\langle \boldsymbol{u}(2), \boldsymbol{u}(2)\rangle^{-1}\boldsymbol{u}^{\mathrm{T}}(2) \\
&= \begin{bmatrix} 0 \\ 4 \end{bmatrix} \left\langle \begin{bmatrix} 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \end{bmatrix} \right\rangle^{-1} \begin{bmatrix} 0 \\ 4 \end{bmatrix}^{\mathrm{T}} \\
&= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}
\end{aligned}$$

类似地,有

$$\begin{aligned}
\boldsymbol{P}_u(3) &= \boldsymbol{u}(3)\langle \boldsymbol{u}(3), \boldsymbol{u}(3)\rangle^{-1}\boldsymbol{u}^{\mathrm{T}}(3) \\
&= \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix} \left\langle \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix} \right\rangle^{-1} \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix}^{\mathrm{T}} \\
&= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.8 & 0.4 \\ 0 & 0.4 & 0.2 \end{bmatrix}
\end{aligned}$$

(5) 由于用$\boldsymbol{u}(n)$一步前向预测$\boldsymbol{v}(n)$, 故$\boldsymbol{P}_{1,m}(n) = \boldsymbol{P}_{1,1}(n) = \boldsymbol{P}_u(n)$, 且

$$\hat{\boldsymbol{v}}(n) = \boldsymbol{P}_{1,1}\boldsymbol{v}(n) = \boldsymbol{P}_u(n)\boldsymbol{v}(n)$$

故

$$\begin{aligned}
\hat{\boldsymbol{v}}(2) &= \boldsymbol{P}_u(2)\boldsymbol{v}(2) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \\
\hat{\boldsymbol{v}}(3) &= \boldsymbol{P}_u(3)\boldsymbol{v}(3) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.8 & 0.4 \\ 0 & 0.4 & 0.2 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 3.2 \\ 1.6 \end{bmatrix}
\end{aligned}$$

(6) 前向预测误差向量

$$\begin{aligned}
\boldsymbol{e}_1^{\mathrm{f}}(2) &= \boldsymbol{v}(2) - \hat{\boldsymbol{v}}(2) = \begin{bmatrix} 4 \\ 2 \end{bmatrix} - \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \end{bmatrix} \\
\boldsymbol{e}_1^{\mathrm{f}}(3) &= \boldsymbol{v}(3) - \hat{\boldsymbol{v}}(3) = \begin{bmatrix} 4 \\ 2 \\ 4 \end{bmatrix} - \begin{bmatrix} 0 \\ 3.2 \\ 1.6 \end{bmatrix} = \begin{bmatrix} 4 \\ -1.2 \\ 2.4 \end{bmatrix}
\end{aligned}$$

9.14 已知前向和后向预测残差分别为

$$\begin{aligned}
\epsilon_m^{\mathrm{f}}(n) &= \langle \boldsymbol{x}(n), \boldsymbol{P}_{1,m}^{\perp}(n)\boldsymbol{x}(n)\rangle \\
\epsilon_m^{\mathrm{b}}(n) &= \langle z^{-m}\boldsymbol{x}(n), \boldsymbol{P}_{0,m-1}^{\perp}(n)z^{-m}\boldsymbol{x}(n)\rangle
\end{aligned}$$

和偏相关系数$\Delta_{m+1}(n)=\langle \boldsymbol{e}_m^{\rm f}(n), z^{-1}\boldsymbol{e}_m^{\rm b}(n)\rangle$。证明

$$\epsilon_{m+1}^{\rm f}(n)=\epsilon_m^{\rm f}(n)-\frac{\Delta_{m+1}^2(n)}{\epsilon_m^{\rm b}(n-1)}$$

$$\epsilon_{m+1}^{\rm b}(n)=\epsilon_m^{\rm b}(n-1)-\frac{\Delta_{m+1}^2(n)}{\epsilon_m^{\rm f}(n)}$$

证明： 由递推公式

$$\langle \boldsymbol{z}, \boldsymbol{P}_{U,u}^{\perp}\boldsymbol{y}\rangle=\langle \boldsymbol{z}, \boldsymbol{P}_U^{\perp}\boldsymbol{y}\rangle-\langle \boldsymbol{z}, \boldsymbol{P}_U^{\perp}\boldsymbol{u}\rangle\langle \boldsymbol{P}_U^{\perp}\boldsymbol{u}, \boldsymbol{P}_U^{\perp}\boldsymbol{u}\rangle^{-1}\langle \boldsymbol{u}, \boldsymbol{P}_U^{\perp}\boldsymbol{y}\rangle \tag{9.12}$$

(1) 令上式中$\boldsymbol{z}=\boldsymbol{y}=\boldsymbol{x}(n)$，$\boldsymbol{U}=\boldsymbol{X}_{1,m}(n)$和$\boldsymbol{u}=z^{-m-1}\boldsymbol{x}(n)$，则易知

$$\boldsymbol{P}_U^{\perp}\boldsymbol{y}=\boldsymbol{P}_{1,m}^{\perp}\boldsymbol{x}(n)$$

$$\boldsymbol{P}_U^{\perp}\boldsymbol{u}=\boldsymbol{P}_{1,m}^{\perp}z^{-m-1}\boldsymbol{x}(n)$$

利用正交投影矩阵的性质,又可得到

$$\begin{aligned}
\langle \boldsymbol{z}, \boldsymbol{P}_U^{\perp}\boldsymbol{u}\rangle &= \langle \boldsymbol{x}(n), \boldsymbol{P}_{1,m}^{\perp}(n)z^{-m-1}\boldsymbol{x}(n)\rangle \\
&= \langle \boldsymbol{P}_{1,m}^{\perp}(n)\boldsymbol{x}(n), \boldsymbol{P}_{1,m}^{\perp}(n)z^{-m-1}\boldsymbol{x}(n)\rangle \\
&= \langle \boldsymbol{P}_{1,m}^{\perp}(n)\boldsymbol{x}(n), z^{-1}\boldsymbol{P}_{0,m-1}^{\perp}(n)z^{-m}\boldsymbol{x}(n)\rangle \\
&= \langle \boldsymbol{e}_m^{\rm f}(n), z^{-1}\boldsymbol{e}_m^{\rm b}(n)\rangle \\
&= \Delta_{m+1}(n) \\
\langle \boldsymbol{P}_U^{\perp}\boldsymbol{u}, \boldsymbol{P}_U^{\perp}\boldsymbol{u}\rangle &= \langle \boldsymbol{P}_{1,m}^{\perp}(n)z^{-m-1}\boldsymbol{x}(n), \boldsymbol{P}_{1,m}^{\perp}(n)z^{-m-1}\boldsymbol{x}(n)\rangle \\
&= \langle z^{-1}\boldsymbol{P}_{0,m-1}^{\perp}(n)\boldsymbol{x}(n), z^{-1}\boldsymbol{P}_{0,m-1}^{\perp}(n)\boldsymbol{x}(n)\rangle \\
&= \langle z^{-1}\boldsymbol{e}_m^{\rm b}(n), z^{-1}\boldsymbol{e}_m^{\rm b}(n)\rangle \\
&= \epsilon_m^{\rm b}(n-1) \\
\langle \boldsymbol{u}, \boldsymbol{P}_U^{\perp}\boldsymbol{y}\rangle &= \langle z^{-m-1}\boldsymbol{x}(n), \boldsymbol{P}_{1,m}^{\perp}(n)\boldsymbol{x}(n)\rangle \\
&= \langle \boldsymbol{P}_{1,m}^{\perp}(n)z^{-m-1}\boldsymbol{x}(n), \boldsymbol{P}_{1,m}^{\perp}(n)\boldsymbol{x}(n)\rangle \\
&= \langle z^{-1}\boldsymbol{e}_m^{\rm b}(n), \boldsymbol{e}_m^{\rm f}(n)\rangle \\
&= \Delta_{m+1}(n)
\end{aligned}$$

由上面的式子可直接得到前向预测残差的阶数和时间更新公式为

$$\epsilon_{m+1}^{\rm f}(n)=\epsilon_m^{\rm f}(n)-\frac{\Delta_{m+1}^2(n)}{\epsilon_m^{\rm b}(n-1)}$$

(2) 在式(9.12)中令$\boldsymbol{z}=\boldsymbol{y}=z^{-m-1}\boldsymbol{x}(n)$，$\boldsymbol{U}=\boldsymbol{X}_{1,m}(n)$ 和$\boldsymbol{u}=\boldsymbol{x}(n)$，则有

$$\begin{aligned}
\langle \boldsymbol{z}, \boldsymbol{P}_U^{\perp}\boldsymbol{u}\rangle &= \langle z^{-m-1}\boldsymbol{x}(n), \boldsymbol{P}_{1,m}^{\perp}(n)\boldsymbol{x}(n)\rangle \\
&= \Delta_{m+1}(n)
\end{aligned}$$

另一方面,有

$$\begin{aligned}\langle \boldsymbol{P}_U^{\perp}\boldsymbol{u}, \boldsymbol{P}_U^{\perp}\boldsymbol{u}\rangle &= \langle \boldsymbol{P}_{1,m}^{\perp}\boldsymbol{x}(n), \boldsymbol{P}_{1,m}^{\perp}(n)\boldsymbol{x}(n)\rangle \\ &= \langle \boldsymbol{x}(n), \boldsymbol{P}_{1,m}^{\perp}(n)\boldsymbol{x}(n)\rangle \\ &= \epsilon_m^{\mathrm{f}}(n) \\ \langle \boldsymbol{u}, \boldsymbol{P}_U^{\perp}\boldsymbol{y}\rangle &= \langle \boldsymbol{x}(n), \boldsymbol{P}_{1,m}^{\perp}(n)z^{-m-1}\boldsymbol{x}(n)\rangle \\ &= \Delta_{m+1}(n)\end{aligned}$$

将上式代入式(9.12),并加以整理,即得到

$$\epsilon_{m+1}^{\mathrm{b}}(n) = \epsilon_m^{\mathrm{b}}(n-1) - \frac{\Delta_{m+1}^2(n)}{\epsilon_m^{\mathrm{f}}(n)}$$

9.15 令$\boldsymbol{U}$是$n \times N$实矩阵,并且满列秩,则

$$\boldsymbol{K}_U = \langle \boldsymbol{U}, \boldsymbol{U}\rangle^{-1}\boldsymbol{U}^{\mathrm{T}}$$

称为列空间$U = \mathrm{Span}(\boldsymbol{U})$的横向滤波器算子。考虑新的列空间$Uu = \mathrm{Span}\{\boldsymbol{U}, \boldsymbol{u}\}$,试证明横向滤波器算子的下列递推公式:

$$\boldsymbol{K}_{Uu} = \begin{bmatrix}\boldsymbol{K}_U \\ \boldsymbol{0}_N^{\mathrm{T}}\end{bmatrix} + \left(\begin{bmatrix}\boldsymbol{0}_N \\ 1\end{bmatrix} - \begin{bmatrix}\boldsymbol{K}_U\boldsymbol{u} \\ 0\end{bmatrix}\right)\langle \boldsymbol{P}_U^{\perp}\boldsymbol{u}, \boldsymbol{P}_U^{\perp}\boldsymbol{u}\rangle^{-1}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{P}_U^{\perp}$$

证明:

$$\boldsymbol{K}_{Uu} = \left(\begin{bmatrix}\boldsymbol{U}^{\mathrm{T}} \\ \boldsymbol{u}^{\mathrm{T}}\end{bmatrix}\begin{bmatrix}\boldsymbol{U} & \boldsymbol{u}\end{bmatrix}\right)^{-1}\begin{bmatrix}\boldsymbol{U}^{\mathrm{T}} \\ \boldsymbol{u}^{\mathrm{T}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} & \boldsymbol{U}^{\mathrm{T}}\boldsymbol{u} \\ \boldsymbol{u}^{\mathrm{T}}\boldsymbol{U} & \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{U}^{\mathrm{T}} \\ \boldsymbol{u}^{\mathrm{T}}\end{bmatrix}$$

用高斯消元法计算等式右边的逆矩阵

$$\left[\begin{array}{cc|cc}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} & \boldsymbol{U}^{\mathrm{T}}\boldsymbol{u} & \boldsymbol{I} & \\ \boldsymbol{u}^{\mathrm{T}}\boldsymbol{U} & \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} & & 1\end{array}\right] \longrightarrow \left[\begin{array}{cc|cc}\boldsymbol{I} & (\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{u} & (\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1} & \\ \boldsymbol{u}^{\mathrm{T}}\boldsymbol{U} & \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} & & 1\end{array}\right] \longrightarrow$$

$$\left[\begin{array}{cc|cc}\boldsymbol{I} & (\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{u} & (\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{u}^{\mathrm{T}}\boldsymbol{P}_U^{\perp}\boldsymbol{u} & -\boldsymbol{u}^{\mathrm{T}}\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1} & 1\end{array}\right]$$

上式利用了$\boldsymbol{P}_U^{\perp} = \boldsymbol{I} - \boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}}$。在下面的推导中将多次用到这个结论。

令 $\sigma = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{P}_U^{\perp}\boldsymbol{u} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{P}_U^{\perp}\boldsymbol{P}_U^{\perp}\boldsymbol{u} = \langle \boldsymbol{P}_U^{\perp}\boldsymbol{u}, \boldsymbol{P}_U^{\perp}\boldsymbol{u}\rangle$,进而有

$$\left[\begin{array}{cc|cc}\boldsymbol{I} & (\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{u} & (\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & 1 & -\frac{1}{\sigma}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1} & \frac{1}{\sigma}\end{array}\right] \longrightarrow$$

$$\left[\begin{array}{cc|cc}\boldsymbol{I} & \boldsymbol{0} & (\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1} + \frac{1}{\sigma}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1} & -\frac{1}{\sigma}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{u} \\ \boldsymbol{0} & 1 & -\frac{1}{\sigma}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1} & \frac{1}{\sigma}\end{array}\right]$$

则

$$\begin{aligned}
\boldsymbol{K}_{Uu} &= \begin{bmatrix} (\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1} + \frac{1}{\sigma}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1} & -\frac{1}{\sigma}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{u} \\ -\frac{1}{\sigma}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1} & \frac{1}{\sigma} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}^{\mathrm{T}} \\ \boldsymbol{u}^{\mathrm{T}} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{K}_U + \frac{1}{\sigma}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{P}_U - \frac{1}{\sigma}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}} \\ -\frac{1}{\sigma}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{P}_U + \frac{1}{\sigma}\boldsymbol{u}^{\mathrm{T}} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{K}_U - \langle \boldsymbol{P}_U^{\perp}\boldsymbol{u}, \boldsymbol{P}_U^{\perp}\boldsymbol{u}\rangle^{-1}\boldsymbol{K}_U\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{P}_U^{\perp} \\ \langle \boldsymbol{P}_U^{\perp}\boldsymbol{u}, \boldsymbol{P}_U^{\perp}\boldsymbol{u}\rangle^{-1}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{P}_U^{\perp} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{K}_U \\ \boldsymbol{0}_N^{\mathrm{T}} \end{bmatrix} + \left(\begin{bmatrix} \boldsymbol{0}_N \\ 1 \end{bmatrix} - \begin{bmatrix} \boldsymbol{K}_U\boldsymbol{u} \\ 0 \end{bmatrix} \right) \langle \boldsymbol{P}_U^{\perp}\boldsymbol{u}, \boldsymbol{P}_U^{\perp}\boldsymbol{u}\rangle^{-1}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{P}_U^{\perp}
\end{aligned}$$

命题得证。

9.16 记

$$\begin{aligned}
\{\boldsymbol{U}, \boldsymbol{u}\} &= \{\boldsymbol{x}(n), z^{-1}\boldsymbol{x}(n), \cdots, z^{-N+1}\boldsymbol{x}(n), \boldsymbol{\pi}(n)\} \\
&= \{\boldsymbol{X}_{0,N-1}(n), \boldsymbol{\pi}(n)\}
\end{aligned}$$

证明

$$\boldsymbol{K}_{0,N-1,\pi}(n) = \begin{bmatrix} \boldsymbol{K}_{0,N-1}(n-1) & \boldsymbol{0}_N \\ \boldsymbol{y}^{\mathrm{T}}(n-1) & 1 \end{bmatrix}$$

式中，“$0, N-1, \pi$” 表示在$\boldsymbol{X}_{0,N-1}(n)$的最后一列之后追加$\pi(n)$，$\boldsymbol{y}(n-1)$是一任意向量。(提示：使用$\boldsymbol{P}_{0,N-1,\pi}(n)$的递推公式。)

证明： 由$\boldsymbol{P}$的定义显然有

$$\boldsymbol{P}_{0,N-1,\pi}(n) = \begin{bmatrix} \boldsymbol{P}_{0,N-1}(n) & \boldsymbol{0}_{N-1} \\ \boldsymbol{0}_{N-1}^{\mathrm{T}} & 1 \end{bmatrix}$$

由$\boldsymbol{K}$的定义有

$$\boldsymbol{K}_{0,N-1}(n)\boldsymbol{P}_{0,N-1}(n) = \boldsymbol{K}_{0,N-1}(n)$$
$$\boldsymbol{K}_{0,N-1,\pi}(n)\boldsymbol{P}_{0,N-1,\pi}(n) = \boldsymbol{K}_{0,N-1,\pi}(n)$$

则必有

$$\boldsymbol{K}_{0,N-1,\pi}(n) = \begin{bmatrix} \boldsymbol{K}_{0,N-1}(n-1) & \boldsymbol{0}_N \\ \boldsymbol{y}^{\mathrm{T}}(n-1) & 1 \end{bmatrix}$$

$\boldsymbol{y}^{\mathrm{T}}(n-1)$位任一向量等式都成立。

9.17 假设在同步CDMA中有K个用户同时在通信，CDMA的扩频增益为N，在接收机处，第k个用户的接收功率为A_k，第k个用户的扩频波形为$\boldsymbol{s}_k$，且$\|\boldsymbol{s}_k\| = 1$，则接收信号的等效基带信号可以表示为[51]

$$\boldsymbol{r} = \boldsymbol{SAb} + \boldsymbol{n}$$

其中$b=[b_1,b_2,\cdots,b_K]^{\mathrm{T}}$ 为K个用户传输的信息比特，$\boldsymbol{A}=\mathrm{diag}(A_1,A_2,\cdots,A_K)$，$\boldsymbol{S}=[\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_K]$，$\boldsymbol{n}$是高斯噪声向量。则解相关输出为

$$\begin{aligned}\hat{\boldsymbol{y}}&=\boldsymbol{S}^{\dagger}\boldsymbol{r}=\left(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S}\right)^{-1}\boldsymbol{S}^{\mathrm{H}}(\boldsymbol{S}\boldsymbol{A}\boldsymbol{b}+\boldsymbol{n})\\&=\boldsymbol{A}\boldsymbol{b}+\left(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S}\right)^{-1}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{n}=\boldsymbol{A}\boldsymbol{b}+\boldsymbol{\nu}\end{aligned}$$

证明解相关检测器等价于斜投影。

证明： 不失一般性，考虑第一个用户的检测结果。解相关检测器的第一个用户的检测结果为

$$\hat{y}_1=A_1b_1+\nu_1$$

下面计算ν_1。由于

$$\left(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S}\right)^{-1}=\begin{bmatrix}\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{s}_1 & \boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{U}\\ \boldsymbol{U}^{\mathrm{H}}\boldsymbol{s}_1 & \boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}\end{bmatrix}^{-1}=\begin{bmatrix}\left(\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{P}_U^{\perp}\boldsymbol{s}_1\right)^{-1} & -\left(\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{P}_U^{\perp}\boldsymbol{s}_1\right)^{-1}\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}\right)^{-1}\\ \times & \times\end{bmatrix}$$

则有

$$\begin{aligned}\nu_1&=\left(\left(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S}\right)^{-1}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{n}\right)_1=\left(\left(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S}\right)^{-1}\begin{bmatrix}\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{n}\\ \boldsymbol{U}^{\mathrm{H}}\boldsymbol{n}\end{bmatrix}\right)_1\\&=\frac{\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{n}}{\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{P}_U^{\perp}\boldsymbol{s}_1}-\frac{\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{n}}{\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{P}_U^{\perp}\boldsymbol{s}_1}\\&=\frac{\boldsymbol{s}_1^{\mathrm{H}}\left(\boldsymbol{I}-\boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{H}}\right)\boldsymbol{n}}{\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{P}_U^{\perp}\boldsymbol{s}_1}\\&=\frac{\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{P}_U^{\perp}\boldsymbol{n}}{\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{P}_U^{\perp}\boldsymbol{s}_1}\end{aligned}$$

即解相关检测器第一个用户的检测信号可以表示为

$$\hat{y}_1=A_1b_1+\frac{\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{P}_U^{\perp}\boldsymbol{n}}{\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{P}_U^{\perp}\boldsymbol{s}_1}$$

以下再计算斜投影检测的结果。记干扰子空间为$\boldsymbol{U}=[\boldsymbol{s}_2,\boldsymbol{s}_3,\cdots,\boldsymbol{s}_K]$，则沿干扰子空间$\boldsymbol{U}$到信号子空间$\boldsymbol{s}_1$的斜投影矩阵可以表示为

$$\boldsymbol{E}_{\boldsymbol{s}_1|U}=\boldsymbol{S}_1\left(\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{P}_U^{\perp}\boldsymbol{s}_1\right)^{-1}\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{P}_U^{\perp}$$

斜投影算子第一用户的检测结果为

$$\begin{aligned}\hat{z}_1&=\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{E}_{\boldsymbol{s}_1|U}\boldsymbol{r}=\frac{\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{s}_1\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{P}_U^{\perp}}{\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{P}_U^{\perp}\boldsymbol{s}_1}\left(A_1b_1\boldsymbol{s}_1+\sum_{k=2}^{K}A_kb_k\boldsymbol{s}_k+\boldsymbol{n}\right)\\&=A_1b_1+\frac{\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{P}_U^{\perp}\boldsymbol{n}}{\boldsymbol{s}_1^{\mathrm{H}}\boldsymbol{P}_U^{\perp}\boldsymbol{s}_1}\\&=\hat{y}_1\end{aligned}$$

即第一个用户的解相关的结果，等价于接收信号沿干扰子空间到第一个用户的扩频波形组成的空间的投影，即在同步CDMA中，解相关检测等价于斜投影。

这个结论最先在参考文献[6]中给出证明，而本题中采用的证明方法更简单明了。

参考文献

[1] Magnus J R, Neudecker H. Matrix Differential Calculus with Applications in Statistics and Econometrics. Revised ed. Chichester: Wiley 1999

[2] Kumaresan R. Estimating the paramaters of exponentially damped or undamped sinusoidal signals in noise: [Ph.D. dissertation]. RI: University of Rhode Island, 1982

[3] Hendeson H V, Searle S R. On deriving the inverse of a sum of matrices. SIAM Review, 1981, 23: 53~60

[4] Golub G H, Van Loan C F. Matrix Computation. 2nd ed. Baltimore: The John Hopkins University Press, 1989

[5] Kayalar S, Weinert H L. Oblique projections: Formulas, algorithms, and error bounds. Math of Control, Signals, and Systems, 1989, 2(1): 33~45

[6] Eldar Y C. On geometric properties of the decorrelator. IEEE Commun, Lett. 2002, 6: 16~18

[7] Manolakis D G, Ingle V K, Kogon S M. Statistical and Adaptive Signal Processing. Boston: McGraw-Hill, 2000

[8] Haykin S. Adaptive Filter Theory. 3rd ed. Prentice Hall, 1996

[9] Li X L, Zhang X D. Non-orthogonal joint diagonalization free of degenerate solution. IEEE Trans. on Signal Processing, Accepted

[10] Johnson L W, Riess R D, Arnold J T. Introduction to Linear Algebra. 5th ed. New York: Prentice-Hall, 2000

[11] Lay D C. Linear Algebra and Its Applications, 2nd Edition. New York: Addison-Wesley, 2000

[12] Pham D T. Joint approximate diagonalization of positive defnite matrices. SIAM J. on Matrix Anal. and Appl., 2001, 22(4): 1136~1152

[13] Poularikas A D. The Handbook of Formulas and Tables for Signal Processing. New YorK: CRC Press, Springer, IEEE Press, 1999

[14] Berberian S K. Linear Algebra. New York: Oxford University Press, 1992

[15] De Moor B, Golub G H. The restricted singular value decomposition: Properties and Applications. SIAM J Matrix Anal. Appl., 1991, 12: 401~425

[16] Fernando K V, Hammarling S J. A product induced singular value decomposition (PSVD) for two matrices and balanced relation. In: Proc Conference on Linear Algebra in Signals, Systems and Controls, Society for Industrial and Applied Mathematics (SIAM). PA: Philadelphia, 1988, 128~140

[17] Barkwell L, Lancaster P. Overdamped and gyroscopic vibrating systems. Trans. ASME J. Appl. Mech., 1992, 59(1): 176~181

[18] Jain S K, Gunawardena A D. Linear Algebra: An Interactive Approach. Thomson Learning, 2003

[19] Krabill D M. On extension of Wronskian matrices. Bell Amer Math Soc. 1943, 49: 593~601

[20] Lancaster P. Tismenetsky M. The Theory of Matrices with Applications. 2nd ed. New York: Academic, 1985

[21] Silvery S D. Statistical Inference. Penguin books, 1970

[22] De Moor B. Total least squares for affine structured matrices and the noisy realization problem. IEEE Trans. on Signal Processing, 1994, 42: 3104~3113

[23] Searle S R. Matrix Algebra Useful for statistics. New York: John Wiley & Sons, 1982

[24] Davis C, Kahan W M. The rotation of eigenvectors by a perturbation III. SIAM J Nume Anal, 1970, 7: 1~46

[25] Doyle J C. Analysis of feedback systems with structured uncertainties. Proc IEE. 1982, 129: 242~250

[26] Takagi T. On an algebraic problem related to an analytic theorem of Caratheodory and Fejer and on an allied theorem of Landau. Japanese J Math, 1925, 1: 83~93

[27] Lütkepohl H. Handbook of Matrices. New York: John Wiley & Sons, 1996

[28] Zha H. The restricted singular value decomposition of matrix triplets. SIAM J Matrix Anal. Appl., 1991, 12: 172~194

[29] Lancaster P. Quadratic eigenvalue problems. Linear Algebra Appl., 1991, 150: 499~506

[30] Lancaster P. Lambda-Matrices and Vibrating Systems. Oxford: Pergamon Press, 1966

[31] Bellman R. Introduction to Matrix Analysis. 2nd ed. New York: McGraw-Hill, 1970

[32] Alexander S T. Adaptive Signal Processing: Theory and Applications. New York: Springer-Verlag, 1986

[33] Buckley K M, Xu X L. Spatial-spectrum estimation in a location sector. IEEE Trans. on Acoust, Speech, Signal Processing, 1990, 38(11): 1842~1852

[34] Milliken G A, Akdeniz F. A theorem on the difference of the generalized inverses of two nonegative matrices. Communications in Statistics, 1977, A6: 73~79

[35] Kantorovich L V. Function analysis and applied mathematics. Uspekhi Mathematicheskikh Nauk, 1948, 3: 89~185. Translated from Russian by C D Benster, National Bureau of Standards, Report 1509, 7 March 1952

[36] Neudecker H. A matrix trace inequality. J. Mathema tical Analysis and Applications, 1992, 166: 302~303

[37] Tsatsanis M K, Z. Xu. Performance analysis of minimum variance CDMA receivers. IEEE Trans. on Signal Processing, 1998, 46: 3014~3022

[38] Yang J F, Kaveh M. Adaptive eigensubspace algorithms for direction or frequency estimation and tracking. IEEE Trans. on Acoust, Speech, Signal Processing, 1988, 36: 241~251

[39] Saad Y. Numerical Methods for Large Eigenvalue Problems. New York: Machester University Press, 1992

[40] Van Loan C F. Generalizing the singular value decomposition. SIAM J Numer Anal, 1976, 13: 76~83

[41] Meerbergen K. Locking and restarting quadratic eigenvalue solvers. SIAM J Sci Comput, 2001, 22(5): 1814～1839
[42] Komzsik L. Implicit computational solution of generalized quadratic eigenvalue problems. Finite Elemnets in Analysis and Design, 2001, 37: 799～810
[43] Hanchez Y, Dooren P V. Elliptic and hyperbolic quadratic eigenvalue problems and associated distance problems. Linear Algebra and Its Applications, 2003, 371: 31～44
[44] Pease M C. Methods of Matrix Algebra. New York: Academic Press, 1965
[45] Tisseur F, Meerbergen K. Quadratic eigenvalue problem. SIAM Review, 2001, 43(2): 235～286
[46] Jennings A, McKeown J J. Matrix Computations. New York: John Wiley & Sons, 1992
[47] Meyer C. Matrix Analysis and Applied Linear Algebra. PA: SIAM, 2000
[48] 黄琳. 系统与控制理论中的线性代数. 北京：科学出版社，1984
[49] 俞正光，李永乐，詹汉生. 线性代数与解析几何. 北京：清华大学出版社，1998
[50] 张贤达. 现代信号处理. 北京：清华大学出版社，2002
[51] 张贤达，保铮. 通信信号处理. 北京：国防工业出版社，2000
[52] 张贤达. 矩阵分析与应用. 北京：清华大学出版社，2004